

新型细胞自动机规则空间的参数化^{*})

曹兴芹^{1,2} 王能超¹

(华中科技大学计算机学院 武汉 430074)¹ (新疆师范大学数理信息学院 乌鲁木齐 830054)²

摘要 细胞自动机在复杂性科学研究中占有重要地位,规则空间的规律性是细胞自动机理论研究的核心问题。本文首次引入沃尔什正交变换,结合计算机实验,提出了刻画细胞自动机规则空间规律性的 δ 参数。通过对所有初等细胞自动机的分析,并与朗顿 λ 参数的对比研究,结果表明该参数能更好地刻画细胞自动机规则空间中从有序到混沌的相变,不存在关于相变临界值的争论。

关键词 细胞自动机,规则空间,相变,参数,沃尔什正交变换

A Novel Parameterization for the Space of Cellular Automata Rules

CAO Xing-Qin^{1,2} WANG Neng-Chao¹

(College of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074)¹

(School of Math-physics and Information Science, Xinjiang Normal University, Urumqi 830054)²

Abstract Seeking for regularity of the cellular automata rule space is the center issues for cellular automata theoretical research. By introducing Walsh orthogonal transform and analyzing Walsh spectrum, the elementary cellular automata rules space is investigated, and a parameter δ is proposed and compared with Langton's λ parameter. The results show that the parameter δ is better to character the cellular automata rule space and describe its phase transitions from order to chaos, and no polemics were existed on the critical value of δ .

Keywords Cellular automata, Rules space, Phase transition, Parameter, Walsh orthogonal transform

1 引言

细胞自动机(cellular automata,简记CA)以其结构简单、内涵丰富等特征成为复杂性科学研究的主要工具^[1]。CA的应用已经是近年来国内外研究的热点,然而其理论基础十分薄弱,它们之间的巨大差距大大限制了CA模型的广泛应用,并制约了复杂性研究的深入开展^[2]。

目前关于CA的理论研究国内还没有见到任何相关的报道,国外进展也十分缓慢,主要原因在于作为时空扩展的动力学模型,CA空间的全离散特征难以用经典理论来处理^[1]。

CA理论研究一般限定在线性CA中^[2],而对有巨大应用潜力、占90%以上的非线性CA,迄今为止最好的结果依然是20世纪80年代朗顿(C. G. Langton)的 λ 参数理论^[2,3]。但该理论存在两个关键问题:(1)关于CA空间从有序到混沌的相变点,即临界值 λ 。至今依然存在争论^[4,5];(2)CA规则空间并不严格遵守 λ 参数规律性^[3],当规则空间较小的时候,表现尤为突出^[6]。正因为如此,就连朗顿自己也认为 λ 参数可能不是刻画CA规则空间的最好参数^[3]。

初等细胞自动机(Elementary cellular automata, ECA)是元胞状态数 $k=2$ 、邻域半径 $r=1$ 的一维CA,共有256(k^{2r+1})个规则,本质上包含所有类型的CA,是CA理论研究的典范^[1]。本文采用计算机实验和现代数学方法相结合,引入沃尔什正交变换对ECA规则空间进行分析。在考察线性CA频谱特征及演化行为关系的基础之上,根据其统计性质和函数逼近思想,寻找CA规则空间的规律性,力求更好地参

数化CA规则空间,刻画规则空间从有序到混沌的相变,为有效地搜索适合特定应用问题的CA奠定基础。

2 CA规则空间与沃尔什正交变换

CA是一系列有限自动机在规则空间上的均匀排列。称元胞在空间排列的维数为CA的维数,记为 d 。用 k, r 分别表示元胞状态数和邻域半径。根据维数 d 和邻域半径 r 定义邻域大小,记为 N_r ,因此局部演化规则是有限域 $GF_2^{N_r}$ 上的函数。可能的规则有 k^{N_r} 个,若定义一个距离,则它们组成CA的规则空间。称 $k=2$ 的CA为二值CA,设 GF_2 表示二元域,则二值CA的局部映射 f 可表示为

$$f: GF_2^{N_r} \rightarrow GF_2 \quad (1)$$

令 $N=2^{N_r}$,将 f 的函数值按自变量的字典顺序从小到大排列成一 N 维向量

$$(f(0), f(1), \dots, f(N-1)) \quad (2)$$

它是一个 N 维矢量,与 GF_2 上的 N_r 元布尔函数 $f(x)$ 一一对应,称其为二值CA的向量表示。沃尔什变换是研究布尔函数的重要工具,许多特征都可用沃尔什谱表示。CA规则 f 的沃尔什变换为

$$S_{(f)}(k) = \frac{1}{N} \sum_{j=0}^{N-1} (-1)^{f(j)} W_{jk} \quad (3)$$

其中 W_{jk} 是 N 阶沃尔什矩阵,它有多种排序方式。本文采用沃尔什序,称 $S_f(k)$ 为CA的沃尔什谱,它满足如下公式:

$$\sum_{k=0}^{N-1} S_{(f)}^2(k) = 1 \quad (4)$$

^{*})国家自然科学基金资助项目(编号:60473015)。曹兴芹 博士生,主要研究方向为细胞自动机、复杂系统;王能超 教授,博士生导师,主要研究方向为计算数学、演化数学。

因此, $S_{i0}^2(k)$ ($k=0 \sim N-1$) 可看作 f 的沃尔什谱密度, 反映了 f 与沃尔什线性函数的相似程度, 简称 $S_{i0}^2(k)$ ($k=0 \sim N-1$) 为 CA 的谱密度。

ECA 规则可用一个二进制数或者十进制数表示^[1], 它们与式(2)给出的向量表示一一对应, 例如, 90 号 ECA 的向量表示为 (0,1,0,1,1,0,1,0)

3 线性 CA 的特征与 δ 参数

当局部映射 f 是线性函数时, 称对应的 CA 为线性(或加法)CA。显然, 这样的 CA 共有 $k^{2^{r+1}}$ 个。对于二值 CA, 根据对称性思想, 应还有 $k^{2^{r+1}}$ 个 CA 与上述线性 CA 具有同样的

演化性质, 本文把它们通称为线性 CA, 故在 ECA 中共有 16 个。通过计算机实验, 在随机初始状态(初始密度 $\rho_0 = 0.5$) 下, 不同的线性 ECA 规则演化结果如图 1 所示。容易发现, 线性 CA 的演化行为要么具有很强的周期性, 要么具有明显的自相似结构, 不存在复杂型和局部混沌型等难以分类的 CA。而且在不同的初始状态下, 演化特征相对稳定。同时, 结果发现其中 8 个具有典型的周期结构, 而另外 8 个具有非周期行为。关于进一步的描述, 本文采用前人对沃尔弗拉姆修改的分类方案^[6], 具体分为幂零型、固定点型、周期型、局部混沌型、混沌型和复杂型共 6 类, 线性 CA 的演化行为仅仅只有除局部混沌和复杂型外的其它 4 种类型, 参见表 1。

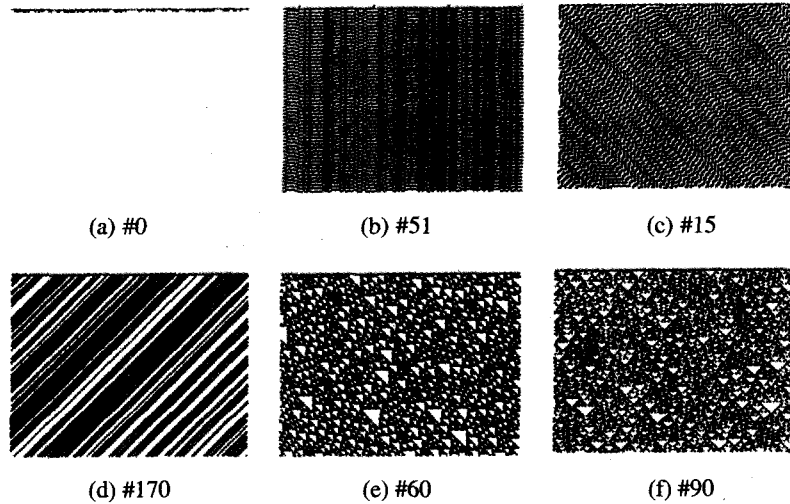


图 1 线性 CA 的演化行为

本文考察了线性 CA 的沃尔什谱特征。按式(3)计算线性 CA 规则的沃尔什谱, 结果发现它们具有单一不为零的谱点, 其最大区别在于谱点的位置不同, 见表 1。这说明不同类型的 CA 规则相似于不同的沃尔什函数。

表 1 线性 CA 的沃尔什谱及演化特征

序号	CA	频谱 $S_{i0}^2(k), k=0, 1, \dots, 7$								演化性态
		0	1	2	3	4	5	6	7	
1	0	1								幂零型
2	255	-1								周期型
3	15		-1							周期型
4	240		1							固定点型
5	51			-1						周期型
6	204			1						固定点型
7	85					-1				周期型
8	170					1				固定点型
9	60		1							非周期行为
10	195		-1							自相似结构
11	102				1					非周期行为
12	153				-1					自相似结构
13	105					-1				非周期行为
14	150					1				非周期行为
15	90						1			非周期行为
16	165						-1			非周期行为

注: 若表中谱系数未给出, 则其值为零。

为了叙述方便, 本文把 8 个周期行为的 CA 对应的沃尔什谱不为零的谱点称为周期谱点, 而 8 个产生非周期行为的 CA 对应的沃尔什谱不为零的谱点称为非周期谱点, 即周期

谱点是 0、1、3 和 7, 非周期谱点是 2、4、5 和 6。特别地, 称序号为 0 的周期谱点为幂零型周期谱点。进一步分析发现, CA 的周期谱点分布于 $2^i - 1$ ($i=0 \sim 2r+1$) 位点上, 而且当邻域半径 $r > 1$ 时, 这些位点上具有单一不为零谱值对应的线性 CA 的演化行为依然具有明显的周期性, 因此这些位点依然是 CA 的周期谱点, 故定义 δ 参数为

$$\delta(f) = 1 - \sum_{i=0}^{2r+1} S_{i0}^2(2^i - 1) \quad (5)$$

它反映了细胞自动机演化行为的非周期性。显然, 线性 CA 中, 产生周期行为的 CA 有 δ 等于零, 而非周期 CA 的 δ 值为 1。

4 参数化 ECA 规则空间

首先, 本文通过计算机实验考察 ECA 的演化行为, 并根据前述 6 类分类方案进行分类。结果表明, 与线性 CA 相比, 除了周期 CA 和混沌型非周期 CA 外, 还有一些 CA 是周期与非周期行为的有机组合(参见图 2), 它们被称为局部混沌型或复杂型。其次, 本文通过计算所有 ECA 规则(256 个)的沃尔什谱密度, 进而分类统计不同类型 CA 的谱密度分布, 结果发现: 不同类型的 CA 在谱点分布上明显不同(参见表 2), 其中前 3 类 CA 在周期型谱点上的值较其它位点高, 而后 3 类 CA 的谱密度更集中于非周期型谱点上; 在周期 CA 中, 幂零型 CA 与另外两类相比, 谱值更加集中于幂零型周期位点; 与局部混沌和混沌型相比, 复杂型 CA 的谱密度分布明显不同, 局部混沌和混沌型 CA 的谱密度分布更加集中于非周期位点, 而复杂型 CA 在第 2、4 位非周期位点上的谱值明显高于其它位点, 在第 5 位非周期位点和第 3 位周期位点上的谱值次高。

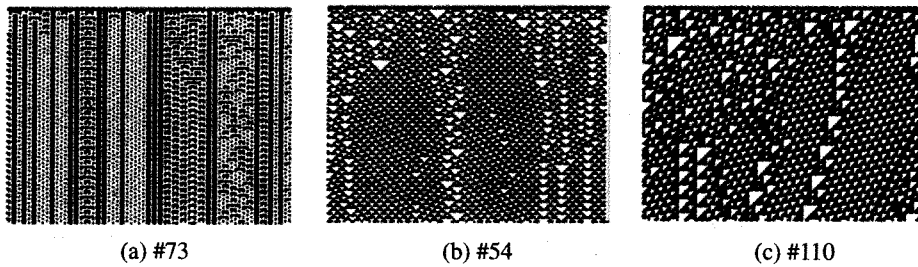


图2 局部混沌型和复杂型 CA 的演化行为

表2 ECA 的分类平均沃尔什谱密度

类别	规则数	平均沃尔什谱密度 $S_{f_j}^2(k)$							
		0	1	2	3	4	5	6	7
1	24	0.3854	0.1354	0.0729	0.0729	0.0729	0.0729	0.0521	0.1354
2	107	0.1285	0.1379	0.1238	0.1519	0.1238	0.0864	0.1098	0.1379
3	79	0.0886	0.1392	0.1266	0.1519	0.1266	0.1266	0.1013	0.1392
4	10	0.0375	0.0875	0.0875	0.0375	0.0875	0.2375	0.3375	0.0875
5	30	0.0458	0.0625	0.1458	0.0292	0.1458	0.2625	0.2458	0.0625
6	6	0.0417	0.0417	0.2917	0.1250	0.2917	0.1250	0.0417	0.0417

注:类别1~6分别表示幂零型、固定点型、周期型、局部混沌型、混沌型和复杂型,下同。

上述结果表明,CA 演化规则的沃尔什谱密度分布与对

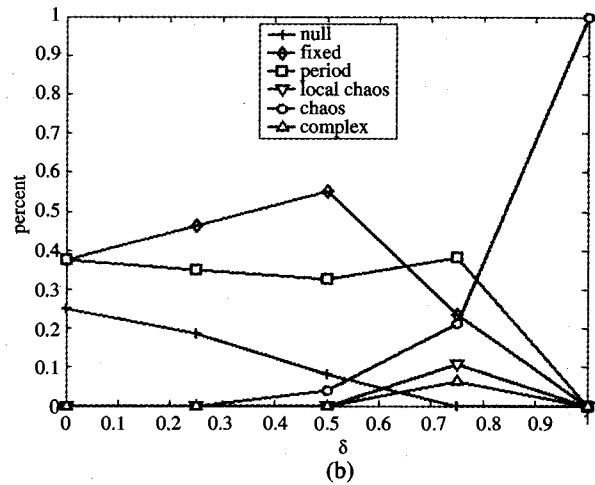
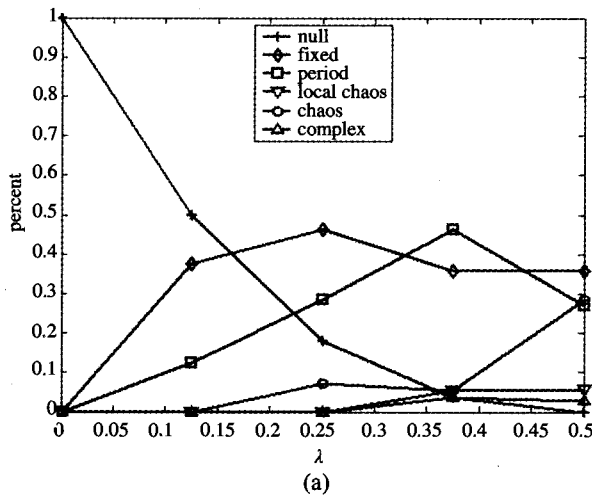


图3 ECA 规则空间中不同类型 CA 所占的百分比随(a) λ 参数和(b) δ 参数的变化

同时本文计算了 δ 参数、 λ 参数和分类之间的相关性, 结果发现 δ 参数与分类之间的相关系数比 λ 参数与分类之间的相关系数要高, 见表 3。这进一步说明 δ 参数能更好地刻画 CA 规则空间的规律性。

表3 CA 类型与 λ 参数、 δ 参数之间的相关性

相关系数	类型	λ	δ
类型	1.0000	0.4241	0.5546
λ	0.4241	1.0000	0.3496
δ	0.5546	0.3496	1.0000

5 讨论

朗顿的 λ 参数提出近 20 年来, 一直被认为是反映 CA 规则空间规律性的最重要参数^[2,4,5,7,8], 被评价为可以与统计物

应的演化行为之间具有很强的一致性, 即该谱密度分布反映了 CA 演化行为的周期性特征, 故据此定义了 δ 参数。本文进一步考察 ECA 规则空间中不同类型 CA 所占的百分比随 λ 参数和 δ 参数的变化情况, 如图 3a 和 b 所示。结果表明: 首先, 在整体趋势上 δ 参数具有 λ 参数所具备的性质, 即随着 δ 的增加, 周期型 CA 逐渐减少, 非周期型 CA 越来越多, 而且 $\delta < 0.5$ 时 CA 都属周期型, $\delta = 1$ 时 CA 都属混沌型; 其次, 混沌型 CA 除了 18 号和 183 号的 δ 值为 0.5 以外, 其 δ 值均大于或等于 0.75, 而且复杂型和局部混沌型 CA 全部集中在一个点上, 即 $\delta = 0.75$, 说明不同类型的 CA 在 δ 参数下的分布比在 λ 参数下更加集中。因此, 与 λ 参数相比, δ 参数能更好地刻画 CA 规则空间从有序到混沌的相变过程, 反映 CA 的本质特征。

理学中的温度或者动力系统非线性度相比^[2]。 λ 参数关于 CA 规则空间的相变理论认为, 随着 λ 参数从零变化到最大值, CA 规则空间将发生类似于物质从固态到液态的相变, 即 CA 的动力学从具有高度有序过渡到高度无序, 而复杂型 CA 则位于从有序到无序的临界点上^[3]。按照相变理论, 如果能够找到这个类似于热力学极限的 λ 参数的临界值 λ_c , 那就可以找到具有通用计算能力的复杂型 CA。然而, 进展并不顺利, 实验发现 λ_c 具有很大的变化性^[4,5]。

过去经典数学方法在分析 CA 中取得了一些重要成果, 然而它也把研究对象严格限定在一定范围内, 即线性 CA^[2]。换一个角度, 本文对 CA 规则进行沃尔什谱分析, 结合计算机实验, 从线性 CA 出发分析了 CA 的沃尔什谱特征, 进而提出

(下转第 164 页)

当用户输入参数为聚类数的时候,根据图 4,通过以下算法进行分裂聚类:

输入:整体相关连通图;聚类数 k
 输出: k 个相关连通子图
 步骤:
 While (连通子图数目 < 聚类数 k)
 {
 选择相关度最小的连接;
 断裂该连接;
 }

4 实验

我们在 Google 的基础上进行了多组不同类型关键词的自动聚类实验,均取得了令人满意的效果。例如,对关键词组 {李宇春,何洁,周笔畅,刘德华,张学友},{香蕉,电视,葡萄,衣服,冰箱,鞋子} 分别输入不同类型的参数进行聚类实验。

对每个关键词获取搜索引擎返回的前 200 条记录进行排名相关度以及主题词相关度的计算。相关度计算公式中取 $k_1=0.6, k_2=0.2, k_3=0.2$ 。

对关键词组 {李宇春,何洁,周笔畅,刘德华,张学友}, 我们输入相关度参数 $\lambda=0.08$, 得出的结果为:

类别	关键词	部分主题词
1	李宇春,何洁,周笔畅	超级,超女,演出,活力,女生
2	刘德华,张学友	江湖,香港,黎明,郭富城,天王

对关键词组 {香蕉,电视,葡萄,衣服,冰箱,鞋子}, 输入聚类数 $k=3$ 得出以下结果:

类别	关键词	主题词
1	香蕉,葡萄	果汁,橙汁,爾必思,金桔,木瓜,牛奶,哈密瓜,水果
2	电视,冰箱	行业,家电,价格,产品,洗衣机,空调
3	衣服,鞋子	店铺,服装,购物,女人,时尚,精品

实验结果表明,该方法能够有效地对关键词组进行自动聚类,不需要任何背景知识以及学习过程,可以方便地应用于其他计算机智能程序以及为用户提供了一些有趣的信息,使用户能够更进一步利用互联网中蕴含的丰富知识,为未来互联网的开发利用提供了一个突破口。

结论 当前人们利用互联网的信息主要是通过搜索引擎

进行信息检索,这种方式只是对存在于互联网某一网站或某一网页中的信息进行快速的查找与简单的重现,并没有发现互联网中隐含的知识和规律。本文把当前的搜索引擎称之为信息搜索引擎并提出了一个新的概念:知识搜索引擎。知识搜索引擎通过对互联网信息的分析挖掘,产生更高层次的知识,这些知识不是静态的存在于互联网的某处,而是信息提炼升华的结果。本文通过一个具体的应用“基于搜索引擎的关键词自动聚类”来对知识搜索引擎的技术以及应用进行探索。基于搜索引擎的关键词自动聚类方法分析搜索引擎返回的关键词相关网页的链接结构以及文本信息,发现关键词间隐含的联系从而对关键词实现智能自动分组。实验结果表明该方法具有良好的效果,能够为其它计算机程序提供智能的预处理过程以及为用户提供更丰富和更有趣的知识。该方法使得用户可以更进一步利用互联网信息,是一个全新的研究。

研究知识搜索引擎其他的应用,例如通过搜索引擎,对互联网中具体某领域进行分析挖掘,发现某种产品或事件间的发展趋势以及规律。

参考文献

- 1 Brin S, Page L. The Anatomy of a Large-Scale Hypertextual Web Search Engine
- 2 Liang Jiu-zhen. Chinese Web Page Classification Based on Self-Organizing Mapping Neural Networks. In: Proceedings of the Fifth International Conference on Computational Intelligence and Multimedia Applications (ICCI'03)
- 3 Shen Dou, Cong Yan, Sun Jian-ta, Lu W-chang. STUDIES ON CHINESE WEB PAGE CLASSIFICATION. In: Proceedings of the Second International Conference on Machine Learning and Cybernetics, Wan, November 2003
- 4 Yu H, Han Jiawei, Chang K C-C. PEBL: Web Page Classification without Negative Examples. IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, 2004, 16(1)
- 5 张伟. 基于 WWW 的聚类搜索引擎的研究: [2003 年全国优秀博士学位论文]
- 6 Zeng Hua-Jun, He Qi-Cai, Chen Zheng, Ma Wei-Ying. Learning To Cluster Search Results. In: The 27th Annual International ACM SIGIR Conference (SIGIR'2004), July 2004
- 7 Mobasher B, Cooley R, Jaideep S. Creating Adaptive Web Sites Through Usage-Based Clustering of URLs
- 8 Abawajy J H, Hu M. J. A new Internet meta-search engine and implementation. IEEE, 2005
- 9 Kawano H. Overview of Mondou web search engine using text mining and information visualizing technologies. IEEE, 2001

(上接第 147 页)

了刻画 CA 规则空间规律性的 δ 参数。本文结果表明,随着 δ 从 0 变化到 1, CA 规则空间发生从有序到混沌的相变。由于 $\delta=0.5$ 出现了混沌型 CA, 而复杂型和局部混沌型 CA 都出现在 $\delta=0.75$ 处, 可以推测 CA 规则空间的相变包含多种途径, 或者发生从有序到混沌的尖锐突变, 或者相变过程中存在临界规则, 这与朗顿等人提出的一阶相变和二阶相变一致; 一阶相变和二阶相变的临界点分别发生在 $\delta=0.5$ 和 0.75 处, 不存在关于临界值的争论, 表明 δ 参数比 λ 参数能更好地刻画 CA 规则空间规律性和反映 CA 的本质特征。从应用的角度, δ 参数使得从巨大的 CA 规则空间中搜索适合特定应用问题的 CA 变得有规可寻。本文的研究表明, 计算机实验和现代数学方法相结合研究 CA 是行之有效的办法。

当 r 值增大的时候, 沃尔什谱密度的分布特征有待于进行更广泛的计算机实验和分析; 对 $k>2$ 的 CA, 采用相同的基本思路, 扩展二值的沃尔什变换到多值正交变换中, 是值得进

一步研究的问题。

参考文献

- 1 Wolfram S. A new kind of science. USA ; Wolfram Media Inc, 2002
- 2 Niloy G, Biplab K S, Andreas D, et al. A survey on cellular automata; [Technical report]. Centre for High Performance Computing, Dresden University of Technology, 2003. 1~28
- 3 Langton C G. Computation at the edge of chaos: Phase transitions and emergent computation. Physica D, 1990, 42: 12~37
- 4 Packard N H. Adaptation towards the edges of chaos. In: Kelso J A S, Mandell A J, Shlesinger M F, eds. Dynamic patterns in complex systems. Singapore, World Scientific, 1988. 293~301
- 5 Mitchell M, Hraber P T, Crutchfield J P. Revisiting the edges of chaos ; Evolving cellular automata to perform computations. Complex Systems, 1993, 7(1): 89~130
- 6 Li Wentian, Packard N H, Langton C G. Transition phenomena in cellular automata rule space. Physica D, 1990, 45: 77~94
- 7 Li Wentian, Packard N H. The structure of the elementary cellular automata rule space. Complex Systems, 1990, 4(3): 281~298
- 8 Wootters W K, Langton C G. Is there a sharp phase transition for deterministic cellular automata? Physica D, 1990, 45: 95~104