ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 10, October 2015

A Survey on Evaluation and Improvement Techniques for Recommendation Systems

Sonali Karmalkar, Dr. Shanta Sondur

PG Student, Dept. of I.T., V.E.S. Institute of Technology, Mumbai, India

Professor, Dept. of I.T., V.E.S. Institute of Technology, Mumbai, India

ABSTRACT: Recommender systems are now becoming increasingly important to individual users, businesses and specially e-commerce for providing personalized recommendations. Recommender systems have been evaluated and improved in many, often incomparable, ways. In this paper, we review the evaluation and improvement techniques for improving overall performance of recommendation systems and proposing a semantic analysis based approach for clustering based collaborative filtering to improve the coverage of recommendation .

KEYWORDS: recommendation system, evaluation, collaborative filtering, clustering.

I. INTRODUCTION

Recommendation systems (RSs) are known as the most popular applications of Web personalization. The RSs aim to provide users with personalized services or products that are relevant to their needs and interests. Recent research studies show that existing personalized online services adopt several RSs approaches. These approaches are classified into four main categories, including content-based (CB) filtering, collaborative filtering (CF), knowledge-based filtering and hybrid recommendation [1]. Over the last decade, different recommender systems were developed and used in a variety of domains.

The primary goal of recommenders is to provide personalized recommendations so as to improve user's satisfaction. As more and more recommendation techniques are proposed, researchers and practitioners are facing the problem of how to estimate the value of the recommendations. In previous evaluations, most approaches focused only on the accuracy of the generated predictions based systems. However, a few recent works argue that accuracy is not the only metric for evaluating RS and that there are other important aspects we need to focus on in future evaluations. In [2], they analyzed the evaluation of RS focusing on the quality of recommendations rather than only on their predictive accuracy of algorithms. Quality of a RS has been extensively discussed over the last decades and various definitions can be found in a wide range of literature. Our focus is to survey on RS improvement techniques by improving their evaluation metrics. Current works suggest that other than accuracy there are variety of other metrics such as Quality, Coverage, Diversity, Scalability, User Preferences, Reliability and Similarity etc. that should be considered when evaluating RS [2]. Concretely, as a critical step in traditional CF algorithms, to compute similarity between every pair of users or services may take too much time, even exceed the processing capability of current RSs.

Consequently, service recommendation based on the similar users or similar services would either lose its timeliness or could not be done at all. In addition, all services are considered when computing services, rating similarities in traditional CF algorithms while most of them are different to the target service. The ratings of these dissimilar ones may affect the accuracy of predicted rating. One of the existing system is high-dimensional parameter-free, divisive hierarchical clustering algorithm that requires only implicit feedback on past user purchases to discover the relationships within the users. Based on the clustering results, products of high interest were recommended to the users. However, implicit feedback does not always provide sure information about the user's preference. It does not consider large coverage of large amount of RS.

II. EVALUATING RECOMMENDATION SYSTEMS

There are two existing methodologies for evaluation of recommendation system. First is the system oriented evaluation and second is the user oriented evaluation. System oriented evaluation also known as offline evaluation because users are not involved in the Evaluation process. In this method, a data set is partitioned into training and test sets. Using the training set, data points in the test set are predicted. User-oriented evaluation is also known as online evaluation because in this method feedback from users interacting with the system is collected by explicit questions or implicit observing [3].

