

# Improving Recommender Systems in E-Commerce Using Similar Goods

Majid Khalaji<sup>\*</sup>, Keramat Mansouri<sup>2</sup>, S. Javad Mirabedini<sup>3</sup>

<sup>1</sup>Department of Computer Engineering, Islamic Azad University, Arak Branch, Arak, Iran; <sup>2</sup>Department of Computer Engineering, Amirkabir Polytechnic University, Tehran, Iran; <sup>3</sup>Department of Computer Engineering, Islamic Azad University, Central Tehran Branch, Tehran, Iran.

Email: <sup>\*</sup>m\_khalaji@yahoo.com

Received November 2<sup>nd</sup>, 2011; revised December 10<sup>th</sup>, 2011; accepted December 30<sup>th</sup>, 2011

## ABSTRACT

Due to developments of information technology, most of companies and E-shops are looking for selling their products by the Web. These companies increasingly try to sell products and promote their selling strategies by personalization. In this paper, we try to design a Recommender System using association of complementary and similarity among goods and commodities and offer the best goods based on personal needs and interests. We will use ontology that can calculate the degree of complementary, the set of complementary products and the similarity, and then offer them to users. In this paper, we identify two algorithms, CSPAPT and CSPOPT. They have offered better results in comparison with the algorithm of rules; also they don't have cold start and scalable problems in Recommender Systems.

**Keywords:** Recommender System; Ontology; Similarity; Complementary; Association Rule; Collaborative Filtering

## 1. Introduction

Today, the world wide web is a great place of digital documents considering the developments of information technology. With increasing information references, numbers of choices have become more and more to find required goods. Also, it is difficult to maintain this information and this problem is related to the developments of information technology and is known as the expensive overload. Personalization is one of the solutions. Personalized systems benefit from a Recommender System and try to settle dynamic pages based on personal interests. Recommender Systems are important research references. The first related paper was presented in 1990 and was about collaborative filtering [1]. In recent decades, researches have been done that led to the developments of new methods for Recommender Systems. There are examples of applications such as products of Amazon site and Video sites. Recommender Systems have a significant effect on the success of websites the uncontrolled growth of electronic market as well as web-based systems. In information technology, the success of a website depends on customers and visitors attractions and definition of user's needs is necessary for improving applications of a website. Recommender Systems try to predict interests and needs of users using data collections and offer a list of user's needs. These Recommender Systems have many

problems and led to some problems in great websites about offering goods and commodities to users. The relationships of goods have important function in designing Recommender Systems. The most important relationships can be named complementary association and similarity.

On the other hand, ontology is used to determine the relationships among contents. Considering the structure of ontology, we can use it for determining the relationship between complementary association, similarity and other associations. Many problems of algorithms can be solved by the association between ontology and commodity.

The purpose of this paper is to design a Recommender System that can offer personal needs and interests. The offers are not the only response to the user's needs, but also they pay so attention to user's interests and tastes. The designed systems should have a good speed and accuracy that lead to customer's satisfaction. Also, the other purpose of this paper is to solve some problems, such as cold start and scalability. These problems are involved in many Recommender Systems.

The most important achievements of this paper can be outlined as follows:

- Implementation and evaluation of the Recommender Systems using ontology and similarity association among goods.
- Offering goods to users using Recommender Systems in websites and selling through E-shops and user's satisfaction.

<sup>\*</sup>Corresponding author.

- Presentation of a new model using data structures to enhance the offers of personalized systems, as well as implementation and evaluation of offering method in this model.

In Section 2, we identify Recommender Systems. In Section 3, we explain complementary procedure and define subsections and the features of complementary and the way of calculation. In Section 4, we design the Recommender System based on complementary goods. In Section 5, we evaluate the designed Recommender System and finally in Section 6, we get a conclusion.

## 2. Recommender System

Web producers raise personalized web for better communication between customers and producers. Although web personalization is not an essential part of business, it has many applications in market. The purpose of personalization system is considering the user's needs in the web. After personalization, Recommender Systems were presented according to the user's needs and tastes. Recommender Systems are divided into 3 categories based on the structure:

- Rule-based Systems: In rule-based systems, decisions were made based on the rules extracted automatically or manually through user's information. The purpose of this system is discovering elements that affect user's preferences in choosing a product. ISCREEN is one of the rule-based systems. It's used for filtering text messages by manually produced rules of users.
- Content-based Filtering: This algorithm presents the offers based on the items that a user has sold before. Using the methods like Machine Learning and buying history, it will be possible to identify the attractive and non-attractive goods for the customers [2,3]. In these systems, items were offered based on content associations and user's interests [4].
- Collaborative Filtering (CF): These systems offer the products based on behavioral similarities and user's application patterns. They perform statistical analyses by using data mining in databases, monitoring user's behaviors, rating goods and buying history.

