doi: 10.3969/j.issn.1672-6073.2022.02.015

基于谱聚类的城市轨道交通车站间客流分型研究

姚振康[1]，高国飞[2]，郑　汉[2]，黄兆察[3]
（1. 苏州轨道交通集团有限公司，江苏苏州 215004；2. 北京城建设计发展集团股份有限公司，北京 100037；3. 北京交通大学，北京 100044）

摘　要：在轨道交通运营管理过程中，不同车站间客流的时间分布特性及规律直接决定了客流组织方案。明晰车站间客流分型及特征，对合理配置客流组织方案大有裨益。相对于单车站客流类型，多车站间客流类型的影响因素多样且复杂，为此，从时间、空间和结构 3 个角度对车站间客流分类特征进行分析，并通过谱聚类方法压缩搜索空间，从而达到更加精准的类型划分。利用轮廓系数与戴维森堡丁指数对比不同方法的分类结果，证明所提出的谱聚类方法相对于 k-means 等其他方法具有更好的分类效果。以苏州地铁 2020 年数据为例，通过提出的方法寻找出 7 种车站间客流分型，该结果可应用于预测模型训练等领域。

关键词：城市轨道交通；车站间客流；谱聚类；多特征分析
中图分类号：U231　　**文献标志码：**A　　**文章编号：**1672-6073(2022)02-0099-06

Time Distribution Types of Passenger Flow Between Urban Rail Transit Stations Based on Spectral Clustering

YAO Zhenkang[1], GAO Guofei[2], ZHENG Han[2], HUANG Zhaocha[3]
(1. Suzhou Rail Transit Group Co., Ltd., Suzhou 215004; 2. Beijing Urban Construction Design & Development Group Co., Ltd., Beijing 100037; 3. Beijing Jiaotong University, Beijing 100044)

Abstract: In the process of rail transit operation management, the characteristics and rules of the time distribution of passenger flow between different stations directly determine the passenger flow organization scheme. The clarification of the interstation passenger flow time distribution types is beneficial to the reasonable configuration of the passenger flow organization scheme. This study analyzes the classification characteristics of interstation passenger flow from three perspectives of time, space, and structure, and compresses the search space of classification using a spectral clustering method to achieve a more accurate type classification. The classification results of different methods are compared using the silhouette coefficient and Davies-Bouldin index are used to demonstrate that the proposed spectral clustering method has better classification results compared with other methods such as k-means. Taking Suzhou Metro 2020 data as an example, seven types of interstation passenger flow time distributions are found by the proposed method, and the results can be applied to areas such as prediction model training.

Keywords: urban rail transit; passenger-flow between station; spectral clustering; multi-feature analysis

---

收稿日期：2021-06-10　　**修回日期：**2021-07-17
第一作者：姚振康，男，本科，政工师，研究方向为交通运输，yaozhenkang@sz-mtr.com
通信作者：高国飞，男，博士，高级工程师，gaoguofei_2007@sina.com
基金项目：苏州市科技发展计划(SS201830)
引用格式：姚振康，高国飞，郑汉，等. 基于谱聚类的城市轨道交通车站间客流分型研究[J]. 都市快轨交通，2022，35(2)：99-104.
YAO Zhenkang, GAO Guofei, ZHENG Han, et al. Time distribution types of passenger flow between urban rail transit stations based on spectral clustering[J]. Urban rapid rail transit, 2022, 35(2): 99-104.

经过多年的发展，城市轨道交通已成为大型及特大城市的有机组成部分，起到沟通城市区域的骨干作用。由于城市功能分区的差异性，城市轨道交通系统所服务的车站间客流对象自然地存在差异。例如，联通商业区与居住区间的客流呈现出潮汐性规律，旅游区与枢纽区之间的客流与周末节假日高度相关。为实现轨道交通系统中客流的精准管理与预测，了解不同车站间客流在时间上的分布类型及特性尤为关键，精准的车站间客流类型划分可以细化预测等技术，支撑精细客流组织方案。然而，相对于单车站客流分类，车站间客流类型受到更多因素的影响，例如，李向楠[1]对客流站点高峰小时乘降量、车站规模等 11 个特征因素进行聚类分析，最终将成都地铁 1 号线现运营 16 个站点划分为 5 大类；Chen 等[2]选取 36 个特征因子作为聚类分析的初始变量，进行聚类分析；高勃等[3]利用 k-means 对站点重要度进行聚类分析。以上研究都仅限于站点的聚类分析。同时，已有较多的聚类方法应用于轨道交通领域的研究，包括 k-means[1]、改进 k-means[4]、基于密度聚类[5]、模糊聚类[6]以及谱聚类[7]等。

