University of Technology, Sydney

FACULTY OF ENGINEERING AND INFORMATION TECHNOLOGY

Personalised E-Customer Relationship Management Models and System

Hua Lin

A thesis submitted for the Degree of Doctor of Philosophy of the University of Technology, Sydney

June, 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

ACKNOWLEDGEMENTS

During the preparation and completion of my research, several people have provided invaluable assistance and support over the past three and a half years. First, I would especially like to extend my respectful thanks to my principal supervisor, A/Prof. Guangquan Zhang of the University of Technology, Sydney, for all the time and effort he has put in, which has significantly contributed to my research and to the preparation of this thesis. His depth of knowledge and research experience was a true gift during the construction of this study. His critical comments, as well as his precious and patient guidance have strengthened each stage of this study significantly. It will always be a great privilege to have received his enduring guidance and encouragement.

I also would like to address my sincere thanks to my co-supervisor, Prof. Jie Lu, for her enormous help and support. Her generous knowledge and professional experience has been beneficial to this study in many aspects: research methodology, experiments, and thesis writing, which have greatly improved the final quality of this thesis.

I must also thank all the other faculty staff members for their participation in every presentation I have given, and for their valuable comments and critical questions during the course of this study. Specially, I would like to thank Ms. Sue Felix for helping me to identify and correct grammar and syntax problems in my thesis. I am also grateful to the many anonymous reviewers across the world for their valuable comments on each of the manuscripts submitted for publication. Last but not least, I would like to express my highest appreciation to my family: my husband and son. Without their encouragement, patience and support, my research could not have been completed.

ABSTRACT

Electronic Customer Relationship Management (eCRM), and analytical eCRM in particular, is currently one of the most active research topics in the area of customer marketing and customer analytics. The goal of this research is to develop an integrated analytical eCRM framework with a personalised intelligent recommendation approach through the analysis of telecom and banking products/services and customer needs. The development of the personalised recommendation approach combines two technology streams: 1) data mining techniques, which enable the prediction of customer buying behaviour patterns, and 2) fuzzy measure and community-based collaborative filtering recommendation techniques, which automatically combine the predicted buying behaviours and needs-preference requirements of customers to provide relevant and needs-based offers in an uncertain environment. The delivery of this framework effectively improve the quality of customer relationship management in terms of reducing the cost of customer acquisition, increasing customer retention and maximising customer lifetime value, which will enable and support a business strategy to build a long-term, profitable relationship with specific customers. The research contributes to both recommender system research and eCRM research and develops frontier technologies that can be applied across industries.

TABLE OF CONTENTS

TABLE OF FIGURES

List of Tables

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND, RESEARCH QUESTIONS AND MOTIVATION

Customer analytics has blossomed over the last decade, and the trend appears to be accelerating, because it has proven its value in many organisations, and in particular, drives the 'customer-centric' strategy of organisations. Customer analytics has made headway in marketing campaign management, segmentation, predictive modelling, measuring customer value and data mining. In recent years, customer analytics has become a hot area within the broader domain of Customer Relationship Management. Customer Relationship Management (CRM) simply means to manage all customer interactions. In practice, this requires using all the available information about existing and prospective customers to effectively interact with all customers at different stages of the organisation relationship with their customers. These stages have been referred to as the customer life cycle, and CRM helps organisations to improve the profitability of these interactions with customers (Coltman 2007; Stone, Woodcock & Wilson 1996).

Many organisations have invested in expensive CRM infrastructure which is mostly focused on sales and call centre operations; that is, operational CRM which can fulfil a wide variety of functions, including marketing automation (campaign management, cross-selling, customer segmentation, customer retention), sales force automation (contact management, lead generation, sales analytics, generation of quotes, product configuration) and call centre management (call management, integration of multiple contact channels, problem escalation and resolution, metrics and monitoring, logging interactions, etc.). This kind of operational CRM has been implemented and connected with customer touch points, and has worked effectively to simplify, organise and manage customer information to create a customer database that presents a consistent picture of the customer's relationship with the organisation, as well as provide specific application information (Bijmolt et al. 2010; Buttle 2012).

Over the past few years, organisations have realised that the raw data from each operational CRM system can be integrated, enabling the storage of a wealth of data about current customers, potential customers, suppliers and business partners in a data warehouse. However, the inability of the data warehouse to discover valuable information hidden in the data prevents the transformation of these data into insightful, meaningful and actionable knowledge. Also, the challenge is to transform the data into a set of customer-specific behaviour measurements and use detected customer behaviour patterns to analyse and choose channels to target customer marketing; further, to enhance the performance of front line communication and help organisations to leverage their existing investments and boost their return on investment. Data mining has played a key role in achieving this and has enjoyed great popularity in the CRM area (Rygielski, Wang & Yen 2002). In fact, data mining is a process that applies a variety of data analysis and statistical techniques to discover patterns and relationships in an enormous database that may be used to make accurate predictions. Analytical CRM uses data mining with certain built-in advanced analytics functions, such as predictive analytics and campaign optimisation, within the existing CRM system (Ngai 2005; Ngai, Xiu & Chau 2009).

Analytical CRM emerged in the last decade to reflect the central role of the strategic customer value proposition of an organisation. Analytical CRM takes a holistic view of customers: it encompasses all the measures for understanding customers and exploiting this knowledge to design and implement customer marketing activities, align production and coordinate the supply-chain. It has been proven that analytical CRM is particularly useful in measuring customer profitability and value, which can be built on a historic, potential or lifetime basis. This is an important part of customer analytics because, presumably, marketing investments increase customer value over time, and one way to measure the effectiveness of marketing is to understand what impact it has on customer value (Coltman 2007; Foss, Stone & Ekinci 2008; Hoekstra & Huizingh 1999).

Customers today increasingly contact organisations via email and Web, thus more and more organisations are implementing e-business applications and CRM strategies using the Internet. It is therefore necessary to provide an integrated solution for the centralisation of customer contact methods into a single enterprise system, such as a data warehouse. Electronic CRM (eCRM) is an efficient operational CRM system that integrates online sales, marketing and service strategies, enabling the identification, attraction and retention of customers (Romano & Fjermestad 2009; Turban 2008; Turban et al. 2007). Basically, eCRM involves all front-end, back-office, and third-party processes that 'touch' customers, from the contact centre that handles a customer's orders, for example, to the customer's bank for credit card authorisation, to the inventory system to check for product viability, and the warehouse for fulfilment and delivery. It improves and increases communication between an organisation and its customers by creating and enhancing customer interaction through innovative technology. The eCRM system works as a process to enable customers to do business with the organisation in the way the customer wants, at any time, via any channel, and to make all customers feel that they are dealing with a single, unified organisation that recognises them at every step of the way (Hwang 2009; Hwang & Kim 2007). At the same time, eCRM concerns all forms of managing relationships with customers by making use of IT; it is an enterprise that uses IT to integrate internal organisation resources and external marketing strategies to understand and fulfil customers' needs.

The key benefit of eCRM when compared with traditional CRM is that it enables an organisation to streamline processes and provide sales, marketing and service personnel with better, more complete customer information (Hadaya & Cassivi 2009). The result is that eCRM allows organisations to build more profitable customer relationships and decrease operating costs. Also, integrated eCRM systems are well documented. The better and more comprehensive the data that organisations can compile about their customers thorough automated customer management processes, the more effectively their global sales and customer support forces will be able to collaborate and share data in real time, manage that data throughout the enterprise, and access it at any time. The biggest challenge, therefore, is to offer communication and information on the right topic, in the right amount, and at the right time that fits the customer's specific needs.

Most eCRM applications are currently facing similar challenges, which include minimal research on analytics solutions that are able to guide customer interactions; lack of effective recommendation methods; lack of effective customer support and failure to enhance the 'human' quality of web experiences. An organisation invests in an eCRM system with the expectation that it will reduce the risk of customer churn and help to achieve the goal of customer retention. From this point of view, analytical eCRM offers the means to provide the most relevant and suitable products and services (hereafter referred to as 'products/services') to customers, through the right channel and at the right time, based on customer analytics. Analytical eCRM expands the traditional analytical CRM techniques by integrating the technologies of new electronic channels, such as Web, wireless and voice technologies. The analytical eCRM system can create a central business intelligence for customer and marketing analytics, and provides a portal on each front end computer system that allows access to customer information in the way that is insightful, meaningful and actionable; hence, front end staff can use this intelligent information to help customers to make the right decisions when buying products or services. By using this system, an organisation can build its capacity to understand customers, products and performance results by using real time information across the business (Bijmolt et al. 2010; Gustafsson, Johnson & Roos 2005; Kumar 2010; Nemati, Barko & Moosa 2004).

In the field of business-to-customer e-service, one problem in the current organisation of e-services is the tendency to only deliver undifferentiated product information to customers without addressing specific requests. Personalisation offers great opportunities to businesses to make communication more effective and efficient, to infer and predict customers' behaviour, and even to influence behaviour (Kim, Zhao & Yang 2008). Personalisation of business e-services can be seen as the adaptation of an e-service to a single customer based on related information about that particular customer, that is, the traditional personal one-to-one relationship. Therefore, personalised customer e-services can provide customers with exactly those services they need, further increasing customer satisfaction levels, which is a priority for organisations, and can then help them to achieve their CRM goals. A key function in the analytical eCRM framework is its recommendation ability, which is a new direction of business intelligence that gives business the advantage of being able to make offers to customers based on those customers' real needs; that is, to answer such queries as 'which product/service is the best a particular new or existing customer?'

This research is focused on answering the following questions;

(1) What is the appropriate analytics eCRM framework to support organisations to achieve an end-to-end integrated customer-centric online sales, marketing and service strategy?

(2) How does the proposed framework improve the quality of customer relationship management in terms of reducing the cost of customer acquisition, increasing customer retention and maximising customer lifetime value?

(3) How can an excellent customer service experience that guides customers to choose the most appropriate products/services be provided?

1.2 Objectives

The objective of this research is to develop a personalised analytical eCRM framework, and in particular to implement it by applying data mining techniques and recommendation techniques to support end-to-end integrated customer-centric online sales, marketing and service strategies. This research will also provide an excellent customer service experience to guide customers in the selection of the most appropriate products/services, as an important part of business intelligence (BI) across industries. The development of the personalised recommendation approach combines two stream technologies: 1) data mining techniques which enable the prediction of customer buying behaviour patterns, and 2) fuzzy measure and community-based collaborative filtering recommendation techniques that automatically combine the predicted buying behaviours and needs-preference requirements of customers to provide relevant and needs-based offers in an uncertain environment. The delivery of this framework will effectively improve the quality of customer relationship management in terms of reducing the cost of customer acquisition, increasing customer retention and maximising customer lifetime value, all of which support a business strategy to build a long term, profitable relationship with specific customers.

To achieve the overall objective, this research is comprised of four specific objectives:

(1) Propose an effective personalised analytical eCRM framework through data mining and recommendation techniques;

(2) Propose the intelligent customer behaviours predictive modelling frameworks for eCRM personalisation;

(3) Develop the novel hybrid recommendation approaches and related algorithms based on the combination of fuzzy measure-based collaborative filtering recommendation methods, and the combination of user-based and item-based collaborative filtering techniques with fuzzy set techniques;

(4) Build a working process of a personalised recommender system for the best recommendation, using a telecom business as a case study to demonstrate a real case implementation.

Aligned with the four specific objectives of this thesis, the main contributions of the thesis are:

(1) A personalised analytical eCRM framework;

(2) Two intelligent hybrid predictive modelling methodologies;

(3) Two hybrid recommendation approaches and related algorithms based on the combination of fuzzy measure-based collaborative filtering

recommendation methods, and the combination of user-based and item-based collaborative filtering techniques with fuzzy set techniques;

(4) A personalised recommender system to support intelligent personalised product/service recommendation approaches.

1.3 SIGNIFICANCE

This study has the following significance:

Significance 1: This research proposes a new research and development topic in business intelligence: personalised analytical eCRM

The most effective eCRM systems include active plans to capture new customers and retain existing customers. Used in conjunction with a decision-support technique, data mining techniques can process all customer information with mathematical precision to analyse customer relationship data and extract information, such as discovering correlations between demographic factors, income and product preferences, and the propensity for particular customer segments to respond to marketing campaigns. A personalised analytical eCRM using intelligent recommendation techniques is a relatively new phenomenon. It can be seen as a new development of Business Intelligence. Most current personalisation research efforts only focus on developing web personalisation, and most BI research focuses on data mining techniques. This research will integrate personalisation and BI to provide a strategic guide for eCRM project managers and developers. It will help to take business e-services to a new level - the ability to offer personalised services to customers. The research results will provide a more effective framework and a system to build personalised customer services, enabling improved customer satisfaction levels related to business products/services.

Significance 2: This research develops a novel intelligent product/service personalised recommender system that combines data mining techniques and recommendation techniques.

The proposed personalised recommender system has two novel distinguishing features:

(1) The system uses hybrid data mining techniques with fuzzy set to predict customer behaviours in a very sophisticated way, adopting a more comprehensive perspective on customer churn by examining ‘who are they’, ‘how do they behave’, ‘when are they likely to churn’, ‘why do they churn’, and ‘how can the organisation prevent customers from taking the decision to leave’. This developed predictive technique can be directly applied to predict customer buying patterns; it helps organisation to improve their understanding of customer behaviour. Hence, the system allows businesses to select the right prospects on whom to focus, offer the right additional products to existing customers, and identify good customers who may be about to leave. Most research has focused on a particular modelling technique to predict customer churn or take-up product, and very little has paid attention to the comprehensibility and justifiability of customer churn or buying prediction models. In practice, combination modelling methods are very tough, due to the magnitude of the observations and the variables. The proposed predictive method provides a new novel approach in the predictive analytics area.

(2) The system uses fuzzy measure-based collaborative filtering to deal with uncertainties in linguistic terms and item similarity in a hierarchical product system; also, it uses both item-based and user-based collaborative filtering methods with fuzzy set techniques and knowledge-based methods to handle the customer neighbourhood. As customers normally have multiple interests in receiving products/services, a properly defined neighbourhood of customers plays an important role in identifying preferences. Existing collaborative filtering approaches are inappropriate for dealing with such complex situations.

The proposed combined recommendation approaches provide a very effective way for personalised analytical eCRM to drive an organisation’s customer relationship management strategies.

Significance 3: This research directly supports Business Intelligence development and customer retention and acquisition.

This proposed Personalised Recommender System (PRS) for best recommendation can be tailored to specialise in the banking or telecom industry to develop a related implementation approach. The working recommender system can be applied to offer the most suitable products/services to both existing and prospective customers. It incorporates an extension of closed loop feedback into models and recommendations, and implement into any eCRM system would prove to add value to an organisation's customer marketing strategy. The system could help organisations to find their most suitable products/services for individual customers with minimal search effort and time. Also, by increasing customer satisfaction, it will in turn attract more customers.

1.4 RESEARCH METHODOLOGY

This research adopts design methodology, in which the problem is first formulated and the objectives are identified (see Section 1.3), following which a solution is proposed and implemented. Finally, the proposed solution will be tested and evaluated to ensure that the solution achieves the expected outcomes of this research. The evaluation process tests and modifies the solution iteratively until satisfactory results are obtained. Figure 1-2 gives the construction framework of this study.

Step 1 – Propose a personalised analytical electronic customer relationship management (eCRM) framework. (Refer to specific objective (1))

In this step, a conceptual personalised analytical eCRM framework is proposed which includes the main components and each function within those components in terms of defining related data mining and personalised models. It proposes a relevant data mining model to support each component of the analytical eCRM framework. The proposed framework includes a customer knowledge base, integrated analytical eCRM, personalised recommender system and necessary interfaces.

There are four functions that define an analytical eCRM system:

(1) Customer Acquisition – This phase involves profiling customers and classifying customers into different segments, and identifying the targeted segment of those who are most likely to become customers or to be most profitable to the organisation. Data mining model-clustering is a popular method of segmenting a population into a number of more homogenous clusters or segments. Common tools for clustering include neural network and discriminant analysis.

(2) Customer Marketing – Once potential customers are identified in (1), an organisation can put effort and resources into marketing the target customer segments, including setting up a multi-contact e-marketing strategy that includes, for example, email, SMS and internet. In this phase, data mining techniques can be used to forecast the effects of e-marketing (response rates to marketing campaigns); for instance, survival analysis, neural network and the Markov chain model can be used to estimate future value and uncaptured value.

(3) Customer Retention – This phase is the crucial element of analytical eCRM. Customer satisfaction is a key fundamental driver for customer retention. Customer satisfaction refers to the comparison of customers' expectations with customers' perceptions of being satisfied. Customer retention includes personalised marketing campaigns, loyalty programs and complaints management. Personalised marketing campaigns are driven by the analysis, detection and prediction of changes in customer behaviours, hence, customer profiling, propensity modelling, and recommender systems are related to personalise marketing. Loyalty programs involve campaign activities whose aim is to maintain a long term relationship with customers. Specifically, attrition analysis and service satisfaction support such loyalty programs.

(4) Customer Development – This phase involves the consistent expansion of transaction intensity, transaction value and individual customer profitability. Customer development includes customer lifetime value analysis, up/cross selling and market basket analysis. Customer lifetime value analysis is defined as the prediction of the total net income a company can expect from a customer. Up/cross-selling refers to promotion

activities which aim to augment the number of associated or closely related services that a customer uses within an organisation. Market basket analysis aims to maximise customer transaction intensity and value by revealing regularities in the purchase behaviour of customers. Sequence analysis aims to discover purchase patterns over time to understand which product/service the customer is likely to buy next (the so-called 'next best offer' model).

These four comprehensive functions can be seen as a closed loop in an integrated analytical eCRM system. They work together to create a deeper understanding of customers to maximise customer value in the long term. Data mining techniques can help to achieve this vision by extracting or detecting hidden customer characteristics and behaviours from large databases.

Step 2 – Development of a hybrid customer segmentation method. (Refer to specific objective (2))

Customer segmentation is an effective way to understand customers, and segmentation methodology has therefore been employed extensively in the last decade. Conventional data mining techniques, such as cluster analysis, have been used to deal with customer segmentation. Segmentation can be a starting point in relationship management to measure the true value of customers, since customer marketing as a whole is deployed toward targeted customers and profitable customers. However, customer segmentation faces the critical challenge of migrating the segmentation solution based on external data sources into an internal customer database. In this step, a hybrid customer segmentation model is proposed to tackle segmentation migration problems, and uses a telecom business as the study case.

Steps 3 and 4 – Propose two different customer risk prediction frameworks for customer retention based on banking and telecom cases. (Refer to specific objective (2))

In these steps, two different customer risk prediction frameworks are developed to deal with the non-linearity associated with customer attrition or customer churn by mining enormous amount of data. The use of different data mining methods and statistical modelling techniques to develop prediction models of customer attrition or customer churn is demonstrated, and their prediction power is compared using data based on bank and telecom cases.

Steps 5 and 6 – Propose two hybrid personalised intelligent recommendation approaches for acquisition and retention. (Refer to specific objective (3))

In these two approaches, the item-based collaborative filtering (CF) approach is first used to analyse the customer profile and product transaction behaviour matrix to identify similarity relationships between different products, and the user-based CF approach is then used to identify similarities between customers.

These two steps tackle the challenge of defining similarities in customer profiles, because unlike comparing two books, comparing two customers cannot simply be based on their names. The similarities between two customers will be based on their profiles and preferences, and because customers' preferences are subjective opinions, they are often expressed in linguistic terms by customers. For instance, a linguistic term set {Strongly Interested, More Interested, Interested, Less Interested, and Not Interested} can be used to express customers' preferences. Since these linguistic terms reflect the uncertainty, inaccuracy and fuzziness of humans, it is appropriate to directly apply fuzzy numbers techniques to deal with them. These hybrid recommendation approaches use these relationships to indirectly compute recommendations for customers, and can be described by four main steps: Calculate CF fuzzy similarity between products; Calculate semantic fuzzy similarity between customers; Integrate the item-based CF fuzzy similarity between products with the semantic fuzzy similarity between customers; Generate prediction values for recommendation.

Step 7 – Development of a personalised recommender system for the best recommendation. (Refer to specific objective (4))

To verify the proposed personalised analytical eCRM framework in Step 1 and the new predictive modelling techniques in Steps 3 and 4 in conjunction with the two new personalised intelligent recommendation approaches in Steps 5 and 6, a personalised recommender system is proposed that will generate the best recommendation, in terms of matching the right customer with the right product/service, at the right time and via the right channel. This development seeks to build a conceptual online e-service recommender system based on the specific personalised analytical eCRM framework developed in Step 1 and the new predictive modelling techniques proposed in Steps 3 and 4, using the two hybrid recommender system techniques proposed in Steps 5 and 6. The implementation of this personalised recommender system combines predictive models, recommender system techniques and business roles to optimise every customer contact at every touch point or interaction, which is an ongoing process that enables front line staff to provide relevant products and services to customers. This recommender system consists of three main components: 1) a Data Collector which collects customer preference and profile information, as well as product information; 2) a Database (DB) Builder which develops a product relevance database, customer profile database and customer ratings database; 3) a Recommendation Engine which generates a recommendation list of top-*k* products to individual customers according to their profiles and preferences.

Step 8 – Propose a personalised recommendation engine. (Refer to specific objective (4))

A personalised recommendation engine is designed to fulfil two functions. One function is to provide recommendations for new or prospective customers, and the other function is to provide recommendations for existing customers. The two functions work closely together based on predictive modelling, a fuzzy matching-based recommendation approach and hybrid fuzzy base recommendation approach, and certain business roles can be applied to this recommendation process to optimise the recommendation. Figure 1-1 shows the logics of the recommendation process.

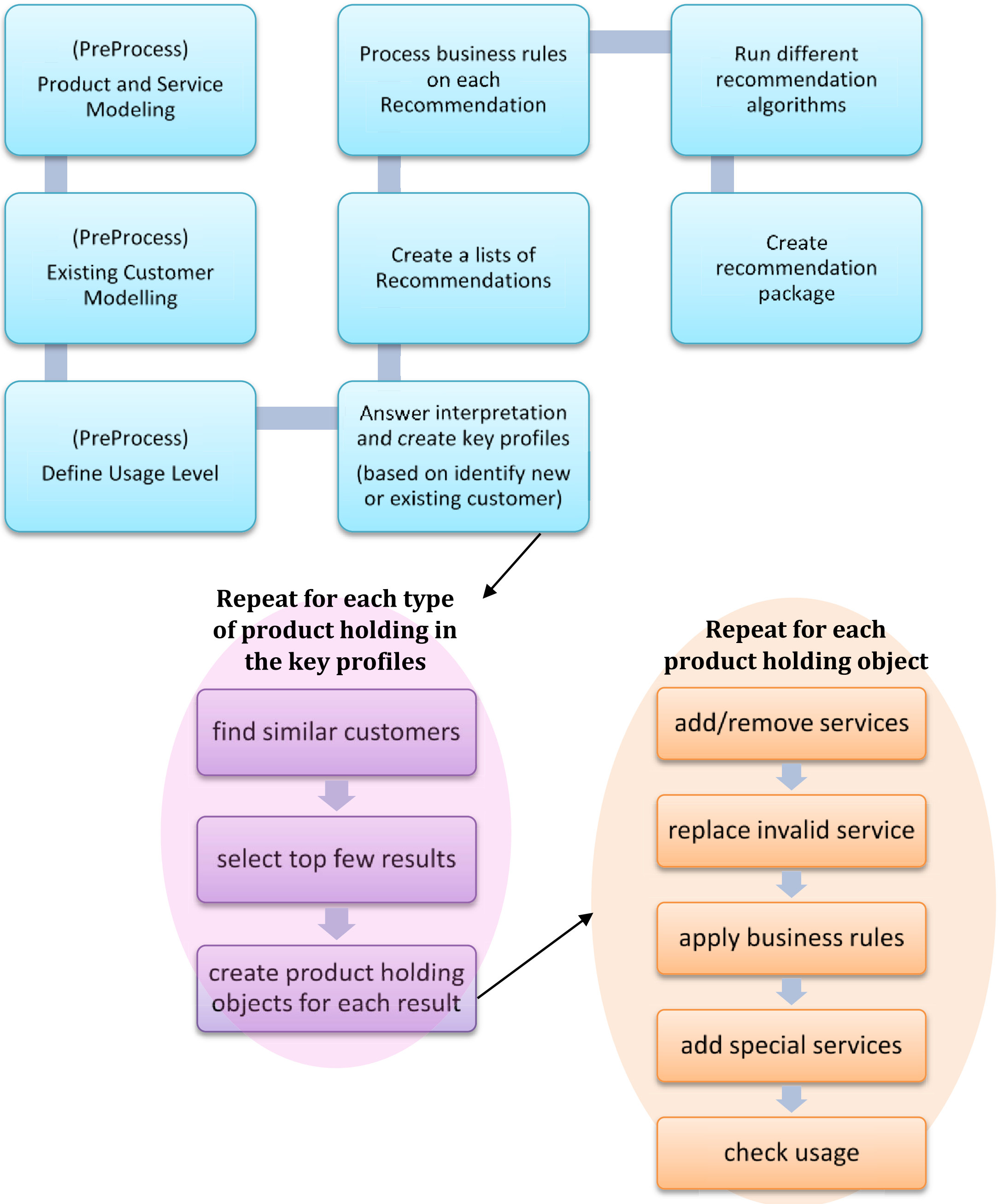


Figure 1-1 Recommendation Process Flow Chart

Step 9 – Experimental evaluation as a case study: fuzzy based telecom product recommender system architecture and development finalisation. (Refer to specific objective (4))

Related experiments are designed and conducted to evaluate the performance of the personalised recommender system using modified telecom data. The outcome results feed back into the system and all the algorithms can be refined in an ongoing refinement cycle designed to enhance the recommendation algorithms.

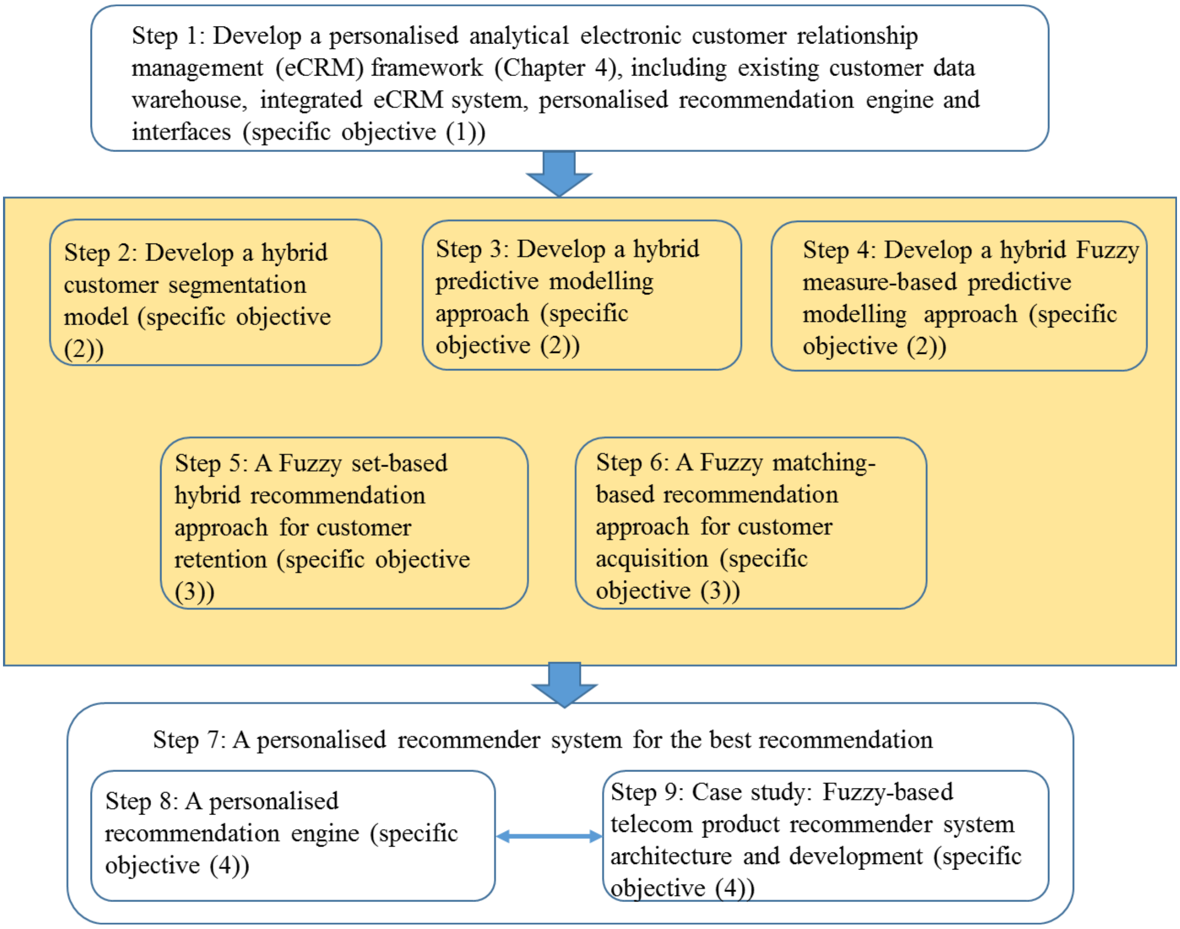


Figure 1-2 Research Methodology

1.5 Thesis Structure

This thesis consists of ten chapters (see Figure 1-3):

(1) Chapter 1 presents an overview of this research, including research issues, research objectives, research significance and research methodology.

(2) Chapter 2 reviews related research areas, and especially addresses the research background with regard to eCRM, data mining techniques, web personalisation, recommender systems, and fuzzy set techniques in recommender systems.

(3) Chapter 3 reviews a set of concepts as a preliminary, including decision tree, logistic regression, survival analysis, factor analysis and fuzzy set.
(4) Chapter 4 proposes a personalised analytical eCRM framework with a hybrid customer segmentation model that combines classification techniques and fuzzy set techniques to estimate the value of missing attributes.
(5) Chapter 5 demonstrates a prediction framework by developing a predictive modelling approach that combines three different modelling techniques to deal with the non-linearity associated with customer attrition when handling an enormous amount of data.
(6) Chapter 6 proposes a comprehensive advanced customer churn management model, including a customer churn profile model, a customer action model, a customer experience model and a fuzzy customer risk model.
(7) Chapter 7 presents the development of a personalised recommendation approach that combines both item-based and user-based collaborative filtering methods with fuzzy set techniques and knowledge-based methods.
(8) Chapter 8 proposes a fuzzy matching technique to deal with linguistic terms. This approach effectively matched customers' requirements to existing product features.
(9) Chapter 9 presents a Personalised Recommender System which combines the proposed predictive modelling approaches and personalised recommendation approaches developed in Chapters 7 and 8. A case study is demonstrated based on the implementation of the proposed personalised recommender system.
(10) Chapter 10 summarises the whole thesis and discusses future research topics and interests.

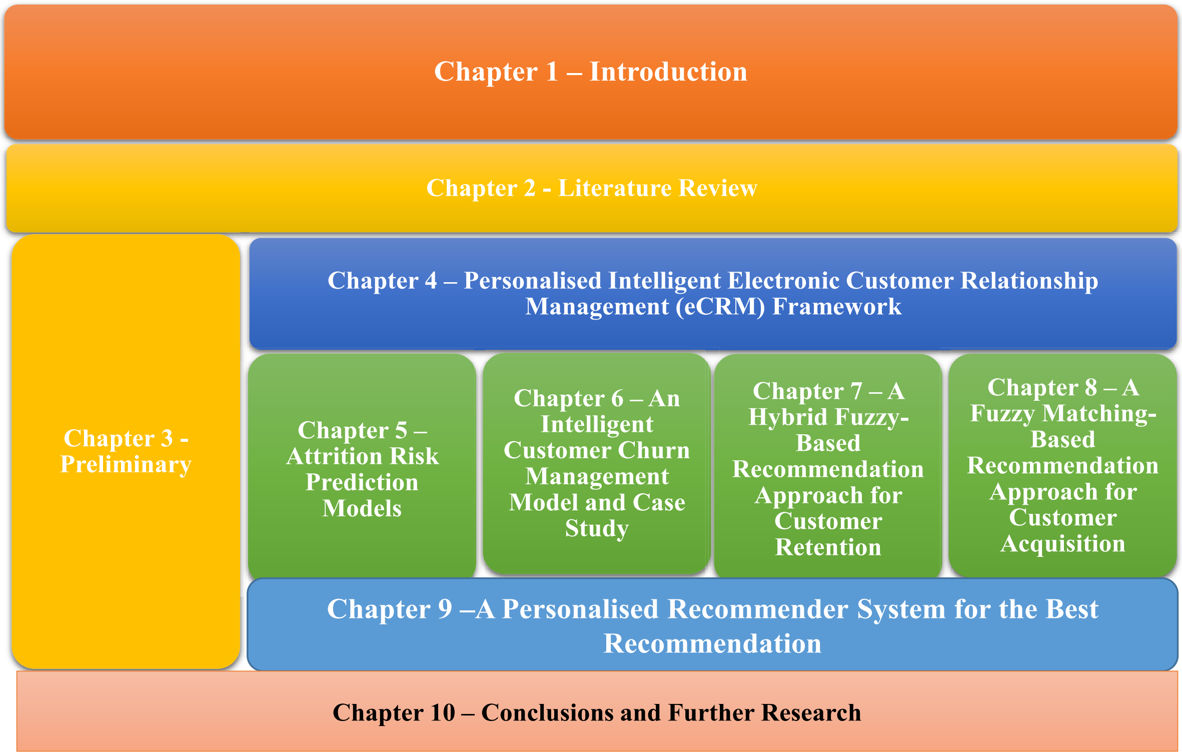


Figure 1-3 Thesis Structure

CHAPTER 2

LITERATURE REVIEW

2.1 ANALYTICAL ELECTRONIC CUSTOMER RELATIONSHIP MANAGEMENT

2.1.1 CUSTOMER RELATIONSHIP MANAGEMENT

Customer relationship management can be defined as "*a management approach that enables organisations to identify, attract and increase retention of profitable customers, by managing relationships with them*" (Bradshaw & Brash 2001). The definition does not mention process management technology or the means of communication or its channels, whether 'traditional' (mail, telephone, face to face) or 'new' (email, web, wireless devices, interactive television). In fact, CRM has become widely recognised as an important business approach, yet there is no universal definition of CRM. Customer relationship management (CRM) has also been defined an "*enterprise approach to understanding and influencing customer behaviour through meaningful communications in order to improve customer acquisition, customer retention, customer loyalty, and customer profitability*" (Swift 2001). Kincaid (2003) viewed CRM as "*the strategic use of information, processes, technology, and people to manage the customer's relationship with your company (Marketing, Sales, Service, and Support) across the whole customer life cycle*" (Kincaid 2003). Parvatiyar and Sheth (2001) defined CRM as "*a comprehensive strategy and process of acquiring, retaining, and partnering with selective customers to create superior value for the company and the customer. It involves the integration of marketing, sales, customer service, and the supply chain functions of the organisation to achieve greater efficiencies and effectiveness in delivering customer value*" (Parvatiyar & Sheth 2001). These definitions place emphasis on CRM as a comprehensive process of acquiring and retaining customers, with the help of business intelligence, to maximise customer value to the organisation. CRM is more than just a set of technologies; it comprises a set of processes and enables systems to support a business strategy to build long term, profitable relationships with specific customers (Ling & Yen 2001). IT professions will play a key role in supporting CRM with information and applications. In addition, the rapid growth of the Internet and its associated technologies has greatly increased the opportunities for marketing and has transformed the ways in which relationships between organisations and customers are managed (Ngai 2005).

A CRM system with build-in analytical function is defined as analytical CRM. Analytical CRM has been identified as a key aspect of a CRM business model, which is an enterprise-wide approach used to optimise the customer's experience and improve marketing efficiency. Databases are used to gather quantifiable data, and the ability to analyse and use this data to improve customer relationships is a driving force behind CRM adoption (Xu & Walton 2005).

2.1.2 ELECTRONIC CUSTOMER RELATIONSHIP MANAGEMENT

As the Internet becomes more and more important in business life, many organisations consider it as an opportunity to reduce customer-service costs, tighten customer relationships and, most importantly, personalise marketing messages. Electronic customer relationship management (eCRM) is being adopted by organisations to manage customer relationship via the Internet, as it increases customer loyalty and customer retention by improving customer satisfaction, which results in long-term profits for online retailers because they incur fewer costs in recruiting new customers, plus they experience an increase in customer retention. Also, electronic methods are used to gather data and analyse customer information (Dhingra & Dhingra 2013). Hence, analytical techniques such as data mining, predictive modelling and supporting decision making, particularly in eCRM, have increasingly been in high demand in recent years (Mahdavi et al. 2008) (Pan & Lee 2003).

With the current turbulent economic environment and highly competitive market, eCRM plays an important key role in retaining customers and maximising profit in many different industries. In the banking industry, for instance, customer retention is one of the key drivers of business strategy, so eCRM becomes more important than ever before. The difference between eCRM and normal CRM is that all the traditional customer contact methods are used in addition to use of the Internet, email, wireless technologies, and systems (created for external use) that are designed according to customer needs, and web applications that are designed for enterprise-wide use. Another important difference is that personalised individual views based on purchase history and preferences can be presented, and individuals have the ability to customise the view. The advantages of eCRM are reductions in time and cost, and the fact that implementation and maintenance can take place at one location and on one server (Adebanjo 2003; Romano & Fjermestad 2009).

eCRM consists of processes that an organisation uses to track and organise its contacts with current and prospective customers. Business intelligence (BI), particularly intelligent eCRM software, is normally in place to support these processes. In other words, eCRM can be defined as those activities that manage customer relationships by using the Internet, web browsers or other electronic touch points (Harrigan, Ramsey & Ibbotson 2011). The challenge lies in how to offer communication and information on the right topic, in the right amount, and at the right time, to fit the customer's specific needs (Kumar 2010).

2.1.3 Analytical Electronic Customer Relationship Management

In Section 2.1.2, it has been identified that eCRM works as a process to enable a customer to do business with an organisation in the way the customer desires. It is very important to make the customers feel that they are dealing with a single, unified organisation, and that the organisation understands their needs at every step of the interaction.

It is quite challenging to create an effective eCRM system within an organisation, as it is a new way of doing business. Internally, organisations want to decrease costs and streamline business processes. Externally, organisations must maintain relationships with customers. To achieve both these goals, data mining techniques have achieved great popularity in eCRM, which uses a data mining process defined as analytical eCRM and allows organisations to learn more about their customers (Azila & Noor 2011; Chang & Wu 2011; Van Dyke, Nemati & Barko). The development of new capabilities in analytical eCRM now promise even more integration of diverse techniques. Organisations are moving aggressively into eCRM, and will therefore appreciate powerful analytical eCRM frameworks to drive their day-to-day business processes (Nemati, Barko & Moosa 2004). There are four important elements in an analytical electronic customer relationship management system;

1) The first phase is customer acquisition, which involves profiling prospective customers and subdividing them into different segments, identifying which are most likely to become customers and to be the most profitable, and understanding the key characteristics of each customer segment. To achieve this, a data mining technique such as clustering (Adebanjo 2003; Hillenbrand & Money 2009; Hoekstra & Huizingh 1999; Hosseni & Tarokh 2011; Khandelwal & Mathias 2011; Lefait & Kechadi 2010; Pesonen 2012; Takano et al. 2010; Teichert, Shehu & von Wartburg 2008; Tsai & Chiu 2004) is a popular method for segmenting populations into a number of more homogenous clusters or divisions (Buttle 2012; Namvar, Gholamian & KhakAbi 2010).
2) The second phase is customer marketing, which puts effort and resources into attracting target customer segments after identifying the various groups of potential customers. This involves setting up a multi-contact e-marketing strategy. It is the promotion process for motivating customers to place orders through various electronic channels, including email, SMS and Internet (online). In this phase, data mining techniques can be used to forecast the effects of direct marketing (the response rates of marketing campaigns); for instance, survival analysis, neural network or Markov chain model (Au, Guangqin & Rensheng 2011; Baesens et al. 2002; Efron 1988; Miller Jr 2011; Ranjan & Bhatnagar 2011; Sharma & Panigrahi 2011; Smith & Gupta 2003; Tan, Sim & Yeoh 2011)

can be used to estimate customers' future value and uncaptured value (Cespedes & Smith 2012; Kim & Street 2004; Weinstein 2004; Wilson & Gilligan 2012).

3) The third phase is customer retention, which is the crucial part of eCRM. Customer satisfaction is the essential element of customer retention; it refers to the comparison of customers' expectations with customers' perceptions of being satisfied. There are several contributing elements to customer retention, such as personalised marketing campaigns, loyalty programs and complaints management. Personalised marketing campaigns are driven by analysing, detecting and predicting changes in customer behaviours, hence, customer profiling, propensity modelling, and recommender systems are related to personalised marketing, while loyalty programs aim to maintain a long term relationship with customers (Ngai, Xiu & Chau 2009; Nitzan & Libai 2011; Rust & Zahorik 1993; Sashi 2012; Stahl et al. 2012; Tabaei & Fathian 2012; Thorleuchter, Van den Poel & Prinzie 2012; Van den Poel & Larivière 2004; Verhoef 2003; Wang & Lei 2010; Wu 2011).
4) The fourth phase is customer development, which involves maximising individual customer profitability in terms of consistent expansion on transaction intensity, transaction value and customer lifetime value, including up/cross-selling propensity analysis, customer lifetime value analysis and market basket analysis. Market basket analysis aims to maximise the customer transaction intensity and value by revealing regularities in the purchase behaviour of customers. Sequence analysis and up/cross-selling propensity modelling aim to discover purchase patterns over time to understand which product/service the customer likely to buy next: the so-called next best offer model (Bijmolt et al. 2010; Chen & Fan 2013; Hoekstra & Huizingh 1999; Hosseni & Tarokh 2011; Hwang, Jung & Suh 2004; Liu & Yang 2012; Reinartz & Kumar 2003; Sashi 2012; Shih, Liu & Hsu 2010; Tabaei & Fathian 2012; Thorleuchter, Poel & Prinzie 2010; Tsai & Chiu 2004; van Bentum & Stone 2005; Verbeke et al. 2012; Wang & Lei 2010; Zorrilla et al. 2011).

