基于 PSOGWO-SVM 的网络入侵检测方法

陈　晨[1,2]，刘　曙[1]，王艺菲[1]，宋亚飞[1]，祝　彦[2]
(1. 空军工程大学防空反导学院，西安，710051；2. 西安卫星测控中心，西安，710043)

摘要　针对 SVM 算法的核函数及参数选择不科学会导致检测的准确率比较差的问题，提出了一种融合粒子群搜索的灰狼优化算法，利用 PSOGWO 算法优化 SVM 的参数，确定 SVM 分类器的最优检测模型，并基于 NSL-KDD 数据集进行对比实验。结果表明：基于 PSOGWO-SVM 的入侵检测方法实现了 SVM 的参数最优化，而且在检测率、收敛速度、模型平衡性等方面有明显提升，该方法在网络入侵检测方面具有更好的性能。

关键词　入侵检测；粒子群优化算法；灰狼优化算法；支持向量机；参数优化
DOI　10.3969/j.issn.1009-3516.2022.02.015
中图分类号　TP393.08　　**文献标志码**　A　　**文章编号**　1009-3516(2022)01-0097-09

A Network Intrusion Detection Method Based on PSOGWO-SVM

CHEN Chen[1,2], LIU Shu[1], WANG Yifei[1], SONG Yafei[1], ZHU Yan[2]
(1. Air Defense and Antimissile School, Air Force Engineering University, Xi'an 710051, China;
2. Xi'an Satellite Control Center, Xi'an 710043, China)

Abstract　Aimed at the problems that the selection of kernel function and parameter adjustment in SVM algorithm are not scientific, and the detection is poor in accuracy of classification, a Grey Wolf Optimization Algorithm based on Particle Swarm Optimization (PSOGWO) algorithm is proposed to improve the Intrusion Detection System (IDS) based on SVM. This method is to utilize PSOGWO algorithm for optimizing the parameters of SVM to improve the overall performance of intrusion detection based on SVM. The optimal detection model of SVM classifier is determined by the fusion of PSOGWO algorithm and SVM. The comparison experiments are made based on NSL-KDD dataset, and the results show that the intrusion detection method based on PSOGWO-SVM achieves the optimization of the parameters of SVM, improving significantly the detection rate, the convergence speed and the model balance. And this algorithm is feasible.

Key words　intrusion detection; particle swarm optimization algorithm; grey wolf optimization algorithm; support vector machine; parameter optimization

随着信息和网络技术的高速发展和日益普及，　网络安全已经成为一个全球严重关切的问题。目前

收稿日期：2021-08-20
基金项目：国家自然科学基金(61703426，61806219，61876189)；陕西省高校科协青年人才托举计划(20190108)；陕西省创新能力支撑计划(2020KJXX-065)
作者简介：陈　晨(1988—)，男，湖北郧县人，硕士生，研究方向为网络运维及安全防护、网络安全态势评估。E-mail：ssnw1899@qq.com

引用格式：陈晨，刘曙，王艺菲，等. 基于 PSOGWO-SVM 的网络入侵检测方法[J]. 空军工程大学学报(自然科学版)，2022，23(2)：97-105.
CHEN Chen, LIU Shu, WANG Yifei, et al. A Network Intrusion Detection Method Based on PSOGWO-SVM[J]. Journal of Air Force Engineering University (Natural Science Edition), 2022, 23(2): 97-105.

已有的安全防护方案有防火墙、数据加密、IDS 等技术。其中,IDS 是一种积极主动的安全防护技术[1],能够自动警告和主动拦截网络入侵,具有很高的实用价值。

为了区分攻击和正常网络访问,IDS 采用了不同的机器学习方法。SVM 是一种基于统计学习的机器学习算法,能够有效解决非线性、小样本、高维度等问题,越来越多的研究人员将其应用在入侵检测领域。核函数选择与参数调节直接影响了 SVM 算法的分类效果,传统 SVM 往往是随机选择相关参数,导致直接使用入侵检测数据训练 SVM 算法难以实现参数的优化,造成了多分类准确率不高。相关研究人员针对这个问题开展了一系列研究:文献[2~3]使用改进 PSO 算法对 SVM 的参数进行了优化,但是实验使用的 KDD99 数据集存在重复记录等许多固有问题,会对算法效能评估结果造成干扰;文献[4]设计了一种基于模糊 SVM 模型的入侵检测方法,但未对多分类入侵检测进行研究;文献[5]提出了用改进遗传算法优化基于 SVM 的 IDS,有效降低了入侵检测中恶意流量的特征维度,但收敛速度过慢;文献[6]中蚁群优化算法被用于优化 SVM 参数,但由于蚁群算法容易陷入局部最优,导致 SVM 的检测率下降。

基于此,本文提出一种具有更佳寻优能力和收敛速度的基于 PSOGWO-SVM 的网络入侵方法。

1 理论基础

1.1 PSOGWO 算法

GWO 算法是 Mirjalili 等提出的一种新型群体智能优化算法,它模仿了自然界中灰狼群体的社会等级和捕食行为,灰狼群体的社会等级由高到低分别为 α 狼、β 狼、δ 狼和 ω 狼,通过对狼群跟踪、包围、追捕、攻击猎物等过程进行模拟,重现灰狼群体捕食行为,从而实现优化[8]。

1)包围。狼群包围猎物可以表示为:

$$\begin{cases}\boldsymbol{D}=|\boldsymbol{C}\boldsymbol{X}_p(t)-\boldsymbol{X}(t)|\\ \boldsymbol{X}(t+1)=\boldsymbol{X}_p(t)-\boldsymbol{A}\boldsymbol{D}\end{cases} \tag{1}$$

式中:t 为当前迭代次数;$\boldsymbol{A}=2\boldsymbol{a}\boldsymbol{r}_1-\boldsymbol{a}$ 和 $\boldsymbol{C}=2\boldsymbol{r}_2$,为系数向量($\boldsymbol{a}$、$\boldsymbol{r}_1$ 和 $\boldsymbol{r}_2$ 均为常数);$\boldsymbol{X}_p(t)$ 表示猎物的位置向量;$\boldsymbol{X}$ 表示灰狼的位置向量。

