

引用格式:江泽涛,翟振宇. 基于二次决策的深度学习入侵检测模型[J]. 微电子学与计算机, 2020, 37(4): 32-36. [JIANG Z T, ZHAI Z Y. Based on twice decisionfor deep learning intrusion detection model [J]. Microeletronics & Computer, 2020, 37(4): 32-36.]

基于二次决策的深度学习入侵检测模型

江泽涛, 翟振宇

(桂林电子科技大学 广西图像图形与智能处理重点实验室, 广西 桂林 541000)

摘 要: 针对深度神经网络用于入侵检测方法时存在训练过程中由于数据不平衡和特征冗余两大问题而导致的低检测率和高误报率,提出一种基于二次决策的深度学习模型(TDDL). 该模型由深度堆栈自动编码器(DSAE)和神经网络结合,包括二个阶段特征学习,其中第一阶段使用 DSAE 对特征压缩并加入区分异常数据的概率值特征,第二阶段使用神经网络(NN)接收第一阶段的特征并训练,从而降低特征冗余和平衡对正常数据的偏向,以提高检测效果. 经 KDDCUP99 数据集进行实验测试,仿真实验结果表明,该模型能有效提升深度神经网络在入侵检测数据上特征学习的效果,使其具有更高的准确率的同时,还具有较低的误报率.

关键词: 深度学习;入侵检测;自动编码器;特征学习

中图分类号: TP309

文献标识码: A

文章编号: 1000—7180(2020)04—0032—05

Based ontwice decision for deep learning intrusion detection model

JIANG Ze-tao, ZHAI Zhen-yu

(Guilin University of Electronic Technology Guangxi Key Laboratory of Imageand
Graphic Intelligent Processing,Guilin University of Electronic Technology, Guilin 541000,China)

Abstract: Aiming at the intrusion detection method of deep neural network, there are two problems of data imbalance and feature redundancy in the training process, resulting in low detection rate and high false positive rate. Based on Twice Decision Deep Learning model(TDDL) is proposed; The model is a combination of Deep Stack Autoencoder(DSAE) and neural network, including two-stage feature learning, in which the first stage uses DSAE to compress features and add probability value features that distinguish abnormal data, and the second stage uses neural networks (NN) receives the characteristics of the first stage and trains, thereby reducing the feature redundancy and balance bias on normal data to improve the detection effect. The experimental results of KDDCUP99 dataset show that the model can effectively improve the effect of deep neural network on feature detection of intrusion detection data, so that it has higher accuracy and lower false positive rate.

Key words: deep learning; intrusion detection; self-encoder; feature learning

1 引言

网络入侵检测系统(NIDS)是一种强大的防御工具,可以抵御复杂的攻击和威胁^[1]. 但是网络入侵检测目前还在许多问题和挑战^[2,3,4]. 首先恶意攻击

和威胁的多样性和可变性. 现有的入侵检测方法无法应对网络威胁形势的不断演变和新威胁的出现;其次在网络入侵检测方法中使用的经典机器学习方法存在一些困难,例如由于不相关或冗余的特性而导致的过度拟合和高偏差,以及网络流量的类分布

收稿日期: 2019—07—22; 修回日期: 2019—09—08

基金项目: 国家自然科学基金(61876049,61762066, 61572147)、广西科技计划项目(AC16380108)、广西图像图形智能处理重点实验室项目(GIIP201701、GIIP201801、GIIP201802、GIIP201803)广西研究生教育创新计划项目(2019YCX043)

不平衡^[5].其次与标记用于开发 NIDS 的流量数据集的复杂性有关^[6],在一段时间内,需要大量的工作来生成这样的标记数据集.这些问题导致入侵检测方法在大规模网络环境下无法有效地检测入侵行为,同时也无法有效地学习特征表示,以构建更复杂的学习过程.

Hinton 等人^[7]提出了深度学习的概念,以优化学习问题,通过在分类任务之前以无监督的方式对网络的每一层进行预先培训,以学习数据的歧视性表示.此外,Vincent 等人^[8]提出了一种基于叠加去噪自动编码器(SDAE)的方法,使神经网络能够学习有用的特征表示.系统模型训练数据中的特征冗余是造成实时 NIDS 过拟合问题的主要原因之一,但利用有效的特征学习,减少了训练数据中可能导致过拟合问题的冗余或无关特征.

为了从大量未标记的特征中学习更有用的特征表示,提高现有网络入侵检测系统的灵活性,本文提出一种基于二次决策的深度学习模型,该模型由深度堆栈自动编码器和神经网络结合,可以在正常和异常状态检测两种操作模式下检测网络入侵,以及检测正常状态和其他类型的攻击.

