

一种扩展的 Post 类函数及其对基因 调控机制的影响

许 鹏¹, 欧阳宏嘉¹, 沈良忠², 刘文斌¹

(1. 温州大学物理与电子信息工程学院, 浙江温州 325035; 2. 温州大学城市学院, 浙江温州 325035)

摘 要: 已有的理论分析表明,有序函数对网络的运行及其动态行为具有重要的影响. Post 类函数是目前发现的数量最大的一种有序函数,其缺点是它仅反映了基因之间的正调控机制. 本文在 Post 类函数的基础上,将其扩展为一种同时具有正负调控机制的函数类. 仿真结果表明:扩展的 Post 类函数在维持网络的有序性及长程相关性方面,和原有的 Post 类函数以及渠化函数类似. 同时与已有的 Post 类和渠化函数相比,新扩展的 Post 类函数数量最多. 从进化的角度,新扩展的函数更有可能成为自然进化选择的调控方式. 通过对八种真实生物系统布尔网络模型函数分布的研究,也部分的支持这一结论.

关键词: 布尔网络; 动态行为; Post 类函数; 基因调控网络

中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112 (2016)07-1689-05

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2016.07.024

The Study of an Extended Post Class Functions and its Influence on the Gene Regulatory Mechanism

XU Peng¹, OUYANG Hong-jia¹, SHEN Liang-zhong², LIU Wen-bin¹

(1. Department of Physics and Electronic Information Engineering, Wenzhou University, Wenzhou, Zhejiang 325035, China;

2. City College, Wenzhou University, Wenzhou, Zhejiang 325035, China)

Abstract: Theoretical analysis reveals that ordered functions play an important role in the running and dynamical behavior of networks. Post function class has been known as the maximal ordered function. However, it only describes the positive regulatory mechanism appeared in genetic regulatory networks. In this paper, we extend the Post function class so that it contains both the positive and negative regulatory mechanism. Simulation results show that the extended Post function class can present similar dynamical characters as the original Post class and Canalizing function class. It has the ability to keep ordered behaviors and can emerge the long-range correlated dynamics. From the point of view of natural evolution, we conclude that such abundant ordered functions are more possible to be selected as the interactions among genes. Finally, the distribution of function classes on eight real systems also partially supports our conclusion.

Key words: Boolean network; dynamic behavior; post class function; gene regulatory network

1 引言

基因调控网络的研究已经成为系统生物学的一个热点问题,基因调控网络的模型主要有布尔网络、贝叶斯网络,及微分方程模型等^[1]. 布尔网络模型是一种最简单的模型^[2],基因的表达状态为“0”和“1”二种状态,“0”表示抑制或不表达,“1”表示激活或表达. 这种模型虽然简单,但是仍然能够反映网络运行过程中复杂的

动态行为. 已有的研究表明:真实基因调控网络处于有序和无序的边界. 具有这种动态行为的网络在各种噪声下既能保持一定的稳定性,同时又能对环境变化做出适当的调整,具有适应性^[3]. 探知产生这种临界性的原因就成为一个重要的问题. 一方面,人们认为包括基因调控网络在内的生物网络具有无尺度或小世界特征,如 Jeong 等发现酵母的代谢网络服从 $\gamma_{in} \approx \gamma_{out} \approx 2.2$ ^[4,5]. 另一方面,基因之间的作用方式对于系统的稳

定性同样具有重要的影响. 渠化函数在阻止无序行为方面具有重要的作用, 增加网络中渠化函数的比例, 可以促使网络穿越临界界限并进入有序状态. 大量事实表明在高等脊椎动物的基因调控系统中, 渠化函数非常普遍^[6]. 近年来大量文献表明, 真核生物的转录调控规则具有很高的渠化性倾向^[7]. 2003 年, Shmulevich 在渠化函数的基础上提出了一种 Post 类的有序函数. 这种函数的封闭性能够保证它的各种复合函数仍然属于 Post 类. 并且当连通度 $k \geq 5$ 的时候, 其数量远大于渠化函数^[3]. 因此, 他认为 Post 类函数更有可能成为基因之间相互作用的方式. Post 类函数实际上反映的是基因之间的一种正调控关系, 然而, 细胞网络中不仅有正调控, 同时也存在着负调控关系. 从生物学角度来看, 正调控是完成细胞分化的必要部分, 而负调控则有助于维持生物体内平衡^[8]. 本文将 Post 类函数扩展为同时具有正负调控机制的一类函数. 然后, 从数量上、有序性和长程相关性三个方面研究了渠化函数, Post 类函数和扩展 Post 类函数的关系及其对基因调控机制的影响. 最后, 我们以八种真实生物为例, 讨论了它们的布尔网络模型中三种函数类的分布情况.

