

考虑特征差异的地铁乘客满意度研究

王晓佳¹, 吴可²

¹武汉市交通运输局客运事业发展中心, 湖北 武汉

²交通运输部科学研究院, 北京

收稿日期: 2023年1月20日; 录用日期: 2023年3月8日; 发布日期: 2023年3月15日

摘要

基于武汉地铁乘客满意度调查中获取到的基本特征, 采用K-means++算法识别出外地乘客、男性通勤、女性通勤以及生活休闲四类群体, 并对各群体主要特征进行分析。依据满意度调查中各项指标收集到的具体问题, 利用LDA主题模型识别出其主题分布与主题词, 明确各类群体的诉求, 并提出了相应的建议。研究结果可为武汉地铁提升服务质量提供参考和依据。

关键词

地铁乘客满意度, K-Means++, LDA, 服务质量

A Study of Metro Passenger Satisfaction Considering Differences in Characteristics

Xiaojia Wang¹, Ke Wu²

¹Passenger Transport Development Center of Wuhan Municipal Transport Bureau, Wuhan Hubei

²China Academy of Transportation Sciences, Beijing

Received: Jan. 20th, 2023; accepted: Mar. 8th, 2023; published: Mar. 15th, 2023

Abstract

Based on the basic characteristics obtained from the Wuhan Metro passenger satisfaction survey, the K-means++ algorithm is used to identify four groups of foreign passengers, male commuters, female commuters and life and leisure, and the main characteristics of each group are analyzed. According to the specific problems collected from each index in the satisfaction survey, the LDA topic model is used to identify their topic distribution and subject terms, clarify the demands of each group, and put forward corresponding suggestions. The research results can provide refer-

文章引用: 王晓佳, 吴可. 考虑特征差异的地铁乘客满意度研究[J]. 交通技术, 2023, 12(2): 113-119.

DOI: 10.12677/ojtt.2023.122014

ence and basis for Wuhan Metro to improve its service quality.

Keywords

Subway Passenger Satisfaction, K-Means++, LDA, Service Quality

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着人民生活水平的提升,越来越多的地铁乘客对于地铁的需求由“快建设”转变为“高质量”。2019年9月,中共中央、国务院和交通运输部先后印发了《交通强国建设纲要》和《数字交通发展规划纲要》,为城市轨道交通的高质量发展指明了方向。城市轨道交通高质量发展包含很多方面,其中服务质量的提升是乘客较为关注的一点,而乘客满意度调查作为服务质量评价的主要内容之一,可以直接收集乘客意见,反映乘客诉求。在此背景下,各城市轨道交通运营企业如何精确识别乘客满意度调查中的主要诉求,针对性地改进不足,提升运营线路的服务质量,是实现高质量发展的必要手段。

李世伟[1]从列车运行、窗口服务及设备设施3个方面选取了58个指标构建了满意度指标体系,利用层次分析法(Analytic Hierarchy Process, AHP)确定指标权重,并利用模糊算法计算各项指标得分。刘晓娟[2]从进出站、环境与秩序、设施运行、换乘、咨询与投诉五个方面选取了15个指标构成评价指标体系,采用AHP确定各指标权重,利用模糊综合评价法计算各项指标得分。梁铭真[3]从感知质量、乘客期望、感知价值、乘客满意、乘客抱怨和乘客忠诚6个方面选取了47个指标构成评价指标体系,采用AHP-因子分析法对各层次指标赋权,最终利用模糊综合理论以及四分图法对各项指标进行评价。李林波[4]从拥挤、舒适、时效、便利、安全、经济6个方面选取了24个指标构成评价指标体系,采用客观权重赋权法CRITIC (Criteria Importance Though Intercriteria Correlation)确定指标权重,并采用云模型对乘客满意度进行评价。

