

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)

#gohawks:

Total number of tweets 188135

Average number of tweets per hour 276.669118

Average number of retweets 0.209164

Average number of followers of (77168) users 1720.634084

#gopatriots:

Total number of tweets 26231

Average number of tweets per hour 58.551339

Average number of retweets 0.026838
Average number of followers of (18005) users 1559.278200

#nfl:

Total number of tweets 259019
Average number of tweets per hour 420.485390
Average number of retweets 0.050938
Average number of followers of (75167) users 4399.303205

#patriots:

Total number of tweets 489710
Average number of tweets per hour 736.406015
Average number of retweets 0.091462
Average number of followers of (326173) users 1865.903974

#sb49:

Total number of tweets 826905
Average number of tweets per hour 1531.305556
Average number of retweets 0.178023
Average number of followers of (590066) users 2247.285607

#superbowl:

Total number of tweets 1348766
Average number of tweets per hour 2207.472995
Average number of retweets 0.136686
Average number of followers of (689690) users 4228.627878