

中国交通运输业发展区域分布特征与聚类分析

雷思途, 张矢宇, 韦文景, 韦金汎

(武汉理工大学 交通学院, 湖北 武汉 430063)

摘要: 由于自然地理条件差异及地区经济发展状况不同, 中国各地区的交通运输业发展存在很大差异, 明确中国交通运输业区域分布特征及发展情况对制定相应地区交通发展战略至关重要。文中选取社会经济、交通运输产运量及交通运输生产要素建立交通运输业发展分析指标体系, 根据各指标变异系数的变化趋势分析中国交通运输业发展区域分布差异化的动态特征, 并对 31 个地区进行聚类分析, 将 31 个地区分为交通发达、较发达和欠发达三类区域, 同时对各类区域的交通运输业发展提出相关建议。

关键词: 交通运输; 区域分布特征; 聚类分析; 相似性; 差异性

中图分类号: U491

文献标志码: A

文章编号: 1671-2668(2019)04-0019-05

中国幅员辽阔, 区域间自然地理条件差异大, 人口分布、经济发展不平衡, 区域间交通运输业呈现明显差异。众多学者对交通运输业发展的区域特征与聚类分析进行了研究, 并分析了区域特征与交通运输业发展之间的联系。如侯平路、张伟、姚新胜等运用聚类分析探讨了全国省际综合发展水平的区域相似性和差异性; 谢霖检运用聚类分析对甘肃省 15 个地州市进行分类, 确定了重点枢纽城市带动次枢纽城市发展的思想; 周志龙选取区域交通与区域经济指标进行聚类分析, 利用推拉效用模型, 分析了中国各区域综合交通运输结构的变化规律。该文利用变异系数把握交通运输业发展区域差异的新动

态, 并根据指标对聚类分析的贡献度筛选指标, 使聚类结果更优。

1 指标选取

鉴于交通运输业本身的社会性、系统性和复杂性, 单一地选取某一类指标难以把握各地区交通运输业的分布差异, 考虑数据的完整性、可靠性, 在众多指标中选取能刻画交通运输业基本状况的社会经济指标、产运量指标和生产要素指标建立交通运输业发展分析指标体系(见表 1)。鉴于航空运输线路的特殊性和管道运输主要运输液态或气态货物的特性, 在此不予讨论。

表 1 交通运输业发展分析指标体系

指标类型	具体指标
社会经济	地区人均 GDP; 交通运输业固定资产投资; 交通运输业的地区生产总值; 交通事故直接财产损失; 居民消费水平; 城市化水平
生产要素	公路线路密度; 高速公路线路密度; 铁路线路密度
产运量	货运总量、公路货运量、铁路货运量、水路货运量; 客运总量、公路客运量、铁路客运量、水路客运量; 货物平均运距、公路货物平均运距、铁路货物平均运距、水路货物平均运距; 旅客平均运距、公路旅客平均运距、铁路旅客平均运距、水路旅客平均运距

2 基于变异系数的区域差异化动态分析

对 2016 年交通运输业各地区主要指标数据的集中趋势和离散程度进行统计分析, 发现大部分地区的交通运输业指标值低于全国平均水平, 且集中在比平均值更小的数值范围, 地区间的差

