

基于PCA-聚类分析的高铁旅客购票行为特性研究

刘帆涪^{a,b}, 彭其渊^{*a,b}, 梁宏斌^a, 傅志坚^a, 张斌^a

(西南交通大学 a. 交通运输与物流学院; b. 综合交通运输智能化国家地方联合工程实验室, 成都610031)

摘要: 铁路运输中旅客购票行为是铁路客运运营策略制定的重要基础。旅客购票行为直接影响着列车能力的占用过程, 是铁路客运票额组织的重要依据。根据贵阳—广州高速铁路的旅客购票统计数据, 以高铁购票旅客为样本, 运用主成分分析法购票行为的特征属性进行综合分析, 获得购票行为中重要特征变量。结合单次购票强度提出基于模糊C均值的双重聚类算法, 对购票旅客进行聚类, 并利用模糊聚类有效性指标 Xie-beni 和分离系数法确定最佳聚类数。结果表明, 高铁旅客购票行为的关键特性为单次出行旅客人数、购票提前天数、出行OD城市人均GDP和购票渠道; 不同旅客类型的购票行为有明显特性。

关键词: 铁路运输; 购票行为; 主成分分析; 模糊C均值; 双重聚类; 购票旅客

High-speed Railway Passenger Ticketing Behavior Characteristics Based on PCA and Clustering

LIU Fan-xiao^{a,b}, PENG Qi-yuan^{*a,b}, LIANG Hong-bin^a, FU Zhi-jian^a, ZHANG Bin^a

(a. School of Transportation and Logistics; b. National United Engineering Laboratory of Integrated and Intelligent Transportation, Southwest Jiaotong University, Chengdu 610031, China)

Abstract: The ticketing behavior is an important basis for the railway transportation operation planning. It influences the occupation process of the train capacity. Based on the ticketing statistics data from Gui-Guang high-speed railway, principal component analysis (PCA) is used to analyze the behavior characteristics comprehensively to obtain the critical characteristic variables, while the ticketing passengers are considered as the samples. Considering the tickets probability density function for a single purchase, the ticketing passengers are clustered by double clustering based on fuzzy C-means algorithm and the optimal number of cluster K is determined by fuzzy clustering effectiveness index: Xie-beni and Separation Coefficient. The result shows that the number of passengers per trip, the number of ticketing days in advance, GDP per capital of original-destination (OD) cities and pre-sell approach are the significant factors. The ticketing behaviors performed distinctively in different passenger types.

Keywords: railway transportation; ticketing behavior; PCA; fuzzy C-means; double clustering; ticketing passenger

收稿日期: 2017-06-20

修回日期: 2017-09-11

录用日期: 2017-09-19

基金项目: 国家自然科学基金/National Natural Science Foundation of China (U1234206, 61571375); 中央高校基本科研业务专项计划资助项目/Fundamental Research Funds for the Central Universities(2682013BR021).

作者简介: 刘帆涪(1985-), 女, 重庆长寿人, 工程师, 博士生.

*通信作者: piyuan-peng@home.swjtu.edu.cn

0 引言

目前,我国铁路运营里程已达14万 km,其中高速铁路线路里程已达2.6万 km.随着高铁路网规模的迅速扩大,旅客出行的选择方案也越来越多元化.从铁路运输企业角度考虑,研究高铁旅客出行选择行为特性,对于合理分配高铁列车运输能力、提高运营效益和社会效益都具有重要意义.

近年来,基于铁路旅客出行行为的研究获得了一系列成果.文献[1]对高速列车和高频率列车换乘等旅客出行模式选择的角度进行了研究.文献[2]从旅客主体特性、列车特性和随机因素3方面,对列车效用值分析了旅客乘车选择行为特征.文献[3]采用各种聚类算法对旅客就不同等级客票产品选择进行了研究.文献[4]对城际列车旅客出行时段和出行方式的选择进行了分析.文献[5]利用模糊综合评价方法对城际旅客的选择行为进行了分析.文献[6]提出高速铁路旅客出行时间选择的多项Logit模型,分析了影响旅客出行时间选择行为的关键因素.以上研究主要对铁路旅客出行的某种选择行为(出行时间选择、出行模式选择和客票产品选择等)进行研究.然而所有出行选择的最终表现形式都是以“购票行为”来实现.旅客购票行为是基于旅客出行各种选择行为的结果,体现了旅客出行的选择偏好,能更加全面地反映旅客的出行特性.因此研究高铁旅客购票行为特性有利于高铁售票组织策略的合理制定和列车能力利用率的提升.

