基于优化支持向量回归的工业互联网安全态势预测方法

胡向东[1],吕高飞[1],白 银[2]
(1. 重庆邮电大学自动化学院,重庆 400065; 2. 重庆邮电大学先进制造工程学院,重庆 400065)

摘 要: 作为支撑智能制造等的新型工业基础设施,工业互联网的安全态势预测是一个关键性需求和应用新挑战. 本文提出一种基于优化支持向量回归的工业互联网安全态势预测方法,即利用差分进化算法和自适应参数调整策略克服灰狼优化算法计算速度慢、优化精度低的缺点;再利用改进的灰狼优化算法优化支持向量回归参数;最后,利用最优化参数组合建立支持向量回归预测模型,实现工业互联网环境下的安全态势预测. 仿真实验结果表明,在容许偏差为0.05或0.1时,本文方法的预测准确率分别为90%和100%,预测结果的绝对误差均小于0.07,相比于对比方法有更高的预测准确率和预测精度.

关键词: 工业互联网;安全态势预测;支持向量回归;灰狼优化算法;差分进化算法

基金项目: 教育部-中国移动科研基金(No.MCM20150202, No.MCM20180404);重庆市高校创新研究群体(No.CXQT20016)

中图分类号: TP391.9;TN918.91 **文献标识码:** A **文章编号:** 0372-2112(2023)02-0446-09

电子学报 URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20210558

A Method of Security Situation Prediction for Industrial Internet Based on Optimized Support Vector Regression

HU Xiang-dong[1], LÜ Gao-fei[1], BAI Yin[2]
(1. *School of Automation, Chongqing University of Posts and Telecommunications, Chongqing* 400065, *China*;
2. *School of Advanced Manufacturing Engineering, Chongqing University of Posts and Telecommunications, Chongqing* 400065, *China*)

Abstract: The Industrial Internet is an emerging modern infrastructure for supporting smart manufacturing. Accurate security situation prediction of industrial Internet is nowadays still a key demand and challenge for industrial application. To this aim, a novel method of security situation prediction for industrial Internet based on optimized support vector regression is proposed in this paper. The proposed method is a three-step procedure: in the first step, an improved gray wolf optimizer algorithm, based on differential evolution and adaptive parameter adjustment strategy, with high calculation speed and optimization accuracy is proposed; then, the optimized parameters of support vector regression are obtained; after that, accurate security situation prediction model for industrial Internet is established. The simulation results show that the prediction accuracy rate of the proposed method are 90% and 100% when the allowable deviations are 0.05 or 0.1, respectively, and the corresponding absolute errors are less than 0.07, and thus the proposed method can enhance the accuracy rate and precision of prediction, in contrast to the existing methods.

Key words: industrial Internet; security situation prediction; support vector regression; grey wolf optimizer; differential evolution algorithm

Foundation Item(s): Joint Research Foundation of the Ministry of Education of the People's Republic of China and China Mobile (No. MCM20150202, No. MCM20180404); Chongqing University Innovation Research Group (No. CXQT20016)

1 引言

随着新一代信息技术、互联网与工业系统的深度融合,支持智能决策与智能制造的工业互联网可极大提升工业生产和管理效能,作为新型工业基础设施,正

收稿日期:2021-04-30;修回日期:2022-01-04;责任编辑:孙瑶

在引领新一轮工业变革,成为如制造业、医疗、能源等传统工业创新发展的主流方向[1]. 工业互联网以新技术助力工业生产降低成本和减小资源消耗的同时,仍存在诸多挑战,尤其是面临的网络信息安全性风险. 与传统互联网安全相比,工业互联网安全有三个特征:一是防护对象扩大,安全场景更丰富;二是连接范围更广,威胁延伸至物理世界;三是网络安全和生产安全交织,安全事件危害更严重. 基于工业互联网的应用,工业企业的数据价值得到不断提升,但工业通信协议本身可能存在漏洞,安全防御机制并不足够健全. 如果这些安全漏洞被攻击者利用,企业在转型升级过程中会面对层出不穷的网络威胁,给企业造成巨大损失[2,3]. 例如,在2018年8月,台积电的半导体工厂突然遭受病毒攻击,旗下多个生产线被迫停产,造成了巨大损失. 作为新型信息基础设施,工业互联网正在快速发展,随着5G和NB-IOT等技术与工业互联网的融合应用,网络攻击门槛不断降低,针对工业互联网的攻击日益增加,手段更加繁杂,各国政府相继加强对网络安全的研究与管理[4].

工业互联网具有可靠性要求高、实时性要求严、数据量大等特点,建立工业互联网安全防护体系是工业企业智能化发展的重要基石. 以传统入侵检测系统为代表的网络安全防护措施依赖基于误用的方法,难以应对工业互联网当前面临的各种信息安全风险. 网络态势感知技术(Cyberspace Situation Awareness, CSA)[5]能够从复杂多变的工业互联网环境中提取安全状态信息,通过分析历史和当前安全事件,进一步对未来安全发展趋势进行预测,从而帮助安全决策者制定安全防御决策. Bass[6]在20世纪末将态势感知应用于网络安全,给出了一种安全态势感知框架,这为网络态势感知研究提供了重要基础.

