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

Dihedral Product Recommendation System for E-commerce Using Data Mining Applications

Yahya Dorostkar Navaei[1], Dr. Mehdi Afzali[2*]

MSc, Department of Electrical, Computer & IT, Zanjan Branch, Islamic Azad University[1],

Associate Professor in Department of Electrical, Computer & IT, Zanjan Branch, Islamic Azad University, Zanjan, Iran[2]

y.dorostkar@gmail.com[1]

Abstract: With the development of communication networks, online access to information among much dense similar information has become a big problem. Therefore finding customers' desired products in e-commerce is more difficult now. Product Recommendation System (PRS) tries to solve this problem and reduce the overhead of communication networks by giving recommendations to customers. The purpose of this study is clustering products and create groups of products that have similar characteristics.Thus access to products with common attributes becomes easier and it prevents customers from searching in confusion or wasting time. In this article, data collected from electronic stores is clustered and grouped using C-Means algorithm. Another goal is to predict whether or not the customers purchase accessories related to the products they tend to buy. To explore the relationship between products, you must use the customer's behavior and their purchase history using association rules .These rules use data mining to discover the relationship. This relationship eventually will lead to recommendations to customers when they purchase the product The relationship between products helps to increase the accuracy of recommendations and also increases the likelihood of selling related products in electronic transactions .More detailed recommendations will lead to carefully selected customers. The results indicate that the accuracy of recommendations in the proposed PRS is more than other RPSs.

Keywords: product recommendation system, dihedral RPS, e-commerce, clustering, association rules.

1. Introduction

In recent years, rapid advances in data collection and storage technology have enabled organization to accumulate vast amount of data. Data from experimental and scientific experiments needs to store. On the other hand, data storage has related complexity and the cost. Also, extracting useful information has proven extremely challenging [1].Besides the input data gives statistical information in the first level to us, it has additional information embedded inside itself that are not easily accessible.

Today with using of data mining and applying the rules in this field, we can create models on data that reveal in this implicit knowledge and embedded information to us. Data mining applications are growing increasingly in various sciences such as Internet and Web [2]. These applications can cover many aspects of the web and extract implicit information on the web. This information can show way of use the web for users to reach their goal. Undoubtedly with access to this information we can design a flexible web interface that users enjoy from interact with this interface and they can achieve their goals ease and fast.

In this paper, we use data mining applications in e-commerce. Using this application can give information about available products at electronics to customers. They can access to desired products easily and without wasting time for search. Also with monitoring commercial transactions can mine patterns as association rules to discover the potential relationship between the products in the store and suggest these related products to customers. To achieve these goals, PRSs have been developed in e-commerce field.

PRS is a service that 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 [3].

This paper is organized in five sections. The Second section includes related work about PRS. The third section contains proposed PRS. The fourth section includes implementation of proposed PRS. The fifth section concludes of this paper and discusses about future work.

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 for e-commerce with combining similarity, trust and relationship is proposed 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.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 based on Association classification for personalization in B2C e-commerce applications is presented that based on associative classification method, for the product recommendation Issue can build an evolving system [2]. In this system first, historical requirement data are selected and transformed into proper phrase datasets. Then Data mining procedure then starts to search for a set of associated, frequently occurring phrase patterns (classifiers). In this regard, stemming algorithm and a common stop word list in English are adopted to reduce the dimensions of the text documents and improve the efficiency of the classifier extraction [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 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. 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.

Another PRS is presented [9] that introduced a prototype of e-commerce portal, called e-Zoco, of which main features are: (i) a catalogue service intended to arrange product categories hierarchically and describe them through sets of attributes, (ii) a product selection service able to deal with imprecise and vague search preferences which returns a set of results clustered in accordance with their potential relevance to the user, and (iii) a rule-based knowledge learning service to provide the users with knowledge about the existing relationships among the attributes that describe a given product category [9].

3. Proposed Dihedral PRS

Dihedral PRS proposed in this work, as the name suggests, 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 properties of products into clusters are very similar to each other. However, the products are in separate clusters have different properties.

After clustering products, an important challenge that seems is addition of new products to the system after clustering products. New products can add to electronic store in two ways:

- New type of products
- Different type of available products

First case is simple because products have clustered by type and addition of a new type of product is easy.

Second case is more complex than other one. These products after adding to the system should be in correct place between the clusters to accordance with other products in the system. It becomes a classification issue that we solve it using decision tree that it's a data mining application.

Next, the PRS tries to identify customers' requirements and criteria. In order to this work we will use an online form in electronic store that gives information about type, quality, price, brand, etc. of product that costumer wants to buy.

After gathering this information, customers are leaded to related cluster to choose products that close to their preferences.

In next step we collect information about history of shopping behavior of customers from the electronic store. This information includes items that often when a product is purchased other products to be purchased with this. In fact can be said products to be identified that participate in electronic store transactions as relate products. With achieving this information we can explore relate between products by a data mining application that called association rule mining. Also we can set buying pattern using this rules. Eventually these relations and rules will lead to creation of product buy recommendations to customers. The relationship between the products will increase the likelihood of buying the products with together.Figure1 illustrates the overall frame work of proposed PRS.

