Abstract

It is difficult to sidestep Big Data today, as the industry is abuzz with its promises. The trend is towards data-driven decision-making in all aspects of businesses because making sense out of data is very profitable and valuable. People tend to use social media, especially Twitter, to tweet about their opinions and sentiments. However, due to the prevalence of data that might be noisy, varied, unfiltered, and the impractical state of manually labeling large number of tweets to train classifiers, data acquisition for training sentiment analysis classifiers is becoming more and more of a challenge. This paper proposes a solution to easily acquire automatically labeled, filtered, and huge training data from Twitter in order to be given as input to a support vector machine classifier. The recommended solution discusses the workaround of unlabeled data through using Twitter hashtags to automatically induct the sentiment of a tweet (positive or negative). Neutral class is trained using tweets generated by newspapers accounts. A test study was conducted to show the accuracy of the applied features on the classifier. As a result, tweets trending on Twitter can now be analyzed to induce their sentiments which helps organizations in future data-driven decisions.
References


12. J. Milgram, M. Cheriet and R. Sabourin, "One Against One" or "One Against All": Which One is Better for Handwriting Recognition with SVMs?," in enth International Workshop on Frontiers in Handwriting Recognition, 2006.


Index Terms

Computer Science Artificial Intelligence
Keywords

Big Data, data-driven, Twitter, automatically labeled, training, support vector machine, unlabeled data, hashtags, newspapers.