

数据挖掘的主网继电保护故障识别研究

邬小坤, 万春竹, 牛 静, 石 柯, 齐雪雯

(贵州电网有限责任公司电力调度控制中心, 贵州 贵阳 550002)

摘要: 主网继电保护故障类型多, 而且变化具有显著的随机性, 使得传统技术无法进行准确的主网继电保护故障识别, 导致主网继电保护故障识别成功率较低, 经常出现误识别现象, 为了提高主网继电保护故障识别正确率, 提出基于数据挖掘的主网继电保护故障识别方法。首先根据当前国内外主网继电保护故障识别相关文献, 找到导致主网继电保护故障识别错误多的原因; 然后采集主网继电保护故障数据, 提取主网继电保护故障识别特征; 最后引入数据挖掘技术描述主网继电保护故障类型和特征之间的变化特点, 建立主网继电保护故障识别模型, 并与其他主网继电保护故障识别方法进行了对比测试。该方法的主网继电保护故障识别正确率平均值超过95%, 而且主网继电保护故障识别效率更优, 相对于其他方法, 主网继电保护故障识别结果具有明显的优越性。

关键词: 主网继电保护; 故障识别; 特征提取; 数据挖掘技术; 随机变化

Research on fault identification of main network relay protection based on Data Mining

WU Xiaokun, WAN Chunzhu, NIU Jing, SHI Ke, QI Wenxue

(Power dispatching control center of Guizhou Power Grid Co., Ltd., Guiyang 550002, China)

Abstract: There are many types of relay protection faults in the main network, and the changes are random, which makes the traditional technology unable to carry out accurate relay protection fault identification in the main network, resulting in the low success rate of relay protection fault identification in the main network, and the phenomenon of false identification often occurs. In order to improve the accuracy rate of relay protection fault identification in the main network, a data mining based relay protection fault identification in the main network is proposed. First of all, according the current domestic and foreign related literature of main network relay protection fault identification, the causes of the main network relay protection fault identification errors are found out, then the main network relay protection fault data are collected, the main network relay protection fault identification features are extracted, and finally the data mining technology to describe the main network relay protection fault types and features of the changes between the characteristics are introduced, and the main network relay protection fault identification model established Compared with other main network relay fault identification methods, the accuracy of this method is more than 95%, and the efficiency of main network relay fault identification is better. Compared with other methods, the main network relay fault identification results have obvious advantages.

Key words: main network relay protection; fault identification; feature extraction; data mining technology; random change characteristics

0 引言

随着经济的不断发展, 人们生活水平不断

提高, 家里的电器种类越来越多, 使用更加频繁, 人们对电网稳定性提出了更高的要求^[1-3]。

在实际生活中,电网出现故障是不可避免的。主网继电保护故障识别是进行故障诊断的前提,其可以给故障处理员提供有价值的决策信息,因此提高主网继电保护故障识别效果具有一定的实际和理论意义^[4-6]。

最初主网继电保护故障识别通过专家系统实现,其利用专家推理方法的计算机模型来解决主网继电保护故障识别问题。在实际应用中,专家知识库的建立和完备性验证十分困难,知识库的维护难度大,无法满足大规模主网继电保护故障识别要求,目前主要用于离线的主网继电保护故障识别,应用范围很小^[7-8]。随后有学者提出了基于人工神经网络的主网继电保护故障识别方法,相对于专家系统,人工神经网络的容错能力和学习能力更强,而且主网继电保护故障识别效率更高,但是其同样存在缺陷,如:主网继电保护故障识别效果与样本完备性密切相关,如果样本比较完备,那么主网继电保护故障识别效果好;如果样本不完备,那么主网继电保护故障识别效果差,而主网继电保护故障识别样本的采集十分困难,使得主网继电保护故障识别结果极不稳定^[9-10]。有学者提出了Petri网的主网继电保护故障识别方法,将主网继电保护故障看作是一种离散事件的动态系统,通过电压变化以及一些保护动作实现主网继电保护故障识别,但对于大规模的主网继电保护故障,识别效率极低,而且Petri网的容错能力较差,经常出现主网继电保护故障误识别现象^[11-12]。近年来,数据挖掘技术的研究不断深入,出现了许多有效的方法,如支持向量机就是最常用方法,它没有人工神经网络的主网继电保护故障样本完备性限制,在样本不完备的条件下可以得到较好的主网继电保护故障识别效果,但是其参数与主网继电保护故障识别效果密切相关,如何确定最优参数是当前一个亟需解决的难题。

