

文章编号:1001-9081(2009)06-1539-05

一种新颖的基因调控网络结构学习方法

杜智华, 王宜伟

(深圳大学 德州仪器 DSPs 实验室, 广东 深圳 518060)

(duzh@szu.edu.cn)

摘要: 基于贝叶斯模型的结构学习是 NP 完全问题。给出了一种新颖的基于粒子群优化算法 (PSO) 的无序条件下的结构学习方法。迭代中, 新算法用粒子表示网络结构, 根据 PSO 群体规则进行更新, 同时利用互信息进行修正, 将所得到的最佳网络结构作为 K2 算法的先验知识。实验结果表明, 新算法可以在无序的条件下, 取得稳定的贝叶斯分数值。同时, 在与 K2 算法和 BN-PSO 算法的比较中, 新算法取得了更优异的网络结构和更快的收敛速度。

关键词: 基因调控网络; 贝叶斯网络; 粒子群优化算法; K2 算法

中图分类号: TP18 文献标志码: A

Novel structure learning method for constructing gene regulatory network

DU Zhi-hua, WANG Yi-wei

(Texas Instruments DSPs Laboratory, Shenzhen University, Shenzhen Guangdong 518060, China)

Abstract: Structure learning of Bayesian network is a NP hard problem. A structure learning method without an ordering requirement based on novel Particle Swarm Optimization (PSO) was proposed. In this method, the structures were represented by particles, and updated according to swarm intelligence of PSO. Then the net structures were modified with mutual information. Finally the K2 algorithm was applied with the global best particle as the ordering and updating the best particle for next generation. Experimental results demonstrate that the performance of this new strategy is better than that of K2 and BN-PSO, especially it has higher convergence rate and better Bayesian score without a previous ordering.

Key words: Gene regulatory network; Bayesian network; Particle Swarm Optimization (PSO); K2 algorithm

0 引言

基因芯片是近十年来在分子生物学领域迅速发展起来的一项高新技术, 并成为探索生命奥秘的重要工具之一。基因芯片可快速获得海量的生物信息数据。如何分析此类数据已成为该技术应用的瓶颈, 也成为后基因组时代研究的重点所在。

基因调控网络描述一组调控因子, 如 DNA、RNA 以及蛋白质等, 如何调控基因表达的过程^[1]。其从分子层次对生物系统进行研究, 并揭示基因组的功能和行为, 有助于从基因组层次对生物过程进行解释。因此基因调控网络的研究有着重要意义, 同时基因芯片的发展又极大地推动了研究进展。在国外, 迄今为止, 基因调控网络研究的模型主要有布尔模型^[2]、线性组合模型^[3]、贝叶斯模型^[4-5]、微分方程^[6]以及信息熵关联网络模型^[7]等。贝叶斯模型可以更好地处理原始数据中的噪声和缺失, 同时又易于融入先验知识, 所以贝叶斯模型更适合基因调控网络的学习。

贝叶斯网络模型的学习分为结构学习与参数学习两个方面。结构学习是指在先验知识的指导下, 利用训练样本集确定合适的网络拓扑结构。参数学习是指在网络结构确定的情况下学习每个节点的条件概率表。结构学习是一个 NP 完全问题^[8], 所以对结构学习的研究更加困难同时也更加具有意义。结构学习的方法分为基于评分的方法和基于条件独立性的方法。基于评分的方法将贝叶斯网络看成是含有属性之间联合概率分布的结构, 给出评价网络结构的评分函数, 搜索出

与数据拟合最好的结构, 即得分函数最好的结构。常用的基于评分的结构学习方法有 K2 算法^[9]。K2 算法学习速度快, 但是要求已知网络中节点的次序, 对先验知识的依赖性很大。近年来许多启发式算法被应用到结构学习中^[5,10-11], 并取得良好的性能。粒子群算法作为一种新兴的启发式算法^[12], 也被应用在结构学习当中, 并取得优于遗传算法 (Genetic Algorithm, GA) 的性能^[11,13]。

