Volume 7, Issue 4, April 2017 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Review of Techniques for Recommender Systems

Harbhajan Kaur*
Computer Science Department
North West Institute of Engineering &Technology, Moga, India

Dr. Mohita Garag
Computer Science Department
Principal of North West Institute of Engineering &Technology, Moga, India

Amanjot Kaur
Computer Science Department
North West Institute of Engineering &Technology, Moga, India

Abstract—In electronic commerce, recommender systems are used to help customers to choose products according to their needs. These systems suggest products automatically to users by learning their requirements. Recommendations provided by these systems depends upon users purchase probability and preferences. In this paper, different techniques used for recommender systems are studied.

Keywords—Recommendation Systems, E-commerce, Content Based Filtering, Collaborative Filtering, Hybrid Methods

I. INTRODUCTION

Recommender systems touch our lives every day, from searching on Google to shopping on at any major online retailer. Their sophisticated algorithms attempt and often succeed at showing us the information and products we seek. The information consumers give to service providers and retailers is expanding rapidly, which makes recommendation systems both more complex, and potentially more powerful. Online behaviour including customer metadata, transaction histories and communications allows companies to understand shoppers better, sense their similarities, and address needs they may not even know they have. Simultaneously, we are better able to analyse information about the items sought, including images, sounds in the case of music, and descriptions, which allows companies to refine the ways they cluster products, as well as whom they target as potential buyers. Recommendation systems are at the center of retail both online and off, showing ads on web sites as well as through dynamic displays in brick and mortar stores. By using facial recognition and in-store video cameras, retailers are able to group their customers instantly by gender, age and other demographics in order to show them immediately the goods they are most likely to buy. Given the explosion of consumer goods as well as the rapid increase in the number of vendors on the Internet, the main problem facing customers is how to find the object that they seek, when it is buried beneath a mountain of irrelevant information. A recommender system that can instantly personalize ads in order to solve the customer's search problem is doing both the customer, and the vendor, a favour. New goods are constantly introduced, and new fads sweep the nation, altering shoppers' behaviour. Only a recommender system that is constantly able to learn new patterns can serve consumer needs.

II. LITERATURE REVIEW

In [1] authors provided an extensive empirical evaluation of stability for six popular recommendation algorithms on four real world datasets. This paper provided experiments, results of which suggest that stability performance of individual recommendation algorithm is consistent across a variety of datasets and settings. In this paper, it is suggested that modelbased recommendation algorithm consistently demonstrate higher stability than neighborhood-based collaborative filtering technique. Analysis of important factors such as sparsity of original rating data, normalization of input data, the number of new incoming ratings, the distribution of incoming ratings, the distribution of evaluation data, etc. to report their impact on recommendation stability is performed in this paper.

In [2] a survey of neighborhood-based methods for the item recommendation problem is presented. Nearest neighbors is most popular approach among collaborative recommendation approaches. It is the simple and efficient approach. The main advantages and their important characteristics are described by the authors. While implementing neighborhood-based systems, the required essential decisions and practical information on how to make such decisions are suggested in this document. Large recommendation systems have issue of sparsity and limited coverage. This problem is also discussed and solutions are provided in this paper.

In [3] various efficient techniques available among collaborative filtering are discussed. A framework is suggested which will combine these techniques to obtain good predictions. The methods provided in this paper are compared against Netflix Prize dataset It is proposed that the set of predictors with addition of biases to the regularized SVD, post processing SVD with kernel ridge regression is extended using a separate linear model for each movie and using methods similar to SVD but with fewer parameters. All predictors and selected 2-way interactions between them are combined using linear regression on a holdout set.

In [4] the comparison is made on recommenders based on a set of properties that are relevant to application. A few algorithms are compared using some evaluation metric instead of absolute benchmarking of algorithms. The experimental settings are suggested to make decisions between algorithms. Three experiments namely Offline settings where there is no interaction from user, a small group of subjects experiment and provides experience and large scale online experiment where real user populations interact with the systems are reviewed. Different questions to ask, protocols to use in experimentation etc. are suggested. A largest of properties is reviewed and systems are evaluated in the given relevant properties. A large set of evaluation metrics in the context of the property that is evaluated is also surveyed.

In [5], the computational complexity of user-based collaborative filtering is discussed. Model-based recommendation techniques developed to address this problem analyze the user-item matrix to discover relations between the different items and use these relations to compute the list of recommendations. One of these model-based recommendation algorithm is presented in this paper. This algorithm determines the similarities between the various items and then uses them to identify the set of items to be recommended. The method used to compute the similarity between the items, and the method used to combine these similarities in order to compute the similarity between a basket of items and a candidate recommender item are two key steps in this type of algorithm. It is experimentally presented that these item-based algorithms are two times faster than traditional user-neighborhood based recommender systems. It is also proved by experiment that the algorithm provides better quality recommendations.