When developing a recommender system, either a new algorithm or a new application, it is useful to be able to evaluate how well the system works. When we evaluate the quality of RS, most approaches only focus on the predictive accuracy of these systems. But as discussed in introduction, there are other metrics like Quality, Coverage, Diversity, Scalability, User Preferences, Reliability and Similarity etc. that must be evaluated when evaluating RS [2]. There are various techniques to evaluate the performance of RS. In [4], they have presented and explained a range of common metrics used for the evaluation of recommendation systems in software engineering. Based on a review of current literature, they derived a set of dimensions that are used to evaluate individual recommendation systems or in comparing it against the current state of the art. For the dimensions, they have provided a description as well as a set of commonly used metrics and explored relationships between the dimensions. In [5], they discuss how to compare recommenders based on a set of properties that are relevant for the application. They focus on comparative studies, in this a few algorithms are compared using some evaluation metric, rather than absolute benchmarking of algorithms. They describe experimental settings appropriate for making choices between algorithms. They review three types of experiments, starting with an offline setting, where recommendation approaches are compared without user interaction, then reviewing user studies, where a small group of subjects experiment with the system and report on the experience, and finally describe large scale online experiments, where real user populations interact with the system. In [6], Effective and meaningful evaluation of RS is challenging. To date, there has been no published attempt to synthesize what is known about the evaluation of RS, nor to systematically understand the implications of evaluating RS for different tasks and different contexts. They also have attempted to overview the factors that have been considered in evaluations as well as introduced new factors that they believe should be considered in evaluation. In addition, they have introduced empirical results on accuracy metrics that provide some initial insight into how results from different evaluation metrics might vary. They hope that this article will increase the awareness of potential biases in reported evaluations, increase the diversity of evaluation dimensions examined where it is necessary, and encourage the development of more standardized methods of evaluation. In [2], they focus on two crucial metrics in RS evaluation: coverage and serendipity. Based on a literature review, they first discuss both measurement methods as well as the tradeoff between good coverage and serendipity. Then they analyze the role of coverage and serendipity as indicators of recommendation quality, present novel ways of how they can be measured and discuss how to interpret the obtained measurements. Overall, they argue that their new ways of measuring these concepts reflect the quality impression perceived by the user in a better way than previous metrics thus leading to enhanced user satisfaction.

III. IMPROVING RECOMMENDATION SYSTEMS

In [7], they conclude that, the first RS were focused on improving recommendation accuracy through filtering. Most memory-based methods and algorithms were developed and optimized in this context (e.g., kNN metrics, aggregation approaches, singular value decomposition, diffusion-based methods, etc.). At this stage, hybrid approaches (primarily collaborative–demographic and collaborative content filtering) improved the quality of the recommendations. In the second stage, algorithms that included social information with previous hybrid approaches were adapted and developed (e.g., trust-aware algorithms, social adaptive approaches, social networks analysis, etc.). Currently, the hybrid ensemble algorithms incorporate location information into existing recommendation algorithms. Evaluation of the predictions and recommendations has evolved since the origins of RS, which weighted prediction errors (accuracy) heavily. They also recognized the convenience of evaluating the quality of the top *n* recommendations as a set; evaluation of the top *n* recommendations as a ranked list was then incorporated. Currently, there is a tendency to assess new evaluation measures, such as diversity and novelty. The authors in [8], describes various limitations of current recommendation methods and discusses possible extensions that can improve recommendation capabilities and make RS applicable to an even broader range of applications. These extensions include, among others, an improvement of understanding of users and items, incorporation of the contextual information into the recommendation process, support for multcriteria ratings, and a provision of more flexible and less intrusive types of recommendations. In [9], Social tagging systems pose new challenges to developers of RS. As observed by recent research, traditional implementations of classic recommender approaches, such as collaborative filtering, are not working well in this new

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 10, October 2015

context. To address these challenges, a number of research groups worldwide work on adapting these approaches to the specific nature of social tagging systems. In joining this stream of research, they have developed and evaluated two enhancements of user-based collaborative filtering algorithms to provide recommendations of articles on CiteULike, a social tagging service for scientific articles. In [10], they showed that a CF framework can be used to combine personal Information Filtering (IF) agents and the opinions of a community of users to produce better recommendations than either agents or users can produce alone. It also shows that using CF to create a personal Combination of a set of agents produces better results than either individual agents or other combination mechanisms. One key implication of these results is that users can avoid having to select among agents; they can use them all and let the CF framework select the best ones for them. Following is the metrics wise literature survey to improve the RS:

A. Accuracy

In [11], this paper presents a metric to measure similarity between users, which is applicable in collaborative filtering processes carried out in RS. The proposed metric is formulated via a simple linear combination of values and weights. Values are calculated for each pair of users between which the similarity is obtained, whilst weights are only calculated once, making use of a prior stage in which a genetic algorithm extracts weightings from the RS which depend on the specific nature of the data from each RS. The results obtained present significant improvements in prediction quality, recommendation quality and performance. Thus the improvements can be seen in the system's accuracy. In [12], they examine an advanced collaborative filtering method that uses similarity transitivity concepts. By propagating "similarity" between users, in a similar way as with "trust", we can significantly expand the space of potential recommenders and also improves the recommendation's accuracy. In [13], they propose several new approaches to improve the accuracy of recommendations by using rating variance (which, as we show, is inversely related to the recommendation accuracy) to gauge the confidence of recommendations. They then empirically show how these approaches work with different recommendation techniques. We also show how these approaches can generate more personalized recommendations, as measured by the coverage metric (described later in the paper in more detail). As a result, users can be given a better control to choose whether to receive recommendations with higher coverage or accuracy. In [14], despite its success, similarity-based collaborative filtering suffers from some significant limitations, such as scalability and sparsity. This paper introduces trust to the domain of collaborative filtering to overcome these limitations. Compared with the similarity-based CF, introduction of trust does improve the performance of CF in terms of coverage, prediction accuracy, and robustness in the presence of attacks. Experimental results based on a real dataset are illustrated as evidences to support their claim. In [15], they have presented two contributions to the RS field, both of them based on a semantic approach. The common goal of their work has been to improve collaborative filtering recommendations in e-commerce, in terms of accuracy and reliability. To this aim, their strategies rely on an ontology that formalizes the semantic descriptions of commercial products.

B. Coverage

In [11], they proposed a use of genetic algorithms that can be applied for obtaining optimal similarity functions. Those similarity functions obtained provide better quality and faster results than the ones provided by the traditional metrics. Improvements can be seen in the system's accuracy (MAE), in the coverage and in the precision and recall recommendation quality measures. The proposed use of GAs applied to the RS is a novel approach and has the main advantage that it can be used in all CF-based RS, without the need to use hybrid models which often cannot be applied, as in many cases no reliable demographic information or content-based filtering information is available. In [12], they proposed a novel similarity propagation scheme to confront the data sparsity problem in RS and evaluate their method over two datasets with different characteristics, exhibiting a much higher recommendation coverage and better accuracy than classical collaborative filtering methods even under very sparse data conditions. In [13], using a simple filtering approach they have demonstrated that prediction accuracy can be significantly improved by filtering out recommendations above a minimum rating standard deviation threshold. However, there was also a corresponding decrease in the coverage of recommendations. They then proposed the smart and safe approaches which generate recommendations of greater value by providing a good balance of prediction accuracy and coverage. New approaches are especially useful, since they can confidently improve the accuracy of recommendation, and in addition to that, a user can control the balance between the accuracy and coverage of recommendations. In [14], it has been shown by the

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 10, October 2015

experimental results that the trust metrics and corresponding prediction making approach do improve the performance of traditional similarity-based CF in terms of coverage, prediction accuracy and robustness.

