

基于 GAF-DenseNet 的航空发动机虚假数据 注入攻击检测

黄鹏程¹, 陈丽丹², 祁 恬³, 张 哲⁴, 马永良⁵, 高 明⁵

- (1. 昆明理工大学 信息工程与自动化学院, 昆明 650504;
2. 广州航海学院 轮机工程学院, 广州 510725;
3. 华南理工大学 自动化学院, 广州 510640;
4. 人工智能与数字经济广东省实验室, 广州 510335;
5. 广州城市理工学院 电气工程学院, 广州 510800)

摘 要: 提出一种基于格拉姆角场 (Gramian angular field, GAF) 和密集连接卷积网络 (densely connected convolutional networks, DenseNet) 的航空发动机系统遭受虚假数据注入攻击的机器学习检测方法。首先, 基于美国国家航空和宇宙航行局的商用模块化航空推进系统仿真数据集 (commercial modular aero-propulsion system simulation, C-MAPSS), 构建了连续和间隔虚假数据注入两种攻击模型; 其次, 通过 GAF 算法, 在保留原始航空发动机传感器获得的时序信号的时间依赖性的前提下, 对时间序列数据进行唯一编码, 并设计了 DenseNet-121 网络对图像阵列中内含的传感器信息进行深层挖掘, 进而检测航空发动机是否遭受虚假数据注入攻击及攻击类型识别; 最后, 融合 GAF-DenseNet 方法在 T24、T50 和 P30 传感器上的平均分类准确率为 98.46%, 与长短期记忆、门控循环单元和卷积神经网络对比分别提高了 1.91%、3.82% 和 0.38%。

关 键 词: 航空发动机; 商用模块化航空推进系统仿真数据 (C-MAPSS); 虚假数据注入攻击;
格拉姆角场 (GAF); 密集连接卷积网络 (DenseNet)

中图分类号: V263.6; TP18

文献标志码: A

False data injection attacks detection for aero engine system based on GAF-DenseNet

HUANG Pengcheng¹, CHEN Lidan², QI Tian³, ZHANG Zhe⁴,
MA Yongliang⁵, GAO Ming⁵

- (1. Faculty of Information Engineering and Automation,
Kunming University of Science and Technology, Kunming 650504, China;
2. Faculty of Marine Engineering, Guangzhou Maritime University, Guangzhou 510725, China;
3. School of Automation, South China University of Technology, Guangzhou 510640, China;
4. Guangdong Artificial Intelligence and Digital Economy Laboratory, Guangzhou 510335, China;
5. Department of Electrical Engineering,
Guangzhou City University of Technology, Guangzhou 510800, China)

Abstract: A machine learning detection method for aero-engine system false data injection attacks based on Gramian angular field (GAF) and densely connected convolutional networks (DenseNet) was

收稿日期: 2022-08-28

基金项目: 智能电网四川省重点实验室开放基金(2021-IEPGKLS-PKFA01)

作者简介: 黄鹏程(1999—), 男, 硕士生, 主要研究方向为深空探测中的小行星着陆图像分析。

通信作者: 陈丽丹(1981—), 女, 副教授, 博士, 主要研究方向为电动汽车与交通、能源网交互、智能技术应用。E-mail: chenld@gcu.edu.cn

引用格式: 黄鹏程, 陈丽丹, 祁恬, 等. 基于 GAF-DenseNet 的航空发动机虚假数据注入攻击检测[J]. 航空动力学报, 2023, 38(7): 1691-1702.

HUANG Pengcheng, CHEN Lidan, QI Tian, et al. False data injection attacks detection for aero engine system based on GAF-DenseNet[J]. Journal of Aerospace Power, 2023, 38(7): 1691-1702.

proposed. Firstly, two attack models of continuous and interval spurious data injection were constructed based on the simulation dataset of NASA's commercial modular aero-propulsion system simulation (C-MAPSS). Secondly, the GAF method was proposed to transform the timing signal obtained by the aero-engine sensors into the image signal, and a DenseNet-121 network was designed to detect whether the aero engine was subject to false data injection attack and the type of attack was identified. Finally, the average classification accuracy of GAF-DenseNet method on T24, T50, and P30 sensors was 98.46%, which was 1.91%, 3.82%, and 0.38% better compared with long and short-term memory, gated recurrent units, and convolutional neural networks, respectively.

Keywords: aero engine; commercial modular aero-propulsion system simulation (C-MAPSS); false data injection attack; Gramian angular field (GAF); densely connected convolutional networks (DenseNet)

航空发动机作为飞行器的核心设备,其运行状态健康与否直接决定了飞机能否安全飞行^[1]。剩余寿命(remaining useful life, RUL)预测是发动机故障预测与健康管理(prognosis and health management, PHM)的重要组成部分^[2],依据当前状态监测数据或历史数据等信息,通过特征提取和趋势预测技术,构建寿命预测模型,得到航空发动机的剩余寿命。RUL能够用于动态感知航空发动机健康状况的未来变化,为预防性维修提供技术支撑^[3]。但是,倘若网络攻击者对航空发动机监测数据实施虚假数据注入攻击(false data injection attack, FDIA)^[4],可能会导致通过原有PHM预测出的航空发动机RUL出现增大且大于最大维护时间的结果,从而造成飞机正常飞行时间航空发动机出现因未及时检修的故障,最终危及飞行器的安全飞行。所以,面对航空发动机传感器及PHM受到网络攻击挑战,探索FDIA机理,建立识别FDIA的检测方法,从而实现及时维护航空发动机各部件,保障发动机运行安全具有重要意义。

然而,目前学术界和工业界研究人员较多关注于数据驱动的RUL预测方法及准确性^[5-9],忽视了FDIA的影响,已发表文献中航空发动机FDIA及其检测方法的研究较少。FDIA是指通过注入虚假数据或改变信号发送策略,进而导致系统行为失常或失稳而不被探测机制发现的网络攻击,其概念最初是在智能电网领域提出的^[10]。隐蔽的FDIA不易被系统检测,严重危害电力、航空等系统的安全可靠运行。随着传感器为基础的工业4.0和物联网技术的不断发展和应用,一些除电力系统外的如无线传感器网络^[11]、停车场管理系统^[12]、航空发动机系统^[13-16]等也成为FDIA的对象。文献[13]模拟了FDIA对航空发动机剩余

寿命预测准确性的影响,结果表明,即使在少数传感器实施FDIA也会严重破坏RUL预测准确性。文献[14-16]关注航空广播式自动相关监视系统(automatic dependent surveillance-broadcast, ADS-B)安全性,ADS-B的作用是要求飞机定期广播其位置和速度,但攻击者可对ADS-B系统发起各种无线攻击,在面对FDIA时,特别是在当前飞机运行状态通信依赖的背景下,攻击者有很大可能修改、阻止或发出虚假ADS-B消息,以欺骗控制员和监视系统。ADS-B受到网络攻击的挑战,文献[14]和文献[15]模拟了欺骗干扰攻击,并采用基于长短记忆网络(long short-term memory network, LSTM)的异常检测算法,取得较好效果。Cretin等^[16]对ADS-B系统进行FDIA,进行了建模和预测研究,考虑了幽灵飞机数据注入(包括单架和多架飞机信息)、虚假轨迹修改攻击、假警报攻击和飞机失踪攻击等多类型FDIA,该研究同样运用了LSTM作为FDIA的检测系统。