Collaborative filtering is classified in two categories-based on users, items, and how it is planned [5,6].

In the first classification, they find a user's system that would be similar to the target system, then offer the goods to target users that is more attractive to them. In the second classification, the system finds the items that would be similar to the higher cost items and then offers them to user and recognizes the similarities of items based on given rates, instead of finding the neighbor target users.

Recommender Systems have been designed based on concept associations. They pay too much attention to user's interests. But there is another considered factor

that is user's needs. For this reason, the similarity cannot recognize user's needs. Therefore, another association which is required complementary association. Complementary has many applications in economical sciences [7]. The complementary association of products means that the products are costumed altogether like hamburgers and hamburger buns. In many application programs of Electronic Market, these kinds of associations are more beneficial than the other associations. For example, offering a similar machine or bicycle to someone who buys a Mercedes Benz is not too beneficial. But offering a CD-player or an alarm system is more beneficial. After investigating these associations, Recommender Systems are designed using suitable associations correctly. In a designed system, firstly, complementary products are to users, then for increasing accuracy of a system and satisfaction of user, they should be added by user's taste about suggestions. It's required to similarity. At last, Recommender Systems are designed for offering complementary products based on user's tastes. Both complementary and similarity are used in a system that is going to be designed.

Complementary can be calculated based on the association of products and helping a new property called need. In fact, we use ontology.

The purpose of this paper is to extract user's needs through product associations and relationships. Product similarity is beneficial for determining user's interests, but it cannot determine the user's needs. Therefore, to a relationship, it is required that is called Product Complementary in economy.

Economically, complementary product is a needed product for using better the main product. For example, product *A* is complement of the product *B*, if it's required to *A* for using the better product *B*. The examples of complementary products are hamburgers and hamburger buns.

In this paper, semantic procedure and semantic complementary are used. In this procedure, products are modeled by OWL language and complementary products are extracted through ontology, and then Recommender System is planned based on them. Similarity illustrates the similarity between two products considering the associations of products. But complementary is another association that is different from similarity completely. For example, consider the association between car and gasoline. In other words, consider the relationship between a car and a bicycle. Car and bicycle have more similarities, but gasoline is more related to car (gasoline is complement of car). It means that using a car requires gasoline [8].

## 3. Complementary Procedure

### 3.1. Product Complementary Property

Complementary is not a bilateral relation. It means that if *A* is complement of *B*, we cannot say that *B* is comple-

ment of  $A$ . For example, we need a video-player for using a video game; but we don't need a video game for using a video player. So, we can get the conclusion that the video game is complement the video player.

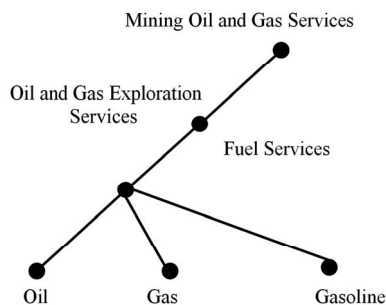
Complementary has a transitive relation. It means that if  $A$  is complement of  $B$  and  $B$  is complement of  $C$ , then  $A$  is complement of  $C$ .

### 3.2. Determining Complementary Degree

First, products should be divided based on standard UNSPSC. The standard indicates Is-a property in all items. In this step, the classes and the associations are determined. Every class indicates that it comes from ontology tags based on OWL language [9]. A part of catalogue is related to fuel service that's planned based on standard UNSPSC. See **Figure 1**.

Classes have a property called need. It has an important function in calculating the complementary algorithm degree. This property would be OWL in ontology's language using property definitions. For example, a user needs a paper and an eraser for a pencil usage, so the need property includes an eraser and a paper.  $Com(A, B)$  illustrates complementary degree of  $A$  and  $B$ , that is between 0 and 1. The more closer to 1, the more need to  $B$ . For determining  $Com(A, B)$ , we require product associations and need properties. If  $B$  is one of the needs of  $A$ , so  $Com(A, B)$  is 1; if not we use coefficient  $k$  and the algorithm with  $B$ . This algorithm continues with the high class  $B$ . If high class  $B$  is a member of the need product, so  $B$  is complement of  $A$  using  $k$ . If high class  $B$  is not a member of the needed  $A$  and high class  $B$  is not the root, this procedure is repeated with higher class  $B$ . The more distance between  $B$  and needed  $A$  in catalogue, the less amount of  $Com(A, B)$  and if  $B$  and higher classes are not members of the needed  $A$ . So, the amount of  $Com(A, B)$  is equal to 0. The Equation (1) indicates that:

$$com(A, B) = \begin{cases} \frac{1}{k} * com(A, B \cdot H) & B \notin A \cdot N \\ 1 & B \in A \cdot N \\ 0 & B \cdot H = \Phi \end{cases} \quad (1)$$