考虑到谱聚类在高维特征聚类方面具有优势，因此本文选取谱聚类方法对车站间客流(车站 OD 客流)进行聚类分析，并以苏州地铁 2020 年数据为例，寻找出 7 种车站间客流时间分布类型，该结果可应用于预测模型训练等领域。

1 客流 OD 分类特征

轨道交通客流 OD 的分类取决于乘客出行的时间特征、空间特征和客流结构，可拓展为乘客平均旅行时间、不同时段出行量、发生吸引站的用地特征、车站特点、客票种类等具体指标。

1.1 客流 OD 时间特征

客流 OD 的时间特征包括乘客乘坐轨道交通的平均旅行时间(T)，不同时段下的出行量($F=\{F_{\mathrm{MP}},F_{\mathrm{EP}},F_{\mathrm{WD}},F_{\mathrm{WE}}\}$)等。其中乘客的平均旅行时间通过一段时间内 AFC 数据中各个 OD 乘客进出站的时间来确定，由于不同 OD 的乘客等待时间、在车时间、换乘时间等都不相同，不同 OD 的平均旅行时间也会有所差异。同时，不同 OD 在不同时段下的出行比例也不相同，根据乘客出行时段特点，对一天内早高峰(F_{MP})、晚高峰(F_{EP}) 2 个时段的客流；一周内，工作日(F_{WD})和休息日(F_{WE})的客流进行统计和分析。

1.2 客流 OD 空间特征

客流 OD 在空间上的特征为：客流发生、吸引车站的用地特征(LF^{o}、LF^{d})、与市中心的距离(D_{o}、D_{d})，以及 OD 直线距离(L)。

轨道交通站点与市中心距离则为轨道交通站点至城市中心点的直线距离，用于描述轨道交通车站与城市中心的相对位置；OD 平均距离指该 OD 对中乘客乘坐轨道交通出行的平均距离；用地特征根据用地分类标准[8]并结合轨道交通车站特点，对站点用地特征分类时，计算车站中心点半径 500 m 范围内各类用地类型面积，设置用地面积占比截断阈值 p_m，确定每个车站的用地特征主导因素，根据面积占比大小确定轨道交通站点的用地主导类型：①对于只有一种类型用地占比超过截断阈值 p_m，则将其作为该区域的主导用地类型，如办公主导型(LF_0)、居住主导型(LF_1)、商业主导型(LF_2)、枢纽主导型(LF_3)、旅游主导型(LF_4)；②对于存在多种类型用地占比超过阈值 p_m 时，则判定为混合型，主要的混合型包括办公居住混合型(LF_5)、办公商业混合型(LF_6)和居住商业混合型(LF_7)；③对于涉及类型较多，且各类型用地占比均未超过阈值 p_m 时，归为综合型(LF_8)。

1.3 客流结构特征

客流结构特征主要通过客票种类(T)进行区分，在苏州轨道交通系统中，客票种类主要可以归纳为以下 5 种：一卡通(T_1)、单程票(T_2)、公交卡(T_3)、次票(T_4)、App(T_5)等。

2 OD 分型聚类与评价方法

2.1 基于谱聚类的 OD 分型算法

轨道交通乘客出行的 OD 类型受到时间特征、空间特征和客流特征 3 大类的影响，其影响指标多(共计 16 个)，特征复杂，因此描述 OD 类型的特征向量的维度也较高。如此高维的复杂数据在使用传统的聚类方法(如 k-means)进行直接聚类时，往往达不到预期的效果。为了实现高维数据的聚类分析，本研究利用谱聚类算法，对 OD 特征数据进行聚类分析。

