An Improved Recommendation Approach based on User's Standard and Interests

Rajesh Shukla
PhD Scholar
RGPV Bhopal

Sanjay Silakari, PhD
Professor and Head
Dept of CSE, UIT RGPV, Bhopal

P K Chande, PhD
Ex Professor (IS)
Indian Institute of Management, Indore, India

ABSTRACT
Nowadays, the field of web personalization is growing exponentially. From e-mail, e-trading, and internet forums to social networking based websites, directly or indirectly utilize the concept of web personalization and recommendation system for providing customized services and attention-grabbing offers to their users. By now, a wide range of recommendation systems have been proposed by various researchers and still the research is going on so as to attain the user's expectations. It is a general thought that a user should be recommended the best product from his/her favorite product categories. In this work, we show theoretically that a recommendation system that follows the above criterion does not recommend appropriate products to all types of users. In addition, we propose a user's standard based recommendation system that overcomes this limitation of conventional approach, and provides improved recommendations covering all types of users. For evaluation purpose, we have employed the KDD MovieLens dataset and developed a movie recommendation system based on the proposed approach. We introduce the term 'Gain' for measuring the difference created by the proposed approach as compared to conventional approach. Experimental results reveal that the proposed approach benefited 56% of users and improved 23% of total recommendations.

General Terms
Web Usage Mining and recommendation system

Keywords
Data mining, KDD dataset, recommendation system, web personalization

1. INTRODUCTION
Ecommerce applications use link personalization to recommend items based on the clients buying history or some categorization of clients based on ratings and opinions. Personalization is a relatively new and challenging field for web content delivery. In order to meet expectations of visitors, customers and loyal users, web world is struggling to offer excellent customized services during their interaction with the system. The impact of personalization and recommendation system can be experienced by the rapid popularity that this area has gained in the last few years. Customers preferably choose to visit those websites, which understand their needs, provide them rapid value added customized services and easy access to required information in simple understandable format. Web personalization and recommendation system plays a major role to meet these goals.

Corporate world looks towards the huge volume of transactional and interaction data generated by the Internet for R&D that facilitates the creation of new innovative competitive services and products [2]. In today's e-business world, most of the major e-commerce players have adapted the web personalization and recommendation system including Yahoo!, Amazon, eBay, Netflix, NewsWeeder, IBM and many more.

A recommendation system learns from a customer’s behavior and recommends product(s) in which a user may be interested. It helps to build up a long lasting relationship with loyal users of website. Various vendors offer web personalization tools that can be employed in existing systems to achieve personalized web system.

IBM Product Recommendations tool helps to present the most relevant, effective and timely recommendations wherever customers are in the buying process [8]. Adobe (Omniture) Test & Target is an optimization tool that identifies and executes unlimited A/B and multivariate tests (MVTs) whenever required. It also measures the effectiveness and relevance of contents across any online channel and increase content relevance through segmentation, targeting, and automated personalization [6]. BT Buckets is a free personalization tool and behavioral targeting tool that can be integrated with Google Analytics. Another tool Magiq Dynamic Personalization allows marketing teams to conduct website content personalization campaigns, and optimize their digital marketing communications to the individual customer [7]. The WP Greet Box plug-in enables a website to display different greeting message to each visitor depending upon the web pages visited by him previously. ATG's eStara enables customers to initiate phone calls with the help of agents that assist customers with their order.

So far, a variety of recommendation approaches have been proposed to generate correct predictions which include collaborative or social filtering, content based, knowledge based, and hybrid [12]. Collaborative systems aggregate data about customers’ purchasing habits and/or preferences, and make recommendations to other users based on similarity in overall purchasing patterns. This technique was introduced by Goldberg et al in 1992. In 1998, Basu et al [3] introduced content based filtering systems, which are classifier systems derived from machine learning research. The knowledge
based recommender systems use knowledge about visitors and/or products to reason about what product(s) meet the user’s requirements and generating the recommendations [13]. Hybrid recommender systems combine more than one of the above mentioned techniques together in order to their disadvantages and obtain better recommendations.

