



基于Transformer重建的时序数据异常检测与关系提取

孟恒宇, 李元祥

(上海交通大学 航空航天学院, 上海 200240)

摘要: 现有时序异常检测方法存在计算效率低和可解释性差的问题。考虑到Transformer模型在自然语言处理任务中表现出并行效率高且能够跨距离提取关系的优势, 提出基于Transformer的掩膜时序建模方法。建立时序数据的并行无方向模型, 并使用掩膜机制重建当前时间步, 从而实现整段序列的重建。在存储系统数据集和NASA航天器数据集上的实验结果表明, 与基于长短期记忆网络模型的检测方法相比, 该方法可节约80.7%的计算时间, Range-based指标的F1得分达到0.582, 并且其通过可视化关系矩阵可准确反映人为指令与异常的关系。

关键词: 时序数据; 注意力机制; 异常检测; 关系提取; 自动编码器

开放科学(资源服务)标志码(OSID):



中文引用格式: 孟恒宇, 李元祥. 基于Transformer重建的时序数据异常检测与关系提取[J]. 计算机工程, 2021, 47(2): 69-76.

英文引用格式: MENG Hengyu, LI Yuanxiang. Anomaly detection and relation extraction for time series data based on Transformer reconstruction[J]. Computer Engineering, 2021, 47(2): 69-76.

Anomaly Detection and Relation Extraction for Time Series Data Based on Transformer Reconstruction

MENG Hengyu, LI Yuanxiang

(School of Aeronautics and Astronautics, Shanghai Jiao Tong University, Shanghai 200240, China)

[Abstract] Existing Anomaly Detection (AD) methods for time series are faced with inefficient computation and poor interpretability. As the Transformer model shows the high parallel efficiency and the ability to extract relations regardless of distance in Natural Language Processing (NLP) tasks, this paper proposes a Transformer-based method, Masked Time Series Modeling (MTSM). The parallel model with no direction of time series data is constructed, and the mask strategy is used for the reconstruction of the current timestep and then the whole sequence. Experimental results on the storage system dataset and NASA spacecraft dataset show that the proposed method saves about 80.7% time cost compared with the detection method based on Long-Short Term Memory (LSTM) model, and achieves 0.582 in F1 score for Range-based index. Moreover, it can visualize the relation matrix to reflect the relation between anomalies and human instructions accurately.

[Key words] time series data; attention mechanism; Anomaly Detection (AD); relation extraction; Auto-Encoder (AE)

DOI: 10.19678/j.issn.1000-3428.0057092

0 概述

异常检测是时序数据分析最成熟的应用之一, 定义为从正常的时间序列中识别不正常的事件或行为的过程^[1]。有效的异常检测被广泛用于现实世界的多个领域, 如量化交易、网络安全检测^[2]、汽车自动驾驶^[3]和大型工业设备的日常维护等。以在轨航天器为例, 由于航天器昂贵且系统复杂, 不能及时检测到危险可能

导致严重甚至无法弥补的损害, 异常随时可能发展为严重故障。因此, 准确、及时地检测出异常能够帮助航天工程师在第一时间给出应对措施。

基于重建的异常检测是目前时序数据异常检测的主流方法。文献[4]将深度学习用于异常检测任务中, 使用基于长短期记忆(Long-Short Term Memory, LSTM)网络的自动编码器(Auto-Encoder, AE)来检测时序数据异常, 该方法被成功应用于许多工业领域^[5-7]。基于

基金项目: 国家自然科学基金“海洋环境动力学和数值模拟”(U1406404)。

作者简介: 孟恒宇(1995—), 男, 硕士研究生, 主研方向为航空数据挖掘、图像处理; 李元祥(通信作者), 副教授、博士。

收稿日期: 2020-01-02 **修回日期:** 2020-02-13 **E-mail:** london@sjtu.edu.cn

重建的异常检测算法首先使用正常数据训练自动编码器来重建序列,然后通过相同的模型重建异常序列,获得重建误差序列,最后利用阈值或其他方法分割出异常发生的子序列。

然而,LSTM序列式传播的特点导致检测算法难以实现并行化,大幅增加了计算耗时,尤其是随着数据量指数级增长,过多的训练耗时会影响模型的及时更新^[8-9]。此外,LSTM等深度学习的“黑箱”特点使得提取到的特征很难有对应的物理含义,如在记录人为操作的故障数据集中,需要找出人为指令与异常的对应关系,此关系难以体现在LSTM模型中^[10]。

近年来,Transformer模型成为完成自然语言处理(Nature Language Processing, NLP)任务的主要方法。自然语言同样是一种序列数据,但其中语义存在跳跃性,这类似于时序数据中因人为因素而导致的状态变化,而Transformer模型可以很好地处理这一问题。因此,本文以注意力机制作为时序数据异常检测手段,提出一种基于Transformer模型的时序数据异常检测方法——掩膜时序建模(Mask Time Series Modeling, MTSM)。使用自注意力机制并行地重建数据以加快速度,通过可视化关系矩阵分析人为因素与异常发生的关系,并在存储数据集和NASA航天器数据集上进行实验,比较不同工况下的重建及异常检测效果。