These four phases can be worked together as a closed loop within an analytical electronic customer relationship management system to create a deeper understanding of customers and customer behaviours, to maximise customer value in the long term.

The first two business areas to directly benefit from such a system are sales and marketing. Sales people are given access to customer information to assist them in up-selling and cross-selling, while customer needs can be identified by data mining processes. These findings can then be shared throughout the organisation to create a complete view of each customer. At the same time, marketing managers are provided with new ways to generate leads and target advertising to appropriate prospects, a process which is also heavily based on data mining (Nemati, Barko & Moosa 2004; Ngai, Xiu & Chau 2009).

2.2 Data Mining Related Methods In Customer Analytics

Data mining is defined as a process that *"uses statistical, mathematical, artificial intelligence and machine-learning techniques to extract and identify useful information and subsequently gain knowledge from large databases"* (Turban et al. 2007).

Data mining is a process that uses a variety of data analysis and modelling techniques to discover patterns and relationships in data that may be used to make accurate predictions. It helps organisations to improve their understanding of customer behaviour, and this better understanding allows them to select the right prospects on whom to focus, to offer the right additional products to existing customers, and to identify good customers who may be about to leave (Peacock 1998; Rygielski, Wang & Yen 2002; Thorleuchter, Poel & Prinzie 2010; Wong et al. 2005).

2.2.1 SEGMENTATION

To achieve success in contested turbulent markets, a best practice approach to marketing activities will drive superior returns. Customer insight is at the heart of this approach, because all customers are different, and translating customer insight, segmentation and modelling into an end-to-end execution is the key step to putting an organisation on the path to success (Wilson & Gilligan 2012; Wu et al. 2005; Wu 2011; Zhang 2010). Segmentation is the basis for an organisation's marketing campaign managers to develop targeted and effective marketing plans. Furthermore, analysis and profiling of defined segments enables decisions to be made about the intensity of marketing activities in particular segments.

It is essential to develop a segmentation framework with clear business objectives and a systematic approach, which is both insightful and practical, to form a solid foundation for an organisation's marketing strategy. A segment-orientated marketing approach generally offers a range of advantages for both business and customer (Amiri 2006; Pesonen 2012; Sashi 2012; Takano et al. 2010; Wang & Lei 2010; Weinstein 2004; Woo, Bae & Park 2005).

It has been found that some segmentations are too broad or too abstract to be useful. In many cases, highly specific segmentation would be preferable; for example, business wants to know why customers take up one product but not another; which behaviour patterns drive customers to buy; and according the behaviour pattern, which kinds of products or services can be sold to customers. Hence, segmentation should be built upon a sophisticated propensity model to predict the customers who are most likely to take up a particular product (Buttle 2012; Chen et al. 2007; Han, Lu & Leung 2012; Hillenbrand & Money 2009; Hosseni & Tarokh 2011; Kim, Jung, et al. 2006; Mazzoni, Castaldi & Addeo 2007; Miguéis, Camanho & Falcão e Cunha 2012).

Customer segmentation is a crucial, fundamental methodology for understanding customers. It is a key function of customer relationship management (CRM) which enables businesses to provide customers with expected products and better services in competitive markets (Cooil, Aksoy & Keiningham 2008). Customer segmentation has been extensively studied across all industrial sectors, such as telecom (Mazzoni, Castaldi & Addeo 2007; Miguéis, Camanho & Falcão e Cunha 2012), finance (Hillenbrand & Money 2009), insurance (Hosseni & Tarokh 2011), information and communication technology (ICT) facilities (Gil-Saura & Ruiz-Molina 2009), the airline industry (Keramati & Ardabili 2011), healthcare (Khandelwal & Mathias 2011), and tourism (Kim, Wei & Ruys 2003). Methods are built on both internal organisation datasets and external datasets, such as samples from historical customer transactions in database or data warehouse (Lefait & Kechadi 2010), demographic and socioeconomic data from government agencies, and customer surveys (Amiri 2006; Keramati & Ardabili 2011; Miguéis, Camanho & Falcão e Cunha 2012). Internal data sources are veritable glod mines of customer profiles and behaviour patterns; however, they are often missing some several important features for effective customer segmentation. The ideal case is for both datasets to provide similar customer-related attributes; however, this is quite a challenge in real world applications. Hence, many such methods have *"fallen flat when used in marketing and advertising campaigns"* (Kim, Jung, et al. 2006).

Customer segmentation is a process to discover and *"recognise groups of customers who share the same or similar needs"* (Lefait & Kechadi 2010). Essentially, customer segmentation is a cluster analysis; therefore, the majority of existing methods are based on clustering techniques and algorithms; for example, the K-means clustering algorithm. Experience description and statistical analysis also form the main basis of a segmentation method (Chen et al. 2007); however, because combining multiple clustering or learning algorithms often outperforms a single algorithm, many works adopt more than one algorithm in a method.

The key challenge is to migrate customer segmentation that has been developed and defined from one data source to another. The design of a customer segmentation method is based on customer demography, geographic location, behaviours, cost benefit relations, and lifestyle. These variables may change frequently and affect the consistency of a segmentation (Lefait & Kechadi 2010). Customer segmentation can also be conducted based on customer attitudinal data, which focuses on customer perceptions. The attributes from an attitudinal dataset contain a huge number of subjective expectations or preferences for customers, which are hard to find in a customer behaviour dataset. Currently, there is no significant research on migrating segmentation results from one dataset to another.

2.2.2 PREDICTIVE MODELS

More and more organisations are changing their product focus from mass marketing champion strategies to customer-centric, targeted marketing strategies. Customer retention management is an important discipline within customer-centric organisations, for which a key role is the prevention of customer churn, that is, the propensity of customers to end the relationship with the organisation. As it is more profitable to retain existing customers than constantly attract new customers (Coussement, Benoit & Van den Poel 2010; Coussement & Van den Poel 2008a, 2008b; Verbeke et al. 2012), customer churn or attrition management has attracted intense attention across industries. The term 'churn' or 'attrition' refers to the loss of customers who switch from one provider to another during a given period. In the banking industry, customer attrition analysis and customer attrition rates have been used as key business metrics, because the cost of retaining an existing customer is far less than the cost of acquiring a new one. Also, lost customers can create an exponential revenue impact through viral, negative word of mouth. As a result, understanding the key factors that lead to customer attrition is crucial. Banks, for example, can use this understanding as a basis to effectively drive customer retention programs to prevent customers from leaving (Parasuraman 1997) (Ganesh, Arnold & Reynolds 2000; Larivière & Van den Poel 2004, 2005).

The attrition prediction models that predict the likelihood of customer churn that has become such a hot topic across industries, can be used to assess the risk of a customer churning. In the banking industry, for example, customers at risk are those who are likely to close their accounts with their existing bank. The attrition prediction models can generate a small prioritised list of potential 'at risk' customers; the bank can then focus proactively on customer retention marketing programs. Furthermore, attrition prediction models also can be used to drive cost-effective customer retention programs which only target the subset of the customer base who are most likely to leave (Hung, Yen & Wang 2006).

The practice of customer churn prediction is a very important approach to improve customer retention. In the last decade, researchers have devoted effort to the evaluation of different modelling techniques for the prediction of customer churn, such as a classification model that has been proposed to identify those customers most likely to demonstrate churning behaviour (Xie, Li, Ngai, & Ying, 2009). Technically, customer churn prediction involves binary classification, which generalises the relationship between churning behaviours and the information describing the customer in a model that can be used for prediction purposes. In the literature of predictive models, a variety of learning methods, such as Neural Networks (Datta et al. 2000; Gedikli & Jannach 2013; Sharma & Panigrahi 2011), Clustering (Abbasimehr, Setak & Tarokh 2011; Popović & Bašić 2009), Decision Tree (Nie et al. 2011), Regression (Owczarczuk 2010), generalized additive models (GAMs) (Guisan, Edwards Jr & Hastie 2002), Support Vector Machines (De Bock & Poel 2011) (Coussement & Van den Poel 2008a; Kim et al. 2012), and ensemble of hybrid methods (Au, Guangqin & Rensheng 2011) Kumar and Ravi 2008) have been used as basic models to handle customer churn or attrition. In addition, existing studies on customer churn have taken a variety of directions, such as building predictive models (Chen, Fan & Sun 2012; Kim et al. 2012), exploring new features for churn prediction (Coussement & Van den Poel 2008a; Huang, Kechadi & Buckley 2012), and dealing with class imbalance (Burez & Van den Poel 2009). There are two limitations in the current research and practice of customer churn prediction and management in a real-world application:

(1) Even though the direction taken by the methods mentioned above have been different, more have focused on a particular modelling technique, less have paid attention to combination modelling methods, and the objectives of most studies are identical, that is, they all aim to provide more accurate churn prediction. However, organisations not only face the question of how to identify potential churners, they also need to find an appropriate way to convert this risky

relationship into a stronger relationship. It would appear that the latter topic has received relatively little research attention;

(2) According to a literature review on customer churn management, fuzzy logic is one other area of soft computing that has never been investigated in either churn management or CRM research (Hadden et al. 2007). Due to the fuzziness of the problem, a fuzzy risk assessment model is required to manage customer churn effectively.

2.3 PERSONALISED RECOMMENDER SYSTEMS

One of the most important enhancements to analytical eCRM that could meet the growing needs and preferences of customers would be to offer personalised services. Personalised intelligence is widely used in e-commerce applications and has achieved a many improvements and benefits for both customers and service providers in terms of service delivery. It aims to provide customers with what they want without asking them explicitly about their needs (Goy, Ardissono & Petrone 2007) and can provide the most relevant information expected to meet customers' demands (Schubert, Uwe & Risch 2006; Sunikka & Bragge 2008). A typical examples in e-commerce websites is Amazon (www.amazon.com).

2.3.1 CONCEPT OF RECOMMENDATION SYSTEMS

A recommendation system is the most popular technique for implementing personalisation (Deshpande & Karypis 2004; Ricci & Shapira 2011). Such systems can be defined as programs which attempt to recommend items to customers by predicting a customer's interest in an item based on various types of information. The aim of a recommendation system is to provide the right information about products/services that are relevant to the needs and interests of the right customers at the right time. This can be achieved by automatically filtering out unrelated products and suggesting only relevant products (Goy, Ardissono & Petrone 2007). In a recommender system, the information used to make recommendations is acquired from the demographic data of customers; from an analysis of the past purchasing behaviour of customers as a prediction for future buying behaviour; or from the top overall sellers on a site (Schafer, Konstan & Riedi 1999). A recommendation system can either predict whether a particular customer is interested in a particular product, or it can identify a set of products that a particular customer may be interested in (Deshpande & Karypis 2004). Most personalisation research efforts focus on developing web personalisation applications by using recommendation systems in the e-commerce domain (Goy, Ardissono & Petrone 2007; Schubert, Uwe & Risch 2006). Very few researches have discussed personalisation applications in the eCRM field.

Recommender systems can be designed to automatically make helpful recommendations to customers across a range of products and services (Ricci & Shapira 2011). Such systems can make recommendations according to customer profiles or preferences, or they can rely on the choices of other people who could be useful referees. The advantage of recommender systems is that they suggest the right items (products or services) to particular users (customers, suppliers, salespeople, etc.) based on their explicit and implicit preferences by applying information filtering technologies (Manouselis & Costopoulou 2007).

Three main types of recommendation system techniques are content-based, collaborative-filtering and knowledge-based techniques (Burke 2002).

2.3.2 CONTENT-BASED RECOMMENDATION TECHNIQUES

Content-based (CB) recommendation techniques recommend items that are similar to items previously preferred by a specific customer. The basic features of CB recommendation systems are that they:

(1) Analyse the description of the preferred items by a particular customer to discover the main attributes (preferences) in common that can be used to distinguish these items. The attained preferences are stored in a customer profile;
(2) Compare each item's attributes with the customer profile so that as a result only items that have a high degree of similarity with the customer profile will be recommended (Pazzani & Billsus 2007).

Two well-developed CB recommendation techniques are information retrieval methods, such as cosine similarity measure, and machine learning methods that learn customers' interests from the customers' historical data (training data) using decision tree, naïve Bayesian and k-nearest neighbour methods. The advantages of this latter type of recommendation system is that it adopts the semantic content of items and recommends items to a specific customer that are similar to the preferred items in his/her profile. As a result, a CB recommender system can recommend new items and less popular items. Furthermore, it can provide a clarification of recommended items by listing the content features on the basis of which an item is to be recommended. It does not need to have information about the preferences of other customers to make recommendations and thus does not suffer from the sparsity problem associated with collaborative-filtering systems. However, these methods tend to rely heavily on textual descriptions of items, leading to several unsolved problems such as limited information retrieval, new customer problems, and overspecialization. Hence, it is not able to offer accurate recommendations for new customers who have very few rated items, and is thus challenged in meeting the objective that the companies always need to provide appropriate products/services to new customers (Nazim Uddin, Shrestha & Geun-Sik 2009).

2.3.3 Collaborative-Filtering Recommendation Techniques

Collaborative-filtering (CF) recommendation techniques help people to make choices based on the opinions of other people who share similar interests and therefore try to provide the right information to the right customer (Shardanand & Maes 1995). It has been proven that the CF recommendation approach is the most successful and widely used approach for recommendation systems (Huang, Zeng & Chen 2007; Schafer et al. 2007). Unlike CB methods, CF methods do not involve user profiles and item features when making recommendations. CF methods help people to make *t* choices based on the opinions of other people who share similar interests (Sinisalo et al. 2007). Collaborative-filtering-based recommendation systems have been developed and used in many fields including recommending news, articles, movies, music, products, books, web pages and many more. Literature has reported a number of CF algorithms that can be used to generate recommendations, and they can be divided into three main types: user-based (memory-based), item-based, and model-based CF (Schafer et al. 2007).

(1) The user-based algorithms are formally known as the nearest neighbour algorithms (Sarwar et al. 2001). These algorithms recommend new items to a particular user using statistical-based methods. To use these algorithms, all items and user ratings are stored in the memory, hence these algorithms are called user-based or memory-based.
(2) The item-based CF algorithms suggest new items to a particular user. They aim to recommend a new item, the *i*th item, which has not been rated by the target user.
(3) The model-based CF algorithms use existing ratings to build a model which is then used to make predictions for unrated items.

The most popular CF methods are user-based CF and item-based CF (Sarwar et al. 2001). As explained above, a user-based CF method uses the ratings of customers that are most similar to the target user (customer) to predict the ratings of unrated items. More specifically, when making a recommendation, the user-based CF recommender system will first calculate the similarities of all customers to the target user by analysing the previous ratings of all users. The system will then select a certain number of most similar customers as references, following which it will use the ratings of the selected customers on the target item (the unrated item of the target user) to predict the rating of this item for the target customer. By contrast, the item-based CF method uses the similarities of items to predict ratings. The major limitations of CF methods are the cold start problem for new customers and new items, the sparsity problem (Adomavicius & Tuzhilin 2005), and the long tail problem (Park & Tuzhilin 2008). These problems have attracted much attention from researchers. A kernel-mapping recommender has been proposed (Ghazanfar, Prügel-Bennett & Szedmak 2012), and the recommendation algorithm performs well in handling these problems. Park et al. (2008) used a clustering method to solve the long tail problem.

Various machine learning algorithms are used to accomplish the model building process. The main advantages of using CF recommendation techniques are that they work for any type of item without the need to extract the features related to items. These techniques not only suggest similar items to match the customer's interests and preferences, they also suggest new items based on the choices of other people who have the same or similar tastes and interests as the customer. The major limitations of CF methods include sparsity, scalability and cold-start problems (Adomavicius & Kwon 2012; Adomavicius & Tuzhilin 2005; Schafer et al. 2007). In a user-based CF algorithm, computing a customer's neighbourhood is expensive because it is necessary to perform a comparison with all other customers in real time. The number of customers and items in existing systems can reach thousands or even millions, and to locate the potential neighbours of a specific customer, it is to search the entire dataset. Therefore, the time and memory requirements of user-based CF algorithms grow linearly with the number of customers and items (i.e. a long computation time is required to find a neighbourhood). Clearly, this scalability problem needs to be overcome in existing applications. The cold-start problem refers to the inability of a CF approach to make useful recommendations for both new customers and new items (Papagelis & Plexousakis 2005; Schafer et al. 2007). Organisations often develop new products/services and have new customers at the same time, and in such situations, there is a lack of ratings based on previous purchases and lack of ratings for new items, so it is difficult to make good recommendations about such items. The content-based (CB) methods and collaborative filtering (CF) methods are the most popular techniques adopted in recommender systems (Iaquinta et al. 2007).

2.3.4 KNOWLEDGE-BASED RECOMMENDATION TECHNIQUES

Knowledge-based (KB) recommendation techniques offer items to users based on knowledge about both users and items. Usually, KB recommendations retain a functional knowledge base that describes how a particular item meets a specific user need, which can be performed based on inferences about the relationship between a user's need and a possible recommendation (Burke 2002). Case-based reasoning technique is the most common example of a KB recommendations technique (Smyth 2007). Knowledge-based recommendation systems can overcome the cold-start problem because new customers can obtain recommendations based on simple knowledge of their interests. Recommendations are generated by computing the similarities between existing cases and the customer's request, so the customer is not required to rate or purchase many items to generate good recommendations.

The KB recommendations also can be based on business knowledge (business rules) and inferences about a customer's needs and preferences, and because a KB system has functional knowledge about how a particular item meets a particular customer need, it is able to reason about the relationship between a need and a potential recommendation (Burke 2002; Manouselis & Costopoulou 2007). Some KB systems employ case-based reasoning techniques for recommendation. These types of recommenders solve a new problem by looking for a similar past solved problem. The KB approach has some limitations, however; for instance, it needs to retain information about items and customers, as well as functional knowledge, to make recommendations. It also suffers from the scalability problem because it requires more time and effort to calculate the similarities in a large case base than other recommendation techniques (Goossen et al. 2011; Wu et al. ; XiaoYan, HongWu & SongJie 2008).

2.3.5 OTHER RECOMMENDATION TECHNIQUES

The hybrid-based recommendation approach is a combination of two or more of the aforementioned approaches to emphasize the strengths of these approaches and to achieve the peak performance of a recommender system (Adomavicius & Tuzhilin 2005). Burke (2001) proposed a classification of hybrid recommender systems, listing seven basic hybridization mechanisms for building such systems (Burke 2002). Iaquinta et al. (2007) incorporated CB methods into a CF model for calculating customer similarities, using customer profiles built using machine learning techniques. Su et al. (2007) built a model using multiple experts including both CB and CF approaches which adopted different strategies in different situations. All these methods are largely based on the rating structure. To increase the accuracy and performance of recommender systems, many researchers have tried non-ratings techniques such as data mining, machine learning and intelligent agents, according to the circumstances (Hofmann 2003; Pazzani & Billsus 2007). For example, Su et al. (2007) proposed a sequential mixture CF (SMCF) which first uses the predictions from a TAN-ELR (Greiner et al. 2005) content-based predictor to fill in the missing values of the CF rating matrix to form a pseudo rating matrix, and then predicts customer ratings by using the Pearson CF algorithm instead of weighted Pearson CF on the pseudo rating matrix. Su et al. also proposed a Joint Mixture CF (JMCF) which combines the predictions from three independent experts: Pearson correlation-based CF, a pure TAN-ELR content-based predictor, and a pure TAN-ELR. The results have been compared with Pearson correlation-based CF (a kind of memory-based CF), model-based CF algorithm, content-based predictor, combination of CB and CF (Su et al. 2007). Rodriguez hybridized a collaborative system and a knowledge-based system to solve the cold start problem (Rodríguez et al. 2010). It has been proven that the CF recommendation approach, or its combination with another technique, is the most successful and widely used approach for recommender systems (Bobadilla, Serradilla & Bernal 2010; Herlocker et al. 1999; Schafer et al. 2007). The literature particularly shows that the combination of a user-based CF and an item-based CF may achieve good performance in a big-user-set and big-item-set environment (Sarwar et al. 2001).

A hybrid recommendation technique may be proposed to gain higher performance and to avoid the drawbacks of typical recommendation techniques (Burke 2007). The most common practice in current hybrid recommendation techniques is to combine the CF approach with other recommendation techniques in an attempt to avoid cold-start, sparsity and/or scalability problems (Adomavicius & Tuzhilin 2005; Kim, Li, et al. 2006; Lu 2012; Lu et al. 2010, 2013; Takács et al. 2009).

One major challenge is that recommendations to customers in selecting the most suitable products/services are often made with incomplete and uncertain information. The similarity between items or between customers is naturally with fuzziness. Furthermore, to make a recommendation to a customer we must have some information about the customer's preferences which can essentially be obtained in two different ways: extensionally or intentionally expressed. The first is based upon the past experiences of the customers and the second is based on the actions of similar customers, that is, finding the 'general neighbourhood'. All experiences and actions may be described by linguistic terms such as 'good', or 'very good'. Fuzzy set theory lends itself well to handling fuzziness and uncertain issues in recommendation problems. Recent research efforts have indicated that fuzzy sets, fuzzy logic and fuzzy relations are potentially within the domain of recommender systems. For example, Chen and Duh (2008) developed a personalised intelligent tutoring system based on the proposed fuzzy item response theory, which is capable of recommending courseware with suitable difficulty levels for learners according to a learner's uncertain responses (Chen & Duh 2008). Cornelis and colleagues (Cornelis et al. 2005; Cornelis et al. 2007) developed a conceptual framework for recommending one-and-only items using fuzzy logic techniques to overcome the limitations of existing recommendation techniques in an uncertain information processing and matching environment. Leung, Chan and Chung (Leung, Chan & Chung 2006) proposed the use of fuzzy association rules to solve the sharp boundary problem of the CF approach. In more complex situations, such as telecommunications, telecom products are in a hierarchical structure and each customer can purchase a package with a set of products. Clearly, earlier techniques could not deal with complex situations in CRM and thus we need to develop hierarchical fuzzy measure-based recommendation techniques.

Community-based recommendation technique is another kind of hybrid recommendation method that has evolved to replace and broaden CF, such recommendation systems that provide best-seller lists based on the predictions of customer preferences that provide community opinions (Chen, Chou & Kauffman 2009; Kamahara et al. 2005). Through community-based recommender systems, customers with interests in similar products can interact, and their interactions become trusted information which reduces information asymmetries; they also induce more trust and foster confidence in their purchasing decisions. Research has found that digitized word-of-mouth is central to the power of trust and reputation and can be defined as the main driver of sales, although online review sites also have a secondary influence (Biancalana et al. 2013). Based on observations of real-world community based recommender systems, which can be classified them based on the recommended targets: systems supporting product review; systems for product providers, and systems for product sellers (Li & Kao 2009; Lu et al. 2010, 2013; Shambour & Lu 2012; Wang et al. 2013). However, how to identify customer communities (especially in banking and telecom industries) and how to use such a community's opinions to improve recommendation accuracy has not been well studied in the literature.

2.3.6 RECOMMENDER SYSTEM APPLICATION

Recommender systems are the most successful implementation of web personalisation and can be defined as personalised information filtering technology that is used to automatically predict and identify a set of interesting items on behalf of customers according to their personal preferences (Bobadilla, Serradilla & Bernal 2010; Piao, Zhao & Zheng 2009). Recommender systems use the concept of rating to measure customers' preferences and a range of filtering techniques, and can be classified in multiple ways according to the nature of the input information.

Web personalisation can be defined as the ability to provide tailored products and services, or information relating to products or services, to individuals based on their preferences and behaviours (Gao, Liu & Wu 2010). In recent years, significant steps have been taken towards providing personalised services for a wide variety of web-based applications in e-commerce, e-business, e-learning and e-government (Gao, Liu & Wu 2010; Guo et al. 2013; Lu 2012; Lu, Ruan & Zhang 2006; Lu et al. 2010, 2013; Piao, Zhao & Zheng 2009; Porcel, López-Herrera & Herrera-Viedma 2009; Shambour

& Lu 2011; Shi, Ye & Gong 2008; Wei, Huang & Fu 2007; XiaoYan, HongWu & SongJie 2008). Successful applications using recommendation techniques have involved various product and service areas such as recommending news, movies, books, videos, exhibitions, and business partners (Alonso et al. 2012; Untema et al. 2010).

There are three main types of web personalisation application: personalised search, adaptive website, and recommender systems (Guo & Lu 2007; Lu et al. 2010; Schubert, Uwe & Risch 2006). Personalised search seeks to tailor the search results according to each customer's personal needs. The literature suggests that it is a personalised mapping framework that automatically maps a set of known customer interests onto a group of categories in the open directory project, which categorises and personalises search results according to a web customer's interests. Adaptive website, also known as website customisation, offers customers the ability to build their own web interface by selecting from channels of information; in so doing, it modifies the content and structure of websites according to individual customers' preferences. The literature also reports a number of website customisation models that personalise the site's contents and structure according to a particular web customer's needs by learning from the customer's interests, which are identified and described through the customer's website navigation records. A recommender system, as a personalised information filtering technology, uses explicit and implicit information to either predict whether a particular customer will like a particular item, or to identify a set of items that will be of interest to a particular customer (Manouselis & Costopoulou 2007).

2.3.7 FUZZY SET TECHNIQUES IN RECOMMENDER SYSTEMS

In many studies, item ratings are specified on a scale of values; for example, on a scale of 1 to 5, where 1 indicates the lowest preference and 5 indicates the highest preference for an item by a specific customer. Some researchers have also introduced other preference models in specific application fields (Benferhat et al. 2006). In practical situations, customers like to express their preferences in linguistic terms, such as 'very interested', or 'not interested' for the features of a mobile product/service. Therefore, recommendations to online customers are often generated on the basis of uncertain or vague information (Herrera-Viedma & Porcel 2009). The similarities between items or between customers are naturally fuzzy, which attracts many researchers to apply fuzzy set theory, fuzzy logic and fuzzy relations to recommender systems in an attempt to achieve more accurate and effective recommendations. For example, Cao and Li (Cao & Li 2007) proposed a fuzzy-based recommender system for the consumer electronics area to retrieve optimal products. Porcel et al. (2009) developed a fuzzy linguistic-based recommender system based on both content-based filtering and fuzzy linguistic modelling techniques. However, there has been no report on the implementation of a recommender system for the complex situations in banking or telecom products/services recommendation. Telecom businesses, for example, offer hundreds of different mobile products and services such as handsets, mobile plans and broadband to customers and are constantly exploring new products that will support customers in their selection and purchase of products and services on the Internet. Telecom products are always linked with services, referred to hereafter as 'products/services', and have very complex structures and a huge number of choices.

Recommender systems are designed to resolve this problem by automatically making helpful recommendations about various products and services to customers (Ricci & Shapira 2011). Such systems can make recommendations according to customer profiles or preferences, or they can rely on the choices of other people who could be useful referees. The advantage of recommender systems is that they suggest the right items (products or services) to particular customers (suppliers, salespeople, etc.) based on their explicit and implicit preferences by applying information filtering technologies (Manouselis & Costopoulou 2007) (Lu et al. 2010; Untema et al. 2010).

CHAPTER 3

PRELIMINARIES

3.1 STATISTICAL METHODS

This section reviews a set of related statistical methods employed in this study, including logistic regression, survival analysis and factor analysis.

3.1.1 BOOSTING

Boosting refers to a general and provably effective method that attempts to 'boost' the accuracy of any given learning algorithm (Freund and Schapire 1999). Although boosting is not algorithmically constrained, most boosting algorithms involve iteratively learning and adding weak classifiers to come up with a final strong classifier. Each added weak classifier is usually weighted according to its accuracy, and trained with re-weighted training data.

One of the earliest and best-known boosting algorithms is AdaBoost (Freund and Schapire 1999). The AdaBoost algorithm takes a training set $S = \{(x_i, y_i)\}$ as inputs, where $i = \{1,2,\ldots,N\}$, $x_i = (x_1, x_2, \ldots, x_N) \in X$ and $y_i \in Y = \{-1,+1\}$, which works by repeatedly training a base classifier based on a weighted training set, and synthesising these trained classifiers. Initially, all training samples are weighted equally, but the weights of incorrectly classified samples are increased for the next round, so that the base classifier is forced to focus on examples with higher weights in the training set. The pseudo-code for AdaBoost is given in the following process. As this algorithm uses a discrete base classifier (hypothesis) $h: X \to \{-1,+1\}$, the algorithm is also called Discrete AdaBoost in later literature. The algorithm is defined as follows;

Algorithm 3-1:

(1). Initialize the weight of sample s_i, $D_1(i) = 1/N$.

(2). For $t = 1,\ldots,T$:

Train a base classifier $h_t : X \to \{-1,+1\}$, using weighted training set D_t

Compute the estimated error of h_t, $\varepsilon_t = \Pr_{i\sim D_t}\left[h_t(x_i) \neq y_i\right]$

Choose $\alpha_t = \frac{1}{2}\ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right)$

Set $D_{t+1}(i) = \frac{D_t(i)}{Z_t} \times \begin{cases} e^{-\alpha_t} & if \quad h_t(x_i) = y_i \\ e^{\alpha_t} & if \quad h_t(x_i) \neq y_i \end{cases}$

$= \frac{D_t(i)\exp(-\alpha_t y_i h_t(x_i))}{Z_t}$

where Z_t is a normalize factor, so that $\sum_i D_{t+1}(i) = 1$

(3). Output the classifier $H(x) = sign\left[\sum_{t=1}^{T}\alpha_t h_t(x)\right]$

Boosting has been studied in a more generalized framework (Schapire & Singer 1999). Schapire and Singer proposed using a base classifier $f : X \to$R to replace the discrete classifier $h(x)$, where $|f(x)|$ represents the confidence of its prediction. They also improved the choice of α_t by Equation 1. When $W^0 = 0$, the choice of α will be identical to Discrete AdaBoost, as the latter can be reformed as $\alpha_t = \frac{1}{2}\ln\left(\frac{W_t^+ + \frac{1}{2}W_t^0}{W_t^- + \frac{1}{2}W_t^0}\right)$ according to their generalized framework.

$$\alpha_t = \frac{1}{2}\ln\left(\frac{W_t^+}{W_t^-}\right) \tag{3.1}$$

where $W_t^b = \sum_{i:y_i f_t(x_i)=b} D_t(i)$, $b \in \{-,0,+\}$.

From a statistical point of view, (Friedman, Hastie & Tibshirani 2000) proposed the Gentle AdaBoost algorithm that fits additive logistic regression. Using Newton stepping to minimize the criterion $E\left[e^{-yF(x)}\right]$, where $p(x)=\dfrac{e^{2F(x)}}{1+e^{2F(x)}}$, the Gentle AdaBoost does not require the computation of the log ratios which may lead to very large updates. Thus, Gentle AdaBoost is believed to be more reliable and stable. The detail of Gentle AdaBoost is given in the following process.

Algorithm 3-2:

(1). Initialize the weight of sample s_i, $D_1(i)=1/N$, $F(x)=0$.

(2). For $t=1,\ldots,T$:

a) Fit the function $f_t(x)$, using weighted least-squares regression of y_i to x_i with weight $D_t(i)$

b) Set $F(x)=F(x)+f_t(x)$.

c) Set $D_{t+1}(i)=\dfrac{D_t(i)\exp(-y_i f_t(x_i))}{Z_t}$

where Z_t is a normalize factor, so that $\sum_i D_{t+1}(i)=1$

(3). Output $F(x)=\sum_{t=1}^{T} f_t(x)$.

Boosting algorithms are intentionally omitted for classification with multiple classes. In this study, there are only two classes in the case studies, churner or non-churner. For a comprehensive review of boosting algorithms, refer to (Chatrchyan et al. 2012; Friedman, Hastie & Tibshirani 2000).

3.1.2 LOGISTIC REGRESSION

Logistic Regression (Anderson 1982; Collins, Schapire & Singer 2002) is a simple, but effective analytic method that is used to describe and test hypotheses about relationships between a categorical variable and one or more categorical, or continuous variables. Given a sample set $S=\{(x_i,y_i)\}$, where $i=\{1,2,\ldots,N\}$, and

$x_i = (x_1, x_2, \ldots, x_n) \in X$, to evaluate the relationship between a set of independent variables (inputs) $x_i \in R^n$ and a corresponding target label $y \in Y = \{-1, +1\}$, the logistic regression estimates the probability of $P(y = 1 \mid x_i) = \hat{p}$ by

$$\hat{p} = \frac{1}{1 + \exp(-\sum_{j=0}^{n} \beta_i x_i)} \tag{3.2}$$

where $\sum_{i=0}^{n} \beta_i x_i = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n = \ln\left(\frac{\hat{p}}{1 - \hat{p}}\right) = \operatorname{logit}(\hat{p})$ is called the regression equation, with intercept β_0, and regression coefficients $\beta_i, i = \{1, 2, \ldots, n\}$. (Hosmer & Lemeshow 2004)

In its application, the maximum likelihood estimation is used to maximize the likelihood of the regression coefficients given a set of observations (samples).

3.1.3 *Survival Analysis*

Survival Analysis is a statistical method that incorporates time-varying covariates and accounts for the sequential nature of the data (Cox 1972). An example is in the modelling of an occurrence and the timing of events, as in customer churn. Assume the churn time T for a particular mobile customer is a random variable with a cumulative distribution function, $F(t)$, which is also known as the failure function, and probability density function, $f(t)$., the survivor function $S(t)$ is defined as

$$\Pr(T > t) = 1 - F(t) \equiv S(t) \tag{3.3}$$

where t is the elapsed time since censoring of each customer at time 0

The hazard rate $h(t)$, is defined as:

$$h(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)} \tag{3.4}$$

As $f(t)$ is the slope of $F(t)$:

$$f(t) = \frac{\partial F(t)}{\partial t} = \lim_{\Delta t \to 0} \frac{\Pr(t \le T < t + \Delta t)}{\Delta t} \tag{3.5}$$

Thus, for tiny Δt, $h(t)\Delta t$ is akin to the probability of a customer churning at exactly the time of t, conditional on survival up to time t.

3.1.4 FACTOR ANALYSIS

Factor Analysis is a statistical method used to analyse the inter-relationships between numbers of observed, correlated variables of interest and explain those variables in terms of a potentially lower number of unobserved, uncorrelated factors (Brown 2010; Fabrigar et al. 1999). In factor analysis, we assume that a set of p random variables $X = [x_1, x_2, \ldots, x_p]^T$, with means $\mu = [\mu_1, \mu_2, \ldots, \mu_p]^T$, can be expressed as a linear combination of a set of independent, uncorrelated common factors $F = [F_1, F_2, \ldots F_k]^T$,

$$X - \mu = LF + \varepsilon \tag{3.6}$$

where $L_{p \times k}$ is the loading matrix, and $\varepsilon = [\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p]^T$ is a set of independently specific factors.

If there are n observations, then it will have the dimensions $X_{p \times n}$ and $F_{k \times n}$. Each column of X and F denote values for one particular observation, and the loading matrix L does not vary across observations.

In Factor Analysis, the following assumptions are usually imposed: 1) The specific factors ε_i are independently distributed with zero mean and finite variance, so that $E(\varepsilon_i) = 0,\ Var(\varepsilon_i) = \sigma_i^2$; 2) The unobservable factors F_i are independent of one another and of the specific factors, so that $E(F_i) = 0,\ Cov(F) = I$.

3.2 FUZZY TECHNIQUES PRELIMINARIES

3.2.1 *FUZZY SETS*

Fuzzy sets were introduced by Zadeh (1965) as an extension of the classical notion of set to manipulate ambiguous, uncertain and imprecise values in real life. A conventional set is dichotomous, while a fuzzy set is characterized by a membership function, which assigns a grade of membership ranging from 0 to 1 to each object. A mathematical definition of a fuzzy set is given as follows:

Definition 3.1 (Fuzzy set): Let X be a universal set, then a fuzzy set $\widetilde{A}$ of X is defined by its membership function.

$$\mu_{\widetilde{A}} : X \to [0,1],\ x \mapsto \mu_{\widetilde{A}}(x) \in [0,1] \tag{3.7}$$

The value of $\mu_{\widetilde{A}}(x)$ represents the grade of membership of x in X and is interpreted as the degree to which x belongs to $\widetilde{A}$; therefore, the closer the value of $\mu_{\widetilde{A}}(x)$ is to 1, the more it belongs to $\widetilde{A}$.

A crisp, or ordinary set A of X can also be viewed as a fuzzy set in X with a membership function as its characteristic function, i.e.,

$$\mu_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases} \tag{3.8}$$

A fuzzy set $\widetilde{A}$ can be characterized as a set of ordered pairs of elements x and grade $\mu_{\widetilde{A}}(x)$ and is noted by

$$\widetilde{A} = \{(x, \mu_{\widetilde{A}}(x)) \mid x \in X\} \tag{3.9}$$

where each pair $(x, \mu_{\widetilde{A}}(x))$ is called a singleton.

When X is a countable or finite set, a fuzzy set $\widetilde{A}$ on X is expressed as

$$\widetilde{A} = \sum_{x_i \in X} \mu(x_i)/x_i \tag{3.10}$$

When X is a finite set whose elements are $x_1, x_2, \ldots, x_n$, a fuzzy set $\widetilde{A}$ on X is expressed as

$$\widetilde{A} = \{(x_1, \mu_{\widetilde{A}}(x_1)), (x_2, \mu_{\widetilde{A}}(x_2)), \ldots, (x_n, \mu_{\widetilde{A}}(x_n))\} \tag{3.11}$$

When X is an infinite and uncountable set, a fuzzy set $\widetilde{A}$ on X is expressed as

$$\widetilde{A} = \int_X \mu(x)/x \tag{3.12}$$

These expressions mean that the grade of x is $\mu_{\widetilde{A}}(x)$ and the operations 'Σ,' and '$\int$' do not refer to ordinary addition and integral, but are a union, and '/' does not indicate an ordinary division, but is merely a marker.

To describe the proposed approach used in Chapters 7 and 8, based on Zadeh (1965), basic notions of fuzzy sets, fuzzy numbers, positive and negative fuzzy numbers, and linguistic variables are described as follows, and a related theorem is given (Zhang and Lu 2003). These notions are used in a linguistic term similarity calculation in the proposed recommendation approach in Chapter 7 and 8.

Definition 3.2 A fuzzy set $\tilde{A}$ in a universe of discourse X is characterized by a membership function $\mu_{\tilde{A}}(x)$ which associates with each element x in X a real number in the interval [0, 1]. The function value $\mu_{\tilde{A}}(x)$ is termed the grade of membership of x in $\tilde{A}$. A fuzzy number $\tilde{a}$ is a fuzzy set, which is defined in a set of all real numbers R.

Definition 3.3 The λ-cut of fuzzy number $\tilde{a}$ is defined

$$\tilde{a}_\lambda = \{x \colon \mu_{\hat{a}}(x) \geq \lambda, x \in R\} \tag{3.13}$$

where $\tilde{a}_\lambda$ is a nonempty bounded closed interval contained in *X* and can be denoted by $\tilde{a}_\lambda = [a_\lambda^-, a_\lambda^+]$, a_λ^- and a_λ^+ are the lower and upper bounds of the closed interval, respectively.

Definition 3.4 A triangular fuzzy number $\tilde{a}$ can be defined by a triplet (a_0^-, a, a_0^+) and the membership function $\mu_{\tilde{a}}(x)$ is defined as:

$$\mu_{\tilde{a}}(x) = \begin{cases} 0, x < a_0^- \\ \dfrac{x - a_0^-}{a - a_0^-}, a_0^- \leq x \leq a \\ \dfrac{a_0^+ - x}{a_0^+ - a}, a \leq x \leq a_0^+ \\ 0, a_0^+ < x \end{cases} \tag{3.14}$$

From Definition 3, we can deduce that $a = a_1^- = a_1^+$.

Definition 3.5 If $\tilde{a}$ is a fuzzy number and $a_\lambda^- > 0$ for any $\lambda \in [0,1]$, then $\tilde{a}$ is called a positive fuzzy number. Let $F_+^*(R)$ be the set of all finite positive fuzzy numbers on *R*.

Definition 3.6 For any $\tilde{a}, \tilde{b} \in F_+^*(R)$ and $0 < \alpha \in R$,

$$\tilde{a} + \tilde{b} = \cup_{\lambda \in [0,1]} \lambda\left[a_\lambda^- + b_\lambda^-, a_\lambda^+ + b_\lambda^+\right] \tag{3.15}$$

$$a\tilde{a} = \cup_{\lambda \in [0,1]} \lambda[aa_\lambda^-, aa_\lambda^+] \tag{3.16}$$

$$\tilde{a} \times \tilde{b} = \cup_{\lambda \in [0,1]} \lambda\left[a_\lambda^- \times b_\lambda^-, a_\lambda^+ \times b_\lambda^+\right] \tag{3.17}$$

Definition 3.7 Let $\tilde{a}$ and $\tilde{b}$ be two fuzzy numbers, then $\tilde{a} = \tilde{b}$ if $a_\lambda^- = b_\lambda^-$ and $a_\lambda^+ = b_\lambda^+$ for any $\lambda \in [0,1]$.

