Faculty of Engineering and Information Technology
University of Technology, Sydney

Coupled Behavior Informatics: Modeling, Analysis and Learning

A thesis submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

by

Can Wang

October 2013

CERTIFICATE OF AUTHORSHIP/ORIGINALITY

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signature of Candidate

Production Note:
Signature removed prior to publication.

Acknowledgments

First and foremost, I would like to express the deepest appreciation to my supervisor, Prof. Longbing Cao, for providing me with continuous support throughout my PhD study and research. Without his professional guidance and persistent help, this dissertation would not have been possible.

Thanks to my co-workers, Jinjiu Li and Wei Wei, for being so supportive especially when I was struggling through hard times. We three enrolled in the same semester at a close time, through all these years, we shared our joy of achievements and the anxiety of struggling. Luckily, we have got over the hard times and shall step into our new period of life.

I am also grateful to my colleagues Xin Cheng and Chunming Liu, for their hard working in our collaborated papers. Thanks to all other colleagues in AAI, their warm help and insightful suggestions have definitely helped me. Also I would like to thank all the support staffs in AAI.

I would like to in particular thank my boyfriend Zhong She, without his dedicated support, selfless help and endless love, I can not survive and outlast the tough times during my PhD period. I would like to attribute my productive phase, i.e. the last two years, to him without a hesitate.

Last but not the least, I would like to thank my parents. Without their encouragement, finishing this dissertation would be impossible.

Can Wang
June 2013 @ UTS

Contents

List of Figures

List of Tables

Abstract

Behavior refers to the action or property of an actor, entity or otherwise, to situations or stimuli in its environment. The in-depth analysis of behavior has been increasingly recognized as a crucial means for understanding and disclosing interior driving forces and intrinsic cause-effects on business and social applications, including web community analysis, counter-terrorism, fraud detection and customer relationship management, etc. Currently, behavior modeling and analysis have been extensively investigated by researchers in different disciplines, e.g. psychology, economics, mathematics, engineering and information science. From those diverse perspectives, there are widespread and long-standing explorations on behavior studies, such as behavior recognition, reasoning about action, interactive process modeling, multivariate time series analysis, and outlier mining of trading behaviors.

All the above emerging methods however suffer from the following common issues and problems to different extents: (1) Existing behavior modeling approaches have too many styles and forms according to distinct situations, which is troublesome for cross-discipline researchers to follow. (2) Traditional behavior analysis relies on implicit behavior and explicit business appearance, often leading to ineffective and limited understanding on business and social activities. (3) Complex coupling relationships between behaviors are often ignored or only weakly addressed, which fails to provide a complete understanding of the underlying problems and their comprehensive solutions. (4) Current research usually overlooks the checking of behavior interactions, which weakens the soundness and robustness of models built for complex behavior applications. (5) Most of the classic mining and learning algorithms follow the fundamental assumption of independent and identical distribution (i.e. IIDness), but this is too strong to match the reality and complexities in practical applications.

With the deepening and widening of social/business intelligences and their networking, the concept of behavior is in great demand to be consolidated and formalized to deeply scrutinize the native behavior intention, lifecycle and impact on complex problems and business issues. In the real-world applications, group behavior interactions (i.e. coupled behaviors) are widely seen in natural, social and artificial behavior-related problems. The verification of behavior modeling is further desired to assure the reliability and stability. In addition, complex behavior and social applications often exhibit strong explicit or implicit coupling relationships both between their entities and properties. They can not be abstracted or weakened to the extent of satisfying the IIDness assumption. These characteristics greatly challenge the current behavior-related analysis approaches. Moreover, it is also very difficult to model, analyze and check behaviors coupled with one another due to the complexity from data, domain, context and impact perspectives.

Based on the above research limitations and challenges, this thesis reports state-of-the-art advances and our research innovations in modeling, analyzing and learning coupled behaviors, which constitute the coupled behavior informatics. Coupled behaviors are categorized as qualitative coupled behaviors and quantitative coupled behaviors, depending on whether the behavior involved is qualified by actions or quantified by properties.

In terms of the qualitative coupled behavior modeling and analysis, we propose an Ontology-based Qualitative Coupled Behavior Modeling and Checking (*OntoB* for short) system to explicitly represent and verify complex behavior relationships, aggregations and constraints. The effectiveness of *OntoB* system in modeling multi-robot behaviors and their interactions in the Robocup soccer competition game has been demonstrated.

With regard to the quantitative coupled behavior analysis and learning, we carry out explorations on three tasks below. They are under the non-IIDness assumption of entities or properties or both of them, which caters for the intrinsic essence of real-world problems and applications.

For *numerical coupled behavior analysis*, we introduce a framework to address the comprehensive dependency among continuous properties. Substantial experiments show that the coupled representation can effectively model the global couplings of numerical properties and outperforms the traditional way. For *categorical coupled behavior analysis*, we present an efficient data-driven similarity learning approach that generates a coupled property similarity measure for nominal entities. Intensive empirical studies witness that the coupled property similarity can appropriately quantify the intrinsic and global interactions within and between categorical properties for especially large-scale behavior data. For *coupled behavior ensemble learning*, we explicate the couplings between methods and between entities in the application of clustering ensembles, and put forward a framework for coupled clustering ensembles (*CCE*). The *CCE* is experimentally exhibited to capture the implicit relationships of base clusterings and entities with higher clustering accuracy, stability and robustness, compared to existing techniques. All these models and frameworks are supported by statistical analysis.

Finally, we provide a consolidated understanding of coupled behaviors by summarizing the qualitative and quantitative aspects, extract the multi-level couplings embedded in them, and then formalize a coupled behavior algebra at its preliminary stage. Many open research issues and opportunities related to our proposed approaches and this novel algebra are discussed accordingly.

Under varying backgrounds and scenarios, our proposed systems, algorithms and frameworks for the coupled behavior informatics are evidenced to outperform state-of-the-art methods via theoretical analysis or empirical studies or both of them. All these outcomes have been accepted by top conferences, and the follow-up work has also been recognized. Therefore, coupled behavior informatics is a promising though wholly new research topic with lots of attractive opportunities for further exploration and development.

Chapter 1

Introduction

1.1 Background

At first, two basic concepts throughout this thesis are introduced: behavior and coupled behaviors. We then present the modeling, analysis and learning of coupled behaviors, which form the coupled behavior informatics.

1.1.1 Behavior

Behavior is an important concept and a key component in the scientific, societal, economic, cultural, political, military, living and virtual world. In Wikipedia, "behavior" is the range of actions and mannerisms made by organisms, systems or artificial entities in conjunction with their environment, which includes the other systems or organisms around as well as the physical environment. It is the response of system or organism to various stimuli or inputs, whether internal or external, conscious or subconscious, overt or covert, and voluntary or involuntary[1]. In dictionaries, "behavior" refers to manner of behaving or acting, and the action or reaction of any material under given circumstances (Cao & Philip 2012).

Generally, behavior is the action, reaction or property of an entity, hu-

[1] http://en.wikipedia.org/wiki/Behavior

man or otherwise, to situations or stimuli in its environment (Cao 2010). It is ubiquitous and can be widely seen anywhere at any time in any form. In different applications and scenarios, behaviors exhibit respective characteristics and features. A qualitative behavior consists of two important ingredients: actor and action, while a quantitative behavior is characterized by entity and property. For instance, from the qualitative perspective, the "buy" order in a stock market and the "kick of a goal" action in robot soccer game are behaviors conducted by respective actors: investors and robots. From the quantitative aspect, the records or observations represented by rows of an information table can also be regarded as behaviors described by a range of properties or variables. For example, a plant object depicted by its lengths and widths of sepal and petal is regarded as a quantitative behavior exhibiting the plant entity, a movie object featured by its director, actor and genre is also treated as a quantitative behavior that manifests the movie entity. However, these two entities differ from each other in terms of the property type. The former (i.e. plant entity) is described by numerical properties, while the latter (i.e. movie entity) is expressed by categorical properties. Thus, the quantitative behavior can be further divided into numerical behavior and discrete behavior, which correspond to the plant entity and the movie entity, respectively.

To obtain an integrated understanding, actor and entity can be unified as the body of behavior, action and property are consolidated as the depictor of behavior for they appear in different forms to describe the body. Accordingly, the above mentioned plant entity and movie entity are the body, and the relevant actions and properties are the depictors. From a broad perception, the actions/operations, properties, responses or presentations associated with the corresponding actors and entities form a concrete, substantial and rich concept: behavior (Cao & Yu 2009).

With the fast development and deep engagement of social and digitalized life by means of advanced computing technology, in particular, virtual reality, multimedia information processing, visualization, machine learning and pattern recognition techniques, behaviors in the virtual and social world are emerging increasingly. Additionally, behaviors in the traditional spheres and environment are becoming more and more complex with the involvement of both the virtual and social world. They are widely seen on the internet, social and online networks, multi-agent systems, brain systems, and information database.

The in-depth understanding of complex behaviors has been increasingly recognized as a crucial means for disclosing underlying working mechanism, interior driving forces, causes and impact, dynamic and evolution on businesses as well as social systems in handling many challenging issues, such as intrusion detection (Vigna, Valeur & Kemmerer 2003), social computing (Wang, Carley, Zeng & Mao 2007), fraud detection (Fast, Friedland, Maier, Taylor, Jensen, Goldberg & Komoroske 2007), event analysis (Weiss & Hirsh 1998), outlier detection (Hodge & Austin 2004), and group decision-making (Cao, Ou & Yu 2012), etc. This forms the need and emergence of behavior informatics, i.e. understand behaviors from computing and information perspective (Cao 2010), which has been recently studied to "formalize", "quantify", "compute", and "learn" complex behavioral applications and social communities.

1.1.2 Coupled Behaviors

As indicated above, behavior is an essential and critical activity which has been increasingly investigated in diverse fields, from social and behavioral sciences to computer science (Pierce & Cheney 2004, Zacharias & MacMillan 2008, Liu, Salerno & Young 2008, Getoor & Taskar 2007). Although there is an emerging focus on deep behavior studies, such as social network analysis (Hogg & Szabo 2008), periodic behavior analysis (Cao, Mamoulis & Cheung 2007) and behavior informatics approach (Cao 2010), previous research work has mainly focused on individual behaviors without considering the interactions of them. However, with increasing network and community-based events as well as their applications, e.g. group-based crime and social network interactions, coupling relationships between behaviors contribute to the intrinsic causes and impacts of eventual business and social problems.

In addition, most of the current theories and systems in statistics, data mining and machine learning are built on the IIDness[2] hypothesis, assuming the independent and identical distribution in the underlying objects, properties and/or values of behaviors. This works perfectly well in abstract problems and simplified business applications with weakened and avoidable interactions and heterogeneity, and acts as the foundation of classic mining and learning algorithms. Nevertheless, complex behavioral and social applications often exhibit strong coupling relationships either explicitly or implicitly, which are beyond the usual dependency or relationship. They also embody the heterogeneity between objects, object properties and attribute values of behaviors, which can not be weakened or abstracted to the extent of satisfying the IIDness assumption.

Couplings may be presented in different forms between actors or actions of behaviors and distinct levels between objects, properties and/or values of behaviors. Heterogeneity is also embodied through multiple or mixed structures and distributions within or between behaviors as well as their properties. This makes it necessary and unavoidable to consider the coupling and heterogeneity in behavior and social informatics. That is to say, coupling relationship analysis inevitably emerges as a crucial issue while it even has not been much studied or realized in the knowledge representation, artificial intelligence, statistics, data mining and machine learning communities.

In both natural and social sciences with their applications, accordingly, behaviors from one or multiple actors often interact with one another, which are called coupled behaviors. In other words, coupled behaviors refer to the activities of one to many actors who are associated with one another in terms of certain relationships. They play a much more fundamental role than individuals in the driving cause, dynamics and effect of business problems, enterprise applications and social communities (Park & Chang 2009, Liu

[2] Refer to http://en.wikipedia.org/wiki/Iid

et al. 2008). Effective approaches for analyzing coupled behaviors are not available, since existing methods mainly focus on individual behavior analysis or follow the IIDness assumption (Cao 2010).

While very limited research outcomes can be identified in the literature, coupled behaviors are widely observed in terms of typical applications (Cao et al. 2012) such as group-based criminal behaviors, cross-reference citation analysis, cross-market manipulation, car transport system, social network interactions, intrusion detection, and multi-agent systems. Below, we illustrate coupled behaviors and behavior interactions with two examples respectively from qualitative and quantitative perspectives.

In this thesis, qualitative coupled behaviors are the qualitative behaviors by additionally involving the coupling relationships in any form, thus composed by three important elements: actor, action and coupling. Similarly, quantitative coupled behaviors correspond to the quantitative behaviors together with their diverse interactions, and consisting of entity, property and coupling. Accordingly, the qualitative coupled behaviors mainly focus on reasoning about the temporal and inferential interactions of behaviors conducted by the same or different actors, and the quantitative coupled behaviors mostly address the quantified correlations within or between specific properties and involved entities of coupled behaviors. Following the same vein in Section 1.1.1, the quantitative coupled behaviors are further partitioned as numerical coupled behaviors and categorical coupled behaviors depending on the type of associated properties[3].

Example 1.1.1 (Qualitative Aspect: Multi-robot Soccer Game)

As shown in Figure 1.1, two teams participate in a Robocup soccer competition with four Sony AIBO robots in each group. Each robot is treated as an actor of the qualitative behavior, the actions of each robot are regarded as operations of the qualitative behavior. The robot players operate on their

[3] In this thesis, we only work on the coupled behaviors quantified by purely numerical properties and purely categorical properties. We will address the mix-type quantitative coupled behaviors in the future work.

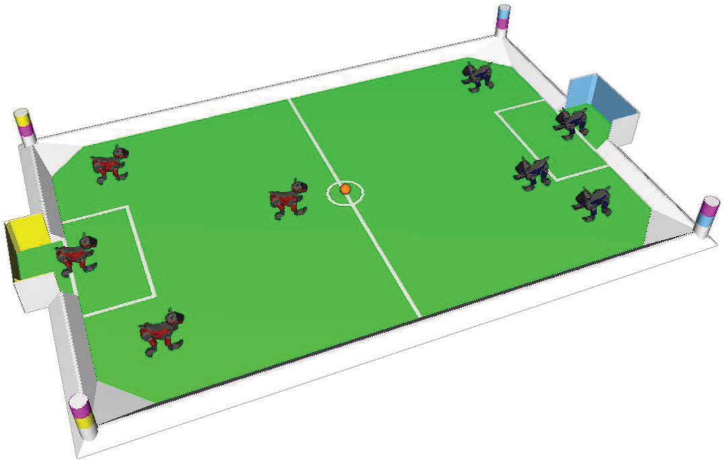
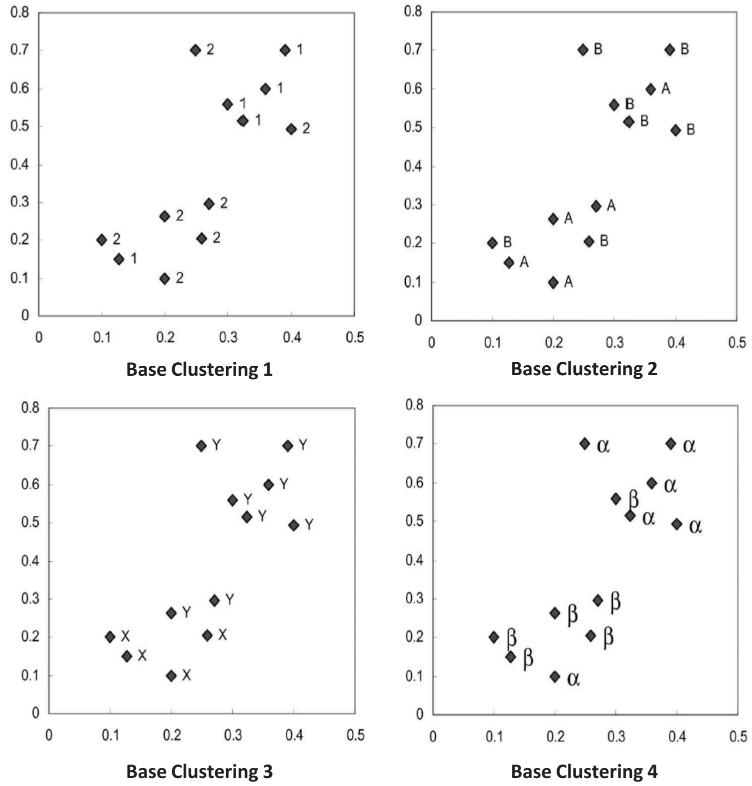


Figure 1.1: Qualitative coupled behaviors: Robocup soccer competition.

own without any external control, either by humans or by computers. They communicate with one another by wireless or by using the speakers and microphones. Their couplings or interactions include the collaborations between different actions of the same robot, e.g., one of the robots kicks the ball after it gets a message (interaction I); and distinct operations conducted by different robots, such as sending messages between different players (interaction II). If a robot undertakes tasks without appropriate arrangement and coordination with other robots, the Robocup is likely to be unsuccessful, even though every robot performs perfectly well.

This example shows that group actors and behaviors by the same or different actors within the team are often coupled in different forms of interactions (Cao et al. 2012), such as serial coupling reflected by interaction I and causal coupling exemplified by interaction II. The whole multi-robot system is an instance of the qualitative coupled behaviors.

In this multi-robot soccer game, there are several important issues to consider, namely, how to visually and formally represent such qualitative interactions? How to aggregate and reason about those group actors and related actions? How to verify and refine the group behavior model to guarantee a



Base Clustering 1

Base Clustering 2

Base Clustering 3

Base Clustering 4

Figure 1.2: Quantitative coupled behaviors: clustering ensemble.

stable and robust system? We will propose an Ontology-based Qualitative Coupled Behavior Modeling and Checking to solve all these issues in Chapter 3.

Example 1.1.2 (Quantitative Aspect: Clustering Ensemble)

Figure 1.2 (Topchy, Jain & Punch 2005) shows four possible base clusterings of 12 data objects into two clusters. Different partitions use different sets of labels. The target of clustering ensemble is to obtain a final clustering based on these four base clusterings.

Here, we regard each object or observation as an entity of the quantitative behavior. Each base clustering is treated as a property of the quantitative behavior, and the clustering result of each base clustering is the correspond-

ing property value or attribute value of the quantitative behavior. The base clusterings are expected to have interactions with one another, such as the co-occurrence of their cluster labels over the same set of objects, since they are all conducted on the same data objects. This kind of interactions embodies the coupling relationships between the properties of behaviors, which is one coupling aspect of the quantitative coupled behaviors. In addition, each object has the neighborhood records as its environment. Thus, how this neighborhood impacts the clustering performance reflects another coupling aspect of the quantitative coupled behaviors in terms of the interactions among objects. Both aspects deliver an example of the quantitative coupled behaviors. In particular, this example illustrates an application of the categorical coupled behaviors, since the associated property of base clustering is in essence a categorical attribute with cluster labels to be its values.

For this clustering ensemble problem, Chapter 6 will introduce a coupled framework of clustering ensemble to formalize and learn both the coupling relationships between base clusterings (i.e, properties) and between data objects (i.e. entities), which are based on non-IIDness assumption. Note that the "IID" in this thesis is short for the independent and identical distribution, so the non-IIDness suggests the dependent (or relational) and heterogenous distribution in behavior data.

The above two instances provide an intuitive picture about what kind of behaviors we are interested in and how we regard the qualitative coupled behaviors as well as quantitative coupled behaviors. In the following Chapter 3 and Chapter 6, we will model and analyze these two examples in detail. In reality, there exist a diversity of either qualitative or quantitative coupled behaviors. Another two case studies are given in Chapter 4 and Chapter 5 to show how we address the coupling relationships between behaviors.

Chapter 3, Chapter 4, Chapter 5 and Chapter 6 explore the qualitative coupled behaviors and quantitative coupled behaviors individually. Accordingly, in Chapter 7, an integrated understanding is proposed that coupled behaviors are composed of four elements: body, depictor, consolidated coupling and context. The body is a unified concept for actor and entity. The depictor is a general notion that embodies the action/operation, property/attribute and all other characteristics of body. Stated differently, the operation, the numerical property, the categorical property to be discussed in this thesis are the concrete manifestations of depictor under different scenarios. The consolidated coupling is teased out by analyzing either the explicit or implicit interactions and relationships among a collection of behaviors, and it also shows the hierarchy and diversity of coupling relationships. The context specifies the environment and styles in which the couplings are considered, such as what sorts of couplings are examined at what level including body-body interaction and depictor-depictor interaction, etc.

1.1.3 Modeling, Analysis and Learning

Due to the emerging popularity and importance of coupled behaviors, the representation, modeling, analysis, mining and learning, and determination of coupled behaviors are becoming increasingly useful, essential and challenging in ubiquitous behavioral applications and problem-solving techniques. They inevitably and undoubtedly constitute new computing opportunities, necessity and technology innovations, we refer to them as coupled behavior informatics, which is an important branch of behavior computing and analytics (Cao & Philip 2012).

Coupled behavior informatics consists of methodologies, techniques and practical tools for exploring human, organizational, artificial and virtual, qualitative and quantitative behaviors, their interactions and relationships, the formation and decomposition of behavior-oriented groups, and collective intelligence. Such observations and discussions motivate this research thesis *Coupled Behavior Informatics: Modeling, Analysis and Learning*. This thesis reports state-of-the-art advances and our research innovations in representing, modeling, analyzing and learning coupled behaviors from both the qualitative and quantitative aspects.

(1) **Coupled Behavior Modeling**: refers to develop representation and modeling mechanisms, languages and tools to capture behavior characteristics, intrinsic and contextual properties of behaviors, behavior dynamics, and internal and external communications among behaviors. Those techniques and methods can also be used to understand interaction, causality, convergence, divergence, selection, decision, evolution, emergence, and intelligence of behavior entities, behavior properties, behavior networks, and behavior impact. Both formal and visual specifications are discussed to represent coupled behaviors and behavior interactions. Case studies are demonstrated to model complex coupled behaviors in a multi-agent system, shown in Figure 1.1.

(2) **Coupled Behavior Analysis**: denote proposing effective methods, techniques and tools for emergent areas and domains in analyzing coupled behaviors and their properties. Model checking technique is utilized to verify the coupled qualitative behavior model with desired requirements, and to further refine the model. Coupled similarities are also introduced to characterize the quantitative behavior interactions in terms of coupling relationships among between properties (i.e. attributes, features, variables) and/or entities (i.e. objects, records, observations). Algorithms and case studies are discussed to analyze behaviors correlated with one another based on mixed properties and complex coupled interactions. The analytical results will be used for detection, prediction, intervention, and grouping of coupled behaviors as well as their interactive relationships.

(3) **Coupled Behavior Learning**: to identify clusters and patterns among the quantitative behavior entities and networks, such as partition and classify a group of coupled behaviors. Algorithms and case studies are proposed to induce the non-IIDness learning by teasing out the attributes coupling, objects coupling, and clusters coupling under the scenario of numerical and categorical entity clustering, classification

and ensemble learning tasks. One of the examples, which is clustering ensemble, is exhibited in Figure 1.2. We aim to show that the non-IIDness issues are manageable and the involvement of non-IIDness can result in substantially improved outcomes. These strategies can be widely used and expanded for analyzing and learning complex behavioral and social problems.

All the above three aspects compose what we call coupled behavior informatics, which contributes to the in-depth understanding, discovery, applications, and management of coupled behavior intelligence. The coupled behavior analysis with applications (Cao et al. 2012) and the overview of non-IIDness learning (Cao 2013) proposed by Cao et al. are the instances of preliminary investigation for this promising research topic. This thesis creates an important opportunity and detailed explorations to broaden current research to areas that consist of coupled behaviors. It aims to serve as the first dedicated source of references for the theory and applications of coupled behavior computing, establishing current research work, disseminating our latest research discoveries, and providing a ground-breaking textbook to researchers with interest in this field.

1.2 Limitations and Challenges

Behaviors can be seen everywhere in business and social life. There exist a diversity of ways to explore and investigate behaviors from multiple disciplines and areas covering psychology, economics, mathematics, engineering and information science.

On one hand, many qualitative approaches have been proposed to model and analyze behaviors, such as belief-desire-intention model (Wooldridge 2000), situation calculus (Giacomo, Lesperance & Pearce 2010), human-machine interaction (Kobsa 2001), reasoning about action (Gu & Soutchanski 2007), behavior recognition (Gabaldon 2009) and simulation (Subramanian 2010). On the other hand, a lot of quantitative methods including sequence analysis (Ayres, Flannick, Gehrke & Yiu 2002), activity monitoring (Fawcett & Provost 1999) and mining (Cao, Zhao, Zhang & Zhang 2008), customer behavior analysis (Dasgupta, Singh, Viswanathan, Chakraborty, Mukherjea, Nanavati & Joshi 2008) and web user behavior analysis (Flesca, Greco, Tagarelli & Zumpano 2005) have been introduced to facilitate the behavior computing. In addition, ontological engineering and semantic web (Breitman, Casanova & Truszkowski 2007), representation and reasoning (Brachman & Levesque 2004) and model checking (Baier & Joost 2008) are also helpful techniques and theories to enable the in-depth behavior analysis.

Despite such great progress and development, limited efforts have been made in deep modeling and analysis of coupled behaviors. Several slightly relevant areas are coupled Hidden Markov model (Oliver & Pentland 2000), multivariate time series based approaches (Yoon, Yang & Shahabi 2005) and social network analysis (Hogg & Szabo 2008). However, demographic data rather than the genuine behavior structure is the main focus in those methods. Apart from this, these research suffers from the lack of explicit behavior representation and the stability verification of their models. Limited work can be identified on formalizing and checking complex behavior structures and interactions. For example, based on the multi-robot soccer game shown in Figure 1.1, though Ros and Veloso (Ros & Veloso 2007) designed two kinds of evaluations to show the superiority of their case-based coordination mechanism, they neither explicitly represent the interactions among robots nor verify the stability of that system, which is not convincing enough.

In addition, most of the classic theories, algorithms, systems and tools in artificial intelligence, statistics, data mining and machine learning are built based on the fundamental assumption of IIDness, which believes the independence and identical distribution between underlying objects (i.e. entities), attributes (i.e. properties) and/or values. For instance, the traditional *k-means* or *k-modes* algorithm to perform clustering and the classical *KNN* algorithm to conduct classification (Gan, Ma & Wu 2007). Nevertheless, increasingly complex behavioral applications and social problems often exhibit strong couplings and heterogeneity between objects, attributes and values (i.e., non-IIDness), explicitly or implicitly. This fundamentally challenges the IIDness-based learning methodologies and techniques. For the example of clustering ensemble in Figure 1.2, traditional clustering ensemble methods such as *CSPA* and *MCLA* (Strehl & Ghosh 2002), which lack the consideration of any coupling relationship, usually randomly distribute several controversial objects in either an identical cluster or different groups. However, those objects can actually be correctly allocated if non-IIDness assumption is followed.

Detailed introductions and evaluations of the related work are given in Chapter 2. Below, we summarize and list the main limitations and challenges of current research work on behavior computing.

- Existing behavior modeling approaches have too many styles and forms according to distinct situations. There is very limited research on formalizing the concept of behavior and its elements, which is too weak to reveal that behavior plays the key role of an internal driving force for social and business activities. Additionally, it is ineffective or even impossible to deeply tease out native behavior intention and impact on complex issues and business problems. There are no formal behavior representation models stated from a general perspective and providing a comprehensive understanding of behavior constitution.

- Traditional behavior analysis is usually built on customer demographics and business usage related transactions directly. It mainly relies on implicit behavior and explicit business appearance from behavioral and social sciences, often leading to ineffective and limited analysis in understanding business and social activities deeply and accurately. With behavior implied in demographic and transactional data, it is not possible to support in-depth analysis on behavior interior surrounded by behavioral elements, but on behavior exterior such as service usage. The behavior implication in transactional data also determines that it fails to to scrutinize behavioral intention and impact on business appearance and applications.

- State-of-the-art research work is in lack of explicitly modeling and analyzing complex interactions of group behaviors directly. Complex coupling relationships between behaviors are often ignored or only weakly addressed. However, behaviors are often observed to be correlated in terms of certain coupling relationships, for instance, serial or parallel, conjunction or disjunction. Such coupling relationships greatly challenge existing behavior representation methods, since they involve multiple behaviors from different actors, constraints on the interactions and behavior evolution, which are often not obvious and exhibit large complexities. However, a deep exploration of interactive relationships is necessary for us to understand how behaviors are correlated and how those coupled behaviors drive and impact business and social problems.

- Current research often overlooks the checking of behavior modeling, which weakens the soundness and robustness of models built for complex behavior applications. The quality of behavior interactions are not checked through verification techniques. Little related work is ready for the formalization and verification of coupled behaviors, including elaborating and representing behavioral elements, specifying behavior interactive relationships, and checking the modeling of multiple behavior couplings. The engagement of verification in behavior analysis may make the findings much more stable and robust for problem-solving.

- Most of the existing mining and learning algorithms assume that behavior entities (i.e. objects) and their properties (i.e. attributes) follow the independence and identical distribution (i.e. IIDness), which means the involved observations and variables do not have any connections among one another. They often overlook or abstract the couplings and heterogeneity, by taking this strong hypothesis. However, the assumption and abstraction taken in IIDness learning techniques are too

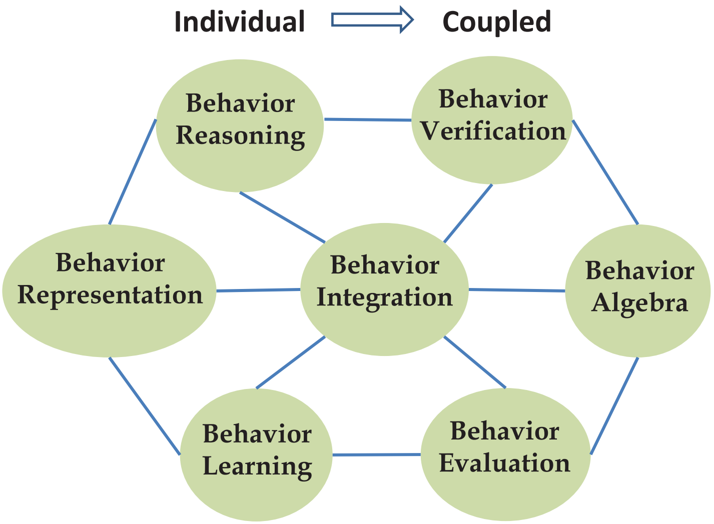


Figure 1.3: Research issues.

strong and seriously mismatch the reality and complexities in behavioral/social systems and applications. For example, in social media, users are more or less inter-related or inter-influenced by one another in terms of various aspects and reasons. Thus, the great demand of catering for heterogeneity and couplings urge the development of non-IIDness learning strategies.

This thesis aims to break through and overcome those limitations, introduce new models and novel frameworks to explicitly formalize the concept: coupled behaviors, analyze their contextual properties, and learn the intrinsic non-IID data structure and characteristics.

1.3 Research Issues and Objectives

Based on the aforementioned research limitations and challenges, we propose several research issues and objectives from the following aspects, which are also shown in Figure 1.3.

– **Coupled Behavior Representation**: or called coupled behavior modeling, is to develop behavior-oriented specifications and formalizations to describe coupled behaviors (i.e., behaviors from either the same, or different actors are often coupled with each other) and the relationships among them. It provides a unified and formalized mechanism for describing, presenting and aggregating behavior interactions, desired requirements or properties, behavior impact and patterns.

- **Coupled Behavior Reasoning and Verification**: With the formal representation of coupled behaviors, the qualitative analytics addresses the task of behavior reasoning and verification, which are used to check and verify complex behavioral elements, relationships, aggregations, properties and constraints. It accordingly refines sensitive and problematic model proposals, and then guarantees the robustness and stability of coupled behavior representation schemes.

- **Coupled Behavior Learning and Evaluation**: Inspired by the non-IIDness hypothesis, the quantitative research targets behavior learning and evaluation via exposing the coupling relationships between behavioral objects, between behavioral variables, and/or between behavioral attribute values. It can be applied in data mining and machine learning algorithms, such as clustering, classification and ensemble learning. Those methodologies and techniques are then evaluated by accuracy, mutual information, stability and statistical significance to exhibit whether the non-IIDness based learning is helpful and effective or not.

- **Coupled Behavior Algebra and Integration**: To make the qualitative coupled behaviors and quantitative coupled behaviors as one general concept, an appropriate way must be chosen to integrate both the qualitative reasoning and verification with the quantitative learning and evaluation to obtain a comprehensive understanding of the implicit complex coupled behaviors. A coupled behavior algebra is in great demand to be proposed, formalized and generalized.

In the following chapters, we aim to clarify and solve all the above research issues. Chapter 3 focuses on coupled behavior representation, reasoning and verification. Chapter 4, Chapter 5 and Chapter 6 addresses coupled behavior learning and evaluation. Chapter 7 lays particular emphasis on coupled behavior algebra and integration.

1.4 Research Contributions

In this thesis, we mainly work on the coupled behavior informatics in terms of modeling, analysis and learning on coupled behaviors. The proposed coupled behavior informatics covers our accepted or published research work listed in Appendix A as follows.

Formalization and verification of coupled behaviors are introduced from the qualitative perspective (Wang & Cao 2012). Coupled numerical attributes are analyzed for continuous quantitative coupled behaviors (Wang, She & Cao 2013*a*). Coupled categorical attributes are also explored for discrete quantitative coupled behaviors (Wang, Cao, Wang, Li, Wei & Ou 2011). A coupled framework of clustering ensembles is designed for applications on quantitative coupled behaviors (Wang, She & Cao 2013*b*). A coupled discretization algorithm (Wang, Wang, She & Cao 2012), a coupled recommendation system (Yu, Wang, Gao, Cao & Chen 2013) and a coupled document clustering approach (Cheng, Miao, Wang & Cao 2013) are proposed as well.

This thesis focuses on the first four research topics mentioned above (i.e. formalization and verification of coupled behaviors, coupled numerical attributes analysis, coupled categorical attributes analysis, and coupled framework of clustering ensembles), and the last three research issues (i.e. coupled discretization algorithm, coupled recommendation system, and coupled document clustering approach) are solved based on the first four topics. The contributions of such tasks are listed individually as below. Note that object and the entity of coupled behaviors are interchangeable, attributes and base clusterings are equivalent to the properties of coupled behaviors.

1.4.1 Coupled Behavior Formalization and Verification

We build an Ontology-based Qualitative Coupled Behavior Modeling and Checking (*OntoB* for short) for representing and verifying complex behavior relationships and interactions. With knowledge representation techniques, the *OntoB* system offers the following characteristics and capabilities:

(1) *Systematic*: The built-in behavior ontology combines the features of entity-relationship with theoretical semantics to explicitly capture behavior interactions, various intra-couplings (interactions between behaviors from the same actor) and inter-couplings (interactions between behaviors from different actors) between distinct behaviors as well as their aggregation and behavior constraints, in terms of behavioral perspective rather than transactional aspect.

(2) *Solid*: The inclusion of model checking (Baier & Joost 2008) in the *OntoB* system makes reliable models, and outperforms the manual proof and test with simulations in terms of nondeterminism and automation;

(3) *Generic*: The building blocks in the *OntoB* system are generic and can be used for modeling behavior-oriented applications with a variety of coupling relationships; and

(4) *Flexible*: The semantic mappings of intra-coupled and inter-coupled syntax in *OntoB* are flexible according to specific requirements when interpreted as the corresponding aggregations, which means the way we provide here for verification is only an alternative option.

We introduce the *OntoB* system, and illustrate it by modeling and verifying behaviors of all robots as well as their interactions from temporal, inferential and party-based aspects in the multi-robot soccer game, shown in Figure 1.1.

1.4.2 Numerical Coupled Behavior Analysis

We propose a framework of the coupled attribute analysis on numerical quantitative coupled behaviors, in which the continuous properties of coupled behaviors are assumed to follow non-IIDness. The key contributions of this part are listed as follows:

(1) We consider both the intra-coupled interaction within an attribute, captured by the correlations between every attribute and its own powers; and the inter-coupled interaction between different attributes, quantified by the correlations between each attribute and the powers of others.

(2) A coupled representation scheme is introduced for quantitative objects to integrate the intra-coupled and inter-coupled interactions with the original information table representation via Taylor-like expansion in a global way.

(3) The proposed coupled representation method is compared with the traditional representation approach by applying data structure analysis, clustering and classification, revealing that the couplings of continuous attributes are essential to the learning applications.

1.4.3 Categorical Coupled Behavior Analysis

We explicitly discuss the data-driven intra-coupled similarity and inter-coupled similarity, as well as their global aggregation in unsupervised learning on nominal quantitative coupled behaviors. Here, the categorical properties of coupled behaviors are under the assumption of non-IIDness. The key contributions of this part are listed in the following:

(1) We propose a Coupled Attribute Similarity for Objects ($CASO$) measure based on the Coupled Attribute Similarity for Values ($CASV$) measure, by considering both the Intra-coupled and Inter-coupled Attribute Value Similarities ($IaASV$ and $IeASV$), which capture the attribute value frequency distribution and the attribute dependency aggregation respectively with a high accuracy and relatively low complexity in a global way.

(2) We compare the accuracy and efficiency of the four proposed measures for *IeASV* in terms of four relationships: power set, universal set, join set and intersection set; and obtain the most efficient candidate based on the intersection set (i.e. *IRSI*) from theoretical and experimental aspects.

(3) A method is proposed to flexibly define the dissimilarity metrics with the proposed similarity building blocks according to specific requirements.

(4) The proposed measures are compared with the state-of-the-art metrics on a range of benchmark categorical data sets in terms of the internal and external clustering criteria, as well as the classification accuracy. All the results are statistically significant.

(5) Two new coupled categorical clustering algorithms, which are *CROCK* and *CLIMBO*, are accordingly proposed and verified based on the existing algorithms *ROCK* and *LIMBO*.

1.4.4 Coupled Behavior Ensemble Learning

We propose an effective framework for coupled clustering ensembles (*CCE*) as an application on quantitative coupled behaviors by involving non-IIDness to uncover the intrinsic coupling relationships between base clusterings (i.e. properties) and between objects (i.e. entities). The key contributions of this part are as follows:

(1) The non-IIDness nature is described from the perspectives of clustering-based, object-based, and cluster-based algorithms, and reveals that the couplings are essential to the clustering ensemble.

(2) Both the couplings between base clusterings and between data objects are considered in a coupled framework *CCE* of clustering ensembles to support an integrated coupling.

(3) We propose several similarity measures that incorporate the couplings of base clusterings and objects, and they exhibit an impressive ability to capture the implicit relationships within the data.

(4) Our proposed framework *CCE* is evaluated against the eight existing clustering ensemble methods and two categorical clustering algorithms on a variety of benchmark data sets in terms of accuracy, stability, robustness, and statistical significance.

(5) In addition, we empirically explore the relationship between the data characteristics of base clusterings and the degree of improvement in the final clustering quality.

In this part, the aforementioned Figure 1.2 is used to illustrate how the *CCE* framework is applied to conduct clustering ensembles.

1.5 Thesis Organization

The profile and structure of research work in this thesis are exhibited in Figure 1.4. The thesis is organized as follows.

Chapter 1 provides an introduction to coupled behavior informatics in terms of modeling, analysis and learning. It describes the research background and motivation, current limitations and challenges, research issues and objectives, as well as key contributions.

Chapter 2 reviews the related literatures on qualitative behavior modeling and analysis, numerical attribute analysis of quantitative behaviors, categorical attribute analysis of quantitative behaviors, and clustering ensemble of quantitative behaviors individually.

Chapter 3 proposes an ontology based behavior modeling and checking system to explicitly represent and verify complex behavior relationships, aggregations and constraints. It mainly focuses on the modeling and analysis of qualitative coupled behaviors.

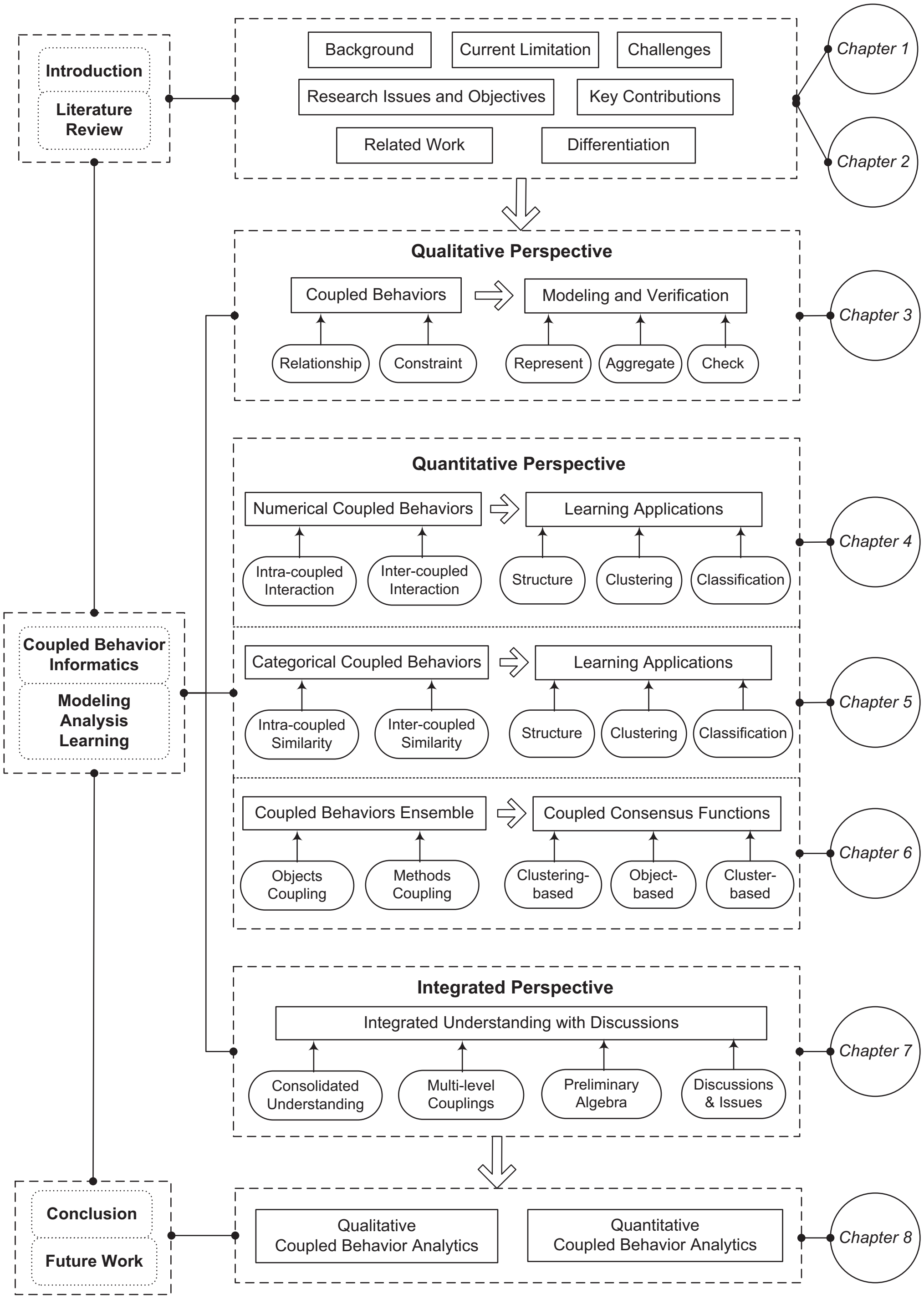


Figure 1.4: The profile of research work in this thesis.

Chapter 4 introduces a numerical coupled behavior analysis framework to capture the global dependency or non-IIDness of continuous properties. Chapter 5 puts forward a coupled attribute similarity measure for nominal entities with coupling relationships or non-IIDness between categorical properties. Chapter 6 discusses the problem of explicating the non-IIDness between base clusterings and between objects in clustering ensembles, which is an application on quantitative coupled behaviors. All these chapters address the analysis and learning of quantitative coupled behaviors.

Chapter 7 integrates the modeling and analysis of qualitative coupled behaviors with the analysis and learning of quantitative coupled behaviors to obtain a unified and comprehensive understanding of complex coupled behaviors with multi-level couplings therein, abstracts a preliminary coupled behavior algebra, and discusses the proposed models, methods and frameworks in this thesis with open research issues.

Chapter 8 concludes the thesis and outlines the scope for future work.

Appendix A shows a list of publications during my PhD studies. Appendix B lists the main denotations in this thesis.

Chapter 2

Literature Review

In this chapter, the related work and literatures are reviewed in terms of behavior modeling and analysis, numerical behavior analysis, categorical behavior analysis, and a behavior learning application on clustering ensembles. These four parts overview the related research work in Chapter 3, Chapter 4, Chapter 5 and Chapter 6, respectively.

The first part on behavior modeling and analysis mainly focuses on state-of-the-art representation schemes and analysis strategies for behaviors in general. The second part on numerical behavior analysis, in particular, addresses the current research on the continuous properties (i.e. attributes) of a collection of quantitative entities (i.e. objects); while the followed part on categorical behavior analysis evaluates the existing work on the discrete properties of entities. Further, the last part on behavior application explicates the research progress in terms of clustering ensemble learning, in which each object is regarded as the entity of behavior, and each base clustering is treated as the property of behavior. Figure 2.1 shows the framework of the literature review for behavior studies. The last column in Figure 2.1 lists the limitations of current work. The detailed reviews and evaluations of the related research work are specified as follows.

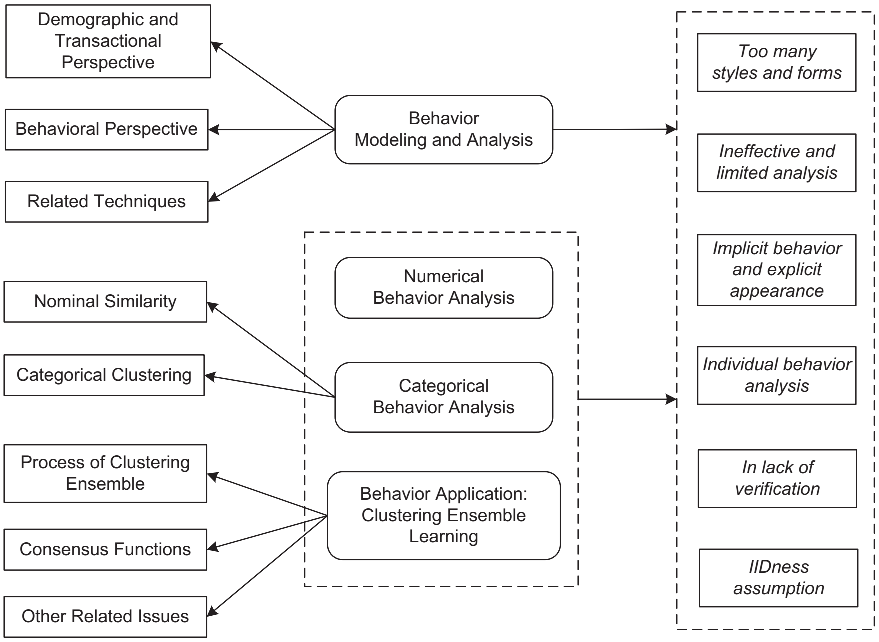

Demographic and Transactional Perspective
Behavioral Perspective
Related Techniques
Behavior Modeling and Analysis
Nominal Similarity
Categorical Clustering
Process of Clustering Ensemble
Consensus Functions
Other Related Issues
Numerical Behavior Analysis
Categorical Behavior Analysis
Behavior Application: Clustering Ensemble Learning
Too many styles and forms
Ineffective and limited analysis
Implicit behavior and explicit appearance
Individual behavior analysis
In lack of verification
IIDness assumption

Figure 2.1: The framework of literature review.

2.1 Behavior Modeling and Analysis

Behavior is ubiquitous in business and social life. Behavior modeling and analysis have been extensively studied by researchers in psychology, economics (Holcombe 1989), mathematics, engineering (Zacharias & MacMillan 2008) and information science (Wilson & Walsh 1997) among others.

Typical concepts have been proposed to model and analyze a range of behaviors, including activity monitoring (Fawcett & Provost 1999), customer behavior analysis (Dasgupta et al. 2008), user modeling (Kobsa 2001), web user behavior patterns (Flesca et al. 2005, Kwan, Fong & Wong 2005), action reasoning and composition (Gu & Soutchanski 2007), belief-desire-intention model (Wooldridge 2000), situation calculus (Giacomo et al. 2010), and behavior compositions (Sardina, Patrizi, Giacomo & Universita 2008). Additional techniques and theories such as ontological engineering and semantic web (Breitman et al. 2007, Razmerita 2011), representation and reasoning

(Brachman & Levesque 2004), reality mining (Eagle, Pentland & Lazer 2008), sequence analysis (Zaki 2001, Ayres et al. 2002) and model checking (Baier & Joost 2008) are also helpful for the behavior simulation and computing.

In the sections below, these work on behavior modeling and analysis is introduced, sorted and evaluated from different perspectives.

2.1.1 Demographic and Transactional Perspective

Most of the existing behavior analysis approaches have been directly conducted on customer demographic and transactional data (Cao et al. 2012, Wang & Cao 2010), in which behavior-oriented elements are hidden in routinely collected information. For example, in stock markets, transactions mainly record and manage the price, volume and index related information. For this kind of data, behavior is hidden and its properties are separately kept. In the following, we enumerate the behavior modeling and analysis methods from the demographic and transactional perspective.

In outlier mining of trading behaviors, price movement is usually addressed to detect abnormal trading (Donoho 2004*a*, Yamanishi & Takeuchi 2002). In capital markets, relational data associated with brokers and securities as well as the corresponding price information is used to analyze security and capture fraud (Donoho 2004*b*, Fast et al. 2007). In telecom churn analysis, service subscriber demographics and their usage, billing, credit, application and complaint history are analyzed to classify customers based on the dynamics of usage change (Mozer, Wolniewicz, Grimes, Johnson & Kaushanky 2000, Nanavati, Gurumurthy, Das, Chakraborty, Dasgupta, Mukherjea & Joshi 2006, Dasgupta et al. 2008). In human-computer interaction (Kobsa 2001), user modeling aims at designing cognitive models of human users, including modeling their skills with declarative knowledge and performing user testing. In web usage and preference analysis (Srivastava, Cooley, Deshpande & Tan 2000, Flesca et al. 2005), online access and session information for navigational location, history and experience is mainly adopted to simulate and model web user behaviors. In online customer behavior analysis (Kwan et al. 2005), product awareness and exploration of purchase commitment are investigated to discover customer patterns.

These approaches are only capable of demographic and transactional modeling, analysis and learning. In scrutinizing the above examples, we realize that the so-called behavior analysis is actually not based on intrinsic and genuine behavioral elements, but on straightforward customer demographic data and in particular on business usage related transactions generated during business processes (Cao 2010). Behavior is a very weak object and behavior analysis is widely used without a unified definition of behavior. Thus, more methods and strategies are expected to be explored from the behavioral perspective.

2.1.2 Behavioral Perspective

Among limited genuine behavior analysis, behavior modeling from the behavioral and social sciences mainly relies on either quantitative or qualitative methods (Peterson, Cumming & Carpenter 2003, Pierce & Cheney 2004).

On a quantitative aspect, for example, Cao proposed a behavior informatics approach for in-depth human behavior understanding and use (Cao 2010); Bi et al. (Bi, Tsimhoni & Liu 2011) modeled human performance by using the support vector regression approach; and Kim et al. (Kim, Breslin, Decker & Kim 2011) represented and mined user interest under the scenario of tagging practices. From the perspective of qualitative modeling, behavior-oriented modeling and analysis has been proposed and studied in areas, including scenario planning (Peterson et al. 2003) and knowledge representation (Razmerita 2011). Other typical concepts such as action reasoning and composition (Gu & Soutchanski 2007, Giacomo et al. 2010), coordination and planning (Koenig, Keskinocak & Tovey 2010, Edelkamp & Kissmann 2009), and modeling systems rather than behaviors (Nagy & Vargas-Vera 2011, Ros & Veloso 2007, Serrano & Saugar 2010) can also be found in a number of references. For instance, Serrano and Saugar (Serrano & Saugar 2010) exploited the application-independent software connector to specify multi-agent societies rather than agent behaviors. Gu and Soutchanski (Gu & Soutchanski 2007) discussed action reasoning based on a modified situation calculus. Sardina et al. (Sardina et al. 2008) considered behavior compositions when failure presents.

In this chapter, we mainly concentrate on the qualitative model and analysis of genuine behavioral elements, which capture the intrinsic components of behavior. We note that many qualitative works on "behavior modeling" actually refer to behavior recognition (Gabaldon 2009) and simulation (Subramanian 2010) instead of representation, which differs from our focus here. Although several abstract models have been proposed as well, e.g. belief-desire-intention model (Wooldridge 2000) and situation calculus (Giacomo et al. 2010), there is little exploration to formalize the concept of behavior and its elements (Cao 2010). In addition, Ros and Veloso (Ros & Veloso 2007) designed two kinds of evaluations to show the superiority of their case-based coordination mechanism, but they did not verify the stability of that system. Limited work can be identified on formalizing and checking complex behavior structures and interactions.

All the above methods only concern each individual behavior. Also mentioned by (Cao et al. 2012) that previous research has mainly focused on individual behaviors, e.g. sequence analysis including Spade (Zaki 2001) and Spam (Ayres et al. 2002), interactive process modeling (Bickhard 2000), activity mining (Cao et al. 2008).

An increasing number of researchers however argue that coupled behaviors, i.e., behaviors from either the same, or different actors, are often associated with one another in terms of some coupling relationships. They are widely seen, although very limited research outcomes can be identified in the literature. Slightly relevant approaches are coupled Hidden Markov model (CHMM) (Oliver & Pentland 2000), multivariate time series analysis (Chandrakala & Chandra Sekhar 2008, Yoon et al. 2005) and social network analysis (Hogg & Szabo 2008). The CHMM (Oliver, Rosario & Pentland 2000) is introduced to model multiple processes with interactions.

The CHMM is composed of more than one HMM chains denoting different processes. The state of any HMM chain at time t depends on both the state of its own HMM chain and the states of other HMM chains at time $t-1$. Regarding the multivariate time series analysis, a density-based clustering approach in the kernel feature space is proposed to cluster multivariate time series behavior data with varying lengths (Chandrakala & Chandra Sekhar 2008). Yoon, Yang and Shahabi (Yoon et al. 2005) suggested to select a feature subset from multivariate time series by adopting common principal component analysis. For the social network analysis, Hogg and Szabo (Hogg & Szabo 2008) proposed and evaluated plausible mechanisms to explain behaviors in online community site. Although all the methods have somewhat considered interactions, demographic data rather than the genuine behavior structure is the main focus, which is not aimed at behavior analysis and overlooks aspects like internal-driven coupling relationships and behavior actions/properties.

2.1.3 Related Techniques

In this section, we briefly introduce two important techniques (i.e. ontology and model checking) to be used in Chapter 3 when building the coupled behavior modeling and checking system.

Ontology

The word "ontology" is used with different meanings in distinct communities (Breitman et al. 2007). Originally, the ontology proposed by Aristotle, is the philosophical study of the nature of being, existence or reality in general, as well as the basic categories of being and their relations. In computer and information sciences, ontology is a formal representation of the knowledge by a set of concepts within a domain and the relationships between those concepts. For artificial intelligence systems, what exist is what can be represented, we can describe the ontology of a program by defining a set of representational terms. The most frequently quoted definition of ontology is

Gruber's: "An ontology is a formal, explicit specification of a shared conceptualization" (Gruber 1993). Here, conceptualization represents an abstract model, formal means that the mathematical specification should be machine processable, and explicit indicates that the elements must be clearly defined. Basically, there are four types of relationship between distinct concepts and individuals: part-of, kind-of, instance-of (is-a), attribute-of. These are only the basic ones, a lot more relationships are to be revealed according to varied situations and application domains.

Due to its strong ability of representation and description, ontology has witnessed its wide applications in many areas such as artificial intelligence (Lim, Suh & Suh 2011), semantic web (Nagy & Vargas-Vera 2011), software engineering (Sicilia 2007), biomedical informatics (Yu 2006) and information architecture (Wessel & Möller 2009). In this thesis, thus, we take advantage of the ontology to model and represent coupled behaviors in Chapter 3.

Model Checking

System verification techniques are applied to the design of information and communication technology in terms of software and hardware verification. Among all the techniques, model checking (Baier & Joost 2008) is currently attracting considerable attention, and was given the ACM Turing Award 2007 in recognition of the paradigm shifting work on this topic initiated a quarter century ago. This verification technology provides an algorithm for determining whether an abstract model, i.e. a hardware or software design, satisfies a formal specification normally expressed as a temporal logic formula (Alur, Henzinger & Kupferman 2002). For instance, the specification contains safety requirements (i.e. the absence of deadlocks) and similar critical states that can cause the system to crash, model checking is then a technique for automatically verifying correctness properties of finite-state systems. Model checking can manage unknown parameters and the nondeterminism caused by uncertain communication delays when concurrent agents interact with one another. Automation also makes model checking far more attractive than other verification techniques such as manual proof of mathematical arguments (Wang, Hidvegi, Bailey & Whinston 2000) and interactive computer aided theorem proof (Kaufmann, Manolios & Moore 2000).

Following the same vein, we develop the modeling and verification for various behavior interactions between coupled behaviors in Chapter 3. Specifically, our approach models and checks interactions for coupled behaviors by formally describing the involved behaviors and required properties as transition systems and temporal logic formulae, respectively. Though Lomuscio, Qu and Raimondi (Lomuscio, Qu & Raimondi 2009) proposed a model checker for the verification of multi-agent systems, they overlook the explicit representation of interactions among agents and the unified formalization of them, which is different from our focus in this chapter.

To the best of our knowledge, there is no existing work on both modeling and verifying the coupled behaviors systematically from a qualitative perspective. In Chapter 3 we therefore develop a framework, namely Ontology-based Qualitative Coupled Behavior Modeling and Checking (*OntoB*) system for modeling coupled genuine behaviors, which differs from the current related research in three aspects: we consider the genuine behavioral elements on top of demographic and transactional inputs; we spell out the coupling relationships between behaviors explicitly; and we enhance the modeling quality by formal verification.

2.2 Numerical Behavior Analysis

Numerical behavior refers to the behavior data described by continuous properties or attributes. In numerical behavior analysis, the entity of behavior is represented by each object or observation, the property of behavior is quantified by each continuous attribute.

The traditional way to deal with the numerical behavior data is to treat each property independently. When calculating the similarity or distance between the entities of numerical behavior, we usually use the Manhattan distance, Euclidean distance and Minkowski distance (Gan et al. 2007), which are all in lack of the interactions between properties.

An increasing number of researchers however point out that the assumption of independent and identical distribution (IIDness) on attributes often leads to a mass of information loss, and several papers have addressed the issue of attribute interactions. In addition to the basic Pearson's correlation (Gan et al. 2007), Jakulin and Bratko analyzed the attribute dependency by information gain (Jakulin & Bratko 2003), but they involve the label information which is only eligible in supervised learning. While a rank-correlated measure (Calders, Goethals & Jaroszewicz 2006) has been proposed to mine frequent patterns, it only considers the pairwise relationship in a local way and works on nonintuitive ranks rather than attribute values. More recently, Wang et al. put forward the coupled nominal similarity in unsupervised learning (Wang et al. 2011), but only for categorical data. A relational classifier was investigated by using multiple source relations (Bollegala, Matsuo & Ishizuka 2011), which yet merely contributes to classification tasks. Plant presented the dependency clustering by mapping objects and attributes in a cluster-specific low-dimension space (Plant 2012), however, the interaction mechanism is embedded in the modeling process of clustering and not explicitly defined. Despite the current research progress, no work has been reported that systematically takes into account the global relationships (i.e. non-IIDness) among continuous attributes.

2.3 Categorical Behavior Analysis

Categorical behavior refers to the behavior data described by nominal properties or attributes. In categorical behavior analysis, the entity of behavior is also embodied by each object or observation, the property of behavior is reflected by each discrete attribute.

Some surveys, in particular (Gan et al. 2007, Boriah, Chandola & Kumar 2008), discuss the similarity between categorical attributes. The usual practice is to binarize the data and then use binary similarity measures rather than directly considering nominal data. Cost and Salzberg (Cost & Salzberg 1993) proposed *MVDM* based on labels, Wilson and Martinez (Wilson & Martinez 1997) performed a detailed study of heterogeneous distance functions for instance based learning, and Figueiredo et. al (Figueiredo, Rocha, Couto, Salles, André Gonçalves & Meira Jr 2011) introduced word co-occurrence features for text classification. Unlike our focus, their similarities are only designed for supervised approaches.

2.3.1 Nominal Similarity in Unsupervised Learning

There are a number of existing data mining techniques for the unsupervised learning of nominal data (Boriah et al. 2008, Ahmad & Dey 2007). Well-known metrics include *SMS* (Kaufman & Rousseeuw 1990) and its diverse variants such as Jaccard coefficients (Ribeiro & Harder 2011), which are all intuitively based on the principle that the similarity measure is 1 with identical values and 0 otherwise and are not data-driven. More recently, the frequency distribution of attribute values has been considered for similarity measures (Boriah et al. 2008), such as *OF* and *Lin*. Similarity computation has been incorporated into the learning algorithm without explicitly defining general measures (Gibson, Kleinberg & Raghavan 2000). Neighborhood-based similarity (Houle, Oria & Qasim 2010, Guha, Rastogi & Shim 2000) was also explored to measure the proximity of objects by using functions that operate on the intersection of two neighborhoods. They present the similarity between a pair of objects by considering only the relationships among data objects, which are built on the similarity between attribute values simply quantified by the variants of *SMS*. However, we focus on the couplings between attributes to further develop the similarity between attribute values and then between data objects. Our proposed similarity measure between attribute values are incorporated with the neighborhood-based similarity between data objects to more precisely describe the neighborhood of an object. It represents the neighborhood-based metric as a meta-similarity measure

(Boriah et al. 2008) in terms of both the couplings between attributes and between objects.

All the above methods are attribute-independent (i.e. IIDness) since similarity is calculated separately for two categorical values of individual attributes. However, an increasing number of researchers argue that the attribute value similarity is also dependent on the couplings of other attributes (Boriah et al. 2008, Cao et al. 2012). The Pearson's correlation coefficient (Houle et al. 2010) measures only the strength of linear dependence between two numerical variables. Das and Mannila put forward the Iterated Contextual Distances algorithm, believing that the attribute, object and sub-relation similarities are inter-dependent (Das & Mannila 2000). They convert each object with binary attribute values to a continuous vector by a kernel smoothing function, and define the similarity between objects as the Manhattan distance between continuous vectors (Das & Mannila 2000). By contrast, we directly consider similarity for categorical values to maintain the least information loss. Andritsos et al. (Andritsos, Tsaparas, Miller & Sevcik 2004) introduced a context sensitive dissimilarity measure between attribute values based on the Jensen-Shannon divergence. Similarly, Ahmad and Dey (Ahmad & Dey 2007) proposed an algorithm *ADD* to compute the dissimilarity between attribute values by considering the co-occurrence probability between each attribute value and the values of another attribute. Though the dissimilarity metric leads to high accuracy, the computation is usually very costly (Ahmad & Dey 2007), which limits its application in large-scale problems. In addition, Ahmad and Dey's (Ahmad & Dey 2007) approaches only focus on the interactions of different attributes, whereas our proposed measure also considers the couplings within each attribute globally.

