

文章编号: 2095-2163(2020)01-0071-06

中图分类号: TP183

文献标志码: A

# 基于图论的上海轨道交通站点客流等级预测研究

于莉, 刘志钢, 石志峰, 陈颖雪, 张香明

(上海工程技术大学 城市轨道交通学院, 上海 201620)

**摘要:**科学的站点分级利于新线车站参考同等级现有车站,快速、合理确定设施设备、人员岗位规模及规划客流组织方案。图论法可准确高效地挖掘线网潜在的规律和特点。引入距城市中心最短距离作为周边土地利用的替代指标,并以度值、接近中心性值及节点介数值作为等级评价指标来描述站点重要度,通过K-means聚类将线网站点分为10级。上海轨道交通车站按重要度等级分类来进行车站客流水平预测,等级误差在2以内的车站数目达264个,占总数的86.8%。

**关键词:**城市轨道交通;图论;站点分级;K-means聚类;客流等级预测

## Research on passenger flow grade prediction of urban rail transit station based on graph theory

YU Li, LIU Zhigang, SHI Zhifeng, CHEN Yingxue, ZHANG Xiangming

(School of Urban Rail Transportation, Shanghai University of Engineering Science, Shanghai 201620, China)

**[Abstract]** Scientific station classification is conducive to the station of new line to refer to the existing stations of the same grade, and to quickly and reasonably determine the facilities, equipment, staff size and passenger flow organization plan. The graph theory method can accurately and efficiently excavate the potential rules and characteristics of line network. The shortest distance from the center of the city is introduced as a substitute index for the surrounding land use, and degree value, near-centrality value and node intermediate value are used as grade evaluation indexes to describe the importance of the site, and K-means clustering is used to classify the site points into 10 levels. The passenger flow level of Shanghai rail transit stations is predicted according to the importance level classification. The number of stations with the grade error within 2 reaches 264, accounting for 86.8% of the total number.

**[Key words]** urban rail transit; graph theory; site classification; K-means clustering; passenger flow level prediction

### 0 引言

随着中国大中型城市轨道交通建设快速发展,轨道交通线网的拓扑结构越来越复杂。对轨道交通车站的特性进行分类,可以有目的地了解不同种类站点之间的差异,科学合理的站点分级体系能够为地铁车站人员配置、资源分配及管理提供决策依据。应用图论的方法可以将城市轨道交通的线网用数学模型方法进行描述,不仅可以区别于传统的以车站性质和周边用地性质来进行站点等级划分,又可以在复杂的线网中准确地提取线网自身的规律和特点。

由于城市轨道交通线网的复杂性,国内外学者主要研究图论中的点、线以及整体性指标对线网结构改变后的评价方法及结果。慈立坤等人<sup>[1]</sup>使用图论的方法研究上海轨道交通运营网络的特性并建

立了线网可靠性评价模型,论证了当网络节点破坏不同出行路径影响大小不同。四兵锋等人<sup>[2]</sup>从图论的角度出发,构建了多层次的交通系统网络结构并进行描述,获取乘客对于网络进行的差异性选择并建立以乘客出行需求为基础的广义费用模型和路阻函数。在站点分级的研究中,WEI等人<sup>[3]</sup>选取客流量、位置和客运服务指标对车站进行打分,并提出客流量等级公式;余丽洁等人<sup>[4]</sup>在确定影响轨交站点属性因素参数的基础上,以西安地铁2号线为例,采用多种聚类算法对站点分类效果进行评述。陈扶崑等人<sup>[5]</sup>选取行车组织方案、车站客运服务、客流量和地理位置四个指标,应用定量分析方法,建立城市轨道交通车站等级划分体系标准。

本文从图论的角度出发,针对新线开通初期,引入站点客流等级预测方法,通过车站图论等级与客

**作者简介:**于莉(1994-),女,硕士研究生,主要研究方向:城市轨道交通运营与管理;刘志钢(1974-),男,博士,教授,上海工程技术大学城市轨道交通学院院长,主要研究方向:城市轨道交通运营管理优化及安全技术;石志峰(1994-),男,硕士研究生,主要研究方向:城市轨道交通运营与管理;陈颖雪(1983-),女,博士,讲师,主要研究方向:城市轨道交通运营管理优化;张香明(1994-),女,硕士研究生,主要研究方向:城市轨道交通运营与管理。

