

能源数字对象压缩流转变时间尺度异常检测方法

杨 挺¹, 陈士威¹, 宋继勳², 沈子奇², 赵兴安², 郭 经²

(1. 天津大学电气自动化与信息工程学院, 天津 300072; 2. 国家电网有限公司大数据中心, 北京 100052)

摘 要: 针对能源数字对象在供用能主体间流转效率低、异常检测实时性和准确度不高的问题, 提出了一种基于压缩感知的数据流转和异常检测算法。首先, 基于压缩感知理论实现了能源数字对象数据压缩, 并采用基于动态基追踪和最优稀疏度估计的广义正交匹配追踪(GOMP-DBP-OSE)算法进行数据重构, 提高了数据在不同主体间的流转速度; 其次, 提出了多时间尺度下双向长短期记忆-Transformer(BiLSTM-Transformer)融合模型的数据异常滑动检测算法, 同时解析数据流的长相关性和短相似性, 提升了异常检测的实时性和准确度; 最后, 以某地区电动汽车-充电桩-电网互动的真实数据作为能源数字对象实验实例, 验证了本文所提方法的有效性。实验结果表明, 所提压缩算法的压缩率相较于对比算法平均提升 14.2%, 重构速度相比传统变步长稀疏度自估计子空间追踪(VSSESP)和压缩采样匹配追踪(CoSaMP)算法分别提升了 32.8%和 44.5%, 所提多时间尺度融合模型的异常检测准确率超过 95%, 相较于其他异常检测算法性能优越。

关键词: 能源数字对象; 压缩感知; 异常检测; 双向长短期记忆

中图分类号: TM732

文献标志码: A

文章编号: 0493-2137(2026)02-0155-09

Variable Time-Scale Anomaly Detection Method for Energy Digital Object Compression and Flow

Yang Ting¹, Chen Shiwei¹, Song Jimeng², Shen Ziqi², Zhao Xing'an², Guo Jing²

(1. School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China;

2. Big Data Center of State Grid Corporation of China, Beijing 100052, China)

Abstract: To improve the flow efficiency of energy digital objects between energy suppliers and users and enhance low real-time performance and accuracy of anomaly detection, a data flow and anomaly detection algorithm based on compressed sensing is proposed. First, the data compression of energy digital objects is realized according to the theory of compressed sensing, after which the data reconstruction is carried out using the generalized orthogonal matching pursuit based on dynamic basis pursuit and optimal sparsity estimation (GOMP-DBP-OSE) algorithm, thus improving the speed of data flow between different subjects. Second, a data anomaly sliding detection algorithm is proposed for the bidirectional long short-term memory Transformer (BiLSTM-Transformer) fusion model under multi-time scales, thereby resolving the long correlation and short similarity of the data flow at the same time. This solution improves the accuracy of real-time anomaly detection. Finally, the real data of electric vehicle-charging pile-grid interaction in a region are used as an experimental example of an energy digital object to verify the effectiveness of the method proposed in this paper. The experimental results revealed that the compression rate of the proposed algorithm is on average 14.2% higher than that of the comparison algorithms, whereas the reconstruction speeds increase by 32.8% and 44.5% compared with those of the traditional variable step sparsity estimation subspace pursuit (VSSESP) and compressive sampling matching pursuit (CoSaMP) algorithms, respectively. Moreover, the accuracy of anomaly detection of the proposed multi-time scale fusion model exceeds 95%, which is superior to the performance of

收稿日期: 2025-01-22; 修回日期: 2025-04-22.

作者简介: 杨 挺 (1979—), 男, 博士, 教授.

通信作者: 杨 挺, yangting@tju.edu.cn.

基金项目: 国家电网有限公司总部科技资助项目 (5700-202390591A-3-2-ZN).

Supported by the Science and Technology Project of State Grid Corporation of China (No. 5700-202390591A-3-2-ZN).

other anomaly detection algorithms.

Keywords: energy digital object(DO); compressed sensing; anomaly detection; bidirectional long short-term memory (BiLSTM)

随着能源互联网的快速发展,能源系统的数字化转型正在加速推进^[1]. 传统能源基础设施通过物联网设备和智能传感器不断产生海量实时数据,这些数据涵盖了能源产业等全生命周期的关键信息^[2-3]. 为了有效管理和利用数据资源,国家数据局以《数字中国建设整体布局规划》为指导,牵头制定数字对象(digital object, DO)这一新型的结构化数据封装形式^[4-5],其通过对唯一标识符、元数据和实际数据内容的系统整合,构建了物理实体与数字空间的精准映射机制,为能源资产管理和数据交互提供了标准化载体. 在能源领域,数字对象可以实现发电机组、输变电设备、用电终端等各类关键基础设施的数字映射^[6],有效支撑能源生产、传输、消费等全流程业务的高效流转^[7]. 例如,国家数据局已在数字对象的接入、标识、发现、交换、使用及管控等功能上取得突破,推动了中国工业互联网数据要素登记和国家科学数据发现与共享等多个领域的示范应用,推动了以数字对象为核心的数联网基础设施的建设和规模化应用.

然而,能源数字对象在实际应用场景中面临着严峻挑战. 首先,能源数据采集和存储的节点分布广泛,数据在不同节点间流转效率要求较高,而数字对象包含大量数据,导致节点间数据流转较慢;其次,在能源数字对象采集、封装和传输过程中,可能存在数据错误等异常^[8],需要准确率较高的异常检测算法,且能源互联网对数据实时性要求较高,传统异常检测方法难以满足系统的快速的响应需求^[9].

