A two-level Product Recommender for E-commerce Sites by Using Sequential Pattern Analysis

Shahram Jamali[1,*], Yahya Dorostkar Navaei[1]

[1]Department of Computer Engineering
 University of Mohaghegh Ardabili, Ardabil, IRAN.

Abstract: With the development of communication networks and rapid growth of their applications, huge amount of information have been produced. Major part of these information are in electronic stores, and hence it's really hard to find desired products inside *huggermugger*. Product Recommendation System (PRS) tries to solve this problem by giving appropriate and fast recommendations to the customers. This paper proposes a two-level product recommender for E-commerce sites. At first, the available products are clustered by using C-Means algorithm to create groups of products with similar characteristics. Then, the second level considers the customers' behavior and their purchase history for drawing the relationships between products by using Sequential Pattern Analysis (SPA) method. These relationships, eventually, will lead to appropriate recommendation for customers and also increases the likelihood of selling related products in electronic transactions. Extensive numerical simulations over *UCI transactions 10k* dataset indicates that 87% of records in mined sequential patterns are predicted correctly and the accuracy of recommendations is more than other RPSs.

Keywords: Product recommendation system, two-level RPS, e-commerce, clustering, sequential pattern analysis

1. Introduction

In recent years, rapid growth in products of electronic stores causes problems for online products search among a lot of same products [1]. So customers get confused and often can't find desired products and it causes wasting time. At this time, it's valuable to solve this problem with a product suggestion. A product suggestion is a recommendation that advises some products according to customers' need to guide them for purchase. It can be extracted of prior transactions by data mining techniques.

Nowadays, with using of data mining and applying the rules, we can create models on data to reveal implicit knowledge and embedded information. Data mining applications are growing increasingly in various fields [2]. These applications can cover many aspects of the web such as electronic stores, online purchase and so on. This information can show how to use the web services for users to reach their goals. Undoubtedly with access to this information, we can design a flexible web interface that users enjoy interact it and they can achieve their purposes easily and fast. Product recommendation system is one of the flexible interfaces that can be created relied on mined information [2].

PRS is a service that takes a set of the user's criteria as input and looks for products among the items in the database according to the criteria and finally suggests the list of products to user as output [3]. Users can find require products by this recommendation without wasting time and confusion and make more accurate decisions

about purchase product. In addition, PRS can predict products that are closely related to purchase products by monitoring history of customer's purchase behavior.

In this paper, we use data mining techniques to provide a recommender system in E-commerce sites. Using this application can give information about available products in electronic stores to customers. Also commercial transactions monitoring can mine rules as sequential pattern to discover potential relationship between the products in the store and suggest these related products to customers. To achieve these aims, PRSs have been developing in e-commerce field.

This paper is continued by five sections. Section 2 includes related work about PRS. Section 3 contains proposed PRS. Implementation of proposed PRS comes in section 4 and finally concluding remarks of this paper and discusses about future works are given in section 5.

2. Related Works

The rapid development of the e-commerce platforms has made marketers to devise online PRS to assist the customers in their purchase process and persuade them to make decisions. In the other hand, customers have demanded more personalized information delivery services. Hence, all PRSs try to provide more personalized services. So, we can find many product PRSs that work as online guide for customers. Also, we can divide online PRSs into two approaches: review-based and feature-based.

Review-based approach focuses on other customers' opinion about a special product. This approach recommends items to a consumer based on other customers' purchase decisions that have similar preferences.

This approach is introduced as content-based technique [4] that recommends items similar to those had been bought by other customers. These kinds of PRSs have low quality recommendations [4], which is due to three major issues. First, these cannot provide recommendations unless multiple-item purchasing profiles for a number of consumers, or at least for the consumer currently using the system, are available. Second, preference estimates based on purchasing profiles are inaccurate when, as is often the case, these profiles contain products purchased as gifts for or on behalf of other consumers. Third, purchasing profiles are historical data, revealing past but not necessarily current preferences [4] for example a social recommendations system with combining similarity, trust and relationship that detects priority of the members through close friends and social network[3]. The basic idea of trust and reputation systems is to obtain a score for users. According to these results, other users can decide whether they are traded by a trusted user or not. [3]. Other PRS is proposed in [5] that it's a solution on personalized products recommendation based on user-contributed photos and corresponding textual descriptions from social media sites. Then the PRS recommends related products with these descriptions to user [5]. Other one based on clustering comments of user and reviews that related comments about products presented by reviewers and users are clustered [6].

Other PRS is presented based on associative classification method that, for the product recommendation issue builds an evolving system [7]. In this system first, customer required products data are gathered and transformed into proper phrase datasets. Then data mining procedure starts to search for a set of associated, frequently occurring phrase patterns (classifiers). Finally products related to extracted pattern are advised to customer [7]. Also other PRS is developed [8] that called HOPE, which integrates CF-based recommendation using implicit rating and SPA-based recommendation that calculates explicit rating for each user. Then it finds k neighbors that have similar rating for each target user.

