基于 ELM-SVM 的入侵检测模型研究

贾一帆[1,*]，宋孟月[1]，王莉芬[1]，孙岚琪[1]，闫芮铵[2]
[1]河南工业大学人工智能与大数据学院，河南郑州，中国
[2]科大讯飞股份有限公司，安徽合肥，中国

【摘要】传统的基于支持向量机（SVM）的入侵检测系统分类思想简单，分类效果也表现出鲁棒性，但是在大数据时代，随着数据量的增加，以及入侵攻击的多样性和未知性，基于 SVM 的入侵检测系统出现了较高的误检率和漏检率，为了解决这些问题提出一种基于极限学习机和 SVM 的入侵检测系统（ELM-SVM），利用极限学习机训练速度快的优点，快速训练不同结构的极限学习机模型并产生新的数据集，然后利用 SVM 对数据进行分类处理，实验利用 NSL-KDD 数据集进行仿真实验，实验结果表明 ELM-SVM 的准确率和召回率为 98.2%，其结果与 ELM 模型和 SVM 模型相比都有所提高，并且 ELM-SVM 有较快的训练速度，因此 ELM-SVM 模型具有更好的鲁棒性。

【关键词】入侵检测；支持向量机；极限学习机；ELM-SVM

1.引言

入侵检测技术（Intrusion Detection Systems，简称 IDS）是网络安全系统中的第二道防线，它主要通过对网络流量进行采集，然后按照一定的关联规则来对异常数据进行检测并进行实时报警。随着互联网时代的到来，网络传输的流量大大增加，同时黑客用来入侵的方式和技术得到了更多的提升，因此越来越多的网络用户和公司受到了网络攻击[1]。例如 11 月 5 日，西班牙的两家大型公司 Everis（NTT Data Group 旗下的 IT 咨询公司）和 Cadena SER（西班牙最大的无线电网络公司）在同一天受到了网络勒索软件攻击；11 月 12 日 SmarterASP。NET 公司遭到勒索软件的攻击，该攻击对公司客户数据进行了加密，对该公司造成了巨大的损失。因此网络入侵技术受到了更加广泛的关注，通常网络入侵检测技术主要分为滥用检测和异常检测，异常对于处理未知的攻击类型表现有良好的鲁棒性，因此一般的网络入侵检测技术主要是基于异常检测[2,3]。如随着攻击类型的多样性，对于如何能够降低入侵检测系统的误检率和提高其泛化能力是网络安全领域研究最需解决的问题[4]。

随着人工智能技术的发展，越来越多的学者将机器学习技术应用到入侵检测中，基于机器学习的入侵检测不仅可以检测到已知的攻击类型，还能检测到未知的攻击类型[5]。钱亚冠等[6]提出一种基于 Support Vector Machine（简称 SVM）的入侵检测技术，并取得了较高的识别率。随着神经网络技术的发展，蒋东荣等[7]等将神经网络技术应用到入侵检测系统中，结果显示提高了检测系统的准确率，虽然以上方法都提高了检测的准确率，但是在大数据时代，由于数据量的增加，SVM 和神经网络的入侵检测会出现维度灾难以及训练速度下降，无法满足网络安全的实时性，因此为了提高检测系统的训练速度，李熠等[8]提出一种基于极限学习机的入侵检测技术，并减少了模型的训练时间，但是该技术带来的另外一个问题就是入侵检测率较低，为了解决这些问题设计了一种 Extreme Learning Machines-Support Vector Machine（简称 ELM-SVM）的入侵检测系统，该系统主要通过将数据输入不同结构的 ELM 模型，然后输出新的低维数据，然后将这些新生成的低维数据输入到 SVM 分类器最后得到分类结构，该系统解决了 ELM 的准确率较低，以及 SVM 的维度灾难等问题，最后到了较高的分类结果和泛化能力。通过实验证明，ELM-SVM 模型较 ELM 模型和 SVM 模型有更高的准确率和泛化能力。因此该模型对于入侵检测有更好的鲁棒性。

