Big Data Project Update

Analyzing Real-Time Tweets to Generate Game Report

By
Clouded Minds

Project Goal

Game Report generation based on tweets posted by users watching game live.

- Important moments detection
  - will be detected by analysis of spikes in the volume of tweets.
  - sentence analysis is performed on those moments to describe the moment (What exactly happened at that moment).
  - Sentence analysis: Clustering on tweets of important moments.
- Sentiment Analysis
  - Collected tweets will be divided into two groups for each team based on hashtags.
  - Sentiment value for each team is calculated.
  - Correlation of sentiment value to actual match result (by what margin one team lost or won).

Thus far

- Data Collection
  - English Premier League Soccer games.
  - Collected tweets of about 10 games.
  - Filtered using official game hashtags, team names and their short names.
- Data Preprocessing
  - Stored in mongodb.
  - Conversion of twitter json format into csv format.
  - tweet timestamp translation into game minutes.

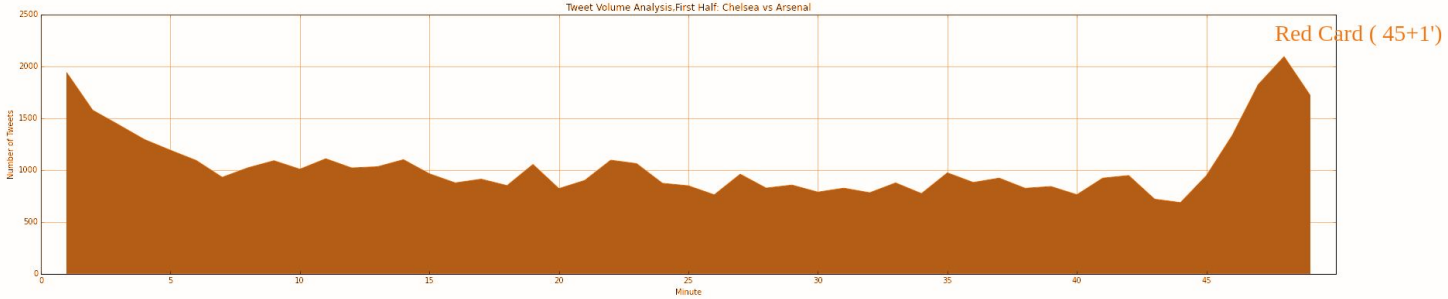
Thus far...

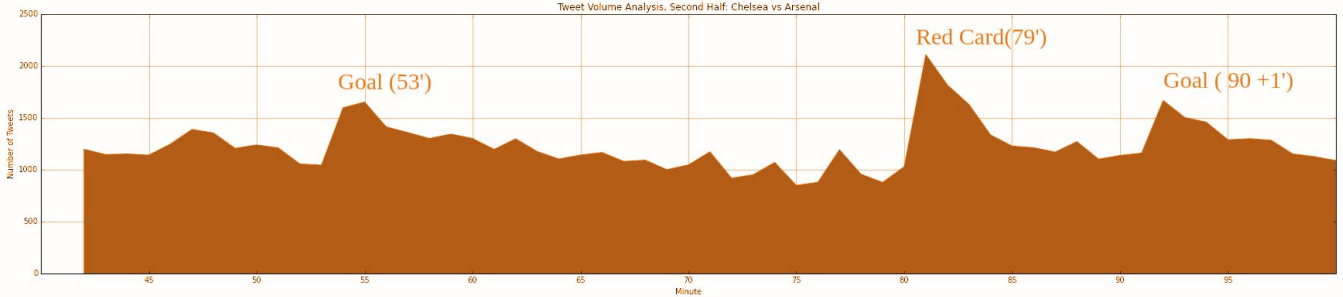
Analysis

- Based on tweet volume
- Tweet Count:
  - Gives high accuracy in detection of top moments of the game.
  - Goals, Red Cards and missed chances are detected with high accuracy.
- Results
  - Chelsea vs Southampton
  - Chelsea vs Arsenal

Thus far...

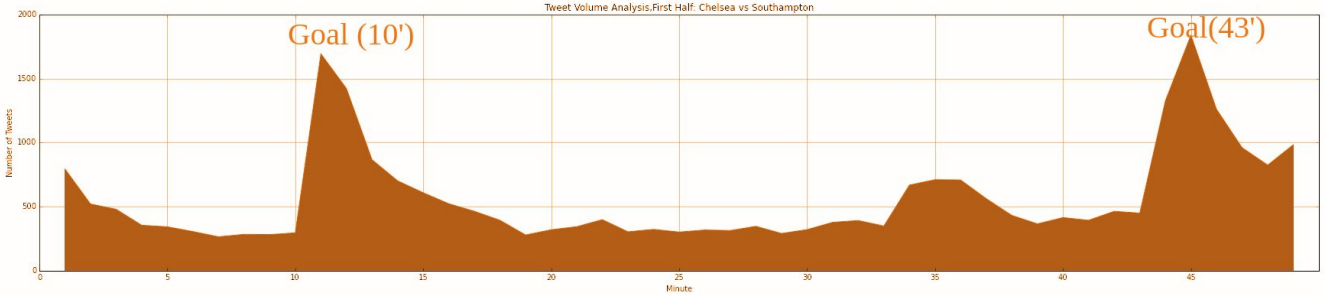
Chelsea vs Arsenal

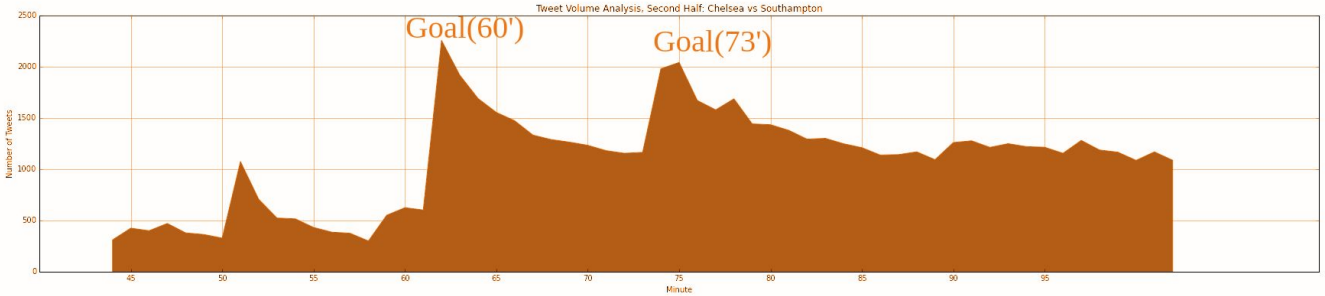




Thus far...

Chelsea vs Southampton





Thus far...

Analysis

- Sentiment Analysis
  - Attempted to use online sentiment analysis API.
  - Due to the limit on number of requests, still searching for alternatives.

To do

Detailed volume analysis

- Tweet Count
  - May not sufficient to detect less important moments such as yellow cards, substitutions.
  - But, evaluation of slope of each spikes in the volume of tweets can give better result for less important moments.
  - Spikes can be detected using sliding window on tweets over the time.
- Sentence Analysis by clustering on tweets of detected moments
- Sentiment Analysis

To do...

Noise Reduction

- Removing off-topic tweets in the detected moments is huge challenge.
- Similarity detection for words like ‘Goal’ or ‘Goooooaaal’ for sentence analysis.
- Removing spellings errors before sentiment analysis.