

## Classification of debit card customers based on AHP and K-means

Chunxin Jiang<sup>1,a,\*</sup>, Weidi Dai<sup>1</sup>, Wenjun Wang<sup>1</sup> and Yunqiang Jia<sup>1</sup>

<sup>1</sup>School of Computer Science and Technology, Tianjin University, Tianjin 300072, China;

<sup>a</sup>dangyang@tju.edu.cn

**Abstract:** Debit card business is a very important business, bank debit card and bank users the most on the one hand, so every year the debit card user transaction data is very huge, in what is now the era of big data, data is wealth, we have huge debit card transaction data, you need to use big data technology to deal with the data analysis, found that there were a rule, and conducive to the development of bank data and results are obtained. In this article, through analysis of transaction data of debit card users, debit card users are classified, and the bank large customers and high quality clients and has great potential customers, and puts forward the time were significantly abnormal in the transaction data

**Keywords:** Debit card, customer classifies.

### 1. Introduction

In the current market environment, who can grasp the customer demand trends, explore effective customer resources, improve the relationship between the customers and the bank, will have an inherent advantage. It is well known that there exist "and" law of banking, the bank customer contribution of 80% of the profits by 80%, Banks can take advantage of the data to find their own 20% of high-quality customers, and through a variety of preferential services to improve their loyalty, thus improve the efficiency of <sup>[1]</sup>.

Customer classification using clustering analysis technology, clustering analysis, as a branch of statistics, has been widely studied for many years, mainly in the distance-based clustering analysis, based on k-means, K - center and some other methods of clustering tool has been added to the many statistical analysis software package or <sup>[2]</sup> in the system.

In the face of the many attributes of bank customers, we cannot blindly determine the weight of each attribute. This article USES AHP to evaluate the weights of each property. The component factors, AHP is decomposed into complex problems according to these factors will govern relations constitute a class hierarchy, by paired comparison method determine the importance of various factors in the layer, and then comprehensive decision makers, determine the weight of their <sup>[3, 4]</sup>.

### 2. Datasets

We are using a bank debit card water data, we first analyze the table structure of a data set, we need to get our customer classification properties, we programmed for each customer's different amount, frequency, and stability data. Here we used the bank for about 60 million records over three years. It is found that the number of withdrawal times of the user is consistent with the retailed distribution.

### 3. Methods

Because the bank debit card Numbers, the statistics found that users of different difference is bigger, and does, k means clustering amount is the difference of the serious influence the reliability of classification, according to amount differences problem, we combined with the actual situation to the deposit amount from ten thousand to ten thousand between the customer classification.

Here we calculate the amount, the amount, and the stability of the user's deposit and withdrawal. Then, the weights of the properties are calculated by using the AHP hierarchical clustering method, and the user's deposit and withdrawal scores are calculated, and the user is classified by the k-means cluster<sup>[5,6,7]</sup>.

### 3.1 Analytic hierarchy process (AHP)

Analytic hierarchy process (AHP) is operations research by the United States home A.L.Saaty in the early 1970 s, at the time of the us department of defense research topic proposed a hierarchical weighted decision analysis method. Analytic hierarchy process is a systematic, hierarchical analysis method which is qualitative and quantitative. It is because of its practicability and effectiveness in dealing with complex problems that are soon to be valued and applied worldwide.

Application of the analytic hierarchy process (AHP)<sup>[8,9,10]</sup> can be divided into three steps: first step, set up for the class hierarchy, to solve the problem is mainly to various factors of the causality analysis, and then isolated elements levels to implement the structure to establish; The second step is to construct the judgment matrix, which is achieved through the importance of comparing each element with two or two. The third step is to calculate each element's weight

### 3.2 Customer classification

We've already received a credit card for debit card users in 2013, 2014 and 2015, meaning that every debit card user has two points (withdrawal, deposit). Now, we need to do the k-means method to cluster the user, and by k-means we get the following three graphs for user categories:

