

基于粒子群优化相关向量机的网络入侵检测

吴良海

(茂名学院 实验教学部, 广东 茂名 525000)

摘 要: 构建计算机网络的入侵检测系统, 对于保护网络中的信息免受各种攻击显得非常重要. 为了克服支持向量机的缺点, 提出了一种基于粒子群优化相关向量机(RVM)网络入侵检测方法. 相关向量机是一种建立在支持向量机上的稀疏概率模型. 与支持向量机相比, 它不仅具有较高检测精度, 还具有较好的实时性. 粒子群优化算法用于确定相关向量机的核参数. 最后结合试验将提出的方法同支持向量机算法、BP 神经网络进行了比较, 结果表明提出的相关向量机相比于支持向量机、BP 神经网络有着更高的入侵精度.

关键词: 相关向量机; 网络异常; 入侵检测; 粒子群

中图分类号: TP391

文献标识码: A

文章编号: 1000-7180(2010)05-0181-04

Network Intrusion Detection Based on Relevance Vector Machine Optimized by Particle Swarm Optimization Algorithm

WU Liang-hai

(Department of Experimental Teaching, Maoming University, Maoming 525000, China)

Abstract: It is significant to protect the information of network and avoid it under attack by constructing network intrusion detection system. In order to overcome the drawbacks of support vector machine, network intrusion detection based on relevance vector machine optimized by particle swarm optimization algorithm (PSO-RVM) is presented in the paper. Relevance vector machine is a sparse probability model based on support vector machine. Relevance vector machine has not only higher detection accuracy, but also better real-time than support vector machine. In the study, particle swarm optimization algorithm is used to determine nuclear parameter of relevance vector machine. Finally, the case data are used to testify and analyze the performance of the proposed model. The experimental results show that PSO-RVM has greater prediction accuracy than PSO-SVM, BP neural network.

Key words: relevance vector machine; network anomaly; intrusion detection; particle swarm

1 引言

随着信息科技和网络技术的发展和普及, 计算机网络中的信息安全问题越来越突出. 由于计算机网络自身存在的局限性, 使得网络和计算机系统的硬件、软件, 以及信息数据等遭到破坏、更改、泄露, 严重时会引起系统的崩溃瘫痪, 造成巨大的经济损失^[1-2]. 因而, 构建计算机网络的入侵检测系统, 以保护网络中的信息免受各种攻击显得非常重要. 入侵检测系统通过监测网络数据流以检测出系统中未经授权的使用或滥用计算机资源的行为, 实现对内、

外部入侵攻击进行实时监测. 目前, 入侵检测理论和技术在不断发展, 神经网络是一种常用的网络入侵检测方法, 然而由于其存在收敛速度慢且容易陷入局部极小等问题限制了它的应用. 支持向量机(Support Vector Machine, SVM)是一种基于统计学习理论的机器学习方法, 它能有效解决非线性、小样本问题, 在入侵检测系统中的应用越来越广泛^[3-5].

尽管 SVM 在实际中取得广泛应用, 但也存在不足之处, 如 SVM 的惩罚参数 C 必须人为确定且难以选取优化的惩罚参数, 此外 SVM 的核函数必须符合 Mercer 条件^[6]. 相关向量机(relevance vector

machine, RVM)^[7-10]是一种建立在支持向量机上的稀疏概率学习模型,与支持向量机相比,它需要更少的参数和核函数约束。

为此,文中提出了一种基于粒子群优化相关向量机(RVM)网络入侵检测方法.相关向量机的训练是在贝叶斯框架下进行的,它能产生的具有较高检测精度的决策函数,并提高了系统的实时性,粒子群优化算法^[11-12]用于确定相关向量机的核参数.文中建立了基于 RVM 的网络入侵检测,并采用 MIT Lincoln 实验室的 KDDCup99 数据库评估所提出的入侵检测技术的性能。