针对上述能源数字对象在流转中的挑战,研究人员开展了初步研究:在数据压缩方面,文献[10-12]分别基于前缀深度编码(prefix depth code, PDC)算法、基于霍夫曼编码的 LZSS 压缩算法(LZSS compression algorithm based on Huffman coding, LZSSCH)和基于 LZW 的混合压缩编码(hybrid compression programming based on LZW, ZHCP)算法,基于编码方式实现数据压缩,但数据压缩率较低;文献[13-14]应用压缩感知理论对数据进行压缩,并分别采用 CoSaMP 和 VSSESP 算法进行重构,有效减小了数据体量,降低了对存储和传输资源的需求,但在重构迭代过程中未能充分利用上一时刻的有效信息,重构效率较低;在异常检测方面,文献[15-17]分别提出了基于卷积神经网络(convolutional neural network, CNN)、长短期记忆(long short-term memory, LSTM)等深度学习的

异常数据诊断方法,但上述方法主要关注单一时间尺度的异常模式,忽略了多时间尺度上的复杂异常特征.

基于上述分析,本文提出了一种基于压缩感知的能源数字对象流转及异常检测算法:首先,基于压缩感知理论对数据进行有效压缩,实现数字对象高效流转,并采用 GOMP-DBP-OSE 算法对数据快速重构;其次,提出了基于 BiLSTM-Transformer 融合模型的异常滑动检测算法,实现多时间尺度下的数字对象异常检测,提升了异常检测的实时性和准确度;最后,以某地区电动汽车-充电桩-电网互动的真实数据作为本文实验数据集,实验结果表明,所提方法能够有效提升能源数字对象流转效率和异常检测的实时性与准确度.

1 能源数字对象建模

能源数字对象是在能源行业用于共享和流通的最小逻辑单元,指包含标识、元数据、数据实体 3 个部分的数据整体,其实际的数据内容在下文中统称为“数据”. 能源数字对象结构如图 1 所示.

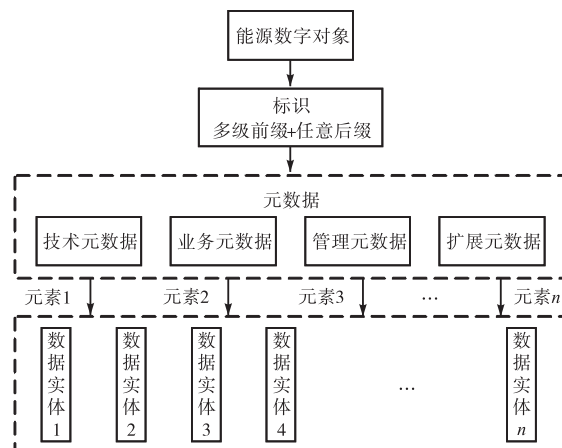


图 1 能源数字对象结构

Fig.1 Structure of energy digital object

由图 1 可知,能源数字对象包括以下部分.

(1) 标识:用于唯一识别每个能源数字对象,确保其不同系统和平台中不重复或混淆.标识通常采用唯一的编码或标记,不随存储方式或访问环境等变化^[6].能源数字对象标识为多级前缀/任意后缀的形式,前缀表示对象的不同层级属性,例如国家、地区、公司或业务领域等,后缀表示本地空间唯一命名.

(2) 元数据:包括该对象的描述性信息,可分为

多级元数据,如技术元数据、业务元数据、管理元数据、扩展元数据等,且元数据又分为许多元素。元数据帮助用户理解对象的内容和属性,为数据的组织和检索提供基础。

(3)数据实体:包含具体的能源数据内容,为实际流通数据,是数字对象中最核心的部分。数据实体分为结构化数据和多模态数据,其中结构化数据为数据表中的数据,多模态数据指文本、图像或视频等类型文件^[18-19]。以某地某充电桩为例,封装成的能源数字对象实例如图2所示。

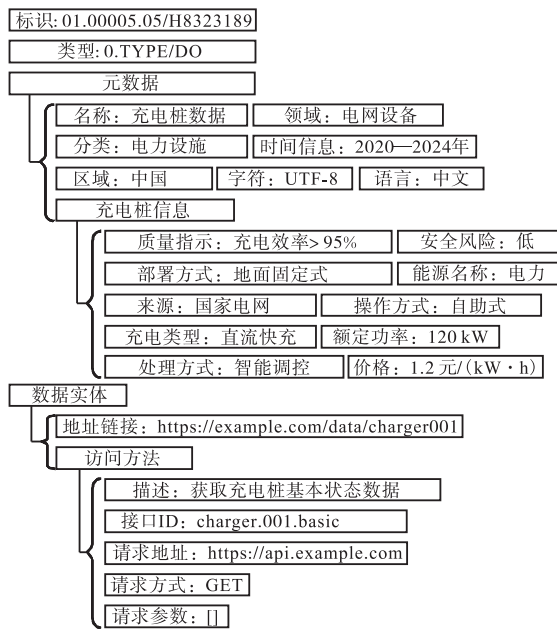


图2 某地充电桩能源数字对象

Fig.2 Energy digital object for a certain charging pile

2 基于压缩感知和多尺度模型融合的能源数字对象压缩流转同步异常检测

由于能源数字对象流转信息庞大,导致传输过程缓慢,目标节点不能及时获取信息,而且数据的异常检测任务时效性不高。针对上述问题,结合压缩感知算法和数据传输特点,提出基于压缩感知理论和多时间尺度模型融合的能源数字对象压缩流转同步异常检测算法。

首先利用压缩感知理论对数据进行压缩,然后在流转目标节点直接对压缩后的数据进行异常检测,避免了重构的等待时间,且减少了异常检测模型的输入特征,有效缩短了异常检测所需时间。本文采用数据流形式压缩传输数据^[20-21],在目标节点进行数据重构,并部署多时间尺度下的异常检测模型,实现了能源数字对象的快速流转同步异常检测。