态势评估和态势预测是CSA体系的两大关键技术. 其中,态势评估能够实时地评价网络安全风险,进一步定量计算工业互联网系统整体安全态势和遭受的某种攻击. 王益丰等人[7]利用抗体浓度构建模型计算网络安全风险,能够实时进行安全态势检测和评估. 态势预测是指在安全测试规则下,凭依历史和当前安全数据,预测下一个相同时间段的安全状况. 基于工业互联网本身的特点,其安全预测面临的对象和场景更复杂、动态性更强、资源等制约因素更多、达成目标的难度更高,导致其挑战度巨大. 工业互联网安全态势预测技术通过采集数据信息、设备资产信息和第三方外部数据,综合分析工业互联网的网络行为和未来发展趋势. 与传统互联网安全相比,该技术更加注重控制安全和智能设备安全,并在出现安全威胁时更加注重利用工业互联网中各类设备之间的协同机制进行联合抵制. 该技术有助于主动了解、分析和掌控工业互联网安全风险变化情况,是一种从宏观层面为企业安全管理人员提供决策依据,辅助制定防范策略的方法.

自态势预测技术问世以来,大量科研人员提出了不同的安全态势预测模型,机器学习也为解决此类问题提供新方法. 李方伟等人[8]利用自适应聚类算法优化径向基(Radial Basis Function, RBF)神经网络的参数,增强了算法的泛化能力;但该模型的应用条件单一,不适用于复杂的网络系统. Hu等人[9]利用约束协方差矩阵自适应进化策略优化模型参数,提出一种基于云信念规则库的态势预测方法;该方法考虑安全状态的模糊性和随机性,提高了预测精度. Zhang等人[10]基于安全事件的突发性和不确定性,利用小生境遗传算法改进小波神经网络参数,建立的预测模型在非线性预测中具有一定优势,但其泛化能力有限. Kou等人[11]通过深入分析网络攻击意图与网络配置之间的关联,利用攻击图对安全事件进行因果分析,并结合漏洞和网络连通性,有效预测下一阶段的攻击;但是该方法依赖已有的攻击,缺乏针对新型攻击方式的通用性.

本文在研究支持向量回归(Support Vector Regression, SVR)算法的基础上,利用灰狼优化算法(Grey Wolf Optimizer, GWO)选取SVR的相关参数;但GWO算法求解速率慢、易出现局部最优,故本文利用自适应参数调整策略和差分进化(Differential Evolution, DE)算法进行改进;然后利用改进的GWO算法优化SVR参数,提出一种优化SVR算法的工业互联网安全态势预测方法,该方法可取得相较于对比方法更好的预测性能.

2　支持向量回归预测

设数据集 $X=(x_i,y_i)$; $x_i,y_i\in\mathbf{R}_n$; $i=1,2,\cdots,N$,其中 x_i 为输入样本,y_i 为 x_i 对应的期待样本输出(预测值),N 为样本总数. 样本在新特征空间中的待拟合预测函数为

$$f(x)=\omega\cdot\varphi(x)+b \tag{1}$$

其中,ω 表示权重;b 为偏置量.

为了确定 ω 和 b,引入松弛变量 ξ_i,ξ_i^* 和惩罚因子 C,将求解回归问题转换为求最小值[12],即

$$\begin{aligned}&\min_{\omega,b,\xi,\xi^*}\ \frac{1}{2}\|\omega\|^2+C\sum_{i=1}^{N}(\xi_i+\xi_i^*)\\&\text{s.t.}\begin{cases}y_i-(\omega\cdot\varphi(x)+b)\leqslant\varepsilon+\xi_i\\(\omega\cdot\varphi(x)+b)-y_i\leqslant\varepsilon+\xi_i^*\end{cases}\end{aligned} \tag{2}$$

其中,$C>0,\xi_i\geqslant0,\xi_i^*\geqslant0,i=1,2,\cdots,N$;$\varepsilon$ 为不敏感损失参数,用于确定拟合精度的大小.

构造Lagrange函数:

$$\begin{aligned}&L(\omega,b,\xi_i,\xi_i^*,\alpha_i,\alpha_i^*,\mu_i,\mu_i^*)\\&=\frac{1}{2}\|\omega\|^2+C\sum_{i=1}^{N}(\xi_i+\xi_i^*)\\&\quad-\sum_{i=1}^{N}(\mu_i\xi_i+\mu_i^*\xi_i^*)+\sum_{i=1}^{N}\alpha_i(\varepsilon+\xi_i-y_i(\omega\cdot x_i+b))\\&\quad+\sum_{i=1}^{N}\alpha_i^*(\varepsilon+\xi_i^*-y_i(\omega\cdot x_i+b))\end{aligned}\tag{3}$$

其中，α_i,α_i^*为非负 Lagrange 乘子；$\mu_i\geqslant 0,\mu_i^*\geqslant 0$.

针对 $L(\omega,b,\xi,\xi^*,\alpha_i,\alpha_i^*,\mu_i,\mu_i^*)$，在对 ω,b,ξ,ξ_i^* 求极小值的条件下，再求对 α_i,α_i^* 的极大值，最后根据最优化条件得到回归预测函数为

$$f(x)=\sum_{i=1}^{N}(\alpha_i-\alpha_i^*)K(x_i,x)+b\tag{4}$$

其中，$K=(x_i,x)$为核函数；N为样本数量.

核函数对 SVR 的作用体现在两个方面：一是利用一个非线性变换 $\varphi(x)$将原输入样本映射到新的 Z 空间上，从而使原空间内的超曲面变化为 Z 空间内的超平面；二是在 Z 中用解决线性拟合问题的方法去训练输入样本，得到一个线性拟合模型. 本文选取 RBF 核函数如下：

$$K(x_i,x)=\exp(-g\|x_i-x\|^2)\tag{5}$$

其中，g 是自由参数.