Feature-based approach focuses on the features of products. This approach predicts matched products with customers' criteria for example a PRS with the dynamic templates [9]. Since users have different needs at different times, so behavior of the users during the lifecycle is considered and the related products in each period of lifecycle are recommended [9]. Also a PRS introduced a prototype of e-commerce portal called e-Zoco that contains hierarchically catalogue of products, a product selection service and a rule-based knowledge learning service. It recommends products to the users with knowledge about the existing relationships among the attributes that describes a given product category [10].

Another PRS based on redesign promotion strategy for e-commerce competitiveness through pricing and recommendation is presented [11] that discounts products and encourages customers to purchase them.

PRSs try to deal some challenges that faced involved such as demand variations, seasonal fluctuations, and stockless policy in inventory management and the risks such as lost sales, lost customers, low customer satisfaction associated with the same [12,15]. Absolutely, a PRS cannot solve all challenges but each PRS tries to improve some aspects of the problem.

In this paper we employ the feature-based approach and measure the relation between products to provide accurate recommendation to the customers.

3. Proposed Product Recommendation system

In this section, we describe the clustering and SPA methods briefly and present their reasons and algorithms that are used in our research.

3.1 Preliminaries

In this subsection, we explain C-means algorithm to classify products and Freespan algorithm to extract frequent patterns in following subsections.

3.2 C-Means algorithm

Generally clustering be used to create groups of objects based on their features in such a way that the objects belonging to the same groups are similar and those belonging in different groups are dissimilar. Also we want to separate products by these types and create groups with similar features.

So we employ C-Means clustering algorithm to classify in products flexible clusters. These clusters can be achieved based on fuzzy logic and its variants have been proved to be competitive to conventional clustering algorithms. The advantage of this algorithm is that it doesn't consider sharp boundaries between the clusters, thus allowing each feature vector to belong to different clusters by a certain degree. The degree of membership of a feature vector to a cluster is usually considered as a function of its distance from the cluster centroid points.

It is based on minimization of the following objective function [13]:

$$J_m = \sum_{i-1}^{N} \sum_{i-1}^{C} u_{ij}^{m} \left\| x_i - c_j \right\|^2, 1 \le m \le \infty \qquad (1)$$

Where m is any real number greater than 1, u_{ij} is the degree of membership of x_i in the cluster j, x_i is the ith of d-dimensional measured data, c_j 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 u_{ij} and the cluster centers c_j by [13]:

$$u_{ij} = \frac{1}{\sum_{K=1}^{C} \left[\left\| x_i - c_j \right\| / \left\| x_i - c_K \right\| \right]^{\frac{2}{m=1,\ldots,c_j}}}$$

$$= \left[\frac{\sum_{i=1}^{N} u_{ij}^{m} x_i}{\sum_{i=1}^{N} u_{ij}^{m}} \right] \qquad (2)$$

This iteration will stop when [13]:

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

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 J_m [13].

3.3 Freespan algorithm

SPA [14] discovers subsequences that occur in a sequence database with frequency bigger than a user-specified threshold. The formal definition of SPA is presented as follows:

Given a set of sequences, where each sequence consists of a list of elements and each element consists of a set of items, and given a user-defined min_support threshold, sequential pattern mining is to find all frequent subsequences, i.e., the subsequences whose occurrence frequency in the set of sequences is no less than min_support [14].

Freespan mines sequential patterns by partitioning the search space and projecting the sequence sub-databases recursively based on the projected itemsets [14]. It calculates *min_support* value for all items that appear in transactions. Items with the *min_support* value less than the threshold are eliminated and remain considered as i-frequent itemsets that i=1,2,3,… .This algorithm makes a 2-dimension matrix that all items placed in both horizontal and vertical axis. Each element of matrix in intersection of two items includes pair *(i,j,k)* that *i* denotes number of appearance first item before second one and *j* denotes number of appearance first item after second one and *k* denotes number of appearance two items together. Each of *i, j* is less than threshold will be eliminated. Finally rules with support more than threshold will be considered as sequential patterns.

The detailed presentation of the Freespan algorithm, the proof of its completeness and correctness, and the performance study of the algorithm are in [14].

3.4 Proposed Framework and Algorithms

Proposed PRS includes two levels of product recommendation that first level is recommended before product purchase and other one after purchasing. This PRS initially recommends products that closed to customer's criteria to avoid wasting time. In second level it recommends associated products to purchased product to complete buying process and make the customer aware of potentially related products. Fig. 1 illustrates the overall framework of proposed PRS.

