Coupled Attribute Analysis on Numerical Data

Can Wang, Zhong She, Longbing Cao
Advanced Analytics Institute, University of Technology, Sydney, Australia
{canwang613, zhong2024, longbing.cao}@gmail.com

Abstract

The usual representation of quantitative data is to formalize it as an information table, which assumes the independence of attributes. In real-world data, attributes are more or less interacted and coupled via explicit or implicit relationships. Limited research has been conducted on analyzing such attribute interactions, which only describe a local picture of attribute couplings in an implicit way. This paper proposes a framework of the coupled attribute analysis to capture the global dependency of continuous attributes. Such global couplings integrate the intra-coupled interaction within an attribute (i.e. the correlations between attributes and their own powers) and inter-coupled interaction among different attributes (i.e. the correlations between attributes and the powers of others) to form a coupled representation for numerical objects by the Taylor-like expansion. This work makes one step forward towards explicitly addressing the global interactions of continuous attributes, verified by the applications in data structure analysis, data clustering, and data classification. Substantial experiments on 13 UCI data sets demonstrate that the coupled representation can effectively capture the global couplings of attributes and outperforms the traditional way, supported by statistical analysis.

1 Introduction

Real-world data sets predominantly consist of quantitative attributes in diverse domains [Saria *et al.*, 2011], such as finance and bioinformatics. The basic knowledge representation of numerical data is an information table [Kaytoue *et al.*, 2011], which comprises columns designating "attributes" and rows denoting "objects". Each table cell thus 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 independence of them.

Taking the fragment data of Iris (Table 1) as an example, 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

Table 1: A Fragment Example of Iris Data Set

Iris	Sepal.L (a_1)	Sepal.W (a_2)	Petal.L (a_3)	Petal.W (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

does not reflect any interaction with other attributes. Based on this classical representation, many data mining techniques and machine learning tasks [Plant, 2012; Li and Liu, 2012] including clustering and classification have been performed. One of the critical parts in such applications is to study the 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".

The above instance shows 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.*, 2011]. 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. 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 cou-
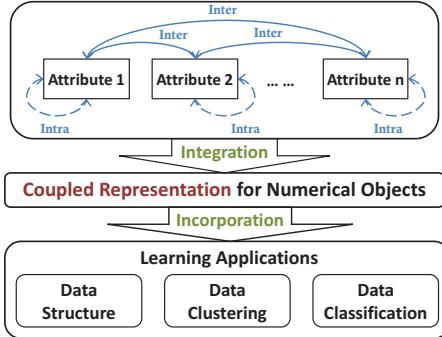
Figure 1: A framework of the coupled attribute analysis.

plings 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 1, if we adopt the method in [Kalogeratos and 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.

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 *et al.*, 2007] is a special case, since function log can be expressed by its Taylor series. For Table 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.

Accordingly, this paper proposes a framework of the coupled attribute analysis on numerical data, shown in Figure 1, to address the aforementioned research issues. The key contributions are as follows:

– 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 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.

– 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 paper is organized as follows. In Section 2, we briefly review the related work. Section 3 specifies the coupled interactions of numerical attributes. We formalize the coupled representation for objects in Section 4. The effectiveness of coupled representation is demonstrated in Section 5 with extensive experiments. Finally, we end this paper in Section 6.

2 Related Work

An increasing number of researchers point out that the independence assumption 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 and Bratko, 2003], but they involve the label information which is only eligible in supervised learning. While a rank-correlated measure [Calders *et al.*, 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 *et al.*, 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 among continuous attributes.

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-coupled 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 $a_j(1 \le j \le n)$; and $f = \bigcup_{j=1}^{n} f_j, f_j : U \to V_j$ is an information function which assigns a particular value of attribute a_j to each object. For instance, Table 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 $f_4(u_1) = 0.2$.

Based on S, we aim to capture the interactive relationships within a numerical attribute (intra-coupled) and among different continuous attributes (inter-coupled). A common method

Table 2: The Extended Information Table of Iris Data Set

$\tilde{A}$	$\langle a_1 \rangle^1$	$\langle a_1 \rangle^2$	$\langle a_2 \rangle^1$	$\langle a_2 \rangle^2$	$\langle a_3 \rangle^1$	$\langle a_3 \rangle^2$	$\langle a_4 \rangle^1$	$\langle a_4 \rangle^2$
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

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

$$Cor(a_j, a_k) = \frac{\sum_{u \in U}(f_j(u) - \mu_j)(f_k(u) - \mu_k)}{\sqrt{\sum_{u \in U}(f_j(u) - \mu_j)^2}\sqrt{\sum_{u \in U}(f_k(u) - \mu_k)^2}},$$
(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 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 and 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 $\langle a_j \rangle^p$ ($1 \le p \le L$) 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 2 is an extended information table of the original Table 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

$$R\_Cor(a_j, a_k) = \begin{cases} Cor(a_j, a_k) & \text{if p-value} < 0.05, \\ 0 & \text{otherwise.} \end{cases}$$
(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, $1 \le p, q \le L$, $a_j = \langle a_j \rangle^1$.

