

神经模糊理论在入侵检测系统中的应用研究

周豫苹^{1,2}, 方建安²

(1 漳州师范学院 计算机科学与工程系, 福建 漳州 363000; 2 东华大学 信息科学与技术学院, 上海 201620)

摘 要: 构建了一类二层式神经模糊推理系统的入侵检测模型, 能同时实现误用检测和异常检测. 系统采用可进化的模糊 C 均值聚类法(FCMm)获取并管理模糊规则. 构造底层的基于神经模糊网络的多个并行神经模糊分类器结构, 顶层的 Mamdani 模糊推理系统实现最终入侵行为判断. 仿真结果证明该检测模型的有效性.

关键词: 模糊推理; 入侵检测; 神经网络; 软计算

中图分类号: TP393.08

文献标识码: A

文章编号: 1000-7180(2010)09-0126-04

Application of Neuro-Fuzzy Method in Intrusion Detection System

ZHOU Yu-ping^{1,2}, FANG Jian-an²

(1 Department of Computer Science and Engineering, Zhangzhou Normal University, Zhangzhou 363000, China;

2 School of Information Science and Technology, Donghua University, Shanghai 201620, China)

Abstract: This paper constructs a Hierarchical Neuro-Fuzzy Inference intrusion detection system which can implement either misuse or anomaly detection. In the proposed system the input data is clustered by applying an enhanced Fuzzy C-Means clustering algorithm to create and extract fuzzy rules. The system developments two level neuro-fuzzy inference system to detect intrusion. Finally a publicly available DRAPA/KDD99 dataset is used to demonstrate the approaches and the results show their accuracy.

Key words: fuzzy inference; intrusion detection; neural network; soft computing

1 引言

近几年,一些互相协作的入侵行为不断出现,研究新型高效的入侵检测系统是十分必要的. 软计算(Soft computing, SC)作为一种创建计算机智能系统的新颖方法,与人脑相对应,具有在不确定环境中进行推理和学习的卓越能力^[1]. 在解决实际计算问题时,协同地而不是互斥地采用多种软计算技术通常具有较大的优越性^[2],所产生的系统被称为互补的混合智能系统. 神经-模糊计算是这类智能系统的精髓,其中神经网络负责识别模式和按变化的环境进行自适应调节,模糊推理系统包含对人类知识进行推理和决策. 文中构建一类层次式神经模糊推理入侵检测系统,能同时实现误用和异常检测,具有较好动态更新模糊推理规则的能力,并进一步提高

在线监测网络数据流的准确性.

2 系统结构

系统由两个主要模块组成:规则创建和管理模块(RCM)和一个两层次的模糊推理分类器模块(FIS),结构如图 1 所示. 其中 RCM 模块由可进化的 C 均值模糊聚类(ECMm)算法得到的聚类结果作为输入,获取并管理模糊规则,这里每一个聚类群代表一类模糊规则,并衍生出若干条具有不同置信度 CF 的模糊规则,把创建的模糊规则加入对对应分类器构建神经-模糊分类器的结构. 在选择规则时,根据隶属度随机初始化置信度,在特定连接中计算置信度乘积,当置信度值超过一定阈值,则模糊规则被选择. 在检测入侵时,对于在训练中没有出现过的新入侵类型,则 ECMm 算法将产生新的聚类群,并

收稿日期: 2009-07-09; 修回日期: 2009-09-15

基金项目: 国家自然科学基金项目(60874113);漳州师范学院科学研究项目(SK09006)