In this section, we describe the overall framework of proposed PRS. First, we collect products' data from electronic store. We separate the products according to their type. Then, these products are clustered based on their numerical attributes in three separate clusters of high, medium and low quality by C-means algorithm. The properties of products into clusters are very similar to each other. However, the products in separate clusters have different properties.

After clustering products, an important challenge that seems is adding new products to the system. In this paper, we have solved this problem by using decision tree classification method that will be presented in next section.

Next, the PRS tries to identify customers' requirements and criteria. In order that, we have used an online form in electronic store that takes information about product such as type, quality, price, brand, etc. Thus, this information is used to assign an appropriate cluster to the customer.

In next step, we collect information about history of customers' shopping behavior from the electronic store. This information is used to explore relations between products by Freespan algorithm of SPA method. Eventually, these relations and patterns will be provided as product recommendations to buy to customers. The relationships between the products will increase the likelihood of buying the products together.
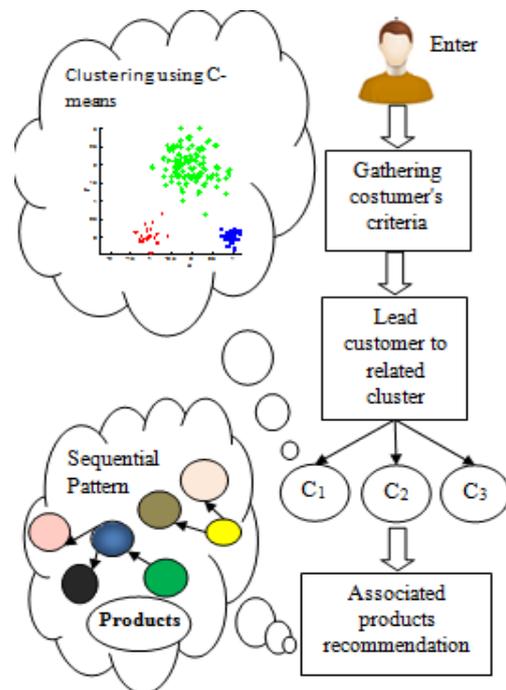

Fig. 1 The overall framework of the proposed PRS.

4. Results and Discussion

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 e-commerce 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) 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. Table 1 shows a part of data about *cellphone* dataset that are available in Tebian electronic store. This table 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 Fig. 2. 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 Fig. 3.

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. Fig. 4 illustrates final place of centroid points in each clusters that c_1, c_2, c_3 are means of

Low quality, *Middle quality* and *High quality* clusters, respectively.

Then each point is assigned to cluster based on u_{ij} 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 Fig. 5. Also table1 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.

4.3 The Proposed PRS Development

Due to the increasing variety of products in the online store, recommendation system must be able to respond it. It means, the products that are added after clustering to online store should be placed in related clusters correctly else either product will be remained as no-clustered or clustering will be false, and provided recommendations will be incorrect.
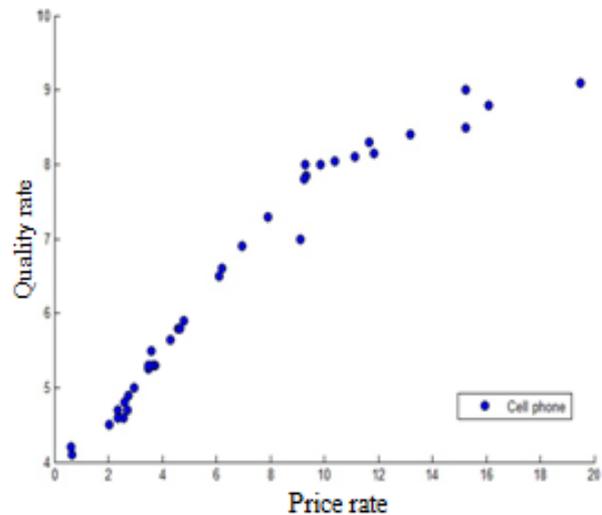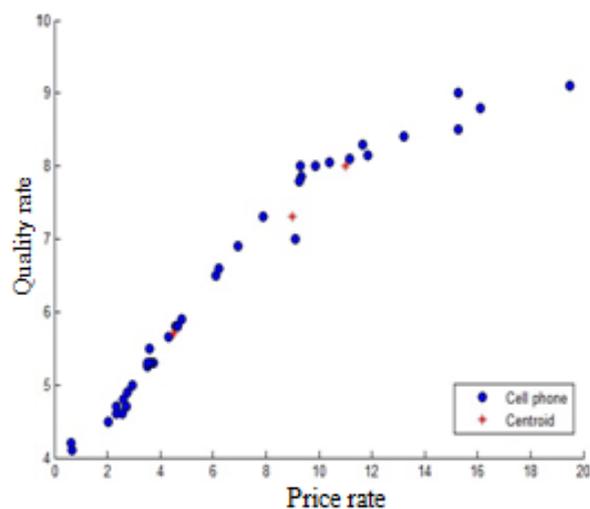

Fig. 2 Scattering data of cell phone.

