

广电营销智能推荐

1 项目背景

广播电视行业，是指专业从事广电设备的生产，研究，探索，销售的单位，主要包括：摄、录、监、采、编、播、管、存等主要方面。伴随互联网和移动互联网的快速发展，各种网络电视 / 视频应用（爱奇艺、腾讯视频、乐视视频、芒果 TV 等）遍地开花，越来越多人群的电视观看行为正发生变化，由之前的传统电视媒介向电脑、手机、平板端的网络电视转化。

在这种新形势下传统广播电视运营商感受到的危机也越来越明显。另一方面，“三网融合”为传统广播电视运营商带来发展机遇，特别是随着超清 / 高清交互数字电视推广，广播电视运营商可以和家庭用户实现信息实时交互，家庭电视也逐步变成多媒体信息终端。

本项目的广电运营商是经市委、市政府授权，专门负责建设、维护和管理该行政区域内广播电视网络的主要运营机构。经过十多年的发展，该集团已成为华南地区最大规模的广播电视网络运营商之一。

目前该集团已建成完整覆盖各区（县级市）的有线传输与无线传输互为延伸、互为补充的广电宽带信息网络，实现了城区全程全网的双向覆盖，为广大市民提供有线数字电视、互联网接入服务、高清互动电视、移动数字电视、手机电视、信息内容集成等多样化、跨平台的信息服务。其信息数据收集过程如图 1-1 所示，首先每个家庭收看电视节目都需要有一个机顶盒来进行收视节目的接受和交互行为（如点播行为、回看行为）的发送，然后会发送至每个区域的光机设备（进行数据传递的中介）汇集该区域的信息数据，最后再由各光机设备发送至数据中心进行数据整合、存储在大数据平台中。

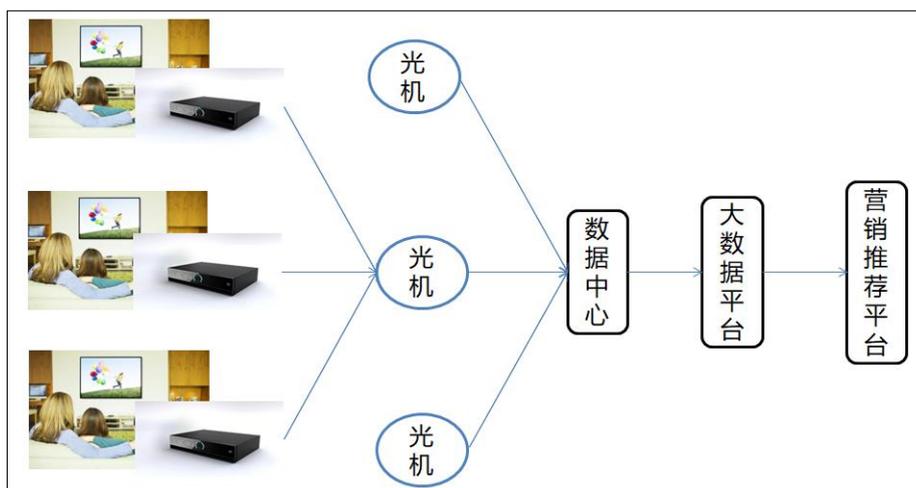


图 1-1 信息数据传递过程

由于在已建设的大数据平台积累了大量用户基础信息和用户观看记录信息等数据，需在此基础上进一步挖掘出数据价值、形成客户画像，以提升客户体验，并实现精准的营销推荐。总而言之，智能营销推荐服务可以为用户提供个性化的服务，改善用户浏览体验，增加用户黏度，从而使用户与企业之间建立稳定交互关系，实现客户链式反应增值。

2 项目目标

1. 构建 Popular 流行度推荐模型
2. 构建协同过滤推荐模型

3 项目步骤

3.1 新建数据源

进入 PB 平台，点击左侧“数据源”导航栏，点击“新建数据源”，在弹出的下拉窗口中选择“数据来源于数据库”，如图 3-1 所示。



图 3-1 新建输入源

在弹出的窗口中填入如下信息：

1. 新建目标表：`case_tv_recommend`
2. 数据库类型选择 PostgreSQL
3. URL 链接：`jdbc:postgresql://192.168.0.95:5432/case_tv`
4. 用户名：`postgres`
5. 密码：`root`
6. SQL 语句：`SELECT * FROM (SELECT phone_no, program_title, ROW_NUMBER () OVER (PARTITION BY phone_no) AS ROW FROM media4) T WHERE ROW <= 5`

