**MOVIE RATING SYSTEM BASED ON SENTIMENTAL ANALYSIS OF TWEETS**

*By Nahush Raichura, Justin Lee, Haein Oh (Team 2)*

*School of Informatics, Computing, and Engineering*

*Indiana University*

*Bloomington, IN 47405*

**I. Introduction**

Twitter has over 300 million active monthly active users, the world sees Twitter as a social media platform to express their feeling towards anything be it politics, sports or be in our case movies that are in trend or any past instances. This is done simply by writing text limited to 280 characters accompanied by pre-defined or new hashtags. These hashtags are the backbone of any tweet for researchers and developers that are using these tweets for their research or analysis based on tweets. The hashtag gives us a gist of what the tweet is about such as movie names, a person name and team name etc.

Also, there are about 330 million tweets on a daily basis from all parts of the society including the president, politician, musician, sports-person, student, engineer, or a doctor etc. This provides us with a platform which provides us with data from varied users. A very common method of analyzing the tweets is detecting the tone of the user, which is called sentiment analysis. The tone can be good, bad, angry, sad and many other forms of sentiments. Major sentimental analysis algorithms classify a tweet as bad or good.

The “Big 2” movie rating platform IMDB and Rotten Tomatoes demand their users or members to rate a movie, which over the users have been the go-to platforms for users to check any movie rating, not only because of their accurate rating about their movies but also because of their huge database of movies they have reviewed over the time. This process, however, can take some time as it takes few days after the release for any movie to accumulate reviews in platforms, and reach stable average ratings. In terms of any other platforms in this field, this is even slower since it is an uphill task to attract such a large number users to rate the movies on their platform. However, Twitter has distinct advantage on its volume with magnificent speed, but also contains real-time customer reviews that treasures unfiltered genuine thoughts, which can be closer to the true feeling towards the movie than the reviews in the movie platform.

Hence, we came up with this idea of determining movie rating using only twitter data. This method not only overcomes the difficulty of attracting users to come and rate movies on your platform but you also have a humongous amount of quality data available to you at your disposal as the number of monthly active users on Twitter is almost twice as big to that of IMDB or Rotten Tomatoes.

**II. Related Work**

There have been several studies that analyzed movie reviews using Twitter. Kesharwani, A., & Bharti, M. R. (2017) conducted a sentiment analysis to predict existing movies ratings for Twitter data. [1] This study simply classifies tweets referring to movie titles based on a positive and negative bag of words and calculates rating scores as the number of total positive tweets divided by total negative tweets and positive tweets. [1] The limitation of this study was ignoring possibilities of same users tweeting the movie several times, tweeting without seeing the movie, and the number of neutral tweets while calculating the movie rating scores or etc. [1] Also, it didn’t allow any rational numbers like 3.7 and didn’t consider geographical bias, which movies are rated unreasonably high in specific location, while it’s mediocre movie globally. [1] Dooms, S., De Pessemier, T., & Martens, L. (2013) collects a movie rating dataset based on Twitter which is automatically generated from IMDb. [5] This method retrieves a lot of up-to-date data from unique users and a possibly useful dataset of movie titles with ratings, but by ignoring pure review of the data, it was hard to find the relationship between the tweets and rating scores. [5]

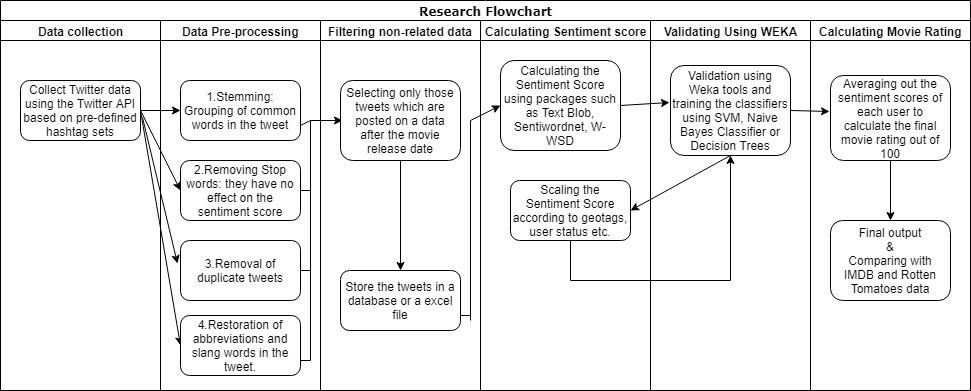
We looked at previous studies of movie other than Twitter data to understand the mechanics of sentiment analysis and other platforms. Chaovalit, P., and Zhou, L. (2005) uses movie reviews retrieved from IMDb to compare supervised and unsupervised methods. [4] It collects positive and negative reviews as their training corpus for N-gram classifiers and Semantic Orientation (SO) approach. [4] Pang et al. (2002) devised sentiment classification using machine learning methods. [7] This paper is a leading paper which was cited by numerous papers about the movie rating system and sentiment classification. [7] It also collected IMDb data for their sentiment corpus and implemented three machine learning methods - Naive Bayes, maximum entropy classification, and Support Vector Machines with bag-of-features. [7] Interestingly, it used n-gram, POS and position feature to enhance the accuracies of three classifiers, but their performances were not better than traditional topic-based categorization. [7]

