

## 求解大规模 TSP 问题的自适应归约免疫算法\*

戚玉涛<sup>†</sup>, 刘芳, 焦李成

(西安电子科技大学 智能信息处理研究所, 陕西 西安 710071)

### Immune Algorithm with Selfadaptive Reduction for Large-Scale TSP

QI Yu-Tao<sup>†</sup>, LIU Fang, JIAO Li-Cheng

(Institute of Intelligent Information Processing, Xidian University, Xi'an 710071, China)

+ Corresponding author: E-mail: qi\_yutao@163.com

Qi YT, Liu F, Jiao LC. Immune algorithm with selfadaptive reduction for large-scale TSP. *Journal of Software*, 2008,19(6):1265-1273. <http://www.jos.org.cn/1000-9825/19/1265.htm>

**Abstract:** Analyses on factors which impact the performance of the multi level algorithms have been made, and on the basis of which an immune algorithm with selfadaptive reduction has been proposed for the TSP problems. By using an evolutionary reduction set, the proposed algorithm refines the reduction edges which gradually increase in the number and enhance in the forecasting accuracy. As a result, the probability that the refined algorithm finds the global optimal solution can be improved. Experimental results show that the proposed algorithm can achieve better solutions than other approaches.

**Key words:** TSP; artificial immune; clonal selection; selfadaptive reduction

**摘要:** 从理论上分析了影响多级算法性能的因素,并以此为依据构造了求解 TSP 问题的自适应归约免疫算法. 该算法借助归约集的进化使归约集规模自适应增长,归约边的预测精度不断提高,从而提高了算法在归约后找到全局最优解的概率. 实验结果表明,该算法比其他算法获得了质量更高的解.

**关键词:** TSP; 人工免疫; 克隆选择; 自适应归约

中图法分类号: TP301 文献标识码: A

旅行商问题(TSP)由于其广泛的应用价值长期以来一直吸引着许多研究者的关注. 目前, TSP 问题在交通控制、网络路由、VLSI 布图、机器人控制等方面有着广泛的应用背景.

TSP 问题可以形式化描述为: 已知  $N$  个城市  $C = \{C_1, C_2, \dots, C_N\}$ , 以及任意两城市之间的距离  $d(C_i, C_j)$ , 求一条经过  $C$  中所有城市一次且仅一次的闭合路径  $C_\pi = \{C_{\pi(1)}, C_{\pi(2)}, \dots, C_{\pi(N)}\}$ , 使得总行程  $\sum_{i=1}^{N-1} d(C_{\pi(i)}, C_{\pi(i+1)}) + d(C_{\pi(N)}, C_{\pi(1)})$  最小<sup>[1]</sup>.

TSP 问题属于 NP-hard 组合优化问题, 难以找到在多项式计算时间内得到最优解的有效算法, 因此, 对于大规模的问题人们更倾向于寻求在有限的时间内找到可接受的近似解算法. 求解 TSP 问题的近似算法分为两类:

\* Supported by the National Natural Science Foundation of China under Grant Nos.60575037, 60502043 (国家自然科学基金); the National High-Tech Research and Development Plan of China under Grant No.2006AA01Z107 (国家高技术研究发展计划(863))

Received 2007-05-14; Accepted 2007-11-05

环路构造算法和环路改进算法.环路构造算法从某个非法解出发,逐步改变路径,直到得到一个合法路径为止.这类算法包括:最近邻算法、贪心算法、Clarke-Wright 算法、Christofides 算法等<sup>[1]</sup>.环路改进算法则在给定初始合法解之后,使用某种策略寻找质量更好的解.这类算法包括:局部搜索策略(2-Opt,3-Opt,LK,LKH,循环 LK<sup>[2]</sup>等)、禁忌搜索<sup>[1]</sup>、模拟退火<sup>[1]</sup>、遗传算法<sup>[3]</sup>、蚁群算法<sup>[4]</sup>、粒子群算法<sup>[5]</sup>、多级算法<sup>[6]</sup>、多级归约算法<sup>[7]</sup>、免疫算法<sup>[8]</sup>等.

在求解大规模 TSP 问题中,多级(multilevel)算法取得了相当大的成功<sup>[6,7]</sup>.多级的思想是以一定的策略确定路径中的部分边,将多个城市归约为一个节点,并将被归约城市与外界城市间的约束边转移到此节点上,从而把原问题转化为较小规模的 TSP 问题求解.如此逐层归约,直到问题的规模降到足够小,能够用已有的算法迅速求解为止<sup>[9]</sup>.每一层归约的结果是对上一层问题的近似逼近.归约完成之后,迅速求解最后一层近似问题,然后进行反向逐层细化.细化过程将下一层问题的解作为上一层问题的初始解,如此逐层细化直到第 1 层,即问题为原问题为止<sup>[10]</sup>.根据这一思想,Chris Walshaw 用 LKH 算法作为细化算法构造了多级 LKH 算法(MLLKH)<sup>[6]</sup>,该算法将搜索的重点放在了细化步骤上.邹鹏等人利用多个局部最优路径的交集对最优路径中的边进行估计,将搜索放在归约步骤之前,构造了求解 TSP 问题的多级归约算法(MR)<sup>[7]</sup>.与多级 LKH 算法不同,该算法将计算的重点放在了归约城市的选择上.无论是自下而上细化的 MLLKH 算法,还是自上而下细化的 MR 算法,多级算法可以分为两个步骤:归约和细化.

