

基于非一致性自适应变异的克隆选择算法

钟燕飞¹ 张良培¹ 李平湘¹

(1 武汉大学测绘遥感信息工程国家重点实验室,武汉市珞喻路 129 号,430079)

摘要:提出了一种基于非一致性自适应变异的克隆选择算法。该算法根据抗体的亲和力自适应地确定相应抗体的变异率,同时采用非一致性变异方法来提高算法的效率。实验证明,与传统的克隆选择算法相比,本文提出的方法所需要的收敛时间更少,且能快速地找到最优解,具有一定的实用价值。

关键词:人工免疫系统;克隆选择;非一致性变异

中图法分类号:P237.3

克隆选择算法(clonal selection algorithm, CSA)是人工免疫系统^[1]的一种重要模型,是 Castro 于 2002 年根据免疫系统中的克隆选择原理而提出的^[2]。CSA 作为一种新型的优化算法,在组合优化^[2]、多峰函数优化^[3]、网络入侵^[4]、模式识别^[5]和特征选择^[6]中得到了广泛的应用。然而,由于 CSA 中的参数数量太多,导致算法对参数较敏感,因此,算法的精度在应用时常常取决于人们的经验,算法收敛较慢,应用范围受到一定的限制。针对以上情况,本文提出了一种基于非一致性自适应变异的克隆选择算法(clonal selection algorithm based on non-uniform adaptive mutation, AMCSA),该方法的变异率通过自适应方法获得,无需人为设置,大大提高了算法的实用性。同时,AMCSA 继承了传统人工免疫系统和 CSA 的众多属性,具有自组织、自学习、自识别、自记忆的能力^[7,8],能快速提供达到最优解的搜索范围,而且可以保存各个局部最优解,从而可得到全局最优解。

1 基于非一致性自适应变异的克隆选择算法

传统克隆选择算法需要以下参数:最大迭代次数 T 、初始抗体种群大小 N 、最优抗体个数 n 、克隆率 β 、变异率 m 、替换个数 d 。AMCSA 中仅

需设定 T 和 N 的值,变异率通过自适应方法获得,其余参数根据前期对传统 CSA 的分析后提前设置,因此减少了大量的人为干预和对先验知识的依赖,具有更高的实用性和智能性。同时,采用非一致性变异算子对抗体进行变异,将变异结果与进化代数结合起来,使得算法能更好地获得最优解。本文以应用于优化问题的 AMCSA 算法描述其算法过程。

1.1 初始抗体群的产生

对于一个优化问题来说,需要求解一个最优值来使得某个评价函数最大,即 $\max(J)$ 。被求解的参数或者变量有一个取值范围 $[V_{\max}, V_{\min}]$,初始抗体种群则在该取值范围内随机获得。而对于优化问题来说,抗原这一概念则可以忽略,其亲和力则通过评价函数 J 来获得。抗体集合 B 分为非记忆抗体集合 B^s 和记忆抗体 M ,其中, $B^s = \{b_i^s, i=1, 2, \dots, N\}$, N 为非记忆抗体集合的抗体总数;记忆抗体 M 即为最优的抗体,也就是问题的候选解。本文对所有问题都采用实数编码,该编码方式在保证算法精度的同时,可以加快其运算速度。

1.2 亲和力函数值计算

对于优化问题来说,亲和力通常根据具体的工程问题来确定,如待优化的函数。 b_j 表示抗体个体, N 表示抗体数,则抗体个体亲和力表示为:

$$f_j = J(b_j), j = 1, 2, \dots, N \quad (1)$$

亲和力越高,表示该抗体越靠近最优解。 $J(b_j)$ 表示对于某个抗体个体 b_j 的评价函数,如最大化函数应用中的优化函数,这里用评价函数来表示抗体个体亲和力。

1.3 选择操作

从抗体集 B 中选择亲和力最高的 n 个抗体,产生一个新的集合 B^{new} 。对于优化问题来说,通常需要尽可能地覆盖整个解空间,然后才能避免获得局部最优解,因此,这里的 n 通常直接取值为 N ,也就是说,所有非记忆抗体全部参与算法操作。

