ISSN : 0974 - 7435 Volume 10 Issue 13

2014

BioTechnology

An Indian Journal

FULL PAPER

BTAIJ, 10(13), 2014 [7136-7141]

The improved credit card customer behavior clustering analysis based on ant colony algorithm

Yingchun Liu
Beihang University, Beijing 100083, (CHINA)
E-mail : liuyingchun2@126.com

ABSTRACT

With the development of credit card, the banks need to classify the credit card customers with the advanced data mining technology, and then take different measures to target different customers. An improved K-means algorithm based on pheromone is proposed, which works with the transformation probability to realize the clustering and has reduced the number of the parameters and improved the speed of clustering. At last, the proposed algorithm is tested and used to analyze bank credit card customer spending behavior.

KEYWORDS

Credit card; Clustering analysis; Ant colony algorithm; Pheromone.

INTRODUCTION

The phenomenon about "the data explode but knowledge is poor" caused by rapid development of information technology and network technology is serious day by day. Under this environment, data mining was put forward as a kind of new technology, which demonstrated its strong vitality more and more[1,2]. In some domestic primary website, data mining technology was chosen one of the most popular technologies in the future[3]. Data mining is a research field of crossing multi-disciplines, and its development will influence the process of the global informatization greatly[4]. So it is the desirability of informatization development to have a research on data mining technology in an all-round and systematical way. Clustering analysis as one of data mining analysis methods is the performance is prominent, including the pattern recognition, the data analysis, the imagery processing, as well as the marketing research and so on many domains obtain the widespread application[5]. This dissertation deeply studies and analyses the data mining technology, especially the one for clustering analysis. And then some ideas and improvements have been proposed.

In the next section, we introduce basic idea of K-means clustering algorithm[6-7]. In section 3, improved K-means scheme based on pheromone of ant colony algorithm [8] is proposed. In section 4, cluster analysis of customer behavior is given. In section 5, the paper is concluded and some remarks is given.

K-MEANS CLUSTERING ALGORITHM

The clustering algorithm based on division can be described as follows. Given a space of d dimension, define a evaluation function $C:\{X:X\subseteq S\}\to R+$, which evaluate each clustering quantitatively. For a object set S of input R_a and an integer. The output is a division of S, $S_1,S_2,\cdots,S_k$, which makes target function minimum.

$$J=\sum_{i=1}^{k}C(S_i)$$

The process of k-means algorithm is as follows.

(1) Random selection of k objects are taken as the initial clustering center.

(2) Repeat (3) and (4), until every clustering center has no longer any change;

(3) Calculate the distance between each object to the center of the cluster, the object is attributed to the class, which has the nearest distance to the object.

(4) Calculate the new clustering center, that is calculate the average value of each class object. Use average value point of the object as a new clustering center.

But the traditional K-means algorithm exists two inherent drawbacks.

(1)The random selection of the initial clustering center may lead to different clustering results, and even there exists the situation of no solution.

(2)This algorithm is climbing mountain search algorithm, which is based on the tonsure decline algorithm, thus it is inevitable to fall into local minima.

The two faults greatly limits its application scope. Here an improved K-means method based on the pheromone is proposed.

IMPROVED K-MEANS ALGORITHM BASED ON PHEROMONE

When ants foraging, the process of information transmission is mainly done through the pheromone diffusion, and pheromone determines the movement direction of the ants. When the front ants affect the behavior of the back ants, the back ants are not on the trajectory of the front ants. They keep a certain distance with this trajectory. When the distance is small, the behavior of back ant is

affected greatly, and this feature is useful for data clustering. Using this principle, this paper proposes an improved K-means algorithm based on pheromone. The algorithm takes data objects as ants with different attributes, clustering center is taken as the food source that the ants what to look for, in this way, data clustering process can be regarded as the process of ants looking for food.

Improved K-means algorithm based on pheromone uses the characteristics of the ant colony distributed search to improve the traditional K-means algorithm, which is easy to fall into local optimum. When the ants are on the way form food source i to food source j, if they find the right path, they releases the corresponding concentration of pheromone. The pheromone on the one hand directly affect the ants of the two clustering centers, on the other hand it will spread taking this path as the center. It can affect the behavior of the other ants nearby, make them choose this path with greater probability in the next step when looking for a path. Through this way of collaboration based on pheromone, other data objects are affected slightly when choosing the clustering center, thus can improve the convergence speed of the algorithm. In the next round of cycle, the clustering deviation is introduced to update the clustering center, until deviation meets certain condition. The proposed algorithm is as follows. $X=\{X_i \mid X_i=(x_{i1},x_{i2},\ldots,x_{in}), i=1,2,\ldots,n\}$ represents the data set for clustering. $\overline{C_j}$ represents clustering center, the initial value of which is randomly distributed. At time t, for other data object l, the amount of pheromone of the k-th ant from food source i to food source j is $\tau_{ij}^k(t)$. At this time, ants spread pheromone taking i and j as the center. The information of data object l produced by ant k is $\tau_{il}^k(t)$, $\tau_{jl}^k(t)$.

