Cold-Start Problem in Recommender Systems A Literature Review

Azhar Yousf Mir
Department of Computer Sciences University of Kashmir, Srinagar, J&K, India

Majid Zaman Scientist,
Directorate of Information Technology & Support Systems, University of Kashmir, Srinagar, J&K, India

S. M. K. Quadri
Professor, Department of Computer Science, Jamia Millia Islamia, India

Muheet Ahmed Butt
Scientist ,PG Department of Computer Sciences, University of Kashmir, Srinagar, J&K, India

Abstract— Recommender systems are software tools / Applications used to extract most relevant information from a large set of data for a particular user based on user's history and preferences. But most of the Recommender Systems are facing 'Cold - Start Problem' i.e., when the user is searching for new item its rating is zero or has not history available, hence is not recommended to any of the users. In this paper we will try to find the different methods that have been used overcome this problem.

Key words *— Recommender Systems, Cold Start Problem, Hybrid Recommender Systems.*

I. INTRODUCTION

INFORMATION organizations and Communication Technology (ICT) has influenced the people by facilitating effective retrieval of information with the help of various Search Engines. These Search Engines are facing a serious challenge due to the huge storage of data particularly due the access and availability of Word Wide Web. The availability of data on web also demands a proper filtering mechanism for common users in order to fetch relevant information to avoid any sort of confusion while searching. This makes ICT community rethink and develop alternate mechanism to enhance the significance of Search Engines in the present scenario. The Artificial Intelligence (AI) community has carried out the great deal of work on how artificial intelligence can help the people to find what they exactly want on the web and as a result the idea of Recommender Systems (RS) has been introduced and widely accepted.

Recommender Systems are Software tools which represent user preferences for suggesting items to the user for its use. The suggestions relate to various decision-making processes, such as what items to buy, what online news to read , or what music to listen. “Item” is the general term used to denote what the system recommends to users. A RS normally focuses on a specific type of item (e.g., Books, Movie, Audio Track, CDs, or news) and accordingly its design, its graphical user interface, and the core recommendation technique used to generate the recommendations are all customized to provide useful and effective suggestions for that specific type of item.

Recommender Systems are usually classified according to their approach of rating estimation. Recommender Systems are of following types

1. Content Based Approach (CB) : In this approach the system recommends items to the user that are similar to the ones that the user liked in the past. The similarity of items is calculated based on the features associated with the compared items. For example, if a user has positively rated a movie that belongs to the comedy genre, then the system can learn to recommend other movies from this genre. Some of the central problems concerning content-based recommender systems are limited content analysis, overspecialization and the new user problem.

2. Collaborative Filtering Approach (CF) : The simplest and original implementation of this approach recommends to the active user the items that other users (neighbours) with similar tastes liked in the past. The similarity in taste of two users is calculated based on similarity relations like Correlation , Cosine and other similarity in the rating history of the users. This is the reason why collaborative filtering is refered to as “people-to-people correlation.” Collaborative filtering is considered to be the most popular and widely implemented technique in RS. Collaborative recommender systems exhibit the new user problem and first have to learn user preferences to make reliable recommendations. Beside the new user cold start problem collaborative approaches also exhibit the new item problem, which means that a new item needs to be rated by a sufficient number of users in order be recommended accurately by the system.

3. Hybrid Recommender System : Another type of recommendation system called Hybrid Recommendation system, which combines the advantages of CF and CB and reduces their limitations. Hybrid recommender systems are composed of basic recommender systems like Content-Based, Collaborative Filtering systems.

The main problem with Collaborative Recommender System is that when a user is searching for some new item for the first time, its rating is zero and hence can not be recommended to any user called Cold Start Problem (a.k.a Sparse Rating or First Rater Problem) , it reduces the performance of a Recommendation System.

II. BACKGROUND / RELATED WORK:

In collaborative filtering (CF), user ratings are used to generate recommendations for any user. It works by taking into consideration the taste of other users called neighbours. CF has been extensively studied and tested on movie domain [1] but same methodology can be applied to a number of other domains as well, jokes [2], news, newsgroups [3], advertisements, and more. CF algorithms vary from basic "memory-based" methods [4] to more advanced model-based methods in which we first train a model for example, a classifier [5],—based on some historical data, and then use this trained model to generate recommendations. So Many CF algorithms have been devised and tested, which includes machine learning (ML) methods [6], graph-based methods [7], etc. In addition to CF many hybrid methods which combine Content-Based (CB) and CF techniques have also been proposed [8], these systems are very productive when user ratings are sparse, for example in cold-start situations. Also CF methods cannot be used at all if user has not rated any item, in such cases CB methods or hybrid CB/CF methods are required. Hyung Jun Ahn (2007) in his paper presented a new heuristic similarity measure that focuses on improving to suggest recommendations under cold-start conditions where only a few ratings are available for similarity calculation for each user.

Heung-Nam Kim , Ae-Ttie Ji, Inay Ha, Geun-Sik Jo (2009) presented a new recommendation algorithm via collaborative tags of users to enhance recommendation quality and to overcome some of the limitations in collaborative filtering systems. The experimental results demonstrated that the proposed algorithm offers significant advantages both in terms of improving the recommendation quality for sparse data and in dealing with cold-start users as compared to traditional CF algorithms.

Heung-Nam Kim , Abdulmotaleb El-Saddik , Geun-Sik Jo (2011) have proposed a method of building models derived from explicit ratings and we apply the models to Collaborative filtering recommender systems. The proposed method first predicts actual ratings and subsequently identifies prediction errors for each user. From this error information, pre-computed models, collectively called the error-reflected model, are built. We then apply the models to new predictions. Experimental results show that our approach obtains significant improvement in dealing with cold start problems, compared to existing work.

