

一种基于混合核函数 PSO_SVR 的网络安全态势预测方法

李方伟, 罗 嘉, 朱 江, 张海波

(重庆邮电大学 移动通信技术重庆市重点实验室, 重庆 400065)

摘 要: 为了对错综复杂的网络安全形势做出可靠的预测, 提出了一种基于混合核函数 PSO_SVR 的网络安全态势预测模型. 本模型针对基于传统支持向量机(SVR)的网络安全态势预测模型精度不够高, 其核函数的选择及参数的设定没有统一标准的情况, 构造了一种兼顾插值能力和外推性能的混合核函数. 并引入粒子群算法(PSO)对基于混合核函数的 SVR 进行参数寻优, 有效地提高了 SVR 预测能力. 通过仿真实验表明, 该模型相比与传统的网络安全态势预测方法, 预测精度上更有保障.

关键词: 态势预测; 网络安全; 混合核函数; 粒子群算法; 支持向量机

中图分类号: TP393.08; TP18

文献标识码: A

文章编号: 1000-7180(2015)12-0110-06

A Method of Network Security Situation Prediction Based on Hybrid Kernels PSO-SVR

LI Fang-wei, LUO Jia, ZHU Jiang, ZHANG Hai-bo

(Chongqing Key Lab of Mobile Communications Technology, Chongqing University of Posts
and Telecommunications(CQUPT), Chongqing 400065, China)

Abstract: In order to predict the complicated network security situation reliably, a hybrid network security situation predictive model based on kernel function PSO_SVR is proposed. To solve the problem that the accuracy of prediction is low and there is no uniform to set the parameters, we construct a hybrid kernel function whose interpolation and extrapolation performance is good. The particle swarm optimization (PSO) based on hybrid kernel function is introduced to search the optimization parameters. The simulation results show that the model compared with the traditional network security situation prediction method is more secure than in the prediction accuracy. The predictions are more scientific and reliable.

Key words: situation prediction; network security; hybrid kernels; PSO; SVM

1 引言

随着信息网络中各种新型技术的广泛运用, 网络安全态势感知技术应运而生. 网络安全态势感知主要分为网络安全态势要素获取、网络安全态势评估、网络安全态势预测三个方面^[1]. 其中态势预测是态势感知中关键技术. 网络安全态势预测能够主动地掌握网络中将要出现的变化, 在为网络决策者提供主动的决策依据的同时有效地减轻管理者的压

力. 网络安全态势预测实际是一个具有非线性、时变性等特点的回归问题^[2], 一些传统的时间序列预测方法不能准确反映出网络当前状态和未来状态的关系. 随着人工智能算法的发展, 目前态势预测方法主要包括基于神经网络^[3]、马尔可夫链^[4]、支持向量机(SVM)^[5]等方法. 这些方法预测精度都能够有所保证, 但是通过研究发现: 基于神经网络的预测方法具有网络收敛速度慢和易陷入局部最优的缺点; 基于马尔可夫的预测方法难以建立准确预测模型, 在实

收稿日期: 2015-03-12; 修回日期: 2015-05-09

基金项目: 国家自然科学基金项目(61271260); 教育部科学研究重点项目(212145); 重庆市教委科学技术研究项目(KJ1400405)

际应用中难以建立。

支持向量回归算法(SVR)相对于其他预测算法有着收敛速度快、不易陷入局部寻优等优点。在SVR中核函数和参数的选择是影响预测精度的关键问题,SVR没有固定的选取核函数与参数的方法,一般都是通过实验比较调节。现在对核函数的选择与优化在态势预测的运用中研究不多,而核函数的选择决定了支持向量机中的学习能力与泛化能力(对未知样本的预测能力)。

本文在分析各个核函数的特点的基础上,采用了一种局部函数与全局函数相结合的混合核函数,它在学习能力与泛化能力两方面取得良好的折衷,并以此构造的SVR提出了基于混合核函数SVR的网络安全态势预测模型,为了得到更精确的预测数据,采用粒子群优化算法(PSO)对新构造的混合核函数SVR的参数进行优化。

