第六次实验

58119125
蒋卓洋

一．配置与调试：

1．问题解决：

（1）虚拟机扩容失败后打不开了：重装虚拟机重新配环境

（2）配置过程中忘记换环境，导致后续操作错乱：删除 anaconda 相关文件，从头开始配置。

2．样例测试结果：

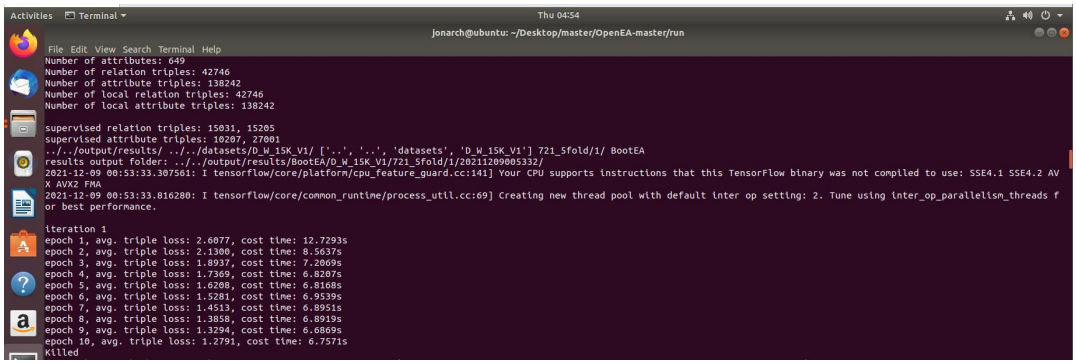

图 1.样例测试

二．分别使用 EN_FR_15K_V2 的 split1 和 EN_DE_15K_V2 的 split2 来运行 MTransE，记录使用的命令和结果

1．EN_FR_15K_V2 的 split1 运行 MTransE：

（1）需求分析：

①测试代码（python 源文件）：main_from_args.py

②模型参数：./args/mtranse_args_15K.json

（embedding 模式为 MtransE，数据集规模为 15K）

③训练集：EN_FR_15K_V2

④训练集具体文件夹：1

（2）命令：

```
python main_from_args.py ./args/mtranse_args_15K.json EN_FR_15K_V2 721_5fold/1/
```

图 2.命令 1

（3）结果：

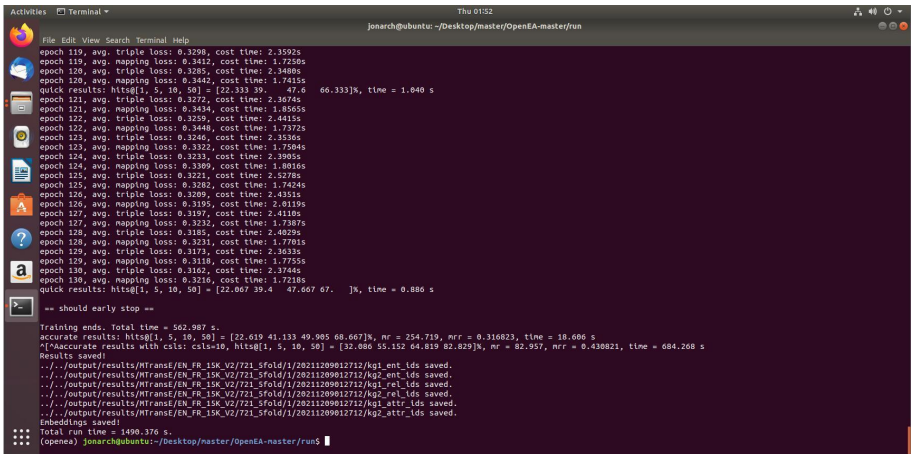

图 3.运行结果 1

```
 == should early stop ==

Training ends. Total time = 562.987 s.
accurate results: hits@[1, 5, 10, 50] = [22.619 41.133 49.905 68.667]%, mr = 254.719, mrr = 0.316823, time = 18.606 s
^[^Aaccurate results with csls: csls=10, hits@[1, 5, 10, 50] = [32.086 55.152 64.819 82.829]%, mr = 82.957, mrr = 0.430821, time = 684.268 s
Results saved!
../../output/results/MTransE/EN_FR_15K_V2/721_5fold/1/20211209012712/kg1_ent_ids saved.
../../output/results/MTransE/EN_FR_15K_V2/721_5fold/1/20211209012712/kg2_ent_ids saved.
../../output/results/MTransE/EN_FR_15K_V2/721_5fold/1/20211209012712/kg1_rel_ids saved.
../../output/results/MTransE/EN_FR_15K_V2/721_5fold/1/20211209012712/kg2_rel_ids saved.
../../output/results/MTransE/EN_FR_15K_V2/721_5fold/1/20211209012712/kg1_attr_ids saved.
../../output/results/MTransE/EN_FR_15K_V2/721_5fold/1/20211209012712/kg2_attr_ids saved.
Embeddings saved!
Total run time = 1490.376 s.
```

图 4.运行效果 1

图 5.对齐结果 1

2．EN_DE_15K_V2 的 split2 运行 MTransE：

（1）需求分析：

①测试代码（python 源文件）：main_from_args.py

②模型参数：./args/mtranse_args_15K.json

（embedding 模式为 MtransE，数据集规模为 15K）

③训练集：EN_DE_15K_V2

④训练集具体文件夹：2

（2）命令：

```
python main_from_args.py ./args/mtranse_args_15K.json EN_DE_15K_V2 721_5fold/2/
```