谱聚类是在谱图理论的基础上，将聚类问题转化为图的划分问题，从而提高聚类算法的效率，适用于高维数据的聚类分析。对于谱聚类算法而言，OD 特征数据集中的每一条 OD 特征数据都是图 $G(V, E)$中的一个点，其中，V 是数据点的集合，E 为所有边的集合。为了建立邻接矩阵 $\boldsymbol{W}$ 和相似矩阵 $\boldsymbol{S}$，本文采用

基于高斯径向核函数 RBF 的全连接法对任意两点 v_i 和 v_j(v_i，$v_j \in V$)间的权重 w_{ij} 进行计算，此时权重 w_{ij} 和距离 s_{ij} 相等。根据 Ncut 的图切割方法，利用标准化的 Laplacian 矩阵、矩阵的度和主成分分析(PCA)思想来优化图的切割效果，计算特征矩阵，最后对特征矩阵进行 k-means 聚类。

谱聚类的计算流程包括：①以 OD 特征数据样本集为输入，构建样本相似矩阵；②以 Ncut 为图划分准则，计算 Laplacian 矩阵及其特征向量，形成标准化后的特征矩阵；③利用 k-means 算法对降维后的特征矩阵进行聚类分析，最终输出 OD 类型划分。谱聚类算法具体步骤如下所示。

步骤 1：输入 OD 特征数据集 $Set_OD=(od_1, od_2, \cdots od_n)$，$od_i$ 为 k 维向量；

步骤 2：计算邻接矩阵和相似矩阵的元素 $w_{ij}=s_{ij}=\exp\left(-\frac{\left\|od_i-od_j\right\|_2^2}{2\sigma^2}\right)$，从而构建邻接矩阵 $\boldsymbol{W}$ 和相似矩阵；

步骤 3：计算 Laplacian 矩阵 $\boldsymbol{L}=\boldsymbol{D}-\boldsymbol{W}$，$\boldsymbol{D}$ 为度矩阵；

步骤 4：标准化 Laplacian 矩阵 $\boldsymbol{L_b}=\boldsymbol{D}^{-1/2}\boldsymbol{L}\boldsymbol{D}^{-1/2}$；

步骤 5：计算 $\boldsymbol{L_b}$ 最小的 k_1 个特征值所各自对应的特征向量 f，$k_1<k$；

步骤 6：标准化特征向量 f，形成特征矩阵 $\boldsymbol{F}_{n\times k_1}$；

步骤 7：对特征矩阵 $\boldsymbol{F}$ 进行 k-means 聚类，聚类维数为 k_2，$k_2<k_1$；

步骤 8：输出 OD 类型划分 $Type_OD=\left(t_1, t_2, \cdots t_{k_2}\right)$。

2.2 聚类结果评价指标

一般而言，紧凑性和发散性是评价聚类结果质量的两个重要方面[9]，为了评价谱聚类算法得到的聚类结果，本文引入轮廓系数[10](Silhouette coefficient)和戴维森堡丁指数[11](Davies-bouldin index)，对聚类结果进行评价，这两者均兼顾了聚类结果的紧凑性和发散性，具体计算方法如式(1)和式(2)所示。

$$SC=\frac{1}{N}\sum_{i\in I}\frac{b(i)-a(i)}{\max\{a(i),b(i)\}} \tag{1}$$

其中，i 为簇中任意一点；N 为总样本数；$a(i)$为 i 到同一簇内其他点不相似程度的平均值；$b(i)$为 i 到其他簇的平均不相似程度的最小值。

$$DBI=\frac{1}{n_c}\sum_{i=0}^{n_c-1}\max_{j\neq i}\left(\frac{\overline{SP}_i+\overline{SP}_j}{v_i-v_j}\right) \tag{2}$$

其中，$\overline{SP}_i$、$\overline{SP}_i$ 为第 i、j 个簇的分散程度；v_i、v_j 分别为第 i、j 类中心点的属性向量；v_i-v_j 为第 i 类与第 j 类中心的距离；n_c 为分类后的簇类数。

