International Journal of Engineering and Technical Research (IJETR)
ISSN: 2321-0869, Volume-2, Issue-10, October 2014

Study of Recommender Systems Techniques

Shreya Gangan, Khyati Pawde, Niharika Purbey, Prof. Sindhu Nair

***Abstract*— Recommender systems provide a way to make the user's search for required data from a huge reservoir of data easier. This also benefits the E-learning and E-commerce, which host large databases with a large number of products. This paper attempts to study the basics of the recommender systems and understand the transitions in the trends of approaches like the individual approaches of content-based, collaborative, knowledge-based, utility-based and demographic and their combinations given by hybrid approaches. It mainly focuses on two most successfully used techniques - Collaborative Filtering and Hybrid Systems, as well as the superiority of the latter over the former. The recent developments in hybridization in the field of Recommender Systems are also analysed in an attempt to track their progress.**

***Index Terms*—Recommender Systems, Content-Based Filtering, Collaborative Filtering, Hybrid Recommender Systems.**

I. INTRODUCTION

The internet comprises of enormous amounts of data, and the amount of data is increasing at an accelerated pace. Hence, it is imperative to develop information retrieval techniques that are efficient and capable of handling the copious amount of data. For this purpose, recommender systems were invented. With its applications spreading across numerous fields, the development of various kinds of recommender systems has been phenomenal. A recommender system aims at making personalized recommendations or suggestions using various knowledge discovery algorithms. In other words, it is a tool or a technique that was devised to predict the ratings that a user would give to an item or a product.

However, there are several challenges and problems faced by recommender systems that need to be examined and resolved. Sparsity of data is a common problem wherein many of the fields remain empty i.e. without any evaluation. Many researchers have attempted to mitigate this problem; however this issue was never completely resolved. Another type of the sparsity problem[1,2] is the cold start problem[3] wherein there is very little information about new users or items. Hence in such a case, it becomes challenging for the recommender system to generate recommendations. Furthermore, the issue of spam[4] i.e. malicious users or systems can undesirably influence the system. As a result of which, the recommender system could engender erroneous and false recommendations.

Manuscript received October 06, 2014.

Shreya Gangan, Khyati Pawde, Niharika Purbey, Prof. (Mrs.) Sindhu Nair, Computer Engineering,, Dwarkadas J Sanghvi College of Engineering, Mumbai, India

Our goal was to study the different approaches used for recommendation systems. This paper mainly discusses and studies the collaborative approach and the hybrid approach.

II. RECOMMENDER SYSTEM TECHNIQUES

A. Classification of Recommender Systems

The two major basic architectures of recommender system are Collaborative Filtering and Content-based Recommender systems. Content-based filtering approaches recommend items based on the items the user has rated positively in his user history. They relate these items by applying similarity measures on their properties. Collaborative filtering approaches recommend items to users based on the responses of other users that are similar to the user in question (who have rated similar items and have similar responses).

There are other approaches like knowledge-based, utility-based and demographic techniques. The knowledge-based approach is based on understanding the relation between the user's needs or requirements and the product in question. Since the needs of the user vary over the time, knowledge-based approach doesn't allow creation of a long-term model. The demographic approach groups the users on the basis of their demographic details and gives the recommendations accordingly. The utility-based approach calculates the utility of every item for the user and makes recommendations by assessing the utility against some constraints. These approaches are not suitable for long-term models for Recommendation Systems. Another type of Filtering has emerged to overcome the shortcomings that the two basic approaches pose, which is called Hybrid Filtering[5]. Hybrid Filtering combines the techniques of content-based filtering with collaborative filtering approaches, or others.

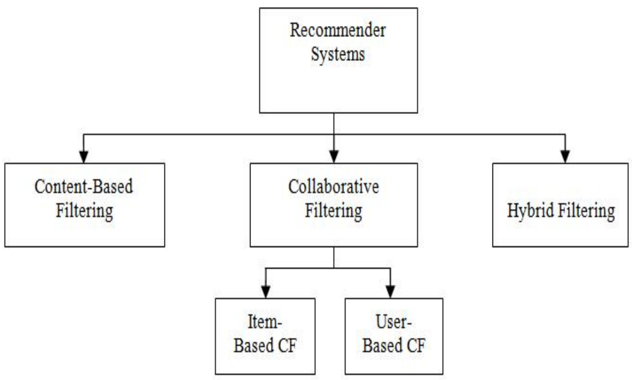

Fig. 1. Major Classification of Recommender Systems

B. Information for the Recommender Systems

Personal Recommender Systems need certain information to base the recommendations on and consequently need to maintain data about clients (users), products (items), sales data and ratings. In general, the user profile and additionally the item profile are maintained by a recommender system. Information about the users' preferences is stored in the user profile. An item profile, which is especially used in content-based recommenders, is a set of records that denotes important characteristics of an item that help relating the items to each other, in terms of their similarity. These profiles can be represented in terms of vectors which represent the presence of each feature in terms of boolean or real-valued or integer-valued components. Apart from these a utility matrix is defined which stores the relation between users and items (the ratings or evaluation of the items by the user). These evaluations can be those representing only the items user shows preference for (unifying) or evaluations for both good and bad items (binary).