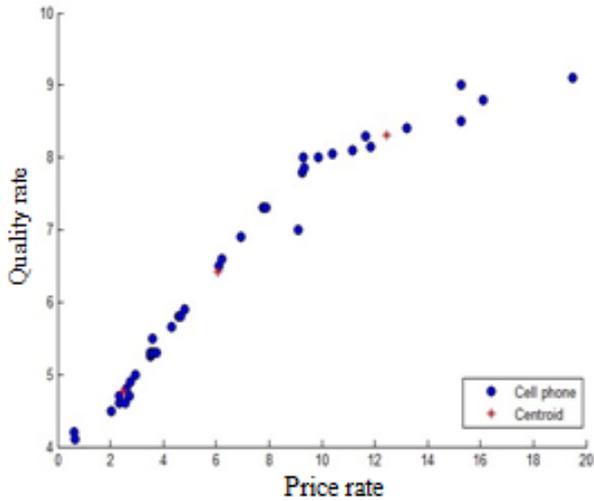Fig. 3 Selection centroid points random.
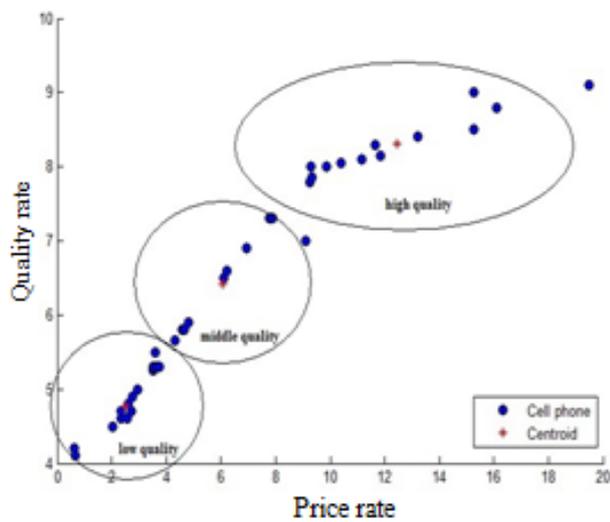
Fig. 4 Correct place of each centroid points.

Fig. 5 Clustering for data related to cellphone.

The new products can be added to the system in two ways.

i.   New type of products
ii. Different type of available products

First way solves as clustering a new type of products and second one must be classified.

For classification we use Decision Tree approach. We should calculate the membership value of each product to the three clusters. Each of this membership values is higher; so the product is allocated to this cluster.

To do this, we will calculate the following relations [13]:

$$A = \frac{1}{\left[\|x_i - c_1\|/\|x_i - c_k\|\right]^2} \tag{4}$$

$$B = \frac{1}{\left[\|x_i - c_2\|/\|x_i - c_k\|\right]^2} \tag{5}$$

$$C = \frac{1}{\left[\|x_i - c_3\|/\|x_i - c_k\|\right]^2} \tag{6}$$

Where A, B, C are membership value of c_1, c_2, c_3, respectively. Above formulas are the modified form of equation (2) [10].

4.4 Colleting Customer's Criteria

After clustering products, it's time to collect customer's criteria and requirements. Actually customer's criteria are information about the required products that help us to assign closed cluster to customers.

In this paper, we prepare an online form that has certain options that customer can choose one of the available options between several options. The advantage of this form is to prevent the redundant and noisy information. This form contains three categories of information as following.

i.   *Type Information*: in this part, customer selects the type of required product.
ii.   *Cluster Information*: in this part, customer selects the quality and price of required product.
iii. *Additional Information*: in this part, customer selects the brand and manufacturing date of required product.

Customers fill this form and give correct information about desired product, so PRS suggests related cluster to them.

Fig. 4 Decision tree for PRS development.

