机器学习在网络安全入侵检测中的应用

吴 乔
（延安大学，陕西 延安 716000）

摘 要：网络入侵检测技术能够对系统进行实时检测或事后检测，及时发现、防范各种复杂多变的网络攻击企图、攻击行为与攻击结果。同时网络入侵检测技术也是网络空间安全中一个具有挑战性的网络安全问题，机器学习对入侵检测技术性能的提高成为热门研究内容。文章综述了近些年来几种典型的群体智能优化算法，及其与支持向量机相结合的网络入侵检测技术，最后对机器学习在网络安全入侵检测中的未来发展前景进行展望。

关键词：机器学习；入侵检测；网络安全
中图分类号：TP181 **文献标志码**：A **文章编号**：2095-2945(2020)25-0001-04

Abstract: Network intrusion detection technology can detect the system in real time or after the detection, timely find and prevent various complex and changeable network attack attempts, attack behavior and attack results. At the same time, the network intrusion detection technology is also a challenging network security problem in the network space security. This paper summarizes several typical swarm intelligence optimization algorithms in recent years, and its network intrusion detection technology combined with support vector machine. Finally, the future development prospect of machine learning in network security intrusion detection is prospected.

Keywords: machine learning; intrusion detection; network security

1 概述

新时代下，随着云计算，大数据，物联网，5G 网络等技术的迅猛发展，网络流量和网络攻击日益增长且越来越复杂。入侵检测系统[1]（Intrusion Detection System）在计算机网络收集、分析、处理系统或网络的日志文件、网络流量、系统目录和文件的异常变化、程序执行中的异常行为等能够可靠检测到准确的数据信息，及时发现违反安全策略、攻击的行为。因此，网络安全管理员必须快速找到可以改进和提高入侵检测系统的新方法来避免网络遭受黑客攻击。

目前，机器学习（Machine Learning）广泛应用于各学科领域，在数据挖掘[2,3]，计算机视觉[4]，语音识别[5]，自然语言处理[6]，机器人应用[7,8]，材料科学[9]，医学信息[10]，网络安全[11]等领域备受关注，均取得了不同程度的成功。另一方面，网络存在的漏洞、风险、异常、未知恶意软件越来越多，其防御过程难度增大，机器学习改进算法在入侵检测技术中的应用成为很多学者研究关注的热点问题，本文选取了近些年来几种改进的群体智能优化算法与支持向量机融合的网络入侵检测技术开展了综述。

2 群体智能优化算法

K 最邻近（k-Nearest Neighbor，KNN）、模糊聚类（Fuzzy Clustering）、贝叶斯优化（Bayesian Optimization）、随机森林（Random Forest）等经典算法[12]不能够准确得到非线性、多极值等复杂函数的最优解，随着优化理论的不断发展，学者们通过对粒子群、蚁群、鸟群、蜂群、萤火虫群等社会生物的群体活动行为的研究，衍生出了一系列智能群体优化算法，比如遗传算法[13]、粒子群优化算法[14]、蚁群优化算法[15]、人工蜂群优化算

表 1 群体智能优化算法比较

算法	优点	缺点	应用领域
遗传算法	收敛速度快，通用性强	实现复杂，易陷入早熟收敛，依赖于初始种群	函数优化和组合问题
粒子群优化算法	收敛速度快，效率高，算法简单	只有正反馈机制，易陷入局部最优，处理离散问题效果不佳	求解连续函数优化问题
蚁群优化算法	思想简单，易于实现，鲁棒性和搜索能力较强	计算量大，匮乏初始信息素，易陷入局部最优	求解组合优化问题
人工蜂群优化算法	全局寻优能力强，收敛速度快	易陷入局部最优解，迭代后期搜索能力差	求解多变量函数优化问题
人工萤火虫优化算法	算法原理简单，全局极值求解和多极值搜索能力强	实现过程复杂，发现率低，求解精读不高，求解速度慢	求解多极值函数问题

作者简介：吴乔（1989-），女，本科，工程师，研究方向：计算机与网络安全。

法[16]、人工萤火虫群优化算法[17]等。基于群体智能优化算法改进机器学习算法的技术在网络入侵检测领域广泛应用，各种群体智能优化算法比较如表 1 所示。

3 智能融合的支持向量机算法

入侵检测系统(Intrusion Detection System,IDS)可以分为基于误用、异常和混合网络入侵检测系统[18],入侵检测系统分类如图 1 所示。由于 DOS、R2L、U2R、probe 等网络攻击行为在快速演变,识别“零日”攻击日益重要,而异常入侵检测成为检测的主要研究方向的首要原因是异常入侵检测能够识别未知的网络威胁。学者们研究了多种机器学习检测模型,针对高维、小样本的网络数据集,支持向量机(Support Vector Machine，SVM)模型具有相当的优势,不会面临像聚类算法高错误率、神经网络“维数灾”、过拟合等难题,有更好的泛化能力,能够适用于其他数据集,但是由于 SVM 算法的惩罚参数 C 和核宽度参数 σ 的寻优主要依赖于经验方法，难以保证参数的准确性,而参数直接影响 SVM 模型识别网络攻击分类的准确性。为了能够获得较优参数的 SVM 模型,高效、准确的检测网络攻击行为,诸多学者提出了智能优化算法与 SVM 融合的入侵检测技术。

