

分布估计算法综述

周树德¹ 孙增圻¹

摘要 分布估计算法是进化计算领域新兴起的一类随机优化算法,是当前国际进化计算领域的研究热点.分布估计算法是遗传算法和统计学习的结合,通过统计学习的手段建立解空间内个体分布的概率模型,然后对概率模型随机采样产生新的群体,如此反复进行,实现群体的进化.分布估计算法中没有传统的交叉、变异等遗传操作,是一种全新的进化模式;这种优化技术能够通过概率图模型对变量之间的关系进行建模,从而能有效的解决多变量相关的优化问题.根据概率模型的复杂性,本文按照变量无关、双变量相关、多变量相关等三类分别介绍相应的分布估计算法.作为一篇综述性文章,本文旨在全面系统的向国内读者介绍这一新技术,并总结分布估计算法的研究现状和未来的研究方向.

关键词 分布估计算法,遗传算法,统计学习,概率模型
中图分类号 TP18

A Survey on Estimation of Distribution Algorithms

ZHOU Shu-De¹ SUN Zeng-Qi¹

Abstract Estimation of distribution algorithms (EDAs) are a class of novel stochastic optimization algorithms, which have recently become a hot topic in field of evolutionary computation. EDAs acquire solutions by statistically learning and sampling the probability distribution of the best individuals of the population at each iteration of the algorithm. EDAs have introduced a new paradigm of evolutionary computation without using conventional evolutionary operators such as crossover and mutation. In such a way, the relationships between the variables involved in the problem domain are explicitly and effectively exploited. According to the complexity of probability models for learning the interdependencies between the variables from the selected individuals, this paper gives a review of EDAs in the order of interactions: dependency-free, bivariate dependencies, and multivariate dependencies, aiming to bring the reader into this novel filed of optimization technology. In addition, the future research directions are discussed.

Key words Estimation of distribution algorithms(EDAs), genetic algorithm, statistical learning, probabilistic model

1 引言

最近几年,在进化计算领域兴起了一类新型的优化算法,称为分布估计算法(Estimation of distribution algorithms, EDAs),并迅速成为进化计算领域的研究热点和解决工程问题的有效方法.分布估计算法的概念最初在1996年提出^[1],在2000年前后迅速发展,成为当前进化计算领域前沿的研究内容,2005年在进化计算领域权威的国际期刊Evolutionary Computation上出版了分布估计算法的专刊,近年来国际上进化计算领域的各大学术会议,如ACM SIGEVO、IEEE CEC等,都将分布估计算法作为重要专题予以讨论.

分布估计算法提出了一种全新的进化模式.在

传统的遗传算法中,用种群表示优化问题的一组候选解,种群中的每个个体都有相应的适应值,然后进行选择、交叉和变异等模拟自然进化的操作,反复进行,对问题进行求解.而在分布估计算法中,没有传统的交叉、变异等遗传操作,取而代之的是概率模型的学习和采样.分布估计算法通过一个概率模型描述候选解在空间的分布,采用统计学习手段从群体宏观的角度建立一个描述解分布的概率模型,然后对概率模型随机采样产生新的种群,如此反复进行,实现种群的进化,直到终止条件^[1~3].

根据概率模型的复杂程度以及不同的采样方法,分布估计算法发展了很多不同的具体实现方法,但是都可以归纳为下面两个主要步骤^[2]:

1) 构建描述解空间的概率模型.通过对种群的评估,选择优秀的个体集合,然后采用统计学习等手段构造一个描述当前解集的概率模型.

2) 由概率模型随机采样产生新的种群.一般的,采用蒙特卡罗方法,对概率模型采样得到新的种群.

分布估计算法作为一种新型的进化算法,它的科学价值主要体现在以下几个方面.首先,从生物进化的数学模型上来看,分布估计算法与传统进化

收稿日期 2006-3-20 收修改稿日期 2006-6-12
Received March 20, 2006; in revised form June 12, 2006
国家重点基础研究专项基金(G2002cb312205)资助
Supported by National Key Project for Basic Research of P. R. China (G2002cb312205)
1. 清华大学计算机系智能技术与系统国家重点实验室 北京 100084
1. State Key Lab of Intelligent Technology and Systems, Department of Computer Science and Technology, Tsinghua University, Beijing 100084
DOI: 10.1360/aas-007-0113

算法不同: 传统进化算法是基于对种群中的各个个体进行遗传操作(交叉、变异等)来实现群体的进化的, 是对生物进化“微观”层面上的数学建模; 而分布估计算法则是基于对整个群体建立数学模型, 直接描述整个群体的进化趋势, 是对生物进化“宏观”层面上的数学建模. 其次, 分布估计算法给人类解决复杂的优化问题提供了新的工具, 它通过概率模型可以描述变量之间的相互关系, 从而对解决非线性、变量耦合的优化问题更加有效, 试验表明, 分布估计算法能更加有效的解决高维问题, 降低时间复杂性, 例如, 贝叶斯优化算法(分布估计算法的一种)可以通过与问题规模成多项式数量级的采样求得一类 GA(Genetic algorithm) 难问题的最优解^[3]. 最后, 分布估计算法是一种新的启发式搜索策略, 是统计学习理论与随机优化算法的结合, 与其他智能优化算法的混杂设计, 将极大丰富混杂优化算法的研究内容, 给优化算法的研究提供了新的思路.

本文作为一篇综述性文章, 旨在向读者介绍一种新型的优化方法, 为研究人员尽可能详细地总结该领域的研究现状. 后续篇幅安排如下: 首先, 将通过一个简单的算例介绍分布估计算法的实现过程, 使读者对该方法有一个直观的认识; 然后将按照概率模型的复杂程度对分布估计算法进行分类, 分别介绍变量无关的 PBIL、UMDA 和 cGA 算法, 双变量相关的 MIMIC、BMDA 算法和多变量相关的 ECGA、FDA、BOA 等算法; 之后, 将介绍连续域分布估计算法的相关研究; 第 5 节综述了分布估计算法的理论研究现状; 第 6 节介绍了基于概率模型的其他优化算法; 第 7 节综述了分布估计算法目前存在的关键问题和难点, 指出了该领域的发展趋势; 最后对分布估计算法进行总结.

在后文的叙述中, EDA(Estimation of distribution algorithms 的缩写) 表示分布估计算法.

2 一个简单的分布估计算法

下面通过一个简单的 EDA 算例, 介绍该方法独特的进化操作, 使读者对 EDA 方法有一个直观的认识.

假设用分布估计算法求解函数 $f(x) = \sum_{i=1}^n x_i$ 的最大值, $x \in \{0, 1\}^n$, $n = 3$. 在这个例子中, 描述解空间的概率模型用简单的概率向量 $\mathbf{p} = (p_1, p_2, \dots, p_n)$ 表示, \mathbf{p} 表示群体的概率分布, $p_i \in [0, 1]$ 表示基因位置 i 取 1 的概率, $1 - p_i$ 表示基因位置 i 取 0 的概率.

第一步, 初始化群体 B_0 . 初始群体在解空间按照均匀分布随机产生, 概率向量 $\mathbf{p} = (0.5, 0.5, 0.5)$. 群体大小为 8, 通过适应值函数 $f(x) = \sum_{i=1}^n x_i$ 计

算各个个体适应值. B_0 表示如表 1.

表 1 初始群体, 8 个个体按照均匀分布在解空间内随机产生
Table 1 The initial population with 8 individuals that are uniformly generated in solution space

	x_1	x_2	x_3	f
1	0	0	1	1
2	1	1	0	2
3	0	0	0	0
4	0	1	1	2
5	0	1	0	1
6	1	0	0	1
7	1	0	1	2
8	1	1	1	3

第二步, 选择适应值较高的 4 个个体更新概率向量 \mathbf{p} , 如表 2. X_S 表示选择后的优势群体, 概率向量 \mathbf{p} 通过式 $p_i = P(x_i = 1|X_S)$ 更新, 例如 $p_1 = P(x_1 = 1|X_S) = 0.75$, 这样得到新的概率向量 $\mathbf{p} = (0.75, 0.75, 0.75)$.

