

## 燃气蒸汽联合循环机组运行优化调控系统

万安平<sup>1</sup>, 杨洁<sup>1,2</sup>, 王景霖<sup>3</sup>, 王文晖<sup>1,4</sup>, 缪徐<sup>1</sup>,  
陈挺<sup>1</sup>, 纪杨建<sup>2</sup>, 李客<sup>2,5</sup>(1.浙大城市学院机电系, 浙江 杭州 310015; 2.浙江大学机械工程学院, 浙江 杭州 310027;  
3.故障诊断与健康安全技术航空科技重点实验室, 上海 201601;  
4.安徽理工大学机械工程学院, 安徽 淮南 232001;  
5.中信重工机械股份有限公司, 河南 洛阳 471039)

**[摘要]** 针对电厂机组运行管理过程中人工调控运行参数耗时费力、效率低、准确率低等问题, 设计了一种基于数据挖掘的电厂机组运行优化调控系统。首先, 采用数据挖掘算法对机组历史和实时运行数据进行参数的整合和相关性分析, 得到影响机组运行的关键参数, 作为机组健康状态评估的指标。然后, 利用长短时记忆 (LSTM) 神经网络模型对机组健康状态特征模块中确定的特征值进行训练, 预测参数随时间的变化趋势, 实现机组智能调控。最后, 开发了一套燃气蒸汽联合循环发电机组运行调控系统, 并应用于浙江某电厂。运行结果表明: 该系统可指导机组运行优化, 提高机组的运行可靠性及经济性; 优化后电厂机组输出功率提升了0.412 5%, 年电能产量增加4 806 MW·h, 机组年收益增加约326万元。

**[关键词]** 数据挖掘; 燃气蒸汽联合循环机组; 机组健康状态; LSTM神经网络; 运行优化

**[中图分类号]** TK39 **[文献标识码]** A **[DOI编号]** 10.19666/j.rld.202104092

**[引用本文格式]** 万安平, 杨洁, 王景霖, 等. 燃气蒸汽联合循环机组运行优化调控系统[J]. 热力发电, 2021, 50(12): 21-29.  
WAN Anping, YANG Jie, WANG Jinglin, et al. Operation optimization control system of gas-steam combined cycle generator unit[J]. Thermal Power Generation, 2021, 50(12): 21-29.

## Operation optimization control system of gas-steam combined cycle generator unit

WAN Anping<sup>1</sup>, YANG Jie<sup>1,2</sup>, WANG Jinglin<sup>3</sup>, WANG Wenhui<sup>1,4</sup>, MIU Xu<sup>1</sup>,  
CHEN Ting<sup>1</sup>, JI Yangjian<sup>2</sup>, LI Ke<sup>2,5</sup>(1. Department of Mechanical Engineering, Zhejiang University City College, Hangzhou 310015, China;  
2. School of Mechanical Engineering, Zhejiang University, Hangzhou 310027, China;  
3. Aviation Key Laboratory of Science and Technology on Fault Diagnosis Health Management, Shanghai 201601, China;  
4. College of Mechanical Engineering, Anhui University of Science and Technology, Huainan 232001, China;  
5. CITIC Heavy Industries Co., Ltd., Luoyang 471039, China)

**Abstract:** In view of the time-consuming, low efficiency and low accuracy of manual acquisition and adjustment of operating parameters in the process of power plant units operation management, this paper designs an optimization and control system of power plant units operation based on data mining. Firstly, the data mining algorithm is used to integrate and analyze the parameters of the historical and real-time operation data of the units, and the key parameters affecting the units operation are obtained, which are used as the indicators of the unit health status evaluation. Then, the long-term and short-term memory (LSTM) neural network model is employed to train the eigenvalues determined in the health state feature acquisition module, and the change trend of parameters with time is predicted, so as to realize the intelligent control of units. Finally, a set of gas-steam combined cycle generator operation control system is developed and applied to a power plant in Zhejiang Province. The operation results show that, this system can guide the optimal operation of the unit and improve the operation reliability and economy of the unit. After optimization, the output power of the unit increased by 0.412 5%, the annual power output increased by 4 806 MW·h, and the annual revenue of the unit increased by 3.26 million yuan.

