Classification of Discrete Customer and Re-Raking of Classified Data

C. Selvi[1], A. Aadhishri[2], J. Arun[3], K. Deepika[4]

[1]Associate Professor, Computer Science and Engineering, VCET, Tamilnadu, India. Email: selviilango.cse@gmail.com

[2]Student, Department of Computer Science and Engineering, VCET, Tamilnadu, India. Email aadhishri1599@gmail.com

[3]Student, Department of Computer Science and Engineering, VCET, Tamilnadu, India. Email: arunjarone@gmail.com

[4]Student, Department of Computer Science and Engineering, VCET, Tamilnadu, India. Email: deepikakamaraj1998@gmail.com

Abstract - Collecting correct information under the background of big data can help industry to classify customers more accurately. Outlier data includes important customer information. In order to know about the customer classification problem based on customer asset outlier data, a customer classification model based on outlier data analysis concerning customer asset is constructed successfully. The model has Variables in 4 dimensions that are frequency of transactions, product types and services traded amount of transaction and Age of client. And using clustering before dividing the twenty-five types of customer data into four categories and corresponding marketing strategies also are put forward according to different classification of customer company's data. It also presents a flexible and effective re-ranking method, called CR-Re-ranking, to improve the retrieval effectiveness. To offer high accuracy on the top-ranked results, multi modal fusion re-ranking approach is used. Experimental results show that the quality, especially on the top-ranked results, is improved significantly.

Keywords: Discrete data, Customer Classification, Customer Assets, Data Mining.

I. INTRODUCTION

With the arrival of the era of big data, enterprises' data has formed a certain scale in the field of marketing. Its diversity, low-value density and real-time complexity are both challenges and opportunities for marketing. In marketing management classifying customer management is one of the core issues of enterprise operation. Identifying and owning excellent customers, and developing and maintaining customers in a targeted manner, not only avoids the waste of resources and higher costs caused by the decentralization of energy, but also reduces the huge risk of blind marketing. By collecting the data generated by customers and enterprises at the contact end, the effective customer classification not only filters the data interference of customers who do not have transaction relationship with enterprises in the market, but also avoids the legal risk of infringing customer privacy. In the process of customer data mining, outlier data are often encountered. They are inconsistent with the law embodied in the overall data representation level. They are free from most of the intervals and are usually considered as noise data or abnormal data to be eliminated. However, as objective data, the way of processing is obviously inappropriate. Therefore, how to filter data from mass data and how to use data mining algorithm to complete the value "purification" of customer data and find important customers have become the urgent problems to be solved in the marketing field under the background of large data. On the basis of relevant research, a customer classification model is constructed in this project based on customer asset outlier data analysis, and corresponding marketing strategies are put forward for different customer classification. Based on the traditional customer classification model, the age dimension is added to the three dimensions including transaction frequency, types of products or services traded and transaction amount, and the customer classification model based on customer asset outlier data analysis is constructed and customer information is deeply excavated from the perspective of outlier data, and customer classification is carried out. If we consider that the final aim of search engines is to meet users' information needs, it is reasonable to take user satisfaction and user behavior into account when designing a search engine. According to the analysis, users are rarely patient to go through the entire result list. Instead, they usually check the top-ranked documents. Analysis on click-through data from a very large Web search engine log also reflects such preference. Therefore, it is more crucial to offer high accuracy on the top-ranked documents than to improve the whole search performance on the entire result list. As an alternative scheme, the reranking method can improve search quality by reordering the initial result list. Although the total number of relevant documents remains fixed after reranking, the precision improvement at the low depth of the result list can be expected by forcing true relevant documents to move forward.

II. SCOPE AND OBJECTIVES

- To make the fusion is carried out to the intermediate clusters only and so the calculation time and overhead is reduced much.
- To improve accuracy of Re-ranking of result.
- To take location data of customer so that village, town or city based customer grouping is possible.
- To make suitable even if the retailer is situated in any type of location (village or town or city).

