

基于遗传算法的景区公共设施和配送中心选址——以南京市为例¹

张鑫 李柯宜 赵越

东南大学成贤学院

摘要：2022 年 6 月，国家推出《“十四五”国民健康规划》，提出扎实推进健康中国建设。为响应政策号召，解决居民对医疗物资和服务等公共卫生设施日常和应急需求，有必要对其选址进行研究。本文利用 POI 兴趣点，基于聚类分析和改进的遗传算法构建数学模型，实现多目标的选址任务，并创新地进行组合优化。考虑到南京市景区较多，本文以风景名胜相关数据为需求点，并对其进行实证分析，最终得到了 541 个设施选址点和 180 个医疗中心选址点，实现“一站多点”的选址模式，从而缩小区域间医疗资源配置、服务能力和健康水平差异。

关键词：多目标规划；POI 数据；聚类分析；遗传算法；

1 引言及文献综述

近年来，重大公共卫生事件频发，对各国人民的安全和社会的经济造成了严重的影响，如何应对这些紧急事件已成为了国内外学者探讨研究的焦点。随着我国《“十四五”国民健康规划》的提出以及对全面推进健康中国指导思想建设的强调，基层的医疗卫生系统的解决刻不容缓。与此同时，随着旅游业的复兴和游客数量的暴增，景区是否能在短时间内及时提供物资和援助直接关系到应急医疗工作的时效性及准确性，因此科学正确地对物资发放点和医疗站进行选址非常关键。

从国内外相关学者的研究进展来看，公共设施选址问题涉及大数据的应用，汪晓春、熊峰等人(2021)在武汉市养老设施规划布局的实例研究中，通过对 Python 导出并转换格式的 POI 数据进行聚类分析和网格分化，结合机器学习算法，得到了养老设施的选址。由此可见，POI 信息点为规划人员在数据挖掘上提供了更全面和科学的方式，优化了选址规划的过程，使结果更具可靠性和参考价值。在基础经典选址模型研究中，张金凤(2021)在其论文研究中提到，选址问题的目标分为成本型、利润型、需求型目标、环境型目标四类。而随着选址问题研究的深入，对选址有了更高的要求，需要同时兼顾多种目标，面对优化改进问题，郭昌勇(2022)在覆盖模型的基础上，利用非支配排序遗传算法，以效益最大化和覆盖水平最大化为优化目标，建立了多目标二次规划模型。整理现有研究发现，鲜少有文献将设计医疗设施类选址的方法论科学系统的搭建起来并加以定量数据分析。

本文将以南京市风景名胜作为研究对象，在结合收集南京市 POI 数据的基础上，运用改进的遗传算法在规定的参数下进行数学建模反复迭代直到得出满足目标的解。

2 理论分析

¹ 作者简介：张鑫(2003-)，汉族，女，江苏盐城人，东南大学成贤学院本科在读，研究方向：管理科学与工程。；

基金：江苏省大学生实践创新训练项目“基于多目标规划的公共卫生设施投放研究”(SCX22025)；

2.1 数据来源

根据研究需求, POI 数据是描述城市各类功能空间分布的基本数据, 通常分为公共设施、科教文化服务、风景名胜、公司企业等。本文所使用的南京市 POI 数据是从高德地图为开发者所提供的 GIS 获取的, 如表 1 所示, 为统计得到的南京市各区 POI 兴趣点个数, 其中江宁区数据最多, 秦淮, 玄武和鼓楼区相对较多, 与其行政区面积和发展水平有关。

表 1 南京市 POI 信息点汇总

区域	POI 信息点个数	区域	POI 信息点个数
秦淮区	57161	浦口区	56520
玄武区	36086	雨花台区	34009
鼓楼区	60101	高淳区	22294
建邺区	33869	栖霞区	44001
六合区	51448	江宁区	111222

2.2 聚类分析

聚类分析是一种典型的无监督学习, 用于对未知类别的样本进行划分, 将样本中的数据按照一定的规则划分成若干个类簇, 把相似的样本聚在同一个类簇中, 把不相似的样本分为不同类簇, 从而揭示样本之间内在的性质以及相互之间的联系规律。不同的聚类方法适用于不同的条件和数据集。

(1) K-means 聚类:

计算样本点与类簇质心的距离, 与类簇质心相近的样本点划分为同一类簇。K-均值通过样本间的距离来衡量它们之间的相似度, 两个样本距离越远, 则相似度越低, 否则相似度越高。该方法适用于较大的数据集, 且对 K 的取值很敏感。

(2) 密度 (DBSCAN) 聚类:

密度聚类算法利用密度思想, 将样本中的高密度区域划分为簇, 将簇看作是样本空间中被稀疏区域 (噪声) 分隔开的稠密区域, 通过局部密度和局部距离的计算, 实现自适应挑选聚类中心的功能, 避免在一处很密集的地方有多个中心, 又赋予了偏远地区中出现中心的可能。

3 模型设计

3.1 数据处理

根据南京市各区面积和入流量，通过 SPSS 聚类分析将各区的 POI 信息点数量按比例缩减至算法可简便运行范围内，保留下来的有效数据既可以表现各区数据分布特点，也避免了接下来算法运行中因数据过于密集而无法合理设置参数的问题。

要得到准确的设施备选点坐标，需要通过密度聚类对现有坐标点进行分析，根据所得簇类数确定备选设施点的个数。将上述所得均值聚类中心点代入 MATLAB 分析计算，在噪声点数最小的情况下，可得结果为 1087 簇，即再进行均值聚类时，应当输入的 K 值为 1087，最终可以得到 1087 个聚类中心的坐标点，这 1087 个坐标点即为设施选址的备选点。

3.2 多目标优化算法-遗传算法的应用

基本遗传算法是一种通过模拟自然进化过程搜索最优解的方法。遗传算法通过二进制编码的“染色体”表示，将原问题的结构变换为染色体的位串结构，从初始化一个编码表示基因种群出发，采用优胜劣汰、适者生存的自然法则选择个体，对个体的适应性进行度量，并通过杂交、变异等选择操作来产生新一代种群，并反复地迭代进化，直到满足目标为止。

针对此项目研究，做出如下假设和约束条件：(1) 一个设施点可为多个需求点提供货物，且一个设施点至少覆盖一个需求点（需求点为风景名胜的 POI 数据）。(2) 运输成本与时间无关。(3) 设施点到需求点的平均范围最少，设施点数量最少，需求点覆盖量范围最多。(4) 备选点的配送能力与需求点的需求量有关。

该问题的数学建立模型如下：

$$\min F = \sum_{i \in N} \sum_{j \in M_i} \omega_i d_{ij} C_{ij} \quad (1)$$

$$s. t. \begin{cases} \sum_{j \in M_i} C_{ij} = 1, i \in N & (2) \\ C_{ij} \leq h_j, i \in N, j \in M_i & (3) \\ d_{ij} \leq s & (4) \\ C_{ij}, h_j \in \{0, 1\}, i \in N, j \in M_i & (5) \\ \sum_{j \in M_i} h_j = n & (6) \end{cases}$$

其中， C_{ij} 表示为设施备选点坐标； h_j 为 0、1 整数变量，当 h_j 为 1 时表示第 j 个设施点被选中，当 h_j 为 0 时表示第 j 个设施点没有被选中； d_{ij} 表示需求点 i 到设施点 j 的运输距离； ω_i 表示需求点 i 到设施点 j 的运输成本；公式 (1) 表示满足条件的最少设施点的值；公式 (2) 表示需求点与设施点的分配关系，确保每一个设施点至少覆盖一个需求点；公式 (6) 表示被选为设施点的数量为 n 。

3.3 改进的交叉算子遗传算法

基本的遗传算法解决这类优化存在进化速度慢和出现局部最优解的问题。因此将具有自适应的记忆功能的禁忌搜索算法对基本遗传算法进行混合，可以增加遗传算法的全局最优解的能力。具体通过改进交叉算子提高遗传算法的全局搜索的效果。通常防值个体的交叉操作对算法解的结果有较大影响，一般交叉概率 P_{cross} 取 0.5。在此问题中定义交叉概率为：

$$P_{cross} = \frac{1}{1+e^{k(f_{avg}-f_{max})}} \quad (7)$$

其中，k 为常系数，f_{max} 为最大适应度值，f_{avg} 为平均适应度值。再对算法中设置新的收敛条件，在反复地迭代过程中，父代基因池最优解未被替换，因此可算出全局最优解。

