International Journal of Database Theory and Application
Vol.9, No.7 (2016), pp.1-8
http://dx.doi.org/10.14257/ijdta.2016.9.7.01

Research on the Method of Personalized Information Recommendation Based on Social Tagging

Chaoyang Ji

Xuchang University
Xuchang, China
jcy@xcu.edu.cn

Abstract

In the social tagging systems, tags are results of use annotations. Combining social tagging with personalized information recommendation, this paper designs a personalized information recommendation method based on social tagging on the basis of the existing method of personalized information recommendation. This method clusters user tags and resources tags to form tagging clouds, and then calculate the similarity between the user community tagging cloud and resources community tagging cloud to make personalized recommendation. The experiment results show that the method has achieved a good expected effect and can raise the accuracy and efficiency of personalized information recommendation.

Keywords*: social tagging system; personalized information recommendation; tagging cloud*

1. Introduction

Personalized information recommendation is mining the users' interests and potential information demand through analysis, to recommend information resource satisfied the users' preference and the latest information needs in a doable way to the relevant users, its essence is to recommend the information that users mostly want timely, to meet the information demand of them. With the explosive growth of network information, the information overload becomes a problem for the current network user, the existing information processing technology and method is difficult to find the ideal data and information for the majority of users, and no one can fully grasp the entire Internet information. This leads to a serious dilemma of rich information and poor knowledge. To deal with the increasing massive information in network space and personalized information needs for the majority of users, the research of personalized recommendation technology and methods becomes research hotspot in information retrieval and information search field, and it is accepted as the effective way to solve the information overload.

The key of personalized information recommendation method is how to accurately obtain the users' individual preference information, and the current research mainly focuses on using the invisible knowledge of users, the potential relevance of users, and users' preference information for personalized information recommendation. Such as personalized information recommendation method based on analysis of social network proposed by Xu Baoxiang is using social network analysis tool to mine common preference of network community and personalized preferences of individual, and then combine with the wisdom of crowds and individual preference information to recommend collaboratively; multidimensional information recommendation method based on scene similarity proposed by Yang Jun is blending situational factors in the mining of user behavior, to optimize content and manner of information recommendation through catching the dynamic changes of user preferences in different scenarios, greatly

ISSN: 2005-4270 IJDTA

enhancing the personalization and effects of information recommendation; personalized recommendation system model based on trusting social network proposed by Walter and blog recommendation method proposed by Li combined trust and social relations and semantic analysis are build user trust relationship network through social network to obtain users' personalized information, to realize personalized recommendation by the trust relationship Network.

Accurately obtaining users personalized information, building user interest model based on the personalized information, and on this basis to realize personalized information recommendation based on collaborative or personalized information recommendation based on the content is the focus of the research. In this paper, on the basis of the above research, combining the open architecture concept of take the customer as the center and the user participation in the Web 3.0, integrating social tagging into personalized information recommendation, has researched and designed personalized information recommendation method based on social tagging. The method is to form communities through cluster analysis of users and resources of social tagging system, to obtain target user tag clouds and resource tag clouds representing each community, so as to the gained tag clouds can not only reflect the preferences of the user, but also reflect the characteristics of the resource content, in the process of resource recommendation to the user, only need to compare the similarity between the user community tag clouds and resources community tag clouds, can greatly reduce time complexity of the recommendation method in the similarity calculation. The experimental results show that, this method can not only better mining user interest preference and resource content characteristics, but also can effectively improve the accuracy and efficiency of the personalized information recommendation.

2. The Recommendation Method Based on Social Tagging

Social tagging system is the typical application that users participate and create Internet information content in Web 3.0 times, it not only allows users to freely label the concerned information resources, but provides an open social environment for the user sharing resources or emotional tendencies. Since Joshua Schachter created the first website del.icio.us to offer social tagging service, the social tagging sites have been rapidly developed and widely used, have appeared a lot of new application and experience such as Twitter, Youtobe, and Flickr. In these social tagging systems, personalized tags added to relevant information resources can not only reflect users' description and understanding of the resources, but also can reflect the preference of the user's.

