﻿Multi-sensor Planetary gearbox fault diagnosis using Multi-layer Convolutional Gated Recurrent Unit Networks

**abstract**

**keywords: intelligent diagnosis, gearbox fault diagnosis, deep learning, time-frequency features**

[【深度好文】齿轮箱故障简易诊断方法详解](https://baijiahao.baidu.com/s?id=1701448745077712363&wfr=spider&for=pc)

Planetary gearbox fault diagnosis using bidirectional-convolutional LSTM networks(多传感器)

一、Introduction

随着城轨列车服役年限的增加，齿轮、轮对、轴承等关键部件逐渐进入损耗失效期，故障率不断提高，针对城市轨道交通安全运营的重要需求，亟待对其关键部件的健康状态进行监测并诊断。由于实际应用的安全需求，城轨列车的运行数据存在故障数据小且故障种类少的问题，这种不平衡的小样本集难以对故障诊断模型进行有效训练，致使故障特征的提取较为困难。因此，开展城轨列车齿轮箱智能化、精确化、快速化的故障诊断研究已成为城市轨道交通安全运营迫切需要解决的问题。

可将现有的齿轮箱故障诊断方法分为四个类别，即基于物理模型的方法、基于信号处理的方法、基于机器学习的方法及其混合方法。基于物理模型检测方法对检测人员的领域知识要求非常高，并且需要对机器构成有着非常深入的了解，导致在动态和嘈杂的复杂环境中很难为复杂的机械设备构建相对准确的物理模型。扎实的故障表示理论和数学基础是基于信号处理检测方法的前提，其旨在探索先进的信号去噪和滤波技术以此对故障特征进行有效提取。作为数据驱动方法的典型代表，机器学习检测方法随着现代工业的发展活跃而辉煌。传统浅层机器学习模型，通常需要专业人员手动提取与选择任务相关的特征，这在复杂的大数据分析中非常具有局限性。其次，特征提取和决策是相互分开的，其中异步的最优化过程需要消耗大量的时间，并且由于结构浅，也难以有效挖掘高维特征。随着机械结构复杂性的增加，若需要获得更好的诊断性能往往需要增加传感器的数量与种类，从而导致数据维度和动态性的增加。综上所述，传统故障检测方法难以获得满意的诊断结果。

机器学习中最热门的分支，深度学习，见证了计算机图像、自然语言处理等多个领域的蓬勃发展。除了强大的数据处理、特征提取能力与网络架构创新等内在因素外，工业大数据的爆发式增长、硬件的突破等多方面的刺激也是不可忽视的外部因素。自然，深度学习在过去的五年中也掀起了智能故障诊断的热潮。

其中卷积神经网络，自动编码器等深度学习方法由于其可以自动提取空间特征，占当前齿轮箱故障诊断的主导地位。[引用文献]。总之，CNN可以从震动幅度的变化和频谱不同层次中提取空间特征，并且可通过识别测量值之间的相关性来融合多个传感器的数据源。但不能提取包含故障敏感信息的时间特征，如异常相应对应的时间步的顺序或时间间隔。

循环神经网络，LSTM等深度学习方法逐渐被应用于自动提取时间特征以进行齿轮箱故障诊断。[引用文献]。总之，大部分现有方法通过捕获故障引起的异常响应起止时间与正常响应之间的时间间隔来捕捉时序数据的时变变化。然而由于网络结构的局限性，这些深度学习方法无法在一个时间步长内确定多个观察值之间的短期空间相关性。

由于齿轮箱结构复杂，为了提高分类准确性accelerometers [7], shaft encoders [8], and acoustic sensors [9]等多种传感器设备已被用于监测齿轮箱的健康状况以及对故障进行分类。不同种类的传感器具有不同的可测频率范围。一般来说，接触式传感器中，速度型传感器适用于测量不平衡、不对中、松动接触等引起的低频振动，用它测量振动位移，可以得到稳定的数据；加速度传感器适用于测量齿轮、轴承故障等引起的中、高频振动信号，但用它测量振动位移往往不太稳定。因此，加速度传感器测量仪一般只用于测振动速度，其优点是能测到高频振动信号。随着传感器种类与数量的增加，齿轮箱的故障诊断逐渐与多传感器数据融合技术相结合。传统的数据融合方法从单个观测量中提取特征，然后根据物理特性与统计分析进行特征提取。[引用文献]。但不同传感器信号之间往往体现出不同的物理特性，例如accelerometers与acoustic sensors的信号几乎没有相关性，导致大多数现有研究只能提取相同类型的不同传感器之间的关系。并且传统方法需要从的那个传感器源中提取手工特征，导致效率较低。