2)追捕。狼群中 α 狼、β 狼、δ 狼距离猎物最近,据此可以先求出其他灰狼与这 3 只狼之间的距离,继而确定攻击猎物的最佳方向。

$$\boldsymbol{D}_k=|\boldsymbol{C}_i\boldsymbol{X}_k(t)-\boldsymbol{X}(t)|$$

$$\boldsymbol{X}_i=\boldsymbol{X}_k-\boldsymbol{A}_i\boldsymbol{D}_k$$

$$\boldsymbol{X}_p(t+1)=\frac{\boldsymbol{X}_1+\boldsymbol{X}_2+\boldsymbol{X}_3}{3} \tag{2}$$

式中:$k=\alpha、\beta、\delta$,$i=1,2,3$,$\boldsymbol{X}_p(t+1)$ 表示灰狼进化到第 $t+1$ 代时猎物的位置向量。

3)攻击。捕食行为的最后一步是攻击,该阶段狼群需要捕获猎物,即 GWO 算法获得最优解。攻击行为主要依据 $\boldsymbol{a}$ 从 2 递减到 0 来实现,即 $\boldsymbol{KA}$ 在 $[-2\boldsymbol{a},2\boldsymbol{a}]$ 中取任意值。若 $|\boldsymbol{A}|\leqslant 1$,灰狼群体对猎物集中攻击,找到最优解;若 $|\boldsymbol{A}|>1$,狼群在原位置散开,进行全局搜索。

为提高 GWO 算法的寻优能力和收敛速度[9],引入 PSO 算法中对粒子自身运动最佳位置信息进行记忆的方法,用粒子位置更新替代灰狼个体位置更新,使其能够记忆自身进化过程中的最佳位置信息。通过改变惯性常数 ω 来协调混合算法平衡全局搜索及局部开发的能力,ω 的变化范围为[0.5,1]。则速度和位置的更新为:

$$\begin{cases}\upsilon_i^{k+1}=\omega[\upsilon_i^k+c_1r_1(x_1-x_i^k)+\\ c_2r_2(x_2-x_i^k)+c_3r_3(x_3-x_i^k)]\\ x_i^{k+1}=x_i^k+\upsilon_i^{k+1}\end{cases} \tag{3}$$

式(2)中的第 1 式变为:

$$\boldsymbol{D}_k=|\boldsymbol{C}_i\boldsymbol{X}_k(t)-\omega\boldsymbol{X}(t)| \tag{4}$$

1.2 SVM 算法

SVM 是 Cortes 和 Vapnik 提出的一种基于统计学习理论和结构风险最小化原理的监督式机器学习算法,其基本思想是构造一个最优决策超平面,使得该平面两侧距离最近的两类样本之间的距离最大化[10-11]。在 SVM 无法分隔两个类的情况下,通过使用核函数将输入数据映射到高维特征空间中来解决此问题。在高维空间中,可以创建允许线性分离的超平面(对应于低维输入空间中的曲面)。此时,样本数据在高维空间中线性可分的条件可等价为以下最小化问题:

$$\min\left[\frac{1}{2}\|\boldsymbol{\omega}\|^2+C\sum_{i=1}^{N}(\xi_i+\xi_i^*)\right]$$

$$\text{s.t.}\begin{cases}y_i-[\boldsymbol{\omega}^{\mathrm{T}}\varphi(x_i)+b]\leqslant\varepsilon+\xi_i\\ \boldsymbol{\omega}^{\mathrm{T}}\varphi(x_i)+b-b_i\leqslant\varepsilon+\xi_i^*\\ \xi_i,\xi_i^*\geqslant 0,i=1,2,\cdots,N\end{cases} \tag{5}$$

式中:$\boldsymbol{\omega}$ 为权系数向量;C 为惩罚因子;ξ_i 和 ξ_i^* 为松弛变量;ε 为不灵敏参数。通过引入 Lagrange 乘子 a_i^* 以及满足 Mercer 定理的高斯核函数 $K(x_i,x_j)$,最优分类超平面函数可表述为:

$$f(x)=\operatorname{sgn}\left[\sum_{i=1}^{N}a_i^*y_iK(x_i,x_j)+b^*\right] \tag{6}$$

核函数的选择是影响分类模型效果的一个关键,在线性、多项式、高斯等众多核函数中,高斯核函

数能够实现非线性映射，同时仅有 2 个参数 C 和 g 可以适当降低模型复杂度。通过调节核函数参数 g，可实现线性与非线性两种分类器，具有很好的灵活性，因而应用最为广泛[11]。考虑到网络入侵特征向量与入侵行为类型之间存在一定的随机性和非线性，为此本文将采用高斯核(RBF)：

$$\begin{cases} K(x_i,x_y)=\exp\left(-\dfrac{||x_i-x_j||^2}{2\sigma^2}\right) \\ g=\dfrac{1}{2\sigma^2} \end{cases} \tag{7}$$

2　PSOGWO-SVM 算法

利用 PSOGWO 方法寻优能力强、收敛速度快，不易陷入局部最优的特点，对 SVM 中的参数 C 和 g 进行寻优，来确定 SVM 模型，进而提出一种基于 PSOGWO 算法优化 SVM(PSOGWO-SVM)的网络入侵检测模型。该模型的具体构建过程如下：

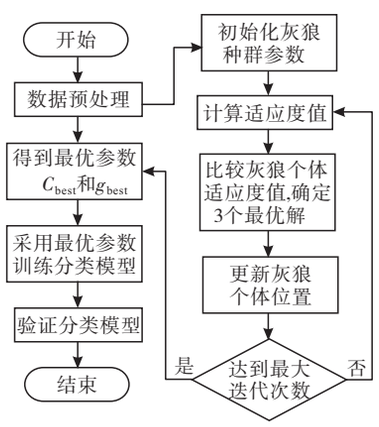


图 1　PSOGWO-SVM 网络入侵检测模型的工作流程

1)抽取 NSL-KDD 数据集中的部分数据，分别获得训练数据集和测试数据集，并对数据集进行归一化处理。

2)设定狼群规模与最大迭代次数，将 SVM 中的参数 C 和 g 设定为狼群个体位置的二维坐标，随机初始化 C 和 g。

3)通过训练集样本对 SVM 模型进行训练，以计算适应度值。狼的等级(α 狼、β 狼、δ 狼和 ω 狼)根据适应度值进行分类。计算适应度值为：

$$F=1-\frac{y_t}{y_t+y_f} \tag{8}$$

式中：y_t 是正确分类个数；y_f 是错误分类个数。

4)根据公式(3)更新灰狼的位置。

5)计算更新后的个体的适应度值，将当前代数的最优适应度值记录为 F_{best}。如果 $F_{best}>F_\alpha$(α 狼的适应度)，则 F_α 被更新为 F_{best}，并记录相应的位置；如果 $F_\beta<F_{best}<F_\alpha$，则将 F_{best} 赋值给 β 狼，相应的位置更新为 β 狼的位置；若 $F_\delta<F_{best}<F_\beta$，则将 F_{best} 赋值给 δ 狼，对应位置也更新为 δ 狼的位置。显然，迭代到当前代数时，α 狼的位置就是种群的最优位置。

