Enhancing E-commerce Recommendations with Amazon Shopping Session

Jiayu Shi, Minwoo Sohn

Abstract

This project focuses on developing an advanced recommendation system for Amazon by leveraging Neural Networks (NNs), including Long Short-Term Memory Recurrent Neural Networks (LSTM-RNNs), to enhance the e-commerce experience. Utilizing a dataset with over 1.5 million product entries and 1.04 million session records, the initiative employs both traditional neural networks and LSTM models. The methodology integrates a graph-based approach to dynamically predict user purchases based on session data by incorporating various product features. Initial results showed minimal accuracy from the baseline NN model; however, integrating price features into the LSTM significantly improved performance. These outcomes illustrate the substantial potential of neural networks, particularly LSTM, to refine recommendation systems, although the notable performance discrepancies indicate a need for deeper investigation into data preprocessing and feature engineering.

Introduction

Modeling customer shopping intentions is pivotal for enhancing user experience and engagement on e-commerce platforms. The ability to accurately predict a customer's next purchase based on their current session data not only presents a significant technological challenge but also a substantial opportunity for innovation in the field of e-commerce. Traditional recommendation systems often rely on long-term user profiles or past purchasing histories, which may not always reflect the immediate and dynamic preferences of users. In contrast, session-based recommendation systems represent a more immediate approach by analyzing the data generated during a single browsing session. By analyzing data from a single browsing session, these systems offer real-time personalization, adapt quickly to user interactions, ensure greater privacy by avoiding long-term data, and respond promptly to emerging trends (Wu et al. 2023). For this project, we aim to harness the Amazon KDD Cup 2023 dataset, applying advanced neural network models to enhance our recommendation capabilities.

Objective

The overarching goal of this project is to enhance the Amazon e-commerce experience by developing an advanced recommendation system that accurately predicts the next product a customer will be interested in. The specific objectives are:

- Construct a graph-based model utilizing session data to predict users' next purchases dynamically.
- Improve prediction accuracy, targeting higher accuracy values to ensure relevant product recommendations are effectively prioritized.
- Leverage comprehensive product attributes and session data to derive insights into user preferences and behaviors, enhancing the personalization of recommendations.
- Refine the model to manage large data volumes efficiently, maintaining high performance as Amazon's marketplace continues to expand.

Method

Dataset

The datasets under analysis are organized into two main tables: Product and Session, each serving distinct roles in the development of the recommendation system.

1. Product Dataset

This dataset comprises 1,551,057 entries, each described by 11 attributes: ID, Locale, Title, Price, Brand, Color, Size, Model, Material, Author, and Description. Critical attributes like ID, Locale, Title, and Price are fully populated, ensuring reliable product information across entries. However, there are some missing values. Several attributes such as Color, Size, Model, Author, and Description have significant missing data (ranging from 19,371 to 789,922 missing entries), which may reflect attributes that do not apply to all products or inconsistencies in data entry.

2. Session Dataset

Our dataset consists of approximately 1.04 million session records. The session dataset includes key interaction data such as previous items (prev_items) and the next purchased item (next_item), providing insights into customer browsing and purchasing sequences. Data is segmented by the locale of the Amazon store, with the UK, DE, and JP locales showing the highest session counts. This geographic distribution will be essential for analyzing market-specific behaviors. Analysis of session length reveals a predominance of sessions with fewer than four interactions, indicating a trend towards brief, focused browsing sessions. This aspect is critical for optimizing the recommendation system to be efficient in handling shorter user interactions.

Exploratory Data Analysis (EDA)

This project aims to develop an advanced recommendation system for Amazon by performing an in-depth Exploratory Data Analysis (EDA) on the 'product_train' and 'session_train' datasets. The goal is to identify patterns across several product attributes that will inform the development of a recommendation model.

1. Most products are priced under 20€, with Japanese products generally below 45€, using outlier boundaries defined by IQR methods.
2. Popular brands include Amazon Basics, LEGO, APPLE, and others, indicating a market with diverse consumer preferences.
3. The majority of product listings are from DE, UK, and JP, with the UK selected for initial model testing due to language familiarity and data volume.
4. Indicated a strong preference for standard color schemes like Black, White, Multicolor, Blue, and Gray, underscoring their versatility and broad appeal across product categories. Additionally, "One Size" or "Free Size" is predominant, highlighting a significant market for products not requiring precise sizing.

