第38卷 第3期 华 中 科 技 大 学 学 报(自然科学版) Vol.38 No.3
2010年 3月 J. Huazhong Univ. of Sci. & Tech. (Natural Science Edition) Mar. 2010

基于支持向量机的信息安全风险评估

党德鹏 孟 真

(北京师范大学 信息科学与技术学院, 北京 100875)

摘要: 针对信息安全风险评估中训练数据数目少、方法主观性大、求解最优值困难等问题,提出了基于支持向量机(SVM)的信息安全风险评估方法. 与传统学习方法相比,SVM 分类器对小样本测试环境的适应能力强,具有较好的分类准确率,能有效防止过学习. 通过分析影响信息系统安全的主要因素,结合支持向量机思想,设计并实现了基于支持向量机的信息安全风险评估模型,通过多类核函数构造出不同的分类面以及分类函数,然后对样本数据进行测试,最终得到问题的最优分类解.

关 键 词: 支持向量机; 网络安全; 信息安全; 风险评估; 特征空间; 核函数

中图分类号: TP309 **文献标识码**: A **文章编号**: 1671-4512(2010)03-0046-04

Assessment of information security risk by support vector machine

Dang Depeng Meng Zhen

(College of Information Science and Technology, Beijing Normal University, Beijing 100875, China)

Abstract: In order to solve the difficult issues in information security risk assessment such as having less number of training samples, methods solving the problems are subjective difficulties in solving the optimal value and so on, an approach based on support vector machine (SVM) is proposed. Compared with traditional learning methods, it is more adaptable for small sample test environment, has good classification accuracy, can effectively prevent over study, so SVM classifier has good application prospects in the field of information security risk assessment. By analyzing the main factors that impact the security of information system, and combining with the thinking of SVM, a model of information security risk assessment that based on SVM is designed and realized. In the model, different classification of facial function is constructed by some types of nuclear function and then sample data is tested, ultimately receive the optimal classification result.

Key words: support vector machine; network security; information security; risk assessment; feature space; nuclear function

随着电子政务和电子商务的蓬勃发展,信息安全风险评估已经得到政府、军队、企业和科研机构的高度重视[1~3],现有信息安全风险评估主要依赖专家经验,样本数据数目少,方法主观性大. 为此,本文提出了基于支持向量机(SVM)的信息安全风险评估模型. 通过多类核函数构造出不同的分类面以及分类函数,然后对样本数据进行测试,最终得到该问题的最优分类解. 基于支持向量机的信息安全风险评估模型对小样本测试环境的适应能力强,具有较好的分类准确率,能有效防止过学习,具有很好的应用前景.

1 评估要素及 SVM 的适应性

信息安全指信息系统的部件、程序和数据不因偶然的或恶意的原因而遭到破坏、更改或泄密,

收稿日期: 2009-06-25.

作者简介: 党德鹏(1970-),男,副教授, E-mail: ddepeng@bnu.edu.cn.

基金项目: "十一五"国家科技支撑计划重大项目(2006BAK01A07); "十一五"国家科技支撑计划重点项目(2006BAC18B06); 国家自然科学基金资助项目(60940032).

系统连续可靠地正常运行,服务不中断. 问题的本质在于信息系统的资源存在价值和脆弱性,并容易由此引发风险,其中系统的脆弱性是引发安全问题的内在原因,系统面临的危险则是外在原因[1]. 信息安全风险评估就是评估威胁的存在以及由于系统存在易于受到攻击的脆弱性而引发的潜在损失,是对一些不确定事件在一定的时间周期内发生的概率和引发的潜在损失进行定量或定性的测量. 图 1 所示即为信息安全风险评估的 3 个主要因素[1]:威胁识别,脆弱性识别,资产识别.

