

基于系统聚类法的城市轨道交通 建设能力排序

牛圣宽¹ 伊盼盼²

(1. 长江水利委员会长江勘测规划设计研究院, 430010 武汉;

2. 中国科学院武汉岩土力学研究所, 430074 武汉 // 第一作者, 助理工程师)

摘要 城市轨道交通项目具有投资大、成本高、投资回收期长等特点, 为了防止各城市盲目修建城市轨道交通项目, 国家提出坚持有序发展的方针。就城市的各项指标, 通过应用主成分分析法对指标体系进行简化, 应用系统聚类分析法对各城市进行聚类, 得到各城市轨道交通建设能力的排序, 为我国各城市有序地修建轨道交通提供参考。

关键词 城市轨道交通; 建设能力; 系统聚类分析

中图分类号 O 223 TU 984 191

Sequence of Urban Rail Transit Building Capacity on Hierarchical Cluster Analysis

Niu Shengkuang Yi Panpan

Abstract There are some characteristics of the rail transit items such as large investment, high costs, long capital return period and so on. In order to prevent the blind construction of rail transit, the Central Government has promulgated the guideline for the orderly development. This paper analyzes the various urban indicators and presents a method of principal component analysis to simplify the indicator system. With the application of the hierarchical cluster analysis, the rail transportation capacity is scheduled in regular sequence which will ensure the orderly construction of urban rail transit in China.

Key words urban rail transit building capacity; hierarchical cluster analysis

First author's address Changjiang Water Resources Commission, Changjiang Institute of Survey, Planning, Design and Research, 430010 Wuhan, China

0 引言

随着我国经济的快速发展, 目前, 我国人口在100万以上的大城市已发展到30多个。城市人口急剧增加, 大量流动人口涌进城市, 使城市交通面临着严峻的形势。当前, 全国大中城市普遍存在着道

路拥堵的现象。为了缓解愈来愈严峻的交通压力, 部分大城市相继修建了一批轨道交通项目, 使城市交通状况有了明显改善。因此, 越来越多的城市也提出了修建大中运量的城市轨道交通的计划。虽然城市轨道交通有很多优点, 但其投资大、成本高、投资回收期长, 政府需要每年投入很大的财力来维持其运营。目前我国的经济实力, 还不容许所有的大中城市都发展轨道交通。因此, 城市轨道交通的兴建应坚持量力而行、有序发展的方针。

城市轨道交通的修建与否同城市的发展水平密切相关。城市的发展水平可用市辖区GDP、市辖区总人口、财政可支配收入、市辖区人口密度、客运总量等指标来描述。本文从影响我国城市轨道交通建设的各项指标入手, 应用主成分分析法对指标体系进行简化, 并应用系统聚类分析法对我国大城市进行分类, 间接对各城市建设轨道交通的必要性进行排序, 以供各城市有序地修建城市轨道交通作参考。

1 用主成分分析法简化城市发展水平指标体系

1.1 主成分分析法的原理

主成分分析法是多元统计分析的重要组成部分, 是美国心理学家、统计学家 Charles Spearman 于1904年提出的。

在研究问题的过程中, 常常会遇到很多变量。这些变量之间又可能有一定的相关性, 因此希望从中综合出一些主要的指标。这些指标所包含的信息量很多, 使分析复杂问题时能抓住主要矛盾, 实现问题的简化。主成分分析法就是抓住有代表性的少数几个指标去进行评价, 达到了指标筛选的科学化。

设有 m 个初始的评价指标, n 个评价的数据样本, x_{ij} 表示第 i 个样本第 j 个评价指标的值 ($1 \leq i \leq n$)

$n \leq i \leq m$), 得到的原始资料矩阵为:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}$$

为了避免量纲的不同而带来的一些不合理的影响, 比较合理的做法就是对原始数据进行标准化处理。为使每个指标的平均数为 0 方差为 1, 数据标准化公式为:

$$\bar{x}_{ij} = \frac{x_{ij} - \bar{X}_j}{S_j} \quad (1 \leq i \leq n, 1 \leq j \leq m) \quad (1)$$

式中:

$$\bar{X}_j \text{——每列的均值, } \bar{X}_j = \frac{1}{n} \sum_{k=1}^n x_{kj} \quad (1 \leq j \leq m);$$

$$S_j \text{——每列的方差, } S_j = \frac{1}{n-1} \sum_{k=1}^n (x_{kj} - \bar{X}_j)^2 \quad (1 \leq j \leq m).$$