1 相关工作

本节阐述以Transformer作为特征提取器的自动编码器工作原理及其优点,并介绍从重建误差序列中获取异常动态阈值分隔的过程。

1.1 嵌入注意力机制的自动编码器

注意力机制起源于人类眼球的注意力^[11-12],该机制能够帮助眼球从繁冗复杂的图像细节中获取最重要的部分。广义的注意力机制是指从大量信息中将权重分配给重要信息的过程。针对时序数据,不同时间步的数据其重要性应当是不同的。

假定输入序列为 $X=[x_1, x_2, x_3]$,输出序列为 $Y=[y_1, y_2, y_3]$,目标为根据输入序列 X 得出输出序列 Y 。假定有一组序列对 $\langle X, Y \rangle$,则自动编码器的任务是学习该序列对的映射关系。自动编码器由编码器Encoder和解码器Decoder组成,不同的任务使用不同的特征提取器^[13]。本文以NLP领域的Transformer模型作为自动编码器的Encoder和Decoder。

Encoder针对不同的输入 x_i 映射出不同的隐含编码 c_i ,将这种非线性映射定义为:

$$\begin{aligned} y_1 &= g(c_1) \\ y_2 &= g(y_1, c_2) \\ y_3 &= g(y_1, y_2, c_3) \\ &\vdots \\ y_i &= g(y_1, y_2, \dots, y_{i-1}, c_{i-1}) \end{aligned} \quad (1)$$

Decoder对 c_{i-1} 做符合要求的解码,得到输出序列 Y 。

每一个 i 时刻的 y_i 根据 c_{i-1} 和 i 时刻之前的 y_1, y_2, \dots, y_{i-1} 来生成,其对应的非线性映射函数定义为:

$$y_i = g(y_1, y_2, \dots, y_{i-1}, c_{i-1}) \quad (2)$$

由此,可以根据输入的 X 序列依次得到输出的 Y 序列。可以看出,不同于一般的自动编码器,不断更新的隐含编码 c_i 利用了注意力机制,在计算不同时间步时使用了不同的 c_i 权重向量,也就是分配了不同的注意力。在计算 c_i 时,第 t 个输入 x_t 对 y_i 的权重计算公式为:

$$c_{i,t} = \frac{y_i^T x_t}{\sqrt{n}} \quad (3)$$

其中, $1/\sqrt{n}$ 是加速收敛系数(n 是通道数)。

Transformer是完全基于自注意力机制的特征提取器,其两个关键结构如图1所示^[14]。图1(a)所示为注意力计算部分,其中,Query和Key经过矩阵相乘缩放到0和1之间,再通过Softmax得到注意力得分,之后计算Value的注意力得分加权,得到最终输出。图1(b)所示为Transformer编码器结构,其中,输入序列经过一层嵌入后成为一个多维序列,经过位置嵌入后的数据输入多层的转换层中,每一层转换层由多头注意力机制、层正则化和前向传播模块组成,同时加入了残差连接。

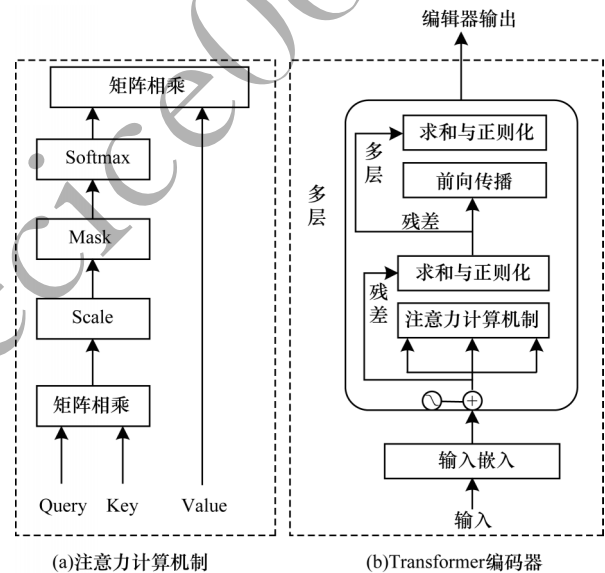


图1 注意力计算机制与Transformer编码器结构

Fig.1 Attention computing mechanism and Transformer encoder structure

与传统序列传播的LSTM从前到后依次输入序列不同,严格来说,Transformer没有方向性,其直接输入整段序列。此特性可使模型同时学习当前时间步周围的时间步信息。

