HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY
SCHOOL OF ELECTRICAL AND ELECTRONIC ENGINEERING

PROJECT II

TITLE: A Neural Collaborative Filtering Approach for Personalized Book Recommendation Using Implicit Feedback

TRUONG TUAN HUNG
hung.tt224284@sis.hust.edu.vn

Major: Digital Communication and Multimedia Engineering

Instructor: **Ph.D. Duong Tan Nghia**

Hanoi, 02/2026

TABLE OF CONTENTS

ABSTRACT

Recommender systems play an essential role in helping users navigate large-scale digital platforms by providing personalized suggestions. In the context of online book platforms, the abundance of available books makes it increasingly difficult for readers to discover content that aligns with their preferences. This project presents a Book Recommendation System based on implicit feedback, where user preferences are inferred from observed interactions rather than explicit ratings.

The proposed system employs Neural Collaborative Filtering (NeuMF), a deep learning–based recommendation model that integrates Generalized Matrix Factorization (GMF) and a Multi-Layer Perceptron (MLP) to capture both linear and non-linear user–item interaction patterns. Explicit rating data are first transformed into implicit feedback by considering high ratings as positive interactions, followed by negative sampling to generate training instances for binary classification. The model is trained using binary cross-entropy loss and optimized with the Adam optimizer.

To evaluate the effectiveness of the recommendation system, top-K ranking metrics, including Hit Ratio@10 (HR@10) and Normalized Discounted Cumulative Gain@10 (NDCG@10), are adopted. Experimental results demonstrate that the proposed NeuMF-based model achieves a Hit Ratio@10 of 0.66 and an NDCG@10 of 0.40, indicating that the system is capable of retrieving relevant books for users with a reasonable ranking quality. Furthermore, a recommendation module is implemented to generate personalized top-N book suggestions for individual users.

Overall, this project highlights the applicability of neural collaborative filtering techniques for book recommendation tasks using implicit feedback and provides a foundation for future improvements, such as hybrid recommendation models and more advanced negative sampling strategies.

CHAPTER 1: INTRODUCTION

With the rapid expansion of digital content platforms, users are increasingly confronted with information overload. In online book platforms, thousands of new titles are continuously added, making it difficult for readers to efficiently discover books that match their interests and reading preferences. As a result, recommender systems have become a critical component in modern information systems, enabling personalized content delivery and improving user experience by filtering and ranking items based on individual preferences.

Book recommendation poses several unique challenges compared to other recommendation domains. User–book interaction data are often sparse, as most users interact with only a limited number of books. Moreover, explicit feedback such as numerical ratings is not always reliable or readily available, since many users prefer passive interactions (e.g., browsing or purchasing) over providing ratings. These limitations have led to increased interest in implicit feedback–based recommendation, where user preferences are inferred from observed behaviors rather than explicit judgments.

Traditional collaborative filtering approaches, including memory-based methods and matrix factorization techniques, have demonstrated effectiveness in recommendation tasks by modeling user–item interactions. However, these methods typically rely on linear interaction assumptions, which may limit their ability to capture complex and non-linear user preference patterns. Recent advances in deep learning have introduced neural-based recommender systems, which enhance representation learning and interaction modeling through neural network architectures.

In this project, we propose a Book Recommendation System based on Neural Collaborative Filtering (NeuMF), a deep learning framework that combines the strengths of Generalized Matrix Factorization (GMF) and Multi-Layer Perceptrons (MLP). By integrating linear and non-linear interaction modeling, the NeuMF architecture provides greater expressive power than traditional matrix factorization methods. The system is designed for an implicit feedback setting, where explicit ratings are transformed into binary preference signals, and negative samples are generated to facilitate supervised learning.

The objectives of this project are threefold. First, we aim to preprocess and transform book rating data into an implicit feedback dataset suitable for neural recommendation models. Second, we implement and train a NeuMF-based model to learn personalized user–book representations. Finally, we evaluate the effectiveness of the proposed system using top-K ranking metrics, including Hit Ratio@10 and Normalized Discounted Cumulative Gain@10, which are more appropriate for recommendation tasks than traditional classification accuracy.

CHAPTER 2: RELATED WORKS

Research on recommender systems has evolved significantly over the past two decades, progressing from traditional collaborative filtering techniques to advanced neural network–based models. This section reviews three representative studies that are closely related to the methodology adopted in this project, with a particular focus on implicit feedback modeling and neural collaborative filtering.