高铁旅客购票行为属性的特征变量较多,分析关键变量是研究购票行为的重要基础.传统的旅客出行选择行为研究,主要依据研究者经验选择部分特征变量.但各种特征变量之间是相互制约和关联的,不同特征变量均可导致旅客不同的购票行为,因此需要综合分析各种特征属性.由于特征变量较多,要实现每个旅客的购票行为进行分析实则不易.运用PCA可在综合分析的基础上挖掘关键特征变量;为了在不知先验信息的情况下,通过购票旅客样本挖掘不同购票行为特性的内在联系,采用无监督学习的聚类分析可有效实现将具有类似购票属性的旅客进行分类,从而发现不

同类旅客的购票行为特性.

本文以贵广高速铁路D类与G类列车的购票旅客统计数据为样本,研究高铁旅客购票行为特性.旅客购票行为属性包含多种特征变量,首先利用主成分分析法(PCA)^[7],以 k 个主成份近似表达所有特征变量,且尽可能多地包含原信息.考虑单次购票强度对实际购票行为的影响,提出基于模糊C均值聚类^[8]的双重聚类分析法;以模糊聚类有效性指标Xie-Beni指标^[9]和分离系数指标^[10]进行有效性评价,确定最佳聚类数,使聚类结果最大化保留各类购票旅客的主要特征,将具有相似购票行为特性的旅客归类,从而为优化高铁售票组织策略、合理利用高铁列车运输能力的研究奠定理论基础.

1 高铁旅客购票行为统计分析

通过对贵广高铁本线列车进行调研,选取非特殊季节2016年11月21~27日期间运营的所有列车,自预售-60天至发车之日的旅客购票相关信息,共计 $7 \times 61 = 427$ 天的不同售票信息.通过多次调研,先后共获取到该线20个乘降作业车站的相关数据,164种不同车次列车,共计67 009次购票,一共售出229 543张客票.

首先对预售期内列车售票情况进行统计分析,得到一周内每日发车的累计售票量百分比(图1).由此可知,高铁旅客的购票行为具有一定特征分布规律,其特征属性是影响分布的主要原因,需首先分析购票旅客行为的特征属性.