1.4 克隆操作

对 n 个被选择出来的抗体进行克隆操作,产生克隆集合 C_j ;克隆过程中,以一定的克隆率 β 进行克隆。从前期的研究来看,克隆选择算法通常对于克隆倍数较为敏感。根据前期的实验^[5], β 取值在 $[1,3]$ 较为合适。在本文中, β 取值为 2,其克隆抗体总数为:

$$N_{\text{clone}} = \sum_{i=1}^n \beta \cdot n \quad (2)$$

1.5 变异操作

对克隆集合 C_j 进行变异操作,即得到产生变异后的抗体集合 \hat{C}_j 。其中,亲和力越高的个体,其变异机会越小。通常,该变异率 m 需要人为设定,而且对于所有的克隆抗体,变异率都是相同的。但从实际生物免疫系统中的理论来说,亲和力越高的个体变异由于与抗原更为接近,其变异率应该越小;而亲和力越低的个体,变异概率应该越大。为了实现该理论过程和减少对人为设定的依赖,在本文中,该值通过下面的方法自适应获得:① 每个抗体的亲和力被归一化:

$$f'_j = [f_j - \min(f_j)] / [\max(f_j) - \min(f_i)] \quad (3)$$

② 确定每个抗体 b_j 的变异率 m_j :

$$m_j = 1 - f'_j \quad (4)$$

通过以上过程,不仅可以自适应地得到每个抗体的变异率 m_j ,而且可以使得亲和力越高的抗体其变异率越小,而亲和力越低的抗体其变异率越大。

在确定完变异率后,需要对其进行变异操作。传统的变异算子的作用和变异代数是没有直接关系的,当算法进化到一定代数以后,由于缺乏局部搜索,很难通过传统的变异算子获得最优解。而在本文算法中,拟采用非一致性变异算子,该变异算子将变异算子的变异结果与进化代数联系起来,使得在进化初期,变异算子的范围相对较大,

而随着进化的推进,变异的范围越来越小,从而起到一种对系统微调的作用。

抗体 b_j 的变异过程通过变异函数 $\text{mutate}(b_j)$ 来实现,其中需要输入该个体的变异率。如前所述,通过式(3)和式(4)能自适应地根据个体的亲和力来获得其相应的变异率 m_j 。具体过程如下:

1) 将变异率 m_j 输入到变异函数 $\text{mutate}(b_j)$ 中, $b_j = \{b_{jk} | k=1, \dots, r\}$, k 表示抗体 b_j 的向量维数。如对于只有一个参数的最优函数问题, $k=1$ 。对于任意的 $k \in [1, r]$,执行以下变异步骤。

2) 根据该个体的变异率 m_j ,判读 b_j 的第 k 维是否需要变异。若不需要,则直接执行步骤 3);如果需要,则利用式(5)的非一致性变异方法进行变异。变异过程如下:

```
mutate( $b_j$ )
{
    foreach( $b_{jk}$ )
    do
        rd_mr = random(0,1)
        rd_to = random(-1,1)
        if(rd_mr <  $m_j$ )
        if(rd_to >= 0 )
             $b_{jk} = b_{jk} + \Delta(t, \max - b_{jk})$ 
        else
             $b_{jk} = b_{jk} - \Delta(t, b_{jk} - \min)$ 
    done
    return  $b_j$ 
}
```

其中,函数 $\text{random}(a, c)$ 为随机函数; \max 、 \min 分别表示抗体 b_j 第 k 维取值的最大、最小值。变异过程通过变异量函数 $\Delta(t, u)$ 来获得变异后的抗体:

$$\Delta(t, u) = u(1 - r^{(1-t/T)^\lambda}) \quad (5)$$

其中, r 是 $[0, 1]$ 上的一个随机数; T 为最大迭代数; t 为当前迭代数; λ 是决定非一致性程度的一个参数,它起着调整局部搜索区域的作用,其取值一般为 2 到 5,本文取值为 2。 $\Delta(t, u)$ 的取值范围为 $[0, u]$,变异后的结果考虑当前迭代次数 t 的影响,当 t 较小时, $\Delta(t, u)$ 的取值接近于 0 的可能性越小,使得算法在进化初期能搜索到较大范围;而当 t 越接近于 T 时, $\Delta(t, u)$ 的取值接近于 0 的可能性增大,则主要进行局部搜索。这正好符合了进化初期由于候选解与最优解差别较大,需要进行大变异,而在进化后期,由于接近于最优解,需要进行局部搜索的进化规则。

3) 输出结果。如果 k 取值等于 r ,则 b_j 所有

维的分量都经历了变异过程,重复步骤 1),执行下一个抗体的变异。

变异是 AMCSA 算法的关键步骤。它使得某个特征随机变化,避免算法陷入局部最优,这些变化值将被用到下一代抗体的运算中。同时,在每次克隆过程中,为了避免变异后的最优抗体比变异前的最优抗体更差,一个原始的未变异的抗体在变异过程中被保存下来,保证算法朝着最优的方向进化。