Through the analysis of personalized tags labeled by the user, to gain potential interest and preference of the user, provides the possibility to realize personalized information recommendation, such as through social tagging system to mine emotional tendency or attitude of users, or using social tagging system to establish the relational network between users and resources, to realize social recommendation based on clustering, matrix and semantic; or combining social tagging system with traditional recommendation methods and using social tagging system to provide user preference information, to solve the problem of user demand preference information singularity in traditional recommendation methods. Research shows that, to mine user interest preference information and resource preference information through social tagging system, to carry out personalized recommendation by using the potential preference information, is the key to improve the user information demand satisfaction and recommendation quality. Therefore, blending social tagging system into the personalized information recommendation, this paper has researched and designed a personalized information recommendation method based on social tagging. The method includes three phases:

(1) The acquisition of user tag cloud based on the user tag similarity

The premise of the acquisition of user tag cloud based on the user tag similarity is similar users tending to have similar preferences. Let T denote the reference tag set, the tag vector for user u_j:

$$u_j = (wt_{1j}, wt_{2j}, ..., wt_{ij}, ..., wt_{nj})$$

Among them, wt_{ij} is the weight of tag t_i for user u_j, and it is calculated by using tf-idf formula:

$$wt_{ij} = tf_{ij} * idf_i = \frac{freq_{ij}}{\sum_{i=1}^{n} freq_{ij}} * \log\frac{m}{n_i}$$

Wherein, $freq_{ij}$ represents the total number of resources labeled by tag t_i used by the users u_j, $\sum_{i=1}^{n} freq_{ij}$ represents the total frequency labeled by reference tag set T used by user u_j, m represents the total number of users, n_i represents the total number of users who collect and use tag t_i. In the social tagging system, although user interests preference and its changes in different period can be reflected by the tag used by the user, the higher the frequency of a tag, the stronger the preferences to the resource, the resources marked the tag should also have a higher weight; but because the social tagging system is a dynamic system, user tag will change with time, therefore, in the process of user clustering, need to consider the influence of temporal information on user interests preference. In general, the resources concerned by the user recently should be more valuable than before, also can manifest the user current interest preference easier; in addition, the longer the user concerned a tag, the higher the interest preference for the tag. In view of the above situation, this paper blend time factors into the user tag vector, the calculation formula is as follows:

$$pwt_{ij} = \frac{r_{ij} + f_{ij} + d_{ij}}{\chi}$$

Among them, r_{ij} represents the difference value between the last time user u_j using t_i tag to label resources and the current time, the less the different value, the more the tag t_i can be used to represent the current user's preferences; f_{ij} represents the number of times a user u_j using the tag t_i to label resources in a certain time range, the more the times, the higher the users' interest of tag t_i; D_{ij} represents duration times user u_j using the tag t_i to label resources, the longer, the easier can reflect the current user preferences, the calculation method is:

$$d_{ij} = \frac{date_{last} - date_{begin}}{date_{now} - date_{begin}}$$

Among them, $date_{last}$ represents the last time of tagging the tag, $date_{begin}$ represents the first time of tagging the tag, $data_{now}$ represents the current time. The result of tag vector of user u_j corrected by using the above formula is:

$$u_j* = (wt_{1j} \times pwt_{1j}, wt_{2j} \times pwt_{2j}, ..., wt_{ij} \times pwt_{ij}, ..., wt_{nj} \times pwt_{nj})$$

After all the user tag represented using vector, adopt cosine similarity calculation method to calculate the similarity between user u_i and user u_j:

$$Sim_{ij}(u_i, u_j) = \frac{\overrightarrow{u_i^*} \times \overrightarrow{u_j^*}}{\left\|\overrightarrow{u_i^*}\right\| * \left\|\overrightarrow{u_j^*}\right\|}$$

Set the default threshold value is e, if $Sim_{ij}(u_i, u_j) > e$, u_i and u_j will be divided into the same community, the user tag cloud of this community is :

$$T_U = T_{u_i} \cup T_{u_2} \cup ... T_{u_p}$$

Among them, T_U says the tag cloud of user community, T_{u_p} says the tag set used by user u_p.

