

模拟退火算法在动态建立基准地价模型中的应用

祝国瑞¹ 唐旭¹ 王平¹

(1 武汉大学资源与环境科学学院, 武汉市珞喻路 129 号, 430079)

摘要: 阐述了传统基准地价评估模型建立流程将地价样本检验与模型建立分离的不足, 引入了模拟退火算法思想, 对目标函数、初始状态、新状态产生、新状态接受、初温、温度更新函数以及抽样准则、算法终止准则等关键问题进行设计, 将地价样本检验、优化组合与模型建立有机结合, 并设计了基于模拟退火算法的动态建立基准地价模型的流程, 最后应用实例验证了方法的可行性。

关键词: 模拟退火算法; 基准地价模型; 土地估价

中图分类号: P273

在建立传统的基准地价模型过程中, 首先根据级别内的因素总分值区间, 将样本按照级别划分为 t 个子集 $P_1、P_2、\cdots、P_t$, 其中 $P_1 \cup P_2 \cup \cdots \cup P_t = S$, 且 $P_i \cap P_j = \emptyset (i \neq j)$; 然后根据各土地级别内的样点地价的统计分布, 采用相应的数理统计方法 (t 检验法或均值-方差法) 对样点进行检验剔除, 得到各级有效样本子集 $P'_1、P'_2、\cdots、P'_t$, 则有效样本集合为 $S' = P'_1 \cup P'_2 \cup \cdots \cup P'_t$; 最后根据有效样本集合 S' 中的样本分值和地价 (或各级样本子集中的样本地价均值和因素作用分均值) 的二维关系, 采用最小二乘法进行曲线拟合, 建立基准地价模型。但是, 受城市发展、规划限制以及外业样本采集等人为因素的影响, 不同区位发生的交易样本无论在数量、密度、地价水平以及分布特性上都有较大差异。样本按照级别检验并且与基准地价模型建立过程分离, 不利于建立优化的基准地价模型。对此, 笔者曾提出采用局部搜索算法实现地价样本检验与模型建立的有机结合, 并设计了相应的流程, 在具体的应用中效果较好。但是, 由于局部搜索算法的搜索性能依赖于邻域函数的设计和初始解, 搜索过程容易陷入局部最小, 亦存在不足。

1 模拟退火算法的应用思路

模拟退火算法 (simulated annealing, SA) 是一种

基于 Monte Carlo 迭代求解策略的随机寻优算法。在模拟固体退火过程中, 在某一初温下, 算法通过应用 Metropolis 抽样策略具有在局部最优解能概率性地跳出的特性。伴随着温度的下降, 不断重复样本的选取过程, 最终在解空间中得到目标函数的全局最优解。如图 1 所示, 在建立基准地价模型的过程中, 有效样本子集与基准地价模型之间是一一映射关系。

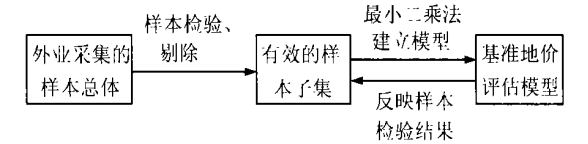


图 1 样本子集与基准地价模型的映射关系
Fig. 1 Relationship Between Sample Subset and Land Benchmark Price Model

因此, 基准地价模型的优化实际上是地价样本子集的组合问题。传统的基准地价模型建立时, 把样本检验剔除和模型建立分离, 不能克服由于有效样本子集分布的不确定性和样本地价分布不均对“地价 \rightleftharpoons 因素总分值”关系分布规律的影响。从 Metropolis 准则可以看出, 应用模拟退火算法接受新解时, 除了接受优化解以外, 还在一定的范围内接受恶化解。开始, 控制参数 t 值较大, 可接受较差的恶化解; 随着 t 值的递减, 只能接受较好的恶化解; 最后当 t 值趋近于 0 时, 就不再接受任何恶化解, 从而达到全局最优, 解决了局部搜索

算法容易陷入局部最小的缺陷。因此,应用模拟退火算法的思想,将样本检验剔除与地价模型建立动态结合,在顾及样本地价总体分布的前提下,摆脱局部最优解的“陷阱”,对样本进行最佳优化组合,获取模型最优解的思想是可行的。