2.1 能源数字对象压缩重构

2.1.1 数据压缩

对于文本数据,其在封装时以 GBK、UTF-8、UTF-16 等方式编码后以数字形式存储。本文将数据按照一定长度 N 截取后以数据流形式分段压缩传输,设某段能源数字对象数据为 $\mathbf{r}(n)$,在压缩感知框架中,通过 $M \times N$ 大小的测量矩阵 α 将其压缩为 M 维,即

$$\mathbf{z}(m) = \alpha \cdot \mathbf{r}(n) \quad (1)$$

式中 $\mathbf{z}(m)$ 为压缩后数据, $m = 1, 2, \dots, M$ 。

首先利用部分正交化的联合混沌系统生成测量矩阵,优点是具有良好的相关性控制能力和对初始值的高敏感性。初始联合混沌系统采用 Logistic 映射和 Tent 映射,初始化参数 a_1 和 b_1 ,按照规则生成数据序列为

$$\begin{cases} a_{g+1} = \mu a_g (1 - a_g) \\ b_{g+1} = \begin{cases} \mu b_g & b_g < 0.5 \\ \mu(1 - b_g) & b_g \geq 0.5 \end{cases} \end{cases} \quad (2)$$

式中 g 为 a 、 b 下标数量索引。

混沌序列元素可以为 $c = a \times b$,随后将 c 映射到 $[0, 1]$,并组成大小为 $M \times N$ 的矩阵。进而对此矩阵前 d 列进行正交化处理,得到部分正交化的联合混沌系统生成测量矩阵 α 为

$$\alpha = \begin{bmatrix} c'_{11} & c'_{12} & \dots & c'_{1l} & c_{1(l+1)} & \dots & c_{1N} \\ c'_{21} & c'_{22} & \dots & c'_{2l} & c_{2(l+1)} & \dots & c_{2N} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ c'_{M1} & c'_{M2} & \dots & c'_{Ml} & c_{M(l+1)} & \dots & c_{MN} \end{bmatrix} \quad (3)$$

然后通过式(1),利用测量矩阵对数据 $\mathbf{r}(n) \in \mathbf{R}^{N \times 1}$ 进行压缩,得到压缩数据 $\mathbf{z}(m)$ 。

2.1.2 数据重构

在第2.1.1节中,单段的 N 个数据点被压缩为 M 个,数据被截取分段传输至目标分布式节点,则在目标节点进行分段重构。采用傅里叶变换的稀疏基,即

$$\begin{bmatrix} \mathbf{S}(1) \\ \mathbf{S}(2) \\ \vdots \\ \mathbf{S}(N) \end{bmatrix} = \frac{1}{N} \begin{bmatrix} 1 & 1 & \dots & 1 \\ e^{j\frac{2\pi}{N} \times 1 \times 1} & e^{j\frac{2\pi}{N} \times 1 \times 2} & \dots & e^{j\frac{2\pi}{N} \times 1 \times N} \\ e^{j\frac{2\pi}{N} \times 2 \times 1} & e^{j\frac{2\pi}{N} \times 2 \times 2} & \dots & e^{j\frac{2\pi}{N} \times 2 \times N} \\ \vdots & \vdots & \vdots & \vdots \\ e^{j\frac{2\pi}{N} \times (N-1) \times 1} & e^{j\frac{2\pi}{N} \times (N-1) \times 2} & \dots & e^{j\frac{2\pi}{N} \times (N-1) \times N} \end{bmatrix} \cdot \begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(N) \end{bmatrix} \quad (4)$$

式中: $\mathbf{S}$ 为经傅里叶变换后的稀疏信号; 等式右侧的左乘矩阵的逆即为傅里叶变换下的稀疏基 $\boldsymbol{\varphi}$. 则感知矩阵 $\boldsymbol{\psi} = \boldsymbol{\alpha} \times \boldsymbol{\varphi}$.

本文采用基于动态基追踪和最优稀疏度估计的广义正交匹配追踪 (generalized orthogonal matching pursuit based on dynamic basis pursuit and optimal sparsity estimation, GOMP-DBP-OSE) 算法进行压缩数据的重构^[22]. 主要步骤如下.

步骤 1 输入压缩数据 $\mathbf{z}(m)$, 稀疏度 k , 感知矩阵 $\boldsymbol{\psi}$, 误差阈值 ε , 初始化稀疏信号 $\mathbf{S}_0 = 0$, 迭代次数 $i = 0$, 残差 $\boldsymbol{\theta}_0 = \mathbf{z}(m)$, 索引集 $D_0 = \emptyset$, 支撑集 $l_0 = \emptyset$.

步骤 2 $i = i + 1$, 计算残差 $\boldsymbol{\theta}_i$ 和感知矩阵 $\boldsymbol{\psi}$ 所有列 $\boldsymbol{\psi}_i$ 的内积, 挑选最大的 k 项列的索引放入集合 D_i , 即 $D_i = \text{indexmax}(k, |\langle \boldsymbol{\psi}_i, \boldsymbol{\theta}_{i-1} \rangle|)$.

步骤 3 更新支撑集 $l_i = l_{i-1} \cup D_i$.

步骤 4 求解稀疏信号 $\mathbf{S}_i = \arg \min \|\mathbf{z}(m) - \boldsymbol{\psi}_i \mathbf{S}_i\|_2$, $\boldsymbol{\psi}_i$ 为 $\boldsymbol{\psi}$ 中支撑集 l_i 对应的索引列组成的矩阵.

步骤 5 根据式 (5) 进行最优稀疏度估计, 即

$$k = \arg \min_k \left\| (k)_{l_i} \right\|_{\lambda} \quad \|\mathbf{z}(m) - \boldsymbol{\psi} \cdot \mathbf{S}_i\| \leq \varepsilon \quad (5)$$

式中采用可调 λ 范数控制稀疏度, 当 $\lambda < 1.0$ 时, 范数倾向于促进稀疏性, 使得更多的小系数趋于零; 当 $\lambda > 1.0$ 时, 范数更倾向于减小稀疏性. 通过调整适当 λ 值使模型更好地适应数据的稀疏结构, 提高模型的性能和泛化能力, λ 值基准为 1.0, 变化范围一般取 0.3 ~ 3.0.