4.5 SPA Implementation Details

In this section we discover patterns and relations between clustered products by using SPA method. So, we use extracted transactions from standard dataset. In these transactions each product is shown by product identification number. These transactions are probed and Table 5 is gained as available items with their *support*. The *support of an item* is simply the relative frequency of occurrence of an itemset in the transaction set. The

19

Table 1 Distance of each point to clusters.

product ID	Price (Rials)	quality rate	price rate	Type	Distance			Cluster
					Centroid1 (2.4767, 4.7654)	Centroid2 (6.0754, 6.4115)	Centroid3 (12.4669,8.3115)	
78277	2700000	4.7	2.7	cell phone	0.2391	3.7854	10.413	low quality
78282	4600000	5.8	4.6	cell phone	2.3680	1.5971	8.2581	middle quality
78286	10400000	8.05	10.4	cell phone	8.5833	4.6246	2.0834	high quality
79025	4650000	5.8	4.65	cell phone	2.4130	1.5510	8.2105	middle quality
79029	6100000	6.5	6.1	cell phone	4.0232	0.0919	6.6196	middle quality
79129	7900000	7.3	7.9	cell phone	5.9924	2.0294	4.6776	middle quality
76358	13200000	8.4	13.2	cell phone	11.329	7.3969	0.7384	high quality
76361	6200000	6.6	6.2	cell phone	4.1568	0.2260	6.4964	middle quality
76369	3500000	5.25	3.5	cell phone	1.1383	2.8252	9.4751	low quality
76382	6950000	6.9	6.95	cell phone	4.9626	1.0018	5.6946	high quality
76384	3630000	5.3	3.63	cell phone	1.2773	2.6862	9.3359	low quality
76386	2950000	5	2.95	cell phone	0.5343	3.4294	10.414	low quality
76395	15250000	9	15.25	cell phone	13.463	9.5328	2.8670	high quality
76460	650000	4.1	0.65	cell phone	1.9378	5.8973	12.545	low quality
76469	630000	4.2	0.63	cell phone	1.9249	5.8773	12.531	low quality

Table 2 A part of data about cellphone dataset.

product ID	product name	quality range	price range
78277	Huawei Ascend Y210D	4.7	2.7
78278	Huawei Ascend Y210	4.6	2.55
78279	Huawei Ascend Y300 D	5.3	3.75
78282	Huawei Ascend G510	5.8	4.6
78286	Huawei Ascend Mate	8.05	10.4
78290	Huawei Ascend P6	8	9.85
79025	G510	5.8	4.65
79029	Huawei Ascend G610	6.5	6.1
79129	Huawei Ascend G700	7.3	7.9
79204	Huawei Ascend Y220	4.6	2.35
76353	C3312 Duos	4.5	2.02
76355	Galaxy Ace S5830	5.65	4.3
76357	Galaxy Mini 2 S6500	5.5	3.59
76358	Galaxy Note II N7100 - 16GB	8.4	13.2
76361	Galaxy Wonder I8150	6.6	6.2
76362	Galaxy Y S5360	4.9	2.75
76367	Galaxy Ace S7500	5.9	4.8
76369	Galaxy Mini S5570	5.25	3.5
76382	Galaxy S Advance I9070-8GB	6.9	6.95
76383	Galaxy S III I9300	8	9.3
76384	Galaxy Y Duos S6102	5.3	3.63
76386	S3850 Corby II	5	2.95

confidence of a rule measures the likelihood of occurrence of the consequent of the rule out of all the transactions that contain the antecedent of the rule. Confidence provides the reliability measure of the rule [16].

In this work, the *min-support=50%* and *min-confidence=80%* are considered. *Freespan* algorithm is applied on these transactions for mining sequential pattern and drawing the relationships between products.

As explained in the previous section, the candidate i-itemset is calculated for i=1,2,3,…, and then according to the algorithm, patterns with high confidence selected as candidate frequent patterns. Then according to the time ordering candidate frequent patterns will change to the sequential patterns. In table 3 items are sorted in descending order of support. Items with support less than the threshold are pruned.

In the next step, two-dimensional matrix is created that frequent items are written at the beginning of each row and column, and the confluence of each items indicates a 2-length itemset. Each element of the matrix consists of three values (A, B, C) that A represents the number of occurrences of first item before second item and B represents the number of occurrences of first item after second item and C represents the number of occurrences of two Items together. Then, 2- length frequent itemsets are pruned according to predefined *min-confidence* and remained itemsets are considered as candidate frequent patterns. Finally, sequential patterns are extracted from candidate frequent patterns with high confidence by considering the occurrences time of items in transactions. As shown, using *Freespan* algorithm, candidate frequent itemsets were extracted from transactions and by these itemsets sequential patterns with high confidence were achieved.

Table 4 shows a summary of sequential patterns that extracted from candidate frequent patterns. In this table, directed arrows indicate time order in occurrences time of items in transactions. In other word, it shows the order of products purchasing in electronic store.

Now using these patterns, we discover the relations between products and create second level of product recommendation. Fig. 6 shows relations between products. When customer selects one of the products, other linked products are suggested to customer as second level of product recommendation.

Where in this figure each product is shown with an icon and arrows indicate a close relationship between these products. Since, sequential patterns mining consider to the time of items occurrence in the transaction, arrows are drawn as unidirectional to indicate choice ordering. The double arrows indicate equal number of items occurrence before and after each other. Related product suggestion not only does cause prevent of wasting time in product finding, but also increase sales and profitability of the electronic store.

