

doi: 10. 6046/gtzyyg. 2020. 04. 23

引用格式: 张莉,谢亚楠,屈辰阳,等. 基于 K - Means 城市分类算法的夜光遥感电力消费估算[J]. 国土资源遥感,2020,32(4):182 - 189. (Zhang L,Xie Y N,Qu C Y,et al. Estimation of electric power consumption using nighttime light remote sensing data based on K - Means city classification algorithm [J]. Remote Sensing for Land and Resources ,2020,32(4):182 - 189.)

基于 K - Means 城市分类算法的夜光遥感电力消费估算

张 莉^{1,2}, 谢亚楠¹, 屈辰阳^{2,3}, 汪鸣泉^{2,3}, 常 征^{2,3,4}, 王茂华^{2,3}

(1. 上海大学特种光纤与光接入网重点实验室,特种光纤与先进通信国际合作联合实验室,上海先进通信与数据科学研究院,上海 201210; 2. 中国科学院上海高等研究院,上海碳数据与碳评估中心,上海 201210; 3. 中国科学院低碳转化科学与工程重点实验室,上海 201210; 4. 中国科学院洁净能源创新院,大连 116023)

摘要: 为了减小利用夜间灯光影像估算城市电力消费量时的误差,需要考虑样本地区的发展状况,在估算之前对样本进行分类。选取 2015 年中国大陆 263 个地级市的 NPP - VIIRS 夜间灯光数据对城市电力消费量进行估算。提出了基于灯光结构而非传统统计数据数据的 K - Means 城市分类算法。利用该方法将样本分为 5 类并估算电力消费量,与其他分类方法的估算结果相比可知:该方法估算值的平均相对误差和均方根误差分别为 32. 02% 和 57. 04,较不分类时分别减小 25 和 3. 39 百分点;估算中的高精度城市比例为 53. 99%,较不分类时增加了 13. 59 百分点,且为所有方法中的最高比例;相较不分类时的估算结果,有 152 个城市的估算误差有所降低。该方法性能与其他分类方法的最优性能相似。

关键词: NPP/VIIRS; 电力消费量; 城市分类; K - Means 算法

中图法分类号: TP 79 **文献标志码:** A **文章编号:** 1001 - 070X(2020)04 - 0182 - 08

0 引言

电力消费量(electric power consumption, EPC)是衡量一个地区电力发展情况的重要社会经济参数,它被誉为是经济发展的先行指标^[1]。快速准确地估算地区 EPC,能够为经济发展和能源供需政策制定与调整提供及时地依据与支撑^[2]。

美国新一代 Suomi 地对地观测卫星(National Polar - Orbiting Partnership, NPP)搭载的可见光红外成像辐射仪(Visible Infrared Imaging Radiometer Suite, VIIRS)能够记录城市的低强度夜间灯光,可以作为估算城市 EPC 等社会经济参数的理想数据源^[3-5]。在不同尺度的 EPC 估算研究中发现,经济发达地区的估算精度高于经济欠发达地区^[6-9]。为减小估算误差,需要考虑到样本地区的发展状况,在估算前对样本城市进行分类。

Xie 等^[10]在估算国家 EPC 的研究中,将样本国家按地理位置划分为 7 个区域,并建立 EPC 对数估算模型,发现除中东和非洲地区,其他区域的估算模型 R^2 较不分类时的模型 R^2 提高了 4% 以上;李通等^[11]在进行中国城市的 EPC 估算时,按照经济分区将样本划分为东部、中部和西部城市,估算结果显示,除西藏拉萨市以外的 30 个省会城市(含直辖市)的平均相对误差为 33. 49%。除传统分类方法之外,施建刚等^[12]提出波士顿矩阵法,可以同时考虑相对指标与绝对指标,能客观地反映城市周边环境与城市发展水平;李熙等^[13]基于波士顿矩阵,采用相对城镇人口比例和城镇人口比例增长率将城市划分为 4 种类型,其 EPC 估算结果的平均相对误差为 34. 04%,较不分类时减小 18. 3 百分点;Li 等^[14]基于分部门就业数据,使用聚类分析将中国城市按功能分为“服务型城市”、“工业型城市”和“技术及教育型城市”,分类后的估算模型 R^2 较不分类

收稿日期: 2020 - 01 - 06; 修订日期: 2020 - 04 - 03

基金项目: 中国科学院洁净能源创新研究院合作基金项目“变革性洁净能源关键技术对我国碳排放达峰目标的贡献及其减排路径研究”(编号: DNL180101)、国家自然科学基金项目“面向低碳城市规划的碳排放评价方法研究”(编号: 51778601)、国家重点研发计划项目“行业碳排放核算与效益成本评估模型研究”(编号: 2016YFA062603)、国家重点研发计划项目“基于碳卫星数据的全球大气中 CO₂ 浓度估算与预测模型研究”(编号: 2016YFA062602)和国家重点研发计划项目“世界主要国家碳排放因子研究”(编号: 2017YFA065300)共同资助。