C. Diversity

There is increasing awareness in the RS field that diversity is a key property that enhances the usefulness of recommendations. In [16], they argued that as new types of recommendation domains and tasks emerge, this blind faith in the similarity assumption begins to seem flawed. They showed that very often recommendation diversity is important and that traditional recommendation systems are marred by poor diversity characteristics. They evaluate a new class of diversity-preserving algorithm capable of addressing this without compromising similarity or efficiency. In [17], they introduce and explore a number of item ranking techniques that can generate substantially more diverse recommendations across all users while maintaining comparable levels of recommendation accuracy. Comprehensive empirical evaluation consistently shows the diversity gains of the proposed techniques using several real-world rating datasets and different rating prediction algorithms. In [18], Genre information can serve as a means to measure and enhance the diversity of recommendations and is readily available in domains such as movies, music or books. In this work they propose a new Binomial framework for defining genre diversity in RS that takes into account three key properties: genre coverage, genre redundancy and recommendation list size-awareness. They show that methods previously proposed for measuring and enhancing recommendation diversity including those adapted from search result diversification fail to address adequately these three properties. They also propose an efficient greedy optimization technique to optimize Binomial diversity. In [19], they present topic diversification, an algorithmic framework to increase the diversity of a top-N list of recommended products. In order to show its efficiency in diversifying, they also introduced their new intra-list similarity metric. Contrasting precision and recall metrics, computed both for user-based and item-based CF and featuring different levels of diversification, with results obtained from a large scale user survey, they showed that the user's overall liking of recommendation lists goes beyond accuracy and involves other factors, e.g., the users' perceived list diversity. They were thus able to provide empirical evidence that lists are more than mere aggregations of single recommendations, but bear an intrinsic, added value. In [20], in most cases, new techniques are designed to improve the accuracy of recommendations, whereas the recommendation diversity has often been overlooked. In particular, they showed that, while ranking recommendations according to the predicted rating values provides good predictive accuracy, it tends to perform poorly with respect to recommendation diversity. Therefore, in this paper, they proposed a number of recommendation ranking techniques that can provide significant improvements in recommendation diversity with only a small amount of accuracy loss. In addition, these ranking techniques offer flexibility. In [21], they propose a neighbor diversification collaborative filtering algorithm to improve the recommendation lists. By using Movielens dataset for empirical analysis, they investigated the influence of neighbor diversity to the recommendation accuracy, diversity, novelty and coverage. Intensive experimental results proved the efficiency of their proposed algorithm for improving recommendation lists.

D. Quality

In [22], Information Filtering and Collaborative Filtering techniques have been used for selecting information based on the user's previous preference tendency and the opinion of other people who have similar tastes with the user. Combining both Information Filtering and Collaborative Filtering, or hybrid systems, they have also been proposed to get better recommendation results. In this paper, they present an improved recommendation method that copes with the sparsity problem of the hybrid systems and increases the quality of recommendation results. In [23], this paper involves a prefiltering process that eliminates the least representative users from the k-neighbour selection process and retains the most promising ones. The improvements obtained are always positive with respect to both the prediction quality measures and the recommendation quality measures. This demonstrates that certain users should not be included among the active user's neighbours and that the traditional similarity measures are not capable of detecting them. The favourable results obtained here can be considered generally applicable due to the broad margin of improvement observed and their testing on the two most representative databases of collaborative filtering RS. In [11], they have presented a genetic algorithm method for obtaining optimal similarity functions. The similarity functions obtained provide better quality and quicker results than the ones provided by the traditional metrics.

E. Scalability

In [24], they presented and experimentally evaluated a new approach in improving the scalability of RS by using clustering techniques. Their experiments suggest that clustering based neighborhood provides comparable prediction quality as the basic CF approach and at the same time significantly improves the online performance.

F. User Preferences

In [25], many RS employed in commercial web sites use collaborative filtering. The main goal of traditional collaborative filtering techniques is improvement of the accuracy of recommendation. Nevertheless, such techniques present the problem that they include many items that the user already knows. These recommendations appear to be good when they consider accuracy alone. On the other hand, when they consider users' satisfaction, they are not necessarily good because of the lack of discovery. In their work, they infer items that a user does not know by calculating the similarity of users or items based on information about what items users already know. They seek to recommend items that the user would probably like and does not know by combining the above method and the most popular method of collaborative filtering. In [26], they have presented an approach to improve traditional RS which are special types of expert systems able to select automatically the most relevant information for each individual by exploiting the knowledge of an expert in a particular domain and the users' preferences. Specifically, their strategy prevents from selecting fake neighborhoods in collaborative RS. This problem appears in domains such as e-commerce where there are a wide range of products, and the different categories contain items of very different nature (such as books, music, cloths or food). In these contexts, taking into account all the preferences registered in user's profiles when estimating their similarity of interests can lead to the selection of fake neighbors, that is, neighbors that have dissimilar interests with the target user with respect to the target product, but with similar preferences to him/her regarding lots of items of other categories.