步骤 6 更新残差 $\boldsymbol{\theta}_i = \mathbf{z}(m) - \boldsymbol{\psi}_i \mathbf{S}_i$.

步骤 7 若 $\|\boldsymbol{\theta}_i\|_2 \leq \varepsilon$, 则输出 $\mathbf{S}_i$, 转步骤 8, 否则转步骤 2.

步骤 8 重构出原始数据 $\mathbf{y}(n) = \boldsymbol{\varphi} \times \mathbf{S}_i$.

步骤 9 对下一时刻的压缩数据进行重构并执行步骤 1, 由于相邻时刻数据具有较强的时间相关性, 因此采用 DBP 动态基追踪, 即将上一分段计算的支撑集求解结果作为此段数据的初始支撑集, 从而减小冗余计算, 重复执行步骤 2 ~ 7, 实现数据重构, 以此类推进行分段数据流快速动态重构.

2.2 基于多时间尺度模型融合的能源数字对象异常滑动检测算法

基于第 2.1 节所述, 能源数字对象被压缩后向目标节点流转, 在实际流转中, 数据是以数据包形式发送. 一方面, 由于能源数字对象流转异常检测对实时性要求较高, 需要在较短时间内完成异常识别, 即当发送少量数据包时应完成检测; 另一方面, 数据异常在较短数据流内可能发生漏检情况, 且部分数据异常在短时间内难以体现, 需要增加数据包检测的长度进行二次识别.

因此, 针对上述问题, 本文提出基于多时间尺度滑动模型的能源数字对象异常检测算法, 直接对压缩后的流转数据进行异常检测. 针对不同时间尺度的数据包异常检测需求, 本文提出了基于变时间尺度的异常检测算法. 对于短时间尺度数据包, 采用 BiLSTM 网络模型进行检测, 其通过双向门控机制有效捕捉相邻数据点之间的局部时序依赖关系, 具有参数量少、计算效率高的优势; 而对于长时间尺度数据包, 则采用 Transformer 模型进行检测, 其自注意力机制能够建立跨时间步的全局关联, 通过多头注意力层提取长时间序列中的异常模式特征, 从而实现对累积数据包中异常的滑动检测. 多时间尺度模型融合的能源数字对象异常滑动检测框架如图 3 所示.

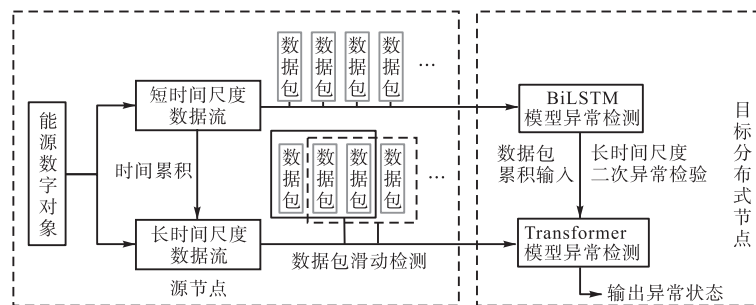


图 3 多时间尺度模型融合的能源数字对象异常滑动检测框架

Fig.3 Multi-time scale model fusion framework for anomaly sliding detection of energy digital objects

2.2.1 基于 BiLSTM 的短时间尺度异常检测模型

本文采用 BiLSTM 网络作为短时间尺度数据流异常检测的模型, LSTM 由于其神经元特点, 在处理文本等序列数据时具有很大优势. BiLSTM 是一种改

进的 LSTM 架构, 通过引入两个 LSTM 层, 分别以正向 (从前到后) 和反向 (从后到前) 处理序列数据, 从而更全面地捕捉上、下文信息.

设能源数字对象数据的输入为 $\mathbf{x} =$

$[x_1 \ x_2 \ \cdots \ x_t]$, 首先经 BiLSTM 双向传递后输出为 $h = [h_1 \ h_2 \ \cdots \ h_t]$, BiLSTM 具体结构如图 4 所示。

能源数字对象数据经 BiLSTM 结构单元输出后, 再经 ReLU 层、全连接层、概率计算层、分类层后输出数据是否异常, 具体结构如下。

ReLU 激活层为

$$H_1 = f_{\text{ReLU}}(h) \quad (6)$$

全连接层为

$$H_2 = f_{\text{fc}}(H_1) \quad (7)$$

概率计算层为

$$H_4 = f_{\text{sm}}(H_2) \quad (8)$$

分类层为

$$Y = f_{\text{cls}}(H_3) \quad (9)$$

式中分类层输出的 Y 为 0(数据正常)或 1(数据异常)。然后设置损失函数为

$$L = -\frac{1}{G} \sum_{p=1}^G \sum_{q=1}^C y_{pq} \lg y'_{pq} \quad (10)$$

式中: G 为训练批样本数; C 为类别数(两类); y_{pq} 为在第 p 个样本对应第 q 个类别的真实标签; y'_{pq} 为第 p 个样本的第 q 个类别的预测概率, 通过概率计算层 H_3 得到。最后采用 Adam 随机梯度下降算法进行模型训练。