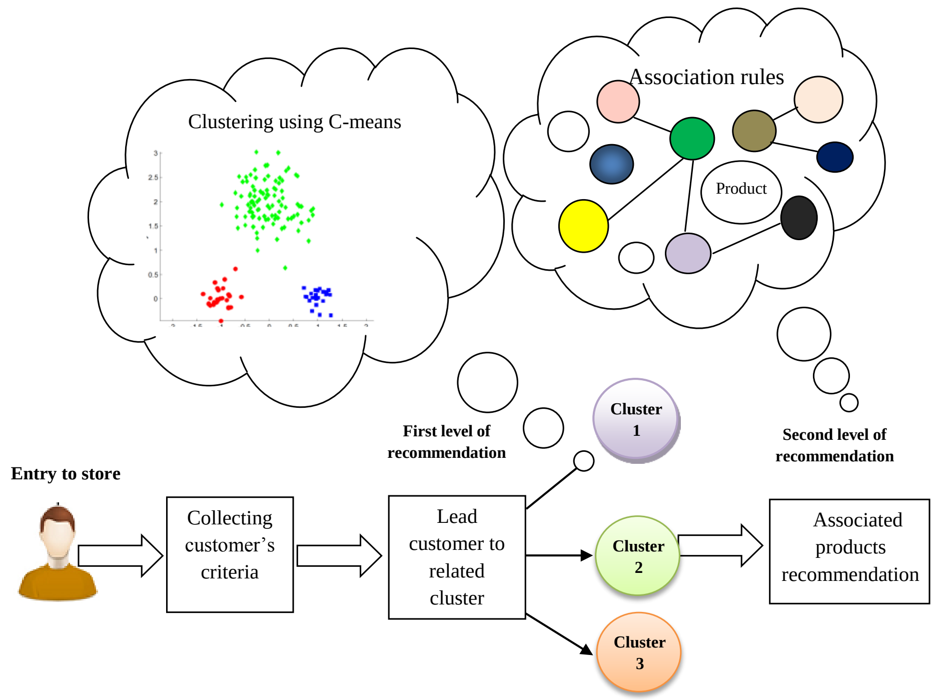

Figure 1: The Overall Framework of Proposed PRS

3.1. Clustering

Clustering means creating 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. Clustering is one of the standard workhorse techniques in the field of data mining. Its intention is to systematize a dataset into a set of groups, or clusters, which contain similar data items, as measured by some distance function. The major applications of clustering include document categorization, scientific data analysis, customer/market segmentation and www. The other areas include pattern recognition, artificial intelligence, information technology, image processing, biology, psychology, and marketing [10].

Data clustering has attracted the attention of many researchers in different disciplines. It is an important and useful technique in data analysis. A large number of clustering algorithms have been put forward and investigated. Clustering is an unsupervised learning technique. Unlike classification, in which objects are assigned to predefined classes, clustering does not have any predefined classes.The main advantage of clustering is that interesting patterns and structures can be found directly from very large data sets with little or none of the background knowledge. The cluster results are subjective and implementation dependent. The quality of a clustering method depends on the similarity measure used by the method and its implementation; its ability to discover some or all of the hidden patterns and the definition and representation of cluster chosen by the user [10].

A variety of data clustering algorithms are developed and applied for many applications domain in the field of data mining. Clustering techniques have been applied to a wide variety of research problems. So can be said, whenever one needs to classify a "mountain" of information into manageable meaningful piles, cluster analysis is of great utility [10].

In this paper we use C-means algorithm to clustering data that collected from electronic store.

3.1.1. C-means Algorithm

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. The comparative advantage of these approaches is that they do not consider sharp boundaries between the clusters, thus allowing each feature vector to belong to different clusters by a certain degree (the so-called soft clustering in contrast to hard clustering produced by conventional methods). The degree of membership of a feature vector to a cluster is usually considered as a function of its distance from the cluster centroids or from other representative vectors of the cluster. 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 pattern belongs to one and only one cluster. Fuzzy clustering extends this notion to associate each pattern with every cluster using a membership function. The output of such algorithms is a clustering, but not a partition some times. 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 widely used clustering algorithm implementing the fuzzy philosophy is FCM,initially developed by Dunn and later generalized by Bezdek, who proposed a generalization by means of a family of objective functions. Despite this algorithm proved to be less accurate than others, its fuzzy nature and the ease of implementation made it very attractive for a lot of researchers that proposed various improvements and applications. 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 two or more clusters. This method is frequently used in pattern recognition. It is based on minimization of the following objective function [10]:

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

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 [10]:

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

This iteration will stop when [10]

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

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 J_m. The algorithm is composed of the following steps [10]:

Step 1: Initialize U= [u_{ij}] matrix, $U^{(0)}$.