2 概念

2.1 布尔网络及其灵敏度

一个布尔网络 $G(V, F)$ 由节点集 $V = \{x_1, \dots, x_n\}$ 和函数集 $F = \{f_1, \dots, f_n\}$ 组成, 其中 $f_i: \{0, 1\}^k \rightarrow \{0, 1\}$ 为基因 x_i 的布尔函数, k_i 指调控 x_i 的基因的个数, 通常也指 x_i 的入度. x_i 在 $t+1$ 时刻的状态完全由其调控基因 $x_{j_1}, x_{j_2}, \dots, x_{j_{k_i}}$ 在 t 时刻状态确定, 可以写为 $x_i(t+1) = f_i(x_{j_1}(t), x_{j_2}(t), \dots, x_{j_{k_i}}(t))$. 网络的演化通常采取同步演化的方式. Shmulevich 提出了网络灵敏度的概念刻画网络的有序性, 灵敏度 S 反映了 1 比特的扰动对于网络演化的影响: $S < 1$ 对应于有序状态; $S > 1$ 对应于混沌状态; $S = 1$ 对应介于二种之间的临界状态.

2.2 长程相关性

长程相关性是一种衡量复杂系统运行平滑性的动态行为. 具有长程相关性的系统通常具有一种“记忆”特性或“自相似性”. 前者指系统的演化过程通常受其前面很长时间范围行为的影响; 后者指系统的行为或特征在不同的时间尺度上具有相似性. 通常采用标度指数 α 表征时间序列信号的长程相关性. $\alpha = 1.5$ 对应布朗噪声, 表示信号在小时间范围内具有很强的相关性. $\alpha = 0.5$ 时对应白噪声, 表示信号完全无规律性. $\alpha = 1$ 对应于 $1/f$ 过程, 表示信号局部呈无序, 而宏观上具有一定相关性. 通常都认为 $0.9 \leq \alpha \leq 1.1$ 为 $1/f$ 过程, 其他两侧区域分别对应布朗噪声和高斯白噪声. DFA 已经被广泛的应用于 DNA 序列、神经颠覆信号、人的步

态、心脏跳动信号等^[9]. 很多医学信号的研究表明: 偏离 $1/f$ 波动往往与系统功能的失调相关^[10,11].

2.3 渠化函数

当某个变量取特定值时可以唯一确定函数的输出值, 这种函数就是渠化函数. 渠化函数的定义为: 如果存在 y 和 z , 使得对于所有输入 $(x_1, x_2, \dots, x_n) \in \{0, 1\}^n$, 当 $x_i = y$ 时, $f(x_1, x_2, \dots, x_n) = z$, 则称布尔函数 $f: \{0, 1\}^n \rightarrow \{0, 1\}$ 是渠化函数, x_i 为渠化变量. 渠化函数可以有一个或多个渠化变量. 例如在函数 $f(x_1, x_2, x_3) = 01110011$ 中, x_2 就是一个渠化变量.

2.4 Post 类函数^[12,13]

对于 $\mu \geq 2$, 如果任意 μ 个使函数 f 取值为 1 的向量的某一个分量均为 1, 则称该函数属于 A^μ 类. 类似地, 如果任意 μ 个使函数 f 取值为 0 的向量的某一个分量均为 0, 相应的函数类为 a^μ . 由 $A^\mu \cup a^\mu$ ($2 \leq \mu \leq k$) 构成的函数类称为 Post 类函数. Post 类函数要求函数的输出值与其输入变量的值要满足一致性, 即都同时为 0 或 1. 这其实是一种正调控机制. Post 类函数具有一个重要的性质就是封闭性, 即任意多个 Post 类函数的合成仍然属于 Post 类函数.

2.5 扩展 Post 类函数

本文将 Post 类函数作如下扩展, 使其也包含中负调控机制关系. 如果任意 μ 个使函数 f 取值为 1 的向量的某一个分量均为 0, 则称该函数属于 A_{1-0}^μ 类. 类似地, 如果任意 μ 个使函数 f 取值为 0 的向量的某一个分量均为 1, 得到相应的函数类为 a_{0-1}^μ . 显然, 扩展的 A_{1-0}^μ 和 a_{0-1}^μ 函数类要求函数的输出与其输入变量的值必须满足相反性. 因此, 它们反映的是一种负调控机制. 我们将 A_{1-0}^μ 和 a_{0-1}^μ 函数加如到 Post 函数类就得到扩展 Post 类函数.

