International Journal of Computer & Information Technologies (IJOCIT)
www.ijocit.ir & www.ijocit.org
ISSN = 2345-3877

Products Recommendation System in E-Commerce using Generalized Sequential Pattern

[1]Mohsen Dindar, [2]Shahram Jamali
[1]*Computer engineering Department, Germi Branch, Islamic Azad University, Germi, Iran*
[2]*Computer networking lab, Department of Computer Engineering, University of Mohaghegh Ardabili, Ardabil, Iran.*
Vbcars@gmail.com

Abstract:- Development of communication networks cause information overhead that the electronic stores have faced this problem. Therefore Product Recommendation System (PRS) tries to solve this problem by giving recommendations to customers. In this paper, a product recommender system is introduced that uses clustering products through fuzzy clustering C-means algorithm, to advise clusters of similar products to customer requirements. Then it uses sequential patterns analysis to suggest potentially associated products with customer's chosen product according to choice time of transaction. As the number of transactions in the electronic store is very high, the patterns of the relationship between the products can be redundant that impose additional computational complexity on the system. In this study to address the problem, the generalized patterns method has been used. The implementation results show pruning the redundant patterns makes the complexity of system that it was in previous methods of order O(n2) significantly reduced and the degree of complexity of the proposed system be order of O(n log n).

Keywords: Product recommendation system, Sequential pattern analysis, Clustering, Most generalization, Frequent closed itemset lattice.

1. Introduction

Product Recommendation System (PRS) as a web-service takes a set of the user's criteria as input and searches among the items in the database for products in accordance with the user's criteria and finally suggests the list of products to the user that match with user’s criteria as the output. Users by this recommendation can find require products without wasting time and confusion and make more accurate decisions about purchase product. In addition, PRS can monitor the history of customer purchase behavior, preferences and predict customer needs and products that are closely related with products which proposed to customers. Also users can recognize products based on the recommendations that associated with their priorities, and decide to buy. Hence, the recommendation system can help to users to identify suitable products for their needs and preferences in an effective way and solve the problem of information overload in e-commerce and help to growth sales [4].

The customers without having help to online access to information about their desired product, they will be confused. In this time users are willing to get advice from an expert or experienced person.

In this case the use of a product recommender system which recommends relative products to users according to user preferences is necessary. In the proposed product recommender system, in the first stage we collect relevant data to available products in the online store. Then, products are classified in terms of the type of separation, the obtained data are clustered by use of the C-mean clustering approach. In the next stage, information related to history of customer purchase and behavior is collected through standard dataset. With obtaining this data, we use sequential pattern [8] mining and explore connections between the products. These associations will ultimately lead to make recommendations in the time of buying products to customers. The connection between products will increase the likelihood of purchase of that product together [9].

Since e-stores have gained great popularity among customers, rising of trading volume and transactions, so patterns that are achieved across the previous stage can be redundant and duplicated. Therefore, waste rules should be removed to deacrease useless patterns and the computational complexity. For this purpose, in this study more generalization [10] approach is presented in order to reduce the number of candidate patterns and minimize the volume of calculations.

2. Related Work

Undoubtedly customers have demanded more personalized information delivery services. Therefore many PRSs are introduced in literatures that try to provide more personalized services.

A social recommendations system in electronic commerce combines similarity, trust and relationship and detects priority of the members through close friends and social network [3].The main idea is to obtain a rate for users. According to this rating, other users can decide whether they are traded by a trusted user or not. In fact this mark establishes the reliance on the recommendations of any person that is obtained by close friends in social network [3].

Other recommendation system in order to personalize the applications in B2C e-commerce uses association classification for the product recommendation issue [2]. In this system customers' requirement data are collected and changed into proper phrase datasets. Then association rule mining method starts to search for a set of related phrases that frequently occurring phrase patterns (classifiers). In order that, a list common stop word in English are adopted and removed by stemming algorithm to reduce the volume of documents dataset and improve the accuracy of the rukes mining [4].

Other one based on clustering comments of user and reviews is proposed in [5]. In this work, related comments about products presented by reviewers and users are clustered. Product reviewers also considered as the part of users and affect their preferences in the clustering. This paper recommends closed products with cluster apriority to users which into clusters. The customers that shopping for the first time can use reviews and other customer comments cluster as guidance in choosing the product. Products that have already been offer to this cluster, present to new customers [5].

A PRS with the dynamic templates is introduced in [6]. Since users have different needs at different times and customer's purchase due to variety of requirements at different times are vary considerably, hence the constant PRS cannot Responder the needs of customers. In this paper, the behavior of the users during the lifecycle is considered and the needs of customers in each period manually classified. This classification facilitates recommendation process and helps to personalization product recommendation [6].

Other PRS is proposed in [7] that it's a solution on personalized products recommendation based on user-contributed photos from social media sites. The input of this approach is user shared photos of the same webpage and their corresponding textual descriptions. Then the PRS recommends related products with these descriptions to user.

Other PRS is developed [8] that called HOPE, which combines collabrative fillter based recommendation using implicit rating and sequential pattern analysis based recommendation that calculates explicit rating for each user. Then it finds k neighbors that have similar rating for each target user. Then it discovers sequential pattern with monitoring on neighbors' purchase behavior and assigns normal weight to products. Finally it advises n products with high normal weight.

