

一种基于粗糙遗传算法的缩放模式双聚类分析方法

李 刚 苗夺谦 王睿智

(同济大学计算机科学与技术系 国家高性能计算机工程中心同济分中心
同济大学嵌入式系统与服务计算教育部重点实验室 上海 201804)

摘 要 研究了基因表达数据的缩放模式发现问题,给出了一种缩放模式双聚类评价函数,并提出了一种基于粗糙遗传算法的双聚类分析方法。该方法先以启发式算法及随机方法生成初始种群,再基于粗糙遗传算法对种群进行迭代,以达到全局优化的目的。在 Yeast 数据集上进行的测试表明,该算法能对启发式算法的结果进行良好的修正。生物显著性分析表明所发现的缩放模式双聚类具有生物学意义。

关键词 双聚类分析,缩放模式,粗糙遗传算法,基因表达数据

Biclustering Method for Detecting Scaling Patterns with Genetic Rough Algorithm

LI Gang MIAO Duo-qian WANG Rui-zhi

(Department of Computer Science and Technology, Tongji University, Tongji Branch, National Engineering & Technology Center of High Performance Computer, The Key Laboratory of "Embedded System and Service Computing" Ministry of Education, Shanghai 201804, China)

Abstract This paper addressed the problem of detecting scaling patterns in gene expression data. A mean ratio residue as a merit function for Scaling patterns was presented. Based on the mean ratio residue a biclustering method was proposed under genetic rough framework. In this method, initial population is generated by Heuristic Rough Biclustering Algorithm as well as random choice. And then it adjust the seeds with Genetic Rough Algorithm. We tested this method on yeast expression data. The experimental results show that the Genetic Rough based method well improves the performance of heuristic algorithm and biclusters found on the yeast data are biologically significant using online GO Term Finder.

Keywords Biclustering, Scaling patterns, Genetic rough algorithm, Gene expression data

1 引言

以基因芯片为代表的高通量分子生物技术的产生,使得同时检测成千上万个基因表达值成为现实。基因芯片技术为基因功能研究提供了一种强有力的工具,高通量特点使其在基因表达分析、疾病诊断和治疗、新药发现等众多领域得到广泛应用。但基因芯片应用过程中产生了海量的、关系复杂的实验数据。如何对这些海量数据进行准确而合理的分析,并从中挖掘出有意义的生物信息,已成为限制该技术进一步发展的主要“瓶颈”,也是当前生物信息学的主要研究内容和主要发展方向。

聚类是最早用于基因表达数据分析的无监督学习方法。然而,传统聚类分析方法执行全空间的单向聚类,揭示基因关于条件的全局信息,无法发现局部信息。2000年,Cheng Yizong 和 George M. Church^[1]第一次将双聚类的概念引入到了基因表达数据挖掘之中。双聚类的出现使得基因表达数据分析得到了飞跃的发展。大量的文献采用了均方残基作为双聚

类评判函数的方法来深入研究基因表达数据中的平移模式^[8-10]。然而基因表达数据中的另一模式——缩放模式的研究虽然已受到关注,但对它的深入研究却较少。

所谓缩放模式,是指双聚类簇中每行(列)元素值可由该双聚类簇中其他行(列)元素乘以一个偏移常量而得到。缩放模式是一致波动模式之一,而一致波动模式有助于研究基因共调控网络,因此缩放模式双聚类也是生物学家较为感兴趣的一种模式。

缩放模式发现问题是指在给定的基因表达数据矩阵中,寻找满足缩放波动一致性约束的最大双聚类簇。缩放模式发现问题也是一个 NP-Complete 问题。

目前对缩放模式的研究大都采用均方残基作为评判标准^[1,3],但是研究表明,均方残基对于发现缩放模式存在明显的不足之处。针对上述问题,本文提出了一种粗糙比率残基,并以此为评价函数设计了基于粗糙遗传机制的算法,用以发现基因表达数据中的缩放模式双聚类簇。算法在 Yeast 表达数据集上进行了测试,结果表明粗糙比率残基对发现基因表

到稿日期:2009-02-13 返修日期:2009-04-29 本文受国家自然科学基金项目(60475019,60775036)以及高等学校博士学科点专项科研基金(20060247039)资助。

李 刚(1984—),男,硕士生,主要研究方向为双聚类分析、粗糙集、数据挖掘,E-mail:iGangLi.LG@gmail.com;苗夺谦(1964—),男,教授,博士生导师,主要研究方向为人工智能、模式识别、数据挖掘、粗糙集、主曲线和粒计算等;王睿智(1968—),女,讲师,主要研究方向为聚类分析、粗糙集、粒计算、Web 智能。

