

基于扩张卷积神经网络的异常检测模型

高治军,曹浩东,韩忠华

(沈阳建筑大学电气与控制工程学院,辽宁 沈阳 110168)

摘要 目的 提出一种基于 DCNN-MiLSTM 的异常检测模型,解决传统的网络异常检测模型难以处理具有时序特征网络流量数据的问题。方法 对原始流量数据的时间标签进行重定义,利用扩张卷积神经网络对整体特征进行提取,同时引入 Mogrifier LSTM 网络,对时序信息进行深层次挖掘。结果 与其他异常检测模型相比,DCNN-MiLSTM 模型的准确率达到 99.12%,召回率为 98.94%, F_1 值为 99.03%,各项指标均优于其他常见模型,提升了检测异常网络流量数据的能力。结论 DCNN-MiLSTM 模型可以更好地处理具有时序特征的流量,捕捉流量数据中的时间依赖关系和趋势,更有效地检测和识别出异常数据。

关键词 网络异常检测;扩张卷积神经网络;标签重定义;时序特性

中图分类号 TP393

文献标志码 A

引用格式:高治军,曹浩东,韩忠华.基于扩张卷积神经网络的异常检测模型[J].沈阳建筑大学学报(自然科学版),2024,40(4):738-744.(GAO Zhijun, CAO Haodong, HAN Zhonghua. Anomaly detection model based on extended convolutional neural network[J]. Journal of Shenyang jianzhu university (natural science), 2024, 40(4): 738-744.)

Anomaly Detection Model Based on Extended Convolutional Neural Network

GAO Zhijun, CAO Haodong, HAN Zhonghua

(School of Electrical and Control Engineering, Shenyang Jianzhu University, Shenyang, China, 110168)

Abstract: A DCNN-MiLSTM-based anomaly detection model is proposed to solve the problem that traditional network anomaly detection models are difficult to handle network traffic data with temporal characteristics. The timestamps of the original traffic data are redefined, and the overall features are extracted by using an expansive convolutional neural network, while the Mogrifier LSTM network is introduced for deeper mining of temporal information. Compared with other anomaly detection models, the DCNN-MiLSTM model achieves an accuracy of 99.12%, a recall of 98.94%, and a F_1 of 99.03%, which are better than other common models in all metrics, and

收稿日期:2023-02-14

基金项目:国家重点研发计划项目(2018YFF0300304-04);辽宁省重点研发计划项目(2020JH2/10100039);辽宁省教育厅高等学校基本科研重点项目重点项目(LJKZ0583);沈阳市中青年科技创新人才支持计划项目(RC200026)

作者简介:高治军(1978—),男,教授,主要从事智慧建筑信息化技术和网络安全技术等方面研究。

improves the ability of detecting anomalous network traffic data. The DCNN-MiLSTM model can better deal with traffic flows with temporal characteristics, capture the time in traffic data dependencies and trends in traffic data, and more effectively detect and identify anomalous data.

Key words: network anomaly detection; extended convolutional neural network; label redefinition; temporal characteristics

互联网的快速发展提高了人们的生活水平,给人们带来了全新的生活方式。与此同时,互联网的安全问题也不容忽视。近些年来网络安全事件大幅度增加,已经成为了严重危害社会的安全问题之一,网络安全已提升到国家安全的层面,因此网络安全解决方案的社会需求日益增多。

目前已存在一些网络入侵的检测方式,为了提高检测数据库中已知攻击的检测准确率,L. Zhang 等^[1]提出了一种基于签名的入侵检测系统,使用一组特定的规则来识别已知模式,提高了检测的准确率,缺点是很难观察单个规则如何与整个系统一起工作,且无法检测未知的攻击。M. Kumar 等^[2]提出了一种使用区块链和云计算基础设施的分布式检测系统,能以更快、更简单的方法来识别跨多个网段协调的攻击并轻松发现复杂的攻击模式从而采取预防措施,使得系统能更全面地识别出各种入侵,但该系统的通信延迟较高,成本较大,难以普及。为了准确检测大规模数据集,J. V. A. Sukumar 等^[3]提出了一种改进的遗传 k 均值聚类算法模型,该模型不用事先固定集群的数量,在使用大规模集群时准确率非常高,但是在使用较小的数据集时,准确率较低。C. Chen 等^[4]提出了一种具有粒子群优化灰狼优化器的模型,以提高基于 SVM 检测的整体性能,但是容易陷入局部最优解。S. Zheng 等^[5]提出了通过改进集成分类器,减少数据流不平衡对集成分类器性能的影响,对于未知攻击的识别率高,但该算法波动较大,不适用于实际检测。

