

文章编号: 1674-9057(2020)03-0546-05

doi:10.3969/j.issn.1674-9057.2020.03.013

# 顾及空间相关特征的最大熵 DEM 插值权系数计算

陈天伟, 康传利, 陈 明, 段青达, 韦海福

(桂林理工大学 广西空间信息与测绘重点实验室, 广西 桂林 541006)

**摘 要:** 针对传统格网 DEM 插值建模时由于主观因素的干扰, 在解算权系数时存在负权现象的问题, 提出了一种解算 DEM 插值权系数的最大熵模型算法。该方法以熵函数作为目标函数, 设置非负权值和参考点数据的 0、1、2 阶统计矩作为约束条件, 为了揭示参考点的空间自相关特征, 建立最小最大权系数约束条件, 通过最大化熵值来求解格网 DEM 插值的非负权系数。最大熵法解得权系数大小比例与点位关系相适应, 在 MATLAB 软件平台利用遗传算法求解, 并与传统插值方法进行了比较, 其估值精度优于传统插值方法。

**关键词:** 最大熵; 负权; 约束条件; 相关特征; 遗传算法

**中图分类号:** O29; P207

**文献标志码:** A

## 0 引 言

格网 DEM 作为国家基础地理信息基础, 得到了广泛的应用, 具有极其重要的意义。然而其精度仍有进一步提高的需求, 特别是在数字地形分析领域。一般, 规则格网 DEM 由数学插值方法得到。插值方法有多种, 包括多项式插值、多面函数插值、地统计插值、加权平均插值等算法<sup>[1]</sup>。由插值核函数解算插值权系数, 以求得待插点高程值, 其高程误差来源于原始数据误差和模型逼近误差两方面, 插值模型与实际地表符合的程度决定了逼近误差的大小。如地学界一般认为, 地统计学插值方法是在最优线性无偏估计的准则下, 对未知点空间属性及插值精度进行估计<sup>[2-3]</sup>, 在平稳假设的前提下, 理论上可以得到很高的插值精度; 然而, 地形表面的变化时常不符合平稳假设, 此时构建的插值模型与实际不符就会显著降低插值精度。如由地统计学的杨赤中插值法求解参考点的权系数时, 可能会出现负权系数, 导致估值结果的不合理, 引起 DEM 建模的失真甚至错误。有国内学者对此类负权现象展开研究, 认为出现负值与参考点、待估点

的点位关系以及插值函数模型等因素相关<sup>[4]</sup>。而笔者则认为更主要的是由于传统插值方法在构建数学模型时存在主观因素的干扰。由于空间抽样数据(如高程值)存在随机性, 依据最大熵原理<sup>[5]</sup>, 要使数学模型对样本试验数据依赖程度最大, 那么熵值应该最大, 使得插值模型更能反映空间数据的离散性, 则模型会最稳健, 解算模型则可以避免负权系数的产生, 笔者在文献[6-7]中验证了上述观点的正确性。另外, 注意到地形离散点之间客观存在着空间相关性, 在建模时顾及空间数据的相关特征, 应可以使 DEM 模型更加符合客观实际。

## 1 模型建立

### 1.1 最大熵模型的基本形式

最大熵方法是以空间域随机变量的信息熵最大为准则, 其数学模型包括目标函数和约束条件。目标函数即熵值最大, 约束条件包括样本点(参考点)高程值的数学期望和方差等, 故建立最大熵模型<sup>[6-7]</sup>。

模型系统的熵值

$$S[p] = - (p_1 \ln p_1 + p_2 \ln p_2 + \cdots + p_n \ln p_n) / \ln 2, \quad (1)$$

约束条件

收稿日期: 2019-01-11

基金项目: 广西自然科学基金项目(2017GXNSFAA198308); 广西空间信息与测绘重点实验室主任基金项目(15-140-07-09)

作者简介: 陈天伟(1965—), 男, 硕士, 教授, 大地测量学与测量工程专业, 675916588@qq.com。

引文格式: 陈天伟, 康传利, 陈明, 等. 顾及空间相关特征的最大熵 DEM 插值权系数计算 [J]. 桂林理工大学学报, 2020, 40(3): 546-550.

