

教育资源是如何聚集的？以成都市重点小学为例

汪诗璐，李旭升

(西南交通大学经济管理学院)

摘要：尽管以往研究已利用宏观经济指标对我国教育资源的聚集性进行了探讨，但该类指标并不能从微观角度有效应用于城市内部基础教育资源的空间聚集性研究。本文则借助于地理统计学方法与机器学习算法，利用标准差椭圆和 DBSCAN 聚类算法，从聚集方向、聚集度与聚集区域三个维度，对成都市重点小学的聚集情况进行了评估。结果表明，成都市重点小学存在两个聚集区域，且较普通小学空间分布更为聚集，并具有更强的方向性。

关键词：地方教育；教育资源聚集；资源配置；教育公平

中图分类号：G527

How are Educational Resources Gathered? Example of The Key Primary Schools in Chengdu

WANG Shilu, LI Xusheng

(School of Economics and Management of Southwest Jiaotong University, Chengdu 610031)

Abstract: Although previous studies have used macroeconomic indicators to discuss the agglomeration of educational resources in my country, such indicators cannot be effectively applied to the study of spatial agglomeration of basic educational resources within cities from a microscopic perspective. In this paper, with the help of geostatistical methods and machine learning algorithms, using standard deviation ellipse and DBSCAN clustering algorithm, from the three dimensions of clustering direction, clustering degree and clustering area, the clustering situation of key primary schools in Chengdu is evaluated. The results show that there are two clustering areas in key primary schools in Chengdu, and the spatial distribution is more clustered and more directional than ordinary primary schools.

Key words: local education; educational resource aggregation; resource allocation; educational equity

0 引言

教育公平作为社会公平的一部分，是现代社会的发展目标，也是维护国家安全与发展的基本保障。公平是对利益关系调整 and 资源配置合理性的价值判断^[1]，优质教育资源的空间聚集被认为是配置不合理的表现之一，严重影响了教育公平的实现。

研究人员致力于评估教育资源的空间聚集情况。鉴于我国教育事业已具有经济产业的性质^[2]，大量研究利用衡量产业聚集度的宏观经济指标评估教育资源聚集性，例如赫芬达尔-赫希曼指数^[3]、产业集中度指数^[4]、教育基尼系数^[5]等。

然而，这些研究有两个特点：一是囿于宏观经济指标的局限性，以往研究主要是以全国为研究范围，以各省市或城乡之间的教育资源空间分布情况为研究对象，鲜有对于城市内部教育资源聚集性的探讨，由此带来的 MAUP 问题 (Modifiable Areal Unit Problem, 可塑性面

作者简介：汪诗璐 (1998-)，女，硕士研究生，主要研究方向：数据分析，房地产大数据

通信联系人：李旭升 (1970-)，男，博士，副教授，硕导，主要研究方向：数据挖掘、工程管理、房地产管理。E-mail: xsli@swjtu.edu.cn

积单元问题,即分析结果随基本面积单元定义的不同而变化的问题)可能会影响结果的可信度,且结果无法为城市管理者提供有效参考;二是主要以间接变量而不是学校本身代表教育资源,如刘善槐以学龄人口数代表义务教育资源^[6],罗华玲以在校生人数、师生比等计算教育基尼系数和 HHI 指数^[7],这种间接变量的有效性与说服力值得商榷。

近年来,已有研究利用 DO 指数和标准差椭圆法等基于坐标的地理统计学方法^[8-10],以及 K-means 与 DBSCAN 等机器学习聚类算法^[11,12]衡量产业聚集度。这两类方法能有效控制上文所述的两个问题,但尚未应用于教育产业聚集性的研究中。

因此本文将分别利用标准差椭圆法与 DBSCAN 算法,以成都市重点小学为研究对象,从聚集方向、聚集度和聚集区域三个维度评估其聚集情况,以弥补该研究领域的空白,并为城市决策者提供参考。

1 研究区域与数据来源

1.1 研究区域

新一线城市,是《第一财经周刊》根据商业资源集聚度、城市枢纽性、城市人口活跃度、生活方式多样性和未来可塑性五大指标综合评比后划分的名单。而成都市作为西南地区的经济文化中心,近年来经济稳定发展,人口稳定增长,商业繁荣,社会和谐,已连续六年进入新一线城市行列,且位于榜单首位。在某种程度上,成都市可以代表中国城市经济社会的发展趋势。

鉴于成都市占地面积较大,使得周边辖区与中央城区之间经济、交通等发展程度差别很大。因此本文仅选取成都市“小成都”区域,即成都市最初设立的五个主要市辖区:金牛区、成华区、锦江区、武侯区、青羊区作为本文的研究区域。这五个辖区设立历史相同,经济发展情况相似,范围小,距离近,能有效排除一些无关的文化、经济等因素给模型带来的误差。

1.2 研究数据

由于中学的入学资格主要仍来自于统一考试,而小学升学则因“就近入学”政策所带来的与房地产市场的捆绑关系,受到家庭经济背景的影响,继而影响教育获得的公平性,因此本文以重点小学为研究对象。

基于上述情况,本文从成都市教育局官方网站获取了成都市五城区 206 所小学的基本信息,包括名称、地址、公民办性质;并从我要搜学网(<http://www.51sxue.com/>)获取了成都市五城区 44 所市重点与省重点小学名录;接着利用高德地图开放平台(<https://lbs.amap.com/>)提供的 API 获取了各小学的经纬度信息。

对获取到的学校数据进行描述性统计,得到学校数据分布的基本信息如表 1。由表 1 可知,成都市五城区重点小学有 44 所,约占学校总数量的 21.4%;其中,锦江区学校数量虽仅占成都市小学数量的 17%,但其重点小学数量占比却高达 36%,可见在总体上成都市重点小学呈现出分布不均衡的情况,且主要集中在锦江区。

表 1 成都市小学描述性统计

Tab. 1 Descriptive Statistics of Chengdu Primary Schools

区域	类别	数量/所	占成都市总体比例
成华区	普通小学	31	
	重点小学	6	14%
	合计	37	18%
金牛区	普通小学	46	
	重点小学	5	11%
	合计	51	25%
锦江区	普通小学	20	
	重点小学	16	36%
	合计	36	17%
青羊区	普通小学	27	
	重点小学	11	25%
	合计	38	18%
武侯区	普通小学	38	
	重点小学	6	14%
	合计	44	21%
成都市	普通小学	162	
	重点小学	44	100%
	合计	206	100%

2 研究方法

2.1 标准差椭圆法

80 标准差椭圆是研究要素空间分布方向性的经典方法之一，它可以从多维度分析对象的空间分布特征，最早是由美国社会学教授 Lefever 于 1926 年提出^[13]。基于该方法的有效性，许多学者对将其应用于各领域的研究中，尤其是经济产业的时空分布研究^[14-17]。

标准差椭圆包含三个参数：椭圆的偏角（方位角），即以 x 轴的正北方向为基准顺时针旋转和椭圆长轴形成的夹角；椭圆的最大标准差距离 σ_x ，即椭圆的长轴长度；椭圆的最小距离 σ_y ，即椭圆的短轴长度。长轴和短轴分别为点状要素空间分布最多的和最少的方向。
85 标准差椭圆的面积大小反映了要素的聚集程度，椭圆的偏角则表示要素的聚集方向。

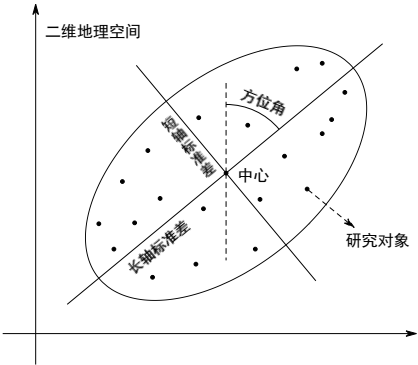


图 1 标准差椭圆的基本参数示例图
Fig. 1 Example Diagram of Basic Parameters of Standard Deviation Ellipse
标准差椭圆中各参数的计算公式如下：

90

1) 偏角

$$\tan \theta = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 - \sum_{i=1}^n (y_i - \bar{y})^2 + \sqrt{\left[\sum_{i=1}^n (x_i - \bar{x})^2 - \sum_{i=1}^n (y_i - \bar{y})^2 \right]^2 + 4 \left[\sum_{i=1}^n (x_i - \bar{x}) \sum_{i=1}^n (y_i - \bar{y}) \right]^2}}{2 \sum_{i=1}^n \sum_{i=1}^n (x_i - \bar{x}) \sum_{i=1}^n (y_i - \bar{y})}$$

公式 1

2) 长轴

$$\sigma_x = \frac{\sqrt{\left[\sum_{i=1}^n (x_i - \bar{x}) \cos \theta - \sum_{i=1}^n (y_i - \bar{y}) \sin \theta \right]^2}}{n}$$

95

公式 2

3) 短轴

$$\sigma_y = \frac{\sqrt{\left[\sum_{i=1}^n (x_i - \bar{x}) \sin \theta + \sum_{i=1}^n (y_i - \bar{y}) \cos \theta \right]^2}}{n}$$

公式 3

其中, x_i 和 y_i 分别为研究对象的空间坐标, \bar{x} 和 \bar{y} 为对象在 x 轴、 y 轴上坐标的平均值, 在一些研究中, 也采用加权的平均中心代替 \bar{x} 和 \bar{y} , 本文则仍然采用坐标的均值。

100

2.2 DBSCAN 算法

105

DBSCAN 是一种基于密度的聚类算法, 它主要通过样本密度来考察样本间的可连接性, 其中簇的形成主要基于样本间距离的定义以及“邻域参数”。它的基本思路是, 通过检查数据集中每点的 Eps 邻域来搜索簇, 如果点 p 的 Eps 邻域包含的点多于 $MinPts$ 个, 则创建一个以 p 为核心对象的簇; 然后, DBSCAN 迭代地聚集从这些核心对象直接密度可达的对象, 这个过程可能涉及一些密度可达簇的合并; 当没有新的点添加到任何簇时, 该过程结束。

在 DBSCAN 算法中, 所有的对象都分为三类: 核心点、边界点与离群点。

1) 核心点

若某一对象在其半径 Eps 的邻域内包含了至少 $MinPts$ 个对象, 则该对象为一个核心点。

110

2) 边界点

若某一对象在其半径 Eps 的邻域内包含了少于 $MinPts$ 个对象, 并且位于其他核心点的邻域内, 则该对象为边界点。

3) 离群点

若某一对象既不是核心点也不是边界点, 则为离群点。

115

对于图 3, 当 DBSCAN 参数中半径为 Eps , $MinPts$ 为 5 时, 由于 A_1 、 A_2 、 A_3 的领域内超过了 5 个对象, 即为核心点; B_1 、 B_2 的领域内对象数量小于 5, 但其位于 A_1 的领域内, 故属于边界点; 而 C_1 、 C_2 既不属于核心点也不属于边界点, 则为离群点。

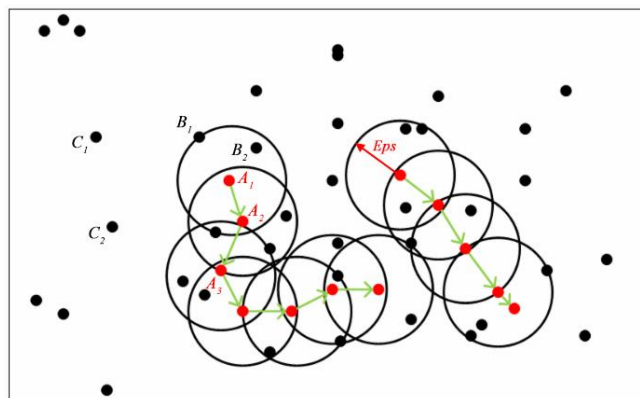
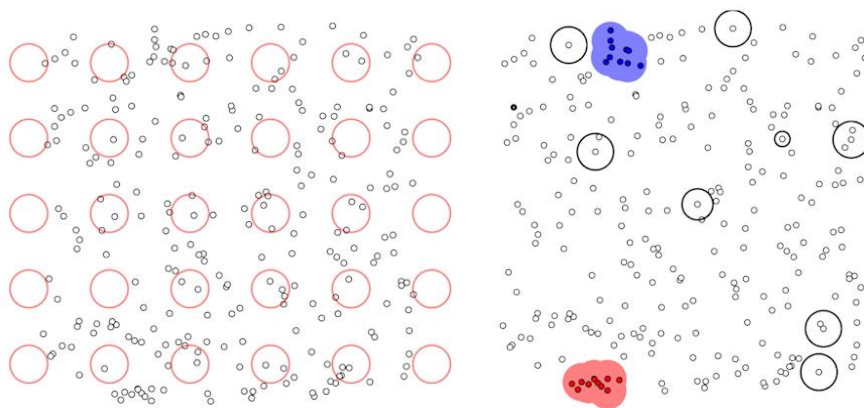


图 2 DBSCAN 算法示例图

Fig. 2 DBSCAN Algorithm Example

对于一个动态聚类过程，设置 Eps 为 1， $MinPts$ 为 5 时，其初始状态如图 4-a)；当聚类过程开始，计算机将随机从多点开始寻找满足条件的核心点、边界点及其领域，如图 4-b)；随着过程的不断进行，最终得到如图 4-c)的由核心点与边界点组成的 8 个簇类。



a) 初始状态

b) 聚类过程

c) 聚类结果

图 3 DBSCAN 聚类过程图

Fig.3 DBSCAN Clustering Process

3 结果与分析

3.1 聚集方向与聚集度

将成都市小学的经纬度代入标准差椭圆的公式，并利用 ArcGIS 软件进行计算和可视化分析，以此得到成都市重点小学 的聚集度与聚集方向，表 1 与图 5 分别为标准差椭圆的计算结果和可视化图像。

根据表 1 与图 5，在聚集度上，普通小学标准差椭圆面积为 0.0136，重点小学则为 0.0063，约为普通小学面积的一半，这意味着成都市重点小学的空间分布更为集中，聚集度更高，且主要集中在市中心区域。

在聚集方向上，成都市重点小学标准差椭圆的偏转角约 108.68°，主要呈西北—东南向聚集；

进一步的，根据椭圆离心率的定义：

$$e = \frac{\sqrt{\sigma_x^2 - \sigma_y^2}}{\sigma_x}$$

公式 4

本文计算了成都市普通小学与重点小学标准差椭圆的离心率，得到普通小学离心率为 0.63，重点小学为 0.74。表明相较于普通小学，重点小学的分布在空间上更具有方向性，而普通小学在各方向的资源配置则更为均衡。

表 2 成都市小学标准差椭圆基本参数
Tab. 2 Chengdu Primary School Standard Deviation Ellipse Parameters

变量名	项目	值	
		普通小学	重点小学
Shape_Length	周长	0.4179	0.2906
Shape_Area	面积	0.0136	0.0063
CenterX	椭圆中心坐标经度	104.0581	104.0678
CenterY	椭圆中心坐标纬度	30.6623	30.6592
XStdDist	椭圆长轴标准差	0.0745	0.0547
YStdDist	椭圆短轴标准差	0.0579	0.0368
Rotation	偏转角	91.6464	108.6808

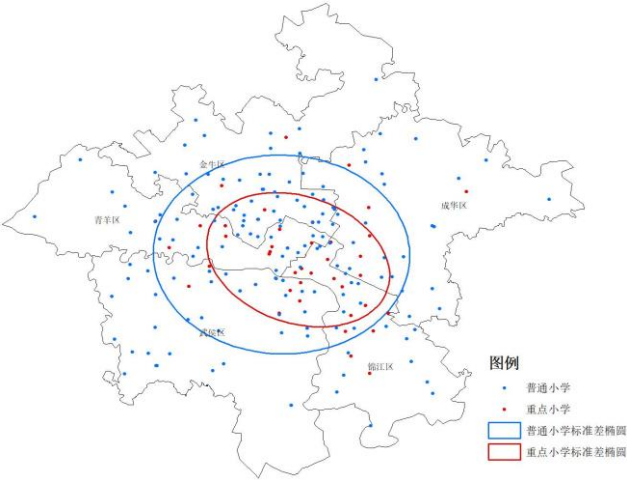


图 4 成都市小学空间分布标准差椭圆
Fig.4 Chengdu Primary School Standard Deviation Ellipse

3.2 聚集区域

DBSCAN 算法无需提前设置聚类数量，但需设置定义密度时的初始参数邻域半径 Eps ，以及邻域内所包括的最小对象数量 $MinPts$ 。在本研究中，邻域半径 Eps 代表重点小学聚集区的最大半径， $MinPts$ 则代表该区域至少有多少个重点小学才被认为是聚集区。这两个参数通常需要凭经验设定，并进行多次估计以选择合适的参数值，本文最终设定的邻域半径 Eps 为 1km，最小对象数量 $MinPts$ 为 3。

本文基于 Python 语言编写 DBSCAN 算法对成都市重点小学的空间位置进行聚类，得到了两个聚集区域：

聚集区 A：

表 3 聚集区 A 中的学校
Tab. 3 Schools in Cluster A

学校	经纬度
成都市泡桐树小学	104.052692, 30.666978
成都市实验小学（本部）	104.061803,30.66615
成都市少城小学	104.051619,30.663991
成都市胜西小学	104.051189,30.66272

聚集区 B：

表 4 聚集区 B 中的学校
Tab. 4 Schools in Cluster B

学校	经纬度
成都师范附属小学（成师附小）	104.070129,30.654197
成都市盐道街小学	104.066171,30.651735
成都龙王庙正街小学	104.075477,30.65151
成都市龙江路小学（南区）	104.069482,30.646044
四川师范大学附属小学（川师附小）	104.063386,30.64156
成都市磨子桥小学	104.0689169,30.635454

将两个聚集区绘制在图上，得到成都市重点小学聚集区的分布图。根据图 6，两个资源聚集区位置相近，且都位于市中心区域，由此可见，成都市优质资源的确出现了明显的配置不均现象。

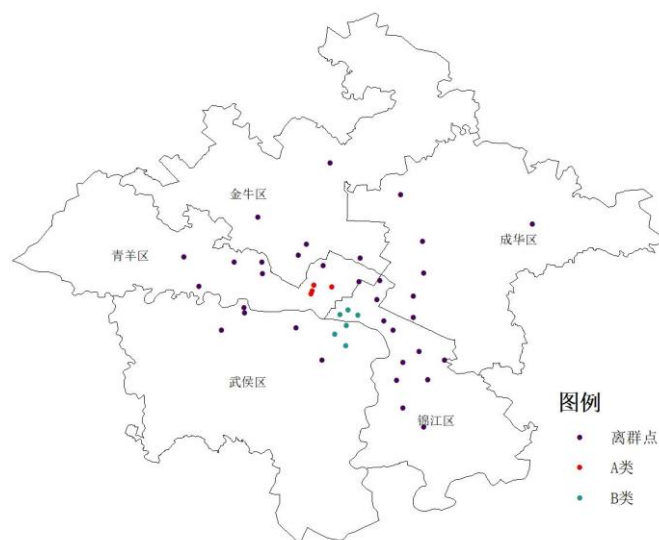


图 5 成都市重点小学聚集区

Fig. 5 Cluster Distribution Map of Key Primary Schools in Chengdu

4 结论

本文首次将地理统计学方法与机器学习算法应用于教育资源的聚集性研究中,并对成都市的重点小学进行了实证研究。

本研究应用标准差椭圆法,发现成都市重点小学主要呈西北—东南向聚集,且聚集度和方向性都远高于普通小学;通过 DBSCAN 聚类算法得到了包括四所小学在内的聚集区域 A,和包括六所小学的聚集区域 B,在我国“就近入学”政策的背景下,这两个聚集区域的家庭将通过购买房产获得巨大的优质教育资源竞争优势,由此给教育公平的实现甚至房地产市场的平稳发展都带来了极大的冲击。

总体而言,尽管成都市的基础教育资源已得到了较为全面的发展,但优质资源仍处于一种聚集的而非均衡的分布状态。如何更好地进行空间资源配置,以及各项教育政策的实际执行情况,亟待政府的进一步思考。

[参考文献] (References)

- [1] 褚宏启. 关于教育公平的几个基本理论问题[J]. 中国教育学报, 2006(12): 1-4.
- [2] 罗燕. 教育产业化的制度分析--新制度主义社会学的视角[J]. 教育与经济, 2006(01): 46-50.
- [3] 迟景明, 任祺. 基于赫芬达尔-赫希曼指数的我国高校创新要素集聚度研究[J]. 大连理工大学学报(社会科学版), 2016, 37(04): 5-9.
- [4] 迟景明, 任祺. 基于赫芬达尔-赫希曼指数的我国高校创新要素集聚度研究[J]. 大连理工大学学报(社会科学版), 2016, 37(04): 5-9.
- [5] 王辉, 延军平, 彭邦文, 等. 中国中等职业教育空间集聚及其影响因素[J]. 热带地理, 2020, 40(03): 525-538.
- [6] 刘善槐, 王爽. 我国义务教育资源空间布局优化研究[J]. 教育研究, 2019, 40(12): 79-87.
- [7] 罗华玲. 云南省义务教育资源空间配置研究[D]. 云南: 云南师范大学, 2018.
- [8] 袁海红, 张华, 曾洪勇. 产业集聚的测度及其动态变化--基于北京企业微观数据的研究[J]. 中国工业经济, 2014(09): 38-50.
- [9] 丁鹏飞. 基于 DO 指数的上海市制造业空间集聚测度研究[J]. 经济论坛, 2019(01): 41-47.
- [10] 刘华军, 王耀辉, 雷名雨. 中国战略性新兴产业的空间集聚及其演变[J]. 数量经济技术经济研究, 2019, 36(07): 99-116.

[11] Govindaraju P , Achter S , Ponsignon T , et al. Finding Demand Patterns in Supply Chain Planning - Two Clustering Approaches for Semiconductor Production Industry[J]. atp magazin, 2018, 60(8):54.

205 [12] 马小宾, 侯国林, 李莉, 等. 基于 DBSCAN 算法的民宿集群识别、分布格局及影响因素--以南京市为例 [J]. 人文地理, 2021, 36(01): 84-93.

[13] Lefever D W . Measuring Geographic Concentration by Means of the Standard Deviatonal Ellipse[J]. American Journal of Sociology, 1926, 32(1):88-94.

[14] 单良, 张涛. 中国产业结构与就业结构协调性时空演变研究[J]. 中国人口科学, 2018(02): 39-49+127.

210 [15] 李细归, 吴清, 刘大均. 武汉城市圈农业旅游目的地空间分布及影响因素[J]. 热带地理, 2014, 34(03): 422-428.

[16] 张杰, 唐根年. 浙江省制造业空间分异格局及其影响因素[J]. 地理科学, 2018, 38(07): 1107-1117.

[17] 赵璐, 赵作权. 基于特征椭圆的中国经济空间分异研究[J]. 地理科学, 2014, 34(08): 979-986.