2 网络安全态势预测架构

针对网络系统具有复杂的性质,其产生的未来态势值有不确定性和非线性等特点。网络安全态势预测是利用已知态势值之间相互隐藏的关系来预测未来时刻态势值,本质是将原始目标问题(一段时间的网络安全态势通过某种方法量化为态势值进行预测)抽象成一个连续时间序列上的回归预测问题^[9]。在一般的网络安全态势预测中,都是基于某种构架对网络安全进行有效预,本文采用的构架主要由以下四个模块组成:(1)态势要素获取模块;(2)态势值量化及预处理模块;(3)态势预测模块;(4)决策策略模块。系统结构图如图1所示。

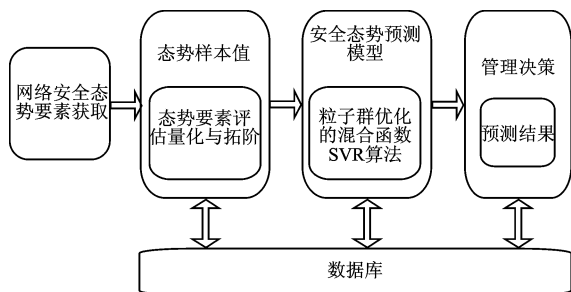


图1 网络安全态势预测架构

(1) 态势要素获取模块

态势要素可通过基于系统运行信息与系统配置信息两大类进行获取。系统运行信息可以通过IDS、Firewall、VDS等提供的多源安全信息得到;系统配置可以通过扫描漏洞技术等来得到系统的服务、网络以及主机等漏洞信息。

(2) 态势值计算及预处理模块

将获取的态势要素信息根据态势评估量化方法得到量化评估后的态势值,其值的大小取决于态势要素的攻击的次数、危险程度、发生的频率等因素^[1]。为了使网络安全预测模型能够高速有效的对态势数据进行预测,将态势数据进行极值标准化处理。

(3) 态势预测模块

态势预测模块主要由粒子群优化算法与基于混合核函数SVR组成。首先选取一组数据预处理好的训练样本与测试样本,将训练样本输入到粒子群算法与混合核函数SVR相结合的预测模型中进行训练,得到初始的预测模型。将得到预测值与真实值之间的误差作为适应度函数对粒子群算法的终止条件进行判断。通过不断迭代寻找到适应值最小的回归函数,并得到相对应的最优参数。根据得到的参数建立最终预测模型。

(4) 决策策略模块

根据态势预测模型得到的结果,通过网络管理员的经验,对得到的态势结果进行网络安全决策。

3 网络安全预测模型

3.1 数据预处理与拓阶

由于网络安全态势值的跨度大,为了避免其对改进的SVR预测模型产生负面影响,同时为了加快算法的收敛速度,因此对所得的态势值做极值标准化处理。对于态势值数据 $X = (x_1, x_2, \dots, x_n), x \in R$ 。由极值标准化公式:

$$x'_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad i = 1, 2, \dots, n \quad (1)$$

式中, x_i 与 x'_i 分别为处理前后的网络安全态势值, $x_{\min}$ 与 $x_{\max}$ 分别代表处理前所有网络安全态势值的最小值与最大值,通过处理后得到的网络安全态势数据 $X' = (x'_1, x'_2, \dots, x'_n), x' \in R$ 是一组一维时间序列。为了精确地进行预测,在输入预测模型之前需要对 X' 进行拓阶和时间延迟,让它变换成一组多维网络安全数据。本文设定时间延迟 τ 为1,维数按2, 3, ..., 等逐步进行拓阶试凑。最后确定拓阶维数为 m 。进行拓阶后转换的成多维数据如表1所示。

表1 重构后多维态势数据

学习样本输入	输出期望值
$x'_1, x'_2, \dots, x'_{m-1}$	x'_m
$x'_2, x'_3, \dots, x'_m$	x'_{m+1}
...	...
$x'_{n-m}, x'_{n-m+1}, \dots, x'_{n-1}$	x'_n

3.2 混合核函数支持向量机设计

支持向量回归算法具有收敛速度快,不易陷入局部寻优等特点,针对网络安全态势数据具有高度复杂性的非线性等特点,其优势明显.