2.相关知识

2.1 极限学习机（ELM）

极限学习机（ELM）首次被 Huang 等[9]提出，ELM 是一种快速学习算法，传统的 BP 神经网络主要是通过不断的反向传播来

对权重进行调整，而极限学习机通过随机设置输入层和隐藏层的连接权重，并且一旦设定就不再更改，而隐藏层和输出层的连接权重则是通过解方程组方式一次性确定。

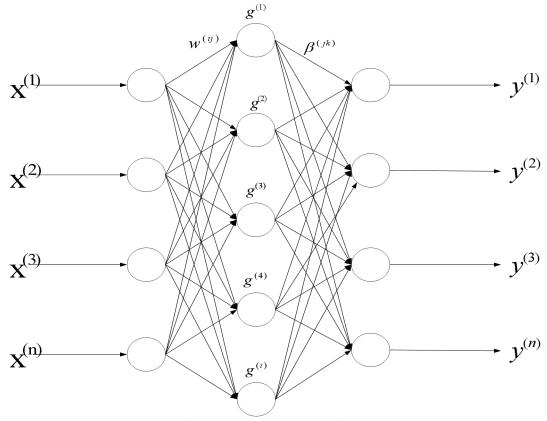

图 1.单隐藏层的神经网络

对于一个单隐藏层的神经网络（如图 1），可以表示为公式 1：

$$\sum_{i=1}^{L}\beta_i g\left(W_i\cdot X_j+b_i\right)=o_j,j=1,\cdots,N \quad (1)$$

其中 $g(x)$ 表示激活函数，W_i 表示输入层与隐藏层的连接权重，β_i 表示隐藏层与输出层的连接权重，X_i 表示输入数据的特征，o_j 表示预测标签值，b_i 表示偏差。

为了简化我们可以将式 1 表示为式 2：

$$H\beta=O \quad (2)$$

其中 H 表示隐藏层的输出，β 表示隐藏层与输出层的连接权重，O 表示预测标签值。

神经网络的学习主要是为了使预测值与真实值误差最小如式 3：

$$\min\|H\beta-T\| \quad (3)$$

其中 T 表示输入数据的真实标签。

传统神经网络中主要利用梯度下降法算法，来求式 3 的解，但是利用梯度的学习算法求解的过程中需要进行不断的迭代来调整所有参数，直到满足要求，而在 ELM 算法中输入层和隐藏层的连接权重已经被随机初始化，因此只需要利用解方程组的方法求式 3 的解即可，其结果如式 4：

$$\hat{\beta}=H^{\dagger}T \quad (4)$$

其中 $H^{\dagger}$ 表示矩阵的 Moore-Penrose 广义逆。并且可以证明求得的解 $\hat{\beta}$ 的范数是最小的并且唯一。

2.2 支持向量机（SVM）

SVM 是基于统计学习方法中的一种学习算法，SVM 可以用来解决线性问题和非线性问题。其处理分类问题的典型模式是寻找使各类别分类间隔最大的最优分类超平面，对于 SVM 分类算法可以看作是一种约束优化问题如式 5：

$$\begin{aligned}&\min_{\mathbf{w},b}\ \frac{1}{2}\mathbf{w}^T\cdot\mathbf{w} \quad (5)\\&\text{s.t.}\quad y^{(i)}\left(\mathbf{w}^T\cdot\mathbf{x}^{(i)}+b\right)\geq 1\quad \forall\, i=1,2,\cdots,m\end{aligned}$$

其中 $y^{(i)}$ 和 $\mathbf{x}^{(i)}$ 分别表示样本的标签和属性特征，$\mathbf{w}$ 和 b 表示需要学习的参数。

对于线性的不可分的数据，需要为每一个实例引入一个松弛变量 $\xi^{(i)}\geq 0$，$\xi^{(i)}$ 表示第 i 个实例上允许间隔违例的程度，因此对于非线性的约束优化问题可表示为式 6：

$$\begin{aligned}&\min_{\mathbf{w},b}\ \frac{1}{2}\mathbf{w}^T\cdot\mathbf{w}+C\sum_{i=1}^{N}\xi^{(i)} \quad (6)\\&\text{s.t.}\quad y^{(i)}\left(\mathbf{w}^T\cdot x^{(i)}+b\right)\geq 1-\xi^{(i)}\quad \forall\, i=1,2,\cdots,m\end{aligned}$$

