An Enhanced Hybrid Approach towards Recommender Systems

Vakkalanka Aparna[1], B.Hanumanthu[2]
[1]Student, Master of Technology, KITS, Warangal
[2]Assistant Professor,Department of CSE, KITS, Warangal

Abstract

A specific set of utilizations that is increasing wide interest is recommender systems. Recommender systems provide their users with recommendations on variety of personal and relevant items or activities. They can assume a critical role in the present life whether in E-commerce or for every day decisions that we need to make. We introduce a cross breed approach for taking care of the problem of finding the evaluations of unrated items in a user-item positioning framework through a weighted blend of user-based and item-based collaborative filtering. The proposed technique provides improvements in addressing two noteworthy challenges of recommender systems: precision of recommender systems and sparsity of information by simultaneously consolidating users' correlations and items ones. The evaluation of the system indicates superiority of the arrangement compared to remain solitary user-based collaborative filtering or item-based collaborative filtering.

Index Terms: Collaborative filtering, Recommender system, Model-based recommender, Memory-based recommender

I. INTRODUCTION

Mobile computing systems and applications are constantly increasing much interest from academia and industry given their effect on people's every day lives and given the technological advances in manmade brainpower, integrated hardware and processor speeds. Specific applications that are catching a great deal of attention are recommender systems due to their importance in helping people with their life-related decisions, for example, what book to read, what movie to watch, what music to listen to, where to eat, which connects to visit on the web [1] and several other circumstances where a decision is required.

The essential idea for a recommender system is to discover the appraisals of unrated items and to suggest top recommended items for a given user, given the evaluations of different items.

In this paper, a new hybrid approach inside collaborative filtering is proposed for estimating evaluations for unrated items based on a weighted blend of user-based collaborative filtering and item-based collaborative filtering. Given a user-item positioning framework with unrated items, we go for finding these appraisals and in this way, recommend the best ranked items. The proposed technique provides the appraisals of unrated items based on the two aspects of closeness simultaneously, user-user similitude and item-item comparability.

In the rest of the paper, we begin by presenting in section II the state of the craftsmanship techniques adopted for recommender systems and calling attention to the restrictions they incorporate. In section III, we propose a method which depends on a blend of user-based and item-based collaborative filtering. This method takes into account the calculation of missing evaluations by joining user-user and item-item similarities. In section IV, we evaluate our proposed scheme and compare it to state of the craftsmanship techniques, specifically user-based collaborative filtering and item-based collaborative filtering. In section V, we final our paper by outlining the results achieved.

II. LITERATURE REVIEW

In this area a survey is led on collaborative filtering techniques, client based and thing based ones, content-based recommender systems, half breed recommender systems and inclination based recommender systems. We include the techniques utilized and consolidating the difficulties of recommender systems.

A. Techniques of Collaborative Filtering

Sarwar et al. presents in [4] a technique that makes use of collaborative filtering. This technique assumes a rundown of m users $U = \{u_1, u_2, \ldots, u_m\}$ and a rundown of n items$I=\{i_1, i_2, \ldots, i_n\}$. Each user ui has rated a rundown of items noted by I_{ui}. Two approaches for collaborative filtering, which are fundamentally user-based and item-based, can be distinguished as takes after: user-based collaborative filtering utilizes the similitude computed between the active user and every other user. For instance, two users who gave close evaluations to the same set of items will in all likelihood have a similitude measure close to 1, whereas two users who have different appraisals for the same set of items are more likely to have closeness measure close to 0. The definition of closeness can be found through the Pearson correlation coefficient, the cosine similitude or the adjusted cosine likeness as described in [5]. An example of Pearson correlation coefficient is given in. For users*u* and *v*, I_{uv} represents the set of items rated by the two users *u* and *v*, $r_{u,i}$ the rating of user u to item I and the average of evaluations provided by user u. be treated as an order problem.

