

基于非对称卷积自编码器和支持向量机的入侵检测模型

王佳林, 刘吉强, 赵迪, 王盈地, 相迎宵, 陈彤, 童恩栋, 牛温佳

(北京交通大学智能交通数据安全与隐私保护北京市重点实验室, 北京 100044)

摘 要: 网络入侵检测系统在防护网络安全中占据重要地位, 随着科技不断发展, 目前的入侵技术没有考虑到检测技术的可扩展性、可持续性以及训练时间长短, 无法应对现代复杂多变的网络异常流量。针对这些问题, 提出了一种新的深度学习方法, 使用无监督的非对称卷积自编码器, 对数据进行特征学习。另外, 提出了一种新的基于非对称卷积自编码器和多类支持向量机相结合的方法。在 KDD99 数据集上进行了实验, 实验结果表明, 该方法取得了良好的结果, 与其他方法相比显著减少了训练时间, 进一步提高了网络入侵检测技术。

关键词: 入侵检测技术; 卷积自编码器; 支持向量机; 网络安全

中图分类号: TP18

文献标识码: A

doi: 10.11959/j.issn.2096-109x.2018086

Intrusion detection model based on non-symmetric convolution auto-encode and support vector machine

WANG Jialin, LIU Jiqiang, ZHAO Di, WANG Yingdi, XIANG Yingxiao,
CHEN Tong, TONG Endong, NIU Wenjia

Beijing Key Laboratory of Security and Privacy in Intelligent Transportation, Beijing Jiaotong University, Beijing 100044, China

Abstract: Network intrusion detection system plays an important role in protecting network security. With the continuous development of science and technology, the current intrusion technology cannot cope with the modern complex and volatile network abnormal traffic, without taking into account the scalability, sustainability and training time of the detection technology. Aiming at these problems, a new deep learning method was proposed, which used unsupervised non-symmetric convolutional auto-encoder to learn the characteristics of the data. In addition, a new method based on the combination of non-symmetric convolutional auto-encoder and multi-class support vector machine was proposed. Experiments on the data set of KDD99 show that the method achieves good results, significantly reduces training time compared with other methods, and further improves the network intrusion detection technology.

Key words: intrusion detection technology, convolutional auto-encoder, support vector machine, network security

收稿日期: 2018-09-07; 修回日期: 2018-10-24

通信作者: 牛温佳, niuwj@bjtu.edu.cn

基金项目: 中央高校基础研究基金资助项目 (No.2017RC016, No.2018JBZ103); 国家自然科学基金资助项目 (No.61672092); 信息保障科技实验室基金资助项目 (No.614200103011711); 北京优秀人才培养基金资助项目 (No.BMK2017B02-2); 国家留学基金委资助项目 (No.201807095014)

Foundation Items: The Fundamental Research Funds for the Central Universities (No.2017RC016, No.2018JBZ103), The National Natural Science Foundation of China (No.61672092), Science and Technology on Information Assurance Laboratory (No.614200103011711), Beijing Excellent Talent Training Project (No.BMK2017B02-2), China Scholarship Council (No.201807095014)

1 引言

随着科技水平不断发展, 网络空间安全的防护技术也在不断提升, 但仍然面临着许多新的入侵威胁。根据赛门铁克 2018《互联网安全威胁报告》^[1]指出, 2018 年网络中恶意入侵的发生率增长了 200%, 针对物联网攻击的增长率更是达到 600%, 严重威胁着人民的权益, 因此防护网络空间安全在网络空间中占据重要位置。但网络空间安全面对现代网络的复杂变化存在许多新的挑战。例如, 第一, 现代物联网以及云服务的不断普及, 使网络数据量急剧增长, 并且这种情况将继续持续下去; 第二, 对于大量的网络入侵流量, 在有限带宽的影响下, 人们对于入侵检测的要求不仅仅停留在较高的准确率方面, 更是对检测的效率提出了新的要求; 第三, 在现代网络流量中增加了很多新的协议, 这给网络安全检测系统增加了很多检测难度和复杂性。因此, 需要不断调整网络防护措施来适应不断变化的网络状况。

现存的防护网络安全的措施主要有放置在网关处的防火墙, 它可以静态地监视入侵活动。为了动态地防护网络安全, 入侵检测系统 (IDS) 被用作第二道防线, 它可以主动地防护计算机中的日志以及系统文件等, 并且通过分析日志信息、文件的变动得出攻击的迹象, 检测出计算机中的正常或者异常流量, 甚至是一些未知的新的攻击^[2]。IDS 又可以分为基于主机的入侵检测系统和基于网络的入侵检测系统。基于主机的入侵检测系统使用由单个计算机系统收集的信息, 而基于网络的入侵检测系统收集原始网络数据分组作为网络数据源并分析入侵标志^[3], IDS 在防护网络空间安全中占据重要地位。

近年来, 机器学习和深度学习的方法引起了广泛关注。许多研究人员使用它们对网络入侵检测系统进行了研究, 并且在准确率方面取得了不错的成绩。目前, 基于网络的入侵检测系统最大的挑战是提高它的泛化性能和效率。但是, 由于很难获得可靠的训练数据、训练数据的寿命以及高错误率等, 大多数已有的入侵检测系统仍然停留在基于已有标签的检测^[4], 无法追赶现在层出不穷的异常流量, 这种方式高度依赖于数据的可

用性, 需要人类专家对数据进行过滤, 依靠人类专家的指导, 这是一个十分耗费时间和成本的过程, 还可能限制计算机的能力。并且这种方式的检测系统基本达到了饱和的状态, 如果一直依赖这种方式应对现在日新月异的网络异常流量, 将会出现高错误率以及无效检测的状况^[7]。因此, 应该及时调整检测方式, 提出一种能够被广泛接受的方式, 来打破现代网络流量变化的约束。

