A Hybrid Approach with Collaborative Filtering for Recommender Systems

Gilbert Badaro[1], Hazem Hajj[1], Wassim El-Hajj[2], Lama Nachman[3]

[1]Electrical Engineering Department
[2]Computer Science Department
American University of Beirut, Lebanon
{ggb05, hh63, we07}@aub.edu.lb

[3]Intel Corporation
Santa Clara, CA 95052 USA
lama.nachman@intel.com

Abstract-- The proliferation of powerful smart devices is revolutionizing mobile computing systems. A particular set of applications that is gaining wide interest is recommender systems. Recommender systems provide their users with recommendations on variety of personal and relevant items or activities. They can play a significant role in today's life whether in E-commerce or for daily decisions that we need to make. We introduce a hybrid approach for solving the problem of finding the ratings of unrated items in a user-item ranking matrix through a weighted combination of user-based and item-based collaborative filtering. The proposed technique provides improvements in addressing two major challenges of recommender systems: accuracy of recommender systems and sparsity of data by simultaneously incorporating users' correlations and items ones. The evaluation of the system shows superiority of the solution compared to stand-alone user-based collaborative filtering or item-based collaborative filtering.

Keywords- Recommender system, Collaborative filtering, Memory-based recommender, Model-based recommender

I. INTRODUCTION

Mobile computing systems and applications are continuously gaining much interest from academia and industry given their impact on people's daily lives and given the technological advances in artificial intelligence, integrated circuitry and processor speeds. Particular applications that are capturing a lot of attention are recommender systems due to their importance in helping people with their life-related decisions such as: what book to read, what movie to watch, what music to listen to, where to eat, which links to visit on the web [1] and several other situations where a decision is required. Moreover, recommender systems are becoming more necessary since the amount of information displayed to the user is getting wider and thus, an automated system can assist the users by displaying what is relevant to their needs [2].

Several techniques have been proposed in the literature for recommender systems, where most rely on the ratings provided by the users on different items. The primary idea for a recommender system is to find the ratings of unrated items and to suggest top recommended items for a given user, given the ratings of different items. The major approaches for recommender systems can be classified into: collaborative techniques, content-based techniques and hybrid models that cross both [3].

In this paper, a new hybrid approach within collaborative filtering is proposed for estimating ratings for unrated items based on a weighted combination of user-based collaborative filtering and item-based collaborative filtering. Given a user-item ranking matrix with unrated items, we aim at finding these ratings and thus, recommend the top ranked items. Unlike existing collaborative filtering techniques that focus on either user-based similarity or item-based similarity, the proposed technique provides the ratings of unrated items based on both aspects of similarity simultaneously, user-user similarity and item-item similarity.

In the rest of the paper, we start by presenting in section II the state of the art techniques adopted for recommender systems and pointing out the limitations they incorporate, the challenges that are currently being faced, and the new applications for recommender systems. In section III, we propose a method that is based on a combination of user-based and item-based collaborative filtering. This method allows for the computation of missing ratings by incorporating user-user and item-item similarities. In section IV, we evaluate our proposed scheme and compare it to state of the art techniques, specifically user-based collaborative filtering and item-based collaborative filtering. In section V, we conclude our paper by summarizing the results achieved.

II. LITERATURE REVIEW

In this section a literature review is conducted on collaborative filtering techniques, user-based and item-based ones, content-based recommender systems, hybrid recommender systems and preference-based recommender systems. We highlight the techniques used and summarizing the challenges of recommender systems.

A. Collaborative Filtering Techniques

Sarwar et al. presents in [4] a technique that makes use of collaborative filtering. This technique assumes a list of

978-1-4673-2480-9/13/$31.00 ©2013 IEEE

m users $U = \{u_1, u_2, \dots, u_m\}$ and a list of n items $I=\{i_1, i_2, \dots, i_n\}$. Each user u_i has rated a list of items noted by I_{ui}. The purpose of this technique is to predict the ratings of unrated items by a given user and recommend the Top-N items.