使用 SVR 进行回归预测，其性能主要依托于参数 C 和 g 的选取. 通常可使用专家经验选择 2 个参数的大小，有一定的主观性和盲目性，无法保证预测模型的精度，本文选用改进的灰狼优化算法优化 SVR 算法参数.

3 工业互联网安全态势预测方法

3.1 灰狼优化算法原理

灰狼优化算法[13]参数较少且容易满足搜索条件，在全局寻优方面优势明显，现已运用于物联网数据传输优化、作业车间调度优化等多个领域[14]. GWO 算法的基本原理如图 1 所示.

(1)社会等级分层行为：将种群大小为 $N=\{x_1,x_2,\cdots,x_n\}$的灰狼种群，依据个体适应度值大小，分为 α,β,δ 这 3 种等级.

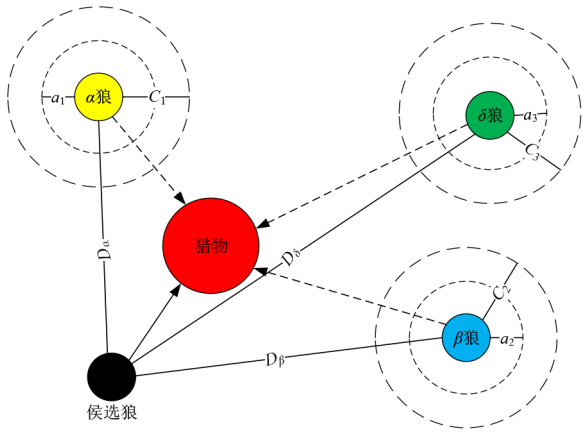

图 1 灰狼捕食原理图

(2)包围行为：捕猎前，灰狼们先分散，再以猎物为中心，逐渐形成包围网，即

$$D(t)=\left|c\circ x_p(t)-x(t)\right|\tag{6}$$

$$x(t+1)=x_p(t)-A_i\circ D(t)\tag{7}$$

$$c=2r_1\tag{8}$$

$$A_i=a_i(2r_2-1)\tag{9}$$

其中，$D(t)$表示搜索个体与目标的相对位置信息；$\circ$表示 Hadamard 乘积操作；$x_p(t)$为猎物的位置向量信息；$x(t)$是 α 狼的方位向量信息；t 表示当前迭代次数；参数 a_i 的值一般从 2 线性减为 0；A_i 和 c 为协同因子；$r_1,r_2\in\text{rand}(0,1)$.

针对参数 A_i，当 $|A|\geqslant 1$，算法进行全局性搜索；当 $|A|<1$，算法进行局部寻优.

(3)捕猎行为：狩猎行为主要依靠灰狼对自身和目标相对距离的辨认能力，并根据 α 狼的位置信息更新自身位置，即

$$D_k=\left|c_i\circ x_k-x_k(t)\right|\tag{10}$$

$$x_i=x_k-A_i\circ D_k\tag{11}$$

$$x(t+1)=\frac{1}{3}\sum_{i=1}^{3}x_i\tag{12}$$

其中，k 依次取 α,β,δ，代表前 3 类优势灰狼；x_k 表示第 k 种灰狼的位置；D_k 表示候选 α 狼与其他 3 种优势狼的直线距离；$x(t+1)$为更新后候选优势灰狼个体的位置.

(4)围攻行为：包围网形成后，同时袭击猎物.

由式(8)可知，参数 c 具有随机性，可在一定程度上防止算法陷入局部最优. 但是 GWO 算法在整个寻优过程中存在 3 个潜在最优解(α,β,δ). 如果当前最优解 α 在某一环节为局部最优，在包围后期所有捕食狼将会接近局部最优个体所在位置区域，使算法出现过早收敛而导致寻优过程提前结束.

3.2 灰狼优化算法的改进

为避免 GWO 算法陷入局部最优，需要优化选取最优解(即 α 狼).

3.2.1 变异进化策略

DE 算法[15]是一种模拟生物进化的算法，计算能力强，可用于实时参数及实值可测函数的优化. 该算法保留了基于种群的全局搜索和并行分布的特点，不易陷入局部最优解，能够不依托样本特征信息，具有较强的收敛能力，并且可利用自身具备的记忆功能动态地跟踪整个过程，尤其在复杂环境下对参数的依赖性弱.

DE 算法的实现过程包括差分变异、重组交叉、优势选择[16]. 首先，随机产生一个种群，再随机挑选两个个体向量，计算得到二者的矢量差值；其次，将其缩放加权后，按照某种方法和待变异体求和，得到中间突变

体;然后,与初始化的新个体进行交叉混合;最后,比较更新最优个体. 通过不断地迭代,保留到最后的种群个体就是最优解,从而实现参数最优化.

在搜索空间中,建立初始规模满足$N>4$的种群,且种群内个体x_i的维数为D,一般维数区间为$[5D,10D]$. 依据式(13)对个体随机初始化,有

$$x_i^k=x_l^k+(x_h^k-x_l^k)\text{rand}(0,1) \tag{13}$$

其中,i表示种群内个体;k表示当前维数,且$k=1,2,\cdots,D$;x_h^k和x_l^k分别表示个体初始化时,第k维的上、下限;rand(0, 1)为一个在(0, 1)上的随机数.

