

基于 K-means 聚类算法的航空旅客候机行为偏好分析

姚诗坤

南京航空航天大学, 中国·江苏 南京 211106

摘要: 分析航站楼中旅客候机行为, 可以优化航站楼布局, 有助于满足旅客多方面需求。通过某机场航站楼案例, 采用 K-means 聚类算法, 分析了旅客的候机行为, 明确旅客的路径聚类模式。结合旅客出行方式, 建立了机场航站楼对旅客吸引力模型。数据分析表明, 旅客收入水平和机场对于非职能区域选择概率呈现正面影响; 中青年旅客在非职能区域很容易发生消费活动; 航站楼商业设施对首次乘机的旅客有较强吸引力, 这对于优化航站楼空间布局, 提高管理水平有积极的作用。

关键词: K-means 聚类算法; 旅客行为; 旅客细分; 空间行为模式

Analysis of Airline Passenger Waiting Behavior Preferences based on K-means Clustering Algorithm

Shikun Yao

Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu, 211106, China

Abstract: Analyzing the waiting behavior of passengers in the terminal can optimize the layout of the terminal and help meet the diverse needs of passengers. Through a case study of a certain airport terminal, the K-means clustering algorithm was used to analyze the waiting behavior of passengers and clarify their path clustering patterns. A model of airport terminal attractiveness to passengers was established based on their travel modes. Data analysis shows that passenger income level and airport have a positive impact on the probability of non functional area selection; young and middle-aged travelers are prone to engage in consumption activities in non functional areas; the commercial facilities in the terminal have a strong appeal to first-time passengers, which has a positive effect on optimizing the spatial layout of the terminal and improving management level.

Keywords: K-means clustering algorithm; passenger behavior; passenger segmentation; spatial behavior pattern

0 前言

随着航空出行在现代城市生活中的日益普及, 旅客对于机场服务的需求已经从传统的“办票—值机—安检—乘机”流程型服务向多元化、个性化体验转变。现代航站楼不仅是交通中转空间, 更承载着购物、餐饮、娱乐、休憩等复合功能, 旅客候机行为也呈现出多样化特征。精准了解不同旅客群体的空间活动模式, 优化航站楼布局, 可以有效改善机场服务质量, 提高资源配置效率。为此, 论文基于实际运行数据, 采用 K-means 聚类算法, 分析旅客在航站楼内的路径行为提取其停留时间、活动区域密度、行为频次等多维指标, 构建空间行为特征矩阵; 结合可获得的旅客基本身份特征, 分析了旅客类型与行为偏好的关系特性。结果表明, 旅客根据收入、年龄、旅客性别等方面因素, 以及旅客出行目的地、频次等要素, 将其被划分为高流动性旅客、商业消费型旅客、休闲放松型旅客等多个类别, 各类别的行为偏好和功能区使用强度存在显著差异, 为机场运营者提供数据支撑, 提高了航站楼资源规划的科学性, 也有助于推动航站楼

由经验驱动型管理向数据驱动型智能化服务的转型。

1 基于 K-means 算法旅客运动行为聚类模型

1.1 数据采集

本研究基于 2023 年某国际机场 T2 航站楼国内出发旅客行为数据, 构建 K-means 聚类模型, 进一步挖掘航空旅客候机行为的潜在模式。调查在 2023 年 4 月 18 日与 4 月 21 日开展, 采集旅客在控制区内外的行为路径信息, 形成 968 条完整的行为轨迹数据, 主要包括旅客在航站楼内的活动时间、空间转移路径、功能区域停留特征。为了提高数据分析的精细化程度, 利用航站楼功能属性将其国内出发区域划分为子空间, 包括职能服务区、零售业态、餐饮设施、休息等候区、休闲娱乐场所等类别, 确保旅客行为路径能在空间维度上实现高分辨率映射。通过数据预处理提取旅客在各区域的驻留时长、移动频次及功能区转换率等多维指标, 构建标准化行为特征向量, 并结合密度峰值聚类算法优化初始聚类中心选择, 提高 K-means 模型的收敛效率, 增强旅客行为异质性的识别能力 (见图 1)^[1]。

