

改进支持向量机在网络入侵检测中的应用

于 静, 王 辉

(南京工业大学 电子与信息工程学院, 江苏 南京 210009)

摘 要: 为提高网络安全性, 提出一种改进支持向量机的网络入侵检测算法. 首先采用核主成分分析提取网络数据重要特征, 加快网络入侵检测速度, 然后采用粒子群算法对支持向量机参数进行优化, 提高网络检测正确率. 仿真实验结果表明, 改进支持向量提高网络入侵检测正确率, 降低漏检率, 同时加快了网络入侵检测速度, 是一种有效、实时性较强的网络入侵检测算法.

关键词: 粒子群优化算法; 核主成分分析; 支持向量机; 入侵检测

中图分类号: TP309

文献标识码: A

文章编号: 1000-7180(2012)03-0010-04

Application of Improved Support Vector Machines in Network Intrusion Detection

YU Jing, WANG Hui

(College of Electronics and Information Engineering, Nanjing University of Technology, Nanjing 210009, China)

Abstract: In order to improve the security of the network, this paper proposes an improved support vector machine algorithm for network intrusion detection. Firstly, kernel principal component analysis is used to extract data network feature, accelerate the network intrusion detection rate, then the particle swarm optimization algorithm is used to optimize the support vector machine parameters, improve the network detection accuracy rate. The simulation results show that, the proposed method has improved network intrusion detection rate, reduce the failure rate, and speed the network intrusion detection, it is an efficient, real-time network intrusion detection algorithm.

Key words: particle swarm optimization; PCA; support vector machines; intrusion detection

1 引言

网络入侵检测^[1]主要有误用和异常入侵检测两大类检测方法. 误用检测采用预先设定入侵模式, 然后根据模式匹配对入侵进行检测, 针对性强, 可以建立高效检测系统, 对于已知入侵可达到很高的检测准确率, 但是无法对未知或变异入侵行为进行有效检测, 检测结果没有实际意义. 异常检测技术利用正常用户和系统运行模式建立一种模型, 然后根据模型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, 因此异常入侵检测方法是当前网络入侵检测主要研究方向^[2].

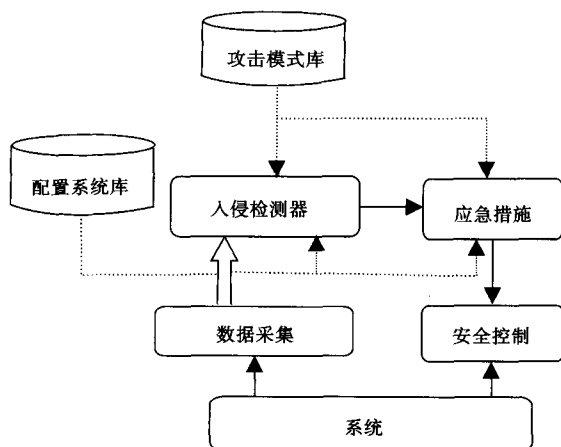
当前异常入侵检测技术主要有传统统计、贝叶斯网络、数据挖掘等方法^[3]. 传统统计异常入侵检测

是最原始、最早网络入侵检测技术, 其只适合于小规模网络入侵检测, 对于现代大规模网络, 其检测速度慢, 效率低^[4]. 贝叶斯网络入侵检测速度快, 误警率高, 自适应能力较差, 不适合当前多变的网络入侵检测^[5]. 神经网络具有较强非线性处理能力, 能够识别大量未知的、新网络入侵检测, 检测正确高, 然而其采用经验风险最小化理论设计, 大样本情况下入侵检测正确率高, 样本数量较少时, 检测效率不佳^[6]. 网络入侵样本实际上是一个典型小样本问题, 神经网络检测结果经常出现拟合现象, 限制其在网络入侵检测中的应用^[6]. 最小二乘支持向量机 (Least Squares Support Vector Machine, LSSVM) 是一种专门是针对小样本分类和预测问题, 非常适合于小样本的网络入侵检测^[6].

为了提高网络入侵检测准确率,保证网络安全,提出一种基于核主成分分析(KPCA-GA-LSVM),并采用粒子群算法对LSSVM参数进行优化,加快网络入侵检测速度。

2 网络入侵检测系统结构

网络入侵检测是一种主动的安全防护措施,它从系统内部和各种网络资源中主动采集信息,分析可能的入侵攻击行为,有效地扩展系统管理员的安全管理能力,提高信息安全基础结构的完整性。网络入侵检测系统结构如图1所示^[7]。



3 KPCA-GA-LSVM 算法

3.1 核主成分分析(KPCA)

核主成分分析(KPCA)是一种非线性数据处理方法,网络入侵特征多,特征间呈典型非线性关系,传统主成分分析(PCA)不能很好的提取网络入侵检测特征。因此本文采用KPCA在网络入侵检测之前,对输入数据特征进行主成分分析,有效提取对网络检测检测重要的特征^[6]。

