State-of-Art Survey on Recommender Systems: Techniques and Issues

S. Archana[1], K. Indira[2]
[1]PG Student, [2]Assistant Professor

Abstract - Recommender systems are the techniques used to predict the rating of one individual will give to an item or social entity. There are wide ranges of items on which individuals have different choices. It aims to deliver users with personalized online item or service recommendations to handle the increasing online information overhead problem and develop customer relationship management. In this paper, we have highlighted the techniques of recommender systems with its merits, demerits, survey on issues with its solutions and its research issues.

Keywords: Recommender systems, issues, techniques.

I. INTRODUCTION

Recommender systems have become very popular in modern years and used in various web applications. They have become an important research area. They can be defined as programs that recommend the most apt items to specific users by predicting a user's interest in an item based on allied information about the items, the users and the relations between items and users [1].

Recommender systems are a subdivision of information filtering system used to forecast 'rating' or 'preferences' that a user would give to an item or social factor. Recommendation systems concern data mining techniques and prediction algorithms to foresee users' curiosity on information and products among the tremendous amount of offered items. These recommender systems are used widely to overcome the information overload problem.

The first recommender system was developed by Golderg, Nichols and Terry in1992. Tapestry an electronic messaging system which allowed users to either rate messages as "good or bad". The applications of recommender systems include recommending images, music, movies, websites, television programs, books, documents, conferences, news, tourism spots and learning materials [2]. It also adapts the areas of e-commerce, e-learning, e-government, e-library and e-business services.

II. RECOMMENDER SYSTEMS CLASSIFICATION

Recommender systems are classified into variety of approaches Fig 1. that have been used to provide recommendations for users.

1) Content-based recommender systems

Content-based (CB) approach recommends articles or products that are similar to items previously preferred by a particular user. The vital principles of CB recommender systems are:-

i. To explore the depiction of the items preferred by a specific user to determine the primary common preference that can be used to differentiate these items. These preferences are stored in a user profile.
ii. To evaluate each item's preferences with the user profile so that only items that have a high degree of similarity with the user profile will be recommended [3].

This recommender system uses two techniques to generate recommendations. One technique is using information retrieval methods and other technique is using statistical learning and machine learning methods.

2) Collaborative filtering-based recommender systems

Collaborative filtering (CF)-based approach help specific user to make choices based on the ratings of other people who share similar interests. The CF technique can be classified as:-

2.1 Memory-based Collaborative filtering approach

This approach uses item-to-item or user-to-user correlations to make prediction for user on upcoming items. Prediction is done using the complete or a section of the user-item database. The similarity among users or items can be calculated by cosine-based similarity, Pearson correlation-based similarity [4], constrained Pearson correlation-based similarity. This approach can be classified into *two* types:

- *User-based CF-* specific user will receive recommendations of items liked by similar users.
- *Item-based CF-* specific user will receive recommendations of items that are related to those they have liked in the earlier period.

2.2 Model-based Collaborative filtering approach

This approach gives recommendations to users based on learned models. It analyzes the training data, review the complicated patterns into the learned models, and then make predictions based on those models. This approach can be classified into two types:

- *Cluster Model*- collection of preferences is taken by definite groups or kind of users that are similar in their clusters[8].
- *Bayesian Network*-items are represented as nodes and the feasible rating value is determined from the state of each node.

3) *Knowledge-based recommender systems*

Knowledge-based (KB) approach recommends items to users based on the explicit knowledge about the users, items and/or their relationships [5]. Common expressions of KB approach are retaining functional knowledge base, case-based reasoning, ontology-based.

4) *Hybrid recommender systems*

Hybrid approach recommends users by combining best features of two or more recommendation methods into a hybrid method to achieve synergism between them. Depending on data characteristics and domain, several hybridization techniques are weighted, mixed, switching, feature combination, cascade, feature augmentation and meta-level are used. Different ways of hybridization [6]:

i. Implementing CF and CB individually and combine their applications
ii. Incorporating several content based characteristics into collaborative approach
iii. Incorporating several collaborative characteristics into content based approach
iv. Construct a general consolidative model that incorporate both CF and CB characteristics

5) *Demographic recommender systems*

Demographic recommendation approach uses information about user only. The demographic types are age, gender and knowledge of languages, ethnicity, disabilities, employment status, mobility, home ownership and location. This approach recommends items to the users according to the demographic similarities of the users[7].