达数据中的缩放模式存在一定的优势。

2 缩放模式双聚类描述

对于给定的矩阵 $A=(X,Y)$, 一个完美的缩放模式双聚类 (I,J) 是由因子子集 $I \subseteq \{1, \dots, n\}$ 和条件子集 $J \subseteq \{1, \dots, m\}$ 构成的子矩阵, 其中元素 a_{ij} 遵循

$$a_{ij} = \mu \times \alpha_i \times \beta_j$$

其中, μ 是该双聚类中的背景效应, α_i 是第 i 行的缩放因子, β_j 是第 j 列的缩放因子。所以缩放模式又称为乘法模型。

缩放模式双聚类存在如下性质:

给定一个 $I \times J$ 的缩放模式双聚类簇 $B=(a_{ij})$, 如果 $a_{ij} \neq 0$, 则 $\log |a_{ij}| - \log |a_{ij}| - \log |a_{ij}| + \log |a_{ij}| = 0$, 其中 $a_{ij} = \frac{1}{|J|} \sum_{j \in J} a_{ij}$ 为行均值, $a_{ij} = \frac{1}{|I|} \sum_{i \in I} a_{ij}$ 为列均值, $a_{ij} = \frac{1}{|I||J|} \sum_{i \in I, j \in J} a_{ij}$ 为该双聚类簇的均值。

证明:

$$\text{设 } B = (a_{ij}) = \begin{pmatrix} \pi_{11}\alpha_1 & \pi_{12}\alpha_1 & \dots & \pi_{1|J|}\alpha_1 \\ \pi_{11}\alpha_2 & \pi_{12}\alpha_2 & \dots & \pi_{1|J|}\alpha_2 \\ \dots & \dots & \dots & \dots \\ \pi_{|I|1}\alpha_1 & \pi_{|I|2}\alpha_1 & \dots & \pi_{|I||J|}\alpha_1 \end{pmatrix},$$

其中, $\pi_j \neq 0, j=1, \dots, |J|; \alpha_i \neq 0, i=1, \dots, |I|$ 。

因为, $a_{ij} = \pi_j \alpha_i, a_{ij} = \alpha_i \mu_j, a_{ij} = \pi_j \mu_\alpha, a_{ij} = \mu_\alpha \mu_\alpha$, 其中 μ_α 为 $\pi_1 \dots \pi_{|J|}$ 的均值, μ_α 为 $\alpha_1 \dots \alpha_{|I|}$ 的均值。当 $\mu_\alpha \neq 0$ 时, 有 $\frac{a_{ij}}{a_{ij}} = \frac{\pi_j \alpha_i}{\pi_j \mu_\alpha} = \frac{\alpha_i}{\mu_\alpha}, \frac{a_{ij}}{a_{ij}} = \frac{\alpha_i \mu_j}{\mu_\alpha \mu_\alpha} = \frac{\alpha_i}{\mu_\alpha}$, 所以 $\frac{a_{ij}}{a_{ij}} = \frac{a_{ij}}{a_{ij}}$ 成立。如果 $a_{ij} \neq 0$, 则 $\log |a_{ij}| - \log |a_{ij}| - \log |a_{ij}| + \log |a_{ij}| = 0$ 。

3 粗糙比率均值残基

根据缩放模式双聚类的上述性质, 本文先给出了比率残基。然后, 基于粗糙集上下近似的思想, 提出一种适用于缩放模式重叠结构双聚类问题的评价函数——粗糙比率均值残基。

3.1 比率均值残基

在一个非完美双聚类 (I,J) 中, 元素 a_{ij} 的比率残基 r_{ij} 定义为

$$r_{ij} = \log |a_{ij}| - \log |a_{ij}| - \log |a_{ij}| + \log |a_{ij}|$$

该双聚类 (I,J) 的比率均值残基定义为

$$R'(B_k) = \frac{1}{|B_k|} R(I,J) = \frac{1}{|I||J|} \sum_{i \in I, j \in J} |r_{ij}| \quad (1)$$

