第 37 卷 第 1 期 电 子 科 技 大 学 学 报 Vol.37 No.1
2008 年 1 月 Journal of University of Electronic Science and Technology of China Jan. 2008

在子集条件约束下的最大熵分布及其应用

姜 群[1]，李祖枢[1]，欧 阳[1]，董世都[1]，邱小平[2]

(1. 重庆工学院计算机科学与工程学院 重庆 九龙坡区 400050; 2. 重庆工学院计算机中心 重庆 九龙坡区 400050)

【摘要】介绍了有关熵的概念及计算方法，并将其应用于构建一类新的分布估计算法(EDAs)。该类分布估计算法用基于最大熵估计种群中的模式概率分布和从最大熵分布中抽样取代遗传算法(GA)的交叉和变异，产生新的种群。在该类算法中，二阶连接模式算法由于只使用了连接模式，在解决变量之间相互作用趋向于发生在串中相互靠近的变量之间的一类问题时，比遗传算法更好。

关 键 词 熵；抽样；模式；模式族

中图分类号 T P391 文献标识码 A

Maximum Entropy Distribution Subject to Subset Constraints with Application

JIANG Qun[1]，LI Zu-shu[1]，OU Yang[1]，DONG Shi-du[1]，QIU Xiao-ping[2]

(1. College of Computer Science & Engineering, Chongqing Institute of Technology Jiulongpo Chongqing 400050;
2. Computer Center, Chongqing Institute of Technology Jiulongpo Chongqing 400050)

Abstract After introducing some concepts and computations of entropy, a new type of estimation of distribution algorithms (EDAs) is developed by using principle of maximum entropy. This type of algorithms replaces the crossover and mutation operators used by genetic algorithm (GA) with the estimation of the maximum entropy distribution of schema in the population and sampling from maximum entropy distribution to generate new population. Among this type of algorithms, only contiguous schemata are used in order-2 contiguous schemata algorithm. Therefore, order-2 contiguous schemata algorithm may work better than GA when interactions between variables tend to be between variables that are located close to each other on the string.

Key words entropy; sampling; schema; schema family

熵的概念起源于物理学，并广泛用于估计数据的概率分布，它的主要特点是与数据关联的概率分布应该反应已知的数据特征，并可表示“最大程度的不定性”。在新的数据变为已知时，允许最大可能的“自由度”来校准分布，从而得到更有效的分布估计。

GA通过选择、交叉与变异来实现进化过程。选择只增加高适应度解的频率，不会在种群中增加新的解，而交叉和变异从当前种群中产生新的解来扩大从搜寻空间中抽样。与GA不同，分布估计算法不使用交叉和变异[1]，而是根据当前种群中较好的个体估计概率分布，用概率分布引导对搜寻空间的探索。由于EDAs较GA有更强的理论基础，已成为当前进化计算的研究热点[2-4]。

本文将最大熵原理应用于构建分布估计算法，其主要思想是：用已知种群中模式的频率去约束所估计的概率分布，使分布的熵随之达到最大，再从该分布中取样产生新的种群。初步实验结果表明，本算法在解决某些复杂问题时，比遗传算法有明显的优势。

1 有关熵分布及子集条件约束

如果 X 是一个有限集， $p(x)$ 是 X 上的概率分布，那么 $p(x)$ 的熵定义为[5]：

$$S(X)=\sum_{x\in X}-p(x)\log_2 p(x)$$

在 $p(x)$ 上的一个模式约束 (Q, a) 是由集合 $Q\subset X$ 与一个实数 $a\in[0,1]$组成，使 $\sum_{x\in Q}p(x)=a$ 。模式约束族由一对字母 (T,g) 表示，T 表示 X 上的模式集；g 是一个从 T 到单位区间[0,1]的函数。每

收稿日期：2007 – 01 – 30；修回日期：2007 – 06 – 17
基金项目：重庆市自然科学基金(CSTC2006BB2397)
作者简介：姜 群(1959 –)，女，硕士，副教授，主要从事进化计算方面的研究.

