IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 7, June 2022
Impact Factor: 6.252

Clustering of Customer Transaction Data

Ms. S. L. Mortale[1], Mrs. M. A. Parlikar[2], Mrs. M. M. Mali[3], Ms. T. R. Shinde[4], Mrs. A. A. Sawlkar[5]
I/C HOD, Department of Information Technology[1]
Lecturer, Department of Information Technology Department[2,3,4,5]
Pimpri Chinchwad Polytechnic, Pune, Maharashtra, India

Abstract:*Clustering of customer transaction data and finding patterns using Apriori Algorithm is useful task in data mining to formulate market strategies and maximize profit. We apply an Apriori Algorithm for finding patterns. We use output of Customer Transaction Clustering Algorithm as an input to Apriori Algorithm. We have transaction tree which represents the customer's transaction records. Distance between transaction trees is calculated. A customer transaction clustering algorithm is used for clustering of transaction data of customers. The most frequent customers are selected as representatives of customer groups. Clustering is performed by assigning customer to the nearest neighborhood. Finally, the clustering results are forwarded to Apriori Algorithm for finding patterns.*

Keywords: Customer Transaction Clustering Algorithm, Apriori Algorithm, Transaction Tree, Clustering, etc.

I. INTRODUCTION

Clustering is the method of arranging objects into groups. Objects with similar characteristics are placed into one group. These groups are known as clusters. Customer Segmentation is a technique in which the customers are clustered on the basis of certain characteristics.The main objective of this paper is to find the best number of clusters and these resultant clusters are used for finding patterns using Apriori Algorithm

In clustering, we use large amount of raw and unorganized data as an input and determine similarities in input data.The clustering of transaction data of customers is essential phase to identify customer activities in retailand ecommerce firms [1].As there is rapid growth in customer behavior data, scientists are now focusing on clustering of transaction data of customers [2].

Basically, the transaction data is the information of the daily transactions of customers. It contains information about, what type of product or set of products purchased by customers. There are three common problems of clustering the data. One is the how to show customer and customertransaction data. Second is how to calculate the distance between different customers, and third is how to divide a customer into a certain number of customer groups.

We apply Apriori algorithm for finding patterns. We use output of Customer Transaction Clustering Algorithm as an input to Apriori Algorithm. We have transaction tree which represents the customer's transaction records. Distance between transaction trees is calculated. A customer transaction clustering algorithm is used for clustering of transaction data of customers. The most frequent customers are selected as representatives of customer groups. Clustering is performed by assigning customer to the nearest neighborhood. Finally, the clustering results are forwarded to Apriori Algorithm for finding patterns.Apriori algorithm is used to identify patterns. The rules are derived from the association.These rules must satisfy the minimum support threshold and the minimum confidence threshold.

We use transaction tree distance to compare customers at all levels of the Item(Product)Tree. However, transaction data of customer are very big, even after data is compressed by transaction trees. So, the speed of Clustering of customer transactiondata is very important. In real applications, it is very hard to use a hierarchical clustering method because of the high computational complexity. For complex product tree data, it is very difficult to apply the fast k-means algorithm. In this paper, we use clustering method called Customer

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 7, June 2022
Impact Factor: 6.252

Transaction Clustering for clustering of large amount of transaction tree data and uses Apriori algorithm which further processes the output of Customer Transaction Clustering algorithm and discovers thepatterns.
In Customer Transaction Clustering algorithm, we use a separate density for ranking a transactiontree as a representative tree. The most frequent customers are selected as representatives of customer groupsand clustering is performed by assigning each customer to the nearest neighborhood. Finally, clustering results are given as an input to Apriori Algorithm for finding patterns
The paper is organized in the following way.Literature Review is given in Section II. Proposed methodology is given in Section III. Result and discussion are given in section IV.Finally, we provide conclusion in section V.

II. LITERATURE REVIEW

Here we present the literature review of existing techniques:

In paper [1], transaction data of customers are compressed into set of purchase trees. Distance between two purchase trees is calculated. They implemented PurTreeClust clustering algorithm for performing clustering.

In this paper [2], they considered transaction data with high capacity and dimension. The author used heuristic approach to increase the ratio of width and height of cluster histogram. They developed fast and scalable algorithm named as CLOPE.

In order to predict the yearly sales of supermarket, SPPS tool and K-means clustering is used to create online and real time system for supermarket [3].

