IJCSS International Journal of Complex Systems in Science

Int. J. Complex Systems in Science
vol. 1 (2011), pp. 10–15

Identifying the Role of Social Tags and its Application in Recommender Systems

Zi-Ke Zhang[1,†] and Chuang Liu[1,2,3]

[1] *Department of Physics, University of Fribourg, Chemin du Muse, Fribourg 1700, Switzerland*

[2] *School of Business, East China University of Science and Technology, Shanghai 200237, P. R. China*

[3] *Engineering Research Center of Process Systems Engineering (Ministry of Education), East China University of Science and Technology, Shanghai 200237, P. R. China*

Abstract. The past few years have witnessed the great success of a new family of paradigms, social tagging networks, which allows users to freely associate social tags to items and efficiently manage them. Thus it provides us a promising way to effectively find useful and interesting information. In this paper, we consider two typical roles of social tags: (i) an accessorial tool helping users organize items; (ii) a bridge that connects users and items. We then propose a hybrid algorithm to integrate the two different roles, expecting to obtain better recommendation performance.

Keywords: Personalized recommendation, Social tagging networks, Hybrid
MSC 2000: 97P70, 91D30, 05C85

1. Introduction

The past few years have witnessed an explosion era of information: we face too much data to be able to filter out irrelevant information [1]. Recently, *Personalized Recommendation*, which provides a promising way to solve this dilemma, has attracted increasing attention [2, 3, 4, 5, 6, 7, 8], among which the most successful one is *Collaborative Filtering* (CF) [9, 10]. CF tends to produce recommendations by evaluating most similar users or items to the target user based on his/her historical activities. Despite its great success, CF has also some drawbacks: (i) a huge number of items that is far beyond users' ability to measure even a small fraction of them; (ii) many users are not willing to give explicit rate to items. Therefore, a variety of researches have been devoted to use accessorial information to obtain better performance of recommendation under the framework of CF. Examples include user profiles [11], item attributes [12] and content descriptions [13]. However, profile-based methods are restricted by user privacy rules, attribute-aware algorithms are limited by vocabulary which are defined by domain experts, and various applications(e.g. videos, images) are lack of contents.

Recently, the advent of Web2.0 and its affiliated applications bring a new form of paradigms, *Social Tagging Systems*, introducing a novel platform for users' participation. A social tagging system allows users to freely assign tags to annotate their collections into baskets, requires no specific skills for users' participation, broadens the semantic relations among users and items, and thus has attracted much attention from the scientific community. A considerable number of researches have been done to study its usage patterns [14], structure [15], evolution and dynamics [16, 17]. Besides, social tagging systems have also been found wide applications in recommender systems. Szomszor *et al.* [18] considered the tag frequency as weight in movie recommender systems, and the result showed that with the consideration of weights the recommendation accuracy can be enhanced. Schenkel *et al.* [19] proposed an incremental threshold algorithm taking into account both the social ties among users and semantic relations of different tags, which performed remarkably better than the algorithm without tag expansion. In addition, Zhang *et al.* [20] and Shang *et al.* [21] proposed an item-based and user-based hybrid tag algorithm, respectively, harnessing diffusion-based method [6] to obtain better recommendations. Furthermore, Shang *et al.* [22] considered the tag usage frequency as edge weight in bipartite networks and improved the accuracy of recommendation.

In this paper, we firstly identify the two roles of social tags, and then we propose a hybrid recommendation algorithm which considers these two basic roles of social tags: (i) an accessorial tool helping users organize items; (ii) a bridge that connects users and items.

2. Hybrid Algorithm and Performance Metrics

A social tagging network considered in this paper consists of three sets, respectively of users $U = \{U_1, U_2, \cdots, U_n\}$, items $I = \{I_1, I_2, \cdots, I_m\}$, and tags $T = \{T_1, T_2, \cdots, T_r\}$. A tripartite graph representation can be described by three adjacent matrices: A, A' and A'' for user-item, item-tag and user-tag relations. If user U_i has collected item I_j, we set $a_{ij} = 1$, otherwise $a_{ij} = 0$. Analogously, we set $a'_{jk} = 1$ if I_j has been assigned by the tag T_k, and $a'_{jk} = 0$ otherwise. Furthermore, we consider that users have personal preferences on tags, that is, a weighted adjacent matrix A'' is constructed, of which each element, a''_{ik}, is denoted as how many times that U_i has used tag T_k.