$$\begin{cases} p_1 + p_2 + \cdots + p_n = 1; \\ h_1 p_1 + h_2 p_2 + \cdots + h_n p_n = E_1; \\ (h_1 - E_1)^2 p_1 + \cdots + (h_n - E_1)^2 p_n = E_2. \end{cases} \quad (2)$$

其中： $p_i$  为空间域变量的产生概率（即权系数）； $p_i > 0$ ； $i = 1, 2, \dots, n$ ； $S[p]$  为系统熵值。约束条件第1式即0阶矩约束（或称自然矩约束）；第2式即1阶矩约束： $E_1 = E(h)$  是高程变量原点矩的期望值（随机变量  $h_i$  的原点矩等于相应的样本矩）<sup>[8]</sup>，则  $E(h)$  可利用参考点与估值点距离反比例赋权，对参与估值计算的参考点权系数取加权平均值算出；第3式即2阶矩约束， $E_2$  是高程变量2阶中心矩的期望值，利用高程样本值和  $E_1$  按计算方差估值的公式算出。

## 1.2 空间变量相关特征的表达

上述约束条件的本质是通过期望值和方差反映研究区域高程值变量的分布中心以及离散度。根据地理学第一定律，地貌在一定空间范围内具有相关性，应在约束条件中表达出这些特征。

在地统计学理论中，认为高程属于区域化变量，满足2阶平稳假设情况下，在一定的范围内，点和点之间是相关的<sup>[9]</sup>，相关程度与距离成反比。克里金法采用了半变异函数等指标描述此相关特征，杨赤中法采用了综合随机特征函数进行描述<sup>[10]</sup>。常用的反距离加权插值法亦是基于地理学第一定律的原理，即在一定区域，距离越近事物越相似。那么在解算 DEM 插值模型得到的各参考点权系数的大小，其实也就表达了插值点（待估点）与其邻域的参考点（已知点）的相关程度，即权系数大小与参考点、估值点间距平方成反比。为了简化最大熵模型的复杂度，在设置约束条件中仅仅增设最大最小权约束条件，即距离插值点最近的参考点权系数最大；反之，距离插值点最远的参考点权系数最小。

如图1所示，某待估点（最中心的点）对应的参考点有5个，其中1<sup>#</sup>点距离待估点最远，3<sup>#</sup>点最近，根据空间数据自相关的特性，1<sup>#</sup>点权系数应该最小，3<sup>#</sup>点权系数应该最大。

相应的空间相关约束条件：①最小权约束条件  $p_1 - p_2 < 0$ ， $p_1 - p_3 < 0$ ， $p_1 - p_4 < 0$ ， $p_1 - p_5 < 0$ ；②最大权约束条件  $p_2 - p_3 < 0$ ， $p_4 - p_3 < 0$ ， $p_5 - p_3 < 0$ 。转换为矩阵表达式

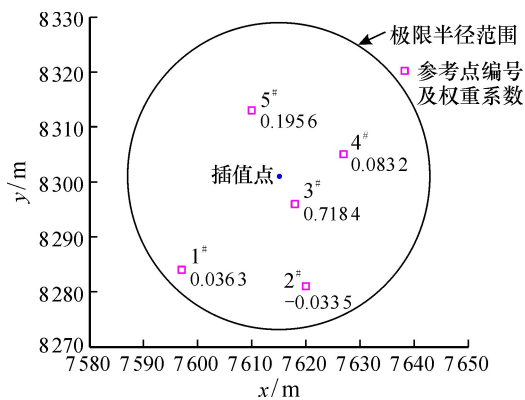


图1 参考点、插值点的点位分布

Fig. 1 Distribution of reference points and interpolation point

$$\begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \end{bmatrix} < \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad (3)$$

$$\begin{bmatrix} 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \end{bmatrix} < \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \quad (4)$$

此处，2<sup>#</sup>参考点在杨赤中法计算结果为负数，这是因为该点到待估点的方向被3<sup>#</sup>点屏蔽的缘故。2<sup>#</sup>、4<sup>#</sup>、5<sup>#</sup>参考点与待估点的距离介于最大、最小之间，本文不作类似的空间相关性约束，以免增加问题的复杂程度从而影响模型解算的效率。它们的权系数由以上模型自然解得。可见，上述有关空间相关性的约束表达式相当于给最大熵模型提供了最重要的空间相关特征信息，并进一步缩小了可行解范围，有利于加快模型解算收敛速度，求得最优解。