Chapter 5 mainly focuses on the similarity computation by involving the coupling relationships (i.e. non-IIDness) among discrete properties, specifically for clustering. Below, various categorical clustering methods are briefly reviewed. However, our proposed coupled similarity can also be extended to supervised learning, which is illustrated in Section 5.7.2.

2.3.2 Categorical Clustering

Clustering algorithms (Guha et al. 2000), including partition-based methods such as *k-means* and hierarchy-based methods like divisive approaches (Gan et al. 2007), are more suitable for clustering data with numerical attributes than categorical data.

Clustering of categorical data (categorical clustering for short) is a difficult, yet important task. Many fields, from statistics to psychology, deal with categorical data. Despite this fact, categorical clustering has received limited attention with only a handful of relevant publications. Guha et al. (Guha et al. 2000) proposed a robust hierarchical clustering algorithm *ROCK*, which uses the link-based similarity measure to measure the similarity between two categorical data points and between two clusters. Gibson et al. (Gibson et al. 2000) first construct a hypergraph according to the database, and then cluster the hypergraph using a discrete dynamic system *STIRR*. Andritsos et al. (Andritsos et al. 2004) introduced a scalable hierarchical categorical clustering algorithm *LIMBO* that builds on the Information Bottleneck (IB) framework for quantifying the relevant information preserved when clustering. An incremental algorithm called *COOLCAT* (Barbara, Couto & Li 2002) was proposed to cluster categorical attributes using entropy; however, it is based on the assumption of independence between attributes. Clustering with sLOPE (*CLOPE*), presented by Yang et al. (Yang, Guan & You 2002), uses a global criterion function instead of a local one defined by the pairwise similarity to cluster categorical data, especially transactional data. Rather than aiming for a measure of similarity, *CLICKS* (Zaki, Peters, Assent & Seidl 2005) uses a graph-theoretic approach to find k disjoint sets of vertices in a graph constructed for a particular data set. The last three algorithms, i.e. *COOLCAT*, *CLOPE*, and *CLICKS*, have a different focus from our proposed coupled similarity.

Therefore, in the experiments, we compare the clustering quality of *ROCK*, *STIRR*, and *LIMBO* with the coupled versions of them, i.e., by using our proposed coupled similarity measure to replace the original similarity measure between attribute values in *ROCK* and *LIMBO*.

2.4 Behavior Application: Clustering Ensemble Learning

In this section, we introduce the related work on the process of the clustering ensemble as a behavior application, differentiate the existing consensus function based clustering ensemble from our proposed framework, and discuss other related issues. Here, each object stands for the entity of behavior, and each base clustering is treated as the property of behavior.

2.4.1 Process of Clustering Ensemble

In general, the whole process of the clustering ensemble consists of three steps: building base clusterings, aggregating base clusterings, and post-processing clustering. Various heuristics have been proposed to build the ensemble members, e.g. random initializations (Christou 2011), data sampling (Kuncheva & Vetrov 2006), random projection and random hyperplane splits (Topchy et al. 2005). The combination of base clusterings can be constructed by three kinds of method: the consensus functions (Strehl & Ghosh 2002), the categorical clusterings (Gionis, Mannila & Tsaparas 2007), and the direct optimizations (Christou 2011). The consensus functions focus on the total agreement of all the base clusterings from different perspectives (Li, Ogihara & Ma 2010). The clustering ensemble can also be converted to the problem of clustering categorical data (categorical clustering (Guha et al. 2000, Andritsos et al. 2004), for short) by viewing each attribute as a way of producing a base clustering of the data. However, the direct optimizations (Christou 2011) are substantially performed on the original objective function of clustering rather than exploring the agreement among partial solutions. Finally, the post-processing clustering algorithms are conducted on the consensus building based on the essence of the aggregation structure. For instance, partition-based (e.g., *k-means* (Gionis et al. 2007)) and hierarchy-based (e.g., *single linkage* (Kuncheva & Vetrov 2006)) algorithms are associated with the consensus pairwise matrix, while spectrum-based (e.g., *SPEC* (Fern & Brodley 2004)) and graph-based (e.g., *METIS* (Strehl & Ghosh 2002)) are applicable to the relevant consensus graphs or hypergraphs (Fern & Brodley 2004). The performance of clustering ensemble can be greatly enhanced if the algorithms of these steps are carefully organized.

Here, we focus on building proper consensus functions to aggregate base clusterings, which is the essential element in the clustering ensemble. A consensus function seeks a combination of multiple base clusterings to provide a prior superior input for post-processing clustering. We can construct consensus functions by the following approaches: direct best matching (Li et al. 2010), graph-based mappings (Fern & Brodley 2004, Strehl & Ghosh 2002), statistical mixture models (Topchy et al. 2005), pairwise comparisons (Gionis et al. 2007, Li et al. 2010) and a number of other models. They are all built on the co-associations or pairwise agreements between clusterings (e.g., partition difference (Li et al. 2010) and *QMI* (Topchy et al. 2005)), between data objects (e.g., *CSPA*[1] (Strehl & Ghosh 2002)), or between clusters (e.g., *MCLA* (Strehl & Ghosh 2002)). While the clustering ensemble based on consensus functions largely captures the common structure of the base clusterings, and achieves a combined clustering with better quality than individual clusterings, it also faces several issues that have not been explored well in the consensus design. In the next section, we analyze the problems inherent in the existing work which motivate us to propose a new effective ensemble framework.

2.4.2 Problems on Consensus Functions

Several papers (Strehl & Ghosh 2002, Gionis et al. 2007) address the issue of consensus function for the clustering ensemble. Heuristics including *CSPA*,

[1] Note that the above categories of approach could overlap; for example, *CSPA* is both a graph-based mapping and pairwise comparison.

HGPA and *MCLA* (Strehl & Ghosh 2002) solve the ensemble problem by first transforming the base clusterings into a hypergraph representation and then developing consensus functions. Based on *CSPA* and *MCLA*, Fern and Brodley (Fern & Brodley 2004) proposed *HBGF* to consider the similarity between objects and the similarity between clusters collectively. By defining an appropriate distance measure between objects, Gionis et al. (Gionis et al. 2007) mapped the clustering aggregation problem to the weighted correlation clustering problem with linear cost functions. In addition, Topchy et al. (Topchy et al. 2005) introduced a mixture probability model *EM* and an information-theoretic consensus function *QMI* to effectively combine weak base clusterings. Based on the *EM* model, Nguyen and Caruana (Nguyen & Caruana 2007) presented an *EM*-like consensus algorithm with variations, but they follow the IIDness assumption. Most of the existing research on the consensus function has been summarized in (Li et al. 2010), in which the equivalence is revealed between the basic partition difference (*PD*) algorithm and other advanced methods such as Chi-square based approaches.

All of the above methods either fail to address the interactions between base clusterings and between objects (e.g. *CSPA*, *QMI*) or assume independence between them (e.g. *EM*), thus they are IIDness based. Further, the weighted correlation clustering solution proposed in (Gionis et al. 2007) fails to partition the objects if their distance measures are equally 0.5. However, an increasing number of researchers argue that the clustering ensemble is also dependent on the relationship between input partitions (Iam-On, Boongoen, Garrett & Price 2011, Punera & Ghosh 2007, Domeniconi & Al-Razgan 2009). Punera and Ghosh (Punera & Ghosh 2007) put forward soft cluster ensembles, in which they used a fuzzy clustering algorithm for the generation of base clusterings. The weighted distance measure (Domeniconi & Al-Razgan 2009) represented a soft relation between a pair of objects and clusters. Unlike our proposed framework, those refined solutions of different base clusterings are stacked up to form the consensus function without explicitly addressing the relations among input clusterings. More recently, Iam-On et al. (Iam-On et al. 2011) presented a link-based approach, and its improved model (Iam-On & Boongoen 2012) involves cluster-cluster similarity based on the interaction between clusters.

However, so far no framework has been proposed to consider comprehensive couplings, including intra-coupling within and inter-coupling between base clusterings and objects. In Chapter 6, we propose a general and effective framework for uncovering the non-IIDness nature in ensemble clustering.

2.4.3 Other Related Issues

The clustering ensemble can also be mapped to categorical clustering by treating each base clustering as an attribute (Gionis et al. 2007). Similar to Section 2.3.2, let us explore the widely used categorical clustering algorithms again here. Guha et al. (Guha et al. 2000) proposed *ROCK*, which uses the link-based similarity between two categorical objects. Andritsos et al. (Andritsos et al. 2004) introduced *LIMBO* which is built on the information bottleneck framework for quantifying the relevant information preserved when clustering. In summary, *ROCK* considers the relationship between objects by linkage; *LIMBO* concerns the interaction between different attributes. Neither of them takes couplings between attributes and between objects into account together, but our proposed framework considers both.

In our previous work (Wang et al. 2011) and Chapter 5, we proposed a coupled nominal similarity measure to specify the coupling relationship between attributes. In Chapter 6, we introduce a coupled framework for clustering ensemble, which addresses the problem of seeking the global consensus among base clusterings and also involves the couplings between objects.

2.5 Summary

In this chapter, we present the literature review regarding behavior modeling and analysis, numerical behavior analysis, categorical behavior analysis, and a behavior application on clustering ensembles. The conclusions from the above literature review include but are not limited to:

(1) Traditional behavior modeling approaches exhibit too many styles and forms based on different situations. There is very limited research on formalizing the concept of behavior and its elements. There are no formal behavior representation models stated from a general perspective and providing a comprehensive understanding of behavior constitution.

(2) Current behavior modeling that mainly relies on qualitative methods from behavioral and social sciences often leads to ineffective and limited analysis in understanding social activities deeply and accurately. In addition, many qualitative works on "behavior modeling" actually refer to behavior recognition and simulation rather than representation, which is different from our focus in this thesis.

(3) General behavior expressiveness is too weak to reveal that behavior plays the key role of an internal driving force for social activities. Current behavior-oriented analysis is usually conducted on customer demographic and transactional data directly, which is not organized in terms of behavior but entity relationships closely related to particular business problems.

(4) Complex coupling relationships between group behaviors, either from the same actor or different actors, are often ignored or only weakly addressed. Few building blocks are available to explicitly model complex interactions of group behaviors. Effective approaches for analyzing coupled behaviors are not available, since existing methods mainly focus on individual behavior analysis.

(5) The existing work often overlooks the checking, verification and amendments of behavior modeling, which weakens the soundness and robustness of models built for complex behavior-oriented applications. Limited work can be identified on formalizing, checking and amend-

ing complex behavior structures and interactions, in particularly for coupled behaviors with their constraints.

(6) For most of the existing theories, tools, models and systems in statistics, data mining and machine learning, it is usually assumed that methods, objects, attributes and values are independent and identically distributed (i.e. IIDness). This works well in simple business applications and abstract problems with weakened and avoidable relationships and heterogeneity, and serves as the foundation of classic analytics, mining and learning theories, algorithms, systems and tools.

The above summary is also consistent with the limitations and challenges listed in Section 1. As a matter of fact, behavioral and social applications are ubiquitous, ranging from business, online to social and organizational domains. With the ever-increasing and continuous development of such applications, an emerging demand is to explore and establish an in-depth, robust and comprehensive understanding of underlying driving force, working mechanism, constraints, dynamics and evolution within a behavioral and/or social system, as well as the impact on business and its context. In those applications and domains, coupled behaviors play a much more fundamental role than individuals in the cause, dynamics and effect of business and social problems. In addition, complex behavioral and social applications often exhibit strong but implicit coupling relationships and heterogeneity between objects, object attributes and attribute values, which cannot be degenerated or weakened to the extent of satisfying the IIDness assumption.

Therefore, building on the classic theories and algorithms available in behavioral science, social science and computer science, coupled behavior informatics has been proposed and studied in this thesis to “formalize”, “represent”, “verify”, “analyze”, “learn”, “compute” and “abstract” complex behavioral and social applications.

Chapter 3

Formalization and Verification of Group Behavior Interactions

In this chapter, we model and verify the qualitative coupled behaviors in terms of group behavior interactions. Below, action and operation are interchangeable, they are both used to describe the characteristics of actors in the qualitative coupled behaviors. The behavior addressed in this chapter particularly refers to the qualitative behavior, and the group behaviors with their interactions correspond to the qualitative coupled behaviors.

Group behavior interactions (i.e. qualitative coupled behaviors), such as multi-robot teamwork and group communications in social networks, are widely seen in both natural, social and artificial behavior-related applications. Behavior interactions in a group are often associated with varying coupling relationships, for instance, conjunction or disjunction. Such coupling relationships challenge existing behavior representation methods, because they involve multiple behaviors from different actors, constraints on the interactions, and behavior evolution. In addition, the quality of behavior interactions are not checked through verification techniques.

This chapter proposes an Ontology-based Qualitative Coupled Behavior Modeling and Checking (*OntoB* for short) system to explicitly represent and verify complex behavior relationships, aggregations, and constraints. The

OntoB system provides both a visual behavior model and an abstract behavior tuple to capture behavioral elements, as well as building blocks for the qualitative coupled behaviors. It formalizes various intra-coupled interactions (behaviors conducted by the same actor) via transition systems, and inter-coupled behavior aggregations (behaviors conducted by different actors) from temporal, inferential and party-based perspectives. *OntoB* converts a behavior-oriented application into a transition system and temporal logic formulae for further verification and refinement. We demonstrate the effectiveness of the *OntoB* system in modeling multi-robot behaviors and their interactions in the Robocup soccer competition game. We show that the *OntoB* system can effectively model complex behavior interactions, verify and refine the modeling of complex group behavior interactions in a sound manner.

3.1 Background and Overview

Behavior refers to the action or reaction of any material under given circumstances and environment[1]. It is intrinsic in many areas, and behavior analysis has become a fundamental topic which has been increasingly investigated as an essential activity in many fields, from social and behavioral sciences to computer science (Pierce & Cheney 2004, Zacharias & MacMillan 2008, Liu et al. 2008, Getoor & Taskar 2007). In Google, the keyword "behavior" attracts 281,000,000 hits while "behavior interaction" achieves 76,700,000 results[2]. In both natural and social sciences and applications, multiple behaviors from one or multiple actors often interact with one another, which are called qualitative coupled behaviors or group behavior interactions. These coupled behaviors and behavior interactions may form interior driving forces that shape underlying businesses, such as in online community and social networks (Liu et al. 2008, Hogg & Szabo 2008), or may even cause challeng-

[1] http://dictionary.reference.com

[2] These results are searched on 13th June 2013

ing problems like group-based manipulation by a group of traders (Cao & Yu 2009). With the deepening and widening of complex networking, coupled behaviors or group behavior interactions are increasingly seen in both mainstream and emerging situations, in particular, in enterprise applications, organizations, complex systems, online and social communities.

We illustrate the coupled behaviors and behavior interactions using the example of multi-robot soccer game in Figure 1.1 mentioned in Chapter 1. Figure 1.1 shows that two teams participate in a Robocup soccer competition (*Sony Four Legged Robot Football League Rule Book* 2004) with four Sony AIBO robots in each group. As indicated in the scenario described by Ros and Veloso (Ros & Veloso 2007), a team of robots intelligently cooperate with one another and self-adjust their own activities; the successful task execution and problem resolution rely on the proper implementation of an individual robot's activities as well as collaborative interactions between robots. If a robot undertakes tasks without appropriate arrangement and coordination with the other robots, the Robocup is likely to be unsuccessful, even though every robot performs perfectly. This example shows that group actors and behaviors by the same or different actors within the group are often coupled in different forms of interactions (Cao et al. 2012), and it is essential to identify, represent and verify how the robots interact to ensure the performance of a multi-robot system.

To enable the above behavior interaction-oriented systems to work properly, a fundamental task is to develop effective behavior representation tools to capture, formalize and verify behavioral elements, coupling relationships, and interactions between behaviors, in both qualitative and quantitative aspects (Cao 2010). This challenges the existing behavior representation research[3], including user behavior modeling (Kim et al. 2011, Razmerita 2011), periodic behavior analysis (Cao et al. 2007), social network analysis (Hogg & Szabo 2008), behavior learning (Subramanian 2010), reasoning about action (Gu & Soutchanski 2007), behavior composition (Sardina et al. 2008),

[3] http://brimsconference.org

action recognition (Gabaldon 2009), and modeling a system (e.g., modeling a multi-agent system (Ros & Veloso 2007, Serrano & Saugar 2010)). Most of the existing research models a single behavior or analyzes behavior groups by focusing on either demographic or transactional perspectives. For instance, for the robot soccer game, Lim et al. (Yoshimura, Barnes, Ronnquist & Sonenberg 2003) built a dynamic formation mechanism based on individual robots. Though Ros and Veloso (Ros & Veloso 2007) designed a multi-robot framework with implicit interactions, they did not explicate behavioral elements. Kaminka and Frenkel (Kaminka & Frenkel 2005) presented a flexible teamwork architecture for teams of robots by protocols, but without the verification of the protocols. Little related work is available for the explicit formalization and verification of coupled behaviors (Cao et al. 2012), including elaborating and representing behavioral elements (Brachman & Levesque 2004, Wang & Cao 2010, Cao 2010), specifying behavior interaction relationships, and checking the modeling of multiple behavior couplings (Baier & Joost 2008). A detailed review of the related work on behavior modeling and analysis can be found in Section 2.1.

In summary, the existing work is not effective for modeling and checking group behaviors with interactions due to the following major issues:

(1) While several abstract models have been proposed, such as the belief-desire-intention model (Wooldridge 2000) and the situation calculus (Giacomo et al. 2010), there is very limited research on formalizing the concept of behavior and its elements (Cao 2010). In fact, behavior is a fuzzy concept in existing studies; there are no formal behavior representation stated from a general perspective and providing a comprehensive understanding of behavior constitution.

(2) Complex coupling relationships between group behaviors are often ignored or only weakly addressed; few building blocks are available to explicitly model complex interactions between group behaviors (Cao et al. 2012).

(3) The existing work often overlooks the checking of behavior modeling and behavior interactions, which weakens the soundness and robustness of models built for complex behavior applications.

In this chapter, we build an Ontology-based Qualitative Coupled Behavior Modeling and Checking (*OntoB* for short) system for representing and verifying complex relationships and interactions of the qualitative coupled behaviors. The built-in behavior ontology combines the features of entity-relationship with theoretical semantics to explicitly capture behavior interactions, various intra-couplings (interactions between behaviors from the same actor) and inter-couplings (interactions between behaviors from different actors) between distinct behaviors as well as their aggregation and behavior constraints, in terms of behavioral perspective rather than transactional aspect. Further, the inclusion of model checking (Baier & Joost 2008) in the *OntoB* system makes reliable models, and outperforms the manual proof and test with simulations in terms of nondeterminism and automation. As a matter of fact, the building blocks in the *OntoB* system are *generic* and can be used for modeling behavior-oriented applications with a variety of coupling relationships. In addition, the semantic mappings of intra-coupled and inter-coupled syntax in *OntoB* are *flexible* according to specific requirements when interpreted as the corresponding aggregations, which means the way we provide here for verification is just an alternative option.

Throughout this chapter, we present the *OntoB* system, and illustrate it via modeling and verifying behaviors of all robots as well as their relationships from temporal, inferential and party-based perspectives in the multi-robot soccer game, exhibited in Figure 1.1.

The rest of the chapter is structured as follows. In Section 3.2, we provide a framework to explain the building blocks of our system. Section 3.3 introduces the behavior descriptors in terms of a visual model and formal specification. The behavior aggregators including intra-coupling, inter-coupling and their combination are presented in Section 3.4. Case study of a complex system on formalization strategy is provided in Section 3.5. We specify the behavior constraint indicator in Section 3.6. Section 3.7 presents the behavior checking system. The behavior model refiner and behavior model exporter are described in Section 3.8. We end this chapter in Section 3.9.

3.2 Coupled Behavior Modeling Framework

3.2.1 Modeling and Checking System

Previous behavior studies on model checking for rational agents (Bordini, Fisher, Wooldridge & Visser 2004), multi-agent behavior modeling (Sun 2007), behavior informatics (Cao 2010), coupled behavior analysis (Cao et al. 2012), and ontology-based frameworks (Lim et al. 2011, Razmerita 2011) provide a solid foundation for forming a unified behavior modeling and checking framework, to describe group behaviors with coupling relationships from the low-level visual description to the high-level formal abstraction, as well as to verify the aforementioned required properties in the resulting behavior models.

This chapter introduces such a general framework, namely an Ontology-based Qualitative Coupled Behavior Modeling and Checking (*OntoB*) system, to represent group behaviors and behavior interactions (i.e. qualitative coupled behaviors). *OntoB* includes several building blocks: behavior descriptor, behavior aggregator, behavior constraint indicator, behavior checker, behavior model refiner, and behavior model exporter. A schematic overview of *OntoB* is shown in Figure 3.1.

The five components in *OntoB* play different roles according to their specifically assigned responsibilities.

- *Behavior Descriptor*: The qualitative coupled behaviors are represented in terms of both a visual structure model and a formal abstract model, by extracting and constructing the core behavioral properties/elements and interactions.

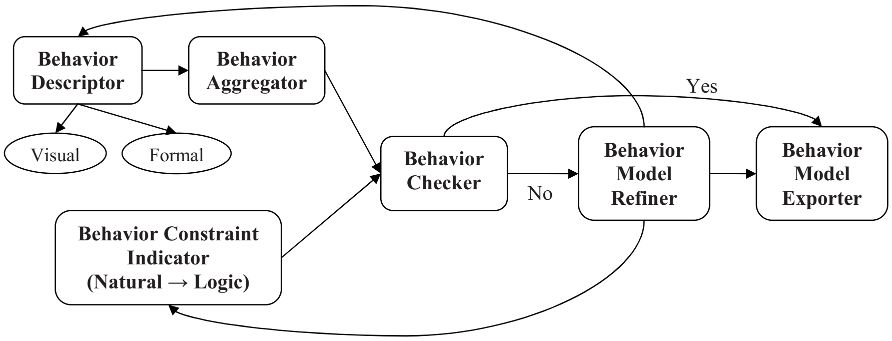


Figure 3.1: General framework of the *OntoB* system.

- *Behavior Aggregator*: Hierarchical and hybrid combinations of the qualitative coupled behaviors are aggregated to generate and convert into the semantics of transition systems in terms of intra-coupling and inter-coupling relationships in group behaviors.

- *Behavior Constraint Indicator*: The natural language description about the constraints or properties of a behavior-based system is formalized into logic formula.

- *Behavior Checker*: This checker verifies whether the resulting behavior model satisfies the properties and constraints imposed by behavior constraint indicator.

- *Behavior Model Exporter*: A stable and desired behavior model is generated and exported as a result of a perfect modeling and checking process engaging the above components.

- *Behavior Model Refiner*: The behavior model is further revised to fix any problem within the model or to address inaccurate descriptions of the constraints until no counter example arises.

The Behavior Descriptor provides the coupled behavior representation mechanism to enable behavior verification in terms of visual description and formal abstraction, while the Behavior Aggregator generates a combined

Table 3.1: List of Main Notations in Chapter 3

Variable	Explanation
$\mathbb{B}_c$	Qualitative coupled behaviors
$\mathscr{C}$	Coupling
$\{\mathscr{A}_1, \ldots, \mathscr{A}_I\}$	The set of I actors
$\{\mathscr{O}_{i1}, \ldots, \mathscr{O}_{iJ_i}\}$	The set of J_i operations conducted by actor $\mathscr{A}_i$
$\theta(\mathbb{B})$	Intra-coupling function
$\eta(\mathbb{B})$	Inter-coupling function
$FM(\mathbb{B})$	Behavior feature matrix
$\mathbb{B}^\theta(\mathscr{A}_i)$	Intra-coupled behaviors of actor $\mathscr{A}_i$
$\mathbb{B}^\eta(\mathscr{A}_i)$	Inter-coupled behaviors of actor $\mathscr{A}_i$

transition system for interpreting multiple behavior interactions. The Behavior Constraint Indicator incorporates predefined logic properties so that the resulting behavior model satisfies the domain specification and priori knowledge. Building on the above three components, the Behavior Checker verifies the resultant behavior model in terms of certain constraints. If nothing is detected by the Behavior Checker, the Behavior Model Exporter outputs a behavior model for further exploration such as reasoning and inference; otherwise, the Behavior Model Refiner either fixes the errors in the model or corrects the misunderstanding of constraints until no further improvement of the verification is needed.

The Behavior Descriptor, the Behavior Aggregator, and the Behavior Constraint Indicator are the prerequisites of the Behavior Checker. The Behavior Model Refiner and the Exporter adjust the outputs based on the verification results from the Behavior Checker. As a result of the Behavior Descriptor, the visual model is the intuitive and perceptible understanding about behaviors in a system, such as multi-agent systems and stock trading behaviors. In contrast, the corresponding formal model abstracts syntactic behavior-related concepts to enable further verification, reasoning or inference. In addition, the Behavior Aggregator interprets the implicit semantics

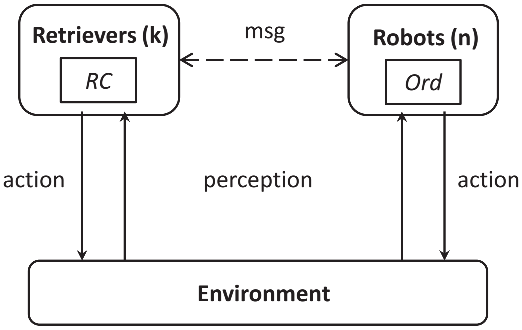


Figure 3.2: A case-based multi-robot system with n robots and k retrievers.

about the abstracted coupled behavior syntax for verification. As indicated in the schematic framework (Figure 3.1), all the above components are built upon others, logically and reasonably, to form an integrative system for modeling and analyzing behaviors. Note that the main notations in this chapter are listed in Table 3.1.

3.2.2 Case Study Description

We illustrate the proposed coupled behavior modeling and checking framework in terms of the case study: the multi-robot soccer game in Section 1. As defined in (Ros & Veloso 2007) and shown in Figure 3.2, this multi-robot architecture is composed of n robots, including k retrievers. All the robots interact with the environment. During the interactions, they perceive the world, perform actions, and communicate messages with one another for a collaboration. A team of retriever robots RCs and robot players Ords communicate with one another and try to put the ball in the opponent's goal as frequently as possible, while the opponent's robots have the same goal. When a new situation arises, a distinguished set of k retrievers RCs take charge of selecting cases from a case space and then inform the rest of the ordinary robot players Ords. Also as the coordinators, RCs send messages (msg) to all Ords and instruct them to conduct the corresponding actions. If timeout expires, or messages or cases are lost in the interactions, Ords abort the executions at any moment based on their own perceptions.

Below, we discuss all modules in the proposed *OntoB* in detail, and illustrate them by concrete examples from the above case study system.

3.3 Behavior Descriptor

According to (Cao 2010), *behaviors* refer to actions, operations or events as well as their combinations such as processes, procedures and activity sequences conducted by either an individual or a group of actors within certain contexts and environments in either a virtual or physical organization. The goal of behavior modeling is to form concepts and facilities that enable representation and reasoning about behaviors in an organization. For this, we propose both a *visual* model and a *formal* model to represent the qualitative coupled behaviors.

3.3.1 Behavior Visual Descriptor

In *OntoB*, coupled behavior modeling and checking is based on the behavior ontology (Staab & Studer 2009), which typically consists of a number of classes, relations, instances and axioms to enable the representation and reasoning of behaviors. As clarified by Gruber (Gruber 1993), an ontology is a formal, explicit specification of a shared conceptualization. This definition is the most frequently quoted for the concept: ontology, where a set of objects and their relationships is described by a vocabulary. In this chapter, we propose the behavior ontology composed of three core units, namely *actor*, *operation* and *coupling* (Cao 2010), as shown in Figure 3.3. This kind of behavior that examines coupling is also called the qualitative coupled behaviors. Below, an explicit specification of this shared conceptualization (i.e. qualitative coupled behaviors) is given as the behavior visual descriptor. In next section, the formal description will be abstracted to complement this visual representation.

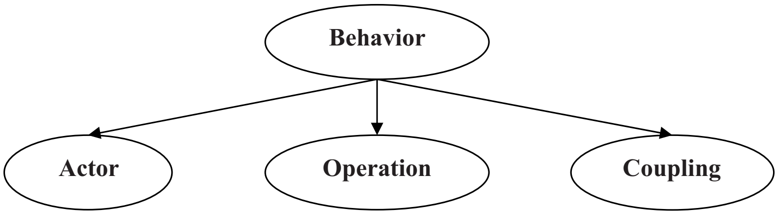


Figure 3.3: The behavior ontology that involves coupling (i.e. coupled behaviors).

- *Actor*: refers to the subject(s) or object(s) of coupled behaviors, for example, organizations, departments, systems, agents and people involved in an activity or activity sequence.

- *Operation*: represents activities, actions or events in coupled behaviors or coupled behavior sequence. For example, in the robot soccer game, kicking the ball is an operation; in the stock market, placing a buy order is an operation too.

- *Coupling*: refers to the interactions in coupled behaviors, including connections between actors and/or operations of either one or multiple actors. For example, all actions conducted by a robot player are associated with one another, while interactions between behaviors of robot players in the same team are also coupled.

We categorize coupling relationships into *intra-coupling relationship* and *inter-coupling relationship. Intra-coupling relationship* refers to the association between behaviors from the same actor, while *inter-coupling relationship* exists in behaviors from different actors.

Based on the various communication strategies within a concurrent system (Baier & Joost 2008) and knowledge representation mechanisms (Brachman & Levesque 2004), we classify the behavior coupling relationships into three categories: *temporal*, *inferential* and *party-based* couplings, as shown in Figure 3.4. Those nodes that represent couplings can be further zoomed in to see the specific relationships stated below by illustrating with the robot soccer game.

In addition, three types of relationships are displayed in Figure 3.4:

- *Instance Of* (indicated by "−·→"), connecting instances to their corresponding classes;
- *Subclass Of* (shown by "⟶"), linking a subclass to its parent class; and
- *Unit Feature* ("--➝"), denoting the relationships between instances, between an unit and its features, or between features. Here unit refers to a behavior or its actor, operation or coupling.

The above relationships associate coupled behaviors and their units into a hierarchical ontology structure. For instance, two relationships: "*Behavior 1* −·→ *Behavior*" and "*Behavior 2* −·→ *Behavior*" depict that *Behavior 1* and *Behavior 2* are instances of the root *Behavior*. Relationships "*Intra-coupling* ⟶ *Coupling*" and "*Inter-coupling* ⟶ *Coupling*" specify that *Intra-coupling* and *Inter-coupling* are subclasses of *Coupling*, which further indicates the interaction between *Behavior 1* and *Behavior 2*. Relationship "*Actor 1* --➝ *Behavior 1*" and "*Operation 1* --➝ *Behavior 1*" indicate that *Actor 1* and *Operation 1* are the features that characterize *Behavior 1*, in which *Actor 1* conducts *Operation 1*, and the completion of *Operation 1* has impact on *Actor 1*. Correspondingly, units *Temporal*, *Inferential* and *Party-based* are constituents of either *Intra-coupling* or *Inter-coupling* between behaviors.

In the following sections, we illustrate each category of coupling relationships by the relevant instances in the robot soccer game from temporal, inferential and party-based coupling aspects.

Temporal Coupling

From the temporal perspective, behaviors form into sequences. Such behavior sequences may take different forms, from serial combinations to more complex

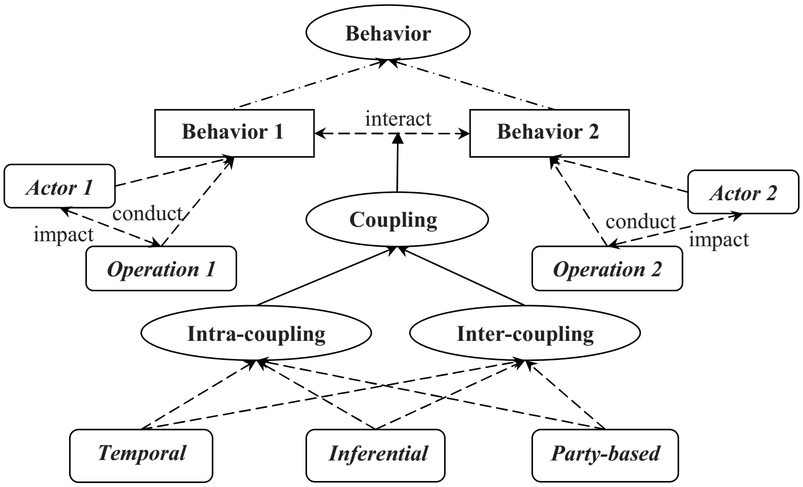


Figure 3.4: Relationships between the qualitative coupled behaviors: *Instance Of* ("−·→"), *Subclass Of* ("⟶"), *Unit Feature* ("- - →").

settings, such as *serial coupling*, *parallel coupling*, *synchronous coupling*, and *asynchronous coupling*.

- *Serial coupling*, denoted by $\{\mathbb{B}_1; \mathbb{B}_2\}$, showing the situation in which behavior $\mathbb{B}_2$ follows behavior $\mathbb{B}_1$. For instance, a robot player *Ord* kicks the ball to score after capturing it.

- *Parallel coupling*, by which behaviors happen in varying concurrent manners, including synchronous coupling and asynchronous coupling.

  - *Synchronous* relationship, denoted by $\{\mathbb{B}_1 \| \mathbb{B}_2\}$, indicating that $\mathbb{B}_1$ and $\mathbb{B}_2$ present at the same time based on certain communication protocols. For instance, a retriever *RC* and a robot player *Ord* both give up execution simultaneously when messages are lost.

  - *Asynchronous coupling*, showing that two behaviors $\mathbb{B}_1$ and $\mathbb{B}_2$ interact with each other at different time points. Here, we focus on three typical kinds of asynchronous interactions: *interleaving*, *shared-variable*, and *channel system*.

- ∗ *Interleaving*, denoted by $\{\mathbb{B}_1 : \mathbb{B}_2\}$, representing the involvement of independent complex behaviors by nondeterministic choice, i.e., the involved behaviors happen independently. For example, two Ords independently execute the actions received from one RC;
- ∗ *Shared-variable*, denoted by $\{\mathbb{B}_1|||\mathbb{B}_2\}$, signifying that the relevant behaviors have variables in common. An instance is that two RCs coordinate with each other to retrieve a case to conduct the other robot players; and
- ∗ *Channel system*, denoted by $\{\mathbb{B}_1|\mathbb{B}_2\}$, is a parallel system in which complex behaviors communicate via a channel, for instance, first-in and first-out buffers. An example is that a RC sends an execution message to a robot player Ord through a channel.

Inferential Coupling

Inferential coupling shows that behaviors are associated with certain logic reasoning relationships, in particular, *causal*, *conjunction*, *disjunction* and *exclusive* couplings.

- *Causal coupling*, represented as $\{\mathbb{B}_1 \rightarrow \mathbb{B}_2\}$, meaning that behavior $\mathbb{B}_1$ causes behavior $\mathbb{B}_2$. For instance, every message sent from a retriever RC will lead to a response from the robot player Ords.

- *Conjunction coupling*, $\{\mathbb{B}_1 \wedge \mathbb{B}_2\}$, specifying that $\mathbb{B}_1$ and $\mathbb{B}_2$ take place together. For instance, all Ords must complete their allocated actions before a new case is retrieved.

- *Disjunction coupling*, $\{\mathbb{B}_1 \vee \mathbb{B}_2\}$, by which at least one of the associated behaviors must happen. For instance, any robot player can abort the execution if the message is lost.

- *Exclusive coupling*, $\{\mathbb{B}_1 \otimes \mathbb{B}_2\}$, indicating that if $\mathbb{B}_1$ happens, $\mathbb{B}_2$ will not happen, and vice versa. For instance, defensive and offensive strategies cannot be conducted by an Ord at the same time.

The above inferential interactions correspond to the following operators in the propositional logic: IMPLY, AND, OR and XOR, respectively.

Party-based Coupling

The party-based coupling reflects a partner relationship associating actors and operations. We consider three types of party-based couplings: *One-Party-Multiple-Operation* (OPMO), *Multiple-Party-One-Operation* (MPOO), and *Multiple-Party-Multiple-Operation* (MPMO) couplings. Each describes the amount of behaviors and correspondingly affiliated parties. Formally,

- *One-Party-Multiple-Operation*, represented as $\{(\mathbb{B}_1, \mathbb{B}_2)^{[\mathscr{A}_1]}\}$, depicts that distinct behaviors $\mathbb{B}_1$ and $\mathbb{B}_2$ are performed by the same actor $\mathscr{A}_1$;

- *Multiple-Party-One-Operation*, shown as $\{(\mathbb{B}_1)^{[\mathscr{A}_1 \mathscr{A}_2]}\}$, represents that multiple actors $\mathscr{A}_1$ and $\mathscr{A}_2$ implement the same behavior $\mathbb{B}_1$ to achieve their own intentions;

- *Multiple-Party-Multiple-Operation*, presented as $\{(\mathbb{B}_1,\ \mathbb{B}_2)^{[\mathscr{A}_1 \mathscr{A}_2]}\}$, describes that different behaviors $\mathbb{B}_1$ and $\mathbb{B}_2$ are carried out by distinct actors $\mathscr{A}_1$ and $\mathscr{A}_2$, respectively.

For example, the situation in which a robot player Ord sends acknowledgment to the retrievers RCs after receiving an offensive message is like an OPMO. However, MPOO describes the scenario in which two Ords execute block simultaneously, and MPMO refers to that in which two Ords complete their designated tasks at the respective stages of waiting and sending acknowledgment.

The above three types of coupling relationships provide a foundation for representing complex interactions between behaviors either directly or through the composition of some of the above couplings, as needed.

3.3.2 Behavior Formal Descriptor

With the behavior visual model, we further establish the behavior formal descriptor to enable the verification and reasoning in the subsequent stage. The behavior visual and formal descriptors complement each other to support the complete formulation of behavior interactions.

The previous sections focus on revealing the explicit description of the behavior elements in a visual way. In this section, we first introduce an abstract behavior model by specifying the concepts and relationships involved in the qualitative coupled behaviors. Further, we propose a formal behavior model to represent the various relationships based on the ontology specification. Inspired by the abstract behavior model proposed in (Cao et al. 2012), several relevant definitions are given, followed by illustration of their use in modeling behaviors in the robot soccer game system.

Definition 3.3.1 (Qualitative Coupled Behaviors) ***Qualitative coupled behaviors*** $\mathbb{B}_c$ *are described as a triple tuple* $\mathbb{B}_c = (\mathscr{A}, \mathscr{O}, \mathscr{C})$*,*

- *Actor* $\mathscr{A}$ *is the entity that issues a behavior or on which a behavior is imposed.*
- *Operation* $\mathscr{O}$ *is what an actor conducts in order to achieve certain goals.*
- *Coupling* $\mathscr{C} = \langle \theta(\mathbb{B}), \eta(\mathbb{B}) \rangle$ *is a tuple that reveals complex interactions including intra-coupling (i.e.* $\theta(\mathbb{B})$*) and inter-coupling (i.e.* $\eta(\mathbb{B})$*).*

Note that the specific semantic meanings of intra-coupling $\theta(\mathbb{B})$ and inter-coupling $\eta(\mathbb{B})$ are clarified in Section 3.4. Here, we only propose the syntax of the concept: qualitative coupled behaviors.

For instance, in a stock market, the qualitative coupled behaviors can be represented as “an investor places a buy order”. The involved *actor* is the “investor” himself or herself, the *operation* is the transaction of “buy”. The third component *coupling* exposes the intra-relationship between this behavior and this investor’s sell order on the other day, together with the inter-relationship between this behavior and another investor's buy order on the same day. We tackle the coupled behaviors from either one or different actors, denoted as intra-coupling and inter-coupling, respectively.

Intuitively, suppose there are I actors $\{\mathscr{A}_1, \mathscr{A}_2, \ldots, \mathscr{A}_I\}$, an actor $\mathscr{A}_i$ undertakes J_i operations $\{\mathscr{O}_{i1}, \mathscr{O}_{i2}, \ldots, \mathscr{O}_{iJ_i}\}$ individually. From this perspective, we have a *Behavior Feature Matrix* $FM(\mathbb{B})$ as follows:

$$FM(\mathbb{B}) = \begin{pmatrix} \mathscr{O}_{11} & \mathscr{O}_{12} & \ldots & \mathscr{O}_{1J_{max}} \\ \mathscr{O}_{21} & \mathscr{O}_{22} & \ldots & \mathscr{O}_{2J_{max}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathscr{O}_{I1} & \mathscr{O}_{I2} & \ldots & \mathscr{O}_{IJ_{max}} \end{pmatrix}, \qquad (3.3.1)$$

where $J_{max} = max\{J_1, J_2, \cdots, J_I\}$, for an operation set $\{\mathscr{O}_{ij} | J_i < J_{max}\}$, the corresponding element $\mathscr{O}_{ij}$ is recognized as $\emptyset$ when $J_i < j \leq J_{max}$. In fact, each (i, j) element in this matrix $FM(\mathbb{B})$ is the jth operation conducted by actor $\mathscr{A}_i$. Further, the intra-coupling, which reveals the complex couplings within an actor's distinct behaviors, is the relationship inside one row of the above matrix, such as the interaction between $\mathscr{O}_{11}$ and $\mathscr{O}_{12}$; while the way multiple behaviors of different actors interact is embodied in each column of $FM(\mathbb{B})$, indicated as inter-coupling, for example, the relationship between $\mathscr{O}_{11}$ and $\mathscr{O}_{21}$. In this chapter, operation and behavior are interchangeable if there is no ambiguity about the involved actors.

In detail, two formal definitions about *intra-coupled behaviors* and *inter-coupled behaviors* are abstracted hierarchically in a recursive way to show the basic syntax of behavior couplings, followed by the combined syntactic concept of coupled behaviors (Cao et al. 2012). Note that the syntax below is proposed to reflect how the couplings to be considered, and the semantic meanings of such syntax are interpreted in Section 3.4.

Definition 3.3.2 (Intra-Coupled Behaviors) *Actor $\mathscr{A}_i$'s operations $\mathscr{O}_{ij}$ $(1 \leq j \leq J_{max})$ are intra-coupled with each other in terms of coupling func-*

tions $\theta_j(\mathbb{B})$, which are called ***intra-coupled behaviors*** *$\mathbb{B}^\theta(\mathscr{A}_i)$.*

$$\mathbb{B}^\theta(\mathscr{A}_i) ::= FM(\mathbb{B})_{i\cdot}|\sum_{j=1}^{J_{max}} \theta_j(\mathbb{B}) \odot_a \mathscr{O}_{ij}, \tag{3.3.2}$$

where $\odot_a$ means that the operations $\mathscr{O}_{ij}$ are intra-coupled via $\theta_j(\mathbb{B})$, and $\sum_{j=1}^{J_{max}} \odot_a$ is the connection of them.

For instance, in the stock market, the investor will place a sell order at some time after buying his or her desired instrument due to a great rise in the trading price. This is, to some extent, one way to express how these two behaviors are intra-coupled with each other.

Definition 3.3.3 (Inter-Coupled Behaviors) *Actor $\mathscr{A}_i$'s operations $\mathscr{O}_{ij}$ $(1 \leq i \leq I)$ are inter-coupled with each other in terms of coupling functions $\eta_i(\mathbb{B})$, which are called* ***inter-coupled behaviors*** *$\mathbb{B}^\eta(\mathscr{A}_i)$.*

$$\mathbb{B}^\eta(\mathscr{A}_i) ::= FM(\mathbb{B})_{\cdot j}|\sum_{i=1}^{I} \eta_i(\mathbb{B}) \odot_e \mathscr{O}_{ij}, \tag{3.3.3}$$

where $\odot_e$ means that the operations $\mathscr{O}_{ij}$ are inter-coupled via $\eta_i(\mathbb{B})$, and $\sum_{j=1}^{J_{max}} \odot_e$ is the connection of them.

For instance, a trading happens successfully only when an investor sells the instrument at the same price as the one for which the other investor buys this instrument. This is another example of how to trigger the interactions between inter-coupled behaviors.

In practice, behaviors may interact with one another in both ways of intra-coupling and inter-coupling. Thus, their appropriate combinations are considered in the following by integrating Definition 3.3.2 and Definition 3.3.2 with Definition 3.3.1.

Corollary 3.3.1 (Qualitative Coupled Behaviors) *Actor $\mathscr{A}_{i_1}$'s operations $\mathscr{O}_{i_1j_1}$ and actor $\mathscr{A}_{i_2}$'s operations $\mathscr{O}_{i_2j_2}$ $((i_1 \neq i_2) \vee (j_1 \neq j_2) \wedge (1 \leq i_1, i_2 \leq I) \wedge (1 \leq j_1, j_2 \leq J_{max}))$ are coupled via function $h(\theta(\mathbb{B}), \eta(\mathbb{B}))$, which are* *the qualitative coupled behaviors* $\mathbb{B}_c$.

$$\mathbb{B}_c = (\mathbb{B}^\theta(\mathscr{A}_{i_1}))^\eta \circ (\mathbb{B}^\theta(\mathscr{A}_{i_2}))^\eta ::= FM(\mathbb{B})_{ij}| \sum_{i_1,i_2=1}^{I} \sum_{j_1,j_2=1}^{J_{max}} h(\theta_{j_1}(\mathbb{B}), \theta_{j_2}(\mathbb{B}), \eta_{i_1}(\mathbb{B}), \eta_{i_2}(\mathbb{B})) \odot_c (\mathscr{O}_{i_1j_1}, \mathscr{O}_{i_2j_2}), \tag{3.3.4}$$

where $h(\theta_{j_1}(\mathbb{B}), \theta_{j_2}(\mathbb{B}), \eta_{i_1}(\mathbb{B}), \eta_{i_2}(\mathbb{B}))$ *is the coupling function denoting the corresponding relationships between* $\mathscr{O}_{i_1j_1}$ *and* $\mathscr{O}_{i_2j_2}$. $\sum_{i_1,i_2=1}^{I} \sum_{j_1,j_2=1}^{J_{max}} \odot_c$ *means the connection of operations* $\mathscr{O}_{i_1j_1}$ *coupled with* $h(\theta_{j_1}(\mathbb{B}), \eta_{i_1}(\mathbb{B}))$ *and operations* $\mathscr{O}_{i_2j_2}$ *coupled with* $h(\theta_{j_2}(\mathbb{B}), \eta_{i_2}(\mathbb{B}))$.

As shown above, we take into account the intra-relationship and inter-relationship hierarchically embedded with each other to form the comprehensive couplings between behaviors. For instance, we consider both the successful trading between investor $\mathscr{A}_1$ (buy) and investor $\mathscr{A}_2$ (sell), and then the selling behavior conducted by $\mathscr{A}_1$ after he or she has bought the instrument at a relative low price.

So far, we have presented an abstract syntax for representing the qualitative coupled behaviors. Concrete mappings and semantic interpretations will be discussed in Section 3.4 to give specific meanings for intra-coupling and inter-coupling. The case study introduced in Section 3.1 is again described to briefly explain the concepts involved.

Example 3.3.1 (Robot scenario model) *We model the multi-robot system in the context of the robot soccer game as shown in Figure 3.2. In this scenario, for the robot's behavior tuple* $\mathbb{B} = (\mathscr{A}, \mathscr{O}, \mathscr{R})$, *the involved actors* $\mathscr{A}$ *are robots RCs (k retrievers), Ords* ($n - k$ *other robot players) for the red team; the operations* $\mathscr{O}$ *are the actions of the corresponding actors, for instance, a robot RC "retrieve case" from the case space to lead the whole team; and the couplings* $\mathscr{R}$ *are the interactions among these robots. They try to get the ball and kick it to score; the subject actors RCs retrieve the most appropriate case and coordinate the object actor Ords to collaborate with one another in order to win the match. The strategies of all the robots can be* *either defensive (*def*) or offensive (*off*). The intra-coupled behaviors* $\mathbb{B}^{\theta}(\mathscr{A}_i)$ *indicate the behavior interactions within robots* RC*s,* Ord*s respectively, e.g., the robot* RC *could send messages to its team members only after it has already retrieved a case from case space; while the relationships between different robots refer to the inter-coupled behaviors* $\mathbb{B}^{\eta}(\mathscr{A}_i)$*, such as the coupling between one* RC *and one* Ord*, e.g., a robot* Ord *sends acknowledgement to one of its captains* RC *when it receives a defensive command from this* RC *through communication channels.*

The above examples show that coupled behaviors $\mathbb{B}_c$ are properly represented in our proposed model from both intra-coupled and inter-coupled aspects. In this case, we are concerned with both individual and group-based scenarios including how each player acts individually, and how all of these robots interact with one another. In the following sections, for simplicity, $\mathbb{B}$ is used to denote coupled behaviors $\mathbb{B}_c$.

3.4 Behavior Aggregator

The description of qualitative coupled behaviors in Section 3.3.2 can be given with different semantic explanations in relation to distinct purposes, such as fraud detection (Cao et al. 2012), relational learning (Getoor & Taskar 2007), machine learning (Wang et al. 2011), inference, and reasoning.

In this section, we conduct behavior aggregations to interpret the interactions of intra-coupled and inter-coupled behaviors. The outcomes of the behavior aggregations form the basis of behavior verification. Three types of aggregations will be discussed in the following: intra-coupled aggregation is used to specify the intra-coupled behaviors with function $\theta_j(\mathbb{B})$, inter-coupled aggregation is disclosed to described the inter-coupled behaviors with function $\eta_i(\mathbb{B})$, and combined aggregation is made to interpret the coupled behaviors with function $h(\theta(\mathbb{B}), \eta(\mathbb{B}))$.

3.4.1 Intra-coupled Aggregation

For the behaviors conducted by the same actor, we interpret the behavior dynamics in terms of a transition system (TS). The transition system (Baier & Joost 2008) is often used in computer science for modeling the behavior dynamics of a system. A transition system consists of directed graphs, in which nodes represent system states and edges represent model transitions, i.e. state changes of a system. A state describes the behavior status at a certain moment of system dynamics.

In particular, the TS interpretation of intra-coupled behaviors $\mathbb{B}^{\theta}(\mathscr{A}_i)$ for actor $\mathscr{A}_i$ is the tuple $(\mathrm{St}, \mathrm{Act}, \rightarrow, \mathrm{In})$ simplified from (Baier & Joost 2008), where $\theta_j(\mathbb{B})$ is the intra-coupling function introduced in Definition 3.3.2, and $1 \leq i \leq I$, $1 \leq j \leq J_{max}$.

- $\mathrm{St} = \{\theta_j(\mathbb{B})\}$ is a set of states.
- $\mathrm{Act} = \{\mathscr{O}_{ij}\}$ is a set of actions or operations.
- $\theta_{\mathrm{j}}(\mathbb{B}) \xrightarrow{\mathscr{O}_{\mathrm{ij}}} \theta_{\mathrm{j+1}}(\mathbb{B})$ is a transition relation, and $\rightarrow$ denotes the connection operator $\odot_a$ in Equation (3.3.2).
- $\mathrm{In} = \{\theta_0(\mathbb{B})\}$ is a set of initial states.

In this behavior-oriented transition system, we regard an operation as a corresponding action in TS; and the intra-coupling function $\theta_j(\mathbb{B})$, which links intra-coupled behaviors, represents the associated states in TS to connect all the involved operations. In fact, every actor is interpreted by an independent transition system unless two actors perform exactly the same in the observation. Stated differently, each row of the behavior feature matrix $FM(\mathbb{B})$ defined in Equation (3.3.1) is mapped as a TS by teasing out the involved transitions of behaviors. Below, we illustrate this transition-system-based behavior modeling for intra-coupled behaviors in the robot soccer game via logic and temporal relationships, in particular for robots RC and Ord.

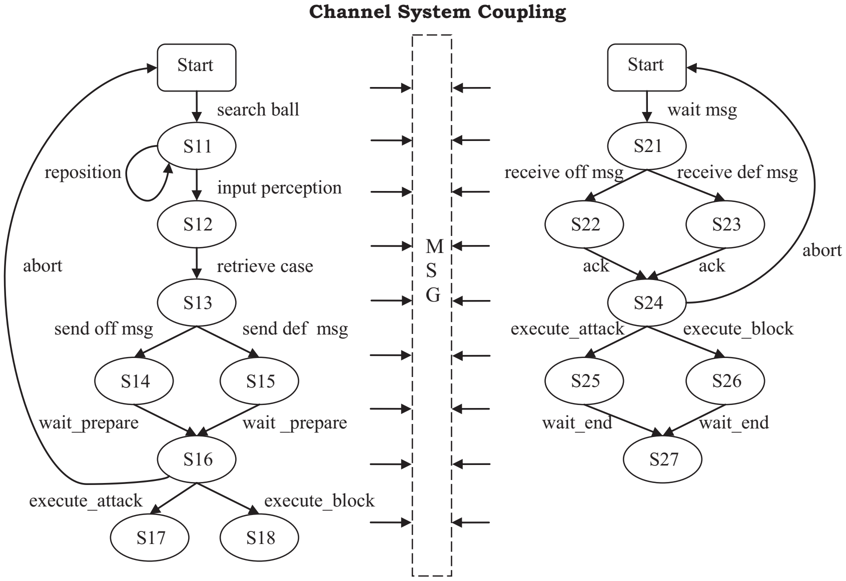


Figure 3.5: Transition system models $TS(\mathbb{B}(RC_p))$ and $TS(\mathbb{B}(Ord_q))$.

Example 3.4.1 (Intra-coupling of robots) *The robot RC plays the roles of both retriever and coordinator. Specifically, RCs first search for a ball on the field; they then retrieve cases from the case space according to the current situation and the coordinations among all the RCs; subsequently, they send either a* def *message or an* off *message to other ordinary robot players for their information, which is followed by the coordination by RCs of all robot players. All the operations of the pth* $(1 \leq p \leq k)$ *retriever* RC_p *and their precedence relationships are shown in the left side of Figure 3.5, which reveals the logic and temporal interactions among operations conducted by the retriever robot* RC_p*.*

The ordinary robot Ord first waits for a message from one RC; it then executes the corresponding action after sending acknowledgement to its coordinator RC. However, it is entitled to abort the whole execution if a case is lost, the message is lost, or time is out. Since all the robot players Ords have the same action procedures, their transition systems are equal but have different implementation orders. The entire intra-coupled behaviors of the *qth ($k+1 \leq q \leq n$) robot Ord_q are indicated on the right side of Figure 3.5, i.e., how the robot player Ord_q acts individually. All the operations involved consist of* {search ball, reposition, input perception, retrieve case, send off msg, send def msg, *wait_prepare*, *execute_attack*, *execute_block*, wait msg, receive off msg, receive def msg, ack, *wait_end*, abort}*. All the states shown in this figure refer to the corresponding intra-coupling functions, i.e., S_{15} denotes function θ_{15}, S_{16} denotes function θ_{16}, and the arrow → in between represents $\odot_a$.*

3.4.2 Inter-coupled Aggregation

Apart from the intra-coupled behaviors, inter-coupling (i.e. $\mathbb{B}^{\eta}(\mathscr{A}_i)$) refers to interactions between operations by different actors. On the basis of the temporal, inferential, and party-based couplings discussed in Section 3.3.1, the specific semantic mapping about the inter-coupling is defined as follows. Note that all the notations in this definition are referred to Section 3.3.1.

Definition 3.4.1 (Inter-coupling Function) *The behavior inter-couplings are essentially the various interactions among multiple behaviors. Let $\mathbb{B}_1$ and $\mathbb{B}_2$ be two behaviors, then the* ***inter-coupling function*** *$\eta_i(\mathbb{B})$ is defined as*

$$\eta_i(\mathbb{B}) ::= \mathbb{B}_1;\mathbb{B}_2 \mid \mathbb{B}_1\|\mathbb{B}_2 \mid \mathbb{B}_1 : \mathbb{B}_2 \mid \mathbb{B}_1|||\mathbb{B}_2 \mid \mathbb{B}_1|\mathbb{B}_2 \mid \mathbb{B}_1 \rightarrow \mathbb{B}_2| \; \mathbb{B}_1 \wedge \mathbb{B}_2 \mid \mathbb{B}_1 \vee \mathbb{B}_2 \mid \mathbb{B}_1 \otimes \mathbb{B}_2 \mid f(\mathbb{B}_1)^{[\mathscr{A}_1]}. \tag{3.4.1}$$

Regarding Equation (3.3.3), the connection operator $\odot_e$ is interpreted as the composition of the above functional symbols. Here, we consider the interactions between different actors from the perspectives of temporal, inferential and party-based aspects. Though they may not describe the complete situations, they are sufficient to characterize the basic features of our concerned behavior systems: dynamic, concurrent, logic, and hierarchical. Specifically, the temporal operators (*serial*, *synchronous*, *interleaving*, *shared-variable*, and *channel system*) induce a dynamic behavior system with concurrent processes (Baier & Joost 2008); the inferential operators (*causal*, *conjunction*, *disjunction*, and *exclusive*) take into account the logic preferences; and part-based operators (*OPMO*, *MPOO*, and *MPMO*) facilitate a hierarchical integrated behavior system. For instance, stock market and multi-agent systems both possess these features which mostly concern us during analysis.

Below, we illustrate inter-coupled behaviors in the robot soccer game.

Example 3.4.2 (Inter-coupling of robots) *In the robot soccer game, RC interacts with Ords over channels by sending execution, acknowledgement, and abort messages, shown in the middle part of Figure 3.5; the retrievers RCs communicate with one another by the way of shared-variable coupling, in which the case to be chosen is the shared variable; while the robots Ords interact with one another in an interleaving way, i.e., they act individually and independently from other Ords. For several operations such as* $\{execute\_attack, execute\_\ block, abort\}$*, robots RCs and Ords interact synchronously through a channel system with capacity 0.*

Formally, the inter-relationships among all robots are:

$$\begin{aligned} &((\mathbb{B}(RC_1)|\mathbb{B}(Ord_{k+1}))|||\cdots|||(\mathbb{B}(RC_k)|\mathbb{B}(Ord_{k+1}))) \\ &\quad : ((\mathbb{B}(RC_1)|\mathbb{B}(Ord_{k+2}))|||\cdots|||(\mathbb{B}(RC_k)|\mathbb{B}(Ord_{k+2}))) \\ &\quad :\cdots: ((\mathbb{B}(RC_1)|\mathbb{B}(Ord_n))|||\cdots|||(\mathbb{B}(RC_k)|\mathbb{B}(Ord_n))). \end{aligned} \tag{3.4.2}$$

Here, the set $\{\eta_i(\mathbb{B})\}_{1\leq i\leq n}$ *of inter-coupling functions is indicated as a collection of functional symbols* $\{|, |||, :\}$*, and the connection operator* $\odot_e$ *composes those functions.*

In fact, the three communication strategies, i.e. temporal, inferential, and party-based couplings, are embedded with intra-coupled and inter-coupled aggregations in different ways by taking advantage of either transition system semantics or communication operators. Furthermore, combined aggregation integrates intra-coupled and inter-coupled couplings to systematically represent the interactions between behaviors.

3.4.3 Combined Aggregation

With the intra-coupled and inter-coupled interactions defined, we develop the combined aggregation of the qualitative coupled behaviors to model complex behavior-oriented applications. This is achieved through three steps: *behavior combination*, *rule reduction*, and *TS conversion* to represent the complex combined interactions of qualitative coupled behaviors, namely $h(\theta_{j1}(\mathbb{B}), \theta_{j_2}(\mathbb{B}), \eta_{i_1}(\mathbb{B}), \eta_{i_2}(\mathbb{B}))$. In addition, the connection operator $\odot_c$ in Equation (3.3.4) is interpreted to concatenate these three steps.

First, we consider the extension of behavior sequences towards hierarchical and hybrid combinations, in which behaviors are associated in a hierarchical structure that consists of different relationships, e.g., $f(\mathbb{B}_1, g(\mathbb{B}_2, \mathbb{B}_3)) = \{\mathbb{B}_1; (\mathbb{B}_2 \| \mathbb{B}_3)\}$ indicates that behavior $\mathbb{B}_1$ is followed by the handshaking (i.e. $g(\cdot)$) of $\mathbb{B}_2$ and $\mathbb{B}_3$; $\{f(\mathbb{B}_1).g(\mathbb{B}_2)\}$, $\{f\ (\mathbb{B}_1)^\star\}$, and $\{f(\mathbb{B}_1)^\omega\}$ characterize the concatenation of $\mathbb{B}_1$ and $\mathbb{B}_2$, finite repetition of $\mathbb{B}_1$, and infinite iteration of $\mathbb{B}_1$, respectively. For example, some robots execute their actions sent from the coordinator, while other robots often abort executions due to timeouts.

Second, *interaction rules* (IR) are induced to support appropriate combinational reduction of multiple coupling relationships. Formally, a rule IR is defined in terms of SOS-notation (Baier & Joost 2008) as follows.

Definition 3.4.2 (Interaction Rule) *An* ***interaction rule***

$$IR : \mathbb{B}_1 \times \cdots \times \mathbb{B}_n \to \frac{f(\mathbb{B}_1, \cdots, \mathbb{B}_n)}{g(\mathbb{B}_1, \cdots, \mathbb{B}_n)} \tag{3.4.3}$$

is the combinational equivalence and reduction about the coupling relationships among behaviors $\mathbb{B}_i(1 \le i \le n)$, where $f(\cdot)$ and $g(\cdot)$ are two coupling expressions for the involved behaviors.

In the above SOS-notation based interaction rule (see 3.4.3), if the numerator formula holds, then the denominator part holds as well. With interaction rules, we can perform reasoning about behaviors to simplify and conclude critical rules as in (Kazakov 2009). For instance, four interaction rules are induced as follows (where $*, *_1, *_2$ are the coupling operators discussed in (3.4.1) in Section 3.3.1).

$$IR_1 : \frac{(\mathbb{B}_1 * \mathbb{B}_2) * \mathbb{B}_3}{\mathbb{B}_1 * (\mathbb{B}_2 * \mathbb{B}_3)}, \tag{3.4.4}$$

$$IR_2 : \frac{(\mathbb{B}_1 *_1 \mathbb{B}_2) *_2 (\mathbb{B}_3 *_1 \mathbb{B}_2)}{(\mathbb{B}_1 *_2 \mathbb{B}_3) *_1 \mathbb{B}_2}, \tag{3.4.5}$$

$$IR_3 : \frac{(\mathbb{B}_1 * \mathbb{B}_2) * (\mathbb{B}_2 * \mathbb{B}_3)}{\mathbb{B}_1 * \mathbb{B}_2 * \mathbb{B}_3}, \tag{3.4.6}$$

$$IR_4 : \frac{(\mathbb{B}_1 *_1 \mathbb{B}_2) *_2 (\mathbb{B}_1 *_1 \mathbb{B}_3)}{\mathbb{B}_1 *_1 (\mathbb{B}_2 *_2 \mathbb{B}_3)}. \tag{3.4.7}$$

IR_1 reveals the associative law, IR_2 and IR_4 specify the distributive law, and IR_3 describes the absorption law. They are all the basic equivalence rules for induction and reduction of behavior aggregation and reasoning (reasoning about behaviors is not the focus in this chapter).