This paper is organized as follow: section.2 presents a brief overview of recently emerged recommendation systems. Proposed recommendation approach is described in section.3. It includes the outline, description and the algorithm of proposed approach for improving the recommendations. Section.4 shows the experimental setup for evaluating the effectiveness of proposed recommendation system. Major findings, results and discussions are presented in section.5. Finally, section.6 concludes the paper.

2. AN OVERVIEW OF RELATED RECENT WORKS

[Dimitrios Pierrakos et al., 2012] presented a system that builds and maintains community web directories by employing a web usage mining framework that offers a range of personalization functionalities. It was named as OurDMOZ, which includes adaptive interfaces and web page recommendations. [A. Vaishnavi, 2012] proposed a technique for developing web personalization system using Modified Fuzzy Possibilistic C Means (MPPCM). The author claims that this approach raises the possibility that URLs presented before a user will be of his interest.

[Bin Xu et al., 2012] extends the traditional clustering collaborative filtering models. They formulate the Multiclass Co-Clustering (MCoC) problem and propose an effective solution to it in order to find meaningful subgroups. They also propose a unified framework that extends the traditional collaborative filtering algorithms by utilizing the subgroups information, for improving their top-N recommendation performance.

So far, collaborative filtering has been the most successful technique in the design of recommender systems, where a user is recommended those items, which people with similar tastes and preferences liked in the past. Although the studies of tag-aware recommender systems have achieved fruitful goals, but there are still few challenges that are yet to meet, which are highlighted in [Zi-Ke Zhang et al., 2012]. [Tamas Jambor et al., 2012] applied the principles of modern control theory to recommendation system, and suggested how to construct and maintain a stable and robust recommender system for dynamically evolving environments.

[Ronaldo Lima Rocha Campos et al., 2011] proposed a multi-agent based system application model for indexing, retrieving and recommendation learning objects that are stored in different and heterogeneous repositories. In order to improve the accuracy, they have come up with an information retrieval model, which is based upon the multi-agent system approach and an ontological model.

[Joseph A. Konstan et al., 2012] have shown that the embedding of the algorithm in the user experience dramatically affects the value of the recommender to user. They argue that evaluating the user experience of a recommender requires a broader set of measures than have been commonly used, and suggest additional measures that have proven effective. [Vivek Arvind. B et al., 2012] proposed an intelligent recommendation system that utilizes the boosted item based collaborative filtering (for the efficient rating of predicted items) and association rule mining technique (for making a personalized recommender system for the target user). This system improves the overall web recommendation precision.

[Ilham Esslimani et al., 2011] presented a new densified behavioral network based collaborative filtering model (D-BNCF), based on the BNCF approach, which uses navigational patterns to model relationships between users. This approach achieves a high precision when new links are exploited to compute the predictions.

Recommendation services play a vital role in E-commerce. After analyzing the feasibility to combine Case-based Reasoning (CBR) and web log mining with recommendation system in E-commerce, [Ya-min Wang et al., 2011] integrated the two techniques into E-commerce recommendation system effectively. CBR is a paradigm that takes advantage of knowledge obtained from past experience and the state of a specific problem. It uses past similar cases and their solutions in the state of the new problems for solving them. In this work, they proposed a framework for recommendation system that is closer to human thinking mode.

A challenging problem in recommendation systems deals with unvisited or newly added pages. [Rana Forsati et al., 2009] addresses this problem by introducing a novel Weighted Association Rule mining algorithm. This method can improve the overall quality of web recommendations. Collaboration filtering techniques used for recommendation require accumulation of vast amount of historical user-preference information, which is queried to provide a personalized experience. Model-based collaborative filtering techniques are preferred over the somewhat more accurate memory-based collaborative filtering techniques primarily due to their higher efficiency and scalability. Next section describes the proposed recommendation approach in detail.