本文利用美国国家航空航天局(简称NASA)公开的飞行器涡轮风扇发动机从正常运行到失效的全寿命采集的实验数据集C-MAPSS^[17],研究了航空发动机数据FDIA的建模及其识别方法。在数据检测约束下构建了连续FDIA和间隔FDIA两种不同的攻击方式;FDIA检测方法上,分别利用长短期记忆网络(LSTM)、门控循环单元网络(gate recurrent unit network, GRU)和卷积神经网络(convolutional neural network, CNN)算法对原始航空发动机传感器采集的时序数据是否遭受虚拟数据注入攻击进行识别;并进一步探索了利用GAF数据变换方法将C-MAPSS数据中的时序数据转换为图像数据,从而构建密集连接卷积网络DenseNet-121,深度挖掘时序数据特征进行FDIA

识别。

1 问题描述

1.1 航空发动机预测性维护系统

航空发动机预测性维护系统如图 1 所示, 通

常由机载系统与地面系统两部分组成, 机载系统的任务是监测并采集航空发动机各部件的原始数据(通过传感器传输), 对各部件的使用与损伤状况进行跟踪; 地面系统则主要根据机载系统采集到的数据, 利用算法, 对各部件 RUL 预估。

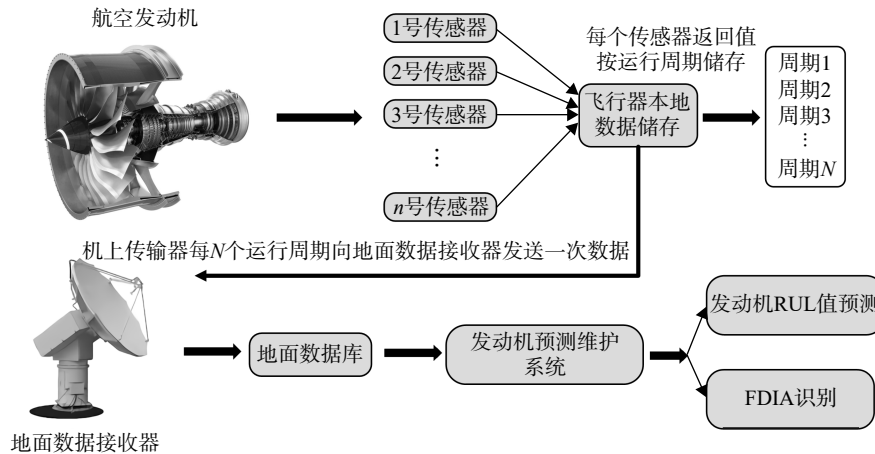


图 1 航空发动机预测性维护系统

Fig. 1 Aero engine health monitoring system architecture

1.2 RUL 预测数据集

目前较多采用 NASA 的 C-MAPSS 数据集、挑战赛数据集 PHM08^[18] 用于航空发动机 RUL 研究, 并且以前者为多。C-MAPSS 数据集包含 4 个子数据集(FD001、FD002、FD003 和 FD004), 每个子数据集均由训练数据和测试数据组成。需要注意, C-MAPSS 数据集中的传感器和执行器没有动态特性, 只是统一增益, 没有计算时间延迟也不会出现错误或偏差, 其目的是保证系统的快速执行速度^[19]。测试数据集示例如表 1 所示, 表中 ID(identity document)表示身份标识。

表 1 C-MAPSS 数据集示例
Table 1 Samples of dataset of C-MAPSS

发动机 ID	运行周期	第 1 列	第 2 列	...	第 26 列
1	1	-0.000 7	-0.000 4	...	23.419 0
1	2	0.001 9	-0.000 3	...	23.423 6
1	3	-0.004 3	0.000 3	...	23.344 2
...	...	...	...	...	...
1	192	0.000 9	0	...	22.964 9
2	1	-0.001 8	0.000 6	...	23.458 5
2	2	0.004 3	-0.000 3	...	23.408 5
...	...	...	...	...	...
2	287	-0.000 5	0.000 6	...	23.084 8
3	1	0.000 8	0.000 5	...	23.320 5
...	...	...	...	...	...
100	200	-0.003 2	-0.000 5	...	23.052 2

每个数据集的每一行都是一个时间周期(相当于一个飞行小时, 直到航空发动机寿命结束^[13], 时间周期结束), 一个时间周期均有 26 列, 其中第 1 列表示发动机 ID 编号, 第 2 列表示当前操作周期号, 第 3 到 5 列表示 3 个操作设置项, 第 6 到 26 列表示 21 个发动机传感器返回值。时间序列数据只有在出现故障时才会终止, 例如表 1 中 ID 为 1 的发动机有 192 个时间周期的数据, 这意味着该发动机在第 192 个时间周期时故障。

1.3 问题描述

本文将研究针对图 2 所示系统中传感器返回值(以 C-MAPSS 为对象)实施 FDIA 和其检测的问题。

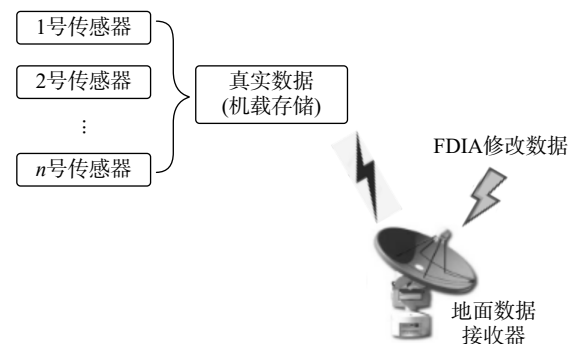


图 2 航空发动机 FDIA 示意图

Fig. 2 Schematic diagram of FDIA of aero engine

FDIA 的方式多种多样,主要的方式有直接攻击物理层面的传感器返回值、攻击传感器进行数据传输的数据链路层和攻击进行传感器数据处理的程序层。航空发动机 FDIA 的具体方式为,网络攻击者向发动机传感器中注入错误数据,本文实现的 FDIA 出现在发动机传感器向中央发动机控制系统传输返回值时,且攻击频率主要取决于攻击意图,攻击频率可以看作被攻击系统的一个运行周期(本文中可看作飞行器的一个常规飞行周期,即起飞到降落的过程)^[20]。对航空发动机而言,遭受 FDIA 后会导致发动机 RUL 预测值升高且高于最大安全运行时间,最终导致发动机运行超出安全时限,不能得到及时的维护,严重者会导致发动机的一些零部件失灵,甚至导致航空器在飞行途中发动机失效、飞机坠落等重大事故的发生。

2 FDIA 模型

2.1 航空发动机 FDIA 数学描述

航空发动机受到 FDIA 可以描述为:攻击者在原始时序数据中注入一连串的错误信息。例如,设 X_i 为第 i 个传感器传输的信息,其形式为 $X_i = [x_1, x_2, \dots, x_k]$,表示第 i 个传感器所得到的包含 k 个时间序列的时序数据。通过向原始传感器数据中注入相同维数的 FDIA 向量即可实现对原始数据的污染。设 $F_i = [\lambda_1, \lambda_2, \dots, \lambda_k]$ 为第 i 个传感器所受到的数据污染向量,则式(1)表示被攻击后的最终传感器数据:

$$Z_i = X_i + F_i \quad (1)$$

一般地,可将 FDIA 分为两种攻击方式,一种为不受限攻击,即攻击者可以选择任意传感器进行虚假数据注入,达到完全掌控传感器和其通信链路层的目的;另一种方式为受限攻击,即攻击者只可在有限的传感器中进行数据攻击操作。第2种方式与现实更为贴近,本文 FDIA 研究主要考虑该种方式,即在多维传感器返回值中选择其中几维进行数据注入攻击。