设 c_{ij} 表示 X^* 第 i 列和第 j 列的协方差,

$$c_{ij} = \frac{1}{n-1} \sum_{k=1}^n (\bar{x}_{ki} - \bar{X}_i^*)(\bar{x}_{kj} - \bar{X}_j^*) \quad (2)$$

可求出经过标准化处理后的数据矩阵 X^* 的协方差矩阵 C

求出 X^* 的相关系数矩阵 R 其中:

$$r_{ij} = c_{ij} / \sqrt{S_i^* S_j^*} \quad (1 \leq i \leq m, 1 \leq j \leq m)$$

$$R = \begin{pmatrix} 1 & 0.6386 & 0.9628 & 0.8079 & 0.04339 & 0.4923 & 0.8594 & 0.1153 & 0.7861 & 0.9289 \\ & 1 & 0.7135 & 0.9241 & 0.0910 & 0.8430 & 0.6196 & 0.3615 & 0.6456 & 0.8105 \\ & & 1 & 0.8738 & 0.1348 & 0.5152 & 0.7888 & 0.1179 & 0.7747 & 0.9701 \\ & & & 1 & 0.1826 & 0.7328 & 0.7040 & 0.1879 & 0.7348 & 0.9359 \\ & & & & 1 & -0.3716 & -0.1038 & 0.1208 & -0.1457 & 0.1076 \\ & & & & & 1 & 0.6148 & 0.3751 & 0.7186 & 0.6345 \\ & & & & & & 1 & 0.4412 & 0.8845 & 0.8194 \\ & & & & & & & 1 & 0.4444 & 0.1954 \\ & & & & & & & & 1 & 0.8054 \\ & & & & & & & & & 1 \end{pmatrix}$$

计算相关系数矩阵的特征值和方差贡献率 见表 1。

由表 1 可知, 主成分个数为 4 时累计贡献率达到 96.999%, 故选择前 4 个主成分, 不须计算其他的成分。这 4 个主成分基本上保留了原来 10 个指标 ($x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}$) 的信息。这 4 个主成分的特征向量分别为:

$$f_1 = (0.3531, 0.3394, 0.3598, 0.3655, 0.0035,$$

求解行列式 $|R - \lambda I| = 0$ 的 m 个根, 即相关系数矩阵 R 的特征值 $\lambda_1, \lambda_2, \dots, \lambda_m$ 。为了能用尽量少的主成分个数反映原始变量间的变异信息, 将特征值按从大到小顺序排列, 计算它们的累积贡献率, 即为 $\sum_{i=1}^k \lambda_i / \sum_{i=1}^m \lambda_i (1 \leq k \leq m)$ 。

选取累积贡献率在 90% 以上的前几个特征值, 其对应的特征向量为 $f = (f_1, f_2, \dots, f_n) (1 \leq i \leq m)$ 。则可以求出相应的主成分 $P_i = \sum_{k=1}^n f_{ki} x_k$ 然后进行分析。

1.2 城市发展水平评价指标体系的简化

为了较准确地描述城市的经济发展和规模, 本文选取了 10 项主要指标: x_1 ——市辖区 GDP, x_2 ——市辖区总人口, x_3 ——财政可支配收入, x_4 ——市辖区非农业人口, x_5 ——市辖区人口密度, x_6 ——市辖区行政区域土地面积, x_7 ——建成区面积, x_8 ——客运总量, x_9 ——常规道路公交客运总量, x_{10} ——固定资产投资总额。

本文选取我国上海、北京、广州、苏州、深圳、天津、重庆、杭州、成都、青岛、大连、武汉、南京、沈阳、长春、西安等 16 个大城市进行分析。

应用式 (1) 对各个城市的原始数据矩阵进行标准化处理, 得到标准化数据矩阵。然后根据式 (2) 计算指标间的相关系数矩阵 R 得到:

$$0.3025, 0.3480, 0.1398, 0.3504, 0.3770)^T;$$

$$f_2 = (0.1814, -0.0185, 0.2449, 0.16256,$$

$$0.7355, -0.4338, -0.1076, -0.2790,$$

$$-0.1823, 0.1689)^T;$$

$$f_3 = (-0.2248, 0.1463, -0.1784, -0.0238,$$

$$0.4805, 0.0213, 0.0284, 0.8065,$$

$$0.0368, -0.1086)^T;$$

$$f = (-0.2869 \ 0.5281 \ -0.1290 \ 0.3320 \\ 0.0858 \ 0.4296 \ -0.4283 \ -0.2291 \\ -0.2899 \ -0.0013)^T$$

