

DOI: 10.7667/PSPC170304

# 基于数据挖掘的输配网一体化事故响应决策

赵晋泉<sup>1</sup>, 杨婷<sup>1</sup>, 吴凡<sup>1</sup>, 韩佳兵<sup>2</sup>, 林昌年<sup>2</sup>, 魏文辉<sup>2</sup>

(1. 河海大学能源与电气学院, 江苏 南京 210098; 2. 北京科东电力控制系统有限责任公司, 北京 110179)

**摘要:** 针对含大量分布式电源电网的事故处理, 提出了一种基于数据挖掘的输配网一体化事故响应决策方法。首先, 利用蒙特卡洛仿真生成海量样本, 建立运行情景集。继而通过输配网一体化故障最优恢复得到不同运行情景发生预想故障的控制策略, 建立控制策略知识库。最后, 采用基于随机森林算法的数据挖掘技术建立情景决策规则库。当实际电网发生故障时, 进行快速决策匹配得到符合电网运行状态的控制策略。利用 R 软件对 IEEE30 和 IEEE33 节点构成的电网进行仿真, 结果验证了所提方法的有效性。

**关键词:** 事故响应决策; 数据挖掘; 输配网一体化故障最优恢复; 随机森林; 情景决策规则库

## Data mining based accident response and decision-making for integrated transmission and distribution grid

ZHAO Jinquan<sup>1</sup>, YANG Ting<sup>1</sup>, WU Fan<sup>1</sup>, HAN Jiabing<sup>2</sup>, LIN Changnian<sup>2</sup>, WEI Wenhui<sup>2</sup>

(1. College of Energy and Electrical Engineering, Hohai University, Nanjing 210098, China;

2. Beijing Kedong Electrical Control System Co., Ltd., Beijing 110179, China)

**Abstract:** In order to deal with the accidents in power grid containing a large number of distributed generations, a data mining-based accident response and decision-making method for integrated transmission and distribution grid is proposed. Firstly, massive samples are generated by Monte Carlo simulation and an operation scenario set is established. Secondly, the control strategies with respect to anticipated accidents of different operation scenarios are obtained by optimal service restoration of integrated transmission and distribution grid, and the control strategy knowledge database is established. Finally, the scenario decision-making rule database is established by using the random forest algorithm. When an accident occurs in a real power grid, the control decision corresponded to the practical state of the power grid is obtained by quickly scenario-decision matching. Simulation results on a power grid combining IEEE30 & IEEE33 bus test system by using R software show the effectiveness of the proposed method.

This work is supported by National Natural Science Foundation of China (No. 51577049) and Science and Technology Project of State Grid Corporation of China “Research on the key technologies of integrated analysis and fast simulation of main and distribution network with DGs”.

**Key words:** accident response and decision-making; data mining; optimal service restoration of integrated transmission and distribution grid; random forest; scenario decision-making rule database

## 0 引言

电网发生故障后进行快速准确的事故响应决策是电网安全运行的重要内容。随着大量分布式电源(DG)接入配电网, 现有事故响应决策系统<sup>[1-2]</sup>正面

临以下挑战:

1) 输电网和配电网事故后的控制策略是孤立制定的, 未能计及彼此间的相互支援和相互影响。随着配网 DG 渗透率的不断提高, 这种做法的经济性和有效性均无法保障, 进行输配网相互协调的事故响应决策非常必要。

2) 目前电网故障恢复决策主要基于专家经验和离线决策, 辅以在线分析。一方面, 规则粗放, 经济性差。大量不可控、间歇性 DG 的接入使电网

**基金项目:** 国家自然科学基金项目(51577049); 国家电网公司科技项目“面向分布式电源接入的主配网一体化分析及快速仿真关键技术研究”

的运行工况更加复杂,粗放型运行规则已不再适用;另一方面,随着电网规模的不断扩大、输配网一体化分析计算时间长,决策快速性无法保障。

由于数据挖掘技术能够在大规模数据中自动发现潜在规则,已在电力系统多个领域得到应用,包括电网稳定性分析<sup>[3-6]</sup>、配电网负荷预测<sup>[7]</sup>、电力系统故障诊断<sup>[8]</sup>和电力系统异常监测<sup>[9]</sup>等。其中,清华大学孙宏斌等人提出的智能机器调度员架构和电力系统精细化规则的建立机制<sup>[3-6]</sup>,表明数据挖掘技术应用于电网辅助决策具有良好前景。因此,本文将数据挖掘技术应用到输配网一体化事故响应决策中。

为适应大量 DG 接入电网后事故处理的快速性和有效性,本文提出基于随机森林算法的事故响应决策方法。首先,利用蒙特卡洛仿真生成海量样本,建立运行情景集;继而通过输配网一体化故障最优恢复得到不同故障情景下的最优控制策略,建立控制策略知识库(以下简称知识库);最后,采用随机森林算法建立情景决策规则库(以下简称规则库)。当实际电网发生故障时,进行快速决策匹配得到符合电网运行状态的控制策略。利用 R 软件对 IEEE30 和 IEEE33 节点构成的电网进行仿真,结果验证了本文所提方法的有效性。

