ORIGINAL RESEARCH PAPER
Volume - 15 | Issue - 08 | August - 2026 | PRINT ISSN No. 2277 - 8179 | DOI : 10.36106/ijsr
INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH

MULTI-TASK LEARNING FOR MOVIE RECOMMENDATION SYSTEMS: JOINT RATING AND GENRE PREDICTION

Computer Science

Saarah Patkar Singapore International School

ABSTRACT

Movie recommendation systems traditionally focus on single objectives, primarily rating prediction. However, user preferences are inherently multi-dimensional and influenced by semantic attributes such as genre, which are often modeled independently. This paper proposes a multi-task learning (MTL) framework that jointly predicts user ratings and movie genres using a shared neural network architecture. By leveraging the complementary nature of these tasks, the proposed approach learns richer and more robust representations. The framework consists of shared embedding layers followed by task specific output heads and is trained on the MovieLens100K dataset. Experimental results demonstrate that the multi-task model achieves a Root Mean Squared Error (RMSE) of 0.9123 for rating prediction, representing a 3.5% improvement over a single-task neural baseline, while simultaneously achieving 87.4% label-wise accuracy in genre prediction. Additional analysis shows that genre prediction acts as an implicit regularizer, improving generalization and reducing overfitting. The model is computationally efficient and suitable for real-time recommendation scenarios.

KEYWORDS

Multi-Task Learning, Recommender Systems, Deep Learning, Rating Prediction, Genre Classification

INTRODUCTION

Recommender systems play a critical role in modern digital platforms by filtering large volumes of content and presenting personalized suggestions to users. Traditional recommendation approaches primarily focus on predicting user ratings or preferences, treating the task as a standalone regression problem. While effective in controlled settings, such models often fail to capture the multi-faceted nature of user preferences observed in real-world systems.

User decisions are influenced not only by historical interactions but also by semantic properties of items, such as genre, theme, and narrative structure. In movie recommendation, genre information provides an important semantic signal that reflects both user taste and item characteristics. Despite this, rating prediction and genre modeling are typically handled as separate tasks, limiting the model's ability to exploit shared
information.

Multi-task learning (MTL) provides a principled mechanism for jointly learning related tasks using shared representations. By optimizing multiple objectives simultaneously, MTL introduces an inductive bias that can improve generalization and reduce overfitting. Motivated by these observations, this paper explores joint learning of rating prediction and genre classification within a unified neural framework.

In summary, the following are the contributions:

- Multi-task neural architecture with joint rating and genre prediction.
- Demonstrated empirically that genre prediction supplements the primary task of rating prediction on the Movielens-100K dataset.
- Ablation study and detailed analysis of performance.

Related Work

Collaborative filtering methods, particularly matrix factorization techniques, have been widely adopted for recommendation tasks due to their simplicity and effectiveness [1]. However, such methods rely exclusively on user-item interaction data and suffer from sparsity and cold-start problems.

Content-based recommendation methods incorporate item attributes to alleviate these issues [2], while hybrid approaches combine collaborative and content-based signals [3].

Despite improvements, most hybrid systems treat different learning objectives independently.

Recent advances in deep learning have enabled more expressive recommendation models. Neural Collaborative Filtering replaces linear interactions with neural networks to capture non-linear user–item relationships [4]. Autoencoder-based models [5] and recurrent architectures [6] further enhance representation learning.

Multi-task learning has been shown to improve generalization by sharing representations across related tasks[7]. In recommender systems, MTL has been applied to ranking, engagement prediction, and conversion modeling [8, 9, 10]. However, joint modeling of rating prediction and genre classification remains relatively under-explored, motivating the present work.

