Volume 3, Issue 3, March 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Techniques for Novelty in Recommender System

Nisha Nandal
Computer science Department,
M.D.University, Rohtak, Harayana-124001
India.

Chhavi Rana
Computer science department,
M.D.University, Rohtak, Haryana124001
India.

Abstract— ***Recommendation systems help the people to find things according to their interest and they are used widely with the development of electronic commerce. Many recommendation systems employ the collaborative filtering (CF) techniques, which are proved to be one of the most successful in recommender systems in recent years. CF techniques can be applied in entertainment domain also, for example in Amazon.com. CF provides recommendation based on the individual preferences. Because of the obvious increase of products in electronic commerce systems, the time consumed in searching the real time requirement of the customer sometimes results in failure. At the same time it suffers from poor novelty in case the user database increases. Less novelty is the major reason of increase in user frustration at times when he is looking for something new. To solve this problem in collaborative based filtering techniques, this paper propose a technique that enhance the novelty in a data set using the item cluster matrix and user cluster matrix, which contains item to user relationships. This technique is more efficient and accurate than the traditional one.***

Keywords— Recommender Systems, Collaborative basedfiltering, Novelty, User cluster, Item cluster.

I. Introduction

As there is growing popularity of using recommender systems in e-commerce, a variety of recommender algorithms have been proposed over time. The goal of recommender systems is to reduce information overload and provide personalised recommendation for users. Over the years recommendation systems have been widely applied in e-commerce, for example, book recommendation on Amazon or movie recommendation on Netflix [10].

This paper evaluates different aspects of recommender systems focusing on the quality of recommendations rather than only on their predictive accuracy of algorithms. Quality as a concept has been extensively discussed over the last decades and various definitions can be found in a wide range of literature. A definition that is especially prevalent in marketing and the service industries is the one of Gronroos [3]. Gronroos defines quality as meeting and/or exceeding customer's expectations. Another well-known definition was introduced by [6], who defined quality as "fitness for use". From these definitions we are able to conclude that (1)a high-quality recommendations need to be fitting their intended purpose and (2) the actors behind this purpose are the ultimate judges of the quality of recommendations. One particular problem in the RS domain is that the purpose of a RS and the actors behind (costumers and sellers) take different positions and has different perspectives on the purchase act. Thus, a "good" recommendation can be considered at the same time one that makes the customer happy or points him to an interesting item as well as one that maximizes the sales margin. The whole point of recommendation is linked to a notion of discovery, because recommendation makes most sense when it exposes the user to a relevant experience that would not have found, or thought of by him - obvious, however accurate recommendations are generally of little use. The general problem to find a subset of items that best match a query or request issued by a system user is common across many applications of information retrieval. The items in question may be products offered for purchase in a recommender system, or documents returned by a search engine for instance. Not only does a varied recommendation provide in itself for a richer user experience. Given the inherent uncertainty in user interest prediction-since it is based on implicit, incomplete evidence of interests, where the latter are moreover subject to change avoiding a too narrow array of choice is generally a good approach to enhance the chances that the user is pleased by at least some recommended item.In addition, a number of changes in terms of content that affect the type of recommendation can be implemented in the RS [8]. However the changes in the recommender systems behaviour go beyond temporal factors and involve context, novelty, serendipity, real-time dynamics as well as diversity. All these parameters contribute to the dynamic behaviour of Dynamic Recommender Systems. The contribution work in this field is used in different categorisation of recommender system i.e. content based filtering, collaborative filtering and hybrid system [7].The novelty of an item could be another promising one, as this information is in general available for every item. Besides that, new items are the ones that the user has possibly little knowledge about and for which it is important for a recommender to be able to make the user aware of them and promote them. For items the user already has heard a lot about (think of currently top-selling, generally-liked items) even an accurate recommendation would not be too meaningful. Yet another possible measure is effectiveness which can be approximated by comparing what has been recommended and what has actually been purchased, or the click-through rate. Using novelty is another dimension added by RS. Sometimes, finding something new in specific space give better results that interests the user as it discovers new items for the user. This enhances the novelty in specific set of data objects. Sales novelty may enhance business as well, leveraging revenues from market. It is easy to increase novelty by giving up accuracy; the challenge is to enhance these aspects while still achieving a fair match of the user's interests. The authors introduce an interesting evaluation approach consisting of the biased selection of novel test items, whereby evaluating for novelty is achieved by studying the accuracy on such difficult items.The proposed technique roots recommendation novelty metrices on a few ground concepts and formal models. We identify three essential concepts: choice, discovery and relevance, upon which the framework is built. The metric scheme takes at its core an item novelty model –discovery-based or distance-based– which mainly determines the nature of the resulting recommendation metric. Item rank and relevance are introduced through a probabilistic recommendation browsing model, building upon the same three basic concepts. Based on the combination of ground elements, and the assumptions in the browsing model, different metrics and variants unfold. We provide model estimation approaches on available observations of the interaction between users and items, thus providing for the practical computation of the metrics upon both explicit and implicit data. We report experimental observations validating and illustrating the properties of the proposed metrics.

