文献引用格式： 周利均 . 基于竞争机制粒子优化算法选择 NIDS 系统特征子集 [J]. 通信技术，2020，53（08）：2006-2013.

ZHOU Li-jun.Competitive Particle Optimization Algorithm-based Feature Subset Selection of NIDS System[J].Communications Technology,2020,53(08):2006-2013.

doi:10.3969/j.issn.1002-0802.2020.08.028

基于竞争机制粒子优化算法选择 NIDS 系统特征子集 *

周利均

（中国电子科技网络信息安全有限公司，四川 成都 610041）

摘　要： 网络入侵检测系统通过分析网络连接数据发现入侵行为，网络流量数据中包含了大量冗余、不相关特征，严重影响了 NIDS 系统训练模型精度。因此，提取有用特征搭建检测模型，对实现动态防御至关重要。针对特征选择问题，研究竞争机制粒子优化算法在海量数据中的特征选择问题，通过稳定性理论对优化模型进行稳定性分析，利用 S- 函数对更新后的速度进行二值化处理，以分类错误率和特征数的线性组合为适应度函数，通过优化算法选择特征子集构建分类器，对 KDDCUP[99] 数据集中的训练集（10%）进行数据预处理，采用交叉验证方法验证分类模型，并采用循环迭代法获取最佳控制参数。测试结果表明，竞争机制粒子优化算法能够降低数据维度，获取特征子集，使得搭建的训练模型具有较高精度。

关键词： 机器学习；二值化处理；竞争机制；粒子优化算法；数据预处理；特征选择

中图分类号： TN918　　**文献标志码：** A　　**文章编号：** 1002-0802(2020)-08-2006-08

Competitive Particle Optimization Algorithm-based Feature Subset Selection of NIDS System

ZHOU Li-jun

(China Electronics Technology Cyber Security Co. Ltd., Chengdu Sichuan 610041, China)

Abstract: The network intrusion detection system discovers intrusions by analyzing network connection data. The network traffic data contains a large number of redundant and irrelevant features, which seriously affects the accuracy of the NIDS system training model. Therefore, it is very important to extract useful features and build detection model for realizing dynamic defense. Aiming at the feature selection problem, this paper discusses the feature selection of the competitive mechanism particle optimization algorithm in massive data. The stability analysis of the optimization model is carried out through stability theory, and the S-function is used to binarize the updated speed, the linear combination of the classification error rate and the number of features is used as the fitness function. The classifier is constructed by selecting feature subset through the optimization algorithm, and data preprocessing is done on the training set (10%) in KDDCUP99 data set. A cross-validation method is used to verify the classification model, and a loop iteration method is used to obtain the best control parameters. The test results indicate that the particle optimization algorithm of the competition mechanism can reduce the data dimension and obtain the feature subset, thus making the established training model have fairly high precision.

* **收稿日期：** 2020-04-20；**修回日期：** 2020-07-08　Received date:2020-04-20;Revised date:2020-07-08
通讯联系人： 631441191@qq.com　Corresponding author: 631441191@qq.com

Key words: machine learning; binary processing; competition mechanism; particle optimization algorithm; data preprocessing; feature selection

0 引　言

入侵检测系统是计算机系统的关键组成部分，一般分为外部异常检测和内部误用检测，近年已有相关的系统研究和综合性论述[1−5]。异常检测核心是通过训练数据搭建分类模型来实时分析检测网络入侵行为，重点是检测的实时性、准确性以及兼容性等。网络入侵检测系统通过分析 TCP/IP 数据流来挖掘分析潜在的异常攻击行为，在万物互联的网络时代显得尤为重要。攻击者攻击手段已经向着自动化、智能化方向发展[6−7]。面对攻击手段更加隐蔽、攻击方式更加多样的情况，本文研究在构建 NIDS 系统分类器模型中的特征优化选择问题，以便适应海量网络流量数据的挖掘分析。

构建 NIDS 系统分类器核心是选择较为合适的特征子集[8−14]，具有 D 个特征的流量数据，会有 2^D-1 种不同的特征子集组合。合适的特征子集既可以消除冗余信息，也可以降低计算的复杂度[15]。根据评价方式，特征选择主要分为 3 种[16]。

