Research project proposal

Your Name
Date

Sentiment Trend Analysis of Public Opinion using emoticons

Sentiment Analysis aims to determine the attitude of a speaker or a writer with respect to some topic or the overall contextual polarity some text. Sentiment analysis including news, social networking and blogs is used for monitoring trends and opinions. Extracting public opinion from social media is a motivating research and as people increasingly use emoticons in text in order express their opinion, it is important for sentiment analysis tools to account for graphical cues for sentiment.

For sentiment analysis, I plan to reuse the algorithm that Brendan O'Connor, Ramnath Balasubramanyam, Bryan R. Routledge and Noah A. Smith designed. My problem considers twitter messages which uses emoticons significantly. Since it is important to consider emoticons during sentiment analysis, I plan to implement a system to detect emoticons in the algorithm to have an improved version of lexicon-based sentiment classifier.

Background

From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series

From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series (O'Connor et. al.). O'Connor et. al. designed metrics to connect public opinion measured from polls and tweets. They analyzed surveys on consumer confidence and political opinion, using Gallup's Economic Confidence Index and Michigan's Index of Consumer Sentiment, over the period of 2008 and 2009 and correlated the sentiment word frequencies. And, concluded that text streams can substitute and supplement traditional polling methods.

Though their correlations were as high as 80% and captured important large scale trends, they only counted instances of positive and negative sentiments in each message which resulted in a message being counted as both positive and negative. One of the major disadvantage in their sentiment analysis was that they did not use a system to detect alternately spelled words and emoticons. Since it involved analyzing twitter messages and lot of such messages contain informal social media dialect, exploiting emoticons in sentiment analysis would have led to an improved version of the lexicon-based sentiment classifier.

Data sources

I plan to use Twitter data using the Twitter public REST API https://dev.twitter.com/rest/public

I plan to use Twitter's REST API using R.
http://www.r-bloggers.com/talking-to-twitters-rest-api-v1-1-with-r/

We will also be archiving the "Fire hose" real-time stream. See
https://dev.twitter.com/streaming/firehose

Algorithms

I plan to use the following algorithms:

- Naive bayes algorithm - https://en.wikipedia.org/wiki/Naive_Bayes_classifier
- Maximum Entropy Classifier -
  https://en.wikipedia.org/wiki/Principle_of_maximum_entropy
- Decision Tree - https://en.wikipedia.org/wiki/Decision_tree
- Support Vector Machines - https://en.wikipedia.org/wiki/Support_vector_machine

For text processing – the following packages can be used:

- tm
- tau
- language
- scrapeR

For sentiment analysis –

I plan to use R's package "sentiment" -
https://sites.google.com/site/miningtwitter/questions/sentiment

Evaluation of the project

The project evaluation can be divided into two parts –

The first part consists of successfully evaluating the existing algorithm by measuring my results against previous results obtained using this sentiment classifier.

The second part consists of extending this algorithm to an emoticon based sentiment analysis and checking if there is any improvement in the sentiment analysis.

The third part consists of evaluating the algorithm versus a number of public opinion datasets (see Evaluation Data)

Evaluation Data

There are a number of public opinion datasets:

Consumer Confidence Index from the Consumer Board https://www.conference-board.org/
Index of Consumer Sentiment (ICS) from the Reuters/University of Michigan Surveys of Consumers at http://www.sca.isr.umich.edu/
pollster.com http://www.pollster.com/polls/us/08-us-pres-ge-mvo.html
Gallup Poll Data – daily economic indexes http://www.gallup.com/poll/122840/gallup-daily-economic-indexes.aspx
Gallup Daily: Obama Job Approval http://www.gallup.com/poll/113980/Gallup-Daily-Obama-Job-Approval.aspx

I will continue to look for more polling data sets.

References

O'Connor, Brendan, et al. "From tweets to polls: Linking text sentiment to public opinion time series." Proceedings of the International AAAI Conference on Weblogs and Social Media. 2010.

Alexander Hogenboom , Daniella Bal , Flavius Frasincar , Malissa Bal , Franciska de Jong , Uzay Kaymak, Exploiting emoticons in sentiment analysis, Proceedings of the 28th Annual ACM Symposium on Applied Computing, March 18-22, 2013, Coimbra, Portugal

http://www.sentiment140.com/