计算机应用研究
Application Research of Computers
ISSN 1001-3695,CN 51-1196/TP

《计算机应用研究》网络首发论文

题目：基于深度迁移学习的网络入侵检测
作者：卢明星，杜国真，季泽旭
DOI：10.19734/j.issn.1001-3695.2019.05.0147
收稿日期：2019-05-07
网络首发日期：2019-08-29
引用格式：卢明星，杜国真，季泽旭．基于深度迁移学习的网络入侵检测[J/OL]．计算机应用研究. https://doi.org/10.19734/j.issn.1001-3695.2019.05.0147

网络首发时间：2019-08-29 10:41:37
网络首发地址：http://kns.cnki.net/kcms/detail/51.1196.TP.20190828.1140.014.html

基于深度迁移学习的网络入侵检测 *

卢明星[1]，杜国真[1]，季泽旭[2]

(1. 河南护理职业学院 网络管理中心，河南 安阳 455000; 2. 中国科学技术大学 计算机科学与技术学院，合肥 230026)

摘 要：为解决网络入侵检测问题，提高检测准确率和降低误报率等性能，提出一种基于深度迁移学习的网络入侵检测方法，该方法使用非监督学习的深度自编码器来进行迁移学习，实现网络的入侵检测。首先对深度迁移学习问题进行建模，然后对深度模型进行迁移学习。迁移学习框架由嵌入层和标签层实现编解码，编码和解码权重由源域和目标域共享，用于知识的迁移。嵌入层中，其中通过最小化域之间的嵌入实例的 KL 散度来强制源域和目标域数据的分布相似；在标签编码层中，使用 softmax 回归模型对源域的标签信息进行编码分类。实验结果表明，该方法能够实现网络入侵检测，且性能优于其他入侵检测方法。

关键词：深度自编码器；迁移学习；入侵检测；嵌入层；标签层
中图分类号：TP391 doi: 10.19734/j.issn.1001-3695.2019.05.0147

Network intrusion detection based on deep transfer learning

Lu Mingxing[1], Du Guozhen[1], Ji Zexu[2]

(1. *Network Management Center*, *Henan Vocational College of Nursing*, *Anyang Henan* 455000, *China*; 2. *School of Computer Science & Technology*, *University of Science & Technology of China*, *Hefei* 230026, *China*)

Abstract: In order to solve the problem of network intrusion detection, improve detection accuracy and reduce false positive rate, this page proposed a network intrusion detection method based on deep transfer learning. This method used unsupervised learning deep self-encoder for transfer learning to realize network intrusion detection. Firstly, the deep transfer learning problem was modeled, and then the deep transfer learning problem was modeled. The transfer learning framework implemented encoding and decoding by embedding layer and label layer, and the weight of encoding and decoding was shared by source domain and target domain for knowledge transferring. In the embedding layer, the distribution of source domain and target domain data was compelled to be similar by minimizing the KL divergence of embedded instances between domains; in the label coding layer, the label information of source domain was coded and classified by using the softwaremax regression model. The experimental results show that this method can implement network intrusion detection, and its performance is better than other intrusion detection methods.

Key words: deep self-encoder; migration learning; intrusion detection; embedded layer; label layer

0 引言

网络入侵检测(network intrusion detection, NID)是网络系统管理员检测组织网络内各种安全漏洞的重要手段[1~2]。NID 能够对网络设备及流量进行监视和分析，并在发现入侵时发出警报[3]。目前常用的网络入侵检测分为基于签名的检测方法和基于异常检测的方法[4]，在基于签名的入侵检测方法中，预先安装攻击签名，针对安装的签名执行模式匹配以检测网络中的入侵；而在基于异常检测的方法中，在观察到与正常流量模式的偏差时将网络流量分类为入侵[5]。

基于签名的检测方法可有效检测已知的攻击并显示出高检测精度，并且误报率较低。 但是，由于可以预先在 IDS 中安装的攻击签名的限制，其性能在检测到未知或新攻击时会受到影响[6]。

基于异常检测的入侵检测方法对于未知的攻击和新的攻击效果较好，但是会产生较高的假阳性率[7]。目前，网络中新型入侵检测和未知入侵检测越来越多，这使得基于异常的检测得到越来越多的关注。目前已经有许多关于网络入侵检测的方法，陈虹等人[8]提出基于优化数据处理的深度信念网络模型的入侵检测方法，其创新之处是将经过概率质量函数编码和归一化处理的数据应用于深度信念网络模型中，使得该方法具有良好的自适应性。袁琴琴等人[9]提出一种基于改进蚁群算法与遗传算法组合的网络入侵检测，提高对最优特征的选择效果，能够获得更好的入侵特征。封化民等人[10]提出 SMOTE 和 GBDT 组合的网络入侵检测方法，提高检测正确率的同时降低漏报率。