上述学者进行乘客满意度研究的一般方法是选取指标构建乘客满意度指标体系,确定指标权重,最终计算各项指标得分。该方法一方面建立的乘客满意度指标体系不统一,难以对行业进行规范化指引,另一方面还忽视了乘客自身差异导致的不同诉求。本文采用《城市轨道交通服务质量评价规范》(后文简称《规范》)[5]中规定的乘客满意度指标构建指标体系,在乘客群体分类的基础上对乘客满意度进行分析,重点挖掘乘客实际诉求,有利于城市轨道交通运营企业针对性提升运营服务质量。

2. 问卷设计与数据预处理

2.1. 问卷设计

本文需根据乘客基本特征对其进行分类,并借助主题模型根据每类乘客“文本”形式的建议识别具体需求,因此需要收集乘客基本特征数据和乘客对于各项满意度指标的“文本”形式的建议。乘客基本特征方面,需要包含性别、年龄、乘坐次数、出行目的和居住地。乘客对于各项指标的满意度方面,根据《规范》中规定,包含7个一级指标和12个二级指标,一级指标分别为进出站、环境与秩序、设施运行、换乘、咨询、投诉、安全感以及各项指标存在的具体问题。如图1所示。

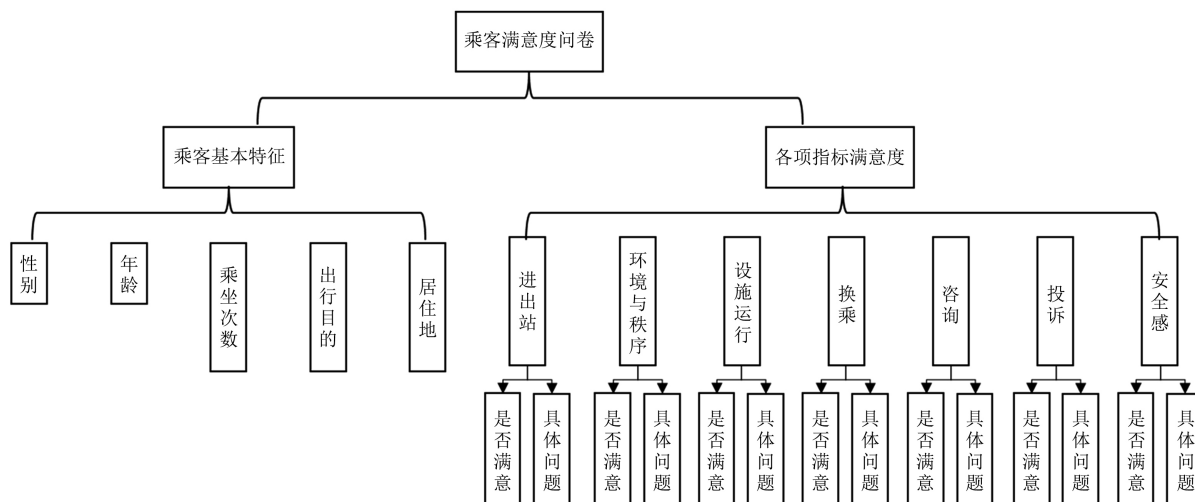


Figure 1. Passenger satisfaction questionnaire question composition

图 1. 乘客满意度问卷问题构成

2.2. 数据获取与预处理

2020 年 12 月至 2021 年 1 月, 以线上和线下形式在武汉市轨道交通开展了乘客满意度调查工作, 并将问卷数据录入到数据库, 最终形成了包含 17,367 条记录的武汉地铁乘客满意度数据集。本文第 3 节利用乘客基本特征数据识别乘客群体, 然而获取到的基本特征数据均为文本形式, 无法输入到无监督聚类算法 K-means++ 中, 因此需要对其进行离散化数值编码。此外, 由于乘客漏填、换乘站数据重复记录等原因, 会存在数据重复和数据缺失的情况。针对上述情况, 采用以下步骤进行数据预处理。

