

基于混合小生境人工免疫系统的流量矩阵估测算法

孙 斌¹, 王新勇²

(1 内蒙古科技大学 经济与管理学院, 内蒙古 包头 014010; 2 河南科技大学 电子信息工程学院, 河南 洛阳 471023)

摘 要: 为了解决网络流量矩阵估测精度不高的问题, 基于混合小生境人工免疫系统建立一种新的算法 HNAIS (Hybrid Niche and Artificial Immune System Algorithm). 该方法首先结合链路负载和连通度矩阵定义了流量矩阵模型, 并利用人工免疫系统方法对模型进行求解, 同时引入小生境技术来优化计算结果, 避免陷入局部最优. 最后, 通过利用 OPNET 和 MATLAB 进行仿真实验, 深入研究了影响该算法的关键因素. 实验结果表明, 相比于其他算法, HNAIS 具有较好的适应性.

关键词: 流量矩阵; 估测; 精度; 小生境; 人工免疫系统

中图分类号: TP393

文献标识码: A

文章编号: 1000-7180(2014)01-0008-05

The Traffic Matrix Estimation Algorithm Based on Hybrid Niche and Artificial Immune System

SUN Bin¹, WANG Xin-yong²

(1 School of Economics and Management, Inner Mongolia University of Science and Technology, Baotou 014010, China;

2 College of Electronic Information Engineering, Henan University of Science and Technology, Luoyang 471023, China)

Abstract: In order to mitigate the estimation accuracy of traffic matrix, a new algorithm HNAIS (Hybrid Niche and Artificial Immune System algorithm) is presented by Hybrid niche and artificial immune system. In this algorithm, the Traffic Matrix model is proposed with link load and connectivity matrix at first, which is solved by artificial immune system. Then, the calculation result is optimized with niche technology to avoid falling into the local optimum. At last, a simulation with OPNET and MATLAB was conducted to study the key influence factors. The result shows that, compared to other algorithm, HNAIS has better adaptability.

Key words: traffic matrix; estimation; accuracy; niche; artificial immune system

1 引言

随着网络规模和网络流量的迅速增加, 有效地预防和处理网络故障显得尤为重要. 而流量矩阵作为描述网络全局状态的关键因素, 因此, 如何提高下一时刻流量矩阵的预测精度成为目前研究的热点^[1-3]. 在大规模网络中, 直接测量流量矩阵存在较大困难, 但是可以通过链路负载和连通度矩阵来间接获取流量矩阵信息. 对此, 国内外学者提出了大量解决方法, 典型算法如重力模型方法、独立连接模型方法、最优化技术及路由变化方法和统计推断方法

等. 文献[4]中阐述了流量矩阵的基本预测原理, 并详细讨论各种算法的性能和预测误差. 文献[5]中基于路由矩阵提出了一种估算流量矩阵的新方法. Liang G 和 Yu B^[6]利用伪似然估计方法, 将流量矩阵问题被分解为若干个包含一个 OD 对的子问题, 从而有效地降低了计算复杂度. 但是这些方法通常都需要建立关于 OD 流的数学模型, 并结合 OD 流的直接测量来推导流量矩阵, 由此会产生较大的计算量.

针对上述问题, 本文首先定义流量矩阵模型, 并利用混合小生境人工免疫系统算法对流量矩阵进行

收稿日期: 2013-03-18; 修回日期: 2013-05-03

基金项目: 国家自然科学基金(60263002); 内蒙古科技大学创新基金(2011NCL029)

求解^[7],同时通过仿真实验深入研究影响该算法的关键因素.

2 流量矩阵模型

假设存在如图1所示的一个网络拓扑结构 G , 该网络被划分为 m 个社团 (community), 社团内部节点具有紧密的连接关系, 而社团之间则相对稀疏, 主要通过某几个核心节点进行连接. 2002 年, Girvan 和 Newman 提出了基于链路介数的社团划分算法 GN, 通过引入介数来衡量网络流量大小. 不难发现, 当链路的介数越高对社团之间连通性越重要. 为了研究方便, 可以将社团与社团之间的流量抽象为介数最高的链路之间的流量和.