近年来, 一些浅层的 K 近邻 (KNN, k -nearest neighbor) 算法^[8]、SVM 算法等被应用在入侵检测领域^[9], 并且展现出了良好的性能。但这样的浅层学习算法也有一定的局限性, 对数据样本的限制以及复杂函数的表示能力有一定的限制, 针对复杂问题, 它们的泛化能力受到一定约束。为了克服浅层学习的问题, 一些研究者已经证明, 分层的深度学习算法和浅层分类器相比较, 能够更好地对网络数据进行学习以及分类^[10]。

为了解决上述问题, 本文提出了一种新的无监督的基于非对称卷积自编码器 (NCAE) 和支持向量机 (SVM) 的入侵检测方法, 结合了深层学习和浅层分类器对数据进行分析 and 识别, 适应现代网络流量的变化, 使用 KDD99 数据集进行实验, 实验结果表明, 该方法有效提高了入侵检测技术的效率和检测能力。

本文的主要贡献总结如下。

1) 使用无监督特征学习的新的非对称卷积自编码器 (NCAE) 技术, NCAE 方法与典型的自动编码器方法不同, 它提供了非对称数据降维, 解决了卷积神经网络和自编码器的缺陷。因此, 与深度信念网络 (DBN) 以及栈式自编码器 (SAE) 等领先方法相比, 本文所提技术能够改进分类结果。

2) 利用深度的 NCAE 和 SVM 相结合的分类算法。通过结合两者的深度和浅层学习技术, 开发各自的优势, 减少分析开销, 能够更好或至少匹配类似研究的结果, 同时显著减少训练时间。

2 相关工作

目前深度学习被应用在很多领域, 如医疗^[11]、自动驾驶^[12]、图像识别^[13]以及自然语言处理^[14]等, 并且都展现出了良好的效果。在入侵检测领

域,也有很多研究人员使用深度学习算法对入侵检测进行了研究。

Zhao 等在文献[15]中提出了一项深度学习应用的最新研究。他们将传统的机器学习方法与4种常见的深度学习方法(自动编码器、受限玻尔兹曼机、卷积神经网络和递归神经网络)进行了实验比较。他们的研究得出结论,深度学习方法比传统方法更准确。Dong 等^[16]对传统的NIDS技术和深度学习方法的NIDS进行了比较并得出结论,基于深度学习的方法可以提高在各种样本数量和异常流量类型下的入侵检测准确率。同时,他们使用了合成少数过采样技术(SMOTE),证明了用过采样的方法可以克服与不平衡数据集相关的问题。

本文还对入侵检测算法中一些先进的深度学习算法进行了调研。Kim 等^[4]提出了一种利用深度神经网络(DNN)提高入侵检测技术的方法,使用了100个隐藏单元,对数据进行归一化预处理,采用ADAM优化器对模型进行优化。该方法在KDD CUP 99数据集上进行测试,准确率达到99%。同时,作者表示循环神经网络以及长短期记忆模型都是未来网络防御的需要。

Wang 等^[5]提出了一个基于层次时空特征的入侵检测系统(HAST-IDS),首先利用深层卷积神经网络(CNN)学习浅层特征,使用长短期记忆网络学习深层特征,整个过程是一个高效的自动化过程,减少了入侵检测系统的误报率。该方法应用在标准数据集DARPA1998和ISCX2012进行评估,证明该方法在特征学习方面效果显著。

Jia 等^[6]使用卷积神经网络构造入侵检测系统,将输入数据映射为二维灰度图像进行处理。在测试阶段,使用10个测试数据集对模型进行性能测试,实验结果与其他入侵检测系统进行比较,基于卷积神经网络的模型具有较高的检测率和精度,作者证明了卷积神经网络在入侵检测中的可行性。

深度学习方法在入侵检测领域取得了良好的效果,一些研究者已经证明分层的深度学习算法和浅层分类器相结合能够更好地对网络数据进行学习以及分类。Shone 等^[7]提出了一种深度学习方法用于入侵检测,使用一种非对称的栈式自编码

器对数据进行特征学习,随机森林对数据进行分类。他们使用KDD99和NSL-KDD数据集进行实验,准确率分别达到了97.75%和85.42%,还分别对模型五分类和十三分类进行了实验。作者表明深度神经网络和浅层分类器结合表现出的效果更好,更能适应现代网络的变化。

综上所述,尽管一些IDS已经取得了不错的成绩,但针对不同方向仍有很多改进的地方,包括监督学习对于数据的需求、需要消耗大量人力和成本、大量数据训练时间的优化,以及对不平衡数据中小样本的检测准确率低、检测模型的顽健性和可解释性等。大多数的研究人员仍然从已有标签数据出发研究问题,本文希望利用无监督的卷积自编码器的方式对数据进行处理,从提高入侵检测系统对现代网络流量的适应性以及检测效率的角度出发,进一步提高入侵检测的性能。

3 方法介绍

3.1 自编码器

自编码器是深度学习流行的算法之一,它是一种无监督的学习方式,和主成分分析的方法比较相近,但能够提供比主成分分析更强大的性能,已经被广泛应用在各个领域进行降维或者特征学习,如医疗^[17]、汽车驾驶^[18]等。数据经过训练之后自编码器能尝试将输入尽可能地复制到输出。自编码器中有输入层、隐藏层、输出层、组成深层的神经网络模型,其中,每个层级之间有连接,但每个层中的单元之间没有连接。隐藏层的维度通常比输入层维度小,输出层的维度通常和输入层的维度相同,构建出输入与输出之间复杂的非线性关系。自编码器主要由2部分组成:一个由函数 $h=f(x)$ 表示的编码器和一个由 $r=g(h)$ 表示生成重构的解码器。自编码器不是简单地对数据进行复制 $g(f(x))=x$,这样的输出是毫无意义的,需要对自编码器加一些约束,使它只能对数据进行近似的复制,它会自动优先选取重要的特征进行学习,因此适用于数据中的特征学习。将输出数据与输入数据进行对比,得到的误差进行反向传播计算,调整参数,优化模型。一个典型的自编码器如图1所示。