Step 1 删除缺失值: 由于乘客漏填等原因, 可能会导致某些基本特征缺失。调用 pandas 中的 dropna 函数, 将性别、年龄、乘坐次数、出行目的和居住地中存在缺失值对应的数据记录删除。

Step 2 删除重复数据: 乘客在换乘站所填数据将被重复记录, 调用 pandas 中的 drop_duplicates 函数删除重复的数据记录。

Step 3 对基本特征进行编码: 乘客基本特征数据采用问卷方式获得, 均为文本形式, 为了更好地输入聚类算法中, 将文本进行离散化数值编码, 以出行目的为例, 该基本特征编码规则见图 2。

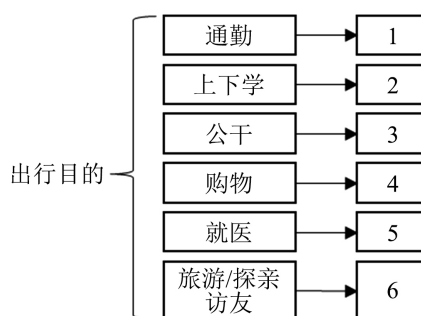


Figure 2. Discrete numerical codes for travel purposes

图 2. 出行目的离散化数值编码

3. 基于 K-Means++ 的乘客群体识别

本节选取性别、年龄、乘坐次数、出行目的和居住地作为基本特征指标, 基于 K-means++ 算法实现

乘客群体识别。

3.1. 乘客群体识别

K-means++是对 K-means 算法中选取初始聚类中心方法的一种改进, 该算法首先随机选择一个点作为第一个聚类中心, 并将所有点按照距离该聚类中心的距离的平方的概率加入聚类中心的集合中。然后, 循环直到聚类中心的数量达到所需的数量 k , 在每次循环中随机选择一个点并将其加入聚类中心的集合中。最后, 返回聚类中心的集合作为最终的聚类中心。确定聚类中心后, 依据各样本与聚类中心的距离判断样本类别。该算法确定初始聚类中心的伪代码如下所示:

```

input: 数据集 D (大小为 n), 聚类个数 k
output: k 个聚类中心

1. 初始化空列表 S 来存储选定的聚类中心
2. 从 D 中随机选择一个点 p 并将其添加到 S 中
3. 对于 D 中的每个点 x:
   a. 计算 x 和 S 中最近聚类中心之间的距离 d(x,S)
   b. 以 d(x,S)2 的概率将 x 添加到 S 中
4. 当|S| < k 时:
   a. 从 D 中随机选择一个点 p
   b. 将 p 添加到 S 中
5. 返回 S 作为聚类中心集合
end
    
```

由于 K-means++属于无监督聚类算法, 聚类前需指定聚类数量。常用的选取最佳聚类数量的方法是肘部法则。肘部法则的原理为: 当聚类数量 K 小于最佳聚类数时, K 的增大会大幅增加每个簇的聚合程度, 误差平方和 SSE 的下降幅度会很大, 而当 K 到达最佳聚类数时, 再增加 K 所得到的聚合程度回报会迅速变小, SSE 的下降幅度会骤减, 当二者关系处于曲率突变点时对应的 K 值为最佳聚类数量。图 3 展示了本算法中 K 与 SSE 的关系, 根据图 3 可以确定最佳聚类数量为 4。