1.6 亲和力的计算

利用式(1)计算所有的变异抗体集合 \hat{C}_j 的亲和力 \hat{f}_j , \hat{f}_j 表示变异后抗体的亲和力。

1.7 重新选择

从变异克隆抗体集 \hat{C}_j 中选择 n 个最高的亲和力来替换初始抗体种群中 n 个具有最低亲和力的抗体。在 \hat{C}_j 中,选择具有最高亲和力的抗体 b_{best} 作为候选记忆细胞。如果 b_{best} 的亲和力高于原始记忆细胞, $f_{b_{best}} > f_M$, 则 b_{best} 将替换原有的记忆细胞,从而变成一个新的记忆细胞 M 。

1.8 替换最低亲和力抗体

为了增加抗体的多样性,随机选择 5% ($d = 5\%$) 的抗体替换 B^* 中的同等数量的最低亲和力的抗体。该步增加了抗体种群的多样性。替换抗体数目太多,会导致算法接近于随机算法;如太少,则导致算法过早收敛。

1.9 终止条件判断

当迭代次数达到预先定义的迭代次数 T 或者两次相邻迭代次数中的记忆细胞的变化小于一个阈值范围时,算法执行过程完成。如果不满足终止条件,则返回 § 1.2。

算法结束后,AMCSA 输出记忆细胞,该记忆细胞即为最后要求解的最优值。

2 实验与分析

为了验证 AMCSA 的有效性,本文通过两个实验将 AMCSA 与常用的优化算法,如遗传算法(GA)、模拟退火算法(SA)、遗传模拟退火算法(GASA)和传统的克隆选择算法(CSA)进行了比较分析。

2.1 函数优化

为了进行函数优化测试,选用典型的多峰值函数作为测试函数,其形式如下:

$$f(x) = 1 - |\sin(30x)| (1 - x/2), x \in [0, 1]$$

(6)

该函数存在 10 个极小值。对于 CSA 和 ACSA,

需要将其转换为求解最大值问题,即式(1)被设置为:

$$f_j = J(b_j) = -f(x) \tag{7}$$

各算法参数设置如下:所有算法的最大迭代次数都为 200,其中,GA、GASA 选用 22 位染色体,保证精确到小数点后 6 位,初始种群数为 50,交叉率为 0.65,变异率为 0.01;SA 中的衰减参数为 0.95,步长因子为 0.02;CSA 和 ACSA 的初始抗体为 50;CSA 采用的变异率为 0.1。表 1 列出了五种测试算法的实验结果。

从表 1 可以看出,除了 GA 算法外,其他三种算法都能够求解到全局最优解。在实验中发现,SA 能够将大部分的峰搜索到,因此,相信只要温度降低得足够缓慢,就可以找到所有的峰值点。而且 SA 在 16 代就能找到最优解,速度很快。但是在实际应用中,不可能每次都会有足够多的训练次数给 SA 训练,那么当训练次数有限或者温度降低得不够缓慢时,它就可能会陷入局部范围寻优。因此,SA 应该和其他全局性较好的优化算法相结合,才能达到最佳效果(如 GA、CSA)。从实验中可以看出,基本遗传算法的多峰搜索能力和局部搜索能力都是最差的。其很难搜索到局部最优解的原因在于存在二进制的通常问题——“海明悬崖”问题,即两个值只是相差 1,可是在二进制中,不同的位数却差很多,这导致当 GA 搜索到离最优解比较近时,却很难搜索到最优解,因为“海明悬崖”使得 GA 的跳跃性太大,导致局部搜索能力较弱。GASA 在 GA 的基础上加入了 Boltzman 机制保存差解来加强其局部搜索能力,算法得到改进,找到最优解的能力较基本遗传算法有一定的提高。

表 1 五种算法函数优化结果比较
Tab. 1 Results Comparison of Five Algorithms for the Function Optimization

算法	收敛最优值	收敛代数	时间/s
	(0.051 790, 0.026 037 4)		
GA	(0.051 761 2, 0.026 037 7)	122	0.390
SA	(0.051 789 6, 0.026 037 4)	114	4.141
GASA	(0.051 791 4, 0.026 037 4)	84	0.391
CSA	(0.051 788 9, 0.026 037 4)	171	0.373
AMCSA	(0.051 789 1, 0.026 037 4)	85	0.215

从图 1 可以发现,CSA 和 AMCSA 两种克隆选择算法能够搜索到最优解,而且多峰搜索能力也要强于 SA,算法所需的收敛次数和时间也要少于 SA。其原因在于该类 CSA 算法具有以下优点:① 在记忆单元基础上运行,确保了快速收敛

