基于数据挖掘技术的电气化铁路电能质量控制研究

徐元成

(中国铁建电气化局集团北方工程有限公司，山西 太原，030053)

摘要：针对电气化铁路电能质量控制准确度较低等问题，提出基于数据挖掘技术的电气化铁路电能质量控制方法。首先分析当前电气化铁路电能质量控制方法的研究现状，并提取电气化铁路电能质量变化特征；然后将特征输入到数据挖掘技术中的深度学习网络进行训练，建立相应的电气化铁路电能质量控制模型；最后与其他方法进行了电气化铁路电能质量控制仿真对比实验，结果表明，所提方法提高了电气化铁路电能质量控制准确度，获得了更优的电气化铁路电能质量控制结果，具有广泛的应用前景。

关键词：数据挖掘技术；深度学习网络；电气化铁路；电能质量

Research on power quality control of electrified railway based on data mining technology

XU Yuancheng

（North Engineering Co., Ltd. of the Electrification Bureau Group, CRCC, Taiyuan 030053, China）

Abstract: Aiming at the low accuracy of power quality control in electrified railway, a power quality control method based on data mining technology is proposed. Firstly, the current research status of power quality control of electrified railway is analyzed, and the change characteristics of power quality of electrified railway are extracted. Then, the characteristics are input into the deep learning network of data mining technology for training, and the corresponding power quality control model of electrified railway is established. Finally, the simulation experiment of power quality control of electrified railway is carried out with other methods, and the results are shown in the table. The results show that this method can improve the accuracy of power quality control of electrified railway and obtain better results of power quality control of electrified railway, which has a wide application prospect.

Key words: data mining technology; deep learning network; electrified railway; power quality

0 引言

随着铁路技术的不断发展，电气化铁路对国民经济发展起着重要的作用[1]。在电气化铁路的运营过程中，电能质量十分关键，直接影响电气化铁路的正常运营，因此电能质量控制是电气化铁路研究领域的一个热点问题[2]。

针对电气化铁路电能质量控制问题，国外一些发达国家的研究历史比较悠久，电气化铁路电能质量控制比较成熟，而我国对电气化铁路电能质量控制研究的时间比较短，但是发展速度快，涌现出了许多有效的电气化铁路电能质量控制方法[3-4]。当前电气化铁路电能质量控制方法可以划分为两类[5]：一类是传统方法，如基于小波变换的电气化铁路电能质量控制方法等；另一类为现代方法，主要为人工智能和机器学习算法，如基于BP 神经网络的电气化铁路电能质量控制方法。当前这些方法在实际应用中均存在一些不足，如电气化铁路电能质量控制准确度低，电气化铁路电能质量控制的实时性差等，电气化铁路电能质量控制有待进一步研究[6]。

深度学习网络(Deep Neural Network，DNN)

是一种新兴起的数据挖掘技术，也是一种具备非线性映射、深层结构的多层无监督神经网络学习方法，其具备完成复杂函数逼近的能力。DNN 可通过无监督训练逐层学习算法，获取输入数据的主要驱动变量和数据的分布式特征。为了提高电气化铁路电能质量控制准确度，本文提出了基于数据挖掘技术的电气化铁路电能质量控制方法，并通过仿真实验分析了其性能。

1 基于数据挖掘技术的电气化铁路电能质量控制方法

1.1 深度学习网络

采用深度学习理论构建 DNN，为得出数据的分布式特征，通过低层特征组合描述高层特征。该方法利用非监督式贪婪训练法对传感器采集到的电气化铁路电能质量信号进行训练，依次完成各层初始化，并通过 BP 算法实现微调，完成电气化铁路电能质量控制。降噪自编码网络是 DNN 的一种，其结构如图 1 所示。为获取含有噪声的样本数据并对其进行编码，该网络可通过一定几率将输入层节点值置为零。编码后网络对样本进行解读和学习，获取样本的原始形式，以此使降噪自编码网络具备鲁棒性，可降低其对微小随机搅动的敏感性。因此，降噪自编码网络可有效减少异步电动机的机械变化以及环境噪声等因素对异步电动机状况信息提取的影响。

