

Data Mining Analysis and Modeling for Marketing Based on Attributes of Customer Relationship

Xiaoshan Du

Data Mining Analysis and Modeling for Marketing Based on Attributes of Customer Relationship

Xiaoshan Du

School of Mathematics and Systems Engineering, Växjö University
SE-351 95 Växjö, Sweden
Supervisor: Joakim Nivre

Abstract. With the rapid growing marketing business, Data Mining technology is playing a more and more important role in the demands of analyzing and utilizing the large scale information gathered from customers. To predict the consequent business strategy by using Data Mining, the Customer Relationship Management (CRM) nowadays is required to evaluate the customer performance, discover the trends or patterns in customer behavior, and understand the factual value of their customers to their company. In this paper, we present an effective model to apply Data Mining to the CRM problem of categorizing the customers in marketing to search for potential clients based on their properties by (1) computing Distance in Cluster Analysis and Lift in Association Rules according to the Attributes of Customer Relationship (ACR) including Self-Reliance Index, Impact Index and Matrix for customer value, and (2) in the Data Mining modeling theory, constructing the Regression Model in the ACR and implementing the corresponding algorithm to mine the most profitable customer group.

1 Introduction	- 5 -
1.1 Background and motivation	- 5 -
1.2 Problem State	- 6 -
1.3 Thesis Outline	- 6 -
2 Data Mining Methods and Models	- 7 -
2.1 Data Mining Process.....	- 7 -
2.1.1 Data Preparation	- 7 -
2.1.2 Knowledge discovery in database.....	- 8 -
2.1.3 Model Explain and Estimate.....	- 9 -
2.2 Categorization of Data Mining Methods.....	- 9 -
2.2.1 Categorization Based on Mining Tasks.....	- 9 -
2.2.2 Categorization Based on Mining Objects	- 10 -
2.2.3 Categorization Based on Mining Techniques	- 10 -
2.3 Analysis and Modeling for Data Mining	- 12 -
2.3.1 Fundamentals of Model.....	- 12 -
2.3.2 Structures of Predictive model.....	- 13 -
2.3.3 Linear Regression Model.....	- 13 -
2.3.4 Predictive model for Classification.....	- 15 -
2.3.5 Stochastic Parts of Data Mining Model	- 15 -
2.3.6 Summary.....	- 16 -
2.4 Data Mining in Marketing.....	- 17 -
2.4.1 Application of Data Mining in Marketing.....	- 17 -
2.5 Application of Data Mining in CRM.....	- 19 -
2.5.1 Introduction of CRM.....	- 19 -
2.5.2 Concept of aCRM.....	- 20 -
2.6 Summary.....	- 22 -
3 Modeling Based on Attributes of Customer Relationship (ACR)-	23 -
3.1 Problem statement.....	- 23 -
3.1.1 Criterion of Customer Value.....	- 23 -
3.1.2 Discussion based on Customer Classification.....	- 24 -
3.2 Segmentation of Customer Value.....	- 24 -
3.3 Concept of Attributes of Customer Relationship (ARC).....	- 26 -
3.4 Dissimilarity in Cluster Algorithm	- 28 -
3.5 Lift in Association Rules.....	- 29 -
3.6 Search Reference Method for Network Relation.....	- 31 -
4 Implementation of Data Mining Model	- 32 -

4.1 Symbolic System.....	- 32 -
4.2 Estimate of Purchase Probability.....	- 32 -
4.3 Evaluate Customer Value	- 35 -
5 Experiments.....	- 37 -
5.1 Experimental Setup.....	- 37 -
5.1.1 Data Set.....	- 37 -
5.1.2 Main Functions Name and Description	- 39 -
5.1.3 Mining Algorithm.....	- 40 -
5.2 The evaluated results	- 40 -
5.3 Discussion.....	- 41 -
6 Conclusion and Future Works.....	- 43 -
7 Acknowledgement.....	- 44 -
8 References.....	- 45 -

1 Introduction

From this Chapter you will get a main idea of this thesis, include its background, why we choose topic, and which problems will be solved in this thesis, as well as how can we solved them.

1.1 Background and motivation

Traditional Large-scale sales pattern is the most familiar sales pattern for companies. Based on this Patten, companies usually aim at their produces, products and then give all the customers same sales promotion. However, this kind of sales promotions neglects the differences among customers. In most cases, these promotions cost a lot, but only get few real profits from customers. That means many promotions are waste.

In the meanwhile, data mining technologies become more and more popular in commercial terrain, such as in banking industry, insurance industry and retail trade. Data mining can solve many typical commercial problems, such as Database Marketing, Customer Segmentation and Classification, Profile Analysis, Cross-selling, Churn Analysis, Credit Scoring, Fraud Detection, and so on.

Since those data mining technologies appeared, companies have changed their sales target from products to customers. How to classify customers? How to find out the common character of customers from database? How to dig up the potential customers? How to find out the most valuable customers? These kinds of questions become the most popular data mining applications in marketing.

Nevertheless, the recent customer relation analyses have some serious drawbacks. The most important one is that based on those analyses company usually consider the customer as an isolated object and having value only when he/she deals with this company. Neglect the network value of each customer and the value from potential purchase probability.

In this paper, which is based on the application of data mining in marketing and recent research result of Customer Relationship Management, we would like to try to use new visual angle to improve these drawbacks.

1.2 Problem State

In this paper, two main problems will be solved. The first one is how to generally classify the customers by their value? The second one is how data mining techniques can be used to estimate the value of a customer given a database containing information about his/her name, age, profession, etc? This is the most important one.

In this paper the value of a customer should be considered as how much profit he/she can bring to the company. This value should be calculated as a numerical scale. It can be primarily defined like this: $V = rP_m - rP_n - C$. Where “r” is the profit brought from a specific produce when customer buys it. “ P_m ” is the customer’s purchase probability when there is a sales promotion, while “ P_n ” is the customer’s purchase probability when there is no sales promotion. “C” is the cost of the sales promotion.

To get the value of a customer “V”, the customer’s purchase probability must be calculated at first. In this paper, Attribute of Customers Relation (ACR) will be defined to represent the probability

At last, the new model and algorithms should be experimented and the result should be evaluated.

1.3 Thesis Outline

The first Chapter is an introduction of the paper, including background, motivation and thesis outline.

In the second Chapter, we are going to introduce some related data mining methods and models. Furthermore, give some introductions of data mining in marketing. Give a short introduction of CRM, including its categorization and applications of data mining in CRM.

In third Chapter, we would like to focus on modeling based on ACR. We are going to define some new concepts and algorithms in this Chapter, in order to get ready for the following chapter as well.

In the fourth Chapter, the implementation of our data mining model will be defined. After that, the fifth Chapter will give out the general idea of experiments.

In addition, we put conclusion and future works, acknowledgement, and references at the end of this paper.

2 Data Mining Methods and Models

First of all, we'd like to give out the definition of Data Mining. Data Mining is the process of identifying hidden patterns and relationships within data. [11] In another word, Data Mining is the process finding hidden information in a database. [12] From the definition of data mining, we learn that it is a kind of technologies that can help us know the useful things hidden in the data. Therefore, data mining should be an interesting work.

2.1 Data Mining Process

There are three main steps of data mining process.

2.1.1 Data Preparation

In the whole data mining process, data preparation is somehow a significant process. Some book says that if data mining is considered as a process then? Data preparation is at the heart of this process. However, nowadays databases are highly susceptible to noise, missing and inconsistent data. So preprocessing data improve the efficiency and ease of the data mining process, this becomes an important problem. Several consulting firms, such as IBM, have approved that data preparation costs 50%~ 80% resource of the whole data mining process. From this view, we really need to pay attention to data preparation.

There are three data preprocessing techniques should be considered in data mining:

1) Data cleaning

a) Inconsistent data:

Not all the data we get is "clean". For example, a list of Nationality may have the values of "China", "P.R.China", and "Mainland China". These values refer to the same country, but are not known by the computer. Therefore, this is a consistency problem.

b) Missing values

Data from a company's database often contains missing values. Sometimes the approaches require rows of data to be complete in order to mine them, but the database may contain several attributes

with missing values. If too many values are missing in a data set, it becomes hard to gather useful information from this data.

c) Noisy data

Noise is a random error or variance in a measured variable.

2) Data integration

Usually the data analysis task will involve data integration. It combines data from multiple sources into a coherent data store. Those sources include multiple database or flat files. Several issues should be considered during data integration, such as schema integration, correlation analysis for detecting redundancy, and detection and resolution of data value conflicts. Careful integration of the data can help improve the accuracy and speed of the mining process.

3) Data reduction

If you select data from a data warehouse, you probably find the data set is huge. Data reduction techniques can be applied to obtain a reduced representation of the data set. Mining on reduced data set should be more efficient yet produce the same analytical results. It includes several strategies, such as data cube aggregation, dimension reduction, data compression, numerosity reduction, and discretization and concept hierarchy generation.

2.1.2 Knowledge discovery in database

As a core data mining techniques, knowledge and information discovery has several main components:

1) Determine the type of data mining tasks

We must confirm that the functions and tasks to be achieved by recent system belong to which kind of classification or clustering.