Definition 3.8 A linguistic variable is a variable whose values are words or sentences in a natural or artificial language. A linguistic variable is characterized by a quintuple $(\chi, T(\chi), U, G, M)$ in which χ is the name of the variable; $T(\chi)$ is the term-set of χ, that is, the collection of its linguistic values; U is a universe of discourse; G is a syntactic rule which generates the terms in $T(\chi)$; and M is a semantic rule which associates with each linguistic value X its meaning, $M(X)$, where $M(X)$ denotes a fuzzy subset of U Zadeh (1975).

Definition 3.9 Let $\tilde{a}, \tilde{b} \in F_+^*(R)$ then the vertex method is defined to calculate the distance between them as

$$d(\hat{a}, \hat{b}) = \sqrt{\frac{1}{3}[(a_0^- - b_0^-)^2 + (a - b)^2 + (a_0^+ - b_0^+)^2]} \tag{3.18}$$

Definition 3.10 Let $\tilde{a}, \tilde{b} \in F^*(R)$ then fuzzy number $\tilde{a}$ is closer to fuzzy number $\tilde{b}$ as $d(\tilde{a}, \tilde{b})$ approaches 0.

In general, fuzzy numbers are applied to deal with sets of linguistic terms; for instance, a set of five linguistic terms {Strongly Interested (SI), More Interested (MI), Interested (I), Less Interested (LI), Not Interested (NI)} is used to describe the customer ratings. Essentially, any form of fuzzy numbers, called general fuzzy numbers, can be used to describe these linguistic terms. The fuzzy numbers related to these linguistic terms are shown in Table 3-1. Their membership functions are illustrated in Figure 3-1.

Table 3-1 Linguistic Terms and Related Fuzzy Numbers

Linguistic Terms	Triangular Fuzzy Numbers
Strongly Interested (SI)	(4,5,5)
More Interested (MI)	(3,4,5)
Interested (IN)	(2,3,4)
Less Interested (LI)	(1,2,3)
Not Interested (NI)	(1,1,2)
N/A	-

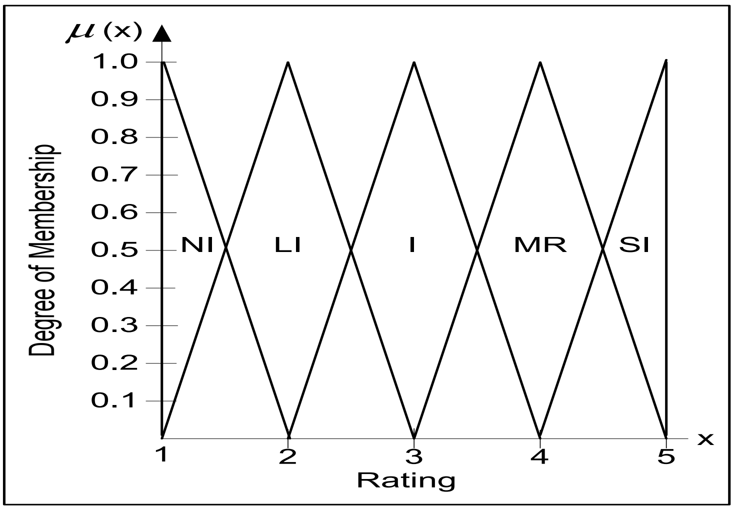


Figure 3-1 Fuzzy Sets and Membership Functions for Table 3-1

CHAPTER 4

A PERSONALISED ANALYTICAL ELECTRONIC CUSTOMER RELATIONSHIP MANAGEMENT FRAMEWORK

4.1 OVERVIEW

In Chapter 2, a review of the literature has proven that analytical eCRM is particularly useful in measuring customer profitability and value, which can be developed on an historic, potential or lifetime basis. This is an important part of customer marketing analytics, because the purpose of marketing investments is to increase customer value over time, and one of the ways to analyse the effectiveness of marketing is by understanding what impact it has had on customer value. Further, organisations can use it to drive their customer retention and acquisition strategies.

In this chapter, a personalised analytical eCRM framework is proposed to enable the application of data mining, statistical modelling and recommendation techniques to eCRM operations, to allow predictive analytics to play a critical role in pinpointing the drivers of short-term and long-term customer value, and to establish how best to increase this customer value in such a way that the offer best suited to the customer is the one recommended. A customer segmentation approach is also developed that combines classification techniques and fuzzy set techniques to populate the values of missing variables to an internal database. As segmentation is the very first step in understanding the drivers of customer value, all kind of marketing programs might be deployable against different types of customer segments. Segmentation also can result in a major overhaul of product pricing, sales incentive structures and call centre scripts which, in turn, lead to significant savings and increases in profitability.

The remainder of this chapter is organised as follows. Section 4.2 describes the proposed analytical eCRM framework. Section 4.3 outlines the main steps of the Hybrid Migrating Customer Segmentation (HMCS) method, and Section 4.4 illustrates an application of the HMCS method in a telecom customer segmentation under the proposed analytical eCRM framework. Section 4.5 summarises the work presented in this chapter.

4.2 A Personalised Analytical eCRM

In this section, a personalised analytical Electronic Customer Relationship Management (eCRM) framework is proposed for the purpose of applying predictive modelling and recommendation techniques to operational eCRM. This proposed framework can make offers or recommendations according to customer profiles, propensity and preferences. Apart from likelihood of offer acceptance (the propensity), it can also draw on the choices of other people as useful referees. The advantage of this framework is that it will suggest the right products or services to particular customers based on their explicit and implicit preferences, and at same time, the offer communication will be based on the likelihood of customer acceptance.

The four main components to take into account in the design of the personalised analytical eCRM framework are:

- Define which customers, called ‘prospects’, are in the target market, differentiating the types of customer based on behavioural and demographic characteristics through data mining classification and clustering analysis. This process is part of customer acquisition.
- Identify the responders from marketing campaigns, sales offers and ratings to determine which customers are interested in a product/service by using predictive modelling and profiling. This activity is defined as part of customer marketing.
- Understand the buying patterns of existing customers via sequential/association analysis, propensity modelling and customer rating, which will add great value to customer development.
- Prevent customer churn by defining triggers, through survival analysis and propensity modelling. This aims to achieve customer retention goals.

These four dimensions can be defined as a closed loop of a customer lifecycle management system. They work together to create a deep understanding of customers which will maximise customer value in the long term.

The proposed framework of the personalised analytical eCRM is shown in Figure 4-1. It has four main areas: existing customer data warehouse, integrated analytical eCRM system, personalised recommendation engine and interfaces.

(1) The customer data warehouse contains all the information about products and services, customer profiles, behaviours and transaction information.

(2) The integrated analytical eCRM system is the combination of marketing, sales, and customer support, which, in conjunction with data mining techniques, supports the functions of marketing campaign automation, sales-force automation, customer service and data capture. Figure 4-1 shows that this area consists of four functionalities;

- Marketing function; this function identifies targeted customer e-marketing driven by analytics, including customer segmentation and propensity modelling, in preparation for the targeted marketing campaigns.

- Support function; this function mainly executes all targeted customer marketing campaigns by using e-marketing channels, and captures all information from all customer interaction points.
- Sales function; this function fulfils sales orders, and by understanding the responders from marketing campaigns, suggests sales offers and ratings that will enhance the selling system, using predictive modelling techniques.
- Content management function; this function is the set of processes and technologies for collecting, managing, and publishing information in any form or medium; for example, digital content may take the form of text (such as electronic documents), or multimedia files (such as audio or video files).

These four functions work together as an integrated analytical eCRM system.

(3) The recommender system generates recommendations by combining predictive modelling and recommendation techniques. There are two main functions in the recommendation engine, as shown in Figure 4-1.

- Predictive analytics solutions: this function combines all the inputs from data mining classification and clustering analysis, predictive modelling and profiling, also survival analysis or risk analysis, to generate the analytics solutions to fit into the web recommendation engine.
- Web recommendation engine: this function applies recommendation techniques based on customer preference profiles, in conjunction with predictive analytics solutions, to generate the most relevant personalised recommendation to customers.

(4) Interfaces are the web connections that access customers' requests, product holdings, existing customer information, and the interface that captures the feedback information from all customer touch points. In Figure 4-1, two main interfaces are shown. The existing user (customer) web interface allows existing customers to access their personal information to self-manage their account and also to search relevant information for their purchase. During this search process, the recommendation engine will generate personalised recommendations to customers. The other interface shown is

the new user (customer) web interface, which enables a new customer (or prospective customer) to access the organisation's product/service information to search for purchases. The search process purposely guides customers to use their own information match to existing customers' profiles and preferences, and the recommendation engine will generate personalised recommendations to these new customers.

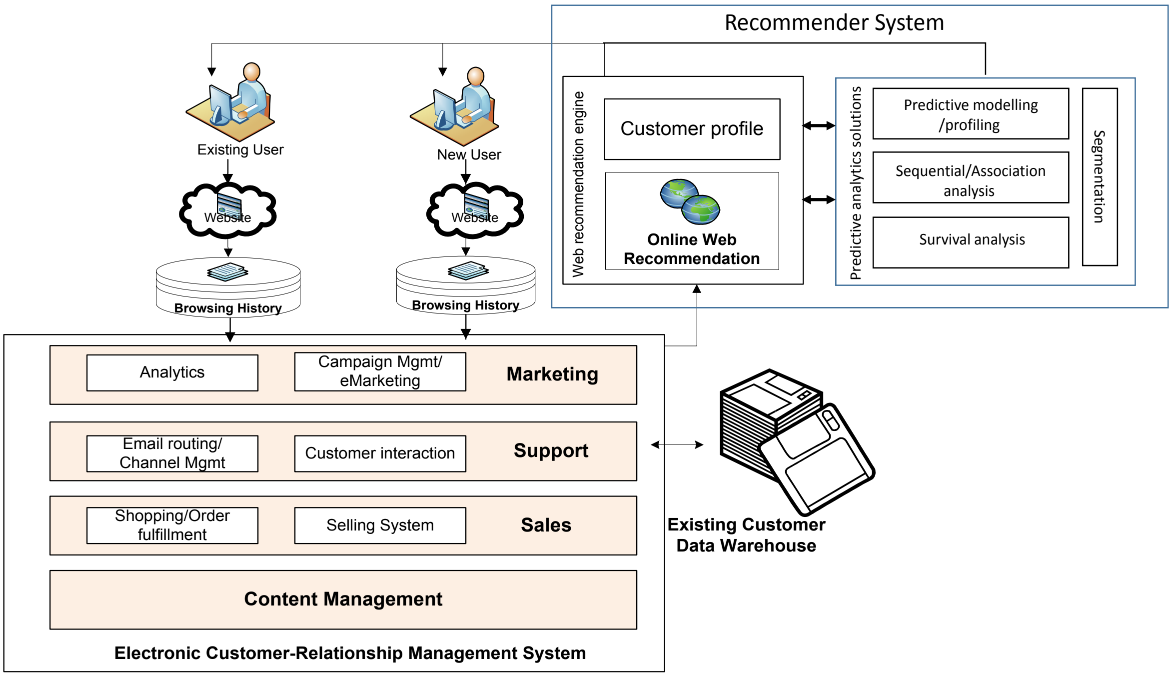


Figure 4-1 Proposed Personalised Analytical eCRM Framework

As mentioned in points (2) and (3) above, customer segmentation aims to understand customer characteristics by allocating customers to different marketing target groups, which is the starting point for understanding customer profiles. Conventional segmentation practices commence with the organisation's internal customer database, which is a very efficient way to understand the profiles of existing customers. However, it is also necessary to understand the prospects' profiles: how different or similar are they to the profiles of existing customers? This kind of segmentation can be achieved by combining internal and external data sources. Based on this sophisticated understanding, marketing teams can start to design targeted campaigns, then based on responders from campaigns, sales offers and customer ratings, statistical predictive modelling can be used to predict the likelihood of customers being interested in or taking up a product/service. Finally, predictive modelling can also be effectively applied in customer churn or attrition risk management, which plays very important role in customer retention.

The following section proposes a hybrid customer segmentation approach to tackle the challenge of customer segmentation by combining internal and external data sources, and in particular, migrating customer segmentation results from external data to the internal database.

4.3 A HYBRID CUSTOMER SEGMENTATION MODEL

Customer segmentation is very significant and effective way to understand customer information. Customer segmentation can be conducted on an organisation's internal and external data sources. Internal data sources are very valuable for customer profiles and behaviour patterns; however, there are still some important variables not available for effective customer segmentation. For instance, a telecom organisation can hold internal information about their mobile customers' contracts and billing history, as well as details of the handset models in its enterprise database, but it cannot practically store information about a customer's professional background, marital status and business relationship which contribute more information for better customer segmentation. Hence, an organisation often needs to collect customer-related information from external data sources for segmentation and migration to its internal data sources. Customer-related variables in external data sources seldom matching those in internal data, however, and it is challenging to apply segmentation results obtained on external data sources to an internal database. Therefore, it is necessary to find a practical solution to the problem of migrating segmentation results on external data sources to an internal database.

Internal customer data is often objective information such as a customer's contract terms, billing history, and spend amount; a customer's subjective expectations or preference changes would not normally be found in the database. To obtain customers;

subjective expectations and preferences, customer surveys are widely used, and customer segmentation is often conducted based on the survey data. If participants in a customer survey could be identified in the internal customer database, it would be easier to migrate customer segmentation results from survey data into the internal customer database. However, this is prohibited by privacy law or privacy policy in most cases; hence, a direct link between external and internal data sources could be problematic. Furthermore, even if such a link exists, the inconsistent variables between two data sources needs to be solved for migration to take place. In fact, the variables of customer segmentation based on survey data are not necessary every variable can be found in the internal database. Because of these challenges, the migration of customer segmentation often fails in real applications.

In the following sections, a five-step Hybrid Migrating Customer Segmentation (HMCS) method is explored, with the aim of migrating the customer segmentation result from the external data to the internal database. The model combines classification techniques and fuzzy set techniques to populate the values of missing variables to the internal database, and then implements the customer segmentation result defined by the external data source into the internal database. The developed model is applied to a real world customer segmentation problem.

4.3.1 PROBLEM DESCRIPTION AND FORMALISATION

In this section, the Migrating Customer Segmentation problem is described and formalised based on a case study.

A telecom organisation aims to develop several new products to retain existing customers and attract potential customers, especially business customers. An initial analysis has found that the industrial sector background, number of employees, and owner's preference will impact the customer's choice of a certain product or service. Hence, it is necessary to segment the customer base into several groups and develop corresponding products and services for each group. For various historical, legal or technical reasons, there is a lack of important demographic, behavioural and preference data in the internal database. A number of randomly selected customers have therefore been surveyed and customer segmentation based on the collected survey data has been conducted. The developed and defined customer segments need to be assigned to the customer base, but the survey data cannot be fully matched to the internal customer behaviour database because there is no link to certain customer-related indicators which are used in the segmentation model. It is difficult to assign the segments to the customer database directly and it is necessary to find a way to apply the results effectively. This is called a Migrating Customer Segmentation (MCS) problem. The MCS problem does not only exist in telecoms but also in many other industrial sectors, such as finance and insurance. A definition of the MCS problem is given below.

Definition 4.1: Suppose C is an enterprise's internal dataset, which contains customer-related records. Each customer-related record is depicted through m variables $a_1, \dots, a_m$. Let S be another dataset, the customer survey data in the above example, obtained externally from the enterprise. Within S, K customer segments are defined through n variables $x_1, \dots, x_n$ and labelled as $G_1, \cdots, G_K$, i.e., $G_k = g_k(x_1, \cdots, x_n)$ for any $k = 1, \dots, K$. The MCS problem needs to address the question of how to migrate $G_1, \cdots, G_k$ from S to C under the constraint that $\{a_1, \cdots, a_m\} \cap \{x_1, \cdots, x_n\} \neq \emptyset$ and $\{a_1, \cdots, a_m\} \neq \{x_1, \cdots, x_n\}$.

An MCS problem can be formalised in a more generalised form;

Definition 4.2: Let S and T be the source and target datasets, respectively. Elements of S and T are represented by variable sets X and A, respectively; and $X \neq A$ and $X \cap A \neq \emptyset$. Let G be a label set with K labels $G_1, \cdots, G_K$, which represents knowledge learnt from S. A mapping g is defined on S such that for any $s \in S$, $g(s) \in G$. An MCS problem is how to define a mapping f on T such that for any $t \in T$, $f(t) \in G$.

4.3.2 A FIVE-STEP HYBRID MIGRATING CUSTOMER SEGMENTATION METHOD

In this section, a Hybrid Migrating Customer Segmentation (HMCS) method is presented to solve the MCS problem. This model contains five steps as described below. Figure 4-2 gives its main steps.

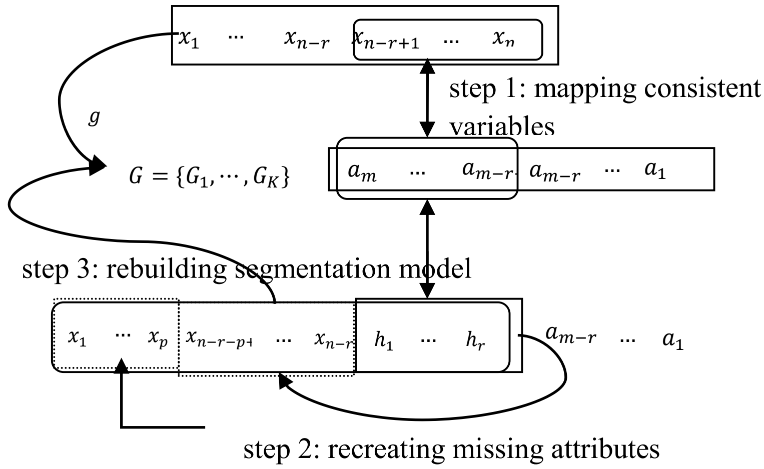


Figure 4-2 Main steps of the HMCS Method

Table 4-1 Outline of the HMCS Method

Outline of the HMCS method
Step 1: Mapping consistent variables between source and target datasets
Step 2: Recreating missing variables and populating their values
Step 3: Rebuilding segmentation model on source dataset
Step 4: Applying segmentation model to target dataset
Step 5: Evaluating segmentation model

Step 1: Mapping consistent variables between source and target datasets.

Let variable x in the source dataset be consistent with variable a in the target dataset if both x and a refer to the same feature of a customer and may have different value forms. For instance, 'month spending' is a variable used in most telecom customer surveys and is often given in the form of a number of spending ranges (intervals of spending amounts). In a telecom company's customer database, a customer's ‘monthly billing amount’ records the real spend of the customer and is often recorded as a real number. Although they are expressed in different forms in the source and target database, these two variables describe the same thing, i.e., a customer's spend on a telecom service in approximately a month. Hence, ‘month spending’ (from a source dataset) is consistent with ‘monthly billing amount’ (in a target dataset). In the following sections, two consistent variables are matched to each other.

Below, the same symbol has been used to replace the consistent variables between X and A ; and rewrite X and A as: $X = \{x_1, \dots, x_{n-r}, h_1, \dots, h_r\}$, $A = \{h_1, \dots, h_r, a_{m-r}, \dots, a_m\}$. Let $H = \{h_1, \cdots, h_r\}$, where h_r is a matching variable and H is called the matching variable set.

Each matching variable indicates a common customer feature in both the source and target datasets. For each h_r, a mapping m_r is build, such that:

(1) If h_r has categorical values in both source and target datasets.

$$m_r\big(V_S(h_r)\big) \subseteq V_T(h_r) \tag{4.1}$$

where $V_S(h_r)$, $V_T(h_r)$ are the values of h_r occurs in S and T, respectively.

(2) If h_r has categorical values in source dataset but continuous values in target dataset.

$$m_r\big(V_{T(h_r)}\big) \subseteq V_S(h_r) \tag{4.2}$$

By this step, the matching variables are aligned.

Step 2: Recreating missing variables on the source dataset and populating the values of missing variables to the target dataset.

A missing variable in the target dataset is a variable which only exists in the source dataset and does not have a consistent (matching) variable in the target dataset. A typical example is a customer's 'gender'. Gender is a common variable used in many customer-oriented survey datasets, but it is seldom a variable stored in an enterprise's database.

Because a missing variable does not exist in the target dataset but is used in the segmentation mapping$g$, this step tries to build a mock variable for the target dataset. Consider the matching variable set H which is shared between the source and target datasets, H is used to generate the missing variable. Without loss of generality, suppose the number of $n - r - p$ missing variables $x_{n-r-p+1}, \dots, x_{n-r}$ can be generated from H. For each $x_j, j = n - r - p + 1, \cdots, n - r$, a subset S_{x_j} of the source dataset with variables $H \cup x_j$ is obtained where H can be seen as condition variables and x_j can be seen as decision variable (class/category variable). Therefore, a classification algorithm L_j, such as decision tree or support vector machine (Yang et al. 2011) can be implemented to learn x_j from H, which can be then used on the target dataset to populate the values of missing variables $x_{n-r-p+1}, \dots, x_{n-r}$.

Since the fact that not all missing variables can be generated by H, another method is also used to generate missing variables. Missing variables $x_1, \dots, x_p$, which are not generated from H, will be populated into the target dataset based on the nature of their values. If a variable x focuses on a customer's objective feature, such as geographical location, then its value is populated following the probability distribution of those values. If a variable y focuses on a customer's subjective features, such as 'how likely a customer will select a competitor's service', fuzzy set and fuzzy logic technique will be used (Zhang & Lu 2003) to summarize its values, define a fuzzy set, and populate the fuzzy memberships of those values. To explain this method, an example is given below.

Example. Suppose a missing variable x in a telecom's customer survey is 'previous service provider' with values 'company A', 'company B', 'company C', and 'company D'. All four values will be populated into the target dataset following their frequency distribution (probability). Suppose another missing variable y in the same survey is 'how likely you will select another service provider?' and with five values 'Definitely', 'Very likely', 'More likely', 'Unlikely' and 'Definitely not'. Then a fuzzy set F will be defined on these five values as 'degree of likely to leave'; the fuzzy membership degree of each value is calculated, then the fuzzy membership degrees can be populated into the target dataset.

Considering that a fuzzy set is not uniquely determined, a small disturbance can be added belong to a fuzzy membership degree when populating it to the target dataset.

In Step 2, the variables missed in the target dataset have been artificially generated. After this step, the target dataset T contains all the variables in the source dataset S except that previously missed variables take artificial values. Before using the original and generated variables to implement customer segmentation, a model rebuilding (retraining) procedure on the source dataset is needed; this is the main task in Step 3.

Step 3: Rebuilding segmentation model on source dataset.

In this step, a model retraining is implemented by using artificial data for some variables and a classification algorithm, which is conducted on the source dataset.

Notice that in Figure 4-2, variables $x_{n-r-p+1}, \cdots, x_{n-r}$ can be learned by variables $h_1, \cdots, h_r$, and replaced by

$$x_j = L_j(h_1, \cdots, h_r),\ j = n - r - p + 1, \dots, n - r. \qquad (4.3)$$

Variables $x_1, \dots, x_p$, are reassigned artificially and their generated values are based on either probability distribution or the fuzzy membership degree following the method given in Step 2. For these variables, $y_1, \cdots, y_p$ are used to replace them. As the customer segmentation has been conducted and the segmentation result is known, variable z is used to record the segmentation result.

Based on the above preparation, a classification model g^* is built where variables $h_1, \cdots, h_r, y_1, \cdots, y_p$ are condition variables and the variable z is the decision variable, i.e.,

$$g^*(h_1, \cdots, h_r, y_1, \cdots, y_p) = z. \tag{4.4}$$

Note that the model is completely built on variables which now exist in the target dataset; model g^* is therefore applied to the target dataset.

Step 4: Applying the segmentation model on the target dataset.

In this step, the model g^* is applied to the target dataset after populating values to the artificially generated variables. Because the target dataset does not contain the artificially generated variables and cannot provide any information about them, a value pool for each of those variables is first generated based on the variable's probability distribution in the source dataset. A value is then randomly picked from the generated value pool for each record in the target dataset to build an applicable record as the input of model g^*. Formally, the data population procedure is shown in Figure 4-3.

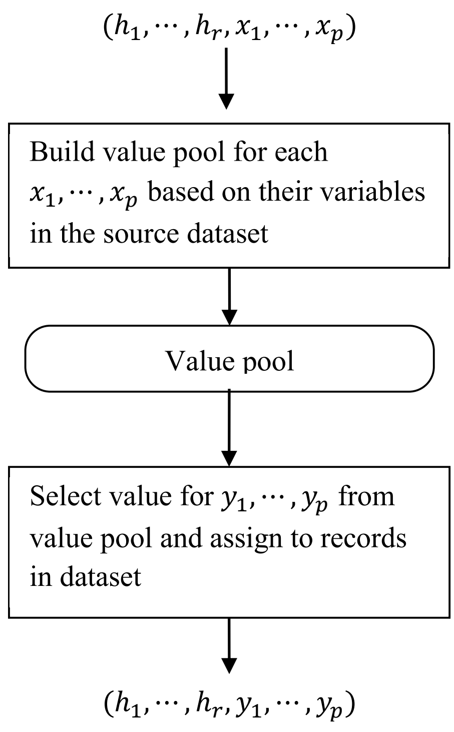


Figure 4-3 Data Population from Source Dataset to Target Dataset

Step 5: Evaluating segmentation results.

The evaluation is conducted on the source dataset and is also conducted manually on a sample from the target dataset. On the source dataset, cross validation is adopted. On the target dataset, a sample set is randomly selected and evaluated by the domain experts, focusing mainly on the approximate distribution of the segmentation result rather than the individual record.

4.4 EXPERIMENT AND ANALYSIS

The HMCS model presented in Section 4.3.2 has been implemented using MySQL database and KNIME (the Konstanz Information Miner, www.knime.org) tool on a Dell Latitude D6500 laptop with 3GB RAM running Fedora 16 Linux system. In this section, the experiment result is briefly introduced and analysed.

4.4.1 EXPERIMENT DATA

The experiment data employed a company's customer survey. The customer details have been delimited and this sample is only used for research purposes. The survey contains 42 customer-related questions and covers a total of 2000 customers. Of the 2000 customers, 1542 customers currently contact with the telecom company and 1519 customers have answered all relevant survey questions. Hence, all 1519 have been selected as valid records and form the source dataset. The target datasets are three samples from the customer database with 102555 (target-1), 109743 (target-2), and 103013 (target-3) customer records, respectively; also the customer details have been delimited.

4.4.2 INITIAL CUSTOMER SEGMENTATION RESULT

A customer segmentation has been developed based on the 1519 record of survey data. The total of 1519 customers has been segmented into five groups which are labelled 'segment-1', 'segment-2', 'segment-3', 'segment-4', and 'segment-5', respectively. Table 4-2 shows the record numbers of all five groups.

Table 4-2 Record Distribution over Five Segments in Source Dataset

Group Label	segment-1	segment-2	segment-3	segment-4	segment-5
Record Number	432	533	247	158	149
Percentage (%)	28.4	35.1	16.3	10.4	9.8

The segmentation result is built upon five variables (denoted by x_1, x_2, x_3, x_4, x_5) extracted from five questions from a total of 42 surveys. Of the five variables, three (x_3, x_4, x_5) have objective measurements and the other two (x_1, x_2) are subjective opinions. Of the three objective variables, two (x_4, x_5) have counterparts in the target dataset. Furthermore, statistical analysis of correlation indicates that variables x_1, x_2, x_3 cannot be estimated or learned from variables (x_4, x_5); therefore, a value pool for each of them needs to be generated to populate their values into the target dataset as well as the source dataset, as shown in Step 2 and Step 3.

4.4.3 RESULTS AND ANALYSIS

To evaluate the presented model, three experiments are conducted. The first experiment, Experiment 1, compares the model's segmentation result with the original segmentation result on the survey dataset through each segment's distribution. The second experiment (Experiment 2) compares the model's segmentation result on the same target dataset (target-1) with different value pools in populating values of missing variables. The third experiment, Experiment 3, compares the model's segmentation results on three sample target datasets.

The result of Experiment 1 is shown in Figure 4-4, which x-axis represents percentage of population proportion and y-axis represents segments The result indicates that the first three segments in both the original and the model segmentation occupy the majority of the 1519 records and have similar distributions, particularly the first two segments. Although as a whole, segment-4 and segment-5 in both segmentations occupy almost the same percentages of the total dataset, their distribution in the two segmentations are significantly different. By checking the segmentation result individually, it is noted that about 50-60% records are segmented to the same segment by both models.

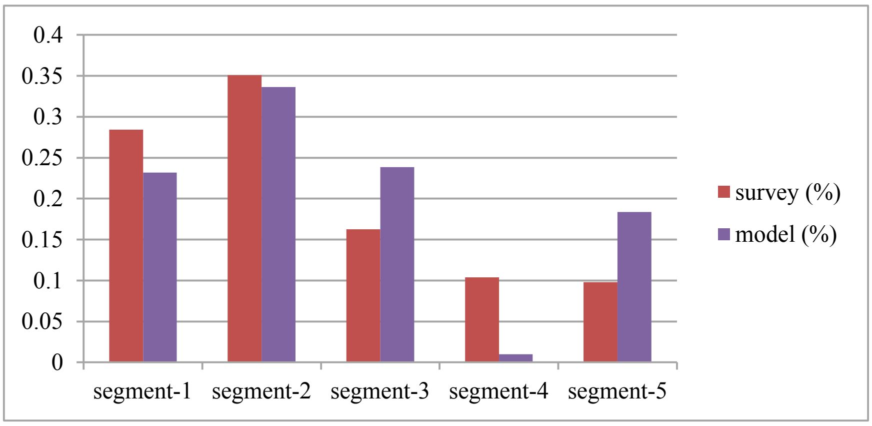


Figure 4-4 Result of Experiment 1

The result of 2 is shown in Figure 4-5. Same as in Figure 4-4, x-axis represents percentage of population proportion and y-axis represents segments The result indicates that the model has produced similar segmentation results by using different value pools in populating missing values to the target dataset. However, it still shows a difference, particularly for segment-4.

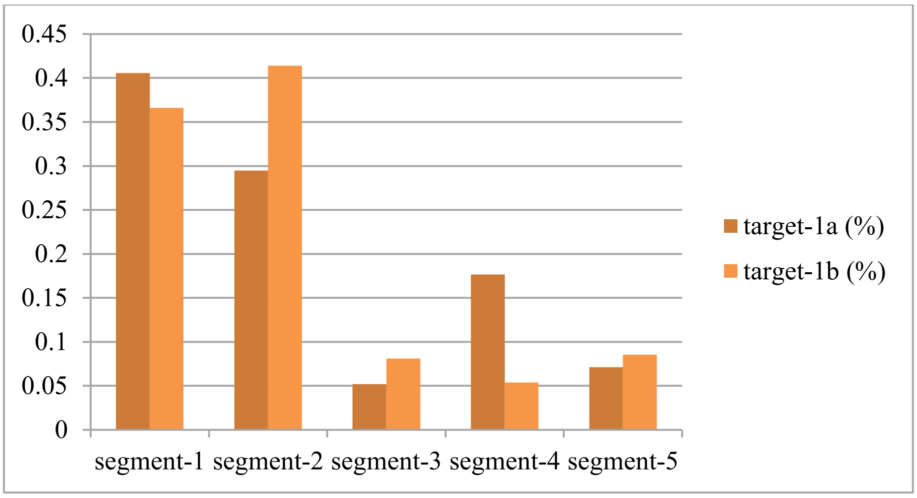


Figure 4-5 Result of Experiment 2

The result of Experiment 3 is shown in Figure 4-6, x-axis represents percentage of population proportion and y-axis represents segments. It indicates that the model has produced similar segmentation results on different sample sets, although significant differences still exist among these results.

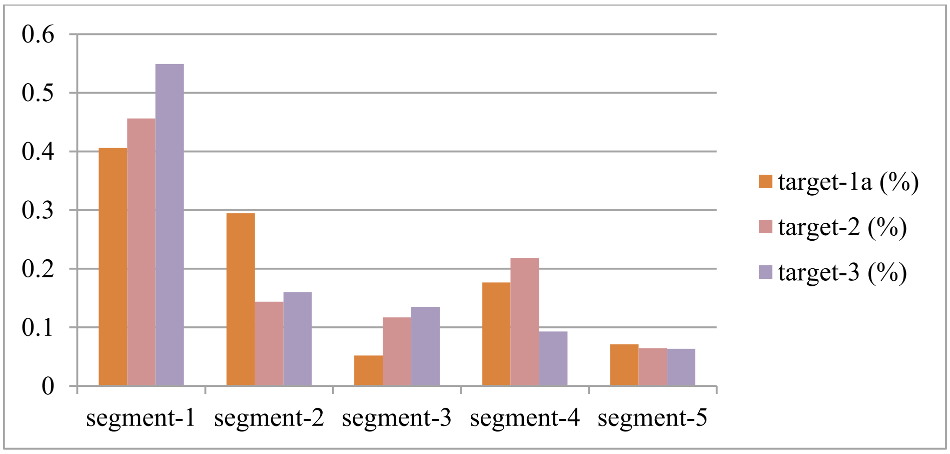


Figure 4-6 Result of Experiment 3

To evaluate the model, a number of domain experts have been consulted. They are satisfied with the segmentation result based on the telecom organisation's enterprise data, but have pointed out the limitation of some segments, such as segment-4 and segment-2.

There are many reasons for the model's limitation: the incompleteness of the enterprise data would be the main reason. Due to missing variables in the target dataset, the value population procedure of those variables could be another reason. Moreover, the inconsistent data between the source and target datasets also plays a role in the limitation. To overcome this limitation and improve the performance of the model, more work is needed.

4.5 SUMMARY

A personalised analytical eCRM framework has been developed. This framework is based on a very solid approach in developing a combination of predictive modelling and recommendation techniques within the integrated eCRM system, which has shown the building up blocks or process with the proven results in coming chapters. Also in this chapter, a hybrid customer segmentation model has been developed to tackle the migration of customer segmentation based on survey data into the customer behaviour database, which itself is a challenging topic in customer relationship management. The developed five-step hybrid segmentation method (HMCS method) particularly focuses on missing variables in the internal database. A model has been developed to replace the missing variables and populate their values to the target dataset. Experiments have shown the capability of the model to solve this kind of problem. Due to the complexity of the MCS problem, more work still needs to be done in the presented HMCS method to improve its performance. Firstly, although the obtained segmentation results are acceptable, further improvement to the artificially generated values for missing variables is still needed. More theoretical analysis is possibly required. Secondly, the model's segmentation result has significant differences in some segments compared to the initial segmentation result. How to the difference can be reduced requires more study. Finally, the essence of the MCS problem is a clustering algorithm; hence, how to build an appropriate clustering algorithm for this kind of problem is also an important issue to be studied.

CHAPTER 5

ATTRITION RISK PREDICTION MODELS

5.1 OVERVIEW

In many industries, customer attrition or churn has always been the top business priority because it is more profitable to retain existing customers than to seek new customers. It is very important to understand the key factors that lead to customer attrition. Organisations can use this understanding to effectively drive customer retention programs to prevent customers leaving, and to retain the most profitable customers.

There are many different modelling techniques for the prediction of customer churn, as reviewed in Chapter 2. Several data mining methods may be used to construct models to estimate customer attrition, such as the logistic method, the Cox regression method and the tree-based classification method. However, each may be better suited to one particular application over another. Most researches focus on a particular modelling technique in terms of how to improve the model accuracy and but few attempts have been made to apply hybrid modelling techniques. As mentioned earlier, customers may leave an organisation for multiple reasons (e.g., account cancellation or switching to a competitor), and the combination of different reasons is not linear. Thus a critical issue is how to efficiently build a model one or more methods to maximise the possibility of capturing key risk factors which related to customer attrition. Although all modelling methods are straightforward in theory, combination modelling methods are very tough in practice, as a result of the magnitude of the observations and variables.

In this chapter, a prediction framework that deals with non-linearity associated with customer attrition by using an enormous amount of data is proposed. Three predictive models are developed in this framework, using different data mining methods and statistical modelling techniques to develop prediction models of customer attrition and compare their prediction power, using data from a major bank.

The structure of this chapter is as follows: three customer attrition segments, namely regular transactors, new transactors and sporadic transactors based on the behaviour of customer externally transfer funding out have been identified in Section 5.2, which assist in establishing the risk prediction framework for bank customer attrition. In Section 5.3, a sporadic risk prediction approach is presented by using combined prediction modelling techniques. Two other predictive models are developed in Section 5.4. The validation of the models is shown in Section 5.5, and Section 5.6 gives a summary of this chapter.

5.2 A Risk Prediction Framework For Bank Customer Attrition

Customer Attrition is a function of customer transaction demographics, account holding and service related characteristics, and is also a combination of actions related to cancellation and switching to a competitor. When these causes cannot be separated, it is necessary to combine them into the model as a single measure of attrition. The prediction of customer attrition is an important business intelligence application and has attracted the attention of researchers for a long time (Au, Ma & Li 2003). In general, in fitting a customer attrition model, all customer transaction information, the time of change to an account's status, and customer/service/demographic characteristics will be used as initial inputs into modelling the preliminary analysis. It is subsequently necessary to identify the association between customer attrition and transaction behaviours, as well as other characteristics (Buckinx & Van den Poel 2005).

One of the main objectives of modelling customer attrition is to determine the causal factors, so that the bank can attempt to prevent attrition from occurring in the future. Some banks want to prevent their valued customers from transferring their funds to external competitors, which is a strong signal that those profitable customers will eventually be lost. Existing bank reports have shown that more than 40% of bank customers transfer their funds to another finance institution every year. They also indicate that this is a key risk factor in relation to customer attrition, and may be for numerous reasons, such as unmet expectations, low perceived value, competitive attraction, or unexpressed and unresolved complaints.

The major challenge is how to effectively identify and understand the behaviour of customers who transfer funds externally so that the decision making of financial managers can be supported effectively. To appropriately identify funds at risk of flowing externally, the various customer segments not only require different modelling approaches but also require different overall marketing programs. This chapter uses the case of a real bank to establish a risk prediction framework to counter the risk of funds flowing externally. This framework will enable the bank to retain deposit funds that would otherwise be lost to a competitor, albeit at a reduced margin, without cannibalising funds already held and without risk of loss to a competitor. Another contribution in this chapter is that it develops a risk prediction approach which has three steps and a combined sporadic risk prediction model. Real world dataset-based experiments fully support the developments.

A sample of customers from a major bank was used to produce an initial analysis and consisted of customers who had held at least one type of saving account and were still active at the end of the observation period. Looking at the previous seven months transaction window, it was been found that about 44% of funds had flowed externally and that recent high value deposits were key drivers for identifying customers 'at risk' of external migration, or 'attrition'. Some analysis outcomes are as follows:

1. Identified about 10% of customers that were most likely to drive about 44% the majority of value loss or cannibalisation for existing deposit products, also, this group customers would exit bank in certain period time.

2. This high risk group of customers that drive value loss, have different frequency of transferring funds out during the observed 7-months period, a relatively small proportion of customers actually just started to exhibit funds flow behaviour.
3. The analysis found that 80% of customers that had an external transaction (ET) during a particular month would continue to make an external transaction in the following month. Out of these customers, 97.5% have made regular external transactions during the entire 7-month period. That is, the customers with consecutive external transactions over two months can be defined as regular transactors.
4. As a result, the total pool of high risk customers, those who are more likely to actually exhibit funds flow behaviour need to be identified with greater precision to maximise marketing effectiveness. Also, the relationship between what drives value loss and what drives attrition needs to be understood.

From the above analysis, it is clear that customer attrition is normally found within in three discrete customer segments;

Segment1: customers who transfer funds externally - Recent and Repeat (regular transactors)
Segment2: customers who transfer funds externally - Sporadic (irregular transactors)
Segment3: customers who have just started to transfer funds externally (new transactors)

These three customer segments can be identified by the process shown in Figure 5-1.

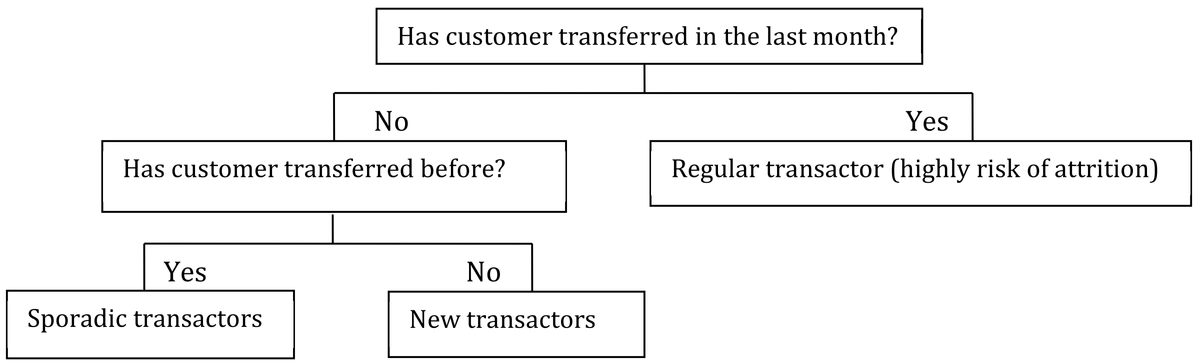


Figure 5-1 High Level of the Split Points of the Three Identified Segments

Based on the flow chart above, customers who regularly transact their funds out should be contacted automatically for Winback, and customers who sporadically transact their funds out should be contacted proactively prior to their next transfer. Ideally, customers who might start to transact their funds out should be contacted proactively prior to the first transfer. Therefore, the flow chart would help to identify who should be treated as a 'Winback' customer and who should be a 'pre-empt' customer with predictive modelling.