本文从理论上分析了影响多级算法性能的因素,并以自上而下细化的多级算法 MR 为例分析了多级算法中存在的问题.针对这些问题,我们设计了求解 TSP 问题的自适应归约免疫算法(immune algorithm with self-adaptive reduction for TSP,简称 SRIA).经过分析我们发现,MR 算法对 TSP 问题的每一次归约都会迅速降低算法找到全局最优解的概率,尤其当问题规模很大时,MR 算法在归约后找到全局最优解的概率几乎为 0.为了克服这一缺点,SRIA 放弃了多级算法逐层归约的思想,每次迭代都借助当前候选解种群的一个随机子集重新构造归约边集合对原始 TSP 问题进行新的一级归约.候选解种群的迭代进化带来了归约边集合的进化,而归约边集合的进化正是 SRIA 的核心思想.与 MR 算法不同,SRIA 的重点不是借助逐层归约不断扩大归约边集合的规模以达到降低 TSP 问题规模的目的,而是通过归约边集合的进化提高算法在归约后找到全局最优解的概率.实验结果表明,随着迭代的进行,归约边集合不断进化,归约边集合的规模自适应地增长,归约边集合中的边在全局最优解路径中出现的概率逐渐递增并趋近于 1.归约边集合的进化使得归约后算法找到全局最优解的概率始终保持在一一定的水平并呈递增趋势.

## 1 人工免疫系统和克隆选择算法

人工免疫系统(artificial immune system,简称 AIS)是模仿生物免疫系统的免疫应答、免疫记忆、免疫调节等机理,构造出的一类高性能、自组织、鲁棒性好的人工智能系统.目前,人工免疫系统的研究已经受到学者们越来越广泛的关注,人工免疫算法也在实际工程应用中得到了推广.人工免疫成为继神经网络、模糊逻辑和进化计算后人工智能领域的又一研究热点<sup>[11]</sup>.

1958 年,Burnet 等人提出了著名的免疫克隆选择学说.克隆选择原理认为,淋巴细胞除了扩增或者分化成浆细胞以外,也能分化成生命期较长的 B 记忆细胞.当再次遇到相应的抗原时,记忆细胞将预先被免疫系统选择出来,并迅速活化、增殖、分化为效应细胞,执行高效而持久的免疫功能.生物免疫过程中所体现出的学习、记忆、抗体多样性等特性,正是人工免疫系统所借鉴的.受免疫克隆选择学说启发,1995 年,Hunt 等人提出了第一个人工免疫系统模型——B 细胞网络模型<sup>[12]</sup>.2000 年,De Castro 等人提出了一种较为简洁的克隆选择算法(CSA),并成功地用于解决模式识别、数值优化和 TSP 问题<sup>[13]</sup>,建立了克隆选择机理用于随机搜索算法的框架.2000 年,王磊等人在遗传算法中引入了免疫算子,构造了一类免疫遗传算法(IGA),并用于求解 TSP 问题<sup>[8]</sup>.

本文提出的 SRIA 算法从一个初始记忆 B 细胞网络出发,当受到抗原刺激时,适应度高的记忆 B 细胞会克隆增殖,产生抗体种群.抗体种群中的抗体经历若干次克隆选择进化操作之后,适应度最大的抗体会分化成新的记忆细胞加入记忆 B 细胞网络中去.为保持记忆 B 细胞网络规模稳定,长期没有受到抗原刺激的记忆 B 细胞会

逐渐消亡.已有的免疫克隆选择算法对 B 细胞和抗体不作区别,而在 SRIA 中两者有所不同.SRIA 中,记忆 B 细胞是对原问题的解的编码,而抗体是对当前归约前提下的逼近问题的解的编码.

SRIA 采用城市的遍历次序作为抗体的编码,记忆 B 细胞  $\mathbf{B}=(b_1, b_2, \dots, b_n)$  ( $n$  为原 TSP 问题的城市数,  $b_i$  为城市编号)的亲合度定义为式(1).其中,  $L$  为包含所有城市的最小正方形边长,  $\text{length}(\mathbf{B})$  为 B 细胞对应路径长度.抗体  $\mathbf{A}$  的亲合度记为  $\text{aff}(\mathbf{A})$ .求抗体的亲合度时,将抗体对应的逼近问题的解恢复到原问题再用式(1)求得.

$$\text{aff}(\mathbf{B}) = 76.5 \times L \times \sqrt{d} / \text{length}(\mathbf{B}) \tag{1}$$

## 2 多级算法分析

邹鹏等人提出的求解 TSP 问题的多级归约算法(MR)<sup>[7]</sup>是一种典型的多级算法,我们以 MR 为实例分析影响多级算法性能的因素.邹鹏等人研究发现,在用 LK 算法获得的局部最优解路径中,80%的边将在全局最优解中出现,而且这个概率和问题的规模无关.他们还发现,多个局部最优解路径的交集能够更加准确地估计全局最优路径中的边,随着局部最优解数量的增加,交集中边的个数缓慢减少,同时,全局最优解路径中的边所占的比例迅速增长,并很快趋近于 1.基于这个统计结果,他们提出了 MR 算法.MR 用 LK 算法得到的若干个局部最优解路径的公共边集合来估计全局最优解路径中的边,每一次归约将当前公共边集合中的边固定,把问题的规模降低,搜索在剩下的子空间中进行,直到问题能够用成熟的算法得到较好的解为止.最后,按照归约的反序逐步拼接固定的边,最终得到一个合法的解.影响 MR 算法性能的因素有 3 个方面:对全局最优解路径中边的预测精度,归约边集合的规模和局部搜索算法的搜索性能.为了便于分析,我们引入如下概念.