Step 2: At k-step: calculate the centers vectors $C^{(k)} = [c_j]$ 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 has to build an appropriate matrix named U whose factors are numbers between 0 and 1, and represent the degree of membership between data and centers of clusters. In general introducing the fuzzy logic in k-Means clustering algorithm is the FCM algorithm. FCM 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. This algorithm is similar in structure to the k-Means algorithm and also behaves in a similar way [10].

3.2. Association Rules Mining

Another interesting application of data mining is discovering hidden relationships among a large set of data. Although data stored in data sets display statistical information, these may have useful implicit information for data collectors. Exploring this information can cause improvement and promotion in preference and more utilization in variety fields such as Industry, providing services, trade, bioinformatics, medical diagnostics, web search, analysis of scientific data and etc.

This interesting application called association rules that identification is the discovery of item sets which occur mutually in a given data set. The rules are based on the frequency number of an itemset, which occurs alone, or in combination with other sets in a database. Association rule is mostly used to identify 'interesting' hidden relationships among attributes of huge database. Generally, a standard association rule is expressed in X →Y form, where X is the antecedent and Y is the consequent, which signifies that X will occur with Y for the same instance in a database with a minimum level of significance. Note that each rule can have multiple items,i.e., a set of items, as antecedent and consequent [11].Actually association rules mining finds repeated items in a set of transactions as frequent patterns.

A number of association rule mining algorithms and techniques have been developed in the last few years. In this paper we use Apriori algorithm to discovering relationship between products available in electronic store.

3.2.1. Apriori Algorithm

Apriori algorithm has been used to find out strong association rules among itemsets of the incident data collected. The apriori algorithm to find out frequent itemsets and generation of association rule from frequent item set. The main two measurements of rule effectiveness are support and confidence, which reflect the usefulness and certainty of discovered rules, respectively. Support (S) is the measurement of the proportion occurrence of any itemset or combination of itemsets (e.g., X and Y) in a database (D) [11]. Support is expressed as follows [1]:

$$\text{Support, } s(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{N}\ ; \qquad (4)$$

Confidence (C) is defined as the conditional probability (P) of occurrence of the consequent of an itemset given that the antecedent of that itemset has occurred [11]. Confidence is expressed as follows [1]:

$$\text{Confidence, } c(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)}\ ; \qquad (5)$$

The main idea of the Apriori algorithm theory as follows:

Theorem1 (Apriori Principle). If an itemset is frequent, then all of its subsets must also be frequent [1].

Conversely, if an itemset is infrequent, then all of its supersets must be infrequent too and the entire subgraph containing can be pruned immediately. This strategy of trimming the exponential search space based on the support measure is known as support-base pruning. Such a pruning strategy is made possible by a key property of the support measure, namely, that the support for an itemset never exceeds the support for its subset [1].

Apriori is the first association rule mining algorithm that pioneered the use of support-based pruning to systematically control the exponential growth of candidate itemsets. So initially, every item is considered as a candidate 1-itemset. After counting their supports, the itemsets that have support less than determined threshold is discarded. In the next iteration, candidate 2-itemsets are generated using only the frequent 1-itemset because the Apriori principle ensures that all supersets of the frequent 1-itemsets must be infrequent. Remaining itemsets will be used to generate candidate3-itemsets. More level of candidate itemsets will be generating in same way [1].

The apriori algorithm terminates when there are no new frequent itemset .This algorithm uses as a level-wise approach for generating association rules, where each level corresponds to the number of items that belong to the rule consequent. Initially, all the high confidence rules that have only one item in the rule consequent are extracted [1].

4. Implementation of Proposed PRS

Implementation of this PRS includes three stages that described in follow.

4.1. First Level of Product Recommendation

Proposed dihedral PRS provides two level of product recommendation to users 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. For implementation this level we cluster a type of the collected data from electronic store using C-means algorithm based on numerical adjectives of products such as price and quality ratings. Considering that the quality rating is announced by company that creates the product, so reliable is. Therefore 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 that are available in Tebian electronic store. This table includes six field, *namely product ID, product name, Price (Rials), quality range, price range (Rials/M), Type*. For the price and quality ratings are in a same range, we divide value of price to 1000000 Rials and achieve price range (Rials/M) field.

In first step of C-means algorithm, scattering data of cell phone is showed in Figure 2. After this step, C-means algorithm is applied on data. In this paper, the number of clusters is considered three, 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 3.

