**目录**

[运行环境 1](#_Toc27357)

[界面介绍 2](#_Toc7021)

[系统逻辑 6](#_Toc9814)

[后台操作 6](#_Toc25472)

[项目运行 7](#_Toc14148)

## 运行环境

一、软件环境

1. GUI部分依赖于py2neo4.3.0，其他版本未经测试可能无法成功动态展示知识图谱，neo4j3.5.0版本，py2neo4.3.0不支持neo4j4.0及以后的版本，可以修改源码运行，但可能导致出乎意料的错误。
2. 推荐算法部分依赖tensorflow1.14框架，python3.7。
3. Window11，显卡RTX3070
4. 数据结构

根据网络资源、公务员录用考试公共科目教材周盈主编的《申论》，政法干警招录培养考试专用教材李永新主编的《申论（本硕类）》等查询知识点，建立三元组关系的组合作为数据。分为行测和申论两个模块。

2.1、制作数据

entity1是作为父节点，entity2是作为用分号分隔的若干子节点，关系“relation”在行测的知识点是“包括”，在申论模块内容是“包括（申论）”。

|  |  |  |
| --- | --- | --- |
| entity1 | relation | entity2 |
| 言语理解与表达 | 包括 | 片段阅读；语句表达；逻辑填空；篇章阅读 |
| 图形推理 | 包括 | 位置规律；样式规律；属性规律；数量规律；特殊规律；空间重构；空间类 |
| 位置规律 | 包括 | 题型特征（位置规律）；平移；旋转；翻转 |
| 题型特征（位置规律） | 包括 | 元素组成相同 |
| 平移 | 包括 | 方向（平移）；步数；路径；粉笔提示（平移） |
| 方向（平移） | 包括 | 上下、左右、顺逆时针；粉笔提示（方向） |

2.2 GUI要求的json格式数据

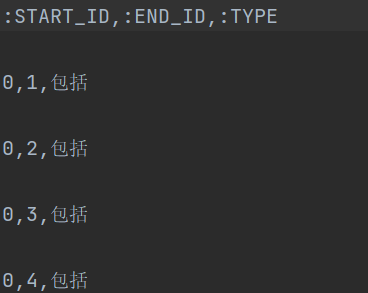
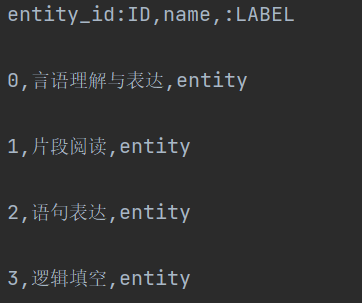
为了GUI以树形结构展开，以递归形式展示了节点与节点之间的父子关系和爷孙关系等，如下：

{'content': '定义判断', 'children': [{'content': '读定义', 'children': [{'content': '提问方式（读定义）', 'children': ['属于/不属于', '正确/不正确', '违背/不违背']}, {'content': '读题技巧（读定义）', 'children': ['单定义', '多定义', '故事类']}]}, {'content': '看选项（定义判断）', 'children': [{'content': '关键信息（看选项）', 'children': ['主体、客体、本质', '目的、方式、原因、时间、地点']}, '选项比较', {'content': '同构选项', 'children': ['正误一致，打包处理']}]}]}

解释：“定义判断”包括“读定义”，“看选项（定义判断）”等，而“读定义”又包括“提问方式（读定义）”，“读题技巧（读定义）”等，而“提问方式”读定义包括“属于/不属于”，“正确不正确”等，以此类推。

2.3 neo4j的输入数据

Neo4j作为在存储图关系组织的数据时在查询修改等都较为高效，在大量三元组情况下，能极大提高效率，选用了目前仍在维护且较多人使用的neo4j3.5.0。将制作的数据转为下图所示的数据格式：



Entity\_id为实体对应的编号，label在行测模块是entity，在申论模块entity\_shenlun，第二张图是编号为:START\_ID的实体包括:END\_ID的实体，以这种格式导入知识图谱，并运行neo4j。