另外,考虑到真实 FDIA 的隐蔽性质,即注入后的 FDIA 应不被航空器防御监测系统捕获,达到欺骗航空器预测维护(predictive maintenance, PdM)系统的效果,因而本文建模过程中考虑 FDIA 时不超过一定的传感器测试边界范围的约束,具体如式(2)所示:

$$Z_{\min} \leq Z_i \leq Z_{\max} \quad (2)$$

式中 $Z_{\min}$ 和 $Z_{\max}$ 为常数,为该传感器所能检测的边界范围(FDIA 偏差攻击的隐蔽性^[20]),本文中实施的 FDIA 范围大小设定为原信号的 -0.2% ,另外为模拟数据注入过程中的信号误差,加入原数据的 $0.01\% \sim 0.05\%$ 大小的信号噪音,建模方式参考了 Mode 等的做法^[13]。

2.2 连续 FDIA 模型

连续 FDIA 的含义是从原始时序数据的某一时刻开始进行虚假数据注入,直到原始时序数据结束,即整个攻击过程是连续的,一旦攻击操作开始,就破坏该传感器的后续所有读数。例如,攻击在 $A_{\text{start}}=3$ 的瞬间开始,并在 A_{end} 结束, F_i 可表示为 $F_i = [\lambda_1, \lambda_2, \lambda_{A_{\text{start}}}, \dots, \lambda_{A_{\text{end}}}]$,其中 $A_{\text{start}} \geq 1$, $A_{\text{end}} = k$ 。

图3为向发动机 ID 为 1 的 T24 传感器(传感器编号为 7)原始数据中注入连续 FDIA 之后的数据对比示例,其中实线代表原始时序数据,短虚线表示受到连续 FDIA 后的虚假数据。攻击持续周期为 130~192 周期。

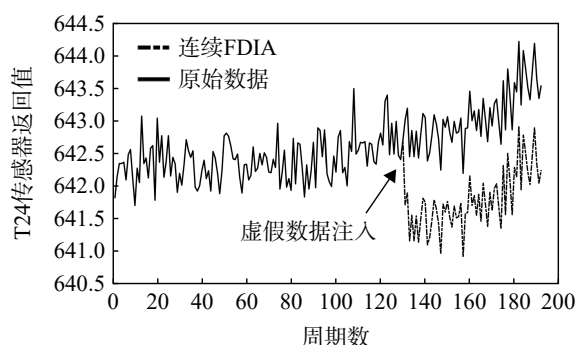


图3 连续 FDIA

Fig. 3 Continuous FDIA

2.3 间隔 FDIA 模型

间隔 FDIA 的含义是从原始时序数据的某一时刻开始进行虚假数据注入,持续一段时间周期,即整个攻击过程是间隔的,攻击操作开始后,只破坏该传感器的后续一段时间内的读数。例如,在总时间长度为 100 的原始数据下,如果攻击在 $A_{\text{start}}=3$ 的瞬间开始,并在 $A_{\text{end}} = k$ 时刻结束,那么 F_i 可以表示为 $F_i = [\lambda_1, \lambda_2, \lambda_{A_{\text{start}}}, \dots, \lambda_{A_{\text{end}}}, \dots, \lambda_{100}]$ 。

图4为发动机 ID 为 1 的 T24 传感器(传感器编号为 7)原始数据中注入间隔 FDIA 之后的数据对比示例,其中实线代表原始数据,短虚线表示受到间隔 FDIA 后的虚假数据。攻击持续周期为 130~150 周期。

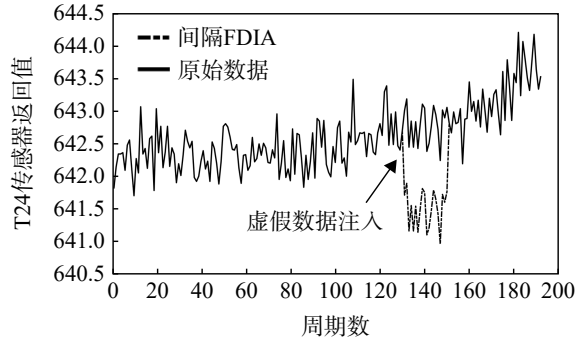


图4 间隔 FDIA

Fig. 4 Intermittent FDIA

3 基于 GAF-DenseNet 的 FDIA 检测算法

3.1 数据预处理 GAF 图像编码

基于坐标变换的思想, GAF^[21]可在保留数据时间依赖性的前提下, 将一维时间序列编码为二维灰度图像, 包括格拉姆角场相合(GASF)和格拉姆角场相差(GADF)方法^[22]。其基本原理如下: ①提取数据中的 N' 个点的平均值, 该 N' 个时间序列进行聚合操作以减少数据大小, 聚合过程采用分段聚合近似算法(piecewise aggregation approximation, PAA); ②将聚合后的数据映射到 $[0, 1]$ 区级内, 并通过时序信号中的时间周期数作为绘制极坐标图的半径, 同归一化后的数据进行反余弦计算, 生成极坐标值; ③利用格拉姆矩阵, 分为相合和相差的方法, 将每对值相加或相减, 最后取余弦值后进行求和汇总, 得到时序数据转化为的图像数据。格拉姆矩阵定义如式(3), 式(4)表示在二维空间上, 两个向量 u 和 v 之间的内积。

$$G = \begin{pmatrix} \langle u_1, v_1 \rangle & \langle u_1, v_2 \rangle & \cdots & \langle u_1, v_{n'} \rangle \\ \langle u_2, v_1 \rangle & \langle u_2, v_2 \rangle & \cdots & \langle u_2, v_{n'} \rangle \\ \vdots & \vdots & & \vdots \\ \langle u_{n'}, v_1 \rangle & \langle u_{n'}, v_2 \rangle & \cdots & \langle u_{n'}, v_{n'} \rangle \end{pmatrix} \quad (3)$$

$$\langle u, v \rangle = u_i \cdot v_i + u_j \cdot v_j \quad (4)$$

以如图5正弦时序数据为例说明 GAF 变换的效果, 原始时间序列一维数据(1D)经过分段聚合逼近和缩放(PAA)后, 进行极坐标变换, 再利用格拉姆矩阵, 通过相合或相差操作, 得到原始一维时间序列数据的二维图像。

格拉姆矩阵中由于时间随着位置从左上角到右下角的移动而增加, 所以时间维度被编码到矩阵的几何结构中, 其较好地保留了时间依赖性,

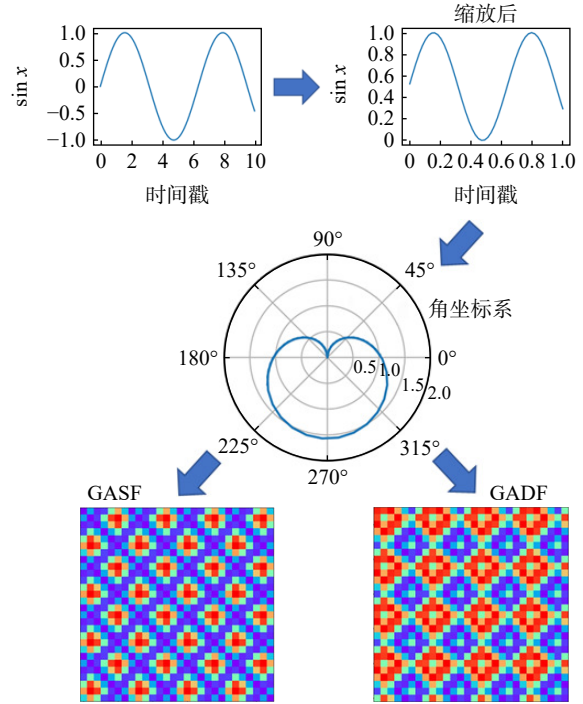


图5 格拉姆数据变换

Fig. 5 Data transformation based on GAF

矩阵主对角线包含了原始值和角度信息, 可有效表示原时间序列的时间相关性。

3.2 DenseNet 模型

近年来, 深度学习日益发展, 应用于实际生活中的语音识别、图像识别等各类识别。DenseNet 模型沿用残差网络(residual network, ResNet)中短路连接的基本思想^[23], 在此基础上强化了前层与后层之间的密集连接, 通过通道间的拼接以实现特征重用。其网络特点是被定义为密集块(dense block)的核心模块, 结构如图6所示。图6(a)中的C代表卷积层(convolution)。