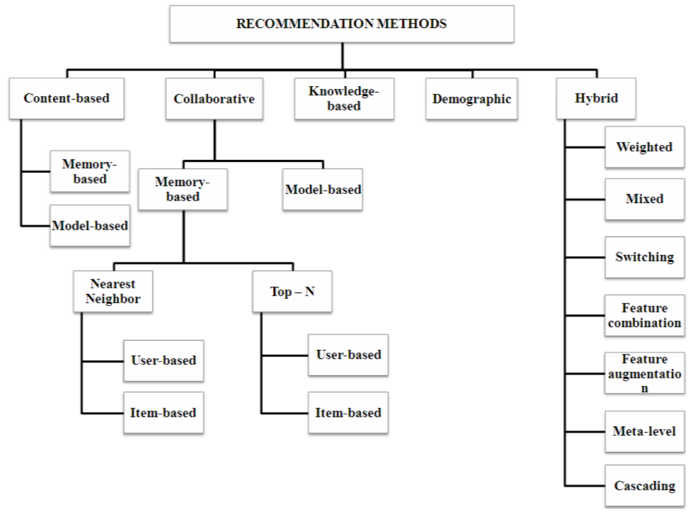

Fig 1. Recommender Systems classification

III. STATE-OF-ART SURVEY ON ISSUES AND SOLUTIONS

1) *Cold start*

Cold start problem refers to the state when a new user or item just enters the system [6]. Kinds of cold start problems are: - new user problem and new item problem. It's challenging to give recommendations for new user as there is very little information about the user and also for new item as there are no ratings for that item.

TABLE 1. SURVEY ON COLD START

Paper Name/Year	Author	Problem Characteristics	Methodology
Enhancing Collaborative Filtering by User Interest Expansion via Personalized Ranking, IEEE 2012	Qi Liu et al [11]	Exploiting the information about the user's interaction with the systems; the information about latent user interests is largely underexplored	iExpand , collaborative filtering-based recommender system by user interest expansion via personalized ranking
Cold-Start Recommendation Using Bi-Clustering and Fusion for Large-Scale Social Recommender Systems,	Daqiangzhang et al [12]	Cold-start problem that denotes a situation that social media sites fail to draw recommendation for new items, users or both	Bi-clustering and fusion(BiFu) a newly fashioned scheme

IEEE 2013			
Personalized Recommendation Combining User Interest and Social Circle, IEEE 2014	Xueming Qian et al	The new factors of social network like interpersonal influence and interest based on circles of friends bring opportunities and challenges for recommender system (RS) to solve the cold start	Three social factors, personal interest,interpersonal interest similarity, and interpersonal influence, fuse into a unified personalized recommendation model based on probabilistic matrix factorization

2) *Sparsity*

Usually, data of recommender system like MovieLens is represented in form of user-item matrix colonized by ratings given to movies and as number of users and items increases the matrix extent and sparsity evolves. The main reason of data sparsity is that the majority of users do not rate most of the items and the offered ratings are usually sparse [9]. Collaborative filtering is dependent over the rating matrix in most cases hence it leads to this problem.

TABLE 2. SURVEY ON SPARSITY

Paper Name/Year	Author	Problem Characteristics	Methodology
A Cross-Domain Recommendation Model for Cyber-Physical Systems, IEEE 2013	Sheng gao et al[13]	Suffer from the data sparsity problem	Cross-domain recommendation model, domain specific rating patterns in each domain involving discriminative information propitious to performance improvement.
TST: Threshold Based Similarity Transitivity Method in Collaborative Filtering with Cloud Computing, 2013	FengXie et al	Inaccurate similarities derived from the sparse user-item associations would generate the inaccurate neighborhood for each user or item	Threshold based Similarity Transitivity (TST)method
An Efficient Non-Negative Matrix-Factorization-Based Approach to Collaborative Filtering for Recommender Systems, 2014	XinLuo et al	Current non-negative MF (NMF) models are mostly designed for problems in computer vision, while CF problems differ from them due to their extreme sparsity of the target rating matrix	NMF-based CF model with a single-element-based approach
A Memory-based Collaborative Filtering Algorithm for Recommending Semantic Web Services, IEEE 2013	Coello et al [14]	Sparse rating data	Similarity between users is computed using the Pearson correlation coefficient, extended to consider also the ratings of users for similarity services

3) *Overspecialization*

Overspecialization problem is the system that can only recommend items based on a user's profile; the user is limited to being recommend items that are similar to those already rated.