个模式约束用 $(Q, g(Q))$ 对一些 $Q \in T$ 的形式表示。如果有概率分布满足所有模式约束,则称该族 (T, g) 是可行的。如果对每一个 $Q \in T$，$g(Q)$ 定义为 Q 在一个已知种群中的频率，那么 T 是可行的。

引理 1　如果 $p(x)$ 是 X 上在模式约束下的最大熵分布，元素 x_1、x_2 相对于这些约束是不可区分的(它们在同一个模式中)，则 $p(x_1) = p(x_2)$。

证明　设元素 x_1，$x_2 \in A$，使得 $p(x_1) \neq p(x_2)$。定义概率分布 q 使得 $q(x_1) = q(x_2) = (p(x_1) + p(x_2))/2$ 和 $q(x) = p(x)$ 对所有 $x \notin \{x_1, x_2\}$。显然，q 满足所有模式约束。由于 $-p(x)\log_2 p(x)$ 是 $p(x)$ 的凹函数，$q(x)$ 的熵大于 $p(x)$ 的熵，与 $p(x)$ 是最大熵分布矛盾。

模式族是具有共同被定义位置的模式集合，可用长度为 L 定义在元素集{D, $*$}上的串来表示，D 表示被定义位置(该位置为0或1)；$*$表示变量位置。

设搜寻空间 $\Omega = A_i = A_1 \times A_2 \times \cdots \times A_L$，其中 A_i 表示有限字符集中的字符在位置 i (每个在位置 i 的字符取自某个有限字符集)；L 为串的长度。如搜寻空间是串长为 L 的二元串集，则 $\Omega = \{0,1\} \times \{0,1\} \times \cdots \times \{0,1\}$。当把 Ω 写成笛卡儿乘积时，如 $\Omega = X \times Y$，每一个因子对应一些 A_i 的积，即对应于一个字符串位置集合，而这些集合构成字符串位置集合的一个划分。如果 $\Omega = X \times Y$，Π_X 和 Π_Y 分别表示 Ω 在 X 和 Y 上的投影，假设 Q 是任意一个模式，则 $\Pi_X(Q)$ 表示把 Q 投影到 X 上，去掉所有对应于 Y 的字符串位置。如 X 对应于前三个字符串位置，Y 对应于后三个字符串位置，那么 Π_X (1$*$0$*$01)=1$*$0。如果所有对应于 Y 的位置是变量$*$，则一个模式是 Y 的变量。如果模式 Q 是 Y 的变量，则 $Q = \Pi_X(Q) \times Y$。

假设 $\Omega = X \times Y$，而 $p(x, y)$ 是 Ω 上在模式约束族 (T, g) 下的最大熵分布，这里任意 $Q \in T$ 是 Y 的变量。由于 $\Pi_X(Q) \times Y = Q$，定义 $g(\Pi_X(Q)) = g(\Pi_X(Q) \times Y)$，$g$ 自然地被扩大到集 $\Pi_X(Q)$中。从引理1可知对于任意的 $x \in X$，y_1 和 $y_2 \in Y$ 有：

$$p(x, y_1) = p(x, y_2)$$

因此，在约束 $(\Pi_X(Q), g)$ 下，X 上的最大熵分布也是 $p(x, y)$ 的边值分布 $p(x)$，该分布称为 X 上的在 $\Pi_X(T, g)$约束下的最大熵分布。

在 $X \times Y$ 上的概率分布 $p(x, y)$ 的条件熵为[6]：

$$S(X \mid Y) = \sum_Y p(y) S(X \mid y) = -\sum_Y \sum_X p(x, y) \log_2 p(x \mid y)$$

引理 2[6]　$S(X/Y) \leqslant S(X)$，当且仅当 X 和 Y 相互独立即 $p(x, y) = p(x)\ p(y)$ 时等号成立。

引理 3[6-7]　$S(X,Y) = S(X/Y) + S(Y) \leqslant S(X) + S(Y)$，当且仅当 X 与 Y 概率独立时等号成立。

定理 1　(用于非重叠模式) 设($T_X \cup T_Y, g$)是一个 $X \times Y$ 上的模式约束集，使得每一个 $Q \in T_X$ 是 Y 上的变量，而每一个 $Q \in T_Y$ 是 X 上的变量。那么在 $X \times Y$ 上的最大熵分布 $p(x, y)$ 等于 X 上在约束 $\Pi_X(T_X, g)$下的最大熵分布 $p(x)$ 与 Y 上在约束 $\Pi_Y(T_Y, g)$ 下的最大熵分布 $p(y)$ 之积 $p(x, y) = p(x)\ p(y)$。