（1）过滤式（Filter）：先选择特征再训练分类器，主要有卡方检验、信息增益以及相关系数法等。

（2）包裹式（Wrapper）：分类器作为分类评价函数选择最优子集，主要有 GA、PSO、DE 以及 ABC 等算法。

（3）嵌入式（Embedding）：特征选择与分类器学习融合，主要有正则化方法等。

以上 3 种方式比较，Wrapper 方法具有更好的分类性能[17]，因此启发式优化算法在该领域有广泛应用。GA 算法通过生产染色体种群解，通过每一代的选择、交叉和突变完成进化，已有相关特征选择研究[18−19]，在大数据等特征选择领域也有相关研究[20−23]。人工蜂群算法通过初始化蜂群，采用不同角色之间的信息交流、转换以及协作来实现优化，该算法也在特征选择有相关研究[24−25]。以上方法的参数设置对优化结果有较大影响。

粒子群优化算法通过初始化粒子种群，每个粒子在搜索空间中独立搜索其局部最优解，速度代表向最优解靠近的快慢，用位置代表移动方向，通过记忆局部最优并采取全局共享机制获取全局最优解，所有粒子根据其局部最优和当前全局最优更新其位置和速度。PSO 算法用于特征选择问题已有相关研究[26−29]，但是 PSO 因其采用的经验学习机制，存储的个体最优极易使得算法陷入局部最优解，且针对海量数据时会消耗大量的计算资源。

基于 PSO 算法的思想，提出了一种基于竞争机制的粒子优化算法。该算法随机选取粒子对竞争，竞争中的优胜者直接进入下一代，失败者向优胜者学习。学习机制舍弃了对局部最优和当前全局最优解的依赖，因此相对 PSO 算法不易陷入局部最优解，且每次迭代只有一半的粒子向优胜粒子学习，因此在解决大规模数据优化方面更具有优势。本文以 KDDCUP^{99} 的训练集（10%）和测试集[30]来对算法进行验证，利用错误率和选择特征数的线性组合为适应度函数，准确率更重要，权重值设置更大，错误率来源于分类结果，采用 KNN 等分类算法计算。

本文共分为 5 个部分：第 1 部分为序言，主要回顾该方向前期研究成果以及本文研究思路；第 2 部分主要介绍竞争机制粒子优化算法原理，并对算法模型进行稳定性分析，求得控制参数取值范围，再分析了二值化方法的更新函数及流程；第 3 部分主要分析 KDDCUP^{99} 数据集，并研究该数据集的预处理方法及流程；第 4 部分是实验设计及验证，利用提出的优化算法选择合适的特征子集，利用选择的特征子集搭建训练模型，采用交叉验证方法的训练模型的检测准确度进行验证；第 5 部分是结论，根据本文提出的算法模型仿真得出的实验结果得出结论，结合该方向分析和展望后续研究重点。

1 竞争机制粒子优化算法

1.1 竞争机制粒子优化算法描述

竞争机制粒子群优化算法框架见图 1，其中 $P(t)$ 和 $P(t+1)$ 分别代表第 t 和 $t+1$ 代粒子群，每一代粒子群共有 m 个粒子，每个粒子位置和速度为 n 维，第 i 个粒子位置和速度分别表示为：

$$X_i(t)=[x_{i,1}(t),x_{i,2}(t),\cdots,x_{i,n}(t)] \quad (1)$$

$$V_i(t)=[v_{i,1}(t),v_{i,2}(t),\cdots,v_{i,n}(t)] \quad (2)$$

在每一次迭代过程中，粒子群 $P(t)$ 被随机分成种群大小为 $m/2$ 的两部分（若 m 为奇数，随机划分时竞争优胜者比失败者多 1 个粒子，从优胜者子群里随机选取 1 个补充到失败者子群），从两个子群中分别选择粒子成对竞争，优胜者直接进入下一次迭代，失败者通过学习机制向优胜者学习更新，再

进入下一代，即每一次迭代每个粒子只参与 1 次竞争，只有 $m/2$ 的粒子通过学习机制更新。可见，该算法计算消耗主要是在失败粒子的更新上，计算复杂度为 $O(mn)$。

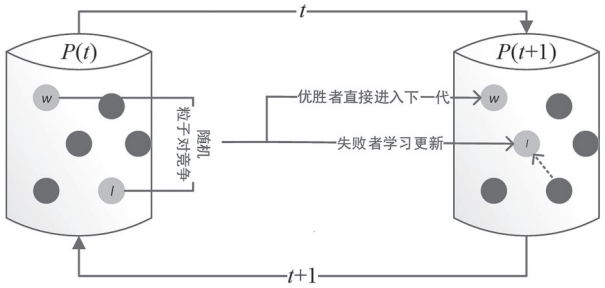


图 1 竞争机制粒子优化算法框架

令 $X_{l,k}(t)$ 和 $V_{l,k}(t)$ 表示第 t 次迭代中第 k 轮竞争中失败粒子的位置和速度；$X_{w,k}(t)$ 和 $V_{w,k}(t)$ 表示第 t 次迭代中第 k 轮竞争中优胜粒子的位置和速度；$R_1(k,t),R_2(k,t),R_3(k,t) \in [0,1]^n$ 表示第 t 次迭代中，第 k 轮竞争中失败者向优胜者学习的随机数向量，其中 k=1,2,3,⋯,m/2。