3. PROPOSED WORK

3.1 Outline

In case of online shopping, reputation of a product is determined on the basis of its feedback, provided by those customers who purchased and used it previously. This feedback is generally collected in the form of user rating where users are given the flexibility to give marks to the product that they purchased earlier according to its usefulness. Other users see the average ratings of relevant products and try to judge them on the basis of it. Recommender systems also utilize the ratings provided by various users in predicting the product(s) which the users may find interesting. But it is not always a good strategy to pick the highest rated top k products among user’s favorite category and recommend them to him.
It is because not all the users are interested in purchasing the best quality products as most of the time, they are costly. Rather each user has its own standard, which we define as follow: The standard of a customer can be determined by averaging the reputation (rating) of all the products purchased by a user so far. The basic idea is that while recommending the product(s) to a customer, not only customer's interest but also his standard must be taken into account while recommending products to him.

3.2 Description
Generally, an e-trading website presents a number of products, which are categorized according to certain criteria. Like other approaches, the proposed approach also determines the level of interest of individual user in various product categories. There can be several criteria that can be used to determine user interest such as number of products purchased in each category, number of products rated in each category, and products checked by the visitors. After determining interest of each user, the product categories are sorted for each user according to his/her interest in various categories.

It can be explained with a simple example where there are three categories c1, c2 and c3 of available products, and the two users represented by U1 and U2. Let us suppose U1 purchased two products from category c2 and one product from c1, whereas U2 purchased four products from c3 and two products from c1 then the User-Interest table can be represented as follow:

### Table.1. The User-Interest Table for above scenario

<table>
<thead>
<tr>
<th>User</th>
<th>Interest</th>
</tr>
</thead>
<tbody>
<tr>
<td>U1</td>
<td>&lt;c2,c1,c3&gt;</td>
</tr>
<tr>
<td>U2</td>
<td>&lt;c3,c1,c2&gt;</td>
</tr>
</tbody>
</table>

Once the User-Interest Table is obtained, which represents the interest of each user in various categories of products; it is time to determine the standard of each customer. The standard of ith user can be obtained according to following formula:

$$\text{Standard}(U_i) = \frac{\sum_{P=1}^{n} \text{Average rating of product } P \text{ purchased by } U_i}{n}$$

Where Ui represents ith user, and n denotes the total number of products purchased by Ui.

After obtaining the interest and standard of each user, the recommender system is ready to recommend the products. Let the recommender has to recommend N products to a user that best suits to his/her profile then first of all it searches the products whose reputation (rating) is equal or nearby the standard of that user.

We define a variable called Precision Value (PV) that tells recommender, how much difference between user’s standard and product’s reputation should be ignored and considered as a match. If the recommender system has to recommend best matching products to a user Ui then the recommender considers only product P as a candidate to be added into Recommendation List (RL) that satisfies following two conditions:

$$P \in \text{category } c_j; \ [\text{User-Interest Table contains } c_j \text{ for } U_i]$$

and

$$\left| \text{Standard}(U_i) - \text{PV} \right| \leq \text{Reputation of Product } P \leq \left| \text{Standard}(U_i) + \text{PV} \right|$$

All such products that satisfy above two criteria will be added into RL until either the size of list becomes N or all the categories for user Ui in User-Interest Table have been searched. If the number of products in RL equals to N then all those products are recommended to the user. If only M products are found in user’s most favorite category such that M<N then his/her next favorite category is searched for remaining N-M products and products matching the Standard(Ui) are again added into RL. This process continues until either all the categories are searched or when N recommendations are generated.

