

文章编号:1001-9081(2015)11-3087-05

doi:10.11772/j.issn.1001-9081.2015.11.3087

## 基于树突细胞算法与对支持向量机的入侵检测

梁 鸿, 葛宇飞\*, 陈 林, 王雯娇

(中国石油大学 计算机与通信工程学院, 山东 青岛 266580)

(\* 通信作者电子邮箱 geyfupc@163.com)

**摘要:**针对入侵检测技术在处理大规模数据时存在的高误报率、低训练速度和低实时性的问题,提出了一种基于树突细胞算法与对支持向量机的入侵检测策略(DCTWSVM)。利用树突细胞算法(DCA)对威胁数据进行初始检测,在此基础上利用对支持向量机(TWSVM)进行检测结果的优化处理。为了验证策略的有效性,设计性能对比实验,实验结果表明,相较于DCA、支持向量机(SVM)、反向传播(BP)神经网络,DCTWSVM策略的检测精度提高了2.02%、2.30%、5.44%,误报率分别降低了0.26%、0.46%、0.90%,训练速度相较于SVM提高了两倍且只需耗费极少的训练时间,可以更好地适用于大规模数据下的实时入侵检测环境。

**关键词:**树突细胞算法; 对支持向量机; 入侵检测; 大数据

中图分类号: TP393.08 文献标志码:A

### Intrusion detection based on dendritic cell algorithm and twin support vector machine

LIANG Hong, GE Yufei\*, CHEN Lin, WANG Wenjiao

(College of Computer and Communication Engineering, China University of Petroleum, Qingdao Shandong 266580, China)

**Abstract:** In order to solve the problem that network intrusion detection was weak in training speed, real-time process and high false positive rate when dealing with big data, a Dendritic Cell TWin Support Vector Machine (DCTWSVM) approach was proposed. The Dendritic Cell Algorithm (DCA) was firstly used for the basic intrusion detection, and then the TWin Support Vector Machine (TWSVM) was applied to optimize the first step detection outcome. The experiments were carried out for testing the performance of the approach. The experimental results show that DCTWSVM respectively improves the detection accuracy by 2.02%, 2.30%, and 5.44% compared with DCA, Support Vector Machine (SVM) and Back Propagation (BP) neural network, and reduces the false positive rate by 0.26%, 0.46%, and 0.90%. The training speed is approximately twice as the SVM, and the brief training time is another advantage. The results indicate that the DCTWSVM is suitable for the comprehensive intrusion detection environment and helpful to the real-time intrusion process.

**Key words:** Dendritic Cell Algorithm (DCA); TWin Support Vector Machine (TWSVM); intrusion detection; big data

### 0 引言

入侵检测一直是计算机网络安全中重要的研究热点之一<sup>[1]</sup>。由于当前网络安全威胁形式呈现多样化,因黑客攻击、行业竞争等原因引发的安全问题无一不在威胁着计算机网络下的系统终端用户<sup>[2]</sup>。入侵检测系统( Intrusion Detection System, IDS)是一种集成了入侵行为过程的软件系统,并常与入侵防御系统( Intrusion Prevention System, IPS)并称为入侵检测防御系统( Intrusion Detection Prevention System, IDPS)。在网络环境中,入侵检测的延迟报警并不具备较高的实用性,但由于当前检测技术大都依赖于网络环境下产生的历史审计数据(Audit Data)进行分析,所以实时入侵检测的实现、提高检测正确率与效率也是当下重要的研究问题。

生物免疫系统( Immune System, IS)是生物体内保护生物免受病原体危害及保障生物稳态性的一种免疫机制<sup>[3]</sup>,该系统拥有动态性和自适应性等诸多特性。当病原体侵入人体

后,将会引发免疫细胞的一系列活动来保障人体稳态性<sup>[4]</sup>。近些年通过对危险理论( Danger Theory, DT)的深入研究<sup>[5-6]</sup>,业界开始针对树突细胞( Dendritic Cell, DC)生物学来开拓免疫机制的新思路以应对日益严峻的安全形势<sup>[7]</sup>。由此衍生的树突细胞算法具备多项优势,如良好的实时性、较小的资源需求、较少的训练样本、精简的训练过程和优质的检测精度等。