**Figure 1. Fuel service catalog.**

In the above formula,  $H$  is the high class product.  $B \cdot H$  is high class  $B$ .  $N$  is needed property, that  $A \cdot N$  indicates needed  $A$ .  $k$  is variable coefficient that shows the increasing speed rate of complementary and equals 2 [10].

**Example 1.** Complementary Degree of Hamburger and Hamburger Buns.

First, draw a classification figure, then determine the products that are the amount of needed hamburger and hamburger buns.

Burger,  $N = \{\text{Burger buns, Vegetable, Sause}\}$

$Com(\text{Burger, Burger buns}) = 1$  Burgerbuns  $\in$  Burger  $\cdot N$

## 4. To Design a Recommender System

Recommender Systems are divided into 3 groups based on content filtering, correlative filtering and combined filtering. The designed system of this paper is not offered for increasing the algorithm accuracy like previous similar products, but it is offered for combined complementary and similarity of user files. Also, it is used to conceptualize method to determine complementary products for solving some limitations and problems such as the first user, cold start and scalable. The whole purpose of this paper is to design a Recommender System for solving many problems of systems with the suitable accuracy and speed.

In these systems, users are divided into 2 classes. The first class is the users who have purchased the products before, and the second class is the users who didn't ever purchased the products from the market.

The system offers two classes of complementary and similarity products. To users' first, while a user enters a shop, the system offers some suitable products based on the user's buying background. But when the second group enters the shop, the system cannot offer any product. Because, There is not any background. However along with a shopping, the system determines user's needed properties and offers suitable products. For this reason, complementary products are divided into two categories:

- **CSPAPT**  
Complementary and Similarity Product after Purchase Time.
- **CSPOPT**  
Complementary and Similarity Product over Purchase Time.

### 4.1. CSPAPT Algorithm

When a user enters a shop for the next time, the system offers the products based on CSPAPT algorithm. This algorithm checks previous purchases at the exact time that a user enters. Then it offers complementary and similarity products to previous products, and it clearly determines colors, brands and sizes of the products. For instance, it offers a complementary eraser to a user buy-

ing a pencil. Considering the user who buys the pencil with identified color, size and brands, the system offers an eraser complement and similar with the pencil properties.

In this algorithm, the needed products (complementary products) are determined, and then it offers the products which have the most similarity with the user's interests and tastes. For example, the favorite color of a customer is blue and usually he/she buys office supplies with the brand X, the middle size and the quality Y. Therefore, the system offers to the customer who buys a pencil, a blue, brand X, quality Y and middle size eraser or it offers an eraser that most of the customers buy.

#### 4.2. CSPOPT Algorithm

When a user is buying a product, CSPOPT algorithm offers the complementary and similarity products. When a user enters a shop, chooses some products and adds them to a basket, the system checks the basket and offers complementary and similarity products. For example, when a user adds a hamburger to a basket, the system checks the basket and offers the complementary hamburger (hamburger buns). If the hamburger is of the best quality, the system offers the best hamburger buns.

The Recommender System works as a salesperson and tries to introduce the complementary and similarity products with purchased products to customers for more satisfaction and selling.

### 5. Evaluation

In this paper, a dataset of a supermarket is used. The data has been collected from January 2008 to August 2010, based on invoices of the company that exist as an access database. The data consists of 2154 customers, 3125 products invoices for simplifying to design and execute. The products are divided into 20 main classes, according to classes including milk, yoghurt, cheese, butter, bread, hamburger, rise, sauce, ice cream, chocolate, and so on.

It requires ontology for calculating complementary degree of both products. Product properties are in a table. The table is as a product's ontology. The main column of this table is as following:

- Id: Code of every product based on UNSPSC standard
- Name: Name of every product
- Is-a: High class of every product
- Need: Needed property of a product

Two popular measurement criteria used in Recommender Systems, called "precision" and "recall", have been invoked in this paper to evaluate the designed system. Finally, the standard criterion of  $F_1$  metric has been obtained by the combination of these two criteria [11].

