基于改进 LSTM 的高速列车牵引系统微小渐变故障诊断

冒泽慧¹, 顾彧行¹, 姜斌^{1*}, 许德智², 孙秀文^{1,3}, 刘文静⁴

1. 南京航空航天大学自动化学院, 南京 211106

2. 江南大学物联网工程学院, 无锡 214122

3. 南京图灵人工智能研究院, 南京 210046

4. 中国空间技术研究院, 北京 100094

* 通信作者. E-mail: binjiang@nuaa.edu.cn

收稿日期: 2020-05-28; 修回日期: 2020-07-30; 接受日期: 2020-10-30; 网络出版日期: 2021-05-18

国家自然科学基金项目 (批准号: 61922042, 61973140) 和江苏高校“青蓝工程”, 111 引智基地 (批准号: B20007) 资助

摘要 高速列车牵引系统的微小渐变故障诊断是保障高速列车安全可靠运行的一项重要任务, 传统的数据驱动方法通常难以在故障初期实现微小渐变故障的诊断. 本文考虑到传感器时序信号的非平稳特性和微小渐变故障具有的慢时变特性, 提出了一种基于改进长短期记忆网络 (long short-term memory, LSTM) 的微小渐变故障无监督诊断方法. 首先, 我们提出了一种基于 LSTM 的改进网络结构, 并结合门控循环单元 (gated recurrent unit, GRU) 解码器, 构造了一种对非平稳特性具有鲁棒性的自编码器模型以提取故障特征. 进而, 采用 t-SNE (t-distributed stochastic neighbour embedding) 对提取的故障特征向量进行降维, 使用 DBSCAN (density-based spatial clustering of applications with noise) 进行故障聚类与识别, 实现了对微小渐变故障的无监督诊断. 最后, 我们基于 CRH₂ 牵引系统的半物理仿真平台数据对上述方案进行了验证. 实验结果表明, 该无监督诊断方法对同时包含已知和未知故障的情况能够达到 95% 以上的诊断准确率.

关键词 故障诊断, 微小渐变故障, 无监督学习, 改进的 LSTM 网络

1 引言

高速列车因其高效性和乘坐的舒适性而越来越普及, 为此, 人们广泛研究其控制问题以保障铁路运输系统的安全性和可靠性^[1~4]. 机械和电子元件的老化以及在复杂环境下的长时间运作等各种因素, 都可能导致故障发生, 引发高速列车减速甚至停车, 这促进了高速列车故障检测和容错控制的设计研究^[5~8].

引用格式: 冒泽慧, 顾彧行, 姜斌, 等. 基于改进 LSTM 的高速列车牵引系统微小渐变故障诊断. 中国科学: 信息科学, 2021, 51: 997–1012, doi: 10.1360/SSI-2020-0382
Mao Z H, Gu Y X, Jiang B, et al. Incipient fault diagnosis for high-speed train traction systems via improved LSTM (in Chinese). Sci Sin Inform, 2021, 51: 997–1012, doi: 10.1360/SSI-2020-0382

牵引系统是高速列车的核心部件, 由牵引电机、逆变器、脉宽调制器、整流器和相关机械传动装置组成, 为整个列车提供动力. 牵引系统中的故障一般是来自电容、电阻、绝缘栅双极晶体管、传感器等. 目前, 已有很多研究致力于检测和诊断牵引系统中的这些故障^[9~12], 其中, 主要使用的方法可分为两种: 基于模型的方法和基于数据的方法. 众所周知, 基于模型的方法可以诊断未知故障, 但需要建立牵引系统的数学模型^[13~15]; 而基于数据的方法虽然不需要数学模型, 却需要大量的历史数据, 且通常无法解决未知故障的问题^[16~19]. 由于牵引系统具有多耦合、闭环、故障种类多, 以及传播复杂的特点, 因此难以建立完整的故障模型. 所以基于数据的故障诊断方法是一种较有效地处理牵引系统故障诊断的方法, 且可以解决一些未知故障.

牵引系统中的电气元件多退化性故障, 也称为微小渐变故障. 这类故障初期的时候对系统性能的影响较小, 不会立即导致系统损坏, 但会随着时间累积, 发展成系统性故障, 最终导致系统功能丧失. 由微小渐变故障产生的影响往往在传感器测量上产生的偏差很小, 且对其特征的提取也是一个难题. 在过去的研究中有一些针对微小渐变故障诊断问题的研究, 如文献 [20] 针对滚动轴承的微小渐变故障诊断, 提出了一种新的基于带宽经验模态分解和自适应多尺度模态分析的信号处理方案. 文献 [21] 提出了一种两层故障检测方法利用有限批次信息监测多批次信息中的微小渐变故障问题. 文献 [22] 采用机器学习的 3 种工具 (支持向量机、卷积神经网络和递归神经网络), 利用直流电机中最常见和最易测量的电流信号来对直流电机进行微小渐变故障的早期监测和识别. 文献 [23] 提出了一种深度主成分分析法以实现高速列车电气传动系统微小渐变故障的实时检测与诊断. 然而, 这些研究工作中都没有考虑到数据信号未知情况下的未知故障诊断问题. 针对未知故障, 如 TMF^[24], PCA^[25], HMM^[26] 等无监督学习方法也常被用于处理无标签数据, 可以引入到高速列车的未知微小渐变故障的诊断方案设计中, 进行微小渐变故障诊断方案的研究, 特别是针对未知故障的无监督故障诊断方法.

本文提出了一种高速列车牵引系统的微小渐变故障诊断方案, 首先通过引入加权的比例微分反馈改进了 LSTM (long short-term memory) 网络, 并结合 GRU (gated recurrent unit)^[27] 构造了一个自编码器模型, 以实现微小渐变故障特征的提取. 基于改进的自编码器模型^[28], 结合 t-SNE (t-distributed stochastic neighbour embedding)^[29] 和 DBSCAN (density-based spatial clustering of applications with noise)^[30], 设计了微小渐变故障的诊断方案, 并通过中车株洲电力机车有限公司的 CRH₂ 列车牵引系统半实物平台的数据^[10,31] 进行了实验验证.

2 问题描述

通常, 高速列车的电气牵引系统包含 5 个主要部分: 中间直流环节、整流器、牵引变压器、逆变器和牵引电机. 为了监测牵引系统的运行状况, 高速列车配备了多种传感器来采集电压、扭矩、速度和电流等信号, 如表 1 所示, 其采样频率为 2500 Hz.

