大数据分析第八次作业——知识图谱补全

【目的】

对知识图谱进行补全，预测出三元组中缺失的部分，从而使知识图谱变得更加完整.根据补全类型分类，使得三元组的部分得到补全，具体到freebase数据集中，就是补全每个人的出生等与自身关系的知识图谱。

【数据集】

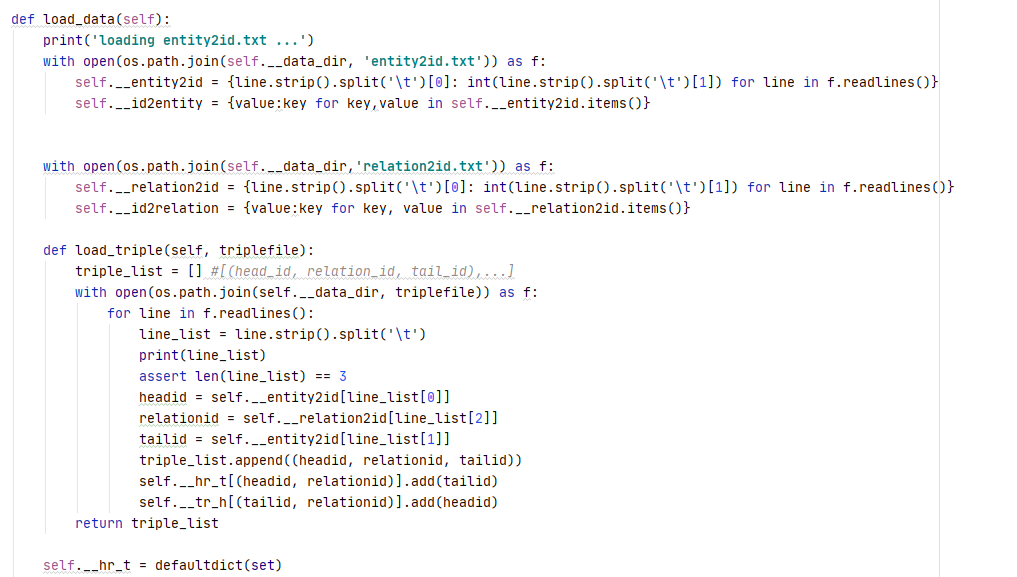
Freebase15k

【使用方法】

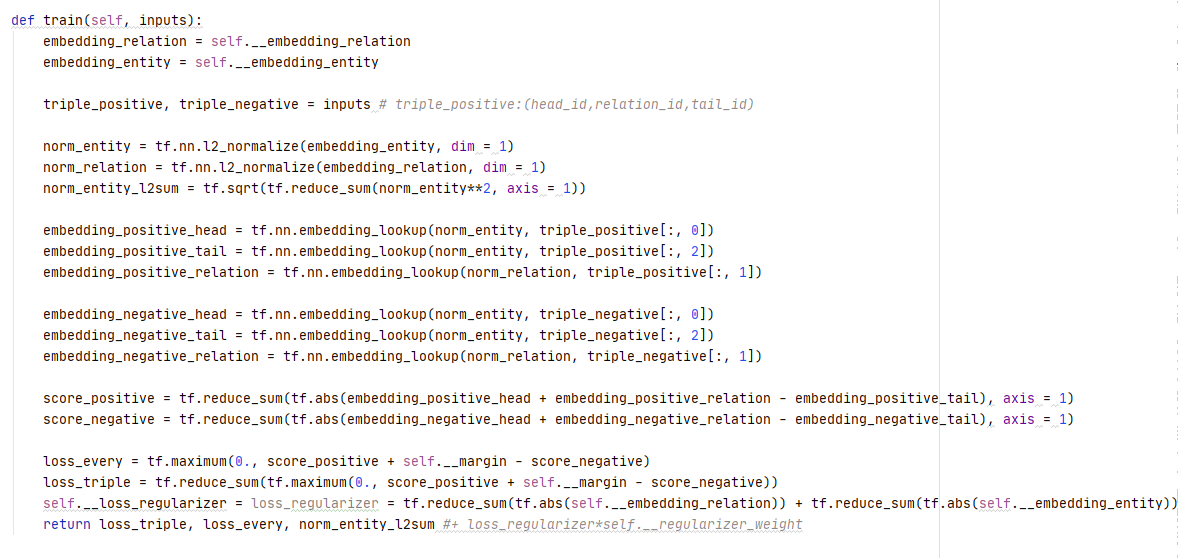
使用TransE加tensorflow模拟神经网络来实现，TransE算法，是Bordes等人2013年发表在NIPS上的文章提出的算法，根据我找到的Bordes的论文，TransE方法，就是基于实体和关系的分布式向量表示，将每个三元组实例（head，relation，tail）中的关系relation看做从实体head到实体tail的翻译（其实我一直很纳闷为什么叫做translating，其实就是向量相加），通过不断调整h、r和t（head、relation和tail的向量），使（h + r） 尽可能与 t 相等，即 h + r = t。基本原理：