收稿日期: 2019-09-29

流等级的相关性分析,两者成强正相关,建立相应的预测标准。

## 1 上海轨道交通线网模型建立

图论是专门用于研究图的一个数学分支,提供了描述网络的语言和研究的平台,通过对图的研究可以得到具体的实际网络的相关拓扑性质<sup>[6]</sup>。城市轨道交通线网是由各个车站和各条线路组成的拓扑网络,可以抽象为由点、线组成的基本图形,车站作为一个点,两站点间的线路作为图形中的边。用图论的方法表示城市轨道交通线网可以清晰地描绘出线网内部的各种关联关系,通过计算机准确地刻画出线网模型并完成进一步的定量处理。城市轨道交通线网图可表示为:

$$G = (V(G), E(G), \varphi_c). \quad (1)$$

其中,  $G$  表示城市轨道交通线网图;  $V(G)$  表示车站集;  $E(G)$  表示两站间线路集;  $\varphi_c$  表示关联函数。  $E(G)$  中的每条边都有  $V(G)$  中的一对点与之相对应。在本文中将轨道交通线网视为无向图,可使图  $G$  的每条边对应于  $G$  无序顶点对。

由于 Space L 方法构建的模型简单却能够很好

地重现线网的实际拓扑结构,因此本文采用 Space L 方法构建线网模型,如图 1 所示。

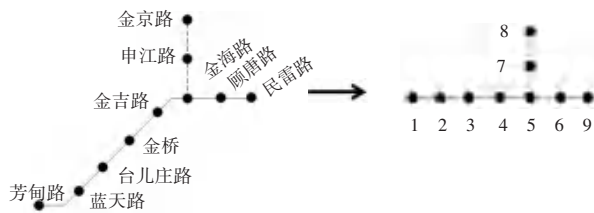


图 1 Space L 模型实例

Fig. 1 Space L model example

为了获得图论角度下上海轨道交通线网的一般拓扑结构规律及特征,在建模时对线网进行了简化:首先,不考虑实际运营中站点的区位以及距离等信息,假定各站之间的距离为 1;其次共线区段合并为一条线路(主要为 3、4 号线的宝山路站-宜山路站及 2、10 号线的虹桥火车站至虹桥 2 号航站楼);最后,地铁线网拓扑结构图为无向无权图(不区分上下行)。

