

量子粒子群和最小二乘支持向量机相结合的网络异常检测

姚 晔

(辽宁行政学院, 辽宁 沈阳 110161)

摘 要: 为了提高网络安全性的异常入侵检测的准确率, 提出一种量子粒子群算法(QPSO)优化最小二乘支持向量机(LSSVC)的网络异常检测方法(QPSO-LSSVC). 首先利用量子粒子群处算法对 LSSVC 模型参数进行搜索, 选出最优参数, 然后采用泛化能力优异的 LSSVC 对网络入侵进行建模和检测. 选取 KDDCUP99 数据对 QPSO-LSSVC 性能进行测试, 实验结果表明, QPSO-LSSVC 提高了网络异常检测准确率, 降低了误报率, 为网络安全提供了有效保证.

关键词: 量子粒子群算法; 最小二乘支持向量机; 网络异常; 检测

中图分类号: TP393

文献标识码: A

文章编号: 1000-7180(2012)03-0039-04

Network Anomaly Detection by Combination of QPSO and LSSVM

YAO Ye

(Liaoning School of Administration, Shenyang 110161, China)

Abstract: In order to improve the network security intrusion detection accuracy, proposes a quantum particle swarm optimization (QPSO) optimized least square support vector machine (LSSVC) network anomaly detection method (QPSO-LSSVC). The first use of the quantum particle swarm algorithm for the parameters of LSSVC model search, optimal parameters, and then the generalization ability of excellent LSSVC modeling and detection for network intrusion. Select the KDDCUP99 data on QPSO-LSSVC performance testing, the experimental results show that, QPSO-LSSVC improves the network anomaly detection accuracy, reduce the rate of false positives, provides effective guarantee for network security.

Key words: QPSO; Least Squares Support Vector Machine; network anomaly; detection

1 引言

随着网络技术和信息化飞速发展, 网络环境变得日益复杂, 来了许多复杂难以解决的问题, 各种攻击和入侵方式变化多样, 网络入侵和攻击检测成为安全信息安全研究的一个热点问题^[1].

为了给用户提供更安全可靠的网络服务, 学者们提出了一些网络入侵检测系统(Intrusion Detection System, IDS), 分为误用检测和异常检测两种^[2]. 误用检测采用预先设定入侵模式对其进行模式匹配来检测, 可以针对性建立高效检测系统, 检测准确率高, 但是不能对未知或变异入侵行为进行有效检测^[3]. 异常检测技术利用建立的正常用户和系统的

行为模式对系统异常度进行检测, 能够检测出新的和未知的网络入侵行为^[4-5]. 具有自组织、自学习和容错能力强等的人工神经网络不断成熟, 在网络异常入侵检测中得到了广泛的应用^[5]. 然而, 人工神经网络存在着自身的缺陷, 如: 易陷入局部极值、学习速度慢和过拟合等, 导致网络异常入侵检测准确率, 远不能满足网络入侵检测要求. 支持向量机(Support Vector Machine, SVM)方法能解决人工神经网络的过拟合、局部极值等缺陷, 检测准确率比较高^[6]. 最小二乘支持向量机(Least Squares Support Vector Machine Classifier, LSSVC)是一种改进的支持向量机, 加快了学习和检测效率. 大量实验表明, LSSVC 的检测准确率与训练参数存在一定的

关系. 目前很多训练参数优化方法都有各自的不足, 影响了 LSSVC 在网络入侵检测中的应用^[7].

为提高 LSSVC 在网络入侵检测中的正确率, 提出采用 QPSO 对 LSSVC 参数进行优化的方法 (QPSO-LSSVC), 并通过仿真测试验证算法有效性和可行性.

2 网络入侵检测系统的工作原理

网络入侵检测首先通过对计算机或网络系统关键点信息进行采集, 然后采用一定检测算法, 鉴别该信息是否包含有恶意攻击行为. 网络入侵检测系统具体工作原理如图 1 所示.

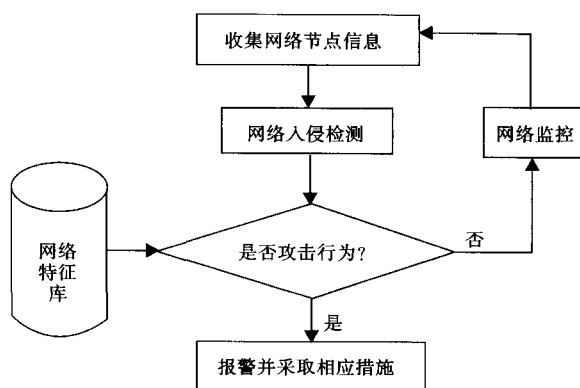


图 1 网络入侵检测系统的工作原理

从图 1 可知, 网络入侵检测系统中最为核心的模块为网络入侵检测算法, 网络入侵数据相当于正常行为数据, 是一种小样本数据, 为了获得更高的网络入侵检测准确率, 本文采用专门针对小样本的非线性分类方法——LSSVC 作为网络入侵检测算法.

