

文章编号:1001-9081(2015)09-2513-06

doi:10.11772/j.issn.1001-9081.2015.09.2513

多层极限学习机在入侵检测中的应用

康松林*, 刘乐, 刘楚楚, 廖锐

(中南大学 信息科学与工程学院, 长沙 410083)

(*通信作者电子邮箱 sunkang@mail.csu.edu.cn)

摘要:针对神经网络在入侵检测应用存在的维度高、数据大、获取标记样本难、特征构造难、训练难等问题,提出了一种基于深度多层极限学习机(ML-ELM)的入侵检测方法。首先,采用多层网络结构和深度学习方法抽取检测样本最高层次的抽象特征,用奇异值对入侵检测数据进行特征表达;然后,利用极限学习机(ELM)建立入侵检测数据的分类模型;其次,利用逐层的无监督学习方法解决入侵检测获取标记样本难的问题;最后采用KDD99数据集对该方法的性能进行了验证。实验结果表明:多层极限学习机的方法提高了检测正确率,检测漏报率也低至0.48%,检测速度比其他深度模型的检测方法提高了6倍以上。同时在极少标记样本的情况下仍有85%以上的正确率。通过多层网络结构的构建提高了对U2L、R2L这两类攻击的检测率。该方法集成深度学习和无监督学习的优点,能对高维度,大数据的网络记录用较少的参数得到更好的表达,在入侵检测的检测速度以及特征表达两个方面都具有优势。

关键词:入侵检测;高维度;大数据;标记样本;特征构造;训练;多层极限学习机

中图分类号: TP393.08 文献标志码:A

Intrusion detection based on multiple layer extreme learning machine

KANG Songlin*, LIU Le, LIU Chuchu, LIAO Qin

(School of Information Science and Engineering, Central South University, Changsha Hunan 410083, China)

Abstract: In view of high dimension, big data, the difficulty of getting labeled samples, the problem of feature expression and training existed in the application of neural network in intrusion detection, an intrusion detection method based on Multiple Layer Extreme Learning Machine (ML-ELM) was proposed in this paper. Firstly, the highest level abstract features of the detection samples were extracted by multi-layer network structure and deep learning method. The characteristics of intrusion detection data were expressed by singular values. Secondly, the Extreme Learning Machine (ELM) was used to establish the classification model of intrusion detection data. Then, the problem that hard to obtain labeled samples was solved by using a layer by layer unsupervised learning method. Finally, the KDD 99 dataset was used to test the performance of ML-ELM. The experimental results show that the proposed model can improve the detection accuracy, and the false negative rate of detection is low to 0.48%. The detection speed can be improved by more than 6 times compared with other depth detection methods. What's more, the detection accuracy is still more than 85% in the case of a few labeled samples. The detection rates of U2L attack and R2L attack are improved by constructing multi-layer network structure. The method integrates the advantages of deep learning and unsupervised learning. It can express these features of high dimension and large data well using fewer parameters. It also has a good performance in intrusion detection rate and characteristic expression.

Key words: intrusion detection; high dimension; big data; labeled sample; feature expression; training; Multiple Layer Extreme Learning Machine (ML-ELM)

0 引言

随着网络与信息技术日新月异的发展,网络安全直接关系到国家安全和社会稳定。基于网络安全的迫切需要和现有入侵检测系统的弊端,入侵检测技术的发展与革新势在必行。

现有的入侵检测主要分为以下几种:现如今最常见的入侵检测方法就是模式匹配、统计协议分析、机器学习以及级联入侵检测等。这些方法在原有的基础上有了一定幅度的改善但是仍存在一些缺点。比如在文献[1]和文献[2]中提到的就是典型时间序列分析模型的入侵检测方法。虽然统计分析

模型可以智能地找出用户操作行为的规律,但用户的行为是非常复杂的,如何选择一个合适的特征量进行监测是很难有一个通用的准则。又比如文献[3]中提出用模糊逻辑分类器与遗传算法相结合应用于入侵检测,虽然可以达到较高的正确率,但存在模糊描述语义不容易被理解,遗传算法对处理高维大规模数据有很大的局限性等问题。还有一种应用广泛的是基于进化计算的入侵检测。比如文献[4]中基于粒子群优化算法的特征选择,特征选择和不选择代码0和1,但光是KDD(Knowledge Discovery in Database)数据集有41个属性,如果所有的特征用它编码,计算量太大,很容易造成信息冗

收稿日期:2015-04-20;修回日期:2015-06-14。 基金项目:国家自然科学基金资助项目(60773013)。

作者简介:康松林(1968-),男,湖南新化人,副教授,硕士,主要研究方向:网络信息安全; 刘乐(1991-),女,湖南邵阳人,硕士研究生,主要研究方向:网络信息安全; 刘楚楚(1990-),女,湖南湘潭人,硕士研究生,主要研究方向:网络信息安全; 廖锐(1990-),男,湖南长沙人,硕士研究生,主要研究方向:数据挖掘。

余。还有一种综合多种分类器的级联入侵检测系统^[5-6]。这种系统集成多种分类器的优点,但是会造成时间与成本的增加,并且并不一定适用于所有的攻击类型。

每种方法各有它的优缺点。其中,神经网络因具有很好的智能性,被广泛应用到入侵检测中。但神经网络在入侵检测中的应用存在几个问题:

