Dean&Francis
ISSN 2959-6157

Movie Rating Prediction based on Data Mining and Machine Learning

Shiyu Nie[1, *], **Zuole Wu**[2] and **Zhihan Zhong**[3]

[1] Faculty of Innovation Engineering, Macau University of Science and Technology, Macau, China
[2] College of Computer and Software Engineering, Xihua University, Chengdu, Sichuan, China
[3] College of mechanical engineering, Taiyuan University of Technology, Taiyuan, Shanxi, China
*Corresponding author: ldy87392@gmail.com

Abstract:

In personalized recommendation, accurately predicting the user's rating of movies can significantly improve the quality of recommendations, enhance user experience and platform stickiness. In the analysis of the film industry, it is helpful for film market performance evaluation, audience preference analysis and content creation guidance; In the field of academic research, it is also one of the important benchmarks for testing the performance of data mining and machine learning models. This paper systematically reviews the research progress of film rating prediction based on data mining and machine learning. This paper first reviews the traditional prediction methods relying on metadata and collaborative filtering, and analyzes their advantages and limitations. Then, the sentiment feature extraction and modeling method based on text comments is discussed, and the role of sentiment information in improving the interpretability and accuracy of prediction is expounded. Furthermore, the multimodal fusion strategy of integrating metadata, text, vision, audio and other multi-source information is discussed, and the technical characteristics and research status of different fusion levels are summarized. This paper also summarizes the common datasets and evaluation indicators in this field, and points out that there are still deficiencies in data integrity, modal fusion depth and interpretability. Finally, this paper looks forward to future research in the directions of deep multimodal interaction, cold start mitigation, and model interpretability enhancement, in order to provide a systematic reference for researchers in related fields.

Keywords: Multimodal Fusion, Machine Learning, Sentiment Analysis, Recommender Systems.

1. Introduction

As an important part of the global cultural industry, the market value and social influence of film continue to grow. According to statistics, the global film market has exceeded 100 billion US dollars, and the rapid development of streaming media platforms has spawned massive data on movie content and user interaction [1]. In this context, film ratings have become a key quantitative indicator to measure the artistic quality, audience acceptance and market performance of films. Ratings not only directly affect the audience‘s viewing choices and film box office revenue, but also profoundly affect all aspects of the film industry chain: high ratings can improve the scheduling rate and on-demand rate of films, accumulate word-of-mouth capital for directors and actors, and then affect the investment and creative direction of subsequent projects; For the platform, accurate score prediction is the core of building a personalized recommendation system, which is directly related to the efficiency of user retention, content operation and business monetization.

However, film scoring is essentially a complex product of audience subjective emotions, aesthetic preferences and group consensus, and its generation and evolution are affected by multiple factors such as film content, audience characteristics, social culture, etc., showing a high degree of non-linearity and dynamics. Traditional scoring prediction methods mostly rely on historical scoring matrices or limited metadata, and although they achieve prediction functions to a certain extent, they generally face shortcomings such as data sparsity, cold start problems, and insufficient utilization of multi-source information. For example, a collaborative filtering method based solely on a user-item rating matrix struggles to handle predictions of newly released movies or newly registered users; However, the method of using only structured metadata such as director, actor, and genre cannot fully capture the rich emotional and stylistic signals contained in unstructured content such as commentary text, poster visuals, and trailer audio [2, 3].

In recent years, the rapid development of data mining and machine learning technology has opened up a new path for film rating prediction research [4, 5]. Researchers began to try to automatically learn effective feature representations from multimodal and multi-source data, and build an integrated prediction model that integrates content, emotion, context and user behavior [6, 7]. These methods not only improve the accuracy of predictions but also enhance the explainability of „why users like a certain movie“. However, the research results in this field are scattered in multiple academic branches such as recommendation systems, natural language processing, and multimedia computing, and lack systematic combining and comparison. At the same time, existing studies still have obvious limitations in modal alignment, deep semantic fusion, and real-time prediction [8, 9, 10].

Therefore, this paper aims to systematically review the prediction research of film ratings based on data mining and machine learning. This paper will first review the traditional methods based on metadata and collaborative filtering, analyze their design principles and scope of application, then focus on how to extract sentiment features from commentary texts and use them to enhance prediction models, and then delve into the multimodal fusion method that fuses text, visual, audio and other multi-source information, and summarize its technical framework and evolution trend. In addition, this paper will summarize the commonly used datasets and evaluation index systems in this field, point out the shortcomings of existing research, and look forward to future research directions from the dimensions of deep fusion architecture, sparsity mitigation, and interpretability enhancement. The review aims to provide researchers in academia and industry with a clear and comprehensive map of technological evolution, and promote the development of film score prediction in a more accurate, intelligent and credible direction.