牵引系统中故障主要为电容、电阻和传感器故障, 如表 2 所示, 本文主要对这些故障进行诊断研究. 其中, t_f 表示故障发生时间, C_f 表示故障 I 的电容值, C_h 表示健康状态电容值, α 表示电容的衰减指数; R_f 表示故障 II 的电阻值, R_h 是无故障时的电阻值, β 代表电阻的衰减指数; N_f 表示速度传感器的缺失脉冲数, γ 表示缺失参数; Ω_f 表示存在速度传感器漂移故障的测量转速, Ω_h 表示无故障情况的实际转速, λ 表示损失参数. 故障类型中包含微小渐变故障: 故障 I — 中间电容退化故障, 故障 II — 中间电阻退化故障, 故障 III — 速度传感器退化故障. 此外, 我们还考虑了微小突变故障, 故障 IV — 速度传感器偏置故障.

表 1 传感器信号
Table 1 The sensor signals

Symbol	Description	Unit
U_{net}	Transformer input voltage	V
$U_{d1,d2}$	Bridge arm voltage	V
T_{oq}	Motor torque	N·m
$I_{a,b,c}$	A current of the three-phase input current of the motor	A
ω_m	Motor speed	rpm

表 2 4 种微小故障
Table 2 The considered four kinds of faults

Category	Typical description	Parameter	Fault type
Health	Healthy data	None	Health
Fault I	$C_f = C_h e^{-\alpha t_f}$	$\alpha = 0.01, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.5$	Incipient time-varying fault
Fault II	$R_f = R_h e^{-\beta t_f}$	$\beta = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$	Incipient time-varying fault
Fault III	$N_f = \gamma \log(t_f + 1)$	$\gamma = 1, 3, 5, 7, 10, 13, 15, 17$	Incipient time-varying fault
Fault IV	$\Omega_f = \Omega_h(1 + \lambda)$	$\lambda = 0.005, 0.01, 0.02, 0.05, 0.1, 0.2$	Abrupt fault

本文用矩阵 $X \in \mathbb{R}^{N_v \times L}$ 来表示传感器信号:

$$X = \{x_1, x_2, \dots, x_{L-1}, x_L\}, \quad (1)$$

$$x_i = \{I_n, U_{\text{net}}, U_{d1}, U_{d2}, T_{\text{oq}}, I_a, I_b, I_c, W_r\}^T, \quad x_i \in X, \quad i = 1, 2, \dots, L, \quad (2)$$

其中, N_v 表示数据集的维数, L 表示时序信号的长度, 其中每个信号数据 x_i 包含 8 个监测的传感器信号. 这些传感器时序信号 X 将用来诊断表 2 中的牵引系统微小渐变故障, 考虑到未知故障, 我们将提出一种具有无监督诊断能力和高泛化性的故障诊断方案.

3 改进 LSTM 单元

从原始数据中提取特征是基于数据的故障诊断中的关键步骤, 该步骤决定了故障诊断的准确性. 微小渐变故障是缓时变的, 因此需要使用基于时序的方法进行特征提取.

传统的循环神经网络 (recurrent neural network, RNN) 作为深度学习的一种, 具有在时序数据中挖掘信息的能力. 然而 RNN 存在长期依赖的问题, 难以提取微小渐变故障的特征. 为解决上述问题, Hochriter 等^[32]提出了一种由多层 LSTM 单元构成的 RNN 网络, 即长短期记忆网络 LSTM. LSTM 用记忆存储单元替代了传统 RNN 的隐藏层, 这允许单元利用上下文信息在时间序列上构造一个鲁棒性更强的模型. 尽管 LSTM 网络对非平稳特征具有鲁棒性, 但它难以有效地从缓时变的数据信号和不稳定的故障变量中提取特征, 这促使我们对其网络结构进行改进.

3.1 引入比例微分反馈的 LSTM 单元

传统的 LSTM 单元^[33]可以看作是一个离散的非线性时变动态系统, 其输入为时序信号 $\tilde{c}_t$, 输出为 h_t , 状态空间模型可表示为

$$c_t = G_{t|t-1}c_{t-1} + H_{t-1}\tilde{c}_t, \quad (3)$$

图 1 改进 LSTM 单元结构

Figure 1 The framework of improved LSTM cell

$$h_t = K_{t-1} \tanh(c_t), \quad (4)$$

其中, 下标 t 表示时间步长, 构造的参与状态变量 $\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c)$ 是控制器的输入, c_t 表示控制器的状态变量, h_t 表示控制器的输出. $G_{t|t-1}$, H_{t-1} 和 K_{t-1} 分别表示该动态系统的转换矩阵, 输入矩阵和输出矩阵.

由系统 (3) 和 (4) 的结构可以看出, 传统的 LSTM 的每个神经元中误差信号不直接传递到每个门, 当输出门关闭时, 只有单元输出是可观的, 其值接近于零, 一些有价值的信息可能因此丢失.

我们在当前时刻与前一个时刻的单元状态偏差中加入比例微分 (proportional derivative, PD) 反馈来改进 LSTM 网络, 使得误差在输出门关闭的情况下仍处于可观状态, 同时实现了用过去状态的变化量控制下一时刻的状态. 改进后的 LSTM 单元结构如图 1 所示.

3.2 前向传播

下面, 我们对改进 LSTM 单元内信号传播的过程进行分析. 由改进的 LSTM 单元结构图 1 可见, 每个 LSTM 单元包含 3 个门, 即输入门、输出门和遗忘门, 网络输入量为上一单元的输出 h_{t-1} 与当前时刻信号 x_t , x_t 为原始时序信号输入. 对于本文研究的 CRH₂ 型列车, 其第 t 步传感器信号即为式 (1) 中的向量 x_t , $t = 1, 2, \dots, L$. 网络状态量、反馈量和输出量初始值分别设置为 $c_0 = 0$, $p_0 = 0$ 和 $h_0 = 0$.

改进的 LSTM 单元的遗忘门可以看作一个自连接的线性单元, 遗忘门决定了单元应当删除的信息内容, 称之为记忆存储模块中的常量误差传输带 (constant error carousel, CEC). 过去信息在输入当前单元时首先会被存储至 CEC 中, 由遗忘门选择判断对过去信息进行保留或遗忘, 然后用逻辑 $\sigma(\cdot)$ 函数压缩到 $[0, 1]$, 遗忘门的激活函数 $f_t(\cdot)$ 表达式为

$$\begin{aligned} f_t(\cdot) &= \sigma(\hat{f}_t) = \sigma(W_f \cdot [h_{t-1}, x_t, c_{t-1}, p_{t-1}] + b_f) \\ &= \sigma(W_{hf}h_{t-1} + W_{xf}x_t + W_{cf}c_{t-1} + W_{pf}p_{t-1} + b_f), \end{aligned} \quad (5)$$