Subsequently, we introduce the diffusion-based hybrid algorithm. It can be easily understood that, in a real social tagging system, a user might have two different tagging behaviors [17]: (i) one might be aware of an item via random web surfing and save it as his/her favorites by adding some related words (tags) to describe them; (ii) s/he might firstly input one (or several compound) tag(s), then pick up some items s/he is interested in among all the possible retrieval results related to the required tag(s). The former one indicates social tags play as an organization tool for users, and the latter suggests tags as an information retrieval bridge. Given a target user U_i, the above three algorithms will generate final score of each item, f_j, that are pushed into recommendation list for him/her, are described as following:

(Algorithm I): Supposing that a kind of resource, a_{is} for user U_i and item I_s, is initially located on items, each item will averagely distribute its resource to all neighboring tags, and then each tag will redistribute the received resource to all the items associated with it. The final resource vector, $\vec{f}$, after the two-step diffusion is:

$$f_j = \sum_{l=1}^{r} \frac{a'_{jl}}{k(T_l)} \sum_{s=1}^{m} \frac{a_{is}}{k'(I_s)}, \ j = 1, 2, \cdots, m, \tag{1}$$

where $k(T_l) = \sum_{j=1}^{m} a'_{jl}$ is the number of neighboring items for tag T_l, $k'(I_s) = \sum_{l=1}^{r} a'_{sl}$ is the number of neighboring tags for item I_s.

(Algorithm II): Different from algorithm I, the initial resource, a''_{il} for user U_i and tag I_l, are located on tags according to their frequencies used by the given user U_i. Then each tag will distribute the initial resource directly to all its neighboring items. Thus, the final resource vector, $\vec{f'}$, reads:

$$f'_j = \sum_{l=1}^{r} \frac{a'_{jl} a''_{il}}{k(T_l)}, \ j = 1, 2, \cdots, m. \tag{2}$$

We then adopt a linear superposition to combine the above two algorithm, the final resource can be written as:

$$\vec{f^*} = (1 - \lambda)\vec{f} + \lambda\vec{f'}, \tag{3}$$

where $\lambda \in [0, 1]$ is a freely tunable parameter, $\vec{f}$ is the vector obtained from Eq. 1, $\vec{f'}$ is the vector derived from Eq. 2, and both the vectors are attained for the same target user. In the extremal cases $\lambda = 0$ and $\lambda = 1$, the hybrid algorithm degenerates to algorithm I and algorithm II, respectively. After generating the final resource vector by Eq. 3, all the items that s/he has not collected are ranked in a descending order, and the top L items will be recommended to the target user.

To measure the performance of the proposed hybrid algorithm, we design to employ two basic metrics, including accuracy and diversity:

1. *Ranking Score (RS)* [6, 8].— In the present case, for a particular user, RS is calculated as the rank of the deleted item in testing set divided by the number of all uncollected items for this user. Apparently, the lesser RS is obtained, the higher accuracy the algorithm is. Then an overall $\langle RS \rangle$ is generated by averaging over all the pairs in the testing set.

2. *Diversity (D)* [7, 23, 8].— D measures the differences of different users' recommendation lists, thus can be understood as the inter-user diversity. Denote I_R^i the set of recommended items for user U_i, then

$$D = \frac{2}{n(n-1)} \sum_{i \neq j} \left(1 - \frac{|I_R^i \cap I_R^j|}{L} \right), \tag{4}$$

where $L = |I_R^i|$ for any i is the length of recommendation list. Greater or lesser values of D mean respectively greater or lesser personalization of users' recommendation lists.

3. Conclusions and Discussions

In this paper, we have proposed a hybrid recommendation algorithm based on two roles of social tags. (i) an accessorial tool helping users organize items; (ii) a bridge that connects users and items. Apparently, we only provide a simple start point in understanding the roles of social tags, as well as making use of them to design recommendation algorithms. A couple of open issues remain for future study. First, we lack quantitative understanding of the structure and evolution of social tagging networks to well understand the observed roles. In addition, the two pure algorithms might reduce some information of social tagging networks. The hypergrpah theory [24, 17] is expected to harness the whole network structure without losing any information and thus provide a promising way to obtain better recommendation performance.