1) 随着网络用户的激增,海量的网络记录无疑给人侵检测加上了大数据的标签。对于传统的入侵检测系统而言,则需要添加大量的训练样本以应对复杂多变的攻击,而这样大规模的样本数据会造成训练时间和样本储存空间的急剧增长,给训练增加难度。近年来,在大数据的时代背景下,深度学习持续升温^[7]。相比浅层构架建模以及表现能力的受限,深度学习通过组合低层特征形成更加抽象的高层表示属性类别或特征,对事物具有更本质的刻画。深度学习已在语音识别、图像识别等领域有着大量的应用,但鲜有基于深度学习的入侵检测的研究,并且深度神经网络如深信度网络(Deep Belief Network, DBN)^[8]、深度玻尔兹曼机(Deep Boltzmann Machine, DBM)^[9]的训练速度是非常慢的。

2) 神经网络因其结构与训练方法的限制,收敛速率低,训练时间过长。为了解决权值难训练的问题,2004 年 Huang 等^[10]提出了一种操作简单、高效稳定的极限学习机算法。因为极限学习机具有很好的学习速度与泛化能力,在很多领域都得到了应用,比如图像识别、故障诊断等。但极限学习机在网络安全领域鲜有研究。

3) 在实际的入侵检测中,攻击特征之间常呈现高维度的特点。一种常见的方法是对高维空间进行主成分提取达到降维的目的。比如 Kuang 等^[11]提到的用基于核的主成分分析法进行特征提取实现降维,得到了较好的分类效果,但是在求解主成分分析法的过程中,当网络的原始数据维数极高时,会使自协方差矩阵维数过高,从而在求解特征值和特征向量时耗时过多,导致此方法效率不高。还有一些方法是根据记录的统计信息进行特征选择达到降维的目的。比如 Arya 等^[12]提到用互信息删除冗余特征实现降维,效率得到了提高,但是单纯用网络记录的统计信息进行特征选择不具有很好的智能性,无法应对当前隐蔽性高、复杂度高、变化快速的攻击环境。再比如一些流形学习算法在入侵检测特征提取中的应用,如局部线性嵌入(Locally Linear Embedding, LLE)算法^[13-14],因其存在计算量大且因受邻域参数影响较大导致稳定性较弱等问题而有待改进。还有基于复杂网络理论和最小生成树算法^[15]的特征提取方法,虽然训练时间与主成分分析法相比有所减少,但是从判别效果来看并没有主成分分析法的好。

4) 在现实网络环境中,大部分数据是不包含标签属性的,标签的获取需要大量的人力物力。而有监督学习算法的准确性对标签属性有很大的依赖性。例如徐琴珍等^[16]提出了一种有监督局部决策分层支持向量机的异常检测方法,虽然通过局部决策分层的方式把高维多分类的问题化繁为简、分而治之,但是忽略了标记样本获取难的问题。所以针对现实网络环境中获取标记样本难的问题,半监督学习和无监督学习算法大量运用在入侵检测当中。比如 Haweliya 等^[17]把半监督支持向量机应用到入侵检测中,虽然在一定程度上解决了获取标记样本难的问题,但是在现实网络环境中的记录数据通常都是有噪声的,而半监督学习算法抗干扰性比较弱,

往往会使正确率大打折扣。又如 Casas 等^[18]提到的基于子空间聚类的异常检测方法和在文献[19]中提到的无监督流量监控的方法。虽然避免了人工标记样本的工作,但是在无监督学习中,没有使用有效的方法构造多层网络,容易出现欠拟合的现象。并且准确率不高。

针对以上问题,本文提出一种基于深度多层极限学习机(Multiple Layer Extreme Learning Machine, ML-ELM)算法应用于入侵检测。首先,入侵检测的数据通常是高维度的。而多层极限学习机可以用奇异值(Singular Value)^[20]对入侵特征进行表示,相比传统的主成分分析法(Principal Component Analysis, PCA)求解时间缩短;其次,把基于深度学习思想的多层极限学习机应用于入侵检测,可以抽取检测样本最高层次的抽象特征,实现对检测数据更本质的表达,并且由于其具有较快的训练速度,符合入侵检测数据庞大以及检测速度要求快的实际情况。最后,基于多层极限学习机的检测方法可以利用逐层的无监督学习方式解决入侵检测获取标记样本难的问题。综上所述,基于深度多层极限学习机的入侵检测方法能同时有效地解决当今入侵检测中维度高、数据大、获取标记样本难、构造特征难、训练难等问题。

1 相关工作

1.1 极限学习机

极限学习机器(Extreme Learning Machine, ELM)是神经网络研究中的一种算法,是一种泛化的单隐层前馈神经网络(Single-hidden Layer Feed forward Network, SLFN)^[10]。输入权值和隐层阈值进行随机赋值,输出层权值则通过最小二乘法直接计算。整个学习过程一次完成,无需迭代,因而能达到极快的学习速度。

假设一个 q 类 d 维训练数据集 $\{x_i, y_i\}, i = 1, 2, \dots, N$, $x_i \in \mathbf{R}^d, y_i \in \mathbf{R}^q$, 则 SLFN 的输出可用 L 个隐藏节点表示为:

$$O_i = \sum_{j=1}^L \beta_j G(a_j, b_j, x_i); \quad j = 1, 2, \dots, N \quad (1)$$

其中, a_j, b_j 是随机设置的与第 j 个隐藏节点关联的输入权值和隐层阈值。 $\beta_j \in \mathbf{R}^q$ 是第 j 个隐藏节点的输出权值。 G 为任何一个无限可微分的激励函数。

