

The Efficacy of Tags in Social Tagging Systems

Khasfariyati Razikin, Dion Hoe-Lian Goh, Elizabeth Kian Cheow Cheong,
and Yi Foong Ow

Wee Kim Wee School of Communication and Information,
Nanyang Technological University
{khasfariyati, ashlgoh, w060022, w060050}@ntu.edu.sg

Abstract. Social tagging systems are a popular means for sharing resources. However, social tagging depends on individual knowledge. We evaluate the effectiveness of tags in describing the resources using support vector machines via classification. We achieved precision and recall at 90.22% and 99.27% respectively, with an average accuracy of 89.84%. Our results show that tags may help users' group resources into broad categories.

Keywords: Social Tagging, Support Vector Machines, Machine Learning.

1 The Study

Social tagging is a process of annotating a resource with keywords by anyone without being limited to a set of vocabulary [1]. Users are able to share their tags and their associated resources with others [2]. The tagging process depends largely on an individual's knowledge and might not follow a standard taxonomy. The effectiveness of tags as document descriptors is thus very much dependent on the individual. Additionally, a social tagging system does not have the same characteristics found in a taxonomic classification system. As such, an evaluation of their efficacy in a social tagging system is worth exploring.

Our goal is to determine the ability of the tags to describe their associated documents in a social tagging system. A total of 20 tags and 1385 English-language web pages associated with each tag were randomly downloaded from del.icio.us, a popular social tagging site. Pre-processing involved stop word removal and stemming. TFIDF was used to weight the remaining terms. Two-thirds of the data were used for training a SVM classifier¹, and the remainder were for testing. Table 1 shows the results from the classifier. Precision rates of more than 80% were obtained indicating that the documents were very relevant to their associated tags. In terms of recall, all tags provided highly pertinent pages. On average, we obtained 90.22% for precision, 99.27% for recall and 89.84% for accuracy.

¹ <http://svmlight.joachims.org>

Table 1. Results from experiment

Term	Accuracy (%)	Precision (%)	Recall (%)
Mobile Learning	87.50	87.50	100.00
Text classification	82.09	81.54	100.00
Disneyland	100.00	100.00	100.00
Tennis player	93.22	96.43	96.43
Decision making	96.97	96.97	100.00
Tourism in Egypt	100.00	100.00	100.00
LCD TV	100.00	100.00	100.00
Apple iPod	100.00	100.00	100.00
Dementia	81.82	81.82	100.00
Interferometry	100.00	100.00	100.00
Coffee production	72.73	80.00	88.89
Tin Toy	75.00	72.73	100.00
Car market	80.00	80.00	100.00
Nutritional science	90.00	90.00	100.00
Children abuse	90.00	90.00	100.00
Fashion paris	90.00	90.00	100.00
Journalism	87.50	87.50	100.00
Tsunami	90.00	90.00	100.00
Electronic Game Market	90.00	90.00	100.00
Internet programming	90.00	90.00	100.00
Average	89.84	90.22	99.27

2 Conclusion

Our results suggest that users typically use tags that appropriately characterise the content of the associated documents. Tags may thus help users group their documents into broad categories, and in turn, users may rely on tags to retrieve relevant content.

Similarly, Brooks and Montanez [3] analysed the value of tags for classifying blog entries from Technorati. However, our investigations diverge based on the domain. With blogs, the documents are more sentiment based. This is in contrast with the pages in our dataset, where it is more general.

A limitation is that the sample size for the study is small and generic. On-going work includes analysing with a larger data set and including tags from other domains.

References

1. Macgregor, G., McCulloch, E.: Collaborative tagging as a knowledge organisation and resource discovery tool. *Library Review* 55(5), 291–300 (2006)
2. Marlow, C., Naaman, M., Boyd, d., Davis, M.: HT06, tagging paper, taxonomy, Flickr, academic article, to read. In: *Proceedings of the seventeenth conference on Hypertext and Hypermedia*, ACM Press, Odense, Denmark (2006)
3. Brooks, C.H., Montanez, N.: Improved annotation of the blogosphere via autotagging and hierarchical clustering. In: *Proceedings of the 15th international conference on World Wide Web*, ACM Press, Edinburgh, Scotland (2006)