表 2 选择操作后的优势群体 X_S , 用来更新概率向量 \mathbf{p}
Table 2 The selected individuals X_S used to update probability vector \mathbf{p}

	x_1	x_2	x_3	f
2	1	1	0	2
4	0	1	1	2
7	1	0	1	2
8	1	1	1	3

第三步, 由概率向量 \mathbf{p} 产生新一代群体. 概率向量描述了各个可能解在空间的分布情况, 产生任意解 $\mathbf{b} = (b_1, b_2, \dots, b_n)$ 的概率是

$$P(\mathbf{b}) = P(x_1 = b_1, x_2 = b_2, \dots, x_n = b_n) \\ = \prod_{i=1}^n P(x_i = b_i) = \prod_{i=1}^n |1 - b_i - p_i| \quad (1)$$

例如, $\mathbf{b} = (1, 1, 0)$, 则 $P(1, 1, 0) = 0.75 \times 0.75 \times 0.25 \approx 0.14$. 通过随机采样的方法产生新的群体, 如表 3 所示, 可以发现新产生的群体的个体适应值有了显著的提高.

至此, 分布估计算法完成了一个周期. 然后返回第二步, 从表 3 当前群体中选择最优秀的 4 个个体, 建立新的概率模型 $\mathbf{p} = (1.00, 0.75, 0.75)$, 然后再对概率模型随机采样产生新一代群体. 本例中, 最优解为 $(1, 1, 1)$, 可以发现随着分布估计算法的进行, $(1, 1, 1)$ 的分布概率由最初的 $0.075(0.5 \times 0.5 \times 0.5)$ 变为 $0.42(0.75 \times 0.75 \times 0.75)$, 然后为 $0.56(1.00 \times 0.75 \times 0.75)$, 适应值高的个体的出现概

表 3 经过一代 EDA 操作之后, 产生的新的群体
Table 3 The new generated population after one EDA generation

	x_1	x_2	x_3	f
1	1	1	1	3
2	1	1	0	2
3	1	1	0	2
4	0	1	1	2
5	1	1	0	2
6	1	0	1	2
7	1	1	1	3
8	1	0	0	1

率越来越大. 按照上面的步骤, 改变个体在解空间的概率分布, 使适应值高的个体分布概率变大, 适应值低的个体分布概率变小, 如此反复进化, 最终将产生问题的最优解.

通过上述简单例子, 可以了解分布估计算法的基本过程. 图 1 归纳了一般分布估计算法的过程, 并与遗传算法进行了对比. 在分布估计算法中, 没有遗传算法中的交叉和变异等操作, 而是通过学习概率模型和采样操作使群体的分布朝着优秀个体的方向进化. 从生物进化角度看, 遗传算法模拟了个体之间微观的变化, 而分布估计算法则是生物群体整体分布的建模和模拟.

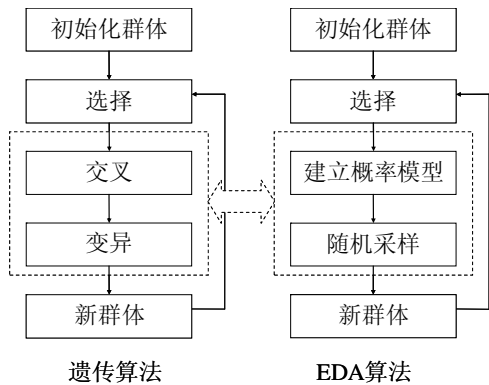


图 1 分布估计算法和遗传算法的基本流程
Fig. 1 The basic procedures of EDA and genetic algorithm

本节通过简单算例, 使读者对分布估计算法有一个直观的认识. 在分布估计算法的研究领域, 有各种各样的具体实现, 下一节将对各种分布估计算法进行较为详细的介绍.

3 基于不同概率模型的分布估计算法

分布估计算法是一种基于概率模型的进化算法,

它可以用概率模型描述变量之间的相互关系, 因此可以解决传统遗传算法难以解决的问题, 特别是对于那些非线性、变量耦合的高维问题, 表现出了很好的性能. 根据优化问题的复杂性, 人们设计了很多种不同的概率模型表示变量之间的关系, 本文将分布估计算法分为变量无关、双变量相关和多变量相关三种, 分别予以介绍. 本节内容只讨论二进制编码的算法, 对于实数编码 (连续空间) 的相关算法将在下一节介绍.

3.1 变量无关的分布估计算法

在 EDA 领域的研究中, 最简单的情况是变量之间无关. 在这种情况下, 一般的可以通过一个简单的概率向量表示解的分布, 变量之间关系可以用图 2 表示. 设定待解决问题为 n 维问题, 每个变量取值是二进制, 变量无关性使得任意解的概率可表示为

$$P(x_1, x_2, \dots, x_n) = \prod_{i=1}^n P(x_i) \quad (2)$$

EDA 领域最早的算法就是针对变量无关的问题提出的, 比较有代表性的算法包括 PBIL 算法 (Population based incremental learning)^[4~6]、UMDA 算法 (Univariate marginal distribution algorithm)^[7, 8]、cGA 算法 (compact genetic algorithm)^[9] 等, 下面分别介绍.

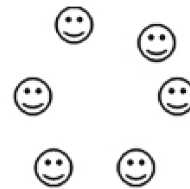


图 2 变量相互无关的概率图模型
Fig. 2 The probabilistic graphical model without interdependencies

PBIL 算法由美国卡耐基梅隆大学的 Baluja 在 1994 年提出^[4], 用以解决二进制编码的优化问题. 虽然在 1994 年“分布估计算法”这一概念还没有在学术界提出, 但是 PBIL 算法被公认是最早的分布估计算法模型. 在 PBIL 算法中, 表示解空间分布的概率模型是一个概率向量 $\mathbf{p}(\mathbf{x}) = (p(x_1), p(x_2), \dots, p(x_n))$, 其中 $p(x_i)$ 表示第 i 个基因位置上取值为 1 的概率. PBIL 算法的过程如下, 在每一代中, 通过概率向量 $\mathbf{p}(\mathbf{x})$ 随机产生 M 个个体, 然后计算 M 个个体的适应值, 并选择最优的 N 个个体用来更新概率向量 $\mathbf{p}(\mathbf{x})$, $N \leq M$. 更新概率向量的规则, 采用了机器学习中的 Heb 规则^[4], 用 $\mathbf{p}_l(\mathbf{x})$ 表示第 l 代的概率向量, $x_l^1, x_l^2, \dots, x_l^N$ 表

示选择的 N 个个体, 则更新过程如下

$$p_{l+1}(x) = (1 - \alpha)p_l(x) + \alpha \frac{1}{N} \sum_{k=1}^N x_k^l \quad (3)$$

其中 α 表示学习速率.

PBIL 算法作为最简单的 EDA 模型引起了学术界的广泛关注, 在算法改进、理论研究和应用方面都有很多论文发表. Baluja 在 1995 年对 PBIL 算法和其他启发式算法做了试验比较^[5].

UMDA 算法由德国学者 Mühlenbein 在 1996 年提出^[7, 8]. UMDA 算法与 PBIL 算法的不同在于概率向量的更新算法. 在本文第 2 节的 EDA 简单算例中, 所使用的算法就是 UMDA 算法. 算法描述如下:

- 1) 随机产生 M 个个体作为初始群体 $D_l, l = 0$;
- 2) 计算 M 个个体的适应值, 如果符合终止条件, 算法结束, 否则继续进行;
- 3) 进行选择操作, 选择 $N < M$ 个个体作为优势群体 D_l^S ;
- 4) 由优势群体 D_l^S 构建概率模型, 估计联合概率分布

$$\begin{aligned} p_l(\mathbf{x}) &= p(\mathbf{x}|D_l^S) = \prod_{i=1}^n p_l(x_i) \\ &= \prod_{i=1}^n \frac{\sum_{j=1}^N \delta_j(X_i = x_i | D_l^S)}{N} \end{aligned} \quad (4)$$

其中 n 为问题维数 (变量个数), $\delta_j(X_i = x_i | D_l^S) = \begin{cases} 1 & X_i = x_i \\ 0 & \text{其他} \end{cases}$

- 5) 从概率向量 $p_l(\mathbf{x})$ 中采样 M 次, 得到新一代群体. 返回 2).