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

Definition 3.1 *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,*

$$\mathbf{R^{Ia}}(a_j) = \begin{pmatrix} \theta_{11}(j) & \theta_{12}(j) & \dots & \theta_{1L}(j) \\ \theta_{21}(j) & \theta_{22}(j) & \dots & \theta_{2L}(j) \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{L1}(j) & \theta_{L2}(j) & \dots & \theta_{LL}(j) \end{pmatrix}, \quad (3.3)$$

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 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 $(k \ne j)$. Accordingly, we have

Definition 3.2 *The **Inter-coupled Interaction** between attribute a_j and other attributes a_k $(k \ne j)$ is quantified as an $L \times L \cdot (n-1)$ matrix $\mathbf{R^{Ie}}(a_j | \{a_k\}_{k \ne j})$, 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,*

$$\mathbf{R^{Ie}}(a_j | \{a_k\}_{k \ne j}) = \quad (3.4)$$

$$\begin{pmatrix} \eta_{11}(j|k_1) \dots \eta_{1L}(j|k_1) \dots \eta_{11}(j|k_{n-1}) \dots \eta_{1L}(j|k_{n-1}) \\ \eta_{21}(j|k_1) \dots \eta_{2L}(j|k_1) \dots \eta_{21}(j|k_{n-1}) \dots \eta_{2L}(j|k_{n-1}) \\ \vdots \quad \ddots \quad \vdots \quad \ddots \quad \vdots \quad \ddots \quad \vdots \\ \eta_{L1}(j|k_1) \dots \eta_{LL}(j|k_1) \dots \eta_{L1}(j|k_{n-1}) \dots \eta_{LL}(j|k_{n-1}) \end{pmatrix},$$

where $\{a_k\}_{k \ne j} = \{a_{k_1}, \cdots, a_{k_{n-1}}\}$ 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 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 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 2, $\widetilde{\mathbf{u_1}}(a_4) = [0.20, 0.04]$, $\widetilde{\mathbf{u_1}}(\{a_1, a_2, a_3\}) = [5.50, 30.25, 4.20, 17.64, 1.40, 1.96]$.

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

Definition 4.1 *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,*

$$\mathbf{u^c}(a_j | \widetilde{A}, L) = \widetilde{\mathbf{u}}(a_j) \odot \mathbf{w} \otimes [\mathbf{R^{Ia}}(a_j)]^T \quad (4.1)$$

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

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 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 (4.1) 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

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

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

$$+ \frac{\theta_{pL}(j)}{L!} \tilde{f}_j^L(u) + \sum_{i=1}^{n-1} \frac{\eta_{pL}(j | k_i)}{L!} \tilde{f}_{k_i}^L(u), \quad (4.2)$$

---

[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.

Table 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

which resembles the Taylor expansion [Jia and Zhang, 2008] of functions. The right side of the above Equation (4.2) 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.

$$\mathbf{u^c}(a_j | \widetilde{A}) = \lim_{L \to +\infty} \mathbf{u^c}(a_j | \widetilde{A}, L). \quad (4.3)$$

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.

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

$$\mathbf{u^c}(\widetilde{A}, L) = [\mathbf{u^c}(a_1 | \widetilde{A}, L), \mathbf{u^c}(a_2 | \widetilde{A}, L), \cdots, \mathbf{u^c}(a_n | \widetilde{A}, L)]. \quad (4.4)$$

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 1, the coupled information table shown in Table 3, is the new representation.

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 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.

5 Empirical Study

In this section, several experiments are performed on 13 UCI data sets (i.e. Table 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

Table 4: Description of Data Sets

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
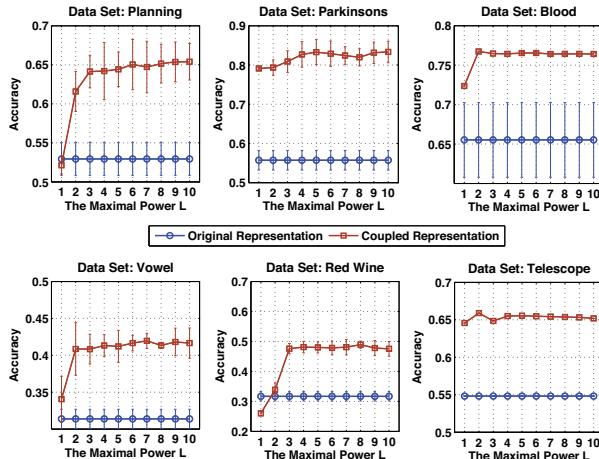
Figure 2: The performance of L on six data sets: the average accuracy with $\pm$ sample standard deviation error bars.