粒子群优化算法 (Particle Swarm Optimization, PSO)^[12]是一种基于群智能的全局优化进化算法, 它模拟了自然界中鸟类捕食行为。粒子由速度来决定飞行的方向和距离, 由优化函数确定的适应值来决定位置的优劣。粒子根据自身所搜索到的最好位置和当前最优粒子的位置而调整自身速度, 从而在解空间中向全局最优解靠近。PSO 的优势在于参数设置简单同时又容易实现, 因而被广泛地应用于各个领域。

本文提出了一种新的贝叶斯网络结构学习算法, 利用改进的粒子群算法进行结构搜索。新算法中, 每个网络结构表示为一个粒子, 在无序的条件下, 根据适应度函数向全局最优粒子及个体历史最优粒子学习。每次更新结束之后, 根据贝叶斯网络的限制条件对所得网络进行修正。运用 K2 算法对修正后的网络进行学习, 将学习结果与粒子群体的全局最优值进行比较, 选出较优的结构作为新的全局最优粒子, 进而指导下一步的 PSO 更新。PSO 算法的快速收敛性能和鲁棒性, 使得新算法可以在没有给定先验节点序的情况下快速稳定的得到最优网络结构。实验结果表明, 新算法无论在经典统计学网络还是真实基因网络中, 均取得了优于 BN-PSO 算法^[13]

收稿日期: 2008-12-29; 修回日期: 2009-03-04。 基金项目: 国家自然科学基金资助项目(60572100)。

作者简介: 杜智华(1977-), 女, 湖北武汉人, 讲师, 博士, 主要研究方向: 生物信息学、生物启发式计算; 王宜伟(1985-), 男, 江苏连云港人, 硕士研究生, 主要研究方向: 智能优化算法、生物信息学。

和 K2 算法的网络分数。在酵母基因调控网络中,新算法更加准确还原出符合生物学信息的基因网络。

1 结构学习

1.1 网络结构

贝叶斯网络结构可以用有向无环图 (Directed Acyclic Graph, DAG) 表示,DAG 中的节点表示网络的变量,边表示变量之间的因果关系。公式表示为: $G = \langle X, E \rangle$, G 表示有向无环图, X 表示变量, 在基因网络中可表示基因表达水平, E 表示边。令 $X = \{x_1, x_2, \dots, x_i, \dots, x_n\}$, $\forall i = 1, 2, \dots, n$, 其中 n 表示基因数量, x_i 表示基因 i 的基因表达水平。对于每一个 x_i , 有一个条件分布 $P(x_i | parents(x_i))$, 这里 $parents(x_i)$ 表示在 G 中此变量直接前驱集, 即父节点集。有向无环图 G 和条件分布 $P(x_i | parents(x_i))$ 共同构成贝叶斯网络。根据马尔可夫假设, 即对于每一个 G 中的节点, 在给定其所有父节点的情况下, 条件独立于其非子节点, 则贝叶斯网络的联合概率分布可分解为:

$$P(X) = \prod_{i=1}^n P(x_i | parents(x_i)) \quad (1)$$

给定数据集 D , 则结构学习的目的为找出最匹配数据的网络结构 G 。在基于评分的结构学习中, 即找出评分函数的最优值。最常用的评分函数为 BDe (Bayesian Dirichlet Equivalent) 分数^[9], 如式(2)所示:

$$BDe = P(D | G) = \prod_{i=1}^n \prod_{j=1}^{q_i} \frac{\Gamma(N'_{ij})}{\Gamma(N'_{ij} + N_{ij})} \prod_{k=1}^{r_i} \frac{\Gamma(N'_{ijk} + N_{ijk})}{\Gamma(N'_{ijk})} \quad (2)$$

其中: $N_{ij} = \sum_{k=1}^{r_i} N_{ijk}$, $N'_{ij} = \sum_{k=1}^{r_i} N'_{ijk}$ 。先验分布的指系数 $N'_{ijk} = \frac{S}{r_i q_i}$ (S 为等效样本容量)。 N_{ijk} 为数据集 D 中满足 $x_i = x_i^k$ (表示 x_i 取第 k 个值) 且 $parents(x_i) = j$ (表示节点集取第 j 个排列) 情况下的数目。