第一作者: 张 莉(1995 -),女,硕士研究生,研究方向为卫星数据处理和城市电力消费模拟。Email: zhangli02@sari. ac. cn。

通信作者: 王茂华(1977 -),男,硕士,高级工程师,研究方向为碳数据与碳评估。Email: wangmh@sari. ac. cn。

时提高 6.4%。

在 EPC 估算时,前人研究中的样本分类方法大多依赖于统计数据。且对分类前后的模型对比没有统一标准。由于数据年份和研究区域的差异,无法进行各方法的横向对比。本研究选择中国大陆区域 263 个城市作为研究对象,提出不依赖统计数据的基于灯光结构的 K - Means 城市分类算法,对分类后的样本城市进行 EPC 估算。并与传统城市分类方法的 EPC 估算结果进行了对比。

1 数据源概况

本研究中使用的数据如表 1 所示,其中,卫星遥感数据为 2015 年 NPP/VIIRS 卫星夜间灯光影像。统计数据包含 2014 年和 2015 年人口、市辖区 GDP、产业结构等社会经济数据。

表 1 数据来源及简述
Tab.1 Data source and description

数据名称	数据描述	数据来源
NPP - VIIRS 卫星夜间灯光影像	年数据,空间分辨率为 15" (弧度)	NOAA - NGDC (http://ngdc.noaa.gov/viir)
全社会用电量	市辖区,亿 kWh	《中国城市统计年鉴》
年末总人口	市辖区,全市,万人	《中国城市统计年鉴》
地区 GDP	市辖区,全市,万元	《中国城市统计年鉴》
第一、二、三产业 GDP 比例	市辖区,全市,%	《中国城市统计年鉴》
常住人口城镇化率	各省份,城市,%	各省市统计年鉴
2015 年中国地级市行政区划边界文件	.shp 格式文件	中国科学院资源环境数据云平台 (http://www.resdc.cn/data.aspx?DATAID=201)

1.1 NPP - VIIRS 夜间灯光遥感数据

NOAA 官网提供了 2015 年的 NPP - VIIRS 年度夜间灯光影像,该影像去除了月度数据中的背景噪声。在没有夜间灯光的地方,像素值为 0。且不存在饱和现象^[15-16]。但仍有一些异常值点需要校正^[17]。参考李雪萍等^[18]、Shi 等^[19]的研究,假设中国大陆地区最大亮度出现在北京、上海和广州 3 个城市。计算北京、上海、广州地区的最大亮度值 $DN_{max} = 264.468$,如果有 $DN > DN_{max}$,将其置为 DN_{max} 。

1.2 统计数据

本研究所使用到的统计数据除常住人口城镇化

率外,均来源于《2016 年中国城市统计年鉴》,该年鉴中记录了 2015 年地级市的全社会用电量、年末总人口和地区 GDP 等统计数据。除全社会用电量数据为市辖区数据外,其他数据均含市辖区和全市数据。

常住人口城镇化率数据来源于各省份和城市统计年鉴,若没有记录,则通过 2015 年末城镇人口除以年末常住人口获得。

1.3 研究区域

到 2015 年底,中国内地共包含地级市 290 个。由于西藏、新疆等地数据缺失严重,本文选取《中国城市统计年鉴》中除西藏、新疆和港澳台以外的其他省份中数据完整的 263 个地级市进行研究。

2 研究方法

2.1 EPC 估算方法

统计每个城市的夜间灯光总量 (total nighttime light, NTL) 为:

$$NTL_i = \sum x_p \text{ , } \tag{1}$$

式中: x_p 为 p 位置的像元值; i 为第 i 个城市。

2.2 城市分类方法

2.2.1 传统城市分类方法

传统的城市分类方法可以按照经济分区将城市划分为东部(82)、中部(76)、西部(81)和东北(24)城市;也可以按照地理分区将城市划分为东北(27)、华北(26)、华东(77)、华南(34)、华中(38)、西北(31)和西南(30)城市;另外,按照《中国城市年鉴》的统计,以 2015 年市区平均人口标准,可以将城市划分为 400 万及以上人口(13)、[200,400) 万人口(12)、[100,200) 万人口(112)、[50,100) 万人口(82)和 50 万以下人口城市(44)。括号中为研究区域内该类型的城市个数。

2.2.2 基于波士顿矩阵的分类方法

波士顿矩阵(图 1)是统计学和经济学中常用的分类方法。李熙等^[13]的研究中,采用“城镇人口增长率”和“相对城镇人口比例”分别作为“绝对指标”

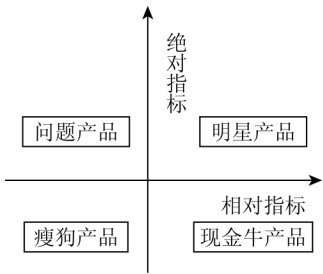


图 1 波士顿矩阵
Fig.1 Boston matrix

和“相对指标”,构建波士顿矩阵,将城市分为 4 类进行讨论。但常住城镇人口比例这一参数并未有全国统一的官方统计数据。且城市的 EPC 与城市的人口、经济和产业结构等都有一定联系^[20-21]。因此,我们提出以“市辖区人口”、“市辖区 GDP”和“市辖区第三产业 GDP”来代替“城镇人口”构建波士顿矩阵。分类后每一类的城市个数如表 2 所示。

表 2 基于波士顿矩阵的城市分类结果
Tab. 2 City classification results based on Boston matrix (个)

