

基于 QPSO 的属性约简在 NIDS 中的应用研究

汪世义^{1,2}, 陶 亮¹, 王华彬¹

(1 安徽大学 智能计算与信号处理教育部重点实验室, 安徽 合肥 230039; 2 巢湖学院 计算机系, 安徽 巢湖 238000)

摘 要: 支持向量机作为一种优良的分类算法应用在网络入侵检测系统中, 但是训练时间过长是它的主要缺陷. 文中提出了基于量子粒子群优化的属性约简和支持向量机(SVM)的入侵检测方法, 利用量子粒子群优化的属性约简算法对训练样本集进行属性约简, 剔除了对入侵检测结果影响较小的冗余特征, 从而使入侵检测系统在获取用户特征的时间减少, 整个入侵检测系统的性能得到提高. 实验结果表明, 该方法是有用的.

关键词: 网络入侵检测系统; 支持向量机; 量子粒子群优化; 属性约简

中图分类号: TP393

文献标识码: A

文章编号: 1000-7180(2010)01-0120-03

Application of Attribute Reduction Based on QPSO in NIDS

WANG Shi-yi^{1,2}, TAO Liang¹, WANG Hua-bin¹

(1 Ministry of Education Key Laboratory of Intelligence Computing and Signal Processing, Anhui University, Hefei 230039, China; 2 Department of Computer, Chaohu College, Chaohu 238000, China)

Abstract: The SVM is one of the most successful classification algorithms in network intrusion detection (NIDS) area, but its long training time limits its use. This paper presents a method for enhancing the training time of SVM using attribute reduction optimized by quantum-behaved particle swarm optimization (QPSO), specifically when dealing with large training data sets in NIDS. The reduction algorithm based on attribute reduction optimized by QPSO is used to eliminate the redundant features of sample data set, with the attributes of the raw data are reduced, the SVM training time are reduced. The NIDS based on attribute reduction optimized by QPSO and SVM has better performance. Experimental results show that this method is efficient.

Key words: NIDS; SVM; QPSO; attribute reduction

1 引言

随着网络技术的快速发展, 网络安全问题变得越来越重要. 入侵检测是一种安全防护措施, 它从计算机网络或计算机系统内的若干关键点收集信息, 并通过对其进行分析从而发现网络或系统中是否有违反安全策略的行为和遭受攻击的迹象, 同时做出相应的响应.

支持向量机(Support Vector Machine, SVM)^[1]是 Vapnik 等人根据统计学习理论提出的一种新的机器学习方法, 是对结构风险最小化原则的近似, 支持向量机在解决小样本非线性及高维模式识别问题

中表现出许多特有的优势, 已经在人脸识别、特征选择和蛋白质结构预测等方面取得了较好的效果^[2-3]. 与传统的人工神经网络相比, SVM 不仅结构简单, 而且各种技术性能尤其是泛化能力明显提高, 这已经被大量实验证实.

入侵检测问题本质上属于有限样本、非线性模式识别问题, 因此, 把分类性能卓越的 SVM 作为入侵特征分类器是一种合理的自然选择. 支持向量机作为分类器实质上是将模式分类问题归结为一个二次规划问题, 而这类问题的求解算法多趋于复杂. 由于入侵检测问题规模较大, 通常需要耗费相当长的计算时间和存储空间, 很难达到入侵检测系统的实

收稿日期: 2008-11-11; 修回日期: 2009-02-20

基金项目: 国家自然科学基金项目(60572128); 安徽省高校自然科学基金研究计划项目(KJ2008B38ZC, KJ2007B239)

时性要求.因此研究快速学习算法是支持向量机在入侵检测中急待解决的问题之一.

粗糙集理论是一种新的处理模糊和不确定性知识的数学工具,它的主要思想是在保持分类能力不变的前提下通过知识约简导出问题的决策和分类规则.量子粒子群优化(QPSO)算法的知识约简算法改进了粒子进化策略,使粒子具有更大搜索空间,可更好地避免陷入局部最优,利用 QPSO 优化粗糙集约简算法对样本集进行特征约简,除去对入侵检测结果影响较小的冗余特征,得到了最小约简结果,从而有效地降低了样本集的维数,减小了 SVM 训练时间,使入侵检测的实时性要求得到提高.

2 支持向量机理论

支持向量机是一种建立在统计学习理论基础上的机器学习方法.其主要的特点是根据 Vapnik 结构风险最小化原则,提高学习机的泛化能力,即由有限的训练集样本得到小的误差,仍然能够保证对独立的测试集保持小的误差.另外由于支持向量机算法是一个凸优化问题,所以局部最优解一定是全局最优解.

