Analysis of User Tags in Social Music Sites: Implications for Cultural Differences

Li Chen

Department of Computer Science

Hong Kong Baptist University

China

lichen@comp.hkbu.edu.hk

Ho Keung Tsoi

Department of Computer Science

Hong Kong Baptist University

China

hktsoi@comp.hkbu.edu.hk

Abstract

With the tremendous popularity of social media sites over Internet, many researchers have started to study users' behavior while the users contribute items to the sites. Among them, tagging behavior is one vital area as the results from studying it can be beneficial to existing collaborative computing applications (e.g., recommender systems). Various metrics have been proposed to evaluate tag patterns, but few studies have dealt with the important cultural dimension associated with the tagging manner. Therefore, in this paper, we have particularly investigated two social music sites, SongTaste and Last.fm, which are respectively popularly used by Chinese and European. Analysis of user tags from them indicates significant differences between the two typical cultural groups in respect of their tag classes, tag agreements, and tag usefulness.

Keywords: Social music sites, tagging behavior, cross-cultural analysis

ACM Classification Keywords: H.5.3. Group and organization interfaces: Collaborative computing.

General Terms: Measurement, human factors.

Introduction

In social networking sites, tags have formed folksonomy [13], a mechanism to encourage users to describe items with their own annotations. In recent years, they have been increasingly adopted into recommender systems in order to compensate the limitations of pure rating based collaborative filtering methods (i.e., rating sparsity problem) [14, 15]. Some researchers have also started to analyze users' tagging trends in order to identify their inherent behavior patterns [4,7,9,11].

Unfortunately, few researches have so far taken the "culture" as one potentially influential factor in users' tagging behavior and involved it into tag-based recommendation generation, though it has been recognized that users from different ethnic countries will use the Web for different purposes and will possess varied preferences on the interface objects [2, 6]. Only till recently, some investigators have started to examine the relationship between content of tags and participants' nationalities [3]. However, since the related study was conducted in a lab setting and specific to digital images, it is still not clear about how online users have actually behaved in the real social environment and whether their discovered cultural differences would be also valid in other product domains.

Thus, driven by existing limitations, we have performed a large-scale analysis of users' tagging records, as extracted from popular social music sites, with the aim to reveal whether the tags that users associated with songs exhibit different properties between different cultural groups of users. The reasons that we emphasized social music sites are not only because they are typical domains in which recommender systems have been often applied, but also their data resources (e.g.,

tags from Last.fm) have been frequently used by researchers to test their recommender algorithms. Therefore, we believe that, with better knowledge of users' tag characteristics in such sites, existing recommender algorithms can be potentially enhanced in terms of better exploiting the tag data.

In the following, we will first introduce the experiment method including datasets we used and evaluation metrics, and then present the results analysis. We will finally make the conclusion of our major findings and indicate their practical implications to the research area.

Experiment Setup

Data Sets

In this experiment, we mainly collected tag data from users of two cultures: China and Europe. Given that it is hard to find a site that is popular in both origins, we have crawled user data from two sites: SongTaste (www.songtaste.com) and Last.fm (www.last.fm), respectively. SongTaste is one of most popular music-sharing sites in China and Last.fm is prominently used by European users. Another reason of choosing the two sites is that they possess similar functional features: providing music rankings, and supporting users to listen to any songs and then leave tags/comments on them, so that the confounding effect of factors such as the design difference of two websites can be minimized.

For each site, we first identified a set of 200 most popular songs (according to the popularity ranking) based on which the group of users who have tagged at least one song in the set is determined. A pre-filtering process was then carried out to leave only users whose nationality is China (for SongTaste), and European

countries (for Last.fm)[1]. The morphological variations (e.g., "ballad" vs. "ballads", "UK" vs. "uk") were also solved based on [14]. Then, to ease the comparison between the two user groups, we sorted all users by their tagging frequencies and finally selected the top 6,500 most active users to be contained in each group, so that the two compared groups are with equal size. The averagely applied number of tags per user was found to be 10.3 in the Chinese group (henceforth CG), and 63.1 in the European group (henceforth EG), which first indicates that EG is more active in creating tags.

We also prepared another dataset that was used for the analysis of tag classes (see the definition of "tag classes" later). It consists of 20 songs that appear in both websites' most popular songs' lists and were associated with at least one user tag (e.g., the song "Let it be" was tagged in both sites). With this dataset, we were hence able to compare, for the same song, the types of tags that were respectively assigned by CG and EG.

