International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 2 Issue 2, February- 2013

REAL : Recommender System For Resources And Educational Assistants For Learners

Khin Nila Win
Department of Computer Operation, University of Computer Studies (Mandalay), Myanmar

Abstract

One of the key challenges in web is information overloading due to extraction of irrelevant data. Recommendation techniques are developed to overcome this problem. Collaborative Filtering techniques are the most popular techniques in Recommender Systems. In this paper, Item-based collaborative filtering approach is used by applying modified adjusted cosine similarity to get more accurate item similarity. Modified weighted sum is used for prediction to recommend the useful links matched with the user's interests. The proposed system recommends scholar links and the useful links for learners willing to learn in various fields based on learner-related data such as degree of students and their interests, and the recommendations of other learners with similar characteristics. The user needs to register by giving their proficiency level and interested subjects. Then, the system also considers not only the ratings of the similar users but also item demographic data according to the topic similarity.

1. Introduction

Most recently, Web usage mining has been proposed as an underlying approach for Web personalization [1]. The goal of personalization based on Web usage mining is to recommend a set of objects to the current (active) user, possibly consisting of links, ads, text, products, or services, tailored to the user's perceived preferences as determined by the matching usage patterns. To date, a variety of recommendation techniques has been developed. Several related algorithms that are often used in recommendation processes are Rule-based filtering, Content-based Recommendation, and Collaborative filtering Recommendation [2]. Collaborative filtering systems, also referred to as social filtering, match the rating of a current user for items with those of similar users in order to produce recommendations for items not yet rated or seen. Collaborative filtering has been known to be the most successful recommendation technique that has been used in a number of different applications such as recommending web pages, movies, and products [3]. However, applying the user-based collaborative filtering approach has two challenges in terms of scalability and sparsity. The proposed system uses the item-based collaborative filtering (CF) approach instead of traditional user-based approach to overcome these two challenges. Moreover, the system predetermines the user's interests by applying the stereotypes collaborative filtering approach. Therefore, the system can get both the advantages of item-based CF and stereotypes CF and overcome the cold-start problem for new user.

The inexperienced learners, who are willing to get international scholarships, need to know in advance which university is suitable to apply for and which qualifications are needed to fulfil the admission requirements of the appropriate university. The learners have difficulties to find the learning resources that are suitable for their current proficiency level because these resources are scattered in various web links according to topics. The proposed system organizes these resources such as e-books links or online course links and recommends to the user. The main purpose of the system is to provide the meaningful recommendations about scholarship links that are offered by international universities.

We proposed REAL for the following purposes:

1) To provide the most efficient way of searching the usable educational links for new learners
2) To reduce the time for learners in searching the educational links and e-books
3) To utilize the knowledge of the experienced users for the new users
4) To enhance the accurate recommendation of the item-based similarity

The rest of the paper is organized as following. In section 2, overview of recommendation techniques is described. Section 3 explains architecture of the system and the procedures of the proposed system, REAL recommender system, are summarized. And the paper discusses about the system in Section 4. Evaluation of the system is represented in Section 5 and Section 6 presents the conclusion.

2. Overview of Recommendation Techniques

There are varieties of recommendation techniques. The most commonly used recommendation techniques are Content-Based Recommendation and Collaborative Filtering Recommendation.

2.1. Content-Based Recommendation

Content-based recommender systems work with profiles of users that are created at the beginning. A profile has information about a user and his taste. Taste is based on how the user rated items. Generally, when creating a profile, recommender systems make a survey, to get initial information about a user in order to avoid the new-user problem [6]. Content-based filtering systems recommend items that are considered sufficiently similar to the content descriptions in the user profile.

2.2. Collaborative Filtering Techniques (CF)

There are various types of collaborative filtering:

1) *User-based CF* -Users who rated the same item similarly probably have the same taste. Based on this assumption, this technique recommends the unseen items already rated by similar users.
2) *Item-based CF* -Focus on items, assuming that the items rated similarly are probably similar. It recommends items with the highest correlation (based on ratings for the items).
3) *Stereotypes or Demographics CF* -Users with similar attributes are matched, and then it recommends items that are preferred by similar users (based on user data instead of ratings).

The process of collaborative filtering is illustrated in figure 1.