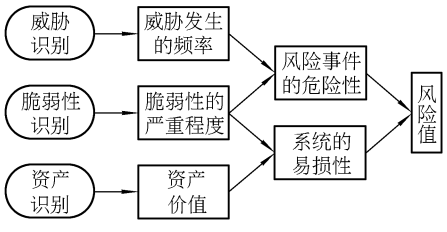

图 1　信息安全风险评估要素

评估总是基于一定的技术手段与评估模型,而评估方法的选择直接影响到评估过程中的每个环节,甚至可以左右最终的评估结果,所以需要根据系统的具体情况,选择合适的风险评估方法. 这里基于支持向量机,对信息安全风险进行评估[4~10],主要基于如下原因:**a**. 信息安全风险评估本质上属于有限样本(数据)的、非线性模式识别问题,支持向量机是专门针对有限样本情况的,它的目标是得到现有信息下的最优解而不仅仅是样本数趋向于无穷大时的最优解. **b**. 风险评估追求的是在现有信息情况下的最优解,支持向量机能将分类问题最终转化成一个二次寻优问题,从理论上将得到全局最优解,解决了在神经网络方法中无法避免得到局部极值的情况. **c**. 风险评估对问题解决方法的泛化能力以及简单性要求较高. 支持向量机能将实际问题通过非线性变换转到高维特征空间,在高维空间中构造线性判别函数使得原始空间具有了非线性判别函数的功能,不仅保证了模型的推广能力,并且解决了维数灾难问题.

2　评估模型

构造基于 SVM 信息安全风险评估模型的基本思路是:通过某种事先选择的非线性映射(核函数)将信息系统样本各个要素特征向量映射到一个高维特征空间,在这个空间中构造最优分类超平面. 通常,信息系统风险要素及安全性分若干个级别,即若干类. 先构造信息安全风险评估二分类评估模型,再在二分类评估模型的基础上,构建信息安全风险评估多分类评估模型,并给出基于 SVM 信息安全风险评估模型的实现方案.

2.1　二分类评估模型

假定有 n 个信息系统 $s_i\ (1 \le i \le n)$ 用作安全风险评估训练数据,它们的安全性分两类 $r_i\ (r_i \in \{1, -1\})$,即训练数据集为 $\{(s_i, r_i)\}$,$i = 1, 2, \cdots, n\}$,若 $s_i \in \mathbf{R}^{(\eta)}$ 属于第 1 类,则标记为正 $(r_i = 1)$;若属于第 2 类,则标记为负 $(r_i = -1)$. 目标是构造一个决策函数,将测试数据尽可能正确分类.

当训练集线性可分时,超平面 $\mathrm{k}s_i + m = 0$ (k 为权值向量,m 为分类阈值) 使两类样本完全分开;当训练数据线性不可分时,引入松弛变量 δ_i 和惩罚参数 π,则超平面优化问题变成

$$\min \psi(s_i) = \frac{1}{2} \| \mathrm{k} \|^2 + \zeta \sum_i^n \delta_i,$$

$$s_i(\mathrm{k}r_i + m) - 1 + \pi_i \ge 0, \pi_i \ge 0.$$

当训练集为非线性时,通过一个非线性函数 $\zeta = \phi(r_i)$ 将训练集数据 r_i 映射到一个高维线性特征空间,在映像空间(ζ 空间)中构造最优分类超平面,并得到评估模型的决策函数(如图 2 所示).

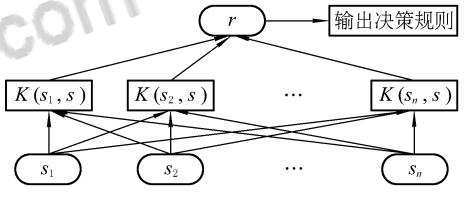

图 2　SVM 网络(核函数将输入空间映射到不同的特征空间)

因此,分类超平面为 $\mathrm{k}\phi(s_i) + m = 0$,决策函数为 $\tilde{f}(r_i) = \mathrm{sign}(\mathrm{k}\phi(r_i) + m)$,最优分类超平面问题的对偶最优化描述:

$$\max_{\alpha} L_{\mathrm{D}} = \sum_i^l \beta_i - \frac{1}{2} \sum_i^n \sum_j^n \beta\beta_j r_i r_j K(s_i, s_j);$$

$$0 \le \beta \le \pi, \quad \sum_i^n \beta r_i = 0,$$

式中 $K(s_i, s_j) = \phi(s_i)\phi(s_j)$ 称为核函数,此时决策函数可表示为 $\tilde{f}(s_i) = \mathrm{sign} \sum_{i=1}^{l} \beta r_i K(s_i, s_j) + m$.

