

文章编号:1001-9081(2009)01-0127-03

基于免疫克隆算法的容量受限工厂选址问题研究

漆 杨,秦子玄,陈 霞,于中华

(四川大学 计算机学院,成都 610065)

(sheepqy@gmail.com)

摘 要:容量受限的工厂选址问题属于 NP 难题,较难得到最优解和满意解。针对现有方法存在的不足,提出了基于免疫克隆的容量受限工厂选址算法,该算法采用不同于一般免疫克隆算法的编码、抗体产生、克隆选择、体细胞高频变异、克隆抑制、抗体循环补充策略。实验结果表明,免疫克隆算法在解决容量受限的工厂选址问题上,能快速收敛于全局最优解,克服了遗传算法易陷入局部最优解和收敛速度慢的缺点。

关键词:克隆;免疫算法;工厂选址

中图分类号: TP18 **文献标志码:** A

Study on capacitated plant location problem based on immune clone algorithm

QI Yang, QIN Zi-xuan, CHEN Xia, YU Zhong-hua

(College of Computer Science, Sichuan University, Chengdu Sichuan 610065, China)

Abstract: The Capacitated Plant Location Problem (CPLP) is a NP problem and hard to get an optimal and satisfactory solution. Regarding the shortage of the existing solutions, an Immune Clone (IC)-based algorithm was presented in this paper. The algorithm adopted different coding, antibody generation, clonal selection, somatic hypermutation, clonal restraint, and antibody supplement strategies from the general IC algorithm. Experimental results show that the IC-based algorithm could converge to global optimal solution rapidly and solve the Capacitated Plant Location Problem effectively. It can avoid the convergence to the local optimal solution and overcome the slow convergence of the Genetic Algorithm (GA).

Key words: clone; immune algorithm; plant location

0 引言

工厂地理位置不仅对企业当前的生产经营发挥重要作用,也对企业未来的战略决策产生重要影响。好的工厂选址可以优化建厂的固定运营成本,同时为企业节省大量的运输费用。工厂选址问题实际上是固定费用问题的变型,可以描述为:有 n 个接收站,供应来自 m 家潜在的工厂或仓库,目标是选择开办工厂的位置和产品的分配,以最小化总成本,包括固定成本和源点与终点之间的运输成本。工厂选址问题属于 NP-难问题。

由于建厂费用有容量约束的工厂选址问题 (Capacitated Plant Location Problem, CPLP) 在公共事务、制造业、商业、物流、供应链、电信网络、交通等领域应用广泛,因此国内外学者对其进行了大量的研究。文献[1]针对不确定环境下的工厂选址问题提出了一个综合的优化决策方法,并设计了 Bender 分解算法。Benders 分解算法类似于拉格朗日算法,属于传统的数学寻优方法,求解过程较为复杂,计算量大,效率不高。文献[2]开发了针对容量限制的工厂选址问题的分支定界算法。分支定界算法的效率很大程度上依赖线性规划问题如何被快速地求解。文献[3]将蚁群系统和拉格朗日启发式算法进行组合,对单源的容量限制工厂选址问题进行了研究。尽管启发式方法一般能产生优秀的解,但是它们依赖大量的数据。文献[4]提出了基于两种表示方案的遗传算法,用于求解具有单个源约束的容量限制工厂选址问题。遗传算法尽管能很好地解决传统数学寻优中的效率问题,避免了大量复杂

的计算,但存在早熟收敛和收敛速度慢的问题。为此,本文提出了基于免疫克隆原理的工厂选址算法,并进行了对应的实验验证。结果表明,在解决容量受限的工厂选址问题上,免疫克隆算法能快速收敛于全局最优解,克服了遗传算法易陷入局部最优解和收敛速度慢的缺点。