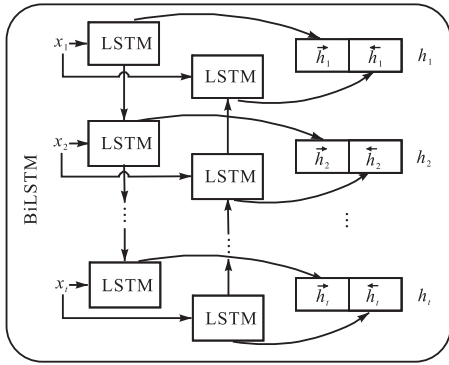


图 4 BiLSTM 结构

Fig.4 Structure of BiLSTM

2.2.2 基于 Transformer 的长时间尺度异常滑动检测模型

在进行长时间尺度数据异常检测时, 由于数据点个数的增加, 数据序列的长期相关性表征难度增大, 为此, 本文采用 Transformer 模型, 利用其注意力机制更好地挖掘长时间尺度下的数据异常特征, 模型结构包括数据嵌入、位置编码、多头注意力、前馈网络、残差连接和归一化等。

1) 数据嵌入

数据嵌入用于将原始数字对象压缩的离散数据转换为稠密的向量表示。由第 2.1.1 节中的短时间尺

度数据 x 知, Transformer 模型的输入为多个数据 x 的时间累积叠加, 即输入数据为 $X = (x_1, x_2, \dots, x_u)$, u 为累积倍数。设原始数据包含 v 个数据点, 每个点经嵌入后会将其映射到一个 d 维的连续空间, 即

$$E = f_{\text{emb}}(X) \quad (11)$$

式中 E 为原始数据 X 经嵌入后大小为 $v \times d$ 的矩阵。

2) 位置编码

由于 Transformer 无隐状态传递的顺序信息, 它需要显式地加入位置编码来传递序列的顺序信息。位置编码是一个和输入序列长度相同的矩阵, 每个位置的编码是基于该位置的数字索引通过使用正弦和余弦函数来表示, 即

$$\begin{cases} P_{(p_{\text{os}}, 2f)} = \sin\left(\frac{P_{\text{os}}}{10000^{2f/d}}\right) \\ P_{(p_{\text{os}}, 2f+1)} = \cos\left(\frac{P_{\text{os}}}{10000^{2f/d}}\right) \end{cases} \quad (12)$$

式中: p_{os} 为位置索引; f 为维度索引; d 为嵌入空间的维度。

位置编码会和输入的嵌入向量相加, 从而得到包含位置信息的输入为

$$E_{\text{pos}} = E + f_{\text{pos}}(E) \quad (13)$$

3) 多头注意力机制

注意力机制用于计算输入序列中每个元素与其他元素的关系(即注意力)。在 Transformer 中, 采用了多头注意力机制, 即将输入分为多个头, 每个头独立计算注意力并最终合并结果。对于上层位置编码输出矩阵 E_{pos} , 首先将其线性变换为 3 个矩阵: 查询矩阵 $Q = E_{\text{pos}} W^Q$, 键矩阵 $K = E_{\text{pos}} W^K$, 值矩阵 $V = E_{\text{pos}} W^V$, W 为需要学习的权重矩阵。然后计算每个头的注意力分数, 通过计算查询和键的点积, 再通过 Softmax 函数获得权重为

$$W = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (14)$$

式中 d_k 为键的维度。

最后将各个头的结果拼接, 并通过一个线性变换得到最终的多头注意力输出矩阵 A 。

4) 前馈网络

前馈网络一般由两个线性层和一个激活函数组成, 其计算过程为

$$F = \max\{0, AW_1 + B_1\}W_2 + B_2 \quad (15)$$

式中: W_1 、 W_2 为权重矩阵; B_1 、 B_2 为偏置矩阵。

5) 残差连接归一化

将模型多头注意力输出和前馈网络输出使用残

差连接并经过归一化处理,有效缓解梯度消失问题.

$$\mathbf{O} = f_{\text{norm}}(\mathbf{F} + \mathbf{A}) \quad (16)$$

6) 分类输出

最后使用和短时间尺度 BiLSTM 模型中相同的 softmax 分类层和损失函数,根据第 2.2.1 节中式 (7) ~ (10),对残差归一化输出 $\mathbf{O}$ 进行相同操作,采用 Adam 随机梯度下降算法进行模型训练,则最终长时间尺度下的 Transformer 模型如图 5 所示.

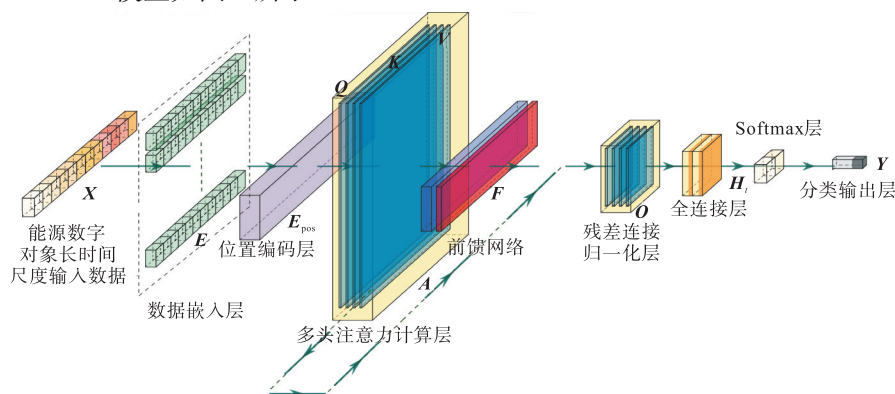


图 5 Transformer 模型

Fig.5 Transformer model

3 实验分析

以某地区电动汽车-充电桩-电网互动真实数据作为本文能源数字对象实验实例,实际交互数据的数字对象建模如图 2 所示.

3.1 压缩重构效果对比

分别使用 PDC 算法、LZSSCH 算法和 ZHCP 算法和本文压缩感知算法对 4 个文本数据集进行压缩,压缩率对比如表 1 所示.

