基于变分自编码器和三支决策的工控入侵检测算法*

王 晨,张迪明,韩 斌
(江苏科技大学 计算机学院,江苏 镇江 212100)

摘 要:为了更精确地提取工控入侵数据集特征和更精准地分类恶意数据,使得入侵检测方法满足当前工业控制网络的安全需求,提出了基于变分自编码器(Variational Autoencoder,VAE)和三支决策理论(Three-way Decisions,TWD)的新型工业控制网络入侵检测算法(VAE-TWD)。该算法利用变分自编码器强大的感知能力对高维数据进行降维映射和特征提取,再对正常和恶意数据利用三支决策理论进行即刻决策,划分入正向决策域和负向决策域。而对于边界域内不确定的数据,将通过不同粒度的特征,选择适当数据构成新的训练集并扩充到原有数据集中。然后重复决策过程,直至决策域中数据为空,规避盲目决策的风险。实验结果表明VAE-TWD算法提升了对工控入侵检测的特征提取能力和分类能力,且在准确率、检出率、误报率、*F*1得分等指标上均优于对比算法,有效提高了工控入侵检测的性能。

关键词:变分自编码器;三支决策;特征提取;工控入侵检测
中图分类号:TP393 **文献标识码**:A DOI:10.19358/j.issn.2096-5133.2022.06.002
引用格式:王晨,张迪明,韩斌. 基于变分自编码器和三支决策的工控入侵检测算法[J].信息技术与网络安全,2022,41(6):10-17.

An industrial intrusion detection algorithm based on variational autoencoder and three-way decisions

Wang Chen, Zhang Diming, Han Bin
(School of Computer, Jiangsu University of Science and Technology, Zhenjiang 212100, China)

Abstract: In order to extract the characteristics of industrial control intrusion data set and classify malicious data more accurately, make the intrusion detection methods meet the security needs of the current industrial control network, a novel VAE-TWD algorithm based on variational auto-encoders(VAE) and three-way decisions theory(TWD) is proposed. The algorithm uses the powerful perceptive ability of variational autoencoder to reduce dimension mapping and extract feature for high-dimensional data; and then makes instant decision for normal and malicious data by using three-way decision theory, divides them into positive decision domain and negative decision domain. For the uncertain data in the boundary region, the new training set will be constructed by selecting appropriate data with different granularity features and then extended to the original data set. Then the decision-making process is repeated until the data in the decision-making domain is empty to avoid the risk of blind decision-making. The experimental results show that VAE-TWD algorithm improves feature extraction ability and classification ability of industrial control intrusion detection, is superior to the comparison algorithms in accuracy, detection rate, false positive rate, *F*1 score and other indicators, and effectively improves the performance of industrial control intrusion detection.

Key words: variational autoencoder; three-way decisions; feature extraction; industrial control intrusion detection

0 引言

工业控制网络其核心是将互联网技术同自动化控制技术相结合。随着工业化的推进,虽然越来越多的网络模块和控制器优化了工控系统并提升了生产效率,但是高度复杂的工控系统同样增加了其暴露高危漏洞的风险[1]。如今,工控安全是网络安全领域亟待解决的热点问题。

在工控安全的研究领域中,学者们针对不同的

*基金项目:国家自然科学基金(61702234)

工业生产环境，设计出了不同的入侵检测算法模型。赵智阳等人[2]提出了一种基于卷积神经网络的电网工控系统入侵检测算法，在神经网络结构中加入级联卷积层提升了特征提取能力。庄卫金等人[3]提出了基于特征提取的电力工控系统入侵检测方法，通过堆叠稀疏编码器并在训练过程中引入迁移学习进行参数优化，提升了对数据关键特征提取的能力。Shang 等人[4]通过一类支持向量机(One-Class Support Vector Machine，OCSVM)概念建立正常的通信行为模型，并设计粒子群优化算法对 OCSVM 模型参数进行优化，设计了工控系统中基于 OCSVM 的入侵检测算法。Liu 等人[5]使用两级检测结构，结合 CNN 特征提取来构建入侵检测的正常状态过程转移模型，提出了一种基于 CNN 和过程状态转换的工业控制系统入侵检测算法。Brugman 等人[6]通过使用软件定义网络将流量路由到云，以使用网络功能虚拟化进行检查，提出了一种使用软件定义网络的基于云的工控入侵检测方法。根据上述研究成果可以得到，大多数算法模型关注到了特征提取对于工控入侵检测的重要意义，并通过相应的特征提取方法进行了实验，取得了相应的成果。但依然存在一定的局限性：

