Data Science Practice Assignment 3

Problem Description

This assignment aims to reproduce the paper: Neural collaborative filtering by latest version of PyTorch. The assignment consists of the following tasks:

- Version of latest PyTorch
- Implement three method (MLP, GMF, NeuMF) without redundant code
- Follow the training and testing setting of the paper
- Compare the three method using metric of HR@10 and NDCG@10
- Reproduce the ablation study in Table 3 (MLP method with different layer)

Solution Procedure

Q1:

The version of PyTorch is shown:

```
Name: torch
Version: 2.3.0a0+6ddf5cf85e.nv24.4
Summary: Tensors and Dynamic neural networks in Python with strong GPU acceleration
Home-page: https://pytorch.org/
Author: PyTorch Team
Author-email: packages@pytorch.org
License: BSD-3
Location: /usr/local/lib/python3.10/dist-packages
Requires: filelock, fsspec, jinja2, networkx, sympy, typing-extensions
Required-by: flash-attn, lightning-thunder, torch-tensorrt, torchdata, torchtext, torchvision, transformer-engine
```

Q2:

For implementation of the three method:

> MLP: follow the default setting of the paper: the embedding size is 16 (8 + 8), the layer constructed is 32 → 16 → 8 and the factor is 8.

> GMF: the dimension of embedding for each one-hot vector is 8, and factor is 8

> NeuMF: this method is the combination of MLP and GMF, and the setting of this method corresponds to the previous methods. The embedding dimension is set as 8, and the factor is 16 (since we need to concatenate the output of two modules).

Q3:

The training and testing setting strictly follows subsection 4.1 of the paper. To be more specifically, for each evaluation, the latest interaction between user and item is selected and 100 negative samples are sampled for testing. For training, for every positive interaction, four negative interactions are sampled and combined with the previous positive interaction to obtain a set of training sample.

Q4:

The comparisons are shown:

	HR@10	NDCG@10
MLP	0.6528	0.3804
GMF	0.6402	0.3672
NeuMF	0.6856	0.4118

Q5:

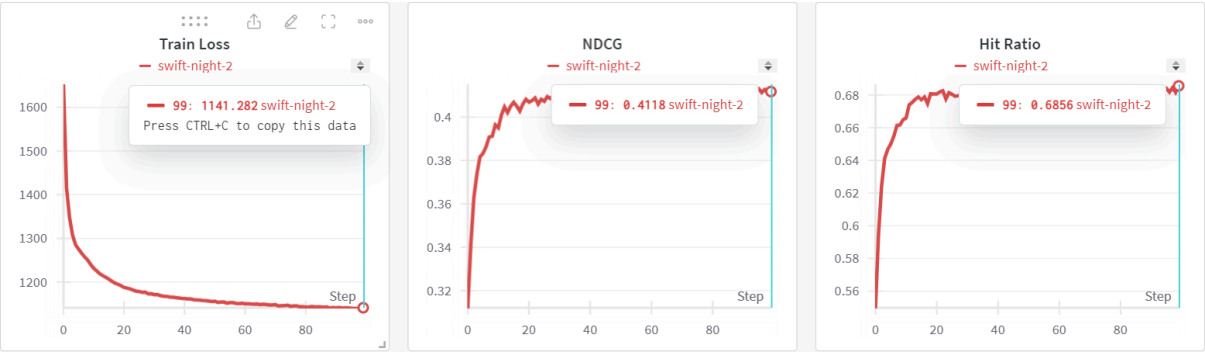
Comparison of ablation study of MLP layers used in MLP method:

Factors (8)	MLP0	MLP1	MLP2	MLP3	MLP4
HR@10	0.4465	0.5715	0.5829	0.5949	0.5906
NDCG@10	0.2483	0.3282	0.3256	0.3391	0.3398

