IJCSC International Journal of Computer Science & Communication Vol. 1, No. 2, July-December 2010, pp. 429-432

Improvement of Personalized Recommendation Algorithm based on Hybrid Collaborative Filtering

Boddu Raja Sarath Kumar[1], Barre John Ratnam[2] & Maddali Surendra Prasad Babu[3]

[1]CSE Department, Lenora College of Engineering, Rampachodavaram-533288, Andhra Pradesh, India
[2]Lenora College of Engineering, Rampachodavaram-533288, Andhra Pradesh, India
[3]CS&SE Department, Andhra University College of Engineering, Visakhapatnam, India
Email: iamsarathphd@gmail.com

ABSTRACT

The explosive growth and availability of data on the internet has caused information overload. Searching for a query is not easy in the sources of information available for the interest of an individual user. Collaborative filtering systems recommend items based upon opinions of people with similar tastes. Collaborative filtering overcomes some difficulties faced by traditional information filtering by eliminating the need for computers to understand the content of the items. Further, collaborative filtering can also recommend articles that are not similar in content to items rated in the past as long as like-minded users have rated the items. Collaborative filtering (CF) is one of the most frequently used techniques in personalized recommendation systems. But currently used CF techniques are based on item rating prediction. We proposed an improved personalized recommended CF algorithm. Hybrid recommender systems or content-boosted technologies are quickly produce high quality recommendations. We have explored content-boosted CF technique which analyzes the user-item matrix to identify relationships between different items, and then use these relationships to indirectly compute recommendations for users. In this paper we analyze different Memory - based CF and Model-based CF techniques. Finally, we experimentally evaluate our results and compare them. The testing results show that in most cases, the improved algorithm that we put forward can improve recommendation quality.

Keywords: Collaborative Filtering, Personalized Recommendation Algorithm, Content-boosted Filtering

1. INTRODUCTION

Recent years have seen the explosive growth in the amount of information available through internet. www has created the world as global village. Collaborative filtering is a general approach to personalized information filtering and automates the process of recommending items to a user based upon the opinions of people with similar tastes. In most cases, the filtering system determines which users have similar tastes by using standard formulae for computing statistical correlations.

Many collaborative filtering techniques use a form of weighted average to determine a prediction for a user. These techniques use the correlation as the weights. Every user receives a prediction for all items and submits a rating of how well she likes an item after reading it. This feedback given by her is used along with similar feedback from other users to calculate a rating prediction.

CF systems work by collecting user feedback in the form of ratings for items in a given domain and exploit similarities and differences among profiles of several users in determining how to recommend an item. On the other hand, content-based methods provide recommendations by comparing representations of content contained in an item to representations of content that interests the user. Content-based methods can uniquely characterize each user, but CF still has some key advantages.

In this paper, we present the framework for this new hybrid approach, Content-Boosted Collaborative Filtering or Hybrid Collaborative Filtering (HCF). We apply this framework in the domain of movie recommendation and show that our approach performs better than both pure CF and pure content-based systems. In traditional collaborative filtering algorithms include user-based collaborative filtering algorithm and collaborative filtering algorithm based on item rating prediction. User-based collaborative filtering algorithm produces recommendation list for object user according to the view of other users. It is based on these assumptions: if the ratings of some items rated by some users are similar, the rating of other items rated by these users will also be similar. Collaborative filtering recommendation system uses statistical techniques to search the nearest neighbors of the object user and then basing on the item rating rated by the nearest neighbors to predict the item rating rated by the object user, and then produce corresponding recommendation list.

2. HYBRID COLLABORATIVE FILTERING (HCF)

HCF systems combine CF with other recommendation techniques to make predictions or recommendations.

Hoping to avoid limitations of either recommender system and improve recommendation performance, HCF recommenders are combined by adding content-based characteristics to CF models, adding CF characteristics to content-based models, combining CF with content-based or other systems, or combining different CF algorithms.

3. IMPROVED ALGORITHM

- Retrieve all the items.
- Retrieve all the user-item rating.

For each user and user-item rating:

- Compute the set which have been rated by user.
- Compute the set which have not been rated by user.
- Compute the predicted rating of item rated by user.

Further:

- Select several users for each item and put them in array.
- Select suited books and store it.

On the basis of this system, we respectively use user-based CF recommendation algorithm based on item rating prediction and the improved CF algorithm to realize personalized recommendation, the method adopted and analysis results are as the following.

4. EXPERIMENTS

We realize that to test content based collaborative filtering algorithm, it is essential to be able to get both the ratings of real users and the content of the articles they rate. The EachMovie database has data only on the ratings of the movies seen by the users.. Movies are represented only by article numbers and not names. This makes it unusable for our experiments on the integration of content and collaborative filtering experiments. This system allows the users to rate these articles and collects the ratings given. We will use this system to collect the ratings given by the user and the predictions computed by the system.

Dataset

We used the following datasets as test data.

EachMovie data Set

The EachMovie service was part of a research project at the DEC Systems Research Center. 72,916 users entered numeric ratings for 1,628 movies with ratings from 0 to 1.

Jester Joke Data Set

Jester is a web based online joke recommendation system, which has been developing at University of California, Berkeley. This data has 73,421 users collected between April 1999 and May 2003 with a rating from -10 to +10.

Book Crossing Data Set

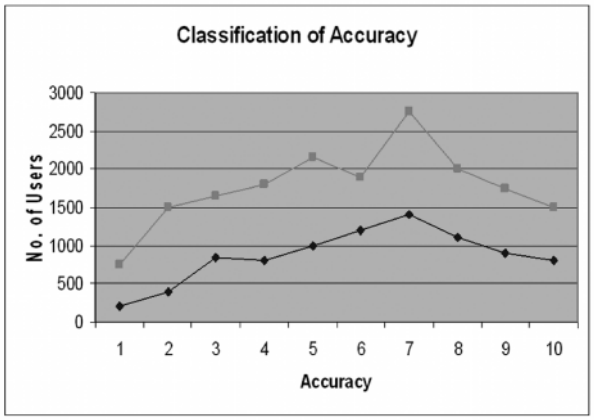
The BookCrossing dataset was collected by Cai-Nicolas Ziegler in a 4-week crawl (August/September 2004) from the Book-Crossing community with kind permission from Ron Hornbaker, CTO of Humankind Systems. It contains 278,858 users providing 1,149,780 ratings about 271,379 books.

Movie Lens Data Set

10 million ratings and 100,000 tags for 10681 movies by 71567 users.

5. RESULTS

Figure 1 shows the classification accuracy of our experiment with EachMovie dataset. Note that since there are 72,916 we limited it to 2000 users. The results show that the maximum accuracy of 72.3%. It seems to be sensitive to the number of selected features and it is getting worse in proportion to the number of features significantly.



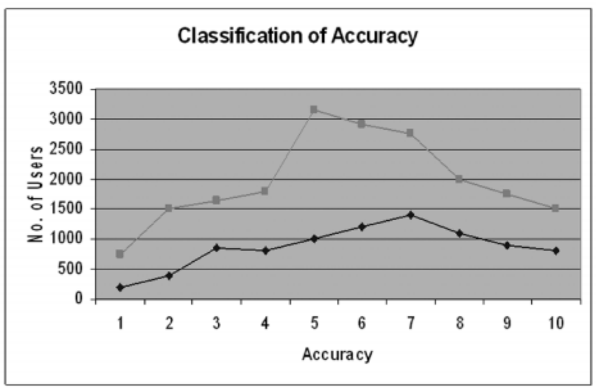


Figure 2 shows the classification accuracy of our experiment with jester joke data set with 73421 users but we limited it to 2000 users. The results shows that the maximum accuracy of 56%.

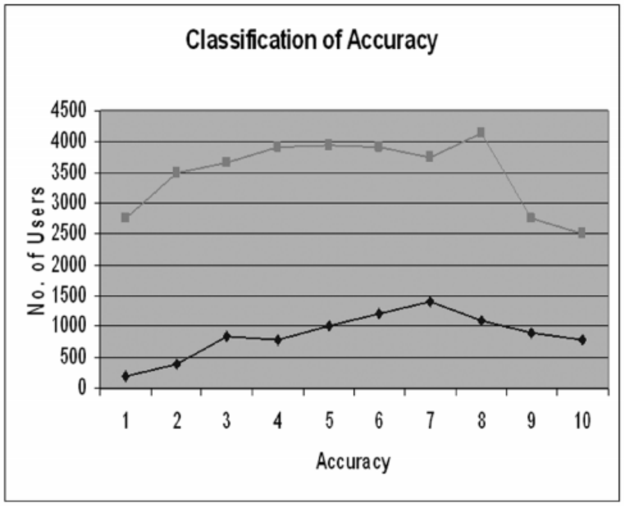


Figure 3 shows the classification accuracy of our experiment with book crossing data set with 278858 users but we limited it to 2000 users. The results shows that the maximum accuracy of 78%.

