

DOI: 10.20040/j.cnki.1000-7709.2024.20232028

基于 CD-BSMOTE 的 D-S 证据融合变压器故障诊断

鲁 玲¹, 高 诚¹, 熊 威², 龚 康², 马 辉¹, 张 鑫¹

(1. 三峡大学电气与新能源学院, 湖北 宜昌 443002; 2. 国网宜昌供电公司, 湖北 宜昌 443000)

摘要: 针对变压器油中溶解气体数据集不均衡特性对故障诊断结果的影响, 提出一种基于清除临界点改进的边界合成少数类过采样算法均衡数据集和 Pearson 冲突距离改进 D-S 证据融合的变压器故障诊断模型。首先, 对少数类样本进行均衡化处理, 根据 K-means 聚类结果清除处于临界位置的样本; 其次, 搭建梯度提升树、随机森林、BP 神经网络的故障诊断模型, 实现变压器故障初步诊断; 接着引入 Pearson 冲突距离改进 D-S 证据融合模型, 实现诊断结果的融合决策; 最后, 经实际算例分析, 诊断精确率达到 92.65%。结果表明, 所建模型能有效解决数据不平衡对诊断结果的影响, 提高故障诊断精度。

关键词: 故障诊断; 油中溶解气体分析; 边界合成少数类过采样; Pearson 冲突距离; D-S 证据融合

中图分类号: TM411

文献标志码: A

文章编号: 1000-7709(2024)05-0192-05

1 引言

变压器对电能的转换和输送起重要作用, 研究变压器故障诊断方法, 对电力系统运行安全具有重要意义^[1]。油中溶解气体分析(DGA)可有效反映变压器早期故障, 在实际工程中应用广泛。目前, 通常将梯度提升树(GBDT)、随机森林(RF)和 BP 神经网络等智能算法应用到变压器故障诊断中, 然而, 单一智能算法难以挖掘出特征关系^[2]。张婷等^[3]基于组合权重的云模型搭建 D-S 证据融合评估模型, 有效描述了不确定性的信息, 实现了变压器状态的准确评估。但未能考虑不确定性信息容易导致描述冲突的现象, 造成融合结果可信度不高。变压器 DGA 数据正常类别与故障类别样本数量不平衡, 导致分类模型诊断结果偏移^[4-6]。因此, 提出使用欠采样和过采样方法均衡数据集, YONG S 等^[7]提出边界合成少数类过采样(BSMOTE)均衡数据集, 增强区域特征。但对非线性数据均衡化处理时, BSMOTE 算法合成样本可能越过类别边界, 造成类别边界模糊。对此, 本文提出清除临界点改进 BSMOTE 数据均衡方法, 考虑单一智能算法提取特征信息的局限性, 搭建 GBDT、RF 和 BP 神经网络的 D-S 证据融合模型, 引入 Pearson 冲突距离对其改进;

最后, 以实际变压器 DGA 数据作为算例, 验证该方法的有效性。

2 数据均衡处理

2.1 边界合成少数类过采样 BSMOTE 算法

BSMOTE 算法基于 SMOTE 的合成样本规则, 引入边界机制改进而来。BSMOTE 算法生成新样本逻辑为: 在少数类数据集中选取样本点, 通过 K 近邻算法搜索其近邻样本。由近邻样本和多数类样本的位置关系生成边界样本集, 对边界集使用 SMOTE 算法随机合成新样本。均衡后的数据集在样本数量上相对一致, 避免分类模型在少数类样本上特征提取不足。BSMOTE 具体步骤如下。

步骤 1 划分边界点集。设小样本数据集为 $X = \{x_1, x_2, \dots, x_n\}$, 利用 K 近邻算法搜索 x_i 样本的 k 个近邻样本, k 个近邻样本属于多数类样本的个数记为 k' 。则有:

$$\begin{cases} k' = k \\ k/2 \leq k' < k \\ 0 \leq k' < k/2 \end{cases} \quad (1)$$

当 $k' = k$, 即样本点 x_i 的 k 个近邻样本全部属于多数类, 则 x_i 被划分为超越边界的噪声点; $k/2 \leq k' < k$, x_i 被视作边界点; $0 \leq k' < k/2$, x_i

收稿日期: 2023-12-03, 修回日期: 2023-12-27

基金项目: 国家自然科学基金项目(52377191); 国网湖北省电力公司管理科技项目(5215H0220002)