3. Proposed PRS

Proposed PRS in this work includes tow level of product recommendation that first level recommended before product purchase and other one after buying. In fact, this PRS initially recommends products that have closely related with customer's criteria and requirements to avoid wasting time. In second level it recommends associated products to purchased product by customers to complete buying process and to aware them from potentially related products with their desired products.

In this section describe the overall framework of proposed PRS. First, we collect products' data from electronic store.

These data includes name, type, price, quality rating and e.g. Then we separate the products according to these type (for example products such as furniture, books, software, and cellphone have different labels).Then these products according to their numerical characteristics that can be price and quality rating are clustering in three separate clusters of high, medium and low quality by C-means algorithm. The products into clusters have very close distance in aspect of similarity criteria,

pp. 95- 110

however, the products are in separate clusters are so far.

3.1. C-means Algorithm

Clustering creats groups of objects according to their features that the similar objects belonging to the same groups and dissimilar ones belonging to different groups. Clustering is one of the most popular techniques to categorize objects in the field of data mining. Its purpose is to systematize classifying a dataset into a set of clusters, which include close data objects as measured by some distance function. The major applications of clustering include document categorization, scientific data analysis, customer/market segmentation and etc[14].

Clustering approaches based on fuzzy logic, such as FCM and its variants have proved to be competitive to conventional clustering algorithms, especially for real-world applications. These approaches assign each feature vector to different clusters by a certain degree, so dislimn sharp boundaries between the clusters. The degree of membership of a feature vector to a cluster is usually determined as a function of its distance from the center cluster or from other representative vectors of the cluster's center. The fuzzy features of the k-Means algorithm are sometimes referred as Fuzzy C-Means algorithm.Traditional clustering approaches generate partitions; in a partition, each point belongs to one and only one cluster. Fuzzy clustering extends this notion to relate each point with every cluster using a membership function. Fuzzy clustering is an extensively applied method for obtaining fuzzy models from data.It has been applied successfully in various fields including geographical surveying, finance or marketing. The most applicable clustering algorithm implementing is the FCM. Despite this algorithm proved to be less accurate than others, it's fuzzy nature and the ease of implementation made it popular and very attractive for a lot of researchers. Usually FCM is applied to unsupervised clustering problems. The basic structure of the FCM algorithm is discussed below. The Algorithm FCM is a method of clustering which allows one piece of data to belong to more than one cluster. It is based on minimization of the following objective function [8]:

$$J_m = \sum_{i-1}^{N} \sum_{j-1}^{c} u_{ij}^{m} \parallel x_i - c_j \parallel^2, 1 \le m \le \infty \quad (1)$$

Where m is any real number greater than 1, uij is the degree of membership of xi in the cluster j, xi is the ith of d-dimensional measured data, cj is the d-dimension center of the cluster, and ||*|| is any norm expressing the similarity between any measured data and the center. Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above, with the update of membership uij and the cluster centers cj by [14]:

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} [\parallel x_i - c_j \parallel / \parallel x_i - c_k \parallel]^{\frac{2}{m=1,\dots c_j}}} = \frac{\sum_{i=1}^{N} u_{ij}^{m} x_i}{\sum_{i=1}^{N} u_{ij}^{m}}$$

(2) This iteration will stop when [8]

$$max_{ij}\{\left|u_{ij}^{(k=1)} - u_{ij}^{(k)}\right|\} < \xi \quad (3)$$

Where ξ is a termination criterion between 0 and 1, whereas k is the iteration steps. This procedure converges to a local minimum or a saddle point of Jm. The algorithm is composed of the following steps [14]:

Step 1: Initialize U= [uij] matrix, U(0).

Step 2: At k-step: calculate the centers vectors C(k) = [cj] with U(k).

Step 3: Update U(k), U(k+1).

Step 4: If $\|U(k+1) - U(k)\| < \xi$ then STOP; otherwise return to step 2.

In this algorithm, data are bound to each cluster by means of a membership function, which represents the fuzzy behavior of the algorithm. To do that, the algorithm builds a membership matrix named U whose factors are numbers between 0 and 1, and represent the degree of membership between data and centers of clusters. FCM is similar in structure to the k-Means algorithm and also behaves in a similar way. It updates the clusters' central points with data point assignment and redirect these to appropriate places in cluster. This clustering techniques are based on fuzzy behavior and provide a natural technique for producing a clustering where membership weights have a natural (but not probabilistic) interpretation [14].

3.2. Sequential Pattern Analysis

Sequential pattern mining discovers subsequences that appear in a sequence database with frequency more than a user-specified threshold. A sequence database contain a number of records, where all records are ordered sequences of events, with or without concrete notions of time. Examples of sequences include retail customer transactions, DNA sequences, and web log data. A subsequence, such as buying first a PC, then a digital camera, and then a memory card, if it occurs frequently in a customer transaction database, is a (frequent) sequential pattern [11].

