

Serendipitous Bookmark Recommendations in a Social Bookmarking System using Behavioral Data Mining

Hiral R. Patel¹ Shital A. Solanki²

¹PG Scholar ²Assistant Professor

^{1,2}Department of Information Technology

^{1,2}L.D.College of Engineering, Ahmedabad, Gujarat-India

Abstract— Social media systems allow users to share resources with the people connected to them. In order to handle the exponential growth of the content in these systems and of the amount of users that populate them, recommender systems have been introduced. A form of social media, known as Social Bookmarking System, allows users to share bookmarks in a social network. It also allows users to use tags 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 resources in common,) accurate recommendations can be produced. This paper proposes a bookmark recommender system that operates in the social bookmarking application domain and is based on behavioral data mining. Experimental results show how this type of mining is able to produce serendipitous bookmark recommendations. Using this approach, the impact of the “interaction overload” and the “over-specialization” problems is strongly reduced.

Key words: Behavioral data mining, Bookmark recommendation, Serendipity, Social bookmarking

I. INTRODUCTION

Bookmark Recommender System that operates in the social bookmarking application domain is based on behavioral data mining, i.e., on the exploitation of the users activity in a social bookmarking system. This type of mining is able to produce accurate bookmark recommendations, allowing users to get to know bookmarked resources that are both novel and serendipitous. Currently, there are no efficient approaches able to produce bookmark recommendations in the Social Bookmarking Systems domain. In this paper, a Bookmark Recommendation Algorithm for a Social Bookmarking System, using tags and bookmarks to produce serendipitous recommendations which are novel, interesting, relevant and surprising is proposed. When users tag in the same way and are also interested in the same content, their bookmarks can be recommended to their friends.

The rest of the paper is organized as follows: in Section 2 we present the categories of recommender systems in social environments; Section 3 describes user behavior mining to produce bookmark recommendations; in Section 4 we propose Bookmark Recommender System algorithm; Section 5 illustrates measurements and evaluations of recommendation systems. Section 6 illustrates the performed experiments and outlines main results; Section 7 contains comments and conclusion.

II. RECOMMENDER SYSTEMS

This section presents related work on bookmark recommendation in the social domain. These systems can be classified into three categories, based on the source of data used to build the recommendations:

A. 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.

The content-based recommender system recommends the items based on their content where user profile is in the form content in which the user is interested. For example, the keywords of purchased books of a user are used to find other books that contain the similar keywords.

Content based filtering is unable to evaluate the quality of an item. It cannot distinguish the bad items from a bad one.

B. 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.

The collaborative recommender system has become the most widely used method to recommend items for users. It makes recommendation according to the similar users with the active user or the similar items with the items which are rated by the active user. It includes memory based/ user based and model based methods.

Collaborative filtering is not so sensitive to the problems that are mentioned with content based filtering. The collaborative filtering will recommend items that have received high ratings by other users with same interests. In collaborative filtering, the user profile is nothing but the set of ratings that are given to items by different users. These ratings are stored by observing the behavior of a user with the system.

C. 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.

III. MINING USER BEHAVIOR TO PRODUCE BOOKMARK RECOMMENDATIONS

This section aims at analyzing user behavior in a social bookmarking system from a bookmark recommendation point of view. In particular, we study how the bookmarking activity of a user is related to that of the others. This analysis has been conducted on a Delicious dataset, which contains:

- 1867 users;
- 69226 URLs;
- 53388 tags;
- 7668 bi-directional user relations;
- 437593 tag assignments (i.e., tuples [user, tag, URL]);
- 104799 bookmarks (i.e., distinct pairs [user, URL]).

In order to focus on the possible connections among users, which might lead to bookmark recommendations in this system, we calculated the number of common tags and resources between the users in the dataset. [1]

IV. BOOKMARK RECOMMENDER SYSTEM

- A social bookmarking system is composed by:
- 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.

We are given a Social Bookmarking System, defined as a tuple $Q = \{U, R, T, A, C\}$, where:

- U, R and T are sets of Users, Resources and Tags
- A is a ternary relation between the set of Users, Resources and Tags, i.e., $A \subseteq U \times R \times T$
- C is a binary relation between the users, i.e., $C \subseteq U \times U$

Objective is to define a function $f : U \times U \rightarrow C$, which allows the system to derive if, given two users $u \in U$ and $m \in U$, there is an undirected connection $c \in C$ among them.

Given a target user $u_t \in U$, the system recommends the users with high tag-based user similarity and a high percentage of common resources. This in future helps to recommend serendipitous bookmarks (URLs).

This analysis has been conducted on delicious dataset.

The system works in following steps:

- 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. A user can be profiled according to his/her use of the tags, by considering the relative frequency of each tag.
- 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. This step builds another user profile, based on the resources bookmarked by each user. A user can be profiled, according to his/her bookmarked resources, by considering the fact that he/she bookmarked a resource (i.e., he/she expressed interest in it.)
- 3) Tag-based similarity computation: Pearson's correlation is used to derive the similarity. It compares the frequencies of all the tags used by the considered users. The similarity values range from 1.0, which indicates complete similarity, to -1.0, which indicates complete dissimilarity.
- 4) User Interest computation: 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 the candidate set of users to target user u_t with both a tag-based similarity and a user interest higher than a threshold value.
- 6) The candidate users ordering of each common item (resource): According to the Diffusion of Innovations Theory, innovators venture to be the first one to adopt new items or products, and introduce them into the community. Other members get to know the innovation from them. In accordance with the innovators' characteristics of early adoption, we identify potential innovators for a target user based on two principles: (i) For the co-selected item, innovators should annotate it ahead of the target user. The earlier s/he has annotated it, the more likely that s/he is an innovator for the target user; (ii) For the candidate innovators, s/he should be active to try out newly released bookmarks (URLs). In other words, an innovator should annotate the bookmarks (URLs) soon after the bookmarks (URLs) were released. Note that, however, in this interpretation, the concept of innovator can be relative to a user. That is, each user can have his own personal innovators, who are the trend-leaders in his area of interest, who have the greatest influence for that user.
- 7) The more frequently the user selected the item soon after it released, the higher the active score s/he achieves.
- 8) InterUser Score between the candidate innovator and target user: InterUser Score is the score between the candidate innovator and a target user that measures how much earlier candidate user annotated the co-selected bookmark ahead of target user. It indicates the lead time that a candidate innovator annotates a bookmark ahead of a target user.
- 9) Calculate the Likelihood that candidate user is an innovator from the perspective of target user: The candidate user who shares similar interests with the target user and is active to try out newly released bookmarks would have a higher chance to be an innovator for the target user.
- 10) Recommendation Generation: It locates the top K innovators for a target user from previous step. Bookmarks that have already selected by an innovator but not yet by the target users then constitute the candidate recommended bookmarks set for the target user. [1][20]

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. [9]

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. [9] Here, we introduce few common evaluation methods. [21]

- 1) Precision: Precision = (Relevant resources Retrieved) / (Total resources Retrieved)
- 2) Recall: Recall = (Relevant resources Retrieved) / (Total Relevant resources)
- 3) F-Measure: F-Measure = $2 * ((\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall}))$
- 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.

Serendipity is one of the metrics used in the evaluation of recommender systems. Precisely, serendipity measures how surprising the successful recommendations are. Serendipity can be seen as a way to introduce diversification in the recommendation, in order to allow users to discover new items that they did not know they were interested in and to improve their knowledge.

VI. EXPERIMENTS

A. Bookmark Recommendations for User ID: 8

User ID	Bookmark	Recommendation Priority (RP)
8	www.designishistory.com	3
8	www.frankchimero.com	4
8	www.slideshare.net	5
8	booktwo.org	4
8	paulirish.com	2

Table 1: Bookmark Recommendations for User ID 8

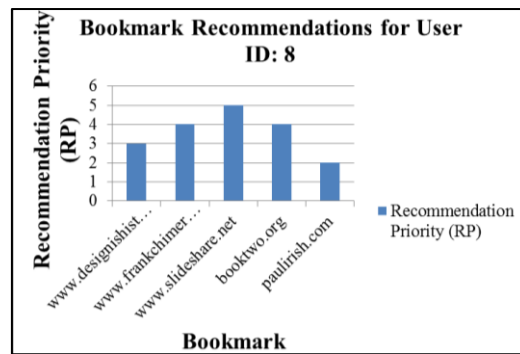


Fig. 1: Bookmark Recommendations for User ID: 8

VII. CONCLUSION

In this research work, a Bookmark Recommender System based on a form of behavioral data mining of the users in a social bookmarking system is proposed, whose intent is to suggest bookmarks with high accuracy and that allows users to come across novel and serendipitous bookmark recommendations.

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] Peyman Nasirifard, Sheila Kinsella, Krystian Samp and Stefan Decker, "Social People-Tagging vs. Social Bookmark-Tagging," Springer Berlin Heidelberg 2010, pp 150-162.
- [7] 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.
- [8] Matteo Manca, Ludovico Boratto, Salvatore Carta, "Producing Friend Recommendations in a Social Bookmarking System by Mining Users Content," IARIA 2013.
- [9] Song Chen, Samuel Owusu, Lina Zhou, "Social Network Based Recommendation Systems: A Short Survey," IEEE 2014.
- [10] 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.
- [11] Harith Alani, Christian Bauchhage, Christian Bizer, Ciro Cattuto, Alexander Lser, Nicolas Maisonneuve, Peter Mika, Ian Mulvany, Alexandre Passant, "Mining for Social Serendipity," Dagstuhl Seminar Proceedings: Social Web Communities 2008.
- [12] B. Smyth and P. McClave, "Similarity vs. diversity," in Case-Based Reasoning Research and Development. Springer, 2001, pp. 347-361.
- [13] Matteo Manca, Ludovico Boratto and Salvatore Carta, "Using Behavioral Data Mining to Produce Friend Recommendations in a Social Bookmarking System," Springer 2015.
- [14] John Riedl, "Research Challenges in Recommender Systems," Tutorial sessions Recommender Systems Conference ACM RecSys. 2009.
- [15] 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.
- [16] 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
- [17] Prem Melville and Vikas Sindhwani, "Recommender Systems," Springer 2010
- [18] Francesco Ricci, Lior Rokach and Bracha Shapira, "Introduction to Recommender Systems Handbook," Springer 2011
- [19] Iman Avazpour, Teerat Pitakrat, Lars Grunke and John Grundy, "Dimensions and Metrics for Evaluating Recommendation Systems," Recommendation Systems in Software Engineering. Springer Berlin Heidelberg, 2014. 245-273
- [20] Qianru, Zheng, "Making Serendipitous Recommendation by Providing Less Obvious Items," City University of Hong Kong, 2014
- [21] Hiral R. Patel, Shital A. Solanki, "A Survey on Friend Recommendations in a Social Bookmarking System Using Serendipity," IJLTEMAS Dec 2016, pp. 99-102.