IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 12, December 2015

Effective Evaluation of Transactional Data Items - A KDD Approach toward the customer rating

Mr. Pragnyaban Mishra[1], Mr. Sudhanshu Shekhar Bisoyi[2]

Department of CSE, Gandhi Institute of Engineering & Technology, Gunupur, Rayagada, India[1]

Ph.D Scholar, Department of Computer Science & Engineering , IGIT Sarang, Utkal University, BBSR, India[2]

Abstract: In the Market Basket Analysis Frequent pattern mining plays the vital role in finding occurrence frequency of the items in a transactional dataset. By this frequent pattern mining analysis every retail company are aimed at maximizing the social relationship with the customer. If we use only frequent pattern mining then it generate a large amount of itemset which is very difficult to perform the in-memory analysis. So we have proposed a new approach of finding the item frequency by associating a weight or utility factor based upon the usage, demand etc. of the item. Finally the clustering techniques have been used to classify the customer from the final transactional dataset which will help to strengthen the social relationship between the retail organization and their customer.

Keywords: CRM, Min-Sup, Cluster of the customers, data mining.

1. INTRODUCTION

Now a day's customer rating became an emerging issue to promote more products, establishing social relation with the customer, adding more numbers of customers etc. On the other hand the organization must establish proper relationship with the customer not only maximizing the benefits, but also the level of satisfaction of the customer can include 'N' numbers of customers with the organization. In general the traditional association rule mining is based on the support and confidence as a measure for the rule generation. On the other way the CRM model explains a lot about different models and prototype of the customer relationship. Various organizations need to make a plan to reach or attract their customer. In order to achieve it the organizations should not concentrate on the entire customer rather they have to be specific to few customer by analyzing the buying habits.

It has been realized that, we are rich by data and poor by information. Our approach explains how to find out a good customer in terms of rating so that, the organization can connect to them through various mean and familiar with their need.

In our research work we have used Association Rule Mining and Classification to find out a suitable rating to each of the products or items from the previous transactions which will make in analyzing the rules generated from the large transactional data set.

2. RELATED WORK

The impact of data mining in CRM [1,4,5] are take important role to find out dynamic information for future references. The mining approach can be customer centric to find out appropriate information from the raw data various data mining applications [2,6,7] named as Association rule mining, Clustering etc. can be adopted. Useful results have been found in various fields [3,8,10,12,13,15] of real world to serve the mankind.
Some of the review work have been performed in these related areas and became a source for new development [9,11,14]

3. OUR FINDINGS

The association rule mining from a large transactional data set not only generates a large number of valid association rules but also leads to a difficulty during the analysis. If the mining derives the rule not only from the frequency but also with the utility of the items then probably that will represent in more meaningful way.

In our study we have derived useful information by associating item frequency and weighted utility factor with the transactional items.

3.1 Proposed algorithm

Step 1: Collect all the transactional data set .
Step 2: Find the frequency of all the items used in the transactions.
Step 3: Discard the items from the transactions below Minimum Support and the table is called as : $T_{r,fil}$
Step 4: Build the utility list of the itrems in $T_{r,fil}$
Step 5: Associate the utility factor with each item in $T_{r,fil}$ *which are the weights of the items.*
Step 6: Do the classification by considering the utility factor C_{utl}^{T} *and prepare the report as per findings*

3.2 Study of the Algorithm

In our model we assumed a small data set containing ten transactions among the five items. In the table 1's are the items reflected in the transaction and 0's are not.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 12, December 2015

Tr id	I1	I2	I3	I4	I5
Tr1	1	0	1	1	1
Tr2	1	1	1	1	0
Tr3	1	1	0	0	0
Tr4	1	1	1	0	0
Tr5	1	1	0	0	1
Tr6	0	0	0	1	0
Tr7	1	0	1	1	0
Tr8	1	0	0	1	1
Tr9	1	1	0	0	1
Tr10	1	1	0	1	0

Table 1: Transactions contains various items purchased.

The item frequencies in various transactions have been counted and represented in the table 2.

Items	Frequency
I1	9
I2	6
I3	4
I4	6
I5	4

Table 2: Frequencies of single item set

We applied the minimum support measure of the association rule mining for single item set and rejected those items whose frequency are less than *Min-Sup*. In this case the predefined *Min-Sup* is 45% .

The items I3, I5 are with frequencies 4 which are less then *Min-Sup* so they are discarded from the computations. After removing the frequencies of I3,I5 from the table 2 is given in table 3.

Items	Frequency
I1	9
I2	6
I4	6

Table 3: Frequencies of single item set after applying *Min-Sup*

As because the frequency of I3 and I5 are less then *Min-Sup* so their occurrence in the transaction doesn't have any significant. It is better to remove I3 and I5 from the transaction database for the further analysis.

The new transaction data base after removing the I3 and I5 became as follows:

Tr id	I1	I2	I4
Tr1	1	0	1
Tr2	1	1	1
Tr3	1	1	0
Tr4	1	1	0
Tr5	1	1	0
Tr6	0	0	1
Tr7	1	0	1
Tr8	1	0	1
Tr9	1	1	0
Tr10	1	1	1

Table 4: The new Transaction history after removing I3, I5.

The transaction History now represented in a form of 10 *X* 3 matrix after removing the occurrence of the item whose support is less then *Min-Sup* i.e 45% given as follows:

$$\begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

Matrix A: Representing Transactions after *Min-Sup*.

