计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2018,27(7):188−192 [doi: 10.15888/j.cnki.csa.006251]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于支持向量机和自适应权重的网络安全态势评估模型①

胡　柳[1], 周立前[2], 邓　杰[1], 李　瑞[3], 赵正伟[4]

[1](湖南信息职业技术学院 计算机工程学院, 长沙 410200)
[2](湖南工业大学 计算机学院, 株洲 412007)
[3](河北省眼科医院 信息科, 邢台 054000)
[4](广西民族大学 信息科学与工程学院, 南宁 530006)

摘　要: 网络安全是近年来国内的重点研究领域, 没有网络安全就没有国家安全. 针对网络安全中数据源多样性、复杂性等特点, 提出了一种基于支持向量机及自适应权重的网络安全态势评估模型. 该模型由训练和预测模块组成, 训练模块采用先验知识方法获取网络安全关注的重点数据, 结合支持向量机和权重策略寻求评估模型. 预测模块进行实时网络安全态势评估. 通过实验过程和结果分析, 表明该模型较好的支持小型网络安全态势的实时预测评估.

关键词: 网络安全; 支持向量机; 安全态势; 评估模型

引用格式: 胡柳,周立前,邓杰,李瑞,赵正伟.基于支持向量机和自适应权重的网络安全态势评估模型.计算机系统应用,2018,27(7):188–192. http://www.c-s-a.org.cn/1003-3254/6251.html

Evaluation Model of Network Security Situation Based on Support Vector Machine and Self-Adaptive Weight

HU Liu[1], ZHOU Li-Qian[2], DENG Jie[1], LI Rui[3], ZHAO Zheng-Wei[4]

[1](College of Computer Engineering, Hunan College of Information, Changsha 410100, China)
[2](College of Computer, Hunan University of Technology, Zhuzhou 412007, China)
[3](Information Section, Hebei Eye Hospital, Xingtai 054000, China)
[4](College of Information Science and Engineering, Guangxi University for Nationalities, Nanning 530006, China)

Abstract: Network security has become a priority research field in China in recent years, for there would be no national security without network security. In consideration of the characteristics of data source in network security like diversity and complexity, this study proposes an evaluation model of network security situation based on Support Vector Machine (SVM) and self-adaptive weight. The model is composed of training module and prediction module. The training module is used to obtain the key data concerned by network security through prior knowledge method and build evaluation model with a combination of SVM and weight strategy. The prediction module is used for real-time network security situation evaluation. Analysis of experimental process and results indicates that the proposed model can favorably support the real-time prediction and evaluation of the security situation of small-size networks.

Key words: network security; Support Vector Machine (SVM); security situation; evaluation model

网络系统面临不断变化的蠕虫病毒、DDOS、数据库注入、挂马等常规风险, 传统的网络安全设备, 如网络防火墙、应用网关、IDS、IPS 等设备, 具有相应的防护及保护能力, 在获取到有危险的攻击时, 需要进行多层处理, 从而导致众多不确定性且存在误报率和漏报率. 据国家互联网应急中心发布的 CNCERT 互联

① 基金项目: 湖南省自然科学基金 (2016JJ5036)
Foundation item: Natural Science Foundation of Hunan Province (2016JJ5036)
收稿时间: 2017-06-15; 修改时间: 2017-07-12; 采用时间: 2017-07-14; csa 在线出版时间: 2018-06-27

网安全威胁报告 (2016 年 10 月) 显示, 当月境内感染网络病毒的终端数量为 194 万余个, 境内被篡改的网站数量为 4524 个, 国家信息安全漏洞共享平台 (CNVD) 收集整理信息系统安全漏洞 1284 个[1]. 网络安全态势评估技术能综合各方面的因素, 从全局分析并响应网络安全状况, 为网络安全提供可靠的参照依据. 近年来, 针对网络安全态势有大量研究人员进行了详细的实证研究, 设计并实现了大量网络安全态势评估模型或系统.