深度学习具有适应大数据处理等优势，受到越来越多的关注，已经有研究者将深度学习应用到网络入侵检测中。Yin 等人[11]提出一种基于递归神经网络的入侵检测深度学习方法，该方法非常适合于高精度的分类模型建模，其性能优于传统的机器学习分类方法。Shone 等人[12]提出一种深度自编码器的网络入侵检测方法，该方法性能优于模支持向量机和人工神经网络的性能。但是深度学习的入侵检测模型往往伴随着较高的计算复杂度。

迁移学习技术是一种机器学习方法，将知识从一个任务迁移到另外一个任务中去，其优点是不需要重新提出解决方案，只需要从相关问题迁移知识，用以解决自定义问题，通过迁移学习，可以利用已有资源，如训练数据等，减少解决方法的成本[13]。迁移学习已经得到越来越多的应用，如故障诊断[14]、预测等。任俊等人[15]提出一种基于 SDA 与 SVR 混

收稿日期：2019-05-07；**修回日期：**2019-06-25 **基金项目：**2016 年河南省教育厅高等学校青年骨干教师培养计划项目(2016GGJS-285)

作者简介：卢明星，副教授，硕士，主要研究方向为网络安全、信息化建设(luxianzh@163.com)；杜国真，讲师，学士，主要研究方向为网络安全、机器学习；季泽旭，讲师，硕士，主要研究方向为软件工程.

合模型的迁移学习预测算法，首先使用混合模型进行迁移预训练，再利用目标域小样本数据对混合模型进行微调，提高大数据时代下小样本数据预测精度。刘振等人[16]提出基于卷积神经网络与迁移学习的油茶病害图像识别方法，利用深度卷积神经网络自动学习油茶病害特征，然后使用迁移学习将在 ImageNet 图像数据集上学习得到的知识迁移到油茶病害识别任务，提高了识别效率。

本文在迁移学习与深度学习的基础上，提出一种基于深度迁移学习的网络入侵检测方法，在本文方法中，迁移学习对深度学习模型做了改善，使得深度学习在网络入侵上的检测更加准确。将本文方法与现有其他方法进行比较，本文方法能够实现网络入侵检测，且性能优于其他入侵检测方法。

1　迁移学习与深度自编码器

1.1　迁移学习

迁移学习是将待训练分类模型通过添加层，或者在不同的数据集上重新训练得到基本特征，并将其迁移到新的任务中去。迁移学习具有以下优点：a)训练数据量需求量更小；b)降低深度学习的入门门槛；c)可以改进模型的泛化能力；d)使训练过程更加稳定。

将迁移学习数学化可表示为：域 $D=\{X,P(X)\}$ 由特征空间 X 和边际概率分布 $P(X)$ 组成，给定一个域，任务 $T=\{Y,f(\cdot)\}$ 由一个标记空间 Y 和一个预测函数 $f(\cdot)$ 组成，该预测函数为 $x\in X$ 和 $y\in Y$ 建模。给定源域 D_S 和学习任务 T_S，以及目标域 D_T 和学习任务 T_T，迁移学习使用来自 D_S 和 T_S 的知识改进 T_T 中目标预测函数 $f_T(\cdot)$ 的学习，其中 $D_S\neq D_T$，$T_S\neq T_T$。

当给定标记的源域数据 D_S 和目标域数据 D_T 时，迁移学习特别相关，此时有 $|D_T|\ll|D_S|$。在本文的设置中，对两个以上的学习任务感兴趣，通过将多源目标关系表示为迁移学习图，首先定义迁移学习问题 $P=\{D,T\}$ 作为域任务对，将迁移学习图定义为 $G=\{V,\varepsilon\}$，是有向无环图，其中顶点 $V=\{P_1,P_2,...,P_e\}$ 是迁移学习问题，$\varepsilon=\{(P_{i_1},P_{j_1}),(P_{i_2},P_{j_2}),...,(P_{i_e},P_{j_e})\}$ 是边集。对于每个迁移学习问题 $P_i=\{D_i,T_i\}$，目的是利用 $\bigcup_{(j,i)\in\varepsilon}P_j$ 中的知识改进 T_i 中的目标预测函数 $f_i(\cdot)$ 的学习。