Jesus Bobadilla , Fernando Ortega, Antonio Hernando, Jesus Bernal (2011) presented a new similarity measure perfected using optimization based on neural learning, which exceeds the best results obtained with current metrics. The metric has been tested on the Netflix and Movielens databases, obtaining important improvements in the measures of accuracy, precision and recall when applied to new user cold start situations.

Chien Chin Chen , Yu-Hao Wan, Meng-Chieh Chung, Yu-Chun Sun (2012) have proposed a cold start recommendation method for the new user that integrates a user model with trust and distrust networks to identify trustworthy users. The suggestions of these users are then aggregated to provide useful recommendations for cold start new users. Experiments based on the well-known Epinions dataset demonstrate the efficacy of the proposed method. Moreover, the method outperforms well-known recommendation methods for cold start new users in terms of the recall rate, F1 score, coverage rate, users coverage, and execution time, without a significant reduction in the precision of the recommendations.

Laila Safoury and Akram Salah (2013) have suggested using of user demographic data for providing better recommendations instead of rating history to avoid cold start problem. They presented a framework for evaluating the usage of different demographic attributes, such as gender, age, location, and occupation for generating recommendations.

Haitao Zou, Zhiguo Gong, Nan Zhang, Wei Zhao and Jingzhi Guo (2013) have presented TrustRank recommendation algorithms which combine with user–user similarity to generate the recommendations, and experimentally evaluated these algorithms. Our results showed that the TrustRank recommendation algorithms provide reasonably accurate recommendation that are better than those provided by User-Based CF. Furthermore, the algorithms are significantly faster and scalable to a system with a larger number of users.

Yang Yujie, Zhang Huizhi and Wang Xianfang (2013) have proposed a novel approach to solve the new user problem in collaborative filtering based on the social network analysis theory. First, we compute the similarities between each pair users and build user relation network. Then cliques are divided from the network by using the UCINET software. For measuring the similarity between new users and existing users, the papper gives the ability of user information, which can distinguish different users. According to preferential attachment characteristic of the BA scale free network model and the cliques, the nearest neighbor of the new user can be selected. Finally, user-based nearest neighbor recommendation is used to recommend items to new users. The experiments show that the proposed approach can effectively alleviate the new user problem.

Blerina Lika, Kostas Kolomvatsos , Stathes Hadjiefthymiades (2014) have presented a method to alleviate the new user cold start problem for RSs applying CF. The proposed system adopts a three-phase approach in order to provide predictions for new users. We adopt a mechanism that takes into consideration their demographic data and based on similarity techniques finds the user's 'neighbours'. We define as 'neighbours', users having similar characteristics with the new user. The idea is that people with a common background and similar characteristics have more possibilities to have similar preferences. Hence, each novel users is classified in a group and accordingly a rating prediction mechanism is responsible to result ratings for items. The final ratings are calculated through a weighted scheme where developers can pay attention to specific attributes or select a more 'fair' approach.

Mi Zhang , Jie Tang Xuchen Zhang , Xiangyang Xue (2014) have tackled the cold-start problem by proposing a context-aware semi-supervised co-training method named C-SEL. Specifically, we use a factorization model to capture fine-grained user-item context. Then, in order to build a model that is able to boost the recommendation performance by leveraging the context, we propose a semi-supervised ensemble learning algorithm.

Andre Luiz Vizine Pereira , Eduardo Raul Hruschka (2015) presented a hybrid approach that combines the features of collaborative filtering recommendations with demographic information. The approach is based on an existing algorithm "Simultaneous Co-Clustering and Learning" (SCOAL), and provides a hybrid recommendation approach that can address the cold start problem, where no ratings are available for new users.

Nowadays from a business

REFERENCES

[1] N. Good, J. B. Schafer, J. A. Konstan, A. Borchers, B. M.Sarwar, J. L. Herlocker, and J. Riedl. "Combining collaborative filtering with personal agents for better recommendations.", In AAAI/IAAI, pages 439–446, 1999.

[2] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. Reidl, "Item-based collaborative filtering recommendation algorithms.",In WWW, pages 285–295, 2001.

[3] J. A. Konstan, B. N. Miller, D. Maltz, J. L. H. L. R. Gordon, and J. Riedl. GroupLens, "Applying collaborative filtering to Usenet news", Communications of the ACM, 40(3):77–87, 997

[4] P. Resnick, N. Iacovou, M. Suchak, P. Bergstorm, and J. Riedl. GroupLens, "An Open Architecture for Collaborative Filtering of Netnews.", In ACM CSCW , pages 175–186, 1994.

[5] K. Miyahara and M. J. Pazzani., "Collaborative filtering with the simple bayesian classifier.", In PRICAI, pages 679–689, 2000.

[6] J. Rennie and N. Srebro., "Fast maximum margin matrix factorization for collaborative prediction.",In ICML, 2005.

[7] Z. Huang, H. Chen, and D. Zeng., "Applying associative retrieval techniques to alleviate the sparsity problem in collaborative filtering", ACM TOIS , 22(1):116–142, Jan 2004

[8] MubbashirNazir, Umar Bashir Mir and Anuj Yadav, "A Mechanism for Mining Top ImpactedSegments in Web Analytics based on the Variants used for AB Testing", International Journal ofApplication or Innovation in Engineering & Management (IJAIEM),pp66-71vol 3, Issue2,February 2014.

[9] Francesco Ricci · Lior Rokach · Bracha Shapira ,Paul B. Kantor "Recommender Systems Handbook" Page No. 35-37, Springer Science+Business Media, LLC 2011.

[10] Andre Luiz Vizine Pereira , Eduardo Raul Hruschka , "Simultaneous co-clustering and learning to address the cold start problem in recommender systems",Knowledge-Based Systems 82 (2015) 11–19