Sentiment Analysis on Soccer Players using Twitter Data

Proposal 2

Aravind Padmanabhan
Computer Science
Binghamton University
Binghamton, New York
apadman1@binghamton.edu

Ridhuvarshan Natarajan
Computer Science
Binghamton University
Binghamton, New York
rnatara2@binghamton.edu

Chethana Gopinath
Computer Science
Binghamton University
Binghamton, New York
cgopina1@binghamton.edu

ABSTRACT

In the initial report of our project, we had planned to analyze racist tweets on soccer players. But as mentioned in the previous report, there weren't a considerable number of tweets directly racist for us to perform our analysis. Only tweets that contained information about racist attacks were present (ex: news articles and general discussions about the said events). So, we decided to broaden our line of research, and the research questions we are trying to answer include

- Finding out the popularity of the English Premier League in other countries.
- The reaction of the fans of particular clubs(in Twitter) to the major happenings around a team, - for example, sacking of a manager, players stripped off captaincy, injury of star players during the season or transfer of players to other teams, match results between said team and an opposing team and such.

Our teams of interest include Arsenal, Men United, Man city, Chelsea, Liverpool, Tottenham Hotspur. We would also be following the performance of these teams across Champions League, Europa League and Premier League. This list is not exhaustive, we may also include teams who are current season favorites and who may not be the usual favorites.

IMPLEMENTATION

As previously mentioned, we collect data from the hashtags which we are using to track the various teams and the general opinion and emotions around the teams mentioned. To answer our research questions, we wanted to find a powerful parser as well as a tool that can aid us in performing sentiment analysis on the tweets. To find out the popularity of the English Premier league, we would be working with using geotagged tweets and based on specific tags of clubs along with the "Tweepy" module in Python. After that, we would mostly be able to derive meaningful results by generating either a heat map or a data frame from the results. As for the sentiment analysis, we would need to utilize a good NLP tool to parse and classify our huge chunk of Twitter data collected based on our hashtags. This, we decided to implement using "spaCy", an NLP library which we will use to classify our twitter stream data. We will use spaCy to classify the text and based on certain value, will classify fan's reactions as either "positive", "neutral" or "negative" based on an assigned integer value. There may be other parameters which we may take into consideration. We may also end up doing a fine-grained sentiment analysis on our data. Following any major happening, as discussed in the abstract, there usually comes a burst of reactions from Twitterati who follow the clubs. So, hopefully as this is during the season, we would have enough data to derive meaningful insights from fans of the clubs we are following. Based on the analysis of the tweets, we would be able to provide our results in either a bar graph that shows the polarity of the fans for that particular club.

So far, we have been collecting tweets based on the hashtags provided in the earlier report. Based on the effectiveness and success of our approach, we may extend our research to tweets pertaining to NBA as well.

REFERENCES

https://monkeylearn.com/sentiment-analysis/

https://spacy.io/usage/rule-based-matching

https://www.iflexion.com/blog/sentiment-analysis-python