$$V_{l,k}(t+1)=R_1(k,t)*V_{l,k}(t)+R_2(k,t)*(X_{w,k}(t)-X_{l,k}(t))+\varphi*R_3(k,t)(\overline{X}_k(t)-X_{l,k}(t)) \quad (3)$$

$$X_{l,k}(t+1)=X_{l,k}(t)+V_{l,k}(t+1) \quad (4)$$

$\overline{X}_k(t)$ 为相关粒子的平均位置，相关粒子平均位置可以使全局平均 $\overline{X}^l_{t,k}(t)$，由此可见 $\overline{X}_k(t)$ 的更新机制里面避开了对局部最优解和全局最优解的依赖，增强了粒子的多样性，避免单个粒子扰动或噪声干扰，利于规避过早收敛问题。

1.2 算法稳定性分析

根据式（3）和式（4），改写得到动态方程形式：

$$\begin{bmatrix}V_{l,k}(t+1)\\X_{l,k}(t+1)\end{bmatrix}=\begin{bmatrix}R_1(k,t) & -\varphi(R_3(k,t)+R_2(k,t))\\R_1(k,t) & 1-\varphi(R_3(k,t)+R_2(k,t))\end{bmatrix}\times\begin{bmatrix}V_{l,k}(t)\\X_{l,k}(t)\end{bmatrix}+\begin{bmatrix}R_2(k,t) & \varphi\\R_2(k,t) & \varphi\end{bmatrix}\begin{bmatrix}X_{w,k}(t)\\\overline{X}_k(t)\end{bmatrix} \quad (5)$$

改写成如式（6）所示的形式，成一对一对应关系：

$$y(t+1)=\boldsymbol{A}y(t)+\boldsymbol{B}u \quad (6)$$

其中，$\boldsymbol{A}$ 为状态转移矩阵，$\boldsymbol{B}$ 为输入矩阵。

根据特征方程定义，$|\lambda\boldsymbol{I}-\boldsymbol{A}|=0$，求得如下特征方程如下：

$$\lambda^2+\lambda(\varphi(R_3(k,t)+R_2(k,t)-R_1(k,t)-1)+R_1(k,t)) \quad (7)$$

根据劳斯稳定性判据 [31]，上述动态方程稳定的充分必要条件是：劳斯表中第 1 列各值为正，如果劳斯表第 1 列中出现小于零的数值，系统不稳定，且第 1 列各系数符号的改变次数，代表特征方程（7）的正实部根的数目。

列出特征方程（7）的劳斯表，如表 1 所示。

表 1 特征方程劳斯表

λ^2	1	$R_1(k,t)$
λ^1	$\varphi(R_3(k,t)+R_2(k,t))-R_1(k,t)-1$	0
λ^0	$R_1(k,t)$	0

根据稳定的充分必要条件，第 1 列各值为正，$R_1(k,t)$ 介于 [0,1] 之间是正数，只需满足 $\varphi(R_3(k,t)+R_2(k,t))-R_1(k,t)-1>0$，求得：

$$\varphi>\frac{1+R_1(k,t)}{R_3(k,t)+R_2(k,t)} \quad (8)$$

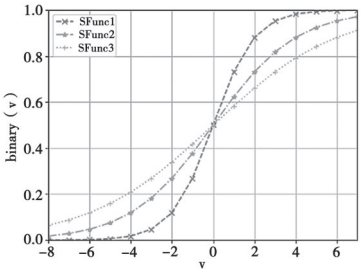
由于 $R_1(k,t),R_2(k,t),R_3(k,t) \in [0,1]^n$，显然不等式（8）右侧取得最小值时，分母趋近于 2，分子趋近于 1，因此其最小值趋近于 0.5，即 φ>0.5。此时，上述动态系统是稳定的，可以实现解空间内的快速收敛。

1.3 二值化方法

竞争机制粒子优化算法每次迭代时更新粒子速度，对粒子速度利用 S 函数 [32] 进行二值化处理，并根据规则对粒子位置进行 0、1 的离散空间处理，其中 1 代表选择该特征，0 代表不选择该特征。