3.2 粗糙比率均值残基

3.2.1 粗糙集上下近似

粗糙集理论是 1982 年由 Z. Pawlak^[2] 提出的一种处理模糊和不确定知识的工具。设 U 为论域, $R \subseteq U \times U$ 是论域上的一个等价关系。二元对 $A=(U,R)$ 称为近似空间。等价关系 R 将论域 U 划分成若干个子集, 划分表示成 $U/R=(E_1, E_2, \dots, E_n)$, 其中 E_i 是关系 R 上的等价类。如果两个元素 $u, v \in U$ 属于同一个等价类 $E \subseteq U/R$, 则称 u 和 v 是不可区分的。关系 R 上的等价类称为近似空间 $A=(U,R)$ 上的基本集合。

同一个等价类中的元素是不可区分的, 因此根据近似空间 A 中的基本集合不能对任意的集合 $X \subset U$ 获得精确的表示。但是, 任意一个 X 都可以用它的上近似和下近似描述。 X 的下近似 $\underline{A}(x)$ 是 X 中所有包含在 X 中的子集的并集, X 的上近似 $\bar{A}(x)$ 是所有和 X 相交不为空的基本子集的并集。

二元对 $(\underline{A}(x), \bar{A}(x))$ 是近似空间 $A=(U,R)$ 上普通集合 X 的描述, 即 X 的粗糙集合。 X 下近似中的元素确定地属于 X ; X 上近似中的元素不一定属于 X , 它可能属于 X , 也可能不属于 X 。

在聚类过程中, 可用下近似和上近似分别描述聚类簇的最小轮廓和最大轮廓。这样, 簇的下近似集成为簇的核心部分, 反映了簇的位置和形状; 上近似中去掉下近似得到的边界区域由于可能同时属于多个簇, 因此可以发现基因表达数据中可重叠的簇。

3.2.2 粗糙比率均值残基

设给定的基因表达矩阵 A 中存在 K 个双聚类簇, 第 k 个双聚类簇 $B_k=(I_k, J_k)$ 的粗糙比率均值残基为

$$R'(B_k) = \frac{1}{|B_k|} \sum_{a_{ij} \in B_k} |\log_2 |a_{ij}| - \log_2 |a'_{i,j}| - \log_2 |a'_{i,j}| + \log_2 |a'_{i,j,k}|| \quad (2)$$

其中,

$$a'_{i,j,k} = \begin{cases} \omega_{\text{low}} \times \frac{\sum_{a_{ij} \in B_k} a_{ij}}{|J_k|} + \omega_{\text{up}} \times \frac{\sum_{a_{ij} \in (B_k - \underline{B}_k)} a_{ij}}{|J_k - \underline{J}_k|}, & \text{if } i \in (\bar{I}_k - \underline{I}_k) \\ \frac{\sum_{a_{ij} \in B_k} a_{ij}}{|J_k|}, & \text{otherwise i.e. } i \in \underline{I}_k \end{cases} \quad (3)$$

$$a'_{i,j,k} = \begin{cases} \omega_{\text{low}} \times \frac{\sum_{a_{ij} \in B_k} a_{ij}}{|I_k|} + \omega_{\text{up}} \times \frac{\sum_{a_{ij} \in (\bar{B}_k - \underline{B}_k)} a_{ij}}{|\bar{I}_k - \underline{I}_k|}, & \text{if } j \in (\bar{J}_k - \underline{J}_k) \\ \frac{\sum_{a_{ij} \in B_k} a_{ij}}{|I_k|}, & \text{otherwise i.e. } j \in \underline{J}_k \end{cases} \quad (4)$$

$$a'_{i,j,k} = \begin{cases} \omega_{\text{low}} \times \frac{\sum_{a_{ij} \in B_k} a_{ij}}{|B_k|} + \omega_{\text{up}} \times \frac{\sum_{a_{ij} \in (\bar{B}_k - \underline{B}_k)} a_{ij}}{|\bar{B}_k - \underline{B}_k|}, & \text{if } \bar{B}_k - \underline{B}_k \neq \emptyset \\ \frac{\sum_{a_{ij} \in B_k} a_{ij}}{|B_k|}, & \text{otherwise} \end{cases} \quad (5)$$

其中, $1 \leq k \leq K, \omega_{\text{low}}$ 为下界权值, ω_{up} 为边界区域权值, $\omega_{\text{low}} + \omega_{\text{up}} = 1$ 。

显然, 当每个行子集、列子集及双聚类簇的上下界相等时, 粗糙比率均值残基式(2)就退化为比率均值残基式(1)。可见, 粗糙比率均值残基是比率均值残基的泛化。粗糙比率均值残基越小, 找到的可重叠双聚类缩放波动一致性越好。下面, 将粗糙比率均值残基简记为 Rscore。