Two approaches for collaborative filtering, which are mainly user-based and item-based, can be distinguished as follows: user-based collaborative filtering utilizes the similarity computed between the active user and all other users. Item-based collaborative filtering makes use of the similarity available between two items. The similarity measures rely on the ratings available for two queried items. For instance, two users who gave close ratings to the same set of items will most likely have a similarity measure close to 1, whereas two users who have different ratings for the same set of items are more likely to have similarity measure close to 0. The definition of similarity can be found through the Pearson correlation coefficient, the cosine similarity or the adjusted cosine similarity as described in [5]. An example of Pearson correlation coefficient is given in (1). For users u and v, I_{uv} represents the set of items rated by both users u and v, $r_{u,i}$ the rating of user u to item i and $\bar{r}_u$ the average of ratings provided by user u.

$$sim(u,v) = \frac{\sum_{i \in I_{uv}} (r_{u,i} - \bar{r}_u)(r_{v,i} - \bar{r}_v)}{\sqrt{\sum_{i \in I_{uv}} (r_{u,i} - \bar{r}_u)^2 \sum_{i \in I_{uv}} (r_{v,i} - \bar{r}_v)^2}} \quad (1)$$

For each of the two above approaches, user-based and item-based collaborative filtering, we describe two major algorithms: memory-based collaborative filtering algorithms and model-based collaborative filtering algorithms. Memory-based algorithms exploit the entire user-item database to make a prediction. Statistical techniques are employed to find the nearest-neighbors for a user if we consider a user-based approach or for an item if we consider an item-based approach. The predicted rating r is given in (2) based on a user-based approach where v belongs to the user in the neighborhood of the active user u:

$$\hat{r}_{u,i} = \bar{r}_u + \frac{\sum_{v \in N(u)} sim(u,v) * (r_{v,i} - \bar{r}_v)}{\sum_{v \in N(u)} |sim(u,v)|} \quad (2)$$

In order to normalize the result, the mean rating of the user v is subtracted from his/her rating for item i and is divided by the sum of the absolute value of the computed similarities in order to make sure that the predicted rating fall within the rating range, for example 1 to 5. Item-based collaborative filtering was proved to have better accuracy in terms of lower MAE (Mean absolute Error) compared to user-based collaborative filtering [4].

On the other hand, model-based algorithms suggest an item recommendation by first developing a model of user ratings using different machine learning techniques such as Bayesian network, clustering and rule-based approaches [4-5] and thus the collaborative filtering will be treated as a classification problem. An example of the user-based collaborative filtering process can be summarized in Fig. 1.

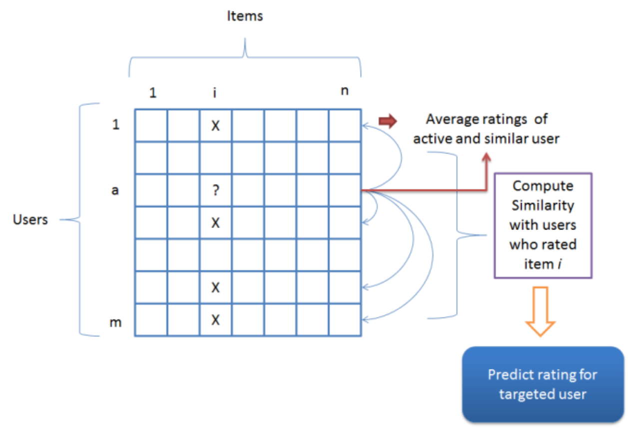

Figure 1. The Collaborative Filtering Process using user-based approach.

In order to improve the accuracy of typical collaborative filtering technique, several methods have been proposed to enhance it such as aggregating external ratings [6], or applying matrix factorization [7]. We now move to over viewing content-based recommender systems.

B. Content-Based Recommender Systems

In [3], the authors describe content based recommendation methods which are inspired from information retrieval and information filtering research. The content-based recommender system looks at the content of a certain item depending on its type and tries to analyze the commonalities among the items that the user has highly rated. Then, based on the analysis, the system will detect items with high degree of similarity to the user's preferences. Typical applications for content-based recommender systems are the ones that deal with items with textual information, such as documents, web sites or movies descriptions.