基于上述研究,笔者提出一种 DCNN-MiLSTM 异常检测模型,先将原始流量数据

进行集成,对时间标签进行重定义,并利用扩张卷积神经网络训练数据,对整体特征进行提取,同时引入 Mogrifier LSTM 网络,对时间信息进行深层次挖掘;研究表明:DCNN-MiLSTM 模型在准确率、召回率和 F_1 值等方面均优于其他常见模型,可以增加网络流量数据的检测能力。

1 DCNN-MiLSTM 异常检测模型

1.1 扩张卷积神经网络

在深度学习中,特征图每个单元的值是由卷积输入的一个区域决定的,这个输入的区域就叫做感受野^[6]。感受野就是 CNN 每层的特征图在原图上映射出来的像素大小^[7]。CNN 的神经元不能对原始图像的所有信息进行感知,这是因为在这些结构中都会使用 convolution 层和 pooling 层,两个层之间采用局部相连的方式^[8]。模型的卷积核越大,神经元就能提取出更整体、更全面的信息;但变化的卷积核也意味着参数的急剧增多^[9]。扩张卷积神经网络能很好地解决这个问题^[10]。

扩张卷积神经网络和普通的卷积神经网络相比,加入了一个新的参数,即扩张率。扩张率是定义卷积核中心之间间距的参数^[11]。当扩张率为 1 时,它与普通的卷积神经网络相同;当扩张率变为 2 时, 3×3 的卷积核所具有的视野将与 5×5 的卷积核相同。在和 3×3 的卷积核使用相同算力的前提下,能提取更大的整体信息^[12]。

用 r_n 来表示第 n 个卷积核中每个单元的感受野,其计算式如下:

$$r_n = r_{n-1} + (k_{n-1}) \prod_{i=1}^{n-1} s_i \sqrt{b^2 - 4ac}. \quad (1)$$

式中: k_{n-1} 为第 $n-1$ 个卷积层的卷积核大小; s_i 为第 n 个卷积层的步长; a, b, c 分别为卷积核的高度、宽度和通道数。

使用这种缩小后的卷积核进行运算,可以减少参数量,节省计算时间,提高训练的效率^[13]。

1.2 形变长短期记忆网络

LSTM 共包含 4 个门控系统,分别是遗忘门、输入门、候选记忆细胞和输出门^[14]。

LSTM 各个参数的计算公式如下:

$$F_t = \sigma(W_f x_t + W_{f'} h_{t-1} + b_f). \quad (2)$$

$$I_t = \sigma(W_i x_t + W_{i'} h_{t-1} + b_i). \quad (3)$$

$$\tilde{C}_t = \tanh(W_c x_t + W_{c'} h_{t-1} + b_c). \quad (4)$$

$$O_t = \sigma(W_o x_t + W_{o'} h_{t-1} + b_o). \quad (5)$$

$$C_t = F_t \odot C_{t-1} + I_t \odot \tilde{C}_t. \quad (6)$$

$$H = O_t \odot \tanh(C_t). \quad (7)$$

式中: $F_t, I_t, \tilde{C}_t, O_t, C_t$ 和 H 分别为遗忘门、输入门、候选记忆细胞、输出门、记忆细胞、新一轮的隐藏状态; σ 为 sigmoid 运算; $W_f, W_{f'}, W_i, W_{i'}, W_c, W_{c'}, W_o, W_{o'}$ 为权重矩阵; x_t 为输入序列的第 t 个时间步的输入; h_{t-1} 为第 $t-1$ 个时间步的隐藏状态; b_f, b_i, b_c 和 b_o 为偏置。