3 结果和讨论

下面我们从函数数量、有序性和长程相关性三个方面, 探讨渠化函数、Post 类函数以及扩展 Post 类函数对系统动态行为的影响. 对每种函数类, 我们分别产生 100 个灵敏度 S 为 0.7 到 1.3 的布尔网络, 其中最大连通度 $k = 4$. 由于 $\mu = 2$ 时的 Post 类或者扩展的 Post 类函数的数量最大, 本文 μ 值均为 2.

3.1 函数数量分布

图 1 给出了连通度 $k = 2, 3, 4$ 时, Post 类函数、渠化函数以及扩展 Post 类函数的数量分布情况. 可以看出, 在任所有情况下扩展 Post 类函数都是最多, 而 Post 类函数只有当 $k \geq 5$ 时才会超过渠化函数的数量. 此外, 由于对称性, 扩展的 Post 类函数数量大约总是 Post 类函数的两倍.

这三种函数都具有有序性, 因而它们之间存在一些

交集. 表 1 列出了 $k=2,3,4,5$ 时, 三种函数及其交集的数量. 可以看出, 新扩展的 Post 类函数与渠化函数的交集正好等于渠化函数的数量, 它与 Post 类函数的交集正好等

于 Post 类函数. 因此, 扩展的 Post 类函数正好包含了这两类函数. 结合以上分析, 从数量上自然界更有可能选择这种扩展的 Post 类函数来维持系统的有序性.

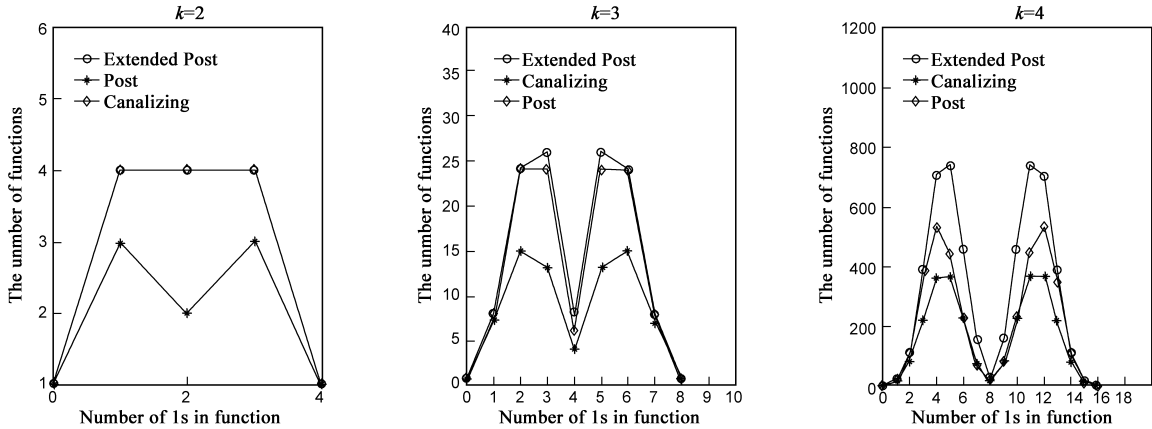


图1 连通度 $k=2,3,4$ 时,渠化函数、Post类和扩展Post类函数的数量分布

表 1 三种函数在连通度 $k=2,3,4,5$ 时的数量及其交集数量 (A, B 和 C 分别表示渠化函数、Post 类函数以及扩展 Post 类函数)

连通度 k	2	3	4	5
函数总数	16	256	16384	4294967296
渠化函数数量	14	120	3514	1292276
Post 类函数数量	10	76	2740	2629551
扩展 Post 类函数数量	14	126	5202	5241500
$A \cap B \cap C$	10	73	1896	654819
$A \cap B$	10	73	1896	654819
$A \cap C$	14	120	3514	1292276
$B \cap C$	10	76	2740	2629551

图 2 给出了连通度 $k=2,3,4,5$ 时, 每个函数类占总的布尔函数的比例. 可以看出, 当 $k > 4$ 时有序函数占全部函数的比例将急剧减少. 如果生物网络中有序函数占大多数的假设成立, 则基因之间的连通度不应该太大, 否则将会增加自然选择的压力. 因此, 有序性假设与生物网络的幂律分布以及小世界特征假设恰好一致. 在基因调控网络的模型推理中, 大部分的研究通常都将连通度取值为 $k=3,4$, 从这一点上来说也是较为

