Group 5: Topic 1: Recommender System

Adham Mokhtar-20398545, Manar Elghobashy - 20398569
Yara Elzahy - 20398570, Yara Hassan - 20398563

1 General Requirements

We want to make sure each of the team members takes care of one model or a pipeline. So that everyone can benefit from the task. You can treat a group-based project as a learning group given a specific task, but each of you still needs to implement your own DL pipeline.

The project proposal must clearly describe which member is taking care of what task and get it approved by TAs.

There are multiple goals for the group-based projects:

1) Get familiar with experiment design and dataset preparation for a specific deep-learning task.
2) Get familiar with SOTA models for the specific task.
3) Able to replicate SOTA models and apply them to new datasets.
4) Able to compare and conclude on the performance of SOTA models.
5) Collaborate with others to boost your knowledge on the specific task.
6) Presentation (we won't ask for a written report, we ask for a recorded presentation and a replicate package for all experiments, your code will tell the methodology, and your results will tell your findings).
7) For the replication package, we recommend once you have done all experiments, create a neat replication package in a notebook file (ipynb) and upload it into your google drive and onQ. Use sections and descriptions to make your replication package easy to follow. One notebook file would be fine for each team. The replication package is only required for the final evaluation (it is not required for the project proposal)

In your project proposal, you will provide four sections as described bellow.

2 Motivation and Problem Statement

Problem Statement: This project aims to develop an effective session-based recommendation system. The input of the problem would typically consist of user-session information, including details such as user interactions, timestamps, and contextual information. The expected outcome of the CTR prediction problem would be a prediction or probability indicating the likelihood of a user clicking on the next item in a session from the QK-video-1M dataset.

Motivation: In today's overwhelming online landscape, traditional recommendation systems that focus on long-term user preferences may not suffice. Session-based recommendation systems offer real-time, personalized recommendations by analyzing users' behaviors within shorter sessions. This motivates the need for accurate and context-aware recommendations that adapt to users' changing preferences within a session.

The main user groups that would benefit from the outcome of CTR prediction in the proposed problem include e-commerce companies, advertisers, marketers, recommender system developers and users.

3 Related Work

Related Work 1: Sun et al. introduced "BERT4Rec: Sequential Recommendation with Bidirectional Encoder Representations from Transformer" [2]. The paper adapts the BERT model to the session-based recommendation problem and proposes BERT4Rec, which utilizes masked self-attention to capture both short- and long-term user preferences. The effectiveness of BERT4Rec is demonstrated on multiple datasets.

Related Work 2: The paper "Recurrent Neural Networks with Top-k Gains for Session-based Recommendations" by Hidasi et al. [1] introduces a session-based recommendation model that utilizes recurrent neural networks (RNNs) with top-k gains. The model aims to improve the accuracy of session-based recommendations by focusing on the top-k items that are most likely to be relevant to the user's current session. The paper provides insights into the effectiveness of incorporating top-k gains into RNN-based recommendation models.

Related Work 3: Chuhan Wu et al. proposed "Session-based Recommendations with Recurrent Neural Networks" [3], where they utilized SASRec, a recurrent neural network (RNN) model, for session-based recommendation. The paper explores the effectiveness of SASRec in capturing sequential patterns and making accurate recommendations in session-based scenarios.

Related Work 4: The paper "NextItNet: Next Item Recommendation with Neural Item Embedding and Multitask Learning" by Yuan et al. [4] presents NextItNet, a deep learning model for next item recommendation. NextItNet incorporates neural item embedding and multitask learning to capture sequential patterns and improve recommendation accuracy. The paper explores the effectiveness of Nex-

tItNet on various datasets and demonstrates its superior performance compared to other state-of-the-art models.

4 Basic Data Exploration

Tenrec dataset: The Tenrec dataset is a benchmark dataset for click-through rate (CTR) prediction in e-commerce. It was collected from an e-commerce platform and includes user interactions, demographics, and item attributes. The dataset is large-scale and contains interconnected files capturing various aspects of user-item interactions. It serves as a valuable resource for developing and evaluating CTR prediction models in recommender systems research.

The QK-video dataset was used for the experiment. The dataset was filtered to remove sessions with a length shorter than 10, which is a common practice in CTR prediction. The maximum session length was set to 30, and sessions shorter than 30 were padded with zeros. The dataset consisted of 928,562 users, 1,189,341 items, and 37,823,609 clicking interactions. For testing, the last item in each session was used. The second to last item in each session was used for validation. The remaining interactions were used for training the CTR prediction models.

5 Experiment Design and Work Distribution

In this experiment, we will use the Tenrec dataset for session-based recommendation systems. Our plan for achieving the goals of implementing a session-based recommendation system using the Tenrec dataset is as follows:

- Model/Pipline implementation.
- Training and evaluation.
- Hyperparameter tuning.
- Comparison and analysis.

Each team member is assigned to implement a specific model:

- Manar elghobashy implements BERT4Rec,
- Yara elzahy implements GRU4Rec
- Yara mohammed implements SASRec
- Adham gamal implements NextItNet.

The team is going to work together on the hyperparameter tuning and the comparison and analysis tasks.

References

[1] B. Hidasi, A. Karatzoglou, L. Baltrunas, and D. Tikk. Recurrent neural networks with top-k gains for session-based recommendations. *arXiv preprint arXiv:1706.03847*, 2018.

[2] F. Sun, J. Liu, M. Wu, W. Li, and Y. Zhou. Bert4rec: Sequential recommendation with bidirectional encoder representations from transformer. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. ACM, 2019.

[3] C. Wu, F. Wu, M. An, J. Huang, Y. Huang, and X. Xie. Session-based recommendations with recurrent neural networks. In *Proceedings of the 2016 ACM Conference on Recommender Systems*. ACM, 2016.

[4] F. Yuan, A. Karatzoglou, I. Arapakis, and J. M. Jose. Nextitnet: Next item recommendation with neural item embedding and multitask learning. *arXiv preprint arXiv:1808.05163*, 2018.