作者简介: 鲁玲(1979-), 女, 副教授, 研究方向为电力设备状态监测与智能决策, E-mail: luling@ctgu.edu.cn

被视作少数样本的内点。因此,边界点构成的集合为边界点集 $B = \{b_1, b_2, \dots, b_v\}$ 。

步骤 2 对边界点集 B 中的样本做均衡化处理,新样本合成规则为:

$$b_{\text{new}} = b_j + \lambda(b_k - b_j) \quad (2)$$

式中, b_{new} 为合成的新样本; λ 为区间 $(0, 1)$ 的随机数; b_k 为 b_j 的近邻样本。

2.2 改进 BSMOTE 算法

传统 BSMOTE 算法易导致合成的新样本偏离原始样本特征,产生冗余样本,而影响分类器的训练及判断。因此,引入 K-means 确定原始样本特征区域,清除临界样本改进 BSMOTE 算法(CD-BSMOTE),提高样本合成质量。

利用 K-means 算法对原始样本聚类并确定簇心位置,记 x_{km} 。根据原始样本、合成的新样本与簇心的平均欧氏距离定位临界点,欧氏距离为:

$$d = \frac{1}{n+v} \left[\sum (x_i - x_{\text{km}}) + \sum (b_{\text{new}} - x_{\text{km}}) \right] \quad (3)$$

式中, d 为经均衡后的样本集偏离簇心的平均欧氏距离; n 为原始小样本数量; v 为边界点集数量; x_i 为原始小样本数据集 X 的样本点; b_{new} 为新合成的样本点,并记录 b_{new} 最大偏离距离为 d_{max} 。

若某合成样本偏离簇心的距离 d 满足 $d > (d_{\text{max}} + d)/2$, 则该样本为临界点,不纳入训练样本和测试样本。

3 融合故障诊断模型搭建

3.1 D-S 融合决策模型

D-S 证据融合理论基于概率论原理,用数据集的方式,将多源证据信任度融合为全局信任度,完成模糊性问题决策任务,能够处理因结构差异造成的信任度函数不一致问题。但是对大规模数据融合处理时,模型的计算复杂度升高,对证据体可信度描述易出现矛盾性量化问题。因此,本文选择 PCR5 融合规则,确定 GBDT、RF 和 BP 诊断结果的置信度,完成 3 类故障诊断模型的融合决策。PCR5 规则融合模型的融合步骤如下。

步骤 1 识别空间 Θ 和证据体:

$$\Theta = \{p_1, p_2, \dots, p_k\} \quad (4)$$

式中, p_i 为识别空间集合中的一个元素,这里表示变压器故障类型;证据体为可以描述识别空间的模型,这里指诊断模型。

步骤 2 信任度函数。信任度函数为 D-S 证据融合理论描述识别空间中元素的基本信任度,

记为 mass 函数。本文 mass 函数来自 GBDT、RF 和 BP 的诊断结果, D-S 证据融合的信任度函数确定证据体对某一命题的信任度。

步骤 3 PCR5 融合规则。设识别空间的两个信任度函数 m_1 、 m_2 , 对命题 A 的置信度为 $m(A)$ 。根据 PCR5 融合规则有:

$$M_{12}(A) = \sum \frac{m_1^2(A)m_2(X)}{m_1(A) + m_2(X)} + \frac{m_1(X)m_2^2(A)}{m_1(X) + m_2(A)} \quad (5)$$

$$m_{12}(A) = \sum m_1(A_1)m_2(A_2) \quad (6)$$

$$m(A) = M_{12}(A) + m_{12}(A) \quad (7)$$

其中 $A \cap X = \emptyset; A_1 \cap A_2 = A$

式中, $M_{12}(A)$ 为 PCR5 对命题 A 冲突部分分配的权重; $m_1(A)$ 、 $m_2(A)$ 分别为信任度函数 m_1 、 m_2 对命题 A 的信任度; $m_1(X)$ 、 $m_2(X)$ 分别为信任度函数 m_1 、 m_2 对命题 X 的信任度; $m_{12}(A)$ 为 D-S 证据融合理论对命题 A 的融合信任度; $m(A)$ 为 D-S 证据融合理论对命题 A 的信任度,即分类器对样本数据的诊断结果。

