

# A Regression Approach to Movie Rating Prediction using Multimedia Content and Metadata

Hossein A. Rahmani  
University of Zanjan  
srahmani@znu.ac.ir

Yashar Deldjoo  
Politecnico di Bari  
deldjooy@acm.org

Markus Schedl  
Johannes Kepler University  
markus.schedl@jku.at

## ABSTRACT

This paper presents the submission of the team MASlab-ZNU to the MMRRecSys movie recommendation task, as part of MediaEval 2019. The task involved predicting average movie ratings, standard deviation of ratings, and the number of ratings by using audio and visual features extracted from trailers and the associated metadata. In the proposed work, we model the rating prediction problem as a regression problem and employ different learning models for the prediction task, including ridge regression (RR), support vector regression (SVR), shallow neural network (SNN) and deep neural network (DNN). The results of fairly large amount of experiments on various models and features indicate that combination of DNN+tag features produce the best results for prediction of avgRating and StdRating while for numRating (popularity) it is the combination of RR+tag that significantly outperforms the other competitors, with a large margin.

## 1 INTRODUCTION AND RELATED WORK

The global revenue obtained from the media and entertainment (M&E) market in 2017 was approximately 2 trillion US dollars. The four main verticals of the industry in the US include: film (40%), music (6%), book publishing (12%) and video games (8%) [1, 2]. The movie industry not only has a large cultural and sociological impact on people but also occupies an important part of the business market in the M&E industry. Producing a new movie means that the company is betting on this movie's success. Being able to predict into the future whether a movie will be successful or not is therefore crucial and requires machine learning techniques. The results of the predictions will be used by producers and investors to decide whether or not to adopt the production of similar movies.

The paper at hand describes the solution by the team MASlab-ZNU for the 2019 Multimedia for Recommender System Task: MovieREC and NewsREEL at MediaEval [6], **the movie recommendation subtask**. The task involved using of audio, visual features extracted from trailers of the corresponding movies coming from MMTF-14K dataset [4] and metadata to predict three scores: avgRating (representing user appreciation and dis-appreciation of the content), stdRating (characterizing agreement and disagreement between user opinions/ratings), and numRating (reflecting popularity of movies). The audio and visual features have been used to solve various tasks in movie recommendation, see e.g., [3, 5].

In the context of movie popularity prediction, Szabo et al. in [10] predict the future popularity of a video based on the number of previous views on YouTube using predictive models based on linear regression. Pinto et al. in [9] extend the work by Szabo et al. by proposing a multivariate linear model and sampling the number of

views at regular time slots. Recently, Moghaddam et al. [8] address the problem of movie popularity prediction using visual features of movie trailers.

## 2 EXPERIMENTS

This section presents the experiments carried out towards addressing the prediction task. For performance evaluation, we randomly split the original train set into 2 non-overlapping subsets, where we consider 80% train data and 20% validation data. We refer to these datasets as *trainSet* and *validSet* throughout the paper. In a preprocessing step, we normalize all features in the dataset, using min-max normalization. We perform a hyper parameter search and report all the results under the best setting. We model the task in question as regression and use the following regression models:

- **Linear model using Ridge Regression (RR):** We use linear regression to serve as a simple yet standard approach to model the relation between dependent (prediction scores) and independent variables (features) in a linear fashion [7]. The model is given by  $y = \theta^T x$  where  $x$  and  $y$  represent feature vectors and scores, respectively,  $\theta^T$  contains the linear model coefficients estimated by RR  $\min_{\theta^T} \frac{1}{2} \|y - \theta^T x\|_2^2 + \lambda \|\theta\|_2^2$  where  $\ell_2$  regularization is applied to avoid overfitting of the coefficients.
- **Support Vector Regression (SVR):** SVR is the regression type of Support Vector Machine (SVM). SVR makes use of a nonlinear transformation function to map the input features to a high-dimensional features space given by  $y = \sum_{i=1}^n (a_i - a_i^*) \cdot K(x_i, x) + b$  where  $K = \exp(-\|x - x'\|^2)$  is the Radial Basis Function (RBF) kernel,  $b$  is intercept, and  $a$  and  $a^*$  are Lagrange multipliers.
- **Shallow Neural Network (SNN):** To model the relationship between features in a non-linear way, we also apply a neural network to predict the scores. Here, we consider a simple (shallow) neural network model. SNN has a hidden layer with 24 neurons and Rectified Linear Unit (*ReLU*) as activation function. As for the output layer we use the *Sigmoid* activation function.
- **Deep Neural Network (DNN):** This method is similar to the SNN but we have more hidden layers to consider deeper relations between features. In our deep model, we have 3 hidden layers; the first layer has 128 neurons with *ReLU* as activation function; the second layer has 64 neurons and uses a *Sigmoid* activation function; the third one has 32 neurons and uses *ReLU*.

## 3 RESULTS AND ANALYSIS

The regression results using the proposed approaches are presented in Table 1. Regarding the comparison of learning models, we can see

**Table 1: Results of regression in terms of RMSE, calculated on the validation set. The first and second best result for each score are shown in bold and *italic* respectively.**