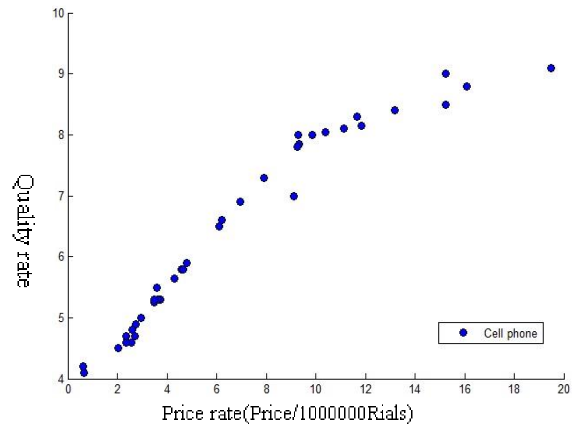

Figure 2: Scattering Data of Cell phone

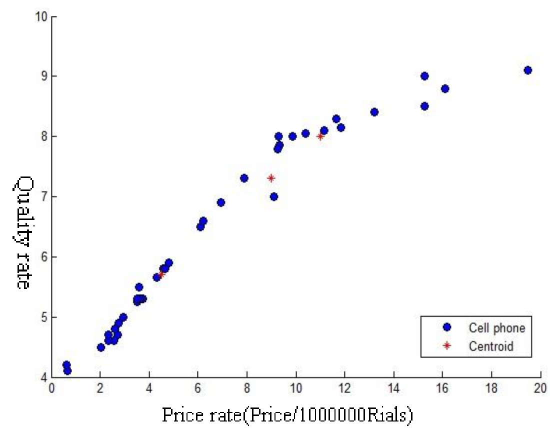

Figure 3: Selection Centroid Points Randomly

Table 1: A Part of Data about Cellphone

product ID	product name	Price (Rials)	quality range	price range (Rials/M)	Type
78277	Huawei Ascend Y210D	2700000	4.7	2.7	cell phone
78278	Huawei Ascend Y210	2550000	4.6	2.55	cell phone
78279	Huawei Ascend Y300 D	3750000	5.3	3.75	cell phone
78282	Huawei Ascend G510	4600000	5.8	4.6	cell phone
78286	Huawei Ascend Mate	10400000	8.05	10.4	cell phone
78290	Huawei Ascend P6	9850000	8	9.85	cell phone
79025	G510	4650000	5.8	4.65	cell phone
79029	Huawei Ascend G610	6100000	6.5	6.1	cell phone
79129	Huawei Ascend G700	7900000	7.3	7.9	cell phone
79204	Huawei Ascend Y220	2350000	4.6	2.35	cell phone
76353	C3312 Duos	2020000	4.5	2.02	cell phone
76355	Galaxy Ace S5830	4300000	5.65	4.3	cell phone
76357	Galaxy Mini 2 S6500	3590000	5.5	3.59	cell phone
76358	Galaxy Note II N7100 - 16GB	13200000	8.4	13.2	cell phone
76361	Galaxy Wonder I8150	6200000	6.6	6.2	cell phone
76362	Galaxy Y S5360	2750000	4.9	2.75	cell phone
76367	Galaxy Ace S7500	4800000	5.9	4.8	cell phone
76369	Galaxy Mini S5570	3500000	5.25	3.5	cell phone
76382	Galaxy S Advance I9070-8GB	6950000	6.9	6.95	cell phone
76383	Galaxy S III I9300	9300000	8	9.3	cell phone
76384	Galaxy Y Duos S6102	3630000	5.3	3.63	cell phone
76386	S3850 Corby II	2950000	5	2.95	cell phone
76388	Samsung Galaxy Star S5282	2350000	4.7	2.35	cell phone
76393	Samsung Galaxy Young S6312	3500000	5.3	3.5	cell phone
76395	Samsung Galaxy Note N7000 - 16GB	15250000	9	15.25	cell phone
76460	Samsung E1200M	650000	4.1	0.65	cell phone
76463	Samsung Galaxy Pocket S5300	2590000	4.8	2.59	cell phone
76468	Samsung I9192 Galaxy S4 Mini Dual Sim	9100000	7	9.1	cell phone
76469	Samsung E1200M	630000	4.2	0.63	cell phone

