

继电保护系统的状态评估与风险评估

汪胜和¹, 叶远波¹, 刘宏君², 谢民¹, 丛雷², 陈军²

(1. 国网安徽省电力有限公司, 合肥 230022; 2. 长园深瑞继保自动化有限公司, 广东 深圳 518057)

摘要:为提升继电保护系统评估准确性以及评估方法的适应性,提出了一种基于半监督的马氏距离机器学习算法的继电保护系统风险评估模型。首先,针对智能变电站的网络拓扑,对继电保护系统评价指标进行了分析;其次,在层次分析法的基础上,将评价结果作为训练集进行模糊综合评价,从准确性、处理时间和可行性方面对比有监督的多元回归分析算法、无监督的 k-means 算法与文章提出的新算法的性能。对比结果表明,提出的算法评估准确性能比另外两种算法好,且适应能力更强。

关键词:继电保护;风险评估;半监督马氏距离;准确率

DOI:10.15938/j.jhust.2023.01.011

中图分类号: TM774 **文献标志码:** A **文章编号:** 1007-2683(2023)01-0088-09

State Evaluation and Risk Assessment for Relay Protection System

WANG Shenghe¹, YE Yuanbo¹, LIU Hongjun², XIE Min¹, CONG Lei², CHEN Jun²

(1. State Grid Anhui Electric Power Limited Company, Hefei 230022, China; 2. CYG SUNRI Co., Ltd. Shenzhen 518057)

Abstract: In order to improve the evaluation accuracy of the relay protection system and the adaptability of the evaluation method, a risk evaluation model of the relay protection system based on the semi-supervised Mahalanobis distance machine learning algorithm is proposed. First, according to the network topology of the smart substation, the evaluation indicators of the relay protection system are analyzed; second, on the basis of the analytic hierarchy process, the evaluation results are used as the training set for fuzzy comprehensive evaluation. The performance of supervised multiple regression analysis algorithm and unsupervised k-means algorithm is compared with that of the new algorithm proposed in this paper in terms of accuracy, processing time and feasibility. The comparison results show that the proposed algorithm has better evaluation accuracy and stronger adaptability compared to the other two algorithms.

Keywords: relay protection; risk assessment; semi-supervised Mahalanobis distance; accuracy

0 引言

近年来,越来越多的智能电力电子设备和新能源设备接入电力系统,提高电力系统设备维护效率对于电力系统稳定运行具有重要意义^[1-2]。

继电保护装置的动作正确、快速、可靠将有效避免电力系统故障后扩大故障范围,从而有效提高电

力系统的稳定性,减少经济损失^[3-5]。文[6]提出了一种由于隐藏故障引起的继电保护系统故障概率模型。ORNL-PSerc-Alaska(OPA)模型是橡树岭国家实验室提出的一种停电模型。文[7]利用 OPA 模型,以风险值(value at risk, VaR)和条件风险值(conditional VaR, CVaR)为例,分析了停电对继电保护 VaR 和 CVaR 的影响。文[8]使用改进 9 总线系统和适用于电力系统极端情况的 IEEE 69 总线系统

收稿日期: 2021-08-26

基金项目: 国家电网公司科技项目(521200190081); 国家科技重大专项(2018YFB0904900)。

作者简介: 汪胜和(1971—),男,硕士,高级工程师;

叶远波(1973—),男,硕士,教授级高级工程师。

通信作者: 刘宏君(1974—),男,硕士,教授级高级工程师, E-mail: liuhj@sznari.com。

研究了所提出的风险评估方法的有效性。文[9] 解决了在谐波污染环境中运行的保护继电器的可靠性评估问题,提出了寻找简化数字距离中继模型的马尔可夫曲线的转换率矩阵的问题解决方案。文[10] 提出了一种全面的马尔可夫可靠性模型,对在谐波污染环境中运行的继电器保护组件的功能状态进行分类。

国内学者陈海涛等^[11]提出了一种改进的遗传算法优化支持向量机参数的分类方法,利用云模型的随机性和稳定的趋向性来实现健康度与评语域之间的不确定性映射,使得评估结果更加符合实际情况。王文焕等^[12]提出了基于故障树和贝叶斯网络的继电保护故障风险评估方法,实现了继电保护系统故障原因的后验定位及溯源。汪洋^[13]提出了一种保护装置多维度状态评估法,在建立装置多维度状态量数据库的基础上,通过 K 近邻算法对其进行训练和识别,以实时识别装置异常状态;但此方法仅在理论中进行了分析研究,未能在实际操作中进行验证。郭采珊^[14]构建智能变电站继电保护系统的网络拓扑,从信息在网络拓扑中传输的角度,用改进简化的可达矩阵算法分析信息在源宿之间的可达性,合理区分不同故障的风险影响大小;此方法在研究过程中给存在一定的主观性。肖金星^[15]综合考虑了设备历史运行信息和状态监测信息,提出了一种基于比例强度模型 (proportional intensity model, PIM) 的继电保护系统故障风险成本优化评估方法;此方法评估计算时间较长。

通过对国内外研究现状汇总,现有方法有不可忽视的局限性,主要存在于以下几方面:需要先进的算法来提高模式识别的准确性;评价方法的适应性和自我修正能力差;方案主观性强,没有太多数据支持;某些模型的模拟结果可能对实际操作和维护帮助不大。

为了解决现有评价方法的不足,本文研究了一种基于半监督的马氏距离 (Mahalanobis distance, MD) 机器学习算法的继电保护系统风险评估模型。本文的主要内容包括:①表征和量化继电保护系统的状态评估模型,以提供考虑动态和静态指标的训练集。②研究 MD 机器学习算法,得到更适合实际情况的继电保护系统风险评估结果。③比较分析有监督、无监督和半监督机器学习方法的诊断结果。④根据评估结果提出相应的维护策略。