其中,

$$p_{t-1} = c_{t-1} - c_{t-2} \quad (6)$$

是 $t-1$ 时刻的状态向量 c_{t-1} 与前一时刻的状态量 c_{t-2} 的误差反馈,也就是我们引入的比例微分反馈的误差项.此外,式(5)中下标 t 表示时间步长, W_{ij} 表示权重矩阵, $i \in \{h, x, c, p\}$, $j \in \{c, o, r, f\}$, b_k 为传播偏置量, $k \in \{c, o, r, f\}$. $f_t(\cdot)$ 是内部状态量 c_t 的自循环连接的权重, $f_t(\cdot)$ 越接近 0 时,表示需要删除或只保留较少的过去状态量 c_{t-1} . 在状态初始值 $c_0 = 0$ 的初始条件下,遗忘门激活函数 $f_t(\cdot)$ 在训练开始时非常接近于 1.0, c_t 约等于上一时间步的状态量 c_{t-1} . 换言之,当误差信号抵达存储记忆单元时,尽管输入门的权重仍在改变,但这些误差将不再向前传播.作为存储记忆单元中 CEC 的输入,当遗忘门关闭时,这些误差可以在其中循环传播而不发生改变.这意味着 LSTM 网络可以在遗忘门被激活前保留任意的内部信息,实现在网络中保留有效信息的作用.

LSTM 的输入门决定了单元如何处理新增信息.一方面,输入门选择需要保留的新增信息,同时创建一个新的候选参与量 $\tilde{c}_t$,并使用 $\tanh(\cdot)$ 函数压缩,来与前一步状态量 c_{t-1} 共同更新当前单元状态量 c_t . 这里的输入门激活函数 $r_t(\cdot)$ 与候选量 $\tilde{c}_t$ 构建为

$$\begin{aligned} r_t(\cdot) &= \sigma(\hat{r}_t) = \sigma(W_r \cdot [h_{t-1}, x_t, c_{t-1}, p_{t-1}] + b_r) \\ &= \sigma(W_{hr}h_{t-1} + W_{xr}x_t + W_{cr}c_{t-1} + W_{pr}p_{t-1} + b_r), \end{aligned} \quad (7)$$

$$\begin{aligned} \tilde{c}_t &= \tanh(\hat{c}_t) = \tanh(W_c \cdot [h_{t-1}, x_t, p_{t-1}] + b_c) \\ &= \tanh(W_{hc}h_{t-1} + W_{xc}x_t + W_{pc}p_{t-1} + b_c), \end{aligned} \quad (8)$$

其中,各参数同(5)中定义. W_{ij} 表示权重矩阵, $i \in \{h, x, c, p\}$, $j \in \{c, r\}$, b_c 和 b_r 为传播偏置量.

经过遗忘门与输入门处理的信息汇总后相加就得到存储单元的内部状态量 c_t :

$$c_t = f_t \odot c_{t-1} + r_t \odot \tilde{c}_t, \quad (9)$$

其中, $\odot$ 表示 Hadamard 乘积.

LSTM 单元的输出门基于单元状态信息 c_t 决定该 LSTM 单元的输出信息,输入当前 x_t 与上一时刻的输出 h_{t-1} 进行预测后用逻辑 $\sigma(\cdot)$ 函数压缩,然后与被 $\tanh(\cdot)$ 压缩至 $[-1, 1]$ 的当前状态量 c_t 相乘,得到包含所需信息的输出,从而只输出想要输出的部分,即

$$\begin{aligned} o_t(\cdot) &= \sigma(\hat{o}_t) = \sigma(W_o \cdot [h_{t-1}, x_t, c_t, p_t] + b_o) \\ &= \sigma(W_{ho}h_{t-1} + W_{xo}x_t + W_{co}c_t + W_{po}p_t + b_o), \end{aligned} \quad (10)$$

$$h_t = o_t \odot \tanh(c_t), \quad (11)$$

其中,各参数同(5)中定义. W_{io} 表示权重矩阵, $i \in \{h, x, c, p\}$, b_o 为传播偏置量.此时,当前输出 h_t 是一个基于当前单元状态量 c_t 的滤波信号且其状态可观. h_t 与下一步时刻的信号 x_{t+1} 作为下一步单元遗忘门的输入,实现了 $t+1$ 时间步的单元状态向量预测,以此类推,构成了完整的改进 LSTM 网络.

根据式(5)~(10)可见,各个门激活函数都可以看作各自输入量的权重矩阵.前向传播通过在通路结构的各个门赋权值与偏差量对原始输入信号进行重编码,提取到含有更多有效信息的特征向量.然而,各个门的权值在初始时并非最佳权值,故需要设计反向传播实现最优权重参数的更新,后文将介绍改进 LSTM 网络的反向传播设计.

3.3 反向传播

由于网络初始权重参数为随机值,单次的前向通路传播往往不能直接得到理想输出,故需要使用反向传播来更新权值矩阵 W_{ij} , $i \in \{h, x, c, p\}$, $j \in \{c, o, r, f\}$. 当网络前向传播的输出未得到期望输出

值时, 我们通过反向传播, 计算设定的损失函数对各神经元权值的偏导梯度来更新权值, 学习过程在误差趋于平缓时结束. 在改进的 LSTM 网络中, 全部隐含层输出为 $H = \{h_1, h_2, \dots, h_{L-1}, h_L\}$, 其中 h_t 为第 t 步输出, $t = 1, 2, \dots, L$. 设最终输出结果是由 H 得到的长度为 m 的 $g(H)$, $g(\cdot)$ 为将隐含层输出转换为最终输出的非线性函数, D 表示 $g(H)$ 对应的期望输出, 其长度也为 m . 我们设置损失函数 $J = J(D, g(H))$ 来表示模型输出值与期望值的一致程度, 其中 $J(\cdot)$ 表示损失函数的计算方式, 对于不同的分类或回归需求可以选择不同的函数. 损失函数 J 越小, 表示模型输出 $g(H)$ 越接近期望输出 D . J 采用随机梯度下降法来实现最小化.

