

基于BI-GRU改进的Seq2Seq网络的变压器油中溶解气体浓度预测方法

汤健¹, 侯慧娟¹, 陈洪岗², 王劲菁², 盛戈皞¹, 江秀臣¹

(1. 上海交通大学 电气工程系, 上海 200240; 2. 国网上海市电力公司电力科学研究院, 上海 200437)

摘要:基于门控循环单元(GRU)构建双向多层门控循环单元,并引入编码器-解码器结构搭建Seq2Seq网络模型,通过优化神经元及神经网络结构提取时序数据依赖关系。同时引入注意力机制和Scheduled Sampling算法,自动获取与当前时刻预测输出显著相关的关键输入时间点,提高长时间预测的精度。变压器正常运行状态下的气体浓度预测算例结果表明,与基于简单GRU模型及简单Seq2Seq模型的方法相比,所提方法的预测误差更低且预测的发展趋势更符合真实值;变压器异常运行状态下的气体浓度预测算例结果表明,所提方法的平均相对误差和最大相对误差相比长短期记忆(LSTM)网络方法分别降低了0.73%和2.31%。

关键词:电力变压器;油中溶解气体;门控循环单元;Seq2Seq;注意力机制;Scheduled Sampling算法

中图分类号:TM 41

文献标志码:A

DOI:10.16081/j.epae.202111017

0 引言

变压器的安全稳定运行关系着电网能否可靠供电^[1]。油浸式电力变压器在运行过程中,会产生各种低分子烃类等气体。当变压器出现异常状态时,油中溶解气体含量及其发展趋势会发生相应的变化^[2]。因此,对变压器油中溶解气体浓度的历史数据^[3]进行时间序列建模,可对其发展趋势进行合理的预测,进而评估变压器运行状态,对制定变压器故障预防性措施具有重要参考意义^[4]。

目前,油中溶解气体浓度预测方法可归纳为统计预测方法^[5]、组合预测方法^[6]和人工智能预测方法等^[7]。其中,统计预测方法包括时间序列模型和灰色模型,但该方法的拟合函数单一,精度受时间序列本身的分布规律影响较大;组合预测方法受权值选择的影响较大,且无法挖掘时序信息间的依赖关系;人工智能预测方法的典型代表有支持向量机回归SVR(Support Vector machine Regressive)模型和多层感知机MLP(MultiLayer Perceptron)模型。然而上述方法割裂了输入时序信息间的依赖关系,因此被适应性和稳定性更好的循环神经网络RNN^[8](Recurrent Neural Network)模型取代。其中长短期记忆LSTM(Long Short-Term Memory)网络^[9]以及门控循环单元GRU^[10](Gate Recurrent Unit)由于引入了门控单元,可以更有效地处理时间序列中的长期依赖问题。文献^[11]指出:GRU相比LSTM在小规模任务上的性能并无降低;同时由于门控单元数量的减少,运算速度更快。文献^[12]提出了双向RNN(BI-RNN)模

型,分别从前向和后向同时处理时序数据。上述时序分析模型都只能对输入、输出定长的序列进行预测,文献^[10]和文献^[13]所提出的Seq2Seq(Sequence to Sequence)模型通过引入编码器-解码器结构很好地解决了该问题,并提高了对于未来较长时间段预测的精度。基于Seq2Seq模型的注意力机制算法^[15]可增强关键时间点在预测中的表达。文献^[16]提出的Scheduled Sampling算法使训练阶段解码器的预测输出有一定概率地取代真实标签作为下一时刻的RNN输入,从而提升测试时的预测性能。

本文以双向门控循环单元(BI-GRU)为基本单元构建Seq2Seq模型,通过双向多层GRU结构以及Seq2Seq模型特殊的编码器-解码器网络结构,实现对时序数据间依赖关系的深度提取;引入注意力机制和Scheduled Sampling算法,一方面能够自动提取关键时间节点对当前值进行预测,实现较长时间的气体浓度预测;另一方面使模型在训练和测试时的解码器输入数据分布趋于一致,从而提升模型实际预测时的准确性和鲁棒性。将本文方法应用于某500 kV变电站主变压器在线监测油色谱数据的预测分析,将9种特征参量,即氢气浓度、甲烷浓度、乙烯浓度、乙烷浓度、一氧化碳浓度、二氧化碳浓度、顶层油温、环境温度、总烃浓度的监测数据输入模型进行预测。结果表明:进行正常状况下的甲烷气体浓度预测时,本文方法相比简单GRU网络以及简单Seq2Seq模型具有更低的误差水平,且预测趋势更贴合真实的发展趋势;进行异常状况下的氢气浓度预测时,本文方法的预测精度显著高于MLP以及SVR模型,且与LSTM方法相比,本文方法能够提前进行准确预测,平均相对误差和最大相对误差分别减少了0.73%和2.31%。

收稿日期:2021-02-10;修回日期:2021-09-29

基金项目:上海交通大学新进青年教师启动计划基金资助项目
Project supported by Startup Program of Young Faculties of SJTU

1 BI-GRU

GRU的结构图如图1所示。GRU将LSTM网络的遗忘门和输入门统一替换为更新门(update gate),并将cell state与hidden state进行合并。图中, x_t 为当前时刻(即 t 时刻)的输入; h_{t-1} 和 h_t 分别为上一时刻(即 $t-1$ 时刻)及当前时刻GRU的状态输出; \tilde{h}_t 为当前时刻的候选状态; r_t 和 z_t 分别为重置门(reset gate)和更新门在当前时刻的输出。GRU的输出计算公式如式(1)所示。

$$\begin{cases} r_t = \sigma(W_r[h_{t-1}, x_t] + b_r) \\ z_t = \sigma(W_z[h_{t-1}, x_t] + b_z) \\ \tilde{h}_t = \tanh(W_d[r_t h_{t-1}, x_t] + b_d) \\ h_t = (1 - z_t)h_{t-1} + z_t \tilde{h}_t \end{cases} \quad (1)$$

$$\tanh(x) = (e^x - e^{-x}) / (e^x + e^{-x}) \quad (2)$$

式中: W_r 和 b_r 分别为重置门的权重和偏置; W_z 和 b_z 分别为更新门的权重和偏置; W_d 和 b_d 分别为门控单元的权重和偏置; $\sigma(\cdot)$ 、 $\tanh(\cdot)$ 分别为sigmoid函数和双曲正切激活函数。