为了提高主网继电保护故障识别效果,提出了数据挖掘的主网继电保护故障识别方法,并通过具体对比实验分析了本文主网继电保护

故障识别方法的可行性和优越性。

1 数据挖掘的主网继电保护故障识别方法

1.1 支持向量机

支持向量机是一种当前最流行的数据挖掘技术,其工作的基本思想是找一个超平面,将其作为问题的决策平面,这个超平面不但可以正确区分所有样本,同时使每一类样本与超平面之间的距离最大,这样可以将主网继电保护故障识别问题换成一个二次规划问题。设有两类主网继电保护故障,主网继电保护故障特征为 $X_i \in R^m$, $i=1, \dots, n$, n 表示主网继电保护故障的训练样本数量; $\{Y_i\}$, $Y_i \in R$, 且 $Y_i = -1$ 或 1 ,表示主网继电保护故障标签,两类主网继电保护故障的最优超平面如图1所示,其中距离超平面最近的主网继电保护故障样本称为支持向量。

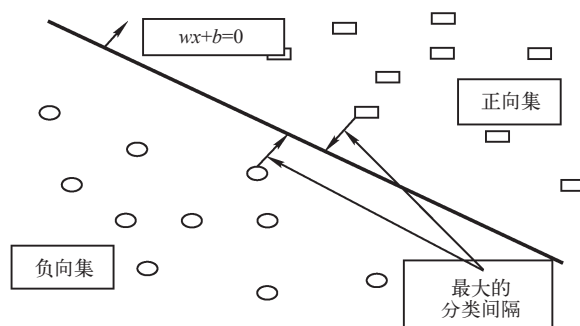


图1 两类主网继电保护故障的最优超平面

Fig. 1 Optimal hyperplane of two types of main network relay protection faults

通过支持向量机的训练得到主网继电保护故障识别最优超平面决策函数为

$$f(x) = \text{sgn}[(w^*x) + b^*] \quad (1)$$

$$= \text{sgn} \sum_{\text{支持向量}} [\alpha_i^* y_i (x_i x) + b^*]$$

其中

$$w^* = \sum_{\text{支持向量}} \alpha_i^* y_i x_i \quad (2)$$

$$b^* = \frac{1}{2} w^* (x_r + x_s) \quad (3)$$

式中, x_r 、 x_s 为支持向量,它们对应的拉格朗日乘子 $\alpha_i > 0$ 。

对应的拉格朗日函数为

$$L(w, b, \alpha) = \frac{1}{2}(ww) - C \sum_{i=1}^n \alpha_i [y_i(w_i x_i + b) - 1 + \xi_i] + \sum_{i=1}^n \mu_i \xi_i \quad (4)$$

式中, C 为惩罚参数。

在支持向量机的主网继电保护故障识别过程中, 首先要选择最合理的核函数 $k(x_i, x_j)$, 其表达式为

$$k(x_i, x_j) = \exp\left(-\frac{|x_i - x_j|^2}{\sigma^2}\right) \quad (5)$$

式中, σ 为核函数的宽度参数。

在支持向量机的主网继电保护故障识别过程中, 其中参数 σ 和 C 直接决定了主网继电保护故障识别效果, 因此本文选择粒子群算法确定最优的主网继电保护故障识别模型参数。

