Movie Recommendation System Using Transformers Project Proposal

Dylan Newman, Xuanjia Bi, Zhanyang Gong, Wentao Shao
University of California, Los Angeles
Los Angeles, CA 90095
xuanjiabi@ucla.edu

1 Problem Statement

Recommendation systems are a cornerstone in marketing and keeping a user engaged with your platform. One of the most popular applications is for efficiently recommending movies on streaming platforms. Companies like Netflix, Hulu, and Max have all invested millions of dollars to keep their users on the platform and engaged with the content. Our goal with this project is to take existing recommendation systems and expand upon them with the hopes of creating an even more accurate and efficient way of finding similarities between a user's movie history and ratings and what they would be interested in watching next. To do this we are using a dataset from GroupLens Research which has taken data from MovieLens, a popular movie recommendation site, and compiled movies, their ratings, tags, and tag relevance to each movie. We can take an existing Transformer-based model like BERT4Rec, which specializes in capturing user interactions for sequential recommendations, and enhance it through various explorations to build a hybrid model that integrates collaborative filtering. This is to train the model with segmented data and to experiment with different data embedding techniques. We will also be comparing the results to various other models to see how much we improved. These models include Self-Attentive Sequential Recommendation (SASRec), a traditional BERT model, and a naive approach using cosine similarity.

2 Literature Review

The extraordinary performance of transformer-based models, such as GPT and BERT, has led to tremendous success in a variety of tasks, including text classification, translation, summarization, reasoning, and generation. Recommendation systems have also been shown to be effective when using transformer-based pre-trained models, due to their capability of handling large volumes of textual or embedded information in both unidirectional and bidirectional contexts.

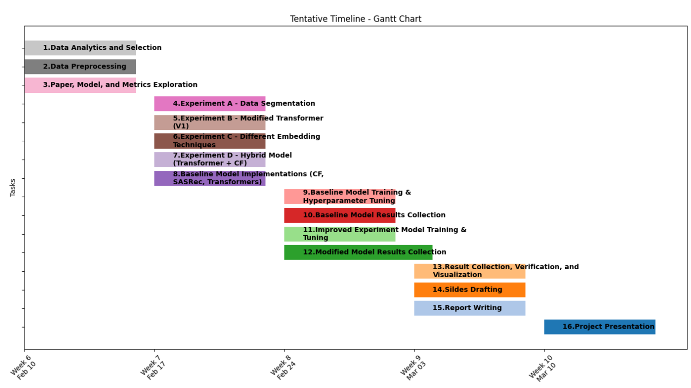
Previous work by Sun et al. (2019) developed BERT4Rec, a model inspired by the pre-trained bidirectional self-attention architecture that captures users' sequential behaviors through a random masking objective. The model demonstrated success in sequential recommendation tasks when compared with several state-of-the-art baseline models, ranging from the simplest matrix factorization approaches to unidirectional single-head attention frameworks. In other work, Moreira et al. (2021) proposed Transformers4Rec, leveraging the Hugging Face Transformers library and modifying its building blocks to fit recommendation system prediction tasks. Even though the implementation of transformer-based recommendation systems has not yet seen widespread practical use, it has been identified as a promising direction for future research, including multimodal applications (Islam et al., 2023).

As a result, the objective of this work is to explore and evaluate possible enhancements to current transformer-based recommendation models, while comparing them with state-of-the-art traditional approaches commonly used in recommendation systems.

Petrov and Macdonald (2022) conducted a systematic review and replicability study of BERT4Rec, a Transformer-based sequential recommendation model. Their study found significant inconsistencies in BERT4Rec's reported performance across different research papers, with some studies failing to reproduce its original results. Through analyzing multiple BERT4Rec implementations, they discovered that the original model requires up to 30x longer training than its default configuration to achieve its claimed effectiveness. To address this, they developed a new implementation using Hugging Face Transformers, which replicated the original results on 3 out of 4 datasets while reducing training time by up to 95%. Additionally, they demonstrated that alternative Transformer architectures (e.g., ALBERT, DeBERTa) can enhance BERT4Rec's performance by up to 9%. Their work highlights the importance of proper training configurations in sequential recommendation models and provides a more efficient, reproducible implementation for future research.