Problem Formulation

Let U denote the set of users and I denote the set of movies. Given a user–item pair (u,i), the primary objective is to predict the rating $\hat{r}_{ui}$. In addition, each movie i is associated with a multi-label genre vector, and the secondary objective is to predict the corresponding genre probabilities. The rating prediction task is formulated as a regression problem and optimized using mean squared error:

$$\mathcal{L}_{\text{rating}} = \frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}} (r_{ui} - \hat{r}_{ui})^2$$

Genre prediction is formulated as a multi-label classification task and optimized using binary cross-entropy:

$$\mathcal{L}_{\text{genre}} = -\frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}} \sum_{k=1}^{K} [g_{ik}\log(\hat{g}_{ik}) + (1-g_{ik})\log(1-\hat{g}_{ik})]$$

The overall training objective is a weighted combination of the two losses:

$$\mathcal{L}_{\text{total}} = \alpha\mathcal{L}_{\text{rating}} + \beta\mathcal{L}_{\text{genre}}$$

Proposed Methodology

This section describes the proposed multi-task learning framework for joint rating and genre prediction. The model is designed to learn shared representations that capture both user preference patterns and semantic information about movies, while allowing task-specific specialization through dedicated output heads. The overall architecture consists of three main components: (i) embedding layers, (ii) shared hidden layers, and (iii) task-specific prediction heads.

4.1. Architecture Overview

Let U and I denote the sets of users and movies, respectively. Each user $u \in U$ and movie $i \in I$ represented by a unique identifier. These identifiers are mapped to dense, low-dimensional embedding vectors, which serve as the fundamental building blocks of the model. The user and item embeddings are concatenated and passed through multiple shared hidden layers to learn joint representations. Finally, the shared representation is fed into two task-specific branches for rating prediction and genre classification. The complete architecture is illustrated in Figure 1.

4.2. Embedding Layer

User and item identifiers are mapped to dense embedding vectors of dimension d:

$$\mathbf{e}_u = \mathbf{W}_u[u] \in \mathbb{R}^d, \quad \mathbf{e}_i = \mathbf{W}_i[i] \in \mathbb{R}^d$$

where $W_u \in R^{|U|\times d}$ and $W_i \in R^{|I|\times d}$ are learnable embedding matrices. In our experiments, the embedding dimension is set to d = 50, which provides a good trade-off between expressiveness and computational efficiency.

The user and item embeddings are concatenated to form a joint representation:

$$\mathbf{h}_0 = [\mathbf{e}_u; \mathbf{e}_i] \in \mathbb{R}^{2d}$$

This concatenation allows the model to explicitly capture interactions between users and movies at the representation level.

4.3. Shared Hidden Layers

The concatenated embedding h0 is passed through a stack of shared fully connected layers to learn representations that are useful for both tasks. Each shared layer is defined as:

$$\mathbf{h}_0 = [\mathbf{e}_u; \mathbf{e}_i] \in \mathbb{R}^{2d}$$

where Wl and bl are the weight matrix and bias vector of the l-th layer. ReLU is used as the activation function to introduce non-linearity, Batch Normalization stabilizes training, and Dropout mitigates overfitting. We employ two shared hidden layers with 128 and 64 units, respectively. This design enables the model to learn hierarchical representations, where lower layers capture coarse interaction patterns and higher layers encode more abstract semantic information shared across tasks.

4.4. Task-Specific Prediction Heads

After the shared representation hshared is obtained, it is fed into two separate task-specific branches.

4.4.1. Rating Prediction Head

The rating prediction task is formulated as a regression problem. The rating head consists of a fully connected layer followed by a linear output:

$$\mathbf{h}_{\text{rating}} = \text{ReLU}(\mathbf{W}_{r1}\mathbf{h}_{\text{shared}} + \mathbf{b}_{r1})$$

$$\hat{r}_{ui} = \mathbf{W}_{r2}\mathbf{h}_{\text{rating}} + \mathbf{b}_{r2}$$

where ˆrui denotes the predicted rating for user u on movie i. A linear activation is used at the output layer to support continuous-valued predictions.