The Recommender systems can be classified into two major types based on the means of acquiring the information viz. - explicit and implicit systems. While the explicit systems directly ask the user to hand out the information, the implicit systems keep the user oblivious to their existence by attaining the information about the user's access and traversal without his or her knowledge.

III. Evaluation and Literature Review

A. Collaborative Filtering Algorithms

We can classify the Collaborative Filtering Algorithms depending on the technique used to process the data in the Utility matrix as memory-based such as Pearson Correlation-based CF algorithm[6,7] and model-based filtering such as Bayesian belief net *CF* algorithms[8] and clustering algorithms. Memory-based algorithms use the entire utility matrix to make recommendations. These are used online and don't require any further data. Each prediction is made on the basis of the calculations performed using the complete table (utility matrix). Memory based algorithms search for similarities between users and or items using similarity measurements. Most of these algorithms can be subsequently classified as user-based algorithms or item-based algorithms. In the model-based algorithms a model is constructed in advance and is used to indicate user behavior. This expected model is used to give predictions. The parameters of this estimated model are assessed offline using the data in the matrix.

Each of these approaches has their own advantages and disadvantages: While Memory-Based algorithms are simple and don't have the complexity or overhead of creating a model, they are vulnerable to the classic problems of sparsity, Cold-start and Spam. They take time to predict the results and are not scalable since the entire utility matrix must be used each time to make predictions. On the other hand, the Model-based algorithms are faster to predict and easily find underlying features in the data. They have higher scalability and are less susceptible to the classic problems. But these algorithms suffer from complexity of the models, long time required for construction of models and sensitivity to changes in the data.

Table 1: Comparison between Memory-Based and Model-Based Algorithms

Memory-Based	Model-Based
• Simple	• Complex
• Used Online	• Used Offline
• Slower in Prediction time	• Faster in Prediction time
• Difficult to find underlying characteristics in the data	• Ability to find underlying characteristics in the data
• Poor Scalability	• Higher Scalability
• More susceptible to problems of Sparsity, Cold-start and Spam.	• More sensitive to changes in data
• More Dynamic and Adaptable	• Less Dynamic and adaptable
• No overhead for model creation	• Long construction times for the model

To overcome these shortcomings and to benefit from advantages of both the approaches hybrid techniques which combine the two approaches are used.

B. Hybrid Filtering Algorithms

Hybrid recommender systems combine two or more recommendation techniques in order to increase the overall performance. The main idea is using multiple recommendation techniques to suppress the drawbacks of an individual technique in a combined model.

Hybridization methods can be broadly classified as:

- Weighted: Rating for a given item is computed as the weighted sum of ratings produced by a pool of recommenders (CBF and CF). The weights are determined by training on previous ratings of the user and they may be adjusted as new ratings arrive.
- Mixed: Recommendations from different Recommenders are presented together.
- Switching: The System switches between the recommenders depending on the present condition.
- Feature Combination: Only one recommendation component is employed, which is supported by a second passive component. Here features from different Recommendation data sources are put together into this single system.
- Feature Augmentation: Similar to Feature Combination, the output of one recommender is passed as the input of the secondary one.
- Cascade: The concept is similar to feature augmentation techniques. However, cascade models make candidate selection exclusively with the primary recommender, and employ the secondary recommender simply to refine item scores.
- Meta-Level: Meta-level hybrids feed the constructed model by one recommender to another as input.The constructed model is denser in information when compared to a single rating. Hence in meta level hybrids, more information is carried from one recommender to another.

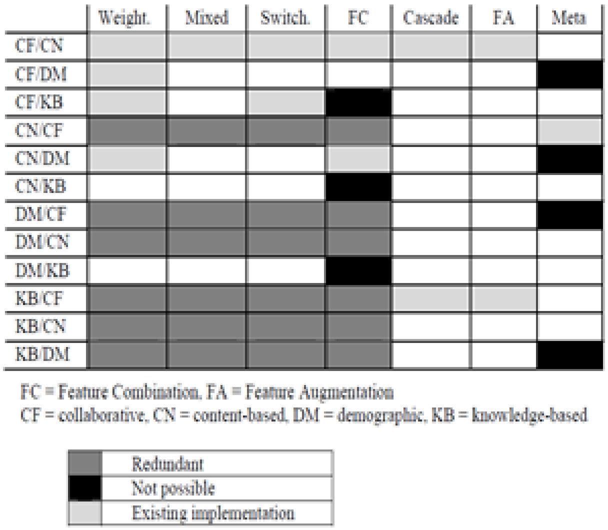

Fig. 2. Hybrid recommender systems combinations (adapted from [9])

Although there exist five recommendation approaches and the seven hybridization techniques, the number of combinations achieved is greater than 35, it is about 53, since many on them are order-sensitive.