(1)对于特征提取部分仍然有提升的空间，例如对于级联卷积层的加入难以避免运算成本大和过拟合风险；对于堆叠稀疏编码器的应用，编码器只是单一地表征不同数据在隐空间的特质而忽视了其概率分布。

(2)多数算法模型的核心设计在于如何更好地进行特征提取，而忽视提取特征后的样本分类步骤，大多采用传统的二支决策分类器进行分类，存在盲目决策的风险。

针对上述问题，本文提出了一种基于变分自编码器(Variational Autoencoder，VAE)和三支决策(Three-way Decisions，TWD)的工业控制网络入侵检测算法(VAE-TWD)。该算法利用深度学习中的变分自编码器理论[7]，先针对输入数据的密集表征进行学习和编码，通过属性映射，在降低输入数据的同时进行特征提取。在训练过程中，成本函数迫使编码在隐空间内移动。然后在由均值和标准差生成的高斯分布中随机采样，并使用解码器解码成重构数据。训练完成后，编码器生成的数据即是降维后的特征。最后基于三支决策理论[8]对决策域中由于暂时信息不足而无法决策的数据进行延时决策，当获得更多粒度特征后再进行决策。三支决策理论极大程度上弥补了传统的二支决策中容错能力差，且不能依靠特征粒度的信息来对网络数据行为做出动态决策的缺点。

本文提出的 VAE-TWD 入侵检测算法模型主要优势在于：

(1)在特征提取部分利用变分自编码器在损失函数中添加一个正则项，并返回数据在隐空间中的概率分布，进而将隐空间组织得更好，解决了仅仅描述单一数据属性而导致的隐空间不规则性，使提取的特征包含更多的数据属性。

(2)采用三支决策分类器，避免传统决策器的盲目分类以满足工控网络的安全性要求。

1 相关理论

1.1 变分自编码器

变分自编码器是自动编码器的一个重要分支。但变分自编码与其他编码器相比，具有独特的优势：(1)VAE 是概率自动编码器，这就表明当神经网络训练以后，VAE 的输出会部分由概率决定(这点与降噪自编码器在训练过程中仅仅依靠随机性相反)；(2)VAE 是生成式自动编码器，这意味着 VAE 生成的数据与原始数据的重构误差很小，即非常接近原始数据。

这两个属性使变分自编码器与受限玻尔兹曼机(Restricted Boltzmann Machine，RBM)[9]类似，但是 VAE 的优点在于训练适用性更强，同时采样过程所花的时间成本更少。在使用 RBM 时，网络“热平衡”稳定的限制条件将迫使新实例的采样时间大幅度提高。同时变分自编码器执行变分贝叶斯推理，这是执行近似贝叶斯推理的有效方法。图 1 显示了变分自编码器的基本结构，在基础自动编码的结构上，

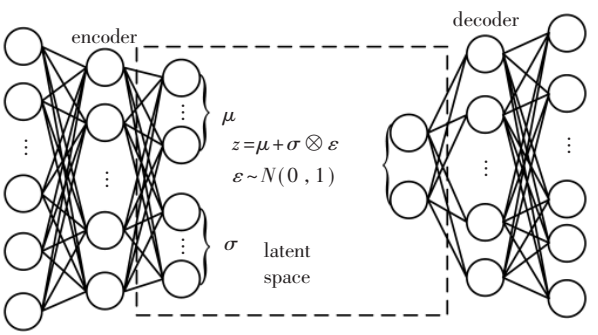

图 1 变分自编码器结构

变分自编码对编码器部分进行了改进，通过编码器在隐空间中生成平均值 μ 和标准差 σ，然后实际编码是从均值 μ 和标准差 σ 的高斯分布中随机采样得到的，而不是将原始输入进行直接编码。最后，解码器进行相反的逆操作即可将编码通过解码得到重构数据。

由图 1 可得，虽然输入的数据可能繁多且复杂，但是变分自动编码器会通过均值和标准差来生成从高斯分布中随机采样的编码：在训练过程中，首先从先验概率分布 $p_\theta(z)$ 中采样得到样本 z，成本函数迫使编码逐渐地在编码空间内移动，最终看起来像高斯点云，而后在高斯分布中采样一个随机编码，根据条件概率分布函数 $p_\theta(X_i|z)$ 解码即可得到重构出的新的网络样本实例。由于解码参数 θ、隐变量 z 和隐变量 z 的后验概率 $p_\theta(z|X_i)$ 无法直接得到，因此，VAE 的主要思想是通过一个网络 $q_\varphi(z|X_i)$ 去逼近。

根据边缘概率公式，对于单个数据样本的似然函数可以表示成：

$$\log p(X_i)=D_{\mathrm{KL}}[q_\varphi(z|X_i)||p_\theta(z|X_i)]+L(\varphi,\theta,X_i) \quad (1)$$

