Journal of Theoretical and Applied Information Technology
20th February 2013. Vol. 48 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A NEW ALGORITHM OF ONINE CUSTOMER CLASSIFICATION BASED ON ANT COLONY ALGORITHM

[1]**LI XINWU**

[1] Prof., Electronic Business Department, Jiangxi University of Finance and Economics, Nanchang, China

E-mail: [1]liyue7511@163.com

ABSTRACT

Effective and correct customer classification plays a key role in analyzing and estimating customer consumption patterns, establishing personalized marketing service system and differentiation management. A new customer classification algorithm is advanced based on ant colony algorithm. Firstly, based on consumer characteristics and behaviors analysis, an indicator system is designed which includes 21 indicators belongs to customer characteristics type variables and customer behaviors type variables respectively. Secondly, ant colony algorithm is improved through adjusting its transition probability dynamically to speed up model's convergence and to build the selection strategy more conducive to the overall search, and then a new algorithm for customer classification is deduced and analyzed. Finally the experimental results verify not only the problem of convergence speed has been solved, but also the simplicity of the model structure and the accuracy of the classification are ensured when the new algorithm are used in electronic customer relationship management practically.

Keywords: *Customer Analysis (CA), Customer Classification (CC), Ant Colony Algorithm (ACA), Consumer，Behaviors Analysis (CBA)*

1. INTRODUCTION

Customer analysis is the basis of customer relation management, and customer classification is an important element for customer analysis. What we call customer classification means that according to customers' attributes, divide all the customers into different types, carry out categorized study on them, make corresponding service strategies, and reasonably allocate service resources, thus achieving customer retention and customer satisfaction improvement to the greatest extent [1].

To conduct customer relation management, we should divide enterprise's customers into different types through customer classification, then make and carry out corresponding customer service strategy on account of characteristics of different types of customers. According to the study of Sherden W. A., the most important 20% customers of the enterprise creates 80% profit of the enterprise, among which, however, 50% profit is offset by 30% negative profit customer. Due to limited resources of enterprise, it cannot provide service that satisfies all the customers, which leads to the situation that while striving to develop new customers, enterprise loses some old customers due to unsatisfactory service. Moreover, the cost for enterprise to develop a new customer is five times than that for maintaining an old customer, and the cost for persuading an unfaithful customer is ten times than that for maintaining an old customer; the profit will increase 25%~85% if the customer retention rate increases 5%. From this, it can be seen that customer retention exerts an amazing impact on profitability of enterprise. Hence, whether enterprise can or cannot hold old customers especially quality customers is the key to determine the extent of probability of enterprise, also an integral part of core competitiveness of enterprise. Nevertheless, the goal for customer retention is not to pursue zero loss, nor that the maximization of customer retention rate equals to the maximization of enterprise's profit. Among all the customers of enterprise, some may bring large profit for enterprise at present or in the future, which should be maintained; while others, no matter at present or in the future, not only cannot bring profit for enterprise, but also will become the burden of enterprise, for such customers, enterprise should reduce their service cost. So customer classification is the key for enterprise to conduct targeted marketing and service strategy [2, 3.4].

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

2. SUMMARIZATION OF CUSTOMER CLASSIFICATION STUDY METHODS

As to the study of enterprise customer classification methods, most foreign and domestic literatures have developed from the traditional qualitative and quantitative classification model to customer classification model based on data mining.

① Traditional customer classification methods mainly include qualitative customer classification model and quantitative customer classification model. (1) Traditional qualitative customer classification model, traditional qualitative customer classification method, mainly carries out the customer classification from the macroscopic view and customer one-sided information, or from the different emphases of value cognized by customers. Multi-factor customer classification method more often used method at present is RFM (Recency, Frequency, Monetary) analysis method, which is a factor analysis method based on customer transacting behavior analysis; RFM model lays too much emphasis on customers' behavior to differentiate customers, thus neglecting the significance of customer value, and resulting in error in customer classification. (2) Traditional quantitative customer classification model has more studies on customer classification based on customer value, which establishes customer classification model through selecting such three major factors influencing customer value as customer life cycle, customer average consumption each time and customer average consumption cycle [1].

However, with the intensifying competition among modern enterprises, business environment is increasingly tensed. Plus the common requirements on self resource limit and many other aspects, enterprise has to dig the value information behind the immense amount of customer information, classify customers precisely, and implement differentiation competitive strategy, only by this can keep the competitive advantage. Under the background of increasing data resources, the general statistical techniques adopted by traditional classification method to build classification model are obviously unable to meet management demand, especially that network enterprises are faced with enormous number of customers. Thus, customer classification method based on data mining emerges at the right moment.