美国 UIUC 大学的 Harik 等人提出了紧致遗传算法 (compact genetic algorithm, cGA)^[9]. 在 cGA 算法中, 由概率向量表示各个解的概率分布, 它与 PBIL 算法和 UMBA 算法的不同不仅在于概率模型的更新算法, 而且 cGA 群体规模很小, 只需要很小的存储空间. cGA 算法中, 每次仅由概率向量随机产生两个个体, 然后对两个个体进行比较, 按照一定的策略对概率向量更新. 下面是 cGA 的算法描述:

- 1) 初始化概率向量 $p_0(\mathbf{x}) = (p_0(x_1), p_0(x_2), \dots, p_0(x_n)) = (0.5, 0.5, \dots, 0.5), l = 0$;

- 2) 对概率向量 $p_l(\mathbf{x})$ 进行随机采样, 产生两个个体, 并计算它们的适应值, 较优秀的个体记为 x_1^l ,

较差的个体记为 x_2^l ;

- 3) 更新概率向量 $p_l(\mathbf{x})$, 使其朝着 x_1^l 的方向改变. 对概率向量中的每一个值, 如果 $x_{1,i}^l \neq x_{2,i}^l$, 按照下面策略进行更新, $\alpha \in (0, 1)$ 一般取作 $\alpha = \frac{1}{2K}$, K 为正整数. 如果 $x_{1,i}^l = 1$, 则 $p_{l+1}(x_i) = p_l(x_i) + \alpha$; 如果 $x_{1,i}^l = 0$, 则 $p_{l+1}(x_i) = p_l(x_i) - \alpha$.

- 4) $l = l + 1$, 检测概率向量 $p_l(\mathbf{x})$, 对任意 $i \in \{1, \dots, n\}$, 如果 $p_l(x_i) > 1$, 则 $p_l(x_i) = 1$, 如果 $p_l(x_i) < 0$, 则 $p_l(x_i) = 0$.

- 5) 如果 $p_l(\mathbf{x})$ 中, 对任意 $i \in \{1, \dots, n\}$, $p_l(x_i) = 0$ 或 1, 则算法终止, $p_l(\mathbf{x})$ 就是最终解. 否则转 2).

cGA 实现更加简单, 是一种适合于硬件实现的分布估计算法^[9].

3.2 双变量相关的分布估计算法

PBIL、UMDA 和 cGA 算法都没有考虑变量之间的相互关系, 算法中任意解向量的联合概率密度可以通过各个独立分量的边缘概率密度相乘得到. 而在实际问题中, 变量并不是完全独立的, 在分布估计算法研究领域, 最先考虑变量相关性的算法是假设最多有两个变量相关. 这类算法比较有代表性的包括由 De Bonet 等于 1997 年提出的 MIMIC (Mutual information maximization for input clustering) 算法^[10], Baluja 于 1997 年提出的 COMIT (Combining optimizers with mutual information trees) 算法^[11], Pelikan 等人于 1999 年提出的 BMMA 算法^[12, 13] 等. 在这类分布估计算法中, 概率模型可以表示至多两个变量之间的关系.

MIMIC 算法, 是由美国 MIT 人工智能实验室的 De Bonet 等人于 1997 年提出的一种启发式优化算法^[10]. 在 MIMIC 算法中, 变量之间的相互关系是一种链式关系, 结构如图 3 所示, 描述解空间的概率模型写作

$$p_l^\pi(\mathbf{x}) = p_l(x_{i_1}|x_{i_2})p_l(x_{i_2}|x_{i_3}) \cdots p_l(x_{i_{n-1}}|x_{i_n})p_l(x_{i_n}) \quad (5)$$

其中 $\pi = (i_1, i_2, \dots, i_n)$ 表示变量 (x_1, x_2, \dots, x_n) 的一种排列, $p_l(x_{i_j}|x_{i_{j+1}})$ 表示第 i_{j+1} 个变量取值为 $x_{i_{j+1}}$ 的条件下第 i_j 个变量取值为 x_{i_j} 的条件概率. 在 MIMIC 算法中构建概率模型时, 我们期望得到最优的排列, 使得 p_l^π 与试验中得到的每代优势群体的概率分布 $p_l(x)$ 最接近.



图 3 MIMIC 算法中双变量相关的链式结构的概率图模型
Fig. 3 Probabilistic graphical model in MIMIC with chain-like pairwise dependencies

衡量两个概率分布之间的距离, 可以采用 K-L 距离 (Kullback-Leiber divergence), 定义如下

$$H_l^\pi(\mathbf{x}) = h_l(X_{i_n}) + \sum_{j=1}^{n-1} h_l(X_{i_j}|X_{i_{j+1}}) \quad (6)$$

其中, $h(X) = -\sum_x p(X=x) \log p(X=x)$, $h(X|Y) = -\sum_y h(X|Y=y)p(Y=y)$, $h(X|Y=y) = -\sum_x p(X=x|Y=y) \log p(X=x|Y=y)$. MIMIC 算法在每一代中要根据选择后的优势群体构造最优的概率图模型 $p_l^\pi(\mathbf{x})$, 也就是搜索最优的排列 π^* 使 K-L 距离 $H_l^\pi(\mathbf{x})$ 最小化. 为了避免穷举所有 $n!$ 个可能排列, 在 De Bonet 的文章中, 提出了一种贪心算法, 搜索变量的近似最优排列.

在 MIMIC 算法中, 每一次循环, 都要根据选择的优势群体构造概率模型 p_l^π , 然后由 p_l^π 采样产生新的群体. 由于模型的复杂化, 采样方法也与变量无关的分布估计算法不同. 其基本思想是按照 π^* 的逆序, 对第 i_n, i_{n-1}, \dots, i_1 个变量依次采样, 构造一个完整的解向量, 描述如下:

- 1) $j = n$, 根据第 i_j 个变量的概率分布 $p_l(x_{i_j})$, 随机采样产生第 i_j 个变量;
- 2) 根据第 i_{j-1} 个变量的条件概率分布 $p_l(x_{i_{j-1}}|x_{i_j})$, 随机采样产生第 i_{j-1} 个变量;
- 3) $j = j - 1$, 如果 $j = 1$, 则一个完整的解向量构造完成; 否则转 2).

COMIT 算法是 PBIL 算法的提出者 Baluja 于 1997 年提出的^[11], 用于解决双变量相关的优化问题. COMIT 算法与 MIMIC 算法的最大不同在于 COMIT 算法的概率模型是树状结构, 如图 4 所示. 首先随机产生一个初始的群体, 从中选择比较优秀的个体集作为构造概率模型的样本集, 概率模型的构造方法采用机器学习领域中的 Chow 和 Liu 提出的方法, 然后按照 MIMIC 算法介绍的采样方法对概率树由上到下遍历, 反复采样构造新的种群.

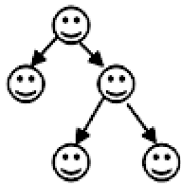


图 4 COMIT 算法中树状结构的概率图模型

Fig. 4 Probabilistic graphical model in COMIT with tree structure

除了 MIMIC 算法和 COMIT 算法外, 这类算法中比较代表性的还有 Pelikan 等人于 1999 年提出的 BMGA 算法^[12, 13]等, 限于篇幅本文不予详细介绍.

3.3 多变量相关的分布估计算法

最近几年研究更多的是多变量相关的分布估计算法. 在这种算法中, 变量之间的关系更加复杂, 需要更加复杂的概率模型来描述问题的解空间, 因此也需要更加复杂的学习算法来构造相应的概率模型. 这类算法中, 比较有代表性的是 ECGA (Extended compact genetic algorithm) 算法^[14]、FDA (Factorized distribution algorithms) 算法^[15~19]、贝叶斯优化算法 (Bayesian optimization algorithm, BOA)^[20~22]等等.

FDA 算法由德国学者 Mühlenbein 于 1998 年提出^[15~19], 可以解决多变量耦合的优化问题. 针对特定问题, FDA 算法用一个固定结构的概率图模型表示变量之间的关系, 进化过程中仅需要根据当前群体更新概率模型的参数. 这种算法的不足在于, 需要专家事先给出变量之间的关系. 对于数学形式已知的优化问题, 我们可以预先得到描述变量关系的概率图结构, 然后采用 FDA 算法, 反复进行参数学习和随机采样, 对问题进行求解. 但是对于很多数学形式复杂或者黑箱的优化问题, 则不能直接采用 FDA 算法进行求解.