2 自动编码器

深度堆栈式自动编码器模型(DSAE)的基本单元是自动编码器(AE).AE本质上是一个前馈神经网络,类似于多层感知器(MLP),由三个部分组成,即输入层、一个或多个隐藏层和输出层,其中输出层和输入层中的神经元数量相等.图1显示了由单个隐藏层组成的简单AE的典型结构.

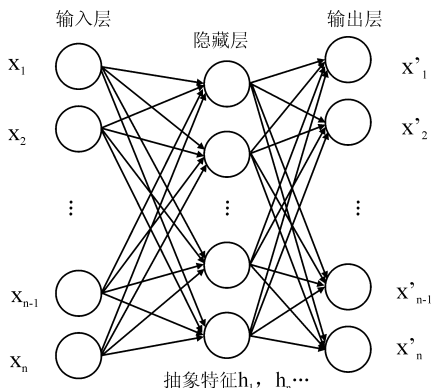


图1 AE的基本架构

自动编码器的学习是无监督学习,因为自动编码器学习抽象的压缩数据表示来重建原始输入数据,而不是猜测输入的目标输出.自动编码器包括两个过程:编码过程发生在输入层和隐藏层之间,解码

过程发生在隐藏层和输出层之间.

输入数据和隐藏层具有抽象压缩数据的节点.输出层的重构值近似等于输入值,其表达式如下:

$$x_i \cong \bar{x}_i, \quad \text{for } i \in n \quad (1)$$

在AE的编码过程中,抽象压缩数据的计算如下:

$$h = g(xw + b_x) \quad (2)$$

式中, w, b_x 和 x 分别是输入层的权重矩阵、偏移向量和输入向量.另一方面,AE的解码过程可用于计算重建数据 $\bar{x}$:

$$\bar{x} = g(hw^T + b_y) \quad (3)$$

式中, w^T, b_y 和 x 分别是输入层的权重矩阵、偏移向量和隐藏层的抽象压缩数据向量.式(1)和式(2)中使用的函数可以是线性或非线性激活函数.在本模型中,使用了一个sigmoid函数,计算如下:

$$g(x) = 1/(1 + e^{-x}) \quad (4)$$

3 TDDL 模型

本文提出的TDDL是由两个神经网络结构组成的深度学习模型(TDDL模型总框架如图2);其创新思想在于采用特征降维和异常概率输出的第一阶段和神经网络分类的第二阶段.第一阶段是使用深度堆栈式自动编码器(DSAE),由于其性能好和实时分类的速度快,将其用于初始阶段,网络流量可分为正常流量和异常流量,并具有概率分值.该概率值作为一个附加特征,用于分类的最终阶段.主要分为两个任务,构造有监督学习的DSAE模型和输出异常概率值P:首先以无监督的学习方式学习了无标签数据的 w 和 b 值,并将其应用到无标签数据上,进行逐层的无标签预训练;然后以有监督的学习方式使用 w 和 b 值,将预训练所得的特征与标签向量一起传入基于softmax回归分类器,使用反向传播算法对标记数据进行分类 (x, y) ,在这一阶段完成对预训练所得的 w 和 b 值进行微调,同时,通过softmax分类器输出一个异常概率值P.下面是从模型的总框架到各个阶段的详述:

3.1 TDDL模型总框架

3.2 预处理阶段

KDDCUP99训练数据的每个连接记录包含了41个固定的特征属性和1个标签.类标识用来表明这条记录是正常的还是异常的.41个固定的属性中9个特征属性是离散型,其余均为连续的.模型训练

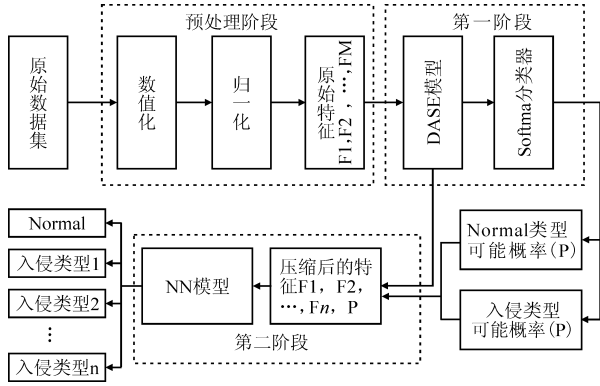


图2 TDDL模型总框架图

之前,首先要进行原始数据做预处理,分为以下两步:

(1)数值化处理:其中的三种字符型特征,协议类型(Protocol Type),网络服务(Service)以及连接状态(flag),分别转化为数值.如3种Protocol Type类型分别编码为0-2;TCP编码为0,UDP为1,TCMP为2;同理,67种Service类型分别编码为0-66,11种flag类型编码为0-10,最后5类数据的标签则是0-4.

(2)标准化和归一化处理:

为了消除不同特征值度量单位的差异对实验结果产生的影响.将已经过(1)步骤的数据,进行标准化和归一化处理.