2. Traditional Methods

Movie rating prediction is usually regarded as a supervised learning problem, that is, based on historical „user-movie-rating“ records, the mapping relationship between user characteristics and movie characteristics is learned, so as to predict the possible ratings of unwatched films by users. Before deep learning and multimodal methods were widely used, the field mainly relied on traditional machine learning methods and collaborative filtering frameworks. Traditional machine learning methods are usually based on artificially constructed structured features, such as film genre, release year, director, starring, etc.; The collaborative filtering method is directly based on the user-item scoring matrix, and user preferences are modeled through similarity calculation or matrix decomposition [4].

Compared with the subsequent deep model, the traditional method has a relatively simple structure, low computational overhead, and a certain degree of interpretability, so it has been widely used in early recommendation systems. However, this type of method is highly dependent on feature engineering, making it difficult to fully mine high-dimensional sparse data and unstructured information such as text and images, and the prediction accuracy is limited in complex application scenarios.

In studies based on regression and tree models, score

ISSN 2959-6157

prediction is directly modeled as a regression problem. Researchers usually use video metadata to construct structured feature vectors and use user scores as supervisory signals to train models such as linear regression, support vector machines, decision trees, and random forests. Relevant studies show that such methods have good stability when the features are relatively complete and the data scale is moderate, and it convenient to analyze the influence of each feature on the score.

Based on Netflix program and IMDb score data, Fariha et al. extracted characteristics such as genre, number of seasons and premiere year, and trained k nearest neighbors, linear regression and random forest models respectively [1]. On the one hand, Marović et al. compared the prediction performance of decision trees, support vector machines, and k-nearest neighbor models under metadata features and found that the error differences between models were limited, while feature selection and data preprocessing had a more significant impact on the results [4].

Collaborative filtration and matrix decomposition are another core type of technology in traditional methods. Collaborative filtering completes missing scores by calculating the similarity between users or items, while the matrix decomposition method maps the scoring matrix to the low-dimensional latent space, learns latent factor vectors for users and movies respectively, and approximates the score value by inner product. This method has achieved significant results on large-scale datasets such as the Netflix Prize, but it relies heavily on explicit scoring matrices, limits performance in cold-start users and new movie scenes, and makes it difficult to directly integrate multimodal information such as text and images.

3. Approaches Based on Textual Sentiment Features

As user review data grows, emotions, polarities, and subjective expressions often directly reflect the audience‘s true preferences for movies, making them valuable in predicting ratings. Its Chinese review is one of the most valuable unstructured features in movie rating predictions. As the most direct and free form of expression for users after watching the movie, text comments contain rich subjective emotions and a detailed viewing experience. Compared with structured features (e.g., director, actor, genre, etc.), review text can more accurately reveal user satisfaction, emotional tendencies, concerns, and potential preferences, which are often highly correlated with the final score.

In his research, Ramos clearly pointed out that the sentiment characteristics (including emotional polarity, emotional intensity, opinion weight, etc.) extracted based on comment text can effectively improve the accuracy of rating prediction. By combining emotional cues with traditional text features, they validate the importance of emotional features in modeling user attitudes and capturing fine-grained evaluation signals. On this basis, Fariha further demonstrated the feasibility of combining emotional signatures with traditional machine learning methods [1]. They added information about emotional propensity extracted from reviews or synopses when predicting IMDb ratings for Netflix titles, and the results showed that emotion-related features significantly enhance the model‘s ability to explain differences in user ratings, especially in scenarios where metadata is limited and user text is highly subjective.

Overall, the introduction of emotional features provides a more expressive way to portray user preferences for movie rating prediction, which can compensate for the lack of signal when relying only on structured metadata. Emotional cues not only improve the sensitivity of the predictive model to users‘ authentic attitudes but also enhance the model‘s robustness in diverse textual environments. Therefore, whether combined with traditional machine learning methods or deep learning models, emotional features play an important role in modern film rating prediction tasks, and are becoming one of the key directions for building high-performance models.

4. Multimodal Integration

A “modality” is like a human sense, the single-modal model is like only seeing the picture, not hearing the sound, and having no touch or smell, while the multimodal model perfectly encompasses these senses. In movie rating prediction research, “modality” refers to different sources of information or data types that describe a movie. Traditional models often rely on a single modality, such as using only the metadata of the film (e.g., director, actor, genre) [1] or analyzing only the user’s historical rating matrix (collaborative filtering). However, a film is a complex multimedia product, and its final score is based on the audience’s consideration of multiple aspects. Therefore, multimodal fusion technology came into being.

