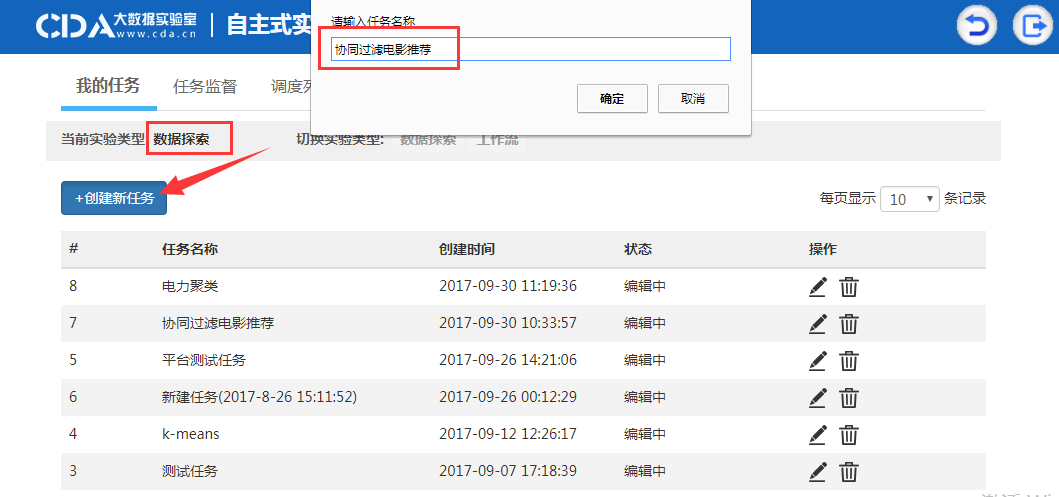
数据探索操作说明之协同过滤电影推荐

本实验是基于矩阵的协同过滤电影推荐算法的使用，主要是将100多万条的原始电影数据（用户名：userid、电影名：movied、偏好度：rating）进行协同过滤模型训练，然后使用建立好的模型对3万条的电影测试数据进行模型预测，预测这3万条电影数据的偏好度，以此来进行高偏好度电影推荐。

1、创建新任务

点击“数据探索”中的“创建新任务”，编辑任务名称为：协同过滤电影推荐，然后点击“确定”。



2、配置第一个组件：读数据表

（1）新建组件：读数据表

从左侧将“读数据表”组件拖拽到工作界面。



（2）配置组件

双击进行配置或右击选择“编辑”进行配置。

数据库：default

数据表：user\_movie\_full

单击“确定”。



（3）右击选择的“查看数据”，即可查看导入的数据。



本数据包含用户id：userid，电影名id：movieid，以及用户对电影的评分数据：rating，评分数值从1-5分为5个级别，评分越高说明用户对该电影的兴趣度越高。



