

网络入侵检测系统的最优特征选择方法

王 树¹, 杜启军¹, 余桂贤², 余生晨¹, 李广平³, 徐亚飞³, 薛 阳⁴, 王晓伟¹

(1. 华北科技学院计算机系, 北京 101601; 2. 标旗集团, 北京 101028;

3. 皖北煤电集团钱营孜煤矿, 淮北 234000; 4. 北京政法职业学院, 北京 102600)

摘要: 用于网络入侵检测系统(IDS)的特征(变量)数量太多或太少都会降低 IDS 识别入侵者的正确率。为解决这一矛盾, 提出一种选择最优特征的方法。计算每个特征或组合成的新特征对 IDS 的“贡献”值, 选择少量“贡献”值较大的特征(最优特征)作为 IDS 识别入侵者的特征, 既减少特征数量又基本保留了原始特征组所提供的信息。实验证明该方法实用且识别入侵者的正确率较高。

关键词: 入侵检测系统; 最优特征; 反向传播神经网络

Method of Choosing Optimal Characters for Network Intrusion Detection System

WANG Shu¹, DU Qi-jun¹, YU Gui-xian², YU Sheng-chen¹, LI Guang-ping³, XU Ya-fei³, XUE Yang⁴, WANG Xiao-wei¹

(1. Department of Computer, North China Institute of Science and Technology, Beijing 101601;

2. Biao-Qi Co. Ltd., Beijing 101028;

3. Qianyingzi Coal Mine, Wanbei Coal and Electricity Co. Ltd., Huaibei 234000;

4. Beijing Management College of Politics and Law, Beijing 102600)

【Abstract】 Using too many or too too few characters(variable) in Intrusion Detection System(IDS) leads to reduce recognizing correctness of IDS. To resolve the contradiction and to improve the whole performance of IDS, an approach of choosing optimal characters used to IDS is presented. With the approach, new characters made of original characters, “contributions” of new characters for recognizing intruders are computed, and the characters with larger “contributions” value are chosen as the characters of IDS. Number of the characters used to IDS is reduced, and the information belonging to original characters are kept largely to improve recognizing correctness. The characters with larger “contributions” are optimal characters. Tests show that the approach is useful.

【Key words】 Intrusion Detection System(IDS); optimal character; Back Propagate(BP) neural network

1 概述

入侵检测系统(Intrusion Detection System, IDS)是一个试图检测针对一个系统或网络的入侵并发出警报的系统。它包括^[1]: (1)数据收集, (2)特征提取, (3)行为分类, (4)报告和反应。当 IDS 识别入侵者的正确率降低时, 一般会增加特征数量, 以便提供更多的分类识别信息, 但是特征过多会增加计算的复杂性, 降低 IDS 的正确率; 特征数量太少会导致信息量过少, 也会降低 IDS 识别入侵者的正确率^[2]。这是个矛盾。

本文将原始的 n 个特征的信息(IDS 识别入侵者的信息)保留在 m ($m < n$) 个新特征中, 然后用这些在数量上少于 n 的新特征 m 识别入侵者, 从而加快识别速度, 提高识别入侵者的正确率。同时计算这 m 个新特征对 IDS 的“贡献”, 选择那些“贡献”较大的新特征用于 IDS 识别入侵者(摒弃那些“贡献”为负或为零的特征)。“贡献”较大的新特征称为“最优特征”。“贡献”越大, 特征越“优”。实验证明该方法识别入侵者的正确率较高。

2 最优特征选择方法

2.1 最优特征的求法

本文中“最优特征”是指尽可能多地保留 n 个原始特征的信息, 但在数量上又少于 n 的那些新特征 m 。最优特征的求法如下:

n 维的特征向量 x (即 IDS 选取的 n 个原始特征)可以用 n 个基向量的加权和表示:

$$x = \sum_{i=1}^n \alpha_i \times \varphi_i \tag{1}$$

其中, φ_i 为基向量; α_i 为加权系数。

式(1)还可以用矩阵形式表示为

$$x = (\varphi_1, \varphi_2, \dots, \varphi_n) \times \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix} = \Phi a \tag{2}$$

其中, 向量 $\Phi = (\varphi_1, \varphi_2, \dots, \varphi_n)$; 向量 $a = (\alpha_1, \alpha_2, \dots, \alpha_n)^T$ 。