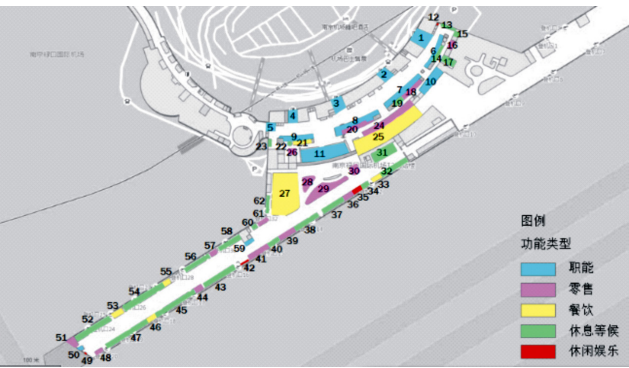


图 1 区域分布及功能类型

1.2 聚类要素

在基于航空旅客候机行为模式的 K-means 聚类模型构建中，研究旅客在航站楼功能区的选择行为，研究其潜在的行为特征。通过调查问卷获取某国际机场 T2 航站楼内旅客的停留区域信息，涵盖入口、值机柜台、购物区、餐饮区、休闲娱乐等功能单元，将旅客在移动过程中的每一次有目的性停留为一次空间决策，如购物消费、休憩就餐、信息查询等行为全部纳入统计分析范畴。在此基础上，构建旅客行为路径的空间转换模型，将原始行进轨迹映射为离散化的功能区访问序列，其中旅客是否进入某一功能区作为二元变量进行编码处理，即访问记为“1”，未访问记为“0”，由此生成的 0~1 数据链保留了个体行为的空间分布特征，结构化表达复杂路径行为。该方法解决旅客基本属性和时间维度的影响，注重分析空间要素的响应机制，加强功能性空间配置的认知程度，将每一个旅客的行为模式全部转化为一个 62 维的二值向量，代表旅客在航站楼内各功能区的选择倾向。接着，采用 K-means 聚类算法，识别出具有相似空间行为特征的群体，形成无数个典型候机行为模式，全面提高旅客动线组织的理解深度^[2]。

1.3 实现步骤

基于航空旅客候机行为模式的 K-means 聚类模型构建过程中，采用距离作为衡量样本相似性的核心指标，利用迭代优化，开展数据对象的空间聚类。本研究中，数据集 D 包含 n 个旅客行为路径样本，每个样本由 d 维属性描述其功能区选择特征，如停留时长、访问频次、动线转换率等。聚类目标是将具有相近行为偏好的旅客归为一类，形成 k 个互斥簇 C_i ($i=1,2,\dots,k$)，将每一个旅客行为路径仅归属在统一的聚类中，并且不同簇之间没有明显交集 ($C_i \cap C_j = \phi$)。

算法流程如下：
①从 n 个样本中随机选取 k 个初始聚类中心，作为迭代起点。

②计算所有样本和不同类别中心的距离，并且依据最小欧氏距离原则分配样本至最近簇，其距离计算公式为：

$$P(X_i, X_j) = \sqrt{\sum_{s=1}^d (X_{is} - X_{js})^2}$$
 (1)

式中， $P(X_i, X_j)$ 为对象 X_i 和 X_j 之间的相似度； X_i 和 X_j 为不同旅客的行为向量，s 为第 s 项属性维度。

③重新计算各簇均值作为新的聚类中心，并且将结果和前一次结果进行对比，如果在中心点发生变动，则返回步骤二继续迭代，直至聚类中心稳定不变。

④在聚类收敛性判断方面，采用误差平方和测度函数 E，科学评估聚类质量，其定义为：

$$E = \sum_{i=1}^K \sum_{p \in C_i} dist(p, m_i)^2$$
 (2)

式中，p 为第 i 簇中的样本；E 表示等待聚类的数据集 D 中所有数据对象的距离平方和； m_i 为簇 C_i 的中心。该函数反映样本和其所属簇中心之间的离散程度，数值越小表明聚类效果越好。当 E 值变化稳定，即达到收敛状态时，输出最终聚类结果，精准识别旅客候机行为的典型模式，提升机场运营效率（见图 2）。