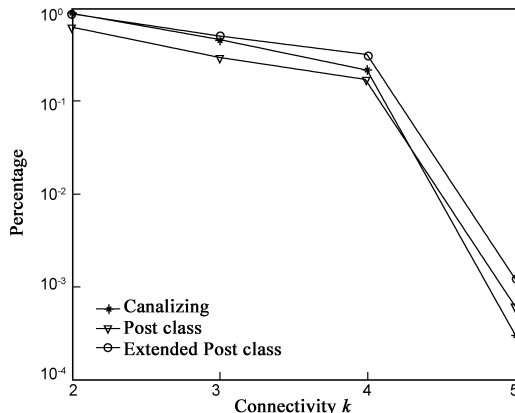


图2 渠化函数、Post类和扩展Post类函数类在 $k=2,3,4,5$ 的比例

合理的.

3.2 有序性

给定 t 时刻的两个状态 $x^{(1)}(t) = (x_1^{(1)}(t), \dots, x_n^{(1)}(t))$ 和 $x^{(2)}(t) = (x_1^{(2)}(t), \dots, x_n^{(2)}(t))$, 它们的归一化汉明距离为其中不相同基因状态所占比例, 即:

$$\rho(t) = \frac{1}{N} \sum_{i=1}^N (x_i^{(1)}(t) \oplus x_i^{(2)}(t))$$

其中 \oplus 是模为 2 加法. 设 $x^{(1)}(t+1)$ 和 $x^{(2)}(t+1)$ 分别表示状态 $x^{(1)}(t)$ 和 $x^{(2)}(t)$ 的后继状态. 类似的, $\rho(t+1)$ 是这二个后继状态的归一化汉明距离. Derrida 曲线是一种表示 $\rho(t+1)$ 随 $\rho(t)$ 变化的曲线^[14]. 在无序网络中, 即使很小的扰动也会迅速扩大. 对应的 Derrida 曲线通常位于主对角线之上, 且在原点附近的斜率大于 1. 反之, 在有序网络中, 很小的扰动会迅速消失. 对应的 Derrida 曲线通常位于主对角线以下, 且在原点附近的斜率小于 1. 从图 3 可以看出, 由三种有序函数构造的网络的 Derrida 非常接近, 而且在原点附近的曲线斜率

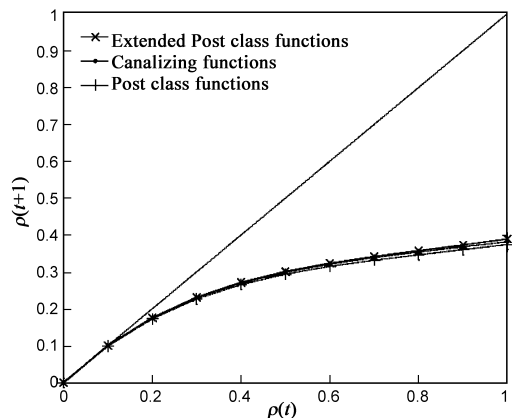


图3 由扩展Post类、渠化函数和Post类函数构造的网络的 Derrida 曲线

均接近 1. 这说明新扩展的 Post 函数与 Post 函数和渠化函数类似, 具有抑制噪声传播并维持系统稳定的能力.

3.3 长程相关性

图 4 给出了分别由三种函数类构造的网络的长程动态相关性在噪声和灵敏度空间 (η, S) 中的相变图. 图 4 中的中部区域对应 $1/f$ 波动的噪声范围, 可以看出, 三种函数的相变区间大致相似. 其中 Post 类函数构成的网络的 $1/f$ 波动的噪声区间最长, 这可能与其封闭性有关. 扩展 Post 类函数与渠化函数构成的网络的 $1/f$ 波动的噪声区间非常接近. 因此, 由扩展 Post 类函数构造的网络, 同样能够产生类似的 $1/f$ 波动.

