

数据挖掘的铁路接触网设备运行状态识别研究

谈 强

(中国铁建电气化局集团第二工程有限公司, 山西 太原 030000)

摘要: 铁路接触网设备运行状态识别研究具有十分重要的理论意义和实际价值, 当前铁路接触网设备运行状态识别方法存在识别成功率低、误差大等不足。为了提高铁路接触网设备运行状态识别效果, 提出了基于数据挖掘的铁路接触网设备运行状态识别方法。首先分析了当前铁路接触网设备运行状态识别的研究进展, 找到各种方法的局限性; 其次采集铁路接触网设备运行状态信号, 从中提取铁路接触网设备运行状态识别特征; 然后采用数据挖掘技术对铁路接触网设备运行状态进行建模与分析, 建立铁路接触网设备运行状态识别的分类器; 最后在MATLAB 2018平台进行了铁路接触网设备运行状态识别的仿真实验。结果表明, 数据挖掘的铁路接触网设备运行状态识别正确率高达90%以上, 铁路接触网设备运行状态识别速度快, 获得比当前其他方法更好的识别结果。

关键词: 铁路接触网; 数据挖掘; 设备运行状态; 识别模型; 仿真测试

Research on operation status identification of railway catenary equipment based on data mining

TAN Qiang

(China Railway Construction Electrification Bureau Group No.2 Engineering Co., Ltd., Taiyuan 030000, China)

Abstract: The research on the operation state identification of railway catenary equipment has very important theoretical significance and practical value. However, the current operation status identification methods of railway catenary equipment have some shortcomings, such as low identification success rate and large error. In order to improve the identification effect of railway catenary equipment operation status, a method based on data mining is proposed. This paper first analyzes the current research progress of railway catenary equipment operation status identification, finds out the limitations of various methods, and then collects the operation status signals of railway catenary equipment, extracts the operation status identification features of railway catenary equipment, and then uses data mining technology to model and analyze the operation status of railway catenary equipment, and establishes the operation status classifier of railway catenary equipment. Finally, a simulation experiment is carried out on MATLAB 2018 platform. The results show that the accuracy rate of the operation status identification of the railway catenary equipment is more than 90%. The identification speed of the operation status of the railway catenary equipment is fast, and the identification result is better than other methods.

Key words: railway catenary; data mining; equipment running state; identification model; simulation test

0 引言

近些年来, 我国铁路发展十分迅速, 其承担

了越来越繁重的客运任务, 当前铁路都是电气化的, 电气化铁路成为交通系统的重要组成部分。

接触网将牵引变电所的电能直接输送到动车组，其设备稳定性对铁路的快速、安全运行具有十分重要的影响。因此如何对动车接触网设备运行状态进行正确识别至关重要，成为了人们关注的新课题^[1-3]。

最初铁路接触网设备运行状态识别通过人工方式，需要花费大量的人力、物力，铁路接触网设备运行状态识别成本高，而且结果的主观性强，无法对铁路接触网设备的异常运行状态及时识别^[4-6]。随着信息处理技术、自动化技术的发展，一些学者提出铁路接触网设备运行状态自动识别方法，如基于红外图像的铁路接触网设备运行状态识别方法，首先采用红外装置对铁路接触网设备运行状态图像进行在线采集，并对图像中的目标进行分析，实现铁路接触网设备运行状态识别，但受天气、环境影响大，铁路接触网设备运行状态识别结果不稳定^[7-9]；有学者提出基于信号处理技术的铁路接触网设备运行状态识别方法，首先采集铁路接触网设备运行状态的信号，然后从信号中提取特征，采用模式识别技术进行铁路接触网设备运行状态识别，如支持向量机、深度学习网络和BP神经网络等，它们的铁路接触网设备运行状态识别效果较好，在实际应用中，该类方法也存在各自的不足，如支持向量机的铁路接触网设备运行状态识别效率低，深度学习网络的铁路接触网设备运行状态识别错误率高，BP神经网络的参数难以确定，使得铁路接触网设备运行状态识别结果的可靠性差^[10]。

为了提高铁路接触网设备运行状态识别正确率，加快铁路接触网设备运行状态识别速度，提出了基于数据挖掘的铁路接触网设备运行状态识别方法，并在MATLAB 2018平台对铁路接触网设备运行状态识别结果进行了测试。结果表明，本文方法是一种正确率高、速度快的铁路接触网设备运行状态识别方法。