绝对指标(阈值)	相对指标 (阈值=1)	第一类 城市	第二类 城市	第三类 城市	第四类 城市
城镇人口增长率(0.02)	相对城镇人口比例	26	81	121	35
市辖区人口增长率(0.04)	相对市辖区人口比例	15	14	93	141
市辖区 GDP 增长率(0.02)	相对市辖区 GDP 比例	15	57	51	140
市辖区第三产业 GDP 增长率(0.01)	相对市辖区第三产业 GDP 比例	15	49	56	143

2.2.3 基于聚类算法的城市分类方法

机器学习常用聚类算法有 K 均值聚类(K-Means)^[22]、K 中心点聚类(K-Medoids)、模糊 C 均值聚类(FCM)和高斯混合聚类(GMM)几种^[23]。K-Means 算法是最常用的聚类算法之一,其算法简单、收敛速度快。K-Medoids 与 K-Means 的不同之处在于其选用类中位置居于最中心的对象来作为下一次迭代的新聚类中心。FCM 算法采用模糊划分,使得每个数据点用[0,1]区间的隶属度来确定其属于各个类的程度。GMM 算法采用概率模型对原型进行刻画,在每个维度用均值和标准差来描述簇的形状^[24]。

通常使用紧密性(compactness, CP)和分离性(separation, SP)来评价以上几种分类方法。前者描述各点到聚类中心的平均距离,越小说明同一类别紧密度越高,效果越好^[25];后者描述各聚类中心两两之间的平均距离,越大说明不同类间隔性越高,效果越好^[26]。2 种指标的计算方法分别为:

$$\overline{C} = \frac{1}{K} \sum_{i=1}^K \overline{C_i} \quad (2)$$

$$\overline{C_i} = \frac{1}{|\Omega_i|} \sum_{x_m \in \Omega_i} \|x_m - w_i\| \quad (3)$$

$$\overline{S} = \frac{2}{K^2 - K} \sum_{i=1}^K \sum_{j=i+1}^K \|w_i - w_j\| \quad (4)$$

式中: $\overline{C_i}$ 为第 i 个聚类集合的 CP 值; Ω_i 为第 i 个聚

类集合; w_i 和 w_j 分别为第 i 个和第 j 个聚类中心; x_m 为集合 Ω_i 中的第 m 个样本; K 为聚类中心个数; $\overline{C}$ 和 $\overline{S}$ 分别为聚类结果的 CP 和 SP 值。

2.2.4 估算结果的评估标准

采用相对误差、均方根误差来评估估算结果^[27]。相对误差指测量的绝对误差与被测量的真值之比,某个样本城市的估算结果相对误差 R 计算公式为:

$$R = \frac{|E_e - E_r|}{E_r} \times 100\% \quad (5)$$

式中: E_e 和 E_r 分别表示某样本城市 EPC 的估算值和真实值。

某区域范围内的估算结果平均相对误差计算公式为:

$$\overline{R} = \frac{|\sum E_e - \sum E_r|}{\sum E_r} \times 100\% \quad (6)$$

式中: $\overline{R}$ 为该区域内的估算结果平均相对误差; $\sum E_e$ 和 $\sum E_r$ 分别为该区域内所有样本城市的 EPC 估算值和真实值之和。

均方误差是估算结果与真实值平方差的期望值,均方根误差 RMSE 是均方误差的算术平方根,可以评价数据的变化程度, RMSE 越小,说明模型描述样本数据具有更好的精确度。计算公式为:

$$RMSE = \sqrt{\frac{1}{n} \sum (E_e - E_r)^2} \quad (7)$$

式中 n 为样本个数。

3 结果与分析

3.1 聚类算法城市分类

3.1.1 聚类算法比较

城市夜间灯光的亮度分布也携带一些城镇化信息^[28-29],统计研究区域灯光亮度值发现,最小亮度值 $DN_{\min} = 0.497$ 且约 90% 以上的像素满足 $DN < 40$ 。因此计算某样本城市的特征值为:

$$P_i = \frac{\text{num}(DN_{(e^{-3+0.1i}, e^{-2+0.1i})})}{\text{num}(DN > 0)} \quad (8)$$

式中: P_i 为样本城市的第 i 个特征值,每个城市拥有 $P_0 \sim P_{84}$ 共 85 个特征值; $\text{num}(DN_{(e^{-3+0.1i}, e^{-2+0.1i})})$ 表示亮度值 DN 在 $(e^{-3+0.1i}, e^{-2+0.1i})$ 范围内的像素个数; $\text{num}(DN > 0)$ 表示亮度值大于 0 的像素个数。

每个城市的特征值满足 $P_0 + P_1 + \cdots + P_{84} = 1$ 。

设聚类中心个数 $K = 4$ 时,使用以上 4 种聚类算法对样本进行了 10 次聚类,其 CP 和 SP 指标变化如表 3 和图 2 所示。综合考虑 2 个指标及稳定性,选取 K - Means 聚类算法对城市进行分类。

