

基于客流特征的城市轨道交通站点类型分类

——以南京地铁为例

1 简介

城市轨道交通车站是城市轨道网络中的关键节点，是服务于乘客乘降、换乘及候车的场所。轨道交通车站是乘客与轨道交通系统发生联系的纽带，轨道交通线网客流特征实质是车站客流时空分布叠加及交换的效果体现。车站客流量存在差异性，表明了城市不同功能分区对应的城市生活的时空的不均衡。由于站点所在的城市区位、交通功能以及周围土地利用性质不同，站点客流存在差异性，同样也存在着相似性，科学合理地对站点分类，有助于进一步研究站点与土地利用、人口岗位之间的相互关系。

城市轨道交通站点的分类一般分为定性和定量两种方法。定性分类法一般是根据车站周边一定范围内的主要用地性质进行划分。例如傅搏峰等根据影响区域内土地利用性质将上海郊区车站划分交通枢纽类、文教区类、景点区类、公共中心区类、居住区类、工业园区类及混合区类站点。这种方法的优点在于其直观性、便捷性，缺点是说服力不足、主观性强，对于周边用地混合开发的车站判断存在模糊性。

定量分析方法一般采用聚类分析法，通过站点客流相关指标进行类型划分，例如陈艳艳等根据车站进站客流的早晚高峰系数、客流量级，将北京地铁站点分为4类。选择全日不同时段的进出站客流量作为聚类分析指标。余丽洁等根据客流特征、站点规模、公交接驳线路、土地利用等参数，利用谱聚类对站点进行分类。

本文以南京地铁刷卡数据为例，统计南京地铁2017年某工作日的分时段进出站客流量，分析不同典型站点客流的时空分布特征，利用K-Means聚类分析将113个站点分类并分析其特征，旨在为轨道交通站点客流预测、规划建设、运营管理等提供一些参考。

2 进出站客流时间分布特征

轨道客流出行按照出行目的划分主要可以分为上班和上学的通勤出行、生活性弹性及公务性弹性出行、出差及旅游类出行。不同出行目的的出行时间分布特征差别较大，一般体现在通勤类出行早晚高峰系数较高，非高峰时段出行较少；弹性出行一般早高峰现象并不明显，全日相对均衡；差旅类主要集中在白天，各时段比例相对均衡。

城市轨道交通沿线的居住用地和就业岗位分布将直接影响轨道交通客流，居住用地较为集中的车站成为主要的客流发生点，而就业岗位较为集中的成为主要客流吸引点。城市轨道交通周边用地性质的差异性，各出行目的所占比例有所不同，因而不同类型站点进出站客流时间分布特征也存在较大差异。周边用地性质趋近一致的车站，其进出站客流时间分布特征具有一定相似性。

按照站点周边用地性质基本可以分一下几类：居住类、办公类、商业商务类、对外枢纽类、高校类、混合类等。选取南京地铁站点符合此类划分的典型站点进出站客流特征进行分析，根据站点客流特征差异性选取合适指标作为聚类分析变量。