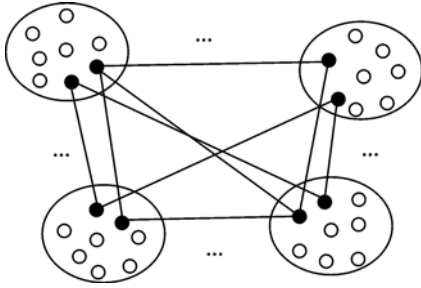


图1 社团划分示意图

令 m 个社团之间存在 k 条介数最高的链路, 定义 t 时刻下的网络流量矩阵为 $\mathbf{F}(t) = [f_1(t), f_2(t), \dots, f_m(t)]$, 链路负载为 $\mathbf{L}(t) = [l_1(t), l_2(t), \dots, l_k(t)]$, 社团之间的连通度矩阵 $\mathbf{R} = [r_{ij}]$, r_{ij} 为社团 i 与社团 j 之间直接相连的链路数, 如果没有直接相连则 $r_{ij} = 0$, 如式(1)所示:

$$\mathbf{R} = [r_{ij}] = \begin{bmatrix} 0 & r_{12} & \dots & r_{1m} \\ r_{21} & 0 & \dots & r_{2m} \\ \dots & \dots & \dots & \dots \\ r_{m1} & r_{m2} & \dots & 0 \end{bmatrix} \quad (1)$$

那么, t 时刻下网络流量矩阵 $\mathbf{F}(t)$ 、链路负载 $\mathbf{L}(t)$ 和连通度矩阵 $\mathbf{R}$ 之间满足如下关系:

$$\mathbf{F}(t) = \mathbf{L}(t)\mathbf{R} \quad (2)$$

通常情况下, 式(2)存在无穷组解. 为了获得这无穷组解中的最优解, 可以通过观察实际链路负载变化情况, 并采用流量矩阵估计方法来获取. 同时, 由于传统的线性规划或者启发式算法对于求解此类问题具有一定的局限性 (比如约束条件太多, 计算复杂度太大, 等等), 因此, 本文提出采用一种新的方法——混合小生境人工免疫系统算法来求解网络流量矩阵 $\mathbf{F}(t)$.

3 混合小生境人工免疫系统算法

3.1 人工免疫系统

人工免疫系统 (Artificial Immune System, AIS) 利用生物免疫系统原理, 通过高度并行的自学习、自适应等操作, 构造自组织、高性能和鲁棒性较好的人工智能系统. 本文结合人工免疫系统来对网络流量矩阵 $\mathbf{F}(t)$ 进行刻画. 这里将流量矩阵元素看作自体 Self, 噪音看作非自体 Nonself, 链路负载看作抗原 Q , 那么这三者之间满足如下关系:

$$\begin{cases} \text{Self} \cap \text{Nonself} = \varphi \\ \text{Self} \cup \text{Nonself} = Q \end{cases} \quad (3)$$

AIS 演化策略中, 会经历自体演化和耐受、成熟免疫细胞生命周期以及免疫记忆等过程. 这三步演化过程可描述如下:

Step 1: 对未成熟细胞进行自体演化和耐受操作, 用来消除流量在传输过程中因为噪音等干扰而产生的误差. 未成熟细胞如果能够成功经历一个自体耐受期 Ω , 则进化成为成熟细胞, 否则该细胞将被淘汰. 具体的自体演化和耐受操作可采用式(4)和式(5)来刻画:

$$\begin{aligned} Q(t) = & \begin{cases} \text{Self}(0), & t = 0 \\ Q_{\text{new}}(t), & t > 0, \text{mod}(t, \Omega) = 0 \\ Q(t-1) - Q_{\text{Nonself}}(t), & t > 0, \text{mod}(t, \Omega) \neq 0 \end{cases} \end{aligned} \quad (4)$$

$$\begin{aligned} \text{Self}_{\text{dead}}(t) = & \begin{cases} \varphi, & |\text{Self}(t-1)| + |Q(t-1)| \leq L \\ \text{Self}_{Q(t-1)}(t-1), & |\text{Self}(t-1)| + |Q(t-1)| > L \end{cases} \end{aligned} \quad (5)$$

