

基于k-means算法的高速公路服务区分类研究

毛俊飞¹, 张冲², 高超², 刘兴国², 王光亚²

¹山东交通学院, 轨道交通学院, 山东 济南

²山东交通学院, 交通物流学院, 山东 济南

收稿日期: 2024年1月6日; 录用日期: 2024年3月2日; 发布日期: 2024年3月12日

摘要

截至2023年6月, 我国高速公路的建设规模已达到17.73万公里, 位居全球之首。根据国家公路网的长远规划预计到2035年, 我国的高速公路总长度将达到16.2万公里, 现如今已经提前完成规划, 基本形成一个现代化、高质量的国家公路网络。这不仅在新的发展格局中发挥着重要作用, 而且对推动经济发展具有关键性的支持作用。在这个背景下, 高速公路服务区的分类成为了一个研究重点。它对于理解不同类型服务区周边的土地利用、客流变化规律、发展趋势和建设规划具有重要意义。为了深入分析这一课题, 本文采用了聚类分析方法。通过收集服务区附近的兴趣点(POI)、交通枢纽等关键设施的数据, 作为研究的主要变量。在对数据进行Z-Score标准化处理后, 本研究运用了主成分分析和k-means聚类方法对高速公路服务区进行了细致的分类。根据我们的分析结果, 这些服务区可以分为四类: 综合型服务区、融合型生产服务区、旅游型服务区以及工业型生产服务区。这一分类不仅为我们后续的深入研究奠定了基础, 而且为服务区未来的发展方向提供了参考。

关键词

聚类分析, 主成分分析, k-means聚类, 高速公路服务区

Research on Classification of Expressway Service Area Based on k-means Algorithm

Junfei Mao¹, Chong Zhang², Chao Gao², Xingguo Liu², Guangya Wang²

¹School of Railway Transportation, Shandong Jiaotong University, Jinan Shandong

²School of Transportation and Logistics Engineering, Shandong Jiaotong University, Jinan Shandong

Received: Jan. 6th, 2024; accepted: Mar. 2nd, 2024; published: Mar. 12th, 2024

Abstract

As of June 2023, China has achieved a remarkable milestone in infrastructure development by

文章引用: 毛俊飞, 张冲, 高超, 刘兴国, 王光亚. 基于 k-means 算法的高速公路服务区分类研究[J]. 交通技术, 2024, 13(2): 73-80. DOI: 10.12677/ojtt.2024.132009

constructing 177,300 kilometers of highways, the largest network globally. According to the long-term plan for the national highway network, the total length of highways in China is projected to reach 162,000 kilometers by 2035, the planning has been completed ahead of schedule, and a modern, high-quality national highway network has basically been formed. This significant development plays a crucial role in shaping the new development pattern and serves as a pivotal support for economic growth. In this context, the classification of highway service areas has become a research focus. It is of significant importance for understanding the land use around different types of service areas, the patterns of passenger flow changes, development trends, and construction planning. To deeply analyze this topic, this paper adopts the method of cluster analysis. By collecting data on points of interest (POI), transportation hubs, and other key facilities near the service area, as the main variables for the study. After the data was standardized using the Z-Score method, this study applied principal component analysis and k-means clustering methods to perform a detailed classification of highway service areas. According to our analysis results, these service areas can be divided into four categories: comprehensive service areas, integrated production service areas, tourism-oriented service areas, and industrial production service areas. This classification not only lays the foundation for our subsequent in-depth research but also provides a reference for the future development direction of service areas.

Keywords

Cluster Analysis, Principal Component Analysis, k-means Clustering, Motorway Service Area

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

习近平总书记强调“交通成为中国现代化的开路先锋”，赋予了交通运输新的历史使命。党的二十大报告指出，加快建设交通强国，加快交通运输结构调整优化，构建现代化基础设施体系。