2 相关向量机

2.1 支持向量机理论

支持向量机是一种统计学习算法,它基于结构风险最小化准则,具有较好的泛化能力.它能利用非线性映射函数将原始输入数据映射到高维特征空间,然后利用线性分类器在该空间进行分类.在高维特征空间中构造最优分类超平面:

$$f(x) = w \cdot \varphi(x) + b = 0 \quad (1)$$

定义核函数为

$$k(x, x_i) = \varphi(x) \varphi(x_i) \quad (2)$$

SVM 分类器函数为

$$f(x) = \text{sign}(\sum_{i,j=1}^n \alpha_{ij} y_j k(x_i, x) + b) \quad (3)$$

2.2 相关向量机理论

相关向量机是由 Tipping 提出的一种建立在支持向量机上的稀疏概率学习模型.它的训练是基于贝叶斯估计理论,能在概率意义下的进行合理划分,使得分类函数针对于训练集似然函数值最大.相关向量机克服了 SVM 的惩罚参数 C 的主观设置,以及 SVM 的核函数必须符合 Mercer 条件等问题。

给定训练样本集 $\{x_i, y_i\}_{i=1}^n, x_i \in R^n$ 为输入矢量, $y_i \in \{-1, 1\}$ 表示 2 个待分类的模式. RVM 的模型输出定义为

$$f(x, w) = \sum_{i=1}^n w_i k(x, x_i) + w_0 \quad (4)$$

式中, w_i 为模型权值, $k(x, x_i)$ 是核函数.在文中主要采用 RBF 高斯核函数.

$k(x, x_i) = \exp\left[-\frac{\|x - x_i\|^2}{2\sigma^2}\right]$, σ 表示核的宽度。

假设噪声 ϵ_i 服从均值为零,方差为 σ^2 的高斯分布,则

$$d_i = f(x_i, w) + \epsilon_i \quad (5)$$

设 d_i 服从独立分布,则整个样本的似然函数为

$$p(d_i | a, \sigma^2) = (\pi\sigma^2)^{\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2} \|d - \theta\|^2\right] \quad (6)$$

$$\theta = \varphi w, d = (d_1, d_2, \dots, d_n)^T, w = (w_1, w_2, \dots, w_n)^T, \Phi = [\varphi(x_1), \varphi(x_2), \dots, \varphi(x_n)]^T, \varphi(x_i) = [1, k(x_i, x_1), k(x_i, x_2), \dots, k(x_i, x_n)]^T$$

为了防止模型中存在过多的相关向量,进而提高模型的泛化能力, RVM 为每一个权值定义了高斯先验概率分布:

$$p(w | a) = \prod_{i=1}^n n(w_i | 0, a_i^{-1}) \quad (7)$$

式中, a 是决定权值先验分布的超参数,它决定了模型的稀疏特性。

对于非线性分类器,存在一个从向量 x 到高维空间 H 的非线性映射函数 $\varphi(x)$ 使得 $k(x, x_i) = \varphi(x) \varphi(x_i)$. RVM 分类器可以写成:

$$f(x) = \Phi^T(x) \left[\sum_{i=1}^n a_i \Phi(x_i) \right] \quad (8)$$

3 RVM 分类器参数确定

在利用 RVM 分类器进行分类以前,需要确定模型的参数,由于选择 RBF 高斯函数为核函数,因此 RBF 高斯核的宽度为 RVM 需要确定的参数.这里采用遗传算法与 5 折交叉验证准则进行 RVM 参数确定.基于 PSO 的参数优化过程见图 1。

步骤 1: 粒子初始化. 由参数 σ 组成一个粒子,并随机产生一组粒子的初始位置和速度. 确定迭代次数、惯性权重、加速系数等参数。

步骤 2: 适应度评估. 采用 5 折交叉验证准则进行适应度评估. 5 折交叉验证中随机地将训练样本分成 5 等份. 对于每一个参数设置, 训练模型 5 次, 在每次训练中, 以 4 个子集进行训练, 剩余的一个子集进行测试, 记录每一次的分类结果. 最后将 5 次的分类结果进行平均作为泛化性能的评估。