3.2 改进 D-S 融合模型

基于 PCR5 规则的 D-S 证据融合模型处理冲突性证据体时,无法辨识各证据间的重要程度,导致融合结果可信度不高。对此,引入 Pearson 冲突距离,根据各证据间的冲突性确定修正权重。Pearson 冲突距离是基于 Pearson 相关系数^[8]的一种定量描述向量之间冲突性大小的方法, Pearson 冲突距离修正公式为:

$$r_{i,j} = \frac{\text{cov}(m_i, m_j)}{\sqrt{D(m_i)} \sqrt{D(m_j)}} \quad i \neq j \quad (8)$$

$$m'_i = \left(\sum_{i \neq j} 1 - r_{i,j} \right) m_i \quad (9)$$

式中, $\text{cov}(m_i, m_j)$ 为两个信任度函数的协方差; D 为方差; \sqrt{D} 为标准差; m'_i 为 Pearson 距离修正后的信任度函数; $r_{i,j}$ 为 Pearson 相关系数; $1 - r_{i,j}$ 为 Pearson 距离。

3.3 数据均衡的故障诊断模型

在 CD-BSMOTE 算法对原始数据集中的少数类样本进行均衡化处理的基础上,通过 GBDT、RF、BP 算法提取 DGA 数据与故障之间非线性特征关系,输出预测故障概率分布序列,实现故障初次分类。在初次信任度函数分配序列中,分别计算 Pearson 冲突距离,加权修正信任度函数分布序列,缓解证据体间的冲突性,根据 PCR5 规则实现信任度融合,输出结果。基于 CD-BSMOTE 数据均衡的 D-S 证据融合变压器故障诊断模型见

图 1。

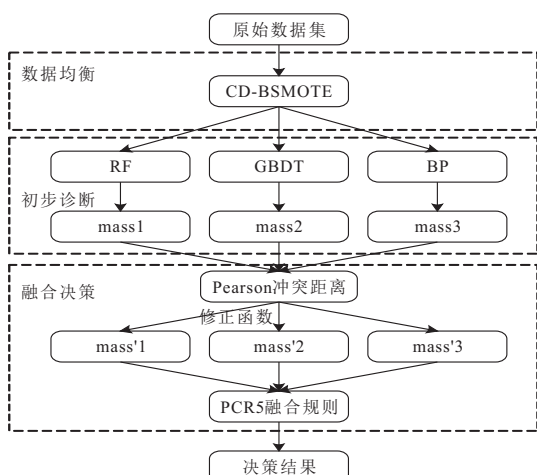


图 1 数据均衡的故障诊断模型

Fig. 1 Troubleshooting model for data equalization

4 案例分析

4.1 试验数据

DGA 原始数据由湖北某供电公司提供的智能变电站在线监测样本共计 703 例, DGA 样本为氢气(H_2)、甲烷(CH_4)、乙烯(C_2H_4)、乙炔(C_2H_2)、乙烷(C_2H_6)特征气体含量, 示例数据见表 1。

表 1 DGA 示例数据(3 例)

Tab. 1 DGA example data

示例	故障状态	五种特征气体含量/($\mu L \cdot L^{-1}$)				
		H_2	CH_4	C_2H_6	C_2H_4	C_2H_2
1	正常	34.20	1.01	0.34	0.12	0.00
2	高能放电	151.00	43.80	15.50	45.30	6.40
3	高温过热	210.00	540.00	250.00	900.00	2.20

参考行业最新执行标准^[9]将 DGA 技术监测的变压器故障分为正常(N)、低能放电(LD)、高能放电(HD)、低温过热(LT)、高温过热(HT)和局部放电(PD), 对应样本数量和状态标签见表 2。

表 2 原始数据集

Tab. 2 Original data set

状态类型	样本数量/个	状态标签	状态类型	样本数量/个	状态标签
正常	203	N	低温过热	97	LT
低能放电	87	LD	高温过热	105	HT
高能放电	93	HD	局部放电	118	PD

4.2 评价指标及参数设置

评价指标精确率 I_{ac} 可评估诊断模型综合分类能力, 与故障判断可信度呈正相关。但对于多分类任务, 精确率 I_{ac} 不能反映模型识别样本的敏感性, 因此采用召回率 I_{recall} 评估模型对样本的识别能力, 该指标与识别灵敏度呈正相关。

本文在一台 Windows10-64 位操作系统的电

脑上进行试验分析, 其装配 CPU 为 Intel(R)Core (TM) i5-7200U@2.50GHz, 机带 RAM 为 8 GB, GPU 为 GeForce940 MAX。试验软件环境中编程语言为 Python3.8, 开发环境为 Pycharm2020。对 GBDT、RF 和 BP 神经网络参数设定见表 3。