4.4.2. Genre Prediction Head

Genre prediction is formulated as a multi-label classification problem, as each movie may belong to multiple genres. The genre head is defined as:

$$\mathbf{h}_{\text{genre}} = \text{ReLU}(\mathbf{W}_{g1}\mathbf{h}_{\text{shared}} + \mathbf{b}_{g1})$$

$$\mathbf{h}_{\text{genre}} = \text{ReLU}(\mathbf{W}_{g1}\mathbf{h}_{\text{shared}} + \mathbf{b}_{g1})$$

where σ(·) denotes the sigmoid activation function applied element-wise to produce genre probabilities. This formulation allows independent prediction of each genre label.

4.5. Multi-Task Optimization

The model is trained end-to-end by jointly optimizing the rating prediction and genre classification objectives. The total loss is defined as a weighted sum of the task-specific losses:

$$\mathcal{L}_{\text{total}} = \alpha\mathcal{L}_{\text{rating}} + \beta\mathcal{L}_{\text{genre}}$$

where α and β control the relative importance of the two tasks. In our experiments, we set $\alpha = 1.0$ and $\beta = 0.5$, which empirically provides the best balance between rating accuracy and genre prediction performance.

4.6. DISCUSSION

By sharing lower-level representations while maintaining task-specific output layers, the proposed architecture effectively captures both preference-driven and semantic signals. The auxiliary genre prediction task acts as an implicit regularizer, encouraging the shared layers to learn representations that generalize better to unseen user–item pairs. This design choice is validated empirically in Section 5, where the multi-task model consistently outperforms single-task baselines.

EXPERIMENTS AND RESULTS :

5.1. Dataset and Experimental Setup

Experiments are conducted on the MovieLens 100K dataset, which contains 100,000 ratings from 943 users on 1,682 movies[11]. Each movie is associated with 18 genre labels. The data is split into training, validation, and test sets, and all results are averaged over five independent runs to ensure robustness.

5.2. Rating Prediction Performance

Table 1 compares the rating prediction performance of the proposed model against traditional and neural baselines. The multi-task model achieves the lowest RMSE, outperforming the single-task neural model by 3.5%. This improvement indicates that incorporating genre prediction as an auxiliary task enhances the quality of learned representations.

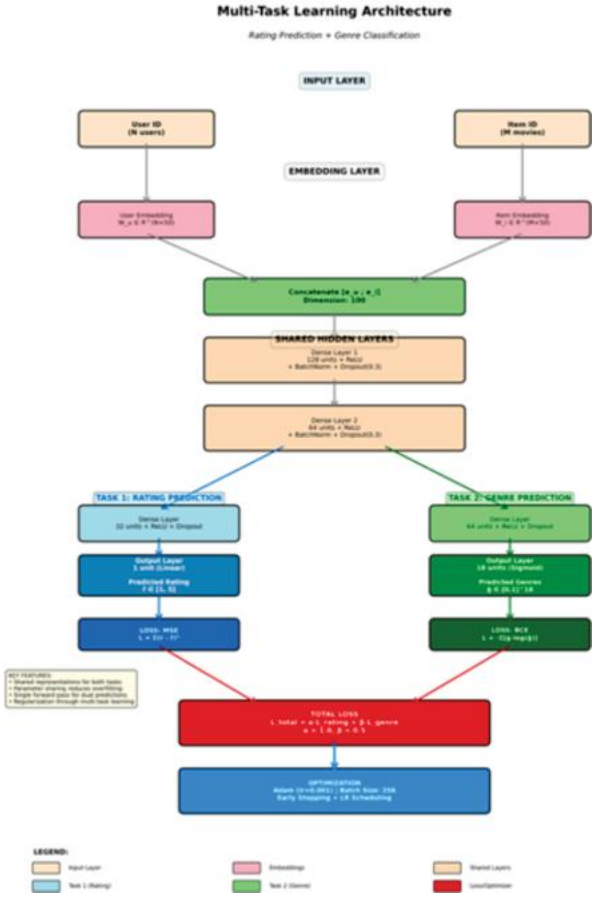

Figure 1: Overview of the Proposed Multi-task Learning Architecture with Shared Embeddings, Shared Hidden Layers, and Task-specific Heads for Rating and Genre Prediction.