When all the categories for user Ui in User-Interest Table have been searched and the number of selected products is less than N then PV is automatically increased by a small value ΔPV and then the whole process is repeated so that remaining recommendations can be determined. If number of elements in RL is still less than N then Standard(Ui) factor is simply ignored and first N products from user’s most favorite category and are recommended to the user. In our previous example, if for user U1, the Standard(U1) is 3.5, PV is 0.5, N is 5 and ΔPV is set to 0.5 where the ordered set <c2,c1,c3> represents that user U1 is most interested in products that belong to category c2, then those belonging to c1 and he/she is least interested in products that belong to category c3. Then according to proposed approach, the recommender system will firstly search for that product from category c2, which satisfies the following criterion:

$$[3.5 - 0.5] \leq \text{Reputation of Product } P \leq [3.5 + 0.5]$$

If five such products are successfully obtained from c2 then the search is over otherwise remaining products will be searched in next favorite category i.e. c1 according to the same criterion. This search continues until either searching from all the categories i.e. c2, c1 and c3 is complete or five products are found that satisfy above criterion. If number of products satisfying the above criterion, found in all three categories is less than five then PV is raised by 0.5 i.e. value of new PV becomes 4 and the same procedure is repeated again. If five products can still not be found then the simply top five products from user’s most favorite category (which is c2 in this case) will be selected for recommendation. Next subsection presents the algorithm for proposed recommender system.
3.3 Algorithm

Let $ARP_i$ denote the Average Rating of $i^{th}$ product, and $UI_i$ be the ordered set that contains product categories in which the $i^{th}$ user i.e. $U_i$ is interested. Remember that $UI_i$ is represented by an ordered set containing the elements in decreasing order of user interest. Term $Standard(Ui)$ represents the average standard of user $U_i$. Precision value $PV$ and $\Delta PV$ are floating point variables whose values are pre-defined by the administrator. Assuming that the recommender has to search for $K$ products, which match the user profile, the algorithm of proposed approach can be given as follow:

**Step.1. Calculate Average Rating of Products:**

For all $i$, calculate the $AR(P_i)$ according to following formula:

$$AR(P_i) = \frac{\sum_{j=1}^{m} rating\ given\ to\ P_i\ by\ U_j}{number\ of\ users\ who\ rated\ P_i}$$

Where, $P_i$ denotes $i^{th}$ product among all available products, and $m$ denotes the total number of users.

Note: If $U_j$ did not rate $P_i$ then a zero value is assumed

**Step.2. Determine User's Interest in various product categories**

For each user, determine the level of interest in various product categories as follow:

2.1. Calculate the number of products purchased by user $U_i$ in each category. For a product category $C_p$, let this value be $N_{C_p}$

2.2. Generate ordered set of product categories such that if $C_p$ lies on left of $C_q$ in the set if and only if $N_{C_p} > N_{C_q}$

**Step.3. Determine User's Standard**

Determine the level of standard for each user using the following formula:

$$Standard(Ui) = \frac{\sum_{i=1}^{n} AR(P_i) purchased\ by\ U_i}{n}$$

Where $Ui$ represents $i^{th}$ user and $n$ denotes the total number of products purchased by $U_i$

**Step.4. Predict the products to be recommended**

Determine the size of ordered set determined in Step.2 and store in variable $SIZE$

Round=1; index=1; numofrecom=0;

UNTIL index<=size

Select those products from category $C_index$ and add to Recommendation List RL that satisfy the following condition:

$$[Standard(Ui) - PV] \leq \text{Reputation of Product} \leq [Standard(Ui) + PV]$$

If size of RL==K then exit from loop

[This situation has been termed as HIT]

Else continue in the loop

End If

End of loop

If size of RL<K, then

[This situation has been termed as MISS]

If Round==1 then set Round=2 and $PV=PV+\Delta PV$

Repeat Step.4 once again

End If

**Step.5. Check whether K predictions are generated successfully or not**

If size of RL==K then recommend the products available in RL and to user $U_i$ and exit

Else

Select $K$ products from first category present in ordered set of user $U_i$ that have highest reputation and recommend them to $U_i$