ECGA 算法由美国 UIUC 大学的 Harik 等人提出^[14]. ECGA 算法是 cGA 的扩展, 在 ECGA 算法中, 将变量分成若干组, 每一组变量都与其他组变量无关. 如果用 $P(S_i)$ 表示第 i 组变量的“联合”概率分布, 由于任何两组变量之间无关, 那么所有变量的联合概率分布可以表示为 $P(X) = \prod_{i=1}^k P(S_i)$, 其中 k 表示变量可以分为 k 组, 并且 $\bigcup_{i=1}^k S_i = X$, X 表示所有变量组成的集合, $\forall i, j \in \{1, \dots, k\}, i \neq j, S_i \cap S_j = \emptyset$. 这种算法对于解决变量组之间无交叠的问题很有效, 但是对于变量组有交叠的问题, 则将构造出不准确的概率模型, 表现较差的性能.

BOA 算法是由美国 UIUC 大学的 Pelikan 等提出的^[3, 20~22], 2005 年 Pelikan 出版了关于 BOA 算法的第一本专著^[23]. 在 BOA 算法中, 由选择后的优势群体作为样本集构造贝叶斯网络, 然后对贝叶斯网络采样产生新一代群体, 反复进行. 贝叶斯网络是一个有向无环图, 可以表示随机变量之间的相互关系. 通过贝叶斯网络可以求得联合概率密度 $p(\mathbf{X}) = \prod_{i=1}^n p(X_i|\Pi_{X_i})$, 其中 $\mathbf{X} = (X_1, X_2, \dots, X_n)$ 表示问题的一个解向量, Π_{X_i} 表示贝叶斯网络中 X_i 的父节点集合, $p(X_i|\Pi_{X_i})$ 表示给定 X_i 父节点条件下 X_i 取值的条件概率. BOA 算法可以简单描述如下:

- 1) 随机产生初始群体 $P(0)$, $t = 0$;
- 2) 计算 $P(t)$ 中各个个体的适应值, 并选择优势群体 $S(t)$;
- 3) 由 $S(t)$ 作为样本集构造贝叶斯网络 B ;

4) 贝叶斯网络可以表示解的概率分布, 对贝叶斯网络 B 反复采样产生新的个体, 部分或者全部替换 $P(t)$, 生成新的群体 $P(t+1)$;

5) $t = t + 1$, 如果终止条件不满足, 转 2); 否则算法结束.

BOA 算法中最重要的是学习算法和采样算法. 贝叶斯网络的学习, 包括结构的学习和参数的学习. 结构的学习是指学习网络的拓扑结构, 参数的学习是指给定拓扑结构后学习网络中各个节点的条件分布概率. 贝叶斯网络的结构学习是一个 NP 难问题^[24], 在统计学习领域有着广泛深入的研究, 在文^[23]中采用的是贪心算法, 时间复杂度为 $O(n^2N + n^3)$, 其中 n 是问题维数, N 是样本个数. 贝叶斯网络的采样, 按照从父节点到子节点的顺序依次随机采样生成, 时间复杂度为 $O(n)$ ^[3, 20].

除此之外, 西班牙学者提出的 EBNA (Estimation of Bayesian network algorithm) 算法^[1, 25, 26]、德国学者提出的 LFDA (Learning factorized distribution algorithm) 算法等都采用了贝叶斯网络模型^[1, 26], 与 BOA 算法类似.

4 连续域的分布估计算法

前一节介绍的分布估计算法都是针对二进制编码问题的, 在实际工程和科学研究中, 研究定义域为实数的优化算法有着重要的意义. 遗传算法最初研究的是二进制编码的算法, 后来发展了实数编码的遗传算法. 分布估计算法的发展和遗传算法相似, 随着二进制编码算法的发展, 人们开始研究具有更普遍意义的、解决连续域问题的分布估计算法. 连续 EDA 是在离散 EDA 的基础上发展起来的, 很多算法的思想来源于或借鉴离散 EDA. 由于连续空间概率模型的复杂性给设计有效的分布估计算法增加了难度, 因此连续 EDA 的发展相对缓慢.

UMDA_c 算法是 UMDA 算法的扩展^[27], PBIL_c 算法是 PBIL 算法的扩展^[28], 这两种算法是比较有代表性的变量无关分布估计算法, 都采用了高斯分布作为描述连续解空间的概率模型. UMDA_c 算法和 PBIL_c 算法的不同在于采用了不同的构造方法更新高斯分布模型. 除此之外, SHCLVND (Stochastic hill climbing with learning by vectors of normal distributions) 算法也是研究较早的基于高斯分布的随机优化算法^[29]. 除了高斯分布外, 直方图分布是日本学者提出的另外一种描述连续解空间概率模型的有效方法^[30].

多变量相关性的连续域 EDA 算法主要有以下工作. 在 EMNA (Estimation of multivariate normal algorithm) 算法中, 采用多变量的高斯模型表

示解的概率分布, 在进化过程中的每一代采用最大似然估计方法, 估计多变量高斯分布的均值向量和协方差矩阵^[27, 31]. EGNA (Estimation of Gaussian networks algorithm) 算法是一种基于高斯图模型的分布估计算法, 高斯图模型中的有向边表示变量之间的关系, 每个变量都由一个高斯分布表示其概率密度. 在进化过程中, 需要根据当前群体重新构造高斯图网络结构^[32]. EMNA 算法和 EGNA 算法中, 都是采用了单峰的概率模型, 对于形状复杂的优化问题, 单峰的高斯分布模型不能有效地描述解在空间的分布. IDEA 算法^[33]是一种基于混合高斯分布和高斯核函数的分布估计算法, 这种方法虽然能够在解决多峰问题中一定程度上克服了 EMNA 算法和 EGNA 算法的缺点, 但是 IDEA 算法也没有充分考虑变量之间的关系.

连续域 EDA 算法的设计还面临很大困难. 主要原因在于: 每一个连续变量都有无限的取值, 这使得优化算法的搜索空间很大; 其次, 通过尽可能少的样本构造连续空间的概率模型是一个很难的问题, 特别是随着维数的增加, 构造连续域概率模型将变的更难. 通过有限样本集, 构造表示连续变量之间的耦合关系的概率图模型, 仍然需要进一步的研究.

5 分布估计算法的理论研究

分布估计算法是一种基于试验分析的随机优化算法, 其理论研究相对比较薄弱. 总的来看, 理论研究主要集中在二进制编码的分布估计算法, 并且主要针对概率图模型比较简单的算法, 主要涉及算法的收敛性分析、EDA 算法的时空复杂性等.

针对比较简单的变量无关的 EDA 算法的理论研究主要包括以下工作. 在 1997 年, Hohfeld 等人分析了 PBIL 算法中概率向量的随机变化过程, 证明了对于线性的二进制优化问题, PBIL 算法能保证群体收敛至全局最优解, 同时指出对于非线性问题 PBIL 算法可能陷入局部极值^[34]. Cristina 等人采用离散的动态系统对 PBIL 算法建模, 理论证明在学习速率趋向于 0 的情况下, PBIL 算法的概率向量的变化遵循动态系统的演化过程, 并指出 PBIL 算法只有在求解单模态问题时才能全局收敛^[35]. Mühlenbein 等人则分析了 UMDA 算法的收敛性, 得到与 PBIL 算法相同的结果^[8]. Shapiro 等人研究了 UMDA 算法中学习参数与问题之间的关系^[36].

针对 FDA 算法, Mühlenbein 等人证明了其解决可加性分解问题时的收敛性, 基本思想是将可加性分解问题转化为广义的线性问题, 借鉴 UMDA 收敛性理论成果, 证明如果一个函数可以分解为若干

分离的子函数,那么 FDA 算法能全局收敛^[15]. Q. Zhang 等对种群规模无穷大的 EDA 算法进行数学建模,理论证明了如果概率模型能精确反映选择后群体的分布,那么在采用比例选择、截断选择、二个体锦标赛选择这三种选择算子的情况下 EDA 算法是全局收敛的.该理论成果从侧面说明了概率模型在分布估计算法中的重要地位,实际上随着问题结构的复杂,很难建立精确的概率模型^[37, 38]. 2004年, Q. Zhang 通过对比 UMDA 和 FDA 两种算法,理论分析了高阶统计量对于分布估计算法性能的影响,证明了 FDA 算法在解决可加性分解的优化问题时能理论上收敛于全局最优解^[39]. R. Rastegar 等^[40]给出了种群规模无限大的情况下, EDA 算法收敛到全局最优所需要的代数.以上理论研究的局限性在于都对 EDA 算法做了某种假设,例如种群规模无限大、概率模型能精确反映问题结构等.