步骤 3: 将每个粒子的适应度值与其经历过的最好位置 pbest 作比较, 如果较好, 则将其作为当前的最好位置 pbest。

步骤 4: 将每个粒子的适应度值与全局所经历的最好位置 gbest 作比较, 如果较好, 则将 gbest 设置为最好粒子的当前位置。

步骤 5: 根据式(6)和式(7)更新粒子的速度和位置。

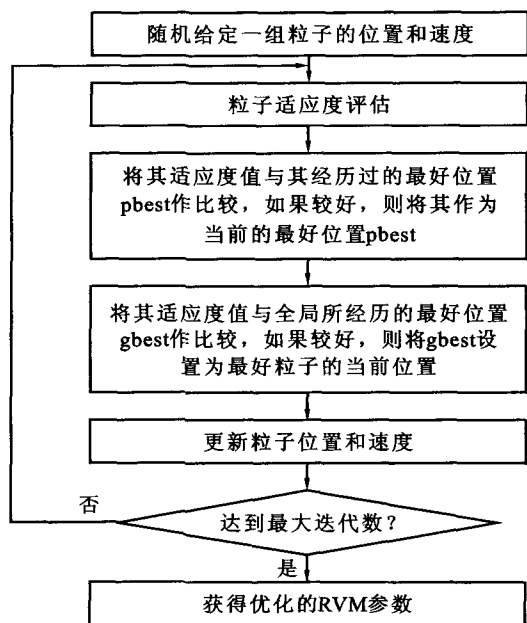


图 1 基于 PSO 的参数优化过程

$$v_{iq}(t + 1) = w \cdot v_{iq}(t) + c_1 \cdot \text{rand} \cdot (\text{pbest}_{iq}(t) - p_{iq}(t)) + c_2 \cdot \text{rand} \cdot (\text{pbest}(t) - p_{ig}(t)) \quad (9)$$

$$p_{iq}(t+1) = p_{iq}(t) + \beta \cdot v_{iq}(t+1) \quad (10)$$

式中, t 为进化代数; rand 为 0 到 1 之间的随机数; w 为惯性权值, 用于平衡全局搜索和局部搜索; β 为约束因子, 用于控制速度的权重; c_1, c_2 是学习因子.

步骤6:检验迭代次数是否达到了给定的最大次数,如果符合,则停止迭代;否则转至步骤2.

4 实验分析

文中采用 MIT Lincoln 实验室的 KDDCup99 数据库评估提出的入侵检测技术的性能。该数据集包含了四种攻击类型:拒绝服务(Denial of Service, DOS)攻击,来自远程主机的未授权访问(User to Root, U2R)攻击,未授权的本地超级用户特权访问(Remote to Location, R2L)攻击,端口扫描(Probing)攻击,其中 DOS 攻击占多数,这与实际网络运行时的情形相符。

[illegible]

如表 1 所示,从 KDDCup99 数据库随机选取 3 120 条数据(其中包括 2 000 条正常数据、800 条 DOS 入侵数据、30 条 U2R 入侵数据、80 条 R2L 入侵数据和 210 条 Probing 入侵数据)组成训练数据集,另外选取 2330 条数据(其中包括 1600 条正常数据、500 条 DOS 入侵数据、20 条 U2R 入侵数据、60 条 R2L 入侵数据和 150 条 Probing 入侵数据)组成测试数据集。文中采用 4 个 PSO-RVM 识别网络的这 5 种状态,包括正常、DOS 入侵、U2R 入侵、R2L 入侵以及 Probing 入侵,这 4 个 PSO-RVM 以二叉树形式布置,以提交计算效率。表 2 为 PSO-RVM、PSO-SVM 和 BPNN 三种模型的人侵检测结果比较,RVM 检测精度为 98.240,SVM 检测精度为 96.481,BPNN 检测精度为 88.026,可以看出 PSO-RVM 检测精度优于 PSO-SVM,BPNN,具体非常好的应用前景。

表 1 构建的实验数据集

类型	训练数据集	测试数据集
正常	2 000	1 600
DOS	800	500
U2R	30	20
R2L	80	60
Probing	210	150
总计	3 120	2 330

表 2 PSO-RVM, PSO-SVM, BPNN 检测精度比较

分类算法	平均分类 精度/%	正确检出 个数	检测精度 /%
PSO-RVM	DOS	493	98.240
	U2R	18	
	R2L	56	
	Probing	143	
PSO-SVM	DOS	482	96.481
	U2R	13	
	R2L	51	
	Probing	138	
BPNN	DOS	412	88.026
	U2R	9	
	R2L	42	
	Probing	103	

5 结束语

文中提出了一种基于粒子群优化相关向量机(RVM)网络入侵检测方法,一种建立在支持向量机

上的稀疏概率学习模型,它的训练是在贝叶斯框架下进行的,它能产生的具有较高检测精度的决策函数,粒子群优化算法用于确定相关向量机的惩罚参数.文中建立了基于粒子群优化 RVM 的网络入侵检测模型,并采用 MIT Lincoln 实验室的 KDD-Cup99 数据库评估提出的入侵检测技术的性能.从实验结果可以看出 PSO-RVM 检测精度优于 PSO-SVM 和 BPNN,具体非常好的应用前景.