3 种 S 函数表达式、S 函数图形、粒子位置离散化公式详见表 2。

表 2 S 函数表达式及图形

函数名称	函数表达式	函数图形
SFunc1	$\frac{1}{1+e^{-v}}$	
SFunc2	$\frac{1}{1+e^{-v/2}}$	
SFunc3	$\frac{1}{1+e^{-v/3}}$	
$\begin{cases}x=1, v>r\\x=0, v\leqslant r\end{cases}$	r 是介于 [0，1] 之间的随机数，大于 r 为 1，小于 r 为 0，	

连续空间到离散空间的映射关系详见图 2。

2 KDDCUP99 数据集预处理

2.1 KDDCUP99 数据集分析

从源 IP 地址到目标 IP 地址，通过 TCP 或 UDP 协议建立连接，每个网络连接分为正常或异常。异常类型共分为 4 大类共 39 种攻击类型，其中 22 种攻击类型在训练集中，另外 17 种攻击类型在测试集中。4 种异常类型分类如下。

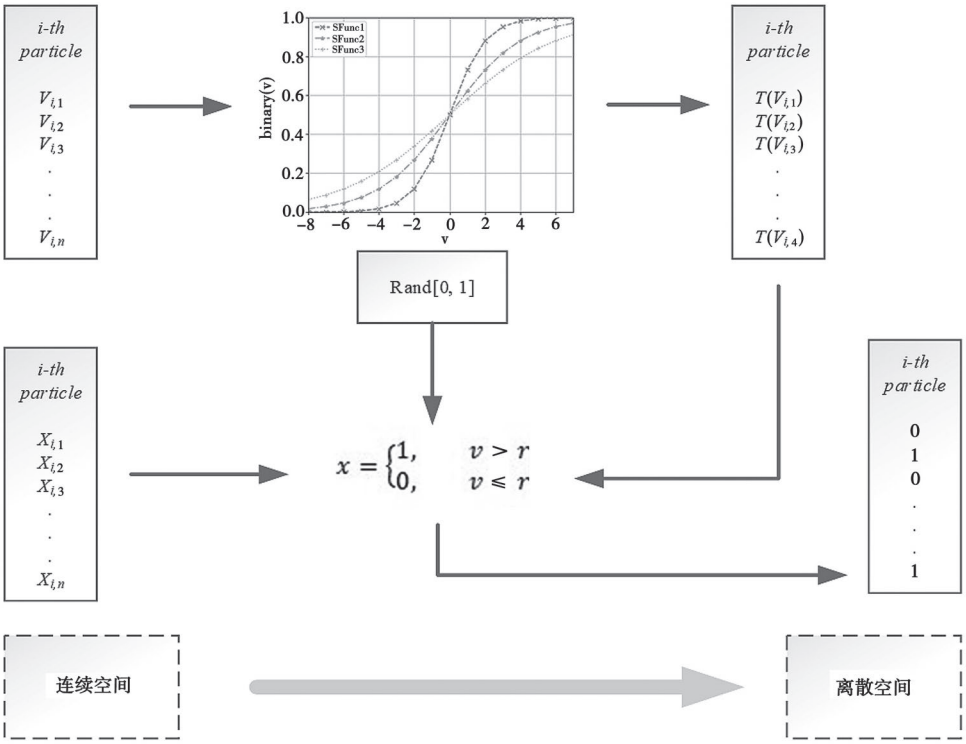


图 2　连续空间到离散空间映射关系

（1）Denial of Service Attacks（DoS）：拒绝服务攻击，有 back、land、neptune、pod、smurf 和 teardrop 共 6 种。

（2）User to Root Attacks（U2R）：来自远程主机的未授权访问，有 ipsweep、nmap、portsweep 和 satan 共 4 种；

（3）Remote to Local attacks（R2L）：未授权的本地超级用户特权访问，有 ftp_write、guess_passwd、imap、multihop、phf、spy、warezclient 和 warezmaster 共 8 种。

（4）Probing attacks：端口监视或扫描，以躲避系统安全控制获取信息，有 buffer_overflow、loadmodule、perl 和 rootkit 共 4 种。

KDDCUP[99] 每一条连接记录由 41 个特征组成，分为 4 种类型（基本特征，内容特征，内容特征，流量特征），第 42 位是特征标签，特征数据结构见图 3。KDDCUP[99] 由约 500 万条记录组成，同时还包括 10% 的训练子集和测试子集，样本类别分布如图 3 所示。部分特征属于字符型，在数据预处理时需要转换成数值型；部分连续性数据需进行数据标准化处理，避免出现数据覆盖现象。