每个主成分计算公式为:

$$F_i = R \times f \quad (1 \leq i \leq 4) \quad (3)$$

应用式 (3) 可以得到主成分组成的矩阵 $F = (F_1, F_2, F_3, F_4)$

表 1 相关系数矩阵的特征值及其贡献率

序号	特征值	贡献率 /%	累计贡献率 /%
1	6.5190	65.190	65.190
2	1.3431	13.431	78.621
3	1.0931	10.931	89.552
4	0.7447	7.447	96.999
5	0.1410	1.410	98.409
6	0.0934	0.934	99.343
7	0.0330	0.330	99.673
8	0.0156	0.156	99.830
9	0.0109	0.109	99.938
10	0.0062	0.062	100.000

2 用系统聚类法确定各城市轨道交通建设的次序

2.1 系统聚类的原理

系统聚类法的基本思路是: 先将需要聚类的样本或者指标各自看成一类, 然后确定类与类之间的相似性统计量, 并选择最近的 2 类或若干类合成一个新类; 计算新类与其他类之间的相似性统计量, 再选择最接近的两类或若干类合成一个新类, 直到所有的样品或指标都合成一类为止。这一归类过程可以用一张聚类图形象地表示出来。由聚类图可以明显地看出分类过程。这一聚类过程就提供了表征城市发展水平相似程度的一个指标。以此指标就可以对城市建设轨道交通的能力进行排序。本文聚类计算中采用欧式距离作为统计距离, 聚类方法采用类平均法。

2.2 城市的分类结果

对主成分矩阵 F 进行聚类计算, 得到聚类图 (如图 1 所示)。

我国城市综合实力最强的是北京和上海。从这两个城市的城市轨道交通运营来看, 确实缓解了城

市交通压力, 对促进城市的经济和社会发展起了重要作用。事实证明了发展轨道交通的必要性。任何城市的市情与北京, 上海的市情越接近, 就说明发展城市轨道交通的必要性越大。由图 1 系统聚类树图, 可以得到各个城市修建城市轨道交通的顺序为: 北京和上海—重庆—广州和深圳—成都—武汉和天津—青岛—南京—苏州—长春—杭州和大连—沈阳和西安。这个结果和目前正在修建轨道交通的顺序基本上一致。由此也可证明, 本文的分析方法可应用于实践, 以指导我国各城市轨道交通建设的有序发展。

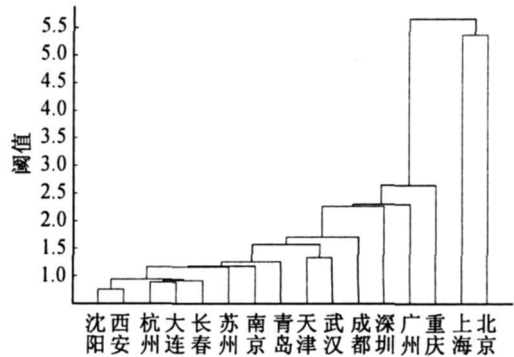


图 1 系统聚类树图

3 结语

本文应用系统聚类法对我国部分城市修建轨道交通的必要性进行排序。排序结果和实际情况基本一致, 说明了这种方法可以作为对各城市修建轨道交通必要性进行排序的一种有效方法, 可为我国各城市有序地修建城市轨道交通提供参考。

参考文献

- [1] 张建平. 基于主成分分析的区域经济发展水平的综合评价 [J]. 农业与技术, 2007(3): 125
- [2] 卢晓春, 李明惠, 吴毅洲, 等. 用主成分分析法优化高速公路顾客评价指标体系 [J]. 交通标准化, 2004 2(3): 93.
- [3] 李伟, 王黎勇, 杨瑞贞. 运用系统聚类法综合评价农村社区卫生服务中心功能 [J]. 中国医院统计, 2006 13(3): 201
- [4] 李艳双, 曾珍香, 张闯, 等. 主成分分析法在多指标综合评价方法中的应用 [J]. 河北工业大学学报, 1999 28(1): 95.
- [5] 李平. 主成分分析法在企业综合经济效益评价中的应用 [J]. 南京邮电学院学报, 1992 12(4): 127

(收稿日期: 2008-03-16)