1 电气化铁路接触网组成

当前铁路主要牵引方式为电力牵引，与电力牵引相关的所有电力传输及其附属设备组成牵

引供电系统，主要负责为高速动车组的正常运行提供电能，其结构示意图如图1所示。从图1可以看出，接触网是铁路牵引供电系统的重要组成部分，主要包括基础、支柱、接触悬挂及定位装置和支持装置等，这些设备工作在室外环境中，易出现故障，如断线故障、接触线零部件故障和外力造成的故障，一旦出现故障，就会影响整个铁路牵引供电系统的正常运行，因此对铁路接触网设备运行状态进行准确、及时识别，可以提高接触网的可靠性，及时发现安全隐患，从而保证铁路安全运行。

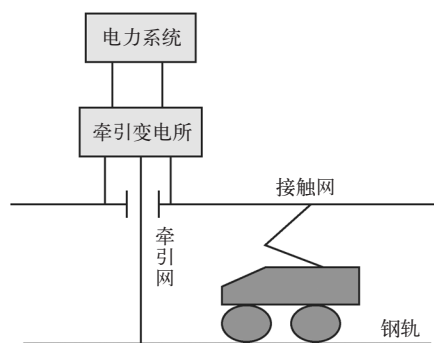


图1 铁路牵引供电系统的示意图

Fig. 1 Schematic diagram of railway traction power supply system

2 数据挖掘的铁路接触网设备运行状态识别方法

2.1 铁路接触网设备运行状态信号采集及特征提取

无线传感器的工作环境要求低，对信号采集实时性强，本文采用无线传感器采集铁路接触网设备运行状态信号。

由于铁路接触网设备运行状态信号具有一定的非平稳性，采用小波分析无法提取其有效特征，因此本文首先采用经验模态分解法对铁路接触网设备运行状态信号进行处理，然后采用希尔伯特谱分析提取处理后的信号铁路接触网设备运行状态信号识别特征。

2.2 经验模态分解法处理状态信号的步骤

1) 经验模态分解法确定原始铁路接触网设备运行状态信号局部极值点，并将最大局部极值

点连接在一起, 形成一个上包络线, 采用相似的原理, 产生下包络线。

2) 计算铁路接触网设备运行状态信号与上、下包络线均值的差, 具体如下

$$y_1(t) = x(t) - \mu_1 \quad (1)$$

式中, $x(t)$ 为原始铁路接触网设备运行状态信号; μ_1 为上、下包络线均值。

3) 若 $y_1(t)$ 不满足本征模态分量的条件, 那么就把 $y_1(t)$ 当原始信号, 重复执行步骤1) 和步骤2); 如果满足本征模态分量的条件, 那么有

$$c_1(t) = y_1(t) \quad (2)$$

式中, $c_1(t)$ 为第1个本征模态分量。

4) 将第1个本征模态分量从原始铁路接触网设备运行状态信号去除, 则有

$$r_1(t) = x(t) - c_1(t) \quad (3)$$

5) $r_1(t)$ 作为原始铁路接触网设备运行状态信号, 执行步骤1) ~ 步骤3), 产生第2个本征模态分量, 重复 n 次得到

$$\begin{cases} r_1(t) = r_1(t) - c_1(t) \\ \vdots \\ r_n(t) = r_{n-1}(t) - c_n(t) \end{cases} \quad (4)$$

6) 综合上述可以得到

$$x(t) = \sum_{i=1}^n c_i(t) + r_n(t) \quad (5)$$

2.3 提取铁路接触网设备运行状态识别特征

采用希尔伯特谱分析法提取铁路接触网设备运行状态识别特征, 具体步骤如下:

1) 采用希尔伯特谱分析对 $y_1(t)$ 进行处理, 得到

$$y_i(t) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{y_1(\tau)}{t - \tau} d\tau \quad (6)$$

2) 对 $y_1(t)$ 进行解析操作, 可以得到

$$z_i(t) = y_i(t) + jy_i(t) = a_i(t)e^{j\varphi_i(t)} \quad (7)$$

3) 幅值函数和瞬时频率计算公式为

$$a_i(t) = \sqrt{y_i^2(t) + \hat{y}_i^2(t)} \quad (8)$$

$$f_i(t) = \frac{1}{2\pi} \omega_i(t) = \frac{1}{2\pi} \times \frac{d\varphi_i(t)}{dt} \quad (9)$$

4) 希尔伯特谱为

$$\begin{aligned} x(t) &= RP \sum_{i=1}^n a_i(t) e^{j\varphi_i(t)} \\ &= RP \sum_{i=1}^n a_i(t) e^{j \int \varphi_i(t) dt} \end{aligned} \quad (10)$$

5) 希尔伯特边际谱为

$$h(\omega) = \int_0^T RP \sum_{i=1}^n a_i(t) e^{j \int \varphi_i(t) dt} dt \quad (11)$$