为了解决这些问题，同时捕获时空特征与融合不同种类的传感器数据并进行特征提取，我们提出了一种新的深度学习方法并应用于不同工作状况下的齿轮箱故障诊断。首先，将采样信号划分为多个时间步，对每个时间步中的信号通过小波包分解对频带信息进行划分。根据海森堡不确定性原理，分解后的时频域信息中时域分辨率会降低，同时频域分辨率将升高。以此可将一维时域信号张成一个包含时频信息的特征矩阵。将多个传感器分解得到的特征矩阵在深度维度进行拼接，可得到一个分别以时域分辨率、频域分辨率、传感器数量为长、宽、高的三维矩阵。其次，用卷积运算替换矩阵乘法对GRU网络中的门控单元进行改进，进而得到ConvGRU单元。ConvGRU单元中的卷积操作旨在提取同一个时间步内的短期信号特征，而在ConvGRU单元之间的前向传播过程旨在通过不同时间步之间特征的长期时序依赖关系来捕获异常相应。

The main contributions in this paper are summarized as follows.

1. 在几乎不依赖专家知识和使用特定信号处理方法进行特征选择的情况下，所提出方法使用原始时序信号作为输入，以端到端的方式进行齿轮箱故障诊断。
2. 提出的MCGRU网络模型通过卷积操作有效的将多传感器数据进行融合，使该模型对传感器数量与类别具有扩展性，并且使用多传感器数据时诊断正确率有较大提升。随着循环神经网络层数的升高，通过池化等操作，获取更高层次的特征信息的同时降低特征维度，相对于传统卷积神经网络与循环神经网络参数数量大量减少。
3. 通过结合卷积神经网络与循环神经网络的优势，提出的MCGRU网络模型将三维特征矩阵作为输入信息与中间特征在循环神经网络中进行计算，在捕获隐藏在序列数据中的长期时序依赖关系的同时保留了信号在空间与频域上的特征信息。 将所提出网络模型学习得到的依赖项应用于不同工况下的故障分类任务。实验结果表明，所提出方法明显优于或媲美现有的智能诊断方法。

The rest of this paper is organized as follows.

Section~\ref{sec2} introduces the technical background used in the proposed method.

Section~\ref{sec3} illustrates the framework of the proposed MCGRU method in detail.

Section~\ref{sec4} presents a case study to demonstrate the effectiveness of the proposed framework.

Finally, we conclude the paper and future work in Section~\ref{sec5}.

﻿二、Background and preliminaries

1. **小波包分解**

傅里叶变换适用于处理平稳信号，只能获取一段信号总体上包含哪些频率的成分，但是对各成分出现的时刻并无所知。 因此时域相差很大的两个信号，可能频谱图一样。小波分析和小波包分解适合对非平稳信号分析，相比较小波分析，利用小波包分解可以对信号分析更加精细，将时频平面划分的更为细致，并且可以根据信号的特征，自适应的选择最佳小波基函数，以便更好的对信号进行分析，所以小波包分解应用更加广泛。

小波分析只对信号的低频部分做进一步分解，而对高频部分也即信号的细节部分不再继续分解，所以小波变换能够很好地表征一大类以低频信息为主要成分的信号，不能很好地分解和表示包含大量细节信息，细小边缘或纹理的信号，如非平稳机械振动信号、遥感图像、地震信号和生物医学信号等。

小波包分解既可以对低频部分信号进行分解，也可以对高频部分进行分解，而且这种分解既无冗余，也无疏漏，所以对包含大量中、高频信息的信号能够进行更好的时频局部化分析。将频带部分多层次划分, 对多分辨率分析没有细分的高频部分进一步分解, 并能够根据被分析信号的特征, 自适应地选择相应的频带, 使之与信号频谱相匹配, 从而提高了时频分辨率。