其中：

$$L(\varphi,\theta,X_i)=-D_{\mathrm{KL}}(q_\varphi(z|X_i)||p_\theta(z))+E_{q_\varphi}[\log p_\theta(X_i|z)] \quad (2)$$

称之为变分下界，这就是变分自编码器网络的损失函数(loss function)。式(1)中第一项即为 KL 散度[10]，可以通过优化 KL 散度来达到防止模型的过拟合；第二项为重构损失函数项，其作用是衡量原始数据与重构后数据之间的差异。

1.2 三支决策理论

三支决策理论[8]是由粗糙集理论衍生而来的，相较于传统的二支决策问题，三支决策在正向决策域和负向决策域的基础上(可用状态集 $\Omega=\{X,\overline{X}\}$ 来表示)加入了边界决策域。

通过 $D=\{D_{\mathrm{P}},D_{\mathrm{N}},D_{\mathrm{B}}\}$ 来指代三支决策的决策集，其中 D_{P} 表示正向决策域，D_{N} 表示负向决策域，D_{B} 表示边界域。对于划分数据对象属于指定区域的标准取决于有限条件集，满足接受条件的划分入正向决策域，满足拒接条件的划分入负向决策域，而对于低于接受条件而又高于拒绝条件的对象，则划分入边界域。在整个三支决策的决策过程中，针对边界域中的数据对象，在获取其他信息以后，将重新即时决定是否划分入 D_{P} 或 D_{N} 中，对于当前不符合条件的对象，将继续存放在 D_{B} 中，然后重复决策过程，直至所有数据对象划分入 D_{P} 或 D_{N} 中为止。

根据上述两个可能的状态和三个可能的决策，列出相应的决策损失函数，如表 1 所示。

表 1 决策损失函数

	属于	不属于
正向决策域	λ_{PP}	λ_{PN}
负向决策域	λ_{BP}	λ_{BN}
边界域	λ_{NP}	λ_{NN}

假定 $0\leqslant\lambda_{\mathrm{PP}}\leqslant\lambda_{\mathrm{BP}}<\lambda_{\mathrm{NP}}$，$0\leqslant\lambda_{\mathrm{NN}}\leqslant\lambda_{\mathrm{BP}}<\lambda_{\mathrm{PN}}$，式(3)和式(4)中的阈值计算公式可根据文献[11]的推倒证明而得到。

$$\alpha=\frac{\lambda_{\mathrm{PN}}-\lambda_{\mathrm{BN}}}{(\lambda_{\mathrm{PN}}-\lambda_{\mathrm{BN}})+(\lambda_{\mathrm{BP}}-\lambda_{\mathrm{PP}})} \quad (3)$$

$$\beta=\frac{\lambda_{\mathrm{BN}}-\lambda_{\mathrm{NN}}}{(\lambda_{\mathrm{BN}}-\lambda_{\mathrm{NN}})+(\lambda_{\mathrm{NP}}-\lambda_{\mathrm{PP}})} \quad (4)$$

其中，$0\leqslant\beta\leqslant\alpha<1$。根据工业控制网络环境，设置如下三条检测策略：

(1)如果 $P(X|[x])>\alpha$，将该数据划分入正向决策域，并认定该数据行为是入侵行为；

(2)如果 $P(X|[x])<\beta$，将该数据划分入负向决策域，并认定该数据行为是正常行为；

(3)如果 $\beta\leqslant P(X|[x])\leqslant\alpha$，将该数据划分入边界域，即表示当前的信息缺乏，不足以对该数据行为进行正负域的即时决策。

其中，属性集中样本的等价类通过 $[x]$ 进行表示，$P(X|[x])$ 指将等价类 $[x]$ 分为 X 的概率。

将三支决策与工控入侵检测结合，可以避免当目前的信息不足以支持做出明确的决策定论时而产生的误分类情况。显然，在工业控制网络的入侵检测中，如果将恶意的网络数据错分决策域，将导致致命性的后果。

2 入侵检测模型 VAE-TWD

2.1 算法整体流程

VAE-TWD 算法的整体流程图如图 2 所示，主要包括三个模块：数据预处理模块、变分自编码器特征提取模块和三支决策分类模块。

VAE-TWD 算法的具体步骤如算法 1 所示。

算法 1：VAE-TWD 算法

输入：数据 $X=\{x_1,x_2,\cdots,x_n\}$；三支决策阈值 α,β。

输出：网络数据分类结果。

初始化：网络参数 φ,θ；代价函数 $\lambda_{\mathrm{PP}},\lambda_{\mathrm{PN}},\lambda_{\mathrm{NP}},$