4 粗糙遗传算法描述

遗传算法是一种借鉴生物界自然选择和自然遗传机制的随机搜索算法^[1]。与传统算法不同, 它能同时对搜索空间中的多个解进行评估, 因此具有较好的全局搜索能力, 不容易陷入局部最优。本文对基本遗传算法进行了改进, 提出了粗糙遗传算法的概念。

4.1 算法描述

4.1.1 启发式算法生成种子

采用简单的种子发现技术产生双聚类种子, 即行(基因)和列(条件)分别单向 K -Means 聚类后其结果的组合, 对得到的不相交的子矩阵做贪心点删除操作, 删去比率残基最大的行和列, 得到给定大小的种子, 选出缩放模式一致性较好的部

分种子。

输入:预处理后 $n \times m$ 的基因表达矩阵 $A = (X, Y)$, 初始种子规模为 $n_s \times m_s$, 种子粗糙比率均值残基阈值 δ_s , 粗糙双聚类中上下界权重对 (w_{low}, w_{up}) , 基因聚类数 $k_g = \lceil n/n_s \rceil$, 条件聚类数 $k_c = \lceil m/m_s \rceil$, 基因聚类时相对阈值 ζ_g , 条件聚类时相对阈值 ζ_c , 粗糙比率均值残基阈值 δ , 行增益差阈值 ϵ_g , 列增益差阈值;

输出: K 个可重叠 δ 双聚类 $B_k, k=1, \dots, K$ 。

4.1.2 粗糙遗传算法对种子进行全局优化

启发式算法是局部搜索, 虽然速度快, 但是容易陷入局部最优, 得到的是次优解。文献[4-7]中分别采用各种方法对次优解进行了优化, 这些方法大都是基于均方残基构造遗传算法的适应度函数, 本文则采用基于粗糙比率均值残基构造遗传算法的适应度函数进行全局优化, 算法中对基本遗传算法进行了一系列的改进。下面详细介绍算法。

a) 染色体编码

本文采用二进制进行编码, 考虑粗糙双聚类是一个广义粗糙集 (B, \bar{B}) , 因此同时对上下近似进行编码:

1	0	1	1	0	1	1	0	1	1	1	0	1	1	1	0
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

编码策略: 数据矩阵 A 中某行或列属于一个双聚类的边界或下界, 则染色体字符串中对应的位置为 1, 否则为 0。

编码顺序依次是: 行下近似空间、列下近似空间、行上近似空间、列上近似空间。

二进制编码具有编码、解码简单, 交叉、编译方便, 字符集小等优点, 在整个遗传算法中操作方便。

b) 初始种群

种群的初始化是遗传算法的重要方面。初始种群的特性对计算结果和计算效率均有重要影响。要实现全局最优, 初始种群在解空间中应尽量分散。标准遗传算法是用随机方法产生初始种群, 这可能导致初始解群体在解空间分布不均匀, 从而影响算法的性能。本文的初始种群 50% 来自启发式算法生成的种子, 50% 来自随机生成的种子, 已达到种子多样性的目的。

c) 适应度函数

在遗传算法中, 使用适应度来度量群体中各个个体在优化计算中所能达到或接近于找到最优解的优良程度。本文的目标是从数据矩阵中找到最优的可重叠缩放模式双聚类, 具体地是寻找粗糙比率均值残基小于给定阈值 δ 的最大体积缩放模式双聚类。根据此目标, 定义双聚类 $B_k = (I_k, J_k)$ 的适应度函数为

$$Fitness(B_k) = \begin{cases} |I_k| \cdot |J_k|, & \text{if } R'(B_k) \leq \delta \\ \frac{\delta}{R'(B_k)}, & \text{otherwise} \end{cases} \quad (6)$$

d) 遗传算法的基本操作

选择: 采用锦标赛选择算子。

交叉: 采用均匀交叉算子。均匀交叉是指两个配对个体的每个基因座上的基因都以相同的交叉概率进行交换, 从而形成两个新的个体。

变异: 采用基本位变异中的一位变异。

e) 防止乱伦机制

为了维持种群的多样性, 选择配对个体时, 本文采用防止乱伦机制, 限制相似个体交配, 即从当前种群 C 中, 以概率选择两个个体, 若这两个个体间的海明距离小于给定的阈值, 则交叉; 否则, 放回重取。

f) 精英策略

为了保留历代中适应度最高的个体, 本文采用精英策略, 将当代种群中适应度值最高的个体完整地复制到下一代种群中, 不参与交叉和变异操作。

4.2 算法时间复杂性分析

假定基因表达数据矩阵中基因个数为 n , 条件个数为 m , 种子规模为 $n_s \times m_s$ 。种子生成阶段采用的启发式算法包括基因聚类、条件聚类、双向聚类结果组合、双聚类种子改进、贪心种子生长 5 个步骤。

