

# 基于 BIC 测度和遗传算法的 TANC 结构学习

蒋望东<sup>1</sup>, 林士敏<sup>1</sup>, 鲁明羽<sup>2</sup>

(1. 广西师范大学 计算机科学系, 广西 桂林 541004;

2. 清华大学 智能技术与系统国家重点实验室, 北京 100084)

**摘要:**树扩展朴素贝叶斯分类器(TANC)是实用性较强的一种分类器,其性能优于朴素贝叶斯分类器。现有的 TANC 结构学习算法有基于互信息测度的相关性分析方法和贝叶斯信息测度(BIC)的搜索打分方法。将遗传算法引入 TANC 结构学习,用 BIC 作为评价函数,提出了基于 BIC 测度和遗传算法的 TANC 结构学习算法 GA-TANC,并以此构建分类器,用分类准确率衡量算法的性能。实验结果表明,GA-TANC 算法有更高的分类准确率,从而说明 GA-TANC 结构学习算法是准确有效的。

**关键词:**贝叶斯分类器;树扩展朴素贝叶斯分类器;遗传算法;结构学习

**中图分类号:**TP301.6

**文献标识码:**A

**文章编号:**1673-629X(2007)04-0096-04

## Structure Learning of TANC Based on BIC and Genetic Algorithms

JIANG Wang-dong<sup>1</sup>, LIN Shi-min<sup>1</sup>, LU Ming-yu<sup>2</sup>

(1. Computer Science Department, Guangxi Normal University, Guilin 541004, China;

2. State Key Laboratory for Intelligent Technology & System, Tsinghua University, Beijing 100084, China)

**Abstract:** Tree Augmented Naive Bayesian Classifier (TANC) is a type of quite applied classifier, its performance is superior to Naive Bayesian Classifier. Existing TANC structure learning algorithms are based on relativity analysis using mutual information criterion or based on search & scoring using Bayesian information criterion. Using BIC as evaluate function, this paper introduces genetic algorithm into TANC structure learning, and proposes a new TANC structure learning algorithm based on BIC and genetic algorithm. Using classification accuracy to scale classification performance. Experiment results show that GA-TANC is accurate and effective.

**Key words:** Bayesian classifier; TANC; genetic algorithm; structure learning

### 1 树扩展朴素贝叶斯分类器

贝叶斯分类器以其坚实的理论基础、自然的表示方式、灵活的推理能力和方便的决策机制受到越来越多的重视<sup>[1,2]</sup>。设有变量集  $U = \{A_1, A_2, \dots, A_n, C\}$ , 其中  $A_1, A_2, \dots, A_n$  是实例的  $n$  个属性变量, 实例可用向量  $x_i = (a_1, a_2, \dots, a_n)$  表示, 其中,  $a_i$  是  $A_i$  的值,  $C$  为类变量,  $c$  表示  $C$  的值。应用贝叶斯定理, 实例  $x_i$  属于类  $c_j$  的概率为:

$$p(c_j | a_1, a_2, \dots, a_n) = \frac{p(a_1, a_2, \dots, a_n | c_j) p(c_j)}{p(a_1, a_2, \dots, a_n)} = \alpha p(a_1, a_2, \dots, a_n | c_j) p(c_j)$$

$$(c_j) \quad (1)$$

其中,  $\alpha$  是正则化因子,  $p(c_j)$  是类  $c_j$  的先验概率,  $p(c_j | a_1, a_2, \dots, a_n)$  是类  $c_j$  的后验概率。依据概率的链式规则, 式(1)可以表示为:

$$p(c_j | a_1, a_2, \dots, a_n) = \alpha p(c_j) \prod_{i=1}^n p(a_i | a_1, a_2, \dots, a_{i-1} | c_j)$$

实例  $e = (a_1, a_2, \dots, a_n)$  为  $c$  类的概率为  $p(c | e) = \frac{p(a_1, a_2, \dots, a_n | c) p(c)}{p(a_1, a_2, \dots, a_n)}$ 。实例  $e$  被分到  $c$  的最大后验概率的类  $C^*$  中,  $g(e) = \arg \max_c p(c | a_1, a_2, \dots, a_n)$ ,  $g(e)$  称为贝叶斯分类器。

