Project Name: Courtside Tweets

Team Members: Audrey Cole, Henry Moyerman, Danny Munro, Bradley Settle

Project Description

Our team is looking to create a tool designed to help people or companies determine when to tweet about a topic in order to get the most retweets. We are looking at focusing our software on predicting the best time for sports related posts. Since the Basketball season will continue through the end of the semester, it provides an excellent source of data for our preliminary program. In order to build this program, we are going to use an Amazon EC2 running a LAMP stack to serve a website that will allow the user to interact with the dataset. The web interface will allow a user to select a team from the NBA and will be told when would be the best time before its next game for any brand to tweet about the team. In order to collect our data we will be using the streaming API provided by Twitter. As suggested by Dr. Miller, we also considered a stationary data set for training purposes named Stanford SNAP, however the data has unfortunately been removed online at Twitter’s request. A python program will subscribe to this stream of data and get the tweets related to each team. The information obtained will be inserted into a database that we will use to do our analysis. Our team has made significant progress on the setup of our system. We have setup an EC2, configured the LAMP stack, and created a python program that listens to Twitter’s API and get live JSON formatted tweets.

Our primary objective is to maximize visibility of each tweet - this would be beneficial both for individuals looking for retweets and for companies looking to promote their brand (however the application will be targeted toward companies). Since fans are most excited about a game during the time leading up to it, and are likely more inclined to buy food and souvenirs in preparation for the game, leading time prior to the game is particularly fertile time for brand advertising. While it would be theoretically possible for such entities to regularly send tweets per unit time, this would likely annoy twitter followers and diminish the tweets’ meaning. In a twittersphere which has become increasingly crowded, the importance of high-quality (including timing) of target tweets will likely outweigh the importance of high-quantity tweets.

We used the Python Twitter API in order to fetch tweets using the Streaming API. We created a Twitter Dev account in order to get access\_token’s and consumer\_key’s to connect to this streaming api and get a live sample of tweets. We ran a python script for about a week or more during game season to fetch enough data. The data is in JSON format in a text file, and was collected on the basis of containing certain NBA team-related keywords such as “Bulls”, “Spurs” and “Knicks”. We parsed this data to import it into the database through a Python script. From there we will use PHP and SQL queries to derive an optimal time to send a tweet for a given team and season (of course, we are only collecting data for one season). The user will also have the option to search for the time for each individual game that yielded the highest visibility, or the highest visibility over all of them. Visibility will be determined by the sum of retweets and favorites for a tweet.