II. Existing Techniques

Novelty in recommender systems is matured with a few algorithms being developed, tested and compared with each other. They collect and establish profiles, and determine the relationships among the data according to models developed. If a system recommends items based on their popularity, it is likely not doing a task the user could not have done by himself –even if the user happens to like the items, the chances that he had already heard about them are high, whereby the recommendation is of very marginal –if any– use. As another case, a very accurate system could return a set of monothematic items matching the user known themes or interests. This approximation may also fail since, albeit accurately matching the user's preferences, the whole set of recommended items may be perceived as one –consider the case of a music recommendation algorithm that only returns songs of the same artist. The key in these situations is that novelty and diversity should be also considered in the quality assessment of a RS. The possible categories of the data in the profiles include user preferences, user behaviour patterns or item properties and accordingly, rating can be done using user data set and item data set. The user-feature vector defines the extent to which novelty can be enhanced. Several existing works found to be very useful for this new dimension. It is argued that recommendation evaluation needs to move beyond conventional accuracy metrics.

Bergmann [17] has addressed the question of case utility in case-based systems; that is, the usefulness of cases to the problem at hand, as opposed to simply their similarity to the current problem. In this article, we take the view that novel items have greater utility and focus on strategies to recommend such items. Indeed, novelty is discussed in Herlocker [6] as an important evaluation dimension for collaborative filtering recommender systems, although no concrete means to carry out such an evaluation is given. The authors propose three heuristic algorithms for selecting retrieval sets that combine similarity and diversity. A new angle on novelty by considering the effect that the passing of time may have on known items, which may regain some of their novelty value as past user experience fades away and is to some degree forgotten by users. They explore the positive effect that exploiting such oblivion discovery-based recommendation novelty metrics. processes may have on the diversity of recommendation. Novelty in recommendation is especially relevant to exploit the Long Tail effect, i.e., the situation where a few items are extremely popular and there is the rest of them are much less known. As stated by Anderson [3], recommender systems may benefit from selling less of more, that is, recommending less wide-known items to more users instead of focusing on highly-popular items. Recommending long-tail items, which few users have accessed to, is a common way in which novelty is understood. Zhou et al [18] define novelty as the average self-information of recommended items, which amounts to the average log inverse ratio of users who like the item (also known as "inverse user frequency"). They target this metric by means of hybrid strategies combining collaborative filtering with graph spreading techniques. Celma and Herrera [12] take an interesting alternative view on long-tail novelty. Rather than assessing novelty just in terms of the long-tail items that are directly recommended, they analyse the paths leading from recommendations to the long tail through similarity links.

George, et al. [5] considered a novel collaborative filtering algorithm which is based on a weighted co-clustering algorithm. It involves simultaneous clustering of items and users. An efficient real time collaborative filtering framework has been built using incremental and parallel versions of co-clustering algorithm. Their empirical evaluation of the proposed approach on large dataset of movie and book shows that accuracy can be obtained upto some extent using matix approach. Xue et Al. [23] present a novel approach which mix up the advantages of memory based collaborative filtering and model based collaborative filtering techniques by using a smoothing based approach. In this approach, the clusters so generated from training dataset provide the basis for data smoothing and neighbourhood selection. As a result, they provide increased efficiency in recommendations. Their empirical studies on two datasets namely EachMovie and MovieLens demonstrate that their new proposed approach consistently outperforms other user based traditional approaches. Santos et al. [18] gives a novel algorithm for diversification is presented. The xQuAD (*explicit query aspect diversification*) algorithm makes use of query reformulation provided by commercial web search engines to derive new sub-queries that will cover the possible aspects of the initial query. So,given an ambiguous query and a ranking of retrieved documents *R*, this algorithm will greedily select and inserting in a new ranking randomly in the document for maximizing the probability.