在空间复杂性方面, Y. Gao 等人研究了分布估计算法中 FDA 和 BOA 的空间复杂性,以可加性分解的函数作为研究对象,采用图和树表示可加性分解函数的变量关系,理论证明了这两种算法具有与问题规模呈指数级增长的空间复杂性.该研究说明了问题结构(即变量之间耦合关系)对问题求解的重要性,同时也说明了分布估计算法在解决变量耦合关系很复杂的问题时具有局限性^[41].在时间复杂度方面, Pelikan 等人研究了 BOA 算法的可扩展性,证明了 BOA 算法在解决一类可加性分解的黑箱优化问题时,适应值函数的计算次数与问题规模呈二次多项式或次二次多项式关系,并讨论了 BOA 算法中种群大小设置与问题规模之间的关系^[22, 23, 42].

另外,在分布估计算法的理论分析方面,熵理论成为重要的理论工具. Mühlenbein 等人采用最大熵逼近理论分析 EDA 算法,指出 EDA 算法的目标是最小化当前群体概率分布与目标概率分布之间的 K-L 距离;基于最大熵的理论框架,给出了分布估计算法的动态特性与统计物理中最大熵分布的关系^[43]. 2005年,古巴学者 Roberto 基于统计物理中的 Kikuchi 逼近理论,分析了采用马尔可夫随机场表示可分解的概率分布的理论基础,并提出基于无向图概率模型的分布估计算法 MN-EDA (Markov network estimation of distribution algorithm)^[44]. 2006年, Jiří O. 提出了基于熵的收敛性度量,并在此基础上给出了基于熵度量的 EDA 算法终止判定^[45].

分布估计算法的理论研究还有很多工作要做.以上研究成果都是针对二进制编码的分布估计算法,针对连续域 EDA 算法的理论研究还很少;目前的研究大都对 EDA 算法进行了假设,例如算法迭代无限次、种群规模无穷大、概率模型能精确反映问题结

构等,而在实际算法中这种假设是不成立的. EDA 算法的理论分析仍然是一个有挑战的课题.

6 基于概率模型的其他相关算法

最近几年,在进化计算领域涌现出了很多基于概率模型的进化算法,这些算法的共同特点是通过一个概率模型表示解的分布,然后通过采样产生新的解.这类算法中除了 EDA 相关文献中经常提到的上述算法外,还有很多其他的算法也是通过将统计学习和进化机制相结合而产生的有效的搜索算法.本文将这些算法归为分布估计算法,并将按照 EDA 的统一框架对这些算法进行简述.这类算法包括 SGA (Selfish gene algorithm) 基因算法^[46]、随机遗传算法 (StGA)^[47]、量子遗传算法^[48~50]等.

SGA 基因算法是意大利学者受 Selfish Gene 理论启发,于 1998 年提出的一种进化算法^[46]. SGA 算法用于解决二进制编码的问题.在算法中,把群体看成虚拟群体,虚拟群体由一个概率向量产生,然后通过竞争机制产生惩罚基因和奖励基因,用于更新概率向量,反复进行,直到产生满意解.可以发现 SGA 算法与 PBIL、UMDA 算法很相似.

随机遗传算法 (StGA) 是新加坡学者于 2004 年针对实数域优化问题提出的.在这种算法中,作者提出了一种随机编码策略,使得传统遗传算法中“点到点”的搜索变为“区域到区域”的搜索过程.高斯分布函数用于表示“区域”,区域内的个体通过随机采样产生. StGA 算法有与 EDA 算法相似的地方,与进化策略 (ES) 也比较相似^[47].

量子进化算法是由韩国学者 Han 等人提出的,借鉴了量子信息的表达方式和信息处理模式.这种算法与量子计算机上的量子算法不同,它是一种在传统计算机上运行的优化算法.在量子进化算法中,用量子比特串来表征描述解空间的概率模型,由量子比特串的“坍塌”操作(本质上是对量子比特串随机采样)产生传统意义上的个体,然后评价产生的个体并构造“量子门”,实现对量子比特串的更新^[48~50].

7 分布估计算法的研究热点

分布估计算法的研究方兴未艾,在算法设计、理论研究和实际应用方面仍需要做很多工作.作为一种新型的优化技术,分布估计算法的核心是解空间的概率模型,针对特定的优化问题,综合考虑搜索空间的结构、概率模型的学习算法、样本产生算法等,选择合适的概率模型,是发挥分布估计算法性能的关键所在.分布估计算法的本质是统计学习与进

化算法的结合,通过统计学习手段对群体中的数据进行分析 and 建模,以挖掘搜索空间结构的相关知识,更好提高搜索效率.实际上,在进化计算领域,采用机器学习的方法分析数据、指导搜索已经成为设计新算法的趋势^[51, 52].但是值得注意的是,统计学习的引入也会给进化计算带来了新的时间和空间开销,因此在科学研究或工程应用中,必须平衡统计学习和进化搜索的关系,设计算法的目的是更精确、快速的求解问题.

下面将分别介绍分布估计算法研究领域的不足之处和研究热点,并期望能通过下面的介绍使读者了解该技术的关键和难点,把握它的发展趋势.

1) 分布估计算法与其他搜索技术的混杂.智能混杂算法是当前智能优化算法的研究热点,可以融合多种优化算法的优势,提高算法的性能.目前文献中,已经出现了很多种基于分布估计算法的混杂算法,例如独立成分分析(ICA)与EDA的混杂^[53],DE(Differential evolution)算法与EDA的结合^[54],传统遗传算法与EDA的结合^[55],再励学习加强EDA算法的搜索能力^[56],爬山算法加强EDA的局部寻优能力^[57],基于最大熵理论的EDA^[58],采用集成学习方法增强EDA性能^[59]等等.还有很多与EDA相关的混杂算法,这里不一一列举.其他智能手段与分布估计算法的混杂,将提高EDA算法的搜索能力,更好的平衡算法的收敛性和群体多样性.

2) 分布估计算法的理论研究.如本文第5节所述,分布估计算法的理论研究相对比较薄弱,并且目前的研究工作仅仅局限于二进制编码分布估计算法.目前的研究大都对EDA算法进行了假设,例如算法迭代无限次、种群规模无穷大、概率模型精确反映问题结构,而在实际中这种假设是不成立的.理论研究的热点包括EDA算法的收敛性分析、时空复杂性分析等.探讨EDA算法与传统遗传算法之间的关系、深入了解EDA算法适合解决哪类问题(即问题结构与概率模型的关系)也是值得研究的问题.EDA算法理论基础方面仍需要做很多的工作,以更好的分析和挖掘这种算法的潜力.

3) 针对多目标优化和多模态优化的分布估计算法的研究.多目标优化算法是当前进化计算领域的研究热点.随着分布估计算法的发展,以及该算法在解决问题时所表现的优越性能,人们开始着手设计解决多目标优化问题的分布估计算法.多目标EDA算法的基本思想是,将多目标进化算法(MOEA)中处理多目标优化的策略引入到EDA算法中,并结合EDA算法的特点设计求解Parato解集更有效的算子.例如,mBOA将NSGA-II中的选择策略与BOA算法相结合,在试验中用mBOA解决欺骗的

多目标优化问题表现出了比NSGA-II更好的性能^[60, 61];H. Li等提出了混杂的多目标EDA算法,采用了加权平均的方法将多目标问题转化为单目标问题^[62];Laumanns等人将多目标进化算法SPEA2的选择策略引入到BOA中,用于解决多目标的01背包问题^[63];D. Thierens等人的工作是将连续域的分佈估计算法进行扩展,解决连续域多目标优化问题,采用混杂的高斯分布表示Pareto解集的分佈,并设计了专门的截断选择算子,保持群体的多样性^[64, 65].在多模态优化方面,代表性工作是Peña J. M等人提出的UEBNA算法,用于解决二进制编码的多模态优化问题,其特点在于设计了贝叶斯网络的无监督学习策略,避免基因流失,试验分析表明了该算法在解决多模态优化问题的优良性能^[66].

4) 强化分布估计算法的研究.分布估计算法的核心是概率图模型.但是我们可以看到,随着待解决问题的复杂化和概率图模型的复杂化,分布估计算法中概率模型的学习占用了大部分的时间和空间开销^[26],这必然将成为分布估计算法发展的瓶颈.因此,强化和改进分布估计算法是该领域研究的难点和热点.这方面的发展趋势主要包括分布估计算法的并行化、稀疏的概率模型学习策略、针对特定问题的专家知识的引入和辅助建模.