定义 $\delta h_t = \frac{\partial J}{\partial h_t}$, $\delta c_t = \frac{\partial J}{\partial c_t}$. 根据式 (11) 中前向传播的输出 h_t , 可以计算出损失函数对各参数的梯度量 δo_t , δc_t , δr_t , $\delta \tilde{c}_t$, δf_t , δp_t 和 δc_{t-1} :

$$\delta o_t = \frac{\partial J}{\partial o_t} = \frac{\partial J}{\partial h_t} \cdot \frac{\partial h_t}{\partial o_t} = \delta h_t \cdot \tanh(c_t), \quad (12)$$

$$\delta c_t = \delta c_{t-1} + \delta h_t \odot \delta o_t \odot (1 - \tanh^2(c_t)), \quad (13)$$

$$\delta r_t = \frac{\partial J}{\partial r_t} = \frac{\partial J}{\partial c_t} \cdot \frac{\partial c_t}{\partial r_t} = \delta c_t \odot \tilde{c}_t, \quad (14)$$

$$\delta f_t = \delta c_t \odot c_{t-1}, \quad (15)$$

$$\delta p_t = \delta c_t \odot \frac{\partial c_t}{\partial p_t}, \quad (16)$$

$$\delta \tilde{c}_t = \delta c_t \odot r_t, \quad (17)$$

$$\delta c_{t-1} = \delta c_t \odot f_t. \quad (18)$$

根据式 (12)~(18), 可以得到当前状态量的梯度值 δc_t :

$$\delta c_t = \delta h_t \odot o_t \odot (1 - \tanh^2(c_t)) + \delta c_{t+1} \odot f_{t+1}. \quad (19)$$

在获得当前状态量的梯度值 δc_t 后, 根据随机梯度下降法, 权重矩阵 W_{ct} 可以计算为

$$\delta W_{ct} = \frac{\partial J}{\partial \tilde{c}_t} \cdot \frac{\partial \tilde{c}_t}{\partial W_{ct}}, \quad (20)$$

其中各权值参数为

$$\delta W_{cij} = \delta \hat{c}_{it} \cdot \frac{\partial \hat{c}_{it}}{\partial W_{cijt}} = \delta \hat{c}_{it} \cdot x_{jt}. \quad (21)$$

根据式 (20) 和 (21) 可以得到

$$\delta W_{ch} = \begin{bmatrix} \delta \hat{c}_{1t} x_{1t} & \delta \hat{c}_{1t} x_{2t} & \cdots & \delta \hat{c}_{1t} x_{nt} \\ \delta \hat{c}_{2t} x_{1t} & \delta \hat{c}_{2t} x_{2t} & \cdots & \delta \hat{c}_{2t} x_{nt} \\ \vdots & \vdots & & \vdots \\ \delta \hat{c}_{dt} x_{1t} & \delta \hat{c}_{dt} x_{2t} & \cdots & \delta \hat{c}_{dt} x_{nt} \end{bmatrix} = \delta \hat{c} \odot (x_t)^T. \quad (22)$$

其他权重矩阵 W_{ij} 的梯度参数可以同 W_{ch} 计算, 其中, $i \in \{h, x, c, p\}$ 且 $j \in \{c, o, r, f\}$. 如果输入的数据信号 $X = \{x_1, x_2, \dots, x_{L-1}, x_L\}$ 有 L 个时间步, 那么其所有的贡献累加得到的权重矩阵可以计算为

$$W_{ij} = W_{ij0} + \sum_{t=1}^L \delta W_{ijt}, \quad (23)$$

其中 W_{ij0} 表示各权重矩阵的随机初始值, 上述权重矩阵通过随机梯度下降法进行更新, 当 W 值趋于平缓时迭代结束, 可以认为此时损失函数 J 取得最小值, 网络输出 $h_k(t)$ 最为接近目标值 $d_k(t)$, 改进的 LSTM 网络完成了训练.

在改进后的 LSTM 网络中, 基于过去单元状态变化的反馈使得该网络对于微小渐变故障信号特征更加敏感, 编码后提取的特征信息可以解码为原始信号, 包含了全部有效信息, 具有微小渐变故障的特征.

4 无监督微小渐变故障诊断

有监督的故障诊断流程通常包含 3 个步骤: 特征提取、特征向量降维和故障分类. 然而有监督学习模型对于新增的未知故障缺乏泛化能力, 未知故障会被分类到相似度最高的已知故障内. 考虑到系统运行时会产生一些无标签的数据, 且运行过程中可能会发生未经训练的新增故障, 这就要求我们设计一种无监督的故障诊断方案, 并用无标签的聚类代替有标签的分类. 在无监督诊断方案中, 网络无法通过标签信息得到损失函数来完成训练, 因此, 我们引入了自编码器模型以解决“无导师的反向传播”问题^[34]. 微小渐变故障的特征提取需要对非平稳的缓时变故障信号敏感, 第 3 节中改进的 LSTM 网络将被用于自编码器的编码部分; 对于自编码器的解码器, 仅需要对于编码效果进行验证, 因此我们选择了一种基于 LSTM 合并了遗忘门和输入门以减少参数计算的 GRU 网络来提高效率. 综上所述, 该无监督微小渐变故障的诊断方案, 如图 2 所示, 采用基于改进 LSTM 网络与 GRU 网络的自编码器进行特征提取后, 对特征进行降维与聚类.

4.1 基于自编码器的微小渐变故障特征提取

在自编码器框架中, 输入为传感器数据信号 $X = \{x_1, x_2, \dots, x_{L-1}, x_L\} \in \mathbb{R}^{N_v \times L}$, 其中, N_v 表示向量的维度, L 表示该时序信号的长度. 在 t 时刻, 编码器网络将当前输入 x_t 与前一时刻的单元输出 h_{t-1} 通过函数 $F(\cdot)$ 得到该时刻的单元输出 h_t , 并通过非线性函数 $q(\cdot)$ 将所有时刻单元的输出编码至最终的特征向量 c_{t_L} 中, 即

$$h_t = F(h_{t-1}, x_t), \quad (24)$$

$$c_t = q(h_1, h_2, \dots, h_L), \quad (25)$$

其中, $h_t \in \mathbb{R}^n$ 是 t 时刻的隐藏状态, c_t 是由该隐含层生成的特征向量, $F(\cdot)$ 和 $q(\cdot)$ 是自编码器中的非线性转换函数. 我们使用 LSTM 网络 (5)~(11) 作为 $F(\cdot)$ 函数, 编码器基于状态向量 c_t 和过去的编码器输出 $\{y_1, y_2, \dots, y_{t-1}\}$ 作为 y_t 的状态预测, 输出预测概率最大的 y_t , 表示为

$$y_t = \operatorname{argmax} P(y) = \prod_{t=0}^T p(y_t | \{y_1, \dots, y_{t-1}\}, c_t), \quad (26)$$

进一步, GRU 网络解码器中, 预测的输出 y_t 可以被表示为

$$y_t = g(\{y_1, \dots, y_{t-1}\}, c) = g(y_{t-1}, h'_t, c_t), \quad (27)$$

其中, $g(\cdot)$ 是关于输出最大预测概率的 y_t 的非线性函数, h'_t 是 GRU 网络的隐含层向量. 如果该自编码器模型的输出向量 $y_1, y_2, \dots, y_t$ 非常接近于输入向量, 则说明该提取到的特征向量 c_t 已包含了原始数据中的最有效的特征, 输入信号 $X = \{x_1, x_2, \dots, x_{L-1}, x_L\}$ 时, 改进 LSTM 编码器的最后一层单元

图 2 (网络版彩图) 无监督微小渐变故障诊断方案的总体框架

Figure 2 (Color online) The overall framework of the unsupervised incipient fault diagnosis scheme

状态为 c_{t_L} . GRU 解码器输出时序信号 $\hat{X} = \{\hat{x}_L, \hat{x}_{L-1}, \dots, \hat{x}_2, \hat{x}_1\} \in \mathbb{R}^{N_v \times L}$, 则自编码器模型的损失函数表达为

$$J_{\text{ed}} = \sum_{X \in D_N} \sum_{t=1}^L \|x_t - \hat{x}_t\|^2, \quad (28)$$

其中 D_N 表示模型的训练集, 该自编码器模型结构与训练过程如图 3 所示. 根据反向传播更新规则 (式 (12)~(23)) 对自编码器特征提取模型进行迭代训练. 经过改进 LSTM 网络特征提取得到的特征向量 c_{t_L} 将被用于后续过程并实现故障诊断.