Finally, concurrent transition systems (TSs) are constructed to specify complex interactions by utilizing temporal, inferential, and party-based couplings to describe, combine and aggregate the coupling relationships, as specified in Section 3.3.1. The relationships among TSs are concerned since complex behaviors are represented as TSs. Assume that there are n complex behaviors (TSs) associated with one another in terms of different coupling relationships, which are further represented as follows:

- *Serial Coupling*: $TS_1; TS_2; \cdots ; TS_n$
- *Synchronous Coupling*: $TS_1 \parallel TS_2 \parallel \cdots \parallel TS_n$
- *Interleaving Coupling*: $TS_1 : TS_2 : \cdots : TS_n$
- *Shared-variable Coupling*: $TS_1 ||| TS_2 ||| \cdots ||| TS_n$
- *Channel System Coupling*: $TS_1 \mid TS_2 \mid \cdots \mid TS_n$
- *Causal Coupling*: $TS_1 \rightarrow TS_2$
- *Conjunction Coupling*: $TS_1 \wedge TS_2$

- *Disjunction Coupling*: $TS_1 \vee TS_2$
- *Exclusive Coupling*: $TS_1 \otimes TS_2$
- *Hierarchical Coupling*: $f(g(TS_1, TS_2, \cdots, TS_n))$
- *Hybrid Coupling*: $f(TS_1).g(TS_2)$, $f(TS_1)^\star$, $(TS_1)^\omega$
- *OPMO Coupling*:$f(TS_1, TS_2, \cdots, TS_n)^{[A_1]}$
- *MPOO Coupling*: $f(TS_1)^{[A_1 A_2 \cdots, A_n]}$
- *MPMO Coupling*: $f(TS_1, TS_2, \cdots, TS_n)^{[A_1 A_2 \cdots A_n]}$

The combined aggregation of coupled behaviors reflects the semantics of behavior coupling and interaction. Following Example 3.4.2, the instance below illustrates the combined behavior aggregation in the soccer game.

Example 3.4.3 (Robot behavior aggregation) *All the behavior couplings involved in (3.4.2) are aggregated and formalized as the following interactions (3.4.8) between the qualitative coupled behaviors according to the right and left distributive laws (IR_2 and IR_4, respectively), and the TS transformation:*

$$\begin{aligned}(TS(\mathbb{B}(RC_1))|||TS(\mathbb{B}(RC_2))|||\cdots|||TS(\mathbb{B}(RC_k))) \\ |(TS(\mathbb{B}(Ord_{k+1})) : TS(\mathbb{B}(Ord_{k+2})) : \cdots : TS(\mathbb{B}(Ord_n))).\end{aligned} \tag{3.4.8}$$

Here, the coupling function $h(\theta(\mathbb{B}), \eta(\mathbb{B}))$ is embodied by the above Equation (3.4.8), and the connection operator $\odot_c$ is regarded as the whole process (i.e. three steps) to obtain the final transformation result.

3.5 Case Study: Representing Complex Behavior Systems

In the above, we detail building blocks for representing group behavior interactions, and illustrate them in terms of specific examples in the case study

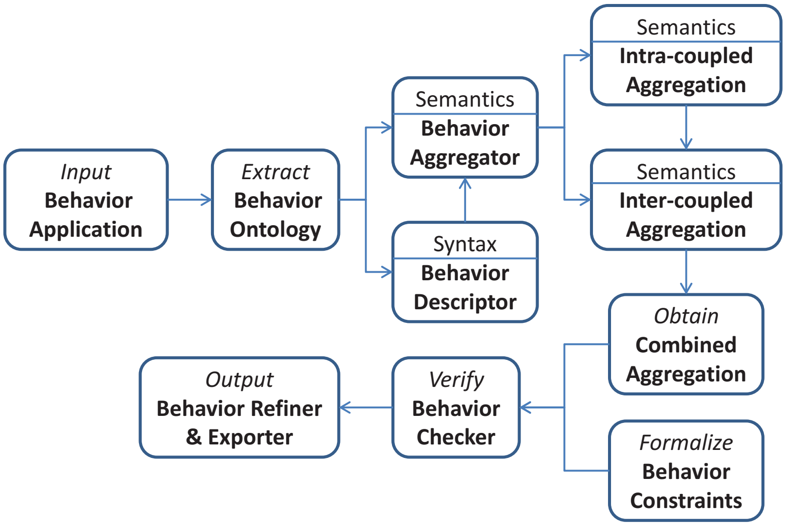


Figure 3.6: Group behavior representation and verification procedures.

system: multi-robot architecture (Figure 3.2). This section provides a more complete case study, showing the use of the proposed modeling techniques for representing a complex behavior system.

Figure 3.6 illustrates the process of applying *OntoB* for representing complex behavior interactions. After a *behavior ontology* is extracted from the given *behavior-related application*, the *behavior syntax* is explicated by visualizing and formalizing the *behavior descriptor*. Based on the syntax formalization, the *behavior semantics* are interpreted through the *intra-coupled aggregation*, *inter-aggregation*, and then *combined aggregation* to enable verification. Subsequently, the *behavior checker* is triggered to conduct the verification with the obtained *combined aggregation* by engaging the formalized *behavior constraints*. Finally, a desired behavior representation system is provided by the *behavior refiner and exporter*. This process guides the building of a representation system for a complex behavior application. Below, we illustrate the above process and the use of *OntoB* to represent a specific scenario in the robot soccer game.

The robot architecture presented in Figure 3.2 can be simplified as a special case-based multi-robot system (Robocup soccer competition) with four

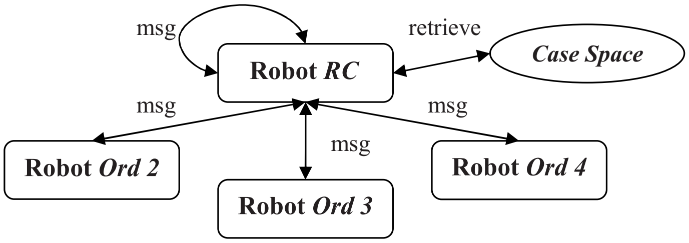


Figure 3.7: A simplified case-based multi-robot system.

robots and one retriever, as shown in Figure 3.7. As pointed out in Section 3.2.2, the multi-robot architecture consists of n robots, which include k retrievers. There must be at least one retriever in this system, for simplicity, Ros and Veloso (Ros & Veloso 2007) only consider one retriever and three ordinary robot players, i.e. $k = 1$ and $n = 4$. In detail, as described in Figure 3.7, robot RC firstly retrieves a case from the case space and then informs the rest of the Ords robot players. Once the robots Ords successfully receive the messages from RC, they send acknowledgments back to the retriever RC for confirmation. As clarified in Section 3.2.2, the RC also coordinates all the other robot players including itself to defeat the opponent. All the robots, no matter RC or Ord, could abort the executions at any moment if timeout expires, or messages or cases are lost in the interactions. Although this is a simplified case of the multi-robot architecture proposed in Section 3.2.2, the inherent behavior system still exhibits several complex features, e.g. distributed behaviors, concurrent actions, uncertain situations, collaborated strategies, and nonstop operations. Our purpose here is to demonstrate that our proposed formalization techniques and the modeling procedures depicted in Figure 3.6 can be adapted accordingly to capture the complexities in this scenario (Ros & Veloso 2007).

The given behavior-oriented application is the case-based multi-robot soccer game with four players in each team, as shown in Figure 3.7. At first, we extract the *behavior ontology*, including *actor*, *operation*, and *coupling*, for this behavior system. In terms of the *Visual Behavior Descriptor*, it is trivial that the *actors* involved are the robots including the retriever robot and three ordinary robot players, i.e., $\mathscr{A} = \{RC, Ord_2, Ord_3, Ord_4\}$. The retriever RC plays the role of a coordinator which also retrieves cases from the case space for the problem solving; while the other robots Ords just wait for the commands from the retriever RC, implement the received orders, and send acknowledgment back to the retriever RC. These are the main *operations* involved. The *couplings*, i.e., interactions, between different actions of the same robot (namely *intra-coupled behaviors* $\mathbb{B}^{\theta}(\mathscr{A}_i)$), and between actions of distinct robots (*inter-coupled behaviors* $\mathbb{B}^{\eta}(\mathscr{A}_i)$), are the sequential or parallel operations of one robot and the communications among them, respectively. Specifically, the retriever RC interacts with the other robots Ords through channel systems (i.e., $\mathbb{B}(RC)|\mathbb{B}(Ord)$), and the other robots Ords are coupled with one another independently and take their own actions (i.e., $\mathbb{B}(Ord_i) : \mathbb{B}(Ord_j)$, where $2 \leq i, j \leq 4$).

Accordingly, by taking advantage of the *Formal Behavior Descriptor* (3.3.4), the *syntax* of qualitative coupled behaviors between retriever RC and ordinary players Ords can be represented as:

$$\mathbb{B}(RC, Ords) = (\mathbb{B}^{\theta^{(RC)}})^{\eta^{(RC,Ords)}} * (\mathbb{B}^{\theta^{(Ords)}})^{\eta^{(RC,Ords)}}, \tag{3.5.1}$$

where $\mathbb{B}^{\theta^{(RC)}}$ and $\mathbb{B}^{\theta^{(Ords)}}$ refer to the *intra-coupled behaviors* of retriever RC and robot players Ords, respectively; $(\cdot)^{\eta^{(RC,Ords)}}$ indicates the *inter-coupled behaviors* between the RC and each Ord; and $\mathbb{B}(RC, Ords)$ means the *combined behaviors* between RC and each Ord, which are built on top of the former two types of behaviors.

To facilitate the purpose of verification, the syntax of behavior descriptor is further given the *semantic* explanations by using behavior aggregations. With respect to the *Behavior Aggregator*, syntax symbols $\theta^{(RC)}$ or $\theta^{(Ords)}$, $\eta^{(RC,Ords)}$, and $*$ are interpreted as *intra-coupled aggregation*, *inter-coupled aggregation*, and *combined aggregation* respectively to enable behavior checking. In detail, behaviors of the retriever RC (*intra-coupled aggregation* $\theta^{(RC)}$) can be modeled as $TS(\mathbb{B}(RC))$ (since only one retriever RC is involved, this transition system is independent with itself), and the qth ($2 \leq q \leq 4$) robot player's (Ord_q) behaviors (*intra-coupled aggregation* $\theta^{(Ord_q)}$) can be represented as $TS(\mathbb{B}(Ord_q))$ in an interleaving way. The specific transition systems are described in Figure 3.5, where $p = 1$ and $n = 4$. Based on the types of couplings revealed by the *Visual Behavior Descriptor* above, we get the formal *inter-coupled aggregation* $\eta^{(RC,Ords)}$ of all the robots as follows:

$$(\mathbb{B}(RC)|\mathbb{B}(Ord_2)) : (\mathbb{B}(RC)|\mathbb{B}(Ord_3)) : (\mathbb{B}(RC)|\mathbb{B}(Ord_4)). \qquad (3.5.2)$$

Further, followed by three steps, i.e. behavior combination, rule reduction (IR_4 here), and TS conversion, the *combined aggregation* $h^{(RC,Ord)}$ can be abstracted as the following expression:

$$TS(\mathbb{B}(RC))|(TS(\mathbb{B}(Ord_2)) : TS(\mathbb{B}(Ord_3)) : TS(\mathbb{B}(Ord_4))). \qquad (3.5.3)$$

Subsequently, the pre-formalized *Behavior Constraints* and the obtained *combined aggregation* are entered into the *Behavior Checker*, and then the *Behavior Refiner and Exporter* are exemplified to follow this specific case study based on (Ros & Veloso 2007) to verify and refine potential problems in it. The required constraints are to be illustrated in Section 3.6 below.

The above process shows that our system *OntoB* can effectively model the scenario based on (Ros & Veloso 2007) of the multi-robot architecture. The proposed formalization techniques can effectively model complex behavior-oriented applications exhibiting similar features, such as distributed behaviors, concurrency, uncertainties, and nonstop operations.

3.6 Behavior Constraint Indicator

In order to improve the robustness and stability of behavior model, a simulation can be conducted prior to the behavior checking. For the verification purpose, the behavior model under consideration needs to be accompanied by a relevant constraint specification that is to be verified. Constraints, i.e. requirements of prior simulations, can be used effectively to get rid of the simpler categories of modeling errors. For instance, a business constraint in stock markets is that investors are not allowed to make transactions after trading hours. To make a rigorous verification possible, constraints should be described in a precise and unambiguous manner. This is typically done through a constraint specification language.

We take advantage of the propositional logic and temporal logic to express the constraints of the desired model. The temporal logic (Baier & Joost 2008) is basically an extension of traditional propositional logic with operators that refer to the behaviors of systems over time. It allows for the specification of a broad range of relevant behavior constraints such as functional correctness, reachability, safety, liveness, fairness, and real-time properties. The underlying nature of time in temporal logic can be either linear (Linear Temporal Logic, LTL) or branching (Computation Tree Logic, CTL). At this stage, we mainly focus on LTL; the involved temporal modalities are basic operators, such as *next* ($\bigcirc$), *until* ($\cup$), *eventually* ($\Diamond$), and *always* ($\Box$). CTL can then be considered for extension. LTL may be used to express the timing for the class of synchronous and asynchronous behavior couplings in which all components proceed in a lock-step fashion. In LTL, one can encode formulae about the future of paths such that a condition will eventually be true, and a condition will be true until another fact becomes true, and so on.

Here, in our consideration of the behavior constraint indicator, the focus is essentially on four important, though relatively simple, types of constraint: *Ontology Axiom*, *Inferential Coupling*, *Desired Constraint*, and *Forbidden Constraint* (Baier & Joost 2008, Breitman et al. 2007). Different constraints can be categorized into these four classes with distinct emphasis.

- *Ontology Axiom*: The predefined ontology axiom that must be satisfied by a behavior ontology model.
- *Inferential Coupling*: The behavior coupling relationships on inferential reasoning, including causal, conjunction, disjunction, and exclusive couplings.

- *Desired Constraint*: Features of the desired patterns which are supposed to be satisfied by the behavior.

- *Forbidden Constraint*: Wicked characteristics that the behavior should get rid of.

Specifically, the ontology axiom is the constraint that the TS interpretation should consider, otherwise there will be problems with this inherent ontology. Consequently, if we build a set of TS models of behavior ontology, this kind of axiom should be satisfied absolutely. The constraints induced by inferential couplings are directly formalized according to logic expressions. The desired constraint is the property that the transition system ontology need to have. If the constraint is satisfied, the behavior ontology is rather stable. Otherwise, risky factors that require further exploration must exist. The forbidden constraint is a property we do not want to see. We are not sure about the result of either of these two latter types; therefore, the relevant analysis is critical.

In the robot soccer game, the constraints and their formal expressions are presented as follows.

Example 3.6.1 (Robot constraint) *Based on the categories of constraints, note that CR represents the coordinator robot while Ord_i denotes an individual robot player as previously indicated.*

C.1 $\Box(\neg(execute\_attack^{[Ord_i]} \wedge execute\_block^{[Ord_i]}))$

It is never the case that any Ord can implement the executions of attack and simultaneously block opponent players.

C.2 $\Box((TS(retrieve\ case)^{[CR]} \wedge \bigcirc TS(send\ msg)^{[CR]}) \rightarrow \Diamond(TS(receive\ msg)^{[Ord_i]} \wedge \bigcirc TS(ack)^{[Ord_i]}))$

If the case is successfully retrieved by CR, then eventually the message sent is received and the acknowledgment is sent by Ord.

C.3 $\Box(wait\_end^{[Ord_i]} \cup (\wedge_{j \neq i}) wait\_end^{[Ord_j]})))$

The execution of a case will not be done until all Ords have completed their actions.

C.4 $\Box\Diamond(\vee_i abort^{[Ord_i]})$

Ord will infinitely often abort the execution.

Within these constraints, constraint *C.1* (*Ontology*) and constraint *C.3* (*Desired*) reflect the safety properties, which are *always* or *never* claims, justifying that something bad never happens; while constraint *C.2* (*Inferential*) and constraint *C.4* (*Forbidden*) are categorized into the liveness properties, i.e., *eventually* claims, specifying that something good will eventually happen in the future. The safety properties are violated in finite time, whereas the liveness properties are destroyed in infinite time. The constraints stated above are just a few examples to illustrate how behavior checking is managed in Section 3.7.

3.7 Behavior Checker

There are different types of formal verification, from the manual proof of mathematical arguments (Wang et al. 2000) to interactive computer aided theorem proof (Kaufmann et al. 2000), and automated model checking (Baier & Joost 2008). Manual proofs are time-consuming, error-prone, and often not economically viable. Computer-aided theorem provers require significant expert knowledge. In contrast, model checking (Baier & Joost 2008) is an automated technique that, given a finite-state model of a system and a formal property, can systematically check whether or not this property holds for that model. If not, model checkers can help to identify the input sequence that triggers the failure. This verification technology provides an algorithmic means of determining whether an abstract model—representing, for example, a hardware or software design—satisfies a formal specification. Model checking can handle the nondeterminism caused by uncertain communication delays and unknown parameters when concurrent agents interact with one another, besides which, automation makes model checking far more attractive to multi-agent systems. In *OntoB*, we take advantage of model checking to verify the behavior systems.

To test the expressiveness and stability of our model, we design a behavior model checker to analyze those predefined constraints. We use SPIN (Simple Promela Interpreter) (Holzmann 2003) for the checking. SPIN is a typical LTL model checker, highly recommended for automation, especially in fields such as security protocol verification, control system verification, software verification, and optimization schedule.

The behavior ontology model and behavior constraints introduced above are mapped to TS and linear temporal logic (LTL) formulae, respectively. Formally, for temporal couplings, the corresponding TS semantics are revealed to unfold the compositions (i.e. *serial*, *interleaving*, *shared-variable*, *channel system*, and *synchronous couplings*) of TSs into a single TS. The conversion rules can be obtained by the corresponding mappings of transition system semantics on modeling concurrent systems in (Baier & Joost 2008). For the party-based couplings, the relevant TSs can be taken into account based on different groups to form intra-coupled interactions separately. Figure 3.8 illustrates the steps performed for the *OntoB* system as an integrated framework, and is specified below.

In particular, if the given constraints are *ontology axioms*, *inferential couplings*, or *desired constraints*, the verification process follows the corresponding LTL formulae. A counter example will be produced when the constraints are not satisfied. Refinement is made to improve this model where appropriate. The verification process with *forbidden constraints* is conducted as per the negative forms of LTL formulae to check whether any path exists that complies with the unwanted constraints. For instance, as stated in Section 3.6, the forbidden constraint 4 "Ord will abort the execution infinitely often" is actually checked with the form $\neg(\Box\Diamond(\vee_i abort^{[Ord_i]}))$.

Once the behavior-related application and its constraints are given by a domain expert, we use the proposed formal components to transform them

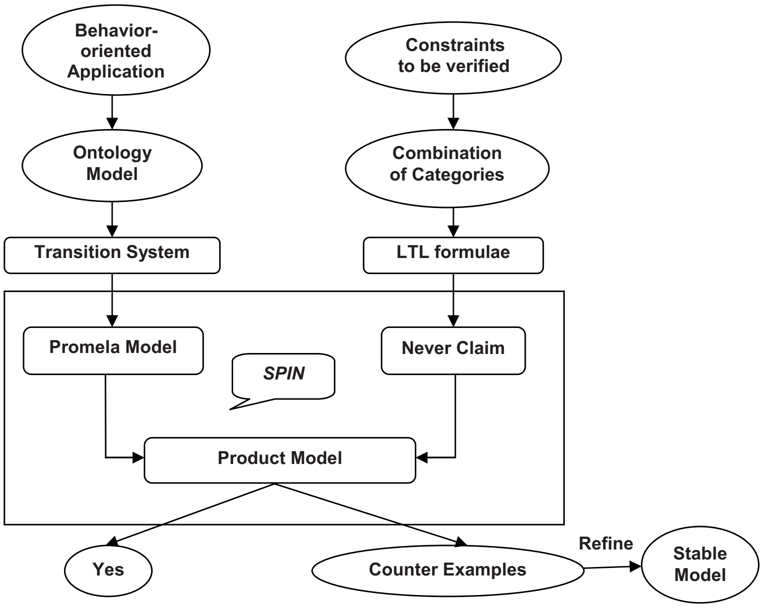


Figure 3.8: Ontology-based Qualitative Coupled Behavior Modeling and Checking (*OntoB*).

into a behavior ontology model and the combination of relevant categories of relationships, respectively. In the next stage, TS and LTL are further explored. TS is then translated to Promela using the approach discussed in (Baier & Joost 2008). Note that the language Promela (process metalanguage) is the input language for the prominent model checker SPIN (Holzmann 2003). Promela supports communication over shared variables and message passing along either synchronous or buffered FIFO-channels. Simultaneously, a *never* claim is made to indicate the negative form of LTL formula. A product model is built from the combination of the Promela model and *never* claim formalizations. These three steps are achieved within SPIN. Accordingly, the outputs of SPIN are two alternative answers: one is "Yes" which means there is no activity or action sequence dissatisfied with the constraint or ¬ constraint; while in contrast, counter examples are given for refinements if any mistakes exist.

The robot soccer game is a behavior-oriented application. The following

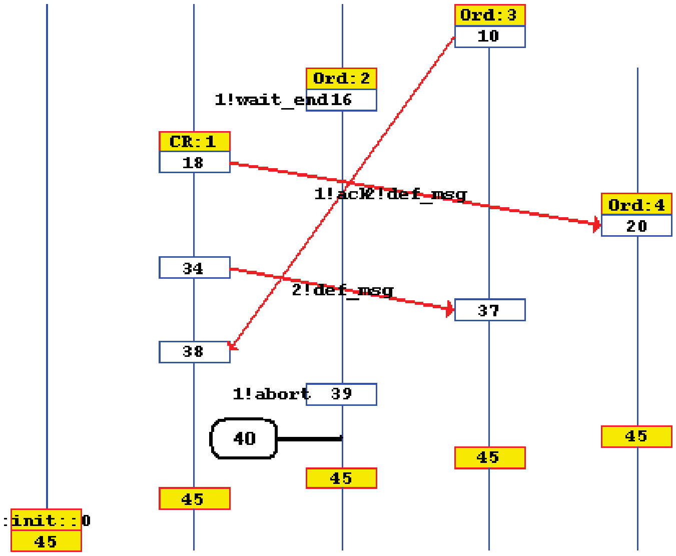


Figure 3.9: A scenario in a soccer game that causes deadlock.

example shows the detailed checking process of a related case study in the robot soccer game with the given constraints indicated in Section 3.6.

Example 3.7.1 (Robot behavior checking) *We conduct the verification of the formalized robot soccer game system. SPIN is used to perform checking of the corresponding $TS(\mathbb{B})$ and constraints. Subsequently, we obtain the results that the constructed TS satisfies the first three constraints, while a counter example arises for the last forbidden constraint $\Box\Diamond(\vee_i abort^{[Ord_i]})$. The graphical interface of the counter example process with XSPIN (Holzmann 2003) is shown in Figure 3.9, which is based on a Message Sequence Chart window of XSPIN. The vertical lines represent robot behaviors, boxes represent states, and arrows represent messages sent. Behavior 0 (init : 0) does nothing but initiates the behaviors of $RC(1)$, $Ord_i(i)$ $(i = 2, 3, 4)$. Formally, we have the following states:*

- *State 10: $ack^{[Ord_3]} \rightarrow wait\_prepare^{[RC]}$*
- *State 18: send def $msg^{[RC]} \rightarrow wait$ $msg^{[Ord_4]}$*
- *State 34: send def $msg^{[RC]} \rightarrow wait$ $msg^{[Ord_3]}$*

- State 39: $\Box(receive\ msg^{[Ord_2]} \rightarrow abort^{[Ord_2]})$

- State 45: $\Box(\wedge_i wait\_end^{[Ord_i]} \wedge wait\_prepare^{[RC]})$

At State 39, the robot player Ord_2 aborts the execution whenever it receives messages from RC. Consequently, at State 45, Ord_2 and RC wait for each other, resulting in an infinite wait loop while the executions of other robots are interrupted simultaneously, which is the so-called deadlock. A typical deadlock scenario occurs when components mutually wait for each other to progress.

Deadlocks are one of the most common problems in multi-agent systems, and they are often difficult to find and reproduce. In the following section, we will refine the model to resolve the deadlock.

3.8 Behavior Model Refiner and Exporter

When behavior checking is done, as indicated in the general framework of *OntoB* (see Figure 3.2.1), there are two options from which to choose. One is to provide the behavior model directly by the behavior model exporter if every constraint is satisfied; and the other is to revise this model by the behavior model refiner according to the errors pointed out by the behavior checker, and then deliver the modified model via the behavior model exporter. The behavior model refiner plays the role of improving the behavior model and make it stable and robust, while the behavior model exporter outputs the desired model to end the entire behavior modeling and checking system. It is encouraging that even the subtle errors or mistakes that remain unrevealed via emulation, testing and simulation can potentially be discovered by using the behavior checker (Baier & Joost 2008). Accordingly, we can replay the violating scenario with a simulator, in this way obtain useful debugging information and then adapt the original model.

Below, we continue with the robot soccer game to explain the refinement and exportation of modeling robot behaviors.

Example 3.8.1 (Robot model refiner and exporter) *After analyzing the deadlock scenario in Example 3.7.1, we introduce an additional state called "hold_on" to break the loop, formalized as the follow formula:*
- State 40: State 39$\rightarrow$ *hold_on*$^{[Ord_i] \vee [RC]}$

When such a deadlock happens, the next state will be "hold_on", which means that the other two robot players Ord_3 *and* Ord_4 *will continue their execution as usual. RC continues to retrieve cases and send messages without receiving ack from* Ord_2 *until the behaviors of* Ord_2 *become normal. If this does not occur, there must be design flaws in* Ord_2*, which should be explored by robot experts. In fact, "State 40" serves as a* Behavior Model Refiner *in* OntoB*. It refines the design by correcting the errors identified by the formal verification, and guarantees a robust and stable multi-agent system to satisfy all the required constraints.*

Finally, a refined system (in addition with State 40) will be provided by the Behavior Model Exporter *to ensure that all the given constraints are valid in this multi-agent system, and to end the Ontology-based Qualitative Coupled Behavior Modeling and Checking (*OntoB*) system as the last component. If the original model has no problem with the required constraints, this block will output the model directly with a rather high confidence. In this example, the behavior model exporter delivers the desired system without any revision in terms of the first three constraints, since they all have been satisfied already.*

3.9 Summary

In this chapter, we have presented a generic and robust system, i.e. an Ontology-based Qualitative Coupled Behavior Modeling and Checking (*OntoB*) system, for modeling and checking complex couplings among behaviors for both individual and group actors. Unlike existing behavior representation systems, *OntoB* consists of comprehensive and solid components for modeling and verifying behavior elements, couplings, aggregations, and constraints. It delivers a generic behavior ontology model to capture behavioral elements, as well as building blocks for combining and aggregating behavior intra-couplings and inter-couplings for modeling complex interactions in behavior-oriented applications. The intra-coupled aggregations are specified in terms of transition systems, while inter-coupled aggregations are depicted from temporal, inferential, and party-based perspectives. The Qualitative coupled behaviors are modeled as the combined aggregations in terms of behavior combination, rule reduction and TS conversion. The behavior ontology and constraints are transformed to a transition system and logic form respectively to verify and refine behavior models. *OntoB* eventually outputs a refined and stable behavior model after verification by the model checker.

We have exemplified the successful use of *OntoB* in modeling and checking of multi-robot behaviors in the robot soccer game in terms of both visual and formal modeling by the proposed representation modules and verification in SPIN. Although the case study looks simple, it embodies complex behavior relationships and primary characteristics of behaviors and group behavior interactions in a typical behavior-oriented application, including distributed behaviors, concurrency, uncertainties, and nonstop operations. *OntoB* system will be useful for research peers to apply the proposed formalisms and verification techniques in group behavior interactions, such as multi-agent configuration, stock market analysis, transportation control.

[***Note***] *Two Preliminary versions of this chapter have been published in the first two items below, and a full journal version has been submitted to the third item.*

- ***Can Wang****, Longbing Cao (2012), "Modeling and Analysis of Social ctivity Process". Behavior Computing: Modeling, Analysis, Mining and Decision, Springer, pp. 21-35.*
- ***Can Wang****, Longbing Cao (2010), "SAPMAS: Social Activity Process Modeling and Analysis System". The International Workshop on Behavior Informatics held in conjunction with The 14th Pacific-Asia Con-*

*ference on Knowledge Discovery and Data Mining (**BI with PAKDD 2010**).*

- ***Can Wang**, Longbing Cao (2013), "Formalization and Verification of Group Behavior Interactions". IEEE Transactions on Systems, Man, and Cybernetics–Part A: Systems (**TSMC-A**).*

Apart from the above papers, the following published conference paper is based on the work introduced in this chapter to address the multi-agent system configuration.

- *Chayapol Moemeng, **Can Wang**, Longbing Cao (2011), "Obtaining an Optimal MAS Configuration for Agent-Enhanced Mining Using Constraint Optimization". The 7th International Workshop on Agents and Data Mining Interaction held in conjunction with the 10th International Conference on Autonomous Agents and Multiagent Systems (**ADMI with AAMAS 2011**), pp. 46-57.*

Chapter 4

Numerical Coupled Behavior Analysis

In this chapter, we explore and analyze the quantitative coupled behaviors by revealing the coupling relationship between numerical properties. Here, numerical behavior refers to the behavior data described by continuous properties or attributes, and numerical coupled behaviors further examine the couplings/interactions among a group of numerical behaviors. Throughout the chapter, object and the entity of coupled behaviors are interchangeable, numerical attributes indicate the continuous properties of coupled behaviors, and accordingly attribute values denote the property values of coupled behaviors.

The usual representation of numerical behaviors is to formalize them as an information table, which assumes the independence of properties/attributes. In the real-world data, properties are more or less interacted and coupled via explicit or implicit relationships. Limited research has been conducted on analyzing such property interactions, which only describe a local picture of numerical property couplings in an implicit way.

A framework of the numerical coupled behavior analysis is introduced to capture the global dependency of continuous properties. The coupling of numerical properties integrates the intra-coupled interaction within a property (i.e. the correlations between attributes and their own powers) and inter-coupled interaction among different properties (i.e. the correlations between attributes and the powers of others) to form a coupled representation for numerical entities by the Taylor-like expansion. This work makes one step forward towards explicitly addressing the global interactions of continuous properties/attributes, verified by the applications in data structure analysis, clustering and classification. Substantial experiments on 13 UCI data sets demonstrate the coupled representation can effectively capture the global couplings of numerical properties and outperforms the traditional way, supported by statistical analysis.

4.1 Background and Overview

Real-world data sets predominantly consist of quantitative attributes in diverse domains (Saria, Duchi & Koller 2011), such as finance and bioinformatics. The usual recommendation of numerical data is to deliver it as an information table (Kaytoue, Kuznetsov & Napoli 2011), which is a basic knowledge representation framework comprising a table with columns designating "attributes" and rows designating "objects". Each table cell therefore stands for the value of a particular attribute for a particular object. This traditional representation scheme only describes each object by associated variables and assumes the independent and identical distribution (IIDness) of them.

The fragment data of Iris (Table 4.1) is an example that six plant objects are characterized by four numerical attributes (i.e. "Sepal Length", "Sepal Width", "Petal Length", and "Petal Width"), and divided into three classes. For instance, the petal width of plant object u_1 is 0.2cm, which does not reflect any interaction with other attributes. Based on this classical representation, many data mining techniques and machine learning tasks (Plant 2012, Li & Liu 2012) including clustering and classification have been performed. One of the critical parts in such applications is to study the

Table 4.1: A Fragment Example of Iris Data Set

Iris	**Sepal.L** ($\mathbf{a_1}$)	**Sepal.W** ($\mathbf{a_2}$)	**Petal.L** ($\mathbf{a_3}$)	**Petal.W** ($\mathbf{a_4}$)	**Class**
u_1	5.5 cm	4.2 cm	1.4 cm	0.2 cm	Setosa
u_2	5.0 cm	3.4 cm	1.5 cm	0.2 cm	Setosa
u_3	6.1 cm	2.9 cm	4.7 cm	1.4 cm	Versicolor
u_4	6.2 cm	2.2 cm	4.5 cm	1.5 cm	Versicolor
u_5	6.3 cm	2.7 cm	4.9 cm	1.8 cm	Virginica
u_6	6.0 cm	2.2 cm	5.0 cm	1.5 cm	Virginica

pairwise distance between plant objects. A variety of distance metrics have been developed for numerical data, such as Euclidean and Minkowski metrics (Gan et al. 2007). Since plant objects u_4 and u_6 have identical values of "Sepal.W" and "Petal.W", the normalized Euclidean distance between them is only 0.493, which is much smaller than that between u_4, u_3 (i.e. 0.950) and nearly half of that between u_6, u_5 (i.e. 0.982). It indicates that u_4 and u_6 stand a good chance to be clustered into the same group. However, in fact, u_4 and u_3 belong to "Versicolor", u_6 and u_5 are labeled as "Virginica". Similar cases can also be observed by the normalized Euclidean distance between u_3 and u_5 (i.e. 0.75), which is smaller than both the distances between u_3, u_4 and between u_5, u_6.

Both instances show that it is often problematic to analyze the numerical data by assuming all the continuous attributes are independent, while the traditional data representation schemes fail to capture the genuine couplings of attributes. In the real world, business and social applications such as investors in capital markets and members in social networking almost always see quantitative attributes coupled with each other (Cao et al. 2012). It is very in demand from both practical and theoretical perspectives to develop effective representation method for analyzing continuous variables by considering the relationships among attributes (i.e. non-IIDness of numerical properties). A conventional way to explore the interaction of continuous attributes is to measure the agreement of shapes between variables via Pearson's correlation coefficient (Gan et al. 2007). Nevertheless, it only caters for the linear relationship between two variables. More often, numerical variables are associated with each other via nonlinear relationships, such as exponential and logarithmic functions. Our motivation is to consider both linear and nonlinear relationship functions, such couplings among variables are called global interactions or global dependency. In contrast, any method to study either the linear relationship or some specific nonlinear function only captures a local picture of the coupling relationships among variables, such as the Pearson's correlation. For Table 4.1, if we adopt the method in (Kalogeratos & Likas 2012) by treating each correlation as the pairwise similarity entry, we then obtain the normalized Euclidean distance between u_4 and u_6 as 0.223, which is still smaller than that between u_4 and u_3 (i.e. 0.329) but only a little larger than that between u_6 and u_5 (i.e. 0.218). It means the coupling relationships are only partially revealed with limited improvement on the distance.

So based on the traditional information table, how to describe the global interactions with the least information loss? The idea of Taylor expansion inspires us that we can use a Taylor-like series to quantify the global dependency, since any analytic function can be approximated by its Taylor polynomials. Therefore, we propose to represent the global coupling relationships by Taylor-like expansion on attribute values, in which the Pearson's correlations between attributes and their extended powers (i.e., each extended attribute value is the power of the original one) play the role of function derivatives. From this perspective, the Pearson's correlation just reflects the first-order Taylor-like expansion of the global dependency; and the mutual information based attribute interdependency (Nazareth, Soofi & Zhao 2007) is a special case, since function *log* can be expressed by its Taylor series. For Table 4.1, the distance between plant objects is then revised by explicitly capturing the intrinsic correlations between attributes and their powers. That is to say, the greater difference in "Petal.L" is expected to remedy the little differences in other attributes since they are correlated significantly.

A detailed review of the related work on numerical behavior analysis can be found in Section 2.2. Though most of the current strategies are based on the hypothesis of IIDness, great efforts have been made to reveal the implicit interactions between properties, such as Pearson's correlation (Gan et al. 2007), rank-correlated measure (Calders et al. 2006), dependency clustering (Plant 2012). Nevertheless, no work that systematically and explicitly considers the global coupling relationships (i.e. non-IIDness) among continuous attributes has been reported .

Accordingly, this chapter proposes a framework of the coupled attribute analysis on numerical data to address the aforementioned research issues, without the IIDness assumption. We consider both the intra-coupled interaction within an attribute, captured by the correlations between every attribute and its own powers; and the inter-coupled interaction among different attributes, quantified by the correlations between each attribute and the powers of others. A coupled representation scheme is then introduced for quantitative objects to integrate the intra-coupled and inter-coupled interactions with the original information table representation via Taylor-like expansion in a global way. Finally, the proposed coupled representation method is compared with the traditional representation approach by applying data structure analysis, clustering and classification, revealing that the couplings of continuous attributes are essential to the learning applications.

The chapter is organized as follows. A framework of coupled attribute analysis is proposed in Section 4.2. Section 4.3 specifies the coupled interactions of numerical attributes. We formalize the coupled representation for objects in Section 4.4. The effectiveness of coupled representation is demonstrated in Section 4.5 with extensive experiments. Finally, we end this chapter in Section 4.6.

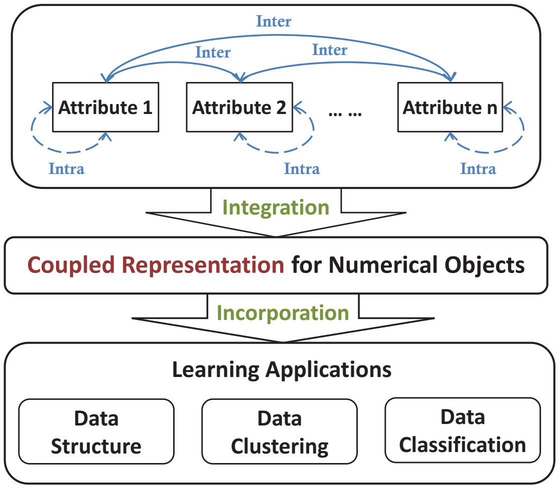


Figure 4.1: A framework of the coupled attribute analysis.

4.2 Framework of Coupled Attribute Analysis on Numerical Data

In this section, a framework of the coupled attribute analysis on numerical data is proposed in terms of the intra-coupled interaction within an attribute, the inter-coupled interaction among multiple attributes, and the integration of them.

The intra-coupled interaction is considered by examining the correlation between each numerical attribute and its extended powers. For the inter-coupled interaction, the relevant correlations between each continuous attribute and the extended powers of other attributes are studied. In the framework described in Figure 4.1, the couplings of numerical attributes are therefore revealed via the pairwise correlations between attributes and their powers. Further, those correlations are integrated to form a coupled representation for numerical data. Finally, three learning tasks are explored for the data structure, the data clustering, and the data classification by incorporating the coupled interactions, revealing that the couplings of continuous attributes are essential to the learning applications in empirical study.

Given a set of m objects $U = \{u_1, \cdots .u_m\}$ and a set of n continuous attributes $A = \{a_1, \cdots, a_n\}$, we specify each building block in this framework in the following sections.

4.3 Coupled Interactions of Attributes

The couplings of continuous attributes are proposed in terms of both intra-coupled and inter-coupled interactions. Below, the intra-coupled and inter-coupling relationships, as well as the integrated coupling, are formalized and exemplified.

The usual way to represent data is to use an information table $S =< U, A, V, f >$, where universe $U = \{u_1, \cdots, u_m\}$ consists of finite data objects; $A = \{a_1, \cdots, a_n\}$ is a finite set of continuous attributes; $V = \bigcup_{j=1}^{n} V_j$ is a collection of attribute value sets, in which $V_j = \{a_j.v_1, \cdots, a_j.v_{t_j}\}$ is the set of t_j attribute values from attribute $a_j (1 \leq j \leq n)$; and $f = \bigcup_{j=1}^{n} f_j$, $f_j : U \rightarrow V_j$ is an information function which assigns a particular value of attribute a_j to each object. For instance, Table 4.1 is an information table composed of six objects $\{u_1, \cdots, u_6\}$ and four attributes $\{a_1, a_2, a_3, a_4\}$, the attribute value of object u_1 on attribute a_4 is $f_4(u_1) = 0.2$, and the set of all attribute values from a_4 is $V_4 = \{0.2, 1.4, 1.5, 1.8\}$.

Based on information table S, we aim to capture the interactive relationships within a numerical attribute (intra-coupled) and among different continuous attributes (inter-coupled). A common method to explore the relationship between continuous attributes is to calculate the Pearson's correlation coefficient (Gan et al. 2007), which measures the agreement of shapes between variables. In detail, the Pearson's product-moment correlation coefficient between attributes a_j and a_k is formalized as

$$Cor(a_j, a_k) = \frac{\sum_{u \in U} (f_j(u) - \mu_j)(f_k(u) - \mu_k)}{\sqrt{\sum_{u \in U} (f_j(u) - \mu_j)^2} \sqrt{\sum_{u \in U} (f_k(u) - \mu_k)^2}}, \qquad (4.3.1)$$

where μ_j, μ_k are the respective mean values of a_j, a_k.

However, the Pearson's correlation coefficient only describes the linear

Table 4.2: The Extended Information Table of Iris Data Set

$\widetilde{A}$	$\langle a_1\rangle^1$	$\langle a_1\rangle^2$	$\langle a_2\rangle^1$	$\langle a_2\rangle^2$	$\langle a_3\rangle^1$	$\langle a_3\rangle^2$	$\langle a_4\rangle^1$	$\langle a_4\rangle^2$
u_1	5.50	30.25	4.20	17.64	1.40	1.96	0.20	0.04
u_2	5.00	25.00	3.40	11.56	1.50	2.25	0.20	0.04
u_3	6.10	37.21	2.90	8.41	4.70	22.09	1.40	1.96
u_4	6.20	38.44	2.20	4.84	4.50	20.25	1.50	2.25
u_5	6.30	39.69	2.70	7.29	4.90	24.01	1.80	3.24
u_6	6.00	36.00	2.20	4.84	5.00	25.00	1.50	2.25

relationship between two variables. It is insufficient if we consider this coefficient just between each pair of continuous attributes. So we expect to expand the numerical space spanned by n continuous attributes with more dimensions, and then expose the coupling relationships of continuous attributes by exploring the correlation between every two updated attributes. The idea of increasing dimensionality is also consistent with (Li & Liu 2012), which extends attribute information but lacks the dependency therein.

Firstly, we lodge some more attributes to the original continuous space. Each attribute a_j is accompanied with $L-1$ more attributes: $\langle a_j\rangle^2, \langle a_j\rangle^3, \cdots,$ $\langle a_j\rangle^L$. The attribute value of $\langle a_j\rangle^p$ $(1 \leq p \leq L)$ is the p-th power of the corresponding value of attribute a_j. That is to say, $\langle a_j\rangle^p.v_t = (a_j.v_t)^p$ for all the attribute values $a_j.v_t \in V_j$. For example, the values of $\langle a_j\rangle^2$ and $\langle a_j\rangle^3$ are the square and cube of the attribute values in V_j, respectively. In this way, data can then be represented as an $m \times L \cdot n$ extended information table, in which the $(L \cdot (j-1) + p)$-th column corresponds to the updated attribute $\langle a_j\rangle^p$. Here, the denotations a_j and $\langle a_j\rangle^1$ are equivalent. For instance, Table 4.2 is an extended information table of the original Table 4.1 if we set $L = 2$ for simplicity.

Next, the correlation between each pair of the updated $L \cdot n$ attributes is calculated. It reflects the global coupling relationships of continuous attributes from both the linear and nonlinear aspects, based on the modeling of variables. Below, we actually use the revised correlation coefficient by taking into account the p-values for testing the hypothesis of no correlation between attributes. Each p-value is the probability of getting a correlation as large as the observed value by random chance, when the true correlation is zero. If p-value is small, say less than 0.05, then the correlation $Cor(a_j, a_k)$ is significant. Thus, the revised correlation coefficient is defined as

$$R\_Cor(a_j, a_k) = \begin{cases} Cor(a_j, a_k) & \text{if p-value} < 0.05, \\ 0 & \text{otherwise.} \end{cases} \tag{4.3.2}$$

In this way, the revised correlation is endowed with the statistical significance, which makes the correlation between variables more reasonable and reliable. That is to say, we only consider those significant coupling relationships of attributes rather than simply involving all of them. The reason is that in the latter case, the over-fitting problem on modeling the coupling relationships may arise, which will inevitably violate the inherent interaction mechanism of attributes. Based on this revised correlation, we propose the intra-coupled interaction and inter-coupled interaction of continuous attributes. Below, L is the maximal power, $1 \leq p, q \leq L$, $a_j = \langle a_j \rangle^1$.

On one hand, the intra-coupled interaction is quantified as the correlations between attribute a_j and its powers $\langle a_j \rangle^p$. Formally, we have

Definition 4.3.1 (Intra-coupled Interaction) *The **Intra-coupled Interaction** within numerical attribute a_j is represented as an $L \times L$ matrix $\mathbf{R^{Ia}}(a_j)$, in which the (p, q) entry describes the correlation between the updated attributes $\langle a_j \rangle^p$ and $\langle a_j \rangle^q$. Specifically,*

$$\mathbf{R^{Ia}}(a_j) = \begin{pmatrix} \theta_{11}(j) & \theta_{12}(j) & \ldots & \theta_{1L}(j) \\ \theta_{21}(j) & \theta_{22}(j) & \ldots & \theta_{2L}(j) \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{L1}(j) & \theta_{L2}(j) & \ldots & \theta_{LL}(j) \end{pmatrix}, \tag{4.3.3}$$

where $\theta_{pq}(j) = R\_Cor(\langle a_j \rangle^p, \langle a_j \rangle^q)$ is the Pearson's correlation coefficient between $\langle a_j \rangle^p$ and $\langle a_j \rangle^q$.

Based on Table 4.1, for attribute a_4, we then have $\mathbf{R^{Ia}}(a_4) = \begin{pmatrix} 1 & 0.989 \\ 0.989 & 1 \end{pmatrix}$ as the intra-coupled interaction within a_4. It means the correlation coefficient between the attribute "Petal.W" and its seconder power is as high as 0.989, which signifies that they are rather closely related.

On the other hand, the inter-coupled interaction captures the correlations between each attribute a_j and all the powers of other attributes a_k $(k \neq j)$. Accordingly, we have

Definition 4.3.2 (Inter-coupled Interaction) *The **Inter-coupled Interaction** between attribute a_j and other attributes a_k $(k \neq j)$ is quantified as an $L \times L \cdot (n-1)$ matrix $\mathbf{R^{Ie}}(a_j|\{a_k\}_{k\neq j})$, in which the $(p, (i-1)\cdot L + q)$ entry represents the correlation of the updated attributes $\langle a_j\rangle^p$ and $\langle a_{k_i}\rangle^q$. Specifically,*

$$\mathbf{R^{Ie}}(a_j|\{a_k\}_{k\neq j}) = \tag{4.3.4}$$
$$\begin{pmatrix} \eta_{11}(j|k_1) & \dots & \eta_{1L}(j|k_1) & \dots & \eta_{11}(j|k_{n-1}) & \dots & \eta_{1L}(j|k_{n-1}) \\ \eta_{21}(j|k_1) & \dots & \eta_{2L}(j|k_1) & \dots & \eta_{21}(j|k_{n-1}) & \dots & \eta_{2L}(j|k_{n-1}) \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \eta_{L1}(j|k_1) & \dots & \eta_{LL}(j|k_1) & \dots & \eta_{L1}(j|k_{n-1}) & \dots & \eta_{LL}(j|k_{n-1}) \end{pmatrix},$$

where $\{a_k\}_{k\neq j} = \{a_{k_1}, \cdots .a_{k_{n-1}}\}$ is the set of attributes other than a_j, and $\eta_{pq}(j|k_i) = R\_Cor(\langle a_j\rangle^p, \langle a_{k_i}\rangle^q)$ is the Pearson's correlation coefficient between $\langle a_j\rangle^p$ and $\langle a_{k_i}\rangle^q$.

For instance, in Table 4.1, we have the inter-coupled interaction of attribute a_4 with others (i.e. a_1, a_2 and a_3) to be

$$\mathbf{R^{Ie}}(a_4|\{a_1, a_2, a_3\}) = \begin{pmatrix} 0.939 & 0.945 & -0.850 & -0.854 & 0.984 & 0.982 \\ 0.925 & 0.933 & 0.000 & -0.813 & 0.951 & 0.952 \end{pmatrix}.$$

Thus, we capture the hidden relationship that "Petal.W" has negative correlation with "Sepal.W", but is positively and closely related with "Sepal.L" and "Petal.L" as well as their second powers, which are consistent with our intuition. In particular, there is no significant correlation between the second power of "Petal.W" and "Sepal.W", indicating the relevant p-value must be at least as large as 0.05. This shows that the involvement of both the intra-coupled interaction and the inter-coupled interaction largely enriches the global coupling than the correlation coefficient which only considers every pair of the original attributes.

4.4 Coupled Representation for Objects

In this section, a coupled representation scheme for numerical objects is proposed by integrating the intra-coupled and inter-coupled interactions of continuous attributes.

In the extended information table $\widetilde{S}$, each quantitative object is described by $L \cdot n$ updated variables $\widetilde{A} = \{\langle a_1\rangle^1, \cdots, \langle a_1\rangle^L, \cdots, \langle a_n\rangle^1, \cdots, \langle a_n\rangle^L\}$. The updated information function $\tilde{f}_j^p(u)$ assigns the corresponding value of attribute $\langle a_j\rangle^p$ to object u. The attribute values of a_j and its powers for u are presented as a vector $\widetilde{\mathbf{u}}(a_j) = [\tilde{f}_j^1(u), \cdots, \tilde{f}_j^L(u)]$, while the attribute values of other attributes and their powers for u are summarized in another vector $\widetilde{\mathbf{u}}(\{a_k\}_{k\neq j}) = [\tilde{f}_{k_1}^1(u), \cdots, \tilde{f}_{k_1}^L(u), \cdots, \tilde{f}_{k_{n-1}}^1(u), \cdots, \tilde{f}_{k_{n-1}}^L(u)]$. For instance, in Table 4.2, $\widetilde{\mathbf{u_1}}(a_4) = [0.20, 0.04]$, we have $\widetilde{\mathbf{u_1}}(\{a_1, a_2, a_3\}) = [5.50, 30.25, 4.20, 17.64, 1.40, 1.96]$.

Further, the coupled interactions are incorporated into a new object representation scheme reflecting the coupling relationships within and between numerical attributes.

Definition 4.4.1 (Coupled Representation) *The* ***Coupled Representation*** *for numerical object u on the continuous attribute a_j is a $1 \times L$ vector $\mathbf{u^c}(a_j|\widetilde{A}, L)$, in which the $(1, p)$ component corresponds to the updated attribute $\langle a_j\rangle^p$. Specifically,*

$$\mathbf{u^c}(a_j|\widetilde{A}, L) = \widetilde{\mathbf{u}}(a_j) \odot \mathbf{w} \otimes [\mathbf{R^{Ia}}(a_j)]^T \tag{4.4.1}$$
$$+\widetilde{\mathbf{u}}(\{a_k\}_{k\neq j}) \odot \underbrace{[\mathbf{w}, \mathbf{w}, \cdots, \mathbf{w}]}_{n-1} \otimes [\mathbf{R^{Ie}}(a_j|\{a_k\}_{k\neq j})]^T,$$

where $\mathbf{w} = [1/(1!), 1/(2!), \cdots, 1/(L!)]$ *is a constant* $1\times L$ *vector,* $[\mathbf{w}, \mathbf{w}, \cdots, \mathbf{w}]$ *is a* $1\times L\cdot(n-1)$ *vector concatenated by* $n-1$ *constant vectors* $\mathbf{w}$*. "*$\odot$*" denotes the Hadamard product*[1]*, and "*$\otimes$*" represents the matrix multiplication.*

For instance, in Table 4.1, we calculate that $\mathbf{u_1^c}(a_4|\widetilde{A}, 2) = [10.92, 14.50]$, where 10.92 and 14.50 are the respective values of $\langle a_4\rangle^1$ and $\langle a_4\rangle^2$. Below, the reason to choose such a coupled representation method is clarified. If the above Equation (4.4.1) is expanded, for instance, we obtain the $(1, p)$ element (corresponds to $\langle a_j\rangle^p$) of the vector $\mathbf{u^c}(a_j|\widetilde{A}, L)$ as

$$\mathbf{u^c}(a_j|\widetilde{A}, L).\langle a_j\rangle^p = \theta_{p1}(j)\cdot \tilde{f}_j^1(u) + \sum_{i=1}^{n-1} \frac{\eta_{p1}(j|k_i)}{1!}\tilde{f}_{k_i}^1(u) \tag{4.4.2}$$

$$+\frac{\theta_{p2}(j)}{2!}\tilde{f}_j^2(u) + \sum_{i=1}^{n-1}\frac{\eta_{p2}(j|k_i)}{2!}\tilde{f}_{k_i}^2(u) + \cdots + \frac{\theta_{pL}(j)}{L!}\tilde{f}_j^L(u) + \sum_{i=1}^{n-1}\frac{\eta_{pL}(j|k_i)}{L!}\tilde{f}_{k_i}^L(u),$$

which resembles the Taylor expansion (Jia & Zhang 2008) of functions. The right side of the above Equation (4.4) is expected to accurately exhibit the intrinsic complete coupled representation $\mathbf{u^c}(a_j|\widetilde{A})$ for object u on the updated attribute $\langle a_j\rangle^p$, when the maximal power L tends to infinity, i.e.

$$\mathbf{u^c}(a_j|\widetilde{A}) = \lim_{L\to+\infty} \mathbf{u^c}(a_j|\widetilde{A}, L). \tag{4.4.3}$$

Further, it is a common practice to approximate a function by using a finite number of terms of its Taylor series. Thus, we intend to approximate the intrinsic complete coupled representation by fixing a positive integer L to largely capture the global interactions of attributes with a tolerable residual error. In the empirical study followed, the maximal power L is evaluated according to the clustering accuracy.

At last, when all the n original attributes are considered, we obtain the global coupled representation for numerical object u to be a concatenated

[1] Hadamard product is a binary operation that takes two row vectors of the same size, and produces another vector where each $(1, i)$ element is the product of the $(1, i)$ elements of the original vectors.

vector:

$$\mathbf{u^c}(\widetilde{A}, L) = [\mathbf{u^c}(a_1|\widetilde{A}, L), \mathbf{u^c}(a_2|\widetilde{A}, L), \cdots, \mathbf{u^c}(a_n|\widetilde{A}, L)]. \tag{4.4.4}$$

Therefore, each object is now represented as a $1 \times L \cdot n$ numerical vector incorporated with the couplings of continuous attributes. We then obtain an $m \times L \cdot n$ coupled information table S^c when all the objects in universe U follow the above steps. For instance, based on Table 4.1, the coupled information table shown in Table 4.3, is the new representation.

Table 4.3: The Coupled Representation of Iris Data Set

$\widetilde{A}$	$\langle a_1\rangle^1$	$\langle a_1\rangle^2$	$\langle a_2\rangle^1$	$\langle a_2\rangle^2$	$\langle a_3\rangle^1$	$\langle a_3\rangle^2$	$\langle a_4\rangle^1$	$\langle a_4\rangle^2$
u_1	22.99	23.00	10.74	10.74	10.05	10.04	10.92	14.50
u_2	20.09	20.10	6.70	6.69	10.80	10.77	11.48	14.30
u_3	41.20	41.27	−7.76	−8.58	34.40	34.32	35.10	36.92
u_4	41.13	41.20	−9.35	−10.30	36.35	36.24	37.03	38.22
u_5	44.66	44.74	−9.87	−11.21	38.55	38.45	39.28	40.86
u_6	42.32	42.39	−11.84	−12.79	37.92	37.82	38.52	39.63

So far, we have obtained the global coupled representation S^c for continuous data. The coupled representation for numerical objects reflects the mutual influence and interactions of attributes, and reserves far more coupling relationships from continuous data than the original representation. Back to the case discussed in Section 4.1, we obtain that the normalized Euclidean distance between u_4 and u_6 is 0.448 based on S^c, larger than both the normalized distances between u_4, u_3 (i.e. 0.354) and between u_6, u_5 (i.e. 0.419). Similarly, the normalized distance between u_3, u_5 (i.e. 0.830) is also greater than them. It means that u_4, u_6 and u_3, u_5 are unlikely to be clustered together, which is consistent with the real situation and verifies that our proposed coupled representation is effective in capturing the implicit relationships.

4.5 Empirical Study

In this section, several experiments are performed on 13 UCI data sets (i.e. Table 4.4) to show the effectiveness of our proposed coupled representation scheme for numerical objects. Two data representation schemes are considered and compared: the original representation as an information table S and the coupled representation as a coupled information table S^c. Each column of S and S^c is normalized to have zero mean and standard deviation as one, so as to eliminate value differences in the order of magnitudes.

The experiments are divided into two categories: parameter estimation and learning applications. Note that the number of runs is set to be 100 to obtain the corresponding average results with their sample standard deviations. The number of clusters is fixed to be the number of real classes , i.e. the fourth column in Table 4.4.

Table 4.4: Description of Data Sets in Chapter 4

Data Set	Object	Attribute	Class	Short Form
Iris	150	4	3	Ir
Planning	182	12	2	Pl
Parkinsons	195	22	2	Par
Seeds	210	7	3	See
Segment	210	19	7	Seg
Ionos	351	34	2	Io
Patient	583	9	2	Pat
Blood	748	5	2	Bl
Vowel	990	10	11	Vo
Red Wine	1599	11	6	Rw
Waveform	5000	21	3	Wa
Navigation	5456	24	4	Na
Telescope	19020	10	2	Te

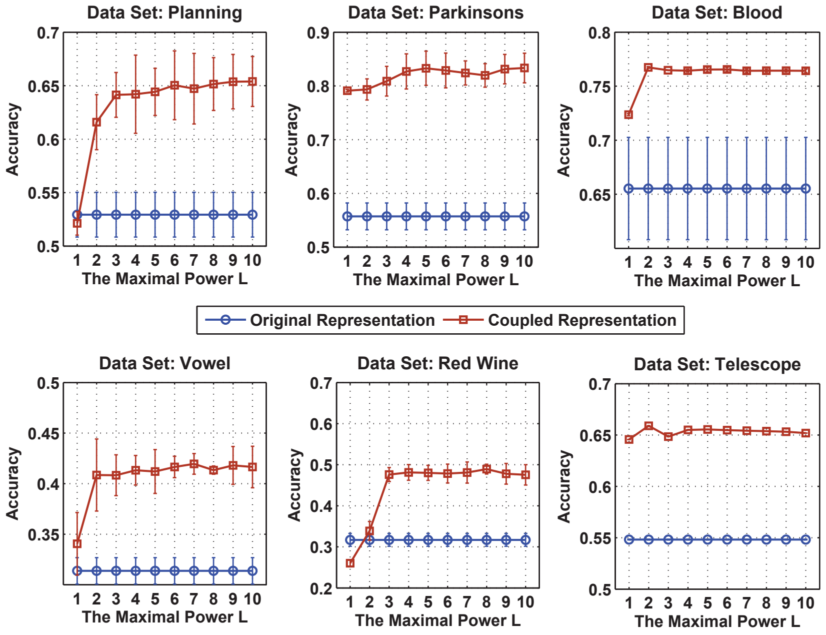


Figure 4.2: The performance of L on six data sets: the average accuracy with $\pm$ sample standard deviation error bars.

4.5.1 Parameter Estimation

As indicated in Equation (4.4), the proposed coupled representation for numerical objects is strongly dependent on the maximal power L. Here, we conduct several experiments to study the performance of L with regard to the clustering accuracy of *k-means*. The maximal power L is set to range from $L = 1$ to $L = 10$ since $L!$ becomes extremely large when L grows, which means $L = 10$ is probably large enough to obtain most of the information in Equation (4.4).

Figure 4.2 shows the performance of L on six data sets in terms of the clustering accuracy of *k-means* based on different representations. It is clear that the clustering accuracy of coupled representation generally reaches to a stable point when L takes the value 3 or 4, which means that $L = 3$ or $L = 4$ is empirically large enough to capture the global couplings of attributes. As a general trend, the accuracy goes up when L increases. Only except Blood and Telescope, the correlation coefficients between the accuracy and L are significantly around 0.75 for the rest data. But the increasing rate of accuracy gets smaller as L grows. This is consistent with Equation (4.4), since a large value of $L!$ acting as the denominator makes the corresponding item rather small. In the experiments followed, we fix L to be 3 or 4 and report the better results between them.

Another important observation is *k-means* based on the coupled representation always outperforms that built on the original representation when $L \geq 2$, though a small deviation exists. That is to say, our proposed representation is useful and effective to discover the coupling relationships embedded in the continuous attributes. In addition, the null hypothesis that *k-means* with the coupled representation is better than the original *k-means* in terms of the accuracy is accepted. However, we can see that our coupled method does not perform stably well when $L = 1$. The reason is the case of $L = 1$ just reflects the linear relationship among attributes and only captures a local picture of the global interactions.

4.5.2 Learning Applications

In this part, three groups of experiments are conducted on extensive data sets for machine learning applications.

Cluster Structure Analysis

Experiments are performed to explicitly specify the internal structures for the labeled numerical data. Clusterings are normally evaluated by assigning the best score to the algorithm that produces clusters with the highest similarity within a cluster and the lowest similarity between clusters based on a certain data representation scheme. We work in a different way, in which data representation methods are evaluated with the given labels and the clustering internal descriptors: Relative Distance (RD), Davies-Bouldin Index (DBI) (Davies & Bouldin 1979), Dunn Index (DI) (Dunn 1974), and Sum-Distance

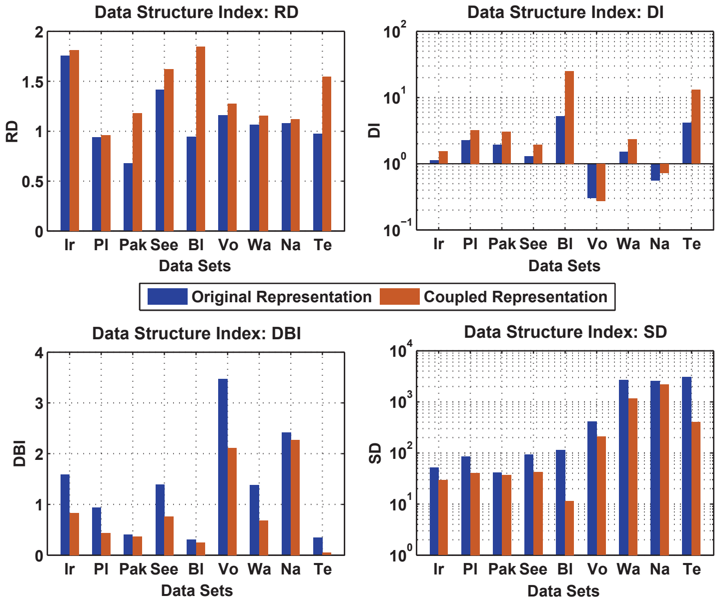


Figure 4.3: Data structure index comparisons on nine data sets.