Multimodal fusion aims to integrate different sources and different types of data to obtain a more comprehensive and accurate prediction model than any single modality through information complementarity. In the field of film score prediction, the main multimodal data includes: metadata modality, text modality, visual modality and audio modality. Metadata modalities include basic information about a movie, such as genre, director, actors, release year, duration, etc., which are structured data that is easy

to process. Text modalities include movie synopses, user reviews, professional film reviews, etc. Text data contains rich emotional, thematic, and subjective evaluation information, and the main features of the film can be extracted through abstract processing of a large number of texts. Visual modalities include movie posters, promotional videos, keyframes in feature films, etc. Visual information can intuitively convey the visual style and tone of the film (such as bright and warm or dark and cold).

Audio modalities include movie soundtracks, sound effects, dialogue tones, etc. Audio characteristics help to judge the atmosphere of the scene at the time of the movie (e.g., passionate, soothing, or horror).

The core challenge and key technology of multimodal fusion lies in how to effectively integrate these heterogeneous data. According to the integration level, the fusion strategy can be divided into early fusion (feature level), late fusion (decision level) and intermediate fusion (model level). Among them, the importance of intermediate fusion using attention mechanisms and other technologies to realize modal interaction within the model and dynamically learn different modalities is the most promising research direction at present.

Earlier research focused on mining the predictive power of specific single modalities; however, the limitations of single modality prompted researchers to turn to richer data sources. In their study, Marović et al. systematically compared collaborative filtering (based on a user-rating matrix) with content-based methods (based on film metadata) and found that a probabilistic latent semantic analysis model performed best. Although this work does not introduce multimedia content, it reveals an important signal of collective user behavior patterns and indirectly indicates the need to mix different information sources (user and object characteristics) [4].

The study of multimodal comparative analysis of multimedia content in the true sense comes from Rahmani et al. [2]. In a key work, they systematically compared for the first time audio and visual characteristics extracted from trailers with metadata (e.g., genre, tags) in predicting film average ratings, rating divergences, and popularity. The conclusion is enlightening: user-generated label data performs best in all prediction tasks, and deep neural networks show an advantage in processing these features. This work not only proves the differences in the prediction information contained in different modalities, but also demonstrates the huge potential of multimodal fusion more powerfully, and points out the direction for subsequent research.

To sum up, film ratings are the result of the combined action of multiple attributes of movies, and no single modality can perfectly predict movie ratings. Research trends are shifting from evaluating the effectiveness of monomodality to designing more sophisticated architectures to achieve deep multimodal fusion. Future research focuses on building a unified model that can adaptively weigh multi-source information such as metadata, text emotion, visual style, audio emotion, and user group behavior, so as to achieve more accurate and stable prediction of film scores.

5. Datasets and Evaluation Metrics

5.1 Common Datasets

Currently, movie rating prediction research is mainly carried out on several public datasets, among which the MovieLens series is one of the most widely used. This dataset is maintained by the GroupLens team and includes multiple versions of MovieLens 100K, 1M, 10M, and 25M, with data sizes ranging from 100,000 to tens of millions, providing basic information such as user ID, movie ID, rating value, movie type, release year, etc., with a standardized data structure and few missing, suitable as standard benchmark data for collaborative filtering and matrix decomposition methods.

The IMDb Open Dataset focuses on providing content and production-level information for movies and TV shows, including film titles, durations, genres, production countries, directors, actors, and average ratings based on user votes. In rating prediction research, IMDb is often used in conjunction with MovieLens or Netflix datasets to complement the structured features of the movie side and enrich the movie profile.

The Netflix Prize dataset is based on a large-scale user-movie rating matrix and contains about 100 million rating records, making it a classic dataset that drives the development of collaborative filtering and latent factor models. The dataset contains a few additional content features, but it is ideal for evaluating the generalization and computational efficiency of the model in large-scale, high-sparsity scenarios.

In addition, the Kaggle platform provides a variety of datasets that contain both ratings and user text comments, which belong to the combined data of „rating label + review text“. This type of data provides important support for studying the relationship between sentiment characteristics, thematic characteristics and scores, and is suitable for the construction and validation of emotion enhancement models and multimodal models.

5.2 Evaluate the Indicators

Movie score prediction is often considered a regression

ISSN 2959-6157

task, so most studies employ regression error metrics to measure the difference between predicted and true scores, with common metrics including mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2).

MSE measures the overall bias by calculating the square mean of the prediction error, which is more sensitive to large error samples. RMSE is the square root of MSE, and its dimension is consistent with the original score, which has more intuitive physical significance and is one of the most commonly used evaluation indicators in recommendation system studies. MAE calculates the absolute error average between the predicted and true values, is less sensitive to outliers than MSE, and has better stability in the presence of noisy data.

