Anime Rating Prediction Using Tag systems and Neural Networks

Jiaxin Huang
Department of Computer Science
The University of Hong Kong
jiaxin.huang@connect.hku.hk

Che Shen
Department of Computer Science
The University of Hong Kong
shenche@connect.hku.hk

Abstract

This paper explores an approach to predicting the ratings and popularity of anime using a sophisticated tag system. By leveraging the descriptive power of tags associated with each anime, our model offers a unique capability to forecast both the qualitative and quantitative success metrics of anime, notably its ratings and vote numbers, prior to its release. This predictive model is rooted in a detailed analysis of a comprehensive dataset gathered from a popular anime review platform. Utilizing advanced neural network techniques, we demonstrate the effectiveness of this model in accurately estimating the potential success of anime based on pre-release attributes. The integration of such a model into anime-related websites can provide significant benefits, offering real-time, data-driven insights to users and creators alike. The outcomes of this study not only showcase the practical applications of our model but also open avenues for future research in predictive analytics within the anime industry, emphasizing the potential of tag-based predictive models in enhancing user experience and guiding production decisions.

1 Introduction

Movies and cinemas have long been regarded as not only sources of entertainment and recreation but also as domains of art and sociological exploration. In the late 20th century, the advent of data analysis brought a fresh perspective to the study of the film industry. The widespread use of the internet, in particular, has led to the generation of vast amounts of user-generated content (UGC) related to movies, including ratings, reviews, and critiques. Prominent UGC databases such as *IMDB (Internet Movie Database)*[1], *Rotten Tomatoes*[2], and *Bangumi*[3] serve as references for potential movie viewers in deciding which movies to watch or whether to watch a specific film. Movie producers also turn to these databases to gain insights that can help them create films that attract a larger audience.

One significant application of movie data analysis is the prediction of viewer ratings, which represent a 10-point scale reflecting the quality of a movie. Different viewers may assign different ratings to the same movie based on their individual tastes and preferences. Additionally, factors such as the movie's directors, starring actors, and genres also impact its rating. Considerable research has been dedicated to predicting user ratings by leveraging encoded preference information of the user, as well as ratings provided by other users for the same movie [1, 2]. Some models even incorporate text analysis tools to include review articles in the analysis [3, 4]. These models serve as valuable tools for providing personalized movie recommendations to individual users.

However, from the perspective of filmmakers, such models offer limited utility. Filmmakers are interested in understanding what makes a movie successful. Specifically, they want to predict whether

[1] https://www.imdb.com
[2] https://www.rottentomatoes.com
[3] https://bangumi.tv

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

the films they are producing will perform well at the box office or how they can modify their productions to increase profitability. This objective reveals two main challenges. First, when a movie is still in production, user review information is unavailable. Second, rating scores may not always be perfectly correlated with box office performance. In other words, highly rated movies do not always achieve significant commercial success. Consequently, there is a need to develop new models that directly predict box office performance.

In this project, we focus on addressing these two challenges. Our approach involves excluding viewer review information from the model input and relying solely on attributes related to the movie itself. To enrich and diversify the set of movie attributes, we adopt a "tag" system that incorporates keywords associated with each movie. Obtaining accurate box office data can be challenging; therefore, we substitute it with the number of user ratings, assuming a linear relationship between the number of viewers who visit a cinema and those who leave ratings on a review platform.

The remainder of this paper is organized as follows. Section 2 examines the existing literature on movie rating predictions, highlighting their technical strengths and weaknesses. Section 3 discusses the dataset structure, and Section 4 discusses the model architecture employed in this project. Section 5 presents the experimental results and provides analysis. Finally, in Section 6, we conclude by discussing future prospects and potential extensions of this work.

2 Related Work

Considerable research has been conducted on predicting movie ratings using viewer reviews. Two major approaches have emerged using traditional machine learning: content-based methods and collaborative methods [1]. Content-based methods focus on utilizing information about the user and the movie itself for prediction, while collaborative methods incorporate ratings from similar users on similar movies. A notable example of content-based methods is the decision tree model developed by Li and Yamada, where separate trees are constructed for each user, and tree labels represent movie features [2]. In [1], Marovic et al. provide a comprehensive comparison of various models within these two approaches, including k-nearest neighbors [5], personality diagnosis [6], among others.