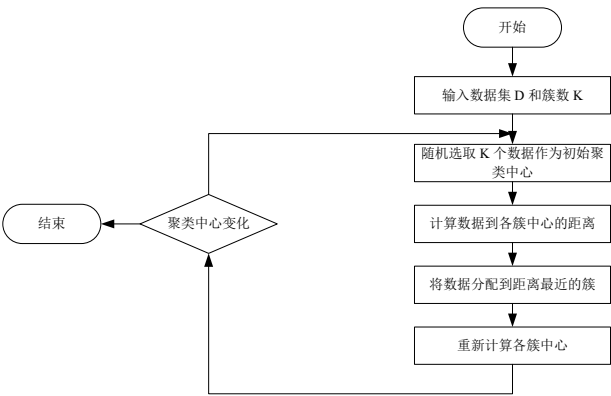


图 2 K-Means 聚类算法流程图

1.4 聚类结果的实例

根据肘部法（Elbow Method）确定最优聚类数目 k=5，将 968 名旅客划分为五类典型候机行为模式。各类别在功能区访问偏好上表现出显著差异，具体见表 1。

表 1 聚类结果实例分析

聚类类别	特征描述	访问偏好	占比（%）
C1	短时高频型：频繁穿梭于值机与零售区，停留时间短但路径复杂	值机柜台、免税商店	21.5
C2	消费主导型：侧重于购物和餐饮区域，消费意愿强烈，动线集中	零售业态、餐饮设施	17.8
C3	休憩导向型：主要停留在休息区与贵宾室，行为模式稳定	休息等候区、VIP 休息室	26.2
C4	时间驱动型：受航班时间限制影响较大，仅短暂访问必要功能区	出发大厅、安检通道	22.7
C5	休闲探索型：主动探索航站楼空间，访问儿童区、数字互动区等非核心功能区	休闲娱乐场所、互动展示区	11.8

进一步分析显示，C1 类旅客对零售业态的依赖程度最高，平均访问次数达 3.2 次，而 C3 类旅客在休息区的驻留时间最长，平均停留时长达 56 分钟。此外，C5 类旅客表现出较强的空间探索倾向，其功能区转换率为 0.78，显著高于其他类别。上述聚类结果验证了旅客候机行为的多样性，有助于机场运营方制定差异化服务策略。例如，针对 C1 类旅客可增设智能导航指引系统，提升值机与购物之间的路径效率；对 C5 类旅客应加强互动装置部署，提高航站楼空间吸引力。

2 不同旅客的行为偏好分析

2.1 旅客类型划分

结合航站楼中国出发区域空间布局特征，将旅客群体划分为 6 个样本集，指标共有 12 个，包括值机区访问频次、安检通道停留时长、购物行为强度、餐饮区使用频率、VIP 休息室利用率、信息咨询点访问率、路径复杂度指数、平均驻留时长、动线覆盖密度、数字互动区访问记录、儿童活动区使用情况、航班前最后停留区。

2.2 旅客空间行为偏好模式

基于 K-means 聚类算法，实证研究航空旅客候机行为模式，加强各簇样本分布的均衡性。将旅客初始分为 6 个样本集，即 K 值取 6，通过迭代聚类分析停留时间、动线密度、功能区访问频次等指标，最终形成具有显著差异的行为模式分类结果，如表 2 所示^[3]。

表 2 初始聚类结果

区域	旅客候机行为模式聚类						差异度 F	显著性 sig
	1	2	3	4	5	6		
1	0	0	0	0	0	0	10.426	0.046
2	0	0	0	0	1.000	0	23.833	0.003
3	1.000	0	0	0	0	1.000	12.663	0
4	0	0	0	0	0	0	13.711	0
5	0	0	0	1.000	0	0	25.312	0
6	0	1.000	0	0	0	0	11.383	0.038
7	0	0	0	0	0	1.000	12.121	0.012
8	0	1.000	0	0	0	0	13.411	0.031
9	1.000	1.000	0	0	1.000	0	28.711	0.012
10	0	0	0	1.000	0	0	30.426	0.019
11	0	0	0	0	0	1.000	18.540	0.059
12	0	0	0	0	0	0	15.435	0.012