1 容量受限工厂选址问题的数学模型

CPLP 旨在使用最小的总成本,包括生产、运输和工厂定址的固定成本,确定工厂的最佳位置。形式化地,假设有 m 个源点 (或设备地址) 为 n 位顾客生产单一的产品,每位需求量为 $b_j (j = 1, 2, \dots, n)$ 个单位,如果某个特殊源点 i 是开放的 (或装有设备),则它存在固定成本 $d_i \geq 0$ 和相关的生产能力 $a_i > 0$ 。从源点 i 到顾客 j 也有一个正的单位运输成本 c_{ij} 。目标是在满足生产能力限制和需求的情况下,确定工厂的最佳位置,以使总成本最小。CPLP 可以归结为如下的混合整数规划问题^[5]:

$$\text{CPLP: } \min z(x) = \sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij} + \sum_{i=1}^m d_i y_i \quad (1)$$

$$\text{s. t. } \sum_{i=1}^m x_{ij} = b_j; j = 1, 2, \dots, n \quad (2)$$

$$\sum_{j=1}^n x_{ij} \leq a_i y_i; i = 1, 2, \dots, m \quad (3)$$

$$x_{ij} \geq 0; \forall i, j \quad (4)$$

$$y_i = 0 \text{ 或 } 1; i = 1, 2, \dots, m \quad (5)$$

其中:变量 x_{ij} 和 y_i 分别表示从工厂 i 到仓库 j 的运输量,工厂

收稿日期:2008-07-14;修回日期:2008-10-07。

作者简介:漆杨(1983-),男,四川成都人,硕士研究生,主要研究方向:数据挖掘、自然语言处理; 秦子玄(1983-),女,四川阆中人,硕士研究生,主要研究方向:数据挖掘、自然语言处理; 陈霞(1984-),女,四川眉山人,硕士研究生,主要研究方向:数据挖掘、自然语言处理; 于中华(1967-),男,黑龙江齐齐哈尔人,副教授,博士,主要研究方向:数据挖掘、自然语言处理。

是开放的(或定址)($y_i = 1$)或关闭的($y_i = 0$)。

2 基于免疫克隆的工厂选址算法

2.1 免疫克隆算法框架

免疫克隆原理是用来描述免疫应答基本特征的理论,该理论认为只有那些可以识别抗原的细胞才会被选择,发生增殖,而那些不能识别抗原的则不会被选择。被选择的细胞受制于亲和成熟过程。巴西 Campinas 大学的 De Castr 博士基于免疫克隆基本原理提出了免疫克隆算法^[6],并且验证了该算法在解决复杂机器学习问题,如模式识别和多模式优化上有很好的效果。该算法具体步骤如下:

- 1) 产生初始群体 P , 包括记忆群体 M 。
- 2) 根据亲和度大小从 P 中选择 n 个最好个体。
- 3) 克隆这 n 个最好个体, 克隆的规模随亲和度的大小而改变。
- 4) 克隆后的个体产生突变, 突变概率与抗体的亲和度成正比/反比。
- 5) 在新产生的群体中重新选择一些好的个体构成记忆群体。母体中的一些个体被新群体中的其他好于母体的个体取代。
- 6) 插入新个体替换 n 个低亲和度个体, 以保持群体多样性。

2.2 设计与改进

1) 编码。

现有的编码方法可以分为三大类: 二进制编码方法, 浮点数(真值)编码方法, 符号编码方法^[7]。考虑到运输问题中数据及约束条件的复杂性, 同时为了表示上的直观, 采用浮点数(真值)编码方法。具体地, 由于假定货物具有不可分割性, 因此采用整数编码方法。