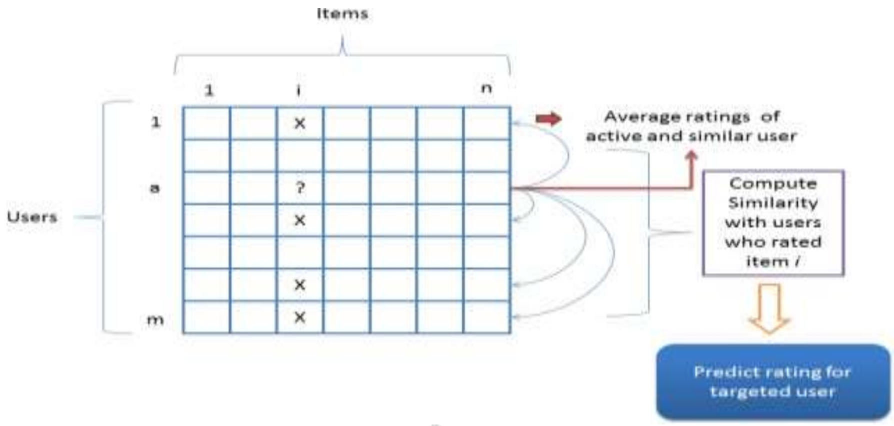

Figure 1. Using user-based approach for Collaborative Filtering Process.

In order to improve the exactness of run of the mill collaborative filtering technique, several methods have been proposed to enhance it, for example, aggregating external evaluations [6]. We now move to over viewing content-based recommender systems.

B. Content-Based Recommender Systems

For each of the two above approaches, user-based and item-based collaborative filtering, we describe two noteworthy calculations: model-based collaborative filtering calculations and memory-based collaborative filtering calculations. Memory-based calculations exploit the entire user-item database to make a prediction. Measurable techniques are employed to discover the nearest-neighbors for a user in the event that we consider a user-based approach or for an item in the event that we consider an item-based approach. The content-based recommender system takes a gander at the content of a certain item depending on its type and tries to analyze the commonalities among the items that the user has very rated. Then, based on the investigation, the system will detect items with high degree of closeness to the user's preferences.

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 aggregate of the absolute value of the computed similarities in order to make sure that the predicted rating fall inside the rating range.

Then again, model-based calculations suggest an item recommendation by first developing a model of user appraisals utilizing different machine learning techniques, for example, Bayesian network, clustering and rule-based approaches and in this manner the collaborative filtering will representing the extracted features: $Content(s_j)=(w_{1j},w_{2j},\ldots,w_{kj})$, where w_{1j} represents the value for the given feature 1 out of the k features. Based on the item profile vector, $ContentBasedProfile)=(w_{c1},w_{c2},\ldots,w_{ck})$ 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. Like collaborative filtering, comparability measures can be used to locate the highest scoring item, or probabilistic techniques can be employed.

C. Hybrid Models

The two techniques described in sections II-An and II-B bring about some confinements and can be summarized basically by four problems. The new user problem makes it difficult for the system to learn about the new user preferences especially on the off chance that he/she didn't rate enough items; this is otherwise called frosty begin users [8] and was discussed in [9]. The new item problem is likewise an issue for collaborative filtering since even if the item has a high evaluating, the recommender system won't be able to recommend it unless a base number of users have rated it. This can affect negatively the exactness of the recommender system.

Due to the above restrictions, hybrid models were developed as discussed in [10] where the two techniques described in section II-An and II-B are merged in four different ways. The first works by joining the separate recommenders appraisals utilizing a linear blend or a voting scheme which fundamentally selects the recommendation that is seen better in terms of value and more consistent with past user's evaluations. With respect to the second method, it will add the content-based characteristics to collaborative models. The third route is to add collaborative characteristics to content-based models. The fourth mode is to develop a single bringing together recommendation model based on content-based and collaborative characteristics utilizing probabilistic approaches, for example, 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 evaluations provided by the users, it identifies dynamic features and relations based on the user profile. For instance, based on the profession of the user, his area, his gender and other preferences, the preference-based recommender system builds theoretical relationships. The advantage of the proposed technique is that it does not require item appraisals but rather instead it relies on user behavior and sessions' attendance for this specific case. Other possible types of lattice projection that are used to learn theoretical relationship about users or items based on tangible information are Singular Value Decomposition, Principal Component Analysis and Vector Quantification. The preference-based model was adopted by several e-commerce companies, for example, Amazon.com and Netflix [12].