(SD). In detail, RD is the ratio of average inter-cluster distance upon average intra-cluster distance for all the cluster labels; SD is the sum of object distances within all the clusters. Since the internal criteria seek the clusters with a high intra-cluster similarity and a low inter-cluster similarity, larger RD, larger DI, smaller DBI, and smaller SD indicate a better characterization of the cluster differentiation capability, which corresponds to a superior data representation scheme.

The cluster structures produced by the original and coupled data representation schemes are then analyzed on nine data sets in different scales. The results after normalization are shown in Figure 4.3, which shows that, with the exception of only one item (i.e. Vowel on DI), the corresponding RD and DI indexes for the coupled representation are larger than those for the original representation; while the associated DBI and SD indexes for the former are always smaller than those for the latter. It shows that our proposed coupled representation, which effectively captures the global interactions of attributes, is superior to the original method in terms of differentiating objects in distinct clusters.

Table 4.5: Clustering Comparisons on Six Data Sets with ± Sample Standard Deviation

Data Set		Iris	Parkinsons	Seeds	Segment	Vowel	Navigation	Avg
Accuracy	*LC-OR*	0.660 ± 0.00	0.744 ± 0.00	0.348 ± 0.00	0.157 ± 0.00	0.111 ± 0.00	0.404 ± 0.00	0.404
	LC-CR	**0.690 ± 0.00**	**0.799 ± 0.00**	**0.382 ± 0.00**	**0.229 ± 0.00**	**0.224 ± 0.00**	**0.495 ± 0.00**	**0.470**
	SC-OR	0.783 ± 0.08	0.728 ± 0.05	0.852 ± 0.13	0.496 ± 0.06	0.337 ± 0.01	0.400 ± 0.00	0.599
	SC-CR	**0.905 ± 0.03**	**0.770 ± 0.06**	**0.891 ± 0.00**	**0.566 ± 0.06**	**0.423 ± 0.06**	**0.485 ± 0.01**	**0.673**
NMI	*LC-OR*	0.579 ± 0.00	0.005 ± 0.00	0.011 ± 0.00	0.034 ± 0.00	0.064 ± 0.00	0.000 ± 0.00	0.116
	LC-CR	**0.628 ± 0.00**	**0.050 ± 0.00**	**0.050 ± 0.00**	**0.229 ± 0.00**	**0.365 ± 0.00**	**0.034 ± 0.00**	**0.226**
	SC-OR	0.606 ± 0.05	0.016 ± 0.03	0.657 ± 0.17	0.464 ± 0.07	***0.397 ± 0.01***	0.010 ± 0.00	0.358
	SC-CR	**0.752 ± 0.02**	**0.056 ± 0.11**	**0.703 ± 0.00**	**0.497 ± 0.05**	0.394 ± 0.11	**0.037 ± 0.00**	**0.407**
Precision	*LC-OR*	0.500 ± 0.00	0.566 ± 0.00	***0.669 ± 0.00***	0.307 ± 0.00	0.254 ± 0.00	0.548 ± 0.00	0.474
	LC-CR	**0.837 ± 0.00**	**0.670 ± 0.00**	0.650 ± 0.00	**0.453 ± 0.00**	**0.391 ± 0.00**	**0.608 ± 0.00**	**0.602**
	SC-OR	0.783 ± 0.09	0.590 ± 0.06	0.860 ± 0.12	0.496 ± 0.08	0.357 ± 0.01	0.365 ± 0.00	0.575
	SC-CR	**0.905 ± 0.04**	**0.730 ± 0.16**	**0.907 ± 0.00**	**0.641 ± 0.10**	**0.418 ± 0.16**	**0.413 ± 0.02**	**0.669**
Specificity	*LC-OR*	0.830 ± 0.00	0.243 ± 0.00	0.674 ± 0.00	0.860 ± 0.00	0.904 ± 0.00	**0.596 ± 0.00**	0.685
	LC-CR	**0.840 ± 0.00**	**0.273 ± 0.00**	**0.681 ± 0.00**	**0.871 ± 0.00**	**0.905 ± 0.00**	**0.596 ± 0.00**	**0.694**
	SC-OR	0.892 ± 0.04	0.284 ± 0.11	0.926 ± 0.06	0.916 ± 0.01	0.924 ± 0.00	0.618 ± 0.00	0.760
	SC-CR	**0.952 ± 0.02**	**0.325 ± 0.03**	**0.965 ± 0.00**	**0.928 ± 0.01**	**0.936 ± 0.03**	**0.635 ± 0.01**	**0.790**

Data Clustering Evaluation

Two classical clustering approaches are single linkage based agglomerative algorithm (*LC*) (Ackerman & Ben-David 2011) and spectral clustering (*SC*) (Luxburg 2007). Here, these two methods are evaluated when incorporating the original (i.e. *LC-OR* and *SC-OR*) and coupled (i.e. *LC-CR* and *SC-CR*) data representation schemes individually. The external clustering quality measures here include Accuracy, Normalized Mutual Information (NMI), Precision, and Specificity. As described in (Cai, He & Han 2005, Figueiredo et al. 2011), the larger these indexes, the better the clustering.

Table 4.5 reports the results for the four approaches on six data sets in terms of the above four clustering quality measures. The higher measure scores of each experimental setting are highlighted in boldface, when *LC-CR* is compared with *LC-OR* and *SC-CR* is compared with *SC-OR*. This table indicates that the adapted *LC-CR* and *SC-CR* respectively outperform their baseline algorithms *LC-OR* and *SC-OR* on almost all the evaluation measures for all the data sets, only except the two italic bold values. The maximal average improvement rate across all the data sets is 95.67%, while the minimal is 1.44%. Statistical testing also supports the results that *LC-CR* performs better than *LC-OR* and *SC-CR* performs better than *SC-OR*, at a 95% significance level. Another interesting observation is that *SC* is mostly superior to *LC*, which is also consistent with such a statement in (Luxburg 2007).

Data Classification Evaluation

To further verify the superiority of our proposed coupled method, we use the k-nearest neighbor (*KNN*) algorithm (Figueiredo et al. 2011) to compare the classification quality when using different representation schemes. *KNN* is a type of instance-based learning, classifying objects based on the closest training examples in the attribute space. We carry out experiments on six data sets from different domains. As we know, a better data representation approach corresponds to a better classification result, i.e. higher

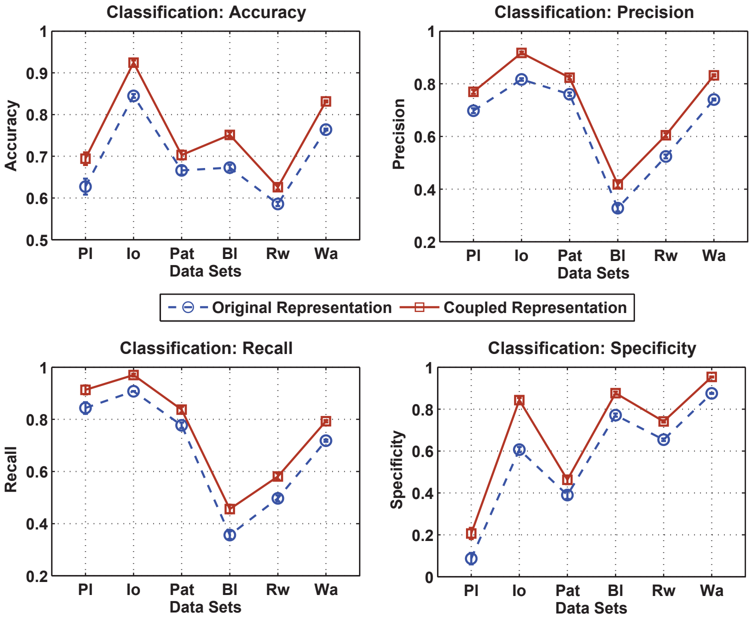


Figure 4.4: Data classification comparisons on six data sets: the average values with ± sample standard deviation error bars.

Accuracy, higher Precision, higher Recall, and higher Specificity (Figueiredo et al. 2011). We use the 10-fold cross-validation with $K = 4$.

The results of *KNN* based on he original and coupled representations are shown in Figure 4.4. *KNN* upon the coupled representation remarkably outperforms the original *KNN* for all the data sets in terms of all the evaluation measures. As can be seen, there is a remarkable improvement in the results with our proposed coupled scheme compared to the original method with respect to the classification quality. The maximal relative improvement rate across all the data sets is 138.89%, while the minimal rate is 5.51%. All the results are supported by a statistical significant test at 95% significance level. Similar results can also be observed by *KNN* when K takes other integers, which again suggests the effectiveness and superiority of our proposed coupled method.

It is also noted that the improvement on Patient is relatively small with respect to all the measures. The reason is the coupled interactions among the attributes of Patient is weak. Only around 45% pairs of attributes and their powers have significant coupling relationships, compared to the average percentage of around 78% on other data sets.

4.6 Summary

We have proposed a novel coupled representation scheme for objects via teasing out the interactions of numerical attributes (i.e. non-IIDness of continuous properties). Those interactions are quantified by analyzing the Pearson's correlation coefficients between attributes and their powers, in which the intra-coupled interaction within an attribute is described by the correlation between each attribute and its own powers, and the inter-coupled interaction among different attributes is characterized by the correlation between every attribute and the powers of others. Both interactions are integrated with the traditional information table in a global way to form the Taylor expansion of a coupled representation scheme for quantitative objects. The selection of the maximal power is empirically studied in terms of the clustering accuracy, reporting that $L = 3$ or $L = 4$ is large enough to capture the coupling relationships of numerical attributes. Substantial experiments have verified that our proposed coupled representation scheme outperforms the original representation method from the perspectives of data structure, data clustering, and data classification. Statistical analysis supports our conclusion.

[***Note***] *A conference version of this chapter has been accepted already and will be published soon by AAAI Press as below.*

- ***Can Wang****, Zhong She, Longbing Cao (2013), "Coupled Attribute Analysis on Numerical Data". The 23rd International Joint Conference on Artificial Intelligence (****IJCAI 2013****), full paper accepted.*

Chapter 5

Categorical Coupled Behavior Analysis

In this chapter, we investigate and analyze the quantitative coupled behaviors via teasing out the coupling relationship between categorical properties. Here, categorical behavior refers to the behavior data described by nominal properties or attributes, and categorical coupled behaviors further explore the couplings/interactions among a group of categorical behaviors. Throughout the chapter, data object and the entity of coupled behaviors are interchangeable, categorical attributes indicate the discrete properties of coupled behaviors, and accordingly attribute values denote the property values of coupled behaviors.

Limited research has been conducted on the similarity analysis for categorical behaviors, which mainly assumes the independence of nominal properties, especially in unsupervised learning. Recent work on the attribute value frequency distribution and attribute dependency aggregation introduces the frequency and co-occurrence of attribute values to explore the categorical property coupling, but they are considered separately to exhibit only a local picture in analyzing the similarity of categorical behaviors. Such a local picture is not effective for deep analysis, and the integration of frequency and co-occurrence in defining property similarity is not a trivial task.

An efficient data-driven similarity learning approach for categorical coupled behaviors is presented. It generates a coupled property similarity measure for nominal entities with property couplings to capture a global picture of attribute similarity. It involves the frequency-based intra-coupled similarity within a property and the inter-coupled similarity upon value co-occurrences between properties, as well as their integration on the entity level. In particular, four measures are designed for the inter-coupled similarity to calculate the similarity between two categorical values by considering their relationships with other attributes in terms of power set, universal set, join set and intersection set. The theoretical analysis reveals the equivalent accuracy and superior efficiency of the measure based on the intersection set, particularly for large-scale data sets. Substantial experiments on 20 UCI data sets verify the theoretical conclusions. In addition, intensive experiments of data structure, clustering and classification algorithms incorporating the coupled dissimilarity metric achieve a significant performance improvement on state-of-the-art measures and algorithms on 12 UCI data sets and bibliographic data, which is confirmed by the statistical analysis. The experiment results show that the proposed coupled property similarity for categorical behaviors is generic, and can effectively and efficiently capture the intrinsic and global interactions within and between properties for especially large-scale categorical behavior data sets.

5.1 Background and Overview

Similarity analysis has been a problem of great practical importance in several domains for decades, not least in recent work, including behavior analysis (Cao et al. 2012), document analysis (Figueiredo et al. 2011) and image analysis (Wang, Hoiem & Forsyth 2012). A typical aspect of these applications is clustering, in which the similarity is usually defined in terms of one of the following levels: between clusters, between attributes, between data objects, or between attribute values. The similarity between clusters is often built on top of the similarity between data objects, e.g. centroid similarity. Further, the similarity between data objects is in general derived from the similarity between attribute values, e.g. Euclidean distance and simple matching similarity (Kaufman & Rousseeuw 1990). The similarity measured between attribute values assesses the relationship between two data objects and even between two clusters: the more two objects or clusters resemble each other, the larger is the similarity (Gan et al. 2007). The other similarity between attributes (Das & Mannila 2000) can also be converted into the difference of similarities between pairwise attribute values (Li et al. 2010). Therefore, the similarity between attribute values plays a fundamental role in similarity analysis.

The similarity measures for attribute values are sensitive to the attribute types, which are classified as discrete and continuous. The discrete attribute is further typed as nominal (categorical) or binary (Gan et al. 2007). The nominal data, a special case of the discrete type, has only a finite number of values; while the binary variable has exactly two values. In this chapter, we regard the binary data as a special case of the nominal data.

Compared to the intensive study on the similarity between two numerical variables, such as Euclidean and Minkowski distance, and between two categorical values in supervised learning, e.g. Heterogeneous Distance Functions (Wilson & Martinez 1997) and Modified Value Distance Matrix (*MVDM*) (Cost & Salzberg 1993), the similarity for nominal variables has received much less attention in unsupervised learning on unlabeled data. Only limited efforts (Gan et al. 2007) have been made, including Simple Matching Similarity (*SMS*, which uses 0s and 1s to distinguish the similarity between distinct and identical categorical values), Occurrence Frequency (*OF*) (Boriah et al. 2008) and Information-theoretical Similarity (*Lin*) (Boriah et al. 2008, Lin 1998), to discuss the similarity between nominal values. The challenge is that these methods all follow a classic but strong assumption of independent and identical distribution (i.e. IIDness). They are too rough to precisely characterize the similarity between categorical attribute values,

Table 5.1: An Instance of the Movie Database

Movie	**Director**	**Actor**	**Genre**	**Class**
Godfather II	Scorsese	De Niro	Crime	l_1
Good Fellas	Coppola	De Niro	Crime	l_1
Vertigo	Hitchcock	Stewart	Thriller	l_2
N by NW	Hitchcock	Grant	Thriller	l_2
Bishop's Wife	Koster	Grant	Comedy	l_2
Harvey	Koster	Stewart	Comedy	l_2

only deliver a local picture of the similarity, and are not data-driven. In addition, none of them provides a comprehensive picture of similarity between categorical attributes by combining relevant aspects. Below, we illustrate the problem with *SMS* and the challenge of analyzing the categorical data similarity.

As shown in Table 5.1, six movie objects are divided into two classes with three nominal attributes: director, actor and genre. The *SMS* measure between directors "*Scorsese*" and "*Coppola*" is 0, but "*Scorsese*" and "*Coppola*" are very similar[1]. Another observation by following *SMS* is that the similarity between "*Koster*" and "*Hitchcock*" is equal to that between "*Koster*" and "*Coppola*"; however, the similarity of the former pair should be greater because both directors belong to the same class l_2.

The above examples show that it is much more complex to analyze the similarity between nominal variables than between continuous data. *SMS* and its variants fail to capture a global picture of the genuine relationship for nominal data. With the exponential increase of categorical data such as that derived from social networks, it is important to develop effective and efficient measures for capturing the similarity between nominal variables.

The similarity between categorical values is sensitive to the data characteristics. In general, two attribute values are expected to be similar if they present analogous frequency distributions within one attribute (e.g. *OF* and

[1] A conclusion drawn from a well-informed cinematic source.

Lin) (Boriah et al. 2008, Lin 1998); this reflects the *intra-coupled similarity* within attributes. For example, two directors are very similar if they appear with almost the same frequency, such as "*Scorsese*" with "*Coppola*" and "*Koster*" with "*Hitchcock*". However, the reality is that the former director pair is more similar than the latter. Ahmad and Dey (Ahmad & Dey 2007) introduced the co-occurrence probability of categorical values from different attributes and compared this probability for two categorical values from the same attribute. This means that the similarity between directors relates to the dependency of "director" on other attributes such as "actor" and "genre" over all the movie objects: namely, the *inter-coupled similarity* between attributes. They both capture local pictures of the similarity from different perspectives. A detailed review of the related work on categorical behavior analysis can be found in Section 2.3. No work has been reported on systematically considering both intra-coupled similarity and inter-coupled similarity. The incomplete description of the categorical value similarity leads to tentative and less effective learning performance. In addition, it is usually very costly to consider the similarity between values in relation to the dependency between attributes and the aggregation of such dependency (Ahmad & Dey 2007), which is verified in Section 5.5.

In this chapter, we explicitly discuss the data-driven intra-coupled similarity and inter-coupled similarity, as well as their global aggregation in unsupervised learning on nominal data, reflecting our non-IIDness design in this research task. We propose a Coupled Attribute Similarity for Objects ($CASO$) measure based on a Coupled Attribute Similarity for Values ($CASV$) measure, by considering both the Intra-coupled and Inter-coupled Attribute Value Similarities ($IaASV$ and $IeASV$), which globally capture the attribute value frequency distribution and the attribute dependency aggregation respectively with high accuracy and relatively low complexity. Then, we compare the accuracy and efficiency of the four proposed measures for $IeASV$ in terms of four relationships: power set, universal set, join set and intersection set, and obtain the most efficient candidate based on the intersection set (i.e. *IRSI*) from theoretical and experimental aspects. In addition, the proposed measures are compared with the state-of-the-art metrics on various benchmark categorical data sets in terms of the internal and external clustering criteria, as well as the classification accuracy. All the results are statistically significant. During the whole process, a method is proposed to flexibly define the dissimilarity metrics with the proposed similarity building blocks according to specific requirements. Finally, it is also remarkable to note that two new coupled categorical clustering algorithms, which are *CROCK* and *CLIMBO*, are accordingly proposed and verified.

The chapter is organized as follows. Preliminary definitions are specified in Section 5.2. Section 5.3 proposes the framework of the coupled attribute similarity analysis. Section 5.4 defines the intra-coupled similarity, inter-coupled similarity, and their aggregation. The theoretical analysis is given in Section 5.5. We describe the *CASO* algorithm in Section 5.6. The efficiency and effectiveness of *CASO* are empirically studied in Section 5.7, two new categorical clustering methods (*CROCK* and *CLIMBO*) are introduced, and a flexible method to define dissimilarity metrics is also developed. Finally, we conclude this work and address future work in Section 5.8.

5.2 Preliminary Definitions

A large number of data objects with the same attribute set can be organized by an information table $S =< U, A, V, f >$, where universe $U = \{u_1, \cdots, u_m\}$ is composed of a nonempty finite set of data objects; $A = \{a_1, \cdots, a_n\}$ is a finite set of attributes; $V = \bigcup_{j=1}^{n} V_j$ is a collection of attribute value sets, in which V_j is the set of attribute values from attribute $a_j (1 \leq j \leq n)$; and $f = \bigcup_{j=1}^{n} f_j$, $f_j : U \to V_j (1 \leq j \leq n)$ is an information function which assigns a particular value of attribute a_j to every object. For instance, Table 5.2 is an information table consisting of six objects $\{u_1, \cdots, u_6\}$ and three attributes $\{a_1, a_2, a_3\}$, the attribute value of object u_1 for attribute a_2 is $f_2(u_1) = \mathcal{B}_1$, and the set of all attribute values for a_2 is $V_2 = \{\mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3\}$.

Table 5.2: An Example of Information Table

A U	a_1	a_2	a_3
u_1	$\mathcal{A}_1$	$\mathcal{B}_1$	$\mathcal{C}_1$
u_2	$\mathcal{A}_2$	$\mathcal{B}_1$	$\mathcal{C}_1$
u_3	$\mathcal{A}_2$	$\mathcal{B}_2$	$\mathcal{C}_2$
u_4	$\mathcal{A}_3$	$\mathcal{B}_3$	$\mathcal{C}_2$
u_5	$\mathcal{A}_4$	$\mathcal{B}_3$	$\mathcal{C}_3$
u_6	$\mathcal{A}_4$	$\mathcal{B}_2$	$\mathcal{C}_3$

Generally speaking, the similarity between two objects $u_x, u_y (\in U)$ can be built on top of the similarities between their attribute values $v_j^x, v_j^y (\in V_j)$ for all attributes $a_j \in A$. Here, v_j^x and v_j^y indicate the respective attribute values of objects u_x and u_y for the attribute a_j, for example, $v_2^1 = \mathcal{B}_1$ and $v_1^2 = \mathcal{A}_2$. By proposing a coupled attribute value similarity measure, we define a new object similarity for categorical data. The basic concepts below facilitate the formulation for a coupled attribute value similarity measure. They are exemplified by Table 5.2. Below, an information table S is given, and |set| is the number of elements in a certain set.

Definition 5.2.1 (SIF) *Two* ***Set Information Functions (SIFs)*** *are defined as:*

$$F_j : 2^U \to 2^{V_j}, \quad F_j(U') = \{f_j(u_x) | u_x \in U'\}, \tag{5.2.1}$$

$$G_j : 2^{V_j} \to 2^U, \quad G_j(V_j') = \{u_i | f_j(u_i) \in V_j'\}, \tag{5.2.2}$$

where $1 \leq j \leq n$, $1 \leq i \leq m$, $U' \subseteq U$ *and* $V_j' \subseteq V_j$.

These *SIF*s describe the relationships between objects and attribute values from different levels. Function $F_j(U')$ assigns the associated value set of attribute a_j to the object set U'. Function $G_j(V_j')$ maps the value set V_j' of attribute a_j to the dependent object set. For example, based on the attribute a_2, $F_2(\{u_1, u_2, u_3\}) = \{\mathcal{B}_1, \mathcal{B}_2\}$ collects the attribute values of u_1, u_2 and u_3;

and $G_2(\{\mathcal{B}_1, \mathcal{B}_2\}) = \{u_1, u_2, u_3, u_6\}$ returns the objects whose attribute values are $\mathcal{B}_1$ and $\mathcal{B}_2$.

Note that in the two definitions below, the superscripts x and y of v_j are omitted, since any attribute value $v_j \in V_j$ used here is independent of the objects u_x and u_y. However, v_j^x and v_j^y are reused when defining the similarities in the following sections.

Definition 5.2.2 (IIF) *The **Inter-information Function (IIF)** obtains a value subset of attribute a_k for the corresponding objects, which are derived from the value v_j of attribute a_j. It is defined as:*

$$\varphi_{j\to k} : V_j \to 2^{V_k}, \quad \varphi_{j\to k}(v_j) = F_k(G_j(\{v_j\})). \tag{5.2.3}$$

This *IIF* $\varphi_{j\to k}$ is the composition of F_k and G_j. The involved subscript $j \to k$ means that this mapping φ is performed from attribute a_j to attribute a_k. Intuitively, $\varphi_{j\to k}(v_j)$ computes the set of attribute values from attribute a_k that co-occurs with a particular attribute value v_j from attribute a_j. For example, $\varphi_{2\to 1}(\mathcal{B}_1) = \{\mathcal{A}_1, \mathcal{A}_2\}$ specifies that the attribute values $\mathcal{B}_1$ of attribute a_2 and $\{\mathcal{A}_1, \mathcal{A}_2\}$ of attribute a_1 are related by the corresponding objects: u_1 and u_2.

Definition 5.2.3 (ICP) *The value subset $V_k'(\subseteq V_k)$ of attribute a_k, and the value $v_j(\in V_j)$ of attribute a_j, then the **Information Conditional Probability (ICP)** of V_k' with respect to v_j is $P_{k|j}(V_k'|v_j)$, defined as:*

$$P_{k|j}(V_k'|v_j) = \frac{|G_k(V_k') \bigcap G_j(\{v_j\})|}{|G_j(\{v_j\})|}. \tag{5.2.4}$$

Intuitively, when given all the objects with the value v_j of attribute a_j, *ICP* is the percentage of common objects whose values of attribute a_k fall in subset V_k' and whose values of attribute a_j are exactly v_j as well. For example, $P_{1|2}(\{\mathcal{A}_1\}|\mathcal{B}_1) = 0.5$.

All these concepts and functions form the foundation for formalizing the coupled interactions within and between categorical attributes, as presented below. The main notations in this chapter are listed in Table 5.3.

Table 5.3: List of Main Notations in Chapter 5

Variable	Explanation
$\{u_1, \cdots, u_m\}$	The set of m objects U
$\{a_1, \cdots, a_n\}$	The set of n attributes A
$l(\in L)$	Any label in the label (class) set L
$V_j'(\subseteq V_j)$	The subset of value set V_j of attribute a_j
$R(=\max \lvert V_j \rvert)$	The maximal number of values of each attribute
$v_j^x, v_j^y(\in V_j)$	Specific values of attribute a_j for objects u_x, u_y
$v_k(\in V_k)$	Any value of attribute a_k

5.3 Framework of the Coupled Attribute Similarity Analysis

In this section, a framework for coupled attribute similarity analysis is proposed from a global perspective of the intra-coupled interaction within an attribute, the inter-coupled interaction among multiple attributes, and the integration of both.

With respect to the intra-coupled interaction, the similarity between attribute values is considered by examining the occurrence frequencies of them within one attribute. For the inter-coupled interaction, the similarity between attribute values is captured by exposing the co-occurrence dependency of them on the values of other attributes. For example, the coupled value similarity between B_1 and B_2 (i.e. values of attribute a_2) concerns both the intra-coupling relationship specified by the repeated times of values B_1 and B_2: 2 and 2, and the inter-coupled interaction triggered by the other two attributes (a_1 and a_3). Next, the coupled interaction is derived by the integration of intra-coupling and inter-coupling. In this way, the couplings of attributes lead to more accurate similarity ($\in [0, 1]$) between attribute values, rather than a rude assignment of either 0 or 1.

In the framework described in Figure 5.1, the couplings of attributes

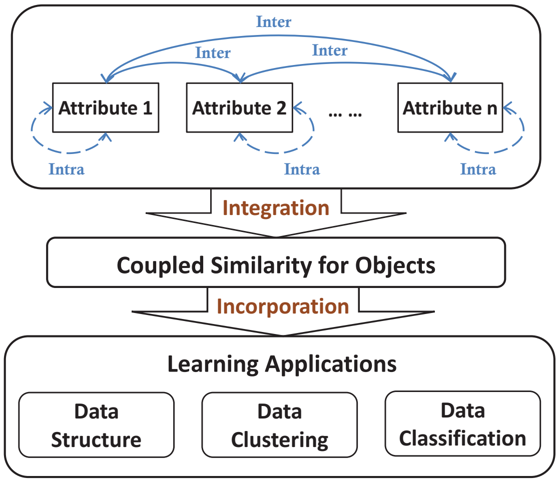


Figure 5.1: A framework of coupled attribute similarity analysis, where $\dashleftarrow\dashrightarrow$ indicates intra-coupling and $\longleftrightarrow$ refers to inter-coupling.

are revealed via the similarity between attribute values v_j^x and v_j^y of each attribute a_j by means of the intra-coupling and inter-coupling. Further, the coupled similarity for objects is built on top of the pairwise similarity between attribute values according to the integration of couplings. Finally, three learning tasks are explored for the data structure, data clustering, and data classification by incorporating the coupled interactions, revealing that the couplings of attributes are essential to learning applications in empirical studies.

Given an information table S with a set of m objects U and a set of n attributes A, we specify those interactions and couplings individually in the following sections.

5.4 Coupled Attribute Similarity

The attribute couplings are proposed in terms of both intra-coupled and inter-coupled similarities. Below, the intra-coupled and inter-coupling relationships, as well as the integrated coupling, are formalized and exemplified.

5.4.1 Intra-coupled Interaction

According to (Gan et al. 2007), the discrepancy in attribute value occurrence times reflects the value similarity in terms of frequency distribution. It reveals that greater similarity is assigned to the attribute value pair which owns approximately equal frequencies. The higher these frequencies are, the closer the two values are. Different occurrence frequencies therefore indicate distinct levels of attribute value significance.

These principles are also consistent with the similarity theorem presented in (Lin 1998), in which the commonality corresponds to the product of frequencies and the full description relates to the total sum of individual frequencies and their product. In addition, a comparative evaluation on similarity measures for categorical data has been done in (Boriah et al. 2008), delivering *OF* and *Lin* as the two best similarity measures among 14 existing measures on 18 data sets. Both these measures assign higher weights to mismatches or matches on frequent values, and the maximum similarity is attained when the attribute values exhibit approximately equal frequencies (Boriah et al. 2008).

Thus, when calculating attribute value similarity, we consider the relationship between the attribute value frequencies of an attribute, proposed as intra-coupled similarity to satisfy the above principles.

Definition 5.4.1 (IaASV) *The **Intra-coupled Attribute Similarity for Values (IaASV)** between values v_j^x and v_j^y of attribute a_j is:*

$$\delta_j^{Ia}(v_j^x, v_j^y) = \frac{|G_j(\{v_j^x\})| \cdot |G_j(\{v_j^y\})|}{|G_j(\{v_j^x\})| + |G_j(\{v_j^y\})| + |G_j(\{v_j^x\})| \cdot |G_j(\{v_j^y\})|}. \qquad (5.4.1)$$

Since $1 \leq |G_j(v_j^x)|, |G_j(v_j^y)| \leq m$ and $2 \leq |G_j(v_j^x)| + |G_j(v_j^y)| \leq m$, then $\delta_j^{Ia} \in [1/3, m/(m+4)]$ is obtained according to Proof (a) in the Appendix of Section 5.9. For example, in Table 5.2, both $\mathcal{B}_1$ and $\mathcal{B}_2$ are observed twice, $\delta_2^{Ia}(\mathcal{B}_1, \mathcal{B}_2) = 0.5$.

Note that there is still an issue in the above definition: if two attribute values v_j^x and v_j^y have the same frequency, then we have $\delta_j^{Ia}(v_j^x, v_j^x) = \delta_j^{Ia}(v_j^x, v_j^y)$. This is somewhat intuitively problematic, but the inter-coupled similarity proposed in the next section remedies this issue because the inter-coupled similarities between v_j^x, v_j^x and between v_j^x, v_j^y are overwhelmingly distinct.

By taking into account the frequencies of categories, $IaASV$ characterizes the value similarity in terms of attribute value occurrence times.

5.4.2 Inter-coupled Interaction

$IaASV$ considers the interaction between attribute values within an attribute a_j. It does not involve the couplings between attributes (e.g. between attributes $a_k(k \neq j)$ and a_j) when calculating attribute value similarity. For this, we discuss the dependency aggregation, i.e. inter-coupled interaction.

In 1993, Cost and Salzberg (Cost & Salzberg 1993) presented a powerful new method $MVDM$ for measuring the dissimilarity between categorical values. $MVDM$ takes into account the overall similarity of classification of all objects on each possible value of each attribute. The dissimilarity $D_{j|L}$ between two attribute values v_j^x and v_j^y for a specific attribute a_j regarding labels L is:

$$D_{j|L}(v_j^x, v_j^y) = \sum_{l \in L} |P_{l|j}(\{l\}|v_j^x) - P_{l|j}(\{l\}|v_j^y)|, \qquad (5.4.2)$$

where $l(\in L)$ is a label in the information table S. $P_{l|j}$ is the ICP defined in (6.2.1) by replacing the attribute a_k with the label l, the attribute value subset V_k' with the label subset $L' \subseteq L$ (here $L' = \{l\}$), in which $g_l^*(L')$ refers to the set of objects whose labels fall in L'. $D_{j|L}$ indicates that values are identified as being similar if they occur with the same relative frequency for all classes. According to the principle (Gibbs & Su 2002) that, for the categorical data distribution, the sum of L1 dissimilarities and twice the total variation dissimilarity are equivalent, we have:

$$D_{j|L}(v_j^x, v_j^y) = 2 \cdot \max_{L' \subseteq L} |P_{l|j}(L'|v_j^x) - P_{l|j}(L'|v_j^y)|. \qquad (5.4.3)$$

The detailed proof on the equivalence between Equations (5.4.2) and (5.4.3) is specified by Proof (b) in the Appendix of Section 5.9.

In the absence of labels, the above (5.4.3) is adapted to satisfy our target problem by replacing the class label information with other attribute knowledge to enable unsupervised learning. We regard this interaction between attributes as inter-coupled similarity in terms of the co-occurrence comparisons of *ICP*. The most intuitive variant of (5.4.3) is *IRSP*:

Definition 5.4.2 (IRSP) *The* ***Inter-coupled Relative Similarity based on Power Set (IRSP)*** *between values v_j^x and v_j^y of attribute a_j based on another attribute a_k is defined as $\delta_{j|k}^P(v_j^x, v_j^y, V_k)$ (below $\delta_{j|k}^P$ for short):*

$$\delta_{j|k}^P = \min_{V_k' \subseteq V_k} \{2 - P_{k|j}(V_k'|v_j^x) - P_{k|j}(\overline{V_k'}|v_j^y)\}, \tag{5.4.4}$$

where $\overline{V_k'} = V_k \backslash V_k'$ is the complementary set of a set V_k' under the complete value set V_k of attribute a_k.

The main differences between (5.4.4) and (5.4.3) are: 1) the multiplier 2 in (5.4.3) is omitted; 2) labels are replaced with other values of a particular attribute a_k, i.e., V_k' and V_k are substituted for L' and L, respectively; 3) a complementary set $\overline{V_k'}$ rather than the original set V_k' is concerned for v_j^y in *ICP*, note that $P_{k|j}(\overline{V_k'}|v_j^y)\} = 1 - P_{k|j}(V_k'|v_j^y)\}$; and 4) dissimilarity is considered rather than similarity: the new dissimilarity measure

$$D'_{j|k}(v_j^x, v_j^y) = \max_{V_k' \subseteq V_k} |P_{k|j}(V_k'|v_j^x) + P_{k|j}(\overline{V_k'}|v_j^y) - 1| \tag{5.4.5}$$

is obtained by following the previous three steps, then we have $\delta_{j|k}^P = 1 - D'_{j|k}(v_j^x, v_j^y)$. The detailed conversion process and relevant proof are provided in Proof (c) in the Appendix of Section 5.9. In fact, two attribute values are closer to each other if they have more similar probabilities with other attribute value subsets in terms of co-occurrence object frequencies.

In Table 5.2, by employing (5.4.4), we want to obtain $\delta_{2|1}^P(\mathcal{B}_1, \mathcal{B}_2, \{\mathcal{A}_i\}_{i=1}^4)$, i.e. the similarity between two attribute values $\mathcal{B}_1, \mathcal{B}_2$ of attribute a_2 regarding attribute a_1. As shown in Table 5.4.2, the set of all attribute values of

Table 5.4: Example of Computing Similarity Using *IRSP*

V_1'	$\overline{V_1'}$	$P_{1\|2}(V_1'\|\mathcal{B}_1)$	$P_{1\|2}(\overline{V_1'}\|\mathcal{B}_2)$	$2-P_{1\|2}(V_1'\|\mathcal{B}_1)$ $-P_{1\|2}(\overline{V_1'}\|\mathcal{B}_2)$
$\varnothing$	$\{\mathcal{A}_1,\mathcal{A}_2,\mathcal{A}_3,\mathcal{A}_4\}$	0	1	1
$\{\mathcal{A}_1\}$	$\{\mathcal{A}_2,\mathcal{A}_3,\mathcal{A}_4\}$	0.5	1	0.5
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$\{\mathcal{A}_1,\mathcal{A}_2,\mathcal{A}_3,\mathcal{A}_4\}$	$\varnothing$	1	0	1

attribute a_1 is $V_1 = \{\mathcal{A}_1,\mathcal{A}_2,\mathcal{A}_3,\mathcal{A}_4\}$. The number of all power sets within V_1 is 2^4, i.e., the number of the combinations consisting of $V_1' \subseteq V_1$ and $\overline{V_1'} \subseteq V_1$ is 2^4. The minimal value among them is 0.5, which indicates that the corresponding similarity $\delta^P_{2|1}$ is 0.5.

This process shows that the combinational explosion brought about by the power set needs to be considered when calculating attribute value similarity by *IRSP*. For a given set of attribute values, the power set considers all the subsets, the universal set concerns all the elements involved, and the join and intersection sets focus on parts of the elements. We start with the power set-based *IRSP*, and will proceed to the universal set-based *IRSU*, the join set-based *IRSJ*, and the intersection set-based *IRSI* to see whether the problem can be reduced in this way. We therefore try to define three more similarity metrics *IRSU, IRSJ, IRSI* based on *IRSP*.

Definition 5.4.3 (IRSU, IRSJ, IRSI) *The* ***Inter-coupled Relative Similarity based on Universal Set (IRSU), Join Set (IRSJ), and Intersection Set (IRSI)*** *between values v_j^x and v_j^y of attribute a_j based on another attribute a_k are defined as $\delta^U_{j|k}(v_j^x, v_j^y, V_k)$, $\delta^J_{j|k}(v_j^x, v_j^y, V_k)$ and $\delta^I_{j|k}(v_j^x, v_j^y, V_k)$ (below $\delta_{j|k}$, $\delta^J_{j|k}$, and $\delta^I_{j|k}$ for short), respectively:*

$$\delta^U_{j|k} = 2 - \sum_{v_k \in V_k} \max\{P_{k|j}(\{v_k\}|v_j^x), P_{k|j}(\{v_k\}|v_j^y)\}, \tag{5.4.6}$$

$$\delta^J_{j|k} = 2 - \sum_{v_k \in \bigcup} \max\{P_{k|j}(\{v_k\}|v_j^x), P_{k|j}(\{v_k\}|v_j^y)\}, \tag{5.4.7}$$

$$\delta^I_{j|k} = \sum_{v_k \in \bigcap} \min\{P_{k|j}(\{v_k\}|v_j^x), P_{k|j}(\{v_k\}|v_j^y)\}, \tag{5.4.8}$$

where $v_k \in \bigcup$ *and* $v_k \in \bigcap$ *denote* $v_k \in \varphi_{j\to k}(x) \bigcup \varphi_{j\to k}(y)$ *and* $v_k \in \varphi_{j\to k}(v_j^x) \bigcap \varphi_{j\to k}(v_j^y)$, *respectively.*

In the above, each value $v_k (\in V_k)$ of attribute a_k, rather than its value subset $V_k' \subseteq V_k$, is considered to reduce computational complexity. As shown in Table 5.5, the similarity $\delta_{2|1}^U$ based on *IRSU* is $\delta_{2|1}^U(\mathcal{B}_1, \mathcal{B}_2, \{\mathcal{A}_i\}_{i=1}^4) = 2 - 0.5 - 0.5 - 0 - 0.5 = 0.5$. Since *IRSU* only concerns all the single attribute values rather than exploring the whole power set, it solves the combinational explosion issue to a great extent. In *IRSU*, *ICP* is merely calculated 8 times compared with 32 times by *IRSP*, which leads to a substantial improvement in efficiency.

Table 5.5: Computing Similarity Using *IRSU*

v_k	$P_{1\|2}(\{v_k\}\|\mathcal{B}_1)$	$P_{1\|2}(\{v_k\}\|\mathcal{B}_2)$	max
$\mathcal{A}_1$	0.5	0	0.5
$\mathcal{A}_2$	0.5	0.5	0.5
$\mathcal{A}_3$	0	0	0
$\mathcal{A}_4$	0	0.5	0.5

IIF (5.2.3) is used to further reduce the time cost of *ICP* with two more similarity measures: *IRSJ* (5.4.7) and *IRSI* (5.4.8). With (5.4.7), the calculation of $\delta_{2|1}^J$ is further simplified since $\mathcal{A}_3 \notin \varphi_{2\to 1}(\mathcal{B}_1) \bigcup \varphi_{2\to 1}(\mathcal{B}_2)$. As shown in Table 5.6, we obtain $\delta_{2|1}^J(\mathcal{B}_1, \mathcal{B}_2, \{\mathcal{A}_i\}_{i=1}^4) = 2 - 0.5 - 0.5 - 0.5 = 0.5$, which reveals the fact that it is enough to compute *ICP* with $w \in V_1$ that belongs to $\varphi_{2\to 1}(\mathcal{B}_1) \bigcup \varphi_{2\to 1}(\mathcal{B}_2)$ instead of all the elements in V_1. From this aspect, *IRSJ* further reduces the complexity compared to *IRSU*.

Based on *IRSU*, an alternative *IRSI* is concerned. With (5.4.8), the calculation of $\delta_{2|1}^I$ is once again simplified as in Table 5.7 since only $A_2 \in \varphi_{2\to 1}(\mathcal{B}_1) \bigcap \varphi_{2\to 1}(\mathcal{B}_2)$. Then, we easily get $\delta_{2|1}^I(\mathcal{B}_1, \mathcal{B}_2, \{\mathcal{A}_i\}_{i=1}^4) = 0.5$. In this case, it is sufficient to compute *ICP* with $\mathcal{A}_2 \in V_1$ which only belongs to $\varphi_{2\to 1}(\mathcal{B}_1) \bigcap \varphi_{2\to 1}(\mathcal{B}_2)$. It is trivial that the cardinality of intersection $\bigcap$ is no larger than that of join set $\bigcup$. Thus, *IRSI* is more efficient than *IRSU* due to the reduction of intra-coupled relative similarity complexity.

Table 5.6: Computing Similarity Using *IRSJ*

v_k	$P_{1\|2}(\{v_k\}\|\mathcal{B}_1)$	$P_{1\|2}(\{v_k\}\|\mathcal{B}_2)$	max
$\mathcal{A}_1$	0.5	0	0.5
$\mathcal{A}_2$	0.5	0.5	0.5
$\mathcal{A}_4$	0	0.5	0.5

Table 5.7: Computing Similarity Using *IRSI*

v_k	$P_{1\|2}(\{v_k\}\|\mathcal{B}_1)$	$P_{1\|2}(\{v_k\}\|\mathcal{B}_2)$	min
$\mathcal{A}_2$	0.5	0.5	0.5

Intuitively, *IRSI* is the most efficient of all the proposed inter-coupled relative similarity measures: *IRSP*, *IRSU*, *IRSJ*, *IRSI*. In fact, all four measures lead to the same similarity result, such as 0.5 in our example. These measures are mathematically equivalent to one another. This assumption is proved in Section 5.5.

Accordingly, the similarity between the value pair (v_j^x, v_j^y) of attribute a_j can be calculated on top of these four optional measures by aggregating all the relative similarity on attributes other than a_j.

Definition 5.4.4 (IeASV) *The **Inter-coupled Attribute Similarity for Values (IeASV)** between attribute values v_j^x and v_j^y of attribute a_j is:*

$$\delta_j^{Ie}(v_j^x, v_j^y, \{V_k\}_{k\neq j}) = \sum_{k=1,k\neq j}^{n} \alpha_k \delta_{j|k}(v_j^x, v_j^y, V_k), \qquad (5.4.9)$$

where α_k is the weight parameter for attribute a_k, $\sum_{k=1,k\neq j}^{n} \alpha_k = 1$, $\alpha_k \in [0,1]$, and $\delta_{j|k}(v_j^x, v_j^y, V_k)$ is one of the inter-coupled relative similarity candidates.

Therefore, $\delta_j^{Ie} \in [0,1]$. For the parameter α_k, in this chapter, we simply assign $\alpha_k = 1/(n-1)$. For example, in Table 5.2, we then have $\delta_2^{Ie}(\mathcal{B}_1, \mathcal{B}_2, \{V_1, V_3\}) = 0.5 \cdot \delta_{2|1}(\mathcal{B}_1, \mathcal{B}_2, \{\mathcal{A}_i\}_{i=1}^4) + 0.5 \cdot \delta_{2|3}(\mathcal{B}_1, \mathcal{B}_2, \{\mathcal{C}_i\}_{i=1}^3) = 0.25$ if α_1 and α_3 equal to 0.5.

5.4.3 Coupled Interaction

So far, we have built formal definitions for both $IaASV$ and $IeASV$ measures. $IaASV$ emphasizes the attribute value occurrence frequency, while $IeASV$ focuses on the co-occurrence comparison of ICP with four inter-coupled relative similarity options. Then, the *Coupled Attribute Similarity for Values (CASV)* is naturally derived by simultaneously considering both measures.

Definition 5.4.5 (CASV) *The* ***Coupled Attribute Similarity for Values (CASV)*** *between attribute values v_j^x and v_j^y of attribute a_j is:*

$$\delta_j^A(v_j^x, v_j^y, \{V_k\}_{k=1}^n) = \delta_j^{Ia}(v_j^x, v_j^y) \cdot \delta_j^{Ie}(v_j^x, v_j^y, \{V_k\}_{k\neq j}), \tag{5.4.10}$$

where $V_k(k \neq j)$ is a value set of attribute a_k different from a_j to enable the inter-coupled interaction. δ_j^{Ia} and δ_j^{Ie} are IaASV and IeASV, respectively, which will be detailed in the following sections.

As indicated in Equation (5.4.10), $CASV$ gets larger by increasing either $IaASV$ or $IeASV$. Here, we choose the multiplication of these two components. The rationale is twofold: (1) $IaASV$ is associated with how often the value occurs while $IeASV$ reflects the extent of the value difference brought by other attributes, hence intuitively, the multiplication of them indicates the total amount of attribute value difference; (2) the multiplication method is consistent with the adapted simple matching distance introduced in (Gan et al. 2007). Alternatively, in our future work, we could consider other combination forms of $IaASV$ and $IeASV$ according to the data structure, such as $\delta_j^A(v_j^x, v_j^y, \{V_k\}_{k=1}^n) = \beta \cdot \delta_j^{Ia}(v_j^x, v_j^y) + \gamma \cdot \delta_j^{Ie}(v_j^x, v_j^y, \{V_k\}_{k\neq j})$, where $0 \leq \beta, \gamma \leq 1$ $(\beta + \gamma = 1)$ are the corresponding weights. Thus, $IaASV$ and $IeASV$ can be controlled flexibly to display in which cases the former is more significant than the latter, and vice-versa.

Additionally, $\delta_j^A = \delta_j^{Ia} \cdot \delta_j^{Ie} \in [0, m/(m+4)]$ since we have $\delta_j^{Ia} \in [1/3, m/(m+4)](m \geq 2)$ as well as $\delta_j^{Ie} \in [0, 1]$. For example, in Table 5.2, the $CASV$ of attribute values $\mathcal{B}_1$ and $\mathcal{B}_2$ is $\delta_2^A(\mathcal{B}_1, \mathcal{B}_2, \{V_1, V_2, V_3\}) = \delta_2^{Ia}(\mathcal{B}_1, \mathcal{B}_2) \cdot \delta_2^{Ie}(\mathcal{B}_1, \mathcal{B}_2, \{V_1, V_3\}) = 0.5 \times 0.25 = 0.125$. For the Movie data set, then $\delta_{Director}^A(Scorsese,$

$Coppola) = \delta^A_{Director}(Coppola, Coppola) = 0.33$, and $\delta^A_{Director}(Koster, Coppola)$ $= 0$ while $\delta^A_{Director}$ $(Koster, Hitchcock) = 0.25$. They correspond to the fact that "*Scorsese*" and "*Coppola*" are very similar directors just as "*Coppola*" is to himself, and the similarity between "*Koster*" and "*Hitchcock*" is larger than that between "*Koster*" and "*Coppola*", as clarified in Section 5.1.

In the following theoretical analysis in Section 5.5, the computational accuracy and complexity of the four inter-coupled relative similarity options are analyzed.

5.5 Theoretical Analysis

This section compares the proposed four inter-coupled relative similarity measures (*IRSP*, *IRSU*, *IRSJ* and *IRSI*) in terms of their computational accuracy and complexity.

1) Accuracy Equivalence

According to the set theory, these four measures are equivalent to one another in calculating value similarity. We therefore have the following theorem, which is deduced by Proof (d) in the Appendix of Section 5.9.

Theorem 5.5.1 *Similarity measures IRSP, IRSU, IRSJ and IRSI are all equivalent to one another.*

The above theorem indicates that *IRSP*, *IRSU*, *IRSJ*, and *IRSI* are equivalent to one another in terms of the information and knowledge they present. It also explains the similarity result in Section 5.4.2. Thus, these measures can induce exactly the same computational accuracy in different learning tasks including classification and clustering.

2) Computational Complexity Comparison

When calculating the similarity between every pair of attribute values for all attributes, the computational complexity linearly depends on the time cost of *ICP*, which is quantified by the calculation counts of *ICP*. This reflects the efficiency difference between distinct similarity measures. Table

Table 5.8: Time Cost of *ICP*

Metric	Calculation Times of ICP	$\delta_{2\|1}(\mathcal{B}_1, \mathcal{B}_2)$
IRSP	$2 \cdot 2^{\|V_k\|}$	32
IRSU	$2 \cdot \|V_k\|$	8
IRSJ	$2 \cdot \|\varphi_{j\to k}(v_j^x) \bigcup \varphi_{j\to k}(v_j^y)\|$	6
IRSI	$2 \cdot \|\varphi_{j\to k}(v_j^x) \bigcap \varphi_{j\to k}(v_j^y)\|$	2

5.8 summarizes the time costs of the four inter-coupled relative similarity measures.

Let $|ICP_{j|k}^{(M)}|$ represent the time cost of *ICP* for $\delta_{j|k}^{M}(v_j^x, v_j^y)$ with the associated measure $M = \{P, U, J, I\}$, whose elements are *IRSP*, *IRSU*, *IRSJ*, and *IRSI*, respectively. From Table 5.8, $|ICP_{j|k}^{(P)}| \geq |ICP_{j|k}^{(U)}| \geq |ICP_{j|k}^{(J)}| \geq |ICP_{j|k}^{(I)}|$ holds constantly. It demonstrates the competitive efficiency of *IRSI* compared to the other three measures. In Table 5.2, 32 calculation counts of *ICP* are required in *IRSP*, compared with only two calculation counts when using *IRSI*.

Suppose the maximal number of values for all the attributes is $R(= \max_{j=1}^{n} |V_j|)$. In total, the number of value pairs for all the attributes is at most $n \cdot R(R-1)/2$, which is also the number of calculation steps. For each inter-coupled relative similarity, we calculate *ICP* for $|ICP_{j|k}^{(M)}|$ times. As we have n attributes, the total *ICP* time cost for $CASV$ is $2 \cdot |ICP_{j|k}^{(M)}| \cdot (n-1)$ flops per step. The computational complexity for calculating all four options of $CASV$ is shown in Table 5.9.

As indicated in Table 5.9, all the measures have the same calculation steps, while their flops per step are sorted in descending order since $2^R > R \geq R_{\cup} \geq R_{\cap}$, in which $R_{\cup}$ and $R_{\cap}$ are the cardinality of the join and intersection sets of the corresponding *IIF*s, respectively. This evidences that the computational complexity essentially depends on the time cost of *ICP* linearly with given data. Specifically, *IRSP* has the largest complexity $O(n^2 R^2 2^R)$, compared to the smaller equal ones $O(n^2 R^3)$ presented by the other three measures (*IRSU*, *IRSJ*, and *IRSI*). Of the latter three candidates, though

Table 5.9: Computational Complexity for $CASV$

Metric	Calculation Steps	Flops per Step	Complexity
IRSP	$nR(R-1)/2$	$2(n-1)2^R$	$O(n^2R^22^R)$
IRSU	$nR(R-1)/2$	$2(n-1)R$	$O(n^2R^2R)$
IRSJ	$nR(R-1)/2$	$2(n-1)R_\cup$	$O(n^2R^2R)$
IRSI	$nR(R-1)/2$	$2(n-1)R_\cap$	$O(n^2R^2R)$

they have the same computational complexity, *IRSI* is the most efficient due to $R_\cap \leq R_\cup \leq R$. In fact, the dissimilarity *ADD* that Ahmad and Dey (Ahmad & Dey 2007) used for mixed data clustering corresponds to the worst measure *IRSP*.

Considering both the accuracy analysis and complexity comparison, we conclude that *IRSI* is the best performing measure because it indicates the least complexity but maintains equal accuracy to present couplings.

5.6 Coupled Similarity Algorithm

In previous sections, we have discussed the construction of $CASV$ and its theoretical comparison among the inter-coupled relative similarity candidates. In this section, a coupled similarity between objects is built based on $CASV$. Below, we consider the sum of all these $CASV$ measures, following the Manhattan dissimilarity (Gan et al. 2007).

Definition 5.6.1 (CASO) *Given an information table S, the* ***Coupled Attribute Similarity for Objects (CASO)*** *between objects u_x and u_y is $CASO(u_x, u_y)$:*

$$CASO(u_x, u_y) = \sum_{j=1}^{n} \delta_j^A(v_j^x, v_j^y, \{V_k\}_{k=1}^n), \tag{5.6.1}$$

where δ_j^A is the CASV measure defined in (5.4.10), v_j^x and v_j^y are the attribute values of attribute a_j for objects u_x and u_y respectively, and $1 \leq x, y \leq m$, $1 \leq j \leq n$.

For *CASO*, all the *CASV*s with each attribute are summed up for two objects. For example the similarity between u_2 and u_3 in Table 5.2 is $CASO(u_2, u_3) = \sum_{j=1}^{3} \delta_j^A(v_j^2, v_j^3, \{V_k\}_{k=1}^3) = 0.5 + 0.125 + 0.125 = 0.75$.

CASO has the properties of *non-negativity* because $CASO(u_x, u_y) \in [0, mn/(m+4)]$, in particular $CASO(u_x, u_x) \in [n/3, mn/(m+4)]$, and *symmetry*, i.e. $CASO(u_x, u_y) = CASO(u_y, u_x)$, but it does not guarantee the property of *triangle inequality*. So *CASO* is a non-metric similarity measure.

We then design an algorithm *CASO_IRSI*, given below, to compute the coupled object similarity with *IRSI* (i.e. the best inter-coupled relative similarity). The whole process of this algorithm is summarized as follows: (1) Compute the *IaASV* for values v_j^x and v_j^y of attribute a_j (Line 5); (2) Compute the *IeASV* for attribute values v_j^x and v_j^y based on *IRSI* (Line 10 to Line 20); (3) Compute the *CASV* for attribute values v_j^x and v_j^y (Line 6); (4) Compute the *CASO* for two objects u_x and u_y (Line 7).

Before the similarity calculation is performed, some data preprocessing is conducted to enable this algorithm. In detail, all the categories of each attribute need to be encoded as numberings, starting at one and increasing to the maximum, which is the respective number of attribute values. To reduce unnecessary iterations in Line 7, pairwise *CASV* similarity for any two values of the same attribute, rather than the only two values involved of each attribute, is pre-calculated for reuse when computing the object similarity. Explicitly, this pseudocode also embodies the fact that the computational complexity for *IRIS* is indeed $O(n^2R^3)$. However, it might not be very attractive for extremely large data sets with attributes that take too many values. Thus, we are working on strategies of attribute reduction to effectively reduce the number of coupled attributes.

5.7 Experiments and Evaluation

In this section, several experiments are performed on extensive UCI data sets and bibliographic data (i.e. Table 5.10) to show the effectiveness and effi-

Algorithm 5.1: Coupled Attribute Similarity for Objects *CASO_IRSI*

```
Data: Data set $S_{m\times n}$ with $m$ objects and $n$ attributes, object
      $u_x, u_y (x, y \in [1, m])$, and weight $\alpha = (\alpha_k)_{1\times n}$.
Result: Coupled Similarity for objects $CASO(u_x, u_y)$.
begin
    // Compute pairwise similarity for any two values of the
       same attribute.
    for attribute $a_j$, $j = 1 : n$ do
        for every value pair $(v_j^x, v_j^y \in [1, |V_j|])$ do
            $U_1 \longleftarrow \{i | v_j^i == v_j^x\}$, $U_2 \longleftarrow \{i | v_j^i == v_j^y\}$;
            // Compute intra-coupled similarity for values $v_j^x, v_j^y$.
            $\delta_j^{Ia}(v_j^x, v_j^y) = (|U_1||U_2|)/(|U_1| + |U_2| + |U_1||U_2|)$;
            // Compute coupled similarity for two attribute
               values $v_j^x$ and $v_j^y$.
            $\delta_j^A(v_j^x, v_j^y, \{V_k\}_{k=1}^n) \longleftarrow \delta_j^{Ia}(v_j^x, v_j^y) \cdot IeASV(v_j^x, v_j^y, \{V_k\}_{k\neq j})$;
    // Compute coupled similarity between two objects $u_x, u_y$.
    $CASO(u_x, u_y) \longleftarrow sum(\delta_j^A(v_j^x, v_j^y, \{V_k\}_{k=1}^n))$;
    end
Function $IeASV(v_j^x, v_j^y, \{V_k\}_{k\neq j})$
begin
    // Compute inter-coupled similarity for two values $v_j^x, v_j^y$.
    for attribute $(k = 1 : n) \wedge (k \neq j)$ do
        $\{v_k^z\}_{z\in U_3} \longleftarrow \{v_k^x\}_{x\in U_1} \bigcap \{v_k^y\}_{y\in U_2}$;
        for intersection $z = U_3(1) : U_3(|U_3|)$ do
            $U_0 \longleftarrow \{i | v_k^i == v_k^z\}$;
            $ICP_x \longleftarrow |U_0 \bigcap U_1| / |U_1|$;
            $ICP_y \longleftarrow |U_0 \bigcap U_2| / |U_2|$;
            $Min_{(x,y)} \longleftarrow min(ICP_x, ICP_y)$;
        // Compute IRSI for $v_j^x$ and $v_j^y$.
        $\delta_{j|k}^I(v_j^x, v_j^y, V_k) = sum(Min_{(x,y)})$;
    $\delta_j^{Ie}(v_j^x, v_j^y, \{V_k\}_{k\neq j}) = sum[\alpha(k) \times \delta_{j|k}^I(v_j^x, v_j^y, V_k)]$;
    return $\delta_j^{Ie}(v_j^x, v_j^y, \{V_k\}_{k\neq j})$;
```

Table 5.10: Description of Data Sets in Chapter 5

Data Set	Object	Attribute	Class
Movie	6	3	2
MMR	10	6	3
Shuttle	15	6	2
Balloon	20	4	2
Lense	24	4	3
Corral	32	6	2
Soybean-small	47	35	4
Zoo	101	16	7
Molecular	106	57	2
DNA	106	59	2
Hayesroth	132	4	3
Led24	200	24	10
SPECT	267	22	2
Soybean-large	307	35	19
Voting	435	16	2
Breastcancer	699	9	2
Tic	958	9	2
Solar	1389	10	3
Car	1728	6	4
Letter	2341	16	3
Chess	3196	36	2
Mushroom	8124	22	2
Adult	30718	13	2
Bibliographic Data	720	4	2

ciency of our proposed coupled similarity measures. All the experiments are conducted on a Dell Optiplex 960 equipped with an Intel Core 2 Duo CPU with a clock speed of 2.99 GHz and 3.25 GB of RAM running Microsoft Windows XP. The experiments are divided in two categories: coupled similarity comparisons and *CASO* applications. For simplicity, we assign the weight vector $\alpha = (\alpha_k)_{1\times n}$ with values $\alpha(k) = 1/(n-1)$ in Definition 5.4.4.

5.7.1 Coupled Similarity Comparison

To compare efficiency, we conduct extensive experiments on the inter-coupled relative similarity measures: *IRSP*, *IRSU*, *IRSJ*, and *IRSI*. Experiments are first performed for efficiency comparison, followed by scalability analysis. The time cost of *ICP* is quantified by the calculation counts of *ICP*.

Efficiency Comparison

The goal in this set of experiments is to show the obvious superiority of *IRSI* compared with the most time-consuming measure *IRSP*. As discussed in Section 5.5, the computational complexity linearly depends on the time cost of *ICP* with given data. Thus, we consider the comparison of complexity represented by the time cost of *ICP* from the following two aspects.

In terms of a single attribute, the time costs of *ICP* on *Movie* (Ahmad & Dey 2007), *MMR*, *Soybean-small* and *Zoo* data sets are shown in Figure 5.2. We only consider the attributes whose number of values is more than 1, thus, there are only 24 attributes for *Soybean-small* rather than 35. The horizontal axis refers to the ordinal number of nominal attributes, e.g., 1 indicates attribute a_1; while the vertical axis indicates the total time cost (i.e. calculation counts) of *ICP* for all value pairs of each attribute with four options: *IRSP*, *IRSU*, *IRSJ*, *IRSI*. The results show that for any individual attribute, *IRSI* always has the smallest time cost, followed by *IRSJ* and *IRSU*, while *IRSP* is far more time-consuming.

In more detail, we observe that the complexity of *IRSP* for each attribute is around three or four times the size of *IRSU* for these four data sets.

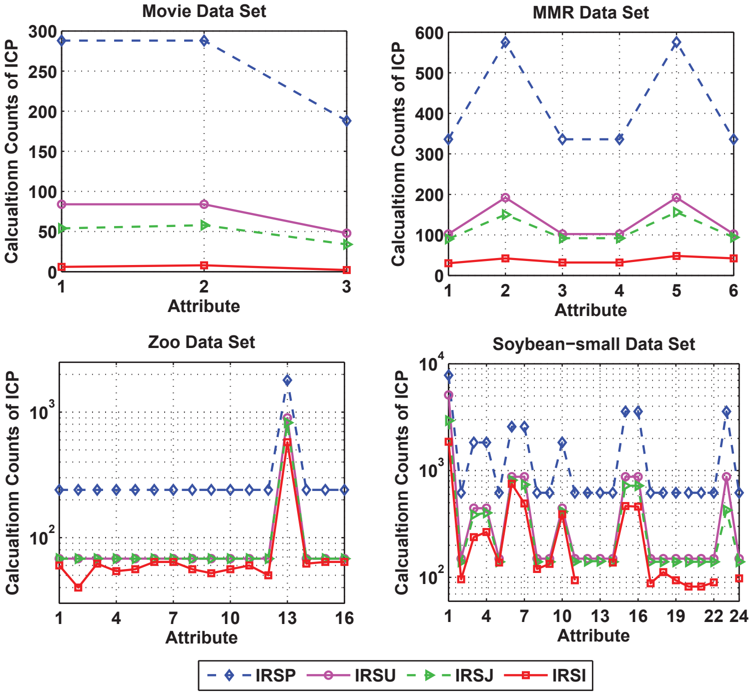


Figure 5.2: Complexity on individual attributes.