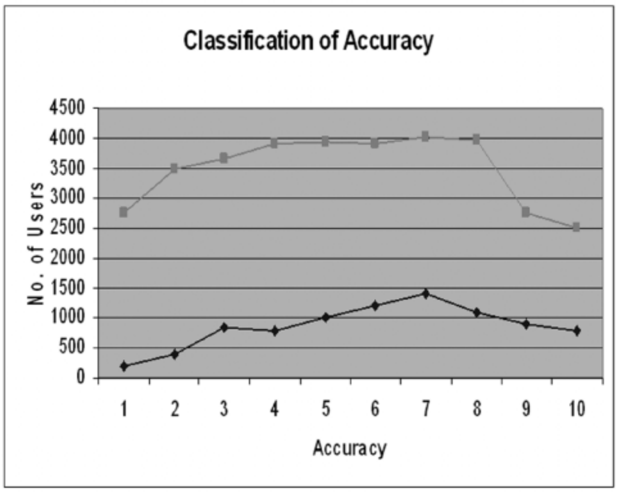


Figure 4 shows the classification accuracy of our experiment with movie lens data set with 71567 users but we limited it to 2000 users. The results shows that the maximum accuracy of 73%.

CONCLUSION

Collaborative filtering (CF) is one of the most successful recommender techniques. Broadly, there are many memory-based CF techniques, but hybrid CF techniques combine CF methods with content-based techniques or other recommender systems to alleviate shortcomings of either system and to improve prediction and recommendation performance. Besides improved performance, hybrid CF techniques rely on external content information that is usually not available, and they generally have increased complexity. There are many evaluation metrics for CF techniques. Due to the nature of our hybrid approach, we believe that improving the performance of the individual components would almost certainly improve the performance of the whole system. Although HCF performs consistently better than pure CF, the difference in performance is not very large. The performance of our system can be boosted by using the methods described earlier.

REFERENCES

[1] Chun Zeng, Chunxiao Xing, and Lizhu Zhou, "A Survey of Personalization Technology",*Journal of Software*, **13**, pp.1952 -1961, 2002.

[2] Yan Zeng, Yonghao Mai, "Collaborative Filtering Recommen-dation Based on Content and Item Rating Prediction", *Computer Applications*, **24**, pp.111-113, 2004.

[3] Ailin Deng, Yangyong Zhu, and Baile Shi, "A Collaborative Filtering Recommendation Algorithm Based on Item Rating Prediction", *Journal of Software*, **13**, pp. 1621-1628, 2003.

[4] Li YU, "Research on Personalized Recommendations in E-business", *Computer Integrated Manufacturing Systems*, **10**, pp. 1306-1313, 2004.

[5] Xiuyan Gu, Linfeng Jiang, and Ziyi Zhang, "Study on User's Browse Behavior to Measure the User's Browse Interest", *Network and Communication*, **15**, pp.43-45, 2005.

[6] Huihong Zhou, Yijun Liu, Weiqing Zhang, and Junyuan Xie, "A Survey of Recommender System Applied in E-commerce", *Computer Application Research*, **1**, pp.8-12, 2004.

[7] Zhi Zhao, Bing Shi, "An Adaptive Algorithm for Personal Recommendation", *Journal of Changchun University*, **15**, pp.26-29, 2005.

[8] G. Adomavicius and A. Tuzhilin, "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**, 2005, 634–749.

[9] R. Bell and Y. Koren, "Scalable Collaborative Filtering with Jointly Derived Neighborhood Interpolation Weights", *IEEE International Conference on Data Mining (ICDM'07)*, pp. 43–52, 2007.

[10] R. Bell, Y. Koren and C. Volinsky, "Modeling Relationships at Multiple Scales to Improve Accuracy of Large Recommender Systems", *Proc. 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2007.

[11] J. Bennet and S. Lanning, "The Netflix Prize", *KDD Cup and Workshop*, *2007*, www.netflixprize.com.

[12] D. Blei, A. Ng, and M. Jordan, "Latent Dirichlet Allocation", *Journal of Machine Learning Research,* **3,** 2003, 993–1022.

[13] M. Deshpande, G. Karypis, "Item-based top-N Recommendation Algorithms", *ACMTrans. Inf. Syst.*, **22** (2004), 143-177.

[14] S. Funk, "Netflix Update: Try This At Home", http://sifter.org/˜simon/journal/20061211.html, 2006.

[15] D. Goldberg, D. Nichols, B. M. Oki and D. Terry, "Using Collaborative Filtering to Weave an Information Tapestry", *Communications of the ACM 35 (1992)*, 61–70.