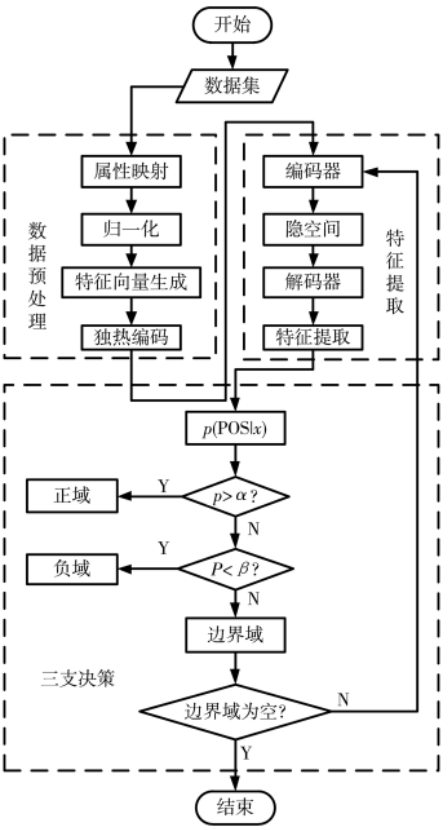

图 2 基于 VAE 和 TWD 的工控入侵检测算法流程图

λ_{BP}, λ_{BN}, λ_{NN}。

(1)数据预处理;

(2)训练变分自编码器,训练结束后编码器部分输出提取的特征;

(3)将步骤(2)中提取的特征输入三支决策分类模块;

(4)训练三支决策分类器,计算出每个数据样本属于正向决策域、负向决策域和边界域的具体概率p;

(5)将概率 $P(X|[x])$ 同三支决策的阈值 α,β 进行比较得出相应的决策结果,划分数据进入对应的决策域中;

(6)当边界域中不为空时,重复步骤(2)~(5)。

2.2 VAE 特征提取算法

变分自编码器利用概率理论将隐空间中经过编码后的特征概率表示出来,所以 VAE 中的编码器职责转化为将输入的数据通过编码转换为在隐空间的概率分布,而不是单一地描述每个数据在隐空间的属性。

假设输入数据为 $X=\{x_1, x_2, \cdots, x_n\}$,经过变分自编码器重构出的数据为 $X'=\{x_1', x_2', \cdots, x_n'\}$。将高维数据输入编码器,通过学习到的权重信息,进行降维编码到隐空间。通过解码器将数据解码为重构数据,尽可能地将其还原成原本数据。

学习过程中将解码得到的重构数据与原数据进行比较,然后利用梯度下降法动态地调整 VAE 中网络的权重和偏置,实现降低网络损失函数的目的,迫使 VAE 进一步减小重构误差。学习完成后,所需的低维特征就是经过编码器降维后的数据。

选用 SELU 函数作为 VAE 网络的激活函数,RMSprop 为优化器,针对 VAE 网络的权重参数通过梯度下降法来进行动态更新。

特征提取模块算法如下所示。

算法 2:变分自编码器特征提取算法

输入:数据集 X;网络循环次数 M。

输出:X'(输入数据 X 的低维特征数据)。

(1)随机初始化网络参数 φ,θ。

(2) For $i=1:M$

将获得的数据以概率 $p(k)$ 传递给第 k 个网络,计算相应的均值 μ_φ 与方差 σ_φ;

计算后验分布 $q_\varphi(z|X_i)$;

计算 KL 散度 $D_{KL}[q_\varphi(z|X_i)||p_\theta(z|X_i)]$;

从 $q_\varphi(z|X_i)$ 中采样 z;

计算重构误差损失 $E_{q_\varphi}[\log p_\theta(X_i|z)]$;

计算损失函数 $L(\varphi, \theta, X_i)$;

使用梯度下降法更新参数 φ,θ;

End for

(3)根据步骤(1)~(2)得到网络参数,获取高维特征的低维表示,完成特征提取。

2.3 TWD 分类算法

当降维后的特征进入三支决策分类器时,通过设定好的阈值 α,β 对不同的特征进行即时决策,划分入对应的决策域中。同时,参照文献[12]定义三支决策域中关于损失函数的经验值,如表 2 所示。

表 2 损失函数的经验值

	属于	不属于
正向决策域	0	0.7
负向决策域	0.3	0.3
边界域	1	0

根据表 2 所示的经验值,可以推导出式(3)和式(4)的相关阈值 α,β。

假设所有的数据样本集为 $X=\{x_1, x_2, \cdots, x_n\}$,则需

要计算出数据x_i属于正向决策域的概率$P(\mathrm{POS}|x_i)$，其中$i=1,2,\cdots,n$。并将阈值α,β与p值加以对比：如果$p>\alpha$，则将其分入正向决策域；假如$p<\beta$，则将其分入负向决策域；若$\beta\leqslant p\leqslant\alpha$，则分入边界域。

