Project: Classification of Tweets

Max Savery

April 4, 2018

A while ago I built a twitter scraper and plan to use data I will generate or have already generated for this project. The scraper is pretty flexible, and can stream tweets based on hashtags, @mentions, users, or words/phrases, or can download some specified amount of individual users or groups of users tweets. I built the scraper so I could easily get a general idea of what people are thinking/talking about on twitter, and no matter what data I end up using, that's the question I will have in mind.

As far as machine learning goes, I would like to use multinomial naive bayes or/and SVM to classify tweets into topic groups. Algorithm recommendations would be appreciated... Another option is to classify tweets by emotional content e.g., frustrated, cranky, worried. A package, TextBlob, makes that pretty easy and has a pretrained method I can use. But if I do the subtler emotional categories, I will have to hand annotate a training set of tweets, which, depending on whose tweets they are, might be fun or might not be.