TABLE 3. SURVEY ON OVERSPECIALIZATION

Paper Name/Year	Author	Problem Characteristics	Methodology
Providing Entertainm	Yolanda Blanco	Due to the use of syntactic	Applying reasoning

ent by Content-based Filtering and Semantic Reasoning in Intelligent Recommender Systems, 2008	Fernández et al	similarity metrics, these systems elaborate overspecialized recommendations including products very similar to those the user already know.	techniques borrowed from the Semantic Web
Enhancing Collaborative Filtering by User Interest Expansion via Personalized Ranking, IEEE 2012	Qi Liu et al [11]	Exploiting the information about the user's interaction with the systems; the information about latent user interests is largely underexplored	iExpand , collaborative filtering-based recommender system by user interest expansion via personalized ranking

4) *Performance & Scalability*

The important issues for recommendation systems are performance and scalability as e-commerce websites must be capable to find out recommendations even though the increase of users and items. Scalability is the property of system indicates its ability to handle growing amount of information in a smooth manner [10]. With enormous growth in information over internet it is obvious that the recommender systems are having an explosion of data.

TABLE 4. SURVEY ON PERFORMANCE & SCALABILITY

Paper Name/Year	Author	Problem Characteristics	Methodology
A Hybrid Recommendation Method with Reduced Data for Large-Scale Application, 2010	Sang Hyun Choi et al	Scalability and sparsity are major problems in large-scale recommendation systems.	Hybrid recommendation algorithm HYRED, which combines CF using the modified Pearson's binary correlation coefficients with CB filtering using the generalized distance-to-boundary-based rating
KASR: A Keyword-Aware Service Recommendation Method on MapReduce for BigData Applications, 2014	Shunmei Meng et al [16]	Present the same ratings and rankings of services to different users without considering diverse users' preferences, and therefore fails to meet users' personalized requirements	Keyword-Aware Service Recommendation method, named KASR
LARS*: An Efficient and Scalable Location-Aware Recommender System, 2014	Mohamed Sarwat et al [17]	Spatial ratings for non-spatial items, non-spatial ratings for spatial items and spatial ratings for spatial items	LARS*, a location-aware recommender system

5) *User input consistency*

Recommendation methods like collaborative filtering or demographic that work with user-to-user correlations depend on more correlation coefficients among the users in a dataset. Users can be classified into three types based on their correlation coefficients with other users.

i. White sheep- there is high rating correlation with many users
ii. Black sheep- there are only few or no correlating users
iii. Gray sheep- low correlation coefficients with many users are resulted from users having dissimilar opinions or an unusual taste.

6) *Privacy*

Privacy has been the most important problem in recommender systems. To provide personalized recommendations, recommendation services must know something regarding to users. The more the systems know, the more accurate the recommendations can be obtained. These privacy impinge on both the collection of explicit and implicit data. Concerning explicit data, users are not interested to reveal information about themselves and their interests. If questionnaires get too delicate, users may give bogus information in order to shield their privacy.

TABLE 5. SURVEY ON PRIVACY

Paper Name/ Year	**Author**	**Problem Characteristics**	**Methodology**
An algorithm for efficient privacy-preserving item-based collaborative filtering	Dongsheng et al [15]	Privacy issue arises in this process as sensitive user private data are collected by the recommender server	Efficient privacy-preserving item-based collaborative filtering algorithm is used, which can protect user privacy during online recommendation process without compromising recommendation accuracy and efficiency

IV. PROS AND CONS OF RECOMMENDATION METHODS

Techniques	***Advantages***	***Disadvantages***
Content-based filtering	• No need for data of other users • No cold start and sparsity • Able to recommend new and unpopular items • Does not depend on the user ratings of items in the database	• Overspecialization • Limited content analysis • New user problem • Unable to exploit quality judgments of other users.
Collaborative filtering	• Recommend most appropriate items to users which are personalized at the same time • Accuracy of their prediction increases enormously as and when more user preferences are added to the database	• System would not be very effective when user preferences change unexpectedly • Data sparsity • User input consistency • Gray sheep problem
Knowledge-based	• Sensitive to changes of preference • Mapping from user needs to products	• User must input utility function • Knowledge engineering is required
Demographic	• Not based on user-item ratings, it gives recommendation before user rated any item • Domain independent since item feature is not needed	• Privacy issues occur because of gathering demographic data • Gray sheep problem

V. FUTURE EXERTION

Based on CB, CF and hybrid recommendation methods several recommendation systems have been proposed and as of now most of them are able to solve the problems while providing better recommendations. However, due to information overhead it is required to work on this research area to explore and provide new methods that can provide recommendation in a wide range of applications.

