A Friend Recommendation in a Social Bookmarking System using Behavioral Data Mining

Bhavesh Jain[1], Tejas Kadiya[2], Sumit Chaudhary[3], Ramesh Prajapati[4]
[1, 2, 3, 4] Department of CSE
[1, 2, 3, 4] IIST-Cadila Group, Ahmedabad, Gujarat

Abstract- *Now a day Social media system becoming more popular. In this System the amount of content shared increases fast, users has arise social interaction overload problem. To overcome Social interaction overload problem, Social Recommendation system have been design and develop to filter the content and suggested to user only interesting item. Over Specialization problem is related to the recommendation which are too similar to the items that already consider by the users and never receive suggestion for unexpected & novel item. This approach is able to produce accurate recommendation but it is not useful recommendation. Experimental results show how this type of behavioral data mining to generate proper friend recommendation, users has to get bookmarks resource are both novel and serendipitous.*

Keywords- Novel and Serendipitous Resources, Behavioral mining, Friend Recommendation.

I. INTRODUCTION

Social media system are web-based services that allows users to generate a public or semi-public form within bounded system, clear a list of other users with whom they share a connection, and view and cross their list of connections and those made by other within the system.

Social media system is also known as Social Bookmarking System that allow user to use tags to describe resource that are interested for them, helping to organization and share these resource with other user in the network. The famous widely known example of social bookmarking system is Delicious [1].

Social media systems are characterized by an ever-growing amount of content and users, that leads to two problems that arise in cascade [1].

Social Interaction Overload Problem Social interaction overload is a problem related to the huge amount of users and items that each user can interact it. So user does not have to choose your appropriate item that might be interesting. To run-over social interaction overload problem, Social Recommender System have been design and develop in order refine the content and suggested users only interesting items.

Serendipity/Over-specialization problem Over Specialization problem is related to the recommendations which are too similar to the items that already consider by the users and never receive unexpected and novel item. This approach is able to produce accurate recommendation but it is not useful recommendation. It is known that the serendipity problem affects both the content-based recommender system and the collaborative recommender system.

Our Contributions In this paper we present a friend recommendation system that operates in the Social bookmarking domain, designed and expand to face the social interaction and serendipity problem.

When a user adds another user as a friend, she/he always receives updates anytime a new bookmark is shared by the friend. So it is important that these bookmarked are related to resources that target user has not already bookmarked. Accuracy of the friend recommendation is a main property.

Therefore, we marked at generating a system able to:
- Recommend friend with a high accuracy [1].
- Recommend friends whose bookmarked are novel and serendipitous [1].

The rest of the paper is organized as follows: Section 2 presents related work; Section 3 describes the friend recommendation system; Section 4 describes Novelty and Serendipity in a friend recommender system; Section 5 illustrated the conducted experiments; Section 6 contains conclusion and future work.

II. RELATED WORK

This section presents related work on user recommendation in the social bookmarking domain. These systems can be divided into three categories, based on the source of the date used to build the recommendation.

1. Systems based on the analysis of social graphs, which examine the set of people connected to the object user, in order to produce recommendation. These system suggested nearest users in the graph, like friend of friend and followers of follower (the "people you may know" feature offered by face book is the most known example of this approach) [4].
2. Systems that analyze the interaction of the users with the content of the system. In order to accomplishment the user interest, these systems usually build a user profile by giving a structured form to content [4].
3. Hybrid system, which consider both the social graph and the interaction of the users with the content [4].

III. DESIGNING FRIEND RECOMMENDER SYSTEM

This section presents the friend recommender system developed in our proposal. We first current an analysis of the user behavior in a social bookmarking system (section 3.1), the algorithms (section 3.2) that compose our system are presented.

1. User behavior in a social bookmarking system

In this section analyzing user behavior in a social bookmarking system from a friend recommendation opinion. This analysis has been conducted on a Delicious dataset, Distributed for the HetRec 2011 workshop [1], which contains:

- 1867 users;
- 69226 URLs;
- 53388 tags;
- 7668 bi-directional user relations;
- 437593 tag assignments;
-104799 bookmarks

2. Algorithms

In order to build an architecture for a friend recommender system in the social bookmarking system. The system works in five steps:

1. **Tag- based user profile.** In this step create a user profile, based on the frequencies of the tags used by the user.

$$v_{uj} = \frac{\#A(u, t_j)}{\#A(u)}$$

This equation estimates the number of times the tag was used, divided by the number of tag assignment of u [3].

2. **Resource-based user profile.** In this step user have to create their resource base user profile based on the resource bookmarked by each user.

$$v_{uj} = \begin{cases} 1 & \text{if the resource } r_i \text{ was bookmarked by u} \\ 0 & otherwise \end{cases}$$

This equation calculates the interest of a user u in a resource . If the resource was bookmarked by user u it return 1 otherwise it return 0 [3].