Duda 和 Hart<sup>[3]</sup>于 1973 提出了基于贝叶斯公式的朴素贝叶斯分类器(NBC)。NBC 假设所有的属性都条件独立于类变量  $C$ , 即每个属性变量都以类变量作为唯一的父结点。但在实际的应用领域中, 各个属性相互独立的假设很难成立, 成为进一步提高其精度的主要障碍。Friedman<sup>[4]</sup>在 NB 结构的基础上提出了 TAN 结构, 类结点没有父结点, 属性结点的最大父结点数是

收稿日期: 2006-06-30

基金项目: 国家自然科学基金资助项目(60473115)

作者简介: 蒋望东(1971-), 男, 湖南永州人, 湖南财经高等专科学校讲师, 硕士研究生, 主要研究方向为人工智能、机器学习; 林士敏, 教授, 硕士研究生导师, 主要研究方向为知识工程、数据采掘; 鲁明羽, 副教授, 博士后, 研究方向为数据采掘、网络挖掘。

## 2. 基于 TAN 结构的分类器就是 TANC。

建立 TAN 结构的方法是:在 NB 结构的基础上,在属性之间增添扩展弧,以消除朴素贝叶斯关于条件独立的假设。如何增添一组最好的扩展弧是问题的关键,这相当于学习以类变量  $C$  为根结点的最优的贝叶斯网络。TANC 的建构过程也分成结构学习和参数学习两个阶段,应用贝叶斯定理,类结点的最大后验概率  $C_{\text{TAN}} = \arg \max_{c \in C} p(c_j) \prod_{i=1}^n p(a_i | \prod a_i)$ ,  $C_{\text{TAN}}$  表示 TANC 输出的目标值。

由文献[4]可知,学习 TANC 结构的时间复杂性为  $O(n^2 * m)$ , 其中,  $n$  是属性个数,  $m$  是实例个数。TANC 的分类效果要优于 NBC。

在计算属性对之间关系的函数  $f(A_i, A_j, D)$  时, Chow<sup>[5]</sup> 等人用信息论中的互信息 MI (Mutual Information) 测度衡量属性  $Y$  提供给属性  $X$  的信息量。互信息测度只考虑属性结点  $X$  和  $Y$  之间的关系,而忽略了类结点  $C$  所提供的信息。Friedman 提出用条件互信息 CMI (Conditional Mutual Information) 测度计算  $f(A_i, A_j, D)$ , 在全局结构上更好地考虑了  $Y$  与  $X$  之间的依赖关系,使分类的精度得到提高。程泽凯和林士敏<sup>[6,7]</sup> 将贝叶斯标准信息测度(BIC)引入到 TANC 的结构学习中,用以衡量属性结点之间的相关性,提出了 TANC-BIC 结构学习算法,从而使分类精度进一步得到提高。BIC 测度是 Schwarz 1978 年提出的,其公式为:

$$Q_{\text{BIC}}(B, D) = \sum_{i=1}^n \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} (N_{ijk} + \alpha_{ijk} - 1) \cdot \log \frac{N_{ijk} + \alpha_{ijk} - 1}{N_{ij} + \alpha_{ij} - r_i} - \frac{1}{2} \text{Dim}(B) \log N \quad (2)$$

其中,属性  $A_i$  有  $r_i$  个状态,其双亲集合(可能不止一个)有  $q_i$  个状态,  $N_{ijk}$  是满足属性  $A_i$  的第  $k$  个状态,且  $A_i$  的双亲集合的第  $j$  个状态的记录数目;  $N_{ij}$  是满足属性  $A_i$  的双亲的第  $j$  个状态的记录数目,即  $N_{ij} = \sum_{k=1}^{r_i} N_{ijk}$ ;  $\alpha_{ijk}$  和  $\alpha_{ij}$  是先验信息,  $\alpha_{ij} = \sum_{k=1}^{r_i} \alpha_{ijk}$ 。Cooper 等人建议无信息均一先验  $\alpha_{ijk} = 1$ ; Buntine 认为先验值依赖于结点的状态数和其双亲的状态数,也称之为 BDeu (likelihood equivalent uniform Bayesian Dirichlet) 先验,其公式为  $\alpha_{ijk} = \text{priors} / (r_i \cdot q_i)$ , 变量  $\text{priors}$  是先验值。基于 BIC 测度的 TANC 结构学习算法与现有算法的区别在于  $f(A_i, A_j, D)$  的计算不同: TANC 主要是用互信息和条件互信息测度计算  $f(A_i, A_j, D)$  的, TANC-BIC 算法考虑属性结点之间的相关性,所设计的计算公式如下:

$$f(A_i, A_j, D) = \text{score}(A_i, A_j) - \text{score}(A_i)$$

其中,  $\text{score}()$  是按式(2)对  $i-j$  属性对打分的函数。

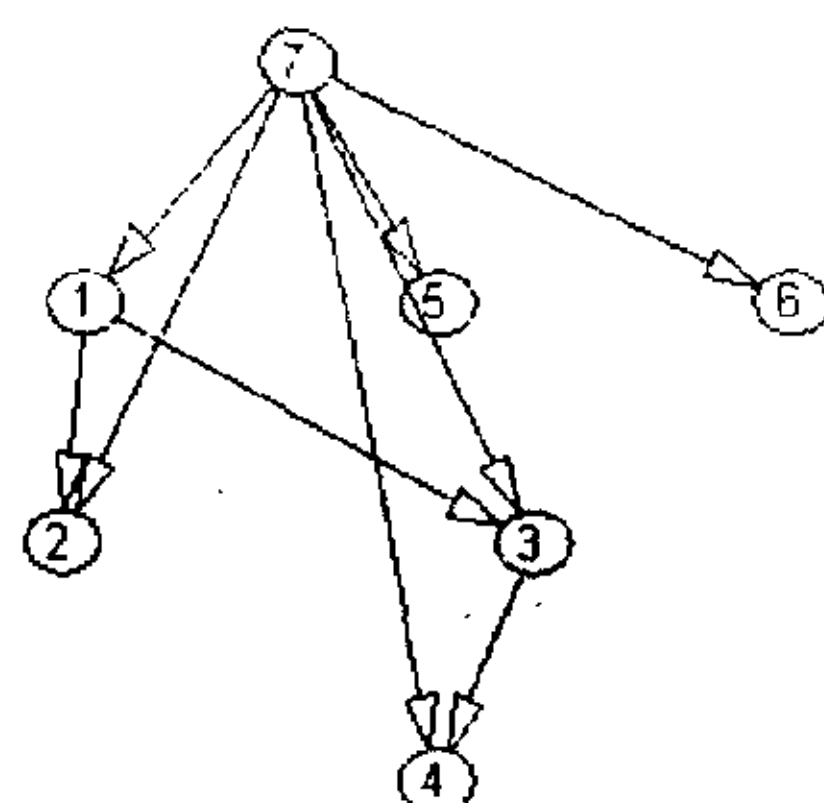
## 2 学习 TANC 结构的遗传算法设计

遗传算法<sup>[8,9]</sup>的创始人是美国著名学者、密西根大学教授 John H. Holland。Holland 于 20 世纪 50 年代末期开始研究自然界的自适应现象,并希望能够将自然界的进化用于求解复杂问题的智能程序设计。他认为,可以用一组编码来模拟一组计算机程序,并且定义了一个衡量每个“程序”的度量:“适应值”。Holland 模拟自然选择机制对这组“程序”进行“进化”,直到最终得到一个正确的“程序”为止。编码方式有:二进制编码、十进制编码和符号编码等方法。二进制编码是最常用的编码方法,用在很多方面;十进制编码包括整数编码和实数编码。实数编码一般用于连续变量的优化计算,整数编码与符号编码一般用于与顺序有关的组合优化方面的问题。

### 2.1 TANC 结构的编码表示

#### (1) 矩阵表示法。

即用矩阵的形式来表示 TAN 网络结构,若结点  $i$  到结点  $j$  有一条弧,则对应矩阵中  $(i, j)$  值为 1, 否则为 0。如图 1 所示的 TAN 网络,其结构的矩阵表示形式如图 2 所示。



$$\begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 \end{bmatrix}$$

图 1 7 个变量的 TAN 网

图 2 矩阵表示法

#### (2) 邻接表链表示法。

如图 3 所示的对图 1 网络的表示,其中每一行表示与一个结点相关的局部网络结构,每一行的第一个变量表示网络结构中的第  $i$  个结点,其余的变量表示网络结构中该结点的父结点。图 3 只是为了清晰说明编码的方法,实际算法的内部编码只表示每个结点的父结点集。结点编号由位置确定,即图 3 中的编码省略各行的第 1 列,第  $i$  行表示结点  $i$  的父结点集。

