

基于群体智能的网络信息自组织方式*

刘 燕 帅典勋 柴震川

(华东理工大学计算机科学系 上海200237) (清华大学智能技术与系统国家重点实验室 北京100084)

A Novel Self-Organizing Approach to Network Information Exploitation Based on Swarm Intelligence

LIU Yan SHUAI Dian-Xun CHAI Zhen-Chuan

(Department of Computer Science, East China University of Science and Technology, Shanghai 200237)

(State Key Laboratory of Intelligence Technology and System, Tsinghua University, Beijing 100084)

Abstract We design and implement a novel information self-organization model based on Generalized Cellular Automata (GCA), to accomplish network information content self-organization employing the idea of swarm intelligence. Through constructing correspondent cell rules and the mapping of complex network environment to our GCA, reasonable distribution of network information from different information sources can be achieved on different nodes according to dynamic variation of local network circumstances. Simulation experiment results show many advantages of our methodology over present approaches in terms of efficiency, adaptability, reliability, and easy hardware implementation.

Keywords Information exploitation, Generalized cellular automata, Swarm intelligence, Self-organization

1 引言

网络环境下的信息利用指人们把网络上的信息资源转换为知识的过程。基于内容的信息利用的当前模式主要有传统信息检索、数据挖掘、知识发现和基于 MAS(多智能 Agent 系统)的网络分布式计算模型等。通过传统信息检索、分析、传输和集成而得到的,一般是信息浅层加工的结果,是基于客户主动请求、服务器消极被动服务的模式。通过数据挖掘和知识发现^[1,2]所得到的,是面向特定数据集中隐含的规则或知识,在不同的信息来源如数据库、站点之间,不同的用户之间是孤立的,没有任何机制去重新组合不同信息源的知识,为不同用户产生新的知识结构。而在基于 MAS 的新一代网络分布式计算模型^[3]的研究中,主要是通过大量理性化的 Agent 间的协调和对环境的适应来实现 Agent 社会的协作,但没有提供实用的基于社会智能的分布式问题求解模型和方法。这些方法大多是从传统信息处理中借用和移植过来的,没有考虑复杂的网络环境,不同信息利用行为之间基本上是相互无关的,信息利用者必须自己确定其网络信息利用途径。而网络环境下的信息利用是在整个网络环境下进行的,其处理的对象大多是分布的、动态的和非结构化的信息,因而它的利用与开发要求更高层次的、更接近应用的和更具智能化的方法,是网络信息资源更充分有效的利用形态。

CA(Cellular Automaton, CA)近来在网络行为的研究中得到了广泛的应用^[4~6],成为研究复杂系统的重要工具。其结构是由许多基本单元——细胞(cell)组成的,每个细胞的状态有限,随时根据相邻细胞的状态并行改变自己的状态。CA 是一个完全分布的离散的动力学系统,利于计算机模拟实现。

本文提出基于广义细胞自动机(Generalized Cellular Automaton, GCA)群体智能的网络信息内容自组织的模型和方法。我们把复杂网络环境映射到广义细胞自动机上,通过信息利用行为群体智能形成网络信息按内容的自组织。模式对不同网络节点上来自不同信息源的信息实现合理的分布与组合,从而达到整个网络的信息动态合理利用。通过仿真实验,我们得到面向网络信息利用的广义细胞自动机状态随时间变化的情况,并对网络信息自组织前后的信息分布情况以及信息利用的平均响应效率的变化情况进行比较和分析。实验结果表明,我们提出的基于广义细胞自动机的网络信息自组织模型,可以改善信息分布,大大提高信息利用的平均响应效率。

2 基于广义细胞自动机的信息自组织模型

2.1 基于网络信息自组织的信息利用模式的概念及定义

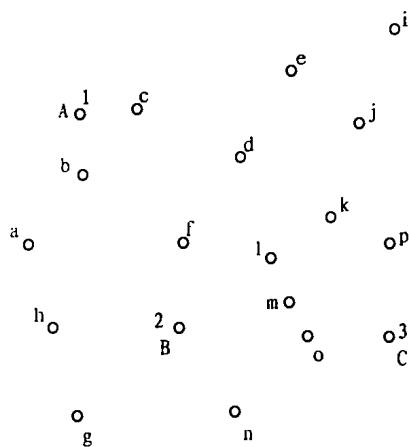
基于网络信息自组织的信息利用模式的简单例子如图1(a)和图1(b)所示。图1(a)表示自组织前的信息分布结构,它涉及16个信息利用地址 $a \sim p$, 3个信息源地址 $A \sim C$ 分别提供三种信息内容1~3。而图1(b)表示网络提供的关于信息利用行为 $\epsilon_i(t)$ 的自组织结构图,它由 t 时刻之前网络中的信息利用行为自组织产生,可对每种信息内容提供多个信源地址。其信息源之间形成了分布式关联,提高了信息利用行为的可靠性和容错性。

用于网络信息自组织的广义细胞自动机由细胞阵列 GCA 组成,每个细胞有自己的细胞核、细胞状态、细胞输入和

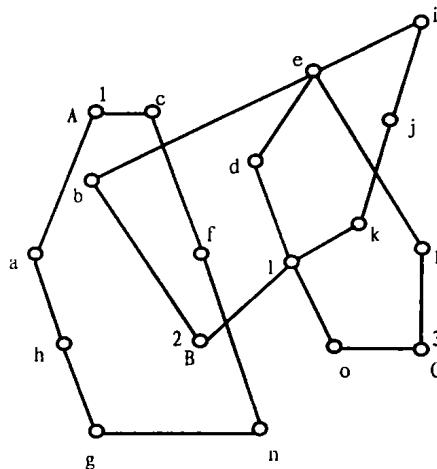
* 本课题得到国家“九七三”重点基础研究发展规划项目(编号 G1999032307)、国家自然科学基金重点项目(编号 60135010)、国家自然科学基金项目(编号 60073008)以及清华大学智能技术和系统国家重点实验室开放课题基金的资助和支持。刘 燕 硕士研究生,研究方向为分布式人工智能。帅典勋 博士,教授,博士生导师,目前从事人工智能,知识工程,分布并行处理等方面的研究。柴震川 硕士研究生,研究方向为分布式人工智能。

细胞输出。细胞阵列中的所有细胞状态都遵循相同的细胞动力学规则变化。我们的目的是,广义细胞自动机中大量细胞的

并行演化结果,体现出广义细胞自动机的群体智能,达到网络自组织信息利用。



(a) 用户指定的一种信息利用行为, 它为 3 种信息内容的 16 个信息利用目的地址指定 3 个信息源地址。



(b) 网络提供的信息自组织结构, 其中 3 个由直线组成的环分别为 3 种信息内容的多个信源地址。它由 t 时刻之前网络中的信息利用行为自组织产生, 可由多个信源中选择代价函数最小的信源提供信息内容。

图1 基于计算机网络信息自组织的信息利用模式的概念(图(a)(b))

定义1 在时刻 t , 由第 i 个网络信息利用者指定的信息利用行为 $\epsilon_i(t)$ (exploitation) 为五元组, 即 $\epsilon_i(t) = (\omega_{ij}(t), \varphi_{ij}(t), \psi_{ij}(t), d_{ij}(t), t)$, 其中 $\omega_{ij}(t) \in \Omega$, $\varphi_{ij}(t) \in \Phi$, $\psi_{ij}(t) \in \Psi$, $d_{ij}(t) \in D$, $t \in T$; 其中 Ω 为表示网络信息内容的特征向量的集合, Φ 为信息源地址的集合, Ψ 为信息利用目的地址的集合, $d_{ij}(t)$ 为行为 $\epsilon_i(t)$ 的信息扩散地址, 初始值为信息利用目的地址 $\Psi_{ij}(t)$, t 为建立信息利用行为的时间, θ_i 为判断信息利用行为是否过期的时间阈值。

定义2 信息利用行为 $\epsilon_i(t)$ 的代价函数 $u_i(t)$ 定义为 $u_i(t) = \sum_{j=1}^{l_i} l_j q_j$, 其中 l_j 表示从信息源 $\varphi_{ij}(t)$ 到信息目标 $\Psi_{ij}(t)$ 传输信息内容 $\omega_{ij}(t)$ 的距离(或链路数量), q_j 为相应的传输数据包数量, 记 $l_j = \|\varphi_{ij}(t) - \psi_{ij}(t)\|$ 。

定义3 若信息地址 $a \in \Phi$ 中有信息内容 $\omega \in \Omega$, 记为 $\omega < a$ 。对于给定的 $\omega \in \Omega$, Φ 中含有 ω 的地址集合记为: $A(\omega) = \{a | \omega < a, a \in \Phi\} \subseteq 2^\Phi$ 。

定义4 广义细胞自动机细胞阵列 GCA 由 $N \times N$ 个细胞组成。定义细胞 C_{ij} 的邻域 $N_{ij} = \{c_{(i-1)j}, c_{(i+1)j}, c_{i(j-1)}, c_{i(j+1)}\}$ 。细胞 c_{ij} 在时刻 t 的细胞状态 $S_{ij}(t)$, 细胞核 $K_{ij}(t)$ 分别定义如下:

$$S_{ij}(t) \in \{0, -1, +1\};$$

$$K_{ij}(t) = (\omega_{ij}(t), \varphi_{ij}(t), \psi_{ij}(t), d_{ij}(t), t), \omega_{ij}(t) \in \Omega, \varphi_{ij}(t) \in \Phi, \psi_{ij}(t) \in \Psi, d_{ij}(t) \in D, t \in T;$$

定义 u_{ij} 和 v_{ij} 分别为细胞 c_{ij} 的细胞核与邻域 N_{ij} 中其他细胞的细胞核之间的信息内容距离及物理地址距离, θ_u 和 θ_v 分别为它们的阈值。

定义5 细胞的收缩过程 $\mathcal{S}$ 。对于细胞 c_{ij} , 若 $\exists c_{r,f} \in N_{ij}, u_{r,f} \leq \theta_u, v_{r,f} < \theta_v$, 细胞 c_{ij} 进行收缩, 并计算新的信息扩散地址 $d_{ij}(t+1)$, 使得

$$\|d_{ij}(t+1) - \Psi_{ij}(t)\| \leq \|\varphi_{ij}(t) - \Psi_{ij}(t)\|, \text{ 且 } \|d_{ij}(t+1) -$$

$$\Psi_{r,f}(t)\| \leq \|\varphi_{r,f}(t) - \Psi_{r,f}(t)\|$$

同时细胞 $c_{r,f}$ 死亡, 即置细胞 $c_{r,f}$ 状态为 0。

定义6 自组织信息利用模式信息扩散过程 $\mathcal{D}$: 在 GCA 达到平稳分布后, 对于所有的非空细胞 c_{ij} , 令信息内容 $\omega_{ij}(t)$ 的扩散地址 $\rho_{ij}(t)$ 为 $\rho_{ij}(t) = d_{ij}(t)$, 并将通过细胞的收缩过程 $\mathcal{S}$ 得到的 $d_{ij}(t)$ 加入到 $A(\omega_{ij}(t))$ 之中, 即令 $\omega_{ij}(t) < d_{ij}(t)$ 。

定义7 自组织信息利用模式源址选择过程 $\mathcal{K}$: 对于网络信息利用行为 $\epsilon_i(t)$, 源址选择过程 $\mathcal{K}$ 确定从集合 $A(\omega_{ij}(t))$ 中选择信息源地址的概率, 使得代价函数 $u'_i(t)$ 满足以下条件:

$$u'_i(t) = \sum_{j=1}^{l'_i} l'_j q_j \leq u_i(t) = \sum_{j=1}^{l_i} l_j q_j \quad (1)$$

其中, $l'_j = \sum_{a_r \in A(\omega_{ij}(t))} p_r \|a_r - \Psi_{ij}(t)\|$ (2) p_r 表示选择 a_r 作为信息内容 $\omega_{ij}(t)$ 的信息源的概率, $A(\omega_{ij}(t))$ 由 t 时刻之前网络信息自组织的信息扩散过程 $\mathcal{D}$ 形成。

注释1:

(a) 简单起见, 在细胞收缩过程中, 我们计算新的信息扩散地址 $d_{ij}(t+1)$ 时, 可选择目的地址 $\Psi_{ij}(t)$ 与 $\Psi_{r,f}(t)$ 的中点。

(b) 为了简化起见, 同时不失一般性, 我们可选择概率

$$p_r = \begin{cases} 1 & \text{if } \|a_r - \Psi_{ij}(t)\| = \min_{a_k \in A(\omega_{ij}(t))} \|a_k - \Psi_{ij}(t)\| \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

(c) 源址选择过程 $\mathcal{K}$ 在按式(1), (2)决定关于 $\epsilon_i(t)$ 中信息 $\omega_{ij}(t)$ 的源址选择概率 p_r 时, 可以考虑源址集合 $A(\omega_{ij}(t))$ 中信息传输路径的拥塞情况及代谢情况。

(d) 网络信息利用者指定的信息利用行为 $\epsilon_i(t)$, 对于信息内容 $\omega_{ij}(t)$, 通常只指定一个信源 $\varphi_{ij}(t)$ 。但在我们的信息利用模式中, 网络将对 $\epsilon_i(t)$ 中的 $\omega_{ij}(t)$ 提供通过自组织形成的一个信源集合 $A(\omega_{ij}(t))$, 而且 $A(\omega_{ij}(t))$ 及从中选择信源的过程对提出 $\epsilon_i(t)$ 的信息利用者是完全透明的, 不被意识到

的网络在响应 $\epsilon_i(t)$ 之后,对原来的 $A(\omega_{ij}(t))$ 动态更新和扩充。

2.2 网络到 CA 的映射

由于信息利用 GCA 与实际网络构造有很大的不同,我们必须把实际网络参数通过某种方式映射到我们的 GCA 上。本文在信息自组织利用的信息扩散过程中,将信息内容相近且物理地址也相近的细胞核依据规则形成自组织,并算出信息扩散节点的位置,因此我们需要将网络从信息内容和 IP 物理地址两方面同时映射到 GCA 上。

(a)根据信息内容的处理 显然,信息自组织过程必须对不同信息内容得到不同结果。简单起见,我们将信息利用行为的内容特征向量用关键字分类,在我们的 GCA 中又用数字 1~100 代表 100 类不同的信息。若 $u_{i,j} < \theta_u$, 我们说 c_{ij} , $c_{i',j'}$ 信息内容接近。在自组织的过程中,信息内容接近的细胞核将以较大的概率收缩形成自组织。

(b)根据物理地址的处理 由于 GCA 中相邻的细胞核仅仅代表在 CA 上的位置接近,并不代表它们的 IP 物理地址接近,而且我们希望在自组织的过程中由相近地址的细胞核收缩形成信息扩散地址,因此同时必须注意收缩的细胞核 IP 地址必须接近。本模型中,定义每个节点物理地址范围为 x 轴 1~256, y 轴 1~256,在系统初始化时随机给出。若 $v_{i,j} < \theta_v$, 我们说 c_{ij} , $c_{i',j'}$ 物理地址接近。同样,信息内容相同而物理地址接近的细胞核将以较大的概率进行收缩。

2.3 GCA 状态转移规则

广义细胞自动机的细胞阵列 GCA 中,每个细胞的细胞动力学的规则形式为:

[t 时刻的细胞状态, t 时刻的细胞核] $\xrightarrow[\text{时刻 } t \text{ 采取的行动}]{\text{时刻 } t \text{ 的输入条件}}$ [(t+1)时刻的细胞状态, (t+1)时刻的细胞核]

细胞动力学规则中的细胞状态转移图如图 2 所示。

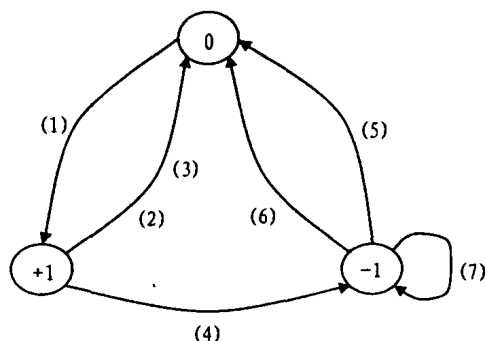


图2 细胞动力学规划中的细胞状态转移图
(弧线旁的数字为细胞动力学规则序号)

对于 GCA 中任一细胞 c_{ij} , $i, j=1, \dots$, 其细胞动力学规则 R 定义如下:

- (1) $[0, \emptyset] \xrightarrow[\text{从环境接受 } K_{ij}]{t=0 \text{ 初始化}} [+1, K_{ij}]$
- (2) $[+1, k_{ij}] \xrightarrow[\text{随机选择 } c_{i',j'}, K_{i',j'} \text{ 传往 } c_{i',j'}]{\exists c_{i',j'} \in N_{ij}, u_{i',j'} > \theta_u, v_{i',j'} > \theta_v} [0, \emptyset]$
- (3) $[+1, K_{ij}] \xrightarrow{t_{ij} > \theta_t} [0, \emptyset]$
- (4) $[+1, K_{ij}] \xrightarrow[\text{收缩 } c_{i',j'}]{\exists c_{i',j'} \in N_{ij}, u_{i',j'} < \theta_u, v_{i',j'} < \theta_v} [-1, K_{ij}]$
- (5) $[-1, K_{ij}] \xrightarrow[\text{随机选择 } c_{i',j'}, K_{i',j'} \text{ 传往 } c_{i',j'}]{\exists c_{i',j'} \in N_{ij}, u_{i',j'} > \theta_u, v_{i',j'} > \theta_v} [0, \emptyset]$
- (6) $[-1, K_{ij}] \xrightarrow{t_{ij} > \theta_t} [0, \emptyset]$

$$(7) [-1, K_{ij}] \xrightarrow[\text{收缩 } c_{i',j'}]{\exists c_{i',j'} \in N_{ij}, u_{i',j'} < \theta_u, v_{i',j'} < \theta_v} [-1, K_{ij}]$$

注释2:

(a)GCA 细胞状态转移规则(2)和(5)的目的是,如果细胞 c_{ij} 上的细胞核信息内容特征向量或物理地址与邻域(上、下、左、右)内的细胞核信息内容或物理地址差值大于一定的阈值, c_{ij} 随机选择一个 CA 上的新位置 $c_{i',j'}$, 并选择让细胞核 K_{ij} 跳转到新位置 $c_{i',j'}$ 。

(b)GCA 细胞状态转移规则(4)和(7)的目的是,如果细胞 c_{ij} 上的细胞核内容特征向量及地址与邻域内的细胞核差值小于阈值,细胞 c_{ij} 进行收缩,并计算新的信息扩散地址。

2.4 网络信息自组织广义细胞自动机并行算法 GAA

并行步1. GCA 中所有细胞的状态置为 0, 细胞核置为 $\emptyset$ 。

并行步2. 网络信息利用行为 $\epsilon_i(t)$ 中的 $k_i(t)$ 个三元组随机映射到 GCA 中 $k_i(t)$ 个状态为 0 的细胞, 并行执行细胞动力学规则 R1。

并行步3. GCA 各细胞根据自身状态和输入条件并行执行细胞动力学规则 R2~R7。

并行步4. 重复并行步3, 直至达到平稳概率分布。

并行算法 GAA 中, 所有细胞是并行工作的。算法本身, 以及细胞阵列的结构, 细胞动力学规则等都和问题规模无关, 即与网络中发生的信息利用行为的数量和突发程度无关。

3 算法及仿真实验

3.1 算法

为了模拟实际网络情况, 我们每隔一段时间 (time interval) 把这段时间产生的多个信息利用行为随机映射到 CA 上, 这时的细胞所代表的信息是杂乱无章的。每个细胞根据规则做出判断, 选择跳转, 收缩, 或是保持原位。只要细胞自动机上的已有信息利用行为总数小于最大数目, 或者超时的信息利用行为被废弃, 每隔一个时间段产生的新信息利用行为就可以不断地映射到细胞自动机上, 形成新的自组织群体。多个细胞按照这一规则反复运作, 最后剩下的细胞核代表的是一类信息自组织群, 而其信息扩散地址 d_i 表示一定区域内此类信息自组织群的中心节点地址。

由于此群落中心地点距发出信息利用行为的节点有相对较短的平均距离, 而且我们可以预测将来将要发出此类信息利用行为的节点很有可能也来自这一地区, 因此以后类似的行为可以得到更快更多的响应。此外, 自组织模式不是被动的、集中式的, 随着时间推移, 信息的分布将朝着经常需要此类信息的节点移动分布。

我们得到的这些新的信息源对于大多数查询是不可见的, 也就是说, 发出查询的节点仍然只知道以前的信息源节点, 因此我们为每个信息源建立信息拷贝节点表 (KCNT, key copied node table)。在本文中, 表中每一项 KCN 存储的是信息内容 K 及拷贝自本信息源且信息内容特征向量为 K 的新的信息源物理地址。当内容为 K 的数据包到达时, 通过检索信息拷贝节点表可知道距发出信息请求节点最近的信息源节点, 并把信息利用行为转交到新的信息源节点。

为了判断我们的模型是否有效, 及是否形成了信息自组织, 我们以平均响应效率是否有一定程度的提高为依据。我们将分配一定数目的站点以一定的概率随机产生信息利用行为, 访问信源, 每隔一段时间计算信息利用行为的平均响应时间。

3.2 实验环境

实验选用 100×100 细胞自动机网络,最大信息利用行为数目 $\max$ 为4000;信息内容种类为100;信息内容相似度阈值 θ_s 为0.01;信息地址接近度阈值 θ_a 为0.3;信息利用行为过期阈值 θ_t 为50个 time-interval。

3.3 实验结果及评价

本实验中用时间作种子产生 $\max$ 个信息利用行为,随机分布在 $\max$ 个细胞上。图3(time=0)即为细胞自动机的初始

状态示意图,x和y轴代表空间(100×100),z轴代表信息利用内容种类($k, k \in \{1 \dots 100\}$)。可以看出自组织聚类之前的广义细胞自动机比较杂乱,不同种类的信息随机地散落在细胞自动机的各个地点,time=1~5分别为细胞自动机按信息内容和物理地址自组织聚类1~5次后的状态示意图,在time=5时由图可见此时信息内容接近的细胞核已经聚集成一个中心细胞核,其地址即为新的信息扩散地址。

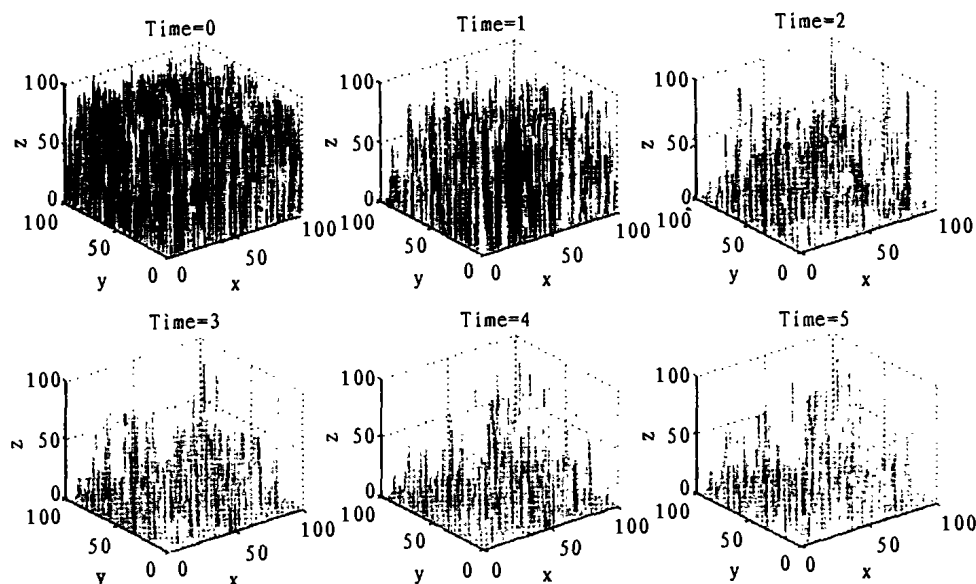
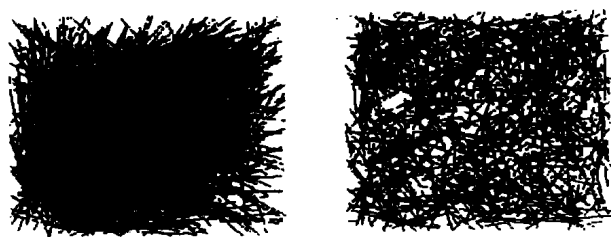


图3 并行算法GAA执行过程中,广义细胞自动机阵列各细胞状态的变化情况
(GCA 细胞数量10000;信息内容的类别数量100;信息请求数量4000;由用户指定的信息源数量5;
自组织后网络可提供的信息源数量256)

为了计算此算法是否使平均查询效率有了一定程度的提高,我们对比网络信息目的地址与信息源地址在产生自组织现象前后的距离(如图4)。图中一条直线代表由信息利用者指定的一次网络信息利用行为,直线的两端点分别为该信息利用行为的目的地址与信息源地址。不难看出,形成信息自组织之后,平均距离有了较大幅度的减少。

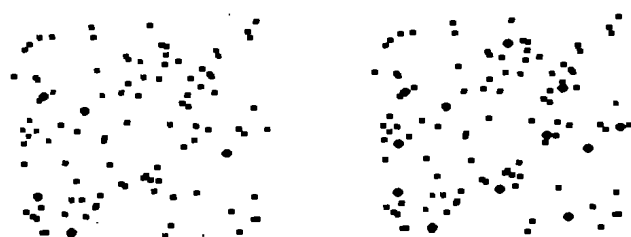


(a) 自组织前

(b) 自组织后

图4 网络信息目的地址与信息源地址距离前后对比图
(GCA 细胞数量10000;信息内容的类别数量100;信息请求数量4000;由用户指定的信息源数量5;自组织后网络可提供的信息源数量256)

换一个角度,我们给出信息自组织利用前后对比图(如图5)。某次网络信息利用行为的信息目的节点与信息源节点初始分布情况如左图。而右图为按照本文基于GCA的网络信息自组织利用模式形成的自组织分布结构。由图5可看出按照网络信息利用需求生成了新的信息源节点,简化了网络实际信息利用途径。



(a) 自组织前

(b) 自组织后

图5 网络信息目的地址与信息源地址分布前后对比图
(GCA 细胞数量10000;信息内容的类别数量100;信息请求数量4000;由用户指定的信息源数量5;自组织后网络可提供的信息源数量256)

根据大量仿真实验,我们得到平均响应效率曲线(如图6)。不难看出,平均响应效率与信息源数目有关系。随着信息源数目的增长,也就是说,信息源越来越分散,平均查询效率呈较快的下降趋势。这是可以理解的,如果信息源数目多且分散,大多数的信息利用行为将在临近信息源节点得到响应。

结论 本文提出的网络信息自组织利用模式是各个网络节点根据需求变化自发形成的信息分布,每个节点上的内容基本上能反映附近节点的信息需求,也就是说,总体上看大多数的信息请求能在附近节点得到响应。此外,系统在信息内容和信息源之间建立了一种多对多的动态的分布式表示,对来自不同信息源的相关知识进行重新组合,为不同用户产生新

(下转第125页)

结果表明,ARABR 算法的性能略优于算法 A,且多数情况下可获得最小属性约简,这与理论分析结论是一致的,因而 ARABR 算法是有效可行的。

结束语 本文在引入相容决策规则概念之后,提出了基于规则的属性约简算法 ARABR,算法 ARABR 不仅为决策表的最小属性约简提供了一种新的框架,而且还提供了一种度量候选属性与已选择属性集相关度的新方法,使相关度最大的候选属性得以扩展,并使搜索空间显著减小,克服了算法 A 单纯利用单属性的近似精度来扩展候选属性所存在的不足。理论分析表明,算法 ARABR 比 Jelonek 算法具有更低阶的计算复杂性,且可克服算法 A 存在的不足,因而是有效可行的。如何对属性约简集进行增量式维护为下一步的研究目标。

参考文献

- 1 Pawlak Z. Rough sets. *International Journal of Information and Computer Science*, 1982, 11(5): 341~356
- 2 Pawlak Z. *Rough Set-Theoretical Aspects of Reasoning about Data*. Dordrecht, Kluwer Academic Publishers, 1991
- 3 杨明,孙志挥,季小俊. 基于 Rough Set 的缺省加权规则挖掘算法.

东南大学学报, 2002, 32(1): 115~118

- 4 Pawlak Z, Slowinski R. Rough set approach to multiattribute decision analysis. *Invited Review. European Journal of Operational Research*, 1994, 72: 443~459
- 5 Skowron A, Rauszer C. The discernibility matrices and functions in information systems. In: Slowinski R, ed. *Intelligent Decision Support—Handbook of Applications and Advances of the Rough Sets Theory*. Dordrecht, Kluwer: Academic Publishers, 1992. 331~362
- 6 王珏,王任,苗夺谦,等. 基于 Rough Set 理论的“数据浓缩”. *计算机学报*, 1998, 21(5): 393~400
- 7 Wong S K M, Ziarko W. On optimal decision rules in decision tables. *Bulletin of Polish Academy of Sciences*, 1985, 33: 693~696
- 8 苗夺谦,胡桂荣. 知识约简的一种启发式算法. *计算机研究与发展*, 1999, 6(36): 681~684
- 9 Hu X H, Cerccone N. Learning in relational databases: a rough set approach. *Computational Intelligence*, 1995, 11(2): 323~338
- 10 Jelonek J, et al. Rough Set reduction of attributes and their domains for neural networks. *Computational Intelligence*, 1995, 11(2): 339~347
- 11 叶东毅. Jelonek 属性约简算法的一个改进. *电子学报*, 2000, 12(12): 81~82

(上接第87页)

的知识结构,不同信息来源之间也可以建立联系,进行信息交换。网络存取的基本形态不再是链状的,可以分布表达和传递信息。由于信息的分布更为合理,并能根据网络的情况和用户的请求状况动态地调整,因此可以抵抗许多突发事件。大量实验结果证明了本模型的优越性。

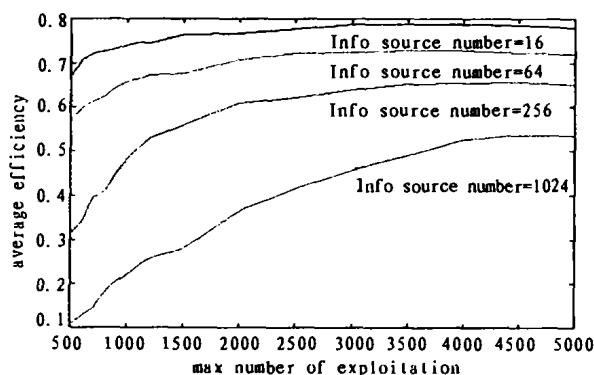


图6 网络信息自组织利用模式的平均响应效率曲线
(GCA 细胞数量10000;信息内容的类别数量100)

参考文献

- 1 Kohavi R, Masand B, Spiliopoulou M. Web mining. *Data Mining and Knowledge Discovery*, 2002, 6(1): 5~8
- 2 Tan P, Kumar V. Discovery of web robot sessions based on their navigational patterns. *Data Mining and Knowledge Discovery*, 2002, 6(1): 9~35

- 3 Ferber, Jacques. *Multi-agent systems: towards a collective intelligence*. Reading, MA: Addison-Wesley, 1998
- 4 Benjaafar, Saifallah, et al. *Cellular automata for traffic flow modeling*. Univ. of Minnesota, Minneapolis, 1997
- 5 Shuai D X. A new parallel-by-cell approach to undistorted data compression based on cellular automaton and genetic algorithm. *Journal of Computer Science and Technology*, 1999, 14(6): 572~579
- 6 Shuai D X, Gu J. The faster high-order cellular automaton for hyper-parallel undistorted data compression. *Journal of Computer Science and Technology*, 2000, 15(2): 126~135
- 7 Feldmann A, Greenberg A, et al. NetScope: traffic engineering for IP networks. *IEEE Network*, March/April, 2000. 11~19
- 8 Casetti C, Meo M. A new approach to model the stationary behavior of TCP connections. *IEEE Infocom*, 2000. 367~375
- 9 Liou C, Tai W. Conformality in the self-organization network. *Artificial Intelligence*, 2000, 116: 265~286
- 10 Liu C, Peek J, Jones R, et al. *Managing Internet Information Service*. O'Reilly Associates, Inc. 1994
- 11 Kumar G P, Venkataram P. Artificial Intelligence approaches to network management: recent advance and a survey. *Computer Communication*, 1997, 20(15): 1313~1322
- 12 Bernardo A, Rajan M. Social Dilemmas and Internet Congestion. *Science*, 1997, 277: 535~537
- 13 Chua L O, Roska T. The CNN paradigm. *IEEE Trans. on Circuits and Systems*, 1993, 40(3): 147~156
- 14 Trinh H, Aldeen M. Distributed state observer scheme for large-scale interconnected systems, *IEE Pro. —Control Theory Appl.*, 1998, 145(3): 331~337