

建立在一般结构 Gauss 网络上的分布估计算法

钟伟才 刘静 刘芳 焦李成

(西安电子科技大学智能信息处理研究所 西安 710071)

摘要: 提出了一种建立在一般结构 Gauss 网络上的分布估计算法。一方面, 它无需进行 Gauss 网络结构的学习, 从而大大减少了计算量; 另一方面, 一般结构 Gauss 网络不是近似网络, 因而可获得精度很高的联合概率密度函数。针对该网络, 采用了一种无需计算条件概率密度函数的产生样本方法, 有效地减少了网络参数学习的计算开销。实验结果表明, 与已有建立在非一般结构 Gauss 网络上的高阶分布估计算法相比, 本文算法具有更高的稳定性和更强的寻优能力。

关键词: 进化计算, 分布估计算法, Gauss 网络

中图分类号: TP391 **文献标识码:** A **文章编号:** 1009-5896(2005)03-0467-04

Estimation of Distribution Algorithm Based on Generic Gaussian Networks

Zhong Wei-cai Liu Jing Liu Fang Jiao Li-cheng

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

Abstract Estimation of Distribution Algorithms (EDAs) available in continuous domains are based on non-generic Gaussian networks. The computational cost for learning this kind of networks is very great, moreover the low accuracy of the joint pdf will be resulted because the greedy algorithm is used to learn the Gaussian networks. To overcome these disadvantages, an Estimation of Distribution Algorithm based on generic Gaussian Networks (GN-EDA) is presented. It leads to the low computational cost by no structure learning of Gaussian networks. In the meanwhile, a generic Gaussian network is not an approximate one, so the joint pdf is of high accuracy. Due to an effective sampling is adopted, the computational cost for parameters learning is great reduced. The experimental results show that GN-EDA achieves a more stable performance and a stronger ability in searching the global optima.

Key words Evolutionary computation, Estimation of distribution algorithm, Gaussian networks

1 引言

由于交叉和变异算子不能保证积木块假设成立, 遗传算法对于某些问题, 如骗问题, 表现出了很差的性能, 这就导致了另一类进化算法的出现——分布估计算法(Estimation of Distribution Algorithms, EDAs)。与传统进化算法不同, 它不使用交叉和变异算子, 而是根据当前种群中较好的个体建立概率分布模型, 然后根据这个模型产生个体以引导算法的搜索。由于 EDAs 较传统进化算法有更强的理论基础, 已成为当前进化计算的研究热点^[1-9]。

EDAs 的思想是从当前种群 $Pop(k)$ 中按照一定的选择方式选出一部分较好的个体 $Pop^s(k)$, 并建立反映个体 $x=(x_1, \dots, x_n)$ 内 n 个变量相互作用的概率分布模型, 然后利用所建的分模型产生新的个体, 以形成下一代种群 $Pop(k+1)$, 如此迭代, 直到满足终止条件。设 $p^s(x, k)$ 为种群 $Pop^s(k)$ 的实际分布, 文献[7]证明如果下一代种群分布 $p(x, k+1)=p^s(x, k)$

且选择方式为比例、截断或锦标赛选择, 则 EDAs 收敛于全局最优解。因此 EDAs 的关键问题是使 $p(x, k+1)$ 在合理的计算代价下尽量逼近 $p^s(x, k)$ 。为此提出了各种算法, 按照处理变量间相互作用的能力, 可分为一阶分布估计算法, 如 PBIL^[1] (Population-Based Incremental Learning), UMDA^[2] (Univariate Marginal Distribution Algorithm)等; 二阶分布估计算法, 如 MIMIC^[3] (Mutual Information Maximization for Input Clustering), BMDA^[4] (Bivariate Marginal Distribution Algorithm)等; 及高阶分布估计算法, 如 BOA^[5] (Bayesian Optimization Algorithm)、EGNA_{ee}^[6] (Estimation of Gaussian Networks Algorithm by edge exclusion)、EGNA_{BGe}^[6] (Estimation of Gaussian Networks Algorithm by BGe metric)等。

实际中的复杂问题, 其变量间往往存在着较强的相互作用, 因而采用高阶分布估计算法比较合适。EGNA_{ee}、EGNA_{BGe} 是两种建立在 Gauss 网络上的处理连续数据的高阶分布估计