1 继电保护系统状态评价

结合南方电网公司某地区继电保护设备运行记录,本节提出了基于层次分析法 (analytic hierarchy process, AHP) 的继电保护系统运行状态评价模型。计算结果可以作为机器学习算法的训练集,从而有助于在定量计算结果的基础上制定维护策略。

1.1 评价指标

根据 DL/T860 工程实施标准,继电保护系统智能变电站的网络拓扑如图 1 所示,分为三层,即站控层、间隔层和过程层^[16]。

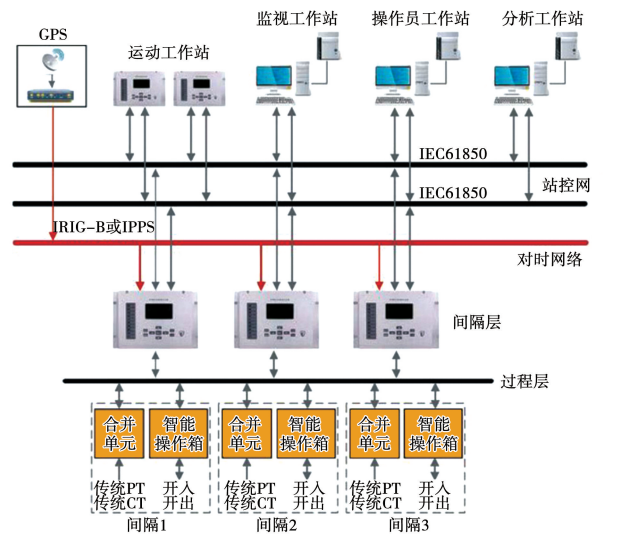


图 1 继电保护系统的网络拓扑

Fig. 1 Network topology of the relay protection system

站控层:站控层包括后台监视器、遥控机、时钟系统、通信电源系统等,主要实现对运行状况、报警信息、CPU、内存利用率、硬盘利用率、网络接口等的监控功能。

间隔层:间隔层设备包括保护装置、测控装置、故障录波仪、网络分析仪等。其主要功能是监测 A/D、通用面向对象变电站事件 (GOOSE)、采样值 (SV) 和 I/O。

过程层:过程层由合并单元、智能终端、互感器等组成,其主要任务是采集时钟匹配、芯片状态、光功率等信息。

参考国家电网《继电保护状态评价导则》和《变电站继电保护信息规范》^[17],选择了反映继电保护系统运行状态的静态和实时参数。某厂家生产的 P 型线路保护装置的评价指标如表 1 所示。

1.2 指标权重值

1.2.1 继电保护系统评价指标

继电保护系统运行考核指标由当前运行状态和历史状态数据。微机保护装置的运行状态来自于装置的自检,主要包括 CPU 状态、采样状态、存储器状态、定值自检以及装置的开入和开出状态等。历史状态数据为一段时间内的装置运行信息,为装置运行状态的历史值。

继电保护的评估状态包括:正常、告警、故障与隐患^[18-19]。

1.2.2 判断矩阵

不同的二次设备,其评价指标的个数不一致,评价指标用式(1)表示^[20]:

$$\mathbf{X} = [x_1 \ x_2 \cdots x_n] \tag{1}$$

式中: n 为某种二次设备总的评价指标个数; x_i 为第 i 个指标值。

表 1 线路保护的评价指标

Tab. 1 Evaluation indicator of line protection		
序号	指标值	等级
1	ROM 故障	5
2	RAM 故障	5
3	定值自检出错	4
4	开入异常	4
5	PT 断线告警	2
6	CT 断线告警	2
7	MMS 通信中断	2
8	接收光功率	3
9	SV 接收中断	3
10	SV 接收告警	2
11	识别码不对应	1
12	时钟自检出错	1
13	GOOSE 接收中断	3
14	GOOSE 接收告警	2
15	Flash 故障	5
16	配置自检错误	4
17	AD 自检出错	5
18	装置内通信出错	3
19	电源电压	3
20	装置温度	3
21	对时错误	1

为了得到包含 n 个指标的 X 对 Y 的影响程度,通过两两比较建立重要度比较矩阵 $\mathbf{A} = (a_{ij})_{n \times n}$,矩阵 $\mathbf{A}$ 中的值的定义参见表 2, $\mathbf{A}$ 中元素的 a_{ij} 定义为第 i 个指标值相对于第 j 个指标值的重要度, $a_{ij} = \frac{1}{a_{ji}}$ 。

表 2 1-9 标度法
Tab. 2 Definition for 9 scales

值	含义	备注
9	极端重要	严重故障类
7	重要得多	一般故障类
5	明显重要	严重告警
3	稍微重要	一般告警
1	同样重要	状态数据

求解矩阵 $\mathbf{A}$ 的最大特征根 $\lambda_{\max}$ 和所对应的特征向量 $\mathbf{W}$ 。

将矩阵中的元素归一化,如式(2):

$$\bar{a}_{ij} = \frac{a_{ij}}{\sum_{k=1}^n a_{kj}} \quad (i, j = 1, 2, \cdots, n) \tag{2}$$

对归一化的矩阵求和,如式(3):

$$\bar{w}_i = \sum_{j=1}^n \bar{a}_{ij} \quad (i = 1, 2, \cdots, n) \tag{3}$$

将 $\bar{w}_i$ 归一化,求特征向量 $\mathbf{W}$,如式(4):

$$\left. \begin{aligned} \mathbf{W} &= [w_1, w_2, w_3, \cdots, w_n] \\ w_i &= \frac{\bar{w}_i}{\sum_{j=1}^n \bar{w}_j} \quad (i = 1, 2, \cdots, n) \end{aligned} \right\} \tag{4}$$

