基于改进支持向量机的网络安全态势评估模型的研究

陈小永[1],赵 飞[2]

(1. 安徽电子信息职业技术学院 软件工程学院,安徽 蚌埠 233000;
2. 安徽工业大学 计算机科学与技术学院,安徽 马鞍山 243032)

[摘要]在信息安全领域,网络安全态势评估已成为信息安全领域中的一个主要研究方向。在以往研究基础上,将支持向量机与鱼群算法相结合,研究面向信息系统的信息安全态势评估模型,提出一种基于遗传算法的优化方法,并将其应用于支持向量机中的参数 C、参数 g 的搜索。利用 SVM 进行分级,选择径向基函数作为支持向量机的核心,采用基于遗传算法的多目标优化算法,并与传统的多目标优化算法进行了比较。结果显示,基于鱼群算法的改进支持向量机信息融合方法能够对现有的信息系统进行快速、准确和实时的计算机网络安全态势评估,具有一定的研究及应用价值。

[关键词]计算机网络安全;态势评估;支持向量机;鱼群算法
[中图分类号]TP393.08 [文献标识码]A [文章编号]1671 - 5330(2025)02 - 0017 - 06
DOI:10.16140/j.cnki.1671-5330.2025.02.007

0 引言

计算机网络安全态势评估旨在通过对信息数据的分析与建模,对当前的信息安全形势进行评价,并利用该方法对信息的安全情况进行预警,以实现对信息环境的有效控制,对信息系统的安全具有十分重要的意义[1]。目前主要采用的评价手段包括:因子分析法、历史比较法和逻辑分析法、量化评价法、故障树分析法、聚类分析法、熵系权系数法等定性和定量评估方法[2-3]。通过近几年的研究,国内外都取得了较大的进展,Vinod 等根据蜂巢中的资料,给出了一种计算网络攻击价值的算法[4]。李欣等将隐马尔可夫方法用于对各级别的安全状况进行训练,通过对各级别的安全级别的转移概率的计算,从而对其进行威胁程度安全的评价[5]。赵文涛等提出了一种基于层次式的攻击行为识别模型,并基于 LAMBDA 的时空因子表示算法,对信息系统的信息安全状态进行了预测[6]。刘晋等通过层层递进的安全危险度指标来量化评价,对评价指标的定量化具有重要意义[7]。但在数据量日益增加的海量信息时代,传统的信息安全态势分析方法面临巨大挑战,迫切需要一种新的方法来解决这一问题。在这一背景下,提出一种改进支持向量机的网络安全态势评估模型,实现对网络安全态势的预测。

1 人工鱼群算法与支持向量机

1.1 人工鱼群算法

人工鱼群算法是一种基于遗传优化的自下向上搜索方法,该方法包含了鱼类的觅食行为、群集行为和追尾行为等多个方面的优化问题。通过与遗传算法和蚁群算法比较,发现该算法不但具有

[收稿日期]2024 - 05 - 23
[基金项目]2024 年度安徽省高校自然科学研究重点项目“基于图神经网络 GNN 的漏洞检测算法研究与应用”(项目编号:2024AH050095);安徽省质量工程项目(项目编号:2022kcsz020)。
[作者简介]陈小永(1980—),男,安徽蚌埠人,讲师,主要研究方向为计算机网络;赵飞(1984—),男,安徽合肥人,博士,副教授,主要研究方向为网络安全。

良好的全局寻优性能,而且具有快速收敛和高精度等优点[8]。鱼群算法假定 x 是一条人造鱼的目前的方位, 视觉上的参量 visual 是其能见度,地点 x_{best} 是它在某个时间点上的视角距离,对目前地点中的食物浓度 σ 进行判定,如果食物的密度比目前的地点高,那么就向前一步,达到地点 x_{nest}。巡逻的次数足够大后,即可能更好地理解周边的环境,避免陷入局部极值,为寻求更好的解决方案提供依据。