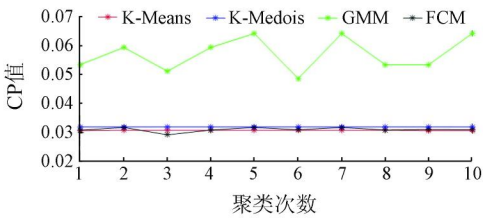
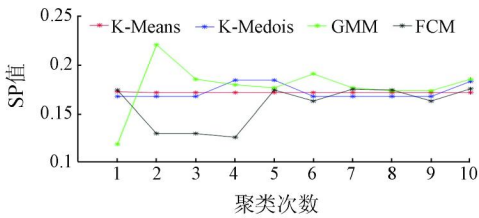


图 2 10 次聚类的 CP 和 SP 值
Fig. 2 CP and SP values of 10 times clustering

表 3 CP 与 SP 的均值与标准差				
Tab. 3 Mean and standard deviation of CP and SP				
聚类方法	CP 均值	CP 标准差	SP 均值	SP 标准差
K - Means	0.031	0.000 0	0.18	0.000 4
K - Medois	0.032	0.000 0	0.17	0.007 9
FCM	0.031	0.000 1	0.17	0.025 2
GMM	0.057	0.000 7	0.16	0.021 3



3.1.2 聚类中心个数

选取 K - Means 聚类算法进行聚类。随着聚类中心个数 K 的增加,计算其聚类方差,即

$$\sigma^2 = \frac{\sum \|x_m - w_i\|^2}{n}。$$
 (9)

结果如图 3 所示。

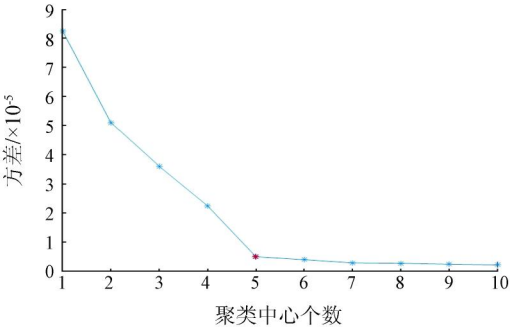


图 3 聚类中心方差随个数变化
Fig. 3 Variance of clustering center varies with the number
3.1.3 聚类结果

图 3 中,聚类方差在 $K = 5$ 时有 1 个拐点,所以设聚类中心个数 $K = 5$,K - Means 聚类后聚类中心特征分布如图 4 所示。各类别城市的灯光分布特征分析及代表性城市如表 4 所示。

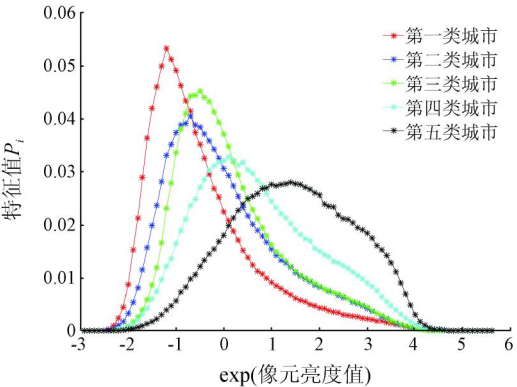


图 4 聚类中心特征分布
Fig. 4 Distribution features of cluster central

表 4 K - Means 城市分类结果			
Tab. 4 Result of K - means city classification method			
类别	个数	灯光分布特征	代表性城市
第一类城市	44	像元亮度值集中分布在 e^{-1} 以下,城区灯光几乎很少	张掖、平凉、绥化、黑河
第二类城市	65	像元亮度值集中分布在 $e^{-1} \sim e^1$ 之间,城区灯光基本明朗	鞍山、吉林、黄山、西宁
第三类城市	91	像元亮度值集中分布在 $e^{-1} \sim e^0$ 之间,城区灯光基本明朗,和周边非城区的亮度差异明显	石家庄、徐州、北海、兰州
第四类城市	45	像元亮度值集中分布在 $e^{-1} \sim e^2$ 之间,城区高亮度灯光较多,部分低亮度灯光分布于郊区	北京、天津、合肥、重庆
第五类城市	18	数量较少,城市化基本完成,几乎没有低亮度地区	上海、南京、无锡、舟山、福州、成都、深圳、珠海

3.2 EPC 估算结果

估算结果如图 5 所示。在地级市尺度上,263 个城市的 NTL 与 EPC 相关性较高($R^2 = 0.795 9$)。建立 EPC 线性估算模型,即

$$EPC_i = 0.003 NTL_i + 31.91 \quad , \quad (10)$$

式中: EPC_i 为城市 i 的 EPC 估算值。

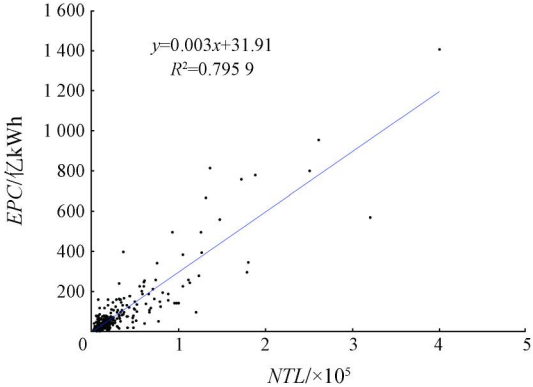


图 5 城市 NTL 与 EPC 的相关关系
Fig. 5 Correlation between the city's NTL and EPC

基于 K - Means 算法对城市进行分类后,图 6 为每一类型城市 NTL 与 EPC 的线性相关关系,可以看出,第五类城市的 NTL 与 EPC 相关性最高。 R^2 达到了 0.924。其次是第一类城市和第四类城市, R^2 分别达到了 0.851 6 和 0.848 1,均高于不分类时的 R^2 (0.795 9)。第二类城市和第三类城市的 NTL 与 EPC 相关性较差。建立分类后的 EPC 线性估算

