

一种基于粒子群算法的配电网低电压诊断模型研究

李占英¹, 马福兰², 马伟兵³

- (1. 国网青海省电力公司黄化供电公司, 青海 尖扎 811200;
- 2. 国网青海省电力公司海北供电公司, 青海 海晏 812200;
- 3. 国网青海省电力公司海南供电公司, 青海 共和 813000)

摘 要:随着智能电网建设的发展,传统的基于检测技术的配电低电压原因诊断已变成基于数据挖掘的电力大数据分类技术,而着眼于低电压故障原因的数据分类研究在国内尚处于起步阶段,为此该文提出一种采用改进聚类算法和支持向量机分类算法的配电网低电压诊断模型。该模型首先采用 Canopy-Kmeans 的聚类算法基于配电网历史运行数据进行低电压原因的聚类分析并得出可能存在的低电压原因,然后采用经粒子群算法对支持向量机数据分类算法进行参数优化,最后使用结果参数优化的支持向量机算法对智能电表所采集的配电网实时运行数据进行低电压原因分类并最终输出低压故障原因的诊断结果。实验表明,采样基于粒子群优化的支持向量机诊断模型能够实现 90% 的低电压原因诊断准确度。

关键词:低电压; Canopy-Kmeans; 粒子群优化(PSO); 支持向量机(SVM)

文章编号:2096-4633(2019)05-0049-07 **中图分类号:**TM744 **文献标志码:**B

电压是电能质量的测量标准之一。配电网低压问题不仅严重影响社会经济发展和人民生活,也对优化低压投资方案、明确低电压投资方向、为低压管理提供决策支持具有重要意义^[1]。因此如何有效建立电网低压故障诊断模型是本文研究的重点。

目前,国内学者和电工主要依靠经验或简单的数据图表进行分析。由于缺乏分析大量电力数据的能力,因此对低压发生的原因识别和分类存在较大困难。文献[2]分析了农村低压问题的成因,指出主要问题是配电网薄弱,无功补偿容量不足,供电半径长,供电能力不足配电变压器不足,低压线路供电能力不足,新设备和新技术几乎没有投入应用。文献[3]总结了短时持续低电压的原因,解释了不同原因的持续低电压现象的电压特性。

近年来,随着智能电表应用的大规模推广和现代配电网的积极建设,系统每天都会产生大量的能源使用、设备运行等方面的数据。这使得电能质量研究逐渐成为数据密集型的科学研究。一些学者已经开始将大数据和数据挖掘方法应用于电能质量数据的存储和电能质量的预测。其中,文献[4]设计

并实现了电能质量的大数据云平台。该平台可以存储和分析大功率质量数据。文献[5]介绍了数据挖掘及其在电能质量分析中的应用。文献[6]提出了一种基于数据挖掘技术的稳态电能质量预测预测方法。尽管上述研究为数据挖掘构建了大型电能质量数据平台或电能质量数据,但目前少有针对低电压成因诊断的研究。

针对上述问题,本文提出一种采用了改进聚类算法和 PSO-SVM 的低电压诊断模型。该模型在使用 Canopy 算法对 K-means 聚类算法进行改进的基础上,实现对低电压原因的分类和识别,然后依据智能电表所采集的电力数据通过 PSO-SVM 算法实现对可能出现的低电压事件及其成因的预测。数值模拟验证了该模型的可行性。

1 诊断流程

低电压原因的诊断主要分为四个步骤。首先,建立基础数据模型。然后由 Canopy-Kmeans 算法对低电压的原因进行聚类。之后对低电压的原因进行识别。最后,利用聚类结果作为训练样本对分类模型进行训练,从而实现对低电压的诊断。具体过程如图 1 所示。