设密集块网络包含 L 层, 每层卷积后均输出 g 个特征图, g 作为超参数被定义为增长率, 假定输入密集块的特征图通道数为 g_0 , 则第 L 层输入的通道数为 $g_0 + g_{L-1}$ 。各层输出的特征图大小一致, 特征图在通道维度上进行拼接, 第 L 层的输出 x_L 由前面所有层拼接而得, 即

$$x_L = H_L([x_0, x_1, \cdots, x_{L-1}]) \quad (5)$$

式中 $H_L(\cdot)$ 为非线性转化函数, 为一个组合操作, 包括批标准化(batch normalization, BN)层、线性修正单元(rectified linear unit, ReLU)和卷积(convolution, conv)层组成, 特征传递方式是直接将前面所有层的特征连接后输入到下一层。

密集块结构作为基本单元, 与过渡层(trans-

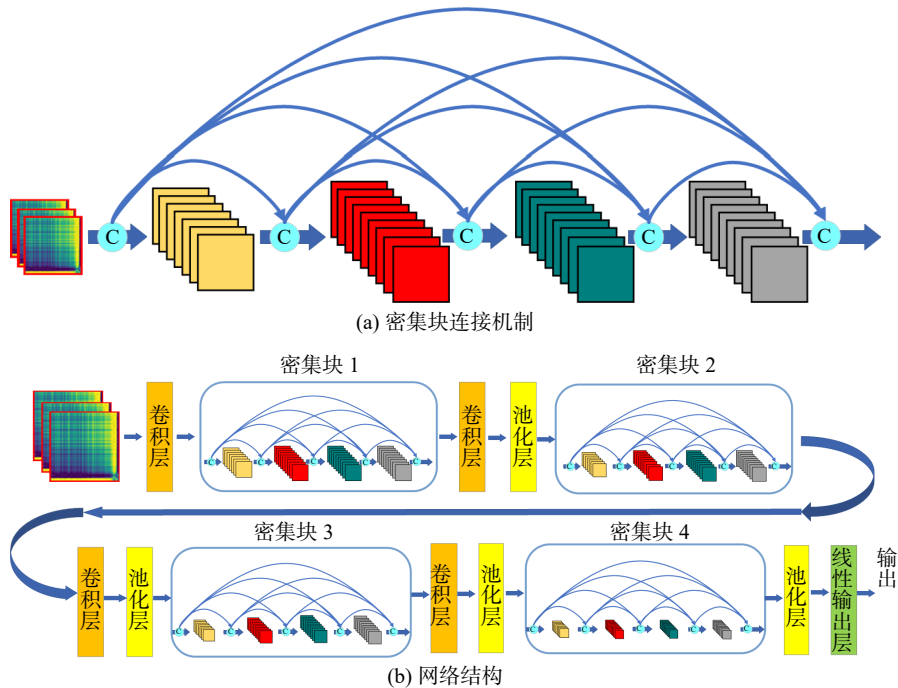


图 6 DenseNet 的网络结构

Fig. 6 Structure of DenseNet

ition layer)交替串联构成 DenseNet 模块。分析其结构,优势主要体现在:①具有更强的梯度流动,由于密集连接方式, DenseNet 提升了梯度的反向传播,使得网络更容易训练;②减少了训练的参数数量;③保持了低维度的特征,在标准的卷积网络中,最终输出只会利用提取最高层次的特征,而在 DenseNet 中,它使用了不同层次的特征,倾向于给出更平滑的决策边界。因而,实际问题中即便训练数据较少的情况下, DenseNet 也能有较好的表现。

3.3 GAF-DenseNet 模型

在 DenseNet 模块的基础上,为改善特征图信息密度不足的问题,通过 GAF 编码算法在时间、幅值两个维度上对原始航空发动机传感器获得的

时间序列进行解构,将一维时间序列重组为二维灰度图像矩阵,实现特征信息密度的有效扩充,并使特征信息在二维图像空间中排列规则化,从数据层面间接提升分类精度。最终,将 GAF 编码算法和 DenseNet 融合形成 GAF-DenseNet 检测模型,整体结构如图 7 所示。图 7 中 1D 和 2D 分别表示一维和二维。

GAF-DenseNet 检测分类模型由 GAF 编码器、DenseNet 模块和分类器共 3 个模块组成。GAF 编码器将输入模型的多通道航空发动机传感器数据序列编码为二维灰度图像阵列, DenseNet 模块依靠卷积层提取二维灰度图像阵列为特征图,并在通道维度上对特征图进行特征重用和密集连接,以优化特征的传递机制和利用效率。其中,为提

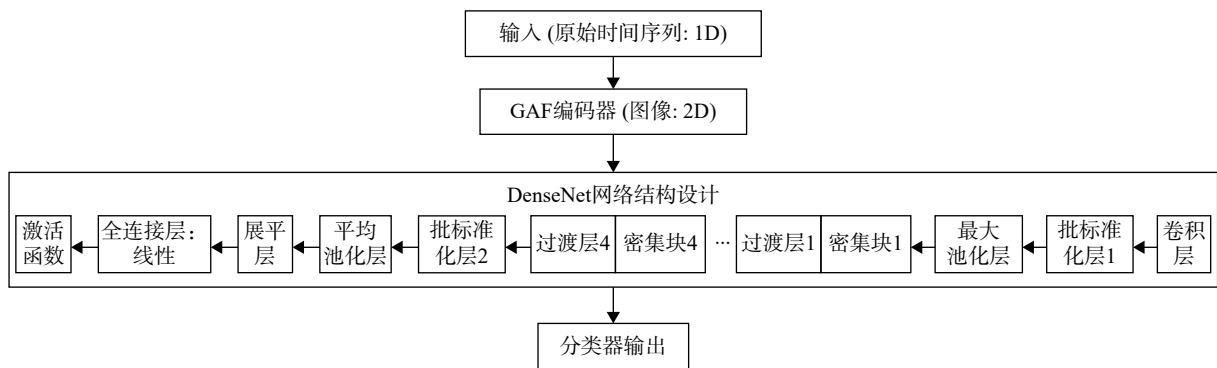


图 7 GAF-DenseNet 模型整体结构

Fig. 7 Overall structure of GAF-DenseNet model

高网络紧凑性,引入过渡层以缩减特征图大小。过渡层包含批标准化层、 1×1 卷积层和 2×2 平均池化层,对于密集块所输出的 m 个特征图,经过过渡层转换后可产生 θ_m 个特征, $\theta \in (0,1]$ 为压缩系数。

DenseNet 模块输出的特征进入分类器后,经过全局平均池化进入 3 层全连接网络。其中,最后一个全连接层使用 Softmax 分类器输出分类预测结果。本文所提的 GAF-DenseNet 模型参数配置如表 2 所示,表中 max pool 为最大池化层,average pool 为平均池化层, stride 为步长。