2.1 不同类型站点客流时间分布特征

(1) 居住类

居住类车站早晚高峰进出站不均衡性比较突出，早高峰进站量很大，而出站量很少，晚高峰正好相反，以出站量为主，平峰时段进出量均很小。

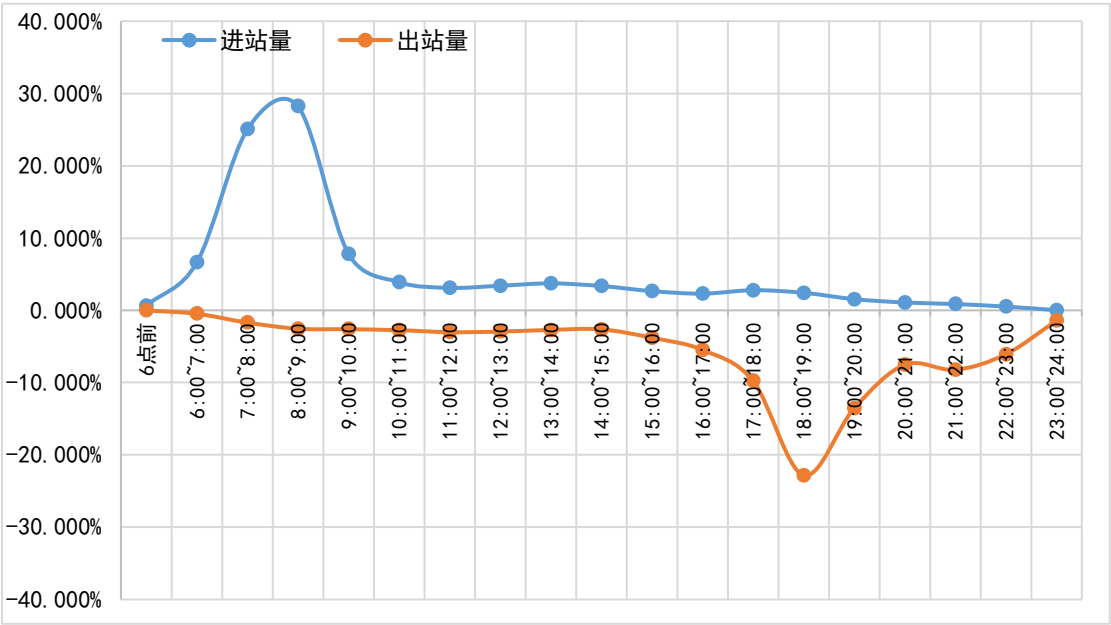


图1 居住类车站客流时间分布特征（油坊桥）

(2) 办公类

办公类车站早高峰进出站不均衡性也较为突出，早高峰以出站为主，而晚高峰主要以进站为主，平峰时段进出站量较小。

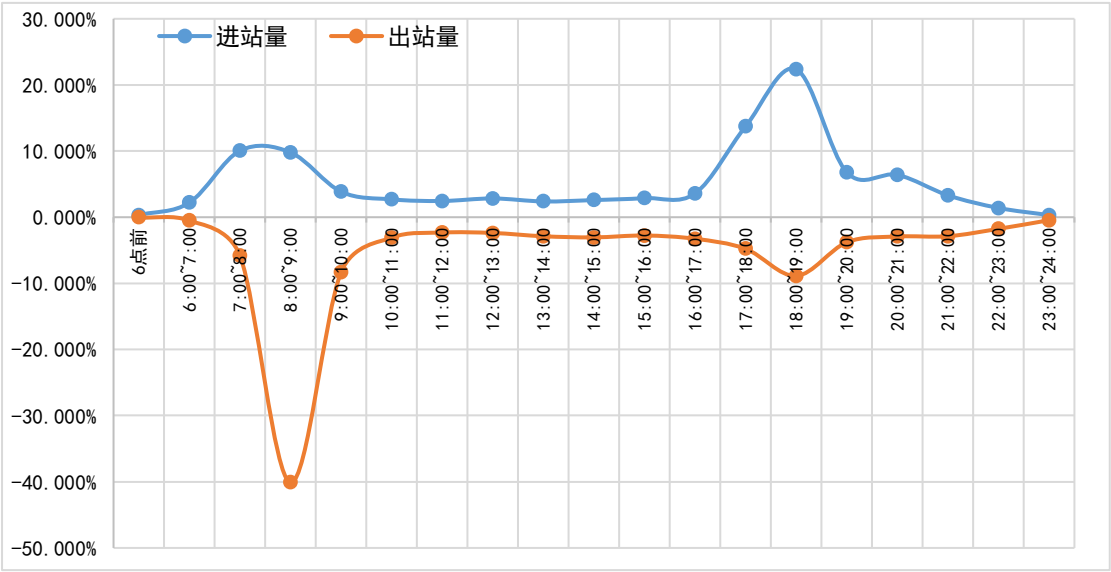


图2 办公类车站客流（软件大道）

(3) 商业类

商业类车站早高峰以出站为主，晚高峰以进站为主。相比于办公类车站，商业类车站高峰小时系数较低，但持续时间较长；平峰时段小时系数达到 5%左右，而办公类平峰小时系数只有 3%左右。