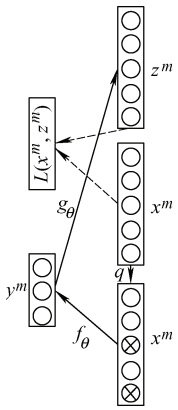

图 1 降噪自编码网络结构

Fig. 1 Structure of noise reduction self coding network

设 $\{x^1,...,x^m,...,x^M\}$ 为降噪自编码网络的输入层，其中参数的上角标表示样本的数量，共 m 个，训练样本数量用 M 表示；$\{z^1,...,z^m,...,z^M\}$ 表示输出层。则用式(1)表示网络结构的损失函数

$$L(x^m,z^m)=\frac{1}{M}\sum_{m=1}^{M}\left\|x^m-z^m\right\|^2 \tag{1}$$

降噪自编码网络在编码阶段时，原始输入电气化铁路电能质量信号可用第一层网络结构的输入层表示，则输入电气化铁路电能质量信号表达式为

$$y^m=f_\theta(\tilde{x}^m)=s_r(W\tilde{x}^m+b) \tag{2}$$

式中，$\tilde{x}^m$ 为 x^m 加入噪声后的输出电气化铁路电能质量信号；s_r 为降噪自编码网络响应函数；编码网络的参数集合用 $\theta=\{W,b\}$ 表示，其中编码网络的连接权值和偏置参数分别用 W 和 b 表示。

降噪自编码网络在解码阶段时，没有加入噪声的原始电气化铁路电能质量信号重构公式为

$$z^m=g_{\theta'}(y^m)=s_g(W'y^m+d) \tag{3}$$

式中，y^m 为编码层输出值；s_g 为解码网络的响应函数；输出层激活函数用 $g(\cdot)$ 表示；解码网络的参数集合用 $\theta'=\{W',d\}$ 表示，其中解码网络的连接权值和偏置参数分别用 W' 和 d 表示。

根据上述内容可知，降噪自编码器作为非监督学习，其获取编码层连接权值和偏置参数时，无须任何标签，依据电气化铁路电能质量信号间的重构误差即可获取，并且该误差属于最小化输出层和输入层。将稀疏性惩罚因子引入降噪自编码器的损失函数中，以此可避免结构过拟合增强特征的聚类能力和判别能力，并获取损失函数为式(4)的稀疏降噪自编码器，即

$$L(x^m,z^m)=\frac{1}{M}\sum_{m=1}^{M}\left\|x^m+z^m\right\|^2+\frac{\lambda}{2}\sum_{i=1}^{s}\sum_{j=1}^{s+1}\left(W_{ji}^{(l)}\right)^2 \tag{4}$$

式中，惩罚因子用 λ 表示，网络的稀疏程度通过其实行控制。

天然聚类能力是稀疏自编码网络的特点，可

使数据所含有的信息被充分运用，有效去除数据中的冗余信息，实现数据利用最大化，可将原始数据本质特征通过高层网络层表达，并且判别能力较高。

1.2 深度学习网络的训练与微调

由于降噪自编码网络的浅层网络结构函数存在一定的表达能力的局限性，因此将其按照顺序实行堆叠，构成叠降噪自编码器(Stacked De noising Autoencoder，SDA)降噪自编码网络，即深度网络结构，其结构如图 2 所示。降噪自编码网络的每一层均符合归一化标准，将上一个降噪自编码网络的编码层作为下一个降噪自编码网络的输入，以此类推，自适应非监督特征的提取通过逐层贪婪训练完成，获取逐层最优结果。基于此，为实现堆叠后分类器整体最优，利用 BP 算法对带有标签的样本数据实行监督式训练，完成微调。为避免传统深度结构容易陷入局部极小值，采取各层预先分别训练。