With the advent of neural networks, the distinction between content-based and collaborative methods has become less clear. In neural networks, all available data is fed into the model during training, eliminating the explicit separation of model parameters representing specific movie or user information. Moreover, advancements in text and sentiment analysis have given rise to analytical tools that leverage Google Trends [7] and public opinions on social media [8, 9, 4].

While these techniques rely on information that is only available post-movie release, there is a need to consider objective information about the movie itself to aid decision-making for movie studios prior to release. Some studies have explored the use of data such as studio name, director name, genre, or country [10–12]. However, the limited categories of such data restrict their applicability to refining ongoing movie projects. For instance, production teams cannot change studio names, directors, or genres to improve box office performance. Therefore, we emphasize the importance of the "tag" system, which provides a detailed and nuanced description of movies, allowing production teams the flexibility to fine-tune and enhance their movies.

3 Dataset

In this section, we provide a detailed overview of our dataset, which incorporates a comprehensive tag system to capture the intricate characteristics of movies. The tag system encompasses various attributes typically associated with movies, such as genre, director names, and actor names. But it also includes any keywords commonly associated with movies, such as "stream", "cinema", "adopted from novels," "love", "back in time," and many more. Such systems can quantitatively reflect the diverse aspects and themes of movies, serving as an accurate account of the emotional perceptions of viewers.

The dataset is structured as a table consisting of 11,187 rows and 407 columns. Each row corresponds to a distinct movie entry, while each column represent a specific attribute of the movies. The first four columns contain basic information such as the movie ID, rating, the number of user ratings, and the publication year. The remaining 403 columns represent 403 most frequently used tags on the selected

Table 1: A snapshot of the dataset.

ID	rating	# of votes	pub year	stream	cinema	...	love	back in time	...
28389	7.2	630	2000	5	89	...	32	0	...
208576	5.3	146	2017	23	29	...	16	21	...
31480	7.6	4383	2022	679	3	...	1	0	...

movie review platform. The values in these columns indicate the number of users who assigned a particular tag to a given movie. We present a snapshot of a sample portion in Table 1, showcasing a subset of movie entries and their corresponding attributes. The columns starting from the fourth column are . The "stream" and "cinema" tags indicate the audience's preference for watching these movies either through streaming services or in cinemas; The "love" and "back in time" tags provide insights into the extent to which these movies incorporate themes related to romance and time travel, respectively.

To create this dataset, we utilized a web crawler to retrieve anime information from Bangumi, a popular anime review platform where users can assign tags to animes. It's important to note that although these tags are assigned by users after watching the animes, they solely describe information about the animes themselves. In theory, anime studios could compile a comprehensive list of appropriate tags for their animes and utilize our model to infer box office information based on these tags.

4 Methodology

4.1 Phase I: Integrated Model Approach

In the first phase of our experiment, we focused on developing an integrated model using a neural network. The methodology was structured as follows:

1. **Data Preparation and Normalization:** We began by importing our dataset using the pandas library. To normalize the data, the 'score' was divided by 10, and 'votes' were divided by their maximum value. Additionally, we applied Min-Max scaling to the 'year' feature and all features from the 5th column onwards.
2. **Dataset Splitting:** The dataset was then split into features (X) and targets ('score' and 'votes'). Using scikit-learn's `train_test_split` function, we divided the data into training (80%), validation (10%), and testing (10%) sets.
3. **Neural Network Architecture:** Our neural network, designed using PyTorch, comprised three fully connected layers with 256, 128, and 64 neurons. We employed ReLU activation for the hidden layers and a Sigmoid activation for the output layer. The training parameters were set to 200 epochs and a batch size of 128, using an Adam optimizer with a learning rate of 1e-5 and a weight decay (L2 regularization) of 1e-5.
4. **Data Conversion and Batch Processing:** The DataFrame was converted to NumPy arrays, followed by transformation into PyTorch Tensors. We then created DataLoader objects for batch processing of 'scores' and 'votes'.
5. **Training Loop with Early Stopping:** A training loop was implemented, incorporating early stopping based on validation loss to train the model for predicting 'scores' and 'votes' simultaneously.
6. **Model Evaluation:** Finally, the model was evaluated on the test set to obtain losses for 'scores' and 'votes'.