1.2　深度自编码器

深度自编码器是一个前馈神经网络，具有输入层、输出层和一个或多个隐藏层，一个深度自编码器框架通常包括编码和解码过程。给定一个输入 x，深度自编码器编码通过几个编码过程到一个或多个隐藏层，然后解码隐藏层以获得输出 $\hat{x}$。深度自编码器尝试最小化 $\hat{x}$ 与输入 x 的偏差，具有一个隐藏层的深度自编码器的过程可以概括为式(1)中的编码过程与式(2)中的解码过程为

$$\xi=f(\boldsymbol{W}_1x+\boldsymbol{b}_1) \tag{1}$$

$$\hat{x}=f(\boldsymbol{W}_1'x+\boldsymbol{b}_1') \tag{2}$$

其中：f 表示非线性激活函数(在本文采用 sigmoid 函数)；$\boldsymbol{W}_1\in\mathbb{R}^{k\times m}$ 和 $\boldsymbol{W}_1'\in\mathbb{R}^{m\times k}$ 表示权重矩阵；$\boldsymbol{b}_1\in\mathbb{R}^{k\times1}$ 和 $\boldsymbol{b}_1'\in\mathbb{R}^{m\times1}$ 表示偏置向量，$\xi\in\mathbb{R}^{k\times1}$ 表示隐藏层的输出。给定一组输入 $\{x_i\}_{i=1}^n$，重建误差可以通过 $\sum_{i=1}^n\|\hat{x}_i-x_i\|^2$ 来得到。深度自编码器的目标是学习权重矩阵 $\boldsymbol{W}_1$ 和 $\boldsymbol{W}_1'$，和偏差向量 $\boldsymbol{b}_1$ 和 $\boldsymbol{b}_1'$ 通过最小化重建误差如下所示。

$$\min_{\boldsymbol{W}_1,\boldsymbol{b}_1,\boldsymbol{W}_1',\boldsymbol{b}_1'}\sum_{i=1}^n\|\hat{x}_i-x_i\|^2 \tag{3}$$

2　深度自编码器的迁移学习

2.1　问题模型化

给定两个域 D_s，D_t，$D_s=\{x_i^{(s)},y_i^{(s)}\}|_{i=1}^{n_s}$ 是具有标记数据 $x_i^{(s)}\in\mathbb{R}^{m\times1}$ 和 $y_i^{(s)}\in\{1,...,c\}$ 的源域，而 $D_t=\{x_i^{(t)}\}|_{i=1}^{n_t}$ 是未标记数据的目标域，其中 n_s 和 n_t 分别是 D_s 和 D_t 中实例数量。

如图 1 所示，本文有三个因素要考虑表示迁移学习，因此，本文提出的迁移学习框架中要最小化的目标可以形式化为

$$J=J_r(x,\hat{x})+\Gamma(\xi^{(s)},\xi^{(t)})+L(\theta,\xi^{(s)})+\Omega(\boldsymbol{W},\boldsymbol{b},\boldsymbol{W}',\boldsymbol{b}') \tag{4}$$

其中：$J_r(x,\hat{x})$ 表示源域和目标域数据的重构误差，定义为

$$J_r(x,\hat{x})=\sum_{r\in\{s,t\}}\sum_{i=1}^{n_r}\|x_i^{(r)}-\hat{x}_i^{(r)}\|^2 \tag{5}$$

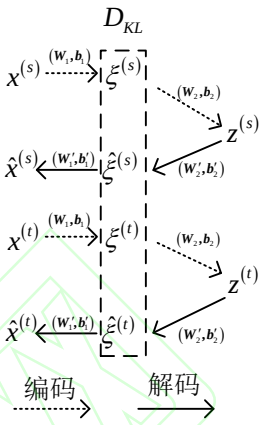

图 1　本文深度自编码器迁移学习结构

Fig. 1　Deep self-encoder migration learning structure in this paper

图 1 中，$x_i^{(s)}$，$x_i^{(t)}$ 表示源域和目标域的第 i 个实例，$\hat{x}_i^{(s)}$，$\hat{x}_i^{(t)}$ 表示 $x_i^{(s)}$ 和 $x_i^{(t)}$ 的重构，$\xi_i^{(s)}$，$\xi_i^{(t)}$ 为 $x_i^{(s)}$ 和 $x_i^{(t)}$ 的隐藏表示，$\hat{\xi}_i^{(s)}$，$\hat{\xi}_i^{(t)}$ 表示 $\xi_i^{(s)}$ 和 $\xi_i^{(t)}$ 的重建，$z_i^{(s)}$ 和 $z_i^{(t)}$ 为 $\xi_i^{(s)}$ 和 $\xi_i^{(t)}$ 的隐藏表示，则有：$\xi_i^{(r)}=f(W_1x_i^{(r)}+b_1)$，$z_i^{(r)}=f(W_2\xi_i^{(r)}+b_2)$，$\hat{\xi}_i^{(r)}=f(W_2'z_i^{(r)}+b_2')$，$\hat{x}_i^{(r)}=f(W_1'\hat{\xi}_i^{(r)}+b_1')$。