②Customer classification model based on data mining, data mining acts as new technology of knowledge extraction and finding from data, among which K_means and neural network are the best performance and most widely-used clustering method, providing a new approach for customer classification. The advantage of this kind of method is mainly embodied in the strong capacity for processing mass data and high classification accuracy [2, 3, and 4].

Customer classification models based on data mining have high classification accuracy but leaves behind the question of slow convergence speed of its algorithm. Therefore, it is hard to put into effect in customer classification. The paper improves ant colony algorithm to control the accuracy of customer classification. In so doing, not only the problem of convergence speed of BPNN has been solved [5, 6], but also the simplicity of the model structure and the accuracy of the classification are ensured.

3. ONLINE CUSTOMER CLASSIFICATION INDICATOR SYSTEM DESIGN

Many scholars have carried out numerous studies on online customer value evaluation indicator system, and established the evaluation indicator system with different emphases and many similarities. According to the systematicity, scientificity, comparability, operability, relative independence among indicators and other basic principles, referring to foreign and domestic literatures, and on the basis of actual survey of network enterprise, and in view of the nature of trading and own characteristics of online trading, this paper adopts customer characteristics type variable and customer behaviors type variable in the specific selection of customer classification variables [2].

3.1 Selection Of Online Customer Characteristics Type Variable

Online customer characteristics type variable is mainly used for getting the information of customers' basic attributes. Such variable indicators as geographical position, age, sex, income of individual customer play a key role in determining the members of some market segment. This kind of variables mainly comes from customers' registration information and customers' basic information collected from the management system of banks, the contents of which mostly indicate the static data of customers' basic attributes, the advantage of which is that most of the contents of variables are easy to collect. But some of the basic customer-described contents of variables are lack of differences at times.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Based on analyzing and summarizing existing literatures, the customer characteristics type variables designed in this paper include: Customer No., Post Code, Date of Birth, Sex, Educational Background, Occupation, Monthly Income, Time of First Website Browsing, and Marital Status.

3.2 Selection Of Online Customer Behaviors Type Variables

Online customer behaviors type variables mainly indicate a series of variable indicators related to customer transacting behavior and relation with banks, which are used to define the orientation which enterprises should strive for in some market segment, and are the key factors for ascertaining target market. Customer behaviors type variables include the records of customers buying services or products, records of customer service or production consumption, contact records between customers and enterprises, as well as customers' consuming behaviors, preferences, life style, and other relevant information [4].

Based on analyzing and summarizing existing literatures, the customer behaviors type variables designed in this paper include Monthly Frequency of Website Login, Monthly Website Staying Time, Monthly Times of Purchasing, Monthly Amount of Purchasing, Type of Consumer Products Purchased, Times of Service Feedback, Service Satisfaction, Customer Profitability, Customer Profit, Repeat Purchases, Recommended Number of Customers, Purchasing Growth Rate.

4. DERIVATION OF ALGORITHM

4.1. Ant Colony Algorithm

Ant colony algorithm is a random search algorithm addressed by Scholar Dorigo and others from Italy in the 1990s, which solves TSP by manual simulation of ants search process, and achieves better results. It is an intelligent heuristic search algorithm applicable for combinatorial optimization problems after the simulated annealing algorithm, genetic algorithm, tabu search algorithm, neural network algorithm, etc. Not only can Ant colony algorithm perform intelligent search and global optimization, but also have the features of robustness, positive feedback, distributed computation and easy combination with other algorithms. At the same time, such characteristics as discreteness and parallelism of ant colony algorithm are very applicable to deal with digital image, and its clustering features and image recognition process have greater similarities, so theoretically it is feasible to study the image segmentation based on ant colony algorithm.

The study found that ants could leave a material called pheromone in the path to guide their own movement direction, and tend to move toward the high-strength material. Therefore, the collective behavior of a large number of ants can display the positive feedback of information: the more there are ants in certain path, the greater the probabilities that the subsequent ants select this path. Meanwhile the pheromone will disappear gradually over time, so the pheromone strength has the relations with the path length and the number of ants in the path. Ants search foods by such information exchange. Ant colony algorithm is a process to simulate the real ants to look for foods [7, 8].

But ant colony algorithm uses a random selection strategy in the process of construction solution. This selection strategy reduces evolutionary rate and easily causes stagnation, that is, the solutions that all individuals found are exactly the same not to search the solution space further and find a better solution after the search reached a certain extent. As for the problem, this paper adopts the selection strategy of the dynamic adjustment to improve the overall search speed and capabilities of ant colony algorithm [9].