小波降噪是通过小波变换对分解得到的小波系数进行分类处理。经小波分解后，有用信号的小波分解系数较大，噪声的小波分解系数较小。因此，通过设置阈值函数进行去噪处理，大于阈值的分解系数保留下来，而小于阈值的分解系数则通过置零操作予以消除。小波包分解以小波变换为基础，对小波变换中未处理的高频分量再次细化分解，相比于小波降噪，小波包分解具有更高的频率分辨率， 可以进一步消除高频部分存在的噪声余量，提高去噪精度。

根据帕塞瓦尔（Parseval）定理，一个信号在其时域上的总能量与在频域上的总能量是相等的。对于小波包分解，只是将信号高频和低频成分分离出来，在信号形式上发生了变化，但是分解前后的总能量始终保持相等。

1. **CNN(卷积、池化、激活、BN操作)**

**卷积：**

一般我们接触过的都是一维信号的卷积，也就是



在信号处理中，是输入信号，是单位响应。于是输出信号就是输入信号响应的延迟叠加。这也就是一维卷积本质：加权叠加/积分。

那么对于二维信号，比如图像，卷积的公式如下所示：



可以看出，其实二维卷积一样也是加权叠加/积分。需要注意的是，其中 Convolution Kernel 进行了水平和竖直方向的翻转，二维卷积常用来提取空间特征。

**池化：**

池化层基于局部相关性的思想，通过从局部相关的一组元素中进行采样或信息聚合，从而得到新的元素值。Pooling通过对 Feature Map 降维，有效减少后续层需要的参数。并且pooling操作可以引入Translation Invariance，它表示当input中的元素在邻域发生微小位移时，Pooling Layer 的输出是几乎不变的，从而增强了网络的鲁棒性，并且有一定抗扰动的作用。

**激活函数：**

神经网络中每一层的输入输出都是一个线性求和的过程，下一层的输出只是承接了上一层输入函数的线性变换，所以如果没有激活函数，那么无论构造的神经网络多么复杂，最后的输出都是输入的线性组合，并不能够解决更为复杂的问题。而引入非线性的激活函数之后，使得神经网络可以逼近其他的任何非线性函数，进而增强网络的表示能力和学习能力。

神经网络的激活函数通常具有以下几点性质：

* 在定义域上连续并可导，允许少数点上不可导，可导的激活函数可以直接利用数值优化的方法来学习网络参数；
* 激活函数及其导数要尽可能简单一些，太复杂会降低网络计算效率；
* 激活函数的导函数值域要在一个合适的区间内，否则会影响训练的效率和稳定性。

**BN：**

深层神经网络在做非线性变换前的激活输入值随着网络深度加深或者在训练过程中，其分布逐渐发生偏移或者变动，之所以训练收敛慢，一般是整体分布逐渐往非线性函数的取值区间的上下限两端靠近，所以这导致反向传播时低层神经网络的梯度消失，这是训练深层神经网络收敛越来越慢的本质原因，而BN就是通过一定的规范化手段，对于每个隐层神经元，把逐渐向非线性函数映射后向取值区间极限饱和区靠拢的输入分布强制拉回到均值为0方差为1的比较标准的正态分布，使得非线性变换函数的输入值落入对输入比较敏感的区域，以此避免梯度消失问题。这样输入的小变化就会导致损失函数较大的变化，从而导致梯度变大，避免梯度消失问题产生，而且梯度变大意味着学习收敛速度快，能大大加快训练速度。

1. **GRU for Sequence Modeling**

经典的神经网络很难捕捉时间序列中时间间隔较长的依赖，因为通常会产生梯度消失或者梯度爆炸的现象，其中梯度消失更加常见。梯度裁剪虽然可以有效解决梯度爆炸的问题，但无法处理梯度消失。门控循环神经网络的提出，是为了更好地捕捉时间序列中的长期时序依赖关系。它通过引入reset gate和update gate来控制信息流动，其中重置门决定了如何将新的输入信息与先前的记忆相结合，更新门定义了先前记忆保存到当前时间步记忆的量。

重置门和更新门的计算方式如下式所示：





随后门控神经单元同通过计算候选隐藏状态（candidate activation）来辅助隐藏状态的计算。





其中, ​ 表示候选隐藏状态。重置门控制了上一步隐藏状态如何流入到当前的候选隐藏状态；更新门控制了包含当前时间步信息的候选隐藏状态如何更新当前隐藏状态。

﻿三、Diagnosis framework with Multi-layer Convolutional Gated Recurrent Unit Networks (MCGRU)