R^2 is used to describe the degree of interpretation of the variance of the observed data, and the closer the value is to 1, the stronger the model's ability to fit the score change. In some studies, ranking indicators such as Precision@K, Recall@K, and NDCG are also introduced to supplement the evaluation of model performance from the perspective of recommended ranking quality. The combination of multiple metrics can help to comprehensively evaluate the performance of the scoring prediction model from different perspectives.

6. Existing limitations and future prospects

Current research still has some limitations. First, at the data level, existing datasets generally have the problem of incomplete modalities. For example, although the study of Rahmani et al. compared multiple modalities, the multimedia features used were not fully integrated and validated on larger datasets, such as metadata and commentary text [2]. In addition, as pointed out by Marović et al. [4], the sparsity of user rating data has always been a major challenge for collaborative filtering methods.

Secondly, at the model level, the existing fusion methods mostly stay in feature splicing or decision weighting. Ramos et al.'s research shows that even better hybrid models fail to dig deep into the deep semantic associations between text emotion and other modalities such as visual style. Rahmani et al.'s work also mainly compares the effects of inputting models separately into different modalities, rather than end-to-end deep fusion [3].

Looking ahead, this field can be explored in depth from the following directions:

Deep multimodal fusion: Based on the findings of Rahmani et al., future research should focus on designing more advanced fusion architectures (such as attention-based mechanisms) to adaptively capture complex interactions between modalities rather than simple combinations [3].

Mitigating data sparsity: Exploring the use of transfer learning or generative models to supplement data to cope with the cold start of new users and new movies, in view of the characteristics of relying on user or item features in the studies of Marović et al. and Fariha et al.

Enhanced interpretability: Combined with Ramos et al.'s idea of extracting explainable emotional features from text, future models should not only predict scores, but also identify which modalities (such as „great performance" or „stunning visual effects") dominate the predictions [3].

7. Conclusion

This paper systematically reviews the research on film score prediction based on data mining and machine learning. This article reviews the technological evolution path: from traditional methods relying on metadata and collaborative filtering, to sentiment analysis methods that dig deep into commentary texts, to multimodal fusion methods aimed at comprehensively utilizing multi-source information. The results show that the fusion of multiple modal information is the key direction to improve prediction performance. However, current research still generally faces challenges such as insufficient integrity of multi-source data and insufficient deep semantic fusion between modalities. Future research should focus on developing deeper multimodal fusion models and effectively solving classic problems such as data sparseness and cold start. The work of this review provides a systematic reference for subsequent research in this field.

Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

[1] Fariha F, Rahman M, Hasan M. Prediction of IMDb rating in Netflix shows using machine learning algorithms. Proceedings of the ACM International Conference on Computing and Information Technology, 2024: 85-92. ACM Digital Library.

[2] Rahmani H A, Haddad M, Boudiaf M. A regression approach to movie rating prediction using multimedia content and metadata. Proceedings of the MediaEval Workshop, 2019: 50-56. CEUR Workshop Proceedings, Vol. 2670.

[3] Ramos J, Silva T, Costa L. Movie rating prediction using sentiment features. Proceedings of the Workshop on Sentiment, Emotion, and Social Signals (SALD), 2022: 14-21. Association for Computational Linguistics.

[4] Marović M, Mihoković M, Mikša M, Pribil S, Tus A, et al. Automatic movie ratings prediction using machine learning.

Proceedings of the MIPRO Conference on Computers in Technical Systems, 2017: 1002-1007.
[5] Walker E, Chen I, Evans L, et al. Leveraging deep learning for accurate movie rating prediction.
[6] Zheng Y, Tang B, Ding W, et al. A neural autoregressive approach to collaborative filtering. International Conference on Machine Learning, 2016: 764-773.
[7] Wu L, He X, Wang X, et al. A survey on neural recommendation: From collaborative filtering to content and context enriched recommendation. arXiv preprint arXiv:2104.13030, 2021.
[8] Tsai Y H H, Bai S, Liang P P, et al. Multimodal transformer for unaligned multimodal language sequences. Proceedings of the Annual Meeting of the Association for Computational Linguistics, 2019: 6558.
[9] Song B, Zhou R, Ahmed F. Multi-modal machine learning in engineering design: A review and future directions. Journal of Computing and Information Science in Engineering, 2024, 24(1): 010801.
[10] Wu S, Sun F, Zhang W, et al. Graph neural networks in recommender systems: A survey. ACM Computing Surveys, 2022, 55(5): 1-37.