III. EXISTING SYSTEM

In the existing system, the customer classification in the modeling idea is being done as follows:

- The determination of discrete customers.
- The existence significance of reasonable customers.
- Management Strategy of Four Categories of Customers.

During building customer classification model, customer information forms includes transaction amount, products and services traded, transaction frequency and age segments to form aggregate customer number set and single customer data set. The total number of customers set is X ={x1, x2…xn}, where N is the total number of customers. The single customer data set is xi=[e1,e2,e3,e4]T, which represents the transaction amount, products and service types, transaction frequency and age segment of the first customer. Based on amount of transactions (big or small), product types (more or less), frequency of transactions (high or low) and age of customers (low, medium or high) customers are group into four major categories and 25 sub categories inside those four categories using various unions of sets.

IV. DRAWBACKS OF EXISTING SYSTEM

- Static union rule is applied during customer grouping.
- Importance is not given among the grouped customer result set.
- Location data of customer is not taken so that village, town or city based customer grouping is not possible.

V. PROPOSED SYSTEM

- The proposed system presents a flexible and effective re-ranking method, called CR-Re- ranking, to improve the retrieval effectiveness.
- To produce higher accuracy on the highly top- ranked results, CR-Re-ranking uses a cross- reference (CR). Like the existing system, the records are classified but low, medium and high option is given for amount of transactions, product types, frequency of transactions, age of customers and location of customers.
- Specifically, multimodal features are first used separately to re-rank the initial returned results at the cluster level, and then all highly ranked clusters from different modalities are cooperatively used to infer the shots with high relevance. Experiment results show that the quality of search, mainly on the top-ranked results, is improved very much.

A. Advantages of the Proposed System

- Fuses the clustering results so that the results are more effective and relevant to end user's requirement.
- The new search mechanism satisfies the user's information needs.
- Accurate Re-ranking of result.
- Location data of customer is taken so that village, town or city based customer grouping is not possible.
- Suitable even if the retailer is situated in any type of location (village or town or city).

VI. MODULE DESCRIPTION

A. Data Collection

In this module, customer information including transaction amount, products and services traded, transaction frequency of transactions, age segments and location data to form aggregate customer number set and single customer data set. The total number of customers set is X ={x1, x2…xn}, where N is the total number of customers. The single customer data set is xi = [e1, e2, e3, e4] T, which represents the transaction amount, products and service types, transaction frequency and age segment of the first customer.All the data set columns are numeric but the last column 'Location' is categorical (Village, Town, City).

B. Dimensionless Treatment

In this module, the extreme value method is applied to deal with the single customer data for infinite tempering. The formulas are as follows:

$$e_{ij} = \frac{e_{ij} - min e_{ij}}{max e_{ij} - min e_{ij}}$$

C. Determining the type of Customer

In this module, customers are grouped as follows.

Firstly, the following three sets are defined. The first kind of set, namely the set of initial judgement of scope, is:

A1＝{i｜eij≤0.5,j＝1}
A2＝{i｜eij≥0.5,j＝1}
A3＝{i｜eij≤0.5,j＝2}
A4＝{i｜eij≥0.5,j＝2}
A5＝{i｜eij≤0.5,j＝3}
A6＝{i｜eij≥0.5,j＝3}

The second kind of set, namely discrete judgment set, is:

B1＝{i｜eij≤a,j＝1}
B2＝{i｜a≤eij≤b,j＝1}
B3＝{i｜eij≥b,j＝1}
B4＝{i｜eij≤a,j＝2}
B5＝{i｜a≤eij≤b,j＝2}
B6＝{i｜eij≥b,j＝2}
B7＝{i｜eij≤a,j＝3}
B8＝{i｜a≤eij≤b,j＝3}
B9＝{i｜eij≥b,j＝3}

The third kind of set, namely age sub section set is:

Y1＝{i｜eij≤c,j＝4}
Y2＝{i｜c≤eij≤d,j＝4}
Y3＝{i｜eij≥d,j＝4}

Secondly, 25 kinds of customers are synthetically described by three sets.

