Study on Techniques to Improve Diversity in Recommender Systems

Vaishnavi.S[1], Jayanthi.A[2] , Karthik.S[3]

Abstract— **Recommender systems play a significant role in E-Marketing. Many companies got increase in their sale due to establishing their products in the sites like Amazon.com, Netflix.com, and Movielens.com etc. These sites help the customers to find relevant product to their interest. They play as a place where customers can find all kinds of items. These systems should give qualitative recommendations e.g. in book recommendations, it should not always tell the books of same authors. Two important qualities are accuracy and diversity. Increasing the diversity decreases the accuracy. Trade-off between accuracy and diversity should be minimized. In this paper, we gave a survey about the techniques which aims to increase the diversity with minimal accuracy loss.**

*Index Terms—***Recommender Systems, E-Marketing, Accuracy, Diversity, Social Diversity, Long-Tail products**

I.INTRODUCTION

In this electronic world, E-Marketing plays a vital role in promoting sale of products. It helps for the growth of business by using computer technologies. E-Marketing comprises core business process of buying, selling goods, services and information over the internet. Its usage is growing day-by-day.

It is used to market the products in online. In simple, company produces goods; sale is done electronically through E-Marketing. Due to this, people enjoy shopping from their home or office or anywhere by using just a computer or laptop. This has evolved after the emerging of internet. Many companies provide information about their products through internet. Using internet, people may get more information about lots of products in less time i.e. faster access to knowledge about products. It transfers the traditional way of marketing into computer oriented marketing [1].

Traditional marketing strategies are as follows:
1. Printing banners, notices, brochures etc
2. Using medias such as TV, Fm, email etc

Using (1), it can cost lot of money if there is mistake in brochures, banners after printing is over. Using (2), it does not involve how people respond to those advertising technique for e.g. people may or may not respond to email. In case of E-marketing, response of people to their advertisement technique can be felt. It is done by knowing how much people have visited the company's website.

There are many websites, in which millions of products are being given by those websites. So information available in the internet about products is rapidly increasing. Customers approach some website and search for the product by giving name, and then website offer products related to customers search produced by many vendors. Products like books, articles, CDs, groceries, electronic goods are all being sold in this website.

II.RECOMMENDER SYSTEMS

One problem is users expect that the product should be matched with his interest i.e. difficult to find relevant content. The key approach developed to this problem is "Recommender systems (RS)". It helps consumer to choose the product among so many options [2] choosing the right item among millions of products is challenging for the consumer.

Recommender Systems finds relevant items from number of alternatives. It has a high commercial value. This has been used by popular website like Amazon.com, Netflix, Movielens, and Facebook etc. It provides personalized recommendation to users. Many approaches and algorithms are being developed since it is related with the real world practical applications.

Recommendation algorithms are designed in such a manner to generate a top-list of items. Recommender systems uses past history about users and items to give a recommendation. They allow users also to know about "Long-Tail" items. "Long-Tail" items are items which are not frequently expressed to users [5]. It aims to rate the product which have not yet been discovered by users earning more profit..

III.RECOMMENDER SYSTEMS IN ADVERTISEMENT

Advertisements promote sales of a product by customers. Using RS help advertisement through following ways
1. By recommendation the product which not known by user yet.
2. By giving bundle purchases[3]

IV.IMPORTANCE OF RS

They are important to both individual users and for business people. Business people by using recommender systems can obtain high sales of their products. Thus they can attain high profit. For individual users, it helps them to find relevant items according to their interest. For persons'

---

providing products in their websites, RS satisfies them by providing right product to customers.

V.RECOMMENDER SYSTEMS VS SEARCH ENGINE

S.No	RS	SE
1)	They cannot use Key words	Can use Keywords
2)	Users don't know what they want to search	Users know what they want to search
3)	They look for needs of user actively	They are passive.