2 关键问题设计

1) 目标函数的设计。目标函数是衡量解的优劣的函数。样点地价和因素总分值主要有线性和指数两种关系,基准地价模型包括线性模型 $y=a+bx$ 和指数模型 $y=a(1+b)^x$ 。因此,可以采用模型回归显著性检验的相关参数来设计目标函数。

设 $\theta=(\hat{a}, \hat{b})$ 表示地价模型系数的估计量, R^2 为线性回归方程 $F(x)$ 的判定系数,将目标函数设为:

$$C(\theta)=a(1-R^2)$$

式中, a 为常数; R^2 为判定系数:

$$R^2=S_{\text{回}}/S_{\text{总}}=\sum(\hat{y}_i-\bar{y})^2/\sum(y_i-\bar{y})^2$$

目标函数 $C(\theta)$ 的值越小, R^2 越接近 1, 则解 θ 越优, 地价回归方程拟合效果越显著。

2) 初始状态的确定。初始状态的确定有以下两种方法:① 令 $S'=S$, 以所有样本地价和因素总分值的均值建立回归模型, 作为初始解 θ_0 ; ② 以各级别样本子集 P_i 中样本地价的均值和因素总分值的均值作为二维变量建立回归模型, 作为初始解 θ_0 , 由 θ_0 计算对应的初始目标函数值 $C(\theta_0)$ 。

3) 新状态产生。新样本子集的产生是目标函数状态转移的关键。首先根据前一状态求得的基准地价模型, 求取样本总体中每个样点的地价偏差 $d_i=|y_i-\hat{y}_i|, i=1, 2, \cdots, n, i$ 表示样点编号。设 p_i 表示每个样点被选取的概率, 则 $P_i=1-\sqrt{d_i}/\sqrt{d_{\max}}$, 其中 d_{\max} 为 d_i 中的最大值。样点的选取概率如图 2 所示。

将随机函数 $\text{Random}[0, 1]$ 作为样本的选取控制函数。抽取一个样本, 当 $P_i>\text{Random}[0, 1]$ 时, 则该样本为有效样本。按照这种方法在样本集合 S 中随机选取一定比例且保证能反映 S 的总体分布的样点组成有效样本子集 S' (选取比例可根据工作经验和数据抽样检查的可信度来确定)。新样本子集产生新的基准地价模型, 当前解 θ_k 改变, $C(\theta_k)$ 也随之改变, 算法状态由 E_i 转为 E_j 。

以选取概率为衡量准则来进行样点的选取, 在

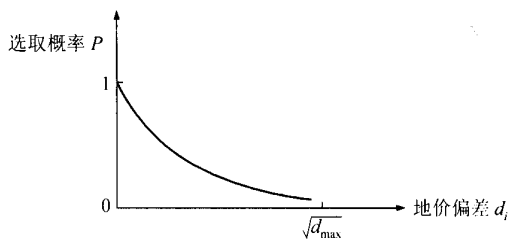


图2 样点选取的概率关系图

Fig. 2 Probability of Selected Samples

抽取样本产生新状态时, 保证正常样本在下次状态中出现的概率较大, 有利于加速算法的收敛。

4) 状态接受条件。求取状态 E_j 对应的目标函数值 $C(\theta_j)$, 计算改变后的目标增量 $\Delta C=C(\theta_j)-C(\theta_k)$ 。如果 $\Delta C<0$, 则接受新状态 E_j , 令 $\theta_k=\theta_j$, 并判断 $C(\theta_j)<C(\theta_{t_k}^*)$ ($\theta_{t_k}^*$ 是当前温度 t_k 下的最优解) 是否成立, 若成立, 则令 $\theta_{t_k}^*=\theta_j$; 如果 $\Delta C>0$, 若 $\exp\{-\Delta C/t_k\}>\text{Random}[0, 1]$, 则接受新状态 E_j , 否则拒绝。

5) 初温的确定。初始温度 t_0 常用的确定方案是随机产生一组状态, 确定两个状态间的最大目标值差 $|\Delta_{\max}|$, 然后依据一定的函数如 $t_0=-\Delta_{\max}/\ln p_r$ 来确定初温, 其中 p_r 为初始接受概率 (理论上应接近于 1)。为提高算法优化效率, p_r 可定为 1/3, 作为 Metropolis 准则接受较差解的状态转移概率的阈值。

6) 温度更新函数。目前最常用的温度更新函数为指数退温函数, 即 $t_{k+1}=\lambda t_k$, 其中 λ 为退温速率。本文采用线性下降方法进行降温处理, 这里取 $\lambda=0.9$ 。