As stated by Zhou [21] that by plugging the popularity-based item novelty models in the general metric scheme we get discovery-based recommendation novelty metrices, which we label as expected popularity complement (EPC). This may lead to alternative formulations, to which we shall refer as expected inverse popularity (EIP), and expected free discovery (EFD), respectively. All three metrics provide a measure of the ability of a system to recommend relevant long-tail items. EPC can be read as the expected number of seen relevant recommended items not previously seen. EIP and EFD can be read as the expected IUF and ICF of (relevant and seen) recommended items, respectively. Note that the mean self-information (MSI) of the recommended items, a metric reported in [17]. If we take a distance-based novelty model relative to the set of items the target user has interacted with the items in his profile, we get an alternative novelty measure consisting of the expected distance between the recommended items and the items in the user profile, which we label as the expected profile distance (EPD). Recommendation Novelty Metrices: In this case, each term in the summation is doubly weighted by the relevance of the involved item pair, and only once by the rank distance function. The metric provides a user-relative measure of novelty which, as far as we are aware of, has not been reported in the literature. Item novelty is the core element in the definition of recommendation novelty. Item novelty can be understood and defined in different ways, depending on which the resulting metrics differ considerably [10]. We identify two main relevant approaches to model item novelty, based on discovery and distance respectively. The framework is nonetheless open to the modular integration of alternative models.

Meanwhile, Jung et al. [8] hierarchical structures are employed to describe the relationships among users. The preferences of each user can be described in a hierarchical structure. The structure represents the index of categories, which are the labels of the nodes. Matching one structure to another with all category labels results in that each node contains a group of users with similar preferences. Hierarchical structures can also be applied on similarity computations for items [15]. Edges in the structure clearly define how items are related for the item-to-item relationships. A hierarchical structure, a tree, specified the relative weights for the edges provide information on how much two items are related. A method of the order-based similarity measurement has been proposed for building a personal computer recommendation system (PCFinder) [14]. Instead of using 0/1 for the search, this method uses the concept in Fuzzy Logic to estimate the similarity.

Celma and Herrera [13] present two methods named, item and user centric to evaluate the quality of novel recommendation. They observe that though CF recommend less novel item than CBF, user's perceived quality is higher. This is because CF is biased towards popularity, effecting novelty and network topology while CBF is not affected at all. Park and Tuzhilin [22] deals with the concept of novelty in a whole new way.They attempt to study the long tail problem of recommender systems where many items in the long Tail have only few ratings, thus making it hard to use them in recommender systems. They are rarely recommended but have got potential to interest user at times, finding which not a trivial task is. Abbassi [26] examines the case of over-specialization in recommender systems, which results from returning items that are too similar to those previously rated by the user. They develops an algorithm Outside The Box (OTB), that attempts to identify regions that are underexposed to users, by taking some risk to help users make fresh discoveries, while maintaining high relevance. On the other hand, Vargas and Castells [15] noted that there is lack of well defined evaluation metrics in this area that take into account their ranking. Therefore, they proposed a framework built upon three ground concept namely choice, discovery and relevance and generalizes several state of the art metrics using them. In contrast, much work on recommendation algorithms is focused on the rating prediction problem. For example, Wang [8] propose a method to unify the user-based and item-based CF approaches by similarity fusion, Xue [23] combine the advantages of memory-based and model-based approaches by introducing a smoothing-based method, in order to improve the accuracy of the predictions, etc. In recent years, the Netflix prize has resulted in much effort been spent on the development of algorithms that reduce the root mean-squared prediction error. Note that algoritnms that perform well from a classification perspective. Indeed, once a rating prediction is made for a set of items, it still remains to decide which of these to recommend. One approach, which we can refer to as a predict-and-select-highest strategy, is to recommend those items with the highest predicted ratings. On the datasets that we have studied, algorithms that focus specifically on generating a top-N list, such as the item-based kNN algorithm proposed in Karypis [4], give a much better performance from the point of view of precision than the predict-and-select.