支持向量机回归算法设计的核心是核函数及其相关参数的选择.其中核函数的引入使得函数的求解绕开了特征空间,直接在原空间计算,从而避免了非线性映射的计算.核函数实现了非线性算法的同时算法的复杂度不会增加,说明了核函数是研究支持向量机回归算法的关键.

设重构后的网络安全态势样本数据为 $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, 其中 x_i 为输入样本, y_i 为输出期望样本值.支持向量回归算法原理是通过一个非线性映射 φ 函数,将非线性输入样本 x_i 映射到高维特征空间 H ,在该高维空间构造一个线性回归公式:

$$f(x) = \omega^T \varphi(x) + b \quad (2)$$

$$\varphi: R^n \rightarrow G, \omega \in G$$

式中, ω 表示权值向量, b 表示偏置值.

因此,为了寻找最小 ω 支持向量,回归预测问题转化为式(3)、式(4):

$$\min_{\omega, b, \xi_i, \xi_i^*} = \frac{1}{2} \|\omega\|^2 + c \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (3)$$

约束条件为

$$\begin{cases} y_i - \omega \cdot x_i - b \leq \varepsilon + \xi_i \\ \omega \cdot x_i + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i \geq 0, \xi_i^* \geq 0, c \geq 0 \end{cases} \quad i = 1, \dots, n \quad (4)$$

求解以上公式,需要利用对偶定理.在引入拉格朗日乘子 a_i 和 a_i^* 后,将得到拉格朗日函数将其转化成一个二次规划问题.

$$L(\omega, \xi, \xi^*) = \frac{1}{2} \|\omega\|^2 + c \sum_{i=1}^n (\xi_i + \xi_i^*) - \sum_{i=1}^n a_i (\varepsilon + \xi_i + y_i - \langle \omega, x_i \rangle - b) - \sum_{i=1}^n a_i^* (\varepsilon + \xi_i^* + y_i - \langle \omega, x_i \rangle - b) - \sum_{i=1}^n (\eta_i \xi_i + \eta_i^* \xi_i^*) \quad (5)$$

根据 KKT 条件等,经过一系列推导得到回归预测函数表达式为

$$f(x) = \sum_{i=1}^n (a_i - a_i^*) (\varphi(x_i), \varphi(x)) + b \quad (6)$$

在支持向量回归算法中通过引入核函数来简化线性逼近,核函数 $k(x, x_i)$ 满足 $k(x, x_i) = \langle \varphi(x_i), \varphi(x) \rangle$. 这样式(6) 变为

$$f(x) = \sum_{i=1}^n (a_i - a_i^*) k(x, x_i) + b \quad (7)$$

只有满足 mercer 条件的函数才能成为核函数,核函数由于具有将线性方法推广到非线性方法的强有力的能力,对工程应用具有极大意义.目前主要研究的核函数有以下三种.

多项式核函数:

$$k(x, x_i) = [(x \cdot x_i) + 1]^d;$$

径向基核函数:

$$k(x, x_i) = \exp(-g \sum_{i=1}^n (x - x_i)^2);$$

Sigmoid 函数:

$$k(x, x_i) = \tanh(k \cdot x \cdot x_i - d).$$

目前支持向量机的核函数主要分为两类:全局核函数与局部核函数.在对应的插值理论中全局核函数具有插值能力较弱但外推能力较强(对未知样本的预测能力)的特性;局部核函数具有较好的插值能力但外推能力较弱的特性^[6].其中常用的局部核函数为径向基核函数,全局核函数为多项式核函数.