The first group of customers:

C1＝（B3∪ B6∪ B9）∩（A2∩A4∩A6∩Y1
C2＝（B3∪ B6∪ B7）∩（A2∩A4∩A5∩Y1
C3＝（B3∪ B4∪ B9）∩（A2∩A3∩A6∩Y1
C4＝（B3∪ B4∪ B7）∩（A2∩A3∩A5∩Y1
C5＝（B3∪ B6∪ B9）∩（A2∩A4∩A6)∩Y2
C6＝（B3∪ B6∪ B7）∩（A2∩A4∩A5)∩Y2
C7＝（B3∪ B4∪ B9）∩（A2∩A3∩A6)∩Y
C8＝（B3∪ B4∪ B7）∩（A2∩A3∩A5)∩Y2

The second group of customers:

C9＝（B3∪ B6∪ B9）∩（A2∩A4∩A6）∩Y3
C10＝ (B3∪ B6∪ B7）∩（A2∩A4∩A5）∩Y3
C11＝(B3∪ B4∪ B9）∩（A2∩A3∩A6）∩Y3
C12＝（B3∪ B4∪ B7 ∩ (A2∩A3∩A5）∩Y3

Thrid group of customers are:

C13＝（B1∪ B6∪ B9）∩（A1∩A4∩A6）∩Y1
C14＝（B1∪ B6∪ B7）∩（A1∩A4∩A5）∩Y1
C15＝（B1∪ B4∪ B9）∩（A1∩A3∩A6）∩Y1
C16＝（B1∪ B4∪ B7）∩（A1∩A3∩A5）∩Y1
C17＝（B1∪ B6∪ B9）∩（A1∩A4∩A6）∩Y2
C18＝（B1∪ B6∪ B7）∩（A1∩A4∩A5）∩Y2
C19＝（B1∪ B4∪ B9）∩（A1∩A3∩A6）∩Y2
C20＝（B1∪ B4∪ B7）∩（A1∩A3∩A5）∩Y

The Fourth Group of Customers are:

C21＝（B1∪ B6∪ B9）∩（A1∩A4∩A6）∩Y3
C22＝（B1∪ B6∪ B7）∩（A1∩A4∩A5）∩Y3
C23＝（B1∪ B4∪ B9）∩（A1∩A3∩A6）∩Y3
C24＝（B1∪ B4∪ B7）∩（A1∩A3∩A5）∩Y3
C25＝(B2∩B5∩B8)

