Personalized Recommendations for Online Educational Resources: A Novel Clustering Method to Increase Effectiveness for Search and Use

Abstract: The idea of personalized recommendations has drawn great attention in the last decade due to the exponential growth of metadata available through web sites. In the realm of Internet-based educational materials, a need exists to provide effective access to the growing body of resources available to educators and students. In this paper we propose a new technique to combine discriminative weights with a novel sparse matrix clustering method to improve current personalized recommendation algorithms. Two current popular personalized recommendation methods are employed: Collaborative Filtering and Bipartite Graph Projection. A famous benchmark *MovieLens* is used to evaluate the performance. Experimental results indicate that the proposed method achieves improvements over traditional methods, with potential to create more accurate and more useful searches that benefit all audiences accessing open educational materials.

Introduction

Due to the exponential growth of educational-related data available via web sites, it is often tedious for a student or teacher to explore and discover relevant items through standard methods. Well-studied and commercialized search engines like Google will often help users to find what they are seeking. However, if those searching do not know exactly what they are looking for, or they do not know the "proper" words to describe what it is that they want, the searching results returned are often unsatisfactory. Further, most search engines do not take into account personalized information—such as known preferences or proclivities of the individual searcher--and thereby produce the same result for users with different interests. For example, if the learner's interest is different from the mainstream, the search result will be less meaningful to the learner. For these reasons, personalized recommendation systems have drawn great attention of e-business web sites to improve user's satisfaction and retention (Brusilovsky et al. 2007; Chu et al. 2009). Those searching for specific educational resources, and the owners of repositories of educational resources, face the same challenges as those within the e-business realm. Certainly, the archiving, retrieving and social nature of Internet-based educational resources is of great concern and the focus of study to those within learning sciences (e.g., Johnson-Henson 2008, Seow 2008).

An effective personalized recommendation system will automatically gather and store the former behaviors of users (in profiles), analyze those behaviors and based on the analysis results, designate different interests for different users. The engines then recommend the right item to the right user (Mobasher et al. 2005). To simplify, personalized recommendation systems map items to users based on the analysis of the users' profiles. Different users will receive different recommendation lists. To do so, the system assigns items to users with recommendation scores. The higher the recommendation score, the higher the item should appear in the recommendation list. Finally, a method is employed to shrink the recommendation list to the appropriate length, often referred to as "top-N method" or "threshold."

Profiles used for personalized recommendations are usually the ratings that the users give to the items. For instance, Youtube.com employs a "5 star" rating system for videos while The Internet Movie Database (IMDB) uses a "10 star" system for rating movies. Researchers are giving significant attention to efforts at designing efficient personalized recommendation algorithms due to the performance of the economy in the last decade (Belkin 2000; Montaner et al. 2003). However, these researchers have ignored the importance of the quantization of those ratings. The common strategy for recommendation algorithms is to equalize the ratings into the same interval. In this paper, we will propose a new method to mapping the ratings to discriminative weights. The experimental results showed that the new method dramatically improved the performance of the personalized recommendation system. To show the generalization of the discriminative weights and compare the results in different algorithms, we employed and improved two personalized recommendation approaches: one is the widest applied Collaborative Filtering (CF (Resnick et al. 1994)) method and the other one is the recently developed Bipartite Graph Projection method (BG

(Zhou et al. 2007)). The impact of the new algorithm, when applied to the effective retrieval of educational resources, may impact how communities form around educational interests and applications as well as increase the effective use of educational materials.

Personalized Recommendations Based on CF and BG

Collaborative Filtering is based on similarity analysis among users' profiles. It can be categorized into two classes: memory-based (Jin et al. 2003) and model-based (Xue et al. 2005). Memory-based algorithms perform the computation on the entire database to identify the active user to similar users while model-based algorithms first group the users into classes and then perform the computation within each class. Compared with memory-based approaches, model-based approaches improve the scalability as well as the accuracy, especially for large datasets. However, little effort has been made to design an appropriate clustering method for this specific personalized recommendation problem. In this paper we will propose a novel clustering algorithm. Combined with discriminative weights, it achieves the highest performance.

Bipartite Graph Projection considers the relationship between users and items as a bipartite graph and employs graph projection theory to compute the recommendation list. BG uses the entire database in its computation. In this paper, however, we show that it also achieves better performance by partitioning the database into different groups and executing projection within groups. The same novel clustering method is employed to partition the database. Experimental results show that BG also gets better results when it is combined with the discriminative weights.