The process of ants foraging is also a constant clustering process actually. Things are the cluster center. Ant colony algorithm is applied in the clustering problem, of which, main ideas are follows.

If $X=\{X_i \mid X_i=(x_{i1},x_{i2},\cdots,x_{im}), i=1,2,\cdots,N\}$ is the data set to analyze the clustering, r is the cluster radius, $ph_{ij}(t)$ is the pheromone concentration in the path from data X_i to data X_j in the time point t, d_{ij} is the Euclid distance with weight of the samples and the cluster center, and p is the weighted factor that can be determined according to the impact degree in the re-clustering of each component. See Formula 1.

$$d_{ij}=\|P(X_i-X_j\|=\sqrt{\sum_{k=1}^{m}p_k(x_{ik}-x_{jk})^2} \quad (1)$$

If $ph_{ij}(0)=0$ is the initial information quantity, see Formula 2.

$$ph_{ij}(0)=\begin{cases}1 & d_{ij}\le r\\ 0 & d_{ij}\succ r\end{cases} \quad (2)$$

According to the pheromone concentration in the path between the sample and the cluster center,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Formula 3 shows the probability ph_{ij} that X_i is merged to X_j.

$$ph_{ij} = \begin{cases} \dfrac{ph_{ij}^{\alpha}(t)\eta_{ij}^{\beta}(t)}{\sum\limits_{s\in S} ph_{ij}^{\alpha}(t)\eta_{ij}^{\beta}(t)} & j \in S \\ 0 & others \end{cases} \tag{3}$$

Of which, $\eta_{ij}(t)$ is the heuristic guide function, which represents the expectation degree that ant X_i is transferred to X_j. The use of heuristic function can reflect the similarity among the pixels. The greater the heuristic function, the greater the probability that the pixel is merged to the same cluster. α and β are the information accumulated in the process of pixels cluster and the impact factor that heuristic function selects the path respectively. is the $S=\{X_s \mid d_{sj} \le r, s=1,2\cdots, j, j+1\cdots, N\}$ set of feasible path.

If $P_{ij}(t) \ge P(0)$, X_i is merged to X_j. If $C_j = \{X_k \mid d_{kj} \le r, k = 1,2,\cdots J\}$, C_j is all data sets merged to X_j. See Formula 4 for the optimal cluster center.

$$\overline{C_j} = \frac{1}{J}\sum_{k=1}^{J} X_k \qquad X_k \in C_j \tag{4}$$

With the movement of ants, there is the change in the amount of pheromone from each path. The pheromone from each path is adjusted in accordance with the overall adjustment rules after one cycle. See Formula 5.

$$ph_{ij}(t) = \rho p_{ij}(t) + \Delta ph_{ij} \tag{5}$$

In Formula 5, E is the attenuation coefficient of the pheromone with the passage of time, generally from about 0.5 to 0.99. E is the pheromone increment of the path in this cycle. See Formula 6.

$$\Delta ph_{ij} = \sum_{k=1}^{N} \Delta ph_{ij}^{k} \tag{6}$$

Δph_{ij}^{k} shows the pheromone amount that the k ant left in the path in this cycle.

4.2. Improvement Of Ant Colony Algorithm By The Overall Search

4.2.1. Dynamic adjustment strategy

Stagnation is the fundamental cause resulting in the inadequacy of ant colony algorithm. Based on the deterministic and random selections, this paper adjusts the transition probability dynamically to build the selection strategy more conducive to the overall search.

The pheromone in the path has continuous change in the evolutionary process. The pheromone of better solution searched is strengthened to increase the selection possibility of next iteration, and some better solutions is forgotten gradually because fewer ants pass in the start-up phase so as to affect the overall search capabilities of the algorithm. If the ants are stimulated properly to try the path occasionally in the selection strategy, it is conducive for the overall search of the solution space. Thus, the inadequacy of basic ant colony algorithm is overcome effectively. See Formula 7 for the improved selection strategy in this paper. when $q \le q_i$, $j \in allowed_k$, there is formula 7, and when $j \in allowed_k, others$, there is formula 8

$$P_{ij}^{k}(t) = \arg\max\{|\tau_{ij}(t)|^{\alpha} \cdot |\eta_{ij}(t)|^{\beta}\} \tag{7}$$

$$P_{ij}^{k}(t) = \frac{[\tau_{ij}(t)]^{\alpha} \cdot [\eta_{ij}]^{\beta} \cdot X_{ij}(t)}{\sum\limits_{k\in allowed_k} [\tau_{ik}(t)]^{\alpha} \cdot [\eta_{ik}]^{\beta} \cdot X(t)} \tag{8}$$

