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 Popuular 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。

	Python数据挖掘建模平台			• 0							
▼ 首页	我的数据源 共享数据源	我的数据源 共享数据源									
数据源	+ 新增数据源 ▲										
「五程	■ 数据来源于文件	选	释状态 ▼ 请选择的	创建时间	(③ 搜索					
€	● 数据来源于数据库	创建人	数据来源	同步状态		创建时间					
₩ Call	● 数据来源于非结构化文件	teacher009	结构化文件	同步完成	201	19-03-27 11:45:					
模型	air_data	teacher009	结构化文件	同步完成	201	9-03-27 10:38:					
TEI 任务	user_data	teacher009	结构化文件	同步完成	201	9-03-21 19:27:					

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
- 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。

	Python数据挖掘函	建模平台		4 8 🗭	8	
▼ 首页	我的数据源 共享	致循環				
; 数据源 ·		新建数据源			×	 큧
	1	2		3		
父 个人组件	连接属性 * 用户名	预览数据 postgres		字段设置		4
(◆ 模型	* 密码					i
日子	* SQL	SELECT * FROM (SELECT phone_no, program_title, ROW_NUMBER () OVER (PARTITION BY phone_ WHERE ROW <= 5	_no) AS ROW F	ROM media4) T		i i
	存储有效期(天)	60 一 十 预览设置 预览前100条	Ŧ]	°	i
			重置	测试连接	5	

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。

	Python数据挖	掘建模平	石			4		8
谷 首页	我的数据源	共享数据源						
数据源					新建数据源			× ^索
「工程	1				2		3	
	连接属性				预览数据		字段设置	i
个人组件	phone_no	р	rogram_title	row				i
模型	1680127479	2 世 实	世界男排联赛波兰 5:法国-中国	1			1	i
日子	1680127479	2 绍	2典	2				i
	1680127479	2 1	王播港澳台	3				i
	1680127479	2 芳	家来了	4				i
	1680127479	2 –	-周财经观察	5				i
								i
						上一	-步 下步	

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。

	Python数据	它掘	建模平台										\bigtriangleup		0	
▼ 首页	我的数据源	共享	数据源													
数据源							新建数据	亰							×	52
「工程	1						2							3		
	连接属性 注意:【字段:	818	能是以字母开头 由	小写英文字母	数字 下划约	彩印成	预览数据	2						字段设置	i	
	原字段		字段名	类	型		长度		精	度		备注			i	
▼ 模型	phone_no		phone_no	字符		25	55 —	+	0		+				i	
	program_title		program_title	数值		25	55 —	+	0		+				i	
任务	row		row	字符		25	55 —	+	0		+				i	
				日期											i	
				时间											i	
															i	
													Ŀ	一步 确立		

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。

	Python数据挖掘建模 ^I	平台				消息	8		0	
▲ 首页	我的数据源 共享数据源				消息通知				ł	操作
数据源	+ 新增数据源 ▼		请输入表名	选择状态	数据表case_tv 据总计: 9940	数据表case_tv_recommend同步成功,同步数 据总计: 9940条,共耗时: 25秒			Û	
「工程	表名	创建人	数据来源	同步状态	创建时间		操作			
	case_tv_recommend	eco_teacher001	关系型数据库	同步完成	2019-04-12 14:4	1:02	• 1	4	C 📽	i
个人组件	case_tv_userevents	eco_teacher001	关系型数据库	同步完成	2019-04-12 10:3	6:35	• 1	4	c 🐝	i –
♥ 模型	case_tv_billevents	eco_teacher001	关系型数据库	同步完成	2019-04-12 10:3	5:47	• 1	4	c 🕫	i
	case_tv_r	eco_teacher001	关系型数据库	同步完成	2019-04-12 10:34	4:47	• 1	4	c 🕫	i
任务	case_tv_c	eco_teacher001	关系型数据库	同步完成	2019-04-12 10:3	3:16	• 1	4	c 🐝	i
	case_tv_f	eco_teacher001	关系型数据库	同步完成	2019-04-12 10:3	1:47	• 1	4	C 🕫	i –
	user_data	eco_teacher001	关系型数据库	同步完成	2019-04-04 15:00	6:32	• 1	4	c 🕫	i
	air_data_all	eco_teacher001	关系型数据库	同步完成	2019-04-04 15:0	5:30	• 1	4	c 🕫	i
	air_data	eco_teacher001	关系型数据库	同步完成	2019-04-04 15:03	3:43	• 1	4	C 08	i