**定义 1.** 在一次归约中被固定下来的边的集合  $E_C = \{e_1, e_2, \dots, e_m\}$  称为该次归约的归约集,  $e_i (i = 1, 2, \dots, m)$  称为归约边.归约集  $E_C$  中归约边的个数  $m$  称为该次归约的归约度.用  $E_C$  对 TSP 问题进行的归约操作称为归约集是  $E_C$  的  $m$  归约,简称  $m$  归约.

规模为  $n$  的 TSP 问题,其解空间大小为  $n!/2$ .经过一次  $m$  归约后,解空间缩小为  $(n-m)!/2$ .图 1 描述了归约度  $m$  对搜索空间规模的影响.对于 100 个城市的 TSP 问题,解空间在  $10^{157}$  数量级,当归约度  $m=n/2$  时,子问题的搜索空间下降到  $10^{64}$  数量级.可见,归约算法能够有效地缩小解空间的规模.

**定义 2.** 归约边  $e_i$  在全局最优解路径中出现的概率  $P(\theta_i)$  称为归约边  $e_i$  的预测精度,  $\varepsilon_i = 1 - P(\theta_i)$  称为归约边  $e_i$  的预测误差.  $\bar{\varepsilon} = \{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m\}$  称为归约集  $E_C$  的预测误差向量.

**定义 3.** 归约前 TSP 问题的解空间称为原空间,在固定了  $m$  个归约边之后,TSP 问题的解空间称为原空间  $m$  归约后的自由搜索子空间.如果在原空间中搜索到全局最优解路径的概率为 1,那么经过一次归约集为  $E_C$  的  $m$  归约后,算法在自由搜索子空间中能够搜索到全局最优解路径的概率  $P(\theta)$  为

$$P(\theta) = \prod_{i=1}^m P(\theta_i)P(\theta|\theta_1, \theta_2, \dots, \theta_m) \tag{2}$$

其中,  $P(\theta|\theta_1, \theta_2, \dots, \theta_m)$  是  $E_C$  中的边全部出现在全局最优解路径中的前提下,算法搜索到全局最优解路径的条件概率.称  $P_m = \prod_{i=1}^m P(\theta_i)$  为自由搜索子空间的自由度.在实际计算中,我们统计归约集  $E_C$  的统计预测误差记为  $\varepsilon$ , 则  $P(\theta_i) = 1 - \varepsilon_i$ ,  $P_m = (1 - \varepsilon)^m$ .由于  $0 < P(\theta|\theta_1, \theta_2, \dots, \theta_m) \leq 1$ , 故有  $P(\theta) \leq P_m$ .因此,自由度  $P_m$  描述了原 TSP 问题经过一次  $m$  归约后算法能够搜索到全局最优解路径的概率上限.

图 2 描述了自由度  $P_m$  和归约边统计预测误差  $\varepsilon$  的数量级以及归约度  $m$  之间的关系.从图中可以看出,当归约度  $m$  的值增大时,  $m$  归约后自由搜索子空间的自由度迅速下降.例如,当  $m$  的值为  $10^4$  数量级时,如果  $E_C$  中归约边统计预测误差  $\varepsilon$  大于  $10^{-3}$ , 那么算法搜索到全局最优解路径的概率接近于 0.当归约边统计预测误差在  $10^{-4}$  数量级时,算法能以 36%左右的概率上限找到全局最优解.即使归约边统计预测误差在  $10^{-5}$  数量级,算法找到全局最优解的概率上限也只能达到 90%左右.

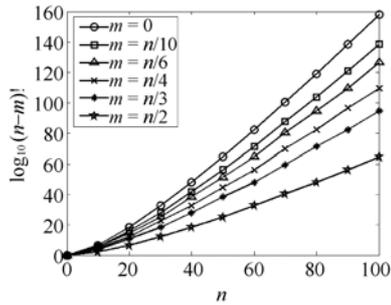
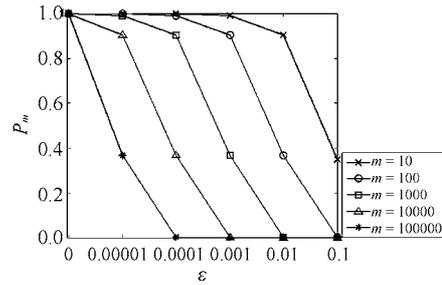
Fig.1 Effect of  $m$  on searching space图 1  $m$  对搜索空间的影响Fig.2 Effect of  $m$  on the degree of freedom after reduction图 2  $m$  归约后自由搜索子空间的自由度

表 1 统计了 MR 在求解 TSPLIB 中已知最优路径的 3 个实例,进行第 1 次归约操作的平均归约度  $m$ ,统计预测误差  $\varepsilon$  和经过  $m$  归约后自由子空间的平均自由度.表中数据为 10 次独立实验的平均结果.可以看出,通过增加局部最优路径的个数能够降低归约边的统计预测误差,但是付出的计算代价却很难换取预测误差跨数量级的降低.MR 算法进行第 1 级归约操作后,自由搜索子空间的自由度接近 0.也就是说,MR 的第 1 次归约就使得算法以极大的概率陷入局部最优.