2.4 推荐算法要求的输入数据

Kg\_xtt.txt是生成的三元组记录，编号为第一列的实体包括编号为第三列的实体，第二列使用0开始计数表示实体关系。

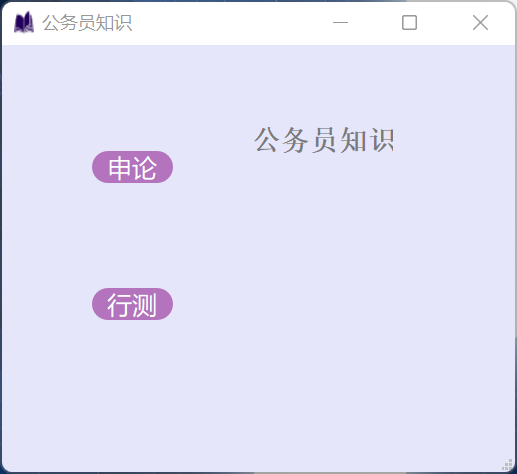
Ratings\_final.txt是用户对某个实体的点击与否，第三列1代表点击0，代表不点击，第一列从0开始编号表示用户，第二列则是实体的编号。

2.5 数据库postgres存储

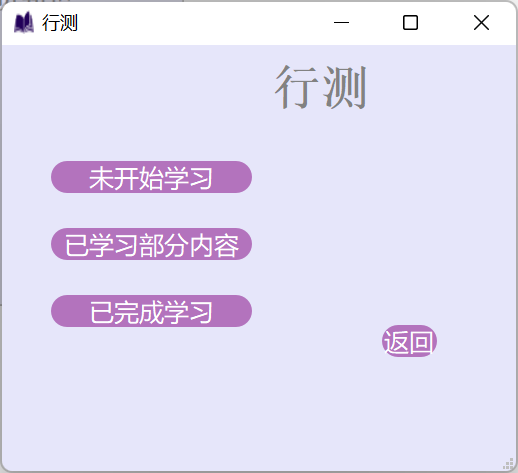
每条记录存储了时间戳，点击的实体编号，是否点击（点击为1，不点击为0），行测部分点击数据存在dataBase中，申论部分点击数据存储在dataBase\_shenlun中。实现逻辑是，在数据库点击数据数目少于三千条时，不使用推荐产生的结果（因为过少）。在已完成部分模块中，可以选择已完成学习的部分和未完成学习的部分，会在剩余的item中推荐。如果是点击我要学习模块，则是所有item都在推荐列表中。推荐界面点击后选确定为有效点击，负样本是前后相邻的各三个item，加起来六个item，假设下一次点击的是和上一次点击相邻，把数据库一天内标未为0不感兴趣的样本删去表示为1，即感兴趣。刷新按钮是如果对推荐的不感兴趣，就换一批最可能感兴趣的，一页显示20个item。

## 界面介绍

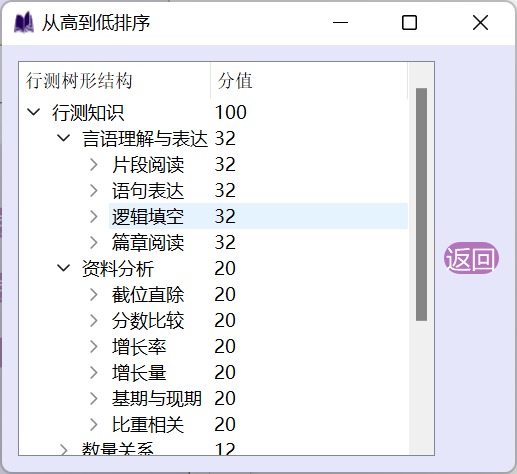
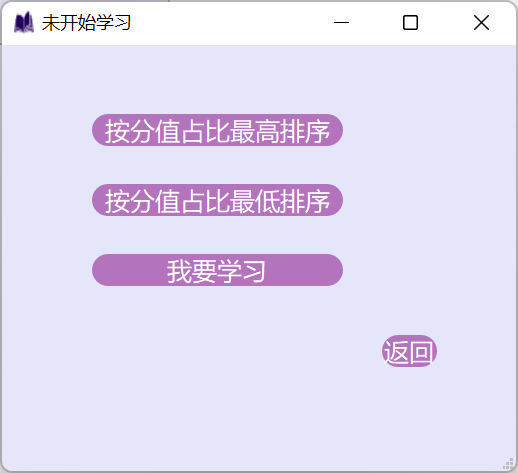
1. 首页是供选择的行测和申论两个模块：