3 基于苏州轨道交通客流 OD 的实例分析

3.1 苏州轨道交通客流 OD 特征

本研究以 2020 年 12 月 AFC 数据作为研究对象进行分析。截至 2020 年 12 月，苏州轨道交通运营车站共计 135 个，因此乘坐城市轨道交通出行的 OD 最多存在 135×134=18 090(对)，而通过对异常值、缺失值的筛选，除去同站进出、员工卡、超出运营时间的 OD 数据，最终统计得到有效 OD 共计 15 212 对。日均客流 74.71 万人次，平均旅行时间 0.65 h，其中早高峰时段客流平均16.59万人次，晚高峰时段平均17.58万人次。与其他城市不同，苏州轨道交通周末客流量依旧较大，工作日与周末日均客流比例为 1.06∶1。从客票种类来说单程票占比最多，占所有客票种类的 43.5%，其次市民卡占比 22.9%。在划分车站主导用地类型时，确定用地截断阈值为 0.4。

选取该月较有特点的两组 OD(A 组：苏州火车站—察院场；B 组：独墅湖南—月亮湾)，对其时间特征、空间特征及客流特征进行展开分析(见表 1、表 2 和表 3)。如表 1 所示，在时间特征上，A 组是全月客流最大的 OD 对，晚高峰小时客流量远大于早高峰小时客流量，且周末日均客流约为工作日日均客流的两倍，相比之下，B 组客流总量略低于 A 组，但其早晚高峰客流潮汐性明显，且工作日日均客流也远大于周末日均客流；在空间特征上，苏州火车站是苏州重要客运枢纽，辐射范围广，承接苏州周边及其他城市的客流，而察院场位处苏州重要的商业中心，周边商业发达，两者距市中心较近，独墅湖南和月亮湾周围用地则分别是居住主导型用地和办公主导型用地，两地相距市中心较远；在客流特征上，A 组的单程票客流占到 77%，其余票种较少，而 B 组虽然单程票的比重依然最大，但市民卡和一卡通的比例大幅增加，其比例为 46%(如图 1 所示)。

通过 OD 时间特征、空间特征和客流结构特征可以推测苏州火车站—察院场以旅游客流为主，而相比之下，独墅湖南—月亮湾客流则以通勤客流为主。

表 1　两组客流 OD 时间特征数据

Table 1　Temporal dimensions in the two groups of origin-destination passenger flow data

OD	总量/人次	平均旅行时间/s	早高峰时段日均客流量/人次	晚高峰时段日均客流量/人次	工作日日均客流/人次	周末日均客流/人次
苏州火车站—察院场	39 073	725	48	261	1 097	1 729
察院场—苏州火车站	36 749	692	68	233	884	2 050
独墅湖南—月亮湾	31 569	572	697	22	1 243	372
月亮湾—独墅湖南	28 388	572	10	502	1 114	343

表 2　两组客流 OD 空间特征数据

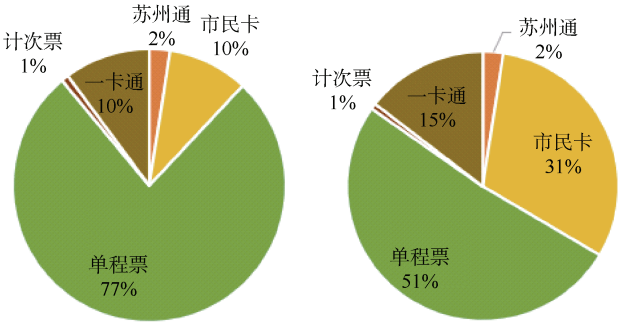
Table 2　Spatial dimensions in the two groups of origin-destination passenger flow data

OD	发生站类型	吸引站类型	发生站距中心距离 D_o/km	吸引站类型距中心距离 D_d/km	OD 直线距离/km
苏州火车站—察院场	枢纽主导型	商业主导型	3.560	2.941	2.31
察院场—苏州火车站	商业主导型	枢纽主导型	2.941	3.560	2.31
独墅湖南—月亮湾	居住主导型	办公主导型	13.063	14.757	4.81
月亮湾—独墅湖南	办公主导型	居住主导型	9.3961	13.063	4.81

表 3　两组客流 OD 结构特征数据

Table 3　Structural dimensions in the two groups of origin-destination passenger flow data

OD	苏州通	市民卡	单程票	计次票	一卡通
苏州火车站—察院场	960	3 719	30 084	340	3 967
察院场—苏州火车站	977	3 523	28 258	201	3 782
独墅湖南—月亮湾	766	9 818	16 180	219	4 586
月亮湾—独墅湖南	770	9 226	13 888	284	4 219