2.1 Neural Collaborative Filtering (He et al., 2017)

He et al. (2017) [1] introduced Neural Collaborative Filtering (NCF), a pioneering framework that replaces the inner product used in traditional matrix factorization with neural network architectures. Within this framework, the authors proposed NeuMF, which integrates two complementary components: Generalized Matrix Factorization (GMF) for modeling linear user–item interactions and a Multi-Layer Perceptron (MLP) for capturing non-linear interaction patterns.

A key contribution of this work lies in its flexibility to learn complex user preference structures directly from data, without relying on manually designed interaction functions. Furthermore, the authors demonstrated that NeuMF consistently outperforms conventional matrix factorization methods and standalone neural models on several benchmark datasets under an implicit feedback setting.

2.2 Collaborative Filtering for Implicit Feedback Datasets (Hu et al., 2008)

Hu et al. (2008) [2] provided one of the earliest and most influential formulations of collaborative filtering for implicit feedback datasets. Unlike explicit feedback scenarios where user preferences are directly expressed through ratings, this study emphasized that implicit feedback data only indicate whether an interaction has occurred, not the degree of user preference. To address this challenge, the authors proposed a confidence-based matrix factorization framework that assigns varying confidence levels to observed and unobserved interactions.

This work is highly relevant to the present project as it establishes the theoretical foundation for treating missing interactions as potential negative signals rather than neutral data points. The concept that unobserved interactions can be interpreted as weak negative feedback motivates the use of negative sampling in modern neural recommender systems.

2.3 Deep Matrix Factorization Models for Recommender Systems (Xue et al., 2017)

Xue et al. (2017) [3] explored deep matrix factorization models, extending traditional latent factor approaches by stacking multiple non-linear layers to learn hierarchical representations of user–item interactions. Their study demonstrated that deep architectures can significantly enhance representation capacity compared to shallow matrix factorization, particularly in sparse recommendation settings.

Unlike NeuMF, which separates linear and non-linear interaction modeling into distinct components, the approach proposed by Xue et al. focuses on learning deep latent representations through successive transformations. Experimental results showed improved recommendation accuracy over conventional collaborative filtering methods, highlighting the advantages of deep learning in recommender systems.

2.4 Summary

In summary, Hu et al. (2008) laid the theoretical groundwork for implicit feedback modeling, while Xue et al. (2017) demonstrated the effectiveness of deep learning in extending matrix factorization methods. Building upon these foundations, He et al. (2017) proposed NeuMF, which unifies linear and non-linear interaction modeling in a single neural framework. Motivated by these studies, this project employs the NeuMF model to construct an effective book recommendation system under an implicit feedback setting.

CHAPTER 3: METHODOLOGY

This section describes the methodology adopted to build the proposed book recommendation system. The overall pipeline consists of data preprocessing, implicit feedback modeling, negative sampling, neural collaborative filtering model construction, and training procedures.

3.1 Problem Formulation

The book recommendation task is formulated as a top-N recommendation problem under an implicit feedback setting. Given a set of users U, a set of books I, and observed user–book interactions, the objective is to predict a ranked list of books that a user is likely to prefer.

Instead of predicting explicit rating values, the model estimates the probability that a user will interact positively with a given book. This formulation aligns with real-world recommendation scenarios, where explicit feedback is often unavailable or unreliable.

3.2 Implicit Feedback Construction

The original dataset contains explicit numerical ratings provided by users. To adapt the data to an implicit feedback setting, explicit ratings are transformed into binary preference signals:

- Ratings greater than or equal to a predefined threshold are treated as positive interactions.

- All other interactions are discarded and not directly modeled.

To reduce data sparsity and noise, users and books with very few interactions are filtered out. This preprocessing step improves training stability and enhances the quality of learned representations.

3.3 User and Item Encoding

Neural recommender models require user and item identifiers to be represented as continuous indices. Therefore, each unique user ID and book ID is mapped to a sequential integer index. These indices serve as inputs to the embedding layers of the neural network.

This encoding allows the model to learn latent representations of users and books through trainable embedding matrices.

3.4 Data Splitting Strategy

To ensure a realistic evaluation scenario, the dataset is split on a per-user basis:

- One interaction per user is reserved for testing.
- One interaction per user is used for validation, when available.
- The remaining interactions are used for training.