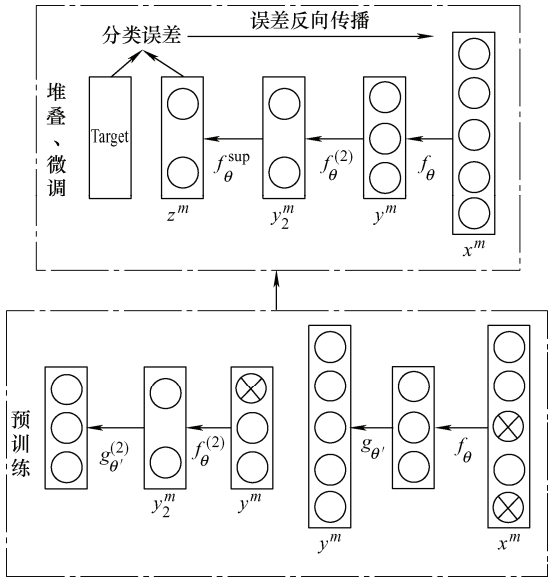

图 2 叠降噪自编码器结构

Fig. 2 Structure of stacked demoise auto encoder

1.3 多响应线性回归模型的电气化铁路电能质量控制

根据多个线性回归模型构建多响应线性回归模型，对上文的输出结果进行集成决策，获取电气化铁路电能质量控制结果。用 $\{x^{(m)}\}_{m=1}^{M}$ 表示样本训练集；故障类别 j(j=1,2,…,c)可用对应样本标签 y 和 y_j 表示。C_i 表示基分类器，其获取的样本预测结果公式为

$$P_{C_i}\left(x^{(m)}\right)=\left(P_{C_i}\left(y_1|x^{(m)}\right),P_{C_i}\left(y_2|x^{(m)}\right),\ldots,P_{C_i}\left(y_c|x^{(m)}\right)\right)^{\mathrm{T}} \tag{5}$$

式中，$P_{C_i}\left(x^{(m)}\right)$ 为 C_i 对 $x^{(m)}$ 预测的概率；$P\left(x^{(m)}\right)$ 为一个 $c\times L$ 维列向量，且由每个降噪自编码网络的 Softmax 类概率输出组成，该向量公式为

$$P\left(x^{(m)}\right)=\begin{pmatrix}P_1^{(1)}\left(x^{(m)}\right),P_2^{(1)}\left(x^{(m)}\right),\ldots,P_c^{(1)}\left(x^{(m)}\right),\ldots,\\ \left(P_1^{(L)}\left(x^{(m)}\right),P_2^{(L)}\left(x^{(m)}\right),\ldots,P_c^{(L)}\left(x^{(m)}\right)\right)\end{pmatrix} \tag{6}$$

多响应线性回归模型可将原来包含 c 个类的分类问题转化成 c 个回归问题。选上文提取的特征预测的 $x^{(m)}$ 属于类 y_j 的概率作为输入特征，实现多响应线性回归模型构建。解决对应于 y_j 的回归问题时，如果类标签位于个体中，其输出为 1。如果其余的类标签是个体的输出，其输出为 0。因此，每个类 y_j 的公式为

$$\mathrm{LR}_j\left(x^{(m)}\right)=\sum_{i=1}^{L}\alpha_j^{(i)}\left(x^{(m)}\right)\quad \alpha_j^{(i)}\geqslant 0\quad j=1,2,\ldots,c \tag{7}$$

式中，α 和 LR()分别为非负系数最小二乘算法预测得出的线性回归模型的参数和线性回归函数。

求解式(7)是预测新样本 x^m 的前提，将计算结果预测为 $\mathrm{LR}_j(x^{(m)})$的类，并且该类具备最大值，则

$$y=\arg\max \mathrm{LR}_j\left(x^{(m)}\right)\quad j=1,2,\ldots,c \tag{8}$$

综上所述，基于数据挖掘技术的电气化铁路电能质量控制流程如图 3 所示。堆叠稀疏降噪自编码网络可以有效完成联合特征的处理，使样本信息最大程度保留。其中叠加多个稀疏降噪自编码，保证输入样本的联合特征被有效提取，再将联合特征输入多响应线性回归模型，实现电气化铁路电能质量控制。因此，将电气

化铁路电能质量信号作为输入，采用堆叠稀疏降噪自动编码器提取信号的联合特征，以多响应线性回归模型建立电气化铁路电能质量控制模型，可以实现基于深度学习特征提取和多响应线性回归模型状态控制的电气化铁路电能质量控制。