2) 抗体产生。

常见的抗体表示方法包括线性表示法、矩阵表示法及 Prüfer 数生成树等。考虑到展现的直观性, 本文采用矩阵表示形式。

随机产生的抗体, 其基因值可能违反问题的约束条件, 例如, 个体 $v[i][j]$ 为工厂 i 提供给客户 j 的货物量, 它必须不大于工厂 i 的供货量 $a[i]$ 和客户 j 的需求量 $b[j]$, 而且还与 $v[i][j]$ 之前个体已经提供的货物量有关。为了保证随机产生的抗体满足给定的约束条件, 采用如下算法来生成抗体矩阵:

```

input: array  $a[m], b[n]$ 
output: an array( $v$ ) $ij$  such that  $v_i \geq 0$  for all  $i$  and  $j$ 
 $\sum_{j=1}^n v_{ij} = a[i]$  for  $i = 1, 2, \dots, m$  and
 $\sum_{i=1}^m v_{ij} = b[j]$  for  $j = 1, 2, \dots, n$ 
procedure initialization;
begin
  set all numbers from 1 to  $n * m$  as unvisited
  repeat
    select an unvisited random number  $q$ 
    from 1 to  $n * m$  and set it as visited
    set (row)  $i = \lfloor (q - 1) / n \rfloor$ 
    set (column)  $j = (q - 1) \bmod n$ 
    set  $val = \min(a[i], b[j])$ 
    set  $v_{ij} = val$ 
    set  $a[i] = a[i] - val$ 
    set  $b[j] = b[j] - val$ 
  until all numbers are visited.
end

```

3) 亲和度计算。

对于某个抗体对应的选址方案, 我们通过亲和度来判断其优劣, 也就是要计算其总成本目标函数值。总成本目标函数值为 $F(i)$, 为了将目标函数求极小值问题转化为求极大值问题, 定义亲和度函数 $aff(i)$ 为:

$$aff(i) = \frac{1}{F(i)} \quad (6)$$

4) 克隆选择。

在一般免疫克隆算法的第二步和第三步, 如果整个群体亲和度都很高, 都被选择克隆, 那么搜索过程就蜕变成贪婪搜索。为了避免贪婪搜索, 将每次循环开始抗体群 P 分割为固定大小的两部分 $P = P_{\text{优}} \cup P_{\text{差}}$, 每次对固定规模的 $P_{\text{优}}$ 进行克隆操作。克隆规模函数定义如下:

$$N_c = \sum_{i=1}^{N_{\text{优}}} \text{round}\left(\frac{\beta \times N_{\text{优}}}{i}\right) \quad (7)$$

得到新抗体群 $P_{\text{优}}$, 其中 $N_{\text{优}}$ 为 $P_{\text{优}}$ 抗体群的抗体数, β 为克隆系数, 这两个参数均根据抗体群 P 的大小人为设定为一定的值。 N_c 为克隆后新抗体群的抗体总数。

5) 抗体变异。

抗体变异通常是单个抗体按照一定的突变概率随机选取抗体中的一个或多个点, 将这些点上的基因抽出, 再插入原抗体随机的某个位置, 形成新的抗体。对于矩阵形式的抗体, 这种方法没有充分考虑抗体中各基因之间的关联。为此, 本文采取的策略是: 随机从原矩阵抽出一定的行和列形成一个子矩阵, 对子矩阵进行合法变异之后, 再返还给原矩阵对应的基因位。

假定 $\{i_1, i_2, \dots, i_p\}$ 为 $\{1, 2, \dots, m\}$ 的子集, 且 $\{j_1, j_2, \dots, j_q\}$ 为 $\{1, 2, \dots, n\}$ 的子集, 有 $2 \leq p \leq m, 2 \leq q \leq n$, 以 $(m \times n)$ 矩阵 $V = (v_{ij})$ 表示变异的一个亲体。那么可以从矩阵 V 的全体元素按照下面的方式产生 $(p \times q)$ 子矩阵 $W = (w_{ij})$: 元素 $v_{ij} \in V$ 并且在 W 中, 当且仅当 $i \in \{i_1, i_2, \dots, i_p\}$ 及 $j \in \{j_1, j_2, \dots, j_q\}$ (如果 $i = i_r$ 及 $j = j_s$, 那么元素 v_{ij} 被放置在矩阵 W 的第 r 行、第 s 列)。

现在可以为矩阵 W 分配对应的约束 $a_w[i]$ 和 $b_w[j]$ ($1 \leq i \leq p, 1 \leq j \leq q$):

$$a_w[i] = \sum_{j \in \{j_1, j_2, \dots, j_q\}} v_{ij}; 1 \leq i \leq p \quad (8)$$

$$b_w[j] = \sum_{i \in \{i_1, i_2, \dots, i_p\}} v_{ij}; 1 \leq j \leq q \quad (9)$$