#### (3) 定长的邻接表链表示法。

如图 4 (a) 所示,其方法与(2)相同,不同的是定长实际就是为每个结点规定了最大父结点数。由于第七个结点即类结点没有父结点,可以省略。又由于在 TAN 中,每个结点的父结点除类结点外,最多还有一个父结点,故又可直观地表示成图 4 (b) 所示的形式,其中 0 表示父结点只有类结点。

以上三种表示方法都很直观,且各有优势,矩阵表示法和定长的邻接表链表示法适合于以矩阵运算见长

的编程语言,邻接表链表示法适合于以指针表示数据结构的编程语言。笔者选择的编程工具是 Matlab, Matlab 的基本数据单元是一个维数不加限制的矩阵,用它编写遗传算法程序,比用 C 等其它高级语言要简单、灵活、快捷,程序篇幅也将缩小许多。所以选用第一种表示方法。



图 3 邻接表链表示法 图 4 定长的邻接表链表示法

## 2.2 染色体的选择操作

染色体的选择方法一般有:

(1)适应度比例法,又称轮盘赌选择法。即根据适应度函数计算各染色体的适应度值,求出所有染色体的适应度值的总和,计算每个染色体的选择概率,计算每个染色体的累积概率  $q_i$ ,产生随机数  $r$ ,若  $r < q_1$ ,则选择第一个染色体。重复选取,直到选择的染色体总数等于初始染色体总数。显然,在这种染色体选择方法中,适应值越大的染色体即结果越好的染色体被选择的概率也越大,但这种方法容易引起过早收敛;完全依赖概率随机地进行寻优操作虽然可以避免陷入局部极小,但受寻优条件的限制,一般只能得到全局范围内的次优解,很难得到最优解。

(2)最佳个体保留法。即把群体中适应度最高的个体不经过配对交叉直接复制到下一代中,但这种方法的全局搜索能力差,仅适用于只有一个峰值的搜索空间,对于具有多个峰值的搜索空间不适用。

(3)竞争选择法。先采用适应度比例法进行选择,交叉后产生下一代,再利用最佳个体保留法将上一代的最佳个体直接保存下来,然后从新群体中淘汰一个适应度最差的个体。这种方法综合了上述两种方法的优点并克服了它们的缺点,文中就采用这种方法进行染色体的选择。

## 2.3 染色体的交叉操作

将选择操作选出的个体随机配对,并按概率  $P_c$  决定是否进行交叉操作。

交叉方法有:(1)单点交叉。选中个体对应网络中的某一个结点,将两个个体对应该结点的进出弧全部交换,也可以只交换进弧,还可以只交换出弧。(2)多点交叉。即选中多个结点,将两个个体对应的多个结点分别进行单点交叉。

在遗传算法中如果只使用一种交叉方法,由于交叉方式单一,容易引起过早收敛,即“早熟”。因此,采用以上两种交叉方法,每种方法均等概率使用,从而避免过早收敛,扩大遗传算法的搜索范围,得到全局最优解。

## 2.4 染色体的变异操作

染色体的变异算子有:(1)增加弧。即随机选取一个结点  $i$ ,若该点除类结点外没有其它父结点,则随机选取另一个结点  $j$ ,添加一条  $j \rightarrow i$  的弧。(2)删去弧。即随机选取两个结点  $i$  和  $j$ ,若两点间存在  $i \rightarrow j$  或  $j \rightarrow i$  的弧,则删去此弧。(3)弧反向。即随机选取两个结点  $i$  和  $j$ ,若两点间存在  $i \rightarrow j$  的弧且  $i$  除类结点外没有其它父结点,则将此弧改为  $j \rightarrow i$ 。

在学习 TAN 网络结构的遗传算法中灵活采用以上三种变异方法,均适用情况下每种方法均等概率使用,扩大搜索范围,避免过早收敛,得到全局最优解。

## 2.5 交叉和变异时异常情况的处理和优化

所要学习的贝叶斯网络结构要求是有向无环的,但在执行交叉和变异操作后,往往会产生有回环的网或非 TAN 网。对此一般有四种处理方法:(1)用上代种群中最好的 TAN 网代替;(2)随机产生一个 TAN 网代替;(3)给这个网络以极小的适应度,使之被选中的机率趋近于 0,因为非法的结构中可能蕴含着较好的局部结构,通过选择、交叉、变异等操作遗传给下一代,但这样操作容易使种群中的染色体两极分化,从而过早收敛,得不到全局最优解;(4)修改这个网络,使之无环并为 TAN 网,但这样做通常很繁杂。文中在编程时等概率使用(1)、(2)两种方法,(3)、(4)两种方法另外专文研究。