2) Choose suitable technologies for data mining

We can choose the appropriate data mining technologies based on the tasks we have confirmed. Such as, classification model often use learning neural network or decision tree to realize; while clustering usually use clustering analysis algorithms to realize; association rules often use association and sequence discovery to realize.

3) Choose the algorithms

Based on the technologies have been chosen, we can select a specific algorithm. Furthermore, a new efficient algorithm can be designed by the specific mining tasks. To choose data mining algorithms, we should determine the hidden pattern in selecting data.

4) Mining data

We are supposed to use the selected algorithms or algorithms portfolio to do repeated and iterative searching. Extract the hidden and innovative patterns from data set.

2.1.3 Model Explain and Estimate

Explain and estimate the patterns got from data mining, get the useful knowledge. For instance, remove some irrespective and redundant patterns, after filtration the information should be presented to customers; Use visualization technology to express the meaningful model, in order to translate it into understandable language for users. A good application of data mining can change primal data to more compact and easily understand form and this form can be defined definitely. It also includes solving the potential conflict between mining results and previous knowledge, and using statistical methods to evaluate the current model, in order to decide whether it is necessary to repeat the previous work to get the best and suitable model.

The information achieved by data mining can be used later to explain current or historical phenomenon, predict the future, and help decision-makers make policy from the existed facts.

2.2 Categorization of Data Mining Methods

There are several data mining methods, and there are some different ways to classify them as well.

2.2.1 Categorization Based on Mining Tasks

Based on the different mining tasks, we can categorize data mining methods as classification, clustering, regression, association rules, sequence discovery, prediction, and so on. [13]

1) Classification

Classification maps data into predefined group or classes. Because the classes are determined before examining the data, classification is often considered as supervised learning. Classification algorithms require that the classes be defined based on data attribute values. They often describe these classes by looking at the characteristics of data which are already known to belong to the classes.

2) Clustering

Clustering is similar to classification; the difference is the groups are not predefined. It is alternatively referred to as unsupervised learning. It is usually achieved by determining the similarity among the predefined attributes of the data. The most similar data are grouped into clusters.

3) Regression

Regression is used to map a data item to a real valued prediction variable. Regression assumes that the target data fit into some known type of functions and then determines the best function of this type. A simple example of regression is the standard linear regression.

4) Association rules

Association rules alternatively referred to as affinity analysis. An association rule is a model that identifies specific types of data associations. They are usually used in the retail sales community to identify items which are often purchased together.

5) Sequence discovery

It is used to determine sequential patterns in data. Those patterns are based on a time sequence of actions. The relationship of those patterns is based on time, and they are similar to associations.

6) Prediction

Based on past and current data, many real-world data mining applications can be considered as predicting future data states. Prediction is viewed as a type of classification. The difference is that prediction is predicting a future state rather than a current state. Prediction applications include flooding, speech recognition, machine learning, and pattern recognition.

2.2.2 Categorization Based on Mining Objects

If we categorize based on mining objects, the data mining methods can be divided into based on Relational Database, Object Oriented Database, Spatial Database, Text Data Sources, Temporal Database, Multimedia Database, Heterogeneous Database, and Web source.

2.2.3 Categorization Based on Mining Techniques

There are many different techniques used to achieve DM tasks, so we can primarily divided DM methods into Machine learning methods, Statistical

methods, and Neural Networks methods. And then subdivision them as follows:[13]

1) Machine Learning methods

a) Decision Trees

Decision tree is one of the most popular classification algorithms in current Machine Learning. A decision tree is usually used in classification, clustering and prediction tasks as a predictive modeling technique. They are ideal methods for making financial decisions where lots of complex information needs to be taken into account.

b) Genetic Algorithms

Genetic algorithms are a method of breeding computer solutions to optimization problems by simulated evolution. The processes based on crossover and mutation repeatedly applied to a population of binary strings. Time after time, the better fit individuals and average individuals are created, and a good solution to the problem is found. Genetic algorithms usually are used to predict and used to replace the missing attributes. It means when there are some attributes missed, we can analyses its specimens using genetic algorithms to get the possible value and replace the missing one.

2) Statistical Methods

a) Statistical Analysis

Statistical analysis is one of the most mature and proven data mining methods. The key of this method is construct appropriate statistical models and mathematical models to interpret the data models. This approach requires the user has abundant knowledge of this field. Generally, statistical analysis has two steps: firstly, the user chooses the appropriate data from data warehouse. Secondly, the user uses visualization functions and analysis functions provided by statistical analysis tools, in order to find the relationship between the data and statistical models and constructed mathematical models to interpret data. The second step needs to be repeated and continuous carefully.

b) Cluster Analysis

Cluster analysis classification is based on their characteristics in order to discover typical pattern. When the data that should be analyzed miss the describe information or can not be organized into any classification model, the use of cluster analysis will be automatically divided into categories according to certain

characteristics. The substance of cluster analysis is an overall optimal problem, commonly used in the market subdivision, customer orientation, performance evaluation, and other aspects.

3) Neural Networks

Neural Networks is an information processing system that consists of a graph representing the processing system as well as various algorithms that access this graph. It is structured as a graph with many nodes and arcs between them. It can be considered as a directed graph with input, output and internal nodes.

After accept a variety of input, every nerve calculates the total input value and then uses filtering mechanisms to compare the total input, in order to determine its own output value. When change the link weight between two nerves or two layers, neural network is on a study or "training." After "training" the neural network can be used to predict the likely outcome of the existing cases, the analysis could also be applied in customer relations, or other fields.

2.3 Analysis and Modeling for Data Mining

In the previous sections we briefly described the process of data mining and data analysis methods. As some of design modeling will involved in this paper, this subsection will be used to give more in-depth discussion about the concept of modeling, and inspect several major types of models for data mining, in order to provide a theoretical basis for the follow-up sections.

Model is high-level data sets, and a global summary. It usually treats the integer through a large group of samples. Models can be descriptive, which induce data using a concise manner; they can also be rational, which allowing making some certain inferences for the data integer or future data. In this paper, we will integrate several forms of theoretical models as a basis, such as Linear Regression Models, Hybrid Models, and Markov Model.

In modeling, it must be noted that when summarized data we should take account into factors, such as time factor. If the used method has certain limitations, it may lead to a distortion in the model. Therefore, a model is good or bad, that needs to be tested by reality.

2.3.1 Fundamentals of Model

Models are abstract description of the real world process. We start from the simplest model to expatiate its meaning. For example, a simple model might take the form of $Y = aX + c$, where Y and X are variables and a and c are parameters of the model (constants determined during the course of the data mining exercise). Here we would like to say that the functional form of the model is linear, since Y is a linear function of X . We can set $a=1$, $c=2$ to simplify this model. More generally, we also can improve this model to $Y = aX + c + e$, where e is a random variable component of the mapping from X to Y . Normally, a , c are considered as the parameters of the model, and we often use the notation θ to express a generic parameter or a parameters set, where $\theta = \{a, c\}$. Given the form of structure of a model, we choose the appropriate values for its parameters. This is achieved by minimizing or maximizing an appropriate score functions measuring the fit the model to the data.

Modeling is to discuss issues from a theoretical perspective, but we must be recognized that the theory and practical phenomenon are always different. We should recognize that "all models are not perfect, but some are useful." For example, we may assume that the existence of a linear model abstracts a process, but the reality is that there are always some nonlinear roles. This is what the models can not take into account. What we want is to find a model which can summarize the main features of a process.

The following are examples of several common model structures [16].

2.3.2 Structures of Predictive model

In a predictive model, the variable is expressed by a function of the other variables. We take

$Y = aX + c + e$ as an example. The values of the response variable Y are predicted from given values of predictor variables X . Generally, the response variable in predictive models is often denoted by Y , and the p predictor variables are denoted by x_1, x_2, \dots, x_p . The model will be yield predictions, $y = f(x_1, x_2, \dots, x_p; \theta)$, where y is the prediction of the model and θ represents the parameters of the model structure. When Y is quantitative, this task of estimating a mapping from p -dimensional X to Y is known as Regression. When Y is categorical, the task of learning a mapping from X to Y is called Classification learning. Both of them can be referred as Function Approximation problems in which we are learning a mapping from a p -dimensional variable X to Y .

2.3.3 Linear Regression Model

Next, we are going to discuss a linear predictive model. Its structure is simple and easy to understand. Its response variable is a linear function of the predictor variables:

$$Y = a_0 + \sum_{j=1}^p a_j X_j$$

Where $\theta = \{a_0, a_1, \dots, a_p\}$. We note that the model is purely empirical, so that the existence of a well fitting and highly predictive model does not imply any causal relationship.

We can retain the additive nature of the model, while generalizing beyond linear functions of the predictor variables:

$$Y = a_0 + \sum_{j=1}^p a_j f_j(X_j)$$

Where the f_j functions are smooth functions of X_j . f_j could be log, square-root, or related transformations of original X variables. The model assumes that the dependent variable Y depends on the independent variables X of the model in an additive fashion. This may be a strong assumption in practice, but it will lead to a model in which it may be easy to interpret the contribution of each individual X variable. This can be found in our first model in this paper.