模型,即

$$EPC_i = \begin{cases} 0.003\ 3\ NTL_i + 10.381\ 8, & \text{第一类城市} \\ 0.002\ 9\ NTL_i + 20.479\ 6, & \text{第二类城市} \\ 0.003\ 8\ NTL_i + 24.897\ 1, & \text{第三类城市} \\ 0.003\ 2\ NTL_i + 30.034\ 5, & \text{第四类城市} \\ 0.002\ 9\ NTL_i + 55.043\ 1, & \text{第五类城市} \end{cases} \quad (11)$$

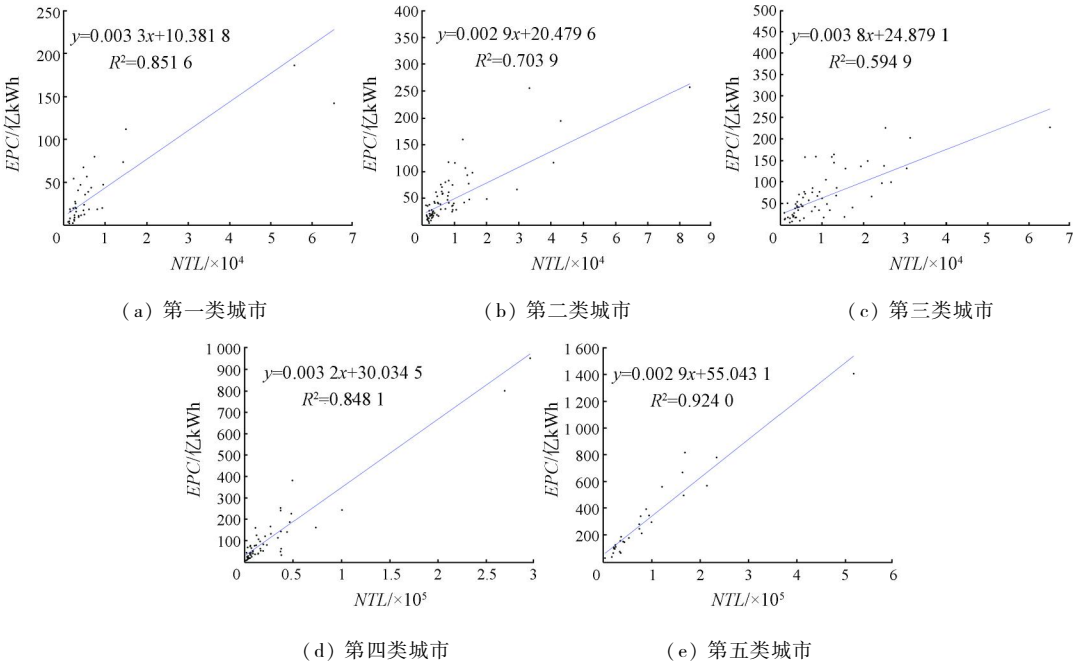


图 6 5 类城市的 NTL 与 EPC 的相关性

Fig. 6 Correlation between NTL and EPC in five types of cities

3.3 与不分类时的估算结果对比

如表 5 所示,K - Means 分类之后,第五类城市的估算误差最小,29 个城市的平均相对误差(R) 为 15.22% ; 其后依次是第四类、第三类、第一类和第二类城市。所有城市的 R 为 32.02% ,较不分类时减小 25 百分点。将不分类时的 EPC 估算结果也按照 K - Means 算法的城市分类进行分类,可以看到,除第二类城市外,K - Means 分类后估算结果的

R 较分类之前都有减小。且 2 种方法获得的第二类城市估算结果 R 相差不大。第一类城市和第三类城市的 R 在 K - Means 分类之后分别减小了 23.38 百分点和 12.7 百分点。从模拟结果来看,这 2 类城市中分布在黄土高原等中西部地区的城市相对误差(R)有明显下降。

定义某城市估算结果相对误差 R 满足 $R \leq 25\%$ 为高精度城市, $25\% < R \leq 50\%$ 为中精度城市, $R > 50\%$ 为低精度城市,统计不同精度的城市比例。如表 6 所示,K - Means 分类之后高精度城市占 53.99% ,较不分类时增加了 13.59 百分点。中精度和低精度城市比例分别为 26.99% 和 19.01% ,

表 5 K - Means 城市分类法与不分类时的 EPC 估算结果对比

Tab. 5 Comparison between the EPC estimation results in K - means city classification method and no classification			
城市类型	城市个数	R / %	
		不分类	K - Means 分类
第一类城市	44	57.49	44.05
第二类城市	65	44.64	45.17
第三类城市	91	48.10	41.96
第四类城市	45	35.17	33.05
第五类城市	18	18.50	15.22
合计/平均值	263	42.37	32.02