(1)差分变异:种群初始化后,在当前迭代过程中,通过第i个子代个体和随机选取的父代个体之间重组产生中间变异体,如图2所示.

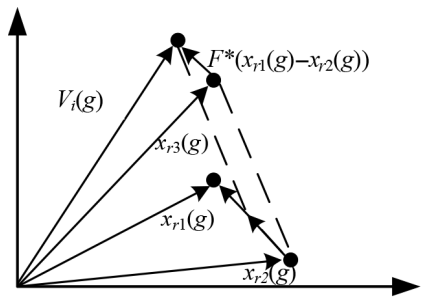

图2　标准差分变异过程二维示意图

具体产生中间体的过程可用式(14)表示:

$$V_i(g+1)=x_{r3}(g)+\lambda(x_{r1}(g)-x_{r2}(g)) \tag{14}$$

其中,$x_{r1}(g)$, $x_{r2}(g)$, $x_{r3}(g)$为随机选择的3个搜索个体且$r1\neq r2\neq r3\neq i$;$V_i(g+1)$为变异搜索个体;λ为缩放参数;g为当前迭代数值.

(2)重组交叉:将得到的变异个体与原个体比较,只有换了部分元素才能得到重组交叉后的实验个体$U_i(g+1)$. 交叉操作过程如图3所示. 其中,$x_i(g)$表示第g代种群个体;$V_i^k(g+1)$表示第g+1代中间变异搜索个体$V_i(g+1)$的第k位可进行变异基因;$U_i^k(g+1)$表示重组交叉后得到实验个体$U_i(g+1)$的第k位基因.

具体进行重组交叉的过程可表示为

$$U_i^k(g+1)=\begin{cases}V_i^k(\text{g}+1), & \text{rand}(0,1)\leqslant \text{CR}或k=\text{rand}(i)\\ x_i^k(g), & 其他\end{cases} \tag{15}$$

其中,CR$\in$[0.1, 1.0],表示交叉因子;k表示维度;rand(i)从$\{1,2,\cdots,n\}$中随机选取.

(3)优势选择:采用贪婪算法,当新产生的中间体比种群内当代优势体优秀时将其替换,保证下一代种群适应性更好,有

$$x_i(g+1)=\begin{cases}U_i(g), & f(U_i(g))\leqslant f(x_i(g))\\ x_i(g), & f(U_i(g))>f(x_i(g))\end{cases} \tag{16}$$

其中,$x_i(g+1)$是选择出最佳的新个体;$f(\cdot)$表示适应性

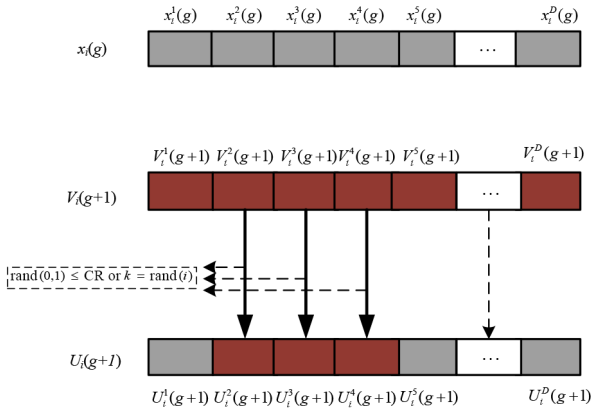

图3　重组交叉示意图

评价函数,只有当$f(U_i(g))\leqslant f(x_i(g))$时才表明当前迭代进化成功.

3.2.2　自适应参数调整策略

为了避免DE算法收敛过快,利用式(17)生成自适应缩放因子λ. 在算法后期,λ值的大小将趋于初始值λ_0,以保证得到全局最优解,即

$$\lambda=\lambda_0 2^{\exp\left(1-\frac{G}{G+1-g}\right)} \tag{17}$$

其中,$\lambda\in(\lambda_0,2\lambda_0)$且$\lambda\in(0,1)$. G为最大迭代次数,g为当前迭代次数.

为了使当前种群内的每个搜索个体能够最大限度地发生进化,需要对每个搜索个体进行监控. 在T次迭代中,当某个个体三代内未产生更好的后代或未发生过突变时,利用式(18)更新得到新的交叉因子CR′,即

$$\text{CR}'=\begin{cases}\text{CR}+\text{CR}\cdot\text{rand}(0,1), & \text{CR}\leqslant 0.5\\ \text{CR}-\text{CR}\cdot\text{rand}(0,1), & \text{CR}>0.5\end{cases} \tag{18}$$

此外,为了平衡GWO算法整体和局部的优化能力,利用式(19)更新计算收敛因子,即

$$a_i=\frac{2f_i}{\sum_{i=1}^{3}f_i} \tag{19}$$

其中,$a_i(i=1,2,3)$分别表示更新α,β,δ这3种个体信息时的收敛因子;$f_i(i=1,2,3)$分别表示α,β,δ这3种个体的适应度值.

改进后的GWO算法描述如算法1所示.

3.3　工业互联网安全态势预测模型

工业互联网安全环境结构复杂,工业网络安全数据种类繁多、流向复杂,同时存在安全漏洞等安全问题,对工业互联网安全数据的处理要求更高[17]. 工业互联网安全包括智能设备、网络设备、应用程序、WEB应用、数据库、安全产品和操作系统等方面的安全.

