The Data Excavation Model in CRM Based on Fuzzy Rule

Cao jie[1]; Wu mingzan[2]
([1]E-commerce Department, Nanjing University of Finance and Economics, Nanjing, 210003)
([2]Automatization College, Nanjing University of Science and Technology, Nanjing, 210094)

***Abstract*- It is needed for the strengthening of enterprises competed abilities to excavate customer resources and realize individual services. The thesis uses a swift fuzzy rules excavated method in the customer relationship management of E-commerce. Practice simulation show its authenticity, and then the thesis gain the intelligent customer relationship management model that faces to electronic business.**

Key words: E-commerce; customer relationship management; data excavation; fuzzy rules

I. INTRODUCTION

In the era of E-commerce, the core meaning of enterprise management is stretching further. Besides traditional enterprise financial management, storage management, sales management, stock management, sales management, stock management and many parts involved in the entire enterprise value chain are required to be included in the management category. With the intensification of market competition, it's hard to tell absolute advantages in the enterprises' products and services. Those who can grasp the needs of customers and strengthen communications with customers can gain competition advantage take an initiative.

So we need the Customer Relationship Management (CRM) which can enhance the income, profit and the satisfaction of customers. CRM is a management concept which's core is to take customers as the most important enterprise sources and to guarantee customers' lifetime value by fulfilling customers' needs through complete service and deep customer analysis.

CRM is a new management mechanism aimed to improve the relationship between enterprises and customers. It is implemented in marketing, sales, service, technical support and many other fields concerning customers. By providing comprehensive and personalized customer information to professional staff in the fields of sales, markets, customer services and strengthening its ability to track service and analyze information. It can help them to construct and maintain a series of "one-to-one" relationships between enterprises and customers and between enterprises and cooperative partners. So by this way enterprises can provide better and quicker service, improve the satisfaction of customers, attract and retain more customers in order to increase turnovers and on the other hand reduce operational costs by sharing information and optimizing business flows. The implementation of CRM requires constructing enterprises, consummating infrastructure that can respond quickly to customers' needs, standardizing customer service-focused transaction flows, building customer droved product/service design with a customer-focused concept, so as to cultivate customers' loyalty to the brand and enhance profitability ratio.

CRM is also a kind of management software and technology, which combines the business practice with data excavation, data warehouse, one-to-one marketing, sales automation and other information technologies tightly and provides a solution to transaction automation for the fields of sales, best customer service, decision-making support and so on. It makes enterprises have a customer oriented forefront based on E-commerce, so they can make a successful transition from traditional enterprise modes to modern enterprise modes based on E-commerce.

With the increase of commerce data, data excavation will become the key technology in CRM in the comprehensive of customers. Enterprises will use customer resources and develop the analysis and forecast of customer behaviors based on the technology of data excavation which can classify customers, analyze customer's payoff power, search potential valuable customers, development individuation service and enhance the degree of satisfaction and loyalty of customers. That is propitious to deepen the management for customers, scout the market motions, forecast customers' consume trend, and research the individuation product meeting customers' need [1].

In this paper, we will use the swift method of data excavation with fuzzy rule which can deal with the incertitude things in various commerce data. So we will excavation the fuzzy rule in numerical value data and use it in the CRM of E-commerce [2].

II. THE DESCRIPTION OF THE EXCAVATED METHOD ABOUT FUZZY RULE OF NUMERICAL DATA

A. The system of fuzzy rule

Considered that a multi-input/single output system:

$$\{(x_p, y_p) \mid p = 1, 2, \cdots, k\} \tag{1}$$

And $x_p = \{x_{p1}, x_{p2}, \cdots, x_{pm}\}$ is the pth input in m dimension, m is the number of input data, y is the single output data [3] [4].

