Jeffrey Wang

Web/Mobile Senior Lab

Period 5 Kosek

6 October 2017

Research Proposal

**Using Movie Ratings to Predict Box Office Success**

**Abstract:**

**Problem/Purpose/Engineering Goals:**

The purpose of this project is to determine to what extent can user-based movie ratings predict the box office performance of movies. A web application will be developed to allow users to view retroactive predictions for various movies released in the past four years. The prediction system will utilize machine learning techniques.

**Background: (outline for now)**

* In recent years, machine learning has been rapidly increasing in popularity.
* For supervised learning, there are many different classification algorithms that can be used. Examples are decision trees, support vector machines, and Naive Bayes. (Source 1)
* There are many applications of machine learning, including statistical analysis to provide predictions for business decisions. (Source 2)
* Another application of machine learning is sentiment analysis. In the context of business, sentiment analysis can be effective in measuring brand loyalty and predicting consumer trends. (Source 3)
* In particular, sentiment analysis can be applied to predict the popularity of movies. The box office total is a useful measure of this popularity. (Source 4)
* With the rising popularity of social media, word of mouth on platforms such as Twitter can greatly impact the success of movies. (Source 5)
* User ratings posted online can be used as a reflection of sentiment towards movies.

**Research Techniques/Methods:**

To begin the project, several movie datasets will be downloaded from the Internet, such as MovieTweetings and MovieLens. These datasets will contain various information about a list of movies, with special focus directed toward ratings, budget, and box office numbers. Python will be the main programming language used to extract necessary information from these datasets. Additionally, there are various modules for Python that can be installed to perform machine learning. From a randomly selected sample of movies, a program will be trained to classify each movie as a success, neutral, or flop.

After data analysis is complete, a web application will be developed to provide the results to the user. The user may select a movie title to get relevant information about that movie such as user ratings and box office numbers. The predicted classifier will also be displayed to the user.

**Materials:**

* MovieTweetings dataset
* Other movie datasets (e.g. TMDB 5000, MovieLens)
* Web Development service (e.g. Cloud 9, Heroku)
* IDLE as the Python development environment
* Python machine learning module (unsure which one)
* Personal laptop

**References: (links for now, will add MLA later)**

1. <https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/>
2. <http://www.datasciencecentral.com/profiles/blogs/top-9-machine-learning-applications-in-real-world>
3. <http://blog.paralleldots.com/technology/deep-learning/applications-of-sentiment-analysis-in-business/>
4. <https://s3.amazonaws.com/academia.edu.documents/36689032/vol3.no3.46.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1507252446&Signature=gDkRSFZHa%2Fk%2B4UnluQSvT45pPRM%3D&response-content-disposition=inline%3B%20filename%3DPrediction_of_Movie_Success_using_Sentim.pdf>
5. https://go2cinema.com/stories/the-twitter-effect-8