

本文引用格式: 陈彦彬,杨泽华,谢佳.融合多源特征的电梯广告推荐系统[J].自动化与信息工程,2023,44(2):27-35.

CHEN Yanbin, YANG Zehua, XIE Jia. Elevator advertising recommendation system integrating multi-source features[J]. Automation & Information Engineering, 2023,44(2):27-35.

融合多源特征的电梯广告推荐系统*

陈彦彬^{1,2} 杨泽华² 谢佳²

(1.揭阳职业技术学院实训与信息中心, 广东 揭阳 522051

2.广东博华科技有限公司, 广东 揭阳 522000)

摘要: 针对电梯传媒终端广告精准投放面临的采集广告受众个人隐私数据难的问题, 提出融合电梯交通流量、广告主行为时空特征、广告主题特征、广告主评分行为等多源特征的电梯广告推荐算法。首先, 利用差分函数算法从电梯运行数据中提取电梯交通流量峰值特征; 然后, 通过广告主广告行为数据提取广告主行为时空特征; 接着, 将电梯交通流量峰值特征与广告主行为时空特征融合, 并利用 ReliefF 算法进行特征筛选; 最后, 设计融合多源特征的电梯广告推荐系统, 实现电梯广告节目的精准投放。实验结果表明: 融合多源特征的电梯广告推荐算法的 *Precision*、*Recall* 和 ROC 曲线的 AUC 值等评价指标均明显提高; 在一定程度上解决了冷启动、数据稀疏等问题。该系统无需采集广告受众的个人隐私数据, 具有较强的实用性。

关键词: 多源特征; 电梯广告; 推荐算法; 精准投放

中图分类号: TP 301

文献标志码: A

文章编号: 1674-2605(2023)02-0006-09

DOI: 10.3969/j.issn.1674-2605.2023.02.006

Elevator Advertising Recommendation System Integrating Multi-source Features

CHEN Yanbin^{1,2} YANG Zehua² XIE Jia²

(1. Training and Information Center, Jieyang Polytechnic, Jieyang 522051, China

2. General Manager, Guangdong Bohua Technology Co., Ltd., Jieyang 522000, China)

Abstract: In response to the difficulty in collecting personal privacy data of advertising audiences for precise advertising placement in elevator media terminals, a elevator advertising recommendation algorithm is proposed that integrates multi-source features such as elevator traffic flow, spatiotemporal characteristics of advertiser behavior, advertising theme characteristics, and advertiser rating behavior. Firstly, the difference function algorithm is used to extract the peak characteristics of elevator traffic flow from elevator operation data; Then, extract the spatiotemporal characteristics of advertisers' behavior through their advertising behavior data; Next, the peak characteristics of elevator traffic flow are fused with the spatiotemporal characteristics of advertiser behavior, and the ReliefF algorithm is used for feature selection; Finally, design an elevator advertising recommendation system that integrates multi-source features to achieve accurate placement of elevator advertising programs. The experimental results show that the evaluation indicators such as *Precision*, *Recall*, and AUC value of the ROC curve of the elevator advertising recommendation algorithm that integrates multi-source features are significantly improved; To some extent, it has solved problems such as cold start and data sparsity. This system does not need to collect personal privacy data of advertising audiences, and has strong practicality.

Keywords: multi-source features; elevator advertising; recommendation algorithm; accurate placement

0 引言

随着计算广告尤其是电梯传媒终端广告的快速

发展, 精准投放变得越来越重要。推荐算法是计算广告精准投放的关键, 对提高流量分发和广告效益起到

* 基金项目: 揭阳市科技计划项目(2019070); 广东省科技厅驻镇帮镇扶村农村科技特派员重点派驻任务(KTP20210400)。

重要作用。传统的推荐算法主要包括基于内容、用户、物品、标签的推荐等^[1-3]。李剑锋等^[4]在协同过滤的基础上，提出融合个性化和大众化认同度的近相邻改进算法，在一定程度上改善了推荐算法的效果。王英博等^[5]在协同过滤的基础上，通过处理 3 种类型的用户项目子空间，形成 3 棵邻居用户树，计算相似用户，实现协同过滤推荐。文献[6-8]在标准协同过滤框架上整合用户信任关系，改善了推荐效果。文献[4-8]提出的推荐算法均基于传统的推荐算法进行改进，虽然在一定程度上优化了推荐效果，但本质仍是协同过滤思想，需要用户评分等行为数据作为支撑，因此在冷启动、数据稀疏等方面存在瓶颈。

目前，针对计算广告的研究大都集中于对在线广告点击通过率的预测；户外广告的研究则以出租车 LED 屏^[9]、广告牌等为主，侧重于广告设计的美化和提升等策略研究^[10]，较少涉及计算技术。当前大多数电梯传媒终端广告的投放，由于采集广告受众个人隐私数据难、缺少互动场景等原因，使广告受众的行为分析缺乏数据支撑，导致广告投放精准度及效益相对较低。