这样,在求解最优化问题和计算决策函数时只需计算核函数,避免了特征空间维数灾难问题.

2.2　核函数

基于 SVM 的信息安全风险评估模型就是通过对训练数据集学习获得最优分类面,然后用该最优分类面对未知类别的样本进行风险评估,结合图 3 所示学习与评估过程,可以看出核函数是模型的核心. 不同的核函数,其评估模型的性能完全不同;另外,核函数中二次规划参数(如 δ 参数

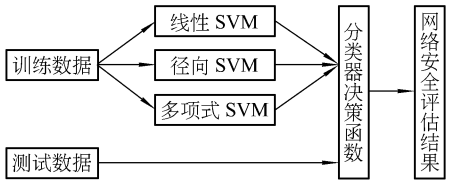

图 3 学习与评估过程

和 π 参数)对评估模型的泛化能力也有重要影响,因此,选择核函数至关重要. 本文采用线性函数 $K(s_i, s_j) = s_i s_j$,多项式函数 $K(s_i, s_j) = (s_i\ s_j + 1)^{\delta}$,径向基函数 $K(s_i, s_j) = \exp(-\|s_i - s_j\|^2/\pi^2)$,共 3 类核函数对训练数据进行测试,构造出不同的分类函数,并对分类结果进行了比较.

2.3 多分类评估模型

前述为信息安全风险的二分类评估模型,通常,信息系统风险要素及安全性分若干个级别,即分若干类,应该构建多分类评估模型. 对此本研究在训练阶段采用一对一的模式,即在 n 类样本数据中任取 2 类训练数据进行训练,构造出所有可能的二分类评估模型,进而构造得到信息安全风险的多分类评估模型.

2.4 模型实现

多分类评估模型的实现如图 4 所示,其中训练数据与测试比例可以根据实际情况进行设定.

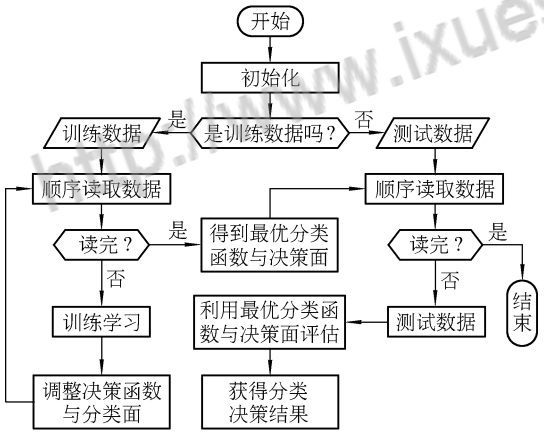

图 4 模型实现

3 仿真实验

把影响信息安全风险评估的 3 个主要因素(威胁识别,脆弱性识别,资产识别)分成 16 个具体因素,即:信息被窃、删除或丢失,网络资源被破坏,信息滥用、讹用或篡改,服务的中断和禁止,信息泄露,硬件缺陷,软件缺陷,网络脆弱性,通信协议脆弱性,环境恶化,数据泄露,通信被干扰,信息丢失,信息、服务恢复,中断、延迟,削弱. 对每个因素赋予权值,作为输入数据,将评估结果分为安全、疑似和危险 3 个类别,作为输出数据.