Brand new customers are the ones whose behaviour may be changed. The new attrition model seeks to predict the likelihood of a customer starting to transfer their funds externally and to prevent these customers becoming 'Recent and Repeat Offenders – Winback'. Sporadic offenders may be identified by to predict the timespan until the likelihood of the customer transferring their funds externally recurs. The customers who are identified for Winback are very likely to continue to repeat this behaviour and make up the vast majority of the volume (both in the number of accounts and dollar value). One way to prevent this is to use triggers to drive customer contact. Thus, the Online Saver model is built to predict which customers are most likely to open an Online Saver account and deposit funds into it; this model can be used to prioritise the Winback campaign and identify opportunities to acquire regular saving customers from the broader deposit customer base.

It was found that about 80% of the ET population are likely to be regular transactors, and that this group of customers would be most at 'risk' of attrition. The regular transactors group very likely to continue to repeat this behaviour and makes up the vast majority of the volume (both number of accounts and $ value) for which triggers can be used to drive customer contact to prevent attrition or attempt Winback.

Another 15% of the ET population as are sporadic transactors. The actions of this group should be pre-empted with predictive modelling and may be identified by predicting the time when a customer is likely to make their next external funds transfer.

The remaining 5% are new-start transactors. The predictive model can be built to predict the customer likelihood of starting to transfer their funds externally and used to prevent these customers becoming regular transactors.

As a result, three models need to be built in the risk prediction framework (shown in following chart) to handle the three discrete customer segments identified. In each case, the purpose of our framework is to act pre-emptively, and we therefore call our risk prediction framework the Pre-emptive Attrition Framework (PAM). The three models we shall build within this framework are:

1) Regular Attrition Model;

2) Sporadic Model;

3) New Customer Attrition Model

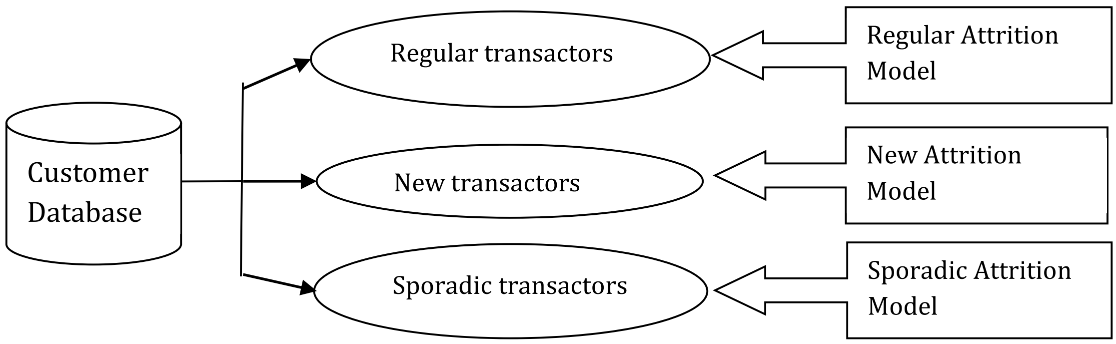


Figure 5-2 A Pre-emptive Attrition Framework for Bank Customers

To appropriately identify these customer segments not only requires different modelling approaches, but also different overall marketing approaches. It is intended that outbound phone calls will be made to those customers that score highly for both value and incidence. Whilst pro-active contact will be made, offers of high interest rate deposit products are likely to be offered only reactively, enabling the bank to retain deposit funds that would have been lost to a competitor, albeit at a reduced margin, without cannibalising funds already held and without the risk of losing the customer.

In the next section, a sporadic attrition model is developed, followed by the regular attrition model and the new customer attrition model.

5.3 A SPORADIC RISK PREDICTION APPROACH AND A COMBINED PREDICTION MODEL

In this section, a predictive modelling process is developed by combining three different modelling methods that are most often used to predict customer attrition: 1) tree-based method, 2) logistic method, and 3) Cox regression method.

5.3.1 SPORADIC ATTRITION MODEL DESCRIPTION

There are three steps to achieving this prediction model, which predicts the likelihood of customers sporadically transferring their funds externally.

Step 1: Narrow down the number of variables

Through data analysis, 3000+ variables are initially collected from the customer database, which include all accounts (e.g. credit card, home loan, saving, cheque, personal loan, etc.) daily transactions records for each customer. Let Y is a dependent variable and $X_i = (x_{1n1}, x_{2n2}, x_{3n3}, \dots, x_{ink})$ defined as the groups of independent variables that represent every daily transactions detail records for all kinds of accounts; so that $X_1 = (x_{11}, x_{12}, x_{13}, \dots, x_{1n1})$ is defined as all daily transactions records for account 1, $X_2 = (x_{21}, x_{22}, x_{23}, \dots, x_{2n2})$ is defined as all daily transactions records for account 2, $X_3 = (x_{31}, x_{32}, x_{33}, \dots, x_{3n3})$ is defined as all daily transaction records for account 3, etc. $X_i = (x_{i1}, x_{i2}, x_{i3}, \dots, x_{ink})$ is defined as all daily transaction records for account *i*. where $n1+n2+n3+\dots+nk = 3000$, i refer to the number of accounts. Each customer has different number of accounts.

As the number of variables is far too many for the prediction process, the first step is to reduce the variable dimension to an acceptable range. Chi-squared Automatic Interaction Detection (CHAID) is used for this purpose as it is a highly efficient statistical technique for segmentation (Kass 1980). It is a classification method for building decision trees by using chi-square statistics to identify optimal splits. CHAID evaluates all of the values of a potential independent variable. It merges values that are judged to be statistically homogeneous (similar) with respect to the dependent variable and maintains all other values that are heterogeneous (dissimilar).

Let $h = \{x_p, y_p\}_{p=1}^{N}$ the whole sample size, w_p is the case weight associated with case p and f_p is the frequency weight associated with case *p*. Non-integral positive value is rounded to its nearest integer.

The following CHAID algorithm has defined by Kass (Kass 1980), which only accepts nominal or ordinal categorical dependent variables. When dependent variables are continuous, they are transformed into ordinal dependent variables before using the following algorithm.

Merging step;

For each independent variable X_k, merge non-significant categories. Each final category of X_i will result in one child node if X_i is used to split the node. The merging step also calculates the adjusted *p-value* that is to be used in the splitting step.

1. If X_i has 1 category only, stop and set the adjusted *p-value* to be 1.

2. If X_i has 2 categories, go to step 8.

3. Else, find the allowable pair of categories of X_i (an allowable pair of categories for ordinal independent variable is two adjacent categories, and for nominal independent is any two categories) that is least significantly different (i.e., most similar). The most similar pair is the pair whose test statistic gives the largest *p-value* with respect to the dependent variable Y.

4. For the pair having the largest p-value, check if its p-value is larger than a user-specified alpha-level α_{merge} (*alpha_merge*). If it does, this pair is merged into a single

compound category. Then a new set of categories of X_i is formed. If it does not, then go to step 7.

5. (Optional) If the newly formed compound category consists of three or more original categories, then find the best binary split within the compound category which *p-value* is the smallest. Perform this binary split if its *p-value* is not larger than an alpha-level $\alpha_{split-merge}$ (*alpha_split-merge*).

6. Go to step 2.

7. (Optional) Any category having too few observations (as compared with a user-specified minimum segment size) is merged with the most similar other category as measured by the largest of the *p-values*.

8. The adjusted *p-value* is computed for the merged categories by applying Bonferroni adjustments that are to be discussed later.

Splitting step;

The "best" split for each independent variable is found in the merging step. The splitting step selects which independent variable to be used to best split the node. Selection is accomplished by comparing the adjusted *p-value* associated with each predictor. The adjusted *p*-value is obtained in the merging step.

1. Select the independent variable that has the smallest adjusted *p-value* (i.e., most significant).

2. If this adjusted *p-value* is less than or equal to a user-specified alpha-level split α_{split} (*alpha_split*), split the node using this independent variable. Else, do not split and the node is considered as a terminal node.

Stopping step;

The stopping step checks if the tree growing process should be stopped according to the following stopping rules.

1. If a node becomes pure; that is, all cases in a node have identical values of the dependent variable, the node will not be split.

2. If all cases in a node have identical values for each independent variable, the node will not be split.

3. If the current tree depth reaches the user specified maximum tree depth limit value, the tree growing process will stop.

4. If the size of a node is less than the user-specified minimum node size value, the node will not be split.

5. If the split of a node results in a child node whose node size is less than the user-specified minimum child node size value, child nodes that have too few cases (as compared with this minimum) will merge with the most similar child node as measured by the largest of the *p-values*. However, if the resulting number of child nodes is 1, the node will not be split.

After applying above CHAID algorithm to the data, the initial 3000+ variables were reduced to 52 variables.

Step 2: Investigate relationships between variables and their suitability

These 52 variables were split based on the optimum divisions obtained from the decision tree in Step 1. These splits have been used to define each group as a dummy variable or to give a relative likelihood index. The interaction variables are obtained by looking at the correlations of the coefficients of the modelled variables. If two modelled variables have correlated coefficients, this indicates that there is an interaction, and as such the interaction-terms are formed.

The interaction-terms are based, at the 52 variables level, on the optimum splits from the decision tree. The intersection of these splits for two variables creates the interaction-term split. Once the splits had been obtained, the new interaction variables were analysed in identical fashion to the 52 variables and were also assigned their dummy variables. These 52 significant variables plus some interaction variables were used as inputs for performing logistic regression by using a step-wise method to optimize significant variables. As a result, a further 23 predictor variables were removed from the logistic regression process. Table 5-1 a shows the outcome of the step-wise selection process;

Table 5-1 Analysis of Maximum Likelihood Estimates

Variables	**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	-5.1532	0.1019	2559.4458	<.0001
x_1	1	-0.4275	0.1919	4.9637	0.0259
x_2	1	0.127	0.0512	6.1465	0.0132
x_3	1	0.3437	0.0972	12.5147	0.0004
x_4	1	0.2669	0.051	27.4098	<.0001
x_5	1	0.2099	0.049	18.3741	<.0001
x_6	1	0.2404	0.0849	8.0208	0.0046
x_7	1	0.5764	0.0383	226.4298	<.0001
x_8	1	0.4724	0.0579	66.6457	<.0001
x_9	1	0.703	0.1798	15.2943	<.0001
x_{10}	1	0.889	0.0929	91.6689	<.0001
x_{11}	1	0.2597	0.0991	6.8705	0.0088
x_{12}	1	0.3591	0.1132	10.0655	0.0015
x_{13}	1	0.21	0.0801	6.8735	0.0087
x_{14}	1	0.4352	0.094	21.4435	<.0001
x_{15}	1	0.5258	0.0657	63.9963	<.0001
x_{16}	1	0.9067	0.0489	344.3651	<.0001
x_{17}	1	0.2795	0.1387	4.063	0.0438
x_{18}	1	0.7595	0.1285	34.9453	<.0001
x_{19}	1	0.2835	0.0811	12.2262	0.0005
x_{20}	1	0.9194	0.1157	63.1478	<.0001
x_{21}	1	0.9407	0.109	74.528	<.0001
x_{22}	1	1.0534	0.1862	31.9911	<.0001
x_{23}	1	1.1417	0.1632	48.9086	<.0001

Step 3: Prediction model

All the remaining variables were optimised significant variables and were used to perform Cox regression to predict the time until the sporadic ET customers were next likely to transfer their funds externally. These variables will confirm the key factors that impact customers in their decision to sporadically transfer funds externally. The next question to address would be when customers will next transfer their funds externally.

To use the Cox regression model (Cox 1972) for prediction, a time variable needs to be added into the model, which is the number of weeks from a given point in time until a customer engages in sporadic external transactions (ET). The Cox regression is then used to predict when sporadic ET customers will next transfer their funds. Cox was the first to suggest models in which lifetime factors have a multiplicative effect on the hazard function. These models are called proportional hazards models. Under the proportional hazards assumption, the hazard function of t given x is of the form:

$$h(t|x) = h_0(t) \qquad (5.1)$$

where x is a known vector of predictor variables associated with the individual, β is a vector of unknown parameters, and $h_0(t)$ is the baseline hazard function for an individual with $x = 0$. Hence, for any two covariate sets x_1 and x_2, the log hazard functions $h(t|x_1)$ and $h(t|x_2)$, should be parallel across time.

When a factor does not affect the hazard function multiplicatively, stratification may be useful in model building. Suppose that individuals can be assigned to one of m different strata, defined by the levels of one or more factors. The hazard function for an individual in the j^{th} stratum is defined as

$$h_j(t|x) = h_{0j}(t)e^{x'\beta} \qquad (5.2)$$

where the regression parameter β and the baseline hazard function $h_{0j}(t)$ are two unknown components in the model.

5.3.2 *Model Fit Evaluation*

The model sample was the ET customer at one particular time, who had held at least one type of saving account and was still active at the end of the observation period. The model sample was split into a development sample (70% of the model sample – used to develop the model) and a hold-out sample (30% - used to validate the model). Adjustments were made for this sampling in the creation of model scores and results.

In predicting customer attrition, it has been assumed that the subjects are in one of two basic states: at risk or not at risk. Because of the diverse activity and speculation (i.e., there is no clear-cut distinction between the subjects at risk and the subjects not at risk) and because of the limited information associated with the outcome, the subject at risk and the subject not at risk may not always be predicted correctly. The performance of a prediction model may vary depending on the specific threshold used. Thus, an objective evaluation of a risk-predicting model should examine the overall performance of the model under all possible decision thresholds, not only one particular decision threshold. To achieve this, a useful tool has been adopted, Cumulative Gains and Lift Charts, to evaluate the performance of the predictive models. The lift chart compares the predictive performance of the mining model with an ideal model and a random model. Note that the overall predictive accuracy of this model is close to the ideal model. The greater the area between the lift curve and the random model (baseline) is, the better the model is. In other words, the response of a targeted population predicted by the model is much better than average for the population as a whole (DeLong, DeLong & Clarke-Pearson 1988).

The combined model has been compared with three stand-alone models to examine the power of prediction of these models, and the results are presented in the lift charts in Figure 5-3. From the lift chart, it clearly be seen that the combined model has 10% more lift than other models at the 1st decile, and 20% more lift in the 2nd decile (a deciles is defined as a rank probability score from largest to smallest, divided into 10 even groups; 1st decile means the top 10% of the scored population). It is confirmed that the combined models method is significantly better than the other three individual models. Also, there are other indicators that prove our combined model is well fitted:

(1) The residual Chi-squared test for model fit was 0.4738 – where values greater than 0.05 indicate a significant fit;

(2) The Hosmer and Lemeshow Goodness-of-Fit (Hosmer & Lemeshow, 2004). Test for model was 0.3152, that indicates the model is quite robustness.

(3) The -2 Log-Likelihood test for model fit is < 0.001 – where values less than 0.05 indicate a significant fit;

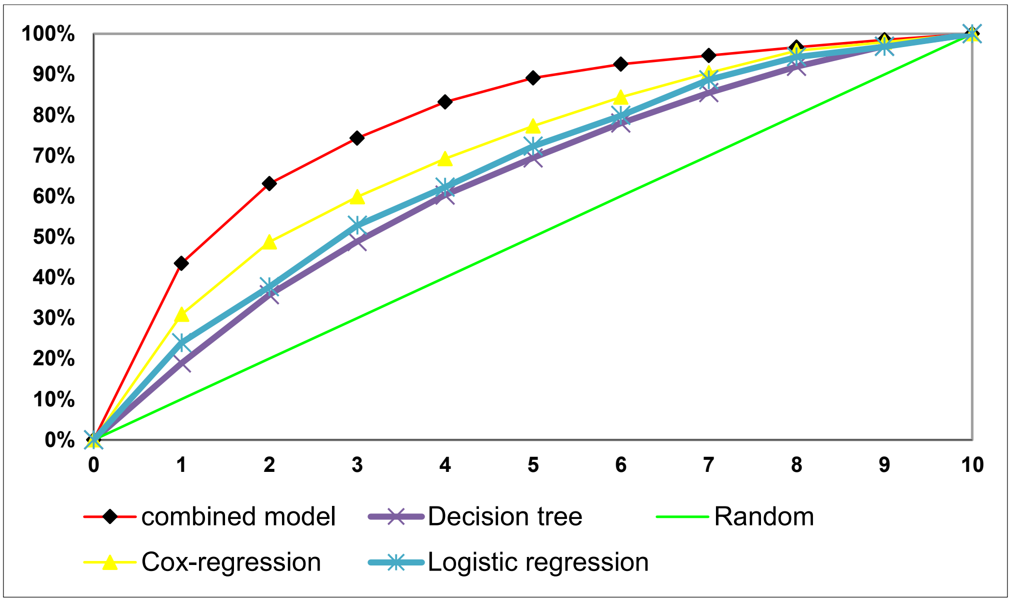


Figure 5-3 Comparison between the Combined Model and Three Other Models

5.3.3 *TOP SIGNIFICANT VARIABLES*

The model output used combined three methods and has shown the final selected variables and their relative predictive power. This power is the variable's ability to explain the variation in the data, compared to other variables. This Sporadic model shows that 80% of the variations can be explained by the top 5 variables (x_1, x_2, x_3, x_4, x_5) which are described as follows:

x_1 - Customers are more likely to transfer funds out within 4 weeks of the balance of their personal savings account having a higher growth rate over the previous 3 months;

x_2 - Customers with little change in their savings balance are more likely to transfer externally within 8-12 weeks' time;

x_3 - The higher growth rate in debit transaction volume, the lower the likelihood of transferring externally;

x_4 - Younger customers (age less than 31) are more likely to transfer their funds out early (in two weeks' time) - older customers (age great than 50) are more likely to move their money out in 11 to 14 weeks;

x_5 - Generally, the higher the variation of direct transaction volume, the more likely it is that the customer will transfer early time (1-5 weeks). Conversely, the lower the variation, the more likely it is that the transfer will occur (10-15 weeks).

5.4 Regular Attrition Model and New Customer Attrition Model

To assist in the prioritisation of the prevention campaign and identify opportunities to acquire online saver customers, it is necessary at same time to target those customers who regularly deposit externally, prevent new customers from starting to transfer funds externally. For the regular attrition case, the Online Saver model can be built to predict those customers who are most likely to open an Online Saver account and deposit funds into it. The New Customer Attrition Model seeks to predict the likelihood of customers starting to transfer their funds externally.

The model sample was defined in Section 5.3.2, and the sample was split into a development sample (70% of the model sample – used to develop the model) and a hold-out sample (30% - used to validate the model). Adjustments were made for this sampling in the creation of model scores and results. The new attrition model sought to target those customers who transferred funds for the first time in defined particular month, and continued to transfer funds in the following month. The online model would target those customers who opened an Online Saver account and deposited funds in the defined month and also in the following month.

5.4.1 New Customer Attrition Model And Online Saver Description

A similar modelling process to that described in 5.4.1 was used to produce the two prediction models that predict the likelihood of customers starting to transfer their funds externally, and the likelihood of customers regularly transferring their funds externally.

As demonstrated in Section 5.3.1, the first step is to reduce the variable dimension to an acceptable range. The Chi-squared Automatic Interaction Detection (CHAID) is used for this purpose. CHAID evaluates all the values of a potential predictor field. Therefore, initial 3000+ variables were reduced to 43 variables for the new attrition model and to 38 variables for the Online Saver model.

These 43 variables and 38 variables were split respectively based on the optimum divisions obtained from the decision tree in the first step. Using the exact method described in Section 5.3.1, these splits were used to define each group as a dummy variable or to give a relative likelihood index. The interaction-terms are based on the optimum splits from the decision tree, at the 43 variables and 38 variables levels. Once the splits had been obtained, the new interaction variables were also assigned their dummy variables or index. Thus these 43 significant variables of the new attrition model plus some interaction variables, and the 38 significant variables of the Online Saver model plus some interaction variables were used as inputs to perform logistic regression by using a step-wise selection process. As a result, a further 34 predictor variables were removed from the logistic regression process for predicting new attrition. Table 5-2 shows the outcome of the step-wise selection process. A further 28 predictor variables were also removed from the logistic regression process for predicting the Online Saver. Table 5-3 shows the outcome from the step-wise selection process.

Table 5-2 Analysis of Maximum Likelihood Estimates -New Attrition Model

Variables	**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
x_1	1	0.2835	0.0811	12.2262	0.0005
x_2	1	0.2597	0.0991	6.8705	0.0088
x_3	1	0.3591	0.1132	10.0655	0.0015
x_4	1	0.21	0.0801	6.8735	0.0087
x_5	1	0.4352	0.094	21.4435	<.0001
x_6	1	0.5258	0.0657	63.9963	<.0001
x_7	1	0.9067	0.0489	344.3651	<.0001
x_8	1	0.2795	0.1387	4.063	0.0438
x_9	1	0.7595	0.1285	34.9453	<.0001

Table 5-3 Analysis of Maximum Likelihood Estimates- Online Saver Model

Variables	**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	-5.1532	0.1019	2559.4458	<.0001
z_1	1	-0.4275	0.1919	4.9637	0.0259
z_2	1	0.127	0.0512	6.1465	0.0132
z_3	1	0.3437	0.0972	12.5147	0.0004
z_4	1	0.2669	0.051	27.4098	<.0001
z_5	1	0.2099	0.049	18.3741	<.0001
z_6	1	0.2404	0.0849	8.0208	0.0046
z_7	1	0.5764	0.0383	226.4298	<.0001
z_8	1	0.4724	0.0579	66.6457	<.0001
z_9	1	0.703	0.1798	15.2943	<.0001
z_{10}	1	0.889	0.0929	91.6689	<.0001

5.4.2 MODEL FIT EVALUATION

Cumulative Gains and Lift Charts have been used to evaluate the performance of the predictive models (see Section 5.3.2). Figure 5-4 shows the lift chart from the New Customer Attrition Model, from which it can be seen that the top 10% of the development sample (or top decile) has an uptake rate 3.1 times higher than random.

Similarly, it can be seen that the top two deciles (top 20%) capture almost 50% of the targets and produce a cumulative rate of twice that of a random sample.

The Gini index represents the degree of separation between the targets and non-targets achieved by the model, with a Gini index of 0 indicating no separation and an index of 25 or more indicating significant separation.

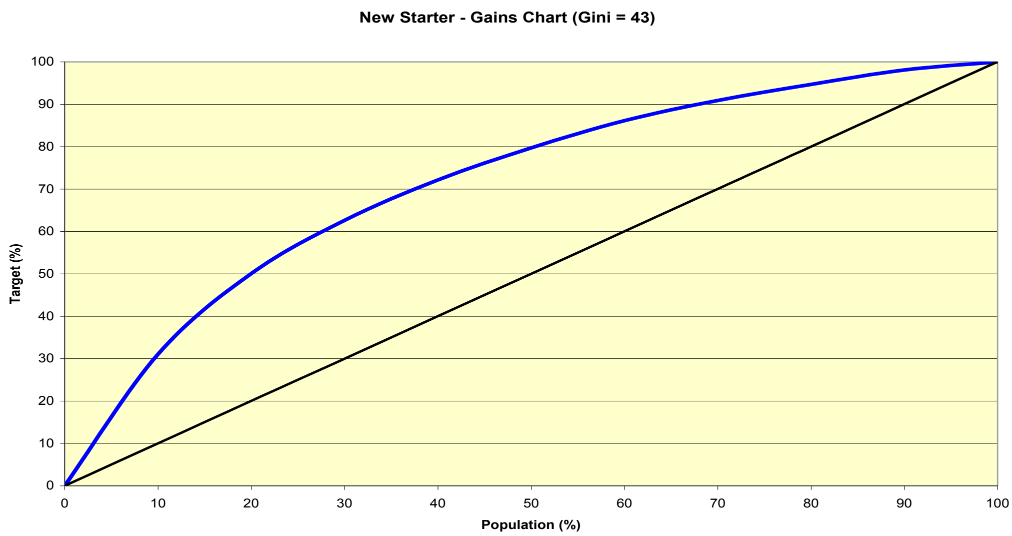


Figure 5-4 Lift Chart Based on the New Attrition Development Sample

Figure 5-5 illustrates the lift chart from the Online Saver model, which shows that the top 10% of the development sample (or top decile) for the Online Saver has an uptake rate 4.7 times higher than random. Similarly, it can be seen that the top two deciles capture about 65% of the targets and produce a cumulative rate of three times that of a random sample.

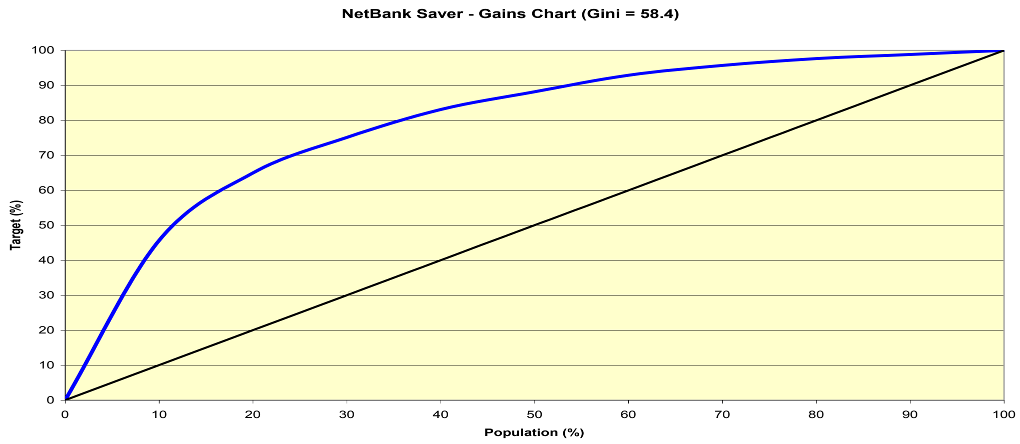


Figure 5-5 Lift Chart Based on the Online Saver Development Sample

The main points to note in the output of the two models, listed below, are further proof that the two models are well fitted:

New Customer Attrition Model

(1) The residual Chi-squared test for model fit is 0.5391, where values greater than 0.05 indicate a significant fit;

(2) The Hosmer and Lemeshow Goodness-of-Fit Test for the model is 0.4152, indicating that the model is quite robust.

(3) The selected variables and their estimates will be used in conjunction with their indices to score future populations.

Online Saver Model

(1) The residual Chi-squared test for model fit is 0.4569 – where values greater than 0.05 indicate a significant fit;

(2) The Hosmer and Lemeshow Goodness-of-Fit Test for the model is 0.0789, indicating that the model is quite robust.

(3) The -2 Log-Likelihood test for model fit is < 0.001, where values less than 0.05 indicate a significant fit.

5.4.3 SIGNIFICANT VARIABLES FROM MODELLING

The output result after using combined methods shows the final selected variables and their relative predictive power. This power is the variable's ability to explain the variation in the data, compared to the other variables. The new attrition model shows that 95% of the variation is explained by the top 5 variables (x_1, x_2, x_3, x_4, x_5), which are described as follows;

x_1 - Younger customers (aged between 20 and 29) and never married are most likely to move their funds out.

x_2 - Customers with more phone transactions and fewer debit transactions per week from their personal savings account are more likely to transfer funds out.

x_3 - Customers with higher annual contents insurance and who pay a higher level of tax annually are more likely to move their funds out. This suggest that customers with higher salaries have higher wealth possessions.

x_4 - Customers with a higher number of online debit transactions are more likely to transfer funds out. Customers are less likely to transfer externally if their average number of credit transactions is higher.

x_5 - The customer takes out less money per week from the ATM and he/she is more likely to transfer funds out.

The results from the Online Saver model show that 85% of the variation can be explained by the top 5 variables (z_1, z_2, z_3, z_4, z_5), which are described as follows;

z_1 - Customers are most likely to take up the Online Saver account and deposit their funds in it if there is a variation in the transaction volume of their personal saving accounts via the Online Saver account per week

z_2 - Customers with a higher predicted balance of for the next month for their personal saving account, and a lower balance in the current month are more likely to take up the Online Saver account and deposit funds in it.

z_3 - Customers with a longer period since their last credit transactions from personal saving accounts are more likely to take up the Online Saver account.

z_4 - Customers with high variation in direct credit transaction amounts are more likely to take up an Online Saver account. Customers are less likely to take up such an account if the average amount of direct credit transactions does not change.

z_5 - Customer with higher the number of online debit transactions are more likely to transfer funds out. Customers are less likely to transfer externally if the average number of credit transactions is higher.

5.5 EXPERIMENTS AND ANALYSIS

To evaluate the prediction power of the model, two independent samples have been built. The model is first developed using a 'learning' (or 'train') sample and then validated by applying it to a 'validation' (or 'test') sample to determine the extent to which the model may be generalized beyond the original 'learning' sample. This is a standard procedure for fitting scoring models (models that estimate the probability (score) of an event, such as attribution) using combined statistical models and data mining techniques.

Figure 5-6 compares the validation distributions with the development decile distributions for the three models. The decile distribution for the development sample is calculated by ranking customers by the modelled propensity and forming ten equal groups. These cut-offs (model score bands) are then applied to the validation sample to compare the discrimination of the model on the validation sample.

The plots show that while the validation distribution varies somewhat from the development distribution, it still maintains effective discrimination, particularly in the higher propensity deciles (one through to four). The variation from the development distribution may be explained by the relatively low number of targets in the validation sample, and hence the high weightings required. Overall, it demonstrates that the three models obtained from the development sample also perform well on the hold-out validation sample.

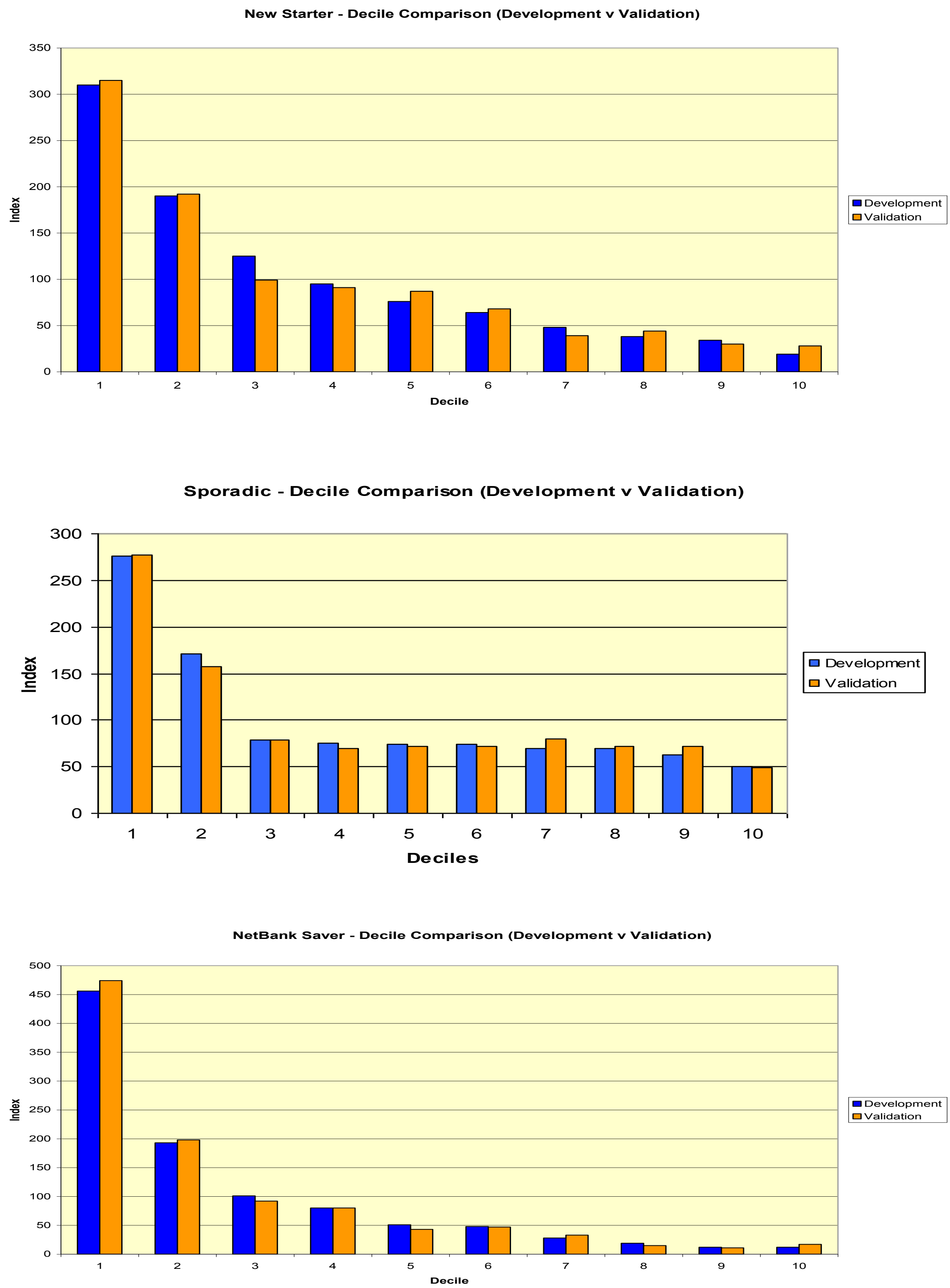


Figure 5-6 Decile Comparison between the Training Sample and the Validation Sample

The New Customer Attrition Model based on the validation (or hold-out) sample shows two important features which indicate that the model performs well on the validation sample. Firstly, the cut-offs obtained from the development sample produce a relatively uniform decile distribution in the scored validation sample. This indicates that there are no large biases in the modelled development sample. Secondly, the index effectively discriminates the validation sample (as shown graphically and explained above).

Figure 5-7 shows the lift chart achieved in both the development sample and the validation sample. It is clear from the chart that the gains are equal for both populations and the model performs equally well on the validation sample.

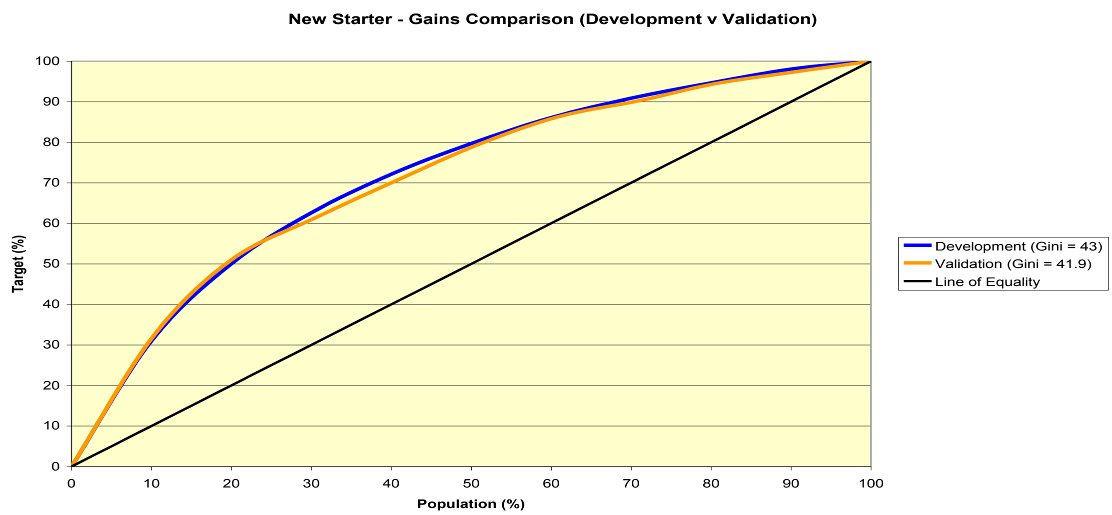


Figure 5-7 Gains Chart Based on the New Customer Attrition Validation Sample

Similarly, the Online Saver model based on the validation (or hold-out) sample indicates that the model also performs well on the validation sample. Figure 5-8 shows the lift chart achieved a great lift in both the development sample and the validation sample. It is clear from the chart that the gains are equal for both populations and the model performs equally well on the validation sample.

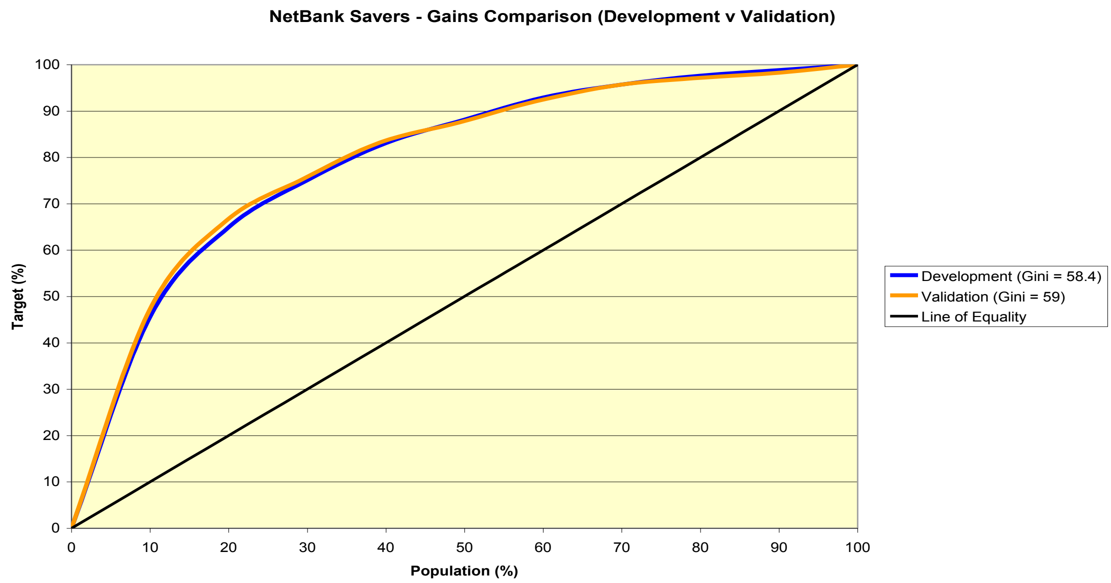


Figure 5-8 Gains Chart Based on the Online Saver Validation Sample

As is known, the sporadic model is designed to predict the time until customers are next likely to transfer their funds externally. Table 5-4 compares the predicted time versus actual time. It is clear from this table that the model prediction time at 1-4 weeks has an 84% accuracy rate and a 71% accuracy rate at 9+ weeks.

Table 5-4 Validation of Predicted Time versus Actual Time

Predicted Time	**Actual Time**		
	1-4 weeks	**5-9 weeks**	**9+ weeks**
1-4 weeks	84%	2.40%	1.00%
5-9 weeks	17%	57.80%	16.20%
9+ weeks	10%	2.40%	71.10%

The following tables demonstrate the prediction power by of recency groups.

Table 5-5 Prediction of Attrition between 1 and 4 Weeks

Recency	Sample Population	Actual Attrition (Week1-4)	Predicted Attrition (Week1-4)	Model Prediction Rate	Predicted Recency	Base Recency
2	11040	3895	5564	70.00%	50.40%	40.70%
3	6040	1495	2464	60.67%	40.79%	18.90%
4	3652	903	1507	59.92%	41.27%	13.30%
5	2775	592	1054	56.17%	37.98%	10.80%
6	2560	339	745	45.50%	29.10%	9.40%
Total	26067	7224	11334	63.74%	43.48%	18.62%

Table 5-6 Prediction of Attrition between 5 and 8 Weeks

Recency	Sample Population	Actual Attrition (Week 5-8)	Predicted Attrition (Week5-8)	Model Prediction Rate	Predicted Recency	Base Recency
2	11040	1068	2504	42.65%	22.68%	10.70%
3	6040	845	1669	50.63%	27.63%	11.90%
4	3652	368	940	39.15%	25.74%	13.30%
5	2775	303	746	40.62%	26.88%	10.80%
6	2560	228	657	34.70%	25.66%	9.40%
Total	26067	2812	6516	43.16%	25.00%	11.22%

Table 5-7 Prediction of Attrition greater than 9 Weeks

Recency	Sample Population	Actual Attrition (Weeks 9+)	Predicted Attrition (Weeks 9+)	Model Prediction Rate	Predicted Recency	Base Recency
2	11040	1783	2972	59.99%	26.92%	10.70%
3	6040	1058	1879	56.31%	31.11%	9.90%
4	3652	604	1205	50.12%	33.00%	11.30%
5	2775	426	975	43.69%	35.14%	10.80%
6	2560	312	1158	26.94%	45.23%	9.40%
Total	26067	4183	8189	51.08%	31.42%	10.42%

Table 5-5 demonstrates that the Sporadic development sample for customers who did not transfer funds during the last month has a predicted rate of 50.40% for funds transference for the next month and 70% of customer attrition within four weeks. The model predicts the likelihood overall of 43.48% customers transferring their funds externally next month, compared with a baseline rate of 18.62%. This is 2.7 times higher than random. The model also predicts the likelihood of customer attrition from this group within 4 weeks, with an accurate attrition rate of 63.74% as shown in Table 5-5. Similarly, there is a 25% likelihood of customer transferring funds next month, with this 25% of customers have a prediction of 43% attrition between 5 and 8 weeks in Table 5-6. The same explanations can be seen in Table 5-7.