随着迭代次数的不断增多，编码器部分获取的特征将含括更多的信息来支撑三支决策分类器的决策，所以在VAE中加入一个多粒度特征结构，用于提取更多的特征信息进一步提升分类器的决策速度。

如果边界域中的数据样本依然存在，将继续重复该过程，直至域内数据清空。

具体算法步骤如算法3所示。

算法3：三支决策理论分类算法

输入：训练集T_x，测试集T_y。

输出：正向决策域POS，负向决策域NEG。

(1)初始化参数：VAE特征提取方式G；初始化分类器f；三支决策阈值α,β；正向决策域(POS)=负向决策域(NEG)=边界域(BND)=Ø。

(2)Do

$T_x=G(T_x)$；$T_y=G(T_y)$；

根据T_x训练模型f：

由模型f得到T_x中特征数据属于正向决策域的概率$P=f(\widetilde{T}_y)$；

For each $p\in P,t_y\in T_y$

If $p>\alpha$

$\mathrm{POS}=\mathrm{POS}\cup t_y$

Else if $p<\beta$

$\mathrm{NEG}=\mathrm{NEG}\cup t_y$

Else

$\mathrm{BND}=\mathrm{BND}\cup t_y$

End

End

$G(\mathrm{BND})\rightarrow T_y$

Until 测试集T_y为空

(3)输出：正向决策域POS，负向决策域NEG

3 实验分析

3.1 实验环境

本文提及的所有实验均在Ubuntu 16.04.7系统上完成，具体硬件配置为：GPU GV100GL[Tesla V100S Pcle 32 GB]，软件配置为：Python 3.9.2。

3.2 数据集介绍

为了评估VAE-TWD算法在工控入侵检测中的综合性能，本文采用美国橡树岭国家实验室发布的天然气真实网络数据集[13]。该数据集的网络流量数据分别来自7种恶意的网络攻击数据和正常的工控网络数据，在这其中已经定义好了1个标签以及26个特征。数据集中不同类别的数据具体分布如表3所示。

表3 数据集描述

攻击种类	数量	类型描述
Nomal	61 156	正常网络流量
NMRI	2 763	简单恶意响应
CMRI	15 466	复杂恶意响应
MSCI	782	恶意状态命令
MPCI	7 637	恶意参数命令
MFCI	573	恶意函数命令
DoS	1 837	拒绝服务攻击
Reconnaissance	6 805	侦察攻击

在实验中，将所有的正常数据作为负类，所有异常数据作为正类(即需要被检测出的类)。本文所有实验均在相同数据集和相同实验环境中进行，同时，每个实验均重复进行5次，并对得到的数据取均值，最终得到相应的实验结果。

3.3 数据预处理及评估指标

实验前进行如下数据预处理：

(1)删除缺失率高于0.7的特征避免信息冗余和资源消耗。

(2)删去唯一值的特征。

(3)通过标准化处理来避免因数据差异较大导致的误差，并采用式(5)将特征值归一化至[0,1]区间。

$$x'=\frac{x-\min(x)}{\max(x)-\min(x)} \tag{5}$$

其中，x为原始值，x'为规范化值。

(4)生成特征向量。

(5)实行独热编码。

实验选用在入侵检测领域代表综合性能的误报率(False Positive Rate，FPR)[14]、准确率(Accuracy，ACC)、检出率(Detection Rate，DR)、精确率(Precison Rate，PR)以及$F1$-得分($F1$-score，$F1$)[15]作为算法性能的评估指标，公式分别如下：

$$\mathrm{FPR}=\frac{\mathrm{FP}}{\mathrm{TN}+\mathrm{FP}} \tag{6}$$

$$\mathrm{ACC}=\frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{FP}+\mathrm{TN}+\mathrm{FN}} \tag{7}$$

$$DR=\frac{TP}{TP+FN} \tag{8}$$

$$PR=\frac{TP}{TP+FP} \tag{9}$$

$$F1=\frac{TP}{TP+\frac{FN+FP}{2}} \tag{10}$$

式中，TN 代表正常的网络行为被正确划分的数量；TP 代表恶意的网络行为被正确划分的数量；FP 表示正常的网络行为被错误划分的数量；FN 表示恶意的网络行为被错误划分的数量。