(a)苏州火车站-察院场　　(b)独墅湖南-月亮湾

图 1　两组客流 OD 客流结构分布

Figure 1　The illustration of passenger flow structures of the two groupd of origin-destination data

为了更好地对所有 OD 的特点进行定量分析，总结乘客出行规律和特征，运用谱聚类算法对此高维特征数据进行聚类分析，并对比其他聚类方法对比聚类效果。

3.2　谱聚类算法实现与指标分析

根据 2.1 谱聚类算法实现流程，编写 python 程序，对 OD 数据进行聚类分析，其中权重计算采用高斯核函数，图切割方法采用 Ncut 方法。为了选取合理的核函数参数 γ，研究计算了不同 γ 下的 *SC* 聚类评价指标(见表 4)，并绘制折线图(见图 2)。

表 4　不同高斯核参数下轮廓系数计算统计

Table 4　Statistical table of contour coefficient calculation with different Gaussian kernel parameters

聚类簇数	γ=0.01	γ=0.1	γ=1	γ=2	γ=3	γ=5
3	0.322 131	0.321 924	0.324 739	0.327 913	0.328 428	0.632 395
4	0.221 09	0.220 615	0.220 552	0.216 102	0.643 752	0.633 15
5	0.204 954	0.205 176	0.198 645	0.236 585	0.637 952	0.628 444
6	0.200 186	0.202 189	0.207 563	0.233 832	0.620 652	0.619 62
7	0.194 419	0.195 563	0.215 8	0.243 812	0.616 88	0.616 604
8	0.192 15	0.195 298	0.171 691	0.188 032	0.616 378	0.532 359
9	0.172 279	0.177 345	0.177 499	0.218 176	0.555 068	0.526 53
10	0.181 833	0.154 374	0.175 094	0.183 159	0.329 283	0.510 188
11	0.162 206	0.157 035	0.158 283	0.180 72	0.229 39	0.510 061
12	0.164 542	0.152 311	0.153 303	0.189 714	0.145 768	0.370 72
13	0.155 896	0.162 321	0.166 06	0.188 174	0.163 995	0.464 258
14	0.152 888	0.148 49	0.161 56	0.181 316	0.198 981	0.351 068
15	0.142 889	0.142 958	0.163 649	0.161 527	0.134 4	0.346 691

从图中结果可以看出，当 γ=3 时，在聚类簇数 $n_c\in[4, 9]$聚类效果较好，且超过其他参数值聚类效果，因此选取 γ=3 作为核函数参数。由于当 $n_c\in[4, 9]$时，*SC* 指标计算结果较为接近，为了更好地确定合理聚类簇数，深入计算了 DBI 指标，计算结果如图 3 所示。

根据图 3 的聚类结果，当 n_c=7 时 DBI 指数出现极小值点(DBI 越小聚类结果越好)，因此综合两个聚类指标的计算结果，确定 n_c=7 为最终的聚类簇数。在此条件下，分别计算 k-means、Mini Batch k-means 以

及层次聚类(Hierarchical clustering)的聚类结果，并对其指标进行计算(见表 5)。

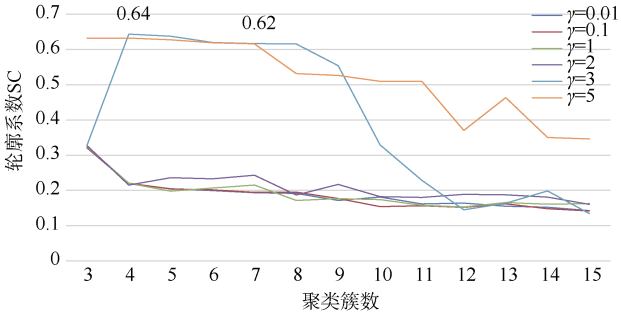


图 2　不同高斯核参数下轮廓系数计算结果

Figure 2　The illustration calculation results of contour coefficients with different Gaussian kernel parameters