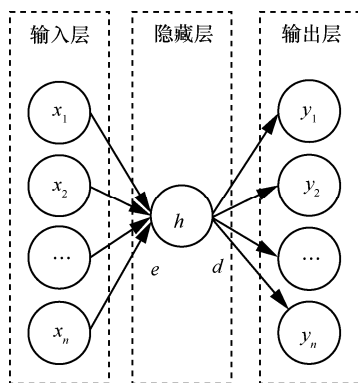


图 1 自编码器

隐藏层的作用是将高维数据转换为低维版本，这个阶段就是编码阶段，得到的数据是自动编码器获得的最显著特征，而不是原始数据本身。

自动编码器的目标函数是

$$y = h_{w,b}(x) \approx x \quad (1)$$

其中， h 是一个非线性的假设， w 和 b 参数分别表示权重和偏差。在整个过程中，它试图将学到的输入数据 y 尽可能地与输入数据 x 保持相似，它们之间的误差进行反向传递，其中，重构误差的函数为

$$L(x, d(e(x))) \quad (2)$$

其中， L 为重构误差函数，表示 $d(f(x))$ 与 x 之间的误差， e 是一个编码函数， d 是一个解码函数。

3.2 卷积自编码器

卷积自编码器是在自编码器的基础上建立起来的，它在自编码器的基础上添加了卷积操作。结合了卷积神经网络和自编码器的优点，解决了卷积神经网络对权重的敏感程度以及对大规模标记数据的依赖。同时也解决了深度信念网络和自编码器这种完全连接网络的一些缺陷，如相邻层之间的完全连接单元具有大量的训练参数。通过卷积核能够有效地将数据中的最优特征提取出来，使用深度训练的方式构造出深度的模型结构，输出高维数据的低维版本。使用多个隐藏层实现深度，每增加一个隐藏层表示更加复杂的特征，可以降低计算的成本、所需的训练数据量以及提供更大程度的精度。每一个隐藏层的输出作为更高阶层的输入，所以通常第一层用来学习输入数据，输出第一阶特征；第二层用来学习和第一阶特征相关的第二阶特征，以此类推，它的训练过

程和自编码器相似。假设的模型有 l 个卷积核，每个卷积核都由参数 ω^l 和 b^l 组成，使用 h^l 表示卷积层，则卷积层表示为

$$h^l = f(x * \omega^l + b^l) \quad (3)$$

其中，符号 $*$ 为卷积操作， f 为激活函数。将得到的 h^l 进行重构特征，可以得到如下表达式

$$y = f(h^l * \hat{\omega}^l + c^l) \quad (4)$$

其中， y 为重构之后的特征， $\hat{\omega}^l$ 表示第 l 个特征的权重 ω^l 的转置， c^l 表示偏置。将输入数据与输出数据结果距离进行比较，这里使用的是欧氏距离，使用梯度下降算法重构误差，调整参数进行优化。

$$E = \frac{1}{2} \sum (x_i - y_i)^2 \quad (5)$$

4 提出的方法

4.1 非对称的卷积自编码器

提高入侵检测效率是现代入侵检测技术的关键。因此笔者的目标是建立一个快速高效的入侵检测系统。本文介绍一种新的网络入侵检测系统，它是非对称的卷积自编码器。从根本上说，它是由编码器-译码器(对称)到只有编码器的过程(非对称)，主要是在特征进行学习的过程中，特征输出映射数目减小，神经网络结构会筛选出最优特征优先输出，从而模型结构学习到每一层的最优特征。并且，如果有正确的学习结构，可以减少计算和时间开销，提高模型的准确率和效率。非对称的卷积自编码器可以作为分层的特征提取器，它可以很好地缩放来适应高维数据的输入，不用担心高维数据的输入。它的主要训练过程和传统的卷积神经网络训练过程相似。图 2 展示了均衡卷积网络和非均衡卷积网络的区别。其中， h 代表维度缩减的隐藏层， e 代表编码阶段， d 代表解码阶段。

NCAE 的输入向量假设为 $x \in \mathbf{R}^1$ ，第一层隐藏层学习输入层的数据映射为 $h_1 \in \mathbf{R}^1$ ，其中，1 表示向量的维度。它的编码函数可以确定为