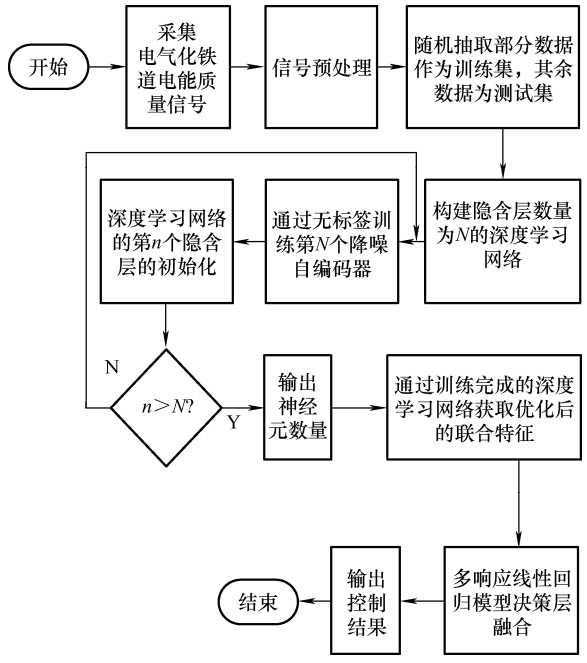

图 3　电气化铁路电能质量控制流程

Fig. 3 Power quality control process of electrified railway

2　仿真测试

2.1　测试环境以及测试对象

为了测试本文提出的数据挖掘技术的电气化铁路电能质量控制方法的性能，采用 MATLAB 2019 进行仿真实验，选择 5 个电气化铁路作为实验对象，采集它们的一段时间电能质量信号进行仿真测试。

2.2　神经元数量对电气化铁路电能质量控制准确度影响

选择第一个电气化铁路的电能质量信号作为测试对象，隐含层层数为 1 时，不同神经元数量下模型的电气化铁路电能质量控制准确度结果如图 4 所示。根据图 4 的测试结果可知：随着神经元数量的增加，电气化铁路电能质量控制的准确度逐渐提高，当神经元数量为 10 时，电气化铁路电能质量控制准确度不再增加，这表明神经元数量的增加可提升电气化铁路电能质量控制准确度，但仅增加神经元数量，无法达到更好的电气化铁路电能质量控制准确度，为此本文实验过程中的神经元数量为 10。

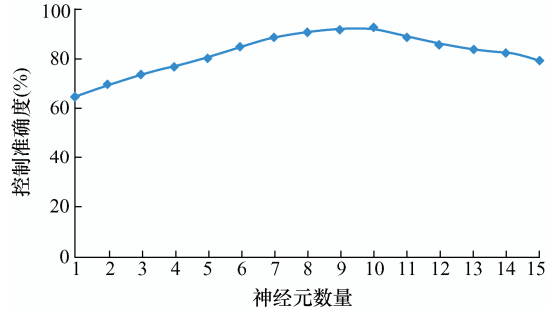

图 4　神经元数量和电能质量控制准确度间的变化关系

Fig. 4 Relationship between the number of neurons and the accuracy of power quality control

2.3　隐含层层数对电气化铁路电能质量控制准确度影响

不同隐含层层数的条件下，电气化铁路电能质量控制准确度统计结果如图 5 所示。根据图 5 的测试结果可知：隐含层的层数逐渐增加，模型的电气化铁路电能质量控制准确度逐渐提高，但是电气化铁路电能质量控制时间越来越小，在保证电气化铁路电能质量控制效率的条件下，隐含层层数为 5 最佳，为此本文实验过程中的隐含层层数量设置 5。

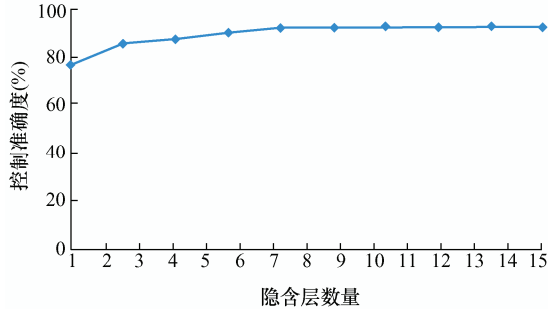

图 5　隐含层层数与电能质量控制准确度间变化关系

Fig. 5 Relationship between the number of hidden layers and the accuracy of power quality control