G. Reliability

Collaborative RS select potentially interesting items for each user based on the preferences of like-minded individuals. Particularly, e-commerce has become a major domain in this research field due to its business interest, since identifying the products the users may like or find useful can boost consumption. During the last years, a great number of works in the literature have focused in the improvement of these tools. Expertise, trust and reputation models are incorporated in collaborative RS to increase their accuracy and reliability. However, current approaches require extra data from the users that is not often available. In [15], they present two contributions that apply a semantic approach to improve recommendation results transparently to the users. On the one hand, they automatically build implicit trust networks in order to incorporate trust and reputation in the selection of the set of like-minded users that will drive the recommendation. The common goal of their work has been to improve collaborative filtering recommendations in e-commerce, in terms of accuracy and reliability. To this aim, their strategies rely on an ontology that formalizes the semantic descriptions of commercial products. The exploitation of semantics enables reasoning about the data stored in the users' personal profiles and inferring new knowledge.

IV. CONCLUSION AND FUTURE WORK

Evaluation and Improvement techniques used for various metrics of Recommendation Systems has discussed in detail. There are two methods for evaluating RS, first is the system oriented evaluation that is also known as offline evaluation and second is the user oriented evaluation that is also known as the online evaluation. RS can be improved with the help of improving various metrics of RS such as accuracy, coverage, diversity, quality, scalability, user preferences, reliability, etc.

Future study suggest, Semantic Analysis based approach for clustering based collaborative filtering which improves the coverage of recommendation and in turn improves the quality and usefulness of the RS. In this approach the system will perform semantic analysis on the description text of service. In this way, more semantic-similar services may be clustered together, which will further increase the coverage of recommendation.