3.4 实验过程及结果分析

为了体现 VAE-TWD 算法模型对比其他算法模型的性能优势，本文进行了 3 个实验。实验一：限定特征提取方式为 VAE 进行提取，对比不同决策分类器（二支决策与三支决策）的分类表现；实验二：限定 TWD 为决策分类器进行分类，对比主流的特征提取方法与 VAE 进行特征提取时的性能差异；实验三：对比 VAE-TWD 算法模型与学术领域中其他研究人员所提出的几种算法模型之间的性能差异。

（1）实验一

此实验中，主要研究工控入侵检测和三支决策分类结合进行寻找恶意网络数据的优势，在确保变分自编码器不变的前提下，对比三支决策分类器与二支决策分类器：K 最近邻（K-Nearest Neighbor，KNN），支持向量机（Support Vector Machine，SVM），深度神经网络（Deep Neural Networks，DNN）和随机森林（Random Forest，RF）[16]在工业控制网络入侵检测领域实现网络监测的综合性能。

表 4 列出了在本文使用的数据集进行实验的性能对比结果。

表 4 不同决策方法的实验结果对比

	ACC	DR	FPR	PR	*F*1
VAE-TWD	0.982	0.965	0.041	0.984	0.976
VAE-KNN	0.768	0.643	0.038	0.856	0.742
VAE-SVM	0.879	0.735	0.049	0.937	0.869
VAE-DNN	0.885	0.865	0.053	0.838	0.796
VAE-RF	0.938	0.820	0.047	0.945	0.835

通过表 4 的结果可得，相较于传统的基于二支决策方法，本文提出的 VAE-TWD 算法模型在工控入侵检测领域，在准确率、检出率、精确率和 *F*1 分数上具有明显优势。从上述结果可以看出将三支决策理论应用于工控安全入侵检测具有积极意义。

图 3 显示了上述方法的 ROC 曲线图。从图中结果发现，本文提出的 VAE-TWD 算法模型的曲线更加接近（0，1）点，且 VAE-TWD 算法模型的曲线下方面积明显大于其他的几种方法，进一步表明了三支决策理论在分类上的优越性。

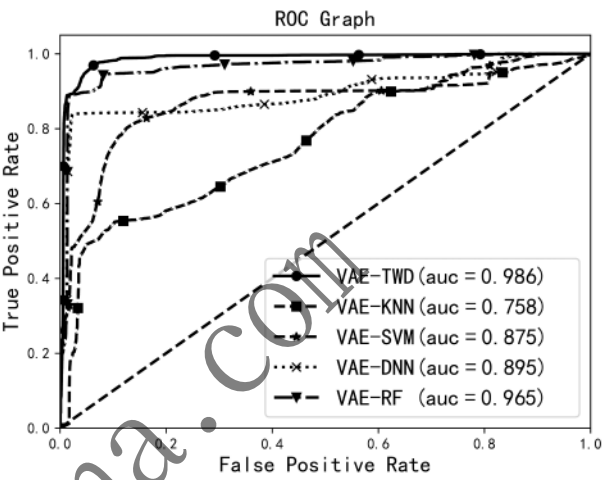

图 3 不同决策方法的 ROC 曲线图

（2）实验二

此实验中，主要探究限定 TWD 为决策分类器进行分类的前提下，当前主流的特征提取方法与 VAE 进行特征提取时的性能差异。对比的特征提取方法有奇异值分解（Singular Value Decomposition，SVD）[17]、核主成分分析（Kernel Principal Component Analysis，KPCA）[18]和深度信念网络（Deep Belief Network，DBN）。实验结果如表 5 所示。

表 5 不同特征提取模型的实验结果对比

	ACC	DR	FPR	PR	*F*1
VAE-TWD	0.982	0.965	0.041	0.984	0.976
SVD-TWD	0.954	0.893	0.037	0.928	0.933
KPCA-TWD	0.957	0.932	0.038	0.945	0.949
DBN-TWD	0.951	0.914	0.045	0.936	0.947

通过观察表 5 可得，本文所提算法 VAE-TWD 有更高的准确率、检出率、精确率以及 *F*1 得分。结果表明，通过变分自编码器中的编码器来描述每个隐属性的概率分布并提取特征的方法优于其他特征提取方法。