Table 3 Available items with their support

f-list item	Support	f-list item	Support	f-list item	Support
111	31	234	12	134	3
113	30	223	11	311	3
322	29	241	11	342	3
114	28	242	9	444	3
112	27	244	9	434	3
323	26	224	8	424	3
332	22	341	8	122	2
331	22	314	7	131	2
123	20	433	7	133	2
143	19	321	6	141	2
324	19	243	6	333	2
343	18	313	6	432	2
144	17	431	6	441	2
344	17	423	6	121	1
142	16	422	6	214	1
232	16	124	5	312	1
222	15	442	5	334	1
233	15	212	4	413	1
221	14	443	4	411	1
211	12	414	4	412	0
231	12	132	3	421	0

4.6 Performance Evaluation

Performance evaluation is applied to determine accuracy of extracted sequential patterns on the test dataset. It is a part of the standard data that is used to explore sequential patterns in the previous section.

We first determine the number of extracted patterns on dataset in table 5. It's important to note, with increased number of transactions, the sequential patterns grow slowly, and most of the patterns are extracted in primary steps. So, the mined sequential patterns have enough accuracy to recommend. Also this result illustrates in Fig. 7. It shows growth graph of sequential patterns according to transactions increase.

Next, the mined sequential patterns are applied on the test dataset. Then to determine the percentage of correct prediction of sequential patterns, we will make a test. Table 6 shows the number of test dataset that mined sequential patterns are applied on them and the number of transactions that are predicted inaccurately. These transactions define as *Errors*. It's noticeable that the number of errors has light speed growth with increase of the test dataset. It shows the high performance accuracy of the proposed PRS. Fig. 8 shows the error graph and Fig. 9 shows the hit rate on data tests.

Table 4 Extracted sequential patterns.

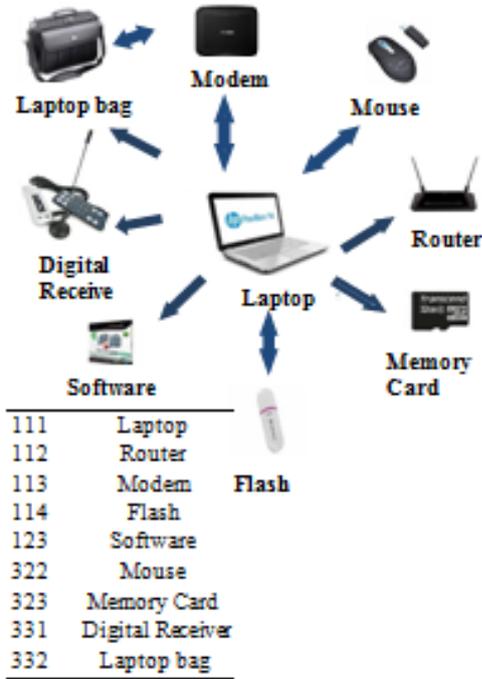| Number	Sequential Patterns	Number	Sequential Patterns
1	‹111›↔‹113›	13	‹113›→‹331›
2	‹111›↔‹322›	14	‹332›→‹113›
3	‹111›↔‹114›	15	‹113›→‹123›
4	‹111›→‹112›	16	‹322›→‹114›
5	‹111›→‹323›	17	‹322›→‹323›
6	‹111›→‹331›	18	‹322›→‹331›
7	‹332›→‹111›	19	‹114›→‹112›
8	‹111›→‹123›	20	‹114›→‹323›
9	‹113›↔‹322›	21	‹114›→‹331›
10	‹113›→‹114›	22	‹332›→‹114›
11	‹113›→‹112›	23	‹112›→‹323›
12	‹113›→‹323›	24	‹332›→‹323›

Fig.6 Relations between products

Table 5 Extracted patterns number on dataset.

Transactions number	Extracted patterns number
60	24
100	27
150	29
200	31
300	32
400	33
500	33
800	34
1000	35

Table 6 Errors on data test.

Data test number	Errors number
30	4
50	6
100	12
150	17
200	25
300	37
400	52
500	68
800	100
1000	128

