

代码说明

笔记本： 草稿

创建时间： 2021/2/25 18:13

更新时间： 2021/2/25 19:14

作者： 1657954133@qq.com

同学你好，整个项目代码组织地不太好所以有点臃肿，我尽量分模块地讲下这些代码的作用

1) 整体idea流程：整体idea的出发点是对sequential recommendation model在进行训练的时候，加入利用gan生成数据，从而提升模型精度的目的。原始数据是序列，因此核心就是，如何利用gan来进行序列生成。老师提出的方法是利用segan来进行序列生成，同时加入两个创新点，第一点seqgan生成的序列同时使得原始sequential recommendation model的损失函数更大，第二点是自步学习，逐步提升生成的序列长度。

2) seqgan的代码我借鉴的是<https://github.com/ZiJianZhao/SeqGAN-PyTorch>中的代码。sequential recommendation model的实现借鉴的是伯乐推荐系统中的实现，具体来说是选了四个model，gru4rec，stamp，narm，sasrec模型。

3) 代码中最核心的两个文件是run_seq_model 和 seqgan，其中run_seq_model用来训练序列模型，接受三个参数：第一个参数，选择数据集（yoochoose1/64 diginetica）；第二个参数：选择序列模型（gru4rec，stamp，narm，sasrec）；第三个参数：ori或者gan，如果是ori，就是直接训练序列模型，如果是gan，就是论文的方法，这时候训练序列模型，会读取gan生成的数据，加入到训练集中

4) seqgan就是用来生成数据的，这部分其实就是论文内容的实现，它接受三个参数，第一个参数选择数据，第二个参数选择序列模型，第三个参数是batch size，它的目的是为对应模型利用gan生成数据。

5) 关于 run_seq_model和seqgan的总结，如果只是想运行baseline，直接用ori模式的run_seq_model，如果想运行论文的方法，假如想运行narm模型在yoochoose1/64数据集上的论文效果，先运行segan（数据集选yoochoose1/64，模型narm），再运行run_seq_model（数据集选run_seq_model，模型narm，模式选'gan'），seqgan是负责生成数据的，run_seq_model需要读取seqgan生成的数据

下面是剩余的文件的介绍

1) abstract_recommender，layer，gru4rec，stamp，narm，sasrec这六个文件，是从伯乐上面借鉴来的四个模型的实现，因为通过伯乐库调用这四个模型，比较难，所以我把他们的代码直接截取了下来。

2) data文件夹，用来存放数据集文件，目前支持两个数据集yoochoose1/64，diginetica，每个数据集包含train.pickle，test.pickle，valid.pickle三个文件，对应训练集，测试集，验证集，比例是80%，10%，10%

2) gan文件夹包含了实际上是seqgan的生成器，判别器的一些实现内容，也就是<https://github.com/ZiJianZhao/SeqGAN-PyTorch>中的代码

3) gan_model和gan_data会保存seqgan训练好生成器，判别器，以及生成的数据

4) log文件夹，保存的是run_seq_model训练序列模型时候产生的日志文件

5) saved_model，保存的训练好的序列模型

其它还存在的一些问题

1) 我写的seqgan.py其中并没有包含自步学习过程，因为当时和老师讨论了比较多版本的自步学习，我后来觉得我实现的自步学习可能跟老师写的是有出入的，所以怕误导并没有放上去。

2) 我写的seqgan.py内容比较凌乱，因为是在原始seqgan上进行修改的，代码结构并不是很好，我建议最好你可以先看下<https://github.com/ZiJianZhao/SeqGAN-PyTorch>这个原版的实现，然后再按照自己理解来添加老师说的创新点，不然对我的代码进行改写难度太大。