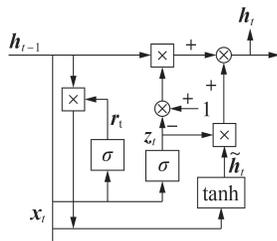


图1 GRU结构图

Fig.1 Structure diagram of GRU

BI-GRU克服了单向GRU仅能处理历史状态信息而无法结合未来信息的缺陷,通过对时间序列同时进行正向和反向计算加强对原序信息的提取效果,提高模型输出结果的准确度。BI-GRU由2个GRU组合而成,它们的状态共同决定了网络输出。

附录A图A1(a)为具有BI-GRU结构的网络示意图。图中,最下方、最上方的节点分别代表输入、输出;中间为具有BI-GRU结构的3层隐藏层。图A1(b)为某层具有BI-GRU结构的按时间步展开的网络示意图,从下到上依次为该层的输入、前向GRU隐藏层、后向GRU隐藏层和该层的输出。在反传过程中共有 $W_1 \sim W_6$ 这6组权重矩阵需要不断更新数值,其中 W_1 、 W_3 分别为输入到前、后向GRU隐藏层的权重矩阵; W_2 、 W_5 分别为前、后向GRU隐藏层间的权重矩阵; W_4 、 W_6 分别为前、后向GRU隐藏层到输出层的权重矩阵。可以看出,每一时刻,当前单元的输出由前向GRU和后向GRU共同计算得到,如式(3)所示。

$$\begin{cases} h_t = f(W_1 x_t + W_2 h_{t-1} + b) \\ h'_t = f(W_3 x_t + W_5 h_{t+1} + b') \\ o_t = g(W_4 h_t + W_6 h'_t + p_o) \end{cases} \quad (3)$$

式中: h'_t 为 t 时刻后向GRU隐藏层的输出; b 、 b' 和 p_o 为偏置; o_t 为 t 时刻神经元输出; $f(\cdot)$ 和 $g(\cdot)$ 为激活函数,本文分别选取双曲正切函数和softmax函数,softmax函数见式(4)。

$$g(x) = \frac{e^{x_d}}{\sum_{d=1}^N e^{x_d}} \quad (4)$$

式中: x_d 为 N 维向量 x 的第 d 个元素。

2 Seq2Seq模型

2.1 编码器-解码器结构

Seq2Seq模型通过2组GRU(本文中所有RNN单元均采用GRU结构)形成编码器-解码器结构。如附录A图A2所示。图中, $x_1 \sim x_5$ 组成输入序列; $y_1 \sim y_3$ 组成输出序列;左、右组GRU分别作为编码器、解码器,解码器在初始时刻接收输入编码<start>作为开始标记,在结束时刻则输出编码<stop>作为结束标记。左组GRU编码器对输入序列进行编码,将最后时刻(或全部时刻)的输出映射为表征输入序列信息的特征向量 c 传递给右组GRU解码器,如式(5)所示。

$$c = d(Uh_T) \quad (5)$$

式中: U 为权重矩阵; h_T 为最后时刻 T 的编码器输出; $d(\cdot)$ 为激活函数,本文选取tanh函数。

将编码器最后时刻的隐藏层状态作为解码器初始隐藏层状态,此后解码器GRU每一时刻的隐藏层状态都传入下一时刻。在每一时刻,解码器GRU接收3组输入,即特征向量 c 、上一时刻解码GRU输出 y_{t-1} 、上一时刻解码GRU隐层状态 h_{t-1} ,如式(6)所示。

$$\begin{cases} h_t = f_{s2s}(G[h_{t-1}, y_{t-1}, c] + b_h) \\ o_t = g_{s2s}(Vh_t + p) \end{cases} \quad (6)$$

式中: G 和 V 为连接权重矩阵; b_h 和 p 为偏置向量; $f_{s2s}(\cdot)$ 和 $g_{s2s}(\cdot)$ 为激活函数。

基于上述原理,Seq2Seq模型能够更有效地发掘和利用序列之间的时序关系,使输出结果具有更强的鲁棒性。

2.2 注意力机制

在编码过程中,时间序列的每个输入时刻点对当前时刻预测值的贡献度并不均等,因此引入注意力机制分配注意力概率分布,提取对当前时刻预测更为重要的输入信息,提高预测的准确度和长期预测性能的稳定性。注意力机制原理具体分解步骤如下。

1) 编码器编译输入信息得到输出时间序列 $\{h_j\}_{j=1}^{T_x}$,其中 T_x 为编码器序列长度。

2) 解码器在 i 时刻,根据来自上一时刻的最顶层

隐藏层状态输出 s_{i-1} , 计算得到编码器 j 时刻的输出在 i 时刻的对应权重 e_{ij} , 如式(7)所示。

$$e_{ij} = V_a^T \tanh(W_a s_{i-1} + U_a h_j) \quad (7)$$

式中: h_j 为编码器 j 时刻的输出状态; V_a^T 、 W_a 与 U_a 为计算注意力权重时全连接层对应的权重, 将在反传过程中更新数值。

3) 根据式(8)对各时刻权重 e_{ij} 进行归一化处理, 得到编码器各时刻输出对解码器在 i 时刻的预测值的重要程度 a_{ij} 。

$$a_{ij} = \frac{\exp(e_{ij})}{\sum_{j=1}^{T_x} \exp(e_{ij})} \quad (8)$$

4) 综合编码器所有状态信息, 与步骤3)的计算结果进行加权求和, 得到解码器在 i 时刻的特征向量 c_i , 如式(9)所示。

$$c_i = \sum_{j=1}^{T_x} a_{ij} h_j \quad (9)$$

通过注意力机制得到的特征向量既包含编码器所有输出状态信息, 又蕴含解码器当前时刻的特征时序注意力关联信息, 且每一时刻互相独立。

2.3 Scheduled Sampling 算法

在 Seq2Seq 模型测试阶段, 每一时刻解码器 GRU 的输出都会作为下一时刻 GRU 单元输入的一部分。而在训练阶段, 通常采用 Teacher Forcing 算法, 即将目标序列中的真实标签作为对应编码器每一步的输入, 以加速模型收敛。此种情况下, 模型在训练和测试阶段解码器输入数据的概率分布并不一致, 很大程度上影响了模型预测性能。