1. 以点击“行测”为例，有未开始学习和已学习部分，已完成学习模块，“返回”按钮是返回首页，点击“已完成学习”会展示可查询的知识图谱界面可动态展开节点，点击“已学习部分内容”是可选择已学习的部分或者未学习的部分进行推荐，点击“未开始学习”则有按分值展开模块，也可根据所有的实体进行推荐。

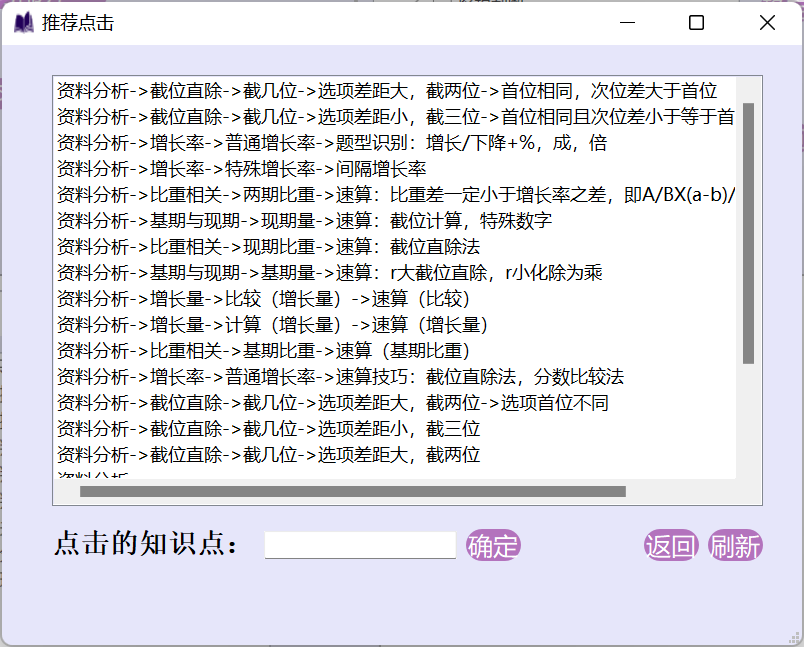
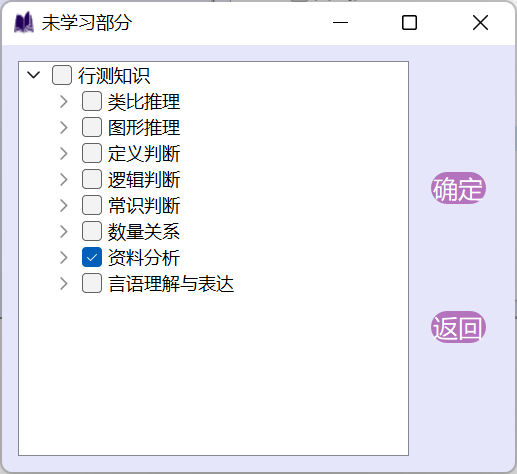
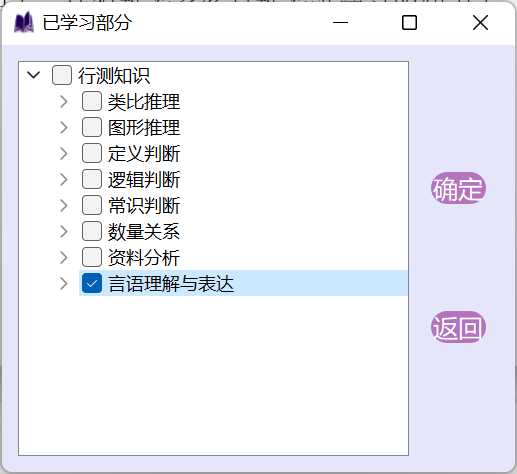
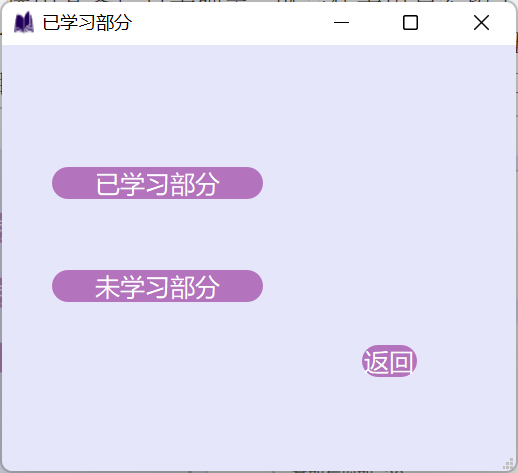