This paragraph presents the formal definition of the sequential pattern mining problem, and its associated notations. Let $I = \{i_1, i_2, \dots, i_n\}$ be a set of all *items*. An *itemset* (or *event*) is a subset of items and denoted by $(i_1 i_2 \cdots i_m)$, where i_k is an item. It is assumed that items in an itemset are sorted in lexicographic order. A *sequence* is an ordered list of itemsets. A sequence s is denoted by $\langle s_1 s_2 \cdots s_l \rangle$ where s_j is an itemset. An item can't occur more than once in an itemset of a sequence, but can occur multiple times in different itemsets of a sequence. The *length* the of the sequence is the number of instances of items in a sequence. A sequence with length l is called an l-sequence. A sequence $\alpha = \langle a1a2 \cdots an \rangle$ is called a subsequence of another sequence $\beta = \langle b_1 b_2 \cdots b_m \rangle$ and β is a super-sequence of α, denoted by $\alpha \sqsubseteq \beta$, if there exist integers $1 \le j1 < j2 < \cdots < jn \le m$ such that $a1 \subseteq bj1, a2 \subseteq bj2, \dots, an \subseteq bjn$.

A *sequence database D* is a set of tuples *sid*, *s*, where *sid* is a sequence_id and *s* is a sequence. A tuple *sid*, *s* is said to contain a sequence α, if α is a subsequence of s. The *support*(α) is the number of occurness of a sequence α in a sequence database D. A sequence α is a *sequential pattern* if *support*(α) be more than a positive integer *min_support* as the *minimum threshold of appearence* in database D. The set of frequent l-sequences is denoted by Fl. If there exists no proper super-sequence of a sequential pattern α with the same support as

α, α is called a *closed sequential pattern* (or a *frequent closed subsequence*) in sequence database D. Furthermore, a sequential pattern α is called a *maximal sequential pattern* (or a *frequent maximal subsequence*) if it is not contained in any other sequential pattern [15].

4. Implementation of Proposed Method

In this section, implementation and performance evaluation of proposed PRS are described as following

4.1. Implementation Issues

As mentioned above, the proposed PRS has two levels. The first level clusters available products in electronic store and the second level draws the existing relations between products. This section describes how the clustering is performed by using C-Means algorithm and also how the relationships are extracted by using SPA method. Our implementation employs UCA transactions 10k dataset to examine the performance of the proposed PRS.

In following, we first bring the clustering details used in this research and discuss how the SPA is used for extracting potentially relation between products.

4.2. Clustering Implementation Details

The proposed PRS initially recommends products that closely are related to customer's criteria and requirements to avoid wasting time. For implementation this level, we cluster a type of the collected data from electronic store using C-means algorithm based on numerical attributes of products such as price and quality ratings. Considering that the quality rating is announced by company that creates the product, so reliable is. It's important to note that all products in ecommerce store must have price and quality rating but we can assume some products with difference attribute. In this case, we can use Principle Component Analysis (PCA) [13] algorithm to reduce dimensionality of data and provide two relevant features in order to clustering these products. This algorithm finds a projection that captures the largest amount of variation in data and eigenvectors of the covariance matrix of data that define the new attributes.

After clustering products that are in a cluster have similar properties while products are in different clusters have different properties. It's a result of clustering. The collected cell phone's dataset from Tebian electronic store includes six fields, namely product ID, product name, Price, quality range, price range, Type.

In first step of C-means algorithm, scattering data of cell phone is showed in Figure 1. After this step, C-means algorithm is applied on data. In this paper three clusters are considered, namely Low quality, Middle quality and High quality. In next step, three centroid points are selected for three clusters randomly that is showed in Figure 2. C-means algorithm changes place of centroid points according to mean of data. Algorithm repeats this replacement until there is no movement in centroid points. At this time each centroid has placed in correct point at mean of cluster. Figure 4 illustrates final place of centroid points in each clusters that c1, c2, c3 are means of Low quality, Middle quality and High quality clusters, respectively.

pp. 95- 110

International Journal of Computer & Information Technologies (IJOCIT)
Corresponding Author: *Mohsen Dindar*
November, 2016

Then each point is assigned to cluster based on uij membership degree that mentioned in (2). Each point with high membership is assigned to cluster with the nearest centroid. Finally clustering for data related to cell phone is shown in Figure 4. Also Table1 1 shows the distance of points to the centroid of the clusters according to membership function. The shortest distance causes to allocate each point to a cluster.