3、配置第二个组件：全表统计

（1）新建组件：全表统计

从左侧将“全表统计”组件拖拽到工作界面，并与“读数据表”节点连接。



（2）配置组件

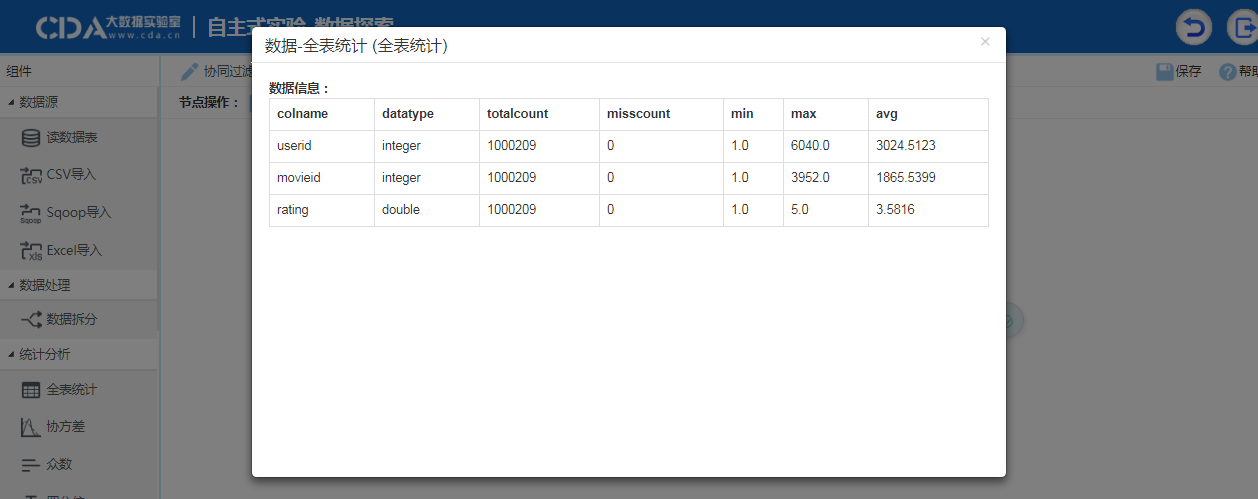
双击进行配置或右击选择“编辑”进行配置。勾选所有字段。



（3）右击选择“运行此节点”，等待任务运行完成。



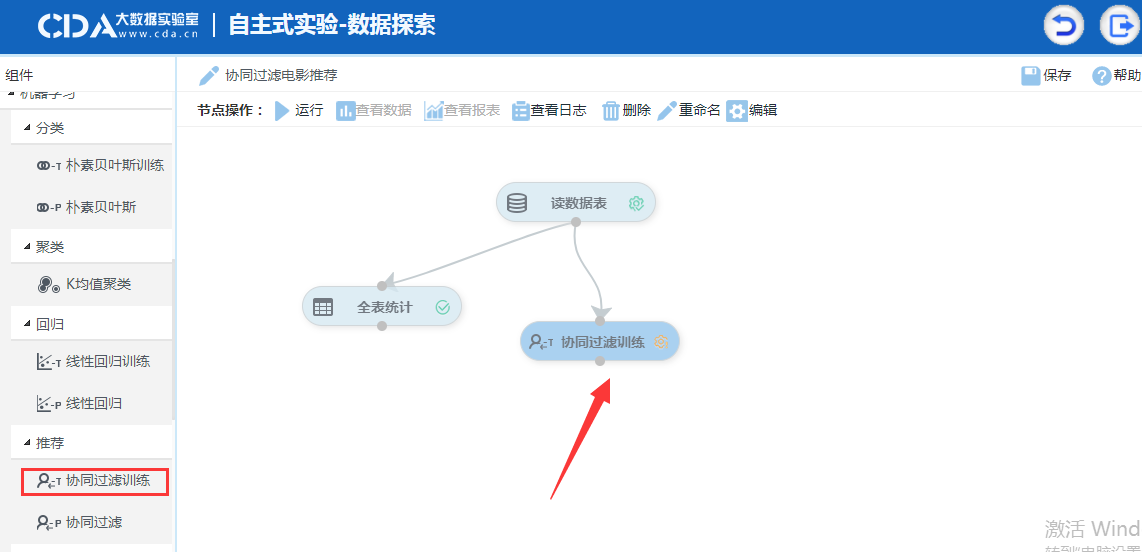
（4）右键点击“查看数据”，可以看到数据量，也可以看到各个属性的最大值、最小值、平均值。