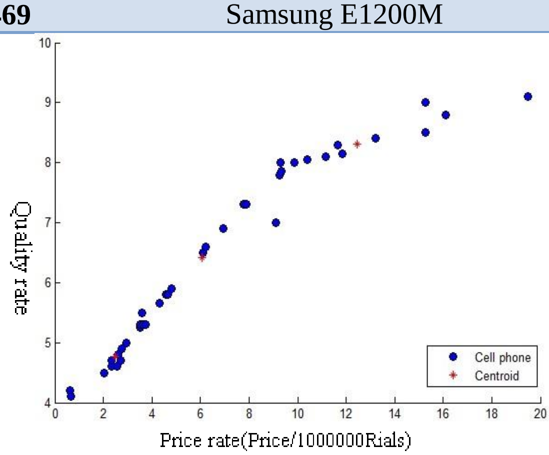

Figure4: Correct Place of Each Centroid

Table2: Coordinates of Each Centroid Points

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

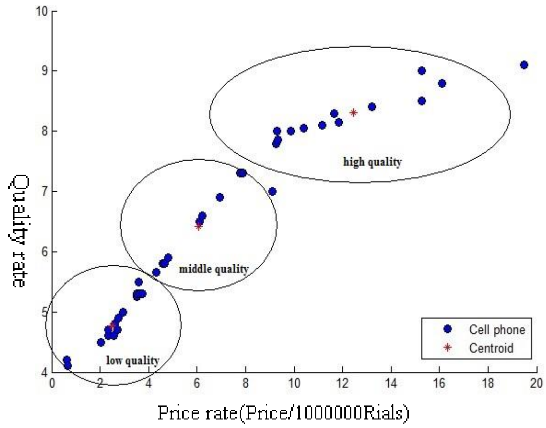

Figure5: Clustering for Data Related to Cellphone

C-means algorithm changes place of centroid points according to mean of data. Algorithm repeats this replacement until there is no move 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. Coordinates of each centroid points are presented in Table 2 that c_1, c_2, c_3 are means of *Low quality*, *Middle quality* and *High quality* clusters, respectively.

Point's assignment to clusters is based on u_{ij} membership degree that mentioned in (2).Each point with high membership is assigned to cluster with nearest centroid. Finally clustering for data related to cell phone is shown in Figure5.

4.2. PRS Development

Due to the increasing variety of products in the online store development system should also be considered. In this case, the development of this system can be expressed as adding products to the system. In fact, the system clustering must be able to respond to increasing of product variety. That means is the products that add after clustering to online store should be placed in related cluster correctly else either product will remain as no-clustered or clustering will be false. And provided recommendations will be incorrect.

For this purpose we discussed earlier, the products can be added to the system in two ways. 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 these membership values is higher; the product is allocated to this cluster. To do this we will calculate the following relations [10]:

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

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

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

(6)

Where A, B, C are membership value of c_1, c_2, c_3, respectively. Top relations are the modified form of equation (2) [10]. Table .3displays a small set of data test for cellphone that should specify their cluster label. In fact, we want to classify this data test using the decision tree shown in Figure6.

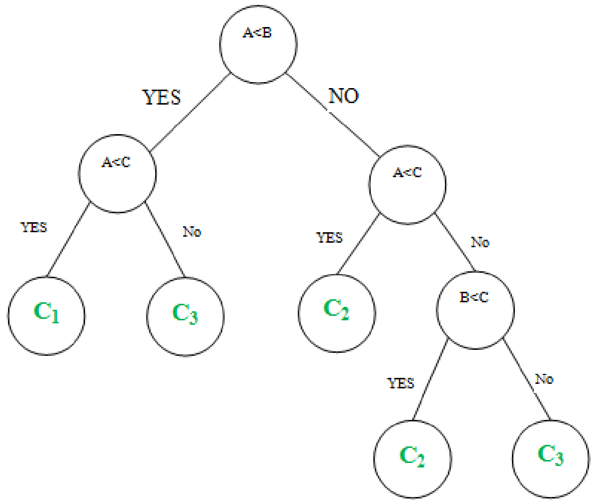

Figure 5: Decision Tree for PRS development

4.3. Colleting Customer's Criteria

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

In order to that 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.

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

Customers with filling this form give correct information about desired product and PRS leads them to related cluster with this product.

Table 3: Data Test Points for Cellphone

product ID	product name	Price (Rials)	quality rate	price rate	Type	Cluster
78213	Xperia P	7850000	7.3	7.85	cell phone	Middle
78214	Xperia SL	9320000	7.85	9.32	cell phone	Middle
78426	Xperia P	7830000	7.3	7.83	cell phone	Middle
78427	Xperia Ion	9250000	7.8	9.25	cell phone	Middle
78428	Xperia S	9250000	7.8	9.25	cell phone	Middle
78429	Sony Xperia M	5550000	6.25	5.55	cell phone	Low
78443	Sony Xperia Z Ultra	19500000	9	19.5	cell phone	High
78444	Sony Xperia L	7450000	7.25	7.45	cell phone	Middle
78445	Sony Xperia M	6950000	6.9	6.95	cell phone	Low
78885	Sony Xperia SP	10100000	8.4	10.1	cell phone	High
78886	Sony Xperia ZR	13000000	8.7	13	cell phone	High
78887	Sony Xperia Z1	21000000	9.5	21	cell phone	High

4.4. Second Level of product Recommendation

In this section we discover rules and relations between clustered products. In order to, we collect all transactions in electronic store during two months. These transactions are probed and Table .4 gained as summary of all transactions table.