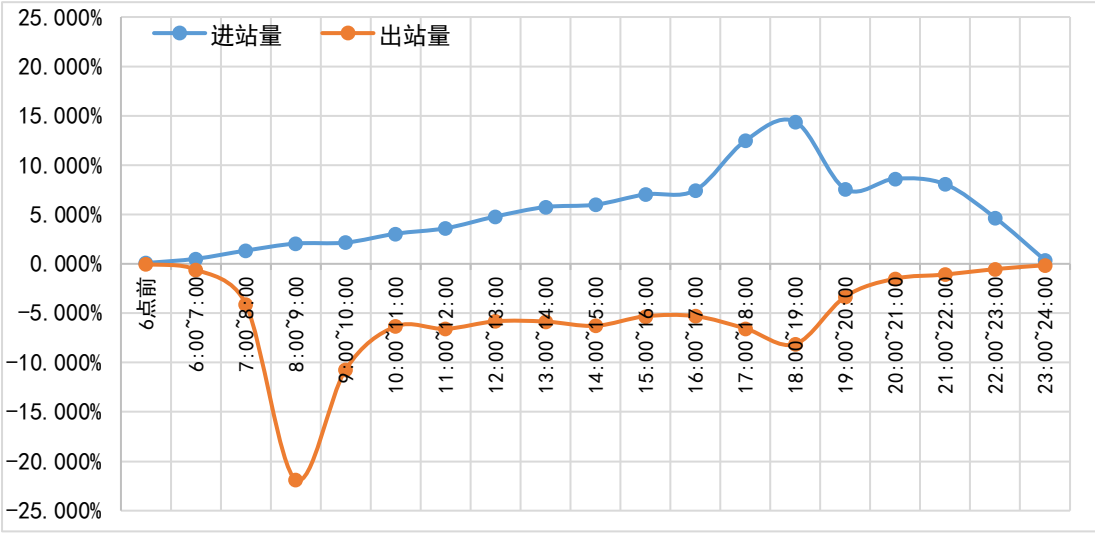


图3 商业类车站客流（新街口）

(4) 对外港站类

对外港站类车站客流早晚高峰现象不明显，全日客流进出站相对均衡，平峰小时系数均在 5%以上，且进出站量差异不大。

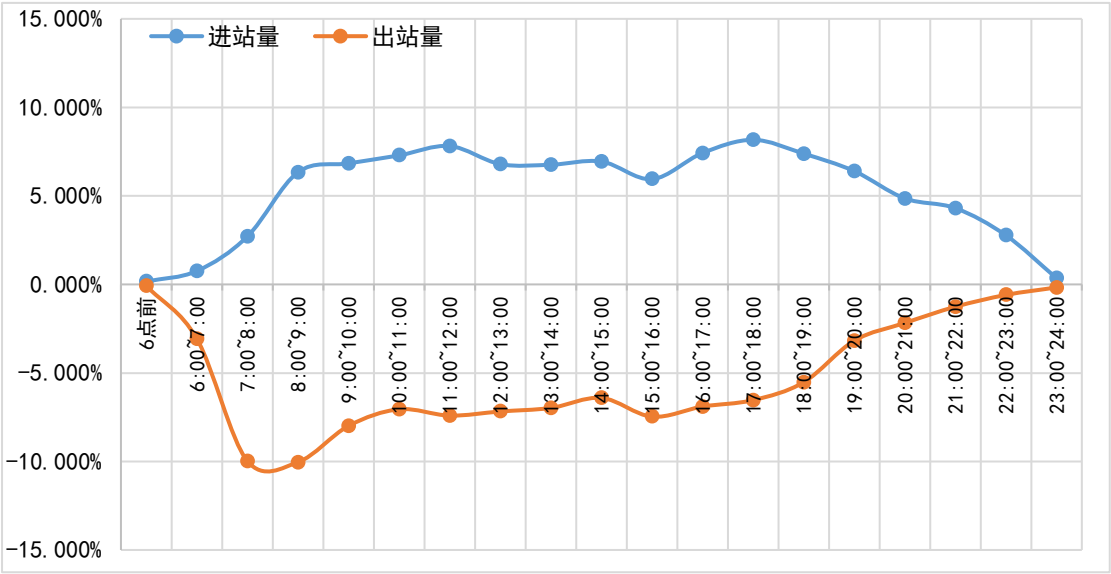


图4 对外枢纽类车站客流（南京南站）

(5) 混合类

混合类车站是指车站功能性不唯一，周围用地并不是以纯居住、纯办公、或者商业为主，而是各用地性质混合。混合类车站进出站客流量相对均衡，有一定的早晚高峰现象，高峰小时系数相对较低。