收稿日期: 2021-04-23

基金项目: 国家自然科学基金项目(51705455); 航空科学基金(20183333001); 浙大城市学院科研培育基金资助课题(J-202113)

Supported by: National Natural Science Foundation of China (51705455); Aviation Science Foundation(20183333001); Scientific Research Foundation of Zhejiang University City College(J-202113)

第一作者简介: 万安平(1983), 男, 博士, 副教授, 主要研究方向为复杂装备健康管理及寿命预测, wanap@zucc.edu.cn.

通信作者简介: 陈挺(1989), 男, 博士后, 主要研究方向为联合循环机组余热回收利用, chenting@zucc.edu.cn.

**Key words:** data mining; gas-steam combined cycle unit; unit health status; LSTM neural network; operation optimization

近年来,传统的燃煤发电厂由于其结构体系和发电过程容易产生环境污染,而燃气轮机联合循环机组相较于前者具有更高的发电效率、环保性以及更为灵活的电网负荷调节能力,对我国能源互联网的发展产生了深远的影响。

伴随人工智能的迅速发展,数据挖掘作为大数据领域一项重要方法,其在电力系统的故障诊断、参数优化等方面的应用愈发广泛。随着国内电厂信息化建设的不断完善,越来越多的电力生产数据被存储到电厂信息化系统实时数据库中,电力生产企业因此积累了海量的历史数据。从这些海量历史数据中找到改善机组运行的信息,并开发和利用数据背后的潜在价值,具有重要的意义。

数据挖掘(data mining)又称为数据库文件的专业知识发觉(knowledge discovery in database, KDD),即从海量存储数据中通过算法搜索获得合理、新奇、潜在有效和可解释的信息的过程。数据挖掘的发展为提高机组能效打开了新的大门<sup>[1]</sup>,在国内外大型电厂的状态参数监测及系统优化方面得到广泛应用。采用数据挖掘可以提取电厂机组运行过程中关键参数的潜在特征,以反映机组运行状态,并结合相关系统做实时合理调整,例如 Tony Ogilvie 等人<sup>[2]</sup>将关联规则挖掘算法应用于火电厂分散控制系统(DCS),并建立专家系统对电厂运行过程进行监测,为技术人员提供合理的检修策略。Kusiak 等人<sup>[3]</sup>通过数据挖掘技术实现了对火电厂化学水处理系统的运行监测及故障检测和诊断,提出了一种简单、鲁棒性高的报警系统,适用于历史故障数据稀疏的设备的实时监测。Burns 等人<sup>[4]</sup>通过机器学习算法从数据中获取知识以确定使锅炉燃烧效率提高的工艺参数范围,提出了适用于不同类型燃烧器和燃料类型的数据挖掘方法。何怀昌等<sup>[5]</sup>利用燃气-蒸汽联合循环机组历史和实时运行数据,采用大数据挖掘技术进行参数整合和相关性分析,建立诊断系统,提高机组运行效能和经济性。王秋平等<sup>[6]</sup>通过对比近年来电厂优化常用的关联规则挖掘算法,并从数据预处理、数据离散化、工况划分、属性约简等 4 个方面进行了算法优化。郭华新等<sup>[7]</sup>利用基于 Clementine 数据挖掘技术分析电厂 DCS 采集的大量生产数据,计算煤质特性参数与煤处理设备、磨煤机、风机静电除尘器、烟气脱硫装置等

辅机设备能耗特性关系,提供了机组在不同负荷及不同煤质特性条件下的最优运行方式与参数控制方法。以上研究虽然将大数据挖掘应用于电厂机组关键参数的分析优化中,在一定程度上解决了机组运行过程中参数的调优问题,但面对复杂的燃气轮机联合循环机组海量的运行数据,其提取的关联度较低,特征也较多,难以有效提取具有强关联性的特征参数。