异明显存在。为准确把握地区分布差异化动态特征, 以各地区为样本, 对主要指标的变异系数进行动态分析。图 1 为近 17 年主要指标变异系数动态分析结果。

由图 1 可知中国交通运输业总体发展的地区差异: 1) 人均 GDP 的变异系数下降, 各地区的经济差

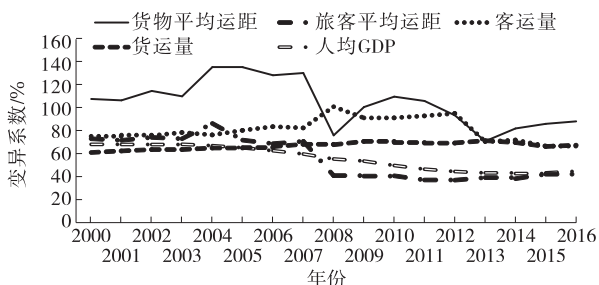


图1 交通运输业主要指标变异系数变化趋势

异呈现缩小趋势,2006—2015年降幅高达20%,表明近十几年地区经济协调发展,地区经济相对差异逐渐缩小。2) 货、客运量反映地区的运输需求和供给能力,其变异系数总体呈现先增后减的趋势,各地区的运输总量差异逐渐缩小,变异程度较缓。3) 客货平均运距反映地区客货运输强度,地区间旅客运输强度比货物运输强度低。地区间旅客运输强度差异不明显,且差异趋于缩小,2008年后地区间差异保持在较低水平;而地区货物运输强度差异较大,在2008、2013年差异趋小,但均有回弹趋势,地区间货物运输强度分布不均,与交通运输业基础设施建设不平衡密切相关。4) 地区间交通运输业的差异比地区经济差异更大,产业分布不均、区域交通需求差异是造成地区间交通运输业发展差异较大的原因。

综上所述,地区间运量差异较小,运输强度差异较大,且差异变化起伏不定,说明地区间的运输存在不合理分配。交通运输业的地区差异变化程度不一,近10年差异正在逐渐减小,其变化趋势与“先富带动后富”政策相适应,在部分地区交通运输业达到一定发展水平时开始注重总体发展,以均衡地区差异。

3 基于Ward法的系统聚类分析

以2016年的主要指标数据对地区进行分类,进一步划分区域,分析区域的差异性特征。

3.1 Ward Method 模型原理

Ward系统聚类法又称离差平方和法,适用于多因素、多指标的分类和特征识别。其以欧氏距离为标准,先将集合中每个样本自成一类,再计算类重心间方差,将离差平方和增加幅度最小的两类首先合并,由此依次将所有类别逐级合并。

算法如下:将 n 个区域样本分成 k 类,表示为 $G_1, G_2, G_3, \dots, G_k$, $X_j^{(i)}$ 为 G_i 中的第 j 个样本($X_j^{(i)}$ 为 P 维向量,即有 P 个系统聚类指标), n_i 为 G_i 中的样本个数, $\bar{X}^{(i)}$ 为 G_i 的重心(即该类样本的均值),则 G_i 中样本的离差平方和 S_i 为:

$$S_i = \sum_{j=1}^n (X_j^{(i)} - \bar{X}^{(i)})^2 \quad (1)$$

k 个类的类内差平方和 S 为:

$$S = \sum_{i=1}^k S_i = \sum_{i=1}^k \sum_{j=1}^{n_i} (X_j^{(i)} - \bar{X}^{(i)})^2 \quad (2)$$

3.2 系统聚类方法模型运算

Ward法的测量区间必须使用欧氏距离,使用标准差标准化法对指标进行标准化,聚类分析过程通过SPSS 22.0软件实现。

3.2.1 模型试算及优化

在初步分析时,指标体系中的25个指标均参与聚类分析,未经过比对和筛选,因指标过多,影响分类的聚集,且一些指标对分类的贡献值很低,故需对指标进行筛选、优化。对各指标进行单因素方差分析,各指标检验值见表2。

表2 指标贡献值比对

指标	贡献值 F	显著性 P	指标	贡献值 F	显著性 P
地区人均GDP	17.194	0.000	公路货物平均运距	1.882	0.156
货运总量	13.844	0.000	铁路货物平均运距	2.304	0.099
公路货运量	16.301	0.000	水路货物平均运距	2.528	0.078
铁路货运量	1.362	0.276	旅客平均运距	2.080	0.126
水路货运量	3.637	0.025	公路旅客平均运距	1.753	0.180
客运总量	20.592	0.000	铁路旅客平均运距	8.529	0.000
公路客运量	18.187	0.000	水路旅客平均运距	0.858	0.475
铁路客运量	12.881	0.000	交通运输业地区生产总值	11.715	0.000
水路客运量	4.544	0.011	交通运输业固定资产投资	9.383	0.000
公路线路密度	14.829	0.000	地区居民消费水平	21.712	0.000
高速公路线路密度	30.545	0.000	城市化水平	18.438	0.000
铁路线路密度	58.251	0.000	交通事故直接财产损失	14.814	0.000
货物平均运距	4.449	0.012			