截至 2021 年底，山东全省高速公路通车里程达 7477 公里[1]，形成以“九纵五横一环七射多连”高速公路网为主骨架的格局，路网密度和通达能力居全国前列。国外如美国将服务区定义为保障高速公路系统安全运行的重要设施，提供车辆停靠区域、休息桌椅、卫生间、水以及其他供给，还可为车辆的短暂停留提供野餐场所。在服务区的建设过程，他们注重资源节约和无公害环保技术的使用，并根据当地环境和场地，对服务区建筑风格、材料、颜色、植物和景观小品进行设计，体现当地特色，因此划分出几种特殊类型的服务区，如游客信息中心、服务中心，短信区、路边停靠区、卡车服务站。国内现有关于高速公路服务区规划布局与分类分级的研究成果多集中于新建服务区，贾珍珍[2]等利用 GIS 技术分析新建服务区规划布局；周畅[3]等对山区旅游型服务区建设提出了一系列对策；周舒灵[4]等对浙江省高速公路服务区改扩建在具体技术措施上提出了建议。刘洋[5]等基于层次分析法对高速公路服务区进行了分类分级研究。目前的研究多为增量方面，然而随着我国近十几年来基础设施建设的迅猛发展，优化存量、激活现有设施也是当前服务区建设的一个重要方面。

本文以山东省 132 个高速公路服务区为例，从服务区周围的交通枢纽、POI 点分布数量、POI 点分布类型进行聚类分析[6]。基于服务区聚类结果，分析总结各类服务区的普适性特色规律，为高速公路路线规划发展总结经验[7]。

2. 关于聚类研究

2.1. 聚类因素选取和数值标准化

2.1.1. 选取原则

选取高速公路服务区进行聚类所使用的变量因素需要满足以下条件。

1) 首先需要明确高速公路服务区的影响范围，是将因素数值标准化的基础前提。

2) 需要反映服务区附近城镇的经济、交通发展程度、产业资源和地域特色。

3) 需要能够反映出服务区自身的特色，既要反映出服务区自身属性，同时又能显现出周围地域的属性特色。

2.1.2. 聚类因素选取

根据实际情况需要，本文选取了服务区附近 5 公里范围作为服务区分标准[8]，同时考虑选取变量的原则，选定了以下的四类数据作为聚类分析使用的参数变量。

1) 选定范围内的各种交通数据，包括公交站台数、地铁站点数和港口码头、长途汽车站点、机场和火车站点等各类交通枢纽，交通物流类公司，交通最能够直接体现出附近城镇发展的情况。

附近城镇人口分布及教育情况的数据，通过统计商务住宅数量计算城镇人口的居住占比，根据教育设施统计计算教育文化产业占比。

2) 附近城镇发展倾向的数据，通过各类消费、娱乐场所 POI 数量计算商业占比，各种企业、工厂数量计算商业占比，展现出服务区附近经济发展趋向。

3) 附近的旅游景点数量，能够体现出附近城镇特色产业发展情况。

综合以上，本文此次聚类研究一共选了四大类数据、8 个变量因素作为初始输入参数，各类变量如表 1 所示。

Table 1. Initial clustering variables

表 1. 初始聚类变量

变量序号	变量名称	变量描述
1	商业比	服务区附近 5 km 内商业型 POI 占附近所有 POI 点比例
2	居住比	服务区附近 5 km 内商务住宅占附近所有 POI 点比例
3	工业比	服务区附近 5 km 内公司企业占附近所有 POI 点比例
4	公交站点数	服务区附近 5 km 内公交站点 POI 的数目
5	枢纽类数量	服务区附近 5 km 内公交站类 POI 的数目
6	景点类数量	服务区附近 5 km 内景点类 POI 的数目
7	文化比	服务区附近 5 km 内科教文化类 POI 占附近所有 POI 点比例
8	物流类点数量	服务区附近 5 km 内物流类型的 POI 的数目

2.1.3. 聚类因素数值标准化

在进行聚类分析之前，通常需要首先对数据进行标准化预处理。数据的标准化能够提高聚类分析的准确性和稳定性，同时也有助于提高算法的收敛速度。为了实现这一目标，通常会采用一些常用的标准化算法，如 Z-Score 标准化、极差标准化法、log 函数标准化法等。在本文选择了 Z-Score 方法来进行数据的标准化预处理。标准化结果如表 2 所示。

Table 2. Standardized results of standard initial variables (partial)**表 2.** 标初始变量标准化结果(部分)

序号	商业比	居住比	工业比	公交站点数	枢纽类数	景点数	文化比	物流类数
1	1.35941	-0.83910	-0.36417	-0.41439	-0.50741	-0.08182	-1.11534	-0.37518
2	-0.82567	-0.12338	1.90238	0.51761	-0.50741	-0.08182	-0.12123	2.97257
3	1.66123	-0.46755	-1.03921	-0.59191	-0.50741	0.21543	-0.51315	-0.14430
4	-0.10328	0.49288	1.66105	-0.50315	-0.50741	-0.08182	-0.59932	0.02886
5	1.04260	-0.86097	0.20841	-0.85819	-0.50741	-0.37907	-0.38277	-0.25974
...