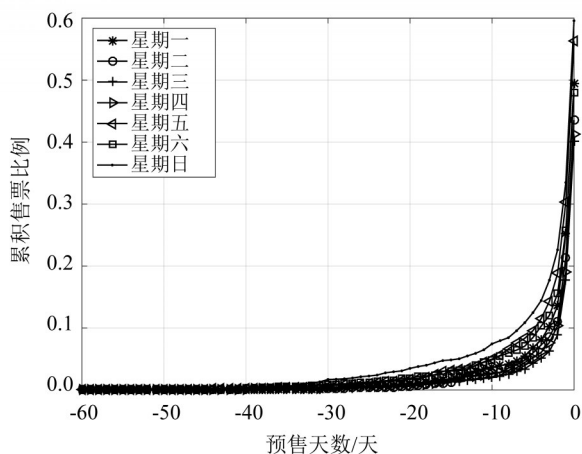


图1 预售期列车累计售票比例

Fig. 1 Rate of the calculative tickets in pre-sale period

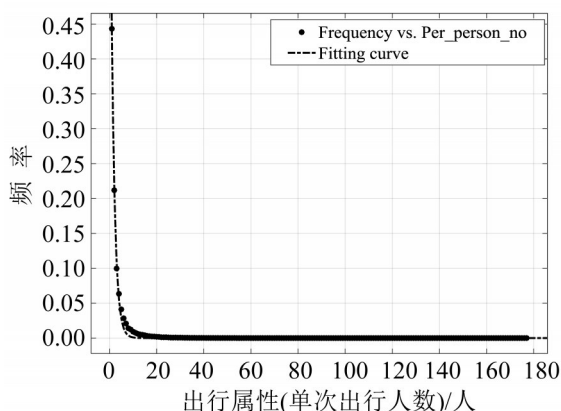


图2 预售期单次出行人数分布曲线

Fig. 2 The amount of passengers for a single trip

2 旅客购票特征属性

高铁旅客出行特征属性分为:出行因素(旅客出行OD与出行日期等非特殊情况下不会改变的因素)、购票因素(旅客为实现出行计划,在购票时可进行选择的因素)和个体因素(旅客年龄和性别等个体属性),如表1所示。

表1中,“单次出行人数”是指每次出行的旅客人数,在实际购票过程中的反映旅客的单次购票量。旅客单次购票量的概率密度函数(单次购票强度)可通过对购票数据的拟合得到,其结果呈指数分布(图2),表达式为

$$f(x) = 0.87 \cdot \exp(-0.68x) \quad (1)$$

表1 特征变量

Table 1 Characteristic variables

类别	序号	特征变量	特性	备注
出行因素	1	出行日期	定性	21~27日: 1~7
	2	上车城市人均GDP/万元	定量	贵阳北6.72; 龙里北3.52; 贵定3.28; 都匀东3.81; 三都县2.18; 榕江1.62; 从江1.51; 三江南1.31; 桂林北4.03; 阳朔4.08; 恭城2.44; 钟山西1.90;
	3	下车城市人均GDP/万元	定量	贺州1.39; 怀集2.14; 广宁2.46; 肇庆东3.06; 三水南18.05; 广州南13.96
购票因素	4	购票提前天数	定性	0~60
	5	列车类型	定性	1: G类列车; 0: D类列车
	6	客票等级	定性	高级动卧: 5 特等座: 4 商务座: 3 一等座: 2 二等座: 1
	7	票价	定量	—
	8	到达时刻	定量	—
	9	出发时刻	定量	—
个体因素	10	购票渠道	定性	窗口: 6 代售点: 5 自售机: 4 手机: 3 互联网: 2 电话: 1
	11	单次出行人数	定量	—
	12	性别	定性	男: 1 女: 0
	13	年龄	定量	—

3 特征属性的主成分分析

3.1 主成分分析原理

为避免因包含过多影响较小的变量影响分析和解决问题,首先对特征变量采用主成分分析。主成分分析主要利用降维思想,最终采用少数变量尽可能多地反映原来变量的信息,再根据相关系数发现强相关的特征变量,并作为关键因素给予更多的考虑和关注。

记矢量 $X^T = \{X_1, X_2, \dots, X_p\}$, X_i 为特征变量,共 p 个。利用特征方程 $R - \lambda I_p = 0$ 得到特征根 λ 及相应的正交化单位特征向量 a_i ; 找到一组特征变量的线性组合 Y_i :

$$Y_i = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ip}X_p \quad (2)$$

当 Y_i 为所有线性组合中方差最大的,则 Y_i 为第一主成分。当 Y_i 不能代表原特征变量绝大部分信息时,选取线性组合 Y_{i+1} ,且 Y_i 与 Y_{i+1} 的信息相互独立,

互不相关 $Cov(Y_i, Y_{i+1})=0$, 则 Y_{i+1} 为第二主成分.按此方式得到其余主成分^[7].由式(3)计算对应特征值累积贡献率 G ,选择大于95%时对应的主成份个数 m .

$$G(m) = \frac{\sum_{i=1}^m \lambda_i}{\sum_{k=1}^n \lambda_k} \quad (3)$$

求得所有主成份特征变量 i 与 j 的相关系数 r :

$$r_{ij} = \frac{\sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{\sqrt{\sum_{k=1}^n (x_{ki} - \bar{x}_i)^2} \cdot \sqrt{\sum_{k=1}^n (x_{kj} - \bar{x}_j)^2}} \quad (4)$$

3.2 关键特征变量的获取

根据以上分析方法可得到图3,进而得到当主成分数 $b=2$ 时,贡献值可满足阈值条件(大于90%).