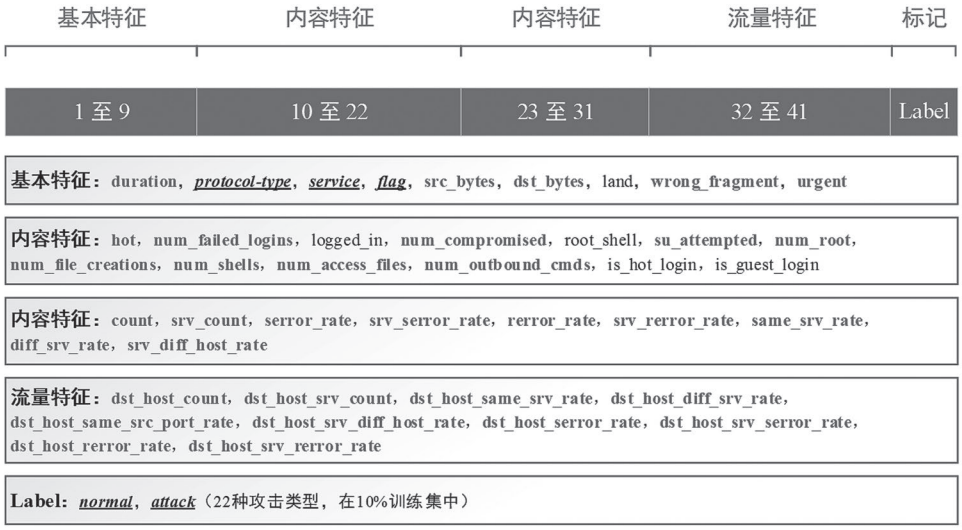


图 3　特征数据结构

2.2 字符型转换成数值型

图 3 为 KDDCUP[99] 数据集特征数据的数据结构，斜体下划线标注的是字符型特征，根据其在列表中出现的顺序值进行数值替换。以 protocol-type 为例，有 3 种协议类型，依次为 tcp、udp 和 icmp，根据上述原则一次将其替换成 0，1，2。同样，service 共有 70 种网络服务类型，转换成数值 0 ~ 69；flag 共有 11 种网络连接状态，转换成数值 0 ~ 10；label 共有 22 种攻击类型，外加正常状态标识 normal（放在攻击类型之前），转换成数值 0 ~ 22。

以上将 KDDCUP[99] 数据集中字符型转换成了数值型，便于后续进一步进行数据处理。

2.3 数据标准化

KDDCUP[99] 数据集中特征数据较为分散，特征数据的数值类型有连续型和离散型，连续型特征数据之间的数值差也较大。为了避免数值差异较大导致的数据覆盖现象，提高模型训练速度，需对数据进行标准化处理。本文采用极值法，即带正向指标的极差变换法。通过式（9）处理后，数据被标准化至 [0，1] 区间。

$$X_{norm}=\frac{X-X_{\min}}{X_{\max}-X_{\min}} \tag{9}$$

3 实验设计及验证

3.1 适应度函数

本文适应度函数采用分类错误率与特征数的线性组合，通过调整权重系数 α 的值分配分类错误率和特征数影响比重：

$$Fitness=\alpha * ErrorRate+(1-\alpha)*\frac{d}{D} \tag{10}$$

ErrorRate 代表分率错误率，由 python 自带 KNN 分类器计算；d 代表第 i 个粒子所选取的特征数；D 是每个粒子总的特征数，针对 KDDCUP[99] 数据集，D 为 41；α 为权重值，用于调节错误率和选择特征数权重。

3.2 参数设置

经过数据预处理的 KDDCUP[99] 数据集共有 494 021 条数据，利用 train_test_split 函数将该数据集随机分成训练集和测试集。在进行模型训练和预测中，为了便于快速计算，随机从训练集和测试集中各选取 1 000 条数据，数据标签选取相对应的标签。

用预测精度、最优适应度值、最差适应度值、平均适应度值以及选择特征来评估模型预测结果。

参数设置如表 4 所示。

表 4 参数设置

参数	值
粒子种群大小	50
粒子维度	41
控制参数 φ	0.6
权重值 α	0.8
迭代次数	100
速度区间	[-6，+6]

