A Survey on Friend Recommendations in a Social Bookmarking System Using Serendipity

Hiral R. Patel[1], Shital A. Solanki[2]

[1]*PG Scholar, Information Technology Department, L. D. College of Engineering, Ahmedabad, Gujarat-India*

[2]*Assistant Professor, Information Technology Department, L. D. College of Engineering, Ahmedabad, Gujarat-India*

***Abstract*— Obtaining recommendations from trusted sources is a critical component of the natural process of human decision making. Social media systems allow users to share resources with the people connected to them. In order to handle the fast growth of the content in these systems and of the increasing amount of users, recommender systems have been introduced. A form of social media, known as Social Bookmarking System, allows to share bookmarks in a social media. It also allows users to use tags (keywords) to describe resources that are of interest for them, helping to organize and share these resources with other users in the network. By analyzing users with a similar behavior (i.e. users who have a large amount of tags and bookmarks in common), accurate friend recommendations can be produced that are both novel and serendipitous.**

***Keywords*— Recommender systems, Social bookmarking, Serendipity, Tags, Bookmarks**

I. INTRODUCTION

Recommender systems help users to find out interesting and useful items from a large amount of information pool. To produce this, a Friend Recommender System that operates in the social bookmarking application domain and is based on behavioral data mining, i.e., on the exploitation of the users activity in a social bookmarking system must be developed. This type of mining is able to produce accurate friend recommendations, allowing users to get to know bookmarked resources that are novel, useful and serendipitous. When users' tagging behavior is same and are also interested in the same content, they can be recommended as friends.

The rest of the paper is organized as follows. Related literature and challenges are explained in section II. Friend recommendations in a social bookmarking system are presented in section III. Recommendation systems as a research problem are presented in section IV. Measurements and evaluations of recommendation systems are explained in section V. Discussions are given in section VI.

II. RELATED LITERATURE AND CHALLENGES

Recent literatures evaluate different types of recommendation system technologies, and their real life applications. [3] introduces and classifies recommendation system into three main techniques: content-based, collaborative and hybrid.

Content-based recommender systems continue to collect the user's information and preference; analyze and establish the user profile; finally the user will be recommended items which are similar with that profile.

Collaborative recommender systems continue to collect the user's information; analyzes and calculates similarity, the similarity is used to find out who have similar interests with that user; finally the user will be recommended items that according to the preference and access records of people who have similar preferences.

Hybrid recommender systems combine content-based and collaborative methods. There are different kinds of methods implementing hybrid recommender systems. First, they implement content-based and collaborative methods respectively, and then combining recommendations from each method. Second, they incorporate some content-based features into the collaborative method. Third, they incorporate some collaborative features into the content-based method. Fourth, they construct general models that incorporate both content based and collaborative methods.

Privacy preserving in recommendation system has been a challenge nowadays. Recommendation systems exploit users' data to generate personalized recommendations. This clearly has negative impacts on the privacy of the users [10].

Diversity of the items recommended to a target user is another issue to discuss. In a recommended list, it is more likely that the user will find a suitable item if there is a certain degree of diversity among the items. There are many situations, in the early stage of a recommendation process, in which the users want to explore new and diverse directions. In such cases, the user is using the recommender as a knowledge detection tool [13]. This is an issue that needs to be incorporated into the evaluation of system.

III. FRIEND RECOMMENDATIONS IN A SOCIAL BOOKMARKING SYSTEM

Friend Recommender System operates in the social bookmarking application domain and is based on behavioral data mining. In this system, by analyzing users with a similar behavior (i.e., users who have a large amount of tags and

resources in common), accurate recommendations can be produced. These recommendations, at the same time, can lead to novel, useful and serendipitous bookmarks. Using this approach, the impact of the "interaction overload" and the "over-specialization" problems is strongly reduced.

A social bookmarking system is composed by: [2]

- A set of users
- A set of resources: These resources characterize the type of social bookmarking system and they might be of different types (e.g., web pages)
- A set of tags: Which are the keywords used to describe the resources
- A set of bookmarks: Which are represented as triplets (user, resource, tag); these triplets are known either as tag assignments, or as tag applications
- A set of connections among users: Which are represented as couples (user, user). Depending on the type of connection among two users, a couple might be ordered (i.e., users are connected by a follow relation), or not (i.e., users are friends and mutually follow each other).

  Once a user decides to bookmark a resource by tagging it, these bookmarks are shown to the users who are friends with or follow this user.

Social bookmarking systems also offer privacy options, which allow to keep a bookmark private, or to share it only with a limited number of users.

Features that allow to explore the tags and to facilitate the management of the bookmarks, like their export from browsers and the possibility to add a bookmark to the profile by email, are often offered.