根据 95%置信水平下显著性 $P < 0.05$ 则指标变量在组间具有显著性差异、贡献值 F 越大则指标对分类贡献越大的原则,得出铁路线路密度对分类的影响最大,其次是高速公路线路密度,以此类推。为使分类更明确,除去无显著性差异的指标,选取贡献值大于 10 的指标作为最终聚类模型的输入指标,分别为地区人均 GDP、货运总量、公路货运量、客运总量、公路客运量、铁路客运量、公路线路密度、高速公路线路密度、铁路线路密度、

交通运输业地区生产总值、地区居民消费水平、城市化水平、交通事故直接财产损失。将参与聚类分析的指标称为 A 类指标,未参与聚类分析的指标称为 B 类指标。

3.2.2 模型优化后的运算结果

对指标变量进行优化后,使用 SPSS 软件,用 Ward 法模型对 31 个省、市、自治区交通运输业发展差异性进行聚类分析。图 2 为指标优化前后的谱系图输出结果。

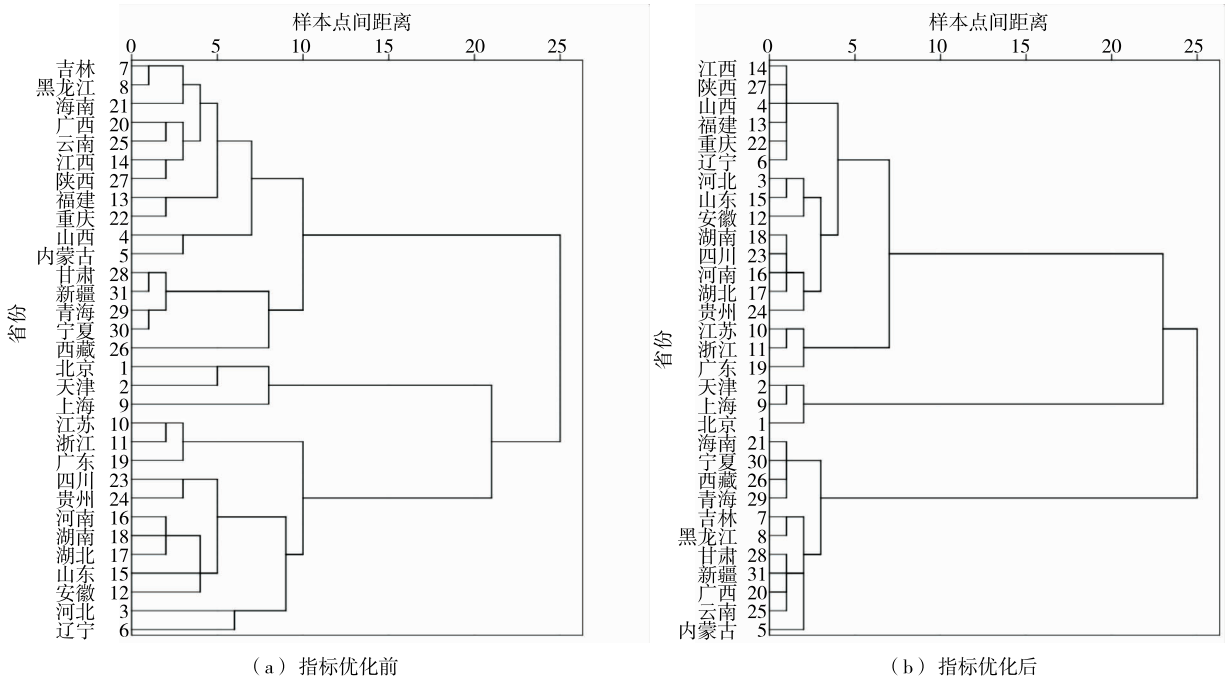


图 2 指标优化前后的谱系图

由图 2 可知:个案之间差距有明显的层次,个案数量也相对分散,聚类过程距离变化幅度较大,说明分类较合理。指标优化后,因剔除了贡献值小的指标的影响,分类的组内距离较小、组间距离大,分类的可靠性得到提升。按 Ward 法将 31 个地区分成 3 类、4 类、5 类均较合理,根据组间距离尽可能大、组内距离尽可能小的原则,分成 3 类为最优。

在谱系图中,当类间距离取 7~25 时,31 个省、市、自治区可合并成三大类,三大类中又可分为 6 个亚类(见表 3)。

3.3 聚类结果验证

根据聚类结果,对每类区域指标的变异系数进行类间对比分析,进一步验证分类的可靠性,对比结果见图 3。

由图 3 可知:整体而言,各类区域指标的变异系数均比全国低,说明区域内指标差异缩小,聚类分析

