

蒋姮博,张剑,方荣超,等. 基于大数据分析的列车轮对故障诊断方法[J]. 湖南科技大学学报(自然科学版), 2021, 36(1): 91-98. doi:10.13582/j.cnki.1672-9102.2021.01.014
JIANG H B, ZHANG J, FANG R C, et al. Research on Fault Diagnosis Method of Train Wheelset Based on Big Data Analysis [J]. Journal of Hunan University of Science and Technology (Natural Science Edition), 2021, 36(1): 91-98. doi:10.13582/j.cnki.1672-9102.2021.01.014

基于大数据分析的列车轮对故障诊断方法

蒋姮博,张剑*,方荣超,欧阳婉卿,罗禹杰,卢晓宇

(湖南科技大学 信息与电气工程学院,湘潭 湖南 411201)

摘要:随着我国列车行驶速度的不断提高,列车的行车安全逐渐得到了人们的重视,海量列车轮对监测数据为分析列车运行状态提供了条件.为了提高列车轮对故障诊断效率和准确性,文章提出一种基于大数据分析的列车轮对故障诊断方法,针对传统列车轮对故障诊断方法在处理大规模监测数据集时存在处理时间长,故障结果不准确等问题.首先设计一个监测数据融合框架,然后将多故障诊断循环神经网络算法与大数据 MapReduce 分布式计算框架相结合,利用循环神经网络算法特征提取能力和 MapReduce 快速计算能力.这样不但能够发挥循环神经网络故障特征提取能力,还能够满足列车轮对故障诊断精确性和实时性的需求,最后通过实例分析,证明了该方法的有效性.

关键词:列车轮对;故障诊断;数据融合;大数据;神经网络

中图分类号:U270.33 **文献标志码:**A **文章编号:**1672-9102(2021)01-0091-08

Research on Fault Diagnosis Method of Train Wheelset Based on Big Data Analysis

JIANG Hengbo, Zhang Jian, FANG Rongchao, OUYANG Wanqing, LUO Yujie, LU Xiaoyu

(College of Information and Electrical Engineering, Hunan University of Science and Technology, Xiangtan 411201, China)

Abstract: With the improvement of the train speed, more and more importance has been attached to the safety of the train. A large number of train wheelset monitoring data provide conditions to analyze the train running status. To improve the efficiency and accuracy of train wheelset fault diagnosis, based on Big Data analysis, an approach for train wheel sets fault diagnosis was put forward. The aim was to solve the problems of the traditional train wheelset fault diagnosis method, such as long processing time, and inaccurate fault results. A framework of data fusion for monitoring was presented at begin. Then, combined the multi fault diagnosis using cyclic neural network algorithm with the big data MapReduce distributed computing framework, the ability of feature extraction of the cyclic neural network algorithm and MapReduce fast computing ability was used. Result showed that the method meets the requirement of the accuracy and real-time of the train wheel sets fault diagnosis. The case analysis also proves the effectiveness at later.

Keywords: train wheelset; fault diagnosis; data fusion; big data; neural network

在我国列车行驶不断提速的同时,列车出现的故障也进一步显露出来,温州动车追尾、武广高铁因故

收稿日期:2020-04-03

基金项目:湖南省自然科学基金资助项目(2020JJ5170);湖南省教育厅一般项目资助(18JC0299)

*通信作者, E-mail: jzhang@hust.edu.cn

障造成旅客滞留情况的出现充分说明了这一点,高速列车虽然在铁路车辆中发生故障较少,但同样不容忽视.近年来,研究学者们对列车故障诊断进行了一定的研究,例如,针对铁路车轮缺陷提出利用模糊逻辑对其进行诊断^[1];文献[2]提出了一种用于滚动轴承故障诊断的深度对抗域自适应模型;利用有限元分析对高速列车变速箱结构的故障进行了识别^[3];文献[4]提出了一种新的深层一维卷积神经网络并将它运用到列车故障诊断中.

列车轮对系统是列车机械部分中关键的部件之一,也是列车走行部的故障多发部位.但传统的列车轮对故障诊断方法主要采用单一数据直接进行诊断,没有将列车轮对系统的运行监测数据、历史监测数据、维护检修数据作为一个整体来考虑,从而造成列车轮对故障诊断效率低和诊断结果不准确^[5].而采用大数据分析技术可以从这些杂乱无章监测数据中挖掘出对列车轮对故障诊断有用的信息,它可以把诊断对象来自多方面的监测数据综合有效地融合在一起,实现对杂乱无章的列车轮对监测数据整理并得出关于诊断对象更详细、全面的分析,使故障诊断结果更加准确^[6-7].