## 1 知识库

### 1.1 运行情景集

电网的运行情景是指电网可能的运行状态,包含电网运行的所有信息,如发电机出力、DG 出力、负荷和电网拓扑等<sup>[10]</sup>。由于风电和光伏 DG 出力与气象因素息息相关,在构建运行情景集前要先建立考虑气象因素的 DG 出力模型。目前,构建运行情景集的方法主要有蒙特卡洛模拟法<sup>[11]</sup>和连续潮流法<sup>[4]</sup>,本文采用蒙特卡洛模拟法构建运行情景集。运行情景集的构建过程如下:

- 1) 分析电网的典型运行方式和典型拓扑结构。
- 2) 在线获取实际电网的运行状态;获得负荷预测和气象数据。
- 3) 生成运行情景集中海量样本。根据负荷预测结果将各节点负荷在负荷预测峰谷范围内按均匀分布抽样;根据气象预测数据得到各 DG 的出力曲线,利用蒙特卡洛仿真对各 DG 出力和各发电机出力在可调范围内抽样。利用输配网一体化潮流计算得到各运行场景下的运行状态。

为了实现输电网全局协调、各配电网分布自治的功能,运行情景集可分为输电网运行情景集和各配电网运行情景集。其中输电网运行情景集为各发

电机出力、各 PCC 点(输电网和各配电网的连接节点)不同功率水平组合而成的多个潮流态;配电网运行情景集为根据气象预测数据得到的各 DG 出力、各节点不同负荷水平组合而成的多个潮流态。

### 1.2 故障集

分析电网的典型故障和历史故障建立故障集,根据故障发生的位置分为输电网故障集和各配电网故障集。其中,输电网故障包括发电机故障、变电站故障、变压器故障、母线故障和输电线路故障等;配电网故障包括线路故障、DG 故障和母线故障。为简化起见,本文暂不考虑失稳故障。

### 1.3 输配网一体化故障最优恢复

输配网一体化故障最优恢复,即当故障发生在输电网时,充分考虑输、配电网侧可控资源进行校正控制;故障发生在配电网时,充分计及输配网间的潮流响应进行故障恢复。输电网故障时输配网一体化故障最优恢复的模型为

$$\min C_T + C_D \quad (1)$$

$$C_T = c_{T1} \sum_{i \in Y} \Delta P_{T,G,i} + c_{T2} \sum_{z \in Z} \Delta Q_{T,C,z} + \quad (2)$$

$$c_{T3} \sum_{w \in W} \Delta K_{T,g,w}$$

$$C_D = c_{D1} \sum_{t \in M} \Delta P_{DG,t} + c_{D2} \sum_{e \in DR} (1 - y_e) P_e \cdot \varphi_e + \quad (3)$$

$$c_{D3} \sum_{n \in N} (1 - y_n) P_n$$

$$\text{s.t. } \mathbf{h}_T(\mathbf{u}_T, \mathbf{x}_T, \mathbf{x}_B) = 0 \quad (4)$$

$$\mathbf{h}_D(\mathbf{u}_D, \mathbf{x}_B, \mathbf{x}_D) = 0 \quad (5)$$

$$\begin{cases} Q_{T,C,\min} \leq Q_{T,C} \leq Q_{T,C,\max} \\ k_{T,g,\min} \leq k_{T,g} \leq k_{T,g,\max} \end{cases} \quad (6)$$

$$\begin{cases} P_{T,G,\min} \leq P_{T,G} \leq P_{T,G,\max} \\ Q_{T,G,\min} \leq Q_{T,G} \leq Q_{T,G,\max} \\ P_{DG,\min} \leq P_{DG} \leq P_{DG,\max} \\ Q_{DG,\min} \leq Q_{DG} \leq Q_{DG,\max} \\ P_{DR} \leq P_{DR,\max} \end{cases} \quad (7)$$

式中:  $C_T$  和  $C_D$  分别为输电网和配电网的控制代价;  $c_{T1}$ 、 $c_{T2}$  和  $c_{T3}$  分别为输电网调节发电机出力、电容电抗器的投切组数和变压器分接头的单位控制代价;  $\Delta P_{T,G,i}$ 、 $\Delta Q_{T,C,z}$  和  $\Delta K_{T,g,w}$  分别为第  $i$  个发电机有功出力的调节量、第  $z$  个电容电抗器调节量和第  $w$  个变压器分接头调节量;  $Y$ 、 $Z$  和  $W$  分别为发电机、电容电抗器和变压器分接头集合;  $c_{D1}$ 、 $c_{D2}$  和  $c_{D3}$  分别为配电网调节 DG 出力、中断可中断负荷和切断刚性负荷(除可中断负荷外的其他负荷)的单位