算法, 其中 $EGNA_{ec}$ 采用根据条件独立关系对弧进行删除的网络结构学习算法, 而 $EGNA_{BGe}$ 则采用将 BGe 作为准则的网络结构学习算法。虽然它们都能够描述变量间的高阶相互作用, 但由于搜索最优网络结构是 NP 困难问题, 一般都采用贪婪搜索, 这就大大降低了联合概率密度函数的准确性, 同时它们的网络参数学习也需要复杂的矩阵运算。我们认为 Gauss 网络可分为两类: 一般结构 Gauss 网络和非一般结构 Gauss 网络 (定义见下节)。在一般结构 Gauss 网络上进行概率推理需要很大的计算量, 因此现有方法都采用非一般结构 Gauss 网络。但由于非一般结构 Gauss 网络是由一般结构 Gauss 网络简化而得的, 显然其联合概率密度函数的准确度没有一般结构 Gauss 网络的高。我们认为分布估计算法并不需要进行概率推理, 只需估计出变量的联合概率密度函数即可, 并且在已假设数据服从高斯分布的前提下, 可以直接采用一般结构的 Gauss 网络作为算法的概率分布模型。所以我们提出了一种建立在一般结构 Gauss 网络上的分布估计算法 (Estimation of Distribution Algorithm based on Generic Gauss Networks, GN-EDA)。GN-EDA 无需进行 Gauss 网络结构的学习, 大大减少了计算量, 而且它可获得高精度的联合概率密度函数。为了减少网络参数学习所需的计算量, GN-EDA 采用了一种不使用条件概率密度函数来产生样本的有效方法。

2 建立在一般结构 Gauss 网络上的分布估计算法

2.1 一般结构 Gauss 网络

分布估计算法的关键在于找到一个能反映 x_1, \dots, x_n 间相互作用的 n 维概率模型。使用图来表示变量间的依赖关系是一种常用的、自然而直观的方式。图中的每个结点代表一个变量, 每条边代表两变量间的关系。若结点 x_i 存在一条指向 x_j 的边, 则称 x_i 是 x_j 的父结点, 由此而得的模型称为概率图模型。若使用的是有向无环图, 则称相应的概率图模型为 Bayes 网络, 见定义 1。

定义 1 一个 Bayes 网络是一个二元组 $\mathcal{B} = \langle G, \Theta \rangle$, 其中

(1) 网络结构 $G = \{\Pi_1, \dots, \Pi_n\}$ 是一个有向无环图 (Directed Acyclic Graph, DAG), 其结点集合为 $V = \{x_1, \dots, x_n\}$, $n \geq 1$, Π_i ($i = 1, \dots, n$) 为结点 x_i 的父结点集合;

(2) 网络参数 $\Theta = \{P(x_i | \Pi_i) | x_i \in V, i = 1, \dots, n\}$ 是一组条件概率的集合。

在该模型中, 变量的联合概率分布可表示为所有变量在其父结点变量下的条件概率的乘积:

$$P(\mathbf{x}) = \prod_{i=1}^n P(x_i | \Pi_i) \quad (1)$$

其中 $P(x_i | \Pi_i)$ 为 x_i 的条件概率。定义 1 主要针对离散随机变

量, 对于连续随机变量, 一般都假设边缘概率分布和条件概率分布为正态分布, 这时称获得的网络为 Gauss 网络。我们认为, 根据网络结构可将 Gauss 网络分为两类, 见定义 2。