结果有效。第三类区域的水路指标的变异系数较大,这是由于地区地理分布存在较大差异,导致水路在分类中的差异难以消除。社会经济类、公路和铁路在各地区中属于均衡差异,而水路属于极端差异。明确差异存在的客观性,有助于进一步了解各区域发展特征。

表 3 聚类分析结果

类别	地区
一类区域	北京、上海、天津
二类区域	江苏、浙江、广东;河北、山东、安徽;湖南、四川、河南、湖北、贵州;江西、山西、陕西、福建、重庆、辽宁
三类区域	吉林、黑龙江、甘肃、新疆、广西、云南、内蒙古;海南、宁夏、西藏、青海

注:大类中用分号隔开的地区又自成一个亚类。

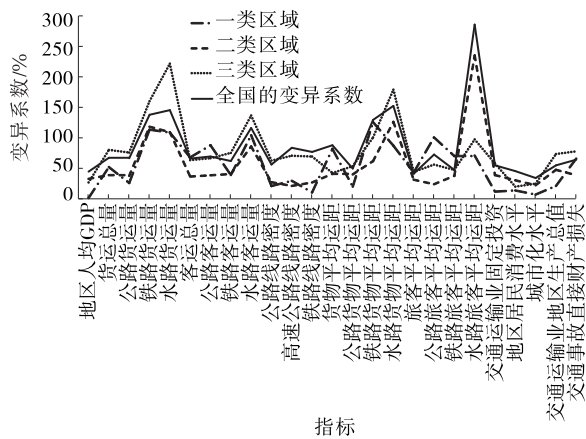


图3 三类区域指标与全国指标变异系数比较

4 主要结论及发展建议

4.1 区域分布特征主要结论

对聚类结果进行讨论和分析,以每类区域中指

标的平均数与全国平均值之比的相对水平,结合地区经济和交通运输业发展水平探讨形成发展状况差异性的原因及影响。各指标在三大类中的分布情况见表4和表5。

三类区域的交通运输业发展情况相对水平层次分明,呈现一定的规律:社会经济类指标在第一类区域的相对水平最高,依次递减;路网建设的密集程度与区域经济发达程度正相关;交通运输业的客货运量主要集中在第二类区域;不同运输方式中货物、旅客平均运距存在差异。由此得出交通运输业发展区域分布特征如下:

(1) 第一类区域属于交通运输业发达且社会经济发达地区。该区域各地区的社会经济指标、线路规模均高于全国平均水平,整体运量和运力都基本适应经济的发展和需求。公路货物平均运距相对水平最低,与该地区的公路线路密度大、路网完善程度

表4 三类区域综合经济和交通运输业发展状况(A类指标)