目前,用于电厂机组运行优化目标值求解的数据挖掘算法主要为遗传算法<sup>[8-9]</sup>、模糊神经网络<sup>[10-11]</sup>和关联规则算法<sup>[12-13]</sup>。本文在对比和总结 3 种算法优缺点的基础上,对基于关联规则的数据挖掘算法进行改进和优化,加入兴趣度衡量标准,结合 Hadoop 中 MapReduce 编程框架,实现对电厂数据的分布并行化处理,使其在电厂机组优化中更加有效。同时基于该算法,提出一种电厂机组运行优化及调控系统,该系统能够依次完成数据提取、数据挖掘、关键参数选择、优化以及调控的功能。并将改进的关联规则数据挖掘算法应用于实际电厂机组数据中,实现了强关联参数的优化。

## 1 机组运行调控模型

### 1.1 特征参数筛选模型

采用可视化工具以及经典的相关性分析方法对联合循环功率与参数间的关联关系进行相关性分析,筛选出与联合循环功率呈明显正相关的运行参数,并计算对应的皮尔森相关系数,通过稳定性判断、极值标准化处理、设置隶属度、划分量化区间、调整最小支持度和最小置信度的值,挖掘出符合要求的关键特征参数。图 1 为优化参数筛选流程。

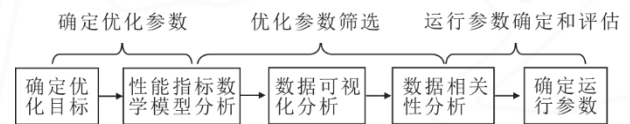


图 1 优化参数筛选流程

Fig.1 Screening process of the optimization parameters

参数选取的过程可分为基于数学模型的性能指标分析、数据可视化分析、相关性分析和运行参数确定 4 个步骤。最终筛选的运行参数为:空气质量流量、天然气摩尔体积、燃烧室进口烟气温度、余热锅炉入口烟气温度、余热锅炉入口烟气压力、

余热锅炉出口烟气温、余热锅炉出口烟气压力、高压过热器出口蒸汽流量、余热锅炉低压蒸汽流量、低压过热器出口蒸汽温度、再热器出口蒸汽温度、中压过热器出口蒸汽温度、高压过热器出口蒸汽温度、透平压比、高压汽包压力、低压省煤器出口水温度、压气机进口空气温度、压气机进口空气湿度、压气机出口空气温度、压气机出口空气压力、透平出口烟气压力、透平出口烟气温度、燃气透平进口湿燃气比焓、燃气透平做功后湿燃气比焓、燃气透平出口冷却比焓、低压过热器出口蒸汽压力、再热器出口蒸汽压力、中压过热器出口蒸汽压力、高压过热器出口蒸汽压力、低压省煤器出口水压力、高压省煤器出口水压力、低压缸排气压力。

### 1.2 机组健康状态聚类分析模型

结合实际生产经验和工况库中数据分布, 确定会导致运行异常临界值的稳定判断指标, 在多个临界值的限定范围内对预处理后的数据进行进一步筛选, 求满足所有限制条件的数据, 得到的筛选结果作为聚类的输入数据。聚类分析采用的是  $k$ -means<sup>[14]</sup>来发现数据集中的  $k$  个工况簇。 $k$  是由用户指定的, 该算法的目标是寻找数据集中和  $k$  个簇的质心, 将数据中各点分别分配给距离最近的质心, 并将该点归入对应质心的类别。图 2 为机组健康状态  $k$ -means 聚类分析流程。

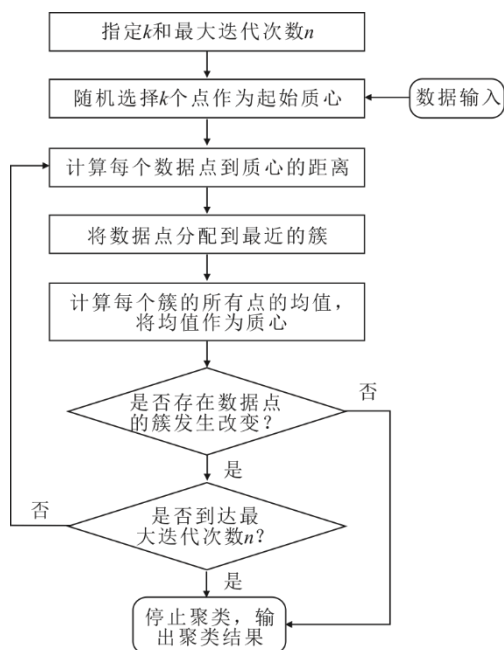


图 2 机组健康状态  $k$ -means 聚类分析流程  
Fig.2 The  $k$ -means clustering analysis process for unit health status