In a more detailed manner, consider the content of an *item profile s* and a set of attributes characterizing item *s* which represent extracted features from the item s. Hence, the content can be represented as a vector of length k representing the extracted features: $Content(s_j)=(w_{1j}, w_{2j}, \dots, w_{kj})$, where w_{1j} represents the value for the given feature 1 out of the k features. Based on the *item profile* vector, $ContentBasedProfile)=(w_{c1}, w_{c2}, \dots, w_{ck})$ which is related to users is a vector of weights where each weight represents the importance of each feature extracted for a certain user c. Similar to collaborative filtering, similarity measures can be used to find the highest scoring item, or probabilistic techniques can be employed to estimate if a certain item belongs to a certain class of users and thus, the problem becomes a classification problem where one can make use of generative approaches that rely on maximum likelihood. The features of the item will be used to compute the conditional probabilities [3].

C. Hybrid Models

The two techniques described in sections II-A and II-B incur some limitations and can be summarized mainly by four problems. The new user problem makes it hard for the system to learn about the new user preferences especially if he/she did not rate enough items; this is also known as cold-start users [8] and was discussed in [9]. The new item problem is also an issue for collaborative filtering since even if the item has a high rating, the recommender system will not be able to recommend it unless a minimum number of users have rated it. Sparsity of data which means that the number of existing ratings is relatively low compared to the number of users and items on the system. This can impact negatively the accuracy of the recommender system. Finally, content-based recommender systems can sometimes be hard to implement for items where feature extraction is not feasible.

Due to the above limitations, hybrid models were developed as discussed in [3, 10] where the two techniques described in section II-A and II-B are merged in four different ways. The first one works by combining the separate recommenders ratings using a linear combination or a voting scheme which basically selects the recommendation that is seen better in terms of quality and more consistent with past user's ratings. As for the second method, it adds the content-based characteristics to collaborative models which can help in overcoming the sparsity problem since we are not only relying on ratings but also on item profiles for our prediction. The third way is to add collaborative characteristics to content-based models where latent factors are introduced to describe the user preferences. The fourth mode is to develop a single unifying recommendation model based on content-based and collaborative characteristics using probabilistic approaches such as rule-based classifier or Bayesian regression models.

D. Preference-based Recommender Systems

A newly introduced approach for recommender systems is the preference-based one that instead of relying on item ratings provided by the users, it identifies abstract features and relations based on the user profile. For instance, based on the profession of the user, his location, his gender and other preferences, the preference-based recommender system constructs abstract relationships that are relatively more general and can capture a larger set of people with similar preferences extracted from their specific profiles. The authors in [11] introduce a preference-based recommender system for the conference recommendation problem that recommends conference sessions for a new set of users based on user profiles and conference themes constructed using nonnegative matrix factorization. The advantage of the proposed technique is that it does not require item ratings but instead it relies on user behavior and sessions' attendance for this specific case. Other possible types of matrix projection that are used to learn abstract relationship about users or items based on tangible data are Singular Value Decomposition, Principal Component Analysis and Vector Quantification. The preference-based model was adopted by several e-commerce companies such as Amazon.com and Netflix [12].

E. Challenges of Recommender Systems

As the amount of information is increasing tremendously on a daily basis and as the number of E-commerce users is also increasing, challenges for existing recommender systems are being more crucial to tackle. Scalability of recommender systems is currently a major point to achieve in order to accommodate for the increasing number of users/items [13]. Moreover, since recommender systems are heavily based on users' interventions on the web and their opinions, privacy issues should not be violated by recommender systems and this makes it challenging for systems that rely for instance, on the cache of a web-browser [14]. Collection of data is sometimes hard to achieve. Hence, implicit and user-friendly ways are required to avoid having sparse data and simplifying the retrieval of ratings from users. Typical techniques for collecting user input on items are performed when the user is signing up for the first time on the website or through a survey. Last but not least, evaluation metrics are being discussed more recently since sometimes a low root-mean-square-error system does not guarantee a good quality recommendation. In fact, users are becoming more curious in knowing why these specific items were recommended and not others. In the same context, research in recommender systems is now considering recommending new items not only based on preference of the users but also on providing them with diversified items that they did not explore before as described in [15].