Model Development

Neural Network

Neural Networks (NNs) simulate the human brain's functionality by processing and storing information through interconnected nodes or neurons, each using an activation function to produce outputs. The neurons are organized into three layers: input units that receive data, output units that present results, and hidden units that perform intermediary processing tasks. The strength of connections between these neurons, represented by adjustable weights, is crucial for information representation and processing. Unlike traditional AI systems that rely on logic-based processing, NNs are adaptive, self-organizing, and capable of real-time learning, enabling them to solve complex problems that defy conventional approaches. In recent decades, NNs have led to breakthroughs in pattern recognition and intelligent robotics, proving essential across various scientific and industrial fields (Wu et al. 2018). A simplified illustration of the neural network is presented below:

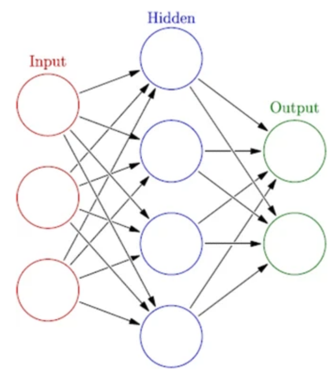

Figure 1. Simplified Neural Network Illustration (Wu et al. 2018)

Long Short-Term Memory (LSTM)

Long Short-Term Memory Recurrent Neural Networks (LSTM-RNNs) are a sophisticated branch of machine learning models that excel in dynamic classification tasks where understanding temporal dynamics is crucial. Originally developed to address the limitations of traditional Recurrent Neural Networks (RNNs), LSTM effectively overcomes the challenges of learning dependencies over varied time intervals. Traditional RNNs are prone to issues like vanishing or exploding gradients, which impede their ability to learn from data where dependencies span across long durations. LSTMs introduce a unique architecture featuring gates that regulate the flow of information, allowing them to retain or forget information selectively over long periods, thus enhancing their capacity to model complex sequences, such as speech or written text (Staudemeyer et al. 2019). By applying LSTM, the study intends to capture not only a customer's immediate interactions but also integrate long-term shopping behavior to generate more accurate and contextually relevant product recommendations. The Architecture of a LSTM Unit is depicted below:

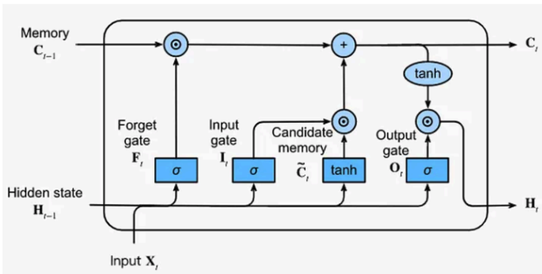

Figure 2. Architecture of LSTM Model (Zhang et al. 2021)

Evaluation Metrics

The accuracy metric is defined by the presence of the exact next product purchased within the top 100 recommended products. A prediction is considered successful and assigned a value of 1 if the next product appears among these top recommendations; if not, the prediction is scored as 0.

Challenges

The initial dataset presented computational challenges, with 1.4 million products and 3.6 million sessions. To address this, the focus was shifted exclusively to the UK market, which comprised around 500,000 products and 1.2 million sessions. This reduction still posed difficulties for processing on standard computing equipment.

The approach to manage the data included two primary filters:

1. Products that appeared in UK sessions more than 100 times were selected, which brought the product count down to 6,000.
2. Sessions were filtered to retain those with at least two previous products, and for sessions with more than two products, the sequences were truncated to include only the most recent two. This adjustment reduced the session count to 52,000.

Feature Engineering

The original product dataset featured columns for price, title, brand, color, size, material, author, and description. For efficiency, the 'price' column was kept unchanged while the remaining attributes were merged into one comprehensive string. This composite string underwent processing using a BERT model, resulting in a 768-dimensional vector that encapsulated the semantic nuances of the product.

Result

Baseline Model

A baseline model was implemented that randomly selects 100 products from the available pool of 6,000, achieving an accuracy rate of 1.49%. This model provides a fundamental benchmark against which more advanced methodologies can be evaluated.

Neural Network Model

For the neural network model, the following steps were undertaken:

- The data sessions were divided with 80% allocated for training and the remaining 20% for testing purposes.
- An adjacency matrix, A, was constructed using the training data exclusively.
- The feature matrix ***X*** was multiplied by A to derive the aggregated feature matrix ***X'***.
- The final input for the neural network was created by concatenating the matrices X and X' to form [X X'], which served as input for a five-layer neural network tasked with predicting the target product features.
- The top 100 product recommendations were ranked using cosine similarity.

The performance of the neural network model yielded the following results:

- When the feature matrix X was limited to only the 'price' feature, the model achieved an accuracy of 2.5%.
- When the feature set was expanded to include both 'price' and the 768-dimensional representations derived from the BERT model, a slight decrease in accuracy to 2.26% was observed. This could be attributed to the increased complexity of the feature set potentially overshadowing the influence of the 'price' feature.