应用 space-L 方法建立上海轨道交通线网拓扑模型,如图 2 所示。

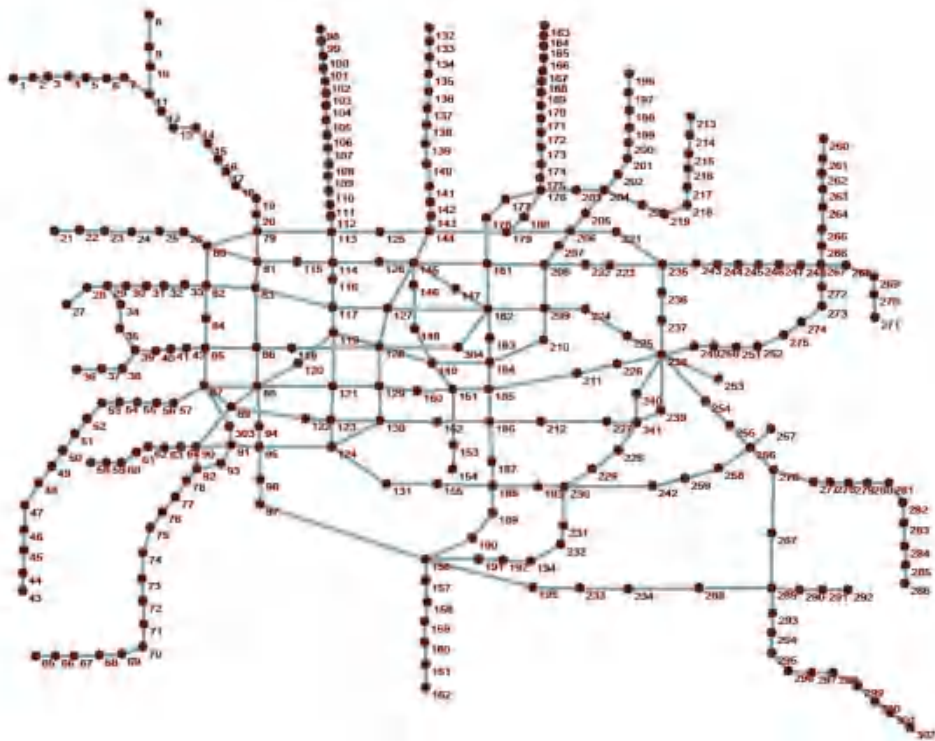


图 2 上海轨道交通线网拓扑结构图

Fig. 2 Topology of Shanghai rail transit network

## 2 轨道交通站点重要度评价指标确定

目前,复杂网络中最主要的统计指标有:度值、接近中心性以及介数值<sup>[7]</sup>等。这些指标从不同的角度对网络结构进行了描述,并且较为全面地反映

线网内部的特性。

对上海轨道交通线网中的站点评价指标进行计算,由于大型城市的中心站点本身影响力较大,会对距离较近的周边站点存在辐射影响,造成同类型站

点间的重要度差异。因此,综合地理位置和用地信息等选取了上海市 4 个城市中心车站,包括位于城市主中心的人民广场站和 3 个位于城市副中心的世纪大道站、中山公园站以及徐家汇站。在不考虑换乘

便捷性的情况下,利用 Dijkstra 算法将各个车站至城市中心站的最短距离求解出来,并将其作为一项重要度指标放入模型中,部分站点指标数值见表 1。

表 1 站点指标数值表(部分)

Tab. 1 Number table of some site indicators( part)

车站	度	接近中心	介数	人民广场	世纪大道	徐家汇	中山公园
人民广场	6	0.105 1	0.147 6	0	4	5	4
南京西路	4	0.104 0	0.135 0	1	5	4	3
虹桥火车站	2	0.125 4	0.203 0	6	12	5	2
世纪大道	8	0.093 4	0.227 1	4	0	8	8
上海火车站	4	0.101 6	0.219 7	3	7	8	5
徐家汇	6	0.165 1	0.139 9	5	8	0	3
中山公园	4	0.154 1	0.048 4	4	8	3	0
打浦桥	2	0.082 0	0.034 6	5	5	3	6
中山北路	2	0.073 8	0.064 0	4	8	9	8
徐泾东	1	0.064 4	0.006 6	11	15	10	7
浦东国际机场	1	0.061 0	0.045 3	8	4	12	22
佘山	2	0.079 2	0.057 8	15	18	10	12
迪士尼	1	0.057 7	0.038 9	17	8	13	16

### 3 基于 K-means 的车站重要度等级聚类及特征分析

由于地铁车站分布于全网,其地理区位不同导致服务客流量差异较大,聚类变量的选取需结合图论中的指标来反映节点空间结构特性,站点距离城市中心站点及城市副中心站点的最短路距离替代土地利用指标,反映站点在实际线网拓扑结构中的重要度。本文引入 K-means 聚类算法模型<sup>[8]</sup>,建立综合评价指标的車站等级划分方法。

度等级更加符合实际的情况,需对部分首末站、大型对外交通枢纽站点指标参数进行适当的提高,以保证站点分级的合理性和有效性。考虑可以更加明确区分各类型车站的特征,选取参数  $k$  为 10 进行聚类,经过多次迭代得到最优的車站等级聚类结果。代表車站聚类结果见表 2。

对 304 个車站的聚类结果进行汇总,根据各类型車站的特征得到轨道交通站点分类结果共 10 类,具体内容见表 3。

上海轨道交通站点数量较多,为了使站点重要

表 2 上海轨道交通代表車站聚类结果(部分)

Tab.2 Clustering results of Shanghai rail transit stations ( part)

编号	車站	等级	编号	車站	等级	编号	車站	等级
182	人民广场	10	150	打浦桥	5	286	浦东国际机场	3
127	南京西路	9	147	新闸路	5	14	桃浦新村	2
28	虹桥火车站	9	33	娄山关路	5	131	后滩	2
238	世纪大道	8	29	虹桥 2 号航站楼	4	289	罗山路	2
145	汉中路	8	142	延长路	4	293	周浦东	2
144	上海火车站	8	277	金科路	4	49	佘山	2
128	陕西南路	7	143	中山北路	4	17	李子园	2
149	新天地	7	148	淮海中路	4	138	共康路	2
185	陆家浜路	7	276	张江高科	4	13	南翔	1
87	宜山路	7	130	大木桥路	3	12	马陆	1
176	虹口足球场	7	18	上海西站	3	284	远东大道	1
88	徐家汇	6	141	上海马戏城	3	290	秀沿路	1
82	中山公园	6	20	枫桥路	3	282	川沙	1
129	嘉善路	6	27	徐泾东	3	292	迪士尼	1