最大特征值通过式(5)求出:

$$\lambda_{\max} = \frac{1}{n} \sum_{i=1}^n \left[\frac{\sum_{j=1}^n a_{ij} w_j}{w_i} \right] \tag{5}$$

1.2.3 权重计算及一致性检查

只有判断矩阵满足一致性检验准则,才认为权重计算结果是正确的。判别矩阵的一般一致性指标 CI 定义为

$$CI = \frac{\lambda_{\max} - n}{n - 1} \tag{6}$$

判别矩阵的随机一致性比率 (consistency ratio, CR) 定义为

$$CR = \frac{CI}{RI} \tag{7}$$

其中: RI 为平均随机一致性指标值,参见表 3。

表 3 平均随机一致性指标值

Tab. 3 Index value of average random consistency					
N	RI	N	RI	N	RI
1	0	11	1.52	21	1.638 5
2	0	12	1.54	22	1.640 3
3	0.58	13	1.56	23	1.646 2
4	0.90	14	1.58	24	1.649 7
5	1.12	15	1.59	25	1.655 6
6	1.24	16	1.594 3	26	1.658 7
7	1.32	17	1.606 4	27	1.663 1
8	1.41	18	1.613 3	28	1.667 0
9	1.46	19	1.620 7	29	1.669 3
10	1.49	20	1.629 2	30	1.672 4

如果 $CI < 0.1$, 认为判别矩阵具有满意的一致性, 说明权重分配是合理的; 否则需要调整判别矩阵, 直到取得满意的一致性。

某厂家生产的 P 型线路保护装置的判断矩阵可通过变换矩阵得出如式(8):

$$A = \begin{bmatrix} 1 & 1 & 3 & 3 & \cdots & 9 \\ 1 & 1 & 3 & 3 & \cdots & 9 \\ 0.33 & 0.33 & 1 & 1 & \cdots & 7 \\ 0.33 & 0.33 & 1 & 1 & \cdots & 7 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0.11 & 0.11 & 0.14 & 0.14 & \cdots & 1 \end{bmatrix}_{21 \times 21} \quad (8)$$

计算出的 P 型线路保护装置的指标权重如表 4 所示, $\lambda_{\max} = 21.922\ 1$, $CI = 0.046\ 1 < 0.1$ 。

表 4 线路保护的指标权重计算

Tab. 4 Index weight calculations offline protection			
序号	权重	序号	权重
1	0.125 0	12	0.008 1
2	0.125 0	13	0.033 0
3	0.066 8	14	0.015 5
4	0.066 8	15	0.125 0
5	0.015 5	16	0.066 8
6	0.015 5	17	0.125 0
7	0.015 5	18	0.033 0
8	0.033 0	19	0.033 0
9	0.033 0	20	0.033 0
10	0.015 5	21	0.008 1
11	0.008 1	—	—

1.3 模糊综合评价

考虑到变量之间存在模糊关系, 在评价继电保护系统运行状态时, 实际运行不能简单地用 0-1 逻辑来描述。模糊关系应该用连续区间 $[0, 1]$ 来表示, 而不是用 0(假)和 1(真)的二元逻辑来表示。

1.3.1 模糊综合评判特征域

继电保护模糊特征域包括指标集、权重集和评价集。用 $I = \{i_1, i_2, \dots, i_n\}$ 表示继电保护系统状态评估的指标集; 根据层次分析法可以计算得到指标权重值, 表示为 $W = \{w_1, w_2, \dots, w_n\}$ 。假设评估集由 m 个评估结果组成, 那么可表示为 $E = \{e_1, e_2, \dots, e_n\}$ 。

1.3.2 数据归一化处理

将评估指标进行数据归一化处理, 将实际值映射至区间 $[0, 1]$ 。指标越靠近 1, 则更优。根据状态评价指标的不同, 将评价指标分为效益型、成本型和区间型三类。

对效益型指标, 价值越大越好。其映射函数可用式(9)表示:

$$x^* = \begin{cases} 0 & x < x_{\min} \\ \frac{x - x_{\min}}{x_{\max} - x_{\min}} & x_{\min} \leq x \leq x_{\max} \\ 1 & x > x_{\max} \end{cases} \quad (9)$$

对成本型指标, 成本越小越好。其映射函数可用式(10)表示:

$$x^* = \begin{cases} 1 & x < x_{\min} \\ \frac{x_{\max} - x}{x_{\max} - x_{\min}} & x_{\min} \leq x \leq x_{\max} \\ 0 & x > x_{\max} \end{cases} \quad (10)$$

对区间型指标, 比如光功率指标, 规定最优运行区间 $[x_{\min_opt}, x_{\max_opt}]$ 和标准规定运行区间 $[x_{\min}, x_{\max}]$ 。在最优区间取值为 1, 超出标准规定运行区间取值为 0。中间线段连接。其映射函数用式(11)表示:

$$x^* = \begin{cases} 0 & x < x_{\min} \text{ or } x > x_{\max} \\ \frac{x - x_{\min}}{x_{\min_opt} - x_{\min}} & x_{\min} \leq x \leq x_{\min_opt} \\ 1 & x_{\min_opt} \leq x \leq x_{\max_opt} \\ \frac{x_{\max} - x}{x_{\max} - x_{\max_opt}} & x_{\max_opt} \leq x \leq x_{\max} \end{cases} \quad (11)$$