3. **Tag-based similarity computation.** In this method calculate the similarity among two tag based user profile with the Pearson's correlation coefficient.

$$ts(u,m) = \frac{\sum_{iCT_{um}} (v_{ui} - \overline{v_u})(v_{mi} - \overline{v_m})}{\sqrt{\sum_{iCT_{um}} (v_{ui} - \overline{v_u})^2 \sum_{iCT_{um}} (v_{mi} - \overline{v_m})^2}}$$

Where Tum represent the set of tags used by the both the users u and m. If the similarity value range from 1.0 which indicate complete similarity and -1.0 which indicate complete dissimilarity [3].

4. **User interest computation.** The user interest of a user u in a user m can be calculated as:

$$ui(u,m) = \frac{\#D(u,m)}{\#R(u)}$$

This equation calculates number of resource bookmarked by both users, divided by the number of resources bookmarked by user u [3]..

5. **Recommendation selection.** The use of keyword and resource in a social bookmarking domain is associated to two different type of behavior.

a) Tag based similarity
b) User interest computation.

Tag based similarity and user interest computation into an individual score would darken the information on how related two users are for each type of behavior. This could lead to potentially incorrect friend recommendation, like two users that use same keyword to describe the different and completely irrelevant type of the resources. So these algorithms filter the user by considering both the type of behavior. This system determines a set of users to recommend to the target user [1].

– Tag based similarity higher than a threshold value α
– User interest computation higher than a threshold value β.

IV. NOVELTY AND SERENDIPITY IN A FRIEND RECOMMENDER SYSTEM

Novelty and Serendipity are two metrics used to check a recommender system. Novelty calculate how many recommendation include item that user did not know. While Serendipity calculate how surprising the accurate recommendation are. A serendipitous recommendation is by definition also novel but novel recommendation might not be serendipitous. Two basic needs for a better serendipitous recommendation (1). Suggestion should be unexpected and surprising to the user (2). It must be useful.

When a user recommended as a friend, we can calculate if a resource bookmarked is serendipitous for candidate user. So, the distance between a recommended resource and resource already bookmarked by the candidate user is based on the tags used to categories the resource.

The similarity $sim(r_i, r_j)$ between two resources r_i and r_j can we define by measuring the Jaccard index [1].

$$sim(r_i, r_j) = \frac{\#(T(r_i) \cap T(r_j))}{\#((T(r_i) \cup T(r_j))}$$

A resource r_i € R can be consider serendipitous resource for a user u if $sim(r_i, r_j) < 0.5$.

Related to the target user there are five similar users. In experiments of the target user and similar five users with help of Jaccard index method many serendipitous resources found. So available resources are not accurate to recommend the target user [1].

To overcome from this problem Inno Serendipitous algorithm is used. Inno is serendipitous recommendation algorithms which generate recommendation from innovators. Recommendation generated from innovators must be unexpected and useful for the user, which contributes to the serendipity of the recommendation.

IV. PROPOSE SYSTEM DIAGRAM AND STEP

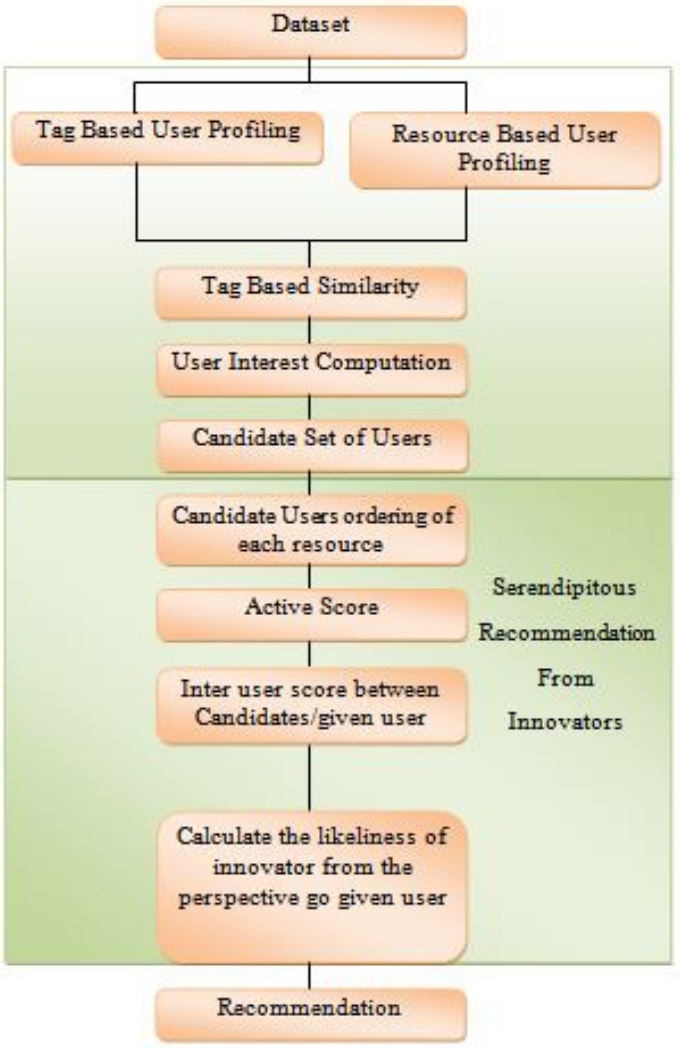

Figure 1.