3.3.1　安全态势指标量化

综合攻击、漏洞和资产等属性,对漏洞种类、威胁程度、危害等级、通用安全漏洞评分系统分数、漏洞被

算法1 DE-GWO组合算法

输入: 交叉因子CR缩放因子初始值λ_0

过程:

1. 初始化(α,β,δ)
2. 计算个体适应度$f(x_i)$
3. 比较选择父代(α,β,δ)
4. **FOR** $g\leftarrow 1$ **TO** G **DO**
5. **FOR** $i\leftarrow 1$ **TO** N **DO**
6. 更新其他个体位置$x_i(g)\leftarrow x_{\alpha,\text{best}}(g-1)-A\circ D_i(g-1)$
7. $k\leftarrow k+1$
8. 产生中间体$V_i(g)\leftarrow x_{r3}(g)+\lambda(x_{r1}(g)-x_{r2}(g))$
9. 进行重组交叉得到$U_i^k(g)$
10. 比较选择优势个体$x_i(g)$
11. **END FOR**
12. **END FOR**
13. 更新最优灰狼信息$x_{\alpha,\text{best}}(g+1)\leftarrow\frac{1}{3}\sum_{i=1}^{3}x_i$

输出: $x_{\alpha,\text{best}}$

用概率、相关资产信息、影响设备类型、安全产品数量等其他信息进行要素提取[18]. 并从以下几个方面考虑安全态势量化.

(1)脆弱性:包括利用漏洞进行攻击的可能性、工控网络防御强度、漏洞被利用后将造成的危害性等.

(2)威胁性:包括漏洞类型、漏洞生命周期、漏洞攻击强度、安全事件发生频率、攻击者可能利用漏洞数量以及网络攻击手段等.

(3)资产价值:包括安全设备数量、影响安全产品类型、被攻击设备重要程度、设备运行状态、专有操作以及应用系统信息等.

在考虑影响工业互联网安全态势各个指标的基础上,将提取到的数据进行预先量化分级处理[19],如表1所示.

而后引入安全态势值SA作为对工业互联网安全状态的量化评价.

表1 指标量化表

量化类型	等级	指标	定义
脆弱性(V)	一般	1	漏洞被利用后对系统影响较小
	低级	2	漏洞被利用后对系统影响中等
	中级	3	漏洞被利用后对系统影响较大
	高级	4	漏洞被利用后对系统影响很大
威胁性(T)	一般	1	出现安全事件的频率较小
	低级	2	出现安全事件的频率中等
	中级	3	出现安全事件的频率较大
	高级	4	出现安全事件的频率很大
资产价值(A)	一般	1	组件的重要程度一般
	中等	2	组件的重要程度中等
	重要	3	组件的重要程度很高

$$\text{SA}=\omega_1 V+\omega_2 T+\omega_3 A \tag{20}$$

其中,SA表示一个周期的安全态势值;$\omega_1,\omega_2,\omega_3$为权重;$V$表示受攻击后系统的脆弱性量化值;$T$表示受攻击后系统的威胁性量化值;$A$表示受攻击资产的价值量化值.

从安全数据库中提炼态势信息并进行量化处理的具体过程如图4所示.

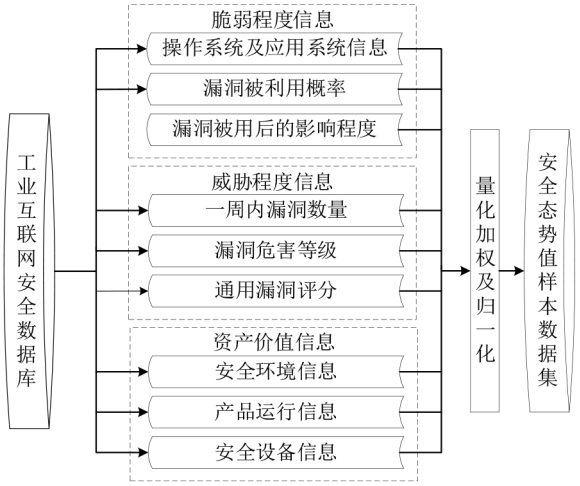

图4 安全态势值量化

3.3.2 安全态势预测

本文通过DE算法和自适应参数调整策略改进GWO算法,然后基于改进的GWO算法智能选取SVR预测模型的惩罚因子C和自由参数g,并建立基于优化SVR算法的工业互联网安全态势预测模型. 该模型的工作流程如图5所示.

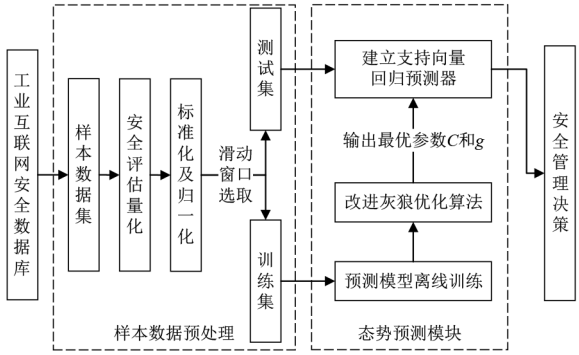

图5 基于优化SVR算法的工业互联网安全态势预测流程图

工业互联网安全态势预测模型以SVR算法为核心完成预测过程,主要步骤包括:

(1)对数据集进行预处理,并以4:1的比例分为训练集和测试集;

(2)设定初始CR值和λ_0值,利用训练集对态势预测模块进行训练,比较更新父代α,β,δ个体,再利用式(7)更新其他个体向量;

(3)利用式(14)使新生子代狼变异,然后将中间突变体利用式(15)进行重组交叉,从而产生变异子代,最后利用式(16)比较得到新α狼,并更新其他个体;

(4)当达到算法循环停止要求,则输出父代α狼的相关信息,得到最优参数组合(C,g),用于构建支持向量回归预测器,并利用测试集计算安全态势值、进行安全态势预测.