因此, 粗糙遗传算法中启发式算法生存种子的时间复杂度近似为 $O((n+m) \times n \times m \times n_s \times m_s)$ 。

遗传算法的时间复杂度为 $O(g \times 2(n+m) \times P)$, 其中 g 是代数, $2(n+m)$ 是染色体编码长度, P 是种群规模。

5 实验结果

实验数据集选用基因表达数据双聚类分析中最常用的 Yeast 表达数据集^[1]。Yeast 数据集是一个 2884×17 的矩阵, 本文对其进行扁平基因的过滤, 过滤后的数据集为 2057×17 的 $[0, 600]$ 实数矩阵。

5.1 参数设置

在启发式算法的搜索过程的 K -means 阶段, 种子的规模大小为 10×5 , 取基因聚类数 $k_g = 150$, 条件聚类数 $k_c = 4$ 。在种子的生长过程中, 双聚类的上下近似的权重选为 $w_{low} = 0.75, w_{up} = 0.25$ 。由于 Yeast 数据集的值变化范围小, 因此基因划分阈值设为 $\zeta_g = 0.06$, 条件划分阈值 $\zeta_c = 0.08$, 粗糙比率均值残基阈值选为 $\delta_1 = 0.04$ 。在遗传算法中, 适应度函数中比率均值残基阈值 $\delta_2 = 0.07$ 。具体的参数设置如表 1 所列。

表 1 参数设置

参数		取值
启发式 搜索阶段	基因聚类数 k_g	150
	条件聚类数 k_c	4
	种子规模 $n_s \times m_s$	10×5
	上近似权重 w_{up}	0.25
	下近似权重 w_{low}	0.75
	基因划分阈值 ζ_g	0.06
	条件划分阈值 ζ_c	0.08
	粗糙比率均值残基阈值 δ_1	0.04
适应度函数中比率均值残基阈值 δ_2		0.07
遗传 算法阶段	种群规模 P	200
	代数 g	100
	配对个体海明距离阈值 δ_3	$0.4 \times \min\{ B_1 , B_2 \}$
	交叉概率 cp	0.7
	变异概率 mp	0.1
锦标赛规模 N		4

5.2 结果分析

选取启发式结果的 100 个种子和随机生成的 100 个种子为初始种群进行粗糙遗传算法。算法运行相关图如图 1—图 3 所示。

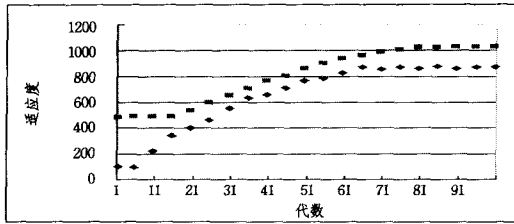


图1 适应度变化

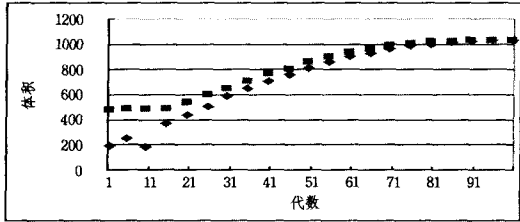


图2 体积变化

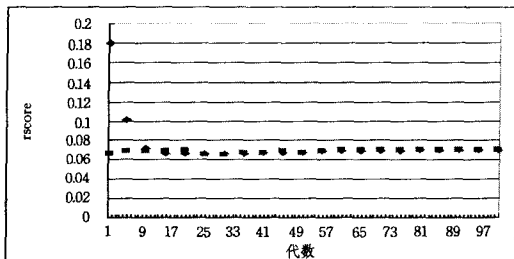


图3 Rscore值

其中,灰色线条表示最优个体值,黑色表示平均值。从图中可以明显看出,粗糙遗传算法对启发式的结果进行了全局优化;种子的适应度随代数的增长而加强,最终趋近于平衡;种子的体积也有同样的效果;而 Rscore 值受到阈值的限制,趋近于 0.07。