设两类线性学习(训练)样本集为 $\{x_i, y_i\}_{i=1}^N$, x_i 是第 i 个学习样本的输入模式,且是 d 维列向量,即 $x_i \in R^d$, $x_i = [x_i^1 \ x_i^2 \ \cdots \ x_i^d]^T$, 另外, $y_i \in \{+1, -1\}$ 是其类别标记(或称为目标输出).两类线性可分情况下的判别函数为

$$g(x) = \langle w, x \rangle + b = w^T x + b$$

式中, x 是一样本向量, w 是权向量, b 为分类阈值, $\langle w, x \rangle$ 表示两向量的内积.如果存在一决策面

$$g(x) = w^T x + b = 0 \quad (1)$$

分开两类样本,使得

$$y_i = 1, \text{ 有 } g(x) = w^T x_i + b > 0 \quad (2)$$

$$y_i = -1, \text{ 有 } g(x) = w^T x_i + b < 0 \quad (3)$$

则称 $g(x) = w^T x + b = 0$ 为分类面或超平面.而最优分类面不但能将两类正确分开,并且要求使分类间隔最大.通过解约束最优化问题得到最优分类函数为

$$\begin{aligned} f(x) &= \text{sign}\{g(x)\} \\ &= \text{sign}(w_{\text{op}}^T x + b_{\text{op}}) \\ &= \text{sign}\left(\sum_{i=1}^s \alpha_i^{\text{sv}} y_i^{\text{sv}} \langle x, x_i^{\text{sv}} \rangle + b_{\text{op}}\right) \end{aligned}$$

3 基于 QPSO 属性约简的入侵检测方法

3.1 粗糙集理论

粗糙集理论是一种新的处理模糊和不确定性知

识的数学工具.它的主要思想是:在保持分类能力不变的前提下,通过知识约简,导出问题的决策和分类规则.

定义 1 令 R 为一等价关系族,并且 $r \in R$ 如果 $\text{ind}(R) = \text{ind}(R - \{r\})$. 则称 r 为 R 中必要的; 否则 r 为 R 中不必要的; 当对 $\forall r \in R$, 如果 r 是必要的, 则簇 R 为独立的.

定义 2 对于属性子集 $P \subseteq R$, 若存在 $Q = P - R$, $Q \subseteq P$, 使得 $\text{Ind}(Q) = \text{Ind}(P)$, 且 Q 为最小子集, 则 Q 称为 P 的约简, 表示为 $\text{Red}(P)$.

定义 3 P 中所有必要关系组成的集合称为 P 的核, 记做 $\text{Core}(P)$. P 的所有约简的交集构成簇 P 的核, 即 $\text{Core}(P) = \bigcap \text{Red}(P)$, 其中 $\text{Red}(P)$ 表示 P 的所有约简.

3.2 基于 QPSO 属性约简算法

知识的简化是粗糙集理论的核心内容之一,也是粗糙集理论在入侵检测技术上的主要应用^[4]. 粒子群优化(Partial Swarm Optimization, PSO)算法属于进化算法的一种,和遗传算法相似,它也是从随机解出发,通过迭代寻找最优解,同样是通过适应度来评价解的优劣.粒子群算法通过追随当前搜索到的最优值来寻找全局最优解,没有遗传算法的“交叉”和“变异”操作,相对遗传算法规则来说更为简单.为提高粒子群算法的搜索能力和优化效率并避免早熟收敛,将量子进化算法融合到粒子群算法中形成量子粒子群优化算法(QPSO).基于量子粒子群优化的属性约简算法^[5]主要包括信息系统的粒子模型的建立、定义适应度函数、种群位置初始化.流程如下:

(1) 初始化参数和种群位置.

(2) 计算每个粒子的适应度值,把每个粒子的适应度值与历史最优位置进行比较,决定是否用新适应度值代替.

(3) 把每个粒子的适应度值与全局最优值进行比较,决定是否更新全局最优值,并记下最优值粒子位置.

(4) 更新粒子位置.

(5) 进行粒子编码转换.

(6) 程序终止条件:迭代次数达到最大值或粒子的适应度值达到预定最优值,则停止,输出全局最优粒子的位置,即所求的最优属性约简,否则转(2).

文中将量子粒子群优化的属性约简算法引入到基于 SVM 的网络入侵检测系统中,提出了一种新的入侵检测模型,取得了比较好的效果.

3.3 基于 QPSO 属性约简入侵检测模型

文中提出的基于 QPSO 属性约简入侵检测模型,如图 1 所示.先对从网络收集的数据进行数据预处理,其中主要包括数据离散化,对决策表进行基于 QPSO 的属性约简,产生约简输出规则,创建安全系统的规则库,产生入侵检测器,选择训练样本,建立初始的入侵检测模型,并且该模型在以后系统运行中得到不断完善和改进,最终达到最好的检测效果.