a. 将数据分成训练数据和测试数据两部分,针对训练数据,若每一个训练数据可表示成 1 × 16 维的行向量,即 $R_m = [A_{m,0}, A_{m,1}, \cdots, A_{m,15}]^{\mathrm{T}}$,则整个信息系统安全性能指标矩阵为 $R = [R_0, R_1, \cdots, R_{M-1}]$. 将这 M 个项目安全性指标矩阵作为 SVM 的训练数据集,把 M 个项目的每个结果按安全性水平 $L_m (m \in [0, M-1])$ 高低分为 3 类:LM 表示安全性水平很低,MM 表示安全性水平中等,HM 表示安全性水平较高. 这些项目的安全性 L_m 作为 r_m,即 $L = Y = [L_0, L_1, \cdots, L_{M-1}]$. 在具体实验中,选取了 14 个已有评估结果的数据作为训练数据,如表 1 所示,其中每个样本的 16 种安全性能指标已经计算得到,表 2 表示 5 个测试样本的 16 个安全性能指标.

b. 通过训练学习,寻找支持向量,构造决策函数,求解最优分类面.

针对表 1 所给训练数据,将该多分类问题利用文中一对一的方法,转化成 3 个二分类问题,对训练数据作非线性变换 $\delta = \phi(s_i)$,使其成为线性可分. 在映像空间(δ 空间)中构造最优分类超平面 $k\phi(s) + m = 0$,并得到评估模型的决策函数 $\tilde{f}(s) = \mathrm{sign}(k\phi(s) + m)$.

c. 利用决策函数和最优分类面对测试样本进行分类实验,考察其评估性能.

对测试数据进行安全性评估时,利用式(2)求出的决策函数和分类面,把每个测试样本的 16 种安全性能指标作为 SVM 的输入向量,输出层即可得到该项目的可维护性水平. 通过与专家评估结果进行对比,考察其评估效果.

在模拟实验中,采用 Matlab 作为测试平台,在进行网络训练时,对于给定数据,主要考察如下 3 种核函数:线性函数 $K(s_i, s_j) = s_i s_j$;基于径向基函数 RBF 内积函数形式 $K(s_i, s_j) = \exp(-\|s_i - s_j\|^2/d)$,假定 $d = 50$;多项式形式的内积函数 $K(s_j, s_i) = (s_j s_i + 1)^d$,假定 $d = 5$.

通过对表 1 中的 14 个训练数据进行网络训练,发现采用径向 SVM,对系统安全性评估的准确率达到了 91 %,高于以线性 SVM 和多项式 SVM 评估效果,且该方法对表 2 中 5 个测试数据的安全性评估准确率达到了 100 %(见图 5). 对于本次实验数据,以径向基函数线性 SVM 相比以线性函数或以多项式形式的内积函数构造的支持向量机具有更高的评估准确率. 利用基于 SVM 的信息安全风险评估模型,可以较为准确地评估出信息系统的风险水平,以便更科学地防范风险.