如图 3-2 所示。

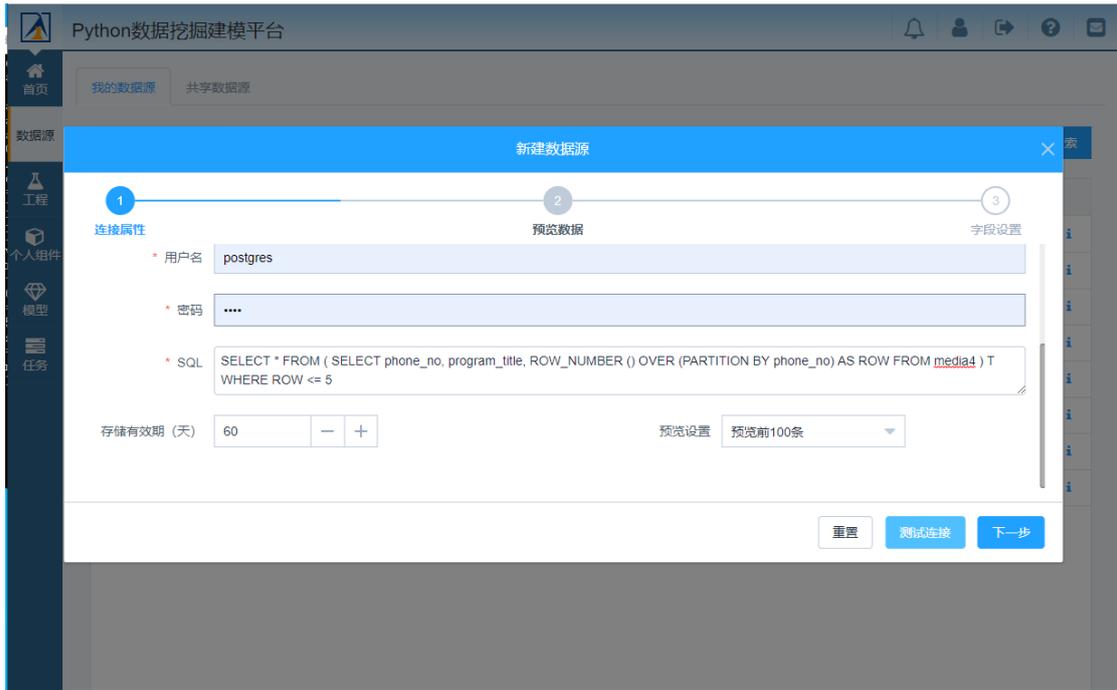


图 3-2 case_tv_recommend 连接属性

点击测试连接，提示连接成功，点击下一步进入数据预览，如图 3-3 所示。

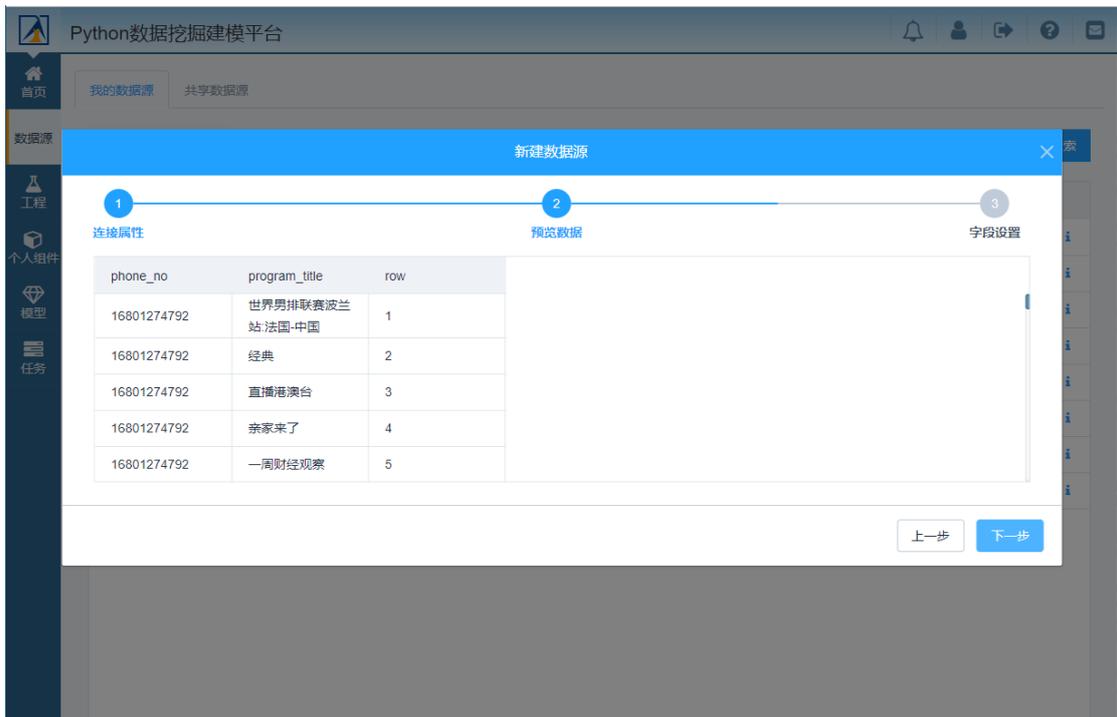


图 3-3 case_tv_recommend 数据预览

点击下一步，进入字段设置界面，将 phone_no 的类型改为字符型，点击确定，创建数据源成功。



图 3-4 case_tv_recommend 字段设置

创建成功后，点击同步按钮，将数据同步至平台，成功后可点击消息铃，查看同步信息，如图 3-5 所示。

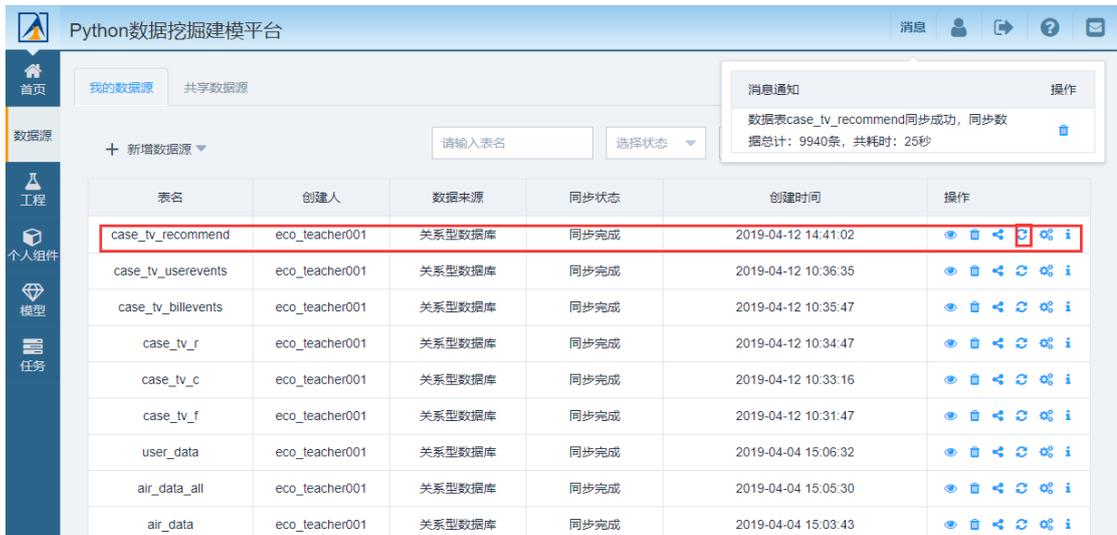