LSTM

For the LSTM model, the following steps were undertaken:

- The dataset was preprocessed to convert item IDs into numerical indices. Unique items were

identified, and two mapping dictionaries (item_to_index and index_to_item) were created for encoding and padding sequences to uniform lengths
- Item embeddings were concatenated with additional feature values (if included in the model), providing enriched inputs for the LSTM layer, while maintaining the baseline model's output layer configuration.
- The model was trained using encoded, padded sequences with cross-entropy loss and the Adam optimizer. Performance was evaluated on the test set using Top-K accuracy (K=5) and overall accuracy metrics.

Baseline LSTM Model

A simple LSTM model was implemented without incorporating any additional features. The model consisted of an embedding layer to convert item IDs into dense vectors, followed by an LSTM layer to capture the sequential patterns, and finally a fully connected output layer to generate the predicted next item. This baseline model relied solely on the sequence of previously interacted items to make recommendations.

LSTM Model with Price Feature

To explore the potential of incorporating additional features, a new LSTM model was developed. This model extended the baseline architecture by concatenating item embeddings with their corresponding price information before feeding them into the LSTM layer. The intuition was that including price data could potentially enhance the model's understanding of user preferences and improve recommendation accuracy.

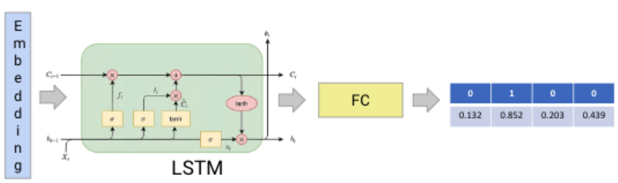

Figure 3. LSTM Network for Product Recommendation

Model Evaluation and Metrics

The performance of the models was evaluated using two primary metrics: Top-K accuracy and simple accuracy. Top-K accuracy measured the fraction of instances where the ground truth item was among the top K predicted items, with K set to 5 in this case. Simple accuracy, on the other hand, calculated the percentage of instances where the model correctly predicted the next item. Separate functions were implemented to compute these metrics on the test set, providing insights into the models' predictive capabilities.

The final result for the LSTM models are as follows:

- The baseline LSTM model achieved an accuracy of 0.00953%, while the top 5 accuracy was 0.009578%, which are disappointing result
- The LSTM model with the price feature demonstrated remarkably higher performance, with a 90.15477% Top-5 accuracy and an impressive 60.90% overall accuracy

The substantial discrepancies in accuracy between the two LSTM models are surprising and suggest potential issues in the implementation or data preprocessing. While the model incorporating the price feature showed significantly higher performance, such a large improvement is unexpected given the constraints of the recommendation task and the modest addition provided by the price feature alone. This warrants a thorough review of the data preprocessing, feature engineering, and model implementation to identify possible errors or anomalies. Additionally, verifying the accuracy calculations and evaluation methods is essential to ensure the reported metrics are free from discrepancies or biases. Further analysis, including replicating the experiments and employing alternative evaluation techniques, is necessary to confirm the reliability of these results and pinpoint any underlying issues.

Conclusion and Future Work

Both the NN and LSTM models surpassed the performance of our baseline. To further enhance model performance, we propose:

- Adjusting the adjacency matrix to retain all session links, improving the contextual linkage in the data.
- Conduct a comprehensive review of data preprocessing, feature engineering, and model implementation to address the unexpected discrepancies in accuracy between the baseline and price-feature LSTM models.
- Enhancing the LSTM model to incorporate a broader array of features, thus capturing a richer dataset.
- Expanding the product graph to include all products if computational resources allow, rather than limiting to frequent items.

These strategies aim to refine our approach, harnessing the full potential of the data for improved predictive accuracy.

References

[1] Wu, Shu; Tang, Yuyuan; Zhu, Yanqiao; Wang, Liang; Xie, Xing; Tan, Tieniu. "Session-Based Recommendation with Graph Neural Networks." 2023. AAAI Conference on Artificial Intelligence. https://arxiv.org/abs/1811.00855v4

[2] Wu, Yc.; Feng, Jw. "Development and Application of Artificial Neural Network." 2018. Wireless Personal Communications 102(5): 1645–1656. https://doi.org/10.1007/s11277-017-5224-x

[3] Staudemeyer, Ralf C.; Morris, Eric Rothstein. "Understanding LSTM -- a tutorial into Long Short-Term Memory Recurrent Neural Networks." 2019. ArXiv:1909.09586 [cs.NE]. Comments: 42 pages, 11 figures, tutorial. https://arxiv.org/abs/1909.09586

[4] Zhang, Aston; Lipton, Zachary C.; Li, Mu; Smola, Alexander J. "Dive into Deep Learning." 2021. ArXiv:2106.11342. https://arxiv.org/abs/2106.11342