Scheduled Sampling 算法在训练早期主要使用真实标签作为编码器输入, 以引导模型快速过渡到合理状态; 在训练后期, 则更多地使用编码器的前一步输出作为下一步的输入, 以解决数据分布不一致的问题, 对模型进行微调直至收敛。对于测试阶段的 Seq2Seq 模型而言, 一旦输出序列出现错误元素, 则误差会传递到后续元素的输入步骤, 导致后续元素的生成受到错误影响, 使误差随序列生成而不断累积。而 Scheduled Sampling 算法由于在训练阶段就有一定概率会将生成数据作为输入, 因此其训练目标仍然为最大化生成真实目标序列的概率, 从而使模型在实际测试时具有较高的容错性能。

以 h 表示训练阶段的迭代次数, ε_h 表示选择真实标签值作为编码器输入的概率, 则常用的 3 种 Scheduled Sampling 算法定义方式如式(10)所示。

$$\varepsilon_h = \begin{cases} \max(\varepsilon_0, k_1 - Ch) & \text{线性衰减} \\ k_2^h & \text{指数衰减} \\ k_3 / [k_3 + \exp(h/k_3)] & \text{反向sigmoid衰减} \end{cases} \quad (10)$$

式中: ε_0 为限制 ε 的最小值; k_1 和 C 为确定的正数;

k_2, k_3 用于控制衰减幅度, $k_2 \in (0, 1]$, $k_3 > 1$ 。本文选取线性衰减 Scheduled Sampling 算法。

3 变压器油中溶解气体预测模型

3.1 数据预处理

本文对算例数据的预处理操作分为 2 步: ①对原始数据进行标准化操作; ②以滑动窗口的形式构建数据集样本。

3.1.1 数据标准化

对 9 种监测特征参量按式(11)进行归一化处理, 将监测数据映射到 $[0, 1]$ 范围内, 从而提升模型的收敛速度和精度。

$$x_k^m = \frac{x_k^m - x_{\min}^m}{x_{\max}^m - x_{\min}^m} \quad m=1, 2, \dots, 9 \quad (11)$$

式中: $k=1, 2, \dots, T_x$, 为样本序号; x_k^m 为第 m 种监测特征参量的第 k 个样本; x_{\min}^m, x_{\max}^m 分别为第 m 种监测特征参量样本中的最小值、最大值。

3.1.2 基于滑动窗口的数据集样本构建

以正常状态为例, 全部训练样本共 770 例(见第 4 节), 而 Seq2Seq 网络编码器时间步长为 60, 解码器时间步长为 30, 若顺次选取, 则可用的训练样本容量仅为 8。由于 Seq2Seq 模型需要大量训练样本优化网络模型, 本文提出一种基于滑动窗口构造样本的方法, 在长度为 T_x 的时间序列上, 以 $L(L=L_x+L_y)$, 其中 L_x, L_y 分别为编码器、解码器的时间步长)作为窗口宽度, 以 1 作为滑动步长, 共可构建 T_x-L+1 组数据样本, 极大丰富了样本容量, 且保持了数据集在时间刻度上的连续性, 可以充分利用数据集信息, 更好地刻画时间序列的时间依赖关系, 提高模型鲁棒性。

利用上述方法将时间序列 $X = \{x_1, x_2, \dots, x_{T_x}\}$ 转化为 $(T_x-L+1) \times L$ 维的矩阵, 如式(12)所示。

$$X = \begin{bmatrix} x_1 & x_2 & \dots & x_{L_x} & x_{L_x+1} & \dots & x_L \\ x_2 & x_3 & \dots & x_{L_x+1} & x_{L_x+2} & \dots & x_{L+1} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ x_{T_x-L+1} & x_{T_x-L+2} & \dots & x_{T_x-L_x} & x_{T_x-L_x+1} & \dots & x_{T_x} \end{bmatrix} \quad (12)$$

由于 X 的每个元素都是 9 维向量, 因此式(12)实际上是一个 $(T_x-L+1) \times L \times 9$ 维的张量。其中, 前 L_x 列作为模型输入数据; 后 L_y 列中预测特征参量序号所对应的 $(T_x-L+1) \times L$ 维矩阵为输入数据对应的真实标签。

3.2 油中溶解气体预测模型

本文建立了以 GRU 为基本单元的 Seq2Seq 网络模型, 如附录 A 图 A3 所示。图中, 编码器采用双层 BI-GRU 单元对原始输入时序数据进行编码; 由于预测结果具有顺次生成的性质, 因此解码器采用单向单层 GRU 单元输出预测气体浓度; g_{t-2}, g_{t-1} 分别为 $t-2, t-1$ 时刻的解码器输出, y_{t-2}, y_{t-1} 分别为 $t-2, t-1$ 时刻的真实值标签, 均有一定概率会作为 $t-1, t$ 时刻的

解码器的实际输入。

每个隐藏层GRU单元神经元个数设置为30,损失函数采用式(13)所示的平均绝对误差 L_1 ,迭代次数为1000次。为防止梯度爆炸,采用梯度裁剪策略,阈值设置为5。为保证模型收敛速率和精度,初始学习率设置为0.001,此后每经过50次迭代学习率衰减为之前的80%。

$$L_1 = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (13)$$

式中: n 为样本数; \hat{y}_i 为预测值; y_i 为样本真实标签。

本文模型采用注意力机制自动提取输入时间序列的关键时刻点并分配相应权重,计算当前时刻对应编码器输出的特征向量,并在训练阶段采用线性衰减Scheduled Sampling算法(ε_0 取0.1, k_1 和 C 分别取1和0.0015),以提高模型预测的准确率和稳定性。