文献[2]提出一种基于时空维度分析的网络安全态势预测方法, 从攻击方、防护方和网络环境 3 方面提取网络安全态势评估要素, 从时间维度和空间维度进行网络安全态势预测, 通过对公用数据集网络的测评分析, 该方法符合实际应用环境, 提高了安全态势感知的准确性. 文献[3]提出基于神经网络的网络安全态势感知方法, 设计了一种基于 BP 神经网络的网络安全态势评估方法, 为了解决要素与结果之间的不确定性及模糊性问题, 提出了基于 RBF 神经网络的网络安全态势预测方法, 并采用自适应遗传算法对网络参数进行优化. 通过真实网络环境验证了方法的可行性和有效性. 文献[4]提出了集对分析的可信网络安全态势评估与预测方法, 采用基于特征库的方法审计网络连接信息、系统管理信息、系统监控信息和应用服务信息, 结合改进的熵权法和层次分析法提取安全态势指标权重, 并利用集对分析方法对安全态势指标进行评估, 进而绘制网络安全态势图. 仿真实验结果表明, 该方法能够准确有效地反映当前及未来的网络安全态势.

另外, 文献[5]提出了一种基于 D-S 证据理论的网络安全态势预测方法, 文献[6]提出了一种基于 Elman 神经网络的网络安全态势预测方法, 文献[7]提出了一种基于 Markov 博弈模型的网络安全态势感知方法, 文献[8]提出了基于信息融合的网络安全态势评估模型, 文献[9]提出了一种基于知识发现的网络安全态势建模与生成框架, 文献[10]提出了一种实时的网络安全态势预测方法等.

针对小型网络内节点数据的特征, 本文提出基于支持向量机及自适应权重的网络安全态势预测模型, 结合先验知识方法提高分类的准确性, 通过在网络中采集各节点网络数据中的特征量作为样本进行实验与分析, 以验证模型的有效性.

1 支持向量机原理

支持向量机 (SVM) 是一种可训练的机器学习方法, 在解决小样本、非线性和高维模式识别方面具有优势. Vapnik 等人在研究统计学理论基础上对线性分类器提出一种设计最佳准则, 从线性可分扩展到线性不可分的情况, 其主要思想有二点[11]:

(1) 它采用非线性映射, 将低维空间转换到高维空间从而实现线性可分, 采用核函数方法实现升维.

(2) 在特征空间中构建最优分割超平面, 得到全局最优的学习器.

通过非线性映射, 将低维样本特征空间映射到高维特征空间, 这一过程可能会出现维数灾难, SVM 利用核函数来解决这一问题, 通常使用的核函数有:

线性核函数:

$$Y(a,b)=a\cdot b \tag{1}$$

RBF 径向基:

$$Y(a,b)=\exp(\frac{-|a-b|^2}{d^2}) \tag{2}$$

多项式核函数:

$$Y(a,b)=[(a\cdot b)+1]\cdot d \tag{3}$$

Sigmoid 函数:

$$Y(a,b)=\tanh(x\cdot(a\cdot b)+y) \tag{4}$$

LibSVM 由台湾大学林智仁等开发设计的一个 SVM 模式识别与回归的软件包, 它能通过简单的配置完成分类问题, 拥有 Java、Matlab、Ruby、Perl、Python、C、LabView 等数十种语言版本[12], 本文将采用 LibSVM 进行样本特征量值进行训练, 在小型网络中采集网络数据核心数据并进行归整, 获取分类模型并进行验证.

2 基于支持向量机及自适应权重的网络安全态势评估模型

2.1 态势评估等级划分

由于网络及其附属设备具有复杂性, 网络中存在大量不确定性的因素, 根据其危害程度可将网络安全评估等级分为四级, 具体级别、安全指数及相关说明如表 1 所示.

表 1 中各级别都对应有网络的脆弱性、可控性、可恢复性、容灾等特性等, 这里不再展开阐述.