结合实际数据的调查，本文对需求点的需求量和运输成本进行了预处理。确定需求量时，对各区景点平日客流量数值进行计算，得出的计算值为需求点的需求量。备选点的运输成本则以南京市各区的地价作为标准来进行计算。在对医疗中心进行选址时，该文将设施选址的最优解作为医疗中心选址的需求点和备选点。备选点的运输成本以南京市各区的地价作为标准。需求点的需求量则看为一个常数，假设医疗中心选址点的需求量都相同，则公式(1)中的 ω_i 可更换为常数 ω 进行计算。继续上文的算法操作，可得出医疗中心选址的最优解。

设施选址和医疗中心选址的结果可视化如图 1 和图 2 所示。

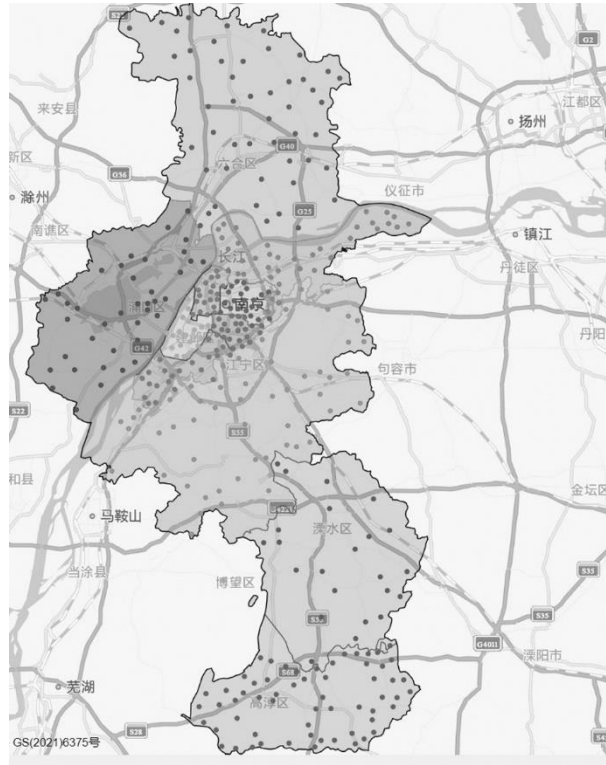


图 1 设施选址点可视化

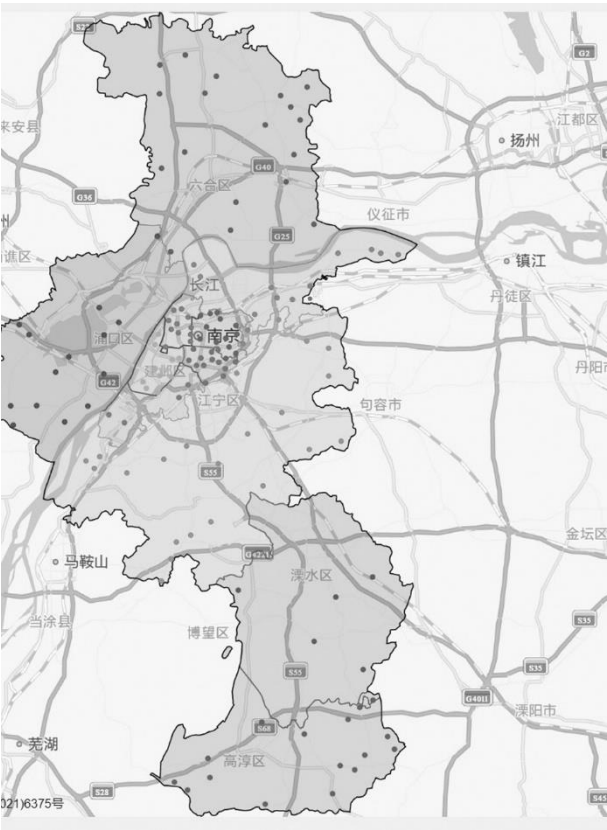


图 2 医疗中心选址可视化

即最终确定的设施选址数为 541, 应急医疗中心选址数为 180, 表 2 为选址结果汇总

表 2 设施选址结果汇总

区名	设施点数量	医疗站数量	区名	设施点数量	医疗站数量
鼓楼区	53	19	雨花台区	28	4
秦淮区	40	24	浦口区	40	14
玄武区	88	31	六合区	48	18
建邺区	41	22	高淳区	59	15
江宁区	54	22	溧水区	23	5
栖霞区	40	13			