4 算例分析

为验证本文模型在预测变压器油中溶解气体浓度时的有效性,本文以某500 kV变电站主变压器在线监测的油色谱数据为例,分别选取变压器不同运行状态下对不同参量^[17]的预测结果进行算例分析。油色谱数据的监测周期为1 d,共有氢气浓度、甲烷浓度、乙烯浓度、乙烷浓度、一氧化碳浓度、二氧化碳浓度、顶层油温、环境温度、总烃浓度9种特征参量,包含2014年9月14日至2017年3月4日共800个时序样本点的正常运行状态数据,以及2017年6月14日至2018年10月27日共500个时序样本点的异常运行状态数据,对2种数据分别进行时长为30 d和20 d的预测。

在评估模型性能方面,采用平均相对误差 δ_{mean} 和最大相对误差 δ_{max} 作为评判指标,分别如式(14)、(15)所示。

$$\delta_{\text{mean}} = \frac{1}{T_y} \sum_{i=1}^{T_y} \left| \frac{\hat{y}_i - y_i}{y_i} \right| \times 100\% \quad (14)$$

$$\delta_{\text{max}} = \max_{1 \leq i \leq T_y} \left(\left| \frac{\hat{y}_i - y_i}{y_i} \right| \right) \times 100\% \quad (15)$$

式中: T_y 为每例预测样本标签总数。

δ_{mean} 可以反映预测结果偏离真实值的整体情况, δ_{max} 可以反映预测结果的稳定性能, δ_{mean} 和 δ_{max} 越小,说明预测结果越接近真实值,模型预测性能越精确和稳定。

4.1 算例1:正常状态参量预测结果分析

正常状态下变压器部分状态参量监测数据如图2所示。由图可见:甲烷浓度稳定增长,年增长率约为 $1.6 \mu\text{L}/\text{L}$;氢气和总烃浓度则呈现周期性增长。

对于正常状态下的变压器数据,选取甲烷作为

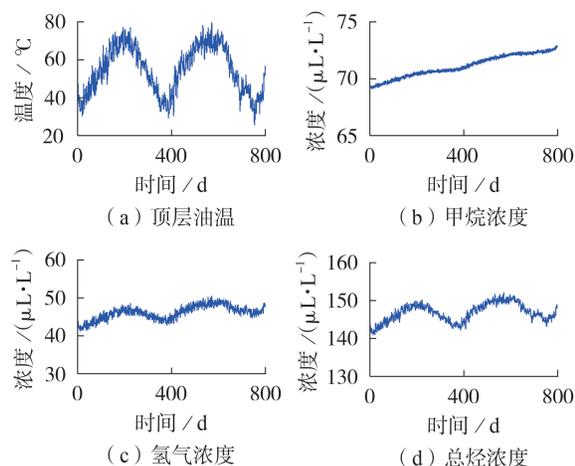


图2 正常状态下变压器部分状态参量监测数据
Fig.2 Partial state parameter monitoring data of transformer under normal condition

特征气体,预测步骤如下:

1)对原始数据进行标准化操作;

2)基于9种特征参量前770 d的监测数据,通过滑动窗口方法构造训练集数据,编码器时间步长取为60,则输入数据维度为 $711 \times 60 \times 9$,解码器时间步长取为30,则输出数据维度为 $711 \times 30 \times 1$,将后30 d的监测数据作为测试集;

3)在训练集上训练模型用前60 d的9种参量全部的历史数据预测后30 d的甲烷浓度;

4)预测后30 d的甲烷浓度,并与测试集中的真实甲烷浓度值比对以评估模型性能。

对气体进行为期1个月的浓度预测,并评估不同预测时长下的效果,为后续现场监测数据提供参考。

本文方法在变压器正常状态下对甲烷气体浓度的预测结果在训练集和测试集上的 L_1 变化如图3所示。由图可见,在训练初始阶段, L_1 快速下降,200次迭代之后下降逐渐趋缓,至1000次迭代完毕可认为模型达到收敛状态。

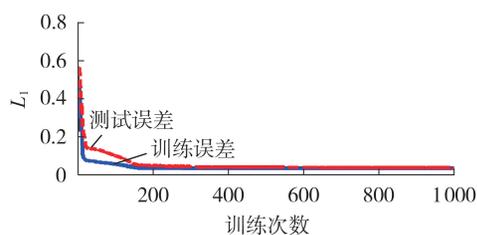


图3 本文方法的预测结果在训练集和测试集上的 L_1 曲线
Fig.3 L_1 curve of prediction results by proposed method on training set and test set

在算例1中,将本文方法与以下3种方法进行对比:①方法A,Scheduled Sampling算法+单向GRU;②方法B,采用BI-GRU,不采用Scheduled Sampling算法;③方法C,采用单向GRU,不采用Scheduled

Sampling算法。

算例1中,本文模型在测试集上对甲烷浓度的预测结果如图4所示,其与方法A—C的预测结果的对比如附录A图A4(a)所示。

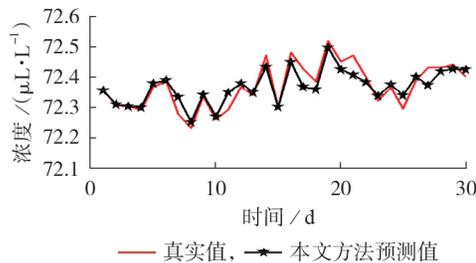


图4 本文方法在测试集上的预测结果

Fig.4 Prediction result of proposed method on test set

为验证本文方法相比简单GRU模型在多步预测时的性能提升,设置预测时间为30 d,在测试集上对比验证本文方法与以下2种基于简单GRU模型的方法:①简单GRU模型一般用于单步预测,本算例将每次单步预测结果融入历史数据,构造新的时间输入序列,以滚动单步预测的形式实现多步预测,记为方法D;②通过修改网络结构,在原输出层后连接神经元个数等于预测时间长度(即30)的全连接层,这也是常用的GRU模型进行多步预测的方法,记为方法E。本文方法与方法D、E的比较结果如附录A图A4(b)所示。

预测天数为7、14、20、30 d时,本文方法和5种对比方法的平均相对误差、最大相对误差分别如表1、2所示。

表1 算例1中6种方法的预测结果平均相对误差

Table 1 Mean relative percentage error of prediction results of six methods in Case 1