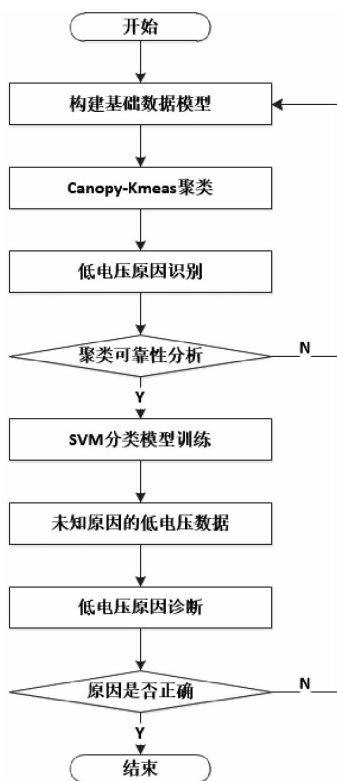


图 1 低电压原因诊断的流程

Fig. 1 Flow of low voltage cause diagnosis

由诊断流程可以发现,该模型能否成功准确和有效实现低电压的诊断,主要难点是聚类算法和分类算法的参数整定上。因此,接下来分别对聚类算法和分类算法的流程和参数整定方法进行重点研究。

2 Canopy-K 均值聚类模型

当发生低电压现象时,会出现许多体现里电压特征的迹象。在传统的方法中,利用实时采集到的测量值来检测低电压现象,并根据相关准则推导出低电压现象的原因。但随着信息化程度的不断提高,越来越多的数据源可以用于低压现象的分析。一方面,数据处理能力受到这些多维和大量实时和非实时数据的挑战。另一方面,需要使用新的以数据为中心的分析来找出超出低电压标准但与低电压现象有关的数据中包含的潜在信息。

本文采用 MapReduce 并行计算框架实现 Canopy-Kmeans 算法,并应用于低电压聚类。

2.1 算法思路

Canopy-Kmeans 算法是一种优化的 Kmeans 算法。该算法的思想是设置初始中心点和面积半径。然后将数据集划分成几个重叠的子集(Canopy),这样

所有对象都属于 Canopy Coverage 范围。重新计算落在同一区域内的物体上的新中心点,并根据物体与新中心点之间的距离重新划分物体的面积。重复“划分 Canopy、计算中心点”的过程,直到 k 点的位置不再变化^[7-9]。然后实现目前状态的稳定分类。

首先,根据特征将所有数据集分为 K 类,每个数据成员具有相同的特征。根据数据存储节点和 Mapper 核的数量将初始数据分为 $N_1, N_2, \dots, N_m$ 个数据集。每个数据集可以独立分配给映射器节点。选择集群中的一个节点作为主节点来执行并行任务的调度。每个映射器内核并行计算相应的任务以通知主节点。然后,主节点调度相应的 Reduce 并返回结果以确定是否进行下一次迭代。

随后通过 Canopy 聚类技术分两个阶段实施聚类。第一步是将数据快速大致分成称为 Canopy 的子集。然后用精确的计算重新聚集每个 Canopy 内的点。它在两个阶段使用不同的距离度量方法来形成重叠 Canopy 中心。在创建 Canopy 之后,第二步是使用 Kmeans 算法对 Canopy 内的点进行聚类。现在只需要关注覆盖聚类中的点,从而避免了所有数据点的传统聚类算法进行精确计算。另外,还有重叠的子集也增加了算法的容错性,并消除了孤立点的作用^[9,10]。Mahout 是一个开源的机器学习软件库。本文基于 Mahout 库所提供的成熟 Canopy 和 Kmeans 算法实现低电压成因的聚类。

2.2 聚类过程

首先是对低压电力数据进行预处理,使其数据格式满足 Canopy-Kmeans 算法的要求。在将输入数据实例化为 Vector 并填充值后,将所有 Vector 文件写入 SequenceFile 格式,以便 Mahout 算法可以读取它们。SequenceFile 是 Hadoop 库中的一种文件格式,由一个键值对组成,必须在 Hadoop 库中作为 WritableComparable 实现。

预处理后的数据用作在 Mahout 库中运行 Canopy-Kmeans 算法的输入,从而产生一些可划分类别的数据集合,并且 Canopy 聚类不需要指定簇的数量。中心的数量仅由距离测量值 T_1 和 T_2 的选择 ($T_1 > T_2$) 确定。对于每个点,它计算到列表中每个 Canopy 聚类中心的距离。如果距离小于 T_1 ,则将 Canopy 聚类中心加入其中。如果距离小于 T_2 ,则将其从数据集合中移除,以便在下一个周期中创建一个新的 Canopy 聚类。