1.2 支持向量机

支持向量机的优点在于能够处理小样本、非线性和高维图像的识别。该方法具有集成度高、学习速度快、理论扎实、自适应能力强、推广效果好等一系列优点[9-10]。支持向量 $X(i)$ 与向量 X 的内乘是支持向量 $x_{(i)}$ 与输出向量 x 的内乘,改善了现有算法存在的过度学习和易陷入局部极值等缺陷;同时,通过引入一种新的核函数,有效地避免了低维非可分离问题,其核心理念在支持向量机的构建中起着至关重要的作用。

2 基于鱼群算法改进支持向量机

2.1 网络安全态势数据

针对我国目前网络动态变化的复杂因素等计算机网络安全方面的问题,研究基于国家互联网应急管理中心发布的计算机网络安全数据,该中心通过事故发现、预警通报、应急处置和测试评估等手段,掌握了国内的安全现状,并以每周、每月、季度和年度的方式,对全国范围内的因特网的安全情况发布报告[10]。基于网络安全报告,对报告中发生的计算机网络安全事件进行了统计与归类。选取了 2020 年第 31 期至 2022 年第 49 期的网络安全信息动态周报告内容,统计了国内被篡改的网站总数、国内被植入后门网站总数、国内网站的仿冒网页数量、新增信息安全漏洞数量、高危漏洞数量以及网络安全态势数值。通过分析这些安全要素的特性,并对其关键数据进行了分析,得出了以下 5 个安全要素:被感染的服务器数量、国内被篡改的网站总数、国内被植入后门网站的总数、国内网站的仿冒网页数量以及新增的信息安全缺陷的数量。将各系统的安全状况分为“优”“良”“中”“差”和“危险”5 个级别,并用数字 5,4,3,2,1 来表示。

2.2 SVM 核函数

采用 C-SVC 分类模型作为支持向量机分类的模型,如公式(1)所示。

$$\min_{\omega,b,\xi} \frac{1}{2}\omega^T\omega + C\sum_{i=1}^{t}\xi_i \tag{1}$$

$$subject \quad to \quad y_i(\omega^T\phi(x_i) + b) \geqslant 1 - \xi_i \tag{2}$$

$$\xi_i \geqslant 0, i = 1,\cdots,l$$

其决策函数为:

$$\mathrm{sgn}\omega^T(\phi(x) + b) = \mathrm{sgn}\sum_{i=1}^{t}(y_i a_i K(x_i + x) + b) \tag{3}$$

在模型中的 C 为参数,其范围为 0 到无穷大。

支持向量机的核心函数包括线性、多项式、径向基函数以及 Sigmoid 核函数等。在选取核心功能时,采用图 1 步骤。

图 1 支持向量机模型核函数选取过程

实验选用 2020 年第 25 周到 2021 年的第 32 周中的数据,一共 99 个样本。横向对照试验采用对核心功能筛选控制变量的方法。对于线性、多项式、高斯径向基核函数等多种核函数,采用相同的样本对各种核函数进行训练。实验拟将 99 样本划分成 2 个样本群,通过选择 svmtrain 中的各种特征来对各核函数进行调控,并利用所得到的样本进行学习,建立支持向量机的分类模型。在此基础上,利用该方法对样本进行训练,以检验其识别精度。用正确率判定函数核的好坏,并对不同类型的核函数进行实验验证,从而判定其稳定性。纵向对照方法,通过比较 4 种核心功能在不同数量的训练样本和检验样本数量上的差异,从而研究检验精度随样本个数的变化规律。

为此,选择不同数量的训练样本及预测样本,探究线性核函数的准确率,实验结果如表 1 所示。结果表明,采用线性核函数方法进行识别,具有较高的精度。实验发现,线性核函数方法具有良好的稳定性,在 60/39、70/29、80/19、90/9 这 4 个系列的纵向对照实验中,其准确率均超过了 78%。显然线性核函数是一种具有更高适合度的备选核函数。

表 1　线性核函数的准确率表

(训练集/预测集)数量	60/39	70/29	80/19	90/9
(训练集/预测集)准确率/%	90.8/78.38	90.1/81.65	91.14/83.13	91.09/88.78

多项式核函数准确性验证同样通过不同数量的训练样本及预测样本进行多项式核函数的模拟实验,如表 2 所示。结果表明,训练集 90 个和预测集 9 个的多项式核函数的识别精度可以超过 90%。然而,60/39、70/29、80/19 等多个系列的多项式核函数的分类准确度较低,最低识别准确率仅为 34.78%。因此,采用多项式核函数并不合适。