III. Challenges

There are still different drawbacks of different approaches described above. First, the lack of information that is sparsity problem would definitely affect the recommendation results. For, relationships in items that are not yet-labelled or not yet-rated can make the novelty in recommender system still less useful. The second problem is it may not cover the exhaustive case. If the scales of user profiles is small or the users have unique tastes, similarity decisions accommodating novelty cannot be established. The third problem is frequency update. If any new information of users has to be included in the recommendation processes in real time, data latency will increase the waiting time for the query result. The fourth problem can be that within the datasets nothing can be said about the novelty of *single item* in response to a query. As novelty is based on some proposed models and metrices , used to build a better recommender system. But still there remain some challenges that need to be met for better fulfilment of user requirements. First and Foremost challenge is the Cold-smart problem that is an item cannot be recommended unless it has been rated by a number of users. This problem applies to new items as well and is particularly detrimental to users. Likewise, a new user has to rate a sufficient number of items before the CF algorithms be able to provide accurate recommendations. The second problem is of Sparsity, there are a few rating of total number of items available in a user-item database. Sometimes, due to high level of data inflow creates missing values because of this rating can't be done properly and that not only hinders the novel recommendation of items but also produces insufficient results to the user. The third problem is of predictions that are deemed successful as long as the predicted set overlaps with the user's real preferences. However it does not account for other qualities of the prediction, in particular, they do not consider the difficulty of the prediction. Taking the perspective of the end-user, we argue that the deficiency of similarity-based retrieval methods is better understood as a failure to retrieve novel but relevant predictions. The biggest challenge for off-line research is the availability of data sets, widely available data set—the EachMovie data set is essential to provide novel but still relevant items. the next challenge is the availability of recommendation engines that are flexible enough to meet specific requirements very quickly in case of on-line search. The remaining challenge is to develop a set of experiment control and data analysis tools, designed to generate community-accepted quality and performance metrics, that can be used to efficiently test hypotheses.

IV. Future Trends

The primary objective of every RS is to satisfy the seller's interests by satisfying the customer's interests. The importance of novelty is particularly manifest in scenarios such as automatic recommendation and exploratory search, in which user needs involve some degree of uncertainty and/or under specification, leaving room for the system's initiative to complete and predict these needs on behalf of –or in collaboration with– the user. In particular novelty have been identified as the key dimension of information utility in real scenarios, and a fundamental research direction to keep making progress in the field like investigating the experimental lack of significant difference between rank-aware metrices and rank- unaware metrices, determining whether the combination of different implicit or explicit features could provide better results for the recommended items. Novelty represents a cutting-edge research area in intelligent information access and delivery, eliciting increasing attention from researchers and practitioners in the recommender systems, information retrieval, and machine learning communities. This special issue seeks to gather a selection of papers reporting leading research on diversity and discovery perspectives in recommender systems and exploratory search, providing a view of the latest advances in this scope.

V. Conclusions

The field of recommender system research is matured with lots of algorithm being developed, tested and compared with each other. However, it maintains its charm with new discoveries and some completely new dimensions of work. This paper explained many techniques for improving the novelty in recommender systems. This paper focuses on several techniques that will help in future research directions. Moreover, it helps in understanding of novel things that guides the user in right direction.

Acknowledgement

Many techniques have been proposed for better recommender systems. But developing recommender systems with increased novelty is a challenging task. Some specific metrices have been introduced for capturing global novelty, user-relative novelty intra-list being all among them that are all aware of concepts such as ranking position and relevance of the items recommended. In this paper several methods for retrieving novel cases in both case-based reasoning systems and collaborative filtering recommender systems have been discussed. An evaluation methodology has been introduced that allows for the first time the performance of different methods to be analyzed from the perspective of their ability to recommend novel but relevant items. The results from different novelty based recommendation systems indicate that the systems can afford to provide novel results to users. Consequently, it can predict user pattern in a much wider scope and enhance novel item recommendation. It also reduces time of searching by comparing the prediction with reality. Although it is true that missing data may lower the accuracy of prediction. This indicates that reduction of the missing or insufficient data is not simple, and that some approximations are required to be performed in order to provide better predictions of user preferences. This emerging field of novelty in recommendation system not only provides better results but also found to be useful in real life applications like business which can use it a core feature to make its sales in a profitable way.

References

1. C. L. A. Clarke, M. Kolla, and O. Vechtomova (2009). An Effectiveness Measure for Ambiguous and Underspecified Queries. ICTIR, 35(8): 188-199.
2. C. Zhai, W. W. Cohen, and J. Lafferty. Beyond independent relevance: Methods and evaluation metrics for subtopic retrieval. SIGIR, 38(3): 10-17.
3. C. Anderson (2006). The Long Tail Effect: Why the Future of Business is Selling Less of More. In the proceedings of 2006 ACM Conference on Recommender systems Recsys, (06), 37(4):455-458.
4. G.Karypis. (2001). Evaluation of item-based top-N recommendation algorithms, In Proceedings of the 10th International Conference on Information and Knowledge Management. ACM SIGMIS, SanDiego, CA, USA, pp. 247–254.
5. G. George and M. George Giaglis (2006). Improving the prediction accuracy of recommendation algorithms: Approaches anchored on human factors, interacting with computers, 18(6):410-431.
6. J. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl (2004). Evaluating collaborative filtering recommender systems. In the proceedings of 2004 ACM Conference on Transactions Information Systems 22(1): 5–53.
7. J. Zobel and A. Moffat (2008). Rank-biased precision for measurement of retrieval effectiveness.In the Proceedings of ACM Transactions on Information Systems, 32(5):1-27.
8. J. Wang, A. P. Varies, and M .J. Reinders (2006). Unifying user-based and item-based collaborative filtering approaches by similarity fusion, In Proceedings of the SIGIR Conference, Chiba, Japan, pp. 501–508.

