A Hybrid Deep Learning Approach for Personalized Movie Recommendations with Content-Based Features

Yehia Mahrous, Omar Elborollosy, Ziad Abdelhafiz
Department of Computer Science
Egypt University of Informatics
Cairo, Egypt

***Abstract*—Movie recommender systems are essential for enhancing user experience on streaming platforms by providing personalized suggestions amidst vast content libraries. This paper introduces a hybrid recommender system that integrates collaborative filtering and content-based approaches using a deep learning model trained on the MovieLens dataset. By leveraging user and movie embeddings alongside genre and keyword features, the system predicts both numerical ratings and binary relevance scores through a dual-task learning framework. Novel enhancements include user historical data analysis to ensure recommendations align with individual preferences and personalized content-based recommendations for movies similar to a user-specified title. Experimental results demonstrate robust performance, achieving a Mean Absolute Error (MAE) of 0.39, Precision@10 of 0.40, and Recall@10 of 0.75. These findings underscore the system's ability to deliver accurate, relevant, and interpretable recommendations, making it a valuable tool for improving user satisfaction and engagement.**

I. Introduction

The proliferation of digital content on streaming platforms has transformed how users consume media, offering unprecedented access to thousands of movies. However, this abundance poses a significant challenge: users often struggle to find content that matches their preferences, leading to decision fatigue and reduced engagement. Recommender systems address this issue by curating personalized suggestions, playing a pivotal role in enhancing user satisfaction and retention on platforms like Netflix, Hulu, and Amazon Prime. These systems traditionally rely on two primary approaches: collaborative filtering (CF), which uses historical user interactions to identify patterns, and content-based (CB) filtering, which leverages item attributes such as genres or descriptions to suggest similar content.

Despite their strengths, traditional methods have notable limitations. CF excels at capturing user preferences through interaction data but struggles with cold-start scenarios—new users or items with insufficient interaction history—resulting in poor recommendation quality. CB methods mitigate this by focusing on item features, yet they often fail to account for the nuanced, evolving preferences of users that extend beyond static metadata. These shortcomings highlight the need for hybrid approaches that combine the complementary strengths of CF and CB to deliver more robust and adaptable recommendations.

Recent advancements in deep learning have opened new possibilities for recommender systems by enabling the integration of diverse data types and modeling complex user-item relationships. This paper proposes a hybrid movie recommender system that synergizes CF and CB techniques using a neural network architecture, trained on the MovieLens dataset. The system employs embeddings for users, movies, genres, and keywords, processed through a dual-task learning framework to predict both numerical ratings (on a 0.5–5.0 scale) and binary relevance scores (rating > 3.5). To enhance personalization and interpretability, two novel features are introduced: (1) a user history analysis tool that displays rated movies, genres, overviews, and ratings to validate preference alignment, and (2) personalized recommendations for movies similar to a user-specified title, tailored to individual tastes. These additions improve recommendation accuracy while fostering transparency and user trust.

The motivation for this research stems from the growing demand for recommender systems that balance accuracy with interpretability, particularly in streaming contexts where user satisfaction drives platform success. By integrating content features and enabling user history analysis, this system delivers recommendations that are precise, relevant, and aligned with individual preferences. The paper is structured as follows: Section II reviews related work, Section III details the methods, Section IV describes the experimental design, Section V presents the results, Section VI concludes, and Section VII lists references.

II. Related Work

Recommender systems have undergone significant evolution since their early development, with foundational approaches rooted in collaborative filtering. Matrix factorization, popularized by Koren et al. [1], decomposes user-item rating matrices into latent factors, effectively capturing user preferences and achieving widespread success in applications like the Netflix Prize. However, CF methods face challenges in cold-start scenarios, where insufficient interaction data for new users or items limits their effectiveness. Content-based methods, as explored by Lops et al. [2], address this gap by leveraging item metadata—such as movie genres, tags, or textual descriptions—to recommend similar items based on user profiles. While effective for new items, CB approaches often struggle to model the dynamic, implicit preferences that emerge from user behavior.