控制代价;  $\Delta P_{DG,t}$ 、 $\varphi_e$ 、 $P_n$  和  $y_n$  分别为第  $t$  个可控 DG 的出力调节量、第  $e$  个可中断负荷的中断比例、第  $n$  个节点的负荷量和第  $n$  个节点的状态(1 表示正常供电、0 表示断开);  $M$ 、 $DR$  和  $N$  分别为可控 DG、可中断负荷和刚性负荷集合;  $\mathbf{u}_T$  和  $\mathbf{u}_D$  分别为输电网和配电网中的可控资源;  $\mathbf{x}_T$ 、 $\mathbf{x}_B$  和  $\mathbf{x}_D$  分别为输电网、PCC 点和配电网的状态变量向量。式(4)和式(5)为输电网和配电网的潮流方程; 式(6)为输电网要满足的约束, 包括电容电抗器的投切约束、变压器分接头的档位约束和发电机有功、无功出力约束; 式(7)为各配电网要满足的约束, 包括可控 DG 的有功、无功出力约束和可中断负荷的中断量约束。除以上约束外, 还应包括输电网和各配电网的节点电压约束和线路传输功率约束。

配电网故障时输配网一体化故障最优恢复的模型为

$$\min F = \lambda_1 \mu(P_{\text{out}}) + \lambda_2 \mu(K_{\text{act}}) + \lambda_3 \mu(P_{\text{DR}}) + \lambda_4 \mu(P_{\text{loss}}) + \lambda_5 \mu(P_{\text{add}}^{\text{DG}}) \quad (8)$$

$$\text{s.t.} \quad h_T(\mathbf{u}_T, \mathbf{x}_T, \mathbf{x}_B) = 0 \quad (9)$$

$$h_D(\mathbf{u}_D, \mathbf{x}_B, \mathbf{x}_D) = 0 \quad (10)$$

$$\begin{cases} P_{\text{DG},\min} \leq P_{\text{DG}} \leq P_{\text{DG},\max} \\ Q_{\text{DG},\min} \leq Q_{\text{DG}} \leq Q_{\text{DG},\max} \\ P_{\text{DR}} \leq P_{\text{DR},\max} \end{cases} \quad (11)$$

$$g \in G \quad (12)$$

式中:  $\lambda_1$ — $\lambda_5$  为各目标的权重;  $\mu$  为隶属度函数;  $P_{\text{out}}$ 、 $K_{\text{act}}$ 、 $P_{\text{DR}}$ 、 $P_{\text{loss}}$  和  $P_{\text{add}}^{\text{DG}}$  分别为刚性负荷丢失量、开关操作次数、可中断负荷中断量、网损量和可控 DG 的有功增加量;  $g$  和  $G$  分别为故障恢复后的网络拓扑结构和允许的辐射状拓扑结构集合。同时, 还应包括输电网和各配电网的节点电压约束和线路传输功率约束。

上述模型充分考虑了输配网间的相互影响, 因此, 所得控制策略较传统孤立计算更有效。

#### 1.4 知识库的建立过程

知识库的建立过程如下:

1) 建立运行情景集和故障集。

2) 将运行情景集和故障集组合模拟不同运行情景下发生不同预想故障, 利用输配网一体化故障最优恢复得到相应的控制策略, 并记录基本特征。包括中枢点信息、裕度特征、灵敏度特征和关键断面信息等。一个运行情景、一条故障信息及相对应的基本特征量和控制策略构成知识库中一条完整知识, 即知识库中一个样本, 其数据结构如表 1 所示。

3) 当所有的运行情景和预想故障组合计算完

毕, 一体化知识库建立完成。

4) 为保持输、配网调度管理模式, 需分别建立输网知识库和各配网知识库。输网知识库是从上述知识库中提取所有输电网的故障信息、运行情景信息、基本特征及对应的控制策略; 配网知识库是从上述知识库中提取相应配网的故障信息、运行情景信息、基本特征及对应的控制策略。

表 1 知识库样本数据结构

Table 1 Structure of the data in the knowledge database

故障信息	运行情景	基本特征	控制策略
故障网络	运行方式	关键断面	输电网控制策略
故障类型	拓扑结构	裕度	
故障位置	气象条件	灵敏度	配电网控制策略
电压等级	潮流态信息	中枢点等	

## 2 规则库

### 2.1 R 软件

R 软件是由 Auckland 大学的 Robert Gentleman 和 Ross Ihaka 等人开发, 以 R 语言为编程语言, 具有数据处理、统计分析和可视化输出功能的开源软件<sup>[12]</sup>。R 软件因其开源性、强大的统计分析和数据挖掘等功能得到广泛应用, 其优势在于提供了大量功能多样的工具包, 尤其是在数据挖掘领域存在大量扩展软件包。因此, 本文的数据挖掘方法将在 R 平台上进行。

### 2.2 数据预处理

为提高数据挖掘的速度和精度, 在数据挖掘之前要进行数据预处理<sup>[12-13]</sup>, 主要分为采样数据预处理、特征选择和专家干预三个部分。

#### 2.2.1 采样数据预处理

若不同故障和不同运行情景对应的控制策略样本数量分布不均衡, 将会造成知识库样本的不平衡性, 从而影响数据挖掘性能。本文采用过采样(SMOTE)技术进行样本类不平衡预处理<sup>[13]</sup>。SMOTE 方法从“相距较近的少类样本之间的样本还是少类样本”的假设出发, 为每个少类样本确定多个相距最近的少类样本, 然后在样本与其近邻样本的连线上合成“人造”样本。其方法为

$$FF_{k,j} = f_{k,j} + (ff_{k,j} - f_{k,j}) \times \text{rand}[0,1] \quad (13)$$

$$k = 1, 2, \dots, p \quad j = 1, 2, \dots, q$$

$$FF_k = \{FF_{k,1}, FF_{k,2}, \dots, FF_{k,q}\} \quad (14)$$

式中:  $f_{k,j}$  为第  $k$  个少类样本的  $j$  个属性值;  $ff_{k,j}$  为  $f_k$  近邻样本的  $j$  个属性值;  $p$  为少类样本总数;  $q$  为总属性个数;  $FF_{k,j}$  为合成的人造样本的  $j$  个属性值;  $FF_k$  为新的少类样本。