This strategy ensures that every user appears in the training set and that evaluation reflects the model's ability to recommend unseen items to known users, which is critical for recommendation tasks.

3.5 Negative Sampling

In implicit feedback datasets, only positive interactions are observed. However, supervised learning requires both positive and negative samples. To address this issue, negative sampling is applied during training.

For each observed positive interaction (u,i), a fixed number of negative samples are generated by randomly selecting books that the user has not interacted with. These sampled interactions are assigned a label of zero.

This process results in a balanced training dataset consisting of both positive and negative user–book pairs. Negative sampling allows the model to learn discriminative patterns between preferred and non-preferred items while keeping the dataset size manageable.

3.6 Neural Collaborative Filtering Model

The proposed recommendation system is based on the Neural Collaborative Filtering (NeuMF) framework, which combines linear and non-linear interaction modeling.

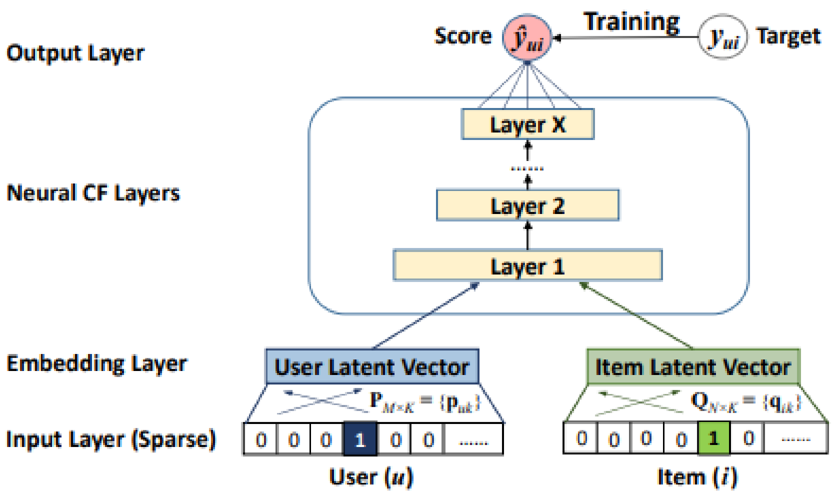

Figure 1: Neural collaborative filtering framework

3.6.1 Generalized Matrix Factorization (GMF)

The GMF component learns latent embeddings for users and books and models their interaction through element-wise multiplication. This structure is analogous to traditional matrix factorization but implemented within a neural framework, allowing parameters to be learned via gradient-based optimization.

3.6.2 Multi-Layer Perceptron (MLP)

The MLP component concatenates user and book embeddings and passes them through multiple fully connected layers with non-linear activation functions. This component captures complex and non-linear interaction patterns that cannot be modeled by linear factorization alone.

Neural Collaborative Filtering (NCF) models user–item interactions through two separate pathways for users and items. Although a straightforward approach is to concatenate the latent features from these two pathways, simple concatenation alone fails to capture meaningful interactions between user and item representations, which is insufficient for collaborative filtering.

To overcome this limitation, the authors propose applying a Multi-Layer Perceptron (MLP) on top of the concatenated user and item embeddings. This allows the model to learn flexible and non-linear interaction functions between user and item latent features, in contrast to Generalized Matrix Factorization (GMF), which relies on a fixed element-wise product.

The MLP consists of multiple hidden layers with learnable weights and biases, culminating in a sigmoid output to predict user preference. Among various activation functions, ReLU is preferred due to its non-saturating property, suitability for sparse data, and reduced risk of overfitting compared to sigmoid and tanh. Additionally, the network adopts a tower structure, where each successive layer has fewer neurons, enabling higher layers to learn more abstract feature representation.

3.6.3 NeuMF Fusion

The outputs of the GMF and MLP components are concatenated to form a unified interaction representation. This fused representation is passed through a final sigmoid layer to estimate the probability that a user prefers a given book.

By integrating both components, NeuMF leverages the strengths of linear interaction modeling and deep non-linear representation learning.

$$\mathbf{h} = \begin{bmatrix} \alpha \mathbf{h}^{GMF} \\ (1-\alpha)\mathbf{h}^{MLP} \end{bmatrix}$$

3.7 Model Training

The model is trained using binary cross-entropy loss, which is suitable for binary preference prediction tasks. Optimization is performed using the Adam optimizer with mini-batch gradient descent.