用于在 FIS 模块中训练新的分类器。

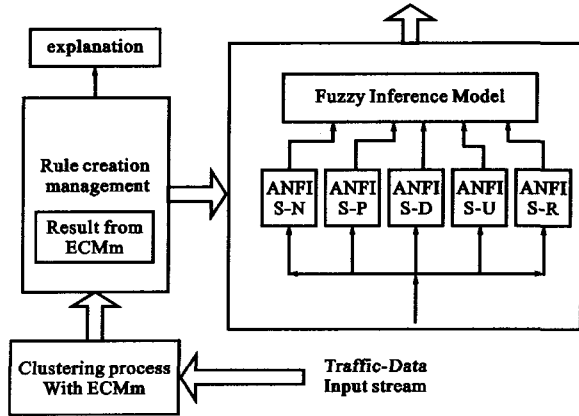


图 1 HNFIS 入侵检测模型系统结构

3 两层次模糊推理分类器

3.1 神经模糊推理

自适应神经模糊推理系统 (Adaptive Neural - Network - Based Fuzzy Interference System, ANFIS) 是一种将模糊逻辑和神经元网络有机结合的新型的模糊推理系统结构,采用反向传播算法和最小二乘法的混合算法调整前提参数和结论参数,并能自动产生 If - Then 规则. Takagi 和 Sugeno 于 1985 年提出了一种 T - S 模糊模型,后来研究者称之为 Sugeno 模糊模型,它是一种非线性模型,宜于表达复杂系统的动态特性,也是最常用的模糊推理模型. 其典型模糊推理规则为:

If x is A and y is B Then $z = f(x, y)$

其中: A 和 B 作为前提的模糊数, $z = f(x, y)$ 为结论中的精确数. 通常 $f(x, y)$ 为 x 和 y 的多项式. 当 $f(x, y)$ 为一阶多项式时,模型称为一阶 Sugeno 模糊模型. 图 2 所示为有两个规则的一阶 Sugeno 模糊推理系统.

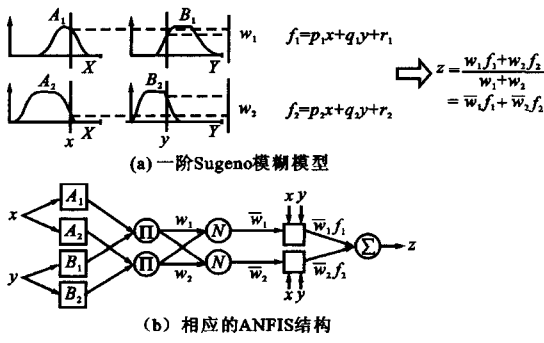


图 2 有两个规则的一阶 Sugeno 模糊推理系统

规则 1: If x is A_1 and y is B_1 , Then $f_1 = p_1x + q_1y + r_1$

规则 2: If x is A_2 and y is B_2 , Then $f_2 = p_2x + q_2y + r_2$

图 2(a) 中输入向量为 $[x, y]$, 权重 w_1 和 w_2 通常由前提中的隶属函数 μ 值乘积得来, 输出 f 为各规则输出的加权平均, $\bar{w}_1$ 和 $\bar{w}_2$ 为各权重在总权重中的比例. 为了实现 Sugeno 模糊模型的学习过程, 一般将其转化为一个自适应网络, 即 ANFIS 见图 2(b). 该自适应网络是一个多层前馈网络, 其中的方形节点需要进行参数学习. 其参数学习可以采用梯度下降法, 但由于该方法速度较慢且易陷入局部最小, 这里采用梯度下降法与最小二乘法的混合算法. 如图 2(b) 所示, 第一层: 负责输入信号的模糊化, 节点 i 具有输出函数: $O_i^1 = \mu_{A_i}(x)$; 其中 x 是节点 i 的输入, A_i 是模糊集. O_i^1 就是 A_i 的隶属函数值, 表示 x 属于 A_i 的程度, 采用高斯 MF 表示, $\text{gaussian}\{x; c, \sigma\} = e^{-\frac{1}{2} \cdot \frac{(x-c)^2}{\sigma}}$, 这里 $\{c_i, \sigma_i\}$ 为前提参数. 第二层: 这一层的节点负责将输入信号相乘. 如: $w_i = \mu_{A_i}(x) \times \mu_{B_i}(y)$, $i = 1, 2$. 每个节点的输出代表一条规则的可信度. 这里的“ $\times$ ”可以是任何满足 T 规范的 AND 算子. 第三层: 第 i 个节点计算第 i 条规则的归一化可信度 $\bar{w}_i = \frac{w_i}{w_1 + w_2}$, $i = 1, 2$. 第四层: 第 i 个节点具有输出: $o_i^4 = \bar{w}_i f_i = \bar{w}_i (p_i x + q_i y + r_i)$, 这里 $\bar{w}_i$ 为第三层的输出, $\{p_i, q_i, r_i\}$ 为结论参数. 第五层: 该层计算总的输出: $o_5^5 = \sum \bar{w}_i f_i = \frac{\sum w_i f_i}{\sum w_i}$ 给定前提参数后, ANFIS 的输出可以表示成结论参数的线性组合:

$$f = (\bar{w}_1 x) p_1 + (\bar{w}_1 y) q_1 + (\bar{w}_1) r_1 + (\bar{w}_2 x) p_2 + (\bar{w}_2 y) q_2 + (\bar{w}_2) r_2 \quad (1)$$

混合学习算法可分为两个步骤:

(1) 确定前提参数的初始值, 用最小二乘法计算结论参数. 由式(1) $f = (\bar{w}_1 x) p_1 + (\bar{w}_1 y) q_1 + (\bar{w}_1) r_1 + (\bar{w}_2 x) p_2 + (\bar{w}_2 y) q_2 + (\bar{w}_2) r_2 = A \cdot X$ 式中, 列向量 X 的元素构成结论参数集合 $\{p_1, q_1, r_1, p_2, q_2, r_2\}$. 若已有 P 组输入输出数据对, 且给定前提参数, 则矩阵 A 、 X 、 f 的位数为 $P \times 6, 6 \times 1, P \times 1$. 一般的, 样本数据个数远大于未知参数的个数 ($P \gg 6$), 使用最小二乘法可以得到均方误差最小 ($\min \|AX - f\|$) 意义下的结论向量的最佳估计 X^* 即

$$X^* = (A^T A)^{-1} A^T f.$$

(2) 根据上一步骤计算得到的结论参数进行误

差计算,采用前馈神经网络中的 BP 算法,将误差由输出端反向传到输入端,用梯度下降法更新前提参数,从而改变隶属函数的形状.

3.2 层次推理系统

FIS 模块是一个两层结构的模糊推理系统^[3],底层由 5 个并行的具有单输出的神经-模糊分类器 ANFIS 构成,它们分别是 ANFIS-N 分类器用于检测正常的数据连接,ANFIS-P 分类器用于检测 Probe 类型攻击,ANFIS-D 分类器用于检测 DOS 类型攻击,ANFIS-U 分类器用于检测 U2R 类型的攻击,而 ANFIS-R 分类器用于检测 R2L 类型的攻击. ANFIS 分类器的输出表示输入数据对应特定类型的隶属度,其取值在 $[-1, 1]$ 范围,1 代表对应类型的确定, -1 代表对应类型的完全否定,输入输出模糊集的隶属函数为高斯函数,训练数据集可通过多次重复学习来训练分类器,使具有最小检测错误的最优 FIS 分类器产生. 顶层采用一个五输入、单输出的 Mamdani 模糊推理系统最终决定入侵行为是否出现. Mamdani 系统应用面积中心法去模糊化策略做最终决策引擎. 五个输入分别来自底层的五个 ANFIS 分类器,每个输入模糊集有二个语言变量值:high 和 low,分别使用四参数的 Gaussian 函数. Mamdani 推理系统的输出表征入侵行为出现的可能性程度,取值范围 $[-1, 1]$,1 代表完全的入侵行为, -1 代表正常行为. 以此类推,正值表示当前模式是入侵模式,当入侵模式被确定,则底层具有最高取值的 ANFIS 分类器决定了入侵的具体类型. Mamdani 推理系统中使用的基于专家经验的模糊规则如表 1 所示.