5.6 SUMMARY

In real-world applications, the main challenge is to mine a huge amount of information from a data warehouse and to identify the key risk factors for customer attrition. This chapter presents a prediction framework for dealing with the non-linearity association with customer attrition where there is an enormous amount of data. It particularly demonstrates the use of the proposed regular transactor, sporadic transactor and new transactor risk prediction approaches, and the combined prediction modelling techniques. These techniques have been successfully applied in a bank's customer attrition management system. Real world experiments have validated the proposed framework, approach and models, and have supported bank managers' decision making in bank customer attrition prediction.

While this chapter has been offered an alternative modelling approach, there are still a number of interesting avenues to pursue. First, the underlying model can be further refined and enriched by admitting various classes of fuzzy sets (membership functions). Second, more experiments using various datasets coming from other industries would be advantageous to make a better assessment of the performance of the models. In the following chapter, a case study using fuzzy sets will be presented to define the level of risk by employing different datasets.

CHAPTER 6

AN INTELLIGENT CUSTOMER CHURN MANAGEMENT MODEL AND CASE STUDY

6.1 OVERVIEW

This chapter proposes an advanced customer churn management model, including a customer churn profile model, a customer action model, a customer experience model and a fuzzy customer risk model. At the same time, an additive logistic regression model is developed by using boosting, which is more robust and has shown success in churn prediction in the banking industry. As the churner usually takes only a fraction of the customer base, the problem of customer churn prediction is always combined with the problem of highly skewed class distribution or lack of churner data. One of the most common techniques for dealing with rarity is sampling (Chen, Fan & Sun 2012). Methods that adopt sampling technique alter the distribution of training examples and generate balanced training sets for building churn prediction models (Abbasimehr, Setak & Tarokh 2011; Burez & Van den Poel 2009; Kim et al. 2012; Nie et al. 2011). By synthesising the results of multiple models of different types, not only can the model predict the churners, but it can also comprehend the issue of customer churn, thereby enhancing insight into customer behaviour.

This chapter is organised as follows. Section 6.2 describes the customer churn management model. Section 6.3 provides the results of the experimental evaluation.

Section 6.4 summarises the significant contribution with a discussion of potential works on this topic.

6.2 Customer Churn Management Model – A Real World Case Study

This section proposes a comprehensive and complete customer churn management model that consists of multiple models to improve modelling accuracy and achieve maximised modelling performance, working with a real world case study.

6.2.1 Case Description

A telecom organisation is currently facing customer retention challenges. One of its key priorities is to prevent customer churn, that is, a customer's decision to end the relationship with the organisation and switch to a competitor. One question emerges: is there an appropriate way to convert this risk relationship into a stronger relationship? The organisation consequently devotes effort to understanding customer churn by examining 'who are they', 'how do they behave', 'why do they churn', 'when do they like to churn' and 'how companies can prevent' customers from taking the decision to leave.

As a result, customer churn prediction is defined as a very important approach to identify the customers that are most likely to demonstrate churning behaviour. In general, the method of the prediction is to generalise the relationship between churning behaviours and customer information in a model that can be used for prediction purposes. It has been known that customer churn is not caused by a single reason, and usually multiple reasons. It is quite difficult to know which reason applies. To detect which customers are about to churn in a short period of time and to know them in depth, an advanced customer churn management model is developed.

In this case study, one sample of mobile customer data from a telecom company has been modified and employed which includes a segment of mobile customers (in the number of millions) that are active at a point in time of the year 2010. Initially, 700+ variables are extracted from the mobile customer database, including mobile plan and contract information, billing, usage, and product holding information, as well as customer services inbound/outbound information. All these variables are defined as independent variables $X_i = (x_1, x_2, \dots, x_n)$. Each mobile customer who churned one month after that point of time of the year 2010, labelled (yes/no), becomes a dependent variable or target variable y_i .

In real business cases, there are two types of churn behaviour: voluntary churn, in which a customer decides to terminate services; and involuntary churn, in which the service provider decides to terminate a customer's services, typically because of financial liability, such as the inability pay accounts (Datta et al. 2000). This case study considers only voluntary churners, because involuntary churners are easier to identify and are of less importance from a retention management perspective.

In contrast to most existing churn prediction models, the developed prediction model allows for an 'Implementation Zone', within which the company is able to perform retention actions. As a result, customers who churn within the Implementation Zone are excluded from the model building process. The modelling timeline is given in Figure 6-1. The timeline below shows that the model, in essence, aims to predict all customers who will churn voluntarily in the next two-month period (in the outcome window), based on their latest three-month information (in the behaviour window).

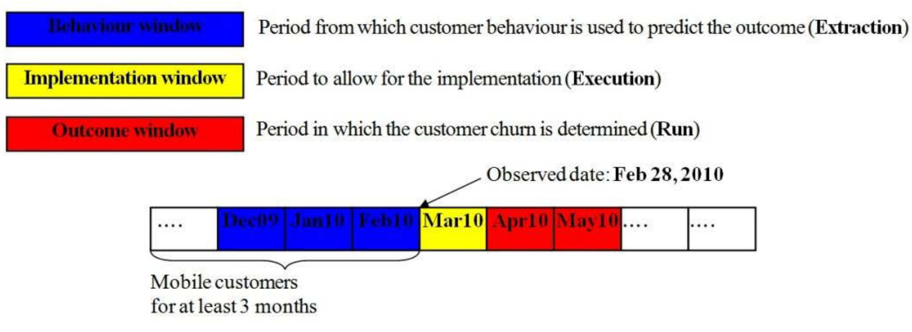


Figure 6-1 Timeline of Churn Prediction Model

With this timeline model, for instance, if data has been extracted from Dec. 2009 to Feb. 2010, the churner label is determined by the two-month period of time after Mar. 2010 (Apr. 2010 – May 2010). Note that this allows for an Implementation Zone.

6.2.2 *MODELLING PRE-PROCESS*

In Section 6.2.1, customer information be extracted and formulated into 700+ variables. Most of these variables are irrelevant, or have little relevance to churn prediction. The excess variables not only burden the computational process in the generation of a model, but they also interfere with the heuristic search for all practical model generation algorithms and result in an inferior model (Schafer, Konstan & Riedi 1999). Thus, it is necessary to continue to evaluate the variable set in an attempt to eliminate some based on their predictive performance, and to produce a more predictive model.

There is a two-step process for variable selection, which is a similar process to that developed in Chapter 5.

Step1: Use CHAID Analysis to narrow down the number of variables from more than 700 variables to 70 variables.

The CHAID Analysis (Kass 1980) is a form of analysis that determines how variables best combine to explain the outcome in a given dependent variable, which incorporates a sequential merge and split procedure based on a chi-square test statistic. CHAID uses a sub-optimal split on each predictor instead of searching for all possible combinations of categories, which reduces computation time. CHAID has been chosen because it is especially useful for data expressing categorised values; also, it evaluates all of the values of a potential predictor field and selects the predictor that is the most significant (smallest p value); its output is highly visual, having a tree image, and is easy to interpret, which is very important from the perspective of checking by expert knowledge and retention management. As a result of CHAID Analysis, 700+ variables were narrowed down to 70 variables. Nevertheless, the number of variables is still far too many for an optimised prediction process.

Step2: Further narrow the number of variables to 21 by using a logistic regression process.

In Step 1, these 70 variables are split according to the optimum divisions obtained from a decision tree. Each split is defined as one group, and each group is assigned a dummy variable. The interaction variables are obtained by looking at the correlations of the coefficients (> 0.3) of the modelling variables. These 70 variables, plus a number of interaction variables, will be the inputs for a logistic regression. A stepwise-selection procedure is employed, in which variables are added to the model one at a time until a pre-set stopping rule is satisfied. Finally, 21 optimised significant variables are identified and used to form a logistic regression model.

6.2.3 CUSTOMER CHURN PROFILE MODEL

In Section 6.2.2, for each customer, a probability, denoted as $p_i(x)$, is calculated from the logistic regression model;

$$p_i(x) = \frac{1}{1 + e^{\sum_1^{21} \beta_i x_i}} \tag{6.1a}$$

which predicts the likelihood that a customer will churn sometime in the future, thereby enabling us to identify customers who are most likely to churn. However, as the training set is highly skewed, it is necessary to improve the predictiveness of the logistic regression model where the churners account for less than 10% of the data, by employing the Gentle AdaBoost (Hong and Weiss 1999) algorithm. The procedure of how Gentle AdaBoost works has already been depicted in Section 2.2.1. To fit a regression equation $g(x)$ using weighted least-squares with sample weight D, the objective function is defined as

$$J = \sum_{i=1}^{N} D(i) \times \left(\frac{1+y_i}{2} - p(x_i) \right)^2 \tag{6.1}$$

where $p(x_i) = \frac{1}{1+e^{-g(x_i)}}$, $y_i = \{+1,-1\}$.

Using the gradient descent algorithm to optimize the regression equation by minimizing J. gives the gradient for the coefficient

$$\nabla J = \frac{\partial J}{\partial \vec{\beta}} \quad \text{where} \quad \frac{\partial J}{\partial \beta_j} = -2\sum_{i=1}^{N} D(i)\left(\frac{1+y_i}{2} - p(x_i)\right)\frac{e^{-g(x_i)}}{\left(1+e^{-g(x_i)}\right)^2} x_{i_j} \tag{6.2}$$

Each round, the regression coefficient vector $\vec{\beta}$ is then updated by

$$\beta_{k+1} = \beta_k - \mu \nabla J_k \tag{6.3}$$

where μ is a constant learning rate, and *k* is the step count in the optimization process.

With Gentle AdaBoost, an additive logistic regression model has been built to predict the likelihood that a customer will churn sometime in the future. Customers with a higher predicted likelihood have a higher propensity to churn. This likelihood can be stored in the customer database, in which each customer is assigned a probability score, denoted as $P_i(X)$, and is ready to apply to all kinds of retention offers. Practically, this predictive model allows a scientific basis for managing business development efforts and therefore optimises marketing costs. The results of the model will be used as a basis for generating lists and prioritising contact customers and offers.

An additive logistic regression model is defined as a customer churn profile model that focuses on answering 'who are they' and 'how do they behave', not only because of its high reported accuracy, but also because of its interpretability for understanding key drivers, which may provide information to set up retention actions. As a result, this model also provides the key factors that lead to customer churn. The top 10 key factors are:

x_1: Customer mobile contract expired

x_2: Monthly bill charged amount decreasing

x_3: Months till contract expire less than four months

x_4: Rate Plan Tenure more than 28 months

x_5: Not Fixed contracts, or month to month contracts

x_6: SMS decreasing dramatically, fewer weekend calls last month

x_7: Not a fixed line

x_8: Handset tenure more than 16 months

x_9: On-net call duration decreasing dramatically, off-net calls increasing significantly

x_{10}: Calls to customer care increasing

All these key factors can be used as customer risk profiles, to be implemented into the retention process to tackle key risk factors and prevent customers from leaving.

6.2.4 CUSTOMER ACTION MODEL

In Section 6.2.3, a customer churn profile model has been developed to provide a way to identify churners. The next question is exactly when will customers be likely to churn. In this section, a customer action model is developed, which resorts to survival analysis to test the impact of on-the-churn likelihood in terms of a period of time; in other words, to identify and predict when customers are likely to churn. Survival Analysis is appropriate to use for design, which involves a time-varying and right-censored data window. Figure 6-2 represents the lifecycle information of two fictitious customers.

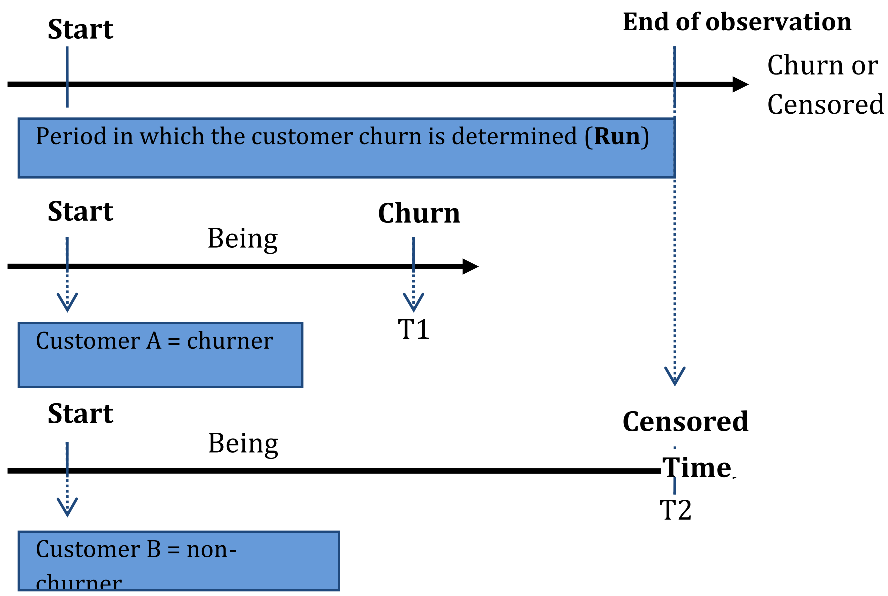


Figure 6-2 Example of Customer Lifecycle

The Proportional Hazard (PH) model introduced by Cox (Cox 1972) makes a separability assumption on the shape of the underlying hazard. In the PH model, the hazard rate for an observation $x_i = (x_{i_1}, x_{i_2}, \ldots, x_{i_n}) \in X$ at time t is given by:

$$h(t \mid x_i) = h_0(t)\lambda = h_0(t)\exp(\beta_1 x_{i_1} + \beta_2 x_{i_2} + \ldots + \beta_n x_{i_n}) \tag{6.4}$$

where $h_0(t)$ represents the baseline hazard function, which depends on t (but not x_i). The baseline hazard is assumed to be common to all observations. $\lambda = \exp(\beta x_i)$ represents a person-specific hazard function that scales the baseline hazard function. x_{i_j} represents the value of covariate j for an observation x_i. β_j represents the coefficient of covariate j.

The proportional hazard model has been employed because of its convenient advantages: first, it allows the incorporation of time varying covariates and both discrete and continuous measurements of event times; second, it can handle observations that did not experience the event (that is, censored observations). The survival function is estimated by calculating the Kaplan–Meier estimator, which is the most widely used method for estimating survival functions and is considered to be an important tool for analysing censored data (Larivière & Van den Poel 2004, 2005). Finally, the hazard rates and survival probabilities are presented as curves.

Figure 6-3 charts two examples, a churner and a non-churner, of hazard rates. The horizontal axis is the observation time t of customers measured in week; the vertical axis is the estimated hazard rates.

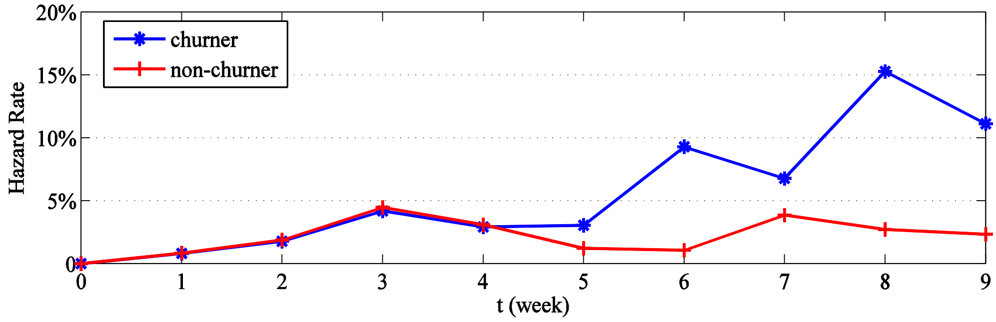


Figure 6-3 Example of Hazard Curves

Figure 6-4 shows the corresponding survival curves. The survival 'curve' is a step function with sudden changes in the estimated probability corresponding to times at which a customer churn may occur.

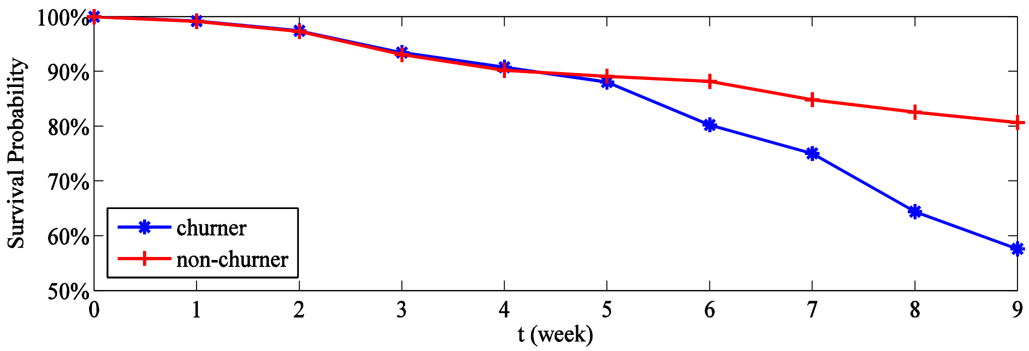


Figure 6-4 Example of Survival Curves

With this model, each customer is assigned a hazardous score $H_i(t)$ to estimate how quickly a customer is going to churn.

6.2.5 CUSTOMER EXPERIENCE MODEL

In Section 6.2.3 and Section 6.2.4, a customer profile model and a customer action model have been established to predict which customers are likely to churn and when. Another important question to answer is: what are the really bad experiences that customers have encountered with the organisation? In this section, Factor Analysis has been used to conduct analysis that is based purely on experience information, such as usage information, network information, customer service information and complaints, to identify and predict the reasons for customer churn.

Factor analysis has been performed on all 100+ experience variables, with the aim of grouping these 100 plus variables into a manageable size of variables, resulting in six factors: billing, network, mobile usage, customer service calls, pricing and handset. Each customer can be scored by these six factors based on factor loadings.

Another logistic regression has been performed to determine which factors have the most impact on customer churn. Unlike the customer churn predictive model, this model predicts the likelihood of the customer churn based only on the experience that customers have encountered with the telecom organisation. The model is defined as:

$$\text{logit}(\hat{p}) = -1.45 - 0.89 * F_1 + 0.01 * F_2 - 1.07 * F_3 - 0.35 * F_4 - 2.04 * F_5 - 0.45 * F_6$$

Again, this model is used to score each customer with a probability; that is, each customer is assigned a probability score that shows the degree of customer dissatisfaction, denoted as $P_i(F)$.

6.2.6 *AN INTELLIGENT CUSTOMER RISK MODEL*

Each customer has now been scored with three risk factors by the previous models: $P_i(X)$, representing the probability to churn for customer *i*; $H_i(t)$, representing how quickly customer *i* is going to churn; and $P_i(F)$, representing the degree of dissatisfaction from a customer experience point of view. The customer risk model will provide an overall customer churn risk assessment.

6.2.6.1 *Risk Integration*

The overall churn risk level of customers is aggregated using a weighted average of normalized scores of these three risk factors:

$$CR_i = \overline{w}_1 P_i'(X) + \overline{w}_2 H_i'(t) + \overline{w}_3 P_i'(F) \tag{6.5}$$

where $P_i'(X)$, $H_i'(t)$, $P_i'(F)$ are normalized risk factors by Equation (14), CR_i denotes the estimated churn risk for customer *i*, $\sum_{j=1}^{3} \overline{w}_j = 1$, $0 \le w_j \le 1$.

$$x' = \ln\left(\frac{x - \min(x)}{\max(x) - \min(x)}(e-1) + 1\right) \tag{6.6}$$

In implementation, at the beginning, a decision panel consisting of n invited experts, denoted by $E = \{E_1, E_2, \ldots, E_n\}$, is established. Then, for each expert E_i, a pair-wise comparison matrix representing the relative importance between risk factors is made. The system estimates the weight set $W^{E_i} = [w_1, w_2, w_3]$ for expert E_i by the Analytic Hierarchy Process (AHP) (Saaty, 1988), which has been widely applied in group decision making. The final weight set estimation is calculated by averaging all experts' weight sets $\overline{W} = \frac{1}{n}\sum_i W^{E_i}$.

6.2.6.2 *Risk Interpretation*

So far, each customer's churn risk can be evaluated by a crisp number between 0 and 1. The intuitive meaning of a crisp number, however, is limited. For instance, given a risk level of 0.5, different people may have different understandings. Hence, fuzzy logic has been employed to transform the crisp results into linguistic terms. The risk level can be defined by five categories: VL (very low risk), L (low risk), M (medium risk), H (high risk), and VH (very high risk). Each category can be viewed as a fuzzy set on [0,1]. Finally, a risk evaluation would be expressed as:

The assessment of churn risk for customer A is: VH - 0.2; H - 0.8.

Again, a decision panel with n invited experts is established, denoted by $E = \{E_1, E_2, \ldots, E_n\}$, to construct the fuzzy membership function for each fuzzy set by the following steps:

Step 1. For each category of risk level, all experts are requested to give their own most appropriate corresponding risk intervals on a scale of 0 to 1, with 1 being the highest risk. For any expert, a given interval side value for a risk category must be greater than any given interval side value of the same side for a lower risk category. Taking H and VH as an example, the intervals given by expert E_i would be $\lfloor H_l^{E_i}, H_r^{E_i} \rfloor$, $\lfloor VH_l^{E_i}, VH_r^{E_i} \rfloor$, where $0 \le H_l^{E_i} \le VH_l^{E_i} \le 1$, and $0 \le H_r^{E_i} \le VH_r^{E_i} \le 1$.

Step 2. Remove an expert's evaluation on a certain risk category if his left side value is greater than the medium value of all given right side values on the same risk category, or, if his right side value is smaller than the medium value of all given left sides values on the same risk category. Again, taking the *VH* as an example, suppose the medium value of the right side is $VH_r^{E_j}$, and the medium value of the left side is $VH_l^{E_k}$, then the evaluation on *VH* from expert E_i will be removed, if $VH_l^{E_i} > VH_r^{E_j}$, or $VH_r^{E_i} < VH_l^{E_k}$.

Step 3. For each risk category, calculate the average midpoint of all remaining evaluations. For example, if there are m remaining evaluations on *VH*, the average midpoint of *VH* is calculated by $VH_m = \frac{1}{m}\sum_i \left(\frac{VH_l^{E_i} + VH_r^{E_i}}{2} \right)$.

Step 4. Remove an expert's evaluation on a certain risk category if his left side value is greater than the average midpoint, or if his right side value is smaller than the average midpoint. For example, the evaluation on *VH* of expert E_i will be removed if $VH_l^{E_i} > VH_m$, or if $VH_r^{E_i} < VH_m$.

Step 5. The membership degrees at the endpoints of any interval given by the experts are determined by the proportion of remaining evaluations that contain this point. For example, at the left point of the estimation on *VH* given by expert E_j, its membership degree to VH is calculated by $\mu_{VH}\left(VH_l^{E_j}\right)=\frac{1}{k}\sum_i f_{VH}\left(VH_l^{E_j}\right)$, where k is the number of remaining evaluations on VH, $f_{VH}(x)=1$ if $VH_l^{E_i} \le x \le VH_r^{E_i}$, otherwise $f_{VH}(x)=0$.

6.2.6.3 An Illustration

Suppose 20 experts are invited to estimate the membership function of high risk (H), given the following table.

Table 6-1 Expert Evaluations on High Risk

Expert	H_l	H_r	Expert	H_l	H_r
1	0.7	0.9	11	0.95	1.0
2	0.6	0.75	12	0.8	0.9
3	0.7	0.8	13	0.7	0.9
4	0.65	0.9	14	0.7	0.85
5	0.7	0.95	15	0.65	0.85
6	0.75	0.9	16	0.75	0.95
7	0.75	0.9	17	0.7	0.9
8	0.7	0.9	18	0.7	0.9
9	0.5	0.65	19	0.6	0.9
10	0.75	0.95	20	0.75	0.9

The medium value of the left side is 0.7, and the medium value of the right side is 0.9. Experts 9 and 11 are removed, because $H_r^{E_9} < 0.7, H_l^{E_{11}} > 0.9.$ with the remaining experts, the medium value $H_m = 0.796$ because $H_r^{E_2} < H_m$, $H_l^{E_{12}} > H_m$, the evaluations of Experts 2 and 12 are discarded. In the final step of this example, a membership function of high risk is drawn in Figure 6-5.

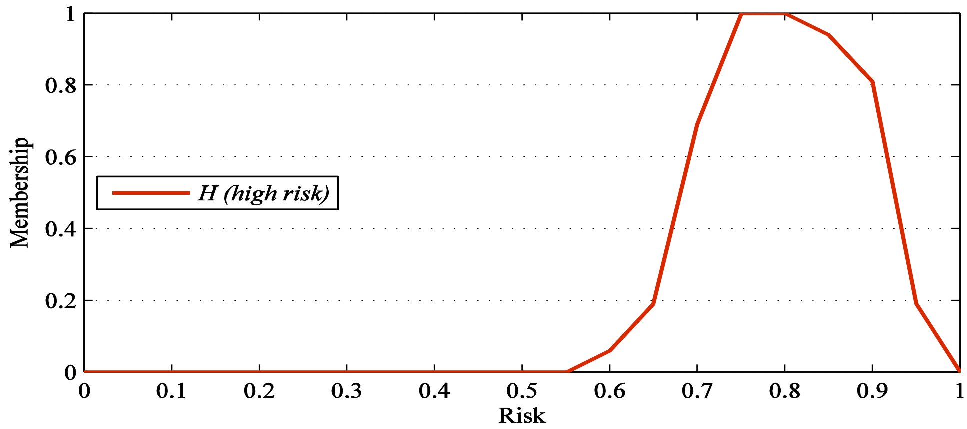
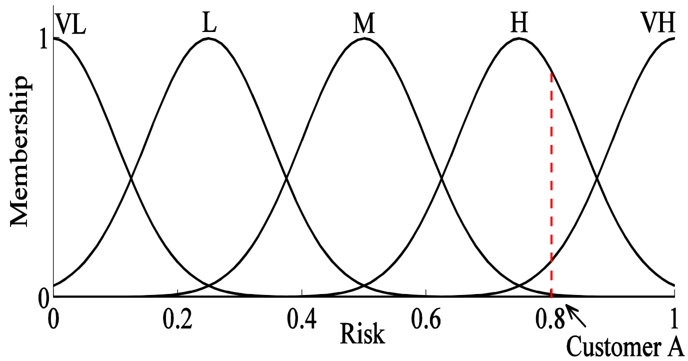


Figure 6-5 Membership Function of High Risk

After the membership functions of all risk categories have been estimated, given a crisp risk evaluation, the overall risk assessment of the customer can be determined, as shown in Figure 6-6.



Churn Risk Assessment

Customer:	A
Churn risk (CR_A):	0.8
Risk level evaluation:	H - 0.85 VH - 0.15

Figure 6-6 Example of Overall Risk Assessment

6.3 EXPERIMENTS AND ANALYSIS

To evaluate the proposed churn management model, a training sample of 7823 customers was selected randomly, including 949 churners, and 6874 non-churners. The model was validated with a test set of six-months' customer information, collected in 2011.

6.3.1 CHURN PREDICTION

A churn prediction model should be measured by its ability to identify churners for marketing purposes (Richter,Yom-Tov and Slonim 2010), thus the Receiver Operating Characteristic (ROC) curve is used to provide a comprehensive evaluation of the customer profile model. Specifically, the area under the ROC curve (AUC) is used as a performance measure, where an AUC close to 1.0 reveals that the model has perfect discrimination, while an AUC close to 0.5 suggests poor discrimination. Figure 6-7 shows the ROC curve of the customer churn predictive model, which provides an adequate discrimination on churn prediction, with an AUC of 0.7.

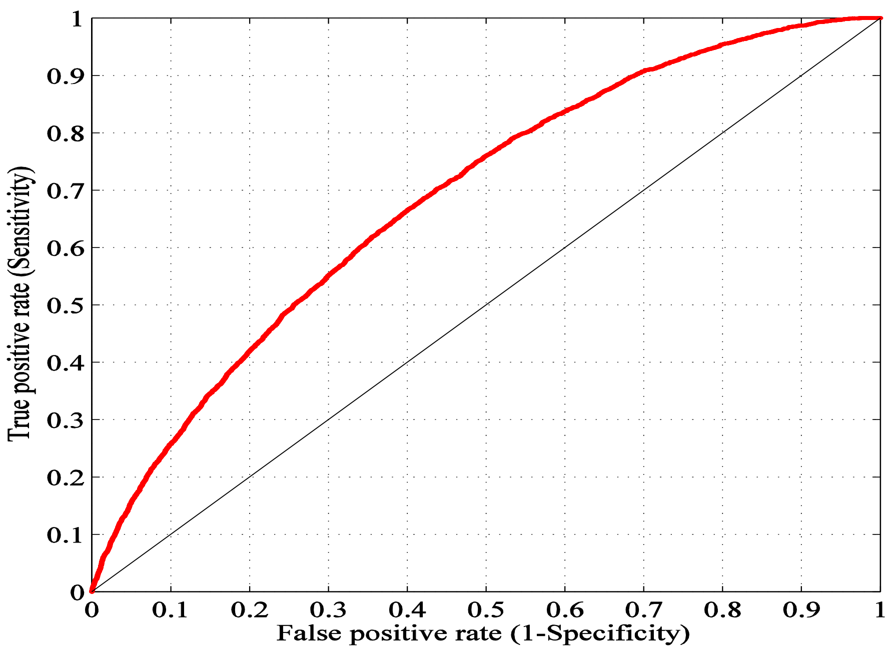


Figure 6-7 Roc Curve of Churn Prediction

The lift values are also used to evaluate the predictive model. The lift value reflects the increase in density with regard to the churn event relative to the density of churners in the customer base (Hadden et al. 2007). The higher the lift is, the better the predictive model is. For example, a top 5% percentile lift of 3 means that the model identifies three times more churners in the top 5% of the population than a random guess would achieve. For marketing purposes, where budgets are often limited and only a small fraction of customers can be targeted for retention actions, the top-quantile-lift is more important and of more practical value than other measurements. As shown in Figure 6-8, our churn predictive model achieves a lift score of 2.6 for the top 5% critical customers (i.e., customers that have the highest churn probability), and a score of 2.3 for the top 10% critical customers.

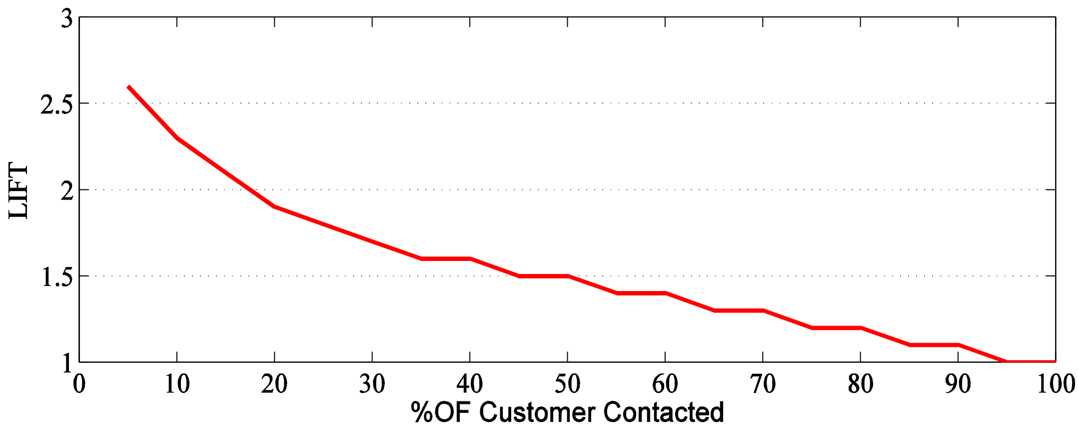


Figure 6-8 Lift Chart for Churn Prediction

The performance of the boosted regression model is evaluated by a training set of customer information that is collected based on a 6-month period. Each customer's churn propensity is monitored and updated according to his latest 3-months information. In this way, the real-world scenario of churn prediction can be simulated.

A churn prediction system should be measured by its ability to identify churners for marketing purposes (Huang, Kechadi & Buckley 2012), thus the Receiver Operating Characteristic (ROC) curve is used and top-quantile-lift values to give a comprehensive evaluation of our prediction model, and to compare the results with logistic regression without boosting. In Figure 6-9, the ROC curve of the boosted logistic regression model is located on the curve of the logistic regression model, with the Area Under Curve

(AUC) increased from 72.39 to 75.14. The improvement is reported to be significant ($z = 1.725$, $p = 0.042$) when $\alpha = 0.1$, which shows that the boosted logistic regression model is able to better distinguish churners from non-churners. The lift value reflects the increase in density with regard to the churn event relative to the density of churners in the customer base (Zhao & Yang 2011). The higher the lift is, the better the predictive model is. For marketing purposes, where budgets are often limited and only a small fraction of customers can be targeted for retention actions, the top-quartile-lift is more important and of more practical value. In Figure 6-10, the boosted logistic regression achieves better lift values than logistic regression, increasing the top-5%-lift from 2.6 to 3.4, and the top-10%-lift from 2.3 to 2.8.

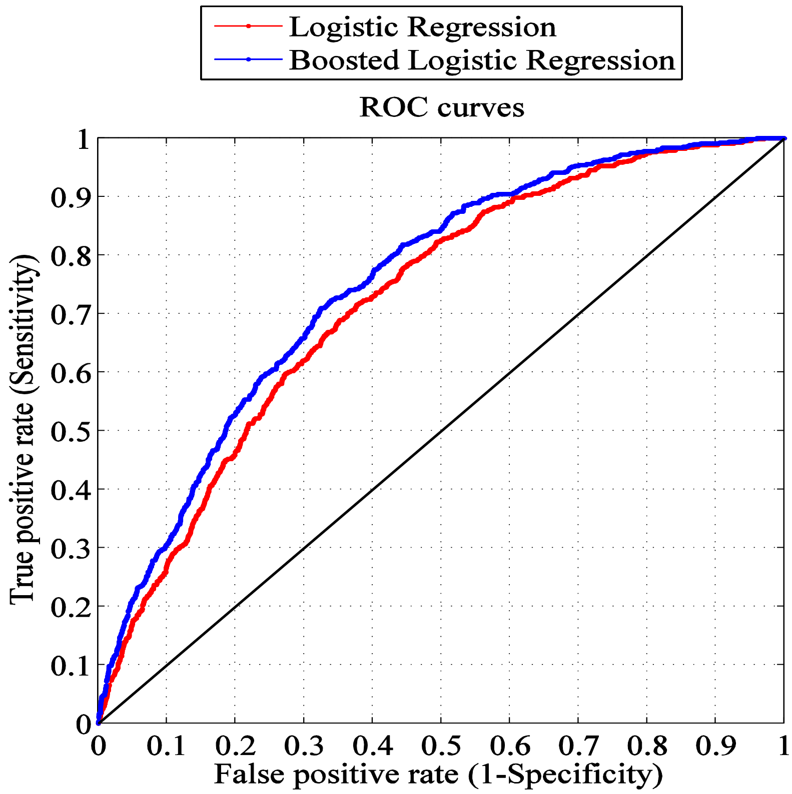


Figure 6-9 Roc Curves of Predictions

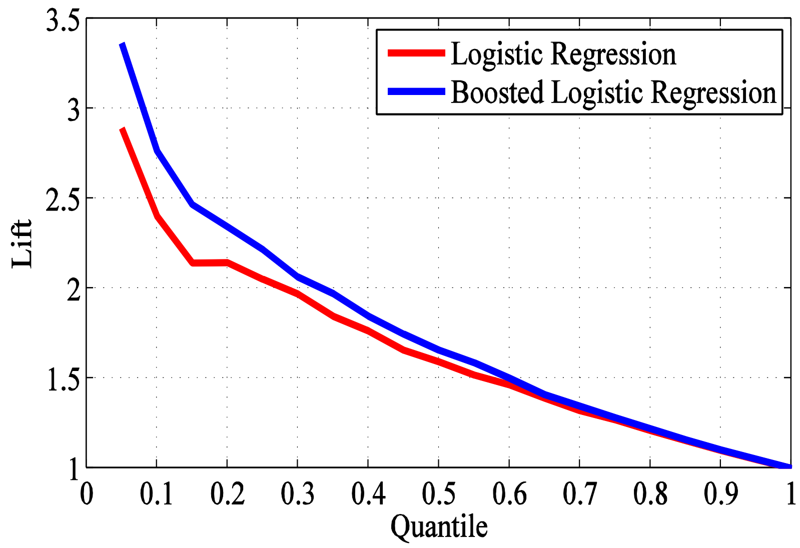


Figure 6-10 Lift Curves of Predictions

6.3.2 RISK ASSESSMENT

The Customer churn prediction model (in Section 6.3.1), Customer action model (Section 6.2.4) and Customer experience model (Section 6.2.5) each scores a customer with a risk factor. To provide a comprehensive insight on the issue of customer churn, in terms of likelihood to churn, eagerness to churn and overall degree of dissatisfaction with levels of service, all three risk factors are synthesised by consulting 20 experts to evaluate the relative importance of each risk factor. This results in the weight set shown in Table 6-2 which provides an example of 10 customer risk factors, as well as overall risk estimations. The five customers of each type (churner and non-churner) are intentionally picked, while the actual data is of a different distribution, as shown in this table.

Table 6-2 Example of Customer Risk Factors

Customer ID	Churner	$P_i(X)$	$H_i(t)$	$P_i(F)$	CR_i
1	No	0.06	0.07	0.07	0.06
2	No	0.06	0.19	0.24	0.06
3	No	0.28	0.44	0.35	0.34
4	No	0.87	0.01	0.02	0.48
5	No	0.91	0.01	0.01	0.49
6	Yes	0.09	0.42	0.37	0.24
7	Yes	0.46	0.27	0.21	0.36
8	Yes	0.60	0.42	0.43	0.52
9	Yes	0.87	0.09	0.12	0.52
10	Yes	0.58	0.79	0.79	0.68

Finally, to interpret a given risk estimation, the five fuzzy sets are built to represent the five different risk levels, with the help of the same 20 experts, as shown in Figure 6-11.

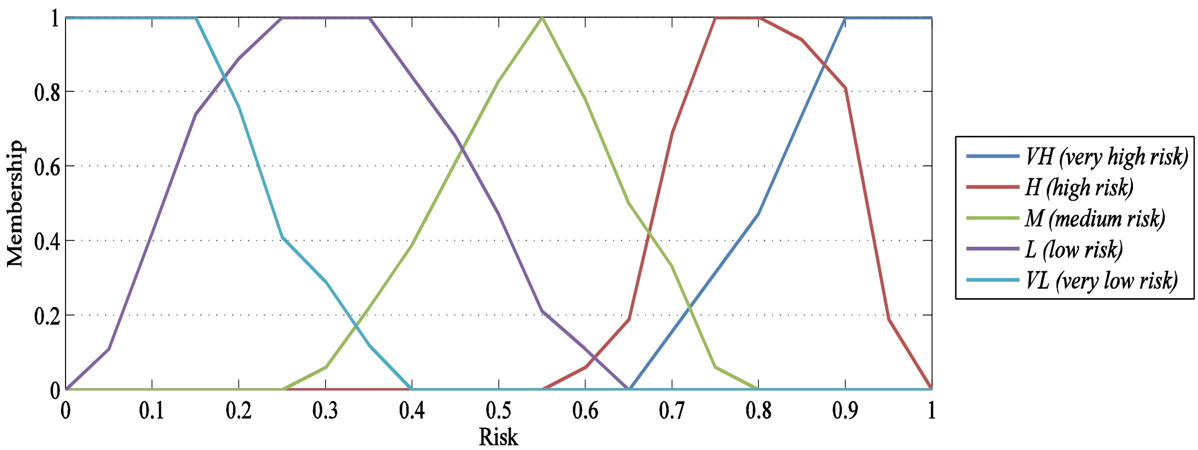


Figure 6-11 Membership Functions of All Risk Levels

Based on the membership degree of each risk level, customers can be segregated into five categories, as summarized in Figure 6-12. Therefore, different retention efforts can be customized according to each risk level.

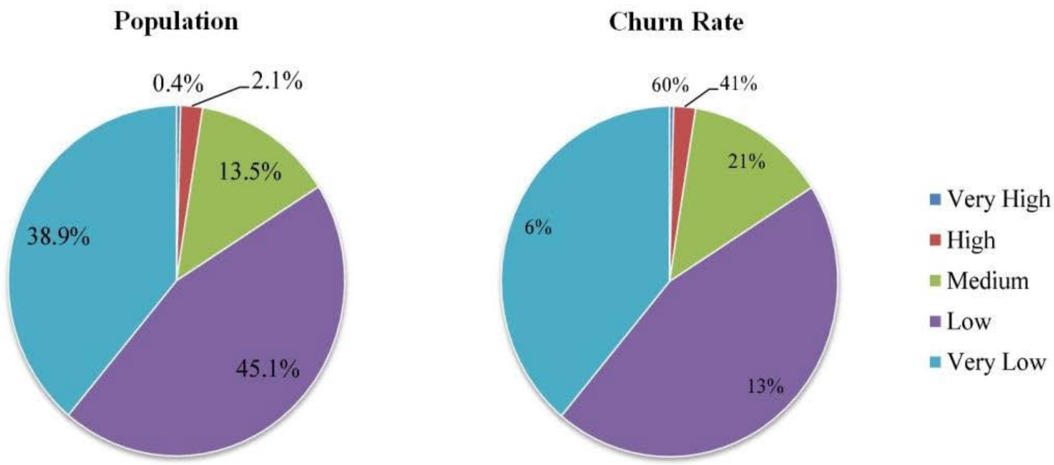


Figure 6-12 Summary of Customer Risk Categories

6.4 SUMMARY

This chapter directly targets the limitations of current studies on customer churn and proposes an intelligent customer churn management model, thereby affording a more comprehensive perspective on the issue of customer churn. Rather than only addressing the issue of customer churn prediction, the proposed customer churn management model utilizes multiple statistical models and is devoted to obtaining detailed information by posing the questions: 'when are they likely to churn', 'why do they churn', and 'how do they behave', to facilitate possible retention efforts. In addition, this chapter also proposed a novel customer risk management process powered by fuzzy logic, which immensely enhances the perception of churn management. Customers are prioritized based on their risk levels with customized retention efforts.