将机器学习与数据挖掘技术应用在入侵检测领域已经取得了较好的成绩<sup>[8-10]</sup>。支持向量机( Support Vector Machine, SVM)技术作为其中的一项主流技术也取得了较多的研究成果<sup>[11-13]</sup>。SVM是根据 Vapnik 的统计学习理论产生,从二分类的研究衍生到多分类问题的研究,究其原理主要是通过求解空间超平面使分类距离最大化来解决分类最优解<sup>[14]</sup>。传统的 SVM 存在训练算法复杂度较高、计算时间较长等问题,因此应用 SVM 处理入侵检测问题时,需要对其进行算法层次的改进或寻找更为简单有效的核函数来简化运算,例如采用

收稿日期:2015-06-17;修回日期:2015-07-17。

基金项目:国家自然科学基金资助项目(61309024);中央高校基本科研业务费专项资金资助项目(15CX02046A)。

作者简介:梁鸿(1966-),男,四川隆昌人,教授,博士,主要研究方向:高性能计算、计算机网络; 葛宇飞(1991-),男,黑龙江牡丹江人,硕士研究生,CCF 会员,主要研究方向:高性能计算、信息安全、Web 数据库; 陈林(1990-),男,内蒙古呼伦贝尔人,硕士研究生,主要研究方向:高性能计算; 王雯娇(1992-),女,四川隆昌人,硕士研究生,主要研究方向:高性能计算。

对支持向量机算法 (Twin Support Vector Machine, TWSVM) 来提升检测速度;另外也可以将 SVM 与其他算法进行结合优化,例如 Shon 利用遗传算法 (Genetic Algorithm, GA) 来优化传统算法<sup>[12]</sup>。

本文提出了一种基于树突细胞算法和对支持向量机 (Dendritic Cell Twin Support Vector Machine, DCTWSVM) 入侵检测策略。该策略有效提高入侵检测的检测准确率,降低误报率,并且在检测数据量大幅提升的情况下可以有效满足检测的实时性要求。

## 1 树突细胞入侵检测模型

基于危险理论而衍生的树突细胞算法 (Dendritic Cell Algorithm, DCA) 在生物免疫学中突破了传统免疫理论中的“自我-非我”(Self-nonself) 免疫思路,转而采取针对危险信号 (Danger Signal) 的识别应答,这使得算法的适用条件比较广泛,如实时计算或半实时计算下的异步处理环境。

DC 细胞对于存在于生物组织中的信息十分敏感,入侵检测的过程中,DC 细胞的主要存在形式有 3 种:未成熟 DC (immature DC, iDC)、半成熟 DC (semi-mature DC, sDC) 以及成熟 DC (mature DC, mDC),通过界定 DC 的 3 种状态可以定义当前环境是否处于危险或者安全状态。DC 细胞组成结构初始化信息包括:生命周期 (lifespan)、初始 signal 值、信号转换权值矩阵 ( $W$ )。DC 进行状态转换的标准主要依据协同刺激信号 (Co-Stimulatory Molecule, CSM)。算法的处理流程如图 1 所示。在图 1 中,输入的信息包括抗原 (Antigen) 与信号 (Signal),信号即安全信号 (Safe Signal) 与 Danger Signal;除此之外还包括病原体相关分子模型 (Pathogenic Associate Molecular Pattern, PAMP) 以及炎症因子 (Inflammatory Cytokines, IC)。在对输入数据进行处理的过程中,作为专职抗原提呈细胞 (Antigen Presenting Cell, APC) 的树突状细胞负责采集 Antigen 产生的信息进行识别、分析、处理并提呈给相关免疫细胞,利用免疫细胞进行病原体入侵识别。

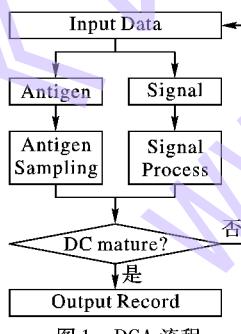


图 1 DCA 流程

在 iDC 采集 Antigen 和 Signal 的过程中,输出信号的计算主要通过下述公式确定:

$$\begin{cases} O_{[csm, semi, mat]} = \frac{WC}{W} \\ WC = W_p * C_p + W_s * C_s + (W_d * C_d) * (1 + IC) \\ W = W_p + W_s + W_d \end{cases} \quad (1)$$

其中,  $O_{[csm, semi, mat]}$  分别代表了 CSM、sDC、mDC 的输出值,  $W_p$  是输入 PAMP 的权值,  $W_s$  是输入安全信号的权值,  $W_d$  是输入危险信号的权值,  $IC$  是炎症因子的值;  $C_p$  是 PAMP 的输入浓度,  $C_s$  是安全信号的输入浓度,  $C_d$  是危险信号的输入浓

