Computer Science and Application 计算机科学与应用, 2019, 9(9), 1732-1737
Published Online September 2019 in Hans. http://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2019.99194

Hans汉斯

Research on Opposite Comprehensive Learning Particle Swarm Optimization for Network Intrusion Detection

Xinkai Yang, Tianzhuo Li

School of Information and Electromechanical Engineering, Shanghai Normal University, Shanghai
Email: xkyang@shnu.edu.cn

Received: Aug. 28th, 2019; accepted: Sep. 12th, 2019; published: Sep. 19th, 2019

Abstract

Aiming at the vulnerability of network systems to external attacks, this paper proposes an anomalous network behavior intrusion detection algorithm based on opposition-based learning mechanism. Firstly, an improved particle swarm optimization algorithm is proposed to optimize the parameters of support vector machine. Then a network intrusion detection model is provided. Experiments show that this method can improve the accuracy of classifier parameter optimization, and increase the accuracy of network intrusion identification.

Keywords

Opposition-Based Learning, Particle Swarm Optimization, Network Intrusion Detection

面向网络入侵检测的反向综合学习粒子群优化算法研究

杨新凯，李天卓

上海师范大学信息与机电工程学院，上海
Email: xkyang@shnu.edu.cn

收稿日期：2019年8月28日；录用日期：2019年9月12日；发布日期：2019年9月19日

摘　要

针对网络系统容易受到外部攻击的问题，本文提出了一种基于反向学习机制的异常网络行为入侵检测算

文章引用：杨新凯，李天卓. 面向网络入侵检测的反向综合学习粒子群优化算法研究[J]. 计算机科学与应用，2019, 9(9): 1732-1737. DOI: 10.12677/csa.2019.99194

法。论文首先提出了改进的粒子群算法，对支持向量机的参数进行优化，进而建立了网络入侵检测模型。实验结果验证了该方法可以提高分类器参数寻优的精度，增加了网络入侵行为识别的准确率。

关键词

反向学习，粒子群优化算法，网络入侵检测

1. 引言

随着互联网的迅猛发展，诸如网络黑客入侵之类的风险与日俱增。为了保证网络系统的完整性和可用性，保护网络系统免受内部滥用和外部攻击，网络入侵检测系统及其相关技术得到了越来越多的关注。网络入侵检测系统通过对网络系统或者网络资源进行实时监测，及时发现网络入侵者、发现防火墙没有检测到的攻击，正在成为一种重要的网络安全工具[1]。

当前，网络入侵检测技术仍然在不断的发展完善过程中，还存在着一些不足的地方。例如，将正常的网络访问误报为入侵行为、检测不到或者漏报入侵网络行为的比例依然较高、智能化程度较低等问题依然存在。在此背景下，本文把最小二乘支持向量机(Least Square Support Vector Machine, LSSVM)分类器模型引入到网络入侵检测系统中，提出了一个基于改进 LSSVM 的网络入侵检测模型。

本文结合反向学习理论改进粒子群优化算法，在种群初始化的过程中引入反向学习机制、在目标粒子选择过程中引入反向学习策略，提出了一种反向综合学习粒子群优化算法(Opposite Comprehensive Learning Particle Swarm Optimization, OCLPSO)，通过改进后的算法对最小二乘支持向量机进行参数寻优。相关实验结果验证了该算法不仅提高了收敛速度，计算精度也有所改善。

下文首先简述相关工作，接着基于反向学习理论给出了改进的综合学习粒子群优化算法，通过参数寻优来改善 LSSVM 分类器的性能，随后给出了基于改进的 LSSVM 的网络入侵检测模型。最后通过实验对提出的模型进行验证，实验结果表明该模型具有良好的性能。

2. 相关工作

网络入侵检测概念出现之后，网络入侵检测相关技术处于不断的发展之中。近些年，基于机器学习的网络入侵检测算法得到了越来越多的关注，基于神经网络、遗传算法等的网络入侵检测模型得到了较为广泛的应用[2] [3] [4]。

2005 年，Tizhoosh 等人[5]在 IEEE 智能计算建模会议上首次提出了反向学习的相关理论，其核心思想主要为：对于一个待解决的问题，不仅考虑问题本身的解与最优解的距离，同时也考虑这个问题的反向解与最优解的距离，通过这种方式可以有效改善算法的运行效率和搜索深度。J. J. Liang [6]提出的综合学习粒子群算法(Comprehensive Learning Particle Swarm Optimization, CLPSO)中通过定义学习概率 P_c，扩大了候选解粒子的选择面，但是传统粒子群优化算法容易出现停滞状态从而陷入局部极值点、且在搜索后期收敛速度慢的问题并未得到有效解决。

