

第六届“泰迪杯”B 题论文附件提交说明

建议参赛队对所给数据集进行划分,例如按时间顺序将用户收视信息数据划分为前期和后期两部分,使用前期的用户收视信息数据来分析用户收视偏好,建立用户和产品标签,在此基础上建立推荐模型给出产品的营销推荐方案,并根据用户后期的收视信息数据来评价所给出的推荐方案,可使用精确率 (Precision), 反馈率 (Recall), F1 系数等评估指标。

问题一:

提交影视产品的营销推荐方案,其中推荐结果按推荐指数排序(最多 Top20)。

要求给出推荐结果表,格式如示例 1 问题一推荐结果表示例所示(要求提交格式为 CSV 文件,命名为队号+问题一推荐结果表)。

用户号	产品名称	推荐指数
11039	一树桃花开	0.45
11039	那年花开月正圆	0.75
11039	食戟之灵 第二季	0.3
11039	摔跤吧!爸爸	0.16
11039	晓说2017	0.14
11039	小马宝莉中文版	0.33
11039	神兽金刚 第一季	0.25
.....
11040	食戟之灵 第二季	0.56
11040	蒙面唱将猜猜猜 第二季	0.23
11040	寰宇视野	0.6

示例 1 问题一推荐结果表示例

问题二:

问题二中需要对用户和产品进行分类(贴标签),并给出营销推荐模型及推荐结果和评估结果,其中推荐结果按推荐指数排序(最多 Top20)。

请给出用户标签体系表和产品标签体系表,格式如示例 2 用户标签体系表示例(要求提交格式为 CSV 文件,命名为队号+用户标签体系表)、示例 3 产品标签体系表示例所示(要求提交格式为 CSV 文件,命名为队号+产品标签体系表)。

一级标签	二级标签	三级标签	四级标签
基本特征	家庭成员	男性		
基本特征	家庭成员	女性		
			
基本特征	观看时间段	早上		
基本特征	观看时间段	上午		
			
收视偏好	体育	足球	英超	
收视偏好	体育	足球	中超	
			

示例 2 用户标签体系表示例

一级标签	二级标签	三级标签
基本特征	剧场	现代剧	
基本特征	剧场	古装剧	
基本特征	剧场	战争剧	
		
适用人群	家庭成员	老人	
适用人群	家庭成员	儿童	
适用人群	家庭成员	青年	
		

示例 3 产品标签体系表示例

建议给出用户收视数据标签如示例 4 用户收视数据标签示例所示(要求提交格式为 CSV 文件,命名为队号+用户收视数据标签表),要求用示例 5 产品数据标签示例(要求提交格式为 CSV 文件,命名为队号+产品数据标签表)和示例 6 用户数据标签示例所示的格式(要求提交格式为 CSV 文件,命名为队号+用户数据标签表)分别给出产品数据标签和用户数据标签,一行数据多个维度可能有多个标签。

10004	20171010	24	中央少儿-高清	2017/10/9 20:31:00	2017/10/10 8:32:00	基本特征	观看时间段	晚上
10004	20171012	24	中央少儿-高清	2017/10/12 20:31:00	2017/10/12 21:27:00	基本特征	家庭成员	儿童
10004	20171013	24	中央少儿-高清	2017/10/13 9:58:00	2017/10/13 18:54:00	收视偏好	娱乐	动画
10004	20171013	14	中央2台-高清	2017/10/13 9:13:00	2017/10/13 9:58:00	基本特征	家庭成员	老人
10004	20171013	15	中央3台-高清	2017/10/13 8:52:00	2017/10/13 9:13:00	收视偏好	娱乐	综艺
10004	20171017	24	中央少儿-高清	2017/10/17 11:41:00	2017/10/17 13:42:00	基本特征	观看时间段	中午
10004	20171020	14	中央2台-高清	2017/10/20 9:07:00	2017/10/20 9:59:00	基本特征	观看时间段	上午
10004	20171020	14	中央2台-高清	2017/10/20 9:07:00	2017/10/20 9:59:00	收视偏好	财经	纪实
10004	20171024	15	中央3台-高清	2017/10/24 11:11:00	2017/10/24 18:33:00	收视偏好	娱乐	综艺
10004	20171025	24	中央少儿-高清	2017/10/25 11:24:00	2017/10/25 13:24:00	收视偏好	娱乐	动漫
10004	20171027	14	中央2台-高清	2017/10/27 9:12:00	2017/10/27 10:33:00	收视偏好	财经	纪实
10004	20171029	28	广东体育-高清	2017/10/29 21:09:00	2017/10/29 21:36:00	收视偏好	体育	篮球
10004	20171029	28	广东体育-高清	2017/10/29 21:38:00	2017/10/29 21:39:00	收视偏好	新闻	体育
10004	20171029	28	广东体育-高清	2017/10/29 20:31:00	2017/10/29 20:35:00	基本特征	家庭成员	男性、中青年