Xia et al. (2024) propose a multi-modal movie recommendation system that leverages pre-trained Transformer models for feature extraction and fusion. The system utilizes BERT for processing textual information (e.g., movie descriptions), ViT (Vision Transformer) for extracting features from movie posters, and a Transformer-based architecture for integrating these modalities to predict user preferences. Their approach addresses the data sparsity problem in traditional recommendation systems by incorporating multi-modal content features, leading to more accurate recommendations. Experimental results on MovieLens 100K and MovieLens 1M datasets demonstrate superior performance compared to collaborative filtering and traditional deep learning models. This research highlights the effectiveness of Transformer-based multi-modal feature fusion in improving recommendation accuracy, which is highly relevant to our project's goal of enhancing movie recommendation systems. Several traditional models are considered in our experiments, including the use of SVD for CF by Zhang et al. (2022). They suggest that matrix factorization for CF works by decomposing the user-item interaction matrix into lower-dimensional latent factors, capturing user preferences and item characteristics to predict missing ratings. Additionally, Kang et al (2018). introduced SASRec, one of the first self-attention-based models for sequential recommendations, which uses a Transformer-like self-attention mechanism to model user interaction sequences.

3 Workload/Schedule



Data Gathering & Preprocessing
Zhanyang Gong: Identify & acquire datasets; initial cleaning; Exploratory Data Analysis
Dylan Newman: Explore enhancement and implementation
Wentao Shao: Explore model choices and baselines
Eric Bi: Set up methodologies and explore metrics
Baseline Model Implementation
Everyone: Implement potential traditional models and analyze results
Transformer-Based Model Setup
Everyone: Develop major transformer model
Results Analysis & Visualization
Everyone: Visualize results for comparison; compile results from all models
Report Writing & Presentation
Zhanyang Gong: Objective Section; results Section for comparison
Dylan Newman: Results; conclusion section; final formatting

Wentao Shao: Introduction; abstract; results
Eric Bi: Methodology Section; results

References

[1] De Souza Pereira Moreira, G., Rabhi, S., Lee, J. M., Ak, R., Oldridge, E. (2021). Transformers4Rec: Bridging te Gap between NLP and Sequential / Session-Based Recommendation. ACM. https://doi.org/10.1145/3460231.3474255

[2] Islam, S., Elmekki, H., Elsebai, A., Bentahar, J., Drawel, N., Rjoub, G., Pedrycz, W. (2023). A comprehensive survey on applications of transformers for deep learning tasks. Expert Systems With Applications, 241, 122666. https://doi.org/10.1016/j.eswa.2023.122666

[3] Kang, W.-C., McAuley, J. (2018). Self-attentive sequential recommendation. arXiv. https://arxiv.org/abs/1808.09781

[4] Petrov, A., Macdonald, C. (2022). A Systematic Review and Replicability Study of BERT4Rec for Sequential Recommendation. In RecSys '22: Sixteenth ACM Conference on Recommender Systems, September 18–23, 2022, Seattle, WA, USA. ACM. https://doi.org/10.1145/3523227.3548487

[5] Sun, F., Liu, J., Wu, J., Pei, C., Lin, X., Ou, W., Jiang, P. (2019). BERT4Rec. ACM, 1441–1450. https://doi.org/10.1145/3357384.3357895

[6] Xia, L., Yang, Y., Chen, Z., Yang, Z., and Zhu, S. 2024. Movie Recommendation with Poster Attention via Multi-modal Transformer Feature Fusion. arXiv preprint arXiv:2407.09157.

[7] Zhang, Y., Pal, A., Khan, M. S., Kumar, A., Harford, S., Eskin, E., Krishnan, S. (2022). Transformer-based approaches for self-supervised learning in speech processing: A survey. arXiv. https://arxiv.org/abs/2203.11026