Lastly, we tried to look at breakthrough research to gain idea of how we can further out our study. Among them, Go, A., Bhayani, R., and Huang, L (2009) conducted a unique study on sentiment analysis with Twitter data. [7] Unlike usual researches, it included emoticons which are often a way to express the users’ mood. [7] This paper also adopted Naive Bayes, Maximum Entropy, and SVM for their machine learning algorithms, and have accuracy over 80%. [7]

**III. Data Collection**

* We plan to use the official Twitter API and the tweepy library in Python to collect twitter data. The tweets that we plan to collect should consist of the hashtags like “Movie Name”, an acronym of that movie, the cast of that movie etc.
* We will also collect the tweets that are re-tweets from user, however; on a condition that the retweeted tweet should not exist in our current database, thus making sure that we collect only unique tweets.
* Limiting the tweets from a single account to a fixed threshold (< 10) so that we avoid taking biased tweets from a fake account or cyborgs.
* The tweets which we get from the Twitter API are in JSON format so we plan to store these tweets in a key-value database system such as MongoDB where the key of that tweet is the primary key of access to all the data possessed by that key. Or we can simply use an excel sheet storing the twitter key in the first column of the excel sheet which will be used as the primary key and the remaining features/data can be stored in the corresponding columns of that key in the excel sheet.

**IV. Methods**



**Figure 1: Research Flowchart**

We designed our approach to differ from the previous studies in following ways:

(1) The system that the authors developed only gave a predefined movie rating to the movie and intermediate ratings like 3,7, 3.8 etc were absolutely ignored which is not a good way to review a movie.

➔ We want to convert this discrete scoring system to a more continuous scoring system.

(2) The authors did not filter out tweets from the same user so the tweets collected were not from unique users.

➔Many accounts on Twitter are programmed to flood twitter with thousands of tweets using a particular set of hashtags. To avoid considering these tweets and creating biases in our final results we will either limit the total number of tweets from a unique account to 3 or 4, or we can average out the result from one particular account so that the biases are reduced.

(3) The authors did not use geotags as people from the same originating company as that of the movie had a better idea about the movie.

➔ In order to make our system to work not only for global movies but also well for regional movies, we want to make use of the geotag feature of Twitter in order to rate the movie. It might be the case that a movie was a hit in a particular region but not elsewhere which speaks a lot about how the movie was a flop globally but a hit only in a particular country. So since the rating of the movie is directly related to the overall success of a movie and not only in a particular region or country, we would like to give a slightly lower score to the tweets of the people whose geotags are in terms with from where the movie originated. This method also overcomes the problem of generating a false score for the ratings because if a movie let’s say originated in China the total population of Chinese people rating the movie highly on Twitter will overpower the global review.

(4) Ignoring tweets that did not actually give a review about the movie but maybe just talked about it or they were just part of a meme game or there are people tweeting without actually watching the movie.

➔ This is one of the most important aspects of computing the ratings of a movie, Using Natural Language Processing (NLP) techniques we will have to find out whether each tweet is actually reviewing about the stated movie or is just talking about say some song, some scene or some clip of the movie. This would also help us avoid movie-generated memes on social media and getting only genuine movie review tweets. Also selecting only those tweets after the movie release date is another solution to filter out non-reviewing tweets.

* Methods that we are going to use for sentimental analysis of the tweets are as follows [2]:
* We get the tweets using the Twitter API and pre-process our data by:

(1) Removing the stop-words such as ‘the’, ‘is’, ‘at’ etc. These words have no effect on the sentiments of a particular tweet.

(2) Stemming of words to enhance accuracy.

* Select the words from the tweet and convert it into an array of tokens.
* Generate features from the array created in the above step. The features can be:

(1) The number of words with positive/negative sentiments.