设置欧几里德距离度量的距离度量。测距时根据不同的应用要求设置阈值 T1 和 T2。由于在文本聚类时稀疏文档向量距离会很大,所以 T1 和 T2 的值需要设置得更大以获得有意义的聚类群集。

在聚类过程中使用 Canopy 聚类改进 K-means 聚类算法。初始 k 聚类中心被 Canopy 算法产生的 Canopies 中心所替代。由于所有的数据点都被 Canopy 聚类覆盖和划分,所以在计算数据聚类中心时不需要计算到所有 k 聚类中心的距离,只需要计算到同一 Canopy 聚类中的 k 个中心的距离以提高算法效率。完成聚类后可以得到不同潜在原因的低电压数据分类。

2.3 聚类结果分析

聚类结果文件夹有 clusteredPoints, clusters-N 和数据文件。存储在 ClusterPoints 中的最终聚类是键值类型。通过 seqdumper 读取的 ClusterPoint 的结果是 (IntWritable, WeightedVectorWritable), 对应于输入数据的顺序, 标记每个数据点的关键值, 聚类中心值。clusters-N 是第 N 个簇的聚类结果, 其数据格式通常是 VL-VXX {n = nxx c = [xx. xxx, xx. xxx, ...] r = [x. xxx, x. xxx]}, 其中 n 是一个类的样本数, c 是每个属性的中心, r 是每个属性的半径。例如, VL-462 {n = 197 c = [29.984, 29.681, ...] r = [3.602, 3.558, ...]} 意味着这个聚类有 197 个数据点, c[...] 表示中心点, r[...] 表示簇半径, clusters-N 的聚类结果类型为(文本, 簇)。

值键是聚类集合的中心和簇。将簇的数据和聚类点的数据结合起来, 得到每个类中包含的聚类结

$$L(w, b, \xi, \alpha) = \frac{1}{2} \|w\|^2 + c \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i [y_i (w^T \varphi(x_i) + b) - 1 + \xi_i] \quad (4)$$

在式 4 中, $\alpha_i (i = 1, \dots, n)$ 是拉格朗日乘子。上述优化问题转化为以下二次规划问题:

$$\max W(\alpha) = -\frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j) + \sum_{i=1}^n \alpha_i \quad (5)$$

约束条件为:

$$\sum_{i=1}^n \alpha_i y_i = 0, 0 \leq \alpha_i \leq c, i = 1, \dots, n \quad (6)$$

$K(x, y)$ 为核函数, $K(x, y) = \varphi(x) \varphi(y)$, SVM 采用核函数法将低维数据非线性映射到高维。常见的核函数有高斯核函数、多项式函数、Sigmoid 函数以及线性函数等^[15]。由于高斯核函数具有很强的泛化能力, 所以本文选用高斯核函数, 其表达式为:

果和数据点。可以用这个方法来分析低电压电力数据聚类的结果。

3 基于粒子群优化的支持向量机分类模型

支持向量机(SVM)是一种常用机器学习方法已被广泛应用于模式识别问题的预测, 分类和求解。在具体的应用过程中, 如何选择关键参数是一个关键问题^[11,12]。参数的选择决定了它的学习能力和泛化能力。目前, 支持向量机参数优化的方法有网格法, 梯度下降法, 遗传算法和粒子群算法^[12,13]。其中粒子群优化(PSO)是一种新的人工智能计算技术。PSO 方法具有收敛快速和只需要少量参数进行调整的优势^[14], 因此本文使用 PSO 对 SVM 参数进行优化整定。

3.1 算法思路

给定训练样本 $\{x_i, y_i\} (i = 1, 2, \dots, n)$ 是样本总数, d 是空间 R^d 的维数, $y_i^1 \in R$ 。非线性映射 $\varphi(\cdot)$ 用于将样本从原始空间映射到高维(k 维, $k > d$) 特征空间。特征空间里的最佳线性回归函数为:

$$f(x) = w^T \cdot \varphi(x) + b \quad (1)$$

其中, x 是样本向量, w 是权向量, b 是分类阈值, 当 SVM 使用误差(允许松散变量的松弛度)作为优化目标中的损失函数, 则优化问题归纳为:

$$\min \varphi(w) = \frac{1}{2} \|w\|^2 + c \sum_{i=1}^n \xi_i \quad (2)$$

约束条件为:

$$y_i [w^T \varphi(x_i) + b] \geq 1 - \xi_i, \xi_i \geq 0, i \in [1, n] \quad (3)$$

