[人工智能]

# 基于人工智能和数据挖掘的电力系统故障分析预测研究

文/尹国梁[1] 张新浩[2]
1. 山东新华设计工程有限公司青岛分公司；2. 青岛青咨工程咨询有限公司

*在电力网络持续扩展和架构日趋复杂的背景下，配电系统故障已成为影响供电质量的重要因素。构建一种创新的三重架构数据分析模型，用于实现配电网络异常状态的智能识别与预警。综合运用K-means聚类、Apriori关联规则和随机梯度下降（SGD）算法，对配电网故障数据进行深度挖掘和分析。实验证明，模型在故障分类预测方面展现出较高的准确率和良好的性能，为配电网络的可靠供电与平稳运行奠定坚实基础，有效增强系统的安全防护能力。*

电能作为一种不可替代的能源形式，其持续稳定的供给对社会经济发展起着决定性作用。然而，由于配电网覆盖范围广、设备众多且运行环境复杂，故障的发生难以完全避免。当配电网发生故障时，快速准确地进行故障分类预测，对于及时采取有效的修复措施、减少停电时间和经济损失具有重要意义。

常规的诊断技术多为人工判断结合基础电气参数分析的方式，依托运维人员的工作经验，配合电压、电流等基本电学量的测量数据进行故障判断。这种方式在面对复杂故障时效率低，且准确性差。在数字化的今天，信息处理技术的快速进步为电力系统故障识别带来新的技术手段。智能分析算法和机器学习方法在电网故障检测中的应用日益深入。其中，信息提取技术能够对庞大的异常记录进行深度分析，从而为系统状态研判提供可靠依据。构建一种三层次信息分析框架，用于电网异常状态的智能识别与预判，通过整合多种智能算法，能够提升系统状态评估的准确性和可靠性。

## 人工智能算法与大数据挖掘技术

### 1. 人工智能算法

（1）K-means聚类算法原理与应用

作为一种典型的划分式聚类方法，K-means算法采用迭代优化策略实现数据分组。首先从样本空间中随机确定$N$个质心位置，通过计算样本点与各质心间的欧氏空间度量，将每个样本归类至最邻近的簇中。这一过程通过不断更新质心位置和重新分配样本，最终达到簇内紧密、簇间分离的聚类效果，计算公式为

$$d(x,y)=\sqrt{\sum_{i=1}^{n}(x_i-y_i^j)^2}$$

（2）Apriori关联规则算法原理与应用

Apriori算法是一种基于迭代搜索的关联关系发现方法。首先识别满足最小支持度阈值的单一特征集合，随后通过特征组合的扩展，逐步构建更高维度的频繁特征集，直至无法产生新的有效组合为止。

（3）随机梯度下降（SGD）算法原理与应用

随机梯度下降（SGD）算法属于一种逐步逼近

的最优解搜索技术，它的计算量较小，收敛速度较快。在配电网故障分类预测模型中，SGD算法用于对经过聚类和关联规则处理后的样本进行回归训练，以建立故障分类预测模型。

2. 大数据挖掘技术

（2）数据挖掘流程概述

数据挖掘基本流程包括确定挖掘目标、建立数据库、数据预处理、建立模型和分析以及模型评估等步骤。明确挖掘目标是对故障进行准确的分类预测的前提；然后建立相应的数据库，存储和管理故障数据；接着对数据进行预处理，提高数据质量；再选择合适的算法建立故障分类预测模型，并对模型进行分析和优化，流程如图1所示。

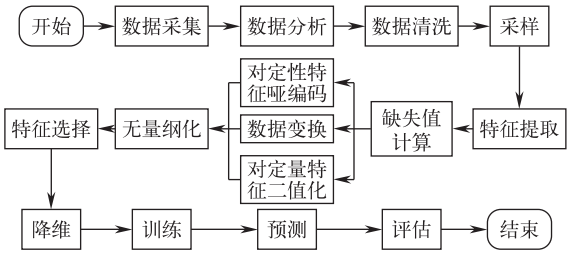


图1　数据挖掘流程图

（2）数据预处理方法

数据预处理主要目标是提升信息的准确性和实用性。常用的方法涵盖信息净化、信息融合与转换以及信息精简等方面。信息净化是对数据集中的空缺值、异常值和干扰信息进行处理；信息融合则是将不同来源的信息进行整合；信息精简则是在保持数据核心特征的前提下，对数据集进行压缩，降低数据规模，提升分析效能。