$$h_i = f(w_i h_{i-1} + b_i), i = 1, \dots, n \quad (6)$$

其中，当 $i=0$ 时， $h_0=x$ ， f 表示激活函数，这里使用 sigmoid 激活函数， n 代表隐藏层的个数。

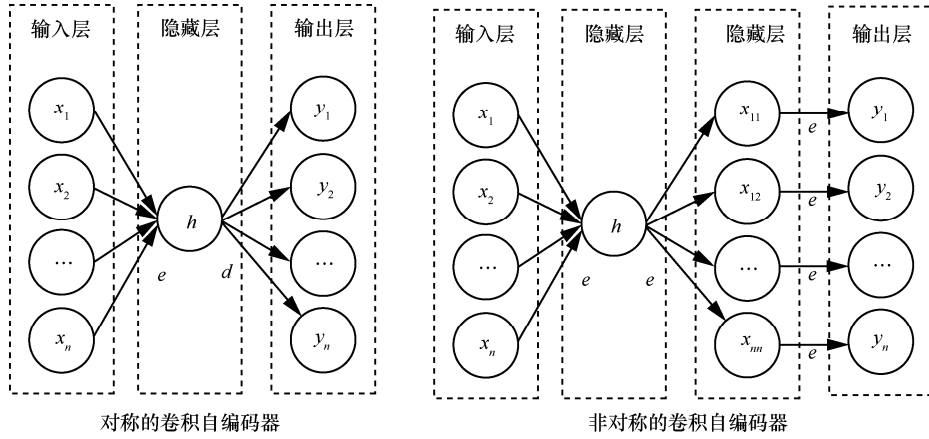


图 2 对称的卷积自编码器和非对称的卷积自编码器对比

Sigmoid 激活函数可以表示为

$$f(t) = \frac{1}{1 + e^{-t}} \quad (7)$$

输出数据可以表示为

$$h_n = f(w_n h_{n-1} + b_n) \quad (8)$$

模型在训练时进行反向传播, 调整误差, 非对称的自编码器的重构误差可以表示为

$$E(\alpha) = \sum_{i=1}^m (x_i - y_i)^2 \quad (9)$$

其中, m 代表训练样本, 模型通过最小化重构误差来调节参数, 从而达到最佳水平。

4.2 基于 NCAE-NSVM 的入侵检测模型

4.1 节详细介绍了非对称的卷积自编码器, 但使用单纯的非对称卷积自编码器相比浅层的分类器 (如 KNN[8] 以及 SVM[9] 算法), 它的准确率并没有很大提升。因此将深度学习和浅层学习算法相结合, 能够提升分类检测的准确率。SVM 算法是目前最流行的机器学习算法之一, 是基于统计学习方法中比传统方法更高效的一种方法。相比传统的其他分类方法, SVM 对小样本数据、高维数据展现出了良好的效果。然而, 随着大数据时代的到来, 数据维度越来越大, 数据量越来越多, 使用 SVM 分类需要的时间较长, 并且存在较高的错误率和低真正率。鉴于上述原因, 本文开发了一种新的基于 NCAE-NSVM 的算法来提高分类检测的准确率。图 3 中展示了模型的主要架构, 它是一个多层的无监督的神经网络, 主要分为 3 个步骤。第一步, 预处理阶段, 将数

据的稀疏特征进行合并, 数值化, 归一化数据。第二步, 特征提取阶段, 使用 4.1 节中提出的非对称卷积自编码器对数据进行特征提取。第三步, 分类阶段, 将提取的最优特征输入基于二叉树构建的多类 SVM 中进行逐层分类。

由于现存的数据以及模型中存在一些无法避免的问题: 1) 对于带标签的数据资源十分稀缺; 2) 在很多深度学习网络中, 误差函数是一个高度非凸的函数, 具有很多局部极值; 3) 在神经网络中易出现梯度弥散问题。因此, 使用无监督的非对称卷积自编码器进行特征提取可以有效解决以上问题。

支持向量机是一种高效的二分类机器学习算法, 相比传统的其他分类方法, SVM 对小样本数据、高维数据展现出了良好的效果。但大多数分类案例都是多分类的。如果使用 SVM 直接对数据进行多分类, 则训练时间通常很长。在本文中, 提取的特征被输入基于二叉树构造的多类 SVM 分类器中。用它来检测 k 分类数据, 需要用到 $k-1$ 个 SVM 分类器, 因此, 本文实现了 4 个 SVM 检测 5 种类型。其中的核函数是高斯核函数, 用来解决非线性样本的分类问题。在这个过程中, SVM 只输出 +1 和 -1。多类 SVM 的分类步骤如下所示。

1) 将得到的特征输入 SVM1 中, SVM1 首先判断数据是正常类型还是攻击类型, 如果是攻击类型, SVM 输出为 -1, 并且将攻击类型的数据输入 SVM2 中, 否则 SVM 的输出为 +1。

2) SVM2 接收到 SVM1 的输出数据之后, 判断得到的数据是 DoS 还是 Probe、U2R、R2L 中

的某个类型, 如果是 Probe、U2R、R2L 中的某个类型, SVM 输出为-1, 并将这类数据输入 SVM3 中, 否则 SVM 输出为+1。

3) SVM3 接收到 SVM2 的输出数据之后, 判断得到的数据是 Probe 还是 U2R、R2L 中的某个攻击类型, 如果是 U2R、R2L 中的攻击类型, SVM 输出为-1, 并将该类数据输入 SVM4 中进行分类, 否则 SVM 输出为+1。

4) SVM4 接收到 SVM3 的输出数据之后, 判断得到的数据是否是 U2R 还是 R2L, 如果是 U2R, 则 SVM 输出为+1, 否则 SVM 输出为-1。

模型中隐藏层的个数、每一层的输出特征映射维数以及 SVM 中的参数(核函数以及惩罚因子)都是由十折交叉验证得到的最优参数, 模型的准确度主要取决于模型的结构以及参数, 因此, 对模型进行参数的调节必不可少。

4.3 模型复杂度以及时效性

本文提出的 NCAE-NSVM 入侵检测模型的计算复杂度主要由非对称的卷积自编码器决定。单个卷积层的时间复杂度为 $O(M^2 K^2 I T)$, 其中, M 表示每个卷积核输出特征映射的边长, K 表示每个卷积核的边长, I 表示每一层卷积层的输入通道数, T 表示每一层卷积层的输出通道数。其中, M 的大小由输入数据的尺寸 L 、卷积核的大小 K 、Padding、卷积的步长 $Stride$ 决定, 即

$$M = \frac{L - K + 2 \times \text{Padding}}{\text{Stride} + 1}.$$

由单层复杂度可以计算得到模型整体的复杂度为 $O(\sum_{i=1}^N M_i^2 K_i^2 I_i T_i)$, 其中, N 表示模型具有的卷积层数, l 表示模型的第几个卷积层。而模型的空间复杂度主要由卷积核的大小 K , 每一层输入输出的通道数 I_i 、 T_i , 以及神经网络的深度决定 N , 与数据的大小无关, 可以表示为 $O(\sum_{i=1}^N K_i^2 I_i T_i)$ 。通过分析模型的时间复杂度和空间复杂度可以看出, 对于入侵检测系统具有较高的可实现性, 可以在任何操作系统上运行。另外, 模型不需要大量人工标记的数据以及预处理过程, 因此在现代网络入侵检测系统中面对未知的入侵事件模型具有较强的竞争力以及时效性。