这个 N 维等式可以写成式子:

$$\mathbf{O} = \mathbf{H}\boldsymbol{\beta} \quad (2)$$

$$\text{其中, } \mathbf{H} = \begin{bmatrix} G(a_1, b_1, x_1) & G(a_2, b_2, x_2) & \cdots & G(a_L, b_L, x_L) \\ G(a_1, b_1, x_2) & G(a_2, b_2, x_3) & \cdots & G(a_L, b_L, x_{L+1}) \\ \vdots & \vdots & \ddots & \vdots \\ G(a_1, b_1, x_N) & G(a_2, b_2, x_N) & \cdots & G(a_L, b_L, x_N) \end{bmatrix}_{N \times L},$$

$$\boldsymbol{\beta} = [\beta_1, \beta_2, \dots, \beta_L]^T, \mathbf{O} = [O_1, O_2, \dots, O_N]^T.$$

\mathbf{H} 称作隐层输出矩阵,其第 i 行表示第 i 个输入 x_i 关于隐层的全体输出,第 j 列则表示全体输入 x_1, x_2, \dots, x_N 关于第 j 个隐藏节点的输出。

输入权重 \mathbf{a} 和阈值 \mathbf{b} 服从某种连续概率分布随机取值,这样式(2)即为以 $\boldsymbol{\beta}$ 为变量的一个线性系统。而求解该线性系统,亦即寻找最小的输出权重 $\boldsymbol{\beta}$,使误差 $\|\mathbf{H}\boldsymbol{\beta} - \mathbf{Y}\|$ 达到最小。ELM 算法采用最小二乘法计算输出权值矩阵 $\boldsymbol{\beta}$,其解可表示为:

$$\boldsymbol{\beta} = \mathbf{H}^* \mathbf{Y} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{Y} \quad (3)$$

其中 \mathbf{H}^* 为 \mathbf{H} 的广义雅克比矩阵的逆。

为了提高泛化能力以及鲁棒性,在 $\boldsymbol{\beta}$ 的基础上增加了一个正则项^[21]:

$$\boldsymbol{\beta} = (\mathbf{I}/C + \mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{Y} \quad (4)$$

其中,C 为正则项系数, \mathbf{I} 是单位矩阵。

1.2 极限学习机自动编码器

在 ELM 基础上,让输入数据等于输出数据 $y = x$,这样便把 ELM 变成了无监督学习,并且使随机选取的隐层节点的权值和阈值都满足正交的条件,这就是极限学习机自动编码器(Extreme Learning Machine Automatic Encoder, ELM-AE)的基础。Widrow 等^[22]在以 ELM 为基础的自动编码器中引进了一种最小均方的方法,这些随机生成的正交隐层参数可以提高 ELM-AE 的泛化能力。

在 ELM-AE 中,这些随机正交的隐层参数把输入数据映射到一个被压缩的维度空间中,如 Johnson-Lindenstrauss 引理^[23]可得到计算公式:

$$\mathbf{H} = \mathbf{G}(\mathbf{a} \times \mathbf{X} + \mathbf{b}); \quad \mathbf{a}^T \mathbf{a} = 1, \mathbf{b}^T \mathbf{b} = 1 \quad (5)$$

其中: $\mathbf{a} = [a_1, a_2, \dots, a_L]$ 是在输入和隐层节点之间正交的随机权值, $\mathbf{b} = [b_1, b_2, \dots, b_L]$ 是正交的随机阈值。

ELM-AE 的输出权值 $\boldsymbol{\beta}$ 负责从特征空间到输入数据的学习转换。可以根据以下式子计算输出权值 $\boldsymbol{\beta}$:

$$\boldsymbol{\beta} = (\mathbf{I}/C + \mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{X} \quad (6)$$

其中: $\mathbf{H} = [h_1, h_2, \dots, h_N]$ 是 ELM-AE 的隐层输出, $\mathbf{X} = [x_1, x_2, \dots, x_N]$ 是它的输入数据。

1.3 ML-ELM

类似于深度网络,ML-ELM 是在 ELM-AE 的基础上进行堆叠而创造一个多层神经网络^[13]。ML-ELM 的第一个特点就是与深度学习网络相比,ML-ELM 不需要微调。ML-ELM 的隐层权值通过 ELM-AE 进行初始化,而 ELM-AE 使用的是逐层贪婪无监督学习算法。其网络结构如图 1 所示:图 1 圆圈中的 1、 p 、 d 、 L^1 、 L^i 表示神经元;图中 \mathbf{h}^k 是第 k 层隐层的输出矩阵;输入层 \mathbf{x} 可以被当成第 0 层隐层;最后一个隐层节点与输出节点 y 的输出矩阵可以用最小二乘法进行计算。