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.

5.1 Parameter Estimation

As indicated in Equation (4.2), 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 (4.2).

Figure 2 shows the performance of L on six data sets. 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
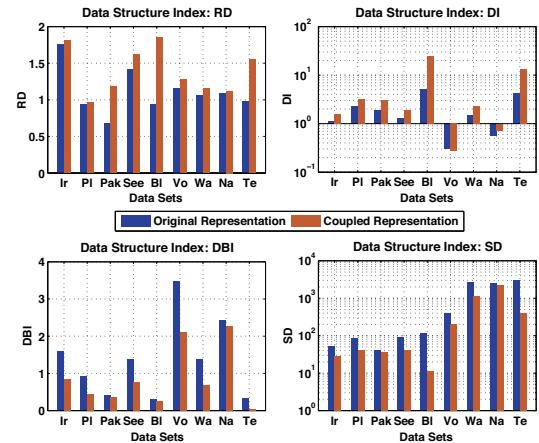

Figure 3: Data structure index comparisons on nine data sets.

the rest data. But the increasing rate of accuracy gets smaller as L grows. This is consistent with Equation (4.2), 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 coupled 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.

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. The data representation methods are evaluated with the given labels and the clustering internal descriptors: Relative Distance (RD), Davies-Bouldin Index (DBI) [Davies and Bouldin, 1979], Dunn Index (DI) [Dunn, 1974], and Sum-Distance (SD). In detail, RD is the ratio of average inter-cluster distance upon average intra-cluster distance; 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 the stronger cluster differentiation capability, which corresponds to a superior representation scheme.

The cluster structures produced by different representation schemes are analyzed on nine data sets. The normalized results are shown in Figure 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

Table 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**

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.

Data Clustering Evaluation

Two classical clustering approaches are single linkage algorithm (*LC*) [Ackerman and Ben-David, 2011] and spectral clustering (*SC*) [Luxburg, 2007]. These two methods are evaluated when incorporating the original (*LC-OR*, *SC-OR*) and coupled (*LC-CR*, *SC-CR*) representation schemes. The external clustering quality measures include Accuracy and Normalized Mutual Information (NMI). As described in [Cai *et al.*, 2005], the larger these indexes, the better the clustering.

Table 5 reports the results of the four approaches on six data sets in terms of external measures. The higher scores are highlighted in boldface, when *LC-CR* compared with *LC-OR* and *SC-CR* compared with *SC-OR*. This table indicates *LC-CR* and *SC-CR* respectively outperform their baseline algorithms *LC-OR* and *SC-OR* on both measures for almost all the data, except only one italic bold value. The maximal average improvement rate across all the data is 95.67%, while the minimal is 12.35%. Statistical testing also supports the results that *LC-CR* performs better than *LC-OR* and *SC-CR* outperforms *SC-OR*, at 95% significance level. Another interesting observation is that *SC* is mostly superior to *LC*, which also agrees 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. *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. As we know, a better data representation approach corresponds to a better classification result, i.e. higher 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 distinct representations are shown in Figure 4. *KNN* upon the coupled representation remarkably outperforms the original *KNN* for all the data sets in terms of all the evaluation measures. 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
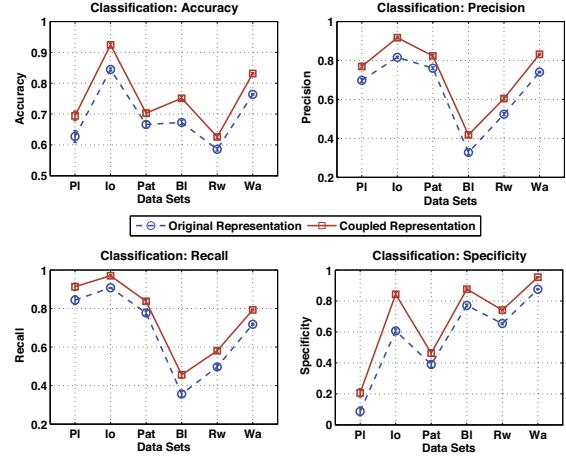
Figure 4: Data classification comparisons on six data sets: the average values with ± sample standard deviation error bars.