1.2 K2 算法

贝叶斯网络模型的结构学习是指利用训练样本集确定网络拓扑结构, 这是一个 NP 完全问题。因此对结构学习的研究更具有意义, 同时难度也更大。K2 算法预先选定一个节点顺序, 并且假定所有网络结构的先验概率相等, 一定程度上缩减了结构候选集的规模。因此被广泛应用在贝叶斯网络的结构学习中。但是 K2 算法同时存在缺点, 即牺牲精确度换取学习速度, 同时要求已知网络节点序, 对先验知识的依赖性很大。

结构学习中, 给定节点序可以大幅度缩减搜索候选集, 节

点数为 n 的候选集规模为 $2^{\binom{n}{2}}$ 。无序条件下的候选集规模可由式(3)得到。表 1 则分别列出有序、无序条件下候选集的规模与节点数的关系。由表 1 我们可以看出无序条件下网络结构候选集随节点数的增加而呈超指数增加, 搜索中的困难也越大。

$$f(n) = \sum_{i=1}^n (-1)^{i+1} \binom{n}{i} 2^{i(n-i)} f(n-i) \quad (3)$$

其中: $f(0) = 1, f(1) = 1$ 。

节点顺序要求只有该节点之后的节点才可以作为其父节

点的备选节点。K2 算法按照节点顺序依次为每个节点寻找父节点直到新增的父节点不能使评分函数的分值更优。算法始终要求在最大化父节点集网络分数的同时满足严格遵循备选节点的条件。

表 1 网络结构搜索候选集

节点数	有序	无序
5	1.024E3	2.928E4
10	3.0518E13	4.175E18
20	1.569E57	2.344E72
30	8.872E130	2.714E158
40	6.359E234	1.124E276

2 PSO 算法在结构学习中的应用

K2 算法牺牲精确度换取学习速度, 同时对先验知识的依赖性大。因此本文提出了一种在无先验知识下, 即无节点序的情况下结构学习算法。由于无序条件下结构规模巨大, 本文提出的新算法引入粒子群优化算法增加搜索效率。

2.1 PSO 算法

在一个 D 维的目标搜索空间中, 由 m 个粒子构成一个群体, 其中第 i 个粒子 ($i = 1, 2, \dots, m$) 的位置可表示为 D 维的矢量 $z_i = (z_{i1}, z_{i2}, \dots, z_{iD})$ 。每次迭代中粒子 i 移动的距离为粒子的飞行速度, 表示为 $v_i = (v_{i1}, v_{i2}, \dots, v_{id}, \dots, v_{iD})$, 粒子迄今为止搜索到的最优位置为 $p_i = (p_{i1}, p_{i2}, \dots, p_{id}, \dots, p_{iD})$, 整个粒子群迄今为止搜索到的最优位置为 $p_g = (p_{g1}, p_{g2}, \dots, p_{gd}, \dots, p_{gD})$ 。每次迭代中, 粒子根据以下公式更新速度和位置:

$$v_{id}^{k+1} = w v_{id}^k + c_1 r_1 (p_{id} - z_{id}^k) + c_2 r_2 (p_{gd} - z_{id}^k) \quad (4)$$

$$z_{id}^{k+1} = z_{id}^k + v_{id}^{k+1} \quad (5)$$

其中: $i = 1, 2, \dots, m; d = 1, 2, \dots, D; k$ 是迭代次数; w 为惯性权重, r_1, r_2 为 $[0, 1]$ 的随机数。 c_1, c_2 为学习因子, 也称加速因子, 使粒子具有自我总结和向群体中优秀个体学习的能力, 从而向自己的历史最优点以及群体内历史最优点靠近。此外速度 v_i 取值范围为 $[-v_{max}, v_{max}]$, 位置 z_i 的取值范围为 $[-z_{max}, z_{max}]$ 。

2.2 改进 PSO 算法

2.2.1 算法描述

本文用邻接矩阵表示 DAG。邻接矩阵 C 是一个 $n \times n$ 阶矩阵, 其元素值只能取 0 或 1, 矩阵元素 $C_{ij} = 1$ 表示节点 j 是节点 i 的父节点, $C_{ij} = 0$ 表示节点 j 与节点 i 无父子关系。新算法中, 粒子位置用邻接矩阵表示, 每个粒子代表一个 DAG。PSO 算法中, 粒子的速度表示粒子的飞行方向及距离, 其数学形式和粒子位置的形式保持一致。因此粒子速度同样用邻接矩阵表示, 其中的 0、1 值随机分配, 如图 1 所示。