在机器学习领域里，最小二乘支持向量机通过将损失函数与误差平方作为算法的损失经验，把传统 SVM 中的二次规划问题转换成求解线性方程组的问题[7]，简化了中间计算步骤，提升了算法运行效率。

最小二乘支持向量机作为分类器具有检测精度高、泛化能力好的特点。但是其分类性能很大程度上受到自身参数选择的影响，如何更好的选择合适的模型参数成为提高该方法检测精度的关键。另外，支持向量机算法主要解决二值分类问题，要处理网络入侵检测中的多分类问题，需要对原有分类器进行改造以适应多分类问题。

3. 反向综合学习粒子群优化算法

在综合学习粒子群算法初始化的过程中，针对每个粒子构造反向粒子，使得离最优粒子距离更近的概率得到上升，能够加快算法的收敛速度，改善进化算法的性能。反向学习的一些相关基本概念如下：

定义 1. 反向数(Opposite Number)

若 $x\in[a,b]$ 且 $x\in R$，则其方向数 $\breve{x}$ 为：$\breve{x}=a+b-x$ (1)

定义 2. 反向点(Opposite Point)

若 $p=(x_1,x_2,\cdots,x_d)$ 是 D 维空间上的一个点，且 $x_1,x_2,\cdots,x_dR$，$x_i\in[a_i,b_i]$，则与点 p 对应点反向点 $\breve{p}$ 为：

$\breve{p}=(\breve{x}_1,\breve{x}_2,\cdots,\breve{x}_d)$，其中 $\breve{x}_1=a_i+b_i-x_i$ (2)

定义 3. 反向优化(Opposite Optimization)

若满足 $p=(x_1,x_2,\cdots,x_d)$ 是 D 维搜索空间上的一个候选群，它对应的反向点为

$$\breve{p}=(\breve{x}_1,\breve{x}_2,\cdots,\breve{x}_d) \tag{3}$$

计算候选解适应度值的评估函数 $f(\ )$，若存在 $f(\breve{p})\le f(p)$，则把 p 替换成 $\breve{p}$。

在种群初始化的过程中引入反向学习机制，在目标粒子选择过程中引入反向学习策略。反向综合学习粒子群优化算法简述如下：

1) 种群初始化

a) 随机生成服从正态分布的种群 $P(n)$，n 为种群的大小。

b) 计算 $P(n)$对应的反向种群 $oP(n)$，其中的第 k 个反向粒子计算公式为：

$$oP_{k,j}=a_j+b_j-P_{k,j}$$

上式中 $k=1,2,\cdots,n,\quad j=1,2,\cdots,N_d$

其中 N_d 粒子的维度，a_j 和 b_j 分别对应第 j 个粒子的上下边界，即有

$$x_j\in[a_j,b_j]$$

c) 从集合 $\{P(n),oP(n)\}$ 中选出 n 个最佳个体作为种群的初始值。

2) 目标学习粒子的选择

a) 首先从种群中随机选择两个未进行速度更新的粒子。

b) 比较这两个粒子和它们的反向粒子的适应度值，即从四个待选粒子中选出适应度值最高的一个粒子作为该维度下的目标学习粒子。

c) 如果一个粒子所有的目标学习粒子都是它自身话，则随机选择一个维度下对应的反向粒子作为该粒子的目标学习粒子。

4. 基于反向综合学习粒子群优化算法的网络入侵检测模型

采用一对多的组合方式来对 LSSVM 进行多类分类器的构造。对于给定 n 个类型标签构造 n 个二值分类器，其中第 i 个二值分类器将第 i 类标签归为正类，将除此类之外的所有标签归为负类，即针对数据

集中的每一个类型标签训练一个二值分类器。对于一个具有 n 个类型标签的问题，共产生 n 个二值分类器。对于一条未知类型标签需要分类的数据，采用投票的策略来确定它的类型标签。例如，使用二值分类器 i 对该条数据进行测试，如果最终分类器返回的结果为正值则认为该条数据归为第 i 个类型标签，随后在该类型标签下记一票；反之，如果分类器返回的结果为负值，则认为当前数据属于除去类型 i 之外的其他类型，随后在类型 i 以外的所有类型标签下记一票。最终统计所有类型标签的记票情况，得票最多的那个类型认定为该条数据所属的类型。