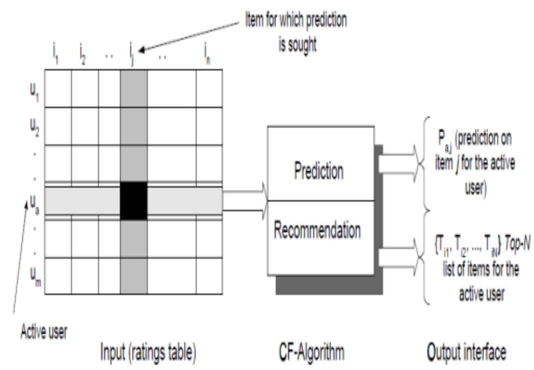

Figure 1. Collaborative Filtering Process

2.3. Challenges of CF Techniques

Improving the scalability and the quality of the recommendation for the users is very important for recommender systems. User-based CF system has been very successful but has two challenges: Sparsity and Scalability.

❖ *Sparsity*

Many commercial recommender systems are used to evaluate large item sets. In these systems, even active users may have purchased over 1% of the items. Recommender systems based on nearest neighbor algorithms may be unable to make any item recommendation for a particular user. As a result, the accuracy of recommendations may be poor.

❖ *Scalability*

Nearest neighbour algorithms require computation that grows with both the number of users and the number of items. With millions of users and items, a typical web based recommender system running existing algorithms will suffer serious scalability problems [4].

❖ *Cold-Start Problem*

New items and new users pose a significant challenge to recommender systems. Collectively these problems are referred to as the cold-start problem. New item problem or the first-rater problem arises in collaborative filtering systems, where an item cannot be recommended unless some user has rated it before. This issue applies not only to new items, but also to obscure items, which is particularly detrimental to users with eclectic tastes. The new-user problem is difficult to tackle, since without previous preferences of a user it is not possible to find similar users or to build a content-based profile. As such, research in this area has primarily focused on effectively selecting items to be rated by a user so as to improve recommendation performance rapidly with the least user feedback [5].

3. Architecture of the System

The architecture of the system is depicted in Figure 2. The system organizes the educational links from the web by using topical crawler. The crawled pages are stored in the database and displays to the user. And then, the system requests the users to rate the pages which is useful and efficient for them. The rating of the experienced users is stored in the database and the recommended pages are displayed to the users according to the profile similarity, rating similarity, and topic similarity.

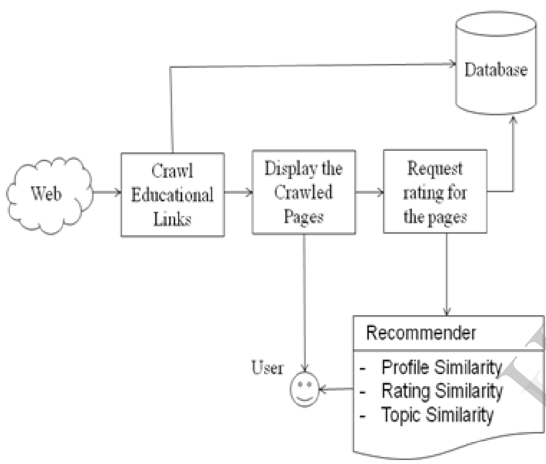

Figure 2. System Architecture

3.1. REAL Recommender System

The proposed system works in the following procedures.

Step 1:

- ❖ Crawls the web using topic based crawler to obtain educational web pages.

Step 2:

- ❖ When the user enters the system, the system checks whether the registered user or new user.
- ❖ For the new users, it allows registration to provide the recommend links.
- ❖ To register, the user needs to enter the following attributes:
  1) Name
  2) Certificate(Local/International)
  3) Degree(Local/International)
  4) Country
  5) Interested Subject
  6) Learning Goal, etc.
- ❖ Save the user profile to the database.

Step 3:

- ❖ Computes Stereotypes or Demographic Similarity for target user.
- ❖ Displays the web pages based on stereotype similarity and recommendation of the similar users.

Step 4:

- ❖ Requests the users to rate his or her interested web pages in the site.

Step 5:

- ❖ Calculates item similarity using modified adjusted cosine similarity.