表 6 K - Means 城市分类法与不分类时的 EPC 估算精度对比

Tab. 6 Comparison between EPC estimation accuracy in K - Means city classification method and no classification (%)			
估算方法	高精度	中精度	低精度
不分类	47.53	29.28	23.19
K - Means 分类	53.99	26.99	19.01

较不分类时也有所减小。

将样本城市分类前后的 R 差值记作 D ,即

$$D_i = R_e - R_r \quad , \quad (12)$$

式中 R_e 和 R_r 分别为 K - Means 分类后和不分类时城市 i 的估算结果相对误差,结果见表 7。

表 7 K - Means 城市分类后与不分类时的 EPC 估算误差对比

Tab.7 Comparison between EPC estimation error in K - Means city classification method and no classification

D_i	(- ■, - 0.25]	(- 0.25, 0]	(0, 0.25]	(0.25, + ■)
比例/%	11.79	46.01	34.98	7.22

如表 7 所示,经 K - Means 分类后,有 57.80% 的城市 R 较不分类时有所降低($D < 0$),其中,11.79% 的城市 R 呈现大幅降低($D \leqslant - 0.25$);46.01% 呈现小幅降低($- 0.25 < D < 0$)。同理,有 34.98% 的城市 R 较不分类时呈现小幅增加($0 < D \leqslant 0.25$);7.22% 的城市 R 呈现大幅增加($D > 0.25$)。

$D > 0.25$ 的城市包括舟山市、丽水市、宁德市、景德镇市、鹰潭市、汕尾市、云浮市、百色市、广元市、内江市、雅安市、巴中市、毕节市、昭通市、咸阳市、嘉

峪关市、张掖市、长治市和齐齐哈尔市。这 19 个城市的 EPC 都在 40 亿 KWh 以下,属于 EPC 较少的城市,在估算时有可能是因为受到 EPC 较大的城市影响而产生了较大误差。如舟山市,因其特殊的岛屿地形,大部分地区均为高亮度灯光,在分类时与上海等城市一起归为第五类城市,其 EPC 的估算受到上海等 EPC 较大的城市影响而产生较大误差。

3.4 与传统分类方法的估算结果对比

如表 8 所示,对比本文方法与 2.2 节中提到的

其他分类方法分类后的 EPC 估算结果。以 $\overline{R}$ 为评价指标,所有方法中,最优分类方法是按照人口规模进行划分。K - Means 城市分类算法的这一性能仅次于最优方法。以 $RMSE$ 为评价指标,最优分类方法是按照地理分区进行划分。K - Means 城市分类算法的这一性能仅次于最优方法和按照人口规模划分时。以估算精度来看,高精度城市占比最高的是 K - Means 城市分类算法(53.99%)。而从分类前后 R 变化来看, $D > 0$,即 R 减小的城市比例最高的是波士顿矩阵 - 市辖区 GDP 比例划分方法(61.21%),而 K - Means 城市分类算法的这一比例(57.80%)仅次于 2 种基于波士顿矩阵的分类方法。

表 8 K - Means 城市分类方法与传统分类方法的估算结果对比

Tab.8 Comparison of EPC estimation results between K - Means city classification method and traditional classification method

分类方法	$\overline{R}$ /%	$RMSE$	高精度/%	中精度/%	低精度/%	$D > 0$ /%
不分类	42.37	59.04	47.53	29.28	23.19	-
经济分区	38.37	63.23	50.19	28.13	21.68	50.57
地理分区	34.05	53.16	52.47	28.13	19.40	52.09
人口规模	31.06	53.76	52.09	28.51	19.40	53.99
波士顿矩阵 - 城镇人口比例	32.62	57.61	51.34	30.41	18.25	56.27
波士顿矩阵 - 市辖区人口比例	32.03	59.20	53.99	27.37	18.64	58.55
波士顿矩阵 - 市辖区 GDP 比例	31.66	59.78	49.42	32.31	18.27	61.21
波士顿矩阵 - 市辖区第三产业 GDP 比例	32.13	58.09	50.95	31.56	17.49	55.13
K - Means 算法	32.02	57.04	53.99	26.99	19.01	57.80

4 结论

本研究提出基于灯光结构的 K - Means 城市分类算法,将城市划分为 5 类,从第一类到第五类,高亮度灯光逐渐增加,在一定程度上可以理解为城镇化水平更高。

使用该方法分类后进行 EPC 估算,估算结果的 $\overline{R}$ 为 32.02% ,较不分类时减小 25 百分点。高精度城市比例达 53.99% ,较不分类时增加 13.59 百分点。有 152 个城市的估算 R 较不分类时有所降低。其中 31 个城市降低了 0.25 个百分点以上。而仅有

19 个 EPC 过小的城市估算 R 较不分类时提高 0.25 个百分点以上。

与前人已提出的分类方法相比,从估算结果的 $\overline{R}$ 和 $RMSE$ 性能来看,本文方法与其他方法相比有较大优势。与其他方法中的最优方法性能相差不大。从高精度城市比例来看,本文方法为所有方法中最优方法。从分类前后 R 变化来看,本文方法仅次于 2 种基于波士顿矩阵的分类方法。

与前人方法相比,本文提出的 K - Means 城市分类算法优点在于,仅依靠灯光本身特征即可进行分类。而不依赖于统计数据。就能获得与前人方法相似甚至更优的估算结果。

本文方法还需改进的地方在于,仅依靠灯光的亮度分布特征进行分类时忽略了灯光的绝对亮度指标,造成如舟山这样的城市因为其特殊的地理环境而被划分到第五类城市。

参考文献(References):

[1] 肖宏伟. 基于卫星灯光数据的我国省域电力消费模拟研究[J]. 中国能源, 2017, 39(1): 6-10, 14.
Xiao H W. Simulation study of provincial electricity consumption based on satellite lighting data[J]. Energy of China, 2017, 39(1): 6-10, 14.

[2] 袁家海, 丁伟, 胡兆光, 等. 电力消费与中国经济发展的协整与波动分析[J]. 电网技术, 2006, 30(9): 10-14.
Yuan J H, Ding W, Hu Z G, et al. Analysis on cointegration and co-movement of electricity consumption and economic growth in China[J]. Power System Technology, 2006, 30(9): 10-14.

[3] 李峰, 米晓楠, 刘军, 等. 基于 NPP-VIIRS 夜间灯光数据的北京市 GDP 空间化方法[J]. 国土资源遥感, 2016, 28(3): 19-24. doi:10.6046/gtzyyg.2016.03.04.
Li F, Mi X N, Liu J, et al. Spatialization of GDP in Beijing using NPP-VIIRS data[J]. Remote Sensing for Land and Resources, 2016, 28(3): 19-24. doi:10.6046/gtzyyg.2016.03.04.

[4] 沈丹, 周亮, 王培安. 基于夜间灯光数据的六盘山连片特困区贫困度识别[J]. 国土资源遥感, 2019, 31(2): 157-163. doi:10.6046/gtzyyg.2019.02.22.
Shen D, Zhou L, Wang P A. Identification of poverty based on nighttime light remote sensing data: A case study on contiguous special poverty-stricken areas in Liupan Mountains[J]. Remote Sensing for Land and Resources, 2019, 31(2): 157-163. doi:10.6046/gtzyyg.2019.02.22.

[5] 肖东升, 杨松. 基于夜间灯光数据的人口空间分布研究综述[J]. 国土资源遥感, 2019, 31(3): 10-19. doi:10.6046/gtzyyg.2019.03.02.
Xiao D S, Yang S. A review of population spatial distribution based on nighttime light data[J]. Remote Sensing for Land and Resources, 2019, 31(3): 10-19. doi:10.6046/gtzyyg.2019.03.02.

[6] He C, Ma Q, Li T, et al. Spatiotemporal dynamics of electric power consumption in Chinese mainland from 1995 to 2008 modeled using DMSP/OLS stable nighttime lights data[J]. Journal of Geographical Sciences, 2014, 22(1): 125-136.

[7] Su Y, Chen X, Wang C, et al. A new method for extracting built-up urban areas using DMSP-OLS nighttime stable lights: A case study in the Pearl River Delta, southern China[J]. Mapping Sciences and Remote Sensing, 2015, 52(2): 218-238.

[8] Cao X, Wang J, Chen J, et al. Spatialization of electricity consumption of China using saturation-corrected DMSP-OLS data[J]. International Journal of Applied Earth Observations and Geoinformation, 2014(28): 193-200.

[9] Zhao N, Ghosh T, Samson E L, et al. Mapping spatio-temporal changes of Chinese electric power consumption using night-time imagery[J]. International Journal of Remote Sensing, 2012, 33(20): 6304-6320.

[10] Xie Y, Weng Q. World energy consumption pattern as revealed by DMSP-OLS nighttime light imagery[J]. GIScience and Remote Sensing, 2016, 53(2): 265-282.

[11] 李通, 何春阳, 杨洋, 等. 1995—2008 年中国大陆电力消费量时空动态[J]. 地理学报, 2011, 66(10): 1403-1412.
Li T, He C Y, Yang Y, et al. Understanding electricity consumption changes in Chinese mainland from 1995 to 2008 by using DMSP/OLS stable nighttime light time series data[J]. Acta Geographica Sinica, 2011, 66(10): 1403-1412.

[12] 施建刚, 裴丽岚. 城市群内城市分级方法比较研究——以成都平原城市群为例[J]. 城市问题, 2009, 173(12): 19-22, 32.
Shi J G, Qiu L L. Comparative study on city classification within urban agglomeration[J]. Urban Problems, 2009, 173(12): 19-22, 32.

[13] 李熙, 薛翔宇. 基于波士顿矩阵的夜光遥感电力消费估算方法[J]. 武汉大学学报(信息科学版), 2018, 43(12): 1994-2002.
Li X, Xue X Y. Estimation method of nighttime light images' electric power consumption based on the Boston matrix[J]. Geomatics and Information Science of Wuhan University, 2018, 43(12): 1994-2002.

[14] Li S, Cheng L, Liu X, et al. City type-oriented modeling electric power consumption in China using NPP-VIIRS nighttime stable light data[J]. Energy, 2019(189): 116040.

[15] Henderson J, Storeygard A, Weil D. Measuring economic growth from outer space[J]. American Economic Review, 2012, 102(2): 994-1028.

