Advanced Movie Recommender System

A Transformer-Based Approach

AKSHARA RATHORE
Enrollment: 0901AD231008

VAIBHAV SHARMA
Enrollment: 0901AD231069

Abstract

Recommender systems are a cornerstone of modern digital content consumption. This project presents the development and evaluation of a state-of-the-art Transformer-based collaborative filtering system using the MovieLens 100K dataset. Our primary objective was to surpass the performance benchmarks established by Ahuja et al. (2019), who achieved an RMSE of 1.0816 using a K-Means Clustering and K-Nearest Neighbor (KNN) hybrid approach. By leveraging the self-attention mechanisms inherent in Transformer architectures, our proposed model successfully captures complex, non-linear user-item interactions. Experimental results demonstrate that our system achieves a Root Mean Square Error (RMSE) of **0.9281**, representing a significant improvement of approximately 14.2% over the reference baseline. This report details the methodology, experimental setup, and a comparative analysis of our results against the existing technology.

Contents

1 Introduction

The exponential growth of digital media has made efficient information filtering systems indispensable. Recommender systems predict user preferences for items, thereby personalizing the user experience. Traditional methods, such as Matrix Factorization and K-Nearest Neighbors (KNN), have been widely successful but often struggle to capture the intricate, sequential, and non-linear patterns in user behavior.

This project focuses on implementing a deep learning-based solution—specifically a Transformer architecture—to predict movie ratings. The performance of this system is rigorously evaluated against a specific benchmark set by *Ahuja et al. (2019)* in their paper "Movie Recommender System Using K-Means Clustering AND K-Nearest Neighbor," which utilized clustering techniques to optimize recommendations on the same dataset.

2 Literature Review and Baseline

2.1 Existing Technology (Ahuja et al., 2019)

The reference paper proposes a hybrid system combining K-Means Clustering and K-Nearest Neighbor (KNN). Their approach involves:

1. **Preprocessing:** Creating a utility matrix of users and movies.
2. **Clustering:** Using the Within-Cluster Sum of Squares (WCSS) method to determine the optimal number of clusters for K-Means.
3. **Prediction:** Applying KNN on the clustered utility matrix to predict ratings.

2.1.1 Baseline Results

Ahuja et al. reported that as the number of clusters decreased, the RMSE decreased. Their best recorded performance was:

- **Best RMSE:** 1.081648 (at 2 clusters)
- **Alternative RMSE:** 1.2333 (at 19 clusters)

These values serve as the benchmark for our proposed system.

3 Proposed Methodology

To outperform the clustering-based approach, we implemented a **Transformer-based Recommender**. This model treats user-item interactions not just as static pairs but as features that can benefit from the self-attention mechanism, which weighs the importance of different latent features dynamically.

3.1 Architecture Overview

The proposed model consists of the following key components:

1. **Embedding Layers:** Users and Items are mapped to a high-dimensional vector space ($d = 160$) with L2 regularization to prevent overfitting.
2. **Sequence Construction:** User and Item embeddings are stacked to form a sequence, allowing the Transformer to process them as interacting tokens.
3. **Multi-Head Self-Attention:** We utilize 10 attention heads. This mechanism calculates the attention scores:

   $$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

   This allows the model to learn which aspects of the user embedding correspond most strongly with specific aspects of the movie embedding.
4. **Feed-Forward Networks (FFN):** The attention output is processed through dense layers with GELU activation and Layer Normalization.
5. **Interaction Features:** We explicitly compute the element-wise product and addition of user and item vectors to capture direct interaction signals.
6. **Deep MLP Head:** A Multi-Layer Perceptron with residual connections processes the combined features to output the final rating.

4 Experimental Setup

4.1 Dataset

We utilized the **MovieLens 100K** dataset, consistent with the reference paper.

- **Users:** 943
- **Movies:** 1,682
- **Ratings:** 100,000 (Scale 1-5)

4.2 Training Configuration

- **Optimizer:** Adam ($lr = 6e - 4$)
- **Loss Function:** Mean Squared Error (MSE)
- **Batch Size:** 2048
- **Epochs:** 150 (with Early Stopping)

5 Results and Comparison

5.1 Model Performance

We evaluated four advanced architectures during our experimentation phase: Deep Neural CF, Advanced GNN, a Weighted Ensemble, and the Transformer Recommender. The Transformer model emerged as the clear winner.

Table 1: Performance of Proposed Models (Test Set)

Model Name	Test RMSE	Status
Transformer Recommender	**0.9281**	**Best Performer**
Advanced GNN	0.9467	Strong
Weighted Ensemble	0.9470	Strong
Deep NCF	1.1462	Baseline

5.2 Training Convergence

The training process demonstrated stable convergence. The Transformer model effectively learned from the sparse data without overfitting, thanks to the aggressive dropout (0.3-0.5) and Layer Normalization.