（3）模型评估指标

在评估故障分类预测模型的表现时，通常会运用一系列评价指标，如混淆矩阵、准确度、精准度、敏感度以及F1值等。混淆矩阵以表格形式呈现模型预测与实际标签的对应关系，基于此可进一步推算出准确度、精准度、敏感度和F1值等关键指标。分类模型的评估涉及多个关键指标：总体正确率体现模型对所有样本的正确判断能力；查准率则反映模型在预测为正类的样本中的准确程度；查全率衡量模型对实际正类样本的识别效果；F1值作为查准率和查全率的调和平均值，能够综合评估模型的精确识别能力和全面覆盖能力。

## 基于三层数据挖掘结构的配电网故障分类预测模型构建

1. 系统模型介绍

模拟参考三机九节点系统，搭建起电力系统的仿真模型。借助PSCAD软件，对配电网内常见的几种短路故障开展模拟操作。精确设定IEEE9节点的负荷数值、网络参数等数据，并在各个节点安装万用表，用于收集故障发生后各个节点所对应的电压值。

运用时间逻辑元件，将故障发生的时刻设定为0.25 s，同时把故障控制的类型确定为外部控制模式。在故障发生的指定位置（即设置为6节点处），经过反复多次的实验操作，成功采集到不同故障类型下的5 000余组数据。

2. 三层数据挖掘方法

（1）第一层。K-means聚类与自编码预处理

在开展第一层的数据挖掘工作时，首要步骤是对源样本库内的数据进行细致筛查，剔除异常值，填充缺失项。完成之后，针对源样本展开聚类方面的预处理工作。在聚类过程中，通过随机选择初始聚类中心，计算样本到各聚类中心的欧几里得距离，将样本就近分配，并根据组内样本均值更新聚类中心，直到准则函数收敛。

聚类完成后，对数据进行自编码。通过K-means聚类和自编码，将源样本库转换为样本库I，使同一组内的数据相关性增大，数据形式更加简洁，便于后续处理，详见表1。

表1　样本库I部分参数样式

| 样本组序号 | 节点1 | 节点2 | 节点3 | 节点4 | 节点5 | 节点7 | 故障类型 |
|---|---|---|---|---|---|---|---|
| 1 | 105 | 201 | 304 | 409 | 505 | 702 | 3 |
| 2 | 102 | 201 | 306 | 402 | 506 | 701 | 1 |
| 3 | 103 | 204 | 305 | 405 | 501 | 702 | 2 |
| 4 | 104 | 202 | 306 | 406 | 503 | 707 | 2 |
| 5 | 102 | 203 | 302 | 402 | 507 | 704 | 2 |
| 6 | 101 | 207 | 306 | 401 | 507 | 701 | 2 |

（2）第二层。Apriori关联规则筛选样本

在第二层数据挖掘中，使用Apriori关联规则算法对样本库I进行处理。首先初始化两个关键参数：将最小支持度阈值设定为0.3，将最小置信度标准确定为0.7。随后，算法将对样本库I中的各个数据集合进行支持度指标的量化分析。通过计算，系统能

够识别并筛选出支持度值超过预设阈值0.3的数据集合，这些集合将被归类为高频项目集合。然后从频繁项集中筛选出满足置信度大于0.7、提升度大于1条件的强关联规则样本组，详见表2。

表2　样本关联挖掘支持度、置信度和提升度

| 序号 | 关联样本 | 支持度 | 置信度 | 提升度 |
|---|---|---|---|---|
| 1 | {306}→{1} | 0.42 | 0.74 | 1.87 |
| 2 | {205,407}→{1} | 0.33 | 0.85 | 1.75 |
| 3 | {205,701}→{2} | 0.55 | 0.88 | 1.84 |
| 4 | {403,505}→{2} | 0.45 | 0.74 | 1.87 |
| 5 | {305,406}→{3} | 0.57 | 0.76 | 1.77 |
| 6 | {105,406，502}→{3} | 0.55 | 0.88 | 1.89 |
| 7 | {205,506}→{3} | 0.45 | 0.71 | 1.73 |
| 8 | {101,206}→{3} | 0.53 | 0.87 | 1.87 |