取基向量为正交向量, 即 $\varphi_i^T \varphi_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$

基金项目: 华北科技学院博士基金资助项目“网络入侵检测系统的几个关键问题的解决方法”(A0825)

作者简介: 王 树(1966—), 男, 副教授、博士, 主研方向: 网络入侵检测, 信息安全; 杜启军, 讲师、硕士; 余桂贤, 硕士; 余生晨, 教授、博士后; 李广平、徐亚飞, 工程师; 薛 阳, 讲师; 王晓伟, 硕士研究生

收稿日期: 2010-01-09 **E-mail:** yusc5291@sina.com

Φ 由正交向量构成, 所以, Φ 是正交矩阵, 即 $\Phi^T \Phi = I$, I 为单位矩阵, 将式(2)两边左乘 Φ^T , 并考虑到 Φ 为正交矩阵, 得 $\alpha = \Phi^T x$, 即 $\alpha_j = \Phi_j^T x, j=1, 2, \dots, n$ 。

从 n 个本征向量中取出 m 个组成变换矩阵 A , 即 $A = (\varphi_1, \varphi_2, \dots, \varphi_m), m < n$, 这时 A 是一个 $n \times m$ 维矩阵, x 为 n 维的特征向量, 经过 $y = A^T x$ 变换, 得到降维为 $m (m < n)$ 的新向量。问题是选取哪 m 个本征向量构成变换矩阵 A , 使降维的新向量在最小均方误差准则下接近原来的向量 x 。

对于式(1), 现在只取 m 项, 对略去的项用预先选定的常数 b_j 代替, 这时对 x 的估计值为

$$\hat{x} = \sum_{j=1}^m \alpha_j \times \varphi_j + \sum_{j=m+1}^n b_j \times \varphi_j$$

由此产生的误差为

$$\Delta x = x - \hat{x} = \sum_{j=m+1}^n (\alpha_j - b_j) \times \varphi_j$$

均方误差为

$$\varepsilon^2 = E[\|\Delta x\|^2] = \sum_{j=m+1}^n E[(\alpha_j - b_j)^2] \quad (3)$$

要使 ε^2 最小, 对 b_j 的选择应满足:

$$\frac{\partial \varepsilon^2}{\partial b_j} = \frac{\partial}{\partial b_j} \sum_{i=m+1}^n E[(\alpha_i - b_i)^2] = \frac{\partial}{\partial b_j} E[(\alpha_j - b_j)^2] = E[-2(\alpha_j - b_j)] = 0$$

所以,

$$b_j = E[\alpha_j] \quad (4)$$

这就是说, 对于 α 中省略掉的那些分量, 应该用它们的期望值来代替。

在 $y = A^T x$ 变换之前, 如果将模式(所选用的已知入侵者和正常程序)总体的均值向量作为新坐标系的原点, 即在新坐标系中, $E[x] = 0$, 根据式(4)得

$$b_j = E[\alpha_j] = E[\varphi_j^T x] = \varphi_j^T E[x] = 0$$

这样, 由式(3)给出的均方误差变为

$$\varepsilon^2 = \sum_{j=m+1}^n E[\alpha_j^2] = \sum_{j=m+1}^n E[(\varphi_j^T x)(\varphi_j^T x)^T] = \sum_{j=m+1}^n E[(\varphi_j^T x x^T \varphi_j^T)] = \sum_{j=m+1}^n \varphi_j^T R \varphi_j = \sum_{j=m+1}^n \lambda_j$$

其中, λ_j 是 x 的自相关矩阵 R 的第 j 个本征值; φ_j 是与 λ_j 对应的本征向量。综上所述, 求最优特征的步骤如下:

(1) 平移坐标系, 将模式总体的均值向量作为新坐标系的原点;

(2) 求出自相关矩阵 R (即协方差矩阵);

(3) 求出 R 的本征值 $\lambda_1, \lambda_2, \dots, \lambda_n$ 及其对应的本征向量 $\varphi_1, \varphi_2, \dots, \varphi_n$;