1.2 基于重建误差动态阈值分割的异常检测

由重建误差序列可分割出异常子序列^[15],如图2所示。

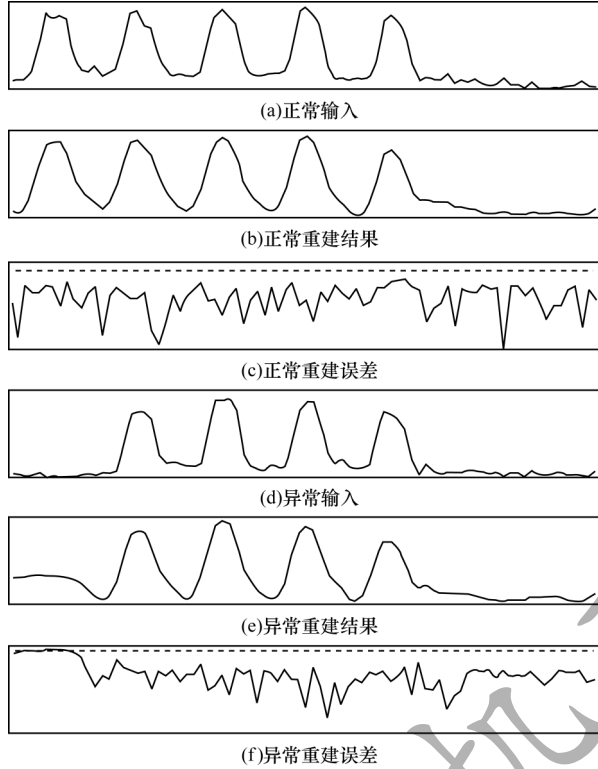


图2 基于重建的异常检测示例

Fig.2 Examples of reconstruction-based anomaly detection

文献[16]提出一种动态阈值分割方法,其设 y_i 为原始序列第 i 时间步的信号值, y'_i 为模型重建出的序列第 i 时间步的信号值,从而计算得到重建误差序列 $e_i = y_i - y'_i$ 。动态阈值分割指根据 e_i 求出动态阈值 ε ,其定义如下:

$$\varepsilon = \mu(e_i) + z\sigma(e_i) \quad (4)$$

其中, $\mu(\cdot)$ 是均值, $\sigma(\cdot)$ 是标准差, z 是权重系数。每一个时间步的误差阈值 ε_i 是动态的,取决于之前整个阈值序列 ε 的最大值,计算公式如下:

$$\varepsilon_i = \operatorname{argmax}(\varepsilon) = \frac{\frac{\Delta\mu(e_i)}{\mu(e_i)} + \frac{\Delta\sigma(e_i)}{\sigma(e_i)}}{|e_a| + |P_j|^2} \quad (5)$$

$$\Delta\sigma(e_i) = \sigma(e_i) - \sigma(\{e_i \in e_i | e_i < \varepsilon\}) \quad (6)$$

$$\Delta\mu(e_i) = \mu(e_i) - \mu(\{e_i \in e_i | e_i < \varepsilon\}) \quad (7)$$

在式(5)中, e_a 是上一个异常发生之后的所有正常序列, P_j 是所有的正常序列,表达式如下:

$$e_a = \{e_i \in e_i | e_i < \varepsilon\} \quad (8)$$

$$P_j = \{e_a\} \quad (9)$$

可以看出,动态阈值检测综合了均值和标准差,其根据误差累积的情况不断更新阈值,检测到异常后再更新阈值。在整个过程中,只需要调整一个参数即权重系数 z 。目前,该方法已经应用到NASA实际的生产环境中。

2 本文方法

为提高基于LSTM模型异常检测的计算效率,同时探索深度模型的特征在物理世界的含义,本文提出以Transformer作为自动编码器特征提取器的掩膜时序建模(MTSM)方法。

2.1 模型的输入与输出

在传统基于循环神经网络(Recurrent Neural Network, RNN)的模型中,对于多数重建的时间步,通常是从整段时间序列中取一段长度来重建下一步(LSTM信息传播的步数是有限的),重建序列即是将分步的预测拼接成一段完整的序列,如图3(a)所示,其中,灰色为模型输入部分,白色为重建部分。可以看出,这一过程是单向的,仅有重建时间步之前的数据影响重建这一时间步。而Transformer模型与基于RNN的模型不同,其一次性读取整段时序数据,在NLP任务中,可以从一个单词的左右文字学习这个单词的作用。因此,本文设计一种新的数据输入输出方式,如图3(b)所示。在获取一段完整序列并分割成小序列时,不是重建这一段小序列的最后一节数据,而是预测这一段小序列的中间一节数据,从而利用重建数据左右两边的信息来实现重建。

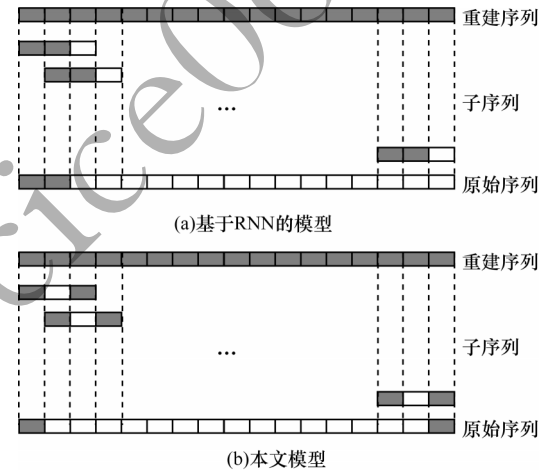


图3 基于RNN的模型与本文模型的输入输出示意图

Fig.3 Schematic diagrams of input and output of RNN-based model and the proposed model

假设被掩膜(Mask)掉的重建序列长度为 L_i (一般取1),重建序列之前的输入序列长度为 L_{i-} ,重建序列之后的输入序列为 L_{i+} ,则每次异常检测要延迟 L_{i+} 个时间步才能进行。当 $L_{i+}=0$ 时,MSTM的输入输出数据退化到一般的输入输出过程,如图3(b)所示。这就带来一个问题,即在线检测中,如果要将目标时间步之后的数据作为模型的输入,势必会影响检测的实时性。但如果类似NLP,加入之后数据可以提升检测效果的话,则可以在工程上做取舍,即需要检测精度高时增长目标时间步之后数据输入的