证明　本文只证明对应于乘积分解 $X \times Y$ 的情形，用归纳法可以导出一般情况。由引理3当且仅当 $p(x, y) = p(x)\ p(y)$ 时等号成立，可得如果定义 $p(x, y)$ 为 $p(x)\ p(y)$，那么 $p(x, y)$ 满足所有的模式约束。

设 $Q \in T_X$，那么：

$$\sum_{(x,y)\in Q} p(x, y) = \sum_{(x,y)\in \Pi_X^{-1}(\Pi_X(Q))} p(x) p(y) = \sum_{x\in \Pi_X(Q)} p(x) \sum_{y\in Y} p(y) = g(\Pi_X(Q)) \cdot 1 = g(Q)$$

对 $Q \in T_Y$ 的情形，同理可证。

引理 4　$S(X, Z \mid Y) = S(X \mid Y, Z) + S(Z \mid Y)$

证明　

$$\begin{aligned} S(X \mid Y, Z) &= -\sum_X \sum_Y \sum_Z p(x, y, z) \log_2 p(x \mid y, z) = \\ &-\sum_Y p(y) \sum_X \sum_Z p(x, z \mid y) \log_2 p(x \mid y, z) = \\ &-\sum_Y p(y) \sum_X \sum_Z p(x, z \mid y) \log_2 \frac{p(x, z \mid y)}{p(z \mid y)} = \\ &-\sum_Y p(y) \sum_X \sum_Z p(x, z \mid y)(\log_2 p(x, z \mid y) - \log_2 p(z \mid y)) = \\ &-\sum_Y p(y) \sum_X \sum_Z p(x, z \mid y) \log_2 p(x, z \mid y) \\ &+\sum_Y p(y) \sum_X \sum_Z p(x, z \mid y) \log_2 p(z \mid y) = \\ &-\sum_Y p(y) \sum_X \sum_Z p(x, z \mid y) \log_2 p(x, z \mid y) \\ &+\sum_Z \sum_Y p(y, z) \log_2 p(z \mid y) = \\ &S(X, Z \mid Y) - S(Z \mid Y) \end{aligned}$$

引理 5　$S(X \mid Y, Z) \leqslant S(X \mid Y)$，当且仅当 $p(x, z \mid y) = p(x \mid y) p(z \mid y)$ 等式成立，或者等价于 $p(x, z \mid y) = p(x) p(z) / p(y)$。

该引理是文献[6]第二章引理6的特例。

定理 2　(用于重叠的模式) 设($T_{X,Y} \cup T_{Z,Y} \cup$

T_Y , g)是一个在 $X\times Y\times Z$ 上的模式约束集，使得 $T_{X,Y}$ 是 Z 上的变量；$T_{Z,Y}$ 是 X 上的变量；T_Y 是 $X\times Z$ 上的变量。又设 $p(y)$ 是 Y 上在约束 Π_Y (T_Y , g)下的最大熵分布， $p(x,y)$ 是 $X\times Y$ 上在约束 $\Pi_{X,Y}$ ($T_{X,Y}$, g)下的最大熵分布， $p(y,z)$ 是 $Y\times Z$ 上在约束 $\Pi_{Y,Z}$ $(T_{Y,Z},g)$ 下的最大熵分布。假设 $p(y)$ 是 $p(x,y)$ 和 $p(y,z)$ 的边值分布，那么在 $X\times Y\times Z$ 上的最大熵分布由 $p(x,y,z)=\dfrac{p(x,y)p(y,z)}{p(y)}$ 给出。