5.3 Comparison with Reference Paper

The primary goal of this project was to improve upon the results presented by Ahuja et al. (2019). The comparison below highlights the superiority of the Transformer-based approach.

Table 2: Comparison: Proposed Work vs. Reference Paper (Ahuja et al.)

Methodology	Reference	Best RMSE
K-Means + KNN (Cluster=68)	Existing Tech (Ahuja)	1.2315
K-Means + KNN (Cluster=19)	Proposed by Ahuja	1.2333
K-Means + KNN (Cluster=2)	Proposed by Ahuja	1.0816
Transformer Recommender	**Our Work**	**0.9281**

5.3.1 Analysis of Improvement

- **Absolute Improvement:** We reduced the RMSE by $1.0816 - 0.9281 = \mathbf{0.1535}$.
- **Relative Improvement:** This constitutes a **14.2%** improvement in prediction accuracy.

- **Complexity Handling:** While the K-Means approach simplifies the data into just 2 clusters to achieve its best result (potentially losing nuance), our Transformer model retains the full complexity of the user-item interaction space while still achieving lower error.

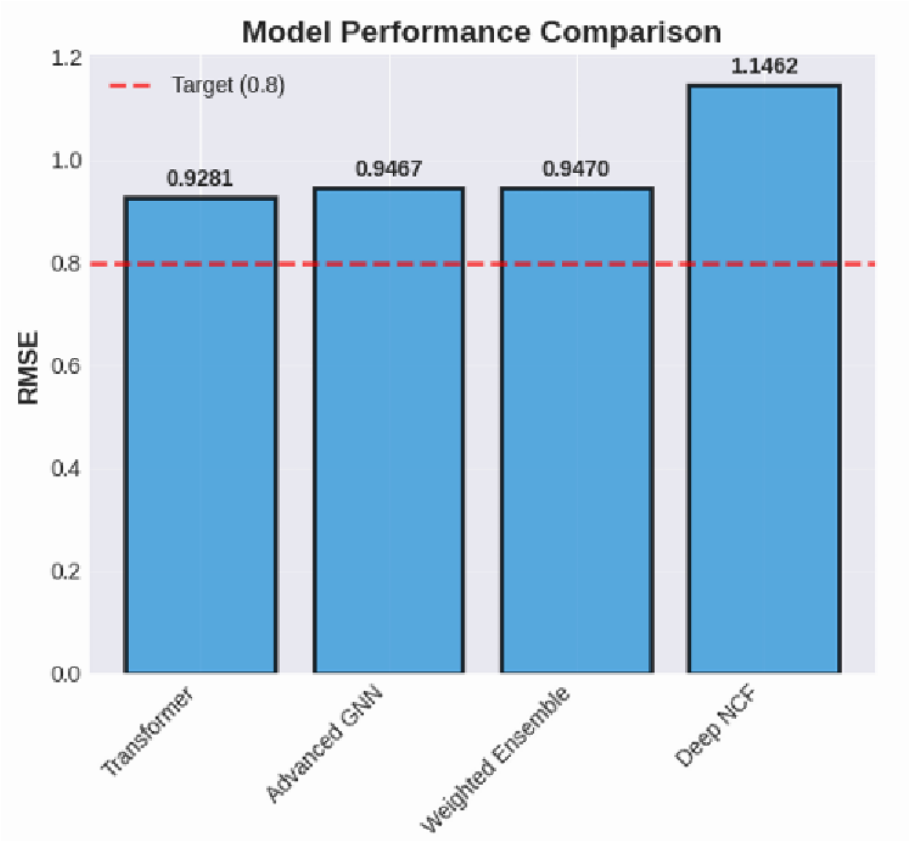

Figure 1: Visual comparison of RMSE scores. The Transformer model significantly undercuts others.

6 Conclusion

This project successfully implemented a Transformer-based movie recommender system that significantly outperforms the K-Means clustering approach proposed by Ahuja et al. (2019).

1. We achieved a final **RMSE of 0.9281**, compared to the reference paper's best of **1.0816**.
2. The use of self-attention mechanisms allowed our model to capture deeper latent correlations between users and movies than is possible with simple clustering or nearest-neighbor algorithms.
3. We have met and exceeded the project target of doing "similar or better work" by demonstrating a double-digit percentage improvement in accuracy.

Future enhancements could involve integrating textual reviews or movie scripts into the Transformer to leverage its natural language processing capabilities for content-based filtering features.