3 QPSO-LSSVC 模型

3.1 LSSVC 算法

设共收集到有 n 个训练样本 $\{(x_1, y_1), \dots, (x_n, y_n)\}$, x_i 表示是输入向量, y_i 表示输入向量, $y_i \in \{1, -1\}$, LSSVC 是在 SVM 的基础上, 采用二次损失函数取代不敏感损失函数, 实质上就是对下列函数进行最小化.

$$J(\omega, \xi) = \frac{1}{2} \|\omega\|^2 + \frac{1}{2} \gamma \sum_{i=1}^n \xi_i^2 \quad (1)$$

其中 ξ 表示松弛变量, ω 表示权向量, 约束条件为:

$$y_i(\omega' \varphi(x_i) + b) - 1 + \xi_i = 0 \quad (2)$$

通过引拉格朗日乘子对式(1)进行优化求解, 即

$$L(\omega, b, \xi, \alpha) = J(\omega, \xi) - \sum_{i=1}^n \alpha_i (y_i(\omega' \varphi(x_i) + b) - 1 + \xi_i)$$

$$b) - 1 + \xi_i) \quad (3)$$

其中, α_i 表示拉格朗日乘子.

对式(3)进行优化, 然后对其进行求偏导得到

$$\begin{cases} \frac{dL}{d\omega} = 0 \rightarrow \omega = \sum_{i=1}^n \alpha_i y_i \varphi(x_i) \\ \frac{dL}{d\xi} = 0 \rightarrow \xi_i = \alpha_i / \gamma \\ \frac{dL}{db} = 0 \rightarrow \sum_{i=1}^n \alpha_i = 0 \\ \frac{dL}{d\alpha_i} = 0 \rightarrow y_i(\omega' \varphi(x_i) + b) + \xi_i - 1 = 0 \end{cases} \quad (4)$$

消除 ω 和 ξ_i , 并转换为求解分块矩阵.

$$\begin{bmatrix} 0 & Y' \\ Y & ZZ' + \gamma^{-1} \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ I \end{bmatrix} \quad (5)$$

这样, 最终得到 LSSVC 分类函数.

$$y(x) = \text{sign} \left(\sum_{i=1}^n \alpha_i y_i k(x, x_i) + b \right) \quad (6)$$

其中, $k(x, x_i)$ 表示核函数, 本文采用高斯核函数建立 LSSVC, 高斯核函数定义为:

$$k(x, x_i) = \exp \left(-\frac{\|x - x_i\|^2}{2\sigma^2} \right) \quad (7)$$

3.2 QPSO 对 LSSVC 参数优化

大量研究和实践表明, LSSVC 的分类精度与其参数关系十分密切, 要获得高性能的 LSSVC 模型必须首先获得最适合的参数. 由于当前 LSSVC 参数优化方法均存在各自不足, 建模效果具有一定的局限性. 量子粒子群优化 (Quantum Particle Swarm Optimization, QPSO) 算法是对粒子群优化 (PSO) 算法进行改进, 具有简单易行、全局搜索能力强的优点^[8], 本文采用 QPSO 对 LSSVC 参数进行寻优, 进一步提高 LSSVC 的网络入侵检测性能.

QPSO 算法粒子的速度和位置信息都归结于 1 个参数, 在 D 维搜索空间中具有 m 个粒子, 第 i 个粒子的位置为: $x_i = (x_{i1}, x_{i2}, \dots, x_{id})$, 其搜索到的最优位置为: $P_{besti} = (p_{bi1}, p_{bi2}, \dots, p_{bid})$, 粒子群的最优搜索位置为: $G_{best} = (bg_1, bg_2, \dots, bg_d)$, 粒子在第 d 维的值 pb_d 为:

$$pb_d = \frac{\varphi_1 \times p_{bid} + \varphi_2 \times g_{bg}}{\varphi_1 + \varphi_2} \quad (8)$$

其中, φ_1, φ_2 表示一随机数.