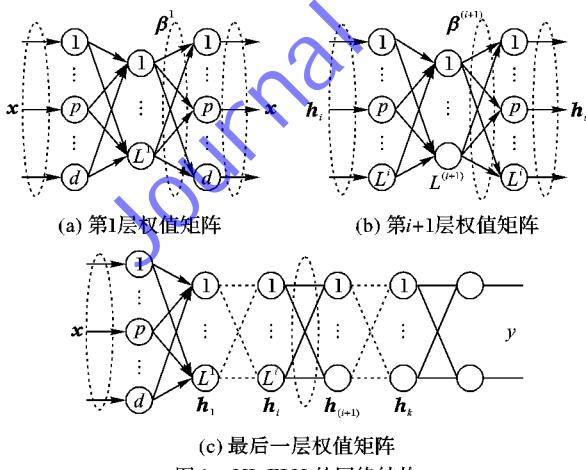


图 1 ML-ELM 的网络结构

ML-ELM 隐层激励函数可以是线性的,也可以是非线性的。如果第 k 层隐层的节点数目 L^k 等于第 $(k-1)$ 层隐层的节点数目 $L^{(k-1)}$,那么激励函数 G 是线性的;否则, G 是非线性的。比如 S 形函数。

深度多层极限学习机(ML-ELM)的第二个特点就是极限学习机自动编码器(ELM-AE)能够用奇异值^[13]表述特征。奇异值分解(Singular Value Decomposition, SVD)是一种常用的特征表示方法。式(4)的奇异值分解为:

$$\mathbf{H}\boldsymbol{\beta} = \sum_{i=1}^N \mathbf{u}_i \frac{d_i^2}{d_i^2 + C} \mathbf{u}_i^T \mathbf{X} \quad (7)$$

则可以得到:

$$\boldsymbol{\beta} = \left(\sum_{i=1}^N \mathbf{u}_i \frac{d_i^2}{d_i^2 + C} \mathbf{u}_i^T \mathbf{X} \right) \mathbf{H}^{-1} \quad (8)$$

其中: \mathbf{u} 是 $\mathbf{H}\mathbf{H}^T$ 的特征向量, d 是 \mathbf{H} 的奇异值,是与输入数据 \mathbf{X} 相关的奇异值分解得到的。经验证^[13],因为 \mathbf{H} 是被 \mathbf{X} 投射的特征空间,所以 ELM-AE 的输出权值 $\boldsymbol{\beta}$ 会通过奇异值去表达输入数据的特征。

通过 ELM-AE 的特征表达,把每一条记录作为一层 ELM-AE 的输入。经过逐层无监督学习后可以直接提供给最后一层的 ELM-AE 用于分类预测,ELM-AE 为基于多层极限学习机的入侵检测奠定了特征表达的基础。

1.4 多层极限学习机的入侵检测

多层极限学习机的入侵检测算法通过逐层无监督学习利用大量未标记样本进行训练得到各隐层的权值输出矩阵,然后利用输出矩阵对需要检测的网络数据进行分类,从而达到区分正常数据与攻击数据的目的。基于深度多层极限学习机的入侵检测方法流程如图 2 所示。

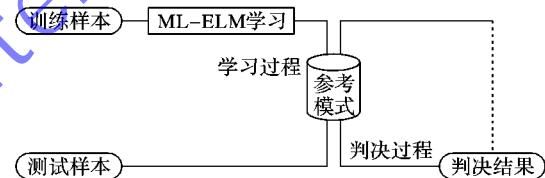


图 2 基于多层极限学习机的入侵检测方法流程

算法 1 基于多层极限学习机的入侵检测训练算法。

输入:训练样本 $\{x_i, y_i\}, i = 1, 2, \dots, N, x_i \in \mathbf{R}^d, y_i \in \mathbf{R}^d$ 。

输出:各隐层节点的输出权值矩阵 $\boldsymbol{\beta}$ 。

1) 设置好 ML-ELM 的网络结构,输入训练样本数据,使 $x = y$ 。

2) 随机设置隐层节点参数:权值 $a_j \in \mathbf{R}^d$, 阈值 $b_j \in \mathbf{R}^d$, $a^T a = 1, b^T b = 1, j = 1, 2, \dots, L$ 。

3) 计算 $\mathbf{H} = \mathbf{G}(a_j \times x_j + b_j)$ 。

4) 用式(8)计算得到 $\boldsymbol{\beta}^{(1)}$ 。

5) 当 $1 \leq i \leq K-1$ 时,循环 2) ~ 4) 计算第 i 层隐层输出矩阵 $\boldsymbol{\beta}^{(i+1)}$ 。

6) 当 $i = K$ 时,用最小二乘法计算最高层隐层输出矩阵 $\boldsymbol{\beta}^{(K+1)}$ 。

算法 2 基于多层极限学习机的入侵检测测试算法。

输入:测试样本 $\{x_i, y_i\}, i = 1, 2, \dots, N, x_i \in \mathbf{R}^d, y_i \in \mathbf{R}^d$, 各隐层节点参数 $(a, b, \boldsymbol{\beta})$ 。

输出:入侵检测分类结果:正常或攻击

1) 设置好 ML-ELM 的网络结构,输入测试样本数据。

2) 当 $1 \leq i \leq K-1$ 时,计算 $\mathbf{H} = \mathbf{G}(a_j \times x_j + b_j)$ 。

3) 计算 $\mathbf{Y} = \mathbf{H}\boldsymbol{\beta}$ 。

4) 当 $\mathbf{Y} = +1$ 时,该样本判定为正常数据;当 $\mathbf{Y} = -1$ 时,

该样本判定为攻击数据。

2 实验结果分析

2.1 实验数据集和仿真环境

实验选用的数据集为 KDD99 数据集。训练集和测试集分别为 494021 条和 311029 条记录。其中包括正常数据和攻击数据, 攻击主要包括 4 大类: PROBE(Probing Attack)、DOS (Denial of Service Attack)、U2R (User-to-Root Attack) 和 R2L (Remote-to-Login Attack)。每条记录包含 41 维特征, 其中最后 1 列为标签属性。本实验是在 2.53 GHz CPU, 4.00 GB RAM 的 Windows 7.0 系统中用 Matlab 2013a 的平台做的仿真。