表 2 DenseNet 结构层次
Table 2 Proposed network structure of DenseNet

层类型 (layers)	输出结构 (output size)	DenseNet-121
卷积 (convolution)	112×112	7×7 conv, stride2
池化 (pooling)	56×56	3×3 max pool, stride2
密集块 1 (dense block_1)	56×56	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$
过渡层 1 (transition layer_1)	56×56	1×1 conv
	28×28	2×2 average pool, stride2
密集块 2 (dense block_2)	28×28	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$
过渡层 2 (transition layer_2)	28×28	1×1 conv
	14×14	2×2 average pool, stride2
密集块 3 (dense block_3)	14×14	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 24$
过渡层 3 (transition layer_3)	14×14	1×1 conv
	7×7	2×2 average pool, stride2
密集块 4 (dense block_4)	7×7	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 16$

4 实验结果及分析

4.1 实验数据集及设置

本文利用 C-MAPSS 数据集,对于 FDIA 场景,选取了 21 个传感器中的 3 个传感器(具体编号分别为 T24、T50 和 P30)的返回时序数据作为 FDIA 的对象。

针对 T24、T50 和 P30,在整个运行周期内的返回值进行提取,并按照发动机 ID 号,从 1 到 100 排列(测试集中包含 100 个发动机的所有传感器数据),且分别为每一个传感器的返回值按运行周期顺序建立各自的数据集。例如, ID 为 1 的发动机有 192 个实际运行周期,则所得到的

T24 号传感器返回值为长度是 192 个周期的时序数据,对所有发动机执行上述相同操作。最终生成 3 个针对传感器 T24、T50 和 P30 返回值的数据列表。文中连续 FDIA 开始时刻设为 130,一直持续到该传感器时序数据记录结束。间隔 FDIA 开始时刻为 130,持续时间为 20,即到该传感器时序数据周期数为 150 时结束。

另外,由于 FDIA 建模需要一定时间长度的时序信息,则需要在原始传感器的返回值中进行筛选,筛选掉过短的时序信号,具体做法是筛选出信号周期大于 150 且小于等于 200 的发动机 ID 号,并将时间序列长度未满 200 个周期的数据末尾通过补 0 的方式填充到 200 个时间周期长度,测试集也用相同方式进行筛选操作。

本文攻击检测仿真和验证运用 Matlab 和 Google 的 Colab 代码云端运行平台,对于数据集的预处理方法以及 FDIA 的建模采用 Matlab 实现,对于 LSTM、GRU、CNN 以及 DenseNet 的构建训练以及测试工作均在 Colab 平台上进行,并利用 Google 提供的免费图形处理器(graphics processing unit, GPU)进行训练集的并行运算。train_batch_size 设为 10,使用二元交叉熵损失(binary cross entropy loss, BCEloss)作为损失,初始学习率为 0.000 1,且随着训练次数增加递减。超参数设置部分参考了文献[22]的相关设置。

4.2 数据预处理

本文将提出的 GAF-DenseNet 航空发动机 FDIA 检测方法与传统基于 LSTM、GRU 和 CNN 网络的机器学习检测算法对比。其中 LSTM、GRU 和 CNN 网络是基于时序数据训练和检测,由于航空发动机每个传感器获得的时序信号量纲不同,因而先将数据集进行标准化,再进行数据可视化分析。而对于 DenseNet,模型输入基础数据应为图像形式,所以需先将时序数据通过 GAF 变换为图像数据。

4.2.1 LSTM、GRU 和 CNN 网络数据预处理

首先,将训练集和测试集数据标准化,将时序数据缩放至区间 $[0, 1]$,如式(6)所示:

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

式中 $\max(x)$ 、 $\min(x)$ 分别为传感器的最大返回值和最小返回值, x 为初始数据。

其次,将数据集集中的原始 26 维的所有数据经过标准化之后进行可视化,即以颜色的深浅对应

每个数值的大小,如图 8 所示。其中纵坐标代表原始数据的总行数,横坐标表示 26 维数据的列数,故可分析原始数据每列之间的相关程度。

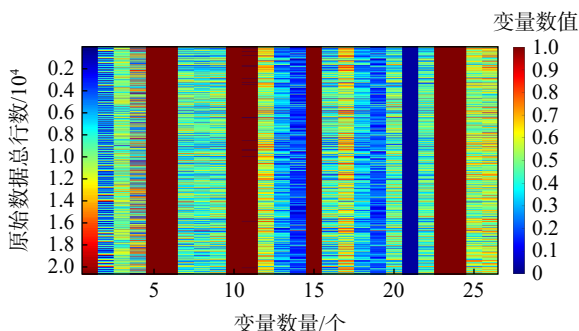


图 8 经标准化处理的原始训练集数据

Fig. 8 Standardized processing of original training set data

4.2.2 GAF-DenseNet 数据预处理

采用 GASF 和 GADF 算法将原始发动机 T24、T50 和 P30 传感器数据受连续 FDIA 和间隔 FDIA 虚假数据进行图像编码。以 T24 传感器为例,其训练集图像数据如图 9 所示。

图 9(a)、图 9(b)和图 9(c)分别为原始时间序列数据、连续 FDIA 后时间序列数据和间隔 FDIA 后时间序列数据经 GASF 变换后的二维图像;图 9(d)、图 9(e)和图 9(f)则分别为对应的 GADF 变换后的二维图像。

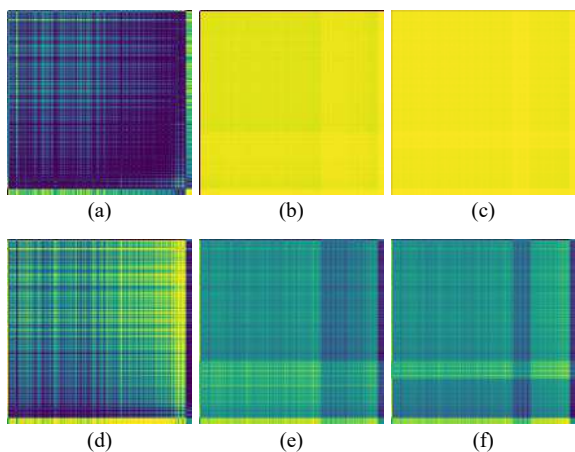


图 9 传感器 T24 原始数据、受攻击数据图像化

Fig. 9 Sensor T24 raw data, attacked data visualization

接着,将每张图像进行大小放缩,缩放成规格为 $224 \times 224 \times 3$ 的图像;再随机选取某些位置的图像进行水平翻转做数据增强,以提高其训练的难度以及训练完成后测试数据时的准确性,即提高模型的泛化性能;随后将图像数据转化为 ToTensor 类型,进行神经网络运算,之后将图像

取均值化、方差化,完成数据图像预处理流程。

4.3 基于时序数据的 FDIA 识别

本文选取了 3 个航空发动机传感器(分别为 T24、T50 和 P30)的返回值作为输入,运用 LSTM、GRU 和 CNN 网络分别对这 3 个传感器所受到的连续 FDIA 和间隔 FDIA 进行虚假数据注入检测和分类,共训练了 9 个网络,其中针对每个传感器分别用 3 种不同的网络实现了判断特定传感器是否受到 FDIA 以及受到何种 FDIA 的分类。