REFERENCES

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 10, October 2015

[1] Malak Al-Hassan, Haiyan Lu , Jie Lu, "A semantic enhanced hybrid recommendation approach: A case study of e-Government tourism service recommendation system" Decision Support Systems 72 (2015) 97–109.
[2] Mouzhi Ge, Carla Delgado-Battenfeld and Dietmar Jannach, "Beyond Accuracy: Evaluating Recommender Systems by Coverage and Serendipity", TU Dortmund 44221, Dortmund, Germany.
[3] Alejandro Bellogin,Ivan Cantador,Fernando Diez, Pablo Castells and Enrique Chavarriaga, "An Empirical Comparison of Social, Collaborative Filtering, and Hybrid Recommenders"
[4] Iman Avazpour, Teerat Pitakrat, Lars Grunske and John Grundy, "Dimensions and Metrics for Evaluating Recommendation Systems", Chapter 12 in Recommendation Systems in Software Engineering Springer, 2014.
[5] Guy Shani and Asela Gunawardana, "Evaluating Recommendation Systems"
[6] Jonathan L. Herlocker, Oregon State University And Joseph A. Konstan, Loren G. Terveen, and John T. Riedl University of Minnesota, "Evaluating Collaborative Filtering Recommender Systems"ACM Transactions on Information Systems, Vol. 22, No. 1, January 2004, Pages 5–53.
[7] J. Bobadilla, F. Ortega, A. Hernando, A. Gutierrez, "Recommender systems survey", Knowledge-Based Systems 46 (2013) 109–132
[8] Gediminas Adomavicius, Member, IEEE, and Alexander Tuzhilin, Member, IEEE, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions", IEEE Transactions on Knowledge and Data Enginnering, VOL. 17, NO. 6, JUNE 2005.
[9] Denis Parra-Santander, Peter Brusilovsky, "Improving Collaborative Filtering in Social Tagging Systems for the Recommendation of Scientific Articles", 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology.
[10] Nathaniel Good, J. Ben Schafer, Joseph A. Konstan, Al Borchers, Badrul Sarwar, Jon Herlocker, and John Riedl , "Combining Collaborative Filtering with Personal Agents for Better Recommendations"
[11] Jesus Bobadilla, Fernando Ortega, Antonio Hernando, Javier Alcala, "Improving collaborative filtering recommender system results and performance using genetic algorithms", Knowledge-Based Systems 24 (2011) 1310–1316
[12] Anna Satsiou and Leandros Tassiulas, "Propagating User's Similarity towards improving Recommender Systems", 2014 IEEE Computer Society DOI 10.1109/WI-IAT.2014.37
[13] Gediminas Adomavicius, Sreeharsha Kamireddy, YoungOk Kwon,"Towards More Confident Recommendations: Improving Recommender Systems Using Filtering Approach Based on Rating Variance"
[14] Jianshu Weng, Chunyan Miao, Angela Goh,"Improving Collaborative Filtering with Trust based Metrics"
[15] Manuel Ramos-Cabrer, Yolanda Blanco-Fernández, Martín López-Nores, "Semantic inference of user's reputation and expertise to improve collaborative recommendations" Expert Systems with Application (2012) 8248–8258
[16] Keith Bradley and Barry Smyth, "Improving Recommendation Diversity"
[17] Gediminas Adomavicius and YoungOk Kwon, "Improving Recommendation Diversity Using Ranking-Based Techniques" Working Paper:Improving Recommendation Diversity Using Ranking Based Techniques.
[18] Saúl Vargas, Linas Baltrunas ,Alexandros Karatzoglou, Pablo Castells," Coverage, Redundancy and Size-Awareness in Genre(style) Diversity for Recommender Systems"
[19] CaiNicolas Ziegler, Sean M. McNee, Joseph A, Konstan, Georg Lausen, "Improving Recommendation Li Through Topic Diversification"
[20] Gediminas Adomavicius, Member, IEEE, and YoungOk Kwon, "Improving Aggregate Recommendation Diversity Using Ranking-Based Techniques" IEEE Transactions on Knowledge and Data Enginnering,VOL.24 NO.5 YEAR 2012.
[21] Chao Yang, Cong Cong Ai, Renfa L, "Neighbor Diversification-Based Collaborative Filtering for Improving Recommendation Lists", 2013 IEEE International Conference on High Performance Computing and Communications & 2013 IEEE International Conference on Embedded and Ubiquitous Computing.
[22] Saranya Maneeroj, Hideaki Kanai and Katsuya Hakozaki, "Combining Dynamic Agents and Collaborative Filtering without Sparsity Rating Problem for Better Recommendation Quality" University of Electro-Communications ,Chofugaoka, Chofu, Tokyo 182-8585 Japan.
[23] Fernando Ortega, Jose-LuisSanchez, JesusBobadilla,Abraham Gutierrez, "Improving collaborative filtering-based recommender systems results using Pareto dominance" Information Sciences 239 (2013)50–61
[24] Badrul M. Sarwar, George Karypis, Joseph Konstan, and John Riedl, "Recommender Systems for Large-scale E- Commerce: Scalable Neighborhood Formation Using Clustering"
[25] Yoshinori Hijikata,Takuya Shimizu,Shogo Nishida, "Discovery-oriented Collaborative Filtering for Improving User Satisfaction"
[26] Manuela I. Martin-Vicente, Alberto Gil-Solla, Manuel Ramos-Cabrer, Jose J. Pazos-Arias, "A semantic approach to improve neighborhood formation in collaborative recommender systems" Expert Systems with Applications 41 (2014) 7776– 7788

BIOGRAPHY

Sonali Eknath Karmalkar is a Junior Research Assistant in the Information Technology Department, from V.E.S. Institute of Technology, Mumbai University. She received Bachelor of Engineering (B.E.COMP) degree in 2013 from PUNE University,Pune, India. Her research interests are data mining (Recommendation Systems).