7) Metropolis 抽样稳定准则, 即内循环终止准则。抽样稳定准则为: 在当前温度 t_k 下的最优解 $\theta_{t_k}^*$ 连续 K_1 次保持不变。

8) 算法终止准则。算法终止准则即外循环准则, 可以定为以下几种: ① 连续 K_2 次退温过程中所得的最优解 θ^* 保持不变; ② 目标函数 $C(\theta_j)<\epsilon$; ③ 算法迭代次数 $K<K_{\max}$ 。只要满足其中一种, 算法即结束。

3 动态逼近的地价模型流程的建立

建立地价模型流程如下。

1) 在给定初温 t_0 和初始状态 (初始基准地价评估模型) 下, 求取对应的目标函数值 $C(\theta_0)$, 令全局最优解 $\theta^*=\theta_0, \theta_{t_k}^*=\theta_0(k=0)$ 。

2) 按照选取准则, 在有效样本总体中重新随

机选取一个样本子集,产生一个新的地价模型,即新状态 E_j ,并与当前状态进行比较,判定是否接受该新状态。

3)检查是否达到该温度下的内循环终止准则,如果达到则转到步骤 4);否则返回步骤 2),继续执行。

4)判断 $C(\theta_{i_k}^*) < C(\theta^*)$ 是否成立,若成立,则令 $\theta^* = \theta_{i_k}^*$ 。检查是否达到外循环终止准则,如果满足则以最优解 θ^* 为最终解输出,算法结束;否则,返回步骤 2),继续执行。

模拟退火算法不断搜索新解的过程,就是地价模型的动态建立过程,即不断地逼近最优地价模型。算法流程图如图 3 所示。

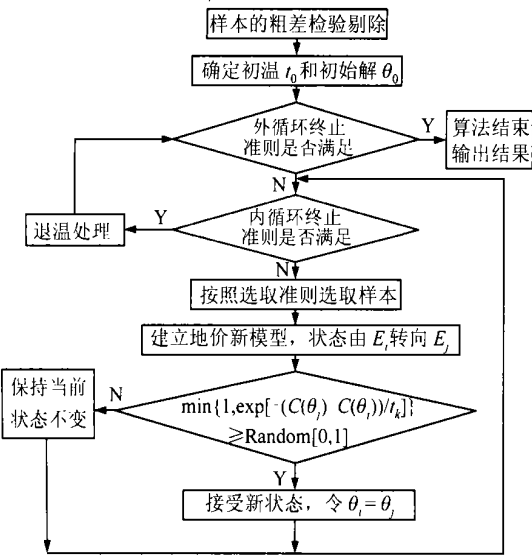


图 3 模拟退火算法流程图

Fig. 3 Process of Simulated Annealing Algorithm

4 应用与分析

4.1 算法的应用

将基于 SA 的基准地价评估模型动态建立算法应用于某市基准地价评估中。已知全市土地级别为 4 级,采集的商业用地样本共 726 个,且根据抽样调查可知,样本的可信度为 80%。样本具体情况见表 1。

表 1 某市商业用地样本数据情况表					
Tab. 1 Sample Data of Commercial Land in a City					
土地级别	样点数目	样本均价 / 元·m ⁻²	样本地价范围 / 元·m ⁻²	级别作用分区间	级别平均分
一级	130	5 577.58	(167, 20 717)	(62.16, 80.85)	69.89
二级	362	4 068.78	(150.45, 815)	(47.09, 62.16)	54.67
三级	200	2 632.10	(43, 12 943)	(29.08, 47.09)	39.38
四级	34	1 500.22	(4, 4 636)	(5.01, 29.08)	24.41

应用基于 SA 的建模方法,对上述样本进行模型建立。其中令目标函数中常数 $a=100$,则可确定初温 $t_0 \approx 100$,内循环准则中 $K_1=20$,外循环准则为算法迭代次数 $K < K_{\max}$,其中令 $K_{\max}=5\,000$ 。另外,样本选取比率为 80%。算法的逼近过程如表 2 所示。

表 2 模拟退火算法的退火过程

Tab. 2 Instance of Annealing Process of the Algorithm

退温次数	当前温度	总循环次数	判定系数
1	100	21	0.556
3	81	77	0.567
5	65.61	127	0.568
6	59.05	166	0.578
41	1.48	1 269	0.579
50	0.57	2 061	0.595
63	0.15	3 535	0.601