5 实验

5.1 实验设计

本文的实验流程如图 4 所示, 首先将 KDD 99 数据集数值化特征其中的字符型数据, 对数据进行归一化操作。将获得的标准数据集输入卷积自编码器中进行特征提取, 使用多类 SVM 分类器进行训练和测试。根据初步实验结果分析模型的不足, 得到优化后的模型。在结果评价方面使用 Accuracy、漏报率、误报率等作

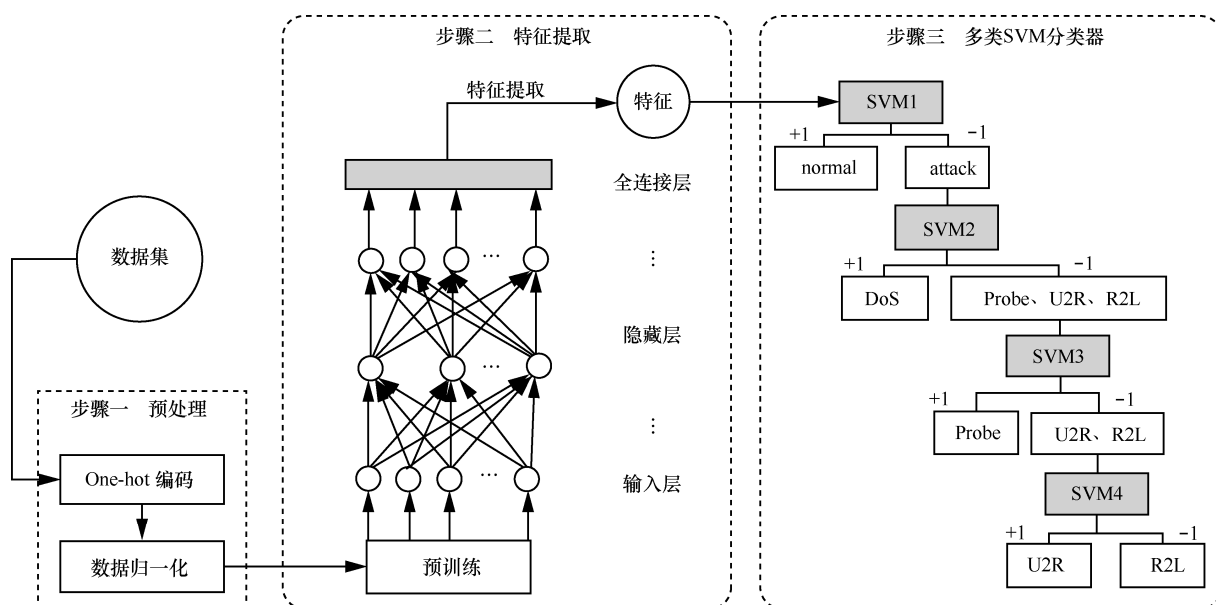


图 3 NCAE-NSVM 模型架构

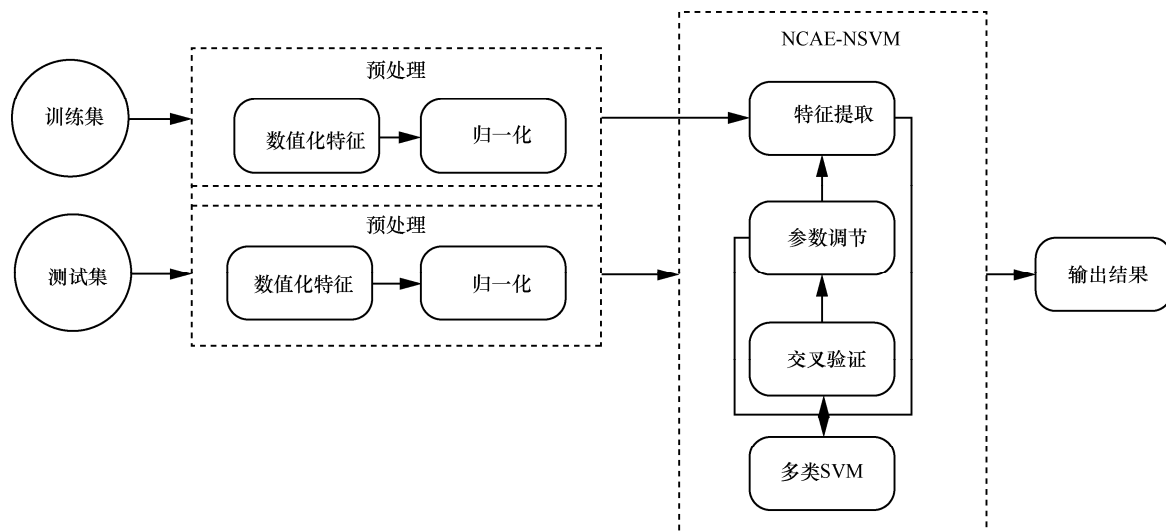


图4 实验流程

为评价标准，最后将本文的结果和其他模型的结果进行对比，展示不同机器学习算法的效果。

5.2 实验数据

本文中使用的 KDD99 数据集是美国空军 9 周收集的网络连接和系统审计数据，通过模拟各种用户类型、各种网络流量和攻击方法，以模拟真实的网络环境^[19]。它的训练数据包括 490 万单个连接数据，测试数据包含 200 万个网络连接数据。本文仅使用数据集的 10%进行训练，训练数据共有 494 021 条记录。这个数据集有 5 种类型：Normal、DoS、R2L、U2R、Probe。攻击类型共 39 种，其中，22 种类型的攻击出现在训练集中，另外 17 种未知类型出现在测试集中。表 1 展示了数据的详细信息。其中，在训练集中 Normal 类型有 97 278 条记录，Probe 类型有 4 107 条记录，DoS 有 391 458 条记录，U2R 有 52 条记录，R2L 有 1 126 条记录。在测试集中 Normal 类型有 60 593 条记录，Probe 有 4 166 条记录，DoS 有 229 853 条记录，U2R 有 228 条记录，R2L 有 16 189 条记录。