于全局最优解;② 有计算亲和性的程序反映了真实免疫系统的多样性;③ 通过促进或抑制抗体的产生,体现了免疫反应的自我调节功能。

为了更好地比较 AMCSA 与 CSA 的结果,图 1(a)描述了 CSA 和 AMCSA 的种群平均亲和力随迭代次数的增加而发生变化的示意图。由于在第 20 次后亲和力发生的变化在图 1(a)中难以明显地表示,因此,将迭代次数从 20 增加到 200 以后的曲线利用图 1(b)单独描绘出来。值得注意的是,图 1 表示的亲和力是用式(7)得到的值,即求得的是其最大值。

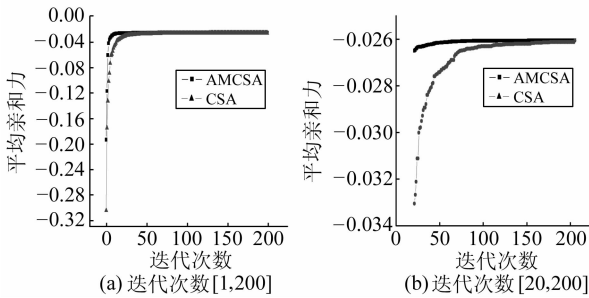


图 1 AMCSA 与 CSA 平均亲和力比较图
Fig. 1 Comparisons of the Average Affinity Values Against the Iteration Using AMCSA and CSA

从图 1 和表 1 可以看出,AMCSA 相比于 CSA 的收敛速度更快,所需的迭代次数更少,仅用了 85 次迭代、0.215 s 就找到了最优解,而 CSA 则需要 171 代、0.373 s 才能找到最优解。其原因在于 AMCSA 通过自适应控制了变异率,使得亲和力较好的个体的变异率较小,而 CSA 则由于采用了固定的变异率,无论抗体亲和力大小,一律用统一的变异率进行变异,这样有可能造成算法在最优解附近时难以收敛,增加了收敛所需的时间。而且 AMCSA 由于采用了非一致性变异方法,当迭代次数较大时,仅在候选解附近进行搜索,从而增加了找到最优解的概率。因此,AMCSA 在保持 CSA 各种优点的同时,减少了对参数的依赖性,而且能够比 CSA 更快地找到最优解,具有更高的实用价值。

2.2 图像分割实验

采用 Lena 影像对算法进行最佳阈值分割。实验中,利用五种算法来进行最佳阈值计算。评价标准利用基于 Otsu 法^[9]的图像分割公式,具体为:

$$f = w_0(u_0 - v)^2 + w_1(u_1 - v)^2 \quad (8)$$

其中, w_0 和 w_1 分别表示其灰度小于和大于门限 k 的概率; u_0 和 u_1 分别表示上面两个区间的平均灰度值; v 表示整幅图像的平均灰度。 f 值越大,

表示分割的质量越好。该影像的最佳阈值为 118。本实验直接利用式(8)中的 f 值作为式(1)的亲和力。

各算法参数设置如下:所有算法的最大迭代次数都为 200,其中,GA、GASA 由于影像灰度在 0~255 之间,采用 8 位染色体长度,精度为 1,初始种群数为 50,交叉率为 0.8,变异率为 0.01。SA 中的衰减参数为 0.95,步长因子为 0.02;CSA 和 ACSA 的初始抗体为 50,CSA 采用的变异率为 0.2。表 2 列出了五种算法所获得的最佳分割阈值和运算所需的时间。

表 2 五种算法 Lena 影像分割结果比较
Tab. 2 Results Comparison of Five Algorithms for Lena Image Segmentation

算法	最佳分割阈值 k	收敛代数	时间/s
GA	117.536 099 6	35	230
SA	117.951 858 5	178	862
GASA	117.950 823 1	113	215
CSA	117.952 005	151	123
AMCSA	117.952 505	103	67

从表 2 可以看出,此次实验再次证明了 AMCSA 算法相比于传统的 CSA 算法在获得同样的实验结果时,其运算速度更快,因此更具有实用性。

3 结 语

本文提出的算法能够自适应地根据抗体的亲和力设置其相应的变异率,相比于传统的 CSA 算法,减少了对使用者经验的依赖,使得算法的实用性和智能性得到很大提高。与其他优化算法相比,本文算法不但能获得最好的结果,而且在多峰搜索能力上也要强于遗传算法,具有很强的优化能力。进一步的工作将通过各种方法来提高算法的运行速度,以使其能够运行到更为复杂的大型计算上。