**Formulation of fault diagnosis 问题建模作为段落起始**

所提出方法的整体结构如图所示。首先，将原始传感器信号分割为若干个时间片段，每个时间片段代表循环神经网络中的一个时间步。对原始时序信号通过小波包分析使频域分辨率与时域分辨率相接近，其中频域分辨率会升高，时域分辨率随之降低。通过将原始GRU单元中的矩阵乘法操作改进为卷积操作来保留一个时间步内的时频信息与不同传感器之间的空间关系，而GRU网络的前向传播过程决定了两个相邻时间步中上述特征的时序依赖关系。从异常相应中提取得到的短期空间相关性与长期时序依赖都会更新至MCGRU网络单元中的隐藏状态。通过更新不同层中网络单元的权重与偏置值来改进特征提取与学习过程。将网络最终层中末端时间步的隐藏状态与诊断结果进行全连接操作，使用交叉熵损失函数确定每次前向传播过程中故障诊断的误差，通过反向传播来优化网络中的权重和偏置值。

1. **Input layer (raw signal transform to time-frequent domain)**

在输入层中，传感器获取的时序信号长度由采样频率与监测时间确定。由于传感器采样频率通常较高，故在时域上存在很多冗余信息。首先使用相同大小的时间窗口将各个传感器获取的信号进行切片，并对窗口内的信号使用小波包分析将时序信号转换为二维时频特征矩阵。该做法降低时域分辨率并提升频域分辨率，减少信息时域冗余信息的同时增加了频域信息，使特征更具多样性。将同一时间步中不同传感器的特征矩阵在深度维度进行拼接，最终得到一个维度为X×Y×D的三维特征矩阵，其中X、Y、D分别为时域分辨率、频域分辨率、传感器数量，这个三维特征矩阵代表了当前时间步t，融合了不同传感器的时频域信号特征。由于不同传感器的采样频率与监测时间可能存在差异，为了确保不同传感器时频域信息维度的一致，使用0对缺少的部分进行填并假设没有关于填充部分的先验知识。每个特征矩阵依次作为MCGRU的输入，以学习每个时间步中的短期空间相关性以及不同时间步间的长期时序相关性。MCGRU一层中单元的数量等于L/(X×Y)。

1. **﻿Convolutional gated recurrent unit cell**

FC-GRU在输入到状态和状态到状态的转换中使用完全连接将数据拼接为一维向量，导致处理多维数据时导致丢失了空间信息。为了解决上述问题，每个时间步输入的数据，细胞单元候选隐藏状态，隐藏状态与门控单元 of the MCGRU 是三维矩阵而不是一维向量，其中前两个维度分别代表时频特征的时域分辨率与频域分辨率。门控单元的计算过程中使用卷积操作来替换直接连接操作，故每个门中的卷积操作融合了高维数据并将高维特征更新至MCGRU单元中的隐藏状态。划分出的每个时间步的长度越长，则三维特征矩阵中可以存储更多具有短期相关性的信息，如图所示。MCGRU中的重置门与更新门通过其本地临域单元的输入和过去状态来确定当前候选隐藏状态，并使用该候选状态对当前隐藏状态进行更新。MCGRU中计算过程的关键方程如下，









其中\*表示卷积算子，表示Hadamard乘积，，，，，，均为权重矩阵，矩阵中的权重与偏置值在每次迭代中都会被更新。

如果将每个隐含状态视为揭示时序信号异常点的某种表示，则具有较大卷积核的MCGRU应该能够捕获时频变化较大的特征，而具有较小卷积核的网络能够捕获时频变化较小的特征。随着网络网络层数的增加，下层隐含状态作为高层的输入进行纵向传播时，使用池化操作对特征进行更高层次的提取，并降低维度。并且在更高层次的卷积操作中，卷积核的大小会随着特征的维度自适应调整，故不同的网络层次可以获得不同层次的特征信息。整个过程可以视为，相同维度的隐含特征在循环神经网络中横向传播，并且每个时间步上都存在一个纵向的卷积神经网络，该卷积神经网络使用低层循环神经网络输出的隐藏状态作为输入，卷积操作后的中间特征作为下一时间步中循环神经网络的输入，池化并激活后的特征作为下一层卷积神经网络的输入，整个过程如图所示。

