Social Media Mining with R. Twitter Analyses

Authors : Diana Codat

Abstract : Tweets' analysis is part of text mining. Each document is a written text. It's possible to apply the usual text search techniques, in particular by switching to the bag-of-words representation. But the tweets induce peculiarities. Some may enrich the analysis. Thus, their length is calibrated (at least as far as public messages are concerned), special characters make it possible to identify authors (@) and themes (#), the tweet and retweet mechanisms make it possible to follow the diffusion of the information. Conversely, other characteristics may disrupt the analyzes. Because space is limited, authors often use abbreviations, emoticons to express feelings, and they do not pay much attention to spelling. All this creates noise that can complicate the task. The tweets carry a lot of potentially interesting information. Their exploitation is one of the main axes of the analysis of the social networks. We show how to access Twitter-related messages. We will initiate a study of the properties of the tweets, and we will follow up on the exploitation of the content of the messages. We will work under R with the package 'twitteR'. The study of tweets is a strong focus of analysis of social networks because Twitter has become an important vector of communication. This example shows that it is easy to initiate an analysis from data extracted directly online. The data preparation phase is of great importance.

Keywords : data mining, language R, social networks, Twitter
Conference Title : ICICT 2018 : International Conference on Information and Computer Technology
Conference Location : Zurich, Switzerland
Conference Dates : January 15-16, 2018