Table 1 The average degree of freedom of subspace after the first reduction in MR

表 1 MR 第 1 次归约后自由搜索子空间的平均自由度

Instance (scale)	Parameters	Number of local optimal tours for intersecting				
		2	5	10	15	20
att532 (532)	$m$	466	424	387	373	369
	$\varepsilon$	0.924 4	0.975 2	0.982 6	0.991 4	0.992 5
	$P_m$	$1.2323 \times 10^{-16}$	$2.3753 \times 10^{-5}$	$1.1 \times 10^{-3}$	$3.99 \times 10^{-2}$	$6.22 \times 10^{-2}$
pcb1173 (1173)	$m$	1 092	982	946	938	930
	$\varepsilon$	0.952 3	0.986 4	0.994 1	0.994 7	0.995 3
	$P_m$	$6.6217 \times 10^{-24}$	$1.4458 \times 10^{-6}$	$3.7 \times 10^{-3}$	$6.8 \times 10^{-3}$	$1.25 \times 10^{-2}$
fl3795 (3795)	$m$	3 072	2 456	2 244	2 118	2 003
	$\varepsilon$	0.847 7	0.923 9	0.949 0	0.960 5	0.972 4
	$P_m$	$3.6307 \times 10^{-221}$	$3.7577 \times 10^{-85}$	$9.669 \times 10^{-52}$	$8.5002 \times 10^{-38}$	$4.5023 \times 10^{-25}$

除去细化搜索算法的搜索能力以外,归约集  $E_C$  的归约度  $m$  和  $E_C$  的预测误差向量  $\bar{\varepsilon}$  是影响多级算法性能的两个重要因素. $m$  和  $\bar{\varepsilon}$  共同决定了  $m$  归约后自由子空间的自由度  $P(\theta)$ ,而  $P(\theta)$  的值制约了细化搜索算法在自由子空间中对解的性能改进的程度.在  $m$  一定的前提下, $\bar{\varepsilon}$  各分量的值越小, $P(\theta)$  越大,细化搜索算法越可能找到全局最优路径.在  $\bar{\varepsilon}$  各分量数量级一定的条件下, $m$  越大, $P(\theta)$  的值越小,细化搜索算法的搜索效率就越低; $m$  越小,自由子空间的规模越大,细化搜索算法的搜索空间规模就越大.因此,要改进多级算法求解大规模 TSP 问题的能力必须从两方面入手:一,降低归约集中各个归约边的预测误差;二,找到一个合适的归约度  $m$ .

### 3 自适应归约免疫算法

从前面的分析可以看出,MR 算法存在以下不足:1) 归约集中各个归约边的预测误差不能满足精度要求;2) 归约度  $m$  的值随着多级归约的进行单调递增;3) 前两个缺陷造成了算法在归约后找到全局最优解的概率迅速下降,几乎为 0.针对 MR 算法的这些缺陷,本文在免疫克隆选择算法的框架下设计了求解 TSP 问题的自适应归约免疫算法.

#### 3.1 算法描述

SRIA 从规模为  $s$  的记忆 B 细胞网络出发,每次迭代由部分记忆细胞产生归约集对原问题进行归约,并产生规模为  $p$  的新抗体种群在归约后的自由子空间中进行克隆选择进化搜索.搜索结束后,将当前抗体种群中的最

优抗体恢复成 B 细胞,更新记忆 B 细胞网络.如此反复,直到算法终止,条件满足.SRIA 的流程如下:

随机初始化记忆 B 细胞网络  $\mathbf{M}(0) = \{\mathbf{B}_1(0), \mathbf{B}_2(0), \dots, \mathbf{B}_s(0)\}, t=0$ ;  
 While (算法终止条件不满足) do  
     轮盘赌策略从  $\mathbf{M}(t)$  中选择  $d\%$  个记忆 B 细胞构成集合  $\mathbf{C}(t)$ ;  
     求  $\mathbf{C}(t)$  中所有记忆 B 细胞对应路径的公共边集合,产生归约集  $\mathbf{E}_c(t)$ ;  
     用  $\mathbf{E}_c(t)$  归约原 TSP 问题,生成新问题;  
     细化搜索:  
         贪婪策略初始化抗体种群  $\mathbf{P}(0) = \{\mathbf{A}_1(0), \mathbf{A}_2(0), \dots, \mathbf{A}_p(0)\}$ ;  
         While (细化搜索算法停止条件不满足) do  
             克隆操作:  $\mathbf{P}'(r) = \text{CL}(\mathbf{P}(r))$ ;  
             免疫基因操作:  $\mathbf{P}''(r) = \text{IG}(\mathbf{P}'(r))$ ;  
             克隆选择操作:  $\mathbf{P}(r+1) = \text{SL}(\mathbf{P}''(r))$ ;  
         End While  
     更新记忆 B 细胞网络,得到  $\mathbf{M}(t+1)$ , 令  $t=t+1$ ;  
 End While

记忆 B 细胞网络为  $\mathbf{M}(t) = \{\mathbf{B}_1(t), \mathbf{B}_2(t), \dots, \mathbf{B}_s(t)\}$ .原 TSP 问题规模为  $n$ ,  $\mathbf{E}_c(t)$  规模记为  $m$ ,则经过归约操作后,新的 TSP 问题的规模为  $n-m$ .在新 TSP 问题解空间上搜索的抗体种群记为  $\mathbf{P}(r) = \{\mathbf{A}_1(r), \mathbf{A}_2(r), \dots, \mathbf{A}_p(r)\}$ .

克隆操作:克隆操作 CL 定义为

$$\text{CL}(\mathbf{P}(r)) = [\text{CL}(\mathbf{A}_1(r)), \text{CL}(\mathbf{A}_2(r)), \dots, \text{CL}(\mathbf{A}_p(r))]^T \quad (3)$$

其中,  $\text{CL}(\mathbf{A}_i(r)) = \mathbf{I}_i \times \mathbf{A}_i(r)$  ( $i=1, 2, \dots, p$ ,  $\mathbf{I}_i$  为  $q_i(r)$  维单位行向量)称为抗体  $\mathbf{A}_i(r)$  的  $q_i(r)$  克隆.  $q_i(r)$  的值与  $\mathbf{A}_i(r)$  的亲合度有关,  $\mathbf{A}_i(r)$  亲合度越大,克隆数量  $q_i(r)$  越大.令  $\mathbf{Y}'_i(r) = \text{CL}(\mathbf{A}_i(r)) = \{y'_{i1}(r), y'_{i2}(r), \dots, y'_{iq_i(r)}(r)\}$ , 则克隆后的种群可以表示为

$$\mathbf{P}'(r) = \{\mathbf{Y}'_1(r), \mathbf{Y}'_2(r), \dots, \mathbf{Y}'_p(r)\} \quad (4)$$

免疫基因操作:免疫基因操作 IG 作用在  $\mathbf{P}'(r)$  中所有抗体之上,SRIA 采用 Inver-over 算子<sup>[3]</sup>进行免疫基因操作.记  $\mathbf{Y}''_i(r) = \text{IG}(\mathbf{Y}'_i(r)) = \{y''_{i1}(r), y''_{i2}(r), \dots, y''_{iq_i(r)}(r)\}$ , 则 IG 操作后的抗体种群记为

$$\mathbf{P}''(r) = \{\mathbf{Y}''_1(r), \mathbf{Y}''_2(r), \dots, \mathbf{Y}''_p(r)\} \quad (5)$$

克隆选择操作:免疫基因操作后的抗体  $\mathbf{A}_i = \{y''_{ij}(r) \mid \max \text{aff}(y''_{ij}(r)), j=1, 2, \dots, q_i(r)\}$ , 其中,  $\text{aff}(y''_{ij}(r))$  表示抗体  $y''_{ij}(r)$  的亲合度.若  $\text{aff}(\mathbf{A}_i) > \text{aff}(\mathbf{A}_i(r))$ , 则用  $\mathbf{A}_i$  替代  $\mathbf{A}_i(r)$  进入下一次迭代, 否则不做任何操作.克隆选择后的种群为  $\mathbf{P}(r+1)$ .

更新记忆 B 细胞网络操作:经过了若干次克隆选择的迭代操作,抗体种群的进化缓慢下来,此时,将抗体种群中的最优抗体取出,用 LK 算法<sup>[2]</sup>对其进行局部搜索,然后将优化后的解细化,生成记忆 B 细胞  $\mathbf{B}_{new}$ .在记忆 B 细胞网络  $\mathbf{M}(t)$  中找到与  $\mathbf{B}_{new}$  距离最近的记忆 B 细胞  $\mathbf{B}_j(t)$  ( $j=1, 2, \dots, s$ ).如果  $\text{aff}(\mathbf{B}_{new}) > \text{aff}(\mathbf{B}_j(t))$  满足, 则用  $\mathbf{B}_{new}$  替代  $\mathbf{B}_j(t)$ , 否则  $\mathbf{B}_j(t)$  继续保留在 B 细胞网络中.这里,  $\mathbf{B}_{new}$  和  $\mathbf{B}_j(t)$  之间的距离如式(6)定义.其中,正整数  $c$  是两个 B 细胞所表示两条路径的距离定义为两者公共边数量.

$$\text{Distance}(\mathbf{B}_{new}, \mathbf{B}_j(t)) = \begin{cases} 1, & c=0 \\ 1/c, & c \neq 0 \end{cases} \quad (6)$$

### 3.2 算法分析

与 MR 不同,SRIA 尝试不同的归约方式对原 TSP 问题进行一级归约,而不是像 MR 那样逐级归约.归约集进化的思想是 SRIA 的重要创新,它为 SRIA 带来了自适应特性,也使得 SRIA 能够克服 MR 算法中的几个缺陷.

如果一条边在多个优秀的 TSP 路径中出现,那么我们有理由相信,这条边很有可能会出现在全局最优解路径中.而且,参与求公共边集合的 TSP 路径质量越高,公共边在全局最优解路径中出现的可能性就越大.SRIA 在已经搜索到的 TSP 路径中随机选取一部分求公共边集合作为归约集,随着迭代的进行,构造归约集的 TSP 路径质量不断提高,因此进化着的归约集中归约边的预测精度也会逐渐提高.

TSP 路径的质量越高,其中包含的优秀边的数量越多.若干个接近全局最优解的路径之间自然有很多相似之处.在 SRIA 中,随着记忆 B 细胞网络的进化,个体解码后的 TSP 路径越来越逼近全局最优解,算法通过求公共边的方式求得的归约集规模(归约度  $m$ )也会逐渐增大.因此,归约度的大小会随着记忆 B 细胞网络的进化而自适应地改变.

SRIA 在归约集进化的初期,归约边的预测误差相对较大而归约度较小.在归约集进化的后期,归约边的预测误差有跨数量级的降低而归约度较大.从归约后自由搜索子空间自由度  $P_m$  的定义可以看出,SRIA 自适应地使  $P_m$  保持在一定的水平上,也就是说,SRIA 在归约后仍然以一定的概率找到 TSP 问题的全局最优解,从而克服了 MR 算法以极大概率陷入局部最优的缺点.后面的实验结果还表明,随着迭代的进行,SRIA 在每次归约后子空间的自由度呈上升趋势,且上升幅度明显,这正是设计 SRIA 的目标.