本文针对传统 BP 神经网络融合算法^[8]在处理大规模数据集时存在数据融合效率低、数据处理时间长等不足^[9],提出了改进的数据融合算法,同时为实现及时对海量列车轮对监测数据处理,将改进的数据融合算法与大数据分布式计算框架相结合,以便于对大规模数据集进行及时处理,满足列车轮对故障诊断实时性需求.

1 列车轮对监测数据采集

列车轮对故障诊断系统采用以 DSP 为核心的结构,能对列车轮对系统进行实时故障诊断;采用总线式分布结构,具有集成度高、抗干扰能力强、稳定性好、成本低等特点.硬件主要包括 DSP 诊断分机、传感器、接线盒等,列车轮对故障诊断系统框架图如图 1 所示.

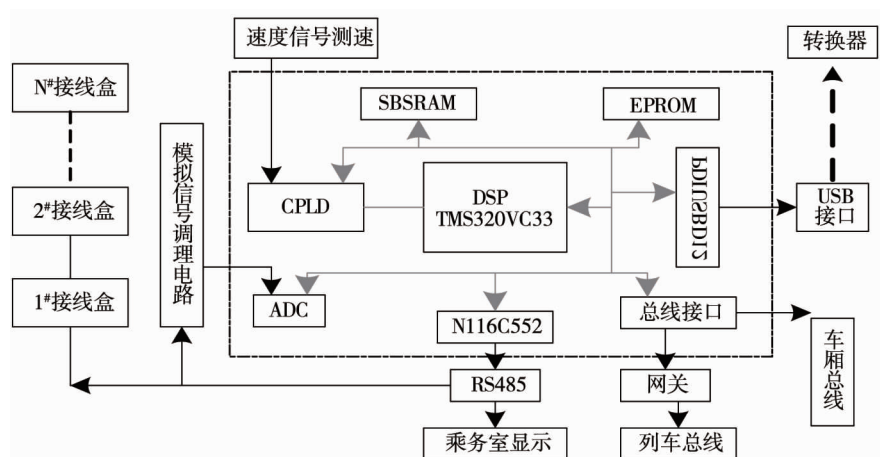


图 1 列车轮对故障诊断系统框架

列车轮对监测数据虽然很多,但只需要一只组合传感器(多传感器)就能将其拾取.为了准确获取列车轮对系统运行状况,通常采用多点测量的方法进行数据采集,图 2 是测量数据采集示意图.

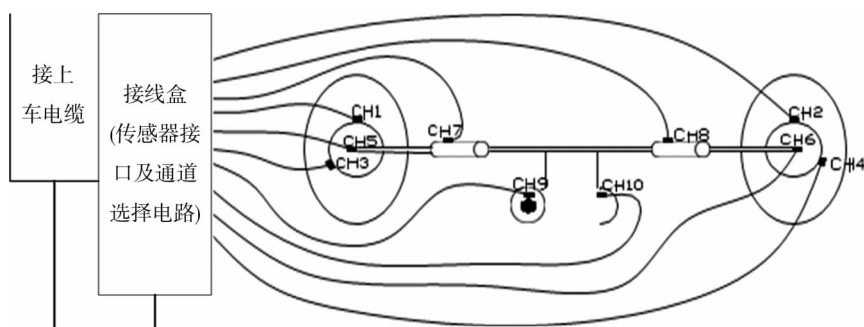


图 2 监测数据采集

如图2所示是在列车轮对系统监测数据采集方式,图中的10个传感器CH1~CH10分别安装在列车轮对各测点,其中传感器CH1,CH2采集列车轮对2个轴承、轮箍及踏面等运行中的垂直向振动信号,传感器CH3,CH4采集列车轮对2个轴承、轮箍及踏面等运行中的水平向振动信号,传感器CH5,CH6采集列车轮对2个轴承、轮箍及踏面等运行的轴向振动信号,传感器CH7,CH8采集列车轮对2个制度盘的工作状态数据,传感器CH9,CH10采集列车轮对牵引机电电压电流数据.通过对列车轮对监测数据的采集,为接下数据融合和故障诊断做好了准备.

2 数据融合架构搭建

安装在列车的传感器是获取列车轮对状态的数据来源,传感器是列车感知轮对状态的重要组成部分,传感器测量数据的准确性直接影响系统的运行,影响分析和决策的正确性^[10].而列车轮对监测数据融合则是将这些传感器监测数据通过适当的融合处理,得到对列车轮对故障诊断有用的信息.列车轮对监测数据经过数据融合后的信息,能精确地反映出轮对列车运行的状态,列车轮对监测数据融合框架图如图3所示.