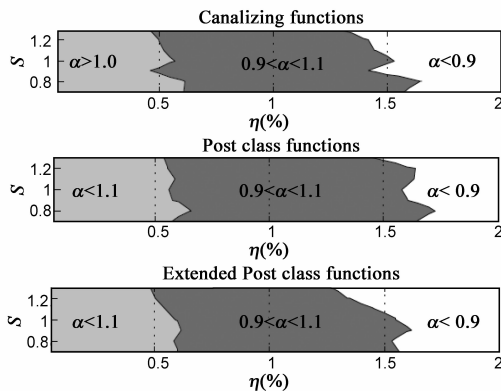


图 4 由渠化函数、Post 类函数和扩展 Post 类函数构造的布尔网络的长程相关性的相变图

4 真实生物网络的研究

布尔网络模型已经广泛应用于酵母细胞周期表达、哺乳动物细胞周期表达、果蝇体节极性网络、花发育形态表达等不同生物的基因调控网络建模和动态行为分析. 下面我们讨论八种典型细胞网络中布尔函数类型的分布情况: 裂殖酵母细胞周期 (Fy)、芽殖酵母细胞周期 (By)、果蝇细胞周期 (Dc)、哺乳动物细胞周期 (Mac)、及哺乳动物体节极性网络 (Ds)、拟南芥的形态网络 (Ar)、人类神经胶质瘤细胞 (Nlc) 和转移性黑色素瘤细胞网络 (Mec). 它们分别代表了细胞周期、细胞分化和肿瘤细胞三种细胞过程. 表 2 列出了八个网络中基因的个数、各种函数类的数量及不同连通度情况下函数的个数^[8,15-19].

可以看出, 在所有模型中扩展的 Post 类函数的数量几乎都超过一半. 而渠化函数在 By、Dc 网络中的数量小于一半, Post 类函数的数量则在 By、Dc、Mac 和 Ds 网络中小于一半. 进一步观察 By、Dc、Mac 和 Ds 等四个网络, 我们发现其中约三分之一节点的连通度 $k \geq 5$, 这是导致其中扩展的 Post 类函数和渠化函数的数量相对较少的主要原因. 由于网络推理技术的限制, 上述模型与真实网络的规模相比还很小, 但是它们还是可以部

分反映实际系统中基因之间的作用关系. 这些模型中各种函数的分布部分的验证了扩展 Post 类函数更有可能成为自然进化选择的作用方式. 同时, 也说明了网络中具有高连通度的 Hub 类节点不应该太多, 否则, 有序函数的数量将会急剧减少.

表 2 八种真实生物网络中的基因个数及各种函数类的个数, 以及各个连通度 k 的函数的个数

细胞网络	Fy	By	Dc	Mac	Ds	Ar	Nlc	Mec
基因个数	14	11	14	10	6	6	14	7
C 渠化函数数量	14	4	4	6	4	6	14	6
Post 类函数数量	12	2	4	3	2	6	7	5
扩展 Post 类函数数量	14	5	9	6	4	6	14	7
连通度 k 的分布	1	6	2	3	2	0	0	6
	2	2	2	0	2	2	1	9
	3	3	2	3	0	2	4	8
	4	1	2	0	2	0	1	0
	5	2	1	5	4	2	0	0
	6	0	2	3	0	0	0	0

5 总结

布尔网络的拓扑结构和函数作用方式直接影响网络的动态行为, 本文在 Post 类函数的基础上, 提出了一种具有正负调控机制的扩展的 Post 类函数, 并进而研究了这种扩展的 Post 类函数对网络动态行为及网络拓扑结构的影响. 首先, 新函数构成的网络与 Post 类函数和渠化函数具有相似的动态行为, 由它们构造的网络都具有维持网络有序行为的能力, 并且其长程相关性的相变过程也很类似. 由于在任意连通度 k 下, 扩展的 Post 类函数的数量最多. 这种函数类更有可能在自然进化过程被选择为基因之间调控的作用方式. 其次, 随着连通度 k 的增加, 三种有序函数占总函数的比例将急剧减少. 因此, 有序性假设正好与生物网络中节点入度的幂律分布的观察一致, 即只有少数关键节点的度很大. 这些度大的节点即使不是有序函数, 对于系统整体的有序性的影响也很小. 在八种真实生物网络中各种函数类型的分布, 也部分支持上述观点.

最后必须指出, 新扩展的 Post 类函数类与原 Post 类函数相比, 虽然具有不仅包含了正、负调控机制, 而且在数量上几乎是后者的二倍的优点. 但是, 负调控机制的引入, 也破坏了原 Post 类函数的封闭性. 由于在布尔网络模型中, 吸引域的半径通常都不大, 因此系统在一个微小的扰动下很快就会回到吸引子状态. 失去封闭性对系统稳定性的影响不会太大, 渠化函数和多数函数也均不满足封闭性.