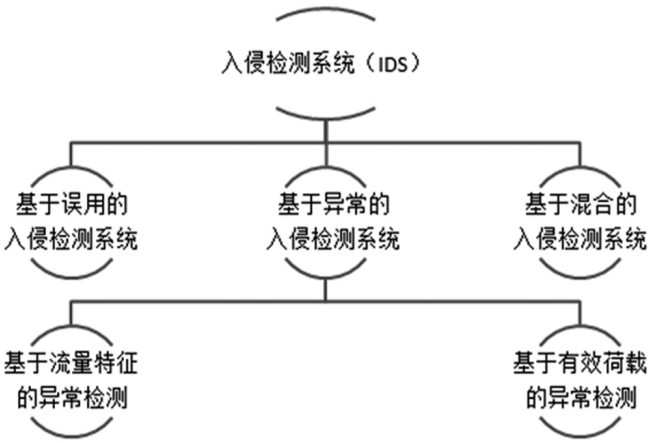

图 1 入侵检测系统分类

3.1 粒子群优化(Particles Swarm Optimization，PSO)算法

PSO 算法是一种收敛速度快、精度高的迭代进化算法,源于 Eberhart 博士和 Kennedy 博士模拟鸟群觅食行为的研究,是类似于模拟退火算法、遗传算法的智能算法。PSO 算法的基本原理步骤如图 2 所示。

余森[19],刘明珍[20]等研究了 PSO 算法和 SVM 融合的网络入侵检测技术。前者研究基于网络系统中的通信数据作为训练集和验证集,KDD 1999 作为测试集,通过采用 PSO 算法确定惩罚参数 C 和核宽度参数 σ 的值,确定网络入侵检测分类器。作者通过试验、对比发现,PSO 算法确定 SVM 参数设计的网络入侵检测分类模型,准确率更高、检测时间更短,降低了

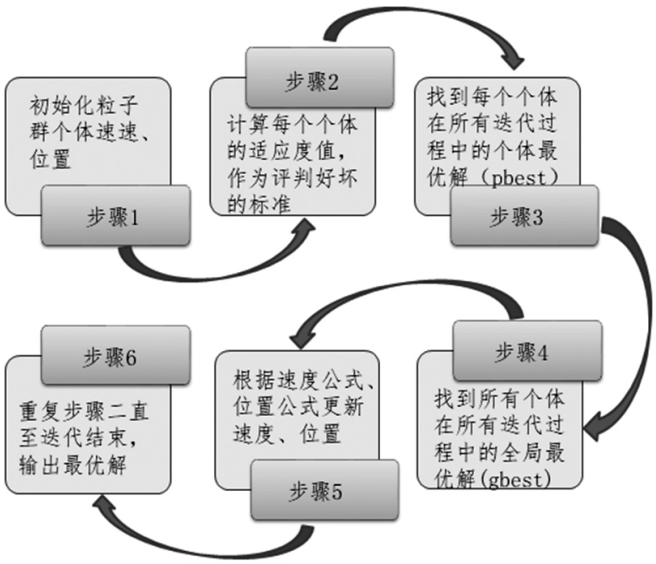

图 2 PSO 算法基本原理步骤

误检率、漏检率,检测效果更好。后者研究基于 KDD 1999 数据集,通过 BPSOA(用二值粒子群优化算法),对数据集的 41 个特征进行特征提取,直至获得最佳特征集,设计网络入侵检测分类器。作者通过试验、对比发现,BPSO 科学的排除了特征集中的噪声特征,同样可以降低检测时间、提高准确率。两者研究均表明,PSO 算法不仅可用于参数寻优,还可用于最优特征集筛选,从而优化网络入侵检测的系统性能。