其中 c 是惩罚因子, 用于控制模型的复杂性。采用拉格朗日方法对上述优化问题进行求解:

$$K(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right) \quad (7)$$

3.2 算法流程

由于 SVM 是利用高斯核函数构造的, 所以模型的参数是惩罚因子 c 的大小、核宽度参数 σ 和函数拟合误差 ε 。其中函数拟合误差 ε 的大小对支持向量机影响不大, 所以本文只考虑 σ 参数和 c 参数。

首先定义粒子群算法的适应度函数, 将直接响应向量机性能的均方误差(MSE)定义为适应度函数。适应度函数为:

$$\sqrt{\sum_{i=1}^N \left(\frac{(y' - y_i)^2}{N} \right)} \quad (8)$$

其中, y 是预测值, y_i 是实际值, N 是训练样本的数量。粒子群算法基于同一组局部信息和不同种群全局信息的交换和迭代实现最优搜索,即最优个体位置向量对应关键参数的最佳组合。

输出参数的最佳组合是 SVM 预测模型的关键参数。具体算法步骤描述如下:

- 步骤 1:对训练样本进行归一化并预测数据。
- 步骤 2:初始化支持向量机 SVM 模型。
- 步骤 3:粒子群算法参数设置。
- 步骤 4:计算每个粒子的适应值
- 粒子群体。
- 步骤 5:根据得到的适应值,更新粒子群中每个粒子的状态,包括粒子的个体最优位置和全局最佳位置。
- 步骤 6:迭代次数通常在迭代次数达到所需的最大迭代次数或粒子适应度满足时终止。否则,继续进行评估和更新,直到达到最大迭代次数或满足粒子适应度。算法流程如图 2 所示。

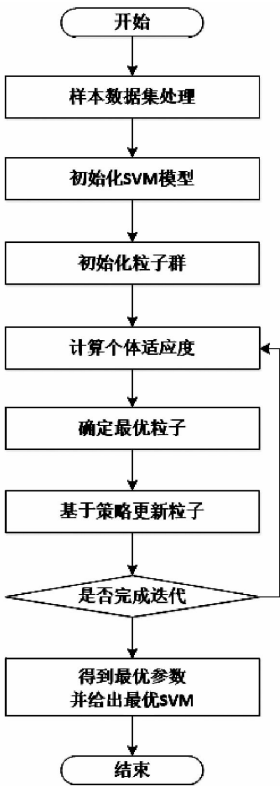


图 2 基于粒子群的 SVM 分类流程图

Fig. 2 SVM classification flow chart based on particle swarm

4 数值模拟

本文以某城区部分居民的低电压数据为例进行

数值模拟实验。实验数据来自智能电表的实时采集。为便于操作和查看,本文截取了 2017 年 6 月 30 日的数据作为低压分析的源数据。通过选择电压数据,如用户端和配电端的电压和电流。然后获得当日低电压 1929 条数据。动态实时电力的采样间隔为 15 分钟。可以每天 24 小时收集 96 个值。实时动态包括用户侧电压、电流、配电侧电压、电流以及功率因数等。使用 PL / SQL 工具将所采集的数据结果导出为 CSV 格式。