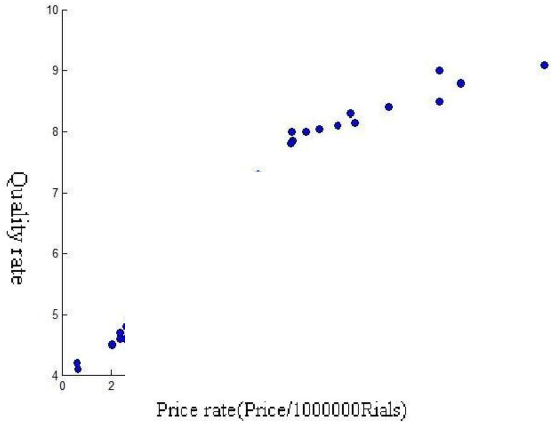

Figure 1: **Scattering Data of Cell phone**

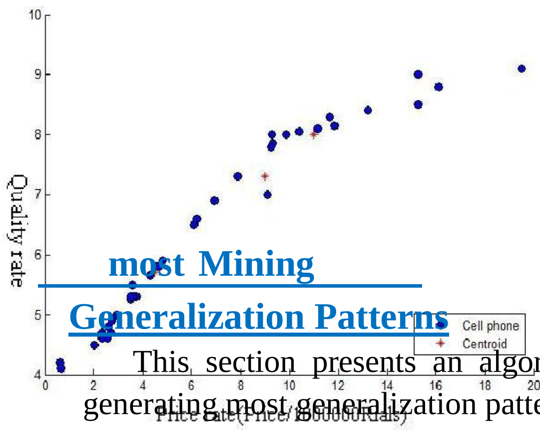

Figure 2: Selection Centroid Points Randomly

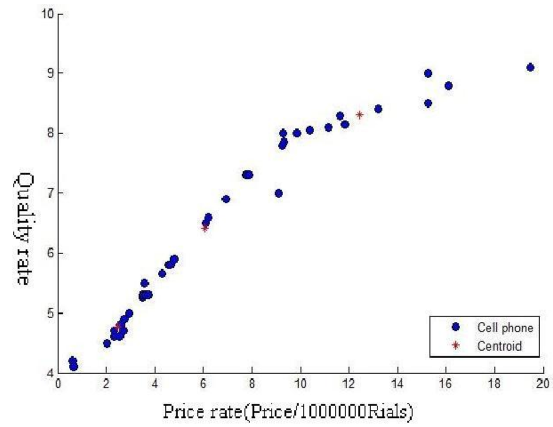

Figure 3: **Correct Place of Each Centroid**

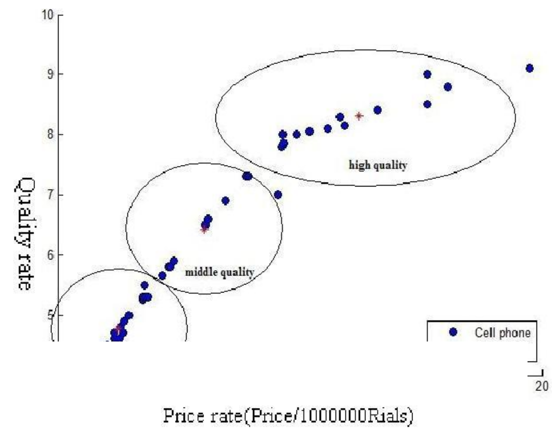

Figure 4: **Clustering for Data Related to Cellphone**

Table 1: Coordinates of Each Centroid Points

Cell phone	Cluster centroid
(2.4767,4.7654)	C_1
(6.0754,6.4115)	C_2
(12.4669,8.3115)	C_3

4.3. Mining most Generalization Patterns

This section presents an algorithm for generating most generalization patterns from a Frequent Itemset Closed Lattice (FCIL). The following theorem is first derived.

Theorem 1. Given three nodes l_1, l_2, and l_3 in FCIL, if l_1 is the parent node of l_2, l_2 is the parent node of l_3, and $\frac{L_2.sup}{L_1.sup} \langle \min conf$, $\frac{L_3.sup}{L_1.sup} \langle \min conf$ [10]. then

International Journal of Computer & Information Technologies (IJOCIT)
Corresponding Author: *Mohsen Dindar*
November, 2016

Proof. Since l_1 is the parent node of l_2 l_3, of node parent the l_2 is and $.L_1.itemstet \subset L_2.itemstet \subset L_3.itemset$

$L_1.\text{sup} \geq L_2.\text{sup} \geq L_3\ \text{sup}$. that implies This

Since $. \frac{L_3.\text{sup}}{L_1.\text{sup}} \geq \frac{L_2.\text{sup}}{L_1.\text{sup}}$ Thus,

implies it , $\frac{L_2.\text{sup}}{L_1.\text{sup}} \langle \min conf$ $\frac{L_3.\text{sup}}{L_1.\text{sup}} \langle \min conf$.

According to Theorem 1, if a lattice node {Y} is a child node of {X} in the FCIL and {Y}.sup/{X}.sup < minConf, then the child nodes of {Y} cannot form rules with {X} [10].

In this study support threshold and confidence threshold are considered respectively 50% and 70%. in this method at first, the degree of support of all items in 100 selected transactions from sum of all collected transactions of e-stores are calculated and then item with degree of support fewer than specified support threshold by user are deleted. In the next stage, item with more support degree is selected as root in network of proposed duplicated items and then its connection with the remaining items from previous stage are determined in the form of a graph. This connection contains the number of the presence of root item and other more frequent items in order of time. The result of these connections is two AA more frequent items. In Figure 5, the resulting graph of the first stage is shown in order to build two hot-swappable network exists no superset items.

In this study support threshold and confidence threshold are considered respectively 50% and 70%. in this method at first, the degree of support of all items in 100 selected transactions from sum of all collected transactions of e-stores are calculated and then item with degree of support fewer than specified support threshold by user are deleted. In the next stage, item with more support degree is selected as root in network of proposed duplicated items and then its connection with the remaining items from previous stage are determined in the form of a graph. This connection contains the number of the presence of root item and other more frequent items in order of time. The result of these connections is two AA more frequent items. In Figure 5, the resulting graph of the first stage is shown in order to build two hot-swappable network exists no superset items.