6)判断迭代次数是否达到最大迭代次数。如果达到，则循环终止，并得到最佳参数 C_{best} 和 g_{best}；否则返回到步骤 4)继续迭代。

7)使用 C_{best} 和 g_{best} 建立 PSOGWO-SVM 模型。

8)使用 PSOGWO-SVM 模型在测试数据集上进行验证，并对模型的预测结果进行性能评价。

3　实验分析

3.1　实验环境

硬件配置：CPU 为 AMD R5-4600H 3 GHz，内存为 16 GiB，硬盘为 512 GB；操作系统：Windows10；测试平台：Matlab R2016b。

3.2　实验数据

1)实验 1：选择 13 个基准函数做测试，其中 $F_1\sim F_7$ 是单峰基准函数，$F_8\sim F_{13}$ 是多模态基准函数，见表 1。

表 1　13 个基准函数

Function	D	range	optimal
$F_1(x)=\sum_{i=1}^{n}x_i^2$	30 100	[−100,100]	0
$F_2(x)=\sum_{i=1}^{n}\lvert x_i\rvert+\prod_{i=1}^{n}\lvert x_i\rvert$	30 100	[−10,10]	0
$F_3(x)=\sum_{i=1}^{n}\left(\sum_{j=1}^{n}x_j\right)^2$	30 100	[−100,100]	0
$F_4(x)=\max\{\lvert x_i\rvert,1\leqslant i\leqslant n\}$	30 100	[−100,100]	0

