TweetVibe: A Realtime Analytics to determine the Popularity of TV Dramas

Kiran Varghese, Mohammed Farhan, S Vaisakh

Problem Statement

Television viewership ratings represent the popularity of TV programs and are an important indicator of the revenue generated by broadcasters from advertisements as well as other sources. Although higher ratings for a given program are beneficial for both broadcasters and advertisers, little is known about the factors that make programs more attractive to viewers.

In this project, we propose a method to predict the ratings of TV dramas before they are broadcasted by considering the cast and the staff involved with them. In order to consider the popularity of actors, we consider the number of hits received by their Wikipedia pages, and tweets related to them on Twitter.

We tested our proposed method using a collection of several TV dramas.

Design

- We formulated a system which fetch the tweets in real time using Tweet Streamer.
- The tweets are forwarded to DStream using a netcat utility.
- The Streaming Job receives the tweets from DStream and determines the sentiment of the tweets.
- The data by analysing various TV Shows is plotted in the form of a graph.

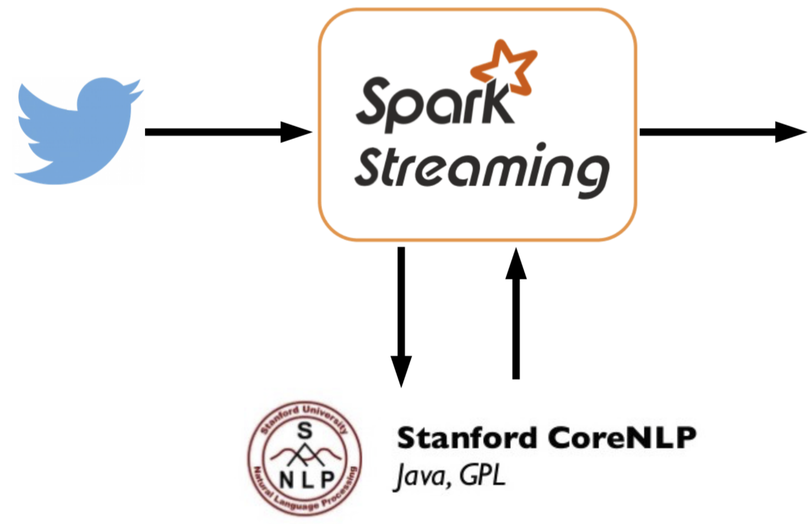
Technology Stack

Apache Spark is an open-source cluster-computing framework. Spark provides an interface for programming entire clusters with implicit data parallelism and fault-tolerance.

Scala is a general-purpose programming language providing support for functional programming and a strong static type system.

Stanford CoreNLP's goal is to make it very easy to apply a bunch of linguistic analysis tools to a piece of text.

Pipeline



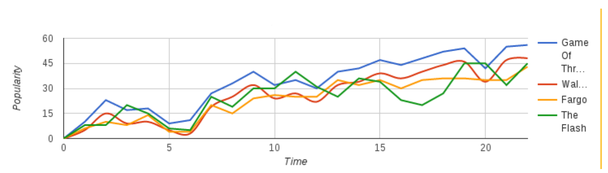
The basic Pipeline of our project is given above.A python script get twitter tweets depending on the movie which the user selects.After that, we piped these tweets using a socket to spark streaming job. The spark streaming jobs use a 60 seconds window to calculate the number of mentions each of the relevant features have, the sentiment of these mentions, how many positive tweets each of the features are receiving, as well as where the tweets are coming from.

Console Output

```
Time: 1490782729000 ms
-------------------------------
("game of thrones",(5,-3,0,2))
("walking dead",(1,-1,0,0))
("fargo",(5,-3,0,2))
("mr robot",(1,0,0,1))

-------------------------------
Time: 1490782729000 ms
-------------------------------
(("CO","game of thrones"),1)
(("GA","game of thrones"),1)
(("FL","game of thrones"),1)
(("MN","fargo"),1)
```

Result



We analysed the popularity of Game Of Thrones,The Walking Dead,The Flash and Fargo.After Performing the Streaming Job,we found that Game of Thrones is the most popular show ,because most of the People were talking positively about it in Twitter than other Shows.

A Future Direction

In research related to our work here, this can be further used to predict the financial success of movies based on the number of editors and viewers contributing to Wikipedia articles on the movies.

Information from social media and search engines is often used to predict not only the popularity of Web content but also events in the world. As research that deals with TV dramas, by using the number of posts, likes, comments, and shares on fan pages of TV dramas on Facebook to predict audience ratings. By considering social networking services-related factors, fluctuations in audience ratings following the first episode of a series can be predicted with high accuracy.

References

1. Stanford CoreNLP âĂŞ a suite of core NLP tools https://nlp.stanford.edu/software/corenlp.shtml
2. Apache SparkâĎć - Lightning-Fast Cluster Computing http://spark.apache.org/
3. The Scala Programming Language https://www.scala-lang.org/
4. Audience Ratings Prediction of TV Dramas Based on the Cast and Their Popularity http://ieeexplore.ieee.org/document/7545038/