度。相关的权值参考表 1。假设状态转移参考阈值为  $T_h$ , 则当  $O_{[csm, semi, mat]}$  大于  $T_h$  时,发生状态转移、将信息输出,反之重新开始采集输入信息。

表 1 基于 DCA 的权值表

输入信号	信号权重		
	csm	semi	mat
$W_p$	2	0	2
$W_s$	1	0	1
$W_d$	2	2	-2

DCA 在异常检测中一项重要的判断标准是上下文成熟度抗原值 (Mature Context Antigen Value, MCAV)。 $MCAV$  代表了在某种环境下完全成熟的抗原数量  $M$  与提呈的抗原总量  $Ag$  的比值,若  $MCAV$  的值接近于 1,则抗原极有可能是异常的,因此  $MCAV$  用于评估输入抗原的异常度。通过界定不同的参考阈值,可以有效提升树突细胞算法的整体检测能力。

$$MCAV = M/Ag \quad (2)$$

$$MCAV_{avg} = \frac{\sum_i M_i}{\sum_i Ag_1^i + \sum_i Ag_2^i} \quad (3)$$

式(2)是 MCAV 求解的标准形式。式(3)是基于式(2)的变形形式,其意义是:由于采集的抗原上下文组合的多样性,若抗原数据在正常状态下收集到抗原上下文,则表示 DC 细胞处于半成熟状态 ( $Ag_1^i$ );若在异常情况下的收集到的抗原上下文,则表示 DC 细胞处于成熟状态 ( $Ag_2^i$ )。 $MCAV_{avg}$  代表了该组序列抗原值。

根据文献[7]的 DCA 形式化描述,树突细胞入侵检测基本的步骤分为 3 个阶段:初始化(第 1 行~3 行)、入侵检测(第 4 行~18 行)、结果分析(第 19 行~23 行)。初始化过程需要设定 DC 细胞数量  $Cell(num)$ 、算法迭代数  $Iteration(max)$ 、以及状态转移阈值  $T_h$ , 经过数据处理、信号转换等过程,最后完成信息提呈,伪代码第 13) 行中的 terminal condition 依据式(1)中  $O_{[csm, semi, mat]}$  的变化而定。

DCA 的过程如下所示:

```

Input: time series data (antigen and signal)
Output: antigen type and MCAV
0) Set Cell(num), Iteration(max), T_h
1) for each DC do
2)   initiate DC
3) endfor
4) for Iteration(max) do
5)   if antigen then
6)     antigen profile update
7)   endif
8)   if signal then
9)     signal transformation
10)  for iDC do
11)    cell lifespan update
12)    signal profile update
13)    if termination condition then
14)      output record
15)    endif
16)  endfor
17) endif
18) endfor
  
```

```

19) for output record do
20)   for antigen type do
21)     calculate MCAV
22)   endfor
23) endfor

```

## 2 基于对支持向量机的入侵检测优化

### 2.1 对支持向量机与入侵检测

传统的 SVM 算法是监督式 (supervised) 的学习方法<sup>[11]</sup>, 在解决非线性分类及高维模式识别等问题中表现出了特有的优势, 在文献 [11–13] 中的研究表明将 SVM 方法应用于入侵检测场景可以收到相对满意的效果。由于支持向量机在训练算法复杂度上并不存在较大的优势, 且算法计算时间较长, 所以若直接利用其来进行入侵检测的离线分析尚且满足要求, 但对于实时性等较高要求, 该方法并不完全满足。关于 TWSVM 与入侵检测, 在文献 [15] 中的研究表明对于传统 SVM, TWSVM 在训练时间上的优势可以有效平衡入侵检测的输出并提高检测率, 但是对于实时性则并未作太多分析。

### 2.2 基于对支持向量机的入侵检测优化

DCA 在很大程度上弥补了 TWSVM 在实时性等方面的优势, 但是在输出结果时存在较高的误报率 (False Positive Rate, FPR)<sup>[7]</sup>。经过分析可得, 产生上述结果的原因主要有以下 3 点: 1) DCA 对于输入数据的序列有一定的依赖性; 2) DCA 对于抗原的危险性需要根据当前设定的参考阈值判断, 且该阈值对于判断结果有直接影响; 3) DCA 对于判断识别率具有一定的随机性<sup>[7]</sup>。对于 1) 本文暂时不予深究, 对于 2) 的影响将通过实验来进行参数优化, 对于 3) 引起的影响将通过 TWSVM 来对 DCA 的检测结果作进一步优化, 从而提高检测结果的准确率, 降低误报率。