1.2 粒子群算法

共有 N 个粒子, 第 i 个粒子位置和速度向量分别可以表示为 $X_i = (x_{i1}, x_{i2}, \dots, x_{iD})$ 和 $V_i = (v_{i1}, v_{i2}, \dots, v_{iD})$, 第 i 个粒子的当前最优解为 $P_i = (p_{i1}, p_{i2}, \dots, p_{iD})$, 粒子群当前最优解为 $P_g = (p_{g1}, p_{g2}, \dots, p_{gD})$, 它们第 i 个粒子的状态式为

$$\begin{aligned} v_i^{t+1} &= \omega v_i^t + c_1 r_1 (P_i^t - x_i^t) + c_2 r_2 (P_g^t - x_i^t) \\ x_i^{t+1} &= x_i^t + v_i^{t+1} \end{aligned} \quad (6)$$

1.3 数据挖掘的主网继电保护故障识别原理

首先采集主网继电保护故障数据, 并提取主网继电保护故障识别特征; 然后建立主网继电保护故障识别的训练样本集合, 引入数据挖掘技术并描述主网继电保护故障类型和特征之间的变化特点, 建立主网继电保护故障识别模型, 采用粒子群算法对主网继电保护故障识别模型参数进行优化; 最后建立主网继电保护故障识别的测试样本集合, 对主网继电保护故障识别模型性能进行测试, 如图2所示。

2 仿真测试

2.1 主网继电保护故障数据来源

为了分析数据挖掘的主网继电保护故障识别效果, 选择5种类型的主网继电保护故障作

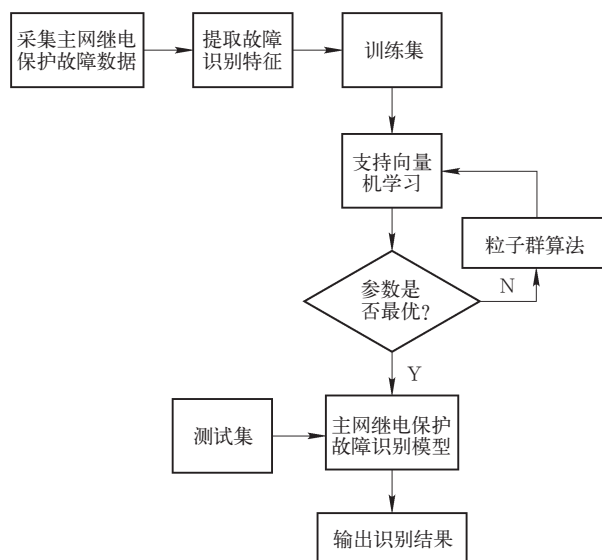


图2 数据挖掘的主网继电保护故障识别原理

Fig. 2 Principle of fault identification of main network relay protection based on Data Mining

为测试样本, 分别采集不同类型的主网继电保护故障数据, 其数据规模如表1所示。

表1 仿真实验用的样本数据

Tab. 1 Sample data for simulation experiment

主网继电保护故障类型标签	训练集
1	15
2	20
3	15
4	20
5	10

2.2 测试平台及对比主网继电保护故障识别方法

分析测试数据挖掘的主网继电保护故障识别方法的优越性, 选择当前经典主网继电保护故障识别方法进行对照测试, 分别为BP神经网络和决策树的主网继电保护故障识别方法, 为了使实验结果具有更好的说服力, 所有方法采用相同的测试平台, 主网继电保护故障识别仿真测试平台设置如表2所示。

2.3 结果与分析

每一类型主网继电保护故障选择5个样本组成测试集, 其他主网继电保护故障数据作为训练集, 建立主网继电保护故障识别模型, 统计不同主网继电保护故障识别方法的实验结果, 如图3所示。其中, 纵坐标表示主网继

表2 主网继电保护故障识别仿真测试平台

Tab. 2 Main network relay protection fault identification simulation test platform

平台参数	具体设置
CPU	Intel Core i5 — 9400F
RAM	镁光 8G DDR4 高频内存
SDD	七彩虹 240 G 固态硬盘
OS	Win10
编程工具	Matlab 2017