5.3 数据预处理

KDD99 数据集中每一条连接由 41 个特征组成，特征中含有很多非数值型的数据以及无用的特征，本文先对数据集进行预处理，将它们有效地转换成可以被模型接受的特征。预处理部分主要可以分为两步：数值化特征以及归一化。

表 1 KDD99 数据集数量

攻击类型	训练集/条	测试集/条
Normal	97 278	60 593
DoS	391 458	229 853
Probe	4 107	4 166
U2R	52	228
R2L	1 126	16 189

1) 数值化特征。基于神经网络的训练需要使用数值型的特征。因此在预处理阶段需要把非数值型的特征转换为数值型的特征。在 KDD 99 数据集中有 3 种协议类型，70 种服务符号取值和 11 种标签符号取值都是非数值型的，本文采用 One-hot 编码方式为它们建立相应的数值映射(如 TCP=[0,0,1], UDP=[0,1,0], ICMP=[1,0,0])转换为数值特征。

2) 归一化。因为 KDD99 数据中存在一些离散或连续的数值，它们的范围不同，使数据在各维度之间不存在可比性，规范化方法使用以下方法映射[0, 1]之间的数字属性。

$$X_k = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (10)$$

其中， X 是数据中某一维度的值， $X_{\min}$ 是该维度的最小值， $X_{\max}$ 是该维度的最大值， X_k 是最后得到的归一化之后的数据。

5.4 实验环境及参数

该模型的实验是在硬件环境 Intel(R) Core

i77700HQ CPU @2.80 GHz, 8 GB RAM, 1 TB 硬盘上以及 Windows10 操作系统上进行的, 使用编程语言 Python 3.5 实现了所提出的入侵检测模型。仿真环境配置如表 2 所示。

在模型中主要的参数变量包含卷积自编码器中每层输出的特征映射、学习率、步长以及支持向量机的核函数的参数和惩罚因子。参数的具体数值如表 2 所示。

表 2 实验变量参数

变量名称	变量值
第 1 层卷积层输出大小	10×10×64
第 2 层池化层输出大小	5×5×64
第 3 层卷积层输出大小	5×5×32
第 4 层池化层输出大小	3×3×32
第 5 层上采样层输出大小	5×5×64
第 6 层卷积层输出大小	5×5×64
第 7 层池化层输出大小	3×3×64
第 8 层卷积层输出大小	3×3×128
第 9 层池化层输出大小	2×2×128
全连接层	512
卷积核大小	3×3
学习率	0.001
卷积核步长	1
核函数的参数 σ	0.000 1
惩罚因子	1 000

5.5 评估指标

本文使用几个评估指标来评估模型的性能：准确率 (A)、精确率 (P)、召回率 (R)、误报率 (F_A) 和漏报率 (M_R)。准确率、误报率和漏报率用来评估整体的性能，精确率和召回率用来评估模型之间的比较。混淆矩阵如表 3 所示，它们的定义如下所示。

表 3 混淆矩阵

混淆矩阵		预测	
		1	0
实际	1	TP	FN
	0	FP	TN

TP: 1 类被正确预测为 1 类。

TN: 0 类被正确预测为 0 类。

FP: 0 类被错误预测为 1 类。

FN: 1 类被错误预测为 0 类。

准确率是指分类器分类正确的样本个数与样本总数之比，计算公式如下。

$$A = \frac{TP + TN}{TP + TN + FP + FN} \quad (11)$$

漏报率是分类器错误预测的样本与实际类别为 0 的所有样本的比率，公式如下。

$$M_R = \frac{FP}{FP + TN} \quad (12)$$

误报率的定义为

$$F_A = \frac{FN}{TP + FN} \quad (13)$$

召回率是指预测类别为 1 的类型被分类为 1，其计算公式如下。

$$R = \frac{TP}{TP + FP} \quad (14)$$

精确率是指预测为 1 的样本中被预测正确的概率，其计算公式如下。

$$P = \frac{TP}{TP + FN} \quad (15)$$

5.6 仿真实验及结果分析

实验 1 模型层数对检测结果的影响

深度神经网络中模型的层数对模型的检测程度都有重要影响。本文分析了非对称卷积自编码器的层数对结果的影响。它们的检测指标是准确率、误报率和漏报率。如表 4 所示，本文设计了 6 种不同的非对称卷积自编码器隐藏层层数，分别为 5、6、7、8、9、10 层，从上到下网络的层数逐渐增多，数据的检测结果一定程度上受到隐藏层个数的影响，随着隐藏层个数的提升，准确率也在提升，误报率、漏报率在减少，主要是因为隐藏层的个数越多，模型对数据的非线性拟合能力越强，越有利于高维数据向低维数据转换，可以提高模型检测的准确率以及分类速度。但是隐藏层的个数有最优值，在 9 层结构中神经网络的检测结果达到最优，准确率达到 97.71%，高于其他模型结构，误报率为 3.11%，漏报率 7.22%，均优于其他模型。因此，本文选择 9 层模型结构作为入侵检测系统中卷积自编码器的结构。这也说明神经网络的深度并不是越深越好，如果层数

过多,反而会降低准确率,提高误报率以及漏报率,使参数调节复杂度变高。

表4 非对称自编码器结构对检测结果的影响

层数	准确率	误报率	漏报率
5	88.27%	4.65%	9.86%
6	89.36%	3.96%	9.75%
7	96.57%	3.75%	8.33%
8	96.20%	3.46%	8.42%
9	97.91%	3.11%	7.22%
10	95.32%	3.23%	8.15%