分布估计算法的并行化可以分为以下几种:a) 概率模型学习的并行化.这部分研究主要集中在概率模型比较复杂的BOA算法方面^[26, 67~69].b) 采样过程的并行化.群体中的个体是由概率模型随机采样产生的,一般的,个体之间是相互独立的,因此群体的产生可以并行进行.这方面的研究较少,主要有Pelikan等研究了BMDA算法中采样的并行化^[70].除此之外,进化计算领域广泛研究的基于多种群的并行化策略、Island模型并行化、适应值函数的并行化计算等都可以引入到分布估计算法的研究领域^[71, 72].

稀疏的概率模型学习策略的基本思想是,在BOA算法中,贝叶斯网络的结构每隔若干代更新一次,贝叶斯网络的参数每一代都更新,这样由于参数学习的时间复杂度远远低于结构学习,稀疏建模将提高分布估计算法的效率^[73].

将先验知识引入到分布估计算法的概率模型学习中^[74~76],主要工作包括:Baluja S.提出将专家知识引入到分布估计算法中,以减少概率模型学习带来的时间开销^[74];Wright A.等通过“联结关系”检测算法,得到变量之间的相互关系,即概率模型的结构,在后续的进化过程中仅仅需要进行概率模型参数学习^[75].

5) 分布估计算法在非二进制编码领域的研究.EDA算法源于对二进制编码的遗传算法的研究,因

此大部分研究成果针对二进制编码的优化问题. 实际上, 大量的优化问题是多值离散的组合优化问题、排列优化问题和连续优化问题, 特别是在排列优化^[77~79]和连续优化问题上, EDA 算法的研究进展相对缓慢. 连续域 EDA 算法研究的难点在于, 对于复杂的多峰的、强耦合的、非线性连续优化问题, 需要复杂的连续域概率图模型表示解空间分布, 目前所采用的简单高斯分布、线性关系的高斯图模型仍有不足; 但是复杂的概率图模型也会给算法的时间复杂性带来问题.

6) 分布估计算法的应用. 试验分析表明分布估计算法在求解问题时表现出了比一般遗传算法更好的性能, 应用分布估计算法解决工程和科学中的复杂优化问题具有很大的潜力. 目前, 分布估计算法已经在众多领域得到了成功的应用, 例如, 基于 EDA 算法的汽车齿轮机械结构的优化设计^[80]、采用贝叶斯优化算法进行特征选择^[81]、不精确图形的模式匹配^[82]、基于 EDA 算法的软件测试^[83]、EDA 算法在癌症分类中的应用^[84]、生物信息学中的特征提取^[85]、军事天线的优化设计^[86]等等. 分布估计算法的应用已经渗透到了模式识别^[81, 82, 84, 85, 87, 88]、运筹学^[79, 89]、工程优化^[80, 86]、机器学习^[88, 90]和生物信息^[85]等众多领域. 毋庸置疑, 使用分布估计算法解决在科学研究和工程应用中碰到的优化问题将是未来几年的研究热点.

8 总结

分布估计算法是进化计算领域内一个崭新的分支, 由于这种算法能够显式的表示变量之间的关系, 能快速、准确、可靠的解决传统遗传算法束手无策的一类优化问题, 在短短的几年时间内, 引起了国际学术界的广泛关注, 因此分布估计算法成为国际进化计算领域的研究热点, 也逐渐成为解决实际工程优化问题的热门工具. 分布估计算法是遗传算法和统计学习的结合, 通过统计学习的手段建立描述解空间个体分布的概率模型, 然后对概率模型随机采样产生新的群体, 如此反复进行, 实现群体的进化. 分布估计算法中没有传统的交叉、变异等遗传操作, 通过对整个群体建立数学模型, 直接描述整个群体的进化趋势, 是对生物进化“宏观”层面上的数学建模. 分布估计算法给人类解决复杂的优化问题提供了新的工具, 它通过概率模型可以描述变量之间的相互关系, 从而对解决非线性、变量耦合的优化问题更加有效. 分布估计算法的研究将极大丰富混杂智能优化算法的研究内容, 给优化问题的求解提供新的思路, 具有巨大的科研价值和应用潜力. 分布估计算法的研究方兴未艾. 希望通过本文介绍, 使读

者对该领域有一个初步的认识, 并能将此方法应用到科学和工程领域中的各种优化问题中去.

References

- 1 Larrañaga P, Lozano J A. *Estimation of Distribution Algorithms. A New Tool for Evolutionary Computation*. Boston: Kluwer Academic Publishers, 2002
- 2 Pelikan M, Goldberg D E, Lobo F. A Survey of Optimization by Building and Using Probabilistic Models. IlliGAL Report No. 99018, University of Illinois at Urbana-Champaign, Illinois Genetic Algorithms Laboratory, Urbana, Illinois, 1999
- 3 Pelikan M, Goldberg D E, Cantú-Paz E. Linkage Problem, Distribution Estimation, and Bayesian Networks. IlliGAL Report No. 98013, University of Illinois at Urbana-Champaign, Illinois Genetic Algorithms Laboratory, Urbana, Illinois, 1998
- 4 Baluja S. Population-Based Incremental Learning: A Method for Integrating Genetic Search Based Function Optimization and Competitive Learning. Technical Report CMU-CS-94-163, Pittsburgh, PA: Carnegie Mellon University, 1994
- 5 Baluja S. An Empirical Comparison of Seven Iterative and Evolutionary Function Optimization Heuristics. Technical Report CMU-CS-95-193, Computer Science Department, Carnegie Mellon University, 1995
- 6 Baluja S, Caruana R. Removing the genetics from standard genetic algorithm. In: *Proceedings of the International Conference on Machine Learning*. San Mateo, USA: Morgan Kaufmann, 1995. 38~46
- 7 Mühlenbein H, Paass G. From recombination of genes to the estimation of distributions I. Binary parameters. *Parallel Problem Solving from Nature - PPSN IV*, Berlin, 1996. 178~187
- 8 Mühlenbein H. The equation for response to selection and its use for prediction. *Evolutionary Computation*, 1997, 5(3): 303~346
- 9 Harik G R, Lobo F G, Goldberg D E. The compact genetic algorithm. In: *Proceedings of the IEEE Conference on Evolutionary Computation*. IEEE, Indianapolis, USA, 1998. 523~528
- 10 De Bonet, J S, Isbell C L, Viola P. MIMIC: Finding optima by estimating probability densities. *Advances in Neural Information Processing Systems*, Cambridge: MIT Press, 1997. 9: 424~430
- 11 Baluja S, Davies S. Using optimal dependency-trees for combinatorial optimization: Learning the structure of the search space. In: *Proceedings of the 14th International Conference on Machine Learning*. San Francisco, CA: Morgan Kaufmann, 1997. 30~38
- 12 Pelikan M, Mühlenbein H. The bivariate marginal distribution algorithm. *Advances in Soft Computing - Engineering Design and Manufacturing*, London, Springer-Verlag, 1999, 521~535
- 13 Pelikan M, Mühlenbein H. Marginal distributions in evolutionary algorithms. In: *Proceedings of the International Conference on Genetic Algorithms*. Brno, Czech Republic: Technical University of Brno, Publisher, 1998. 90~95
- 14 Harik G. Linkage Learning via Probabilistic Modeling in the ECGA. IlliGAL Report No. 99010, Illinois Genetic Algorithms Laboratory, University of Illinois, Urbana-Champaign, Illinois, 1999