在标准遗传算法中,变异概率是个常数,经过多次迭代后,种群中的个体适应度容易趋向一致,出现“近亲繁殖”,对群体的进化会产生不利的影响。文中采用的变异概率是自适应变异概率,变异概率随着群体的进化而不断调整。在进化的初期,变异概率较大,随着迭代次数的增加,变异概率逐步减少。交叉概率和变异概率随进化不断进行自适应调整,种群中个体性能提高时交叉概率提高,反之则变异概率增加。

## 2.6 测度函数与适应度值的计算

如前所述,一般采用贝叶斯信息标准测度(BIC, Bayesian information criterion)来评价所学得的贝叶斯网络结构,因此,引入 BIC 测度来评价 TAN 网络,并用 BIC 测度函数来计算个体对应的适应度值,先求所有个体的 BIC 测度总和,再求每个个体的 BIC 测度值与 BIC 测度总和之比,即为每个个体的适应度值,然后进行遗传算法的操作。

### 3 基于 BIC 测度和 GA 的 TANC 结构学习算法及实验结果分析

#### 3.1 基于 BIC 测度和 GA 的 TANC 结构学习算法

基于以上关于 TANC 结构学习和遗传算法的探讨,设计了基于 BIC 测度和遗传算法的树扩展贝叶斯网络分类器(TANC)结构学习算法 GA-TANC。

(1)初始化,根据数据集大小生成第一代 TAN 种群,种群规模视数据集而定,一般取 20~100。

(2)对生成后的贝叶斯网络结构进行 BIC 测度打分,并将 BIC 测度值转换为相应的适应度值。

(3)循环对上一代种群依据其对应的适应度进行选择、交叉、变异,交叉概率取 0.6~0.9,变异概率取 0.01~0.05,循环代数取 200 代。选择算子选用竞争选择算子,两种交叉算子等概率使用,在有弧的情况下变异时等概率采用去弧和弧反向算子,不存在弧的情况下变异时等概率决定添加  $i \rightarrow j$  的弧或  $j - i$  的弧,并对遗传操作后的种群进行异常情况处理。中止条件是直到 200 代或 10 代内最佳 BIC 测度无变化。

(5)输出最佳个体对应的 TAN 网络结构,进行下一步参数学习、推理和分类。

#### 3.2 实验结果分析

实验是在我们构建的 MBNC 实验平台<sup>[6,10]</sup>上进行的。工作硬件环境是 P42.0G,内存 256M,操作系统是 Windows 2000 专业版,选用 Matlab 的版本是 v6.5。

实验使用的数据集从 UCI(University of California in Irvine)下载。UCI 上很多实验数据集和人工数据集被世界各地相关领域的专家和学者用来进行知识发现、机器学习、概率统计和人工智能以及其它领域的算法的实验研究和比较。下载网址是:<http://www.ics.uci.edu/~mlearn/MLRepository.html>。选取了其中的 30 个标准数据集。表 1 列出了经过预处理后的数据集概况,关于各个标准数据集的具体介绍详见相关网页。

文中主要是针对结构学习的,在 MBNC 实验平台上实现了 GA-TANC 算法。参数学习采用 BDeu 先验,先验值 priors 取 1,推理算法采用全局联合树推理算法。准确性评估采用 CV5 和保留方法,数据集的划分方式与文献一致。准确性评估采用分类准确率指标。实验在相同的环境下完成,均取第 1 次测试的结果。实验结果见表 2。表中第 3 列是文献[4]中朴素贝叶斯分类器的实验结果,第 5 列是文献[4]中基于条件互信息测度的树扩展朴素贝叶斯分类器的实验结果,空格表示文献[4]中没有列出的数据。

从表 2 中实验数据可以看出:(1) NBC 和 TANC 的正确率均比 Friedman 文献[4]的结果高;(2) 基于

MI 测度的 TANC-MI 和基于 CMI 测度的 TANC-CMI 算法及基于 BIC 测度的 TANC-BIC 对标准数据集的分类正确率均高于 NBC, TANC-BIC 的算法和 TANC-MI 算法的分类正确率相当;(3) GA-TANC 算法总体分类效果比较明显,比 NBC 和 TANC 各类算法的准确率要高。