3.3 实验结果

3 个 S 函数所对应的预测精度、最优适应度值、最差适应度值以及平均适应度值如表 5 所示。

表 5 预测精度及适应度值

评估指标	SFunc1	SFunc2	SFunc3
预测精度	99.2	97.1	98.8
最优适应度值	0.054 3	0.045 4	0.055
最差适应度值	0.079 6	0.112 1	0.138
平均适应度值	0.065 98	0.089 3	0.083 6

3 个 S 函数所对应的选择特征数个数及位置如表 6 所示，在每条记录 41 个特征中，0 表示不选择，1 表示选择。

表 6 特征数个数及位置

函数名称	选择特征数	41 个特征
SFunc1	15	00010000110100010000110011101000110100010
SFunc2	17	11010000011101000000101000001011110101001
SFunc3	22	11000001100110010011010001101111101111001

适应度值变化曲线详见图 4。可以看出，初始化适应度值在迭代之处能到达一个较为理想的水平，且 SFunc1 较另外两个转换函数较为平整，SFunc2 和 SFunc3 均有较大突变，突变后趋于稳定。可以看出，基于竞争机制的二值化粒子优化算法具有在短时间内可以到达较为稳定的状况。

以 SFunc1 为例，设置 K 值范围为 1 ~ 20，采用 sklearn 的 cross_val_score 进行交叉验证，求取准

确度的平均值。通过比较准确度值，选取最合适的 K 值，实现参数调优。通过仿真计算，准确度随 K 值变化曲线详见图 5，其中 K=4 时准确度为 0.994 5，此时为模型参数最优。

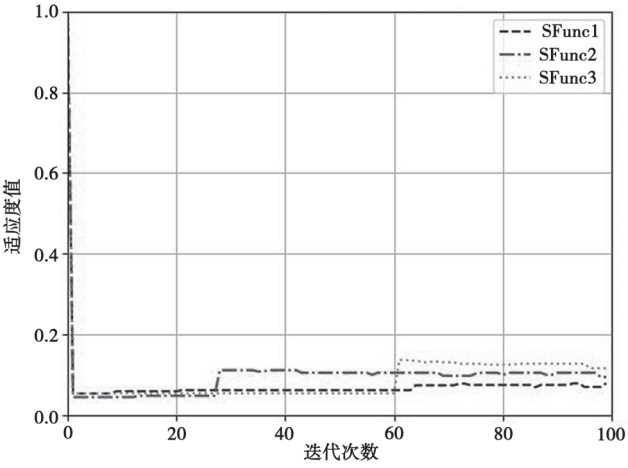


图 4　适应度值变化曲线

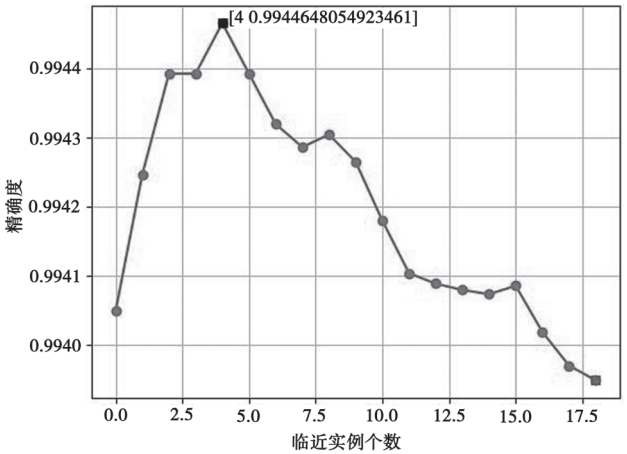


图 5　精确度随 K（实例个数）值变化曲线

4　结　语

本文采用二值化竞争机制粒子优化算法优化选择 KDDCUP[99] 数据集的特征子集，实现数据的降维处理，创新性地采用稳定性理论分析算法模型，选取调控参数取值范围，分析了 KDDCUP[99] 数据集的预处理思路、方法和流程，并采用 sklearn 的 KNN 算法搭建训练模型，利用选择的特征子集进行模型训练，利用交叉验证机制对模型进行精确度校验。实验结果表明，通过优化算法选取的特征子集搭建的训练模型具有较高的精确度，且在处理大数据时具有较大优势。

本文以 KDDCUP[99] 数据集进行验证。KDDCUP[99] 数据集主要是网络流量数据，在入侵检测领域具有较高的代表性。后续将以实际入侵检测模型搭建需求为牵引，将本文研究的优化算法结合实际需求搭建合适的入侵检测模型，对实际的入侵检测行为进行实时的分析感知。

参考文献：

[1] Amparo A B,Noelia S M,F é lix M C F,et al.Classification of Computer Intrusions Using Functional Networks[C]. 15th European Symposium on Artificial Neural Networks,2007.

