Exploring the Influence of Tagging Motivation on Tagging Behavior

Roman Kern[1], Christian Körner[2], and Markus Strohmaier[1,2]

[1] Know-Center, Graz
[2] Graz University of Technology
rkern@know-center.at,{christian.koerner,markus.strohmaier}@tugraz.at

Abstract. The reasons why users tag have remained mostly elusive to quantitative investigations. In this paper, we distinguish between two types of motivation for tagging: While *categorizers* use tags mainly for categorizing resources for later browsing, *describers* use tags mainly for describing resources for later retrieval. To characterize users with regard to these different motivations, we introduce statistical measures and apply them to 7 different real-world tagging datasets. We show that while most taggers use tags for both categorizing and describing resources, different tagging systems lend themselves to different motivations for tagging. Additionally we show that the distinction between describers and categorizers can improve the performance of tag recommendation.

1 Introduction

Tags in social tagging systems are used for a variety of purposes [1]. In this paper, we study the distinction between two different tagging behaviors. The first type of tagging is similar to assign resources to a predefined classification scheme. Users motivated by this behavior use tags out of a controlled and closed vocabulary. These users, named *categorizers*, tag because they want to construct and maintain a navigational aid to resources for later browsing. On the other hand, users who are motivated by description view tagging as a means to accurately and precisely describe resources. Tags produced by this user group resemble keywords that are useful for later searching [2]. This distinction can be exploited for example to improve the performance of tag recommender systems and information retrieval applications. Figure 1 contrasts a tag cloud of a typical categorizer with a tag cloud of a typical describer.

2 Development of Measures

To characterize the extent to which users categorize or describe resources, we present statistical and information-theoretic measures that are independent of the meaning of tags, the language of tags, or the resources being tagged.

Characterizing Categorizers: The activity of tagging can also be viewed as an encoding process, where tags encode information about resources. If this

Fig. 1. Examples of tag clouds of a typical *categorizer* (left) and a typical *describer* (right)

would be the case, users motivated by categorization could be characterized by their encoding quality, where categorizers would aim to maintain high information value in their tag vectors. This intuition can be captured with the conditional entropy of $H(R|T)$, which will be low if the tag distribution efficiently encodes the resources, with R being the set of resources and T the set of tags. For normalization purposes, we relate the conditional entropy to an optimal encoding strategy given by the number of tags, resources and average number of tags per resource: $M_{cat} = \frac{H(R|T) - H_{opt}(R|T)}{H_{opt}(R|T)}$.

Characterizing Describers: Users who are primarily motivated by description would generate tags that closely resemble the content of the resources. As the tagging vocabulary of describers is not bounded by taxonomic constraints, one would expect describers to produce a high number of unique or very rare tags in relation to the number of resources. One way to formalize this intuition is the orphan ratio, a measure capturing the extent to which a user exhibits description behavior: $M_{desc} = \frac{|\{t : |R(t)| \leq n\}|}{|T|}, n = \lceil \frac{|R(t_{max})|}{100} \rceil$

A combination of these measures - $M_{combined}$ - can be defined as the arithmetic mean of M_{desc} and M_{cat}. A detailed description of the measures and a comparison with other measures can be found in [3].

3 Application of Measures

Synthetic Datasets: As a first check of they usefulness of the measures we first applied them on two synthetic datasets, which are designed to resemble extreme categorizing and describing behavior. The synthetic dataset for describer behavior is based on the ESP game dataset, where users describe images (290 users and 29,834 tags). For the extreme categorizers we used the photoset feature from Flickr, where users sort their pictures into albums, just like users would organize their pictures in folders on their hard drives (1,419 users with 39,298 tags). The accuracy with which a measure can identify ESP game data as describers, and Flickr Sets data as categorizers can act as an approximation of its validity. In this simplified setting, $M_{combined}$, M_{desc} and M_{cat} achieve high accuracy values of 99.94%, 99.82% and 99.94% respectively.

Real-World Datasets: We have gathered 7 real-world datasets from different social tagging systems. For each tagging system, we acquired data from users with a minimum number of resources $|R_u|_{min}$. See table 1 for an overview of the size of the datasets. We applied the $M_{combined}$ measure on all datasets to study whether the various tagging systems differ in regard to the two user types. Figure 2 demonstrates that the distribution of describers and categorizers vary

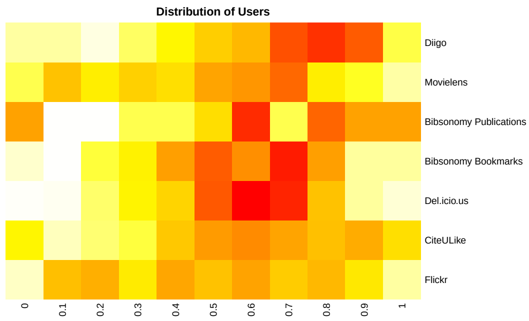

Fig. 2. Distribution of users of the real-world datasets according to the $M_{combined}$ measure. The intensity of the color encodes the relative number of users within a bin. The bins on the left side represent categorizer, while the rightmost bins represent users that display a behavior typical for describers. For example the `Flickr` dataset contains users evenly distributed between the two extremes, whereas the majority of users in the `Diigo` dataset are identified as describers.

between the individual social tagging systems. For example the `Diigo` dataset contains many users that are identified as describers. One possible reason for this might be the fact, that the `Diigo` platform not only offers the possibility to tag resources, but also to create so called bookmark lists, which is better suited to categorize resources.