表 2　多项式核函数的准确率表

(训练集/预测集)数量	60/39	70/29	80/19	90/9
(训练集/预测集)准确率/%	49.8/34.78	84.6/71.45	81.4/64.05	93.36/100

高斯径向基核函数的实验中,选择不同数量的训练样本及预测样本,如表 3 所示。结果表明,径向基核函数具有很高的识别精度。通过实验发现,在 60/39、70/29、80/19、90/9 这 4 个系列的纵向对照实验中,其准确率均大于 87%。当训练样本数目达到 90 个时,其识别准确率可达97.75%。因此,高斯径向基核函数相对于线性的核函数而言,更适合于该网络安全态势分析。

表 3　高斯径向基核函数的准确率表

(训练集/预测集)数量	60/39	70/29	80/19	90/9
(训练集/预测集)准确率/%	97.78/95.8	96.62/87.32	96.4/95.62	97.75/97.55

另外,还对 Sigmoid 核函数进行实验,综合对比以上 4 种核函数的实验结果如表 4 所示。

通过线性、多项式、高斯径向基、Sigmoid 等核函数的比对,证实了线性核函数在识别精度和稳定性上优于传统核函数,但在实际应用中存在着较大的不确定性,Sigmoid 核函数分类预测性能相对较差,径向基核函数在分类性能和稳定性方面都表现出良好的性能。因此,经过比较,选择高斯径向基核函数为该样本集合的最优核,后续实验将选择高斯径向基核函数为支持向量机。

表 4　4 种核函数的准确率对比表

(训练集/预测集)数　量	线性核函数准确率/%	多项式核函数准确率/%	高斯径向基核函数准确率/%	Sigmoid 核函数准确率/%
60/39	90.8/78.38	49.8/34.78	97.78/95.8	74/78.58
70/29	90.1/81.65	84.6/71.45	96.62/87.32	80.52/67.88
80/19	91.14/83.13	81.4/64.05	96.4/95.62	76.41/62.25
90/9	91.09/88.78	93.36/100	97.75/97.55	97.75/97.55

3　鱼群算法优化支持向量机寻优模型

3.1　鱼群算法优化支持向量机

联合鱼群算法和支持向量机,将鱼群算法的寻优能力,用于支持向量机参数选优,提出了一种基于遗传算法的 SVM 优化方法。利用鱼群算法搜索搜寻、群集与追逐等策略,分别构造出最优化的 SVM 惩罚系数 C 与核函数参数 g,并将其应用于搜索参数 C 与 g,使用鱼群算法求出最优的食物浓度坐标 (x_1, x_2),该方法的具体流程如图 2 所示。

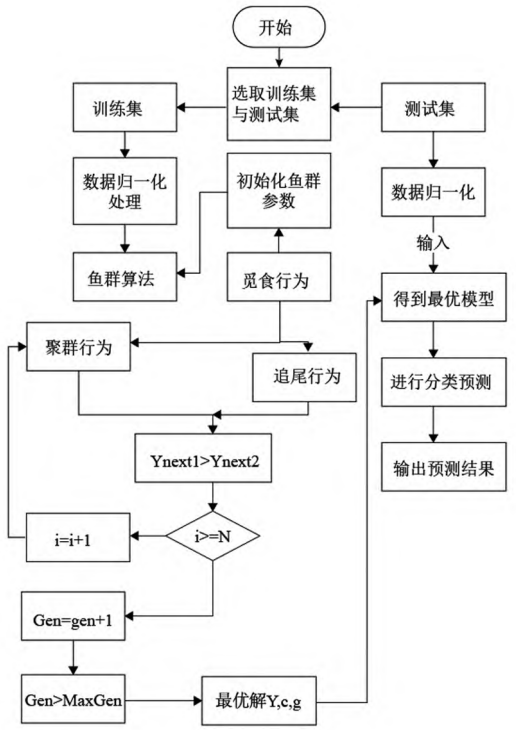

图 2 鱼群算法流程图

人工鱼群算法优化支持向量机网络态势评估模型的基本思路为：

1）先通过对网络安全性评价指数的筛选，将其划分成一个训练集合和一个试验集合，每个集合都是一个输入矢量 $R_m=[x_{m1},x_{m2},x_{m3},x_{m4},x_{m5}]^T$。