实验2 迭代次数对性能的影响

模型的迭代次数决定了模型是否完全将数据中的特征进行学习,这是一个黑盒的过程,因此需要通过实验不断调整合适的参数。从图5可以看出模型迭代次数对检测损失值的影响。当迭代次数在10以内时,漏报率和误报率较高,主要是因为神经网络还没有学到所有的数据特征;当迭代次数在10~20之间时,模型基本能够学习到数据的所有特征,并且准确率保持在一个较高水平,误报率和漏报率保持较低水平;当迭代次数达到20以上时,准确率有所下降,误报率上升,出现了过拟合现象。因此,对于数据集的训练迭代次数应该保持在10~20之间,在这样的迭代次数下,模型能够学习到数据中的全部特征,并且保证模型不过拟合。

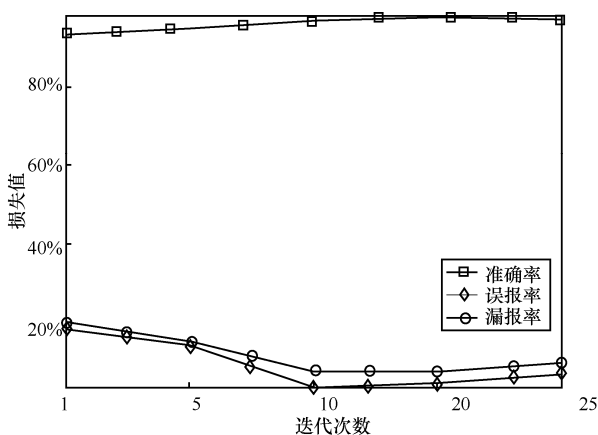


图5 迭代次数与损失值的关系

实验3 与其他模型的性能比较

经过广泛调查,本文选择了3篇与本文使用相同评估标准的文献(文献[7, 20-21])进行比较,

评估标准有准确率、精确率、召回率以及时间,如表5和图6中所示。表5中展示了每种算法对于数据集中每一类样本的分类情况,图6中对于每一类算法的测试时间进行了比较。对于Chi-SVM算法^[20],数据集中大类样本的分类情况不如本文中提出的算法,但是小类样本U2R和R2L的分类情况较好,但这个方法的测试时间较长,达到了10 224 s,本文的算法可以节约 $\frac{1}{10}$ 的时间。与SNDAE^[7]算法进行比较,可以看出在该算法的分类情况中,小类样本的分类情况不好,U2R的精确率、召回率以及准确率都为0,在海量的网络流量中,一些特殊的攻击流量往往在总量中占据少数数量,如果无法及时检测出它们,会给网络安全带来巨大威胁。DBN^[21]算法与本文中提出的算法比较,每一类的分类情况都比本文的算法相对低一些,并且测试时间也相对较长。

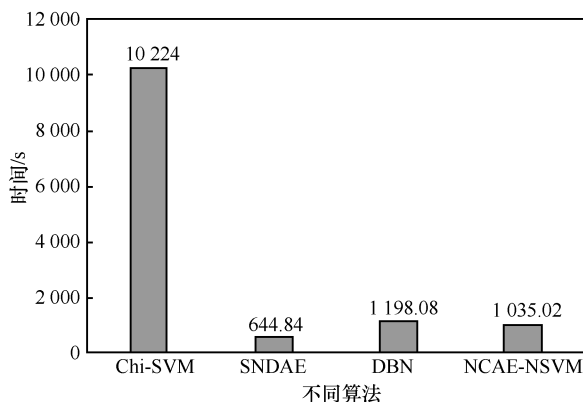


图6 不同算法之间的测试时间比较

综上所述,本文所提出的NCAE-NSVM算法结合了卷积神经网络和自编码器的优点,并且解决了卷积神经网络对权重的敏感程度以及对大规模标记数据的依赖。同时也解决了深度信念网络和自编码器这种完全连接网络的一些缺陷,如相邻层之间的完全连接单元具有大量的训练参数。通过实验证明,该方法摆脱了对已标记数据的依赖,不仅在准确率方面取得了不错的成绩,在训练时间上也有明显提升,在现代网络空间安全中的入侵检测系统中具有一定的优势以及及时性,且在数据集中的小样本检测率方面还有一定程度的提升。

表 5 非对称自编码器结构对检测结果的影响

攻击类型	Chi-SVM ^[20]			SND AE ^[7]			DBN ^[21]			NCAE-NSVM		
	A	P	R	A	P	R	A	P	R	A	P	R
Normal	96.10%	99.50%	99.60%	99.49%	100.00%	99.49%	99.49%	94.51%	99.49%	99.60%	100.00%	99.60%
DoS	98.87%	99.80%	99.90%	99.79%	100.00%	99.79%	99.65%	98.74%	99.65%	99.75%	99.64%	99.76%
Probe	95.80%	99.20%	99.20%	98.74%	100.00%	98.74%	14.19%	86.66%	14.19%	99.65%	99.75%	99.63%
R2L	76.92%	99.20%	98.70%	9.31%	100.00%	9.31%	89.25%	100.00%	89.25%	23.25%	99.32%	87.25%
U2R	96.37%	89.50%	73.90%	0.00%	0.00%	0.00%	7.14%	38.46%	7.14%	8.50%	20.58%	45.53%
合计	—	—	—	97.85%	99.99%	97.85%	97.90%	97.81%	97.91%	97.91%	99.67%	97.91%