表 1 数据集的概况

数据集	属性	类别	训练	测试	数据集	属性	类别	训练	测试
Australiaon	14	2	690	CV5	Lymphography	18	4	148	CV5
Breaks	9	2	683	CV5	Monfn3-7-10	10	2	300	1024
Car	6	4	1880	CV5	Monk1	6	2	124	432
Chess	36	2	2130	1066	Monk2	6	2	169	432
Cleve	10	2	296	CV5	Monk3	6	2	122	432
Corral	6	2	128	CV5	Mushroom	22	2	9120	CV5
Crx	15	2	653	CV5	Nursery	8	2	11025	CV5
Diabetes	8	2	768	CV5	Pima	5	2	768	CV5
Flare	10	2	1066	CV5	Satimage	36	7	4435	2000
German	15	2	1000	CV5	Segment	18	7	1540	770
Glass	9	7	214	CV5	Shuttle-small	8	7	3866	1934
Glass2	9	2	163	CV5	Soybean	35	15	562	CV5
Heart	13	2	270	CV5	Vehicle	18	4	846	CV5
Hepatitis	19	2	80	CV5	Vote	16	2	435	CV5
Iris	4	3	150	CV5	Waveform21	18	3	300	4700

表 2 测试实验结果

数据集\算法	NBC	文献[3] NBC	文献[5] TANC	TANC - MI	TANC - CMI	TANC - BIC	GA - TANC
Australiaon	86.94	86.23	81.30	83.62	84.93	85.36	86.67
Breaks	97.65	97.36	95.75	96.91	96.91	96.62	97.21
Car	87.39			91.81	91.28	92.87	85.80
Chess	87.15			86.57	92.49	88.46	88.76
Cleve	83.39	82.76	79.06	82.37	80.34	84.07	85.08
Corral	86.4	85.88	95.32	84	99.2	84	100
Crx	87.54	86.22	83.77	84.15	85.39	87.69	87.38
Diabetes	77.78	74.48	75.13	78.95	76.99	78.95	79.09
Flare	80.56	79.46	82.74	82.63	83.1	82.72	82.35
German	75.3			72.7	72.8	73.5	75.50
Glass	75.71			74.29	68.57	75.71	74.76
Glass2	85			83.75	83.75	83.13	84.38
Heart	83.33	81.48	82.96	83.7	83.7	83.7	83.70
Hepatitis	90	71.25	85.00	91.25	86.25	91.25	90.00
Iris	94			95.33	95.33	95.33	92.67
Lymphography	84.83	79.72	66.87	79.31	84.14	86.21	85.52
Monfn3-7-10	86.62	86.42	91.70	91.5	90.14	91.5	88.79
Monk1	71.3			93.98	95.83	68.52	100
Monk2	61.57			60.88	68.98	67.36	63.67
Monk3	97.22			97.92	97.22	95.6	96.36
Mushroom	98.36			99.84	99.93	99.97	99.96
Nursery	88.44			91.76	90.75	91.89	90.49
Pima	79.22	75.51	75.13	77.52	76.73	77.52	77.25
Satimage	81.25			85.3	86.8	85.45	84.80
Segment	92.08			95.06	94.81	93.77	92.86
Shuttle-small	98.81			99.48	99.48	99.48	99.48
Soybean	93.21			90.18	91.96	89.82	92.86
Vehicle	64.85	58.28	67.86	72.78	72.66	72.54	69.94
Vote	90.11	90.34	89.20	92.41	95.4	92.41	92.18
Waveform21	78.72			78.02	78.47	78.68	79.38
平均值	84.82			85.93	86.81	85.80	86.90

实验表明所提出的基于 BIC 测度和遗传算法构建  
(下转第 116 页)

据并进行处理,处理的结果被用作本地分析引擎的输入。根据入侵检测算法,这些数据流包括本节点的系统、用户和通信行为,以及和通信范围内的其它节点的通信行为。

本地检测引擎分析来自于本地数据收集模块收集来的异常行为信息,该模块采用基于统计学方法的异常检测技术,其检测结果将被送达协作检测模块。

### 3.2.3 入侵响应

本地入侵响应模块和全局入侵响应模块共同提供入侵响应行为。入侵响应类型因入侵的类型、网络协议和应用的类型以及入侵证据的确凿程度不同而不同。入侵响应应该通知网络中的其它节点。它可能是识别出被俘节点,并将这个节点排除在外重新组织网络,直到该节点重新得到认证为止;也可能只是简单给出告警指示。用户可以根据它选择拒绝和可疑节点连接,当进行合作时可以将这个节点排除在外。本地响应模块将行为传达给本地移动节点;而全局响应模块负责调整相邻节点之间的行为。