遗忘门:对上层的遗忘状态进行控制,输入的为当前时刻的信息。上一时刻的隐藏状态,经过激活函数处理后,得到一个处于 $0 \sim 1$ 的输入,与记忆细胞的数据计算乘积,这个值靠近 0,就忽略记忆的值。

输入门:和遗忘门相似,先经过激活函数的处理,得到一个处于 $0 \sim 1$ 的数值。当输入门靠近 0,遗忘门靠近 1 的时候保存记忆细胞的元素。

候选记忆细胞:和遗忘门不同,候选记忆细胞将激活函数换成了 \tanh 函数,所得数值在 $-1 \sim 1$ 。

输出门:对记忆细胞流动到下一时间的隐藏状态进行控制。

和循环神经网络作比较,长短期记忆网络使用了门的机制,能更好地提取出数据中的时序信息^[15]。

通过上述对长短期记忆网络的分析,我

们能够看到, x_t 和 h_{t-1} 之间没有交互,这对网络的性能产生了一定影响。让两个状态交互之后,便产生了形变长短期记忆网络 Mogrifier LSTM^[16]。

形变长短期记忆网络的相关公式如下:

$$x_t^i = 2\sigma(Q^i h_{t-1}^{i-1}) \odot x_t^{i-2}, i = 1, 2, \dots, r. \quad (8)$$

$$h_{t-1}^i = 2\sigma(R^i x_t^{i-1}) \odot h_{t-1}^{i-2}, i = 1, 2, \dots, r. \quad (9)$$

其中, Q_i, R_i 为交互矩阵; x_t^i, h_{t-1}^i 增加了参数 i, i 是交互需要的一个参数,可以决定两个状态以何种方式进行计算。当交互的轮次为奇数次时使用式(8)进行交互,轮次为偶数次时使用式(9)进行交互。经过激活函数处理后,式(8)和式(9)的值在 $0 \sim 1$,多次计算后会慢慢接近于 0。

对长短期记忆网络性能的限制是上下文没有关联的输入,因此只需要对输入和隐藏状态进行交互,就可以优化网络的性能,提升模型的效果,对整个网络产生正面影响。

1.3 基于 DCNN-MiLSTM 的异常检测模型

针对传统网络异常检测模型无法有效提取流量数据中的时序信息问题,笔者提出一种基于 DCNN-MiLSTM 的异常检测模型,该模型采用了两种不同的网络对流量数据进行训练,深度挖掘流量中的时序信息。首先对数据集进行预处理,包括数据的数值化、归一化和集成。数值化是把模型难以直接识别的字符型特征转化为容易识别的数字型特征,归一化是为了避免原数据各特征值差距过大对模型的影响,数据的集成化是将多条数据进行集成,采用二进制编码组合重定义方法对输入的流量信息进行二次编码,将集成后的数据输入到 DCNN-MiLSTM 网络中,利用集成后的数据包含前后时序关系的特点,使用扩张卷积神经网络对其进行特征提取,将输出的数据输入到 Mogrifier LSTM 中,进一步提取流量中的时序信息,从而解决异常检测中难以处理流量数据之间时序信息的问题,实验结果证明 DCNN-MiLSTM 模型能够提高异常检测的准确率和召回率。图 1 为

DCNN-MiLSTM 模型的工作流程图。

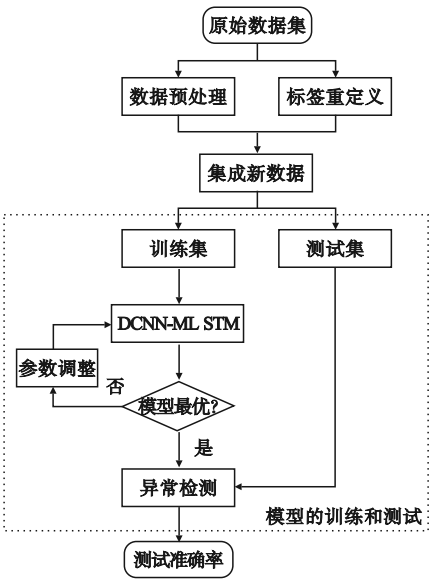


图 1 DCNN-MiLSTM 模型的工作流程

Fig. 1 The process chart of DCNN-MiLSTM model

1.4 数据数值化

异常检测模型输入为数值型特征,因此需要对数据中包含的字符特征进行数值化处理。KDDCUP99 数据集共包含三种类型的协议,即 tcp,udp 和 icmp,采用 one-hot 编码后,分别为[1,0,0],[0,1,0],[0,0,1],处理后的数据维数将增加。