Our risk management model is limited to segmented customers. Future studies might take advantage of this information to compare customers in different risk categories and reveal key determinates of churn for each category. Moreover, applying fuzzy logic in our churn predictive model, customer action model and customer experience model would be another promising field.

CHAPTER 7

A FUZZY SET-BASED HYBRID RECOMMENDATION APPROACH FOR CUSTOMER RETENTION

7.1 OVERVIEW

In Chapter 6, an advanced customer churn management model was developed, based on the telecom industry case. This chapter aims to build a Fuzzy Set-Based Hybrid Recommendation approach that also uses telecom as a case study, but which explores a new area of recommender systems and telecom business intelligence. This proposed approach integrates Item-based Collaborative Filtering (IBCF) and User-based Collaborative Filtering (UBCF) with fuzzy set techniques and a knowledge-based method (business rules). It first uses IBCF to produce predictions to form a dense user-item rating matrix, and based on this matrix, UBCF is applied to generate recommendations. The approach takes advantage of both the horizontal and vertical information in the user-item rating matrix, which can solve the sparsity problem. It also uses fuzzy techniques to tackle linguistic variables which are used in describing customer preference and has the ability to generate recommendations using uncertain information.

This approach can be applied to purposely guide customers to select the most appropriate telecom products/services. As the topic of customer retention is at the top of the telecom organisation's strategic agenda, providing relevant products and offering appropriate services play important roles in retaining customers and maximising profit. This is an important part of customer relationship management and business intelligence across the industry. This proposed approach automatically predicts the behaviour and requirements of existing customers based on their profiles and business knowledge, which will be implemented in a proposed personalised recommender system as a software prototype for telecom products/services recommendation in Chapter 9.

This chapter is organised as follows. The recommendation approach is described in Section 7.2, using a telecom mobile customer as a case study. Based on the suggested recommendation approach, a proposed framework of the recommender system is presented in Section 7.3 and related experiments are shown in Section 7.4. Finally, a summary is given in Section 7.5.

7.2 Case Description For Existing Telecom Customers

Telecom businesses today offer hundreds of different mobile products and services such as handsets, mobile plans, prepaid mobiles, and broadband to customers and are constantly exploring new service models that will support customers in their selection and purchase of products and services on the Internet. Telecom products are always linked with services, referred to hereafter as 'products/services', and have very complex structures and a huge number of choices. A telecom organisation may have more than 500 telecom products/services in several categories for different groups of customers (personal customers and business customers). With such a vast number of choices, it is becoming increasingly difficult for customers to find their favourite products quickly and accurately. Only experienced salespeople with full product knowledge can make suitable personalised recommendations to customers, which is costly and inefficient. To help customers shop online, it is necessary to develop web-based intelligent information technologies that will fully use salespeople's knowledge to help customers select suitable products or services online.

There are three main difficulties in telecom product/service recommendation compared to other industries;

Firstly, telecom products/services have very complex descriptions and features. A complete mobile product/service for a customer includes handsets and related mobile services. A mobile service is a specification of the available sub-services and related prices, discounts and rewards. It is represented by a set of attributes such as the monthly access fee, call rate, data charging, rewards, and so on. A mobile service is often combined with a wireless Internet or a cable Internet service, or other applications. Table 7-1 shows a set of mobile products/services, illustrating the complexity of telecom products/services.

Secondly, mobile services and handsets are two separate features with different price structures, and a mobile plan may have both features or a single feature which comes with different price offers. These plan offers are updated frequently, but a mobile customer only has one product at a time. These factors result in a lack of product rating information from customers, which creates difficulties for comparing telecom products/services and generating recommendations.

Thirdly, telecom products change frequently, but some new products and old products have similarities. Also, telecom customers often express their preferences and interests in online product/service evaluations using linguistic terms, such as 'good', 'very good', and 'interested'.

Table 7-1 Examples of Mobile Products/Services

Product/Service Name	Telecom Product/Service Description
X-Smart $70 Data100MB	$70 included value; 100MB Data Unlimited access within Australia to Facebook, Twitter, LinkedIn, MySpace, eBay and Foursquare within Australia
A-Smart Data 1.5GB	$55 included value; 1.5GB Data Unlimited access within Australia to Facebook, Twitter, LinkedIn, MySpace, eBay and Foursquare within Australia
Smart Data 2GB 24M	$75 included value[1]; 2GB Data Unlimited standard SMS to Australian GSM mobiles; Unlimited access within Australia to Facebook, Twitter, LinkedIn, MySpace, eBay and Foursquare within Australia
Smart SMS/MMS5GB	Unlimited included value; 5GB Data Unlimited standard SMS and MMS to Australian GSM mobiles (excl. Pivotel); Unlimited access within Australia to Facebook, Twitter, LinkedIn, MySpace, eBay and Foursquare within Australia
Smart SMS/MMS6GB	Unlimited included value; 6GB Data Unlimited standard SMS and MMS to Australian GSM mobiles (excl. Pivotel); Unlimited access within Australia to Facebook, Twitter, LinkedIn, MySpace, eBay and Foursquare within Australia
X'Data $19.99	……

To deal with the above difficulties and help a customer to choose the most appropriate telecom products/services, this chapter considers both customer similarity and product similarity in recommendation generation. Because the similarity between products/services or between customers is naturally uncertain, fuzzy set theory lends itself well to handling the fuzziness and uncertain issues in recommendation problems (Cornelis et al. 2007). More importantly, fuzzy set techniques can be applied to tackle linguistic variables, which are used in describing customer preference and have the ability to support recommendation generation using uncertain information.

7.3 APPROACH DESCRIPTION

7.3.1 BUSINESS RULES

There are complex business rules in real world practice which directly impact on recommendation accuracy and customer acceptance of recommendations. This chapter designs and applies five types of business rules: 1) bundle rules, 2) multiple products rules, 3) discount rules, 4) product rules and 5) special offers. Three examples of business rules are:

"Some fixed line products cannot be purchased standalone. They have to be bundled with a fixed broadband product."

"A customer can receive additional discounts for some products, if they are purchased together."

"For a period of time, some products may be on special or the business may be promoting those products".

Business rules are made and maintained by product or sales managers in the organisation, and are taken as an input of the approach. The structure to describe a business rule is in the form of an 'if-then-else' statement. For example, *a* and *b* are similar products; *v* and *w* are similar products in function and price. A customer needs *a* (or *b*) and *v* (or *w*). Since purchasing *a+w* attracts additional discount, the system may recommend it.

7.3.2 A FUZZY SET-BASED HYBRID RECOMMENDATION APPROACH

The proposed approach is described in eight steps as follows.

Input:

n - the number of products (could be services or product bundles) provided.

m - the number of customers (existing customers) in the system.

$r_{i,j}$ - a rating of a customer for a product which could be described in linguistic terms. (i=*1,2,…m*; j=*1,2,..n*).

Business rules - described as if-then rules and stored in a knowledge base.

Output:

p_1, p_2,…p_k - k most appropriate products recommended by this system.

Step 1: Generate a user-item linguistic term-based rating matrix

Each customer is represented by a set of item-rating pairs and the summary of all those pairs can be collected into a user-item rating matrix in which a rating, $r_{i,j}$ is given for the i^{th} customer on the j^{th} item. These ratings are described in the linguistic terms shown in Table 7-2. There are m customers in total in the system and n products are provided. If customer i has not rated item j, then $r_{i,j} = N/A$.

Step 2: Calculate fuzzy item similarity

There are several methods of calculating the similarity between products, among which the Pearson correlation and cosine vector are two popular methods that are applied widely across the field. The cosine vector measures the similarity between two products by calculating the cosine of the angle between the two vectors of the target item x and the comparison item y (Adomavicius & Kwon 2012; Adomavicius & Tuzhilin 2005; Adomavicius & Zhang 2012). The Pearson correlation measures the similarity between two items by calculating the linear correlation between the two vectors (Shi, Ye &

Gong 2008). In this chapter, the Pearson correlation is selected for measuring the similarities between the two items x and y. Since the similarity between two items (telecom products) is naturally uncertain, the ratings collected are linguistic terms, and fuzzy numbers are used in the measure. Therefore the following fuzzy similarity measure is based on definitions given in Section 3.1.5:

$$sim(x, y) = \frac{\sum_{s \in S_{x,y}} \int_0^1 \frac{1}{2}\left[\left(r_{x,s_\lambda}^- - \tilde{r}_{x\lambda}^-\right)\left(r_{y,s_\lambda}^- - \tilde{r}_{y,\lambda}^-\right) + \left(r_{x,s_\lambda}^+ - \tilde{r}_{x\lambda}^+\right)\left(r_{y,s_\lambda}^+ - \tilde{r}_{y,\lambda}^+\right)\right] d\lambda}{\sqrt{\sum_{s \in S_{x,y}} \left(\int_0^1 \frac{1}{2}\left[\left(r_{x,s_\lambda}^- - \tilde{r}_{x\lambda}^-\right) + \left(r_{x,s_\lambda}^+ - \tilde{r}_{x\lambda}^+\right)\right] d\lambda\right)^2} \times \sqrt{\sum_{s \in S_{x,y}} \left(\int_0^1 \frac{1}{2}\left[\left(r_{y,s_\lambda}^- - \tilde{r}_{y,\lambda}^-\right) + \left(r_{y,s_\lambda}^+ - \tilde{r}_{y,\lambda}^+\right)\right] d\lambda\right)^2}} \quad (7.1)$$

where$S_{x,y}$ represents the set of customers that both rated items x and y. r_{x,s_λ} and r_{y,s_λ} represent the ratings of customer s on items x and y under λ -cut respectively, r_{x,s_λ}^- and r_{x,s_λ}^+ are the left-end and right-end of λ-cut respectively, $\bar{r}_{x_\lambda}$ and $\bar{r}_{y_\lambda}$ are the average rating of the customers of $S_{x,y}$ on x and y respectively. This step aims to obtain similarity between items (products).

Step 3: Selection of item neighbours

In most CF methods, a number of neighbours will be selected as references when predicting ratings (Zhang & Lu 2003). According to Shi et al. (2008), two approaches are possible for this task: threshold-based selection or top-*N* techniques. In our approach, the top-*N* technique is used for neighbour selection. By using this method, a certain number of most similar items will be selected as neighbours. The number of neighbours is predetermined before the item neighbour selection process.

Step 4: Predict empty fuzzy ratings using item-based CF with fuzzy number calculation

In this step, all the empty ratings can be calculated using the item-based CF method and all the empty cells in the user-item rating table will be filled except the ratings to the new items which have been rated less than twice. The algorithm for prediction is as follows:

$$F_{p_{x,s}} = \frac{\sum_{y=1}^{c} \tilde{r}_{y,s} \times sim(x,y)}{\sum_{y=1}^{c} sim(x,y)}$$

$$= \bigcup_{\lambda\in[0,1]} \lambda \left[\frac{\sum_{y=1}^{c} r_{y,s_\lambda}^{-} \times sim(x,y)}{\sum_{y=1}^{c} sim(x,y)}, \frac{\sum_{y=1}^{c} r_{y,s_\lambda}^{+} \times sim(x,y)}{\sum_{y=1}^{c} sim(x,y)}\right] \tag{7.2}$$

where $Fp_{x,s}$ refers to the predicted rating of customer s on item x, c is the number of selected neighbours, $\tilde{r}_{y,s}$ is the rating of customer s on item y, and $sim(x,y)$ is the similarity between item x and item y. This step aims to predict customers' (users') rating values to unrated items.

Step 5: Calculate fuzzy user similarity

Besides predicting the ratings based on the similarities of items, the ratings can also be predicted by analysing the similarities between customers. Since the similarity between two customers (users) is also naturally uncertain, fuzzy numbers are used in the similarity measure, similar to Step 2. The Pearson correlation algorithm is used for calculating the customer similarity by

$$sim(s,t) =$$

$$\frac{\sum_{s\in I_{s,t}} \int_0^1 \frac{1}{2}\left[\left(r_{x,s_\lambda}^{-} - \tilde{r}_{x\lambda}^{-}\right)\left(r_{x,t_\lambda}^{-} - \tilde{r}_{s,\lambda}^{-}\right) + \left(r_{x,s_\lambda}^{+} - \tilde{r}_{s\lambda}^{+}\right)\left(r_{x,t_\lambda}^{+} - \tilde{r}_{s\lambda}^{+}\right)\right] d\lambda}{\sqrt{\sum_{s\in I_{s,t}} \left(\int_0^1 \frac{1}{2}\left[\left(r_{x,s_\lambda}^{-} - \tilde{r}_{s\lambda}^{-}\right) + \left(r_{x,s_\lambda}^{+} - \tilde{r}_{s\lambda}^{+}\right)\right] d\lambda\right)^2} \times \sqrt{\sum_{s\in I_{s,t}} \left(\int_0^1 \frac{1}{2}\left[\left(r_{s,t_\lambda}^{-} - \tilde{r}_{s\lambda}^{-}\right) + \left(r_{s,t_\lambda}^{+} - \tilde{r}_{s\lambda}^{+}\right)\right] d\lambda\right)^2}} \tag{7.3}$$

where $sim(s,t)$ is the similarity between customer s and customer t, $I_{s,t}$ is the set of items that are rated by both customer s and customer t, $\tilde{r}_{x,s}$ is the rating of item x from customer s, $\tilde{r}_{x,t}$ is the rating of item x from customer t, $\bar{r}_s$ is the average of all ratings from customer s, $\bar{r}_t$ is the average of all ratings from customer t. This step aims to obtain similarity between customers to help predict customers' ratings of items.

Step 6: Select top-N similar customers

Similar to Step 3, a number of neighbour customers need to be selected to predict ratings. The Top-*N* technique is used in the proposed approach.

Step 7: Recommendation generation with fuzzy number calculation

This step predicts the ratings of every unrated telecom product/service for target customers using user-based CF. The new predicted ratings will replace the ratings predicted in Step 4 and will be regarded as the final results. The applied algorithm is as follows:

$$\begin{aligned} r_{x,s} &= \bar{r}_s + \frac{\sum_{t=1}^{c}(\bar{r}_{x,t}-\bar{r}_t)\times sim(s,t)}{\sum_{t=1}^{c} sim(s,t)} \\ &= \bigcup_{\lambda\in[0,1]} \lambda \left[\bar{r}_\lambda^- + \frac{\sum_{y=1}^{c}(r_{x,t_\lambda}^- - r_{t\lambda}^-)\times sim(x,y)}{\sum_{y=1}^{c} sim(x,y)}, \bar{r}_\lambda^+ + \frac{\sum_{y=1}^{c}(r_{x,t_\lambda}^+ - r_{t\lambda}^+)\times sim(x,y)}{\sum_{y=1}^{c} sim(x,y)}\right] \end{aligned} \tag{7.4}$$

where $r_{x,s}$ is the final predicted rating of item x from customer s, $\bar{r}_s$ is the average of all ratings from customer s, $\bar{r}_t$ is the average of all ratings from customer t, c is the number of neighbours selected in Step 6, $r_{x,t}$ is the rating of item x from customer t, and $sim(s,t)$ is the similarity between customer s and customer t.

Step 8: Final recommendation

The unrated products for the target customer are ranked according to the predicted ratings calculated in Step 7. The top-*K* products (which could be services or product bundles) are selected. Each product is checked to see whether it satisfies the related business rules. For example, if it is proposed to recommend a special fixed line product that has to be bundled with a fixed broadband product, the step will check whether it is bundled.

When all related rules are checked and satisfied, the top *K* products will be recommended directly; otherwise, the product will be revised accordingly. Finally, a set of most suitable products/services/bundles, p_1, p_2,…p_K, is recommended to the target customer.

7.4 A RECOMMENDER SYSTEM FRAMEWORK FOR EXISTING CUSTOMERS

The proposed recommendation approach in Section 7.2 can be applied to produce a personalised recommendation or offer of relevant products and services to existing customers via the right channel at right time. It also can be implemented across an organisation's support and sales channels to introduce the right products to their customers in real time and, therefore, optimise every customer touch-point more effectively. Figure 7-1 shows a framework of the recommender system, by implementing the proposed approach. This framework contains three components: data builder, recommendation engine and interfaces. The data builder contains three pre-developed databases;

a) All products and services database:

The products and services database contains detailed information of all products and customer service information such as, sales calls, care calls, customer inquiry calls and marketing calls.

b) Customer profile, usage, billing, product holding and rating database:

This database contains personal customer information, such as age, address and demographic information. It also includes customer usage, billing, and product holding, as well as product ratings by customers.

c) Business rules database:

This database is designed to store all the business definitions and business rules.

The recommender engine generates recommendations based on the proposed Fuzzy Set-Based Hybrid Recommendation approach. Interfaces connect the customers' request, products holdings, existing customer information, and so on.

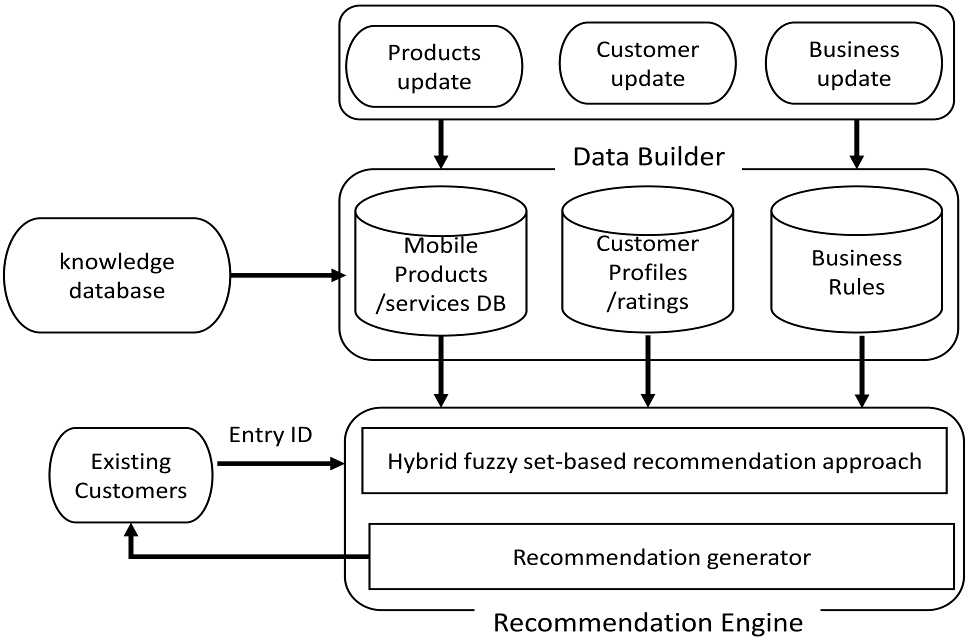


Figure 7-1 A Proposed Recommender System for Existing Customers

This proposed approach can be implemented in an online system to retain existing online customers by providing suitable recommendations based on their needs.

7.5 EXPERIMENTS AND ANALYSIS

Before implementing the proposed approach in an online system, and applying it to a real world telecom application, a set of experiments was conducted using the MovieLens 100K dataset accessible from the GroupLens Research website (http://www.grouplens.org/node/73) to test the prediction accuracy of the proposed hybrid recommendation approach.

7.5.1 DATASET

This dataset is collected by asking new customers to register with the website and rate at least 15 movies that they have watched. The website will then recommend numerous movies to new customers based on their ratings. In this experiment, the MovieLens 100K dataset contains 100,000 rating records for 1682 items from 943 customers, where the rating scale is from 1 to 5, every customer has rated at least 20 movies, and all movies have been rated at least once. To validate the proposed fuzzy set-based recommendation approach in this case, a set of five linguistic terms {Strongly Interested (SI), More Interested (MI), Interested (I), Less Interested (LI), Not Interested (NI)} is used to describe the customer ratings. The related fuzzy numbers to these linguistic terms are shown in Table 7-2. Their membership functions are illustrated in Figure 7-2, and the original rating scale 1 to 5 is fuzzified into NI, LI, I, MI, and SI respectively according to Table 7-2.

Table 7-2 Linguistic Terms and Related Fuzzy Numbers

Linguistic terms	Triangular Fuzzy numbers
Strongly Interested (SI)	(4,5,5)
More Interested (MI)	(3,4,5)
Interested (IN)	(2,3,4)
Less Interested (LI)	(1,2,3)
Not Interested (NI)	(1,1,2)
N/A	-

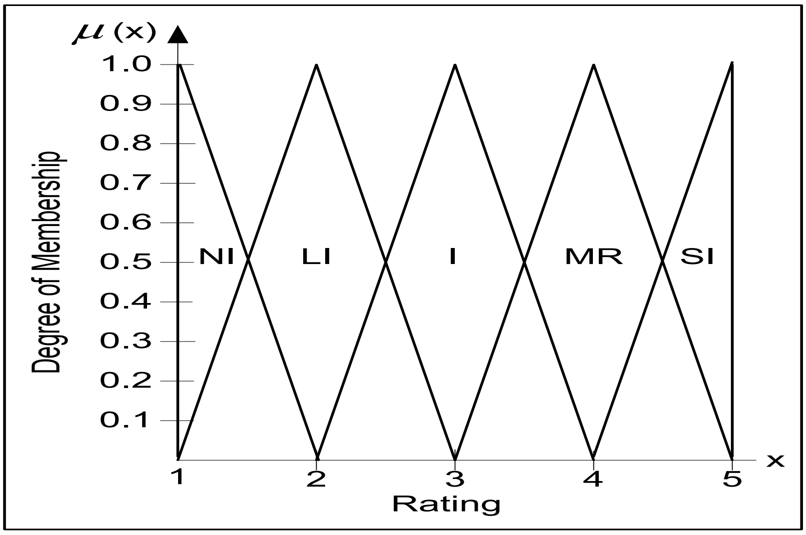


Figure 7-2 Fuzzy Sets and Membership Functions for Table 7.2

7.5.2 EVALUATION METRICS

Two popular methods for measuring recommender systems are statistical accuracy metrics and decision support accuracy metrics. Statistical accuracy metrics compare the predicted ratings with the user-rated ratings. Commonly used statistical accuracy metrics methods include Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Correlation. In the experiments, MAE is selected as the evaluation method because it is easy to interpret directly and is more commonly used than others:

$$MAE = \frac{\sum_{i=1}^{N} d(p_i, r_i)}{N} \tag{7.5}$$

where p_i is the predicted rating value of an item from a particular customer i, r_i is the actual rating of theitem from this customer (i=1,2,…N), $d(.)$ is the distance measure between two fuzzy numbers which is calculated by formula (3.18), and N is the total number of rating pairs compared.

7.5.3 EXPERIMENTAL ANALYSIS

To achieve accurate evaluation results, the training dataset and testing dataset have been randomly selected five times so that five training/testing dataset groups have been formed. For each group, the dataset is divided into one training dataset that contains 80% of all customer ratings and one testing dataset that contains the remaining 20% ratings. For each group, the training dataset is used as the input data for the approach and all unrated ratings are to be predicted; MAE is then applied to compare all the records in the testing dataset with the predicted ratings.

The five training datasets are named u1base, u2base, u3base, u4base and u5base, while the five corresponding testing datasets are named u1test, u2test, u3test, u4test and u5test. To measure the effect of the number of neighbours on the accuracy of the approach, the MAE is calculated four times separately using 5, 10, 20 and 50 neighbours for each training/testing group. The testing result is illustrated in Figure 7-3.

As shown in Figure 7-3, only group 1 has a slightly higher average MAE than the rest of the groups, and the results of the other four groups are all very close to one another. Therefore, the performance of the approach is quite uniform across the MovieLens dataset. Figure 7-3 clearly shows that the average MAE falls, while the number of neighbours increases. As the number of neighbours increases from 5 to 10, there is a significant drop in average MAE which indicates a considerable increase in prediction accuracy. Therefore, considering both accuracy and calculation efficiency, it has been decided that 10 neighbours are most suitable for the system.

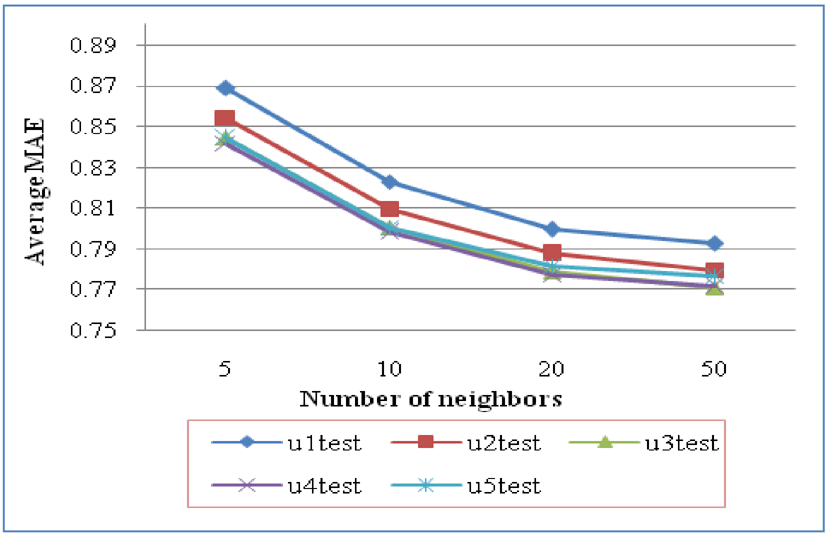


Figure 7-3 Experiment Results (MAE)

Table 7-3 Comparison with Six other Hybrid Collaborative Filtering Approaches

Missing rate %	Pearson CF	Model-Based CF	Content predictor	CBCF	JMCF	SMCF	FTCP-RM
94.64	0.8937	0.8921	0.9178	0.8705	0.8135	**0.7785**	
94.96							**0.784929**
95.59	0.8858	0.8437	0.8669	0.8014	**0.7836**	**0.7335**	

This experiment also compared the results with the other six recommendation approaches given in (Su et al. 2007): Pearson CF (memory-based CF), model-based CF, content-based predictor, combination of CB and CF, Joint mixture CF (JMCF), sequential mixture CF (SMCF), following the performance measurement used, which is MAE. The missing rate of the datasets is calculated as ((1682*943)-80000)/ (1682*943) = 0.9496 = 94.96%. This sparsity rate is between the two rates 94.64% and 95.59% given in (Su et al. 2007); see details in Table 7-3. When using 20 neighbours in the approach, the MAE is 0.784929. This is higher than the MAE of SMCF, very close to that of JMCF, a little lower than that of CBCF, and significantly lower than those of the other three methods. Based on this comparison, it is proven that the accuracy of the approach is competitive with these hybrid recommendation algorithms, and markedly higher than traditional CB and CF recommendation methods. Furthermore, none of the approaches in (Su et al. 2007) can handle uncertainties in customer ranking data, whereas the proposed Recommender System for Existing Customers in this chapter has the ability to deal with the linguistic ratings with fuzzy techniques. Proposed Recommender System for Existing Customers is more suitable for use in telecom product/service recommendation where uncertainty issues exist naturally and business rules are considered.

7.6 SUMMARY

This chapter proposes a hybrid fuzzy set-based recommendation approach which combines user-based and item-based collaborative filtering techniques with fuzzy set techniques and knowledge base for telecom product and service recommendation to customers who already have a relationship with the organisation. It particularly implements the approach in a personalised recommender system proposed in Chapter 9 for telecom products/services called FTGP-RS.

As mentioned in Section 7.1, telecom companies have two groups of customers: personal customers and businesses. This case study only focuses on personal customers. In the future, the recommendation approach and software system will be improved and adapted to develop a mobile product/service recommender system to support business customers. In that situation, a customer (business) may have multiple handsets with different plans, multiple services including fixed-line, SMS, GSM mobiles, access to

Facebook, Twitter, and much more. The similarity between two customers becomes very difficult and has high uncertainty. A new tree-structure fuzzy measure approach needs to be developed and used in a new recommendation approach.

CHAPTER 8

A FUZZY MATCHING-BASED RECOMMENDATION APPROACH FOR CUSTOMER ACQUISITION

8.1 OVERVIEW

In Chapter 7, a Fuzzy Set-Based Hybrid Recommendation Approach is developed. The approach can be used to provide relevant and suitable products or services for existing customers. The case study has shown that there are hundreds of different kinds of mobile products/services (rate plans) with complex features in the telecom industry, and the challenge of selecting an appropriate mobile product/service for a prospective customer (i.e., a person who is not yet a customer for a particular telecom organisation yet, also called new customer) is the same as that for an existing customer.

To achieve this goal, a Fuzzy Matching-Based Recommendation Approach to support new customers in selecting the most appropriate mobile products/services is presented in this chapter. In this approach, a set of characteristic sub-services is abstracted for each mobile product/service. Based on the characteristic sub-services, the requirements of customers and the characteristics of mobile products/services are modelled as linguistic vectors. A fuzzy matching approach is applied to find the mobile products/services that are best matched to customers' requirements. A customer who is not yet a telecom customer (called a prospective customer, or 'prospect') is employed as a case study.

The rest of this chapter is organised as follows. Section 8.2 presents the case description of the customers' requirements models and the characteristics of mobile products/services, and proposes a fuzzy matching-based recommendation approach for new customers. Based on the proposed recommendation approach, a framework of the recommender system for mobile products/services is presented in Section 8.3. Section 8.4 presents an example to illustrate the system. Finally, a summary and discussion are presented in Section 8.5.

8.2 APPROACH DESCRIPTION

8.2.1 CASE DESCRIPTION FOR PROSPECTIVE TELECOM CUSTOMER

As mentioned in Chapter 7, a recommender system is one of the most popular personalisation applications. This type of system filters out uninteresting items or predicts interesting items automatically on behalf of customers, according to their personal preferences, and can thus be used to guide customers in making the right choice. A personalised recommender system can keep track of a customer's every activity and performance at different contact points. It gathers individualised information over time and builds a rich consumer profile. This profile, combined with other customer activity, will form the basis of the recommender system. Telecom mobile products and services are considered for the case study in this chapter.

It was pointed out in Chapter 7 that there are three main challenges for both a sales-representative and a new customer when choosing telecom products/services. This chapter only focuses on mobile prodcuts that have three special features that are slightly different from other products:

(1) A customer faces a number of very complex descriptions of mobile products/services. An example of four mobile services rate plans of a telecom organisation is given in Table 8-1. It shows that a mobile service is a specification of the available sub-services and related prices, discounts and rewards. It is represented by a set of attributes, such as monthly access fee, call rate, data charging, rewards, and so on. Different mobile products/services are described by different attribute sets;

(2) Mobile products/services are updated frequently. New products are introduced very quickly, while some old ones are discontinued;

(3) Customers do not change their mobile services very frequently. There is insufficient rating information on products from customers.

In spite of the above, there is usually rich information on existing customers' profiles and usage records. To design an effective recommendation approach for mobile products/services, all these features should be considered.

Table 8-1 Four Mobile Service Plan Features

	$49 Cap Plan	$49 Rate Plan	Cap Plus 500MB	$129 Timeless Max
Monthly access fee	$49	$49	$79	$129
Total minimum cost	$1176	$1176	$1896	$3096
Standard National call rate	80c/minute	25c/30secs		
Standard National call Connection fee	35c	20c		
National Video call	$1/minute	50c/30secs		
Video call Connection fee	35c	35c		
International Video call		75c/30sec		
Standard National SMS	25c	25c		
Unlimited Standard National SMS Offer			Unlimited	Unlimited
International SMS		50c		
Standard National MMS	25c	50c		
Unlimited Standard National MMS Offer			Unlimited	Unlimited
International MMS		75c		
Call2Anyone Value Offer	$330		$330	
On-net Call Value Offer	$350		$350	

Unlimited standard local, National Voice calls Unlimited National Video calls Offer				Unlimited
Voicemail	Unlimited		Unlimited	Unlimited
Mobile Data Inclusions			Up to 500MB	4GB Data or Unlimited BlackBerry email and browsing
Excess Data usage			15c/MB	35c/MB
International Value Offer				$100
Voicemail		Optional		
Shareplan		Included		

The features of mobile products/services are analysed and a set of characteristic sub-services are abstracted in this chapter. The attributes of each mobile product/service are mapped to the suitable usage amount of the characteristic sub-services. The existing customer usage records are analysed and also mapped to the usage amount of the characteristic sub-services. As the characteristics of mobile products/services are reflected by both their attributes and their customers' usage records, both aspects should be considered. The requirements of prospective (or new) customers are modelled as the needs of the characteristic sub-services and are captured by asking a set of questions. In real situations, a customer's requirements are hard to present with certainty and accuracy, and thus are often expressed by linguistic terms, such as 'highly required', 'do not need it'. Therefore, the requirements of new customers or prospects are represented as linguistic terms in our approach. As seen from Table 8-1, some attributes of products are described in linguistic terms. The usage amount of each product is normally represented as a linguistic term in a customer's requirements, such as, 'high', 'medium high', 'medium', 'medium low' and 'low', because customers are unable to provide an exact amount of usage information. In such cases, precise mathematical approaches are not enough to tackle these linguistic variables, but fuzzy set theory can be applied to deal with the situation. A fuzzy matching optimisation problem (Zhang & Lu 2009) needs to be solved to help prospects to choose the mobile products/services that are best matched their requirements. Although much research has been done in handling the linguistic variables, there is a further need to apply fuzzy matching technique in recommendation for mobile products/services.

8.2.2 MODELLING THE REQUIREMENTS OF CUSTOMERS

As can be seen from Table 8-1, different mobile products/services are described by different attribute sets. However, from the view of customers, all mobile products/services are composed of the following three sub-services: voice services, SMS/MMS services, and data services. Voice services are classified into calls between services within the same billing account, standard local and national calls, international calls, video calls, voice mail, and so on. SMS/MMS services include standard national SMS/MMS and international SMS/MMS. Data services include email services, Internet browsing services and so on. Therefore, the requirements for these characteristic sub-services are used to characterize customers' requirements. In practical situations, the characteristic sub-services should be identified by domain experts. They can be updated as new products and are continuously introduced. As an example, seven characteristic sub-services are abstracted:

S1: standard local and national calls;
S2: calls within the same billing account;
S3: international calls;
S4: standard national SMS/MMS;
S5: international SMS/MMS;
S6: email services;
S7: Internet browsing services.

To capture a new customer's requirements, a customer is asked seven questions (*Q1, ..., Q7*), shown in Table 8-2. These questions correspond to the characteristic sub-services mentioned above. For each characteristic sub-service, a customer's requirements contain two aspects: the usage amount and priority on the sub-service. In most real-world situations, it is difficult for a customer to give precise usage requirements with exact numbers. Customers are likely to measure their requirements in linguistic terms, such as ‘very large amount’, ‘small amount’, and so on. Therefore, it is more appropriate to allow customers to answer these questions using linguistic terms. The linguistic terms in set R are used to describe customers’ requirements here. The customer’s answers to these questions form a linguistic term vector, called a linguistic requirement vector (LRV).

$LRV=(r_1, r_2, ..., r_7)$, $r_i \in R$, $i=1, 2,..., 7$. r_i represents the customer’s requirement on sub-service S_i. The customer’s priorities on the sub-service are obtained by asking him/her to assign weights to these priorities. The weights are described by linguistic terms in set *W*. All the weights on these sub-services form another linguistic term vector, called a linguistic weight vector (LWV). $LWV=(w_1, w_2, …, w_7)$, $w_i \in R$, $i=1, 2,…, 7$. w_i represents the weight of sub-service S_i.

R={*Absolutely small* (*AS*), *Very small* (*VS*), *Small* (*S*), *Medium* (*M*), *large* (*L*), *Very large* (*VL*), *Absolutely large* (*AL*)}

W= {*Very low* (*VL*), *Low* (*L*), *Medium low* (*ML*), *Medium* (*M*), *Medium high* (*MH*), *High* (*H*), *Very high* (*VH*)}.

Fuzzy numbers are applied to deal with both sets of linguistic terms. Based on the results in Section 3.1.5, any form of fuzzy numbers can be used, called general fuzzy numbers, to describe these linguistic terms. In this chapter, $a_1, a_2, ..., a_7$ is defined to describe the terms in *R* and $b_1, b_2, ..., b_7$ is defined to describe the terms in *W*, respectively, a_j, b_j are general fuzzy numbers, $j = 1, 2, ..., 7$.

Table 8-2 Questions to Obtain Customers’ Requirements and Sub-Services

Q	Questions	S
Q_1	You have ____ standard local and national calls per month.	S_1
Q_2	You have ____ calls within your company per month.	S_2
Q_3	You have ____ international calls per month.	S_3
Q_4	You send ____ standard national SMS/MMS per month.	S_4
Q_5	You send ____ international SMS/MMS per month.	S_5

Q_6	You receive ____ emails per month.	S_6
Q_7	You visit ____ web pages per month.	S_7

8.2.3 MODELLING THE ATTRIBUTES OF MOBILE PRODUCTS/SERVICES

Mobile products/services are composed of several characteristic sub-services. Various mobile products/services differ mainly in the discounts and rewards for different sub-services which make them suitable for different usage situations. For example, some mobile products/services are suitable for large amounts of usage of local voice services, while others are more suitable for large amounts of usage of email services. Therefore, these mobile products/services are characterized by the suitable usage amount of the characteristic sub-services. As some attributes of mobile products/services are described by linguistic terms, it is reasonable to also represent their suitable usage amount of characteristic sub-services in linguistic terms. The linguistic terms in set R are used here. This raises the problem of how to map the attributes of the mobile products/services to the suitable usage amount.

First, the attributes of all mobile products/services are analysed and the relevant attribute set for each characteristic sub-service is identified. For example, attributes related to 'standard national SMS/MMS' include 'standard national SMS/MMS rate', 'Call2Anyone Value Offer', 'On-net Call Value Offer' and 'Unlimited Standard National SMS/MMS Offer'. Second, given the attribute set related to each characteristic sub-service, a rule is defined to identify the suitable usage amount depending on the attribute set. This process requires the analysis of all product attributes and should be supported by domain experts. For example, the rule for the suitable usage amount of 'standard national SMS/MMS service' can be defined as:

if "Unlimited Standard National SMS/MMS Offer" is included, it is "Absolute Large";

else if "Call2Anyone Value Offer" and "On-net Call Value Offer" is included,

if the sum of the two offer is larger than $1000, it is "Very Large";

else it is "Large";

else it is "Small".

It should be pointed out that the rule needs to be updated when new mobile products/services are introduced. Each mobile product/service is characterised by a linguistic characteristic vector $LCV=(c_1, c_2, \ldots, c_7)$, $c_i \in R$, $i=1, 2, \ldots, 7$. c_i represents the suitable usage amount on sub-service S_i. Taking the products in Table 8-1 as an example, the *LCV*s of them are illustrated in Table 8-3.

Table 8-3 LCVs of Four Products

$49 Cap Plan	(*L*, *AL*, *VS*, *L*, *VS*, *VS*, *VS*)
$49 Rate Plan	(*S*, *AL*, *S*, *S*, *S*, *VS*, *VS*)
Cap Plus 500MB	(*L*, *L*, *VS*, *AL*, *VS*, *M*, *M*)
$129 Timeless Max	(*AL*, *AL*, *L*, *AL*, *L*, *VL*, *VL*)

8.2.4 MODELLING THE USAGE RECORD

There is usually rich information in the usage records of existing customers for mobile products/services that have been used. This information reflects the characteristics of mobile products/services and can be used to match customers' requirements when generating recommendations. The view structure of customers' usage records is constructed according to the characteristic sub-services. For example, three customers' usage records in one month are illustrated in Table 8-4.

The usage records in Table 8-4 show the exact amount of usage of the characteristic sub-services, while the requirements of a new customer are illustrated in linguistic terms. It is reasonable to map the usage records to linguistic terms in set *R* as well. Then the usage record of the customer is represented by a linguistic usage vector, $LUV=(u_1, u_2, \ldots, u_7)$, $u_i \in R$, $i=1, 2, \ldots, 7$. u_i represents the usage amount on sub-service S_i.

The mapping is defined based on the analysis of the usage record database and the knowledge of domain experts. For example, for the 'standard local and national calls', the mapping from the amount of usage to the linguistic description can be defined as shown in Table 8-5. For other attributes, similar mappings can also be defined. As an example, the usage records of the three customers in Table 8-4 can be described as the following three *LUV*s: LUV_{u1}=(*VS, L, VS, S, VS, AS, AS*), LUV_{u2}=(*L, S, AS, VL, AS, M, VS*), LUV_{u3}=(*L, S, AS, L, AS, M, S*). It should be pointed out that the mapping should be updated periodically, because customers' usage situations change continuously.