Training is conducted for a fixed number of epochs with a large batch size to improve computational efficiency. Validation data are used to monitor training progress and mitigate overfitting.

3.8 Recommendation Generation

After training, the model generates personalized recommendations by predicting preference scores for all candidate books that a user has not previously interacted with. These books are ranked according to their predicted scores, and the top-N items are returned as recommendations.

This procedure enables the system to produce interpretable and personalized book suggestions for individual users.

3.9 Methodology Overview

In summary, the proposed methodology combines implicit feedback modeling, negative sampling, and neural collaborative filtering to construct an effective book recommendation system. The use of NeuMF allows the model to capture both simple and complex user–item interaction patterns, providing a strong foundation for accurate top-N recommendations.

CHAPTER 4: EXPERIMENTS

4.1 Datasets

4.1.1 books.csv – Book Metadata

The books.csv file contains descriptive information about the books available in the system. Each row corresponds to a unique book, identified by a distinct book ID. This dataset serves as the **item catalog** for the recommendation system.

Key attributes include:

- **book_id**: A unique identifier for each book, used as the item index in the recommendation model.
- **title**: The title of the book.
- **authors**: The author(s) of the book.
- *(Other metadata columns may be present but are not directly used in model training.)*

In this project, the metadata is primarily used for:

- Mapping predicted item indices back to human-readable book titles and authors.
- Interpreting and visualizing recommendation results.

Although content-based features such as genres or descriptions are not incorporated into the current model, the availability of metadata allows for future extensions toward hybrid recommendation systems that combine collaborative filtering with content information.

4.1.2 ratings.csv – User–Book Interactions

The ratings.csv file records explicit user ratings for books. Each row represents a single interaction between a user and a book.

Key attributes include:

- **user_id**: A unique identifier for each user.
- **book_id**: The identifier of the book being rated.
- **rating**: An integer value representing the user's rating for the book, typically on a scale from 1 to 5.

This dataset captures users' explicit preferences and forms the basis for constructing the interaction matrix used in training the recommendation model.

4.1.3. Transformation to Implicit Feedback

To align with real-world recommendation scenarios and the NeuMF framework, the explicit rating data are transformed into an implicit feedback dataset.
The transformation process is defined as follows:

- Ratings greater than or equal to a predefined threshold (e.g., rating ≥ 4) are treated as positive interactions, indicating that a user likes a book.
- Ratings below the threshold are discarded and not explicitly modeled.
- Each retained interaction is assigned a binary label of 1, representing positive feedback.

This conversion allows the recommendation task to be formulated as a binary preference prediction problem, which is more suitable for neural collaborative filtering models.

4.2 Libraries

- **Pandas (pandas)**
  Used for data loading, preprocessing, filtering user–book interactions, and organizing recommendation results in tabular form.
- **NumPy (numpy)**
  Provides efficient numerical computation and array operations for indexing, vector manipulation, and batch prediction.
- **TensorFlow (tensorflow)**
  Serves as the core deep learning framework for training and optimizing the Neural Collaborative Filtering (NeuMF) model.
- **Keras (tensorflow.keras)**
  Offers a high-level API for building, compiling, and training neural network models, enabling efficient implementation of the NeuMF architecture.
- **Keras Layers (tensorflow.keras.layers)**
  Supplies essential neural network components such as Embedding, Dense, Multiply, and Concatenate layers for modeling user–item interactions.
- **Random (random)**
  Used to perform negative sampling by randomly selecting non-interacted books for implicit feedback training.
- **Matplotlib (matplotlib.pyplot)**
  Utilized to visualize training dynamics, including accuracy and loss curves across epochs.

The project leverages Pandas and NumPy for data preprocessing, TensorFlow and Keras for implementing the NeuMF deep learning model, Random for negative sampling in the implicit feedback setting, and Matplotlib for visualizing model training behavior.

4.3 Experimental Setup

4.3.1 Data Splitting Strategy

To simulate a realistic recommendation scenario, the dataset is split on a per-user basis:

- One positive interaction per user is reserved for testing.
- One positive interaction per user is used for validation when sufficient data are available.
- All remaining interactions are used for training.

This strategy ensures that evaluation is performed on unseen items for known users, which is consistent with the goal of personalized recommendation.