1.5 数据归一化

经过数值化处理后的数据不同特征之间的差异很大,这会提升个别特征的比重系数,影响模型的训练效果,因此要对网络流量数据进行归一处理。笔者采用 z-score 标准化方法,将数据集中在[-1,1],其公式如下:

$$x'_i = \frac{x_i - \bar{x}}{s}.$$
 (10)

式中: \bar{x} 为样本平均数; s 为样本的标准差; x_i 为该条数据在某一维度的特征值; x'_i 为处理后的特征值。

1.6 数据集成

网络流量数据之间存在着一定的时间相关性,通过将多条数据集成,利用二进制编码的标签重定义方法对原始标签进行重新编码^[17]。

记 $F_i = \{x_{i1}, x_{i2}, \cdots, x_{in}\}$ 为单条数据,原数据对应的时间标签为 t_i , n 为特征的维数, m 为数据条数,集成系数为 q ,利用二进制编码重定义方法,集成后的数据为 $F'_i = \{x_{i1}, \cdots, x_{in}, x_{(i+1)1}, \cdots, x_{(i+1)n}, \cdots, x_{(i+m)n}\}$,对应集成后的标签为 t'_i ,从而产生一组新的数据。标签重定义方法如表 1 所示,其中集成系数 $q \geq 1$ 。

表 1 标签重定义方法

Table 1 The label redefinition method		
集成系数	原标签	集成标签
$q = 1$	(0,0)	0
	(0,1)	1
	(1,0)	2
	(1,1)	3
$q = 2$	(0,0,0)	0
	(0,0,1)	1
	(0,1,0)	2
\vdots	\vdots	\vdots
	\vdots	\vdots

与表 1 类似,当集成系数 q 继续变大时,按相同方式进行标签重定义。同时,在最后的结果输出部分,对数据进行标签还原,还原结果如表 2 所示。

表 2 标签还原方法

Table 2 The label restoration method		
集成系数	原标签	集成标签
$q = 1$	0	(0,0)
	1	(0,1)
	2	(1,0)
	3	(1,1)
$q = 2$	0	(0,0,0)
	1	(0,0,1)
	2	(0,1,0)
\vdots	\vdots	\vdots
	\vdots	\vdots

将数据标签进行还原之后,可以判断单条数据的状态。例如, q 为 1,重定义标签为 2 时,可以得到还原标签为(1,0)的组合。

2 实验仿真

2.1 实验环境

为了验证笔者提出的异常检测模型,建立仿真环境,使用 Keras2. 2. 4 深度学习框架对该模型进行仿真,操作系统为 Windows11 系统,i5-11400H 6 核处理器,内存 16 GB,显卡为 RTX 3050 Ti Laptop GPU,使用 Python3. 5 进行编程,数据集为 KDDCUP99。

2.2 评估方式

使用训练集来训练模型,用测试集来计算模型的准确率^[18]。记将正类预测为正类数为 T_p ,负类预测为负类数为 T_N ,负类预测为正类数为 F_p ,正类预测为负类数为 F_N ,则准确率 P 、召回率 R 和 F_1 可以表示为

$$P = \frac{T_p}{T_p + F_p} \tag{11}$$

$$R = \frac{T_p}{T_p + F_N} \tag{12}$$

$$F_1 = \frac{2T_p}{2T_p + F_p + F_N} \tag{13}$$

准确率是异常检测模型最重要的评判指标。召回率的含义是在实际为正的样本中被预测为正样本的概率^[19],它体现了模型能够找出所有真实目标的能力。准确率和召回率互相影响,无法做到双高, F_1 值就是用来衡量准确率和召回率的指标,它反应了准确率和召回率的加权平均,可以更全面的评估模型的性能。在实际情况下,需要在保证准确率较高的情况下,召回率也不那么低^[20]。