F. Recent Applications for Recommender Systems

Recommender systems are being utilized in different contexts. The authors in [16] present a technique for recommending social events based on the user geographical location extracted from his mobile phone device and based on his movements. From social events to travel packages, the authors in [17] address customizing recommender systems to help the users in selecting their travel packages by exploiting online travel information and focusing on the specific characteristics of this information. A Tourist-Area-Season Topic model is developed and is tested for its effectiveness. Social networks have also got their shares of recommender systems [18]. For instance, TWITOBI, a recommendation system for Twitter using probabilistic modeling for collaborative filtering can recommend top-K users to follow and top-K tweets to read for an active user. Web page recommendation is also a hot topic area for recommender systems as discussed in [19] where a graph based iteration algorithm is proposed to discover the topics of interest for each user and a topic-aware Markov model to learn the navigation patterns for each user. Finally, recommender systems are also proposed to solve patent maintenance related issues that can sometimes induce a high cost on companies or patent owners. In [20], the authors model the patents as heterogeneous time-evolving information network and propose new patent features to build a model for a ranked prediction on whether to maintain or abandon a patent.

III. PROPOSED APPROACH

In this paper, we propose a hybrid model that combines simultaneously user-based collaborative filtering and item-based collaborative filtering by adding the predicted ratings from each technique and multiplying them with a weight that incorporates the accuracy of each technique alone. The proposed approach benefits from correlation between not only users alone or items alone but from both simultaneously. Hence, the predicted result will combine two aspects of similarities: user-user similarities and item-item similarities. The rating $\hat{r}$ for an item given a specific user is given as follows:

$$\hat{r} = \alpha\hat{r}_u + \beta\hat{r}_i \quad (3)$$

Where, $\hat{r}_u$ is predicted using user-based collaborative filtering and $\hat{r}_i$ is predicted using item-based collaborative filtering for the same item and user. α and β are weights given to each of the relative ratings $\hat{r}_u$ and $\hat{r}_i$. α and β are both fractions, satisfying the following conditions:

$$\begin{gathered} \alpha + \beta = 1 \\ \alpha \leq 1 \;;\; \beta \leq 1 \end{gathered} \quad (4)$$

The item rating $\hat{r}_i$ can be computed using the following equation:

$$\hat{r}_i = \bar{r}_i + \frac{\sum_{j\in N(i)} sim(i,j) * \left(r_{u,i} - \bar{r}_j\right)}{\sum_{j\in N(i)} |sim(i,j)|} \quad (5)$$

where *sim(i,j)* is the similarity between active item *i* and item *j*, and *N(i)* is the neighborhood of the item *i*. The user rating $\hat{r}_u$ is computed as shown in (2) where *sim(u,v)* is the similarity between user *u* and user *v*, and *N(u)* is the neighborhood of the user *u*.

Once the ratings of the items are predicted for an active user, the Top *k* items are selected based on the highest ratings. Furthermore, the proposed algorithm deals with cold start users by relying on item similarity and with cold start items by relying on user similarity. We can, therefore, overcome the challenge of cold start users/items too, namely the challenge of data sparsity.

A. On the choice of the Weights

To select the optimum values, the weights α and β have to be selected to maximize the accuracy of the resulting MAE evaluated with the combined rating. Alternatively, the choice of the weights needs to minimize the error resulting from the difference between predicted ratings and actual ratings available in training data.