基于此，本文从影响电梯传媒终端广告投放精准度和个性化的角度进行特征提取，提出融合电梯交通流量、广告主行为时空特征、广告主题特征、广告主评分行为等多源特征的电梯广告推荐算法。首先，介绍融合多源特征的电梯广告推荐系统架构；然后，对多源数据进行特征提取和融合；接着，利用 ReliefF 算法进行特征筛选；最后，设计融合多源特征的电梯广告推荐系统，实现电梯广告节目的精准投放。

1 系统架构

1.1 业务架构

融合多源特征的电梯广告推荐系统本质就是广告节目的推荐系统，主要包括电梯、广告受众、传媒终端（客户端）、广告主、服务器（广告运营商）、数据库等，业务架构如图 1 所示。

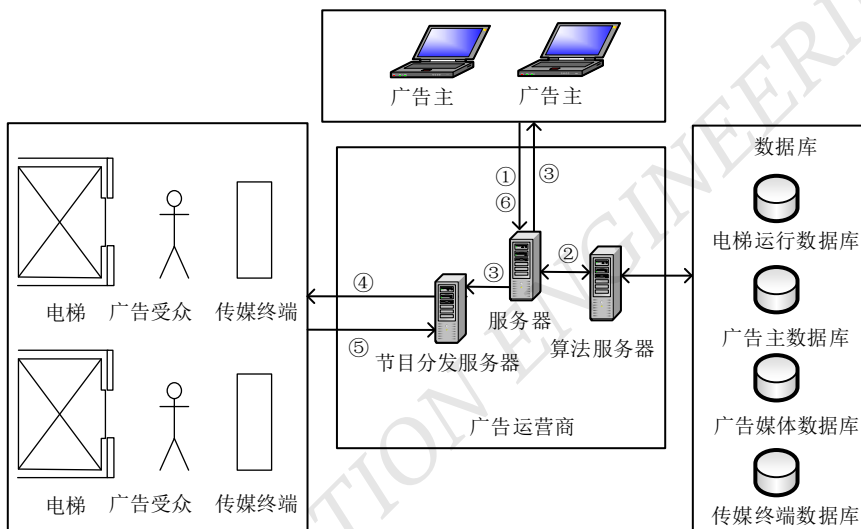


图1 融合多源特征的电梯广告推荐系统业务架构

融合多源特征的电梯广告推荐系统主要业务流程如下：

- 1) 广告主新建广告主题，并提交广告节目推荐列表请求至相应的服务器；
- 2) 服务器调用应用程序接口（application program interface, API）进行广告节目推荐计算；
- 3) 服务器将广告节目推荐列表分发到节目分发服务器；
- 4) 节目分发服务器按照广告节目推荐列表顺序分发广告媒体到电梯传媒终端；
- 5) 电梯传媒终端将广告媒体播出结果及电梯运行数据采集至服务器；
- 6) 广告主对广告节目播放效果进行评分。

1.2 系统框架

融合多源特征的电梯广告推荐系统包括多源数据集、数据预处理及特征提取、特征筛选、模型训练与预测，框架如图 2 所示。

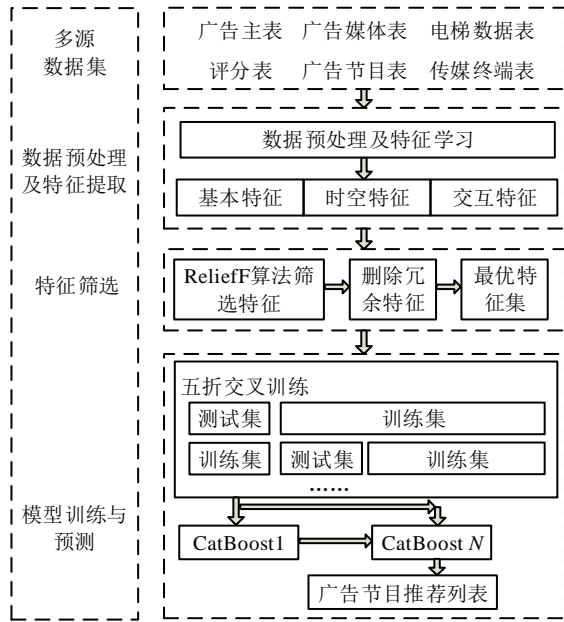


图2 融合多源特征的电梯广告推荐系统框架

2 多源数据处理

2.1 多源数据

本文采用的多源数据为某公司广告推荐系统经脱敏处理后的数据，主要包括广告主、广告主题、电梯、评分记录、传媒终端（电梯）等信息。电梯运行数据表、电梯广告投放日志、广告主信息表、广告主题表、电梯基本信息表、广告主-广告主题评分表分别如表1~表6所示。

表1 电梯运行数据表

| 电梯ID | 轿厢质量/kg | 时间 | 所在楼层 |
|------------|---------|--------------------|------|
| 2020110008 | 300 | 2021/5/8 12:00:18 | 5 |
| 2020110008 | 312 | 2021/5/8 12:00:28 | 10 |
| 2020110008 | 312 | 2021/5/8 12:00:28 | 11 |
| ... | ... | ... | ... |
| 2020110018 | 214 | 2021/5/10 15:00:22 | 10 |