The weights of the items are predefined and can be in a form w_{11}, w_{12}, w_{13} etc. which can be a numeric value. The weights of the items are defined taking the price, salving, seasonal demand, promotional items etc.

Items (I)	Weight (w)
I1	3
I2	3
I4	2

Table 5: The Predefined weights of the Items I1,I2,I3.

This is presented in the matrix B as:
Matrix C: Multiplying both the matrices A and B and represented in the **transpose** form as follows

(5 8 6 6 6 2 5 5 6 8)

The derived weight are used for the respective transactions and represented in table 6.

Tr id	Wts
Tr1	5
Tr2	8
Tr3	6
Tr4	6
Tr5	6
Tr6	2
Tr7	5
Tr8	5
Tr9	6
Tr10	8

Table 6: Weight of the transactions as follows

The table 6 now becomes a base table to retrieve various information and report can be prepared as per the requirement of the organization. Here it has classified all 10 customers into 3 classes based upon their cumulative weight which represent the utility factor of the item in the transaction.

DOI 10.17148/IJARCCE.2015.41250

Excellent (>=8)	Good (>=5 & <= 7)	Avg (<5)
Tr2	Tr1	Tr6
Tr10	Tr3	
	Tr4	
	Tr5	
	Tr7	
	Tr8	
	Tr9	

Table 7: Classification of Customer as per their cumulative weight.

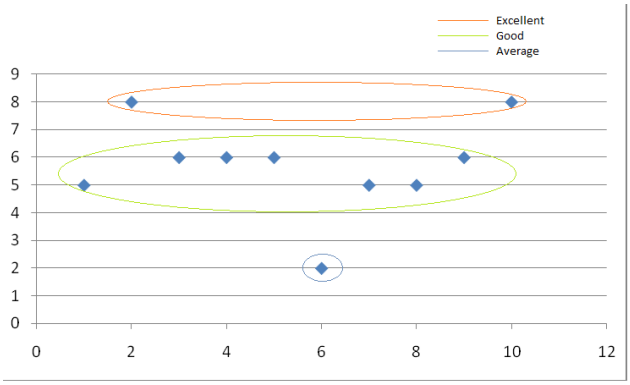

Figure 1: Cluster of the customers(X-axis represents TID, Y-axis represents Cum Wt)

4. CONCLUSION

In our research work we have represented the potential of data mining to find out the proper categorization of customer based upon their buying habits with the utility factor of the items from the transactional dataset. The research finding of this article minimizes the gap between the customer and proprietor to create healthy association among the buyers communities. The future scope of this work is implementation of ANN in a trained data set to find out dynamic weight of items instead of static.

REFERENCES

[1] "A reference model for customer-centric data mining with support vector machines" Stefan Lessmann , Stefan European Journal of Operational Research 199 (2009) PP: 520–530

[2] "An efficient approach for mining association rules from high utility item sets." Jayakrushna Sahoo , Ashok Kumar Das, A. Goswami, Expert Systems with Applications 42 (2015) PP: 5754–5778.

[3] "An integrated data mining and behavioral scoring model for analyzing bank customers" Nan-Chen Hsieh , ROC Expert Systems with Applications 27 (2004) PP: 623–633

[4] "Application of data mining techniques in customer relationship management: A literature review and classification" E.W.T. Ngai, Li Xiu b, D.C.K. Chau.Expert Systems with Applications (36) 2009 PP: 2592–2602

[5] "Application of data mining techniques in customer relationship management: A literature review and classification" E.W.T. Ngai, Li Xiu, D.C.K. Chau Expert Systems with Applications 36 (2009) 2592–2602

[6] "Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method" D.A. Adeniyi, Z. Wei, Y. Yongquan Applied Computing and Informatics 2015.

[7] " Customer Data Clustering Using Data Mining Technique", Sankar Rajagopal International Journal of Database Management Systems (IJDMS) Vol.3, No.4, November 2011

[8] "Customer relationship management in the hairdressing industry: An application of data mining techniques." Jo-Ting Wei a, Ming-Chun Lee b, Hsuan-Kai Chen c, Hsin-Hung Wud, Expert Systems with Applications 40 (2013) PP: 7513–7518.

[9] **"**Feature-based Customer Review Mining" Jingye Wang Heng Ren Stanford University

[10] "Integration of data mining technologies to analyze customer value for the automotive maintenance industry" Yi-Hui Liang Expert Systems with Applications 37 (2010) 7489–7496

[11] "Mining and Summarizing Customer Reviews" Minqing Hu and Bing Liu Department of Computer Science University of Illinois at Chicago

[12] "Mining changing customer segments in dynamic markets" Mirko Bottcher, Martin Spott , Detlef Nauck Rudolf Kruse, Expert Systems with Applications 36 (2009) PP:155–164

[13] "Mining customer knowledge for tourism new product development and customer relationship management" Shu-hsien Liao, Yin-Ju Chen , Min-yi Deng, Expert Systems with Applications 37 (2010) PP: 4212–4223

[14] "Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics" Mohammad Salehan a, Dan J. Kimb, Decision Support Systems 2015.

[15] "Service Quality and Customer Satisfaction in the Restaurant Business Case Study- Sagarmatha Nepalese Restaurant in Vantaa" Abadh Jibi Ghimire, Thesis Central Ostrobothnia Universty of Applied Science Degree Programme in Tourism May 2012

DOI 10.17148/IJARCCE.2015.41250