4、配置第三个组件：协同过滤训练

（1）新建组件：协同过滤训练

从左侧将“协同过滤训练”组件拖拽到工作界面，并与“读数据表”相连接。



（2）配置组件

双击进行配置或右击选择“编辑”进行配置。

输入：用户标识：userid；产品标识：movieid；评价/得分：rating；

参数：设置最大迭代次数：100次

输出：自动为用户配置好输出路径。

（3）右击选择“运行此节点”，等待任务运行完成。



5、配置第四个组件：读数据表

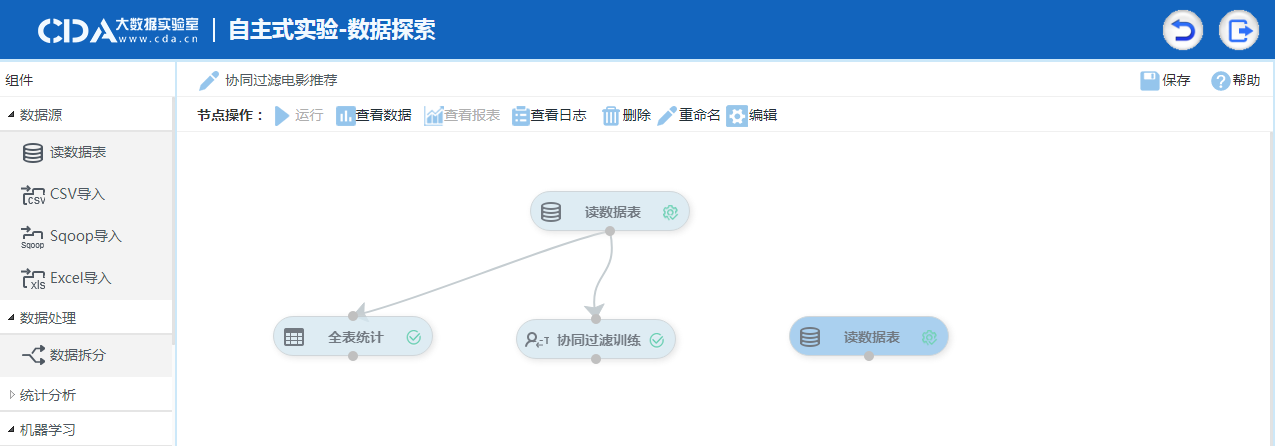
（1）新建组件：读数据表

再新建一个“读数据表”组件。从左侧将“读数据表”组件拖拽到工作界面。此为使用训练好的模型进行电影评分预测的数据，

数据库：default

数据表：user\_movie\_full\_test

点击“确定”。



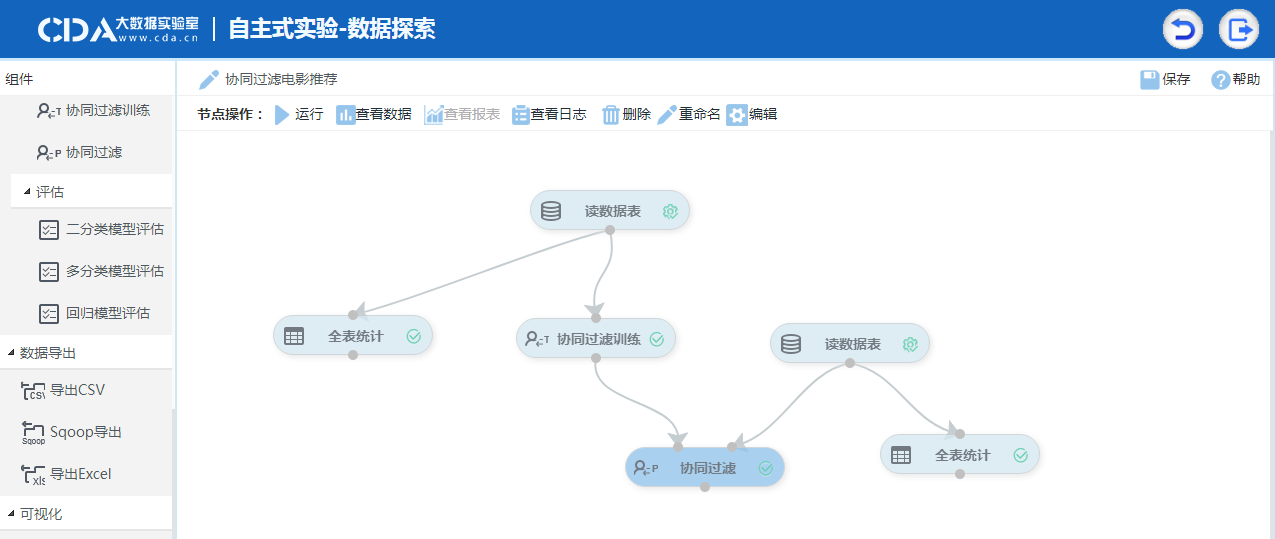


同理可以查看此数据表的数据，可以对此数据表进行全表统计

6、配置第六个组件：协同过滤