图 3-5 同步成功

注意：所有数据源字段设置过程中均需将 phone_no 改为字符型，创建完成后需要进行同步操作。

3.2 工程配置

3.2.1 输入源

点击左侧导航栏中的“工程”，找到“我的工程”，在目录“广电大数据营销推荐”，上右击，点击新建工程“智能推荐”，如图 3-6 所示。

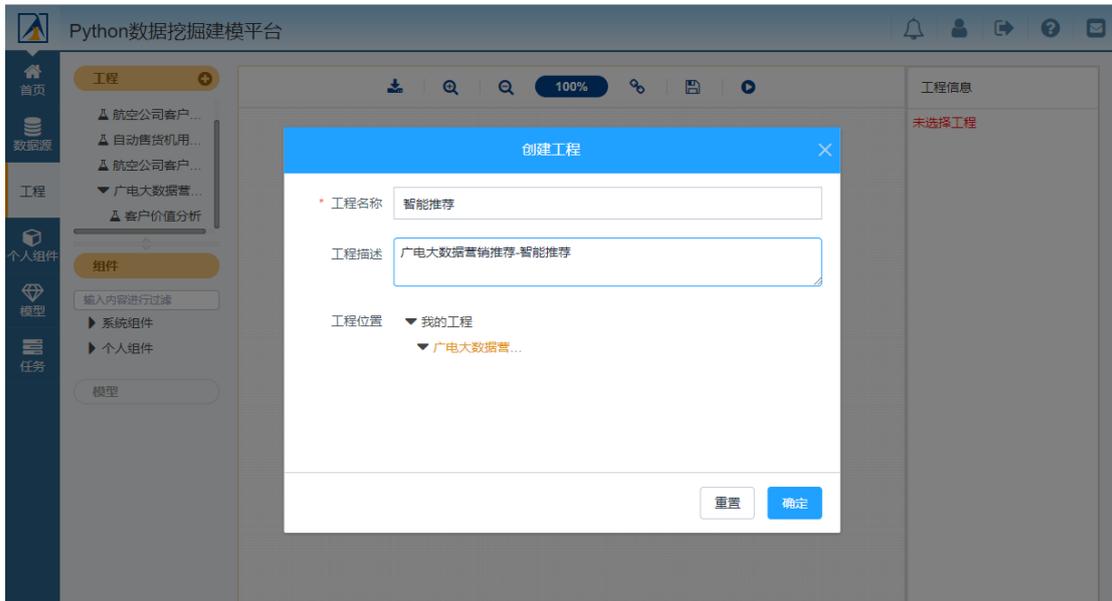


图 3-6 新建工程

点击右侧系统组件，拖入“输入源”组件，在右侧参数设置的“数据表”中填入“case_tv_recommend”，如图 3-7 所示。

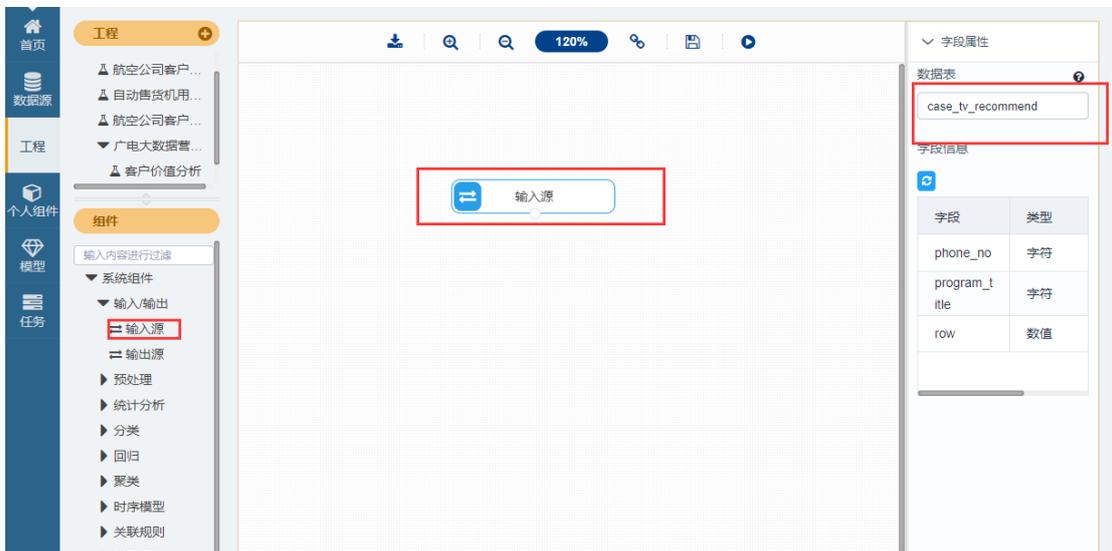


图 3-7 拖入输入源

3.2.2 记录去重

点击右侧“预处理”，拖动“记录去重”，并将其与“输入源”相连，设置字段属性，