E. Challenges of Recommender Systems

As the measure of information is increasing tremendously every day and as the number of E-commerce users is likewise increasing, challenges for existing recommender systems are being more urgent to tackle. Versatility 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 suppositions, security issues ought 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 information is sometimes difficult to achieve. Hence, certain and user-friendly ways are required to abstain from having sparse information and streamlining the retrieval of appraisals from users. Commonplace techniques for collecting user contribution on items are performed when the user is agreeing to accept the first time on the website or through a survey. To wrap things up, evaluation metrics are being discussed more recently since sometimes a low root-mean-square-error system does not guarantee a decent quality recommendation. Truth be told, users are becoming more inquisitive in knowing why these specific items were recommended and not others. In the same context, research in recommender systems is currently considering recommending new items based on preference of the users as well as on giving them diversified items that they didn't 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 get-togethers based on the user geographical area extracted from his mobile phone device and based on his movements. From get-togethers to travel packages, the authors in [17] address modifying recommender systems to help the users in selecting their travel packages by exploiting online travel information and concentrating on the specific characteristics of this information. A Tourist-Area-Season Topic model is developed and is tested for its effectiveness. Informal organizations have likewise got their shares of recommender systems [18]. For instance, TWITOBI, a recommendation system for Twitter utilizing probabilistic modeling for collaborative filtering can recommend top-K users to take after and top-K tweets to read for an active user. Webpage recommendation is likewise a hotly debated issue area for recommender systems as discussed in [19] where a chart based iteration algorithm is proposed to discover the subjects of interest for each user and a point aware Markov model to learn the route patterns for each user. At long last, recommender systems are additionally 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 assemble a model for a ranked prediction on whether to keep up 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 including the predicted appraisals from each technique.

On the choice of the Weights

To select the optimum values, the weights α and β have to be selected to maximize the precision 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 appraisals and real evaluations available in preparing information.

While several measures are possible for assessing the precision of the system, we use mean absolute error (MAE) to measure the deviation of recommendations from their true user-specified values. For each appraising prediction combine $<p_i,q_i>$, pi being the predicted value and qi the correct value available in the preparation information, the absolute error is computed as $|p_i-q_i|$.

The proposed approach benefits from correlation between users alone or items alone as well as from both simultaneously. Hence, the predicted result will combine two aspects of similarities: user-user similarities and item-item similarities. The rating similarities is compared to user similarities and vice versa. Fig. 2 demonstrates the different MAE values obtained for the proposed α and β values. We additionally include a case (α = 2 β, or α=2/3 and β=1/3) to make sure that item-based where sim(i,j) is the likeness between active item I and item j, and N(i) is the neighborhood of the item I. The user rating is computed where sim(u,v) is the likeness between user u and user v, and N(u) is the neighborhood of the user u.

Once the evaluations of the items are predicted for an active user, the Top k items are selected based on the highest appraisals. Furthermore, the proposed algorithm deals with frosty begin users by relying on item closeness and with cool begin items by relying on user comparability. We can, therefore, overcome the challenge of icy begin users/items as well, namely the challenge of information sparsity. collaborative filtering is indeed more accurate than user-based collaborative filtering. It is worth noticing that the two special cases: (α=1, β=0) and (α=0, β=1) correspond to utilizing user-based collaborative filtering alone and item-based collaborative filtering alone respectively.

TABLE I. Values used for testing the proposed method

Case	α	β
$\beta = \alpha$	1/2	1/2
$\beta = 2*\alpha$	1/3	2/3
$\beta = 3*\alpha$	1/4	3/4
$\beta = 4*\alpha$	1/5	4/5
$\beta = 5*\alpha$	1/6	5/6
$\beta = 6*\alpha$	1/7	6/7
$\beta = 7*\alpha$	1/8	7/8
$\alpha = 2*\beta$	2/3	1/3

The MAE for each case is computed by making use of different mix of informational indexes and the ideal weights will be the ones corresponding to the case with the lowest MAE. Moreover, in order to improve the time performance of the system, a fixed neighborhood size N is set. The highest N comparability values are selected for each technique, i.e., the user-user likeness measure and the item-item closeness measure. Since the closest users and items are expected to have the biggest effect on exactness, the effect of picking just N neighbors to perform the estimation of the evaluations to be predicted is expected to be negligible on precision compared to the pick up in performance.