2）对所得到的信息进行标准化，将其作为支撑向量机的一个输入，并以其所得到的信息的大小来表示该系统的安全程度。在此基础上，采用群算法对 SVM 的参数进行优化，以求出最佳惩罚系数 C 及核函数参数 g。

3）鱼群初始化。对每个人工鱼进行初始化，每个人工鱼作为支持向量机最优的参数组合 $(bestX_1,bestX_2)$，以食物浓度 $bestY$ 作为输出，作为最终的学习准确度。为提高搜索精度，并使其符合惩罚系数 $C\geqslant a_i\geqslant 0$，将其取控制在 $[-10,10]$ 的范围，对人工鱼进行随机初始化，将 2 * fish_num 阵列为全局，fish_num 人造鱼为全局优化问题。

4）借鉴鱼群算法中不同类型的个体特征。通过操纵人工鱼的采食、追逐和聚集等行为，以食物浓度作为判断精度的准则，视觉和步长作为运动控制参数。通过反复比较，求出最大的人工鱼食物浓度值，并在此基础上保持优化后的参数。

5）判定结束条件：判定预先设定的最大迭代次数，并将最优的参数输入到预测模型中，得出对应的网络安全风险等级。

其中人工鱼群算法所选取的鱼群算法的参数如表 5 所示。

表 5 鱼群算法参数表

参数名称	参数变量	参数值
鱼群数量	Fishnum	30
最大迭代次数	MAXGEN	40
视野	Visual	1
步长	Step	0.3
拥挤度因子	Delta	0.618

设置鱼群算法的各种参数的目的是通过鱼群算法寻优寻找 SVM 的惩罚系数 C 和核函数参数 g，针对鱼类群体，将惩罚系数 C 和核函数参数 g 转换成鱼类算法中的变量 c、g，以求出具有最好精度的罚因子 C 及核心函数 b，从而得到最优化的参数 c_best、g_best。

3.2 鱼群算法优化支持向量机试验结果

经过核函数选择实验发现，实验准确率与测试和预测个数的比例呈现正相关性，因为，选择已有的数据作为检验用例（80/19，90/9）进行比较，鱼群算法在迭代阶段所得到的结果如图 3 所示。

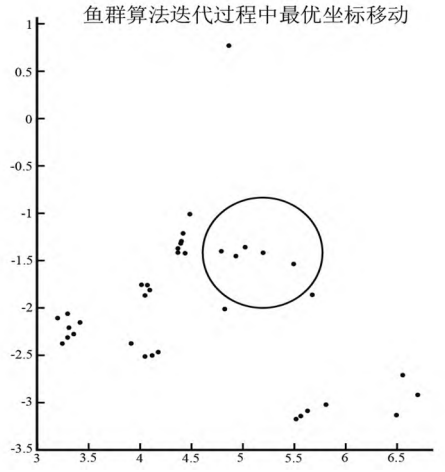

图 3 鱼群算法迭代效果图

优化后选取参数后的结果如图 4 和图 5 所示。

实验结果表明，该方法能有效地解决陷入局部最优解的难题，并能实现全局最优。另外，实验表明，对 80/19 组或 90/9 组，该方法均具有很好的搜索性能，其性能均优于 95%，且在最佳状态

下,90/9 的测试组合可以实现多重预测。因此,仅从实验模型的稳定性以及实验结果的准确性来看,该方法是可行的。

3.3　粒子群优化寻优模型

粒子群算法源于观测和研究鸟类的捕猎行为,利用鸟类寻找猎物的最简便方法进行仿真。该方法将所有的颗粒看作一种可能的求解方法,并且每一种颗粒都有一个适应值。粒子的移动速率决定了它的运动轨迹,并随着时间的推移而不断变化[11]。粒子群算法流程如图 6 所示。