Table 7 Accuracy of proposed system

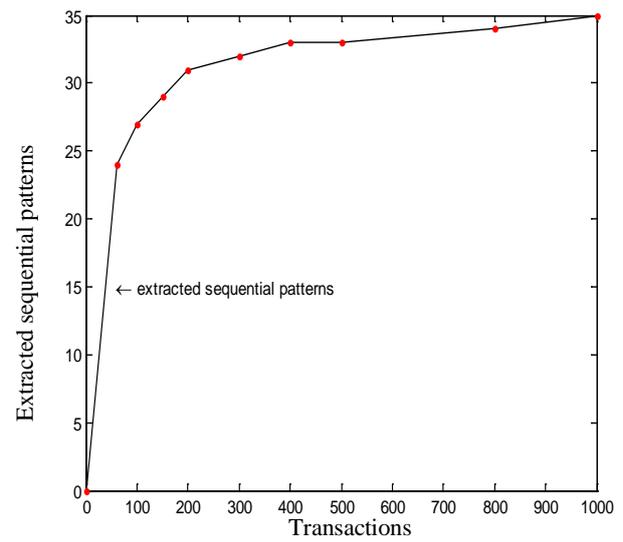| # of test	# of error	Accuracy rate	# of test	# of error	Accuracy Rate
50	5	90%	550	72	86.9%
100	16	84%	600	78	87%
150	24	84%	650	87	86.61%
200	28	86%	700	89	87.28%
250	37	85.2%	750	95	87.33%
300	43	85.67%	800	99	88.38%
350	47	86.58%	850	105	88.35%
400	53	86.75%	900	110	87.78%
450	58	87.11%	950	118	87.58%
500	65	87%	1000	125	87.5%
Average accuracy			86.91%		

Fig.7 Numerical data showing the number of extracted sequential patterns from users' transactions in e-store
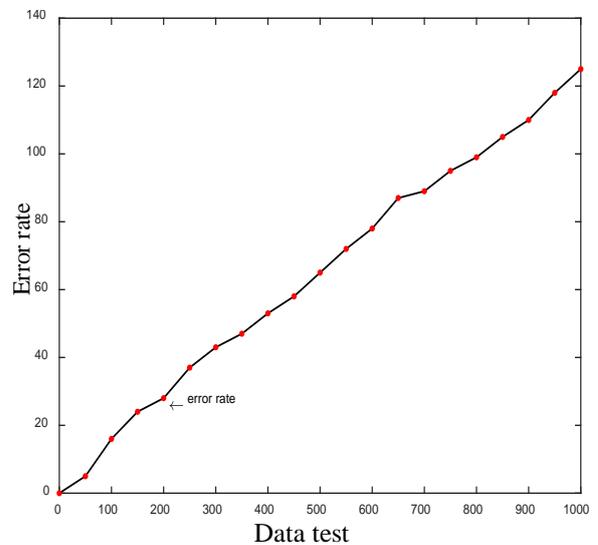

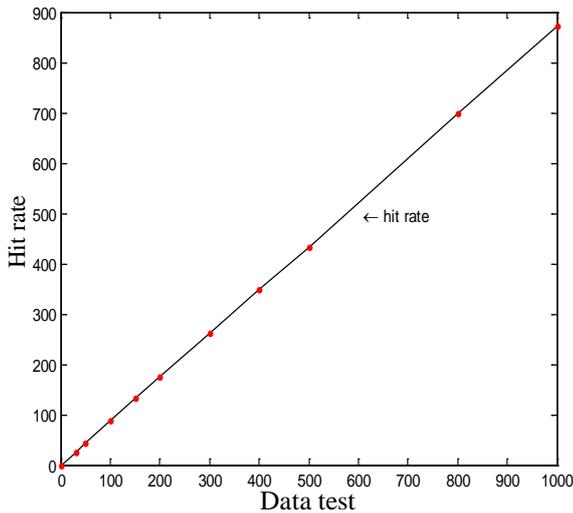Fig.8 Numerical data showing the number of transactions are predicted inaccurately in data test as errors

Fig.9 Numerical data showing the number of patterns that predict correctly in data set as hit rate.
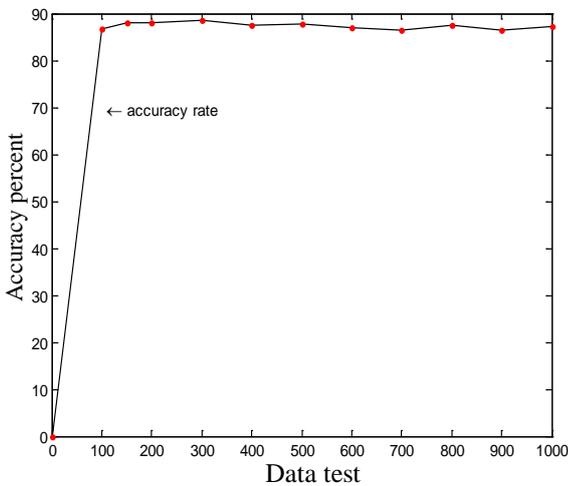

Fig.10 Numerical data showing the accuracy percent of the proposed system on the data test.