1. 点击“未开始学习”，其中点击“按分值占比最高排序”是从分值从高到低会展示行测部分的知识点和对应的分值，点击我要学习部分则会出现“推荐点击”界面，点击某一条后悬空会出现提示本条信息，选中某条后点击确定，就是代表用户希望了解此知识点，会显示在左下角的输出框中，为有效点击存入数据库postgres中。“刷新”按钮会刷新本页所有实体，意在在所有可推荐的条目中按照一定的推荐原则给用户换一批最有可能用户喜欢的知识点。



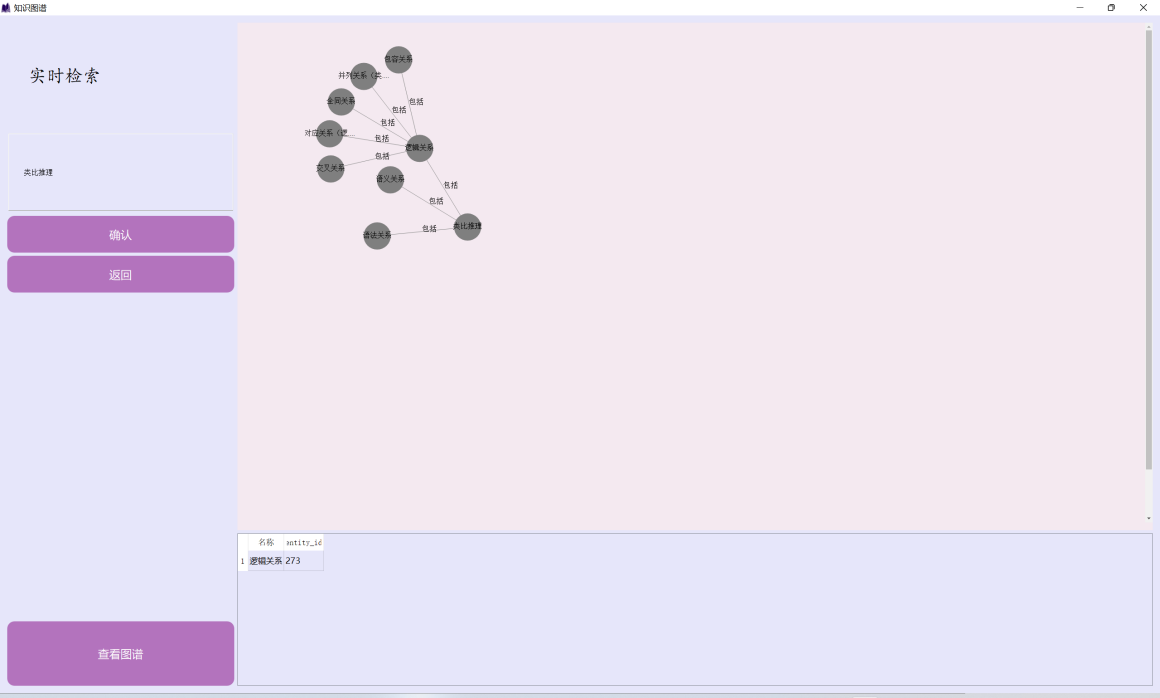


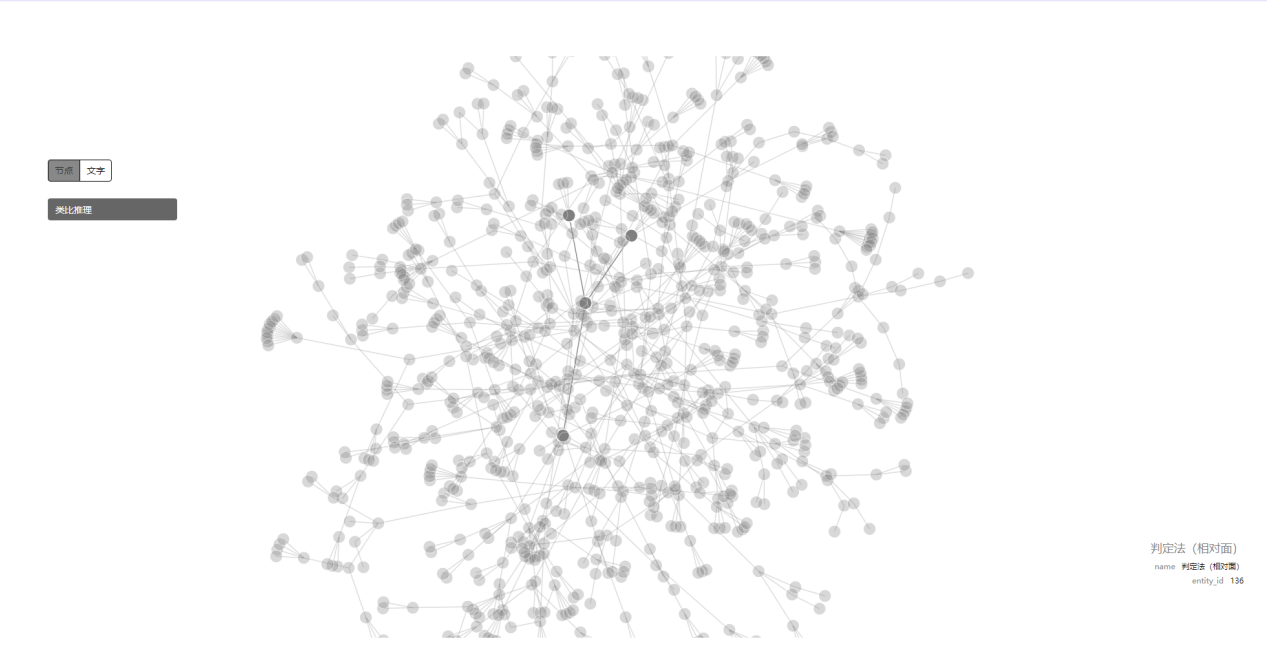
1. 点击“已学习部分”会出现两个模块，再点击“已学习部分”按钮，则会让用户在所有知识点中选择已学习的知识点，点击“确定”按钮后，使得该知识点不再推荐给用户，而点击“未学习部分”按钮是用户选择未学习的知识点，使得推荐系统只推荐没学习的知识点。并且，当用户点击父节点时，该父节点以下的子节点和儿孙节点等都意味着被选中，以避免用户为了只学习一个模块，却要点击所有知识点的尴尬局面。



1. 点击已完成学习部分，在输入框中键入要查询的知识点，会有实体提示补全信息，点击确定后，会在右侧界面弹出该实体对应圆点图形化的信息，并在下侧显示该节点对应在neo4j实体的编号，再次双击后，则会展开当前节点，对当前节点的子节点点击后可再次展开。点击“查看图谱”后可展示neo4j建立的行测部分整个知识图谱信息，图中有搜索框可键入搜索，搜索得到的节点和相邻以外的节点会变半透明，以显示结果，切换到文字的按钮，可显示节点对应文字。