Table 8-4 Usage Records of Three Customers

	Customer 1	Customer 2	Customer 3
Mobile service	$49 Yes Plan	Business Plus 500MB	Business Plus 500MB
standard local and national calls	40mins	620mins	610mins
calls within the same billing account	680mins	220mins	270mins
international calls	25mins		
standard national SMS/MMS	390	760	590
international SMS/MMS	24		
email		450MB	410MB
Internet browsing		60MB	120MB

Table 8-5 Mapping from Usage to Linguistic Description

amount of usage (mins)	0	(0, 200]	(200, 400]	(400, 600]	(600, 800]	(800, 1000]	>1000
linguistic description	*AS*	*VS*	*S*	*M*	*L*	*VL*	*AL*

8.2.5 A Fuzzy Matching-Based Recommendation Approach

A fuzzy matching based recommendation approach is presented in this section.

Input:

n – the number mobile products/services.

$\{p_1, p_2, \ldots, p_n\}$ - mobile product/service available in the market.

Each p_i is described by a linguistic characteristic vector;

$LCV_i=(c_{i1}, c_{i2}, \ldots, c_{i7})$.

If p_i has been used, its customers construct a set $U_i=\{u_{i,1}, u_{i,2}, \ldots, u_{i,m_i}\}$.

For customer $u_{i,j}$, the usage record is described by a linguistic usage vector

$LUV_{ij}=(u_{ij,1}, u_{ij,2}, \ldots, u_{ij,7})$.

Output:

$p_1, p_2, \ldots p_k$— k most appropriate products recommended.

The proposed approach generates recommendations through the following four steps.

Step1: Capturing a customer's requirements

The customer's requirements are obtained by asking the questions in Table 8-2. They are represented by a linguistic requirement vector $LRV=(r_1, r_2, \ldots, r_7)$ and a linguistic weight vector $LWV=(w_1, w_2, \ldots, w_7)$.

Step 2: Weights normalization

Normalized weights on the characteristic sub-services are calculated based on $LWV=(w_1, w_2, \ldots, w_7)$, and denoted as

$$\widetilde{w}_k^* = \frac{\widetilde{w}_k}{\sum_{i=1}^{7} w_{i0}^R} \quad \text{for } k=1, 2\ldots 7 \tag{8.1}$$

Step 3: Computing the matching degree of products to customer requirements

For a mobile product/service p_i, its matching degree to the customer's requirements is evaluated based on both LCV_i and all LUV_{ij} of its customers.

The matching degree of p_i to a customer's requirements based on p_i's linguistic characteristic vector is calculated by the following equation:

$$\tilde{m}_{c,i} = \sum_{j=1}^{7} \tilde{w}_j^* \cdot (1 - d(r_j, c_{ij})) \tag{8.2}$$

where $d(\cdot)$ is the quasi-distance of two finite fuzzy numbers. $\tilde{m}_{c,i}$ is normalized to be a positive fuzzy number, and its range belongs to closed interval [0, 1].

The fuzzy positive-ideal solution is defined as (FPIS, m^*) and the fuzzy negative-ideal solution is defined as (FNIS, m^-): where m^*=1 and m^-=0.The distance between $\tilde{m}_{c,i}$ and m^* is called positive distance, and the distance between $\tilde{m}_{c,i}$ and m^- is called negative distance. The two kinds of distance are calculated by $d_{c,i}^* = d(\tilde{m}_{c,i}, m^*)$ and $d_{c,i}^- = d(\tilde{m}_{c,i}, m^-)$ respectively, where $d(\cdot)$ is the quasi-distance of two finite fuzzy numbers. Then a closeness coefficient of p_i based on LCV_i, $CC_{c,i}$, is defined based on its $d_{c,i}^*$ and $d_{c,i}^-$ as:

$$CC_{c,i} = \tfrac{1}{2}(d_{c,i}^- + (1 - d_{c,i}^*)) \tag{8.3}$$

Considering p_i's customers' usage records, the matching degree of $u_{i,j}$ to the new customer's requirements is calculated by

$$\tilde{m}_{u,ij} = \sum_{k=1}^{7} \tilde{w}_k^* \cdot (1 - d(r_k, u_{ij,k})) \tag{8.4}$$

then a closeness coefficient of $u_{i,j}$ is defined as:

$$CC_{u,ij} = \tfrac{1}{2}(d_{u,ij}^- + (1 - d_{u,ij}^*)) \tag{8.5}$$

where $d_{u,i\,j}^* = d(\tilde{m}_{u,i\,j}, m^*)$ and $d_{u,ij}^- = d(\tilde{m}_{u,ij}, m^-)$.

Considering all the customers using product p_i, U_i, the closeness coefficient of p_i based on its customers, $CC_{u,i}$, is defined as

$$CC_{u,i} = \frac{1}{n}\sum_{j=1}^{n} CC_{u,ij} \tag{8.6}$$

Finally, the closeness coefficient of the product p_i is defined as:

$$CC_i = \alpha_1 \cdot CC_{c,i} + \alpha_2 \cdot CC_{u,i} \tag{8.7}$$

where α_1 and α_2 are weights of the two parts satisfying $\alpha_1 + \alpha_2$=1. If p_i is a new product, α_2=0. Otherwise, α_1 and α_2 are defined depending on the analysis of the data and the knowledge of the domain experts.

Step 4: Generating recommendations

Based on (8.7), the closeness coefficient of all the mobile products/services can be computed. k mobile products/services with the largest closeness coefficient are chosen for recommendation.

8.3 A RECOMMENDER SYSTEM FRAMEWORK FOR THE PROSPECTIVE CUSTOMER

The proposed recommendation approach is applied in the design of a recommender system for mobile products/services. The framework of the recommender system is shown in Figure 8-1. It consists of three main components: data builder, recommendation engine and interfaces. The data builder involves the development of two databases: a mobile products/services database and a customer usage records database. The recommendation engine generates recommendations by applying the fuzzy matching based recommendation approach. The interface part involves the collection of new customers' requirements, product attributes, existing customer usage records, and domain knowledge.

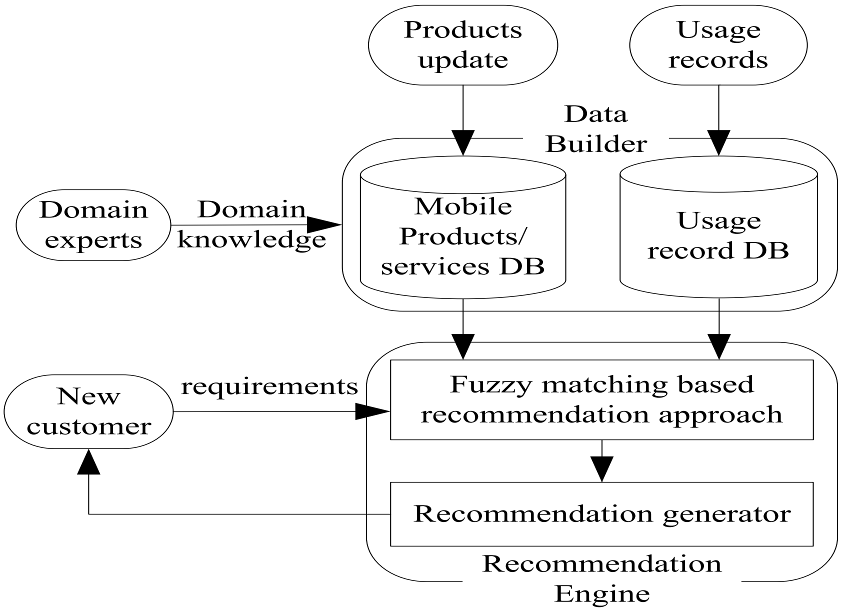


Figure 8-1 Framework of the Recommender System for Prospective Customers

8.4 AN ILLUSTRATION

An example is given to illustrate the proposed approach. The four mobile products/services described in Table 8-1 are: p_1: '\$49 Cap Plan'; p_2: '\$49 Rate Plan'; p_3: 'Cap Plus 500MB'; p_4: '\$129 Timeless Max'. Table 8-5 describes the customers' usage records. Our method is used to recommend appropriate products to a new customer. As an example, a_i and b_i, $i=1, 2, \ldots, 7$, are assigned to describe the linguistic terms with the triangular fuzzy numbers shown in Table 8-6 and 8-7, respectively.

Table 8-6 Linguistic Terms and Related Fuzzy Numbers for Customers' Requirements

AS	*VS*	*S*	*M*	*L*	*VL*	*AL*
(0,0,0.1)	(0,0.1,0.3)	(0.1,0.3,0.5)	(0.3,0.5,0.7)	(0.5,0.7,0.9)	(0.7,0.9,1.0)	(0.9,1.0,1.0)

Table 8-7 Linguistic Terms and Related Fuzzy Numbers for Weights

VL	*L*	*ML*	*M*	*MH*	*H*	*VH*
(0,0,0.1)	(0,0.1,0.3)	(0.1,0.3,0.5)	(0.3,0.5,0.7)	(0.5,0.7,0.9)	(0.7,0.9,1.0)	(0.9,1.0,1.0)

Step 1: by answering the questions in Table 8-2 and assigning the weights of the sub-services, the requirements of the new customer are obtained and represented as two linguistic vectors:

LRV=(*L, M, VS, VL, VS, M, S*), *LWV*=(*VH, VH, L, VH, L, H, ML*).

It can be seen that the customer requires a very large amount of standard national SMS/MMS, a large amount of standard local and national calls, and a medium amount of calls within the same billing account and email services. The customer requires very few other services.

Step 2: weights normalization:

$$\tilde{w}_1^* = \tilde{w}_2^* = \tilde{w}_4^* = \frac{1}{5.1} a_7,\ \tilde{w}_3^* = \tilde{w}_5^* = \frac{1}{5.1} a_2,\ \tilde{w}_6^* = \frac{1}{5.1} a_6,\ \tilde{w}_7^* = \frac{1}{5.1} a_3.$$

Step 3: the closeness coefficient of each mobile product/service based on their LCVs are calculated as: $CC_{c,1}$=0.65, $CC_{c,2}$=0.49, $CC_{c,3}$=0.77, and $CC_{c,4}$=0.55. The closeness coefficient of related customers are calculated as: $CC_{u,21}$=0.5, $CC_{u,31}$=0.78, and $CC_{u,32}$=0.76. Then, $CC_{u,3}$ can be computed by (13), and $CC_{u,3}$=0.77. Let α_1 and α_2 both be 0.5, by 7.13, CC_2=0.5 and CC_3=0.77. As there is no customer using the first and the fourth products, CC_1=$CC_{c,1}$=0.65 and CC_4=$CC_{c,4}$=0.55.

Step 4: based on the closeness coefficient computed in last step, the third product, 'Cap Plus 500MB' is recommended as the most suitable product for the customer. The second most suitable one is '$49 Cap Plan'. The results can help the customer to choose an appropriate product.

8.5 SUMMARY

In this chapter, a fuzzy matching-based recommendation approach is developed to help new customers to choose mobile products/services. In this approach, both the attributes of products and related customer usage records are analysed and used to generate recommendations. A set of characteristic sub-services of mobile products/services is abstracted. The requirements of new customers are obtained by asking them to answer a set of questions related to the characteristic sub-services and to assign weights to them. Customers' requirements are represented by linguistic terms. The attributes of a mobile product/service are mapped to the suitable usage amount on the characteristic sub-services, and represented as a linguistic vector LCV. A customer's usage records are mapped to the usage amount on the characteristic sub-services, and also represented as a linguistic vector LUV. A fuzzy matching approach is applied to find the products that are best matched to customers' requirements. A recommender system framework for prospective customers applying this proposed recommendation approach is developed. An example shows the effectiveness of the approach. This approach can be used in conjunction with the approach from Chapter 7 to build a web-based recommender system, including the design of the databases and customer interfaces (see Chapter 9).

CHAPTER 9

A PERSONALISED RECOMMENDER SYSTEM FOR THE BEST RECOMMENDATION

9.1 OVERVIEW

In Chapter 4, a personalised analytical eCRM framework has been proposed in which all the components of the proposed framework work together to support an organisation's marketing, sales and services strategies. The proposed framework also drives the fundamental principle for better customer understanding and handling. In Chapters 5 and 6, the different predictive approaches have been developed to establish an effective way to predict certain customer behaviours under the proposed personalised analytical eCRM framework. In Chapters 7 and 8, two different recommendation approaches have been proposed which, in conjunction with predictive approaches, can be implemented into the proposed personalised analytical eCRM framework. All these outputs should be seen as insights which can be put together in the context of a Personalised Recommender system with the aim of recommending the right product or service to customers, at the right cost, at the right time, and via right channel to enable the fulfilment of 'personalised' marketing objectives.

This chapter proposes a Personalised Recommender System (PRS). By using the proposed PRS, organisations can decide the next best marketing activity for each customer on a more informed basis, and can then select an 'individualised' approach that might include:

(1) An offer to prevent churn, mainly for high-value, high-risk customers.
(2) A promotion for the right add-on product and a targeted cross- or up-selling offer for customers with growth potential.
(3) The ability to trigger usage limitations and restrictions on customers with bad payment records.
(4) The development of a new product/offer tailored to the specific characteristics of an identified segment or other factor.

This chapter is organised as follows. Section 9.2 presents the proposed Personalised Recommender System (PRS). In Section 9.3, a Fuzzy-based Telecom Product Recommender System (FTCP-RS) is developed by using a telecom case study, including the architecture and design steps. Section 9.4 illustrates an initial application of the proposed FTCP-RS. Finally, a summary is given in Section 9.5.

9.2 A Personalised Recommender System

The five main components that have been taken into account in the design of the Personalised Recommender System (PRS) framework are:

(1) The current and expected or estimated customer profitability and value.
(2) A definition of the type of customer, the differentiating behavioural and demographic characteristics, and the identified needs and attitudes revealed through data analysis and segmentation.

(3) The growth potential estimated by relevant cross- or up-selling models and propensities.
(4) Identification of the defection risk (churn propensity) model.
(5) Design of a recommendation algorithm to make helpful recommendations.

In real world applications, the implementation of PRS is one way of using recommendation approaches to optimise below-the-line marketing campaign performance. It is the concept of 'optimising' every customer contact at every touch point or interaction, to provide additional value to the customer and additional revenue across the business. The recommender system is an ongoing process that helps organisations to ensure they are speaking to customers about products and services that are relevant to them, rather than what is relevant to the organisation. It is a way of providing an organisation's front line and marketing with a palette of optimised marketing offers based on accurate customer level predictions of behaviour and forecasted return on investment.

The proposed PRS is shown in Figure 9-1. There are three main components: (1) Customer data information knowledge base, which includes all information from different sources of organisation, internally and externally; (2) Customer analytics, which pre-processes the data from an enormous data warehouse by applying data mining and advance predictive models to turn data into meaningful, insightful and actionable information; (3) Recommendation engine, which combines all the learning from (2) with hybrid fuzzy-based personalised recommendation approaches to predict the best recommendation for customers.

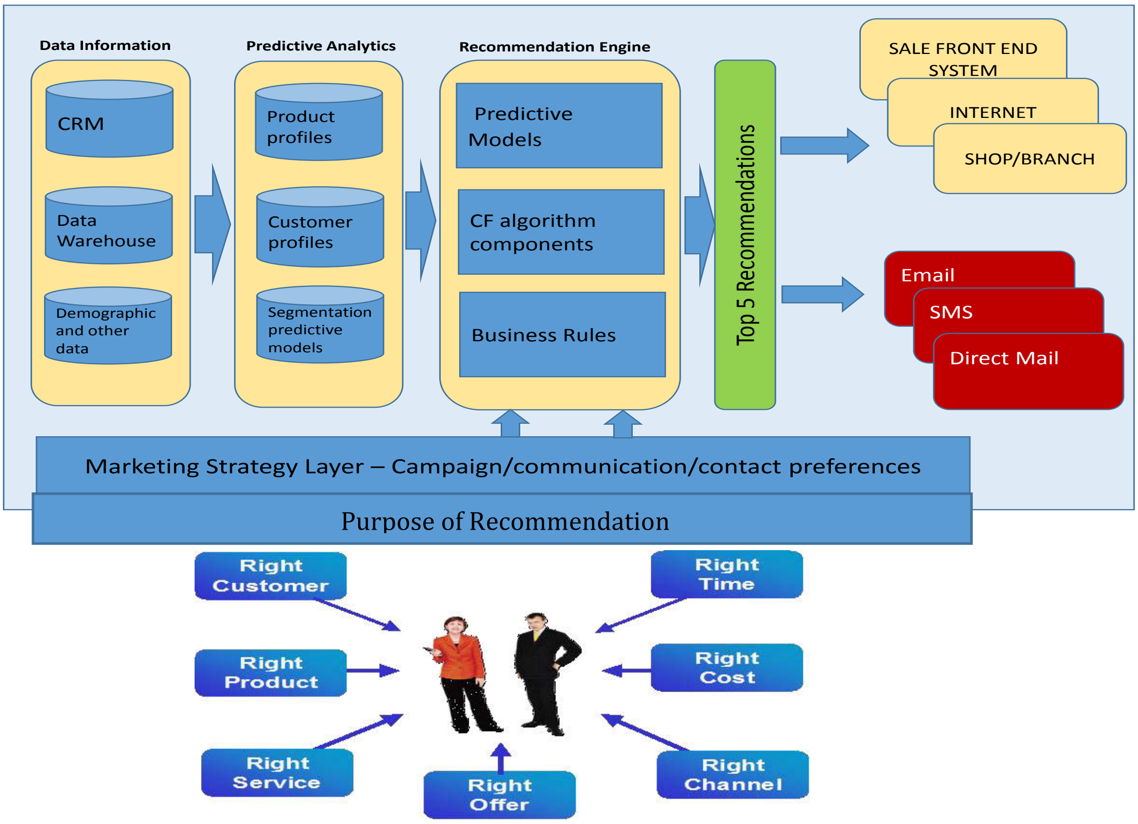


Figure 9-1 Proposed Personalised Recommender System

9.2.1 Recommendation Engine

The recommendation engine in Figure 9-1 plays a key role in generating appropriate recommendations or offers to the right customer, at the right time, via the right channel. This section presents the development of the recommendation engine that combines all the outcomes from previous development in Chapters 5, 6, 7 and 8.

9.2.1.1 Architecture

The architecture of the recommendation engine is illustrated in Figure 9-2 and consists of two components; (1) the pre-processing area, and (2) the recommendation area.

The pre-processing area processes all customer and product information data in terms of customer profiling and segmentation (behavioural and value), product profiling and segmentation. It produces model scores for the propensity to churn and propensity to buy by using the same predictive modelling techniques as were developed in Chapter 5 and Chapter 6.

The recommendation area is the heart of the recommendation engine that combines the two developed recommendation approaches in Chapter 7 and Chapter 8, and the predictive modelling techniques developed in Chapter 5 and Chapter 6,. It works with the pre-process area to process customer data, models and recommendation outcomes in an effective and efficient way. It then generates the top 5-10 recommendations, in conjunction with the business rules, to define the best recommendation for the customer.

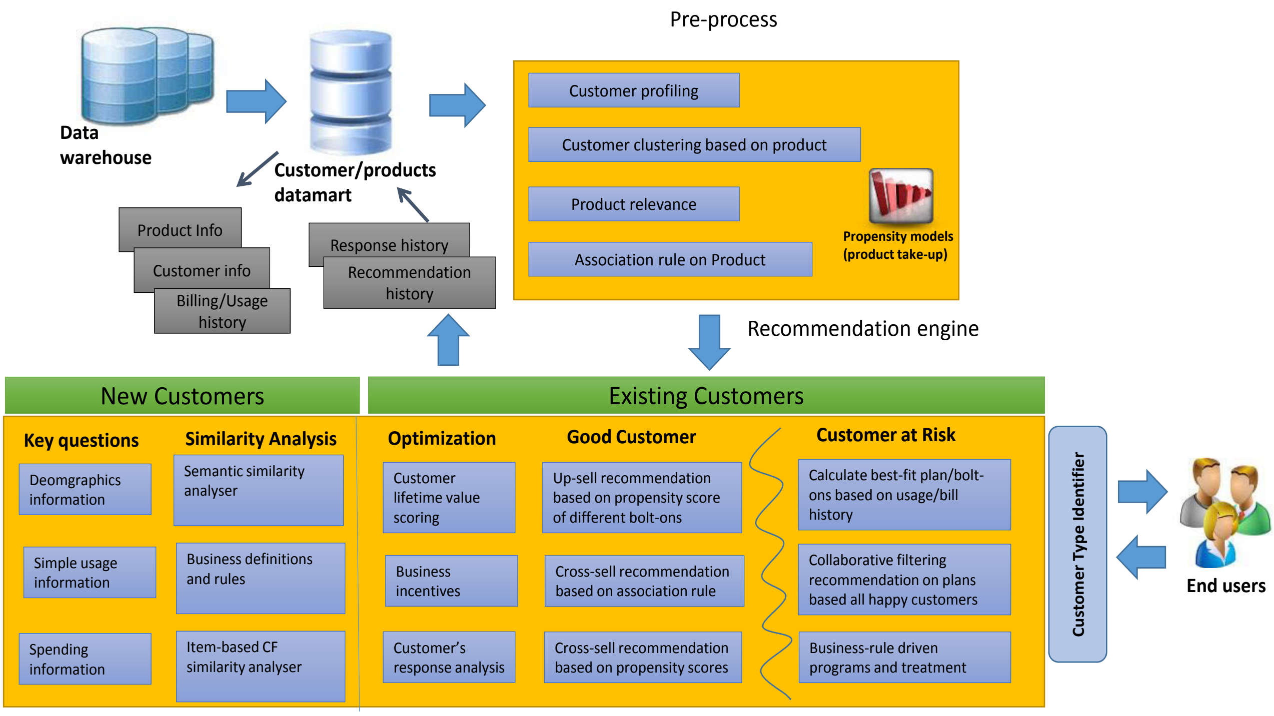


Figure 9-2 Architecture of Recommendation Engine

9.2.1.2 Recommendation Process

This recommendation engine has been developed by the following steps:

Step 1: Identify customer type: new customer or existing customer.

Step 2: Customers who are new to the organisation are directed to the process to gather new customer requirements, and recommendations for new customers are then generated based on the fuzzy matching-based recommendation approach described in Chapter 8.

Step 3: If the customer is already with the organisation, the recommendation engine will identify whether the customer is at risk by using the predictive models proposed in Chapters 5 and 6.

Step 4: If the customer is at risk of churning, using the hybrid fuzzy-set-based recommendation approach described in Chapter 7 is used to define the best offer for the customer.

Step 5: If the customer is at low risk of churning, an up-selling recommendation will be offered to the customer using the combined predictive models (propensity to buy) proposed in Chapters 5 and 6, together with the hybrid fuzzy base recommendation approach demonstrated in Chapter 7. The priority of ranking the cross-sell recommendation is as follows:

(1) Most popular products/plans among similar customers also resulting in high propensity scores;

(2) Most popular products/plans among similar customers not supported by the propensity scores, or missing propensity scores;

(3) Products/plans with high propensity scores which have no popularity among similar customers.

Business rules such as bundling offers and tactical offers are also considered when ranking cross-sell recommendations.

In the following section, a real case implementation of the proposed personalised recommender system is demonstrated, using the telecom environment as the showcase.

9.2.2 CASE STUDY - FUZZY-BASED TELECOM PRODUCT RECOMMENDER SYSTEM ARCHITECTURE AND DEVELOPMENT

In this section, a Fuzzy-based Telecom Product Recommender System (FTCP-RS) is proposed using a telecom application as a case study. This real case study is the implementation of a personalised recommender system based on the proposed Hybrid Fuzzy Set-based Recommendation Approach in Chapter 7, which deals with the challenges presented in Section 7.2 and also combines the predictive modelling techniques proposed in Chapters 5 and 6. This proposed FTGP-RS aims to help a customer to choose the most appropriate telecom products/services via the Internet.

9.2.3 SYSTEM ARCHITECTURE

The FTCP-RS is developed for the telecom industry. The system architecture of FTCP-RS is illustrated in Figure 9-3. It is implemented using a Multi-Tier architecture on a Microsoft .NET 3.5 platform. It consists of three main parts: client, web server and database server.

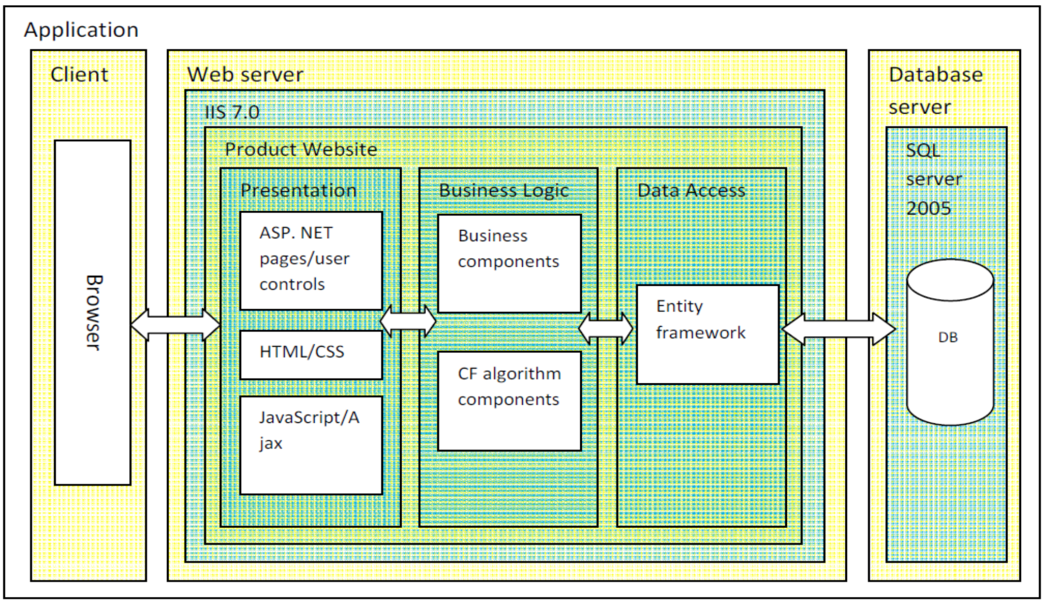


Figure 9-3 Architecture of FTCP-RS

Client

Client is the user interface presented on a web browser. When a customer visits the website of the telecom company, the client browser will send requests to the web server every time the user performs an action, such as login or visiting a new page. When the web server receives the requests, it retrieves the requested resources and sends them back to the client browser.

Web server

Websites are hosted in web servers. A web server consists of two dimensions: the logical server, which is the software that serves the web requests, and the physical server, which is the computer running the logical server and storing all the resources. Based on the web server, the FTCP-RS website can be divided into three layers: the presentation layer, business logic layer and data access layer.

- *Presentation layer:* This layer is responsible for generating the requested web pages and handling the UI logics and events. When a user requests to view a new page, the presentation layer will invoke corresponding methods in the business logic layer, extract the request data, transform the data into HTML page and send it back to the client.
- *Business Logic Layer:* This layer defines the business rules and processes of the application and serves as a mediator between the presentation layer and the data access layer. In FTCP-RS, the business logic layer contains two parts: one part implements the FTCP-RS website business processes and the other part implements the hybrid fuzzy-based telecom product recommendation approach.
- *Data Access Layer:* This layer deals with the data operations of the database and transfers data with the business logic layer. In FTCP-RS, the data access layer is implemented using Entity Framework.

Database Server

The database server is the computer server that runs the database applications. In FTCP-RS, we use SQL Server 2005 as the database application because it is compatible with all the Microsoft technologies we use. The database server can be the same computer as the web server or a separate server running the database application.

9.2.4 FTCP-RS Development Steps

This recommender system has been developed by the following steps:

Step 1: Classification and clustering of existing customers through retrieving and analysing the existing customer profile database. The existing customer profile database has rich profile information about existing customers, such as customer name, customer account(s), current products/services, re-contract time, and customer usage information.

Step 2: Sets up a set of business rules based on business requirements for existing customers. This step uses the case described in Chapter 7, which designes and applies five types of business rules: (1) bundle rules, (2) multiple products rules, (3) discount rules, (4) product rules and (5) special offers. Three examples of the business rules are defined as follows:

"Some fixed line products cannot be purchased standalone. They have to be bundled with a fixed broadband product."

"A customer can receive additional discounts for some products, if they are purchased together."

"For a period of time, some products may be on special or the business may be promoting those products".

Step 3: Establish a customer view from the current customer database. This step involves database information retrieval and incorporation, and the customer view (database) structure design, as well as the physical storage of data in the view (database).

Step 4: Design a set of online data collection pages to obtain existing customers' requirements and web-based interface as well as outputs.

Step 5: Implement the developed recommendation approach given in Chapter 7 and predictive modelling techniques developed in Chapters 5 and 6.

Step 6: Interface design, including customer data collection, propensity models, recommendation list generation and related explanations.

Step 7: System testing and revision. Test cases are conducted to test and evaluate the performance of the developed intelligent recommender system, FTCP-RS, using telecom customer data.

The FTCP-RS site map is presented in Figure 9-3.

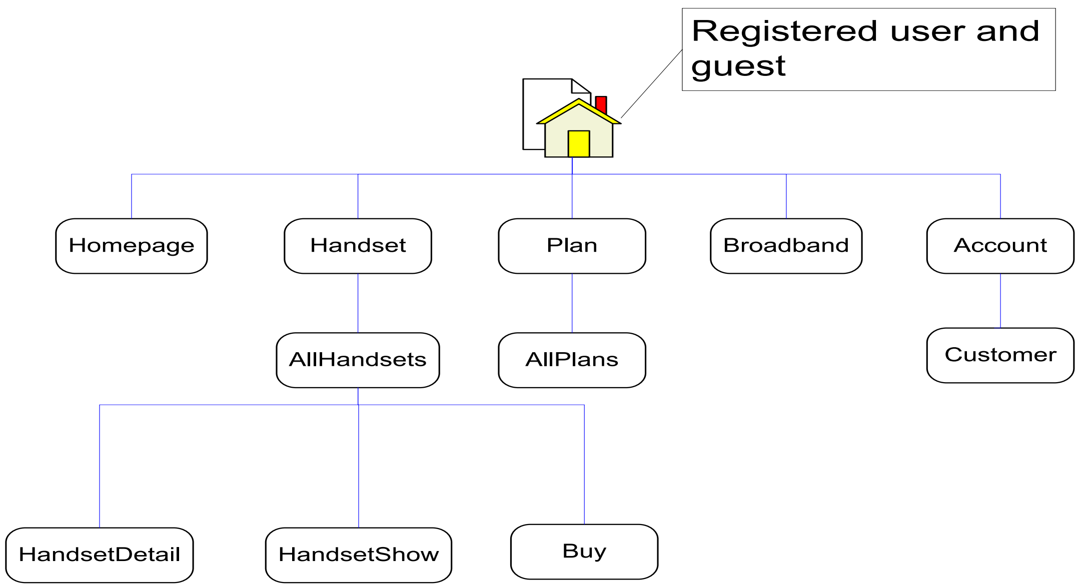


Figure 9-4 FTCP-RS Site Map

9.3 SYSTEM APPLICATION

The main process of recommendation in the use of FTCP-RS is described as follows:

(1) To collect customer information. In this step, the rating data of existing customers are collected in the mobile product/service and handset detail web pages on which a customer can rate a mobile product/service and a handset. The rating value, as well as

the customer ID and mobile product/service ID or handset ID, will then be stored in the database.

(2) To gather data from similar existing customers, including purchase records, usage, website visit history and personal profiles;

(3) To collect related product data and build a product database (current product data is in the form shown in Table 7-1) and determine the main features;

(4) To analyse the collected data (customer and product), business rules, and predict the ratings of unrated products using fuzzy techniques;

(5) To select the top-*K* products with the highest predicted ratings as recommendations for customers. There are two types of recommendation:

(a) *Mobile products/services and handset recommendations*

After a customer logs into the homepage, FTCP-RS is able to generate recommendations to the customer. The system will first read the parameter settings from the configuration file, including parameters such as the number of neighbours and the number of items to be recommended. The system will then load the rating records of all users and use the hybrid method described in Section 7.2 to make recommendations. Finally, the system will return a list of recommended handsets.

(b) *Package recommendation*

For a customer whose contract will expire in four weeks' time, the FTCP-RS will automatically recommend a package which includes handsets, plans and extra telecom services.

Figure 9-5 shows a re-contracting page of the FTCP-RS system which contains a list of telecom product/service contracts. Figure 9-6 presents details of a telecom product/service contract with usage history. Figure 9-7 illustrates a recommendation generated by the FTCP-RS to an existing customer based on their usage.

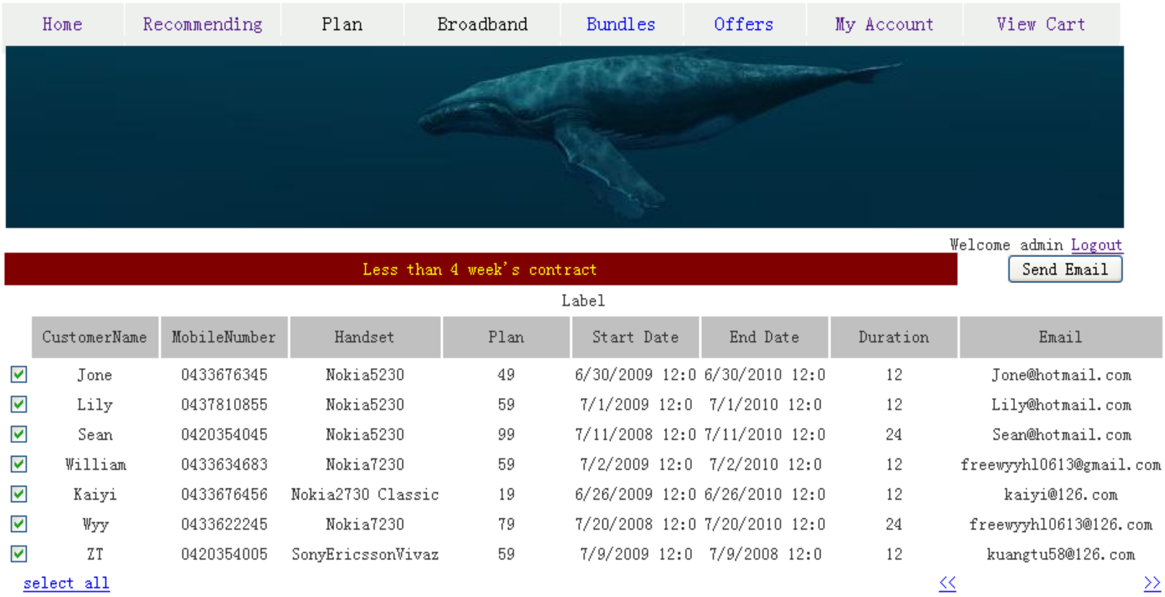


Figure 9-5 List of Telecom Product/Service Contract Generated By FTCP-RS

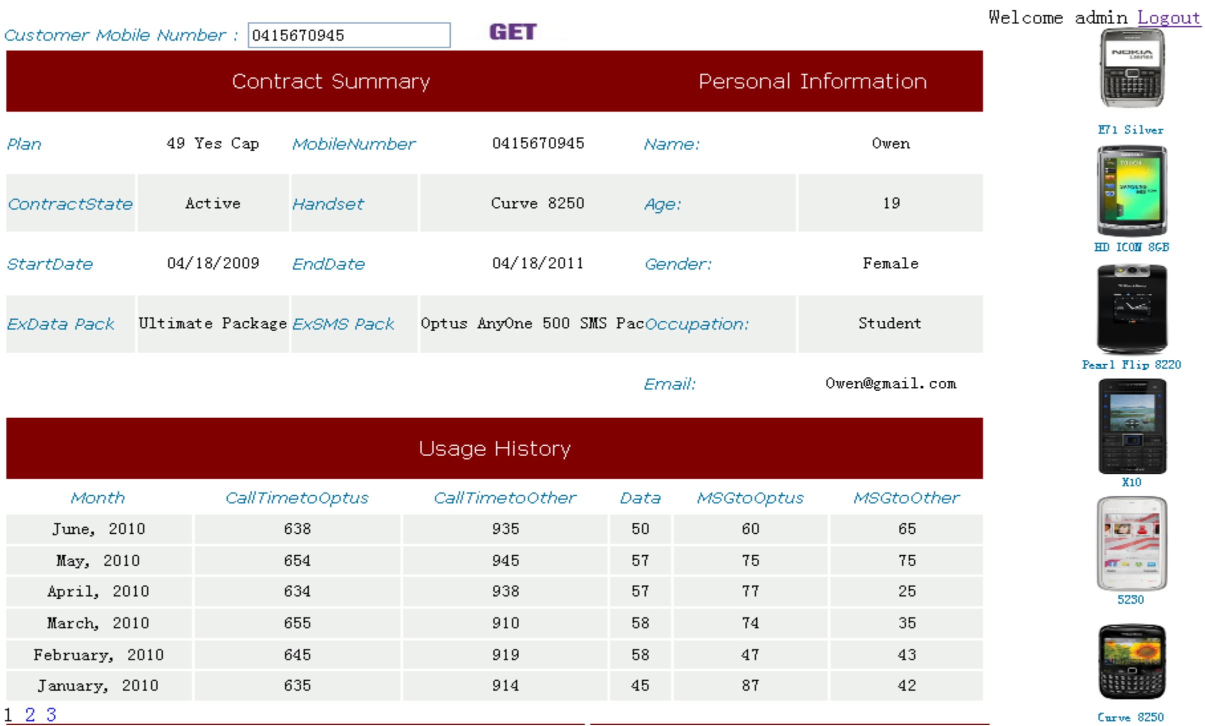


Figure 9-6 Example of a Telecom Product/Service Contract with Usage History

Usage History					
Month	CallTimetoOptus	CallTimetoOther	Data	MSGtoOptus	MSGtoOther
June, 2010	638	935	50	60	65
May, 2010	654	945	57	75	75
April, 2010	634	938	57	77	25
March, 2010	655	910	58	74	35
February, 2010	645	919	58	47	43
January, 2010	635	914	45	87	42

1 2 3

Average useage in last six months			
Calls to Optus	318	Calls to Others	445
TotalTimeToOptus	643	TotalTimeToOther	926
TotalCallValue	1445.95	Message To Optus	70
Message To Other	47	Total Message Value	29.25
Data	54.167	Data Value	832.00
Total Value	2307.20	Actual Fee	862.44

Plan and package upgrade option for you!	
New Plan	99 Timeless Cap
New monthly cost	128.99
Customer gets Unlimited call & message and save 733.45 dollars every month	

Figure 9-7 Recommendations of FTCP-RS

9.4 SUMMARY

This chapter proposes a personalised recommender system (PRS) which combines the hybrid data mining predictive modelling techniques and the hybrid fuzzy set-based recommendation approaches. This system applies fuzzy set techniques and knowledge base for product and service recommendation to all customers who already have, or have yet to develop, a relationship with the organisation. In particular, the detail of the components within the proposed personalised recommender system has been described, including system design, recommendation engine architecture and process. A case of the implementation of a personalised recommender system (FTCP-RS) based on a real world example is demonstrated in Section 9.3. This case study presents the development and implementation processes that focuses on existing customers within the organisation. Due to the constraint imposed by the lack of data for prospective customers, there are challenges in building a closed loop feedback system, and less opportunity to capture customer rating information. Thus, more work still needs to be done to the presented FTCP-RS to improve its performance.

CHAPTER 10

CONCLUSIONS AND FURTHER RESEARCH

This research was motivated by the adoption of personalisation technologies, in particular, the development of new predictive modelling techniques and new proposed recommendation approaches in analytical eCRM applications.

Many methodologies and approaches have been developed in both predictive modelling and recommender systems and have been presented in the literature of recent years. Many researchers have demonstrated the great benefit of using eCRM in conjunction with predictive modelling to recommend relevant offers to customers. However, the literature has paid very limited attention to recommendation techniques in eCRM applications. The rapid development of eCRM has led to increasing customer interaction with organisations via email and the World Wide Web, and these interactions are becoming richer and more varied. Hence, more and more organisations are implementing eCRM applications to drive their personalised marketing and sales, and to implement their retention and acquisition strategies.

By using predictive modelling techniques, the likelihood of a customer taking certain actions can be tracked; however, they cannot take into account customer interests and preferences. Recommendation techniques can tackle these kinds of question, and the combination of these two approaches will add value to personalisation applications, making the process more robust and powerful. There are still some crucial challenges in CF-based recommendation approaches, in particular the problems of data uncertainty and data parity, along with inaccurate recommendations, which need to be investigated more extensively and resolved. This research makes the following main contributions:

(1) It proposes a personalised analytical eCRM framework. This proposed framework will be of great value in personalised recommendation in eCRM research. It will facilitate the transformation of current eCRM, enabling organisations to drive personalised marketing and sales, and effectively achieve their retention and acquisition strategies. The proposed personalised analytical eCRM framework is a unified framework which includes one novel segmentation algorithm, two different novel hybrid recommendation algorithms and two different novel hybrid predictive modelling algorithms. These are capable of dealing with different sources of information (e.g., customer demographic and behaviour data, fuzzy linguistic rating, and multi-criteria rating information) to handle most of the limitations of CF-based recommendation approaches.