In Formula 7 and formula 8, X_{ij} will meets the requirements of Formula 9.

$$X_{ij} = \frac{m \cdot N_c}{m \cdot N_c + \delta \cdot Q_c(i,j) \cdot \eta(i,j)/\max\eta} \tag{9}$$

In Formula 9, m is the number of ants, N_c is the number of current iterations, $\max\eta$ is the maximum of heuristic function $\eta(i,j)$,and $Q_c(i,j)$ is total number of ants in the current path (i,j) from the first iteration. Q_c and η are considered in X. When previous iteration tends to suboptimal solution, the number $Q_c(i,j)$ of ants increases and its X value decreases constantly in spite of constant increase of the pheromone in the suboptimal solution. Therefore, another selection of the path can restrain the excessive increase of the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

pheromone to cause premature convergence, and is conducive to global convergence.

4.2.2. Improved selection strategies

To study the improved selection strategy's impact on convergence rate, the first is to analyze the relation and difference between local update rules and improved selection strategy in the traditional ant colony algorithm. Local update rules are performed after the ants complete a transition, in essence make the subsequent ants avoid repeating the same path, and function the restraint of the same path convergence. Although the improved selection strategy and local update rules restrain premature convergence effectively, their mechanism is not the same and the impacts on the pheromone are also different. Local update rules change the pheromone directly and permanently, but selection mechanism does not change the pheromone directly. In addition, the selection mechanism of dynamic adjustment begins to stimulate the ants to search new path in the path selection; local update rules are to guide the ants to search other path from the feedback by the pheromone adjustment. Thus, the selection mechanism is more advanced than local update rules in the restraint of premature convergence, and faster in the convergence rate. Traditional algorithm mostly uses the local update rules to curb the prematurity. Although the improved ant colony algorithm does not abandon the local update rules completely, the selection strategy of dynamic adjustment is introduced to replace partial effects of local update rules. Dynamic adjustment of choice strategy reduces the effect of local update rules, and new advanced selection strategy is also advanced correspondingly to stimulate the ants, so these are the fundamental reasons to improve the ant colony algorithm and the convergence rate.

Table 1 Customer Classification Result of Some Websites

Customer Type	Number of Customers	Percentage %	Profit Contribution Proportion
Very high	2870	7.00	52.44
High	5933	14.42	31.63
Medium	10321	25.09	13.89
Low	14721	35.79	6.17
Very low	7291	17.72	-4.13
Total	41136	100.00	100.00

Table 2 Classification Performance Comparison of Each Algorithm

Algorithm	Algorithm in This Paper	K-means Algorithm	Ordinary BP Neural Network Algorithm
Accuracy Rate	99.84 %	88.44%	92%
E Value	102.66	159.71	121.23

5 . EXPERIMENT CONFIRMATION

5.1 Classification On Customer Value

According to the selected classification indicators and values after dimensionless disposal on indicators, making use of established ant colony algorithm, we can carry out comprehensive evaluation on customer value. According to customer comprehensive evaluation result, we can classify all the customer value involved in evaluation into such six grades as very high, high, medium low and very low., and their customer values are 0.85—1, 0.70—0.85, 0.50—0.70, 0.5—0.35, under 0.35 respectively.

5.2 Calculation Of Ant Colony Algorithm

This paper, utilizing C language, writes weight training and optimization program (program omitted), taking 41136 customer values of certain network enterprise of 22 indicators as training sample of ant colony algorithm for customer value classification, i.e. dividing correspondingly each single indicator into five grade categories as very high, high, medium, low, extremely low, selecting the same grade classification value as the grade value of classification on customer value. When all 22 single indicators are very high, the customer comprehensive evaluation result will certainly be the highest, which is the same case with other grades. As each indicator is a numerical interval, before conducting network training, indicators must be processed as follows: for the indicator values of the grades of very high, high, medium, ordinary and low, the average value should be taken as input value and for that of the grade of extremely low, the critical value should be taken as input value. The

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

desired output of network is given as below: 0.90, 0.78, 0.60, 0.45, and 0.35 for five grades of very high, high, medium, low and extremely low. The criterion for optimal training effect of ant colony algorithm should be that the sum of difference of squares of evaluation value of the sample network and desired output of given network is less than 10.3. Through ant colony algorithm training and optimization, the algorithm presented in the paper is realized.