1.3.3 模糊综合隶属函数

隶属度属于模糊评价函数里的概念。在工程实践中, 三角分布和岭分布隶属函数都已得到了广泛的应用^[21-22]。三角分布的隶属度函数和岭分布相

比,当指标处于高隶属度位置时,岭分布的隶属度信息更高,更符合继电保护的分类特征。本文采用岭分布隶属函数。继电保护系统的运行状态分为正常、警告、错误和临界 4 个状态,采用岭分布隶属度函数为

$$h_{r_1}(u)=\begin{cases}1 & u\leqslant 0.2 \\ 0.5-0.5\sin\pi(u-0.3)/0.2, & 0.2<u\leqslant 0.4 \\ 0 & u>0.4\end{cases}\quad (12)$$

$$h_{r_2}(u)=\begin{cases}0 & u\leqslant 0.2\cup u>0.6 \\ 0.5+0.5\sin\pi(u-0.3)/0.2, & 0.2<u\leqslant 0.4 \\ 0.5-0.5\sin\pi(u-0.5)/0.2 & 0.4<u\leqslant 0.6\end{cases}\quad (13)$$

$$h_{r_3}(u)=\begin{cases}0 & u\leqslant 0.4\cup u>0.8 \\ 0.5+0.5\sin\pi(u-0.5)/0.2, & 0.4<u\leqslant 0.6 \\ 0.5-0.5\sin\pi(u-0.7)/0.2 & 0.6<u\leqslant 0.8\end{cases}\quad (14)$$

$$h_{r_4}(u)=\begin{cases}0 & u\leqslant 0.6 \\ 0.5+0.5\sin\pi(u-0.7)/0.2, & 0.6<u\leqslant 0.8 \\ 1 & u>0.8\end{cases}\quad (15)$$

结合式(12)~(15),岭分布隶属度函数计算模型如图 3 所示。

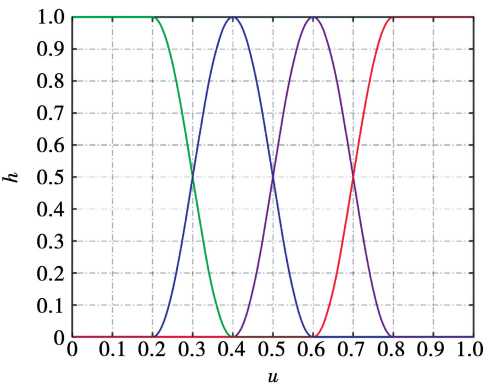


图 2 岭分布隶属度函数

Fig. 2 Ridge distribution membership function

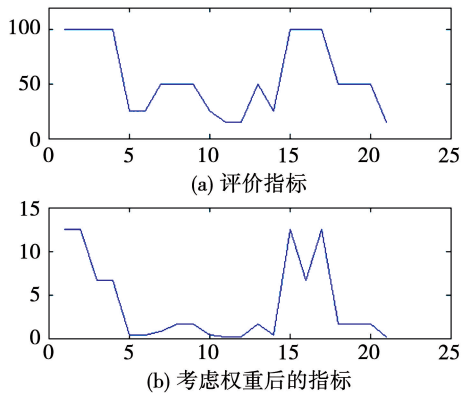


图 3 线路保护运行状态评估图

ig. 3 Operating state assessment diagram of line protection

1.3.4 获取训练集

本文所用的数据信息取自国网安徽省电力公司 2018–2021 年统计数据。以合肥市某几个 220 kV 变电站的 100 个装置为例进行分析。

用于评估线路保护装置的指标数量 $n=21$; 评价等级类型 $m=4$; 权重集参见表 3。

保护装置的数据标准化处理标准如表 5 所示。线路保护运行状态评估计算如图 3 所示。图 3(a) 显示了保护装置的每个指标的对应关系,图 3(b) 显示了考虑了权重后的关系。

表 5 保护装置数据标准化

Tab. 5 Data standardisation criteria of protection device					
序号	权重	$x_{\max}$	$x_{\min}$	$x_{\max_opt}$	$x_{\min_opt}$
1	0.125 0	1	0	0	0
2	0.125 0	1	0	0	0
3	0.066 8	1	0	0	0
4	0.066 8	1	0	0	0
5	0.015 5	1	0	0	0
6	0.015 5	1	0	0	0
7	0.015 5	1	0	0	0
8	0.033 0	−10	−40	−9	−37
9	0.033 0	1	0	0	0
10	0.015 5	10	0	0	0
11	0.008 1	1	0	0	0
12	0.008 1	10	0	0	0
13	0.033 0	1	0	0	0
14	0.015 5	10	0	0	0
15	0.125 0	1	0	0	0
16	0.066 8	1	0	0	0
17	0.125 0	1	0	0	0
18	0.033 0	5	0	0	0
19	0.033 0	5.5	4.5	5.25	4.75
20	0.033 0	70	25	55	35
21	0.008 1	50	0	0	0

本节为机器学习算法提供的训练集由 100 个线路保护装置评估结果组成,如表 6 所示。

S 为机器学习训练集的规模,表示整个样本的比例。训练集包含序列为[1,10 000]的样本。表 5 展示了序列为样本组成的测试集。

表 6 保护装置评估结果

Tab. 6 Assessment result of protective devices

评估时间	指标 1	...	指标 21	得分
2020. 05. 15	0	...	4	94. 99
2020. 06. 11	0	...	10	90. 93
2020. 06. 28	0	...	5	88. 36
⋮	⋮		⋮	⋮
2020. 12. 15	0	...	20	94. 29

2 继电保护系统运行状态的机器学习算法

采用 AHP 模糊综合评价法对线路保护装置进行评价。假设被评估的装置数量为 d , 评估周期为 t , p 代表总样本的比例。 dtp 表示单独设备的训练集数量, $dt(1-p)$ 是指每种设备的一个测试集的数量。为了克服模糊综合算法的主观性缺陷, 本文采用了新的机器学习的方法, 使计算结果更加客观真实。