预测时间/d	$\delta_{\text{mean}} / \%$					
	本文方法	方法A	方法B	方法C	方法D	方法E
7	0.17	0.28	0.19	0.69	0.43	0.54
14	0.23	0.60	0.37	0.79	0.63	0.68
20	0.28	0.61	0.54	0.77	0.75	0.81
30	0.31	0.64	0.57	0.70	0.81	0.86

表2 算例1中6种方法的预测结果最大相对误差

Table 2 Maximal relative percentage error of prediction results of six methods in Case 1

预测时间/d	$\delta_{\text{max}} / \%$					
	本文方法	方法A	方法B	方法C	方法D	方法E
7	0.79	0.79	0.82	1.12	1.28	1.77
14	0.81	1.79	1.58	2.19	1.67	1.79
20	0.82	2.68	1.94	2.19	2.08	2.91
30	0.90	2.68	1.94	2.19	2.08	2.91

由图A4及表1、2可见:本文方法在前15 d的预测结果与真实值基本一致,后15 d两者在趋势上仍保持一致;本文方法的误差也保持在较低水平,综合

来看其在每一时间段上的预测结果都具有最低的平均相对误差和最大相对误差,因此可认为与其他方法相比,本文方法具有更高的预测精度及更好的鲁棒性。

对比分析本文方法与方法A以及方法B与方法C后,可知,由于神经元结构由单向变为双向,未来信息的加入实现了对数据间依赖信息的充分挖掘,因此采用BI-GRU的网络的预测结果误差远低于同条件下采用单向GRU的网络。由图A4(a)可见:方法A的预测结果在第7天之后就出现了严重偏离实际值的情况;方法C的预测结果则是一直处于较小幅度的波动上升状态,无法准确体现真实的甲烷气体浓度的变化趋势和幅度。结合表1、2可知:在第14天,方法A的平均相对误差达到了0.60%,而本文方法的平均相对误差仅有0.23%;在第30天,方法A的最大相对误差为2.68%,而本文方法的最大相对误差仅为0.90%。结合表1、2对照分析本文方法与方法B以及方法A与方法C可知,采用Scheduled Sampling算法后,本文方法的平均相对误差、最大相对误差均得到了一定程度的降低,预测性能更好。

在与简单GRU模型的预测性能比较中,由图A4(b)可看出,方法D、E在预测初期都有明显的滞后性,结合表1、2可知,方法D、E的预测误差随着预测时间的增长而逐渐扩大发散。

综合来看,本文方法的平均相对误差和最大相对误差相比方法C分别减少了0.39%和1.29%,相比方法E分别由0.86%和2.91%减少到0.31%和0.90%。本文模型具有更复杂的神经元结构和网络结构,并加入了注意力机制和Scheduled Sampling算法提升模型性能,所以可以充分发掘时序信息的依赖关系,增强关键时刻点的信息表达,且可以预测气体浓度发展趋势,具有更好的预测效果和稳定性能。

变压器正常状态下,本文方法和基于GRU模型的方法(简称GRU方法)在测试集上对氢气浓度的预测结果、误差分别如附录A图A5、表3所示。可以看出,本文方法的预测结果相比GRU模型的预测结果在趋势上与真实值更吻合,平均相对误差和最大相对误差分别降低了0.64%和1.6%。

表3 正常状态下的氢气浓度预测结果误差

Table 3 Error of H₂ concentration prediction result under normal condition

方法	$\delta_{\text{mean}} / \%$	$\delta_{\text{max}} / \%$
本文方法	0.48	1.13
GRU	1.12	2.73

4.2 算例2:异常状态参量预测结果分析

异常状态下变压器部分状态参量监测数据如图5所示。由图可见:在第1—400天期间,各参量数

值稳定,无明显异常状态;然而从第400天开始,各参量数值均出现异常快速增长,随后由于近3个月的顶层油温呈现异常状态,部分油中溶解气体的浓度快速上升,故变电站对该变压器进行停运检视。该变压器顶层油温上升且甲烷、氢气浓度快速增长,而其他气体浓度无显著增长,因此判定该变压器存在低温过热潜伏性缺陷。

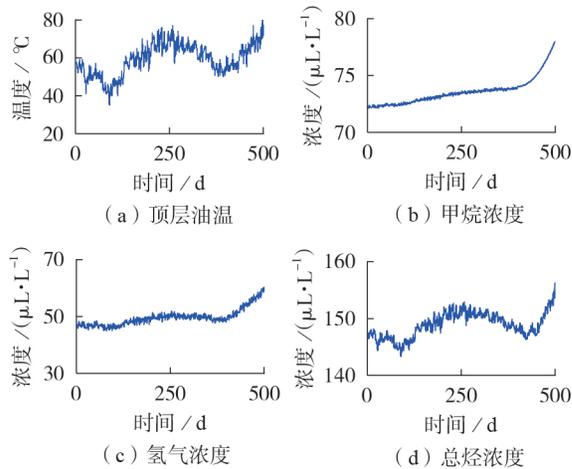


图5 异常状态下变压器部分状态参量监测数据

Fig.5 Partial state parameter monitoring data of transformer under abnormal condition

对于异常状态变压器数据,预测步骤同算例1。以前480 d的监测数据作为训练集样本,后20 d的监测数据作为测试集,选取氢气作为特征气体。由于异常数据样本量较少,因此取编码器时间步长为50,训练时输入数据维度为 $431 \times 50 \times 9$,解码器时间步长为20,输出训练数据维度为 $431 \times 20 \times 1$,即训练模型用前50 d的全部9种参量历史数据预测后20 d的氢气气体浓度,并在不同预测时间下评估模型预测结果,以预测变压器缺陷的发展趋势,提前发现设备的潜伏性隐患,通过后续故障诊断工作预知可能发生的故障类型,为检修做好预警和准备。

将本文方法与LSTM网络方法、基于MLP模型的方法(简称MLP方法)、基于SVR模型的方法(简称SVR方法)进行比较。其中,LSTM网络方法采用[30,30]的双隐藏层结构和滚动多步预测的方法,学习率为0.001,训练周期为100次;MLP方法采用[20,30,20]的三隐藏层结构,迭代次数为1000;SVR方法使用径向基核函数(RBF),惩罚因子设置为1000。4种方法在测试集上的氢气浓度预测结果如附录A图A6所示,预测结果误差如表4、5所示。