表 3 各分类器参数设定值

Tab. 3 Parameter settings for each classifier

分类器	参数	参数设置
GBDT	决策树数量	50
	采样方法	subsample
	损失函数	deviance
RF	决策树数量	30
	采样方法	bootstrap
	节点划分标准	gini 系数
BP	神经元个数/个	128
	学习率	0.01
	优化器	Adam
	最大迭代次数/次	200

4.3 数据均衡试验

由式(1)确定小样本集的边界点集 B, 根据式(2)合成新样本, 最后通过式(3)删除合成新样本中的临界点, 将剩余的新样本与原始样本作为数据集, 直至各状态样本数量与正常状态样本数量一致, 均衡后的数据集共计 1 218 例样本。图 2 中多数类和少数类样本分别为正常状态和低能放电故障下的 CH_4 气体样本。从图 2 中看出, BSMOTE 合成的样本偏离原始数据集中区域, 经改进的 CD-BSMOTE 算法合成的样本更集中分布于原样本区域, 较好保留样本的原始特征。

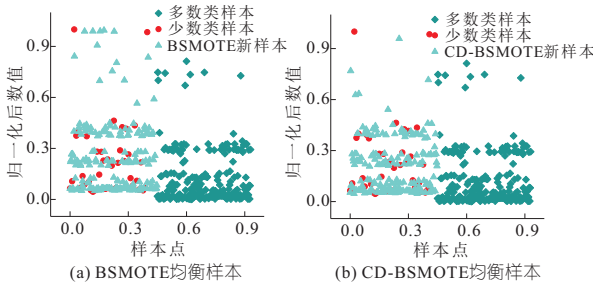


图 2 CD-BSMOTE 均衡 CH_4 样本

Fig. 2 CD-BSMOTE balanced CH_4 sample

选择欠采样算法 NearMiss 和随机欠采样(Random Under-sampling, RUS); 过采样算法 BSMOTE、SMOTE 对比 CD-BSMOTE 合成数据的有效性, 试验中训练集包含 80% 的数据集样本, 测试集包含 20% 的数据集样本。采用 RF 作为分类器首先提取训练集特征信息, 训练分类模型, 然后导入测试集样本, 输出 RF 分类器对测试样本故障状态的预测标签, 完成对测试集样本的故障诊断。诊断结果精确率 I_{ac} 、召回率 I_{recall} 见图 3。

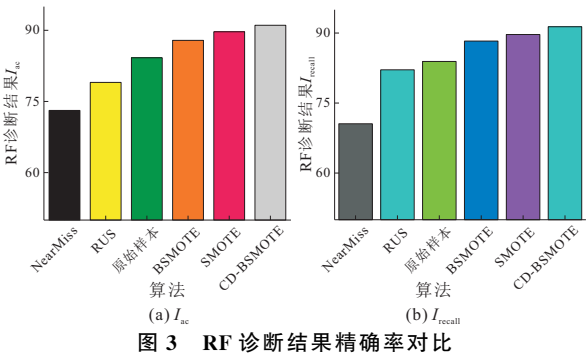


图 3 RF 诊断结果精确率对比

由图 3 可知,采用欠采样算法均衡处理时,诊断结果相较于原始数据集明显降低。因为欠采样算法通过减少抽取多数类样本的方式均衡数据集,使训练的样本较少。因此,分类器因训练不足而造成诊断结果偏低。被过采样算法均衡数据集的试验结果精确率 I_{ac} 和召回率 I_{recall} 均有提升,表明分类器在较均衡的数据集下综合分类能力更好,识别样本的敏感性更优。其中 CD-BSMOTE 清除了临近边界样本,合成样本质量更好,均衡结果在对照组中取得结果最优。

4.4 故障诊断试验

为验证改进 D-S 证据融合故障诊断模型的有效性,与 GBDT、RF、BP 神经网络进行对比分析,试验样本为经 CD-BSMOTE 算法均衡后的数据集,诊断结果见图 4。