Accordingly, we assume that the set of sequential patterns C on S record of test data T with symbol v_s are applied and the accuracy of recommendation system is compared as following:

$$a = \sum_{i=1}^{s} \frac{v_s}{S}$$

$$v_s = \begin{cases} 1 & if \forall v_s \in T_s, \exists v(t_i) \in C \\ 0 & otherwise, i = 1,2,3,... \end{cases}$$

(5)

Where α is called the accuracy of the recommendation that shows the percentage of transactions which are predicted correctly in the test and t_i is items that are present in the record. The ratios of records that are predicted correctly to the total number of records represent accuracy of the testing dataset of mined sequential patterns.

Experimental result indicates that 87% of test dataset records are predicted correctly by mined sequential patterns that it shows high accuracy of the proposed product recommendation system. Fig. 10 shows the accuracy rate of the proposed system on the data test.

5. Summary

In this paper, a new method has been introduced called the two-level PRS in e-commerce sites using data mining applications. This PRS is advised products at two levels. First level of product recommendation is before choosing product where products are clustered using C-means algorithm. Customer product criteria are collected through an online form and customers are leaded to related cluster. Also to add new products to the clusters, an approach through the decision tree classification method is provided.

The second level of product recommendation is performed after product selection where potential relationships between products are discovered using sequential pattern analysis. At this level, customers select the product and related products are advised to them.

Finally to evaluate the performance used standard test dataset. Experimental result indicates that 87% of test dataset records are predicted correctly by mined sequential patterns. This result indicates that the accuracy of recommendations for the proposed PRS is higher than other PRSs.

References

[1] Tan Pang-Ning, Steinbach Michael, Kumar Vipin. Introduction to Data Mining. *Pearson Addison-Wesley,* (2006), chapter 1, 8.

[2] Qinbao Song, Martin Shepperd. Mining Web Browsing Patterns for E-commerce. *Computers in Industry,* volume 57, (2006), pp. 622-630.

[3] Yung-Ming Li, Chun-Te Wu, Cheng-Yang Lai. A Social Recommender Mechanism for E-commerce: Combining Similarity, Trust, and Relationship. *Decision Support Systems,* volume 5, (2013), pp. 740–752.

[4] Zhijie Lin. An Empirical Investigation of User and System Recommendations in Ecommerce. *Decision Support Systems,* (2014), DECSUP 12536.

[5] He Feng, Xueming Qian. Mining user-contributed photos for personalized product recommendation. *Neurocomputing,* volume 129, (2014), pp. 409-420.

[6] Li Chen, Feng Wan. Preference-Based Clustering Reviews for Augmenting E-commerce Recommendation. *Knowledge-Based Systems,* volume 50, (2013), pp. 44–59.

[7] Yiyang Zhang, Jianxin (Roger) Jiao. An Associative Classification-Based Recommendation System for Personalization in B2C E-commerce Applications. *Expert Systems with Applications*, volume 33, (2007), pp. 357–367.

[8] Keunho Choi, Donghee Yoo, Gunwoo Kim, Yongmoo Suh. A Hybrid Online-Product Recommendation System: Combining Implicit Rating-Based Collaborative Filtering and Sequential Pattern Analysis. *Electronic Commerce Research and Applications,* volume 11, (2010), pp. 309–317.

[9] Wenxing Honga, Lei Li, Tao Li. Product Recommendation with Temporal Dynamics. *Expert Systems with Applications,* volume 39, (2012), pp. 12398–12406.

[10] Jose Jesus Castro- Schez, Raul Miguel, David Vallejo, Lorenzo Manuel López. A Highly Adaptive Recommender System Based on Fuzzy Logic for B2C E-commerce Portals. *Expert Systems with Applications,* volume 38, (2011), pp. 2441–2454.

[11] Yuanchun Jiang, Jennifer Shang, Yezheng Liu, Jerrold May. Redesigning promotion strategy for e-commerce competitiveness through pricing and recommendation. *Int. J. Production Economics,* volume 167, (2015), pp.257–270.

[12] Harish Patila; Brig. Rajiv Divekar. Inventory Management Challenges For B2C E-Commerce Retailers. *Procedia Economics and Finance,* volume 11, (2014), pp. 561–571.

[13] T.Velmurugan. Performance Based Analysis Between K-Means and Fuzzy C-Means Clustering Algorithms for Connection Oriented Telecommunication Data. *Applied Soft Computing,* volume 19, (2014), pp. 134–146.

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

[15] Michael Scholz, Verena Dorner, Markus Franz, Oliver Hinz. Measuring consumers' willingness to pay with utility-based recommendation systems. *Decision Support Systems,* volume 72, (2015), pp. 60–71.

[16] Vijay Kotu, Bala Deshpande. Predictive Analytics and Data Mining. *Morgan Kaufmann, 225 Wyman Street, Waltham, MA 02451, USA,* chapter 6, (2015), pp.198-199.