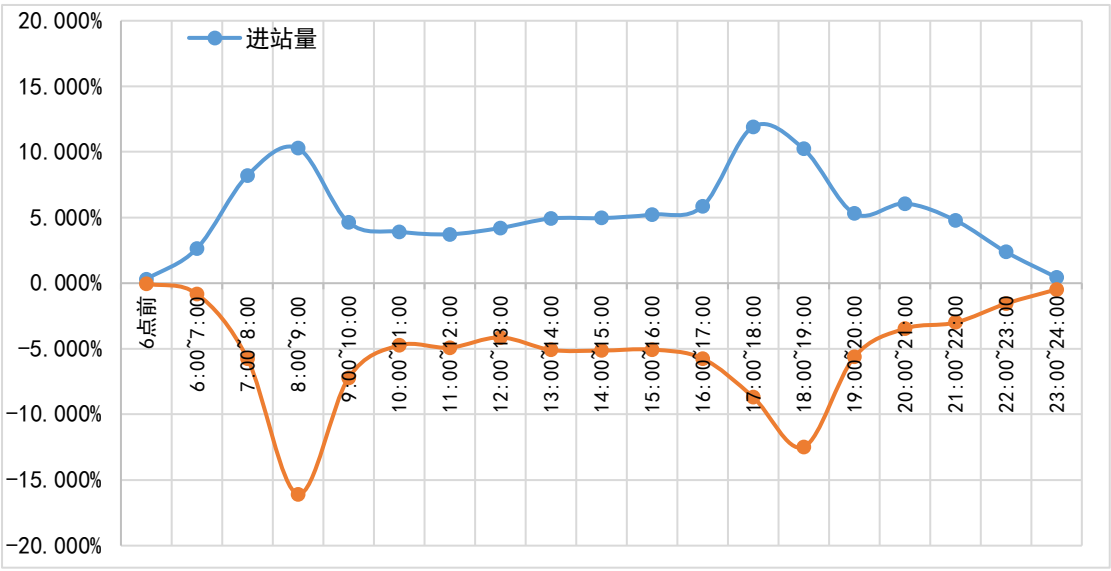


图5 混合类车站客流（三山街）

2.2 对比分析

通过以上不同类型车站客流特征分析，可以看出，典型站点的客流特征差异较大，均有一定的规律可循。

表1 车站客流特征对比

车站	早高峰	晚高峰	平峰
居住类	进站客流为主； 进站早高峰系数大于 15%；	出站客流为主； 出站晚高峰系数大于 15%；	平峰小时系数 低于 5%
办公类	出站客流为主； 出站早高峰系数大于 15%；	进站客流为主； 进站晚高峰系数大于 15%；	平峰小时系数 低于 5%
商业类	出站客流为主； 出站早高峰系数大于 15%；	晚高峰持续时间较长；以进站客流为主；早峰系数在 10%左右；	平峰小时系数 在 5%左右
对外枢纽类	高峰现象不明显	高峰现象不明显	平峰小时系数 大于 5%
混合类	进出站高峰现象明显	进出站高峰现象明显	平峰小时系数 在 5%左右

以上站点类型划分并非完全覆盖到所有站点，特别对于混合类站点，由于影响范围内的人口与岗位比例关系，可以继续划分为偏居住和偏就业类。同时，也存在着文教类、景区类、工业园区内等类型站点，这些站点客流的时间分布特征与上述站点存在一定的相似性。对于这些站点，其客流特征并不明显，很难快速将其划分到某一类中，需要一种简便的方法将其进行迅速划分类型。

2.3 特征指标选取

根据不同类型站点的早晚高峰及平峰进出站客流特征，选取早高峰、晚高峰、平峰进出站小时系数为参数，进行车站类型的聚类分析。

本文以小时为单位，统计南京地铁所有站点的某工作日进站量及出站量，早高峰进出站选取 7:00~10:00 中最大客流量；平峰客流选取 11:00~15:00 的平均进出站客流；晚高峰进出站客流选取 17:00~20:00 时段最大客流量。

$$\text{进站小时系数} = \frac{\text{该小时进站客流量}}{\text{全日进站客流总量}}$$

$$\text{出站小时系数} = \frac{\text{该小时出站客流量}}{\text{全日出站客流总量}}$$

按照公式分别计算早高峰进站小时系数、早高峰出站小时系数、平峰进站小时系数、平峰出站小时系数、晚高峰进站小时系数、晚高峰出站小时系数，作为聚类分析的指标。

3 聚类分析方法

聚类分析（cluster analysis）是一种无监督的机器学习算法，是将数据集划分成若干个子集的过程，并使同一个集合内的数据对象具有较高的相似度，而不同集合中的数据对象存在明显差异。目前，国内外对聚类研究有很大成果，包括 K-Means 算法、K-Medoids 算法、层次聚类法、CLIQUE 算法等，其中，K-Means 是在聚类分析中最常用到的一种聚类方法。