While several measures are possible for assessing the accuracy of the system, we use mean absolute error (MAE) to measure the deviation of recommendations from their true user-specified values. For each rating-prediction pair $<p_i,q_i>$, p_i being the predicted value and q_i the correct value available in the training data, the absolute error is computed as $|p_i - q_i|$. The MAE is then evaluated by examining *N* ratings-prediction pairs, and computing the average error as shown in the equation below:

$$MAE = \frac{\sum_{i=1}^{N} |p_i - q_i|}{N} \quad (6)$$

The lower the MAE the better the accuracy is. As a result, the choice of the weights (α, β) needs to minimize the MAE as shown in (6)

$$\frac{argmin}{\alpha,\beta}\left(\frac{\sum_{i=1}^{N}|p_i - q_i|}{N}\right) \quad (7)$$

To simplify the search space for (α, β), we propose a simplified empirical approach. Based on [4], the accuracy of item-based collaborative filtering was proved to be more accurate than user-based collaborative filtering. As a result, we propose a higher weight, β, be given to the prediction performed by item-based collaborative filtering. Furthermore, since the rating scale in a recommender system consists of integers or decimals with one decimal digit, several cases are considered for the weights α & β and they are represented in Table I. The more we increase β, the more the impact of item similarities is compared to user similarities and vice versa. Fig. 2 shows the different MAE values obtained for the proposed α & β values. We also add a case (α = 2 β, or α=2/3 & β=1/3) to make sure that item-based collaborative filtering is indeed more accurate than user-based collaborative filtering. It is worth noting that the two special cases: (α=1, β=0) and (α=0, β=1) correspond to using user-based collaborative filtering alone and item-based collaborative filtering alone respectively.

TABLE I. Values used for testing the proposed method

Case	α	β
β = α	1/2	1/2
β = 2*α	1/3	2/3
β = 3*α	1/4	3/4
β = 4*α	1/5	4/5
β = 5*α	1/6	5/6
β = 6*α	1/7	6/7
β = 7*α	1/8	7/8
α = 2*β	2/3	1/3

The MAE for each case is computed by making use of different combination of data sets and the optimal weights will be the ones corresponding to the case with the lowest MAE as illustrated in equation (6). Moreover, in order to improve the time performance of the system, a fixed neighborhood size *N* is set. The highest *N* similarity values are selected for each technique, i.e., the user-user similarity measure and the item-item similarity measure. Since the closest users and items are expected to have the biggest impact on accuracy, the impact of choosing only *N* neighbors to perform the calculation of the ratings to be predicted is expected to be negligible on accuracy compared to the gain in performance.

IV. EVALUATION

In order to evaluate the proposed system, experiments are conducted on data selected from MovieLens[21], a web-based research recommender system that debuted in 1997. The data was collected from hundreds of users who visit MovieLens to rate and receive recommendations for movies. Several data sets exist on the site [21], and the 100k ratings was used for the evaluation. The selection of this dataset specifically is made in order to compare our results to user-based collaborative filtering and item-based collaborative filtering performed on the same dataset as described in [4].

The data is stored in text files that we transformed to a user-item matrix. The data is divided into 5 training sets and 5 corresponding testing sets and thus, a 5-fold cross validation approach was applied (i.e. 80% training data and 20% test data) to evaluate our system. The accuracy of the proposed technique was compared to user-based collaborative filtering as stand-alone and item-based collaborative filtering as stand-alone.

We search for the best weights following the proposed values of α & β using an empirical approach by observing the MAE for the different combinations of α & β as described in the previous section and listed in Table I. It was observed that α=1/6 and β=5/6 produced the lowest MAE compared to the other suggested combinations as depicted in Fig. 2. Although the combination α=1/8 & β=7/8 was expected to represent the optimum solution since the weight accorded for item-based collaborative filtering is higher. This behavior can be explained by observing that the relatively reduced α factor hides the intrinsic similarities and relations that can be extracted among users through user-based collaborative filtering. Thus, α=1/6 and β=5/6 were the optimal coefficients as found through empirical analysis.