### 2.2.2 特征选择

特征选择是降维的关键方法，它有助于剔除冗余特征和噪音，提高数据挖掘模型的精度。目前，特征选择方法主要分为有监督<sup>[4]</sup>、无监督<sup>[14]</sup>和半监督<sup>[15]</sup>三种类型，本文采用基于学习模型(随机森林模型)的有监督特征选择方法。该方法通过向原始特征属性中加入随机干扰项生成阴影特征属性，根据平均精确度减少指标得到两者对学习模型的精度，获得所有特征的重要度并通过排序筛选出重要特征，从而建立关键特征子集。本文利用 R 软件进行特征选择的软件包为 Boruta 包，其特征选择过程如下：

1) 构造阴影特征。将知识库样本集中故障信息、运行情景和基本特征定义为原始特征，通过向所有原始特征属性中增加随机干扰项生成阴影特征属性，扩展原样本集得到扩展样本集。

2) 确定原始特征与阴影特征的重要度。利用学习模型训练扩展样本集，采用平均精确度减少指标评估每个特征的重要度，其数学表达式为

$$\gamma(ii)=1-\left(\varepsilon_0-RFA[DL(ii)]\right) \quad ii=1,2,\dots,2 \times nn \quad (15)$$

式中： $\gamma(ii)$  为第  $ii$  个特征的重要度； $\varepsilon_0$  为未经特征提取前的随机森林模型分类精度； $RFA[*]$  为随机森林模型分类精度的获得函数； $DL(ii)$  为打乱第  $ii$  个特征属性顺序的函数； $nn$  为原始特征个数。

3) 提取并排序原始特征中的关键特征。通过判断原始特征与阴影特征重要度的大小进行特征提取。如果原始特征的重要度显著高于阴影特征的重要度，则提取该原始特征。根据重要度对原始特征进行排序获得序列列表，建立关键特征子集。

### 2.2.3 专家干预

为将数据挖掘与电力系统先验知识有机结合，提高算法精度，本文在特征提取后加入了专家干预环节。首先，根据专家经验确定电网事故响应决策中的重要特征，如输电网故障时的灵敏度特征、配电网故障时的根节点电压特征；判断关键特征子集中是否含有这些重要特征，若无则将其加入到关键特征子集。

### 2.3 利用随机森林算法建立规则库

随机森林作为一种数据挖掘算法，其核心是结合单个分类器(决策树)组合成多分类器，各分类器相互独立并最终投票产生分类结果<sup>[16-17]</sup>。由于其降低了模型的复杂度，克服了单分类器精度低的不足，适合高维知识库样本集的数据挖掘。利用随机森林算法进行数据挖掘的过程如图 1 所示。

基于随机森林算法建立规则库的具体过程如下：

1) 将经过数据预处理后的知识库样本集划分

为训练样本集和测试样本集。

2) 利用 Bagging 方法对训练样本集进行抽样，形成多个训练样本子集。由于 Bagging 是有放回抽样，因此约有 37% 的样本不会出现在训练样本集中，这些数据称为袋外(Out of Bag, OOB)数据。

3) 随机抽取每个样本子集中的特征作为决策树的分裂属性，利用 CART 算法对每个样本子集建立 CART 决策树。

4) 不断重复 2)、3) 步骤生成大量决策树，形成随机森林模型。解析模型中的每个决策树，每一条从根节点到叶节点的路径表示一种规则。当模型中所有决策树解析完成即建立完整的规则库。