第一个隐藏层称为嵌入层，其具有 k 节点($k\leq m$)的输出 $\xi\in\mathbb{R}^{k\times1}$，权重矩阵 $\boldsymbol{W}_1\in\mathbb{R}^{k\times m}$，偏置向量 $\boldsymbol{b}_1\in\mathbb{R}^{k\times1}$。第一层的输出是第二个隐藏层的输入，第二个隐藏层称为标签层，其具有 c 个节点的输出 $z\in\mathbb{R}^{c\times1}$(等于类标签的数量)，权重矩阵 $\boldsymbol{W}_2\in\mathbb{R}^{c\times k}$ 和偏置向量 $\boldsymbol{b}_2\in\mathbb{R}^{c\times1}$。本文将 softmax 回归用作源域上的正则化项以合并标签信息。此外，第二层的输出作为目标域的预测结果。第三个隐藏层 $\hat{\xi}$ 是嵌入层的重构，具有相应的权重矩阵 $\boldsymbol{W}_2'\in\mathbb{R}^{k\times c}$ 和偏置向量 $\boldsymbol{b}_2'\in\mathbb{R}^{k\times1}$。最后，$\hat{x}$ 是 x 的重构，其中 $\hat{x}\in\mathbb{R}^{m\times1}$，$\boldsymbol{W}_1'\in\mathbb{R}^{m\times k}$，$\boldsymbol{b}_1'\in\mathbb{R}^{m\times1}$。

式(4)中第二项 $\Gamma(\xi^{(s)},\xi^{(t)})$ 是源域和目标域之间嵌入式实例的 KL 散度，则 $\Gamma(\xi^{(s)},\xi^{(t)})$ 可以表示为

$$\Gamma(\xi^{(s)},\xi^{(t)})=D_{KL}(P_s\|P_t)+D_{KL}(P_t\|P_s) \tag{6}$$

其中：

$$P_s'=\frac{1}{n_s}\sum_{i=1}^{n_s}\xi_i^{(s)},\ P_s=\frac{P_s'}{\sum P_s'} \tag{7}$$

$$P_t'=\frac{1}{n_t}\sum_{i=1}^{n_t}\xi_i^{(t)},\ P_t=\frac{P_t'}{\sum P_t'} \tag{8}$$

D_{KL} 为 KL 散度，也称为相对熵，是一个非对称的两个概率分布之间的差异，给定两个概率分布 $P\in\mathbb{R}^{k\times1}$ 和 $Q\in\mathbb{R}^{k\times1}$，$Q$ 从 P 的 KL 散度是 Q 用于近似 P 时丢失的信息，定义为

$$D_{KL}(P\|Q)=\sum_{i=1}^k P(i)\ln\left(\frac{P(i)}{Q(i)}\right) \tag{9}$$

本文使用对称 KL 散度，$KL(P,Q)=D_{KL}(P\|Q)+D_{KL}(Q\|P)$，来度量分类问题的散度，KL 散度值越小，两种分布就越相似。因此，使用 KL 散度来测量当两个数据域嵌入到相同的潜在空间时它们之间的差异。

最小化源域和目标域之间的嵌入式实例的 KL 散度的目的是确保源和目标数据分布在嵌入空间中是相似的。

式(4)中第三项是 softmax 回归的损失函数，用于将源域的标签信息合并到嵌入空间中，即

$$L(\theta,\xi^{(s)})=-\frac{1}{n_s}\sum_{i=1}^{n_s}\sum_{j=1}^{c}1\{y_i^{(s)}=j\}\log\frac{e^{\theta_j^T\xi_i^{(s)}}}{\sum_{l=1}^{c}e^{\theta_l^T\xi_i^{(s)}}} \tag{10}$$

其中：$\theta_j^T(j\in\{1,...,c\})$ 是 $\boldsymbol{W}_2$ 的第 j 行。最后一项是模型参数的正则化，定义如下：

$$\begin{aligned}\Omega(\boldsymbol{W},\boldsymbol{b},\boldsymbol{W}',\boldsymbol{b}')=&\|\boldsymbol{W}_1\|^2+\|\boldsymbol{b}_1\|^2+\|\boldsymbol{W}_2\|^2+\\&\|\boldsymbol{b}_2\|^2+\|\boldsymbol{W}_1'\|^2+\|\boldsymbol{b}_1'\|^2+\|\boldsymbol{W}_2'\|^2+\|\boldsymbol{b}_2'\|^2\end{aligned} \tag{11}$$