2.1 监督多元回归分析算法

针对不同类型的设备, 保护装置、M-C 装置、合并装置和智能终端的多元回归方程为

$$Y=f(X,B,\varepsilon)$$
 (16)

其中: Y 是不同设备的评价结果; X 是评价指标; B 为回归系数; ε 是随机误差。

为了使历史评估周期中的平方和最小, 提出了如下优化目标函数:

$$\min RSS = \sum_{j=1}^n \varepsilon_j^2 = \sum_{j=1}^n (Y_j - \hat{Y}_j)^2$$
 (17)

其中: Y_j 为观察到的保护装置值; $\hat{Y}_j$ 为保护装置的估计值。考虑到极值原理, 约束函数可以定为式 (18), 其中 j 是训练集的编号。将 100 台保护装置的 100 个评估周期的统计数据作为一个训练集, 训练集比例 $p=80\%$, 利用训练集的监督多元回归分析算法可以得到残差水平图, 当显著性水平 $\alpha=0.05$ 时, 残差图中有 391 个异常样本点, 合格率为 95. 11%, 如图 4(a) 所示。

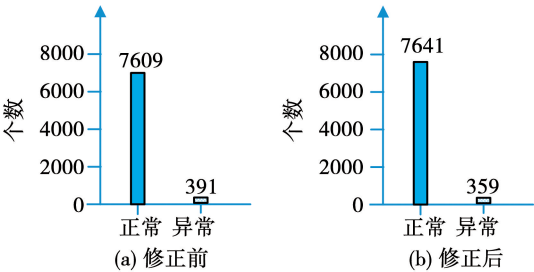


图 4 残差图

Fig. 4 Residual level diagram

s. t.
$$\begin{cases} \frac{\partial RSS}{\partial b_0} = -2 \sum [Y_1 - (b_0 + b_1x_{1j} + b_2x_{2j} + \dots + b_{24}x_{24j})] = 0 \\ \frac{\partial RSS}{\partial b_1} = -2x_1 \sum [Y_1 - (b_0 + b_1x_{1j} + b_2x_{2j} + \dots + b_{24}x_{24j})] = 0 \\ \vdots \\ \frac{\partial RSS}{\partial b_{24}} = -2x_{24} \sum [Y_1 - (b_0 + b_1x_{1j} + b_2x_{2j} + \dots + b_{24}x_{24j})] = 0 \end{cases}$$
 (18)

经过数据清理后的修正残差。修正后的回归模型中的异常样本数为 359, 合格率提高到 95. 51%。如图 4(b) 所示。

2.2 基于欧式距离的无监督 K-means 算法

采用无监督 K-means 算法对线路装置进行评价。对训练集中样本 X_a 和 X_b 之间的欧式距离定义如式 (19):

$$|X_a - X_b|^2 = \sum_{j=1}^n (x_{aj} - x_{bj})^2$$
 (19)

聚类目标函数 D 的定义如式 (20):

$$\min D = \sum_{k=1}^K \sum_{l=1}^{n_k} |X_l - \bar{X}_k|^2$$
 (20)

其中: K 为聚类数目, 在线路保护装置的状态评估中, $K=4$; n_k 为样本编号; $\bar{X}_k$ 为第 k 个族类中样本的均值。

基于欧式距离的无监督 K-means 算法流程如图 5 所示。算法迭代过程结束, 得到最终分组结果。

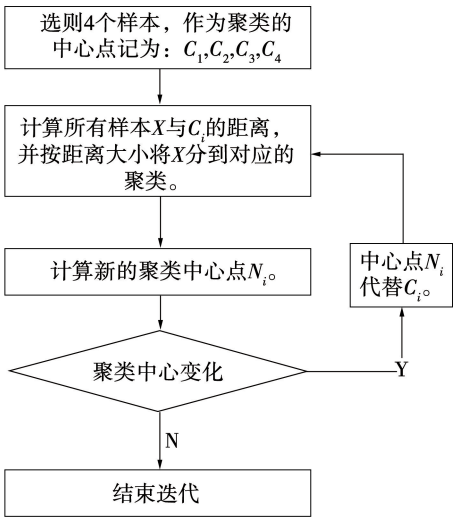


图 5 计算流程图

Fig. 5 Flow chart of calculation

2.3 基于半监督的 MD 算法

与欧式距离相比,基于 MD 算法可用协方差 $\sigma_{ij} = \text{cov}(x_i, x_j)$ 来描述变量之间的相关性,而不受维度的影响。它在处理多变量高维样本之间的距离度量学习方面具有优势。

为了充分利用未标记数据中包含的有用信息,本研究提出了在线顺序 ELM-MD (OSELM-MD)。通过使用具有可逆推理能力的半监督学习算法,将正确标记样本应用于新半标记样本的扩展,判断错误的训练样本被标记为无效样本,然后对参数进行修正。

训练集 T 包含标记集 T_1 和未标记集 T_2 ,半监督 OSELM-MD 的成本函数表示为

$$E = \sum_{j=1}^{N_0} \left(\sum_{i=1}^{\tilde{N}} \beta_i g(\mathbf{w}_i \cdot \mathbf{x}_j^0 - \mathbf{b}_i) - \mathbf{t}_j^0 \right)^2 + \sum_{j=1}^{N_e} \left(\sum_{i=1}^{\tilde{N}} \beta_i g(\mathbf{w}_i \cdot \mathbf{x}_j^e - \mathbf{b}_i) - \mathbf{t}_j^e \right)^2 \quad (21)$$