类别	项目	地区 人均 GDP	货运 总量	公路 货运量	客运 总量	公路 客运量	铁路 客运量	公路 线路 密度	高速 公路 线路 密度	铁路 线路 密度	交通运 输业地 区生产 总值	地区 居民 消费 水平	城市化 水平	交通事 故直接 财产 损失
一类 区域	平均数	11.63	53 188	30 623	31 437	21 728	9 544	1.636	0.099	0.085	1 008	44 919	0.826	2 560
	相对水平	2.06	0.38	0.28	0.53	0.44	1.05	1.74	3.09	3.06	0.89	2.11	1.82	0.66
二类 区域	平均数	5.49	197 659	153 719	86 371	72 942	12 113	1.183	0.034	0.027	1 593	20 599	0.427	5 400
	相对水平	0.97	1.43	1.43	1.45	1.47	1.33	1.26	1.05	0.97	1.41	0.97	0.94	1.39
三类 区域	平均数	4.30	70 740	57 834	26 253	21 596	4 259	0.377	0.011	0.013	456	15 895	0.391	1 934
	相对水平	0.8	0.5	0.5	0.4	0.4	0.5	0.4	0.4	0.5	0.4	0.7	0.9	0.5
总计	平均数	5.66	138 642	107 783	59 723	49 766	9 078	0.941	0.032	0.028	1 133	21 284	0.453	3 895

注:平均数中,地区人均GDP的单位为万元,货运总量、公路货运量的单位为万t,客运总量、公路和铁路客运量的单位为万人,线路密度的单位为km/km²,交通运输业地区生产总值的单位为亿元,交通事故直接财产损失的单位为万元。

表5 三类区域综合经济和交通运输业发展状况(B类指标)

类别	项目	铁路 货运量	水路 货运量	水路 客运量	货物 平均 运距	公路 货物 平均 运距	铁路 货物 平均 运距	水路 货物 平均 运距	旅客 平均 运距	公路 旅客 平均 运距	铁路 旅客 平均 运距	水路 旅客 平均 运距	交通 运输业 固定 投资
一类 区域	平均数	3 131	19 434	166	1 014	89	3 139	1 836	112	140	203	11	811
	相对水平	0.29	0.97	0.19	2.53	0.50	2.52	2.11	0.80	1.64	0.42	0.24	0.51
二类 区域	平均数	12 983	30 957	1 316	350	180	949	1 158	129	65	506	76	2 070
	相对水平	1.21	1.54	1.50	0.87	1.01	0.76	1.33	0.93	0.76	1.05	1.60	1.29
三类 区域	平均数	9 372	3 534	397	314	198	1 188	163	162	102	524	14	1 091
	相对水平	0.87	0.18	0.45	0.78	1.12	0.95	0.19	1.17	1.19	1.08	0.29	0.68
总计	平均数	10 748	20 111	878	401	177	1246	871	139	85	483	48	1 601

注:平均数中,铁路和水路货运量的单位为万t,水路客运量的单位为万人,平均运距的单位为km,交通运输业固定投资的单位为亿元。

高的特点相适应,相应地,货物运输强度较小,而其他运输方式的货物运输强度较大,铁路货物平均运距甚至高达2.52倍;旅客平均运距则相反,公路的旅客运输强度大,而其他运输方式的运输强度较低。

(2) 第二类区域属于交通运输业较发达、经济水平增长快速的地区,主要分布在沿海和中部地区。经济次发达地区经济增长速度快,对交通运输业的依赖更大,运量较大,其相对水平平均大于1;社会经济类指标和产值指标的相对水平与全国平均水平持平;交通运输业的固定资产投资占比最大。

(3) 第三类区域属于交通运输业、社会经济均欠发达地区,主要分布在内陆偏远地区。该区域各地的社会经济和运量指标相对水平均很低,远未达到全国平均水平;公路和铁路的客货平均运距相对水平很高,说明其运量相对较小;运输强度很大,与该区域公路和铁路建设基础薄弱有关。

