A Survey on Recommender System

[1]Ms. Rachana R. Jadhav, [2]Prof. N. R. Wankhade
[1]M. E. Student, [2]Asso. Professor, [1,2]Dept. of Computer Engg. Late G. N. Sapkal COE, Nashik, Maharashtra, India.
[1]*rachana1.jadhav@gmail.com*, [2]*nileshrw_2000@yahoo.com*

Abstract: Recommender Systems are software tools and techniques which provide the suggestions to a user which are closely similar to their needs which makes the ease in the decision-making processes. In this paper, review of research-paper recommender systems available in published literature have been carried out and also presented some survey of the filtering systems. Further, discussion about most commonly used recommendation systems also presented. The literature study reveals that more than half of the recommendation approaches applied for content-based filtering, collaborative filtering, graph-based recommendation concepts including stereotyping, item-centric recommendations and hybrid recommendations.

Keywords: recommender systems, filtering, Stereotyping.

I. INTRODUCTION

With the rapid emergence of big scholarly data, tremendous growth of knowledge is now largely captured in digital form and archived all over the world. Archival materials are also currently being digitized and provided online to people for free or by paying a fee. Such situation aims the problem of information overloading specially in academia. For example if a researcher wants to find the any article related his research he is writing its need to choose and select papers from huge quantity of data. The process of filtering is generally tiresome and time consuming; therefore, so many researchers have focused their attention on the ease of existing process system and make it more useful by providing better recommendation to the users. In the last few years, much more research papers were published about research-paper recommender systems. There are various techniques, algorithms and approaches used by the scholars on this problem.

II. LITRATURE SURVEY

A. Survey of the recommendation classes

Apart from collaborative and content-based filtering, there are many more recommendation classes like feature-based, knowledge-based, behavior-based, citation-based, context based, ruse-based. We consider the following classes to be most commandeered for distinguishing the approaches in the field of research-paper recommender systems:

a) Stereotyping

Stereotyping is one of the earliest user modeling and recommendation classes. Stereotyping[3] techniques allow the definition of a set of differentiating characteristics for a group of users; when a new user is introduced into the system, they can be assigned to a predefined stereotype, based on their personal data, which allows the activation of a set of default preferences that may be further refined over time thanks to user profile adaptation methods [1,7]. Personalization solutions exploiting user characteristics can be used in combination with more sophisticated techniques to provide a first simple step in a hybrid filtering process, or to bootstrap a content based filtering algorithm by using stereotype profiles. Beel et al. [11][12]used stereotypes as a fallback model when different recommendation approaches couldn't deliver recommendations. Author report mediocre performance of the stereotype approach with click-through rates (CTR) around 4%, whereas their content based filtering

approaches achieved CTRs over 6%. instance, Weber and Castillo notice that feminine users were usually sorting out the composer Richard Wagner when they entered the search query 'Wagner' on Yahoo! [14]. In contradiction, male users entering the same query were usually looking for the Wagner paint sprayer. Weber and Castillo modified the search algorithm to show the Wikipedia page for Richard Wagner to female users, and the homepage of the Wagner paint sprayer company to male users searching for 'Wagner.' As a result, user's satisfaction become greater.

Shani et. Al. [4]uses hybrid approach for recommendations, in which user profile is a weighted combination of user stereotypes, created automatically through a clustering process. Each stereotype is defined by an ontology of item attributes. Recommendations are computed based on the stereotype and then normalized given the user affinity to a stereotype.

b) Content-based filtering

Content-based filtering (CBF) is one of the most widely used and researched recommendation class[3]. Content-based recommendation concludes users' interests based upon the contents of the papers and a profile of the user's interests. Chandra sekaran et al. [2] present a profile-based method to recommend papers to users according to their profiles stored in Cite Seer.