Supposed that is the ith continuous attribute, $v_{a_i} = [l_{a_i}, r_{a_i}] \in R$ is the real numerical region of he attribute , and are both real numbers, and the list of data

aggregate which has n data swatches is $xi = (x_{i,1}, x_{i,2}, \cdots, x_{i,n})(i = 1, 2, \cdots, m)$ The clustering center of the list can be gained by subtraction

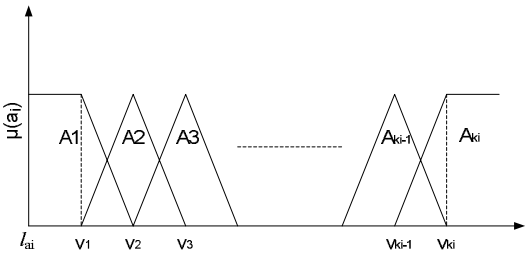

Fig.1 Fuzzy disperse principle of continuous attribute

clustering method that can recognize the number of clustering categories automatically, and k_i is the clustering center number of the ith variable, just as Fig. 1.

The paper uses triangle subjected function, and the subjected function of the ith fuzzy region can form by parameters v_{j_i-1}, v_{j_i}, and v_{j_i+1}, $j_i = 1,2,\cdots,k_i$ When $j_i = 2,3,\cdots,k_{i-1}$, the subjected function can express as

$$\mu_{ij_i} = \begin{cases} \dfrac{x - v_{j_i-1}}{v_{j_i} - v_{j_i-1}} & v_{j_i-1} \le x \le v_{j_i} \\ \dfrac{v_{j_i+1} - x}{v_{j_i+1} - v_{j_i}} & v_{j_i} \le x \le v_{j_i+1} \end{cases} \quad (2)$$

When $j_i = 1$ and k_i, the subjected function can express as

$$\mu_{i1} = \begin{cases} 1 & x \le v_1 \\ \dfrac{v_2 - x}{v_2 - v_1} & v_1 \le x \le v_2 \end{cases} \quad (3)$$

$$\mu_{ij_i} = \begin{cases} \dfrac{x - v_{k_i-1}}{v_{k_i} - v_{k_i-1}} & v_{k_i-1} x \le v_{k_i} \\ 1 & v_{k_i} \le x \end{cases} \quad (4)$$

For the expression after, output variables can divide into N fuzzy interspaces, which can mark as $B_1, B_2, \cdots, B_N$, and their corresponding subjected function $\mu_{B1}, \mu_{B2}, \cdots, \mu_{B_N}$ are defined like （2）~（4）. We define a zero-rank Sugeno fuzzy rule model as:

Rule $R_{j_1\cdots j_m}$:IF x_1 is A_{1j_1} and x_2 is A_{2j_2} …AND x_m is A_{mj_m} THEN y is $b_{j_1\cdots j_m}$. (5)

In the formula above, $R_{j_1\cdots j_m}$ is the mark of $k_1 \times k_2 \times \cdots k_m$ fuzzy rules, and $b_{j_1\cdots j_m}$ is the real number result. For a assured input $x_p = (x_{p1}, x_{p2}, \cdots, x_{pm})$, we define the adapt range of input x_p and rule $R_{j_1\cdots j_m}$ is:

$$\mu_{j_1\cdots j_m}(x_p) = \theta \times \mu_{1j_1}(x_{p1}) \times \mu_{2j_2}(x_{p2}) \times \cdots \times \mu_{mj_m}(x_{pm}) \quad (6)$$

In this, θ is the data reliability, a real number which chooses its value between [0, 1].