## 4 小 结

Ad hoc 网络由于其动态拓扑、无线信道以及各种资源有限等脆弱性特点,其安全性对安全防范提出了更高的要求。入侵检测技术作为其安全防范的第二道防线,是 Ad hoc 网络获得高抗毁性的必要手段,然而

(上接第 99 页)

的 TANC 结构学习算法是有效的和准确的。

## 4 结束语

现有的 TANC 结构学习算法是基于相关性分析的或搜索打分机制的,相关性分析中采用互信息测度和条件互信息测度,基于搜索和打分的 TANC 结构学习中采用 BIC 测度。文中提出了一种新的结构学习方法,用 BIC 测度作评价函数,引入遗传算法学习 TANC 结构。在 MBNC 实验平台上实现了 GA-TANC 算法。实验结果表明,用这种结构学习算法得到的 TANC 是有效的,在某些数据集中的分类准确率要优于 TANC-MI, TANC-CMI 和 TANC-BIC。

### 参考文献:

- [1] Cooper G, Herskovits E. A Bayesian method for the induction of probabilistic networks from data[J]. Machine Learning, 1992, 9: 309-347.
- [2] 林士敏,田凤占,陆玉昌.用于数据采掘的贝叶斯分类器研究[J].计算机科学,2000,27(10):73-76.

其审计数据的准确性、及时性、可靠性是入侵检测获得高效的前提。针对现有 Ad hoc 网络入侵检测模型的不足,将簇技术应用于 Ad hoc 网入侵检测中,该技术增强了入侵检测系统自身的安全性和对分布式攻击的协同检测能力,有效降低了网络的通信负荷。

### 参考文献:

- [1] 冯建新,王光兴. MAIDS—Ad Hoc 网络的多层分布式入侵检测系统[J]. 小型微型计算机系统, 2004(12): 2195-2198.
- [2] 易平,蒋巍川,张世永,等. 移动 ad hoc 网络安全综述[J]. 电子学报, 2005(5): 893-899.
- [3] 王松,王卫红,张繁. 一种新的移动 ad-hoc 网络异常入侵检测技术[J]. 浙江工业大学学报, 2004(12): 696-699.
- [4] Zhang Y, Lee W. Intrusion Detection in Wireless Ad-Hoc Networks[C]//Proceedings of the Sixth Annual International Conference on Mobile Computing and Networking. New York, USA: ACM Press, 2000: 275-283.
- [5] 徐有明,曹元大. 集群技术在消息中间件中的应用[J]. 微机发展, 2005, 15(10): 109-111.
- [6] 程洪,钱乐秋,洪圆. 基于 Linux 集群的 Web 服务的研究和构建[J]. 计算机工程与应用, 2004(34): 158-161.
- [7] Zavgren J. NTDR mobility management protocols and procedures[C]//In: Proceedings of the IEEE Military Communications Conference. Boston, USA: IEEE Press, 1997.
- [3] Duda R O, Hart P E. Pattern Classification and Scene Analysis [M]. New York: John Wiley & Sons, 1973.
- [4] Friedman N, Goldszmidt M. Building classifiers using Bayesian network[C]//In proc. Nation Conference on Artificial Intelligence. Menlo park, CA: AAAI Press, 1996: 1227-1284.
- [5] Chow C K, Liu C N. Approximating discrete probability distributions with dependence trees[J]. IEEE Trans. on Info. Theory, 1968, 14: 462-467.
- [6] 程泽凯. 贝叶斯网络结构学习及 MBNC 实验平台的构建[D]. 南宁: 广西师范大学, 2004.
- [7] 程泽凯,林士敏. TANC-BIC 结构学习算法[J]. 微机发展, 2004, 14(11): 10-12.
- [8] Larranaga P, Poza M, Yurramendi Y, et al. Structure Learning of Bayesian networks by genetic algorithms: A performance analysis of control parameters[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 1996, 18(9): 912-925.
- [9] 刘大有,王飞,卢奕南,等. 基于遗传算法的 Bayesian 网络结构学习研究[J]. 计算机研究与发展, 2001, 38(8): 916-922.
- [10] 程泽凯,林士敏,陆玉昌,等. 基于 Matlab 的贝叶斯分类器实验平台 MBNC[J]. 复旦学报, 2004(5): 729-732.