由图3可得列车轮对监测数据融合框架图有三层结构,可以实现列车轮对监测数据共享、可融合、故障诊断高效率的要求.首先传感测量层通过多个传感器对列车轮对各个部件进行实时的运行监测,并将监测的各类数据传送数据管理层.其次数据管理层将采集到列车轮对监测数据输送到故障特征库,特征库配备了诸如NoSQL,HDFS等工具的数据管理层,能够对数据进一步存储和分布计算处理^[11].最后应用层在MapReduce分布式并行计算编程框架中迁入信息融合算法,把列车轮对不同部件的监测数据融合在一起.在运用层依据相关算法以列车轮对做出故障诊断,同时还能对故障进行分类和故障溯源.

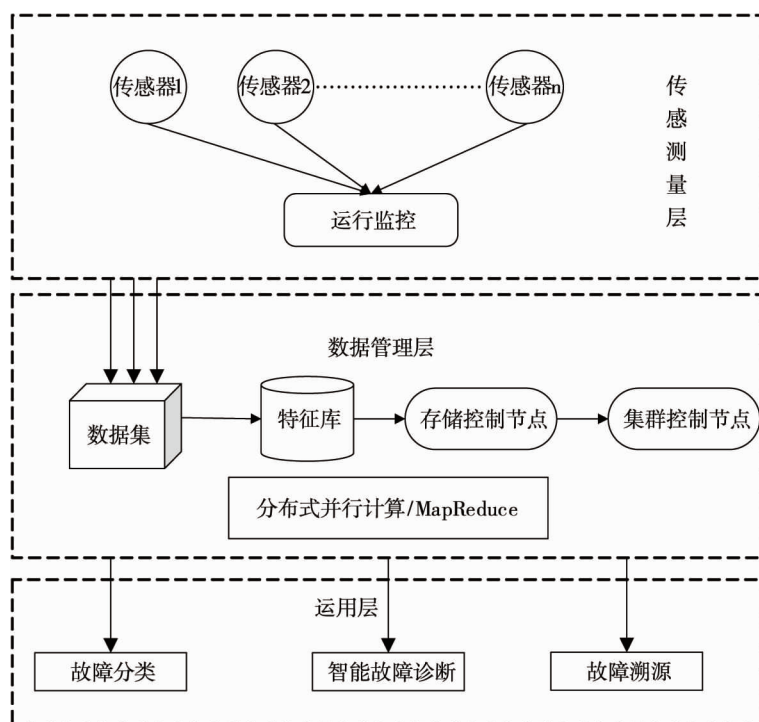


图3 数据融合框架

3 数据融合算法

改进的数据融合算法采用一种循环神经网络算法对监测数据进行故障特征提取,其结构图如图4所示,监测数据由输入层进入到隐层1以后,输入层与隐层1之间各个节点依次顺序连接,网络隐层1与隐层2之间各个节点也作顺序连接,接着,数据传输到输出层,不断训练、迭代、循环.在此过程中,多故障诊断神经网络算法建立在经典循环神经网络的基础上,经过循环神经网络不同的单元节点隐藏层的参数不断更新^[9].

改进的数据融合算法示意图如图4所示.

第一个隐层利用经典循环神经网络的更新方法,

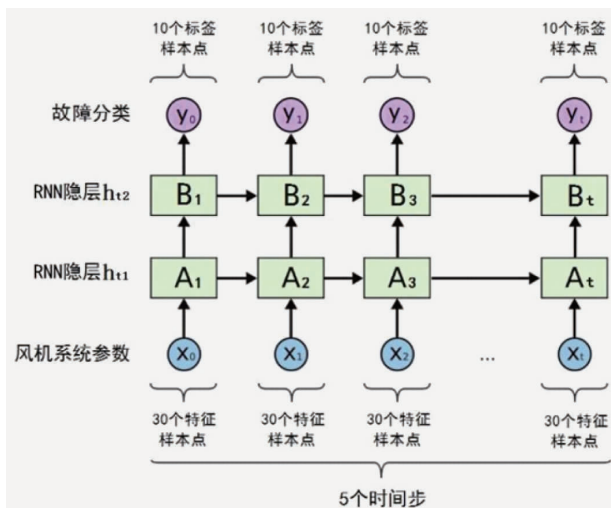


图4 改进的数据融合算法

该隐层 $f_{i1}(t)$ 的结果与前一时刻的状态 $f_{i1}(t-1)$ 和当前时刻的输入 $x(t)$ 相关,该隐层状态的计算方式为

$$f_{i1}(t) = \sigma(\mathbf{W}_x x(t) + \mathbf{W}_{h1} f_{i1}(t-1) + \mathbf{b}). \quad (1)$$

式中: $f_{i1}(t)$ 为第一个网络隐层输出的信息特征; σ 为sigmoid函数; \mathbf{W}_x 为网络输入层到隐层权重矩阵; \mathbf{W}_{h1} 为网络隐层和隐层的权重矩阵; \mathbf{b} 表示偏置向量.