The objective of a Friend Recommender System in the social bookmarking domain should consider the following aspects:

- In order to contain the complexity of a recommender system, no personal information could be used to produce the recommendations. A user can add as a friend or follow another user, in order to receive her/his newly created bookmarks.
- Social media systems grow rapidly. This means that the amount of information added to a social media system and the user population increase at a fast rate. A recommender system that operates in this context needs to build accurate profiles of the users, which have to be up-to-date with the constantly evolving demands of the users.
- The previously carried out behavioral analysis highlighted that both the tags and the resources represent a form of connection among two users. In particular, the number of common tags and resources between users is a small subset of all the tags and resources bookmarked by the users.

The tagging activity of the users reflects their interests. Therefore, the tags used by a user are an important source of information to infer the interests that describe her/him.

The system works in five steps: [1]

1) Tag-based user profiling: Given the tag assignments of each user, this step builds a user profile, based on the frequencies of the tags used by the user.
2) Resource-based user profiling: Given the tag assignments of each user, this step builds a user profile, based on the resources bookmarked by the user.
3) Tag-based similarity computation: The first metric, calculated among a target user u_t and the other users, is based on the tag-based user profile. Pearson's correlation is used to derive the similarity.
4) User interest computation: The second computed metric is the interest of a user toward another user and it is represented by the percentage of common resources among them.
5) Recommendations selection: This step recommends to u_t the users with both a tag-based similarity and a user interest higher than a threshold value.

IV. RECOMMENDATION SYSTEMS AS A RESEARCH PROBLEM

There are some common limitations and problems with recommendation systems and they could be summarized in several general categories:

- Social interaction overhead problem: Social interaction overload is a problem related to the excessive amount of users and items that each user can interact with. This leads to the scarcity of attention, which does not allow a user to focus on users or items that might be interesting for her/him. In order to filter information in the social media domain, in the last few years the research on recommendation has brought to the development of a new class of systems, named social recommender systems. These systems face the social interaction overload problem, by suggesting users or items that the users might be interested in. The filtering of the content in the social domain can be done by means of different metrics. The social similarity between users in a social network can be adopted. [1]
- Over-specialization problem (Serendipity problem): Recommender systems usually suggest items that have a strong match with the user profile, consequently the user always receives recommendations for items very similar to those that

she/he already considered and never receives suggestions for unexpected, surprising, and novel items. This limit of recommender systems, known in the literature as "serendipity problem" or "over-specialization problem", worsens the user experience and does not give the users the opportunity to explore new items and to improve their knowledge. It is known that the serendipity problem affects both the content-based and the collaborative filtering approaches. The accuracy of a recommender system is important but it is only not enough to evaluate it. Therefore, other metrics have to be considered in the evaluation of the system.

Users should be able to discover new items, i.e., diversity among the recommendations is a desired property.

When a recommender system is being designed, the aspects that have to be evaluated should have strong influence when deciding how to implement it.

- Cold-Start problem: Cold-Start problem is related to recommendations for novel users or new items. In case of new users, the system does not have information about their choices in order to make recommendations.
- Shilling: The recommender system is also suffering from the spam attacks due to the users that are interested in misleading the system to recommend a certain product.
- Sparsity: In recommender systems, not all the users rates all the items from the available items. So many cells which are in user-item rating matrix are empty. So in such a case, finding similarity among users or items becomes difficult.

  The number of items sold on major e-commerce sites is exceedingly large. The most active users will only have rated a small subset of the overall database. Thus, even the most popular items have very few ratings.
- Limited Content Analysis: This is similar to the New User Problem and many times we don't have enough information regarding the items [10].

V. MEASUREMENTS AND EVALUATIONS OF RECOMMENDATION SYSTEMS

It is important to use the correct measurements to evaluate if a recommendation system is good or poor. However, there is no single good evaluation method that has been used universally due to the complex nature of recommendation problems. A good evaluation measurement depends on the user tasks, types of analysis and datasets being used [10].

The evaluation of a recommendation system is difficult because of the diversity of the problems and the datasets. The different purposes of recommendations really determine how the systems are going to be implemented [10].

Here, I introduce few common evaluation methods.

1) Precision:
   Precision = (Relevant resources Retrieved) / (Total resources Retrieved)
2) Recall:
   Recall = (Relevant resources Retrieved) / (Total Relevant resources)
3) F-Measure:
   F-Measure = 2 * ((Precision * Recall) / (Precision + Recall))
4) Novelty:
   Novelty = (The sum of novel resources recommended to each user) / (The sum of resources recommended to each user)
5) Serendipity:
   Serendipity = (Useful Items of Unexpected Recommendations) / (Total Unexpected Recommendations)
6) Percentage of satisfied users:
   (The set of users for which a correct recommendations was produced) / (The set of users for which a recommendation was produced) * 100
7) Mean Absolute Error (MAE):
   MAE is a quantity used to measure how close forecasts or predictions are to the eventual outcomes.
   MAE measures the average absolute distance between the predicted ratings and the true ratings.
8) Coverage:
   The coverage of a recommender system is a measure of the domain of items over which the system can make recommendations.
9) Diversity:
   Diversity is the dissimilarity between all pairs of recommendations.