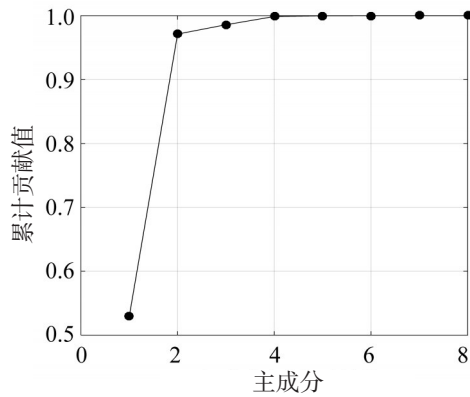


图3 主成分累计贡献

Fig. 3 Principal components contribution

根据PC1和PC2与特征变量的相关系数 r ,选取 $|r|>0.6$ 的特征变量为关键属性:单次出行人数、购票提前天数、出行OD城市人均GDP和购票渠道(图4).

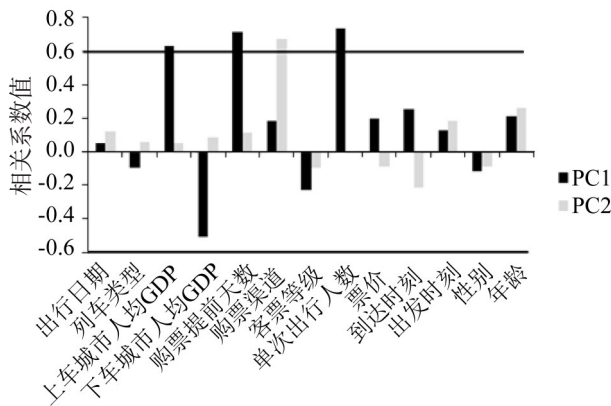


图4 主成分与主要特征变量的相关系数 r

Fig. 4 Correlation values r for principal components and features

4 高铁旅客购票行为的聚类分析

4.1 基于FCM的双重聚类算法

FCM算法是一种基于划分的聚类算法,其思想是使得被划分到同一类的对象之间相似度最大,而各类之间的相似度最小^[8].关于旅客购票行为研究的问题,由于购票旅客的特征数据通常是指以每个出行旅客为样本的购票信息,因此通常情况下FCM聚类分析过程的隶属度 u_{ij} 是关于单个出行旅客购票属性的函数,并非单个购票旅客的行为属性,未在聚类过程中考虑旅客一次购买多张票额的情况,极易忽略铁路客运运营团体类购票的实际需求.为了准确地反应购票旅客的行为,定义单次购票强度 $\gamma_{ij} \in R$ 为旅客单次购买的票额量.由此,隶属度就是关于出行旅客属性与单次购票强度的函数关系

$$\tilde{u}_{ij} = u_{ij} \cdot \gamma_{ij} \quad (5)$$

式中: \tilde{u}_{ij} 为修正后的新的隶属度函数; γ_{ij} 表示对隶属度函数所带来的影响,分布函数如式(1)所示.

在新隶属度函数下,令 \tilde{V} 为聚类中心矩阵, \tilde{U} 为模糊隶属度矩阵.令要划分的集合 $X = \{x_1, x_2, \dots, x_j, \dots, x_n\}$,每个元素 x_j 有 p 个特征,要把 X 分为 c 类($2 \leq c \leq n$).设有 c 个聚类中心 $\tilde{V} = \{\tilde{v}_1, \tilde{v}_2, \dots, \tilde{v}_c\}$, d_{ij} 为样本 x_j 与聚类中心 \tilde{v}_i 的欧式距离, u_{ij} 表示第 j 个样本在第 i 类中的隶属度;设 m 为模糊因子,经验取值范围为 $1.5 \leq m \leq 2.5$.由此,基于FCM双重聚类算法的目标函数定义为