Tag Recommender Finally we implemented two simple tag recommender systems to test whether the distinction between the two users groups could improve the performance of tag recommendation. The first recommender draws tags from the personal tagging history of a user and is labeled as *personomy-based recommender*. The *folksonomy-based recommender* suggests the most frequent tags as used by other describer users. Users were split into a describer and categorizer group according to the $M_{combined}$ measure. The baseline was produced by randomly assigning users to one of the two groups. For the evaluatcion we used the `Del.icio.us` dataset, as the folksonomy-based recommender requires resources tagged by multiple users. Figure 3 depicts the performance of the tag recommenders for different splits of the userbase (from 10% categorizers and 90% describers up to a 90%:10% split). One can see that using the personal tagging history is helpful for categorizers, while describers appear to tags similar to other users (describers) in the folksonomy. Especially of interest is the point where the relative improvement of the two recommenders intersect each other (right chart in figure 3). When developing a production tag recommender, this would be the point to switch from personomy-based tag recommendation for categorizers to a folksonomy-based recommender for describers.

Dataset	$\|U\|$	$\|T\|$	$\|R\|$	$\|R_u\|_{min}$	$\|T\|/\|R\|$
Delicious	896	184,746	1,089,653	1,000	0.1695
Flickr Tags	456	216,936	965,419	1,000	0.2247
Bibsonomy Bookmarks	84	29,176	93,309	500	0.3127
Bibsonomy Publications	26	11006	23696	500	0.4645
CiteULike	581	148,396	545,535	500	0.2720
Diigo Tags	135	68,428	161,475	500	0.4238
Movielens	99	9,983	7,078	500	1.4104

Table 1. Overview of the size and characteristics of the crawled real-world datasets.

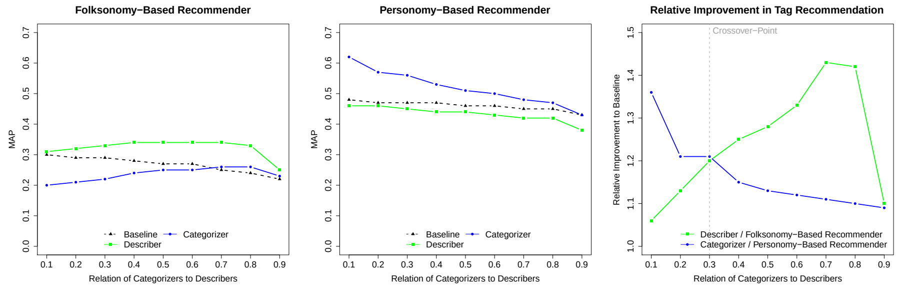

Fig. 3. Suggesting tags also used by other users appears to be a good strategy for describers (left). Categorizer prefer to reuse tags from their personal tagging history (middle). The relative improvements indicates that for about 30% of all users in our `Del.icio.us` dataset the personomy-based recommender is the better choice (right).

4 Conclusion

We showed that different tagging systems lend themselves to different motivations for tagging. Our results reveal that even within tagging systems, tags are adopted in different ways. One of the major implications of our work is that tagging motivation exhibits significant variety, which could play an important part in a range of problems including tag recommendation and information retrieval. In previous work [4], we have demonstrated that the motivation behind tagging influences the performance of semantic acquisition algorithms in folksonomies. Improving existing state-of-the-art tag recommenders by incorporating the tagging motivation is one of the main goals of our future work.

Acknowledgements: The research presented in this work is in part funded by the Know-Center and the FWF Austrian Science Fund Grant P20269. The Know-Center is funded within the Austrian COMET Program under the auspices of the Austrian Ministry of VIT, Austrian Ministry of WA and by the State of Styria.

References

1. Heckner, M., Heilemann, M., Wolff, C.: Personal information management vs. resource sharing: Towards a model of information behaviour in social tagging systems. In: Int'l AAAI Conference on Weblogs and Social Media (ICWSM), San Jose, CA, USA (2009)
2. Strohmaier, M., Körner, C., Kern, R.: Why do users tag? detecting users' motivation for tagging in social tagging systems. In: International AAAI Conference on Weblogs and Social Media (ICWSM2010), Washington, DC, USA, May 23-26. (2010)
3. Körner, C., Kern, R., Grahsl, H.P., Strohmaier, M.: Of categorizers and describers: An evaluation of quantitative measures for tagging motivation. In: 21st ACM SIGWEB Conference on Hypertext and Hypermedia (HT 2010), Toronto, Canada, ACM (June 2010)
4. Körner, C., Benz, D., Hotho, A., Strohmaier, M., Stumme, G.: Stop thinking, start tagging: Tag semantics arise from collaborative verbosity. Proceedings of the 19th Int'l Conf. on World Wide Web - WWW '10 (2010)