待估点  $x_0$  的估计量

$$h^*(x_0) = \sum_{i=1}^n [p_i \cdot h(x_i)]. \quad (5)$$

## 2 模型解算

上述模型的解算本质上属于非线性函数约束优化问题，即寻求全局最优解的问题。所谓全局最优解是指每一次解算某个未知点高程时，先要解算搜索到的参考点的权系数，这些权系数的解有很多组，要找到满足约束条件且对应熵值为最大的权系数解，进而推算未知点高程。

由于目标函数是复杂的非凸函数，只能用数

值计算方法进行优化。传统的优化算法往往求得的是局部最优解，而不是全局最优解，在此利用遗传算法进行解算，再结合最大最小权约束条件，能尽快锁定全局最优解。遗传算法(GA)模仿遗传学自然选择进化过程，主要是通过基因的组合交叉和变异逐步进行问题的优化，避免陷入局部最优解，从而寻找到全局最优解。该算法对于大空间、全局范围内求解最优解具有较强的优势。其基本操作包括编码、选择初始种群、设置适应度函数、遗传算子操作(选择、交叉和变异)、控制参数设定、循环终止设计<sup>[11]</sup>。GA 自身通常解决无约束优化问题，而本文建立的最大熵模型包括 3 个线性等式约束和  $n$  个非负约束， $n$  即参考点个数。实践中常用罚函数法，将有约束问题转化为无约束问题。通过惩罚不可行解，使其目标函数值增大(对求极小值问题)，适应度值减小，经过不断迭代，种群逐渐收敛于可行解，同时迫使无约束问题的极小点逐渐收敛于约束问题的极小点。构造带有惩罚项的适应度函数一般有两种：一种是采用加法形式  $\text{val}(x) = f(x) + p(x)$ ，另一种是乘法形式  $\text{val}(x) = f(x) \cdot p(x)$ 。其中  $p(x)$  是惩罚项。

商业数学软件 MATLAB 附带有遗传算法函数工具箱，MATLAB 的 GA 主函数调用形式为<sup>[12]</sup>  $[x, \text{fval}] = \text{ga}(\text{ObjectiveFunction}, \text{nvars}, \mathbf{A}, \mathbf{b}, \text{Aeq}, \text{beq}, \text{LB}, \text{UB}, \text{ConstraintFunction}, \text{options})$ ，其输入参数包括：目标函数名、变量个数、线性不等式约束系数阵及常量、线性等式约束系数阵及常量、变量的上下界、非线性约束条件函数名、选项。

前述约束条件式(2)的前 3 项对应于线性等式约束 Aeq 和 beq,  $p_i > 0$  非负约束对应于变量的上、下界 UB 和 LB, 而对于组成线性不等式约束系数阵  $\mathbf{A}$  及常量  $\mathbf{b}$ , 由空间相关约束条件式(3)、(4)得到。一般地, 对于某个待估点, 设参考点有  $n$  个, 第  $i$  点距离待估点最远(权最小), 第  $j$  点最近(权最大)。

组成线性不等式约束系数矩阵的方法是先列出最小权系数约束条件, 其不等式系数矩阵为: 在行列数均为  $n - 1$  的负单位阵第  $i$  列之前插入元素为 1 的列向量, 即得线性不等式约束系数阵的第一部分  $\mathbf{A}_1$ , 即式(3)左边的系数矩阵; 再列出最大权系数约束条件, 其不等式系数矩阵为: 在行列数均为  $n - 2$  的单位阵第  $i$  列之前插入元素为 0

的列向量(若  $i$  大于  $n - 2$ , 则在最后列右边加入), 再在新矩阵的第  $j$  列之前插入元素为  $-1$  的列向量(若  $j$  大于  $n - 1$ , 则在最后列右边加入), 即得线性不等式约束系数阵的第 2 部分  $\mathbf{A}_2$ , 即式(4)左边的系数矩阵; 上下组合  $\mathbf{A}_1$  和  $\mathbf{A}_2$  即得  $\mathbf{A}$ , 常量  $\mathbf{b}$  是  $2n - 3$  个 0 元素组成的列向量, 即式(3)、(4)右边的 0 元素列向量。