Furthermore, we can generalize this linear model structure to allow general polynomials in the Xs with cross-product terms to allow interaction among the X_j in the model. Note that by allowing models with higher order terms and interactions between the components of X we can estimate a more complex surface than a simple linear model. However, we note that as the dimensionality p increases, the number of possible interaction terms X in the model increase as a combinatorial function of p . The interpretation and understanding of such a model makes the problem more difficult, moreover it will become more difficult as p increasing. But the response variables compared to the parameters of model are still linear. So the estimate of the parameters will become much easier.

The generalization to polynomials is called the complexity of the model. The more complex models contain the simpler models as special cases. For example, the 1st order $a_1 X_1 + a_0$ model can be considered as a special case of the 2nd order polynomial model $a_2 X_1^2 + a_1 X_1 + a_0$ by set $a_2 = 0$. So it is clear that a complex model can always fit the observed data at least as well as any simpler model can. This raises the complexity of how we should choose one

model rather than others when the complexity of each is different. There is always a question. We may want a model which is closest to some hypothesis, a model that captures the main features of the data without being too complicated, and so on. We must know how to find a model balance both precision and efficiency.

To generalize a linear structure we can transform the predictor variables X . We can also transform the response variable. A good way for further generalization is to assume that Y is locally linear in the X 's, with a different local dependence various regions of the X , that is a piecewise linear model. The piecewise linear model is a good way to solve how we can build relatively complex models for nonlinear phenomena by piecing together simple components. This is also why we use it as an important fundamental of our model.

2.3.4 Predictive model for Classification

By now we have discussed about predictive models which the variable is predicted, where Y is quantitative. Now we consider the case of a categorical variable Y , which only take a few possible categorical values. This is a classification problem. The aim is to assign a new object to its correct Y category on the basis of its observance X value.

In classification problem what we need to do is to set different data to different categories. A classic approach is to use a linear hyperplane in the p -dimensional X space to define a decision boundary between two classes. The model partitions the X space into disjoint decision regions, where the decision regions are separated by linear boundaries. We can use the higher-order polynomial terms, yielding smooth polynomial decision boundaries, to get a more complex model.

2.3.5 Stochastic Parts of Data Mining Model

In the previous discussion, we briefly referred to stochastic parts of data mining model. Now we are going to talk about the functions of stochastic parts. It is very hard to find a perfect functional relationship between the predictor variables X and the response variable Y . The fact is for any given predictor variables x , more than one value of Y can be observed. The distribution of the values Y at each value of X represents an aspect of variation. The variation can be divided into two categories:

- 1) Unexplainable variation

This kind of variation will be reduced by decreasing the complexity of the model. We call it unexplainable or *nonsystematic* or *random* parts of the variation.

2) Explainable variation

This kind of variation is also called *Systematic* variation. The variation in Y can be explained by the X variables.

For example: We have mentioned the regression modeling before. We can extend it to include a stochastic part. We assume that for each X we can observe a particular Y, but the Y is added some noise. So the relationship between X and Y become:

$$y = g(x, \theta) + e$$

Where $g(x, \theta)$ is a deterministic function of X, while e is usually set to zero and assumed to be a random variable with constant variance (σ^2), which is independent from X. The random term e reflects the noise in the measurement process. More generally, e reflects the fact that there are hidden variables. Their affections on Y can not be expressed by the deterministic function of X.

Raising the stochastic parts gives a good way to improve the accuracy of models.

2.3.6 Summary

In the discussion so far, we briefly introduced the modeling theory in data mining. The core principle is incorporating the relatively simple models into complex model, or using different methods generalize the simple model to complex model. According to the modeling theory, none of the models for data mining is absolutely isolated, but interconnected by a variety of relations. This is not hard to understand. As a complex function is incorporated by several basic functions, each model for data mining is a generalization of other models, or a special case of other model. As we all know, in data mining the key of establishing an effective model is to select the best model form in order to solve current problem. This is not only the process of selecting a model, dealing with data, and giving out the results. The process of modeling needs continuous fitting work, and repeated improving work. This process is kind of endless.

2.4 Data Mining in Marketing

In the recent decades, the development of information and communications technologies injects new vitality for enterprise marketing. For example, bar code technology and the emergence of online stores greatly both enhance the efficiency of the enterprise. Resulting, company managers are beginning to face the enormous data. The data is increasing at a very rapid pace, probably 1000 times than five years ago. However, the data and business profits are not directly proportional. Unfortunately, the human brain can not handle so much data. In the meanwhile, data mining technology becomes very mature in theory. Thesis the technology-oriented applications for enterprise decision makers with a new perspective to look at market. Those advanced technologies let enterprises obtain a lot of resources from different channels, and use those effective tools to translate data into unlimited opportunities.

2.4.1 Application of Data Mining in Marketing

DM technology in the marketing is a relatively universal application. Such applications are referred to a Boundary Science, because it sets a variety of scientific theories in all. First, two basic disciplines: Information Technology and Marketing. Another very important basis is Statistics. In addition, it relates to the psychology and sociology as well. The charm of this area is just about the wide scope of disciplines study.

Generally speaking, through the collection, processing and disposal of the large amount of information involving consumer behavior, identify the interest of specific consumer groups or individual, consumption habits, consumer preferences and demand, moreover infer corresponding consumption group and the next group or individual consumption behavior, then based on them sale produces to the identification consumer groups for a specific content-oriented marketing. This is the basic idea.

As automation is popular in all the industry operate processes, enterprises have a lot of operational data. The data are not collected for the purpose of analysis, but come from commercial operation. Analysis of these data does not aim at studying it, but for giving business decision-maker the real valued information, in order to get profits. Commercial information comes from the market through various channels. For example, purchasing process by credit card, we can collect the customer's consumption data, such as time, place, interesting goods or services interested, willing price and the level of reception capacity; when buying a brand of cosmetics or filling in a member form can collect customer purchase trends and frequency. In addition,

enterprises can also buy a variety of customer information from other consulting firms.

Marketing based on data mining usually can give the customer sales promotion according to his convenient purchase records. It should be emphasized data mining is application-oriented. There are several typical applications in banking, insurance, traffic-system, retail and such kind of commercial field. Generally speaking, the problems that can be solved by data mining technologies include: analysis of market, such as Database Marketing, Customer Segmentation & classification, Profile Analysis and Cross-selling. And they are also used for Churn Analysis, Credit Scoring and Fraud Detection. Fig 2.1 shows us the relation between application and data mining techniques clearly and completely.

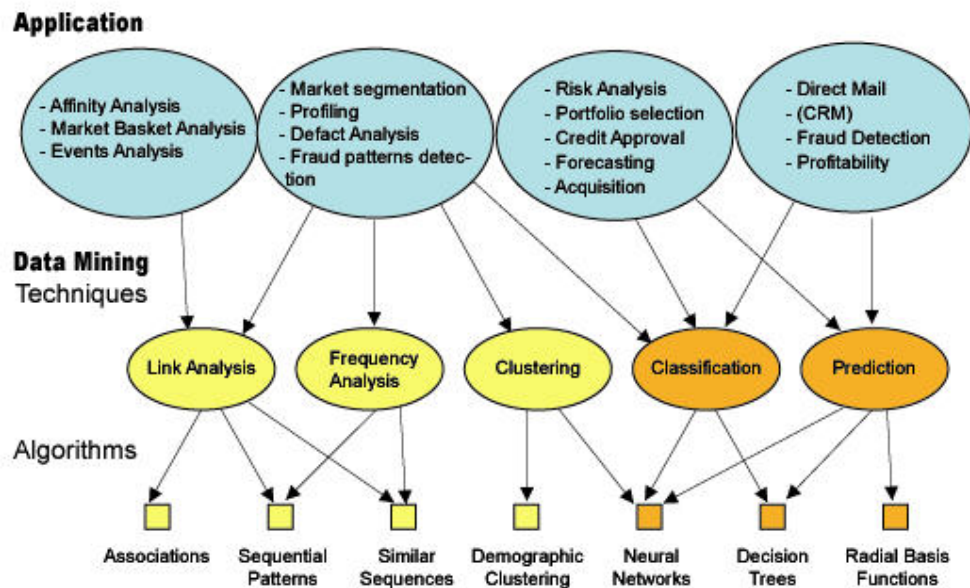


Fig.2.1 Application of data mining for marketing

The basic process of data mining in marketing show as follows: (Fig.2.2 shows the principle of data mining application in marketing)

- a) Prepare primitive data. It includes individual character information (such as age, gender, hobby, background, profession, address, postcode, and income), the previous purchase experience, and

the relationship within customers. The preprocessing of primitive data is very important for selecting potential customers.

b) Establish a certain model. This model may utilize plenty of traditional data mining technologies and many technologies from other related subjects. However, the problem which those technologies should solve is seeking for the best or acceptable market plan, within limited data source, limited time, and limited expense. The three limits are the fundamentality of modeling algorithm.

c) At last, according to the model, utilize testing data to get each pattern or parameter. Ultimately, use this model to select customers and decide marketing plan.

