Feedback Predicition for Blogs

Krisztian Buza

Department of Computer Science and Information Theory
Budapest University of Technology and Economics, Hungary
`buza@cs.bme.hu`

Abstract. The last decade lead to an unbelievable growth of the importance of social media. While in the early days of social media, blogs, tweets, facebook, youtube, social tagging systems, etc. served more-less just as an entertainment of a few enthusiastic users, nowadays news spreading over social media may govern the most important changes of our society, such as the revolutions in the Islamic world, or US president elections. Due to the huge amounts of documents appearing in social media, there is an enormous need for the *automatic* analysis of such documents.

One of the most important properties which distinguishes social media from the classic one, is the uncontrolled, dynamic and rapidly-changing content: e.g. when a blog-entry appears, users may immediately comment this document. In this work, we focus on the analysis of documents appearing in blogs. We present an industrial application which has the following major components: (i) the crawler, (ii) information extractors, (iii) data store and (iv) analytic components. Analytic components allow to explore trends and to predict the number of feedbacks that a document is expected to receive in the next 24 hours. This task is related to opinion mining, however, despite its relevance, there are just a few works on predicting the number of feedbacks that a blog-entry is expected to receive, see e.g. Yano and Smith (2010). In contrast to them, we target various topics (do not focus on political blogs) and perform experiments with many different models. We hope that our observations will motivate research in order to improve classification and regression algorithms.

References

YANO, T. and SMITH, N. A. (2010): Whats Worthy of Comment? Content and Comment Volume in Political Blogs. *4th International AAAI Conference on Weblogs and Social Media, 359–362*

MISHNE, G. (2007): Using Blog Properties to Improve Retrieval. *International Conference on Weblogs and Social Media*

Keywords

SOCIAL MEDIA, BLOGS, FEEDBACK PREDICTION