B. The excavation of fuzzy rule

(1) The result of fuzzy rule. In order to strengthen the non-linearity dealing ability, we define the weight of the pth input and output (x_p, y_p) is:

$$\eta_{j_1\cdots j_m}(x_p) = \left[\mu_{j_1\cdots j_m}(x_p)\right]^{\lambda} \quad (7)$$

In this, the use of index λ is to make the triangle subjected function nonlinearized in order to form a bell-like subjected function. For k data (x_p, y_p), p=1…k, we figure out $\eta_{j_1\cdots j_m}(x_p)$ separately, and then work out $b_{j_1\cdots j_m}$ using formula(8):

$$b_{j_1\cdots j_m} = \left(\sum_{p=1}^{k} \eta_{j_1\cdots j_m}(x_p) \bullet y_p\right) / \sum_{p=1}^{k} \eta_{j_1 \ \ j_m}(x_p) \quad (8)$$

$b_{j_1\cdots j_m}$ is the average weight of y_p, If there is one $\eta_{j_1\cdots j_m}(x_p)$ that is not zero at least, we can use formula (8) to figure out fuzzy rule system formula (5). If all $\eta_{j_1\cdots j_m}(x_p)$ is zero, the relevant fuzzy sub-regions should be combined with the bordered sub-region.

(2) The formation of fuzzy rule language. The former piece parameters $A_{1j_1}, A_{2j_2}, \cdots, A_{mj_m}$ in the fuzzy rule system of formula (5) show the fuzzy interspaces, and back piece parameter $b_{j_1\cdots j_m}$ is expression that can not intelligible. In practice, we can gain a rule that can figure by language variable form a great deal of data in order to be understood easily, which is the aim of data excavation. So, the former piece $A_{1j_1}, A_{2j_2}, \cdots, A_{mj_m}$ and back piece should be expressed as language variable.

Supposed that we gain fuzzy rule system formula (5), and the fuzzy set $B_1, B_2, \cdots, B_N$ can express by language variable. For example, N=5, and $B_1, B_2, \cdots, B_N$ can express as MS，M，ML，L, which is small, a little small, moderate, a little big and big. Because we use the triangle subjected function, on matter what value y choose, there are at most two fuzzy regions to correspond to them, and the set $\{\mu_{Bi}(b_{j_1\cdots j_m}) \mid \mu_{Bi}(b_{j_1\cdots j_m}) \ge 0, i = 1\cdots N\}$ has only two elements, which can mark as $\mu^{(1)}_{B_{j_1\cdots j_m}}(b_{j_1\cdots j_m})$ and $\mu^{(2)}_{B_{j_1\cdots j_m}}(b_{j_1\cdots j_m})$, $\mu^{(1)}_{B_{j_1\cdots j_m}}(b_{j_1\cdots j_m}) \ge \mu^{(2)}_{B_{j_1\cdots j_m}}(b_{j_1\cdots j_m})$. So we can gain:

$$\mu^{(1)}_B(b_{j_1\cdots j_m}) + \mu^{(2)}_B(b_{j_1\cdots j_m}) = 1 \quad (9)$$

Formula (6) can express by two language variable fuzzy rule, just as follows:

Rule $R^{(l)}_{j_1\cdots j_m}$:IF x_1 A_{1j_1} AND x_2 is A_{2j_2} …AND x_m is A_{mj_m} , THEN y is $B^{(l)}$ with

$$CF=\mu_{B^{(l)}_{j1\cdots jm}}(b_{j_1\cdots j_m}),l=1,2 \quad (10)$$

(3) The process of learning and reasoning. So we can gain the model learning process of fuzzy rule system:

① Making n input and output variables dispersive, and dividing them into several fuzzy subjected regions. Generally, it is inadvisable to carve up too many fuzzy regions in order to describe the language conveniently, and if there are great original data, efforts should be made on clustering analyses. That is to divide the original data into several small regions, and then to make it dispersive.

② For every decided input and output (x_p, y_p), P=1…k, p is the number of input and output data. We can figure out $\mu_{ij_i}(x_1),\mu_{ij_i}(x_2),\cdots,\mu_{ij_i}(x_m),$ through （2）~（4）and $\mu_{j_1\cdots j_m}(x_p)$ through formula (6), and then we choose the index λ to find $\eta_{j_1\cdots j_m}(x_p)$ [5].