$$\tau_{ij}^k(t)=\begin{cases}\dfrac{Q}{d_{ij}}, & d_{ij}\le R\\ 0, & d_{ij}>R\end{cases}$$

$$\Delta\tau_{il}^k(t)=\begin{cases}r\cdot Q/d_{ij}\left(1-\dfrac{d_{il}\cdot(d_{il})^w}{\overline{d}^{w+1}}c\tan\theta\right), & d_{il}\le R\\ 0, & otherwise\end{cases}$$

$$\Delta\tau_{jl}^k(t)=\begin{cases}r\cdot Q/d_{ij}\left(1-\dfrac{d_{jl}\cdot(d_{jl})^w}{\overline{d}^{w+1}}c\tan\theta\right), & d_{jl}\le R\\ 0, & otherwise\end{cases}$$

d represents the Euclidean distance between the data object to clustering center $\overline{C_j}$, w represents a constant, r is a constant less than 1, and R represents the preset clustering radius. P_{ij} represents the probability of data object X_i belonging to clustering center $\overline{C_j}$.

$$P_{ij}=\frac{\tau_{ij}^{\alpha}(t)\eta_{ij}^{\beta}(t)}{\sum_{s\in S}\tau_{sj}^{\alpha}(t)\eta_{sj}^{\beta}(t)}$$

$S = \{X_s \mid d_{sj} \le R, s = 1, 2, \ldots, n\}$ represents data object set which is near to the clustering center $\overline{C_j}$. α represents relative important degree of residual information, and β represents relative important degree of expected value. η_{ij} represents some heuristic information of ant from i to j at time t.

$CS_j = \{X_i \mid d_{ij} \le R, i = 1, 2, \ldots, J\}$ represents data object set which belonging to clustering center $\overline{C_j}$ and J represents the number of object in the set. The deviation of the j-th clustering is

$$\overline{C_j} = \frac{1}{J}\sum_{i=1}^{J} X_i \cdot$$

$$\xi_J = \frac{1}{J}\sum_{i=1}^{J} (X_i - \overline{C_j})$$

$$\xi = \sum_{J=1}^{k} \xi_J$$

CLUSTER ANALYSIS OF CUSTOMER BEHAVIOR

Customer behavior analysis is an important part of decision analysis in customer relationship management (CRM). Customer behavior analysis analyzes the overall and specific customer group, and natural social characteristics, the consumer behavior and the trend are known. The customer behavior analysis of customer relationship management can be divided into overall behavior and group behavior analysis. Overall behavior analysis is used to find all customers behavior rule, but only the overall behavior analysis is not enough, customer differ in thousands ways, depending on the customer behavior, they can be divided into different groups, and each group has obvious behavior characteristics. Through customer behavior analysis, CRM users can better understand customers to find the group customer behaviors. Based on the understanding and laws, market experts can develop the corresponding marketing strategy. In customer analysis, it is obviously very important for large customer analysis. In addition, the customer classification has an extreme case, and each category of customer is only one, that is One To One marketing. In this section, clustering algorithm is applied to the bank credit card customer consumption data sets, the described algorithm is used to realize the customer segmentation. Data is taken from some province bank credit card users, that consume more than 100 yuan in a month. Fee includes different purchases of users to use a credit card consumption. Some user uses credit card frequently and consumption is large, some users uses card with low frequency and low consumption. The number of sample is 1000. Because clustering algorithm has characteristics of parallelism, robustness and automatically generated clustering number, combining with bank credit card user consumption data analysis, the customer groups of different characteristics are discovered. It can create conditions for bank to promote the new card and improve the contribution rate of customers. It also can maximize profits of the bank.

