DeepChoice: Enhancing E-Commerce with Hybrid AI Recommendation Systems

Carolina Li, Yaxin Yu, Xuefeng Li

Advisor Prof. Yifan Hu

Northeastern University

1. Background

Driven by improvements in artificial intelligence and machine learning, the digital marketplace is quickly changing. Tech giant like Amazon, TikTok Shop, and Google Shopping has spent leading efforts in personalization, making personalized shopping a must-have. Shoppers now expect highly customized shopping experiences that not only understand their preferences but also predict their needs.

Collaborative filtering and content-based filtering are two common approaches used in recommendation systems. A Content-based recommendation system uses information about the recommended item(item information from the specific user's past behavior), while a collaborative system uses user behaviour data to find a group of similar users. Content-based filtering helps to identify products with similar attributes or content, while collaborative filtering helps to identify products that are popular or liked by users with similar interests.

2. Motivations & Objectives

One of the limitations for traditional recommendation systems is their inability to understand the subtleties of user preferences expressed in textual reviews, which contain rich contextual information. Meanwhile, transformer-based models(including LLMs) excel at deeply analyzing this textual data.

This project is motivated by the potential of transformer-based models, and aims to explore whether transformer-based models can outperform classic filtering methods in item recommendations by leveraging advanced language understanding to capture nuanced sentiments and preferences and provide more personalized and accurate predictions. We will compare the performance of collaborative filtering and content-based filtering models using user review histories and item information, and also compare the performance of classic and transformer-based models.

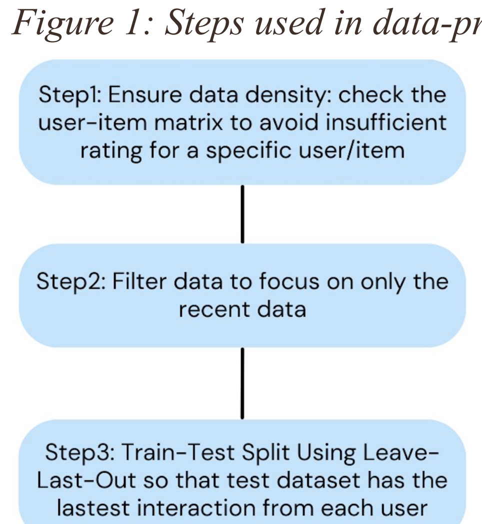
3. Dataset & Preprocessing

We used the large-scale Amazon Reviews dataset, collected in 2023 by McAuley Lab. It consists of both user and item data from 1996 ro 2023, with over 500M reviews, 32 subcategories. Some of the data fields shown as below:

RATING	PARENT_ASIN	USER_ID	TIMESTAMP	HISTORY	TEXT
Rating of the product (from 1.0 to 5.0)	Parent ID of the product	ID of the reviewer	Time of the review (unix time)	parent_asin of items that this user rated before	Text body of the user review.

PARENT_ASIN	TITLE	MAIN_CATEGORY	AVERAGE_RATING	RATING_NUMBER	DESCRIPTION
Parent ID of the product.	Name of the product.	Main category (i.e., domain) of the product.	Average rating of the product.	Number of ratings in the product.	Description of the product.