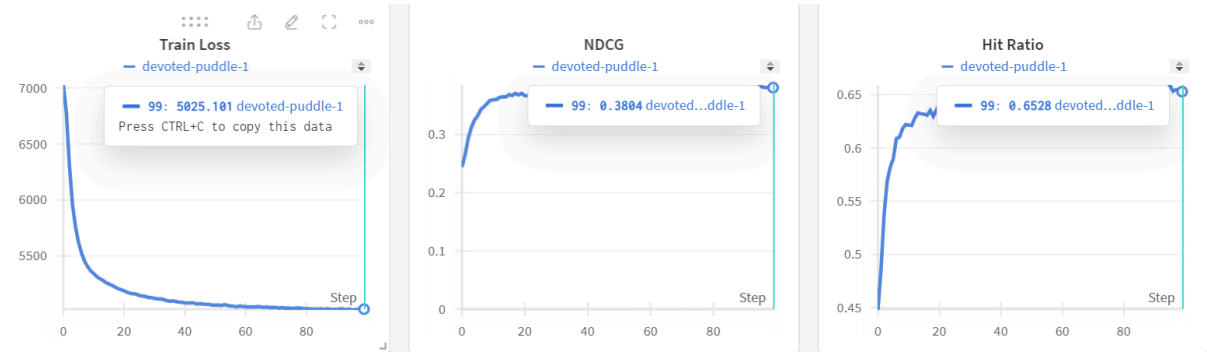
Appendix:

More detailed experiment log is shown:

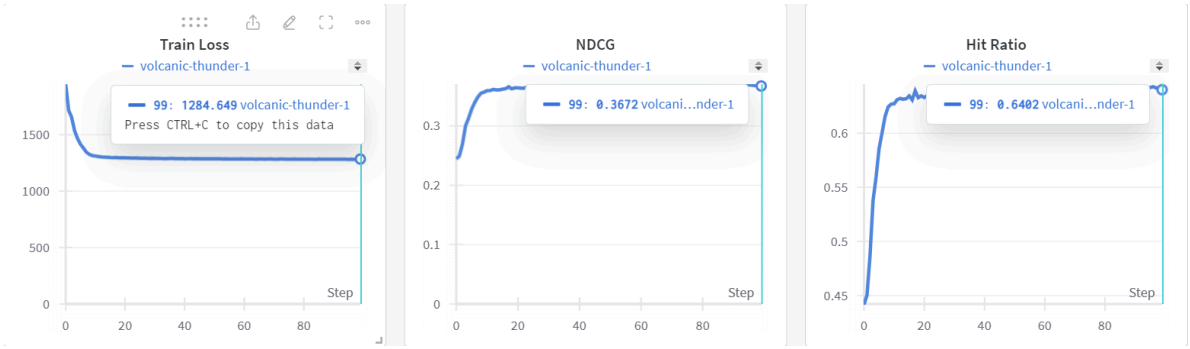
NeuMF_factor8neg4:



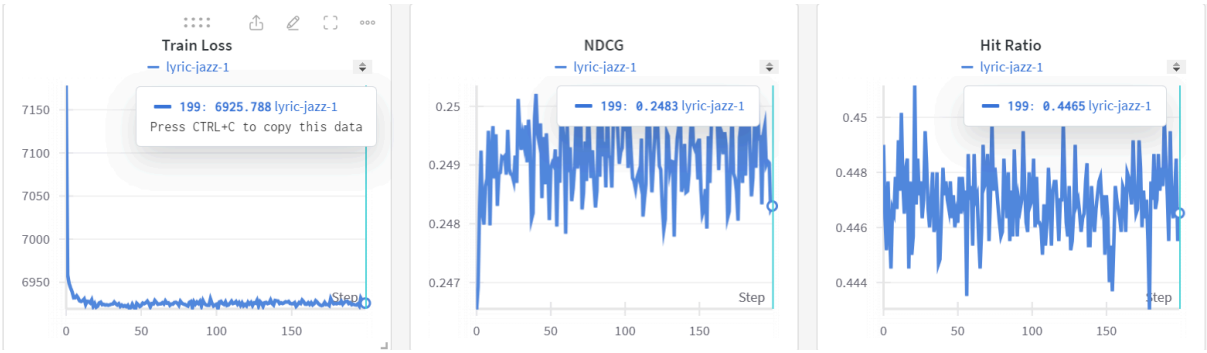
MLP_factor8neg4:



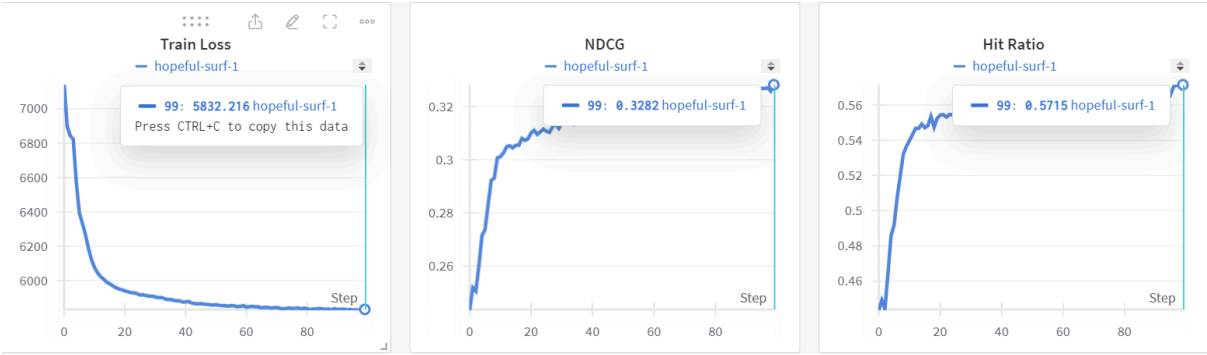
GMF_factor8neg4:



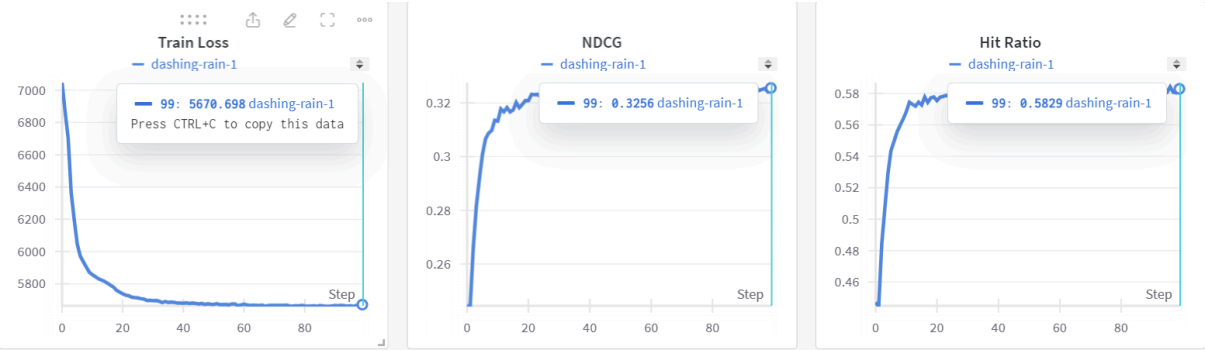
MLP_factor8neg4_layer0:



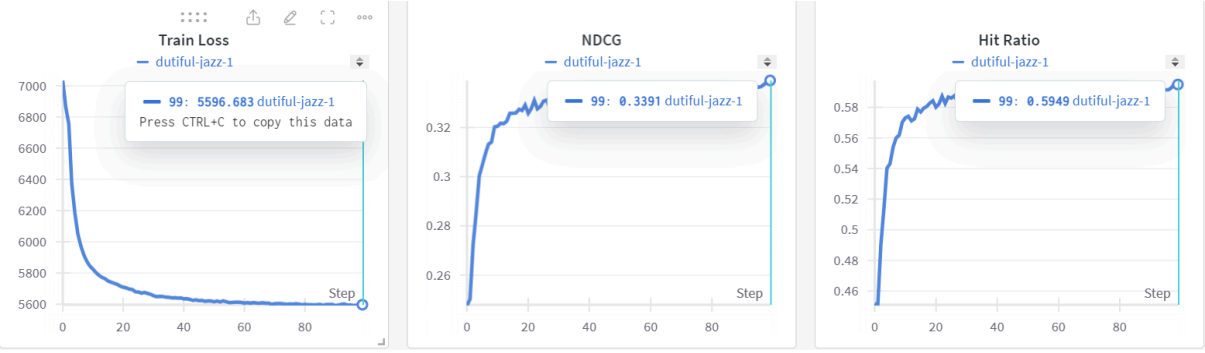
MLP_factor8neg4_layer1:



MLP_factor8neg4_layer2:



MLP_factor8neg4_layer3:



MLP_factor8neg4_layer4: