

DOI: 10.19783/j.cnki.pspc.250534

基于改进决策树的安全约束经济调度的冗余约束识别方法

夏卓廷¹, 张大海¹, 严嘉豪², 李振宇³, 毛文博², 杨大坤¹

(1. 北京交通大学电气工程学院, 北京 100044; 2. 中国电力科学研究院有限公司(南京), 江苏 南京 210003;
3. 国网甘肃省电力公司, 甘肃 兰州 730046)

摘要: 针对安全约束经济调度(security-constrained economic dispatch, SCED)中冗余约束的识别目前尚缺乏快速有效的识别方法。同时, 数据驱动方法容易产生假正例(false positive, FP)误判, 从而影响系统安全性。为此, 提出了一种改进的决策树(decision tree, DT)算法, 即改进的分类回归树(classification and regression tree, CART)算法, 以及改进的错误率降低剪枝(reduced error pruning, REP)算法, 以实现冗余约束的快速识别。首先, 阐述 SCED 模型与 CART 原理。其次, 构建了冗余约束识别的特征工程及数据预处理方法。然后, 提出了融合 FP 惩罚机制的改进 CART 算法及基于 FP 比的 REP 剪枝策略。最后, 通过 SG-126 系统验证了所提改进算法在较好地适应极端 FP 敏感场景的同时, 能够快速、准确地识别冗余约束。冗余约束识别准确率达到 95.13%, FP 误判率为 0, 在削减冗余约束后系统调度时间减少了 88.22%。

关键词: 安全约束经济调度; 冗余约束识别; 分类回归树; 错误率降低剪枝

An identification method for redundant constraints in safety-constrained economic dispatch based on improved decision tree

XIA Zhuoyan¹, ZHANG Dahai¹, YAN Jiahao², LI Zhenyu³, MAO Wenbo², YANG Dakun¹

(1. Department of Electrical Engineering, Beijing Jiaotong University, Beijing 100044, China;
2. China Electric Power Research Institute (Nanjing), Nanjing 210003, China;
3. State Grid Gansu Electric Power Company, Lanzhou 730046, China)

Abstract: Currently, there is a lack of fast and effective methods for identifying redundant constraints in security-constrained economic dispatch (SCED). Moreover, data-driven approaches are prone to false positive (FP) misclassification, which may compromise system security. To address these issues, an improved decision tree (DT) algorithm, namely, an improved classification and regression tree (CART) algorithm combined with an enhanced reduced error pruning (REP) strategy, is proposed for rapid redundant constraints identification. First, the SCED model and the principles of CART are introduced. Second, feature engineering and data preprocessing methods for redundant constraint identification are constructed. Then, an improved CART algorithm incorporating a FP penalty mechanism and a REP strategy based on the FP ratio are proposed. Finally, case studies on the SG-126 system demonstrate that the proposed algorithm can quickly and accurately identify redundant constraints while effectively adapting to extreme FP-sensitive scenarios. The accuracy rate of redundant constraint identification reaches 95.13%, with a FP misclassification rate of zero, and system dispatch time is reduced by 88.22% after eliminating redundant constraints.

This work is supported by the National Key Research and Development Program of China (No. 2022YFB2403400).

Key words: security-constrained economic dispatch (SCED); redundant constraint identification; classification and regression tree (CART); reduced error pruning (REP)

0 引言

随着工业的高速发展, 全球气候也因此遭受不

良影响, 高效使用清洁可再生能源已成为世界各国的共同目标^[1-3]。随着我国“双碳”目标提出, 在我国电力系统中新能源发电占比逐年提高^[4-7], 电力系统的规模也逐年增大, 这都给安全约束经济调度(security-constrained economic dispatch, SCED)带来

基金项目: 国家重点研发计划项目资助(2022YFB2403400)

严峻挑战^[8-9]。

现阶段的 SCED 模型中,在部分场景下需要考虑机组启停等非线性因素,让 SCED 问题的求解变成混合整数规划(mixed-integer programming, MIP)问题,并且大量的约束条件让 SCED 模型的规模变得十分庞大,限制了其高效求解。文献[10-11]表明, MIP 问题中通常存在大量冗余约束,这些约束并不会改变优化问题的最优解和可行域,而只对问题的规模产生影响。因此,识别并削减冗余约束可以显著减小 SCED 规模,进而提升求解效率。

为有效识别调度模型中的冗余约束,国内外学者已开展了大量研究。例如,文献[12]提出了伞约束的概念,通过求解线性规划辅助优化问题得到伞约束并以此构建新的辅助优化问题,其余约束即可作为冗余约束予以删除。删除冗余约束能够缩减调度问题的规模,从而更易于对调度模型进行求解。文献[13]在此基础上进一步提出了一种加速部分伞约束识别的方法,并引入并行计算,显著提高了计算效率。文献[14-15]同样提出了一种基于辅助优化问题的安全约束削减方法,通过松弛辅助优化问题以及采用多线程并行计算等方式,加速了识别过程。辅助优化问题方法为冗余约束识别提供了理论基础,然而其所构建的辅助优化问题往往规模较大、实现复杂且求解速度较慢,因此难以适用于面向短期的 SCED 模型求解。相比之下,迭代法同样能够识别所有起作用的约束,且实现较为简便。文献[16]从初始基态调度问题中抽取子约束,通过迭代代入并求解不同故障态约束,直至找出所有起作用的约束。文献[17]对迭代法进行了改进,提出了一种改进的迭代筛选方法,显著减少了需要处理的约束数量,从而提高了计算效率。尽管迭代法实现简单,但在面对大规模电网调度问题时,其需要多次迭代才能识别全部冗余约束,导致求解时间较长,因此仍不适用于 SCED 模型的快速求解。

此外,文献[10]提出了基于潮流转移关系的冗余约束识别方法,但该方法在动态冗余约束识别方面存在不足,难以适应如 SCED 这类对实时性要求较高的场景。文献[18]提出了一种考虑 $N-1$ 故障的混合整数线性规划模型进行约束削减,但是在基态环境下仅考虑了并联支路约束削减,仍然存在很多冗余约束未考虑到。同样文献[19]提出了一种基于电网安全运行域的故障态约束过滤方法,利用电网故障态约束条件之间的实际物理意义与故障态约束的数学关系进行冗余约束识别。而文献[20]提出了一种边界法进行约束简化,该方法可以高效识别冗余约束,但只能针对线性约束进行简化,适用范围

有限。文献[21]基于神经网络模型实现了对冗余约束的快速辨别。该方法属于数据驱动,具有辨识速度快、覆盖范围广的优点,但难免出现分类误判的情况。然而,该文献并未考虑因冗余约束误判而可能引发的系统安全故障问题。

同样是处理复杂电力数据,文献[22-23]使用决策树(decision tree, DT)算法分别对异常数据和光伏直流串联电弧故障进行识别,识别速度快且准确率高。同时,该算法原理相对简单且可解释性强。

综上,本文对分类回归树(classification and regression tree, CART)算法和错误率降低剪枝(reduced error pruning, REP)算法进行了改进,旨在实现 SCED 模型中冗余约束的快速识别,从而更适合 SCED 模型的求解。首先,本文详细阐述了冗余约束识别的特征工程,以确保方法的复现性。其次,针对冗余约束识别中极端假正例(false positive, FP)敏感问题,对 CART 算法进行了改进。需要说明的是,极端 FP 敏感指的是无法接受将非冗余约束误判为冗余约束。具体改进在于,在 Gini 系数的计算中引入了 FP 惩罚机制,这使得各节点在分裂构建决策树时能够不同程度地考虑降低 FP 误判,惩罚系数越高,降低 FP 误判的效果越显著。然后,采用基于 FP 比改进的 REP 方法对生成的决策树进行处理,以避免过拟合。最后,通过算例验证,在 SG-126 系统中冗余约束识别的准确率达到 95.13%, FP 误判率为 0;削减冗余约束后,系统调度时间减少了 88.22%。

1 SCED 模型

1.1 模型描述

SCED 模型是电力系统优化运行的核心工具,其目标是在满足电网安全约束条件下使运行成本最小化。本文中 SCED 模型考虑了机组启停因素,属于扩展的 SCED 模型,同样也是 MIP 问题。

1.2 目标函数

模型优化的目标函数包含发电成本、运行固定成本和启停成本,如式(1)所示。

$$\min \sum_{i \in G_{\text{th}}} \sum_{t \in T} (c_i^{(2)} P_{i,t}^2 + c_i^{(1)} P_{i,t} + c_u u_{i,t} + c_v v_{i,t}) \quad (1)$$

式中: G_{th} 表示全体常规机组集合; T 表示一个调度周期内的全部时段集合; $P_{i,t}$ 表示 t 时段常规机组 i 的出力; $c_i^{(1)}$ 表示常规机组 i 的一次成本系数; $c_i^{(2)}$ 表示常规机组 i 的二次成本系数; c_u 表示常规机组运行的固定成本系数; $u_{i,t}$ 是 0-1 整型变量,表示 t 时段常规机组 i 的启停状态(1 为运行,0 为停机); c_v

表示机组的启动成本系数; $v_{i,t}$ 是 0-1 整型变量, 表示 t 时段常规机组 i 的启动标志。其中, $c_i^{(2)}P_{i,t}^2 + c_i^{(1)}P_{i,t} + c_{d,i}u_{i,t}$ 表示在 t 时段内机组 i 的运行成本, $c_v v_{i,t}$ 表示在 t 时段内机组 i 的启动成本。

1.3 约束条件

本文考虑的 SCED 模型中约束条件如下。

1) 常规机组出力约束

式(2)、式(3)分别为常规机组最小、最大出力约束。

$$P_{i,t} - R_{i,t}^{\text{down}} \geq P_i^{\text{min}} u_{i,t}, \forall i \in G_{\text{th}}, t \in T \quad (2)$$

$$P_{i,t} + R_{i,t}^{\text{up}} \leq P_i^{\text{max}} u_{i,t}, \forall i \in G_{\text{th}}, t \in T \quad (3)$$

式中: $R_{i,t}^{\text{up}}$ 、 $R_{i,t}^{\text{down}}$ 分别表示 t 时段常规机组 i 的上、下备用功率; P_i^{max} 、 P_i^{min} 分别表示常规机组 i 的出力上、下限。

2) 开停机状态约束

式(4)、式(5)分别为开停机状态初始约束和非初始约束。

$$v_{i,0} - w_{i,0} = u_{i,0} - u_i^{\text{init}}, \forall i \in G_{\text{th}} \quad (4)$$

$$v_{i,t} - w_{i,t} = u_{i,t} - u_{i,t-1}, \forall i \in G_{\text{th}}, t \in T \setminus \{0\} \quad (5)$$

式中: $v_{i,0}$ 表示 0 时刻常规机组 i 的启动标志; $u_{i,0}$ 表示 0 时刻常规机组 i 的启停标志; $w_{i,t}$ 是 0-1 整型变量, 表示 t 时段常规机组 i 的停机标志, 其中 $w_{i,0}$ 表示 0 时刻常规机组 i 的停机标志; u_i^{init} 表示常规机组 i 在初始状态下的启停状态; $T \setminus \{0\}$ 表示一个调度周期内除 0 时刻外的全部时段集合。

3) 最小启停时间约束

式(6)、式(7)表示初始状态下的最小启停时间约束; 式(8)、式(9)表示最小启停时间约束。

$$\sum_{\tau=0}^{T_i^{\text{up,init}}} w_{i,\tau} \leq 1 - v_{i,0}, \forall i \in G_{\text{th}} \quad (6)$$

$$\sum_{\tau=0}^{T_i^{\text{down,init}}} v_{i,\tau} \leq 1 - w_{i,0}, \forall i \in G_{\text{th}} \quad (7)$$

$$\sum_{\tau=t}^{t+T_i^{\text{up}}-1} w_{i,\tau} \leq 1 - v_{i,t}, \forall i \in G_{\text{th}}, t \in T \setminus \{0\} \quad (8)$$

$$\sum_{\tau=t}^{t+T_i^{\text{down}}-1} v_{i,\tau} \leq 1 - w_{i,t}, \forall i \in G_{\text{th}}, t \in T \setminus \{0\} \quad (9)$$

式中: $T_i^{\text{up,init}}$ 、 $T_i^{\text{down,init}}$ 分别表示初始状态下常规机组 i 的最小连续运行、停机时间; T_i^{up} 、 T_i^{down} 分别表示常规机组 i 的最小连续运行、停机时间。

4) 常规机组爬坡率约束

式(10)、式(11)分别为初始状态下的常规机组上、下爬坡率约束; 式(12)、式(13)分别为常规机组上、下爬坡率约束。

$$P_{i,0} - P_i^{\text{init}} + R_{i,0}^{\text{up}} \leq \Delta P_i^{\text{up}} u_i^{\text{init}} + P_i^{\text{min}} v_{i,0}, \forall i \in G_{\text{th}} \quad (10)$$

$$P_{i,0} - P_i^{\text{init}} + R_{i,0}^{\text{up}} \leq \Delta P_i^{\text{down}} u_i^{\text{init}} - P_i^{\text{min}} w_{i,0}, \forall i \in G_{\text{th}} \quad (11)$$

$$P_{i,t} - P_{i,t-1} + R_{i,t}^{\text{up}} \leq \Delta P_i^{\text{up}} u_{i,t-1} + P_i^{\text{min}} v_{i,t}, \quad (12)$$

$$\forall i \in G_{\text{th}}, t \in T \setminus \{0\}$$

$$P_{i,t} - P_{i,t-1} + R_{i,t}^{\text{up}} \leq \Delta P_i^{\text{down}} u_{i,t-1} - P_i^{\text{min}} w_{i,t}, \quad (13)$$

$$\forall i \in G_{\text{th}}, t \in T \setminus \{0\}$$

式中: $P_{i,0}$ 表示 0 时刻常规机组 i 的出力; $R_{i,0}^{\text{up}}$ 表示 0 时刻常规机组 i 的上备用功率; P_i^{init} 表示常规机组 i 的初始出力; ΔP_i^{up} 、 ΔP_i^{down} 分别表示常规机组 i 的上、下爬坡速率限值。

5) 新能源机组出力约束

$$0 \leq P_{r,t} \leq P_{r,t}^{\text{max}}, \forall r \in G_{\text{re}}, t \in T \quad (14)$$

式中: G_{re} 表示新能源机组集合; $P_{r,t}$ 表示 t 时段新能源机组 r 的出力; $P_{r,t}^{\text{max}}$ 表示在 t 时段内新能源机组 r 的最大预测出力。

6) 功率平衡约束

$$\sum_{n \in G} P_{n,t} = \sum_{l \in L} D_{l,t}, \forall t \in T \quad (15)$$

式中: G 表示全体机组集合, 包括 G_{th} 和 G_{re} ; L 表示模型中负荷集合; $D_{l,t}$ 表示 t 时段内负荷 l 的需求功率预测值。

7) 直流潮流约束

实际的 SCED 模型庞大, 计算困难, 通常将潮流约束简化为直流潮流约束, 见式(16)。后续在构建 DT 的特征工程时, 需要重构线性模型以计算对偶变量等重要指标, 使用简化的直流约束更容易计算对偶变量。

$$-\bar{F}_k \leq \sum_{i \in G} G_{k,b(i)} P_{i,t} - \sum_{l \in L} G_{k,b(l)} D_{l,t} \leq \bar{F}_k, \quad (16)$$

$$\forall k \in K, t \in T$$

式中: K 表示模型中所有线路集合; \bar{F}_k 表示线路 k 的热稳定极限; $G_{k,b}$ 表示线路 k 对母线 b 的功率转移分布因子 (generation shift distribution factor, GSDF), 其中 $G_{k,b(i)}$ 、 $G_{k,b(l)}$ 分别表示只与机组 i 、负荷 l 相连节点的 GSDF。

8) 联络线断面约束

$$\sum_{i \in G} G_{k,b(i)} P_{i,t} - \sum_{l \in L} G_{k,b(k_{\text{cross}})} D_{k_{\text{cross}},t} \leq F^{\text{max}}, \quad (17)$$

$$\forall k_{\text{cross}} \in K_{\text{cross}}, t \in T$$

式中: K_{cross} 表示联络线断面内的线路集合; F^{max} 表示联络线断面功率传输极限。

9) 备用容量约束

式(18)、式(19)分别为上、下机组备用容量约束。

$$\sum_{i \in G_{th}} R_{i,t}^{up} \geq \eta \sum_{r \in G_{re}} P_{r,t}^{max}, \forall t \in T \quad (18)$$

$$\sum_{i \in G_{th}} R_{i,t}^{down} \geq \eta \sum_{r \in G_{re}} P_{r,t}^{max}, \forall t \in T \quad (19)$$

式中： η 表示备用容量需求比例系数。

2 基于决策树的冗余约束识别

2.1 CART 算法

决策树分类算法众多，其典型算法包括 ID3、C4.5、CART 和 SLIQ 等。通过这些算法能够对庞大的数据进行分类或回归分析。而冗余约束识别的本质是一个二分类问题，即判断每个系统中的约束条件为冗余或非冗余。基于此，本文选择 CART 二叉树进行冗余约束快速识别。

CART 二叉树本质是先后数次选择不同的分裂特征把数据分成两类，最终精准地将不同数据正确归类的分类器。CART 算法是采用 Gini 系数选择内部节点的分裂特征，Gini 系数表示样本集划分的不纯度，所以 Gini 系数越小，表示样本的纯度越高。该系数的计算方法如下。

1) 样本集 N 在某节点的 Gini 系数 $G(N)$ 为

$$G(N) = 1 - \sum_{e=1}^n p_e^2 \quad (20)$$

式中： p_e 表示类别 e 在某节点样本总数中所占的比例。

2) 如特征 F 将集合 N 在该节点上二分为 N_1 和 N_2 ，则该特征的 Gini 系数 $G_{split}(N)$ 为

$$G_{split}(N) = \frac{N_1}{C_N} G(N_1) + \frac{N_2}{C_N} G(N_2) \quad (21)$$

式中： C_N 表示样本集 N 中的样本总数，其中 $C_N = N_1 + N_2$ ； $G_{split}(N)$ 是特征 F 的 Gini 系数，即基于该特征分裂后的总 Gini 系数。

3) 基于特征 F 进行分裂后，Gini 系数减少量为 $\Delta G(N)$ 。

$$\Delta G(N) = G(N) - G_{split}(N) \quad (22)$$

$\Delta G(N)$ 越大，说明该特征对降低数据的不纯度的影响力越大，其更应该被优先选取。因此，本文将 $\Delta G(N)$ 视为 Gini 重要度。

此外，CART 算法是一种递归二分算法，在每个节点分裂时只能生成两个子节点，从而确保树结构简单。同时采用贪心策略选择分裂点，即选择 $\Delta G(N)$ 最大的组合。依据特征二分规则将样本集递归划分为子集，生成左右子树。最终通过剪枝策略对全树进行后剪枝，以平衡模型复杂度与泛化性能。

综上所述，CART 算法主要包括特征选择和基于剪枝的最优 DT 构建^[24]。

2.2 冗余约束定义

本文中对于冗余约束的定义为：对于一个优化模型，若将某条约束从全体约束集合中移除后，不会改变原模型的可行解集合，也不会影响最优解的存在性和数值，那么这条约束是冗余的。如图 1 所示，不影响最终可行域的 3 条约束即为冗余约束。

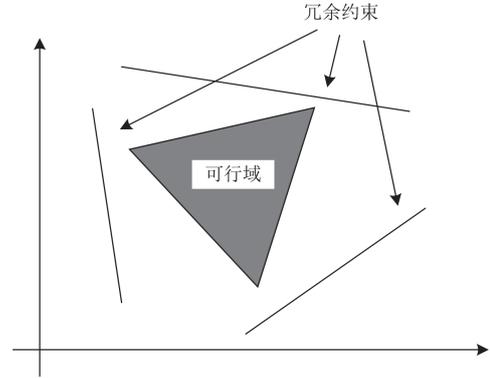


图 1 冗余约束示例图

Fig. 1 Redundant constraints example diagram

2.3 冗余约束识别的特征工程

首先需要明确的是特征工程需要基于已有的数据进行构建。进行冗余约束识别的时间节点是在实行模型调度优化之前，因此能获取到的关于本次优化调度的信息是有限的。

在完成实际调度之前，如发电机实际出力、线路实际潮流等各类实时数据是未知的。所以本文通过历史数据和固定数据作为构建特征工程的依据。在每完成一次调度后，如历史平均负载率此类历史数据会进行滚动更新。基于此，本文把基于历史数据构建的特征设计为滚动特征。对应地，把基于固定数据构建的特征设计为固定特征。

特征量过多可能包含冗余信息，会对模型识别造成误差，降低识别的准确率^[25]。如表 1 所示，本文只选取适用于冗余约束识别的 12 个重要特征。

需要说明的是，导入 CART 算法的数据不能空缺且必须是数字，否则将严重影响训练结果。因此，在进行特征计算时，需要通过独热编码、归一化和全局变量均值替换等方式处理数据。

1) 约束类型

在 SCED 模型中，约束条件集合主要由式(2)一式(19)组成，因此将约束分为最小与最大机组出力约束、常规启停约束、上下爬坡约束、功率平衡约束、直流潮流约束、联络线约束以及机组备用约束等 17 类约束。

为了便于 DT 进行识别，选择对约束类型数据进行独热编码处理。

表 1 适用于冗余约束识别的滚动特征和固定特征

Table 1 Rolling features and fixed features for redundant constraint identification

特征类别	特征名称	选择原因
固定特征	约束类型	在特定系统中, 特定类型的约束可能经常性冗余
	等式约束属性	等式约束通常不可松弛, 冗余概率低; 不等式约束更可能冗余
	线性约束属性	在 MIP 问题中, 选取线性约束属性为特征之一, 符合模型必要性
固定特征	安全必要约束属性	安全必要约束通常认定为非冗余约束
	电气介数	高介数关联的约束通常对系统安全影响较大, 冗余概率低; 反之概率相对较高
	所属节点度	节点度数高的节点通常处于网络枢纽位置, 其约束冗余概率较低; 低度数节点约束冗余概率相对较高
滚动特征	平均新能源高发时期属性	在高比例新能源电力系统中, 新能源出力的波动性会显著影响约束集合的活跃性
	历史平均负载率	长期低负载率表明约束限制宽松, 是冗余的强信号
	峰谷差比	峰谷差比大的约束可能仅在极端场景下有效, 可能是偶发活跃约束
	历史平均松弛绝对值	松弛变量均值较大表明约束极少处于活跃状态, 是冗余的核心指标
	历史平均对偶变量	对偶变量趋于 0 的约束对目标函数影响可忽略, 是冗余的核心指标
	历史约束触发概率	低触发概率的约束在多数场景下不活跃; 高触发概率的约束通常为非冗余约束

2) 等式约束属性

约束是等式时属性为 1, 非等式时属性为 0。

3) 线性约束属性

SCED 模型中包含启停因素, 这使模型中存在一部分非线性约束。当约束是线性约束时属性为 1, 约束是非线性约束时属性为 0。

4) 安全必要约束属性

在 SCED 模型中, 启停逻辑约束式(4)一式(9)是相对独立的约束集, 不存在被其他约束集覆盖的情况, 可以视为安全必要约束, 是一种非冗余约束。当一条约束视为安全必要约束时属性为 1, 否则为 0。

5) 电气介数

电气介数可以量化总线节点或支路等独立元件对电网传输能力的贡献度, 即在网络中所处的重要程度。电气介数越大, 说明输电线路上的功率波动

对于其他线路的影响越大^[26]。

在优化模型中, 与不同总线节点相连的常规机组的介数以及不同支路的介数计算如下。

节点 v 的电气介数 $C_B(v)$ 的计算如式(23)所示。

$$C_B(v) = \frac{1}{(|sh|-1)(|sh|-2)} \sum_{\substack{s,h \in V \\ s \neq v \neq h}} \frac{\sigma_{sh}(v)}{\sigma_{sh}}, \forall v \in V \quad (23)$$

式中: V 表示全体总线节点的集合; $|sh|$ 表示全体总线节点 V 中所有总线节点在系统中两两相连的线路总数; σ_{sh} 表示节点 s 到节点 h 的最小路径总数(即最大导纳路径总数); $\sigma_{sh}(v)$ 表示节点 s 到节点 h 的最大导纳路径中经过节点 v 的路径数。其中, 对输入的数据进行归一化处理, 归一化后数据不存在量纲, 且取值范围在 $[0,1]$ 之间^[27], 确保数值可解释性。

支路 k 的电气介数 $C_B(k)$ 的计算如式(24)所示。

$$C_B(k) = \frac{1}{(|sh|-1)(|sh|-2)} \sum_{\substack{s,h \in V \\ s \neq h}} \frac{\sigma_{sh}(k)}{\sigma_{sh}}, \forall k \in K \quad (24)$$

式中: $\sigma_{sh}(k)$ 表示节点 s 到节点 h 的最大导纳路径中经过支路 k 的路径数。

在 SCED 模型中, 对于不属于特定节点或支路的全局约束, 比如总功率平衡约束, 取平均支路电气介数进行赋值, 以保证数据完整。

平均支路电气介数 \bar{C}_B 的计算如式(25)所示。

$$\bar{C}_B = \frac{1}{|K|} \sum_{k \in K} C_B(k) \quad (25)$$

式中: $|K|$ 表示全体支路的总数。

6) 所属节点度

所属节点度是衡量节点在网络中连通性的核心指标, 节点 v 所属节点度 $S(v)$ 的计算如式(26)所示。

$$S(v) = d(v), \forall v \in V \quad (26)$$

式中: $d(v)$ 表示节点 v 连接的支路数。

同样地, 全局约束集合没有所处节点, 可以使用平均所属节点度 \bar{S} 代替, \bar{S} 的计算如式(27)所示。

$$\bar{S} = \frac{1}{|V|} \sum_{v \in V} S(v) \quad (27)$$

式中: $|V|$ 表示全体总线节点的总数。

7) 平均新能源高发时期属性

该特征的历史数据来源于实际调度模型, 统计每个调度周期的新能源高发时段, 然后再取均值即可计算出平均新能源高发时段。

图 2 为某省 7 月平均新能源出力曲线。若基于该省电网构建特征工程, 新能源高发时段是 10:00—18:00, 在该时间段内属性为 1, 否则属性为 0。

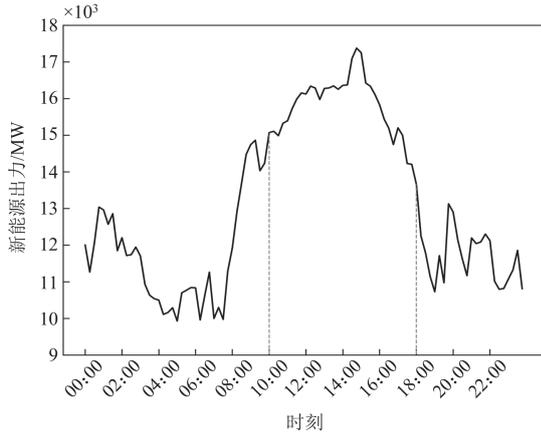


图2 某省7月平均新能源出力曲线

Fig. 2 Average output curve of new energy of a province in July

8) 历史平均负载率

历史平均负载率特征是基于历史负载率计算得到的,不同类型的约束集合需要根据各自的物理意义进行负载率计算。

历史平均常规机组、新能源机组出力负载率 ρ_i^p 、 ρ_r^p 计算分别为

$$\rho_i^p = \frac{1}{|T|} \sum_{t=1}^T \frac{P_{i,t}}{P_i^{\max}} \times 100\% \quad (28)$$

$$\rho_r^p = \frac{1}{|T|} \sum_{t=1}^T \frac{P_{r,t}}{P_r^{\max}} \times 100\% \quad (29)$$

式中: $|T|$ 表示历史数据时间窗口长度。

历史平均机组上、下坡负载率 ρ_i^{up} 、 ρ_i^{down} 计算分别为

$$\rho_i^{\text{up}} = \frac{1}{|T|} \sum_{t=1}^T \frac{|P_{i,t} - P_{i,t-1}|}{\Delta P_i^{\text{up}}} \times 100\% \quad (30)$$

$$\rho_i^{\text{down}} = \frac{1}{|T|} \sum_{t=1}^T \frac{|P_{i,t} - P_{i,t-1}|}{\Delta P_i^{\text{down}}} \times 100\% \quad (31)$$

历史平均直流潮流负载率 ρ_k^{dc} 为

$$\rho_k^{\text{dc}} = \frac{1}{|T|} \sum_{t=1}^T \frac{\sum_{i \in G} G_{k,b(i)} P_{i,t} - \sum_{l \in L} G_{k,b(l)} D_{l,t}}{F_k} \times 100\% \quad (32)$$

历史平均联络线负载率 ρ_k^{cross} 为

$$\rho_k^{\text{cross}} = \frac{1}{|T|} \sum_{t=1}^T \frac{\sum_{i \in G} G_{k,b(i)} P_{i,t} - \sum_{l \in L} G_{k,b(l)} D_{l,t}}{F_k^{\max}} \times 100\% \quad (33)$$

历史平均上、下备用负载率 ρ_R^{up} 、 ρ_R^{down} 分别为

$$\rho_R^{\text{up}} = \frac{1}{|T|} \sum_{t=1}^T \frac{\sum_{i \in G_{\text{th}}} R_{i,t}^{\text{up}}}{\sum_{r \in G_{\text{re}}} P_{r,t}^{\max}} \times 100\% \quad (34)$$

$$\rho_R^{\text{down}} = \frac{1}{|T|} \sum_{t=1}^T \frac{\sum_{i \in G_{\text{th}}} R_{i,t}^{\text{down}}}{\sum_{r \in G_{\text{re}}} P_{r,t}^{\max}} \times 100\% \quad (35)$$

同样地,以防全局约束集合该部分数据空缺,使用历史平均全局约束负载率 ρ_{dc} 代替。

$$\rho_{\text{dc}} = \frac{1}{|K|} \sum_{k=1}^K \rho_k^{\text{corss}} \quad (36)$$

9) 峰谷差比

峰谷差比 r_{PV} 同样通过历史负载率数据计算得出,计算方法如式(37)所示。

$$r_{\text{PV}} = \frac{L_{\max} - L_{\min}}{\bar{L}} \quad (37)$$

式中: L_{\max} 、 L_{\min} 分别表示负载率的最大值和最小值; \bar{L} 表示负载率平均值。

10) 历史平均松弛绝对值

历史平均松弛绝对值是由历史松弛值计算得到,用来量化各约束的历史松弛程度。松弛绝对值 $|X_{\text{slack}}|$ 的计算如式(38)所示。

$$|X_{\text{slack}}| = |X_{\text{rhs}} - X_{\text{lhs}}| \quad (38)$$

式中: X_{rhs} 表示约束式右端项; X_{lhs} 表示约束式左端项。

11) 历史平均对偶变量

历史平均对偶变量是通过计算历史对偶变量的均值得到的。对偶变量的本质是量化约束条件对目标函数的敏感程度,具体表现为约束条件边界的微小变化对目标函数最优值的影响强度。

假设原始问题为

$$\begin{cases} \min & c^T x \\ \text{s.t.} & Ax = b \quad (\lambda) \\ & x \geq 0 \quad (\mu) \end{cases} \quad (39)$$

式中: $\min c^T x$ 表示目标函数; $Ax = b$ 表示全体等式约束集合; $x \geq 0$ 表示全体不等式约束集合。

对应等式对偶变量 λ 和不等式对偶变量 μ 满足式(40)。

$$\frac{\partial(c^T x^*)}{\partial b_i} = \lambda_i, \quad \frac{\partial(c^T x^*)}{\partial x_j} = \mu_j \quad (40)$$

式中: 对偶变量 λ_i 表示第 i 个等式约束的右端项 b_i 增加 1 单位,目标函数值的变化量; 对偶变量 μ_j 表示第 j 个等式约束的下界 x_j 增加 1 单位,目标函数值的变化量; x^* 表示最优解。

12) 历史约束触发概率

历史约束触发概率是判断冗余的核心指标,其通过直接计算各约束触发次数的均值即可得到。

3 基于改进的 CART 和 REP 算法的冗余约束识别

为将决策树方法应用于 SCED 冗余约束识别问题中, 本文将 SCED 模型中每一条约束视为一个样本, 通过设计比如各约束的历史松弛度、对偶变量、所属线路节点等特征作为 CART 决策树的输入特征, 用于分类模型学习。分类标签设定为“是否为冗余约束”, 其中冗余约束的定义来源于 SCED 模型约束削减前后目标函数是否变动的统计结果。因此, CART 与 REP 算法的每一次决策节点划分, 实质是对 SCED 中某一类物理或历史统计特征进行边界切分, 从而达到区分冗余/非冗余的目的。

通过决策树这类数据驱动方法进行冗余约束识别, 本质是学习历史, 进而对未来情况进行预测。与其他需要尽可能提高模型预测精度和避免过拟合的分类问题不同, 冗余约束识别问题是极端 FP 敏感问题, 即不能接受将非冗余约束识别为冗余约束, 对 FP 有严格归零要求。因为一旦出现 FP 误判, 必要约束被误删, 优化结果难以保证其安全性, 进而不允许进行实际模型的约束削减。基于此, 必须使用改进的决策树以适应极端 FP 敏感情况, 从而达到无 FP 误判的预测结果。

3.1 混淆矩阵和 FP 比的定义

通常, 混淆矩阵是个 2×2 的矩阵, 包含表 2 所示的 4 个核心指标。

表 2 混淆矩阵核心指标

Table 2 Confusion matrix core index

	真实为正类	真实为负类
预测为正类	真正例(true positive, TP)	假正例(false positive, FP)
预测为负类	假负例(false negative, FN)	真负例(true negative, TN)

在冗余约束识别问题中, TP 表示被正确识别并删除的冗余约束; FP 表示被错误删除的非冗余约束; FN 表示被错误保留的冗余约束; TN 表示被正确保留的非冗余约束。

FP 比 r_{FP} 定义为

$$r_{FP} = \frac{E_{FP}}{E_{TP} + E_{FP} + E_{FN} + E_{TN}} \quad (41)$$

准确率 P_{Acc} 定义为

$$P_{Acc} = \frac{E_{TP} + E_{TN}}{E_{TP} + E_{FP} + E_{FN} + E_{TN}} \quad (42)$$

式中: E_{TP} 表示被正确识别并删除的冗余约束数量; E_{FP} 表示被错误删除的非冗余约束数量; E_{FN} 表示被错误保留的冗余约束数量; E_{TN} 表示被正确保留

的非冗余约束数量; r_{FP} 表示 SCED 中被误删的非冗余约束数量占总约束数量的比例, 用来量化 FP 在整个预测集中的占比, 作为核心指标来判断预测结果在极端 FP 敏感场景下是否有效。若 r_{FP} 为 0, 代表不存在 FP 误判, 这也是冗余约束识别后能对实际模型进行约束削减的必要条件之一。DT 在可实现 FP 归零的情况下有效; 反之, 则无效。

3.2 基于 FP 惩罚机制改进的 CART 算法

为使构建的 DT 能够适应极端 FP 敏感场景, 以 FP 惩罚机制为核心对 DT 的构建进行改进。具体改进方式是将 CART 算法的核心参数 Gini 系数进行改进, 在原始算法的基础上引入 FP 惩罚机制。在生成 DT 时, 改进算法倾向于选择能够同时降低不纯度和 FP 比的特征, 从而在早期分裂阶段排除存在高 FP 误判风险的路径。

改进后的各系数如式(43)一式(45)所示。

1) SCED 约束样本集 N 在某节点的改进 Gini 系数 $G_{ad}(N)$ 为

$$G_{ad}(N) = \alpha_{FP} \cdot r_{FP}(N) + (1 - \alpha_{FP}) \cdot G(N) \quad (43)$$

式中: $r_{FP}(N)$ 表示 SCED 中被误删的非冗余约束数量占约束样本集 N 中总样本数量的比例, 用以量化 FP 在约束样本集 N 中的比例; α_{FP} 为 FP 比惩罚系数。

2) 若特征 F 将约束样本集 N 在该节点上二分为 N_1 和 N_2 , 则该特征的改进 Gini 系数 $G_{ad,split}(N)$ 为

$$G_{ad,split}(N) = \frac{N_1}{C_N} G_{ad}(N_1) + \frac{N_2}{C_N} G_{ad}(N_2) \quad (44)$$

3) 基于特征 F 进行分裂后, 改进的 Gini 系数减少量 $\Delta G_{ad}(N)$ 为

$$\Delta G_{ad}(N) = G_{ad}(N) - G_{ad,split}(N) \quad (45)$$

在式(43)中的约束样本集 N 即为 SCED 模型中的约束集合(如常规机组出力约束、机组爬坡约束等), 将不同的约束输入 DT 进行后续识别; 在式(44)和式(45)中, 特征 F 是表 1 中的“历史平均松弛绝对值”、“电气介数”等具体特征。

改进的决策树构建将会依据最大 $\Delta G_{ad}(N)$ 原则进行特征选择, 从而满足适应极端 FP 敏感场景的特殊要求。

综上, FP 比惩罚系数 α_{FP} 的取值直接影响 FP 的生成, 即影响模型的 FP 的抑制能力。若 α_{FP} 过大, 可能导致决策树过度保守, 降低整体准确率; 若 α_{FP} 过小, 则无法有效降低 FP 比。惩罚系数 α_{FP} 的调整是为了在 SCED 这个安全至上的场景下强制压降 FP, 其具体的取值需根据算例实际情况进行平衡。

此外, 时间复杂度表示算法在使用过程所需

的时间开销, CART 算法在构建二叉树时的时间复杂度 d_{CART} 的计算如式(46)所示。

$$d_{\text{CART}} = O(m \cdot n \cdot \log n) \quad (46)$$

式中: m 为特征数; n 为样本数; $\log n$ 表示平衡二叉树的平均深度; $O(\cdot)$ 表示时间复杂度计算公式。

式(46)说明 d_{CART} 的取值只与 m 和 n 有关, 在引入 FP 比惩罚机制后, DT 在能够适应极端 FP 敏感场景的同时, d_{CART} 不会增加, 仅有计算 FP 比的额外开销, 改进算法相比原始算法构建 DT 的时间只有微小增加, 适用于求解面向短期的 SCED 模型。

在 DT 构建的原始方法中, 通常通过代价敏感学习的方式处理不平衡问题, 一般通过改变代价权重来实现, 此时 Gini 系数的计算如式(47)所示。

$$G'(N) = 1 - \sum_{e=1}^n w_e p_e^2 \quad (47)$$

式中: w_e 表示类别 e 在某节点的代价权重。

这种方法可以平衡各类别的分类性能, 适用于需要灵活权衡 FP 和 FN 的场景, 可以做到有效抑制 FP, 但在大多数场景中无法进行 FP 归零。若将 FP 强行归零需设定无限代价权重, 这会大幅牺牲总体准确率。而相比之下, 基于 FP 比惩罚机制改进的决策树直接针对 FP 进行优化, 能在保证高准确率的同时, 完成 FP 归零。具体的对比实验将在第 4 节中介绍。

通过调整改进算法中 α_{FP} 参数可以生成若干未剪枝 DT, 将其中可以实现 FP 归零的 DT 移入有效候选集; 反之, 移入无效候选集中。为进一步剪枝做准备。

3.3 基于 FP 比改进的 REP 算法

DT 在训练过程中可能因过拟合导致泛化能力不足, 后剪枝是解决这一问题的有效方法之一。REP 是经典的后剪枝策略, 其通过权衡模型的总体错误率来寻找最优剪枝点^[28]。

传统 REP 的基本思路是, 对于决策树 T 的每棵非叶子树 S , 用叶子 t 代替 S 的方式进行剪枝, 若替换后的新树 T_i 误差损失 $R(T_i)$ 小于 $R(T)$, 则允许剪枝。

根据以上内容, 传统 REP 的优化目标可以表示为

$$\text{Select} T = \arg \min_{T_i} R(T_i) \quad (48)$$

其中, $R(T)$ 的计算可用混淆矩阵的核心参数表示为

$$R(T) = 1 - P_{\text{Acc}} = \frac{E_{\text{FP}} + E_{\text{FN}}}{E_{\text{TP}} + E_{\text{FP}} + E_{\text{FN}} + E_{\text{TN}}} \quad (49)$$

传统 REP 难以有效作用于极端 FP 敏感场景, 无法满足 SCED 对安全约束零误删($\text{FP} = 0$)的刚性要求。由式(49)可知, $R(T)$ 中考虑了 FN。又因为由改进 CART 算法构建的 DT 中, 普遍存在 $E_{\text{FN}} \gg E_{\text{FP}}$ 的情况, 此时 $R(T)$ 可近似表示为

$$R(T) \approx \frac{E_{\text{FN}}}{E_{\text{TP}} + E_{\text{FP}} + E_{\text{FN}} + E_{\text{TN}}} \quad (50)$$

式中: $E_{\text{TP}} + E_{\text{FP}} + E_{\text{FN}} + E_{\text{TN}}$ 总数不变, 可以近似认为 $R(T)$ 与 E_{FN} 呈正相关。可以近似认为传统 REP 在该问题上以减少 FN 为目标进行剪枝, 减少 FN 并不能减少 FP, 反而可能使其增加, 所以这种剪枝策略难以达到减少 FP 的目的, 无法做到在提高泛化性的同时做到 FP 归零。这也是传统 REP 在解决冗余约束识别问题上面临的困境, 需要对 REP 进行改进。

基于此, 本文设计的改进 REP 算法的优化目标表示为

$$\text{Select} T^* = \arg \min_{T_i} [\lambda r_{\text{FP}}(T_i) + (1 - \lambda) r_{\text{FN}}(T_i)] \quad (51)$$

式中: λ 是剪枝系数; $r_{\text{FP}}(T_i)$ 表示 SCED 中被保留的冗余约束数量占树 T_i 中总约束样本数量的比例, 用以量化 FN 在树 T_i 中的比例, $r_{\text{FP}}(T_i)$ 可以理解为安全选择项; $r_{\text{FN}}(T_i)$ 则为效率损失项。 λ 越大表示剪枝更加注重抑制 FP, 不可误删 SCED 中的非冗余约束, 比如式(4)一式(9)此类 SCED 中的关键安全约束; λ 越小表示剪枝更加注重提高识别准确率, 削减被保留的冗余约束, 如式(3)此类 SCED 中的长期闲置的出力上限约束。

其中, FN 比 r_{FN} 的计算可用混淆矩阵的核心参数表示为

$$r_{\text{FN}} = \frac{E_{\text{FN}}}{E_{\text{TP}} + E_{\text{FP}} + E_{\text{FN}} + E_{\text{TN}}} \quad (52)$$

为方便描述, 将这种结合了 FP 和 FN 的代价敏感指标表示为 $R^*(T)$ 。

$$R^*(T) = \lambda r_{\text{FP}}(T_i) + (1 - \lambda) r_{\text{FN}}(T_i) \quad (53)$$

改进的 REP 算法将剪枝决策指标从 $R(T)$ 替换成了 $R^*(T)$, 这使得针对不同的 DT 使用特定的剪枝系数进行剪枝时, 均能保证不增加 FP。

如果参与后剪枝的 DT 来源于有效 DT 候选集, 选择适当小的剪枝系数, 能保持 FP 归零同时减小复杂度; 如果 DT 来源于无效集, 表明在参与后剪枝之前 FP 比不为 0, 选择较大的剪枝系数, 剪枝后有可能令 DT 逐步达成 FP 归零的条件。最后, 将实现 FP 归零的 DT 加入或保留到有效集, 反之加入或丢弃到无效集。以上针对性地小范围参数遍历能够缩短调参时间, 进一步加速优化。

完成改进的 REP 后, 在更新后的有效集中提取 DT 进行寻优, 最终取准确率高者为最优 DT。

3.4 SCED 冗余约束识别的完整流程描述

基于上述介绍, SCED 冗余约束快速识别的完整流程如表 3 所示。

表 3 完整流程描述

序号	步骤
步骤 1	数据准备: 输入历史 SCED 求解结果、约束参数库、电网拓扑数据等
步骤 2	数据预处理: 通过独热编码、归一化等方式处理数据, 以便决策树算法正常运行
步骤 3	特征工程: 将表 1 中各固定特征和滚动特征作为决策树的特征准备
步骤 4	改进 CART 训练: 通过 FP 惩罚机制压降 FP 误判, 遍历惩罚系数 α_{FP} , 通过不同标准的分裂得到 DT 集

续表 3

序号	步骤
步骤 5	DT 集筛选: 从不同 α_{FP} 生成的 DT 中, 选择满足 FP 为 0 的 DT 加入有效集, 反之加入无效集
步骤 6	改进 REP 剪枝: 通过引入剪枝系数 λ 实现剪枝不增加 FP 误判, 根据不同集合遍历不同范围的 λ , 得到新 DT 集
步骤 7	更新 DT 集: 从不同 (α_{FP}, λ) 组合生成的 DT 中, 选择满足 FP 为 0 的 DT 加入或保留到有效集, 反之加入或丢弃到无效集
步骤 8	模型优选: 在有效集中选择准确率最高的为最优 DT
步骤 9	在线应用: 通过最优 DT 对新 SCED 场景进行冗余约束预测, 然后削减冗余约束得到简化 SCED, 实现加速求解
步骤 10	滚动更新: 将新的调度数据加入到历史 SCED 求解结果和滚动特征的计算中, 并重复步骤 1 至步骤 9

SCED 冗余约束快速识别的完整流程图如图 3 所示。

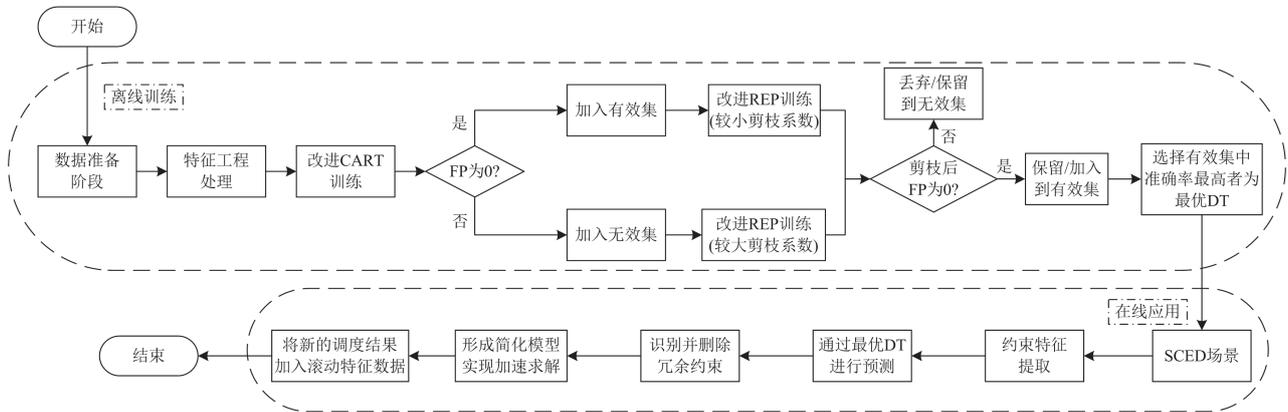


图 3 SCED 冗余约束快速识别的完整流程图

Fig. 3 Complete flowchart for quickly identifying the redundant constraints of SCED

4 算例分析

本次实验编程语言为 python, 编译器为 pycharm, 求解器为 gurobi 12.0.1, 所用计算机配置为 Intel® Core™ i5-14400F CPU@2.50 GHz, 16 GB 内存。

4.1 算例模型

将 SG-126 总线系统作为测算算例, 系统模型信息如图 4 所示。

SG-126 系统统计信息如表 4 所示。

图 4 中, 线路 b43、b44、b53、b29 为区域 1 和 2 之间的联络线, 组成断面。线路 b113、b114、b115、b118、b125 为区域 2 和 3 之间的联络线, 组成断面。

实验过程中, 选择新能源大发的午间时段: 10:00—18:00 时段, 共 8 h, 以 15 min 为时间分辨

率, 共 24 个时刻点。

4.2 结果分析

按照 2.3 节所述的方法进行特征工程构建, 进一步通过改进的 CART 算法进行 DT 初步构建。在实际工作中, 仅通过引入 FP 比惩罚系数进行 FP 削减很难取得理想效果, 往往需配合式(47)改变代价权重来实现 FP 归零。

在不同代价权重下, 改变 α_{FP} 时 FP 比和准确率的变化情况分别如图 5 和图 6 所示。

由图 5 和图 6 可知, 代价权重越倾向于非冗余约束, 即标签为 0, 则 FP 比普遍越低, 实现 FP 归零的临界值 α_{FP}^* 也越小; 但相应的准确率也越低。

当 FP 惩罚系数 α_{FP} 为 0 时, 表示所构建的决策树(DT)采用的是原始的 CART 算法, 未引入 FP 惩罚机制。当 α_{FP} 逐渐变大, 表示 DT 构建时更加倾向于减小 FP 比, 此时 FP 比会变小, 但作为代价, 准