图 4 的 ROC 曲线对比图清楚地显示了上述特征提取方法的性能对比，从中可以看出，VAE-TWD 算法模型的 AUC 面积最大，进一步表明 VAE-TWD

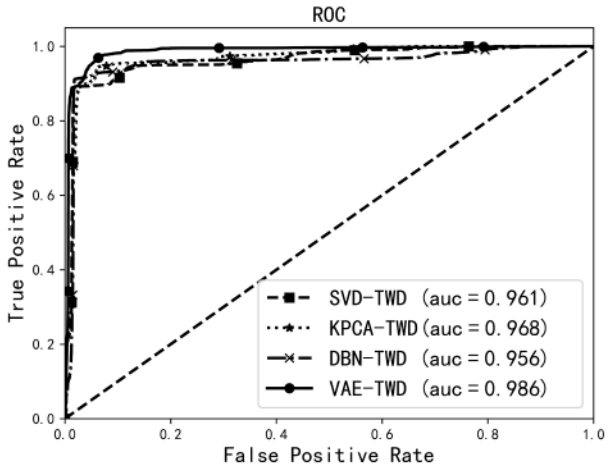

图 4 不同特征提取模型的 ROC 曲线图

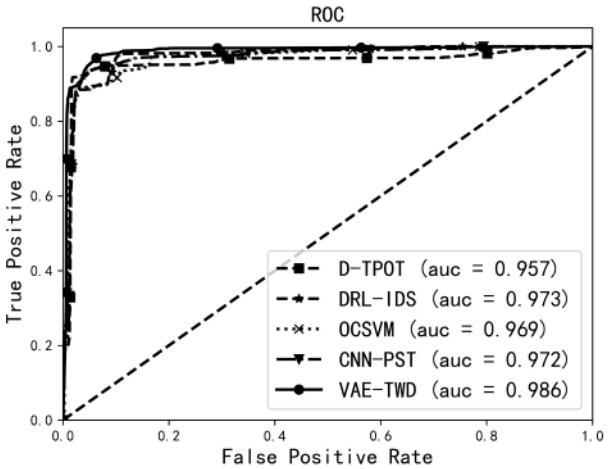

图 5 不同工控入侵检测算法的 ROC 曲线图

算法模型能更加精确地进行特征提取。

(3)实验三

此实验中,主要对比 VAE-TWD 算法模型和在工控入侵领域中其他研究人员研究的算法模型之间的性能差异。选用的算法模型包括:文献[4]提出的基于 OCSVM 的工控入侵检测算法,文献[5]提出的基于 CNN 和过程状态转换的工控入侵检测算法,文献[19]提出的改进 TPOT 的工控入侵检测方法和文献[20]提出的 DRL-IDS——基于深度强化学习的工业物联网入侵检测方法。

表 6 显示了在实验数据集上,本文算法模型与其他算法模型的性能对比。

表 6 不同工控入侵检测算法的实验结果对比

	ACC	DR	FPR	PR	*F*1
VAE-TWD	0.982	0.965	0.041	0.958	0.976
OCSVM	0.968	0.927	0.047	0.968	0.955
CNN-PST	0.979	0.908	0.050	0.917	0.969
D-TPOT	0.968	0.905	0.049	0.919	0.961
DRL-IDS	0.949	0.913	0.051	0.927	0.970

从表 6 的结果可以得到,基于 VAE-TWD 的算法模型在准确率、检出率、误报率以及 *F*1 得分上均表现优秀,特别在检出准确率上,远远地超过了其他算法模型(所有实验均使用同一数据集和相同实验环境)。尽管在精确率上存在较小劣势,但是在总体综合性能上还是占据较大的优势。

本文提出的 VAE-TWD 算法与对比算法的 ROC 曲线如图 5 所示。

从图 5 中可以看出,VAE-TWD 算法模型的 ROC 曲线相较于其他对比模型更加接近于左上角的(0,1)点。同时,在图 5 中 VAE-TWD 算法模型的 AUC 面积大于其他算法模型的 AUC 面积。综合上述结果,进一步佐证了 VAE-TWD 算法模型在工业控制网络入侵检测领域所具有的综合性能优势。

4 结论

提升工业控制网络数据的特征提取能力和分类准确率,对于工控入侵检测具有重要意义。本文提出的基于变分自编码器和三支决策的工业控制网络入侵防御检测算法,在特征提取部分利用变分自编码器在损失函数中添加正则项来返回高维数据在隐空间中的概率分布,再提取成低维抽象特征。在三支决策理论的基础上,划分经过提取的特征进入相应的决策域中,弥补了使用传统二支决策器分类的缺点。经过在真实天然气数据集上的多次实验可得,本文提出的 VAE-TWD 算法的综合性能对比文中其他工控入侵检测算法的表现更优。但是,通过实验一与实验二的结果可知,VAE-TWD 算法在误报率方面存在一定劣势。在后续的工作中,将尝试加强数据间的相关性分析以及扩充负样本数量来降低误报率。未来工作的重点在于如何降低边界域决策时的时间成本,进一步提升决策速率和对更多数据集的泛化适用性。

参考文献

[1] HU L, LI H L, WEI Z H, et al. Summary of research on IT network and industrial control network security assessment[C]//Proceedings of the 3rd IEEE Information Technology, Networking, Electronic and Automation Control Conference (ITNEC). IEEE, 2019: 1203-1210.