对 T24、T50 和 P30 传感器通过预处理得到了 3 个训练集和 3 个测试集数据,训练集结构为一个 $(144, 1, 200)$ 的张量,测试集数据结构为 $(87, 1, 200)$,在总测试集中筛选出了 29 个发动机总周期数符合大于 150 且小于 200 的对应传感器返回值,加上对于连续 FDIA 和间隔 FDIA 后的训练集数据,最终得到了 87 行数据。标签设定与训练集一致,将未遭受数据攻击的原始数据标签定为 0,将受到连续 FDIA 的数据定为 1,受到间隔 FDIA 的数据定为 2。训练集标签数据结构为 $(144, 1)$,其中前 48 行为 0,中间 48 行为 1,最后 48 行为 2,测试集同理结构为 $(87, 1)$,其中前 29 行的标签为 0,中间 29 行为 1,最后的 29 行对应标签数据值为 2。

网络得到的预测结果为数值,对于测试集输出预测结果为 $(87, 1)$ 的数据结构,每一行对应输入信号的最终分类数值,选取了分类界限为 0.5 和 1.5 的两条分界线,当预测值小于 0.5 时判定该行信号为未受 FDIA,当预测值大于 0.5 且小于 1.5 时分类为受到连续 FDIA,当预测值大于 1.5 时,将其分类为受到间隔 FDIA。3 个网络的结构和参数设置如表 3 所示。

对航空发动机 T24、T50 和 P30 传感器的原始时间序列数据实施连续 FDIA 和间隔 FDIA,对比 LSTM、GRU 和 CNN 检测算法的结果分别如图 10、图 11 和图 12 所示。图中 σ 为方均根误差 (root mean square error, RMSE)。

由图 10~图 12 可得在所有网络的分类预测

表 3 LSTM、GRU 和 CNN 网络结构

Table 3 Network structure of LSTM, GRU and CNN

网络	隐藏层数	每层的节点数	随机失活比例	批量大小	激活函数
LSTM	4	100	0.2	48	tanh
GRU	4	100	0.2	48	tanh
CNN	4	64	0.2	48	ReLU

中,只要是遭受了 FDIA 数据攻击,3 个网络都能进行准确地识别,即没有出现将连续 FDIA 的数据或间隔 FDIA 的数据错分类为“未受到 FDIA”。但是 LSTM、GRU 和 CNN 网络对于受到连续 FDIA 后的数据识别中都有误差产生,且对于受到间隔 FDIA 的数据预测中,错分类数据相同,通过原数据对比,得到原因是与该数据周期长度仅比 150 个周期稍大有关。

具体的数据对比如下:对于间隔 FDIA 测试数据的分类,由图 10 可知,86 号(总共有 87 个测试数据)测试数据预测值接近 1,3 个网络的分类结果都将其分类为受到了连续 FDIA,查找实际受到间隔 FDIA 的该数据,发现其时间周期长度正

好为 150,即该数据所受到连续 FDIA 和受到间隔 FDIA 后生成的测试数据是一样的,从图 10 中也可以看到对应的不同网络分类得到的数据预测值相同,就 LSTM 网络来说都为 1.230 601 5 (在图 10(a)中用绿色圈圈出),在另外两个网络中也有相同结果(在图 10(b)和图 10(c)中用绿色圈圈出);同样在测试数据集中分析错将受到间隔 FDIA 数据分类为连续 FDIA 数据的数据特征,得到第 81 号测试数据的时间周期长度为 152,仅比间隔 FDIA 的攻击所在时间周期 130~150 多了 2 个时间周期,从网络的所学习的数据特征来看,81 号测试数据即受到间隔 FDIA 后的数据(在图 10 中用绿色箭头标注),与第 52 号测试数据集

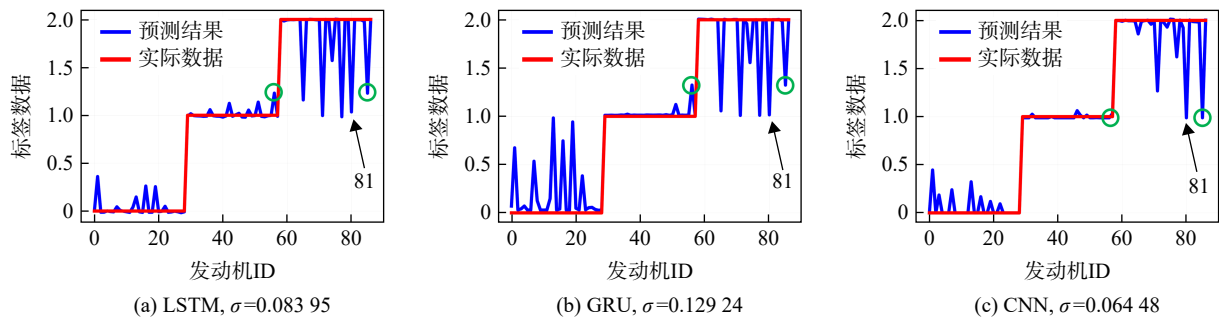


图 10 T24 的 LSTM、GRU 和 CNN 结果

Fig. 10 Results for T24 based on LSTM, GRU and CNN

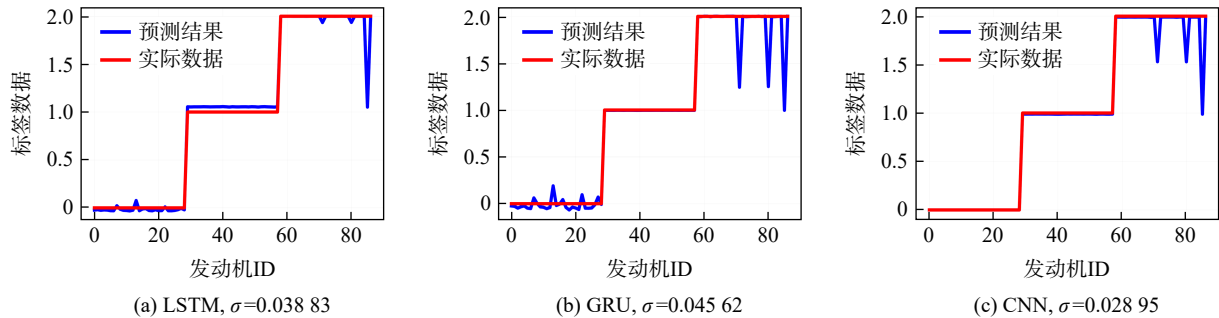


图 11 T50 的 LSTM、GRU 和 CNN 结果

Fig. 11 Results for T50 based on LSTM, GRU and CNN

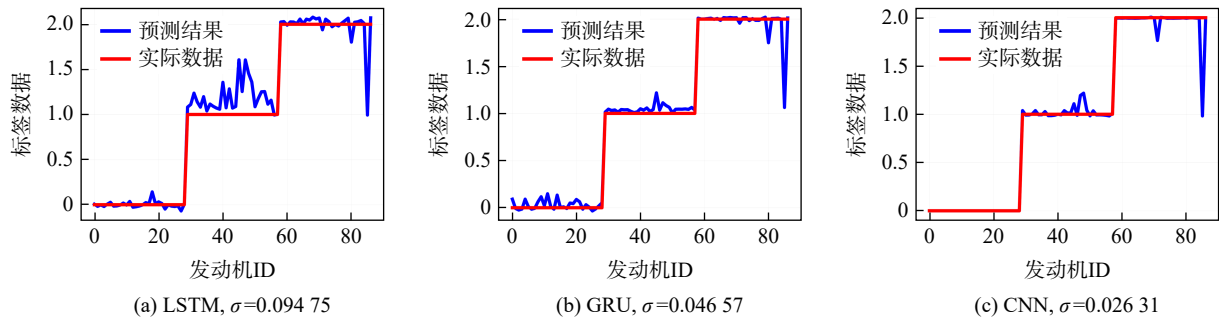


图 12 P30 的 LSTM、GRU 和 CNN 结果

Fig. 12 Results for P30 based on LSTM, GRU and CNN

受到连续 FDIA 后的数据相比仅多了最后 2 个数据点不同,这对网络的特征提取来说确实是很相近的特征,所以就有更大可能将第 81 号测试数据错分类为受到连续 FDIA 的数据,81 号测试数据与 52 号测试数据的由来都是原始未受数据攻击的第 23 号测试数据。