SWCC Subspace Weighted Co Clustering algorithm is developed by the author for high-dimensional expressiondata. To identify different clusters, Subspace weight matrices were presented [4]

In this paper [5], the author presents an automatic two-level variable weighting clustering algorithm for multi view data TW-k-means, which can compute the weight of view and individual variables simultaneously. The algorithm assigns view weights to each view in order to identify the compactness of the view, assigns variable weights to each variable in the view, then using the quantifiable two real data sets, examines the nature of the two types of weight in TW-K-mean, TW-K-mean it is possible to determine the weight of the view. The difference between the weight of the man and the weight of the individual variable weighting method was examined.

The author presents a robust tree edit distance algorithm – RTED [6]. The author introduces the class of LRH (Left-Right-Heavy) algorithms, which includes RTED and the fastest tree edit distance algorithms.

In this paper, data mining techniques are used to provide customer's purchasing patterns of food items. Author uses KMedoids clustering algorithm for clustering of food items. These outputs of clustering are given as an input to the association rule mining based Apriori algorithm and frequent patterns are discovered [7].

The aim of this paper [8] is to recommend the suitable items to the user. A better Rule extraction is needed to recommend the suitable items. Association Rule mining is applied for better rule extraction.The K-means clustering algorithm method is also applied here to cluster the data based on similar characteristics.

In this paper [9], frequent user access patterns are generated from web log entries. Combined efforts of clustering and association rule mining is used to apply pattern discovery.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 7, June 2022
Impact Factor: 6.252

III. PROPOSED APPROACH

A. Problem Statement

To develop a system that performs clustering of customer transaction data and processes resulting clusters to Apriori Algorithm for finding patterns.

B. Proposed System Overview

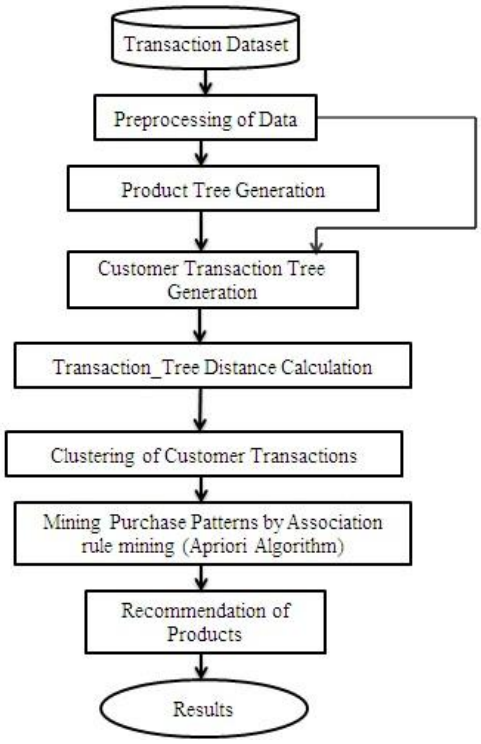

Figure 1: Proposed System Architecture

A detailed description of the proposed system is as follows:

1) Transaction Dataset:

The system uses customer transaction data. The transaction data include the products bought by customers.

2) Pre-processing of Data:

Pre-processing is done on transaction data of customers.

3) Product (Item) Tree Generation:

A Product or Item tree consists number nodes. A child node represents product or item. An internal node represents category of particular item.

4) Transaction Tree Generation:

A Transaction Tree consists of number of nodes. The child nodes represent items bought by customer and internal node represents the category of particular item.

5) Transaction Tree Distance:

Customers do not buy similar products, due to this between any two transaction trees the tree edit distance will produce high distance value. Within the tree edit distance it is very difficult to recover the cluster structure. To solve this issue,Transaction Tree distance metric is used. Transaction Tree distance compares customers from the entire levels of the product tree.

6) Transaction Tree Clustering:

Transaction Tree Clustering Algorithm is used for clustering of transaction tree data.

7) Mining Purchase patterns by association rule mining:

We use apriori algorithm for finding patterns.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 7, June 2022
Impact Factor: 6.252

8) Recommendation of products:
Finally, system recommends the product and gives fast and accurate results.

C. Algorithm

Algorithm1: Cust_Tran_Clustering

Clustering of customer transaction data consists of following steps

1. Generate the Item(product) tree.
2. Generate customer transaction tree for each customer.
3. Calculate the distance between two transaction trees.
4. Estimate the level density of transaction Tree with cover tree: denlCT(p)
5. Calculate the separate distance of object p ∈CSl: sdislCT(p)
6. Calculate separate density of object p ∈CSl:sdenlCT(p)
   sdenlCT(p) = denlCT(p) * sdislCT(p)
7. Select 'k' representative trees as 'k' treeshaving highest separate densities.
8. Perform clustering by assigning eachcustomer to the nearest representative.

