2012 International Conference on Management of e-Commerce and e-Government

Building a Tag Map for Recommendations in Microblogging

Yinghua Xiao
Southwestern University of Finance & Economics
Chengdu, China
mofei1227@gmail.com

Ting Du
Southwestern University of Finance & Economics
Chengdu, China
duttings@sina.com

Wen Zhu
Southwestern University of Finance & Economics
Chengdu, China
vanechu.cs@gmail.com

Qing Li
Southwestern University of Finance & Economics
Chengdu, China
kooliqing@gmail.com

***Abstract*—In its brief history of about two decades, the Web has evolved from a technical framework for information dissemination to more of an enabler of social interactions among its users. This weaves a huge virtual social network for users. Finding friends and targeting useful information are great challenges in such a complicated social network. Different with traditional content-based and collaborative information filtering techniques without considering social connections among users, we propose a naive recommendation approach utilizing user relationships to find friends and hot topics in social media. This is achieved by weaving a tag map from user preferences. Our experiments based on Microblogging data from Sina.com show a promising result in social media recommendation.**

Keywords-tag map; recommender; social media; microblogging

I. INTRODUCTION

Since its inception, the Web has evolved from a technical framework for information dissemination to more of an enabler of social interactions among its users. The Web circa 1990 consisted primarily of static text content expressed in HTML. Nowadays, users of the Web have a variety of advanced techniques including XML and AJAX to access richer and dynamic contents. Such technological advancement has fertilized vibrant creation, sharing, and collaboration among the users [1]. These social interactions make the Web to be the most important vehicle for "social media". In Wikipedia, social media is defined as media for social interaction, using web-based technologies to allow the creation and exchange of user-generated content. The most popular representative of social media is Microblogging. Twitter, Facebook, Plurk, Google+, and Sina Weibo are good examples of Microblogging websites.

Essentially, Microblogging is a broadcast medium in the form of blogging. It allows users to exchange small elements of content such as short sentences, individual images, or video links. Meanwhile, users with similar interests are gradually gathering together and form a huge social network. For instance, Sina Weibo, one of the most popular Microblogging websites in China, reached about 30 million users on May 2012. It becomes the largest virtual social community in China.

With a large scale of users, there are millions of messages disseminating in these social networks. It is a critical problem to find friends and suggest useful information to a target user in these social networks. Traditional information filtering techniques including content-based and collaborative recommenders merely focus on the information to be recommended. Apparently, utilizing the relationship of users in social network will assist a target user to find his or her friends and useful blogging messages. In this work, we build a tag map via user preferences to help recommending friends and messages in Microblogging websites.

The rest of this work is organized as follows. In Section 2, we first provide some background information of recommendations in social media. The design details for tag map based recommendations in Microblogging are described in Section 3. We then conducted experiments to evaluate the performance of such a social network based recommender. This paper is concluded with speculation on how the current prototype can be further improved in Section 4.

II. RELATED WORK

As the beginning of the Web, content-based filtering techniques are the most popular way to provide recommendation services. The common strategies of content-based filtering are to build a topic profile deduced by the targeted textual information to retrieve relevant information or automatically categorize documents by their themes [9]. For instance, Chiang and Chen [2]

978-0-7695-4853-1/12 $26.00 © 2012 IEEE
DOI 10.1109/ICMeCG.2012.29

study a few classifiers for agent-based news recommendations. Yang et al. [12] propose learning algorithms to categorize similar news articles based on the semantic content and temporal aspects of news events. In these studies, the selection of relevant information is determined by the textual similarity between the incoming articles and the targeted article.

Some work on news recommendation takes one step further by incorporating additional metadata including user behaviors. Claypool et al. [3] filter out online newspaper in terms of user ratings on news. Wei et al. [11] combine user's visiting history for collaborative news recommendation.