为了建立其一个同时具有较好的差值能力与外推能力的网络安全预测模型,提出了一种综合两类核函数优点的混合核函数.在 mercer 条件下定义一个权重系数 λ ,将其引入到多项式核函数与径向基核函数之中,组成一种新的核函数,其公式如下:

$$k(x, x_i) = \lambda (x \cdot x_i + 1)^q + (1 - \lambda) \exp(-g \|x - x_i\|^2) \quad (8)$$

式中, q 为多项式系数; g 为径向基系数; λ 为混合权重系数, $\lambda \in [0, 1]$.

3.3 混合核 SVR 参数的选择与优化

在基于支持向量的预测问题中,参数的选择与优化直接影响预测的精度,所以参数的选择与优化成为预测模型的关键.在对混合核的 SVR 的参数分析可知,需要确定的参数:混合核函数权重系数 λ , 多项式核函数参数 q , 径向基核函数参数 g , 惩罚参数 c 以及不敏感系数 ε ^[11]. 为确保良好的外推能力,取敏感系数 $\varepsilon = 0.01$; 多项式参数 $q = 2$; 另外的 (c, λ, g) 三个参数将采用粒子群算法 (PSO) 进行优化获得.

PSO 算法是一种基于群体理论的算法,通过群体中粒子间的合作与竞争产生的群体智能指导优化搜索,建立起“速度—位移”模型来进行全局优化.粒子群算法可描述为:设粒子群在一个 D 维空间中搜索,由 X 个粒子组成种群 $X = \{X_1, X_2, \dots, X_N\}$,

其中每个粒子所处的位置 $X_i = \{x_{i1}, x_{i2}, \dots, x_{in}\}$ 都表示问题的一个解. 粒子通过不断调整自己的位置 x_{id} 来搜索新解. 每个粒子都能记住自己搜索到的最优解, 记做 p_{id} , 以及整个粒子群经历过的最好位置, 即目前搜索到的最优解, 记做 p_{gd} . 此外每个粒子都有一个速度, 记做 $V_i = \{v_{i1}, v_{i2}, \dots, v_{in}\}$, 当两个最优解都找到后, 每个粒子根据式(9)、式(10)来更新自己的速度, 搜索到最优解.

$$v_{id}(t+1) = \omega v_{id}(t) + \eta_1 \text{rand}() (p_{id} - x_{id}(t)) +$$

...

$$\eta_2 \text{rand}() (p_{gd} - x_{id}(t)) \quad (9)$$

$$x_{id}(t+1) = x_{id}(t) + v_{id}(t+1) \quad (10)$$

式中, $v_{id}(t+1)$ 表示第 i 个粒子在 $t+1$ 迭代中第 d 维上的速度, ω 为惯性权重, η_1, η_2 为学习因子, $\text{rand}()$ 为 $0 \sim 1$ 之间的随机数. 为了粒子速度不至于过大, 可设置速度界限 $v_{\max}$ 和 $v_{\min}$ 来限制粒子速度. 通过上述可知权重系数 ω, η_1, η_2 是 PSO 算法的重要参数.

3.4 改进支持向量机的网络安全态势预测流程

构建基于粒子群优化的混合核函数支持向量机网络预测安全态势预测模型的步骤如下所示.

通过系统运行信息与系统配置信息对网络安全态势要素进行收集, 并对收集的异常与不良日志信息进行处理, 得到有效的网络安全态势要素.

将得到的态势要素通过层次化网络安全态势评估量化方法^[1]对收集到的要素信息进行评估量化处理, 生成原始一维网络安全态势值序列. 利用式(1)对所得到的态势值序列进行极值化处理.

对得到的极值化后的态势值序列, 通过拓阶和时间延迟等方法将一维的态势值序列转变成多维的态势值序列样本. 拓阶和时间延迟的赋值分别通过试凑法与经验得到.