Generally, a personalized recommendation system consists of users and items. Each user owns a profile that indicates the ratings that the user has signed to some items. The task of the personalized recommendation system is to utilize those known ratings to predict unknown ratings and based on the prediction scores, recommend different users different recommendation lists.

Denote the user set as $\mathbf{U} = \{\mathbf{u_1}, \mathbf{u_2}, \dots, \mathbf{u_n}\}$ and the item set as $\mathbf{O} = \{\mathbf{o_1}, \mathbf{o_2}, \dots, \mathbf{o_m}\}$. Thus the recommendation system can be considered as an $\mathbf{n} \times \mathbf{m}$ adjacent matrix $\mathbf{A} = \{\mathbf{a_{ij}}\}$, where $\mathbf{a_{ij}}$ refers to the quantization rating that user $\mathbf{u_i}$ signed to item $\mathbf{o_j}$. Note that since each user may only collect a small proportion of the entire item set, the matrix $\mathbf{A}$ will be very sparse. And $\mathbf{a_{ij}} = \mathbf{0}$ doesn't mean the lowest rating that user $\mathbf{u_i}$ signed to item $\mathbf{o_j}$, but it denotes the rating is missing and it is the task of the personalized recommendation system to find out the missing rating.

Collaborative filtering method is based on the similarity measure between users. The similarity between user $\mathbf{u_i}$ and $\mathbf{u_j}$ is denoted as $\mathbf{s_{ij}}$ and can be defined in a Pearson-like equation 1:

$$s_{ij} = \frac{\sum_{l=1}^{m} a_{il} \times a_{jl}}{\min\{\sum_{l=1}^{m} a_{il}, \sum_{l=1}^{m} a_{jl}\}} \tag{1}$$

The numerator of $\mathbf{s_{ij}}$ is the dot production between the $\mathbf{ith}$ and the $\mathbf{jth}$ row of the matrix $\mathbf{A}$ and the denominator of $\mathbf{s_{ij}}$ is normalization factor to make $\mathbf{s_{ij}}$ between $\mathbf{0}$ and 1.

The predicted score r_{ij} which means how much user $\mathbf{u_i}$ will like item $\mathbf{o_j}$ is calculated based on equation 2:

$$r_{ij} = \frac{\sum_{l=1, l \neq i}^{n} s_{il} \times a_{lj}}{\sum_{l=1, l \neq i}^{n} s_{il}} \tag{2}$$

For user $\mathbf{u_i}$, all the recommendation scores $\mathbf{r_{ij}}$ are sorted in descending order. And the top-N best items will be recommended to this person.

On the other hand, Bipartite Graph Projection considers set U and set O as two node sets in a generalized bipartite graph $G(U, O, A)$, where A is the adjacent matrix of the ratings and here denotes the weights of the edges in the graph. Figure1 shows the basic idea of the BG method. Here we will not discuss the algorithm in detail but two important equations are introduced as equation 3 and 4.

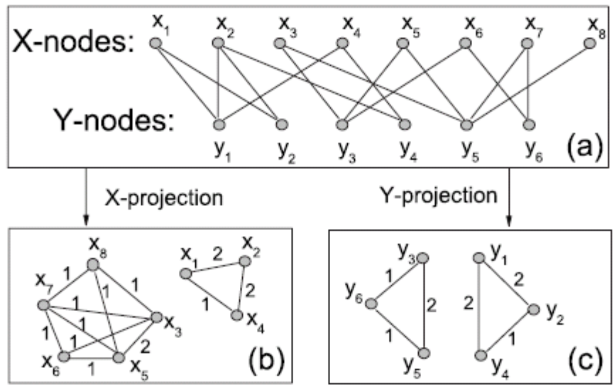

Figure 1. (a) Illustration of a bipartite graph (X $indicates$ U and Y $indicates$ O). (b) X-projection (c) Y-projection. The edge weights in (b) and (c) are the number of common neighbors of Y and X.

Based on the theory of projection, a $\mathbf{n \times n}$ weighted projection matrix $\mathbf{W} = \{\mathbf{w_{ij}}\}$ is created and defined as equation 3:

$$w_{ij} = \frac{1}{\sum_{k=1}^{m} a_{jk}} \sum_{l=1}^{m} \frac{a_{il} a_{jl}}{\sum_{k=1}^{n} a_{kl}} \tag{3}$$

Different with similarity matrix $\mathbf{S}$ of CF, $\mathbf{W}$ is an asymmetric matrix which means $\mathbf{w_{ij}} \neq \mathbf{w_{ji}}$. The weight w_{ij} can be considered as the importance of user $\mathbf{u_i}$ in $\mathbf{u_j}$'s sense, and it is generally not equal to w_{ij} .