表 1 网络安全态势评估等级表

序号	级别	安全指数范围	说明
1	相对安全	0.0–0.2	网络内未出现恶意攻击或病毒, 网络运转正常, 未出现严重的安全漏洞或服务器安全漏洞
2	轻度危险	0.2–0.5	网络内出现轻度安全风险, 出现已知病毒和安全漏洞, 存在恶意攻击但未出现系统重要服务中断
3	中度危险	0.5–0.8	网络内出现中度安全风险, 出现大量病毒和安全漏洞, 存在恶意攻击并已造成部分系统服务中断, 影响网络的正常使用
4	高度危险	0.8–1.0	网络内出现高度安全风险, 出现大量病毒和各类安全漏洞, 存在大量恶意攻击、SQL 注入和系统服务漏洞, 严重影响网络的正常使用

2.2 样本特征选取

2.2.1 数据项选取

数据项的选取直接影响到实验与分析的效果, 在实验过程中将采用基于 Snort 的入侵检测系统、防火墙和 X-Scan 漏洞扫描相结合的环境, 数据项的选取直接从上述三者中获取.

(1) Snort 数据项

Snort 是一种基于 libpcap 的多平台、实时的流量分析工具, 能有效记录 IP 数据包. 其重要的字段有: 动作、协议、源和目标地址、源和目标端口、告警产生时间、告警等级等.

(2) 防火墙数据项

防火墙可以记录内外网络之间的网络通信, 有效保护内部网络不受侵害. 可以获取的重要字段有: 产生时间、源和目的地址、源和目的端口、协议类型、持续时间、发送字节、接收字节、操作行为等.

(3) X-Scan 数据项

X-Scan 主要进行漏洞风险扫描并进行等级评估. 通过在内部进行指定 IP 范围的扫描可以获取以下信息: 操作系统类型与版本、端口状态、端口 BANNER、CGI 漏洞、IIS 漏洞、RPC 漏洞及各类服务器弱口令等.

2.2.2 指标项

网络数据复杂多样, 不同的指标项对实验结果将产生一些偏差, 本文在数据项选取的基础上再增加以下指标:

(1) 连续告警时间: 设 t 时刻网络设备捕捉到告警信息, 并在未来 $t+n$ 时刻内连续告警时间, 连续告警时间 IP_n 为:

$$IP\_n = \sum_{i=1}^{n} D_i^s \cup D_i^f \cup D_i^x \tag{5}$$

其中, n 为连续的时间间隔, D_i^s为 Snort 在 i 时刻捕捉到的告警信息, D_i^f为防火墙在 i 时刻捕捉到的操作告警行为, D_i^x为 X-Scan 在 i 时刻捕捉到的漏洞告警信息. n 的选值范围为$+\infty$, 当未捕捉到告警信息时 n 值清零.

(2) 端口危险程度: 设在固定时间段 s 内, 各项监测数据中捕捉到的危险项端口序列为 $D=\{D_1, D_2, D_3,\cdots, D_n\}$, 在 s 内端口序列中各端口受功击的值为 $P=\{P_1, P_2, P_3,\cdots, P_n\}$.

$$P_n = \sum_{i=1}^{m} D_n | W_i \tag{6}$$

其中, m 为捕捉到的网络数据量, W_i 为当前网络数据中 D_n 端口危险级别为非安全.

(3) 漏洞危险识别指数: 设在固定时间段 s 内, 各项监测数据中捕捉的对漏洞序列 $L=\{L_1, L_2, L_3,\cdots, L_n\}$, 在 s 内漏洞序列中各漏洞受功击的值为 $T=\{T_1, T_2, T_3,\cdots, T_n\}$.

$$T_n = \sum_{i=1}^{m} L_n | W_i \tag{7}$$

其中, m 为捕捉到的网络数据量, W_i 为当前网络数据中 L_n 漏洞危险级别为非安全.