式中: N_0 为标记样本的数量; $\tilde{N}$ 为隐式层节点的数量; $\mathbf{x}_j^0 = [x_{j1}^0, \dots, x_{jm}^0]^T$ 为标记的样本向量; n 为样本向量的维数; $\mathbf{t}_j^0 = [t_{j1}^0, \dots, t_{jm}^0]^T$ 为 x_j^0 的类标记样本向量; C 为网络的输出节点数; $\mathbf{w}_i = [w_{i1}, \dots, w_{in}]^T$ 为输入节点和隐式层节点之间的输入权重向量; $\mathbf{b}_i$ 为节点 i 的偏移量; g 为隐式层节点的激活函数; $\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{ic}]^T$ 为隐式层节点与输出节点之间的输出权重向量; N_e 为扩展样本的数量; $\mathbf{x}_j^e = [x_{j1}^e, \dots, x_{jm}^e]^T$ 为扩展样本向量; $\mathbf{t}_j^e = [t_{j1}^e, \dots, t_{jm}^e]^T$ 为 x_j^e 的类标记样本向量。

通过计算,得到满足式(22)所示的 $\hat{\beta}$:

$$\begin{aligned} \|H_0 \hat{\beta} - T_0\| + \|H_e \hat{\beta} - T_e\| &= \min_{\beta} (\|H_0 \beta - T_0\| + \|H_e \beta - T_e\|) = \\ &= \min_{\mathbf{w}_i, \mathbf{b}_i, \beta_i} \left\| \begin{bmatrix} g(\mathbf{w}_1 \mathbf{x}_1^0 + \mathbf{b}_1) \cdots g(\mathbf{w}_{\tilde{N}} \mathbf{x}_1^0 + \mathbf{b}_{\tilde{N}}) \\ \vdots \\ g(\mathbf{w}_1 \mathbf{x}_{N_0}^0 + \mathbf{b}_1) \cdots g(\mathbf{w}_{\tilde{N}} \mathbf{x}_{N_0}^0 + \mathbf{b}_{\tilde{N}}) \end{bmatrix}_{N_0 \times \tilde{N}} \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_{\tilde{N}}^T \end{bmatrix} - \begin{bmatrix} t_1^{0T} \\ \vdots \\ t_{N_0}^{0T} \end{bmatrix} \right\| \\ &+ \left\| \begin{bmatrix} g(\mathbf{w}_1 \mathbf{x}_1^e + \mathbf{b}_1) \cdots g(\mathbf{w}_{\tilde{N}} \mathbf{x}_1^e + \mathbf{b}_{\tilde{N}}) \\ \vdots \\ g(\mathbf{w}_1 \mathbf{x}_{N_e}^e + \mathbf{b}_1) \cdots g(\mathbf{w}_{\tilde{N}} \mathbf{x}_{N_e}^e + \mathbf{b}_{\tilde{N}}) \end{bmatrix}_{N_e \times \tilde{N}} \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_{\tilde{N}}^T \end{bmatrix} - \begin{bmatrix} t_1^{eT} \\ \vdots \\ t_{N_e}^{eT} \end{bmatrix} \right\| \end{aligned} \quad (22)$$

式中: H_0 为标记样本集的隐藏层输出矩阵; β 为权重的输出矩阵; T_0 为标记样本的类别标记矩阵; H_e 为扩展样本的隐藏层输出矩阵; T_e 为扩展样本的类别标记矩阵。

总体样本 G 指数为 $m = 21$, 高维数据样本的均值向量为 $\mu = [\mu_1, \mu_2, \dots, \mu_{24}]^T$, 总体样本 G 与单个

样本 $\mathbf{X} = [x_1, x_2, \dots, x_{24}]^T$ 之间的 MD 是:

$$d^2(\mathbf{X}, G) = (\mathbf{X} - \mu)^T \Sigma (\mathbf{X} - \mu)^{-1} \quad (23)$$

评价指标之间的协方差为

$$\Sigma = (\sigma_{ij})_{m \times m} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1m} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2m} \\ \vdots & \vdots & & \vdots \\ \sigma_{m1} & \sigma_{m2} & \cdots & \sigma_{mm} \end{bmatrix}_{m \times m} \quad (24)$$

其解公式表示为

$$\begin{cases} \sigma_{ij} = \frac{1}{n-1} \sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j) \\ \bar{x}_i = \frac{1}{n} \sum_{k=1}^n x_{ki} \\ \bar{x}_j = \frac{1}{n} \sum_{k=1}^n x_{kj} \end{cases} \quad (i, j = 1, 2, \dots, m)$$

2.4 对比分析

与欧式距离相比,基于 MD 算法可用协方差 $\sigma_{ij} = \text{cov}(x_i, x_j)$ 来描述变量之间的相关性,而不受维度的影响。它在处理多变量高维样本之间的距离度量学习方面具有优势。

为了验证这些算法的有效性和准确性,对上述有监督、无监督和半监督机器学习方法进行了比较和分析。为保证对比分析的公平性,在同一台电脑上、同一样本完成仿真测试。

以线路保护装置为例,当训练集的规模为 $p = 80\%$ 时,3 种算法的 ROC (receiver operating characteristic) 特性评价曲线结果如图 6 所示。