4.3.2 Model Configuration

The NeuMF model consists of two parallel components:

- Generalized Matrix Factorization (GMF), which models linear user–item interactions using element-wise multiplication.
- Multi-Layer Perceptron (MLP), which models non-linear interactions using multiple fully connected layers.

Key hyperparameters used in the experiments are summarized as follows:

- Embedding dimension (GMF and MLP): 16

- MLP architecture: 64 – 32 – 16
- Loss function: Binary cross-entropy
- Optimizer: Adam
- Batch size: 2048
- Number of epochs: 10

These settings are chosen to balance model expressiveness and training efficiency.

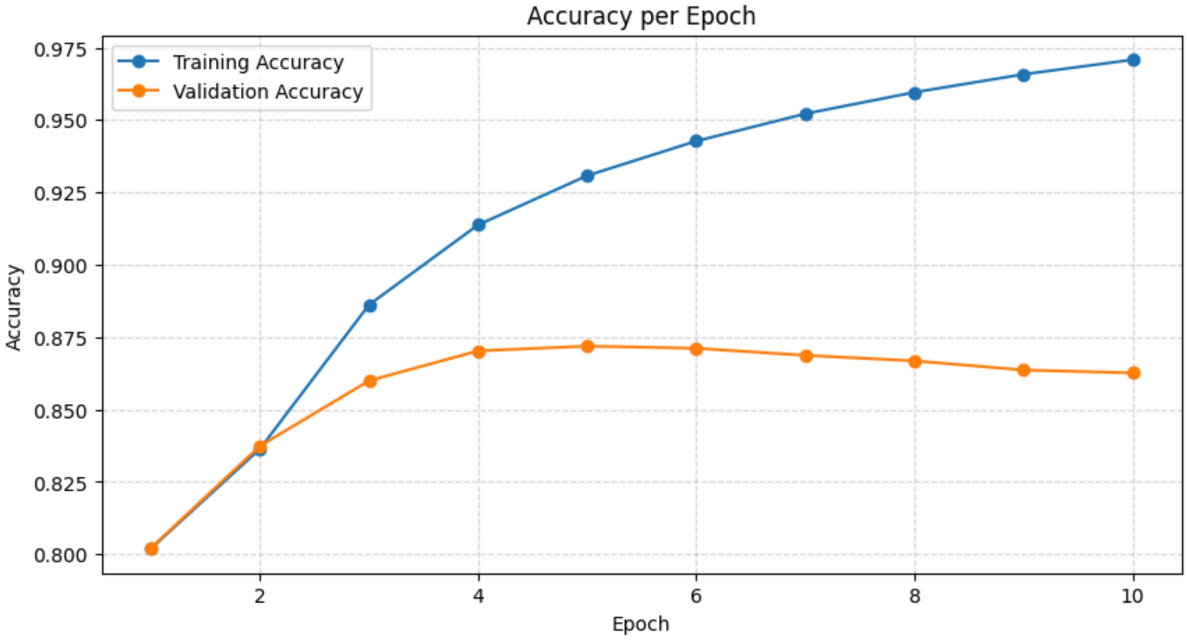

Figure 2: Training and Validation Accuracy of the NeuMF Model Across Epochs

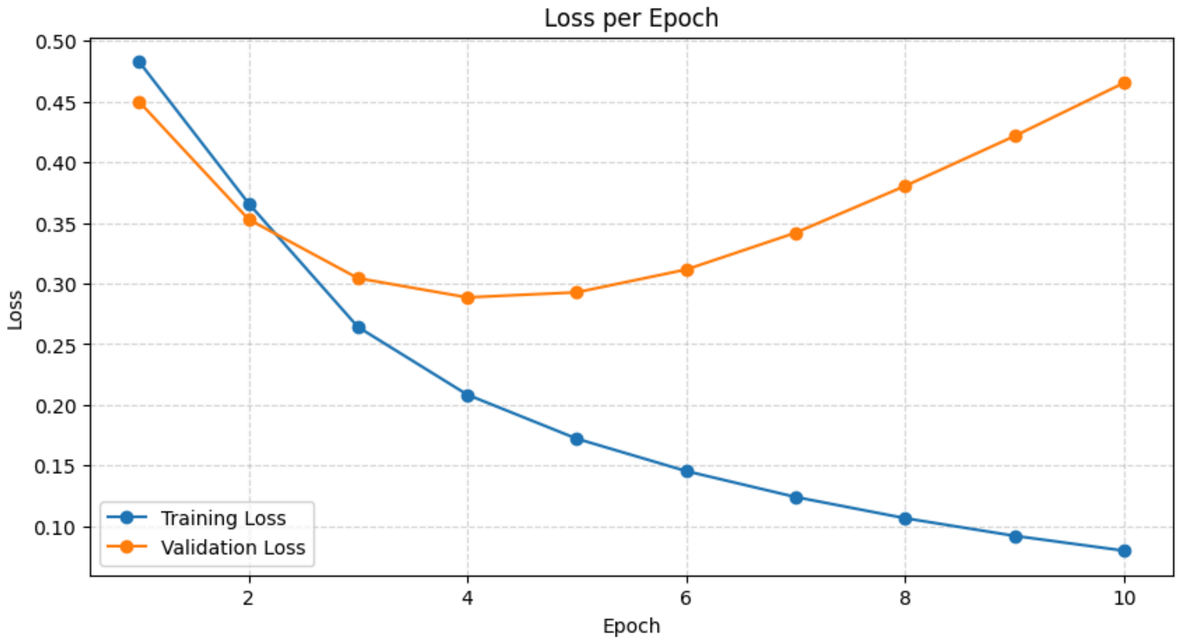

Figure 3: Training and Validation Loss of the NeuMF Model Across Epochs

4.4 Evaluation Protocol

4.4.1 Top-K Recommendation Setting

The model is evaluated using a **top-K recommendation protocol**. For each user in the test set:

- One ground-truth positive item is selected.
- A set of negative candidate items is randomly sampled from books the user has not interacted with.
- The model predicts preference scores for all candidate items.
- Items are ranked according to predicted scores.
- The top-K items are used for evaluation.

This evaluation strategy reflects the practical recommendation scenario in which the system must rank a small number of relevant items among many irrelevant ones.

4.4.2 Evaluation Metrics

Two widely used ranking-based metrics are adopted:

a. Hit Ratio@10 (HR@10)

Hit Ratio@10 measures whether the ground-truth positive item appears within the top 10 recommended items. It reflects the model's ability to retrieve relevant items.
With each user u, define:

$$\mathrm{HR@K}(u) = \begin{cases} 1, & \text{if } i_u^+ \in \mathrm{TopK}(u) \\ 0, & \text{otherwise} \end{cases}$$

Hit Ratio@K average on all user:

$$\mathrm{HR@K} = \frac{1}{|U|} \sum_{u \in U} \mathrm{HR@K}(u)$$

b. Normalized Discounted Cumulative Gain@10 (NDCG@10)