K-Means 聚类分析算法是采用离中心点距离最短的方式进行聚类，同时制定聚类中的 K 值，即聚类后的子集个数。用数学形式描述为：存在一个数据集 U，该数据集由 n 个对象构成，我们通过聚类分析法将这 n 个对象划分为 k 的子集 $D_1, D_2, D_3, \dots, D_k$ 中，假设 $i \geq 1$ ，且 $i \neq j$ ，使得 $D_i \subset U$ 且 $D_i \cap D_j = \emptyset$ 。

K-Means 聚类算法要求研究者先指定需要划分的类别个数 K，然后确定各聚类中心，再计算出各样本到聚类中心的距离，根据距离划分出 K 个类别，然后根据均值特征再产生出一个新的聚类中心，与原聚类中心进行对比，迭代至最后的聚类中心不再产生变化。其计算基本步骤为：

Step1: 从 n 个数据对象任意选择 k 个对象作为初始聚类中心；

Step2: 根据每个聚类对象的均值（中心对象），计算每个对象与这些中心对象的距离，并根据最小距离重新对相应对象进行划分；

Step3: 重新计算每个（有变化）聚类的均值（中心对象）K-Means 算法工作过程如下：首先从 n 个数据对象任意选择 k 个对象作为初始聚类中心，而对于所剩下的其它对象，则根据它们与这些聚类中心的相似度（距离），分别将它们分配给与其最相似的（聚类中心所代表的）聚类；

Step4 计算每个所获新聚类的中心,不断重复这个过程直到标准测度函数满足一定条件,如函数收敛时,算法终止,否则返回 Step2。

K-Means 算法一般采用误差平方作为标准测度函数,具体定义如下:

$$SSE = \sum_{i=1}^k \sum_{p \in D_i} |x - \bar{x}_i|^2$$

式中, SSE 表示集合中所有对象与它所在子集的中心平方误差总和; x: 对象空间中的一个点; xi: 聚类 Di 的均值。

K-Means 算法的相似度量采用欧几里得距离方法,其计算公式为:

$$d(x,y) = \sqrt{\sum_{k=1}^n (x_k - y_k)^2}$$

式中, d(x,y)表示 n 维对象中x(x₁,x₂,x₃,...,x_n)和y(y₁,y₂,y₃,...,y_n)之间的距离。

4 南京地铁车站聚类分析

4.1 数据准备

2016 年南京地铁已开通运营 6 条线路,共计 113 个车站。本次选取南京地铁 9 月中周一至周五的 6 点之前及 6 点到 24 点中各小时进出站刷卡数据,对 5 日的数据计算平均值,减少由于特殊事件引起的某日某站点进出站异常状况。根据处理后的数据分别计算得到各站点早晚高峰及平峰进出站小时系数,通过 python 编写算法过程进行 K-Means 聚类分析。

4.2 聚类分析结果

在对车站进行 K-Means 聚类分析时,需给出聚类分析类别数 K。如果 K 值给的过大,会导致聚类分析类别过多,相似类别站点会被拆分,如果 K 值给的过小,不同性质站点无法被区分开,因此 K 值的预设显得十分重要。参考手肘法对 K 从 1 至 10 之间的 SSE 进行计算,选取本次研究 K 值预设 8,得到聚类分析结果如下。