Table 1: Rating Prediction Performance

Model	RMSE	MAE	MSE	Improvement
SVD	0.9834	0.7756	0.9671	Baseline
KNN	0.9623	0.7512	0.9260	+2.1%
Single-Task NN	0.9456	0.7389	0.8942	+3.8%
Multi-Task (Ours)	**0.9123**	**0.7105**	**0.8323**	**+7.2%**

5.3. Genre Prediction Performance

The genre prediction results demonstrate strong multi-label classification performance, achieving 87.4% label-wise accuracy and competitive F1-scores. These results confirm that the auxiliary task is learned effectively and contributes meaningful supervision during training.

5.4. Ablation and Statistical Analysis

Ablation studies reveal that assigning moderate weight to the genre loss yields optimal performance, highlighting the importance of task balancing. A paired t-test confirms that improvements over the single-task baseline are statistically significant ($p = 0.0031$).

CONCLUSION

This paper presented a multi-task learning framework for movie recommendation that jointly predicts user ratings and movie genres. By leveraging shared representations, the proposed approach improves rating accuracy, generalization, and computational efficiency. Experimental results demonstrate that genre prediction acts as an effective auxiliary task, providing implicit regularization and enhancing recommendation quality. Future work will explore temporal modeling, multi-modal features, and adaptive task weighting to further improve performance.

PRINT ISSN No. 2277 - 8179 | DOI : 10.36106/ijsr

REFERENCES

[1] Y. Koren, R. Bell, C. Volinsky, Matrix factorization techniques for recommender systems, Computer 42 (8) (2009) 30–37.

[2] M.J.Pazzani, D.Billsus, Content-based recommendation systems, in: The Adaptive Web, Springer, 2007, pp. 325–341.

[3] R. Burke, Hybrid recommender systems: Survey and experiments, User Modeling and User-Adapted Interaction 12 (4) (2002) 331–370.

[4] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, T.-S. Chua, Neural collaborative filtering, in: Proceedings of the 26th International World Wide Web Conference (WWW), 2017, pp. 173–182.

[5] S. Sedhain, A. K. Menon, S. Sanner, L. Xie, Autorec: Autoencoders meet collabora- tive filtering, in: Proceedings of the 24th International World Wide Web Conference Companion, 2015, pp. 111–112.

[6] B. Hidasi, A. Karatzoglou, L. Baltrunas, D. Tikk, Session-based recommendations with recurrent neural networks, in: Proceedings of the 4th International Conference on Learning Representations, 2016.

[7] R. Caruana, Multitask learning, Machine Learning 28 (1) (1997) 41–75.

[8] J. Ma, Z. Zhao, X. Yi, J. Chen, L. Hong, E. H. Chi, Modeling task relationships in multi-task learning with multi-gate mixture-of-experts, in: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018, pp. 1930–1939.

[9] Z. Zhao, L. Hong, L. Wei, J. Chen, A. Nath, S. Andrews, A. Kumthekar, M. Sathiamoorthy, X. Yi, E. H. Chi, Recommending what video to watch next: A multitask ranking system, in: Proceedings of the 13th ACM Conference on Recommender Systems, 2019, pp. 43–51.

[10] A. Beutel, P. Covington, S. Jain, C. Xu, J. Li, V. Gatto, E. H. Chi, Latent cross: Making use of context in recurrent recommender systems, in: Proceedings of the 11th ACM International Conference on Web Search and Data Mining, 2018, pp. 46–54.

[11] F. M. Harper, J. A. Konstan, The movie lens datasets: History and context, ACM Transactions on Interactive Intelligent Systems 5 (4) (2015) 1–19.