但基本的 PSO 算法优化 SVM 参数的迭代过程中，种群在搜索空间的聚集度会逐渐提高、多样性会大幅减少,容易导致 PSO 算法陷入局部极值、后期迭代效率低等现象,针对以上不足，诸多学者从改进粒子群优化（Improved Particle Swarm Optimization,IPSO）算法或多种算法融合的角度进行研究,以此优化 SVM 寻优参数的准确率,提高网络入侵检测系统的性能。马占飞[21]等通过在传统的 PSO 算法中设计高斯扰动和适应度方差值,在传统的差分(Differential Evolution,DE)算法中设计缩放因子 F 的自适应动态调整策略,提出了 IQPSO-IDE 入侵检测方法,即改进的粒子群优化和改进的差分(Improved Differential Evolution,IDE)算法融合。该入侵检测模型的设计首先通过在种群粒子的平均位置引入高斯扰动来增加粒子多样性；其次选择能够反映粒子群聚集度的适应度方差 s^2,原理是通过设置阈值来判断算法的搜索能力。s^2 大于阈值,说明粒子群聚集度小,搜索能力较好,未达到局部最优,s^2 小于阈值,说明粒子群聚集度大,搜索能力差,算法陷入局部极值问题;最后利用自适应调整缩放因子 F 的差异策略,原理是动态调整 F 值的大小,在算法的迭代后期,丰富种群粒子的多样性，使其更好的进行 SVM 参数寻优。结果表明,IQPSO-IDE 算法提高了网络入侵检测的正确率、降低了漏报率和误报率。

3.2 蚁群优化(Ant Colony Optimization,ACO)算法

ACO 算法是一种机率型算法，用来在图中寻找优化路径。源于 Marco Dorigo 博士对蚂蚁在觅食过程中,根据其他蚂蚁留下的信息素发现从巢穴到食物源之间最短路径行为的研究,而提出的模拟进化智能算法,具有多样性、正反馈机制、鲁棒性、搜索能力强等特点。ACO 算法的基本原理步骤如图 3 所示。

潘晓君[22]等研究了 ACO 算法与 SVM 融合的网络入侵检测技术。作者将 ACO 与交叉验证两种算法结合,利用蚁群每个个体之间可以共享信息资源的特性,进行局部和全局寻优,有效获得 SVM 参数 C 和 σ 的最优解。基于 KDD 数据集,通过试验发现,针对 DOS、U2R、U2L、Probe 四种网络攻击类型,蚁群优化算法有低误报率、低入侵检测时间、高准确率的优点。

但基本的 ACO 算法存在因发生蚂蚁聚集而陷入局部最优、收敛速度慢、收敛停滞等缺点,为此,王雪松[23],袁琴琴[24]等人提出改进 ACO 算法优化 SVM 的入侵检测技术。前者在蚂蚁进行空间搜索的过程中，为改善因蚁群聚集而导致的局部最优状态,对蚁群个体进行高斯变异,调整蚂蚁的搜索路径,最终选择最优路径,而确定 SVM 的参数;后者利用遗传算法(Genetica1 Algorithm,GA)进行数据特征选择,通过改进节点选择的方法,即在算法迭代后期进行随机节点选择,同时改进信息更新的策略,即增加最优路径信息素、减少最差路径信息素,以便加快算法的收敛速度,最终提出 IACO-GA 网络入侵检测模型。试验验证表明,以上两种改进的 ACO 算法均能较好的克服传统 ACO 算法的缺陷,提升 SVM 寻优效果。

3.3 人工蜂群(Artificial Bee Colony,ABC)算法

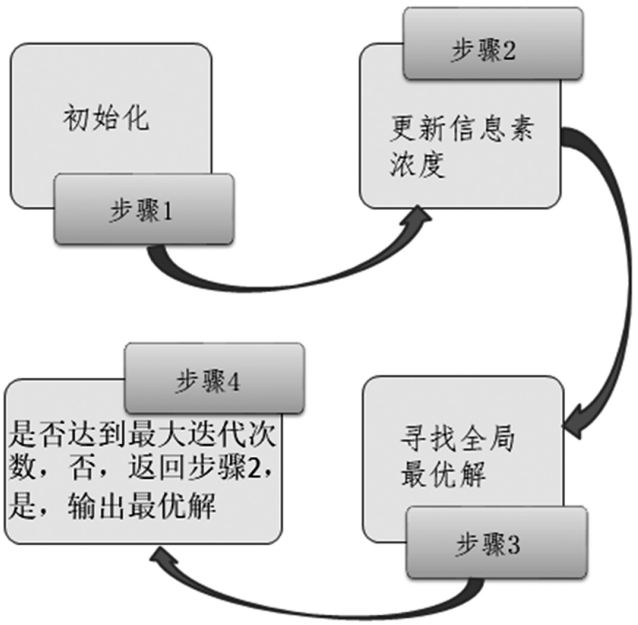

图 3 ACO 算法基本原理步骤

ABC 算法是一种全局搜索能力强、收敛速度较快的群体智能算法，源于 Karaboga 对蜜蜂种群在任何复杂的环境下,能够高效率采蜜行为的研究,根据模拟引领蜂、观察蜂和侦察蜂分配不同的任务进行蜜蜂种群信息交换与共享，最终获得函数全局最优解。ABC 算法的基本原理步骤如图 4 所示。