VI.RECOMMENDATION PROBLEM

Recommendations are made based on preferences that a user will give to an item. This is usually done using ratings. Ratings can be explicit, by asking customer about the product or implicit, observing through purchase behavior of customer. Recommending items to users consists of following steps:
1. Prediction
2. Ranking
3. Recommendation

Prediction involves estimation of ratings of products that are not yet consumed by users using the known ratings. Ranking phase involves arranging the rank recommendations of each user. Recommendation phase involves giving top-N items to the user. Vast amount of information is investigated to find the items with high predicted ratings.

There are two basic entities in recommender systems. They are the following:
1. User (customer)
2. Item (product)

User uses recommender system to obtain products of interest, provides opinion and also obtains the recommendations. Item is bought and rated by users.

Recommendation problem can be given as follows: Let "U" be the set of users and "I" be the set of items. Usefulness of item i€I to user u€U is the rating that a user give to an item or it is the preference given to item by user. This rating can be a binary, ordinal or numerical value. R (u, i) indicates how much user "u" likes item "i". These are known ratings. This rating is used estimate the unknown ratings say R^{*}(u, i). This unknown rating has to be estimated to recommend the products that are not yet discovered.

There are two phases in recommendation. In phase I, unknown ratings are estimated for unknown items. In phase II, with all unknown ratings, top-N recommendations are calculated by using high predicted ratings.

VII.QUALITIES OF RS

Some of the quality aspects of RS are Accuracy, Diversity, Novelty and Serendipity. Accuracy gives the more

relevant items to user i.e. accurate recommendations. Diversity tries to gather more number of distinct items. There are two types of diversity namely individual and aggregate diversity. Novelty of a piece of information generally refers to how different it is with respect to "what has been previously seen", by a specific user, or by a community as a whole. A serendipitous recommendation helps the user to find a surprisingly interesting item that he might not have otherwise discovered.

There is always a compromise between accuracy and diversity. For e.g. if person A is recommended 10 bestselling books of author "Shakespeare" it is accurate but not diverse because person A is not informed about other author's books. Thus, diversity can be achieved with some loss in accuracy.

VIII.APPROACHES OF RS

There are 3 types of approaches namely
1. Content-Based
2. Collaborative
3. Hybrid method

In the following sections, C denotes User space (set of users) and S denotes Item space (set of items).

In Content-Based Methods, User will be recommended items similar to those they preferred in the past. Utility function u(c,s), in which item s€S for user c€C. Utility function u(c,s') in which user c is given item s'€ S. Estimation of ratings for s' is done using rating structure of s because s and s' are similar items. E.g. Movie recommendations done based on the director or actor or category of film.

In Collaborative Method, recommender systems recommend items to users based on similarity with other user and not using similarity between items. Formally, utility u(c,s) of item s for user c is estimated based on the utilities $u(c_j,s)$ assigned to item s by those users c_j€C who are "similar" to user c. E.g. for movie recommendation, collaborative methods recommend items to a user by concerning the similar users. Tastes of users are used to make recommendations. They use the concept of "like-minded users"**.**

In Hybrid Method, it recommends items by combining approaches such as content-based and collaborative. Hybrid approaches can be implemented in several ways: by making content-based and collaborative-based predictions separately and then combining them; by adding content-based capabilities to a collaborative-based approach (and vice versa); or by unifying the approaches into one model [13].

IX.RECOMMENDATION METHODS

The following are the various means to achieve qualitative recommendations.
i) State-Of-Art Methods

This paper [4] describes the methods which form the main categories of recommender systems. They are content-based, collaborative and hybrid methods. They are the main methods mostly used for obtaining recommendations but still there exists some limitations with these methods.

Content-based method has the disadvantage such as

1. Limited features only can be considered since this method concentrates on characteristics of item for recommendation
2. Users are recommended the same type of item already recommended
3. New user problem arises

In collaborative method, following are the limitations:

1. User's preferences have to be known earlier
2. New- item problem arises
3. There needs to be mass number of users

In hybrid approach, new-user and new-item problem arises.