[16] J. L. Herlocker, J. A. Konstan, A. Borchers and John Riedl, "An Algorithmic Framework for Performing Collaborative Filtering", *Proc. 22nd ACM SIGIR Conference on Information Retrieval*, pp. 230–237, 1999.

[17] J. L. Herlocker, J. A. Konstan, and J. Riedl. "Explaining Collaborative Filtering Recommendations", *In Proceedings of the 2000 ACM Conference on Computer Supported CooperativeWork*, ACMPress, pp. 241-250, 2000.

[18] T. Hofmann, "Latent Semantic Models for Collaborative Filtering", *ACM Transactions on Information Systems,* **22** (2004), 89–115.

[19] Z. Huang, D. Zeng and H. Chen, "A Comparison of Collaborative-Filtering Recommendation Algorithms for E-commerce", *IEEE Intelligent Systems*, **22**(2007), 68–78.

[20] G. Linden, B. Smith and J. York, "Amazon.com Recommendations: Item-to-item Collaborative Filtering", *IEEE Internet Computing,* **7**(2003), 76–80.

[21] D.W. Oard and J. Kim, "Implicit Feedback for Recommender Systems", *Proc. 5th DELOS Workshop on Filtering and Collaborative Filtering*, pp. 31–36, 1998.

[22] A. Paterek, "Improving Regularized Singular Value Decomposition for Collaborative Filtering", *Proc. KDD Cup and Workshop*, 2007.

[23] R. Salakhutdinov, A. Mnih and G. Hinton, "Restricted Boltzmann Machines for Collaborative Filtering", *Proc. 24th Annual International Conference on Machine Learning*, pp. 791–798, 2007.

[24] R. Salakhutdinov and A. Mnih, "Probabilistic Matrix Factorization", *Advances in Neural Information Processing Systems,* **20** (NIPS'07), pp. 1257–1264, 2008.

[25] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. Riedl,"Application of Dimensionality Reduction in Recommender System – A Case Study", WEBKDD'2000.

[26] Book-Crossing Dataset http://www.informatik.uni-freiburg.de/~cziegler/BX/.

[27] Bhaskar Mehta, *Unsupervised Shilling Detection for Collaborative Filtering*, AAAI, 2007, pp. 1402–1407.

[28] Bhaskar Mehta, Thomas Hofmann, and Peter Fankhauser, Lies and Propaganda: detecting Spam Users in Collaborative Filtering, IUI '07: Proceedings of the 12th International Conference on Intelligent user Interfaces (New York, NY, USA), ACM Press, 2007, pp. 14–21.

[29] Bhaskar Mehta, Thomas Hofmann, and Wolfgang Nejdl, Robust Collaborative Filtering, In Proceedings of the 1st ACM Conference on Recommender Systems (Joseph A. Konstan, John Riedl, and Barry Smyth, eds.), ACM Press, October 2007, pp. 49–56.

[30] Bhaskar Mehta and Wolfgang Nejdl, "Attack-resistant Collaborative Filtering", *To appear: In Proceedings of the 31st ACM SIGIR Conference*, ACM Press, 2008.

[31] H. Boley, "Object-Oriented RuleML : User-Level Roles, URIGrounded Clauses, and Order-Sorted Terms", *In Proc. Rules and Rule Markup Languages for the SemanticWeb (RuleML-2003)*, Sanibel Island, Florida, LNCS 2876, Springer-Verlag, Oct. 2003.

[32] Jester data, http://shadow.ieor.berkeley.edu/humor/.

[33] Michelle Anderson, Marcel Ball, Harold Boley, Stephen Greene, Nancy Howse Daniel Lemire "RACOFI: A Rule-Applying Collaborative Filtering System" *Proc. IEEE/WIC COLA'03*, Halifax, Canada, October 2003. NRC 46507.

[34] MovieLens data, http://www.grouplens.org/.

[35] J. Sandvig, Bamshad Mobasher, and Robin Burke, "Robustness of Collaborative Recommendation Basedon Association Rule Mining", *RecSys '07: Proceedings of the 2007 ACM Conference on Recommender Systems (New York, NY, USA)*, ACM, 2007, pp. 105–112.

[36] Xiaoyuan Su and Taghi M. Khoshgoftaar, "Review Article A Survey of Collaborative Filtering Techniques" Hindawi Publishing Corporation Advances in Artificial Intelligence Volume 2009, Article ID 421425, 19 pages doi:10.1155/2009/421425.