NDCG@10 considers both the presence and the ranking position of the relevant item. Higher scores are assigned when the correct item is ranked closer to the top, making NDCG sensitive to ranking quality.
These metrics are more appropriate than classification accuracy for recommendation tasks, as they directly evaluate ranking performance.
With each user u, consider item is corrects having rank = r_u (start from 0):

$$\mathrm{NDCG@K}(u) = \begin{cases} \frac{1}{\log_2(r_u+2)}, & \text{if } r_u < K \\ 0, & \text{otherwise} \end{cases}$$

NDCG@K average on all user:

$$\text{NDCG@K} = \frac{1}{|U|} \sum_{u \in U} \text{NDCG@K}(u)$$

4.5 Quantitative Results

Table 4.1 summarizes the experimental results obtained using the proposed NeuMF model.

Metric	Value
Hit Ratio@10	0.6596
NDCG@10	0.4034

The results indicate that the model successfully retrieves relevant books for users in approximately two-thirds of test cases. Moreover, the NDCG score suggests that relevant items are generally ranked within the top positions, although not always at the very top of the recommendation list.

4.6 Discussion

The relatively high Hit Ratio@10 demonstrates the effectiveness of the NeuMF model in identifying relevant books under an implicit feedback setting. The moderate NDCG@10 value indicates that while relevant items are often retrieved, there is room for improvement in ranking them closer to the top positions.

Several factors may influence these results, including the random negative sampling strategy, the limited number of training epochs, and the absence of additional content-based features. Nevertheless, the observed performance confirms the suitability of neural collaborative filtering for book recommendation tasks.

4.7 Qualitative Recommendation Analysis

In addition to quantitative evaluation, the trained model is used to generate personalized top-N book recommendations for individual users. By excluding books that users have already interacted with, the system produces novel recommendations based on learned user–item representations.