2.2　深度自编码器的迁移学习

式(4)中关于 $\boldsymbol{W}_1$，$\boldsymbol{b}_1$，$\boldsymbol{W}_2$，$\boldsymbol{b}_2$，$\boldsymbol{W}_1'$，$\boldsymbol{b}_1'$，$\boldsymbol{W}_2'$，$\boldsymbol{b}_2'$ 的最小化问题是无约束的优化问题，为了解决这个问题，采用梯度下降的方法来解决问题。目标式(4)关于 $\boldsymbol{W}_1$，$\boldsymbol{b}_1$，$\boldsymbol{W}_2$，$\boldsymbol{b}_2$，$\boldsymbol{W}_1'$，$\boldsymbol{b}_1'$，$\boldsymbol{W}_2'$，$\boldsymbol{b}_2'$ 的偏导数可以分别计算如下：

$$\begin{aligned}\frac{\partial J}{\partial W_1}=&\sum_{i=1}^{n_s}2\boldsymbol{W}_1^TA_i^{(s)}\circ(\boldsymbol{W}_2^T(\boldsymbol{W}_2'^TB_i^{(s)}\circ C_i^{(s)}))\circ D_i^{(s)}x_i^{(s)T}+\\&\sum_{i=1}^{n_t}2\boldsymbol{W}_1'^TA_i^{(t)}\circ(\boldsymbol{W}_2'^T(\boldsymbol{W}_2^TB_i^{(t)}\circ C_i^{(t)}))\circ D_i^{(t)}x_i^{(t)T}+\\&\frac{\alpha}{n_s}\sum_{i=1}^{n_s}D_i^{(s)}\circ(1-\frac{P_t}{P_s}+\ln(\frac{P_s}{P_t}))x_i^{(s)T}+\\&\frac{\alpha}{n_t}\sum_{i=1}^{n_t}D_i^{(t)}\circ(1-\frac{P_s}{P_t}+\ln(\frac{P_t}{P_s}))x_i^{(t)T}+2\gamma\boldsymbol{W}_1-\\&\frac{\beta}{n_s}\sum_{i=1}^{n_s}\sum_{j=1}^{c}1\{y_i^{(s)}=j\}(\boldsymbol{W}_2^T-\frac{\boldsymbol{W}_2^Te^{\boldsymbol{W}_2\xi_i^{(s)}}}{\sum_l e^{\boldsymbol{W}_2\xi_i^{(s)}}})\circ D_i^{(s)}x_i^{(s)T}\end{aligned} \tag{12}$$

$$\begin{aligned}\frac{\partial J}{\partial W_{2j}}=&\sum_{i=1}^{n_s}2W_{2j}^T(W_1'^TA_i^{(s)}\circ B_i^{(s)})\circ C_{ij}^{(s)}\xi_i^{(s)T}+\\&\sum_{i=1}^{n_t}2W_{2j}^T(W_1'^TA_i^{(t)}\circ B_i^{(t)})\circ C_{ij}^{(t)}\xi_i^{(t)T}-\\&\frac{\beta}{n_{sj}}(\sum_{i=1}^{n_{sj}}\xi_i^{(s)T}-\sum_{i=1}^{n_s}\frac{e^{W_{2j}\xi_i^{(s)}}}{\sum_l e^{W_{2l}\xi_i^{(s)}}}\xi_i^{(s)T})+2\gamma W_{2j}\end{aligned} \tag{13}$$

$$\frac{\partial J}{\partial W_2'}=\sum_{i=1}^{n_s}2\boldsymbol{W}_1'^TA_i^{(s)}\circ B_i^{(s)}z_i^{(s)T}+2\gamma\boldsymbol{W}_2'+\sum_{i=1}^{n_t}2\boldsymbol{W}_1'^TA_i^{(t)}\circ B_i^{(t)}z_i^{(t)T} \tag{14}$$

$$\frac{\partial J}{\partial W_1'}=\sum_{i=1}^{n_s}2A_i^{(s)}\hat{\xi}_i^{(s)T}+\sum_{i=1}^{n_t}2A_i^{(t)}\hat{\xi}_i^{(t)T}+2\gamma\boldsymbol{W}_1' \tag{15}$$

式 (12)~(15) 中，$A_i^{(r)}=(\hat{x}_i^{(r)}-x_i^{(r)})\circ\hat{x}_i^{(r)}\circ(1-\hat{x}_i^{(r)})$，$B_i^{(r)}=\hat{\xi}_i^{(r)}\circ(1-\hat{\xi}_i^{(r)})$，$C_i^{(r)}=z_i^{(r)}\circ(1-z_i^{(r)})$，$D_i^{(r)}=\xi_i^{(r)}\circ(1-\xi_i^{(r)})$。$\boldsymbol{W}_{2j}$ 是 $\boldsymbol{W}_2$ 的第 j 行，而 n_{sj} 是数字在源域中使用标签 j 的实例。目标式(4)中 $\boldsymbol{b}_1,\boldsymbol{b}_2,\boldsymbol{b}_2',\boldsymbol{b}_1'$ 的偏导数与 $\boldsymbol{W}_1,\boldsymbol{W}_2,\boldsymbol{W}_2',\boldsymbol{W}_1'$ 非常相似，在这里不再给出表达式。基于上述偏导数，给出一种迭代算法，使用以下规则推导出解：

