Big Data Project Update Report

Clouded Minds

by Deepak Gautam, Sandra Lee Gibson, Michael S Nichols, William Wheeler Carter

We have collected tweets for about 10 games of English Premier Soccer League. While this isn't a large dataset yet, it will allow us to create our baseline for future data. Thus, for our project we plan to determine match event using tweets. In order to do so, we will use two separate approaches, tweet volume analysis and sentiment analysis.

Initially, we used Pandas to create a notebook of graphs that showed us the volume of data spread over the period of the game on minute by minute basis. That way, we could better understand how to correlate number of tweets to the events of the game. From there, we attempted to determine game events based on the tweets per minute value. Though, this gives the most important events in a game, this method will miss the smaller, less important events. As a result, we are planning to analyze tweets by calculating slopes of the spikes when volume vs time graph is plotted to get detailed match events. After determining match events, we will be doing sentiment analysis on tweets in those events to determine what exactly happened in those moments.

For sentiment analysis, we plan to group tweets by the favorite team of the twitter users according to their usage of hashtags. Based on the sentiment values of one set of tweets from fans of one team, we are trying to determine if that team lost or won, as well as the final score of the game. Previously, we attempted to analyze the tweets using an online sentiment analysis API provided by Algorithmia, but we felt that the volume of our

data would cause us to lose access or work too slowly due to API request limits. Therefore, we are researching alternatives based around wordlists we have found for sentiment analysis on Twitter, since the text may have short-hand or misspellings. Our goal from this point is to find a library for sentiment analysis based on our wordlists.

In conclusion, we'll be using the tweet volume analysis in combination with the sentiment analysis to determine the results of the games and compare the outputs with real game results to see how accurate our model has been. For now, we need to find a proper sentiment analysis library for our needs before moving on to the next tasks.