对于给定的网络状态特征集合 $F=\{f_1,f_2,\cdots,f_n\}$，采用二进制向量来表示特征选择：$S=\{s_1,s_2,\cdots,s_n\}$，$si\in\{0,1\},i=1,2,\cdots,n$，其中$n$是网络特征集合的大小，1 和 0 分别表示是否选择相应的特征。本文实验采用的数据集中，正常网络数据标记为 normal，攻击类型数据分为 DoS、probe、U2R、R2L 四个大类，分别编码为 0、1、2、3、4，各个类型之间相互独立。以一条 DoS 类型攻击记录为例，通过上述方式构建的 LSSVM 多分类器结构如图 1 所示。在建立多分类器的基础上，结合反向综合学习粒子群算法的网络入侵检测计算步骤如下：

1) 设置模型的相关参数，如最大迭代次数、种群大小、搜索边界等，根据这些参数对种群进行初始化；

2) 根据反向学习策略计算对应的反向种群并归并至初始种群；

3) 将最小二乘支持向量机中的惩罚因子和核参数设置为种群中每一个粒子的位置；

4) 根据速度和位置更新方程进行迭代并对种群中的粒子进行更新；

5) 每轮迭代完成后将得到该轮的最优粒子，该粒子的位置参数作为 LSSVM 的参数代入值，进行分类器训练；

6) 计算适应度函数，判断是否达到终止条件，若达到则终止迭代，当前最优粒子的位置即为最优参数；否则返回(4)中继续进行迭代。

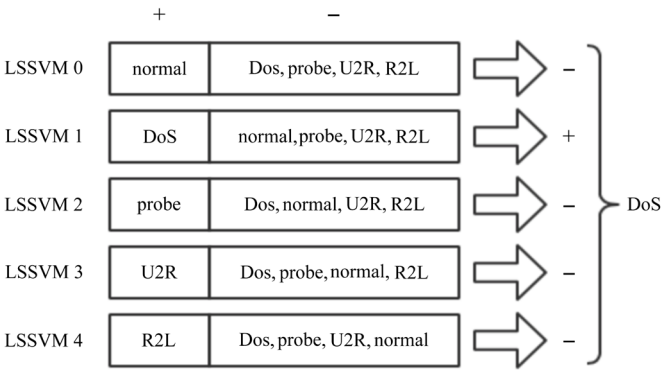

Figure 1. Diagram of multiple classifier with positive DoS example
图 1. DoS 类型为正例的多分类器模型示意图

5. 实验与分析

本文实验所用到的数据集是 1999 年第 3 届国际知识发现与数据挖掘竞赛中所采用的数据集(KDD CUP99)。该数据集中所包含的原始数据主要有两部分组成：其中一部分包含了时间跨度为 7 周约 500 多万条网络连接记录，另外一部分则包含了时间跨度为 2 周约 200 万条的网络连接记录。实验抽取 KDD CUP99 数据集的 10%数据作为训练集，将其中的 corrected 数据集作为测试集。

在对数据进行必要的预处理后，采用不同的基准测试函数来检验改进后的算法对运算时间和求解质量的影响。通过求解基准测试函数的最优解来得到实验结果，其中最大进化代数定为 2000，每种情况下

实验重复 20 次，比较其平均迭代次数、平均花费时间和获得最优解的比率。实验中分别使用三种基准测试函数[8] [9]有：F1 (Griewanks’s function)函数的特点是存在多个波峰极值，F2 (Rastrigin’s function)函数的特点是存在多个局部极值，且很容易出现停滞，F3 (Schwefel’s function)函数的特点是存在多个局部极值，较容易出现停滞。分别对上述三个基准测试函数进行实验，得到实验结果如表 1 所示。

Table 1. Data of benchmark test results
表 1. 基准测试实验结果

	比较算法	平均进化代数	平均收敛时间	最优解率
F1	CLPSO	780.3	587.8 s	95.2%
	OCLPSO	603.5	532.9 s	96.7%
F2	CLPSO	408.5	301.6 s	83.4%
	OCLPSO	303.9	256.7 s	89.1%
F3	CLPSO	2456.7	1687.2 s	71.6%
	OCLPSO	1812.9	1139.2 s	83.3%

由表 1 结果可以看出，在普通函数 F1 的搜索算法中，OCLPSO 只有微弱优势。但在具有一定欺骗性的测试函数中，OCLPSO 存在一定的优势，最优解率提高比较多，说明 OCLPSO 算法更适合具有多分类问题特征的网络入侵检测场景。