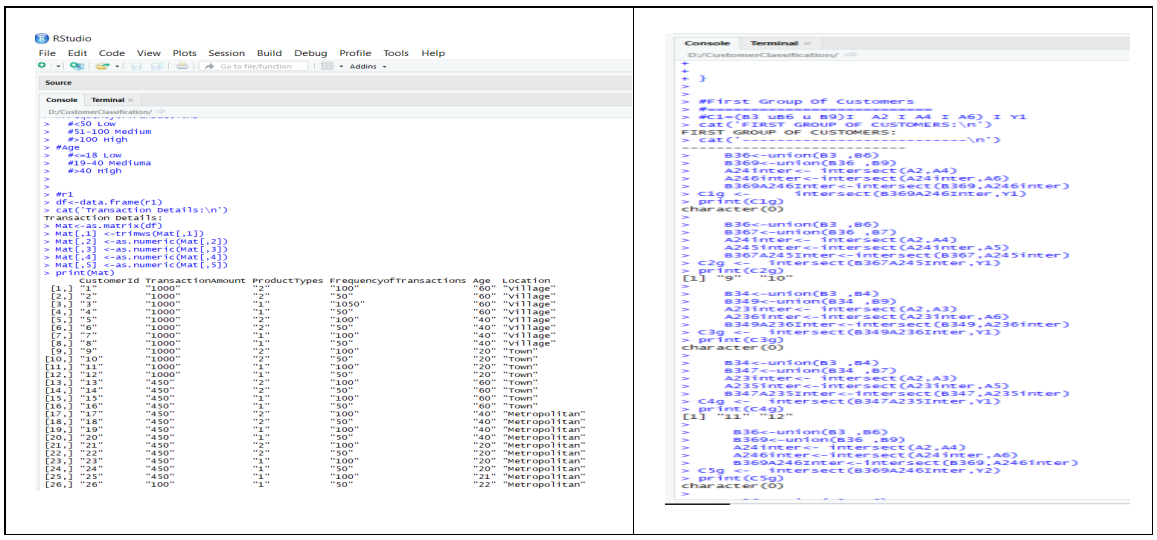

Fig.1 Data Collection, Customer details and First group of customer

Fig.2 second group of customer

Fig.3 Third group of customer

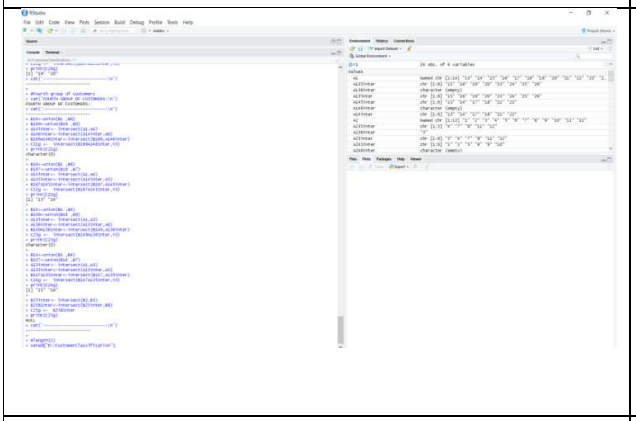

Fig.4 Fourth group of customer

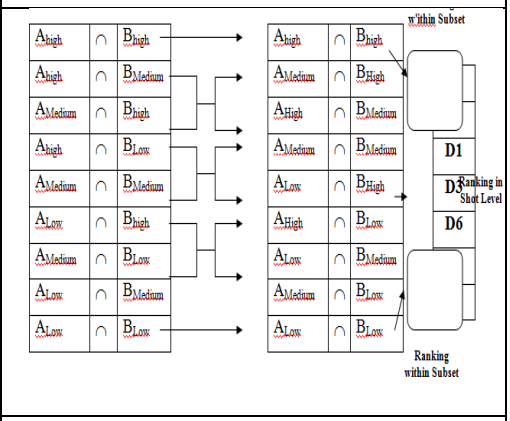

Fig.5 Reranking

D. Cross Reference

In this module,

- Transaction Amount with value ‘1 to 500’ are taken as low, ‘501 to 1000’ as medium and ‘>1000’ are taken as high.
- Product Types with value ‘Low’, ‘Medium’ and ‘High’.Frequency of Transactions with value ‘1 to 50’ are taken as low, ‘51 to 100’ as medium and ‘>100’ are taken as high.

- Age with value '<18' are taken as low, '18 to 40' as medium and '>40' are taken as high.
- Location with value 'Village' is taken as low, 'Town' as medium and 'City' as high.

All the clusters are grouped into three sub groups as High, Medium and Low for each cluster. For example, Cluster A is grouped as Ahigh, Amedium and ALow and Cluster B is grouped as Bhigh, Bmedium and BLow.