(2) Resource tag cloud acquisition based on resource tag similarity

In the social tagging system, although the user's annotation behavior is free, the same resources may be have different tags, but for the same resources, in the exclusion of malicious labeling conditions, these tags in the semantics should be close, therefore, the key of resource tag cloud acquisition based on resource tag similarity is to calculate the similarity of these tags. Suppose the tag vector of resource r_p to be:

$$r_p = (rtagw_{1p}, rtagw_{2p}, ..., rtagw_{ip}, ..., rtagw_{np})$$

Among them, $rtagw_{ip}$ represents the weight of tag tag_i in resources r_p, it is calculated as:

$$rtagw_{ip} = rtagtf_{ip} \times rtagidf_i = \frac{rtagfreq_{ip}}{\sum_{i=1}^{n} rtagfreq_{ip}} \times \log \frac{N_i}{n_i}$$

Among them, $rtagtf_{ip}$ represents the frequency of resources r_p labeled by tag_i, N_i represents the amount of resources, n_i represents the appear frequency of tag tag_i. Then the similarity between resource r_p and r_q is:

$$Sim_{pq}(r_p, r_q) = \frac{\overrightarrow{r_p^*} \times \overrightarrow{r_q^*}}{\left\|\overrightarrow{r_p^*}\right\| * \left\|\overrightarrow{r_q^*}\right\|}$$

Set the default threshold value is e, if $Sim_{pq}(r_p, r_q) > e$, then the r_p and r_q tags are close in semantics, resource r_p and r_q can be considered the same resources, the tag cloud of the resources is:

$$T_R = T_{r_i} \cup T_{ur} \cup ... T_{r_x}$$

Among them, T_R represents the tag clouds of the resources community, T_{r_x} represents the tag set used by the resource r_x.

(3) Personalized recommendation based on the tag cloud

The tag cloud is composed of multiple tags, in the process of personalized recommendation, the tag cloud can be regarded as the document, tags which contains as characteristic concept of the document. Therefore, recommendation for the current users can use the recommendation algorithm based on content analysis:

$$Sim_{U^*,R}(T_{U^*}, T_R) = \frac{\overrightarrow{T_{U^*}} \times \overrightarrow{T_R}}{\left\|\overrightarrow{T_{U^*}}\right\| * \left\|\overrightarrow{T_R}\right\|}$$

3. Experiments and Results Analysis

This paper adopts the comparative experimental analysis to verify the accuracy and efficiency of this method.

(1) Experimental data

The experimental data comes from the typical application del.icio.us of the social tagging system. Selecting recently the most active 150 Internet users by using webpage analytical tools and the website API, and taking these users as the target user to collect related information, including basic information, collected resources and collection time, the species and times of tagging *etc.* After pretreatment to the acquired information, remove the user who use less than 15 species, can get 124 valid users, 37254 resource projects as the experimental data.

(2) The experimental evaluation indexes

The experimental evaluation indexes select recommendation hit-rate and recommendation hit-rank as the basis evaluating various recommendation approach. Hit-

rate index is mainly used to evaluate the number of times that the some recommended resources hit higher score resources in the test data set, hit-rank index is mainly used to evaluate the position of recommended resources in the recommendation ordering set. The formulas for these two kinds of indexes are:

$$hit\text{-}rate = \frac{hit}{N}, \quad hit\text{-}rank = \frac{\sum_{i=1}^{hit} \frac{1}{pos_i}}{N}$$

Among them, *hit* represents the amount of resources meeting user need in the recommendation resources set, *N* represents the amount of resources, pos_i represents the position of recommended resources which are in line with the needs of the users in the recommendation resources set. Through the above formula, the bigger the hit-rate value, the higher the accuracy rate of recommendation method, the better the effect of recommendation; the bigger the hit-rank value, the more forward position of recommended resources which are in line with the needs of the users in the recommendation set, the larger probability used by users, the higher recommendation quality.

(3) The experimental process and results

This paper conducts the comparison experiment in the same experimental data and the evaluation index, the selected comparative method are collaborative filtering recommendation based on users and association recommendation based on resources content. All users constitutes the user set in the experiment process, select 10 users randomly as the target user, then sort the resources collected by target users according to the collection time, and take 80% of the total resources as the training set, 20% as the test set; in order to improve the execution efficiency, set 10 as similar neighbors of target user, the top 40 of the largest resource similarity as the recommendation results. The experimental results as shown in Figure 1 and Figure 2.