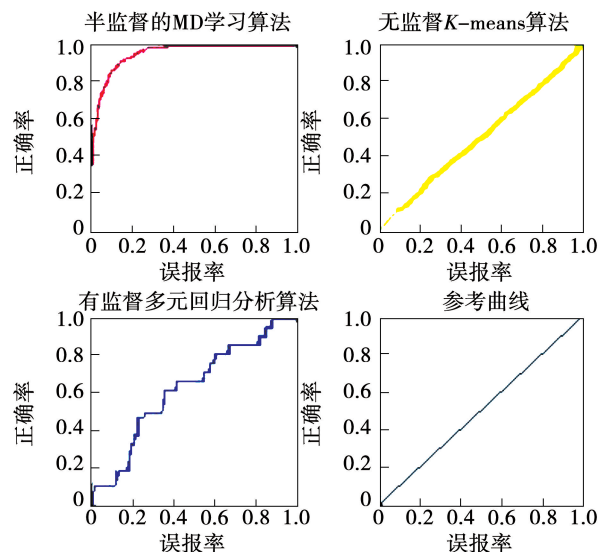


图6 ROC 评估曲线

Fig. 6 ROC evaluation curve

ROC 曲线越靠近左上角,曲线下所覆盖面积 (area under the curve, AUC) 的值越接近 1, 意味着相

应的机器学习方法具有更高的预测性能。通过分析得知,有监督多元回归分析算法的 AUC 值为 $0.6084 \in (0.5,1)$,则预测结果优于随机预测,具有一定的参考意义。基于 ED 的无监督 K -means 算法的 AUC 值为 $0.4952 \sim 0.5$,即预测结果类似于随机预测,无参考意义。基于半监督的 MD 学习算法的 AUC 值为 $0.9464 \sim 1$,即该算法的预测结果非常好,结果较为准确。表 7 展示了 3 种算法在精度、准确度和可靠性的对比分析结果。

表 7 线路保护装置评估结果

Tab. 7 Assessment result offline protection

算法	训练精度	测试集准确度/%	AUC
回归分析	83.69	84.56	0.6084
欧式距离	63.56	61.31	0.4952
MD	94.56	93.06	0.9464

由表 6 可知,与基于半监督的 MD 学习算法相比,有监督的多元回归分析算法和基于 ED 的无监督 K -means 算法的准确率远低于基于半监督的 MD 学习的算法(AUC 值在 0.95 左右,明显优于其他两种方法)。

另外,考虑到训练集的规模是影响机器学习算法精度的主要因素之一,故在同一样本集中选择了 9 个规模在 10% 到 90% 之间的训练子集,其精度分析结果如图 7 所示。

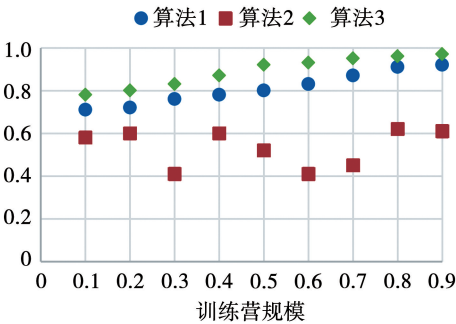


图 7 机器学习算法比较分析

Fig. 7 Machine learning algorithm comparison analysis

观察回归分析算法和基于半监督的 MD 学习算法的变化趋势可以看出训练集规模对计算精度有影响,当样本不足时,精度对训练集的尺度敏感。 $\Delta r = 1 - r_i/r_{\max}$ ($i = 10\%, 20\%, \dots, 90\%$) 代表算法的敏感阈值。当训练集的规模超过 70% 时,算法 1 (多元回归分析算法) 的评估精度将对训练集规模不再

敏感($\Delta r_1 < 5\%$)。当训练集的规模超过 60% 时,算法 3 (基于半监督的 MD 学习算法) 的评估精度对训练规模不再敏感($\Delta r_3 < 5\%$)。纵观算法 2 (基于 ED 的无监督 K -means 算法) 的精度变化趋势,其准确率不受训练规模的限制。从图中可以看出,算法 3 对训练集规模的要求最低。当尺度达到 60% 时,即满足获得优异评价精度的要求。

3 结 论

本文提出了一种基于半监督 MD 机器学习算法的继电保护系统风险评估模型。首先,为机器学习方法提供训练集,选择某型的线路保护装置建立了基于 AHP 模糊综合理论的状态评估模型;其次,为了使评估结果更加客观、准确,将有监督多元回归分析、无监督 K -means 和半监督 MD 算法等机器学习算法应用于继电保护系统的状态评估,并进行了比较。仿真结果表明,本文提出的机器学习算法具有较高的精度和较低的数据规模要求。

参考文献:

[1] 宋旭. 继电保护系统对一次设备可靠性的影响[J]. 通信电源技术, 2021, 38(4): 203.
SONG Xu. Influence of Relay Protection System on Reliability of Primary Equipment [J]. Telecom Power Technology, 2021, 38(4): 203.

[2] 叶远波, 刘宏君, 黄太贵, 等. 变电站继电保护设备状态检修可靠性分析方法研究[J]. 电力系统保护与控制, 2021, 49(10): 170.
YE Yuanbo, LIU Hongjun, HUANG Taigui, et al. Research on the Reliability Analysis Method of Condition-based Maintenance of Relay Protection Equipment in Substation [J]. Power System Protection and Control, 2021, 49(10): 170.

[3] 曹政, 程冀川, 尹相国, 等. 基于继电保护状态分析的电网故障诊断[J]. 现代电子技术, 2018, 41(13): 147.
CAO Zheng, CHENG Jichuan, YIN Xiangguo, et al. TPower Grid Fault Diagnosis Based on Relay Protection State Analysis [J]. Modern Electronics Technique, 2018, 41(13): 147.

[4] 李文学. 继电保护与安全稳定控制系统隐性故障研究综述及展望[J]. 现代工业经济和信息化, 2018, 8(6): 80.
LI Wenxue. Review and Prospect of Hidden Failures in Relay Protection and Safety and Stability Control Systems [J]. Modern Industrial Economy and Informationization, 2018, 8(6): 80.