Algorithm 2: Apriori Algorithm

Following are the steps of Apriori Algorithm:

1. Initialize s=1
2. Generate frequent itemset of size '1'.
3. Generate candidate itemset of size 's+1'from frequent itemset of size 's'.
4. Prune candidate itemsets containing subsetsof size 's' that are infrequent.
5. Count the support of each candidate itemset.
6. Remove candidate that are infrequentkeeping only those that are frequent.
7. Repeat steps 3 to 6 until no new frequentitemsets are identified.

IV. RESULT AND DISCUSSION

A. Expected Result

We used real life supermarket transactional dataset, initially we assumed 500 records. It shows that the Apriori algorithm with Clustering requires less time than Apriori algorithm without Clustering.

Table 1

Comparison of Apriori with Clustering and without Clustering		
Algorithm	***Time Required (Seconds)***	***Accuracy (%)***
Apriori algorithm with Clustering	10 Seconds	90%
Apriori algorithm without Clustering	15 Seconds	80%

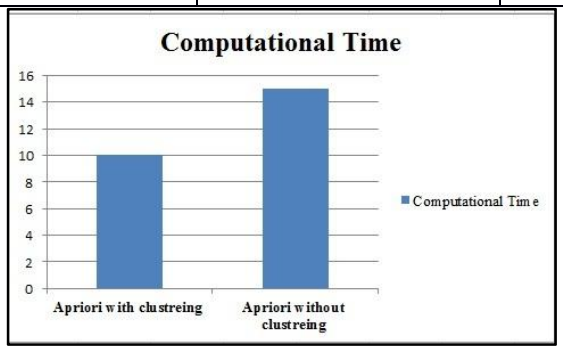

Figure 2: Comparison of Apriori with clustering and without clustering (Computational Time)

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 7, June 2022
Impact Factor: 6.252

V. CONCLUSION AND FUTURE SCOPE

We use Apriori Algorithm for finding patterns. We use output of Customer Transaction Clustering Algorithm as an input to Apriori Algorithm. Apriori Algorithm is used for finding patterns. This is beneficial to increase product sale by identifying relations between combinations of customer's purchase pattern.

A customer transaction clustering algorithm is used for clustering of transaction data of customers. The most frequent customers are selected as representatives of customer groups. Clustering is performed by assigning customer to the nearest neighborhood. Finally, the clustering results are forwarded to Apriori Algorithm for finding patterns.

From the accuracy graph, we can conclude that Apriori algorithm with Customer Transaction Clustering is more accurate and efficient than Apriori Algorithm without Clustering. Resultant patterns are useful to formulate market strategies and maximize profit.

ACKNOWLEDGMENT

The authors would like to thank the researchers as well as publishers for making their resources available and teachers for their guidance. We are thankful to the authorities of Savitribai Phule University, Pune and concern members of cPGCON-2019conferencefor their constant guidelines and support. We are also thankful to the reviewer for their valuable suggestions.

REFERENCES

[1] X. Chen, Y. Fang, M. Yang, F. Nie, Z. Zhao and J. Z. Huang, "PurTreeClust: A Clustering Algorithm for Customer Segmentation from Massive Customer Transaction Data," IEEE vol. 30, no. 3, March 2018.

[2] Y. Yang,X. Guan,J. You, "CLOPE: A Fast and Effective Clustering Algorithm for Transaction Data" in Proceedings of 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,ACM 2002.

[3] Kishana R. Kashwan and C.M.Velu, "Customer Segmentation Using Clustering and Data Mining Techniques",IJCTE Vol. 5.No.6 December 2013.

[4] Q. Wu, X. Chen,J. Z. Huang and M. Yang, "Subspace Weighting Co_clustering of Gene Expression Data", TCBB, 2017.

[5] X. Chen, X. Xu,Y. Ye and J.Z. Huang, "TW-k-means: Automated two-level variable weighting clustering algorithm for multi-view data", IEEE, Vol.25 No.4 Apr. 2013.

[6] M. Pawlik and N. Augsten, "RTED: A robust algorithm for the tree edit distance",Proc.VLDB Endowment,Vol.5,No.4,2011.

[7] Kavita M. Gawande.Mr. Subhash K. Shinde, Mrs. Dipti Patil, "Frequent Pattern Mining Based on Clustering and Association Rule Algorithm", IJARCS,Vol.3,No.3,2012.

[8] Jaimeel. M. Shah, Lokesh Sahu, Recommendation based on Clustering and Association Rules, IJARIIE-ISSN(O)-2395-4396, Vol-1 Issue-2 2015.