设定 PSO 算法中种群大小 m , 初始化惯性权重系数 ω , 学习因子 η_1 与 η_2 , 最大迭代次数 $\text{iter}_{\max}$, 交叉验证次数 K 和粒子群粒子的位置与速度的初始值.

从完成极值化与拓阶操作的态势样本中选择出训练样本与测试样本, 将训练样本输入改进 SVR 模型进行学习. 采用 PSO 算法并结合 k -fold 交叉验证最小误差准则对改进 SVR 预测模型参数进行寻优, 将得到最佳参数建立起最优的预测模型.

将测试样本输入最优预测模型得到预测结果. 将结果通过反极值化公式:

$$x_i = x' \times (x_{\max} - x_{\min}) - x_{\min} \quad (11)$$

得到实际的网络安全态势值, 并结合历史网络安全

态势数据与预测结果, 网络安全管理者对未来的网络安全问题进行主动决策.

综上所述, 基于 PSO 优化的混合核函数 SVR 网络安全态势预测模型流程图如图 2 所示.

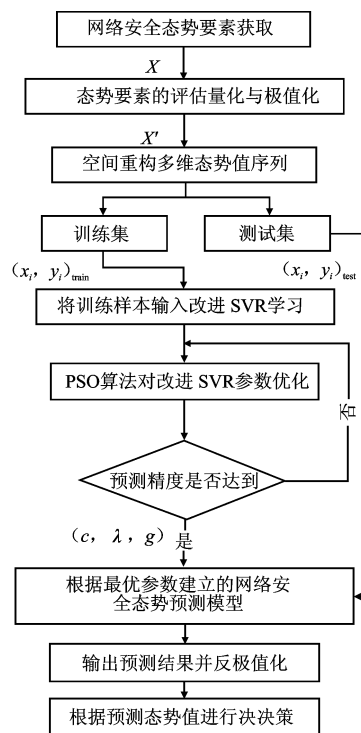


图2 网络安全态势预测模型流程图

4 实验分析

4.1 网络安全态势数据

为了验证上述预测模型的可靠性与合理性, 采用了某公司 10 月 1 日到 12 月 1 日一共 2 个月的网络安全防护日志信息作为原始数据源. 对每天的日志信息进行 5 次采样, 本文将采样得到的日志信息进行网络安全评估量化, 如图 3 所示. 最后得到 300 个态势值组成的一维时间序列.

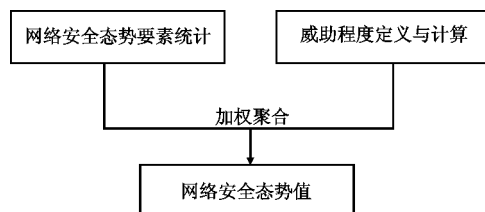


图3 网络安全态势要素评估量化

4.2 网络安全态势预测模型的实现

首先将得到的这组一维时间序列数据进行极值化处理, 然后根据逐步试凑得到数据的最佳拓阶维数, 最终确定最佳拓阶维数 m 为 5. 按照拓阶维数 m

与时间延迟 τ 的值对极值化后的一维时间序列进行重构,将重构生成多维样本分成训练样本与测试样本.其中取前 280 个样本作为训练样本(前 56 天),后 20 个样本作为测试样本(后 4 天).再将训练样本输入预测模型中进行学习.粒子群的参数设置如表 2 所示.