VII. CONCLUSION

Data mining provides a method of outlier analysis aiming at providing modules to help enterprises classify customers according to customer assets, so as to identify customers with good customer assets, and then develop and maintain customers in a targeted manner, which not only avoids the waste of resources caused by decentralization, but also reduces the huge risk brought by blind marketing of enterprises. presents a new re-ranking method that combines multimodal features via a cross-reference strategy. Given the top ranked clusters from all the feature spaces, the cross-reference strategy can hierarchically fuse them into a unique and improved result ranking. Experimental results show that the quality of search,especially on the highly top ranked results, is improved significantly. As analyzed previously, the proposed re-ranking method is sensitive to the number of clusters due to the limitation of cluster ranking. The difficulty in re-ranking of customers is eliminated by using this application. It reduces the re-ranking overheads especially when the number of documents is more. The user interface assists in accurate relevant customers' transactions searching. In future, this project may predict the missed values in the transactions.

REFERENCES

[1] Krishna Rungta, "Learn R Programming in 1 Day: Complete Guide for Beginners". Kindle Edition.
[2] Sandip Rakshit, "R for Beginners" Edition 1, 2017, McGraw-Hill Education, ISBN: 9789352604555, 9352604555
[3] Mike. McGrath, "R for Data Analysis in easy steps: R Programming Essentials", Kindle Edition. Sold by: Amazon Asia-Pacific Holdings Private Limited.
[4] Mark Gardner, "Beginning R: The Statistical Programming Language", Kindle Edition.
[5] http://www.rstudio.com
[6] http://www.rstudio.com/online-learning
[7] https://www.statmethods.net/r-tutorial/index.html
[8] https://www.tutorialspoint.com/r
[9] Peng, Yuhua & Yang, Xiaolan & Xu, Wenli. (2018). Optimization Research of Decision Support System Based on Data Mining Algorithm. Wireless Personal Communications. 102. 10.1007/s11277-018-5315-3. Article in Wireless Personal Communications 102(4) · January 2018 with 14 Reads DOI: 10.1007/s11277-018-5315-3.
[10] Li Ju, Xu Wenbin, and Zhou Bei, Member, IACSIT. "Construction of Customer Classification Model Based on Inconsistent Decision Table" International Journal of e-Education, e-Business, e-Management and e-Learning, Vol. 1, No. 3, August 2011
[11] Er. Jyoti, Er.Amandeep Singh Walia. "Recommendation system with Automated Web Usage data mining using K-Nearest Neighbor (KNN) classification", International Journal of Advanced Research in Computer Science, vol. 8, no. 4, May 2017 (Special Issue). ISSN No. 0976-5697
[12] Y. Cao, J. Xu, T.-Y. Liu, H. Li, Y. Huang, and H.-W. Hon, "Adapting Ranking SVM to Document Retrieval," Proc. 29th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval, 2006.
[13] A. Amir, J. Argillander, M. Campbell, A. Haubold, G. Iyengar, S. Ebadollahi, F. Kang, M.R. Naphade, A. Natsev, J.R. Smith, J. Te si c, and T. Volkmer, "IBM Research TRECVID-2005 Video Retrieval System," TREC Video Retrieval Evaluation Online Proc., 2005.
[14] J.H. Yuan, W.J. Zheng, L. Chen, D.Y. Ding, D. Wang, Z.J. Tong, H.Y.Wang, J.Wu, J.M. Lin, and B. Zhang, "Tsinghua University at TRECVID 2005," TREC Video Retrieval Evaluation Online Proc., 2005.
[15] C. Burges. Ranking as Learning Structured Outputs. Proceedings of NIPS Workshop, 2005.
[16] C. Silverstein, M. Henzinger, H. Marais, and M. Moricz. Analysis of a Very Large AltaVista Query Log. Technical Report SRC 1998-014, Digital Systems Research Center, 1998.
[17] A. Spink, B. J. Jansen, D. Wolfram, and T. Saracevic. e-commerce: web search changes. IEEE Computer, 35(3), pages 107-109, 2002.
[18] A. Spink, D. Wolfram, B. J. Jansen, and T. Saracevic. Searching the web: the public and their queries. Journal of the American Society of Information Science and Technology, 52(3), pages 226-234, 2001.