Step 1: This analysis has been conducted on delicious dataset, Distributed for Het Rec 2011 workshop.

Step 2: Tag Based User Profiling: In this step user has to create your tag base user profile.

$$v_{uj} = \frac{|A(u, t_j)|}{|A(u)|}$$

Step 3: This step build another user profile, based on the resources bookmarked by each user.

$$p_{uj} = \begin{cases} 1 \text{ if } \exists t \in T \mid (u, r_j, t) \in A(u) \\ 0 \text{ otherwise} \end{cases}$$

Step 4: Tag Based Similarity.

$$ts(u, m) = \frac{\sum_{j \subset T_{um}} (v_{uj} - \overline{v}_u)(v_{mj} - \overline{v}_m)}{\sqrt{\sum_{j \subset T_{um}} (v_{uj} - \overline{v}_u)^2} \sqrt{\sum_{j \subset T_{um}} (v_{mj} - \overline{v}_m)^2}}$$

Step 5: User Interest Computation.

$$ui(u,m) = \frac{|D(u,m)|}{|R(u)|}$$

Step 6: the candidate set of users can be defined as:

$$S(u_t) = \{u_i \in U \,|ts(u_t,u_i) > \alpha \,\&\&\, (ui(u_t,u_i) > \beta) \,\|\, (ui(u_i,u_t) > \beta)\}$$

Step 7: the candidate users ordering of each resource.

Step 8: The more frequently the user selected the item soon after it released, the higher the active score s/he achieves.

$$Active(u_i) = avg(1 - \frac{O_{\Pi_j,u_i}}{|O_{\Pi_j}|}), j \in S(u_i)$$

Step 9: Inter User Score between Candidate user and given user.

$$InterUser(u_i,u) = \sum_{j \in S(u_i,u)} \delta_j(O_{\Pi_j,u_i}, O_{\Pi_j,u}) \frac{[(O_{\Pi_j,u} - O_{\Pi_j,u_i})]}{|O_{\Pi_j}|}$$

$$\delta_j(O_{\Pi_j,u_i}, O_{\Pi_j,u}) = \begin{cases} 1, & if O_{(k,u_i)} < O_{(k,u)} \\ 0, & otherwise \end{cases}$$

Step 10: Calculate the Likeliness of u_i is an innovator from the perspective of user u.

$$P(u_i,u) = Active(u_i) \cdot InterUser(u_i,u)$$

Step 11: Rating of innovator for item i.

$$\hat{r}(u,i) = \frac{\sum_{u_k \in Innovators(u,K)} P(u_k,u) \cdot r(u_k,i)}{\sum_{u_k \in Innovators(u,K)} P(u_k,u)}$$

Step 12: Recommendation priority.

$$RP(u,i) = \sum_{u_k \in Innovators(u,K)} P(u_k,u) \cdot r(u_k,i)$$

V. CONCLUSION AND FUTURE WORK

In this paper Behavioral data mining is produce accurate friend recommendation which allow the user to know the bookmarks resources that are Novel and Serendipitous. The Experiment of target user and similar candidate users are producing more serendipitous resources. So available resources are not accurate to recommend the target user. To overcome this problem Inno Serendipitous Recommendation algorithm are used.

REFERENCES

[1] Manca, Matteo, Ludovico Boratto, and Salvatore Carta. "Behavioral data mining to Produce novel and serendipitous friend recommendations in a social bookmarking System." Information Systems Frontiers (2015): 1-15.

[2] Manca, Matteo, Ludovico Boratto, and Salvatore Carta. "Friend recommendation in a Social bookmarking system: Design and architecture guidelines." Intelligent Systems in Science and Information 2014. Springer International Publishing, 2015. 227-242.

[3] Manca, Matteo, Ludovico Boratto, and Salvatore Carta. "Mining User Behavior in a Social Bookmarking System-A Delicious Friend Recommender System." DATA. 2014.

[4] Manca, Matteo, Ludovico Boratto, and Salvatore Carta. "Design and architecture of a Friend recommender system in the social bookmarking domain." Science and Information Conference (SAI), 2014. IEEE, 2014.

[5] Manca, Matteo, Ludovico Boratto and Salvatore Carta. "Producing friend recommendation in a social bookmarking system by mining users content." International Conference on Advances in Information Mining and Management (IMMM'13).2013