Data Analytics Project 4

Counterpoise Pro-ISIS Sentiment Analysis

Text analytics has become an important tool for assessing sentiment in a variety of situations, including customer service, social media, and written works. Sentiment is defined as an immediate emotional response to an issue that may require immediate assistance.

Terrorism becomes more severe these days, especially the attacks sponsored by Islamic State of Iraq and Syria (ISIS) or Daesh. They are experts of using social network for propaganda and recruitment, thus predicting their activities through big social network data will help to predict and avoid more serious attacks.

So-called ISIS Fanboys have been actively tweeting all over the world. A fanboy is a boy or man who ardently supports a single hobby, ideology, movement, etc. Isis released #TweetMovie from twitter, a "normal" day when two Isis operatives murdered a priest saying mass in a French church. Fanboy tweets obviously would suggest certain sentiments. However, there are also many tweets that perhaps counterpoise the ISIS fanboy tweets. Kaggle release a selection of data from the site and made it available to Kaggle users. You can find the dataset in my GitHub repository, with the filename: ISISFanboy. This data set is intended to be a counterpoise to the *How Isis Uses Twitter* data set. That data set contains 17k tweets alleged to originate with "100+ pro-ISIS fanboys". This is not a perfect counterpoise as it almost surely contains a small number of pro-Isis fanboy tweets.

You goal is to perform a counterpoise analysis to pro-ISIS tweets. A counterpoise provides a balance or backdrop against which to measure a primary object, in this case the original pro-Isis data. So, if anyone wants to discriminate between pro-Isis tweets and other tweets concerning Isis you will need to model the original pro-Isis data or signal against the counterpoise which is signal + noise.

These datasets are available in my VIT GitHub directory:
AboutISIS
ISISFanboy
How Isis Uses Twitter

Things to consider:

When a Arabic word is translated to English as "love" or "joy," did the person who wrote the tweet in Arabic have the same meaning in his mind as we do when we read the words in English. In other words, is there a cognitive semantics difference between the Islamic meaning of an Arabic word, which is then translated to English? None of the results I have seen adequate reflect this dilemma.

You should perform this analysis in R or Python. R code is available on Kaggle, but you'll have to

find it as part of your research. It is okay to use the code as is (with proper credit given), but keep in mind that you must examine both the pro-ISIS and anti-ISIS tweets.

Requirements:

Analysis: You may use Rstudio or Python (including Jupyter Notebook) to perform your analysis.

A **white-paper** report (Google what this consist of) with a maximum of five pages. You may add an appendix containing your detailed analysis results (no more than 10 additional pages). An appendix with your model code must be attached. This will give you the foundation of a conference paper.

A **PowerPoint slide deck** consisting of a maximum of about 10 slides:

1. Title slide
2. Business problem statement
3. Modeling objective
4. Model functional form and reasoning for its selection
5. Final model (variables with coefficients and explanation) summary
6. 1 deep-dive slide per model variable (a good model may only have three or four factors, even if all variables are statistically significant—modeling is an "art").

Each slide in a PowerPoint slide deck needs to stand alone yet have few words and contain one graphic (chart, graph, table, picture, etc.) with words, but not verbose. Each slide should tell a story that an executive will understand. If you cannot effectively communicate your analysis, it is worthless—a model is only good if it is used.

Stages of Model Development

1. Business problem development and model requirement definition
2. Model design plan, including modeling objective and model functional from, plus potential data sources
3. Model development
    a. Data preprocessing
    b. Exploratory analysis
    c. Model training
    d. Model testing
4. Model post processing analysis
5. Model validation
6. Model Implementation

Citations: Use ISO 690 with Numerical Reference.