Liming Luo
Justin Murray
Youchen Ren

CS 780 Machine Learning
Project Proposal
April 30, 2014

The movie industry is booming! Over 1.3 billion movie tickets were sold in 2013, generating over 10 billion dollars of revenue, doubling in the last 10 years alone. With the release of more and more mega-budget blockbuster movies every year, it's hard to make sure a movie gets its fair share of the consumer's attention to maximize its profit. This jam-packed schedule of mega-releases can make it hard for movie studios to stray away from the sure bet and produce something a little more different and out of the norm.

Our objective is to model the movie industry in an attempt to know if a movie will turn a profit before it has started production. By using an artificial neural network, we should be able to project the profit or the loss a movie will turn depending on its inputs that would be known before production starts. From this model we should also be able to determine how important each input is to a movie's profitability. For instance, how important is a movie's budget or the lead actors in the movie to its box office ticket sales? Do franchises outperform non franchised movies if everything else remains same?

We will be using an artificial neural network, to project domestic box office revenue and international box office revenue. For our model, each movie will consists of 2 actors and 1 director; with the two highest actors on the movie's bill being taken into account.

Our model will have the following inputs:

- Profitability Per Movie *(normalized)* - one for each person associated with the movie (*see explanation below*)
- Creative Type – the creative type of the movie; as defined by OpusData (*can loosely be thought of as defining the "universe" in which the work's story is set)*
- Budget – budget for the proposed movie
- Franchise – is the movie part of a franchise?

Profitability would be calculated as follows for each member of the movie:

$$\left(\frac{1}{N}\right)\sum_{i=1}^{N} \left(\text{total box office revenue}_i - \text{budget}_i\right)$$

Where *N* is the number of movies the person appears in the credits.

For training and testing data, we will be pulling information from a few sources. Our major source of data will be from OpusData.com. OpusData is a database that contains financial data for over 18,000 movies.

Our estimated timeline would be as follows:

- May 12, 2014 – finished collecting data in correct form to be consumed
- May 13-16, 2014 – train and test model, run interesting cases on trained model
- May 14-18, 2014 – prepare and practice presentation

Our team consists of the following:

- Liming Luo
- Justin Murray
- Youchen Ren

All work will be divided equally amongst the team. Since each movie can contain up to three people, each team member will be responsible for getting social reputation and award data for one person for each movie. This will be the most time consuming, thus the most time has been given to this portion of the project. Since we have all performed the homework on neural networks, we all should have one neural network. This will allow us to discuss which neural network to work with based on correct implementation and configurability via code. Then while the model is being trained and tested, work can begin on the presentation.