4.2 交通运输业区域发展建议

为进一步优化和调整运输结构及运输网络布局,增强交通运输业的效益,使交通运输体系更好地适应社会经济发展,对中国交通运输业区域发展提出以下建议:

(1) 加强交通基础设施建设,形成良好的交通运输发展环境,推动经济发展。第三类区域交通运输基础设施建设情况较差,亟待加强。建议分阶段推进,短期内着力发展建设周期、投资回收周期较短的民航运输,满足短期运输需求,拉动当地经济增长;长期发展需投入大量资金,因地制宜,合理建设公路、铁路基础设施,形成良好交通运输发展环境。

(2) 拓宽路网规模,与经济协调发展。第二类区域处于运输量快速上升期,经济发展势态良好,交通需求较大,应进一步加强交通基础设施建设、完善交通运输系统,为运输由“量”转“质”做准备。

(3) 优化运输线路,转压力为动力,提高运输质量。第一类区域应加强与周围地区城市路网的连接,通过合理优化运输线路,充分发挥转运枢纽的作用,缩短运输距离,减小运输周转量,降低运输强度,使运输资源得到充分合理利用,同时优化网络结构,将自身的运输压力转变为辐射地区的运输动力,带动辐射地区的发展。

5 结语

该文根据中国交通运输业发展状况建立社会经济、产运量、生产要素指标体系,进行各指标变异系

数动态特征分析,运用 Ward Method 模型将 31 个地区分成交通运输业发展水平发达、较发达、欠发达三类区域,并在聚类分析时剔除对聚类结果贡献较小的指标,使分类更优。对三类区域各指标的相对水平进行比较,确保区域分布特征分析结果的可靠性。将变异系数和聚类分析相结合应用于交通运输业发展区域分布特征分析可行,对规划、调整交通运输业的宏观布局具有重要理论价值。

参考文献:

- [1] 侯平路.中国各省区交通发展水平分析[J].科技信息,2010(17).
- [2] 张玮,朱金福.我国交通运输业发展区域差异性研究[J].企业经济,2008(6).
- [3] 姚新胜,王清宇,徐杏.我国交通运输业区域分布特性[J].经济地理,2006,26(4).
- [4] 谢霖铨,谭光霞.聚类分析在道路运输规划中的应用研究[J].铁道运输与经济,2007,29(3).
- [5] 周志龙.考虑区域经济特征的中国综合交通运输结构变化实证分析[D].北京:北京交通大学,2011.
- [6] 杜彩军,周健.我国综合运输统计指标体系构建研究[J].综合运输,2013(10).
- [7] 何玉花,张东水,韩用顺,等.基于 Ward 系统聚类方法的黄河上游干流地区滑坡分类及其特征分析[J].中国地质灾害与防治学报,2018,29(1).
- [8] 项晓敏,金晓斌,杜心栋.基于 Ward 系统聚类的中国农用地整治实施状况分析[J].农业工程学报,2015,31(6).
- [9] 林森,李雪玲.基于因子和聚类分析的市政设施水平评价[J].公路与汽运,2011(4).
- [10] 袁长伟,乔丹,杨颖芳,等.中国省域交通碳排放强度空间分异与聚类分析[J].环境工程,2018(7).
- [11] 熊萍,石义寿.基于泰尔指数的交通运输业对区域经济影响差异分析[J].重庆交通大学学报:社会科学版,2017,17(1).
- [12] 梁振芹.公路运输与区域经济协调发展的对策[J].中国城市经济,2011(24).
- [13] 张锦.中国运输业发展区域差异性及其收敛性研究[J].中国市场,2013(2).
- [14] 杨令宾,邹滨,黄宇萍,等.中国省际综合交通发展水平的定量分析[J].经济地理,2008,28(2).
- [15] 邓童,李泽宇,万美洁.中国城市经济与交通状况的模糊聚类分析[J].商场现代化,2008(25).