[2] Zainal A,Maarof M A,Shamsuddin S M,et al.Ensemble of One-class Classifiers for Network Intrusion Detection System[C].Fourth International Conference on Information Assurance and Security,2008.

[3] 蒋建春，马恒太，任党恩，等．网络安全入侵检测研究综述 [J]. 软件学报，2000,11(11):1460-1466.
JIANG Jian-chun,MA Heng-tai,REN Dang-en,et al. Survey of Network Security Intrusion Detection[J].Acta Software Sinica,2000,11(11):1460-1466.

[4] 杨宏宇，朱丹，谢丰，等．入侵异常检测研究综述 [J]. 电子科技大学学报（自然科学版），2009,38(05):587-596.
YANG Hong-yu,ZHU Dan,XIE Feng,et al.Survey of Anomaly Intrusion Detection Research[J].Journal of University of Electronic Science and Technology of China,2009,38(05):1460-1466.

[5] 刘勇，茅兵．入侵检测研究综述 [J]. 计算机科学，2001,28(07):42-45.
LIU Yong,MAO Bing.Survey on Intrusion Detection Techniques[J].Computer Science,2001,28(07):42-45.

[6] 中国信息通信研究院．人工智能数据安全白皮书 (2019 年)[EB/OL].(2019-08-01)[2020-04-11].http://www.doc88.com/p-5894717480162.html.
China Institute of Information and Communication.White paper on artificial Intelligence Data Security(2019)[EB/OL].(2019-08-01)[2020-04-11].http://www.doc88.com/p-5894717480162.html.

[7] 全国信息安全标准化委员会大数据安全标准特别工作组．人工智能安全标准化白皮书 (2019 年)[EB/OL].(2019-01-04)[2020-04-11].https://max.book118.com/html/2019/0104/7023132111001201.shtm.
Big Data Security Standards Task Force of The National Information Security Standardization Committee.White Paper on Artificial Intelligence Security Standardization

(2019).[EB/OL].(2019-01-04)[2020-04-11].https://max.book118.com/html/2019/0104/7023132111001201.shtm.

[8] 王娟 , 慈林林 , 姚康泽 . 特征选择方法综述 [J]. 计算机工程与科学 ,2005,27(12):68-71.
WANG Juan,CI Lin-lin,YAO Kang-ze.A Survey of Feature Selection[J].Computer Engineering and Science,2005,27(12):68-71.

[9] 姚旭 , 王晓丹 , 张玉玺 , 等 . 特征选择方法综述 [J]. 控制与决策 ,2012,27(02):161-166,192.
YAO Xu,WANG Xiao-dan,ZHANG Yu-xi,et al.Summary of Feature Selection Algorithms[J].Control and Decision,2012,27(02):161-166,192.

[10] 刘艺 , 曹建军 , 刁兴春 , 等 . 特征选择稳定性研究综述 [J]. 软件学报 ,2018,29(09):2559-2579.
LIU Yi,CAO Jian-jun,DIAO Xing-chun,et al.Survey Stability of Feature Selection[J].Journal of Software,2018,29(09):2559-2579.

[11] 王艳丽 , 梁静 , 薛冰 , 等 . 基于进化计算的特征选择方法研究概述 [J]. 郑州大学学报 (工学版),2020,41(01):49-57.
WANG Yan-li,LIANG Jing,XUE Bing,et al.Research on Evolutionary Computation for Feature Selection[J].Journal of Zhengzhou University(Engineering Science),2020,41(01):49-57.

[12] Isabelle G,Andre E.An Introduction to Variable and Feature Selection[J].Journal of Machine Learning Research,2003(03):26.

[13] Yvan S,Iñaki I,Pedro L.A Review of Feature Selection Techniques in Bioinformatics[J].Bioinformatics(Oxford,England),2007(23):2507-2517.

[14] 张雪芹 , 顾春华 . 一种网络入侵检测特征提取方法 [J]. 华南理工大学学报 (自然科学版),2010,38(01):81-86.
ZHANG Xue-qin,GU Chun-hua.A Network Intrusion Detection Feature Extraction Method[J].Journal of South China University of Technology(Natural Science Edition),2010,38(01):81-86.

[15] Lin K C,Hung J C,Wei J.Feature Selection with Modified Lion's Algorithms and Support Vector Machine for High-dimensional Data[J].Appl. Soft Comput.,2018(68):669-676.