标准化处理公式:

$$x' = \frac{x - \bar{x}}{\sigma} \quad (5)$$

公式中, x 为数值化后的特征值, $\bar{x}$ 为特征的平均值, σ 为特征的标准差, x' 为标准化后的特征值.

归一化处理公式:

$$x'' = \frac{x' - x'_{\min}}{x'_{\max} - x'_{\min}} \quad (6)$$

式中, $x'_{\max}$ 是标准化后特征中的最大值, $x'_{\min}$ 是最小值, x'' 为归一化的特征值.

3.3 构造 DSAE 模型

如图3所示,第一阶段将由两个隐藏层组成的 DSAE 与它们上面的 softmax 层分类器结合在一起.为了建立所提议的模型,在培训阶段应该执行两个任务.第一个任务是预培训任务,其中 DSAE 的每一层

都使用无监督学习技术进行了单独的预训练.原始的未标记特征是第一层的输入(Input),输出压缩特征(code)是第二层的输入.当第一层和第二层分别进行预训练时,将它们堆叠在一起,并分别进行

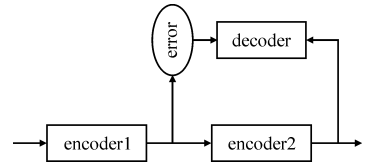


图3 DSAE模型

训练.在每一层,重建输入特征所产生的误差被最小化.

图4是无监督逐层训练过程的原理,无监督学习完成以后,自编码器的解码器就不需要了,直接将encoder1的输出传入encoder2进行第二次编码.

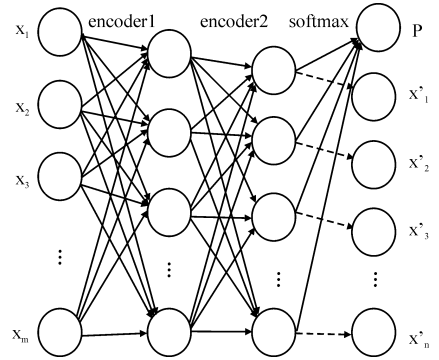


图4 无监督逐层训练过程

最后是微调任务,在顶部添加一个 softmax 层,作为下一个任务的分类器任务.在这项任务中,利用一种基于反向传播算法的监督学习技术,来逐层调整从先前任务中获得的预先训练的 DSAE 模型的参数.因此,该模型采用半监督学习技术进行训练.微调任务的目的是使用数据的标签来最小化预测误差.这样就将自编码器转化为了一个有监督的深度学习,其最后的分类器,可以使用 svm, softmax 等,本文采用的是 softmax 分类器.

3.4 输出第一阶段的特征

p 是 softmax 分类器输出可能是正常和异常数据的概率.对于一个 n 分类问题,给定输入 x 属于第 i 类(y_i)的一种原始度量 $h(x, y_i)$, softmax 计算的属于某一类的概率为:

$$p(y | x) = \frac{e^{h(x, y_i)}}{\sum_{j=1}^n e^{h(x, y_j)}} \quad (7)$$

初始训练阶段完成后,该模型模型能够对未发现的网络流量实例进行分类.第一阶段旨在通过将更多的注意力放在异常流量上,以对其他类型的攻击进行分类,从而减少过度拟合问题并平衡对正常流量的偏向.所以在初始训练阶段使用带标签的数

据进行微调的过程中,仅将标签分为正常和入侵数据两类.

第二阶段由神经网络(NN)处理第一阶段的压缩特征和附加 p 概率特征,理论上来说,只要是神经网络都可以作为这一层的算法处理. 本文使用 BP 神经网络和卷积神经网络(CNN)作为第二阶段的算法,并进行实验对比.

4 实验分析

4.1 数据描述:

本文的数据集采用的是 KDD99 数据集. 该数据集将网络行为分为 5 类: Normal, DOS, R2L, U2R, Probe. 每一个行为用 41 个特征表示. 本文使用 10% 的 KDD99 数据集进行模型的训练. 这部分数据集包含 494 021 条训练样本和 311 029 条测试样本,其中各类入侵行为分布如表 1 所示:

表 1 数据样本描述			
攻击类型	训练集数量	测试集数量	测试集各类比例(%)
Normal	97 278	60 593	19. 48
DOS	391 458	229 853	73. 39
R2L	1 126	16 189	5. 32
U2R	52	228	0. 07%
Probe	4 107	4 166	1. 34

KDDCUP99 数据集是网络入侵检测领域的公认测试数据集. 从表中的数据, 我们可

以看出, 该数据集具有相当的数据不平衡性, 在使用深度神经网络的训练过程中, 会因为数据的不平衡性, 导致对数据量较少的攻击类型具有较低的检测率.

4.2 实验评测

本实验采用准确率和误报率作为评价指标来衡量模型的效果.