Evaluation Metrics
Three metrics were mainly measured in our analysis: tag classes, tag agreements, and tag usefulness. Their definitions were mainly from existing literatures [4,7,9,11], where the metrics have been well theoretically established and successfully applied in the tagging analysis of CiteULike [4, 9], Delicious [7, 11], and Movielens [10].

Tag classes. Tags can be classified into different classes according to their inherent semantics and supported user tasks. Golder et al. [5] proposed seven tag classes (see

Table 1). For example, one class is "identifying what or who it is about", such as "hard rock", and another tag class can be "identifying qualities or characteristics" (e.g., "relaxing"). Based on the 7-class classification scheme, Sen et al. [10] proposed three general tag classes, which are more related to user tasks that tags could support: *factual tags* that help to describe the "facts" and find related items, *subjective tags* that help to express user opinions on an item, and *personal tags* that are often used to organize a user's items (e.g., "my favorites"). In our experiment, according to the two classification schema, we have concretely analyzed tags from CG and EG, with the purpose of knowing whether they would pay different focuses and use semantically different terms to the same item, and whether their tags were created to serve different tasks.

Tag agreements. Friendship and membership are two explicit social relationships that an active SNS user is often involved. For example, on Last.fm, the user can establish friendship with others by "finding people" and have membership by joining interest groups. In recent years, researchers have attempted to fuse such data sources into the recommender system in order to augment the user-user similarity matrix, under the assumption that friends and members can infer more similar interests between users [12]. Driven by this need, we have specifically measured "tag agreements" among friends and among members in CG and EG respectively, so as to identify any differences between them in this regard. The tag agreement is also called "interest sharing" in [9]. The function that they used to assess wide characteristics of interest sharing in social bookmarking sites (e.g., Delicious) is applied by us to identify the sharing pattern of tags in social music sites. Concretely, based on the asymmetric Jaccard coefficient

---

[1] The countries include UK, Germany, Netherlands, Spain and Italy, accessed through the site's provided API.

[9], the formula we used to calculate the tag agreement among friends is:

$$\frac{|T_{user} \bigcap T_{friend_i}|}{|T_{user}|} \quad (1)$$

where T_{user} and $T_{friendi}$ respectively refer to the user's own tag set (i.e., the set of distinct tags given by the user) and her/his i-th friend's tag set, and the average agreement will be then calculated among all of the user's friends. The tag agreement between the user and her/his members is defined in the similar way, for which $T_{friendi}$ is replaced by $T_{memberj}$.

Tag usefulness. The third metric in our analysis contains both tag non-obviousness and tag discrimination, which have been first proposed by [4] as standard metrics to evaluate user tags in CiteULike. *Tag non-obviousness* mainly assesses how often a tag does not occur in the text/content of item that it is associated with (see Formula (2)). The premise is that a non-obvious tag adds more intellectual value to the item than an obvious tag.

$$\frac{|T \notin content|}{|T \in song|} \times 100\% \quad (2)$$

Tag discrimination assesses the ability of individual tags in discriminating the resources that they are assigned to. A highly discriminatory tag should be able to distinguish more items in the collection, bringing more information gain for that tag [4]. Formally, the tag discrimination is calculated as the average number of distinct items that each tag is associated with:

$$\frac{\sum(\#\_of\_distinct\_songs\_for\_each\_tag)}{\#\_of\_tags} \quad (3)$$

Results

Tag Classes

According to [6], the Oriental culture, as mainly influenced by the ancient Chinese culture, is more likely to focus on holistic thought, continuity, and interrelationships of objects. On the contrary, the Western culture, as influenced by the ancient Greek culture, puts greater emphasis on analytical thought, detachment, and attributes of objects. We were thus interested in knowing whether the two cultural groups would also put different attentive focuses for tagging, in form of using different types of words. Based on the 7-classs and 3-class classification schemes respectively defined in [5] and [10], we manually coded and categorized tags from 20 songs that occur in both sites' top song lists. The aim was then to see for the same song, whether the tag classes' distributions by CG and EG would be different.

Concretely, for each song, we first classified its distinct tags into the seven classes [5]. The work was done by two coders. They discussed every tag and reached a consensus finally. Table 1 gives examples of tags in each class. Here, we use the relative percentage (i.e., the amount of tags in every class over all tags for that song) to determine the tag distribution and compare it between CG and EG. For each class, its overall distribution was obtained by averaging its percentages across all 20 songs.