The recommendation score can be calculated based on equation 4:

$$r_{ij} = \sum_{l=1}^{n} w_{li} \times a_{lj} \tag{4}$$

After calculating the recommendation score, the same procedure as CF will be employed to generate the recommendation list. In next section, we will propose a new rating quantization method and modify the current CF and BG algorithm to work with the new idea.

CF and BG with Discriminative Weights

All personalized recommendation systems need quantization of ratings. Quantization means converting ratings to a mentioned in the previous section. Suppose we employ a "5 star" rating strategy. A common interpretation would indicate that a rating above or equal to 3 stars means "like" and below 3 stars means "dislike." There are mainly two ways to quantify such ratings. One method retains the ratings above 3 stars and converts them to a "1" (Zhou et al. 2007), which means this method only considers the "like" ratings and ignores "dislike" ratings. This method suffers high false positive problems. Another method equally quantifies the ratings into intervals with same length between 0 and 1 (0.2, 0.4, 0.6, 0.8, and 1) (Wang et al. 2008). This method retains all information which provides additional power to the analysis. However, one star may indicate "I hate it" instead of "I like it a little bit," and based on this relative rating, it is not wise to give low scores a positive value.

We propose a new quantization approach called discriminative weight. We sign positive values to "like" ratings as well as negative values to "dislike" ratings. For any rating system, for example an n -star system, suppose top-m stars indicate "like," then the quantization value a is calculated based on rating r as showed in equation 5:

$$a = \begin{cases} \frac{1}{n-m+1} \times (r-m+1) & if\ r \geq m \\ \frac{-1}{m-1} \times (m-r) & if\ r < m\ and\ r \neq 0 \\ 0 & if\ r = 0 \end{cases} \quad (5)$$

The proposed method maps ratings into a -1~1 interval. For example, based on 5 stars system $n = 5$, $m = 3$. Then 1-5 stars map to -1, -0.5, 0.33, 0.67, and 1 respectively. Based on the new quantization strategy, a new similarity measure between quantization values is suggested. Suppose a_1 and a_2 are quantization values, the similarity s' between a_1 and a_2 is defined as equation 6:

$$s' = \begin{cases} 1 - |a_1 - a_2|\ if a_1 \neq 0\ and\ a_2 \neq 0 \\ 0 \qquad if a_1 = 0\ or\ a_2 = 0 \end{cases} \quad (6)$$

When a_1 and a_2 are 1 and -1, s' is the smallest value -1 which means they could not be more dissimilar. When a_1 and a_2 are the same, s' achieves the largest value 1 which means they are most similar. Notice that in both the CF and BG algorithms, multiplication is used to calculate value similarity. However, those methods suffer problems when handling discriminative weights. For example, the results are negative when a_1 and a_2 are 0.33 and -0.5 which means they are not similar. Based on the new approach, s' is 0.17 which means they are a little bit similar. Based on the knowledge that 2 stars should be a little bit similar with 3 stars, the new similarity score is more reasonable. Both the similarity measure matrix S (equation 1) and weighted projection matrix W (equation 3) are modified based on the discriminative weights and new measure. The modified equations are equation 7 and 8:

$$s_{ij} = \frac{\sum_{l=1}^{m} \left(1 - \left|a_{il} - a_{jl}\right|\right)}{\min\left\{\sum_{l=1}^{m} |a_{il}|, \sum_{l=1}^{m} |a_{jl}|\right\}} \quad (7)$$

$$w_{ij} = \frac{1}{\sum_{k=1}^{m} |a_{jk}|} \sum_{l=1}^{m} \frac{\left(1 - \left|a_{il} - a_{jl}\right|\right)}{\left(\sum_{k=1}^{n} |a_{kl}|\right)} \quad (8)$$

In equation 7 and equation 8, a_{**} refers to the discriminative weights that are generated from equation 5. The discriminative weights combined with CF and BG are denoted as CFD and BGD respectively.

Discriminative Weights Combined with Clustering

Algorithms we discussed in previous sections are all based on the use of an entire database. However, due to the data scarcity problem, clustering users into different groups will dramatically improve the performance. Furthermore, it is advantageous to recommend to a particular user based on the group opinion rather than the global opinion when the group consists of users that are similar to the particular user.