(4) 将本征值从大到小排序, 如 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m \geq \dots \geq \lambda_n$, 取前 m 个大的本征值所对应的本征向量构成变换矩阵:

$$A = (\varphi_1, \varphi_2, \dots, \varphi_m), m < n$$

(5) 通过 $y = A^T x$ 变换, 将 n 维的原特征向量变换为 $m (m < n)$ 个新特征向量。

m 个新特征基本保留了原始的 n 个特征的信息, 但特征数量少。IDS 使用这 m 个新特征可加快其计算速度, 从而提高其识别入侵者的正确率。

$m (m < n)$ 个新特征对 IDS 的作用即“贡献”是不同的。还可根据这 m 个新特征的“贡献”值, 进一步将新特征数量

减少到 $k (k < m < n)$ 个, 以便进一步加快识别速度。

2.2 新特征对 IDS 识别入侵者“贡献”值的计算

上面求出的 $m (m < n)$ 个新特征都具有: 方差 $\text{var}(y_j) = \lambda_j$, 即, $\text{var}(y_j) = \frac{1}{k} \sum_{i=1}^k (y_j^{(i)} - \bar{y}_j)^2 = \lambda_j$, 式中 k 为样本个数。也就是说, λ_j 越大, y_j 的方差也越大。 y_j 的方差大说明样本(所选用的入侵者和正常程序)在 y_j 坐标轴有较大的区分度。这种区分度大说明该新特征 y_j 对 IDS 的“贡献”大。 $\sum_{i=1}^m \lambda_i / \lambda_j$ 描述了第 j 个新特征提取的信息占总信息的份额。本文称此为第 j 个新特征 y_j 对 IDS 的“贡献率”。“贡献率”大说明 $y_j = A^T x$ 综合原始特征 x_1, x_2, \dots, x_n 所含信息的能力越强。当网络流量较大时, IDS 对一些数据包来不及处理而丢弃。这些丢弃的数据中可能包含入侵者, 所以, 提高 IDS 对数据的处理速度也是提高其识别入侵者正确率的重要途径之一。减少 IDS 所使用的特征相当于减少其计算量, 从而改进 IDS 的性能。这时可选取 λ_j 值较大的新特征作为 IDS 所使用的特征, 从而提高其识别入侵者的正确率。

3 应用实验

为了验证本文所研究的方法, 实验中最初选取了 13 个特征: x_1 = 探测操作系统版本次数, x_2 = 敏感数据访问次数, x_3 = 重要端口扫描次数, x_4 = IP 扫描次数, x_5 = 用户的地址, x_6 = 用户的源, x_7 = 超级用户权限访问次数, x_8 = 非法连接数, x_9 = 用户在磁盘上所建目录数量, x_{10} = 用户程序所占线程数, x_{11} = 用户连接建立成功率, x_{12} = 特殊文件访问数量, x_{13} = 在一定时间段内口令错误的次数。

用本文方法在这 13 个特征中选取了 6 个最优新特征作为反向传播神经网络的输入。该神经网络由 3 层组成, 第 1 层由 6 个节点组成, 第 2 层由 5 个节点组成, 最后一层由 1 个节点组成, 其值为 0 代表正常用户, 为 1 代表入侵者。

用 40 个一般正常用户程序和 50 个不同类型的入侵者程序作为训练样本集。

当该神经网络训练成功时, 即其权系数与阈值确定后, 就可对在实际网络中运行的程序进行监控和评价。当网络流量达到 900 Mb/s 或 1 300 Mb/s 时, 几种不同的 IDS 对入侵者的正确识别率测试结果如表 1 所示^[3-4]。

表 1 不同 IDS 对入侵者的正确识别率比较

不同的网络 IDS	网络流量为	网络流量为
	900 Mb/s 时	1 300 Mb/s 时
	识别正确率/(%)	识别正确率/(%)
用本文选择的“贡献”较大的前 4 个最优特征的 BP 神经网络	97	96
用本文选择的一组最优特征 (6 个) 的 BP 神经网络	96	84
用最初 13 个特征的 BP 神经网络	67	64
Cyber Cop Network (NAI 公司)	38	26
Snort	41	32

当网络流量达到 1 300 Mb/s 时, 几种不同的 IDS 对入侵者的正确识别率都严重下降, 但这时采用“贡献”较大的前 4 个最优特征用于检测入侵者时, IDS (BP 神经网络) 对入侵者的正确识别率基本维持不变 (变化不大), 这是因为特征越少, 识别速度越快, 丢弃的数据包就越少。这说明当网络流量越大时, 适当减少特征数量 (用最优特征) 有利于提高 IDS 对入侵者的正确识别率。

(下转第 144 页)