4.2 降维设计

由于从原始数据信号中得到的特征向量 c_{t_L} 具有高维的特点^[35], 对于高维特征向量, 常采用的非线性降维方法包括 t-SNE, Isomap, kernel PCA 等算法^[29,36,37], 本实验中, 我们将应用 t-SNE 降低提取到的故障特征的维数, 消除冗余信息. t-SNE 法是一种非线性降维方法, 基于重尾 t 分布, 可以提高降维的鲁棒性, 缓解分布过于密集的问题^[29]. 损失函数 J_{SNE} 是通过最小化高维空间中的联合概率分布 P_D 和低维空间中的联合概率分布 Q_D 之间的 Kullback-Leibler ($\mathcal{KL}$) 散度来确定的, 其表达式为

$$J_{\text{SNE}} = \mathcal{KL}(P_D || Q_D) = \sum_i \sum_j p_{ij} \ln \frac{p_{ij}}{q_{ij}}, \quad (29)$$

图 3 (网络版彩图) 改进 LSTM 自编码器模型结构

Figure 3 (Color online) The framework of improved LSTM encoder-decoder model

$$p_{ij} = \frac{\exp(-\|c_{tLi} - c_{tLj}\|^2 / 2\sigma_i^2)}{\sum_{k \neq l} \exp(-\|c_{tLk} - c_{tLl}\|^2 / 2\sigma_i^2)}, \quad (30)$$

$$q_{ij} = \frac{(1 + \|z_i - z_j\|^2)^{-1}}{\sum_{k \neq l} (1 + \|z_k - z_l\|^2)^{-1}}, \quad (31)$$

其中下标 i, j, k 和 l 分别表示第 i, j, k 和 l 个特征向量. $c_{tLi}, c_{tLj}, c_{tLk}$ 和 c_{tLl} 表示高维的原始特征向量, z_i 和 z_j 表示经过降维处理的特征向量, σ_i 表示与特征向量 c_{tLi} 和 c_{tLj} 相关的高斯核宽度, p_{ij} 表示 c_{tLi} 与 c_{tLj} 间的相似度, q_{ij} 表示 z_i 和 z_j 间的相似度, $\mathcal{KL}$ 散度描述了原始特征向量分布与降维后特征向量分布的相似度. 模型训练采用随机梯度下降以最小化 $\mathcal{KL}$ 散度:

$$\frac{\delta J_{\text{SNE}}}{\delta z_i} = 4 \sum_j (p_{ij} - q_{ij})(z_i - z_j)(1 + \|z_i - z_j\|^2)^{-1}. \quad (32)$$

降维后, 特征向量 c_{tL} 的分布变化需要尽可能地小, 从而得到新的特征向量 z 用于聚类. 当损失函数下降到允许范围内时, 降维模型训练完毕, 可以在保证样本的相似性和分布特性的同时达到降维的目的.

4.3 聚类设计

在微小渐变故障诊断方案中, 我们采用改进的 LSTM 编码器提取特征向量 c_{tL} , 高维的特征向量 c_{tL} 经过 t-SNE 降维得到新的特征向量 z , 通过聚类将特征向量 z 识别为相应的故障模式, 实现无监督的微小渐变故障诊断. 我们采用了一种非线性的无监督聚类方法 DBSCAN 来对提取并降维后的特征向量 z 进行聚类以实现无监督的故障诊断.

综上所述, 本文中改进的 LSTM 自编码器模型的特征提取能力主要体现在对无标签的微小渐变故障时序信号的处理上. 构建基于改进的 LSTM 自编码器模型的微小渐变故障的诊断方案的算法, 如算法 1 所示.

Algorithm 1 The improved LSTM-based auto-encoder algorithm

- 1: Process and normalize the initial data signals to push their values between 0 and 1 as $X = \{x_1, x_2, \dots, x_{L-1}, x_L\}$;
 - 2: **Repeat**
 - 3: Input the data signal $X = \{x_1, x_2, \dots, x_{L-1}, x_L\}$ to train the improved LSTM encoder;
 - 4: Extract the cell state C_t from the improved LSTM encoder;
 - 5: Input cell state C_t to the GRU decoder and output the predicted state $\hat{X} = \{\hat{x}_L, \hat{x}_{L-1}, \dots, \hat{x}_2, \hat{x}_1\}$;
 - 6: Calculate the object function through the error between X and $\hat{X}$;
 - 7: Update parameters W with SGD (stochastic gradient descent);
 - 8: Update the cell state C_t as the trained feature vectors;
 - 9: **Until** Convergence of parameters W ;
 - 10: **Return** W and obtain the final C_t ;
 - 11: **Dimension reduction**
 - 12: Utilize the t-SNE approach on C_t to get a low dimensional vector z ;
 - 13: **Clustering**
 - 14: Apply the DBSCAN approach to cluster the low-dimensional feature vectors to achieve the incipient fault diagnosis.
-

5 实验验证

为了验证本文所提出的微小渐变故障诊断方案的有效性, 将用在 CRH₂ 列车牵引系统半实物平台上产生的数据^[10, 31]进行验证. 实验平台包含信号调节器 (signal conditioner)、上位机 (PC)、司控室 (control room)、牵引控制单元 (drive control unit, DCU) 和实时仿真器 (dSPACE real-time simulator) 等.

基于实验平台的仿真信号中 (表 2), 将健康、故障 I 和故障 II 的时序数据信号分割为训练集和测试集, 其中时序信号被排列为向量, 训练集部分被用于训练改进的 LSTM 自编码器. 在收集的约 14000 个时序数据信号样本中, 随机选择其中的 60% (即 8500 个样本) 用于训练改进的 LSTM 自编码器, 其中包含 1180 个健康样本、990 个故障 I 样本和 935 个故障 II 样本. 故障 III 和 IV 的时序数据信号仅作为测试集, 故障 III 和 IV 的样本数分别为 1175 和 1175. 每个样本的长度范围是 20~40, 即时序信号样本的长度 L 在 $[20, 40]$ 之间.