Table 4: Transactions Table

number	Transaction
1	Laptop - Modem - Software - Router
2	Bluetooth Hands free - Tablet - Flash
3	Software - Book - Flash
4	Cell Phone - Memory Card - Software
5	Bilateral pan - Ground cleaner - Digital scale
6	Cell Phones - Hands free - Memory card - Flash
7	Tablet - Software - Book
8	Sofa - Dining table - Coat hanger- Table footstool
9	Lap top - Modem - Router - Mouse
10	Tablet - Software - Memory Card - Flash
11	Lap top - Digital Receiver - Lap top bag
12	Cellphone - Chargers - Gel frame
13	Coat hanger- Table footstool
14	Tablet - Software - Book
15	Laptop - Modem - Digital Receiver
16	Software - Book - Flash
17	Sofa - Dining table - Bilateral pan - Coat hanger
18	Lap top - Modem - Mouse - Flash
19	Cell Phone - Hands free - Gel frame
20	Ice Cream Maker - Juicer - Crusher
21	Lap top - Mouse - Lap top bag
22	Knife - Juicer- Saucepan
23	Router - Mouse - Modem
24	Sofa - Table footstool
25	Software - Book - Memory Card

For mining association rules and the relationship between products Apriori algorithm is applied on these transactions. As explained in the previous section, the candidate i- itemset is calculated for i = 1,2,3,… And then according to the algorithm, rules with high confidence selected as candidate association rules. Considering that in this study the number of transactions and variety of items are high, itemset thresholds can be set down. Candidate 1-itemsets are generated from transactions table and items that have confidence less than threshold are eliminated. Also candidate 2-itemsets are generated from candidate 1-itemsets and candidate 3-itemsets are generated from candidate 2-itemsets by pruning according threshold. Finally, candidate rules are extracted from candidate 3-itemsets with high confidence and there for association rules between products are mined from these candidate rules. Table 5 shows a summary of all candidate rules that extracted from candidate itemsets.

Table 5: Extracted Candidate Rules

Number	Rules
1	{ Laptop}↔{Modem - Router}
2	{Laptop}↔{Mouse - Laptop bag}
3	{Cellphone}↔{Memory Card- Flash}
4	{Cellphone }↔{Chargers - Gel frame}
5	{Tablet }↔{Software - Flash}
6	{Software }↔{Book - Flash}
7	{Sofa}↔{ Coat hanger- Table footstool }
8	{ Laptop}↔{Modem - Router}
9	{ Laptop}↔{Modem - Router}

As shown, using Apriori algorithm, frequent candidate itemsets are extracted from transactions that occur during two month in electronic store and by these itemsets achieved associated rules with high confidence. Now using these rules we discover the relations between products and create second level of product recommendation. Figure 6 shows relations between products. When customer selects one of the products, other products that are linked with this suggested to customer as second level of product recommendation.