Theoretically, this ratio $\xi(P/U)$ can be fixed within an interval based on the given data structure. Suppose we have an information table S with m objects and n attributes. For all the attributes, let $T(=\min_{k=1}^{n}|V_j|)$ and $R(=\max_{k=1}^{n}|V_j|)$ be their minimal and maximal number of values, respectively. Then, for any attribute a_j:

$$\xi_j(P/U) = \frac{|ICP_j^{(P)}|}{|ICP_j^{(U)}|} \in \left[\frac{2^T}{T}, \frac{2^R}{R}\right], \tag{5.7.1}$$

where $|ICP_j^{(M)}|$ is the time cost of ICP for a_j. Proof (e) in the Appendix of Section 5.9 supports this statement. For *Zoo*, $T = 2$ and $R = 6$, and the corresponding multiples ξ_j, which range from 2.0 to 3.5, all fall in $[2, 10.7]$.

With respect to all attributes, all the time costs of ICP for all the attribute value pairs are considered. Table 5.11 reports the total time cost of ICP with four measures on 12 data sets in terms of relative proportion and

Table 5.11: Complexity Comparison on All Attributes

Data Set	**Corral**	**Voting**	**Led24**	**Lense**	**Tic**	**Chess**
R	2	2	2	3	3	3
T	2	2	2	2	3	2
n	6	16	24	4	9	36
$\xi(U/P)$	50.0%	50.0%	50.0%	46.4%	37.5%	49.4%
$\xi(I/J)$	100%	100%	100%	100%	100%	88.7%
$\|ICP^{(U)}\|$	120	960	2208	78	1296	5390
$\|ICP^{(I)}\|$	120	960	2208	78	1296	4774
Data Set	**Movie**	**Hayesroth**	**Molecular**	**Solar**	**Mushroom**	**Letter**
R	4	4	4	7	12	16
T	3	3	4	2	1	10
n	3	4	57	10	22	16
$\xi(U/P)$	27.8%	27.1%	25.0%	20.0%	1.7%	0.1%
$\xi(I/J)$	11.0%	100%	99.2%	82.3%	42.5%	48.4%
$\|ICP^{(U)}\|$	212	468	153216	2544	76020	394294
$\|ICP^{(I)}\|$	16	468	152022	1998	21736	140434

direct frequency, where R and n denote the maximal number of attribute values and the number of attributes, respectively, and $|ICP^{(M)}|$ indicates the total time cost of ICP for all the attributes. Let $\xi(U/P)$ and $\xi(I/J)$ denote the proportions $|ICP^{(U)}|/|ICP^{(P)}|$ and $|ICP^{(I)}|/|ICP^{(J)}|$, respectively. Then $\xi(U/P) \in [R/2^R, T/2^T]$ is deduced according to the proof of Equation (5.7.1). This property can be checked in Table 5.11, 27.1% ∈ [25%, 37.5%] for the data set *Hayesroth*.

These results also show that the efficiency advantage of *IRSU* over *IRSP* becomes more obvious when the maximal number of values R becomes larger, i.e., the proportion $\xi(U/P)$ reduces monotonously from 50% to 0.1% when R increases from 2 to 16. However, due to the fact that *IRSJ* and *IRSI* involve the relevant join set and intersection set respectively, the variation tendency of their relative efficiency ratio $\xi(I/J) \in [0, 1]$ mainly depends on the data structure rather than R and n alone. The probability of achieving a smaller ratio $\xi(I/J)$ increases as R grows, since we have more opportunity to obtain an intersection set smaller than a join set. This can be observed in Table 5.11 by the fact that there is a general decreasing tendency that nevertheless has several disorder ratios.

After fixing R, we consider the variation law for the efficiency of *IRSU* and *IRSI* with the increasing n. It is found that the *ICP* time costs of both measures become greater as n grows. For instance, the calculation frequency of *ICP* for *IRSI* increases from 78 to 4774 when n varies between 4 and 36 with $R = 3$. Similarly, the time costs of the other two options (*IRSU* and *IRSI*) also increase when either n or R goes up. The superiority of *IRSI* becomes more remarkable as the data grows more complicated and bigger compared to the other three metrics. Table 5.11 further evidences that *IRSI* is the most efficient measure in contrast to the worst measure, *IRSP*.

Scalability Analysis

As we have discussed in Section 5.5, the complexity for *IRSP* is $O(n^2R^22^R)$, while the other three have equal smaller complexity $O(n^2R^3)$. Here, scalability analysis is explored in terms of the number of attributes n and the maximal number of attribute values R separately.

From the perspective of the number of attributes n, the *Soybean-large* data set is considered with 307 objects and 35 attributes. Here, we fix R as 7, and focus on n ranging from 5 to 35 with a step length of 5. In terms of the total time cost of *ICP*, the computational complexity comparisons among four measures (*IRSP*, *IRSU*, *IRSJ*, and *IRSI*) are depicted in Figure 5.3 (a). The result indicates that the complexity of all these measures keeps increasing when n becomes larger. The acceleration of *IRSP* (from 3328 to 74128) is the greatest by contrast to the slightest acceleration of *IRSI* (from 632 to 15704). Apart from these two, the scalability curves are almost the same for *IRSU* and *IRSI*, though the complexity of *IRSU* is slightly higher than that of *IRSJ* with varied n. Therefore, *IRSI* is the most stable and efficient measure for calculating the intra-coupled relative similarity in terms of the scalability on n.

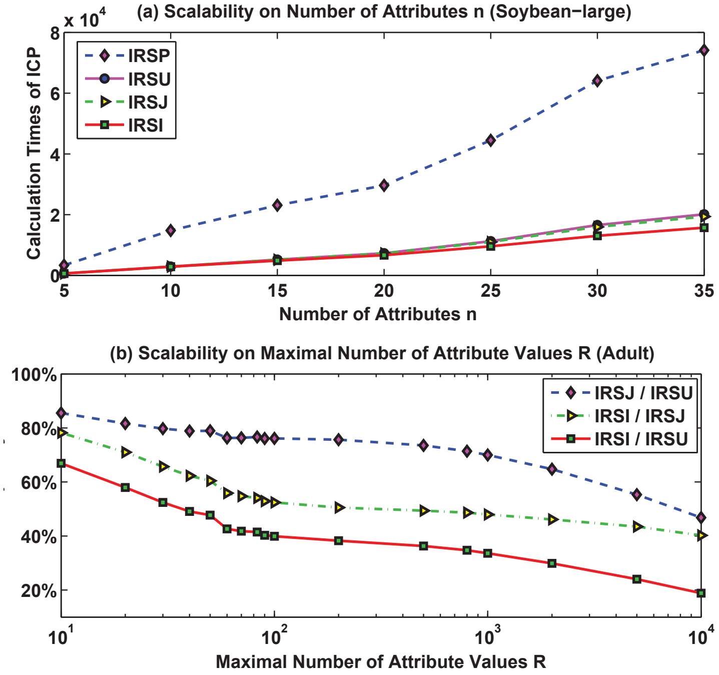


Figure 5.3: Scalability on n and R respectively.

From the perspective of the maximal number of attribute values R, the variation of R is considered when n is fixed. Here, we take advantage of the *Adult* data set with 30718 objects and 13 attributes chosen. Specifically, the integer attribute "fnlwgt" is discretized into different intervals (from 10 to 10000) to form distinct R ranging from 16 to 10000, since one of the existing categorial attributes "education" already has 16 values. The outcomes are shown in Figure 5.3(b), in which the horizontal axis refers to R, and the vertical axis indicates the relative complexity ratios in terms of $\xi(J/U)$, $\xi(I/J)$, and $\xi(I/U)$. From this figure, we observe all the ratios between 10% and 100%, which again verifies the complexity order for these four measures indicated in Section 5.5. Another issue is that all three curves decrease as R grows, which means the efficiency advantage of *IRSJ* over *IRSU* (from 85.5% to 46.8%), *IRSI* over *IRSJ* (from 78.2% to 40.2%), and *IRSI* over *IRSU* (from 66.9% to 18.8%) all become more and more obvious with the increase of R. The general downturn trend of these ratios comes from the fact that there is a higher probability of obtaining a join set smaller than the whole set, and an intersection set smaller than the join set, with larger R. The same conclusion also holds for the ratio $\xi(U/P)$, but this is due to the monotonously decreasing property of $\xi(U/P)$ on R, which has been proved in Proof (f) in the Appendix of Section 5.9. We omit this ratio in Figure 5.3(b) since the denominator $|ICP^{(P)}|$ becomes exponentially large when R grows, e.g., it equals to 5.12×10^{83} when $R = 500$. Hence, *IRSI* is the least time-consuming intra-coupled similarity with regard to scalability on R.

In summary, all of the above experiment results clearly show that *IRSI* outperforms *IRSU*, *IRSJ* and *IRSI* on computational complexity, no matter how small or large, simple or complicated a data set is. In particular, with the increase in the number of either attributes or attribute values, *IRSI* demonstrates superior efficiency compared to the others. *IRSJ* and *IRSU* follow, with *IRSP* being the most time-consuming, especially for large-scale data.

5.7.2 Learning Applications

In this part of our experiments, we focus on two levels of algorithmic accuracy comparison as follows:

1. Compare the proposed four intra-coupled measures: *IRSP*, *IRSU*, *IRSJ*, and *IRSI*.
2. Compare our novel *Coupled Attribute Dissimilarity for Objects (CADO)* induced from *CASO* with existing categorical dissimilarity measures.

Three independent groups of experiments are conducted with extensive data sets based on machine learning applications. In the following, we evaluate the *CADO* which is derived from (5.6.1):

$$CADO(u_x, u_y) = \sum_{j=1}^{n} h_1[\delta_j^{Ia}(v_j^x, v_j^y)] \cdot h_2[\delta_j^{Ie}(v_j^x, v_j^y, \{V_k\}_{k \neq j})], \qquad (5.7.2)$$

where $h_1(t)$ and $h_2(t)$ are decreasing functions. Based on intra-coupled and inter-coupled similarities, $h_1(t)$ and $h_2(t)$ can be flexibly chosen to build dissimilarity measures according to specific requirements. In terms of the capability of revealing the data relationship, the better the induced dissimilarity, the better is its similarity.

We consider $h_1(t) = 1/t - 1$ and $h_2(t) = 1 - t$ here to reflect the complementarity of similarity and dissimilarity measures, since they are both decreasing functions of t. Moreover, the rationale behind these two functions is as follows. The first conversion corresponds to the improved *SMD* with frequency (Gan et al. 2007), if only 0 and 1 are assigned to δ_j^{Ie} (i.e. *SMD* (Hollingsworth, Bowyer & Flynn 2011): dissimilarity 0 for identical values, and otherwise 1). The second transformation guarantees the consistency of *CADO* with the dissimilarity measure *ADD* (Ahmad & Dey 2007), when a constant is fixed for δ_j^{Ia}. In addition, $h_1(t) = 1/t - 1$ is also consistent with the converted measures proposed in (Lin 1998); $h_2(t) = 1 - t$ follows the way of converting *OF* to *OFD* (Boriah et al. 2008), presented in the next section. Both these functions are designed to include existing classical measures as special cases of our proposed coupled similarity. The detailed degenerations to the improved *SMD* and the *ADD* are explained in Section 7.4.2.

Data Structure Analysis

This section performs experiments to explicitly specify the internal structures for the labeled data. Clusterings are normally evaluated by assigning the best score to the algorithm that produces clusters with highest similarity within a cluster and lowest similarity between clusters based on a certain similarity measure. We work in a different way, in which similarity measures are evaluated with the clustering criteria and given labels. In other words, a better cluster structure can be clarified with a better similarity measure on the clustering internal descriptors, such as Sum-Square, Davies-Bouldin Index (DBI) (Davies & Bouldin 1979), and Dunn Index (DI) (Dunn 1974).

To reflect the data cluster structure more clearly, the induced dissimilarity

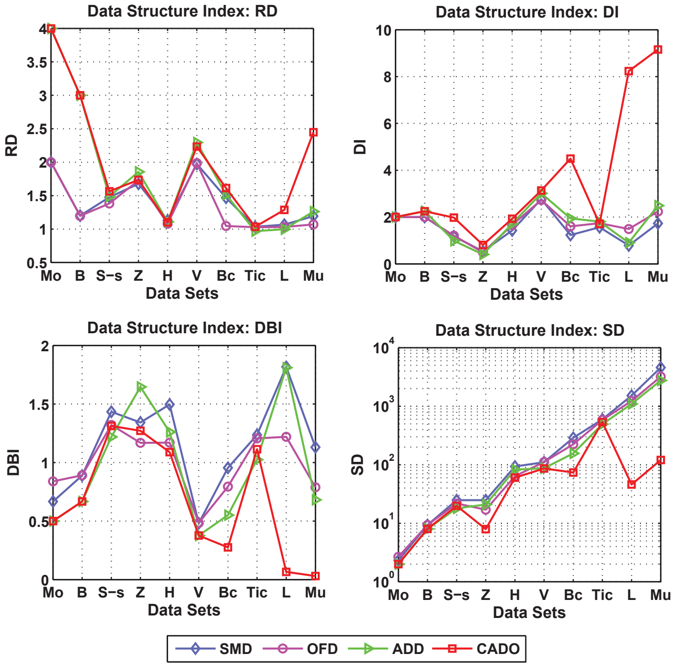


Figure 5.4: Data structure index comparison.

metrics are evaluated by four descriptors: Relative Dissimilarity (RD), DBI, DI, and Sum-Dissimilarity (SD). In detail, RD is the ratio of average inter-cluster dissimilarity upon average intra-cluster dissimilarity for all cluster labels; SD is the sum of object dissimilarities within all the clusters. Since internal criteria seek clusters with high intra-cluster similarity and low inter-cluster similarity, dissimilarity metrics that produce clusters with high RD or DI and low DBI or SD are more desirable.

Four object dissimilarity metrics are considered here: Simple Matching Dissimilarity (Gan et al. 2007) (*SMD*, i.e. Hamming distance (Hollingsworth et al. 2011)), Occurrence Frequency Dissimilarity (*OFD*) (Boriah et al. 2008), *ADD* proposed by Ahmad and Dey (Ahmad & Dey 2007), and *CADO*. *SMD* is a simple, well-known measure for categorical data, while *OFD* considers matching in terms of attribute value frequency distribution, both formalized as the sum of value dissimilarities for all the attributes. Further, attribute value dissimilarities $D_j^{SMD} = D_j^{OFD} = 0$ if $v_j^x = x_j^y$, otherwise they equal

1 and $1-\left[1+log\frac{m}{|G_j(\{v_j^x\})|}\cdot log\frac{m}{|G_j(\{v_j^y\})|}\right]^{-1}$ for *SMD* and *OFD*, respectively. The dissimilarity measure *ADD*, derived from (5.6.1) with the worst inter-coupled relative similarity candidate *IRSP*, considers the sum of inter-coupled interactions between all the corresponding attribute values. These three measures only concern the local picture, while our proposed *CADO* is globally formalized based on (5.7.2).

The cluster structures produced by the above four dissimilarity metrics are then analyzed on 10 data sets in different scales. The results after dissimilarity normalization are shown in Figure 5.4, where the X axis refers to the data sets *Movie*, *Balloon*, *Soybean-small*, *Zoo*, *Hayesroth*, *Voting*, *Breast-cancer*, *Tic*, *Letter*, and *Mushroom*, respectively. They are ordered according to the number of objects involved (i.e. m) to describe distinct data scales, ranging from 6 to 8124. As discussed previously, larger RD, larger DI, smaller DBI, and smaller SD indicate better characterization of the cluster differentiation capability, which corresponds to a better dissimilarity metric being induced. From Figure 5.4, we observe that, with the exception of a few items, the corresponding RD and DI indexes on *CADO* are mostly the largest ones when compared with those on *SMD*, *OFD*, and *ADD*; while the associated DBI and SD index curves on *CADO* are mostly below the other three curves. The results show that our proposed *CADO* is better than *SMD* and *OFD* in terms of differentiating objects in distinct clusters. *ADD* also seems to be slightly better than *SMD* and *OFD* in most cases. The degrees of improvement of *CADO* upon *SMD*, *OFD*, and *ADD* mainly depend on data structure rather than on data scale $|U|(=m)$ alone.

In constructing *CADO*, all four candidates (*IRSP*, *IRSU*, *IRSJ*, and *IRSI*) are used. Just as we proved in Section 5.5, all the indexes are the same regardless of exactly what $\delta_{j|k}(x,y)$ refers to, which directly verifies that these four intra-coupled relative similarity measures present equal accuracy.

Clustering Evaluation

To demonstrate the effectiveness of our proposed *CADO* and *CASO* in clustering applications, we conduct two groups of experiments: *KM* & *SC* and *ROCK* & *CROCK*. The former compares two classical clustering methods based on two dissimilarity metrics on six data sets. The latter considers the clustering quality of the adapted method *CROCK* by integrating our proposed *CASO* with the categorical clustering algorithm *ROCK* (Guha et al. 2000). *CADO* or *CASO* is used with the outperforming measure *IRSI*.

(1) *KM* & *SC*

One of the clustering approaches is the k-modes (*KM*) algorithm (Gan et al. 2007), designed to cluster categorical data sets. The main idea of *KM* is to specify the number of clusters k and then to select k initial modes, followed by allocating every object to the nearest mode. The other is a branch of graph-based clustering, i.e. spectral clustering (*SC*) (Luxburg 2007), which makes use of Laplacian Eigenmaps on a dissimilarity matrix to perform dimensionality reduction for clustering prior to the k-means algorithm. In respect of attribute dependency aggregation, however, Ahmad and Dey (Ahmad & Dey 2007) evidenced that their proposed metric *ADD* outperforms *SMD* in terms of *KM* clustering. Thus, we aim to compare the performances of *ADD* (Ahmad & Dey 2007) and *CADO* (5.7.2) for further clustering evaluation.

We conduct four groups of experiments on six UCI data sets: *KM* with *ADD*, *KM* with *CADO*, *SC* with *ADD*, and *SC* with *CADO*. The clustering performance is evaluated by comparing the obtained cluster of each object with that provided by the data label in terms of accuracy (AC) and normalized mutual information (NMI) (Cai et al. 2005), which are essentially the external criteria compared with the internal criterion analysis in Section 5.7.2. AC $\in [0, 1]$ is a degree of closeness between the obtained clusters and its actual data labels, while NMI $\in [0, 1]$ is a quantity that measures the mutual dependence of two variables: clusters and labels. The larger AC or NMI is, the better the clustering is, and the better the corresponding dissimilarity

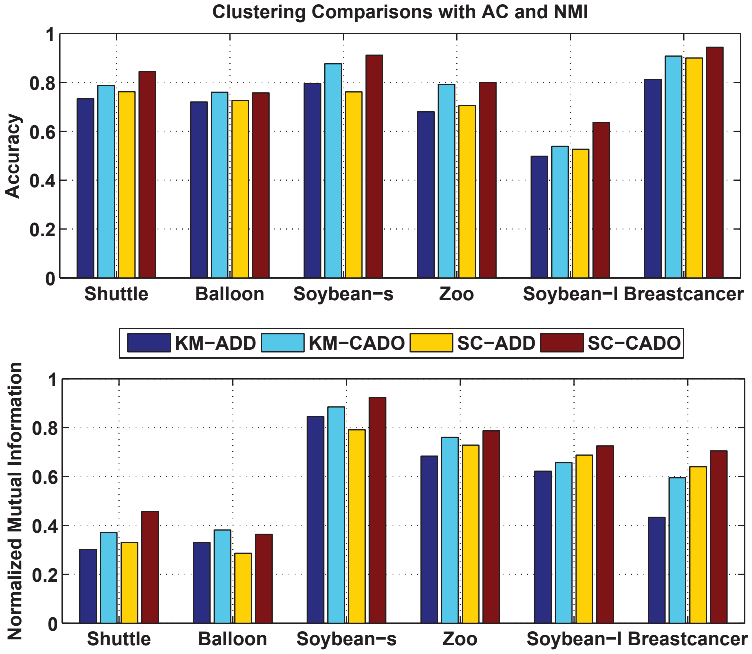


Figure 5.5: Clustering evaluation on six data sets.

metric is.

Figure 5.5 reports the results on six data sets with different $|U|$, ranging from 15 to 699 in the increasing order. The performance of AC and NMI is individually evaluated for *KM-ADD*, *KM-CADO*, *SC-ADD*, and *SC-CADO*. Followed by Laplacian Eigenmaps, the subspace dimensions are determined by the number of labels in *SC*. For each data set, the average performance is computed over 100 tests for *KM* and *SC* with distinct start points.

As can clearly be seen from Figure 5.5, the clustering methods with *CADO*, whether *KM* or *SC*, outperform those with *ADD* on both AC and NMI. That is to say, the dissimilarity metric *CADO* is better than *ADD* for measuring clustering quality. Specifically for *KM*, the AC improving rate ranges from 5.56% (*Balloon*) to 16.50% (*Zoo*), while the NMI improving rate falls within 4.76% (*Soybean-s*, i.e., *Soybean-small*) and 37.38% (*Breastcancer*). With regard to *SC*, the former rate takes the minimal and maximal ratios as 4.21% (*Balloon*) and 20.84% (*Soybean-l*, i.e., *Soybean-large*), re-

Table 5.12: *CROCK* vs *ROCK* on UCI Data Sets

Data Set ($\|U\|$)	*ROCK*			*CROCK*		
	Pr	Re	Sp	Pr	Re	Sp
Movie (6)	0.88	0.75	0.92	**1**	**1**	**1**
Hayesroth (132)	0.43	0.39	0.62	**0.45**	**0.44**	**0.67**
SPECT (267)	**0.73**	0.62	0.57	0.71	**0.76**	**0.62**
Voting (435)	0.88	0.88	0.89	**0.90**	**0.90**	**0.92**
Mushroom (8124)	0.78	0.65	**0.76**	**0.87**	**0.79**	0.74

spectively, however, the latter rate belongs to [5.45% (*Soybean-l*), 38.12% (*Shuttle*)]. Since AC and NMI evaluate clustering quality from different aspects, generally, they take minimal and maximal ratios on distinct data sets. Statistical analysis, namely the t-test, has been done on AC and NMI, at a 95% significance level. The null hypothesis that *CADO* is better than *ADD* in terms of AC and NMI is accepted. Another significant observation is that *SC* mostly outperforms *KM* whenever it has the same dissimilarity metric; this is consistent with the finding in (Luxburg 2007), indicating that *SC* very often outperforms k-means for numerical data.

(2) *ROCK & CROCK*

ROCK, proposed by Guha et al. (Guha et al. 2000), is a robust clustering algorithm for categorical attributes. A link-based similarity measure between two data points is defined based on the neighborhood relation of the two data points, rather than distance or similarity with other data points.

During the process of choosing neighbors for each data object, Guha et al. simply considered the Jaccard coefficient (Ribeiro & Harder 2011) to capture the closeness between each pair of data objects, followed by the determination of neighbors with a user-defined threshold parameter. Their algorithm mainly focuses on the coupling relationship among objects, without any concern for the coupling relationships among attributes and their values. Therefore, we propose to replace the Jaccard coefficient with our proposed coupled nominal similarity *CASO* and to construct a coupled *ROCK*

(*CROCK*) algorithm by considering both coupled objects and coupled attribute values. Specifically, we regard two data objects u_x and u_y to be neighbors if $CASO(u_x, u_y)/n \geq \theta$, instead of $|u_x \bigcap u_y|/|u_x \bigcup u_y| \geq \theta$ presented in (Guha et al. 2000). The other procedures and functions remain the same as (Guha et al. 2000).

Below, we experiment with five real-life data sets, i.e. *Movie*, *Hayesroth*, *SPECT*, *Voting*, and *Mushroom*, to compare the cluster quality between *ROCK* and *CROCK* in terms of three measures: Precision (Pr), Recall (Re), and Specificity (Sp) (Figueiredo et al. 2011, Andritsos et al. 2004). As described in (Andritsos et al. 2004), the larger these indexes, the better the clustering. The number of runs for each experiment here is set to be 20 to obtain corresponding average results for the evaluation measures, due to the high computational complexity.

Table 5.12 shows the results of these two algorithms on the aforementioned quality measures for the five data sets. We choose parameters to obtain the best results, such as $\theta = 0.75$ for *Voting*. As this table indicates, the adapted *CROCK* with our proposed *CASO* outperforms the original *ROCK* on almost all the evaluation measures. Statistical testing also supports the results on Pr, Re and Sp, that *CROCK* performs better than *ROCK*, at a 95% significance level. Thus, *CROCK*'s quality is verified to be superior to that of *ROCK* due to the fact that the former considers both the couplings between attributes with their values (through co-occurrence) and between objects (by links).

Intra-attribute Value Clustering

In this part, we present the results of *CASO* applications to the problem of intra-attribute value clustering. We use the bibliographic data taken from the publicly-accessible bibliographic databases with 720 research papers (Gibson et al. 2000). Some 190 papers focus on database research, and the remaining 530 papers are written on theoretical computer science and related fields. For each paper, we record the name of the first author, the name of the second au-

Table 5.13: Clustering Qualities for First Author

Algorithm	Parameters	Pr	Sp	AC
STIRR	$\mathscr{M} = 9,\ \mathscr{N} = 10$	0.5112	0.236	0.5278
LIMBO	$\phi = 0.0,\ S = \infty$	0.6176	**0.4891**	0.5375
CLIMBO	$\phi = 0.0,\ S = \infty$	**0.8045**	0.4718	**0.7333**

thor, the name of the conference/journal, and the year of publication. We are interested in clustering the first authors, as well as the conferences/journals.

STIRR applies an iterative method based on a linear dynamic system to assign and propagate weights on the categorical values (Gibson et al. 2000) to conduct the intra-attribute value clustering, and *LIMBO* defines a distance between attribute values on the basis of the IB framework to quantify the degree of interchangeability of attribute values within a single attribute to group them (Andritsos et al. 2004). Accordingly, we substitute our proposed *CADO* in Equation (5.7.2) for the distance $\delta I(c_i, c_j)$ described by the information loss (i.e. *Jensen-Shannon* divergence) in (Andritsos et al. 2004), and then propose a coupled version of *LIMBO*, i.e. *CLIMBO*. *LIMBO* reveals that two attribute values are similar if the contexts in which they appear are similar. It is an alternative way to explicate the inter-coupled interactions among different attributes; however, it lacks the consideration of the intra-coupled interactions within each attribute. Thus, the metric *CADO* can be extended to measure the coupled distance between clusters by replacing an object u_i with a cluster c_i, and *CLIMBO* is naturally induced.

Two experiments below are conducted to compare these algorithms for the intra-attribute value clustering. The parameters are specified in the second column of Table 5.13. For *STIRR*, $\mathscr{M}$ and $\mathscr{N}$ are the numbers of initial configurations and iterations, respectively. For *LIMBO* and *CLIMBO*, ϕ indicates the size bound, S refers to the accuracy bound, and the addition operator is used. The experiments in this part only run 20 times to display the average results, since the algorithms itself is computational costly.

The first experiment is designed to cluster the first authors of the 720 academic papers, and the labels for evaluation are the pre-known research fields: database research (190 papers) and theoretical computer science (530 papers). Note also that all authors are identified by their last names so that, for instance, an attribute value "Wang" actually represents several Wangs taken together. In addition, the second author is regarded as being the same as the first author if the research paper has only one author. These two aspects lead to the overall modest clustering quality. *STIRR*, *LIMBO*, and *CLIMBO* are compared on the intra-attribute value clustering results of the attribute "first author" with regard to Precision (Pr), Specificity (Sp) (Figueiredo et al. 2011) and Accuracy (AC) (Andritsos et al. 2004). Table 5.13 shows that *CLIMBO* is the best in terms of Pr and AC, and is comparable with *LIMBO* with respect to Sp. All the results on Pr, Sp and AC are supported by a statistical significant test at a 95% significance level.

We now turn to the problem of clustering the conferences/journals. Figure 5.6 displays the clusters produced by *STIRR*, *LIMBO*, and *CLIMBO*. The x-axis represents the academic papers, while the y-axis denotes publishing venues. The thick horizontal line separates the clusters of conferences/journals, and the thick vertical line distinguishes between database research related papers (on the left) and theoretical computer science related papers (on the right). If an author has published a paper in a particular venue, this is represented by a point. From this figure, it is clear that *CLIMBO* yields the best partition, followed by *LIMBO*, and *STIRR* performs worst. However, even the clustering of *CLIMBO* is slightly mistaken by the conferences/journals between index 50 and 60, which is due to the influence of their co-authors.

The above two experiments therefore reveal that *CLIMBO* is better than *LIMBO* and *STIRR* on the clustering quality of intra-attribute values. Moreover, *LIMBO* can also be clearly observed to outperform *STIRR*, which is consistent with the conclusion drawn in (Andritsos et al. 2004).

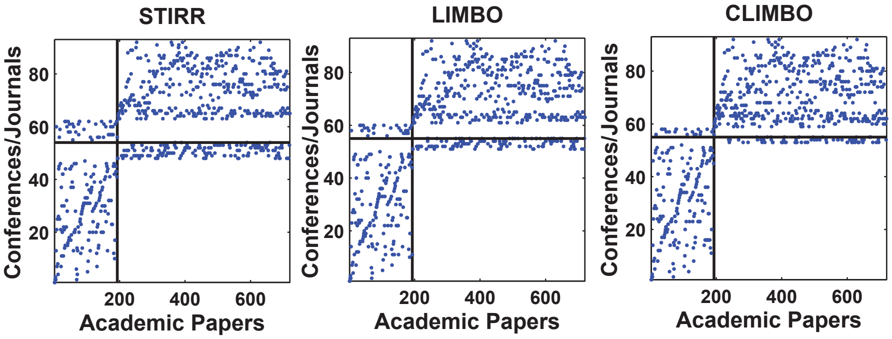


Figure 5.6: Clusterings for conferences/journals.

Classification Evaluation

To further verify the superiority of our proposed method *CADO* for classification, we use the k-nearest neighbor (*KNN*) algorithm (Figueiredo et al. 2011, Garcia, Derrac, Cano & Herrera 2012) to compare the classification accuracy using different dissimilarity measures. *KNN* is a type of instance-based learning, classifying objects based on the closest training examples in the attribute space. The standard of "closest" is characterized by the largest similarity or the smallest distance, and the distance or dissimilarity involved can be measured by *SMD*, *ADD* and *CADO* individually. Though this supervised learning process is chosen for the testing, dissimilarity metrics are all considered without the label information.

We carry out experiments on eight UCI data sets from different domains, where all the data sets are purely categorical. As we know, a better dissimilarity metric gives a better description of the similarity between data objects, and corresponds to a better classification result, i.e. higher classification accuracy. We run 100 tests; in each run, we randomly pick 90% of the data set as the training set, taking the rest as the test set. The results for *1NN* with *SMD*, *ADD*, and *CADO* are shown in Figure 5.7. According to Figure 5.7, we easily discover that *1NN* with *CADO* outperforms *1NN* with *SMD* for all the data, and the average accuracy induced by *1NN* with *CADO* is higher than that induced by *ADD* for most of the data, with the

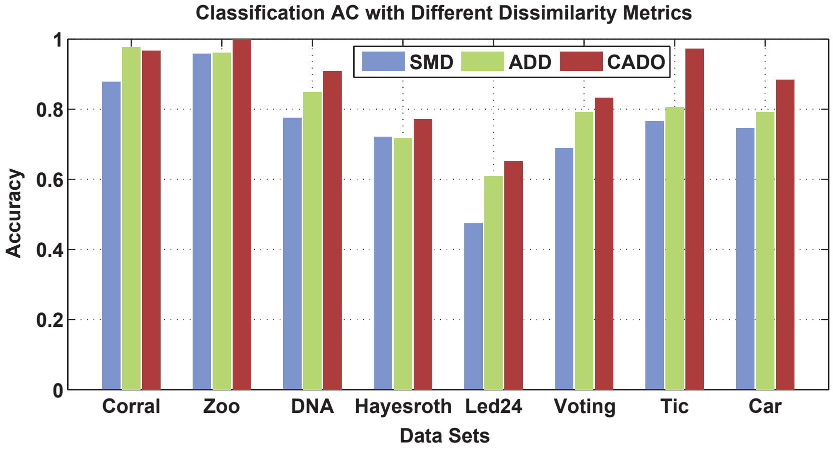


Figure 5.7: Comparisons on classification.

exception of only one data set: *Corral*. As can be seen, there is a remarkable improvement in the results with our proposed *CADO* compared to the other two metrics in terms of accuracy. All the results on accuracy are supported by a statistical significant test at a 95% significance level. Similar results can also be observed with *3NN* and *5NN*, which again suggest the effectiveness and superiority of our method.

In summary, we draw the following conclusions: 1) intra-coupled relative similarity measures *IRSP*, *IRSU*, *IRSJ* and *IRSI* all present the same learning accuracy, but *IRSI* is the most efficient, especially for large-scale data; 2) our proposed object dissimilarity metric *CADO* is better than others, i.e., the traditional *SMD*, frequency distribution only *OFD*, and dependency aggregation only *ADD*, for categorical data in terms of data structure analysis, clustering and classification quality; 3) the existing categorical clustering algorithms such as overlap-based methods (e.g. *k-modes*, *ROCK*), context-based methods (e.g. *STIRR*), and information-theoretic methods (e.g. *LIMBO*), and classification algorithm *KNN*, perform better than the original methods and algorithms when incorporated with *CASO* or *CADO*.

5.8 Summary

We have proposed *CASO*, a novel data-driven and non-IIDness based coupled attribute similarity measure for objects incorporating both intra-coupled attribute similarity for values and inter-coupled attribute similarity for values in unsupervised learning on nominal data. The measure involves both attribute value frequency distribution (intra-coupling) and attribute dependency aggregation (inter-coupling) and the interaction of the two, which captures a global picture of the similarity and has been shown to improve learning accuracy in diverse similarity measures. Theoretical analysis and substantial experiments have shown that the inter-coupled relative similarity measure *IRSI* significantly outperforms the other options (*IRSP*, *IRSU* and *IRSJ*) in terms of efficiency, in particular on a large-scale data set having a huge number of attribute values, while maintaining equal accuracy. Moreover, our derived dissimilarity metric is more general and accurate in capturing the internal structures of the predefined clusters and clustering quality in accordance with intensive empirical results. Very substantial experiments on accuracy and efficiency have been conducted on single attributes and on all attributes, as well as a scalability test on the number of attributes and the maximal number of attribute values, and on the clustering and classification performance by incorporating the proposed similarity. This has clearly shown that the proposed coupled nominal similarity leads to more accurate, efficient and scalable learning performance on large scale categorical data sets, supported by statistical analysis. The reason is that our proposed measure is global as a result of effectively integrating different aspects of the similarity.

5.9 Appendix: Proofs

Proof (a)

Theorem 5.9.1 (a) [Definition **5.4.1**] *Intra-coupled Attribute Similarity for Values (IaASV) between values* v_j^x *and* v_j^y *of attribute* a_j *is* $\delta_j^{Ia}(v_j^x, v_j^y)$*,* *we have* $\delta_j^{Ia} \in [1/3, m/(m+4)]$.

Proof 1 *According to Definition 5.4.1, we have that* $1 \leq |G_j(\{v_j^x\})|, |G_j(\{v_j^y\})| \leq m$ *holds, then*

$$\begin{aligned}\delta_j^{Ia}(v_j^x, v_j^y) &= \frac{|G_j(\{v_j^x\})| \cdot |G_j(\{v_j^y\})|}{|G_j(\{v_j^x\})| + |G_j(v_j^y)| + |G_j(\{v_j^x\})| \cdot |G_j(\{v_j^y\})|} \\ &= \frac{1}{|G_j(\{v_j^y\})|^{-1} + |G_j(\{v_j^x\})|^{-1} + 1} \\ &\leq \frac{1}{2\sqrt{|G_j(\{v_j^y\})|^{-1} \cdot |G_j(\{v_j^x\})|^{-1}} + 1}\end{aligned}$$

On one hand, $\delta_j^{Ia}(v_j^x, v_j^y)$ *is a monotonously increasing function of variables* $|G_j(\{v_j^x\})|$ *and* $|G_j(\{v_j^y\})|$*, respectively. Therefore,* $\delta_j^{Ia}(v_j^x, v_j^y)$ *takes its minimum value* $1/3$ *when* $|G_j(\{v_j^x\})| = |G_j(\{v_j^y\})| = 1$.

On the other hand, because of both $2 \leq |G_j(\{v_j^x\})| + |G_j(\{v_j^y\})| \leq m$ *and the above function property, then* $\delta_j^{Ia}(v_j^x, v_j^y)$ *takes its maximum value* $m/(m+4)$ *when* $|G_j(\{v_j^x\})| = |G_j(\{v_j^y\})| = m/2$.

Thus, considering both aspects above, we have

$$\delta_j^{Ia}(v_j^x, v_j^y) \in [1/3, m/(m+4)].$$

Proof (b)

Theorem 5.9.2 (b) [Definition **5.4.2**] *Equations (5.4.2) and (5.4.3) are equal to each other:* $D_{j|L}(v_j^x, v_j^y) = \sum_{l \in L} |P_{l|j}(\{l\}|v_j^x) - P_{l|j}(\{l\}|v_j^y)| = 2 \cdot \max_{L' \subseteq L} |P_{l|j}(L'|v_j^x) - P_{l|j}(L'|v_j^y)|$ *holds.*

[**Note**] This theorem is deduced from a property in probability theory (Gibbs & Su 2002), which is "The total variation distance between two probability measures $\mathbb{P}$ and $\mathbb{Q}$ on a sigma-algebra $\mathcal{F}$ of the subsets of the sample space Ω is defined via $\delta(\mathbb{P}, \mathbb{Q}) = \sup_{A \in \mathcal{F}} |\mathbb{P}(A) - \mathbb{Q}(A)|$. For a finite alphabet, we can write $\delta(\mathbb{P}, \mathbb{Q}) = \frac{1}{2} \sum_{x \in \Omega} |\mathbb{P}(x) - \mathbb{Q}(x)|$." If we regard $\mathbb{P} = P_{l|j}(\cdot|v_j^x))$ and $\mathbb{Q} = P_{l|j}(\cdot|v_j^y)$, $A = L'$ and $x = l$, then the above theorem holds.

Proof 2 *Assume* $L = \{l_1, l_2, \cdots, l_n\}$ *and* $L' = \{l_1, l_2, \cdots, l_k\}$ *(*$k \leq n$*), then*

$$\begin{aligned} F(L') &= 2 \cdot |P_{l|j}(L'|v_j^x) - P_{l|j}(L'|v_j^y)| \\ &= |2 \cdot \sum_{i=1}^{k} P_{l|j}(\{l_i\}|v_j^x) - 2 \cdot \sum_{i=1}^{k} P_{l|j}(\{l_i\}|v_j^y)|. \end{aligned}$$

Since $\sum_{i=1}^{n} P_{l|j}(l_i|v_j^x) = \sum_{i=1}^{n} P_{l|j}(l_i|v_j^y) = 1$ *holds, then:*

$$\begin{aligned} F(L') &= |[\sum_{i=1}^{k} P_{l|j}(\{l_i\}|v_j^x) + 1 - \sum_{i=k+1}^{n} P_{l|j}(\{l_i\}|v_j^x)] \\ &\quad - [\sum_{i=1}^{k} P_{l|j}(\{l_i\}|v_j^y) + 1 - \sum_{i=k+1}^{n} P_{l|j}(\{l_i\}|v_j^y)]| \\ &= |\sum_{i=1}^{k} P_{l|j}(\{l_i\}|v_j^x) - \sum_{i=1}^{k} P_{l|j}(\{l_i\}|v_j^y) \\ &\quad + \sum_{i=k+1}^{n} P_{l|j}(\{l_i\}|v_j^y) - \sum_{i=k+1}^{n} P_{l|j}(\{l_i\}|v_j^x)| \\ &= |\sum_{i=1}^{k} [P_{l|j}(\{l_i\}|v_j^x) - P_{l|j}(\{l_i\}|v_j^y)] \\ &\quad + \sum_{i=k+1}^{n} [P_{l|j}(\{l_i\}|v_j^y) - P_{l|j}(\{l_i\}|v_j^x)]| \\ \\ &\leq \sum_{i=1}^{k} |P_{l|j}(\{l_i\}|v_j^x) - P_{l|j}(\{l_i\}|v_j^y)| \\ &\quad + \sum_{i=k+1}^{n} |P_{l|j}(\{l_i\}|v_j^y) - P_{l|j}(\{l_i\}|v_j^x)| \\ &\leq \sum_{i \in 1}^{n} |P_{l|j}(\{l_i\}|v_j^x) - P_{l|j}(\{l_i\}|v_j^y)| \\ &= \sum_{l \in L} |P_{l|j}(\{l\}|v_j^x) - P_{l|j}(\{l\}|v_j^y)|) \end{aligned}$$

If there exists $k > 0$*, such that*

$$P_{l|j}(\{l_i\}|v_j^x) \geq P_{l|j}(\{l_i\}|v_j^y)$$

holds for $1 \leq i \leq k < n$ *and*

$$P_{l|j}(\{l_i\}|v_j^x) < P_{l|j}(\{l_i\}|v_j^y)$$

holds for $k+1 \leq i \leq n$*, then* $F(L')$ *takes its maximal value:*

$$\sum_{l \in L} |P_{l|j}(\{l\}|v_j^x) - P_{l|j}(\{l\}|v_j^y)|.$$

If for all $1 \leq i \leq k < n$*,*

$$P_{l|j}(\{l_i\}|v_j^x) < P_{l|j}(\{l_i\}|v_j^y)$$

holds, then we have

$$P_{l|j}(\{l_i\}|v_j^x) \geq P_{l|j}(\{l_i\}|v_j^y)$$

for $k+1 \leq i \leq n$*. Thus, we alternatively consider*

$$F(L'') = 2 \cdot |P_{l|j}(L''|v_j^y) - P_{l|j}(L''|v_j^x)|,$$

where $L'' = L - L'$*. In fact,*

$$\max_{L' \subseteq L} F(L') = \max_{L'' \subseteq L} F(L'')$$

holds. Similar to the above deduction,

$$\max_{L' \subseteq L} F(L') = \max_{L'' \subseteq L} F(L'') = \sum_{l \in L} |P_{l|j}(\{l\}|v_j^x) - P_{l|j}(\{l\}|v_j^y)|.$$

The rest special case is that for $1 \leq i \leq n$*,*

$$P_{l|j}(\{l_i\}|v_j^x) \geq P_{l|j}(\{l_i\}|v_j^y)$$

holds. This is in fact

$$P_{l|j}(\{l_i\}|v_j^x) = P_{l|j}(\{l_i\}|v_j^y)$$

for every possible i*, then* $F(L') = 0$ *takes the maximal value as well (i.e.* $\sum_{l \in L} |P_{l|j}(\{l\}|v_j^x) - P_{l|j}(\{l\}|v_j^y)|$*).*

Therefore, we have

$$\begin{aligned} D_{j|L}(v_j^x, v_j^y) &= \sum_{l \in L} |P_{l|j}(\{l\}|v_j^x) - P_{l|j}(\{l\}|v_j^y)| \\ &= 2 \cdot \max_{L' \subseteq L} |P_{l|j}(L'|v_j^x) - P_{l|j}(L'|v_j^y)|. \end{aligned}$$

Proof (c)

[Definition **5.4.2**] The conversion is conducted from equations (5.4.3) to (5.4.4) via (5.4.5): "$D_{j|L}(v_j^x, v_j^y) = 2 \cdot \max_{L' \subseteq L} |P_{l|j}(L'|v_j^x) - P_{l|j}(L'|v_j^y)|$" to "$\delta_{j|k}^P = \min_{V_k' \subseteq V_k}\{2 - P_{k|j}(V_k'|v_j^x) - P_{k|j}(\overline{V_k'}|v_j^y)\}$".

Proof 3 *The whole conversion procedural is divided into four steps.*

(1) The multiplier 2 *in* $D_{j|L}(v_j^x, v_j^y)$ *is omitted:*

$$D_{j|L}^{(1)}(v_j^x, v_j^y) = \max_{L' \subseteq L} |P_{l|j}(L'|v_j^x) - P_{l|j}(L'|v_j^y)|.$$

(2) Labels are replaced with other values of a particular attribute a_k*:*

$$D_{j|k}^{(2)}(v_j^x, v_j^y) = \max_{V_k' \subseteq V_k} |P_{k|j}(V_k'|v_j^x) - P_{k|j}(V_k'|v_j^y)|.$$

(3) A complementary set $\overline{V_k'}$ *rather than the original one* V_k' *is concerned for* v_j^y *in* ICP, *based on* $P_{k|j}(V_k'|v_j^y) = 1 - P_{k|j}(\overline{V_k'}|v_j^y)$*:*

$$D_{j|k}^{(3)}(v_j^x, v_j^y) = \max_{V_k' \subseteq V_k} |P_{k|j}(V_k'|v_j^x) + P_{k|j}(\overline{V_k'}|v_j^y) - 1|,$$

which is $D'_{j|k}(v_j^x, v_j^y)$ *formalized in equation (5.4.5).*

(4) Dissimilarity is considered rather than similarity, we use $\delta_{j|k}^P = 1 - D'_{j|k}(v_j^x, v_j^y)$ *for simplicity:*

$$\begin{aligned} D_{j|k}^{(4.1)}(v_j^x, v_j^y) &= 1 - D_{j|k}^{(3)}(v_j^x, v_j^y) \\ &= 1 - \max_{V_k' \subseteq V_k} |P_{k|j}(V_k'|v_j^x) + P_{k|j}(\overline{V_k'}|v_j^y) - 1|. \end{aligned}$$

If $P_{k|j}(V_k'|v_j^x) + P_{k|j}(\overline{V_k'}|v_j^y) - 1 \geq 0$*, then we have*

$$D_{j|k}^{(4.2)}(v_j^x, v_j^y) = \min_{V_k' \subseteq V_k} \{2 - P_{k|j}(V_k'|v_j^x) - P_{k|j}(\overline{V_k'}|v_j^y)\}$$

according to the fact that

$$1 - \max(|f(x)|) = \min(1 - f(x))$$

for all $f(x) \geq 0$ $(x \in \mathbb{R})$*, where* $f(x)$ *is a function and* $\mathbb{R}$ *is the real number field.*

If $P_{k|j}(V_k'|v_j^x)+P_{k|j}(\overline{V_k'}|v_j^y)-1<0$, we alternatively use $V_k''=V_k-V_k'=\overline{V_k'}$. Then we have

$$D_{j|k}^{(4.1')}(v_j^x,v_j^y) \;=\; 1-\max_{V_k''\subseteq V_k}|P_{k|j}(V_k''|v_j^x)+P_{k|j}(\overline{V_k''}|v_j^y)-1|$$

Since $P_{k|j}(V_k''|v_j^x)=1-P_{k|j}(V_k'|v_j^x)$ and $P_{k|j}(\overline{V_k''}|v_j^y)=P_{k|j}(V_k'|v_j^y)=1-P_{k|j}(\overline{V_k'}|v_j^y)$, we have

$$P_{k|j}(V_k''|v_j^x)+P_{k|j}(\overline{V_k''}|v_j^y)-1>0.$$

Hence, $D_{j|k}^{(4.2')}(v_j^x,v_j^y)=\min_{V_k''\subseteq V_k}\{2-P_{k|j}(V_k''|v_j^x)-P_{k|j}(\overline{V_k''}|v_j^y)\}$ according to the fact that $1-\max(|f(x)|)=\min(1+f(x))$ for all $f(x)\geq 0$ ($x\in\mathbb{R}$), where $f(x)$ is a function and $\mathbb{R}$ is the real number field.

In fact, we can see that

$$D_{j|k}^{(4.1)}(v_j^x,v_j^y)=D_{j|k}^{(4.1')}(v_j^x,v_j^y).$$

Therefore, we have obtained that

$$D_{j|k}^{(4.1)}(v_j^x,v_j^y) \;=\; D_{j|k}^{(4.1')}(v_j^x,v_j^y)=D_{j|k}^{(4.2)}(v_j^x,v_j^y)=D_{j|k}^{(4.2')}(v_j^x,v_j^y).$$

By following the above four steps, we have successfully converted from Equations (5.4.3) to (5.4.4) via (5.4.5): $D_{j|L}(v_j^x,v_j^y)$ to $D_{j|k}^{(4.2)}(v_j^x,v_j^y)$ or $D_{j|k}^{(4.2')}(v_j^x,v_j^y)$ via $D_{j|k}^{(3)}(v_j^x,v_j^y)$ or $D_{j|k}'(v_j^x,v_j^y)$.

Proof (d)

Theorem 5.9.3 (d) [Theorem **5.5.1**] *IRSP, IRSU, IRSJ and IRSI are all equivalent to one another.*

Proof 4 *Part (I)* IRSP$\Longleftrightarrow$IRSU

Let V_k^* be the value set of attribute a_k that makes

$$P_{k|j}(V_k'|v_j^x)+P_{k|j}(\overline{V_k'}|v_j^y)$$

maximal. Below, we show that for every $v_k \in V_k^*$,

$$P_{k|j}(\{v_k\}|v_j^x) \geq P_{k|j}(\{v_k\}|v_j^y)$$

holds. In fact, if there exists v_k^z $(\in V_k^*)$ *satisfying*

$$P_{k|j}(\{v_k^z\}|v_j^x) < P_{k|j}(\{v_k^z\}|v_j^y),$$

then set $V_k^{**} = V_k^* \backslash \{v_k^z\}$, $\overline{V_k^{**}} = \overline{V_k^*} \bigcup \{v_k^z\}$, *it directly follows that*

$$P_{k|j}(V_k^{**}|v_j^x) + P_{k|j}(\overline{V_k^{**}}|v_j^y) > P_{k|j}(V_k^*|v_j^x) + P_{k|j}(\overline{V_k^*}|v_j^y).$$

This results in the contradiction between V_k^{**} *and* V_k^* *because of the maximal assumption of* V_k^*.

Similarly, for any $v_k \in \overline{V_k^*}$, $P_{k|j}(\{v_k\}|v_j^x) \leq P_{k|j}(\{v_k\}|v_j^y)$ *holds. Hence,*

$$\begin{aligned}
\delta_{j|k}^P(v_j^x, v_j^y) &= \min_{V_k' \subseteq V_k} \{2 - P_{k|j}(V_k'|v_j^x) - P_{k|j}(\overline{V_k'}|v_j^y)\} \\
&= 2 - \max_{V_k' \subseteq V_k} \{P_{k|j}(V_k'|v_j^x) + P_{k|j}(\overline{V_k'}|v_j^y)\} \\
&= 2 - [P_{k|j}(V_k^*|v_j^x) + P_{k|j}(\overline{V_k^*}|v_j^y)] \\
&= 2 - [\sum_{v_k \in V_k^*} P_{k|j}(\{v_k\}|v_j^x) + \sum_{v_k \in \overline{V_k^*}} P_{k|j}(\{v_k\}|v_j^y)] \\
&= 2 - [\sum_{v_k \in V_k^*} \max\{P_{k|j}(\{v_k\}|v_j^x), P_{k|j}(\{v_k\}|v_j^y)\} \\
&\quad + \sum_{v_k \in \overline{V_k^*}} \max\{P_{k|j}(\{v_k\}|v_j^x), P_{k|j}(\{v_k\}|v_j^y)\}] \\
&= 2 - \sum_{v_k \in V_k} \max\{P_{k|j}(\{v_k\}|v_j^x), P_{k|j}(\{v_k\}|v_j^y)\} \\
&= \delta_{j|k}^U(v_j^x, v_j^y)
\end{aligned}$$

Part (II) IRSU$\Longleftrightarrow$IRSJ

Note that in the following Part (II) and Part (III), $v_k \in v_j^x \backslash v_j^y$ *and* $v_k \in v_j^y \backslash v_j^x$ *are the abbreviated forms for* $v_k \in \varphi_{j\to k}(v_j^x) \backslash \varphi_{j\to k}(v_j^y)$ *and* $v_k \in \varphi_{j\to k}(v_j^y) \backslash \varphi_{j\to k}(v_j^x)$, *respectively.*

Given $v_k \notin \varphi_{j\to k}(v_j^x) \bigcup \varphi_{j\to k}(v_j^y)$, *that is*

$$v_k \notin \varphi_{j\to k}(v_j^x) \text{ and } v_k \notin \varphi_{j\to k}(v_j^y).$$

If $v_k \notin \varphi_{j\to k}(v_j^x)$, *we then have*

$$g_k^*(\{v_k\}) \bigcap g_j(v_j^x) = \varnothing,$$

so $P_{k|j}(\{v_k\}|v_j^x) = 0$. *Similarly,* $P_{k|j}(\{v_k\}|v_j^y) = 0$. *Therefore,*

$$\begin{aligned}
\delta_{j|k}^U(v_j^x, v_j^y) &= 2 - \sum_{v_k \in V_k} \max\{P_{k|j}(\{v_k\}|v_j^x), P_{k|j}(\{v_k\}|v_j^y)\}, \\
&= 2 - [\sum_{v_k \in \bigcup} \max\{P_{k|j}(\{v_k\}|v_j^x), P_{k|j}(\{v_k\}|v_j^y)\} \\
&\quad + \sum_{v_k \notin \bigcup} \max\{P_{k|j}(\{v_k\}|v_j^x), P_{k|j}(\{v_k\}|v_j^y)\}] \\
&= 2 - \sum_{v_k \in \bigcup} \max\{P_{k|j}(\{v_k\}|v_j^x), P_{k|j}(\{v_k\}|v_j^y)\} \\
&= \delta_{j|k}^J(v_j^x, v_j^y)
\end{aligned}$$

Part (III) IRSJ$\Longleftrightarrow$IRSI

If $v_k \in \varphi_{j\to k}(v_j^x) \backslash \varphi_{j\to k}(v_j^y)$, *then* $P_{k|j}(\{v_k\}|v_j^y) = 0$. *Accordingly, we have*

$$\max\{P_{k|j}(\{v_k\}|v_j^x), P_{k|j}(\{v_k\}|v_j^y)\} = P_{k|j}(\{v_k\}|v_j^x).$$

Similarly, if $v_k \in \varphi_{j\to k}(v_j^y) \backslash \varphi_{j\to k}(v_j^x)$, *it indicates*

$$\max\{P_{k|j}(\{v_k\}|v_j^x), P_{k|j}(\{v_k\}|v_j^y)\} = P_{k|j}(\{v_k\}|v_j^y).$$

Therefore, we have

$$
\begin{aligned}
\delta^J_{j|k}(v^x_j, v^y_j) &= 2 - \sum_{v_k \in \bigcup} \max\{P_{k|j}(\{v_k\}|v^x_j), P_{k|j}(\{v_k\}|v^y_j)\} \\
&= 2 - [\sum_{v_k \in v^x_j \backslash v^y_j} \max\{P_{k|j}(\{v_k\}|v^x_j), P_{k|j}(\{v_k\}|v^y_j)\} \\
&\quad + \sum_{v_k \in v^y_j \backslash v^x_j} \max\{P_{k|j}(\{v_k\}|v^x_j), P_{k|j}(\{v_k\}|v^y_j)\} \\
&\quad + \sum_{v_k \in \bigcap} \max\{P_{k|j}(\{v_k\}|v^x_j), P_{k|j}(\{v_k\}|v^y_j)\}] \\
&= 2 - [1 - \sum_{v_k \in \bigcap} P_{k|j}(\{v_k\}|v^x_j) + 1 - \sum_{v_k \in \bigcap} P_{k|j}(\{v_k\}|v^y_j) \\
&\quad + \sum_{v_k \in \bigcap} \max\{P_{k|j}(\{v_k\}|v^x_j), P_{k|j}(\{v_k\}|v^y_j)\}] \\
&= \sum_{v_k \in \bigcap} [P_{k|j}(\{v_k\}|v^x_j) + P_{k|j}(\{v_k\}|v^y_j)] \\
&\quad - \sum_{v_k \in \bigcap} \max\{P_{k|j}(\{v_k\}|v^x_j), P_{k|j}(\{v_k\}|v^y_j)\} \\
&= \sum_{v_k \in \bigcap} \min\{P_{k|j}(\{v_k\}|v^x_j), P_{k|j}(\{v_k\}|v^y_j)\} \\
&= \delta^I_{j|k}(v^x_j, v^y_j)
\end{aligned}
$$

Thus, IRSP, IRSU, IRSJ, *and* IRSI *are all equivalent to one another.*

Proof (e)

Theorem 5.9.4 (e) [Experiment **5.7.1**] *For any attribute a_j, the proportion $\xi_j(P/U) \in [\frac{2^T}{T}, \frac{2^R}{R}]$. For all attributes, the proportion $\xi(P/U) \in [\frac{2^T}{T}, \frac{2^R}{R}]$.*

Proof 5 *According to Definitions 5.4.2 and 5.4.3, and Table 5.9, we know*

$$\xi_{j|k}(P/U) = \frac{|ICP^{(P)}_{a_{j|k}}|}{|ICP^{(U)}_{a_{j|k}}|} = \frac{2^{|V_k|}}{|V_k|},$$

where $|ICP^{(P)}_{a_{j|k}}|$ and $|ICP^{(U)}_{a_{j|k}}|$ represent the time costs of ICP *for $\delta^P_{j|k}(v^x_j, v^y_j)$ and $\delta^U_{j|k}(v^x_j, v^y_j)$, respectively. Since $T = \min^n_{k=1}|V_j|$ and $R = \max^n_{k=1}|V_j|$, then $T \leq |V_k| \leq R$ for any set of attribute values V_k. We know $|V_k|$ is a positive integer, so based on Lemma 5.9.1 below, the statement*

$$\xi_{j|k}(P/U) \in [\frac{2^T}{T}, \frac{2^R}{R}]$$

holds. In addition, we have

$$\begin{aligned} \xi_j(P/U) &= \frac{|ICP^{(P)}_j|}{|ICP^{(U)}_j|} = \frac{\sum_{k\neq j}|ICP^{(P)}_{a_{j|k}}|}{\sum_{k\neq j}|ICP^{(U)}_{a_{j|k}}|}, \\ \xi(P/U) &= \frac{|ICP^{(P)}|}{|ICP^{(U)}|} = \frac{\sum_{1\leq j\leq n}|ICP^{(P)}_j|}{\sum_{1\leq j\leq n}|ICP^{(U)}_j|}. \end{aligned}$$

Based on Lemma 5.9.2 below, we then obtain that

$$\xi_j(P/U) \in \left[\frac{2^T}{T}, \frac{2^R}{R}\right] \text{ and } \xi(P/U) \in \left[\frac{2^T}{T}, \frac{2^R}{R}\right].$$

Lemma 5.9.1 *If x is a positive integer, then function $q(x) = 2^x/x$ is a monotonically increasing function.*

Proof 6 *To verify the monotonically increasing property of function $q(x) = 2^x/x$, we only need to look at the derivative of $q(x)$ since $q(x)$ is a continuous function of x, that is*

$$q'(x) = \frac{2^x \cdot \ln 2 \cdot x - 2^x}{x^2} = \frac{2^x \cdot (\ln 2 \cdot x - 1)}{x^2}.$$

If $q'(x) > 0$, then we can guarantee that $q(x)$ is a strictly monotonically increasing function. Here, $q'(x) > 0$ is equivalent to $x > 1/\ln 2$. As x is a positive integer, then $q(x) = 2^x/x$ is a strictly monotonically increasing function when $x \geq 2 > 1/\ln 2$. We also have $q(1) = 2$ when $x = 1$, and $q(2) = 2$ when $x = 2$, so $q(1) \leq q(2)$. Thus, $q(x) = 2^x/x$ is a monotonically increasing function when x is a positive integer.

Lemma 5.9.2 *If $x_1, \cdots, x_n$ are positive integers, where $T = \min_{1\leq i\leq n} x_i$ and $R = \max_{1\leq i\leq n} x_i$, then*

$$\frac{2^T}{T} \leq \frac{2^{x_1} + 2^{x_2} + \cdots + 2^{x_n}}{x_1 + x_2 + \cdots + x_n} \leq \frac{2^R}{R}.$$

Proof 7 *Without loss of generality, we assume $1 \leq x_1 \leq x_2 \leq \cdots \leq x_n$, then $T = x_1$, $R = x_n$, and the question is to prove*

$$\frac{2^{x_1}}{x_1} \leq \frac{2^{x_1} + 2^{x_2} + \cdots + 2^{x_n}}{x_1 + x_2 + \cdots + x_n} \leq \frac{2^{x_n}}{x_n}$$

According to Lemma 5.9.1, we have for $i = 1, \cdots, n$,

$$\frac{2^{x_1}}{x_1} \leq \frac{2^{x_i}}{x_i} \Longleftrightarrow 2^{x_1} \cdot x_i \leq 2^{x_i} \cdot x_1.$$

Then, we can naturally obtain that

$$\sum_{i=1}^{n} (2^{x_1} \cdot x_i) \leq \sum_{i=1}^{n} (2^{x_i} \cdot x_1),$$

which is equivalent to

$$2^{x_1} \cdot (x_1 + x_2 + \cdots + x_n) \leq x_1 \cdot (2^{x_1} + 2^{x_2} + \cdots + 2^{x_n}).$$

Therefore, we have

$$\frac{2^{x_1}}{x_1} \leq \frac{2^{x_1} + 2^{x_2} + \cdots + 2^{x_n}}{x_1 + x_2 + \cdots + x_n}.$$

Similarly, we can prove that

$$\frac{2^{x_1} + 2^{x_2} + \cdots + 2^{x_n}}{x_1 + x_2 + \cdots + x_n} \leq \frac{2^{x_n}}{x_n}$$

Thus, the inequality

$$\frac{2^T}{T} \leq \frac{2^{x_1} + 2^{x_2} + \cdots + 2^{x_n}}{x_1 + x_2 + \cdots + x_n} \leq \frac{2^R}{R}$$

holds for $T = \min_{1\leq i\leq n} x_i$ and $R = \max_{1\leq i\leq n} x_i$.