2.3 模型分析

不同参数的设置会很大程度的影响模型的训练结果,如学习率的设置、激活函数的选择、不同的学习算法等^[21]。将不同的参数分别进行实验,不同学习率的对比实验结果如图 2 所示。

由图 2 可知,不同的学习率对模型的损失率有较大的影响,笔者选择损失率最小时的学习率 0. 01。

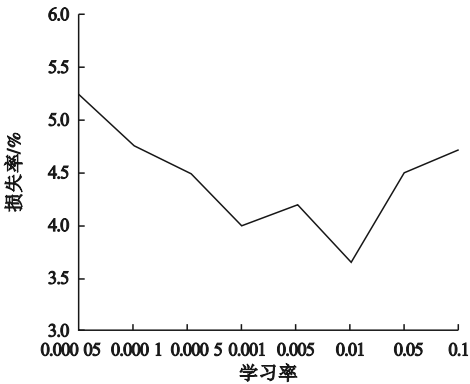


图 2 不同学习率的对比实验

Fig. 2 The comparison experiment result with different learning rates

常用的三种激活函数的对比实验结果如图 3 所示。

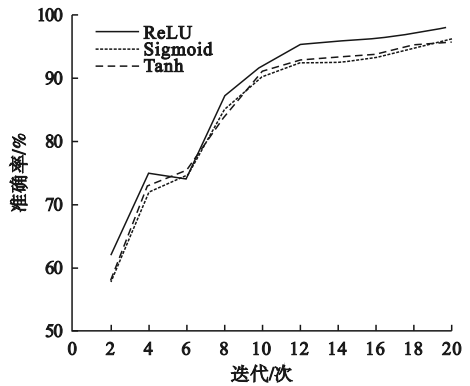


图 3 不同激活函数的对比

Fig. 3 The comparison result with different activation functions

由图 3 可知,使用 ReLU 激活函数有更高的准确率,且 ReLU 激活函数收敛速度比 Sigmoid 激活函数和 Tanh 激活函数更快,不会有梯度消失的问题,因此笔者采用 ReLU 激活函数。

使用深度学习最常用的三种优化算法 SGD、RMSProp 和 Adam 进行实验。不同优化算法的对比实验如图 4 所示。

由图 4 可知,随着迭代次数的增加,三种优化算法的准确率曲线均成上升趋势。在相同迭代次数下,Adam 优化算法的曲线图准确率更高,故该模型使用 Adam 算法。

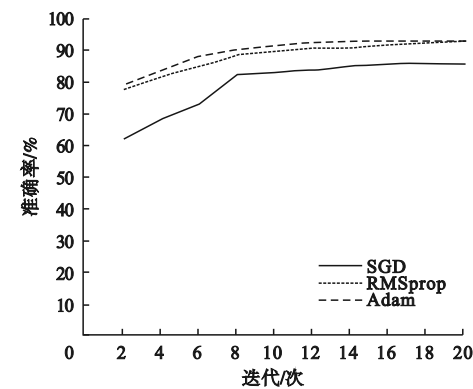


图4 不同优化算法的对比实验

Fig. 4 The comparison result with different optimization algorithms

将笔者提出的 DCNN-MiLSTM 模型与单独使用 DCNN 和 Mogrifier LSTM 进行对比,实验结果如图 5 所示。由图可知,笔者所建模型在准确率上要明显优于其他两种单独网络的模型。