第二个隐层使用第一个隐层训练之后的信息进行训练,在该隐层数据的训练更新为

$$f_{i2}(t) = \sigma(\mathbf{W}_x x(t) + \mathbf{W}_{h2} f_{i2}(t-1) + \mathbf{W}_y y(t-1) + f_{i1}(t)). \quad (2)$$

式中: f_{i2} 为第二个隐层输出的信息特征; $y(t-1)$ 为前一时刻的输出; \mathbf{W}_y 为隐层到输出层的权重矩阵.

另外,使用sigmoid函数作为隐层输出的激活函数,即

$$\sigma(x) = \frac{1}{1 + e^{-x}}. \quad (3)$$

最后,使用SoftMax函数作为输出层的激活函数,即

$$g(Z_m) = \frac{e^{Z_m}}{\sum e^{Z_k}}. \quad (4)$$

由于列车轮对故障诊断监测数据结构各不相同,需要将网络调整为多输入单输出结构,网络输入层为含多样属性的数据样本,因此数据要进行归一化处理,然后利用网络训练结构学习逼近,判断是否满足偏差,如果不满足偏差,则进一步修正权值.

不同类型列车轮对监测数据对应不同学习网络,而且应根据实际情况来调整网络隐含层的数目和神经元的节点数.把一个数据类型不同的学习网络输出层设为1,网络输出实际值不满足期待值时设为0,这样就可以根据设定阈值来对数据进行归类.改进算法后的数据分类融合流程图如图5所示.