如图 3-8 所示。

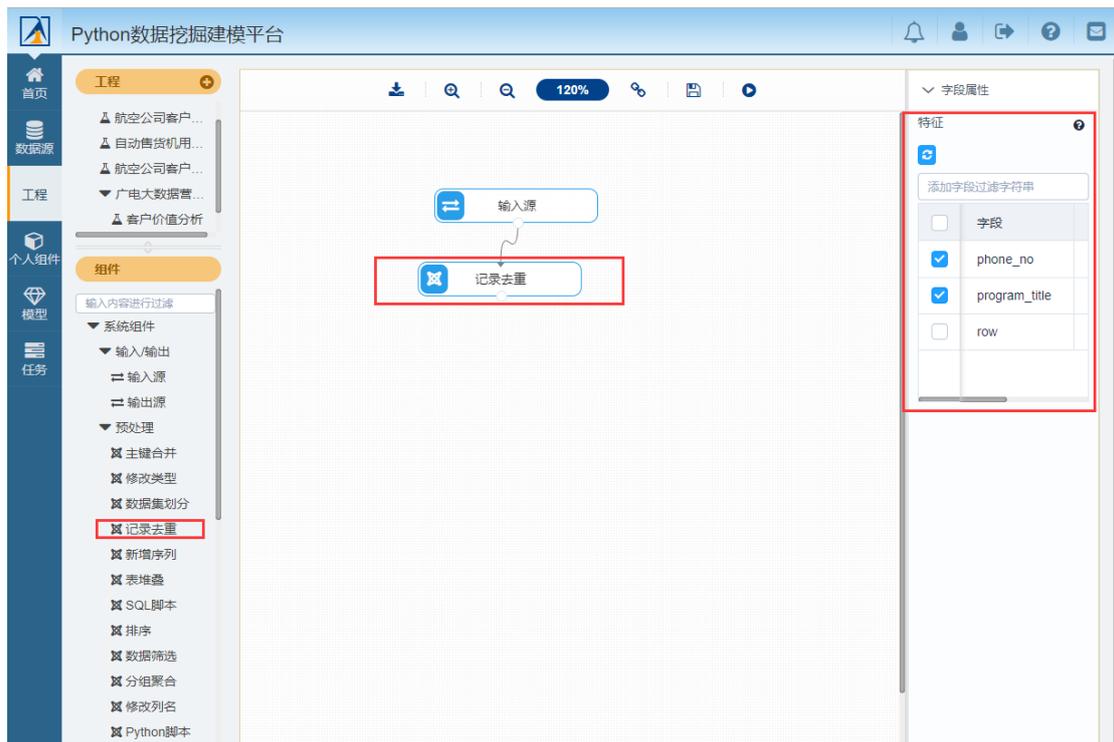


图 3-8 记录去重

3.2.3 频数统计

点击右侧“统计分析”，拖动“频数统计”，并将其与“记录去重”相连，重命名为“Popular”，设置“字段属性”，如图 3-9 所示。

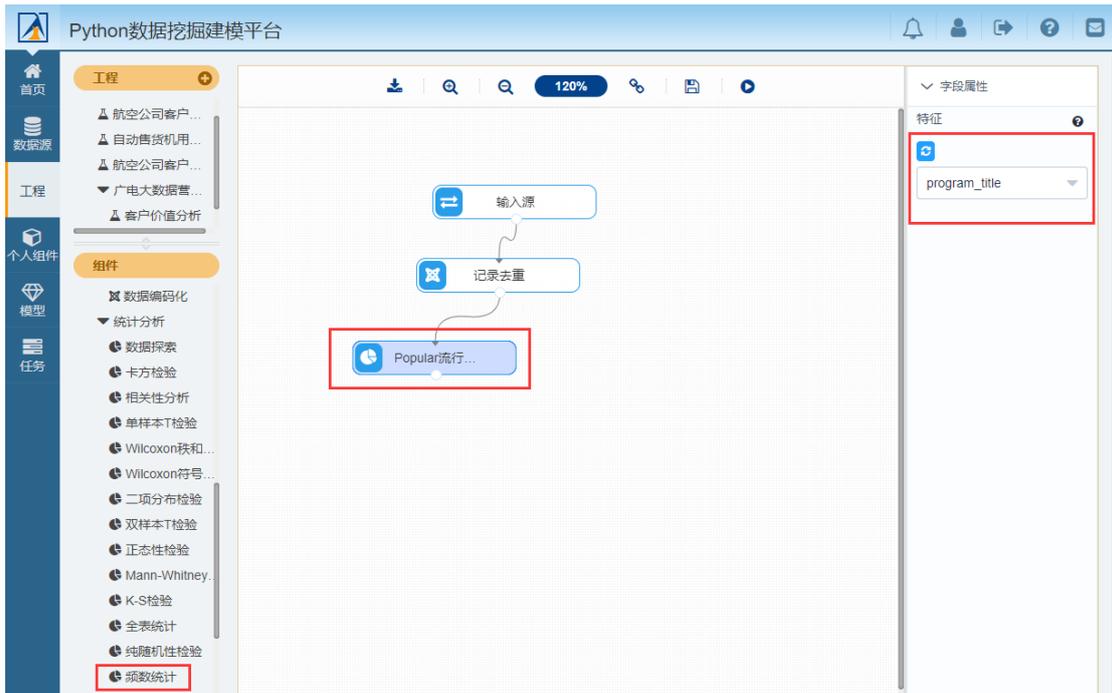


图 3-9 频数统计

3.2.4 Python 脚本

点击右侧“预处理”，拖动“Python 脚本”，并将其与“记录去重”相连，并重命名为“协同过滤”如图 3-9 所示。

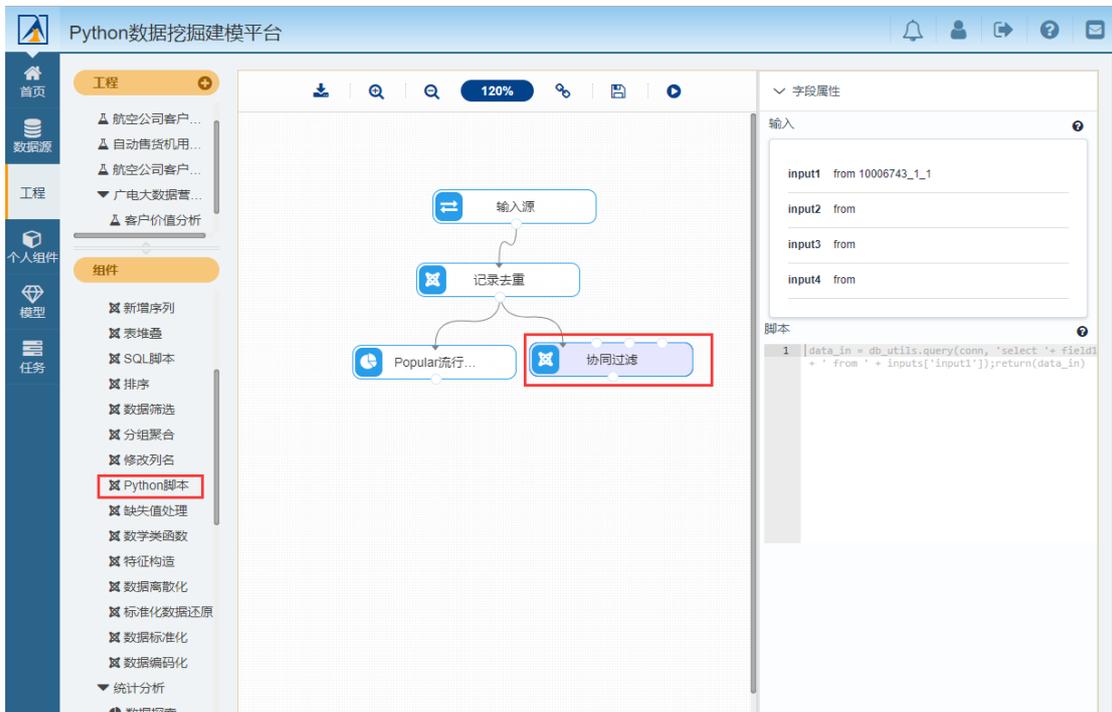


图 3-10 Python 脚本

单机 Python 脚本，在“脚本”中填入代码，如代码 3-1 所示。