式中, $\text{Self}_{\text{dead}}(t)$ 表示 t 时刻被淘汰的自体, $Q_{\text{new}}(t)$ 为 t 时刻产生的新抗原, L 为自体集合上限, $\text{Self}_{Q(t-1)}(t)$ 为 $t-1$ 时刻进化成功的前 $Q(t-1)$ 个抗原.

Step 2: 对上述成功经历自体演化和耐受操作的细胞将进入成熟细胞生命周期. 细胞如果能够在生命周期内能够匹配足够多的抗原, 则进入激活状态, 否则将被淘汰. 令 $T_1(t)$ 为 t 时刻成熟细胞集合 (表示抗原匹配数未达到激活阈值的不与自体匹配的细胞), $T_2(t)$ 为下一代成熟细胞集合, $T_3(t)$ 为新成熟细胞集合 (用来替换被淘汰细胞), $T_4(t)$ 为被淘汰细胞集合, $T_5(t)$ 为处于激活状态的细胞集合, 那么成熟细胞生命周期可由式(6)来刻画:

$$T_1(t) = T_2(t) \cup T_3(t) - T_4(t) - T_5(t) \quad (6)$$

Step 3 对免疫细胞执行免疫记忆操作. 令 $X(t)$ 为 t 时刻的记忆免疫细胞集合, 当记忆细胞匹配上某个抗原 $Y(t)$ 时, 则立即进入激活状态, 进入激活状态后的免疫细胞在协同刺激的帮助下将转化为记忆细胞 $Z(t)$, 如果缺乏协同刺激将被淘汰. 其免疫记忆操作过程可用式(7) 描述:

$$X(t) = X(t-1) - Y(t) \cup Z(t) \quad (7)$$

3.2 小生境原理

小生境(Niche)是指利用同种生物体聚集生活一起形成小的社团环境, 而不同种类生物体将产生分离. 因此, 可以利用小生境原理, 将具有相同特征的数据划分成社团, 减少由于数据大规模集聚而使得种群多样性降低的可能性, 有效地避免陷入局部最优. 本文采用小生境的排挤机制选择策略来刻画流量矩阵. 具体的排挤机制选择策略如下:

首先, 对种群进行初始化, 设置排挤系数 β ;

其次, 在种群中以概率 $1/\beta$ 来选择个体参加排挤运算;

第三, 根据运算规则产生新个体, 并且计算被排挤个体与新个体之间的海明距离;

最后, 结合海明距离和新个体, 淘汰与被排挤个体相似的个体, 同时跳转到第 2 步重复执行, 直至选择结束.

但是在使用小生境的过程中存在一个关键问题: 何时引入小生境来对 AIS 操作结果进行优化. 通过分析, 细胞最有可能出现集群现象是在细胞进行自体演化和耐受操作时. 当 $\text{Self}_{\text{dead}}(t) < 0.05\text{Self}(t-1)$ 时, 引入小生境技术, 随机产生现有成熟细胞数目的 30% 新细胞, 利用排挤机制选择策略将不符合条件的细胞进行淘汰, 以此避免种群陷入局部最优.