表 1 顶层模糊推理系统的模糊规则

Normal	PROBE	Dos	U2R	R2L	Output
high	-	-	-	-	Normal
-	low	low	low	low	Normal
-	high	-	-	-	Acctack
-	-	high	-	-	Acctack
-	-	-	high	-	Acctack
-	-	-	-	high	Acctack
low	-	-	-	-	Acctack

4 可进化的 C-均值模糊聚类算法

模糊推理系统的精华就体现在“分而治之”,即模糊规则定义了局部模糊区域,而后件通过不同的

组成成分描述该区域内的行为. 设计模糊推理系统需要划分输入空间去创建模糊规则. 当模糊推理系统要求能实现动态创建和动态修改,则输入空间的划分就必须满足在线的模式. 我们采用一种在线可进化的 C-均值模糊聚类(ECM)算法对动态神经模糊推理系统输入空间进行划分. ECM 是一种基于距离的快速、一步聚类算法,其聚类中心以在线的方式不断地进化更新. 聚群的最大距离 MaxDist 指聚群中样本点与聚群中心距离的最大值,MaxDist 应小于或等于 Dthr, 而 Dthr 是一个直接影响聚群个数的聚类参数. 算法的基本思想:在聚类过程中,聚类样本点来自网络数据流. 初始时聚群集合为空,当一个新的聚群产生时,它的聚群中心 C_c 被设置,聚群半径 R_u 初始化为 0. 随着样本点一个接一个出现,一些已经创建的聚群将通过更改聚群中心的位置和增加聚群半径来改变状态. 而当前样本点的位置决定了哪些聚群将被改变及如何改变. 如果某个聚群的半径 R_u 达到了特定的阈值 Dthr,则聚群将不再被改变. 在线可进化的 C-均值模糊聚类(ECM)算法:

(1) 如果不是初次,则初始化已用 ECM 聚类算法求得的聚类中心点 C_{cj} , $j = 1, 2, \dots, n$, 否则转 (6) 创建一个新的聚群.

(2) 生成成员隶属矩阵 U , 其中每个矩阵元素 u_{ij} , 根据 $IF P_{x_i} - C_{cj}P \leq P_{x_i} - C_{ck}P$, for $k = 1, 2, \dots, n, j \neq k$, Then $u_{ij} = 1$, else $u_{ij} = 0$.

(3) 应用 Constrained Minimization method^[4] 修改聚群中心点,约束条件为: $P_{x_k} - C_{cj}P \leq Dthr$.

(4) 通过式子 $J = \sum_{j=1}^n J_j$ 计算目标函数 J , 其中 $J_j = \sum_{k, x_k \in c_j} P_{x_k} - C_{cj}P$ 是每个聚群中心点 c_j ($j = 1, 2, \dots, n$) 的对应目标函数.

(5) 如果满足结束条件如相邻结果值之差小于某阈值或循环次数达到规定值等,则算法转 (7), 否则算法转 (2).

(6) 通过从输入数据流中取第一个样本点作为聚群中心 C_{c1} , 创建一个新的聚群 C_1 (如果是第一次), 聚群半径 R_{u1} 初始化为 0.

(7) 如果来自数据流中的所有样本点都处理过,则聚类处理结束,否则当前样本点 x_i 被获取并计算该样本点和所有 n 个已经创建的聚群中心 C_{cj} 的 Euclidean 距离 $D_{ij} = P_{x_i} - C_{cj}P$, $j = 1, 2, \dots, n$.

(8) 如果聚群 C_m 具有聚群中心 C_{cm} , 聚群半径

R_{um} 和距离值 D_{im} , 且满足 $D_{im} = Px_i - C_{cm}P = \min\{D_{ij}\} = \min\{Px_i - C_{cj}P\}$, for $j = 1, 2, \dots, n$; 和 $D_{im} < R_{um}$ 则当前样本点 x_i 属于聚群 C_m , 在这种情况下, 既没有新的聚群产生也没有任何聚群被更改, 算法转(7).

(9) 寻找一个聚群 c_a , 具有聚群中心 C_{ca} , 聚群半径 R_{ua} 和距离 D_{ia} , 满足最小值 $S_{ia} = D_{ia} + R_{ua} = \min\{S_{ij}\}$, $j = 1, 2, \dots, n$.