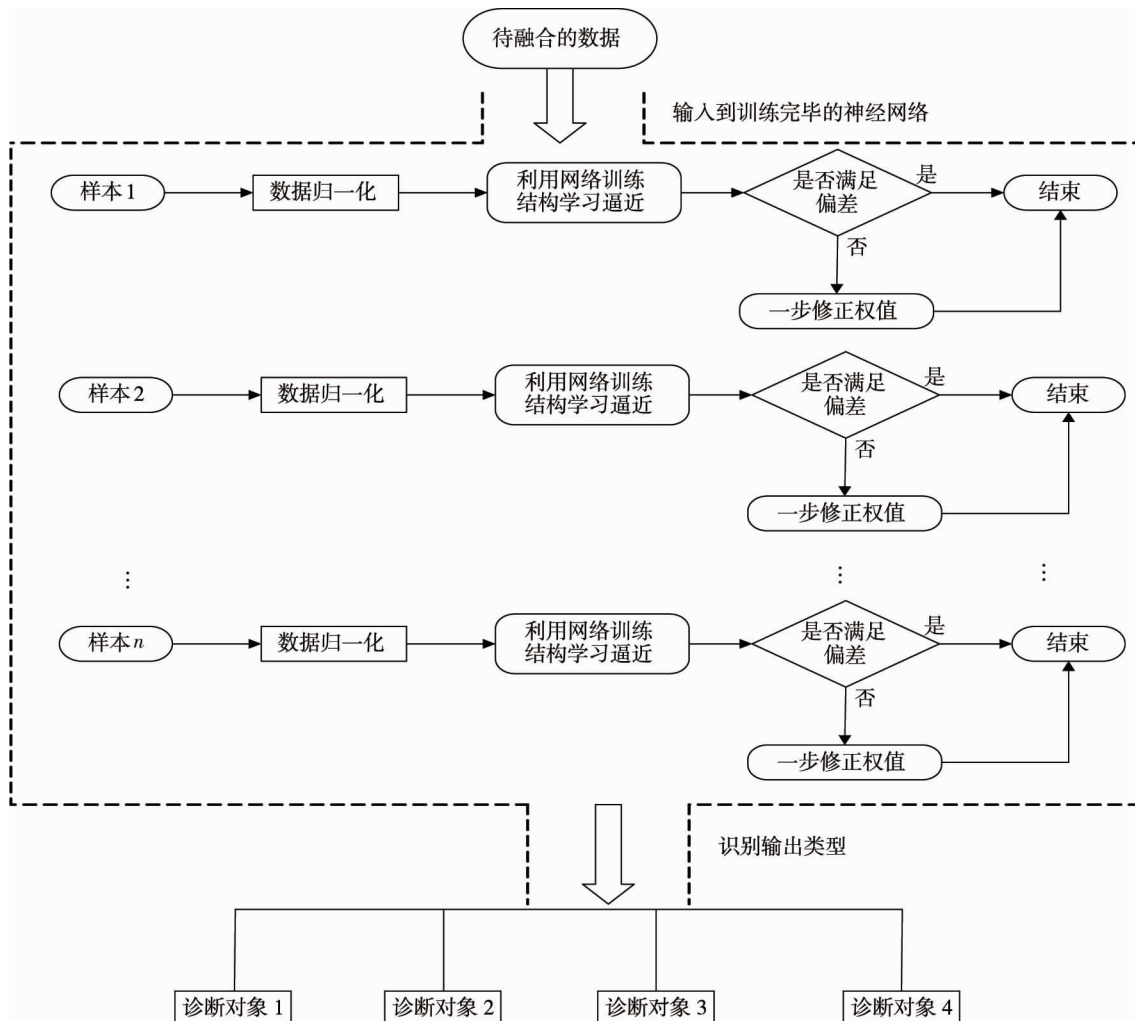


图5 数据分类融合流程

由图 5 可得将待融合的数据输入训练完毕的神经网络,利用网络训练结构学习逼近,根据输出结果将不同数据样本归划到被诊断的对象中,最后分类结果分为诊断对象 1、诊断对象 2、诊断对象 3 和诊断对象 4。

如何依据监测数据实现列车轮对故障诊断,其中监测数据故障特征提取是一个至关重要的步骤,基于循环神经网络多故障诊断算法实现了对杂乱无章的监测数据展开故障特征参数提取,并对列车轮对进行故障诊断.然而基于循环神经网络多故障诊断算法在当面对海量的列车轮对监测数据时,存在收敛速度慢的缺陷,从而导致无法满足列车轮对故障诊断实时性的需求^[12].因此,需要将基于循环神经网络多故障诊断算法与大数据 MapReduce 分布式计算框架相结合,可以实现对大规模数据集进行分布式特征提取,分布式特征提取与传统方法也稍许不同,大数据分析的核心 MapReduce 是在并行框架下采用分布式算法的特征提取.MapReduce 数据处理是向管理者提供了一种简便的分布式计算编程方案.MapReduce 是一种应用广泛的分布式计算框架,其基本思想是将整体数据分成几部分小块数据逐一进行分析计算后,再将计算结果汇总后进行分析,最终得到需要的结果^[13].MapReduce 数据处理模型如图 6 所示。

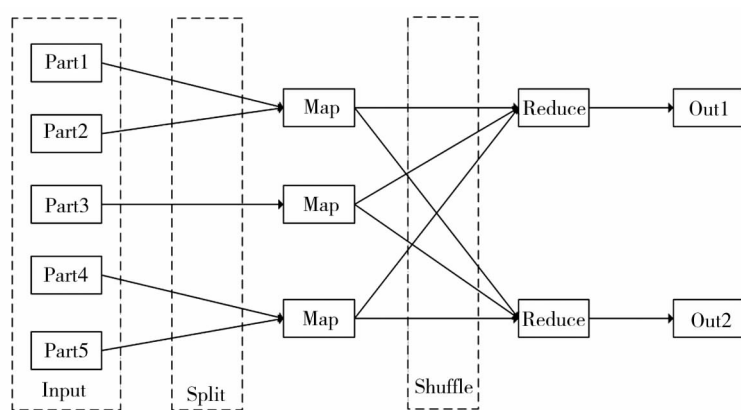


图 6 MapReduce 数据处理模型

利用 MapReduce 分布式计算模型将改进数据融合算法并行化,并将其建立在 Hadoop 分布式集群上,以达到对列车轮对故障的实时诊断.把改进数据融合算法与 Hadoop 分布式计算框架 MapReduce 相结合,并将其命名 MR-PFP 算法,以列车轮对 4 类故障诊断为目标,其流程图如图 7 所示。