按照故障类型分类，形成SGD回归分类预测的样本库II，详见表3。

表3　样本库II部分参数样式

| 序号 | 节点1 | 节点2 | 节点3 | 节点4 | 节点5 | 节点7 | 故障类型 |
|---|---|---|---|---|---|---|---|
| 1 | 103 | 202 | 304 | 405 | 506 | 702 | 1 |
| 2 | 102 | 201 | 305 | 406 | 504 | 705 | 1 |
| 3 | 103 | 202 | 303 | 405 | 506 | 702 | 1 |
| 4 | 104 | 204 | 305 | 404 | 501 | 704 | 1 |
| 5 | 105 | 206 | 304 | 406 | 504 | 704 | 1 |
| 6 | 102 | 204 | 306 | 407 | 501 | 704 | 1 |

（3）第三层。SGD回归训练与模型优化

在第三层的数据挖掘环节，运用交叉验证法搜寻不同故障预测模型的最优参数。以常见的10折交叉验证进行操作，把样本库II里的样本均等地划分成10部分。在每次验证过程中，让一份样本充当交叉验证集，而其余的9份样本作为训练集，如此循环往复地进行10次训练操作，找出在交叉验证中得分最高的那组参数，确定为最优参数。

为了能够使分类预测模型得到更进一步的优化，在执行SGD回归任务时，再次借助交叉验证的方法，对SGD相关的各项参数进行优化调整。

通过对样本库I和样本库II的完整训练数据进行系统性实验，分别在每组参数配置下构建模型，并基于SGD评估其性能得分，可以发现经过聚类和关联规则处理后的样本库II在预测性能上有显著提升，详见表4。

表4　样本库I和样本库II的交叉验证得分

| 序号 | 可供选择的参数组 | 样本库I交叉验证得分 | 样本库II交叉验证得分 |
|---|---|---|---|
| 1 | hinge,L1, 0.1,500 | 0.46 | 0.57 |
| 2 | hinge,L1, 0.05,500 | 0.44 | 0.49 |
| 3 | hinge,L2, 0.1,500 | 0.43 | 0.47 |
| 4 | hinge,L2, 0.05,500 | 0.36 | 0.43 |
| 5 | log,L1, 0.1,500 | 0.44 | 0.65 |
| 6 | log,L1, 0.05,500 | 0.22 | 0.41 |
| 7 | log,L2, 0.1,500 | 0.54 | 0.79 |
| 8 | log,L2, 0.05,500 | 0.36 | 0.45 |

在SGD算法框架下，针对样本库I和样本库II建模过程中，逻辑回归模型被选为最优损失函数。

## 实验结果与分析

### 1. 实验设置与数据处理

将源样本、样本库I和样本库II按9∶1的比例划分训练集和测试集。对不同故障类型的样本进行处理，利用K-means聚类、Apriori关联规则和SGD算法构建故障分类预测模型。在K-means聚类中，通过轮廓系数法确定各节点样本的最佳$K$值；在Apriori关联规则挖掘中，设置合适的支持度、置信度和提升度阈值筛选强关联样本；在SGD回归训练中，通过交叉验证寻优得到最优参数。

### 2. 模型性能评估指标计算

采用原始样本的测试集对各模型性能进行评估，通过构建混淆矩阵来量化模型的整体分类准确率以及各类故障的识别准确度。数据显示，所构建模型的综合分类准确率达到93.6%。单相接地故障的识别率为95.1%；两相相间故障的识别率为91.5%；两相接地故障的识别率同样达到95.1%；而三相故障的识别率为92.8%。

## 结束语

构建三层数据挖掘结构的配电网故障分类预测方法，有机融合K-means聚类算法、Apriori关联规则以及SGD算法，提升了配电网故障分类预测的精准度与效率。模型在配电网故障分类预测领域展现出较高的准确率，同时具备良好的综合性能，能够为配电网的安全、稳定运行提供坚实且有力的支持，有助于保障电力系统的可靠供电。EA

（收稿日期：2025-04-14）