Text Mining of Twitter Data Using a Latent Dirichlet Allocation Topic Model and Sentiment Analysis

Authors : Sidi Yang, Haiyi Zhang

Abstract : Twitter is a microblogging platform, where millions of users daily share their attitudes, views, and opinions. Using a probabilistic Latent Dirichlet Allocation (LDA) topic model to discern the most popular topics in the Twitter data is an effective way to analyze a large set of tweets to find a set of topics in a computationally efficient manner. Sentiment analysis provides an effective method to show the emotions and sentiments found in each tweet and an efficient way to summarize the results in a manner that is clearly understood. The primary goal of this paper is to explore text mining, extract and analyze useful information from unstructured text using two approaches: LDA topic modelling and sentiment analysis by examining Twitter plain text data in English. These two methods allow people to dig data more effectively and efficiently. LDA topic model and sentiment analysis can also be applied to provide insight views in business and scientific fields.

Keywords : text mining, Twitter, topic model, sentiment analysis
Conference Title : ICDM 2018 : International Conference on Data Mining
Conference Location : Dublin, Ireland
Conference Dates : July 23-24, 2018