由图 2 可见, 将关系数据库中经过数据挖掘提取后的待聚类的数据进行  $k$ -means 聚类分析时, 需要设定簇的个数  $k$  和最大迭代次数  $n$ , 然后随机选取  $k$  个数据点作为起始质心, 通过计算每个数据点到质心的距离, 将数据点分配给距离值最小的簇, 并通过均值不断重复更新每个簇的质心, 直至数据点的簇不发生改变或达到最大迭代次数  $n$  时, 结束聚类, 输出结果。实际应用过程中, 技术人员可根据电厂机组确定的优化特征参数, 设定若干类别, 定义每个类别的特征, 将经过特征挖掘后的数据进行分类, 并标定稳定状态与非稳定状态。

### 1.3 机组稳定工况库建立模型

按照对聚类分析中数据状态的定义, 完成对机组运行工况记录的类别标注, 将稳态与非稳态类别标签分别设置为 0 和 1, 从中提取稳定工况, 建立稳定模式工况库。图 3 为机组稳定工况模式库建立过程。

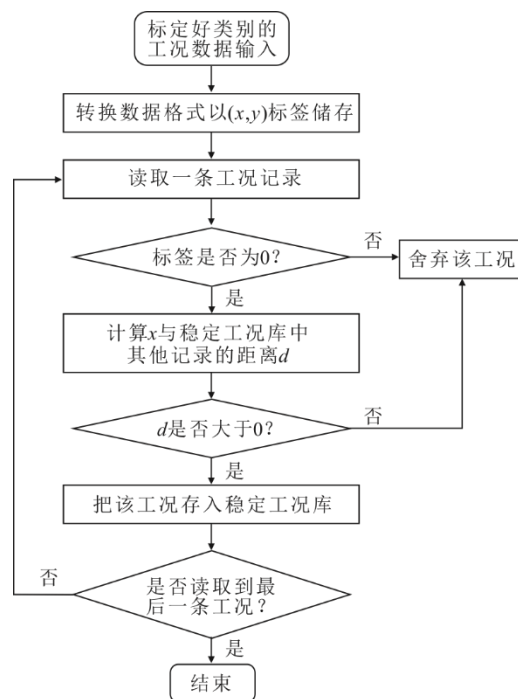


图 3 机组稳定工况模式库建立过程  
Fig.3 The establishment process of mode database for unit stable condition

由图 3 可见, 一条工况记录中包含可控变量  $x$ 、稳定表征变量  $y$  以及类别标签。对每条工况计算  $x$  与工况库中已有工况之间的距离, 若距离为 0, 则认为工况库中已存在该工况, 不再重复记录; 否则, 将该工况加上时间标签, 以向量形式存入稳定工况库。



### 1.4 机组实时特征参数预测模型

利用长短时记忆 (LSTM) 神经网络模型<sup>[15]</sup>对电厂机组健康状态特征模块中确定的特征值进行训练, 预测参数随时间的变化趋势, 以辅助状态判断。LSTM 神经网络结构单元如图 4 所示。