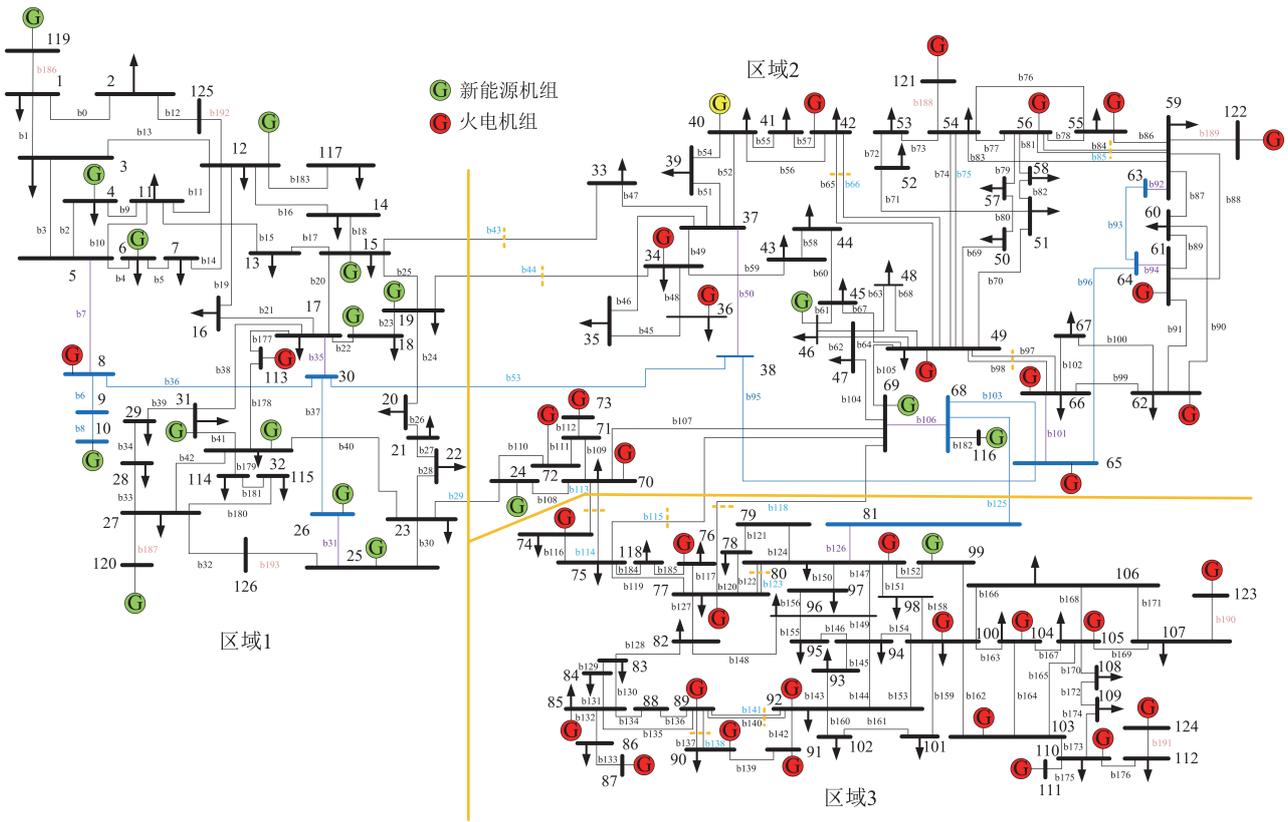


图 4 SG-126 总线系统拓扑图

Fig. 4 Topology diagram of the SG-126 bus system

表 4 SG-126 系统统计信息

Table 4 Statistical information of the SG-126 system

区域	新能源装机/MW	火电装机/MW	负荷总和/MW
1	1115	275	1190
2	380	2483	400
3	100	2570	1193

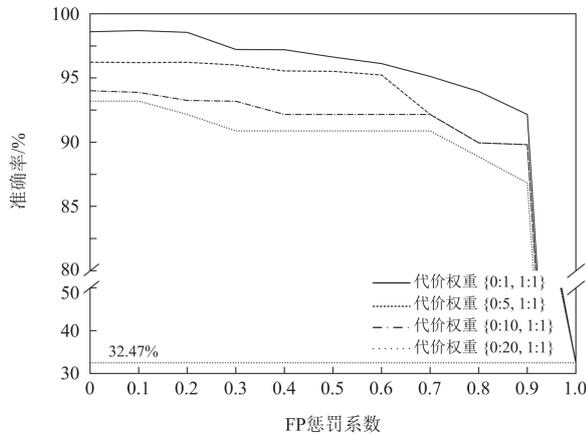


图 5 不同代价权重时惩罚系数和 FP 比的折线关系图

Fig. 5 Line graph of the penalty coefficient and FP ratio at different cost weights

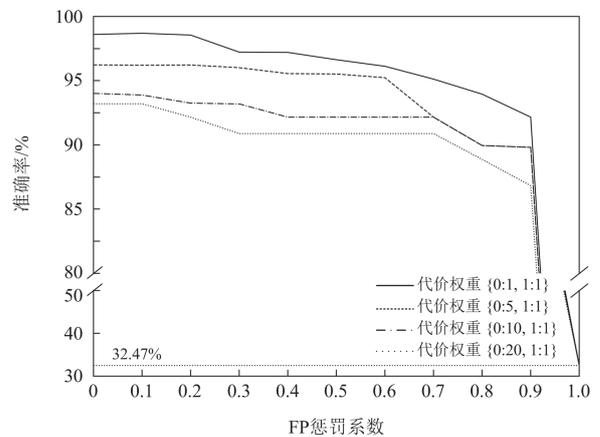


图 6 不同代价权重时惩罚系数和准确率的折线关系图

Fig. 6 Line graph of the penalty coefficient and the accuracy rate at different cost weights

准确率大概率也会变低。当 α_{FP} 变大超过临界值 α_{FP}^* 时, FP 已归零, 准确率将出现不同程度的降低。最后, 当 α_{FP} 逐渐变大到 1 时, 表示 DT 构建时单一地考虑减小 FP 比, 此时不论代价权重是多少均能实现 FP 归零, 但是代价是准确率显著降低, 这个准确率的值为实际非冗余约束总数在约束总数中的占

比, 在该算例中为 32.47%。

当 α_{FP} 大于临界值 α_{FP}^* 时, 此时 DT 可以实现 FP 归零, 则移入上文所述的有效 DT 集; 反之, 移入无效集中。

初步构建 DT 后, 通过改进的 REP 来降低复杂度, 以防过拟合。对有效集和无效集中的典型 DT 进行改进的 REP, 后剪枝过程中改变剪枝系数 λ 时 FP 比、准确率及复杂度的变化如图 7 和图 8 所示。

在未剪枝前, 有效集中的典型 DT1 各项数据为: FP 比为 0.00%, 准确率为 92.12%, 叶子数为 18。

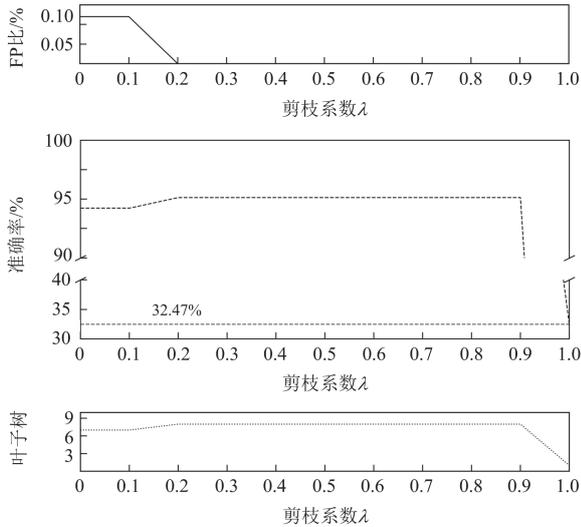


图 7 有效集 DT1 复杂度惩罚系数变化结果

Fig. 7 Result of the variation of the complexity penalty coefficient of the effective set DT1

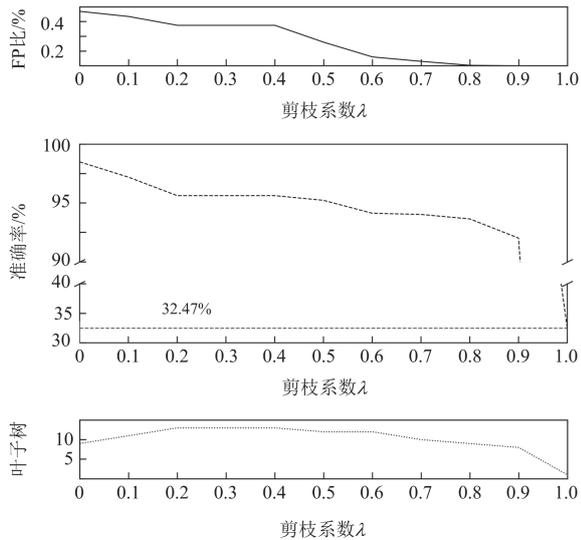


图 8 无效集 DT2 复杂度惩罚系数变化结果

Fig. 8 Result of the variation of the complexity penalty coefficient of the invalid set DT2

无效集中的典型 DT2 各项数据为: FP 比为 1.22%, 准确率为 98.55%, 叶子数为 20。

由图 7 可知, 对 DT1 而言, 当剪枝系数较小时, 此时剪枝更偏向于减小 FN, 相较于剪枝前, 剪枝可能会让 FP 有所升高。当剪枝系数不断变大、剪枝慢慢倾向于减少 FP 时, FP 重新实现归零, 之后一直维持 FP 为 0 直到剪枝系数增大到 1。在此过程中, 准确率的变化受 DT1 本身结构的影响, 有局部增大的可能, 比如图 7 中 DT1 准确率从剪枝前的 92.16% 提升至 95.13%, 并且能够维持 FP 归零, 对 DT1 进行了优化。最终, 当剪枝系数增大到 1 时, 准确率骤降, 这个准确率的值为实际非冗余约束总数在约束总数中的占比, 在该算例中为 32.47%。对叶子数而言, 随着 λ 的增加其值先增大后减小, 最后骤降至 1, 即只保留一个根节点, 这是剪枝策略过度倾向于减少 FP 所造成的结果。

由图 8 可知, 对 DT2 而言, 随着剪枝系数增加, FP 比逐渐减小直到实现 FP 归零, 准确率同样逐渐减小, 最后骤降到 32.47%。叶子数随着剪枝系数的增加先增大后减小。

需要说明的是, 各参数的变化趋势和数值都受特定 DT 结构的影响, 比如在无效集的 DT 中, 当剪枝系数为 0 时, 由于过度偏向减少 FN, 有些 DT 的叶子数会直接骤降为 1, 只有根节点。由图 7、图 8 可知, 改进的 REP 算法可以帮助 DT 减小复杂度, 保证 FP 归零, 同时增加一定的准确率。