应注意, 若最大熵模型不存在非线性约束条件, 则在程序中指定非线性约束条件函数名为空值。

### 3 实例与分析

在某山地区域 1:1 万地形图提取离散特征高程点 60 个<sup>[7]</sup>, 以选取某待估点作插值运算为例, 按极限半径搜索该待估点四周的参考点, 得到 5 个参考点。分别采用杨赤中法、二次规划法、最大熵法 3 种方法计算参考点权重系数进行对比, 结果见表 1。

表 1 不同插值方法参考点权重系数对比						
Table 1 Comparison of reference points weight coefficients with different interpolation methods						
插值方法	参考点编号					权之和
	1 <sup>#</sup>	2 <sup>#</sup>	3 <sup>#</sup>	4 <sup>#</sup>	5 <sup>#</sup>	
杨赤中法	0.036 3	-0.033 5	0.718 4	0.083 2	0.195 7	1.000 1
二次规划法	0.026 1	0.365 8	1.27E-05	0.256 5	0.351 5	0.999 9
最大熵法	0.098 3	0.121 7	0.559 0	0.103 4	0.117 7	1.000 1

根据杨赤中法, 不考虑负权问题即直接解算方程组(1), 解得  $\lambda_2 = -0.033\ 5$ , 结果证明负权问题是存在的。

利用二次规划法解算权系数<sup>[13]</sup>, 参照杨赤中方法算得的权系数, 将负权系数改为 0.001, 设权系数初始值为  $[0.050\ 0.001\ 0.688\ 0.111\ 0.152]$ , 计算结果见表 1 中的二次规划法。

在 MATLAB 平台利用遗传算法函数进行编程, 对最大熵模型进行解算。基本步骤是: 先利用式(2)作为目标函数, 编制适应度评价函数; 再编制主程序及确定初始值  $X_0$ , 初始值即初始权系数, 其值选择恰当, 则能够加快优化计算的收敛速度。根据参考点与估值点的距离反比例计算权系数, 赋权公式为  $p_i = 1/d_i$ , 其中  $d_i$  为参考点  $i$  到估值点的距离。对  $p_i$  归一化得到程序调用的初始权系数为 0.023 9、0.030 2、0.403 6、0.311 2、0.231 1。由此计算得矩约束条件的系数  $E_1$  和  $E_2$ 。



在遗传算法的主程序里，首先指定调用的目标函数，再确定约束条件的各系数矩阵，设置参数及调用遗传算法函数的主要代码为：

```
ConstraintFunction=[];  
% 没有非线性约束条件  
options=gaoptimset('PlotFens',{ @ gaplotbestf,@ gaplot-  
maxconstr},'Display','iter');  
options=gaoptimset(options,'InitialPopulation',X0);  
[x,fval]=ga( ObjectiveFunction,nvars,A,b,Aeq,beq,  
LB,UB, ConstraintFunction,options)
```

运行程序，可解算得各  $p_i$  值即权重系数得到结果(表 1)，可见最大熵法能消除空间权重系数负值现象。

3.1 权值的对比

由表 1 数据可见，杨赤中法存在负权现象(图 1)。二次规划方法虽然能消除负权现象，但是得到的权值分布不合理，离待估点最近的参考点权值接近 0 值，远离待估点的参考点权值却是最大的，此现象不符合空间数据相关的规律。负权值对应的 2#参考点虽然离插值点并不是最远的，但被靠近插值点的 3#参考点所屏蔽，最大熵法求得 2#参考点的权值为非负，且大于最远的 1#参考点的权值，这表明该法是符合空间变量相关特征的。