Figure 1: Steps used in data-preprocessing



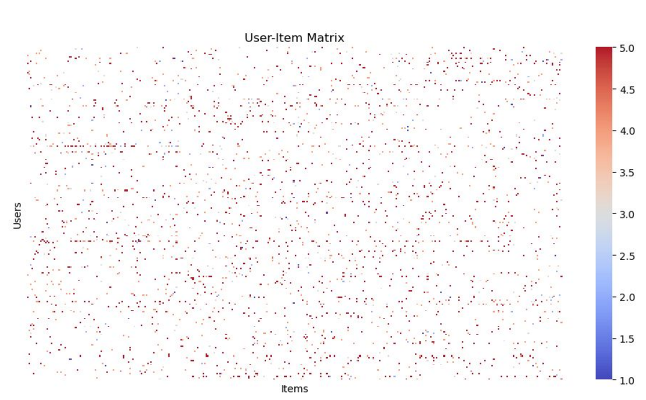


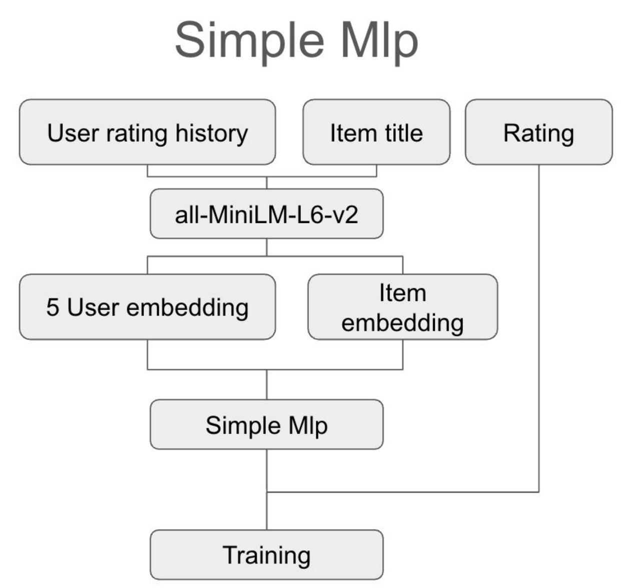
Figure 2: User-Item matrix (an example of dense matrix from beauty subcategory)

