TV Shows Muster and Analysis

Aditya Bhagwat
Binghamton University
abhagwa1@binghamton.edu

Abhimanshu Mishra
Binghamton University
amishr11@binghamton.edu

Sharvari Joshi
Binghamton University
sjoshi14@binghamton.edu

Vinit Bhosale
Binghamton University
vbhosal1@binghamton.edu

ABSTRACT

Online entertainment has become a crucial part of life in the modern age. In recent times, the amount of content and the number of content producers has increased exponentially. With such an increased consumption of content, everyone has opinions about different TV shows and water cooler discussions have become commonplace. People take to different social media forums to express what they think and connect with the world at large. This provides researchers with a unique opportunity to study the pulse of the public about a specific show easily.

This project aims to collect data about TV Shows from different social media platforms on a large scale. We target IMDb and Twitter for the same. IMDb is a movie/tv show rating and review website which is very popular around the world. It is trusted in most circles and people use it for recommendations of content. Twitter is one of the largest social media forums in existence. It has fast become the platform of choice for people to express their feelings and broadcast it on a large scale.

After data collection, appropriate sanity checks and analysis of this data will be performed. Furthermore, we propose an exploration of sentiment analysis techniques using this collected data.

1 DATA SOURCES

1.1 Twitter

Twitter is a micro-blogging and social networking service where users post and interact with messages, "tweets," restricted to 280 characters. Information can be shared using photos, videos and links as and when they're happening, enabling insightful and real-time search results.

With over 330 million active users and 500 million tweets per day, Twitter becomes an excellent source of data collection which can help us discover interesting patterns based on user activities and current trends. Using the Twitter API we can perform complex queries like retrieving tweets, user information including their followers and friends they follow as well as information about trending topics and their location.

1.2 IMDb

IMDb (Internet Movie Database) offers an extensive database of information related to films, television program, video games, and streaming content online – including crew details, plot summaries, trivia, fan and critical reviews.

IMDb currently contains approximately 6.5 million titles (which includes movie and TV show titles), 10.4 million personalities as well as 83 million registered users. Using the OMDb API, we will be gathering data about TV shows, their ratings, votes by users, reviews, genre it belongs to, release dates and other such information.

2 DATA COLLECTION

We will start by running a Daemon script on the virtual machine to collect data using Twitter and OMDb API once in three days. Based on the launch of TV shows on worldwide streaming platforms, the popularity of these shows may vary with time which is one major factor of consideration for the task at hand. In the process of data collection, if we notice a decrease in a show's ratings, we might replace it with another show with a better popularity score for that week.

On an average, the number of tweets to be collected in the given three day period would roughly be around 1000 tweets and about 1500 reviews from OMDb. We estimate about 200 to 450 meaningful tweets and reviews from the collection of every three day data. Overall we expect to collect about 3000-5000 reviews and tweets in order to get a fairly justifiable sample at hand to develop insights.

3 CONCLUSION

Based on the collected data using the Twitter API and the OMDb API , we will be able to study and develop insights about the User reviews, influence of reviews on the users, trending shows based on the popularity ,popularity of the shows region wise and also the deep insight of a review considering retweets and likes as the parameters.