与多数人工免疫算法不同,SRIA 将 B 细胞和抗体区别开来.B 细胞是对原问题解的编码,而抗体是对当前归约前提下新问题解的编码.免疫淋巴细胞除了受到抗原的刺激,还受到免疫网络中邻近抗体的抑制.在更新记忆 B 细胞网络操作中,SRIA 定义了 B 细胞之间的距离测度来量化免疫网络中 B 细胞间的相互抑制.这种策略有利于保持记忆 B 细胞的多样性,从而产生预测精度更高的归约集.

SRIA 考虑了进化类算法的尾重现象(即算法在执行到一定程度后,解的性能改进出现停滞的现象)<sup>[14]</sup>,在种群进化到一定程度时采用重启策略.重启策略的设计是 SRIA 的又一重要特点,重启后的种群是在新的归约方式限制下的子空间上搜索,而不是在原来的子空间上进行重复搜索.不同的归约方式使得细化搜索算法有机会在不同的子空间上进行搜索,因而算法能够保持很好的种群多样性,不易陷入局部最优中无法摆脱.

SRIA 对抗体的免疫基因操作采用了郭涛等人提出的 Inver-over 算子<sup>[3]</sup>.已经证明,Inver-over 算子的性能优于常用的遗传操作算子,如:分映射交叉(PMX)、顺序交叉(OX)以及循环交叉(CX)<sup>[15]</sup>等.高效的细化搜索算法是记忆 B 细胞网络进化的动力,保证了整个算法的搜索性能.Inver-over 算子对于小规模 TSP 问题十分有效,然而,当问题规模较大时则收敛缓慢.SRIA 一方面用归约的方法降低问题规模以发挥 Inver-over 算子的搜索能力,另一方面用 LK 算法对抗体种群搜索到的当前最优解进行局部优化以加速收敛.因此,SRIA 的设计兼顾了算法的求解质量和收敛速度.

记 SRIA 的迭代次数为  $G$ ,细化搜索算法的平均迭代次数为  $g$ ,TSP 问题规模为  $n$ ,归约集规模为  $m$ ,记忆 B 细胞网络规模为  $s$ ,抗体种群规模为  $p$ ,对抗体种群的平均克隆规模为  $c$ .初始化记忆 B 细胞网络的复杂度为  $O(s \times n)$ ,产生归约集的复杂度为  $O(d\% \times s \times n)$ ,初始化抗体种群的复杂度为  $O(p \times (n-m))$ ,克隆操作的复杂度为  $O(c \times p \times (n-m))$ ,免疫基因操作的复杂度最差情况下为  $O(c \times p \times (n-m-1) \times (n-m-2)/2)$ ,克隆选择操作复杂度为  $O(c \times p)$ .因此,每次细化搜索的复杂度最差为  $O(0.5gcp(n-m)^2 + (p-0.5gcp)(n-m) + 2gcp)$ .SRIA 除去更新 B 细胞网络操作的复杂度为  $O(sn + Gd\%sn + 0.5Ggcp(n-m)^2 + G(p-0.5gcp)(n-m) + 2Ggcp)$ ,小于  $O(n^2)$ .而更新 B 细胞网络操作中采用的 LK 算法的复杂度大于  $O(n^2)$ .因此,SRIA 总的复杂度取决于 LK 算法的复杂度.MR 算法由于采用了 LK 算法构造局部最优解,其复杂度也大于  $O(n^2)$ .因此,理论上 SRIA 与 MR 的复杂情况大小相当,均取决于 LK 算法的复杂度.

## 4 仿真实验

实验选用了 TSPLIB 中几个具有代表性的实例,所有算法均用 VC++6.0 实现,各算法均运行在 Intel(R)

Core(TM)2 1.86GHz CPU,2GB 内存,Windows XP 操作系统的 PC 机上.实验中,记忆 B 细胞网络的种群规模为 30,抗体种群规模为 50.算法终止条件设置为:记忆 B 细胞网络的最优个体在 30 次迭代内没有发生改变时算法终止.细化搜索算法的停止条件设置为:当抗体种群的最优个体在 10 次迭代内没有发生改变时停止.每次归约之前,记忆 B 细胞种群的 60%参与归约集的构造.MR 算法中保留两个用于归约的解,当归约后问题的规模小于 100 时,结束归约并用 LK 算法求解当前实例.GT 算法的种群规模为 100,随机翻转概率为 0.02,终止条件为最优个体在 30 次迭代内没有发生改变时算法终止.

4.1 算法性能比较

该实验将 SRIA 和多级归约算法(MR)<sup>[7]</sup>、郭涛算法(GT)<sup>[3]</sup>的搜索性能进行了比较.表 2 和表 3 分别比较了 SRIA 和其他算法在中规模和大规模实例上的搜索性能.表中  $n$  为城市数量,偏差  $\sigma$  的计算公式为

$$\sigma\% = \left( \sum_{i=1}^T (S_i - S_0) \right) / (T \times S_0) \times 100\% \tag{7}$$

其中, $T$  为独立实验次数, $S_i$  为每次实验得到的最优解路径长度, $S_0$  为已知最优解路径长度.