#### 2.2.1 对支持向量机

假设在 TWSVM 中需要的两类超平面分别用  $A$  和  $B$  表示, 则 TWSVM 的求解问题可以转化为两个非平行超平面 (nonparallel hyperplane) 问题的求解过程:

$$\begin{cases} \mathbf{x}^T \boldsymbol{\omega}^{(1)} + \boldsymbol{\lambda}^{(1)} = 0 \\ \mathbf{x}^T \boldsymbol{\omega}^{(2)} + \boldsymbol{\lambda}^{(2)} = 0 \end{cases} \quad (4)$$

式(4)代表了正、负两类超平面的最终求解方程。这里,  $\mathbf{x}$  是一个数据集合,  $\boldsymbol{\omega} \in \mathbb{R}^n$  与  $\boldsymbol{\lambda} \in \mathbb{R}$  分别是两个超平面方程的系数;  $\boldsymbol{\omega}^{(1)}$  与  $\boldsymbol{\lambda}^{(1)}$  属于正类的法向量和偏移量,  $\boldsymbol{\omega}^{(2)}$  与  $\boldsymbol{\lambda}^{(2)}$  属于负类的法向量和偏移量。

TMSVM1:

$$\begin{aligned} & \min_{\boldsymbol{\omega}^{(1)}, \boldsymbol{\lambda}^{(1)}, \mathbf{q}} \left[ \frac{1}{2} (\mathbf{A} \boldsymbol{\omega}^{(1)} + \mathbf{e}_1 \boldsymbol{\lambda}^{(1)})^T (\mathbf{A} \boldsymbol{\omega}^{(1)} + \mathbf{e}_1 \boldsymbol{\lambda}^{(1)}) + c_1 \mathbf{e}_2^T \mathbf{q} \right] \\ & \text{s. t. } -(\mathbf{B} \boldsymbol{\omega}^{(1)} + \mathbf{e}_2 \boldsymbol{\lambda}^{(1)}) + \mathbf{q} \geq \mathbf{e}_2; \mathbf{q} \geq 0 \end{aligned} \quad (5)$$

TMSVM2:

$$\begin{aligned} & \min_{\boldsymbol{\omega}^{(2)}, \boldsymbol{\lambda}^{(2)}, \mathbf{q}} \left[ \frac{1}{2} (\mathbf{B} \boldsymbol{\omega}^{(2)} + \mathbf{e}_2 \boldsymbol{\lambda}^{(2)})^T (\mathbf{B} \boldsymbol{\omega}^{(2)} + \mathbf{e}_2 \boldsymbol{\lambda}^{(2)}) + c_2 \mathbf{e}_1^T \mathbf{q} \right] \\ & \text{s. t. } -(\mathbf{A} \boldsymbol{\omega}^{(2)} + \mathbf{e}_1 \boldsymbol{\lambda}^{(2)}) + \mathbf{q} \geq \mathbf{e}_1; \mathbf{q} \geq 0 \end{aligned} \quad (6)$$

式(5)的 TMSVM1 代表求解一个超平面使其拟合正类样本  $\mathbf{A}$  而远离负类样本  $\mathbf{B}$ ; 式(6)的 TMSVM2 代表求解一个超平面使其拟合负类样本  $\mathbf{B}$  而远离正类样本  $\mathbf{A}$ 。 $\mathbf{q}$  为松弛因子且其元素均为 1,  $c_1, c_2 > 0$  为正负两类样本的惩罚因子,  $\mathbf{e}_1$ 、

$\mathbf{e}_2 > 0$  且其元素均为 1。对于测试样本  $\mathbf{x}$ , 计算并比较它到两个超平面的距离, 即可判断该样本所属类别。

规定如下定义:

$$\begin{cases} \mathbf{H} = [\mathbf{A} \quad \mathbf{e}_1] \\ \mathbf{G} = [\mathbf{B} \quad \mathbf{e}_2] \\ \mathbf{u} = [\boldsymbol{\omega}^{(1)} \quad \boldsymbol{\lambda}^{(1)}] \\ \mathbf{v} = [\boldsymbol{\omega}^{(2)} \quad \boldsymbol{\lambda}^{(2)}] \end{cases} \quad (7)$$