电保护故障类型；横坐标表示主网继电保护故障识别的测试样本编号。

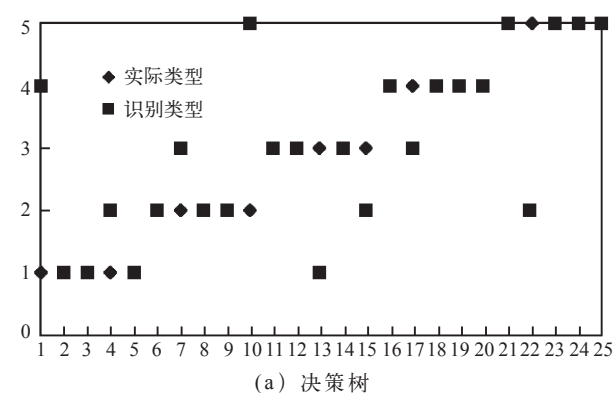
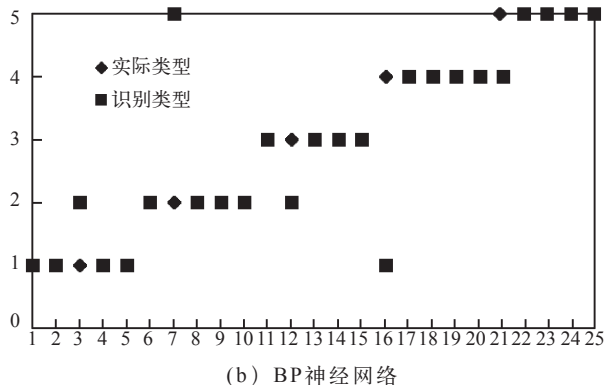
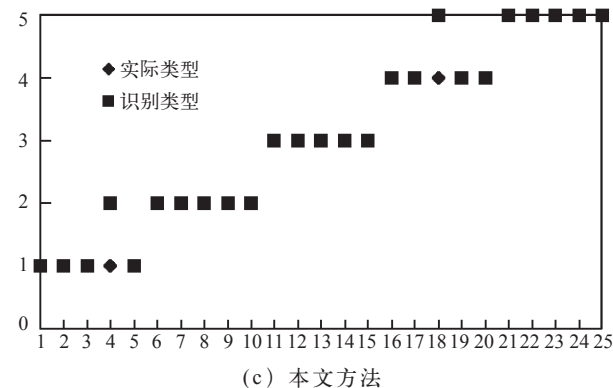


图3 不同方法的主网继电保护故障识别结果比较

Fig. 3 Comparison of fault identification results of main network relay protection with different methods

从图3可以得到如下结论：

1) 决策树的主网继电保护故障识别样本点最多，使得主网继电保护故障识别正确率低，这主要因为决策树是一种线性识别方法，仅可以描述主网继电保护故障和特征之间的线性关系，无法全面描述主网继电保护故障变化特点，使得主网继电保护故障识别结果极不理想。

2) BP神经网络的主网继电保护故障识别结果要好于决策树，主网继电保护故障识别正确率得到了一定的改善，这表明BP神经网络的建模能力也优于决策树，建立更优的主网继电保护故障识别模型，降低了主网继电保护故障识别错误率。

3) 相对于决策树，BP神经网络数据挖掘的主网继电保护故障识别结果更好，大幅度减少了主网继电保护故障识别样本数量，这表明本文方法克服了目前主网继电保护故障识别方法存在的一些不足，获得了十分理想的主网继电保护故障识别结果，同时验证了本文方法的有效性和优越性。

统计不同主网继电保护故障识别方法的平均测试时间，结果如图4所示。对图4分析可知，本文方法的主网继电保护故障识别时间要明显少于BP神经网络和决策树，其中决策树的主网继电保护故障识别时间最长，本文方法可以实现主网继电保护故障线上识别，拓宽了主网继电保护故障识别范围。