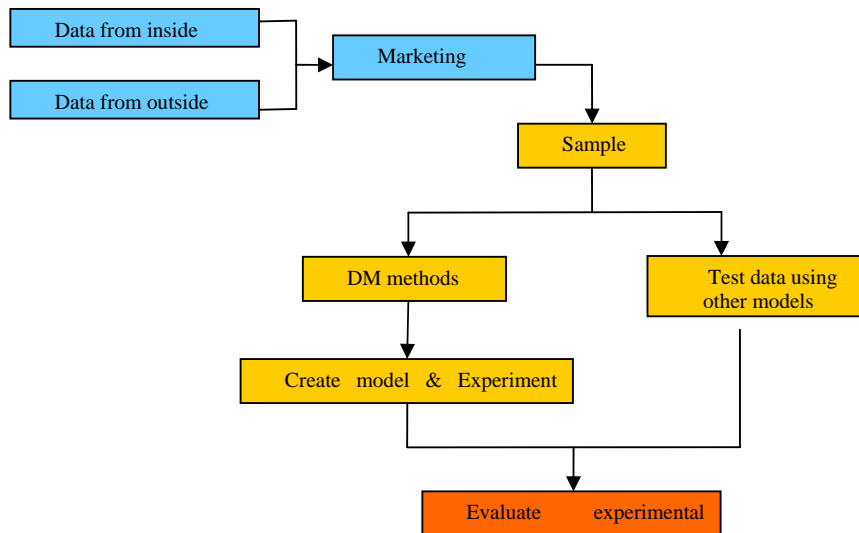


Fig. 2.2. Schematics for DM application in Marketing

2.5 Application of Data Mining in CRM

In this subsection, we give out the introduce of CRM. It includes both oCRM and aCRM. While what we focus on is aCRM.

2.5.1 Introduction of CRM

Customer Relationship Management (CRM) is a strategy to acquire new customers, to retain them and to recover them if they defected. [15] In the recent days, individual customer has brought pressure of change in marketing

practices. One of the main goals of CRM is: Generating additional product benefits by means of communications and services which are designed and delivered to match the individual needs of customers.

There are two kinds of CRM [15]:

1) Operational CRM (oCRM) activity is implemented in the enterprise processes: sales, marketing or service. oCRM involves all activities about the direct customer contact.

2) Analytical CRM (aCRM) provides all components to analyze customer characteristics in order to accomplish oCRM activities, with respect to the customers' needs and expectations. There, the idealistic goal is to provide all information necessary to create a tailored cross-channel dialogue with each single customer on the basis of his or her actual reactions.

We'd better to look at CRM which includes oCRM and aCRM as a cross enterprise process, in order to achieve the goal to show merely one company mapping to a customer. Marketing, sales and service departments have to coordinate their responsibilities, activities, information systems and data. The Fig.2.3 shows the cross functional process of CRM.

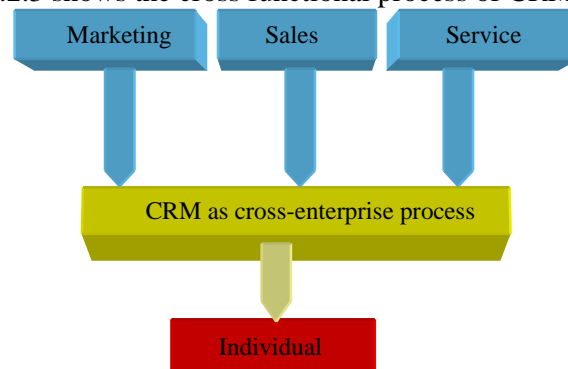


Fig.2.3. CRM as cross functional process

2.5.2 Concept of aCRM

Data mining in the CRM application is primarily embodied in: customer classification, analysis of customer relations, market orientation, and establishing predictive models. That is CRM analysis module. Fig.2.4 shows the structure of aCRM.

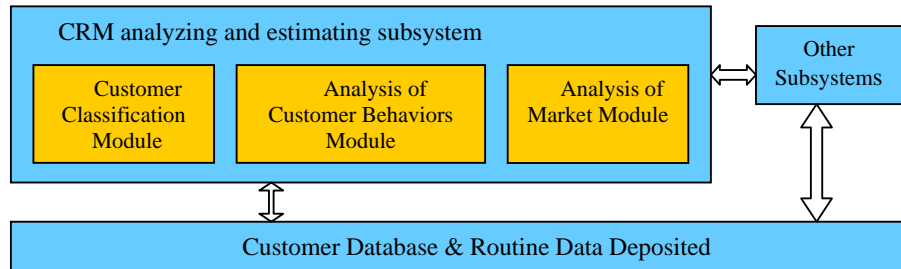


Fig.2.4. Structure of aCRM

1) Customer classification module

This module classifies customers by customers' value and set relevant customer level. It can lead the enterprises to distribute the resources of market, sales, and services to the valuable customers. The enterprises can aim at the valuable customers giving them special sales promotions, and providing more personalized services, in order to get the maximum returns by least investment. This is what we are interested in.

Generally, classification can be considered through three aspects:

a). Exterior attributes

These attributes include the customer's regional distributing, holding produces, and organizational attribute (Customer can be divided to enterprise customer, government customer and individual customer). Usually, this kind of classification is simple and intuitionist. The data is also very easy to get. However, this kind of classification is general. We still don't know that those high valuable customers within in each classification are. What we know is only which category of customers have more purchase power than other category.

b). Inherent attributes

These attributes include age, gender, interest, income, credit, and so on.

c). Classification of consume behaviors

The analysis of consume behavior is usually considered as three aspect. That is RFM: recent consume, frequency of consume, and magnitude of consume. All of these data can be achieved from the accounting system. However, this kind of classification can only be used on existing customers. Since

there is no consume, the potential customers can not be classified by it.

CRM can divide customers into many categories. The customers within one category have the same attributes, while the attributes of customers from different category are certainly different. We can provide different service to customers from different category, in order to enhance the satisfaction. We can easily find the advantages of classification. Even a simple classification can bring the enterprise a satisfying result.

2) Analysis of customer behaviors module

This module mainly process the analysis of customer's satisfaction, loyalty of customer, correspondence of customer, prediction of customer's lost, and cross sales.

3) Analysis of market module

Market is the main goal of enterprise. The enterprise can win the competition only by handling the trend of market. Prediction of market trend includes analyzing and predicting the development of produces, predicting the different consume trend of customer from different region, and predicting the changes appearing as the season's change.

2.6 Summary

While oCRM and aCRM are accepted by enterprises, data mining technologies also get a more important role within aCRM. The key of aCRM is how to find out the most valuable customer and customer group for enterprises by data mining technologies. This is what we are going to solve in the following chapters.

3 Modeling Based on Attributes of Customer Relationship (ACR)

In this chapter, we are going to let you know why we use Attributes of Customer Relationship (ACR) and how does it look like. This chapter is basic information for Chapter 4.

3.1 Problem statement

In this subsection, we focus give a general idea about customer value. This will help us define how to find “big customer” or “more valuable customer”.

3.1.1 Criterion of Customer Value

Application of data mining in CRM is helping enterprise to dig out the most valuable customers. Many managers and marketing decision-makers usually focus on the income-flux brought to enterprises by customers. Commonly, consume quantum is the criterion of customer value. It means that customer with high consume quantum will get more attentions. Moreover, they will get more favorable price and better service.

Nevertheless, this criterion of customer value is doubted lately. More and more companies find that many big customers are not large profitless customers, furthermore, sometimes they bring the company negative profit. The reason is that company didn't use a reasonable criterion to estimate customer value. Using consume quantum as the criterion, the company potentially thought the more consume quantum is, the more value customer have. As a result, they give too many services to the “big customers” which did not get a good result.

Now we give out two customers who have the same consume quantum, after comparing you will find the problem of the criterion mentioned above. Assuming we have two big customers, named A and B. A and B look similar, because they brought almost same purchase to the company in last 12 months, furthermore their consume trends are similar. If we only use income-flux as the criterion of customer value, A and B should have the same value. But the probable fact is A is a loyal customer of this company for years. A do not only buy produces from the company, he also recommend them to his colleagues

and friends. After he bought one produce, his friends maybe buy ten same produces. B as another big customer buys produces, but he is always alone and rarely recommends others. Thus, we can find out the value of A and B is clearly different, but this difference can not be judged by the simple criterion mentioned above.

Via this example, we can find the limit of unitary criterion. Customer A certainly should get more attention from the company. So we should find the more comprehensive criterion of customer value.

3.1.2 Discussion based on Customer Classification

If we are not going to use income-flux as the criterion to evaluate customers, we must find out another principle for customer classification. Customer classification is the customer set partitioned by any attribute of customer. We referred to much information about customer classification. Furthermore, we find that there are many questions when we make the classification:

- 1) What is the difference between customers?
- 2) What is the most comprehensive judgment of customer value?
- 3) What is the difference of the customers who have the same purchase record?
- 4) Which factor will impact the loyalty of customer?
- 5) What fashion does customer classification have? Is there any other classification variable besides the profit brought by customer to company.
- 6) Is classification unitive? Thus, once a customer is classified to a category, all the departments of this company should have the same classification behavior, right?