表2 电梯广告投放日志

| 广告主题ID | 播放开始时间 | 播放结束时间 | 电梯ID |
|------------|-------------------|-------------------|------------|
| 2021050008 | 2021/5/8 12:00:18 | 2021/5/8 12:00:27 | 2020110008 |
| 2021050108 | 2021/5/8 12:00:28 | 2021/5/8 12:00:33 | 2020110008 |
| 2021051009 | 2021/5/8 12:00:34 | 2021/5/8 12:00:40 | 2020110008 |
| ... | ... | ... | ... |
| 2021060018 | 2021/6/7 9:00:22 | 2021/6/7 9:00:31 | 2021110018 |

表3 广告主信息表

| 字段 | 含义 | 字段 | 含义 |
|--------------------------|-------|----------------|------|
| enter_id | 广告主编号 | industry_type | 行业类型 |
| enter_scale | 规模 | enter_addr | 地址 |
| enter_established_time | 成立时间 | enter_level | 级别 |
| enter_registered_capital | 注册资本 | enter_property | 性质 |

表4 广告主题表

| 字段 | 含义 | 字段 | 含义 |
|----------------|--------|------------------|------|
| adver_theme_id | 广告主题编号 | adver_com | 广告评论 |
| adver_time | 广告时长 | product_type | 产品类型 |
| adver_mode | 广告模式 | product_price | 产品价格 |
| adver_audien | 广告受众 | product_indust | 产品行业 |
| adver_period | 广告时段 | product_property | 产品属性 |
| adver_city | 广告区域 | product_brand | 产品品牌 |
| adver_abstract | 广告摘要 | | |

表5 电梯基本信息表

| 字段 | 含义 | 字段 | 含义 |
|---------------------|--------|----------------|------|
| elevator_id | 电梯编号 | elevator_type | 电梯类型 |
| elevator_estab_time | 电梯建设时间 | elevator_price | 电梯价格 |
| elevator_layer | 电梯层数 | elevator_brand | 电梯品牌 |
| elevator_lon | 电梯经度 | elevator_speed | 运行速度 |
| elevator_lat | 电梯纬度 | elevator_load | 电梯载重 |

表6 广告主-广告主题评分表

| | 广告主题1 | 广告主题2 | 广告主题3 | 广告主题4 | ... | 广告主题m |
|------|-------|-------|-------|-------|-----|-------|
| 广告主1 | 1 | 0 | 3 | 5 | ... | 5 |
| 广告主2 | 4 | 5 | 0 | 5 | ... | 2 |
| 广告主3 | 3 | 3 | 4 | 0 | ... | 4 |
| ... | ... | ... | ... | ... | ... | ... |
| 广告主n | 0 | 4 | 5 | 2 | 4 | 4 |

2.2 数据预处理

对多源数据中的缺失值、异常值和重复值进行数据清洗后，采用归一化、二值化、离散化、编码等处理方法，分别对表1~表6中的数据进行数据预处理，形成多源数据集。

2.2.1 连续型特征

多源数据中的连续型特征，如表3中的规模（enter_scale）、注册资本（enter_registered_capital）等，采用归一化方法进行处理，具体如公式(1)所示。

$$V_{new} = \frac{v - v_{min}}{v_{max} - v_{min}} \quad (1)$$

式中： V_{\min} 为特征值的最小值， V_{\max} 为特征值的最大值。

表 5 中的电梯载重 (elevator_load) 需折算为电梯乘客数，作为广告受众的流量特征。假定电梯载客为 75 kg/人，电梯载重折算为电梯乘客数的计算公式为

$$N_{t_i} = \left\lfloor \frac{W_{t_i}}{75} \right\rfloor \quad (2)$$

式中： W_{t_i} 为轿厢质量， N_{t_i} 为折算后的电梯乘客数。

利用坐标转换将表 5 中的电梯经度 (elevator_lon)、电梯纬度 (elevator_lat) 转换为百度地图坐标，与电子地图匹配。

每个广告主题根据播放起止时间进行节目的归属划分处理。

根据广告主行为时空特征的分析结果，利用同类型广告主评分的均值补齐缺省值，解决评分数据稀疏的问题。

2.2.2 离散型特征

多源数据中的离散型特征，如表 3 中的行业类型 (industry_type)、级别 (enter_level)、性质 (enter_property) 等，需从非数值型数据转换为数值型数据，以方便模型训练。本文采用虚拟编码的方法，将同个特征中的 n 个取值转换为 $n-1$ 个特征值，以解决虚拟编码的共线问题，提高模型训练精度。如广告主性质 (enter_property) 的属性有国家机关、事业单位、国有企业、集体企业、有限公司、股份公司、三资企业、私营企业、自然人、个体户及其他企业共 11 个取值，采用虚拟编码的方法将其转换为 10 个特征值，如国家机关用向量表示为 $f = (1,0,0,0,0,0,0,0,0,0)$ ，事业单位用向量表示为 $f = (0,1,0,0,0,0,0,0,0,0)$ ，其他企业用向量表示为 $f = (0,0,0,0,0,0,0,0,0,0)$ 。表 4、表 5 中的其他离散型特征，采用同样的虚拟编码方法进行预处理。