Step 6:

- ❖ Calculates prediction using modified weighted sum.

Step 7:

- ❖ Generates the recommend links to the users.

3.1.1. Similarity Computation

Most of the recommender systems usually use three similarity computing techniques: Cosine-based Similarity, Correlation-based Similarity, and Adjusted Cosine Similarity.

❖ *Cosine-based similarity*

In Eq. (1), similarity between items i and j, sim(i,j) in the m * n ratings matrix is presented.

$$sim(i,j) = \cos(\vec{i},\vec{j}) = \frac{\vec{i}.\vec{j}}{\|\vec{i}\|_2 * \|\vec{j}\|_2} \qquad (1)$$

Where "." denotes the dot-product of the two vectors.

❖ *Correlation-based similarity*

Let the set of users who both rated i and j are denoted by U then the correlation similarity is given by Eq. (2).

$$sim(i,j) = \frac{\sum_{u\in U}(R_{u,i}-\overline{R_i})(R_{u,j}-\overline{R_j})}{\sqrt{\sum_{u\in U}(R_{u,i}-\overline{R_i})^2}\sqrt{\sum_{u\in U}(R_{u,j}-\overline{R_j})^2}} \qquad (2)$$

$R_{u,i}$ denotes the rating of user u on item i, $\overline{R_i}$ is the average rating of the i-th item.

❖ *Adjusted Cosine Similarity*

$$sim(i,j) = \frac{\sum_{u\in U}(R_{u,i}-\overline{R_u})(R_{u,j}-\overline{R_u})}{\sqrt{\sum_{u\in U}(R_{u,i}-\overline{R_u})^2}\sqrt{\sum_{u\in U}(R_{u,j}-\overline{R_u})^2}} \qquad (3)$$

In Eq. (3),
$\overline{R_u}$

is the average of the u-th user's ratings.

Computing similarity using basic cosine measure in item-based case has one important drawback – the differences in rating scale between different users are not taken into account. The adjusted cosine similarity offsets this drawback by subtracting the corresponding user average from each co-rated pair. However, it has one drawback- the different rating styles of the different users are not taken into account.

For example, in the case of the system's range of the rating is 1 to 5, user i sets the rating 3 to his/her most like item t, while the other user j sets the rating 5 to his/her most like items t. In such case, the system can't assume the item t is the user i's most likes while it assumes this item is the user j's most likes. So, the system can't determine the highest rating of the users and can't assume the user's most like even if it is the user's highest rating in the case of not being highest rating of the system. Therefore, the system needs to normalize the rating style to accurately determine which the user most like and which the least even if the users have different rating styles. The proposed system applies the normalized rating to over such problem. The proposed method, modified adjusted cosine similarity, can reduce misunderstanding of the system on the users' likes and dislikes.

❖ *Modified Adjusted Cosine similarity*

$$sim(i,j) = \frac{\sum_{u \in U}(NR_{u,i} - \overline{R_u})(NR_{u,j} - \overline{R_u})}{\sqrt{\sum_{u \in U}(NR_{u,i} - \overline{R_u})^2}\sqrt{\sum_{u \in U}(NR_{u,j} - \overline{R_u})^2}} \quad (4)$$

In Eq. (4),

$\overline{R_u}$ is the average of the u-th user's ratings

And,

$$NR_{u,i} = \frac{HS}{HR_u} * R_{u,i} \quad (5)$$

In Eq. (5),

HS means highest rating scale of the system

HR_u means highest rating scale of the current user

Considering the topic similarity of item,

Where,

$enh\_cor_{ij} = sim_{ij} + sim_{ij} * dem\_cor_{ij}$

sim_{ij} means the similarity of item i and item j from the adjusted cosine similarity after normalizing the user's rating behaviour,

dem_cor_{ij} means the similarity of the item i and item j according to the topic similarity.

3.1.2. Prediction Computation

❖ *Modified Weighted Sum*

Weighted Sum computes the prediction on an item i for a user u by computing the sum of the ratings given by the user on the items similar to i. Each rating is weighted by the corresponding similarity $s_{i,j}$ between items i and j.

In Modified Weighted Sum in Eq. 6, each normalized rating, $NR_{u,N}$ in Eq. 7, is weighted by the enhanced correlation similarity enh_cor_{iN}.