$$\min J(\tilde{U}, \tilde{V}) = \sum_{i=1}^c \sum_{j=1}^n (\tilde{u}_{ij})^m \|x_j - \tilde{v}_i\|_{opt}^2 \quad (6)$$

$$\text{st. } \sum_{i=1}^c u_{ij} = 1, \quad j = 1, 2, 3, \dots, n \quad (7)$$

为了使式(6)达到极小的必要条件,采用拉格朗日乘子法进行求解.设拉格朗日乘子为 $\lambda = [\lambda_1, \lambda_2, \dots, \lambda_n]^T$,由此得到增广泛函为

$$J_{\text{lag}}(\tilde{U}, \tilde{V}) = \sum_{i=1}^c \sum_{j=1}^n (u_{ij} \gamma_{ij})^m \|x_j - \tilde{v}_i\|^2 + \sum_{j=1}^n \lambda_j (\sum_{i=1}^c u_{ij} - 1) \quad (8)$$

由最优一阶条件:

$$\frac{\partial J_{\text{lag}}(\tilde{U}, \tilde{V})}{\partial \lambda} = \left(\sum_{i=1}^c u_{ij} - 1 \right) = 0 \quad (9)$$

$$\frac{\partial J_{\text{lag}}(\tilde{U}, \tilde{V})}{\partial u_{ij}} = m(u_{ij})^{m-1} (\gamma_{ij})^m \|x_j - \tilde{v}_i\|^2 + \lambda_j = 0 \quad (10)$$

可以得到 $J_{lag}(\tilde{U}, \tilde{V})$ 在取得极小值时的 u_{ij} , 类似地方法可以得到对应最优一阶条件下的 \tilde{v}_i 的值, 从而得到迭代计算过程中的 \tilde{u}_{ij} 与 \tilde{v}_i 的迭代计算公式:

$$\tilde{u}_{ij} = \left[\sum_{k=1}^c \left(\frac{\|x_j - \tilde{v}_i\|_{opt}^2}{\|x_j - \tilde{v}_k\|_{opt}^2} \right)^{\frac{1}{m-1}} \right]^{-1} \cdot \gamma_{ij} \quad (11)$$

$$\tilde{v}_i = \frac{\sum_{j=1}^n \tilde{u}_{ij} x_j}{\sum_{j=1}^n \tilde{u}_{ij}} \quad (12)$$

因此, 改进FCM聚类算法与FCM算法的核心思想相同, 其主要步骤:

Step 1 初始化聚类中心矩阵 $\{\tilde{V}^l | l=0\}$, 将设定的聚类类别数 $c(2 \leq c \leq n)$, n 为样本总数, 确定模糊指标 m , 这里取值为2, 停止迭代阈值为 $\varepsilon \in [0.001, 0.01]$, 迭代次数 $t \leq l$, l 为最大迭代次数.

Step 2 根据聚类中心矩阵 \tilde{V}^l , 通过计算各样本 x_j 与聚类中心 \tilde{v}_i 的距离 d_{ij}^l , 利用式(9)对模糊划分矩阵 \tilde{U}^l 进行更新.

Step 3 根据模糊划分矩阵 \tilde{U}^l 和式(10)更新聚类中心矩阵.

Step 4 判定阈值. 对于给定的阈值 ε , 如果 $\|\tilde{V}^{l+1} - \tilde{V}^l\| \leq \varepsilon$ 或当迭代次数达到最大上限 l 时, 停止迭代; 否则令 $l=l+1$, 转到第2步. 迭代停止时, 可得到模糊聚类中心, 以及模糊划分矩阵参数估计值; 最后对样本 x_j 的聚类判决依据为

$$\omega = \arg \max_{i=1, \dots, c} \tilde{u}_{ij} \quad (13)$$

由于本文提出的双重聚类法是基于FCM的模糊聚类, 因此可采用具有代表性的模糊聚类有效性指标 Xie-Beni 指标 $V_{XB}^{[9]}$ 和分离系数指标 $V_{SC}^{[10]}$ 来进行有效性评价, 从而获取最佳聚类数 $K=4$ 时, 两种指标达到最优评价价值.