3.3 算法描述

针对上述提出的 AIS 演化策略与小生境的排挤机制选择策略, 这里提出混合小生境人工免疫系统算法 HNAIS (Hybrid Niche and Artificial Immune System algorithm) 来求解网络流量矩阵, 具体描述如下:

(1) 在 $t=0$ 时刻初始化网络, 划分社团并确定初始链路负载、连通度矩阵以及迭代次数;

(2) 将各个社团流量看作免疫细胞, 根据式(4) 和式(5) 对免疫细胞进行自体演化与耐受, 淘汰不满足条件的免疫细胞 $\text{Self}_{\text{dead}}(t)$, 并且产生成熟免疫细胞集合;

(3) 如果 $\text{Self}_{\text{dead}}(t) < 0.05 \text{Self}(t-1)$, 则引入

小生境技术, 随机产生现有成熟细胞数目的 30% 新细胞, 利用排挤机制选择策略重新淘汰不符合条件的细胞, 并产生新的成熟免疫细胞集合;

(4) 针对成熟免疫细胞集合, 在生命周期内利用式(6) 匹配抗原, 使其进入激活状态;

(5) 对进入激活状态的细胞进行免疫记忆, 根据式(7) 获得记忆细胞, 从而获得下一时刻链路负载的最优估测值;

(6) 结合连通度矩阵和链路负载的最优估测值, 根据式(2) 计算下一时刻流量矩阵最优值;

(7) 令 $t=t+1$, 并判断当前迭代次数是否结束, 如果没有结束则跳转到步骤(2) 继续进行估测下一时刻流量矩阵, 否则跳转到步骤(8);

(8) 算法结束.

4 仿真实验

为了验证算法 HNAIS 的有效性, 这里利用 OPNET 和 MATLAB 进行联合仿真实验. 首先建立图 2 所示的网络拓扑结构图, 整个网络被划分为 5 个社团(即 $m=5$), 各个社团之间的链路状况如图中所示($k=9$), 社团内部各节点之间是全连通的. 同时设置网络参数值: 流量数据包大小为 1 024 bit, 节点缓存容量大小为 50 kbit, 社团内部的链路带宽为 5 Mb/s, 社团之间的链路带宽为 10 Mb/s, 延时 10 ms. 并且初始化 HNAIS 算法参数: 初始自体集合大小为 50, 每步产生 30 个未成熟细胞, 耐受期为 9 s, 迭代次数为 200, 排挤系数 $\beta=2$. 根据图 2 拓扑结构, 连通度矩阵 R 可表示为

$$R = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix} \quad (8)$$

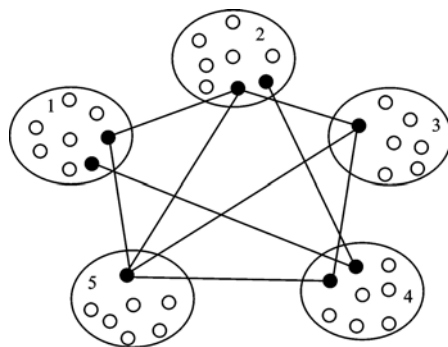


图 2 仿真示意图

同时, 这里将文献[8]提出的基于神经网络方法

的流量矩阵估计算法 MMRMLP (Multi-input and Multi-output RMLP network model) 与本文提出的 HNAIS 算法对比分析, 在图 3 中显示了 100 s 内这两种算法估测结果与实际统计结果的情况. 从图 3 可以看出, HNAIS 算法与实际统计结果较为接近, 而 MMRMLP 则存在较大误差. 通过进行数据分析, HNAIS 算法和 MMRMLP 算法的残差分别为 15.82% 和 23.81%.