As shown in Figure 5, connection between root and next more frequent item is in the form of two directional arrows which on the edge of these arrows has been inserted a number. For example, to interpret the connection between two items 111 and 113, we see that the output flash that ends from item 111 to item 113, a weight is 26. This means that item 111 in all 100 transactions is bought 26 times before item 113. Naturally, the output flash from item 113 that ends to item 111 have the weight of 21 that indicate the frequency of presence of item 113 before than item 111. Connection between root item and other items have been explore in this way. At the end of this stage, those arrows that have less weight have been deleted. In this way, prediction is determined by the chorological order of

International Journal of Computer & Information Technologies (IJOCIT)
Corresponding Author: *Mohsen Dindar*
November, 2016

purchase items together. Network resulting from this stage is shown in Figure 6.

In this stage again the remaining items are compared with determined ensure threshold. The only difference is that in the previous stage threshold 50% evaluated for all transactions but in this stage support threshold is regarded to those transactions which at least one of the items is available on them. On the other hand, the number of transactions are reduced from 100 transactions to 65 transactions which item 111 exist in them and the number of frequency of items must be more than of new support threshold. If the number of presence of items were less than the amount , it were deleted from network of more frequent items and do not go to next stage that is building three AA more frequent items.

In the next stage three AA more frequent items are selected and added to network of more frequent items. Naturally those items that do not apply in support threshold condition wont added to network. Figure 7 show network resulting from adding items.

Created items network in this stage, given that support degree condition has been satisfied, can regarded as sequential patterns. As can be seen in figure 7. Created sequential patterns by proposed way which contain 5 patterns of two AA and 7 patterns of three AA are reduced in comparison with created patterns by freeman methods.

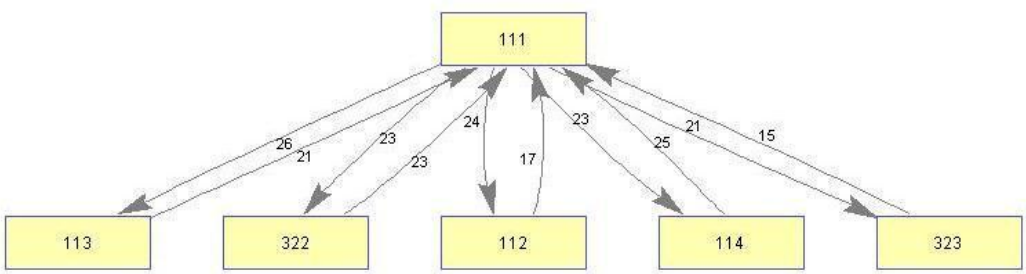

Figure 5: Frequent 2-itemset closed lattice

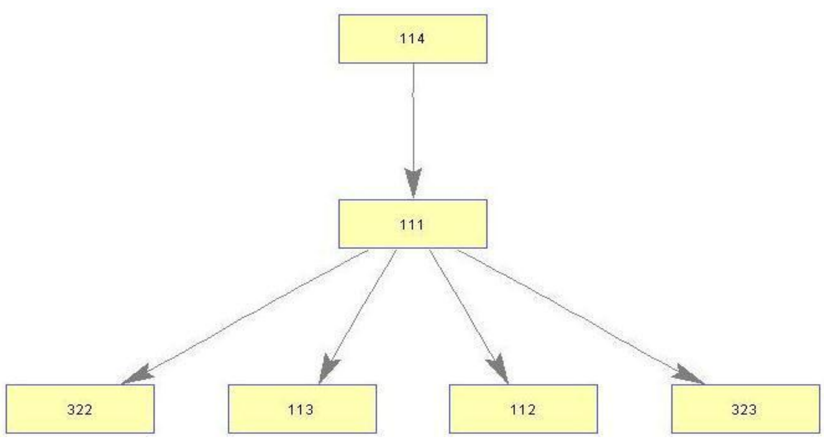

Figure 6: Frequent 2-itemset closed lattice after remove light weight edges

International Journal of Computer & Information Technologies (IJOCIT)
Corresponding Author: *Mohsen Dindar*
November, 2016

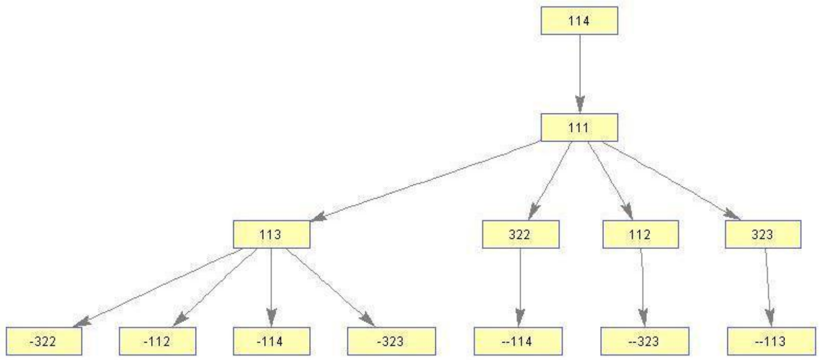

Figure 7: Frequent 3-itemset closed lattice

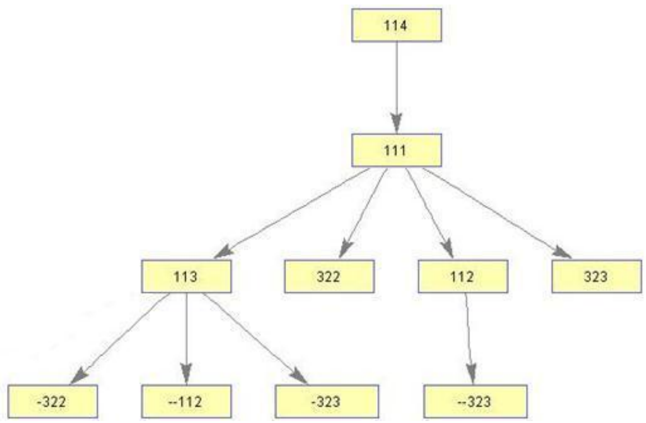

Figure 8: Frequent itemset closed lattice with generalized patterns

Surely, this differences has been seen with an increase in transactions.