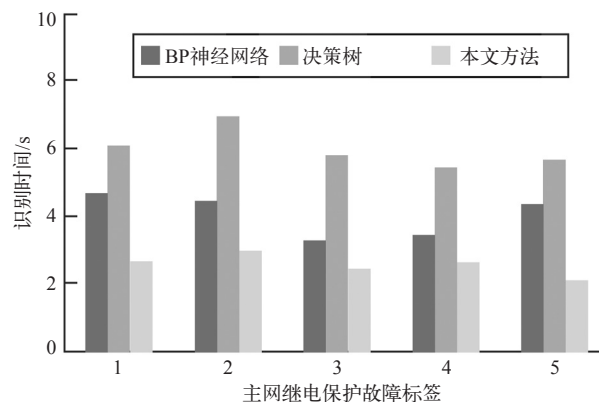


图4 不同方法的主网继电保护故障识别时间比较

Fig. 4 Comparison of fault identification time of main network relay protection with different methods

3 结束语

针对当前主网继电保护故障识别方法存在的局限性,以获得理想的主网继电保护故障识别结果,提出了数据挖掘的主网继电保护故障识别方法,采用数据挖掘中的支持向量机跟踪主网继电保护故障变化趋势,并引入粒子群算法解决主网继电保护故障识别模型参数确定问题。与其他的主网继电保护故障识别对比结果表明,本文方法的主网继电保护故障识别率得到了明显的改善,可以更加准确识别各种主网继电保护故障,而且主网继电保护故障速度也可以满足实际应用的要求,具有很广的应用前景。

参考文献

- [1] 谢民,王同文,徐靖东,等.分布式电源对配网继电保护影响及综合改进保护方案[J].电力系统保护与控制,2019,47(19):78-84.
- [2] 曾奕铭,张慧媛,龚仁敏.一种对配网继电保护影响最小的分布式电源的最优接入方式[J].电测与仪表,2017,54(10):55-62.
- [3] 曾昭炜,段晓瑞,李锦.含分布式电源接入条件下的配网继电保护研究[J].贵州电力技术,2015,18(5):38-39,53.
- [4] 吴义纯,李铁柱,徐结红,等.浅析分布式电源对配网继电保护影响[J].中国设备工程,2019(19):185-186.
- [5] 庄红山,王晓飞,冯小萍,等.基于过电流保护的闭环配网继电保护定值在线校核[J].自动化技术与应用,2019,38(1):102-106.
- [6] 冯小萍,王晓飞,倪宏坤,等.基于区间线

性规划的微网继电保护定值配合优化[J].自动化技术与应用,2018,37(11):109-113.

- [7] 邱金辉,钱海,张道农,等.基于PFIS的继电保护常态特性在线监视与隐性故障诊断[J].电力系统保护与控制,2015,43(8):145-149.
- [8] 熊小伏,陈星田,曾星星,等.基于广义变比辨识的继电保护电流测量回路故障诊断方法[J].中国电机工程学报,2014,34(增刊1):76-84.
- [9] 方勇灵,陆榛,宋福海,等.一种继电保护故障信息智能分析方案及应用[J].电力系统保护与控制,2013,41(5):114-118.
- [10] 高湛军,陈青,王涛,等.基于继电保护时空参数的电网故障诊断模型[J].电力系统自动化,2012,36(13):61-66,91.
- [11] 熊小伏,刘晓放.基于WAMS的继电保护静态特性监视及其隐藏故障诊断[J].电力系统自动化,2009,33(9):11-15,19.
- [12] 吴欣,郭创新,曹一家.基于贝叶斯网络及信息时序属性的电力系统故障诊断方法[J].中国电机工程学报,2005,25(13):14-18.

作者简介



邬小坤

邬小坤(1988-),男,硕士,高级工程师,主要研究方向为电力系统自动化,继电保护。

(收稿日期:2019-09-26) **EA**