Visual Analytics for the Prediction of Movie Rating and Box Office Performance

Mennatallah el Assady, Daniel Hafner, Michael Hund, Alexander Jäger, Wolfgang Jentner, Christian Rohrdantz, Fabian Fischer, Svenja Simon, Tobias Schreck, Daniel A. Keim *

University of Konstanz, Germany

1 Introduction

This paper describes our solution to the IEEE VAST 2013 Mini Challenge 1[1]. The task of the challenge was to create a visual and interactive tool to predict the popularity of new movies in terms of viewer ratings and ticket sales for the opening weekend in the U.S. The data usage was restricted by the challenge organizers to data from the Internet Movie Database (IMDb)[2] and a predefined set of Twitter[3] microblog messages. To tackle the challenge we designed a system together with an analysis workflow, combining machine learning and visualization paradigms in order to obtain accurate predictions. In Section 2 we describe the machine learning components used within the analysis workflow. Next, in Section 3, we describe where and how the human analyst is enabled to enhance the prediction with her/his world knowledge. Finally, Section 4 concludes the paper providing an evaluation of the prediction accuracy with and without human intervention.

2 Machine Learning

In order to predict the performance of ratings and box office takings for upcoming movies, it makes sense to rely on data from past movies. For example, if movies from certain directors or with certain actors have been successful in the past, it is to be expected that also their future movies will be successful. It has been shown that machine learning models, like neural networks, can support the prediction in this way [1]. We experimented with different models for predicting the movie viewer rating, which we trained and tested based on IMDb data from past movies. As input we took into account movies with same cast and crew members or genres. In a 10-fold cross validation test, neural network predictors performed best. For the prediction of box office takings, we applied multinomial regressions of different orders. The input parameters to the regressions were the movie budget and runtime, as those were the only reliable numerical variables available across all movies. Another line of research has shown that social media messages, such as Tweets, can potentially also be exploited in a beneficial way when predicting movie performance [4]. Yet, when predicting the performance of upcoming movies as requested by the VAST 2013 Mini Challenge 1 certain limiting and biasing factors have to be taken into account. First, the challenge demands the prediction of the rating a movie achieves on the opening weekend only. In contrast to that, the IMDb data contains ratings that have established over a longer period of time. Thus, the characteristics of the training data do not match those of the data that is to be predicted. Another related issue is that box office takings from the distant past are probably less meaningful for the prediction due to economic factors such as inflation and developments in ticket pricing and purchasing power. Second, the challenge organizers restricted the sample of Tweets that were allowed to be used to a predefined set. We found out that this set was hardly representative and sometimes also contained many Tweets not related to the corresponding movie. Past research suggests that there are correlations between numerical characteristics that could be derived from the Tweets, such as number of Tweets or Tweet sentiments [4], and the rating performance. However, such correlations did not appear within our restricted Tweet sample and the set of movies to be predicted. Thus, the Tweets were not useful for generating an automatic prediction. Third, the automatic methods lack the integration of world knowledge. Especially in the prediction of box office takings external factors not contained in the data may have a strong impact. For example, the number of cinemas in which a movie is shown, the coincidence of holidays, the weather on the opening Weekend, whether a movie is shown 2D or 3D or both, and which other movies are released on the same weekend running in competition for spectators. For the prediction of viewer ratings important external factors are, for example, whether the movie is based on a book, whether it is a sequel, and the publicity for the movie spread in news or through web channels.

In order to account for these biases it is required to integrate the human into the analysis loop. Still, an automatic prediction provides an indication for what a realistic value or range of values for the final prediction could be. In the next section we will detail on how the human analyst can enhance the automatic prediction.

3 Interactive Adjustment

As mentioned, mere automatic methods fall short of incorporating all possibly availabe factors influencing the prediction. The human analyst has to be integrated into the analysis workflow in order to contribute world knowledge and interpretations of social media content. Our solution is twofold: First, the analyst interactively decides on what s/he considers to be the most useful input for the machine learning. We name this pre-learning interaction phase. Second, after the machine learning process has finished the analyst is enabled to tune the results. We name this post-learning interaction phase. For both phases we offer different interactive visual displays. For the sake of brevity, only some fundamental steps will be described in the following paragraphs. Further details are given in our challenge submission[4] and video[5], which are both available online.

Pre-learning interaction phase First, we provide a graph-based visualization that reveals details on the social media content relating to a certain movie. The graph-structure shows co-occurrences of different persons, concepts, and attributes and also reflects sentiments. The graph visualization is created using VISONE[6]. The structure of the graph is generated as follows: *Nodes* represent different types of keywords (names of actors, adjectives, verbs, nouns, and #hashtags). Each of these types is mapped to

*firstname.lastname@uni-konstanz.de

[1]http://boxofficevast.org/
[2]http://www.imdb.com/
[3]https://twitter.com/
[4]http://bib.dbvis.de/uploadedFiles/MooVisSummaryFinal.pdf
[5]youtu.be/XhJDPa9FNck
[6]http://visone.info/html/about.html

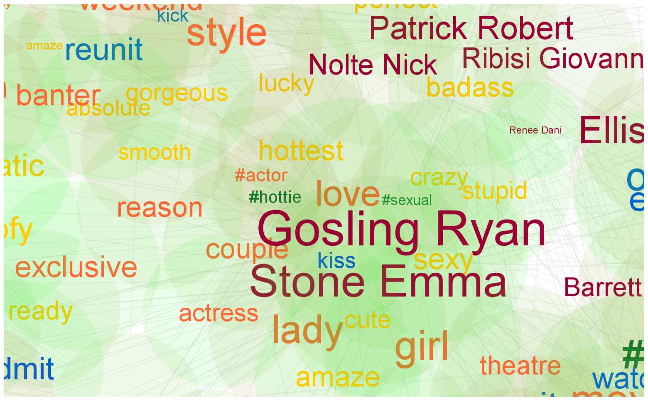
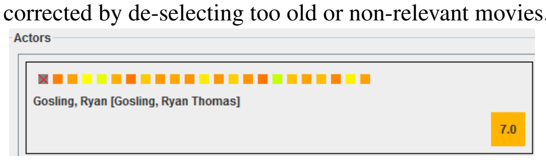