Hybrid recommender systems seek to bridge these gaps by combining CF and CB techniques. Burke [3] classifies hybrid systems into categories like weighted, switching, and feature combination methods, demonstrating their superior performance over standalone approaches. The rise of deep learning has further propelled hybrid systems forward. He et al. [4] proposed Neural Collaborative Filtering (NCF), which employs neural networks to capture non-linear user-item interactions, outperforming traditional matrix factorization in accuracy and flexibility. Wang et al. [5] introduced a collaborative deep learning model that integrates content features with CF, enhancing recommendation quality in sparse datasets. Zhang et al. [6] provide a comprehensive survey of deep learning-based recommender systems, emphasizing their ability to process heterogeneous data, including metadata, text, and images.

Despite these advancements, many existing systems prioritize prediction accuracy over interpretability, often leaving users uncertain about why specific recommendations are made. This lack of transparency can erode trust and engagement, particularly on streaming platforms where user experience is paramount. Few systems incorporate user history analysis or context-aware, personalized content recommendations as core features. This work builds on prior hybrid models by integrating genre and keyword embeddings into a dual-task neural network, while introducing novel tools for user history analysis and personalized similar movie recommendations. These contributions enhance both the accuracy and interpretability of recommendations, addressing critical gaps in the literature and making the system well-suited for real-world streaming applications.

III. Methods

The proposed system is a hybrid recommender model that integrates collaborative filtering and content-based features within a deep neural network framework. It leverages the MovieLens dataset, encompassing movie metadata (titles, genres, overviews), user ratings, and keywords, to generate personalized recommendations.

Data Preprocessing Data preprocessing is critical to ensure quality and consistency. The process involves several steps: - **Movie Cleaning**: The ‘movies metadata.csv‘ file is refined to retain only essential columns: ‘id‘, ‘original title‘, ‘genres‘, and ‘overview‘. Movies with missing overviews are excluded, reducing the dataset from 45,000 to 40,000 entries. Movie IDs are standardized to integers, and duplicates are removed based on ‘id‘ and ‘original title‘ to eliminate inconsistencies. - **Genre Parsing**: Genres are extracted from JSON-like strings using ‘ast.literal eval‘ to safely handle syntax variations. Movies with no genres are dropped, ensuring all entries have valid metadata. - **Keyword Processing**: Keywords are parsed similarly and filtered to the top 1000 most frequent terms, reducing dimensionality while preserving informative features. - **Ratings Integration**: The ‘ratings small.csv‘ file, containing user ratings, is merged with the cleaned movie data via an inner join on ‘movieId‘ and ‘id‘, yielding 80,000 ratings from 700 users. This ensures ratings align with movies possessing complete metadata.

To balance the dataset, negative samples are generated at a 1.5:1 ratio per user, assigning unrated movies a rating of 0.5. Ratings are normalized to [0, 1] to facilitate stable model training.

Model Architecture The neural network comprises four input branches: - **User and Movie Embeddings**: Users and movies are encoded as 256-dimensional embeddings, representing latent preferences and item traits. - **Genre and Keyword Embeddings**: Genres and keywords are embedded into 32-dimensional vectors, with sequences padded to fixed lengths (max genres and 50 keywords, respectively). - **Feature Processing**: User and movie embeddings are flattened, while genre and keyword embeddings undergo global average pooling. These vectors are concatenated into a unified feature set. - **Dense Layers**: The concatenated features pass through four dense layers (128, 64, 32, 16 units) with ReLU activation, batch normalization, and 30% dropout to mitigate overfitting. - **Outputs**: The model generates two outputs: - A rating prediction (0.5–5.0 scale) via a sigmoid-activated layer, scaled appropriately. - A binary relevance score (rating > 3.5) via another sigmoid-activated layer.

The model uses the Adam optimizer (learning rate 0.0001) and a composite loss function: mean squared error (weight 1.0) for ratings and binary cross-entropy (weight 0.85) for relevance. Metrics include MAE for ratings and accuracy for relevance.

Novel Features Two innovative features enhance usability and interpretability: - **User History Analysis**: The ‘display user history‘ function retrieves a user’s rated movies, presenting titles, genres, overviews, and ratings in a table, and logs genre distributions to validate preference alignment (e.g., frequent high ratings for comedies). - **Personalized Similar Movie Recommendations**: The ‘recommend similar movies‘ function uses Jaccard similarity (70% genres, 30% keywords) to find movies similar to a specified title, then ranks them by predicted user ratings, ensuring personalized and contextually relevant suggestions.