## 系统逻辑

使用递归查找了父亲节点，使用并查集来判断是否是同一类数据，以此计算在推荐数据不够时按分数排序的结果。此次文件都在之前几乎都略有改动，get\_tree\_path.py运行可以得到每个节点对应路径，而我已经以字典的形式放在要用的文件里了。make\_num.py生成的要预测的item文件。运行process\_to\_json.py文件，将erec.csv文件转为json文件，以方便GUI的加载。

Mkr模型框架：

MKR由三个主要部分组成：推荐模块、KGE模块、交叉压缩单元。

左侧的推荐模块：将一个user和item作为输入，使用多层感知器(MLP)和交叉压缩单元分别提取user和item的特征，提取出的特征再一起送入另一个多层感知器，输入预测可能性。

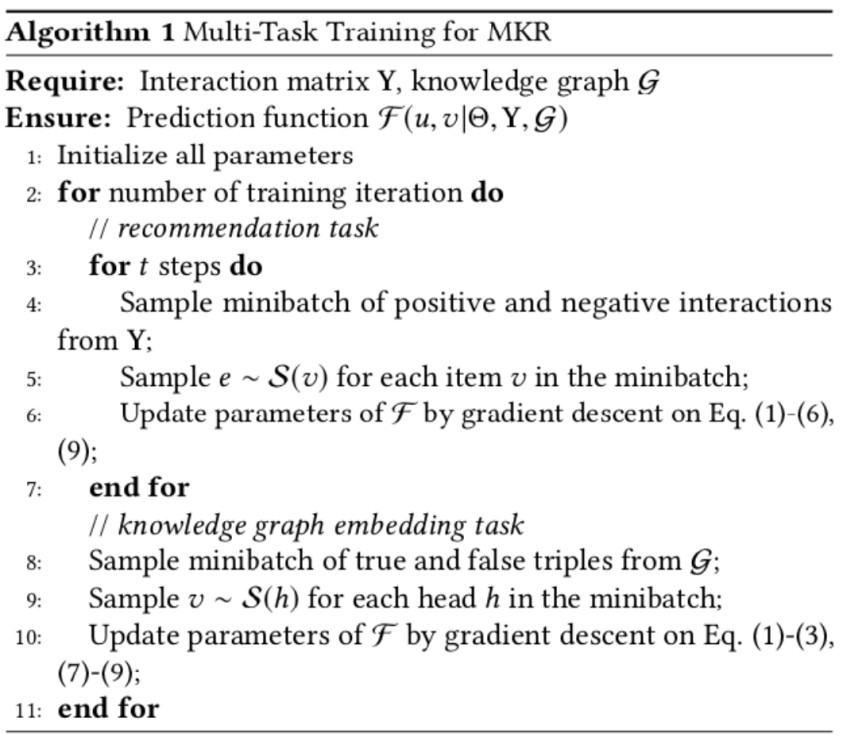
右侧的KGE模块：将一个head和relation作为输入，使用多层感知器(MLP)和交叉压缩单元分别提取head和relation的特征，使用head和relation计算出预测tail的表示，然后使用函数f计算预测tail和实际tail的相似度，作为KGE链路预测的能力分数。