根据式(7)中的定义, 可以得到下方程:

$$\begin{cases} \mathbf{H}^T \mathbf{H} \mathbf{u} + \mathbf{G}^T \mathbf{a} = 0 \\ \mathbf{H}^T \mathbf{H} \mathbf{v} + \mathbf{G}^T \mathbf{b} = 0 \end{cases} \quad (8)$$

在式(8)中,  $\mathbf{a}$  和  $\mathbf{b}$  作为拉格朗日乘子向量依据 Wolfe 对偶问题 (DTWSVM)<sup>[16]</sup> 的求解方式如下:

DTWSVM1:

$$\begin{aligned} & \min_{\mathbf{a}} \left[ \mathbf{e}_2^T \mathbf{a} - \frac{1}{2} \mathbf{a}^T \mathbf{G} (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{G}^T \mathbf{a} \right] \\ & \text{s. t. } 0 \leq \mathbf{a} \leq c_1 \mathbf{e}_2 \end{aligned} \quad (9)$$

DTWSVM2:

$$\begin{aligned} & \min_{\mathbf{b}} \left[ \mathbf{e}_1^T \mathbf{b} - \frac{1}{2} \mathbf{b}^T \mathbf{H} (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{H}^T \mathbf{b} \right] \\ & \text{s. t. } 0 \leq \mathbf{b} \leq c_2 \mathbf{e}_1 \end{aligned} \quad (10)$$

从式(9)和式(10)中解出  $\mathbf{a}$  和  $\mathbf{b}$  的值, 接着根据式(8)可以求出  $\mathbf{u}$  和  $\mathbf{v}$ , 最后利用式(1)和式(7)确定最终的超平面解。在给定样本  $\mathbf{x} \in \mathbb{R}^n$  后, 可以根据式(11)来判断  $\mathbf{x}$  的最终分类:

$$Class_x = \arg \min_{k=1,2} |(\boldsymbol{\omega}^{(k)} \cdot \mathbf{x}) + \boldsymbol{\lambda}^{(k)}| \quad (11)$$

其中  $|\cdot|$  运算表示样本  $\mathbf{x}$  到超平面的垂直距离。

传统的 SVM 算法训练问题本质上就是求解一个二次规划 (Quadratic Programming, QP) 问题, 且时间复杂度在给定样本数为  $m$  后的上限为  $O(m^3)$ <sup>[17]</sup>。比较而言, TWSVM 算法将原本求解的大问题转成两个二次规划问题, 缩小了每个子问题的规模。若每类样本规模数量为  $m/2$ , 则近似的时间复杂度为  $O(m^3/4)$ , 相较于传统 SVM 算法展现了绝对的时间优势。

#### 2.2.2 基于对支持向量机的入侵检测优化算法

鉴于 TWSVM 的分类精度高和训练速度快的优势, 本文利用 TWSVM 对 DCA 的检测结果进行更深层次的优化处理, 同时针对惩罚因子  $c_1, c_2$  通过实验进行参数优化, 进一步提高算法性能。

检测优化的 TWSVM 描述如下 (假设训练集样本中的种类为  $n$ ):

步骤 1 设置  $c_1, c_2$  的初始值。

步骤 2 训练算法。训练分类器  $TWSVM_1$ , 得到两个超平面  $\Pi_1$  和  $\Theta_1$ ; 第  $i$  个分类器  $TWSVM_i$  将第  $i$  类训练样本的类别标记为 +1, 而降其余所有训练样本的类别标记为 -1, 得到的超平面是  $\Pi_i$  和  $\Theta_i$ ; 直至构建第  $n-1$  个分类器  $TWSVM_{n-1}$ 。

步骤 3 测试。将树突细胞检测结果样本经过所有 TWSVM 分类器进行分类, 计算样本  $\mathbf{x}$  到  $TWSVM_i$  的两个超平面  $\Pi_i$  和  $\Theta_i$  的距离为  $d_1$  和  $d_2$ , 若  $d_1 > d_2$ , 则样本  $\mathbf{x}$  被判定为第  $i$  类, 继续遍历直到样本中的所有数据都被判定类别后停止。