1. **Feature space**

传统卷积神经网络搭配循环神经网络的故障诊断模型中，利用CNN作为特征提取器，图像或是信号拼接而成的矩阵中提取空间特征关系。但由于传统循环神经网络只能够处理1D向量，故需要将提取获得的2D空间特征矩阵进行拼接转换为1D空间特征向量。与传统的CNN-RNN网络相结合的方法相比，将CNN集成至GRU网络中得到MCGRU网络，从而可以创建一个3D特征空间，包括2D时频特征和1D传感器空间特征。如图所示，MCGRU单元可以同时从3D输入中提取多传感器融合后的时频特征并捕获时频特征在时序上的长期依赖。

1. **Hidden state and classification layers**

MCGRU网络的最高层中末端时间步的隐藏状态包含了所有时间步中多传感器融合后的时频信息。将该最终状态通过两层全连接层将状态特征连接至训练数据的标签。将全连接层的输出经过softmax操作得到预测分布。使用交叉熵来衡量对于同一个随机变量x的真实分布p(x)和预测分布q(x)之间的差异，且交叉熵的值越小表示两个分布越接近。交叉熵公式如下：

，其中n表示事件可能发生的情况总数。显然使用q(x)来描述样本不如p(x)准确，q(x)需要不断地学习来拟合准确的分布p(x)。

从公式得知，交叉熵关注的是正确类别的预测概率。分类问题中，模型的输出空间是概率分布，但目标输出空间是样例的类别，也就是说我们最终目标是获得正确的类别。交叉熵是假设模型分布为多项式分布时，模型分布的负条件对数似然。

﻿四、Experimental evaluation

**本节将首先对源自传动系统动力学模拟器 (DDS) 的数据集进行介绍。应用数据集进行一系列可视化实验，以获得对模型行为的一些基本理解。网络的最佳设计和调整参数可能会因不同的应用而改变，通过使用不同层数与内核大小运行模型以评估不同网络和调整参数的性能，并选择具有最佳性能的配置作为最终模型。最后，将所提出的方法与几种成熟的方法进行了比较，验证了该方法的有效性。**

4.1 Experimental setup and data description

4.2. Data preprocessing

数据驱动的分析方法建立在有效数据的基础上，然而设备状态监测数据会受传感器的产品/安装特性（灵敏度、线性度、重复性、漂移等）、工作环境、数据传输通道等的影响而出现低精度和异常，进行影响后续分析算法的性能。因此，在数据接入或数据分析之前，通常需要进行预处理。

在本研究中，使用了一种简单但可靠的方式从每个数据通道中去除异常值，其将与中值相差超过三个scaled median absolute deviations (MAD)的点标记为异常值[Alternatives to the median absolute deviation]。其中scaled MAD为定义为：

其中A为输入数据，C为公式，且erfcinv是逆互补误差函数被定义为inverfc(y)= {x : erfc(x) = y}[Improved exponential bounds and approximation for the Q-function with application to average error probability computation]。

原始振动数据可能包含有关部件故障特征的重要信息，因此如RMS等时域特征已广泛用于机械部件故障诊断。但基于深度学习的方法有望在无人工特征提取的数据上取得令人满意的诊断结果。部分文献表明故障特征可能表现为主频率附近边带中 FFT 振幅的增加，本文使用二进小波包变换将不重要（特征不明显）的信息分量去除，达到数据精简的效果。

为了使频域可分性增加，使用Symlets小波作为基函数。由于运行频率的存在，无法在原始时域数据中检测到故障特征。 但是，不同信号频谱存在明显差异。 将小波包变换后不同频段的信号在频率维度与传感器通道维度进行堆叠，拼接成三维张量，直接用于模型中进行自动特征学习和故障诊断。

在许多深度领先方法中，常使用数据增强来增加训练数据集的大小以改善分类精度。该方法如图[]所示，通过使用滑动窗口的方式获取足够数量的训练样本。本文使用的窗口大小为4096，且在信号上随机滑动进行数据的采样。

4.3. Results and discussion

4.3.1. Case 1: Hyper-Parameters’ Effect on Diagnosis Result

4.3.2. Experiment 1: Automatic feature learning for classification

﻿五、Conclusion