[2] 赵智阳,夏筱筠.基于卷积神经网络的电网工控系统入侵检测算法[J].计算机系统应用,2020,29(8):179-184.

[3] 庄卫金,方国权,张廷忠,等.基于特征抽取的电力工控系统入侵检测方法[J].浙江电力,2021,40(9):85-91.

[4] SHANG W L,ZENG P,WAN M,et al. Intrusion detection algorithm based on OCSVM in industrial control system[J].Security and Communication Networks,2016,9(10):1040-1049.

[5] LIU J J,YIN L B,HU Y,et al.A novel intrusion detection algorithm for industrial control systems based on CNN and process state transition[C]//Proceedings of the 37th IEEE International Performance Computing and Communications Conference(IPCCC).IEEE,2018:1-8.

[6] BRUGMAN J,KHAN M,KASERA S,et al.Cloud based intrusion detection and prevention system for industrial control systems using software defined networking[C]//Proceedings of Resilience Week(RWS).IEEE,2019,1:98-104.

[7] ZAVRAK S,İSKEFIYELI M.Anomaly-based intrusion detection from network flow features using variational autoencoder[J].IEEE Access,2020,8:108346-108358.

[8] NAUMAN M,AZAM N,YAO J T.A three-way decision making approach to malware analysis using probabilistic rough sets[J].Information Sciences,2016,374:193-209.

[9] LAROCHELLE H,MANDEL M,PASCANU R,et al. Learning algorithms for the classification restricted Boltzmann machine[J].Journal of Machine Learning Research,2012,13(1):643-669.

[10] 王欢良,韩纪庆,郑铁然.高斯混合分布之间KL散度的近似计算[J].自动化学报,2008,34(5):529-534.

[11] YAO Y Y.Three-way decisions with probabilistic rough sets[J].Information Sciences,2010,180(3):341-353.

[12] 张清华,吕功勋,陈玉洪,等.基于字符型属性值更新的动态三支决策模型[J].电子学报,2019,47(2):344-350.

[13] Oak Ridge National Laboratory.Transportation energy data book[M].United States:Noyes Data Corporation,2008.

[14] 唐永旺,刘欣.基于Bi-LSTM和自注意力的恶意代码检测方法[J].计算机应用与软件,2021,38(3):327-333.

[15] ZHANG D,WANG J,ZHAO X X.Estimating the uncertainty of average F1 scores[C]//Proceedings of International Conference on the Theory of Information Retrieval,2015:317-320.

[16] PAL M.Random forest classifier for remote sensing classification[J].International Journal of Remote Sensing,2005,26(1):217-222.

[17] WALL M E,RECHTSTEINER A,ROCHA L M.Singular value decomposition and principal component analysis[M]//A practical Approach to Microarray Data Analysis.Springer,Boston,MA,2003:91-109.

[18] HOFFMANN H.Kernel PCA for novelty detection[J].Pattern Recognition,2007,40(3):863-874.

[19] 杨瑞君,何立君,程燕.一种改进TPOT的工控入侵检测方法[J].小型微型计算机系统,2022,43(4):767-772.

[20] 李贝贝,宋佳芮,杜卿芸,等.DRL-IDS:基于深度强化学习的工业物联网入侵检测系统[J].计算机科学,2021,48(7):47-54.

(收稿日期:2022-04-28)

作者简介:

王晨(1998-),男,硕士研究生,主要研究方向:信息安全、模式识别与智能系统等。

张迪明(1986-),通信作者,男,博士,主要研究方向:操作系统、形式化验证等。E-mail:seventeen_wzc@163.com。

韩斌(1968-),男,博士,教授,主要研究方向:信息安全、智能检测等。

版权声明

经作者授权，本论文版权和信息网络传播权归属于《信息技术与网络安全》杂志，凡未经本刊书面同意任何机构、组织和个人不得擅自复印、汇编、翻译和进行信息网络传播。未经本刊书面同意，禁止一切互联网论文资源平台非法上传、收录本论文。

截至目前，本论文已经授权被中国期刊全文数据库（CNKI）、万方数据知识服务平台、中文科技期刊数据库（维普网）、JST 日本科技技术振兴机构数据库等数据库全文收录。

对于违反上述禁止行为并违法使用本论文的机构、组织和个人，本刊将采取一切必要法律行动来维护正当权益。

特此声明！

《信息技术与网络安全》编辑部

中国电子信息产业集团有限公司第六研究所