其中 C 表示惩罚因子，主要作用是为了平衡最大间隔与松弛变量。

对于非线性不可分的数据，需要将数据映射到高维空间，从而转化为线性数据，其约束优化问题可表示为式 7：

$$\begin{aligned}&\min_{\mathbf{w},b}\ \frac{1}{2}\mathbf{w}^T\cdot\mathbf{w}+C\sum_{i=1}^{N}\xi^{(i)} \quad (7)\\&\text{s.t.}\quad y^{(i)}\left(\mathbf{w}^T\cdot\phi\left(x^{(i)}\right)+b\right)\geq 1-\xi^{(i)}\quad \forall\, i=1,2,\cdots,m\end{aligned}$$

其中 $\phi\left(x^{(i)}\right)$ 表示非线性函数。

在机器学习中，可以利用核函数来计算 $\phi(a)$ 和 $\phi(b)$ 的点积，利用核函数可以直接基于原始向量 a 和 b 来计算，不需要计算转换函数 $\phi(x)$。常见核函数如式 8：

$$\begin{aligned}&\text{Linear:}\quad K(\mathbf{a},\mathbf{b})=\mathbf{a}^T\cdot\mathbf{b}\\&\text{Polynomial:}\quad K(\mathbf{a},\mathbf{b})=\left(\gamma\mathbf{a}^T\cdot\mathbf{b}+r\right)^d \quad (8)\\&\text{Gaussian RBF:}\quad K(\mathbf{a},\mathbf{b})=\exp\left(-\gamma\|\mathbf{a}-\mathbf{b}\|^2\right)\\&\text{Sigmoid:}\quad K(\mathbf{a},\mathbf{b})=\tanh\left(\gamma\mathbf{a}^T\cdot\mathbf{b}+r\right)\end{aligned}$$

2.3 ELM-SVM 模型

ELM-SVM 模型利用不同结构的 ELM 模型来产生低维数据，然后将新产生的数输入到 SVM 训练模型中输出最终结果如图 2。通过 ELM 模型产生的数据相较于原始数据有更低的维度，很好的解决 SVM 处理大量数据分类效果欠佳以及 ELM 分类准确率较低的问题和训练速度比较慢的问题。该模型融合 ELM 的快速训练的优点，以及 SVM 的分类性能较好的优点，实现了一种应对大规模

ISSN: 3079-4455

数据也可以用较短的训练时间来得到较高的分类准确率和较好的分类误报率。

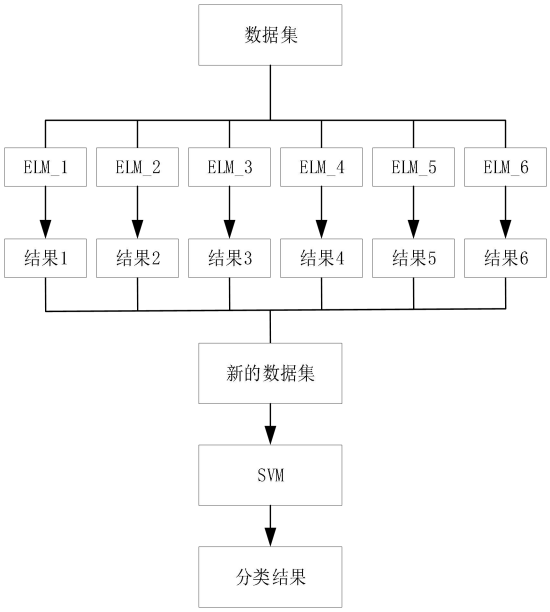

图 2.ELM-SVM 模型

ELM-SVM 的工作步骤如下：

（1）将数据进行预处理；

（2）构建不同结构的 ELM 模型，然后将数据输入；

（3）将不同训练的结果进行保存，并合并为新的数据集；

（4）将经过降维的数据输入到 SVM 模型，并得到最终的分类结果。

3 实验验证与分析

3.1 数据集选取

实验中主要利用 NSL-KDD CUP 数据集，与 KDD CUP99 数据集相比，NSL-KDD CUP 删除了大量冗余数据，与现实世界的入侵行为更加接近，因此更加适合本实验的检测模型。NSL-KDD CUP 数据集有 42 维数据特征，其数据类型主要分为正常类型和 4 中攻击类型。本实验所用训练集和测试集的样例数分别如图 3 和图 4 所示：