③We can figuring out $b_{j_1\cdots j_m}$ by formula (9), and make $A_{1j_1}, A_{2j_2},\cdots, A_{mj_m}$ to be $k_1\times k_2\times\cdots\times k_m$ fuzzy subjected regions. So there are $k_1\times k_2\times\cdots\times k_m$ fuzzy rule through formula (5).

④For every $b_{j_1\cdots j_m}$, $\mu^{(1)}_{B_{j_1\cdots j_m}}(b_{j_1\cdots j_m})$ and $\mu^{(2)}_{B_{j_1\cdots j_m}}(b_{j_1\cdots j_m})$ can be count out, and then form $2\times k_1\times k_2\times\cdots\times k_m$ fuzzy rules on bases of language variables[6].

So for any new input and output x_p, the model output $y_p(x_p)$ can be figured out by formula (11):

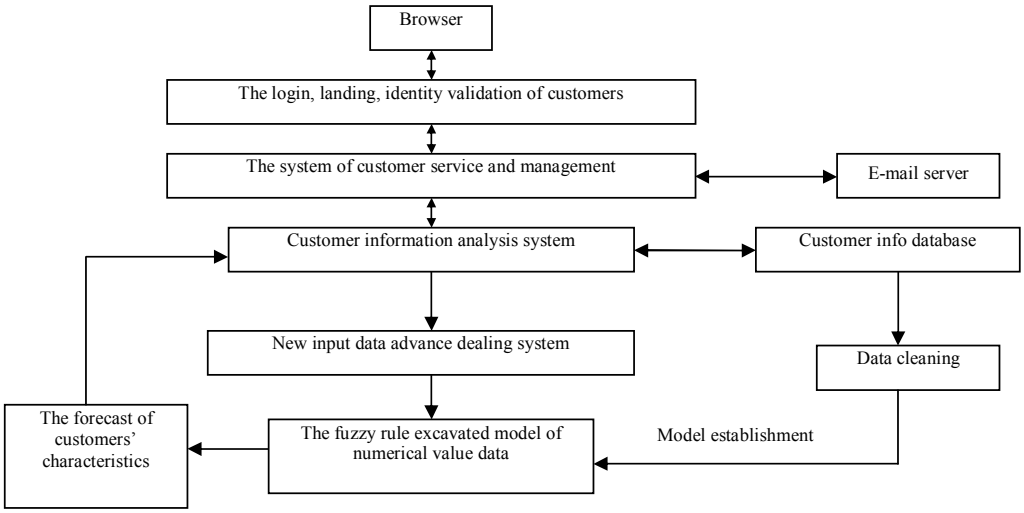

Fig. 2 Excavated system structure of intelligent customer relationship

$$y(x_p)=\left(\sum_{j_1=1}^{k_1}\cdots\sum_{j_m=1}^{k_m}\mu_{j_1\cdots j_m}(x_p)\bullet b_{j_1\cdots j_m}\right)\Bigg/\sum_{j_1=1}^{k_1}\cdots\sum_{j_m=1}^{k_m}\mu_{j_1\ j_m}(x_p) \quad (11)$$

In the arithmetic of fuzzy rules, $b_{j_1\cdots j_m}$ in the rule $R_{j_1\cdots j_m}$ is an m-dimensions data memory, so it must be made to form a relational database. We should turn m-dimensions to one-dimension to store the fuzzy subjected interspaces of corresponding input variables in relational database. Supposed that $R_{j_1\cdots j_m}$ is marked $R^*_t, t\in[1,(k_1\times k_2\times\cdots\times k_m)]$ in the corresponding relational database. We use row priority rule, so:

$$t=\sum_{i=1}^{k_{m-1}} j_i\bullet k_i+j_m \quad (12)$$

We turn $b_{j_1\cdots j_m}$ to b^*_t, replace the fuzzy subjected interspaces $A_{1j_1}, A_{2j_2},\cdots, A_{mj_m}$ with numbers or symbols, and add a language conclusion to express functions and $CF=\mu_{B^{(l)}_{j1\cdots jm}}(b_{j_1\cdots j_m}),l=1or2$ to store $2\times k_1\times k_2\times\cdots\times k_m$ fuzzy rules in relational database.