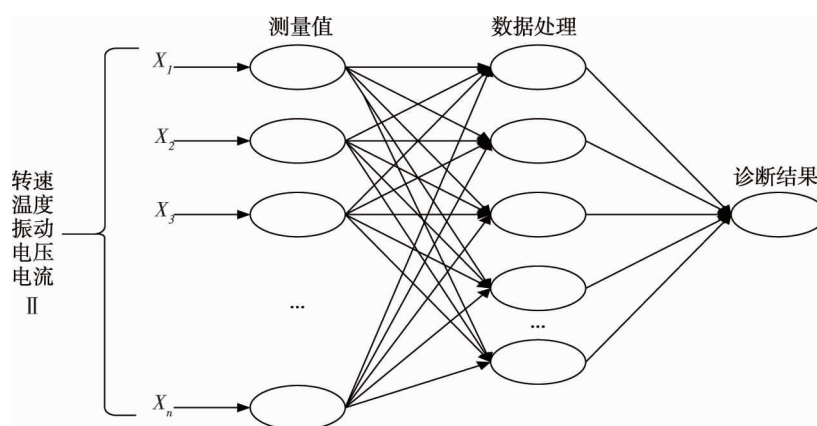


图 7 数据并行融合处理算法

MR-PFP 算法主要过程:

首先,根据列车轮对故障诊断要求,将采集的轮对牵引机电电压电流、轴承温度,列车时速,各部件振动频率数据进行离散化处理.然后将数据进行矩阵化,设 G 矩阵为列车轮对系统某一时刻监测到的数据集 $G = (g_1, g_2, \dots, g_m, t)$. 其中, m 为矩阵的维数, n 为矩阵的数据取值, t 是数据采集的时刻^[12].

$$\mathbf{G} = \begin{bmatrix} G_1 \\ G_2 \\ \vdots \\ G_n \end{bmatrix} = \begin{bmatrix} G_{11} & G_{12} & \cdots & G_{1m} & t_1 \\ G_{21} & G_{22} & \cdots & G_{2m} & t_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ G_{n1} & G_{n2} & \cdots & G_{nm} & t_n \end{bmatrix}. \quad (5)$$

其次,列车轮对系统在某一个时间段内监测数据集矩阵 \mathbf{E} 为

$$\mathbf{E} = \begin{bmatrix} G_{q1} \\ G_{q2} \\ \vdots \\ G_{qi} \end{bmatrix}. \quad (6)$$

最后,通过 MR-PFP 算法对列车轮对进行故障诊断.

4 实例分析

本文实验数据为长沙车辆段检修车间收集的某型号列车 2019 年 6 月份轮对历史监测数据,数据量总共为 10 GB(64 935 000 条数据).本节的故障诊断实验结果分析将从召回率(recall)、准确率(accuracy)、标准误差(RMSE)和平均绝对百分误差(MAPE)4 个方面对本文算法和传统 BP 神经网络算法进行对比.它们具体计算公式为

$$\text{recall} = \frac{T_p}{T_p + F_N}; \quad (7)$$

$$\text{accuracy} = \frac{T_p + T_N}{P + N}; \quad (8)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (X_{p_i} - x_{m_i})^2}{n}}; \quad (9)$$

$$\text{MAPE} = \frac{\left(\sum_{i=1}^n \left(\frac{|X_{p_i} - x_{m_i}|}{x_{m_i}} \right) \right)}{n} \times 100\%. \quad (10)$$

式中: T_p 为列车轮对在真实数据中故障同时在故障诊断中也为故障的总数; T_n 为列车轮对在真实数据中无故障而在故障诊断为无故障的总数; F_n 为列车轮对在真实数据中故障而在故障诊断中为无故障的总数; P 为真实数据中列车轮对为故障的总数; N 为真实数据中列车轮对无故障的总数^[14]; x_{m_i} 为故障参数值; X_{p_i} 为故障诊断值; n 为故障诊断的次数 ($n > 1$)^[15].

根据式(7)~式(10)可得,召回率表示在列车轮对故障数据样本中找出了多少故障;准确率表示在列车轮对故障诊断工程中正确有多高,标准误差越小表示列车轮对故障诊断越精确,平均绝对百分误差越小表示列车轮对故障诊断的结果越接近真实值^[16].

2 种算法故障诊断的召回率如图 8 所示.