[16] 周志华 . 机器学习 [M]. 北京 : 清华大学出版社 ,2016:247-254.
ZHOU Zhi-hua.Machine Learning[M].BEIJING:Tsinghua University Press,2016:247-254.

[17] Xue B,Zhang M,Browne W N.Particle Swarm Optimization for Feature Selection in Classification:A Multi-Objective Approach[J].IEEE Trans. Cybern.,2013(43):1656-1671.

[18] Huang C L,Wang C J.A GA-based Feature Selection and Parameters Optimization for Support Vector Machines[J].Expert Syst. Appl.,2006(31):231-240.

[19] De Stefano C,Fontanella F,Marrocco C,et al.A GA-based Feature Selection Approach with an Application to Handwritten Character Recognition[J].Pattern Recognit. Lett.,2014(35):130-141.

[20] 张文杰 , 蒋烈辉 . 一种基于遗传算法优化的大数据特征选择方法 [J]. 计算机应用研究 ,2020(01):50-52,56.
ZHANG Wen-jie,JIANG Lie-hui.Using Genetic Algorithm for Feature Selection Optimization on Big Data Processing[J].Application Research of Computers,2010(01):50-52,56.

[21] 任江涛 , 孙婧昊 , 黄焕宇 , 等 . 一种基于信息增益及遗传算法的特征选择算法 [J]. 计算机科学 ,2006,33(10):193-195.
REN Jiang-tao,SUN Jing-hao,HUANG Huan-yu,et al.Feature Selection Based on Information Gain and GA[J].Computer Science,2006,33(10):193-195.

[22] 俞研 , 黄皓 . 面向入侵检测的基于多目标遗传算法的特征选择 [J]. 计算机科学 ,2007,34(03):197-200.
YU Yan,HUANG Hao.Feature Selection Using Multi-Objective Genetic Algorithms for Intrusion Detection[J].Computer Science,2007,34(03):197-200.

[23] 黄炜 , 黄志华 . 一种基于遗传算法和 SVM 的特征选择 [J]. 计算机技术与发展 ,2010,20(06):21-24.
HUANG Wei,HUANG Zhi-hua.Feature Selection Based on Genetic Algorithm and SVM[J].Computer Technology and Development,2010,20(06):21-24.

[24] 巢秀琴 , 李炜 . 人工蜂群算法优化的特征选择方法 [J]. 计算机科学与探索 ,2019,13(02):300-309.
CHAO Xiu-qin,LI-Wei.Artificial Bee Colony Algorithm Based on KNEE Points[J].Journal of Frontiers of Computer Science and Technology,2019,13(02):300-309.

[25] Emrah H,Bing X,Mengjie Z,et al.Pareto Front Feature Selection Based on Artificial Bee Colony Optimization[J].Information Sciences:An International Journal,2018(422):462-479.

[26] Liu X,Shang L.A Fast Wrapper Feature Subset

Selection Method Based On Binary Particle Swarm Optimization[J].IEEE Congress on Evolutionary Computation,2013(07):3347-3353.

[27] Zhe Z,Xing L,Ping L,et al.Feature Selection Method with Proportionate Fitness Based Binary Particle Swarm Optimization[J].Simulated Evolution and Learning,2014(8886):582-592.

[28] 郑洪英 , 侯梅菊 , 王渝 . 入侵检测中的快速特征选择方法 . 计算机工程 ,2010,36(06):262-264.
ZHENG Hong-ying,HOU Mei-ju,WANG Yu.Fast Method for Feature Selection in Intrusion Detection[J]. Computer Engineering,2010,36(06):262-264.

[29] Zhang H,Gao H,Wang X.Quantum Particle Swarm Optimization Based Network Intrusion Feature Selection and Detection[C].International Federation of Automatic Control World Congress,2008.

[30] KDD Cup 1999 Data[EB/OL].http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html.

[31] 胡寿松 . 自动控制原理 [M]. 北京 : 科学出版社 ,2007:98-106.
HU Shou-song.Principle of Automatic Control[M]. Beijing:Science Press,2007:98-106.

[32] Mirjalili S,Lewis A.S-shaped Versus V-shaped Transfer Functions for Binary Particle Swarm Optimization[J]. Swarm Evol. Comput.,2013(09):1-14.

[33] Cheng R,Jin Y.A Competitive Swarm Optimizer for Large Scale Optimization[J].IEEE Trans. Cybern., 2015(45):191-204.

作者简介：

周利均（1987—），男，硕士，工程师，主要研究方向为人工智能与大数据安全、网络安全。