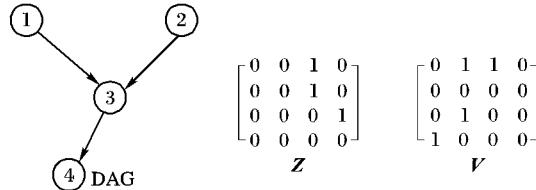


图 1 粒子结构示意图

由于粒子用邻接矩阵表示, PSO 的更新公式中的四则运算不再适用。因此, 新算法中运算符号被重载, 位置更新的运算规则如式(6)所示。

$$\begin{cases} z_2 = z_1 + v \\ Z_2^j = \begin{cases} 0, & Z_1^j = 0 \text{ 且 } V^j = 0 \\ 1, & \text{其他} \end{cases} \end{cases} \quad (6)$$

速度更新的运算较为复杂,可将其分为三步运算。第一步计算 p_i 与 z_i 的差值,设为 v_p ,第二步计算 p_g 与 z_i 的差值,设为 v_g ,第三步计算 v_i 、 v_p 、 v_g 之和。前两步运算可归结为位置的减法运算,其规则如式(7)所示。第三步为速度的加法运算,其规则如式(8)所示。

$$\begin{cases} v = z_1 - z_2 \\ V^j = \begin{cases} 1, & Z_1^j = 1 \text{ 且 } Z_2^j = 0 \\ 0, & \text{其他} \end{cases} \end{cases} \quad (7)$$

$$\begin{cases} v_3 = v_1 + v_2 \\ V_3^j = \begin{cases} 0, & V_1^j = 0 \text{ 且 } V_2^j = 0 \\ 1, & \text{其他} \end{cases} \end{cases} \quad (8)$$

速度的更新是 PSO 算法的精髓。由式(4)我们可以看出,速度更新由三部分组成。第一部分为粒子的原始速度,标志着粒子速度的更新继承了前期成果。第二部分为粒子的历史最优值与粒子当前位置的距离,标志着粒子向历史最优学习。第三部分为粒子群体的全局最优值与粒子当前位置的距离,标志着粒子的群体学习性能,向种群中最优秀的个体学习。每部分都存在系数因子,用以表示该部分对更新的影响作用大小。惯性权重权衡算法的全局最优能力和局部最优能力。在搜索的初期,通常增强群体的全局搜索能力,而搜索的后期,通常增强群体的局部搜索能力。学子因子调节向优秀粒子靠近的步长,同时加入随机数以保持群体的多样性。在新算法中,我们保留了系数因子,并将其影响用式(9)表示。

$$V_{i+1} = \begin{cases} V_i + V_p, & w \geq c_2 r_2 \text{ 且 } c_1 r_1 \geq c_2 r_2 \\ V_i + V_g, & w \geq c_1 r_1 \text{ 且 } c_2 r_2 \geq c_1 r_1 \\ V_p + V_g, & c_1 r_1 \geq w \text{ 且 } c_2 r_2 \geq w \end{cases} \quad (9)$$

2.2.2 算法流程

DAG 表示贝叶斯网络结构,通常存在两个前提条件,即“有向”和“无环”。“有向”明确指 DAG 表示贝叶斯网络结构,通常存在两个前提条件,即“有向”和“无环”。“有向”明确指出了节点直接的父子关系,“无环”则保证了网络结构的正确性。利用邻接矩阵表示 DAG 同样必须遵守这两个前提条件。邻接矩阵的 1 表示了行指向列的父子关系,满足了“有向”的要求。而对于“无环”的要求,新算法中添加了检测并纠正回路的步骤。在位置更新之后,算法检测网络中是否存在回路。如不存在回路,则继续搜索过程,如存在回路,则删除回路中互信息最小的有向边直到不再存在回路。同时在结构学习中,通常设定父节点的个数是受限的。因此在新算法中,如果某个节点的父节点数超过预先设定的最大父节点数,则删除指向其的互信息最小的有向边直到满足条件。