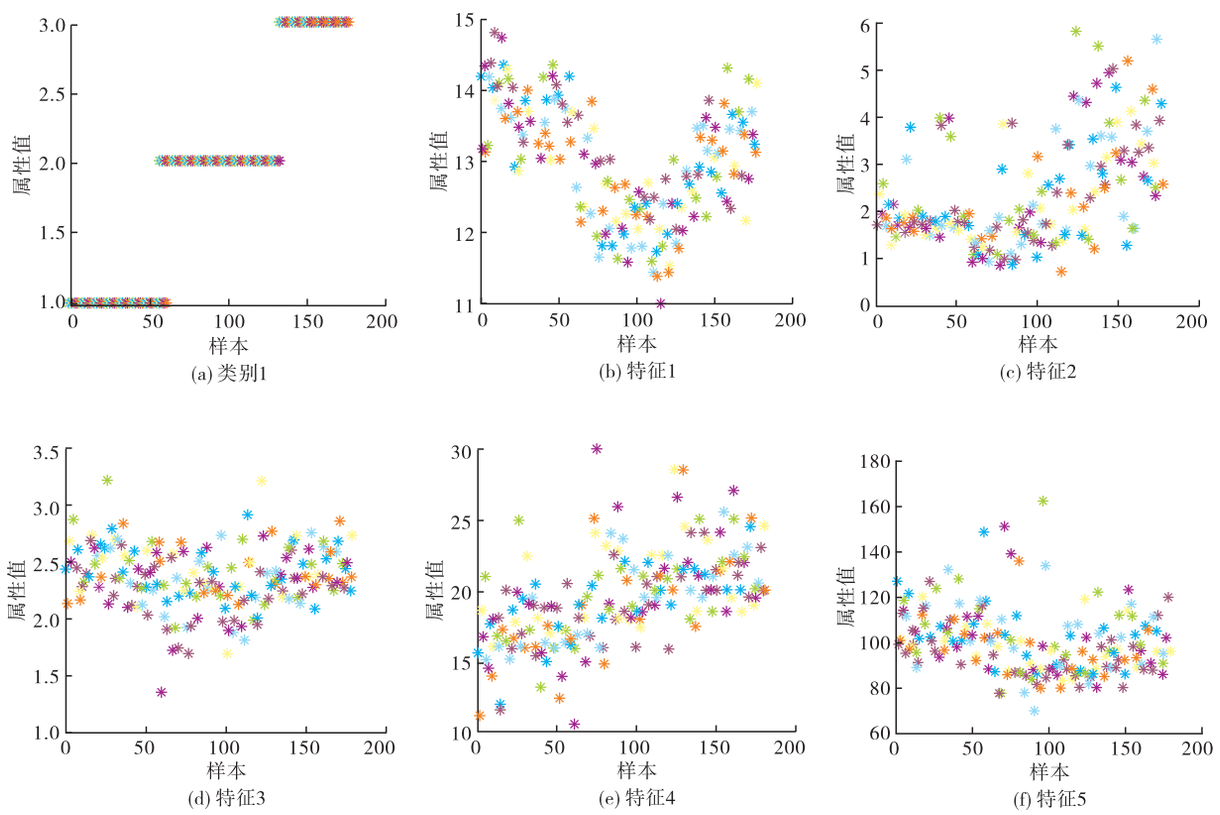
续表 1

Function	D	range	optimal
$F_5(x)=\sum_{i=1}^{n-1}[100(x_{i+1}-x_i^2)^2+(x_i-1)^2]$	30 100	$[-30,30]$	0
$F_6(x)=\sum_{i=1}^{n}(\lvert x_i+0.5\rvert)^2$	30 100	$[-100,100]$	0
$F_7(x)=\sum_{i=1}^{n}x_i^4+\mathrm{random}[0,1]$	30 100	$[-1.28,1.28]$	0
$F_8(x)=\sum_{i=1}^{n}-x_i\sin(\sqrt{\lvert x_i\rvert})$	30 100	$[-500,500]$	$-418.982\ 9n$
$F_9(x)=\sum_{i=1}^{n}[x_i^2-10\cos(2\pi x_i)+10]$	30 100	$[-5.12,5.12]$	0
$F_{10}(x)=-20\exp\left(-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n}x_i^2}\right)-\exp\left(\frac{1}{n}\sum_{i=1}^{n}\cos(2\pi x_i)\right)+20+e$	30 100	$[-32,32]$	0
$F_{11}(x)=\frac{1}{4\ 000}\sum_{i=1}^{n}x_i^2-\prod_{i=1}^{n}\cos\left(\frac{x_i}{\sqrt{i}}\right)+1$	30 100	$[-600,600]$	0
$F_{12}(x)=\frac{\pi}{n}\left\{10\sin^2(\pi y_1)+\sum_{i=1}^{n-1}(y_i-1)^2[1+10\sin^2(\pi y_{y+1})]\right\}+(y_n-1)^2$	30 100	$[-50,50]$	0
$F_{13}(x)=0.1\left\{\sin^2(3\pi x_i)+\sum_{i=1}^{n}(x_i-1)^2[1+10\sin^2(3\pi x_i+1)]+(x_n-1)^2\cdot[1+\sin^2(2\pi x_n)]\right\}+\sum_{i=1}^{n}u(x_i,5,100,4)$	30 100	$[-50,50]$	0

2)实验 2:UCI 数据集详细信息见表 2。wine 数据集是 UCI 数据集中的一个标准测试数据集,能够很好测试各算法的多分类效果。图 2 展示了 wine 数据集中 13 个属性特征的分布情况。

表 2 UCI 数据集

数据集	维度	类别数	样本总数	测试样本数
wine	13	3	178	89



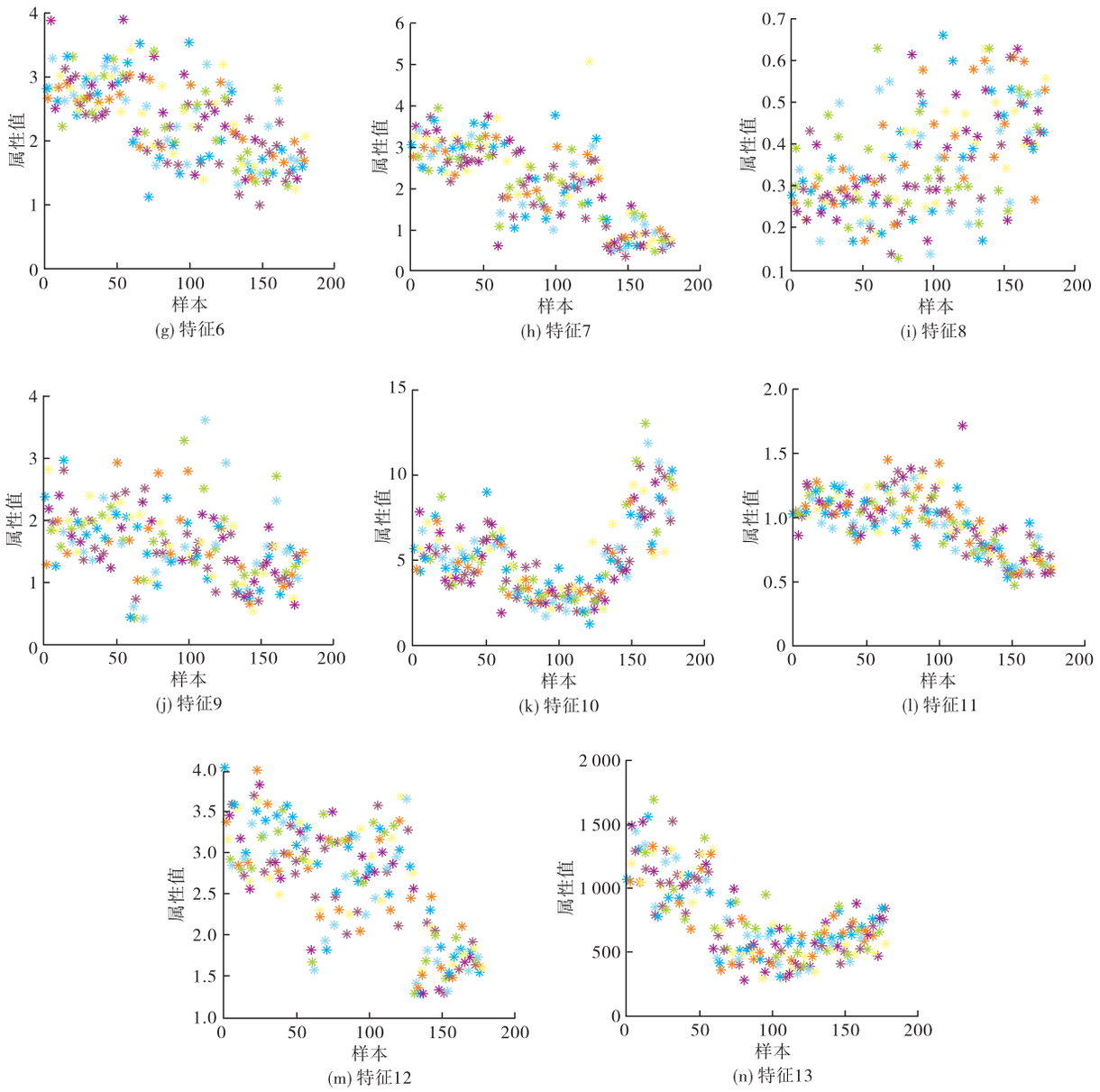


图 2　wine 数据集中 13 个属性特征的箱线图和分布情况

3）实验 3：NSL-KDD 数据集中抽取部分数据，类型构成如表 3 所示。

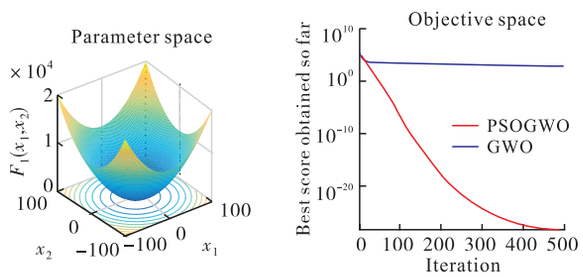
表 3　NSL-KDD 数据类型构成

分类	Normal	Probe	DoS	U2R	R2L	总数
训练集	2 580	330	920	52	118	4 000
测试集	592	120	228	43	17	1 000