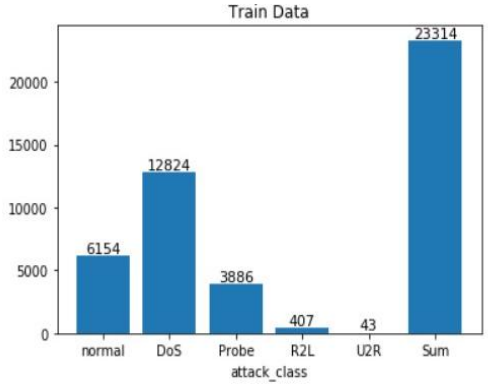

图 3.训练集数据

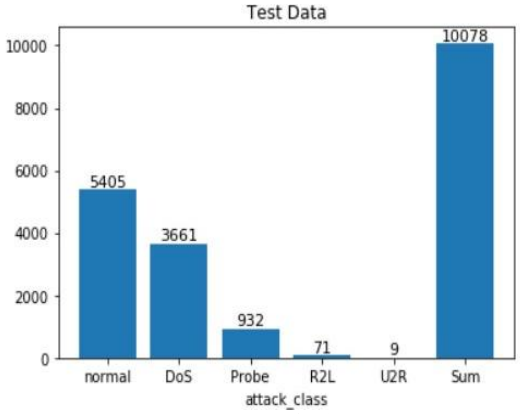

图 4.测试集数据

3.2 数据预处理

NSL-KDD CUP 数据集由 38 个数字型特征和 3 个符号型特征以及一个标签特征组成[10]，因此主要对数据进行如下处理：

（1）符号型特征进行数值化；

（2）数据归一化处理，主要是将数据中特征的取值映射到[0,1]区间内，来加快模型的学习效率以及泛化能力，其归一化操作如式 9：

$$X_{\text{norm}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \qquad (9)$$

其中 $X_{\max}$ 表示特征数据的最大值，$X_{\min}$ 表示特征数据的最小值，X 表示特征数据的原始值，X_{norm} 表示归一化后的数据。

（3）对于标签类型需要将 39 类小攻击类型转化为 4 中大攻击类型，并读取类标识，其中 0 表示 normal，1 表示 DoS，2 表示 Probe，3 表示 R2L，4 表示 U2R。

3.3 实验环境

实验中运行环境配置如表 1 所示。

表 1.实验环境

软硬件	配置
操作系统	win10
CPU	i7-2600 @ 3.40GHz 四核
内存	16.0 GB RAM
开发工具	Pycharm2018
实验语言	Python3.5

3.4 评估标准

本实验主要通过精确率（Precision，P），召回率（Recall，RE）和 F1 值来评估模型。

$$P = \frac{TP}{TP + FP} \qquad (10)$$

$$\text{RE} = \frac{TP}{TP + FN} \qquad (11)$$

$$F_1 = \left(\frac{RE^{-1} + P^{-1}}{2} \right)^{-1} = 2 \cdot \frac{P \cdot RE}{P + RE} \quad (12)$$

其中 TP 表示将正类预测为正类的个数，FP 表示将负类预测为正类数的个数，FN 表示将正类预测为负类的个数。

3.5 实验结果与分析

实验所得的结果如表 2 所示

表 2.每个攻击类型结果

评估指标	样本	ELM	SVM	ELM-SVM
P	normal	0.97	0.99	0.99
	DoS	0.98	0.97	0.99
	Probe	0.86	0.89	0.92
	R2L	0.79	0.92	0.92
	U2R	0.44	0.56	0.44
RE	normal	0.99	0.97	1.00
	DoS	0.95	0.99	0.97
	Probe	0.88	0.98	0.95
	R2L	0.61	1.00	0.87
	U2R	1.00	1.00	1.00
F1	normal	0.98	0.98	0.99
	DoS	0.97	0.98	0.98
	Probe	0.87	0.93	0.93
	R2L	0.69	0.96	0.89
	U2R	0.62	0.71	0.62

normal 类型的标签。ELM-SVM 模型和 SVM 模型的精确率相同而 ELM 模型为 97%有所降低；ELM-SVM 模型的召回率比 ELM 模型和 SVM 模型分别提高了 1%和 2%；ELM 模型和 SVM 模型的 F1 值相同而 ELM-SVM 模型提高了 1%。