c) Collaborative filtering

The simplest and original implementation of this approach recommends to the active user the items that other users with similar tastes liked in the past. The similarity in taste of two users is calculated based on the similarity in the rating history of the users. This is the reason refers to collaborative filtering as "people-to-people correlation." Collaborative filtering is considered to be the most popular and widely implemented technique in Recommender System. Wang and Blei[1,7] proposed a collaborative topic regression model for article recommendation, in which each user was represented with interest's distribution and each article was described using content-based item topic distribution. The collaborative topic regression (CTR) model. CTR combines traditional collaborative filtering with topic modeling to fit a model that uses the latent topic space to explain both the observed ratings and the observed words. In[5], a system is based on Collaborative Filtering where a technique which is used in recommending systems. A set of features is computed for each paper contains both content and stylometric features.

d) Co-occurrence recommendations

To give co-occurrence recommendations, the items those are recommended frequently co-occur with some source items. Wartena et al. [14] recommends Co-occurrence as a stable measure for tag similarities. Distance measure between tags it is straightforward to derive methods to analyze user interest and compute recommendations. Movielens dataset and a dataset of tagged books is used for the evaluations of tag based recommendations

e) Graph based

In the data is represented in the form of a graph in graph based approach where nodes are users, items or both, and edges encode the interactions or similarities between the items and users. In this method various similarity approaches are used. Those are

1. Path-based similarity

In path-based similarity, the distance between two nodes of the graph is evaluated as a function of the number of paths connecting the two nodes, as well as the length of these paths.

2. Random walk similarity

Within a probabilistic framework transitive associations in graph-based methods can also be defined. In this framework, the similarity between users or items is evaluated as a probability of reaching these nodes in a random walk. A random walk in the graph is actually a transition from a vertex to another vertex.

3. Item rank

A recommendation approach, based on the PageRank algorithm for ranking Web pages [11], is ItemRank [16]. This approach ranks the preferences of a user u for new items i as the probability of u to visit i in a random walk of a graph in which nodes correspond to the items of the system, and edges connect items that have been rated by common users.

Liu et al. [10]determine important nodes like authors and venues on a heterogeneous bibliographic graph are the pseudo relevance feedback (PRF) algorithm is used. Then, a random walk algorithm was run. It compute the ranking scores of an article innovative publication ranking method with PRF by leveraging a number of meta-paths on the heterogeneous bibliographic graph. Different meta-paths on the graph address different ranking hypotheses, whereas the pseudo-relevant papers (from the retrieval results) are used as the seed nodes on the graph .Jiang et al.[8] employees the paper citation graph to generate a set of potential relevant papers. H. Liu et al[15] Proposes the citation based scientific article recommendation which combines information of researchers historical preferences and citation relation between article .In this weak citations are first filtered out and then remaining are incorporated into graph based article ranking.

f) Global relevance

In its very simplest form of a recommender system adopts a one fits all approach and recommends items that have the highest global relevance. In the global relevance the relevance is not calculated global measures like overall popularity. For example, a movie-rental system could recommend those movies that were most often rented or that had the highest average rating over all users. In this case, the basic assumption would be that users like what most other users like. From the reviewed approaches, none used global relevance exclusively but many used it as an additional ranking factor. Yupeng Guet al.[16] propose a unified method which can simultaneously learn the weights of multiple content matching signals, as well as global term weights for specific recommendation tasks.

g) Hybrid recommendation approaches

Hybrid recommendation approaches may be the combination of previously introduces recommendation classes. Many recommender systems combine different techniques of collaborative approaches and content based approaches for the better result. The combination of approaches can proceed in different ways [7]:

In 2013M. Kaminskas et al.[6] suggested hybrid approach in which two techniques are combined– one based on representing both POIs and music with tags, and the other based on the knowledge of the semantic relations between the two types of items which exploit orthogonal types of relations between places and music tracks.

III. PROBLEMS AND CHALLENGES

A. Sparsity [17]

In e-commerce shops people purchase few items from the huge amount of items. There are almost all users that have rated just a few items even though the most popular items have very few rating. If a user has access just few items then its quite difficult to determine his choice and he/she could be related to the incorrect neighborhood. Sparsity is the problem of lack of information input from user.