2.2 实验参数的选取

基于 ML-ELM 的入侵检测以测试正确率和测试时间这两个指标作为评判实验结果的标准。

首先确定各多层次神经元结构的层数。先从测试时间来看, ML-ELM 在应用于入侵检测, 当网络结构超过 5 层时, 测试时间迅速增长, 效率偏低, 因此舍弃超过 5 层以上的网络结构。再从测试正确率来看, 4 层要比 3 层的测试正确率要高。因此 ML-ELM 的网络结构选择 4 层。同理, 经实验结果最后确定 DBN 的网络结构为 5 层, DBM 的网络结构分为 4 层。

然后确定各层神经元数目。各层神经元数目除第一层神经元数目为 41, 最后一层神经元数目为 2 保持不变以外, 中间层的神经元数目分别在 5~200 的范围内进行变化, 以 5 为步长依次增加。每次只改变一层中间层的神经元数目, 其他层神经元数目保持不变。经多次实验结果分析可得, 对于 ML-ELM, 当第二层神经元数目为 35, 第三层神经元数目为 100 时, 实验效果最佳。所以基于 ML-ELM 入侵检测的各层的神经元数目分别为 41, 35, 100, 2。对于 DBN, 当第二层和第三层数目为 30, 第四层神经元数目为 100 时, 实验效果最佳。所以基于 DBN 入侵检测的各层的神经元数目分别为 41, 30, 30, 100, 2。针对 DBM, 当第二层神经元数目为 35, 第三层神经元数目为 100 时, 实验效果最佳。所以基于 DBM 入侵检测的各层的神经元数目分别为 41, 35, 100, 2。

最后根据经验选取其他参数。ML-ELM 各层岭参数根据参考文献[23]的经验选取。其中, 第 0 层隐层 S 型激励函数的岭参数为 10-1, 第 1 层隐层 S 型激励函数的岭参数为 10, 第 2 层隐层 S 型激励函数的岭参数为 103。ELM 的隐层神经元数目为 100。支持向量机(Support Vector Machine, SVM) + 核主成分分析(Kernel Principal Component Analysis, KPCA), 即 SVM + KPCA 选用径向基函数(Radial Basis Function, RBF)作为核函数, SVM 的参数 $\sigma = 0.37$, $\varepsilon = 0.0032$, $C = 4.86$ 。SVM 的参数隐层激励函数为 Sigmoid 函数。每次实验循环 20 次求均值和标准差。

2.3 大数据, 高维数据集的适应性

从训练集和测试集中分别选取 100000 和 100000 条作为训练样本和测试样本, 比较 5 种算法在同样训练集情况下的训练时间、测试正确率和测试漏报率。实验数据如表 1。

从表 1 的数据来看, 在处理高维度、大数据集时, 就测试正确率来看, 基于深度模型的 ML-ELM, DBN 和 DBM 较之浅层学习的 ELM 和 SVM + KPCA 有较明显的优势。基于深度模型的三种算法的测试正确率达到了 95% 以上, 而浅层模型

的 ELM 因其算法的优良使其正确率接近 95%, 但 SVM + KPCA 正确率低于 92%。这是因为深度模型能用较少的参数使复杂的函数得到更本质的表达。就测试漏报率而言, ML-ELM、DBN 和 DBM 的测试漏报率已经低于 1%, 且 ML-ELM 的漏报率最低。而 ELM 和 SVM + KPCA 的漏报率相比其他三种方法较高。而就训练时间来看, ML-ELM 的优势明显突出, DBN 和 DBM 的训练时间是 ML-ELM 的几十倍。这是因为 ML-ELM 算法无需迭代, 无需微调, 大大减少了训练时间。可见, 从测试正确率、测试漏报率和训练速度综合来看, 基于深度模型的 ML-ELM 有很突出的表现。在大数据时代背景下, 入侵检测训练样本数量激增, 系统同时兼顾测试正确率、测试漏报率和训练速度这 3 个标准就显得格外重要。从实验结果分析来看, ML-ELM 可以在高维度、大数据的情况下达到快速训练并且低漏报率的要求。

表 1 测试正确率、测试漏报率和训练时间的比较

算法	测试正确率/%	测试漏报率/%	训练时间/s
ML-ELM	96.05 ± 0.247	0.48 ± 0.228	546.655 ± 0.267
ELM	94.37 ± 0.232	1.72 ± 0.304	205.687 ± 0.278
DBN	95.07 ± 0.271	0.83 ± 0.283	30589.378 ± 0.288
DBM	95.83 ± 0.219	0.91 ± 0.279	78346.762 ± 0.282
SVM + KPCA	91.14 ± 0.306	2.68 ± 0.311	4993.853 ± 0.343

2.4 少量标记样本的适应性

为了评价基于 ML-ELM 的攻击检测对于少量标记样本的适应性, 本文将 ML-ELM 和 ELM、DBN、DBM、SVM + KPCA 进行比较, 分别对于训练样本中不同标记样本数目进行了对比实验。每条训练样本为 41 维完整记录, 最后一维为标签属性, 正常类记录属性标签为 +1, 攻击类记录属性标签为 -1。为在实验中增加未标记样本, 随机选取样本把最后一列标签属性值改为 0。实验中, 训练样本 1000 条, 测试样本 1000 条。其中训练样本中由标记样本和未标记样本组成, 标记样本数量为 2^i ($i = 1, 2, \dots, 8$) 进行变化, 剩余为未标记样本, 测试样本全为未标记样本。ML-ELM、DBN、DBM、ELM 的网络结构不变。每次实验循环 20 次求平均值。