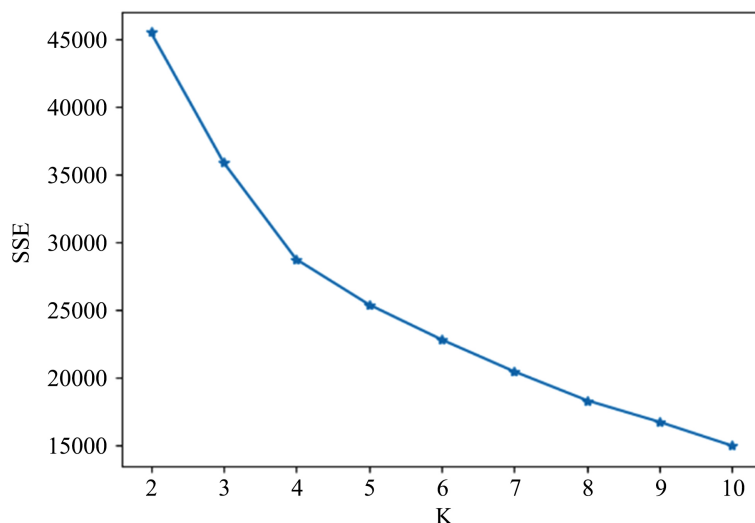


Figure 3. Elbow rule to determine the optimal number of clusters
 图 3. 肘部法则确定最佳聚类数量

3.2. 乘客群体特征分析

确定最佳聚类数量为 4 后, 将由性别编码、年龄编码、乘坐次数编码、出行目的编码和居住地编码形成的数据集输入到 K-means++ 模型中, 最终识别出 4 类乘客群体。根据群体主要特征, 分别将其命名为外地乘客群体、男性通勤群体、女性通勤群体、生活休闲群体。

居住地方面, 外地乘客群体全部由短暂停留的外地乘客构成; 性别方面, 该类群体中男女乘客比例接近, 占比分别为 53.3% 和 46.7%; 年龄方面, 大多数乘客处于 18~30 岁, 占比约为 75%; 乘坐次数方面, 该类群体属于出行频次较低的群体, 有超过 60% 乘客一周乘坐地铁小于 5 次; 出行目的方面, 有大多数乘客出行目的为通勤和旅游/探亲访友, 比例分别占到了 35% 和 25%。

男性和女性通勤群体除了在性别上分别由男性和女性构成以外, 在年龄、乘坐次数、居住地、出行目的等方面基本一致。两类群体均有超过 60% 的乘客处于 18~44 岁之间; 有超过 80% 的乘客一周出行 5 次以上, 属于出行频次较高的群体; 居住地方面, 两类群体均由本地乘客构成; 出行目的方面, 两类群体中绝大多数都是以通勤为目的的出行, 其中女性通勤群体中有接近 87% 的乘客出行目的以通勤为主, 其他类别出行目的仅占 13%。

生活休闲群体中女性乘客比例略高于男性乘客, 男女乘客分别占比为 40% 和 60%; 年龄方面, 与其他三类群体相比, 该类群体处于 18~30 岁的乘客比例下降至 48.6%, 45 岁以上的乘客比例逐步提升; 居住地方面, 该类乘客绝大多数都是本地乘客, 仅有不到 1% 的乘客为短暂停留的外地乘客; 乘坐次数方面, 该类群体出行频率较低, 有接近 80% 的乘客一周出行次数小于 5 次; 出行目的方面, 有接近一半的乘客出行目的是购物, 超过 30% 的乘客出行目的为旅游/探亲访友。

4. 基于 LDA 的乘客群体满意度分析

乘客针对各项满意度指标的提出了“文本”形式具体问题, 因此本节采用 LDA 模型识别不同类别群体对于各项满意度指标具体需求。

4.1. LDA 原理

隐含狄利克雷分布(Latent Dirichlet Allocation, LDA)是一种常见的主题模型, 该模型假设一篇文档的生成过程服从隐含狄利克雷分布, 通过对分布的参数进行标定推断文档的主题分布, 实现语义信息的提取和识别。

LDA 模型描述文档生成的过程中需要用到隐含狄利克雷分布和多项式分布。隐含狄利克雷分布本质上是 Beta 分布推广到高维的一种特殊情况, 其概率密度函数计算过程见式(1)和式(2)。

$$f(x_1, x_2, \dots, x_k; \alpha_1, \alpha_2, \dots, \alpha_k) = \frac{1}{B(\alpha)} \prod_k x_i^{\alpha_i - 1} \quad (1)$$

$$B(\alpha) = \frac{\prod_{i=1}^k \Gamma(\alpha_i)}{\Gamma\left(\sum_{i=1}^k \alpha_i\right)}, \sum_k x_i = 1 \quad (2)$$