表 2 粒子群参数设置表

粒子群参数设置	预设值
学习因子 η_1, η_2	1.5, 1.7
最大迭代数量	200
种群最大数量	20
惯性权重系数 ω	1.0
C, λ 和 g 的范围	(0.1, 100), (0, 1), (0.01, 1000)
交叉验证次数 K	5

4.3 态势预测实验与对比分析

本文所有的实验均在 windows xp 操作系统下的 matlab 7.8 平台下实现.首先将训练样本输入到基于粒子群寻优的混合核函数支持向量机中对模型中的参数组 (c, λ, g) 进行寻优,进行不断迭代并结合 k -fold 交叉验证最小误差准则得到最优解: $c = 35.223, \lambda = 0.873, g = 9.126$.进而将得到的参数建立最佳的网络安全态势预测模型,用该模型对测试数据进行实验.为了验证本网络安全态势预测模型的预测精度,分别采用 SVR(默认核函数为径向基核函数)、参数未进行优化混合核 SVR 两种方法与本文方法进行相同的实验,其参数如表 3 所示.

表 3 预测算法相关参数

预测算法	$c, g, \epsilon / c, g, \epsilon, \lambda, g$
RBF_SVR	10, 1, 0.01
混合核函数 SVR	0.394, 12.45, 0.01, 0.5, 2
混合核函数 PSO_SVR	35.223, 9.126, 0.01, 0.873, 2

得到的预测结果与真实值进行对比.其对比结果如图 4 所示.

由图 4 可以看出:尽管三个方法都能够取得一定的预测效果,但传统的 SVR 预测模型比基于混合核的 SVR 效果要差.由于混合核的 SVR 参数较多,在无法有效地取得最优参数前提下不能得到最好的预测效果.在此情况上通过运用粒子群算法寻优之后得到的本文预测方法相对于其他两种方法,更加符合真实的网络安全态势变化趋势.因为本文的预测模型相对于传统的 RBF 核函数 SVR 预测模型,

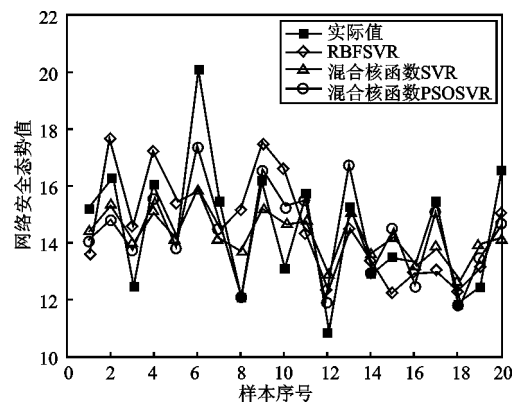


图 4 态势预测结果对比

其外推能力更强(对未知样本的预测能力更好).同时引入粒子算法进行寻优得到最优参数,能够有效提高预测精度.

下面采用定量分析方法对三种预测模型进行比较,首先采取绝对误差指标对三种模型的所有态势预测值分别进行对比,结果如图 5 所示.

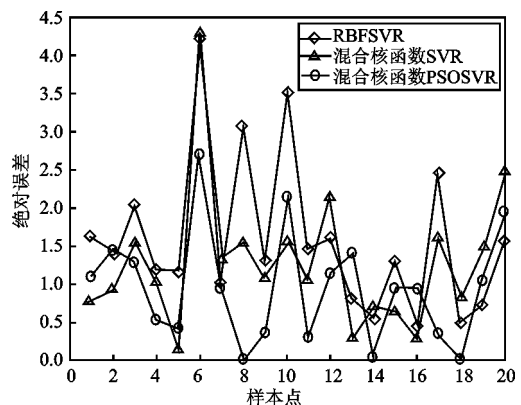


图 5 态势预测误差绝对值对比

由图 5 可以看出:本文方法除了在少部分点外的绝对误差大于其他两种方法,绝大部分的预测误差控制在 2 个态势值以内,预测精度明显好于其他两种方法.

接下来通过平均相对误差(MAPE)与均方根误差(RMSE)两项误差指标来定量模型的精准度.三种模型对比见表 4.其中 MAPE 与 RMSE 值越小表示某种方法精确度越高,MAPE 与 RMSE 如式(12)、(13)所示.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i' - y_i|}{y_i} \quad (12)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i' - y_i)^2} \quad (13)$$

式中, y_i, y_i' 分别表示网络安全态势真实值与预测值.