4.2 购票旅客聚类结果分析

基于PCA分析, 利用本文提出的改进FCM聚类方法和最佳聚类数 $K=4$ 得到聚类结果(图5). 由第1~3类分布可以看出PC2值越大PC1值增加; 第4类较特别, 对应较高的PC1值和较低的PC2值. 根据图3的主成份分析结果, PC1与单次出行人数、

提前购票时间和出行OD城市人均GDP强相关; PC2与旅客购票渠道强相关. 由此可以进一步分析各类旅客的关键特征属性.

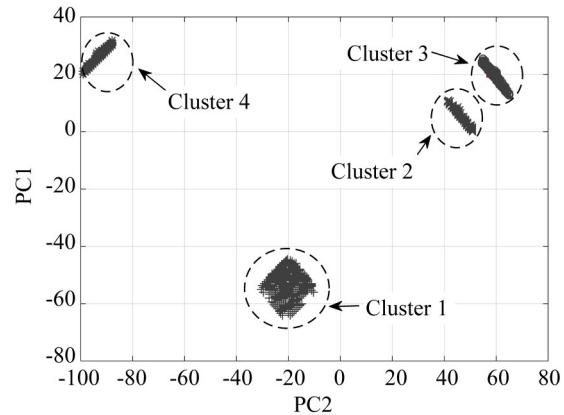


图5 购票旅客聚类结果

Fig. 5 Clustering results of ticketing passengers

图6为4类旅客在预售期内单次出行人数的分布. 第1类(C1)主要包含单次出行人数 $n \in [1, 5]$ 的私人出行旅客类型; 第2类(C2)集中于单次出行人数 $n \in (5, 20]$ 的多人出行旅客类型; 第3类(C3)主要为单次出行人数 $n \in (20, 50]$ 的小团体出行旅客类型; 第4类(C4)主要是单次出行人数 $n \in (50, 200]$ 的团体出行旅客类型. 单次出行人数越多, 购票时间相对提前.

表2 不同OD城市经济情况的旅客出行比例

Table 2 Passenger rate for different O-D economic levels (%)

旅客类别	“出发城市—到达城市”人均GDP水平			
	较高—较高	较高—较低	较低—较高	较低—较低
C1	26.4	27.5	11.9	34.2
C2	24.8	32.4	26.9	15.9
C3	18.1	15.2	35.8	30.9
C4	30.7	24.9	25.4	19.0

表2为不同类型旅客的出行OD城市人均GDP情况的所占比例. 根据国家统计局数据, 2016年全国人均GDP平均值为5.5万元, 当本文研究的城市人均GDP高于全国人均GDP时, 认为该城市处于经济状况较好的水平, 反之则是较低经济水平. 由表2可以看出, 在私人出行旅客(C1)中, 经济发展程度偏低的城市之间的占大多数; 多人出行旅客(C2)主要是从经济发展程度较高的城市到经济发展程度偏低的城市; 小团体旅客(C3)主要是从

经济发展程度偏低的的城市到经济发展程度较高的城市;团体出行旅客(C4)集中于经济发展程度较高的城市之间.