式中, T 为铁路接触网设备运行状态信号长度。

2.4 铁路接触网设备运行状态识别的分类器设计

BP神经网络是当前一种流行数据挖掘技术, 不需要事先给出问题的数学公式, 可以通过模仿人脑思考与处理问题能力对问题进行求解, 因此采用BP神经网络设计铁路接触网设备运行状态识别的分类器。

BP神经网络包含输入层、隐含层和输出层, 其结构如图2所示。

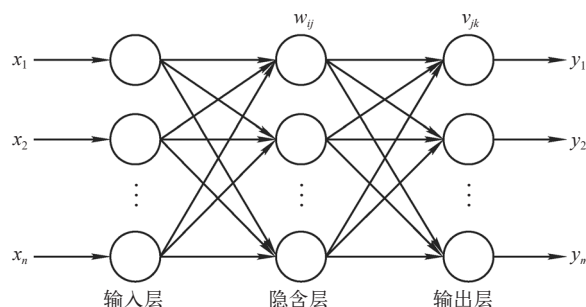


图2 BP神经网络的结构

Fig. 2 Structure of BP neural network

BP神经网络的输入和输出向量分别为 $X = (x_1, x_2, \dots, x_n)^T$ 和 $Y = (y_1, y_2, \dots, y_m)^T$; w_{ij} 和 v_{jk} 分别为隐含层和输出层的权值; θ_i 和 θ_k 分别为隐含层和输出层的阈值; 隐含层和输出层节点的输出分别为

$$u = f\left(\sum_{i=1}^n w_{ij} x_i + \theta_j\right) \quad (12)$$

$$y = f(v_{jk} u + \theta_k) \quad (13)$$

式中, m 为铁路接触网设备运行状态识别特征的数量; $f()$ 为激活函数, 通常定义如下

$$f(x) = \frac{1}{1 + e^{-x}} \quad (14)$$

由于权值和阈值对BP神经网络的训练和学习性能具有重要的影响,如果设置不合理,易陷入局部极小点和收敛慢,为此采用粒子群优化算法求解BP神经网络最优的权值与阈值。

数据挖掘的铁路接触网设备运行状态识别的基本思想为:首先采集铁路接触网设备运行状态信号,然后从信号中提取铁路接触网设备运行状态识别特征,最后采用BP神经网络对铁路接触网设备运行状态进行建模与分析,建立铁路接触网设备运行状态识别的分类器,并采用粒子群算法确定BP神经网络的权值和阈值,具体流程如图3所示。

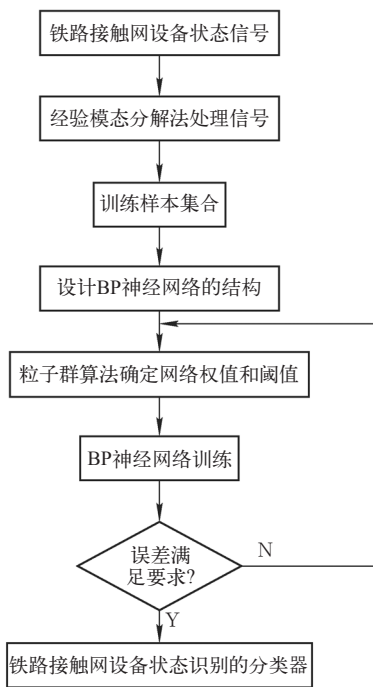


图3 铁路接触网设备运行状态识别流程

Fig. 3 Operation status identification process of railway catenary equipment

3 仿真实验

3.1 实验平台及数据

为了分析数据挖掘的铁路接触网设备运行状态识别效果,选择传统BP神经网络的铁路接触网设备运行状态识别方法,基于红外图像的铁路接触网设备运行状态识别方法进行仿真对照实验,设备运行状态为断线故障、接触线零部件故

障、外力造成的故障和正常状态,它们样本数量如下表所示。采用MATLAB 2018作为实验平台。

表 铁路接触网设备运行状态识别的实验数据

Tab. Experimental data for operation status identification of railway catenary equipment

编号	设备运行状态	样本数量
1	断线故障	10
2	接触线零部件故障	20
3	外力造成的故障	20
4	正常状态	100

3.2 铁路接触网设备运行状态识别正确率对比与分析

三种方法的铁路接触网设备运行状态识别正确率如图4所示。由图4可知,本文方法的铁路接触网设备运行状态识别正确率平均值为91.20%,传统BP神经网络和红外图像的铁路接触网设备运行状态识别正确率平均值分别为87.06%和87.73%,本文方法的铁路接触网设备运行状态识别错误率明显减少,获得更优的铁路接触网设备运行状态识别结果。