所设计的自编码器共有 4 个超参数需要选择, 分别是学习率 (learning_rate)、隐含层大小 (hidden_size)、训练轮数 (epochs) 和教师强制率 (teacher_forcing_ratio). 根据经验规律, 我们将其分别设为 0.01, 64, 训练样本的长度 (len(trainData)) 和 0.5.

5.1 改进的 LSTM 自编码器效果

使用训练集将改进 LSTM 编码器的训练结果与 LSTM 和 GRU 编码器的训练结果进行了比较, 以体现本文所提出的改进的 LSTM 自编码器的性能. 对这些编码器的损失函数 J_{ed} 进行了比较, 结果如图 4 所示, 分别展示了全部 8500 个训练样本的前 1500 轮和后 500 轮的结果. 表 3 显示了训练过程中损失函数 J_{ed} 的最小均值和方差. 可见, 改进的 LSTM 编码器的损失函数 J_{ed} 和方差在图 4 和表 3 中均为最小. 由于损失函数 J_{ed} 衡量了输入的时序信号和解码器重构的输出时序信号之间的相似性,

图 4 (网络版彩图) 基于改进的 LSTM, LSTM 和 GRU 的自编码器损失函数 J_{ed}

Figure 4 (Color online) The loss J_{ed} of the encoders based on improved LSTM, LSTM, and GRU. (a) First 1500 iterations; (b) last 500 iterations

表 3 各编码器中损失的均值和方差

Table 3 The mean and variance of the loss in the encoders

Model	Mean	Variance
GRU encoder-decoder model	0.0513	0.00007
LSTM encoder-decoder model	0.0175	0.00002
Improved LSTM encoder-decoder model	0.0062	0.00006

图 5 (网络版彩图) 降维结果

Figure 5 (Color online) The dimensionality reduction results. (a) t-SNE; (b) PCA; (c) Isomap

因此, 改进的 LSTM 网络结构能够提取输入时序信号的更多有效信息, 也就是, 改进的 LSTM 自编码器在微小渐变故障的特征提取方面比原始 LSTM 和 GRU 具有更好的性能.

5.2 微小故障诊断结果

经过改进 LSTM 网络提取出了特征, 对其使用 t-SNE 和 DBSCAN 进行降维和聚类, 以实现无监督的微小渐变故障诊断. 从故障 I 和 II 的测试集中提取的特征向量将用于验证模型对已知故障的诊断能力, 而从故障 III 和 IV 的测试集中提取的特征向量将用于验证模型对未知故障检测的泛化能力. 图 5 中分别显示了 t-SNE(a), kernel PCA(b) 和 Isomap(c) 的降维结果, 其中蓝点表示健康样本, 红点表示故障 I 的样本, 绿点表示故障 II 的样本. 显然, t-SNE 可以得到更好的降维效果.

进一步, 将由 t-SNE 得到的降维特征通过 DBSCAN 聚类, 其结果如图 6 所示, 也就是对应故障诊断的结果. 从图 5 和 6 中可以看出, 来自健康数据信号的样本和来自其他故障类别的一些样本被聚类

图 6 (网络版彩图) t-SNE 聚类结果

Figure 6 (Color online) The clustering results with t-SNE

图 7 (网络版彩图) 健康、故障 I、故障 II 样本测试集的混淆矩阵

Figure 7 (Color online) The confusion matrix for the test sets of health, fault I and II cases

至同一类别 (红色), 部分来自故障 II 的样本和大部分来自故障 I 的样本被聚类至同一类别 (米黄色), 大部分来自故障 II 的样本被聚类至同一类别 (紫色), 也就是大部分来自不同类别的样本被清晰地分开了。

为了量化故障诊断效果, 我们引入准确率、虚警率和漏警概率作为性能指标, 其定义为

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}}, \quad (33)$$

$$\text{FA} = \frac{\text{FP}}{\text{TP} + \text{FP}}, \quad \text{MA} = \frac{\text{FN}}{\text{TP} + \text{FN}}, \quad (34)$$

其中 Acc, FA 和 MA 分别表示准确率、虚警率和漏警概率, TP 表示本来是正样例, 分类成正样例; TN 表示本来是负样例, 分类成负样例; FP 表示本来是负样例, 分类成正样例; FN 表示本来是正样例, 分类成负样例。

基于以上性能指标, 生成经过聚类的降维后特征向量的混淆矩阵, 如图 7 所示, 用于计算准确率、虚警率和漏警概率, 得到针对已知故障 (健康、故障 I、故障 II 类) 的故障诊断结果, 如表 4 所示。可见, 健康样本的准确率为 100%, 故障 I 的准确率为 85.86%, 其中误分类的样本都被聚类到健康类别中; 故障 II 的准确率为 95.62%, 其中只有一小部分样本被误分类并聚类到健康和故障 I 类别中。尽管故障 I 样本的特征向量与健康样本的特征向量在聚类中具有很高的相似性, 但由于故障 I 的幅值很小, 故障诊断的准确率在工程上是可以接受的。

5.3 未知故障泛化性验证

上述实验结果表明了我们所提出的故障诊断方案能够处理已知的故障。这里, 我们将验证该无监督微小渐变故障诊断方案是否能够实现对于未知故障的识别与诊断, 这意味着来自这些故障的数据信号是未在模型中训练过的。故障 III 和 IV 的测试集样本没有标签, 也未事先经过训练, 以用于验证所提

表 4 已知故障的故障诊断结果 (健康、故障 I、故障 II)

Table 4 The fault diagnosis scheme for the known faults (health, fault I, fault II cases)

Category	Number	Testing accuracy (%)	False alarm rate (%)	Missed alarm rate (%)
Health	1480	100	0.00	0.00
Fault I	990	85.86	0.00	14.14
Fault II	935	95.62	0.21	4.17

图 8 (网络版彩图) 未知故障 (健康、故障 I~IV 样本) 的聚类结果

Figure 8 (Color online) The clustering results for the unknown faults (health, fault I~IV cases)

图 9 (网络版彩图) 健康、故障 I~IV 样本测试集的混淆矩阵

Figure 9 (Color online) The confusion matrix for the test sets of health and fault I~IV cases

出的微小渐变故障诊断方案对于未知故障的泛化性能. 其中, 故障 III 是微小渐变故障, 故障 IV 是微小突变故障.