III. THE EXCAVATED MODEL ON CUSTOMER RELATIONSHIP MANAGEMENT DATA

The customer management of electronic business requests enterprises to build a customer information database, on the bases of which we can serve customers and analyze their demands on internet to realize customer individuation. Fig. 2 is a customer relationship management database found based

on fuzzy rules. The electronic business sends the results that forecast by customers' characteristics to them by mutual net page and e-mail, and make follow services pertinently.

IV. Some Common Mistakes

The elements that effect customer purchasing levels in the use of vehicle electronic business are like these: family income, customers' education level, ages, the number of children, family housing, etc. According to the five chief indexes of customer statistical data, we analyze their purchasing inclines of vehicle on high level, middle level and low level. The following is some typical purchasing investigative tables in vehicle electronic business (table 1). We can realize the arithmetic described in table 1 using MATLAB7.1, and family income, customers' education level, ages, the number of children, family housing are: $x_1, x_2, x_3, x_4, x_5, y$ is the purchasing incline on high level, middle level and low level. When $\lambda = 1$，2，4，5, the errors are 0.211，0.134，0.071，0.037. We choose λ =5.

TABLE I

CUSTOMERS' VEHICLE PURCHASING INVESTIGATIVE TABLE

serial number	Income (yuan)	education level	age	children number	housing （m2）	purchasing level
1	6000	bachelor	35	0	80	low
2	8000	master	36	1	120	middle
3	9000	doctor	38	1	150	middle
4	10000	master	45	1	140	high
5	15000	post doctor	39	1	135	high
6	9000	bachelor	43	2	110	middle
7	7000	bachelor	45	1	100	low
8	12000	doctor	45	1	180	high
9	7000	master	40	1	110	low
10	8000	master	35	0	120	high
11	13000	bachelor	38	1	140	high
12	8000	master	30	0	130	high
13	8500	doctor	35	1	120	middle
14	6500	master	34	1	110	low
15	9500	post doctor	35	1	135	high

V. TAG

In the thesis, a swift fuzzy rules excavate method was used in the customer relationship management of enterprises electronic business. The method learns fuzzy rules from numerical value data developmentally, and makes them lingual, in which there are not any iterated integrals, and few of accounting. The practical simulation obtains a preferable applying effect, at the same time, and gives us an intelligent customer relationship model based on fuzzy rule excavate method of numerical value data which faces to electronic business.

Acknowledgment

This research is supported by the high-tech industry developing project of the Education Department of Jiangsu Province "Research of cooperation e-commerce platform towards the different source data"(JHB05-23); Jiangsu Six Talent Summit Project" Customer Relationship Management based on DSS".

E-mail: caojie690929@163.com

References

[1] Huang jiejun, Wan youchuan. The E-commerce strategy based on the data extraction [J]. Computer application and software, 2004, 21(7): 12-14.

[2] Xie yongfang, Hu zhikun, Gui weihua. The swift method of extraction with fuzzy rule based on the numerical value data [J]. Control project, 2006, 13(5): 442-448.

[3] Zhang shuhong, Zhang quanlong, Chen dejun. The study of data extraction model in CRM faced E-commerce [J]. Industry project and management, 2004, (4): 78-82.

[4] Ding qiulin, Li shiqi. Customer relationship management [M]. Tsinghua Press, 2002:6-7.

[5] France C, Richard L. Constraining the optimization of a fuzzy logic controller using an enhanced genetic algorithm[J].IEEE Trans System Man Cyber net Part B,2000,30(1):31-45

[6] Li R P, Mukaid. Gauss ion clustering method based on maximum fuzzy-entropy interpretation[J]. Fuzzy sets and system, 1999,102(2):253-258