互信息^[14]表示两个节点之间的依赖程度,互信息的值越高表明两个节点之间依赖程度越大。其计算公式如式(10)所示。

$$\begin{aligned} I(X; Y) &= H(X) - H(X | Y) = \\ &- \sum_{i=1}^n P(x_i) \log(P(x_i)) + \\ &\sum_{i=1}^n \sum_{j=1}^m P(X = x_i, Y = y_j) \times \\ &\log(P(X = x_i | Y = y_j)) \end{aligned} \quad (10)$$

综上,新算法的流程如下。

初始化:

FOR $i = 1$ to N_p

随机赋予每个粒子的位置、速度;

修正位置使其满足无环、最大父节点条件;

根据式(2)计算适应度函数;

设定初始位置为粒子的个体最优值;

END

设定群体的全局最优值;

更新:

FOR $l = 1$ to $MaxIter$

FOR $j = 1$ to $Maxloop$

FOR $i = 1$ to N_p 根据式(6)、(7)、(8)更新速度;

根据式(5)更新位置;

修正位置使其满足最大父节点数限制;修正位置使其满足无环限制;

根据式(2)计算适应度函数;

更新粒子的个体最优值;

END

更新群体的全体最优值;

END

根据全体最优值的节点序运行 K2 算法;

根据 K2 算法结果更新群体全体最优值;

END

算法执行 10 次 PSO 算法后,利用其节点序进行一次 K2 算法学习,选出较优的结构作为全局最优进行下一次 PSO 更新。

3 实验结果及分析

3.1 实验准备

本文使用经典的 ASIA 网络^[15]作为实验网络,该网络已被广泛应用于结构学习中^[5,8,10-11,13]。它含有 8 个节点,网络结构如图 2 所示。依据概率逻辑采样,分别取得样本数为 500,1000,5000 和 10000 的数据集作为实验数据。

本文采用 BDe 分数作为适应度函数衡量网络结构的优劣,BDe 数值越大,网络结构越优异。为了测试提出的改进算法,将其与文献[13]提出的 BN-PSO 算法比较。在文献[13]中,BN-PSO 取得了优于 BN-GA^[10]的网络结构得分。

为了让改进 PSO 算法在搜索初期拥有较大的全局搜索能力,而搜索后期拥有较大的局部寻优能力,改进 PSO 算法的参数设定参考文献[16]。其中取 $c_1 = c_2 = 1.3$, w 依照循环次数从 1.8 线性下降到 0.2, $N_p = 10$, $Maxloop = 10$, $MaxIter = 10$ 。根据 ASIA 网络的实际结构,各算法中网络最大父节点数均设置为 2。BN-PSO 的循环次数设置为 100,粒子个数同样设置为 10,其他参数设置参照其参考文献。

3.2 实验结果及分析

分别使用 BN-PSO 算法、K2 算法以及本文提出的改进 PSO 算法(KPSO)对 ASIA 网络进行结构学习,10 次平均结果如表 2 所示。其中原始得分为用标准 ASIA 网络在实验样本下得出的 BDe 分数。表 2 中同时列出各算法对真实基因表达数据进行结构学习的网络分数(见 4.3 节)。由表 2 我们可以看出,KPSO 始终取得优于 BN-PSO 和 K2 算法的 BDe 分数。在样本规模较小的情况下,比如 ASIA500 和 ASIA1000,KPSO 甚至取得优于网络原始得分的分数。出现这种情况的原因是样本规模较小,其原始得分并不能完全反应网络结构。

在样本规模较大的情况下,KPSO 可以取得十分接近原始得分的分数。

BN-PSO 和 KPSO 算法均在无序条件下进行结构学习。实验中,两种算法的循环次数均设置为 100,粒子种群规模均设置为 10。在 10 次实验中,分别选取两种算法各自最优的一组实验结果进行收敛分析,收敛曲线见图 3 所示。由图 3 可以看出,KPSO 算法可以快速收敛到较优位置,体现了其快速收敛性能。搜索后期,种群增强了局部搜索的能力,旨在细致搜索结果。但算法同样存在潜在的陷入局部最优的风险。