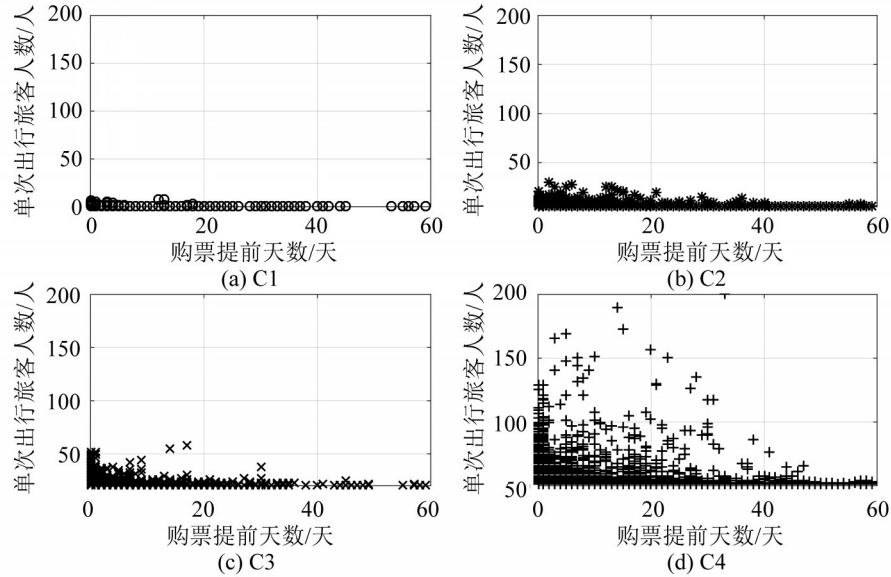


图6 预售期内各类旅客单次出行人数分布

Fig. 6 Distributions of passenger travel characteristic in pre-sell period

图7为4类旅客的购票渠道分布.其中,私人出行旅客(C1)购票渠道涉及面较广,包括窗口、代理点、自售机、手机和互联网等;多人出行和小团体(C2、C3)购票渠道主要集中于窗口、代售点和手机购票;团体出行旅客(C4)主要为窗口和代售点办理.基于以上分析,得到4类旅客类型的购票行为特性,如表3所示.

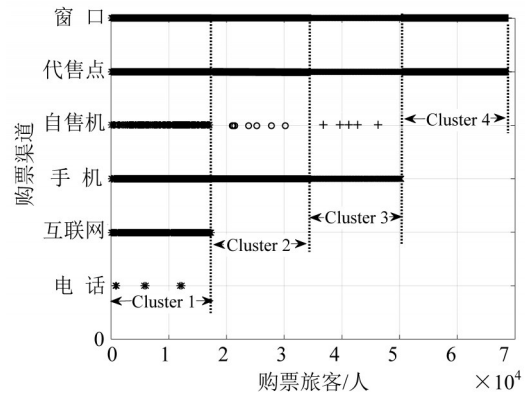


图7 各类购票渠道分布

Fig. 7 Distribution of ticketing approach for clusters

表3 各类旅客购票情况

Table 3 Ticketing situation for different clusters

旅客类别	旅客购票情况				主要购票渠道	售票数量/万张	比例/%
	单次出行人数n	平均提前购票天数	出行城市人均GDP/万元	到达城市人均GDP/万元			
C1	$1 \leq n < 5$	2	7.73	8.74	窗口、代理点、自售机、手机、互联网	14.75	64.2
C2	$5 \leq n < 20$	3	15.85	8.63	窗口、代售点、手机	4.5	19.6
C3	$20 \leq n < 50$	7	8.12	16.21	窗口、代售点、手机	1.75	7.6
C4	$50 \leq n < 200$	9	17.26	17.23	窗口、代售点	1.95	8.4

由表3可知,单次旅客出行人数对购票行为有一定程度的影响,人数越多,出行团体越大,购票时间越早.无论何种旅客类型,窗口和代售点仍是

目前采用最广泛,旅客接受度最高的购票渠道,越小的团体旅客出行更容易选择智能化购票(手机、互联网等).此外,私人出行旅客类型的购票量占绝

大部分(64.2%),由于出行人数较少,主要包含了沿线普通城市之间的来往的旅客;多人出行旅客类型购票量占19.6%,根据属性特征,该类旅客主要体现于回乡的行程;小团体出行旅客类型购票量仅占7.6%,主要体现于返城的行程;团体出行旅客购票量占8.4%,主要活动于沿线大城市之间,由于单次出行人数较多,主要通过窗口和代售点等人工办理。