定义 2 若 Gauss 网络的结构为如图 1 所示的完全连接的有向无环图, 则称其为一般结构 Gauss 网络, 否则称为非一般结构 Gauss 网络。

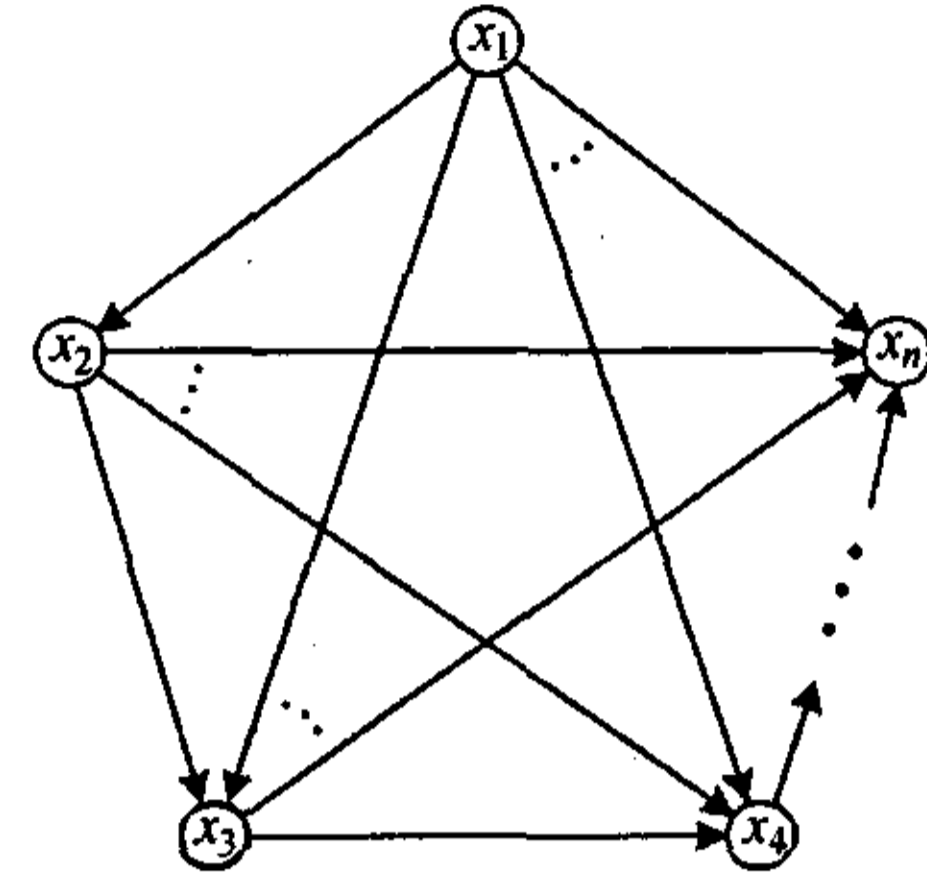


图 1 一般结构 Gauss 网络模型

假设 n 维变量 \mathbf{x} 的联合概率密度函数为多元正态分布:

$$P(\mathbf{x}) \equiv \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \equiv (2\pi)^{-\frac{n}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right\} \quad (2)$$

其中 $\boldsymbol{\mu}$ 是 n 维均值向量, $\boldsymbol{\Sigma}$ 是 $n \times n$ 的协方差矩阵, 而 $|\boldsymbol{\Sigma}|$ 是 $\boldsymbol{\Sigma}$ 的行列式, $\boldsymbol{\Sigma}^{-1}$ 是 $\boldsymbol{\Sigma}$ 的逆矩阵。由概率论中的乘法定理可得式(3):

$$P(\mathbf{x}) = \prod_{i=1}^n P(x_i | x_1, \dots, x_{i-1}) = \prod_{i=1}^n P(x_i | \Pi_i) \quad (3)$$

其中 $\Pi_i = (x_1, \dots, x_{i-1})$ 。 $P(\mathbf{x})$ 就对应于一个一般结构 Gauss 网络, 由定理 1 可知, 这样的网络有 $n!$ 个。

定理 1 设 $\pi = (l_1, l_2, \dots, l_n)$ 是 $(1, 2, \dots, n)$ 的任何一个排列, 则有

$$P_{\pi}(\mathbf{x}) = \prod_{i=1}^n P(x_{l_i} | \Pi_{l_i}) = P(\mathbf{x}) \quad (4)$$

其中 $\Pi_{l_i} = (l_1, \dots, l_{i-1})$ 。

定理 1 由概率论中的乘法定理可得证。该定理显示所有这 $n!$ 个网络的联合概率密度函数是相同的。

定理 2 记所有 $n!$ 种一般结构 Gauss 网络的结构为 $G_1, G_2, \dots, G_{n!}$, 若 G 为任一 Gauss 网络的结构, 则 G 必为 $G_1, G_2, \dots, G_{n!}$ 中某个图的子图。