Proof (f)

Theorem 5.9.5 (f) [Experiment **5.7.1**] *Multi-variant function $\xi(U/P)$ is a monotonously decreasing function on the maximal number of values R.*

Proof 8 *The multi-variant function $\xi(U/P)$ is*

$$\begin{aligned}\xi(U/P) &= \frac{|ICP^{(U)}|}{|ICP^{(P)}|} = \frac{\sum_{1\le j\le n}|ICP_j^{(U)}|}{\sum_{1\le j\le n}|ICP_j^{(P)}|} \\ &= \frac{\sum_{1\le j\le n}\sum_{k\neq j}|ICP_{a_{j|k}}^{(U)}|}{\sum_{1\le j\le n}\sum_{k\neq j}|ICP_{a_{j|k}}^{(P)}|}.\end{aligned}$$

Since $|ICP_{a_{j|k}}^{(U)}| = |V_k|$ and $|ICP_{a_{j|k}}^{(P)}| = 2^{|V_k|}$, then the question is to prove that the function

$$F(x_1, x_2, \cdots, x_n) = \frac{x_1 + x_2 + \cdots + x_n}{2^{x_1} + 2^{x_2} + \cdots + 2^{x_n}}$$

is a monotonously decreasing function on x_n, if we assume integers $1 \le x_1 \le x_2 \le \cdots \le x_n$. To verify this, we only need to look at the partial derivative of F on x_n since F is a continuous function of x_n. If $\partial F/\partial x_n \le 0$, then we can say that F is a monotonously decreasing function on x_n. Suppose $M = \sum_{i=1}^{n-1} x_i$ and $N = \sum_{i=1}^{n-1} 2^{x_i}$, then we have

$$\begin{aligned}\frac{\partial F}{\partial x_n} &= \frac{N + 2^{x_n} - (M + x_n)\cdot 2^{x_n}\cdot \ln 2}{(N + 2^{x_n})^2} \\ &= \frac{(N - M\cdot 2^{x_n}\cdot \ln 2) + 2^{x_n}\cdot(1 - x_n\cdot \ln 2)}{(N + 2^{x_n})^2}.\end{aligned}$$

To judge whether $\partial F/\partial x_n \le 0$, we discuss the following four cases:

1) $2 \le x_{n-1} \le x_n$: In this case, according to Lemma 5.9.2, we have

$$\begin{aligned}\frac{N}{M} \le \frac{2^{x_{n-1}}}{x_{n-1}} \le 2^{x_n}\cdot \ln 2 \quad &\Longleftrightarrow \quad N \le M\cdot 2^{x_n}\cdot \ln 2, \\ \frac{1}{ln2} < 2 \le x_n \quad &\Longleftrightarrow \quad 1 < x_n\cdot \ln 2.\end{aligned}$$

Thus, $\partial F/\partial x_n < 0$ holds.

2) $x_{n-1} = 1$ and $2 \leq x_n$: In this case, according to Lemma 5.9.2, we have

$$\frac{N}{M} \leq \frac{2^1}{1} < 2^2 \cdot \ln 2 \leq 2^{x_n} \cdot \ln 2 \iff N < M \cdot 2^{x_n} \cdot \ln 2,$$
$$\frac{1}{ln2} < 2 \leq x_n \iff 1 < x_n \cdot \ln 2.$$

Thus, $\partial F / \partial x_n < 0$ holds.

3) $x_{n-1} = x_n = 1$: In this case, we have

$$F(x_1, \cdots, x_{n-1}, x_n) = F(1, \cdots, 1, 1) = \frac{n}{2n} = \frac{1}{2}.$$

For $x_{n-1} = 1, x_n = 2$, then

$$F(x_1, \cdots, x_{n-1}, x_n) = F(1, \cdots, 1, 2) = \frac{n-1+2}{2(n-1)+4} = \frac{1}{2}.$$

Thus, $F(x_1, \cdots, x_{n-1}, 2) \leq F(x_1, \cdots, x_{n-1}, 1)$.

4) $2 \leq x_{n-1}$ and $x_n = 1$: This case is impossible since we assume that $x_{n-1} \leq x_n$.

Therefore, we discover that for both $x_{n-1} = 1$ and $2 \leq x_{n-1}$, F is a monotonously decreasing function on x_n. That is to say, multi-variant function $\xi(U/P)$ is a monotonously decreasing function on the maximal number of values R.

[***Note***] *A conference version of this chapter has been published in the first item below, and a full journal version has been submitted to the second item.*

- ***Can Wang**, Longbing Cao, Mingchun Wang, Jinjiu Li, Wei Wei, Yuming Ou (2011), "Coupled Nominal Similarity in Unsupervised Learning". The 20th ACM Conference on Information and Knowledge Management (**CIKM 2011**), pp. 973-978.*

- ***Can Wang**, Longbing Cao (2013), "Coupled Attribute Similarity Analysis on Categorical Data". IEEE Transactions on Neural Networks and Learning Systems (**TNNLS**).*

In addition, the following two published conference papers are based on the work introduced in this chapter, they have used the proposed coupled similarity measures in different applications, such as recommendation system and document clustering.

- *Yonghong Yu, **Can Wang**, Yang Gao, Longbing Cao, Xixi Chen (2013), "A Coupled Clustering Approach for Items Recommendation". The 17th Pacific-Asia Conference on Knowledge Discovery and Data Mining (**PAKDD 2013**), pp. 365-376.*

- *Xin Cheng, Duoqian Miao, **Can Wang**, Longbing Cao (2013). "Coupled Term-Term Relation Analysis for Document Clustering". The 2013 International Joint Conference on Neural Networks (**IJCNN 2013**), full paper accepted.*

Chapter 6

Coupled Behavior Ensemble Learning

In this chapter, we analyze and learn the quantitative coupled behaviors by exploring the coupling relationships between properties and between entities. Here, behavior ensemble refers to the optimization learning tasks (e.g. clustering analysis) based on a collection of behavior learning models (e.g. base clusterings), and coupled behavior ensemble further investigates the couplings/interactions among behavior ensembles. This chapter mainly focuses on the task of clustering ensemble learning. Throughout the chapter, data object and the entity of coupled behaviors are interchangeable; base clusterings, behavior learning methods/models and discrete attributes all correspond to the categorical properties of coupled behaviors; and accordingly attribute values (i.e. the related labels in each base clustering) refer to the property values of coupled behaviors. Therefore, the coupling relationship between base clusterings (i.e. properties) embodies the interactions between different learning strategies as well. In addition, this work is based on the coupled categorical behavior analysis discussed in Chapter 5 since they both deal with the categorical properties of coupled behaviors.

The behavior ensemble learning is a powerful approach for improving the accuracy and stability of individual (base) behavior learning methods. Most of the existing behavior clustering ensembles obtain final solutions based on the IIDness (i.e. independent and identical distribution) assumption which assumes that base clusterings (i.e. behavior learning methods) perform independently of one another and all entities are also independent. However, in real-world data sources, entities are more or less associated in terms of certain coupling relationships. Several base clusterings trained on the source data are complementary to one another since each of them may only capture certain specific aspects rather than the full picture of the behavior data. This forms the non-IIDness nature, which is embodied in strong couplings and heterogeneity between objects and between methods/properties.

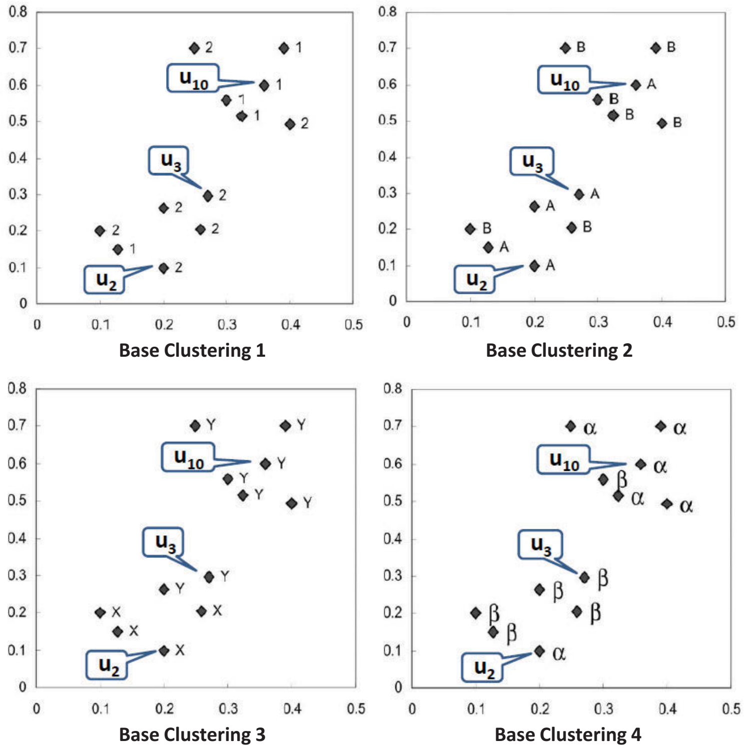
This chapter explicates the non-IIDness or coupling relationships between behavior learning models and between entities in clustering ensembles, and propose a corresponding framework for coupled clustering ensembles (*CCE*). *CCE* not only considers but also integrates the coupling relationships between base clusterings and between objects. Specifically, we examine both the intra-coupling within one base clustering (i.e. cluster label frequency distribution) and the inter-coupling between different base clusterings (i.e. cluster label co-occurrence dependency). Furthermore, we engage both the intra-coupling between two entities by aggregating the interactions of base clusterings and the inter-coupling among other objects by exploring their neighborhood domains. This is the first work to explicitly address the non-IIDness issue in clustering ensembles, verified by the application of such couplings in three types of consensus function: clustering-based, object-based and cluster-based. Substantial experiments on two synthetic and nine UCI data sets demonstrate that the *CCE* framework can effectively capture the interactions embedded in behavior learning methods and entities with higher clustering accuracy, stability, and robustness compared to eleven state-of-the-art techniques, supported by statistical analysis. In addition, we verify that the final learning quality is dependent on the data characteristics (i.e. the quality and consistency) of the initial behavior learning methods.

6.1 Background and Overview

Clustering analysis is a fundamental tools for capturing the structure of a data set. A large number of clustering algorithms (Jain, Murty & Flynn 1999, Chen, Zhang, Liu, Poon & Wang 2012) have been proposed, but the No Free Lunch theorem (Wolpert & Macready 1996) suggests that there is no single, supreme algorithm that fits all cluster shapes and structures perfectly. It is extremely difficult for users to decide which algorithm will be the most effective for a given data set. Consequently, as a recent offshoot of classifier ensemble research (Kittler, Hatef, Duin & Matas 1998, Bi, Guan & Bell 2008, García-Osorio, de Haro-García & García-Pedrajas 2010), the clustering ensemble (Gao, Fan & Han 2010, Vega-Pons & Ruiz-Shulcloper 2011) has exhibited great potential for enhancing clustering accuracy, robustness and parallelism (Strehl & Ghosh 2002) by combining results from various clustering methods. In general, the whole process of the clustering ensemble can be divided into three parts: building base clusterings, aggregating base clusterings, and post-processing clustering. The objective is to produce an overall high-quality clustering that agrees as much as possible with each of the input clusterings. The essence of the clustering ensemble is to aggregate the advantages of each base clustering to give a more complete, global understanding of the underlying data, while each base clustering is assumed to capture the best local and partial picture of a data set.

The clustering ensemble can be applied in various settings (Gionis et al. 2007, Luo & Jennings 2007), such as clustering heterogeneous data, detecting outliers, improving clustering robustness, decision making, and privacy-preserving information. While the clustering ensemble largely captures the common structure of the base clusterings, and achieves a combined clustering with better quality than that of individual clusterings, it also faces several issues that have not been explored well in the consensus design. We illustrate the problem with the related work and also discuss the challenge of the clustering ensemble below.

The clustering ensemble described in Figure 6.1 adapted from Figure



Base Clustering 1

Base Clustering 2

Base Clustering 3

Base Clustering 4

Figure 6.1: Four possible base clusterings of 12 data objects into two clusters; different partitions use different sets of labels.

1.2 in Chapter 1 is an example of four two-cluster partitions of 12 two-dimensional data objects. The target of the clustering ensemble is to obtain a final clustering based on these four base clusterings. As shown in Figure 6.1, the four possible cluster labels for the objects u_2, u_3 and u_{10} are $\{2, A, X, \alpha\}$, $\{2, A, Y, \beta\}$ and $\{1, A, Y, \alpha\}$, respectively. Two of the four base clusterings put each pair of objects in the same group, and the remainder of the two partitions assign different cluster labels to this pair. For instance, the first and second base clusterings distribute u_2 and u_3 in the same cluster, while the last two base clusterings give them distinct labels. In this situation, the traditional clustering ensemble method (i.e., *CSPA* (Strehl & Ghosh 2002)) treats the similarity between each pair of these three objects as 0.5, which is $Sim(u_2, u_3) = Sim(u_2, u_{10}) = Sim(u_3, u_{10}) = 0.5$. In the last stage of post-processing clustering, it is thereby difficult to determine the final label for these objects. This is because the consensus building assumes that all four base clusterings are independent, and that each base clustering also treats all the objects independently based on the "IIDness" (i.e. independent and identical distribution) assumption. A conventional way to solve this dilemma is to randomly distribute them in either an identical cluster or different groups, which will inevitably affect the clustering performance.

If we carefully explore the information provided in Figure 6.1, however, coupling relationships between the base clusterings and between the data objects can be identified, apart from the consensus among initial results proposed by traditional ensemble strategies. Below, we illustrate this observation via the example in Figure 6.1.

On one hand, as indicated in Figure 6.2, objects u_2 and u_3 are considered to have a high similarity value (e.g. 1) in base clusterings 1 and 2, in which they are assigned to the same clusters (i.e. clusters 2 and A, respectively). In contrast, their similarity value is rather low (e.g. 0) if the information in base clustering 4 is all that is used, since they are grouped into different clusters: α and β. However, clusters α and β are intuitively more similar than they appear to be, due to the fact that they connect with two identical clusters in the other base clusterings 1 and 2 via objects u_2 and u_3. Thus, the similarity (i.e. the dashed line) between clusters α and β in relation to other base clusterings should be larger than 0. The same principle also applies to the similarity between clusters X and Y in base clustering 3. Note that here the similarity between the identical clusters (e.g., clusters A or 2) is manually set as 1. In this way, the overall similarity between objects u_2 and u_3 must be larger than 0.5 as the traditional method maintains. Accordingly, objects u_2 and u_3 are more likely to be assigned to the correct identical cluster, rather than depending on the random allocation used in conventional methods (Strehl & Ghosh 2002, Gionis et al. 2007).

On the other hand, the similarity between objects u_2 and u_3 and the similarity between u_2 and u_{10} are identical (i.e., both 0.5). Thus, how do we

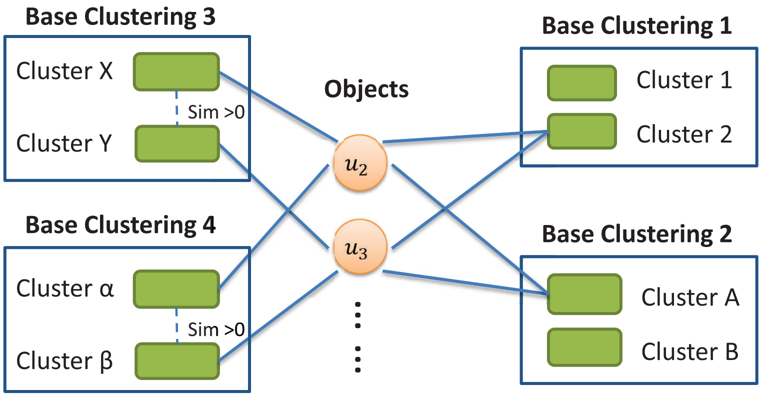


Figure 6.2: A graphical representation of the coupling relationship between base clusterings, where a circle denotes an object, a rounded rectangle represents an cluster, and an edge exists if an object belongs to a cluster.

distinguish between them and assign the correct label to each object? If we only consider the aforementioned coupling relationship between base clusterings, we may fail since both similarities can be enhanced by involving the co-occurrence with clusters in other base clusterings. However, we discover that the discrepancy on the common neighborhood domains of objects u_2 with u_3 and u_2 with u_{10} is capable of differentiating u_2 and u_{10} in distinct clusters. Intuitively, we recognize that, in Figure 6.1, the number of common neighbors of objects u_2 and u_3 is much larger than that of u_2 and u_{10}. From this perspective, it is more probable that objects u_2 and u_3 are in the same cluster and that object u_{10} is in another cluster, which corresponds to the genuine partition.

The above examples and analysis disclose that the IIDness assumption about base clusterings and objects actually causes the aforementioned problems. The corresponding problem-solving solutions essentially raise the following three important research questions.

1. *Clustering Coupling*: There is a likely structural relationship between base clusterings since they are induced from the same data set. How do we describe the coupling relationship between base clusterings?

2. *Object Coupling*: There is a context surrounding two objects which makes them dependent on each other. How do we design the similarity or distance between objects to capture their relation with other data objects?

3. *Integrated Coupling*: If there are interactions between both clusterings and objects, how do we integrate such couplings in the clustering ensemble?

The above questions suggest a very different assumption for clustering ensembles: non-IIDness (Cao 2013) based clustering ensemble. Non-IIDness captures strong couplings and heterogeneity between values, attributes, objects and methods in analyzing and learning complex problems. For the aforementioned toy example, intuitively, the base clusterings are expected to have some interactions with one another, such as the co-occurrence of their cluster labels over the same set of objects, which reflects the non-IIDness assumption of base clusterings. Here, the cluster label refers to the label of a cluster to which an object belongs, i.e. α, β in Figure 6.1, but most of the existing methods such as *CSPA* (Strehl & Ghosh 2002) and *QMI* (Topchy et al. 2005) are based on the hypothesis that base clusterings are independent of one another. Furthermore, the similarity between any two objects within the same cluster may differ and should be distinguished. In existing work, however, the predominant approach is to treat the similarity between objects as roughly 1 if they belong to the same cluster, otherwise 0. Such a binary measure is inadequate in terms of capturing the relationships between objects. In addition, some controversial objects with approximately equal similarity are observed to have different sizes of common neighborhood domains to differentiate them, which alternatively reveals the non-IIDness nature of objects. This issue has not been addressed in current approaches to the clustering ensemble problem, which instead merely consider the similarity between a pair of objects irrespective of other objects.

Recently, a link-based approach (Iam-On et al. 2011) has been proposed to consider cluster-cluster similarity by the connected-triple approach, the progress of which is promising, but it overlooks the interaction between objects. A clustering algorithm *ROCK* for categorical data (Guha et al. 2000) specifies the interaction between objects, but it is only designed for categorical clustering and lacks any consideration of the relationship between base clusterings. A detailed review of the related work on behavior ensemble learning can be found in Section 2.4. For *integrated coupling*, no work has been reported that systematically takes into account the couplings between base clusterings or between data objects.

In the real world, business and social applications such as investors in capital markets and members in social networking almost always see objects coupled with each other (Cao et al. 2012, Wang & Cao 2012, Song, Cao, Wu, Wei, Ye & Ding 2012). There is a great demand from both practical and theoretical perspectives to initiate new mechanisms to explicitly address the non-IIDness both between base clusterings and between objects, and to explicate how to incorporate the non-IIDness for the clustering ensemble based on consensus functions.

In this chapter, we propose an effective framework for coupled clustering ensembles (*CCE*) to address the aforementioned research questions. We consider both the couplings between base clusterings and between data objects, and propose a coupled framework of clustering ensembles to form an integrated coupling (i.e. non-IIDness). We then explicate our proposed framework *CCE* from the perspectives of clustering-based, object-based, and cluster-based algorithms, and reveals that the couplings are essential to the clustering ensemble. During the whole process, we propose several similarity measures that incorporate the couplings of base clusterings and objects, and they exhibit an impressive ability to capture the implicit relationships within the data. In addition, we evaluate our proposed framework *CCE* with the existing eight clustering ensemble methods and two categorical clustering algorithms on a variety of benchmark data sets in terms of accuracy, stability, robustness, and statistical significance. Finally, we empirically explore the relationship between the data characteristics of base clusterings and the degree of improvement in the final clustering quality.

The chapter is organized as follows. Preliminary definitions are specified in Section 6.2. Section 6.3 proposes the coupled framework *CCE*. Coupling relationships between base clusterings and between objects in *CCE* are specified in Section 6.4. Section 6.5 presents the coupled consensus functions for *CCE* together with miscellaneous issues. We describe the *CCE* algorithms in Section 6.6. The effectiveness of *CCE* is shown in Section 6.7 with intensive experiments. We conclude this work and address future work in Section 6.8.

6.2 Preliminary Definitions

The problem of the clustering ensemble can be formally described as follows: $U = \{u_1, \cdots, u_m\}$ is a set of m objects for clustering; $C = \{bc_1, \cdots, bc_L\}$ is a set of L base clusterings, each clustering bc_j consists of a set of clusters $bc_j = \{c_j^1, \cdots, c_j^{t_j}\}$ where t_j is the number of clusters in base clustering bc_j $(1 \leq j \leq L)$. Our goal is to find a final desirable clustering $fc^* = \{c_*^1, \cdots, c_*^{t^*}\}$ with t^* clusters such that the objects inside each cluster c_*^t are close to one another and the objects in different clusters are far from one another.

We construct an information table S by mapping each base clustering as an attribute. Here, v_j^x indicates the label of a cluster to which the object u_x belongs in the jth base clustering, and V_j is the set of cluster labels in base clustering bc_j. For example, Table 6.1 (Topchy et al. 2005) is the representation of Figure 6.1 as an information table consisting of twelve objects $\{u_1, u_2, \cdots, u_{12}\}$ and four corresponding attributes (i.e. base clusterings $\{bc_1, bc_2, bc_3, bc_4\}$). The cluster label α in base clustering bc_4 is mapped as the attribute value v_4^2 of object u_2 on attribute bc_4, and cluster label set $V_4 = \{\alpha, \beta\}$.

Based on this information-table representation, we use several concepts adapted from our previous work (Wang et al. 2011), which has been specified in Chapter 5. The "set information function" $g_j(v_j^x)$ specifies the set of objects whose cluster labels is v_j^x in base clustering bc_j. For example, we have

Table 6.1: An Example of Base Clusterings

C U	bc_1	bc_2	bc_3	bc_4
u_1	2	B	X	β
u_2	2	A	X	α
u_3	2	A	Y	β
u_4	2	B	X	β
u_5	1	A	X	β
u_6	2	A	Y	β
u_7	2	B	Y	α
u_8	1	B	Y	α
u_9	1	B	Y	β
u_{10}	1	A	Y	α
u_{11}	2	B	Y	α
u_{12}	1	B	Y	α

$g_4(v_4^2) = g_4(\alpha) = \{u_2, u_7, u_8, u_{10}, u_{11}, u_{12}\}$. We adopt the "inter-information function" $\varphi_{j\to k}(v_j^x)$ to obtain a subset of cluster labels in base clustering bc_k for the corresponding objects, which are derived from the cluster label v_j^x in base clustering bc_j, e.g., $\varphi_{4\to 2}(\alpha) = \{A, B\}$ derived from object set $g_4(\alpha)$. Added to this, the "information conditional probability" $P_{k|j}(v_k|v_j^x)$ characterizes the percentage of objects whose cluster labels in base clustering bc_k is v_k among those objects whose cluster label in base clustering bc_j is exactly v_j^x, formalized as:

$$P_{k|j}(v_k|v_j^x) = \frac{|g_k(v_k) \cap g_j(v_j^x)|}{|g_j(v_j^x)|}, \tag{6.2.1}$$

where v_k is a fixed cluster label in base clustering bc_k. Note that $|\cdot|$ is the number of elements in the specific set. For example, we have $P_{2|4}(A|\alpha) = 2/6 = 1/3$.

All these concepts and functions form the foundation of the framework

Table 6.2: List of Main Notations in Chapter 6

Variable	Explanation
$\{u_1, \cdots, u_m\}$	The set of m objects U
$\{bc_1, \cdots, bc_L\}$	The set of L base clusterings C
$\{c_j^1, \cdots, c_j^{t_j}\}$	The set of t_j clusters in base clustering bc_j
$\{c_*^1, \cdots, c_*^{t^*}\}$	A final clustering fc^* with t^* clusters
V_j	The set of cluster labels in base clustering bc_j
$v_j^x(\in V_j)$	The cluster label of object u_x in base clustering bc_j
$v_k(\in V_k)$	Any cluster label in base clustering bc_k
δ^{Sim}	The similarity measure
$N_{u_x}^{Sim}$	The neighbor set of object u_x based on δ^{Sim}
$(BC_j)_{m\times m}$	The associated similarity matrix of objects for bc_j

for capturing the coupled interactions between base clustering and between objects. The main notations in this chapter are listed in Table 6.2. In addition, several important abbreviations are defined in Table 6.3 to facilitate the reading of this chapter.

6.3 The Coupled Framework of Clustering Ensembles

In this section, a coupled framework of clustering ensembles CCE is proposed in terms of both interactions between base clusterings and between data objects. In the framework described in Figure 6.3, the couplings between base clusterings are revealed via the similarity between cluster labels v_j^x and v_j^y of each base clustering bc_j; and the couplings between objects are specified by defining the similarity between data objects u_x and u_y. In addition, three models are proposed for clustering-based, object-based, and cluster-based consensus building, revealing that the couplings are essential to the clustering ensemble.

Table 6.3: List of Abbreviations in Chapter 6

Abbreviation	Full Name
IaCSC (δ_j^{IaC})	Intra-coupled Clustering Similarity for Clusters
IeRSC ($\delta_{j\|k}$)	Inter-coupled Relative Similarity for Clusters
IeCSC (δ_j^{IeC})	Inter-coupled Clustering Similarity for Clusters
CCSC (δ_j^C)	Coupled Clustering Similarity for Clusters
IaOSO (δ^{IaO})	Intra-coupled Object Similarity for Objects
IeOSO (δ^{IeO})	Inter-coupled Object Similarity for Objects
CCOSO (δ^{CO})	Coupled Clustering and Object Similarity for Objects
CgC (S_{Cg}^C)	Proposed Clustering-based Coupling
OC-Ia (S_O^{IaC})	Proposed Intra-coupled Object-based Coupling
OC-H (S_O^C)	Proposed Hierarchical Object-based Coupling
CrC-Ia ($S_{Cr}^C + \delta^{IaO}$)	Proposed Intra-coupled Cluster-based Coupling
CrC-C ($S_{Cr}^C + \delta^{CO}$)	Proposed Coupled Cluster-based Coupling

In terms of the *clustering coupling*, relationships within each base clustering and the interactions between distinct base clusterings are induced from the coupled nominal similarity measure *COS* in (Wang et al. 2011), which has also been proposed as *CASO* in Section 5.6. The intra-coupling of base clusterings captures the cluster label frequency distribution, while the inter-coupling of base clusterings considers the cluster label co-occurrence dependency (Wang et al. 2011). *Object coupling* also focuses on the intra- and inter-coupling, in which intra-coupling combines all the results of base clusterings for data objects, whereas inter-coupling is explicated by the neighborhood relationship (Guha et al. 2000) among different data objects. The object coupling also leads to a more accurate similarity ($\in [0, 1]$) between data objects. Moreover, as indicated in Figure 6.3, the data objects and base clusterings are associated through the corresponding clusters, i.e., the position of an object in a clustering is determined by which cluster the object belongs to. Therefore, an integrated coupling is derived by treating each cluster label as an attribute value and then defining the similarity between

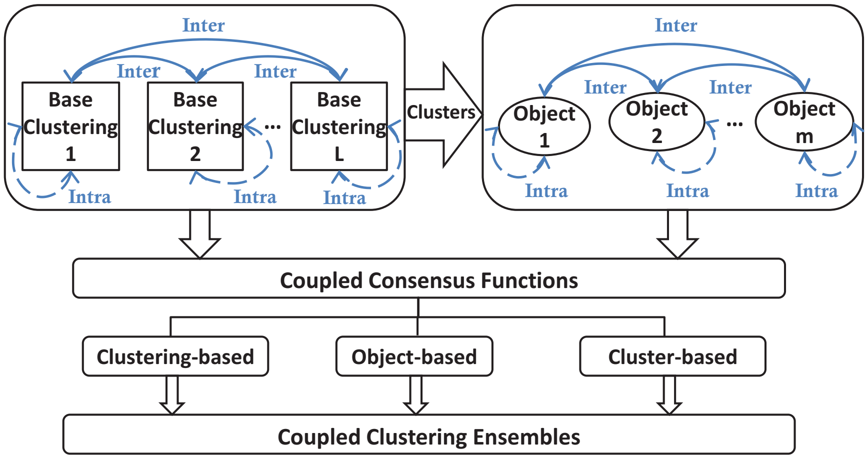


Figure 6.3: A coupled framework of clustering ensembles (CCE), where $\dashleftarrow\dashrightarrow$ indicates the intra-coupling and $\longleftrightarrow$ refers to the inter-coupling.

objects grounded on the similarity between cluster labels over all the base clusterings. Finally, new similarity measures are designed for the clustering-based, object-based, and cluster-based consensus functions by addressing clustering coupling, object coupling and both of them respectively.

Given a set of m objects U and a set of L base clusterings C, we specify those interactions and the coupled consensus functions of CCE in the following two sections.

6.4 Coupling relationships in *CCE*

In this section, we discuss how to describe the coupling of base clusterings and how to represent the coupling of objects.

6.4.1 Coupling of Base Clusterings

Since all base clusterings are conducted on the same data objects, intuitively we assume there must be some relationship among these base clusterings. The coupling of base clusterings is proposed from the perspectives of intra-coupling and inter-coupling. The intra-coupling of base clusterings indicates the involvement of cluster label occurrence frequency within one base clustering, while inter-coupling of base clusterings means the interaction of other base clusterings with this base clustering (Wang et al. 2011). Accordingly, we have:

Definition 6.4.1 (IaCSC) *The* ***Intra-coupled Clustering Similarity for Clusters*** *between cluster labels v_j^x and v_j^y of base clustering bc_j is:*

$$\delta_j^{IaC}(v_j^x, v_j^y) = \frac{|g_j(v_j^x)| \cdot |g_j(v_j^y)|}{|g_j(v_j^x)| + |g_j(v_j^y)| + |g_j(v_j^x)| \cdot |g_j(v_j^y)|}, \tag{6.4.1}$$

where $g_j(v_j^x)$ and $g_j(v_j^y)$ are the set information functions.

By taking into account the frequency of cluster labels, *IaCSC* characterizes the cluster similarity in terms of cluster label occurrence times. As clarified by (Wang et al. 2011) and in Section 5.4.1 as well, Equation (6.4.1) is a well-defined similarity measure $\delta_j^{IaC} \in [1/3, m/(m+4)]$ and satisfies two main principles: greater similarity is assigned to the cluster label pair which owns approximately equal frequencies; the higher these frequencies are, the closer are the two clusters. For example, in Table 6.1, $\delta_j^{IaC}(\alpha, \beta) = 3/4$.

The above two principles are also consistent with the similarity theorem presented in (Lin 1998), in which the commonality corresponds to the product of frequencies and the full description relates to the total sum of individual frequencies and their product. In addition, a comparative evaluation on similarity measures for categorical data has been conducted in (Boriah et al. 2008), delivering *OF* and *Lin* as the two best similarity measures among 14 existing measures on 18 data sets. Both these measures assign higher weights to mismatches or matches on frequent values, and the maximum similarity is attained when the attribute values exhibit approximately equal frequencies (Boriah et al. 2008).

IaCSC considers the interaction between cluster labels within a base clustering bc_j. It does not involve the coupling between base clusterings (e.g. between base clusterings bc_k and $bc_j (k \neq j)$) when calculating cluster label similarity. For this, we discuss the dependency aggregation, i.e. inter-coupled interaction.

Definition 6.4.2 (IeRSC) *The* ***Inter-coupled Relative Similarity for Clusters*** *between cluster labels v_j^x and v_j^y of base clustering bc_j based on another base clustering bc_k is:*

$$\delta_{j|k}(v_j^x, v_j^y|V_k) = \sum_{v_k \in \cap} \min\{P_{k|j}(v_k|v_j^x), P_{k|j}(v_k|v_j^y)\}, \tag{6.4.2}$$

where $v_k \in \cap$ denotes $v_k \in \varphi_{j\to k}(v_j^x) \cap \varphi_{j\to k}(v_j^y)$, $\varphi_{j\to k}$ is the inter-information function, and $P_{k|j}$ is the information conditional probability formalized in Equation (6.2.1).

Definition 6.4.3 (IeCSC) *The* ***Inter-coupled Clustering Similarity for Clusters*** *between cluster labels v_j^x and v_j^y of base clustering bc_j is:*

$$\delta_j^{IeC}(v_j^x, v_j^y|\{V_k\}_{k\neq j}) = \sum_{k=1,k\neq j}^{L} \lambda_k \delta_{j|k}(v_j^x, v_j^y|V_k), \tag{6.4.3}$$

where λ_k is the weight for base clustering bc_k, $\sum_{k=1}^{L} \lambda_k = 1$, $\lambda_k \in [0, 1]$, $V_k (k \neq j)$ is a cluster label set of base clustering bc_k different from bc_j to enable the inter-coupled interaction, and $\delta_{j|k}(v_j^x, v_j^y|V_k)$ is IeRSC.

According to (Wang et al. 2011) and Section 5.4.2, relative similarity $\delta_{j|k}$ is an improved similarity measure derived from *MVDM* proposed by Cost and Salzberg (Cost & Salzberg 1993). It considers the similarity of two cluster labels v_j^x and v_j^y in base clustering bc_j on each possible cluster label in base clustering bc_k to capture the co-occurrence comparison between them. Further, the similarity δ_j^{IeC} between the cluster pair (v_j^x, v_j^y) in base clustering bc_j can be calculated on top of $\delta_{j|k}$ by aggregating all the relative similarity on base clusterings other than bc_j. For the parameter λ_k, in this chapter, we simply assign $\lambda_k = 1/(L-1)$. For example, in Table 6.1, we obtain $\delta_{4|2}(\alpha, \beta|V_2) = 1/3 + 1/2 = 5/6$ and $\delta_4^{IeC}(\alpha, \beta|\{V_1, V_2, V_3\}) = 1/3 \times 5/6 + 1/3 \times 5/6 + 1/3 \times 4/6 = 7/9$ if we take $\lambda_1 = \lambda_2 = \lambda_3 = 1/3$.

Thus, *IaCSC* captures the base clustering frequency distribution by calculating the occurrence times of cluster labels within one base clustering, and *IeCSC* characterizes the base clustering dependency aggregation by comparing the co-occurrence of the cluster labels in objects among different base clusterings. Finally, there is an eligible way to incorporate these two couplings together, specifically:

Definition 6.4.4 (CCSC) *The* ***Coupled Clustering Similarity for Clusters*** *between cluster labels v_j^x and v_j^y of clustering bc_j is:*

$$\delta_j^C(v_j^x, v_j^y|\{V_k\}_{k=1}^L) = \delta_j^{IaC}(v_j^x, v_j^y) \cdot \delta_j^{IeC}(v_j^x, v_j^y|\{V_k\}_{k\neq j}), \tag{6.4.4}$$

where δ_j^{IaC} and δ_j^{IeC} are IaCSC and IeCSC, respectively.

As indicated in Equation (6.4.4), *CCSC* becomes larger by increasing either *IaCSC* or *IeCSC*. For example, in Table 6.1, we could consider the coupled similarity between cluster labels α and β to be $\delta_j^C(\alpha, \beta|\{V_1, V_2, V_3, V_4\}) = 3/4 \times 7/9 = 7/12$.

Here, we choose the multiplication of these two components. The rationale is twofold: (1) *IaCSC* is associated with how often the cluster label occurs while *IeCSC* reflects the extent of the cluster difference brought by other base clusterings. Intuitively, the multiplication of them indicates the total amount of the cluster difference; (2) the multiplication method is consistent with the adapted simple matching distance introduced in (Gan et al. 2007), which considers both the category frequency and matching distance.

Figure 6.4 summarizes the whole process to calculate the coupled similarity for two cluster labels α and β. As indicated here, the similarity value between cluster labels α and β is 7/12, which is larger than 0 suggested by existing methods. Thus, *CCSC* discloses the implicit relationship for both the frequency of cluster labels (intra-coupling) in each base clustering and the co-occurrence of cluster labels (inter-coupling) across different base clusterings. Intuitively, the "intra" here means the calculation of similarity between clusters is limited to only one base clustering, while the "inter" describes how this calculation also considers the involvement of other base clusterings.

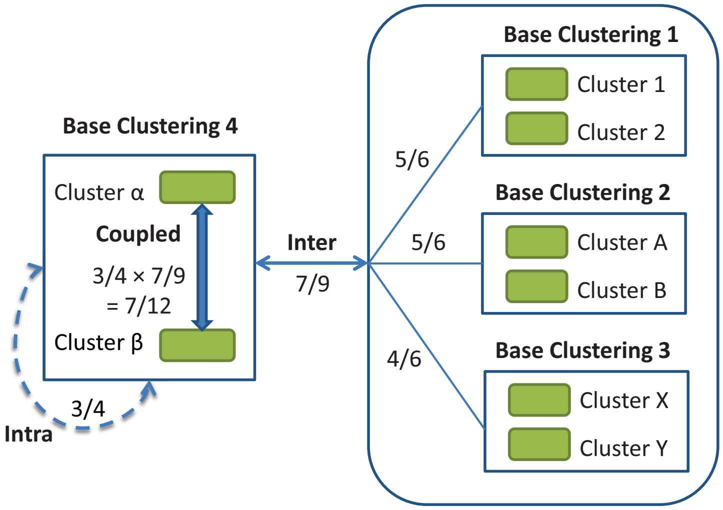


Figure 6.4: An example of the coupled similarity for cluster labels α and β, where ←----→ indicates the intra-coupling and $\longleftrightarrow$ refers to the inter-coupling, with the value along each line being the corresponding similarity.

6.4.2 Coupling of Objects

In the previous section, we presented the couplings of base clusterings from the aspects of intra-coupled similarity and inter-coupled similarity between cluster labels. Here, we proceed by considering the coupling relationships among objects. Similarly, we assume that the objects interact with each other both internally and externally.

In terms of the intra-perspective, the object u_x is coupled with u_y by involving the cluster labels of all the base clusterings for u_x and u_y. The similarity between u_x and u_y could be defined as the average sum of the similarity between the associated cluster labels ranging over all the base clusterings. Formally, we have:

Definition 6.4.5 (IaOSO) *The* ***Intra-coupled Object Similarity for Objects*** *between objects u_x and u_y with respect to all the base clustering* *results of these two objects is:*

$$\delta^{IaO}(u_x, u_y) = \frac{1}{L} \cdot \sum_{j=1}^{L} \delta_j^C (v_j^x, v_j^y | \{V_k\}_{k=1}^L), \tag{6.4.5}$$

where $\delta_j^C(v_j^x, v_j^y, \{V_k\}_{k=1}^L)$ refers to CCSC between cluster labels v_j^x and v_j^y of base clustering bc_j.

In this way, all the *CCSC*s δ_j^C $(1 \leq j \leq L)$ with each base clustering bc_j are summed up for two objects u_x and u_y. Intuitively, the "intra" here represents that the calculation of similarity between objects has nothing to do with other objects. It just involves the two objects to be considered with their internal attributes. For example, the similarity between u_2 and u_3 in Table 6.1 is $\delta^{IaO}(u_2, u_3) = 0.655$ and $\delta^{IaO}(u_2, u_{10}) = 0.684$, which are both larger than 0.5 as provided by the traditional approach. We find that the intra-coupled object similarity between objects u_2 and u_{10} is a little greater than that between u_2 and u_3, which may prove somewhat misleading in terms of the final clustering in the post-processing stage. To solve this problem, we also examine the coupling between objects to further underscore the interaction on the object level.

As indicated in (Guha et al. 2000), the set theory-based similarity measure for categorical values, such as the Jaccard coefficient (Gan et al. 2007), often fails to capture the genuine relationship when the hidden clusters are not well-separated and there is a wide variance in the sizes of clusters. This is also true for our proposed *IaOSO*, since it only considers the similarity between the two objects in question, and it is superior to the Jaccard coefficient because it concerns the interactions among base clusterings while the latter is too rough to characterize the pairwise cluster similarity. However, neither *IaOSO* nor Jaccard coefficient reflects the properties of the neighborhood of the objects. Therefore, we present our new coupled similarity for objects based on the notions of neighboring and *IeOSO* as follows.

Definition 6.4.6 (Neighbor) *A pair of objects u_x and u_y are defined as*

neighbors *if the following holds:*

$$\delta^{Sim}(u_x, u_y) \geq \theta, \tag{6.4.6}$$

where δ^{Sim} *denotes any similarity measure for objects,* $\theta \in [0, 1]$ *is a given threshold.*

In the above definition of neighboring, the similarity measure can be the Jaccard coefficient (Guha et al. 2000) for objects described by categorical attributes, or Euclidean dissimilarity (Gan et al. 2007) for objects depicted by continuous attributes. The neighborhood set of objects u_x is denoted as:

$$N_{u_x}^{Sim} = \{u_z | \delta^{Sim}(u_x, u_z) \geq \theta\}, \tag{6.4.7}$$

which collects all the neighbors of u_x to form an object set N_{u_x}. For example, u_3 and u_{10} are the neighbors of object u_2, since $\delta^{Sim}(u_2, u_3) = \delta^{Sim}(u_2, u_{10}) = 1/3 \geq 0.3$ if we adopt the Jaccard coefficient as the similarity measure and set $\theta = 0.3$, and the neighborhood set of u_2 is $N_{u_2} = \{u_1, u_3, u_4, u_5, u_6, u_7, u_{10}, u_{11}\}$.

Further, we can embody the inter-coupled interaction between different objects by exploring the relationship between their neighborhoods. Intuitively, objects u_x and u_y more likely belong to the same cluster if they have a larger overlap in their neighborhood sets N_{u_x} and N_{u_y}. Below, we use the common neighbors to define the inter-coupled similarity for objects.

Definition 6.4.7 (IeOSO) *The* ***Inter-coupled Object Similarity for Objects*** *between objects* u_x *and* u_y *in terms of other objects* $\{u_z\}$ *is defined as the ratio of common neighbors of* u_x *and* u_y *upon all the objects in* U*, based on similarity* δ^{Sim}*:*

$$\delta^{IeO}(u_x, u_y | U, \delta^{Sim}) = \frac{1}{m} \cdot |\{u_z \in U | u_z \in N_{u_x}^{Sim} \cap N_{u_y}^{Sim}\}|, \tag{6.4.8}$$

where $N_{u_x}^{Sim}$ *and* $N_{u_y}^{Sim}$ *are the neighborhood sets of objects* u_x *and* u_y *based on similarity measure* δ^{Sim}*, respectively.*

Thus, *IeOSO* builds the inter-coupling relationship between each pair of objects by capturing the global knowledge of their neighborhood. Intuitively, the "inter" specifies that the calculation of similarity between objects also concerns other objects if they are in a neighborhood relationship. For example, $\delta^{IeO}(u_2, u_3|U) = 0.583$ and $\delta^{IeO}(u_2, u_{10}|U) = 0.417$ when setting δ^{Sim} to be the Jaccard coefficient and $\theta = 0.3$.

Finally, the intra-coupling and inter-coupled interactions can be considered together to induce the following coupled similarity for objects by exactly specifying the similarity measure δ^{Sim} in (6.4.7) to be *IaOSO* δ^{IaO} as defined in Equation (6.4.5).

Definition 6.4.8 (CCOSO) *The* ***Coupled Clustering and Object Similarity for Objects*** *between objects u_x and u_y is defined when δ^{Sim} is in particular regarded as δ^{IaO}:*

$$\begin{aligned} \delta^{CO}(u_x, u_y|U) &= \delta^{IeO}(u_x, u_y|U, \delta^{IaO}) \qquad (6.4.9) \\ &= \frac{1}{m} \cdot |\{u_z \in U | u_z \in N_{u_x}^{IaO} \cap N_{u_y}^{IaO}\}|, \end{aligned}$$

where the above two sets of objects $N_{u_x}^{IaO} = \{u_z | \delta^{IaO}(u_x, u_z) \geq \theta\}$ and $N_{u_y}^{IaO} = \{u_z | \delta^{IaO}(u_y, u_z) \geq \theta\}$.

In this way, the coupled similarity takes into account both the intra-coupled and inter-coupling relationships between two objects. At the same time, it also considers both the intra-coupled and inter-coupled interactions between base clusterings, since one of the components *IaOSO* of *CCOSO* is built on top of them. Thus, we call this the coupled clustering and object similarity for objects (*CCOSO*). For example, the corresponding neighbors of objects u_2, u_3 and u_{10} are described in Table 6.4 below, here $\theta = 0.65$.

From this table, we observe that the number of common neighbors of objects u_2 and u_3 (i.e., 9) is truly larger than that of objects u_2 and u_{10} (i.e., 7), which correctly corresponds to our claim in Section 6.1. Based on Equation (6.4.9), we obtain $\delta^{CO}(u_2, u_3|U) = 0.75$ and $\delta^{CO}(u_2, u_{10}|U) = 0.5$. This means that the similarity between objects u_2 and u_3 is larger than that between u_2 and u_{10}, which effectively remedies the issue caused by $\delta^{IaO}(u_2, u_3) < \delta^{IaO}(u_2, u_{10})$.

Table 6.4: An Example of Neighborhood Domain for Object

Object	Neighborhood Domain
u_2	$\{u_1, u_3, u_4, u_5, u_6, u_7, u_8, u_{10}, u_{11}, u_{12}\}$
u_3	$\{u_1, u_2, u_4, u_5, u_6, u_7, u_8, u_9, u_{10}, u_{11}, u_{12}\}$
u_{10}	$\{u_2, u_3, u_6, u_7, u_8, u_9, u_{11}, u_{12}\}$
Object Pair	Common Neighbors
u_2, u_3	$\{u_1, u_4, u_5, u_6, u_7, u_8, u_{10}, u_{11}, u_{12}\}$
u_2, u_{10}	$\{u_3, u_6, u_7, u_8, u_{11}, u_{12}\}$

6.5 Coupled Consensus Function in *CCE*

There are many ways to define the consensus function such as pairwise agreements between base clusterings, co-associations between data objects, and interactions between clusters. Some of the criteria focus on the estimation of similarity between base clusterings (Li et al. 2010, Topchy et al. 2005), some are based on the similarity between data objects (Strehl & Ghosh 2002), and others are associated with the similarity between clusters (Fern & Brodley 2004, Iam-On et al. 2011). In the following, we specify the coupled versions of clustering-based, object-based, and cluster-based criteria individually.

6.5.1 Clustering-based Coupling

The clustering-based consensus function captures the pairwise agreement between base clusterings. Note that each base clustering bc_j defines an associated similarity matrix $(BC_j)_{m\times m}$ that stores the information for each pair of objects about their similarity. Each entry $BC_j(x, y)$ of the matrix represents the similarity between objects u_x and u_y within the base clustering bc_j.

The usual way to define the entry $BC_j(x, y)$ of the similarity matrix BC_j is to justify whether objects u_x and u_y are in the same cluster of base clustering bc_j, i.e., whether u_x and u_y have the same cluster label. Formally:

$$BC_j(x,y) = \begin{cases} 1 & \text{if } v_j^x = v_j^y, \\ 0 & \text{otherwise,} \end{cases} \tag{6.5.1}$$

where v_j^x and v_j^y are the cluster labels of u_x and u_y in base clustering bc_j, respectively. Then, given two base clusterings bc_{j_1} and bc_{j_2}, a common measure of discrepancy is the partition difference (PD) (Li et al. 2010):

$$S_{Cg}(bc_{j_1}, bc_{j_2}) = \sum_{1 \le x,y \le m} [BC_{j_1}(x,y) - BC_{j_2}(x,y)]^2, \tag{6.5.2}$$

where x and y refer to the indexes of objects u_x and u_y respectively. However, this traditional way is too imprecise to characterize the similarity between objects, and it assumes independence among the base clusterings.

Alternatively, we can focus on the entry $BC_j(x,y)$ to incorporate the coupling of base clusterings as follows:

$$BC_j^C(x,y) = \delta_j^C(v_j^x, v_j^y | \{V_k\}_{k=1}^L), \tag{6.5.3}$$

$$S_{Cg}^C(bc_{j_1}, bc_{j_2}) = \sum_{1 \le x,y \le m} [BC_{j_1}^C(x,y) - BC_{j_2}^C(x,y)]^2, \tag{6.5.4}$$

where δ_j^C refers to $CCSC$ in Definition 6.4.4. We denote this newly proposed clustering-based coupling to be CgC.

Intuitively, S_{Cg}^C calculates the sum of similarity between objects that belong to different base clusterings bc_{j_1} and bc_{j_2}. A target clustering fc^* thus should be:

$$fc^* = \underset{c^1,\cdots,c^{t^*}}{\arg} \min \sum_{j=1}^{L} S_{Cg}^C(fc, bc_j), \tag{6.5.5}$$

where $fc = \{c^1, \cdots, c^{t^*}\}$ denotes the candidate set of clusters for final clustering fc^*. According to (Topchy et al. 2005), the optimization problem in (6.5.5) then can be heuristically approached by *k-means* operating in the normalized object-label space OL with each entry to be:

$$OL(u_x, v_j^y) = \delta_j^C(v_j^x, v_j^y | \{V_k\}_{k=1}^L) - \mu^y(\delta_j^C), \tag{6.5.6}$$

where u_x is an object, v_j^y is a cluster label in bc_j, and $\mu^y(\delta_j^C)$ is the mean of $\delta_j^C(v_j^x, v_j^y|\{V_k\}_{k=1}^L)$ for every cluster.

Thus, the clustering-based coupling addresses the intra-coupling and inter-coupling of base clusterings to form the coupled consensus function CgC.

6.5.2 Object-based Coupling

The object-based consensus function captures the co-associations between objects. Given two objects u_x and u_y, based on all the base clustering results, a simple and obvious heuristic to describe the similarity between u_x and u_y is the entry-wise average of the L associated similarity matrices induced by the L base clusterings. In this way, an overall similarity matrix BC^* with a finer resolution is yielded (Strehl & Ghosh 2002). Formally, we have:

$$BC^*(x,y) = \frac{1}{L} \cdot \sum_{j=1}^{L} BC_j(x,y). \tag{6.5.7}$$

The entry of the induced overall similarity matrix BC^* is the weighted average sum of each associated pairwise similarity BC_j between objects of every base clustering. However, the common pairwise similarity measure $BC_j(x,y)$ is rather inadequate since only 1 and 0 are considered as defined in Equation (6.5.1). The relationship that is neither within nor between base clusterings (i.e., bc_{j_1} and bc_{j_2}) is explicated. In addition, most existing methods (Christou 2011, Fern & Brodley 2004, Gionis et al. 2007) only use the similarity measure between objects when clustering them, which thus does not involve the context (i.e. neighborhood) of the objects.

To solve the first two issues above, we regard the entry $BC^*(x,y)$ of the overall similarity matrix to be $IaOSO$:

$$S_O^{IaC}(u_x,u_y) = BC^*(x,y) = \delta^{IaO}(u_x,u_y), \tag{6.5.8}$$

where δ^{IaO} is defined in (6.4.5). Here, S_O^{IaC} captures the intra-coupled interactions within two objects as well as both the intra-coupled and inter-coupled interactions among base clusterings. Alternatively, we can also assign $BC_j(x,y)$ of base clustering bc_j to be δ_j^C (6.4.4), in the same way as

Equation (6.5.3); then, the overall similarity matrix BC^* is obtained by averaging the associated similarity matrix BC_j over all the base clusterings according to (6.5.7). Afterwards, *METIS* is applied to the overall similarity matrix BC^* to produce the final clustering fc^*. We denote this newly proposed intra-coupled object-based coupling method as *OC-Ia*.

Further considering the third issue above, both the intra-couplings and inter-couplings of clusterings and of objects are incorporated as follows:

$$S_O^C(u_x, u_y) = BC^*(x, y) = \delta^{CO}(u_x, u_y|U), \tag{6.5.9}$$

where δ^{CO} is defined in (6.4.9). We would like to maximize the sum of $\delta^{CO}(u_x, u_y|U)$ (6.4.9) for data object pairs u_x, u_y belonging to a single cluster, and at the same time minimize the sum of $\delta^{CO}(u_x, u_y|U)$ for u_x and u_y in different clusters. Accordingly, the desired final clustering $fc^* = \{c_*^1, \cdots, c_*^{t^*}\}$ with t^* clusters can be obtained by maximizing the following criterion function:

$$fc^* = \underset{c^1, \cdots, c^{t^*}}{\arg} \max \sum_{t=1}^{t^*} m_t \cdot \sum_{u_x, u_y \in c^t} \frac{S_O^C(u_x, u_y) \cdot m}{m_t^{1+2f(\theta)}}, \tag{6.5.10}$$

where c^t denotes the tth cluster of size m_t, m is the total number of objects, and $f(\theta) = (1-\theta)/(1+\theta)$. The rationale of the above function is twofold: on one hand, one of our goals is to maximize $\delta^{CO}(u_x, u_y|U)$ for all pairs of objects in the same cluster $u_x, u_y \in c^t$; on the other hand, we divide the total *CCOSO* (i.e., $S_O^C = \delta^{CO}$) involving pairs of objects in cluster c^t by the expected sum of δ^{CO} in c^t, which is $m_t^{1+2f(\theta)}/m$ (Guha et al. 2000); and then weigh this quantity by m_t, i.e., the number of objects in c^t. Dividing by the expected sum of δ^{CO} prevents a clustering in which all objects are assigned to a single cluster, and avoids objects with very small coupled similarity value between them from being put in the same cluster (Guha et al. 2000). Subsequently, we adapt the standard agglomerative hierarchical clustering algorithm to obtain the final clustering fc^* by solving Equation (6.5.10) (Guha et al. 2000). We abbreviate this newly proposed hierarchical object-based coupling to *OC-H*.

The intra-coupled object-based coupling thus examines the intra-coupling and inter-coupling of base clusterings as well as the intra-coupling of objects to form the coupled consensus function *OC-Ia*, while the hierarchical object-based coupling considers both the intra-coupling and inter-coupling of base clustering and objects to build the coupled consensus function *OC-H*.

6.5.3 Cluster-based Coupling

The cluster-based consensus function characterizes the interactions between every two clusters. One of the basic approaches based on the relationship between clusters is *MCLA* proposed by Strehl and Ghosh (Strehl & Ghosh 2002). The idea in *MCLA* is to yield object-wise confidence estimates of cluster membership, to group and then to collapse related clusters represented as hyper-edges. The similarity measure of clusters in *MCLA* is the Jaccard matching coefficient (Gan et al. 2007), formally:

$$S_{Cr}(c_{j_1}^{t_1}, c_{j_2}^{t_2}) = \frac{|c_{j_1}^{t_1} \cap c_{j_2}^{t_2}|}{|c_{j_1}^{t_1} \cup c_{j_2}^{t_2}|}, \tag{6.5.11}$$

where $c_{j_1}^{t_1}$ and $c_{j_2}^{t_2}$ are the t_1th cluster of base clustering bc_{j_1} and the t_2th cluster of base clustering bc_{j_2}, respectively.

The above similarity measure S_{Cr} considers neither coupling between base clusterings nor interaction between objects. Therefore, it lacks the capability to reflect the essential link and relationship among data. To remedy this problem, we define the coupled similarity between clusters $c_{j_1}^{t_1}$ and $c_{j_2}^{t_2}$ in terms of both the coupling relationships between clusterings and between objects. The average sum of every two-object pairs in $c_{j_1}^{t_1}$ and $c_{j_2}^{t_2}$ respectively is selected here to specify the coupled similarity between clusters:

$$S_{Cr}^{C}(c_{j_1}^{t_1}, c_{j_2}^{t_2}) = \frac{1}{m_{t_1} m_{t_2}} \sum_{u_x \in c_{j_1}^{t_1}, u_y \in c_{j_2}^{t_2}} S_O(u_x, u_y), \tag{6.5.12}$$

where m_{t_1} and m_{t_2} are the sizes of clusters $c_{j_1}^{t_1}$ and $c_{j_2}^{t_2}$, respectively; $S_O(u_x, u_y)$ is the coupled similarity for objects, and can be either δ^{IaO} (6.4.5) or δ^{CO} (6.4.9). If $S_O = \delta^{IaO}$, the cluster-based coupling includes the intra- and inter-coupled interaction between base clusterings as well as the intra-coupled interaction between objects; if $S_O = \delta^{CO}$, it reveals both the intra- and inter-coupled interactions between base clusterings and between objects. Afterwards, *METIS* is used based on the cluster-cluster similarity matrix to conduct meta-clustering as in (Strehl & Ghosh 2002). We denote the cluster-based coupling as *CrC* (including *CrC-Ia* with δ^{IaO} and *CrC-C* with δ^{CO}).

The intra-coupled cluster-based coupling considers the intra-coupling and inter-coupling of base clusterings together with the intra-coupling of objects to define the coupled consensus function *CrC-Ia*, while the coupled object-based coupling concerns both the intra-couplings and inter-couplings of base clustering and objects to construct the coupled consensus function *CrC-C*.

6.5.4 Miscellaneous Issues

How to Establish the Number of Final Clusters: The automatic identification of the appropriate number of clusters is a deep research problem that has attracted significant attention (Gionis et al. 2007, Kuncheva & Vetrov 2006, Wang, Domeniconi & Laskey 2010). There are four ways to handle this issue: imposing a hard constraint on the number of clusters or on their quality, model selection, finding the size t^* of final clustering by similarity analysis, and nonparametric estimation. In our experiments, for simplicity, the number of clusters t^* is fixed, the same as the number of classes in each data set. The different ways to determine t^* can also be incorporated into our proposed coupled consensus functions.

How to Generate Base Clusterings: There are several methods of providing diverse base clusterings: using different clustering algorithms (Gionis et al. 2007), employing random or different parameters for some algorithms (Iam-On et al. 2011), and adopting random sub-sampling or random projection of the data (Fern & Brodley 2004). Since our focus is mainly on the consensus function, we use *k-means* on random sub-sampling (Fern & Brodley 2004) of the data as the base clustering algorithm in our experiments. The number t^j of base clustering bc_j is pre-defined for each data set and remains the same for all clustering runs.

How to Post-process Clustering: In the proposed *CCE* framework, we mainly focus on the consensus function based on pairwise interactions between base clusterings, between objects and between clusters. Those interactions are described by the corresponding similarity matrices. Thus, a common and recommended way to combine the base clusterings is to re-cluster the objects using any reasonable similarity-based clustering algorithm. In our experiments, we choose *k-means*, *agglomerative algorithm* (Guha et al. 2000) and *METIS* (Strehl & Ghosh 2002) due to their popularity in the clustering ensemble.

How to Deal with Big Data: In case the data set is large, random sampling and labeling enable the pairwise similarity-based *CCE* to reduce the number of objects to be considered, and ensure that the input data set fits the main memory. Efficient algorithms for selecting random samples can be found in (Guha et al. 2000). As indicated in (Gionis et al. 2007), sampling $O(\log m)$ objects is sufficient to guarantee that at least one object in a large cluster will be selected with high probability. Afterwards, *CCE* assigns the remaining data objects to the clusters generated by the sampled objects, according to the similarity between each object and a fraction of objects from every cluster. If the sum of similarity between the object u_x to be labeled and the objects chosen from a final cluster c_*^t is maximum, then the object u_x is allocated to the tth final cluster c_*^t. The similarity of objects here can be either *IaOSO* (δ^{IaO}) or *CCOSO* (δ^{CO}). In our experiments, we will consider *IaOSO* (δ^{IaO}) for simplicity.

6.6 Algorithm and Analysis

In previous sections, we have discussed the coupled framework of clustering ensembles *CCE* from the perspectives of coupling of clusterings, coupling of objects, and coupled consensus functions. They are all based on the intra- and inter-coupled interactions between clusterings and between

Algorithm 6.1: Coupled Similarity for Clusters *CCSC*

```
Data: Object set $U = \{u_1, \cdots, u_m\}$ and $u_x, u_y \in U$, base clustering
      set $C = \{bc_1, \cdots, bc_L\}$, and weight $\lambda = (\lambda_k)_{1\times L}$.
Result: Similarity matrix $CCSC$ between cluster labels.
begin
    maximal cluster label $r(j) \longleftarrow \max(V_j)$
    for every cluster label pair $(v_j^x, v_j^y \in [1, r(j)])$ do
        $U_1 \longleftarrow \{i | v_j^i == v_j^x\}$, $U_2 \longleftarrow \{i | v_j^i == v_j^y\}$
        // Compute intra-coupled similarity between cluster
           labels $v_j^x$ and $v_j^y$.
        $\delta_j^{IaC}(v_j^x, v_j^y) = (|U_1||U_2|)/(|U_1| + |U_2| + |U_1||U_2|)$
        $\delta_j^C(v_j^x, v_j^y | \{V_k\}_{k=1}^L) \longleftarrow \delta_j^{IaC}(v_j^x, v_j^y) \cdot IeCSC(v_j^x, v_j^y)$
    $CCSC(v_j^x, v_j^y) \longleftarrow \delta_j^C(v_j^x, v_j^y | \{V_k\}_{k=1}^L)$
    end

Function $IeCSC(v_j^x, v_j^y, U_1, U_2)$
begin
    for each base clustering $(bc_k \in C) \wedge (bc_k \neq bc_j)$ do
        $\varphi \longleftarrow \{v_k^x | x \in U_1\} \cap \{v_k^y | y \in U_2\}$
        for every intersection $v_k^z \in \varphi$ do
            $U_0 \longleftarrow \{i | v_k^i == v_k^z\}$
            $ICP_x \longleftarrow |U_0 \cap U_1| / |U_1|$
            $ICP_y \longleftarrow |U_0 \cap U_2| / |U_2|$
            $Min_{(x,y)} \longleftarrow min(ICP_x, ICP_y)$
        $\delta_{j|k}(v_j^x, v_j^y | V_k) = sum[Min_{(x,y)}]$
    // Compute inter-coupled similarity between two cluster
       labels $v_j^x$ and $v_j^y$.
    $\delta_j^{IeC}(v_j^x, v_j^y | \{V_k\}_{k\neq j}) = sum[\lambda_k \cdot \delta_{j|k}(v_j^x, v_j^y | V_k)]$
    return $IeCSC(v_j^x, v_j^y) = \delta_j^{IeC}(v_j^x, v_j^y | \{V_k\}_{k\neq j})$
```

Algorithm 6.2: Coupled Similarity for Objects *CCOSO*

```
Data: Object set $U = \{u_1, \cdots, u_m\}$ and $u_x, u_y \in U$, base clustering
      set $C = \{bc_1, \cdots, bc_L\}$, and threshold $\theta \in [0, 1]$.
Result: Similarity $CCOSO(u_x, u_y)$ between objects $u_x, u_y$.
begin
    for each base clustering $bc_j \in C$ do
        $\delta_j^C(v_j^x, v_j^y | \{V_k\}_{k=1}^L) \longleftarrow CCSC(v_j^x, v_j^y)$
    // Compute intra-coupled similarity between two objects $u_x$
       and $u_y$.
    $\delta^{IaO}(u_x, u_y) = 1/L \cdot sum[\delta_j^C(v_j^x, v_j^y | \{V_k\}_{k=1}^L)]$
    neighbor sets $N_{u_x} = N_{u_y} = \emptyset$
    for objects $(u_{z_1}, u_{z_2} \in U) \wedge (u_{z_1} \neq u_x) \wedge (u_{z_2} \neq u_y)$ do
        if $\delta^{IaO}(u_x, u_{z_1}) \geq \theta$ then
            $N_{u_x} = \{u_{z_1}\} \cup N_{u_x}$
        if $\delta^{IaO}(u_y, u_{z_2}) \geq \theta$ then
            $N_{u_y} = \{u_{z_2}\} \cup N_{u_y}$
    // Compute inter-coupled similarity between two objects $u_x$
       and $u_y$.
    $\delta^{CO}(u_x, u_y | U) = 1/m \cdot |N_{u_x} \cap N_{u_y}|$
    $CCOSO(u_x, u_y) \longleftarrow \delta^{CO}(u_x, u_y | U)$
    end
```

objects. Therefore, in this section, we design two algorithms *CCSC*[1] and *CCOSO* to compute the coupled similarity for each pair of cluster labels and the coupled similarity for objects u_x and u_y, respectively.