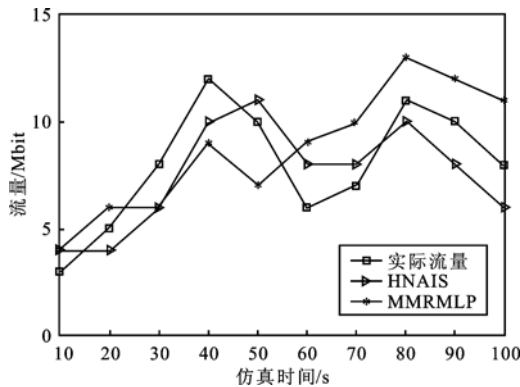


图 3 三种算法估测结果比较

其次, 这里针对影响 HNAIS 算法的关键因素进行深入研究. 图 4 关注小生境技术对估测结果的影响, 给出了 100 s 仿真时间内不同排挤系数 β 下的估测结果与实际统计结果的对比情况. 从图 4 可以看出, 在初始阶段, β 越小对应曲线越接近于实际统计结果. 而在仿真后期, β 越大对应曲线越接近于实际统计结果. 这是由于在初期网络流量不太大, 流量出现集群现象的几率较小, 所以此时排挤系数越小反而有利于提高估测精度. 当仿真后期网络流量增大, 流量出现集群现象的几率增大, 此时提高排挤系数能够有效降低集群现象, 同时避免陷入局部最优.

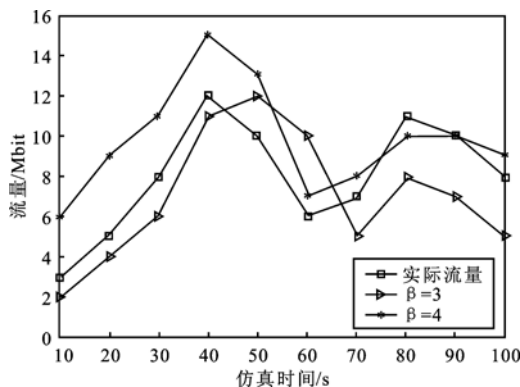


图 4 不同排挤系数的估测结果比较

在图 5 中给出了不同自体集合上限 L 下的估测结果与实际统计结果的对比情况. 从图 5 可以看出, 类似于图 4 中的跳变现象也存在. 当仿真初期自

体集合上限越小, 对应曲线越接近于实际统计结果, 而在仿真后期则现象相反. 这是由于仿真初期流量较小, 自体集合上限越大, 未成熟细胞自体耐受将越不充分, 从而使得成熟细胞识别自体的几率增加, 导致预测误差增加. 而在仿真后期, 当流量逐渐增大时, 自体集合上限越大, 成熟细胞识别自体的几率减少, 估测精度提高.

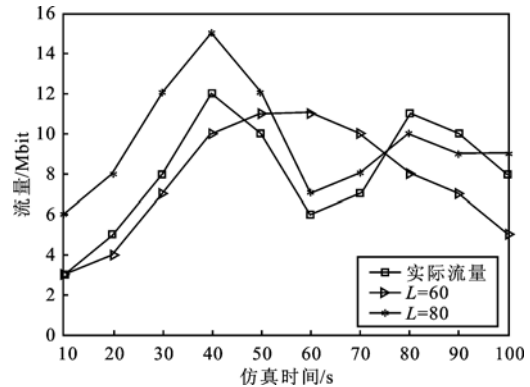


图 5 不同自体集合上限的估测结果比较

体集合上限越小, 对应曲线越接近于实际统计结果, 而在仿真后期则现象相反. 这是由于仿真初期流量较小, 自体集合上限越大, 未成熟细胞自体耐受将越不充分, 从而使得成熟细胞识别自体的几率增加, 导致预测误差增加. 而在仿真后期, 当流量逐渐增大时, 自体集合上限越大, 成熟细胞识别自体的几率减少, 估测精度提高.

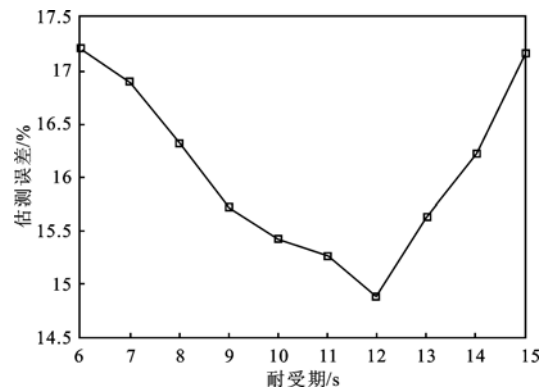


图 6 估测误差与耐受期之间的关系

5 结束语

本文基于混合小生境人工免疫系统算法提出了一种新的算法 HNAIS. 该算法首先结合连通度矩阵和链路负载定义了流量矩阵的表达式, 同时利用人工免疫系统进行求解, 并引入小生境排挤机制选择策略来避免集群现象和局部最优. 仿真实验结果

表明 HNAIS 具有较好的适应性. 在以后的研究中, 可以考虑结合矩阵模型的高度病态性以及智能算法来进一步提高预测精度.