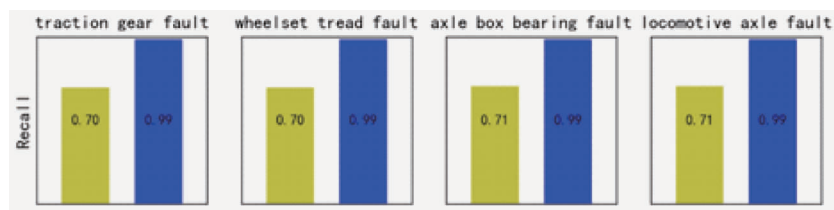


图 8 召回率对比

由图 8 可以看出在召回率上,传统的 BP 神经网络算法在列车轮对故障诊断时召回率只有 70%左右,容易在故障诊断时出现漏诊的现象,而在实际运用中可能会因为漏诊而造成列车重大安全事故,本文算法

在列车轮对 4 种故障诊断中的召回率达到了 99%,可以避免漏诊的情况出现.

2 种算法故障诊断的准确率如图 9 所示.

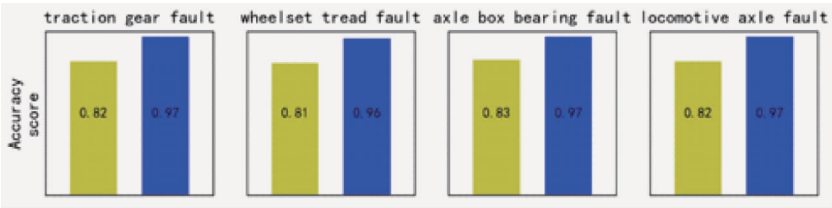


图 9 准确率对比

由图 9 可得在准确率上,本文算法在列车轮对 4 类故障诊断中的准确率要比传统 BP 神经网络高出 15%左右,这表明本文算法在列车轮对故障诊断中的效果更好,基于信息融合的本文算法更适合于轮对故障诊断.

为了对比本文方法与传统的 BP 神经网络算法的列车轮对故障诊断准确度,本文用标准误差和平均绝对百分误差来作为评价指标,2 种方法的标准误差如表 1 所示.

表 1 2 种算法标准误差对比

算法类型	标准误差
BP 神经网络	2.672 8
本文算法	1.253 6

由表 1 可得,本文方法的标准误差比传统 BP 神经网络算法低 1.419 2,这表明使用本文方法的列车轮对故障诊断可靠性更高,为了直观地看出两者之间的区别,两者的平均绝对百分误差的仿真图如图 10 所示.

由图 10 可得 2 种算法的平均绝对百分误差在不同的时间段,本文方法的值都要比传统 BP 神经网络的值要低,这说明使用本文方法的列车轮对故障诊断结果比传统 BP 神经网络算法更加准确^[17].

为了验证本文方法在处理大规模数据集的优势,将列车轮对监测数据集分为 10,50,100,150 G 这 4 组不同容量的数据样本,然后用 2 种算法分别对 4 组数据样本进行列车轮对故障诊断,并分别记录它们的计算处理时间.实验结果表明随着数据容量的上升传统 BP 神经网络算法的处理时间急速上升,而本文方法增幅不大,这说明本文方法于传统 BP 神经网络算法相比,在处理海量列车轮对监测数据时更有优势,能满足列车轮对故障诊断实时性的需求,2 种方法在 4 组不同容量数据处理时间如图 11 所示.

由图 11 可得当监测数据容量为 10 GB 时,本文方法和传统方法的处理时间差别并不大,当监测数据容量超过 10 GB 后,传统方法用时增长很快而本文方法增长缓慢,这说明基于大数据分析的列车轮对故障诊断方法更加适用于对海量列车轮对监测数据的处理.