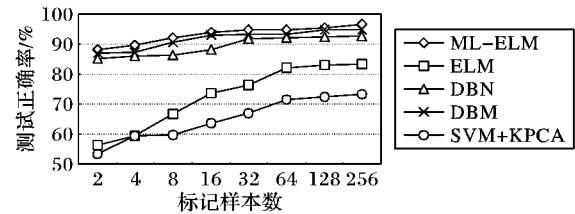


图 3 少量标记样本下测试正确率的比较

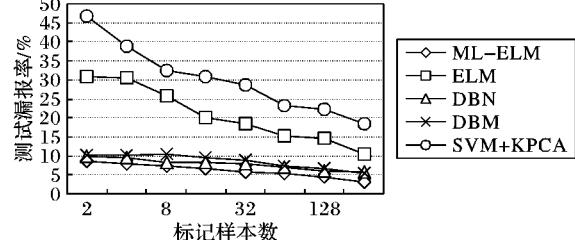


图 4 少量标记样本下测试漏报率的比较

从图 3 的结果可以看出, 在只有极少数标记样本的情况下, 基于深度模型的 ML-ELM、DBN、DBM 的正确率仍保持 80% 以上, 而其他几种基于浅层模型的算法的正确率却低于 60%。虽然随着标记样本数量的增加, 5 种算法的正确率都

有增加。但是在这种少量标记样本的前提下,ELM、SVM + KPCA 的正确率还是大大低于 ML-ELM、DBN、DBN 的正确率。从图 4 的结果可以看出,在只有极少数标记样本的情况下,几种基于浅层模型的算法的漏报率在 30% 以上;其中,SVM + KPCA 的漏报率已经接近 50%。而基于深度模型的 ML-ELM、DBN、DBN 的漏报率还是保持在 10% 左右。虽然随着标记样本数量的增加,5 种算法的漏报率都有增加。但是在这种少量标记样本的前提下,ML-ELM、DBN、DBN 的漏报率还是远远低于 ELM、SVM + KPCA 的漏报率。这得益于深度模型把预训练变成了无监督学习,无监督学习可以利用大量未标记样本进行训练学习。从图 3,4 中还可以看到,随着标记样本的增加,ML-ELM 与 DBN、DBN 相比较,ML-ELM 正确率的变化较平稳,且正确率一直略高于 DBN 和 DBN。而漏报率一直低于其他 4 种算法。这是因为 DBN、DBN 进行训练后需要 BP 算法进行微调,正确率与漏报率受网络结构以及数据变化的影响较大;而 ML-ELM 则省去了微调这一步,算法更简单,鲁棒性更好。因此对于少量标记样本的情况下,ML-ELM 算法有较高的正确率、较低的漏报率,同时也具有很好的鲁棒性。

2.5 对 U2R, R2L 的检测

KDD99 数据集的攻击样本分为 4 类:PROBE、DOS、U2R 和 R2L,其中传统的入侵检测对于 U2R 和 R2L 的检测率较低。为了评价基于 ML-ELM 的攻击检测对于攻击样本特征表达的优势,对 U2R, R2L 这两类攻击进行了检测对比。实验分别在 U2R 和 R2L 中选取 10000 和 10000 条作为训练样本和测试样本,比较 5 种算法对 U2R, R2L 这两类攻击的测试正确率和漏报率。实验数据如图 5 和图 6 所示。

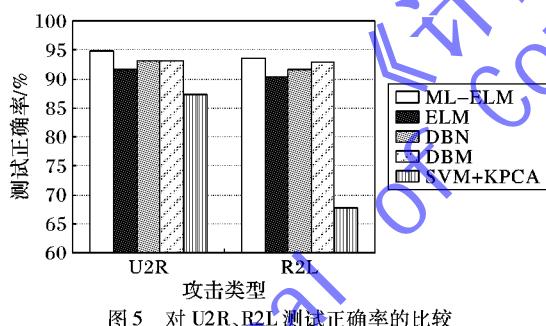


图 5 对 U2R、R2L 测试正确率的比较

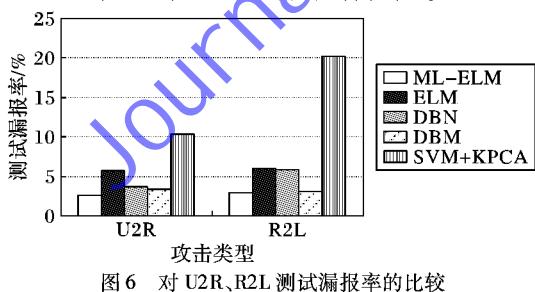


图 6 对 U2R、R2L 测试漏报率的比较

从图 5 的实验结果来看,SVM + KPCA 对于 U2R 和 R2L 的测试正确率偏低,尤其是对于 R2L 的测试正确率已低于 70%;ELM 的正确率也低于 ML-ELM、DBN 和 DBM;另外,基于深度学习的 ML-ELM、DBN、DBM 对于这两种攻击的检测正确率都在 90% 以上。从图 6 的实验结果看,两种浅层模型的方法 SVM + KPCA 和 ELM 的测试漏报率偏高,尤其是 SVM + KPCA 对 R2L 的漏报率高达 20%;而在 3 种基于深层模型的