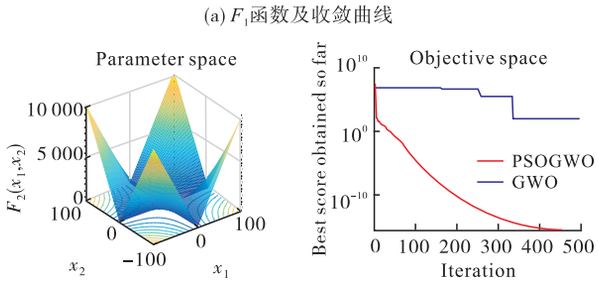
3.3　结果与分析

3.3.1　实验 1：PSOGWO 寻优效果实验

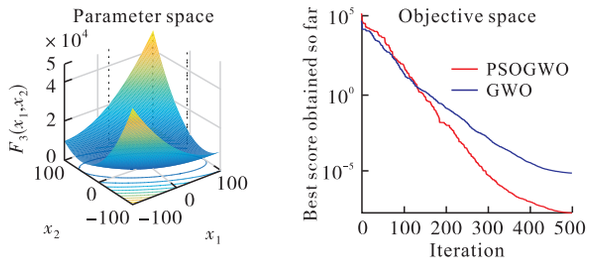
图 3 为 13 个基准函数的图像及其在 30 维度时各算法的收敛曲线。在 PSOGWO 与 GWO 算法中，种群数量为 30，最大迭代次数为 500 次。从图中可以看出，PSOGWO 的寻优效果均比 GWO 的寻优效果好：①当两种算法都寻不到最优值时，PSOGWO 的值更接近最优值；②当二者都可以寻到最优值时，例如 F_5 函数，PSOGWO 的收敛速度更快。