表3 上海轨道交通车站重要度等级

Tab. 3 Importance grade of Shanghai rail transit station

分类	站点特征	客流特征	代表站点
第一类	各线的郊区站	该类站点皆为非换乘站且位置偏僻,重要度最低	嘉定西站、临港大道站、醉白池站等
第二类	各线的郊区中部站点	该类站点同样为非换乘站,拥有较小的度,因此重要度相对较低	祁连山路、佘山、颛桥等
第三类	各线终点站及郊区段接入市区段的前段站点	终点站比一般普通站点的吸引范围更广,且为多条地面公交线路的枢纽站,因此虽然有着较小的图论指标,但是重要度却较大	美兰湖、徐泾东、行知路站等
第四类	各线郊区段接入市区段的过渡站点	该类站点距离城市中心相对较近,受市区段的影响,因此重要度略微有所提高	如真如站、大渡河路站、石龙路站等
第五类~第七类	各线市区段的站点和	该类站点由于换乘情况的不同及距离城市中心的远近,导致各站点介数及接近中心性指标有所不同,因此对于重要度也同样有所区分	如宜山路站、打浦桥站、新天地站等
第八类~第十类	各线位于市中心附近的站点和大型交通枢纽站	该类站点本身即为多线换乘站,拥有较高的度,且位于网络中心,自身的介数及接近中心性指标也同样较大,因此站点也就最为重要	人民广场站、南京东路站、世纪大道站等

#### 4 站点客流等级预测实例研究

为了将复杂的预测方法进行简化,不再仅仅关注站点客流的具体数值,而是研究站点相关客流等级。考虑利用站点在图论中的重要度来推算相应的站点客流等级,从而在新线开通初期可以对各个站点客流情况有个较为简便准确的获取方式。对此拟展开研究阐述如下。

##### 4.1 站点进出站客流分布及分类

由于是根据所建的站点图论重要度等级来进行预测,因此得到的预测结果也应是相应的客流等级。针对各个站点的实际客流量进行站点客流等级的划分,绘制各个站点的历史某天的进站客流量数值图,如图3所示,从实际数据量上可知上海地铁站点的进出站量最小值约在1 400 人/天,而最大值在11 万人/天,进出站量在3 万人/天以下的站点数占了总数的约83%,由此可见客流数据的范围较大且各数据级的分布十分不均衡。

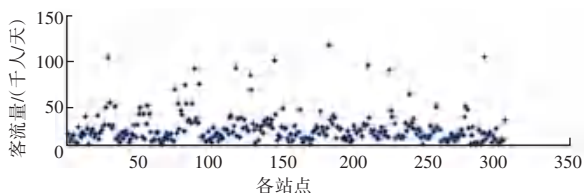


图3 线网各站点进站客流量数值分布图

Fig. 3 Numerical distribution of inbound passenger flow at each station

为了保证客流等级的分类能够较为准确地反映出站点的实际客流量区间,所以在进行客流等级分

类时舍弃均匀分配的方法。针对前六等级的站点以较小的数据间隔5 000 人次进行划分,而后四级则以较大的数据间隔进行划分。由于在该种划分方式下的各客流等级实际间距较大,所以能够在一定范围内忽略站点客流的自然增长情况。具体分类标准见表4。

表4 客流等级具体分类标准

Tab. 4 Specific classification standard of passenger flow grade

客流等级	客流区间/人	客流等级	客流区间/人
1	≤10 000	6	30 000~35 000
2	10 000~15 000	7	35 000~60 000
3	15 000~20 000	8	60 000~90 000
4	20 000~25 000	9	90 000~100 000
5	25 000~30 000	10	≥100 000

利用实际进站客流数据、分类好的各个站点的客流等级数据分别与图论重要度等级进行相关性分析,可以进一步证实该预测方法的准确性。

##### 4.2 相关性分析

结合本次数据的特征,选用 Spearman 相关系数进行分析。根据各个站点的实际进出站量以及上文所构筑的站点图论重要度等级指标来对数据进行初步的相关性分析,可知其相关系数均大于0.7,为强正相关关系。