4.2 Phase II: Separated Model Approach

In the second phase, we adopted a similar methodology as in Project, but instead, we developed separate models for 'scores' and 'votes'. Each model was trained using the same architecture and training parameters as the integrated model. The evaluation for each model was then conducted independently on their respective test sets.

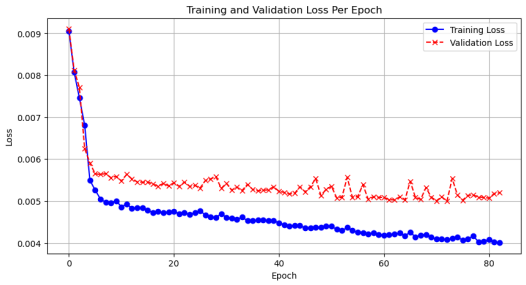
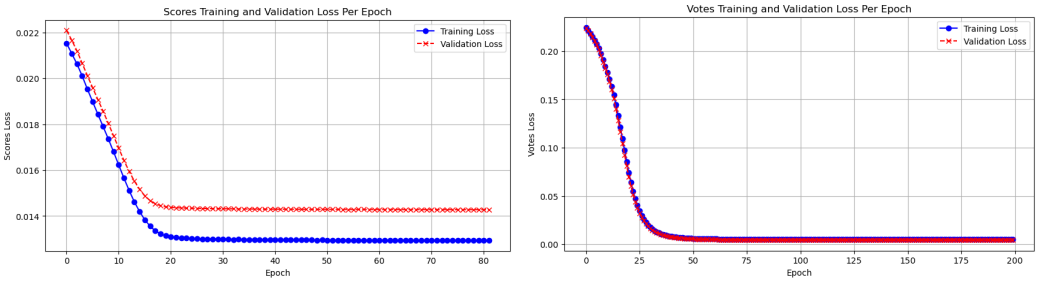


Figure 1: Loss Over Epochs for the Integrated Model



(a) Scores Loss

(b) Votes Loss

Figure 2: Loss Over Epochs for the Separated Model

5 Experiment

5.1 Evaluation of Models

The performance evaluation of our models was meticulously conducted, yielding insightful results:

Integrated Model: The Integrated Neural Network model, designed to predict both 'scores' and 'votes', achieved a notable test loss of 0.0048. This result signifies the model's strong predictive capability and its effectiveness in handling the dual task of predicting scores and votes simultaneously (Figure 1).

Separated Model: In contrast, the Separated Model approach, where 'scores' and 'votes' were predicted using two distinct models, resulted in a test loss of 0.0129 for scores and 0.0056 for votes. This outcome demonstrates the advantage of treating scores and votes as separate entities, enhancing the precision of our predictions (Figure 2a, 2b).

5.2 Comparison with Baseline Models

A comparison with baseline models was crucial to contextualize the performance of our neural networks:

Random Guess Model: Both the Integrated and Separated Models significantly outperformed the Random Guess Model. This comparison underscores the superiority of our neural network models over a naive prediction approach (Table 2, 3).

Linear Model: Interestingly, the Linear Model displayed competitive results, with a test loss of 0.0055 for the Integrated Model and matching test losses of 0.0129 for scores and 0.0056 for votes in the Separated Model. This similarity in performance between the neural networks and the linear model, particularly in the Separated Model, suggests that for certain aspects of our dataset, a simpler linear approach could be nearly as effective as a more complex neural network (Table 2, 3).

5.3 Observations

From these experiments, several key observations were made:

Table 2: Comparison of Test Loss of Integrated Model

Model	Test Loss
Integrated Neural Network	0.0048
Random Guess Model	0.3109
Linear Model	0.0055

Table 3: Comparison of Test Loss of Separated Model

Model	Test Loss for Scores	Test Loss for Votes
Separate Model	0.0129	0.0056
Random Guess Model	0.1078	0.3109
Linear Model	0.0129	0.0056

- The integrated model was effective in predicting both 'scores' and 'votes', but the separate models provided more accuracy, especially for 'votes'.
- The incorporation of early stopping in our training loop was instrumental in preventing overfitting, thereby ensuring the generalization capability of our model.
- Comparing our models with baseline models underscored the superiority of neural networks in handling complex data relationships.