Finally, generalized approach is used for decreasing created patterns in the network of three AA more frequent item. For this purpose according to proposition 1, we need ensure threshold that amount of which is equal to 70%. Those patterns which are not applied in the condition of proposition 1 are pruned so they have not additional computational load on the system. Figure 8 shows network of more frequent items with generalized patterns.

According to Figure 8, patterns which are not applied in condition of proposition one are deleted. It is obvious that with reducing patterns complexity of system reduced and less computation is done.

5. The Accuracy of the Proposed Method

After creating network of more frequent items and pruning this network by using a generalized method, now we evaluate the performance of the proposed method. The aim of evaluation of the performance of the proposed method is measuring degree of accuracy of this method in prediction of users visiting patterns. Also given that generalized method is done for the first time on sequential patterns mining technique, we want to compare degree of priority of this method to traditional one.

For this purpose we test the accuracy of the proposed method on an available standardized data collection. The high accuracy in this test is in the form of presence of at least one sequential patterns in one transaction and if there is no sequential patterns mining in one transaction, that transaction will be considered as training error for the created model. Data collection and testing errors in the proposed Freespan model are shown in Figure 9.

Also in Figure 9, the number of available errors in proposed model during the process of increasing the number of test data are shown.

As shown in Figure 10, the number of errors increase with gentle slope which reflects the fact that with increasing the number of test transactions, existing errors don't exceed the threshold 12%. In fact, the average of performance accuracy of proposed model is about 88%. Figure 26 show accuracy of the proposed model in test data collection.

This is a good accuracy rate and can present useful recommendations in order to guide customers in product purchase process. Surely with increasing of accuracy, product recommender system would be more flexible in providing purchase recommend and attracts many users.

Also, in Figure 11 has been shown the rate of encountering of patterns in test data collection. Patterns Encountering rate is the number of transactions which patterns were predicted rightly the presence of items in this transactions. Predicting of right items result in presenting product purchase recommends with high accuracy.

As shown in Figure 12, patterns encountering rate and indeed number of right prediction of patterns linearly increase with increasing number of available transaction in test data collection, this shows high accuracy of sequential patterns mining.

5.1. Comparison of the Proposed Method with Freespan Algorithm

After assessment of performance accuracy of the proposed method now, let's compare proposed method with algorithm Freespan. Freespan algorithm is one of the base algorithms for sequential patterns mining technique. The important thing about this algorithm is creating more sequential patterns that most of them are a part of redundant patterns and they impose more computational load and performance time. Now for comparing between these two models, we regard number of produced patterns by each model. For mining patterns comparison are done between more frequent items and available transactions that comparison be more with increasing of patterns and so will bring more performance time and computational load for system; so whatever produced patterns by a method are less, so computational volume and performance time of algorithm will be less. However we must bear in mind that decreasing of patterns number to a threshold is possible and if number of patterns mining be less than this threshold affect negatively on system performance accuracy. With reducing accuracy of performance, created recommendations won't be accurate so

pp. 95- 110

International Journal of Computer & Information Technologies (IJOCIT)
Corresponding Author: *Mohsen Dindar*
November, 2016

cause customer dissatisfaction. Table 2 shows number of done comparison between pattern and available transactions between two methods of Freespan and proposed model.

As shown in table above, number of sequential patterns created by proposed method is significantly less than Freespan algorithm. Analysis of the proposed method of calculation times versus Freespan method has been shown in Table 3.

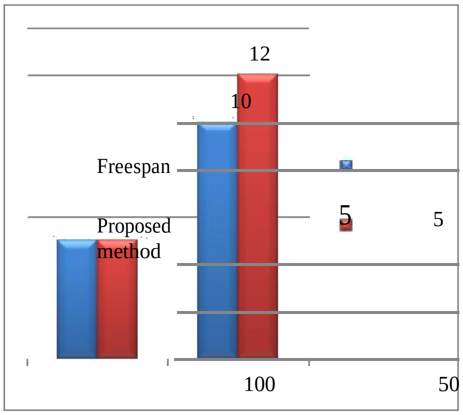

Figure 9: Comparison of the proposed method with Freespan algorithm in 100 data test