(2) It develops a hybrid customer segmentation model (Chapter 4). This proposed segmentation model tackles migrating customer segmentation based on survey data into the customer behaviour database, addressing a challenging topic in customer relationship management. The developed model works as a five-step hybrid segmentation algorithm, particularly focusing on replacing the missing variables in the internal database. Experiments have shown the capability of the model in solving this kind of problem.

(3) It develops two different hybrid predictive modelling approaches and related algorithms (Chapter 5 and Chapter 6) for customer behaviours. One proposed hybrid predictive modelling approach combines conventional methods into a dynamic and time related prediction method. The other hybrid predictive modelling approach applies a fuzzy measure-based dynamic time related prediction method. These two new approaches provide a comprehensive perspective on customer behaviours by examining 'who are they', 'how do they behave', 'when are they likely to churn or buy', 'why do

they churn or buy', and 'how can organisations prevent or recommend customers from taking the decision to leave or buy?. The two approaches help organisations to improve their understanding of customer behaviour.

(4) It develops two different hybrid recommendation approaches and related algorithms (Chapter 7 and Chapter 8). One approach is a hybrid fuzzy matching-based recommendation approach, which aims to support prospective customers (new to the organisation) to select the most appropriate products/services and will add value to an organisation's personalised sales strategy. Another proposed approach is a hybrid Fuzzy Set-Based Hybrid Recommendation approach that combines Item-based Collaborative Filtering and User-based Collaborative Filtering with fuzzy set techniques and a knowledge-based method. This approach guides existing customers in the selection of the most appropriate products/services, which helps the organisation to achieve their retention strategy.

(5) It proposes a personalised recommender system (Chapter 8). The proposed personalised recommender system combines the hybrid predictive modelling techniques and the hybrid fuzzy set-based recommendation approaches in (3) and (4). The outputs gain insight which can be put together in the context of the personalised recommender system. This personalised recommender system provides intelligent recommendations to the right customer about the right product/service, at the right time and via the right channel, which enables organisations to achieve their 'personalised' marketing objectives. A real world case study exemplifies the implementation of the proposed personalised recommender system for recommending relevant products/services to existing customers in a telecom business environment.

Future research can be extended through the proposed personalised recommender system in (5) to implement the recommendation for prospective customers. This enables an ongoing process that helps the organisation to ensure that they are communicating with customers who already have, or have yet to develop, a relationship with the organisation about products and services that are relevant to them (not what is relevant to the organisation). It will provide a palette of optimised marketing offers based on accurate customer level predictions of behaviour and forecasted return on investment to the organisation’s front line customer support and marketing arms. Finally, a number of crucial challenges in CF-based recommendation approaches need to be investigated more extensively and resolved to improve the capability and efficiency of the framework.

REFERENCE

Abbasimehr, H., Setak, M. & Tarokh, M. 2011, 'A neuro-fuzzy classifier for customer churn prediction ', *International Journal of Computer Applications*, vol. 19, no. 8, pp. 35-41.

Adebanjo, D. 2003, 'Classifying and selecting e-CRM applications: An analysis-based proposal', *Management Decision*, vol. 41, no. 6, pp. 570-577.

Adomavicius, G. & Kwon, Y. 2012, 'Improving aggregate recommendation diversity using ranking-based techniques', *IEEE Transactions on Knowledge and Data Engineering,* vol. 24, no. 5, pp. 896-911.

Adomavicius, G. & Tuzhilin, A. 2005, 'Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions', *IEEE Transactions on Knowledge and Data Engineering,* vol. 17, no. 6, pp. 734-749.

Adomavicius, G. & Zhang, J. 2012, 'Stability of recommendation algorithms ', *ACM Transactions on Information Systems (TOIS)*, vol. 30, no. 4, p. 23.

Alonso, K., Zorrilla, M., Inan, H., Palau, M., Confalonieri, R., Vazquez-Salceda, J., Calle, J. & Castro, E. 2012, 'Ontology-based tourism for all recommender and information retrieval system for Interactive Community Displays', *8th International Conference on Information Science and Digital Content Technology (ICIDT), 2012*, vol. 3, IEEE, pp. 650-655.

Amiri, A. 2006, 'Customer-oriented catalog segmentation: Effective solution approaches', *Decision Support Systems*, vol. 42, no. 3, pp. 1860-1871.

Anderson, J. 1982, 'Logistic discrimination', in P. R. Krishnaiah and L. Kanal (Eds.), *Handbook of Statistics, Vol. 2. Elsevier North-Holland, Amsterdam*, pp. 169-191.

Au, S.T., Guangqin, M. & Rensheng, W. 2011, 'Iterative multivariate regression model for correlated responses prediction ', *2011 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC),* pp. 55-59.

Au, T., Ma, G. & Li, S. 2003, 'Applying and evaluating models to predict customer attrition using data mining techniques', *Journal of Comparative International Management*, vol. 6, no. 1.

Azila, N. & Noor, M. 2011, 'Electronic customer relationship management performance: Its impact on loyalty from customers' perspectives', *International Journal of e-Education, e-Business, e-Management and e-Learning*, vol. 1, no. 2.

Baesens, B., Viaene, S., Van den Poel, D., Vanthienen, J. & Dedene, G. 2002, 'Bayesian neural network learning for repeat purchase modelling in direct marketing', *European Journal of Operational Research*, vol. 138, no. 1, pp. 191-211.

Benferhat, S., Dubois, D., Kaci, S. & Prade, H. 2006, 'Bipolar possibility theory in preference modeling: Representation, fusion and optimal solutions', *Information Fusion*, vol. 7, no. 1, pp. 135-150.

Biancalana, C., Gasparetti, F., Micarelli, A. & Sansonetti, G. 2013, 'An approach to social recommendation for context-aware mobile services', *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 4, no. 1, p. 10.

Bijmolt, T.H., Leeflang, P.S., Block, F., Eisenbeiss, M., Hardie, B.G., Lemmens, A. & Saffert, P. 2010, 'Analytics for customer engagement', *Journal of Service Research*, vol. 13, no. 3, pp. 341-356.

Bobadilla, J., Serradilla, F. & Bernal, J. 2010, 'A new collaborative filtering metric that improves the behavior of recommender systems', *Knowledge-Based Systems*, vol. 23, no. 6, pp. 520-528.

Bradshaw, D. & Brash, C. 2001, 'Managing customer relationships in the e-business world: How to personalise computer relationships for increased profitability', *International Journal of Retail & Distribution Management*, vol. 29, no. 12, pp. 520-530.

Brown, J.D. 2010, 'How are PCA and EFA used in language test and questionnaire development?', *Statistics*, vol. 14, no. 2.

Buckinx, W. & Van den Poel, D. 2005, 'Customer base analysis: Partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting', *European Journal of Operational Research*, vol. 164, no. 1, pp. 252-268.

Burez, J. & Van den Poel, D. 2009, 'Handling class imbalance in customer churn prediction', *Expert Systems with Applications*, vol. 36, no. 3, Part 1, pp. 4626-4636.

Burke, R. 2002, 'Hybrid recommender systems: Survey and experiments', *User Modeling and User-Adapted Interaction*, vol. 12, no. 4, pp. 331-370.

Burke, R. 2007, 'Hybrid web recommender systems', in P. Brusilovsky, A. Kobsa and W. Nejdl (Eds.), *The Adaptive Web*, Springer, Berlin-Heidelberg, pp. 377-408.

Buttle, F. 2012, *Customer relationship management*, Routledge.

Cao, Y. & Li, Y. 2007, 'An intelligent fuzzy-based recommendation system for consumer electronic products', *Expert Systems with Applications*, vol. 33, no. 1, pp. 230-240.

Cespedes, F.V. & Smith, H.J. 2012, 'Database marketing: New rules for policy and practice', *Sloan Management Review*, vol. 34.

Chang, W.-L. & Wu, Y.-X. 2010, 'A framework for CRM e-services: From customer value perspective', in R. Sharman, H. Raghav Rao and T. S. Raghu (Eds.), *Exploring the Grand Challenges for Next Generation E-Business*, Springer, Berlin-Heidelberg, pp. 235-242.

Chatrchyan, S., Khachatryan, V., Sirunyan, A., Tumasyan, A., Adam, W., Aguilo, E., Bergauer, T., Dragicevic, M., Erö, J. & Fabjan, C. 2012, 'Inclusive and differential measurements of the charge asymmetry in proton–proton collisions at', *Physics Letters B*.

Chen, C.-M. & Duh, L.-J. 2008, 'Personalized web-based tutoring system based on fuzzy item response theory', *Expert Systems with Applications*, vol. 34, no. 4, pp. 2298-2315.

Chen, P.-Y., Chou, Y.-C. & Kauffman, R.J. 2009, 'Community- based recommender systems: Analyzing business models from a systems operator's perspective ', *42nd Hawaii International Conference on System Sciences, 2009. HICSS'09*, IEEE, pp. 1-10.

Chen, Y., Zhang, G., Hu, D. & Fu, C. 2007, 'Customer segmentation based on survival character', *Journal of Intelligent Manufacturing*, vol. 18, no. 4, pp. 513-517.

Chen, Z.-Y. & Fan, Z.-P. 2013, 'Dynamic customer lifetime value prediction using longitudinal data: An improved multiple kernel SVR approach', *Knowledge-Based Systems*.

Chen, Z.-Y., Fan, Z.-P. & Sun, M. 2012, 'A hierarchical multiple kernel support vector machine for customer churn prediction using longitudinal behavioral data', *European Journal of Operational Research*.

Collins, M., Schapire, R.E. & Singer, Y. 2002, 'Logistic regression, AdaBoost and Bregman distances', *Machine Learning*, vol. 48, no. 1-3, pp. 253-285.

Coltman, T. 2007, 'Why build a customer relationship management capability?', *Journal of Strategic Information Systems*, vol. 16, no. 3, pp. 301-320.

Cooil, B., Aksoy, L. & Keiningham, T.L. 2008, 'Approaches to Customer Segmentation', *Journal of Relationship Marketing*, vol. 6, no. 3-4, pp. 9-39.

Cornelis, C., Guo, X., Lu, J. & Zhang, G. 2005, 'A fuzzy relational approach to event recommendation', *Proceedings of the Indian International Conference on Artificial Intelligence*.

Cornelis, C., Lu, J., Guo, X. & Zhang, G. 2007, 'One-and-only item recommendation with fuzzy logic techniques', *Information Sciences*, vol. 177, no. 22, pp. 4906-4921.

Coussement, K., Benoit, D.F. & Van den Poel, D. 2010, 'Improved marketing decision making in a customer churn prediction context using generalized additive models', *Expert Systems with Applications*, vol. 37, no. 3, pp. 2132-2143.

Coussement, K. & Van den Poel, D. 2008a, 'Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques', *Expert Systems with Applications*, vol. 34, no. 1, pp. 313-327.

Coussement, K. & Van den Poel, D. 2008b, 'Integrating the voice of customers through call center emails into a decision support system for churn prediction', *Information & Management*, vol. 45, no. 3, pp. 164-174.

Cox, D.R. 1972, 'Regression models and life-tables', *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 187-220.

Datta, P., Masand, B., Mani, D.R. & Li, B. 2000, 'Automated cellular modeling and prediction on a large scale ', *Artificial Intelligence Review*, vol. 14, no. 6, pp. 485-502.

De Bock, K.W. & Van den Poel, D. 2011, 'An empirical evaluation of rotation-based ensemble classifiers for customer churn prediction', *Expert Systems with Applications*, vol. 38, no. 10, pp. 12293-12301.

DeLong, E.R., DeLong, D.M. & Clarke-Pearson, D.L. 1988, 'Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach', *Biometrics*, pp. 837-845.

Deshpande, M. & Karypis, G. 2004, 'Item-based top-n recommendation algorithms', *ACM Transactions on Information Systems (TOIS)*, vol. 22, no. 1, pp. 143-177.

Dhingra, M. & Dhingra, V. 2013, 'Determinants of electronic customer relationship management (e-CRM) for customer satisfaction in banking sector in India', *African Journal of Business Management*, vol. 7, no. 10, pp. 762-768.

Efron, B. 1988, 'Logistic regression, survival analysis, and the Kaplan-Meier curve', *Journal of the American Statistical Association*, vol. 83, no. 402, pp. 414-25.

Fabrigar, L.R., Wegener, D.T., MacCallum, R.C. & Strahan, E.J. 1999, 'Evaluating the use of exploratory factor analysis in psychological research', *Psychological methods*, vol. 4, pp. 272-299.

Foss, B., Stone, M. & Ekinci, Y. 2008, 'What makes for CRM system success—Or failure?', *Journal of Database Marketing & Customer Strategy Management*, vol. 15, no. 2, pp. 68-78.

Friedman, J., Hastie, T. & Tibshirani, R. 2000, 'Additive logistic regression: A statistical view of boosting (with discussion and a rejoinder by the authors)', *The Annals of Statistics*, vol. 28, no. 2, pp. 337-407.

Ganesh, J., Arnold, M.J. & Reynolds, K.E. 2000, 'Understanding the customer base of service providers: An examination of the differences between switchers and stayers', *The Journal of Marketing*, pp. 65-87.

Gao, M., Liu, K. & Wu, Z. 2010, 'Personalisation in web computing and informatics: Theories, techniques, applications, and future research', *Information Systems Frontiers*, vol. 12, no. 5, pp. 607-629.

Gedikli, F. & Jannach, D. 2013, 'Improving recommendation accuracy based on item-specific tag preferences', *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 4, no. 1, pp. 11.

Ghazanfar, M.A., Prügel-Bennett, A. & Szedmak, S. 2012, 'Kernel-Mapping Recommender system algorithms', *Information Sciences*, vol. 208, pp. 81-104.

Gil-Saura, I. & Ruiz-Molina, M.-E. 2009, 'Customer segmentation based on commitment and ICT use', *Industrial Management & Data Systems*, vol. 109, no. 2, pp. 206-223.

Goossen, F., IJntema, W., Frasincar, F., Hogenboom, F. & Kaymak, U. 2011, 'News personalization using the CF-IDF semantic recommender', paper presented to the *International Conference on Web Intelligence, Mining and Semantics*, 25-27 May, Sogndal, Norway.

Goy, A., Ardissono, L. & Petrone, G. 2007, 'Personalization in e-commerce applications', in P. Brusilovsky, A. Kobsa and W. Nejdl (Eds.), *The Adaptive Web*, Springer, Berlin-Heidelberg, pp. 485-520.

Greiner, R., Su, X., Shen, B. & Zhou, W. 2005, 'Structural extension to logistic regression: Discriminative parameter learning of belief net classifiers ', *Machine Learning*, vol. 59, no. 3, pp. 297-322.

Guisan, A., Edwards Jr, T.C. & Hastie, T. 2002, 'Generalized linear and generalized additive models in studies of species distributions: setting the scene', *Ecological modelling*, vol. 157, no. 2, pp. 89-100.

Guo, L., Luo, Y., Zhou, Z. & Ji, M. 2013, 'A recommendation trust method based on fuzzy clustering in P2P networks', *Journal of Software*, vol. 8, no. 2, pp. 357-360.

Guo, X. & Lu, J. 2007, 'Intelligent e-government services with personalized recommendation techniques', *International Journal of Intelligent Systems*, vol. 22, no. 5, pp. 401-417.

Gustafsson, A., Johnson, M.D. & Roos, I. 2005, 'The effects of customer satisfaction, relationship commitment dimensions, and triggers on customer retention', *Journal of Marketing*, pp. 210-218.

Hadaya, P. & Cassivi, L. 2009, 'Collaborative e-product development and product innovation in a demand-driven network: The moderating role of eCRM', *Electronic Markets*, vol. 19, no. 2-3, pp. 71-87.

Hadden, J., Tiwari, A., Roy, R. & Ruta, D. 2007, 'Computer assisted customer churn management: State-of-the-art and future trends', *Computers & Operations Research*, vol. 34, no. 10, pp. 2902-2917.

Han, S.H., Lu, S.X. & Leung, S.C. 2012, 'Segmentation of telecom customers based on customer value by decision tree model', *Expert Systems with Applications*, vol. 39, no. 4, pp. 3964-3973.

Harrigan, P., Ramsey, E. & Ibbotson, P. 2011, 'Critical factors underpinning the e-CRM activities of SMEs', *Journal of Marketing Management*, vol. 27, no. 5-6, pp. 503-529.

Herlocker, J.L., Konstan, J.A., Borchers, A. & Riedl, J. 1999, 'An algorithmic framework for performing collaborative filtering', paper presented to the *22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 15-19 August, Berkeley, CA, USA.

Herrera-Viedma, E. & Porcel, C. 2009, 'Using incomplete fuzzy linguistic preference relations to characterize user profiles in recommender systems ', *Ninth International Conference on Intelligent Systems Design and Applications,* pp. 90-95.

Hillenbrand, C. & Money, K. 2009, 'Segmenting stakeholders in terms of corporate responsibility: Implications for reputation management ', *Australasian Marketing Journal (AMJ)*, vol. 17, no. 2, pp. 99-105.

Hoekstra, J.C. & Huizingh, E.K. 1999, 'The lifetime value concept in customer-based marketing', *Journal of Market-Focused Management*, vol. 3, no. 3-4, pp. 257-574.

Hofmann, T. 2003, 'Collaborative filtering via Gaussian probabilistic latent semantic analysis', paper presented to the *26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 28 July-1 August, Toronto, Canada.

Hosmer, D.W. & Lemeshow, S. 2004, *Applied Logistic Regression*, Wiley-Interscience, Hoboken NJ.

Hosseni, M.B. & Tarokh, M.J. 2011, 'Customer segmentation using CLV elements', *Journal of Service Science and Management*, vol. 4, no. 3, pp. 284-290.

Huang, B., Kechadi, M.T. & Buckley, B. 2012, 'Customer churn prediction in telecommunications', *Expert Systems with Applications*, vol. 39, no. 1, pp. 1414-1425.

Huang, Z., Zeng, D. & Chen, H. 2007, 'A comparison of collaborative-filtering recommendation algorithms for e-commerce', *Intelligent Systems, IEEE*, vol. 22, no. 5, pp. 68-78.

Hung, S.-Y., Yen, D.C. & Wang, H.-Y. 2006, 'Applying data mining to telecom churn management', *Expert Systems with Applications*, vol. 31, no. 3, pp. 515-524.

Hwang, H., Jung, T. & Suh, E. 2004, 'An LTV model and customer segmentation based on customer value: A case study on the wireless telecommunication industry', *Expert Systems with Applications*, vol. 26, no. 2, pp. 181-188.

Hwang, Y. 2009, 'The impact of uncertainty avoidance, social norms and innovativeness on trust and ease of use in electronic customer relationship management', *Electronic Markets*, vol. 19, no. 2-3, pp. 89-98.

Hwang, Y. & Kim, D.J. 2007, 'Customer self-service systems: The effects of perceived Web quality with service contents on enjoyment, anxiety, and e-trust', *Decision Support Systems*, vol. 43, no. 3, pp. 746-760.

Iaquinta, L., Gentile, A.L., Lops, P., de Gemmis, M. & Semeraro, G. 2007, 'A hybrid content-collaborative recommender system integrated into an electronic performance support system ', *7th International Conference on Hybrid Intelligent Systems,* pp. 47-52.

Kamahara, J., Asakawa, T., Shimojo, S. & Miyahara, H. 2005, 'A community-based recommendation system to reveal unexpected interests', *11th IEEE International Conference on Multimedia Modelling,* pp. 433-438.

Kanellopoulos, D.N. 2008, 'An ontology-based system for intelligent matching of Needs for Group Package Tours', *International Journal of Digital Culture and Electronic Tourism*, vol. 1, no. 1, pp. 76-99.

Kass, G.V. 1980, 'An exploratory technique for investigating large quantities of categorical data', *Applied Statistics*, pp. 119-127.

Keramati, A. & Ardabili, S.M.S. 2011, 'Churn analysis for an Iranian mobile operator', *Telecommunications Policy*, vol. 35, no. 4, pp. 344-356.

Khandelwal, S. & Mathias, A. 2011, 'Using a 360° view of customers for segmentation', *Journal of Medical Marketing: Device, Diagnostic and Pharmaceutical Marketing*, vol. 11, no. 3, pp. 215-220.

Kim, B.M., Li, Q., Park, C.S., Kim, S.G. & Kim, J.Y. 2006, 'A new approach for combining content-based and collaborative filters', *Journal of Intelligent Information Systems*, vol. 27, no. 1, pp. 79-91.

Kim, C., Zhao, W. & Yang, K.H. 2008, 'An empirical study on the integrated framework of e-CRM in online shopping: Evaluating the relationships among perceived value, satisfaction, and trust based on customers' perspectives', *Journal of Electronic Commerce in Organizations (JECO)*, vol. 6, no. 3, pp. 1-19.

Kim, J., Wei, S. & Ruys, H. 2003, 'Segmenting the market of West Australian senior tourists using an artificial neural network', *Tourism Management*, vol. 24, no. 1, pp. 25-34.

Kim, N., Jung, K.-H., Kim, Y.S. & Lee, J. 2012, 'Uniformly subsampled ensemble (USE) for churn management: Theory and implementation', *Expert Systems with Applications*, vol. 39, no. 15, pp. 11839-11845.

Kim, S.-Y., Jung, T.-S., Suh, E.-H. & Hwang, H.-S. 2006, 'Customer segmentation and strategy development based on customer lifetime value: A case study', *Expert Systems with Applications*, vol. 31, no. 1, pp. 101-107.

Kim, Y. & Street, W.N. 2004, 'An intelligent system for customer targeting: A data mining approach', *Decision Support Systems*, vol. 37, no. 2, pp. 215-228.

Kincaid, J.W. 2003, *Customer Relationship Management: Getting it Right!*, Prentice Hall, Upper Saddle River, NJ.

Kritikou, Y., Demestichas, P., Adamopoulou, E., Demestichas, K., Theologou, M. & Paradia, M. 2008, 'User profile modeling in the context of web-based learning management systems', *Journal of Network and Computer Applications*, vol. 31, no. 4, pp. 603-627.

Kumar, D. 2010, 'Customer Appreciation of E-CRM based Website Services in Banks', *Chief Patron: Mrs. Aarathy Sampathy*, p. 5.

Larivière, B. & Van den Poel, D. 2004, 'The impact of product features and intermediaries on customer retention', *4th International Conference on Data Mining*, vol. 7, Wit Press, pp. 337-346.

Larivière, B. & Van den Poel, D. 2005, 'Predicting customer retention and profitability by using random forests and regression forests techniques', *Expert Systems with Applications*, vol. 29, no. 2, pp. 472-484.

Lefait, G. & Kechadi, T. 2010, 'Customer segmentation architecture based on clustering techniques ', *Fourth International Conference on Digital Society*, pp. 243-248.

Leung, C.W.-k., Chan, S.C.-f. & Chung, F.-l. 2006, 'A collaborative filtering framework based on fuzzy association rules and multiple-level similarity', *Knowledge and Information Systems*, vol. 10, no. 3, pp. 357-381.

Li, Y.-M. & Kao, C.-P. 2009, 'TREPPS: A trust-based recommender system for peer production services', *Expert Systems with Applications*, vol. 36, no. 2, pp. 3263-3277.

Ling, R. & Yen, D.C. 2001, 'Customer relationship management: An analysis framework and implementation strategies', *Journal of Computer Information Systems*, vol. 41, no. 3, pp. 82-97.

Liu, X. & Yang, J. 2012, 'Social buying metanetwork modelling and analysis', *International Journal of Services Technology and Management*, vol. 18, no. 1, pp. 46-60.

Lu, J. 2012, 'Personalized recommender systems for e-government and e-business intelligence ', *Practical Applications of Intelligent Systems*, vol. 124.

Lu, J., Ruan, D. & Zhang, G. (Eds.) 2006, *E-service Intelligence: Methodologies, Technologies and Applications*, Springer Verlag, Berlin-Heidelberg.

Lu, J., Shambour, Q., Xu, Y., Lin, Q. & Zhang, G. 2010, 'BizSeeker: A hybrid semantic recommendation system for personalized government-to-business e-services', *Internet Research*, vol. 20, no. 3, pp. 342-365.

Lu, J., Shambour, Q., Xu, Y., Lin, Q. & Zhang, G. 2013, 'A Web-based personalized business partner recommendation system using fuzzy semantic techniques ', *Computational Intelligence*, vol. 29, no. 1, pp. 37-69.

Mahdavi, I., Cho, N., Shirazi, B. & Sahebjamnia, N. 2008, 'Designing evolving user profile in e-CRM with dynamic clustering of Web documents', *Data & Knowledge Engineering*, vol. 65, no. 2, pp. 355-.

Manouselis, N. & Costopoulou, C. 2007, 'Analysis and classification of multi-criteria recommender systems ', *World Wide Web*, vol. 10, no. 4, pp. 415-441.

Mazzoni, C., Castaldi, L. & Addeo, F. 2007, 'Consumer behavior in the Italian mobile telecommunication market', *Telecommunications Policy*, vol. 31, no. 10–11, pp. 632-647.

Miguéis, V.L., Camanho, A. & Falcão e Cunha, J. 2012, 'Customer data mining for lifestyle segmentation', *Expert Systems with Applications*.

Miller Jr, R.G. 2011, *Survival Analysis*, Wiley-Interscience, New York.

Namvar, M., Gholamian, M.R. & KhakAbi, S. 2010, 'A two phase clustering method for intelligent customer segmentation', *2010 IEEE International Conference on Intelligent Systems, Modelling and Simulation (ISMS),* pp. 215-219.

Nazim Uddin, M., Shrestha, J. & Geun-Sik, J. 2009, 'Enhanced content-based filtering using diverse collaborative prediction for movie recommendation Content-Based Filtering Using Diverse Collaborative Prediction for Movie Recommendation', *First Asian Conference on Intelligent Information and Database Systems,* pp. 132-137.

Nemati, H.R., Barko, C.D. & Moosa, A. 2004, 'E-CRM analytics: The role of data integration', *Business Intelligence in the Digital Economy: Opportunities, Limitations and Risks*, p. 251.

Ngai, E. 2005, 'Customer relationship management research (1992-2002): An academic literature review and classification', *Marketing Intelligence & Planning*, vol. 23, no. 6, pp. 582-605.

Ngai, E.W., Xiu, L. & Chau, D. 2009, 'Application of data mining techniques in customer relationship management: A literature review and classification', *Expert Systems with Applications*, vol. 36, no. 2, pp. 2592-2602.

Nie, G., Rowe, W., Zhang, L., Tian, Y. & Shi, Y. 2011, 'Credit card churn forecasting by logistic regression and decision tree', *Expert Systems with Applications*, vol. 38, no. 12, pp. 15273-15285.

Nitzan, I. & Libai, B. 2011, 'Social effects on customer retention', *Journal of Marketing*, vol. 75, no. 6, pp. 24-38.

Owczarczuk, M. 2010, 'Churn models for prepaid customers in the cellular telecommunication industry using large data marts', *Expert Systems with Applications*, vol. 37, no. 6, pp. 4710-4712.

Pan, S.L. & Lee, J.-N. 2003, 'Using e-CRM for a unified view of the customer', *Communications of the ACM*, vol. 46, no. 4, pp. 95-99.

Papagelis, M. & Plexousakis, D. 2005, 'Qualitative analysis of user-based and item-based prediction algorithms for recommendation agents', *Engineering Applications of Artificial Intelligence*, vol. 18, no. 7, pp. 781-789.

Parasuraman, A. 1997, 'Reflections on gaining competitive advantage through customer value', *Journal of the Academy of Marketing Science*, vol. 25, no. 2, pp. 154-161.

Park, Y.-J. & Tuzhilin, A. 2008, 'The long tail of recommender systems and how to leverage it', paper presented to the *2008 ACM Conference on Recommender Systems*, 23-25 October, Lausanne, Switzerland.

Parvatiyar, A. & Sheth, J.N. 2001, 'Customer relationship management: Emerging practice, process, and discipline', *Journal of Economic and Social Research*, vol. 3, no. 2, pp. 1-34.

Pazzani, M. & Billsus, D. 2007, 'Content-based recommendation systems ', in P. Brusilovsky, A. Kobsa & W. Nejdl (Eds.), *The Adaptive Web*, vol. 4321, Springer, Berlin-Heidelberg, pp. 325-341.

Peacock, P.R. 1998, 'Data mining in marketing: Part 1', *Marketing Management*, vol. 6, no. 4, pp. 8-18.

Pesonen, J.A. 2012, 'Segmentation of rural tourists: Combining push and pull motivations ', *Tourism and Hospitality Management*, vol. 18, no. 1, pp. 69-82.

Piao, C.-H., Zhao, J. & Zheng, L.-J. 2009, 'Research on entropy-based collaborative filtering algorithm and personalized recommendation in e-commerce', *Service Oriented Computing and Applications*, vol. 3, no. 2, pp. 147-157.

Popović, D. & Bašić, B.D. 2009, 'Churn prediction model in retail banking using Fuzzy C-Means algorithm', *University of Zagreb, Faculty of Electrical Engineering and Computing*, vol. 33, pp. 243-247.

Porcel, C. & Herrera-Viedma, E. 2010, 'Dealing with incomplete information in a fuzzy linguistic recommender system to disseminate information in university digital libraries', *Knowledge-Based Systems*, vol. 23, no. 1, pp. 32-39.

Porcel, C., López-Herrera, A.G. & Herrera-Viedma, E. 2009, 'A recommender system for research resources based on fuzzy linguistic modeling', *Expert Systems with Applications*, vol. 36, no. 3, Part 1, pp. 5173-5183.

Ranjan, J. & Bhatnagar, V. 2011, 'Role of knowledge management and analytical CRM in business: Data mining based framework', *The Learning Organization,* vol. 18, no. 2, pp. 131-148.

Reinartz, W.J. & Kumar, V. 2003, 'The impact of customer relationship characteristics on profitable lifetime duration', *Journal of Marketing*, pp. 77-99.

Ricci, F., Rokach, L., Shapira, B. & Kantor, P. B. (Eds.) 2011, *Recommender Systems Handbook*, Springer, New York.

Rodríguez, R.M., Espinilla, M., Sánchez, P.J. & Martínez-López, L. 2010, 'Using linguistic incomplete preference relations to cold start recommendations', *Internet Research*, vol. 20, no. 3, pp. 296-315.

Romano, N.C. & Fjermestad, J.L. 2009, 'Preface to the focus theme on eCRM', *Electronic Markets*, vol. 19, no. 2, pp. 69-70.

Rust, R.T. & Zahorik, A.J. 1993, 'Customer satisfaction, customer retention, and market share', *Journal of Retailing*, vol. 69, no. 2, pp. 193-215.

Rygielski, C., Wang, J.-C. & Yen, D.C. 2002, 'Data mining techniques for customer relationship management', *Technology in Society*, vol. 24, no. 4, pp. 483-502.

Sarwar, B., Karypis, G., Konstan, J. & Riedl, J. 2001, 'Item-based collaborative filtering recommendation algorithms', *Proceedings of the 10th International Conference on World Wide Web*, ACM, pp. 285-295.

Sashi, C. 2012, 'Customer engagement, buyer-seller relationships, and social media', *Management Decision*, vol. 50, no. 2, pp. 253-272.

Schafer, J.B., Frankowski, D., Herlocker, J. & Sen, S. 2007, 'Collaborative Filtering Recommender Systems', in P. Brusilovsky, A. Kobsa & W. Nejdl (Eds.), *The Adaptive Web*, Springer, Berlin-Heidelberg, pp. 291-324.

Schafer, J.B., Konstan, J. & Riedi, J. 1999, 'Recommender systems in e-commerce', *Proceedings of the 1st ACM Conference on Electronic Commerce*, ACM, pp. 158-166.

Schapire, R.E. & Singer, Y. 1999, 'Improved boosting algorithms using confidence-rated predictions', *Machine Learning*, vol. 37, no. 3, pp. 297-336.

Schubert, P., Uwe, L. & Risch, D. 2006, 'Personalization beyond recommender systems ', in R. Suomi, R. Cabral, J.F. Hampe, A. Heikkilä, J. Järveläinen & E. Koskivaara (Eds.), *Project E-Society: Building Bricks*, vol. 226, Springer, pp. 126-139.

Shambour, Q. & Lu, J. 2011, 'Government-to-business personalized e-services using semantic-enhanced recommender system', in K. N. Andersen et al, (Eds.), *Electronic Government and the Information Systems Perspective*, Springer, New York, pp. 197-211.

Shambour, Q. & Lu, J. 2012, 'A trust-semantic fusion-based recommendation approach for e-business applications', *Decision Support Systems*.

Shardanand, U. & Maes, P. 1995, 'Social information filtering: Algorithms for automating “word of mouth”', in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM /Addison-Wesley, pp. 210-217.

Sharma, A. & Panigrahi, P.K. 2011, 'A neural network based approach for predicting customer churn in cellular network services', *International Journal of Computer Applications*, vol. 27, no. 11.

Shi, X., Ye, H. & Gong, S. 2008, 'A personalized recommender integrating item-based and user-based collaborative filtering', *IEEE International Seminar on Business and Information Management,* pp. 264-267.

Shih, M.-J., Liu, D.-R. & Hsu, M.-L. 2010, 'Discovering competitive intelligence by mining changes in patent trends', *Expert Systems with Applications*, vol. 37, no. 4, pp. 2882-2890.

Sinisalo, J., Salo, J., Karjaluoto, H. & Leppäniemi, M. 2007, 'Mobile customer relationship management: Underlying issues and challenges', *Business Process Management Journal*, vol. 13, no. 6, pp. 771-787.

Smith, K.A. & Gupta, J.N. 2003, *Neural Networks in Business: Techniques and Applications*, IRM Press, Hershey PA.

Stahl, F., Heitmann, M., Lehmann, D.R. & Neslin, S.A. 2012, 'The impact of brand equity on customer acquisition, retention, and profit margin', *Journal of Marketing*, vol. 76, no. 4, pp. 44-63.

Stone, M., Woodcock, N. & Wilson, M. 1996, 'Managing the change from marketing planning to customer relationship management', *Long Range Planning*, vol. 29, no. 5, pp. 675-683.

Su, X., Greiner, R., Khoshgoftaar, T.M. & Zhu, X. 2007, 'Hybrid collaborative filtering algorithms using a mixture of experts', *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*, pp. 645-649.

Sunikka, A. & Bragge, J. 2008, 'What, who and where: Insights into personalization', *Proceedings of the 41st Hawaii International Conference on System Sciences,* pp. 283.

Swift, R.S. 2001, *Accelerating Customer Relationships: Using CRM and Relationship Technologies*, Prentice Hall, Upper Saddle River, NJ.

Tabaei, Z. & Fathian, M. 2012, 'Developing online customer satisfaction strategic maps: with Iranian online retailing case studies', *International Journal of Electronic Customer Relationship Management*, vol. 6, no. 1, pp. 87-112.

Takács, G., Pilászy, I., Németh, B. & Tikk, D. 2009, 'Scalable collaborative filtering approaches for large recommender systems', *The Journal of Machine Learning Research*, vol. 10, pp. 623-656.

Takano, Y., Inoue, A., Kurosawa, T., Iwashita, M. & Nishimatsu, K. 2010, 'Customer segmentation in mobile carrier choice modeling', *9th International Conference on Computer and Information Science (ICIS),* pp. 111-116.

Tan, D.-W., Sim, Y.-W. & Yeoh, W. 2011, 'Applying feature selection methods to improve the predictive model of a direct marketing problem', in J. M. Zain, W. M. bt Wan Mohd & E. El-Qawasmeh (Eds.), *Software Engineering and Computer Systems*, Springer, New York, pp. 155-167.

Teichert, T., Shehu, E. & von Wartburg, I. 2008, 'Customer segmentation revisited: The case of the airline industry', *Transportation Research Part A: Policy and Practice*, vol. 42, no. 1, pp. 227-242.

Thorleuchter, D., Van den Poel, D. & Prinzie, A. 2010, 'Mining innovative ideas to support new product research and development ', in H. Locarek-Junge & C. Weihs (Eds.), *Classification as a Tool for Research*, Springer, Berlin-Heidelberg, pp. 587-594.

Thorleuchter, D., Van den Poel, D. & Prinzie, A. 2012, 'Analyzing existing customers' websites to improve the customer acquisition process as well as the profitability prediction in B-to-B marketing', *Expert Systems with Applications*, vol. 39, no. 3, pp. 2597-2605.

Tsai, C.-Y. & Chiu, C.-C. 2004, 'A purchase-based market segmentation methodology', *Expert Systems with Applications*, vol. 27, no. 2, pp. 265-276.

Turban, E. 2008, 'Information technology for management: Transforming organizations in the digital economy -'.

Turban, E., Sharda, R., Delen, D. & Efraim, T. 2007, *Decision Support and Business Intelligence Systems*, Pearson Education, Delhi, India.

Untema, W., Goossen, F., Frasincar, F., Hogenboom, F. & Hogenboom, F. 2010, 'Ontology-based news recommendation ', in *Proceedings of the 2010 EDBT/ICDT Workshops, New York, ACM*, pp. 16-23.

van Bentum, R. & Stone, M. 2005, 'Customer relationship management and the impact of corporate culture: A European study', *The Journal of Database Marketing & Customer Strategy Management*, vol. 13, no. 1, pp. 28-54.

Van den Poel, D. & Larivière, B. 2004, 'Customer attrition analysis for financial services using proportional hazard models', *European Journal of Operational Research*, vol. 157, no. 1, pp. 196-217.

Van Dyke, T.P., Nemati, H.R. & Barko, C.D., 'Leveraging customer data integration for effective E-CRM analytics'.

Verbeke, W., Dejaeger, K., Martens, D., Hur, J. & Baesens, B. 2012, 'New insights into churn prediction in the telecommunication sector: A profit driven data mining approach', *European Journal of Operational Research*, vol. 218, no. 1, pp. 211-229.

Verhoef, P.C. 2003, 'Understanding the effect of customer relationship management efforts on customer retention and customer share development', *Journal of Marketing*, pp. 30-45.

Wang, Z. & Lei, X. 2010, 'Study on customer retention under dynamic markets', in *Second International Conference on Networks Security Wireless Communications and Trusted Computing (NSWCTC),* IEEE, pp. 514-517.

Wang, Z., Sun, L., Zhu, W., Yang, S., Li, H. & Wu, D. 2013, 'Joint social and content recommendation for user-generated videos in online social network '.

Wei, K., Huang, J. & Fu, S. 2007, 'A survey of e-commerce recommender systems', in *2007 International Conference on Service Systems and Service Management,* pp. 1-5.

Weinstein, A. 2004, *Handbook of Market Segmentation*, Haworth Press, New York.

Wilson, R.M. & Gilligan, C. 2012, *Strategic Marketing Management*, (3rd edn.) Routledge, Oxford.

Wong, K.W., Zhou, S., Yang, Q. & Yeung, J.M.S. 2005, 'Mining customer value: From association rules to direct marketing', *Data Mining and Knowledge Discovery*, vol. 11, no. 1, pp. 57-79.

Woo, J.Y., Bae, S.M. & Park, S.C. 2005, 'Visualization method for customer targeting using customer map', *Expert Systems with Applications*, vol. 28, no. 4, pp. 763-772.

Wu, C.-H., Kao, S.-C., Su, Y.-Y. & Wu, C.-C. 2005, 'Targeting customers via discovery knowledge for the insurance industry', *Expert Systems with Applications*, vol. 29, no. 2, pp. 291-299.

Wu, D., Zhang, G., Lu, J. & Halang, W., 'A similarity measure on tree structured business data', in *Proceedings of the 23rd Australasian Conference on Information Systems: Location, location, location*, pp. 1-10.

Wu, L.-W. 2011, 'Satisfaction, inertia, and customer loyalty in the varying levels of the zone of tolerance and alternative attractiveness', *Journal of Services Marketing*, vol. 25, no. 5, pp. 310-322.

XiaoYan, S., HongWu, Y. & SongJie, G. 2008, 'A personalized recommender integrating item-based and user-based collaborative filtering ', in *International Seminar on*, *Business and Information Management,* vol. 1, pp. 264-267.

Xu, M. & Walton, J. 2005, 'Gaining customer knowledge through analytical CRM', *Industrial Management & Data Systems*, vol. 105, no. 7, pp. 955-971.

Yang, X., Zhang, G., Lu, J. & Ma, J. 2011, 'A kernel fuzzy c-means clustering-based fuzzy support vector machine algorithm for classification problems with outliers or noises', *IEEE Transactions on Fuzzy Systems*, vol. 19, no. 1, pp. 105-115.

Zhang, G. 2010, 'Customer Segmentation in customer relationship management based on data mining, *International Federation for Information Processing Digital Library*, vol. 207, no. 1.

Zhang, G. & Lu, J. 2003, 'An integrated group decision-making method dealing with fuzzy preferences for alternatives and individual judgments for selection criteria', *Group Decision and Negotiation*, vol. 12, no. 6,. pp. 501-515.

Zhang, G. & Lu, J. 2009, 'A linguistic intelligent user guide for method selection in multi-objective decision support systems', *Information Sciences*, vol. 179, no. 14, pp. 2299-2308.

Zhao, Y. & Yang, Q. 2011, 'Sampling method for imbalanced distribution in customer churn model', in *2011 International Conference on Computational and Information Sciences (ICCIS),* pp. 503-505.

Zorrilla, M.E., Mazón, J.N., Ferrández, Ó., Garrigós, I. & Florian, D. 2011, *Business Intelligence Applications and the Web: Models, Systems and Technologies*, IGI Global, Hershey, PA.