首先对基础数据进行预处理。然后对预处理后的低电压数据集生成 Canopy 中心,并将距离测量值设置为欧氏距离测量值。由于数据之间的差异,阈值 T1 和 T2 相对较小,经过多次实验,确定 T1 = 15,T2 = 20。使用聚类输出工具所得到的 Canopy 算法结果为:

```
Key :class :class org. apache. hadoop. io. Text ValueClass :
class
org. apache. mahout. clustering. iterator. ClusterWritable
Key :C - 0 : Value :
org. apache. mahout. clustering. iterator. ClusterWritable
@ 365
2831e
Key :C - 1 : Value :
org. apache. mahout. clustering. iterator. ClusterWritable
@ 365
2831e
Key :C - 2 : Value :
org. apache. mahout. clustering. iterator. ClusterWritable
@ 365
2831e
Key :C - 3 : Value :
org. apache. mahout. clustering. iterator. ClusterWritable
@ 365
2831e
Key :C - 4 : Value :
org. apache. mahout. clustering. iterator. ClusterWritable
@ 365
2831e
Key :C - 5 : Value :
org. apache. mahout. clustering. iterator. ClusterWritable
@ 365
2831e
```

```
Key:C - 6: Value:
org.apache.mahout.clustering.iterator.ClusterWritable
@ 365
2831e
Key:C - 7: Value:
org.apache.mahout.clustering.iterator.ClusterWritable
@ 365
2831e
Key:C - 8: Value:
org.apache.mahout.clustering.iterator.ClusterWritable
@ 365
2831e
Key:C - 9: Value:
org.apache.mahout.clustering.iterator.ClusterWritable
@ 365
2831e
Key:C - 10: Value:
org.apache.mahout.clustering.iterator.ClusterWritable
@ 365
2831e
Key:C - 11: Value:
org.apache.mahout.clustering.iterator.ClusterWritable
@ 365
2831e
Key:C - 12: Value:
org.apache.mahout.clustering.iterator.ClusterWritable
@ 365
2831e
Key:C - 13: Value:
org.apache.mahout.clustering.iterator.ClusterWritable
@ 365
2831e
Count:14.
```

从以上结果可以看出,Canopy 算法输出 14 个聚类,即总共 14 个低电压原因。然后使用 Canopy 算法结果作为 Kmeans 输入参数,这样就可以得到低电压数据的聚类结果。执行 Kmeans 聚类算法 8 次,得到聚类 0 到聚类 7 最终的聚类结果。结合最后的聚类结果 cluster-7-final 和聚类点,得到最终结果。得到的其中一个低电压原因的结果如下:
VL - 11 { n = 2 c = [0 : 7. 466 , 1 : 7. 466 , 2 : 7. 466 , ...] r = [188 : 2. 837 , 283 : 1. 801 , 303 : 0. 847 , 316 : 1. 173 ,

```
355 : 2. 067 , ... ] }
Weight : [ props - optional ] : Point :
1. 0 : [ distance = 5. 128051498738058 ] : / 1531. txt =
[ 0 : 7. 466 , 1 : 7. 466 , 2 : 7. 466 , 3 : 7. 466 , ... ]
1. 0 : [ distance = 5. 128051498738236 ] : / 1533. txt =
[ 0 : 7. 466 , 1 : 7. 466 , 2 : 7. 466 , 3 : 7. 466 , ... ]
```

VL - 11 代表一个低电压数据的集合,集合中有两个数据。c 是聚类的聚类中心点,r 是聚类的半径,Weight-Points 代表属于聚类的 2 个数据以及它们与聚类中心的距离。

根据 svm 库的分类函数对聚类后的结果数据进行所需格式的处理,然后将基于粒子群算法的支持向量机分类算法对未知原因的低压数据进行分类。将预测分类结果与实际分类结果进行对比。SVM 测试数据集由 1 929 条电力数据组成,其中有 1 529 条为训练数据,400 条为测试数据。将 PSO-SVM 的预测值与实际值进行对比,对比结果如图 3 所示。