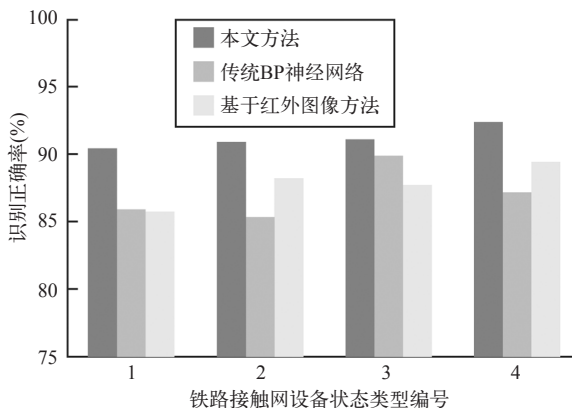


图4 铁路接触网设备运行状态识别正确率对比

Fig. 4 Comparison of identification accuracy of operation status of railway catenary equipment

3.3 铁路接触网设备运行状态识别效率对比与分析

三种方法的铁路接触网设备运行状态识别时间如图5所示。由图5可知,本文方法的铁路接触网设备运行状态识别时间平均值为6.55 ms,传统BP神经网络和红外图像的铁路接触网设备运行状态识别时间分别为11.84 ms和12.38 ms,

本文方法缩短了铁路接触网设备运行状态识别时间，提高了铁路接触网设备运行状态识别效率。

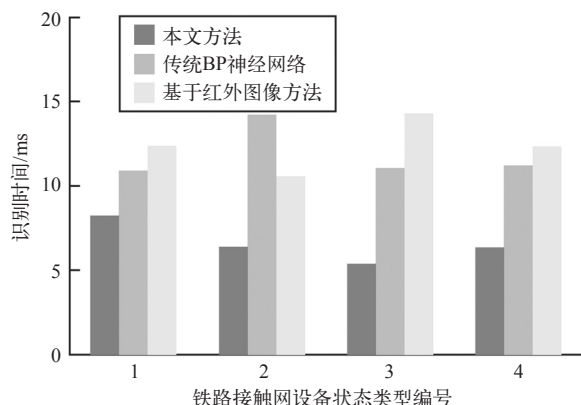


图5 铁路接触网设备运行状态识别时间对比

Fig. 5 Comparison of operation status identification time of railway catenary equipment

4 结束语

铁路接触网设备运行状态识别一直是铁路安全领域中的研究热点，为了解决当前铁路接触网设备运行状态识别过程中存在的问题，提出了基于数据挖掘的铁路接触网设备运行状态识别方法。结果表明，数据挖掘的铁路接触网设备运行状态识别正确率高，效率也要优于其他方法，具有更加广泛的应用前景。

参考文献

- [1] 朱家诚, 吴彬, 吴焱明, 等. 铁路接触网巡检无人车自动定位系统研究[J]. 机械设计与制造, 2018(3): 160-163.
- [2] 赵慧, 肖晓晖, 戚广枫, 等. 高速铁路接触网吊弦的疲劳寿命分析[J]. 武汉大学学报(工学版), 2019, 52(4): 351-357.
- [3] 谭德强, 莫继良, 罗健, 等. 高速铁路接触网定位钩和定位支座失效分析[J]. 中国铁道科学, 2018, 39(5): 111-118.

- [4] 钟俊平, 刘志刚, 陈隽文, 等. 高速铁路接触网悬挂装置开口销不良状态检测方法研究[J]. 铁道学报, 2018, 40(6): 51-59.
- [5] 戚广枫, 赵慧, 肖晓晖, 等. 高速铁路接触网吊弦动应力数值模拟及其疲劳荷载特征分析[J]. 中国机械工程, 2018, 29(9): 1063-1068.
- [6] 陈隽文, 刘志刚, 韩烨, 等. 基于局部特征描述的高速铁路接触网斜撑套筒定位与故障检测[J]. 铁道学报, 2017, 39(11): 30-37.
- [7] 赵峰, 梁丽, 王思华. 基于GO-FLOW法的高速铁路接触网系统可靠性分析[J]. 电工技术学报, 2015, 30(12): 351-356.
- [8] 刘仕兵, 朱雪龙, 张艳伟, 等. 基于灰色聚类 and 组合赋权法的高速铁路接触网健康状态评估研究[J]. 铁道学报, 2016, 38(7): 57-63.
- [9] 周能嘉, 程江洲, 李振华. 基于模糊神经网络的电气化铁路接触网安全评估[J]. 南昌大学学报(理科版), 2016, 40(2): 151-155.
- [10] 贺德强, 江洲, 陈基永, 等. 基于深度卷积神经网络的铁路接触网鸟窝检测方法研究[J]. 机车电传动, 2019(4): 126-130.

作者简介



谈 强

谈强 (1974—), 男, 高级工程师, 主要研究方向为铁道电气化。

(收稿日期: 2020-9-14) **EA**