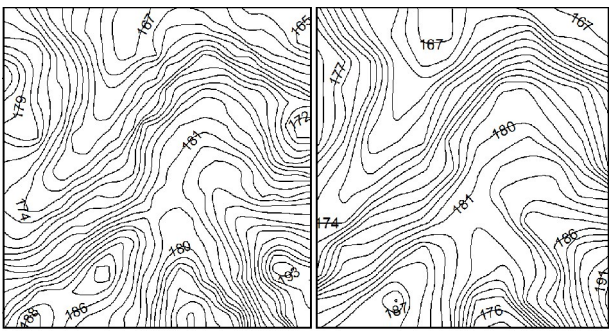
3.2 精度的对比

3.2.1 抽样统计高程插值中误差 对研究区域划分 12.5 m×12.5 m 格网，利用杨赤中、二次规划和最大熵 3 种方法分别进行格网点高程插值，抽样统计格网点的高程插值中误差结果分别为±1.123 6、±2.168 6 和±0.963 2 m。可见，最大熵法算得的估值精度最高，满足国家有关 DEM 建模精度的规范要求<sup>[14]</sup>。

图 2 为上述区域采用杨赤中方法、顾及自相关特征最大熵方法生成的等高线，可见等高线回放效果有一些差异，后者反映出更多的局部地形特征细节。

3.2.2 各种插值方法构建 DEM 的高程比较 将最大熵法插值得到的网格点高程数据在 ArcGIS 软件构建 DEM，并与克里金法、反距离加权法构建的 DEM 对照分析如图 3 所示。

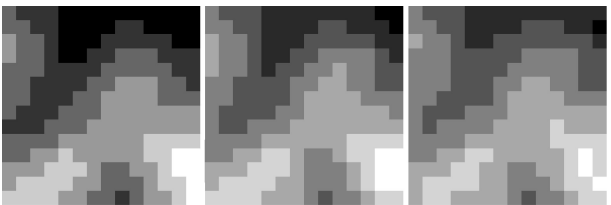
作栅格运算，将最大熵法 DEM 分别减去克里金法和反距离加权法的 DEM，结果如图 4 所示(黑色、灰色、白色分别表示正最大、最小、负最大)。在图 4a 中，高程差值绝对值最大为 2 m，在起伏较大区域，最大熵方法平滑效果比较明显，即在山



(a) 杨赤中法 (b) 顾及自相关特征最大熵法

图 2 不同高程插值方法生成的等高线

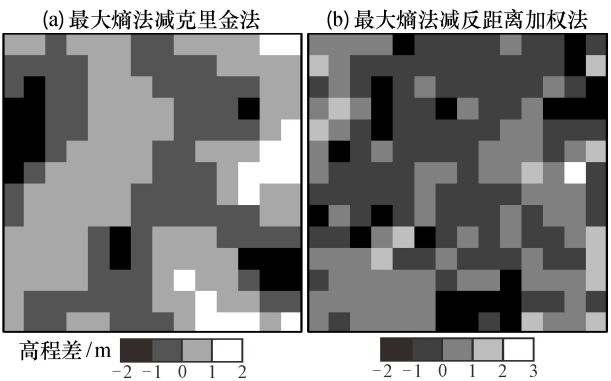
Fig.2 Contour generated by different elevation interpolation methods



(a) 最大熵法 (b) 克里金法 (c) 反距离加权法

图 3 不同高程插值方法的 DEM 栅格图对比

Fig.3 Comparison of DEM raster maps with different elevation interpolation methods



(a) 最大熵法减克里金法 (b) 最大熵法减反距离加权法

图 4 不同高程插值方法栅格相减结果

Fig.4 Results of raster subtraction with different elevation interpolation methods

顶最大熵法高程值低于克里金法；在谷底最大熵法高程值高于克里金法；图 4b 中，高程差值最大为 3 m，且没有太明显的规律，由误差理论分析得知该数值符合 3.2.1 节各种高程插值方法对中误差的描述。

3.3 时间效率的对比

最大熵法、杨赤中法、二次规划法、克里金法、反距离加权法的插值时间效率分别为 3.667、2.985、3.534、3.018、2.365 s。从计算效率来看，由于附加了空间相关性约束条件、缩小了可行解范围，相较于文献 [7] 不带空间相关性约束

条件的最大熵法加快了模型解算收敛速度。由此可见,反距离加权法速度最快,而最大熵法运算效率最慢,主要原因是最大熵方法解算最优值时采用了智能算法 GA,其使用随机搜索并进行了多次的迭代,这是其不足之处。