$$\begin{aligned}&\boldsymbol{W}_1\leftarrow\boldsymbol{W}_1-\eta\frac{\partial J}{\partial\boldsymbol{W}_{11}},\quad\boldsymbol{b}_1\leftarrow\boldsymbol{b}_1-\eta\frac{\partial J}{\partial\boldsymbol{b}_1}\\&\boldsymbol{W}_1'\leftarrow\boldsymbol{W}_1'-\eta\frac{\partial J}{\partial\boldsymbol{W}_1'},\quad\boldsymbol{b}_1'\leftarrow\boldsymbol{b}_1'-\eta\frac{\partial J}{\partial\boldsymbol{b}_1'}\\&\boldsymbol{W}_2\leftarrow\boldsymbol{W}_2-\eta\frac{\partial J}{\partial\boldsymbol{W}_2},\quad\boldsymbol{b}_2\leftarrow\boldsymbol{b}_2-\eta\frac{\partial J}{\partial\boldsymbol{b}_2}\\&\boldsymbol{W}_2'\leftarrow\boldsymbol{W}_2'-\eta\frac{\partial J}{\partial\boldsymbol{W}_2'},\quad\boldsymbol{b}_2'\leftarrow\boldsymbol{b}_2'-\eta\frac{\partial J}{\partial\boldsymbol{b}_2'}\end{aligned} \tag{16}$$

其中：η 是步长，它决定了收敛速度。本文算法的步骤在算法 1 中给出。所提出的优化问题不是凸优的，因此不能保证获得最优的全局解。为了更好地获得梯度下降算法的局部最优解，首先在所有源和目标域数据上运行堆叠自编码器，然后使用堆叠自编码器的输出来初始化编码和解码权重。

算法 1　使用深度自编码器的迁移学习

输入：给定源域 $D_s=\{x_i^{(s)},y_i^{(s)}\}_{i=1}^{n_s}$ 和目标域 $D_t=\{x_i^{(t)}\}_{i=1}^{n_t}$，嵌入层和标签层中的节点数 k 和 c。

输出：标签层 z 和嵌入层 ξ 的结果。

a)通过在源域和目标域上执行堆叠式自编码器初始化 $\boldsymbol{W}_1$，$\boldsymbol{b}_1$，$\boldsymbol{W}_2$，$\boldsymbol{b}_2$，$\boldsymbol{W}_1'$，$\boldsymbol{b}_1'$，$\boldsymbol{W}_2'$，$\boldsymbol{b}_2'$；

b)根据式(12)-(15)计算所有变量的偏导数；

c)使用式(16)迭代更新变量；

d)继续步骤 b)和步骤 c)直至算法收敛；

e)计算嵌入层 ξ 和标签层 z，然后构造目标分类器。

本文算法对于深度自编码器的改进之处是加入迁移学习，将堆栈深度自编码器的内部隐藏层当做一个通用的中间级特征提取器，并分为标签层和嵌入层，然后使用梯度下降算法计算权重和偏置向量的偏导，得到嵌入层和标签层，引入 KL 散度，确保源域与目标域数据分布在嵌入空间中是相似的，在源域数据集上预训练得到的特征可迁移至其他目标域数据集上，实现高精度网络入侵检测。整个深度自编码器迁移学习流程图如图 2 所示。

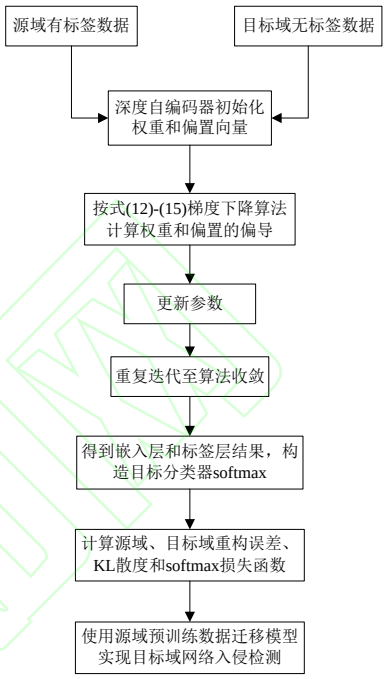

图 2　深度自编码器的迁移学习网络入侵检测流程图

Fig. 2　Transfer Learning of depth self-encode network intrusion detection flow chart

3　实验结果与分析

本文选择 KDDCUP 99 这个网络入侵检测算法研究领域的基准数据集来对本文算法进行有效性验证，该数据集中的训练数据为 490 多万，测试集数据为 310 多万，1 个类标签和 41 个属性组成一条数据，数据集中有 4 类入侵类型，包含了 36 种攻击。表 1 给出了样本数据集分布，本次实验采用该数据集的 10%左右的数据进行实验。

表 1　实验数据集

Tab. 1　Experimental data set

数据类型	训练数据	测试数据
DoS	391458	229853
probe	4107	4166
R2L	1126	16189
U2R	523	2287
normal	97278	60593