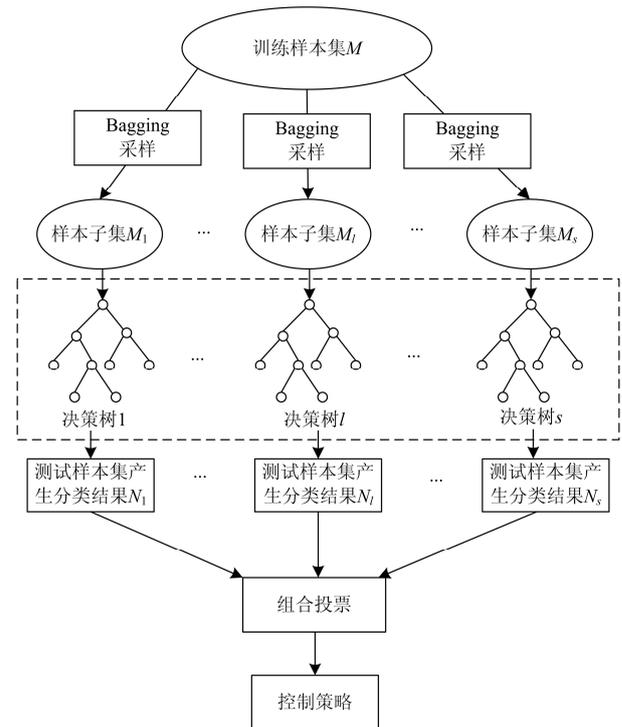


图 1 随机森林的建模过程

Fig. 1 Modeling process of a random forest

基于随机森林模型对测试样本集进行测试，得到各决策树的分类结果，汇总并采用投票方式确定最终分类结果(控制策略)<sup>[18]</sup>。

$$H(x) = \arg \max_B \sum_{l=1}^s I(hh_l(x) = B) \quad (16)$$

式中： $H(x)$  为控制策略预测模型； $hh_l$  是第  $l$  个决策树分类模型； $B$  为目标变量(即该决策树下对应的分类结果)； $I(*)$  为示性函数。

本文中利用 R 软件的程序包 Library(Random Forest)实现随机森林算法，该程序包是基于 Breiman 随机森林理论的 R 语言软件包 Random Forest

4.6-12。利用随机森林算法进行数据挖掘过程中需要设置两个重要参数：森林中决策树的数量(Ntree)和内部节点随机选择属性的个数(Mtry)。

### 3 事故响应决策系统

输配网一体化事故响应决策系统架构如图 2 所示。图中，箭头方向表示事故响应决策系统过程。首先，从实际电网的 EMS/DMS 系统中获得实时运行数据，结合气象和负荷预测数据构建运行情景集，同时构建故障集。其次，利用输配网一体化故障最优恢复得到不同情景和故障对应的控制策略，得到输网和各配网知识库。最后，对输网和各配网知识库进行数据挖掘建立相应的规则库。

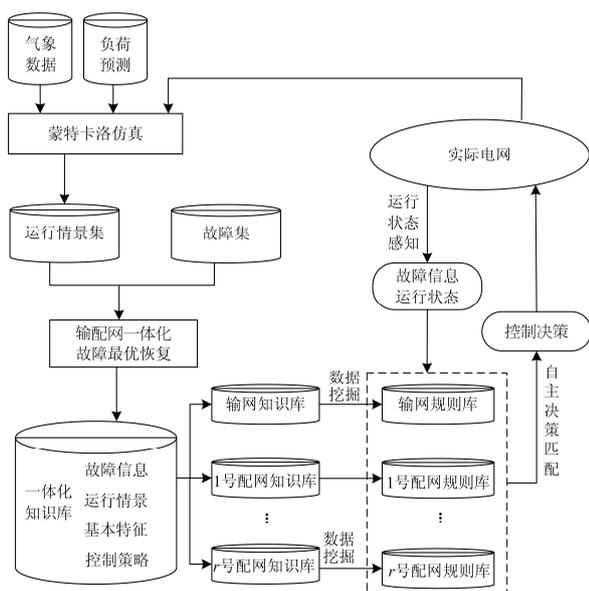


图 2 事故响应决策系统架构

Fig. 2 Framework of accident response and decision-making system

当实际电网发生故障时，通过运行态势感知技术获得电网的状态信息和故障前电网运行状态数据，利用自主策略匹配技术从规则库中获得与电网状态相匹配的控制策略，并作用于实际电网。

本文事故响应决策系统能获得适应于输、配电网的精细规则，只需要在线匹配，其他环节均可通过离线计算进行，缩短了控制时间。通过不断从电网中获取不同运行方式的状态数据，可更新和扩充知识库和规则库，保证控制策略的有效性。

## 4 算例分析

### 4.1 输配网一体化与孤立事故响应决策对比

#### 1) 配电网故障

采用 IEEE30 节点输电网和 IEEE33 节点配电网构造全局电网算例，模拟配电网故障。IEEE30 节点输电网中 19 号节点为 PCC 点，其他节点等值为负荷节点。IEEE33 节点配电网中各节点负荷增大为原来的 2 倍，其含有 5 台 DG，其类型和接入点如表 2 所示。可中断负荷节点为 15、21、24、30 和 31 号节点，可中断比例均为 60%。假设配电网中 3 号支路发生故障，将造成 4 号节点下连所有节点负荷失电和 DG 脱网。输配网一体化与配电网孤立事故响应决策对比结果如表 3 所示。

表 2 配电网中 DG 节点与参数

Table 2 Buses and parameters of DGs in distribution grid

节点	DG 类型	节点类型	额定参数
6	光伏	PV	$P = 800 \text{ kW}$ , $V = 1 \text{ p.u.}$ , $Q_{\max} = 500 \text{ kvar}$ , $Q_{\min} = -500 \text{ kvar}$
17	内燃机	PV	$P = 900 \text{ kW}$ , $V = 1 \text{ p.u.}$ , $Q_{\max} = 500 \text{ kvar}$ , $Q_{\min} = -500 \text{ kvar}$
20	风机	PQ	$P = 800 \text{ kW}$ , $Q = 400 \text{ kvar}$
23	内燃机	PQ	$P_{\max} = 1000 \text{ kW}$ , $Q_{\max} = 600 \text{ kvar}$
31	微型燃气轮机	PQ	$P_{\max} = 1100 \text{ kW}$ , $Q_{\max} = 600 \text{ kvar}$