由图A6和表4、5可见:MLP和SVR方法由于单独学习输入时序点,没有考虑历史数据之间的关联性,导致预测结果偏离实际值较多;MLP方法的预测结果与真实值的吻合度较差,SVR方法的预测结果在5 d后趋于定值,明显不理想,这2种方法的误差

表4 算例2中不同方法的预测结果的平均相对误差

Table 4 Mean relative percentage error of prediction results of different methods in Case 2

预测时间/d	$\delta_{\text{mean}} / \%$			
	本文方法	LSTM网络方法	MLP方法	SVR方法
7	0.52	0.81	1.45	1.15
14	0.68	1.21	1.56	1.29
20	0.82	1.65	1.86	1.90

表5 算例2中不同方法的预测结果的最大相对误差

Table 5 Maximal relative percentage error of prediction results of different methods in Case 2

预测时间/d	$\delta_{\text{max}} / \%$			
	本文方法	LSTM网络方法	MLP方法	SVR方法
7	0.98	1.63	4.11	3.08
14	1.53	2.56	4.11	3.08
20	2.15	4.46	4.24	4.97

水平也始终高于本文方法和LSTM网络方法。LSTM网络方法由于引入了时间记忆单元,其在短期(7 d)预测过程中,平均相对误差不超过1%,最大相对误差也仅有1.63%。而随着预测时间跨度的增加,预测误差累积增加,由图A6可见其在7 d之后的预测结果就已出现了较大偏差,第20天的误差水平已与MLP和SVR方法相当。本文方法的平均相对误差和最大相对误差与LSTM网络方法相比分别由1.65%和4.46%减少至0.82%和2.15%。

本文方法和GRU方法在测试集上对其他特征气体(甲烷、乙烷、乙烯、一氧化碳、二氧化碳)浓度的预测结果误差如表6所示。由表可见,本文方法对各类气体浓度预测结果的平均相对误差均在1.5%以内,最大相对误差均在2.5%以内,相比GRU方法具有更优异的预测结果。其中由于甲烷浓度在异常状况下几乎呈直线增长(见图5(b)),因此预测结果较容易判断,误差值也较低。综上所述,本文方法在对特征气体浓度预测上具有一定的准确性和合理性。

表6 各特征气体浓度的预测误差对比

Table 6 Comparison of concentration prediction error for each characteristic gas

特征气体	$\delta_{\text{mean}} / \%$		$\delta_{\text{max}} / \%$	
	本文方法	GRU	本文方法	GRU
甲烷	0.33	1.20	0.77	3.80
乙烷	1.08	2.44	1.97	4.10
乙烯	0.85	2.39	1.65	3.51
一氧化碳	1.42	3.07	2.38	6.48
二氧化碳	0.57	1.91	1.35	2.87

本文方法由于Seq2Seq模型和注意力机制、Scheduled Sampling算法的引入,能够深入理解时序数据间的依赖关系,较为准确地预测气体浓度未来发展状况,可在实际变压器状态监测中为检修提供时间裕量和数据支持。

5 结论

本文提出一种基于BI-GRU的Seq2Seq模型用于变压器油中溶解气体浓度的预测,并引入注意力机制和Scheduled Sampling算法提升模型预测性能。在变压器正常和异常状态下对各特征气体的浓度进行预测,得到以下结论。

1)由于Seq2Seq模型和注意力机制的引入,本文方法能够进行任意输入、输出长度的时序预测,并深入挖掘时序数据间联系,提取关键时间点进行较长时间的预测。

2)BI-GRU单元的引入可以提升网络挖掘时序信息依赖关系的能力,Scheduled Sampling算法使模型在实际测试时具有较高的容错性能,提升模型预测精度。

3)变压器正常运行状态下的气体浓度预测算例结果表明,本文方法能够准确预测气体浓度的发展趋势,相比简单GRU模型及简单Seq2Seq网络误差更低;变压器异常状态下的气体浓度预测算例结果表明,本文方法相比传统MLP及SVR方法能够充分利用时序关系,从而提高预测性能;相比LSTM网络方法,本文方法能够提前进行准确预测,最大相对误差减少了2.31%。

附录见本刊网络版(<http://www.epae.cn>)。

参考文献:

- [1] 崔宇,侯慧娟,胥明凯,等. 基于双重注意力机制的变压器油中溶解气体预测模型[J]. 中国电机工程学报,2020,40(1):338-347,400.
CUI Yu, HOU Huijuan, XU Mingkai, et al. A prediction method for dissolved gas in power transformer oil based on dual-stage attention mechanism[J]. Proceedings of the CSEE, 2020, 40(1):338-347,400.
- [2] 李洪超,王伟刚,董雪梅. 基于M-LS-SVR的变压器油中溶解气体浓度预测[J]. 电气技术,2016(1):76-80.
LI Hongchao, WANG Weigang, DONG Xuemei. Concentration prediction of dissolved gases in transformer oil based on M-LS-SVR[J]. Electrical Engineering, 2016(1):76-80.
- [3] 林湘宁,黄京,熊卫红,等. 变压器油中溶解气体浓度的区间预测[J]. 电力自动化设备,2016,36(4):73-77.
LIN Xiangning, HUANG Jing, XIONG Weihong, et al. Interval prediction of dissolved-gas concentration in transformer oil [J]. Electric Power Automation Equipment, 2016, 36(4):73-77.
- [4] 江秀臣,盛戈峰. 电力设备状态大数据分析的研究和应用[J]. 高压技术,2018,44(4):1041-1050.
JIANG Xiuchen, SHENG Gehao. Research and application of big data analysis of power equipment condition[J]. High Voltage Engineering, 2018, 44(4):1041-1050.
- [5] 赵文清,朱永利,张小奇. 基于改进型灰色理论的变压器油中溶解气体预测模型[J]. 电力自动化设备,2008,28(9):23-26.
ZHAO Wenqing, ZHU Yongli, ZHANG Xiaoqi. Prediction model for dissolved gas in transformer oil based on improved grey theory[J]. Electric Power Automation Equipment, 2008, 28(9):23-26.
- [6] 张施令,姚强. 基于WNN-GNN-SVM组合算法的变压器油色谱时间序列预测模型[J]. 电力自动化设备,2018,38(9):155-161.
ZHANG Shiling, YAO Qiang. Predicting model of transformer

DGA time series based on WNN-GNN-SVM combined algorithm[J]. Electric Power Automation Equipment, 2018, 38(9):155-161.