Table 2 Performance comparisons of SRIA with MR and GT on midsize instances

表2 SRIA与MR和GT算法在中规模实例上的搜索性能比较

Instance	$n$	$S_0$	$\sigma\%$ (mean)			Average running time (s)		
			SRIA	MR	GT	SRIA	MR	GT
att532	532	27 686	0.000 3	0.001 7	0.715 2	34.12	9.53	187.42
pcb1173	1 173	56 892	0.001 5	0.004 6	1.216 6	152.36	97.16	1 894.38
vm1748	1 748	336 556	0.000 9	0.001 1	1.852 7	556.41	223.81	2 915.43
pr2392	2 392	378 032	0.023 1	0.037 8	2.573 6	837.44	176.47	9 256.74
fl3795	3 795	28 772	0.054 6	0.103 5	2.652 1	1 342.17	236.34	17 749.36
ri5934	5 934	556 045	0.064 3	0.125 4	3.948 3	1 206.17	182.37	15 812.79
Average			0.024 1	0.045 7	2.159 8	688.111 7	154.28	7 969.35

表 2 中数据为 20 次独立实验的统计结果.可以看出,SRIA 在求得解质量上优于其他两种算法.在运行时间上,SRIA 相对于 GT 有明显优势,却稍逊于 MR.虽然 SRIA 的复杂度在理论上与 MR 相当,但是理论分析得到的只是当问题规模充分大时算法复杂度的一个上界.由于 SRIA 收敛速度比 MR 要慢,因此 SRIA 耗时比 MR 要长,平均耗时约为 MR 的 4.46 倍.然而,这种时间的代价却换来了解的质量的提高.

4.2 搜索效率分析

该实验用 3 个 TSPLIB 中的实例考察了 SRIA 随着归约次数的增加归约度  $m$ 、归约集中边的统计预测误差  $\varepsilon$  以及  $m$  归约后子空间自由度的变化情况.算法迭代 100 次后停止,结果为 50 次独立实验的统计值.

图 3 是归约度  $m$  随着归约次数的增加而发生变化的情况.从图 3(a)中可以看出,随着迭代的进行,归约度  $m$  的值逐渐增大.迭代的初期增长速度较快,随着迭代的进行增长速度趋于平缓.图 3(b)是归约后问题规模和原 TSP 问题规模的比值,经过 100 次归约后 3 个实例的规模都有大幅度的降低.

图 4 描述了归约集中边的统计预测误差  $\varepsilon$  的变化情况,图中纵坐标的值为  $\log_{10}\varepsilon$ .从图 4 可以看出,随着迭代的进行,归约集中的边对最优解路径中边的平均预测误差能够下降到  $10^{-3}$  到  $10^{-4}$  数量级.与 MR 一次性决策确定归约边集合的策略不同,SRIA 通过迭代逐渐改进归约集,能使预测误差降低到很小的数量级.

图 5 是每次归约后子空间自由度的变化情况.我们已经分析了 MR 算法随着归约的进行自由子空间的自由度迅速下降并趋近于 0,而从图 5 可以看出,SRIA 中归约后自由子空间的自由度没有下降反而上升了.这是由于:随着迭代的进行,归约集的预测精度逐渐提高,归约集的规模虽然增大了却没有像 MR 那样单调递增,而是自适应地保持在一定的水平上.因此,归约后自由子空间的自由度会呈现递增趋势.自由子空间的自由度越大,细化搜索算法陷入局部最优的可能性越小,算法对解的质量改进的余地也就越大.

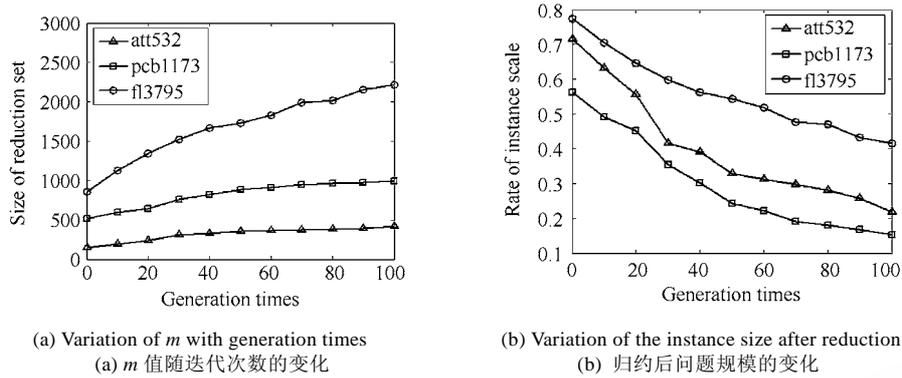
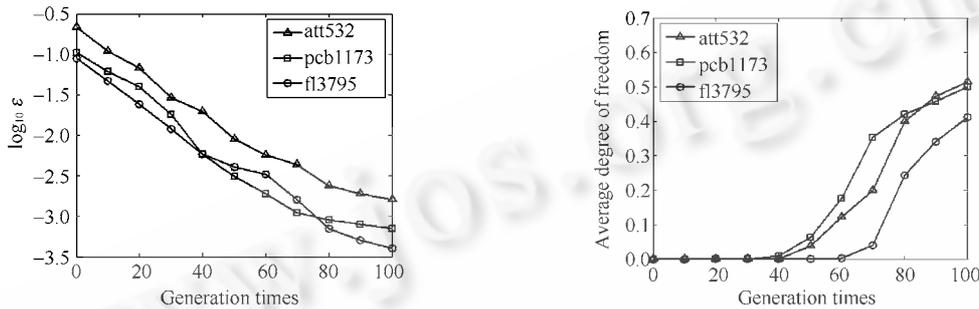
Fig.3 The property of variation of reduction degree  $m$ .图 3 归约度  $m$  的变化特性

Fig.4 Average magnitude order of forecasting errors Fig.5 Variation of the freedom of subspace after reduction

图 4 归约集中边的平均预测误差的变化

图 5 归约后子空间自由度的变化

## 5 总结与展望

大规模 TSP 问题向所有组合优化方法提出了严峻的挑战.如何利用先验知识或者算法自适应地学习到关于问题的知识来降低 TSP 问题的规模,是解决大规模 TSP 问题的一个思路.多级算法就是这类算法中的重要代表.本文对一个典型的多级算法——多级归约算法(MR)进行深入的分析发现,归约集的规模不是越大越好,而是要取一个合适的值才能使算法以较大的概率搜索到问题的全局最优解.因此,如何自适应地确定一个合适的归约度,以提高算法在归约后找到全局最优解的概率是本文工作的出发点.