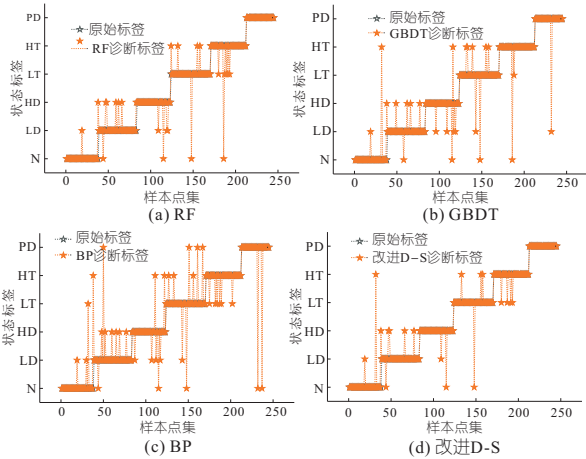


图 4 故障诊断结果

Fig. 4 Fault diagnosis results

由图 4 可知,RF、GBDT、BP 神经网络诊断结果的误判个数分别为 23、24、35 个,诊断精确率分别为 90.61%、90.20%、85.71%。分类器 RF 对高温过热 PD 的故障数据诊断精度为 100%,未出现误判样本。但 GBDT 和 BP 神经网络的诊断结果中却出现 3 例误判样本,表明不同分类器之间存在一定歧义,故采取 D-S 证据融合模型融合对三种分类器进行融合决策。D-S 证据融合模型

诊断结果误判个数和诊断精确率分别为 18 个、92.65%。相较于 RF、GBDT、BP 神经网络的诊断结果误判个数分别减少了 5、6、17 个,诊断精度分别提高了 2.04%、2.45%、6.94%。结果表明,利用 D-S 证据理论融合 RF、GBDT、BP 神经网络分类模型的故障诊断方式能够有效提高变压器的故障诊断精度。

为直观反映 D-S 证据融合模型工作状态,选择表 2 中示例 1 样本进行分析。试验中各诊断模型对示例数据诊断结果服从最大信任规则,即信任度函数的最大值为最终诊断状态。根据最大信任规则,实际状态为 N 的示例样本被 GBDT 分类器诊断为 LD 状态,被 RF 和 BP 神经网络诊断为 N 状态,三种分类器的诊断结果中出现相互冲突的结果。

表 4 为该示例样本在各分类器下信任度函数,GBDT 诊断结果 m_1 ,RF 诊断结果 m_2 ,BP 诊断结果 m_3 ,根据式(5)~(7)求出基于 PCR5 规则 D-S 证据融合结果 $m(1)$,根据式(8)算出 m_1 、 m_2 、 m_3 之间的 Pearson 相关系数 $r_{1,2}=0.522$ 、 $r_{1,3}=0.069$ 、 $r_{2,3}=0.471$,由式(9)可得出改进 D-S 模型融合结果 $m'(1)$ 。由 $m'(1)$ 可知改进 D-S 模型诊断结果与实际状态一致,说明引入 Pearson 冲突距离修正置信度函数,可有效缓解证据体间冲突性,准确诊断样本真实状态。

表 4 改进 D-S 证据融合模型融合结果

fusion model			%					
序号	分类器	状态	信任度函数					
			N	LD	HD	LT	HT	PD
示例 1	m_1	LD	21.13	65.37	1.17	0.51	0.35	11.47
	m_2	N	37.00	30.00	0	0	0	33.00
	m_3	N	33.45	13.14	13.13	13.13	13.13	14.02
	$m'(1)$	N	56.47	23.04	0	0	0	20.49

5 结论

a. 提出了一种基于 CD-BSMOTE 的 D-S 证据融合变压器故障诊断方法,弥补了传统诊断模型处理不平衡数据能力不足的问题。

b. 采用清除临界点的方式改进 BSMOTE 算法,提高样本合成质量,实现数据集均衡,改善变压器故障诊断训练样本的均衡性。

c. 采用 PCR5 融合规则实现 GBDT、RF、BP 神经网络分类模型的融合诊断。引入 Pearson 冲突距离修正信任度函数对 D-S 证据融合模型进行改进,降低证据体间的冲突性。算例验证表明,

诊断精确率达到了 92.65%。该方法对比其他故障诊断方法具有更高的精确率,该模型误判率更低,诊断可靠性更好。

参考文献:

- [1] 王文森,贺馨仪,杨晓西,等. 基于多参量数据回归分析的电力变压器状态监测方法[J]. 电网与清洁能源, 2023, 39(4): 83-90.
- [2] 陈伟根,刘娟,曹敏. 基于信息融合的变压器内部故障诊断方法[J]. 高电压技术, 2015, 41(11): 3797-3803.
- [3] 张婷,赵峰,高锋阳. 基于云模型和 D-S 证据理论的配电变压器健康状态评估[J]. 水电能源科学, 2020, 38(3): 181-185.
- [4] 于海平,吴雪琼,杜天硕. 不平衡数据集下海量配电网量测终端故障诊断方法[J]. 电力系统及其自动化学报, 2023, 35(4): 32-40.
- [5] 余松,胡东,唐超,等. 基于 TLR-ADASYN 平衡化数据集的 MSSA-SVM 变压器故障诊断[J]. 高电压技术, 2021, 47(11): 3845-3853.
- [6] 谢国民,王嘉良. 基于混合采样与 IHBA-SVM 的变压器故障辨识方法[J]. 电子测量与仪器学报, 2022, 36(12): 77-85.
- [7] YONG S, HUAKUN Q, QIANQIAN C, et al. Borderline SMOTE algorithm and feature selection-based network anomalies detection strategy [J]. Energies, 2022, 15(13): 4751.
- [8] 赵源上,林伟芳. 基于皮尔逊相关系数融合密度峰值和熵权法典型场景研究[J]. 中国电力, 2023, 56(5): 193-202.
- [9] 国家能源局. 变压器油中溶解气体分析和判断导则: DL/T 722-2014[S]. 北京:中国电力出版社, 2015.

CD-BSMOTE Based D-S Evidence Fusion Transformer Fault Diagnosis

LU Ling¹, GAO Cheng¹, XIONG Wei², GONG Kang², MA Hui¹, ZHANG Xin¹

(1. College of Electrical Engineering and New Energy, China Three Gorges University, Yichang 443002, China;

2. State Grid Yichang Power Supply Company, Yichang 443000, China)

Abstract: Aiming at the solving the unbalanced characteristics of the dissolved gas data set in transformer oil on the fault diagnosis results, a transformer fault diagnosis model is proposed based on the fusion of critical removal improved boundary synthesis minority class oversampling algorithm equalized data set and Pearson conflict distance improved D-S evidence. Firstly, the minority class samples are equalized, and the samples in critical position are removed according to K-means clustering results. Secondly, the fault diagnosis model of gradient boost decision tree, random forest, and BP neural network is built to realize the preliminary diagnosis of transformer faults. Then, Pearson conflict distance is applied to improve the D-S evidence fusion model to realize the fusion decision of preliminary diagnosis results. Finally, after analyzing the cases, the precision rate of the diagnosis results reached 92.65%. The results show that the proposed model can effectively eliminate the influence of data imbalance on the diagnostic results and improve the fault diagnosis precision.

Key words: fault diagnosis; dissolved gas analysis; borderline synthetic minority over-sampling technique; Pearson conflict distance; D-S evidence fusion

(上接第 165 页)

Traceability Analysis and Protection of Pump Stop Water Hammer Accident in Pump Station

MA Quan¹, JIANG Li-yun¹, LI Zhong-qing²

(1. Zhuzhou Zhuhua Smart Water Technology Co., Ltd., Zhuzhou 412000, China;

2. Lanshan County Hydropower Co., Ltd., Yongzhou 425000, China)

Abstract: In order to study the influence of the closing time and type of the two-stage slow-closing check valve on the pump-stopping water hammer, the pumping station in the Gaotangping separated pumped storage project in Lanshan County is taken as an example. Based on the working principle of two-stage slow closing check valves, a numerical simulation method is used to trace and analyze the causes of pump stop water hammer accidents, and to compare the protective effects of hydraulic control butterfly valves and multi-functional water pump control valves under different working conditions on pump stop water hammer under the same total valve closing time. The results show that the unreasonable valve closing law of the hydraulic control butterfly valve is the main cause of the water hammer accident. Compared with the hydraulic control butterfly valve, the adaptive valve closing mode of the one-stage zero flow rate of the multifunctional pump control valve under different working conditions can significantly reduce the water hammer peak under the pump stopping condition. The water hammer curve of stopping pump after replacing the multi-functional pump control valve is obtained by field measurement. The multi-functional pump control valve effectively solves the problem of water hammer protection of pumping station and improves the safety resilience of pumping station.

Key words: separated pumped storage power station; pump stop water hammer; traceability; multifunctional water pump control valve; hydraulic control butterfly valve