7-class classification scheme		CG	EG	*p*
C1	Identifying what (or who) it is about, e.g. "male vocalist", "hard rock", "love"	74.3%	41.1%	*.000*
C2	Identifying what it is, e.g., "pop", "demo"	0.59%	1.10%	*.404*
C3	Identifying who owns it	0	0	
C4	Refining categories, e.g., "60s", "uk"	3.8%	10.4%	*.005*
C5	Identifying qualities or characteristics, e.g., "great", "sad", "relaxing"	8%	36.6%	*.000*
C6	Self reference, e.g., "my collection"	13.2%	11.4%	*.624*
C7	Task organizing, e.g., "to save", "to listen"	0.13%	1.78%	*.006*
3-class classification scheme		**CG**	**EG**	***p***
FT	Identifying "facts" about an item such as people, places, or concepts, e.g., "metal", "uk", "American", "Lady GaGa", "disco"	75.6%	48.4%	*.000*
ST	Expressing user opinions related to an item, e.g., "brilliant lyrics", "awesome", "funny"	5.34%	36.7%	*.000*
PT	Organizing a user's items, e.g., "favorites"	19%	17.2%	*.718*

Table 1. Distributions over tag classes between Chinese Group (CG) and European Group (EG) (*FT* stands for Factual Tags, *ST* for Subjective Tags, and *PT* for Personal Tags).

It can be seen from Table 1 that in EG, the two more popular classes are C1 ("identify what (or who) the item is about", 41.07%) and C5 ("identifying qualities or characteristics", 36.6%). The less popular ones are C6 ("self reference", 11.36%) and C4 ("refining categories", 10.42%). Relatively, fewer tags were classified into C7 ("task organizing", 1.78%), C2 ("identifying what it is", 1.10%), and C3 ("identifying who owns it", 0). In comparison, in CG, C1 achieves the dominant popularity (74.25%), that is significantly higher than the C1's relative occupancy in EG ($p < 0.01$, t = 7.2 by t-Test). On the other hand, the values of C5 & C4 (8% and 3.8%) in CG are both significantly lower than their corresponding percents in EG.

It thus indicates the obvious disagreement between CG and EG as for their focus distribution among classes C1, C4, C5 and C7, though they share similar tagging pattern in C2, C3 and C6 (i.e., the less usage of tags in these classes). Indeed, CG is more active in using nouns to identify what (or who) the item is about (e.g., " " (i.e., classics), " " (i.e., band)), whereas, besides giving tag for this purpose, EG also frequently applies subjective tags to identify the item's qualities or characteristics (e.g., "perfect", "amazing").

To further verify the difference from the aspect of tag-supported user tasks, we coded tags into the 3-class scheme [10]. Table 1 gives the examples in these classes and Figure 1 illustrates the comparison between CG and EG respecting their class distribution. It can be seen that the percentage of factual tags (FT), i.e., tags are used to "identify 'facts' about an item such as people, places, or concepts", is significantly higher in CG than in EG (75.7% against 48.4%, $p < 0.01$, t = 4.6). However, the percent of subjective tags (ST), which are used to "express user opinions related to an item", is significantly lower in CG (5.3% vs. 36.7% in EG, $p < 0.01$, t = -13.4). As for personal tags (PT) that are used to "organize a user's items", there is no significant difference between the two groups (19% in

CG vs. 17.2% in EG, $p = 0.72$, t = 0.37). The results from this 3-class classification hence show again that Chinese are more likely to apply tags to identify the facts, whereas Europeans also employ tags to express evaluative opinions. Such divergence can be probably due to the different emphases that they put on the same object, as inferred by [6], or the distinction in their usage and interpretation of "tags" while supporting certain tasks. In our future work, we will be engaged in revealing the exact reason through more focused studies.
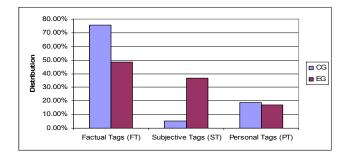

Figure 1. Comparison of tag classes according to t he 3-class classification scheme.

The above finding therefore implies that the tag sources as from different cultural groups could be used differently when being utilized to enhance the music recommenders. Tags from CG will be more suitable to improve similarity metric between items because they contain more facts for expanding an item's content, and tags from EG can be not only incorporated for this purpose, but also used to infer users' item preferences as the subjective tags can reflect users' opinions on the item. Moreover, given that users might have different aims of using the tags (e.g., Chinese more likely apply

tags to describe the facts), the process of recommending tags to a user [13] can be also different for CG and EG, taking into account of their 3-class and even more detailed 7-class tag distributions.