算法,漏报率较低的一直是 ML-ELM,这得益于深度学习构架对于特征表达的优势。虽然 ML-ELM 在正确率和漏报率方面比 DBN、DBM 只有微弱的优势,但是结合 2.2 节的实验,从训练时间来看,基于 ML-ELM 的入侵检测一方面大大缩短了训练时间,另一方面对于攻击样本具有更好的特征表达,从而有利于对入侵检测更高效地的进行判决。

2.6 实验小结

以上实验以训练时间、测试正确率、测试漏报率为衡量标准,从高维度、大数据、小样本和特征表达 4 个角度来比较 ML-ELM、ELM、DBN、DBM 和 SVM + KPCA 这 5 种算法。其中 ML-ELM、DBN、DBM 为深度模型,ELM、SVM + KPCA 为浅层模型。从 2.3 节的实验可得,在处理高维度、大数据的数据集时,深度模型算法在测试正确率、测试漏报率方面有较明显的优势,且 ML-ELM 的训练时间约为其他两种深度模型算法的 1%。从 2.4 节的实验可得,在少量标记样本的情况下,ML-ELM 相比其他 4 种算法具有更高的正确率和较低的漏报率。因为 ML-ELM 为逐层无监督学习,可以利用大量未标记样本进行学习。从 2.5 节的实验可得,基于 ML-ELM 的入侵检测有很好的特征表达的能力,即使对于以往检测率偏低的 U2R、R2L 这两类攻击也具有较高的正确率和较低的漏报率。综上所述,基于深度多层极限学习机(ML-ELM)的入侵检测方法能同时有效地处理现实网络环境中大数据、高维度、获取标记样本难这几个问题。

3 结语

首先,本文把深度模型构架应用于入侵检测来解决以往入侵检测中浅层构架建模能力受限的问题。其次,基于 ML-ELM 的入侵检测方法利用逐层无监督学习以及奇异值特征表达的方式来解决高维度、大数据、获取标记样本难、特征构造难这些实际网络环境中的问题,能够更本质地表达出检测样本的特征。再者,ML-ELM 应用于入侵检测相比其他深度神经网络模型速度更快,算法更加简单,解决训练难的问题。综合以上几点,基于 ML-ELM 的入侵检测方法能同时解决现实网络中高维度、大数据、获取标记样本难、构造特征难、训练难的这几个问题,接下来的工作重点将放在入侵检测中 ML-ELM 网络构架的自动选择以及入侵检测实时性的改进等方面。

参考文献:

- [1] GU X, WANG H, NI T, et al. Detection of application-layer DDoS attack based on time series analysis [J]. Journal of Computer Applications, 2013, 33(8): 2228–2231. (顾晓清, 王洪元, 倪彤光, 等. 基于时间序列分析的应用层 DDoS 攻击检测[J]. 计算机应用, 2013, 33(8): 2228–2231.)
- [2] WANG L, TENG S. Application of clustering and time-based sequence analysis in intrusion detection [J]. Journal of Computer Applications, 2010, 30(3): 699–701. (王令剑, 滕少华. 聚类和时间序列分析在入侵检测中的应用[J]. 计算机应用, 2010, 30(3): 699–701.)
- [3] KAVITHA B, KARTHIKEYAN D S, MAYBELL P S. An ensemble design of intrusion detection system for handling uncertainty using neutrosophic logic classifier [J]. Knowledge-Based Systems, 2012, 28(4): 88–96.
- [4] HUANG H, SUN H. Network intrusion detection based on particle