那么, 粒子的进化方程为:

$$x(t+1) = pb \pm \beta \times |m_{best} - x(t)| \ln(1/\mu) \quad (9)$$

其中, β 表示调节算法的收敛速度, m_{best} 表示粒

子群最优中值。

当进行到第 t 次时, β 的值为:

$$\beta = b + c \times (T_{\max} - t) / T_{\max} \quad (10)$$

QPSO 优化 LSSVC 参数的具体步骤为:

(1) 根据问题的规模确定粒子群中的粒子数 m , 对粒子进行初始化处理, 本文的粒子为二维, 分别代表 LSSVC 的参数 γ 和 σ 。

(2) 确定粒子群算法最大迭代次数和其它参数。

(3) 每一个粒子采用 LSSVC 对训练样本进行训练建, 并计算粒子的适应度函数。

(4) 计算粒子的个体和全局极值, 建立公共信息库。

(5) 通过公共信息库更新粒子群最优个体和全局极值。

(6) 采用式(9)和(10)对每一个粒子的位置进行更新。

(7) 迭代次数增加, 并对结束条件进行判断, 如果满足结束条件, 则转到步骤(8), 不然返回到步骤(3)。

(8) 选择最优粒子作为 LSSVC 的参数, 并建立最优网络入侵检测模型。

4 QPSO-LSSVC 的网络异常检测模型

网络入侵包括多种攻击类型, 因此实质上是一个多分类问题, 而 LSSVC 为两分类设计的, 因此必须通过 LSSVC 两类器构件一个多分类器, 才能对网络入侵进行检测, 本文采用一种“一对一”的方式构建多分类器, 具体为:

(1) 首先采用 LSSVC 将数据分成正常和异常两类。

(2) 然后区分未授权远程访问攻击和其它攻击。

(3) 再区分未授权本地访问攻击和拒绝服务、端口扫描攻击状态。

(4) 最后区分拒绝服务、端口扫描攻击状态。

综合上述可知, QPSO-LSSVC 的网络异常检测模型如图 2 所示。

5 仿真实验

5.1 数据来源

仿验数据来自 KDD CUP 99 数据库, 其四类攻击方式: 拒绝服务攻击(DOS); 远程用户未授权访问攻击(U2R); 未授权使用本地超级权限访问攻击(U2L); 扫描攻击(Probe)。从数据库随着选择正常数据为 1000 条, DOS 攻击样本 100 个; R2L 攻击样

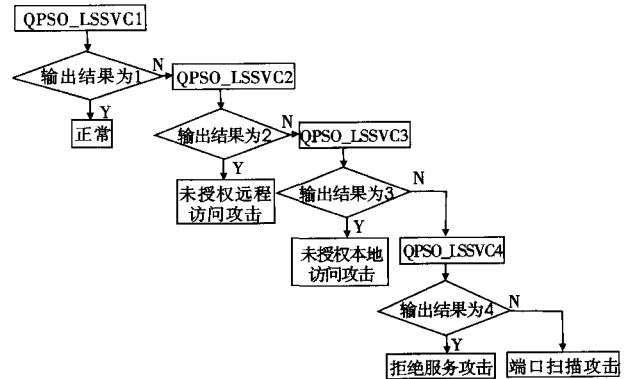


图 2 QPSO-LSSVC 的网络异常检测模型

本 100 个, U2R 攻击样本 120 个, Probe 攻击样本 120 个。

5.2 网络数据预处理

在 KDD CUP 99 数据集中, 样本特征属性范围各异, 而 LSSVC 是根据样本与超平面距离来确定样本的归属, 因此少量特征对 LSSVC 性能影响大, 其它特征影响较小, 因此需要进行预处理来消除这种不利影响, 具体处理方式:

$$x'_i = \frac{x_i - \bar{x}}{x_{\text{std}}}, \quad i = 1, 2, \dots, n \quad (11)$$

其中, n 为训练样本数, x_{std} 为样本特征标准差。

5.3 模型比较与性能分析

5.3.1 与 PSO 参数优化方法性能比较

对于相同数据集, 采用 PSO 作为对比模型, 比较四类入侵检测正确率, 和误报率, 对比结果如表 1 所示。根据表 1 可知, 相于标准的粒子群算法(PSO), QPSO 提高了网络入侵检测准确率, 降低检测误报率, 且检测效率明显提高, 表明基于 QPSO 的 LSSVC 检测结果是有效的, 用于网络入侵检测是可行的。

表 1 QPSO-LSSVC 和 PSO-LSSVC 结果对比

攻击方式	QPSO		PSO	
	检测率(%)	误报率(%)	检测率(%)	误报率(%)
DOS	92.75	5.14	91.035	7.45
U2L	70.88	12.75	66.25	19.73
U2R	90.15	6.20	87.45	9.24
Probing	95.47	3.18	90.78	8.95

5.3.2 与神经网络检测模型性能对比

为了验证 LSSVC 的优越性, 采用 BP 神经网络(BPNN)作为参比模型, 同时采用 QPSO 进行参数优化, 两种算法检测准确率如图 3 所示。从图 3 可