多项式分布是二项分布的高维扩展, 其概率密度函数计算过程见式(3)。

$$P(x_1, x_2, \dots, x_k; n, p_1, p_2, \dots, p_k) = \frac{n!}{x_1! \dots x_k!} p_1^{x_1} \dots p_k^{x_k} \quad (3)$$

LDA 主题模型生成文档的流程如下:

- Step 1: 按照贝叶斯理论中的先验概率 $P(d_m)$ 选中一篇文档 d_m ;
- Step 2: 从隐含狄利克雷分布 α 中抽样生成文档 d_m 的主题分布 θ_m ;
- Step 3: 从主题的多项式分布 θ_m 中抽取文档 d_i 第 j 个词的主题 $Z_{m,n}$;
- Step 4: 从隐含狄利克雷分布 β 中抽样生成主题 $Z_{m,n}$ 对应的词分布 φ_k ;
- Step 5: 从词语的多项式分布 φ_k 中抽样生成单词 $W_{m,n}$ 。

4.2. 乘客群体满意度分析

对 4 类群体各项满意度指标存在的具体问题进行主题识别, 分别列出概率分布最大的主题中的三个权重最大的主题词, 如表 1 所示。

Table 1. Subject lines for specific questions for each indicator from 4 groups

表 1. 4 类群体对各项指标具体问题的主题词

乘客满意度指标	外地乘客群体	男性通勤群体	女性通勤群体	生活休闲群体
进出站	标识、混乱、不连续	安检、形式、拥挤	安检、包、慢	引导、标识、指引
环境与秩序	乞讨、拥挤、推销	温度、上下车、拥挤	温度、上下车、秩序	乞讨、小广告、微信
设施运行	噪声、电梯、扶梯	噪音、电扶梯、维修	电梯、检修、噪声	噪声、吵、直梯
换乘	标识、不方便、不醒目	挤、换乘距离、间隔时间	人多、等车、换乘标识	老年人、无障碍、距离
咨询	问题、态度、冷漠	态度差、不耐烦、恶劣	态度差、暴躁、冷漠	还行、凑合、不主动
投诉	投诉渠道、醒目、速度	投诉渠道、应答、回复	渠道、官话、回复	渠道、电话、回复
安全感	站台、安全隐患、拥挤	车厢、拥挤、踩踏	站台、拥挤、隐患	摔倒、帮助、工作人员

由表 1 可知, 进出站方面, 武汉地铁存在的两大问题是安检和标识。对于不熟悉站内环境的外地乘客群体和年龄相对较大的生活休闲类群体来说, 两类群体更关注进站标识是否连续、清晰、醒目; 对于出行频率高, 对站内环境熟悉的男性和女性通勤群体来说, 乘客更关注早晚高峰时段的安检效率。

环境与秩序方面, 男性和女性通勤群体往往会选择早晚高峰等乘客较多的时段出行, 在拥挤的站台上时往往更关注上下车秩序的问题, 在车厢上时则更关注温度; 生活休闲类群体出行目的则以购物和旅游/探亲访友为主, 往往会避开早晚高峰出行, 而此时也是乞讨卖艺者和广告推销者较多出现的时段, 因此该类群体更关注站内是否存在乞讨卖艺或散发广告的行为; 对于既有一部分乘客选择早晚高峰出行, 又有一部分选择平峰出行的外地乘客群体来说, 往往会同时关注拥挤状态和是否存在乞讨卖艺行为。

设施运行方面, 4 类群体的乘客提出的问题主要集中在电扶梯的运行状态和噪声。其中男性和女性通勤群体反映电扶梯经常检修, 极大影响早晚高峰时段进出站及换乘效率。

换乘方面, 男性和女性通勤群体认为换乘过程中较为拥挤、换乘距离长, 此外, 由于两类通勤群体对于时间敏感度较高, 因此普遍认为换乘后列车间隔时间长; 不熟悉站内环境的外地乘客群体则更为关