IV. EVALUATION

In order to evaluate the proposed system, experiments are conducted on information selected from MovieLens[21], a web-based research recommender system that debuted in 1997. The information was collected from hundreds of users who visit MovieLens to rate and receive recommendations for movies. Several informational collections exist on the site [21], and the 100k appraisals 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 information is stored in text files that we transformed to a user-item framework. The information is divided into 5 preparing sets and 5 corresponding testing sets and in this way, a 5-crease cross approval approach was applied (i.e. 80% preparing information and 20% test information) to evaluate our system. The exactness 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 and β utilizing an empirical approach by observing the MAE for the different blends of α and β as described in the previous section and listed in Table I. It was observed that $\alpha=1/6$ and $\beta=5/6$ produced the lowest MAE compared to the other suggested mixes as depicted in Fig. 2. In spite of the fact that the mix $\alpha=1/8$ and $\beta=7/8$ was expected to represent the optimum arrangement 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 inborn similarities and relations that can be extracted among users through user-based collaborative filtering. In this way, $\alpha=1/6$ and $\beta=5/6$ were the ideal coefficients as found through empirical investigation.

To test the proposed technique, a neighborhood size of $N = 20$ was used based on [4] where a neighborhood size of 20 was ideal in terms of MAE and performance. For larger neighborhood sizes, no huge improvement was obtained in terms of MAE. The recreation was performed in MATLAB on a Windows 7 with Intel I7 2.4GHz as CPU and 6GB RAM. The results are appeared in Fig. 2. 2 where for $\alpha=1/6$ and $\beta=5/6$ we get the lowest MAE compared to the other mixes and hence, this is the ideal arrangement. In order to compare our proposed approach to state of the workmanship techniques, we select the optimum evaluated mix and compare the resulting MAE to the ones measured by utilizing user-based collaborative filtering and item-based collaborative filtering separately. As appeared in Fig. 3, the proposed technique gives a better exactness 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.

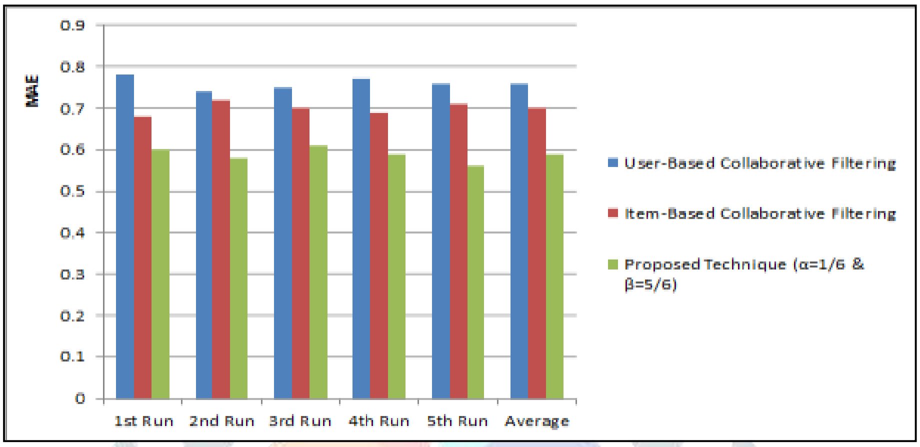

Figure 3. Simulation results for user-based, item-based and our proposed hybrid-base collaborative filtering.

V. CONCLUSION

In this paper, we proposed a new hybrid method for recommender systems based on simultaneous blend of user-based and item-based collaborative filtering. The results showed improvements in exactness compared to utilizing user-based or item-based collaborative filtering separately. Moreover the proposed technique addresses two normal challenges of recommender systems, namely sparsity of information and improved exactness of recommender system by joining 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 tenth International WWW Conference, 2001.
[5]T. Hoffman; "Latent Semantic Models for Collaborative Filtering," ACM exchanges on Information Systems, vol.22, no.1, pp.89-115, 2004.
[6]Umyarov, A.; Tuzhilin, An.; "Enhancing 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: "Digging Social Networks for Recommendation," Simon Fraser University ICDM 2011, Dec. 12, 2011.
[9]Basiri, J.; Shakery, A.; Moshiri, B.; Hayat, M.Z.; "Alleviating the cool begin problem of recommender systems utilizing a new hybrid approach," Telecommunications (IST), 2010 fifth 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, "Consolidating 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 protection preserving recommendation system," Ad Hoc Networking Workshop (Med-Hoc-Net), 2011 The

tenth 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 eleventh 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 tenth 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 tenth 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 eleventh 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 by means of Collaborative Filtering and Topic-Aware Markov Model," Data Mining (ICDM), 2010 IEEE tenth 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 eleventh 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]