4　仿真实验与结果分析

4.1　安全态势数据预处理

本文以工业互联安全应急响应中心和国家互联网应急中心在2018年1月1日至2020年1月19日期间,公开发布的107周的安全数据为实验基础.

从工业互联网安全数据中提取的态势信息进行量化处理,得到107组原始数据. 然后,根据实际工业互联网安全环境对网络安全的要求,可设置一小时、一天、一周或一个月等为一个周期,并利用式(20)得到安全态势值. 其中,利用层次分析法[20]确定权重$\omega_1,\omega_2,\omega_3$分别为0.548 6,0.385 5和0.065 9.

为了加快预测速度,利用式(21)对SA数据进行归一化处理,有

$$x_i' = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{21}$$

其中,x_i和x_i'分别表示处理前后的态势值;$x_{\max}$和$x_{\min}$表示原数据中的最大值和最小值.

数据预处理后,得到107个处于0~1之间的归一化安全态势值,如图6所示. 然后,采用滑动窗口方式重构得到训练集$D=(X_i,Y_i)$和测试集$T=(X_i,Y_i)$.

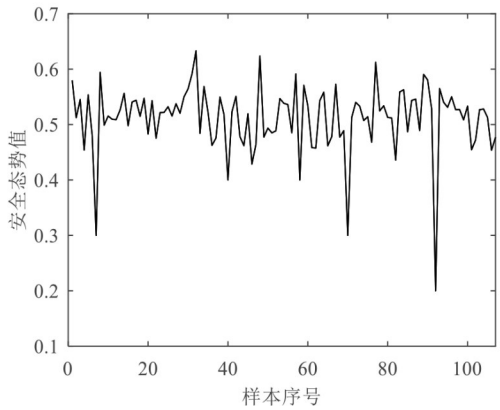

图6　归一化后的安全态势值

本文对归一化后的安全态势值进行重构时,设置滑动窗口为6,每次滑动单位为1,即利用过去5个周期的安全状态预测未来1个周期的安全状态. 所以对107个安全态势数据,分别使用前87个数据和后25个数据进行重构,结果如表2所示. 最后得到包含82组数据的训练集D和包含20组数据的测试集T.

表2　样本数据构建

5维输入	1维输出
X_1, X_2, X_3, X_4, X_5	X_6
X_2, X_3, X_4, X_5, X_6	X_7
…	…
$X_{n-5}, X_{n-4}, X_{n-3}, X_{n-2}, X_{n-1}$	X_n

利用数据集D训练优化SVR算法的参数C和g. 然后,利用数据集T评价模型的预测效果. 初始参数设置如表3所示.

表3　初始化参数设置表

初始参数设置	初始值
种群规模N	30
最大迭代次数G	300
缩放因子初始值λ_0	0.5
交叉概率CR	0.9

在整个迭代过程中,以均方误差衡量个体适应性. 用改进的GWO算法优化得到本文模型的最佳参数.

4.2　预测实验与对比分析

本文方法选取MEA-BP[21]模型、PSO-SVM[22]模型、未改进的GWO-SVR模型和本文方法,分别使用测试集进行预测对比实验.

4.2.1　预测结果对比分析

利用测试集对4种模型进行预测对比实验,结果如图7所示.

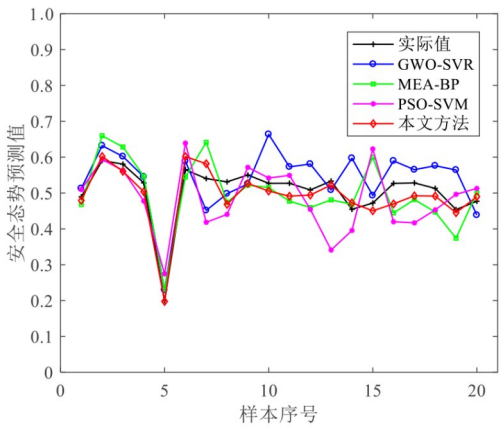

图7　预测结果对比

从图7可以直观地看出,4种模型都有一定的预测能力,但本文方法能够更好地拟合实际的预测变化曲线. 相比其他3种方法,本文方法的预测结果更好,尤其是在局部极值点处能够进行更准确的预测. 例如样本序号为1,5,7,12,15,16,19处,利用本文方法得到的预测值更贴近实际值.

4.2.2 绝对误差对比分析

计算4种模型的每个预测结果相对于实际值的绝对误差(Absolute Error,AE),以直观地对比预测误差的变化趋势和局部极值点处的预测误差情况,结果如图8所示.