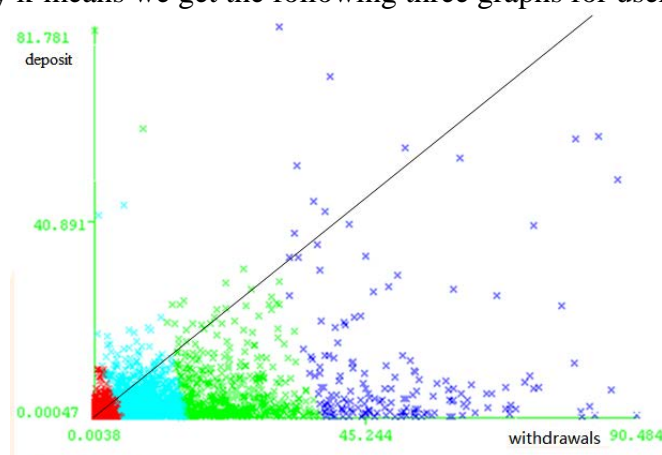


Figure 3-1 Customer clustering in 2013

Figure 3-1 in 2013, is the user accesses a scoring figure, our users can be divided into four categories, including red is a different class of scores are low, belongs to the low value customer, their different amount and the number of less; Shallow blue and green part belong to potential customers, their withdrawal amount and number in the middle, and finally the blue part is high quality clients, their highest score withdrawal. There is a black line in the picture, under the black line is withdrawn score is greater than the scores of users, black line above is score higher than the score users withdrawal of deposit. From the picture we can get some point from the group, they score many points deposit is quite high, but a withdrawal score in the second or third class, and then in the fourth class there are a lot of point in their deposit has a high score, the point is we need to find a good point

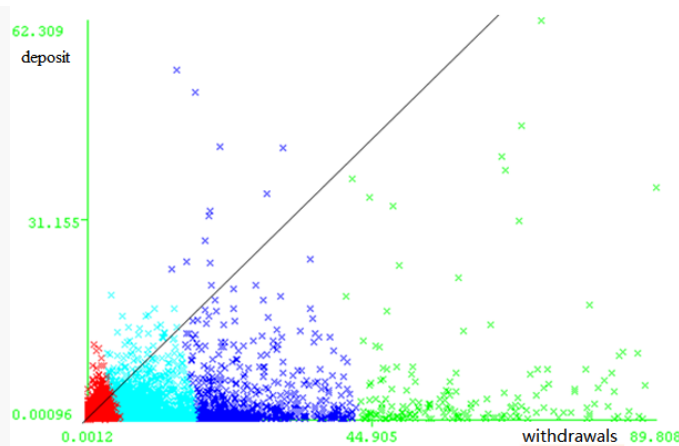


Figure 3-2 Customer clustering in 2014

Figure 3-2 is a different points of view of the user in 2014, we have the same user is divided into four categories, the red part is still low value customers, their different scores are low; And part of the green part is the highest score withdrawal, there are still some outliers in the figure, their deposits score higher than the score of withdrawal, but through the contrast of 13 years and 14 years can be found that 13 years in the fourth class: quality class there were higher than that of the black point, and 14 years in the fourth class is not. And 13 years in 14 points in the fourth class of many years into the third class, their score decreased a lot, withdrawal and deposit score has declined, at the same time, we can found at roughly 14 years overall user accesses are falling.