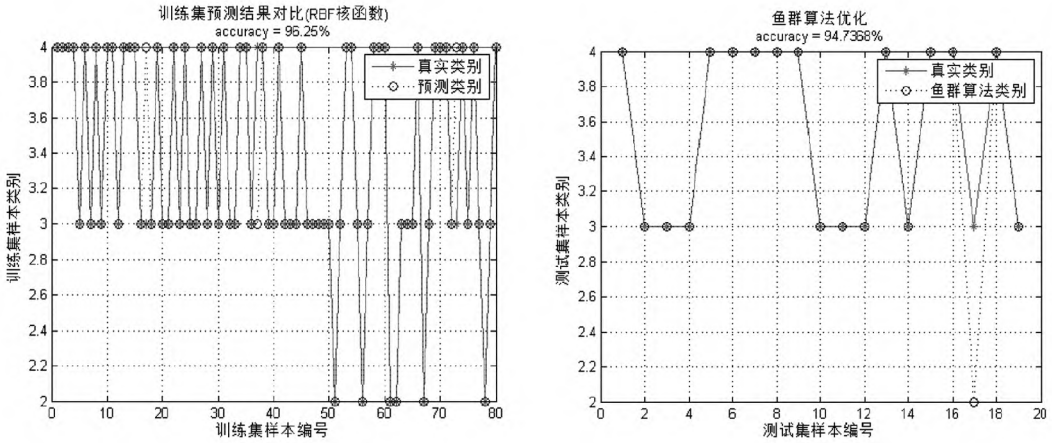

图 4　80/19 试验效果图

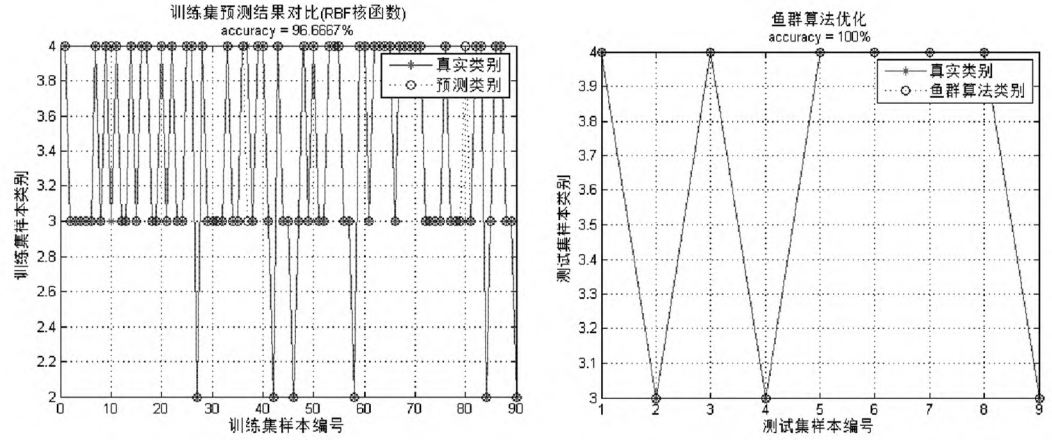

图 5　90/9 试验效果图

使用粒子群算法优化与鱼群算法优化的结果对比如表 6 所示。鱼群算法的迭代次数远比粒子群算法要多,但由于粒子群算法收敛速度太快,无法获得足够的迭代来实现全局优化。

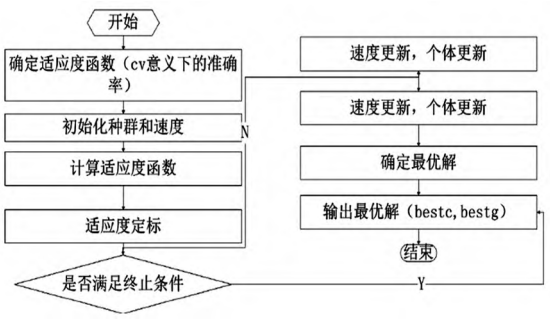

图 6　粒子群算法流程图

表 6　粒子群算法与鱼群算法改进 SVM 结果对比

参数名称	参数名称	训练集/测试集数量	
		89/19	90/9
粒子群算法改进 SVM	gBest(1)	0.0000	0.0000
	gBest(2)	0.0032	0.3425
	TPR/%	85.11	86.77
鱼群算法改进 SVM	C	126.5620	32.0577
	σ	0.1344	0.3174
	TPR/%	94.74	100.00