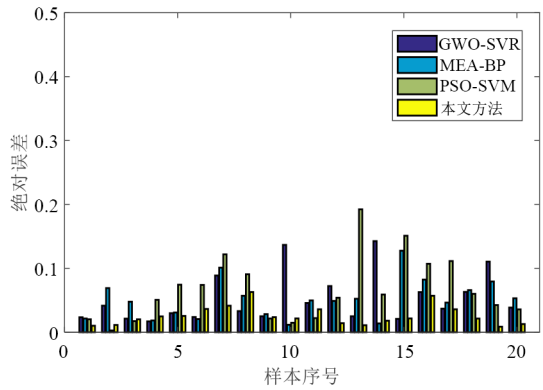

图8 预测结果绝对误差对比

由图8分析得到,整体上本文提出的安全态势预测模型相对于另外3种模型,绝对误差更小,且AE值均稳定控制在0.07之内.这说明本文提出的自适应参数调整策略能够使种群进行有效的变异进化,验证了算法的有效性.

4.2.3 预测准确率对比分析

定义容许偏差Δ,并采用式(22)计算预测模型的准确率,即

$$\mathrm{Acc}=\frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FN}}\times 100\% \tag{22}$$

其中,Acc表示准确率;TN表示$\Delta\leqslant$AE的预测值总数;FN表示$\Delta>$AE的预测值总数.

为了直观地表现本文方法的优势,分别设定Δ值为0.1及0.05,对比结果如表4所示.

表4 模型准确率对比

预测模型	准确率Acc	
	Δ=0.1	Δ=0.05
GWO-SVR	85%	65%
MEA-BP	85%	55%
PSO-SVM	70%	40%
本文方法	100%	90%

在预测准确性方面,本文方法相对于其他3种模型,在容许偏差为0.1时分别提高了15%,15%,30%,在容许偏差为0.05时分别提高了25%,35%,50%,从而证实了本文预测模型的优越性.

4.2.4 预测精度对比分析

对选用的4种模型进行定量分析,采用平均绝对误差(Mean Absolute Error,MAE)、平均绝对百分比误差(Mean Absolute Percentage Error,MAPE)和均方根误差(Root Mean Square Error,RMSE)这3个指标,相关表达式如下:

$$\mathrm{RMSE}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\mathrm{observed}_i-\mathrm{predicted}_i)^2} \tag{23}$$

$$\mathrm{MAE}=\frac{1}{N}\sum_{i=1}^{N}\left|\mathrm{observed}_i-\mathrm{predicted}_i\right| \tag{24}$$

$$\mathrm{MAPE}=\frac{1}{N}\sum_{i=1}^{N}\left|\frac{\mathrm{observed}_i-\mathrm{predicted}_i}{\mathrm{observed}_i}\right| \tag{25}$$

其中,$\mathrm{observed}_i$表示第i个真实值;$\mathrm{predicted}_i$表示对应的预测值;N表示数据总数.

分别计算各模型的3个预测误差指标,结果如表5所示.

表5 预测误差指标对比

预测模型	RMSE	MAE	MAPE
GWO-SVR	0.065 0	0.052 9	0.109 0
MEA-BP	0.058 3	0.049 8	0.101 2
PSO-SVM	0.082 3	0.066 2	0.139 9
本文方法	0.029 1	0.024 6	0.047 2

表5结果进一步证明,本文提出模型的3项预测误差指标均小于0.05,相比其他3种模型的预测误差指标值更小,有助于更好地预测工业互联网安全态势.

4.2.5 模型运行时间对比分析

通过实验记录了不同模型运行总时间和模型预测时间,结果如表6所示.

表6 模型运行时间对比

预测模型	运行总时间/s	预测时间/s
GWO-SVR	6.438	0.123
MEA-BP	0.821	0.263
PSO-SVM	8.621	0.052
本文方法	6.403	0.121

由表6可以看出,本文方法对比未优化的GWO-SVR模型,运行时间上未表现出明显减少,但结合表4和表5可知,本文方法的主要优势在于提升预测效果,也正是因为方法上的改进聚焦预测准确性和精度的提高,算法需要一定的运行时间成本作支撑,导致其难以下降.PSO-SVM模型结构简单,预测时间较短,但预测性能指标最差,难以满足实际需要.

5 仿真实验与结果分析

为了更有效地防护工业互联网安全,支持工业互联网健康发展,本文针对工业互联网安全数据高复杂性、非线性和特异性等特点,提出了一种基于优化SVR算法的工业互联网安全态势预测方法.该方法在对工

业互联网安全数据分析、量化的基础上,首先利用DE算法和自适应参数调整策略改进灰狼优化算法,防止预测算法后期陷入局部最优而导致模型预测精度降低;然后利用改进的灰狼优化算法的快速收敛能力及全局寻优能力优化SVR参数,建立基于优化SVR算法的工业互联网安全态势预测模型;最后利用该模型完成预测过程.

通过仿真对比分析可知,相比于其他方法,在容许偏差为0.05时,本文方法的预测准确率为90%.在容许偏差为0.1时,本文方法的预测准确率达到100%. 而且,预测结果的绝对误差稳定在0.07之内,3项预测误差指标值也均保持在0.05以下,说明了本文方法可取得更精确的预测结果. 所以,本文方法可以为新兴的工业互联网安全防护工作提供新的技术路径.

参考文献

[1] LI J Q, YU F R, DENG G Q, et al. Industrial Internet: A survey on the enabling technologies, applications, and challenges[J]. IEEE Communications Surveys & Tutorials, 2017, 19(3): 1504-1526.

[2] WU Y H, HU X D. Industrial Internet security protection based on an industrial firewall[C]//2021 IEEE International Conference on Artificial Intelligence and Computer Applications. Dalian: IEEE, 2021: 239-247.