证明 由 Gauss 网络是一种 Bayes 网络可知 G 为一个有向无环图。因为给出一个有向图中所有结点的父结点将完全确定一个有向图 G , 所以我们可以将有向无环图 G 表示为 $G = \{\Pi_{A_1}, \Pi_{A_2}, \dots, \Pi_{A_n}\}$ 。按父结点个数从小到大排列, 可得 $G = \{\Pi_{A'_1}, \Pi_{A'_2}, \dots, \Pi_{A'_n}\}$, 其中 $|\Pi_{A'_i}| \leq |\Pi_{A'_{i+1}}|$ 。由 G 是一个无环图可知 $|\Pi_{A'_i}| \leq i-1$ 。设 $G_i \in \{G_1, G_2, \dots, G_{n!}\}$ 且其排列 $\pi = (l_1, l_2, \dots, l_n) = (A'_1, A'_2, \dots, A'_n)$, 则 $|\Pi_{l_i}| = i-1$ 。因此有 $|\Pi_{A'_i}| \leq |\Pi_{l_i}|$, $i = 1, \dots, n$, 由图论可知, G 为 G_i 的子图。

由定理 1 和定理 2 可知, $EGNA_{ec}$ 和 $EGNA_{BGe}$ 所采用的 Gauss 网络结构是 $G_1, G_2, \dots, G_{n!}$ 中某个图的子图, 因而所获得的联合概率密度函数是 $P(\mathbf{x})$ 的一个近似, 而由一般结构 $G_1, G_2, \dots, G_{n!}$ 获得的联合概率密度函数就等于 $P(\mathbf{x})$ 。

2.2 算法描述

式(3)中的概率密度 $P(x)$ 可以写成每一项独立且服从正态分布的条件概率密度的乘积:

$$P(x) = \prod_{i=1}^n P(x_i | \Pi_i) = \prod_{i=1}^n \mathcal{N}(\mu_i + \sum_{j=1}^{i-1} a_{ji}(x_j - \mu_j), \sigma_i) \quad (5)$$

这里 μ_i 是变量 x_i 的无条件均值, σ_i 为给定变量 x_1, \dots, x_{i-1} 条件下的方差, 而 a_{ji} 为变量 x_j 与 x_i 间关系强度的一个线性系数^[10].

(μ, a, σ) 即为这个 Gauss 网络参数。由多元正态分布理论可知, 参数 a 和 σ 可根据条件独立关系和协方差矩阵 Σ 来求出。这样 Gauss 网络的参数变成 (μ, Σ) , 而 (μ, Σ) 可通过极大似然法容易求出。已知 Gauss 网络的结构和参数, 通过式 (5) 可以从一般结构 Gauss 网络中产生样本。实际应用中可以发现, 通过协方差矩阵 Σ 求出参数 a 和 σ 需要很多复杂的矩阵运算, 大大降低了运算效率。由定理 3 可知, 我们可用另一种方式从这个网络中产生样本。

定理 3 设 y_1, y_2, \dots, y_n 为独立标准正态变量, 服从 n 维正态分布 $\mathcal{N}(\mu, \Sigma)$ 的随机变量 x 可以表示如下:

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = A \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} + \mu \quad (6)$$

其中 $\Sigma = AA^T$ 。

定理 3 可由多元正态分布的定义^[11]直接推出。

从式(6)可知, 只要产生 n 个互不相关的服从标准正态分布的随机数 y_1, y_2, \dots, y_n , 通过矩阵 A 的线性组合作用, 便可直接产生满足 n 维正态分布 $\mathcal{N}(\mu, \Sigma)$ 的随机数。这时只需估计出参数 Σ 而无需进一步计算参数 a 和 σ 。算法 1 给出了生成一个服从 n 维正态分布随机数的过程。

算法 1 服从 n 维正态分布 $\mathcal{N}(\mu, \Sigma)$ 的随机数产生算法。

步骤 1 用 Cholesky 分解方法求出 A 使 $\Sigma = AA^T$;

步骤 2 由 $\mathcal{N}(0,1)$ 生成 y_1, y_2, \dots, y_n ;

步骤 3 $x_k = \mu_k + \sum_{i=1}^k A_{ki}y_i, (k=1,2,\dots,n)$ 。

于是, 我们得到了一个建立在一般结构 Gauss 网络上的分布估计算法, 见算法 2。

算法 2 建立在一般结构 Gauss 网络上的分布估计算法 (N 为种群规模, γ 为选择比例)