(2) The number of negations.

(3) Length of a message / tweet.

(4) The number of different parts of speech etc.

* Feature Selection using the following process:

(1) Search procedure: Testing the algorithm by creating a small subset of features and then adding features and removing features to test the algorithm on various combinations of features.

(2) Evaluation procedure: Evaluating a score for the selected subset using statistical methods such as Chi-square, Information Gain etc.

(3) Stopping criteria: Stop the feature selection when:

* Take the subset giving the best results from the search procedure.
* “If the change of feature space does not produce a better subset or if optimal subset was found according to the value of evaluation function” [2].
* We plan to use sentiment analysis libraries like Text Blob, Sentiwordnet, W-WSD [3].
* Validating the labeled sentiments of each tweet using “Validation with Waikato Environment for Knowledge Analysis (Weka)” [3].
* Train the classifier from the above step by using Support Vector Machines (SVM), Naive Bayes Classifier and Decision Trees.
* Select the highest accuracy result of each classifier on each of the sentimental analysis.
* The methods that we are going to use for calculating the movie ratings are as follow:
* Once we get the sentiment score after validating the data we can directly compute our movie rating based on those tweets. For example: say we have ‘n’ sentiment scores of 0.61, 0.50, 0.56…………n for a particular movie then our movie rating would simply be the average of this scores\*100 %.
* Now the other thing that we plan to do is to give extra weight to the sentiment scores of a professional reviewer over that of a rookie. This will be done in the following way:

(1) We first select a number of movies, say 50 with their IMDB scores\*10(to convert it into %).

(2) Collect tweets with their respective scores for each movie from each user.

(3) Now we have a database of 50 movies rated by 1000 unique users if a user is present in more than 30 movies and who’s individual rating is very close to that of IMDB rating in at least 80% of the 30 movies he is present in we categorize him as a ‘super-user’ meaning he is a professional reviewer. But out of those 50 movies if there’s a user whose reviewing only one movie out of 50 we categorize him as a ‘rookie’ and the rest of the users are normal users.

* However, the sentiment scores in the first step above would be affected by external factors such as geo-tags, the category of the user. Say a user ‘X’ has a rated a movie ‘Y’ 0.718(based on the sentiment score) and that user happens to be a ‘super-user’ the score of that user will be multiplied by 1.1 and if it’s a rookie user then the score will be multiplied by 0.98 (Note that the score would be tweaked for that movie only). This would mean that the score of a ‘super-user’ matters more than a rookie user thus giving its score more weight.

**V. Conclusion**

As the technology in managing and computing data has developed, voluminous tweets have become a potential source of recognizing the society, politics, entertainment, and trends etc. Additionally, people use tweets to express their genuine thoughts and feelings, so if implemented appropriately, we will have a better understanding of the world and predict future trends or make better decisions. Our above research will show the plausibility of tweet analysis on movies as the collective review of the movies as we had devised new ways to overcome the limitations that previous movie studies had. We have purposed this study to further direct the future of sentiment analysis, not only in the movies but also encompassing in other various sectors.

**References**

[1] Kesharwani, A., & Bharti, M. R. (2017). *Movie Rating Prediction Based on: Twitter Sentiment Analysis*. LAP LAMBERT Academic Publishing.

[2] Kolchyna, O., Souza, T. T., Treleaven, P., & Aste, T. (2015). Twitter sentiment analysis: Lexicon method, machine learning method and their combination. *arXiv preprint arXiv:1507.00955*.

[3] Hasan, A., Moin, S., Karim, A., & Shamshirband, S. (2018). Machine Learning-Based Sentiment Analysis for Twitter Accounts. *Mathematical and Computational Applications*, *23*(1), 11.

[4] Chaovalit, P., & Zhou, L. (2005, January). Movie review mining: A comparison between supervised and unsupervised classification approaches. In *System Sciences, 2005. HICSS'05. Proceedings of the 38th Annual Hawaii International Conference on* (pp. 112c-112c). IEEE.

[5] Dooms, S., De Pessemier, T., & Martens, L. (2013, October). Movietweetings: a movie rating dataset collected from twitter. In *Workshop on Crowdsourcing and human computation for recommender systems, CrowdRec at RecSys* (Vol. 2013, p. 43).

[6] Go, Alec, Lei Huang, and Richa Bhayani. "Twitter sentiment analysis." *Entropy* 17 (2009): 252.

[7] Pang, B., Lee, L., & Vaithyanathan, S. (2002, July). Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*(pp. 79-86). Association for Computational Linguistics.