这里使用前文的初始矩阵生成算法来为矩阵 W 分配新值, 约束 $a_w[i]$ 和 $b_w[j]$ 分别表示子矩阵 W 对应行和列元素值之和, 是取自矩阵 V 约束 $a[i]$ 和 $b[j]$ 的一部分。子矩阵 W 的变异必须满足这两个约束意味着变异后的子矩阵 W' 的行列元素值之和保持不变, 再返回矩阵 V 对应位置, 那么新矩阵 V' 的行列元素值之和也就不会改变, 必然保持了原矩阵 V 的约束 $a[i]$ 和 $b[j]$ 。

6) 克隆抑制。

若变异后大量抗体都比母体好, 那么母体群中很多个体都会被取代, 这样会出现群体多样性差的问题。为此, 选取固定量的突变后优秀个体替换母体等量抗体。

7) 抗体补充。

一般免疫克隆算法的替换补充操作没有很好地考虑群体多样性的问题, 仅仅替换低亲和度抗体, 而亲和度相同或者相似的抗体却没有考虑。本文采用克隆抑制策略同时考虑替换低亲和度以及亲和度相同或者相似的抗体, 以更好地保持群体多样性。

2.3 算法描述

基于上述分析、设计和改进, 基于免疫克隆的容量受限工

厂选址算法描述如下:

- 1) 初始化抗体群 P , 包括记忆单元 M , 随机产生 N 个抗体。
- 2) 计算亲和度并对抗体进行排序, 将排序后的抗体群分割为 $P_{\text{优}} \cup P_{\text{差}}$ 。
- 3) 对选出的 $P_{\text{优}}$ 抗体群进行克隆, 得到规模为 N_c 的抗体群 $P_{\text{优}c}$ 。
- 4) 克隆后的 $P_{\text{优}c}$ 抗体群进行子矩阵变异和克隆删除, 得到规模为 N_e 的抗体群 $P_{\text{优}e}$ 。
- 5) 在新产生的群体中重新选择一些好的个体构成记忆群体。母体中的固定量的个体被新群体中的其他好于母体的等量个体取代, 即 $P_{\text{优}e}$ 和 $P_{\text{差}}$ 进行替代合并, 形成规模为 N 的抗体群 P' 。
- 6) 插入新个体替换 N_i 个低亲和度以及亲和度相同或者相似的抗体, 形成规模为 N 的抗体群 P'' , 以保持群体多样性。

- 7) 判断终止条件, 若达到迭代次数则停止, 否则返回第 2 步。

3 实例和结果分析

3.1 实例

为了验证算法的有效性, 采用文献[8]介绍的实例作为实验对象, 并以该文提出的算法作为比较对象。文献[8]提出基于矩阵和 Prüfer 数生成树的遗传算法 (m-GA 和 st-GA), 实验参数设置为: 5×10 问题规模, 抗体表示采用矩阵及 Prüfer 数生成树, 种群大小为 100, 变异概率 0.4, 杂交概率 0.2, 最大迭代次数 2000, 计算运行 10 次。本文算法 (m-ICA) 的参数设置为: 5×10 问题规模, 抗体表现采用矩阵。种群大小为 100, $P_{\text{优}}$ 规模选定为 15, β 取值 0.9, 子矩阵大小选定为 3×5 , 最大迭代次数 2000, 计算运行 10 次。表 1 给出了 5×10 测试问题的随机取样数据。

表 1 5×10 测试问题随机取样数据

工厂 i	固定成本 d_i	交货成本 c_{ij}										供应量 a_i
1	304	10	9	8	7	4	6	4	7	8	4	1113
2	381	9	3	6	9	6	9	8	4	4	5	1247
3	459	3	5	8	7	3	10	10	8	5	4	902
4	292	6	6	3	10	8	7	4	4	6	3	912
5	241	10	10	9	4	8	5	3	10	4	6	826
需求量	b_j	729	630	321	293	251	573	207	732	481	783	