表 2 中对比与其他的神经网络结合所带来的优化效果, 实验使用传统的 BP 神经网络和 CNN 分别与这两种网络作为第二阶段的 TDDL 模型比较, 可以看出本文提出的模型将通过第一阶段的预训练过程, 将原有的神经网络作为第二阶段的分类算法, 可以有效提升原有的神经网络的效果. 实验数据用 10 次数据取平均值作为最后的结果.

实验表明, 在提升了两个神经网络检测率的同时降低了其误报率. 特别是 U2R 和 R2L 这两个样本较少的攻击类型的检测率也有了很大幅度的提升. 说明第一阶段使用 DSAE 做预训练, 进行特征

的提取, 同时加入了概率特征 P, 使得第二阶段的神经网络在对入侵数据的分类上有很大的提高.

表 2 神经网络提升效果对比较 (%)							
方案	检测率					准确率	误报率
	Normal	DOS	Probe	U2R	R2L		
BPNN	96. 82	98. 74	85. 38	48. 32	52. 41	93. 54	1. 21
TDDL(BP)	99. 84	99. 19	95. 71	85. 48	90. 59	99. 73	0. 36
CNN	97. 48	98. 95	93. 64	54. 43	63. 32	96. 28	1. 77
TDDL(CNN)	99. 58	99. 20	98. 97	91. 49	85. 44	99. 64	0. 23

表 3 对比了本文的模型与其他机器学习算法的检测效果, 可以看出本文模型的检测效果要高于其他算法的效果, 因此, 本文提出的模型是有效的.

表 3 与部分算法的对比 (%)		
模型	准确率	误报率
DNN4 ^[9]	93. 49	0. 76
KPCA-BP ^[10]	98. 01	0. 97
k-means+CART ^[11]	98. 56	1. 43
TDDL(BP)	99. 73	0. 36
TDDL(CNN)	99. 64	0. 23

5 结束语

本文提出的一种基于二次决策的深度学习入侵检测模型, 能够较大的改善传统的深度神经网络用于入侵检测领域, 在训练过程中由于数据不平衡和特征冗余两大问题, 导致的低检测率和高误报率. 仿真实验结果表明, 该模型能有效提升传统神经网络的效果, 使其具有更高的准确率, 同时降低了误报率.

参考文献:

[1] WU S X, BANZHAF W. The use of computational intelligence in intrusion detection systems: A review [J]. Applied soft computing, 2010, 10(1): 1-35.

[2] CAI Z, WANG Z, ZHENG K, et al. A Distributed TCAM Coprocessor Architecture for Integrated Longest Prefix Matching, Policy Filtering, and Content Filtering [J]. IEEE Transactions on Computers, 2013, 62(3):417-427.

[3] ZHENG K, CAI Z, ZHANG X, et al. Algorithms to speedup pattern matching for network intrusion detection systems [J]. Computer Communications, 2015, 62:47-58.

[4] YU Y, LONG J, LIU F, et al. Machine learning combining with visualization for intrusion detection: a survey[C]//International Conference on Modeling De-

- cisions for Artificial Intelligence. Springer, Cham, 2016: 239-249.
- [5] SOMMER R, PAXSON V. Outside the closed world: On using machine learning for network intrusion detection[C]//2010 IEEE symposium on security and privacy. IEEE, 2010: 305-316.
- [6] YU Y, LONG J, CAI Z. Network intrusion detection through stacking dilated convolutional autoencoders[J]. Security and Communication Networks, 2017: 1-10.
- [7] HINTON G E, OSINDERO S, TEH Y W. A Fast Learning Algorithm for Deep Belief Nets[J]. Neural Computation, 2006, 18(7):1527-1554.
- [8] VINCENT P, LAROCHELLE H, LAJOIE I, et al. Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion[J]. Journal of Machine Learning Research, 2010, 11(12):3371-3408.
- [9] GAO N, GAO L, GA Q, et al. An intrusion detection model based on deep belief networks[C]//2014 Second International Conference on Advanced Cloud and Big Data. IEEE, 2014: 247-252.
- [10] Hongwei D, Liang W. Research on Intrusion Detection Based on KPCA-BP Neural Network[C]//2018 IEEE 18th International Conference on Communication Technology (ICCT). IEEE, 2018: 911-915.
- [11] AUNG Y Y, MIN M M. Hybrid Intrusion Detection System Using K-Means and Classification and Regression Trees Algorithms[C]//2018 IEEE 16th International Conference on Software Engineering Research, Management and Applications (SERA). IEEE, 2018: 195-199.

作者简介:

江泽涛 男,(1961),博士生导师,教授. 研究方向为访问控制与信息安全及计算机视觉.

翟振宇(通信作者) 男,(1994),硕士研究生. 研究方向为信息安全.

E-mail: 731877986@qq.com