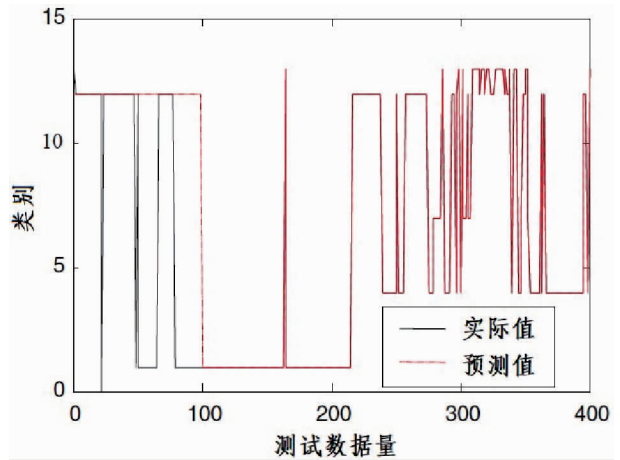


图 3 PSO-SVM 的低电压预测结果

Fig. 3 Low voltage prediction results of PSO-SVM

由图 3 可知,采用粒子群优化的 SVM 预测准确率高达 90%。实验结果表明基于粒子群的支持向量机具有较好的精度和精度。

5 结论

本文从电网低压现象的实际出发,构建了比较完整的低压诊断模型,该模型的实现过程主要包括聚类和分类两部分。在聚类过程中,采用 canopy 算法确定聚类数量和初始聚类数量,然后用 K 均值聚类算法进行迭代运算,收敛得到最终的聚类结果。在分类过程中,采用 PSO-SVM 用于确定低电压的原因。案例测试结果表明,该系统具有高可用性和高

精度,能够满足电力公司低压诊断的需要。

参考文献:

- [1] 李兰芳,李鑫,黄嘉健,等. 基于前推回代法的低电压台区成因数据分析[J]. 机电信息,2018(12):1-3.
LI Lanfang, LI Xin, HUANG Jiajian, et al. Analysis of genesis data of low voltage platform based on forward pushback method [J]. Mechanical and Electrical Information, 2018(12):1-3.
- [2] 刘永亮. 农村低电压治理典型办法与成效分析[J]. 内燃机与配件,2018(06):124-125.
LIU Yongliang. Analysis of typical methods and effectiveness of rural low voltage control [J]. Internal Combustion Engine & Parts, 2018(06):124-125.
- [3] 李沛彧,柳杰. 山西省农村电网“低电压”综合治理分析与研究[J]. 电工电气,2018(01):69-70+73.
LI Pei, LIU Jie. Shanxi Province, rural power grid "low voltage" comprehensive treatment analysis and research [J]. Electrotechnics Electric, 2018(01):69-70+73.
- [4] 任子晖,王琦. 基于优化 DDAGSVM 多类分类策略的电能质量扰动识别[J]. 电力系统保护与控制,2018,46(05):82-88.
REN Zihui, WANG Qi. Power quality disturbance identification based on optimized ddagsvm multiclass classification strategy [J]. Power System Protection and Control, 2018, 46(05):82-88.
- [5] 卢嗣斌,刘汉婕. 分布式电源接入对解决农网低电压问题的研究[J]. 电力大数据,2018,21(08):43-47.
Lu Yibin, Liu Hanqi. Research on Distributed Power Supply Access to Solve the Problem of Low Voltage in Rural Power Network [J]. Power System and Big Data, 2018, 21(08):43-47.
- [6] 李娜源,袁梅,王瑶等. SVM 与 PSO 相结合的电机轴承故障诊断[J]. 重庆大学学报,2018,41(01):99-107.
LI Yuyuan, YUAN Mei, WANG Yao, et al. Fault diagnosis of motor bearing based on SVM and PSO [J]. Journal of Chongqing University, 2018, 41(01):99-107.
- [7] 王凌云,李开成,肖厦颖,等. 基于稀疏分解的复合电能质量扰动分类[J]. 电测与仪表,2018,55(01):14-20+33.
WANG Lingyun, LI Kaicheng, XIAO Xaling, et al. Classification of complex power quality disturbances based on sparse decomposition [J]. Electric Measurement & Instrumentation, 2018, 55(01):14-20+33.
- [8] 林培杰,陈志聪,吴丽君,等. 一种 PSO-SVM 的光伏阵列故障检测与分类[J]. 福州大学学报(自然科学版),2017,45(05):652-658.
LIN Peijie, CHEN Zhicong, WU Lijun, et al. A PSO-SVM photovoltaic array fault detection and classification [J]. Journal of Fuzhou University (Natural Science Edition), 2017, 45(05):652-658.
- [9] 林顺富,谢潮,汤波,等. 数据挖掘在电能质量监测数据分析中的应用[J]. 电测与仪表,2017,54(09):46-51.
LIN Shunfu, XIE Chao, TANG Bo, et al. THE DATA MINING APPLICATION IN THE POWER QUALITY MONITORING DATA ANALYSIS [J]. Electrical Measurement & Instrumentation, 2017, 54(09):46-51.
- [10] 许元斌,李国辉,郭昆,等. 基于改进的并行 K-Means 算法的电力负荷聚类研究[J]. 计算机工程与应用,2017,53(17):260-265.
XU Yuanbin, LI Guohui, GUO Kun, et al. Research on parallel clustering of power load based on improved K-Means algorithm [J]. Computer Engineering and Applications, 2017, 53(17):260-265.
- [11] 张华赢,朱正国,姚森敬,等. 基于大数据分析的暂态电能质量综合评估方法[J]. 南方电网技术,2015,9(06):80-86.
ZHANG Huaying, ZHU Zhengguo, YAO Senjing, et al. Comprehensive assessment method of transient power quality based on big data analysis [J]. China Southern Power Grid, 2015, 9(06):80-86.
- [12] 韩俊,谈健,黄河,等. 基于改进 K-means 聚类算法的供电块划分方法[J]. 电力自动化设备,2015,35(06):123-129.
HAN Jun, TAN Jian, HUANG He, et al. Power-supplying block partition based on improved K-means clustering algorithm [J]. Electric Power Automation Equipment, 2015, 35(06):123-129.
- [13] 马聪,高峰,田昊,等. 适用于并网系统低电压穿越的电压检测算法[J]. 电力系统自动化,2015,39(05):122-127.
MA Cong, GAO Feng, TIAN Hao, et al. A Voltage Detection Method for Low Voltage Ride Through of Grid-connected System [J]. Automation of Electric Power Systems, 2015, 39(05):122-127.
- [14] 刘爱国,薛云涛,胡江鹭,等. 基于 GA 优化 SVM 的风电功率的超短期预测[J]. 电力系统保护与控制,2015,43(02):90-95.
LIU Aiguo, XUE Yuntao, HU Jianglu, et al. Ultra-short-term wind power forecasting based on SVM optimized by GA [J]. Power System Protection and Control, 2015, 43(02):90-95.
- [15] 余长俊,张燃. 云环境下基于 Canopy 聚类的 FCM 算法研究[J]. 计算机科学,2014,41(S2):316-319.
YU Changjun, ZHANG Ran. Research on FCM algorithm based on canopy clustering in cloud environment [J]. Computer Science, 2014, 41(S2):316-319.