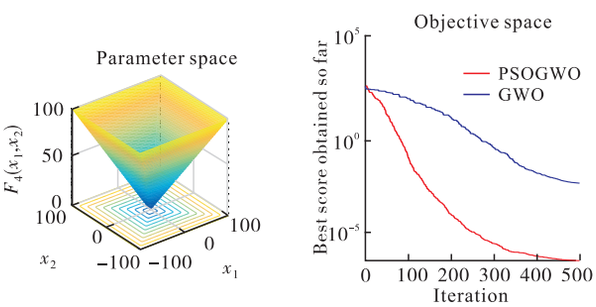
(a) F_1函数及收敛曲线



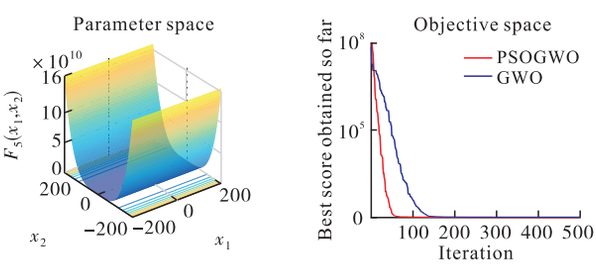
(b) F_2函数及收敛曲线



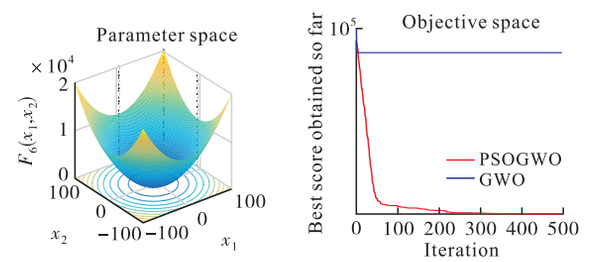
(c) F_3函数及收敛曲线



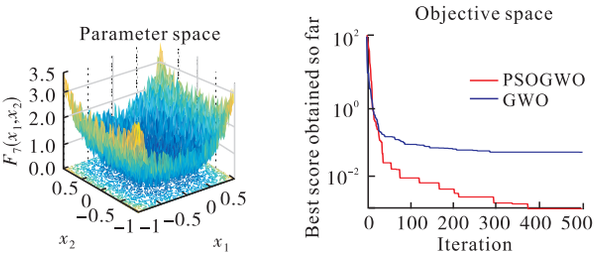
(d) F_4函数及收敛曲线



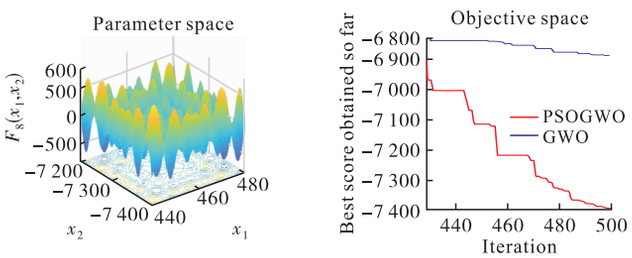
(e) F_5函数及收敛曲线



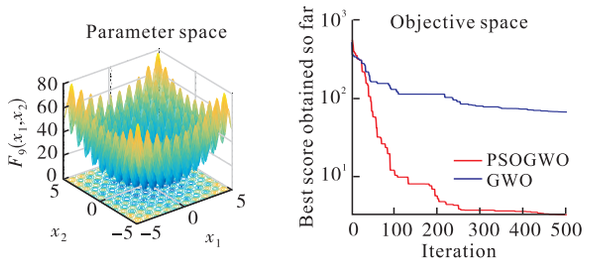
(f) F_6函数及收敛曲线



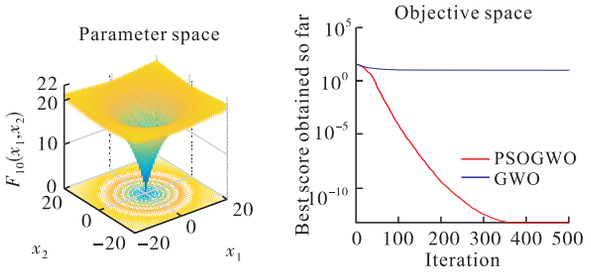
(g) F_7函数及收敛曲线



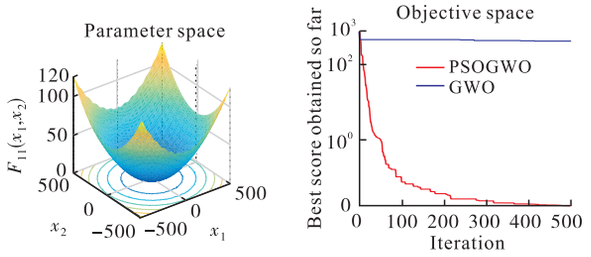
(h) F_8函数及收敛曲线



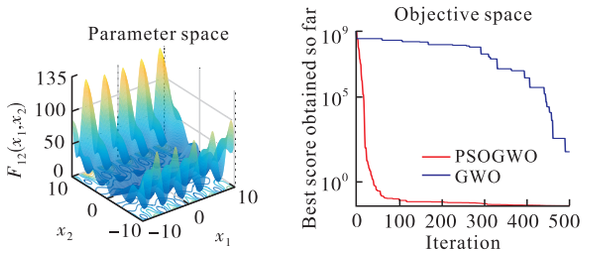
(i) F_9函数及收敛曲线



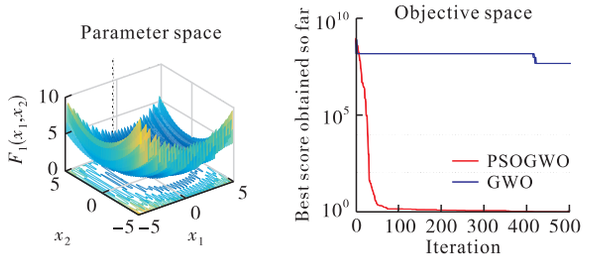
(j) F_{10}函数及收敛曲线



(k) F_{11}函数及收敛曲线



(l) F_{12}函数及收敛曲线



(m) F_{13}函数及收敛曲线

图 3 基准函数图和收敛曲线图

3.3.2 实验 2:UCI 数据集实验

为验证所提算法在真实分类数据集上的有效性,在 UCI 数据集上将 PSOGWO-SVM 与 SVM、GWO-SVM 作对比。在 SVM 算法中,$C=10$、$g=30$;在 PSOGWO-SVM 与 GWO-SVM 算法中,种群数量为 20,最大迭代次数为 200 次,C、g 取值范围为[0.01,100],得到的最优参数 C、g 见表 4。

表 4　各算法准确率

算法	C_{best}	g_{best}	准确率/%
SVM	10	30	49.438 2
GWO-SVM	6.129 9	3.036	98.876 4
PSOGWO-SVM	11.468 8	1.633 1	100

在表 4 中列出了各算法的准确率，PSOGWO-SVM 的准确率最高，可以达到 100%。

3.3.3　实验 3：NSL-KDD 数据集实验

将 Normal、Probe、Dos、U2R、R2L 等 5 类数据类型分别记为 1、2、3、4、5，依次进行分类对比，分类结果对比见表 5～10 和图 4，并在图 5 中给出了混淆矩阵。在 SVM 算法中，$C=10$，$g=30$；在 PSOGWO-SVM 与 GWO-SVM 算法中，种群数量为 20，最大迭代次数为 200 次，C、g 取值范围为[0.01，100]，得到的最优参数 C、g 见表 5。

表 5　各算法的关键参数值和准确率

算法	C_{best}	g_{best}	准确率/%
SVM	10	30	87.1
GWO-SVM	31.243 1	1.036 4	90.4
PSOGWO-SVM	38.304 8	0.017 727 3	94.4

表 6　各算法在 Normal 上的性能评价指标

算法	Precision/%	TPR/%	FPR/%	F 值/%	AUC
SVM	82.77	99.83	30.15	90.51	0.848 4
GWO-SVM	93.99	97.80	9.07	95.86	0.943 7
PSOGWO-SVM	93.03	99.16	10.78	95.99	0.941 9

表 7　各算法在 Probe 上的性能评价指标

算法	Precision/%	TPR/%	FPR/%	F 值/%	AUC
SVM	100.00	42.50	26.59	59.65	0.579 5
GWO-SVM	90.91	66.67	29.89	76.92	0.683 9
PSOGWO-SVM	93.22	91.67	29.09	92.44	0.812 9

表 8　各算法在 Dos 上的性能评价指标

算法	Precision/%	TPR/%	FPR/%	F 值/%	AUC
SVM	100	100	6.61	100	0.967 0
GWO-SVM	100	100	14.38	100	0.928 1
PSOGWO-SVM	100	100	14.25	100	0.928 8

表 9　各算法在 U2R 上的性能评价指标

算法	Precision/%	TPR/%	FPR/%	F 值/%	AUC
SVM	50.00	2.33	30.15	44.44	0.511 1
GWO-SVM	62.96	39.53	0.94	48.57	0.693 0
PSOGWO-SVM	90.48	44.19	14.25	59.38	0.719 9

表 10　各算法在 R2L 上的性能评价指标

算法	Precision/%	TPR/%	FPR/%	F 值/%	AUC
SVM	0	0	0	NaN	0.500 0
GWO-SVM	0	0	0.10	NaN	0.500 0
PSOGWO-SVM	0	0	0.10	NaN	0.500 0

PSOGWO-SVM 的准确率为 94.4%，与其他算法比较是最高的。由于 NSL-KDD 数据集为非平衡数据集，本文根据精确率、TPR、FPR、F 值和 AUC 等 5 个评价指标进一步衡量各算法的入侵检测能力。

从表 10 中可以看出，3 种算法对于 R2L 数据类型的检测效果均不理想，是因为 R2L 数据类型属于入侵行为的稀有类[12]，KDD99 竞赛中的冠军对于稀有类的入侵检测效果也很不理想。故本文主要对 Normal、Probe、Dos 和 U2R 共 4 类数据类型的分类效果进行对比。

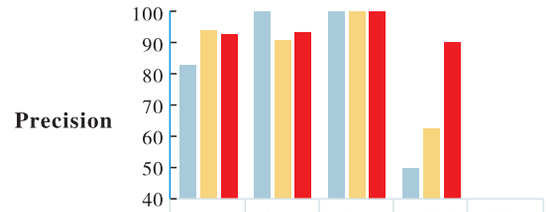
从图 4(a)中可以看出，PSOGWO-SVM 的精确率均高于 90%，而其他 2 种算法对于 U2R 数据类型的精确率均低于 63%，说明 PSOGWO-SVM 较其他 2 种算法，在 4 类数据类型上都能够保持很高的精确率，对数据集具有更广泛的适用性。