4. Methods

4.1 Neural Network:

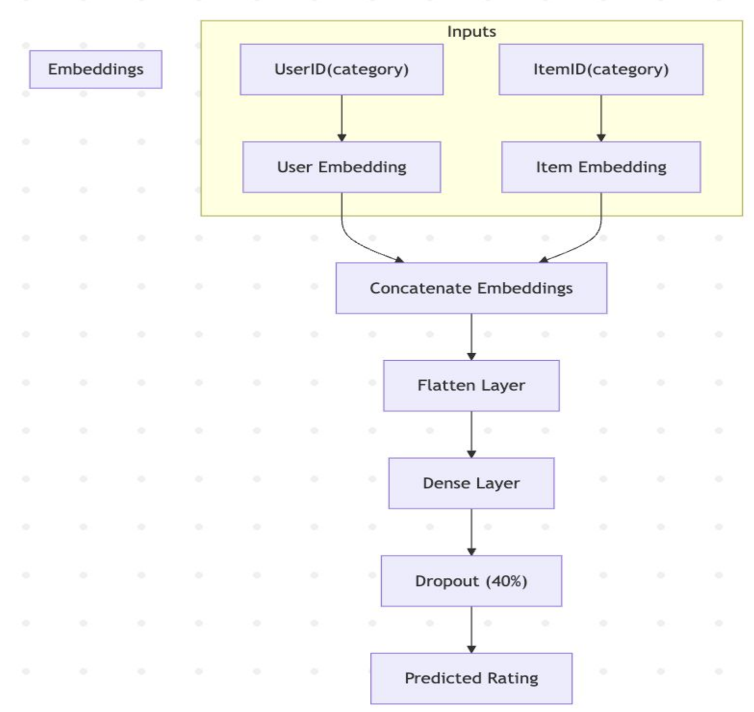
1. **User history rating as features**

Figure 2: Pipeline of Neural Network



2. **User-Item category encoding**

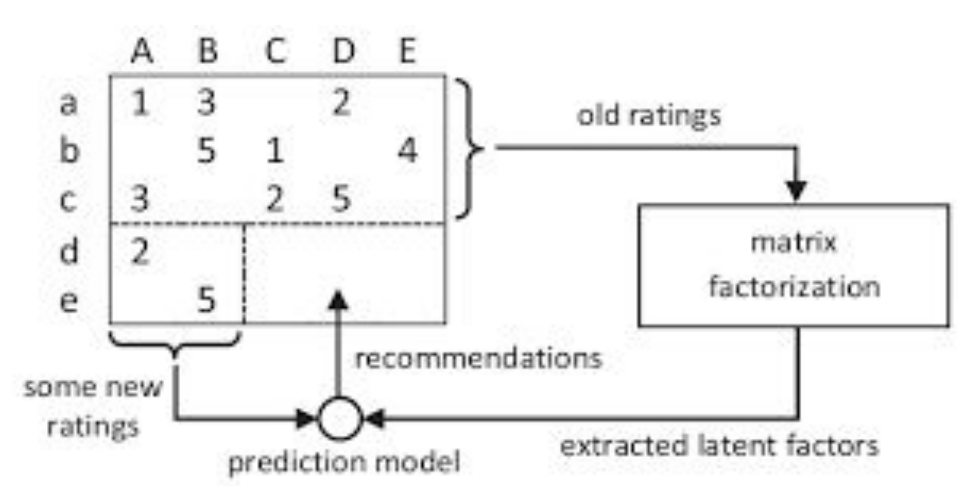
Figure 3: Pipeline of Neural Network



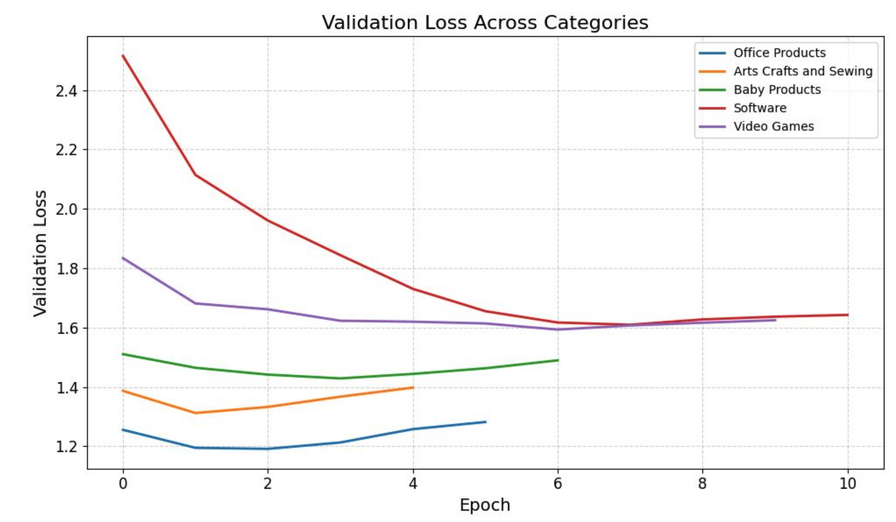
4.2 Baseline Model: Collaborative filtering

- **Latent factor model - SVD**: approximates the user-item rating matrix by decomposing it into three smaller matrices reflecting hidden features underlying the interactions.
- **Neighborhood models -KNN With Means**: Memory Based calculates similarities between users or items based on their ratings histories.

Figure 4: Steps used in SVD



5. Experimental Results



Model Performance Comparison Across Product Categories

Product Category	NN Top-5 Recommendations	LightFM Top-5 Recommendations	Validation RMSE
Office Products	16787, 8701, 52393, 5799, 43476	562, 4274, 41, 105, 47	1.1397
Arts, Crafts & Sewing	68183, 52697, 46901, 26864, 69824	707, 4945, 7757, 1314, 6072	1.2205
Baby Products	7006, 9609, 14078, 20273, 600	768, 2794, 682, 28, 570	1.295
Software	4693, 5258, 777, 2842, 2625	51, 73, 239, 41, 69	1.8148
Video Games	10908, 10256, 10727, 94, 4961	195, 291, 1667, 1737, 1428	1.4877

Result Analysis:

Office Products
- Best performance (RMSE 1.14); NN showed stable, low validation loss.
- Indicates strong generalization and predictable user behavior.

Arts, Crafts & Sewing / Baby Products
- RMSEs: 1.22, 1.30; early flattening of validation loss.
- Suggest consistent rating patterns across users.

Software
- Highest RMSE (1.81) and sharpest loss drop.
- Reflects rating variability or data sparsity.

Video Games
- Moderate RMSE (1.49); validation loss plateaued early.
- LightFM gave more diverse recommendations in sparse context.

Overall
- NN outperformed LightFM in most categories.
- Validation loss stabilized within 5–6 epochs, supporting early stopping.
- Model effectiveness varies by category density and user behavior.

4.3 Baseline Model: LightFM

LightFM is a hybrid collaborative filtering method designed for recommendation tasks. It learns embeddings for users and items simultaneously based on their interactions (ratings).