6 Conclusion

In conclusion, this paper presents a novel approach to predicting anime ratings and popularity using a tag-based system. This method is particularly innovative as it allows for predictions even before an anime is released, utilizing tags that describe the anime's attributes. Our experiments have shown that this model can successfully predict both the quality (related to scores) and popularity (related to votes) of anime.

The integration of this model into anime-related websites can greatly enhance their functionality. By automatically providing predictions on an anime's potential success, these platforms can offer more insightful recommendations to users and assist creators and studios in making informed decisions during the production phase. This predictive capability could revolutionize the way anime is marketed and consumed, offering a more data-driven approach to the anime industry.

Future work could focus on refining the model's accuracy and exploring its applicability in different contexts, such as in streaming services or in specific genres of anime. Additionally, further research could investigate the potential of integrating user feedback post-release to continually update and improve the model's predictions. Overall, the findings of this study pave the way for a more analytical and predictive approach to anime rating and popularity assessment, with significant implications for both consumers and creators in the anime industry.

References

[1] Mladen Marović, Marko Mihoković, Mladen Mikša, Siniša Pribil, and Alan Tus. Automatic movie ratings prediction using machine learning. In *2011 Proceedings of the 34th International Convention MIPRO*, pages 1640–1645, 2011.

[2] P. Li and S. Yamada. A movie recommender system based on inductive learning. In *IEEE Conference on Cybernetics and Intelligent Systems, 2004.*, volume 1, pages 318–323 vol.1, 2004. doi: 10.1109/ICCIS.2004.1460433.

[3] Gayatree Ganu, Noemie Elhadad, and Amélie Marian. Beyond the stars: improving rating predictions using review text content. In *WebDB*, volume 9, pages 1–6, 2009.

[4] Ayşegül Özkaya Eren and Mustafa Sert. Movie rating prediction using ensemble learning and mixed type attributes. In *2017 25th Signal Processing and Communications Applications Conference (SIU)*, pages 1–4. IEEE, 2017.

[5] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. Grouplens: An open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work*, CSCW '94, page 175–186, New York, NY, USA, 1994. Association for Computing Machinery. ISBN 0897916891. doi: 10.1145/192844.192905. URL https://doi.org/10.1145/192844.192905.

[6] David M Pennock, Eric J Horvitz, Steve Lawrence, and C Lee Giles. Collaborative filtering by personality diagnosis: A hybrid memory-and model-based approach. *arXiv preprint arXiv:1301.3885*, 2013.

[7] Deniz Demir, Olga Kapralova, and Hongze Lai. Predicting imdb movie ratings using google trends. *Dept. Elect. Eng, Stanford Univ., California*, 2012.

[8] Andrei Oghina, Mathias Breuss, Manos Tsagkias, and Maarten De Rijke. Predicting imdb movie ratings using social media. In *Advances in Information Retrieval: 34th European Conference on IR Research, ECIR 2012, Barcelona, Spain, April 1-5, 2012. Proceedings 34*, pages 503–507. Springer, 2012.

[9] Abhishek Kesharwani, Rakesh Bharti, and M Tech Scholar. *Movie rating prediction based on: twitter sentiment analysis*. LAP LAMBERT Academic Publishing, 2017.

[10] Sandy Çağlıyor and Başar Öztayşi. Predicting movie ratings with machine learning algorithms. In Cengiz Kahraman, Sezi Cevik Onar, Basar Oztaysi, Irem Ucal Sari, Selcuk Cebi, and A. Cagri Tolga, editors, *Intelligent and Fuzzy Techniques: Smart and Innovative Solutions*, pages 1077–1083, Cham, 2021. Springer International Publishing. ISBN 978-3-030-51156-2.

[11] Warda Ruheen Bristi, Zakia Zaman, and Nishat Sultana. Predicting imdb rating of movies by machine learning techniques. In *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, pages 1–5, 2019. doi: 10.1109/ICCCNT45670.2019.8944604.

[12] Ping-Yu Hsu, Yuan-Hong Shen, and Xiang-An Xie. Predicting movies user ratings with imdb attributes. In Duoqian Miao, Witold Pedrycz, Dominik ŚlĆzak, Georg Peters, Qinghua Hu, and Ruizhi Wang, editors, *Rough Sets and Knowledge Technology*, pages 444–453, Cham, 2014. Springer International Publishing. ISBN 978-3-319-11740-9.