

Prediction of Box Office Movie Profitability Using Data Analytics

By Kasra Zarei

kasra-zarei@uiowa.edu

Two UI researchers have developed an algorithm to predict the success of box office movies, in terms of profitability.

The U.S. film industry is worth billions of dollars, with hundreds of Hollywood movies made every year. The secrets to making profitable movies have evaded producers and investors for decades.

The research of Kang Zhao, University of Iowa assistant professor of Management Sciences, and Michael Lash, University of Iowa graduate student in Computer Science and lead author on the project, could present a solution to this problem that film studios, financial analysts, and other academic teams have unsuccessfully tried to answer for years.

Zhao's research involves data analytics for social and business networks, with his most recent work in collaboration with Lash focusing on predictive modeling of box office movie success.

"We looked at factors that drive the success of movies from a targeted perspective, being profits," Lash said.

By defining novel features of movies and appropriately leveraging them, Zhao and Lash have been able to make pre-production predictions of profitability with over ninety percent accuracy on historical data.

"One of the approaches we took was an interpersonal, connected look at people that are involved in making movies," Lash said. "That is, historically, what level of profit have people seen in their past collaborations."

This research question has been a large undertaking for Zhao and Lash, who have been working on the project together for almost three years.

"The first steps of the project were to acquire historical data," Lash said. Using extracted data from various sites including IMDB and Box Office Mojo, Zhao and Lash iteratively defined the features or attributes of movie success with a profit-centered focus.

Zhao and Lash ran a set of predictive algorithms that use features of movies that influence profitability like big name actors, movie genres, plot synopses and release locations to identify patterns in past data.

“We wanted to find the best predictive model from a global set of classifiers that exists,” Lash said.

Among all the various approaches tested, Zhao and Lash found that their top performing classifier consisted of a “Random Forrest Classifier”, which incorporates a sense of randomness, feature selection, and voting in the profitability decision-making process.

“[Predicting movie success] is a historical problem. People have been looking at this problem for years, getting data from the Internet and running computations on machines,” Lash said.

However, what differentiates Zhao and Lash’s research is the way they define movie success.

“Often times people define success in terms of box office gross, but this is in no way indicative of box office profitability,” Zhao said.

But their recently conducted study has found a loose correlation of gross revenue with profit in the film industry.

“While big-name stars may drive the box office revenue generated, the price tag associated with these stars may not equivocate to profitable returns,” Lash said.

Furthermore, compared to previous approaches that assesses movie success using post-production information, Zhao and Lash use a more realistic approach.

“We define our features, or variables, carefully such that we are not using any information that is available right before or after release. We only use features that are available during the investment stage of a movie so that we can actually help investors make their decision prior to production,” Lash said.

Lash hopes to take a more targeted look at how they can help the different roles of people involved in making a film.

“For instance, if you are a writer, we want to know what could we suggest to you to make your script better,” Lash said.

Speaking about Zhao and Lash’s work, Nick Street, University of Iowa DEO of the Management Sciences Department, had positive words to share about the two researchers.

“[Zhao and Lash’s work] is a great example of the highly interdisciplinary work going on in the Management Sciences Department - applying methods from computer and information science to problems in business, social sciences, biology, and health care,” Street said.