图 6.命令 2

（3）结果：

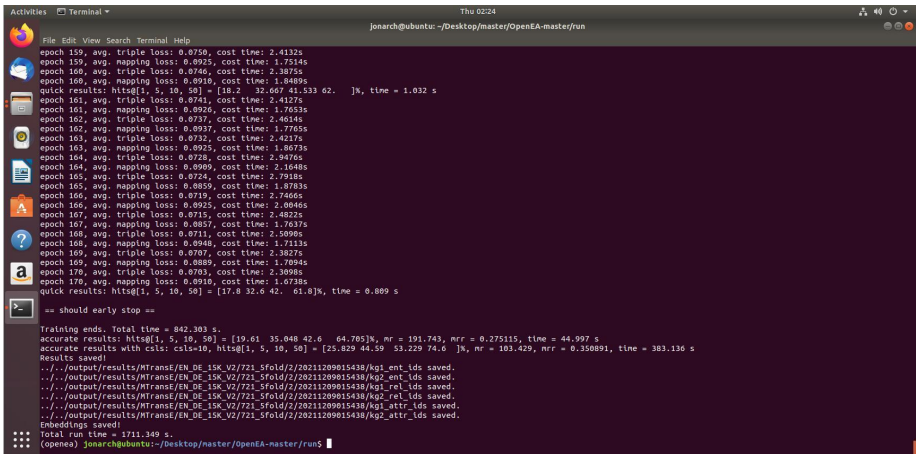

图 7.运行结果 2

```
== should early stop ==

Training ends. Total time = 842.303 s.
accurate results: hits@[1, 5, 10, 50] = [19.61  35.048 42.6   64.705]%, mr = 191.743, mrr = 0.275115, time = 44.997 s
accurate results with csls: csls=10, hits@[1, 5, 10, 50] = [25.829 44.59  53.229 74.6  ]%, mr = 103.429, mrr = 0.350891, time = 383.136 s
Results saved!
../../output/results/MTransE/EN_DE_15K_V2/721_5fold/2/20211209015438/kg1_ent_ids saved.
../../output/results/MTransE/EN_DE_15K_V2/721_5fold/2/20211209015438/kg2_ent_ids saved.
../../output/results/MTransE/EN_DE_15K_V2/721_5fold/2/20211209015438/kg1_rel_ids saved.
../../output/results/MTransE/EN_DE_15K_V2/721_5fold/2/20211209015438/kg2_rel_ids saved.
../../output/results/MTransE/EN_DE_15K_V2/721_5fold/2/20211209015438/kg1_attr_ids saved.
../../output/results/MTransE/EN_DE_15K_V2/721_5fold/2/20211209015438/kg2_attr_ids saved.
Embeddings saved!
Total run time = 1711.349 s.
```

图 8.运行效果 2

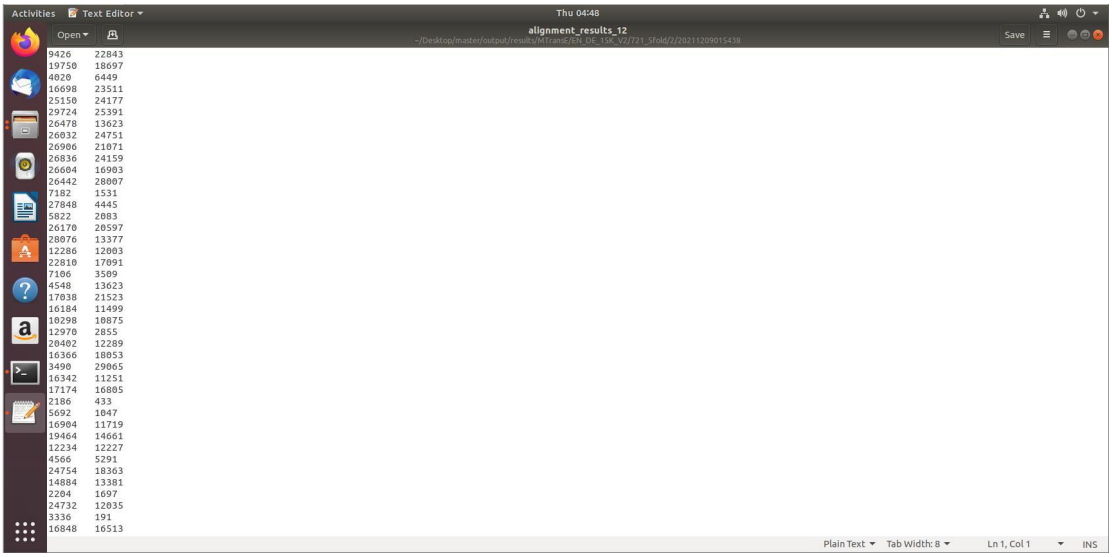

图 9.对齐结果 2

三．mtranse_args_15K.json 和 mtranse_args_100K.json 有何区别，为什么要设置这种区别，而不是直接写一个 mtranse_args.json？

答：区别在于训练时的 batch_size 不同，同时线程数量也对应的不同。

分开写两种 batch size 的模型参数，可以应对不同训练的需求。

四．什么是 earlystop？这个实例中为什么需要 earlystop？

答：早停是用于在深度学习训练时防止过拟合的方法，希望模型泛化能力更强，故在训练错误率达到一定值的时候就停止继续优化。

本实验中 embedding 方法是全连接神经网络，很容易过拟合，故选择用遭停发防止过拟合。