设网络入侵检测训练样本集为 $X = \{x_1, x_2, \dots, x_m\}$, 其中 m 表示样本数, 计算其协方差矩阵, 即有:

$$C = \frac{1}{m} \sum_{j=1}^m \phi(x_j) \phi(x_j)^T \quad (1)$$

其中, ϕ 是一个非线性映射。

对协方差矩阵进行特征分解, 可以获得:

$$\lambda v = Cv \quad (2)$$

特征向量 v 是由 $\phi(x_1), \phi(x_2), \dots, \phi(x_m)$, 且对于所有特性值, 有 $\lambda \geq 0$, 由此可得:

$$\lambda(\phi(x_k), v') = (\phi(x_k), Cv') \quad (3)$$

由于 v' 是 $\phi(x)$ 的线性组合, 那么通过转换可

以得到特征向量: $M\lambda^r c^r = Kc^r$ 。

假设 $c^p, c^{p+1}, \dots, c^M$ 对应于大于 0 特征向量, 对 c^r 进行归一化, 那么 $\phi(x)$ 在 c^r 上投影为:

$$\phi(x_i) = \phi(x_i) - \frac{1}{m} \sum_{i=1}^m \phi(x_i) \quad (4)$$

那么, $g(x)$ 是 $\phi(x)$ 非线性主元分量, 将全部投影值组成一个矢量 $g(x) = [g_1(x), g_2(x), \dots, g_l(x)]^T$, 作为网络入侵训练样本的新特征。

这样, 通过 KPCA 对网络入侵数据的特征维特征向低维特征转化, 从而达到降维的目的, 提高后继的网络入侵速度。

3.2 支持向量机(SVM)算法

支持向量机(SVM)是一种基于统计学理论的机器学习方法, 其最早应用于线性分类问题, 在所有的训练样本中找出一个最优分类平面, 然而在约束条件下对最优分类面求解, 一般采用以下对偶问题进行求解:

$$Q(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \quad (5)$$

式(5)的约束条件为:

$$\begin{cases} \sum_{i=1}^n y_i \alpha_i = 0 \\ \alpha_i \geq 0, i = 1, 2, \dots, n \end{cases} \quad (6)$$

这样就可以在 α_i 求得式(5)的最大值。如果 α_i 为最优解, 那么所要寻找的最优分类平面的权系数向量为:

$$w^* = \sum_{i=1}^n \alpha_i^* y_i x_i \quad (7)$$

α_i 取值不为零的样本就是 SVM 的支持向量, 在所有样本只是占了很少一部分。这样两分类的最优决策函数为:

$$f(x) = \text{sgn} \left(\sum_{i=1}^n \alpha_i^* y_i (x_i \cdot x) + b^* \right) \quad (8)$$

其中, sgn 表示符号函数, b^* 表示分类域值。

如果对于一个线性不可分的两分类问题, 通过引入内积 $K(x, x^T)$ 代最分类平面中的点积。这样不仅防止了复杂的非线性变换, 同时通过线性函数实现了非线性问题的求解。这样式(5)就变为:

$$Q(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j K(x_i \cdot x_j) \quad (9)$$

相应的决策分类函数为:

$$f(x) = \text{sgn} \left(\sum_{i=1}^n \alpha_i^* y_i K(x_i, x_j) + b^* \right) \quad (10)$$

SVM 是针对两分类问题提出来的, 但是网络入

入侵检测是一个多分类问题,就首先需要将网络入侵检测多分类问题转化成多个二分类问题,然后对这些二分类问题分别求解,最后采用一定的方法将这些二分类问题结果组合得到网络入侵检测结果.本文采用一对一的方法实现网络入侵检测多分类问题.具体如图 2 所示.

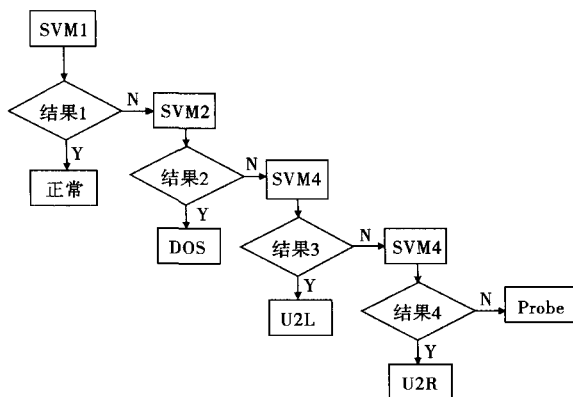


图 2 SVM 的网络入侵检测模型

3.3 粒子群优化算法(PSO)