长度,而在需要及时性时减少甚至取消这部分输入。这一点将在下文实验部分加以讨论。

2.2 模型结构

受 Mask Language Modelling 启发,本文提出 MTSM 方法用于时序数据的重建。

如图4所示,对于输入序列 $X=[x_1, x_2, \dots, x_n]$,选择其中将要重建的时间步(如 x_4),用 [mask] 标志代替,从接下来的模型中屏蔽掉。经过一层位置嵌入后进入 Transformer 编码器,得到编码后的序列 encoded sequence,即上文提到的 memory(但与一般的 memory 不同,此时的 memory 不包含要重建的 x_4),再经过一个简单的 decoder 之后获取序列。损失函数定义为被屏蔽的 x_4 预测值与真实值之间的关系。换言之,MTSM 是从被 Mask 的时间步两边的数据或者说上下文数据重建时间步,这充分利用了 Transformer 无向性的特点。此外,综合利用上下文信息,也可以提高重建的精度。

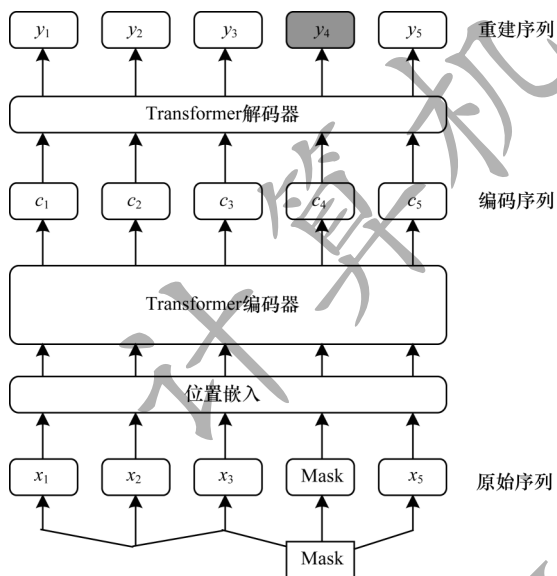


图4 MTSM方法建模过程

Fig.4 Modeling process of MTSM method

2.3 关系提取

Transformer 的一大优点是可以捕捉跨距离的依赖关系。针对 NLP 任务,研究者常利用此特性分析某个词在整句话中的作用,从而捕捉到较为直观的语义特征。本文利用这一特性进行人为指令与异常的关系提取,以探索异常检测中的人为因素影响。

首先对人为指令和异常的关系进行解释。以 NASA 航天器数据集为例,通过一键热编码(one-hot encoding)的形式记录地面操作人员的相关指令。如图5所示,事实上航天器仅传回了一个传感器遥测值,但与此同时 NASA 将地面操作人员的 24 种或 54 种指令的实行情况随信号的时间步编码在一起,这样不仅可以将其作为故障检测的数据集,同时也

可以用来分析人为指令在异常发生中的关系。假设信号在 102 时间步~104 时间步发生异常,计算 102 时间步与 100 时间步、101 时间步人工指令的相关系数,如果 102 时间步与 101 时间步的相关系数较大,则说明 101 时间步造成了 102 时间步的异常。

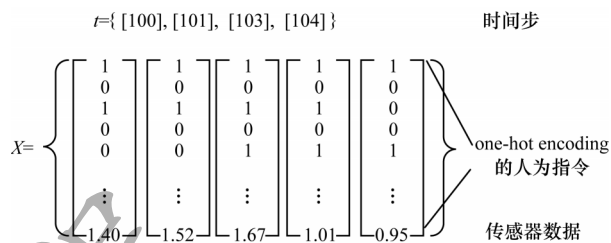


图5 NASA航天器数据集中异常与人为指令的关系

Fig.5 Relationship between anomaly and manual command in NASA spacecraft dataset

关系提取的具体实现过程如下:将 Query 的信号数据全部掩膜,只将人为指令传入模型,而输入保留全部数据,训练完毕后逐层可视化相关系数矩阵,找出异常发生时哪些时间步的人为指令权重较大,从而获得所提取到特征对应的物理含义。

3 实验结果与分析

本文在存储数据集和 NASA 航天器数据集上进行实验,比较不同工况下的重建及异常检测效果,并通过可视化关系矩阵进行所提取特征的物理含义分析。

3.1 不同工况下的时序数据重建

时序数据异常检测数据集的标注需要昂贵的专家成本,尤其是详细标注异常发生的开始和结束时刻的数据集。由于此类数据集规模通常较小,因此基于算法效果检验的实验结论存在一定偶然性。同时,基于重建误差的异常检测要求模型在输入正常数据时重建误差较小,输入异常数据时重建误差较大。因此,本文先使用含有不同工况的存储数据集^[17-18]进行重建效果对比。

本文使用的存储系统数据集来自于 Ostwestfalen-Lippe 应用科学大学的智慧工厂项目。这个高存储系统由 4 条传送带和 2 条导轨组成,每条传送带包含 3 个感应传感器,共有 18 个通道。数据集提供了 4 种工况,即无故障(正常)未优化、无故障(正常)已优化、有故障(异常)未优化和有故障(异常)已优化。基于此数据集,可以在工况是否异常和是否被优化两个维度上比较算法的优缺点。

图6为存储数据集通道3数据的重建结果对比。通过比较4种不同工况下的重建序列可以看出,优化后的数据比未优化的数据更平稳,杂波更少,异常数据比正常数据多了一段急剧上下波动的异常。