B. Privacy [17]

Privacy has been the most important problem. Many online shops offer effective protection of privacy of the users by utilizing specialized algorithms and programs. To receive the most accurate and correct recommendation, the system must receive the most amount of information possible about the user, including demographic data, and data about the location of a particular user. Naturally, the question of reliability, security and confidentiality of the given information arises.

C. Cold Start [17]

The problem of cold start appears at early stages of a recommender system's lifecycle .When there is a less information available, ontologies are a proven tool for knowledge extension. [Middleton 2004] If there is a little information about content then the Cold Start problem affects every recommender system, the content-based filtering will behave poorly; the same applies to collaborative filtering. If it happens to have no information recognizable by content-based methods, and there is no history in the system, the hybrid approach will produce nearly random recommendations as well.

D. Loss of neighbor transitivity [17]

When user 1 correlated with user 2 and user 2 correlated with user 3 then possibility of user 3 also correlated to user 1. Such type of relationships are not captured by recommender systems, but can be captured with knowledge of users from, for instance, ontology.

E. Synonymy [17]

Changing word "user" to "item" in the assumption above, the overseen relations between items A and C might be uncovered in a same manner utilizing ontologies. For example, an item "Laptop ASUS" and an item "Bag for laptops" are both neither laptops nor bags, but both are of class "IT products" and, therefore, have a degree of commonality overseen by content-based and collaborative filtering methods. Ontologies uncover such overseen transitivity between objects.

F. A recommendation quality [17]

It plays one of the important role in recommender systems. Among other details, the user is sensible for incorrect recommendations, which the user does not like. If the recommender system will rate any movie as bad one, but the user likes it, the prediction will be a false negative. In such cases users lose trust in the system and stop using it. Therefore, it is important to keep recommendation quality at the higher t possible level.

G. Scalability [17]

Scalability is another important impact in RSs. As ratings database grows, the performance decreases. It is beneficial to try to make systems, which can handle large amounts of data and produce accurate recommendations quickly. There has been seen a trade-off between performance and the prediction accuracy.

IV. PROPOSED SYSTEM

To propose a novel recommendation method, which incorporates common author relations between articles to help generate better recommendations for relevant target researchers using side-information. Such side-information may be of different kinds, such as document provenance information, the links in the document, user-access behavior from web logs, or other non-textual attributes which are embedded into the text document.

In this research new guidelines are propose which combines classical partitioning algorithms with probabilistic model in order to create an effective clustering approach.

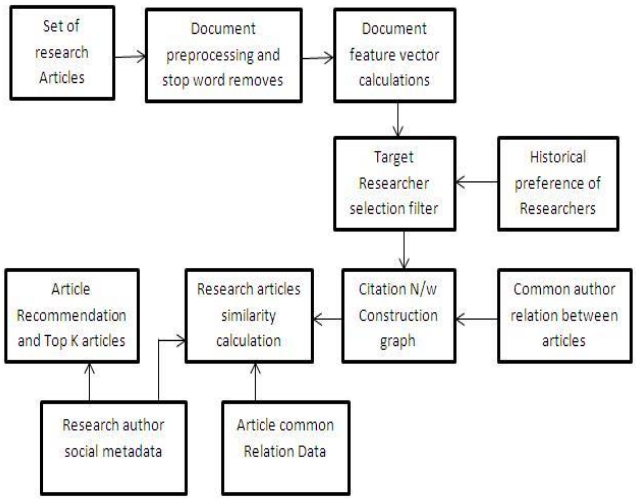

Fig. 1 Proposed System

VI. CONCLUSION

In this paper , thorough review of the various techniques and algorithms used to recommender system are presented .Further various modern recommendation classes such as Collaborative Filtering, Content-based Filtering, Hybrid, Stereotyping, Co-Occurrence, Graph-based and Global

Relevance also studied. The problems and challenges of recommender system such as privacy, sparsity, cold start are also identified.