[3] 陆耿虹, 冯冬芹. 基于粒子滤波的工业控制网络态势感知建模[J]. 自动化学报, 2018, 44(8): 1405-1412.
LU G H, FENG D Q. Modeling of industrial control network situation awareness with particle filtering[J]. Acta Automatica Sinica, 2018, 44(8): 1405-1412. (in Chinese)

[4] 李艳, 王纯子, 黄光球, 等. 网络安全态势感知分析框架与实现方法比较[J]. 电子学报, 2019, 47(4): 927-945.
LI Y, WANG C Z, HUANG G Q, et al. A survey of architecture and implementation method on cyber security situation awareness analysis[J]. Acta Electronica Sinica, 2019, 47(4): 927-945. (in Chinese)

[5] JAJODIA S, LIU P, SWARUP V, et al. Cyber Situational Awareness: Issues and Research[M]. New York: Springer, 2010: 25-34.

[6] BASS T. Multisensor data fusion for next generation distributed intrusion detection systems[C]//1999 IRIS National Symposium on Sensor and Data Fusion. Laurel: The Johns Hopkins University Applied Physics Laboratory, 1999: DOI:10.13140/RG.2.2.20357.96482/1.

[7] 王益丰, 李涛, 胡晓勤, 等. 一种基于人工免疫的网络安全实时风险检测方法[J]. 电子学报, 2005, 33(5): 945-949.
WANG Y F, LI T, HU X Q, et al. A real-time method of risk evaluation based on artificial immune system for network security[J]. Acta Electronica Sinica, 2005, 33(5): 945-949. (in Chinese)

[8] 李方伟, 郑波, 朱江, 等. 一种基于AC-RBF神经网络的网络安全态势预测方法[J]. 重庆邮电大学学报(自然科学版), 2014, 26(5): 576-581.
LI F W, ZHENG B, ZHU J, et al. A method of network security situation prediction based on AC-RBF neural network[J]. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), 2014, 26 (5): 576-581. (in Chinese)

[9] HU G Y, QIAO P L. Cloud belief rule base model for network security situation prediction[J]. IEEE Communications Letters, 2016, 20(5): 914-917.

[10] ZHANG H B, HUANG Q, LI F W, et al. A network security situation prediction model based on wavelet neural network with optimized parameters[J]. Digital Communications and Networks, 2016, 2(3): 139-144.

[11] KOU G, WANG S, TANG G M. Research on key technologies of network security situational awareness for attack tracking prediction[J]. Chinese Journal of Electronics, 2019, 28(1): 162-171.

[12] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016: 133-137.
ZHOU Z H. Machine Learning[M]. Beijing: Tsinghua University Press, 2016: 133-137. (in Chinese)

[13] 顾秋阳, 吴宝, 孙兆洋, 等. 基于改进灰狼优化的复杂网络重要节点识别算法[J]. 通信学报, 2021, 42(6): 72-83.
GU Q Y, WU B, SUN Z Y, et al. Key node identification algorithm for complex network based on improved grey wolf optimization[J]. Journal on Communications, 2021, 42(6): 72-83. (in Chinese)

[14] JIANG T H, ZHANG C. Application of grey wolf optimization for solving combinatorial problems: Job shop and flexible job shop scheduling cases[J]. IEEE Access, 2018, 6: 26231-26240.

[15] WANG S H, LI Y Z, YANG H Y, et al. Self-adaptive differential evolution algorithm with improved mutation strategy[J]. Soft Computing, 2018, 22(10): 3433-3447.

[16] LI Y Z, WANG S H, YANG B. An improved differential evolution algorithm with dual mutation strategies collaboration[J]. Expert Systems with Applications, 2020, 153: 113451.

[17] LIANG W, LI K C, LONG J, et al. An industrial network intrusion detection algorithm based on multifeature data clustering optimization model[J]. IEEE Transactions on

Industrial Informatics, 2020, 16(3): 2063-2071.

[18] LIU J, RUI S P, YANG M, et.al. Software and cyber security—A survey[J]. Journal of Software, 2018, 29(1): 42-68.

[19] 黄家辉, 冯冬芹, 王虹鉴. 基于攻击图的工控系统脆弱性量化方法[J]. 自动化学报, 2016, 42(5): 792-798.
HUANG J H, FENG D Q, WANG H J. A method for quantifying vulnerability of industrial control system based on attack graph[J]. Acta Automatica Sinica, 2016, 42(5): 792-798. (in Chinese)

[20] WANG H, CHEN Z F, FENG X, et al. Research on network security situation assessment and quantification method based on analytic hierarchy process[J]. Wireless Personal Communications, 2018, 102(2): 1401-1420.

[21] XIAO P, XIAN M, WANG H M. Network security situation prediction method based on MEA-BP[C]//2017 3rd International Conference on Computational Intelligence & Communication Technology (CICT). Ghaziabad: IEEE, 2017: 1-5.

[22] 孙卫喜. 用于网络安全态势预测的粒子群与支持向量机算法研究[J]. 计算机应用与软件, 2019, 36(6): 308-316.
SUN W X. Pso and svm for network security situation prediction[J]. Computer Applications and Software, 2019, 36(6): 308-316. (in Chinese)

作者简介

胡向东 男,1971年出生,四川广安人. 重庆邮电大学教授,博士生导师. 主要研究方向为智能感知、网络化测量及工业互联网安全,物联网安全智能理论与技术,复杂系统建模、仿真与优化等.
E-mail: huxd@cqupt.edu.cn

吕高飞 男,1995年出生,河南洛阳人. 主要研究方向为工业互联网安全.