End If

4. EXPERIMENTAL SETUP

For evaluation purpose, we have developed a movie recommender system, which is based upon the proposed approach. We have used the standard MovieLens dataset to recommend the movies to users in which, they may be interested. First of all, we determined the distribution of average ratings of movies, which is shown in Fig.1. Since average ratings of each movie was found to be greater than or equal to 1, hence the distribution is shown only between 1 to 5 (excluding the range between 0 and 1). The average rating of most of the movies rests between the values 3 and 4.5.
Total number of recommendations to be made (i.e. $K$) has been set to 5. The value of $PV$ should be taken wisely because if it is too small then most of the time, recommender may not be able to find sufficient number of recommendations. In other words, the number of miss may become high, which will degrade the performance of recommender system and reduce the accuracy. On the other hand, if this value is kept large then it will diminish the sole purpose. Therefore we have analyzed the number of MISS occurrences when $PV=0.01, 0.02, 0.03$ and $0.05$, which is illustrated in Fig.2.

It can be observed from the graph that MISS occurrence is too high when $PV=0.01$, and it falls down instantly when value of $PV$ is reduced to $0.02$. Further increase in value of $PV$ is not putting any significant effect on MISS occurrence. Due to this reason, we have taken $PV=0.02$.

5. FINDINGS, RESULTS AND DISCUSSIONS

The distribution of individual users' standard i.e. $Standard(U_i)$ is shown in Fig.3, which shows that there is a wide variation in the standards followed by users and maximum distribution lies from level 2.5 to 4.5.

In MovieLens dataset, there are 19 different categories, namely unknown, action, adventures, animation, children, comedy, crime, drama, fantasy, filmnoir, horror, musical, mystery, romance, scific, thriller, war and western; and a movie may belong to multiple categories. Fig.4 shows tabular as well as graphical representation of category wise user interest in various movie categories. It can be observed that maximum users found those movies interesting, which belong to drama, second favorite category is comedy, then action and so on.
A conventional recommender system would ignore the Standard(Ui) factor and recommend simply the top rated movies from user's favorite category, but proposed system recommends tries to recommend those movies, which match the user's standard i.e. that fall within the distance PV from Standard(Ui). Therefore we defined a term Gain that denotes the percentage of movies, which belong to user's most favorite movie category but do not match corresponding user's standard.

In this case, conventional recommender systems would recommend those movies that belong to user's most favorite category but does not match user's standard whereas the proposed recommender system picks up those movies, which belong to user's next favorite categories and also matches user's standard. This will increase the probability that a user would not only visit those movies but also buy/watch them.

Fig.5(a) through (h) shows graphical representation of gain achieved by the proposed recommender system for eight randomly selected users.
Fig. 5. Details of generated recommendations and Gain achieved by proposed system for eight randomly selected users.
For user U1, all the five recommendations have been chosen from user's most favorite category i.e. drama and therefore value of Gain is 0. For U3, only 2 recommendations belong to his most favorite category i.e. drama whereas two recommendations belong to thriller category. Thus the Gain for U3 can be calculated as Gain(U3) = 
\[
\frac{3}{5} \times 100 = 60\%
\]
Similarly, user U20, U80, U103, U252, U346, and U445 have achieved a Gain of 60%, 20%, 40%, 60%, 20%, and 80% respectively.

6. CONCLUSION
In this paper, we presented an enhanced recommender system that improves the quality of recommendations by taking into account, the level of standard of users. The proposed system first determines the standard of each user and then presents the recommendations before them according to it. Since the purpose of a recommender system is not only to recommend those products to the users, which they find attractive but also it should match the individual user's standard so that they move one step forward and purchase them willingly. Results reveal a significant improvement of recommendations as compared to conventional approach. The recommender system presented in this paper is equally beneficial to online customers as well as traders and it can be employed in almost all sorts of e-trading systems.

7. REFERENCES