注换乘标识是否清晰、醒目、连续；而对于整体年龄相对较大的生活休闲类群体来说，则更关注换乘距离及无障碍设施情况。

咨询方面，男性和女性通勤以及外地乘客群体往往认为工作人员态度较差，而生活休闲类群体对工作人员的评价则偏向中性。

投诉方面，4类群体的乘客均认为武汉地铁未清晰的向乘客公布投诉渠道，并且投诉处理速度较慢，投诉回复不及时。

安全感方面，男性和女性通勤以及外地乘客群体更关注由于拥挤状态可能导致的安全隐患，生活休闲类群体则更担心摔倒后是否有工作人员提供帮助的问题。

4.3. 服务质量提升建议

针对四类群体对于各个乘客满意度指标提出的问题，本文对于武汉地铁各方面提出如下建议。

进站方面，一方面需确保进站导向标识连续、清晰、醒目，提升外地乘客群体和生活休闲类群体对于标识的满意度；另一方面，可以通过预约出行等方式，减少男性和女性通勤群体安检排队时间，提升安检效率。环境与秩序方面，一方面需加强早晚高峰时段工作人员的巡查频率，维护上下车秩序；另一方面根据夏季和冬季站外气温和车厢拥挤度动态调整空调温度；最后需加强平峰时段的巡查频率，对乞讨卖艺或推销广告者进行及时劝离。设施运行方面，需加强对于电扶梯的检查力度，确需维修的，应当避免在早晚高峰等时段检修。换乘方面，应系统梳理换乘导向标识，确保标识的连续、清晰、醒目。咨询方面，建议加强对工作人员的培训。投诉方面，武汉地铁应在车站醒目位置公布投诉渠道，并完善投诉处理闭环制度，确保问题的妥善解决与乘客的满意率。安全感方面，车站工作人员一方面应在早晚高峰等时段注意站台和车厢内秩序的维持，另一方面应重点关注中老年人的需求，主动提供帮助。

5. 结语

为精确识别各类地铁乘客对于武汉地铁的主要诉求，首先基于满意度调查问卷中收集到的乘客基本特征识别出外地乘客、男性通勤、女性通勤和生活休闲4类乘客群体，分析各类群体的主要特征，并在此基础上借助LDA主题模型对4类群体对于7项满意度指标存在的具体问题进行了主题识别，归纳核心诉求，最终对武汉地铁提升乘客满意度和服务质量提出了相应建议。论文结果可为武汉地铁从乘客诉求角度出发，精准改进不足，提升运营服务质量提供参考依据。

参考文献

- [1] 李世伟, 辛晓敏, 潘福全. 基于模糊综合评价方法的青岛地铁乘客满意度分析[J]. 城市轨道交通研究, 2020, 23(8): 117-119+124.
- [2] 刘晓娟. 基于模糊综合评价法的广州地铁乘客满意度分析[J]. 交通与运输, 2021, 37(6): 27-31.
- [3] 梁铭真, 陈红. 基于模糊理论的轨道交通乘客满意度指数评价[C]//中国科学技术协会, 交通运输部, 中国工程院, 湖北省人民政府. 2022世界交通运输大会(WTC2022)论文集(轨道交通与水上运输篇), 2022: 41-47.
- [4] 李林波, 郭晓凡, 傅佳楠, 吴兵. 基于云模型的城市轨道交通乘客满意度评价[J]. 同济大学学报(自然科学版), 2019, 47(3): 378-385.
- [5] 交通运输部办公厅. 关于印发《城市轨道交通服务质量评价规范》的通知[EB/OL]. http://www.gov.cn/zhengce/zhengceku/2019-10/29/content_5446219.htm, 2019-04-08.