Score	Regressor	Metadata		Audio		Visual		Average
		Genre	Tag	BLF	I-Vec	AVF	Deep	
<i>avgRating</i>	RR	0.1401	0.1440	0.1495	0.1482	0.1499	0.1520	0.1472
	SVR	0.1466	0.1462	0.1518	0.1588	0.1523	0.1523	0.1513
	SNN	<i>0.1396</i>	0.1457	0.1492	0.1487	0.1496	0.1483	0.1468
	DNN	0.1409	<b>0.1370</b>	0.1491	0.1509	0.1501	0.1509	0.1464
<i>stdRating</i>	RR	0.1410	0.1445	0.1438	0.1421	0.1430	0.1472	0.1436
	SVR	0.1487	<i>0.1396</i>	0.1476	0.1464	0.1495	0.1495	0.1468
	SNN	0.1411	0.1433	0.1429	0.1420	0.1435	0.1423	0.1425
	DNN	0.1411	<b>0.1356</b>	0.1429	0.1429	0.1428	0.1429	0.1413
<i>numRating</i>	RR	0.0515	<b>0.0232</b>	0.0518	0.0523	0.0524	0.0528	0.0473
	SVR	0.1892	0.1142	0.1752	0.1656	0.1954	0.1952	0.1461
	SNN	0.0533	<i>0.0272</i>	0.0524	0.0526	0.0525	0.0523	0.0483
	DNN	0.0529	0.0529	0.0526	0.0527	0.0527	0.0528	0.0527

that DNN and RR are the best predictors, in most cases generating the best performance for each feature while SVR is the worst. The final submitted runs are selected based on the ones performing the best on the *validSet*, which are highlighted in **bold** in Table 1.

**Predicting average ratings:** As can be seen in Table 1, the performance of all audio and visual features, are closely similar to each other. These results with a close margin look similar to the performance of the genre descriptor, e.g. 0.1487 v.s. 0.1466. However, it can be noted that tag features is the best feature to predict the average ratings. These results show that user-generated tags assigned to movies contain semantic information that are well correlated with ratings given to movies by user. One can also observe that the DNN model is the best model to predict the average ratings, though it is only marginally better than the SNN model.

**Predicting standard deviation of ratings:** The results of predicting standard deviations of ratings show the audio-visual features and genre metadata have very similar results. Again we can see the best feature to predict the standard deviation of ratings is tag metadata using the DNN learning model. Average results indicate that the DNN model is the best learning model to predict the standard deviation of ratings.

**Predicting number of ratings:** As for predicting number of ratings, it can be seen that except for the tag feature with the best performance, the rest of audio-visual features and genre metadata yield very similar results. For the tag feature, we observe a substantial superiority in performance compared to all other features; it reduces the RMSE by about 55% (0.0232 v.s. 0.0515 for tag v.s. genre), compared to the second best feature (genre). More interestingly, we can see for predicting number of ratings, the best model is the simple RR learning model followed by SNN but not DNN.

## 4 CONCLUSION

This paper reports the method used by team MASlab-ZNU for the Multimedia for Recommender Systems task at MediaEval 2019 [6]. Results of experiments using different regression approaches are promising and show the efficacy of audio and visual content in comparison with genre metadata but overall it is the tag feature that provides the best prediction quality in all experimental cases.

## REFERENCES

- [1] 2018. 2017 Top Markets Report Media and Entertainment. <https://www.trade.gov/topmarkets/pdf/Top%20Markets%20Media%20and%20Entertainment%202017.pdf>. (2018). Accessed: 2018-12-27.
- [2] 2018. Media and Entertainment Industry Overview. <https://investmentbank.com/media-and-entertainment-industry-overview/>. (2018). Accessed: 2018-12-27.
- [3] Yashar Deldjoo, Mihai Gabriel Constantin, Hamid Eghbal-Zadeh, Bogdan Ionescu, Markus Schedl, and Paolo Cremonesi. 2018. Audio-visual encoding of multimedia content for enhancing movie recommendations. In *Proc. of the 12th ACM Conference on Recommender Systems, RecSys 2018, Vancouver, BC, Canada, October 2-7, 2018*. 455–459.
- [4] Yashar Deldjoo, Mihai Gabriel Constantin, Bogdan Ionescu, Markus Schedl, and Paolo Cremonesi. 2018. MMTF-14K: a multifaceted movie trailer feature dataset for recommendation and retrieval. In *Proceedings of the 9th ACM Multimedia Systems Conference, MMSys 2018, Amsterdam, The Netherlands, June 12-15, 2018*. 450–455.
- [5] Yashar Deldjoo, Maurizio Ferrari Dacrema, Mihai Gabriel Constantin, Hamid Eghbal-zadeh, Stefano Cereda, Markus Schedl, Bogdan Ionescu, and Paolo Cremonesi. 2019. Movie genome: alleviating new item cold start in movie recommendation. *User Model. User-Adapt. Interact.* 29, 2 (2019), 291–343.
- [6] Yashar Deldjoo, Benjamin Kille, Markus Schedl, Andreas Lommatzsch, and Jialie Shen. 2019. The 2019 Multimedia for Recommender System Task: MovieREC and NewsREEL at MediaEval. In *Working Notes Proceedings of the MediaEval 2019 Workshop, Sophia Antipolis, France, 27-29 October 2019*.
- [7] Jinna Lv, Wu Liu, Meng Zhang, He Gong, Bin Wu, and Huadong Ma. 2017. Multi-feature fusion for predicting social media popularity. In *Proceedings of the 25th ACM international conference on Multimedia*. ACM, 1883–1888.
- [8] Farshad B Moghaddam, Mehdi Elahi, Reza Hosseini, Christoph Trattner, and Marko Tkalcić. 2019. Predicting Movie Popularity and Ratings with Visual Features. In *14th International Workshop On Semantic And Social Media Adaptation And Personalization*.
- [9] Henrique Pinto, Jussara M Almeida, and Marcos A Gonçalves. 2013. Using early view patterns to predict the popularity of youtube videos. In *Proceedings of the sixth ACM international conference on Web search and data mining*. ACM, 365–374.
- [10] Gabor Szabo and Bernardo A Huberman. 2010. Predicting the Popularity of Online Content. *Commun. ACM* 53, 8 (2010), 80–88.