2.2.3 数据归一化

为了使数据项和指标值符合支持向量机的参数要求, 通过统一的方式进行数据归一化, 使其范围在[0, 1]范围内, 根据本文实际需要, 数据项和指标值的归一化公式如下:

$$F(i) = \begin{cases} \dfrac{D_i}{100} & , \quad D_i < 100 \\ 1 & , \quad D_i \geq 100 \end{cases} \tag{8}$$

其中, $F(i)$ 为归一化后的指标值, D_i 为当前指标值.

2.3 自适应权重策略

支持向量机较好的支持两类问题的分类, 本文通过将数据项和指标值输入预测模型, 得到的结果为安全和危险, 还不能较好深入区分当前网络的安全态势. 在安全和危险的结果之上, 根据表 1 设计自适应权重策略方法进行深入的安全态势细分.

根据表 1 的安全态势等级划分, 将安全指数

0.0–0.2 之间归为相对安全, 0.2–0.5 为轻度危险, 0.5–0.8 为中度危险, 0.8–1.0 为高度危险, 利用 IP_n、P_n、T_n 的归一值按不同权重策略进行划分, 初始权重分配依次为: 40%、30%、30%. 权重策略计算结果 $D=IP\_n*0.4+P_n*0.3+T_n*0.3$.

在实验过程中通过手动截取的数据, 根据实际网络数据不断优化权重值, 使态势评估结果与实际结果接近. 最终的网络安全态势评估结果判定如表 2 所示.

表 2 网络安全态势评估结果判定表

序号	权重值	分类值	网络安全态势结果
1	0.0–0.2	0	相对安全
		1	轻度危险
2	0.2–0.5	0	相对安全
		1	轻度危险
3	0.5–0.8	0	轻度危险
		1	中度危险
4	0.8–1.0	0	中度危险
		1	高度危险

2.4 模型工作流程

模型工作流程图如图 1 所示.

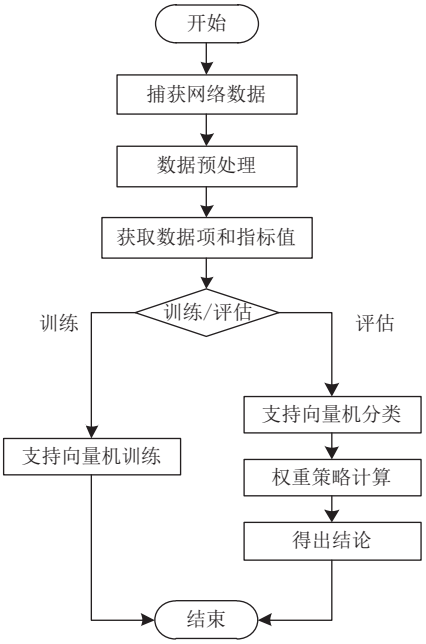

图 1 模型流程图

3 实验与分析

3.1 实验环境

根据本文提出的网络安全态势评估模型, 搭建的实验环境如图 2 所示.

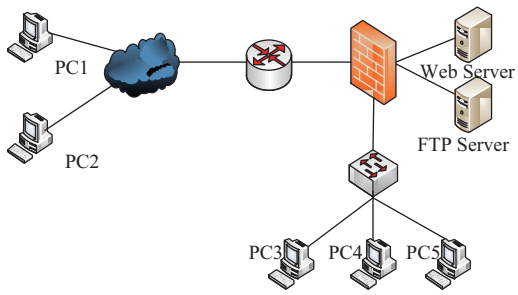

图 2 实验环境

通过在网络内部署监测工具, 同时通过以下攻击方式使监测工具能获取到相应的网络数据:

(1) PC1 上传带有病毒的文件到 FTP Server.

(2) PC1 对 Web Server 进行数据库注入攻击.

(3) PC2 对 PC5 发送大容量数据包.

(4) PC2 嗅探服务器上的漏洞.

(5) PC1 扫描内网中存在的弱口令.

(6) PC2 扫描内网 Server 和 PC3、PC4、PC5 上开放的端口.

通过在真实的网络环境中采集 2.2 节中所涉及的数据项和指标值, 为网络安全态势评估提供可靠的数据源.