- [7] 张磊,杨廷方,李炜,等. 基于LSTSVR模型的边缘计算预测变压器平均油温及绕组热点温度[J]. 电力自动化设备,2020,40(8):197-203.
ZHANG Lei, YANG Tingfang, LI Wei, et al. Prediction of transformer average oil temperature and winding hot spot temperature by edge computation based on LSTSVR model[J]. Electric Power Automation Equipment, 2020, 40(8):197-203.
- [8] VERMAAK J, BOTHA E C. Recurrent neural networks for short-term load forecasting[J]. IEEE Transactions on Power Systems, 1998, 13(1):126-132.
- [9] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural Computation, 1997, 9(8):1735-1780.
- [10] CHO K, VAN MERRIËNBOER B, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Doha, Qatar: Association for Computational Linguistics, 2014:1724-1734.
- [11] GREFF K, SRIVASTAVA R K, KOUTNIK J, et al. LSTM: a search space odyssey [J]. IEEE Transactions on Neural Networks and Learning Systems, 2017, 28(10):2222-2232.
- [12] SUTSKEVER I, VINYALS O, LE Q V. Sequence to sequence learning with neural networks [C]//Proceedings of the 27th International Conference on Neural Information Processing Systems-Volume 2. Montreal, Canada: MIT Press, 2014:3104-3112.
- [13] SCHUSTER M, PALIWAL K K. Bidirectional recurrent neural networks[J]. IEEE Transactions on Signal Processing, 1997, 45(11):2673-2681.
- [14] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: MIT Press, 2017:6000-6010.
- [15] BENGIO S, VINYALS O, JAITLY N, et al. Scheduled sampling for sequence prediction with recurrent neural networks [EB/OL]. (2015-09-23) [2020-12-05]. <https://arxiv.org/abs/1506.03099>; [spm=a2c4e.11153940.blogcont221611.26.72801f465o9PhE&file=1506.03099](https://arxiv.org/abs/1506.03099).
- [16] SHENG G H, HOU H J, JIANG X C, et al. A novel association rule mining method of big data for power transformers state parameters based on probabilistic graph model [J]. IEEE Transactions on Smart Grid, 2018, 9(2):695-702.
- [17] 国家能源局. 变压器油中溶解气体分析和判断导则: DL/T 722—2014[S]. 北京: 中国电力出版社, 2015.

作者简介:



汤健

汤健(1996—),男,江苏南京人,硕士研究生,研究方向为输变电设备状态的评估(**E-mail**: tangjian1@sjtu.edu.cn);

侯慧娟(1984—),女,山东郓城人,助理研究员,博士,通信作者,研究方向为电力设备绝缘的在线监测及故障诊断(**E-mail**: houhuijuan@sjtu.edu.cn);

盛戈峰(1974—),男,湖南株洲人,教授,博士,研究方向为输变电设备状态监测及其智能化(**E-mail**: shenghe@sjtu.edu.cn);

江秀臣(1965—),男,山东郓城人,教授,博士研究生导师,博士,研究方向为电气设备在线监测、状态检修和自动化(**E-mail**: xcjiang@sjtu.edu.cn)。

(编辑 任思思)

(下转第217页 continued on page 217)

Low-voltage PLC network routing method based on improved genetic ant colony algorithm

CUI Ying

(Zhuhai Power Supply Bureau of Guangdong Power Grid Co.,Ltd.,Zhuhai 519000,China)

Abstract: In the networking of low-voltage PLC (Power Line Communication) network, multiple networks coexist for a short time when nodes are energized under the condition that the distance between nodes is relatively far or the channel environment is relatively harsh, which severely affects the communication reliability. Aiming at this problem, the multi-network rapid fusion method based on CSMA/CA+TDMA (Carrier Sense Multiple Access with Collision Avoid+Time Division Multiple Access) hybrid protocol is explored. The proposed method can intelligently identify multiple networks in the area, select the network with the smallest MAC (Media Access Control) address as the multi-network fusion direction, dissolve the network with relatively big MAC address. It can solve the problem of multi-network uncertain fusion. On this basis, to solve the existing problems that the genetic algorithm has poor local search ability under the constraint of QoS (Quality of Service) parameters and is difficult to obtain the optimal solution of on-demand routing, a hot standby routing method based on improved genetic ant colony algorithm is proposed in asymmetric channel environment. Since the source nodes and destination nodes are not involved in crossover and mutation, the generation of invalid chromosomes can be effectively avoided. The optimal retention mechanism is used to find an approximate optimal solution, and then the approximate optimal solution is converted into the initial pheromone of ant colony algorithm to find the global optimal solution. The simulative results show that the proposed method is more effective than the traditional methods.

Key words: Ubiquitous Electric Internet of Things; low-voltage power line communication; network fusion; routing method; improved genetic ant colony algorithm

(上接第202页 continued from page 202)

Concentration prediction method based on Seq2Seq network improved by BI-GRU for dissolved gas in transformer oil

TANG Jian¹, HOU Huijuan¹, CHEN Honggang², WANG Shaojing², SHENG Gehao¹, JIANG Xiuchen¹

(1. Department of Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China;

2. State Grid Shanghai Electric Power Research Institute, Shanghai 200437, China)

Abstract: Based on the GRU (Gate Recurrent Unit), the bidirectional multi-layer GRU is constructed. The encoder-decoder structure is introduced to build a Seq2Seq (Sequence to Sequence) network model. The time series data dependencies are obtained automatically by optimizing neurons and neural network structure. At the same time, the attention mechanism and the Scheduled Sampling algorithm are introduced to automatically obtain the key input time points significantly related to the prediction output at the current moment, so as to improve the accuracy of long-term prediction. In the case of gas concentration prediction under normal operating condition of transformer, compared with the methods based on simple GRU model and the simple Seq2Seq model, the proposed method has lower prediction error, and the prediction development trend is more in line with the true value. In the case of gas concentration prediction under abnormal operating condition of transformer, the average relative error and maximum relative error of the proposed model are respectively reduced by 0.73% and 2.31% compared with the LSTM (Long Short-Term Memory) network method.