These methods can be extended which provides better recommendation capabilities. Some of the possible extensions specified in this paper are:

1. Understanding of users and items
2. Supporting multicriteria ratings
3. By considering multidimesions of recommendations
4. By trying to avoid explicit ratings (non-intrusive)
5. By developing good metrics

These extensions improve the quality of recommendations thereby providing good recommendations to user.

ii) Social Diversity

Social networks are used in this paper [9] to achieve diversity. Social network information is used for recommendations. Social networks are networks connected by many users who can share their views. Vendors can indirectly make money by advertising their product in these networks. Some of social networks are facebook, orkut, YouTube, blogger, and flicker etc. These sites involve interests of many groups of people. Preferences or likes given by these users in different communities are used. Hence, more diversity can be achieved.

In these social networks, for e.g. user see a particular "presentation slides" in YouTube and rate it as "like" or "dislike". This information is used to rate the "presentation".

Here, diversity is achieved by utilizing user's membership in social networks. These networks can be "real-world" or "virtual group". Real-world groups are networks created and explicitly joined by users. Virtual groups are groups created automatically by system based on usage where members are from different geographical location.

Advantages:

1. This diversity can apply to any domains
2. Source of recommendations could be selected
3. Users' can increase or decrease diversity range

iii) Network-Based Interface (NBI)

In [11], NBI involves a bipartite network which has user-object network, object-object network. In these networks, nodes are objects (items) and connection between them is established when two objects are bought or collected by atleast one user. A weight is assigned to each node which relates an object to another. This method provides more accurate recommendations than traditional collaborative filtering (CF) technique.

Major problem in this method is it provides redundancy correlations for any attribute when considering cumulative recommendations. This redundancy reduces accuracy. An improved algorithm is proposed by introducing a free parameter which reduces higher order strong correlations. This gives higher accuracy. This parameter also affects diversity of recommendations.

Here, hamming distance is a concept introduced. It is used to quantify diversity. A better recommendation can be made when hamming distance is large.

iv) "Long-Tail" products

"Long-Tail" products are niche products that are developed for specialized purposes but earn more profit. They are less popular but still has high profit value. As it has less popularity they are kept unrevealed during recommendations. These products should be included during recommendations

If products sold in large amount, it's profit is low since all companies have to sell the product for same price. If product sells less, they earn more profit if given to buyers. So, due to considering "long-tail" products all types of products will be available to users.

In this paper [12], a graph based representation is used to represent the user and item. It contains vertices and edges. Random-walk similarity involves how close a item to a user. To obtain diversity using long-tail products, three concepts were given. They are hitting time, absorbing time and absorbing cost.

These algorithms imply high user interest and high recommendation diversity

v) Graph-Theoretic Approach

Graph-based approach given in this paper [10] helps to attain maximum diversity. It uses max-flow problem and max-bipartite matching problem in graph theory. Max-flow problem is a simple approach. Maximum number of edges gives high recommendation diversity. Maximum-bipartite transforms the general solution of max-flow to specialized solution. It is less complex. It gives solution of top-1 to top-N recommendation. This problem matches the item and user.

Here we achieve high diversity in recommendations.

X. USAGE OF DATA MINING IN RECOMMENDER SYSTEMS

Data mining is used to find patterns from large sets of data. In recommender system applications, data mining is a technique used to find recommendation rules. Data mining helps to find correlations between users and items, two users and also between two items. Algorithms such as clustering, classification and association rule mining etc.

Clustering is used to group users or items according to specified characteristics for e.g. users having similar interests or buying behavior. Classification is a technique used to categorize the data sets for e.g. classifying items based on price. Association rule mining is mostly used because it finds relationship between two data sets for e.g. it finds relationship between two products often bought together.

Thus, data mining plays a significant role in recommender systems [17].

XI. CONCLUSION

A recommender system plays a key role in real life in online marketing. It helps people to buy right product. It can be applied to various domains. It helps to individual user to give personalized recommendations. It gives a collection of

items to customer. Data mining is most helpful for this process. Various techniques is evolving day-by-day to give a right product to right customer. This paper gives a idea of various methods of recommending personalized products.