The prediction $P_{u,i}$ is denoted as

$$P_{u,i} = \frac{\sum_{allsimilaritems,N}(enh\_cor_{iN} * NR_{u,N})}{\sqrt{\sum_{allsimilaritems,N}(|enh\_cor_{iN}|)}} \quad (6)$$

Where,

$$NR_{u,N} = \frac{HS}{HR_u} * R_{u,N} \quad (7)$$

Modifying the weighted sum by enhanced correlation similarity performs the prediction more accurately than the existing systems. Each of the systems considering the item demographic data produces the prediction quality more than 9% higher than the systems which do not consider the item demographic data.

4. Discussion

Most of the earlier learning resources recommender systems find the problems in determining the recommended pages accurately since they ignore the rating style of the current user. The proposed system overcomes this challenge by normalizing the current user's rating style. And in the section of similarity computation, the system considers the rating similarity accompanying with topic similarity of resources pages.

To avoid the cold-start problem for users earlier system encountered, the proposed system uses stereotypes or demographic CF. As a result, the system takes advantages of not only item-based CF and but also stereotypes or demographic CF. Moreover, the system can avoid the scalability and quality bottleneck of the user based CF since it uses item-based Collaborative Filtering Techniques. Modifying adjusted cosine similarity with normalized rating of users and modifying weighted sum with enhanced correlation similarity are not only able to determine accurately which the user's most likes but also able to produce the higher prediction quality than the systems which do not consider the item demographic data and only emphasize the rating of the users. The system can reduce mean absolute error (MAE) between the predicted ratings and actual ratings of the users due to the advantages of modified adjusted cosine similarity and modified weighted sum.

5. Evaluation of the System

The quality of a recommender system can be evaluated by comparing recommendations to a test set of known user ratings. These systems are typical measured using predictive accuracy metrics [2], where the predicted ratings are directly compared to actual user ratings. The most commonly used metric in the literature is Mean Absolute Error (MAE) – defined as the average absolute difference between predicted ratings and actual ratings, give by:

$$MAE = \frac{\sum_{\{u,i\}} \left| P_{u,i} - r_{u,i} \right|}{N} \qquad (8)$$

Where,
$P_{u,i}$ is the predicted rating for user u on item i,
$r_{u,i}$ is the actual rating,
And N is the total number of ratings in the test set.

The proposed system can reduce MAE by applying both demographic correlation and rating similarity of items.

6. Conclusion

The proposed recommender system, REAL, can provide the meaningful educational links to the inexperienced learners. First, it shows the links that the similar existing users rated by computing stereotypes or demographic similarity. To compute the item based similarity, it uses modified adjusted cosine similarity and to predict more usable recommendations, it calculates modified weighted sum. In this system, combining the rating similarity and topic similarity can achieve more accurate predictions for recommendation.

The proposed system can support the learners to achieve the scholarship of their interested subjects that offered by the international universities based on the recommendation of the experienced learners. It can also reduce the amount of time for finding the suitable resources and scholar links.

7. References

[1] Srivastava, J., et al., "Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data". SIGKDD Explorations, 2000.
[2] Jonathan L. Herlocker, Joseph A. Konstan, Loren G. Terveen, and John T.Riedl,"Evaluating collaborative filtering recommender systems". ACM Trans.Inf. Syst., 22(1):5–53, 2004.
[3] Herlocker, J., et al,"An Algorithmic Framework for Performing Collaborative Filtering", Proceedings of the 22nd ACM Conference on Researchand Development in Information Retrieval (SIGIR'99). 1999.
[4] BadrulSarwar, George Karypis, Joseph Konstan, and John Riedl, "Item-based Collaborative Filtering Algorithms",ACM 1-58113-348-0/01/0005, May 1-5, 2001, Hong Kong.
[5] Andrew I. Schein, AlexandrinPopescul, Lyle H. Ungar, and David M. Pennock,"Methods and metrics for cold-start recommendations", SIGIR '02:Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval, pages 253–260, New York,NY, USA, 2002. ACM.
[6] L. Candillia, F. Meyer, F. Fessent, and K. Jack,"State-of-the-art recommender systems", 2009.