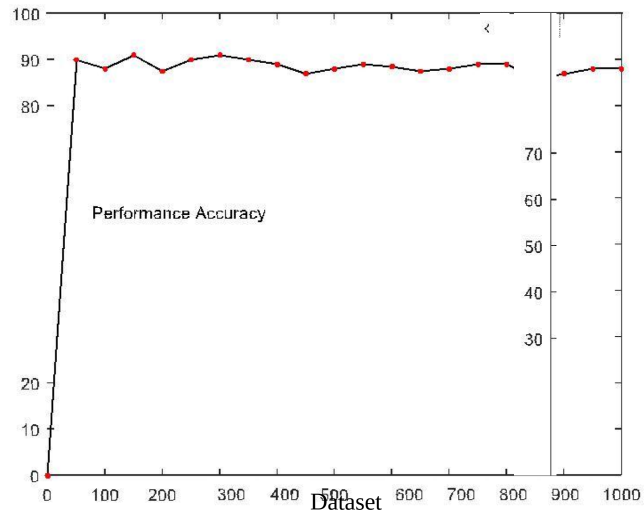

Figure 11: The accuracy of proposed method in test data

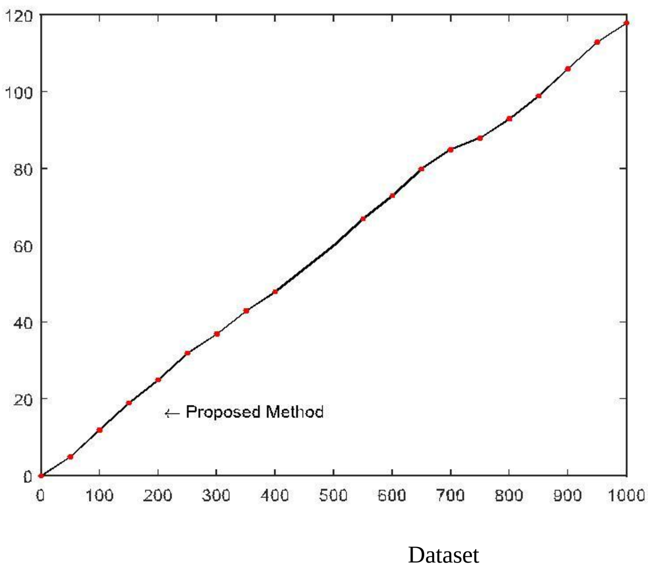

Figure 12: Hit rate in data test

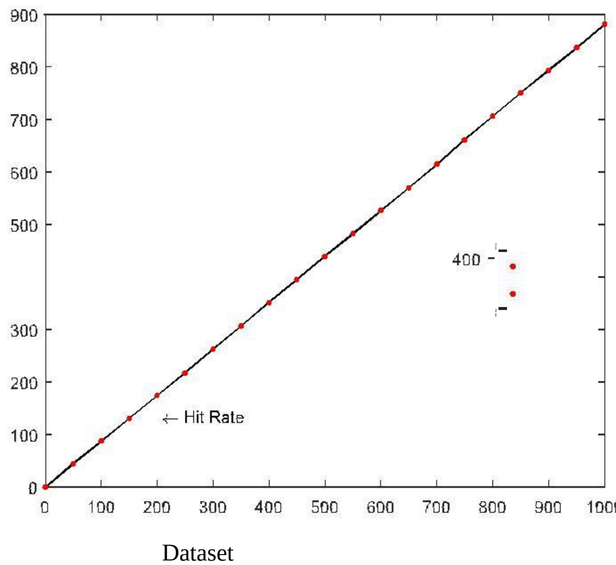

Figure 10: Number of errors in proposed method

Table 2: Comparison of the proposed method with Freespan algorithm

pp. 95- 110

	mined sequential pattern for 100 transactions with n=19 items and support 30%	mined sequential pattern for 100 transactions with n=12 items and support 40%	mined sequential pattern for 100 transactions with n=6 items and support 50%	mined sequential pattern for 100 transactions with n=3 items and support 60%	performance accuracy	system complexity
Freespan	34/	129	23	9	90%	$O(n^2)$
proposed method	62	31	12	4	88%	O(nlogn)

Table 3: Analysis of comparison of the proposed method with Freespan algorithm

Analysis	Proposed method	Analysis	Freespan
n	For i=1 to TR	N	For i=1 to TR
{c	IF Sup(i)>= min_sup	{c	IF Sup(i)>= min_sup
C	Then FI=+i	C	Then FI=+i
}	End	}	End
n	For j=1 to FI	n	For j=1 to FI
{logn	For k= j+1 to FI-j	{n-1	For k= j+1 to FI
c	IF Sup(jk)>= min_conf	c	IF Sup(jk)>= min_conf
c	Then SP=+(jk)	c	Then SP=+(jk)
}	End	}	End
	End		End
	End		End

In table 3, TR represents number of more frequent items and SP represents number of total more frequent items and in relative column to analysis, n represents linear order of time and c represents fixed number which is considered 0(1) in calculating of order of time so order of time of proposed method is calculated I comparison to Freespan method as below:

$$\begin{cases} Freespan: T(n) = n + \frac{n(n-1)}{2} + c \in O(n^2) \\ Proposed\ method: T(n) = n + n(logn) + c \in O(nlogn) \end{cases} \quad (4)$$

So computational complexity of proposed method is O(nlog n) which n is the number of attended items in transactions which in comparison to computational complexity of Freespan algorithm which is O(n2) function highly optimized.

Figure 13 shows comparison of the numbers of produced sequential patterns by proposed method and Freespan algorithm with the increasing the number of more frequent items.