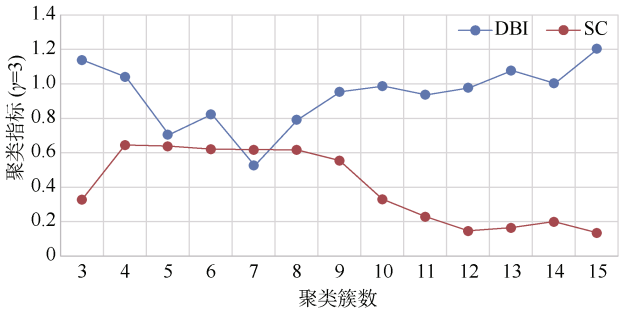


图 3　γ=3 时聚类结果指标对比

Figure 3　Comparison of clustering result indicators when γ=3

从表 5 中可以看出，谱聚类算法相较其他算法的指标计算结果较优，聚类效果较好。

3.3　聚类结果分析

聚类结果压缩成二维后如图 4 所示，部分簇的聚类效果较好，但簇间也存在散点相互交叉的现象；各个簇的散点数不同但无数量级的差异，其中聚类 5 的样本数最多，聚类 1 的样本数最少(见表 6)。为了减少分析过程中离群点的影响，依据用地特征对前 80%的样本进行分析，并从时间特征、空间特征和客流结构进一步分析聚类后各个簇的特征。

聚类 0：包括了独墅湖南—月亮湾等 1 221 个 OD 对。在空间特征上，客流 OD 对的主要用地类型为

表 5　谱聚类与其他算法聚类结果对比

Table 5　Comparison of spectral clustering and other algorithms clustering results

聚类算法	DBI	SC
谱聚类	0.525 3	0.616 7
k-means	1.371 2	0.209 0
Mini Batch k-means	1.521 0	0.209 0
层次聚类	1.456 1	0.173 2

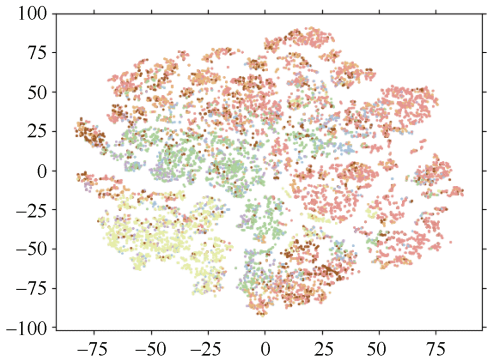


图 4　聚类结果降维散点图

Figure 4　Reduced dimensional scatter plot of clustering results

表 6　聚类结果特征分析

Table 6　Characterization of clustering result

聚类	聚类 0	聚类 1	聚类 2	聚类 3	聚类 4	聚类 5	聚类 6
样本数	1 221	2 239	1 711	1 246	2 316	4 338	2 140
用地特征	混合—混合、居住—办公	枢纽—旅游、枢纽—商业	枢纽—居住、旅游—居住	居住—居住	居住—枢纽、居住—旅游	办公—居住、居住—办公	旅游—枢纽、商业—枢纽
发生地距市中心距离/km	12.65	16.12	11.64	12.44	6.52	3.84	5.94
吸引地距市中心距离/km	14.75	5.58	6.86	12.34	8.94	4.22	15.62
OD 对直线距离/km	16.62	14.15	10.46	2.87	8.93	3.06	13.02
日均 OD 量	20.90	29.05	23.33	113.54	57.22	63.30	29.81
平均旅行时间/s	3 832	2 924	2 442	557	2 321	812	2 842
早高峰时段客流系数	0.28	0.11	0.15	0.22	0.12	0.28	0.16
晚高峰时段客流系数	0.10	0.15	0.18	0.21	0.21	0.29	0.25
工作日日均客流系数	0.18	0.14	0.12	0.15	0.15	0.15	0.14
周末日均客流系数	0.14	0.18	0.20	0.12	0.17	0.13	0.18
单程票占比/%	41.45	69.81	45.69	39.56	49.82	40.65	67.55

居住办公混合—商业办公混合、居住—办公，其发生吸引地距离城区较远，且相对距离较远；在时间特征上，OD 客流的平均旅行时间也最长，工作日客流远大于周末客流，早高峰时段客流大于晚高峰时段客流；在客流结构上，除单程票外，市民卡和一卡通的客流也占据一定的比例。因此可将此类 OD 归纳为市郊进城通勤客流。