Using a restaurant recommendation system called Entrée, which was developed using case-based reasoning;

Fig. 3. Predictions of a query from the Entrée restaurant recommender

Burke attempted to compare the efficiency and performance of different types of hybrid recommendation systems[10]. The suitable restaurant is found using the interactive critique between the Recommender System and the User, a process similar to browsing by shift of focus. The drawback of this system is that its dataset is rather small and mostly contains negative ratings. The experiments were evaluated using ARC (Average Rank of the Correct recommendations) and accuracy of retrieval. Four hybrid algorithms, collaborative Pearson, collaborative heuristic, content-based, and knowledge based were tested while Mixed hybrid and demographic recommendation could not be tested using that dataset. According to their experimental results, the hybrid recommenders showed better performance than basic recommendation systems. Since this alliance was found in situations with smaller session size and sparse density, it was deduced that hybrid recommenders can effectively overcome the problem of cold-start. The experiment also indicted that feature augmentation and cascade were the best hybrids, and the knowledge based technique are appropriate for secondary or contributing components. In feature augmentation a contributing recommender made positive impact on the overall system without affecting the performance of a better algorithm while in cascade hybrids combining of recommender with different strengths makes it proficient. Other studies also conducted experiments on hybrid recommender systems and proved the improvements in the systems[11-14].

IV. Recent Recommender System Approaches

Most frequently used recommendation algorithms in studies were content-based and collaborative filtering recommendation algorithms and the feature augmentation hybrid strategy or its variants for instance, a new system architecture as given in [15] was formulated that supported universal queries by combining the tables (as viewed in relational database). An new approach put forth by Gao Fengrong et al.[16] , based on combining partition-based collaborative filtering, which reduced the dimensions of the utility matrix and meta-information filtering, which solved the low rating problem. This approach achieved high efficiency and performance for digital source management.

Li, Y et al.[17] , proposed a web log mining approach which combined Collaborative Filtering (CF), which finds item sets similar to the content, and Sequential Pattern Mining (SPM), which provides users with recommendations, for recommending learning resources to each active user based on the historical learning path of the user. Results of experiments show good performance of the proposed method.

Salehi, M. et al.[18] , proposed a new material recommender system framework using methods based on concurrent consequence of dynamic interests, multi-preferences and multidimensional attributes of learning materials which revolved around sequential pattern mining and Learner Preference Tree (LPT). This was followed by another approach[19] where techniques like weighted association rules, compact tree (CT), clustering learners and LPT is used. This was again followed by another proposition of hybrid recommender system[20] for learning materials which used Nearest Neighborhood Algorithm (NNA) and Preference Matrix (PM), and showed greatly improved results and can abate the problems of cold-start and sparsity.

V. Conclusion

This paper studies the different approaches that have been used and that are emerging to create an effective recommender system. We have considered various types of recommender systems and their techniques and come to the conclusion that a hybrid system essentially overcomes the inherent problems in an individual approach by amalgamating these approaches. The two most-popular individual approaches viz. Collaborative Filtering (CF) and Content-Based Filtering (CBF) have various disadvantages. While The CBF algorithms have disadvantages like Content description, Over-specialization, Subjective domain problem the CF algorithms have their own such as Early rater problem, Sparsity problem, Gray sheep. The Hybrid Algorithms which can use seven hybridization techniques such as Weighted, Mixed, Switching, Feature Combination, Feature Augmentation, Cascade, Meta-Level allows creation of a system which combines the basic approaches like Content-Based, Collaborative Filtering, Knowledge-based filtering, Utility-based Filtering or Demographic filtering and overcome each other's complementary shortcomings . The experiments on recent algorithms[15-20] proposed for hybrid recommender systems have shown that the Hybrid Recommender Systems have improved performance compared to the native techniques that have been combined therein.

ACKNOWLEDGMENT

We would like to extend our gratitude to our honorable principal Dr. Hari Vasudevan of D.J.Sanghvi College of Engineering and Dr.Narendra Shekhokar, the Head of Department of Computer Engineering for granting us the required amenities for our research.

REFERENCES

[1] Sarwar, B., Karypis, G., Konstan, J., and Reidl, J. 2001. Item-based collaborative filtering recommendation algorithms. In 10th international Conference on World Wide Web. WWW '01. ACM, New York, NY, 285-295. 2001.