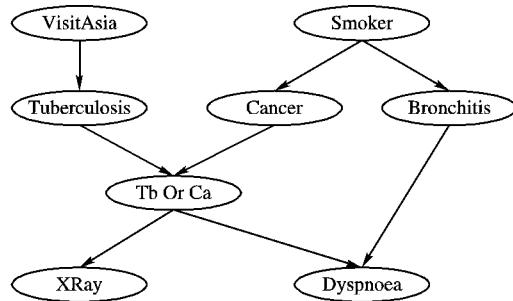


图 2 ASIA 网络结构

表 2 各算法在 ASIA 网络中的 BDe 分数比较

实验样本	BN-PSO	K2	KPSO	原始得分
ASIA500	-1 200.00	-1 173.20	-1 170.40	-1 172.90
ASIA1000	-2 373.40	-2 288.20	-2 279.00	-2 281.10
ASIA5000	-11 622.10	-11 445.40	-11 419.80	-11 215.30
ASIA10000	-22 945.30	-22 634.90	22 620.30	-22 551.30
YEAST77	-176.21	-170.68	-167.82	—

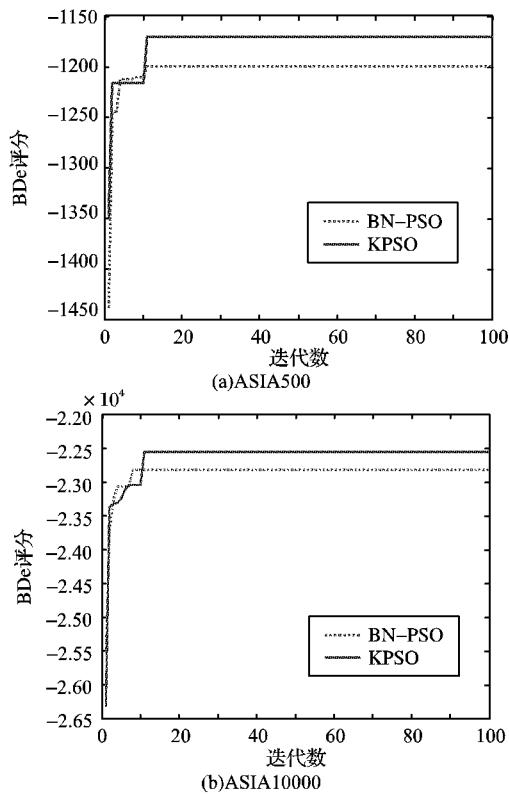


图 3 各算法基于 ASIA 网络的收敛曲线

3.3 基因数据实验结果及分析

文献[17]研究了一个细胞周期内酵母基因调控情况,样本涉及 80 个时间点中的 6000 条基因。研究最终确定了其中 800 条基因参与基因调控过程。在本文设计的实验中,我们

选取其中 8 条在组蛋白的编码中起基础作用的基因(HHT1, HHT2, HHF1, HHF2, HTA1, HTA2, HTB1, HTB2),并研究其在 77 个样本中的表现。生物学家已经发现酵母基因中部分调控信息^[18],对构建基因网络有指导意义。同时,酵母基因是人类已经全部测出基因序列的真核生物之一,对其分析可以对未来人类基因的分析具有尝试性意义。

本文利用 3 值量化^[19]对基因数据进行离散处理,均采用均值一方差的方法。图 4 描述了利用 KPSO 算法对基因实验数据构建的基因网络。算法参数设置中将最大父节点数设为 3。图中实线表示已被证明存在的基因调控边^[18,20-21],其中部分调控边的调控方向与实际相反;虚线表示由 KPSO 算法得出但尚未被证明有效的基因调控边,这些调控边的功能尚有待进一步的生命科学研究。