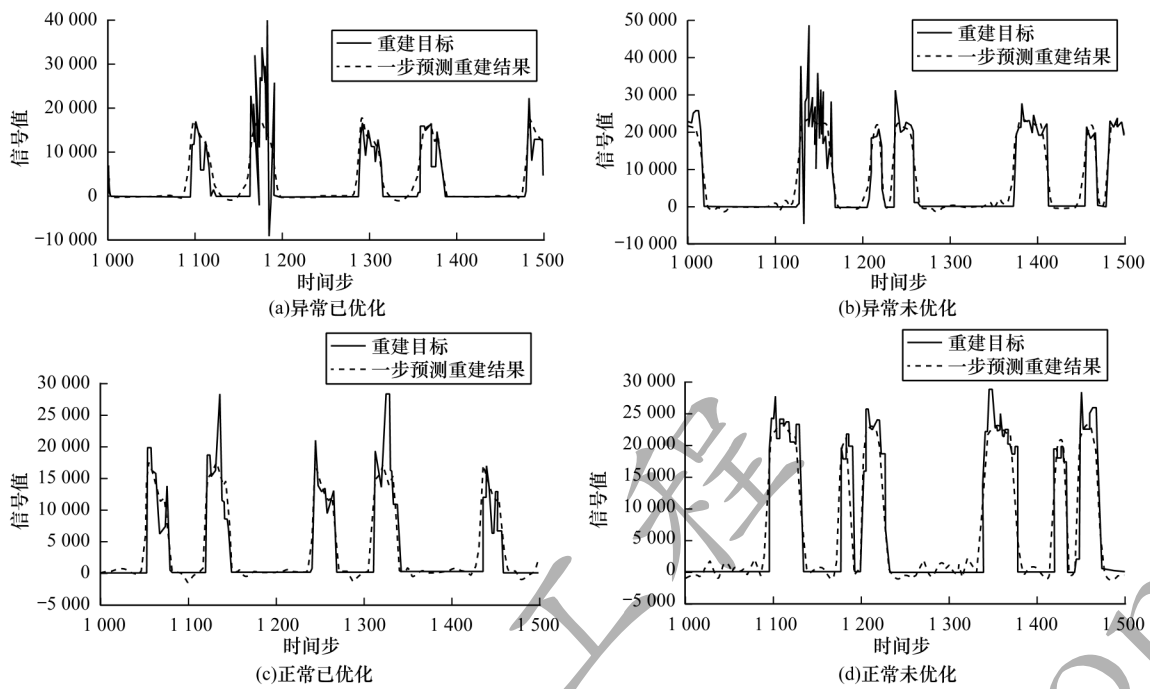


图 6 存储系统数据集上的重建结果

Fig.6 Reconstruction results comparison of storage system dataset

表 1 为 4 个子集上的重建结果绝对误差对比。可以看出,Transformer 方法在有异常的数据集上重建的绝对误差更大,这说明异常发生时 Transformer 的表现更鲁棒。

表 1 存储系统数据集上重建误差对比

Table 1 Reconstruction error comparison in storage system dataset

工况	LSTM 模型	Transformer 模型	
		有 Mask	无 Mask
异常未优化	0.252 9	0.261 2	0.258 5
异常已优化	0.324 6	0.331 8	0.330 2
正常未优化	0.318 6	0.309 6	0.320 1
正常已优化	0.268 5	0.234 7	0.256 9

表 2 是训练到收敛的平均用时对比,相关硬件为 NVIDIA GTX-1080TI、16 GB 内存、Intel B5 处理器。可以看出,由于 Transformer 放弃了 LSTM 的序列式推进,其注意力机制模块完全可以并行,因此用时普遍较少,最多可以节约 80.7% 的时间。

表 2 存储数据集上训练耗时对比

Table 2 Training time comparison in storage system dataset

工况	LSTM 模型	Transformer 模型	min
异常未优化	45.1	8.7	
正常未优化	37.2	10.1	
异常已优化	33.1	8.9	
异常已优化	33.3	9.9	

3.2 时序数据异常检测

3.2.1 NASA 航天器数据集

NASA 开源的专家标注的真实世界航天器故障

数据集^[16]包括火星科学实验室好奇号(Mars Science Laboratory rover, MSL)和土壤水分主动被动探测卫星(Soil Moisture Active Passive, SMAP),共计 82 个通道、105 个故障,数据集描述如表 3 所示。

表 3 NASA 航天器数据集描述

Table 3 Description of NASA spacecraft dataset

项目	SMAP	MSL	总计
序列总数	69	36	105
通道数	55	27	82
特有通道数	28	19	47
信号总长度	429 735	66 709	496 444
平均序列长度	约 6 000	约 2 000	—

3.2.2 Point-based 指标

对于序列数据的异常检测,一种常用的检测指标为 Point-based 指标,即当预测异常与真实值有交集时记为 True Positive,预测异常与任何真实值均无交集时记为 False Positive,真实值与任何预测值均无交集时记为 False Negative,其中,精确率(Precision)、召回率(Recall)与 F1 值的计算均与一般的检测任务相同。