步骤 1 随机产生 N 个个体形成初始种群 $Pop^0, k \leftarrow 0$;

步骤 2 如果终止条件满足, 则输出结果, 停止; 否则

转步骤 3;

步骤 3 从 Pop^k 中选出最优的 $(\gamma \cdot N)$ 个个体, 根据这些个体建立一般结构的 Gauss 网络, 获得网络参数 (μ, Σ) ;

步骤 4 按算法 1 产生 $(N-1)$ 个个体并计算其适应度, 这些个体与当前的最优个体一起形成 $Pop^{k+1}, k \leftarrow k+1$, 转步骤 2。

3 仿真实验

我们首先用一个求非线性方程组实根的问题来验证算法的性能, 参数设置如下: 种群规模 $N=1000$, 选择比例 $\gamma=0.2$, 运行代数 50。三元非线性方程组为

$$\begin{cases} g_1(x) = x_1^3 + e^{x_1} + 2x_2 + x_3 + 1 = 0 \\ g_2(x) = -x_1 + x_2 + x_2^3 + 2e^{x_2} - 3 = 0 \\ g_3(x) = -2x_2 + x_3 + e^{x_3} + 1 = 0 \end{cases} \quad (7)$$

求其在 $\Omega = \{x \in \mathcal{R}^3 | -5 \leq x_i \leq 5, i=1,2,3\}$ 中的实根。对于这一类在科学实践和工程技术中大量出现的问题, 一般采用把非线性方程组转化为最优化问题来求解, 式(7)可以转化为

$$\min f(x) = |g_1(x)| + |g_2(x)| + |g_3(x)|, \quad \text{s.t. } x \in \Omega \quad (8)$$

图 2 给出了这个非线性方程组的根随代数 k 的进化轨迹。从图 2 可以看出, GN-EDA 只需要大约 10 代便找到了方程组的根 $(-0.767760, 0.075331, -1.162152)$, 并且在对该算法进行的 100 次独立试验中, 每次都得到相同的结果。