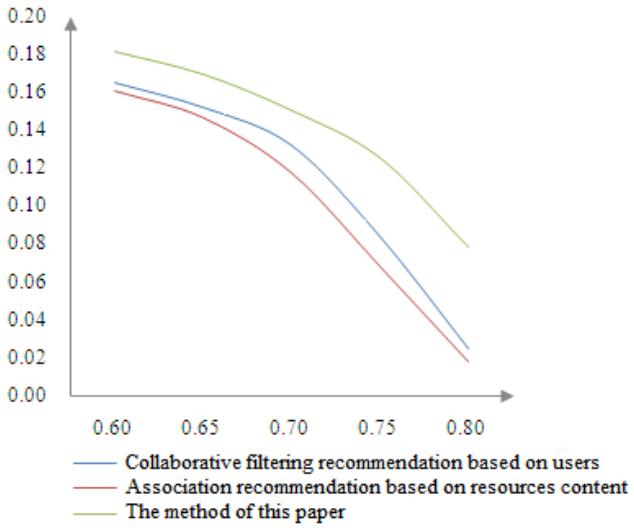

Figure 1. The Hit-Rate Value of Three Kinds of Different Recommendation Method

(4) The analysis of experimental results

From Figure 1 and Figure 2 we can see, these three recommendation methods are affected by the similarity threshold, for different similarity threshold e, hit-rate curve and hit-rank curve have the same variation trend, *i.e.* decreases with the increase of similarity threshold e. Compared to traditional collaborative filtering recommendation method based on the user and association recommendation method based on the resource content,

personalized information recommendation method based on social tagging presented in this paper , on the same similarity threshold, hit-rate value and hit-rank value is the maximum value, and with the increase of similarity threshold e, the variation trend of hit-rate and hit-rank curves is relatively flat, the result shows that the proposed method is better in terms of recommendation quality and effect than collaborative filtering recommendation method based on the user and association recommendation method based on the resource content.

The mainly reason is that the method presented in this paper gives consideration to two aspects of both users and resources, can be more comprehensive access to the user's interest and preference, can be more comprehensive to proceed the user and resource clustering by acquiring the user tag cloud and resource tag cloud, can predict the current user's interest and preference accurately, so as to give a more appropriate recommendation, while collaborative filtering recommendation method based on users only proceed similarity calculation and user interest modeling according to the user's browsing habits, then proceed information recommendation according to the similarity, because of the change of the user's browsing habits, lead to the migration of the users' interest and the lag of the recommendation result. The association recommendation method based on the resource content aggregates resources only according to the content characteristics of the resources, and then recommends based on the use of related resources. Although this method can effectively avoid the recommendation results lag, can recommend related resources according to the user's browsing habits, cannot guarantee the effect and satisfaction of recommendation owing to the inability to accurately predict the user's interest.

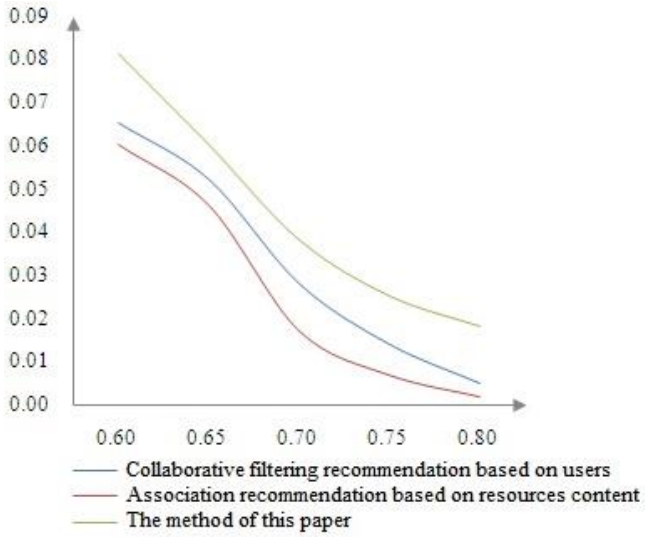

Figure 2. The Hit-Rank Value of Three Kinds of Different Recommendation Method

By using the bridge role of tag in social tagging system, to proceed user similarity clustering based on user tag, can effectively mine preferences and changes of this kind of user, to proceed resource similarity clustering based on resources tag, can effectively achieve the effective integration and clustering of similar resources. To calculate the similarity between the user community tag clouds and resources community tag cloud based on these process, and sort according to the similarity calculation results. The recommendation results according to sorting results will be better than the recommendation results only in the view of user or resources. The experimental results in this paper also illustrate the issue, which proved that personalized recommendation

method based on the social tagging presented in this paper can obtain the ideal application effect.