To test the proposed technique, a neighborhood size of *N* = 20 was used based on [4] where a neighborhood size of 20 was optimal in terms of MAE and performance. For larger neighborhood sizes, no significant improvement was obtained in terms of MAE. The simulation was performed in MATLAB on a Windows 7 with Intel I7 2.4GHz as CPU and 6GB RAM. The results are shown in Fig. 2. 2 where for α=1/6 and β=5/6 we obtain the lowest MAE compared to the other combinations and hence, this is the optimal solution. In order to compare our proposed approach to state of the art techniques, we select the optimum evaluated combination and compare the resulting MAE to the ones measured by using user-based collaborative filtering and item-based collaborative filtering separately. As shown in Fig. 3, the proposed technique gives a better accuracy with an improvement of 23% over user-based collaborative filtering and 16% over item-based collaborative filtering.

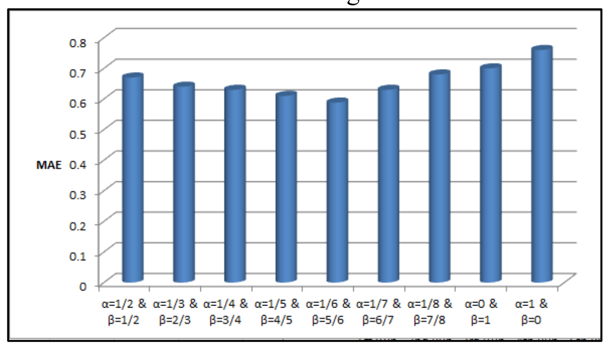

Figure 2. Simulation results for different values of α & β.

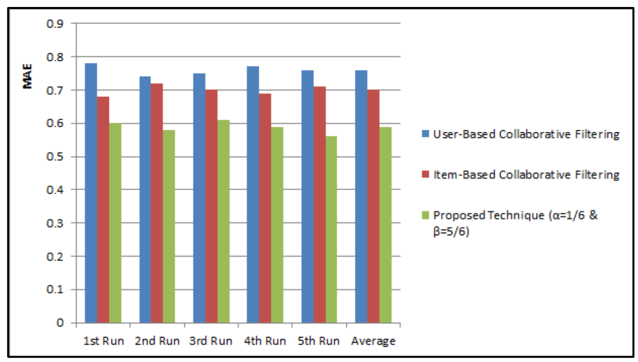

Figure 3. Simulation results for user-based, item-based and our proposed hybrid-base collaborative filtering.

V. CONCLUSION

In this paper, we proposed a new hybrid method for recommender systems based on simultaneous combination of user-based and item-based collaborative filtering. The results showed improvements in accuracy compared to using user-based or item-based collaborative filtering separately. Moreover the proposed technique addresses two common challenges of recommender systems, namely sparsity of data and improved accuracy of recommender system by combining the hidden relations between users and items.

REFERENCES

[1] HengSong Tan; HongWu Ye, "A Collaborative Filtering Recommendation Algorithm Based on Item Classification," *Circuits, Communications and Systems, 2009. PACCS '09 Pacific-Asia Conference*, pp.694-697, May 2009.

[2] Martinez, L.; Rodriguez, R. M.; Espinilla, M., "REJA: A Georeferenced Hybrid Recommender System for

Restaurants," *Web Intelligence and Intelligent Agent Technologies, 2009. WI-IAT '09. IEEE/WIC/ACM International Joint Conferences*, vol.3, pp.187-190, September 2009.

[3] Adomavicius, G.; Tuzhilin, A., "Towards the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions," *IEEE Transactions on Knowledge and Data Engineering 17,* pp.634-749, 2005.

[4] B. Sarwar; G. Karypis, J. Konstan, and J. Riedl. "Item-based Collaborative Filtering Recommendation Algorithms," In Proc. *of the 10th International WWW Conference,* 2001.

[5] T. Hoffman; "Latent Semantic Models for Collaborative Filtering," *ACM transactions on Information Systems,* vol.22, no.1, pp.89-115, 2004.

[6] Umyarov, A.; Tuzhilin, A.; "Improving Collaborative Filtering Recommendations Using External Data," *Data Mining, 2008. ICDM '08. Eighth IEEE International Conference on* , vol., no., pp.618-627, 15-19 Dec. 2008

[7] Chen, Gang; Wang, Fei; Zhang, Changshui; , "Collaborative Filtering Using Orthogonal Nonnegative Matrix Tri-factorization," *Data Mining Workshops, 2007. ICDM Workshops 2007. Seventh IEEE International Conference on*, vol., no., pp.303-308, 28-31 Oct. 2007.