参考文献:

- [1] 章金熔,刘峰,赵志宏,等.数据挖掘方法在网络入侵检测中的应用[J].计算机工程与设计,2009,30(24):5561-5566.
- [2] 张琨,曹宏鑫,刘凤玉,等.一种基于 SVM 的网络入侵检测模型[J].南京理工大学学报,2007,31(4):403-408.
- [3] 王强,陈欢欢,王珽.一种基于多类支持向量机的故障诊断算法[J].电机与控制学报,2009,13(2):302-306.
- [4] 赫英明,王汉杰,姜祝辉.支持向量机在云检测中的应用[J].解放军理工大学学报:自然科学版,2009,10(2):191-194.
- [5] 许少华,何新贵,周继,等.一种过程支持向量机及其在动态模式分类中的应用[J].控制与决策,2009,24(2):309-311.
- [6] Sankar Mahadevan, Sirish L Shah. Fault detection and diagnosis in process data using one-class support vector machines[J]. Journal of Process Control, 2009, 19(10): 1627-1639.
- [7] 张磊,刘建伟,徐翔,等.基于相关向量机的神经活动分类及译码[J].计算机工程,2009,35(20):197-199.
- [8] Subimal Ghosh, Mujumdar P P. Statistical downscaling of GCM simulations to streamflow using relevance vector machine[J]. Advances in Water Resources, 2008, 31(1): 132-146.
- [9] Clodoaldo A M Lima, André L V Coelho, Sandro Chagas. Automatic EEG signal classification for epilepsy diagnosis with relevance vector machines[J]. Expert Systems with Applications, 2009, 36(6): 10054-10059.
- [10] John Flake, Todd K Moon, Mac McKee, et al. Application of the relevance vector machine to canal flow prediction in the sevier river basin[J]. Agricultural Water Management, 2010, 97(2): 208-214.
- [11] 陈白帆,蔡自兴,袁成.基于粒子群优化的移动机器人 SLAM 方法[J].机器人,2009,31(6):513-517.
- [12] Suresh S, Sujit P B, Rao A K. Particle swarm optimization approach for multi-objective composite box-beam design[J]. Composite Structures, 2007, 81(4): 598-605.

作者简介:

吴良海 男,(1978-),硕士,讲师.研究方向为计算机应用技术.

(上接第 180 页)

- [7] 尹学松,胡恩良,陈松灿.基于成对约束的判别型半监督聚类分析[J].软件学报,2008,19(11):2791-2802.
- [8] 宋长新.基于 TLS 的正交小波变换红外图像去噪[J].激光与红外,2009,39(1):74-77.
- [9] Yan B, Domeniconi C. Kernel optimization using pairwise constraints for semi-supervised clustering[R]. Technical Report ISE-TR-06-09, 2006.
- [10] Yan B, Domeniconi C. An adaptive kernel method for semi-supervised clustering[C]// Proceedings of the 17th European Conference on Machine Learning. Germany: Berlin, 2006(4212):521-532.
- [11] Wu K-P, Wang S-D. Choosing the kernel parameters for support vector machines by the inter-cluster distance in the feature space[J]. Pattern Recognition, 2009, 42(5):710-717.
- [12] 宋长新.改进的空间约束加权模糊核聚类红外图像分割[J].微电子学与计算机,2009,26(5):60-63.

作者简介:

吴术路 男,(1964-),副教授.研究方向为图像处理等.
张俊峰 男,(1967-),副教授.研究方向为图像处理及模式识别等.
宋长新 女,(1971-),博士,副教授.研究方向为数据挖掘和图像处理等.