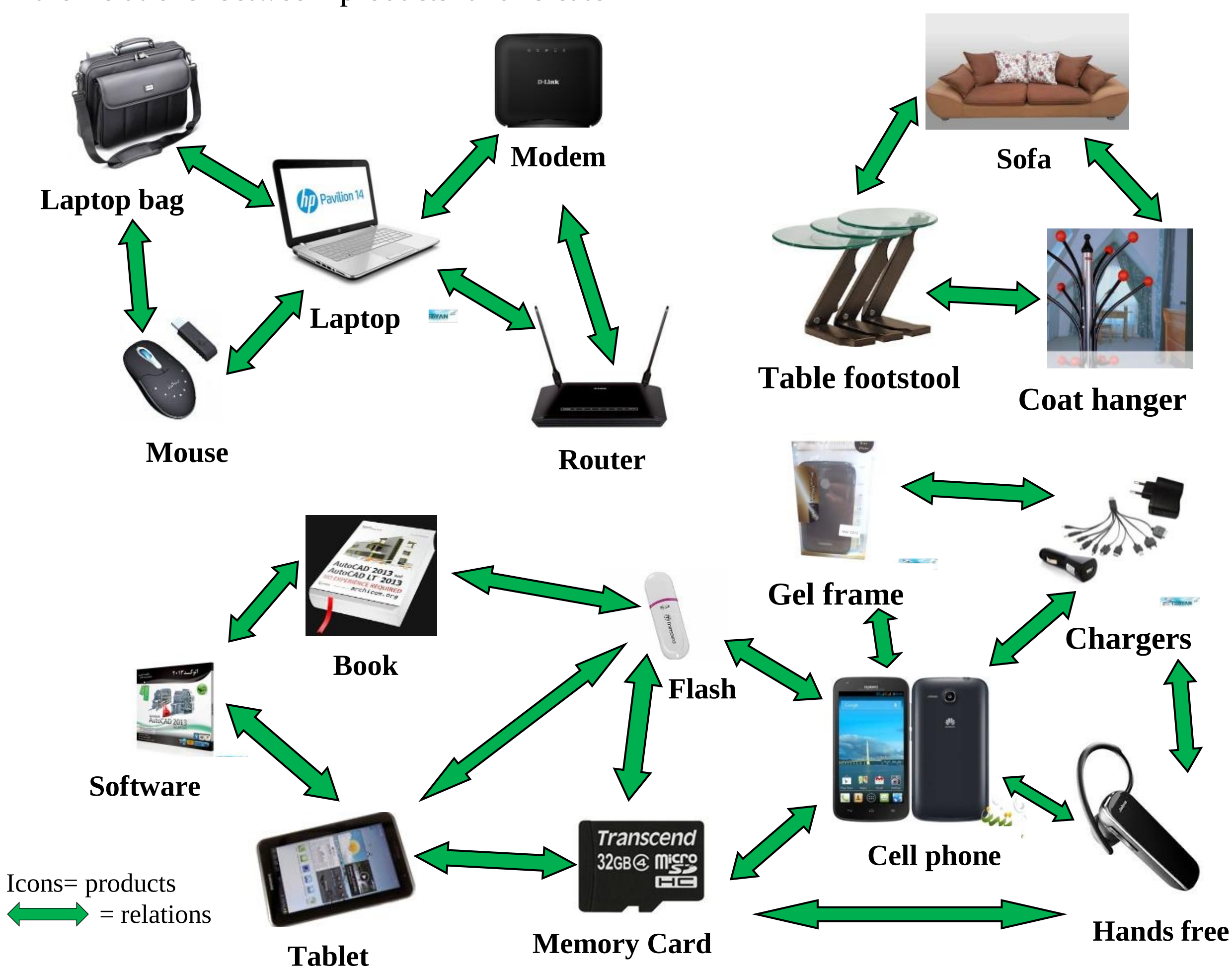

Figure 6: Relations between Products

Where in this figure each product is shown with an icon and double-sided arrows indicate a close relationship between these products. Related product Suggestion causes prevent of wasting time in product finding. Also increasing accuracy of this recommendation will raise the sales and profitability of the electronic store and will increase customer satisfaction rate.

5. Performance Evaluation

As mentioned previous section, with increasing levels of customer satisfaction from received recommendations, store sales will increase that it shows high accuracy of proposed PRS. In order to evaluate the performance of the dihedral PRS we use these two criteria.

5.1. Feedback

The first way is accessible by collect feedback on the performance of the PRS. In fact, this is collection of comments from customers about the system and the effectiveness of the proposed recommendations. Customer response to this survey specifies the performance of the PRS.

Feedback is a method of gathering opinions in online systems, to evaluate the performance and utilization improvements or make changes in the system. Online surveys are one of the most common methods to providing feedback. Survey forms are embedded in most websites and web pages causes this method has become one of the most popular methods across web. Since this study also takes steps to provide online product recommendations to customers, we use online survey form to get feedback. This form contains five options that are *very weak*, *weak*, *middle*, *good* and *excellent* that expresses Customer's feelings about the system. Two first options represent a customer dissatisfaction sense of the system. Third option represents a neutral sense of customer and two last options show the user's sense of satisfaction.

Figure7 shows the opinion graph of 200 customers has purchased from online store. As shown in Figure .7 approximate 75 percentages of customers that received recommendations have sense of satisfaction. This amount for people who have felt neutral is about 15%.Also, according to the Figure we can realize that only 10% of customers are dissatisfied

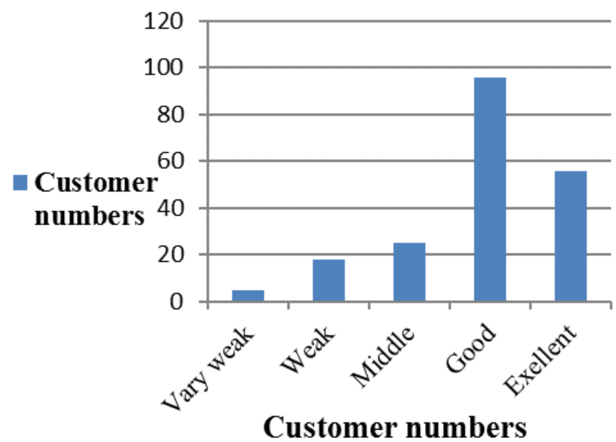

Figure7: Opinion Graphs of 200 Customers

Finally, we find that the PRS relying on customer reviews have good performance in clustering products and discover the relations between products available in stores.

5.2. Increasing Sales of Products

With increasing levels of customer satisfaction will increase the use of this system to buy from stores and this will increase the sales of products. In fact, the accuracy of the PRS should be to increase sales in online store. For this purpose, data on sales during one month before and one month after the implementation of the PRs to evaluate the performance and amount of selling products in the store is collected. Figure 8 compares the selling products for a month before and after the implementation of the PRS.