- 15 Mühlenbein H, Mahnig T. Convergence theory and applications of the factorized distribution algorithm. *Journal of Computing and Information Technology*, 1999, **7**(1): 19~32
- 16 Mühlenbein H, Mahnig T. The Factorized distribution algorithm for additively decomposable functions. In: Second Symposium on Artificial Intelligence. Adaptive Systems. CIMAFA 99, La Habana, 1999. 301~313
- 17 Mühlenbein H, Mahnig T. FDA - a scalable evolutionary algorithm for the optimization of additively decomposed functions. *Evolutionary Computation*, 1999, **7**(4): 353~376
- 18 Kallel L, Naudts B, Rogers A. *Theoretical Aspects of Evolutionary Computing*. London, UK: Springer-Verlag, 2001. 137~176
- 19 Mühlenbein H, Mahnig T, Ochoa A R. Schemata, distributions and graphical models in evolutionary optimization. *Journal of Heuristics*, 1999, **5**: 215~247
- 20 Pelikan M, Goldberg D E, Cantú-Paz E. BOA: The Bayesian optimization algorithm. In: Proceedings of the Genetic and Evolutionary Computation Conference GECCO-99, Orlando, FL: 1999. 525~532
- 21 Pelikan M. A Simple Implementation of Bayesian Optimization Algorithm in C++(Version1.0). Illigal Report No. 99011, Illinois Genetic Algorithms Laboratory, University of Illinois at Urbana-Champaign, Urbana, Illinois, 1999
- 22 Pelikan M, Goldberg D E, Cantú-Paz E. Bayesian Optimization Algorithm, Population Sizing, and Time to Convergence. IlliGAL Report No. 2000001, Illinois Genetic Algorithms Laboratory, University of Illinois at Urbana-Champaign, Urbana, Illinois, 2000
- 23 Pelikan M. *Hierarchical Bayesian Optimization Algorithm: Toward a New Generation of Evolutionary Algorithms*. New York: Springer-Verlag, 2005
- 24 Chickering D M, Geiger D, Heckerman D. Learning Bayesian Networks Is NP-hard. Technical Report MSR-TR-94-17, Microsoft Research, Redmond, Wirginia, 1994
- 25 Larrañaga P, Etxeberria R, Lozano J A, Peña J M. Combinatorial optimization by learning and simulation of Bayesian networks. In: Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence. Stanford: 2000. 343~352
- 26 Jiří O. Parallel Estimation of Distribution Algorithms. [Ph.D. dissertation], Brno University of Technology, 2002
- 27 Larrañaga P, Etxeberria R, Lozano J A, Peña J M. Optimization in continuous domains by learning and simulation of Gaussian networks. In: Proceedings of the 2000 Genetic and Evolutionary Computation Conference Workshop Program. Las Vegas, Nevada, 2000. 201~204
- 28 Sebag M, Ducoulombier A. Extending population-based incremental learning to continuous search spaces. In: Back Th, Eiben G, Schoenauer M, Schwefel H P, editors, Proceedings of the 5th Conference on Parallel Problem Solving from Nature - PPSN V, Springer-Verlag, 1998. 418~427
- 29 Rudlof S, Köppen M. Stochastic Hill Climbing by Vectors of Normal Distributions. In: Proceedings of the First Online Workshop on Soft Computing (WSC1), Nagoya, Japan, 1996
- 30 Tsutsui S, Pelikan M, Goldberg D E. Evolutionary Algorithm Using Marginal Histogram Models in Continuous Domain. IlliGAL Report No. 2001019, University of Illinois at Urbana-Champaign, Urbana, Illinois, 2001
- 31 Larrañaga P, Etxeberria R, Lozano J A, Peña J M. Optimization by Learning and Simulation of Bayesian and Gaussian Networks. Technical Report EHU-KZAA-4/99, University of the Basque Country, Spain, 1999
- 32 Larrañaga P, Lozano J A, Bengoetxea E. Estimation of Distribution Algorithms Based on Multivariate Normal and Gaussian Networks. Technical Report KZZA-IK-1-01, Department of Computer Science and Artificial Intelligence, University of the Basque Country, 2001
- 33 Bosman P A N, Thierens D. Expanding from discrete to continuous estimation of distribution algorithms: The IDEA. In: Proceedings of the 6th International Conference on Parallel Problem Solving from Nature - PPSN VI, Springer-Verlag: 2000. 767~776
- 34 Hohfeld M, Rudolph G. Towards a theory of population based incremental learning. In: Proceedings of the 4th International Conference on Evolutionary Computation. IEEE, 1997. 1~5
- 35 Cristina G, Lozano J A, Larrañaga P. Analyzing the PBIL algorithm by means of discrete dynamical systems. *Complex Systems*, 2001, **12**(4), 465~479
- 36 Shapiro J L. Drift and scaling in estimation of distribution algorithms. *Evolutionary Computation*, 2005, **13**(1): 99~123
- 37 Zhang Q, Mühlenbein H. On the convergence of a class of estimation of distribution algorithms. *IEEE Transactions on Evolutionary Computation*, 2004, **8**(2): 127~136
- 38 Zhang Q. On the convergence of a factorized distribution algorithm with truncation selection[Online], available: <http://cswww.essex.ac.uk/staff/zhang/EDAWEB/>, May 10, 2006
- 39 Zhang Q. On stability of fixed points of limit models of univariate marginal distribution algorithm and factorized distribution algorithm. *IEEE Transactions on Evolutionary Computation*, 2004, **8**(1): 80~93
- 40 Rastegar R, Meybodi M R. A study on the global convergence time complexity of estimation of distribution algorithms. *Lecture Notes in Computer Science*, 2005, **3641**: 441~450
- 41 Gao Y, Culberson J. Space complexity of estimation of distribution algorithms. *Evolutionary Computation*, 2005, **13**(1): 125~143
- 42 Pelikan M, Sastry K, Goldberg D E. Scalability of the Bayesian optimization algorithm. *International Journal of Approximate Reasoning*, 2002, **31**(3): 221~258
- 43 Mühlenbein H, Höns R. The estimation of distributions and the minimum relative entropy principle. *Evolutionary Computation*, 2005, **13**(1): 1~27
- 44 Roberto S. Estimation of distribution algorithms with Kikuchi approximations. *Evolutionary Computation*, 2005, **13**(1): 67~97
- 45 Jiří O. Entropy-based convergence measurement in discrete estimation of distribution algorithms. In: Lozano et al. (Eds): Towards a New Evolutionary Computation: Advances in the Estimation of Distribution Algorithms. Springs-Verlag, 2002. 125~142
- 46 Corno F, Reorda M S, Squillero G. The selfish gene algorithm: a new evolutionary optimization strategy. In: Proceedings of the 13th Annual ACM Symposium on Applied Computing (SAC'98). Atlanta, Georgia, USA: 1998. 349~355
- 47 Tu Z, Lu Y. A robust stochastic genetic algorithm (StGA) for global numerical optimization. *IEEE Transaction on Evolutionary Computation*, 2004, **8**(5): 456~470
- 48 Han K H, Kim J H. Quantum-Inspired Evolutionary Algorithm for a class of combinatorial optimization. *IEEE Transactions on Evolutionary Computation*, 2002, **6**(6): 580~593