The difficulty of employing clustering method to a personalized recommendation system is considering how to find centers of clusters. Different from a traditional clustering problem, the rating matrix is sparse and 0 does not mean numerical 0 but rather "missing information." Therefore, the mean of 0 and 1 is not 0.5, but unknown. If the arithmetic mean is used to find the center of the cluster, poor performance will be achieved. We propose a novel k-centroid clustering method to deal with this problem. The new algorithm is like the popular k-means algorithm except for finding the center and how to calculate distance. The procedure of the proposed approach is:

1. Randomly initial k clustering centroids (k random users).
2. Calculate the distances between users and current centroids and assign users to the nearest cluster.
3. Recalculate the centroids based on the current clusters.
4. If clusters keep the same or iteration numbers exceed a threshold value, stop. Otherwise return to step 2.

In step 2, if we use discriminative weights method, the distance d_{ij} between user $\mathbf{u_i}$ and centroid $\mathbf{c_j}$ is calculated in equation 9:

$$d_{ij}=\frac{\min\left\{\sum_{l=1}^{m}|a_{il}|,\sum_{l=1}^{m}|c_{jl}|\right\}}{\sum_{l=1}^{m}\left(1-|a_{il}-c_{jl}|\right)} \quad (9)$$

Otherwise, the distance d_{ij} between user u_i and centroid c_j is calculated in equation 10:

$$d_{ij}=\left(\min\left\{\sum_{l=1}^{m}a_{il},\sum_{l=1}^{m}c_{jl}\right\}\right)/\left(\sum_{l=1}^{m}a_{il}c_{jl}\right) \quad (10)$$

Note that distance is the reciprocal of similarity score mentioned in section 2 and 3 because the higher the similarity score the more similar it should be while the lower the distance score the closer it should be. In step 3, within each current cluster, the traditional mean vector of the cluster is first clustered. Based on the histogram analysis, if the 0-1 quantization method is used, the distribution of the elements in the mean vector looks similar to an exponential distribution. If the discriminative weights method is used, the elements in the mean vector is divided into two exponential distributions (<=0, >=0, see figure 2 right). The quantile function of exponential distribution is defined as equation 11:

$$Q(q:\lambda)=\frac{-\ln(1-p)}{\lambda} \quad (11)$$

where p denotes the proportion of elements in the vector and λ is the reciprocal of the mean of all elements in the mean vector. Based on the analysis of the rating quantization values within the current cluster, we can calculate the quantiles according to each rating quantization value. Based on the new clustering algorithm, a similarity measure (equation 1) between users of CF and weighted projection matrix W (equation 3) of BG will executed within the groups which are created by the clustering algorithm. The same strategy is employed for CFD and BGD. Here the algorithms combined with clustering are denoted as CFC, BGC, CFDC, BGDC, respectively.

Experimental Results

To evaluate the performance of the proposed approaches, a benchmark data set named *MovieLens* is employed. The data set can be downloaded from the website of *GroupLens*: http://www.grouplens.org. This data set consists of 943 users and 1682 movies. It is "5 star" rating system. There are 100,000 ratings in the data set. Among those ratings, 85% are ratings >=3.

We used a 5-folder cross validation approach to evaluate the proposed algorithms. The data set is randomly divided into 5 groups. Each group contains 20k ratings. Every group will be enacted as test set for a single instance. While one group is used for test, the other 4 groups will be considered as a training set, so every rating will be tested once. Based on this data set, first we provide some visual evidence of the clustering method. The clustering results are shown in Figure 3, based on the training data set of the first folder. Here, we display a 0-1 quantization data set. In the image a bright dot means 1 and dark dot means 0. Each row of the image indicates a profile of a user. One can see the algorithm successfully partition the messy original training data set into 5 groups. Users within the same group have some common interests.