VI. DISCUSSIONS

Social Bookmarking System, using tags and bookmarks can produce serendipitous recommendations which are novel, interesting, relevant and surprising. When users tag in the same way and are also interested in the same content, they can be considered as friends.

Considering the tag similarity between Innovator (or Early adopter) and a target user, serendipitous bookmarks can be generated.

With the growing number of social networks, there are great potential to utilize this information to help with improving the recommendation systems. Good recommendation systems can not only improve the business outcomes but also help with reducing the information barriers for regular users.

New technologies are transforming the user behaviors and new types of social network of information may become available in the future for better recommendations.

It is and will still be a challenge to preserve the user's privacy and at the same time utilize the user information for knowledge and recommendation systems. Many network sites have explored the use of user's network information, such as Google, Facebook, Twitter and LinkedIn. However these usages are limited to their own applications and couldn't be publicly available for broader applications. There is a need for establishing an industry standard for utilizing network information while protecting user privacy [10].

REFERENCES

[1] Matteo Manca, Ludovico Boratto, Salvatore Carta, "Behavioral data mining to produce novel and serendipitous friend recommendations in a social bookmarking system", Springer US 2015.

[2] Matteo Manca, Ludovico Boratto, Salvatore Carta, "Friend Recommendation in a Social Bookmarking System: Design and Architecture Guidelines", Springer 2015.

[3] Yu-Shian Chiu, Kuei-Hong Lin, Jia-Sin Chen, "A Social Network-based Serendipity Recommender System", International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS) 2011.

[4] Feilong Chen, Jerry Scripps, Pang-Ning Tan, "Link Mining for a Social Bookmarking Web Site", IEEE 2008.

[5] Matteo Manca, Ludovico Boratto and Salvatore Carta, "Mining User Behavior in a Social Bookmarking System: A Delicious Friend Recommender System", Third International Conference on Data Management Technologies and Applications DATA 2014.

[6] R. Mehul, "Discrete Wavelet Transform Based Multiple Watermarking Scheme", in *Proceedings of the 2003 IEEE TENCON*, pp. 935-938, 2003.

[7] Peyman Nasirifard, Sheila Kinsella, Krystian Samp and Stefan Decker, "Social People-Tagging vs. Social Bookmark-Tagging", Springer Berlin Heidelberg 2010, pp 150-162.

[8] Ido Guy, Inbal Ronen, Eric Wilcox, "Do You Know? Recommending People to Invite into Your Social Network", Proceedings of the 13th international conference on Intelligent user interfaces ACM USA 2009.

[9] Matteo Manca, Ludovico Boratto, Salvatore Carta, "Producing Friend Recommendations in a Social Bookmarking System by Mining Users Content", IARIA 2013.

[10] Song Chen, Samuel Owusu, Lina Zhou, "Social Network Based Recommendation Systems: A Short Survey", IEEE 2014.

[11] Pierpaolo Basile, Domenico Gendarmi, Filippo Lanubile, Giovanni Semeraro, "Recommending Smart Tags in a Social Bookmarking System", 'Bridging the Gap between Semantic Web and Web 2.0 (SemNet 2007)', pp. 22-29.

[12] Harith Alani, Christian Bauckhage, Christian Bizer, Ciro Cattuto, Alexander Lser, Nicolas Maisonneauve, Peter Mika, Ian Mulvany, Alexandre Passant, "Mining for Social Serendipity", Dagstuhl Seminar Proceedings: Social Web Communities 2008.

[13] B. Smyth and P. McClave, "Similarity vs. diversity," in Case-Based Reasoning Research and Development. Springer, 2001, pp. 347–361.

[14] Matteo Manca, Ludovico Boratto and Salvatore Carta, "Using Behavioral Data Mining to Produce Friend Recommendations in a Social Bookmarking System", Springer 2015.

[15] John Riedl, "Research Challenges in Recommender Systems", Tutorial sessions Recommender Systems Conference ACM RecSys. 2009.

[16] Adomavicius, G., Tuzhilin, A., "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, Vol. 17, Issue 6, pp. 734-749,2005.

[17] Iván Cantador, Alejandro Bellogín, David Vallet, "Content-based Recommendation in Social Tagging Systems", Proceedings of the fourth ACM conference on Recommender systems USA 2010, Pages 237-240

[18] Prem Melville and Vikas Sindhwani, "Recommender Systems", Springer 2010

[19] Francesco Ricci, Lior Rokach and Bracha Shapira, "Introduction to Recommender Systems Handbook", Springer 2011

[20] Iman Avazpour, Teerat Pitakrat, Lars Grunske and John Grundy, "Dimensions and Metrics for Evaluating Recommendation Systems." Recommendation Systems in Software Engineering. Springer Berlin Heidelberg, 2014. 245-273