2.2. 因子分析

在进行因子分析前,通常需要先进行巴特利检验和 KMO 检验。通过巴特利球形检验,可以检测变量间是否具有相关性,确保可以提取公因子。当巴特利球形检验结果显著性小于 0.05 时,说明变量间相关性较强,适合进行因子分析。KMO 检验用于考察变量间的偏相关性,取值 0~1 之间;当 KMO 统计量越接近 1,则变量间的偏向关性越强,因子分析效果越好。一般统计量在 0.5 以下不适合做因子分析,当统计量 > 0.7 时适合做因子分析[9]。

本文通过社会科学统计软件包(SPSS)软件对 8 个标准化变量进行巴特利检验和 KMO 检验,结果如表 3 所示。其中巴特利球形检验结果显著性为 0,小于 0.05,数据间相关性较强[10];KMO 检验值为 0.611,大于 0.6,数据间偏相关性较强,比较适合使用因子分析方法。

Table 3. KMO and Bartlett test results**表 3.** KMO 和巴特利特检验结果

KMO 取样適切性量数	巴特利特球形度检验		
	近似卡方	自由度	显著性
0.611	289.437	28	0.000

使用主成分分析法对预处理后的参数进行数据处理,可以得到降维后的公因子。方差表及总方差解释分别如表 4、表 5 所示。其中有三个公因子初始特征值大于 1,这三个公因子旋转后的累计方差贡献率为 69.379%,提取的公因子能够涵盖约 7 成的原始变量提供的信息,说明提取的三个公因子可以对分析结果进行解释。通过将标准化数据矩阵与成分得分系数矩阵相乘,得到解释初始变量特性的公因子矩阵,完成聚类前数据的所有预处理。

Table 4. Common factor variance table**表 4.** 公因子方差表

变量序号	变量名称	初始	提取
1	商业比	1.000	0.832
2	居住比	1.000	0.567
3	工业比	1.000	0.814
4	公交站点数	1.000	0.775
5	枢纽类数量	1.000	0.520
6	景点类数量	1.000	0.718
7	文化比	1.000	0.666
8	物流类点数量	1.000	0.657

Table 5. Total variance explained
表 5. 总方差解释

成分	初始特征值			提取载荷平方和			旋转载荷平方和		
	总计	方差百分比	累积%	总计	方差百分比	累积%	总计	方差百分比	累积%
1	2.686	33.576	33.576	2.686	33.576	33.576	2.661	33.265	33.265
2	1.604	20.049	53.625	1.604	20.049	53.625	1.593	19.909	53.174
3	1.260	15.754	69.379	1.260	15.754	69.379	1.296	16.205	69.379
4	.681	8.518	77.897						
5	.650	8.119	86.016						
6	.518	6.472	92.487						
7	.386	4.823	97.310						
8	.215	2.690	100.000						

第一主成分中公交站点数、景点数、枢纽类数量、居住比例和物流类数的因子载荷比较大，其中公交、枢纽类和物流数量明显是交通类型的功能，同时景点和人口居住对交通的便利程度需求也高，因此将第一主成分定义成交通便利程度。第二主成分中工业化比例因子载荷比较大，因此将第二主成分定义为工业生产力。第三主成分中文化占比因子载荷较大，居住比也有部分比例，且人口在当地居住越高对文化建设需求也会越高，因此将第三成分定义为教育建设度。旋转后的成分矩阵如表 6 所示。

