

EE 239AS
Special Topics in Signals and Systems
Project 4
Popularity Prediction on Twitter
Winter 2016

Liqiang YU, Kaiming WANG and Jun FENG
904592975, 504592374, 304588434

03-18-2016

Contents

| | | |
|----------|----------------------|----------|
| 1 | Introduction | 3 |
| 2 | Data Analysis | 3 |

1 Introduction

The Twitter website, as the most famous social network, is a good source to predict future popularity of a subject or event. In this project, we analyzed the data from twitter crawled during the period of 2015 superbowl, from two weeks before the game to one week after the game. We trained different regression models for tweets with different hashtags to predict the number of tweets in the next hour. We included features both from the tutorial and our own design. And we used the test set to evaluate our model's prediction results. Finally, we came up with a new idea based on the rich data from the twitter, that is try to describe the emotion change of fans from both teams during and after the game, based on the contents from their tweets.

The report is organized as follows : in section 2, we did a quick scan through the file and got some statistics information about the data....

2 Data Analysis

We have six hashtags for training : #gohawks, #gopatriots, #nfl, #patriots, #sb49, #superbowl. The data information is shown in table 1. The distribution for #superbowl and #nfl are shown in figure 1 and 2. We only consider the time period of interest, which is two weeks before Feb 1st, 2015 and one week after it. From the histograms we can see that both hashtags concentrated during the game time, especially for #SuperBowl. For #NFL, there was another peak during the last weekend before the superbowl final game, which showed that people would like to talk about the NFL game during weekends.

Table 1: The data information for each hashtag

| | #gohawks | #gopatriots | #nfl | #patriots | #sb49 | #superbowl |
|--------------------------------------|----------|-------------|---------|-----------|---------|------------|
| Average number of tweets per hour | 380.84 | 53.43 | 515.98 | 975.52 | 1647.31 | 2692.14 |
| Average number of follower per users | 1544.97 | 1298.82 | 4289.75 | 1650.32 | 2235.16 | 3591.60 |
| Average number of retweets per tweet | 2.01 | 1.40 | 1.54 | 1.78 | 2.51 | 2.39 |

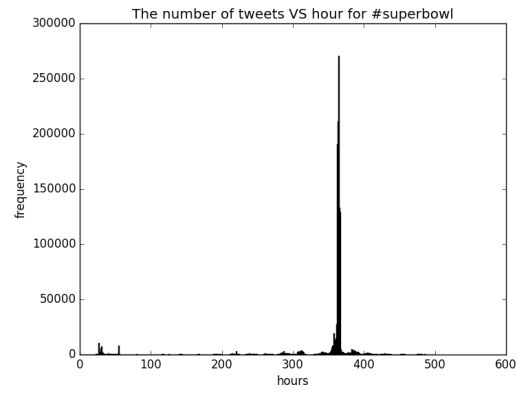


Figure 1: The histogram for #SuperBowl

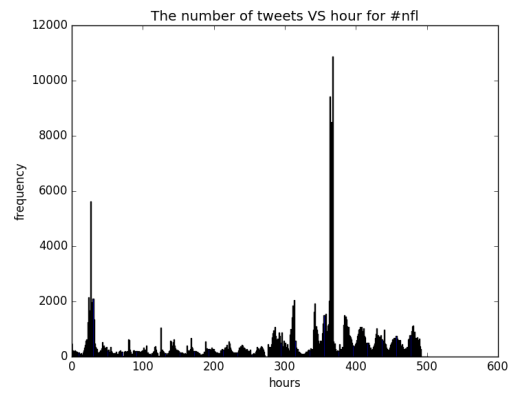


Figure 2: The histogram for #NFL