表 1　训练数据

数据编号	L_0	L_1	L_2	L_3	L_4	L_5	L_6	L_7	L_8	L_9	L_{10}	L_{11}	L_{12}	L_{13}	L_{14}	L_{15}	安全性水平
1	0.1	0.3	0.3	0.5	0.3	0.3	0.1	0.3	0.3	0.1	0.3	0.1	0.3	0.1	0.3	0.3	LM
2	0.5	0.3	0.3	0.5	0.3	0.3	0.5	0.5	0.3	0.5	0.3	0.7	0.5	0.5	0.3	0.3	MM
3	0.5	0.3	0.5	0.3	0.5	0.3	0.5	0.3	0.5	0.3	0.3	0.3	0.3	0.5	0.3	0.5	MM
4	0.7	0.7	0.5	0.7	0.5	0.7	0.7	0.5	0.5	0.5	0.5	0.5	0.3	0.7	0.7	0.5	HM
5	0.1	0.3	0.1	0.1	0.1	0.1	0.3	0.3	0.1	0.1	0.3	0.3	0.7	0.3	0.1	0.3	LM
6	0.3	0.1	0.3	0.3	0.3	0.5	0.1	0.5	0.5	0.1	0.3	0.3	0.1	0.3	0.7	0.3	LM
7	0.5	0.3	0.3	0.5	0.5	0.5	0.1	0.5	0.3	0.5	0.3	0.5	0.5	0.3	0.5	0.3	MM
8	0.3	0.3	0.1	0.3	0.1	0.3	0.5	0.1	0.3	0.1	0.1	0.3	0.3	0.1	0.5	0.1	LM
9	0.1	0.5	0.3	0.3	0.5	0.5	0.5	0.1	0.5	0.3	0.1	0.5	0.5	0.3	0.7	0.5	MM
10	0.7	0.5	0.3	0.3	0.9	0.7	0.5	0.5	0.3	0.5	0.7	0.7	0.5	0.5	0.7	0.5	HM
11	0.3	0.5	0.1	0.3	0.3	0.1	0.3	0.3	0.3	0.1	0.3	0.1	0.3	0.3	0.1	0.3	LM
12	0.3	0.3	0.1	0.3	0.3	0.1	0.3	0.1	0.3	0.1	0.1	0.5	0.1	0.3	0.3	0.5	LM
13	0.3	0.5	0.5	0.3	0.3	0.3	0.5	0.3	0.3	0.3	0.5	0.5	0.3	0.7	0.3	0.1	MM
14	0.7	0.3	0.5	0.5	0.3	0.5	0.5	0.5	0.7	0.7	0.5	0.7	0.5	0.5	0.5	0.7	HM

表 2　测试数据

数据编号	L_0	L_1	L_2	L_3	L_4	L_5	L_6	L_7	L_8	L_9	L_{10}	L_{11}	L_{12}	L_{13}	L_{14}	L_{15}	安全性水平
1	0.3	0.3	0.5	0.3	0.5	0.3	0.3	0.3	0.1	0.3	0.1	0.7	0.1	0.1	0.3	0.5	LM
2	0.5	0.3	0.5	0.7	0.5	0.1	0.3	0.7	0.5	0.3	0.7	0.5	0.5	0.3	0.3	0.3	HM
3	0.5	0.3	0.3	0.3	0.3	0.5	0.3	0.3	0.3	0.7	0.5	0.3	0.3	0.5	0.3	0.5	MM
4	0.5	0.1	0.3	0.5	0.3	0.1	0.3	0.3	0.3	0.3	0.3	0.1	0.3	0.3	0.3	0.5	LM
5	0.1	0.5	0.1	0.1	0.3	0.1	0.1	0.1	0.5	0.3	0.1	0.3	0.3	0.3	0.5	0.3	LM

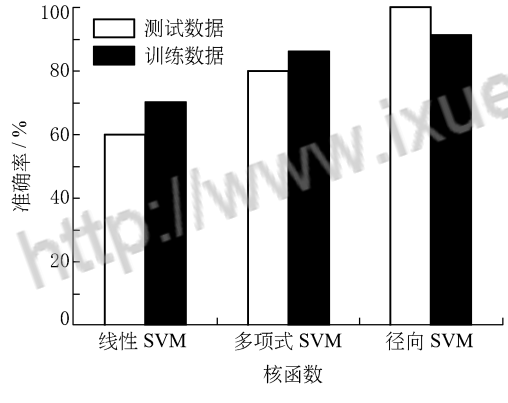

图 5　算法性能比较

参　考　文　献

[1] 冯登国,张　阳, 张玉清. 信息安全风险评估综述[J]. 通信学报, 2004, 25(7): 10-18.

[2] 李红莲,王春花,袁保宗,等. 针对大规模训练集的支持向量机的学习策略[J]. 计算机学报, 2004, 27(5): 715-719.

[3] 张　铭,银　平,邓滞洪,等. SVM + Bi HMM:基于统计方法的元数据抽取混合模型[J]. 软件学报, 2008, 19(2): 358-368.

[4] 康威新,彭喜元. 基于二层 SVM 多分类器的桩基缺陷诊断[J]. 电子学报, 2008, 36(12): 66-70.