参考文献

- [1] I 什穆列维奇, 等. 基因组信号处理[M]. 刘文斌, 高琳,

- 译.北京:科学出版社,2010.
Shmulevich I, et al. Genomic Signal Processing [M]. Liu Wenbin, Gao Lin. Beijing: Science Press, 2010. (in Chinese)
- [2] WANG X, LIU Q, CHENG Y, LI L. Qualitative analysis of gene regulatory networks based on angular discretization [J]. Chinese Journal of Electronics, 2011, 20: 646 – 650.
- [3] Shmulevich I, Lähdesmäki H, Dougherty E R, Astola J, et al. The role of certain post classes in Boolean network models of genetic networks [J]. Proceedings of the National Academy of Sciences, 2003, 100: 10734 – 10739.
- [4] Guan B, Zan X, Xiao B, Ma R, et al. Detecting dense subgraphs in complex networks based on edge density coefficient [J]. Chinese Journal of Electronics, 2013, 22: 517 – 520.
- [5] Jeong H, Tombor B, Albert R, Oltvai Z N, et al. The large-scale organization of metabolic networks [J]. Nature, 2000, 407: 651 – 654.
- [6] Kauffman S. The Origins of Order: Self Organization and Selection in Evolution [M]. Oxford University Press, 1993.
- [7] Harris S E, Sawhill B K, Wuensche A, Kauffman S. A model of transcriptional regulatory networks based on biases in the observed regulation rules [J]. Complexity, 2002, 7: 23 – 40.
- [8] Mendoza L, Thieffry D, Alvarez-Buylla E R. Genetic control of flower morphogenesis in Arabidopsis thaliana: a logical analysis [J]. Bioinformatics, 1999, 15: 593 – 606.
- [9] Penzel T, Kantelhardt J W, Grote L, Peter J-H, et al. Comparison of detrended fluctuation analysis and spectral analysis for heart rate variability in sleep and sleep apnea [J]. IEEE Transactions on Biomedical Engineering, 2003, 50: 1143 – 1151.
- [10] Amaral L A, Díaz-Guilera A, Moreira A A, Goldberger A L, et al. Emergence of complex dynamics in a simple model of signaling networks [J]. Proceedings of the National Academy of Sciences of the United States of America, 2004, 101: 15551 – 15555.
- [11] Goldberger A L, Amaral L A, Hausdorff J M, Ivanov P C, et al. Fractal dynamics in physiology: alterations with disease and aging [J]. Proceedings of the National Academy of Sciences of the United States of America, 2002, 99: 2466 – 2472.
- [12] Post E L. The Two-Valued Iterative Systems of Mathematical Logic (AM – 5) [M]. Princeton University Press, 1941.
- [13] Post E L. Introduction to a general theory of elementary propositions [J]. American Journal of Mathematics, 1921, 43: 163 – 185.
- [14] Derrida B, Pomeau Y. Random networks of automata: a simple annealed approximation [J]. EPL (Europhysics Letters), 1986, 1: 45.
- [15] Fauré A, Thieffry D. Logical modelling of cell cycle control in eukaryotes: a comparative study [J]. Molecular BioSystems, 2009, 5: 1569 – 1581.
- [16] Sánchez Rodríguez L, Chaouiya C, Thieffry D. Segmenting the fly embryo: logical analysis of the role of the segment polarity cross-regulatory module [J]. International Journal of Developmental Biology, 2008, 52: 1059 – 1075.
- [17] Davidich M I, Bornholdt S. Boolean network model predicts cell cycle sequence of fission yeast [J]. PLoS One, 2008, 3: e1672.
- [18] Davidich M, Bornholdt S. The transition from differential equations to Boolean networks: a case study in simplifying a regulatory network model [J]. Journal of Theoretical Biology, 2008, 255: 269 – 277.
- [19] Zhang S-Q, Ching W-K, Ng M K, Akutsu T. Simulation study in probabilistic Boolean network models for genetic regulatory networks [J]. International Journal of Data Mining and Bioinformatics, 2007, 1: 217 – 240.

作者简介



许 鹏 男, 博士研究生, 1986 年出生于江苏泗洪, 2013 年 6 月获温州大学硕士学位, 2013 年 9 月进入东南大学攻读博士学位, 主要研究方向为生物信息学、数据挖掘。
E-mail: dpzc765@163.com



刘文斌 男, 博士, 教授, 1969 年出生于陕西韩城。2004 年获华中科技大学博士学位, 目前感兴趣的研究领域为生物信息学、数据挖掘、DNA 计算等。获得省部级奖励 4 项, 主持国家省部级项目 6 项, 发表学术论文 40 余篇。
E-mail: wbliu6910@126.com