[8] M. Jamali and M. Ester; ICDM 2011, Conference Tutorial, Topic: "Mining Social Networks for Recommendation," Simon Fraser University ICDM 2011*,* Dec. 12, 2011.

[9] Basiri, J.; Shakery, A.; Moshiri, B.; Hayat, M.Z.; "Alleviating the cold-start problem of recommender systems using a new hybrid approach," *Telecommunications (IST), 2010 5th International Symposium on* , vol., no., pp.962-967, 4-6 Dec. 2010.

[10] Claypool, M.; A. Gokhale; T. Miranda; P. Murnikov; D. Netes; M. Sartin, "Combining content-based and collaborative filters in an online newspaper," *ACM SIGIR'99, Workshop on Recommender Systems: Algorithms and Evaluation*, vol., no., Aug. 1999.

[11] Hornick, M.F.; Tamayo, P.;, "Extending Recommender Systems for Disjoint User/Item Sets: The Conference Recommendation Proble," *Knowledge and Data Engineering, IEEE Transactions on,* vol.24, no.8, pp.1478-1490, Aug. 2012

[12] Konstan, J.A.; Riedl, J.; , "Recommended for you," *Spectrum, IEEE,* vol.49, no.10, pp.54-61, October 2012.

[13] "ACM Recommender Systems": http://recsys.acm.org/2012/call_for_papers.html, May 25, 2012 [Apr. 21 2012]

[14] Armknecht, F.; Strufe, T., "An efficient distributed privacy-preserving recommendation system," *Ad Hoc Networking Workshop (Med-Hoc-Net), 2011 The 10th IFIP Annual Mediterranean*, pp.65-70, 12-15 June 2011.

[15] Jinoh Oh; Sun Park; Hwanjo Yu; Min Song; Seung-Taek Park; , "Novel Recommendation Based on Personal Popularity Tendency," *Data Mining (ICDM), 2011 IEEE 11th International Conference on* , vol., no., pp.507-516, 11-14 Dec. 2011.

[16] Quercia, D.; Lathia, N.; Calabrese, F.; Di Lorenzo, G.; Crowcroft, J., "Recommending Social Events from Mobile Phone Location Data," *Data Mining (ICDM), 2010 IEEE 10th International Conference on* , vol., no., pp.971-976, 13-17 Dec. 2010.

[17] Quercia, D.; Lathia, N.; Calabrese, F.; Di Lorenzo, G.; Crowcroft, J., "Recommending Social Events from Mobile Phone Location Data," *Data Mining (ICDM), 2010 IEEE 10th International Conference on* , vol., no., pp.971-976, 13-17 Dec. 2010.

[18] Qi Liu; Yong Ge; Zhongmou Li; Enhong Chen; Hui Xiong, "Personalized Travel Package Recommendation," *Data Mining (ICDM), 2011 IEEE 11th International Conference on* , vol., no., pp.407-416, 11-14 Dec. 2011.

[19] Qingyan Yang; Ju Fan; Jianyong Wang; Lizhu Zhou; , "Personalizing Web Page Recommendation via Collaborative Filtering and Topic-Aware Markov Model," *Data Mining (ICDM), 2010 IEEE 10th International Conference on* , vol., no., pp.1145-1150, 13-17 Dec. 2010.

[20] Xin Jin; Spangler, S.; Ying Chen; Keke Cai; Rui Ma; Li Zhang; Xian Wu; Jiawei Han; , "Patent Maintenance Recommendation with Patent Information Network Model," *Data Mining (ICDM), 2011 IEEE 11th International Conference on* , vol., no., pp.280-289, 11-14 Dec. 2011.

[21] "MovieLens Data Sets": http://www.grouplens.org/node/73, May 5, 2012 [Aug. 24 2011]