In order to consider all the questions above, we'd better find the appropriate criterion of classification, and establish appropriate model to simulate customer group. Furthermore, use data mining technology to partition certain customer group. And then use relative marketing strategy on the most valuable customers.

3.2 Segmentation of Customer Value

Traditional concept of customer classification in CRM usually just considers customers as many individual units or an object. The methods we mentioned before are all based on the purchase probability of potential customer and the profits from which company can get. It assumes customers as a large object.

Every individual customer is isolation and without any connection. It merely considers the profits which customers bring to enterprise. This is what we call self-value. However, customers and enterprises are all in the certain social relationship network. When the customer is going to make the decision of a purchase, the decision is not only depended on his own interest, but also impacted by the opinions from others. In the meanwhile, he can impact on the probability of others of purchase the produce. This is what we call network-value. When two customers have the same self-value, we should give more sales promotions to the one with higher network-value.

In this paper, we are going to add the concept of network-value to the customer considering system of CRM. We divide the large object of customer, in order to do some deeper researches on the complex relationship of customer. This will give CRM a new visual angle of customer classification.

First of all according to the concept of network-value, we primly divide customer value into two dimensionalities. They are customer's self-value and network-value. Furthermore, we give each dimensionality two levels: high level and low level. By now we can divide customers into four groups. This subdivision will be expressed as a matrix, which is Matrix for Customer Value (Fig.5).

Network-value	III	IV
	I	II
	Self-value	

Fig 3.1. Matrix for Customer Value

According to this Fig 3.1, we can find that IV category of customer is the one enterprise should focus on, while I category is opposite. Meanwhile, II and III category of customer should be discussed. Category II is the favorite customer to traditional CRM. They often can get the best prices and services from enterprises. However, category III is usually neglected by traditional CRM customer classification. They rarely get the attentions from enterprise

because of the low self-value. After a long time, this category of customer will lose confidence of the enterprise, so they leave. According to their high network-value, this will cause a large loss for the enterprise.

After the analysis above, we think that evaluating both self-value and network-value of customer carefully is significant to the marketing of enterprises. How to evaluate customers self-value? How to get the network-value by data mining tools? To solve these problems is an interesting work and we will talk about them later.

Thereinafter we are going to do more analysis of the Matrix for Customer Value.

3.3 Concept of Attributes of Customer Relationship (ARC)

As we talked before, convenient sales market has the limit of information communication. The enterprise only considers customer as an isolative unit. As the development of information and network technologies, a complex customer network is formed. According to the matrix from last subsection, self-value is expressed by the purchase after sales promotions. Network-value is expressed by the impact of the purchase to other customer.

We must mention that customer network is an unordered network. The impact is mutual. While the customer affects to others, he also get the impact from others. For example, before many purchases the customers will ask for comments from others who have bought the produce or they will search from websites. And then they will make their own decisions. The same after their purchases they may give out their comments to others via internet.

Within this process, some customers have more self-leading. They mostly make decisions by their own interests. However, there are some others may change their thought after reading comments. At the same time, some of the customers would like to give their comments of produces to others, but some others maybe not like to. So we think the network-value of a customer should be referred to the impact on others. Thus, others' attributes should be used as the criterion of a customer's network-value.

According to this kind of complex situations, traditional CRM analytical strategy looks powerless. We must use a new strategy to evaluate customer value. So we subdivide customers based on the matrix of customer value. We change the dimensionality into self-value and impact-value. Like Fig.3.2:

Impact-value	III Bidirectional	IV Consulting
	I Diving	II Self-centered
	Self-value	

Fig.3.2. Matrix for subdivision of Customer Value

I. We call them Diving customer. This group of customers would like to search for the comments of a produce before they making their purchase decisions. And they may change their mind easily according to other's impact. However, they don't like to give out their comments to others.

II. We call them Self-centered customer. This group of customers has more definite idea. They usually make purchase decisions by themselves and they are not going to make others listen to their opinion.

III. We call them Bidirectional customer. This group of customers would like to listen to others' opinions and also like to give out their own comments.

IV. We call them Consulting customer. This group of customers is very suit to do consulting to potential customer. They have a lot definite ideas. In the meanwhile, they would like to share their purchase experiences with others.

After this categorizing, enterprises should be able to dig out the most valuable customers and give them appropriate sales promotions. However, when we check the existent data we find out that the most valuable customers are not from the same category. For example, seemingly the customers from category III should be give the most sales promotions. Since they are easily to accept others' opinions and they have more opportunities to persuade others to buy some produces. However, this is not the truth. Let's take the customers from category I as an instance. Although they rarely consider others' opinion, they may have very high self-desire to buy a produce. This desire may be

much more than the desire of Bidirectional customer which is gotten after others' impact.

Therefore, we consider the purchase trend of customer as the synthetical exhibition of customer relationship attributes. Thus, customer value depends on both self-reliance index within purchase behavior and impact index after purchase. In this paper, we call them Attributes of Customer Relationship (ACR). ACR is the key of customer classification. And how to effectively dig out these ACMS is the main work of this paper.

3.4 Dissimilarity in Cluster Algorithm

In this subsection we are going to discuss about how to evaluate whether a customer trend for hearing his own idea more. Here we would like to use the concept of Dissimilarity in Clustering Algorithm for data mining, in order to give out an evaluating method which suits to our model.

Dissimilarity in Clustering Algorithm for data mining is used to represent the similar degree of two objects. When the attributes of represented objects are different, the algorithm for dissimilarity is different as well. In the Clustering problem domain, for high dimensional sparse data of binary variables, a dissimilarity measure algorithm named Spare Feature Dissimilarity (SFD) of a set was put forward [4]. SFD of a set represents the similar degree of all the objects in a set.

Definition 3.1 (SFD of a set): Given n objects, each object is described by m attributes. m equals 1 or 0. X is a set of objects, in which the number of objects is denoted as $|X|$, the number of attributes that equal 1 for all objects is indicated by a , and the number of attributes that equal 1 for some objects and equal 0 for other objects is indicated by e . SFD of set X , denoted as $SFD(X)$ is defined as: [18]

$$SFD(X) = \frac{e}{|X| \times a} \quad (3.1)$$

$SFD(X)$ represents the similarity of the objects in set $|X|$. The smaller the SFD is, the more similar the objects are.

This feature of SFD totally accord with the logistic definition of customer's self-reliance index. The more the dissimilarity is, the more trend of hearing his own idea the customer has. Thus, we give out the definition of self-reliance index based on the concept of dissimilarity.

We abstract the behavior of customer based on the purchase record in database as follow:

In database each customer's purchase record relates to produce ID. In order to make the follow data processing convenient, we define each customer as an object. All the produces within database are the attributes of this object. When the customer buying a produce, the relative attribute of this customer will be set as 1, otherwise the attribute will be set as 0. We know that this abstract completely accord with the definition of SFD. So we define self-reliance index (SR index) as follows:

Definition 3.2 (SR index): Given n objects of customers, each object has m attributes for produces, set as 1 or 0. $X=\{x_1, x_2 \dots x_n\}$ is an ordered object subset, in which the number of objects is denoted as $|X|$. The number of attributes, that equal 1 for the first object and equal 0 for the next $|X|-1$ objects and the attributes that equal 0 for the first object and equal 1 for the next $|X|-1$ objects, is indicated by e ; and the number of attributes that equal 1 for the first object and the equal 1 for at least one of next $|X|-1$ objects is indicated by a . Thus, dissimilarity of the first object in X is defined as:

$$SFD(x_1) = \frac{e}{|X| \times a} \quad (3.2)$$

And self-reliance index is defined as

$$SR(x_1) = \frac{SFD(x_1)}{1 + SFD(x_1)} \quad (3.3)$$

So we can learn from this definition: the bigger the SFD is, the bigger the SR index is.

3.5 Lift in Association Rules

By now we have confirmed Self-reliance index for each customer. That means we know how much the customer will hear others' opinion. The following work is evaluating how much the customer can impact others. Now we are going to define our evaluating method based on the concepts from association rules.

Here are some concepts from association rules [4]:

Definition 3.3 (Confidence): In the transaction set D , transaction T support itemset X . There are $C\%$ transactions from T also support itemset Y . $C\%$ is named as **Confidence** of association rule $X \Rightarrow Y$. That is,

$$\text{Confidence}(X \Rightarrow Y) = \frac{|\{T \mid T \in D \text{ And } (X \cup Y) \subseteq T\}|}{|\{T \mid T \in D \text{ And } X \subseteq T\}|} \quad (3.4)$$

Definition 3.4 (ExpectedConfidence): In the transaction set D, e% of transaction T support itemset Y. Thus e% is named **ExpectedConfidence** of association rule $X \Rightarrow Y$. That is,

$$\text{Expectedconfidence}(X \Rightarrow Y) = \frac{|\{T \mid T \in D \text{ And } Y \subseteq T\}|}{|D|} \quad (3.5)$$

Definition 3.5 (Lift): Lift is the rate of confidence and expectedconfidence. That is,

$$\text{Lift}(X \Rightarrow Y) = \frac{\text{confidence}(X \Rightarrow Y)}{\text{expectedconfidence}(X \Rightarrow Y)} \quad (3.6)$$

Lift represents how much impact that the existing of itemset X will affect on the existing of itemset Y.