- [5] 刘一民,章家欢,杨心平,等. 基于宏观微观的继电保护设备评估体系构建方法[J]. 电网技术, 2020, 44(8): 3090.
LIU Yimin, ZHANG Jiahuan, YANG Xinping, et al. Construction Method of Evaluation Index System of Relay Protection Based on Macroscopic and Microcosmic View [J]. Power System Technology, 2020, 44(8): 3090.
- [6] Tuyou S, Jiekang W, Weideng Y, et al. Power Supply Risk Assessment Method for Relay Protection System Faults [J]. Archives of Electrical Engineering, 2016, 65(4): 803.
- [7] MEI S, HE F, ZHANG X, et al. An Improved OPA Model and Blackout Risk Assessment [J]. IEEE Transactions on Power Systems, 2009, 24(2): 814.
- [8] LIU X, SHAHIDEHPOUR M, CAO Y, et al. Risk Assessment in Extreme Events Considering the Reliability of Protection Systems [J]. IEEE Transactions on Smart Grid, 2015, 6(2): 1073.
- [9] JEDRZEJCZAK J, ANDERS G J. Transition Rates Assessment of Protective Relay Reliability Models with Incomplete Data [J]. IEEE Transactions on Power Systems, 2016, 32(1): 809.
- [10] JEDRZEJCZAK J, ANDERS G J, Fotuhi-Firuzabad M, et al. Reliability Assessment of Protective Relays in Harmonic-polluted Power Systems [J]. IEEE Transactions on Power Delivery, 2016, 32(1): 556.
- [11] 陈海涛, 张学敏, 戴伟康, 等. 基于 GA 优化 SVM 参数与云模型的继电保护装置状态评估方法[J]. 智慧电力, 2020, 48(7): 88.
CHEN Haitao, ZHANG Xuemin, DAI Weikang, et al. State Assessment Method of Relay Protection Device Based on GA Optimizaztion SVM Parameters and Cloud Model [J]. Smart Power, 2020, 48(7): 88.
- [12] 王文焕, 郭鹏, 祝洁, 等. 基于故障树及贝叶斯网络的继电保护系统风险评估及故障定位方法[J]. 电力科学与技术学报, 2021, 36(4): 81. WANG Wenhuan, GUO Peng, ZHU Jie, et al. Fault Positioning and Risk Assessment Method of Relay Protection Based on Fault Tree and Bayesian Network [J]. Journal of Electric Power Science and Technology, 2021, 36(4): 81.
- [13] 汪洋. 电力系统继电保护装置多维度状态评估法[J]. 中国新技术新产品, 2020(11): 64. WANG Yang. Multi-Dimensional State Evaluation Method for Power System Relay Protection Devices [J]. China New Technologies and New Products, 2020(11): 64.
- [14] 郭采珊, 蔡泽祥, 潘天亮, 等. 基于信息可达性的智能变电站继电保护系统风险评估方法[J]. 电网技术, 2018, 42(9): 3041.
GUO Caishan, CAI Zexiang, PAN Tianliang, et al. Risk Assessment for Protection System in Smart Substation Considering Information Reachability [J]. Power System Technology, 2018, 42(9): 3041.
- [15] 肖金星, 王法, 熊芬芳, 等. 基于比例强度模型的继电保护系统风险成本优化评估方法[J]. 电气应用, 2020, 39(2): 47. XIAO Jinxing, WANG Fa, XIONG fenfang, et al. The Risk Assessment Optimization Method for the Relay Protection System Based on Proportional Intensity Model [J]. Electrotechnical Application, 2020, 39(2): 47.
- [16] ZHANG Y, WANG L, XIANG Y, et al. Inclusion of SCADA Cyber Vulnerability in Power System Reliability Assessment Considering Optimal Resources Allocation [J]. IEEE Transactions on Power Systems, 2016, 31(6): 4379.
- [17] GB/T 50976-2014, 继电保护及二次回路安装及验收规范[S].
- [18] 南东亮, 王维庆, 张陵, 等. 基于关联规则挖掘与组合赋权-云模型的电网二次设备运行状态风险评估[J]. 电力系统保护与控制, 2021, 49(10): 67.
NAN Dongliang, WANG Weiqing, ZHANG Ling, et al. Risk Assessment of the Operation State of Power Grid Secondary Equipment Based on Association Rule Mining and Combination Weighting-cloud Model [J]. Power System Protection and Control, 2021, 49(10): 67.
- [19] 陈勇, 李胜男, 张丽, 等. 基于改进平均秩次法的保护装置可靠性评估[J]. 电力系统保护与控制, 2020, 48(2): 165.
CHEN Yong, LI Shengnan, ZHANG Li, et al. Reliability Evaluation of Protective Device Based on Improved Average Rank Order Method [J]. J Power System Protection and Control, 2020, 48(2): 165.
- [20] SUN C C. A Performance Evaluation Model by Integrating Fuzzy AHP and Fuzzy TOPSIS Methods [J]. Expert Systems with Applications, 2010, 37(12): 7745.
- [21] 杨国生, 戴飞扬, 王文焕, 等. 基于灰度关联法和 TOPSIS 法的继电保护状态评估综合算法研究与应用[J]. 中国电力, 2019, 52(2): 94.
YANG Guosheng, DAI Feiyang, WANG Wenhuan, et al. Research and Application of Comprehensive Algorithm of Relay Protection Status Assessment Based on Gray Correlation Analysis and TOPSIS Method [J]. Electric Power, 2019, 52(2): 94.
- [22] 徐岩, 王士铭, 刘婧妍, 等. 继电保护装置的一种新型评估方法[J]. 电测与仪表, 2018, 55(5): 119.
XU Yan, WANG Shiming, LIU Jingyan, et al. A New Evaluation Method for Relay Protection Device [J]. Electrical Measurement & Instrumentation, 2018, 55(5): 119.

(编辑: 温泽宇)