5 结 论

以高铁旅客购票行为为研究对象,结合贵阳—广州高铁购票旅客的相关数据,在基于PCA分析购票行为主要特征属性基础上,结合实际购票行为特征提出基于FCM的双重聚类方法对购票旅客样本进行聚类,得出以下结论:

(1) 根据高铁旅客实际购票行为特征,提出基于FCM的双重聚类分析法。该方法能有效结合高铁旅客购票行为的各种特征变量,利于发现旅客购票行为关键特征属性的相互关系,可实现将具有相似购票特性的旅客进行划分,进而获得不同类别旅客购票行为特性,从而简化高铁售票组织策略和列车能力利用的相关问题的研究。

(2) 高铁旅客购票行为的主要特征属性为单次出行人数、购票提前天数、出行OD城市人均GDP和购票渠道。基于PCA分析,利用本文提出的双重聚类法将购票旅客分为4类:私人出行、多人出行、小团体出行和团体出行购票旅客,不同旅客类型有不同的购票行为特性。

基于不同购票行为特性的高铁售票组织策略研究及合理优化列车能力利用的有效方法,将是进一步研究的主要内容。

参考文献:

[1] 王爽,赵鹏. 基于Logit模型的客运专线旅客选择行为分析[J]. 铁道学报, 2009, 31(3): 6-10. [WANG S, ZHAO P. Analysis of passengers choice behavior for

dedicated passenger railway lines based on logit model[J]. Journal of the China Railway Society, 2009, 31(3): 6-10.]

- [2] 史峰,邓连波,霍亮. 铁路旅客乘车选择行为及效用[J]. 铁道学报, 2007, 28(6): 117-121. [SHI F, DENG L B, HUO L. Railway passenger choice behavior and its utility[J]. Journal of China Railway Science, 2007, 28(6): 117-121.]
- [3] 吕红霞,王文宪,蒲松,等. 基于聚类分析的铁路出行旅客类别划分[J]. 交通运输系统工程与信息, 2016, 16(1): 129-134. [LV H X, WANG W X, PU S, et al. Classification of railway passengers based on Cluster analysis[J]. Journal of Transportation Systems Engineering and Information Technology, 2016, 16(1): 129-134.]
- [4] 黄志鹏. 基于旅客出行选择行为的城际列车开行方案研究[J]. 武汉理工大学学报(交通科学与工程版), 2015, 39(6): 1135-1139. [HUANG Z P. Research on inter city train operation plan based on passenger travel choice behavior[J]. Journal of Wuhan University of Technology(Transportation Science & Engineering), 2015, 39(6): 1135-1139.]
- [5] 裴志远,刘万明. 用模糊综合评价方法分析旅客城际出行的选择行为[J]. 交通运输工程与信息学报, 2004, 1(2): 86-91. [PEI Z Y, LIU W M. Applying fuzzy synthetic appraisal method to analyze inter-city passenger's choice behavior[J]. Journal of Transportation Engineering and Information, 2004, 1(2): 86-91.]
- [6] 张航,赵鹏,乔珂,等. 高速铁路旅客出行时间选择Logit模型与分析[J]. 铁道运输与经济, 2017, 39(1): 55-60. [ZHANG H, ZHAO P, QIAO K, et al. Analysis on Logit model of high-speed railway passengers' travel time choice[J]. Railway Transport and Economy, 2017, 39(1): 55-60.]
- [7] WANG C W, JENG J H. Image compression using PCA with clustering[C]. International Symposium on Intelligent Signal Processing & Communications Systems, 2012, 41(11): 458-462.
- [8] BEZDEK J C. Pattern Recognition with Fuzzy Objective Function Algorithms[N]. Plenum Press, 1981, New York.
- [9] XIE X L, BEINI G. A validity method for fuzzy clustering[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 1991, 13(8): 841-847.
- [10] ZAHID N, LIMOURI M, ESSAID A. A new cluster validity for fuzzy clustering[J]. Pattern Recognition, 1999, 32(7): 1089-1097.