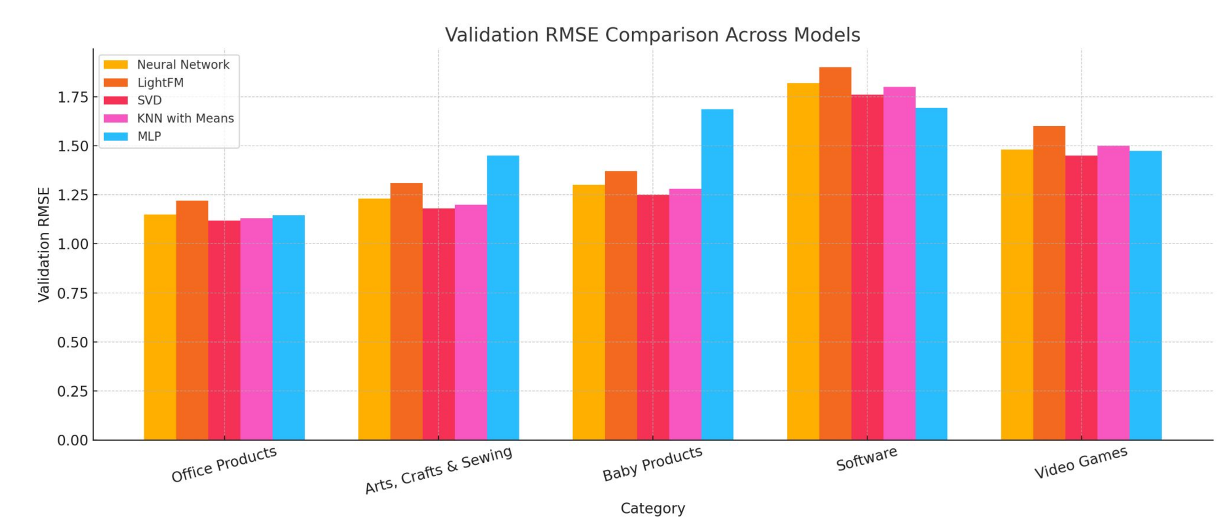
- Encoding Interactions: Converting user-item ratings into an interaction matrix to reflect user preferences explicitly.
- Embedding Learning: Simultaneously learning latent features (embeddings) for both users and items, capturing implicit relationships.
- Pairwise Ranking Optimization: Training with Weighted Approximate-Rank Pairwise (WARP) loss, effectively prioritizing relevant recommendations at the top of the ranked list.

6. Key Learnings and Discussions

Our project compares traditional collaborative filtering methods (SVD, KNN) with deep learning approaches, including a one-hot encoded neural network and a transformer-based model using BERT embeddings. We observed that collaborative filtering consistently achieved lower Mean Squared Error (MSE), indicating superior performance in minimizing prediction error across the dataset. This advantage likely stems from its ability to directly model user-item interactions without the need for extensive feature engineering. While deep learning models offer flexibility and can incorporate rich content features such as product descriptions, they require significantly more data and tuning to outperform simpler baselines. These findings highlight the practical trade-offs between model complexity, interpretability, and performance. In real world settings with structured data and sparse content, traditional collaborative methods remain competitive, whereas content-rich environments may benefit from more advanced models like transformers.

7. Future Work

To further improve recommendation performance, future work will explore refined deep learning architectures and hyperparameter tuning to better capture complex user-item interactions. We also plan to investigate reinforce learning ,to optimize product recommendations based on browsing history, aiming to enhance both accuracy and personalization. Expanding to larger datasets and new product domains will help evaluate the models' scalability and adaptability across diverse user behaviors and item types.



Result Analysis:

We compared a neural network leveraging BERT-based item embeddings against LightFM, SVD, and KNN approaches. The neural network model consistently outperforms LightFM across all product categories, demonstrating better predictive accuracy. Gains were most pronounced in Office Products and Video Games, where content signals enriched user-item understanding. However, SVD and KNN showed performance close to or even better than neural network, indicating competitive predictive capability. These shows that collaborative methods remain competitive in dealing with real-world data.