[16] 李欣欣, 王利, 何飞. 基于 NPP/VIIRS 夜间灯光数据和土地利用数据的人口分布图绘制——以大连金普新区为例[J]. 遥感信息, 2018, 33(4): 35-41.
Li X X, Wang L, He F. Population distribution map drawing based on NPP/VIIRS nighttime light remote sensing images and land use data: A case study in Jinpu new area[J]. Remote Sensing Information, 2018, 33(4): 35-41.

[17] Jing X, Shao X, Cao C, et al. Comparison between the Suomi-NPP day-night band and DMSP-OLS for correlating socio-economic variables at the provincial level in China[J]. Remote Sensing, 2015, 8(1): 17.

[18] 李雪萍, 贡璐. DMSP/OLS 和 VIIRS/DNB 夜间灯光影像的校正及拟合[J]. 测绘通报, 2019(7): 138-146.
Li X P, Gong L. Correction and fitting of night light images of DMSP/OLS and VIIRS/DNB[J]. Bulletin of Surveying and Mapping, 2019(7): 138-146.

[19] Shi K, Chen Y, Yu B, et al. Modeling spatiotemporal CO₂ (carbon dioxide) emission dynamics in China from DMSP-OLS nighttime stable light data using panel data analysis[J]. Applied Energy, 2016(168): 523-533.

[20] Fan Y, Liu L C, Wu G, et al. Analyzing impact factors of CO₂ emissions using the STIRPAT model[J]. Environmental Impact Assessment Review, 2006, 26(4): 377-395.

[21] Zhou Y, Clarke L, Eom J, et al. Modeling the effect of climate change on U.S. state-level buildings energy demands in an integrated assessment framework[J]. Applied Energy, 2014(113): 1077-1088.

[22] Hartigan J A, Wong M A. A K-means clustering algorithm[J]. Journal of Applied Statistics, 2013, 28(1): 100-108.

[23] Bansal N, Blum A, Chawla S, et al. Correlation clustering[J]. Machine Learning, 2004, 56(1-3): 89-113.

[24] Maddah M, Grimson W E L, Warfield S K. Statistical modeling and EM clustering of white matter fiber tracts[C]//Proceedings of the 2006 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Arlington, VA, USA, 2006: 6-9.

[25] Davies D L, Bouldin D W. A cluster separation measure[J]. IEEE Transactions Pattern Analysis and Machine Intelligence, 1979, PA-MI-1(2): 224-227.

[26] Calinski T, Harabasz J. A dendrite method for cluster analysis[J]. Communicertionsin Statistics, 1974, 3(1): 1-27.

[27] Li X, Li D, Xu H, et al. Intercalibration between DMSP/OLS and VIIRS night-time light images to evaluate city light dynamics of Syria's major human settlement during Syrian Civil War[J]. International Journal of Remote Sensing, 2017, 38(21): 5934-5951.

[28] Hu Y, Zhao G, Zhang Q L. Spatial distribution of population data based on nighttime light and LUCC data in the Sichuan - Chongqing region[J]. Journal of Geo - Information Science, 2018, 20(1): 68-78.

[29] Su Y, Chen X, Li Y, et al. China's 19-year city-level carbon emissions of energy consumptions, driving forces and regionalized mitigation guidelines[J]. Renewable and Sustainable Energy Reviews, 2014(35): 231-243.

Estimation of electric power consumption using nighttime light remote sensing data based on K – Means city classification algorithm

ZHANG Li^{1,2}, XIE Yanan¹, QU Chenyang^{2,3}, WANG Mingquan^{2,3}, CHANG Zheng^{2,3,4}, WANG Maohua^{2,3}

(1. Key Laboratory of Specialty Fiber Optics and Optical Access Networks, Joint International Research Laboratory of Specialty Fiber Optics and Advanced Communication, Shanghai Institute of Advanced Communication and Data Science, Shanghai University, Shanghai 201210, China; 2. Shanghai Carbon Data Research Center, Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai 201210, China; 3. CAS Key Laboratory of Low – Coal Conversion Science and Engineering, Shanghai Advanced Research Institute, Shanghai 201210, China; 4. Dalian National Laboratory for Clean Energy, Dalian 116023, China)

Abstract: In order to reduce the error in estimating urban electric power consumption by nighttime light images , it is necessary to consider the development status of sample areas and classify the samples before estimation . In this paper, the NPP – VIIRS nighttime light data from 263 prefecture – level cities in China’s mainland in 2015 were selected to estimate urban electric power consumption . A K – Means city classification method based on light structure rather than traditional statistical data is proposed . The authors used this method to divide the samples into 5 types and estimate the electric power consumption . A comparison of the estimated results with those from other classification methods shows the following regularity : The mean relative error and root mean square error of the estimated results are 32. 02% and 57. 04, decreasing by 25 and 3. 39 percentage points compared with the estimated results without classification respectively . The proportion of high – precision cities in the estimation results is 53. 99% , increasing by 13. 59 percentage points compared with estimated result without classification , and is the highest proportion among values of all methods . Compared with the estimated results without classification , 152 cities have lower estimated errors. The performance of this method is similar to the optimal performance of other classification methods.

Keywords: NPP/VIIRS nighttime light data ; electric power consumption ; city classification ; K – Means algorithm
(责任编辑: 李 瑜)