从图 4(b)中可以看出，3 种算法在 Normal 数据类型上的 TPR 基本持平，在 Probe、Dos、U2R 等 3 类数据类型上，PSOGWO-SVM 的 TPR 均为最高，因此综合表现更好。

从图 4(c)中可以看出，对于 Normal、Probe、Dos 等 3 类数据类型，PSOGWO-SVM 与 GWO-SVM 的 FPR 基本持平，同时远远低于 SVM 的 FPR；PSOGWO-SVM 在 U2R 数据类型上面的 FPR 要高于 GWO-SVM，说明 PSOGWO-SVM 在 FPR 指标上的表现要略低于 GWO-SVM。

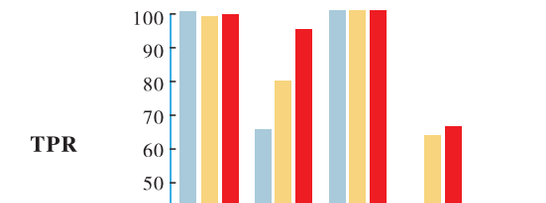
从图 4(d)中可以看出，PSOGWO-SVM 对于 Normal、Probe、Dos 和 U2R 等 4 类数据类型的 F 值均为最高值，说明 PSOGWO-SVM 的模型平衡性最优。

从图 4(e)中可以看出，PSOGWO-SVM 对于 Probe、U2R 等 2 类的 AUC 均为最高值，而对于 Normal、Dos 等 2 类数据类型的 AUC 与最高值相差无几。将 4 类综合起来看，PSOGWO-SVM 的 AUC 总能保持最高或者近似最高的数值，而其他 2 种算法的 AUC 在不同数据类型上的表现有明显起伏，故 PSOGWO-SVM 在多分类效能上具有最优的表现。



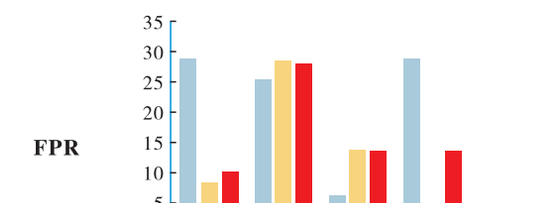
	Normal	Probe	Dos	U2R	R2L
SVM	82.77	100	100	50	0
GWO-SVM	93.99	90.91	100	62.96	0
PSOGWO-SVM	93.03	93.22	100	90.48	0

(a) 各算法的Precision



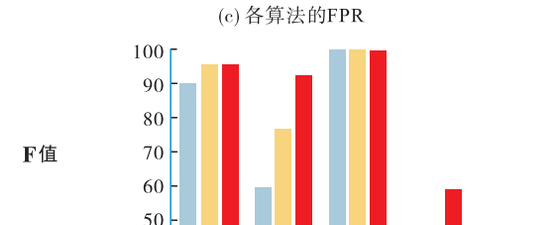
	Normal	Probe	Dos	U2R	R2L
SVM	99.83	42.5	100	2.33	0
GWO-SVM	97.8	66.67	100	39.53	0
PSOGWO-SVM	99.16	91.67	100	44.19	0

(b) 各算法的TPR



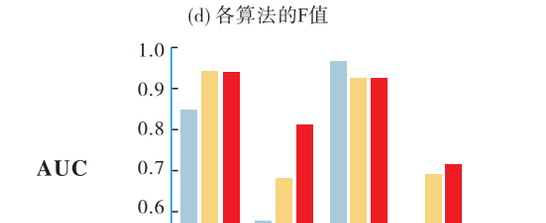
	Normal	Probe	Dos	U2R	R2L
SVM	30.15	26.59	6.61	30.15	0
GWO-SVM	9.07	29.89	14.38	0.94	0.1
PSOGWO-SVM	10.78	29.09	14.25	14.25	0.1

(c) 各算法的FPR



	Normal	Probe	Dos	U2R	R2L
SVM	90.51	59.65	100	44.44	0
GWO-SVM	95.86	76.92	100	48.57	0
PSOGWO-SVM	95.99	92.44	100	59.38	0

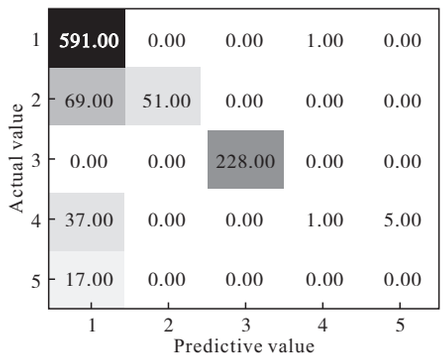
(d) 各算法的F值



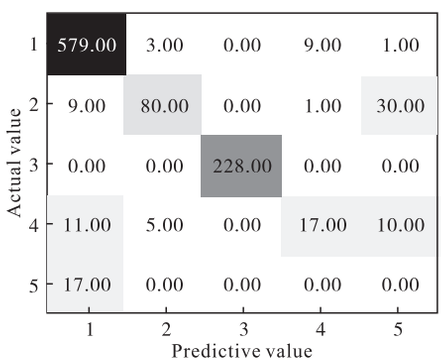
	Normal	Probe	Dos	U2R	R2L
SVM	0.848 4	0.579 5	0.967	0.511 1	0.5
GWO-SVM	0.943 7	0.683 9	0.928 1	0.693	0.5
PSOGWO-SVM	0.941 9	0.812 9	0.928 8	0.719 9	0.5