When the first time we saw this definition, we were very exciting. Because this is just the evaluating method what we need to represent our impact index of customer. Thus, we give out our definition of impact index.

Definition 3.6.1: In the database D, customer X exists in the customer purchase record T, and c% of the situation exists customer Y as well. c% is called the confidence of X and Y purchases a produce at the same time. That is the rate of when customer X purchases one produce customer Y also purchases it.

$$\text{Confidence}(Y|X) = \frac{|\{T \mid T \in D \text{ And } (X \cup Y) \subseteq T\}|}{|\{T \mid T \in D \text{ And } X \subseteq T\}|} \quad (3.7)$$

Definition 3.6.2: In the database D, e% of the customer purchase record T has customer Y. e% is called as the expectedconfidence of X and Y purchases a produce at the same time. Expectedconfidence represents the probability of customer Y purchase one produce without any condition.

$$\text{Expectedconfidence}(Y|X) = \frac{|\{T \mid T \in D \text{ And } Y \subseteq T\}|}{|D|} \quad (3.8)$$

Definition 3.6: Lift represents the rate of confidence and expectedconfidence. In our model, that is impact index. It represents when customer X purchase one produce how much impact will being bring to customer Y.

$$\text{Impact}(Y|X) = \frac{\text{confidence}(Y \mid X)}{\text{expectedconfidence}(Y \mid X)} \quad (3.9)$$

By then, we already can mine the self-reliance index and impact index of a customer. The next task is finding an appropriate model to estimate the value of each customer.

3.6 Search Reference Method for Network Relation

As the network impact has been considered in this paper, this impact should be evaluated. Since it has lot complexities, we are going to learn some related models and try to find inspire from them. After a long time searching, two models were found. The analysis of these two will help us to build our own model.

1) Model based on Markov random field

Pedro Domingos and Matt Richardso gave out a model for evaluating the customer's network value, which is the expected profit from sales to other customers he may influence to buy.[5] Instead of viewing a market as a set of independent entities, they regarded it as a social network and model it as a Markov random field. Their solution is based on modeling social networks, where each customer's probability of buying is a function of both the intrinsic desirability of the product for the customer and influence of other customers.

2) Model based on linear network

We found another model based on linear network was given out later. This model extends the previous techniques, achieves a large reduction in computational cost, and applies them to data from a knowledge-sharing site. [11] They showed how to find optimal viral marketing plans, used continuously valued marketing actions, and reduce computational costs. They employ a simple linear model to approximate the interaction between customers. It simply considers a customer's value is the combination of customer's intrinsic value and network value.

Those references give us a good idea to establish ours.

4 Implementation of Data Mining Model

In this section, we are going to set up a linear regression model for customer value. The data mining modeling theories and linear math models mentioned in Chapter 2 will give us a good academic basic. That is the reason that we introduce them at first.

4.1 Symbolic System

a) To make it simple, we focus on a certain produce when we mine the most valuable customer. The feature set of this produce is $Y = \{y_1, y_2, \dots, y_m\}$.

b) Given n potential customers to this produce. It is set $X = \{x_1, x_2, \dots, x_n\}$. Furthermore, we define the neighbors of customer x_i is $N_i \subset X - \{x_i\}$. This neighbor represents the one who get direct impact from the purchase of x_i .

c) Define mapping $s: X \rightarrow \{0,1\}$ to express purchase states of all the customers. If customer i bought this produce set $s(x_i)=1$, otherwise set $s(x_i)=0$. Simply, we use x_i to represent $s(x_i)$. So we can use $X = \{x_1, x_2, \dots, x_n\}$ to represent the purchase state of each customer. Using N_i to represent purchase state of the recent customer i 's neighbor.

d) At last, define mapping $m: X \rightarrow \{0,1\}$ to express the recent sales promotion of enterprise. $m(x_i)$ represents the sales promotion degree to customer i . We denote $m(x_i)$ as m_i . Let $M = \{m_1, m_2, \dots, m_n\}$ represent the state of the entire sales promotion plan.

4.2 Estimate of Purchase Probability

Now we are going to estimate the purchase probability of a customer. Based on certain sales promotion and purchase state, this produce's purchase probability of each customer is: $P(x_i | X - \{x_i\}, Y, M)$.

As customer i will not be impact by customer $X - N_i$, we have:

$$P(x_i | X - \{x_i\}, Y, M) = P(x_i | N_i, Y, M)$$

Furthermore, we do linear estimate on $P(x_i | N_i, Y, M)$. We consider $P(x_i | N_i, Y, M)$ equal to the linear increased of purchase probability from customer himself and from neighbor's impact. That is,

$$\begin{aligned} & P(x_i | X - \{x_i\}, Y, M) \\ &= P(x_i | N_i, Y, M) \\ &= SR(x_i)P_0(x_i | Y, M) + (1 - SR(x_i))P_N(x_i | N_i, Y, M) \end{aligned} \quad (4.1)$$

$P_0(x_i | Y, M)$ is the purchase probability based on customer's self-willing. We call it embedded probability. It doesn't relate to customer's neighbor. $P_N(x_i | N_i, Y, M)$ is the purchase probability based on impact from neighbor. We call it network probability. While $SR(x_i)$ is the self-reliance index which we mentioned before.

Moreover, we do linear estimate on $P_N(x_i = 1 | N_i, Y, M)$. We consider it as the linear plus of all neighbor's states. That is,

$$P_N(x_i = 1 | N_i, Y, M) = \sum_{x_j \in N_i} impact(x_i | x_j)P(x_j = 1 | Y, M) \quad (4.2)$$

$impact(x_i | x_j)$ represents how much customer j impact on customer i . x_j is the recent state of customer. That is $s(x_j)$. If $j \notin N_i$, set $impact(x_i | x_j) = 0$.

From formula 4.1 and 4.2, we get

$$\begin{aligned} & P(x_i = 1 | N_i, Y, M) \\ &= SR(x_i)P_0(x_i = 1 | Y, M) + (1 - SR(x_i)) \sum_{x_j \in N_i} impact(x_i | x_j)P(x_j = 1 | Y, M) \end{aligned} \quad (4.3)$$

Given $C(N_i)$ is the set of all the possible states of customer from N_i . That is if $\tilde{N} \in C(N_i)$, \tilde{N} will be one of the states of customer from N_i . Generally, all the neighbor states are unknown, so $P(x_i = 1 | Y, M)$ is what we need to calculate. So we have

$$\begin{aligned}
& P(x_i = 1 | N_i, Y, M) \\
&= \sum_{\tilde{N} \in C(N_i)} P(x_i = 1 | \tilde{N}, Y, M) P(\tilde{N} | Y, M) \\
&= \sum_{\tilde{N} \in C(N_i)} SR(x_i) P_0(x_i = 1 | Y, M_i) P(\tilde{N} | Y, M) \\
&+ \sum_{\tilde{N} \in C(N_i)} (1 - SR(x_i)) \sum_{x_j \in N_i} impact(x_i | x_j) \tilde{N}(x_j) P(\tilde{N} | Y, M)
\end{aligned} \tag{4.4}$$

$\tilde{N}(x_j)$ represent the value of x_j based on \tilde{N} . As $P_0(x_i | Y, M_i)$ doesn't relate to the state of N_i , we have

$$\sum_{\tilde{N} \in C(N_i)} SR(x_i) P_0(x_i = 1 | Y, M_i) P(\tilde{N} | Y, M) = SR(x_i) P_0(x_i = 1 | Y, M_i).$$

Since the value of $\tilde{N}(x_j)$ is 1 or 0, we have

$$\begin{aligned}
& \sum_{\tilde{N} \in C(N_i)} (1 - SR(x_i)) \sum_{x_j \in N_i} impact(x_i | x_j) \tilde{N}(x_j) P(\tilde{N} | Y, M) \\
&= (1 - SR(x_i)) \sum_{x_j \in N_i} \sum_{\substack{\tilde{N} \in C(N_i) \\ \text{with } \tilde{N}(x_j)=1}} impact(x_i | x_j) \tilde{N}(x_j) P(\tilde{N} | Y, M) \\
&= (1 - SR(x_i)) \sum_{x_j \in N_i} impact(x_i | x_j) P(x_j = 1 | Y, M)
\end{aligned}$$

At last, there is

$$\begin{aligned}
& P(x_i = 1 | Y, M) = SR(x_i) P_0(x_i = 1 | Y, M_i) \\
&+ (1 - SR(x_i)) \sum_{x_j \in N_i} impact(x_i | x_j) P(x_j = 1 | Y, M)
\end{aligned} \tag{4.5}$$

According to formula 4.5, we find that to iterative calculate all the customers' purchase probability we only need to know the embedded probability of each customer, that is $P_0(x_i = 1 | Y, M_i)$. In the meanwhile, we should use it as the initial value of customer purchase probability.