In [6] collaborative filtering for web service recommendation system is discussed. Missing QoS (Quality-of-Service) values of web services are considered and also performance improvements are suggested. While measuring the similarities between users and between services, existing QoS prediction methods do not consider personalized influence of users and services. Web service QoS factors like response time and throughput depends on the location of the web services and user which is also ignored by the existing web services recommendation systems. In this paper, a location-aware personalized CF method is used for web service recommendation. It utilizes both locations of users and web services when selecting similar neighbors for the target user or service. It provides enhanced measurement for experiments using real world datasets. It improves the QoS prediction accurately and efficiently compared to traditional CF-based methods.

In [7], various collaborative filtering approaches in web services selection and recommendation do not consider the difference between web service recommendation and product recommendation used in e-commerce sites. A hybrid collaborative filtering approach, region KNN is discussed in this paper. It is designed for large scale web service recommendation. It uses the characteristics of QoS by building an efficient region model. Memory-based collaborative filtering algorithm used in this method provides a quicker recommendation. Region KNN is highly scalable, accurate algorithm compared to other traditional CF algorithms.

In [8], voting classification algorithms such as Bagging and AdaBoost are reviewed and experimented with large empirical study comparing several variants in conjunction with a decision tree inducer (three variants) and a Naive-Bayes inducer. These algorithms use perturbation, reweighting, and combination techniques. These techniques affect classification error. It is discussed that why and when these algorithms affect classification error. It is determined that Bagging reduced variance of unstable methods, while boosting methods reduced both the bias and variance of unstable methods but increased the variance for NaiveBayes was very stable. It showcases fundamental differences between AdaBoost and Arc-x4. Some of voting variants are introduced in this paper such as pruning versus no pruning, use of probabilistic estimates, weight perturbations (Wagging), and back fitting of data. It is proved by experiment that Bagging improves when probabilistic estimates in conjunction with no-pruning are used, as well as when the data was back fit. Tree sizes are measured and it is shown that there is a positive correlation between the increase in the average tree size in AdaBoost trials and its success in reducing the error. Voting methods and non-voting methods are compared which indicated that voting methods lead to large and significant reductions in the mean-squared errors. Practical problems are also discussed.

In [9], a scalable, general-purpose iterative smoothing algorithm is proposed in conjunction with different traditional recommendation algorithms. It improves the stability. The experiments with real world rating data proves the proposed approach more stable compared to the original algorithms. By improving stability, it does not sacrifice the predictive accuracy but instead at the same time improved it.

In [10], a software agents Syskill & Webert are described. These are used to rate pages on World Wide Web. It provides the pages those might interest a user. Thus it increases the classification accuracy without seeing many rated pages. Using 'conjugate priors' from Bayesian statistics for probability revision, the user profile is revised when more training data is available. This approach is compared to learning algorithms that do not make use of such background knowledge, and it is found that a user defined profile can significantly increase the classification accuracy.

In [11] recommendation systems are investigated that combines different data sources and contributes in two distinct ways. First, this study created a framework to evaluate performance and feature importance of different recommendation systems in a setting with multiple data sources as input. Second, the framework was empirically validated on eight datasets in an e-commerce setting. Concretely the empirical validation collects four distinct data sources i.e. product data, customer data, raw behavioral data, and aggregated behavioral data, each composed of individual features. A posteriori weighting and factorization machine based feature combination are deployed as hybridization techniques to exhaustively combine all possible data sources into an integrated recommendation system.

In [12] researchers developed a novel reputation measurement approach for web service recommendations. They firstly detected malicious feedback ratings by adopting the cumulative sum control chart, and then reduced the effect of subjective user feedback preferences employing the Pearson Correlation Coefficient. Moreover, in order to defend malicious feedback ratings, they proposed a malicious feedback rating prevention scheme employing Bloom filtering to enhance the recommendation performance. Extensive experiments are conducted by employing a real feedback rating data set with 1.5 million web service invocation records.

In [13] introduced a probabilistic structure to resolve the diversity-accuracy dilemma in recommender systems. They proposed a hybrid model with adjustable level of diversity and precision such that one can perform this by tuning a single parameter. The proposed recommendation model consists of two models: one for maximization of the accuracy and the other one for specification of the recommendation list to tastes of users. These experiments applied on two real datasets show the functionality of the model in resolving accuracy diversity dilemma and outperformance of the model over other classic models.

III. RECOMMENDER TECHNIQUES

Recommender systems or recommendation engines are active information filtering systems that seek to predict the "preference rating" that user would give to items, aiming at recommending items (movies, music, books, web pages) to users. Recommender systems typically take either of two basic approaches: collaborative filtering or content-based filtering. Other approaches, such as hybrid approaches, also exist.

A. **Content Based Filtering Approach: -** Content based filtering approach gives recommendation to a certain user by examines properties of the items that have been rated by the user and the description of items to be recommended. In a content based system, we must construct for each item a profile, which is a record or collection of records representing important characteristics of that item that are interested to the users. A content-based recommender learns a profile of the user's interests based on the features present in objects the user has rated. Schafer, Konstan & Riedl call this "item-to-item correlation." The type of user profile derived by a content-based recommender depends on the learning method employed. Decision trees, neural nets, and vector-based representations have all been used [14].