我们将无标签的故障 I~IV 的信号样本和健康信号样本作为测试集输入至训练好的改进 LSTM 故障诊断模型中, 聚类结果如图 8 所示, 其中黄色点代表故障 III, 绿色点代表故障 IV. 可以看出, 故障 III 和 IV 的样本被分别聚类为两个不同的类别, 并未出现与其他 3 个类别混合的情况. 因此, 训练后的改进 LSTM 自编码器模型对未知故障具有良好的诊断性能. 图 9 显示了健康样本和故障 I~IV 样本集的混淆矩阵, 我们使用式 (33) 和 (34) 来计算准确率、虚警率和漏警概率, 则故障诊断性能如表 5 所示, 其中健康样本的诊断准确度为 100%, 故障 III 和 IV 的准确率分别为 98.04% 和 100%, 故障 III 的样本中只有一小部分被归入健康类别. 实验结果表明, 所提出的早期故障诊断方案能够识别已知和未知故障, 且故障检测准确率达到 95% 以上, 对未知故障识别的准确率也达到了工程要求.

6 小结

本文主要研究高速列车牵引系统微小渐变故障的无监督诊断方案. 提出了一种改进的 LSTM 自编码器结构, 可以从原始信号中提取特征向量, 该特征向量可以表示微小渐变故障的特征. 进一步地,

表 5 未知故障的故障诊断结果 (健康、故障 I~IV)
Table 5 The fault diagnosis results for the unknown faults (health, fault I~IV cases)

Category	Number	Testing accuracy (%)	False alarm rate (%)	Missed alarm rate (%)
Health	1480	100	0.00	0.00
Fault I	990	85.86	0.00	14.14
Fault II	935	95.62	0.21	4.17
Fault III	1175	98.04	0.00	1.96
Fault IV	1175	100	0.00	0.00

我们通过 t-SNE 法对提取的特征向量进行降维, 并将低维特征向量聚类至不同的类别, 以实现无监督的微小渐变故障诊断. 最后, 我们通过中车株洲电力机车有限公司的 CRH₂ 列车牵引系统半实物平台的数据验证了该方案的有效性, 结果表明, 该无监督故障诊断方案能够实现对列车微小渐变故障的初期预警.

参考文献

1 Ning B, Dong H R, Gao S G, et al. Distributed cooperative control of multiple high-speed trains under a moving block system by nonlinear mapping-based feedback. *Sci China Inf Sci*, 2018, 61: 120202

2 Ning B, Dong H R, Zheng W, et al. Integration of train control and online rescheduling for high-speed railways: challenges and future. *Acta Autom Sin*, 2019, 45: 2208–2217 [宁滨, 董海荣, 郑伟, 等. 高速铁路运行控制与动态调度一体化的现状与展望. *自动化学报*, 2019, 45: 2208–2217]

3 Gao S G, Dong H R, Ning B. Characteristic model-based all-coefficient adaptive control for automatic train control systems. *Sci China Inf Sci*, 2014, 57: 092201

4 Ji H H, Hou Z, Zhang R. Adaptive iterative learning control for high-speed trains with unknown speed delays and input saturations. *IEEE Trans Autom Sci Eng*, 2016, 13: 260–273

5 Yang C, Peng T, Tao H W, et al. Review of recent research on fault injection for high-speed train information control systems. *Sci Sin Inform*, 2020, 50: 465–482 [杨超, 彭涛, 陶宏伟, 等. 高速列车信息控制系统故障注入研究进展. *中国科学: 信息科学*, 2020, 50: 465–482]

6 Song Y D, Song Q, Cai W C. Fault-tolerant adaptive control of high-speed trains under traction/braking failures: a virtual parameter-based approach. *IEEE Trans Intell Transp Syst*, 2014, 15: 737–748

7 Guzinski J, Diguët M, Krzeminski Z, et al. Application of speed and load torque observers in high-speed train drive for diagnostic purposes. *IEEE Trans Ind Electron*, 2009, 56: 248–256

8 Jiang B, Wu Y K, Lu N Y, et al. Review of fault diagnosis and prognosis techniques for high-speed railway traction system. *Control Dec*, 2018, 33: 841–855 [姜斌, 吴云凯, 陆宁云, 等. 高速列车牵引系统故障诊断与预测技术综述. *控制与决策*, 2018, 33: 841–855]

9 Wang Y J, Song Y D, Gao H, et al. Distributed fault-tolerant control of virtually and physically interconnected systems with application to high-speed trains under traction/braking failures. *IEEE Trans Intell Transp Syst*, 2016, 17: 535–545

10 Yang X Y, Yang C H, Peng T, et al. Hardware-in-the-loop fault injection for traction control system. *IEEE J Emerg Sel Top Power Electron*, 2018, 6: 696–706

11 Dong H R, Lin X, Gao S G, et al. Neural networks-based sliding mode fault-tolerant control for high-speed trains with bounded parameters and actuator faults. *IEEE Trans Veh Technol*, 2020, 69: 1353–1362

12 Jiang B, Chen H T, Yi H, et al. Data-driven fault diagnosis for dynamic traction systems in high-speed trains. *Sci Sin Inform*, 2020, 50: 496–510 [姜斌, 陈宏田, 易辉, 等. 数据驱动高速列车动态牵引系统的故障诊断. *中国科学: 信息科学*, 2020, 50: 496–510]

13 Zhou D H, Ji H Q, He X, et al. Fault detection and isolation of the brake cylinder system for electric multiple units. *IEEE Trans Control Syst Technol*, 2018, 26: 1744–1757

14 Yin J T, Xie Y F, Chen Z W, et al. Fault tracing method based on fault propagation and causality with its application to the traction drive control system. *Acta Autom Sin*, 2020, 46: 47–57 [尹进田, 谢永芳, 陈志文, 等. 基于故障传播与因果关系的故障溯源方法及其在牵引传动控制系统中的应用. *自动化学报*, 2020, 46: 47–57]