如上文所述,本文模型的检测实时性主要取决于重建序列使用的输入序列长度,尤其是位于重建序列之后的输入序列长度 L_{i+} 。此外,Transformer 的层数和有头机制也会对检测效果造成影响。本节比较了不同参数配置对时间消耗的影响(硬件配置同上),如图 7 和表 4 所示。可以看出,本文模型的 Point-based 指标达到与 LSTM 相近的效果,并且可以大幅减少时间消耗,与 LSTM 模型相比节约了 84.4% 的时间。同时可以看出,多头机制在这个时序

数据任务中几乎没有发挥作用,这可能是因为数据维度过少(只有25维或55维),而一般NLP任务的词嵌入512维度的词向量空间。

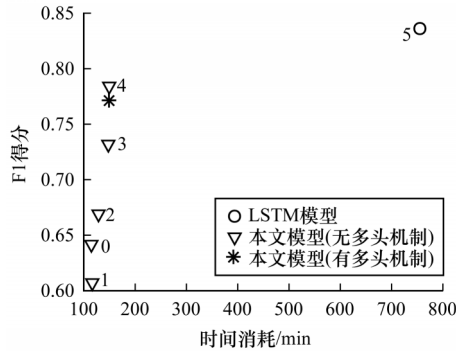


图7 时间消耗与F1得分对比

Fig.7 Comparison of time consumption vs F1 score

表4 图7实验的详细参数

Table 4 Detailed parameters of the experiment in Fig.7

实验序号	模型层数	L_{i-}	L_{i+}	L_i
0	6	100	100	1
1	6	200	0	1
2	8	100	100	1
3	12	100	100	5
4	12	100	100	1
LSTM	—	200	0	1
Multi-head	12	100	100	1

针对异常在线检测问题,由实验数据可以看出,设置 L_{i+} 即延迟在线检测时,异常检测的精度更高。而放弃利用两边信息即 $L_{i+}=0$ 时,虽然检测的精度有所下降,但是避免了在线检测的延迟,这也符合之前假设,即利用两边的信息可以加强建模,从而提升整体效果,并且输入数据长度越长精度越高。因此,针对在线检测问题,需要具体考虑不同场景的需要。

3.2.3 Range-based 指标

文献[19]指出,Point-based指标对时序数据并不准确。图8所示的3个例子均被记作True Positive。从左到右可以看出:第1个例子有一个真实故障,算法检测出1个故障;第2个例子有2个真实故障,检测算法却混淆为1个;第3个例子有1个真实故障,算法却识别为2个。

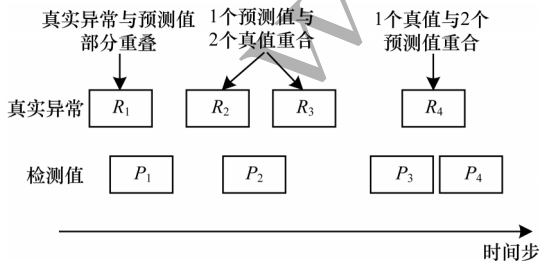


图8 Range-based 时序检测指标示意图

Fig.8 Schematic diagram of Range-based indicator of time series detection

虽然Point-based的精确率和召回率均为100%,但是物理意义不同。因此,文献[19]提出Range-based指标,相较于Point-based指标,其增加了OverlapReward值。第 i 个真实异常 R_i 与预测值集合 P 的OverlapReward计算如下:

$$\text{OverlapReward}(R_i, P) =$$

$$\sum_{j=1}^{N_p} (\omega(R_i, R_i \cap P_j, \delta)) \times \text{CardinalityFactor}(R_i, P) \quad (10)$$

$$\text{CardinalityFactor}(R_i, P) =$$

$$\begin{cases} 1, & \text{如果 } R_i \text{ 最多和一个 } P_j \text{ 相交} \\ \gamma(R_i, P), & \text{否则} \end{cases} \quad (11)$$

其中: $\omega(\cdot)$ 是重叠尺寸函数,用于衡量真实异常子序列与预测异常子序列的重叠大小; $\gamma(\cdot)$ 是重叠基数函数,用于衡量一个真实异常子序列与几个预测异常子序列对应,一般与真实异常重合的预测值越多,得分越低; $\delta(\cdot)$ 是位置偏置函数,如取前缘优先(Front-end Bias)时,预测异常的前缘与真实异常的前缘越接近则得分越高,取中心优先(Middle-end Bias)时,预测异常的中心与真实异常的中心越接近则得分越高。

表5比较了Range-based的不同位置偏置函数设置。如果设置为前缘优先,则Transformers模型优势更明显,即检测出异常的前缘相比真实值更接近。设置为中心优先,则Transformer模型也有一定优势。但如果完全不考虑异常的位置,则Transformer模型不占优势,这体现了Transformer模型在高层次语义上的优势。

表5 Range-based F1得分对比

Table 5 Comparison of Range-based F1 scores

不同位置偏置函数	LSTM 模型	Transformer 模型
Front-end Bias	0.547	0.582
Flat Bias	0.490	0.491
Middle-end Bias	0.524	0.529

3.3 异常人为因素分析

本文选用一个6层Transformer模型的关系矩阵进行可视化,部分实验结果如图9所示。已知这个序列在1 899时间步~2 099时间步、3 289时间步~3 529时间步和4 286时间步~4 594时间步处发生异常,可见注意力矩阵的可视化效果非常直观。对于Anomaly0,1 870时间步处所有的点的相关性都极强。因此,极大可能1 870时间步处发出的人为指令造成了之后的故障。对于Anomaly2,可以发现4 390时间步处点的相关性更大,这比真实异常开始处有延迟,但对应了从4 430时间步开始的预测异