[2] Huang, Z., Chen, H., Zeng, D.: Applying associative retrieval techniques to alleviate the sparsity problem in collaborative filtering. ACM Transactions on Information Systems 22, 1, pp. 116-142. 2004.

[3] Schein, A. I., Popescul, A., Ungar, L. H.: Methods and Metrics for Cold-Start Recommendations. In 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, New York. 2002

[4] Mobasher, B., Burke, R., Bhaumik, R., Williams, C.: Effective Attack Models for Shilling Item-Based Collaborative Filtering System. In WebKDD Workshop (KDD'2005). Springer, Chicago, Illinois, USA (2005)

[5] Basilico, J. and Hofmann, T.: Unifying collaborative and content-based filtering. In Twenty- First International Conference on Machine Learning. vol. 69. ACM, New York, NY, 9. 2004.

[6] K. Goldberg, T. Roeder, D. Gupta, and C. Perkins. *Eigentaste: A Constant Time Collaborative Filtering Algorithm, Information Retrieval*. 4(2), pp. 133-151, 2001.

[7] B.M. Sarwar, G. Karypis, J.A. Konstan, and J. Riedl. *Item-based Collaborative Filtering Recommendation Algorithms*. 10th International World Wide Web Conference, pp. 285-295, 2001.

[8] K. Miyahara, and M.J. Pazzani. *Improvement of Collaborative Filtering with the Simple Bayesian Classifier*. Information Processing Society of Japan, 43(11), 2002.

[9] Burke, R. *Hybrid Recommender Systems: Survey and Experiments*. UMUAI 12 (4), 331- 370. (2002)

[10] Burke, Robin D. Hybrid Web Recommender Systems. The Adaptive Web Lecture Notes in Computer Science Volume 4321, 2007, pp 377-408

[11] P Melville, RJ Mooney, R Nagarajan. Content-boosted collaborative filtering for improved recommendations. AAAI/IAAI, 187-192 ;2002

[12] SE Middleton, NR Shadbolt, DC De Roure. Ontological user profiling in recommender systems. ACM Transactions on Information Systems (TOIS) 22 (1), 54-88 ;2004

[13] R Torres, SM McNee, M Abel, JA Konstan, J Riedl .Enhancing digital libraries with TechLens+. Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries;2004

[14] Q Li, BM Kim. An approach for combining content-based and collaborative filters. Proceedings of the sixth international workshop on Information retrieval with Asian languages-Volume 11;2003

[15] Baudisch.P. *Joining Collaborative and Content-Based Filtering.* Proceedings of the SIGCHI conference on. Human Factors in Computing Systems (CHI '99).Pittsburgh, PA, USA, May 15 – 20, 1999, New York: ACM Press.

[16] Gao Fengrong, Xing Chunxiao,Du Xiaoyong and Wang Shan. *Personalised service system based on hybrid Filtering for digital library*.Tsinghua Science and Technology,12 ,1,2007, pp. 1-8.

[17] Li, Y., Niu, Z., Chen, W. and Zhang, W. *Combining Collaborative Filtering and Sequential Pattern Mining for Recommendation in ELearning Environment*. In: 10th International Advances in Web-Based Learning (ICWL 2011). Hongkong, China, December 8-10, 2011, pp. 305-313.

[18] Salehi, M., Nakhai Kamalabadi. and Ghaznavi Ghoushchi, M.B. *Personalized Recommendation of Learning Material Using Sequential Pattern Mining and Attribute Based Collaborative Filtering, Education and Information Technologies*,17,4 ,2012, pp. 1-23.

[19] Salehi ,M and Nakhai Kamalabadi. *Hybrid Recommendation Approach for Learning Material Based on Sequential Pattern of the Accessed Material and the Learner's Preference Tree, Knowledge-Based Systems*,48,2013, pp. 57-69.

[20] Salehi, M., Pourzaferani, M. and Razavi, S.A.Hybrid *Attribute-Based Recommender System for Learning Material Using Genetic Algorithm and a Multidimensional Information Model*. Egyptian Informatics Journal, 14, 1,2013, pp. 1- 23.

Shreya Gangan, B.E. in Computer Engineering, Dwarkadas J Sanghvi College of Engineering, Mumbai, India.

Khyati Pawde, B.E. in Computer Engineering, Dwarkadas J Sanghvi College of Engineering, Mumbai, India.

Niharika Purbey, B.E. in Computer Engineering, Dwarkadas J Sanghvi College of Engineering, Mumbai, India.

Sindhu Nair, Assistant Professor (Computer Engineering Department), Dwarkadas J Sanghvi College of Engineering, Mumbai, India. Prof. Nair has pursue her BE and ME in computers from Mumbai University. She has a teaching experience of 10 years. And an experience of 4 years as a software developer. Prof Nair has authored 3 national and 3 international papers. She is a member of IETE.