粒子群优化算法(PSO)是对鸟群飞行行为模拟的群智能技术,粒子代表每一个问题可行解,粒子在解空间进行搜索,通过本身与同伴飞行经验调整自己飞行速度和方向,追踪当前和历史最优粒子进行搜索,最终找到问题最优解^[7].在每一次迭代中,粒子最优值度和位置更新公式为:

$$v_{id}(i+1) = \omega \times v_{id}(i) + c_1 \times rand() \times (P_{best} - x_{id}(i)) + c_2 \times rand() \times (g_{best} - x_{id}(i)) \quad (11)$$

$$x_{id}(i+1) = x_{id}(i) + v_{id}(i+1) \quad (12)$$

其中, c_1 、 c_2 表示加速因子; $v_{id}(i)$ 和粒子群优化算法(PSO)分别表示粒子当前和更新后速度; $x_{id}(i)$ 和 $x_{id}(i+1)$ 分别表示粒子的当前和更新后位置; ω 表示惯性权重; $rand()$ 表示一个(0,1)间随机数.

3.4 PSO 对 SVM 参数优化

选择径向基核函数作为 SVM 核函数,在网络入侵检测建模过程,需要优化参数为惩罚参数 C 、径向基核函数宽度 δ ,采用 PSO 对 C 和 δ 进行寻优. PSO 对 SVM 参数优化具体优化流程如图 3 所示.

4 KPCA-PSO-LSVM 的网络入侵检测过程

(1) 采集网络中网络数据,并对其进行预处理,消除不用数据.

(2) 采用 KPCA 对网络原始数据进行主成分分析,降低输入维数,提取有利用网络健的重要特

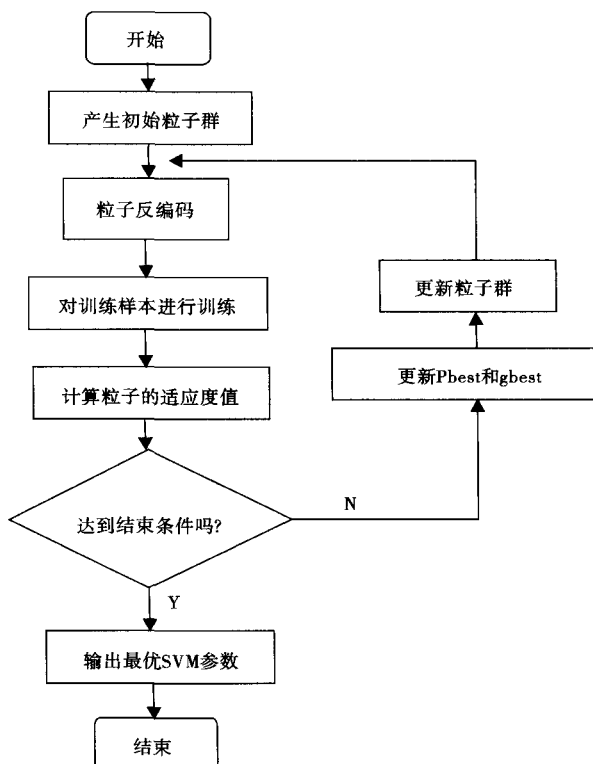


图 3 PSO 对 SVM 参数优化的工作流程

征.

(3) 将经过 KPCA 降维处理的网络数据分成两部分:训练集和测试集,训练集用于参数寻优和建模,测试集用于检验模型性能.

(4) 将训练样本输入到 SVM 进行学习,并通过 PSO 对参数 C 和 δ 进行寻优,采 10 折交叉验证判断参数优劣,最后采用最优参数建立网络入侵的最优检测模型;

(5) 利用建立最优网络入侵检测模型对网络测试集检测,检测建立模型的有效性.

(6) 对获得网络入侵检测结果进行分析和判断,评价模型可行性和优越性.

5 仿真实验

5.1 数据来源

数据来自 KDD CUP 1999 数据集,共包括 4 类入侵:扫描攻击(Probe)、拒绝服务攻击(DOS)、远程用户未授权访问攻击(U2L)、未授权使用本地超级权限访问攻击(U2R).每个网络连接由 41 个特征属性和 1 个标记组成,仿真实验选择数据集见表 1.

5.2 网络数据 KPCA 处理

首先采用预处理方法消除数据中噪声信息,然后采用 KPCA 对其进行降维处理,提取其主要检测

特征,得到的特征值如表2所示.从表2可知,可以选择前7个主要特征就可以代表原始网络入侵检测数据,从而大大提高网络入侵检测效率,加快检测速度.

表1 仿真实验采用各种入侵样本数目

入侵类型	训练样本数	测试样本数
DOS	1000	200
U2L	800	150
U2R	100	20
PROBE	200	40

表2 KPCA 的处理结果

No	特征值	百分率%	累计百分率%
1	4.9717	41.4311	41.4311
2	1.5534	12.9448	54.3759
3	1.2477	10.3975	64.7735
4	1.1261	9.3838	74.1572
5	1.1020	9.1837	83.3409
6	0.8272	6.8935	90.2344
7	0.6156	5.1304	95.3648
...	...	...	...