- 49 Han K H, Kim J H. Quantum-inspired evolutionary algorithms with a new termination criterion, H ϵ gate, and two-phase scheme. *IEEE Transactions on Evolutionary Computation*, 2004, **8**(2): 156~169
- 50 Zhou S, Sun Z. A new approach belonging to EDAs: quantum-inspired genetic algorithm with only one chromosome. In: Proceedings of First International Conference on Natural Computation (ICNC'05), Changsha, China, 2005. 141~150
- 51 Michalski R S. Learnable evolution model: Evolutionary processes guided by machine learning. *Machine Learning* 2000, **38**: 9~40
- 52 Wojtusiak J, Michalski R S. The LEM3 System for Non-darwinian Evolutionary Computation and Its Application to Complex Function Optimization. Reports of the Machine Learning and Inference Laboratory, MLI 05-2, George Mason University, Fairfax, VA, Oct. 2005
- 53 Sun J, Zhang Q, Tsang E. DE/EDA: A new evolutionary algorithm for global optimization. *Information Sciences*, 2005, **169**: 249~262
- 54 Cho D Y, Zhang B T. Evolutionary continuous optimization by distribution estimation with variational Bayesian independent component analyzers mixture model. *Lecture Notes in Computer Science*, 2004, **3242**: 212~221
- 55 Peña J M, Robles V, Larrañaga P, Herves V, Rosales F, Pérez M S. GA-EDA: Hybrid evolutionary algorithm using genetic and estimation of distribution algorithms. *Lecture Notes in Computer Science*, 2004, **3029**: 361~371
- 56 Topon K P, Hitoshi I. Reinforcement learning estimation of distribution algorithm. In: Proceedings of Genetic and Evolutionary Computation Conference 2003. *Lecture Notes in Computer Science*, 2003, **2724**: 1259~1270
- 57 Laurent G, Rodolphe L R, Raphael T H. A study of the effects of dimensionality on stochastic hill climbers and estimation of distribution algorithms. *Lecture Notes in Computer Science*, 2004, **2936**: 27~38
- 58 Wright A, Poli R, Stephens C, Langdon W B, Pulavarty S. An estimation of distribution algorithm based on maximum entropy. In: Proceedings of the Genetic and Evolutionary Computation Conference. *Lecture Notes in Computer Science*, 2004, **3103**: 343~354
- 59 Zhou S, Sun Z. Can ensemble method convert a 'weak' evolutionary algorithm to a 'strong' one? In: Proceedings of International Conference on Computational Intelligence for Modeling, Control and Automation, Vienna, Austria, 2005. **2**: 68~74
- 60 Nazan K, Goldberg D E, Pelikan M. Multi-Objective Bayesian Optimization Algorithm. IlliGAL Report No. 2002009, University of Illinois at Urbana-Champaign, Urbana, Illinois, 2002
- 61 Pelikan M, Sastry K, Goldberg D E. Multiobjective hBOA, clustering, and scalability. In: Proceedings of the 2005 Conference on Genetic and Evolutionary Computation. New York, USA: ACM Press, 2005. 663~670
- 62 Li H, Zhang Q, Tsang E P K, Ford J. Hybrid estimation of distribution algorithm for multiobjective Knapsack problem. In: Proceedings of 4th European Conference on Evolutionary Computation in Combinatorial Optimization. Coimbra, Portugal: 2004, 145~154
- 63 Laumanns M, Ocenasek J. Bayesian optimization algorithms for multi-objective optimization. In: Proceedings of the 7th International Conference on Parallel Problem Solving from Nature. London, UK: Springer-Verlag, 2002. 298~307
- 64 Thierens D, Bosman P A N. Multi-objective mixture-based iterated density estimation evolutionary algorithms. In: Proceedings of Genetic and Evolutionary Computation Conference. San Francisco, USA: Morgan Kaufmann, 2001. 663~670
- 65 Bosman P A N, Thierens D. Multi-objective optimization with the naive MIDEA. In: Towards a New Evolutionary Computation. Advances in Estimation of Distribution Algorithms. Springer-Verlag, Berlin: 2006. 123~157
- 66 Peña J M, Lozano J A, Larrañaga P. Globally multimodal problem optimization via an estimation of distribution algorithm based on unsupervised learning of Bayesian networks. *Evolutionary Computation*, 2005, **13**(1): 43~66
- 67 Lozano J A, Sagarna R, Larrañaga P. Parallel estimation of distribution algorithms. In: Estimation of Distribution Algorithms. A New Tool for Evolutionary Computation. Boston: Kluwer Academic Publishers, 2002. 125~142
- 68 Mendiburu A, Miguel-Alonso J, Lozano J A. Implementation and performance evaluation of a parallelization of estimation of Bayesian networks algorithms. *Parallel Processing Letters*, 16(1), 133~148
- 69 Jiří O, Josef S. The parallel Bayesian optimization algorithm. In: Proceedings of the European Symposium on Computational Intelligence. Košice, SK, Springer, 2000. 61~67
- 70 Pelikan M, Mühlenbein H. The bivariate marginal distribution algorithm. In: Advances in Soft Computing-Engineering Design and Manufacturing. London: Springer-Verlag, 1998. 521~535
- 71 Alba E, Troya J M. A survey of parallel distributed genetic algorithms. *Complexity*, John Wiley & Sons, Inc, 1999, **4**(4): 31~52
- 72 Julio M, Enrique A, Alberto O. A parallel island model for estimation of distribution algorithms. *Studies in Fuzziness and Soft Computing*, 2006, **192**: 159~186
- 73 Pelikan M, Sastry K, Goldberg D E. Sporadic model building for efficiency enhancement of hBOA. In: Proceedings of Genetic and Evolutionary Computation Conference, ACM Press, 2006. 405~412
- 74 Baluja S. Using a priori knowledge to create probabilistic models for optimization. *International Journal of Approximate Reasoning*, 2002, **31**(3): 193~220
- 75 Wright A, Pulavarty S. On the convergence of an estimation of distribution algorithm based on linkage discovery and factorization. In: Proceedings of Genetic and Evolutionary Computation Conference. ACM Press, 2005. 695~702
- 76 Leonardo R E, Traleski R, Pozo A. Using prior knowledge to improve the performance of an estimation of distribution algorithm applied to feature selection. In: Proceedings of the Fifth International Conference on Hybrid Intelligent Systems. IEEE, 2005. 393~398
- 77 Tsutsui S, Pelikan M, Ghosh A. Edge histogram based sampling with local search for solving permutation problems. *The International Journal of Hybrid Intelligent Systems*, 2006, **3**(1): 11~22
- 78 Bosman P A N, Thierens D. Permutation optimization by iterated estimation of random keys marginal product factorizations. In: Proceedings of the 7th International Conference on Parallel Problem Solving from Nature - PPSN VII. Granada, Spain: Springer Berlin Heidelberg, 2002. 331~340
- 79 Aickelin U, Li Jing-Peng. An estimation of distribution algorithm for nurse scheduling. *Annals of OR*, to be published

- 80 Simionescu P A, Beale D G, Dozier G V. Teeth-number synthesis of a multispeed planetary transmission using an estimation of distribution algorithm. *Journal of Mechanical Design*, 2006, **128**(1): 108~115
- 81 I.Inza P, Larrañaga P, Etxebarria B S. Feature subset selection by Bayesian network-based optimization. *Artificial Intelligence*, 2000, **123**(1-2): 157~184
- 82 Cesar R M, Bengoetxea E, Bloch I, Larrañaga P. Inexact graph matching for model-based recognition: Evaluation and comparison of optimization algorithms. *Pattern Recognition*, 2005, **38**(11): 2099~2113
- 83 Sagarna R, Lozano J. On the performance of estimation of distribution algorithms applied to software testing. *Applied Artificial Intelligence*, 2005, **19**(5): 457~489
- 84 Blanco R, Larrañaga P, Inza I. Gene selection for cancer classification using wrapper approaches. *International Journal of Pattern Recognition and Artificial Intelligence*, 2004, **18**(8): 1373~1390
- 85 Saeyns Y, Degroev S, Aeyels D, van de Peer Y, Rouzé P. Fast feature selection using a simple estimation of distribution algorithm: a case study on splice site prediction. *Bioinformatics*, 2003, **19**(s2): 179~188
- 86 Santarelli S, Yu T, Goldberg D E, Altschuler E, O' Donnell T, Southall H. Military antenna design using simple and competent genetic algorithms. *Mathematical and Computer Modelling*, 2006, **43**(9-10): 990~1022
- 87 Yang X, Birkfellner W, Niederer P. Optimized 2D/3D medical image registration using the estimation of multivariate normal algorithm (EMNA). In: Proceedings of Biomedical Engineering, Innsbruck, Austria, 2005, 163~168
- 88 Joaquin R, Roberto S. Improving the discovery component of classifier systems by the application of estimation of distribution algorithms. In: Proceedings of the Students Sessions, ACAI'99. Chania, Greece, 1999. 43~44
- 89 Zhang Q, Sun J, Tsang E P K, Ford J. Combination of guided local search and estimation of distribution algorithm for solving quadratic assignment problem. In: Proceedings

of the Brid of a Feather Workshops, Genetic and Evolutionary Computation Conference. Berlin: Springer-Verlag, 2004. 42~48

- 90 Cantú-Paz E. Pruning neural networks with distribution estimation algorithms. In: Proceedings of the Genetic and Evolutionary Computation Conference GECCO 2003. Berlin: Springer-Verlag, 2003. 790~800



周树德 清华大学计算机系智能技术与系统国家重点实验室博士研究生。2003 年获得西安电子科技大学计算机系学士学位。主要研究方向为进化计算。本文通信作者。E-mail: zsd03@mails.tsinghua.edu.cn

(**ZHOU Shu-De** Ph. D. candidate in State Key Lab of Intelligent Technology and Systems, Department of Computer Science and Technology, Tsinghua University. He received his B. S. degree from Xidian University in 2003. His research interest includes evolutionary computation, etc. Corresponding author of this paper.)



孙增圻 清华大学计算机系智能技术与系统国家重点实验室教授，博士生导师。主要研究方向为机器人、智能控制、进化计算、模糊系统和神经网络。E-mail: szq-doc@tsinghua.edu.cn

(**SUN Zeng-Qi** Professor in State Key Lab of Intelligent Technology and Systems, Department of Computer Science and Technology, Tsinghua University. His research interests include robotics, intelligent control, evolutionary computation, fuzzy system, and neural networks.)