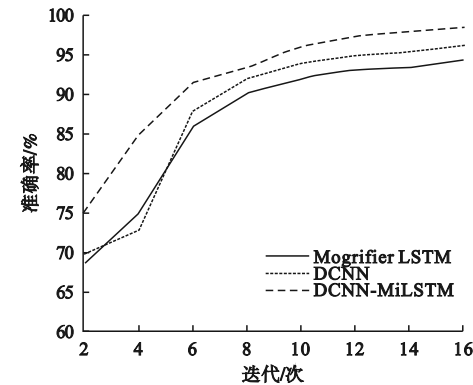


图5 三种深度学习算法对比结果

Fig. 5 The comparison result with three deep learning algorithms

常见的不同异常检测模型之间的正确率对比如表 3 所示,包含 CPSO-SVM、聚类算法、朴素贝叶斯模型、随机森林模型和笔者提出的 DCNN-MiLSTM 模型。

表3 不同模型对比

模型	准确率/%	召回率/%	F_1 /%
CPSO-SVM	96.12	95.23	95.67
聚类算法	93.26	93.12	93.19
朴素贝叶斯	97.06	97.02	97.04
随机森林	98.23	98.11	98.17
DCNN-MiLSTM	99.12	98.94	99.03

由表 3 可知,笔者所提模型在准确率、召回率和 F_1 值上优于其他几种模型。该模型准确率为 99.12%,提升了 0.89%,能够尽可能地检测出正确的结果;在保证较高准确率的基础上,召回率也达到了 98.94%,进一步减少了漏报概率; F_1 值为 99.03%,比其他几种异常检测模型更加稳健,可以有效地增加异常检测能力。

3 结 论

(1)提出了一种 DCNN-MiLSTM 异常检测模型,通过在实际网络流量数据集上的实验验证了其有效性;所提模型不仅在准确率上有所提高,召回率也有所上升,具有更低的漏报率, F_1 值也高于传统模型,因此该检测模型更加稳健,性能更加优秀。

(2)该检测模型收敛速度较快,训练时间较短,可以捕捉流量数据中的时间依赖关系,解决了传统的网络异常检测模型难以处理具有时序特征的数据的问题。

(3)该检测模型关注提取数据的时序特征,对于低时序特征的数据检测能力不足,还需寻找可以升处理低时序特征数据的方法,例如引入 GCN 网络来提升处理能力。

参考文献

[1] ZHANG L, DU H L. Research on SDN intrusion detection based on online ensemble learning algorithm [C]// International conference on networking and network applications (NaNA). Haikou, China: IEEE, 2020:114 – 118.

[2] KUMAR M, SINGH A K. Distributed intrusion detection system using blockchain and cloud computing infrastructure [C]//The 4th international conference on trends in electronics and informatics (ICOEI) (48184). Tirunelveli, India: IEEE, 2020 :248 – 252.

[3] SUKUMAR J V A, PRANAV I, NEETISH M M, et al. Network intrusion detection using improved genetic k -means algorithm [C]// International conference on advances in computing, communications and informatics (ICACCI). Bangalore, India: IEEE, 2018: 2441 – 2446.