3.2 预测模型获取

预测模型是将获取的已知分类数据进行训练, 获取相应的预测模型文件. 通过前期的网络数据采集, 共采集到样本数据 2200 条, 将其以 Libsvm 支持的格式进行存储, 部分样本数据如图 3 所示.

```
1 1:0.793170276 2:0.444317234 3:0.453661944 4:0.704229456 5:0.365277476 6:0.896781249
1 1:0.540333290 2:0.151398311 3:0.494424888 4:0.305281125 5:0.258466836 6:0.496818898
1 1:0.307052545 2:0.838582572 3:0.912151410 4:0.249957024 5:0.405644974 6:0.944440425
1 1:0.320071392 2:0.202339248 3:0.924995866 4:0.919566528 5:0.282501218 6:0.645029008
1 1:0.521185161 2:0.776124462 3:0.378756743 4:0.800672449 5:0.025873546 6:0.210473761
1 1:0.663277257 2:0.347967303 3:0.673658122 4:0.861916409 5:0.749339991 6:0.358234852
1 1:0.856655239 2:0.900832199 3:0.950220810 4:0.166124099 5:0.460986062 6:0.494492068
1 1:0.513943970 2:0.388833146 3:0.209518129 4:0.769899747 5:0.063621746 6:0.646415782
1 1:0.165371857 2:0.636386390 3:0.897165051 4:0.618758008 5:0.354613507 6:0.709649413
1 1:0.851459770 2:0.456389720 3:0.145889770 4:0.730570607 5:0.786065499 6:0.609277933
1 1:0.119191750 2:0.995197830 3:0.138053279 4:0.135194517 5:0.774028901 6:0.668903449
1 1:0.257355854 2:0.967917483 3:0.729577604 4:0.022296936 5:0.654899113 6:0.390861763
1 1:0.070730510 2:0.670798799 3:0.542452158 4:0.145877377 5:0.977270333 6:0.759910852
1 1:0.332560486 2:0.442729414 3:0.056110139 4:0.775782138 5:0.165540503 6:0.465384878
1 1:0.195628302 2:0.201588176 3:0.510894918 4:0.203072966 5:0.398388119 6:0.435716142
1 1:0.158750052 2:0.511868250 3:0.566710508 4:0.891531471 5:0.024454845 6:0.701624608
```

图 3 采集的训练数据

图 4 中的数据是进行了归一化后的采集到的网络数据值, 通过 Libsvm 提供的 Svmtrain 进行训练, 获取预测模型文件.

3.3 实验预测

利用获取的预测模型文件和当前获取的数据项和指标项进行网络安全态势预测, 结合 自适应权重策略进行深入分类, 在固定时间 1 min 内采集到的网络数据

开展预测. 选取参与预测的 10 条记录, 其评估结果如表 3 所示.

表 3 部分预测结果

编号	预测样本	预测结果	权重结果	评估结果
1	0.312 0.844 0.099 0.852 0.964 0.066	1	0.650	中度危险
2	0.135 0.132 0.879 0.879 0.583 0.120	0	0.563	轻度危险
3	0.605 0.795 0.801 0.138 0.166 0.293	1	0.193	轻度危险
4	0.114 0.617 0.724 0.590 0.175 0.069	0	0.309	相对安全
5	0.047 0.535 0.079 0.898 0.770 0.980	0	0.884	中度危险
6	0.593 0.647 0.153 0.070 0.210 0.046	1	0.105	轻度危险
7	0.799 0.979 0.541 0.729 0.192 0.685	1	0.555	中度危险
8	0.286 0.799 0.584 0.106 0.771 0.705	0	0.485	相对安全
9	0.669 0.907 0.017 0.718 0.000 0.965	1	0.577	中度危险
10	0.768 0.707 0.829 0.695 0.910 0.996	1	0.850	高度危险

经过与人工验证样本数据的结果对比, 上述 10 条记录中有 2 条与结果存在出入, 在调整权重值为 42%、25%、33% 之后, 记录中有 1 条与结果存在出入.