Tag Agreements among Friends & among Members
As defined before, the second assessment we did is mainly about the tag agreements among fiends and among members. To reach this objective, for every user, her/his friends (who appear in the user's friend list) and members (who appear in the same discussion group with the user) were first collected. The average numbers of friends and members turned out to be 4.5 and 502.09 respectively in CG, and 32.55 and 1742.86 in EG.

We then conducted the within-group analysis in order to first see within the same group, whether the two types of agreements would be different or not. That is, it can be analyzed whether a user's agreement value with her/his friends is significantly higher or lower than her/his agreement with members. In CG, it is found to be significantly higher among members (0.004 against 0.0017, $p < 0.01$, t = -3.32 by Paired-Sample t-Test; see Table 2). The result is similar in EG (0.064 with members that is significantly higher than 0.027 with friends, $p < 0.01$, t = -33.99). It thus infers that for both user groups, the membership can be more powerful than the friendship to determine the interest sharing of tags among users. The finding will be hence suggestive to the setup of relative weights on the two relationships when they are fused together into a social filtering recommender algorithm for optimizing the accuracy of user-user similarity measure [12].

Between-group analysis first showed that both types of agreements are higher in EG (see Table 2) than in CG.

In addition, the overlapping between an average user's friend list and her/his member list has been also found averagely higher in EG than in CG (7.93 vs. 1.96, $p < 0.01$, t = -38.97). The results infer that the degree of social affiliation can be stronger in EG's dataset, in terms of revealing more similar tag interests and social connections among users.

	Among friends	Among members	*Within-group*
CG	0.0017	0.0044	t = -3.32 ($p < 0. 01$)
EG	0.0272	0.0641	t = -33.99 ($p < 0. 01$)
Between groups	t = -33.77 (*p <* .01*)*	t = -57.4 (*p <* .01*)*	

Table 2. Tag agreements among friends and among members respectively from the two groups.

Tag Non-Obviousness and Discrimination
As defined before, tag non-obviousness is the ratio of tags that do not appear in the item's content to the total number of tags being associated with the item. For each site, 200 most popular songs' contents (including the song's title, singer(s), lyrics, etc.) were compared against their user tags. Resulting from Formula (2), for the average item, 93% and 96% of the tags are found to be non-obvious respectively in CG and EG ($p = 0.004$, t = -2.85). The results indicate that users in both sites are inclined to apply tags that do not appear in the content, which phenomenon is even more intense in EG.

We further calculated the average tag discrimination value with Formula (3). According to [4] and

information theory, the tag with the maximal information gain should be able to discriminate 50% of all items. In our case, the lower bound is 1.0 song/tag (when each tag is only associated with one song) and the upper bound is 200 songs/tag (when each tag is applied to every item in the set). The ideal tag discrimination value is then expected to lie between the two bounds and be close to the middle value, i.e., 100 songs/tag. The data analysis shows that the overall information gain that EG provides through tags is higher than that of CG (53.18/200 = 26.6% against 8.13/200 = 4.07%, $p < 0.01$, t = -27.6).

Since a tag's usefulness can be evaluated by combining its non-obviousness value and discrimination power [4], the least non-obvious tags (e.g., the tag "rock" that already occurs in the song's title) or the least discriminating tags (e.g., the tag "music" is applied to every item in the set) can be considered to be removed from the processing without any intellectual loss. The comparison between CG and EG from the two aspects indeed shows that the tag's quality is on average positively higher in EG, implying that it may demand less effort in the pre-cleaning before its tag dataset can be effectively used (such as to generate recommendations).

Discussion and Conclusion
Thus, through the comparison of tag datasets that we crawled from SongTaste and Last.fm, we discovered the significant differences between their users' music tagging behavior. Given that the two sites' functional features are essentially similar, except that they are respectively popularly used in China and Europe, we expect that the results can be in some sense indicating of the two typical cultural groups' tagging habits, and

hence be suggestive to optimally utilize tag data for developing more effective recommender systems that can tailor to the users' inherent properties.

Specifically, we analyzed user tags from the sites regarding three aspects: tag classes, tag agreements among friends and among members, and tag usefulness (including tag non-obviousness and discrimination). The main findings can be summarized as follows:

1) Chinese tend to apply factual tags to identify what (or who) a song is about, while Europeans also assign subjective tags to express their opinions on the song's qualities or characteristics.

2) Relative to the friendship, the membership is more powerful for revealing tags' interest sharing, within both cultural groups. Moreover, the overall social affiliation seems stronger among European users.

3) The tags assigned by both groups are with high level of non-obviousness and discrimination values. The European tags are even shown more useful in terms of the two aspects, inferring that they are capable of exhibiting more information gain.