Table 6. Component matrix after rotation
表 6. 旋转后的成分矩阵

初始变量	成分 1	成分 2	成分 3
公交站点数	0.879	-0.009	-0.042
景点数	0.836	-0.080	0.116
枢纽类数	0.701	-0.167	0.010
居住比	0.602	0.240	0.384
商业比	0.026	-0.884	-0.225
工业比	-0.084	0.835	-0.331
文化比	0.207	0.005	0.789
物流类数	0.534	0.148	-0.592

2.3. 服务区聚类

聚类方法

k-means 算法是一种常用的聚类算法，实现原理为将数据集划分为具有相似属性的 k 个簇。该算法采用迭代的方式，通过最小化每个样本点与所属簇中心之间的距离来实现聚类[11]。算法具体由以下几个步骤实现：

- (1) 随机选取 n 个对象作为初始的聚类中心；
- (2) 分别计算每个对象到聚类中心的欧氏距离，并将对象划分到最近的聚类中；
- (3) 根据最新生成的聚类，将其各个方向参数相加求平均值，生成新的聚类中心；
- (4) 重复执行(2)~(3)步，直至完成迭代条件。

3. 基于山东高速公路服务站聚类特征分析

3.1. 聚类结果

使用 k-means 算法进行分类后，将不同的类别用不同的颜色标出区分，一共分成了四类。分类结果包括从 X、Y 轴方向观测如图 1~3 所示。其中 X 轴代表交通便利程度，Y 轴代表工业生产力，Z 轴代表教育建设度。详细分类结果如表 7 所示。

高速公路服务区kmeans聚类

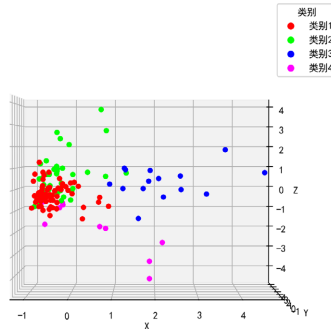


Figure 1. X-axis direction observation classification results
图 1. X 轴方向观测分类结果

高速公路服务区kmeans聚类

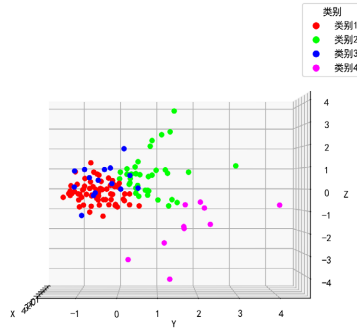


Figure 2. Y-axis direction observation classification results
图 2. Y 轴方向观测分类结果

高速公路服务区kmeans聚类

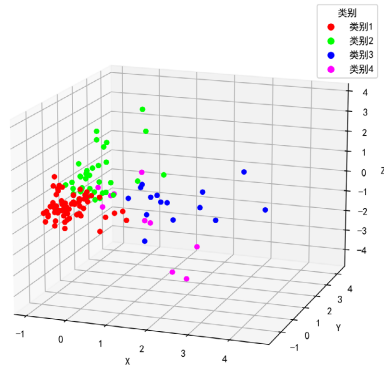


Figure 3. Categorize the results
图 3. 分类结果

Table 7. Detailed classification
表 7. 详细分类情况

服务区类别	服务区	数量
综合型服务区	日兰高速曹州服务区、莘南高速莘县西服务区、青兰高速莘县服务区、日兰高速菏泽服务区、德上高速莘县停车区、德上高速东昌府服务区、青银高速夏津服务区、青兰高速东阿服务区、日兰高速巨野服务区、青银高速高唐服务区…	69
融合型生产服务区	德上高速鄄城服务区、东吕高速高唐西停车区、京台高速德州服务区、青兰高速平阴服务区、济广高速长清服务区、京台高速济南服务区、青银高速齐河停车区、济南绕城高速天桥服务区、岚菏高速峰城服务区、滨德高速阳信服务区…	38
旅游型服务区	董梁高速梁山北服务区、京台高速禹城服务区、日兰高速济宁服务区、京台高速泰安服务区、京台高速曲阜服务区、京台高速枣庄服务区、济南大东环服务区、泰新高速新汶服务区、京沪高速莱芜服务区、滨台高速博山停车区、青银高速淄博服务区、青兰高速沂源服务区、荣潍高速坊子服务区、荣乌高速莱州服务区、荣潍高速平度服务区	15
工业型生产服务区	济聊高速冠县服务区、济聊高速聊城服务区、济广高速梁山服务区、荣乌高速利津服务区、长深高速青州服务区、长深高速临朐停车区、青银高速潍坊西服务区、荣乌高速潍坊北服务区、青银高速青岛服务区、沈海高速福山服务区	10