表 1 不同算法压缩率对比

Tab.1 Comparison of compression ratio among different algorithms

文本数据集	原始数据集大小/Byte	压缩率			
		PDC 算法	LZSSCH 算法	ZHCP 算法	本文算法
1	826 512	1.681	1.272	1.637	1.834
2	301 355	1.598	1.367	1.621	1.873
3	64 654	1.937	1.325	1.572	2.035
4	544 981	1.475	1.299	1.524	1.934

注:压缩率=原始数据集大小/压缩后数据集大小.

由表 1 可知,本文所提压缩算法在 4 个文本数据集上与其余 3 种效果最好的压缩算法相比,压缩率分别提升了 9.1%、15.5%、5.1%、26.9%.

另外,为测试本文 GOMP-DBP-OSE 算法的重构速度,分别使用压缩感知理论中常用的 CoSaMP 算法、VSSSP 算法和本文所提算法对文本一部分压缩

综上,本文设计了基于多时间尺度模型融合的滑动检测方案,使用 BiLSTM 模型对短时间传输的数据包进行异常检测,减少了数据完整传输的等待时间,同时采用 Transformer 模型对长时间传输的数据包进行滑动检测,异常检测频率和短时间尺度模型相同,检测效率较高,且多时间尺度下的双重检测模式进一步提高了检测的准确性和可靠性.

数据进行重构,其重构所需时间随数据压缩体量变化情况如图 6 所示.

由图 6 可以看出,在相同数据体量下,本文所提算法的重构效率更高,平均重构速度相比 VSSSP 和 CoSaMP 分别提升了 32.8% 和 44.5%. 这是因为本文采用所提出的 DBP 算法,以上一时刻解的支撑集作为当前时刻数据恢复的先验信息,在连续重构过程中充分利用数据的时间相关性,减少了冗余计算,加速了重构.

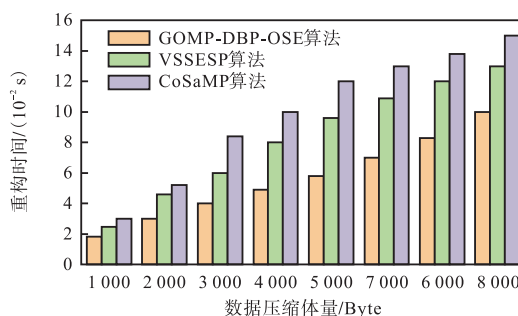


图 6 重构时间随数据压缩体量变化情况

Fig.6 Variation of reconstruction time with data compression volume

3.2 文本数据异常检测

对文本数据集 1 数据压缩后数据进行随机截取,设置短时间尺度文本压缩流转长度为 150 个数据点,长时间尺度样本长度为短时间尺度样本长度的 3 倍,

然后人为设置一定程度的异常样本,按照 7:3 比例进行训练集和测试集划分,样本构造结果如表 2 所示。

表 2 样本数据

Tab.2 Sample data

数据类型	样本长度	训练集		测试集	
		正常	异常	正常	异常
技术数据	短时尺度	8 168	1 920	3 500	824
业务数据	长时尺度	8 166	3 410	3 500	1 462
管理数据	短时尺度	4 534	1 024	1 942	440
扩展数据	长时尺度	4 532	1 786	1 942	766
数据	短时尺度	6 610	1 206	2 832	516
数据	长时尺度	6 608	1 780	2 832	762
数据	短时尺度	2 280	500	976	214
数据	长时尺度	2 278	802	976	344

对样本设置标签,数据正常设为 0,数据异常设为 1,利用训练集对短时间尺度 BiLSTM 模型和长时间尺度 Transformer 模型进行训练,具体训练参数分别如表 3 和表 4 所示。

表 3 BiLSTM 模型主要训练参数

Tab.3 Main training parameters of BiLSTM model

层数	批量大小	隐含神经元数量	最大训练轮次	学习率	优化器
1	300	50	4 000	0.010 0 ~ 0.000 1	Adam

表 4 Transformer 模型主要训练参数

Tab.4 Main training parameters of Transformer model

注意力层数	注意力头数	批量大小	最大训练轮次	学习率	优化器
2	3	300	4 000	0.050 0 ~ 0.001 0	Adam

表 3 和表 4 中,学习率设置为区间变化模式,即训练前期学习率较大,使得模型更快收敛;训练后期学习率逐渐下降,保证了训练的精度。BiLSTM 和 Transformer 模型的训练损失迭代情况如图 7 所示,可以看出当迭代轮数到达 3 500 轮左右时训练损失基本不再下降。

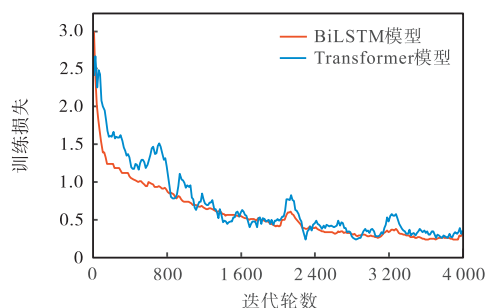


图 7 训练损失迭代情况

Fig.7 Iteration status of training loss

为检验模型准确度,对测试样本进行异常识别。设数据异常为阳性,数据正常为阴性,采用真阳

率(R_{TP})、假阳率(R_{FP})、真阴率(R_{TN})、假阴率(R_{FN})和准确率(A_{CC})5 种性能指标对模型效果进行评价,具体计算公式如下。

$$\begin{cases} R_{TP} = T_p / (T_p + F_n) \\ R_{FP} = F_p / (F_p + T_n) \\ R_{TN} = T_n / (T_n + F_p) \\ R_{FN} = F_n / (F_n + T_p) \\ A_{CC} = (T_p + T_n) / (T_p + T_n + F_p + F_n) \end{cases} \quad (17)$$