证明　由引理4和5可知，如果 $p(x,y,z)$ 定义为 $p(x,y,z)=p(x,y)\ p(y,z)/p(y)$ ，那么它满足所有的约束。首先假设 $Q\in T_{X,Y}$，因 $Q=\Pi_{X\times Y}(Q)\times Z$ ，则：

$$\sum_{(x,y,z)\in Q}\frac{p(x,y)p(y,z)}{p(y)}=$$

$$\sum_{(x,y)\in \Pi_{X\times Y}(Q)}\frac{p(x,y)}{p(y)}\sum_{z\in Z}p(z,y)=$$

$$\sum_{(x,y)\in \Pi_{X\times Y}(Q)}\frac{p(x,y)}{p(y)}p(y)=$$

$$\sum_{(x,y)\in \Pi_{X\times Y}(Q)}p(x,y)=g(Q)$$

式中　$p(y)$是$p(y,z)$的边值分布。

对$Q\in T_{Z,Y}$ 和$Q\in T_Y$ 的情形，同理可证。

2　新的分布估计算法

2.1　一阶模式算法

一阶模式算法描述为：(1) 初始化种群；(2) 产生所有一阶模式集合；(3) 根据个体适应度进行选择；(4) 在模式集中计算所有一阶模式的频率；(5) 用定理1从最大熵分布中抽样产生新的种群；(6) 如果终止条件不满足，返回(3)。

不难看出，一阶模式算法是一个均值多变量分布算法[8-9]。均值多变量分布算法是一种简单的分布估计算法，对应于每一个变量的频率被相对独立地保持着。适用于解决变量之间没有明显相互作用的一类问题，而不适于解决变量之间相互作用较大的评价函数问题[10]。

2.2　二阶连接模式算法

如果一个模式的被确定位置是连接的，则该模式是连接的。在一个模式集中，如被确定的位置是连接的，没有变量位置介于其间，则该模式集是连接的，如模式集{00∗∗,01∗∗,10∗∗,11∗∗,∗00∗,∗01∗,∗10∗,∗11∗,∗∗00, ∗∗01, ∗∗10, ∗∗11}是连接的。如果两个模式有相同的被确定位置，则它们是兼容的，如∗0∗10与10∗1∗是兼容的，而∗0∗01与11∗1∗是不兼容的。

二阶连接模式算法描述为：(1) 初始化种群；(2) 产生所有二阶连接的模式集合；(3) 根据个体的适应度进行选择；(4) 在模式集合中计算模式的频率；(5) 用定理2从最大熵分布中抽样产生新的种群；(6) 如果终止条件不满足，返回(3)。

为了产生一个新的个体，可以从全部位置都是∗的模式开始。然后用定理2逐个地定义更多的位置，直到所有的位置都被定义，再没有∗剩下。当所有位置被定义后，可把对应的个体放入到新的种群中。用模式的频率作为概率，从 $DD*\cdots*=$ {00∗⋯∗,01∗⋯∗,10∗⋯∗,11∗⋯∗}中随机地选取一个模式来定义开始的前两个位置。确定第三个位置时采用定理2和模式族 $*DD*\cdots*$ 。同理，确定第四个位置采用定理2和模式族 $**DD*\cdots*$ 。如此下去，需要 $L-1$ 步完成。如当 $L=4$ 时，新个体的最初形式为模式∗∗∗∗，然后根据概率 $g(00**)$、$g(01**)$、$g(10**)$、$g(11**)$ 的值从模式族 $DD**$ 中选取一个模式。假设01∗∗被选取，新个体的当前形式为01∗∗。然后从模式族 $*DD*$ 中选取一个模式，且该模式必须与01∗∗兼容，所以∗10∗或∗11∗被选。用概率值 $g(*10*)/g(*1**)$ 来决定取舍∗10∗；用概率值 $g(*11*)/g(*1**)$ 来决定取舍∗11∗。如果∗10∗被选用，则新的个体具有形式010∗。用同样的方法可定义最后一个∗，由此便产生一个新个体。

在该算法中，由于只使用了连接的模式，所以适用于解决变量之间相互作用趋向于发生在串中相互靠近的变量之间的一类问题。

3　实验结果

以*NK*-fitness函数对两种算法作性能比较，如图1所示。本文用Java编程实现了二阶连接模式算法。个体用Java BitSets实现，模式被表示为一对BitSets，然后用集合运算判定一个个体是否属于一个模式。适应度函数选用在进化计算中常用作测试函数的*NK*-fitness函数。毗连型的*NK*-fitness函数是*N*元多项式，串被划分为 ℓ 个k+1阶的重叠的毗连块，而函数是定义在每个毗连块上的随机子函数之和。本文取k=3，因而每个块包含4比特。串的长度N=64，变异率为0.01，采用截断选择法，截断参数为0.2，函数评价5 000 次，图1是运行1 000 次的平均结果。结果显示，本文算法优于GA。