2.2.3 文本型特征

多源数据中的文本型特征，如表 4 中的广告摘要 (adver_abstract)、广告评论 (adver_com) 等，采用手工、自动提取相结合的方法选择关键词 key ，并利用词频-逆向文件频率 (term frequency-inverse

document frequency, TF-IDF) 计算其权重值 $weight$ ，构成新的特征向量 ($key, weight$)。

3 特征提取

3.1 电梯交通流量峰值特征

对多源数据集进行分析，发现居民小区、办公楼的电梯交通流量具有明显的周周期性，因此对电梯交通流量以周周期进行分析。假定每部电梯只有 1 个传媒终端，传媒终端 e 的广告以时间段 (5 min) 进行划分，记为广告节目 AD_j^e ($0 \leq j < 2016$)，形成新特征。

假定 N_{t_i} 为 t_i 时刻的电梯乘客数，则电梯乘客数据集 $N = \{N_{t_1}, N_{t_2}, \dots, N_{t_n}\}$ 。通过差分函数算法计算电梯乘客数的极大值，从而提取电梯交通流量峰值特征，计算公式为

$$D(i) = N_{t_{i+1}} - N_{t_i} \quad (0 \leq i < n) \quad (3)$$

$$\text{sign}(x) = \begin{cases} 1 & x > 0 \\ 0 & x = 0 \\ -1 & x < 0 \end{cases} \quad x \in D(i) \quad (4)$$

$$Q_{j_new} = \frac{\text{index}(Q_j)}{\text{count}(Q)} \quad (5)$$

式中： $D(i)$ 为 N_{t_i} 的一阶差分， $\text{index}(Q_j)$ 为电梯交通流量特征 Q_j 的索引值， $\text{count}(Q)$ 为集合 Q 中的元素数量。

电梯交通流量峰值特征的提取步骤如下：

步骤 1 求电梯乘客数据集 N 的一阶差分 $D(N)$ ；

步骤 2 利用公式(4)计算 $\text{sign}(D(N))$ ，记为 E ，

若 $E_{t_{i-1}} = 0$ ，且 $E_{t_i} \geq 0$ ，则 $E_{t_{i-1}} = 1$ ；若 $E_{t_{i-1}} = 0$ ，且 $E_{t_i} < 0$ ，则 $E_{t_{i-1}} = -1$ ；

步骤 3 利用公式(3)计算 E 的一阶差分 $D(E)$ ，记为 Q ，若 $Q_{t_{i-1}} = -2$ ，则 N_{t_i} 为 t_i 时刻的极大值；

步骤 4 根据极大值 N_{t_i} 构建电梯交通流量峰值特征集合 Q ，并按照 t_i 时刻与广告节目 AD_j^e 对应；

步骤 5 求电梯交通流量峰值特征集合 Q 的平均值，作为广告节目的电梯交通流量特征 Q_j ；

步骤 6 对广告节目的电梯交通流量特征 Q_j 进行

排序, 记为 $Q_j = \{Q_1, Q_2, \dots, Q_j\}$, 其中 $Q_k \leq Q_l$, $1 \leq k \leq l \leq J$;

步骤7 利用 $index(Q_j)$ 求出 Q_j 的索引值, 以及集合 Q 中的元素数量 $count(Q)$, 并利用公式(5)进行特征映射转换, 形成电梯交通流量峰值特征 Q_{j_new} 。

3.2 广告主行为时空特征

广告主发布广告主题会形成一系列行为, 包括广告时长、广告模式、广告区域、广告评论等。本文分别从时间和空间进行特征分析, 为广告节目推荐提供基础依据, 同时为新用户提供聚类均值。

3.2.1 时序特征

广告主在时间范围 t 内的行为序列为 $AD = \{AD_1, AD_2, AD_1, AD_3, AD_2, \dots, AD_m\}$, 其中 AD_j 为广告主在连续时间范围 t 内实际投放的广告节目。行为序列 AD 描述了广告主在过去一段时间内广告行为的时间变化。基于此, 本文把行为序列类比为句子, 采用 Word2Vec 中的 Skip-gram 词向量模型构建广告行为时序特征学习算法。

假定由广告主题数据集 AT 、时间窗口 t 、所有广告主的行为序列组成语料库 A , 训练得到 Skip-gram 词向量模型 $S(A)$, 则集合相似度计算公式为

$$\text{Sim}(AD_i, AD_j) = \frac{|L_i \cap L_j|}{|L_i \cup L_j|} \quad (6)$$

式中: L_i 为节目推荐列表, $L_i \cap L_j$ 、 $L_i \cup L_j$ 为节目推荐列表 L_i 、 L_j 的交集和并集。

广告主行为时序特征的提取步骤如下:

步骤1 在广告主题数据集 AT 中, 提取广告主在时间 t 内的行为序列 $AD = \{AD_1, AD_2, AD_1, AD_3, AD_2, \dots, AD_m\}$;

步骤2 计算行为序列 AD 中次数最多的广告节目 AD_m 、广告总时长最长的广告节目 AD_p 、最近一次广告节目 AD_r ;

步骤3 利用 Skip-gram 词向量模型 $S(A)$ 分别计算 AD_m 、 AD_p 、 AD_r 的节目推荐列表 L_m 、 L_p 、 L_r ;

步骤4 利用公式(6)计算 L_m 、 L_p 、 L_r 两两之间的

相似度, 并输出广告主的行为时序特征, 记为 $x = [\text{Sim}(AD_m, AD_p), \text{Sim}(AD_m, AD_r), \text{Sim}(AD_r, AD_p)]$ 。

3.2.2 空间特征

电梯传媒终端广告精准投放时, 广告受众分布的区域特点, 使广告主的行为特征也具有明显的区域特征。本文采用密度峰值聚类算法对广告主行为空间特征进行提取。

广告主在时间范围 T 内投放广告的电梯传媒终端位置集合为 $Z = \{Z_1, Z_2, \dots, Z_n\}$, Z_q 为电梯传媒终端的百度地图坐标。对位置集合 Z 中的任一节点 Z_q , 计算其局部密度 ρ_q 及其到其他具有更高局部密度节点的距离 δ_q , 计算公式为

$$\rho_q = \sum_i \chi(d_{qi} - d_c) \quad (7)$$

$$\delta_q = \begin{cases} \max_i(d_{qi}), & \text{if } q = \arg \max_i(\rho_i) \\ \min_{i: \rho_i < \rho_q}(d_{qi}), & \text{otherwise} \end{cases} \quad (8)$$

式中: $\chi(x) = \begin{cases} 1 & x < 0 \\ 0 & x \geq 0 \end{cases}$, d_{qi} 为节点 Z_q 、 Z_i 之间的欧式距离, d_c 为截断距离。

利用公式(7)计算出所有节点的 ρ_q 和 δ_q 后, 根据公式(9)计算聚类中心点的权值, 具体公式为

$$\gamma_q = \rho_q \times \delta_q \quad (9)$$

$$C_c = \frac{\Delta m_c}{m_c} \quad (10)$$

式中: γ_q 为聚类中心点的权值, 其值越大, 节点 Z_q 越有可能成为聚类中心; c 为聚类的类别数量; $\Delta m_c = |m_c - m_{c+1}|$ 为聚类中心数量变化时样本数量变化的增量; m_c 为算法聚类后归属到各个类别的样本总数量; C_c 反映聚类中心变化时样本数量总体变化的幅度。

广告主行为空间特征提取步骤如下:

步骤1 计算电梯传媒终端位置集合 $Z = \{Z_1, Z_2, \dots, Z_n\}$ 任意 2 个节点之间的欧式距离 d_{qi} 及截断距离 d_c ;

步骤 2 根据公式(7)、(8)、(9)计算各个节点的 ρ_q 、 δ_q 、 γ_q ;

步骤 3 对 γ_q 进行降序排列, 并标注下标号, 令 $c = 1$;

步骤 4 对位置集合 Z 进行聚类, 并通过公式(10)计算 C_c ;

步骤 5 利用 $c = c + 1$ 更新聚类中心数量, 并对位置集合 Z 进行聚类, 继续计算 C_c ;

步骤 6 重复步骤 4~5, 直到 C_c 达到最大值, 返回聚类中心数量 $c = c1$;

步骤 7 分配非聚类中心点到相应的聚类类别;

步骤 8 返回各个节点的类别特征, 即为广告主行为空间特征, 记为 $Z = \{(Z_1, C_1), (Z_2, C_2), \dots, (Z_n, C_c)\}$ 。

4 特征筛选

为进一步提高模型训练效率和准确率, 利用 ReliefF 算法对多源数据集的特征进行筛选, 选取贡献度大的特征作为最优特征集。

ReliefF 算法进行特征筛选的流程如图 3 所示。

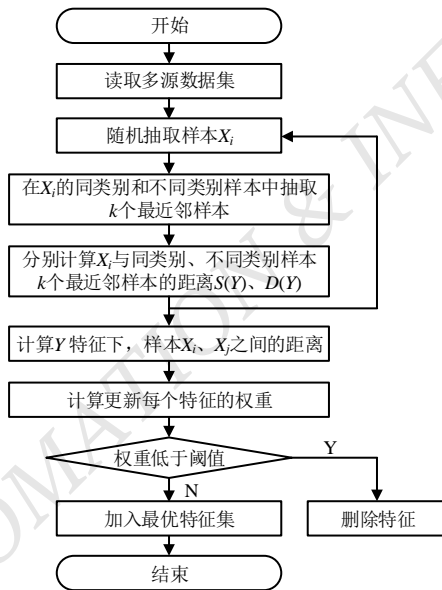


图 3 ReliefF 算法进行特征筛选的流程

特征权重 $W(Y)$ 的更新公式为

$$W(Y) = W(Y) - \frac{S(Y)}{nk} + \sum_{c \in \text{class}(X_i)} \left[\frac{P(c)}{1 - P(\text{class}(X_i))} \times D(Y) \right] \quad (11)$$