ACKNOWLEDGMENT

It gives us great pleasure in presenting the preliminary project report on "A Survey on Recommender System". I would like to take this opportunity to thank my internal guide Assisted Prof. N. R. Wankhade for giving me all the help and guidance I needed. I am really grateful to them for their kind support. Their valuable suggestions were very helpful.

REFERENCES

[1] C. Wang and D. M. Blei, "*Collaborative topic modeling for recommending scientific articles*," in Proceedings of the 17th ACM SIGKDD, 2011, pp. 448– 456.

[2] Chandrasekaran, Kannan, Susan Gauch, Praveen Lakkaraju, and Hiep PhucLuong. *Concept-based document recommendations* for citeseer authors. In Adaptive Hypermedia and Adaptive Web-Based Systems , page 83-92, 2008.

[3] Ricci, F., Rokach, L., Shapira, B., Kantor, P.B. (eds): Recommender Systems Handbook, pp. 1–35. Springer, Berlin (2011)

[4] Guy Shani and AmnonMeisles and Yan Gleyzer. *A Stereotypes-Based Hybrid Recommender System for Media Items*. Association for the Advancement of Artificial Intelligence.

[5] Z. Yang and B. D. Davison, "*Venue recommendation: Submitting your paper with style*," in Machine Learning and Applications (ICMLA), 2012 11th International Conference on, vol. 1, pp. 681–686, IEEE, 2012.

[6] M. Kaminskas, F. Ricci, and M. Schedl, "*Location-aware music recommendation using auto-tagging and hybrid matching*," in 7th ACM Conference on Recommender Systems, 2013, pp. 17–24.

[7] A. Elgohary, H. Nomir, I. Sabek, M. Samir, M. Badawy, and N. A.Yousri, "*Wiki-rec: A semantic-based recommendation system usingwikipedia as an ontology*," in Intelligent Systems Design andApplications(ISDA), 2010 10th International Conference on, 29 2010.

[8] Y. Jiang, A. Jia, Y. Feng, and D. Zhao, "*Recommending academic papers via users' reading purposes,*" in Proceedings of the 6th ACM Conference on Recommender Systems, 2012, pp. 241–2

[9] International Journal on Digital Libraries, pp. 1–19, 2014. [33] C. Wang and D. M. Blei, "*Collaborative topic modeling for recommending scientific articles*," in Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2011, pp. 448–456.

[10] X. Liu, Y. Yu, C. Guo, and Y. Sun, "*Meta-path-based ranking with pseudo relevance feedback on heterogeneous graph for citation recommendation*," in Proceedings of the 23rd ACM International Conference on Information and Knowledge Management, 2014, pp. 121–130.

[11] Hofmann, T.: *Collaborative filtering via gaussian probabilistic latent semantic analysis*. In: SIGIR '03: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 259–266. ACM 2003.

[12] Rich, *E.:Usermodeling via stereotypes*. Cogn. Sci. 3(4), 329–354(1979)

[13] JoeranBeel • BelaGipp • Stefan Langer •Corinna Breitinger," *Research-paper recommender systems:a literature survey*", Springer-Verlag Berlin Heidelberg 2015

[14] Weber, I., Castillo, *C.: The demographics of web searc*h. In: Proceeding of the 33rd international ACM SIGIR conference on Research and development, pp. 523–530 (2010)

[15] H. Liu, Z. Yang, I. Lee, Z. Xu, S. Yu, and F. Xia, "*Car: Incorporatingfiltered citation relations for scientific article recommendation*,"in The 8th IEEE International Conference on Social Computing and Networking (SocialCom), China, Dec 2015.

[16]YupengGu, David Hardtke, Bo Zhao, Yizhou Sun "*Learning Global Term Weights for Content-based Recommender Systems*" in the the International World Wide Web Conference Committee (IW3C2),Canada. April 2016,

[17] Cover, T.,and Hart, P., *Nearest neighbor pattern classification*. Information Theory, IEEE Transactions on, 13(1):21–27, 1967.