根据测试数据的实际预测值和期望值进行对比, 如图 4 所示.

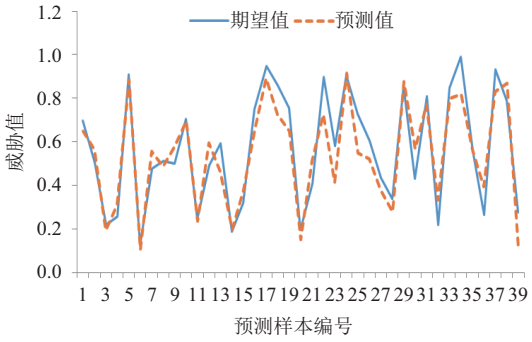

图 4 期望值与预测值的对比

实验结果表明, 利用支持向量机和自适应权重策略的网络安全态势评估模型在小型局域网内对网络安全的预测能进行有效的预测, 期望值与预测值结果比较接近, 能实时对小型局域网络安全进行有效评估与监控.

4 结论

本文构建了一个基于支持向量机和自适应权重策略的网络安全态势评估模型, 通过设计网络数据特征项和指标值, 利用 Libsvm 和自适应权重策略对采集的样本进行训练, 并采集了部分网络数据进行预测分析.

网络态势评估模型设计过程中需要人工干预、分析、整理相关网络数据开展模型训练, 然后实时采集网络中的数据并提取相应的指标值进行评估, 给出评估结果, 为网络管理人员提供了网络安全管理参考. 本实验由于受样本数据量和网络环境的限制, 互联网中存在的网络攻击方式层出不穷, 未来还需要对攻击方式、指标项、预测方法进行深入的分析, 改进计算算法, 实现对网络安全态势更有效的评估.

参考文献

1 国家互联网应急中心. CNCERT 互联网安全威胁报告(2016 年 10 月). http://www.cert.org.cn/publish/main/upload/File/2016monthly10.pdf. [2016-10/2016-12-26].
2 刘玉岭, 冯登国, 连一峰, 等. 基于时空维度分析的网络安全态势预测方法. 计算机研究与发展, 2014, 51(8): 1681–1694. [doi: 10.7544/issn1000-1239.2014.20121050]
3 谢丽霞, 王亚超, 于巾博. 基于神经网络的网络安全态势感知. 清华大学学报 (自然科学版), 2013, 53(12): 1750–1760.
4 吴琨, 白中英. 集对分析的可信网络安全态势评估与预测. 哈尔滨工业大学学报, 2012, 44(3): 112–118. [doi: 10.11918/j.issn.0367-6234.2012.03.022]
5 石波, 谢小权. 基于 D-S 证据理论的网络安全态势预测方法研究. 计算机工程与设计, 2013, 34(3): 821–825.
6 尤马彦, 凌捷, 郝彦军. 基于 Elman 神经网络的网络安全态势预测方法. 计算机科学, 2012, 39(6): 61–63, 76.
7 张勇, 谭小彬, 崔孝林, 等. 基于 Markov 博弈模型的网络安全态势感知方法. 软件学报, 2011, 22(3): 495–508.
8 韦勇, 连一峰, 冯登国. 基于信息融合的网络安全态势评估模型. 计算机研究与发展, 2009, 46(3): 353–362.
9 王春雷, 方兰, 王东霞, 等. 基于知识发现的网络安全态势感知系统. 计算机科学, 2012, 39(7): 11–17, 24.
10 黄同庆, 庄毅. 一种实时网络安全态势预测方法. 小型微型计算机系统, 2014, 35(2): 303–306.
11 胡柳. 基于 SVM 的不良图片过滤研究与系统实现[硕士学位论文]. 株洲: 湖南工业大学, 2013.
12 Chang CC, Lin CJ. LIBSVM-a library for support vector machines. http://www.csie.ntu.edu.tw/~cjlin/libsvm. [2016-12-22/2017-01-03].