### 3 实验与分析

本文采用的是 KDD Cup (1999) 的 10% 数据子集, KDD 数据集是目前应用于计算机网络入侵检测研究中普遍采用的测试数据集。该数据集包含训练集数据 4 898 431 条和测试集数据 311 029 条, 除去正常数据之外, 所有的攻击数据包括以下 4 类: 拒绝服务 (Denial of Service, DoS)、权限提升 (User to Root, U2R)、远程权限获取 (Remote to Local, R2L), 以及端口漏洞扫描 (Probe)。在训练集数据中, 只有 19.85% (约 972 781 条数据) 是正常网络流量数据, 其他均为攻击数据; 在测试集数据中, 有 19.48% (约 60 593 条数据) 是正常的网络流量数据而其他的均为攻击数据。在 KDD Cup 数据集中的每一条记录都可以用 41 种定量且定性的特征进行约束。本文从训练集中选出代表性数据 86 990 条, 从测试集中选出数据 43 130 条, 关于数据描述详见表 2。同时作为算法对比测试, 本文采取单独使用 DCA、SVM、反向传播 (Back Propagation, BP) 神经网络, 与本文使用的 DCTWSVM 一同进行数据的训练和测试。实验参数如下:  $Cell(num) = 200$ ,  $Iteration(max) = 50$ ,  $T_h = 0.65$ ,  $c_1 = 1000$ ,  $c_2 = 1000$ 。实验环境是 Intel Xeon CPU 2.60 GHz, 内存是 32 GB, 所有策略算法均采用 C++ 实现。

表 2 基于 DCA 的权值表

数据类型	训练集数据量	测试集数据量
Normal	41 440	20 680
DoS	39 600	19 500
U2R	450	240
R2L	2 400	1 165
Probe	3 100	1 545

关于评价入侵检测系统的性能, 有相关定义如下:

- 1) 真正 TP (True Positive) 预测为正的正样本;
- 2) 假正 FP (False Positive) 预测为负的负样本;
- 3) 真负 TN (True Negative) 预测为正的负样本;
- 4) 假负 FN (False Negative) 预测为负的正样本。

根据上述定义, 可以计算检测精度 (Acc) 和误报率 (False Positive Rate, FPR):

$$Acc = \frac{\sum_i (TP_i + TN_i)}{\sum_i (TP_i + FP_i + TN_i + FN_i)} \times 100\% \quad (12)$$

$$FPR = FP / (FP + TN) \times 100\% \quad (13)$$

表 3 和表 4 是训练集和测试集在给定的四种检测方法上进行实验后得出的测试结果, 包括 Acc 和 FPR 的平均值。

表 3 入侵检测算法性能比较结果 1 %

检测策略	DCTWSVM		DCA	
	Acc	FPR	Acc	FPR
Normal	99.33	0.52	98.21	0.74
DoS	99.57	0.55	99.26	0.63
U2R	59.95	1.72	14.71	3.57
R2L	79.62	2.33	29.53	5.89
Probe	96.51	0.65	94.95	0.76
平均	98.59	0.59	96.57	0.85

由表 3 和表 4 中的平均水平来看, 与 DCA、SVM、BP 神经

网络相比, 本文的 DCTWSVM 在检测精度方面分别提高了 2.02%、2.30%、5.44%, 在误报率方面分别降低了 0.26%、0.46%、0.90%。

综合分析, DCTWSVM 展现了较高的检测精度, 在处理 DoS、Probe 时相比其他策略均有小幅提高, 在处理 R2L 与 U2R 的检测精度相比 SVM 有一定提升、比 DCA、BP 神经网络有较大提升; 与此同时 DCTWSVM 取得了较低的误报率, 其中 U2R 和 R2L 的误报率相比 DCA 有十分明显的降低。

表 4 入侵检测算法性能比较结果 2 %

检测策略	SVM		BP	
	Acc	FPR	Acc	FPR
Normal	96.79	0.89	95.60	1.17
DoS	98.35	0.84	96.27	1.34
U2R	53.22	4.21	10.28	5.91
R2L	75.54	6.15	15.83	7.54
Probe	94.97	1.57	92.14	2.37
平均	96.29	1.05	93.15	1.49