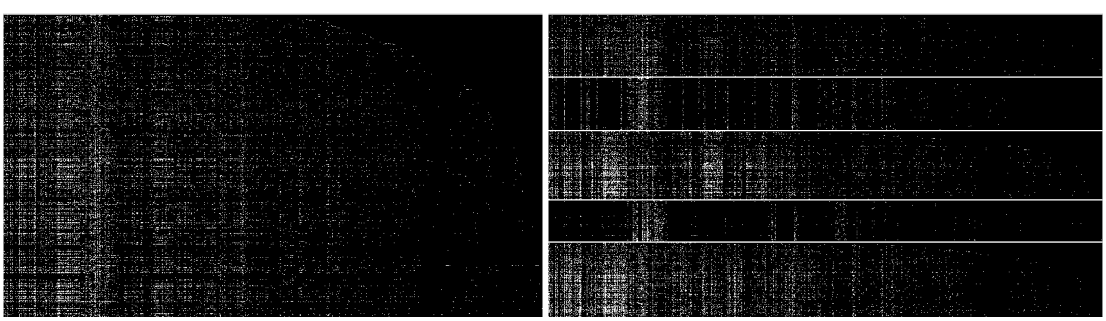

Figure 3. Left is the original dataset while the right is the result after clustering using the proposed method.

Table 1 includes the abbreviations of different algorithms we implemented in the experiments. Here, to avoid complexity, we implement CF* by using the equally quantization method and compared it with CF and CFD.

Based on the training set, a top-N recommendation list will be created for a user. If a movie appears in the top-N, and it also collected by the user in test set with a rating >=3, then we provide a designation as a true positive (TP). If the rating in the test set <3, then we designate it as a false positive (FP). Notice that if the user does not collect the movie in the test set, it is neither a TP nor a FP. The number of TPs and FPs will both increase as the length of recommendation list increase. To make a fair comparison, several efficient metrics are employed.

1. TP rate (recall) is defined as equation 12. P denotes the total number of ratings>=3. FP rate is defined as equation 13. N denotes the total number of ratings<3.

$$recall = \frac{TP}{P} \quad (12) \qquad FP\ rate = \frac{FP}{N} FP\ rate = \frac{FP}{N} \quad (13)$$

2. Based on RP rate and FP rate, receiver operating characteristic (ROC) curves are used to illustrate the performances of the algorithms. FP rate is located on the x-axis while TP rate is on the y-axis. The curve closest to the upper-left corner will represent the best performance, which represents both high recall as well as a low FP rate. If a system recommends all relevant items before non-relevant items, it will draw the perfect curve--a straight upward line from (0,0) to (0,1) and a horizontal line from (0,1) to (1,1).
3. Furthermore, since the data set is highly unbalanced (85% positive, 15% negative), we introduce a metric that is widely applied in bio-informatic area: Mattew's Corellation Coefficients (MCC). MCC is defined as equation 14.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (14)$$

TN refers to true negative and FN means false negative. MCC is between -1 and +1. The higher is the MCC the better is the performance of the system. MCC is effective in dealing with unbalanced data set issues.

Table 1: Abbreviation of different algorithms

Abbreviation	Quantization Method	Clustering	Basic Algorithm
CF	0-1	No	Collaborative Filtering
BG	0-1	No	Bipartite Graph
CFD	Discriminative weights	No	Collaborative Filtering
BGD	Discriminative weights	No	Bipartite Graph
CFC	0-1	Yes	Collaborative Filtering
BGC	0-1	Yes	Bipartite Graph
CFDC	Discriminative weights	Yes	Collaborative Filtering
BGDC	Discriminative weights	Yes	Bipartite Graph
CF*	0.2, 0.4, 0.6, 0.8 and 1	No	Collaborative Filtering

Results from the first experiment show the efficiency of the discriminative weight method (see Figure 3). Here, recommendation length is tested from 10 to 200. BGD and BG get the highest TP rate while CF and CFD get the lowest. Notice that the discriminative weights method does not improve the TP rate significantly. However, the method decreases the FP rate of CFD and BGD dramatically. Also notice that BG and CF* achieve the highest FP rate. Based on ROC curve, one finds that CFD and BGD get better performance compared with other methods since they are closer to the upper-left corner while CF* is the worst. In the final MCC curve map, it is evident that BGD achieves the best performance and CFD provides the second-best performance. Interestingly, CF and BG curves look similar to the MCC curve. Since CF* had the worst performance compared with other methods, we will not perform clustering method based on 1-5 rating quantization.