聚类 1：包括了苏州新区火车站-察院场等 2 239 个 OD 对。在空间特征上，客流 OD 对的主要用地类型为枢纽—旅游、枢纽—商业，且发生地距离城区较远，但吸引地较城区较近，相对距离较远；在时间特征上，OD 客流的平均旅行时间较长，周末客流远大于工作日客流，晚高峰时段客流略高于早高峰时段；在客流结构上，以单程票为主，其余票种数量较少。因此可将此类 OD 归纳为外市进城旅游客流。

聚类 2：包括了山塘街-宝带路等 1 711 个 OD 对。在空间特征上，客流 OD 对的主要用地类型为枢纽—居住、旅游—居住，发生、吸引地距中心城区的平均距离适中，相对距离适中；在时间特征上，OD 客流的平均旅行时间适中，周末客流远大于工作日客流，晚高峰时段客流高于早高峰时段；在客流结构上，以单程票、市民卡为主，其余票种数量较少。因此可将此类 OD 归纳为本市返城旅游客流。

聚类 3：包括了钟南街—独墅湖邻里中心等 1 246 个 OD 对。在空间特征上，客流 OD 对的主要用地类型为居住，发生、吸引地距中心城区的平均距离均较远，但相对距离较近；在时间特征上，OD 客流的平均旅行时间适中，周末客流略大于工作日客流，早晚高峰时段客流相差不大；在客流结构上，以单程票、市民卡、一卡通为主，其余票种数量较少。因此可将此类 OD 归纳为短途出行客流。

聚类 4：包括了宝带路—山塘街等 2 316 个 OD 对。在空间特征上，客流 OD 对的主要用地类型为居住—枢纽、居住—旅游，发生、吸引地距中心城区的平均距离较为适中，相对距离适中；在时间特征上，OD 客流的平均旅行时间适中，周末客流略大于工作日客流，晚高峰时段客流略大于早高峰时段；在客流结构上，以单程票、市民卡为主，并有部分次票，其余票种数量较少。因此可将此类 OD 归纳为本市出城旅游客流。

聚类 5：包括了横塘—人民桥南等 4 338 个 OD 对。在空间特征上，客流 OD 对的主要用地类型为办公—居住、居住—办公，发生、吸引地距中心城区的平均距离均较近，相对距离较近；在时间特征上，OD 客流的平均旅行时间短，工作日客流远大于周末客流，早晚高峰时段客流较大且相差较少；在客流结构上，以单程票、市民卡为主，其余票种数量较少。因此可将此类 OD 归纳为市区通勤客流。

聚类 6：包括了山塘街—高铁苏州北站等 2 140 个 OD 对。在空间特征上，客流 OD 对的主要用地类型为旅游—枢纽、商业—枢纽，发生地距中心城区的距离较近，吸引地较远，相对距离较远；在时间特征上，OD 客流的平均旅行时间较大，周末客流远大于工作日客流，晚高峰客流略高于早高峰时段客流；在客流结构上，以单程票为主，其余票种数量较少。因此可将此类 OD 归纳为外市返城旅游客流。

综上所述，将 7 类客流 OD 归纳为以下 7 种客流(见表 7)。

表 7　客流 OD 归纳

Table 7　Types of origin-destination passenger flows

聚类簇	客流种类
聚类 0	市郊进城通勤客流
聚类 1	外市进城旅游客流
聚类 2	本市返城旅游客流
聚类 3	短途出行客流
聚类 4	本市出城旅游客流
聚类 5	市区通勤客流
聚类 6	外市返城旅游客流

4　结论

本文以苏州轨道交通 OD 客流数据为基础，从 OD 客流的时间特征、空间特征、结构特征 3 个角度对车站间客流特征进行分析，并通过谱聚类算法对高维客流 OD 数据进行聚类分析，从而实现精准分型；利用轮廓系数与戴维森堡丁指数对比不同方法的分类结果，证明谱聚类方法相对于 k-means 等其他方法具有更好的分类效果。通过归纳其聚类特点，最终得出市郊进城通勤客流等 7 种车站间客流时间分布类型。本文为不仅为客流的分类提供了一种更具操作性的定量分析方法，也为后续的客流预测等研究奠定基础。

（下转第 116 页）