常,可以看出模型将4 430时间步之后认为异常并认为4 390时间步处的指令依然关系极大。自上而下,中间层如Layer3提取到的特征就并不直观,可能是

相关的高层语义特征,而到模型的最后一层Layer5,特征又变得直观。以上逐层关系矩阵的变化反映了编码到解码的过程^[20]。

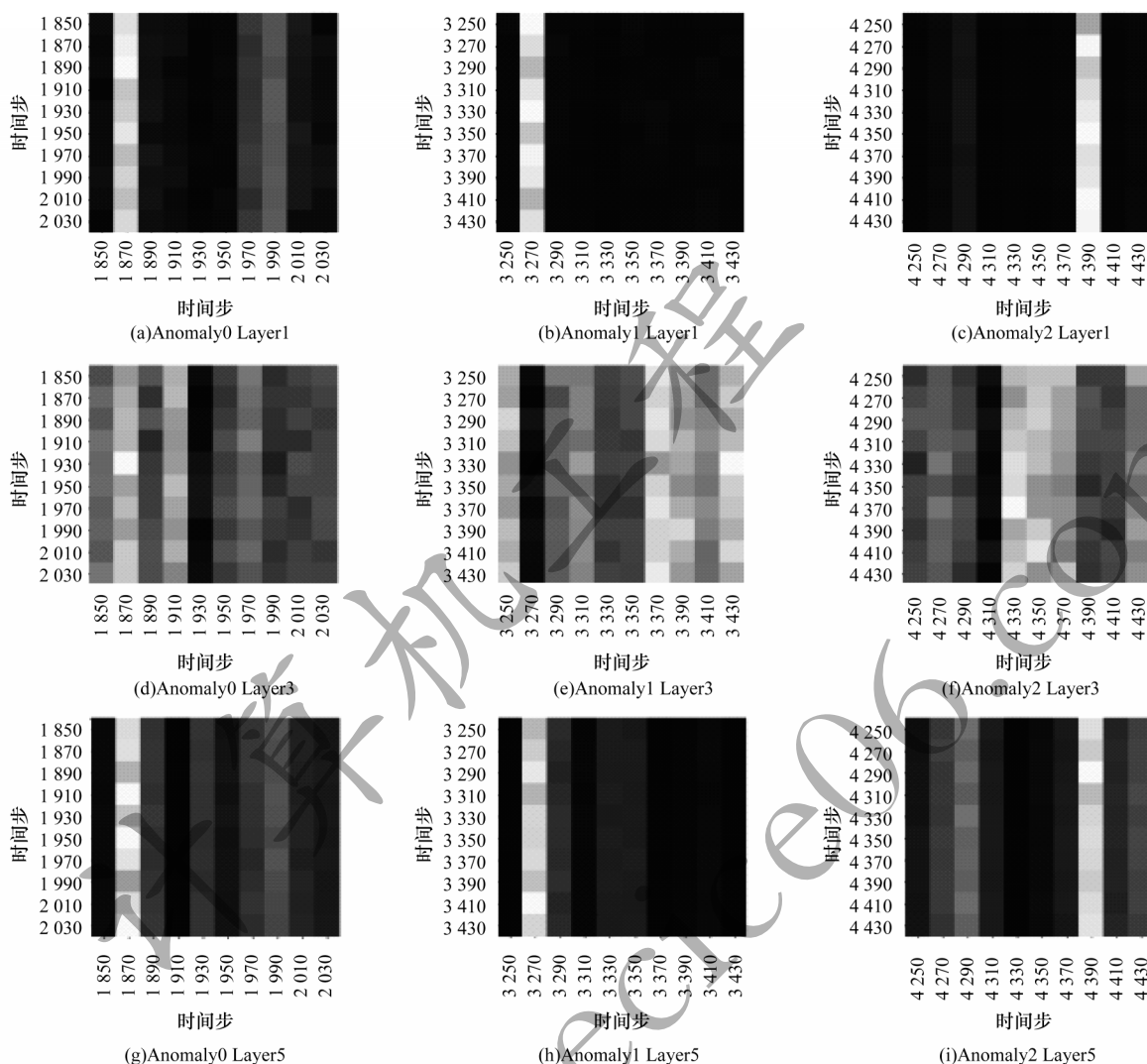


图9 关系矩阵部分可视化结果

Fig.9 Partial visualization results of relation matrix

4 结束语

目前,基于LSTM的异常检测已经被部署到很多实际应用场景中。本文针对现有异常检测方法计算效率低和提取到的特征意义不明显这两个问题,提出基于Transformer重建的MTSM方法。实验结果表明,与LSTM模型相比,Transformer模型减少了80.7%的训练耗时,并且对应了关系矩阵的物理含义,而Mask的使用则提高了重建和检测的精度,在重建任务中建模更鲁棒,Range-based检测指标的F1得分达到了0.582。下一步将继续提高重建精度并建立规模更大的数据集,以避免实验结论的偶然性。

参考文献

- [1] CHANDOLA V, MITHAL V, KUMAR V. Comparative evaluation of anomaly detection techniques for sequence data [C]//Proceedings of International Conference on Data Mining. Berlin, Germany: Springer, 2008: 743-748.
- [2] CHANDOLA V, BANERJEE A, KUMAR V. Anomaly detection: a survey[J]. ACM Computing Surveys, 2009, 41(3): 1-58.
- [3] ANAVA O, HAZAN E, ZEEVI A. Online time series prediction with missing data[C]//Proceedings of International Conference on Machine Learning. [S. l.]: PMLR, 2015: 2191-2199.
- [4] MALHOTRA P, RAMAKRISHNAN A, ANAND G, et al. LSTM-based encoder-decoder for multi-sensor anomaly detection[EB/OL]. (2016-06-01)[2020-01-02]. <https://arxiv.org/pdf/1607.00148.pdf>.