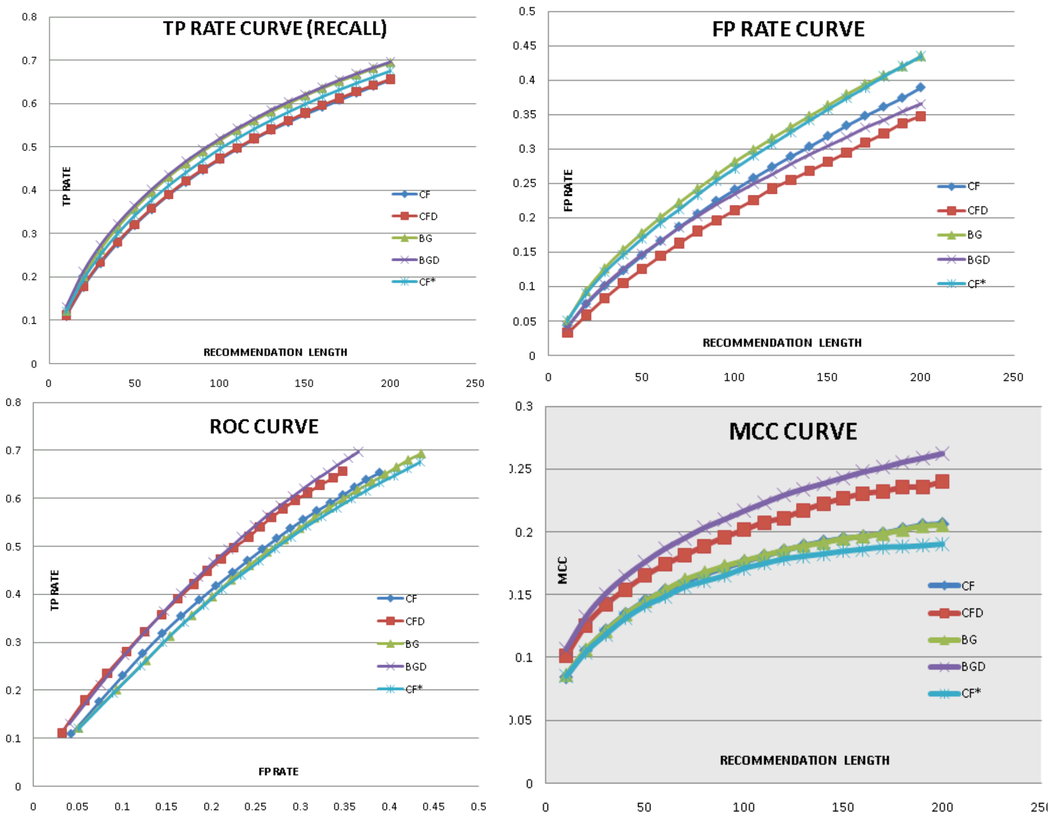

Figure 3. Comparisons between traditional methods and the methods combined with discriminative weights

Figure 4 shows the results of CF and BG based on proposed clustering method. Comparing both ROC curve and MCC curves, BGDC results in the best performance while CFC and BGC improves slightly compared to CF and BG respectively.

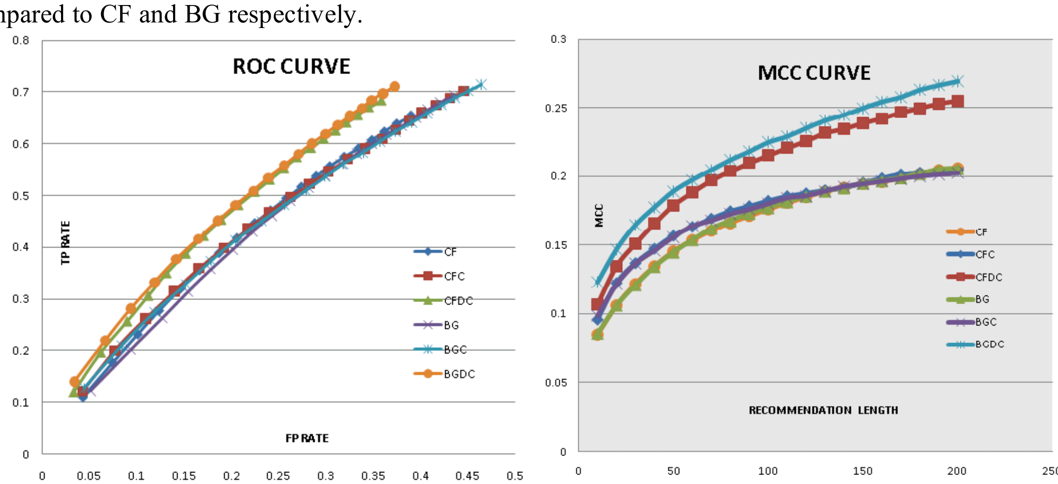

Figure 4. Results of CF and BG based on proposed clustering method

BGDC achieves the best MCC performance while CFDC is the second best. However, CFDC has a relative low FP rate compared with BGDC. Discriminative weights can dramatically decrease the FP rate without changing the TP rate. BG achieves a better TR rate while CF achieves a better FP rate. Equalizing the rating into the same interval is not appropriate because CF* shows the worst performance.