REFERENCES

[1] J.Bobadilla, F.Ortega, A.Hernando, A.Gutiérrez, "Recommender systems survey", Knowledge-Based Systems, 2013
[2] Jie Lu, Dianshuang Wu, Mingsong Mao, Wei Wang, Guangquan Zhang, "Recommender system application developments: A survey", ELSEVIER, 2015
[3] M. Pazzani, D. Billsus, Content-based recommendation systems, The Adaptive Web, Springer, Berlin Heidelberg 2007.

[4] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, J. Riedl, “GroupLens: an open architecture for collaborative filtering of netnews”, Proceedings of the ACM Conference on Computer Supported Cooperative Work, ACM, Chapel Hill, North Carolina,USA, 1994.

[5] R. Burke, “Hybrid recommender systems: survey and experiments, User Modelling and User-Adapted Interaction”, 2002.

[6] G.Adomavicius, A.Tuzhilin, “Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions”, IEEE Transactions on Knowledge and Data Engineering, 2005.

[7] M. Pazzani, “A framework for collaborative, content-based and demographic filtering”, Artificial Intelligence Review, 1999.

[8] Breese. J. S, Heckerman. D, Kadie. C., “Empirical Analysis of Predictive Algorithms for Collaborative Filtering.” Microsoft Research, 1998.

[9] Yibo Chen, Chanle Wu, Ming Xie and Xiaojun Guo, “Solving the Sparsity Problem in Recommender Systems Using Association Retrieval”, Journal of Computers, 2011.

[10] Manos Papagelis, Ioannis Rousidis, Dimitris Plexousakis, Elias Theoharopoulos, “Incremental Collaborative Filtering for Highly-Scalable Recommendation Algorithms”, ISMIS 2005.

[11] Qi Liu, Enhong Chen, Hui Xiong, Chris H. Q. Ding and Jian Chen, “Enhancing Collaborative Filtering by User Interest Expansion via Personalized Ranking”, IEEE Transactions on Systems, man and Cybernetics Vol.42, No.1, February 2012.

[12] Daqiang zhang, Ching hsienhsu, Min chen, Quanchen, Naixuexiongand Jaime lloret, “Cold-Start Recommendation Using Bi-Clustering and Fusion for Large-Scale Social Recommender Systems”, IEEE transactions on Emerging Topics in Computing, 2014.

[13] Sheng gao, Haoluo, Da chen, Shantao li, Patrick gallinari, Zhanyu ma, and Jun guo, “A Cross-Domain Recommendation Model for Cyber-Physical Systems”, IEEE Transactions on Emerging Topics in Computing, 2013.

[14] J. M. A. Coello, Y. Yuming and C. M. Tobar “A Memory-based Collaborative Filtering Algorithm for Recommending Semantic Web Services”, IEEE Latin America Transactions, Vol.11, No.2, March 2013.

[15] Dongsheng Li, Chao Chen, Qin Lv, Li Shanga, Yingying Zhao, TunLu, NingGu, “An algorithm for efficient privacy-preserving item-based collaborative filtering”, ELSEVIER 2014.

[16] Shunmei Meng, Wanchun Dou, Xuyun Zhang, and Jinjun Chen, “KASR: A Keyword-Aware Service Recommendation Method on MapReduce for Big Data Applications”, IEEE Transactions on Parallel and Distributed Systems, Vol.25, No.12, December 2014.

[17] Mohamed Sarwat, Justin J. Levandoski, Ahmed Eldawy, and Mohamed F. Mokbel, “LARS*: An Efficient and Scalable Location-Aware Recommender System”, IEEE Transactions on Knowledge and Data Engineering, Vol.26, No.6, June 2014.