With the popularity of social media, connections of users are utilized for information recommendations. Esmaili et al. [4] applies the linking structure among blogs to recommend blogging messages. Hayes, Avesani, and Bojars [6] explore measures based on blog authorship and reader tagging to improve recommendation. Li and Chen further integrate trust, social relation and semantic analysis [7] for suggestion messages. These works directly apply the social connections of users for recommendation. Due to the dynamic of social connections, Liu and Maes [10] utilize ontology techniques to convert such relationships into an interestmap for robust recommendation. There are more than 21,000 interest descriptors and 1,000 identity descriptors in their ontology for mapping terms with similar concepts.

Follow the track of Liu and Maes [10], we convert user connections into a tag map deducted from user tag preference. To improve the utility of tag map based recommendation, we adopt our previous work on concept unification [8] to map similar tag terms into one concept. Our experiments based on Microblogging data from Sina.com show a promising result of tag map based recommendation.

III. OUR APPROACH

To clearly present our approach for tag map based recommendations, we first introduce the background information of a Microblogging websites and then describe our approach in details.

A. Background of Microblogging

Figure 1. An example of Microblogging from SINA

Typically, a microblogger posts a short message in a microblogging website and his/her followers comment on this message. A microblogger also maintains a personal profile describing his or her interests and friends. Interests of a microblogger are usually presented in tags and friends who prefer to read messages of the microblogger are listed as fans or followers. As shown in Figure 1, the user has 5 tags, i.e., TAMU, Capricorn, SWUFE, Latin dance and movie sicker, which are freely written by the owner. There are 126 friends reading the messages of this user.

With these two typical features of microblogging, i.e., tag and friend, we construct a tag map describing the connections of tag terms. As shown in Figure 2, two tag terms are connected if they are tagged together by a microblogger or tagged separately by two microbloggers who are friends. The connection weight is measured by the count of being connected with the above two conditions. With this tag map, we can easily recommend friends or messages to users.

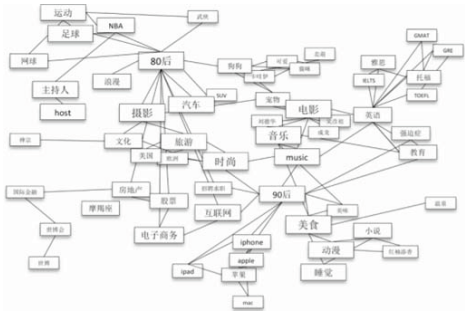

Figure 2. A Snapshot of a tag map for SINA Microblogging website

B. Construction of Tag Map

The construction of a tag map consists of two steps: (a) extract tag terms and unify similar

terms into one concept; (b) calculate the connection weight of two terms.

1) Tag term

Since each microblogger maintains a tag maker describing his or her passions, interests and hobbies etc, we can easily extract tag terms from these tag markers. The challenge work comes from the free setting of tag terms. Each microblogger can select any terms as his or her tag terms. In this case, two microbloggers can express the same interest with the different words. For instance, a microblogger from Southwestern University of Finance & Economics can describe his affiliation in the tag as SWUFE, others can describe it in full name "Southwestern University of Finance & Economics". Some can go even further by express it in Chinese as "西南财经大学" or "西财". Apparently, we should unify these terms into one concept and treat them as one tag term. In this work, we apply our previous study result on concept unification of different words [8] to solve this problem.

To determine whether we should combine tag term T1 and T2 into one concept, we apply three models, i.e., statistical model, semantic model and phonetic model to measure their similarity. Statistical model measures the co-occurrence of these two words in the Web. Semantic model measures the semantic similarity of these two words. For example, For example, there are two tag terms "电影迷" and "电影控" ("movie fan" and "movie sicker"). Since both words share the same root "movie", it tends to be a high semantic similarity. Phonetic model measures the similarity of two phrases in terms of phoneme. Based on the similarity of these three models, we can decide the probability to unify the terms. More details can be found in our previous work [8].