3.2. 服务区分类特征分析

1) 综合型服务区数量最多在结果中用红色表示，这类服务区附近交通便利程度、工业生产力和教育建设程度都比较均衡，主要分布在-1~1 内，分布相对聚集因此将该类定义为了综合类服务区。

2) 融合型生产服务区在结果中用绿色表示，这类服务区附近教育建设程度比较强，工业生产力相对其他类别也偏高，因此将之定义为融合型生产服务区

3) 旅游型服务区在结果中用蓝色表示，这类服务区附近工业生产力和教育建设度比较一般，但是在交通便利度上远超其他服务区，根据交通便利度的特性，将之定义为旅游型服务区。

4) 工业型生产服务区在结果中用粉色表示，这类服务区附近工业生产力普遍高于均值，文化建设程度几乎全在均值之下表明了常驻人口较少的特性，因此将之定义为工业型生产服务区。

对服务区所属类型进行分类研究，有助于更好地理解 and 满足用户需求，提高运营效率，优化资源配置，并制定更有针对性的发展和市场营销策略。

4. 结语

在高速公路服务区的研究中，通过考察服务区周边的发展状况，我们可以对服务区进行有效的分类。本研究以山东省的高速公路服务区为例，采用了定量和定性的分析方法，不仅考虑了交通流量、服务设施的数量和质量等定量指标，也考虑了服务区的地理位置、周边环境、用户需求等定性因素。基于这些分析，我们提出了不同类型的服务区分类，如综合型服务区、融合型生产服务区、旅游型服务区、工业型生产服务区等。

此外，研究还探讨了服务区分类建设策略的应用。例如，如何根据服务区的地理位置和用户需求，设计和提供相应的服务。随着社会经济的发展，人们对高速公路服务区的需求也在不断变化。因此，未来的研究将更加注重对计算模型的优化，以确保服务区分类和服务能够更加精准地满足用户的需求，同时也促进了当地经济的发展。

基金项目

过饱和状态下交叉口信号控制策略研究(2023YK066)。

参考文献

- [1] 国家发展改革委交通运输部. 国家公路网规划[EB/OL].
https://www.gov.cn/zhengce/zhengceku/2022-07/12/content_5700633.htm, 2022-07-14.
- [2] 贾珍珍, 徐静. 基于 GIS 的高速公路服务区布局规划研究[J]. 交通建设与管理, 2020, 475(2): 68-69.
- [3] 周畅, 何畏. 浅谈山区高速公路旅游型服务区规划设计[J]. 公路交通技术, 2021, 37(2): 133-138.
- [4] 周舒灵. 浙江省高速公路服务区改扩建规划设计探研[J]. 城市建筑, 2020, 17(25): 121-126.
- [5] 刘洋, 陈时通, 杨博, 等. 基于层次分析法的高速公路服务区分类分级研究: 以湖南省为例[J]. 中外建筑, 2023(5): 86-92.
- [6] 张玲. POI 的分类标准研究[J]. 测绘通报, 2012(10): 82-84.
- [7] 赵明飞. 高速公路开放式服务区功能分类及设计思路[J]. 交通世界, 2022(19): 97-99.
- [8] 杜艳花. 高速公路服务区分类及用地规模调查研究[J]. 山西交通科技, 2022(2): 100-102.
- [9] 尹芹, 孟斌, 张丽英. 基于客流特征的北京地铁站点类型识别[J]. 地理科学进展, 2016, 35(1): 126-134.
- [10] 陈蔚珊, 柳林, 梁育填. 广州轨道交通枢纽零售业的特征聚类及时空演变[J]. 地理学报, 2015, 70(6): 879-892.
- [11] 夏雪, 盖靖元. 基于 K-Means 聚类算法的城市轨道交通站点分类及客流特征分析[J]. 现代城市轨道交通, 2021(4): 112-118.