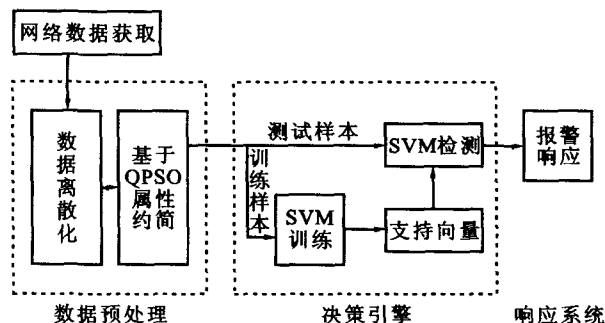


图 1 基于 QPSO 属性约简入侵检测模型

4 实验结果与讨论

4.1 实验数据与环境

系统测试采用的是 KDD99 数据集^[6].每个连接由 41 个标记连接状态的属性和一个标记是否正常的标志性属性组成.为了证明本文提出的入侵检测模型的有效性,实验中采集了各种不同的检测数据,描述了每个网络连接的 41 个属性特征.训练用的每条网络连接都进行了以下标记:normal(正常连接)、DoS(拒绝服务攻击)、probe(探测攻击)、U2R(User to Root 攻击)、R2L(Remote to Local 攻击).对每类标记的数据经属性约简后得到不同数目的属性用于训练,将入侵数据经过数据预处理和属性权重离散化,得到训练数据 17964 条,测试数据 13680 条.训练集中正常连接 7458 条,DoS 攻击 10293 条,probe 攻击 156 条,U2R 攻击 6 条,R2L 攻击 51 条;测试集中正常连接 2692 条,DoS 攻击 10200 条,probe 攻击 314 条,U2R 攻击 13 条,R2L 攻击 461 条.

4.2 测试结果与分析

实验环境操作系统为 Windows XP, Pentium CPU 2.0 GHz, 内存:1.0 GB.程序执行环境采用 Visual C++ 6.0 和执行 QPSO 属性约简算法.实验结果如表 1 所示.

表 1 未经属性约简和经 QPSO 优化属性约简检测数据

	未经属性约简			经 QPSO 优化属性约简		
	学习时间 /s	检测时间 /s	检测率	学习时间 /s	检测时间 /s	检测率
DoS	0.281	0.107	0.981	0.071	0.074	0.994
probe	0.132	0.356	0.931	0.092	0.176	0.985
U2R	0.344	0.450	0.612	0.083	0.312	0.830
R2L	0.276	0.182	0.654	0.032	0.158	0.887

从表 1 可以看出,经过 QPSO 属性约简,支持向量机学习时间减少,在对四种攻击的检测中,入侵检测系统在经过 QPSO 优化属性约简的数据上获得的检测率比未经属性约简的数据的检测率明显提高.

5 结束语

文中将 QPSO 优化属性约简算法和支持向量机结合应用到网络入侵检测,提高了系统的效率.实验中要不断通过调整惩罚因子和核参数使系统获得了较好的性能,研究更好的求解约简算法和在入侵检测系统中对支持向量机的参数和待处理数据的特征同时进行选择是后继的工作.

参考文献:

- [1] Cortes C, Vapnik V. Support vector networks[J]. Machine Learning, 1995, 20(3):273-297.
- [2] 张威, 魏冬生. 基于 Adaboost 与支持向量机的人脸特征提取[J]. 微电子学与计算机, 2007, 24(5):69-72.
- [3] Huang C, Lee Y, Lin D, et al. Model selection for support vector machines via uniform design[J]. Computational Statistics & Data Analysis, 2007, 52(1):335-346.
- [4] 谷雨, 夏虹, 张天军, 等. 基于决策规则格的入侵检测[J]. 微电子学与计算机, 2005, 22(3):53-56.
- [5] 吕士颖, 郑晓鸣, 王晓东. 基于量子粒子群优化的属性约简[J]. 计算机工程, 2008, 34(18):65-66.
- [6] KDD99[EB/OL]. [2009-01-10]. <http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html>.

作者简介:

汪世义 男, (1974-), 博士研究生, 讲师. 研究方向为模式识别与信号处理、网络安全.

陶亮 男, (1963-), 博士, 教授, 博士生导师. 研究方向为数字信号与图像处理、模式识别.

王华彬 男, (1983-), 博士研究生. 研究方向为数字信号与图像处理、模式识别.