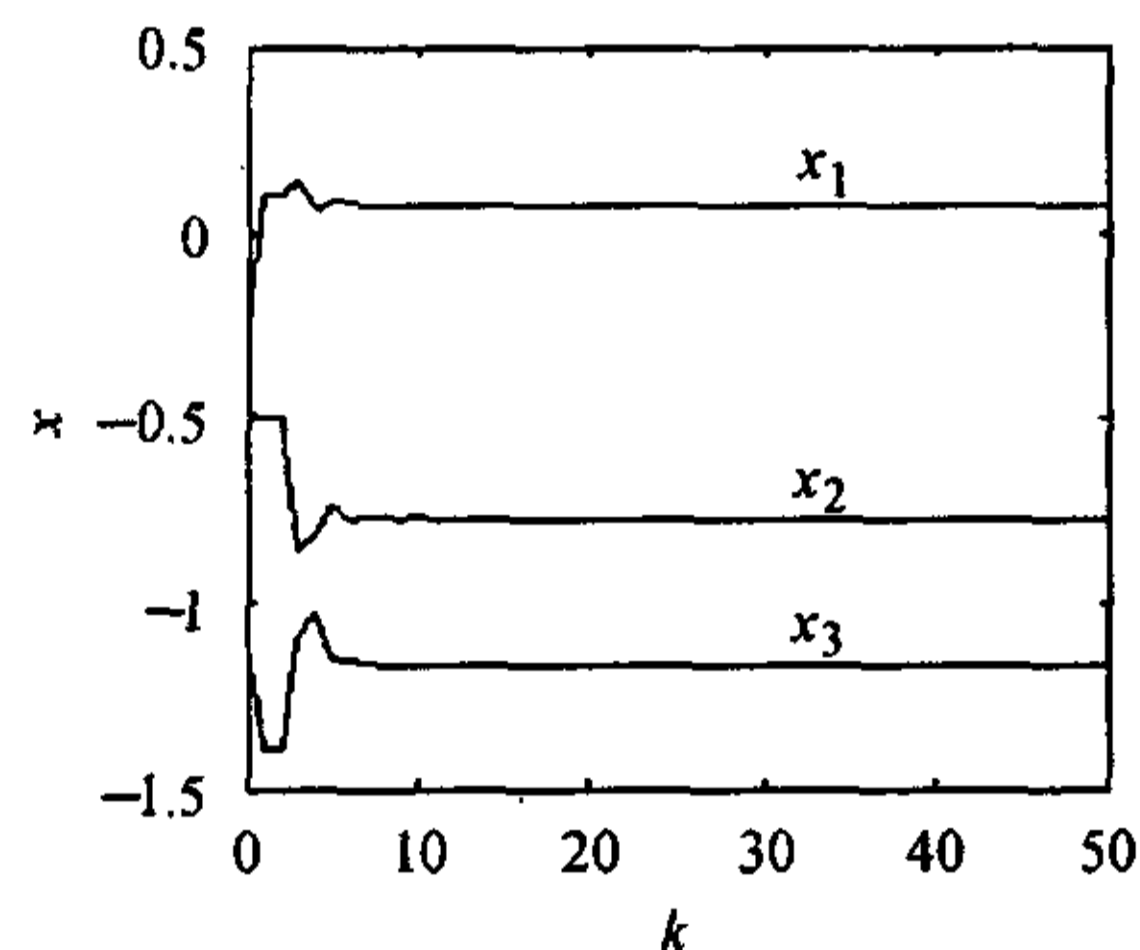


图 2 方程组根随代数 k 的进化轨迹

为了进一步验证 GN-EDA 的性能, 本文对 3 个标准测试函数 f_1, f_2, f_3 进行实验, 并与 $EGNA_{cc}$, $EGNA_{BGe}$ 的结果进行比较, 其中 $EGNA_{cc}$, $EGNA_{BGe}$ 的结果来自文献[6]。参数设置如下: $n=10, N=1000, \gamma=0.2$, 最大运行代数为 300, 精度 $\epsilon=10^{-5}$ 。停止准则为: 如果 $f_{min} \neq 0$, 则考察 $|f_{best} - f_{min}| < \epsilon |f_{min}|$ 是否成立, 否则考察 $|f_{best}| < \epsilon$, 此处 f_{min} 和 f_{best} 分别表示全局最优解和到当前代为止找到的最优解。函数 f_1, f_2, f_3 的全局最小值分别为 $-10^5, 0$ 和 0 。

$$f_1(x) = -1 / \left(10^{-5} + \sum_{i=1}^n |y_i| \right); \quad y_1 = x_1, \quad y_i = x_i + y_{i-1}, \quad i=2, \dots, n; \quad -0.16 \leq x_i \leq 0.16$$

$$f_2(x) = \sum_{i=1}^n \left((x_i - x_i^2)^2 + (x_i - 1)^2 \right); \quad -10 \leq x_i \leq 10$$

$$f_3(x) = \frac{1}{4000n} \sum_{i=1}^n x_i^2 - \prod_{i=1}^n \cos\left(\frac{x_i}{\sqrt{i}}\right) + 1; \quad -600 \leq x_i \leq 600$$

表 1 为 3 个算法对每个函数进行 100 次独立试验的结果。可以看出, 对于函数 f_1 , GN-EDA 与 $EGNA_{cc}$, $EGNA_{BGe}$

的所有试验都找到了全局最优值,但 GN-EDA 的平均评价次数只是 EGNA_{ee}, EGNA_{BGe} 的 1/3。对于函数 f_2 , GN-EDA 100% 地找到了全局最优值,而 EGNA_{ee}, EGNA_{BGe} 都没找到,并且 GN-EDA 的计算量大约只是其他两个算法的 1/2。同时 GN-EDA 的最优值平均达到了 10^{-6} , 其精度远远高于其它两个算法。对于函数 f_3 , GN-EDA 在 100 次试验中都找到了全局最优值,而 EGNA_{ee}, EGNA_{BGe} 分别只找到了 66 次和 57 次,同时 GN-EDA 的计算量只是 EGNA_{ee}, EGNA_{BGe} 的 1/4。另外, GN-EDA 的最优值平均所达到的精度也远远高于其它两个算法。由此可见, GN-EDA 的性能远远优于 EGNA_{ee}, EGNA_{BGe}。另外, GN-EDA 由于无需计算条件概率密度函数,所需的时间很少,在 PIII 667 的 PC 机求解函数 f_1, f_2, f_3 所需的时间分别为 1.17s, 0.93s, 0.82s。