进行上述实验后，设计使用 OCLPSO 作 LSSVM 参数寻优的实验，图 2 为实验时 OCLPSO 算法在进化过程中的适应度函数曲线。从图中可以看出，基于实验所使用的训练样本，迭代在进行到约 100 次的时候就已基本收敛，这与前面进行的基准测试实验相验证，证实了 OCLPSO 算法在收敛性上是有所改善的，可以得到更好的参数，使 LSSVM 获得更加优良的性能。在采用混合核函数对比测试中，本文提出的 OCLPSO 算法的检测正确率是 90.2%，CLPSO 的检测正确率 85.7%，OCLPSO 算法的检测正确率更高一些，证明了该算法的有效性。

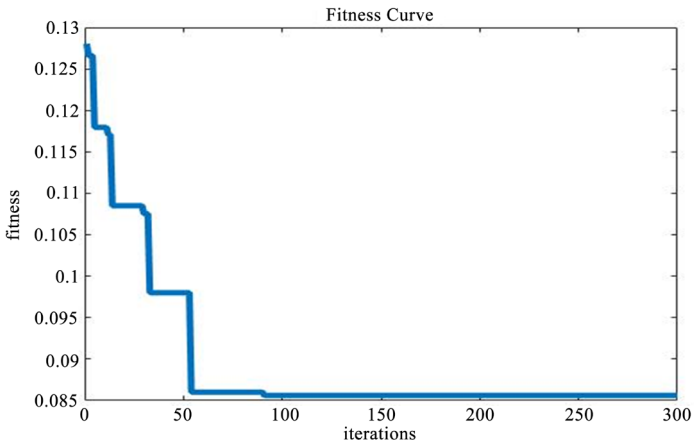

Figure 2. The curve of fitness function during OCLPSO evolution process
图 2. OCLPSO 进化过程中的适应度函数曲线

6. 结束语

基于机器学习的网络入侵检测技术随着机器学习、人工智能技术的发展而处于不断的发展之中。本

文通过引入反向学习机制，给出了一种基于 OCLPSO-LSSVM 的网络入侵检测模型，在粒子群算法和支持向量机的应用方面进行了探索，但是还存在许多问题有待后续进一步的研究探索：本文的数据是基于标准数据集 KDD CUP99 上进行的，而真实的网络环境要复杂的多，需要面向实际网络环境，提取网络特征和属性；网络入侵检测的实时性要求比较高，需要进一步研究网络实时入侵检测的方法和技术；另外在网络特征降维方面，需要对现有的特征降维方法进行更加深入的研究，寻找更好的方法来提高所提算法的性能。

参考文献

[1] 戴英侠, 连一峰, 王航. 系统安全与入侵检测[M]. 北京: 清华大学出版社, 2002.

[2] 崔宝才. 基于 GA 改进 BP 神经网络网络异常检测方法[J]. 现代电子技术, 2016, 39(3): 90-93.

[3] 张颖芳. 基于动态调整的 GA-SVM 多分类算法的研究及应用[D]: [硕士学位论文]. 广州: 华南理工大学, 2017.

[4] 黄芳, 樊晓平. 基于岛屿群体模型的并行粒子群优化算法[J]. 控制与决策, 2016, 21(2): 175-179+188.

[5] Tizhoosh, H.R. (2005) Opposition-Based Learning: A New Scheme for Machine Intelligence. *CIMCA-IAWTIC*06, Vienna, Austria, 28-30 November 2005.

[6] Liang, J.J., Qin, A.K., Suganthan, P.N. and Baskar, S. (2006) Comprehensive Learning Particle Swarm Optimizer for Global Optimization of Multimodal Functions. *IEEE Transactions on Evolutionary Computation*, **10**, 281-295. https://doi.org/10.1109/TEVC.2005.857610

[7] Yang, L., Yang, S., Li, S., *et al.* (2015) Coupled Compressed Sensing Inspired Sparse Spatial-Spectral LSSVM for Hyperspectral Image Classification. *Knowledge-Based Systems*, **79**, 80-89. https://doi.org/10.1016/j.knosys.2015.01.006

[8] 韩冰. 泥石流平均流速的 PSO-SVM 预测模型[J]. 河北地质大学学报, 2018, 41(2): 27-30.

[9] 陈晓康. 基于 Spark 云计算平台的改进 K 近邻算法研究[D]: [硕士学位论文]. 广州: 广东工业大学, 2016.