同时利用 SPSS 对2种划分方法下的分类结果进行进一步的 Spearman 相关性分析,研究可得在图论角度及客流角度下的部分站点的划分结果比较见表5。

表 5 图论角度及客流角度下的站点分类结果表(部分)

Tab. 5 Partial station classification results from the perspective of graph theory and passenger flow

车站	图论	客流	车站	图论	客流	车站	图论	客流
嘉定北	3	4	虹桥路	4	6	水产路	1	1
嘉定西	1	2	交通大学	6	5	淞滨路	1	1
白银路	1	2	宜山路	7	7	张华浜	1	1
嘉定新城	3	3	徐家汇	9	8	淞发路	1	1
马陆	1	2	上海体育馆	6	5	长江南路	2	3
上海西站	3	3	龙华	3	2	西藏北路	5	4
真如	3	1	云锦路	2	1	中兴路	3	3
枫桥路	3	2	龙耀路	2	1	宝山路	3	2
金运路	3	6	美兰湖	3	3	东宝兴路	3	2

经计算得出相关系数为 0.733, 同样为强正相关, 验证说明, 随着图论等级的上升, 客流等级也会相应的提高, 因此能够根据新线开通前后站点重要度的变化来反映实际客流等级的变化, 从而利用图论中的分类等级来预测站点相应的客流等级。

### 4.3 新线开通后站点客流等级预测

9 号线延伸段开通后, 线网的拓扑结构发生了

一系列的变化, 同时各个站点的图论指标也会发生相应变化, 利用上文的程序能够轻松获取新线站点及随之发生变化的既有线站点的图论重要度。根据站点图论重要度等级对新线开通初期的站点客流等级进行预测, 由此求出预测所得的部分既有站点客流等级见表 6。

表 6 既有站点部分车站客流等级预测表

Tab. 6 Forecast of passenger flow grade in some old stations

车站	预测	实际	车站	预测	实际	车站	预测	实际
嘉定北	3	3	交通大学	5	7	惠南	3	2
嘉定西	2	1	宜山路	7	7	华夏中路	1	2
白银路	1	1	徐家汇	8	8	御桥	2	2
嘉定新城	3	3	上海体育馆	5	5	罗山路	1	2
马陆	2	1	漕宝路	6	5	五洲大道	1	1
南翔	6	1	龙漕路	5	5	东靖路	1	1
桃浦新村	3	2	上海南站	8	8	巨峰路	3	2
武威路	1	1	石龙路	3	3	杨高北路	2	1
祁连山路	2	2	上海游泳馆	3	4	金京路	1	1
李子园	3	3	龙华	2	4	申江路	1	1
上海西站	3	3	云锦路	1	4	金海路	1	2
真如	2	3	龙耀路	1	3	五莲路	2	1
枫桥路	2	3	美兰湖	3	3	博兴路	2	1
金运路	6	3	罗南新村	2	1	金桥路	3	2

为了验证预测结果的优劣, 绘制出既有站点的预测等级与实际客流等级间的差值分布, 如图 4 所示。

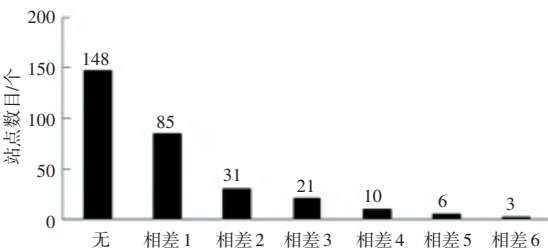


图 4 站点两类分类下的差值分布

Fig. 4 Difference distribution under two categories of stations

由图 4 可以看出, 预测的等级误差在 2 以内的车站数目达 264 个, 占总数的 86.8%, 可见该预测方法能够有效地反映站点的实际客流情况, 具有很强的可用性。

### 5 结束语

从图论的方法出发选择 Space-L 方法对上海轨道交通线网进行拓扑结构建模, 选取衡量站点重要度的 3 个指标, 并从自身、空间土地利用和流量的三个角度评价车站的重要性, 同时引入站点至 4 个城市中心的最短路径指标, 全面地对车站指标进行评估。基于 K-means 聚类算法, 利用 Python 编程求得

(下转第 79 页)