最后, 在剪枝后的有效 DT 集中筛选出总体准确率最高者为最优 DT, 本算例的最优 DT 如图 9 所示。

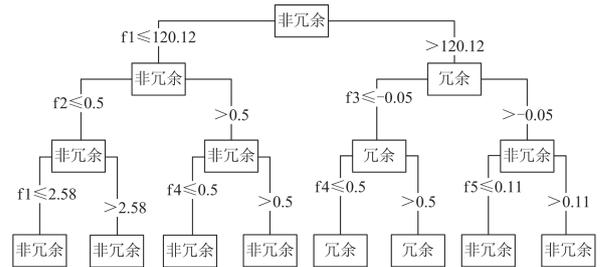


图 9 最优 DT 示意图

Fig. 9 Optimal DT diagram

图 9 中: f_1 表示历史平均松弛绝对值特征; f_2 表示新能源出力下限约束特征(类型约束特征); f_3 表示平均对偶变量特征; f_4 表示平均新能源高发时期属性特征; f_5 表示平均约束触发频率特征。其中, f_2 和 f_4 属于 0-1 整数特征, CART 算法以 0.5 为分界线区分特征取值。

最优 DT 的混淆矩阵如表 5 所示。

表 5 最优 DT 混淆矩阵

Table 5 Confusion matrix of optimal DT

	真实为正类	真实为负类
预测为正类	25 127	0
预测为负类	3356	40 421

通常采用准确率、召回率、精确率以及 F1 指标^[29]评估模型的性能。最优 DT 的各项指标如表 6 所示。

表 6 最优 DT 各项指标

Table 6 Optimal DT indicators

	%			
	准确率	精确率	召回率	F1 值
	95.13	100.0	88.22	93.74

由表 5、表 6 可知，最优 DT 能够实现无 FP 误判，即实现 100% 精准率，同时维持高准确率以确保约束削减后的效果。

通过最优 DT 识别预测后，对原 SCED 问题中的冗余约束进行削减，约束削减前后 SCED 问题各项数据对比如表 7 所示。

表 7 约束削减前后 SCED 问题对比

Table 7 SCED with and without constraints reduction

SCED 模型	原始 SCED	约束削减后的 SCED
目标函数值/元	3 085 554.52	3 085 554.52
起作用约束数量/条	22 373	22 373
总约束数量/条	68 904	24 639
约束削减时间/s	—	16.07
优化求解时间/s	17.57	1.50
总计算时间/s	17.57	2.07

SG-126 系统中，冗余约束占比 67.53%，成功识别 95.13% 的冗余约束，识别出的冗余约束数量占总约束数量的 64.24%。削减冗余约束后，优化求解时间降低 91.47%，总时间减少 88.22%，仿真结果表明改进算法识别速度快，识别准确率高。

本节实验验证了改进算法不仅可以处理普通二分类任务，还能够针对 SCED 模型中不同类别约束的特征，实现“约束削减-调度提速”的闭环流程。

4.3 对比实验

改进算法是针对冗余约束识别问题进行优化的，相比原始的 CART 和 REP 算法更具优势。图 10 直观展示了改进算法和原始算法所构建 DT 的各参数对比。最优 DT(改进算法)明显更适合极端 FP 敏感问题，没有安全约束被误删，保持更高的预测准确率，且复杂度更低，避免了过拟合。

原始 DT 是通过式(47)改变代价权重并结合经典 REP 构建的，各参数针对 FP 比和准确率进行寻

优，是原始算法所能构建的最佳 DT。由图 10 可知，相比于原始算法，改进算法可以实现无 FP 误判且保持更高的总体准确率，同时拥有更小的复杂度以提升泛用性。

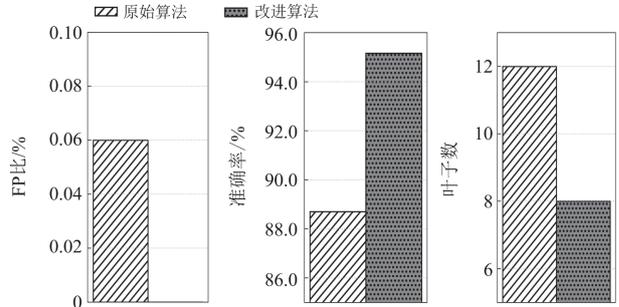


图 10 最优 DT 与原始 DT 对比图

Fig. 10 Comparison chart of optimal DT and original DT

此外，考虑到约束特征是表格数据且无明显时序关系，本文在对比决策树算法改进前后性能的基础上，进一步引入多层感知机(multilayer perceptron, MLP)神经网络作为参照算法开展对比实验。所采用的 MLP 神经网络架构如图 11 所示，具体包含如下技术特征：采用 4 层全连接架构，依次为输入层(12 个神经元，对应输入特征维度)、隐藏层 1(64 个神经元)、隐藏层 2(32 个神经元)以及输出层(1 个神经元，适配单输出任务)，各层之间通过全连接方式实现神经元的信号传递。

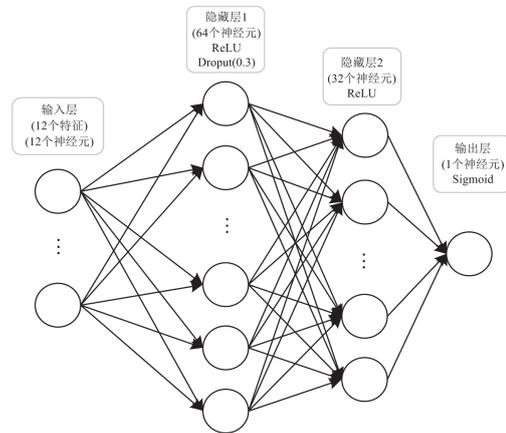


图 11 MLP 神经网络结构图

Fig. 11 Structure diagram of MLP neural network

利用该 MLP 神经网络进行冗余约束识别，该算法特征工程的构建与 DT 相同。然后，将训练集引入 MLP 进行训练即可，总共训练 100 轮，训练过程中的训练损失、准确率和 FP 比变化曲线如图 12 所示。由图 12 可知，训练过程中随着训练轮次增加，训练损失首先骤降然后趋于平稳，整体呈现下降趋

势, 到训练末期只存在少量波动; 准确率首先迅速提升然后趋于平稳, 整体呈现上升趋势, 到训练末期只存在少量波动, 训练集准确率最高达到 97.06%; FP 比并无下降趋势, 在 2.1%~1.6%之间波动。

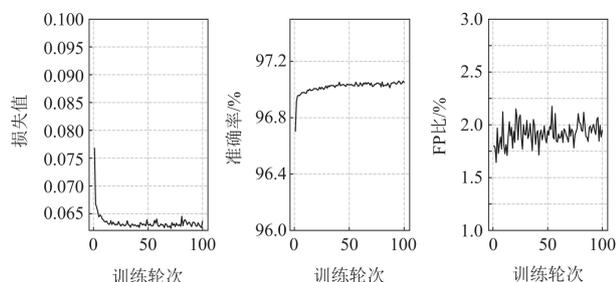


图 12 MLP 的训练集各指标变化图

Fig. 12 Variation graphs of each index in the training set of MLP

最优 DT(改进算法)和 MLP 在测试集中的预测结果对比如图 13 所示。

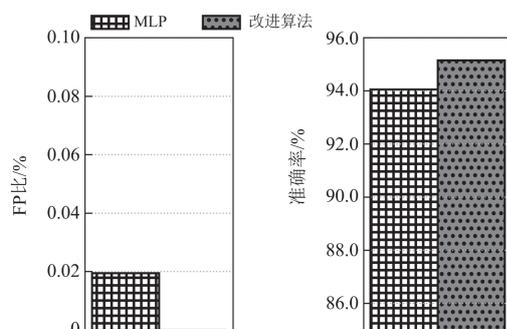


图 13 最优 DT 与 MLP 对比图

Fig. 13 Comparison chart of optimal DT and MLP

由图 13 可知, MLP 经过大量训练后, 对于测试集的预测准确率略低于改进算法的预测准确率。同时, MLP 仍然存在部分安全约束误删(FP 不等于 0)。经过对比实验可得, 改进的决策树算法在冗余约束识别问题中具备明显优势。

5 结论

针对 SCED 冗余约束识别缺乏快速识别方法, 以及数据驱动方法易发生 FP 误判进而影响系统安全的问题, 提出了基于改进 CART 与 REP 算法的冗余约束快速识别方法, 主要工作包括:

1) 给出了针对 SCED 冗余约束识别的 DT 构建所需的特征构建和数据处理方法。

2) 针对极端 FP 敏感场景改进了算法, 在 CART 中引入 FP 惩罚机制, 使其构建的 DT 更容易实现 FP 归零。

3) 基于 FP 比提出了改进的 REP 算法, 使 DT

剪枝后能够在实现零 FP 误判的同时降低模型复杂度, 以防过拟合。

4) 以 SG-126 系统为例进行实验验证, 结果表明改进算法能够快速识别冗余约束: 在实现零 FP 误判的前提下, 识别准确率达到 95.13%; 冗余约束削减后, 系统调度时间减少 88.22%。同时, 通过对比实验进一步证明了改进决策树在冗余约束识别方面的优越性。

综上, 本文提出的改进算法通过在 CART 算法和 REP 算法中引入 FP 惩罚机制以实现 FP 归零, 同时在冗余约束识别问题上保持高预测准确率。与其他方法相比, 该方法识别速度更快、识别约束范围更广, 且能保证削减约束后系统安全运行。

但本文工作仍存在不足, 例如未显式考虑新能源和负荷的不确定性。未来研究可进一步扩展至考虑源荷不确定性的鲁棒 SCED 或随机 SCED 模型的冗余约束识别, 其中针对不确定性的特征工程构建与数据处理等方面仍有待深入探索。

参考文献

- [1] 姜海洋, 杜尔顺, 金晨, 等. 高比例清洁能源并网的跨国互联电力系统多时间尺度储能容量优化规划[J]. 中国电机工程学报, 2021, 41(6): 2101-2115.
JIANG Haiyang, DU Ershun, JIN Chen, et al. Optimal planning of multi-time scale energy storage capacity of cross-national interconnected power system with high proportion of clean energy[J]. Proceedings of the CSEE, 2021, 41(6): 2101-2115.
- [2] 蒋骏骏, 向月, 谈竹奎, 等. 计及需求响应的高比例清洁能源园区储能容量优化配置[J]. 中国电力, 2023, 56(12): 147-155, 163.
JIANG Zhaojun, XIANG Yue, TAN Zhukui, et al. Optimal allocation of energy storage capacity in high proportion clean energy parks considering demand response[J]. Electric Power, 2023, 56(12): 147-155, 163.
- [3] 张郁, 苑波, 黄石成, 等. 考虑高比例可再生能源接入的有源配电网经济调度策略研究[J]. 电测与仪表, 2025, 62(1): 158-166.
ZHANG Yu, YUAN Bo, HUANG Shicheng, et al. Research on economic dispatching strategy for active distribution network considering high penetration of renewable energy source[J]. Electrical Measurement & Instrumentation, 2025, 62(1): 158-166.
- [4] 陈文淑乐, 向月, 彭光博, 等. “双碳”目标下电力系统供给侧形态发展系统动力学建模与分析[J]. 上海交通大学学报, 2021, 55(12): 1567-1576.
CHEN Wenxule, XIANG Yue, PENG Guangbo, et al.