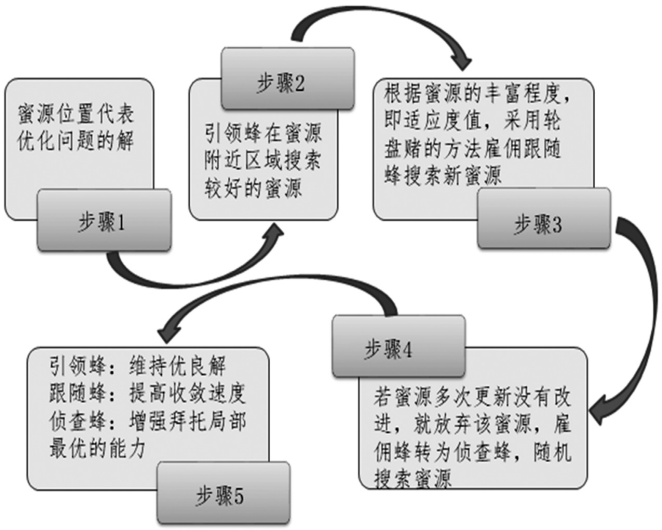

图 4 ABC 算法的基本原理步骤

为了改善 SVM 网络入侵检测模型的参数寻优能力,提高检测正确率,谢伟增[25]提出 ABC-SVM 网络入侵检测技术,通过 ABC 算法将 SVM 惩罚因子 C 和核参数 σ 组合编码成为人工蜂群算法的蜜源,设定算法迭代搜索目标为检测率,多次模拟蜂群寻找蜜源的过程，最终选择较优的 SVM 参数,建立了网络入侵检测模型。试验表明,ABC-SVM 网络入侵检测模型提高了检测正确率,降低了误报率。传统的 ABC 算法易陷入局部最优解且迭代后期搜索能力差,刘铭[26]等人通过引入交叉突变算子,将父代种群与适应度较差的种群进行交叉,提出了基于(CMABCCrossover Mutation ABC)算法的 SVM 网络入侵检测模型,根据适应度值划分蜂群,不仅克服了传统 ABC 算法易陷入局部最优解的缺点,且提高了算法的收敛速度。

4 结束语

网络入侵检测技术是网络安全领域较重要的防护技术,随着突发网络安全事件数量逐渐增加，提高网络入侵检测技术性能尤为重要,本文陈述了基于改进的粒子群优化算法、蚁群优化算法、人工蜂群算法三种群体智能优化算法的支持向量机的网络入侵检测方法,这种融合算法可以取长补短,有效解决收敛速度慢、收敛停滞、易陷入局部最优、种群多样性减少等缺点，大大提升参数寻优能力，设计检测效率较高的模型,减少网络安全事件发生。

但是,就群体智能优化算法和传统的机器算法相融合的入侵检测技术还有很大的提升空间，一是由于网络威胁事件

日益发展和变化,常用检测技术难以应对新的网络攻击,只有继续深究,才能发现识别“零日”攻击的新型技术;二是群体智能优化算法都有各自的优缺点和更适合的应用领域,研究者应同时融合多种算法,才可以扬长避短,提升算法性能,提高检测效率,降低漏检率和误报率,获得精确的网络入侵检测模型。以上两点将是未来网络入侵检测技术研究发展的重要方向。

参考文献:

[1]Kemmerer RA, Vigna G (2002) Intrusion detection: a brief history and overview[J]. Computer,2002,35(4):27–30.

[2]Hitarth Shah,Vishruti Kakkad,Reema Patel,Nishant Doshi. A survey on game theoretic approaches for privacy preservation in data mining and network security [J]. Procedia Computer Science, 2019,155.

[3]GiangNguyen, Stefan Dlugolinsky, Martin Bobák, Viet Tran, Álvaro López García, Ignacio Heredia, Peter Malík, Ladislav Hluchý. Machine Learning and Deep Learning frameworks and libraries for large–scale data mining: a survey [J]. Artificial Intelligence Review, 2019, 52 (1), pp.77–124.

[4]Siddharth Singh Chouhan, Uday Pratap Singh, Sanjeev Jain. Applications of Computer Vision in Plant Pathology: A Survey [J]. Archives of Computational Methods in Engineering: State of the Art Reviews, 2020, 27 (12), pp.611–632.