图 2 展示了在训练集比例从 20% 增加到 100% 后, 4 种策略的训练时间对比。

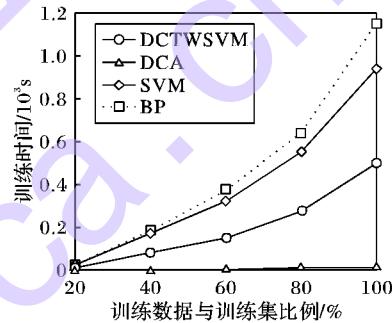


图 2 算法的训练时间对比

从图 2 中看到, DCA 的训练时间很少, 这主要与该算法需要较少的训练样本有关<sup>[7]</sup>。DCTWSVM 的训练时间呈现了比 SVM、BP 神经网络算法更大的优势, 尤其当训练数据规模较大时, 训练速度几乎为 SVM 的两倍。

综合误报率、检测精度和训练时间可以得出: 虽然 DCA 根据其基本原理可以在训练速度上占优<sup>[7]</sup>, 但是本文提出的 DCTWSVM 可以在此基础之上进一步提高入侵检测的检测精度并有效降低误报率。在实际运用中, 若对于实时性要求较高且用于检测 DoS 以及 Probe 攻击时, 可以单独运用 DCA; 但是当数据规模较大且检测种类较多时, 使用 DCTWSVM 可以在牺牲较少的训练时间基础上进一步提高检测精度, 降低整体误报率, 并最终提升检测的整体实时性, 这使得 DCTWSVM 在复杂的应用场景中具备较高的参考价值。

### 4 结语

本文鉴于 DCA 在处理入侵检测过程中具备的较高实时性的优势, 结合 TWSVM 多类分类思想, 提出了一种基于 DCTWSVM 的入侵检测策略。将 DCA 在入侵检测后可能存在的误报率较高的检测结果利用 TWSVM 训练效率高、分类精度高的特点进行结果优化。实验表明, DCTWSVM 不仅保持了较高的检测精确度、较低的误报率, 且在训练速度上相比

一些传统算法有了显著提高,另外在实时性检测能力上有了明显提升,具有一定的实用价值。由于本文仅在 KDD Cup 数据集上进行了对比实验,在今后的工作中,要加强对网络动态环境下产生的数据进行研究;加强 DCA 的优化,降低其检测的误报率;加强对 TWSVM 的优化,进一步减少其训练时间;另外考虑采取更为高效的分类算法与 DCA 进行组合解决复杂网络环境下的入侵检测问题。

#### 参考文献:

- [1] LIAO H, LIN C, LIN Y, et al. Intrusion detection system: a comprehensive review[J]. Journal of Network and Computer Applications, 2013, 36(1): 16–24.
- [2] KREUTZ D, RAMOS F M V, ESTEVES VERISSIMO P, et al. Software-defined networking: a comprehensive survey[J]. Proceedings of the IEEE, 2015, 103(1): 14–76.
- [3] AICKELIN U, DIPANKAR D, FENG G. Artificial immune systems [M]. Berlin: Springer, 2014: 187–211.
- [4] HUA Y, LI T, HU X, et al. A survey of artificial immune system based intrusion detection[J]. The Scientific World Journal, 2014, 2014(3): 156790.
- [5] FANG X, WANG L, KANG J, et al. On dendritic cell algorithm and its theoretical investigation[J]. Computer Science, 2015, 42(2): 131–133. (方贤进, 王丽, 康佳, 等. 树突细胞算法及其理论研究[J]. 计算机科学, 2015, 42(2): 131–133.)
- [6] SALMON H M, FARIA S M D, LOUREIRO P, et al. Intrusion detection system for wireless sensor networks using danger theory immune-inspired techniques[J]. International Journal of Wireless Information Networks, 2013, 20(1): 39–66.
- [7] FENG G, GREENSMITH J, AICKELIN U. The dendritic cell algorithm for intrusion detection[M]. IGI Global: Bio-Inspired Communications and Networking, 2011: 84–102.
- [8] ZONG W, HUANG G, CHEN Y. Weighted extreme learning machine for imbalance learning[J]. Neurocomputing, 2013, 101(3): 229–242.
- [9] LIN W J, CHEN J J. Class-imbalanced classifiers for high-dimensional data[J]. Briefings in Bioinformatics, 2013, 14(1): 13–26.
- [10] HE Q, LI N, LUO W, et al. A survey of machine learning algorithms for big data[J]. Pattern Recognition and Artificial Intelligence, 2014, 27(4): 327–336. (何清, 李宁, 罗文娟, 等. 大数据下的机器学习算法综述[J]. 模式识别与人工智能, 2014, 27(4): 327–336.)
- [11] JAEHAK Y, LEE H, KIM M S, et al. Traffic flooding attack detection with SNMP MIB using SVM[J]. Computer Communications, 2008, 31(17): 4212–4219.
- [12] JIANG C, ZHANG G, LI Z. Abnormal intrusion detection for embedded network system based on genetic algorithm optimised SVM [J]. Computer Applications and Software, 2011, 28(2): 287–289. (姜春茂, 张国印, 李志聪. 基于遗传算法优化 SVM 的嵌入式网络系统异常入侵检测[J]. 计算机应用与软件, 2011, 28(2): 287–289.)
- [13] NIE P, ZANG L, LIU L. Application of multi-class classification algorithm based on twin support vector machine in intrusion detection[J]. Journal of Computer Applications, 2013, 33(2): 426–429. (聂盼盼, 藏洌, 刘雷雷. 基于对支持向量机的多类分类算法在入侵检测中的应用[J]. 计算机应用, 2013, 33(2): 426–429.)
- [14] VLADIMIR V. The nature of statistical learning theory[M]. Berlin: Springer Science & Business Media, 2000: 2–6.
- [15] HE J, ZHENG S. Intrusion detection model with twin support vector machines[J]. Journal of Shanghai Jiaotong University: Science, 2014, 19(4): 448–454.
- [16] MANGASARIAN O L. Nonlinear programming[M]. [S. l.]: Society for Industrial and Applied Mathematics, 1993: 10.
- [17] KURT A K, HALL L O, GOLDGOF D B, et al. Fast support vector machines for continuous data[J]. IEEE Transactions on Systems, Man, and Cybernetics, 2009, 39(4): 989–1001.