中间的交叉压缩单元：是将推荐模块和KGE模块连接起来的关键，这个单元可以自动的学习RS中item和KG中entity的高阶交互特征。

学习算法如下：

在每次的训练迭代中包含两个阶段：推荐任务和KGE任务

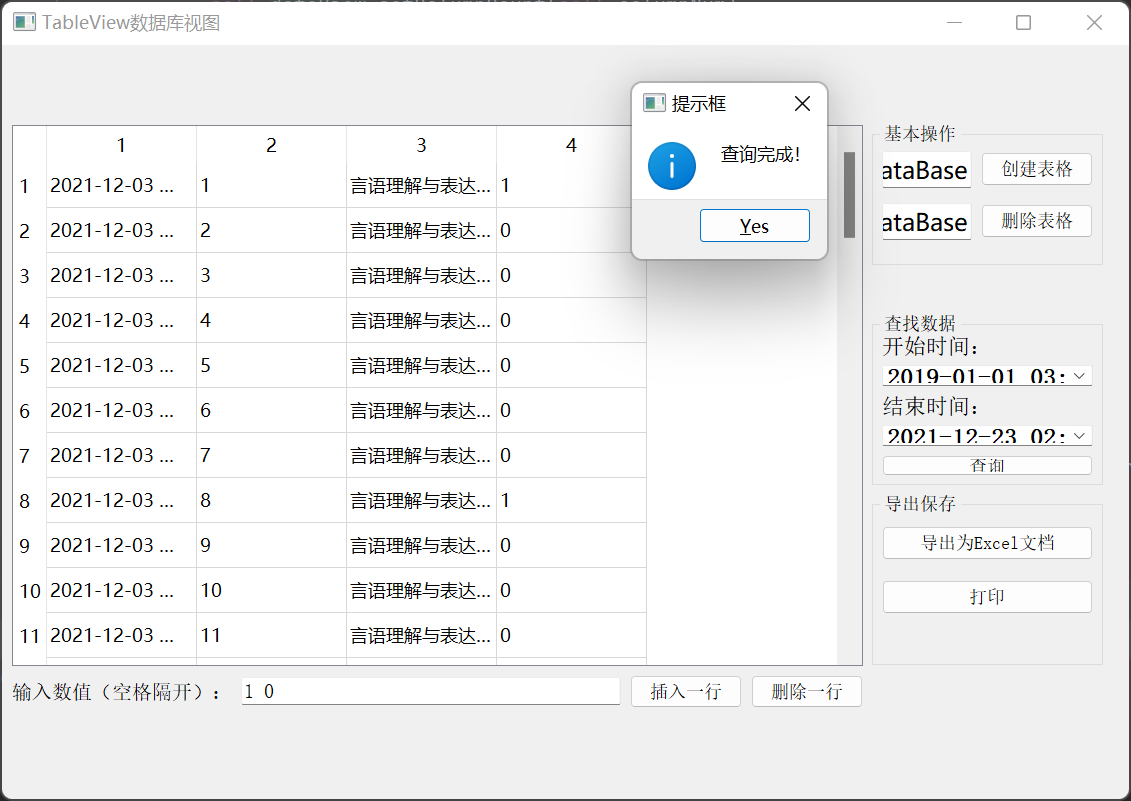
在每次的迭代中我们先重复训练推荐任务 t 次，再训练KGE任务1次，因为我们更加关注提升推荐的性能。



推荐算法融入知识图谱，可以在没有历史数据的情况下推荐内容，并且具有良好的可解释性。但由于需要考虑到不同用户，所以KGCN并不是非常合适，采用了MKR的交替学习。

## 后台操作

如果不想要目前数据库存储的内容，可以使用database\_weihu文件夹下的weihu\_gui.py，点击删除表格，就可以删除创建的数据库，而申论对应数据库名字是dataBase\_shenlun，行测是dataBase，可以在输入框中修改数据库的名字进行创建和修改，选择开始和结束时间可以查询到期间内的数据库记录，可以导出为excel，默认为D:/ExceptionData.csv，“删除一行”按钮可以删除日期最新的一条，而插入一行则只是作为演示。第四列代表点击或者不点击，第三列是点击的带上路径的实体，第一列是日期，第二列是计数。



## 项目运行

安装好requirements.txt里的环境后，和neo4j3.5.0导入数据，有两种处理方式：

1. 在dist文件夹下的zonghe文件夹的exe可以直接运行，要同时运行同目录mkr下的run.py，意在使推荐系统每隔五分钟运行一次，及时处理用户点击数据。
2. 直接运行zonghe.py和项目根目录下的mkr文件夹下的run.py文件。