(下转第123页)

Washington, USA: IEEE, 2003, 19-24: 147-154.
[5] LEVENBERG J. Fast view-dependent level-of-detail rendering using cached geometry[C]//In: Proceedings of Visualization 2002. Boston USA: [s.n.], 2002.
[6] ALEX A. POMERANZ. Roam using surface triangle clusters (rustic)[D]. Davis: University of California, 2000.
[7] LINDSTROM P, PASCUCCI V. Terrain simplification simplified: A general framework for view-dependent out-of-core visualization[J]. IEEE Transaction on Visualization and Computer Graphics, 2002, 8(3): 239-254.
[8] LINDSTROM P, PASCUCCI V. Visualization of large terrains made easy[C]//In: Proceedings of Visualization 2001. San Diego California: [s.n.], 2001.
[9] HOPPE H. Smooth view-dependent level-of-detail control and its application to terrain rendering[C]//In: Proceedings of Visualization '98. Durham USA: [s.n.], 1998.
[10] DAVID CLINE, PARRIS K E. Terrain decimation through quadtree morphing[J]. IEEE Transactions on Visualization and Computer Graphics , 2001, 7(1): 62 -69.
[11] ROTTGER S, HEIDRICH W, SLUSALLECK P, et al. Real-time generation of continuous levels of detail for height fields[C]//In: Proceedings of WSCG 98. Czech Republic: University of West Bohemia, 1998.

编　辑　熊思亮

(上接第96页)

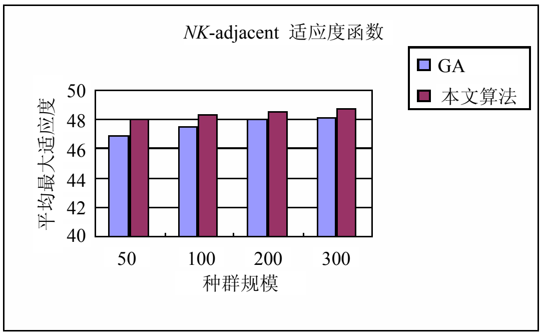

图1　以*NK*-fitness函数对两种算法作性能比较

4 总　结

在介绍了熵的有关理论后给出了用最大熵原理构建的分布估计算法。*NK*-fitness函数初步实验证明了基于最大熵的二阶连接模式分布估计算法在解决某些复杂问题时优于GA，高阶的模式算法将是未来的研究工作。

参 考 文 献

[1] 钟伟才, 刘　静, 刘　芳, 等. 建立在一般结构Gauss网络上的分布估计算法[J]. 电子与信息学报, 2005, 27(3): 467.
[2] MUHLENBEIN H. The equation for response to selection and its use for prediction[J]. Evolutionary Computation, 1998, 5(3): 303-346.
[3] DE BONET J S, ISBELL C L. Finding optima by estimating probability density. Advances in Neural Information Processing System[M]. Cambridge: The MIT press, 1997, (9): 424-431.
[4] 林亚平. 概率分析进化算法及其研究进展[J]. 计算机研究与发展, 2001, 38(1): 43-49.
[5] 张　渝, 周宗放. 商业银行信用风险评价指标的熵权选择方法[J]. 电子科技大学学报, 2006, 35(5): 857.
[6]FEINSTEIN AMEIL. Foundations of information theory[M]. York, PA 17405, USA: The Maple Press Company, 1959.
[7] GUIASU S, SHENITZER A. The principle of maximum entropy[J]. The Mathematical Intelligencer, 1985, (7): 42-48.
[8] HEINZ M U, MAHNIG T. Convergence theory and application of the factorized distribution algorithm[J]. Journal of Computing and Information Technology, 1999, 7(1): 19-32.
[9] HEINZ M U, MAHNIG T. FDA—a scalable evolutionary algorithm for the optimization of additively decomposed functions[J]. Evolutionary Computation, 1999, 7(4): 353-376.
[10] HEINZ M U. The equation for the response to selection and its use for prediction[J]. Evolutionary Computation, 1997, 5(3): 303-46.

编　辑　熊思亮