知,相对于 QPSO-BPNN 算法,基于 QPSO-LSSVC 算法提高了网络入侵的检测准确率,加快检测速度,结果表明 LSSVC 很好的解决了 BPNN 局部极值和收敛速度慢等缺陷,十分适合于现代网络异常检测的实时性和检测准确率高要求。

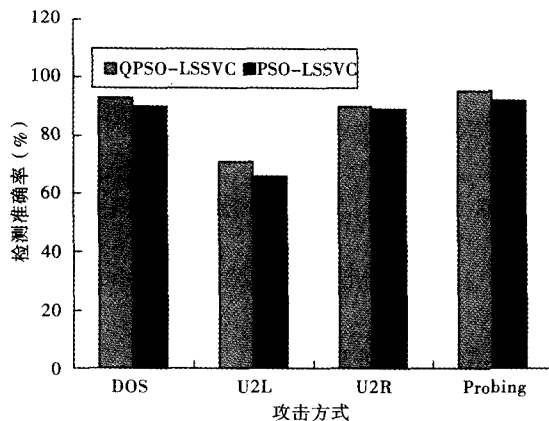


图 3 QPSO-LSSVC 和 QPSO-BPNN 结果对比

6 结束语

针对 LSSVC 在网络入侵检测中参数优化问题,文中提出采用 QPSO 对 LSSVC 参数进行优化,并通过仿真实验进行验证。仿真实验结果表明,QP-SO-LSSVM 解决传统 LSSVC 模型参数优化的难题,提高网络异常入侵检测检测准确率,降低误报率,加快了网络入侵检测速度,在网络安全应用中具有广泛的应用前景。

(上接第 38 页)

化后对规则直接匹配,避免了对数据字符串的逐一匹配,大大缩短了检测耗时,提高了检测效率,实现了高速网络中入侵的快速检测,满足了人们对网络入侵检测的实时性要求。

5 结束语

本文提出一种基于 ACBM 规则匹配的网络入侵算法。通过定义入侵规则,将入侵数据与数据库中的数据规则化,然后根据 ACBM 算法实现入侵规则的匹配,这样就避免了传统算法对单一数据串的逐一匹配,大大缩短了高速网络检测所消耗的时间。实验证明,这种方法能够有效提高网络入侵检测的效率,对高速网络的检测效率也有了很大的改善,满足了人们对检测实时性的要求,具有很高的实用价值。

参考文献:

- [1] Taeshik Shon, Jongsub Moon. A hybrid machine learning approach to network anomaly detection[J]. Information Sciences, 2007, 177 (18): 3799-3821.
- [2] Ozgur Depren, Murat Topallar, Emin Anarim, et al. An intelligent intrusion detection system (IDS) for anomaly and misuse detection in computer networks[J]. Expert Systems with Applications, 2005, 29(4): 713-722.
- [3] 赵俊忠. 入侵检测系统中检测技术的研究[J]. 计算机工程与应用, 2005(2): 11-13.
- [4] 卿斯汉, 蒋建春, 等. 入侵检测技术研究综述[J]. 通信学报, 2004, 25(7): 9-29.
- [5] 刘渊, 马汝辉, 林星. 基于 QPSO 小波神经网络的网络异常检测[J]. 宁波工程技术大学学报: 自然科学版, 2009, 28(2): 261-264.
- [6] 许劲松, 覃俊. 一种基于支持向量机的入侵检测模型[J]. 计算机仿真, 2004, 22(5): 43-55.
- [7] 李旭芳, 王士同. 基于 QPSO 训练支持向量机的网络入侵检测[J]. 计算机工程与设计, 2008, 29(1): 34-36.
- [8] 池越, 夏克文, 陈国鹰, 等. 一种基于 QPSO 和 WLS-SVM 智能方法[J]. 河北工业大学学报, 2009, 38(2): 9-15.

作者简介:

姚 晔 女, (1973—), 硕士, 副教授. 研究方向为计算机应用及网络。

参考文献:

- [1] 王哲, 阮永良. 异常入侵检测技术综述[J]. 电脑知识与技术, 2004(20): 56-59.
- [2] 郁伟, 胡蓓, 陈昌芳. 入侵诱骗系统中的数据捕获[J]. 网络安全技术与应用, 2007(3): 36-38.
- [3] 朱元忠. 入侵检测系统发展研究[J]. 北京工业职业技术学院学报, 2006(1): 8-11.
- [4] 杨光, 巫春玲, 程远征. 基于 RS 和 WSVM 的网络入侵检测算法研究[J]. 计算机仿真, 2011(5): 175-178.
- [5] 彭诗力, 谭汉松. 基于特征值的多模式匹配算法及硬件实现[J]. 计算机工程与应用, 2005(1): 148-150.

作者简介:

谈娟茹 女, (1972—), 硕士, 讲师. 研究方向为计算机硬件、嵌入式系统、软件工程。