5.3 Application Example

Making use of the trained ant colony algorithm, we carry out comprehensive evaluation of 41336 customer values of certain network enterprise. Due to limited space, we only give each indicator value of customer evaluation, ignoring the specific introduction of each customer. Each indicator value of customer evaluation and evaluation results are shown as Table 1.

We can see from Table 1 that in the autonomous learning of algorithm of this Paper, such five factors as the educational background, income, occupation, times of purchasing, and total amount of purchasing of customers have a relatively great influence on customer classification. Through the classification result in Table 1, it can be seen that very high customers take up 7.00% of the total number of customers, while the profit takes up 52.44% of the total profit. These customers play a significant role in the existence and development of enterprises. However, the very low negligible customers account for 17.72%, who not only do not bring profit to enterprises, but also make enterprise lose money. These customers should be either further cultivated or eliminated according to the actual situation.

We can see from table 2 that the cluster accuracy rate of algorithm in this paper is the highest, reaching 99.84 %, obviously higher than K-means algorithm and Ordinary BP Neural Network algorithm; the square errors and E values on customer classification of three algorithms are 102.66, 159.71 and 121.23 respectively. The smaller the E value is, the smaller the possibility of wrong classification is. Thus it can be seen that the square error and E value of the algorithm in this paper during the classification are far more less than ordinary K-means algorithm [4] and BP Neural Network algorithm [6]. Therefore, it shows that the improvement ant colony algorithm in this paper turns out to be a success, with reasonable classification results.

6. CONCLUSION

This paper, through improving ant colony algorithm making use of the advantage in high customer classification accuracy, also overcoming the actual defect in poor algorithm convergence, builds new customer classification model based on ant colony algorithm, also analyzes and establishes a set of classification indicator system for comprehensive evaluation of network customer value, carrying out comprehensive evaluation and classification study on customer value of network enterprises' users. The experimental results show that the model classification results are satisfactory. The model in this thesis has the following superiorities compared with other methods. ① Through self-study on samples involving in the comparison, ant colony algorithm structure can be decided, repeatedly iterating according to the criterion of optimal training, constantly adjusting ant colony algorithm structure, until reaching a relatively stable status, thus, the utilization of that method eliminates many human factors, helping to ensure the objectiveness of the results; ② High accuracy, able to make system error reach the requirement of any accuracy with convergence;③ Good dynamics, self-study and dynamic tracking ability will be stronger with the progress of time and the increase of samples involved in comparison. Hence, there is certain practical application value in that method.

ACKNOWLEDGMENT

This work is supported by the 52nd Chinese Post doctoral Fund under the grant number. 2012M521284.

REFRENCES:

[1] H. Zhou, "Study of Classifying Customers Method in CRM", *Computer Engineering and Design*, Vol. 20, No. 3, March 2009, pp. 659-661.

[2] W. B. Deng and Y. Wang, "B2C Customer Classification Algorithm Based on Based on 3DM", *Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition)*, Vol. 21, No. 4, April 2009, pp. 568-572.

[3] Y. H. Guan, "Application of Improved K-means Algorithm in Telecom Customer Segmentation", *Computer Simulation*, Vol. 28, No. 8, August 2011, pp. 138-140.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[4] X. Y. Qu and, "Application of K-means Based on Commercial Bank Customer Subdivision", *Computer Simulation*, Vol. 28, No. 6, July 2011, pp. 357-360.

[5] B. Z. Yang and T. Gen, "Research on Customer Value Classification Based on BP Neural Network Algorithm", *Science and Technology Management Research*, Vol. 23, No. 12, December 2007, pp. 168-170.

[6] S. F. Ding and W. K. Jia, "An Improved BP Neural Network Algorithm Based on Factor Analysis", *Journal of Convergence Information Technology*, Vol. 5, No. 4, April 2010, pp. 103-108.

[7] Y. D. Zhang and W. Lenan, "A Novel Genetic Ant Colony Algorithm", *Journal of Convergence Information Technology*, Vol. 7, No. 1, January 2012, pp. 268-274.

[8] S. W. Zhang and J. Z. Lin, "A New Hybrid Ant Colony Algorithm for Solving Vehicle Scheduling Problem", *International Journal of Advancements in Computing Technology*, Vol. 4, No. 5, May 2012, pp. 17-23.

[9] J. P. Wang, "A Multi-Task Scheduling Method Based on Ant Colony Algorithm Combined QoS in Cloud Computing", *Advances in Information Sciences and Service Sciences*, Vol. 4, No. 11, November 2012, pp. 185-192.