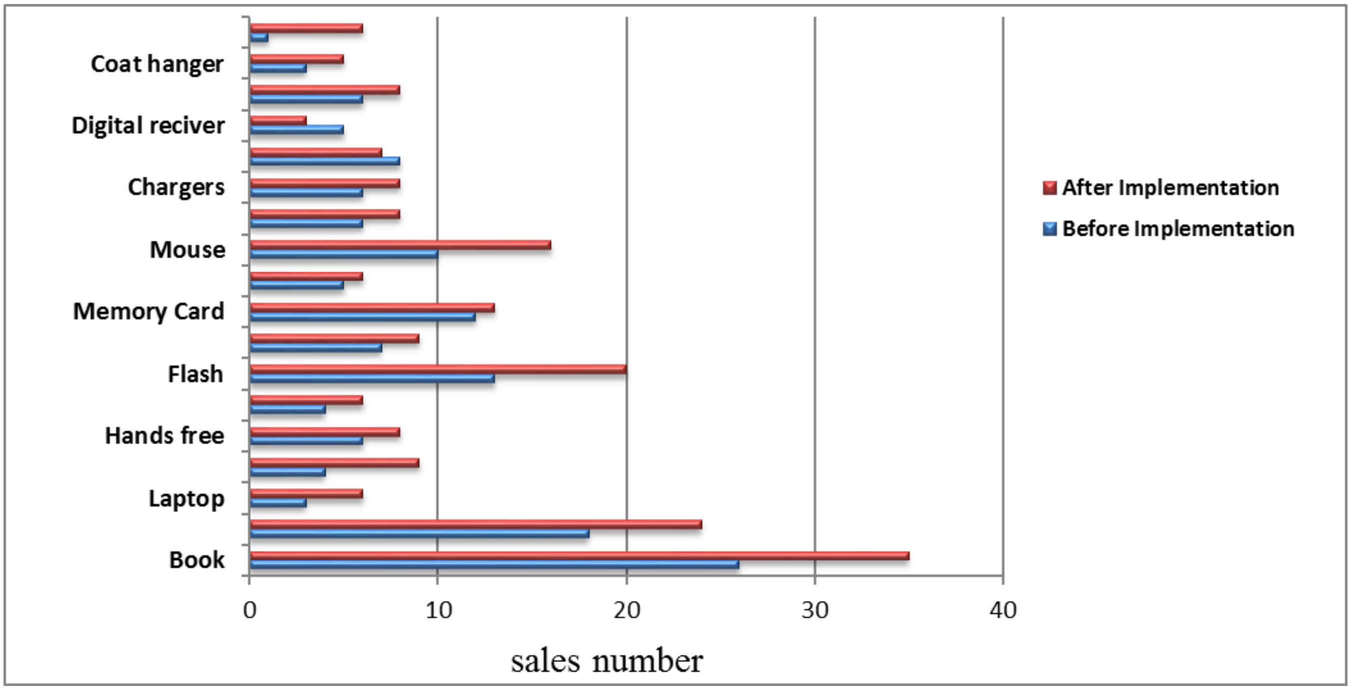

Figure 8: Compares Selling Products for a Month Before and After the Implementation of the PRS

According to Figure.8 can be seen that 76% of the products after the implementation of the PRS have sales growth. On closer inspection you can see that the product recommendations were accurate therefore the amount of sales is increased by implementing proposed PRS in electronic store.

6. Conclusion

In this paper, a new method has been introduced called the dihedral PRS in e-commerce using data mining applications. In this PRS, products recommendation is offered 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 is presented an approach through the decision tree classification method.

The second level of product recommendation is after product selection where potential relationships between products are discovered using association rules mining. At this level, then customers select the product, related products are advised to them.

Finally to evaluate the performance used of the two criteria namely customers feel and increase sales .The results show that 75% of customers, who received buy recommendations from the proposed dihedral PRS in this study, feel satisfied. Also the comparison of sales before and after the implementation of proposed PRS shows 76% of the products are available in store after the implementation of PRS, sales have increased. These results indicate that the accuracy of recommendations for the proposed PRS is high.

For future work, can be used sequential patterns to increase the accuracy of the relationship between products and providing product recommendations. This approach attends to the time ordering of transactions.

References

[1] Tan Pang-Ning, Steinbach Michael, Kumar Vipin, (2006), "Introduction to Data Mining", Pearson Addison-Wesley, chapter1, 8.

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

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

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

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

[6] Wenxing Honga, Lei Li, Tao Li,(2012), "Product Recommendation with Temporal Dynamics", Elsevier Expert Systems with Applications, vol 39,pp. 12398–12406.

[7] He Feng, Xueming Qian,(2014), "Mining user-contributed photos for personalized product recommendation", Elsevier Neurocomputing, vol 129, pp. 409-420

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

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

[10] T.Velmurugan,(2014), "Performance Based Analysis Between K-Means and Fuzzy C-Means Clustering Algorithms for Connection Oriented Telecommunication Data", Elsevier Applied Soft Computing , vol19, pp. 134–146.

[11] A.Verma, S.Das Khan, J.Maiti, O.B. Krishna, (2014), "Identifying Patterns of Safety Related Incidents in a Steel Plant Using Association Rule Mining of Incident Investigation Reports", Elsevier Safety Science, vol70, pp. 89–98.

Authors Profile

Yahya Dorostkar Navaei is an MSc in department of electrical computer & IT in Zanjan Branch Islamic Azad University. His present research interests are Recommendation systems and Data mining.

Mehdi Afzali is an Associate Professor and Chief of Department of Electrical Computer & IT in Zanjan Branch Islamic Azad University. His present research interests are management of Information Technology.