表 2 给出了启发式算法和粗糙遗传算法分别在 Yeast 数据集上进行 10 次实验的最优结果。

表 2 最优结果比较

算法	Rscore	体积
粗糙遗传双聚类	0.0699	1336
启发式搜索	0.0595	284

从表 2 中可以看出,本文提出的粗糙遗传双聚类算法的最优结果与启发式搜索的最优结果相比,虽然在 Rscore 值上有所增长,但是种子的适应度和种子的体积得到了更大的增长。而 Rscore 值的增长是在允许范围(Rscore 设定的阈值)内的,这说明粗糙遗传双聚类能对启发式的结果进行全局优化。

5.3 生物学意义分析

本文采用在线分析工具 GO Term Finder^[12]来考察启发式算法和粗糙遗传算法的结果在联合生物过程方面的生物学意义。在线提交粗糙遗传算法发现的缩放模式双聚类后,得到表 3 所列的分析结果。

表 3 算法发现的双聚类的生物过程 GO 词

Biclust	GO term	Percentage	P-value
B6	biological regulation	65 out of 233 genes, 27.9%	0.00069
B8	Unannotated	21 out of 272 genes, 7.7%	1.37e-20

表 3 是粗糙遗传算法运行 10 次后发现的 2 个双聚类共有的显著生物过程本体词。算法运行第 6 次得到的最优双聚类 B6 的基因主要与 biological regulation 过程有关,而第 8 次实验得到的最优双聚类 B8 中有 21 个基因与某一尚未得到注释的生物过程显著相关($P\text{-value}=1.37e-20$),其统计显著性可由 P -值反映,较小的 P -值表示聚类结果在尚未得到注释的生物过程中有重要的生物学意义。未经标注但仍有生物学显著意义的显现是可解释的,因为在 Yeast 数据集的 2879 个有价值基因中,有 2601 个基因已得到生物过程本体注释,尚有 3.8% 的基因未获得注释。从实验结果中发现, Yeast 数据集得到的双聚类缩放模式不明显,生物显著性分析结果也验证了这一发现。

结束语 本文基于粗糙双聚类的模型,采用粗糙遗传双聚类算法对启发式算法的结果进行了全局优化,并针对 Yeast 基因表达数据中的缩放模式进行了实验。实验结果表明,粗糙遗传双聚类算法能对启发式的结果进行全局优化。

参考文献

- [1] Cheng Y, Church G M. Biclustering of expression data[C]// Proceedings of the 8th International Conference on Intelligent Systems for Molecular Biology (ISMB 2000). La Jolla, CA, August 2000; 93-103
- [2] Pawlak Z. Rough Sets[J]. International Journal of Information and Computer Sciences, 1982, 11: 145-172
- [3] Wang H, Wang W, Yang J, et al. Clustering by pattern similarity in large data sets[C]// Proceedings of the 2002 ACM SIGMOD International Conference on Management of Data, 2002; 394-405
- [4] Bleuler S, Prelic A, Zitzler E. An EA framework for biclustering of gene expression data[C]// Proceedings of Congress on Evolutionary Computation, 2004; 166-173
- [5] Chakraborty A, Maki H. Biclustering of gene expression data using genetic algorithm[C]// Proceedings of the 2005 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB '05), 2005; 1-8
- [6] Divina F, Aguilar-Ruiz J S. Biclustering of expression data with evolutionary computation[J]. IEEE Transactions on Knowledge and Data Engineering, 2006, 18(5): 590-602
- [7] Mitra S, Banka H. Multi-objective evolutionary biclustering of gene expression data[J]. Pattern Recognition, 2006, 39(12): 2464-2477
- [8] Cho H, Dhillon I S. Co-clustering of human cancer microarrays using minimum sum-squared residue co-clustering[J]. IEEE/ACM Transactions on Computational Biology and Bioinformatics (2007 accepted, DOI 10.1109/TCBB.2007.70268)
- [9] Lazzaroni L C, Owen A. Plaid models for gene expression data[J]. Statist Sinica, 2002, 12(1): 61-86
- [10] Yang J, Wang H, Wang W, et al. Enhanced biclustering on expression data[C]// Proceedings of the 3rd IEEE Conference on Bioinformatics and Bioengineering (BIBE 2003), 2003; 321-327
- [11] 雷英杰, 张善文, 李续武, 等. MATLAB 遗传算法工具箱及应用[M]. 西安: 西安电子科技大学出版社, 2005
- [12] <http://www.yeastgenome.org/cgi-bin/GO/goTermFinder.pl>