式中: $\text{class}(X_i)$ 为随机样本 X_i , $S(Y)$ 、 $D(Y)$ 分别为对于某一特定的特征 Y , 随机选取的样本 X_i 与同类别的 k 个最近邻样本、不同类别的 k 个最近邻样本之间的距离, n 为迭代次数, $P(c)$ 为类比比例。

本文设定特征权重的阈值为 0.4。基于此, 本文多源数据集经特征提取后共有 154 个特征, 经 ReliefF 算法筛选出 122 个特征作为本文系统的特征集。

5 系统设计

5.1 功能模块设计

融合多源特征的电梯广告推荐系统主要包括广告模块、受众模块、电梯管理模块、管理员模块等 4 个模块, 如图 4 所示。

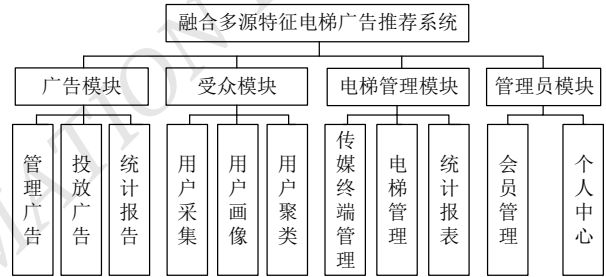


图 4 融合多源特征的电梯广告推荐系统功能模块

融合多源特征的电梯广告推荐系统的用户主要包括广告商、代理商、操作员、管理员等。该系统可进行广告节目、广告主题、广告交易明细等信息的实时统计, 设定定向投放准则, 设置频次控制等。

5.2 系统实现

融合多源特征的电梯广告推荐系统采用 B/S 架构设计开发, 系统服务器端和客户端的界面如图 5 所示。



(a) 服务器端

(b) 客户端

图 5 系统界面

系统使用时，需要将电梯传媒终端的设备号录入服务器端，客户端可在后台注册设备信息。从广告节目服务器获取广告媒体并进行播放。系统以 CatBoost 模型为基础，形成基于广告节目的时间推荐列表、基于传媒终端的空间推荐列表、基于终端-节目的时空推荐列表，经过加权融合后，依据 CatBoost 模型计算结果进行 TOP-N 推荐，即为广告节目推荐列表。

6 实验结果分析

6.1 实验数据集

为验证本文提出的融合多源特征的电梯广告推荐系统的效果，采集某公司广告推荐系统的 2021 年 3 月~8 月的平台数据，经脱敏和预处理后作为实验数据集，如表 7 所示。

表 7 实验数据集表

| 数据 | 规模 |
|---------|-------------|
| 电梯运行数据 | 1 962 000 条 |
| 广告主 | 16 500 家 |
| 广告主题 | 41 800 个 |
| 广告主行为数据 | 356 000 条 |
| 传媒终端 | 2 450 个 |

实验数据集中的每条数据包括电梯及广告主基本信息、电梯交通流量峰值特征、广告主行为时空特征和广告主与广告主题的交互信息等共 122 个特征信息。

将实验数据集划分为训练集和测试集，其中训练集包含数据 224 556 条，测试集包含数据 56 139 条。

6.2 评价指标

本文采用查准率 *Precision*、查全率 *Recall* 和 *F1* 得分等指标对本文算法性能进行评价。其中，*Precision* 表示分类正确的正样本个数占被判定为正样本的样本个数的比例；*Recall* 表示分类正确的正样本个数占真正的正样本个数的比例；*F1* 是对 *Precision*、*Recall* 两个指标的调和平均值。假定 I 表示算法推荐列表 TOP-N 个广告节目集合； A 表示广告主感兴趣的广告节目集合； $|I|$ 表示推荐列表广告节目的数量； $|A|$ 表示广告主感兴趣的广告节目的数量，评价指标计算公

式为

$$Precision = \frac{|I \cap A|}{|I|} \quad (12)$$

$$Recall = \frac{|I \cap A|}{|A|} \quad (13)$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (14)$$

除了上述基本评价指标外，本文还对推荐列表排序采用 P-R 曲线和 ROC 曲线的 AUC 值等评价指标进行推荐效果分析。

6.3 实验过程及结果分析

为验证融合多源特征的电梯广告推荐系统（记为方法 1）的推荐效果，进行 4 组对比实验。假定只考虑时间特征的算法记为方法 2；只考虑空间特征的算法记为方法 3；考虑时空特征但是未使用 ReliefF 进行特征筛选的算法记为方法 4。通过以上 4 种方法对同一数据集进行五折交叉验证训练，并进行 TOP-20 推荐，分别计算 *Precision*、*Recall*、以及 ROC 曲线的 AUC 值等评价指标数值，实验结果如图 6 所示。