DoS 类型的标签。ELM 模型和 SVM 模型的精确率相近而 ELM-SVM 模型有所提升；SVM 模型的召回率率最高相较于 ELM-SVM 模型提高 2%；SVM 模型和 ELM-SVM 模型的 F1 值相同而 ELM 模型降低了 1%。

Probe 类型的标签。ELM-SVM 模型的精确率相较于 ELM 模型和 SVM 模型分别提高了 6%和 3%； ELM 模型的召回率最低，而 SVM 模型比 ELM-SVM 模型提高了 3%；ELM-SVM 的 F1 值与 SVM 模型相同为 93%，而 ELM 模型为 87%。

R2L 类型的标签。SVM 模型和 ELM-SVM 模型的精确率相同而 ELM 模型比较低；SVM 模型的召回率和 F1 值相较于 ELM 模型和 ELM-SVM 模型都有很大的提升。

U2R 类型的标签。SVM 模型的精确率相较于 ELM 模型和 ELM-SVM 模型有很大的提升；召回率三个模型相同；SVM 模型的 F1 值最高。

综上可以看出，对于数据占比比较多的攻击类型利用 ELM-SVM 可以更容易检测，而对于占比比较少的攻击类型利用 SVM 更佳。

ELM-SVM 模型于 SVM 模型的训练时间如表 3 所示：

表 3.训练时间

模型	训练时间（s）
ELM-SVM	177
SVM	209

可见 ELM-SVM 可以快速的训练模型提高了入侵检测的实时性。更加符合现实世界中的入侵检测系统。

对于总体的攻击类型的结果如表 4 所示：

表 4.总体的结果

评估指标	ELM-SVM	ELM	SVM
P	0.982	0.962	0.977
RE	0.982	0.963	0.977
F1	0.982	0.965	0.977

ELM-SVM 模型的总体精确率较 ELM 模型和 SVM 模型都有所提高。因此利用 ELM-SVM 模型相较于 ELM 模型和 SVM 模型更加适用于入侵检测。

4.结束语

设计了通过融合极限学习机和支持向量机算法各自的优点，提出了一种基于极限学习机和支持向量机的入侵检测系统，并通过仿真实验得出对于总体的检测性能 ELM-SVM 入侵检测模型相较于 ELM 和 SVM 有更高的精确率和召回率，并且 ELM-SVM 由较快的训练速度提高了模型检测的实时性。因此 ELM-SVM 模型更加适用于入侵检测系统。但是对于较少数据占比的攻击类型类型 SVM 具有较高的检测性能，因此对于提高 ELM-SVM 在较少数据占比的攻击类型类型中的召回率是下一步需要解决的问题。

参考文献

[1] 池亚平，凌志婷，王志强，杨建喜.基于支持向量机与 Adaboost 的入侵检测系统[J].计算机工程，2019，45（10）：183-188+202.

[2] 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.

[3] Kai Zheng, Zhiping Cai, Xin Zhang,

Zhijun Wang, Baohua Yang. Algorithms to speedup pattern matching for network intrusion detection systems [J]. Computer Communications, 2015, 62(C): 47-58.

[4] 王有金.数据挖掘在入侵检测系统中的应用研究[D].吉林大学，2017.

[5] 刘明川，彭长生.基于机器学习的入侵检测研究[J].计算机工程与设计，2008（11）：2736-2738+2741.

[6] 钱亚冠，卢红波，纪守领，周武杰，吴淑慧，雷景生，陶祥兴.一种针对基于 SVM 入侵检测系统的毒性攻击方法[J].电子学报，2019，47（01）：59-65.

[7] 蒋东荣，陈冠霖，贾勇等.基于元胞自动机与多层前馈神经网络的电力信息物理融合系统安全评估模型[J].重庆理工大学学报（自然科学），2018（3）：17-26.

[8] 李熠，李永忠.基于自编码器和极限学习机的工业控制网络入侵检测算法[J/OL].南京理工大学学报，2019（04）：408-413.

[9] Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: Theory and applications [J]. Neurocomputing, 2006, 70(1-3): 489-501.

[10] 刘月峰，王成，张亚斌，苑江浩.用于网络入侵检测的多尺度卷积 CNN 模型[J].计算机工程与应用，2019，55（03）：90-95+153.