表 3 详细描述了 KPSO 算法实验得出的调控边情况。算法重建的调控网络中,完全符合生物信息的调控边占 27%,调控端符合但调控方向不符合的边占 36%,目前未被证实其生物信息的调控边占 36%。结构学习的性能依赖于基因样本的大小以及量化等级。基因样本数越多,基因之间父子关系越简单,结构学习能够处理的基因数就越多。大规模的基因网络重建目前尚未能实现,现在的算法主要通过关联分析或者聚类分析选取小规模的基因进行实验。同时,基因表达数据离散处理对结构学习的性能也存在影响。量化层数越多,数据保留的信息也越多,但是需要更多的样本数以建立稳定的网络。如果样本数过少的基因数据,比如微阵列数据,过多的量化层数会导致不稳定的网络结构。现阶段常用的量化层数为二层量化和三层量化,本实验中两种量化所得网络结构较为相似。大于三层的量化容易令算法获得的网络失去意义,因为过多的层次让节点难以找到相互之间的依赖关系,极有可能变得相互“绝缘”。

表 3 KPSO 算法的调控边分析

调控边	调控状态	调控信息出处
HTB1—HHF2	调控节点正确 调控方向反向	文献[18]
HTA1—HTA2	调控节点正确 调控方向反向	文献[18]
HTB2—HTA2	调控节点正确 调控方向反向	文献[18]
HTB1—HTA1	调控节点正确 调控方向反向	文献[20]
HHF1—HHT1	调控节点正确 调控方向同向	文献[18]
HTB2—HHF1	调控节点正确 调控方向同向	文献[18]
HHF1—HHF2	调控节点正确 调控方向同向	文献[21]
HTB1—HHT2	未证实	
HTA1—HTB2	未证实	
HTB2—HHT2	未证实	
HHT2—HHT1	未证实	

4 结语

本文给出了一种新的基因调控网络的结构学习算法——KPSO 算法。算法以贝叶斯网络为模型,在无序条件下引入

粒子群算法对网络结构进行分布式搜索,从而取得理想的网络结构。实验结果表明在无需先验节点顺序的条件下,KPSO 算法始终能够取得比 BN-PSO 算法和 K2 更优异的网络得分。在处理基因数据的实验中,KPSO 算法重建的基因调控网络结构在已有生物学知识的前提下取得 64% 的正确率,另外 36% 的有向边尚未被生物信息学领域证明是否存在。

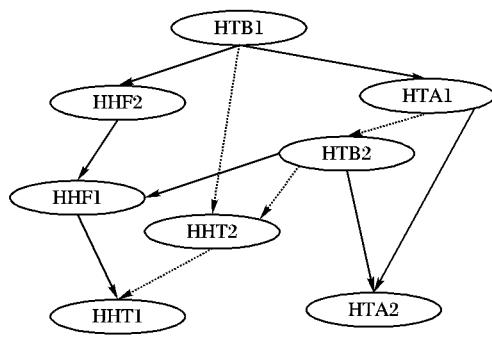


图 4 KPSO 构建的酵母基因调控网络

参考文献:

- [1] de HIDDE J. Modeling and simulation of genetic regulatory system: a literature review [J]. *Journal of Computational Biology*, 2002, 9 (1): 67 – 103.
- [2] KAUFFMAN S A. The origins of order: self-organization and selection in evolution [M]. New York: Oxford University Press, 1993.
- [3] CHEN K C, WANG T Y, TSENG H H. A stochastic differential equation model for quantifying transcriptional regulatory network in *Saccharomyces cerevisiae* [J]. *Bioinformatics*, 2005, 21(12):2883 – 2890.
- [4] FRIEDMAN N, LINIAL M, NACHMAN I. Using Bayesian networks to analyze expression data [J]. *Journal of Computational Biology*, 2000, 7(3):601 – 620.
- [5] CHEN X W, ANANTHA G, WANG X K. An effective structure learning method for constructing gene networks [J]. *Bioinformatics*, 2006, 22(11):1367 – 1374.
- [6] ELOWITZ M B, LEIBLER S . A synthetic oscillatory network of transcriptional regulators [J]. *Nature*, 2000, 403(6767): 335 – 338.
- [7] BUTTE A J, KOHANE I S. Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements [EB/OL]. [2008 – 10 – 10]. <http://helix-web.stanford.edu/psb00/butte.pdf>.
- [8] HECKERMAN D , GEIGER D , CHICKERING D M . Learning Bayesian networks: The combination of knowledge and statistical data [J]. *Machine Learning*, 1995, 20(3):197 – 243.
- [9] COOPER G E, HERSKOVITS E. A Bayesian method for the induction of probabilistic networks from data [J] . *Machine Learning*, 1992, 9(4):309 – 347.
- [10] LARRAAGA P, POZA M, YURRAMENDI Y, et al. Structure learning of Bayesian network by genetic algorithms: A performance analysis of control parameters [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1996, 18(9):912 – 926.
- [11] HENG X C, QIN Z, TIAN L, et al. Learning Bayesian network structures with discrete particle swarm optimization algorithm [C]// IEEE Symposium on Foundations of Computational Intelligence 2007: FOCI 2007. Washington, DC: IEEE Press, 2007: 47 – 52.
- [12] KENNEDY J , EBERHART R C . Particle swarm optimization [C]// Proceedings of the IEEE International Conference on Neural Networks. Piscataway, New Jersey: IEEE Press, 1995: 1942 – 1948.
- [13] DU T, ZHANG S S, WANG Z J. Learning Bayesian networks from data by particle swarm optimization [J]. *Journal of Shanghai Jiaotong University: Science*, 2006, E-11(4):423 – 429.
- [14] PROAKIS J. Digital communications [M]. New York: McGraw-Hill, 2000.
- [15] OLESEN K, ADERS L M. Maximal prime subgraph decomposition of Bayesian networks [J]. *IEEE Transactions on Systems, Man, and Cybernetics*, 2002, 32(1):21 – 31.
- [16] SHI Y , EBERHART R C . A modified particle swarm optimizer [C]// Proceedings of the IEEE International Conference on Evolutionary Computation. Piscataway, NJ, IEEE Press, 1998: 69 – 73.
- [17] SPELLMAN P T, SHERLOCK G, ZHANG M Q, et al. Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization [J]. *Molecular Biology of the Cell*, 1998, 9(12), 3273 – 3297.
- [18] von MERING C, KRAUSE R, SNEL B, et al. Comparative assessment of large-scale datasets of protein-protein interactions [J]. *Nature*, 2002, 417(6887), 399 – 403.
- [19] YU J, SMITH V A, WANG P P, et al. Advances to Bayesian network inference for generating causal networks from observational biological data [J]. *Bioinformatics*, 2004, 20(18):3594 – 3603.
- [20] VENDITTI S, VEGA-PALAS MA, di STEFANO G, et al. Imbalance in dosage of the genes for the heterochromatin components Sir3p and histone H4 results in changes in the length and sequence organization of yeast telomeres [J]. *Mol Gen Genet*, 1999, 262 (2):367 – 377.
- [21] MOSAMMAPARAST N, JACKSONA K R, GOUB Y, et al. Nuclear import of histone H2A and H2B is mediated by a network of karyopherins [J]. *Journal of Cell Biology*, 2001, 153(2):251 – 262.

(上接第 1538 页)

参考文献:

- [1] 左兴权,李士勇.一种新的免疫优化算法及其性能分析[J].系统仿真学报,2003,15(11):1607 – 1609.
- [2] CAO P B, Xiao R B. Assembly planning using a novel immune approach [J]. *International Journal of Advanced Manufacturing Technology*, 2007, 31(7): 770 – 778.
- [3] XIAO R B, TAO Z W, LIU Y. Isomorphism identification of kinematics chains using novel evolutionary approaches [J]. *ASME Journal of Computing and Information Science in Engineering*, 2005, 5 (1):18 – 24.

- [4] JIAO L C, WANG L. A novel genetic algorithm based on immunity [EB/OL]. [2008 – 10 – 10]. http://diuf.unifr.ch/people/wangl/AIS_IEEE.pdf.
- [5] 朱锡华.生命的卫士—免疫系统[M].北京:科学技术文献出版社,2004.
- [6] 张问修,梁怡.遗传算法的数学基础[M].西安:西安交通大学出版社,2007.
- [7] 于瀛,候朝桢.一种克隆选择算法的收敛性分析[J].计算机应用研究,2006, 23(6):96 – 98.
- [8] Universitat Heidelberg. TSPLib [EB/OL]. [2009 – 01 – 02]. <http://www.iwr.uni-heidelberg.de/groups/comopt/software/TSPLIB95/>.