2.4　不同方法的电气化铁路电能质量控制准确度对比

为验证本文设计的电气化铁路电能质量控制

优越性，选择 BP 神经网络的电气化铁路电能质量控制方法进行对比测试，统计它们的电气化铁路电能质量控制准确度，结果如图 6 所示。对图 6 进行分析可知：本文提出的电气化铁路电能质量控制准确度要远远高于对比方法，降低了电气化铁路电能质量控制误差，对比结果验证了本文提出的电气化铁路电能质量控制方法的优越性。

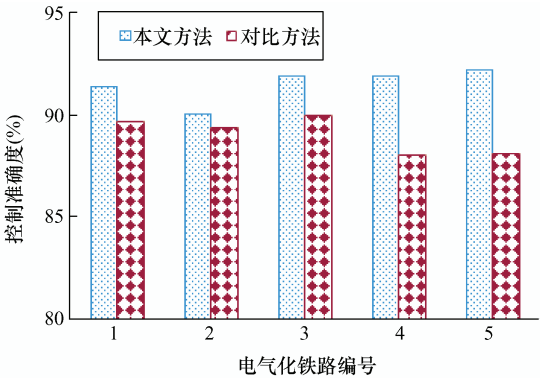

图 6　电气化铁路电能质量控制准确度对比

Fig. 6 Comparison of power quality control accuracy of electrified railway

2.5　不同方法的电气化铁路电能质量控制效率对比

为验证本文提出的电气化铁路电能质量控制效率性，统计两种方法的电气化铁路电能质量控制时间，结果如图 7 所示。分析图 7 可知，相对对比方法，本文所提方法减少了电气化铁路电能质量控制时间，提升了电气化铁路电能质量控制效率。

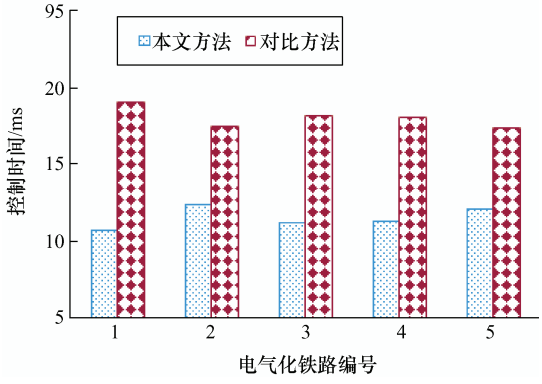

图 7　电气化铁路电能质量控制时间对比

Fig. 7 Comparison of power quality control time for electrified railway

3　结束语

为了准确、实时控制电气化铁路电能质量，本文提出了数据挖掘技术的电气化铁路电能质量控制方法，该方法采用两个步骤完成：一是特征提取；二是电气化铁路电能质量控制。通过构建实验台展开相关测试，结果表明：该方法具备较好的寻优性能，可获取最佳特征结果；具备较好的抗噪性能，并且电气化铁路电能质量控制效果好。**EA**

参考文献

[1] 惠亮亮，张子麒，王江彬，等. 电气化铁路电能质量综合治理实验平台设计[J]. 电力电容器与无功补偿，2020，41(2)：122-127.

[2] 黄小红，赵艺，李群湛，等. 电气化铁路同相储能供电技术[J]. 西南交通大学学报，2020，55(4)：856-864.

[3] 宁志毫，王灿，陈跃辉，等. 铁路统一电能质量控制系统及其工程运行分析[J]. 电气传动，2019，49(10)：70-74.

[4] 陈民武，刘若飞，陈玲，等. 组合式同相供电系统补偿算法与控制策略优化[J]. 电机与控制学报，2019，23(8)：28-34，42.

[5] 王鹏程，李勇，安柏楠，等. 基于不对称接线平衡变压器的电气化铁路电能质量混合调节系统[J]. 电工技术学报，2019，34(21)：4590-4600.

[6] 于丹文，张岩，张青青，等. 计及电气化铁路接入影响的电能质量预测评估方法[J]. 山东电力技术，2020，47(12)：26-31.

作者简介：

徐元成

徐元成(1976—)，男，高级工程师，研究方向：铁路四电。

(收稿日期：2021-1-20)