2) Connection weight

After mapping similar terms into one concept, we can calculate the connection weight of these tag terms. Here, we adopt Pointwise Mutual Information (PMI) [5] to measure the connection weight. The PMI of a pair of terms A and B quantifies the discrepancy between the probability of their coincidence in case of given their joint distribution and the probability of their coincidence of given their individual distributions, assuming independence. The formula to calculate PMI,

$$PMI(a.b) = \log\frac{p(a,b)}{p(a)p(b)} = \log\frac{p(a\mid b)}{p(a)} = \log\frac{p(b\mid a)}{p(b)}$$

As aforementioned, connection weight of terms is calculated based on two cases: (1) two tag terms appear in the same microblogger tag maker. In this case, we can figure out the value of PMI1 easily using the definition of Pointwise Mutual Information. (2) Tag terms appear in a microblogger and his/her friend tag maker respectively. As for this part, we set up two subsets of the target user's tags and his/her followings' tags, where PMI2 needs more complicated calculation based on both subsets.

What is more, we found it possible to gain the reasonable weights for PMI1 and PMI2 utilizing the number of followings and followers, which can reflect the level of user's curiosity and influence in social network.

Finally, in order to compare the final values of PMI, we normalized PMI1 and PMI2 obtained in the two cases. The connection weight of tag maps is defined as,

$$PMI_{AB} = W_1 PMI_1 + W_2 PMI_2$$

where PMI1 denotes the PMI value of A and B owned by one user. PMI1 is the PMI value in case of A and B belongs to a user and one of her/his followings. The overall PMIAB is the final value to express the relations between A and B. W1 and W2 are the weight defined as,

$$W_1 = \frac{m_1}{M_1}, W_2 = \frac{m_2}{M_2}$$

where m1 denotes the average number of users' followers who own a certain pair of tags, and M1 denotes the mean level of number of followers of all the samples; m2 denotes the average number of followings in each pair of words, and M2 reveal the average number of followings calculated from the profile table.

3) Tag map based recommendation

With a tag map, we can easily recommend friends or messages to microbloggers. In particular, once a new microblogger registers in the website, a tag map based recommender can first find top-n tag terms strongly connecting with his/her tag preferences. Users containing these top-n tag terms are recommended to this new microblogger. Messages with a strong textual similarity to the top-n tag terms can be recommended to the user. In fact, there are lots of other applications of this tag map including as commodities recommendation and hot issue tracking.

IV. EXPERIMENT

A. Data Description

In this study, we construct a tag map from the information of SINA VIP microbloggers. We collected 325,392 personal profiles of VIP users from Sina microblog website in February 2012

by our focused crawler. After preprocessing, we extracted 356,614 tag terms from these profiles to construct our tag map.

B. Tag Map Eficiency

We have designed a questionnaire to evaluate the efficiency of our constructed tag map. First, we extract tag terms from our volunteer profiles in the SINA Weibo. Then, we find relevant tags and present them to volunteers. Volunteers check these feedbacks to judge the relevance of these suggested tag terms - "√" for accept, "×" for reject and no marks for can be accepted. In our experiment, there are 200 participants. 86% of the volunteers satisfied (or ticked) with at least one tag term in his/her recommended tag group, whilst 78.4% of the suggested tags have been accepted. Only 5.7% recommended tag group have been definitely rejected. Table 1 is an example of this questionnaire.

TABLE I. A PART OF THE INVESTIGATION RESULTS.

user name	tags	recommendations
Queen_兰儿	旅游 80后 电影 音乐	☐ 大学生 ☐ 双鱼座 ☐ 安静
孙溥123	自由 学生 旅游 旅行 80后	☐ 阅读 ☐ 足球 ☐ 大学生
Ricky_Yao	西南财大 模范丈夫 大学 财经 90后 篮球爱好者	☐ 运动鞋 ☐ 大学生 ☐ 前卫
LWY鲁文宇	减肥 运动 跳舞 90后 听歌 电影	☒ 大学生 ☐ 减肥达人 ☐ 微博
段晓_Irma	自由 依存症 射手座 90后 旅行 吃货	☒ 足球 ☐ 明星 ☐ 定向越野 ☐ 宅女

C. Tag Map based Recommendation

We also suggested 5 potential friends based on top-10 relevant tag terms of the target participants. 85% of these 200 participants add at least one of the recommended friends and 76% recommendations are accepted.