表 3 配电网故障恢复结果对比

Table 3 Service restoration comparison of distribution grid

类别	输配网一体化故障最优恢复结果	配电网孤立故障恢复结果
支路操作	闭合支路 33	断开支路 26; 闭合支路 33、37
可中断负荷中断节点	30、31	30
失电负荷节点	无	无
开关操作数	1	3
网损/kW	100.34	119.49
目标函数值	0.066 78	0.104 31

由表 3 可知，当配电网故障时，输配网一体化故障最优恢复同配电网孤立故障恢复得到的控制策略相比，虽然可中断负荷量增加，但其开关操作数更少、网损更小，因此前者所得的控制策略更优。

#### 2) 输电网故障

采用 IEEE30 节点输电网和 3 馈线、16 节点配电网<sup>[19]</sup>构造全局电网算例，模拟输电网故障。输电网中 21、24 节点为 PCC 点，其他节点等值为负荷节点。输电网中变压器分接头在 0.95~1.05 间分 5 档，初始位置分别为 3、1、3 和 0；无功补偿电容器在 0~0.4 间分 4 组投切，初始组数分别为 2 和 0；配电网中接入 DG 的参数同表 2，接入位置依次更改为 6、7、10、11 和 16 号节点。可中断负荷节点为 5、8 和 9 号节点，可中断比例均为 60%。假设输电网中除 21 和 24 节点外其他各节点负荷增大为原来的 1.15 倍，28 号线路发生故障。输配一体化故障最优恢复与输电网孤立故障恢复得到的故障影响

及控制策略进行对比, 结果如表 4 所示。

由表 4 可知, 当输电网发生故障时, 输配一体化故障最优恢复与输电网孤立故障恢复得到的故障影响与控制策略均不同。由于前者计算过程中充分计及输、配网间的潮流响应, 故其故障影响更准确。前者所得的控制策略可通过中断配网中的可中断负荷消除故障引起的越限, 而不必切除配网中其他负荷, 因此其控制策略更优。

表 4 输电网故障恢复结果对比

Table 4 Service restoration comparison of transmission grid

类别	输配网一体化故障 最优恢复结果	输电网孤立故障最优 恢复结果
故障影响	电压越限节点: 无 潮流越限支路: 1, 27	电压越限节点: 24, 26, 30 潮流越限支路: 1, 27
控制策略	中断配网全部可中断负荷	切除 21 号节点负荷

#### 4.2 基于随机森林算法的数据挖掘

采用 IEEE30 节点输电网和 IEEE33 节点配电网构造全局电网算例进行数据挖掘。IEEE30 节点输电网中 10、16 和 26 号节点为 PCC 点, 其他节点等值为负荷节点。3 个 PCC 点所连配电网为相同的 IEEE33 节点配电网, 其 DG 类型、接入点如表 5 所示。各节点负荷在 0.7~1.3 倍原负荷范围内均匀分布。气象变化引起的光伏和风机有功出力在  $0 \sim P_w$  范围内随机变化; 其他类型 DG 有功出力在  $0.4P_w \sim P_w$  范围内变化; 无功出力随有功出力按额定比例变化。

表 5 IEEE33 节点系统 DG 节点与参数

Table 5 Buses and parameters of DGs in IEEE33 bus system

节点	DG 类型	额定参数
6	光伏	$P_w = 300 \text{ kW}, Q_w = 150 \text{ kvar}$
17	内燃机	$P_w = 500 \text{ kW}, Q_w = 300 \text{ kvar}$
20	光伏	$P_w = 300 \text{ kW}, Q_w = 150 \text{ kvar}$
23	异步风机	$P_w = 200 \text{ kW}, Q_w = 50 \text{ kvar}$
31	微型燃气轮机	$P_w = 400 \text{ kW}, Q_w = 200 \text{ kvar}$

利用 1.1 节中运行情景集生成方法得到 1000 个运行情景, 故障集为配电网 4 号支路故障和 17 号节点 DG 故障。配电网 4 号支路故障将造成 5 号节点下连所有节点负荷失电、DG 脱网, 利用输配网一体化故障最优恢复得到不同运行情景下的控制策略 5 种; 由于配电网 17 号节点处于树状结构末端, 17 号节点所连 DG 故障将造成部分运行情景下的配电网末端电压偏低, 利用输配网一体化故障最优恢复得到不同运行情景下的控制策略 6 种。故障后的控制策略如表 6 所示。3 号配电网知识库中包括 2000 个样本, 其中 1500 个样本作为训练样本集, 500 个样本作为测试样本集。

泛化能力是指数据挖掘算法对新样本的适应能力, 可利用 OOB 误差来估计随机森林算法的泛化误差。以每棵决策树为单位, 将 OOB 数据实际分类结果与决策树分类结果对比得到 OOB 误差率; 将所有决策树的 OOB 误差率取均值得到随机森林的 OOB 误差率<sup>[12]</sup>。由于 OOB 误差估计是无偏估计, 本文将其作为预测控制策略结果的评价指标。样本集预处理前后的 OOB 估计误差结果如表 7 所示。