According to Naïve Bayes Model, we have

$$\begin{aligned}
& P_0(x_i | Y, M_i) \\
&= \frac{P_0(Y | x_i) P_0(M_i | x_i) P_0(x_i)}{\sum_{x_j} P_0(Y | x_j) P_0(M_i | x_j) P_0(x_j)} \\
&= \frac{P_0(x_i | Y) P_0(M_i | x_i)}{\sum_{x_j} (P_0(x_j | Y) P_0(M_i | x_j))}
\end{aligned} \tag{4.6}$$

Now we only need to calculate $P_0(x_i | Y,)$ and $P_0(M_i | x_i,)$. For the first one, we can use $P_0(x_i | Y,) = \frac{n(x_i | Y) + mp}{\sum_{x_i} n(x_i | Y) + p}$ to calculate.

$n(x_i | Y)$ represents the number of customer i 's purchase based on the attribute Y . p is an initial estimate of probability, usually use its experience value. For the second one, we can suppose that

$P_0(x_i = 1 | M_i = 1) = \min(\alpha P_0(x_i = 1 | M_i = 0), 1)$. It means a customer's purchase probability after sales promotion is α times as the probability without sales promotion. That is the estimate of sales promotion's effect. Simply, we consider

$P_0(M_i = 0) = P_0(M_i = 1) = 0.5$, so we get:

$$\begin{aligned} P_0(M_i = 1 | x_i = 1) &= \frac{P(x_i = 1 | M_i = 1)P(M_i = 1)}{P(x_i = 1)} \\ &= \frac{P(x_i = 1 | M_i = 1)P(M_i = 1)}{P(x_i = 1 | M_i = 1)P(M_i = 1) + P(x_i = 1 | M_i = 0)P(M_i = 0)} \quad (4.7) \\ &= \frac{\alpha}{\alpha + 1} \\ P_0(M_i = 1 | x_i = 0) &= \frac{1}{\alpha + 1} \end{aligned}$$

By then, we can use formula 4.5 to calculate purchase probability of each customer.

4.3 Evaluate Customer Value

For evaluating customer value, we are going to use P.Domingos and M.Richardson's evaluating method. This can let us use the same criterion to estimate our experiment result.

At first, we define $f_i^0(M)$ as a sales promotion configuration when $M_i=0$. $f_i^1(M)$ is a sales promotion configuration when $M_i=1$. The cost of sales promotion for one customer is c . The profit brought from the produce when customer buys it is r . So the profit which enterprise can get from customer after sales promotion is:

$$ELP_i^1 = rP(x_i = 1 | Y, f_i^1(M)) - rP(x_i = 1 | Y, f_i^0(M)) - c \quad (4.8)$$

While based on a specific sales promotion plan's state M , enterprise can get the total profit is

$$ELP(Y, M) = \sum_{i=1}^n (rP(x_i = 1 | Y, M) - rP(x_i = 1 | Y, M_0) - c) \quad (4.9)$$

$M_0 = \{0, 0, \dots, 0\}$ means there is no sales promotion.

Thus, customer i 's entire value is

$$TOLV_i = ELP(Y, f_i^1(M)) - ELP(Y, f_i^0(M)) \quad (4.10)$$

Utilize this evaluating method we can translate customer's purchase probability into the value for enterprise.

5 Experiments

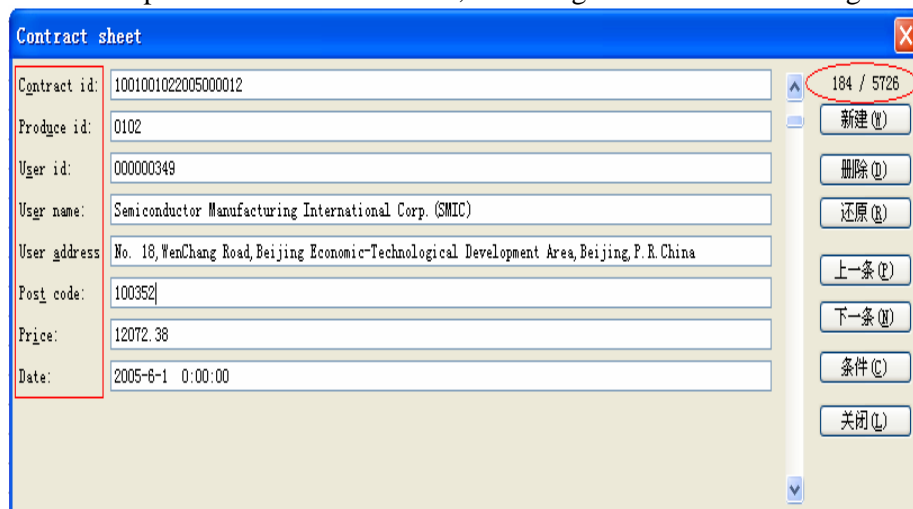
In this chapter, we intent to give you a general introduce about our experiments. It includes from data preparing to calculate result. Those figures and tables will help you get an intuitionistic idea.

5.1 Experimental Setup

The intention of this subsection is to give an abstract description of the experimental setup. It includes the data sets, main functions and main algorithm.

5.1.1 Data Set

We used 5726 purchase records as our experimental data. They are all true data from a real world company. Looking at the original data set, each purchase is expressed as a contract from, including the information as Fig 5.1



The screenshot shows a software window titled "Contract sheet" with a blue title bar and a red close button. The window contains a form with the following fields and values:

Field	Value
Contract id:	1001001022005000012
Produce id:	0102
User id:	000000349
Usgr name:	Semiconductor Manufacturing International Corp. (SMIC)
User address:	No. 18, WenChang Road, Beijing Economic-Technological Development Area, Beijing, P. R. China
Post code:	100352
Price:	12072.38
Date:	2005-6-1 0:00:00

On the right side of the form, there is a vertical stack of buttons: "新建(N)" (New), "删除(D)" (Delete), "还原(R)" (Restore), "上一条(U)" (Previous), "下一条(L)" (Next), "条件(C)" (Condition), and "关闭(L)" (Close). Above these buttons, the text "184 / 5726" is displayed, indicating the current record number and total record count. A red circle highlights this text.

Fig 5.1 Contract sheet of the each purchase

There are about 3800 Users, in our data set, And we also have the detail information of each customer, including User type (company or individual), User id, User name, Post code, Profession type. as Fig5.2

The screenshot shows a window titled 'Sheet 1' with a close button (X) in the top right corner. On the left, there is a vertical list of labels: 'User type:', 'User id:', 'User name:', 'Profession type:', and 'Post code:'. To the right of these labels are input fields containing the following values: '0', '000000120', 'SHANGHAI YIBAI INTERNATIONAL TRADE CO., LTD', '60918', and '200120'. On the far right, there is a vertical stack of buttons: '新建 (N)' (New), '删除 (D)' (Delete), '还原 (R)' (Restore), '上一条 (P)' (Previous), '下一条 (N)' (Next), '条件 (C)' (Condition), and '关闭 (C)' (Close). Above the '新建 (N)' button, the text '43 / 3849' is displayed, indicating the current record number and total record count.

Fig 5.2 User Information Sheet

Fig 5.3 shows the distributing of all the users. The picture can not show each of those 3800 users clearly, but it is very obvious that some of the users purchased a lot of times, others did not. That is easy for the company to suppose there are some big customers in this data set. They should have higher value for the company, but just from this distributing the company can not tell who the most valuable customers are. To solve this problem is just the goal of our model which is defined in the previous sections. Furthermore, our experiment will use the algorithm to set up before to mine the top ten valuable customers in this data set.

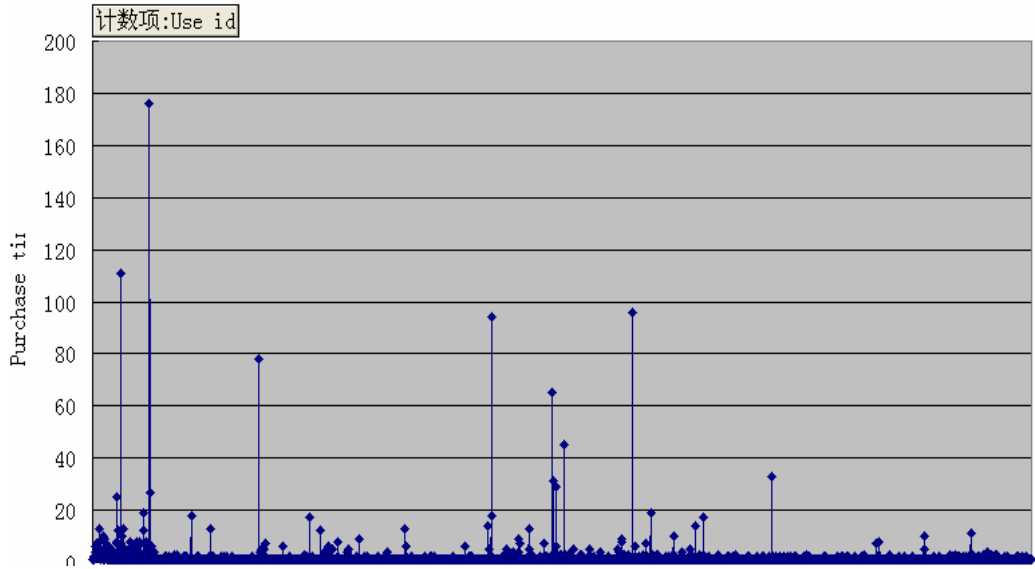


Fig 5.3 Distributing of Customer Purchases

5.1.2 Main Functions Name and Description

Table 5.1 gives out the main functions in the experiment which are used for calculate the value of customers.