由表4可以看出:相对与其他两种预测模型,本文方法得到的预测模型的RMSE与MAPE值明显降低.表明本模型预测性能高于其他两种模型.

表4 三种方法的预测误差指标对比

预测方法	RMSE	MAPE
RBF_SVR	1.869 3	0.109 7
混合核函数 SVR	1.560 9	0.087 2
混合核函数 PSO_SVR	1.190 5	0.064 3

在复杂度上,由于支持向量机核函数的计算复杂度与样本个数相关而与参数的变化无关,其时间复杂度一般为 $O(n^3)$ [7].本文构造的混合核函数是符合mercer条件下引用权重系数相加得到新的核函数.相比于传统的径向基核函数基础上只是单纯地增加了核函数的计算量,在时间复杂度没有产生阶数增加的情况下能够有效地增加SVR模型对网络安全态势预测精度.

5 结束语

网络安全态势预测一直是网络安全态势感知的关键问题,它能够有效地预防大规模网络安全事件的发生.本文从影响SVR的性能的两个方向入手,提出了基于PSO的混合核SVR网络安全态势预测模型.该方法在结合了SVR中局部核函数与全局核函数的优点,同时引入粒子群算法对SVR参数进行优化,得到最佳参数组合.预测结果表明该模型能够有效地提高预测精度,反映出关于未来安全的可靠的发展趋势,有助于网络管理者对网络安全事件进行决策.

参考文献:

[1] Jajodia S, Liu P, Swarup V, et al. Cyber situational

awareness[M]. Berlin:Springer, 2010.

- [2] Yuan Quan Shi, Tao Li, Wen Chen. Network security situation prediction using artificial immune system and phase space reconstruction[C] // Applied Mechanics and Materials. Switzerland: trans tech Publications LTD, 2010:3662-3666.
- [3] 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[J]. 计算机科学, 2012, 39(6):61-63+76.
- [4] 王晋东,沈柳青,王坤,等.网络安全态势预测及其在智能防护中的应用[J]. 计算机应用, 2010, 30(6):1480-1482.
- [5] 陈虹,王飞,肖振久.基于PSO_SVR的网络安全态势预测方法[J]. 计算机应用与软件, 2014, 35(8):292-294.
- [6] Elattar E E, Goulermas J, Wu Q H. Electric load forecasting based on locally weighted support vector regression[J]. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, 2010, 40(4):438-447.
- [7] 孙俊,王艳,金夏明,等.基于MSCPSO混合核SVM参数优化的生菜品质检测[J]. 农业机械学报, 2013, 44(9):209-213.

作者简介:

李方伟 男,(1960-),博士,教授.研究方向为移动通信技术与理论、信息安全技术.

罗嘉 男,(1990-),硕士研究生.研究方向为移动通信理论、网络安全态势感知技术. E-mail:125832324@qq.com;

朱江 男,(1977-),博士,副教授.研究方向为认知无线电技术.

张海波 男,(1979-),博士,讲师.研究方向为无线通信资源管理.

(上接第109页)

- [4] Zhang X. Separable reversible data hiding in encrypted image [J]. IEEE Transactions on Information Forensics and Security, 2012, 7(2):826-832.
- [5] Zhang W, Ma K, Yu N. Reversibility improved data hiding in encrypted images [J]. Signal Processing, 2014, 94(11):118-127.
- [6] Zhang X, Qian Z, Feng G, et al. Efficient reversible data hiding in encrypted images [J]. Journal of Visual

Communication and Image Representation, 2014, 25(2):322-328.

作者简介:

郑淑丽 女,(1974-),博士,副教授.研究方向为多媒体内容安全、无线传感网络等.

李丹丹 女,(1989-),博士.研究方向为图像密文域可逆信息隐藏. E-mail:lidandan110506@126.com.