9. J. Jung, J.S. Yoon, and G.S. Yo (2001). Collaborative information filtering by using categorized bookmarks on the web. Web Knowledge Management and Decision Support. 14th International Conference on Applications of Prolog, 25(35):pp. 237-50.
10. M. Zhang and N. Hurley (2009). Avoiding Monotony: Improving the Diversity of Recommendation Lists.In the proceedings of 2009 conference on Recommender systems RecSys(09), 24(5):123-130.
11. M. Balabanovic and Y.Shoham (1997). Fab: Content-based collaborative recommendation. Communications of the ACM, 40(3): 88-89.
12. N. Lathia, S. Hailes and L. Capra (2009). Temporal collaborative filtering with adaptive neighbourhoods, In the proceedings of 32nd international ACM SIGIR conference on Research and development in information retrieval (SIGIR, 2009), Lausane, Switzerland, pp.796-797.
13. O. Celma and P. Herrera (2008). A new approach to evaluating novel recommendations, In Proceedings of the 2008 ACM conference on Recommender systems (RecSys 08), Lausanne, Switzerland, pp. 179-186.
14. P. Xiao, E. Aimeur, and J. M. Fernandez (2003). PCFinder: an intelligent product recommendation agent for e-commerce, In the Proceedings IEEE International Conference on E-Commerce. IEEE Computer Society. 16(14): 181-8.
15. P. Castells, S. Vargas and J. Wang (2011). Novelty and Diversity Metrices for Recommender Systems: Choice, Discovery and relevance, In International Workshop on Novelty in Document Retrieval (DDr 2011), Las Vegas, Nevada, pp. 45-87.
16. P.Ganesan, H. Garcia-Molina and J. Widom (2003). Exploiting hierarchical domain structure to compute similarity. In the proceedings of ACM Transactions on Information Systems,ACM, 21(1): 64-93.
17. R. Bergmann, M. Richter, S. Schmitt, A. Stahl, and I. Vollrath (2001). Utility-Oriented matching: A new research direction for case-based reasoning, In Proceedings of the German Workshop on Case-Based Reasoning, Paris, France, pp. 265-305.
18. R.Santos, C.Macdonald, and I. Ounis (2010). Exploiting Query Reformulations for Web Search Result Diversification. 19th International Conference on World Wide Web (WWW'10), (pp. 881-890). Raleigh, NC, USA.
19. S. Robertson (2008). A New Interpretation of Average Precision. 31st International ACM SIGIR Conference on Research and Development in Information Retrieval, Singapore,Russia. pp. 689-690.
20. T. Sakai (2006). Evaluating evaluation metrics based on the bootstrap. In the 29th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'06), Seattle, WA, USA pp. 525-532.
21. T. Zhou, Z. Kuscsik, J-G. Liu, M. Medo, J. R. Wakeling, and Y. C. Zhang (2010). Solving the apparent diversity-accuracy dilemma of recommender systems, In Proceedings of the National Academy of Sciences of the United States of America, 107(10): 4511-4515.
22. Y. J. Park and A. Tuzhilin (2008). The long tail of recommender systems and how to leverage it, In Proceedings of the 2008 ACM conference on Recommender systems (RecSys 08), October 2008, Lausanne, Switzerland, pp.11-18.
23. X. Xue, Q. Yang, W. Zeng, Z. Chen (2005). Scalable collaborative filtering using cluster- based smoothing, In Proceedings of the ACM SIGIR Conference, Hongkong, China, pp. 114–121.
24. Y. Koren (2008). Factorization meets the neighborhood: A multifaceted collaborative filtering model, In Proceedings of the 14th ACM Siekd Conference, Barcelona, Spain, pp. 78-93.
25. Z. Abbassi, S. Amer-Yahia, L. V. S., Lakshmanan, S. Vassilvitskii, and C. Yu (2009). Getting recommender systems to think outside the box, In Proceedings of the third ACM conference on Recommender systems RecSys 09, ACM Press, Dublin, Ireland, pp. 285-288.
26. Z.Gronroos(1983). Strategic management and marketing in the service sector. Marketing Science Institute. USA.