表 6 不同运行情景下的控制策略

Table 6 Control strategies of different operation scenarios

故障类型	控制策略
4 号支路故障	闭合支路 37
	断开支路 25; 闭合支路 33、37
	断开支路 7; 闭合支路 33、37
	断开支路 27; 闭合支路 353、37
17 号节点 DG 故障	断开支路 9、27; 闭合支路 33、35、37
	不动作
	断开支路 9; 闭合支路 35
	断开支路 7、11; 闭合支路 33、35
17 号节点 DG 故障	断开支路 10、28; 闭合支路 35、37
	断开支路 14、28; 闭合支路 36、37
	断开支路 7、10、13; 闭合支路 33、34、35

表 7 样本集预处理前后的 OOB 估计误差率

Table 7 Error rate of OOB estimation before and after sample preprocessing

样本集	OOB 估计误差率/%
初始样本集	20.84
预处理后样本集	3.43

由表 7 所知, 本文建立的随机森林模型在经过数据预处理后的 OOB 误差率为 3.43%, 在允许范围内; 未经数据预处理建立的随机森林模型的 OOB 误差率为 20.84%, 验证了建立随机森林模型前进行数据预处理的必要性。

本文对随机森林算法中 Ntree 与 OOB 估计误差率之间的关系进行了试验, 结果如表 8 所示。由表 8 可知, Ntree 的大小直接影响预测模型的精度, 取值过小会使模型倾向决策树模型, 预测精度不高; 取值过大会使随机森林模型分类过细, 增加计算量, 当 Ntree 为 400 时会获得较高的分类精度。

表 8 不同大小森林的 OOB 估计误差率

Table 8 Error rate of OOB estimation of different forest sizes

树的大小 Ntree	OOB 估计误差率/%
100	6.84
200	3.57
300	3.46
400	3.43
500	3.46

### 4.3 不同事故响应决策方法对比

在本文研究基础上,对本文方法和实时计算获得控制策略的方法进行对比,结果如表9所示。

由表9可知,当电网发生故障时,本文离线建立规则库、在线匹配较实时计算的事故响应决策方法用时更短,因此本文方法能够实现快速决策。

表9 不同事故响应决策方法对比结果

Table 9 Accident response and decision-making schemes comparison

方法	获得策略时间/s
实时计算	11.1085
本文方法	0.0101

## 5 结论

本文提出了一种基于数据挖掘的输配网一体化事故响应决策方法。通过随机森林算法离线建立规则库、在线决策匹配实现了事故后的快速响应。经过仿真验证,得出如下结论:1)输配网一体化故障最优恢复较孤立计算得到的控制策略更有效;2)基于生成的知识库样本集,随机森林算法可以对控制策略进行分类,从而建立决策规则库;3)所提离线建立规则库、在线决策匹配较传统的事故响应决策方法更快、更有效。本文研究中仅计及了简单非失稳故障,但是所提方法同样适用于复杂非失稳故障的恢复。

### 参考文献

- [1] 靖宇宸,张钟毓.城市电网智能调度故障辅助决策系统设计开发[J].电力系统保护与控制,2016,44(8):132-136.  
JING Yuchen, ZHANG Zhongyu. Design and development of an auxiliary decision-making system for smart distribution dispatching fault in urban power network[J]. Power System Protection and Control, 2016, 44(8): 132-136.
- [2] 梁俊晖,廖志伟,黄志元,等.广东电网事故处理智能辅助决策专家系统开发[J].南方电网技术,2008,2(6):26-30.  
LIANG Junhui, LIAO Zhiwei, HUANG Zhiyuan, et al. The development of intelligent assistant-decision expert system for accident treatment in Guangdong power system[J]. Southern Power System Technology, 2008, 2(6): 26-30.
- [3] 孙宏斌,谢开,蒋维勇,等.智能机器调度员的原理和原型系统[J].电力系统自动化,2007,31(16):1-6.  
SUN Hongbin, XIE Kai, JIANG Weiyong, et al. Automatic operator for power systems principle and prototype[J]. Automation of Electric Power Systems, 2007, 31(16): 1-6.
- [4] 蒋维勇,孙宏斌,张伯明,等.电力系统精细规则的研究[J].中国电机工程学报,2009,29(4):1-7.  
JIANG Weiyong, SUN Hongbin, ZHANG Boming, et al. Fine operational rule of power system[J]. Proceedings of the CSEE, 2009, 29(4): 1-7.
- [5] 王康,孙宏斌,蒋维勇,等.智能控制中心二级精细化规则生成方法[J].电力系统自动化,2010,34(7):45-49.  
WANG Kang, SUN Hongbin, JIANG Weiyong, et al. A two-level refinement method for operational rules in smart control centers[J]. Automation of Electric Power Systems, 2010, 34(7): 45-49.
- [6] 孙宏斌,赵峰,蒋维勇,等.电网精细规则在线自动发现系统架构与功能设计[J].电力系统自动化,2011,35(18):81-86.  
SUN Hongbin, ZHAO Feng, JIANG Weiyong, et al. Framework and functions of fine operational rules online automatic discovery system for power grid[J]. Automation of Electric Power Systems, 2011, 35(18): 81-86.
- [7] 林顺富,郝朝,汤晓栋,等.基于数据挖掘的楼宇短期负荷预测方法研究[J].电力系统保护与控制,2016,44(7):83-89.  
LIN Shunfu, HAO Chao, TANG Xiaodong, et al. Study of short-term load forecasting method based on data mining for buildings[J]. Power System Protection and Control, 2016, 44(7): 83-89.
- [8] 宋杰,谢海宁,杨增辉,等.基于多源异构数据挖掘的配电网故障信息统计分析[J].电力系统保护与控制,2016,44(3):141-147.  
SONG Jie, XIE Haining, YANG Zenghui, et al. Statistical analysis of the distribution fault based on multiple source and isomerism data mining technology[J]. Power System Protection and Control, 2016, 44(3): 141-147.
- [9] 郭慧东,王玮,夏明超.基于数据挖掘的风电机组变桨系统劣化状态在线辨识方法[J].中国电机工程学报,2016,36(9):2389-2397.  
GUO Huidong, WANG Wei, XIA Mingchao. On-line identification for wind turbine pitch system degradation based on data mining technology[J]. Proceedings of the CSEE, 2016, 36(9): 2389-2397.
- [10] 刘洪,朱振环,马康.基于情景分析的配电网运行风险预警方法[J].电力系统及其自动化学报,2015,27(5):39-44.  
LIU Hong, ZHU Zhenhuan, MA Kang. Distribution network