B. **Collaborative Filtering Approach: -** Collaborative recommender systems aggregate ratings or recommendations of objects, recognize commonalities between users on the basis of their ratings, and generate new recommendations based on inter-user comparisons. A typical user profile in a collaborative system consists of a vector of items and their ratings, continuously augmented as the user interacts with the system over time. Collaborative filtering approach makes recommendations to users not on basis of their own preference profile, but based on similarities between the profiles of the active user and other users in the system. Collaborative filtering approach is further categorized into memory based and model based approaches. The greatest strength of collaborative techniques is that they are completely independent of any machine-readable representation of the objects being recommended, and work well for complex objects such as music and movies where variations in taste are responsible for much of the variation in preferences. Schafer, Konstan & Riedl (1999) call this "people-to-people correlation."[14]

C. **Hybrid Approaches: -** Hybrid approaches can be obtained from a combination of collaborative and content based filtering approaches, try to capitalize on the strengths of each approach and increase the efficiency of recommender systems. Seven categories of hybrid recommendation approaches, weighted, switching, mixed, feature combination, feature augmentation, cascade, and meta-level have been introduced by Burke [2].

Table 1: Hybridization Methods

Hybridization Method	Description
Weighted	The scores (or votes) of several recommendation techniques are combined together to produce a single recommendation.
Switching	The system switches between recommendation techniques depending on the current situation.
Mixed	Recommendations from several different recommenders are presented at the same time.
Feature combination	Features from different recommendation data sources are thrown together into a single recommendation algorithm.
Cascade	One recommender refines the recommendations given by another.
Feature augmentation	Output from one technique is used as an input feature to another.
Meta-level	The model learned by one recommender is used as input to another

IV. CONCLUSIONS

The study shows that all existing systems employed basic techniques such as collaborative and content based techniques. These techniques have their own advantages and disadvantages. To improve these techniques hybrid techniques were introduced, that combines both. Various hybrid techniques were used to improve performance of recommender systems. In this paper, six hybrid techniques are discussed such as weighted, mixed, switching, feature combination, feature augmentation, and meta-level. The survey shows that recommender systems still need to be improved. There is lack of work in field of knowledge based recommender systems.

REFERENCES

[1] Li, L., Chu, W., Langford, J., & Wang, X. (2011, February). Unbiased offline evaluation of contextual-bandit-based news article recommendation algorithms. In Proceedings of the fourth ACM international conference on Web search and data mining (pp. 297-306). ACM.

[2] Amatriain, X., Jaimes, A., Oliver, N., & Pujol, J. M. (2011). Data mining methods for recommender systems. In Recommender Systems Handbook (pp. 39-71). Springer US.

[3] Koren, Y. (2008, August). Factorization meets the neighborhood: a multifaceted collaborative filtering model. In Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 426-434). ACM.

[4] Hofmann, T. (2004). Latent semantic models for collaborative filtering.v ACM Transactions on Information Systems (TOIS), 22(1), 89-115.M. M. Eining, DS. R. Jones, and J. K. Loebbecke, "Reliance on decision aids: an examination of auditors' assessment of management fraud.," Auditing: A Journal of Practice and Theory, vol. 16, pp. 1-19, 1997.

[5] Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001, April). Item-based collaborative filtering recommendation algorithms. In Proceedings of the 10th international conference on World Wide Web (pp. 285-295). ACM.

[6] Tang, M., Jiang, Y., Liu, J., & Liu, X. (2012, June). Location-aware collaborative filtering for QoSbased service recommendation. In Web Services (ICWS), 2012 IEEE 19th International Conference on (pp. 202-209). IEEE.

[7] Schafer, J. B., Konstan, J. A., & Riedl, J. (2001). E-commerce recommendation applications. In Applications of Data Mining to Electronic Commerce (pp. 115-153). Springer US.

[8] Bauer, Eric, and Ron Kohavi. "An empirical comparison of voting classification algorithms: Bagging, boosting, and variants."Machine learning, vol.36, no.1-2 (1999): 105-139.

[9] AdomaviciusSontakke, S., & Chavan, P. Improving Stability, Smoothing and Diversifying of Recommender Systems.

[10] Pazzani, M. J., Muramatsu, J., & Billsus, D. (1996, August). Syskill & Webert: Identifying interesting web sites. In AAAI/IAAI, Vol. 1 (pp. 54-61).

[11] Herlocker, J. L., Konstan, J. A., & Riedl, J. (2000, December). Explaining collaborative filtering recommendations. In Proceedings of the 2000 ACM conference on Computer supported cooperative work (pp. 241-250). ACM.

[12] Belton, V., & Stewart, T. (2002). Multiple criteria decision analysis: an integrated approach. Springer Science & Business Media.

[13] Javari, Amin, and Mahdi Jalili. "A probabilistic model to resolve diversity–accuracy challenge of recommendation systems." Knowledge and Information Systems, Vol. 44, no. 3, 2015, pp. 609-627.

[14] Burke, R. (2005). "Hybrid systems for personalized recommendations". In Intelligent Techniques for Web Personalization (pp. 133-152). Springer Berlin Heidelberg.