受人工免疫系统的启发,我们设计了求解大规模 TSP 问题的自适应归约免疫算法(SRIA).SRIA 引入了归约集进化的思想,通过迭代进化使归约集的规模自适应增长,归约集中归约边的预测精度不断提高,从而实现提高算法在归约后找到全局最优解概率这一设计目标.实验结果表明,随着归约次数的增加,归约集不仅规模自适应增大,归约集中的边对最优解路径的统计预测误差也实现了跨数量级的降低.而且,归约后算法找到全局最优解的概率自适应地保持在一定的水平并呈递增趋势.这说明,SRIA 能够在降低问题规模的同时有效地保持细化搜索算法对解的质量改进的可能性.与 MR 相比,SRIA 用合理的时间代价换取了算法搜索性能的提高,适合求解大规模 TSP 问题.

## References:

- [1] Johnson DS, McGeoch LA. The traveling salesman problem: A case study in local optimization. In: Aarts EH, Lenstra JK, eds. Local Search in Combinatorial Optimization. New York: John Wiley and Sons, 1996.
- [2] Johnson DS. Local optimization and the traveling salesman problem. In: Paterson MS, ed. Proc. of the 17th Colloquium on Automata, Language, and Programming. LNCS 443, Berlin: Springer-Verlag, 1990. 446-461.

- [3] Guo T, Michalewicz Z. Inver-Over operator for the TSP. In: Eiben AE, *et al.*, eds. Proc. of the 5th Parallel Problem Solving from Nature Conf. LNCS 1498, Berlin: Springer-Verlag, 1998. 803–812
- [4] Song XM, Li B, Yang HM. Improved ant colony algorithm and its applications in TSP. In: Proc. of the 6th Int'l Conf. on Intelligent Systems Design and Applications. Washington: IEEE Computer Society, 2006. 1145–1148.
- [5] He XX, Zhu YL, Hu HC, Niu B. A new algorithm for TSP based on swarm intelligence. In: Proc. of the 6th World Congress on Intelligent Control and Automation. Piscataway: Institute of Electrical and Electronics Engineers Inc., 2006. 3241–3244.
- [6] Walshaw C. A multilevel lin-kernighan-helsgaun algorithm for the travelling salesman problem. Technical Report, 01/IM/80, London: University of Greenwich, Computing and Mathematical Sciences, 2001. 1–9.
- [7] Zou P, Zhou Z, Chen GL, Gu J. A multilevel reduction algorithm to TSP. Journal of Software, 2003,14(1):35–42 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/14/35.pdf>
- [8] Wang L, Pan J, Jiao LC. The Immune programming. Chinese Journal of Computers, 2000,23(8):806–812.
- [9] Walshaw C. A multilevel approach to the traveling salesman problem. Operations Research, 2002,50(5):862–877.
- [10] Walshaw C. Multilevel refinement for combinatorial optimization problems. Annals of Operations Research, 2004,131:325–372.
- [11] Dasgupta D, Forrest S. Artificial immune systems in industrial applications. In: Proc. of the 2nd Int'l Conf. on Intelligent Processing and Manufacturing of Materials. Hawaii: IEEE Press, 1999. 257–267.
- [12] Hunt JE, Cooke DE. An adaptive, distributed learning system based on immune system. IEEE Int'l Conf. on System, Man and Cybernetics, Vol.3. Vancouver: IEEE Press, 1995. 2494–2499.
- [13] De Castro LN, Von Zuben FJ. The clonal selection algorithm with engineering application. In: Proc. of the Genetic and Evolutionary Computation Conf. on Workshop on Artificial Immune System and Their Applications. Morgan Kaufmann Publishers, 2000. 36–37.
- [14] Zou P, Zhou Z, Jiang H, Chen GL and Gu J. Analysis of runtime distribution and solution performance distribution of iterated local search for the TSP. Chinese Journal of Computers, 2006,29(1):92–99 (in Chinese with English abstract).
- [15] Larrañaga P, Kuijpers CMH, Murga RH, Inza I, Dizdarevic S. Genetic algorithms for the traveling salesman problem: A review of representations and operators. Artificial Intelligence Review, 1999,13(2): 129–170.

## 附中文参考文献:

- [7] 邹鹏,周智,陈国良,顾钧.求解 TSP 问题的多级规约算法.软件学报,2003,14(1):35–42. <http://www.jos.org.cn/1000-9825/14/35.pdf>
- [8] 王磊,潘进,焦李成.免疫规划.计算机学报,2000,23(8):806–812.
- [14] 邹鹏,周智,江贺,陈国良,顾钧.求解旅行商问题的循环局部搜索算法的运行时间和性能分布分析.计算机学报,2006,29(1): 92–99.



戚玉涛(1981—),男,河南潢川人,博士生,主要研究领域为自然计算,机器学习.



焦李成(1959—),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为自然计算,数据挖掘,图像处理.



刘芳(1963—),女,教授,博士生导师,CCF 高级会员,主要研究领域为自然计算,图像处理.