通过分析上述表格内容，发现聚类中心 1 和中心 5 的数据对象有较强的相似性，说明在选取初始聚类中心中存在的较强的误差，所以在论文中将相似样本进行集成，这时 K 值为 5，再利用 SPSS 软件计算原始数据，得到表 3。发现

12 个区域的 sig 值均小于 0.05，可见两者候机行为模式类型存在较强差异性，无法进行二次分类^[4]。

表 3 优化聚类结果

区域	旅客候机行为模式聚类					差异度 F	显著性 sig
	1	2	3	4	5		
1	0	0	0	0	0	12.066	0.035
2	0	0	0	0	0	40.935	0.001
3	1.000	1.000	1.000	1.000	1.000	337.540	0
4	1.000	1.000	1.000	1.000	1.000	14.824	0.001
5	1.000	1.000	1.000	1.000	0	28.895	0
6	0	0	0	0	0	16.689	0.022
7	0	0	0	0	0	30.131	0.023
8	1.000	1.000	1.000	1.000	1.000	32.122	0.041
9	0	0	0	0	0	19.312	0.003
10	0	0	1.000	0	0	34.110	0
11	0	0	0	0	0	11.066	0.0003
12	0	0	0	0	0	27.623	0.013

2.3 结果分析

6 种旅客空间行为模式在基本属性上表现出明显的差异性，体现出不同类别旅客的行为特征和候机活动之间的内在关联。从性别分布来看，类型 3、类型 5 的女性占比均超过 68.81% 与 69.56%，体现出较强的女性主导性，和其出行目的有直接关联；而类型 1 和类型 2 则以男性为主，比例分别为 56.32% 与 57.74%，多为商务出行旅客。年龄结构分析表明，类型 4、类型 5 中青年群体占比高达 65.23%、66.39%，反映出年轻旅客更倾向于高效候机行为模式，并且对于机场设施具有较强的依赖度，消费意愿相对较高^[5]。而在收入水平方面，类型 5 旅客月收入 > 10000 元的比例达 50.24%，明显高于其他类型，说明高收入人群在机场行为选择上具有强目的性。是否首次乘机的数据显示，类型 4 旅客首次乘机比例仅 45.78%，表明该类旅客已经具备较高的机场熟悉度^[6]。

3 结语

论文在利用中国关于空间行为模式和 K-means 聚类算法相关研究的基础上，划分航站楼国内出发区域功能区，并基于实证问卷数据对旅客候机行为进行聚类分析。通过构建多维特征向量，涵盖旅客停留时长、活动区域分布、功能区访问频次等指标，识别出五类具有显著差异的行为模式，进一步结合旅客基本属性，分析在各类模式中的分布特征。研究发现，不同人口统计学特征的旅客在候机阶段的空间选择呈现明显异质性，反映出航站楼功能布局给旅客行为带来的影响。

参考文献:

- [1] 苏丹.基于非集计模型的航空旅客出行选择行为研究——以石家庄正定国际机场为例[D].河北:河北师范大学,2023.
- [2] 宁慧慧.有限理性视角下旅客对航空预付产品重复购买行为研究[D].天津:中国民航大学,2023.
- [3] 潘定一.基于偏好异质性的航空旅客品牌运价产品选择行为研究[D].天津:中国民航大学,2021.
- [4] 常柯新.品牌态度视角下航空公司网络舆情对旅客购买行为影响研究[D].天津:中国民航大学,2021.
- [5] 田园黄蝶.旅客权益保护视角下航空旅客不循规行为的判定[J].中国民航飞行学院学报,2024,35(3):69-74.
- [6] 刘天亮,高云蕊.航空安全保障信息公开对旅客航班选择的影响——基于意向偏好行为实验的证据[J].东方论坛,2023(3):99-106.

作者简介:姚诗坤(1991-),男,中国安徽合肥人,硕士,从事智慧机场研究。