- System dynamic modeling and analysis of power system supply side morphological development with dual carbon targets[J]. Journal of Shanghai Jiao Tong University, 2021, 55(12): 1567-1576.
- [5] 张智刚, 康重庆. 碳中和目标下构建新型电力系统的挑战与展望[J]. 中国电机工程学报, 2022, 42(8): 2806-2819. ZHANG Zhigang, KANG Chongqing. Challenges and prospects for constructing the new-type power system towards a carbon neutrality future[J]. Proceedings of the CSEE, 2022, 42(8): 2806-2819.
- [6] 卢有飞, 邹时容, 刘璐豪, 等. 基于知识共享的高比例可再生能源系统发电控制[J]. 高压电器, 2024, 60(10): 33-45. LU Youfei, ZOU Shirong, LIU Luhao, et al. Generation control of high percentage renewable energy systems based on knowledge sharing[J]. High Voltage Apparatus, 2024, 60(10): 33-45.
- [7] 李客村, 黄振宇, 游星星, 等. 双碳目标下基于双时间尺度动力学模型的灵活性机组容量演化[J]. 电力建设, 2025, 46(2): 137-150. LI Kecun, HUANG Zhenyu, YOU Xingxing, et al. Capacity evolution of flexible units based on a dual-timescale dynamic model under dual carbon goals[J]. Electric Power Construction, 2025, 46(2): 137-150.
- [8] 易海琼, 赵朗, 李一铮, 等. 市场条件下新能源大基地储能规划双层经济性分析方法[J]. 电力建设, 2025, 46(3): 95-103. YI Haiqiong, ZHAO Lang, LI Yizheng, et al. Economic method for energy storage planning of large new energy bases in the electricity spot market[J]. Electric Power Construction, 2025, 46(3): 95-103.
- [9] 高怡杰, 蔡德福, 俞耀文, 等. 面向电力系统安全约束经济调度的数字孪生方法[J]. 电力自动化设备, 2025, 45(7): 197-203. GAO Yijie, CAI Defu, YU Yaowen, et al. Digital twin method for security-constrained economic dispatch of power system[J]. Electric Power Automation Equipment, 2025, 45(7): 197-203.
- [10] 郭瑞鹏, 边麟龙, 宋少群, 等. 安全约束最优潮流的实用模型及故障态约束缩减方法[J]. 电力系统自动化, 2018, 42(13): 161-168. GUO Ruipeng, BIAN Linlong, SONG Shaoqun, et al. Practical model of security constrained optimal power flow and reduction methods for contingency constraints[J]. Automation of Electric Power Systems, 2018, 42(13): 161-168.
- [11] ZHAI Q, GUAN X, CHENG J, et al. Fast identification of inactive security constraints in SCUC problems[J]. IEEE Transactions on Power Systems, 2010, 25(4): 1946-1954.
- [12] ARDAKANI J, BOUFFARD F. Identification of umbrella constraints in DC-based security constrained optimal power flow[J]. IEEE Transactions on Power Systems, 2013, 28(4): 3924-3934.
- [13] ARDAKANI J, BOUFFARD F. Acceleration of umbrella constraint discovery in generation scheduling problems[J]. IEEE Transactions on Power Systems, 2015, 30(4): 2100-2109.
- [14] 付聪, 王砚平, 刘俊磊, 等. 基于辅助优化问题的安全约束机组组合约束削减方法[J]. 电力系统保护与控制, 2021, 49(21): 9-17. FU Cong, WANG Yanping, LIU Junlei, et al. Constraint reduction method for security-constrained unit commitment based on an auxiliary optimization problem[J]. Power System Protection and Control, 2021, 49(21): 9-17.
- [15] 王砚平. 安全约束机组组合问题的冗余约束削减方法研究[D]. 杭州: 浙江大学, 2021. WANG Yanping. Redundant constraints reduction methods for security-constrained unit commitment[D]. Hangzhou: Zhejiang University, 2021.
- [16] CHEN Y, CASTO A, WANG F, et al. Improving large scale day ahead security constrained unit commitment performance[J]. IEEE Transactions on Power Systems, 2016, 31(6): 4732-4743.
- [17] XAVIER Á S, QIU F, WANG F, et al. Transmission constraint filtering in large-scale security-constrained unit commitment[J]. IEEE Transactions on Power Systems, 2019, 34(3): 2457-2460.
- [18] 王砚平, 鲍威, 李赢, 等. 考虑 $N-1$ 故障的安全约束机组组合模型及约束削减方法[J]. 电力自动化设备, 2021, 41(7): 167-175. WANG Yanping, BAO Wei, LI Ying, et al. Model and constraint-reduction method for security-constrained unit commitment considering $N-1$ contingency[J]. Electric Power Automation Equipment, 2021, 41(7): 167-175.
- [19] 徐渊, 谌艳红, 王凯, 等. 基于有效约束识别的电网热稳边界辨识方法[J]. 电力系统保护与控制, 2021, 49(11): 54-60. XU Yuan, CHEN Yanhong, WANG Kai, et al. Identification method of a power grid thermal stability boundary based on effective constraint identification[J]. Power System Protection and Control, 2021, 49(11): 54-60.
- [20] 许彦平, 赵明欣, 秦晓辉, 等. 基于边界法的大型机组组合模型的简化方法[J]. 系统仿真学报, 2024, 36(2): 296-304. XU Yanping, ZHAO Mingxin, QIN Xiaohui, et al. A simplification method of large-scale unit commitment

- model based on boundary method[J]. *Journal of System Simulation*, 2024, 36(2): 296-304.
- [21] 王麒宁, 陈思远, 徐箭, 等. 基于冗余约束快速辨识的电力系统前瞻调度鲁棒决策方法[J]. *电力系统自动化*, 2025, 49(22): 123-134.
WANG Qining, CHEN Siyuan, XU Jian, et al. Robust decision-making method for look-ahead dispatch in power system based on fast identification of redundant constraints[J]. *Automation of Electric Power Systems*, 2025, 49(22): 123-134.
- [22] 梅玉杰, 李勇, 周王峰, 等. 基于机器学习的配电网异常缺失数据动态清洗方法[J]. *电力系统保护与控制*, 2023, 51(7): 158-169.
MEI Yujie, LI Yong, ZHOU Wangfeng, et al. Dynamic data cleaning method of abnormal and missing data in a distribution network based on machine learning[J]. *Power System Protection and Control*, 2023, 51(7): 158-169.
- [23] 王桐, 史雯, 石浩渊, 等. 基于平稳小波变换的光伏直流串联电弧故障检测[J]. *电力系统保护与控制*, 2024, 52(12): 82-93.
WANG Tong, SHI Wen, SHI Haoyuan, et al. Photovoltaic DC series arc fault detection based on stationary wavelet transform[J]. *Power System Protection and Control*, 2024, 52(12): 82-93.
- [24] 黄南天, 彭华, 蔡国伟, 等. 电能质量复合扰动特征选择与最优决策树构建[J]. *中国电机工程学报*, 2017, 37(3): 776-786.
HUANG Nantian, PENG Hua, CAI Guowei, et al. Feature selection and optimal decision tree construction of complex power quality disturbances[J]. *Proceedings of the CSEE*, 2017, 37(3): 776-786.
- [25] 江永鑫, 陈丽安, 郭梦倩, 等. 基于改进CEEMD和RF的低压串联故障电弧识别方法[J]. *电力系统保护与控制*, 2024, 52(1): 97-108.
JIANG Yongxin, CHEN Li'an, GUO Mengqian, et al. Identification method of low voltage series fault arc based on improved CEEMD decomposition and RF[J]. *Power System Protection and Control*, 2024, 52(1): 97-108.
- [26] 李晖, 汪莹, 罗天, 等. 基于电气介数的电网故障无功规划选址与定容[J]. *控制与决策*, 2019, 34(8): 1609-1615.
LI Hui, WANG Ying, LUO Tian, et al. Fault reactive power location and capacity planning using electric betweenness[J]. *Control and Decision*, 2019, 34(8): 1609-1615.
- [27] 李焱, 贾雅君, 李磊, 等. 基于随机森林算法的短期电力负荷预测[J]. *电力系统保护与控制*, 2020, 48(21): 117-124.
LI Yan, JIA Yajun, LI Lei, et al. Short term power load forecasting based on a stochastic forest algorithm[J]. *Power System Protection and Control*, 2020, 48(21): 117-124.
- [28] 魏红宁. 决策树剪枝方法的比较[J]. *成都: 西南交通大学学报*, 2005, 40(1): 44-48.
WEI Hongning. Comparison among methods of decision tree pruning[J]. *Journal of Southwest Jiaotong University*, 2005, 40(1): 44-48.
- [29] 王娜娜, 栗文义, 李小龙. 基于不均衡小样本 DGA 数据与改进 CatBoost 决策树的油浸式变压器故障诊断方法[J]. *电力系统保护与控制*, 2024, 52(23): 167-176.
WANG Nana, LI Wenyi, LI Xiaolong. An oil-immersed transformer fault diagnosis method based on DGA unbalanced limited sample processing and improved CatBoost[J]. *Power System Protection and Control*, 2024, 52(23): 167-176.

收稿日期: 2025-05-19; 修回日期: 2025-09-10

作者简介:

夏卓延(2002—), 男, 硕士研究生, 研究方向为电力系统调度与优化; E-mail: 24126323@bjtu.edu.cn

张大海(1973—), 男, 通信作者, 博士, 教授, 研究方向为智能电网、继电保护。E-mail: dhzhang1@bjtu.edu.cn

(编辑 魏小丽)