Classifying Short Text in Social Media: Twitter as Case Study

Authors:
Faris Kateb
Jugal Kalita

Abstract

With the huge growth of social media, especially with 500 million Twitter messages being posted per day, analyzing these messages has caught intense interest of researchers. Topics of interest include micro-blog summarization, breaking news detection, opinion mining and discovering trending topics. In information extraction, researchers face challenges in applying data mining techniques due to the short length of tweets as opposed to normal text with longer length documents. Short messages lead to less accurate results. This has motivated investigation of efficient algorithms to overcome problems that arise due to the short and often informal text of tweets. Another challenge that researchers face is stream data, which refers to the huge and dynamic flow of text generated continuously from social media. In this paper, we discuss the possibility of implementing successful solutions that can be used to overcome the inconclusiveness of short texts. In addition, we discuss methods that overcome stream data problems.

References

- Fabian Abel, Qi Gao, Geert-Jan Houben, and Ke Tao. Semantic enrichment of twitter
- Adrian Chen. Can an algorithm solve twitter&amp;apos;s credibility problem?, 2014. [Online; posted 5-May-2014].
- Susan Dumais, John Platt, David Heckerman, and Mehran Sahami. Inductive learning models


- Lila MacLellan. Tweets per minute social media, 2012. [Online; posted 6-Sep-2012].
- M. M. Masud, Qing Chen, L. Khan, C. C. Aggarwal, Jing Gao, Jiawei Han, A. Srivastava, and N. C. Oza. Classification and adaptive novel class detection of feature-evolving data streams. Knowledge and Data Engineering, IEEE Transactions on, 25(7):1484–1497, July 2013.
- M. M. Masud, Jing Gao, L. Khan, Jiawei Han, and Bhavani Thuraisingham. Classification and novel class detection in concept-drifting data streams under time constraints. Knowledge and Data Engineering, IEEE Transactions on, 23(6):859–874, June 2011.
- Benjamin P. Wing and Jason Baldridge. Simple supervised document geolocation with
- Lei Yang, Tao Sun, Ming Zhang, and Qiaozhu Mei. We know what @you #tag: Does the dual role affect hashtag adoption? In Proceedings of the 21st International Conference on World Wide Web, WWW &apos;12, pages 261–270, New York, NY, USA, 2012. ACM.
- Xue Zhang, Hauke Fuehres, and Peter A. Gloor. Predicting stock market indicators through twitter &quot;i hope it is not as bad as i fear?&quot;&apos; Procedia - Social and Behavioral Sciences, 26(0):55 – 62, 2011. The 2nd Collaborative Innovation Networks Conference - fCOINs2010g.

Index Terms
Computer Science
Information Science

Keywords
Social Media Mining  Short Text Classification  Stream Data