式中: T_p 、 F_p 分别代表实际阳性、阴性样本中被预测为阳性的样本数量; F_n 、 T_n 分别代表实际阳性、阴性样本中被预测为阴性的样本数量。

图 8 中分别给出了两种模型测试集真阳/阴率、假阳/阴率和准确率。由图 8 可以看出真阳率和真阴率均达到了 90% 以上的精度,且整体准确率为 95% 以上,说明本文所提模型异常检测效果良好。这是因为本文选取 BiLSTM 作为短时间尺度异常检测模型,充分利用了数据的双向相关性,而在长时间尺度上使用 Transformer 模型,其多注意力机制可以更好地挖掘较长时间序列数据的有效特征。

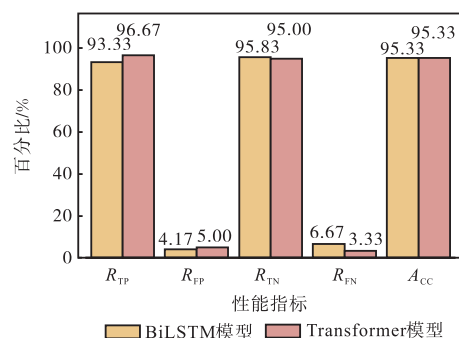


图 8 测试集结果

Fig.8 Results of test set

同时,为验证分布式节点流转异常检测效果,将能源数字对象长序列流数据作为模型输入,用于模拟节点流转传输情况。表 5 分别给出了仅用 CNN、BP、SVM、ELM、LSTM、BiLSTM、Transformer 模型进行短时间尺度、长时间尺度检测和本文所提 BiLSTM-Transformer 多时间尺度异常检验效果对比结果。

从表 5 可以看出,长时间尺度和短时间尺度下 LSTM 和 Transformer 模型的数据异常漏检率均低于其他算法,这是由于 BiLSTM 模型的双向记忆单元可以有效记录和传递时间序列数据的信息,而 Transformer 模型的多注意力机制使其在处理较长时间序列文本时可以更好提取数据之间的相关特征。具体分析,仅用 BiLSTM 模型进行短时间尺度检测漏检率为 11.72%,仅使用 Transformer 模型进行长

时间尺度检测漏检率为 10.17%, 而本文所提多时间尺度滑动检测漏检率仅为 7.31%, 分别下降了 37.63% 和 28.12%, 这是由于部分数据异常体现在较长的时间尺度内, 仅使用单一时间尺度进行异常检测易发生漏检情况, 本文使用变时间尺度的双重检测机制有效减少了异常数据的漏检情况, 提升了能源数字对象流转过程中异常检测的准确度。

表 5 长序列流数据异常检测效果对比

Tab.5 Comparison of anomaly detection performance for long sequence stream data

检测模型	时间尺度	漏检率/%
CNN	短时间尺度	18.65
	长时间尺度	14.98
BP	短时间尺度	30.45
	长时间尺度	32.51
SVM	短时间尺度	28.63
	长时间尺度	33.71
ELM	短时间尺度	28.32
	长时间尺度	22.68
LSTM	短时间尺度	12.53
	长时间尺度	15.47
BiLSTM	短时间尺度	11.72
	长时间尺度	13.72
Transformer	短时间尺度	16.43
	长时间尺度	10.17
BiLSTM-Transformer	多时间尺度	7.31

4 结 语

本文提出了一种基于压缩感知的能源数字对象压缩流转同步异常检测方法, 主要成果如下: 基于压缩感知理论实现了能源数字对象数据压缩, 提高了数据在各节点间的流转速度, 并使用 GOMP-DBP-OSE 算法实现数据高效重构, 提出了结合 BiLSTM 和 Transformer 模型的多时间尺度异常滑动检测方案, 可以捕捉不同时间尺度的异常情况, 整体检测准确率超过 95.00%。通过某充电桩数据的大量实验验证了所提压缩算法和异常检测算法的有效性。未来将针对能源数字对象中的多模态数据(如图像等), 构建更加通用的能源数字对象多模态数据处理体系。

参考文献:

- [1] 余贻鑫. 面向 21 世纪的智能电网[J]. 天津大学学报(自然科学与工程技术版), 2020, 53(6): 551-556.
Yu Yixin. A brief description of the basics of the smart grid[J]. Journal of Tianjin University (Science and Technology), 2020, 53(6): 551-556 (in Chinese).
- [2] 苏万华, 邱俊松, 邬斌扬, 等. 基于运行特性和生命周期理论的综合能源系统优化方法[J]. 天津大学学报(自然科学与工程技术版), 2022, 55(4): 371-382.
Su Wanhua, Qiu Junsong, Wu Binyang, et al. Integrated energy system optimization method based on the equipment operating characteristics and life cycle theory[J]. Journal of Tianjin University (Science and Technology), 2022, 55(4): 371-382 (in Chinese).
- [3] 杨 挺, 刘亚闯, 刘宇哲, 等. 信息物理系统技术现状分析与趋势综述[J]. 电子与信息学报, 2021, 43(12): 3393-3406.
Yang Ting, Liu Yachuang, Liu Yuzhe, et al. Review on cyber-physical system: Technology analysis and trends[J]. Journal of Electronics & Information Technology, 2021, 43(12): 3393-3406 (in Chinese).
- [4] Sharp C. Overview of the digital object architecture (DOA) [EB/OL]. <https://www.internetsociety.org/resources/doc/2016/overview-of-the-digital-object-architecture-doa/>, 2016-11-26.
- [5] Sayao LF, Sales LF, Segundo W. Fair digital object: Expanding the boundaries of data stewardship[J]. Informacao & Sociedade-Estudos, 2023, 33: 1-27.
- [6] Reilly S, Tupelo-Schneck R. Digital object repository server: A component of the digital object architecture [J/OL]. D-Lib Magazine, <https://www.dlib.org/dlib/january10/reilly/01reilly.html>, 2010-01-02.
- [7] Hisseine M A, Chen D J, Xiao Y, et al. A review of digital object architecture and handle system: Development, current applications and prospective[J]. Internet of Things, 2024, 26: 101230.
- [8] 赵少东, 王程斯. 基于多级索引集群均衡的电力能源大数据采集方法研究[J]. 电网与清洁能源, 2024, 40(8): 85-90.
Zhao Shaocong, Wang Chengsi. Research on power and energy big data acquisition methods based on multi-level index cluster equilibrium[J]. Power System and Clean Energy, 2024, 40(8): 85-90 (in Chinese).
- [9] 王 伟, 周少泽, 黄 萌, 等. 构网型技术: 演进历程、功能定位与应用展望[J]. 电力系统自动化, 2025, 49(1): 1-13.
Wang Wei, Zhou Shaoze, Huang Meng, et al. Grid-forming technologies: Evolution history, function, and application prospects[J]. Automation of Electric Power Systems, 2025, 49(1): 1-13 (in Chinese).
- [10] 曾党泉. 基于 PDC 编码的中文文本压缩算法[J]. 计算机工程与应用, 2015, 51(17): 205-209, 227.
Zeng Dangquan. Chinese text compression algorithm

