Name: Abhilash Mandadi, Josh Alphonse, Chris Glynn
Student ID: 001153596
Course Name/number: Data Mining/ 531
Section number: 0001

Crowdsourcing to Predict Current Popular Movie

Team Members

- Abhilash Mandadi Graduate Team Lead
- Josh Alphonse Undergraduate
- Chris Glynn Graduate

Motivation

Potential Applications

Use of this application will be to predict the current popular movie utilizing crowdsourcing.

Users

Potential users will be general population who would like to know the popular movie to go see. Future uses could include movie critics to use in their work. Movie creators could use the features extracted to find out what makes a movie popular.

Major Components

Data Collection

Use Twitter REST Search API Tweepy[1] in conjunction with OnConnect Data Delivery API[2]. OnConnect Data Delievery API will provide a list of current movie titles for a certain geological area. The list of movie titles will be used as the query to the Twitter REST Search API Tweepy. The form of the query text is 'movieTitle OR movieTitle'. In addition to the query text, the query will contain the same general geological location.

Classification

Create a classification model to remove tweets that are not relevant. Use training data to create the classification model necessary to eliminate collected tweets that are not relevant.

Feature Extraction

Extract features that will be used in predicting popular movie. Possible features to include Sentiment Scores, total number of words, frequency of some terms etc.

Brief Survey of Existing Work

The significance of our study is that the data is collected based on geographical location. Our study will focus on textual twitter data as opposed to other social media platforms. This is an identifiable gap between our study and to those of a similar topic. The Spatial data is not independent. The classification model helps eliminate uncertainty through the use of training data. The data collected can of scientific and commercial value to movie theaters and consumers in a specific area. This study aims to explore the predictive capabilities of data mining.

---

[1] (Tweepy, 2017)
[2] (gracenote A NIELSEN Company, 2017)

Name: Abhilash Mandadi, Josh Alphonse, Chris Glynn
Student ID: 001153596
Course Name/number: Data Mining/ 531
Section number: 0001

References

Aggarwal, C. C. (2015). *Data Mining.* Springer.

gracenote A NIELSEN Company. (2017, February 17). Retrieved from http://developer.tmsapi.com/:
http://developer.tmsapi.com/

Tweepy. (2017, February 17). Retrieved from Tweepy: http://www.tweepy.org