示例 4 用户收视数据标签示例

标识	正题名	创建日期	导演	演员	出品年代	内容描述	总集数	分类名称	连续剧分类	声道语种	地区参数	一级标签	二级标签	三级标签
51875	温州一家人(18)	2017/10/30 17:16	孔笙	李立群, 李	2012	该剧由李	36	电视剧场\大陆剧	温州一家人	国语	大陆	基本特征	剧场	家庭剧
51869	晓说2017(22)	2017/10/30 16:20	无	高晓松	2017	新版《晓	54	综艺娱乐	晓说2017	国语	大陆	基本特征	娱乐	脱口秀
51868	精灵梦叶罗丽第	2017/10/30 16:18	金今	无	2017	女王曼多	26	电视剧场\大陆剧	精灵梦叶罗丽	国语	大陆	基本特征	娱乐	动漫
51866	精灵梦叶罗丽第	2017/10/30 16:18	金今	无	2017	女王曼多	26	电视剧场\大陆剧	精灵梦叶罗丽	国语	大陆	适用人群	家庭成员	儿童
51864	捷德奥特曼(17)	2017/10/30 16:17	坂本浩一	滨田龙臣	2017	主人公朝	25	电视剧场\日韩剧	捷德奥特曼	日语	日本	基本特征	娱乐	动漫
51863	外来媳妇本地郎	2017/10/30 15:17	无	无	2017	外来媳妇	1	电视剧场\大陆剧	外来媳妇本地	粤语	大陆	基本特征	剧场	情景剧
51861	超级小郎中(10)	2017/10/30 15:07	杨小波	朱元冰, 李	2017	民国初期	20	电视剧场\大陆剧	超级小郎中	国语	大陆	基本特征	剧场	言情剧

示例 5 产品数据标签示例

用户号	业务品牌	用户状态	状态改变时间	套餐	销售品	资费	入网时间	机顶盒编号	一级标签	二级标签	三级标签
11183	互动电视	正常	20160718	乐惠套餐	互动+联合	互动+联合宽带-59元包-	20160718	11284	基本特征	观看时间段	上午
11183	互动电视	正常	20160718	乐惠套餐	互动+联合	互动+联合宽带-59元包-	20160718	11284	基本特征	家庭成员	老人
10830	互动电视	正常	20160119		互动标准	互动标准包-包月-26.5元	20160119	10980	基本特征	家庭成员	男性
10073	互动电视	正常	20141122		互动标准	互动600预存-26.5元-捆绑	20130704	10551	基本特征	家庭成员	儿童
10536	互动电视	正常	20140630	乐惠套餐	互动+联合	互动+联合宽带-包月-26.	20140630	10716	基本特征	家庭成员	青年
10536	互动电视	正常	20140630	乐惠套餐	互动+联合	互动+联合宽带-包月-26.	20140630	10716	收视偏好	电视剧	古装剧
10082	互动电视	正常	20130711		互动标准	互动600预存-26.5元-捆绑	20130711	11141	收视偏好	电视剧	情感剧
10783	互动电视	正常	20170125	乐惠套餐	互动+联合	互动+联合宽带-包月-26.	20151230	10524	基本特征	入网时长	老用户
10381	互动电视	正常	20160123	[13年底促	互动+宽	互动+宽带]互动(26.5元)	20131213	10167	基本特征	观看时间段	中午

示例 6 用户数据标签示例

要求给出推荐结果表，格式如示例 7 问题二推荐结果表所示（要求格式为 CSV 文件，命名为队号+问题二推荐结果表）。

用户号	一级标签	二级标签	三级标签	推荐指数
10992	收视偏好	电影	动作		0.73
10992	收视偏好	电视剧	军旅片		0.66
10992	收视偏好	电视剧	古装剧		0.43
10992	收视偏好	娱乐	动画		0.33
10992	基本特征	语言	粤语		0.5
10992	收视偏好	娱乐	综艺		0.2

示例 7 问题二推荐结果表示例