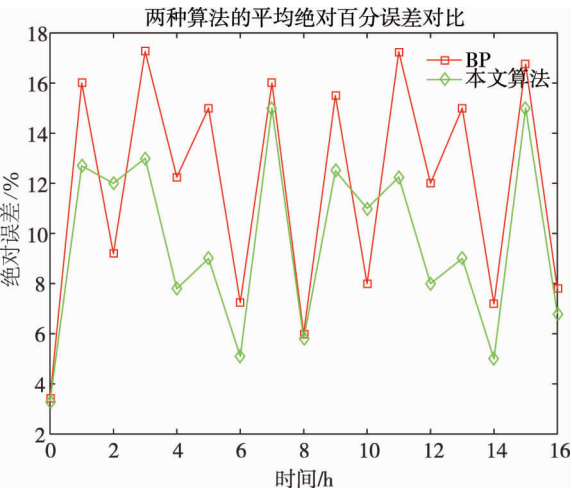


图 10 2 种算法的平均绝对百分误差对比

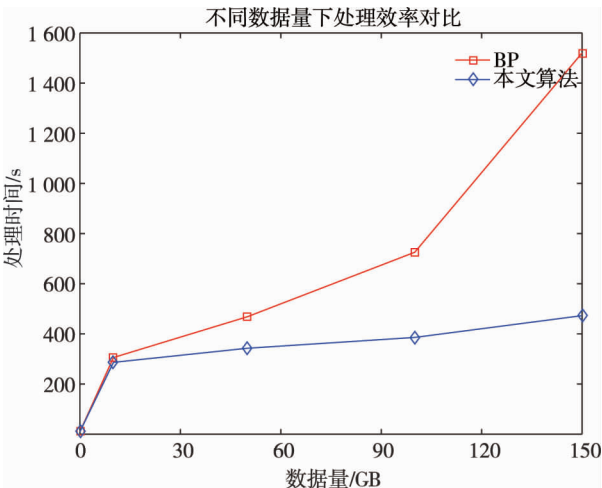


图 11 不同数据量下处理效率对比

5 结论

1) 将改进的数据融合算法与分布式计算框架 Map Reduce 相结合,并对改进的算法并行化处理,提出了一种更加适合于大规模数据集处理的 MR-PFP 算法.

2) 通过实例分析,从召回率、准确率、标准误差和平均绝对百分误差 4 个方面对本文提出的算法和传统 BP 神经网络算法进行了对比,得出 MR-PFP 算法的列车轮对故障诊断结果更加准确,证明了该方法的有效性,提高了列车轮对故障诊断效率和准确度,对大数据背景下的列车轮对故障诊断具有一定意义.

参考文献:

- [1] Skarlatos D, Karakasis K, Trochidis A. Railway wheel fault diagnosis using a fuzzy-logic method[J]. Applied Acoustics, 2004, 65(10): 951-966.
- [2] Liu Z H, Lu B L, Wei H L, et al. Deepadversarial domain adaptation model for bearing fault diagnosis[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2019, 99: 1-10.
- [3] Zhang B, Tan A C C, Lin J. Gearbox fault diagnosis of high-speed railway train[J]. Engineering Failure Analysis, 2016, 66: 407-420.
- [4] Peng D, Liu Z, Wang H, et al. A novel deeper one-dimensional CNN with residual learning for fault diagnosis of wheelset bearings in high-speed trains[J]. IEEE Access, 2018, 7: 10278-10293.
- [5] 张剑,樊晓平,黄采伦,等.列车轮对状态的融合监测系统[J].交通运输工程学报,2008,8(6):13-19.
- [6] 胡智勇. 列车轴承轨边声学监测与诊断方法研究[D].合肥:中国科学技术大学,2016.
- [7] Hu H, Tang B, Gong X, et al. Intelligent fault diagnosis of the high-speed train with big data based on deep neural networks[J]. IEEE Transactions on Industrial Informatics, 2017, 13(4): 2106-2116.
- [8] Fei S, Jun G. The data fusion method of basing BP neural network[J]. Automation and Instrumentation, 2005(5): 63-65.
- [9] Moosavian A, Khazaei M, Najafi G, et al. Spark plug fault recognition based on sensor fusion and classifier combination using Dempster - Shafer evidence theory[J]. Applied Acoustics, 2015, 93: 120-129.
- [10] 邢璐璐. 高速列车减振器状态监测及故障诊断方法研究[D].北京:中国铁道科学研究院,2018.
- [11] Salahshoor K, Kordestani M, Khoshro M S. Fault detection and diagnosis of an industrial steamturbine using fusion of SVM (support vector machine) and ANFIS (adaptive neuro - fuzzy inference system) classifiers[J]. Energy, 2010, 35(12): 5472-5482.
- [12] 化柏林,李广建.大数据环境下多源信息融合的理论与应用探讨[J].图书情报工作,2015,59(16):5-10.
- [13] Dean J, Ghemawat S. MapReduce: a flexible data processing tool[J]. Communications of the ACM, 2010, 53(1): 72-77.
- [14] Blinowska K J, Kuś R, Kamiński M. Granger causality and information flow in multivariate processes[J]. Physical Review E, 2004, 70(5): 050902.
- [15] Shi J, Ding Z, Lee W J, et al. Hybrid forecasting model for very-short term wind power forecasting based on grey relational analysis and wind speed distribution features[J]. IEEE Transactions on Smart Grid, 2013, 5(1): 521-526.
- [16] 赵小乐,吴亚东,张红英,等. 基于密切多项式近似的多项式插值算法框架[J]. 计算机应用, 2015(8): 2266-2273.
- [17] 黄小霞. 列车轮对轴承故障实时检测系统研究[D].成都:西南交通大学,2017.