6 结束语

本文对入侵检测领域中的深度学习算法进行了广泛调查,现代网络流量数量增多,并且出现了很多新的协议,现有的入侵检测系统大多基于已有标记的数据进行检测,无法识别出网络流量中新的攻击类型,对于未来的可扩展性显现出一定的局限性,并且检测效率相对较低,需要专家操作对数据进行标记和预处理,严重威胁了网络空间安全。基于以上原因,提出了一种新的网络入侵检测系统 NCAE-NSVM,该模型同时弥补了卷积神经网络和自编码器的缺点:卷积神经网络对权重过于敏感以及极度依赖大规模标记数据进行训练;自编码相邻层之间参数过多;使用非对称的卷积自编码器的数据进行特征提取,多类 SVM 对数据进行分类。经过实验验证,模型的整体准确率达到 97.91%,同时显著减少了训练的时间。本文的主要贡献在于提出了模型的完整思路以及实验设计,实现了大规模入侵检测样本的特征提取与分类检测的模型,最终经过实验验证取得了良好的效果。未来笔者将继续在这一方面进行深入研究,后续的工作需要提高样本的数量并且尝试使用本文的模型测试其他数据集,在短训练时间内,进一步提升数据中小样本的分类情况,使入侵检测技术得到进一步提升。

参考文献:

- [1] 赛门铁克. 互联网安全威胁报告[R]. 2018. Symantec. Internet security threat report[R]. 2018.
- [2] LIAO H J, LIN C H R., LIN Y C, et al. Intrusion detection system: a comprehensive review[J]. J Netw Comput Appl, 2013, 36(1): 16-24.
- [3] CHOWDHURY M M U, XIN C, LI J, et al. A few-shot deep learning approach for improved intrusion detection[C]//IEEE Uemcon. 2017.
- [4] KIM J, SHIN N, JO S Y, et al. Method of intrusion detection using deep neural network[C]//2017 IEEE International Conference on Big Data and Smart Computing (BigComp). 2017: 313-316.
- [5] WANG W, SHENG Y, WANG J, et al. HAST-IDS: learning hierarchical spatial-temporal features using deep neural networks to improve Intrusion detection[J]. IEEE Access, 2018, 6(99):1792-1806.
- [6] JIA F, KONG L Z. Intrusion detection algorithm based on convolutional neural network[J]. Beijing Transaction of Beijing Institute of Technology, 2017, 37(12):1271-1275.
- [7] SHONE N, NGOC T N, PHAI V D, et al. A deep learning approach to network intrusion detection[J]. IEEE Transactions on Emerging Topics in Computational Intelligence, 2018, 2(1):41-50.
- [8] DAVE D, VASHISHTHA S. Efficient intrusion detection with KNN classification and DS theory[C]//All India Seminar on Biomedical Engineering 2012 (AISOB 2012). Springer, 2013:173-188.
- [9] ABUROMMAN A A, REAZ M B I. A novel SVM-kNN-PSO ensemble method for intrusion detection system[J]. Applied Soft Computing, 2016, 38: 360-372.
- [10] HOU S, SAAS A, CHEN L, et al. Deep4MalDroid: a deep learning framework for android malware detection based on Linux kernel system call graphs[C]//2016 IEEE/WIC/ACM International Conference on Web Intelligence Workshops (WIW). 2016: 104-111.
- [11] SHEN D, WU G, SUK H I. Deep learning in medical image analysis[J]. Annual Review of Biomedical Engineering, 2017, 19(1): 221-248.
- [12] LIU H, TANIGUCHI T, TANAKA Y, et al. Essential feature extraction of driving behavior using a deep learning method[C]//Intelligent Vehicles Symposium. 2015:1054-1060.
- [13] GRM K, ŠTRUC V, ARTIGES A, et al. Strengths and weaknesses of deep learning models for face recognition against image degradations[J]. Iet Biometrics, 2018, 7(1):81-89.
- [14] GARDNER M, GRUS J, NEUMANN M, et al. AllenNLP: a deep semantic natural language processing platform[J]. Computer Science Bibliography, 2018.
- [15] ZHAO R, YAN R, CHEN Z. Deep learning and its applications to machine health monitoring: a survey[J]. IEEE Transactions on Neural Networks and Learning Systems, 2016, 14(8):1-14.
- [16] DONG B, WANG X. Comparison deep learning method to traditional methods using for network intrusion detection[C]//2016 8th IEEE International Conference on Communication Software and Networks (ICCSN). 2016: 581-585.
- [17] XU J, XIANG L, LIU Q, et al. Stacked sparse autoencoder (SSAE) for

nuclei detection on breast cancer histopathology images[C]//IEEE International Symposium on Biomedical Imaging. 2014:119-130.

- [18] DONG W, YUAN T, YANG K, et al. Autoencoder regularized network for driving style representation learning[J]. arXiv: 1701.01272, 2017.
- [19] TAVALLAEI M, BAGHERI E, LU W, et al. A detailed analysis of the KDD CUP 99 data set[C]//IEEE International Conference on Computational Intelligence for Security & Defense Applications. 2009:1-6.
- [20] THASEEN I S, KUMAR C A. Intrusion detection model using fusion of chi-square feature selection and multi class SVM[J]. Journal of King Saud University - Computer and Information Sciences, 2016, 29(4).
- [21] ALRAWASHDEH K, PURDY C. Toward an online anomaly intrusion detection system based on deep learning[C]//15th IEEE International Conference on Machine Learning and Applications (ICMLA). 2016: 195-200.

[作者简介]



王佳林 (1994-), 女, 山西太原人, 北京交通大学硕士生, 主要研究方向为信息安全和人工智能安全。



陈彤 (1993-), 女, 天津人, 北京交通大学博士生, 主要研究方向为信息安全和人工智能安全。



刘吉强 (1973-), 男, 山东海阳人, 博士, 北京交通大学教授、博士生导师, 主要研究方向为隐私保护、可信计算、安全协议设计与分析。



童恩栋 (1986-), 男, 山东聊城人, 博士, 北京交通大学讲师, 主要研究方向为智能信息处理、网络空间安全。



赵迪 (1995-), 女, 河北承德人, 北京交通大学硕士生, 主要研究方向为信息安全和人工智能安全。



牛温佳 (1982-), 男, 宁夏银川人, 博士, 北京交通大学教授、博士生导师, 主要研究方向为人工智能安全。