本文实验的实验环境为笔记本电脑上的 MATLAB 2013a，电脑配置为：8G 内存，1T 硬盘，Intel i7 处理器，Windows 10 系统。本文的性能指标有正确率 CR、误报率 FR、漏报率 CR。实验中分类模型结果如表 2 所示。

表 2　分类模型的结果

Tab. 2　Results of the classification model

	分类为正常样本	分类为异常样本
真实正常样本	TP	FN
真实异常样本	FP	TN

则本文性能指标可以表示为

$$CR=\frac{TP+TN}{TP+TN+FP+FN} \tag{17}$$

$$FR=\frac{FN}{TP+TN+FP+FN} \tag{18}$$

$$CR=\frac{FP}{TP+TN+FP+FN} \tag{19}$$

对于现实中网络入侵检测结果，将网络异常分类为正常比将网络正常分类为异常更不可接受，当算法具有较高的正确率和较低的漏报率时，该算法对于网络入侵的检测更有优势。

为了验证本文方法的有效性与优越性，将本文方法与文献[11]中的基于递归神经网络的入侵检测深度学习方法，文献[12] 中深度自编码器的网络入侵检测方法以及文献[17]中基于卷积神经网络的入侵检测算法进行比较，对于攻击的性能指标结果见表 3 所示。

表 3　不同方法对于网络入侵检测性能指标

Tab. 3　Different methods for network intrusion detection performance indicators

模型	评价指标(%)	DoS	Probe	R2L	U2R
文献[11]	准确率	89.12	82.3	93.54	95.37
	误报率	8.13	13.74	2.9	2.26
	漏报率	2.52	1.72	1.31	1.94
文献[12]	准确率	86.52	79.69	90.92	92.74
	误报率	9.93	15.55	4.72	3.62
	漏报率	3.52	2.73	2.33	2.97
文献[17]	准确率	82.66	77.59	89.71	91.83
	误报率	11.27	19.14	6.75	7.5
	漏报率	6.04	3.24	3.51	4.64
本文	准确率	94.71	90.46	97.68	98.97
	误报率	4.8	10.85	1.56	1.31
	漏报率	0.46	0.7	0.63	0.29

从表中可以看出，本文基于深度自编码的迁移学习检测方法在对网络入侵的攻击进行检测时，在准确率、误报率和漏报率性能方面都优于其他三种方法，说明本文方法的有效性和优越性。图 3 给出了在 normal 情况下不同方法的性能指标。

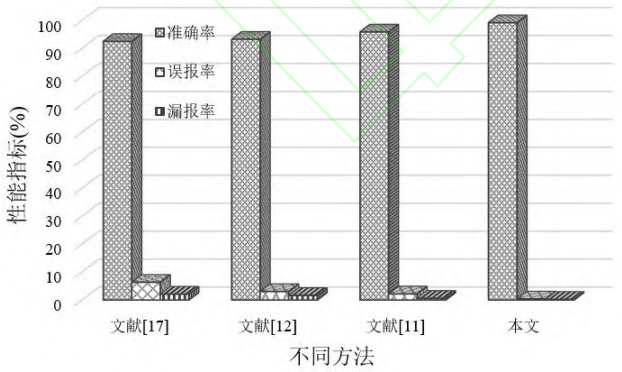

图 3　normal 情况下不同方法的性能指标

Fig. 3　Performance metrics for different methods under normal conditions

从上图可以看出，在 normal 情况下，本文方法在准确率、误报率和漏报率性能方面也都优于其他三种方法，说明本文方法的有效性与优越性。文献[11]、[12]和[17]中深度学习方法能够实现网络入侵检测，这是因为具有学习中间特征的能力，但是现实应用中数据的不充足性导致模型出现过拟合问题，从而使得检测性能下降，本文所提深度自编码器的迁移学习网络入侵检测方法，可以将学习到的特征有效迁移到目标待检测数据中，尽管源域数据和目标域数据存在差异，但是迁移学习仍能够训练处目标数据的特征，在网络入侵检测中达到更优越的检测效果。

4　结束语

提出一种深度自编码器的迁移学习方法来实现网络入侵检测。本文将深度学习与迁移学习结合起来，弥补了深度学习的一些缺点。本文方法中用到两个层来进行编解码，第一个层是嵌入层，通过 KL 散度约束来得到源域与目标域的相似分布，然后通过另一个层，即标签层，来合并来自源域的标签信息。最后在 KDDCUP 99 数据集上对本文方法性能进行验证，实验结果表明，本文在准确率、误报率和漏报率性能方面都优于其他三种网络入侵检测方法，说明本文方法的可行性与有效性。下一步工作是研究将本文方法应用到实际的网络入侵检测中去。

参考文献：

[1] Kevric J, Jukic S, Subasi A. An effective combining classifier approach using tree algorithms for network intrusion detection [J]. Neural Computing and Applications, 2017, 28 (1): 1051-1058.