代码 3-1 Python 脚本

```
data_in = db_utils.query(conn, 'select * from ' + inputs['input1'])
import pandas as pd
from sklearn.metrics.pairwise import pairwise_distances

data_in['value'] = 1
mat1 = data_in.pivot_table(index='phone_no',columns='program_title') # 透视表
mat1.columns = [i[1] for i in mat1.columns]
mat1.fillna(0,inplace=True) # 0 填充
df_matrix1 = mat1
df_matrix1 = df_matrix1 / df_matrix1.sum(axis=0)*5
item_similarity = 1-pairwise_distances(df_matrix1.T, metric='cosine')

# 对角线设为 0
a = range(item_similarity.shape[0])
item_similarity[a,a] = 0
item_similarity = pd.DataFrame(item_similarity)
item_similarity.index = item_similarity.columns = df_matrix1.columns

phone_test = data_in.phone_no.unique()

result1 = pd.DataFrame()
for i in range(len(phone_test)):
    res1 = pd.DataFrame({'phone':[phone_test[i]] * 5,
                        'program':(df_matrix1.iloc[i]
item_similarity).sum(axis=1).sort_values(ascending=False).index[:5].tolist())})
    result1 = result1.append(res1)
return(result1)
```

点击运行，查看推荐程序运行结果，如所示。

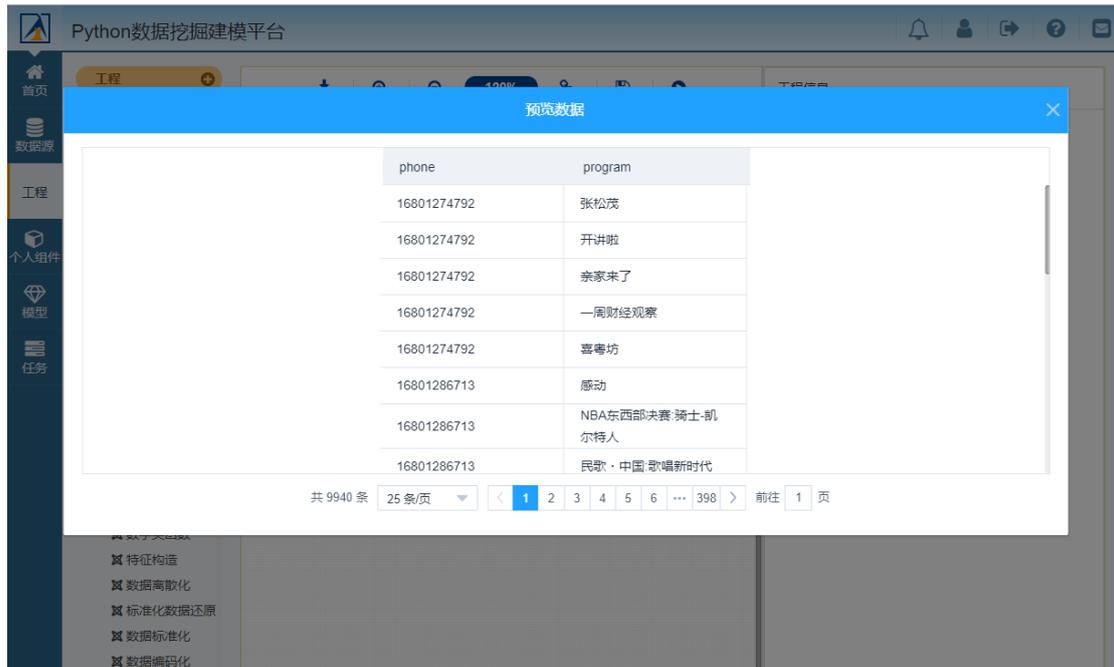


图 3-11 协同过滤结果预览