表 1 GN-EDA 与 EGNA_{ee}, EGNA_{BGe} 的性能比较

		GN-EDA	EGNA _{ee}	EGNA _{BGe}
f_1	找到全局最优解的次数	100	100	100
	平均函数评价次数	84,916	194,980	188,868
	最优值平均	-10^5	-10^5	-10^5
f_2	找到全局最优解的次数	100	0	0
	平均函数评价次数	72,379	128,256	126,357
	最优值平均	6.69×10^{-6}	0.09914	0.0250
f_3	找到全局最优解的次数	100	66	57
	平均函数评价次数	52,968	223,820	238,728
	最优值平均	6.53×10^{-6}	0.008175	0.012605

4 总结

采用非一般结构 Gauss 网络的高阶分布估计算法的计算量较大且估计出来的联合概率密度函数准确度不高,针对这些问题,本文提出了建立在一般结构 Gauss 网络上分布估计算法。由于算法采用一般结构的 Gauss 网络,因此无需网络结构的学习,从而省去了通过数据学习 Gauss 网络模型中最费计算量的部分。另一方面,由于一般结构的 Gauss 网络不是近似网络,因而可获得精度很高的联合概率密度函数。同时一般结构的 Gauss 网络可直接从 $\mathcal{N}(\mu, \Sigma)$ 产生所需的样本,而无需进一步计算参数 μ 和 Σ 。仿真实验结果表明本文算法的性能稳定,且具有很强的全局寻优能力。由于学习 Gauss 网络模型时总是假设数据是服从 Gauss 分布的,因此研究能够处理非 Gauss 分布的数据的分布估计算法将是我们下一步的工作。

参考文献

- [1] Baluja S. Population-based incremental learning: A method for integrating genetic searching based function optimization and competing learning. Carnegie Mellon University, CMU Tech. Rep. CMU-CS-94-163: 1994.
- [2] Muhlenbein H. The equation for response to selection and its use for prediction. *Evolutionary Computation*, 1998, 5(3): 303 – 346.
- [3] De Bonet J S, Isbell C L, Viola P. MIMIC: Finding optima by estimating probability density. *Advances in Neural Information Processing System*, Cambridge: The MIT Press, 1997, Vol.9: 424 – 431.
- [4] Pelikan M, Muhlenbein H. The bivariate marginal distribution algorithm, in R. Roy, T. Furnhashi, P. K. Chandhery. eds., *Advance in Soft Computing Engineering Design and Manufacturing*, London: Springer-Verlag, 1999: 521 – 535.
- [5] Pelikan M, Goldberg D E, Cantu-Paz E. BOA: the Bayesian optimization algorithm. *Proceeding of the Genetic and Evolutionary Computation Conference*, Orlando: Morgan Kaufmann, 1999: 525 – 532.
- [6] Larranga P, Etxeberria R, Lozano A, J. Pefia M. Optimization by learning and simulation of Bayesian and Gaussian networks. *Proceedings of the 2000 Genetic and Evolutionary Computation Conference Workshop Program*, Las Vegas : Morgan Kaufmann, 2000: 201 – 204.
- [7] Zhang Q F, Muhlenbein H. On convergence of a class of optimization algorithms using estimation of distribution, 2000-11. Available at <http://privatewww.essex.ac.uk/~qzhang>.
- [8] 林亚平. 概率分析进化算法及其研究进展. *计算机研究与发展*, 2001, 38(1): 43 – 49.
- [9] 钟伟才, 刘静, 刘芳, 焦李成. 基于 Kalman 滤波的 Helmholtz 机进化算法. *电子与信息学报*, 2004, 26(8): 1190 – 1195.
- [10] DeGroot M. *Optimal Statistical Decision*. New York: McGraw-Hill, 1970: 155 – 189.
- [11] 张尧庭, 方开泰著. *多元统计分析引论*. 北京: 科学出版社, 1999: 65 – 118.

钟伟才: 男, 1977 年生, 博士生, 研究方向为进化计算、数据挖掘、多媒体应用、模式识别。

刘静: 女, 1977 年生, 博士生, 研究方向为进化计算、数据挖掘、视频压缩、机器学习。

刘芳: 女, 1963 年生, 教授, 研究方向为人工智能、模式识别、进化计算。

焦李成: 男, 1959 年生, 教授, 博士生导师, 研究方向为进化计算、神经网络、子波理论。