表2 聚类分析结果

聚类	站点
----	----

1	奥体中心、小行、新模范马路、集庆门大街、仙鹤门、天隆寺、星火路、梦都大街、雨花门、九龙湖、秣周东路
2	元通、张府园、新街口、珠江路、鼓楼、奥体东、上海路、大行宫、西安门、正方中路、吉印大道、长芦、化工园、鸡鸣寺、浮桥、常府街、六合开发区
3	中胜、软件大道、花神庙
4	三山街、玄武门、兴隆大街、汉中门、明故宫、百家湖、南医大江苏经贸学、绿博园、临江、小市、南京林业大学新庄、大明路、明发广场、新模范马路
5	南京站、云锦路、苜蓿园、下马坊、学则路、仙林中心、羊山公园、南大仙林校区、南京南站、中国药科大学、禄口机场、河海大学佛城西路、翠屏山、泰山新村、葛塘、凤凰山公园、金牛湖、林场、夫子庙、武定门、卡子门
6	红山动物园、迈皋桥、孝陵卫、马群、经天路、河定桥、胜太路、竹山路、龙眠大道、南京交院、江心洲、龙华路、文德路、雨山路、翔宇路南、翔宇路北、泰冯路、信息工程大学、大厂、龙池、雄州、沈桥、八百桥、诚信大道
7	安德门、中华门、雨润大街、莫愁湖、钟灵街、金马路、小龙湾、天印大道、浦口万汇城、南京工业大学、高新开发区、方州广场、东大成贤学院、上元门、五塘广场、胜太西路、天元西路
8	油坊桥、双龙大道、卸甲甸、天润城、柳洲东路、宏运大道

表3 最终聚类中心

	聚类							
	1	2	3	4	5	6	7	8
早高峰进站系数	.1391	.0673	.0862	.1115	.1194	.1969	.1774	.2867
早高峰出站系数	.2306	.2429	.4064	.1633	.0926	.0663	.1298	.0370
晚高峰进站系数	.1503	.1669	.2185	.1356	.0926	.0660	.1092	.0415
晚高峰出站系数	.1106	.0687	.0726	.1071	.1092	.1534	.1458	.2230
平峰进站系数	.0373	.0415	.0274	.0451	.0566	.0476	.0403	.0347

平峰出站系数	.0381	.0467	.0281	.0455	.0547	.0456	.0392	.0331
--------	-------	-------	-------	-------	-------	-------	-------	-------

4.3 站点分类

根据表 2 及表 3 给出的结果将南京地铁站点分为以下 8 类：

（1）居住集中型

对应第 8 聚类，该类站点周围分布着大量的居住用地，通勤客流居多，早高峰进站及晚高峰出站客流很大，进出站客流极度不均衡，潮汐现象明显，平峰时段客流很少。

（2）就业集中型

对应第 3 聚类，该类站点周围用地主要为办公用地，集中着大量的岗位。与居住集中型特征相反，早高峰出站客流及晚高峰进站客流很大，高峰潮汐现象突出。同样属于通勤客流，其平峰时段客流量很小。

（3）差旅及文教类

对应第 5 聚类，车站周围用地主要是对外港站、旅游景点及高校等，该类站点早晚高峰现象不明显，进出站高峰小时系数在 10%左右；全日客流相对均衡，平峰进出站系数接近 5.5%。

（4）居住主导型

对应第 6 聚类，该类站点周围用地中居住用地占很大部分，相比于居住集中型，其早高峰进站及晚高峰出站系数低于居住集中型，平峰小时系数在 4%~5%。

（5）就业主导型

对应第 2 聚类，该类型周围用地中以商业商务、工业为主，相比于就业集中型，早晚高峰小时系数相对较低，平峰小时系数相对较高，在 4%~5%。

（6）混合型

混合型车站周围居住人口及就业岗位相对均衡，其进出站早晚高峰系数均高于 10%，但小于 20%。根据人口及岗位的权衡，细分可以分为混合偏居住型（第 7 聚类）、混合偏就业

型（第 1 聚类），根据平峰客流大小，可以划分出混合偏弹性（第 4 聚类），该类平峰小时系数在 0.45 左右，而偏居住和偏就业平峰小时系数低于 4%。

5 分析结论

本文利用南京地铁刷卡进出站数据，分析了典型站点的进出站时间分布特征，选取了表征站点客流特征的特征指标，通过 K-Means++ 聚类分析对南京地铁 113 个站点进行类型划分，将站点分为居住集中类、就业集中型、差旅及文教类、居住主导型、就业主导型以及混合型。

研究表明，站点客流时间分布特征与其周围用地性质紧密相关，用地性质决定了居民出行目的及出行特征。通勤客流往往聚集在早晚高峰，弹性客流分散在全日各时段。进站与出站客流在时间上的分布能较好地表现其这一特点，采取早晚高峰及平峰进出站小时系数作为指标，通过聚类分析能迅速地将地铁站点进行分类。

根据客流特征将城市轨道交通站点进行分类，有助于地铁运营公司对车站实行差异化管理，制定灵活的运营组织模式。