由实例可知,共退温 63 次,判定系数随温度的降低而增大,模型逐步优化。由此可见,算法收敛。

4.2 多种方法的结果比较

分别采用传统地价建模方法(I)和基于局部搜索算法(II)对同一组数据进行模型建立,并将建模结果和基于模拟退火算法(III)的建模结果进行比较,计算结果见表 3。

表 3 几种方法建模结果的比较

Tab. 3 Comparison Between the Results of Several Methods

方法	样本数目	各级别基准地价				地价和作用分值模型	R^2
		一级	二级	三级	四级		
I	135	4 921	2 873	1 673	985	$Y=415.524(1+0.086)^X$	0.461
II	141	4 713	2 633	1 467	827	$Y=325.096(1+0.089)^X$	0.508
III	145	5 361	2 743	1 399	724	$Y=247.278(1+0.045)^X$	0.601

从表 3 中的各级别基准地价可以看出,基于模拟退火算法的建模方法,克服了二级和三级用地样本对建模样本数目占优的影响,较好地反映了样本的总体趋势。例如,由于城市商业用地一级用地范围较小,级内土地区位及质量差异小,因此,反映区域平均地价水平的级别价与处在级内区位较好的沿路样本的样本均价差异应该较小,其他如二级土地的级别地价应该比样本均价低 30%左右。

在保证样本置信水平相近的前提下,基于模拟退火算法的模型复相关系数最高,优于传统建模方法和基于局部搜索算法的基准地价建模方法。在大样本量的建模时,时间效率不及基于局部搜索算法的基准地价建模方法,但在计算机硬件水平不断提高的条件下,这一点已经显得不那么重要了。

(下转第 607 页)

Design of Object-Oriented Survey Adjustment Software and Its Online Service

XIE Zhiying¹ LI Qingquan¹ PENG Junhuan²

(1 Research and Development Center of Spatial Information and Network Communication,
Wuhan University, 129 Luoyu Road, Wuhan, China, 430079)

(2 School of Geodesy and Geomatics, Wuhan University, 129 Luoyu Road, Wuhan, China, 430079)

Abstract: In this paper, the survey adjustment software was redesigned with the method of OOP (object-oriented programming), and a set of classes which embodies the procedure of survey adjustment was implemented with the purely object oriented language as Java.

Key words: object-oriented; survey adjustment; Internet information service

About the first author: XIE Zhiying, Ph. D candidate. He majors in survey engineering, Web GIS and mobile spatial information service.
E-mail: xzy-10@sina.com

(责任编辑: 平子)

(上接第 595 页)

参 考 文 献

1 王 凌. 智能优化算法及其应用. 北京: 清华大学出版社, 2001. 17~29

2 国土资源部土地估价师资格考试委员会. 土地估价理论与方法. 北京: 地质出版社, 2000

3 严 星, 林增杰. 城市地产评估. 北京: 中国人民大学出版社, 2000

4 唐 旭. 基于基准地价评估模型的修正体系编制方法

研究. 中国土地科学, 2002(2)

5 陈幼松, 杨位钦. 实用数理统计方法及应用详解. 北京: 北京科学技术出版, 1998

6 王新生, 姜友华. 模拟退火算法及其在非线性地学模型参数估计中的应用. 华中师范大学学报(自然科学版), 2002, 35(1)

第一作者简介: 祝国瑞, 教授, 博士导师。现主要从事数字地图分析与应用、地理信息三维可视化和土地信息理论与技术的研究。
E-mail: zhugr123@sohu.com

Application of Simulated Annealing Algorithm to Building Land Benchmark Price Model

ZHU Guorui¹ TANG Xu¹ WANG Ping¹

(1 School of Resource and Environment Science, Wuhan University, 129 Luoyu Road, Wuhan, China, 430079)

Abstract: In this paper, the target function is designed, which contains the parameter of significance test for the regression model. After a certain ratio of samples is selected, a new model will be reached by regression analysis and a new value of target function can be compared with that of previous state. Other key processes such as initial temperature, initial state, stabilization rule of metropolis sampling, state subrogation and the termination rule are designed.

Key words: simulated annealing algorithm; land benchmark price model; land valuation

About the first author: ZHU Guorui, professor, Ph. D supervisor. He is engaged in the domain of analysis and application of map visualization of geographic 3D information and theories and technology of land information system, etc.
E-mail: zhugr123@sohu.com

(责任编辑: 晓晨)