As shown in these two algorithms, the computational complexity for *CCSC* is $O(LT^3)$, and the computational complexity for *CCOSO* is $O(L^2T^3+$

[1] All the cluster labels of each base clustering need to be encoded as numbers, starting at one and increasing to the maximum which is the respective number of clusters in this base clustering.

$2m$), where L is the number of base clusterings, T is the maximal number of clusters in all the base clusterings, and m is the total number of objects.

6.7 Empirical Study

This section presents the performance evaluation of the coupled framework *CCE* in terms of the clustering-based (*CgC*), object-based (*OC-Ia* and *OC-H*), and cluster-based (*CrC-Ia* and *CrC-C*) couplings. The experiments are performed on 11 synthetic and real data sets to discover the implicit relationships between base clusterings and between data objects, to validate accuracy, stability, and robustness of various consensus functions, as well as to explore the dependency between data characteristics and final clustering quality. All the experiments are conducted on a Dell Optiplex 960 equipped with an Intel Core 2 Duo CPU with a clock speed of 2.99 GHz and 3.25 GB of RAM running on Microsoft Windows XP.

Table 6.5: Description of Data Sets in Chapter 6

Data Set	m	n	t^p	Source
Sy1	200	2	2	modified from (Strehl & Ghosh 2002)
Sy2	400	6	4	modified from (Kuncheva & Vetrov 2006)
Iris	150	4	3	UCI repository (Frank & Asuncion 2010)
Wine	178	13	3	UCI repository (Frank & Asuncion 2010)
Seg	210	19	7	UCI repository (Frank & Asuncion 2010)
Glass	214	9	6	UCI repository (Frank & Asuncion 2010)
Ecoli	336	7	8	UCI repository (Frank & Asuncion 2010)
Ionos	351	34	2	UCI repository (Frank & Asuncion 2010)
Blood	748	5	2	UCI repository (Frank & Asuncion 2010)
Vowel	990	10	11	UCI repository (Frank & Asuncion 2010)
Yeast	1484	8	10	UCI repository (Frank & Asuncion 2010)

6.7.1 Data Sets

The experimental evaluation is conducted on 11 data sets, including two synthetic data sets (i.e., Sy1 and Sy2, which are 2-Gaussian modified from (Strehl & Ghosh 2002) and 4-GaussianN modified from (Kuncheva & Vetrov 2006), respectively) and nine real-life data sets from UCI (Frank & Asuncion 2010). Table 6.5 summarizes the details of these data sets, where m is the number of objects, n is the number of dimensions, and t^p is the number of pre-known classes. Those true classes are only used to evaluate the quality of the clustering results, not the process of aggregating base clusterings. The number of true classes is only used to set the number of clusters both in building the base clusterings and in the post-processing stage. Since we do not involve the information of attributes after building base clusterings, we order the data sets according to the number of objects ranging from 150 to 1484. Note that the second synthetic data set Sy2 is initially created to follow the two-dimension Gaussian distribution and then added with four more dimensions of uniform random noise in the way presented in (Kuncheva & Vetrov 2006). When the data size, which is the number of objects, exceeds 700 (e.g., Blood, Vowel, and Yeast), we regard it as a big data set and use the method proposed in Section 6.5.4 to deal with it.

6.7.2 Baseline Approaches and Parameters

As previously presented, our experiments are designed from the following three perspectives:

1. Clustering-based: Besides the partition difference (PD) proposed in (Li et al. 2010), QMI is also an effective clustering-based criterion (Topchy et al. 2005), which has proved to be equivalent to *Category Utility Function* in (Li et al. 2010). We will compare the clustering-based coupling (CgC) with its baseline method PD (Li et al. 2010), EM and QMI (Topchy et al. 2005).

2. Object-based: In this group, we will compare the intra-coupled object-

based coupling *OC-Ia* with its baseline method *CSPA* (Strehl & Ghosh 2002), and compare the hierarchical object-based coupling *OC-H* with *CSPA* (Strehl & Ghosh 2002) and with the categorical clustering algorithms: *ROCK* (Guha et al. 2000) (the baseline method of *OC-H*) and *LIMBO* (Andritsos et al. 2004).

3. Cluster-based: Based on *MCLA* (Strehl & Ghosh 2002), *HBGF* is another promising cluster-based criterion (Fern & Brodley 2004). It also collectively considers the similarity between objects and clusters but lacks the discovery of coupling. Iam-On et al. (Iam-On et al. 2011) proposed a link-based approach (*LB*), which is an improvement on *HBGF*. Below, cluster-based coupling *CrC* (including *CrC-Ia* and *CrC-C*) is compared with their baseline methods *MCLA* (Strehl & Ghosh 2002), *HBGF* (Fern & Brodley 2004), and *LB* (Iam-On et al. 2011, Iam-On & Boongoen 2012) (including *LB-P* and *LB-S*[2]).

As indicated in Section 6.5.4, *k-means* on random sub-sampling (Fern & Brodley 2004) of the data is used to produce a diversity of base clusterings; *k-means* and *agglomerative algorithm* are used to post-process the coupled consensus functions *CgC* and *OC-H*, respectively, and *METIS* is adopted to post-process the consensus functions *OC-Ia*, *CrC-Ia* and *CrC-C*. Here, *OC-H* is built based on *ROCK* (Guha et al. 2000), thus *agglomerative algorithm* is adopted to do the post-processing as *ROCK* does. But *METIS* is much more efficient than *agglomerative algorithm*, so we use *METIS* to post-process *OC-Ia*. The parameters in the following are especially important:

- θ: The neighbor threshold in (6.4.6) is defined to be the average *IaOCO* and Jaccard coefficient (Guha et al. 2000) values of pairwise objects for *OC-H* and *ROCK*, respectively.

[2] The performance of the model (i.e., *WTU+SPEC*) proposed in (Iam-On & Boongoen 2012) is between that of *LB-P* (i.e., *CSM+PAM* (Iam-On et al. 2011)) and that of *LB-S* (i.e., *CSM+SPEC* (Iam-On et al. 2011)), so we only report the results of *LB-P* and *LB-S*.

- L: The ensemble size (i.e., the number of base clusterings) is taken to be $L = 10$. The reason for this selecting is explained in Section 6.7.3.
- t^j, t^*: The number of clusters in the base clustering bc_j and final clustering fc^* are both regarded as the number of pre-known classes t^p, i.e., $t^j = t^* = t^p$.
- λ_k: The weight λ_k for base clustering bc_k in Definition 6.4.3 on $IeCSC$ is simplified as $\lambda_k = 1/L = 1/10$.
- NR: The number of runs for each clustering ensemble is fixed as $NR = 50$ to obtain corresponding average results for the evaluation measures.

Other parameters of the compared methods remain the same as the original approaches and models.

Since each clustering ensemble method divides data objects into a partition of t^p (i.e. the number of true classes) clusters, we then evaluate the clustering quality against the corresponding true partitions by using these external criteria: accuracy (AC) (Cai et al. 2005) and normalized mutual information (NMI) (Cai et al. 2005). We also judge the stability of multiple runs by using the combined stability index (CSI) (Kuncheva & Vetrov 2006), as well as the robustness (Topchy et al. 2005) of the clustering ensemble by comparing the average AC, NMI, and CSI scores across different data sets. In brief, AC and NMI describe the degree of approximation between the obtained clusters and the true data classes. CSI reveals the stability between them across $NR = 50$ runs, and reflects the deviation of the results across different runs. In fact, the larger the AC or NMI or CSI is, the better the clustering ensemble algorithm is. Note that the correspondence problem on mapping between the derived clusters and the known classes needs to be solved before evaluation. The optimal correspondence can be obtained using the Hungarian method (Topchy et al. 2005) with $O((t^p)^3)$ complexity for t^p clusters.

Table 6.6: Evaluation Measures on Base Clusterings

Data Set	AC			NMI			CSI
	Max	Avg	Min	Max	Avg	Min	Avg
Sy1	0.955	0.950	0.945	0.745	0.720	0.693	0.714
Sy2	0.503	0.460	0.385	0.406	0.406	0.406	0.698
Iris	0.927	0.827	0.513	0.750	0.656	0.427	0.791
Wine	0.708	0.689	0.556	0.441	0.424	0.388	0.659
Seg	0.586	0.529	0.433	0.548	0.496	0.410	0.820
Glass	0.517	0.479	0.449	0.338	0.307	0.276	0.602
Ecoli	0.687	0.512	0.470	0.539	0.437	0.398	0.530
Ionos	0.712	0.704	0.650	0.131	0.107	0.014	0.670
Blood	0.739	0.709	0.707	0.017	0.016	0.013	0.780
Vowel	0.373	0.354	0.339	0.435	0.415	0.388	0.802
Yeast	0.384	0.332	0.319	0.250	0.220	0.218	0.817

6.7.3 Experimental Results

Based on the evaluation measures (i.e., AC, NMI and CSI), Table 6.6 displays the performance of the base clustering algorithm (i.e., *k-means*) over synthetic and real data sets. Note that Max, Avg, and Min represent the maximal, average, and minimum corresponding evaluation scores among input base clusterings, respectively. Below, we report the experimental results on implicit relationship discovery, clustering-based comparison, object-based comparison, and cluster-based comparison individually.

Implicit Relationship Discovery

Prior to the experiments on the clustering ensemble, we first compare the implicit relationship revealed by different similarity measures between objects. The similarity measures to be compared are listed here: traditional measure (*TO* for short) specified in formulae (6.5.1) and (6.5.7), intra-coupled object similarity (*IaO* for short) defined in formula (6.4.5), coupled clustering and

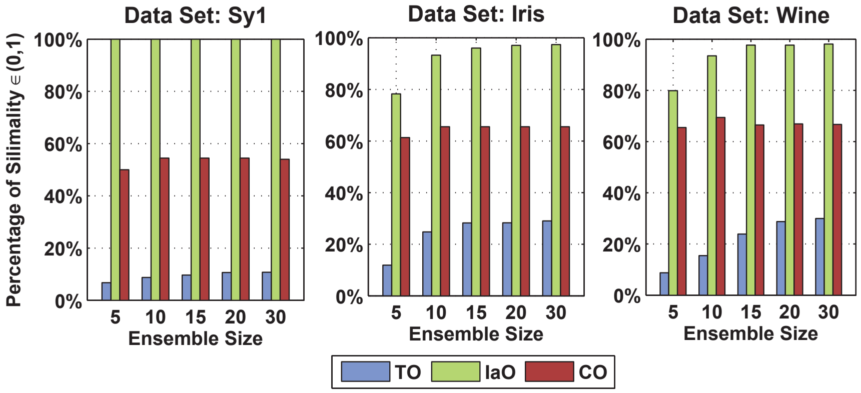


Figure 6.5: Percentage of pairwise similarity $\in (0, 1)$ between objects among all the similarity values.

object similarity (CO for short) formalized in formula (6.4.9). The unknown relationship captured by similarity measures is quantified as the percentage of similarity values that fall within the open interval $(0, 1)$ among all of the pairwise similarities between objects. The ratios reported in this comparison are the average values across 50 runs of generating base clusterings.

Figure 6.5 presents the percentage of similarity values $\in (0, 1)$ (axis y) for three similarity measures (i.e., TO, IaO and CO) on three data sets (i.e., Sy1, Iris, and Wine) with different ensemble sizes L ranging from 5 to 30 (axis x). It is remarkable to note that the proportions of similarity values $\in (0, 1)$ for IaO and CO are much higher than for TO. This empirical evidence signifies that our proposed IaO and CO are capable of discovering the hidden relationships among data objects, while the TO performs rather poorly by mostly assigning 0 and 1 to the similarity between objects.

Another interesting observation is that the percentage score of IaO is larger than that of CO. This is probably due to the fact that the similarity measure CO is filtered and refined from IaO, which means CO may amplify several IaO similarity values and also diminish some IaO values according to the neighborhood coupling. In essence, IaO captures a partial picture of the similarity between objects, while CO provides a global view in terms of the context around objects.

A third discovery is that the ensemble size $L = 10$ is large enough to capture the relationship between data objects, as compared to $L = 15, 20, 30$. It can be also observed that the percentages of TO and IaO have an increasing trend when L goes up, while the ratio of CO keeps fluctuating. The reason is that the likelihood that the TO and IaO values will take 0 become smaller with the increasing number of base clusterings. However, the opportunity for CO to be evaluated as 0 is uncertain since the average threshold for definings neighbors (see formula (6.4.6)) also increases, which probably leads to a smaller set of neighbors (see formula (6.4.7)). Thus, we select $L = 10$ in the experiments below to preserve the ability to discover enough of a relationship but with relatively low computational complexity.

In the following sections, the experimental results are presented and analyzed in three groups: clustering-based comparison which focuses on the evaluation of coupling between base clusterings, object-based comparison which studies the utility of intra-coupling and inter-coupling between objects, and cluster-based comparison which identifies the joint effect of couplings between base clusterings and between objects. We analyze the clustering performance individually by considering the couplings step by step within each group of experiments, although a comparison across these three groups is beyond the scope of this chapter. Note that all the values reported on AC and NMI are the averages across multiple clustering ensembles (i.e., exactly 50 runs). The CSI value reveals the total deviation apart from the average of 50 runs in each experiment, and the improvement rate below refers to the absolute difference value between two evaluation scores.

Clustering-based Comparison

Figure 6.6 shows the performance comparison of different clustering-based ensemble methods over two synthetic and nine real-life data sets in terms of AC, NMI and CSI. It is clear that our proposed CgC usually generates data partitions of higher quality than its baseline model *PD* and other compared approaches, i.e. *EM* and *QMI*. Specifically, in terms of accuracy, the AC improvement rate ranges from 1.59% (*QMI* on Sy1) to 12.71% (*EM* on Vowel), and there has been significant CSI improvement (from 0.69% to 49.83%) except in one case: Glass. Overall, the average improvement rate of *CgC* on AC across all the other methods over all the data sets is 3.85%, and the average improvement rate of *CgC* on CSI is 8.99%. Also, in several data sets such as Sy1, Sy2, Wine, Seg, Ionos, Blood, Vowel and Yeast, the AC measures exceed the maximum of AC in the corresponding base clusterings, i.e. Max(AC) in Table 6.6. All the AC and CSI values of *CgC* are higher than the corresponding average values of base clustering. Another observation is that none of the other three consensus functions compared is an outright winner; *QMI* is the best in most cases, followed by *PD* with *EM* being the worst. However, our proposed *CgC* outperforms all the algorithms compared on almost every data set. A similar situation can also be observed when NMI is used to evaluate clustering quality. Statistical analysis, namely the t-test, has been carried out on the AC and NMI of our *CgC*, at a 95% significance level. The null hypothesis that *CgC* is better than base clusterings and the best result of other methods in terms of AC and NMI is accepted.

In addition, it seems that the improvement level of *CgC* upon other methods is associated with the quality of base clusterings: the better quality of base clusterings corresponds to a relatively smaller level of improvement. This point of view will be justified later in Section 6.7.4.

Hence, we draw the empirical conclusion that clustering accuracy and stability can be further improved with CgC by involving the couplings of clusterings. The improvement rate is dependent on the accuracy of base clusterings.

Object-based Comparison

The evaluations (i.e. AC, NMI and CSI) of distinct object-based ensemble methods are exhibited in Figure 6.7. Eight data sets with smaller size are chosen because of the high computational complexity in this group of exper-

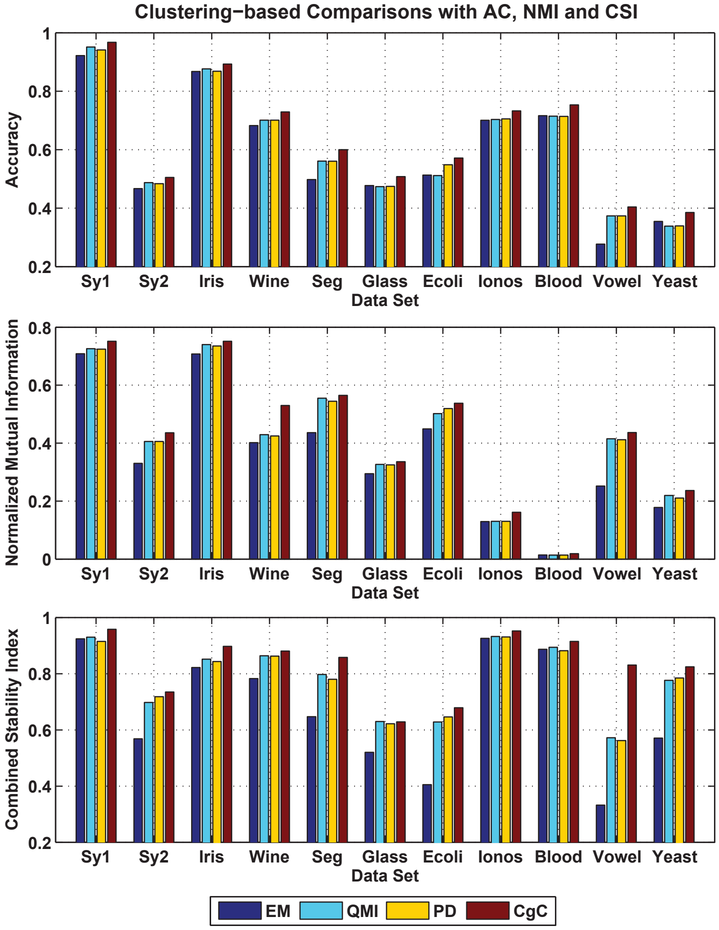


Figure 6.6: Clustering-based comparisons on AC, NMI and CSI.

iments. We observe that, with the exception of a few items, our proposed *OC-Ia* mostly outperforms the ensemble method *CSPA* and categorical clustering algorithm *ROCK* in terms of both NMI and CSI. Our proposed *OC-H* has the largest NMI and CSI values over most of the data sets. Here, it can be clearly seen that our proposed *OC-Ia* and *OC-H* both achieve better clustering quality compared to their respective baseline methods *CSPA* and *ROCK*. The average NMI and CSI improvement rates for the former pair are 4.25% and 6.76% respectively, and those values for the latter pair are 20.80% and 30.10%. When compared with Table 6.6, all the NMI and

CSI values of *OC-Ia* and *OC-H* are greater than the corresponding average values of base clustering, and several NMI values are even larger than the maximum values in the base clustering, e.g. Sy2 and Iris. It is also noteworthy that the evaluation scores of the categorical clustering algorithm *LIMBO* are comparable with our proposed *OC-Ia*, but worse than *OC-H*. The reason is that *LIMBO* also considers the coupling between attributes but from the perspective of information theory, and it lacks the concern of the coupling between objects. However, *ROCK* as a categorical clustering algorithm also leads to poor performance in the clustering ensemble, since it only focuses on the interaction between objects but overlooks the relationship between base clusterings. This discovery is also evidenced by the evaluation results quantified by the AC measure. Statistical testing supports the results on AC and NMI that *OC-Ia* and *OC-H* do not perform worse than *CSPA*, *ROCK*, and *LIMBO*, at a 95% significance level.

Thus, the clustering quality can be enhanced by the involvement of both intra-coupling between objects (e.g. OC-Ia) and inter-coupling between objects (e.g. OC-H) with the latter performing slightly better.

Cluster-based Comparison

Table 6.7 reports the experimental results with the cluster-based ensemble methods by using the evaluation measures: AC, NMI and CSI. The two highest measure scores of each experimental setting are highlighted in boldface. The last column is the average value for associated measures across all the data sets. As the table indicates, our proposed *CrC-Ia* and *CrC-C* mostly hold the first two positions on every individual data set, and their average evaluation scores are the corresponding largest two among all the average values. For AC, the average improvement rate of *CrC-Ia* and *CrC-C* against other methods ranges from 1.84% to 6.79%; for NMI, the minimal and maximal average improvement rates are 2.19% and 6.56%, respectively; for CSI, this rate falls between 2.02% and 12.44%. In addition, the average AC, NMI, and CSI scores of *CrC-Ia* and *CrC-C* across all the data sets are larger than

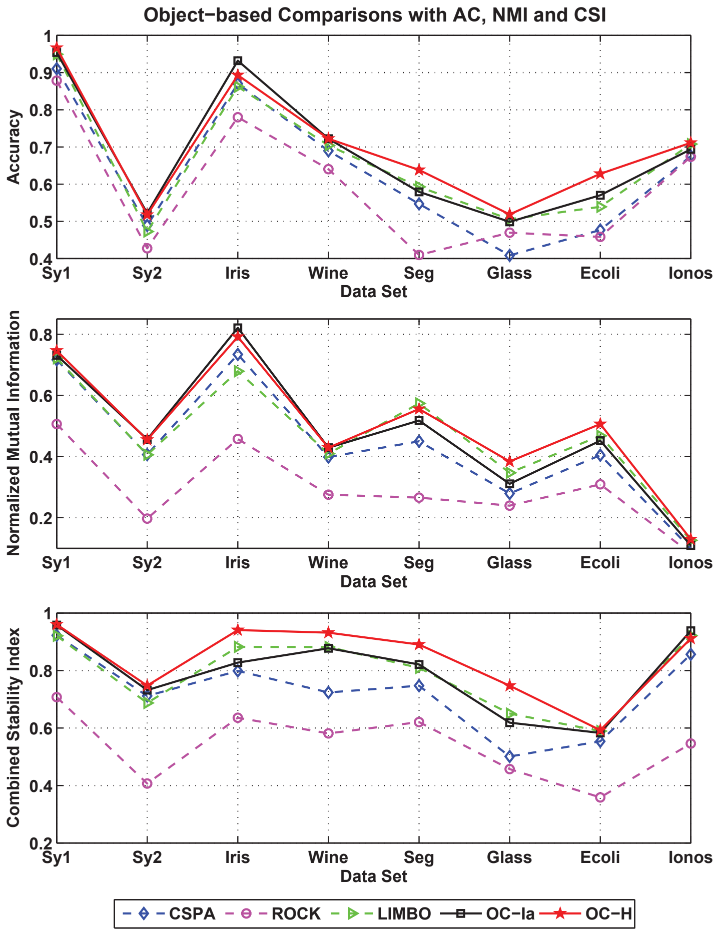


Figure 6.7: Object-based comparisons on AC, NMI and CSI.

those of comparative approaches and are presented in the last column of Table 6.7. Thus, both *CrC-Ia* and *CrC-C* are more robust than other alternatives. Resembling the above comparisons, all the evaluation scores of *CrC-Ia* and *CrC-C* are at least not smaller than the corresponding average values of base clustering, with several AC and NMI values being even greater than the relevant maximal scores in base clustering, e.g., Sy2 and Wine. All the results on AC and NMI are supported by a statistical significance test at a 95% significance level.

Another significant observation is that the average AC and NMI improvement rates of *CrC-C* on *CrC-Ia* are only 1.86% and 1.42% respectively, which are smaller than those of *CrC-Ia* and *CrC-C* on other compared methods. We know that *CrC-C* built on *CrC-Ia* also involves the common neighborhood of objects. When most of the base clusterings have a relatively consistent grouping of objects, the chances of encountering a situation where half of the base clusterings put two objects in the same cluster while the other half separates them into different groups is rare. Therefore, the improvement made by *CrC-C* upon *CrC-Ia* is minor or even negative in this scenario, such as Seg and Yeast whose CSI values across 10 base clusterings are as high as 0.820 and 0.817 in Table 6.6, respectively. However, for a majority of cases, different base clusterings result in a range of results. Thus, *CrC-C* in particular is expected to demonstrate better performance when differentiating those questionable objects, compared to *CrC-Ia*. We will verify this assumption in detail in Section 6.7.4.

Clustering quality consequently benefits from both the couplings between clusterings and the couplings between objects. However, the inter-coupling of objects is dependent on the consistency of base clustering results, which affects the degree of improvement.

Table 6.7: Cluster-based Comparisons on AC, NMI and CSI

Data Set		Sy1	Sy2	Iris	Wine	Seg	Glass	Ecoli	Ionos	Blood	Vowel	Yeast	Avg
AC	*MCLA*	0.945	0.501	0.875	0.702	0.560	0.472	0.528	0.711	0.680	0.365	0.341	0.607
	HBGF	0.949	0.503	0.877	0.690	0.532	0.445	0.468	0.684	0.528	0.379	0.301	0.578
	LB-P	0.952	0.504	0.878	0.703	**0.582**	0.459	0.530	0.711	**0.719**	0.330	0.328	0.609
	LB-S	0.951	0.486	0.844	0.690	0.560	**0.483**	**0.539**	0.711	0.713	0.364	0.332	0.607
	CrC-Ia	**0.954**	**0.513**	**0.893**	**0.731**	**0.579**	0.482	**0.539**	**0.721**	0.713	**0.394**	**0.379**	**0.627**
	CrC-C	**0.969**	**0.518**	**0.902**	**0.764**	**0.579**	**0.511**	**0.587**	**0.742**	**0.723**	**0.430**	**0.378**	**0.646**
NMI	*MCLA*	0.725	0.406	0.744	0.429	0.526	0.318	0.510	0.129	0.015	0.411	0.223	0.403
	HBGF	0.710	0.389	0.706	0.355	0.486	0.316	0.444	0.109	0.007	0.414	0.206	0.377
	LB-P	0.723	0.406	0.745	0.429	**0.548**	0.318	**0.511**	0.130	0.016	0.420	0.221	0.406
	LB-S	0.724	0.363	0.687	0.412	0.531	**0.335**	0.502	0.130	0.015	0.394	0.210	0.391
	CrC-Ia	**0.734**	**0.436**	**0.752**	**0.556**	**0.543**	0.323	**0.511**	**0.164**	**0.018**	**0.445**	**0.226**	**0.428**
	CrC-C	**0.764**	**0.456**	**0.753**	**0.580**	0.540	**0.337**	**0.539**	**0.171**	**0.019**	**0.477**	**0.228**	**0.442**
CSI	*MCLA*	0.950	0.710	0.876	0.828	0.775	0.554	0.640	0.937	**0.897**	0.783	0.774	0.793
	HBGF	0.953	0.703	0.761	0.712	0.716	0.594	0.528	0.839	0.642	0.736	0.742	0.721
	LB-P	0.954	0.713	0.860	0.829	0.840	0.601	**0.673**	0.943	0.893	0.774	0.786	0.806
	LB-S	0.943	0.662	0.787	0.846	0.767	0.601	0.594	0.926	0.892	0.757	0.727	0.773
	CrC-Ia	**0.967**	**0.736**	**0.892**	**0.868**	**0.878**	**0.621**	0.649	**0.955**	**0.897**	**0.808**	**0.817**	**0.826**
	CrC-C	**0.963**	**0.752**	**0.910**	**0.880**	**0.880**	**0.639**	**0.679**	**0.957**	**0.940**	**0.872**	**0.822**	**0.845**

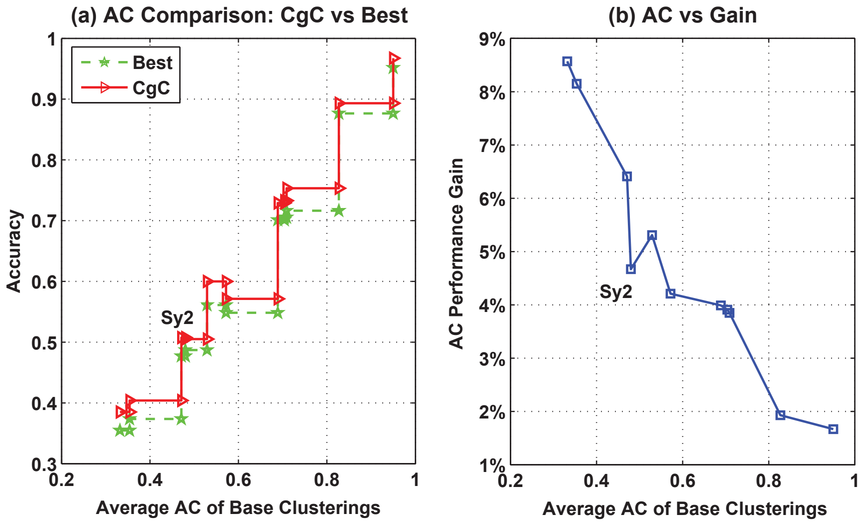


Figure 6.8: Quality of base clusterings and the AC performance gain for the results in Figure 6.6.

6.7.4 Data Characteristics and Performance

Building on the previous quality assessments, here we discuss the data characteristics and performance of our proposed framework *CCE*. Specifically, we address the two assumptions in the previous sections: we aim to discover how the quality of base clusterings affects final clustering accuracy, and how the consistency of base clustering results improves consensus accuracy. Thus, we develop another two groups of experiments to explore the relationship between the data characteristics of base clusterings and the degree of improvement in the final clustering quality.

Quality of Base Clusterings vs Improvement

The first descriptive indicator of data characteristics for base clusterings exhibits the quality of those base clusterings. Here, we use the average AC (i.e. accuracy) or average NMI (i.e. normalized mutual information) of base clusterings generated by *k-means* to represent this indicator to show the quality of base clusterings. In terms of the improvement, the AC performance gain is regarded as the increased proportion of accuracy for *CgC* against the best results among the other three methods (i.e., *EM*, *QMI*, *PD*) considered in Figure 6.6, while the NMI performance gain is described as the increased percentage of NMI for *OC-Ia* against the better results between *CSPA* and *ROCK* compared in Figure 6.7. Note that these ratios are the relative difference value between two evaluation scores, which is different from the improvement rate in Section 6.7.3. Formally, the performance gain is defined as:

$$\text{Performance Gain} = [\tau(*) - \tau(Best)]/\tau(Best), \tag{6.7.1}$$

where τ is either AC or NMI as required, $*$ is the proposed method (e.g. *CgC* or *OC-Ia*), and *Best* is the best comparable algorithm (e.g. $Best \in \{EM, QMI, PD\}$ or $Best \in \{CSPA, ROCK\}$). $\tau(*)$ and $\tau(Best)$ represent the corresponding τ evaluation scores of $*$ and *Best*, respectively.

The results of the relationship between quality and performance gain are reported in Figure 6.8 and Figure 6.9, which correspond to Figure 6.6 and Figure 6.7, respectively. Figure 6.8(a) shows the staircase chart on AC of *CgC* and the best algorithm among *EM*, *QMI* and *PD*. As can be clearly seen from Figure 6.8(b), the larger the average AC of base clusterings (axis x), the smaller the AC performance gain (axis y), for most cases. The only exception is Sy2. This is probably due to the fact that the synthetic data set Sy2 is generated with additional noise, besides which, the Pearson's correlation coefficient between these two variables (i.e. AC of base clusterings and AC performance gain) is -0.9486 with p-value 0.8626×10^{-5} (< 0.05), which means that the correlation is negative at a 95% significance level. We can draw the same conclusion if we consider NMI values.

Similarly, Figure 6.9(a) displays the staircase chart on NMI of *OC-Ia* and the better algorithm between *CSPA* and *ROCK*. Further, Figure 6.9(b) reveals that with the exception of Iris, the larger the average NMI of base clusterings (axis x), the smaller the NMI performance gain (axis y). The great variation of NMI for Iris, which is reflected in Table 6.6 with maximal NMI value 0.750 and minimal NMI value 0.427, probably leads to this exception.

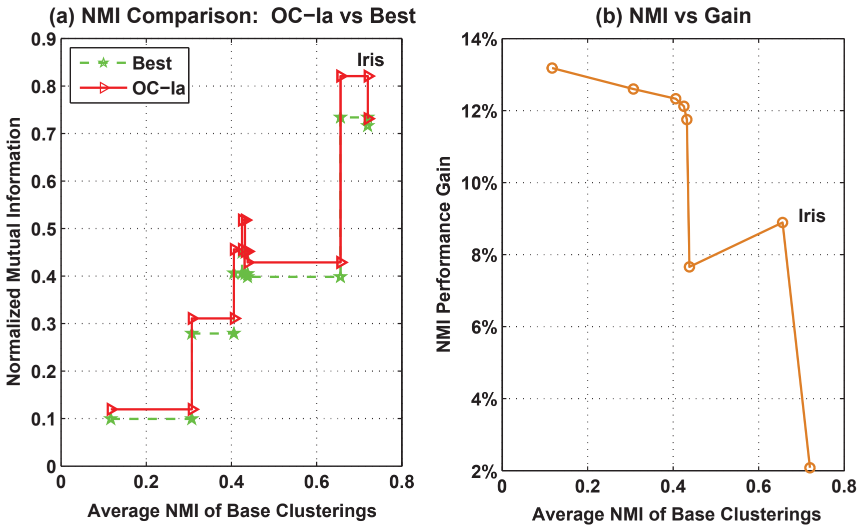


Figure 6.9: Quality of base clusterings and the NMI performance gain for the results in Figure 6.7.

The corresponding Pearson's correlation coefficient here is -0.7953 between two variables: NMI of base clusterings and NMI performance gain, with p-value 0.0183 (< 0.05). It is also revealed that these two variables are significantly associated in anti-correlation at a 95% significance level. Similar results can be obtained when AC scores are concerned instead.

We have therefore verified our first assumption proposed in Section 6.7.3. We conclude that the performance gain brought by the coupling of base clusterings against other ensemble methods is negatively associated with the quality of base clusterings, and the result is statistically significant. Intuitively, this conclusion is easy to understand, since the improvement space will automatically become smaller when the base clusterings have already exhibited better quality.

Consistency of Base Clusterings vs Improvement

The consistency of base clusterings is selected as another descriptive indicator of data characteristics for base clusterings. The consistency here describes

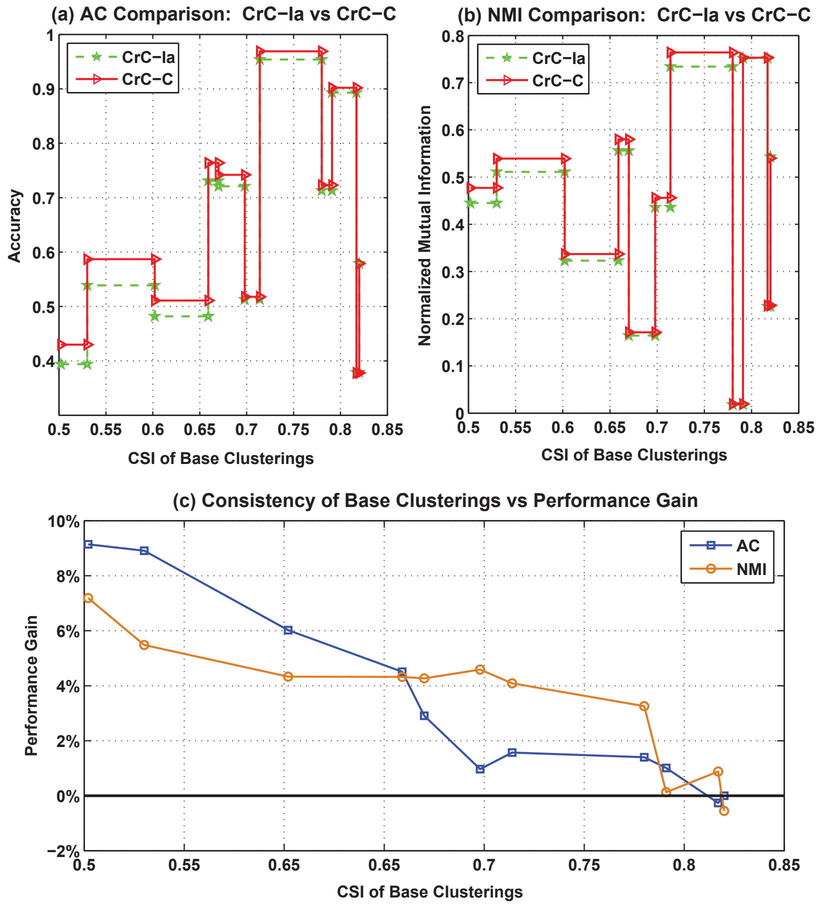


Figure 6.10: Consistency of base clusterings and the AC, NMI performance gains for the results in Table 6.7.

the variation of clustering results among base clusterings. As pointed out in Section 6.7.2, CSI reflects the deviation of clustering results across different runs. Thus, we use the CSI of base clusterings (i.e. the last column in Table 6.6) to represent and quantify the consistency of these results. The larger the CSI, the more consistent the clustering results. Similar to the above Section 6.7.4, AC and NMI performance gains are again adopted to measure the improvement of *CrC-C* upon *CrC-Ia* in Table 6.7. Here, $*$ is *CrC-C* and *Best* is *CrC-Ia* in Equation (6.7.1).

The corresponding results obtained for the dependency between consistency and performance gain are presented in Figure 6.10. In detail, Figure 6.10(a) and Figure 6.10(b) exhibit the staircase charts on AC and NMI of *CrC-C* and *CrC-Ia*, respectively. In Figure 6.10(c), it is clearly observed that both curves, whether they are AC or NMI, have a general tendency to decrease. That is to say, for most cases, the larger the CSI of the base clusterings (axis x), the smaller the AC or NMI performance gain (axis y). This also means that the performance gain of *CrC-C* upon *CrC-Ia* is associated with the consistency of the base clusterings. If the initial base clusterings have more controversial objects for the final grouping, *CrC-C* is more likely to further refine *CrC-Ia* with the inconsistency. Otherwise, *CrC-C* obtains more or less the same clustering results as *CrC-Ia*; sometimes the results of *CrC-C* are even worse than those of *CrC-Ia*. For instance, there are several points located around the horizontal line of 0% in Figure 6.10(c) . Moreover, the Pearson's correlation coefficient between consistency of base clusterings and AC performance gain is -0.9615 with p-value 0 (< 0.05), and the coefficient between consistency and NMI performance gain is -0.8912 with p-value 0.0002 (< 0.05). The statistical test guarantees that the variables of consistency and performance gain are correlated by the negative dependency, significantly with a confidence level at 95%.

The second hypothesis raised in Section 6.7.3 has consequently been confirmed. In conclusion, the performance gain caused by the inter-coupling of objects against other ensemble methods is negatively dependent on the consistency of base clustering results, and this consequence is statistically significant. This conclusion explains that if the initial base clusterings have a relatively high level of inconsistency, a further improvement is necessary by also involving the inter-coupling of objects. This conclusion also conforms to the viewpoint proposed by Kuncheva and Hadjitodorov (Kuncheva & Hadjitodorov 2004) as well as Iam-On (Iam-On et al. 2011): a more accurate partition can be obtained from a diverse ensemble than from the non-diverse case. Here, the diverse ensemble corresponds to the less consistent base clusterings.

In all, we draw the following four conclusions to address the research questions proposed in Section 6.1: 1) Our proposed similarity measures incorporate the couplings of base clusterings and objects, and have an impressive capacity to discover the implicit relationships in the data. 2) Base clusterings are indeed coupled with each other, and the consideration of such couplings can result in better clustering quality; 3) The inclusion of coupling between objects further improves clustering accuracy and stability; 4) The improvement level or performance gain brought by the coupling of base clusterings is negatively associated with the quality of base clusterings, while the further improvement degree or performance gain caused by the inter-coupling of objects is inversely dependent on the consistency of the base clustering results. All the results are accordingly supported by statistical tests.

6.8 Summary

The clustering ensemble has been introduced as a more accurate alternative to individual (base) clustering algorithms. Existing approaches are mostly based on the IIDness assumption, and overlook the couplings between objects. This chapter has proposed a novel framework for coupled clustering ensembles, i.e. *CCE*, to incorporate interactions between base clusterings and between objects. *CCE* caters for cluster label frequency distribution within one base clustering (intra-coupling of clusterings), cluster label co-occurrence dependency between distinct base clusterings (inter-coupling of clusterings), base clustering aggregation between two objects (intra-coupling of objects), and neighborhood relationship among other objects (inter-coupling of objects), which has been shown to improve learning accuracy, stability, and robustness. The proposed similarity measures that involve the couplings of base clusterings and objects have been shown to largely tease out the implicit relationships in the data. Substantial experiments have verified that the consensus functions incorporated with the non-IIDness features significantly outperform ten state-of-art techniques in terms of the clustering-base, object-based and cluster-based ensembles as well as the algorithm to produce base clusterings (*k-means*), supported by statistical analysis.

[***Note***] *A conference version of this chapter has been published in the first item below, and a full journal version has been submitted to the second item.*

- ***Can Wang**, Zhong She, Longbing Cao (2013), "Coupled Clustering Ensemble: Incorporating Coupling Relationships Both between Base Clusterings and Objects". The 29th IEEE International Conference on Data Engineering (**ICDE 2013**), , full paper accepted.*
- ***Can Wang**, Zhong She, Longbing Cao (2013), "Coupled Clustering Ensemble Involving Non-IIDness". Artificial Intelligence (**AI**).*

Chapter 7

Integrated Understanding with Discussions

In this chapter, we summarize the qualitative coupled behaviors and quantitative coupled behaviors in terms of modeling, analysis and learning, and abstract a preliminary behavior algebra, followed by a series of discussions. At first, we provide a consolidated understanding of coupled behaviors. Next, we extract the multi-level couplings embedded in coupled behaviors. After that, we formalize a coupled behavior algebra at its preliminary stage. Finally, our proposed models, systems, measures and frameworks in this thesis are discussed with open research issues.

7.1 Consolidated Understanding

As previously shown in Figure 1.4 in Section 1, the qualitative perspective of coupled behaviors has been addressed in Section 3. The quantitative perspective of coupled behaviors has been studied in Section 4, Section 5 and Section 6. In this part, we aim to give a unified understanding on both qualitative and quantitative coupled behaviors.

Chapter 1 and Chapter 3 clarify that the qualitative coupled behaviors are formalized with three components: actor, operation (or action), coupling.

Similarly, the quantitative coupled behaviors are also characterized by three elements: entity (object or observation or record), property (or attribute or feature), coupling, which has been mentioned in Chapter 1 as well. As a matter of fact, in Chapter 4, the coupled attribute analysis on on numerical data explores the continuous coupled behaviors composed of entity, numerical attribute and coupling. The coupled attribute analysis on categorical data introduced in Chapter 5 focuses on the discrete coupled behaviors, which consist of entity, categorical attribute and coupling. In addition, Chapter 6 works on the coupled clustering ensemble and regards the coupled behaviors made up of entity, base clustering (or method) and coupling.

For either qualitative coupled behaviors or quantitative coupled behaviors, there are accordingly both three elements to formalize them. The actor or entity can be regarded as the body of coupled behaviors.The operation or property can be treated as the depictor of body in coupled behaviors. In fact, qualitative operations, numerical attributes, categorical attributes and base methods are different forms to express depictors that are used to describe the bodies under varied scenarios. The coupling maintains the same as it denotes the relationships or interactions among coupled behaviors. Thus, the coupled behaviors are now characterized by body, depictor and coupling. Similar to the behavior feature matrix proposed in Chapter 3, we here introduce a coupled behavior information table to exhibit our consolidated understanding of coupled behaviors. As shown in Figure 7.1, a large number of behaviors can be organized by a coupled behavior information table $CB = \langle Body, Depictor, Coupling \rangle$. Coupled universe $CU = \{\mathbf{B}_1, \cdots, \mathbf{B}_m\}$ is composed of a nonempty finite set of bodies. Coupled feature $CF = \{\mathbf{D}_1, \cdots, \mathbf{D}_n\}$ is a finite set of depictors, and $\mathcal{V}_{ij}$ is the value of depictor $\mathbf{D}_j$ for body $\mathbf{B}_i$. Coupling is reflected and grouped via the colored curves, which link depictors and depictors by pink curves "1", bodies and bodies by blue curves "2" and orange curves "2.2", depictors and values by green curves "1.2", values and values by red curves "1.1" as well as purple curves "2.1".

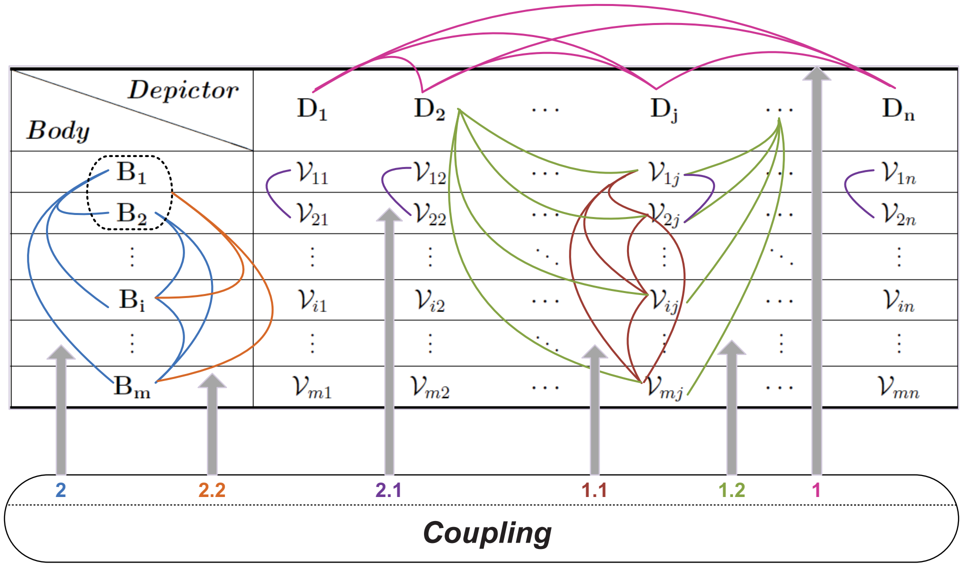


Figure 7.1: A coupled behavior information table.

For instance, regarding the case of multi-robot soccer game studied in Chapter 3, the body $\mathbf{B}_i$ ($1 \leq i \leq 4$) represents each of the four robots participating soccer game, the depictors $\{\mathbf{D}_j\}$ denote the corresponding operations conducted during the whole game such as "retrieve case" and "abort", and the depictor value $\mathcal{V}_{ij}$ refers to whether the designated operation $\mathbf{D}_j$ has been implemented by actor $\mathbf{B}_i$ or not. In addition, the coupling is teased out by the behavior aggregator in its framework, in which the links between depictors (i.e. pink curves "1") are interpreted as transition systems and the connections between bodies (i.e. blue curves "2") are mapped to the temporal , inferential and party-based interactions. In essence, the behavior feature matrix defined in Equation (3.3.1) corresponds to the coupled behavior information table if $\mathscr{O}_{11}$ there is treated as $\mathcal{V}_{ij}$ here in Figure 7.1.

For the toy example of clustering ensemble proposed in Chapter 6, each object is regarded as the body $\mathbf{B}_i$ ($1 \leq i \leq 12$), every base clustering is represented as the depictor $\mathbf{D}_j$, and the depictor value $\mathcal{V}_{ij}$ is quantified as the label that body $\mathbf{B}_i$ has been assigned in depictor $\mathbf{D}_j$. Additionally, the coupling is disclosed from two perspectives, i.e. coupling of clusterings and coupling of objects, which respectively correspond to the links between depictors (i.e. pink curves “1”) and the connections between bodies (i.e. blue curves “2”). Further, the links between depictor values (i.e. red curves “1.1”) of a depictor specify the intra-coupling within this depictor (e.g. $\mathbf{D}_j$), while the connections between depictors and the values of other depictors (i.e. green curves “1.2”) describe the inter-coupling between different depictors (e.g. $\mathbf{D}_2$ and $\mathbf{D}_j$). Apart from this, the purple curves “2.1” connecting depictor values of two bodies exhibit the intra-coupling of two bodies (e.g. $\mathbf{B}_1$ and $\mathbf{B}_2$), whereas the orange curves “2.2” linking a pair of bodies with others expose the inter-coupling of these two bodies.

7.2 Multi-level Couplings

In this section, all the couplings examined in this thesis are synthesized and organized in terms of a hierarchical structure with multi-levels. Here, coupling has much broader meaning than the mostly used “dependency”, which mainly refers to a condition in the statistical sense or to a kind of relationship in the ontological way.

The coupling nature may be embedded in different layers and distinct aspects of an underlying problem. Depending on the focuses and objectives, the couplings may involve one to multiple layers: value, depictor, object, and method which refers to the learning outcomes in a broad sense. The hierarchical structure and relevancy between these aspects are displayed in Figure 7.2 from the very fundamental level (i.e. value) to the output of learning (i.e. method). Below, we briefly summarize and overview the various couplings on these levels from multi-layer, multi-framework and multi-relation aspects. Note that all the denotations and curves mentioned below follow the explanations of Figure 7.1.

- **Value:** For a depictor $\mathbf{D}_j$, there is dependence between its values $\mathcal{V}_{i_1 j}$ and $\mathcal{V}_{i_2 j}$, which is ***value-value relation***. For example, values $\mathcal{V}_{1j}$ to $\mathcal{V}_{m-1,j}$ of depictor $\mathbf{D}_j$ more or less influence its value $\mathcal{V}_{mj}$, shown

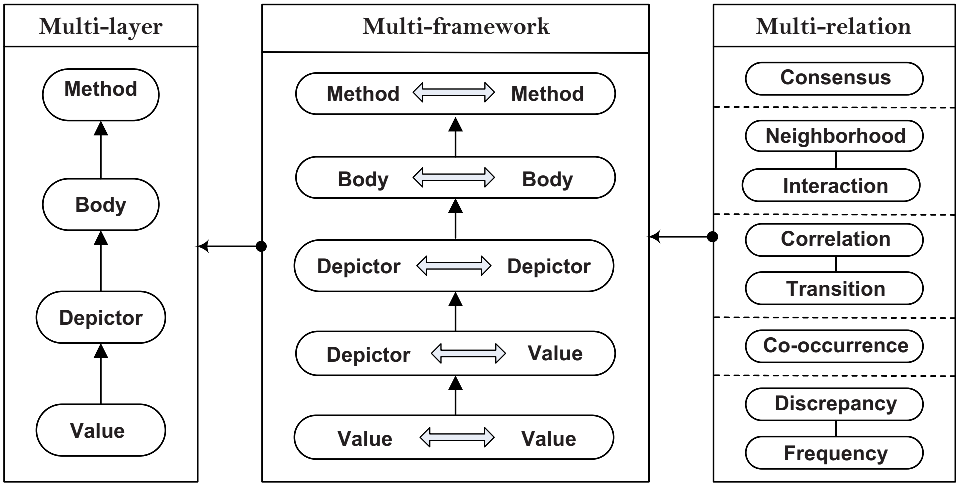


Figure 7.2: Hierarchical structure of couplings.

in Figure 7.1. In this thesis, the frequency count is used to quantify the intra-coupling of categorical attributes in Chapter 5, described by the red curves “1.1”. Alternatively, the relative discrepancy between attribute values is also adopted to build the intra-coupling of objects in Chapter 6, denoted by the purple curves “2.1”.

– **Depictor:** In terms of ***depictor-value relation***, a depictor value $\mathcal{V}_{ij}$ of $\mathbf{D}_j$ is coupled with other depictors $\mathbf{D}_k$ $(k \neq j)$. That is to say, values $\mathcal{V}_{1j}$ to $\mathcal{V}_{mj}$ of depictor $\mathbf{D}_j$ are somehow affected by depictors $\mathbf{D}_1, \cdots, \mathbf{D}_{j-1}, \mathbf{D}_{j+1}, \cdots, \mathbf{D}_n$. In Chapter 5, the co-occurrence of attribute values is proposed to define the inter-coupling of categorical attributes, which corresponds to green curves “1.2”. Note that the intra-coupling of categorical attributes is reflected by the value-value relation above.

With respect to ***depictor-depictor relation***, every pair of depictors $\mathbf{D}_j$ and $\mathbf{D}_k$ are related with each other, visualized as pink curves “1”. These curves are embodied by the intra-coupling of qualitative behaviors via transition systems in Chapter 3, and also quantified as correlations between attributes and their powers to define the intra-coupling and inter-coupling of numerical attributes in Chapter 4.

- **Body:** A body $\mathbf{B}_i$ has interactive coupling relationships with other objects $\mathbf{B}_l$ ($l \neq i$), which is ***object-object relation*** in blue curves "2" and orange curves "2.2". For instance, $\mathbf{B}_1$ is linked by bodies $\mathbf{B}_2, \cdots, \mathbf{B}_m$, displayed in Figure 7.1. In Chapter 3, different interactive schemes ranging from temporal, inferential to party-based aspects are introduced to characterize the inter-coupling of qualitative behaviors. In Chapter 6, the common neighborhood rate (e.g., based on common neighbors of $\mathbf{B}_1$ and $\mathbf{B}_2$) is proposed to define the inter-coupling of objects during the clustering ensemble process, while the intra-coupling of objects is built by the value-value relation above.

- **Method:** A proper method delivers the best choice of learning study for the underlying problem. The generated evaluation scores determine the learning outcome (e.g. a label to cluster an object) of a given method. In multi-method-based learning such as ensemble clustering, there may be coupling between constituent methods (say base clusterings). Each method can be regarded as a depictor and the corresponding results are treated as the depictor values. Thus, the depictor-value relation (green curves "1.2") and depictor-depictor relation (pink curves "1") are directly applied to describe method-result relation and method-method relation. The clustering ensemble studied in Chapter 6 illustrates how the ensemble consensus among different base clustering methods is explored via the method-result relation and the method-method relation.

For the above aspects and levels, we use the word "coupling" to refer to any forms of relations between the underlying components: value, depictor, object, method. In practice, couplings may exhibit in a diversity of forms, styles and formats, such as temporal relation, inferential relation, party-based relation, numeric correlation, categorical frequency and co-occurrence relations, neighborhood relation, which have been emphasized in corresponding levels and frameworks.

It is also remarkable to note that there exists strong hierarchical dependency between the above levels from input values to output patterns, as they are constituents linked with one another in a learning system. In fact, the coupling of depictors are built based on the coupling of values. The coupling of bodies is reflected via the coupling of values and the coupling of depictors. The coupling of methods is further embodied by the coupling of depictors and the coupling of bodies. Thus, we rename the coupling as the consolidated coupling to reflect this hierarchy as well as the diversity mentioned above.

7.3 Preliminary Coupled Behavior Algebra

We have unified a coherent understanding of qualitative coupled behaviors and quantitative coupled behaviors, and teased out a multi-level structure of couplings therein. In the following, a coupled behavior algebra is then formalized, thought at its preliminary stage. The semantic meaning of each component is illustrated with the models and frameworks proposed in this thesis from different perspectives. Note that the following definitions are the general extensions of the definitions proposed in Section 3.3.2 (i.e. Behavior Formal Descriptor) of Chapter 3.

As mentioned in Section 7.1, a coupled behavior information table consists of body, depictor and coupling. In the subsequent Section 7.2, we present a hierarchical structure of multi-level couplings, which can be regarded as the context of coupled behaviors describing what kinds of interactive relationships are involved. The context here includes value-value coupling (V-V), depictor-value coupling (D-V), depictor-depictor coupling(D-D), body-body coupling (B-B) and method-method coupling (M-M). For instance, the context specifies that the coupling in consideration is body-body interaction or depictor-value relationship or both of them. In addition, the coupling has been renamed as the consolidated coupling to show the diverse and hierarchical interactions. Therefore, the integrated coupled behaviors are characterized by body $\mathbf{B}$, depictor $\mathbf{D}$, consolidated coupling $\mathbf{C}$ and context $\mathbf{T}$.

Definition 7.3.1 (Coupled behaviors) ***Coupled behaviors*** $\mathbb{CB}$ *are described as a four-ingredient tuple* $\mathbb{CB} = \langle \mathbf{B}, \mathbf{D}, \mathbf{C}, \mathbf{T} \rangle$*, specifically:*

- *Body* $\mathbf{B}$ *is the actor or entity that issues a behavior or on which a behavior is imposed.*
- *Depictor* $\mathbf{D}$ *is what an actor conducts in order to achieve certain goals as well as the associated properties or attributes.*
- *Consolidated Coupling* $\mathbf{C} = (\theta(\cdot), \eta(\cdot))$ *reveals complex relationships within a depictor or a body (i.e. consolidated intra-coupling* $\theta(\cdot)$*) and those between multiple depictors or bodies (i.e. consolidated inter-coupling* $\eta(\cdot)$*).*
- *Context* $\mathbf{T}$ *specifies what sorts of underlying relationships are under investigation, and* $\mathbf{T}$ *can be selected from but not limited to* {*V-V, D-V, D-D, B-B, M-M*}.

Suppose there are m bodies $\{\mathbf{B}_1, \mathbf{B}_2, \cdots, \mathbf{B}_m\}$ and in total n depictors $\{\mathbf{D}_1, \mathbf{D}_2, \cdots, \mathbf{D}_n\}$ involved. The value of depictor $\mathbf{D}_j$ for body $\mathbf{B}_i$ is $\mathcal{V}_{ij}$. These denotations are all consistent with Figure 7.1.

As clarified in the above section, the coupling of values and the coupling of methods are highly relevant to the coupling of depictors and the coupling of bodies. Therefore, intra-coupling and inter-coupling of depictors and bodies are formally defined below, and they are essentially derived based on the values $\mathcal{V}_{ij}$ $(1 \leq i \leq m, 1 \leq j \leq n)$. Intuitively, the intra-couplings of depictors and bodies specify the internal relationships within those depictors and bodies respectively, while the inter-coupling of depictors and bodies explicate the respective external interactions with other depictors and bodies. We start with the intra-coupling of depictors, then the inter-coupling of depictors, followed by the intra-coupling of bodies, and finally the inter-coupling of bodies. Specifically, we have:

Definition 7.3.2 (Intra-coupling of Depictors) *A depictor $\mathbf{D}_j$ is intra-coupled with itself in terms of intra-coupling functions $\theta_i^D(\cdot)$,*

$$\mathbf{D}_j^\theta ::= \mathbb{CB}_{\cdot j}(\mathbf{B}, \mathbf{D}, \mathbf{C}, \mathbf{T}) | \sum_{i=1}^{m} \theta_i^D(\cdot) \oplus_a^D \mathcal{V}_{ij}, \tag{7.3.1}$$

$$|\theta_i^D(\cdot)| \geq \theta_0^D, \tag{7.3.2}$$

where θ_0^D is the depictor intra-coupling threshold, $\sum_{i=1}^m \oplus_a^D$ means the aggregation of depictor values $\mathcal{V}_{ij}$ intra-coupled with $\theta_i^D(\cdot)$.

Corollary 7.3.1 *If $\theta_i^D(\cdot) < 0$, $\mathbf{D}_j^\theta$ has negative intra-coupling; if $\theta_i^D(\cdot) > 0$ then there is a positive intra-coupling relationship; $\theta_i^D(\cdot) = 0$ indicates none intra-coupling exists.*

Definition 7.3.3 (Inter-coupling of Depictors) *A depictor $\mathbf{D}_j$ is inter-coupled with other depictors in terms of inter-coupling functions $\eta_k^D(\cdot)$,*

$$\mathbf{D}_j^\eta ::= \mathbb{CB}_{i\cdot}(\mathbf{B}, \mathbf{D}, \mathbf{C}, \mathbf{T}) | \sum_{k=1, k\neq j}^{n} \eta_k^D(\cdot) \oplus_e^D \mathcal{V}_{ij}, \tag{7.3.3}$$

$$|\eta_k^D(\cdot)| \geq \eta_0^D, \tag{7.3.4}$$

where η_0^D is the depictor inter-coupling threshold, $\sum_{k=1,k\neq j}^n \oplus_e^D$ means the aggregation of depictor values $\mathcal{V}_{ij}$ inter-coupled with $\eta_k^D(\cdot)$.

Corollary 7.3.2 *If $\eta_k^D(\cdot) < 0$, $\mathbf{D}_j^\eta$ has negative inter-coupling; if $\eta_k^D(\cdot) > 0$ then there is a positive inter-coupling relationship; $\eta_k^D(\cdot) = 0$ indicates none inter-coupling exists.*

Definition 7.3.4 (Intra-coupling of Bodies) *A body $\mathbf{B}_i$ is intra-coupled with itself in terms of intra-coupling functions $\theta_j^B(\cdot)$,*

$$\mathbf{B}_i^\theta ::= \mathbb{CB}_{i\cdot}(\mathbf{B}, \mathbf{D}, \mathbf{C}, \mathbf{T}) | \sum_{j=1}^{n} \theta_j^B(\cdot) \oplus_a^B \mathcal{V}_{ij}, \tag{7.3.5}$$

$$|\theta_j^B(\cdot)| \geq \theta_0^B, \tag{7.3.6}$$

where θ_0^B is the body intra-coupling threshold, $\sum_{j=1}^n \oplus_a^B$ means the aggregation of depictor values $\mathcal{V}_{ij}$ intra-coupled with $\theta_j^B(\cdot)$.

Corollary 7.3.3 *If $\theta_j^B(\cdot) < 0$, $\mathbf{B}_i^\theta$ has negative intra-coupling; if $\theta_j^B(\cdot) > 0$ then there is a positive intra-coupling relationship; $\theta_j^B(\cdot) = 0$ indicates none intra-coupling exists.*

Definition 7.3.5 (Inter-coupling of Bodies) *A body $\mathbf{B}_i$ is inter-coupled with other bodies in terms of inter-coupling functions $\eta_l^B(\cdot)$,*

$$\mathbf{B}_i^\eta ::= \mathbb{CB}_{\cdot j}(\mathbf{B}, \mathbf{D}, \mathbf{C}, \mathbf{T})| \sum_{l=1, l \neq i}^{m} \eta_l^B(\cdot) \oplus_e^B \mathcal{V}_{ij}, \tag{7.3.7}$$

$$|\eta_l^B(\cdot)| \geq \eta_0^B, \tag{7.3.8}$$

where η_0^B is the body inter-coupling threshold, $\sum_{l=1,l\neq i}^m \oplus_e^B$ means the aggregation of depictor values $\mathcal{V}_{ij}$ inter-coupled with $\eta_l^B(\cdot)$.