V. CONCLUSION

Nowadays, Web is one of the most important vehicles for "social media". Essentially, the Web has become a platform for information exchange and user interaction besides information dissemination. Recommendation in social media is a good extension of this fashion. In this work, we present a tag map based approach for recommendations in Microblogging websites.

There are a few interesting ways to extended our study, for example, mining the users' preference not only from tags but also from posted messages. In addition, our techniques can be extended to detect and track hot topics in blogosphere or microblogging. Indeed, its power is yet to be further improved and investigated.

ACKNOWLEDGMENT

This work has been supported by Program for New Century Excellent Talents in University in China (NC1T-11-0694), the National Natural Science Foundation of China (NSFC) (Grant No. 60803106, No. 61170133), Fok Ying-Tong Education Foundation for Young Teachers in the Higher Education Institutions of China (121068) and One Hundred Talents Program in Sichuan Province of China.

REFERENCES

[1] T. Berners-Lee, W. Hal, J. Hendler, N. Shadbolt, and D. Weitzner, D. Creating a Science of the Web. Science, 313(5788), pages 769-771, 2006.

[2] J.-H. Chiang and Y.-C. Chen. An intelligent news recommender agent for filtering and categorizing large volumes of text corpus. International Journal of Intelligent Systems, 19(3):201-216, 2004.

[3] M. Claypool, A. Gokhale, T. Miranda, P. Murnikov, D. Netes, and M. Sartin. Combining content-based and collaborative filters in an online newspaper. In Proceedings of the ACM SIGIR Workshop on Recommender Systems, 1999.

[4] K. S. Esmaili, M. Neshati, M. Jamali, H. Abolhassani and J. Habibi. Comparing performance of recommendation techniques in the blogsphere. In Proceedings of ECAI 2006 Workshop on Recommender Systems, pages 40, 2006.

[5] R. M. Fan. Transmission of Information: A Statistical Theory of Communications, MIT Press, Cambridge, MA (Chapter 2), 1961.

[6] C. Hayes, P. Avesani and U. Bojars. An analysis of bloggers, topics and tags for a blog recommender system. In Proceedings of Workshop on Web Mining (WebMine), pages 1-20, 2007.

[7] Y. M. Li and C. W. Chen. A synthetical approach for blog recommendation: combining trust, social relation, and semantic analysis. Expert Systems with Applications, 36(3):6536 - 6547, 2009.

[8] Q. Li, Y.P. Chen, S.H. Myaeng, Y. Jin, B.Y. Kang Concept unification of terms in different languages via Web mining for Information Retrieval, Information Processing and Management, 45(2):246-262, 2009.

[9] X. Liu, C. Chen, and Y. Liu. Algorithm for ranking news. In Proceedings of the 3rd International Conference on Semantics, Knowledge and Grid, pages 314-317, 2007.

[10] H. Liu, P. Maes. InterestMap: Harvesting Social Network Profiles for Recommendations, Beyond Personalization-IUI, 2005.

[11] C. Wei, Z. Li-jun, C. Chun, and B.Jia-jun Chen. A hybrid phonic Web news recommender system for pervasive access. In Proceedings of International Conference on Communications and Mobile Computing, pages 122 --126, 2009.

[12] Y. Yang, J. G. Carbonell, R. D. Brown, T. Pierce, B. T. Archibald, and X. Liu. Learning approaches for detecting and tracking news events. IEEE Intelligent Systems, 14(4):32-43, 1999.