First, the purchased products are divided into two groups. The first group is Train Set and the other group is Hit Set.

These sets are chosen randomly. First, the given algorithm is executed on the train set and is called Top- $N$ . Then, these products are compared with the test set and shared with both test and Top- $N$  called Hit Set. Finally, an accurate percent of algorithm is determined after giving the test set, train set and hit set using evaluation metric. The Equation (2) indicates that:

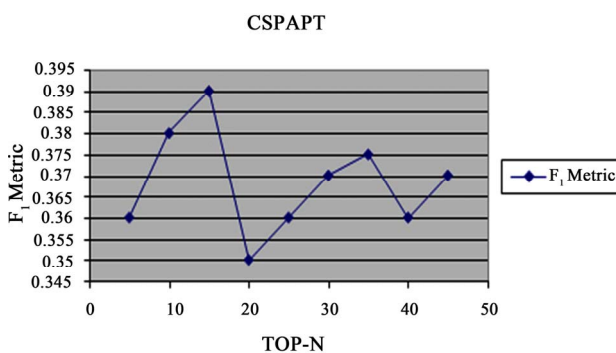
$$F_1 = 2 * \frac{\text{Size of Hit Set}}{\text{Size of Test Set} + \text{Size of Top-}N \text{ Set}} \quad (2)$$

we consider the mean of given  $F_1$  from all users as an accurate algorithm.

The first step of the system evaluation is to calculate complementary degree of products. In the second step, test set and train set are determined. As this sequence, 80% products exist in train set and the rest 20% products exist in test set randomly. In the third step, Top- $N$  set is produced. The algorithm is executed on the train set and  $N$  products are produced for offering to users, the  $N$  products are the highest complementary degree with train set. In the last step,  $F_1$  is calculated for all users after producing train set, test set and hit set, and the mean of  $F_1$  is considered as the final  $F_1$ . We have done the algorithm with different sets, test and train sets 10 times and the result shows the algorithms CSPOPT and CSPAPT.

For evaluating the algorithm CSPAPT, the invoices of products are divided into train and test sets and complementary product sets are classified based on color, size and quality. Then, identified complementary  $N$  product and Top- $N$  set are made. The amount of  $F_1$  are calculated by test, Top- $N$  and Hit sets. The final results have been shown in **Figure 2**. In this diagram, the amount of  $F_1$  varies between 0/35 to 0/39.

For evaluating algorithm CSPOPT, the purchased products are classified based on the purchased date. In these algorithms, the purchased products are divided into train and test sets during a day. A complementary product set is classified based on color, size, quality and  $F_1$  is calculated. The final amounts of  $F_1$  are shown in **Figure 3** and vary between 0/42 and 0/46.



**Figure 2. Diagram  $F_1$  based on variable amount of  $N$  in algorithm CSPAPT.**

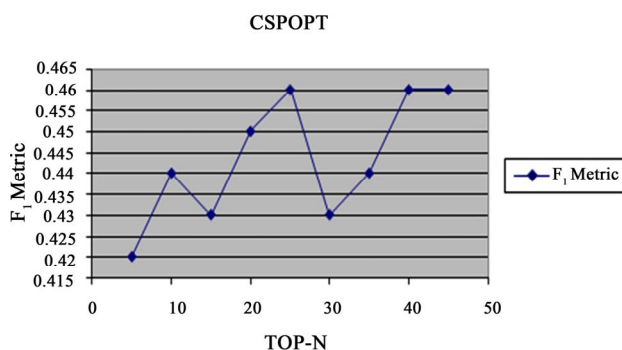


Figure 3. Diagram  $F_1$  based on variable amount of  $N$  in algorithm CSPOPT.

Association Rules are known as algorithms that discover and find the rules and the relations among a big set of data [12]. Discovering and finding the relations, rules of records and data in big databases of companies can be affected better and attract manager's decisions in different field such as catalogue designs, more and effective selling and various business and marketing policy. According to association rule algorithm, the Final amount of  $F_1$  is between 0/13 and 0/16.

Accuracy and result of CSPAPT and CSPOPT is more than association rules. Comparing with  $F_1$  through CSPAPT and association rules, it is clear that the results of association rules are poor. The  $F_1$  amount illustrates the amount of accuracy, system operations and executions. Also, it's determined considering the algorithm comparisons. Recommender algorithms have better accuracy to association rules. The results are shown in **Figure 4**.