- [5] TAO Tao, ZHOU Xi, MA Bo, et al. Abnormal time series data detection of gas station by Seq2Seq model based on bidirectional long short-term memory [J]. Journal of Computer Applications, 2019, 39(3): 924-929. (in Chinese)
陶涛, 周喜, 马博, 等. 基于双向LSTM的Seq2Seq模型在加油站时序数据异常检测中的应用[J]. 计算机应用, 2019, 39(3): 924-929.
- [6] ZHANG Qi, HU Yupeng, JI Cun, et al. Edge computing application: real-time anomaly detection algorithm for sensing data [J]. Journal of Computer Research and Development, 2018, 55(3): 524-536. (in Chinese)
张琪, 胡宇鹏, 嵇存, 等. 边缘计算应用: 传感数据异常实时检测算法[J]. 计算机研究与发展, 2018, 55(3): 524-536.
- [7] GAO Jiawei, LIU Jianmin. An anomaly detection algorithm for time-series data flow oriented to trajectory information [J]. Computer Engineering, 2018, 44(5): 25-32, 46. (in Chinese)
高嘉伟, 刘建敏. 一种面向轨迹信息的时序数据流异常检测算法[J]. 计算机工程, 2018, 44(5): 25-32, 46.
- [8] PAHL M O, AUBET F X. All eyes on you: distributed multi-dimensional IoT microservice anomaly detection [C]// Proceedings of the 14th International Conference on Network and Service Management. Washington D. C., USA: IEEE Press, 2018: 72-80.
- [9] ALRAWASHDEH K, PURDY C. Toward an online anomaly intrusion detection system based on deep learning [C]// Proceedings of International Conference on Machine Learning and Applications. Washington D. C., USA: IEEE Press, 2016: 1-5.
- [10] YU Zhenbo, LIU Guangcan, LIU Qingshan, et al. Spatio-temporal convolutional features with nested LSTM for facial expression recognition [J]. Neurocomputing, 2018, 317: 50-57.
- [11] FENG Hui. Study on visual attention mechanism and its applications [D]. Beijing: North China Electric Power University, 2011. (in Chinese)
冯辉. 视觉注意力机制及其应用研究[D]. 北京: 华北电力大学, 2011.
- [12] DESIMONE R, DUNCAN J S. Neural mechanisms of selective visual attention [J]. Annual Review of Neuroscience, 1995, 18(1): 193-222.
- [13] SHI Weihang, LIN Nan. A joint classification learning algorithm for feature sequences of time-series data [J]. Computer Engineering, 2016, 42(6): 196-200, 207. (in Chinese)
史苇杭, 林楠. 一种联合的时序数据特征序列分类学习算法[J]. 计算机工程, 2016, 42(6): 196-200, 207.
- [14] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [EB/OL]. (2017-06-12) [2020-01-02]. <https://arxiv.org/pdf/1706.03762.pdf>.
- [15] JIANG Hua, ZHANG Hongfu, LUO Yidi, et al. Adaptive threshold network traffic anomaly detection based on KL distance [J]. Computer Engineering, 2019, 45(4): 108-113, 118. (in Chinese)
蒋华, 张红福, 罗一迪, 等. 基于KL距离的自适应阈值网络流量异常检测[J]. 计算机工程, 2019, 45(4): 108-113, 118.
- [16] HUNDMAN K, CONSTANTINOU V, LAPORTE C, et al. Detecting spacecraft anomalies using LSTMs and nonparametric dynamic thresholding [C]// Proceedings of the 24th ACM SIGKDD International Conference. New York, USA: ACM Press, 2018: 387-395.
- [17] BIRGELEN A, NIGGEMANN O. Enable learning of hybrid timed automata in absence of discrete events through self-organizing maps [M]// NIGGEMANN O, SCHÜLLER P. IMPROVE-innovative modelling approaches for production systems to raise validatable efficiency. Berlin, Germany: Springer, 2018: 37-54.
- [18] HRANISAVLJEVIC N, NIGGEMANN O, MAIER A. A novel anomaly detection algorithm for hybrid production systems based on deep learning and timed automata [C]// Proceedings of International Workshop on the Principles of Diagnosis. Denver, USA: [s. n.], 2016: 1-5.
- [19] TATBUL N, LEE T J, ZDONIK S, et al. Precision and recall for time series [C]// Proceedings of the 32nd International Conference on Neural Information Processing Systems. New York, USA: ACM Press, 2018: 1924-1934.
- [20] BAYDOGAN M G, RUNGER G C. Time series representation and similarity based on local autopatterns [J]. Data Mining and Knowledge Discovery, 2016, 30(2): 476-509.

编辑 金胡考