[4] CHEN C, SONG L, BO C, et al. A support vector machine with particle swarm

- optimization grey wolf optimizer for network intrusion detection [C]// International conference on big data analysis and computer science (BDACS). Kunming, China: IEEE, 2021:199 – 204.
- [5] ZHENG S. Network intrusion detection model based on convolutional neural network [C]// The 5th advanced information technology, electronic and automation control conference (IAEAC). Chongqing, China: IEEE, 2021: 634 – 637.
- [6] DU H, ZHANG Y. Network anomaly detection based on selective ensemble algorithm [J]. The journal of supercomputing, 2021, 77 (1): 2875 – 2896.
- [7] RAFI A, MADNI T M, JANJUA U I, et al. Multi-level dilated convolutional neural network for brain tumour segmentation and multi-view-based radiomics for overall survival prediction [J]. International journal of imaging systems and technology, 2021, 31 (3): 1519 – 1535.
- [8] DING S S, WANG Y X, KOU L. Network intrusion detection based on BiSRU and CNN [C]//The 18th international conference on mobile ad hoc and smart systems (MASS). Denver, CO, USA: IEEE, 2021:145 – 147.
- [9] LI F, ZHOU H, WANG Z, et al. ADDCNN: an attention-based deep dilated convolutional neural network for seismic facies analysis with interpretable spatial-spectral maps [J]. IEEE transactions on geoscience and remote sensing, 2020, 59 (2): 1733 – 1744.
- [10] TRUNG N T, TRINH D H, TRUNG N L, et al. Dilated residual convolutional neural networks for low-dose CT image denoising [C]// Asia pacific conference on circuits and systems (APCCAS). Halong, Vietnam: IEEE, 2020:189 – 192.
- [11] KARTHIKEYAN B, MOUNIKA M, PRAVALLIKA S B, et al. Deep CNN with residual learning and dilated convolution for image denoising [C]//The second international conference on electronics and sustainable communication systems (ICESC). Coimbatore, India: IEEE, 2021:1327 – 1333.
- [12] 陈虹, 李泓绪, 金海波. 多尺度卷积与双注意力机制融合的入侵检测方法[J]. 辽宁工程技术大学学报(自然科学版), 2024, 43 (1): 93 – 100.
(CHEN Hong, LI Hongxu, JIN Haibo. Intrusion detection method based on multi-scale convolution and dual attention mechanism [J]. Journal of Liaoning technical university (natural science), 2024, 43 (1): 93 – 100.)
- [13] LUO Yangbiao, WU Muqing, XU Weiyao. Effective fusion of 3DCNN and convolutional GRU for gesture recognition [C]//The 6th international conference on computer and communications (ICC). Chengdu, China: IEEE, 2020:2453 – 2457.
- [14] 王凯, 陈丹伟. 基于LSTM的动态图模型异常检测算法研究[J]. 计算机工程与应用, 2019, 55 (5): 76 – 82.
(WANG Kai, CHEN Danwei. Research on algorithm of dynamic graph anomaly detection based on LSTM [J]. Computer engineering and applications, 2019, 55 (5): 76 – 82.)
- [15] 王健, 易姝慧, 刘浩. 基于注意力机制优化长短期记忆网络的短期电力负荷预测[J]. 中南民族大学学报(自然科学版), 2023, 42 (1): 73 – 81.
(WANG Jian, YI Shuhui, LIU Hao. Attention mechanism optimization long short term memory network based short-term load forecasting [J]. Journal of south-central university for nationalities (natural science edition), 2023, 42 (1): 73 – 81.)
- [16] 赵丽丽, 朱恒伟, 刘聪, 等. 基于形变—时控长短期记忆网络的医学事件表示学习方法[J]. 德州学院学报, 2022, 38 (4): 29 – 35.
(ZHAO Lili, ZHU Hengwei, LIU Cong, et al. Representation learning method of medical event based on mogrifier-time long short term memory [J]. Journal of Dezhou university, 2022, 38 (4): 29 – 35.)
- [17] HE Zhaorong, LUO Xiaonan, ZHONG Yanru. Information extraction method based on dilated convolution and character-enhanced word embedding [C]// International conference on cyber-enabled distributed computing and knowledge discovery (cyberC). Chongqing, China: IEEE, 2020:138 – 143.
- [18] 孙晓璇, 张磊, 李健. 目标检测数据集半自动生成技术研究[J]. 计算机系统应用, 2019, 28 (10): 8 – 14.
(SUN Xiaoxuan, ZHANG Lei, LI Jian. Research on semi-automatic generation technology of object detection datasets [J]. Computer systems and applications, 2019, 28 (10): 8 – 14.)
- [19] PARAMPOTTUPADAM S, MOLDOVANN A N. Cloud-based real-time network intrusion detection using deep learning [C]// International conference on cyber security and protection of digital services (cyber security). Glasgow, UK: IEEE, 2018:1 – 8.
- [20] AYO F E, FOLORUNSO S O, ABAYOMI-ALLI A A, et al. Network intrusion detection based on deep learning model optimized with rule-based hybrid feature selection [J]. Information security journal: a global perspective, 2020, 29 (6): 267 – 283.
- [21] YU C. Shear resistance of cold-formed steel framed shear walls with 0.686 mm, 0.762 mm, and 0.838 mm steel sheet sheathing [J]. Engineering structures, 2010, 32 (6): 1522 – 1529.

(责任编辑:刘春光 英文审校:范丽婷)