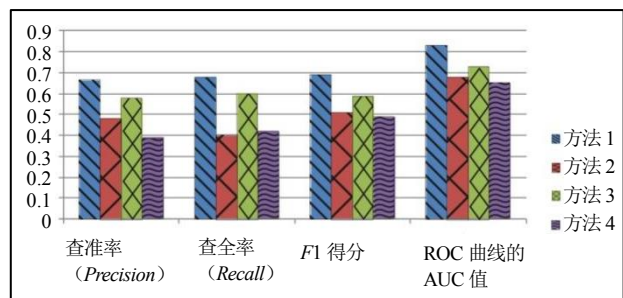


图 6 4 种方法实验结果

由图 6 可知：方法 1 的 *Precision*、*Recall*、*F1* 以及 ROC 曲线的 AUC 值等评价指标明显优于其他方法；方法 3 的各项评价指标优于方法 2，主要原因是方法 2 的电梯广告推荐系统缺少对电梯流量的统计分析，导致广告主在投放电梯广告时仅考虑分布区域；方法 4 的各项评价指标均最低，主要是由于未经过特征筛选，原始特征维度较多，通过训练构建的模型复杂度较高，引起过拟合造成了预测精度不准。

同时，对上述 4 种方法分别进行 P-R 曲线分析，

如图 7 所示。

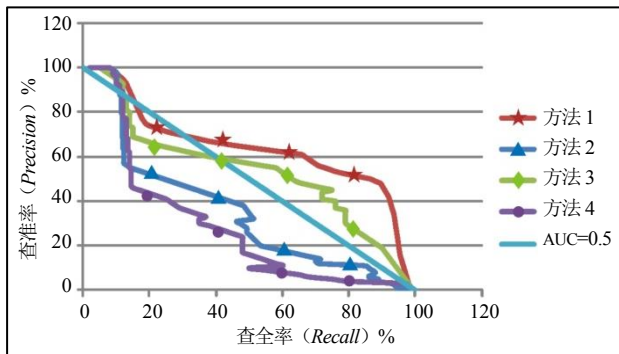


图 7 4 种方法的 P-R 曲线对比

由图 7 可知：方法 1 的 P-R 曲线较为平滑，优于其他方法，且 *Precision*、*Recall* 在 65% 左右时，其推荐性能较为稳定；方法 2 和方法 4 的 P-R 曲线多处出现曲折、不平滑，在一定程度上存在过拟合现象；方法 3 的 *Precision*、*Recall* 在 55% 左右时，推荐性能相对稳定。

综上所述，融合多源特征的电梯广告推荐系统与其他 3 种方法相比，其 *Precision*、*Recall*、*F1* 和 ROC 曲线的 AUC 值等多项评价指标均为最优，达到较好的推荐性能。

本文选用 CatBoost、XGBoost、lighthBGM 算法模型进行测试，分别计算其 AUC 值，如表 8 所示。

表 8 不同推荐算法模型的 AUC 值

| 推荐算法模型 | AUC 值 |
|-----------|---------|
| CatBoost | 0.812 6 |
| XGBoost | 0.798 2 |
| lighthBGM | 0.789 5 |

由表 9 可知，CatBoost 推荐算法模型的 AUC 值为 0.8126，高于 XGBoost、lighthBGM 两种推荐算法，因此本文推荐算法选择 CatBoost 算法模型。

对于新用户冷启动的测试，实验中随机抽取 10 位用户对模型进行训练；随后将其数值型特征用均值代替，删除行为特征，并利用训练好的推荐算法模型进行预测且与真实值进行对比分析，其 ROC 曲线的 AUC 值为 0.725，说明本文方法对解决冷启动具有较好的效果。

7 结论

本文设计一种融合多源特征的电梯广告推荐系统，通过电梯运行数据分析提取电梯交通流量峰值特征；通过广告主广告行为数据提取时空行为特征；运用 ReliefF 算法进行特征筛选；将特征向量输入融合多源特征的电梯广告的推荐系统进行学习训练，形成基于时间特征、空间特征和时空特征的广告节目推荐列表。实验结果表明，本文提出的电梯广告推荐算法的 *Precision*、*Recall*、*F1* 和 ROC 曲线的 AUC 值等评价指标均优于其他推荐算法，并且其 P-R 曲线也较为平滑、稳定，具有较好的推荐性能，同时对解决冷启动、数据稀疏等问题也具有较好的效果。