The qualitative results show that the recommended books are coherent with users' inferred preferences, indicating that the model captures meaningful patterns in user reading behavior.

```
Số sách user đã thích: 19
Số sách có thể recommend: 9974
```

	book_id	title	authors	score
0	5552	QED: The Strange Theory of Light and Matter	Richard Feynman	0.999227
1	7613	Animal Farm	George Orwell	0.994021
2	6537	From Potter's Field (Kay Scarpetta, #6)	Patricia Cornwell	0.993648
3	6749	Oblivion	David Foster Wallace	0.990118
4	5174	Fall on Your Knees	Ann-Marie MacDonald	0.988349
5	8921	The Hound of the Baskervilles	Arthur Conan Doyle, Anne Perry	0.987261
6	5452	Girls in Pants: The Third Summer of the Sister...	Ann Brashares	0.984962
7	7672	Congo	Michael Crichton	0.983358
8	2743	The Lost Boy (Dave Pelzer #2)	Dave Pelzer	0.981985
9	6671	The Wonderful Story of Henry Sugar and Six More	Roald Dahl	0.981071
10	5526	Dear John	Nicholas Sparks	0.979613

Figure 4: Example of Top-10 Personalized Book Recommendations Generated by the NeuMF Model for a Sample User

The figure illustrates an example of personalized book recommendations generated by the Neural Collaborative Filtering (NeuMF) model for a sample user. The user has 19 previously liked books, and recommendations are produced from a candidate set of 9,974 unseen books**.**

The table presents the top-10 recommended books, including the *book_id*, *title*, *authors*, and the predicted *score* for each item. The score represents the estimated preference probability output by the NeuMF model. All recommended books are ranked in descending order of their predicted scores, indicating the model's confidence in each recommendation.

The recommended list contains books from diverse authors and genres, demonstrating the model's ability to capture latent user preferences from implicit feedback data. This qualitative result complements the quantitative evaluation metrics and highlights the practical effectiveness of the proposed recommendation system.

4.8 Summary

In summary, the experiments demonstrate that the proposed NeuMF-based recommendation system performs effectively in an implicit feedback scenario. The model achieves strong retrieval performance as measured by Hit Ratio@10 and produces reasonably well-ranked recommendations according to NDCG@10. These results validate the design choices of the proposed system and provide a solid baseline for future improvements, such as advanced negative sampling strategies, deeper neural architectures, or hybrid recommendation approaches.

CHAPTER 5: CONCLUSIONS

In this project, we developed and evaluated a personalized Book Recommendation System based on the Neural Collaborative Filtering (NeuMF) framework. By transforming explicit user ratings into implicit feedback, the recommendation task was formulated as a top-N ranking problem, which better reflects real-world recommendation scenarios where explicit feedback is limited or unreliable.

The proposed system combines the strengths of Generalized Matrix Factorization (GMF) and Multi-Layer Perceptron (MLP) to model both linear and non-linear user–item interactions. Through negative sampling and neural representation learning, the model effectively captures latent user preferences from sparse interaction data. Experimental results using Hit Ratio@10 and NDCG@10 demonstrate that the NeuMF model is capable of retrieving relevant books for users with reasonable ranking quality. In addition, qualitative analysis of personalized recommendations shows that the system produces meaningful and diverse book suggestions aligned with users' inferred interests.

Despite its effectiveness, the current system has several limitations. The use of random negative sampling may not fully reflect challenging real-world negative examples, and the model relies solely on collaborative signals without incorporating content-based information such as book genres or descriptions. Furthermore, the cold-start problem for new users and new books remains an open challenge.

Overall, this project confirms the suitability of neural collaborative filtering techniques for book recommendation under an implicit feedback setting. The implemented system provides a solid baseline that can be further enhanced in future work by integrating content features, adopting more advanced negative sampling strategies, or extending the model to hybrid and context-aware recommendation frameworks.

REFERENCES

[1] He, X., Liao, L., Zhang, H., Nie, L., Hu, X., & Chua, T.-S. (2017). *Neural Collaborative Filtering*.

[2] Hu, Y., Koren, Y., & Volinsky, C. (2008). *Collaborative Filtering for Implicit Feedback Datasets*.

[3] Xue, H.-J., Dai, X., Zhang, J., Huang, S., & Chen, J. (2017). *Deep Matrix Factorization Models for Recommender Systems*.