Figure 1: Close-up view on the central part of the graph reflecting co-occurrences in Tweets on the movie Gangster Squad.

a different *color*. The *node size* indicates in how many Tweets a keyword occurs. The *edges* represent the association between the actors and their top related keywords, calculated by the likelihood-ratio of both words, as in [3]. This ratio is used to determine the *position* of the nodes in a stress-minimizing layout. As *overlay* (to the node color), the average sentiment value of all tweets containing the corresponding keyword is interpolated between green (positive) and red (negative). Optionally, a Girvan Newman *clustering* [2] can be calculated over the graph to allow a focused exploration of the important entities and all their related keywords.

Figure 1 provides an example for the movie *Gangster Squad* [7]. Because of the dominating green color of the node overlay, it is easy to see, that most of the Tweets related to this movies have a positive sentiment, which will be an important factor when weighting the scores at the end. The most important cast/crew members are visible through their node size. In this example graph, the relation of *Emma Stone* and *Ryan Gosling* is very noticeable due to the positions of the nodes and their surrounding keywords, which include words that indicate that they are well seen as a couple. The keywords *crazy*, *stupid* and *love* indicate, that the users on Twitter have been comparing the couples performance in this movie to their act in the movie *Crazy, Stupid, Love* [8].

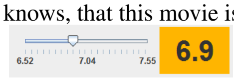
The analyst can use the insight gained from the social media content in order to complement her/his preconception on the movie, identify the prominence of different actors, related movies, and identify prevailing sentiments. Based on this, the user is enabled to select which attributes shall be used as an input for the neural networks to perform the prediction. The detail views allow the user to choose the order of related movies or cast/crew members and de-select non-relevant entities. The score of each cast/crew member, which is the average score of all movies s/he contributed in, can be corrected by de-selecting too old or non-relevant movies.



In the example displayed on top the scores for all previous movies of Ryan Gosling are color-coded. The first movie has been manually deselected by the user. Only selected movies will be considered in the prediction.

Post-learning interaction phase In the machine learning step different complementary predictors are used. In order to predict the user rating several neural networks are used, which have been trained on different numbers of top-ranked cast/crew members, for example, to perform predictions for each genre separately or all combined. In order to predict the box office takings, regressions of different orders are used.

All calculated values of the models have to be combined to arrive at one single prediction value for both the IMDb rating and the box office takings. To integrate his/her knowledge, the analyst may assign different weights to each single score. For example, if a movie is placed in both gernes action and comedy, the user may weight the model trained on the movies with the genre action higher, if s/he knows, that this movie is rather considered to be an action movie.



The final score by default is set to the mean value of all predictions, with the variance indicating the uncertainty range. The user is enabled to change the final prediction within this range to account for the overall popularity of the movie in the social media for example.

[7] http://www.imdb.com/title/tt1321870/

[8] http://www.imdb.com/title/tt1570728/

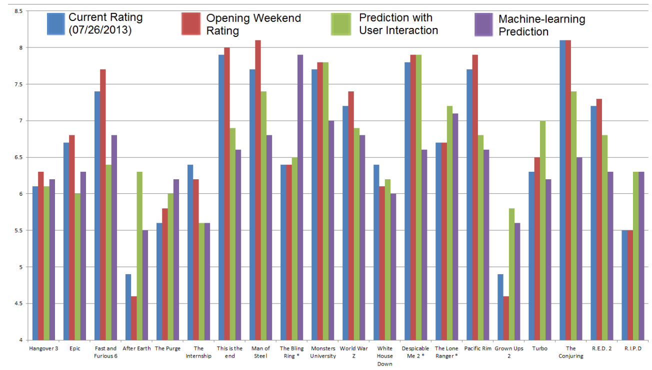


Figure 2: The evaluation shows the prediction results for the VAST Challenge both with respect to short-term and long-term ratings

4 Evaluation

Figure 2 shows the machine-learning and visual analytics predictions of viewer ratings in comparison to the real viewer ratings available after the opening weekend and later. It becomes evident that rather extreme ratings tend to even out on the long run. We assume that the sample of opening weekend raters is not representative for the set of all raters. For the long-term ratings of the 20 movies, that we predicted during the VAST Challenge, the mean squared error was 0.4575. For the short-term ratings the mean squared error was 0.608. Without user interaction this mean squared error was 0.7385. It can be concluded that the user interaction leads to a considerable improvement, especially in the case of animation movies, where the actors only lend their voices and therefore tend to have less influence on the movie success. That the approach has a higher accuracy when predicting long-term ratings than short-term ratings is not quite surprising, as the models have been trained on long-term ratings. In the future, solutions should be investigated that account for this bias.

References

[1] A. Augustine and M. Pathak. User rating prediction for movies. Technical report, University of Texas at Austin, 2008.

[2] M. Girvan and M. E. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826, 2002.

[3] M. C. Hao, C. Rohrdantz, H. Janetzko, D. A. Keim, U. Dayal, L.-E. Haug, M. Hsu, and F. Stoffel. Visual sentiment analysis of customer feedback streams using geo-temporal term associations. *Information Visualization*, 2013.

[4] A. Oghina, M. Breuss, M. Tsagkias, and M. de Rijke. Predicting imdb movie ratings using social media. In *ECIR*, pages 503–507, 2012.