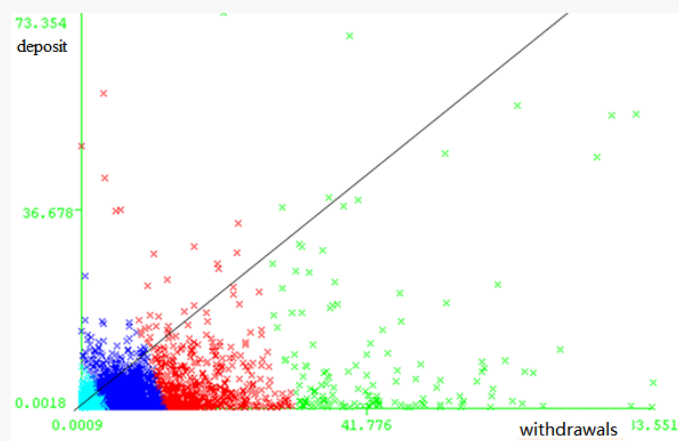


Figure 3-3 Customer clustering in 2015

Figure 3-3 in 2015, is the user accesses a chart, we are also the user can be divided into four categories, including shallow blue for the low value customers, green scores for withdrawal of the highest quality customer, can be found that there are some points in figure in the black line above, and the classification of 13 14 15 years three years on the analysis, we can find that there are a few points more than they have been in the black line in their deposit score higher than the score of withdrawal, and access the score of the floating in the guarantee under the condition of less, their withdrawals have dropped significantly, from 13 to 15 years of the second highest class, these points are very good quality for our bank.

From the above three graphs we can find the classification of each year are very similar, but each year the inside of the four types of users have some differences, by classifying each user every year, we get the list for five respectively, were three years in the first category of users, three years in the second category of users, three years in the third class, three years in the fourth class of users and three years of change in the category of users. Three years 75.36% of the users are the most inactive first class, three years only 1.32% of the users are all in the third and fourth class of the active, but there are three years 18.07% of the users are classified to change.

#### 4. Conclusions

We use AHP to deal with bank debit card users trading data analysis to obtain weights between attributes, then K - means clustering with the customers, through our 13 14 15 years three years of clustering analysis, and got some high quality clients, they are particularly high, the withdrawal of our bank for the customer for the corresponding business promotion to let them reduce the withdrawals, also got some potential clients their deposits and withdrawals are all belong to the middle, we can carry out the corresponding activities increasing their deposits, at the same time we also got the deposit is higher than the high quality of customer withdrawal, every day they brought me a large deposit, we can tell the user to join the monitoring list, such as increased later found to reduce their deposits and withdrawals, warning, let our early found early treatment. In this paper, we use the bank's related data processing analysis, behind we should consider using data from the different aspects of the comprehensive treatment, in order to reflect the all-round ability of the user.

#### 5. Acknowledgments

This work was financially supported by the National Social Science Fund(15BTQ056).

#### 6. References

- [1] Kanungo T, Mount D M, Netanyahu N S, et al. An efficient k-means clustering algorithm: analysis and implementation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2002, 24(7):881-892.
- [2] Arthur D, Vassilvitskii S. k-means++: the advantages of careful seeding[C]// Eighteenth Acm-Siam Symposium on Discrete Algorithms. Society for Industrial and Applied Mathematics, 2007:1027-1035.
- [3] Saaty T L. Modeling unstructured decision problems — the theory of analytical hierarchies[J]. Mathematics & Computers in Simulation, 1978, 20(3):147-158.
- [4] Saaty T L, Vargas L G. The Seven Pillars of the Analytic Hierarchy Process[M]// Models, Methods, Concepts & Applications of the Analytic Hierarchy Process. Springer US, 2001:27-46.
- [5] Hartigan J A, Wong M A. A K-means clustering algorithm[J]. Applied Statistics, 1979, 28(1):100-108.
- [6] Kang S H, Sandberg B, Yip A M. A regularized k-means and multiphase scale segmentation[J]. Inverse Problems & Imaging, 2017, 5(2):407-429.
- [7] Wagstaff K, Cardie C, Rogers S, et al. Constrained K-means Clustering with Background Knowledge[C]// Eighteenth International Conference on Machine Learning. Morgan Kaufmann Publishers Inc. 2001:577-584.
- [8] Tam M C Y, Tummala V M R. An application of the AHP in vendor selection of a telecommunications system[J]. 2001, 29(2):171-182.
- [9] Saaty T. Fundamentals of Decision Making and Priority Theory with the AHP[J]. 1994.
- [10] Felix T. S. Chan, N. Kumar, M. K. Tiwari, et al. Global supplier selection: A fuzzy-AHP approach[J]. International Journal of Production Research, 2008, 46(14):3825-3857.