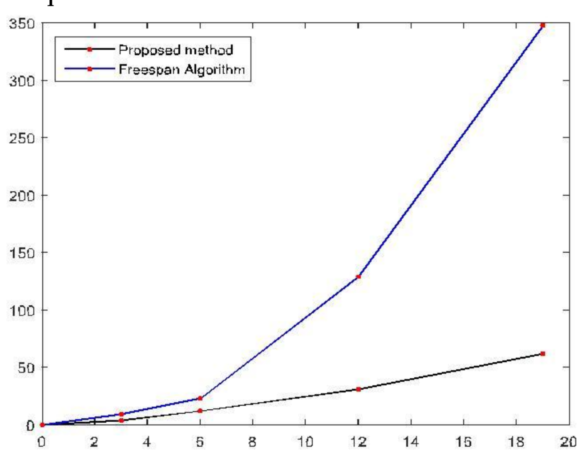

Figure 13: Hit rate in data test

As shown in Figure 13, number of produced patterns in proposed method are significantly reduced with the increasing the number of one A more frequent items to the Freespan algorithm. this reduction causes redundant and duplicate patterns in patterns mining to be wasted while no much damage to the performance accuracy of the system.

6. Conclusion

Overload information about the finding on line information in a communication network has become a serious problem that more efforts have been made to solve the problem. Product recommender system is one of the ways to deal with this problem that has been adopted in the field of e-commerce. the aim of reduce recommender system is providing purchase recommendation in order to guide electronics stores customer to avoid wasting time and confusion among the available products in stores. Providing accurate recommendations increase customer satisfaction and also increase sales of products online store and finally causes further e-commerce boost.

In this study has been introduced product recommender system in e-shop by using data mining applications. In this recommender system, at first, available products in electronic catalog database is clustered by using of C-means method which is one of the clustering methods off data mining applications. In the next stage, generalized sequential pattern mining method based on network approach was used for extracting the more frequent patterns among transactions. Since this method is sensitive to the presence of items in transactions, so follow the chronological order of buying items from electronic stores among patterns. Then a generalized approach is used which according to the ratio of the node support degree and ensure threshold that prune redundant pattern and reduce computational load and time execution system.

Evaluation of the proposed method performance showed that accuracy of this method is about 88% which have high accuracy and provides accurate recommendation to website users. Also, in comparison to Freespan method which is one of the base methods for sequential patterns mining, have been shown that proposed method have higher operating speed and lower computational complexity.

Also patterns reduction don't have so much negative impact on the accuracy of the proposed model.

References

[1] Tan Pang-Ning, Steinbach Michael, Kumar Vipin,(2006),"Introduction to Data Mining", Pearson Addison-Wesley

[2] Qinbao Song, Martin Shepperd. (2006), "Mining web browsing patterns for E-commerce", Computers in Industry, Vol.57, pp.622-630.

[3] Wen-Shan Lina, Nathalie Cassaigne, Tzung-Cheng Huan, (2010), "A framework of online shopping support for information recommendations", Expert Systems with Applications, Vol.37, pp.6874–6884.

[4] Yung-Ming Li, Chun-Te Wu, Cheng-Yang Lai. (2013), "A social recommender mechanism for e-commerce: Combining similarity, trust

pp. 95- 110

and relationship", Decision Support Systems, Vol.5, pp.740–752.

[5] Liyi Zhang ,Jing Zhu ,Qihua Liu. (2012), "A meta-analysis of mobile commerce adoption and the moderating effect of culture". Computers in Human Behavior, Vol.28, pp.1902–1911.

[6] Li Chen, Feng Wan, (2013), "Preference-based clustering reviews for augmenting e-commerce recommendation", Knowledge-Based Systems, Vol.50, pp.44–59

[7] Yu Zhang, Jing Bian, Weixiang Zhu. (2013), "Trust fraud: a crucial challenge for China's e-commerce market". Electronic Commerce Research and Applications, Vol.12, pp.299-308.

[8] Wei Shen, Jianyong Wang, Jiawei Han. (2014), "Sequential Pattern Mining", Springer International Publishing Switzerland, chapter 11.

[9] Yiyang Zhang, Jianxin (Roger) Jiao. (2007), "An associative classification-based recommendation system for personalization in B2C e-commerce applications". Expert Systems with Applications, Vol.33, pp.357–367.

[10] Bay Vo, Tzung-Pei Hong, Bac Le. (2013), "A lattice-based approach for mining most generalization association rules", Elsevier, Knowledge-Based Systems, Vol.45, pp.20-30.

[11] Soon Jae Kwon, Namho Chung. (2010), "The moderating effects of psychological reactance and product involvement on online shopping recommendation mechanisms based on a causal map", Electronic Commerce Research and Applications, Vol.9, pp.522-536.

[12] Lingyun Qiu, Izak Benbasat. (2010), "A study of demographic embodiments of agents in electronic commerce", Human-Computer Studies, Vol.68, pp.669-688.

[13] Tan Pang-Ning, Steinbach Michael, Kumar Vipin,(2006), Introduction to Data Mining, Pearson Addison-Wesley,chapter8 .

[14] T. Velmurugan. (2014), "Performance k-Means between analysis based clustering Fuzzy C-Means and oriented connection for algorithms Elsevier data", telecommunication Computing, Vol.19, Soft Applied pp134–146.

[15] J. Wang. (2009), "Sequential patterns Encyclopedia of Database Systems", LING LIU and M. TAMEROZSU (Eds.), pp.2621–2626.

pp. 95- 110