4. Conclusion

Compared to conventional personalized information recommendation system, social tagging system includes more comprehensive and richer user information. It can be used to mine user's interest and preferences, and provide data for personalized recommendation system. This article combined social tagging system and personalized information recommendation method, and proposed the novel method of personalized information recommendation based on social tagging. The method involves getting each cluster's tag cloud based on the cluster analysis of user tag and resource tag according to the bridging character of tag in social tagging system. Afterwards, the tag cloud is used to fully describe the essential characteristics of users and resources, and then a similarity computation between the user community tag clouds and resources community tag clouds is conducted, and finally the information resources will be ranked and recommended according to the similarity calculation results. The experimental results show that the method has achieved expected effect, and can obtain the user's interest preference and resource content characteristics through effectively mining. The accuracy and efficiency of the personalized information recommendation were also effectively improved.

Acknowledgements

Supported by Science and Technology Research Project of Education Department of Henan Province, Project Number: 13A520746.

References

[1] J. Vig, M. Soukup, S. Sen and J. Riedl, "Tag expression: tagging with feeling", Proceedings of the 23nd annual ACM symposium on User interface software and technology. New York, NY: ACM, **(2010)**, pp. 323-332.
[2] J. Furner, "On Recommending", Journal of the American Society for Information Science and Technology, vol. 53, no. 9, **(2002)**, pp. 747-763.
[3] X. Baoxiang and W. Huijuan, "Research of Personalized Information Recommendation Model Based on Web 3.0", Journal of Information Resources Management, vol. 2, no. 1, **(2012)**, pp. 59-63.
[4] Y. Jun, W. Juhua and A. Danxiang, "A New Method for Multi-dimensional Information Recommendation Based on the Similarity of Context", Journal of the China Society for Scientific and Technical Information, vol. 32, no. 3, **(2013)**, pp. 262-269.
[5] F. E. Walter, S. Battiston and F. Schweitzer, "A Model of a Trust-based Recommending System on a Social Network", Autonomous Agents and Multi-Agent Systems, vol. 16, no. 1, **(2008)**, pp. 57-74.
[6] L. Yungming and C. Chingwen, "A Synthetical Approach for Blog Recommendation: Combing Trust, Social Relation, and Semantic Analysis", Expert Systems with Applications, vol. 36, no. 3, **(2009)**, pp. 6536-6547.
[7] X. Ningxia, S. Yidan, Q. Hua and Z. Min, "Method for personalized user profiling in social tagging systems", Journal of Computer Applications, vol. 31, no. 6, **(2011)**, pp. 1667-1670.
[8] C. Wartena, R. Brussee and M. Wibbels, "Using Tag Co-occurrence for Recommendation", 2009 Ninth International Conference on Intelligent System Design and Application, IEEE, **(2009)**, pp. 273-278.
[9] W. Weiping and W. Jinhui, "Hybrid recommendation method based on tag and collaborative filtering", Computer Engineering, vol. 37, no. 14, **(2011)**, pp. 34-36.
[10] S. Rendle and L. S. Thieme, "Pairwise interaction tensor factorization for personalized tag recommendation", Proceedings of the 3rd ACM International Conference on Web Search and Data Mining, New York, USA, **(2010)**.
[11] G. Weiguang, L. Daofang and Z. Lei, "Personalized resource recommendation method in social tagging system", Computer Engineering and Applications, vol. 47, no. 10, **(2011)**, pp. 240-243.
[12] Y. Tan, C. Yidong and J. Yuehui, "BPR-UserRec: a personalized user recommendation method in social tagging systems", The Journal of China Universities of Posts and Telecommunications, vol. 20, no. 1, **(2013)**, pp. 122-128.

Author

Chaoyang Ji, He received his Master of Engineering in Computer Science from University. Now he is associate professor of computer science at Information Engineer Department, University. Since 2011 he is Member of IEEE. His current research interests include different aspects of Artificial Intelligence and Information Retrieval.