[5] Li Zhifeng. Using support vector machines to enhance the performance of bayesian face recognition [J]. IEEE Transactions on Information Forensics and Security, 2007, 2(2): 174-180.

[6] Singh Y, Kaur A, Malhotra R. Application of support vector machine to predict fault prone classes[J]. ACM, SIGSOFT Software Engineering Notes, 2009, 34(1): 1-6.

[7] Franc V, Lsakov P, Miiller K R. Stopping conditions for exact computation of leave-one-out error in support vector machines[C] //ACM Proceedings of the 25th International Conference on Machine Learning. Helsinki: ACM, 2008: 328-335.

[8] 姜映映,田　丰,王绪刚,等. 基于模板匹配和 SVM 的草图符号自适应识别方法[J]. 计算机学报, 2009, 32(2): 252-260.

[9] 王瑞平,陈　杰,山世光,等. 基于支持向量机的人脸检测训练集增强[J]. 软件学报, 2008, 19(11): 2 921-2 931.

[10] Dlamini M T, Eloff J H P, Eloff M M. Information security: the moving target[J]. Computers & Security, 2009, 28(1): 189-198.

阅读此文的还阅读了：

1. 定性的信息安全评估方法研究与应用
2. 基于支持向量机的企业信用风险评估研究
3. 信息系统(网络)安全分析方法与评价模型
4. 基于支持向量机的综合评估方法的应用研究
5. 基于WOWA的信息安全风险模糊评估
6. 信息安全问题研究与对策
7. 基于支持向量机的房地产投资环境风险评估
8. 基于支持向量机和云模型的网络健康状态评估
9. 基于因子分析和支持向量机的电网故障风险评估
10. 基于支持向量机的桁架桥可靠度评估
11. 基于聚类和支持向量机的个人信誉评估方法
12. 基于主成分分析和支持向量机的商业银行信贷风险评估
13. 基于粗糙集理论与支持向量机的纳税评估模型
14. 网络安全评估方法的研究与实践
15. 基于支持向量机的房地产投资环境风险评估
16. 关于信息安全标准体系建设的思考

17. 货币银行学：基于支持向量机的信用评估模型及风险

18. 基于支持向量机的信息安全风险评估

19. 信息安全风险评估综述

20. 金融时间序列数据预测方法探析

21. 基于支持向量机的企业信用评估模型

22. 基于支持向量机的上市公司信用风险评估

23. 基于支持向量机的舰船建造预付款风险评估

24. 基于支持向量机的环境质量评估方法

25. 基于支持向量机的供应链风险评估研究

26. 基于支持向量机的作战方案评估

27. 基子支持向量机的企业赊销风险评估摸型

28. 基于可拓集的信息安全风险评估

29. 评估安全认证 选择最适合您的认证

30. 基于支持向量机的海战场辐射源威胁评估

31. 基于支持向量机的网络风险评估方法的研究

32. 基于AHP的信息安全风险评估方法研究

33. 基于支持向量机的个人信用评估模型研究

34. 基于粗糙-支持向量机的航运企业客户信用评估

35. 基于网络舆情安全的信息挖掘及评估指标体系研究

36. 基于支持向量机的新能源汽车状态评估

37. 基于模糊支持向量机的客户信用评估研究

38. 信息安全评估标准中电磁兼容检测的实施探讨

39. 基于差分进化支持向量机的作战效能评估方法

40. 信息安全风险评估分析方法简述

41. 一种基于支持向量机的服务质量评估方案

42. 基于支持向量机的溃坝生命损失评估模型及应用

43. 实施信息系统灾难防御评估的过程模型设计

44. 基于粗糙集和支持向量机的商业银行信用风险评估模型

45. 信息安全问题研究与对策

46. 基于粗糙集支持向量机的个人信用评估模型

47. 支持向量机的智能信息安全风险评估模型

48. 基于支持向量机的上市公司信用风险评估研究

49. 基于支持向量机委员会机器的个人信用评估模型

50. 工信部：召开政府机关力公用计算机安全配置试点工作总结评估会议