分类准确性差异更主要是来源于数据之间的特征相似性,进而导致了 3 个网络都对该数据出现错分类情况。

4.4 基于图像数据的 FDIA 识别

对于上述时序数据的识别过程可知,各网络已经能良好检测原始数据是否受到 FDIA,但存在对连续 FDIA 和间隔 FDIA 进行预测分类时常将间隔 FDIA 错分为连续 FDIA,对单对时序数据预测分类准确性不高问题。

首先,将从时序数据按 GAF 算法的两种方式转为图像数据,并将其分为测试集和训练集,对 3 个传感器(T24、T50 和 P30)的返回值分别操作,并为得到的各自传感器所得图像数据的测试集和训练集分别设置 one-hot 标签,未受到 FDIA 时为 [1,0,0],当受到连续 FDIA 时标签为 [0,1,0],当受到间隔 FDIA 时标签为 [0,0,1]。具体过程如下:以 T24 传感器返回值所生产的图像数据为例,将训练集和测试集图像合并后,前 144 张为训练用途,剩余图片作为测试数据,对应标签也是相同处理方式。

对生成的模型本文选取 144 张之后的任意测试数据集进行预测,整个过程与训练时相似,首先选取要预测的数据,然后分别对数据的图像数据和标签数据进行预处理,且过程与训练数据处理时一样,之后导入训练好的模型,得到 3 维的输出结果,分别为 3 种数据分类的可能比率,此时设置权值,当对应预测比率大于 0.5 时,确定该图像数据为对应预测的数据类型。同时利用 cv2.applyColorMap 生成测试图像的特征热力图像,并在图片左上方打印出该图像所对应的攻击类型。以发动机 ID 为 1 的传感器 T24 测试集返回值为例,对应两种 GAF 算法生成图像的 DenseNet 分类预测结果如图 13 所示。

4.5 各检测算法识别结果评价

对于第 4.2 节基于数据的 FDIA 识别结果的评定,本文基于两种数据度量依据,一是得出预测结果与实际标签值的 RMSE 值,二是将预测结果的分类正确率作为数据度量依据进行评价。

对基于时序数据的 FDIA 检测,CNN 网络在

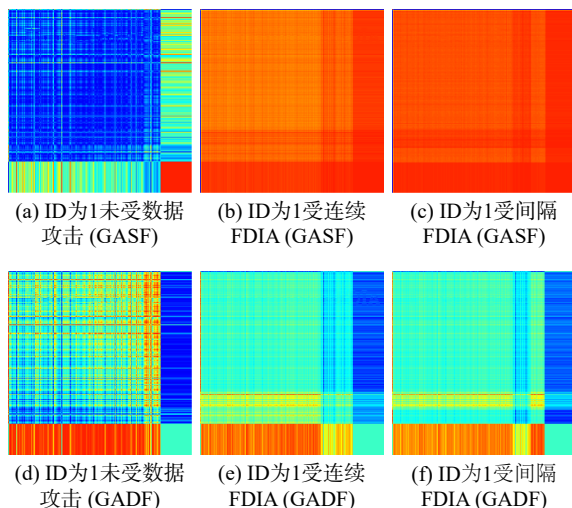


图 13 T24 传感器测试集分类预测图

Fig. 13 T24 sensor test set classification prediction chart

C-MAPSS 数据集上的表现好于 LSTM 和 GRU 网络预测结果,CNN 网络分类预测结果拥有所有 3 个网络预测结果中最高的正确率和最小的 RMSE 值。具体来说,T24 传感器 FDIA 检测时,CNN 网络的 RMSE 值分别为 LSTM 网络对应值的 76.81%,为 GRU 网络对应值的 49.89%;对于 T50 传感器,CNN 网络的 RMSE 值分别为 LSTM 网络对应值的 74.55%,为 GRU 网络对应值的 63.46%;对于 P30 传感器,CNN 网络的 RMSE 值分别为 LSTM 网络对应值的 27.77%,为 GRU 网络对应值的 56.50%。可知 CNN 网络的对时序数据的预测数值拟合性和准确性较高。

而对于文中第 4.3 节基于图像数据的 FDIA 识别结果的评定,一是得出预测结果与实际标签值的 BCEloss,二是将预测结果的分类正确率也作为数据度量依据进行评价。根据 3 个传感器(T24、T50 和 P30)数据以及两种 GAF 算法(GASF 和 GADF)共对应训练 6 个 DenseNet,并通过网络预测分类所得的正确率和 BCEloss 值作为使用 DenseNet 进行分类对各传感器以及不同 GAF 算法的适应性和有效性度量依据。具体预测结果见表 4。

对比时序数据 FDIA 检测与分类,由表 4 可计算,融合 GAF-DenseNet 方法在 T24、T50 和 P30 传感器上的平均分类准确率为 98.46%,与 LSTM、GRU 和 CNN 对比分别提高了 1.91%、3.82% 和 0.38%,且在区分两类 FDIA 类型上更为精确。也存在将间隔 FDIA 错分为连续 FDIA 的情况,但是仅为个别数据。且 DenseNet 对 T24 传

表 4 GAF-DenseNet 分类结果
Table 4 Results of GAF-DenseNet

传感器	GAF 方法	A:未受 FDIA	B:连续 FDIA	C:间隔 FDIA	正确率/%	BCEloss
T24	GASF	全对	全对	错分 2 个到 B	97.70	0.148
	GADF	全对	错分 1 个到 C	错分 1 个到 B	97.70	0.042
T50	GASF	全对	全对	错分 1 个到 B	98.85	0.008
	GADF	全对	全对	错分 1 个到 B	98.85	0.006
P30	GASF	全对	全对	错分 1 个到 B	98.85	0.024
	GADF	全对	错分 1 个到 C	全对	98.85	0.128

传感器的数据进行分类的效果明显好于另外 3 个网络的准确率,就传感器 T24 而言,运用 DenseNet 的分类准确率为 97.70%,而 LSTM 网络的分类准确率为 94.25%,GRU 网络的分类准确率为 88.51%,CNN 网络的分类准确率为 96.55%,可以对比得出 DenseNet 对于另外 3 个网络难以区分的 FDIA 识别效果更为良好,即对攻击的识别弹性也更强。

5 结 论

本文利用 C-MAPSS 数据集,在数据集中选取了 3 个飞机发动机传感器中返回值作为训练数据和测试数据来源,并使用了 4 种不同的机器学习方法对所选传感器返回值的原始数据,构建了连续 FDIA 和间隔 FDIA 模型。同时基于时序数据集采用 LSTM、GRU 和 CNN 网络进行攻击检测,研究表明 CNN 网络表现好于 LSTM 和 GRU 网络预测结果。运用所提的 GAF-DenseNet 方法将时序数据转化为图像数据后输入 DenseNet-121 分类,得出其分类正确率高于直接处理时序数据的 LSTM、GRU 和 CNN 网络的结果,体现 DenseNet-121 能够深度挖掘 GAF 变换后的传感器数据特征的优势。