- based on PDC coding[J]. Computer Engineering and Applications, 2015, 51(17): 205-209, 227(in Chinese).
- [11] 华强. 中文文本压缩的 LZSSCH 算法[J]. 中文信息学报, 1998, 12(1): 50-56.
Hua Qiang. The LZSSCH data compression algorithm for Chinese text files[J]. Journal of Chinese Information Processing, 1998, 12(1): 50-56(in Chinese).
- [12] 朱巧明, 赵英英, 钱培德. 基于中文词编码的压缩算法 ZHCP 的实现[J]. 小型微型计算机系统, 2003, 24(2): 306-308.
Zhu Qiaoming, Zhao Yingying, Qian Peide. To implement a compress algorithm ZHCP based upon Chinese word coding[J]. Mini-Micro Systems, 2003, 24(2): 306-308(in Chinese).
- [13] 杨挺, 尚昆, 袁博, 等. 基于压缩感知的盲源信号分离检测方法[J]. 天津大学学报(自然科学与工程技术版), 2016, 49(11): 1138-1143.
Yang Ting, Shang Kun, Yuan Bo, et al. Blind signal separation detection method based on compressed sensing [J]. Journal of Tianjin University (Science and Technology), 2016, 49(11): 1138-1143(in Chinese).
- [14] 田金鹏, 刘小娟, 郑国莘. 一种变步长稀疏度自适应子空间追踪算法[J]. 自动化学报, 2016, 42(10): 1512-1519.
Tian Jinpeng, Liu Xiaojuan, Zheng Guoshen. A variable step size sparsity adaptive subspace pursuit algorithm[J]. Acta Automatica Sinica, 2016, 42(10): 1512-1519(in Chinese).
- [15] Yang Y Y, Zhu W. Research based on improved CNN-SVM fault diagnosis of V2G charging pile[J]. Electronics, 2023, 12(3): 655.
- [16] 邓丰, 史鸿飞, 冯思旭, 等. CNN-LSTM 全景故障特征挖掘的配电网单端定位方法[J]. 中国电机工程学报, 2023, 43(增1): 114-126.
Deng Feng, Shi Hongfei, Feng Sixu, et al. Single-ended traveling wave location method for distribution network based on CNN-LSTM panoramic fault feature mining[J]. Proceedings of the CSEE, 2023, 43(Suppl 1): 114-126(in Chinese).
- [17] 金萍, 侯娟. 面向新型电力系统的粗糙集和双流网络自动化物联设备故障诊断方法研究[J]. 电测与仪表, 2024, 61(9): 166-171.
Jin Ping, Hou Juan. Research on fault diagnosis method of rough set and dual-stream network automation IoT device for novel power system[J]. Electrical Measurement and Instrumentation, 2024, 61(9): 166-171(in Chinese).
- [18] 王卓瑜, 王磊, 陆婷, 等. 基于逆向运算法的海量大数据安全存储方法[J]. 兵工自动化, 2024, 43(7): 23-26.
Wang Zhuoyu, Wang Lei, Lu Ting, et al. Secure storage method of massive big data based on reverse operation[J]. Ordnance Automation, 2024, 43(7): 23-26(in Chinese).
- [19] 马涛, 王仁智, 陈丰, 等. 车-路协同下路面状态感知技术发展现状与展望[J]. 山东大学学报(工学版), 2024, 54(5): 1-28.
Ma Tao, Wang Renzhi, Chen Feng, et al. Development status and prospects of pavement state sensing technology under vehicle-road interaction[J]. Journal of Shandong University (Engineering Edition), 2024, 54(5): 1-28(in Chinese).
- [20] 杨挺, 杨风霞, 叶芷杉, 等. 基于动态采样压缩感知的超谐波监测方法[J]. 中国电机工程学报, 2023, 43(16): 6278-6288.
Yang Ting, Yang Fengxia, Ye Zhishan, et al. A new measurment method for supra harmonics based on dynamic sampling and compressed sensing[J]. Proceedings of the CSEE, 2023, 43(16): 6278-6288(in Chinese).
- [21] 杨挺, 李大帅, 蔡绍堂, 等. 面向用户隐私保护的用电数据压缩加密方法[J]. 中国电机工程学报, 2022, 42(增1): 58-69.
Yang Ting, Li Dashuai, Cai Shaotang, et al. Power data compression and encryption method for user privacy protection[J]. Proceedings of the CSEE, 2022, 42(Suppl 1): 58-69(in Chinese).
- [22] Wang J, Kwon S, Shim B. Generalized orthogonal matching pursuit[J]. IEEE Transactions on Signal Processing, 2012, 60(12): 6202-6216.

(责任编辑: 孙立华)