- operation risk forewarning based on scenario analysis[J]. Proceedings of the CSU-EPSA, 2015, 27(5): 39-44.
- [11] VOUMVOULAKIS E M, HATZIARGYRIOU N D. A particle swarm optimization method for power system dynamic security control[J]. IEEE Transactions on Power Systems, 2010, 25(2): 1032-1041.
- [12] 曹正凤. 随机森林算法优化研究[D]. 北京: 首都经济贸易大学, 2014.  
CAO Zhengfeng. Study on optimization of random forests algorithm[D]. Beijing: Capital University of Economics and Business, 2014.
- [13] 伊华. 面向高维和不平衡数据分类的集成学习研究[D]. 武汉: 武汉大学, 2012.  
YI Hua. Research of ensemble learning for high-dimensional and imbalanced data classification[D]. Wuhan: Wuhan University, 2012.
- [14] 徐峻岭, 周毓明, 陈林, 等. 基于互信息的无监督特征选择[J]. 计算机研究与发展, 2012, 49(2): 372-382.  
XU Junling, ZHOU Yuming, CHEN Lin, et al. An unsupervised feature selection approach based on mutual information[J]. Journal of Computer Research and Development, 2012, 49(2): 372-382.
- [15] 段洁, 胡清华, 张灵均, 等. 基于邻域粗糙集的多标记分类特征选择算法[J]. 计算机研究与发展, 2015, 52(1): 56-65.  
DUAN Jie, HU Qinghua, ZHANG Lingjun, et al. Feature selection for multi-label classification based on neighborhood rough sets[J]. Journal of Computer Research and Development, 2015, 52(1): 56-65.
- [16] 赵腾, 王林童, 张焰, 等. 采用互信息与随机森林算法的用户用电关联因素辨识及用电量预测方法[J]. 中国电机工程学报, 2016, 36(3): 604-614.  
ZHAO Teng, WANG Lintong, ZHANG Yan, et al. Relation factor identification of electricity consumption behavior of users and electricity demand forecasting based on mutual information and random forests[J]. Proceedings of the CSEE, 2016, 36(3): 604-614.
- [17] 许勇刚, 张建业, 龚小刚, 等. 基于改进随机森林算法的电力业务实时流量分类方法[J]. 电力系统保护与控制, 2016, 44(24): 82-89.  
XU Yonggang, ZHANG Jianye, GONG Xiaogang, et al. A method of real-time traffic classification in secure access of the power enterprise based on improved random forest algorithm[J]. Power System Protection and Control, 2016, 44(24): 82-89.
- [18] 方匡南, 吴见彬, 朱建平, 等. 随机森林方法研究综述[J]. 统计与信息论坛, 2011, 26(3): 32-38.  
FANG Kuangnan, WU Jianbin, ZHU Jianping, et al. A review of technologies on random forests[J]. Statics & Information Forum, 2011, 26(3): 32-38.
- [19] CIVANLAR S, GRAINGER J J, YIN H, et al. Distribution feeder reconfiguration for loss reduction[J]. IEEE Transactions on Power Delivery, 1998, 3(3): 1217-1223.

收稿日期: 2017-03-05; 修回日期: 2017-04-05

作者简介:

赵晋泉(1972—), 男, 通信作者, 博士, 教授, 研究方向为电力系统优化运行、电压稳定分析与控制和电力市场等方面; E-mail: jqzhao2@tom.com

杨婷(1991—), 女, 硕士研究生, 研究方向为电力系统运行优化运行与控制; E-mail: yangtingfly0510@126.com

吴凡(1993—), 男, 硕士研究生, 研究方向为智能配电网自愈控制的理论与方法。

(编辑 葛艳娜)