4 结论及建议

在景区医疗安全突发性事件中，使用多目标规划的医疗设施选址方案及其优化设计确定配送站的选址具有良好的可行性。本文采用聚类分析与改进的遗传算法对南京市医疗设施系统选址进行研究，相关结论与建议如下：

(1)相较于传统的物流配送中心，承担应急医疗站这一功能的配送中心需要克服的核心问题是如何在面对突发性事件时迅速发挥功能且快速调配物资。因此相较于其他的传统的选址方法，基于深入的聚类分析和基于改进目标的遗传算法得出选址方案更加符合南京市的实际情况。将这两种方法分别运用到便民医疗设施选址模型和优化模型后，成功解决了选址的不合理性和方法复杂性，避免了资源的浪费。通过分析发现，改进的聚类分析实现了在突发事件下的紧急处理，与日常配送工作，使用粒子群算法攻克了选址时的多目标协同。

(2)对于当前的大数据时代，静态数据分析远远不够，需要打造动态可视化智能物联网平台以实现增值服务，对于运用企业而言需要联合云计算、互联网、物联网、人工智能等技术搭建动态实时可视化平台，及时发现更新的地理数据，从而整合社会新资源，构建协同网络，实现配送点与配送点之间的交互，提高系统运作效率，实现低碳下的物流运作，提高经济效益，实现城市物联。

(3)结合实证分析结果和长期发展目标，可以发现各区设施点配送点分布呈现显著的差异性。老城区：鼓楼、秦淮、玄武、建邺设施点数量领先于其他区域属于第一梯队，因其发展周期长，商业化程度高、人流量密集、名胜古迹林立而导致应急医疗的需求也随之增加，所以设施点平均覆盖面积更小可以让服务更精准。毗邻老城区的雨花台、栖霞两区人口密集程度有所缓解设施点平均覆盖面积有所增加，医疗站数量有所减少，避免了配送冗余的问题。江宁、浦口、六合等大区处于第三梯队，三区面积广阔人口众多，是南京发展的新势力区，因此配送点医疗站数量反而有所增加。高淳、溧水两区属于第四梯队，因其发展尚未成熟，被辐射能力弱，所以设施点需要补强以应对这种缺陷。

(4)基于多目标规划的医疗设施选址涉及了大量的实时数据，这些难以量化的数据对结果会有影响，数据的难获取度，也降低的结果的精准度，未来要搭建实时数据云平台就要致力于攻克此问题。由于结果会随着需求者的主观意志发生改变，如何对需求进行定位更新也至关重要，本文搭建数学模型时考虑的是单一运输单一成本，之后的研究更需要平衡多个目标的权重，考虑多重成本带来的影响，更加贴合物流与经济流通的实况。

参考文献

- [1] 汪晓春, 熊峰, 王振伟, 等. 基于 POI 大数据与机器学习的养老设施规划布局——以武汉市为例[J]. 经济地理, 2021, 41(06): 2-8.
- [2] 金仙力, 李金刚. 基于遗传算法的多目标路径优化算法的研究[J]. 计算机技术与发展, 2018, 28(2): 2.
- [3] 郭昌勇. 基于大数据分析的城市电动汽车充电设施多目标规划研究[D]. 重庆: 重庆大学, 2022: 9-25.
- [4] 李依麟. 基于 GIS 的零售业选址问题分析[J]. 商经理论, 2016, (4),
- [5] 张金凤. 设施选址问题的建模及优化算法研究[D]. 广州: 广东工业大学, 2011: 15-28.
- [6] 杨叶勇. “互联网+”时代农村物流配送中心选址的优化方案设计[J]. 物流管理, 2017, (20).

[7] 万波, 杨超, 黄松, 等. 基于层级模型的嵌套型公共设施选址问题研究[J]. 武汉理工大学学报(信息与管理工程版), 2014, 34(2): 3-5.

[8] 汤希峰, 毛海军, 李旭宏, 等. 物流配送中心选址的多目标优化模型[J]. 东南大学学报(自然科学版), 2009, 39(2): 384-425.