Key words: power transformers; dissolved gas in oil; gate recurrent unit; Seq2Seq; attention mechanism; Scheduled Sampling algorithm

附录 A

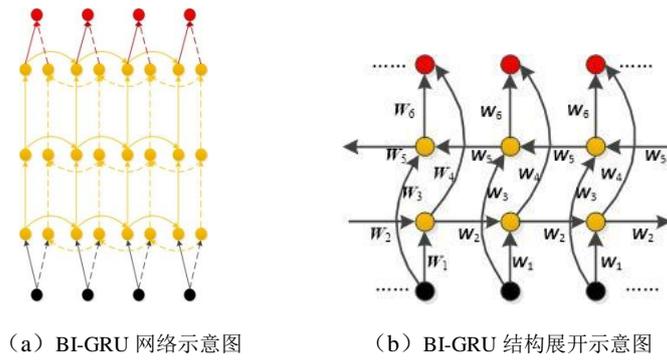


图 A1 BI-GRU 结构示意图

Fig.A1 Schematic diagram of structure of BI-GRU

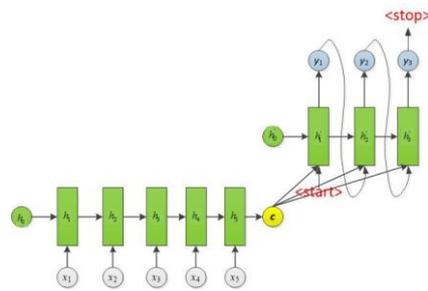


图 A2 Seq2Seq 网络模型示意图

Fig.A2 Schematic diagram of Seq2Seq network model

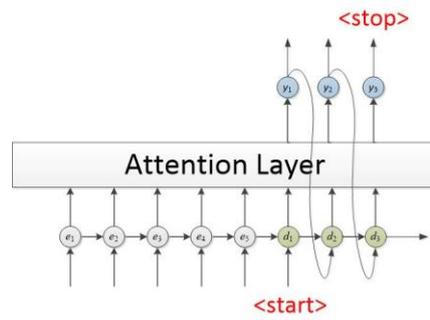
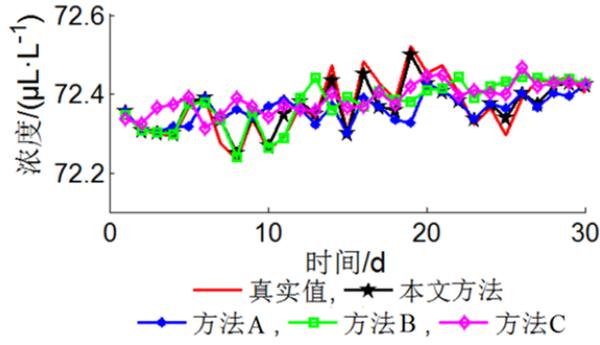
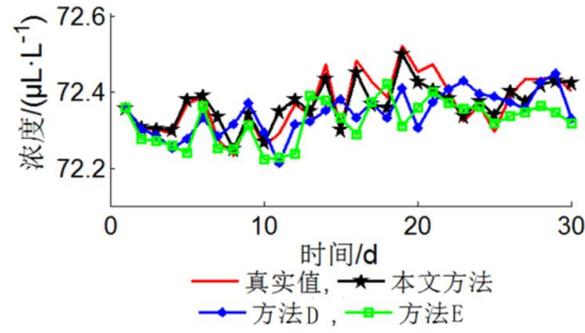


图 A3 本文的 Seq2Seq 网络模型示意图

Fig.A3 Schematic diagram of proposed Seq2Seq model



(a) 本文方法与方法 A—C 的对比



(b) 本文方法与方法 D、E 的对比

图 A4 本文方法与方法 A—E 的预测结果对比

Fig.A4 Comparison of prediction results among proposed method, Method A, Method B, Method C, Method D and Method E

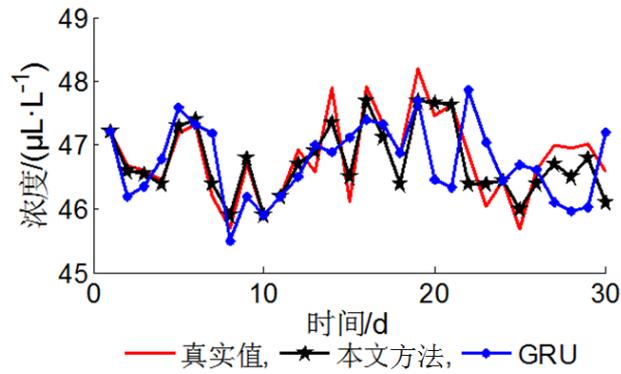


图 A5 正常状态下氢气浓度预测结果

Fig.A5 Result of H_2 concentration prediction under normal condition

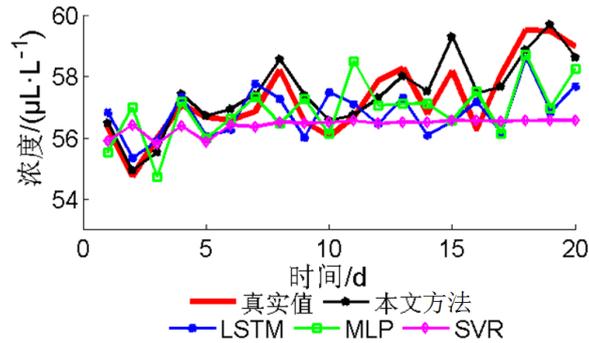


图 A6 本文方法与其他机器学习方法的预测结果对比

Fig.A6 Comparison of prediction results between proposed method and other machine learning methods