EE239AS - Project 4

Yuxin Jin (104195828), Jamie Lee (604589757),
David Hong (204588953) and Nick Cirillo (103834979)

March 18, 2016

1 Popularity Prediction on Twitter

With the increase in popularity of social media and communication, Twitter has emerged as a major platform for online networking which allows users to post and read 140-character messages called tweets. Tweets are publicly viewable via the website and users can subscribe to other user's tweets in the form "following" another user. A word, phrase or topic that is mentioned at a greater rate than others is said to be a "trending topic" and can often be recognized in the form of hashtags ie. #TrendingTopic. These topics become popular usually either through a specific event or topic that prompts people to discuss it or by purely users creating these discussions. This information easily allows users to obtain broad perspectives and recent updates. In addition to users, these trending and bursting topics have sparked recent interest in the scientific community to analyze and predict these topics.

In this project we seek to analyze such data by utilizing current and previous tweet activity for given hashtags in order to predict future tweet activity and behavior. Specifically we analyze tweet data for the 2015 Super Bowl (Seattle Seahawks vs. New England Patriots) from a period of two weeks before the game to a week after the game. We will use from related hashtags to train a regression model and use the same model to make predictions for other hashtags.

1.1 Part 1:

The training data was downloaded and statistics were calculated for each of the hashtags (#gohawks, #gopatriots, #nfl, #patriots, #sb49 and #superbowl). We report data on each hashtag for the total number of tweets, average number of tweets per hour, average number of retweets and average number of followers per users. Statistics for each hashtag can be found in the tables below.

The total tweet count per hour both both the #SuperBowl and the #NFL hashtags are also shown below. Both hashtags noticeably show initial trending prior to the SuperBowl, show significant bursting during the SuperBowl event and both die relatively quickly after the SuperBowl (Feburary 1, 2015).

1.2 Part 2:

Total # of Tweets	188135
Avg. # of Tweets per Hour	276.669118
Avg. # of Retweets	0.209164
Avg # of Followers of (77168) Users	1720.634084

Table 1: #gohawks statistics

Total # of Tweets	26231
Avg. # of Tweets per Hour	58.551339
Avg. # of Retweets	0.026838
Avg # of Followers of (18005) Users	1559.278200

Table 2: #gopatriots statistics

Total # of Tweets	259019
Avg. # of Tweets per Hour	420.485390
Avg. # of Retweets	0.050938
Avg # of Followers of (75167) Users	4399.303205

Table 3: #nfl statistics

Total # of Tweets	489710
Avg. # of Tweets per Hour	736.406015
Avg. # of Retweets	0.091462
Avg # of Followers of (326173) Users	1865.903974

Table 4: #patriots statistics

Total # of Tweets	826905
Avg. # of Tweets per Hour	1531.305556
Avg. # of Retweets	0.178023
Avg # of Followers of (590066) Users	2247.285607

Table 5: #sb49 statistics

Total # of Tweets	1348766
Avg. # of Tweets per Hour	2207.472995
Avg. # of Retweets	0.136686
Avg # of Followers of (689690) Users	4228.627878

Table 6: #superbowl statistics
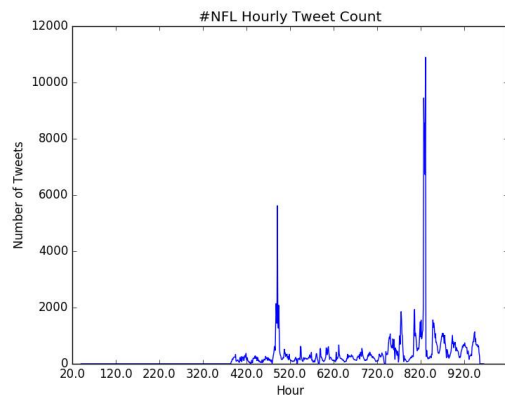
(a) #NFL Hourly Tweet Count
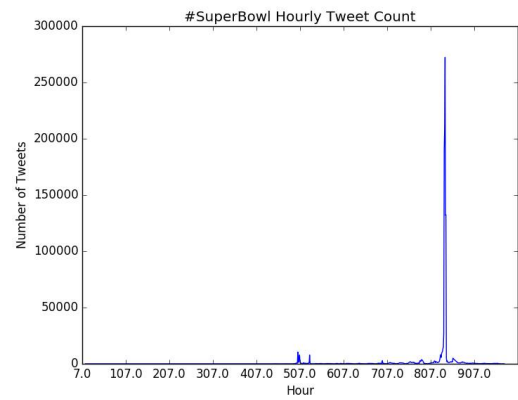
(b) #SuperBowl Hourly Tweet Count

Figure 1: Tweet counts per hour for tweet data obtained from 2 weeks prior to a week after the 2015 Superbowl