REFERENCES

[1] "E-Commerce book", P.T. Joseph, Prentice Hall of India-New Delhi, 2009

[2] Ricci.F, Rokach.L "Recommender systems handbook", Springer, 1st edition, 2011

[3] R. Garfinkel, R. Gopal, A. Tripathi, and F. Yin, "Design of a Shopbot and Recommender System for Bundle Purchases,"Decision Support Systems, vol. 42, no. 3, pp. 1974-1986, 2006.

[4] G. Adomavicius and A. Tuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions," IEEE Trans. Knowledge and Data Eng., vol. 17, no. 6, pp. 734-749, June 2005.

[5] D.G. Goldstein and D.C. Goldstein, "Profiting from the Long Tail," Harvard Business Rev., vol. 84, no. 6, pp. 24-28, June 2006.

[6] Z. Huang, "Selectively Acquiring Ratings for Product Recommendation," Proc. Int'l Conf. Electronic Commerce, 2007

[7] D. Lemire, S. Downes, and S. Paquet, "Diversity in Open Social Networks," technical report, University of Quebec, Montreal, 2008.

[8] B.M. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Analysis of Recommender Algorithms for E-Commerce," Proc. ACM Conf. Electronic Commerce, pp. 158-167, 2000.

[9] Swapneel Sheth, jonathan Bell, "Towards Diversity in Recommendations using Social Networks", Columbia University Computer Science Technical Report,2011

[10] Gediminas Adomavicius, YoungOk Kwon," Maximizing Aggregate Recommendation Diversity: A Graph-Theoretic Approach" Workshop on Novelty and Diversity in Recommender Systems (DiveRS 2011), held in conjunction with ACM RecSys 2011. October 23, 2011, Chicago, Illinois, USA

[11] Tao Zhou, Ri-Qi Su,"Accurate and diverse recommendations via eliminating redundant correlations", New Journal of Physics, 2009

[12] Hongzhi Yin Bin Cui,"Challenging the Long Tail Recommendation",Proc of VLDB Endowment,vol 5, No 9,Aug2012

[13] http://en.wikipedia.org/wiki/Recommender_system

[14] Greg Linden, Brent Smith,"Amazon,com Recommendations", IEEE Internet Computing, industry report.

[15] Rong Hu, Pearl Pu,"Helping Users Perceive Recommendation Diversity" Workshop on Novelty and Diversity in Recommender Systems (DiveRS 2011), held in conjunction with ACM RecSys 2011. October 23, 2011, Chicago, Illinois, USA.

[16] C-N. Ziegler, S.M. McNee, J.A. Konstan, and G. Lausen, "Improving Recommendation Lists through Topic Diversification," Proc. 14th Int'l World Wide Web Conf., pp. 22-32, 2005.

[17] J. Ben Schafer,"The Application of Data-Mining to Recommender Systems" Encyclopedia of Data Warehousing and Mining, Second Edition, 2009

[1] Pursuing Final Year ME Software Engineering in SNS College of Technology. I had completed her Bachelor Degree in Maharaja Prithivi Engineering College. My Area of interests is Data Mining & Cloud Computing.

[2] **Jayanthi.A** - She is working as a Assistant Professor in the Department of Computer Science & Engineering, SNS College of Technology, affiliated to Anna University- Coimbatore, Tamilnadu, India.. Her area of interests is Networking & Software Testing.

[3] **Professor Dr.S.Karthik** is presently Professor & Dean in the Department of Computer Science & Engineering, SNS College of Technology, affiliated to Anna University- Coimbatore, Tamilnadu, India. He received the M.E degree from the Anna University Chennai and Ph.D degree from Ann University of Technology, Coimbatore. His research interests include network security, web services and wireless systems. Dr.S.Karthik published more than 35 papers in refereed international journals and 25 papers in conferences and has been involved many international conferences as Technical Chair and tutorial presenter. He is an active member of IEEE, ISTE, IAENG, IACSIT and Indian Computer Society.