- swarm optimization algorithm and information gain [J]. Journal of Computer Applications, 2014, 34(6): 1686 – 1688. (黄会群, 孙虹. 粒子群选择特征和信息增益确定特征权值的入侵检测[J]. 计算机应用, 2014, 34(6): 1686 – 1688.)
- [5] MRUTYUNJAYA P, AJITH A, MANAS R P. A hybrid intelligent approach for network intrusion detection [J]. Procedia Engineering, 2012, 30(1): 1 – 9.
- [6] HUWAIDA T E, IZZELDIN M O. Alert correlation in collaborative intelligent intrusion detection systems [J]. Applied Soft Computing, 2011, 11(7): 4349 – 4365.
- [7] BENGIO Y. Learning deep architectures for AI [J]. Foundations and Trends in Machine Learning, 2009, 2(1): 1 – 127.
- [8] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks [J]. Science, 2006, 313(5786): 504 – 507.
- [9] SALAKHUTDINOV R, HINTON G. An efficient learning procedure for deep Boltzmann machines [J]. Neural Computation, 2012, 24(8): 1967 – 2006.
- [10] HUANG G B, ZHU Q Y, SIEW C K. Extreme learning machine: theory and applications [J]. Neurocomputing, 2006, 70(1): 489 – 501.
- [11] KUANG F, XU W, ZHANG S, et al. A novel approach of KPCA and SVM for intrusion detection [J]. Journal of Computational Information Systems, 2012, 8(8): 3237 – 3244.
- [12] ARYA A, KUMAR S. Information theoretic feature extraction to reduce dimensionality of genetic network programming based intrusion detection model [C]// Proceedings of the 2014 International Conference on Issues and Challenges in Intelligent Computing Techniques. Piscataway: IEEE, 2014: 34 – 37.
- [13] HOU G, MA X, ZHANG Y. A new method for intrusion detection using manifold learning algorithm [J]. Telkomnika Indonesian Journal of Electrical Engineering, 2013, 11(12): 7339 – 7343.
- [14] CHEN J G, LI Z X. Artificial neural network based intrusion detection method combined with manifold learning algorithm [J]. Applied Mechanics and Materials, 2012, 121 – 126: 3170 – 3174.
- [15] HEYI W, AIQUN H, YUBO S, et al. A new intrusion detection feature extraction method based on complex network theory [C]// Proceedings of the 2012 4th International Conference on Multimedia Information Networking and Security. Piscataway: IEEE, 2012: 852 – 856.
- [16] XU Q, YANG L. A supervised local decision hierarchical support vector machine based anomaly intrusion detection method [J]. Journal of Electronics and Information Technology, 2010, 32(10): 2383 – 2387. (徐琴珍, 杨绿溪. 一种基于有监督局部决策分层支持向量机的异常检测方法[J]. 电子与信息学报, 2010, 32(10): 2383 – 2387.)
- [17] HAWELIYA J, NIGAM B. Network intrusion detection using semi supervised support vector machine [J]. International Journal of Computer Applications, 2014, 85(9): 27 – 31.
- [18] CASAS P, MAZEL J, OWEZARSKI P. Unsupervised network intrusion detection systems: detecting the unknown without knowledge [J]. Computer Communications, 2012, 35(7): 772 – 783.
- [19] CASAS P, MAZEL J, OWEZARSKI P. Knowledge-independent traffic monitoring: unsupervised detection of network attacks [J]. IEEE Network, 2012, 26(1): 13 – 21.
- [20] KASUN L L C, ZHOU H, HUANG G B, et al. Representational learning with ELMs for big data [J]. IEEE Intelligent Systems, 2013, 28(6): 31 – 34.
- [21] HUANG G B, ZHOU H, DING X, et al. Extreme learning machine for regression and multiclass classification [J]. IEEE Transactions on Systems, Man, and Cybernetics — Part B: Cybernetics, 2012, 42(2): 513 – 529.
- [22] WIDROW B, GREENBLATT A, KIM Y, et al. The No-Prop algorithm: a new learning algorithm for multilayer neural networks [J]. Neural Networks, 2013, 37: 182 – 188.
- [23] JOHNSON W B, LINDENSTRAUSS J, SCHECHTMAN G. Extensions of Lipschitz maps into Banach spaces [J]. Israel Journal of Mathematics, 1986, 54(2): 129 – 138.

(上接第 2485 页)

- [9] XU J, LU H, SHI G. Application of restricted velocity particle swarm optimization and self-adaptive velocity particle swarm optimization to unconstrained optimization problem [J]. Journal of Computer Applications, 2015, 35(3): 668 – 674. (许君, 鲁海燕, 石桂娟. 限制速度粒子群优化和自适应速度粒子群优化在无约束优化问题中的应用[J]. 计算机应用, 2015, 35(3): 668 – 674.)
- [10] LALWANI S, SINGHAL S, KUMAR R, et al. A comprehensive survey: Applications of Multi-Objective Particle Swarm Optimization (MOPSO) algorithm [J]. Transactions on Combinatorics, 2013, 2(1): 39 – 101.
- [11] TU K, LIANG Z. Parallel computation models of particle swarm optimization implemented by multiple threads [J]. Expert Systems with Applications, 2011, 38(5): 5858 – 5866.
- [12] MUSSI L, DAOLO F, CAGNONI S. Evaluation of parallel particle swarm optimization algorithms within the CUDA architecture [J]. Information Sciences, 2011, 181(20): 4642 – 4657.
- [13] WAINTRAUB M, SCHIRRÜ R, PEREIRA C M N A. Multiprocessor modeling of parallel particle swarm optimization applied to nuclear engineering problems [J]. Progress in Nuclear Energy, 2009, 51(6/7): 680 – 688.
- [14] ALBA E, TOMASSINI M. Parallelism and evolutionary algorithms [J]. IEEE Transactions on Evolutionary Computation, 2002, 6(5): 443 – 462.
- [15] CLERC M. Standard particle swarm optimisation from 2006 to 2011 [EB/OL]. [2015-01-08]. <https://hal.archives-ouvertes.fr/hal-00764996>.
- [16] AB AZIZ N A, MUBIN M, MOHAMAD M S, et al. A synchronous-asynchronous particle swarm optimisation algorithm [J]. The Scientific World Journal, 2014, 2014: 1 – 17.
- [17] LIANG J J, QU B Y, SUGANTHAN P N. Problem definitions and evaluation criteria for the CEC 2014 special session and competition on single objective real-parameter numerical optimization [EB/OL]. [2015-01-06]. <http://web.mysites.ntu.edu.sg/epnsugan/PublicSite/Shared%20Documents/CEC-2014/Definitions%20of%20CEC2014%20benchmark%20suite%20Part%20A.pdf>.
- [18] ZAMBRANO-BIGIARINI M, CLERC M, ROJAS R. Standard particle swarm optimisation 2011 at CEC-2013: a baseline for future PSO improvements [C]// Proceedings of the IEEE Congress on Evolutionary Computation. Piscataway: IEEE, 2013: 2337 – 2344.