## 4 结束语

对于传统的 DEM 格网插值数学模型,由于其在拟合估值核函数时存在主观因素的干扰,模拟得出的数学模型存在不稳健、不合理的现象,以至于有时出现负权系数。本文采用最大熵为目标函数的建模方法,在约束中增设空间数据的相关性特征约束,最大程度地排除了主观因素,消除了空间权重系数负值现象,得到的权值大小比例分布与参考点位分布协调一致,插值精度更高。这说明该法不仅能客观反映出空间数据的离散性,也使得模型系统更加稳健。从而 DEM 插值精度有了较大的提高。不足之处是,顾及空间相关特征的最大熵法模型解算总的运算速度较慢,下一步研究将转向减少其迭代次数或者采用解析算法以提高效率与精度。

## 参考文献:

[1] 张锦明. DEM 插值算法适应性研究 [D]. 郑州: 解放军

信息工程大学, 2012.

- [2] 陈天伟, 卢献健, 杨子江. 基于遗传算法的杨赤中插值权系数解算 [J]. 桂林理工大学学报, 2015, 35 (3): 514-517.
- [3] 陈天伟. 基于杨赤中法 DEM 对地形断裂线的处理 [J]. 桂林工学院学报, 2009, 29 (3): 378-380.
- [4] 彭军还. 函数模型对杨赤中估值系数的影响 [J]. 桂林工学院学报, 1996, 16 (2): 178-184.
- [5] 姜昱汐, 迟国泰, 严丽俊. 基于最大熵原理的线性组合赋权方法 [J]. 运筹与管理, 2011, 20 (1): 53-59.
- [6] 陈天伟, 卢献健, 江海东. 应用最大熵方法改进克里金负权系数 [J]. 桂林理工大学学报, 2015, 35 (2): 340-342.
- [7] 陈天伟. 最大熵模型联合遗传算法解算 DEM 插值权系数 [J]. 测绘科学, 2017, 42 (1): 25-28.
- [8] 程亮. 最大熵原理与最小熵方法在测量数据处理中的应用 [D]. 成都: 电子科技大学, 2008.
- [9] 侯景儒, 郭光裕. 矿床统计预测及地质统计学的理论与应用 [M]. 北京: 冶金工业出版社, 1993.
- [10] 陈天伟, 彭军还. 基于杨赤中推估法的 DEM 及等值线绘制 [J]. 测绘通报, 2007 (5): 28-29, 53.
- [11] 林于一, 沈阳. 遗传算法在非线性的约束优化问题上的应用 [J]. 机械设计与研究, 1998 (2): 13-14.
- [12] 陈杰. MATLAB 宝典 [M]. 北京: 电子工业出版社, 2007.
- [13] 龚小鹏. 基于数据场的空间数据建模研究 [D]. 北京: 中国地质大学 (北京), 2009.
- [14] 江帆. DEM 表面建模与精度评估方法研究 [D]. 郑州: 解放军信息工程大学, 2006.

# Calculation of DEM interpolation weight modulus based on maximum entropy with self-correlation characteristics

CHEN Tian-wei, KANG Chuan-li, CHEN Ming, DUAN Qing-da, WEI Hai-fu

(Guangxi Key Laboratory of Spatial Information and Geomatics, Guilin University of Technology, Guilin 541006, China)

**Abstract:** Due to the interference of subjective factors in traditional grid DEM interpolation modeling, there is negative weight when solving the weight coefficient. A maximum entropy model algorithm for solving DEM interpolation weight coefficients is proposed. The entropy function is taken as the objective function. Non-negative weights are set 0, 1, and 2 statistical moments of reference point data as constraints. In order to reveal the spatial autocorrelation characteristics of reference points, minimum and maximum weight coefficient constraints are established, and the non-negative weight coefficients of grid DEM interpolation are solved by maximizing the entropy value. The proportion of the weight coefficient obtained by the maximum entropy method adapts to the points. The genetic algorithm is solved by MATLAB, the estimation accuracy is better than traditional interpolation method compared with traditional interpolation method.

**Key words:** maximum entropy; negative weight; constraint condition; related characteristics; genetic algorithm