15 Mao Z H, Tao G, Jiang B, et al. Adaptive actuator compensation of position tracking for high-speed trains with

- disturbances. *IEEE Trans Veh Technol*, 2018, 67: 5706–5717
- 16 Song Q, Song Y D. Data-based fault-tolerant control of high-speed trains with traction/braking notch nonlinearities and actuator failures. *IEEE Trans Neural Netw*, 2011, 22: 2250–2261
 - 17 Cheng C, Qiao X Y, Teng W X, et al. Principal component analysis and belief-rule-base aided health monitoring method for running gears of high-speed train. *Sci China Inf Sci*, 2020, 63: 199202
 - 18 Hu H X, Tang B, Gong X J, et al. Intelligent fault diagnosis of the high-speed train with big data based on deep neural networks. *IEEE Trans Ind Inf*, 2017, 13: 2106–2116
 - 19 Zhao L, Huang D R, Wang H G. Analysis of track irregularity based on the improved hilbert-huang transform. *Control Eng China*, 2019, 26: 1539–1543 [赵玲, 黄大荣, 王宏刚. 改进 Hilbert-Huang 变换在轨道不平顺中的应用. *控制工程*, 2019, 26: 1539–1543]
 - 20 Li Y B, Xu M Q, Liang X H, et al. Application of bandwidth EMD and adaptive multiscale morphology analysis for incipient fault diagnosis of rolling bearings. *IEEE Trans Ind Electron*, 2017, 64: 6506–6517
 - 21 Zhang S M, Zhao C H, Gao F R. Incipient fault detection for multiphase batch processes with limited batches. *IEEE Trans Contr Syst Technol*, 2019, 27: 103–117
 - 22 Munikoti S, Das L, Natarajan B, et al. Data-driven approaches for diagnosis of incipient faults in DC motors. *IEEE Trans Ind Inf*, 2019, 15: 5299–5308
 - 23 Chen H T, Jiang B, Lu N Y, et al. Deep PCA based real-time incipient fault detection and diagnosis methodology for electrical drive in high-speed trains. *IEEE Trans Veh Technol*, 2018, 67: 4819–4830
 - 24 Wu Y, Wang X, Xiao P, et al. Fast algorithm based on triplet Markov fields for unsupervised multi-class segmentation of SAR images. *Sci China Inf Sci*, 2011, 54: 1524–1533
 - 25 Guo W, Dai L R, Wang R H. Feature mapping based on PCA. *Acta Autom Sin*, 2008, 34: 876–879 [郭武, 戴礼荣, 王仁华. 采用主成分分析的特征映射. *自动化学报*, 2008, 34: 876–879]
 - 26 Zhu J L, Ge Z Q, Song Z H. HMM-driven robust probabilistic principal component analyzer for dynamic process fault classification. *IEEE Trans Ind Electron*, 2015, 62: 3814–3821
 - 27 Cho K, van Merriënboer B, Gulcehre C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In: *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, 2014. 1724–1734
 - 28 Meng Z, Zhan X Y, Li J, et al. An enhancement denoising autoencoder for rolling bearing fault diagnosis. *Measurement*, 2018, 130: 448–454
 - 29 van der Maaten L, Hinton G. Visualizing data using t-SNE. *J Mach Learn Res*, 2008, 9: 2579–2605
 - 30 Tran T N, Drab K, Daszykowski M. Revised DBSCAN algorithm to cluster data with dense adjacent clusters. *Chemometr Intell Lab Syst*, 2013, 120: 92–96
 - 31 Yang C H, Yang C, Peng T, et al. A fault-injection strategy for traction drive control systems. *IEEE Trans Ind Electron*, 2017, 64: 5719–5727
 - 32 Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput*, 1997, 9: 1735–1780
 - 33 Zhao R, Yan R Q, Wang J J, et al. Learning to monitor machine health with convolutional Bi-directional LSTM networks. *Sensors*, 2017, 17: 273
 - 34 Srivastava N, Mansimov E, Salakhutdinov R. Unsupervised learning of video representations using LSTMs. In: *Proceedings of the 32nd International Conference on Machine Learning*, 2015. 843–852
 - 35 Bishop C. Pattern recognition and machine learning. *Stat Sci*, 2006, 4: 559–599
 - 36 Balasubramanian M. The Isomap algorithm and topological stability. *Science*, 2002, 295: 7
 - 37 Hofmann T, Scholkopf B, Smola A J. Kernel methods in machine learning. *Ann Stat*, 2008, 36: 1171–1220

Incipient fault diagnosis for high-speed train traction systems via improved LSTM

Zehui MAO¹, Yuxing GU¹, Bin JIANG^{1*}, Dezhi XU², Xiuwen SUN^{1,3} & Wenjing LIU⁴

1. College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China;

2. School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China;

3. Turing AI Institute of Nanjing, Nanjing 210046, China;

4. Science and Technology on Space Intelligent Control Laboratory, Beijing 100094, China

* Corresponding author. E-mail: binjiang@nuaa.edu.cn

Abstract The incipient fault diagnosis of high-speed train traction systems is a fundamental but essential task to guarantee the safety and reliability of high-speed trains. Traditional intelligent data-driven methods require comprehensive expertise and cannot diagnose incipient faults in the early period. This paper focuses on the fault diagnosis approach for incipient time-varying faults and unknown faults using sensor signals. Considering that the sequential signal has non-stationary characteristics and the incipient faults are slow time-varying, an improved LSTM (long short-term memory) network that is an unsupervised recurrent neural network, is developed. Combined with the GRU (gated recurrent unit) decoder, an LSTM-based encoder-decoder model is proposed to extract the feature vectors, which has the robustness to non-stationary characteristics. Then, the t-SNE (t-distributed stochastic neighbor embedding) approach is used to reduce the dimension of extracted feature vectors, and the DBSCAN (density-based spatial clustering of applications with noise) is utilized for clustering and recognizing faults to realize the unsupervised fault diagnosis. The proposed approach is verified using the data from a semi-physical platform of the CRH₂ traction system, in which the known and unknown faults are illustrated. The results show that the proposed unsupervised approach can achieve a fault diagnosis rate of over 95%.

Keywords fault diagnosis, incipient time-varying fault, unsupervised learning, improved LSTM network

Zehui MAO was born in 1981. She received her Ph.D. degree in control theory and control engineering from Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing, China, in 2009. She is currently a professor at the College of Automation Engineering in NUAA. She was a visiting scholar at the University of Virginia. She worked in the areas of fault diagnosis, with particular interests in nonlinear control

systems, sampled-data systems, and networked control systems. Her current research interests include fault diagnosis and fault-tolerant control of systems with disturbance and incipient faults, and high-speed train and spacecraft flight control applications.

Yuxing GU was born in 1994. He received his B.S. degree from the College of Automation Engineering at Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing, China, in 2017. He is currently pursuing his M.S. degree at NUAA.

Bin JIANG was born in 1966. He received his Ph.D. degree in automatic control from Northeastern University, Shenyang, China, in 1995. He is currently a chair professor of Cheung Kong Scholar Program with the Ministry of Education and a vice president of Nanjing University of Aeronautics and Astronautics, Nanjing, China. He was a recipient of the Second Class Prize of National Natural Science Award of China in 2018. He is an IEEE fellow

and a fellow of Chinese Association of Automation. His research interests include fault diagnosis and fault-tolerant control and their applications in aircraft, satellites, and high-speed trains.

Dezhi XU was born in 1985. He received his Ph.D. degree in control theory and control engineering from Nanjing University of Aeronautics and Astronautics, China, in 2013. He was a visiting fellow with the Department of Biomedical Engineering, City University of Hong Kong, Hong Kong, from 2018 to 2019. He joined Jiangnan University from 2014, as an associate professor. His current research interests include data driven control, fault diagnosis and fault-tolerant control, multi-agent systems and CPSs, technologies of renewable energy, motor control, and smart grid.