The other important criterion is the running time of algorithm. On the other word, the time is taken to produce recommender products (Top- $N$  sets). If an algorithm has a suitable and good accuracy and the time is too long, it leads to dissatisfactions. For this reason, the speed is a very important point in E-shops.

CSPAPT and CSPOPT algorithms have more accuracy and speed to compare with association rules and to solve cold start and scalability.

## 6. Conclusion

In this paper, we presented a Recommender System based on complementary and similarity products. Considering the need and taste properties of users, it offers special products. In designed recommender system, two algorithms have been used. CSPAPT offers complementary and similarity products based on previous buying and CSPOPT offers complementary and similarity based on current buying to users. In this paper, we used OWL language to design Electronic Catalogues, calculated the complementary degree based on ontology and offered the need products to users. Recommender Systems can be used

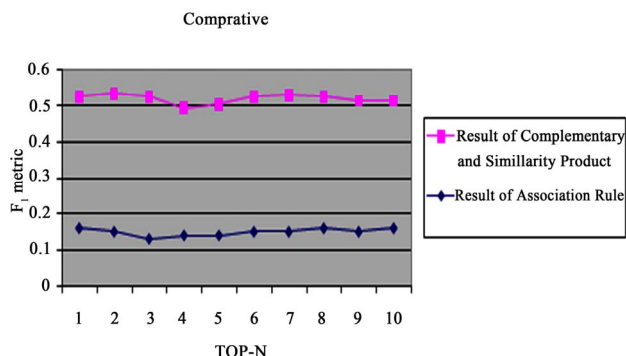


Figure 4. Comparison  $F_1$  in association rules by CSPAPT.

in more accurate applications based on complementary and similarity products, such as promoting selling strategies and designing selling catalogues to increase in selling system operations.

## REFERENCES

- [1] X. H. Sun, F. S. Kong and S. Ye, "A Comparison of Several Algorithms for Collaborative Filtering in Startup Stage," *Proceedings of IEEE Networking, Sensing and Control*, Rome, 19-22 March 2005.
- [2] D. Kalles, A. Papagelis and C. Zaroliagis, "Algorithmic Aspects of Web Intelligent Systems," *WebIntelligence*, Springer, Berlin, 2003, pp. 323-345.
- [3] I. Cantador, *et al.*, "A Collaborative Recommendation Framework for Ontology Evaluation and Reuse," *Proceedings of the International Workshop on Recommender Systems*, 2006, pp. 67-71.
- [4] N. Belkin and B. Croft, "Information Filtering and Information Retrieval," *ACM*, Vol. 35, No. 12, 1999, pp. 29-37. [doi:10.1145/138859.138861](https://doi.org/10.1145/138859.138861)
- [5] G. Adomavicius and A. Tuzhilin, "Toward the Next Generation of Recommender Systems," *A Survey of the State-of-the-Art and Possible Extensions*, *IEEE*, Vol. 17, No. 6, 2005, pp. 734-749.
- [6] M. Papagelis and D. Plexousakis, "Qualitative Analysis of User-Based and Item-Based Prediction Algorithms for Recommendation Agents," *Engineering Applications of Artificial Intelligence*, Vol. 18, No. 7, 2005, pp. 781-789. [doi:10.1016/j.engappai.2005.06.010](https://doi.org/10.1016/j.engappai.2005.06.010)
- [7] L. Kerschberg, W. Kim, *et al.*, "A Semantic Taxonomy-Based Personalizable Meta-Search Agent," *Journal of Innovative Concepts for Agent-Based Systems*, Vol. LNAI 2564, 2003, pp. 3-31.
- [8] R. Knappe, "Measures of Semantic Similarity and Relatedness for Use in Ontology-Based Information Retrieval," Thesis of Doctor, Roskilde University, Roskilde, 2005.
- [9] "OWL Web Ontology Language Reference, W3C Recommendation," 2004. <http://www.w3.org/TR/owl-ref/>
- [10] S. H. Moosavi, M. Nematbakhsh and H. K. Farsani, "A Semantic Complement to Enhance Electronic Market," *Expert System with Application*, Vol. 40, No. 1, 2009, pp. 1-9.

- [11] G. Kowalski, "Information Retrieval Systems: Theory Andimplementation," Kluwer Academic Publisher, Dordrecht, 1997.
- [12] G. I. Webb, "Preliminary Investigations into Statistically Valid Exploratory Rule Discovery," *Proceedings of the Australasian Data mining Workshop AUSDM03*, University of Technology, Sydney, 2003, pp. 1-9.