Corollary 7.3.4 *If $\eta_l^B(\cdot) < 0$, $\mathbf{B}_i^\eta$ has negative inter-coupling; if $\eta_l^B(\cdot) > 0$ then there is a positive inter-coupling relationship; $\eta_l^B(\cdot) = 0$ indicates none inter-coupling exists.*

Below, by integrating the definitions on intra-coupling of depictors, inter-coupling of depictors, intra-coupling of bodies and inter-coupling of bodies with Definition 7.3.1 on the consolidated coupled behaviors, we obtain that

$$\begin{aligned} \mathbb{CB} = (\mathbf{D}_j^\theta)^\eta * (\mathbf{B}_i^\theta)^\eta ::= \mathbb{CB}_{ij}(\mathbf{B}, \mathbf{D}, \mathbf{C}, \mathbf{T})| \\ \sum_{i=1}^{m} \sum_{j=1}^{n} f(\theta_i^D(\cdot), \eta_j^D(\cdot), \theta_j^B(\cdot), \eta_i^B(\cdot)) \oplus_c^{DB} \mathcal{V}_{ij}, \end{aligned} \tag{7.3.9}$$

$$|\theta_i^D(\cdot)| \geq \theta_0^D, |\eta_k^D(\cdot)| \geq \eta_0^D, |\theta_i^B(\cdot)| \geq \theta_0^B, |\eta_l^B(\cdot)| \geq \eta_0^B, \tag{7.3.10}$$

where $f(\cdot)$ represents the way to composite the consolidated intra-couplings and inter-couplings, $\sum_{i=1}^m \sum_{j=1}^n \oplus_c^{DB}$ denotes the aggregation of depictor values $\mathcal{V}_{ij}$ intra-coupled with $\theta_i^D(\cdot)$ and $\theta_j^B(\cdot)$ as well as inter-coupled with $\eta_j^D(\cdot)$ and $\eta_i^B(\cdot)$. To only emphasize the different types of couplings, in the following, we use the denotations $\theta^D(\cdot), \eta^D(\cdot), \theta^B(\cdot), \eta^B(\cdot)$ for simplicity.

In practice and real life, it is not easy to identify the consolidated intra-coupling function $\theta(\cdot)$ and the consolidated inter-coupling function $\eta(\cdot)$ as well as the composite function $f(\cdot)$. Those coupling relationships are often implicit to observe and embedded deeply in the social applications and business problems. However, as we have emphasized throughout the whole thesis, coupled behaviors play a much more fundamental role than individuals in the cause, dynamics and effect of business, social, organizational and behavioral systems and domains.

One of the key aspects in this thesis is to represent the coupled behaviors, which is also called coupled behavior modeling. In terms of this task, we have defined $\theta(\cdot), \eta(\cdot), f(\cdot)$ and given various semantic explanations to them according to different scenario and problems, including multi-agent system in Chapter 3, numerical data analysis in Chapter 4, categorical data analysis in Chapter 5, and clustering ensemble learning in Chapter 6. For instance, in Chapter 3, $\eta^D(\cdot)$ and $\theta^B(\cdot)$ are interpreted by a transition system and $\theta^D(\cdot)$ and $\eta^B(\cdot)$ are used to characterize different interactive schemes for operations, $f(\cdot)$ is then accordingly obtained by a concurrent transition system conversion. The corresponding context $\mathbf{T}$ examines depictor-depictor coupling (D-D) and body-body coupling (B-B). Unlike the multi-agent system, in the process of coupled clustering ensemble, Chapter 6 defines $\theta^D(\cdot)$ as the frequency account for each depictor value, $\eta^D(\cdot)$ as the co-occurrence probability of values from different depictors, $\theta^B(\cdot)$ as the relative discrepancy between depict values, $\eta^B(\cdot)$ as the common neighborhood of bodies, and $f(\cdot)$ as the whole framework to embody both the coupling of depictors (i.e. base clusterings) and the coupling of bodies (i.e. objects). The context $\mathbf{T}$ here includes value-value coupling (V-V), depictor-value coupling (D-V), body-body coupling (B-B) and method-method coupling (M-M).

In addition, the other two key respects are to analyze and learn the coupled behaviors. Formally, we regard these two tasks as the following theorem.

Theorem 7.3.1 (Coupled Behavior Analysis and Learning) *The analysis of coupled behaviors is to build the objective function $g(\cdot)$ under the prerequisite that behaviors are coupled with each other by coupling function $f(\cdot)$,* *and satisfy the following conditions.*

$$f(\cdot) ::= f(\theta^D(\cdot), \eta^D(\cdot), \theta^B(\cdot), \eta^B(\cdot)), and \tag{7.3.11}$$

$$g(\cdot)|(f(\cdot) \geq f_0) \geq g_0, \tag{7.3.12}$$

where g_0 is the given threshold or requirement. The learning of coupled behaviors is to extract the hidden patterns or regularities that optimize the objective function $g(\cdot)$.

In Chapter 3, the function $g(\cdot)$ represents the constraints to be checked and g_0 denotes the desired standard. Coupled behavior analysis in this scenario mainly focuses on the verification of potential constraints, if any problem arises, the model itself will be refined according to the checking results.

In Chapter 4 and Chapter 5, the function $g(\cdot)$ stands for the similarity or distance functions depending on what tasks it aims at, such as clustering and classification, g_0 denotes the domain-driven knowledge for such objectives. Coupled behavior learning in these two chapters is to identify how to group bodies and how to classifier a new body to an existing class.

In Chapter 6, the consensus among different base clustering results is quantified by the objective function $g(\cdot)$, and the performance of base clusterings is reflected by g_0. Coupled behavior learning corresponds directly to the clustering ensemble learning with the aim at producing a superior clustering result compared to the base methods.

7.4 Discussions with Open Issues

In this section, our proposed coupled behavior informatics is discussed from the qualitative, quantitative and integrated perspectives below. Many relevant research issues are accordingly raised and analyzed.

7.4.1 Qualitative Coupled Behaviors

In Chapter 3, we have designed mechanisms for modeling and checking group behaviors in terms of intra-coupled and inter-coupled interactions. We understand the proposed Ontology-based Qualitative Coupled Behavior Modeling and Checking (*OntoB*) system can be expanded to a comprehensive tool for modeling and verifying complex behavior interactions. In this section, we discuss open issues related to *OntoB* from two aspects. The depth extension focuses on the enhancement of the behavior representation and verification, while the breadth extension emphasizes the wide and a range of applicabilities of our framework on complex behavior interactions.

Depth Extension: In our proposed framework, we model a behavior as a triple tuple, and then take advantage of model checking to verify the behavior model. In this system, we limit our work within the range of behavior representation and verification to ensure a stable and robust model. However, we could conduct further reasoning about behaviors to infer other properties of the group behaviors towards a behavior algebra. We consider the categorical communications from temporal, inferential, and party-based perspectives based on the correlations, collaborations and competitions of the same or distinct actors. Behaviors in such coupling relationships may interact in different modes, e.g. peer-to-peer, master-slave, underlying-derivative, and contrast modes (Cao et al. 2012). These interaction modes are helpful for understanding the behavior interactions between behavior sequences once coupling relationships are determined.

Breadth Extension: For the purpose of verification, intra-coupling relationship is interpreted as a transition system, while inter-coupled aspect is explained as inter-coupling operators induced by different actor communication categories. Alternatively, the behavior syntactic framework composed of intra-coupling and inter-coupling can be given different semantics to enable quantitative tasks such as fraud detection (Cao et al. 2012), relational learning (Getoor & Taskar 2007), and machine learning. With regard to the behavior feature matrix (3.3.1), for instance, we may regard the inter-coupled interaction $\eta_i(\mathbb{B})$ as the similarity or distance between every two operations $\mathscr{O}_{i_1 j}$ conducted by actor $\mathscr{A}_{i_1}$ and $\mathscr{O}_{i_2 j}$ performed by actor $\mathscr{A}_{i_2}$; and the relevant intra-coupled interaction $\theta_j(\mathbb{B})$ can be reflected as the co-occurrence couplings between different features of the same actor. Subsequently, both interactions are aggregated based on combined distance for clustering analysis or geometric proximity for anomaly detection. Currently, we are working on this for complex behavior analysis and its application. We have also applied this framework to analyze the coupled nominal similarities, including intra-coupled and inter-coupled similarities, to enhance unsupervised learning accuracy (Wang et al. 2011).

7.4.2 Quantitative Coupled Behaviors

In Chapter 5, we have introduced several coupled nominal similarity measures: Coupled Attribute Similarity for Values ($CASV$), Coupled Attribute Similarity for Objects ($CASO$) and Coupled Attribute Dissimilarity for Objects ($CADO$). In Chapter 6, we have presented a framework for coupled clustering ensembles (CCE). In the following, those coupled similarity measures are further analyzed, and the CCE framework is deeply explored.

Coupled Nominal Similarity

In this part, we discuss the potential opportunities triggered by our proposed $CASV$ (see Equation (5.4.10)), $CASO$ (see Equation (5.6.1)) and $CADO$ (see Equation (5.7.2)) from two aspects. The degenerative aspect discusses the degeneration of $CADO$ and $CASV$ with special cases, while the extended aspect focuses on the direct extension of $CASO$ and $CADO$.

Degenerative Aspect: Many existing similarity measures for attribute values are special cases of our proposed $CADO$ or $CASV$. On one hand, $CADO$ could degenerate as an intra-attribute-independence measure if frequency functions $G_j(\{v_j^x\}), G_j(\{v_j^y\})$ take a nonzero constant value ξ. In this way, the dissimilarity measure ADD between v_j^x and v_j^y proposed by Ahmad and Dey (Ahmad & Dey 2007) is exactly $\xi/2 \cdot CADO$, which considers the interactions between attributes but lacks the couplings within each attribute. On the other hand, an inter-attribute-independence measure could be produced by considering $\delta_j^{Ie}(v_j^x, v_j^y, \{V_k\}_{k=j})$ for $IeASV$, in which $\delta_{j|j}^{I}(v_j^x, v_j^y, V_j)$

replaces $\delta^{I}_{j|k}(v^x_j, v^y_j, V_k)$ $(k \neq j)$ for *IRSI*. Such an example is the improved *SMD* with frequency (Gan et al. 2007). Moreover, an intra-inter-attribute-independence measure could be obtained by specializing $g_j(v^x_j) = g_j(v^y_j) = \xi$ and $\delta^{Ie}_j(v^x_j, v^y_j, \{V_k\}_{k=j})$ both, which corresponds to the classical similarity measure *SMS* and its variants such as Jaccard coefficients (Gan et al. 2007). Therefore, our proposed measures have the capability of generalization on the existing similarity measures which assume independence and partial dependence among attributes.

Extended Aspect: The couplings or relationships between attribute values, attributes, objects, and even clusters should be considered to cater for the interactions among the data. We may naturally induce a range of coupled tasks in data mining and machine learning, such as data discretization and clustering ensemble. We have already proposed a coupled discretization algorithm *CD* (Wang, Wang, She & Cao 2012), which concerns both the information attribute dependency and deterministic attribute relationship to disclose the couplings of uncertainty and certainty degree. We have also developed a coupled framework for clustering ensembles by considering both the relationships within each base clustering and the interactions between distinct base clusterings, in which *CASO* or *CADO* could be applied. In addition, how to appropriately choose the weights α_k for $IeASV$ defined in Equation (5.4.9), rather than simply treating them as equal, is in great need of further exploration. Further, we are also working on a flexible way to control the respective importance of $IaASV$ and $IeASV$ by using corresponding weights β and γ, according to the specific data structure. The other data mining and machine learning tasks, e.g. fraud detection (Cao et al. 2012) and relational learning (Getoor & Taskar 2007), can also be considered to involve coupled interactions.

CCE **Framework**

We discuss the potential and future opportunities related to our proposed framework *CCE* below from two aspects. The depth aspect discusses the extension of current definitions on coupling, while the width aspect explores other approaches apart from the consensus function based clustering ensemble and other stages in the process of the clustering ensemble.

Depth Aspect: According to the conclusion in Section 6.7.4, the improvement on clustering performance is largely dependent on the data characteristics of base clusterings, which are quantified as the quality and the consistency of base clusterings. Hence, we need to consider these two descriptive indicators in the coupled framework of clustering ensembles.

In our implementation, we regard the weight λ_k of base clustering bc_k in Definition 6.4.3 as the same, i.e., $\lambda_k = 1/L$, where L is the number of base clusterings. However, the clustering quality (e.g., AC and NMI) of each base clustering, denoted as q_k, can be adapted to substitute λ_k to differentiate the contributions made by distinct base clusterings. Here, we normalize q_k by $q'_k = q_k / \sum_{k=1}^{L} q_k$ to make $\sum_{k=1}^{L} q'_k = 1$ to satisfy the requirement for λ_k in Definition 6.4.3. In this way, base clustering that performs better will contribute more in the calculation of similarity between cluster labels. Therefore, we can incorporate the indicator q'_k on the quality of base clusterings into our framework.

In Definition 6.4.8, the ratio of the number of common neighbors is used to measure the similarity between two objects. However, the consistency $\mu \in [0, 1]$ (e.g., CIS) of all the base clusterings can be utilized to control the extent to which include the inter-coupling of objects. The purpose here is to adjust the effect of inter-coupled objects according to the consistency of base clusterings. We can then alternatively replace $\delta^{CO}(u_x, u_y|U)$ with $\mu \cdot \delta^{IaO}(u_x, u_y) + (1 - \mu) \cdot \delta^{CO}(u_x, u_y|U)$. Thus, the inter-coupling of objects will be less emphasized when the base clusterings obtain more approximate results. If all the base clusterings ideally perform the same (i.e., $\mu = 1$), the similarity between objects degenerates to δ^{IaO}. Consequently, we can involve the indicator μ on the consistency of base clusterings into our framework.

These two descriptive indicators adapt our framework *CCE* to a soft version *S-CCE*, since *S-CCE* considers how much contribution each base clustering makes and to what extent we involve the inter-coupling of objects. However, the previous indicator requires the label information at the stage of generating base clusterings. If this information is unavailable during the whole process of clustering ensemble, only the second indicator can be used.

Width Aspect: As mentioned in Section 2.4, there are three ways to aggregate the base clusterings: consensus functions, categorical clusterings, and direct optimizations. We mainly focus on the consensus function based clustering ensemble to propose a coupled framework *CCE*. For the second option on categorical clusterings, we have designed the coupled nominal similarity in unsupervised learning (Wang et al. 2011), which induces alternatives to cluster categorical data and also forms a part of our framework *CCE*, besides which, we have already involved the widely used categorical clustering algorithms (i.e., *ROCK* (Guha et al. 2000) and *LIMBO* (Andritsos et al. 2004)) in our experiments. The third group of methods on direct optimizations selects candidates among all the clusters produced by base clusterings and then adjusts them to achieve the minimal cost (Christou 2011). They totally ignore the consensus of base clusterings, and do not rely on the similarity or distance between base clusterings, objects, and clusters. Thus, our proposed *CCE* does not fit the direct optimizations based clustering ensemble. In addition, the direct optimizations based approaches require the detailed information of each object (i.e. the attribute values) to obtain the sum of intra-cluster distance as the cost of each cluster, while our framework *CCE* still works well despite the lack of such prior information and enables the privacy-preserving and distributed mode of data analysis.

Also as introduced in Section 2.4, the whole process of the clustering ensemble is composed of three stages: building base clustering, aggregating base clusterings, and post-processing clustering. In this research, our proposed framework *CCE* is constructed for the second stage, and the first and last stages are fixed as in comparative methods. In reality, base clusterings and post-processing techniques are also shown to affect the performance of clustering ensemble (Iam-On et al. 2011). In our experiments, *k-means* on random sub-sampling with a fixed k is adopted to build base clusterings, and homogeneous results are accordingly obtained. Alternatively, different values of k can be selected, and distinct approaches are also expected to generate heterogeneous base clusterings. The input base clusterings then exhibit a higher level of diversity than those we have used. Note that the consistency pointed out in Section 6.7.4 is only one aspect of diversity among base clusterings. At the post-processing stage, three fundamental clustering algorithms are employed, namely, *k-means*, *agglomerative algorithm*, and *METIS*. However, advanced similarity or distance based clustering algorithms, such as spectral clustering (Luxburg 2007) and affinity propagation (Frey & Dueck 2007), can be applied to further improve the quality of the clustering ensemble. In future studies, therefore, we will also examine the heterogeneous structure of base clusterings and advanced post-processing clustering algorithms in our proposed framework *CCE* to enhance the performance of the whole process.

7.4.3 Integrated Coupled Behaviors

Based on the qualitative and quantitative understandings of coupled behaviors, we have proposed a consolidated concept for coupled behaviors. The coupled behavior representation, analysis and learning act as the basic components to build the coupled behavior algebra. However, the above definitions, corollaries and theorem in Section 7.3 only constitute a preliminary roadmap. There are a lot of opportunities for us to widely explore and strictly define this coupled behavior algebra. Many open issues are worth widely addressing and systematically investigating. These interesting research points include but are not limited to:

- More operators other than $\oplus$, e.g. the adapted scalar multiplication, are to be identified and explored to test whether classic properties such as associativity, commutativity, linearity, continuity and boundedness are satisfied or not.

- The context of coupled behaviors is to be formalized to control the whole process of coupled behavior modeling, analysis and learning according to different requirements. More types of consolidated coupling are to be explored and studied, and the soft computing techniques can be adopted to propose the fuzzy or rough coupled behavior informatics.

- A coupled behavior space is to be structured based on a collection of coupled behavior vectors (e.g. each row of the coupled behavior information table shown in Figure 7.1) and appropriate operators under certain axioms. Several properties, including topology, completeness, duality and best approximation, are to be discussed.

- Analytical problems like the convergence or divergence of coupled behavior vectors is to be defined and intensively studied. Limits of converged coupled behavior sequences are to be clarified, which are essential to calculus and can be used to define continuity, derivatives and integrals.

- Some other research issues, including how to define the bases and dimension of such a coupled behavior space, how to do the space decomposition, how to conduct the linear and nonlinear transformations, are to be addressed and deeply explored.

- Mix-type coupled behaviors are to be modeled, analyzed and learnt. The mix-type here means that the depictors of coupled behaviors can be any combination of qualitative actions, numerical properties and categorical properties, which forms a heterogeneous structure of the coupled behavior informatics.

Chapter 8

Conclusion and Future Work

8.1 Conclusion

Complex behaviors are widely seen on the internet, social and online networks, multi-agent systems, and brain systems. The in-depth understanding of complex behaviors has been increasingly recognized as a crucial means for disclosing interior driving forces, causes and impact on businesses in handling many challenging issues. This forms the need and emergence of behavior informatics, i.e. understand behaviors from computing perspective.

Current behavior modeling methods are however designed rather different from one another according to various backgrounds and scenarios. Moreover, traditional behavior modeling strategies mainly rely on qualitative methods from behavioral science and social science perspectives. The so-called behavior analysis often focuses on human demographic and business usage data, in which behavior-oriented elements are hidden in routinely collected transactional data. As a result, it is ineffective or even impossible to deeply scrutinize native behavior intention, lifecycle, dynamics and impact on complex problems and business issues. In addition, existing methods mainly focus on individual behavior analysis, which only captures a local picture of the interactions among behaviors. Further, state-of-the-art relevant research often overlooks the verification of behavior modeling, which weakens the sound-

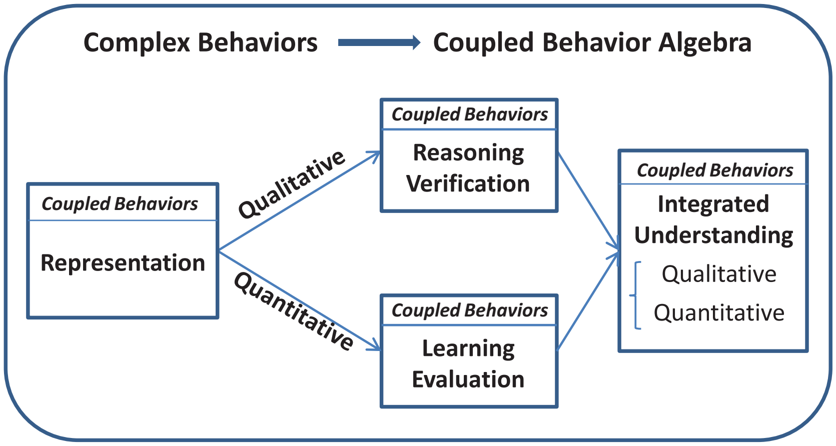


Figure 8.1: Prospects for the coupled behavior informatics

ness and robustness of complex behavior application models. Last but not the least is that most of the existing mining and learning algorithms follow the assumption of independence and identical distribution (i.e. IIDness), which is too strong and seriously mismatches the reality and complexities in behavioral/social problems.

Coupled behaviors refer to the activities of one to many actors who are associated with each other via certain relationships. With increasing network and community-based events and applications, such as group-based crime and social network interactions, behavior coupling contributes to the causes of eventual business problems. Effective approaches for analyzing coupled behaviors are not available, since existing methods mainly focus on individual behavior analysis. This thesis proposes the coupled behavior informatics in terms of modeling, analysis and learning, including the formalization and verification of qualitative group behaviors, the coupled behavior analysis on numerical data, the coupled behavior analysis on categorical data, and the coupled framework for clustering ensembles. They have brought about challenges and opportunities for existing behavior studies and relevant data analytics approaches. We show that the coupled behavior informatics creates new opportunities, directions and means for qualitative and quantitative, formal and systematic modeling, analysis and learning of complex behaviors in both physical and virtual organizations.

As shown in Figure 8.1, we have developed two directions to explicate a global picture of the coupled behaviors informatics: qualitative and quantitative behavior analytics. With the formal representation of coupled behaviors, the qualitative analytics addresses the task of coupled behavior reasoning and verification, while the quantitative research targets coupled behavior learning and evaluation. Finally, an appropriate way could be chosen to integrate these two studies to obtain an integrated understanding of the consolidated complex coupled behaviors from both qualitative and quantitative aspects. During this process, many open issues deserve a systematic investigation along with case studies from aspects such as *coupled behavior reasoning*, *coupled behavior learning*, *coupled behavior evaluation*, *coupled behavior integration* at individual but more on group levels.

In this thesis, Chapter 3 focuses on the qualitative perspective in terms of group behavior formalization and verification. Under the assumption of non-IIDness, Chapter 4, Chapter 5 and Chapter 6 specify the quantitative aspects with regard to the numerical coupled behavior analysis, the categorical coupled behavior analysis and the coupled behavior ensemble learning, respectively. All the new measures, approaches and frameworks introduced in these chapters have been evidenced to outperform the existing methods in terms of theoretical analysis or empirical studies or both of them. In addition, our proposed coupled recommendation system (Yu et al. 2013) and coupled document clustering (Cheng et al. 2013) also belong to the quantitative direction of coupled behavior informatics. Finally, Chapter 7 abstracts an integrated coupled behavior algebra to provide a unified understanding of coupled behaviors. All the current limitations and challenges have been addressed and solved, although to different extents.

Each chapter (i.e. from Chapter 3 to Chapter 6) of this thesis is supported by at least one accepted or published conference papers listed in Appendix A, and then enhanced and supplemented by the corresponding journal submissions. More encouragingly, several novel frameworks proposed in this thesis have been successfully applied in other topics and domains, such as multi-agent configuration, recommendation system and document analysis, with relevant papers recognized by research peers[1]. Therefore, what we have done and propose in this thesis is of great significance to the behavior related research, and the coupled behavior informatics exposes the intrinsic structure and essential nature of behavioral, social and business problems and applications.

8.2 Future Work

In this section, our ongoing work and future directions of coupled behavior informatics are listed and stated below in terms of qualitative coupled behavior analytics, quantitative behavior analytics and their integrated understanding. In the first part, potential tasks of Chapter 3 are addressed. For the quantitative behavior analytics, future work of Chapter 4, Chapter 5 and Chapter 6 is clearly explained. Finally, we reemphasize possible research issues regarding the integrated and consolidated understanding of coupled behaviors in Chapter 7.

8.2.1 Qualitative Behavior Analytics

In Chapter 3, group behavior formalization and verification are analyzed and explored. Our proposed Ontology-based Qualitative Coupled Behavior Modeling and Checking (*OntoB*) system has great potential for designing and analyzing complex behavior-oriented applications with complex behavior interactions. Currently, we are working on the extension of logic expressions for constraints, a behavior algebra to consolidate the techniques for modeling and checking complex behaviors, and behavior aggregation rules for the divergence and convergence of complex couplings. Flexible semantic interpretations of intra-coupled and inter-coupled behaviors are also under study

[1] The detailed information of those papers can be found at the end of each chapter.

on the top of our fundamental behavior building blocks, according to different requirements. The analysis of group behavior interactions brings about great challenges and opportunities in many aspects such as representing, checking, reasoning, learning behavior couplings and interactions, and mining behavior interaction patterns.

8.2.2 Quantitative Behavior Analytics

Potential research issues are clarified one by one from the numerical coupled behavior analysis, the categorical coupled behavior analysis to the coupled behavior ensemble learning.

Numerical Coupled Behavior Analysis

In Chapter 4, we propose a framework of the numerical coupled behavior analysis in terms of the intra-coupled interaction within a property, the inter-coupled interaction between multiple properties, and the integration of them. We are currently enriching this framework of the coupled attribute analysis on numerical data by also addressing the coupling relationships for entities and clusters. In the future, we will work on modeling the coupling relationships for mixed behavior data with both numerical and categorical properties. Additionally, we intend to extend the current work to the tasks of classification and fraud detection etc. by also examining the coupling relationships.

Categorical Coupled Behavior Analysis

In Chapter 5, novel data-driven similarity measures of categorical coupled behavior incorporating both intra-coupled property similarity for values and inter-coupled property similarity for values are introduced. We are currently applying the Coupled Attribute Similarity for Objects (*CASO*) measure with Inter-coupled Relative Similarity based on Intersection Set (*IRSI*) to attribute discretization, classification and other data mining and machine learning tasks. We are working on the assignment of attribute weights, and the flexible engagement of Intra-coupled and Inter-coupled Attribute Value Similarities ($IaASV$ and $IeASV$). We are designing the strategies of attribute reduction to fit extremely large data. Moreover, the proposed concepts *Inter-information Function* and *Information Conditional Probability* have the potential to be used in other applications. One of the clustering criteria, Minimal-Sum-Square, can also be adapted to involve the couplings of categorical behavior data and thus can be improved. Flexible dissimilarity measures can also be built on our fundamental similarity building blocks according to a range of requirements.

Coupled Behavior Ensemble Learning

In Chapter 6, we design a novel framework for coupled clustering ensembles (CCE) to incorporate interactions both between base clusterings (i.e. categorical properties or methods) and entities. This work verifies that non-IIDness is essential to the clustering ensemble problem. The coupling of clusterings can enhance clustering quality in most cases, and the performance gain depends on the quality of the base clusterings. The inter-coupling of entities is associated with the consistency of base clustering results, which leads to fluctuating improvement on the clustering quality. Thus, how should we fix the weights λ_k of base clustering bc_k in $IeCSC$ rather than simply treating them equally? Further, should we introduce a weight to control the couplings of entities during the process of the clustering ensemble? Is there any other way to model the coupling of entities by considering the relative common neighborhood rather than the absolute neighborhood? How do we fix the number of final clusters? We are currently working on these issues, as mentioned in Section 7.4.2, and will also analyze the heterogeneous structure of base clusterings and the advanced post-processing clustering techniques in our framework. Furthermore, we will consider the coupling of clusters and then extend this coupled idea to the supervised learning process.

8.2.3 Consolidated Understanding

In Chapter 7, we abstract a coupled behavior algebra by defining the intra-coupling and inter-coupling of descriptors as well as the intra-coupling and inter-coupling of bodies, but still at its early stage.

As mentioned in Section 7.3, there are a lot of opportunities and directions for us to explore and investigate, including the definition of computing operators and the identification of their properties, the further investigation of the context and consolidated coupling, the construction of a coupled behavior space with different characteristics, the analysis on limits, convergence and divergence of coupled behavior sequences, the base computation of this coupled behavior space, the coupled behavior space decomposition, the linear or nonlinear transformation, and the exploration on mix-type coupled behaviors. After every building block has been defined and explored well, such a coupled behavior algebra is highly expected to create a totally new subject to support cross-multi-discipline research with great significance to behavioral/social/business applications.

All the above open issues and future directions from a diversity of perspectives constitute a large research blueprint with huge opportunities on this promising topic: coupled behavior informatics.

Appendix A

Appendix: List of Publications

Papers Accepted and Published

- **Can Wang**, Zhong She, Longbing Cao (2013), "Coupled Attribute Analysis on Numerical Data". The 23rd International Joint Conference on Artificial Intelligence (**IJCAI 2013**), full paper accepted.

- **Can Wang**, Zhong She, Longbing Cao (2013), "Coupled Clustering Ensemble: Incorporating Coupling Relationships Both between Base Clusterings and Objects". The 29th IEEE International Conference on Data Engineering (**ICDE 2013**), full paper accepted.

- Jinjiu Li, **Can Wang**, Longbing Cao, Philip S. Yu (2013), "Efficient Globally Optimal Rule Selection on Large Imbalanced Data Based on Rule Coverage Relationship Analysis". 2013 SIAM International Conference on Data Mining (**SDM 2013**), full paper accepted.

- Jinjiu Li, **Can Wang**, Wei Wei, Mu Li, Chunming Liu (2013), "Efficient Mining of Contrast Patterns on Large Scale Imbalanced Real-life Data". The 17th Pacific-Asia Conference on Knowledge Discovery and Data Mining (**PAKDD 2013**), pp. 62-73. [**Best Student Paper Award**]

- Yonghong Yu, **Can Wang**, Yang Gao, Longbing Cao, Xixi Chen (2013), "A Coupled Clustering Approach for Items Recommendation". The 17th Pacific-Asia Conference on Knowledge Discovery and Data Mining (**PAKDD 2013**), pp. 365-376.

- Zhong She, **Can Wang** (2013). "Volatility Analysis via Coupled Wishart Process". The 2013 International Joint Conference on Neural Networks (**IJCNN 2013**), full paper accepted.

- Xin Cheng, Duoqian Miao, **Can Wang**, Longbing Cao (2013). "Coupled Term-Term Relation Analysis for Document Clustering". The 2013 International Joint Conference on Neural Networks (**IJCNN 2013**), full paper accepted.

- **Can Wang**, Mingchun Wang, Zhong She, Longbing Cao (2012), "CD: A Coupled Discretization Algorithm". The 16th Pacific-Asia Conference on Knowledge Discovery and Data Mining (**PAKDD 2012**), pp. 407-418.

- **Can Wang**, Longbing Cao (2012), "Modeling and Analysis of Social Activity Process". Behavior Computing: Modeling, Analysis, Mining and Decision, Springer, pp. 21-35.

- Zhong She, **Can Wang**, Longbing Cao (2012), "CCE: A Coupled Framework of Clustering Ensembles". The 26th Conference on Artificial Intelligence (**AAAI 2012**), pp. 2455-2456.

- **Can Wang**, Longbing Cao, Mingchun Wang, Jinjiu Li, Wei Wei, Yuming Ou (2011), "Coupled Nominal Similarity in Unsupervised Learning". The 20th ACM Conference on Information and Knowledge Management (**CIKM 2011**), pp. 973-978.

- Chayapol Moemeng, **Can Wang**, Longbing Cao (2011), "Obtaining an Optimal MAS Configuration for Agent-Enhanced Mining Using Constraint Optimization". The 7th International Workshop on Agents and Data Mining Interaction held in conjunction with the 10th International Conference on Autonomous Agents and Multiagent Systems (**ADMI with AAMAS 2011**), pp. 46-57.

- Juan Zhao, Mingchun Wang, Kun Liu, **Can Wang** (2011), "Discretization based on Positive Domain and Information Entropy". The 7th International Conference on Computational Intelligence and Security (**CIS 2011**), pp. 258-262.

- **Can Wang**, Longbing Cao (2010), "SAPMAS: Social Activity Process Modeling and Analysis System". The International Workshop on Behavior Informatics held in conjunction with The 14th Pacific-Asia Conference on Knowledge Discovery and Data Mining (**BI with PAKDD 2010**).

- Jiayi Feng, Mingchun Wang, **Can Wang**, Longbing Cao (2010), "Enhanced Co-occurrence Distances For Categorical Data in Unsupervised Learning", pp. 2071-2078. The Ninth International Conference on Machine Learning and Cybernetics (**ICMLC 2010**), 2010.

Papers Submitted/Under Review

- **Can Wang**, Zhong She, Longbing Cao (2013), "Coupled Clustering Ensemble Involving Non-IIDness". Artificial Intelligence (**AI**).

- **Can Wang**, Longbing Cao (2013), "Formalization and Verification of Group Behavior Interactions". IEEE Transactions on Systems, Man, and Cybernetics–Part A: Systems (**TSMC-A**).

- **Can Wang**, Longbing Cao (2013), "Coupled Attribute Similarity Analysis on Categorical Data". IEEE Transactions on Neural Networks and Learning Systems (**TNNLS**).

Appendix B

Appendix: List of Symbols

The following list is neither exhaustive nor exclusive, but may be helpful.

$\mathbb{B}_c$	Qualitative coupled behaviors
$\{\mathscr{A}_1, \ldots, \mathscr{A}_I\}$	The set of I actors
$\{\mathscr{O}_{i1}, \ldots, \mathscr{O}_{iJ_i}\}$	The set of J_i operations conducted by actor $\mathscr{A}_i$
$\mathscr{C}$	Coupling
$\theta(\mathbb{B})$	Intra-coupling function
$\eta(\mathbb{B})$	Inter-coupling function
$FM(\mathbb{B})$	Behavior feature matrix
$\mathbb{B}^\theta(\mathscr{A}_i)$	Intra-coupled behaviors of actor $\mathscr{A}_i$
$\mathbb{B}^\eta(\mathscr{A}_i)$	Inter-coupled behaviors of actor $\mathscr{A}_i$
$\{u_1, \cdots, u_m\}$	The set of m entities/objects
$\{a_1, \cdots, a_n\}$	The set of n properties/attributes
$v_j^x, v_t \in V_j$	Values of attribute a_j in their value set V_j
$\langle a_j \rangle^p$	The p-th power of numerical attribute a_j

θ_{pq}, η_{pq}	Pearson's correlation coefficient
$\mathbf{R^{Ia}}(a_j)$	Intra-coupled interaction of numerical attribute a_j
$\mathbf{R^{Ie}}(a_j\|\{a_k\}_{k\neq j})$	Inter-coupled interaction of numerical attribute a_j
$\mathbf{u^c}(\widetilde{A}, L)$	Coupled representation for numerical entity u
F_j, G_j	Set information functions
$\varphi_{j\rightarrow k}$	Inter-information function
$P_{k\|j}$	Information conditional probability
$R(=\max\|V_j\|)$	The maximal number of categorical attribute values
$\delta_j^{Ia}(v_j^x, v_j^y)$	Intra-coupled attribute similarity for nominal values
$\delta_j^{Ie}(v_j^x, v_j^y, \{V_k\}_{k\neq j})$	Inter-coupled attribute similarity for nominal values
$\delta_j^{A}(v_j^x, v_j^y, \{V_k\}_{k=1}^n)$	Coupled attribute similarity for nominal values
$CASO(u_x, u_y)$	Coupled attribute similarity for nominal objects
$CADO(u_x, u_y)$	Coupled attribute dissimilarity for nominal objects
$\{bc_1, \cdots, bc_L\}$	The set of L base clusterings
$\{c_j^1, \cdots, c_j^{t_j}\}$	The set of t_j clusters in base clustering bc_j
$\{c_*^1, \cdots, c_*^{t^*}\}$	A final clustering fc^* with t^* clusters
$N_{u_x}^{Sim}$	The neighbor set of object u_x based on measure δ^{Sim}
$(BC_j)_{m\times m}$	The associated similarity matrix of objects for bc_j
$\delta_j^{IaC}(v_j^x, v_j^y)$	Intra-coupled clustering similarity for clusters
$\delta_j^{IeC}(v_j^x, v_j^y\|\{V_k\}_{k\neq j})$	Inter-coupled clustering similarity for clusters
$\delta_j^{C}(v_j^x, v_j^y\|\{V_k\}_{k=1}^L)$	Coupled clustering similarity for clusters

$\delta^{IaO}(u_x, u_y)$	Intra-coupled object similarity for objects
$\delta^{IeO}(u_x, u_y \vert U, \delta^{Sim})$	Inter-coupled object similarity for objects
$\delta^{CO}(u_x, u_y \vert U)$	Coupled clustering and object similarity for objects
$S_{Cg}(bc_{j_1}, bc_{j_2})$	Proposed clustering-based coupling for base clusterings
$S_O^{IaC}(u_x, u_y),\ S_O^C(u_x, u_y)$	Proposed object-based coupling for objects
$S_{Cr}^C(c_{j_1}^{t_1}, c_{j_2}^{t_2})$	Proposed cluster-based coupling for clusters
$\mathbb{CB}$	Coupled behaviors
$\mathbf{B}$	Body
$\mathbf{D}$	Depictor
$\mathbf{C}$	Consolidated coupling
$\mathbf{T}$	Context
$\mathcal{V}$	Depictor value
$\theta(\cdot)$	Consolidated intra-coupling function
$\eta(\cdot)$	Consolidated inter-coupling function
$\mathbf{D}_j^\theta$	Intra-coupling of depictors
$\mathbf{D}_j^\eta$	Inter-coupling of depictors
$\mathbf{B}_i^\theta$	Intra-coupling of bodies
$\mathbf{B}_i^\eta$	Inter-coupling of bodies

Bibliography

Ackerman, M. & Ben-David, S. (2011), Discerning linkage-based algorithms among hierarchical clustering methods, *in* 'Proceedings of the 22nd International Joint Conference on Artificial Intelligence', pp. 1140–1145.

Ahmad, A. & Dey, L. (2007), 'A k-mean clustering algorithm for mixed numeric and categorical data', *Data and Knowledge Engineering* **63**, 503–527.

Alur, R., Henzinger, T. A. & Kupferman, O. (2002), 'Alternating-time temporal logic', *Journal of the ACM* **49**(5), 672–713.

Andritsos, P., Tsaparas, P., Miller, R. & Sevcik, K. (2004), LIMBO: Scalable clustering of categorical data, *in* 'Proceedings of the 9th International Conference on Extending Database Technology', pp. 123–146.

Ayres, J., Flannick, J., Gehrke, J. & Yiu, T. (2002), Sequential pattern mining using a bitmap representation, *in* 'Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining', pp. 429–435.

Baier, C. & Joost, P. K. (2008), *Principles of Model Checking*, The MIT Press, Cambridge, MA.

Barbara, D., Couto, J. & Li, Y. (2002), COOLCAT: An entropy-based algorithm for categorical clustering, *in* 'Proceedings of the 11th International Conference on Information and Knowledge Management', pp. 582–589.

Bi, L., Tsimhoni, O. & Liu, Y. (2011), 'Using the support vector regression approach to model human performance', *IEEE Transactions on Systems, Man, and Cybernetics–Part A: Systems and Humans* **41**(3), 410–417.

Bi, Y., Guan, J. & Bell, D. (2008), 'The combination of multiple classifiers using an evidential reasoning approach', *Artificial Intelligence* **172**(15), 1731–1751.

Bickhard, M. (2000), 'An interactive process model', *The Caldron of Consciousness: Motivation, Affect, and Self-organization: An Thology* pp. 161–178.

Bollegala, D., Matsuo, Y. & Ishizuka, M. (2011), Relation adaptation: learning to extract novel relations with minimum supervision, *in* 'Proceedings of the 22nd International Joint Conference on Artificial Intelligence', pp. 2205–2210.

Bordini, R. H., Fisher, M., Wooldridge, M. & Visser, W. (2004), 'Model checking rational agents', *IEEE Intelligent Systems* **19**(5), 46–52.

Boriah, S., Chandola, V. & Kumar, V. (2008), Similarity measures for categorical data: A comparative evaluation, *in* 'Proceedings of the 8th SIAM International Conference on Data Mining', pp. 243–254.

Brachman, R. J. & Levesque, H. J. (2004), *Knowledge Representation and Reasoning*, Elsevier, New York.

Breitman, K. K., Casanova, M. A. & Truszkowski, W. (2007), *Semantic Web: Concepts, Technologies and Applications*, Springer, New York.

Cai, D., He, X. & Han, J. (2005), 'Document clustering using locality preserving indexing', *IEEE Transactions on Knowledge and Data Engineering* **17**(12), 1624–1637.

Calders, T., Goethals, B. & Jaroszewicz, S. (2006), Mining rank-correlated sets of numerical attributes, *in* 'Proceedings of the 12th ACM SIGKDD Conference on Knowledge Discovery and Data Mining', ACM, pp. 96–105.

Cao, H., Mamoulis, N. & Cheung, D. W. (2007), 'Discovery of periodic patterns in spatiotemporal sequences', *IEEE Transactions on Knowledge and Data Engineering* **19**(4), 453–467.

Cao, L. (2010), 'In-depth behavior understanding and use: The behavior informatics approach', *Information Sciences* **180**(17), 3067–3085.

Cao, L. (2013), 'Non-iidness learning: an overview', *The Computer Journal* pp. 1–18.

Cao, L., Ou, Y. & Yu, P. (2012), 'Coupled behavior analysis with applications', *IEEE Transactions on Knowledge and Data Engineering* **24**(8), 1378–1392.

Cao, L. & Philip, S. Y. (2012), *Behavior Computing: Modeling, Analysis, Mining and Decision*, Springer.

Cao, L. & Yu, P. (2009), 'Behavior informatics: an informatics perspective for behavior studies', *The Intelligent Informatics Bulletin* **10**(1), 6–11.

Cao, L., Zhao, Y., Zhang, C. & Zhang, H. (2008), 'Activity mining: from activities to actions', *International Journal of Information Technology & Decision Making* **7**(02), 259–273.

Chandrakala, S. & Chandra Sekhar, C. (2008), A density based method for multivariate time series clustering in kernel feature space, *in* 'Proceedings of the 2008 IEEE International Joint Conference on Neural Networks', pp. 1885–1890.

Chen, T., Zhang, N. L., Liu, T., Poon, K. M. & Wang, Y. (2012), 'Model-based multidimensional clustering of categorical data', *Artificial Intelligence* **176**(1), 2246–2269.

Cheng, X., Miao, D., Wang, C. & Cao, L. (2013), Coupled term-term relation analysis for document clustering, *in* 'Proceedings of the 2013 International Joint Conference on Neural Networks', accepted.

Christou, I. (2011), 'Coordination of cluster ensembles via exact methods', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **33**(2), 279–293.

Cost, S. & Salzberg, S. (1993), 'A weighted nearest neighbor algorithm for learning with symbolic features', *Machine Learning* **10**(1), 57–78.

Das, G. & Mannila, H. (2000), Context-based similarity measures for categorical databases, *in* 'Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases 2000', pp. 201–210.

Dasgupta, K., Singh, R., Viswanathan, B., Chakraborty, D., Mukherjea, S., Nanavati, A. A. & Joshi, A. (2008), Social ties and their relevance to churn in mobile telecom networks, *in* 'Proceedings of the 11th International Conference on Extending Database Technology: Advances in Database Technology', pp. 668–677.

Davies, D. & Bouldin, D. (1979), 'A cluster separation measure', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **1**(2), 224–227.

Domeniconi, C. & Al-Razgan, M. (2009), 'Weighted cluster ensembles: methods and analysis', *ACM Transactions on Knowledge Discovery from Data* **2**(4), 17.

Donoho, S. (2004*a*), Early detection of insider trading in option markets, *in* 'Proceedings of the 10th ACM SIGKDD Conference on Knowledge Discovery and Data Mining', pp. 420–429.

Donoho, S. (2004*b*), Early detection of insider trading in option markets, *in* 'Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining', pp. 420–429.

Dunn, J. (1974), 'Well-separated clusters and optimal fuzzy partitions', *Cybernetics and Systems* **4**(1), 95–104.

Eagle, N., Pentland, A. S. & Lazer, D. (2008), Mobile phone data for inferring social network structure, *in* 'Social Computing, Behavioral Modeling, and Prediction', pp. 79–88.

Edelkamp, S. & Kissmann, P. (2009), Optimal symbolic planning with action costs and preferences, *in* 'Proceedings of the 21st AAAI Conference on Artificial Intelligence', pp. 1690–1695.

Fast, A., Friedland, L., Maier, M., Taylor, B., Jensen, D., Goldberg, H. G. & Komoroske, J. (2007), Relational data pre-processing techniques for improved securities fraud detection, *in* 'Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining', pp. 941–949.

Fawcett, T. & Provost, F. (1999), Activity monitoring: noticing interesting changes in behavior, *in* 'Proceedings of the fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining', pp. 53–62.

Fern, X. Z. & Brodley, C. E. (2004), Solving cluster ensemble problems by bipartite graph partitioning, *in* 'Proceedings of the 21st International Conference on Machine Learning', pp. 36–43.

Figueiredo, F., Rocha, L., Couto, T., Salles, T., André Gonçalves, M. & Meira Jr, W. (2011), 'Word co-occurrence features for text classification', *Information Systems* **36**(5), 843–858.

Flesca, S., Greco, S., Tagarelli, A. & Zumpano, E. (2005), 'Mining user preferences, page content and usage to personalize website navigation', *World Wide Web* **8**(3), 317–345.

Frank, A. & Asuncion, A. (2010), *UCI Machine Learning Repository*, University of California, Irvine, School of Information and Computer Sciences.

Frey, B. & Dueck, D. (2007), 'Clustering by passing messages between data points', *Science* **315**(5814), 972–976.

Gabaldon, A. (2009), Activity recognition with intended actions, *in* 'Proceedings of the 21st AAAI Conference on Artificial Intelligence', pp. 1696–1701.

Gan, G., Ma, C. & Wu, J. (2007), *Data Clustering: Theory, Algorithms, and Applications*, ASA-SIAM Series on Statistics and Applied Probability, Philadelphia.

Gao, J., Fan, W. & Han, J. (2010), On the power of ensemble: supervised and unsupervised methods reconciled, *in* 'Proceedings of the 10th SIAM International Conference on Data Mining', pp. 1–163.

García-Osorio, C., de Haro-García, A. & García-Pedrajas, N. (2010), 'Democratic instance selection: a linear complexity instance selection algorithm based on classifier ensemble concepts', *Artificial Intelligence* **174**(5), 410–441.

Garcia, S., Derrac, J., Cano, J. & Herrera, F. (2012), 'Prototype selection for nearest neighbor classification: taxonomy and empirical study', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **34**(3), 417–435.

Getoor, L. & Taskar, B. E. (2007), *Introduction to Statistical Relational Learning*, MIT Press, Cambridge, MA.

Giacomo, G. D., Lesperance, Y. & Pearce, A. R. (2010), Situation calculus-based programs for representing and reasoning about game structures, *in* 'Proceedings of the 12th International Conference on the Principles of Knowledge Representation and Reasoning', pp. 445–455.

Gibbs, A. & Su, F. (2002), 'On choosing and bounding probability metrics', *International Statistical Review* **70**(3), 419–435.

Gibson, D., Kleinberg, J. & Raghavan, P. (2000), 'Clustering categorical data: An approach based on dynamical systems', *The International Journal on Very Large Data Bases* **8**(3), 222–236.

Gionis, A., Mannila, H. & Tsaparas, P. (2007), 'Clustering aggregation', *ACM Transactions on Knowledge Discovery from Data* **1**(1), 1–30.

Gruber, T. R. (1993), 'A translation approach to portable ontology specifications', *Knowledge Acquisition* **5**(2), 199–220.

Gu, Y. & Soutchanski, M. (2007), Decidable reasoning in a modified situation calculus, *in* 'Proceedings of the 20th AAAI Conference on Artificial Intelligence', pp. 1891–1897.

Guha, S., Rastogi, R. & Shim, K. (2000), 'ROCK: A robust clustering algorithm for categorical attributes', *Information Systems* **25**(5), 345–366.

Hodge, V. & Austin, J. (2004), 'A survey of outlier detection methodologies', *Artificial Intelligence Review* **22**(2), 85–126.

Hogg, T. & Szabo, G. (2008), Diversity of online community activities, *in* 'Proceedings of the 19th ACM Conference on Hypertext and Hypermedia', pp. 227–228.

Holcombe, R. G. (1989), *Economic Models and Methodology*, Greenwood, New York.

Hollingsworth, K., Bowyer, K. & Flynn, P. (2011), 'Improved Iris recognition through fusion of Hamming distance and fragile bit distance', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **33**(12), 2465–2476.

Holzmann, G. J. (2003), *The Spin Model Checker: Primer and Reference Manual*, Addison Wesley, Boston.

Houle, M., Oria, V. & Qasim, U. (2010), Active caching for similarity queries based on shared-neighbor information, *in* 'Proceedings of the 19th ACM International Conference on Information and Knowledge Management', pp. 669–678.

Iam-On, N. & Boongoen, T. (2012), Improved link-based cluster ensembles, *in* 'Proceedings of the International Joint Conference on Neural Networks 2012', pp. 1–8.

Iam-On, N., Boongoen, T., Garrett, S. & Price, C. (2011), 'A link-based approach to the cluster ensemble problem', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **33**(12), 2396–2409.

Jain, A., Murty, M. & Flynn, P. (1999), 'Data clustering: a review', *ACM Computing Surveys* **31**(3), 264–323.

Jakulin, A. & Bratko, I. (2003), Analyzing attribute dependencies, *in* 'Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases 2003', pp. 229–240.

Jia, Y. & Zhang, C. (2008), Instance-level semisupervised multiple instance learning, *in* 'Proceedings of the 23rd AAAI Conference on Artificial Intelligence', pp. 640–645.

Kalogeratos, A. & Likas, A. (2012), 'Text document clustering using global term context vectors', *Knowledge and Information Systems* **31**(3), 455–474.

Kaminka, G. & Frenkel, I. (2005), Flexible teamwork in behavior-based robots, *in* 'Proceedings of the 19th AAAI Conference on Artificial Intelligence', pp. 108–113.

Kaufman, L. & Rousseeuw, P. (1990), *Finding Groups in Data: An Introduction to Cluster Analysis*, John Wiley and Sons, New York.

Kaufmann, M., Manolios, P. & Moore, J. (2000), *Using the ACL2 Theorem Prover: A Tutorial Introduction and Case Studies*, Kluwer Academic, Boston, MA.

Kaytoue, M., Kuznetsov, S. & Napoli, A. (2011), Revisiting numerical pattern mining with formal concept analysis, *in* 'Proceedings of the 22nd International Joint Conference on Artificial Intelligence', pp. 1342–1347.

Kazakov, Y. (2009), Consequence-driven reasoning for Horn SHIQ ontologies, *in* 'Proceedings of the 21st International Joint Conference on Artificial Intelligence', pp. 2040–2045.

Kim, H. L., Breslin, J. G., Decker, S. & Kim, H. G. (2011), 'Mining and representing user interests: the case of tagging practices', *IEEE Transactions on Systems, Man, and Cybernetics–Part A: Systems and Humans* **41**(4), 683–692.

Kittler, J., Hatef, M., Duin, R. & Matas, J. (1998), 'On combining classifiers', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **20**(3), 226–239.

Kobsa, A. (2001), 'Generic user modeling systems', *User Modeling and User-Adapted Interaction* **11**(1-2), 49–63.

Koenig, S., Keskinocak, P. & Tovey, C. (2010), Progress on agent coordination with cooperative auctions, *in* 'Proceedings of the 24th AAAI Conference on Artificial Intelligence', pp. 1713–1717.

Kuncheva, L. & Hadjitodorov, S. (2004), Using diversity in cluster ensembles, *in* 'Proceedings of the IEEE International Conference on Systems, Man and Cybernetics', Vol. 2, pp. 1214–1219.

Kuncheva, L. & Vetrov, D. (2006), 'Evaluation of stability of k-means cluster ensembles with respect to random initialization', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**(11), 1798–1808.

Kwan, I. S., Fong, J. & Wong, H. (2005), 'An e-customer behavior model with online analytical mining for internet marketing planning', *Decision Support Systems* **41**(1), 189–204.

Li, D. & Liu, C. (2012), 'Extending attribute information for small data set classification', *IEEE Transactions on Knowledge and Data Engineering* **24**(3), 452–464.

Li, T., Ogihara, M. & Ma, S. (2010), 'On combining multiple clusterings: an overview and a new perspective', *Applied Intelligence* **33**(2), 207–219.

Lim, G. H., Suh, I. H. & Suh, H. (2011), 'Ontology-based unified robot knowledge for service robots in indoor environments', *IEEE Transactions on Systems, Man, and Cybernetics–Part A: Systems and Humans* **41**(3), 492–509.

Lin, D. (1998), An information-theoretic definition of similarity, *in* 'Proceedings of the 15th International Conference on Machine Learning', pp. 296–304.

Liu, H., Salerno, J. & Young, M. J. E. (2008), *Social Computing, Behavioral Modeling, and Prediction*, Springer, New York.

Lomuscio, A., Qu, H. & Raimondi, F. (2009), MCMAS: A model checker for the verification of multi-agent systems, *in* 'Computer Aided Verification', pp. 682–688.

Luo, X. & Jennings, N. R. (2007), 'A spectrum of compromise aggregation operators for multi-attribute decision making', *Artificial Intelligence* **171**(2), 161–184.

Luxburg, U. (2007), 'A tutorial on spectral clustering', *Statistics and Computing* **17**(4), 1–32.

Mozer, M., Wolniewicz, R., Grimes, D., Johnson, E. & Kaushanky, H. (2000), 'Predicting subscriber dissatisfaction and improving retention in wireless telecommunications industry', *IEEE Transactions on Neural Networks* **11**(3), 690–696.

Nagy, M. & Vargas-Vera, M. (2011), 'Multiagent ontology mapping framework for the semantic web', *IEEE Transactions on Systems, Man, and Cybernetics–Part A: Systems and Humans* **41**(4), 693–704.

Nanavati, A., Gurumurthy, S., Das, G., Chakraborty, D., Dasgupta, K., Mukherjea, S. & Joshi, A. (2006), On the structural properties of massive telecom call graphs: findings and implications, *in* 'Proceedings of the 15th ACM Conference on Information and Knowledge Management', pp. 435–444.

Nazareth, D., Soofi, E. & Zhao, H. (2007), Visualizing attribute interdependencies using mutual information, hierarchical clustering, multidimensional scaling, and self-organizing maps, *in* 'Proceedings of the 40th Hawaii International International Conference on Systems Science', pp. 53–53.

Nguyen, N. & Caruana, R. (2007), Consensus clusterings, *in* 'Proceedings of the 7th IEEE International Conference on Data Mining', pp. 607–612.

Oliver, M. & Pentland, A. P. (2000), 'A Bayesian computer vision system for modeling human interactions', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**, 831–843.

Oliver, N. M., Rosario, B. & Pentland, A. P. (2000), 'A Bayesian computer vision system for modeling human interactions', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**(8), 831–843.

Park, Y.-J. & Chang, K.-N. (2009), 'Individual and group behavior-based customer profile model for personalized product recommendation', *Expert Systems with Applications* **36**(2), 1932–1939.

Peterson, G. D., Cumming, G. S. & Carpenter, S. R. (2003), 'Scenario planning: a tool for conservation in an uncertain world', *Conservation Biology* **17**(2), 358–366.

Pierce, W. & Cheney, C. (2004), *Behavior Analysis and Learning*, 3rd edn, Lawrence Erlbaum Associates.

Plant, C. (2012), Dependency clustering across measurement scales, *in* 'Proceedings of the 18th ACM SIGKDD Conference on Knowledge Discovery and Data Mining', pp. 361–369.

Punera, K. & Ghosh, J. (2007), 'Soft cluster ensembles', *Advances in Fuzzy Clustering and Its Applications* pp. 69–91.

Razmerita, L. (2011), 'An ontology-based framework for modeling user behavior–a case study in knowledge management', *IEEE Transactions on Systems, Man, and Cybernetics–Part A: Systems and Humans* **41**(4), 772–783.

Ribeiro, L. A. & Harder, T. (2011), 'Generalizing prefix filtering to improve set similarity joins', *Information Systems* **36**(1), 62–78.

Ros, R. & Veloso, M. (2007), Executing multi-robot cases through a single coordinator, *in* 'Proceedings of the 6th International Conference on Autonomous Agents and Multiagent Systems', ACM Press, pp. 1272–1274.

Sardina, S., Patrizi, F., Giacomo, G. D. & Universita, S. (2008), Behavior composition in the presence of failure, *in* 'Proceedings of the 11th International Conference on Principles of Knowledge Representation and Reasoning', pp. 640–650.

Saria, S., Duchi, A. & Koller, D. (2011), Discovering deformable motifs in continuous time series data, *in* 'Proceedings of the 22nd International Joint Conference on Artificial Intelligence', pp. 1465–1471.

Serrano, J. M. & Saugar, S. (2010), An architectural perspective on multi-agent societies, *in* 'Proceedings of the 11th International Workshop on Agent-Oriented Software Engineering', pp. 85–90.

Sicilia, M.-A. (2007), 'Ontology of systems and software engineering', *Advanced Engineering Informatics* **21**(2), 117–118.

Song, Y., Cao, L., Wu, X., Wei, G., Ye, W. & Ding, W. (2012), Coupled behavior analysis for capturing coupling relationships in group-based market manipulations, *in* 'Proceedings of the 18th ACM SIGKDD Conference on Knowledge Discovery and Data Mining', pp. 976–984.

Sony Four Legged Robot Football League Rule Book (2004), RoboCup Technical Committee.

Srivastava, J., Cooley, R., Deshpande, M. & Tan, P.-N. (2000), 'Web usage mining: discovery and applications of usage patterns from web data', *ACM SIGKDD Explorations Newsletter* **1**(2), 12–23.

Staab, S. & Studer, R. (2009), *Handbook on Ontologies*, 2rd edn, Springer Verlag, Berlin, DE.

Strehl, A. & Ghosh, J. (2002), 'Cluster ensembles–a knowledge reuse framework for combining multiple partitions', *Journal of Machine Learning Research* **3**, 583–617.

Subramanian, K. (2010), Task space behavior learning for humanoid robots using Gaussian mixture models, *in* 'Proceedings of the 24th Conference on Artificial Intelligence', pp. 1961–1962.

Sun, Z. (2007), Multi-agent based modeling: methods and techniques for investigating human behaviors, *in* 'Proceedings of the 2007 IEEE International Conference on Mechatronics and Automation', pp. 779–783.

Topchy, A., Jain, A. & Punch, W. (2005), 'Clustering ensembles: models of consensus and weak partitions', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **27**(12), 1866–1881.

Vega-Pons, S. & Ruiz-Shulcloper, J. (2011), 'A survey of clustering ensemble algorithms', *International Journal of Pattern Recognition and Artificial Intelligence* **25**(3), 337.

Vigna, G., Valeur, F. & Kemmerer, R. A. (2003), 'Designing and implementing a family of intrusion detection systems', *ACM SIGSOFT Software Engineering Notes* **28**(5), 88–97.

Wang, C. & Cao, L. (2010), SAPMAS: social activity process modeling and analysis system, *in* 'Proceedings of the Behavior Informatics Workshop of the 14th Pacific-Asia Conference on Knowledge Discovery and Data Mining', Hyderabad, India.

Wang, C. & Cao, L. (2012), 'Modeling and analysis of social activity process', *Behavior Computing,* pp. 21–35.

Wang, C., Cao, L., Wang, M., Li, J., Wei, W. & Ou, Y. (2011), Coupled nominal similarity in unsupervised learning, *in* 'Proceedings of the 20th ACM Conference on Information and Knowledge Management', pp. 973–978.

Wang, C., She, Z. & Cao, L. (2013*a*), Coupled attribute analysis on numerical data, *in* 'Proceedings of the 23rd International Joint Conference on Artificial Intelligence', p. accepted.

Wang, C., She, Z. & Cao, L. (2013*b*), Coupled clustering ensemble: incorporating coupling relationships both between base clusterings and objects, *in* 'Proceedings of the 29th IEEE International Conference on Data Engineering', p. accepted.

Wang, C., Wang, M., She, Z. & Cao, L. (2012), CD: A coupled discretization algorithm, *in* 'Proceedings of the 16th Pacific-Asia Conference on Knowledge Discovery and Data Mining', pp. 407–418.

Wang, F. Y., Carley, K. M., Zeng, D. & Mao, W. (2007), 'Social computing: From social informatics to social intelligence', *Intelligent Systems, IEEE* **22**(2), 79–83.

Wang, G., Hoiem, D. & Forsyth, D. (2012), 'Learning image similarity from Flickr groups using fast Kernel machines', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PrePrints**.

Wang, P., Domeniconi, C. & Laskey, K. (2010), Nonparametric Bayesian clustering ensembles, *in* 'Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases 2010', pp. 435–450.

Wang, W., Hidvegi, Z., Bailey, A.D., J. & Whinston, A. (2000), 'E-process design and assurance using model checking', *Computer* **33**(10), 48 –53.

Weiss, G. M. & Hirsh, H. (1998), Learning to predict rare events in event sequences, *in* 'Proceedings of the 4th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining', pp. 359–363.

Wessel, M. & Möller, R. (2009), 'Flexible software architectures for ontology-based information systems', *Journal of Applied Logic* **7**(1), 75–99.

Wilson, D. & Martinez, T. (1997), 'Improved heterogeneous distance functions', *Journal of Artificial Intelligence Research* **6**, 1–34.

Wilson, T. D. & Walsh, C. (1997), 'Information behavior: an interdisciplinary perspective', *Information Processing and Management* **33**(4), 551–572.

Wolpert, D. & Macready, W. (1996), No free lunch theorems for search, Technical report, Citeseer.

Wooldridge, M. (2000), *Reasoning About Rational Agents*, MIT, Cambridge, MA.

Yamanishi, K. & Takeuchi, J. (2002), A unifying framework for detecting outliers and change points from non-stationary time series data, *in* 'Proceedings of the 8th ACM SIGKDD Conference on Knowledge Discovery and Data Mining', pp. 676–681.

Yang, Y., Guan, X. & You, J. (2002), Clope: A fast and effective clustering algorithm for transactional data, *in* 'Proceedings of the 8th ACM SIGKDD Conference on Knowledge Discovery and Data Mining', pp. 682–687.

Yoon, H., Yang, K. & Shahabi, C. (2005), 'Feature subset selection and feature ranking for multivariate time series', *IEEE Transactions on Knowledge and Data Engineering* **17**(9), 1186–1198.

Yoshimura, K., Barnes, N., Ronnquist, R. & Sonenberg, L. (2003), Towards real-time strategic teamwork: a RoboCup case study, *in* 'Robot Soccer World Cup 2002', pp. 676–681.

Yu, A. C. (2006), 'Methods in biomedical ontology', *Journal of Biomedical Informatics* **39**(3), 252–266.

Yu, Y., Wang, C., Gao, Y., Cao, L. & Chen, X. (2013), A coupled clustering approach for items recommendation, *in* 'Proceedings of the 17th Pacific-Asia Conference on Knowledge Discovery and Data Mining', pp. 365–376.

Zacharias, G. L. & MacMillan, J. E. (2008), *Behavioral Modeling and Simulation: From Individuals to Societies*, National Academies Press.

Zaki, M. J. (2001), 'SPADE: an efficient algorithm for mining frequent sequences', *Machine learning* **42**(1-2), 31–60.

Zaki, M. J., Peters, M., Assent, I. & Seidl, T. (2005), Clicks: An effective algorithm for mining subspace clusters in categorical datasets, *in* 'Proceedings of the 11th ACM SIGKDD Conference on Knowledge Discovery and Data Mining', pp. 736–742.