The reported results can hence not only compensate the limitation of related works that lacked the tagging analysis study on social music sites, but also be constructive to works on tag-based recommender systems. For instance, the variation related to tag classes between Chinese group (CG) and European group (EG) implies that their tag data should be utilized differently. For CG, the majority factual tags can be adopted to expand items' content and be contributive to augmenting the similarity measure between items. For EG, users' preferences on items can be inferred from their actively applied subjective tags, so as to enhance the user-user similarity accuracy. Item-based or user-based collaborative filtering (CF) [14, 15] can be then accordingly employed to suit the specialty and usage of the respective groups' tag data.

To push our work a step forward, in the future, we will be engaged in analyzing more aspects of tagging patterns from the perspective of cross-cultural analysis. In fact, due to the limited access to the tagging timestamp on the two sites, we did not address the tag reuse and tag growth over time [4]. We will attempt to overcome this by looking into other possible sources. Moreover, motivated by the tagging differences discovered between CG and EG, we are planning to conduct more controlled experiments (e.g., qualitative interviews) in order to in-depth explore the causal reasons. One possible direction can be based on the claim that western countries generally have individualism and a low context culture, whereas eastern countries generally have collectivism and a high context culture [1]. On the other hand, the self-assertiveness has been often correlated to a person's willingness to express their own opinions in interactions [8]. Based on these theories, we will be interested in investigating what could be the leading, influential factors in our scenarios.

References
[1]   Chau, P. Y. K., Cole, M., Massey, A. P., Montoya-Weiss, M. and O'Keefe, R. M. Cultural differences in the online behavior of consumers. *Communications of the ACM 45*, 10 (2002), 138-143.

[2] Chen, L. and Pu, P. A cross-cultural user evaluation of product recommender interfaces. In *Proc. RecSys 2008*, ACM Press (2008), 75–82.

[3] Dong, W. and Fu, W. Toward a cultural-sensitive image tagging interface. In *Proc. IUI 2010*, ACM Press (2010), 313-316.

[4] Farooq, U., Kannampallil, T. G., Song, Y., Ganoe, C. H., Carroll, J. M. and Giles, L. Evaluating tagging behavior in social bookmarking systems: metrics and design heuristics. In *Proc. GROUP 2007*, ACM Press (2007), 351–360.

[5] Golder, S. A. and Huberman, B. A. Usage patterns of collaborative tagging systems. *Journal of Information Science 32*, 2 (2006), 198–08.

[6] Lee, K., Joshi, K. and McIvor, R. Understanding multicultural differences in online satisfaction. In *Proc. ACM SIGMIS-CPR 2007*, ACM Press (2007), 209-212.

[7] Rader, E. and Wash, R. Influences on tag choices in del.icio.us. In *Proc. CSCW 2008*, ACM Press (2008), 239–248.

[8] Richmond, V. P. and McCroskey, J. C. *Communication: Apprehension, Avoidance, and Effectiveness* (4th ed.). Scottsdale, AZ: Gorsuch Scarisbrick, 1995.

[9] Santos-Neto, E., Condon, D., Andrade, N., Iamnitchi, A. and Ripeanu, M. Individual and social behavior in tagging systems. In *Proc. HT 2009*, ACM Press (2009), 183–192.

[10] Sen, S., Lam, S. K., Rashid, A. M., Cosley, D., Frankowski, D., Osterhouse, J., Harper, F. M., and Riedl, J. Tagging, communities, vocabulary, evolution. In *Proc. CSCW 2006*, ACM Press (2006), 181–190.

[11] Wu, C. and Zhou, B. Analysis of tag within online social networks. In *Proc. GROUP 2009*, ACM Press (2009), 21-30.

[12] Yuan, Q., Zhao, S., Chen, L., Ding, S., Zhang, X. and Zheng, W. Augmenting collaborative recommender by fusing explicit social relationships. In *Proc. ACM RecSys Workshop on Recommender Systems and the Social Web*, (2009), 49-56.

[13] Zhang, N., Zhang, Y., and Tang, J. A tag recommendation system for folksonomy. In *Proc. ACM Workshop on Social Web Search and Mining (SWSM '09)*, 9-16.

[14] Zhao, S., Du, N., Nauerz, A., Zhang, X., Yuan, Q., and Fu, R. Improved recommendation based on collaborative tagging behaviors. In *Proc. IUI 2008*, ACM Press (2008), 413-416.

[15] Zhen, Y., Li, W., and Yeung, D. TagiCoFi: tag informed collaborative filtering. In *Proc. RecSys 2009*, ACM Press (2009), 69-76.