参考文献

- [1] ADOMAVICIUS G, TUZHILIN A. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions[J]. IEEE Transactions on Knowledge and Data Engineering, 2005,17(6):734-749.
- [2] LI X F, LI D, YIM J. An improved collaborative filtering recommendation algorithm and recommendation strategy[J]. Mobile Information Systems, 2019:1-11.
- [3] KOREN Y, BELL R, VOLINSKY C. Matrix factorization techniques for recommender systems[J]. Computer, 2009,42(8): 30-37.
- [4] 李剑锋,封林慧,于天一.认同度修正下的近相邻改进推荐算法研究[J].计算机工程与应用, 2022,58(7):116-121.
- [5] 王英博,韩国淼,王铭泽.基于子空间聚类的协同过滤推荐算法[J].计算机工程与应用, 2022,58(3):127-134.
- [6] MORADI P, AHMADIAN S. A reliability-based recommendation method to improve trust-aware recommender systems[J]. Expert Systems with Applications: An International Journal, 2015,42(21):7386-7398.
- [7] LI Y M, WU C T, LAI C Y. A social recommender mechanism for e-commerce: combining similarity, trust, and relationship[J]. Decision Support Systems, 2013,55(3):740-752.
- [8] O'DONOVAN J, SMYTH B. Trust in recommender systems [C]//Proceedings of the 10th International Conference on Intelligent User Interfaces, 2005:167-174.
- [9] 李科,党延忠.出租车新运营模式下的 LED 广告精准投放策略[J].中国管理科学,2020,28(10):220-230.
- [10] 余光华.商场电动扶梯广告设计的互动性研究[D].成都:西

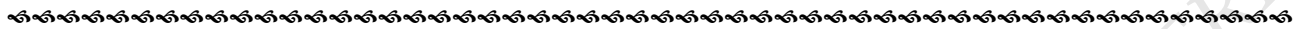
上海交通大学,2012.

作者简介:

陈彦彬,男,1987年生,本科学士,高级工程师,主要研究方向:电子信息系统开发、终端计算广告等人工智能教学与科研。E-mail: chenyanbin01@126.com

杨泽华,男,1967年生,本科学历,总经理、技术员,主要研究方向:计算机技术应用。

谢佳,男,1990年生,本科学历,技术员,主要研究方向:计算机技术应用。E-mail: chenyanbin01@126.com



《自动化与信息工程》征稿启事

《自动化与信息工程》(原《广东自动化与信息工程》)期刊由广东省科学院智能制造研究所、广州市自动化学会联合主办。国内统一连续出版物号 CN 44-1632/TP; 国际标准连续出版物号 ISSN 1674-2605。国内外公开发行,面向各行业及广大自动化与信息用户,欢迎踊跃来稿。

一、刊物性质与栏目:

1. 本刊以报导自动化与信息学科的研究成果、应用成果及国内外信息动态为主。包括:自动控制理论及应用、自动控制系统装置及应用、工业控制及检测技术、自动化仪表及传感技术、机器人控制技术、系统控制科学中的其他重要问题、电子及计算机技术、网络与信息技术、软件开发与应用、自动测控领域的应用研究及其发展动态、通讯技术等。读者对象主要为从事自动化、信息技术的研究人员、工程技术人员及高校师生。

2. 本刊常设栏目有:特约论文、综述、学术研究、开发设计、技术应用、专题报道、新产品新技术介绍、技术讲座、经验交流、问题征解与解答、研究生园地、学术动态等。

二、稿件要求:

1. 投稿论文未公开发表过。

2. 投稿论文论点明确,论述充分,语言通顺,文字简练,图表清晰。

3. 投稿论文格式与要求:文章题目、作者姓名、作者单位、摘要(不超过 200 字)、关键词(3~8 个)、正文、参考文献;英文的题目、单位、摘要、关键词;作者简介。

1). 文中图和表格:须标有图号、图题和表序、表题,所用字符为六号或小五号宋体,图的宽度限制在 8 cm 或 16 cm 以内,随文排入。

2). 参考文献(公开出版物)按文中出现的先后次序排列,书写格式参考 GB7714-2015《信息与文献 参考文献著录规则》。

3). 稿中非标准缩写词(中文或英文)须在首次出现时定义清楚。外文字母必须分清大小写,正、斜体,上、下角;容易混淆的外文字母及符号,要特别注意标清。

4). 计量单位一律采用法定计量单位,名词术语必须规范化、标准化,且前后一致。

5). 文末附作者简介,内容包括:姓名、性别、出生年月、学位、职称。主要研究领域及 E-mail 地址。

6). 投稿论文一般不少于 5000 字(包括图表),综述一般不少于 8000 字(包括图表)。要求:Word 格式的电子文档。

7). 来稿请注明:邮编、详细通信地址及联系电话。

4. 投稿方式:请发电子邮件到邮箱: aie@giim.ac.cn。

三、说明:

1. 本刊原则上只收原始性稿件。已在国内外学术会议上发表或准备发表的文章必须如实说明,酌情刊登。

2. 作者必须对稿件内容的真实性和可靠性负责。

3. 编辑部收到投稿后及时通知作者。来稿若不符合本刊稿约,先退给作者修改后再审稿。

4. 编辑部有权对来稿进行修改或删节。稿件一经采用,赠送当期期刊 3 本。

《自动化与信息工程》编辑部