[2] 戴远飞, 陈星, 陈宏, 等. 基于特征选择的网络入侵检测方法 [J]. 计算机应用研究, 2017, 34(8): 2429-2433. (Dai Yuanfei, Chen Xing, Chen Hong, et al. Feature selection based approach to network intrusion detection [J]. Application Research of Computers, 2017, 34(8): 2429-2433.)

[3] Aziz A S A, Sanaa E L, Hassanien A E. Comparison of classification techniques applied for network intrusion detection and classification [J]. Journal of Applied Logic, 2017, 24: 109-118.

[4] Papamartzivanos D, Mármol F G, Kambourakis G. Dendron: Genetic trees driven rule induction for network intrusion detection systems [J]. Future Generation Computer Systems, 2018, 79: 558-574.

[5] 王笑, 戚湧, 李千目. 基于时变加权马尔可夫链的网络异常检测模型 [J]. 计算机科学, 2017, 44 (9): 136-141. (Wang Xiao, Qi Yong, Li Qianmu. Network Anomaly Detection Model Based on Time-varying Weighted Markov Chain [J]. Computer Science, 2017, 44 (9): 136-141.)

[6] Meng W, Li W, Kwok L F. Towards effective and robust list-based packet filter for signature-based network intrusion detection: an engineering approach [J]. HKIE Transactions, 2017, 24 (4): 204-215.

[7] Aljawarneh S, Aldwairi M, Yassein M B. Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model [J]. Journal of Computational Science, 2018, 25: 152-160.

[8] 陈虹, 万广雪, 肖振久. 基于优化数据处理的深度信念网络模型的入侵检测方法 [J]. 计算机应用, 2017, 37 (6): 1636-1643. (Chen Hong, Wan Guangxue, Xiao Zhenjiu. Intrusion detection method of deep belief network model based on optimization of data processing [J]. Journal of Computer Applications, 2017, 37 (6): 1636-1643.)

[9] 袁琴琴, 吕林涛. 基于改进蚁群算法与遗传算法组合的网络入侵检测 [J]. 重庆邮电大学学报 (自然科学版) , 2017, 29 (1): 84-89. (Yuan Qinqin, Lv Lintao. Network intrusion detection method based on combination of improved ant colony optimization and genetic algorithm [J]. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition) , 2017, 29 (1): 84-89.)

[10] 封化民, 李明伟, 侯晓莲, 等. 基于 smote 和 gbdt 的网络入侵检测方法研究 [J]. 计算机应用研究, 2017, 34(12): 3745-3748. (Feng Huamin, Li Mingwei, Hou Xiaolian, et al. Study of network intrusion detection method based on smote and gbdt [J]. Application Research of Computers, 2017, 34(12): 3745-3748.)

[11] Yin C, Zhu Y, Fei J, *et al.* A deep learning approach for intrusion

detection using recurrent neural networks [J]. Ieee Access, 2017, 5: 21954-21961.

[12] 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.

[13] Zamir A R, Sax A, Shen W, *et al*. Taskonomy: Disentangling task transfer learning [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 3712-3722.

[14] 陈超，沈飞，严如强. 改进 LSSVM 迁移学习方法的轴承故障诊断 [J]. 仪器仪表学报, 2017, 38 (1): 33-40. (Chen Chao, Shenfei, Yan Ruqiang. Enhanced least squares support vector machine-based transfer learning strategy for bearing fault diagnosis [J]. Chinese Journal of Scientific Instrument, 2017, 38 (1): 33-40.)

[15] 任俊，胡晓峰，李宁. 基于 SDA 与 SVR 混合模型的迁移学习预测算法 [J]. 计算机科学, 2018, 45 (1): 280-284. (Ren Jun, Hu Xiaofeng, Lining. Transfer Prediction Learning Based on Hybrid of SDA and SVR [J]. Computer Science, 2018, 45 (1): 280-284.)

[16] 郑群花，段慧芳，沈尧，等. 基于卷积神经网络和迁移学习的乳腺癌病理图像分类 [J]. 计算机应用与软件, 2018, 35 (7): 237-242. (Zheng Qunhua, Duan Huifang, Shen Xiao, *et al.* Breast Cancer Histological Image Classification Based on Convolutional Neural Network And Transfer Learning [J]. Computer Applications and Software, 2018, 35 (7): 237-242.)

[17] 贾凡，孔令智. 基于卷积神经网络的入侵检测算法 [J]. 北京理工大学学报, 2017 (12): 1271-1275. (Jia Fan, Konglingzhi. Intrusion Detection Algorithm Based on Convolutional Neural Network [J]. Transactions of Beijing Institute of Technology, 2017 (12): 1271-1275.)