（1）新建组件：协同过滤

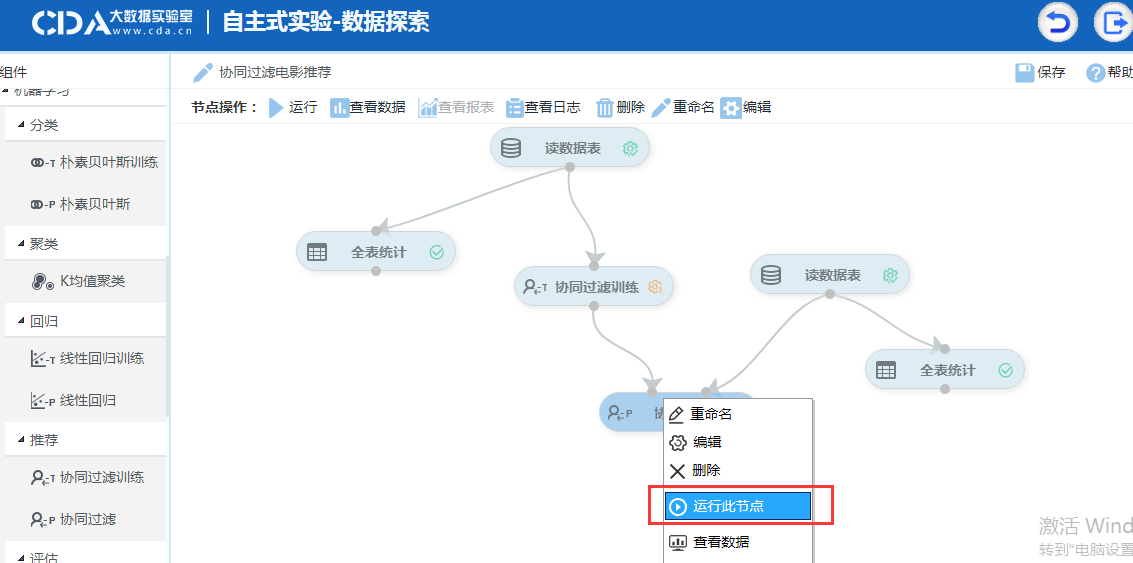
从左侧将“协同过滤”组件拖拽到工作界面，此组件有3个节点，上方左节点用来连接训练模型节点，接收模型参数，右节点用来连接测试数据，也就是需要进行预测的数据。下方节点可以连接可视化组件进行可视化展现，也可连接数据导出组件，将数据导出到不同地方进行进一步处理。

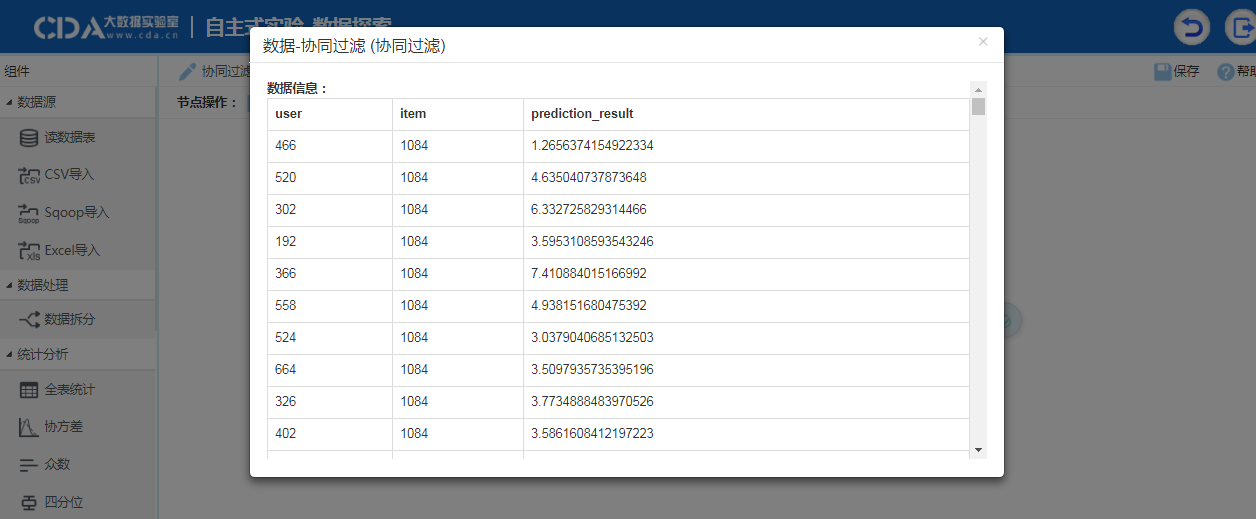
（2）配置组件

双击进行配置或右击选择“编辑”进行配置。此“协同过滤”组件输入输出已自动配置好，直接点击“确定”即可。



（3）右击选择“运行此节点”，运行成功后，再次右击选择“查看数据”，即可看到预测结果，我们可以选取预测值接近5的电影推荐给对应的用户。





7、配置第七个组件：导出Excel

（1）新建组件：导出Excel

从左侧将“导出Excel”组件拖拽到工作界面，并与“协同过滤”节点相连接。可以将实验结果导出到本地Excel中做进一步的处理和操作。