Conclusion

In this paper, we propose a novel method for recommendation systems that when employed with search tools for online educational resources can greatly improve the accuracy of the search. Greater accuracy will result in students and teachers better finding the educational resource they are seeking, in turn improving the effective uses of available educational resources. This method uses combinations of discriminative weights and sparse matrix clustering. Discriminative weights can dramatically decrease the FP rate while retaining the TP rate. The clustering method overcomes the "missing value" problem and groups users with common interests. Based on the experimental results performed on *MovieLens*, the new approaches generate more relevant results while filtering less irrelevant results when compared with the traditional algorithms of the same recommendation length. Researchers are working on systems to support effective use of open educational resources (OERs) including OpenCourseWares (OCWs). Previously, tools have been created to assist in locating and accessing educational resources such as OCW Finder (http://ocwfinder.org) and OER Recommender (http://oerrecommender.org), which have become popular web sites for searching, browsing, and recommending OERs. Those sites index over 110,000 OERs including OERs from the National Science Digital Library, many OCWs, and other resources, including resources in eight languages. Recently, research has begun to add personal recommendation capabilities to the recommender systems. Previously, the system recommended "related resources" based on semantic similarities, and adapted rankings based on click- and time-on-page data, but did not personalize recommendations. New efforts focus on personalizing recommendations to individuals based on metadata gathered about users' attentions. The new system allows a user to register, create a profile, list their bookmark, blog, and comment and share resources with other learners. With data from these sources, in addition to click-and-search metadata, the system will generate personalized recommendations, thereby creating the potential for increasing the effective locating and use of educational resources. Finally, this paper assists in efforts for "learning in the disciplines" by providing informational technology perspectives and insight into search infrastructure for online educational resources.

References

Belkin, N. J. (2000). Helping people find what they don't know. *Communications of the ACM,* 43(8), 58-61.

Brusilovsky, P., Kobsa, A., & Nejdl, W, (Eds.). (2007). *The adaptive web methods and strategies of web personalization*. Berlin: Springer.

Chu, W., & Park, S. T. (2009). Personalized recommendation on dynamic content using predictive bilinear models. In *Proceedings of the 18th international World Wide Web Conference.*

Johnson-Henson, S. (2008). The *folksemantic* web: Tools for a human-centered approach to learning. *International Conference of the Learning Sciences (ICLS) 2008*, Utrecht, Netherlands.

Mobasher, B., & Anand, S. S., (2005). Intelligent techniques for web personalization. In *Intelligent Techniques for Web Personalization*. Bamshad Mobasher and Sarabjot Singh Anand (Eds.), Lecture Notes in Artificial Intelligence (3169) , (pp. 1-37). Springer.

Montaner, M., López, B., & De la Rosa, J. L. (2003). A taxonomy of recommender agents on the Internet. *Artificial Intelligence Review*. 19(4), 285-330.

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., & Riedl, J. (1994). *GroupLens*: An open architecture for collaborative filtering of NetNews. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW)*. ACM Press, New York, 175–186.

Seow, P. (2008). Towards a framework for seamless learning environments. In *Proceedings of the International Conference of the Learning Sciences (ICLS) 2008*. Utrecht, Netherlands.

Si, L., & Jin, R. (2003). Flexible mixture model for collaborative filtering. In *Proceedings of the International Conference on Machine Learning (ICML)*. AAAI Press, DC, 704–711.

Wang, J., de Vries, A. P., & Reinders, M. J. T. (2008). Unified relevance models for rating prediction in collaborative filtering. *ACM Transactions on Information System*. 26(3), 16.

Xue, G.R., Lin, C., Yang, Q., Xi, W., Zeng, H.J., Yu, Y., & Chen, Z. (2005). Scalable collaborative filtering using cluster-based smoothing. In *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM Press, New York, 114–121.

Zhou, T., Ren, J., Medo, M., & Zhang, Y. C. (2007). Bipartite network projection and personal recommendation. *Physics Review E.,* 76.