而对于飞机工况剧烈变化,造成相关传感器发出警报数据变化,则需要飞行员与地面控制中心及时介入解决;对于飞机正常飞行工作过程中的相关传感器数值变动或其所受误差影响,可以未来通过机器学习方法,如本文中神经网络方法进行学习拟合来确定其未受到 FDIA,属于正常工况。

在 FDIA 构建中,不同的攻击起始时间、持续时间对网络结果的识别弹性是否有较大差异是下一步研究工作,另外,在受到 FDIA 后如何检测出 FDIA 所在的航空器运行周期,如何进行原始数据

恢复,以及更多样地考虑实际可能发生的攻击情况,并对其合理建模等问题也值得进一步研究和思考。

参考文献:

[1] 刘畅,陈雯柏.一种基于MSDCNN-LSTM的设备RUL预测方法[J].西北工业大学学报,2021,39(2):407-413.
LIU Chang, CHEN Wenbai. A RUL prediction method of equipments based on MSDCNN-LSTM[J]. Journal of Northwestern Polytechnical University, 2021, 39(2): 407-413. (in Chinese)

[2] 刘伟民,胡忠志.一种基于神经网络的航空发动机剩余寿命预估方法[J].航空发动机,2021,47(3):8-15.
LIU Weimin, HU Zhongzhi. An aeroengine remaining useful life prediction method based on neural network[J]. Aeroengine, 2021, 47(3): 8-15. (in Chinese)

[3] 车畅畅,王华伟,倪晓梅,等.基于1D-CNN和Bi-LSTM的航空发动机剩余寿命预测[J].机械工程学报,2021,57(14):304-312.
CHE Changchang, WANG Huawei, NI Xiaomei, et al. Residual life prediction of aeroengine based on 1D-CNN and Bi-LSTM[J]. Journal of Mechanical Engineering, 2021, 57(14): 304-312. (in Chinese)

[4] ZHAO Z, HUANG Y, ZHEN Z, et al. Data-driven false data-injection attack design and detection in cyber-physical systems[J]. IEEE Transactions on Cybernetics, 2021, 51(12): 6179-6187.

[5] 苗青林,张晓丰,高杨军,等.双通道深度卷积神经网络的航空发动机剩余使用寿命预测方法[J].空军工程大学学报(自然科学版),2022,23(2):12-18.
MIAO Qinglin, ZHANG Xiaofeng, GAO Yangjun, et al. A method of predicting aero-engines remaining useful life based on two-channel deep convolutional neural network[J]. Journal of Air Force Engineering University (Natural Science Edition), 2022, 23(2): 12-18. (in Chinese)

[6] 项德威,郑前钢,张海波,等.基于NN-PSM的航空发动机机载自适应稳态模型[J].航空动力学报,2022,37(2):409-423.
XIANG Dewei, ZHENG Qiangang, ZHANG Haibo, et al. Aero engine on board adaptive steady state model base on NN-PSM[J]. Journal of Aerospace Power, 2022, 37(2): 409-423. (in Chinese)

[7] 史国华.数据驱动的航空发动机RUL预测模型研究[D].辽宁大连:大连理工大学,2021.
SHI Guohua. Research on data-driven aero-engine RUL prediction model[D]. Dalian Liaoning: Dalian University of Technology, 2021. (in Chinese)

[8] 陈保家,郭凯敏,陈法法,等.基于残差NLSTM网络和注意

- 力机制的航空发动机剩余使用寿命预测[J]. 航空动力学报, 2023, 38(5): 1176-1184.
- CHEN Baojia, GUO Kaimin, CHEN Fafa, et al. Prediction of remaining useful life of aero-engine based on residual NLSTM neural network and attention mechanism[J]. Journal of Aerospace Power, 2023, 38(5): 1176-1184. (in Chinese)
- [9] 蔺瑞管, 王华伟, 车畅畅, 等. 基于LSTM分类器的航空发动机预测性维护模型[J]. 系统工程与电子技术, 2022, 44(3): 1052-1059.
- LIN Ruiguan, WANG Huawei, CHE Changchang, et al. Predictive maintenance model of aero engine based on LSTM classifier[J]. Systems Engineering and Electronics, 2022, 44(3): 1052-1059. (in Chinese)
- [10] MUSLEH A S, CHEN G, DONG Z Y. A survey on the detection algorithms for false data injection attacks in smart grids[J]. IEEE Transactions on Smart Grid, 2020, 11(3): 2218-2234.
- [11] LIU J, LABEAU F. Detection of false data injection attacks in industrial wireless sensor networks exploiting network numerical sparsity[J]. IEEE Transactions on Signal and Information Processing over Networks, 2021, 7: 676-688.
- [12] BRILAND M, BOUQUET F. An approach for testing false data injection attack on data dependent industrial devices[J]. Journal of Universal Computer Science, 2020, 27(7): 774-792.
- [13] MODE G R, CALYAM P, HOQUE K A. Impact of false data injection attacks on deep learning enabled predictive analytics[R]. Budapest, Hungary: 2020 IEEE/IFIP Network Operations and Management Symposium, 2020.
- [14] 王文益, 古亭亭. 基于1DCNN-BiLSTM网络的ADS-B欺骗式干扰检测[J]. 信号处理, 2021, 37(6): 984-990.
- WANG Wenyi, GU Tingting. ADS-B deceptive jamming detection based on 1DCNN-BiLSTM network[J]. Journal of Signal Processing, 2021, 37(6): 984-990. (in Chinese)
- [15] 李妮娜. 基于数据驱动的ADS-B异常检测算法研究[D]. 四川 广汉: 中国民用航空飞行学院, 2022.
- LI Nisi. The research on ADS-B anomaly detection algorithm based on data drive[D]. Guanghan Sichuan: Civil Aviation Flight University of China, 2022. (in Chinese)
- [16] CRETIN A, VERNOTTE A, CHEVROT A, et al. Test data generation for false data injection attack testing in air traffic surveillance[R]. Porto, Portugal: 2020 IEEE International Conference on Software Testing, Verification and Validation Workshops, 2020.
- [17] 张公. 基于TrellisNet的航空发动机剩余使用寿命研究[D]. 包头: 内蒙古科技大学, 2021.
- ZHANG Gong. Research on remaining service life of aeroengine based on TrellisNet[D]. Baotou: Inner Mongolia University of Science and Technology, 2021. (in Chinese)
- [18] YANG H, DING K, QIU R C, et al. Remaining useful life prediction based on normalizing flow embedded sequence-to-sequence learning[J]. IEEE Transactions on Reliability, 2021, 70(4): 1342-1354.
- [19] FREDERICK D K, DECASTRO J A, LITT J S. User's guide for the commercial modular aero-propulsion system simulation (C-MAPSS)[R]. NASA/TM-2007-215026, 2007.
- [20] 贾驰千, 冯冬芹. 基于多目标决策的工控系统设备安全评估方法研究[J]. 自动化学报, 2016, 42(5): 706-714.
- JIA Chiqian, FENG Dongqin. Industrial control system devices security assessment with multi-objective decision[J]. Acta Automatica Sinica, 2016, 42(5): 706-714. (in Chinese)
- [21] PAN S J, YANG Q. A survey on transfer learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10): 1345-1359.
- [22] 郑炜, 林瑞全, 王俊, 等. 基于GAF与卷积神经网络的电能质量扰动分类[J]. 电力系统保护与控制, 2021, 49(11): 97-104.
- ZHENG Wei, LIN Ruiquan, WANG Jun, et al. Power quality disturbance classification based on GAF and a convolutional neural network[J]. Power System Protection and Control, 2021, 49(11): 97-104. (in Chinese)
- [23] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[R]. Honolulu, US: 2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017.

(编辑: 秦理曼)