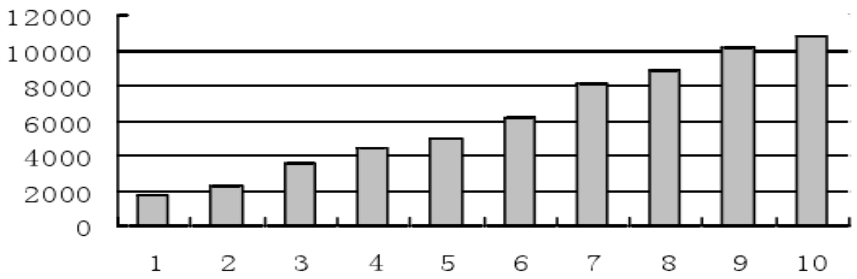

Figure 1 : average consumption clustering results of proposed algorithm

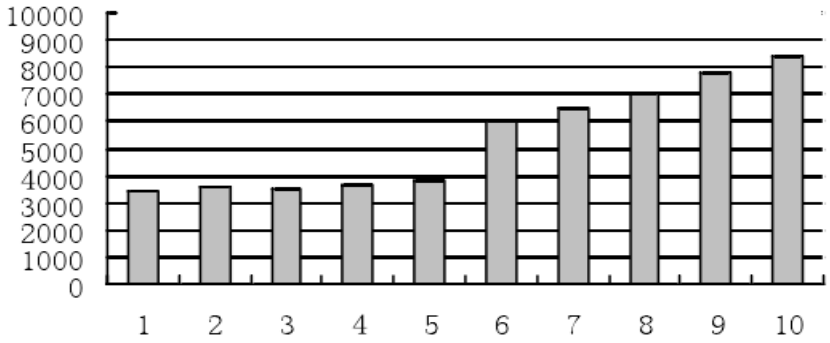

Figure 2 : average consumption clustering results of K-means algorithm

Average consumption clustering results of proposed algorithm is shown in Figure 1 and average consumption clustering results of K-means algorithm is shown in Figure 2. When the number of clustering centers is similar, the clustering algorithm based on ant colony is more clear and has better clustering effect compared with K-means clustering algorithm.

Consumption average values above 8000 are category of 7, 8, 9, and 10. Above 10000 yuan, there are two classes (9, 10). In Figure 2, the average consumption value above 8000 yuan is only class 10, and there is no class above 10000 yuan. Ant clustering algorithm is better for analysis of big customer with large consumption. It will be good for analysis of the special customer in one to one marketing, combining the customer's other natural attributes such as type of work and income, customer group can be analyzed from multiple perspectives.

CONCLUSION

With the development of finance economy, the domestic credit card market has gradually grown up and been more and more prosperous. In recent years, more and more people use credit card to make payments and the credit card has become the most popular non-cash payment instruments in china. As a result, the banks need to classify the credit card customers with the advanced data mining technology, and then take different measures to target different customers, such as the different marketing, service, risk control measures etc. At the same time, the banking system has accumulated a large scale of business data which has provided a basis for the application of data mining. The local commercial bank should seize the opportunity to apply the data mining technology to help the decision making and the providing of different service.

An improved K-means algorithm based on the pheromone is introduced clearly. The algorithm works with the transformation probability to realize the clustering. It has reduced the number of the parameters and improved the speed of clustering. At last, the proposed algorithm is tested by some datasets and used to analyze bank credit card customer spending behavior.

REFERENCES

[1] Nan-Chen Hsieh; “An integrated data mining and behavioral scoring model for analyzing bank customers,” Expert Systems with Applications, **27,** 623-633 **(2004)**.

[2] J.Handl, J.I.Cnowies, M.Dorigo; “Ant-based clustering: a comparative study of its

[3] relative.performance with respect to k-means, average link and ld-som,” IRIDIA,University LibradeBruxelles, **(2004)**.

[4] Bao-Jiang Zhao; “An Ant Colony Clustering Algorithm Machine Learning and Cybernetics,” 2007 International Conference on Publication Date:19-22 **7**, 3933-3938 **(Aug. 2007)**.

[5] M.Dorigo, E.Bonabeau, G.Theraulaz; “Ant algorithm and stigmergy,” Future Generation Computer System, **16(8)**, 851-871 **(2000)**.

[6] Jianfu Li,Wei Zhang; “Solution to Multi-objective Optimization of Flow Shop Problem Based on ACO

Algorithm," Computational Intelligence and Security, 417-420 **(2006)**.

[7] Da-meng Dai, De-Jun Mu; "A Fast Approach to K-means Clustering for Time Series Based on Symbolic Representation," International Journal of Advancements in Computing Technology, **4(5)**, 233-239, **(2012)**.

[8] Xuexia Yang; "Expansion Research on K-means Clustering Algorithm Based on Rough Set," International Journal of Advancements in Computing Technology, **4(10)**, 221-227 **(2012)**.

[9] J.Handl, J.Knowles, M.Dorigo; "Strategies for the Increased Robustness of Ant-Based Clustering," Springer-Verlag Berlin Heidelberg, 90-104 **(2004)**.