(上接第 3058 页)

- [6] TANG F, TANG H, QIAHAN H, et al. A network performance assessment method based on passive measurement[J]. Computer and Digital Engineering, 2013, 282(4): 602–604. (唐飞, 唐海娜, 恰汗·合孜尔, 等. 一种基于被动测量的网络性能评估方法[J]. 计算机与数字工程, 2013, 282(4): 602–604.)
- [7] MENG X, MENG X. Network health evaluation based on SVM and cloud model [J]. Journal of Beijing University of Posts and Telecommunications, 2012, 35(1): 10–14. (温祥西, 孟相如. 基于支持向量机和云模型的网络健康状态评估[J]. 北京邮电大学学报, 2012, 35(1): 10–14.)
- [8] TONG X. A network based estimation method of business system [J]. Journal of Chongqing University of Technology: Natural Science, 2012, 26(8): 101–105, 126. (童晓薇. 一种基于网络的业务系统健康度评估方法[J]. 重庆理工大学学报: 自然科学版, 2012, 26(8): 101–105, 126.)
- [9] MA L. Study and use of modern statistical analysis methods (the third): the standardization of statistical data — dimensionless method[J]. Beijing Statistics, 2000(3): 34–35. (马立平. 现代统计分析方法的学与用(三): 统计数据标准化——无量纲化方法[J]. 北京统计, 2000(3): 34–35.)
- [10] WANG L, ZHANG R, SHA C, et al. A product normalization method for E-commerce [J]. Chinese Journal of Computers, 2014, 37(2): 312–325. (王立, 张蓉, 沙朝峰, 等. 电子商务商品归一化方法研究[J]. 计算机学报, 2014, 37(2): 312–325.)
- [11] YUAN H. Intelligent network management system: research on network performance evaluation model based on BP algorithm [D]. Guangzhou: Guangdong University of Technology, 2006: 33–50. (原慧琴. 智能网络管理系统——基于 BP 算法的网络性能评估模型的研究[D]. 广州: 广东工业大学, 2006: 33–50.)
- [12] CAO W, CHEN G, NIU G, et al. A summarization on technologies of network device test[J]. Electronics Optics and Control, 2014, 21(6): 13–18. (曹文斌, 陈国顺, 牛刚, 等. 网络设备测试技术综述[J]. 电光与控制, 2014, 21(6): 13–18.)
- [13] LUO Y, XIA J, CHEN T. Comparison of objective weight determination methods in network performance evaluation [J]. Journal of Computer Applications, 2009, 29(10): 2624–2626. (罗赟骞, 夏靖波, 陈天平. 网络性能评估中客观权重确定方法比较[J]. 计算机应用, 2009, 29(10): 2624–2626.)