3.2 结果分析

表 2 给出了实验结果对比, 其中标记“Freq.”为找到最优解的次数。解的质量定义如下:

$$\text{解的质量} = \frac{\text{最优解}}{\text{得到的解}} \times 100 \quad (10)$$

从表 2 可以发现, m-ICA 可以稳定地得到问题的最优解, 而且 m-ICA 也比 m-GA 更有可能进化到最优解或近似最优解。该最优解为: 最小总成本 20442, 其中运费成本 18765, 工厂开放固定成本 1677 (即所有工厂全部开放)。

表 2 m-GA 和 m-ICA 之间解质量的比较 (问题规模为 5×10)

算法	最坏	平均	最优	Freq
m-GA	92.43	97.02	100.00	2
m-ICA	100.00	100.00	100.00	10

图 1 给出了 m-ICA 的进化过程曲线, 从中可见, m-ICA 在 50 多代已经基本稳定到最优解。而根据文献[8], 同样的实验数据, st-GA 需要在约 500 代的时候才能稳定到最优解质量。因此, m-ICA 有更优的收敛速度和收敛可靠性。

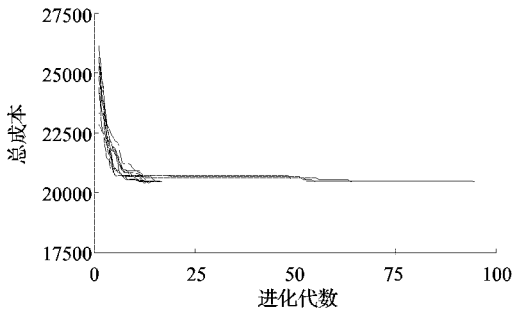


图 1 10 次 m-ICA 进化过程示意图

4 结语

本文提出了基于免疫克隆的容量受限工厂选址算法, 该

算法针对一般免疫克隆算法的不足, 采用了不同的编码、抗体产生、克隆选择、体细胞高频变异、克隆抑制、抗体循环补充策略。实验结果表明, 所提出的算法能够在较少的代数内快速地收敛到全局最优解, 避免了遗传算法存在的早熟收敛和收敛速度慢的缺点。

下一步工作的重点将是对该算法的全局收敛性和收敛速度进一步地进行理论分析和其他实践检验, 以保证此算法的通用有效性。此外, 还将考虑采用 Prüfer 数生成树作为抗体表现形式以节省空间, 进一步提高运算效率。

参考文献:

- [1] REN M M, YANG C, HE B. An integrated optimal approach for facility location and its size decision with uncertain demand[J]. Systems Engineering, 2007, 25(6): 1-5.
- [2] EFROYMSON M, RAY T. A branch-bound algorithm for plant location[J]. Operations Research, 1966, 14: 361-368.
- [3] CHEN C H, TING C J. Combining lagrangian heuristic and ant colony system to solve the single source capacitated facility location problem[J]. Transportation Research Part E: Logistics and Transportation Review, 2008, 44(6): 1099-1122.
- [4] GEN, M, CHOI J O, TSUJIMURA Y. Genetic algorithm for the capacitated plant location problem with single source constraints[C]// Proc of 7th European Congress on Intelligent Techniques & Soft Computing. Aachen: [s. n.], 1999.
- [5] CORTINHAL M J, CAPTIVOV M E. Genetic algorithms for the single source capacitated location problem: A computational study [C]// 4th Metaheuristics International Conference. Porto, Portugal: MIC Press, 2001: 355-359.
- [6] CASTRO L N, ZUBENMM F J. Artificial immune systems: Part I - basic theory and applications[EB/OL]. [EB/OL]. [2007-10-10]. ftp://ftp.dca.fee.unicamp.br/pub/docs/vonzuben/tr_dca/trdca0199.pdf.
- [7] 周明, 孙树栋. 遗传算法原理及应用[M]. 北京: 国防工业出版社, 1999: 32-41.
- [8] 玄光男. 遗传算法与工程优化[M]. 北京: 清华大学出版社, 2003: 246-250.