

Part III Individual Project Brief

Enhanced Content Analysis and Information Retrieval Using Twitter Hashtags

Author:

Jamie Davies

jagd1g11@ecs.soton.ac.uk

Supervisor:

Dr. Nicholas Gibbins

nmg@ecs.soton.ac.uk

Microblogging services such as Twitter¹ have become very popular in recent years. One of the key characteristics of such services is the use of ‘hashtags’ — topical/categorical annotations provided by the authors of the posts (tweets) themselves. These hashtags have been proven to be useful and information-rich in previous microblogging research projects; from using them to crowd-source real-time event detection [2], to using them to train sentiment classifiers [1]. However due to their nature, there is a vast and diverse collection of hashtags for users to choose from, resulting in the information provided by hashtags not being as accurate or as complete as it could be.

For my project, I will create a system that aims to support and enrich the information provided by hashtags. I will use the latest machine-learning techniques to examine and classify the topics and concepts behind the hashtags and in doing so, be able to suggest suitable new hashtags for tweets that are relevant to their content. This will allow users to make a better choice of hashtags when writing their tweets, and therefore refine the information they provide. Furthermore, I will create an extension of this system that will provide a context-aware tweet search facility. This will enable a user to search for a particular hashtag, and instead of only returning tweets containing that hashtag (as current systems do), it will also provide tweets that are contextually relevant to the search term but do not contain that given hashtag.

Finally, I will test the success and importance of my research by implementing a recent research experiment that depends upon hashtags (such as those conducted by Davidov et al. [1] or Sakaki et al. [2]). I will compare the results of the experiment both with and without my enhanced system, and provide validation for the project.

References

- [1] D. Davidov, O. Tsur, and A. Rappoport. Enhanced sentiment learning using twitter hashtags and smileys. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, COLING '10, pages 241–249, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics.
- [2] T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake shakes twitter users: real-time event detection by social sensors. In *Proceedings of the 19th international conference on World wide web*, WWW '10, pages 851–860, New York, NY, USA, 2010. ACM.

¹<http://www.twitter.com>