参考文献:

- [1] Teixeira R, Agarwal S, Rexford J. BGP routing changes:merging views from two ISPs[J]. ACM SIGCOMM Computer Communication Review, 2005, 35(3): 79-82.
- [2] 钱叶魁, 陈鸣. 因特网流量矩阵的流形结构[J]. 电子与信息学报, 2010, 32(12): 2981-2986.
- [3] Cao J, Davis D, Wiel S V, et al. Time-varying network tomography;router link data[J]. Journal of the American Statistical Association, 2000, 452(95):1063-1075.
- [4] 周静静, 杨家海, 杨扬, 等. 流量矩阵估算的研究[J]. 软件学报, 2007, 18(11): 2669-2682.
- [5] Tan L, Wand X. A novel method to estimate IP traffic matrix[J]. IEEE Communications Letters, 2007, 11(11): 907-909.
- [6] Liang G, Yu B. Maximum pseudo likelihood estimation in network tomography[J]. IEEE Transactions on Signal Processing, 2003, 51(8): 2043-2053.
- [7] 戚玉涛, 刘芳, 焦李成. 基于分布式人工免疫算法的数值优化[J]. 电子学报, 2009, 37(7): 1554-1561.
- [8] Jiang D D, Hu G M. Large-scale IP traffic matrix estimation based on the recurrent multilayer perceptron network[C]// 2008 IEEE International Conference on Communication. Washington, DC: IEEE Computer Society, 2008: 229-232.

作者简介:

孙 斌 男, (1974—), 博士, 讲师. 研究方向为网络流量工程.

王新勇 男, (1968—), 硕士, 高级工程师. 研究方向为物联网.

(上接第 7 页)

- [5] Vinay V, Shantanu D, Vishal S. Efficient on-line testing of FPGAs with provable diagnosabilities [C]// 41st Design Automation Conference. USA, California, 2004: 498-503.
- [6] Anita J, Vaathi P. Multiple fault diagnosis with improved diagnosis resolution for VLSI circuits [C]// 2010 International Conference on Computing Communication and Networking Technologies. India, Coimbatore, 2010: 1-6.
- [7] Lin Y, Feng Lu, Kwang-Ting Cheng. Multiple-fault diagnosis based on adaptive diagnostic test pattern generation[J]. IEEE transaction on Computer-Aided Design of Integrated Circuits and Systems, 2007 (26): 932-942.
- [8] Chun-Lung Hsu. Built-in self-test design for fault detection and fault diagnosis in SRAM-Based FPGA[J]. IEEE Instrumentation and Measurement Society, 2009 (50): 2300-2315.
- [9] Shantanu D. Mixed PLB and Interconnect BIST for FPGAs Without Fault-Free Assumption [C]// 24th IEEE VLSI Test Symposium. USA, California, 2006: 38-43.
- [10] San J Ca. The programmable logic data book[M]. USA: Xilinx Inc, 2003.

作者简介:

李文昌 男, (1974—), 博士. 研究方向为 FPGA 设计与测试、电源管理芯片的设计.

万 理 男, (1988—), 硕士研究生. 研究方向为数字 IP 验证.

阮爱武 女, (1967—), 博士, 副教授. 研究方向为 SOC 的验证、FPGA 和 SOPC 设计.

于敦山 男, (1970—), 博士, 教授, 博士生导师. 研究方向为 SOC 设计与测试、电源管理 IP 设计等.