Quantitative Text Analysis: Classwork 10

Data Mining from Social Media

In this exercise you will try out R code for using the Twitter REST and streaming APIs. R source files containing example code are available on your S drive in the zip file twitter.zip.

Instructions

1. Extracting Twitter data.

   (a) Download twitter.zip to your M drive and extract the files. Files with the extension .R contain example R code, and the file my_oauth.Rdata contains the authentication key.

   (b) Open RStudio, and at the console enter the command `getwd()`. The will display the working directory for RStudio. You must copy the authentication file 'my_oauth.Rdata' to this folder. If the working directory is set to a folder you don't have access to, change it with the command setwd(), e.g. `setwd('m:/pc/My Documents/')`.

   (c) Open the R source files in R studio and step through the examples by highlighting each line and clicking 'Run'. Remember that lines beginning with a '#' are comments and will not be run. You must load the libraries containing the functions you need with the command 'library'. If a library is not installed you may see an error like: 'There is no package called 'twitteR'. In this case, install the package you need, e.g. install.packages('twitteR').

   (d) View the documentation for a command by entering the name of the command preceded by a '?' at the console, e.g. `?searchTwitter)`. Use '??' to search the documentation.

   (e) After stepping through the example code, explore parts of the R API packages that interest you and try out new commands and searches.

2. Measuring sentiment in Twitter data

   (a) Download and run the file `phones.R`. The best way to do this is to step through each line and execute it line-by-line, watching what happens in the console. (Note: This file requires that you have internet access to the data files defined as URLs.) The end result of this file is to create an output in comma-separated-value (.csv) format named `twitter_phones.csv`.

   (b) Start QDAMiner, and create a new project from a data file. Locate and load the file `twitter_phones.csv`.

   (c) Inspect the results in the text and variable windows to ensure the import occurred successfully.

   (d) Transform the variables PHONE and USER to categorical variables. This can be done in the VARIABLES window by right-clicking the variable name, and choosing Transform PHONE (or USER): "String $\rightarrow$ Nominal/Ordinal".

   (e) Launch Wordstat after choosing "In relation with" the variables [PHONE] and [USER].

   (f) Try sentiment analysis using one or more of the sentiment dictionaries, and inspect the results in the Cross Tab pane by the variable PHONE. Which type of phone seems to be receiving more positive sentiment?

   (g) Return to QDA Miner and use the Cases: Filter option to select values of the FOLLOWERS variable that are greater than zero, and repeat the previous analysis. Did the results change?

   (h) Experiment with selections on the [USER] variable by Cross Tab. Can you detect any interesting patterns here?