Acknowledgements

This work is partially supported by the Swiss National Science Foundation (Project 200020-121848). ZKZ acknowledges the National Natural Science Foundation of China under the grant no. 60973069. CL and ZKZ acknowledge the Scholarship Program supported by China Scholarship Council (CSC Program).

References

[1] G.-Q. Zhang, G.-Q. Zhang, Q.-F. Yang, S.-Q. Cheng, and T. Zhou *New J. Phys.* **10**, 167 (2007).

[2] Z. Huang, H. Chen and D. Zeng *ACM Trans. Inf. Syst.* **22**, 116 (2004).

[3] J. L. Herlocker, J. A. Konstan, L. G. Terveen and J. T. Riedl *ACM Trans. Inf. Syst.* **22**, 5 (2004).

[4] Y.-C. Zhang, M. Blattner and Y.-K. Yu *Phys. Rev. Lett.* **99**, 154301 (2007).

[5] Y.-C. Zhang, M. Medo, J. Ren, T. Zhou, T. Li and F. Yang *EPL* **80**, 68003 (2007)

[6] T. Zhou, J. Ren, M. Medo and Y.-C. Zhang *Phys. Rev. E* **76**, 046115 (2007).

[7] T. Zhou, L.-L. Jiang, R.-Q. Su and Y.-C. Zhang *EPL* **81**, 58004 (2007).

[8] T. Zhou, Z. Kuscsik, J.-G. Liu, M. Medo, J. R. Wakeling and Y.-C. Zhang *Proc. Natl. Acad. Sci. USA* **107**, 4511 (2010).

[9] M. Balabanović and Y. Shoham *Commun. ACM* **40**, 72 (1997).

[10] B. Sarwar, G. Karypis, J. Konstan and J. Riedl *Proc. the 10th Intl. Conf. WWW* (ACM Press, New York), pp. 295-305 (2001).

[11] P. Kazienko and M. Adamski *Info. Sci* **177**, 2269 (2007).

[12] K. Tso and L. Schmidt-Thieme *Proc. 29th Annual Conf. German Classification Society, Magdeburg, Germany* (2005).

[13] M. J. Pazzani and D. Billsus *Lect. Note. Comp. Sci.* **4321**, 325 (2007).

[14] S. A. Golder and B. A. Huberman *J. Info. Sci.* **32**, 198 (2006).

[15] Z.-K. Zhang, L. Lü, J.-G. Liu and T. Zhou *Eur. Phys. J. B* **66**, 557 (2008).

[16] C. Cattuto, V. Loreto and L. Pietronero *Proc. Natl. Acad. Sci. USA* **104**, 1461(2007).

[17] Z.-K. Zhang and C. Liu *arXiv: 1003.1931.*

[18] M. Szomszor, C. Cattuto, H. Alani, K. Ohara, A. Baldassarri, V. Loreto and V. D. P. Servedio *Proc. the 4th Euro. Semantic Web Conf.* (Innsbruck, Austria), pp. 71-84 (2007).

[19] R. Schenkel, T. Crecelius, M. Kacimi, S. Michel, T. Neumann, J. X. Parreira and G. Weikum *Proc. the 31st Annual Intl. ACM SIGIR Conf. Res. Dev. Info. Retr.* (ACM Press, New York), pp. 523-530 (2008).

[20] Z.-K. Zhang, T. Zhou, and Y.-C. Zhang *Physica A* **389**, 179 (2010).

[21] M.-S. Shang, Z.-K. Zhang, T. Zhou and Y.-C. Zhang *Physica A* **389**, 1259 (2010).

[22] M.-S. Shang and Z.-K. Zhang *Chin. Phys. Lett.* **26**, 118903 (2009).

[23] T. Zhou, R.-Q. Su, R.-R. Liu, L.-L. Jiang, B.-H. Wang and Y.-C. Zhang *New J. Phys.* **11**, 123008 (2009).

[24] G. Ghoshal, V. Zlatić, G. Caldarelli, and M. E. J. Newman *Phys. Rev. E* **79**, 066118 (2009).