(e) 各算法的AUC

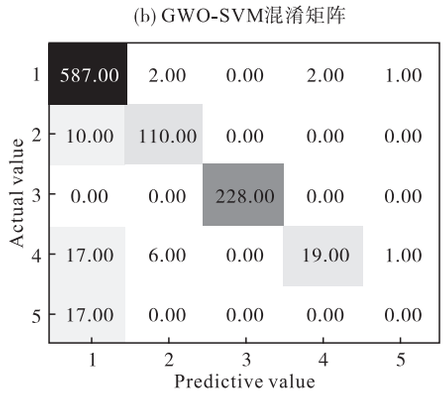
图 4 分类结果对比



(a) SVM混淆矩阵



(b) GWO-SVM混淆矩阵



(c) PSOGWO-SVM混淆矩阵

图 5 各算法混淆矩阵

从图 5(b)、5(c)中可以看出,GWO-SVM 和 PSOGWO-SVM 算法的前 3 种分类的深色区域都集中在对角线上,但是 PSOGWO-SVM 算法对于第 2 类的检测率明显高于其他 2 种算法;对于第 4、5 种分类,3 种算法的结果都不是很明显,一是因为后 2 类属于入侵行为的稀有类,二是与 Dos 和 Probe 相比较,U2R 和 R2L 的数据记录中不具有频繁的序列模式[13]。故 PSOGWO-SVM 具有更好的分类检测效果。

4 结语

为了提高入侵检测的准确性,提出了基于 PSOGWO-SVM 算法的检测模型。首先使用 PSO 算法提升 GWO 算法的寻优能力和收敛速度,然后使用 PSOGWO 算法优化 SVM 的相关参数。最后,本文将其与 GWO-SVM 和 SVM 在 NSL-KDD 数

据集上进行多分类比较实验。实验结果证明，该方法显著提高了入侵检测的正确率。在准确度、精确度、TPR、FPR、F 值和 AUC 等方面的综合表现要优于原始 SVM 和 GWO-SVM，具有更好的检测性能。

由于 PSOGWO-SVM 算法中参数的初始值对寻优能力和收敛速度有很大影响，因此后期工作将研究如何产生合适的参数初始值，从而提高 PSOGWO-SVM 的分类性能。

参考文献

[1] FERRAG M A, MAGLARAS L, MOSCHOYIANNIS S, et al. Deep Learning for Cyber Security Intrusion Detection: Approaches, Datasets, and Comparative Study[J]. Journal of Information Security and Applications, 2020, 50: 102419.

[2] ALSAADI H I, ALMUTTAIRI R M, BAYAT O, et al. Computational Intelligence Algorithms to Handle Dimensionality Reduction for Enhancing Intrusion Detection System[J]. Journal of Information Science and Engineering, 2020, 36: 293-308.

[3] 柯钢. 改进粒子群算法优化支持向量机的入侵检测方法[J]. 合肥工业大学学报(自然科学版), 2019, 42(10): 1341-1345.

[4] 汪生，金志刚. 基于模糊 SVM 模型的入侵检测分类算法[J]. 计算机应用研究, 2020, 37(2): 501-504.

[5] HALIM Z, YOUSAF M N, WAQAS M, et al. An Effective Genetic Algorithm-Based Feature Selection Method for Intrusion Detection Systems[J]. Computers & Security, 2021, 110: 102448.

[6] 刘静，杨正校. 改进 ACO-SVM 在网络入侵检测中的应用[J]. 软件, 2018, 39(10): 57-59.

[7] DAVAHLI A, MAHHOUBEH S, GOLNOUSH A. Hybridizing Genetic Algorithm and Grey Wolf Optimizer to Advance an Intelligent and Lightweight Intrusion Detection System for IoT Wireless Networks[J]. Journal of Ambient Intelligence and Humanized Computing, 2020, 11(11): 5581-5609.

[8] ALMOMANI O. A Feature Selection Model for Network Intrusion Detection System Based on PSO, GWO, FFA and GA algorithms[J]. Symmetry, 2020, 12(6): 1046.

[9] 曹源，高丙朋，张振海. 一种基于 PSO-GWO 的电网故障诊断方法[J]. 电测与仪表, 2021, 58(9): 35-40.

[10] IOANNOU C, VASOS V. Network Attack Classification in IoT Using Support Vector Machines[J]. Journal of Sensor and Actuator Networks, 2021, 10(3): 58.

[11] HU J J. Network Security Situation Prediction Based on MR-SVM[J]. IEEE Access, 2019, 7: 130937-130945.

[12] 杨越翔. 基于日志数据的 U2R 和 R2L 入侵检测研究[D]. 兰州：西北师范大学, 2020.

[13] 全亮亮. 基于数据挖掘算法的入侵检测研究[D]. 武汉：武汉科技大学, 2013.

（编辑：徐敏）

(上接第 96 页)

[10] SUPRUN O, YUDIN O, ZIUBINA R, et al. Devising a Method of Protection Against Zero-Day Attacks Based on an Analytical Model of Changing the State of the Network Sandbox[J]. Eastern-European Journal of Enterprise Technologies, 2021, 109: 50-57.

[11] MILLAR S, MCLAUGHLIN N, RINCON J M D, et al. Multi-View Deep Learning for Zero-Day Android Malware Detection[J]. Journal of Information Security and Applications, 2021, 58(3): 102718.

[12] 白鹏. 文档类型 0 Day 漏洞检测技术的研究与实现[D]. 北京：北京邮电大学, 2015.

[13] JEONG J H, CHOI S G. Hybrid System to Minimize Damage by Zero-Day Attack based on NIDPS and Honey Pot[C]// 2020 International Conference on Information and Communication Technology Convergence. Nanjing: [s. n.], 2020.

[14] 王刚，冯云，马润年. 操作系统病毒时滞传播模型及抑制策略设计[J]. 西安交通大学学报, 2021, 55(3): 11-19.

[15] 张志双. 一类具有双时滞计算机病毒传播模型的动力学性质[D]. 哈尔滨：哈尔滨工业大学, 2013.

[16] YANG F, ZHANG Z. Hopf Bifurcation Analysis of SEIR-KS Computer Virus Spreading Model with Two-Delay[J]. Results in Physics, 2021, 24: 104090.

[17] 王超，杨旭颖，徐珂，等. 基于 SEIR 的社交网络信息传播模型[J]. 电子学报, 2014, 42(11): 2325-2330.

[18] 宋磊，王春雷. 一类时滞网络病毒传播模型的 Hopf 分支[J]. 计算机与现代化, 2016(11): 79-82.

[19] 张子振，储煜桂，KUMARI S，等. 一类具有非线性发生率的无线传感网络蠕虫传播模型的延迟动力学行为[J]. 浙江大学学报：理学版, 2019, 46(2): 168-186, 199.

（编辑：徐楠楠）