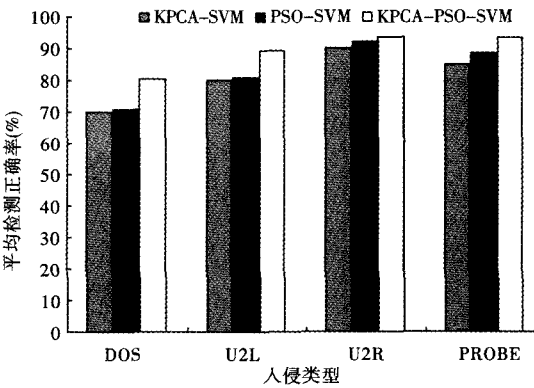
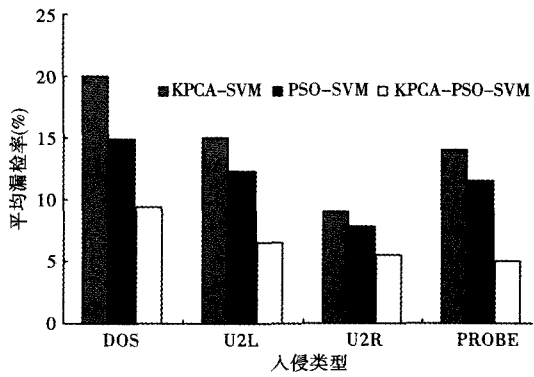
5.3 实验结果与分析

为了检测 KPCA-PSO-SVM 性能,采用 KPCA-SVM、PSO-SVM 作为对比检测算法,采用检测正确和漏检率作为算法评价指标,其检测结果如图4和5所示.从图4的网络入侵检测正确对比可知,基于 KPCA-PSO-SVM 的网络入侵检测系统的检测正确率高于 KPCA-SVM、PSO-SVM.从图5网络入侵检测漏检率对比可知,所有算法中,KPCA-PSO-SVM 检测漏检率最低,主要是由于 KPCA-PSO-SVM 集成了 KPCA、PSO、SVM 的优点,加快了网络入侵检测速度,提高了网络入侵检测正确率高,是一种有效的网络入侵检测工具.

6 结束语

针对传统网络检测算法的缺陷,利用核主成分分析、粒子群优化算法、支持向量机等算法的优点,组成一种新的网络入侵检测算法(KPCA-PSO-SVM).仿真结果表明,KPCA-PSO-SVM 加快了网络入侵的检测速度,提高了网络入侵的检测正确率,检测结果更加可靠,是一种有效的网络安全防范工具.

参考文献:

- 
- | 入侵类型 | KPCA-SVM | PSO-SVM | KPCA-PSO-SVM |
|-------|----------|---------|--------------|
| DOS | ~70 | ~70 | ~80 |
| U2L | ~80 | ~80 | ~90 |
| U2R | ~90 | ~90 | ~95 |
| PROBE | ~85 | ~85 | ~90 |
- 图4 3种算法的网络入侵检测正确率
- 
- | 入侵类型 | KPCA-SVM | PSO-SVM | KPCA-PSO-SVM |
|-------|----------|---------|--------------|
| DOS | ~20 | ~15 | ~10 |
| U2L | ~15 | ~12 | ~6 |
| U2R | ~9 | ~8 | ~5 |
| PROBE | ~14 | ~12 | ~5 |
- 图5 3种算法的网络入侵检测漏检率
- [1] 曾勃炜,徐知海,吴键. 基于粒子群优化和支持向量机的电力负荷预测[J]. 微电子学与计算机, 2011, 28(1):147-149.
 - [2] 戴云,范平志. 入侵检测系统研究综述[J]. 计算机工程与应用, 2002(4):17-19.
 - [3] 郭亚周,高德远. 模糊聚类分析在入侵检测系统中的应用研究[J]. 沈阳理工大学学报, 2005, 24(4):26-28.
 - [4] 饶鲜,董春曦,杨绍全. 基于支持向量机的入侵检测系统[J]. 软件学报, 2003, 14(4):798-803.
 - [5] 曾凡培. 粗集神经网络在网络入侵中的应用研究[J]. 计算机仿真, 2011, 28(7):161-164.
 - [6] 徐涛,孙彤. 基于 KPCA 的非线性 ASM 定位方法研究[J]. 微电子学与计算机, 2010, 27(12):113-117.
 - [7] 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.

作者简介:

于 静 女, (1987—), 硕士. 研究方向为信号与信息处理.
王 辉 女, (1962—), 博士. 研究方向为光通信.