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.

6 Conclusion

We have proposed a coupled representation scheme for objects via teasing out the interactions of numerical attributes. Those interactions are quantified as the intra-coupled relationship described by the Pearson's correlations between attributes and their own powers, and the inter-coupled relationship characterized by the correlations between attributes and the powers of others. Both interactions are integrated to form the Taylor-like expansion based coupled representation for quantitative objects in a global way. The selection of the maximal power is empirically studied, reporting that $L = 3$ or 4 is large enough to capture the global coupling relationships. Substantial experiments have verified that the coupled representation outperforms the original on data structure, data clustering and classification, supported by statistical analysis.

We are currently enriching this framework of the coupled attribute analysis on numerical data by also addressing the coupling relationships for objects and clusters. In the future, we will work on modeling the coupling relationships for mixed data with both numerical and categorical attributes.

Acknowledgments

References

[Ackerman and Ben-David, 2011] M. Ackerman and S. Ben-David. Discerning linkage-based algorithms among hierarchical clustering methods. In *IJCAI 2011*, pages 1140–1145, 2011.

[Bollegala *et al.*, 2011] D. Bollegala, Y. Matsuo, and M. Ishizuka. Relation adaptation: learning to extract novel relations with minimum supervision. In *IJCAI 2011*, pages 2205–2210, 2011.

[Cai *et al.*, 2005] D. Cai, X. He, and J. Han. Document clustering using locality preserving indexing. *IEEE TKDE*, 17(12):1624–1637, 2005.

[Calders *et al.*, 2006] T. Calders, B. Goethals, and S. Jaroszewicz. Mining rank-correlated sets of numerical attributes. In *SIGKDD 2006*, pages 96–105. ACM, 2006.

[Cao *et al.*, 2011] Longbing Cao, Yuming Ou, and Philip S. Yu. Coupled behavior analysis with applications. *IEEE TKDE*, 24(8):1378–1392, 2011.

[Davies and Bouldin, 1979] D.L. Davies and D.W. Bouldin. A cluster separation measure. *IEEE TPAMI*, 1(2):224–227, 1979.

[Dunn, 1974] J.C. Dunn. Well-separated clusters and optimal fuzzy partitions. *Cybernetics and Systems*, 4(1):95–104, 1974.

[Figueiredo *et al.*, 2011] F. Figueiredo, L. Rocha, T. Couto, T. Salles, M. André Gonçalves, and W. Meira Jr. Word co-occurrence features for text classification. *Information Systems*, 36(5):843–858, 2011.

[Gan *et al.*, 2007] G. Gan, C. Ma, and J. Wu. *Data Clustering: Theory, Algorithms, and Applications*. ASA-SIAM Series on Statistics and Applied Probability, SIAM, Philadelphia, ASA, Alexandria, VA, 2007.

[Jakulin and Bratko, 2003] A. Jakulin and I. Bratko. Analyzing attribute dependencies. *PKDD 2003*, pages 229–240, 2003.

[Jia and Zhang, 2008] Y. Jia and C. Zhang. Instance-level semisupervised multiple instance learning. In *AAAI 2008*, pages 640–645, 2008.

[Kalogeratos and Likas, 2012] A. Kalogeratos and A. Likas. Text document clustering using global term context vectors. *Knowledge and Information Systems*, 31(3):455–474, 2012.

[Kaytoue *et al.*, 2011] M. Kaytoue, S.O. Kuznetsov, and A. Napoli. Revisiting numerical pattern mining with formal concept analysis. In *IJCAI 2011*, pages 1342–1347, 2011.

[Li and Liu, 2012] D.C. Li and C.W. Liu. Extending attribute information for small data set classification. *IEEE TKDE*, 24(3):452–464, 2012.

[Luxburg, 2007] U. Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):1–32, 2007.

[Nazareth *et al.*, 2007] D.L. Nazareth, E.S. Soofi, and H. Zhao. Visualizing attribute interdependencies using mutual information, hierarchical clustering, multidimensional scaling, and self-organizing maps. In *HICSS 2007*, pages 53–53, 2007.

[Plant, 2012] C. Plant. Dependency clustering across measurement scales. In *SIGKDD 2012*, pages 361–369, 2012.

[Saria *et al.*, 2011] S. Saria, A. Duchi, and D. Koller. Discovering deformable motifs in continuous time series data. In *IJCAI 2011*, pages 1465–1471, 2011.

[Wang *et al.*, 2011] C. Wang, L. Cao, M. Wang, J. Li, W. Wei, and Y. Ou. Coupled nominal similarity in unsupervised learning. In *CIKM 2011*, pages 973–978, 2011.