实验表明,粒子群算法的搜索效率要高于普通的鱼类搜索算法,但是由于只使用了搜索目标的定位与速度,导致搜索过程中搜索的时间较短,

不能获得全局最优。鱼群算法改进 SVM 虽然比粒子群算法改进 SVM 稍差一些,但其稳健性强,并且能够通过快速的寻优而获得最优。

4 结论

以国家互联网应急中心关于计算机网络安全数据为研究对象,结合多角度实验,对支持向量机的核函数进行实验选定,获得最适合网络安全态势评估数据的 SVM 核函数。在此基础上,将鱼群算法用于支持向量机的参数寻优,对支持向量机进行改进,使其与网络安全态势评估相适配。实验证实,相对于粒子群算法,鱼群算法与支持向量机融合后,应用于计算机网络安全态势评估上,其在寻优速度、寻优精度方面优势较为突出,尤其是在寻优精度和全局寻优能力方面粒子群算法无法与之比拟,充分证明了此算法的创新性和实用性。

[参考文献]

[1]王金恒,单志龙,谭汉松,等. 基于遗传优化 PNN 神经网络的网络安全态势评估[J]. 计算机科学,2021,48(6):338 - 342.

[2]肖鹏,王柯强,黄振林. 基于 IABC 和聚类优化 RBF 神经网络的电力信息网络安全态势评估[J]. 智慧电力,2022,50(6):100 - 106.

[3]郝祖龙,袁睿,郝琦,等. 网络攻击下安全级仪控系统人因失误风险分析初探[J]. 核科学与工程,2022,42(4):959 - 967.

[4]席荣荣,云晓春,金舒原,等. 网络安全态势感知研究综述[J]. 计算机应用, 2012,32(1):1 - 4.

[5]李欣,段泳程. 基于改进隐马尔可夫模型的网络安全态势评估方法[J]. 计算机科学,2020, 47(7): 287 - 291.

[6]赵文涛,殷建平,龙军. 安全态势感知系统中攻击预测的认知模型[J]. 计算机工程与科学,2007(11):17 - 19.

[7]刘晋,程彦斌,齐东川,等. 基于支持向量机的化工工艺安全评价模型构建及优化研究[J]. 中国安全生产科学技术,2022,18(12):154 - 161.

[8]赵冬梅,吴亚星,张红斌. 基于 IPSO-BiLSTM 的网络安全态势预测[J]. 计算机科学,2022,49(7):357 - 362.

[9]张克君,郑炜,于新颖,等. 基于 PSO-TSA 模型的网络安全态势要素识别研究[J]. 湖南大学学报(自然科学版),2022, 49(4):119 - 127.

[10]杨慧娟. 基于 TensorFLow 的个性化推荐系统设计[J]. 粘接,2020,41(2):166 - 169.

[11]钱建,李思宇. 基于 RBF 神经网络的网络安全态势感知预测研究[J]. 网络空间安全,2020,11(5):62 - 67.

Research on Network Security Situation Assessment Model Based on Improved Support Vector Machine

CHEN Xiaoyong[1], ZHAO Fei[2]

(1. School of Software Engineering, Anhui College of Electronic Information Technology, Bengbu 233000, China;
2. School of Computer Science and Technology, Anhui University of Technology, Ma'anshan 243032, China)

Abstract: In the field of information security, network security situation assessment has gradually become a main research direction in the field of information security. On the basis of previous research, the information security situation assessment model for information system is studied by combining support vector machine and fish swarm algorithm, and an optimization method based on genetic algorithm is proposed, and it is applied to search parameters and parameters in support vector machine. Svc is used for classification, radial basis function is chosen as the core of support vector machine, and multi - objective optimization algorithm based on genetic algorithm is adopted, and the methed is compared with traditional multi - objective optimization algorithm. The results show that the improved support vector machine information fusion method based on fish swarm algorithm can evaluate the security situation of computer network quickly, accurately and in real time, and has certain research and application value.

Key words: computer network security; situation assessment; support vector machine; fish school algorithm

[责任编辑:张宏亮 责任校对:张宏亮 张怀涛]