收稿日期:2019-02-23

作者简介:



李占英(1979),男,本科,电力高级工程师,高级技师。主要从事电力运营监测,大数据诊断分析应用研究。

(本文责任编辑:龙海丽)

**A low voltage diagnosis model for distribution network based
on particle swarm optimization algorithm**

LI Zhanying¹, MA Fulan², MA Weibing³

- (1. State Grid Qinghai Electric Power company huanghua Electric Power company, Jianzha 811200 Qinghai, China;
- 2. State Grid Qinghai Electric Power company haibei Electric Power company, Haiyan 812200 Qinghai, China
- 3. State Grid Qinghai Electric Power company hainan Electric Power company, Gonghe 813000 Qinghai, China)

Abstract: With the development of smart grid construction, traditional low-voltage cause diagnosis based on detection technology has become a data big data classification technology based on data mining, and data classification research focusing on the cause of low voltage failure is still in its infancy in China. For this reason, a low-voltage diagnostic model of distribution network using improved clustering algorithm and support vector machine classification algorithm is proposed. The model first uses Canopy-Kmeans clustering algorithm to analyze the low voltage causes based on the historical operation data of the distribution network and obtain the possible low voltage causes. Then the particle swarm optimization algorithm is used to support the SVM data classification algorithm. Parameter optimization, finally using the support vector machine algorithm of the result parameter optimization to classify the real-time operation data of the distribution network collected by the smart meter for low-voltage causes and finally output the diagnosis result of the low-voltage fault. Experiments show that sampling based on particle swarm optimization based support vector machine diagnostic model can achieve 90% low-voltage cause diagnostic accuracy.

Key words: low voltage; Canopy-Kmeans; particle swarm optimization(PSO); support vector machine(SVM)