参 考 文 献

[1] De Castro L N, Timmis J. Artificial Immune systems: A New Computational Intelligence Approach [M]. London: Springer-Verlag, 2002

[2] De Castro L N, Von Zuben F J. Learning and Optimization Using the Clonal Selection Principle [J]. IEEE Transactions on Evolutionary Computation, 2002, 6: 239-250

[3] 焦李成, 杜海峰, 刘芳, 等. 免疫优化计算、学习与识别[M]. 北京: 科学出版社, 2006

[4] Kim J, Bentley P. Towards an Artificial Immune System for Network Intrusion Detection: An Investigation of Clonal Selection with a Negative Selection Operator[C]. Congress on Evolutionary Computation, Seoul, Korea, 2001

[5] 钟燕飞,张良培,龚健雅,等. 基于克隆选择的多光谱遥感影像分类算法[J]. 中国图像图形学报, 2005,10(1):18-24

[6] Zhang Liangpei, Zhong Yanfei, Huang B, et al. Dimensionality Reduction Based on Clonal Selection for Hyperspectral Satellite Imagery [J]. IEEE Transactions on Geoscience and Remote Sensing, 2007, 45(12): 4 172-4 186

[7] Adams D. How the Immune System Works and Why It Causes Autoimmune Diseases [J]. Immunology Today, 1996,17(7): 300-302

[8] 汤凌,郑肇葆,虞欣. 一种基于人工免疫的图像分割算法[J]. 武汉大学学报·信息科学版,2007,32(1):67-70

[9] Otsu N. A Threshold Selection Method from Gray Level Histograms[J]. IEEE Transactions on Systems, Man and Cybernetics, 1979,9(1):62-66

第一作者简介: 钟燕飞,讲师,博士。主要从事遥感影像处理、模式识别、人工智能等方面的研究。
E-mail:zhongyanfei@lmars.whu.edu.cn

A Clonal Selection Algorithm Based on Non-uniform Adaptive Mutation

ZHONG Yanfei¹ ZHANG Liangpei¹ LI Pingxiang¹

(1 State Key Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, 129 Luoyu Road, Wuhan 430079, China)

Abstract: We propose an adaptive clonal selection algorithm, which can adaptively set the mutation rate of each antibody according to its affinity and utilize the non-uniform mutation to decrease the complexity of the algorithm and computational costs. Experiments were performed to evaluate the performance of our proposed algorithm in comparison with other methods, genetic algorithm, simulated annealing algorithm, genetic simulated annealing algorithm. The results show that our proposed algorithm consistently outperforms the traditional algorithms. Compared with traditional CSA, our proposed algorithm can more quickly find the optimal solution and hence provides an effective solution.

Key words: artificial immune system; clonal selection; non-uniform mutation

About the first author: ZHONG Yanfei, lecturer, Ph. D, majors in remote sensing image processing, pattern recognition and artificial intelligence.
E-mail: zhongyanfei@lmars.whu.edu.cn

本刊 2008 年出版基本情况统计

2008 年度,《武汉大学学报·信息科学版》共出版 12 期,合计字数 269 万,比上年增加 9.8%;发表论文 314 篇,比上年增加 39 篇,增加 14.2%;篇均页码 4.17,减少 0.01 面。同期,共收到国内外投稿 887 篇,与上年略有增加;用稿率 35.4%,较上年持平。发表的论文,从投稿到出版,时差最长 202 天,最短 31 天,平均 54.2 天,比上年缩短 16 天。发表的论文中,校内稿件 153 篇,占 48.7%;校外及海外稿件 161 篇,占 51.3%;外稿比例较上年增加 1.8 个百分点。各类基金资助论文 303 篇,基金产文率 96.5%,比上年提高 2.3 个百分点。EI 核心版收录比例继续保持 100%。

已发表论文中,涉及作者 987 人次,比上年增加 2.26%。其中校内作者 421 人次,占 42.7%;校外及海外作者 566 人次,占 57.3%;校外及海外作者减少 7.1 个百分点。按第一作者统计,院士论文 10 篇,占 3.2%;博士生导师 55 篇,占 17.5%;教授、研究员 17 篇,占 5.4%;副教授 37 篇,占 11.8%;博士和博士生 166 篇,占 52.9%;其他 29 篇,占 9.2%。累计高级职称作者论文比例 37.9%,博士论文比例 52.9%,两者合计 90.8%,比上年下降 2.3 个百分点。