【算法细节】

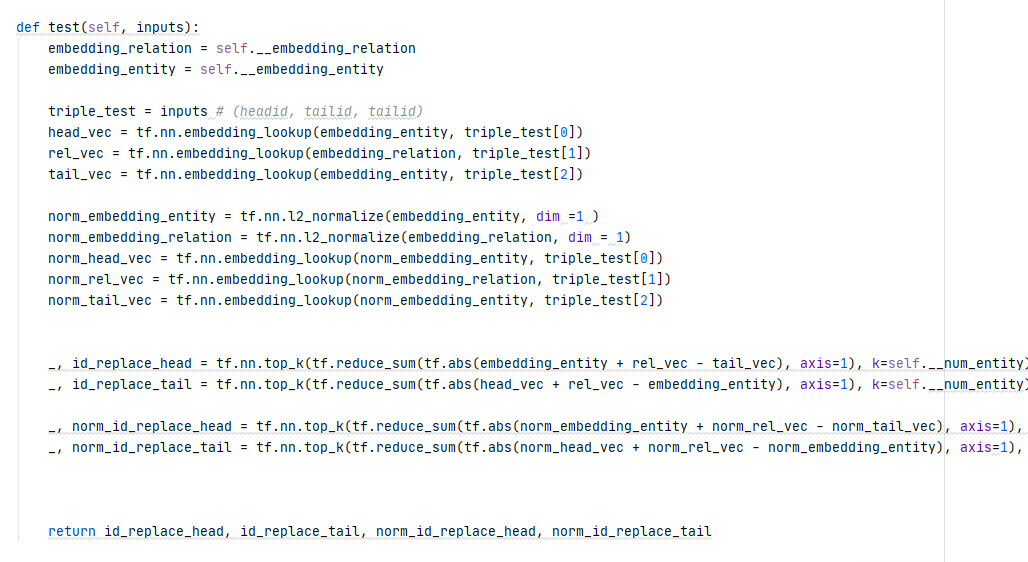
1. 读取数据



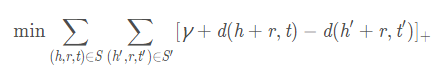
1. 训练



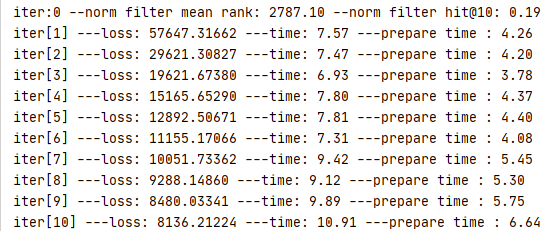
1. 测试



【指标】

在我的算法中，误差（loss）就是SVM中对应的误差模型，误差函数设计如下，另外具有hit与meanrank，hit：即后面的数字个个体中能否遇到真正的实体，meanrank即平均计算到多少个才能遇到正确的实体。

【结论】



如图所示为一个iter的运行结果数据，为每一次迭代的数据。