(10) 如果 S_{ia} 大于 $2 \times Dthr$, 则样本点 x_i 不属于任何已存在的聚群. 则算法转(6) 去创建一个新的聚群.

(11) 如果 S_{ia} 不大于 $2 \times Dthr$, 则通过改变中心点 C_{ca} 和半径 R_{ua} 更新聚群 C_a , 使 R_{ua}^{new} 等于 $S_{ia}/2$, 且新的中心 C_{ca}^{new} 移动到样本点 x_i 和原聚群中心点 C_{ca} 的连线上, 新的中心 C_{ca}^{new} 到样本点 x_i 的距离等于 R_{ua}^{new} , 算法转(7).

5 实验

文中采用 KDD-Cup99 的数据集作为 IDS 系统测试的数据, 取全部训练数据的 10% 作为系统的训练数据, 将连接记录所包含的 22 种攻击类型大体分为四种: DOS、U2R、R2L、Probe 和 normal 类. 实验采用数据库中的 kdd-cup.data_10_percent.gz 文件, 其包含有 494,021 实例的训练数据和 correct-ed.gz 文件其包含有 311,029 实例的测试数据. 每个网络连接具有 41 个特征, 其中 38 个数字型属性特征和 3 个符号型属性特征, 为进一步减少误差, 有必要对试验数据进行预处理, 预处理过程为: (1) 把符号型特征属性数据转化成数字型特征属性数据; (2) 采用式(2)归一化数字型特征属性数据到 $[0, 1]$ 范围.

$$X_{ik} = (x_{ik} - \min x_{ik}) / (\max x_{ik} - \min x_{ik}) \quad (2)$$

系统的性能如表 2 所示, 能同时实现误用检测和异常检测, 具有较高检测率, 较小误警率^[4].

6 结束语

文中通过构建了一类层次式神经模糊推理系统

的入侵检测模型^[5], 使入侵检测系统能同时实现误用和异常检测. 由于采用了 ECMm 算法创建和管理模糊规则, 使系统更容易适应实时、在线的计算环境, 具有较强的自学习、自适应能力. 层次式模糊推理系统的应用可以平滑攻击行为和正常行为的边界, 较大程度地避免尖锐边界问题. 实验结果表明: 该检测系统具有较高的检测率和分类率, 较小的误警率, 检测性能良好. 今后研究重点是如何进一步提高系统的自适应性, 实现更高层次的智能化^[6].

表 2 系统检测性能分析表

Attack Class	DR	FAR
Normal	98.19%	0.29%
DOS	97.91%	0.71%
Probe	98.69%	1.23%
R2L	99.21%	0.78%
U2L	99.51%	0.24%

参考文献:

- [1] Lofti A Zadeh. Fuzzy logic, neural networks and soft computing[D]. Berkeley: University of California 1992.
- [2] 周豫苹, 方建安, 于冬梅. 基于遗传模糊分类器的入侵检测系统研究[J]. 微电子学与计算机, 2009, 26(4): 26-28.
- [3] 危胜军, 胡昌振, 孙明谦. 基于模糊 Petri 网的误用入侵检测方法[J]. 北京理工大学学报, 2007, 27(4): 312-317.
- [4] Grace A. Optimization toolbox user's guide[M]. Natick: The MathWorks, Inc. 1996.
- [5] 孙知信, 徐红霞. 模糊技术在入侵检测系统中的应用研究综述[J]. 南京邮电大学学报, 2006, 26(4): 73-77.
- [6] 曹菊英, 赵跃龙, 王明亮. 异常证据及其检测算法研究[J]. 微电子学与计算机, 2009, 26(4): 111-113.

作者简介:

周豫苹 女, (1976-), 博士研究生, 讲师, CCF 会员. 研究方向为计算机网络、数据挖掘.

方建安 男, (1966-), 博士生导师, 教授. 研究方向为智能控制、计算机网络、混沌理论.