[5]Thales Aguiar de Lima,Múrjory Da Costa–Abreu.A survey on automatic speech recognition systems for Portuguese language and its variations[J]. ComputerSpeech& Language,2020,62.

[6]Cambria E, White B. Jumping NLP curves:A review of natural language processing research [J]. IEEE ComputIntell Mag, 2014, 9: 48–57.

[7]JaeseokKim,NinoCauli, Pedro Vicente,BrunoDamas, AlexandreBernardino,JoséSantos –Victor,Filippo Cavallo.Cleaning Tasks Knowledge Transfer Between Heterogeneous Robots: a Deep Learning Approach[J].Journal of Intelligent &Robotic Systems: with a special section on Unmanned Systems, 2020, 98 (1) :191–205.

[8]Linzhou Pang,Yunzhou Zhang,Sonya Coleman,He Cao. Efficient Hybrid–Supervised Deep Reinforcement Learning for Person Following Robot [J].Journal of Intelligent & Robotic Systems, 2020, 97 (2):299–312.

[9]Jonathan Schmidt, MárioR.G.Marques, Silvana Botti & Miguel A. L. Marques.Recent advances and applications of machine learning in solid–state materialsscience [J].npjComputational Materials volume 5, Article number: 83 (2019).

[10]Bert Heinrichs,Simon B. Eickhoff. Your evidence? Machine learning algorithms for medical diagnosis and prediction [J]. Human Brain Mapping,2020,41(6).

[11]张蕾,崔勇,刘静,等.机器学习在网络空间安全研究中的应用[J].计算机学报,2018,41(09):1943–1975.

[12] Wu X, Kumar V, Ross Quinlan J, et al. Top 10 algorithms in data mining[J]. KnowlInfSyst, 2008, 14: 1–37.

[13]M. Rabe,M. Deininger,A.A. Juan. Speeding Up Computational Times in Simheuristics Combining Genetic Algorithms with Discrete–Event Simulation[J]. Simulation Modelling Practice and Theory,2020.

[14]王尔申,孙彩苗,黄煜峰,等.改进粒子群优化的卫星导航选星算法[J/OL].北京航空航天大学学报:1–7[2020–08–13].https://doi.org/10.13700/j.bh.1001–5965.2019.0644.

[15]Qiang Luo, Haibao Wang, Yan Zheng, Jingchang He. Research on path planning of mobile robot based on improved ant colony algorithm [J].Neural Computing and Applications, 2020, Vol.32 (6), pp.1555–1566.

[16]XiaoyuSong,MingZhao,QifengYan,Shuangyun Xing. A high–efficiency adaptive artificial bee colony algorithm using two strategies for continuous optimization [J]. Swarm and Evolutionary Computation,2019,50.

[17]杨艳,周永权,罗林,等.人工萤火虫群优化算法求解约束优化问题[J].小型微型计算机系统,2014,35(01):185–188.

[18]RaoufBoutaba, Mohammad A.Salahuddin,Noura Limam, Sara Ayoubi,Nashid Shahriar, Felipe Estrada–Solano,Oscar M.Caicedo.A–comprehensive survey on machine learning for networking: evolution, applications and research opportunities [J].Journal of Internet Services and Applications, 2018, 9 (1):1–99.

[19]余森,赵冉.粒子群算法和支持向量机的网络入侵检测[J].微型电脑应用,2019,35(09):143–145.

[20]刘明珍.粒子群优化支持向量机的入侵检测算法[J].计算机工程与应用,2012,48(35):71–74+105.

[21]马占飞,杨晋,金溢,等.基于 IQPSO–IDE 算法的网络入侵检测方法[J].计算机工程与应用,2019,55(10):115–120+204.

[22]潘晓君.基于 IQPSO 的 SVM 参数优化入侵检测研究[J].宁夏师范学院学报,2019,40(10):80–84.

[23]王雪松,梁昔明.改进蚁群算法优化支持向量机的网络入侵检测[J].计算技术与自动化,2015,34(02):95–99.

[24]袁琴琴,吕林涛.基于改进蚁群算法与遗传算法组合的网络入侵检测[J].重庆邮电大学学报(自然科学版),2017,29(01):84–89.

[25]谢伟增.人工蜂群算法优化支持向量机的网络入侵检测[J].微型电脑应用,2017,33(01):71–73.

[26]刘铭,黄凡玲,傅彦铭,等.改进的人工蜂群优化支持向量机算法在入侵检测中的应用[J].计算机应用与软件,2017,34(01):230–235+246.