IV. Experimental Design

The experiments assess the system’s performance in accuracy, relevance, and personalization using the MovieLens dataset.

Dataset The dataset includes: - **Movies**: 45,000 entries, reduced to 40,000 after preprocessing (e.g., removing missing overviews, duplicates, empty genres). - **Ratings**: 100,000 ratings from 700 users, reduced to 80,000 post-merging. - **Keywords**: Filtered to the top 1000 frequent terms.

An 80%-20% train-test split, with a fixed random seed, ensures reproducibility and robust evaluation.

Training The model trains for up to 20 epochs with a batch size of 512, employing: - **Early Stopping**: Halts training if validation MAE stagnates for three epochs, restoring the best weights. - **Learning Rate Scheduling**: Reduces the learning rate by 50% if validation MAE plateaus for two epochs (minimum 1e-6).

These techniques optimize convergence and computational efficiency.

Evaluation Metrics Performance is measured via: - **Mean Absolute Error (MAE)**: Average absolute difference between predicted and actual ratings. - **Precision@10**: Proportion of top-10 recommended movies with true ratings > 3.5. - **Recall@10**: Proportion of relevant movies captured in the top-10 list.

User Analysis Personalization is tested on users 13 and 123, analyzing historical data, predicting ratings (e.g., for ‘Toy Story’), and generating tailored general and similar movie recommendations.

V. Experimental Results

The results highlight the system’s effectiveness in delivering accurate, relevant, and personalized recommendations.

Model Performance Test set performance is summarized below: - **MAE (0.39)**: Indicates high rating prediction accuracy. - **Precision@10 (0.40)**: 40% of top-10 recommendations are relevant. - **Recall@10 (0.75)**: Captures 75% of relevant movies in the top-10.

TABLE I
MODEL PERFORMANCE METRICS

Metric	Value
MAE	0.37
Precision@10	0.39
Recall@10	0.73

User Preference Alignment - **User 123**: Prefers drama (e.g., ‘The Shawshank Redemption’, 4.5); receives drama-heavy recommendations like ‘Forrest Gump’. - **User 13**: Favors action/sci-fi (e.g., ‘The Matrix’, 4.5); gets action-oriented suggestions like ‘Blade Runner’.

For ‘Toy Story’, User 123 receives family-friendly options (e.g., ‘Finding Nemo’), while User 13 gets action-animated films (e.g., ‘The Incredibles’), showcasing personalization.

Robustness and Interpretability The user history tool and similar movie feature enhance transparency and relevance, handling edge cases like missing titles effectively.

VI. Conclusions

This hybrid movie recommender system, achieving an MAE of 0.39, Precision@10 of 0.40, and Recall@10 of 0.75, demonstrates strong performance in delivering accurate and relevant recommendations. Its novel features—user history analysis and personalized similar movie suggestions—improve interpretability and engagement, addressing key needs in streaming platforms. Future enhancements could include additional metadata (e.g., cast, directors) or real-time feedback integration, with potential applications in music or e-commerce.

References

[1] Y. Koren, R. Bell, and C. Volinsky, “Matrix factorization techniques for recommender systems,” *Computer*, vol. 42, no. 8, pp. 30–37, 2009.

[2] P. Lops, M. De Gemmis, and G. Semeraro, “Content-based recommender systems: State of the art and trends,” in *Recommender Systems Handbook*, Springer, 2011, pp. 73–105.

[3] R. Burke, “Hybrid recommender systems: Survey and experiments,” *User Modeling and User-Adapted Interaction*, vol. 12, no. 4, pp. 331–370, 2002.

[4] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T.-S. Chua, “Neural collaborative filtering,” in *Proc. 26th Int. Conf. World Wide Web*, 2017, pp. 173–182.

[5] H. Wang, N. Wang, and D.-Y. Yeung, "Collaborative deep learning for recommender systems," in *Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2015, pp. 1235–1244.

[6] S. Zhang, L. Yao, A. Sun, and Y. Tay, "Deep learning based recommender system: A survey and new perspectives," *ACM Comput. Surv.*, vol. 52, no. 1, pp. 1–38, 2019.