Function Name	Description
CalculateImpact	Used to calculate the impact among customers within one neighbor group.
CalculateNeighbor	Using CABOSFV to classify customers into their neighbor groups.
CalculatePnYM	Used to calculate $P_N(x_i N_i, Y, M)$
CalculatePoM	Used to calculate $P_0(M_i x_i,)$
CalculatePoY	Used to calculate $P_0(x_i Y,)$
CalculatePoYM	Used to calculate $P_0(x_i Y, M_i)$
CalculateSR	Used to calculate each customer's SR
CalculateVOC	Used to calculate customers' value based on both ACR and PM's model

Table 5.1 Description of main functions

5.1.3 Mining Algorithm

1) Code for mining self-reliance index:

```
void CalculateSR() {
    this->CalculateNeighbor();

    this->SR.resize( this->u.size() );
    for(int i=0;i<this->u.size();i++) {
        this->SR[i] = this->GetSR( i );
    }
}
```

2) Code for mining impact index shows as follow:

```
void CalculateImpact() {
    this->IM.resize( this->u.size() );
    char tline[1024];

    for(int i=0;i<this->u.size();i++) {
        this->IM[i].resize( this->u.size() );
        for(int j=0;j<this->u.size();j++) {
            this->IM[i][j] = this->GetImpact(i, j);
        }
    }

    return;
}
```

All the algorithms used by our experiment have been introduced and proved in section 3 and 4. Furthermore, we used the formula 4.5 to calculate each customer's purchase probability and formula 4.10 to calculate the total value of each customer. The evaluated results show in the next subsection.

5.2 The evaluated results

Here we choose 100 customers as a sample data set and did a preprocess for their purchase records, such as changing the user id to 1-100 in order to make it easy to be read.

usr_id	Our VoC	usr_id	PM's VoC
62	9.3023	54	4.6846
75	8.6379	62	4.6512
88	8.3056	6	4.4877
76	7.9734	75	4.3189
79	7.9734	88	4.1528
97	7.6412	52	4.0253
66	6.9767	76	3.9867
73	5.9801	79	3.9867
6	4.7971	8	3.9410
54	4.0140	2	3.9401

Table 5.2 Evaluation of the Value of Customer (VoC)

For the table 5.2, the first two rows are the top ten valuable customers' id and the Value of Customer (VoC) gotten from our algorithm. While the last two rows show the top ten valuable customers' id and Voc mined from PM's model. The important thing that must be emphasized is the promotion costs from these two algorithms are the same.

As we can see now, we have the similar set of most valuable customer in our result and PM's. At the meantime, we have some different customers. All of these show our model's correctness and possible improve. Moreover, when we look at the values we got. We find that the results from PM's model are without big dissimilarity, which means it hasn't given a well estimate among those customers. However, our result can show the difference between customers' values more clearly. In another word, our model gives a prominence to ACR when we using it calculate customer value and we emphasize the differences among them. This result is just what we want at the beginning of this thesis.

5.3 Discussion

The general idea of the experiments in this paper has three steps. The first one is getting excel file data from a company's database and giving preprocessing to the original data. The preprocessing simplified the contract form, and only kept Produce id as an attribute set of User. The second step is realizing our algorithms using VC++ which have been set up in section 3 and 4.

Furthermore, using txt file as an input data we calculate all the customers' value based on ACR. The third step is using the same data but PM's algorithm calculates all the customers' value. Since we know that PM's model is considered as a mature model, we trust its validity and correctness. So when got the results, we take the top ten valuable customer as typical results in order to make a contrast between to results. One the other hand, science the data is from an insurance company their own experiences about big customers also confirm our result as reasonable result.

6 Conclusion and Future Works

By now the two main problems given in the introduction has been solved. A more consummate segmentation of customer has been given out in Section 3. It gave a more careful classification based on the social network environment. Actually, it is also the basic idea of our model for estimating customer value.

In this thesis, the linear regression model was mentioned. It divides the huge object of customer. Moreover, it uses network relationship to give a classification of customers. This model may be not exact, because it used two linear approximations. But we view it as reasonable approximations. As when the probabilities are all small, as in most cases for marketing domains, linear models usually perform well. When data is sparse they can provide significant advantages for computation. Furthermore, by implementing the model we gave in this thesis, the second problem given in the introduction has been solved.

However, we must say that the experiments in this thesis are very preliminary. In the future work, we should try to improve the visualization of our application. And the main contribution of the thesis is in the proposal of methods, which still have to be evaluated on a large scale in realistic settings. And also in the real world social network is not that simple as we discussed in this paper. In the future works, we can still improve our models. For example, we simply consider the impact relationship of customers as one-to-one linear increased. We did not consider the situation of many-to-one impact relationship. So the future work can be extended the linear regression model to a more complex trust relationship as well.

7 Acknowledgement

This paper was finished in my second semester in Vaxjo University. As an exchange student, I want to thank this university to give me such a wonderful opportunity of doing a thesis. Although it was hard, I appreciated it very much. During working on it, I got a lot of help from both teachers and friends.

Thanks to my supervisor Pro. Joakim, he gave me plenty of good ideas at the beginning of this paper, helped me to decide the exact problem domain which should be solved, and gave me continuous advises during the work. Thanks to my teacher Mr. Mathias. He always tracks our work to make sure there is no problem, and if there is, he would give you immediate help to solve the problem. And he also gave me a good idea on how to write the thesis and what is the process of this course. Thanks to my friends, they gave me lots of encourage and help on both study and life. Thanks to my parents, they are always let me know they love me which gives me huge motility. At last, I want to thank this exchange program. I got abundant of wonderful memories from here.

8 References

1. Jedid-Jah Jonker, Nanda Piersma and Rob Potharst, **Direct Mailing Decisions for a Dutch Fundraiser**, Econometric Institute Report EI 2002-19
2. Keith C.C. Chan, Wai-Ho Au, Berry Choi, **Mining Fuzzy Rules in A Donor Database for Direct Marketing by A Charitable Organization**,
3. Rob Potharst, Uzay Kaymak and Wim Puls, **Neural networks for target selection in direct marketing**, in ERIM REPORT SERIES RESEARCH IN MANAGEMENT, March 2001, ERS-2001-14-LIS
4. M. Bastian, **Data warehousing and data mining**
5. P. Domingos and M. Richardson, **Mining the Network Value of Customers**, In Proceedings of the Seventh International Conference on Knowledge Discovery and Data Mining, pages 57-66, San Francisco, CA, 2001. ACM Press.
6. Fayyad, U. Et al., **Knowledge Discovery and Data Mining Towards a Unifying Framework**, KDD-96 Proc. 2nd Intl. Conf. On Knowledge Discovery & Data Mining, AAAI press, 1996
7. Charles X. Ling and Chenhui Li : **Data Mining for Direct Marketing Problems and Solutions**, KDD-98
8. Peter Van Der Putten, **Data Mining In Direct Marketing Databases**
9. **Data Mining in the Insurance Industry**, Solving Business Problems using SAS® Enterprise Miner™ Software
10. Haughton, D. and S. Oulabi (1993). **Direct marketing modeling with CART and CHAID**, Journal of Direct Marketing, 7, 16-26.
11. Pedro Domingos, Matt Richardson, **Mining Knowledge-Sharing Sites for Viral Marketing**
12. Robert Groth, **Data Mining** (A hands-On Approach for Business Professionals)
13. Margaret H. Dunham, **Data Mining** (Introductory and Advanced Topics)
14. Michael J. Shaw a,b,c), Chandrasekar Subramaniam a, Gek Woo Tan a, Michael E. Welge b, **Knowledge management and data mining for marketing.**

15. **Data Mining for Marketing Applications**, 12th European Conference on Machine Learning (ECML'01) *and* 5th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'01), September 7th, 2001 Freiburg, Germany
16. David J. Hand, Heikki Mannila and Padhraic Smyth, **Principles of Data Mining**.
17. Peter Van and Der Putten, **Data Mining in Direct Marketing Databases**, World Scientific, October 15, 1998
18. Sen Wu and Xuedong Gao, **CABOSFV algorithm for high dimensional sparse data clustering**



Växjö
universitet

Matematiska och systemtekniska institutionen
SE-351 95 Växjö

tel 0470-70 80 00, fax 0470-840 04
www.msi.vxu.se