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。

	Python数据挖掘建模	平台 · · · · · · · · · · · · · · · · · · ·	≙ 8 6 8 8
▲ 首页		🛓 I Q I Q 🚺 100% % I 🖹 I O	工程信息
》 数据源	▲ 航空公司客户 ▲ 自动售货机用 ▲ 航空公司客户	创建工程	未选择工程
工程	▼ 广电大数据营	* 工程名称 智能推荐	
● 个人组件	组件	工程描述 广电大数据营销推荐-智能推荐	
模型	输入内容进行过滤 ▶ 系统组件 ▶ 众人组件	工程位置 ▼ 我的工程 ▼ 「由+25週等	
任务	模型		
		重置機定	

图 3-6 新建工程

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

徐 首页	工程 〇	📩 🛛 Q 📄 Q 👥 120% 🔗 🖉 📄 💽	~ 字段属性	
● ● ● ● ● ● ● ●	▲ 航空公司客户 ▲ 自动售货机用 ▲ 航空公司客户		数据表 case_tv_recom	@ mend
工程	▼ 广电大数据营…	10 X00	字段信息 2	
个人组件	组件		字段	类型
愛愛	 输入内容进行过速 ▼ 系统组件 ▼ 输入/输出 		phone_no program_t itle	字符 字符
任务	 ➡ 輸入源 ➡ 輸出源 ▶ 预处理 		row	数值
	 > 统计分析 > 分类 > 回归 > 聚关 > 四生体型 			
	 ▶ 所序模型 ▶ 关联规则 ▶ ###107.00 			

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。

	Python数据挖掘建模平台		↓ ▲ 🗭 0 🖸
▼ 谷 首页	工程	초 Q Q 120% % 🖹 D	◇ 字段属性
上前 9000 题源 工程 ● 今人供 ● 一楼 1846 66	上程 C ▲ 航空公司寄户 ▲ 航空公司寄户 ▲ 航空公司寄户 ✓ 广电大数据常 ▲ 客户价值分析 ● 好价值分析 ● 好价值分析 ● 新人內衛进行过途 ● 系统组件 ● 輸入衛出 二 输入第出 二 输入第出 二 输入第出 二 输入第出 ※ 系统组件 ● 輸入第出 二 输出源 ● 放处理 M 信政关型 ▲ 新增序列 ※ SQL與本 ※ 封那湾 ※ 数型筛选 ※ 公理 ※ SQL與本 ※ 小理序 ※ 授加 ※ 公理 ※ SUL與本 ※ 小生 ※ 第 ※ 公員 ※ 第 ※ 「任 ※ 「任 ※ 「任 ※ 「任 ※ 「任		 > 字段屬性 特征 ⑦ 0 0
	XX Python脚本		

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。

	Python数据挖掘建模平台		↓ 🌡 🕩 🛛 🖾
▲		🛓 I Q I Q 120% % I 🖺 I 🛛	> 字段属性
首页 ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ●	▲ 航空公司案户 ▲ 目动售货机用 ▲ 加空公司案户 ● 广电大数据营 ▲ 客户价值分析	へのの目的に くのの目的に ・・・・・・・・・・・・・・・・・・・・・・・・・・・・・	◇ 字段属性 \$\$\$\phi\ input1 from 10006743_1_1 input2 from input3 from input4 from Bp本 ● 1 [data_in = db_utlls.query(conn, 'select '+ fieldit] 1 [data_in + inputs('input1'));return(data_in)
	C 数据探索	-	

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。

```
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)
```

*

```
代码 3-1 Python 脚本
```

点击运行,查看推荐程序运行结果,如所示。

	Python数据挖掘建模平台				• 0	
▼ 谷 首页		420%		Tipjéé		
劉据源		1				
		phone	program			
工程		16801274792	张松茂			
		16801274792	开讲啦			
		16801274792	亲家来了			
模型		16801274792	一周财经观察			
		16801274792	喜粤坊			
1±95		16801286713	感动			
		16801286713	NBA东西部决赛:骑士-凯 尔特人			
		16801286713	民歌・中国:歌唱新时代			
	共 9940 条	25条/页 🔻 🤇 1 2	3 4 5 6 398 >	前往 1 页	_	
	 スリンに取 資 特征均迭 減 数据商数化 資 标准化数据还原 減 数据标准化 				ľ	

图 3-11 协同过滤结果预览