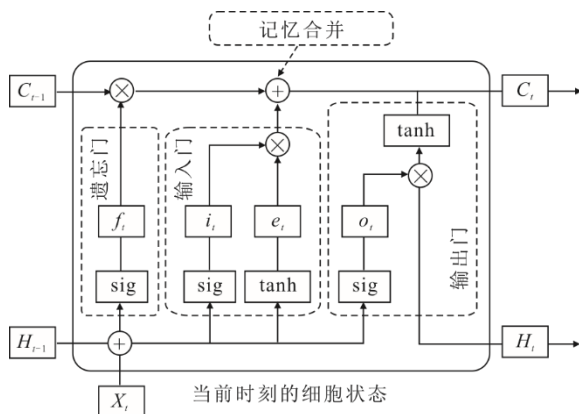


图 4 LSTM 神经网络结构单元

Fig.4 Structural unit of the LSTM neural network

由图 4 可见, 将经由数据挖掘后的待优化参数的时间序列输入 LSTM 神经网络模型, 经过输入门、遗忘门、输出门的选择性保留, 最终预测下一个时间点该运行参数可能存在的状态。LSTM 神经网络的原理即为将时间序列按时间顺序依次输入, 定义  $X_t$  为当前时刻的输入数据, 其输出结果  $H_t$  是上一时刻输入  $X_{t-1}$  经过遗忘门选择性遗忘后的结果与  $X_t$  共同作用的结果。同时, 该时刻输入  $X_t$  通过遗忘门过滤也将作为下一时刻输入的一部分, 即上一时刻数据特征会影响下一时刻输出, 从而提取随时间的变化某一时刻输出的规律性, 进而预测未来某一时刻的结果。

### 1.5 机组运行智能调控模型

在电厂机组运行指标中的参数出现异常时, 启动调控程序, 程序会从稳定模式库中搜索调控目标, 返回距离当前状态最近的点作为待选工况。之后, 比较当前状态与待选工况的差异, 统计当前状态调至待选目标时需要调控的参数、需要调控的幅度以及调控参数个数, 从这 3 个维度在待选工况中确定 1 个调控目标。调控目标的选取原则是调控个数尽可能少, 调节幅度尽可能小。最后, 确定调整目标后, 按照设定的调节幅度进行参数的调整, 对可控变量进行调整, 直到参数达到目标值。在调控过程中会监测稳定指标的变化走势, 如果指标没有回归正常, 可以随时切断调控进程, 进入人工调控环节。

图 5 为电厂机组运行智能调控模块流程。当模块中稳定判断指标的取值偏离正常范围, 会自动触发稳定状态自动调控。

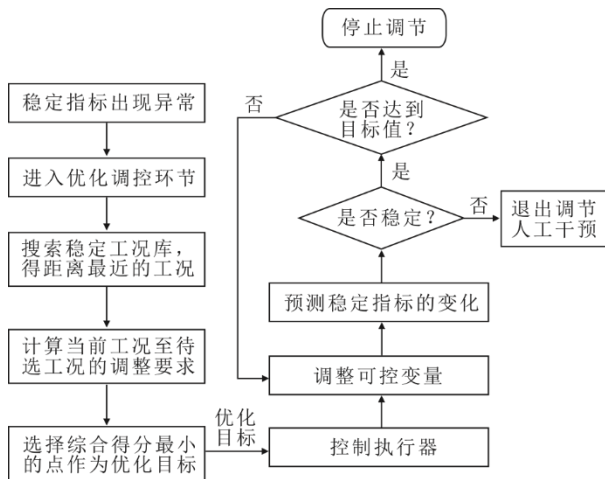


图 5 机组运行智能调控模块流程

Fig.5 Process of intelligent control module for the unit operation

## 2 系统组成及分析

本文基于机组运行调控模型, 设计并开发了一套燃气蒸汽联合循环发电机组运行调控系统, 其总体结构框架如图 6 所示。由图 6 可见, 系统主要包括数据预处理模块、机组健康状态评估指标挖掘模块、机组运行关键特征参数筛选模块、机组健康状态聚类分析模块、机组稳定工况库建立模块、机组状态评估指标特征获取模块、机组实时特征参数预测模块和机组运行智能调控模块等 8 个功能模块。

1) 数据预处理模块 对采集的机组数据进行异常值处理、缺失值处理、离散化处理和归一化处理, 为后续的数据挖掘与分析做好准备。数据异常值处理、空值处理通过基于 Python 数据库的数据筛选清洗功能实现。离散化处理和归一化处理由特征简约和数据变换实现。

2) 机组健康状态评估指标挖掘模块 利用改进的关联规则挖掘算法对运行参数数据进行挖掘分析, 得到影响机组运行的关键参数, 作为机组健康状态评估的指标。

3) 机组运行关键特征参数筛选模块 采用可视化工具以及经典的相关性分析方法对联合循环功率与参数间的关联关系进行相关性分析, 筛选出与联合循环功率呈明显正相关的运行参数。

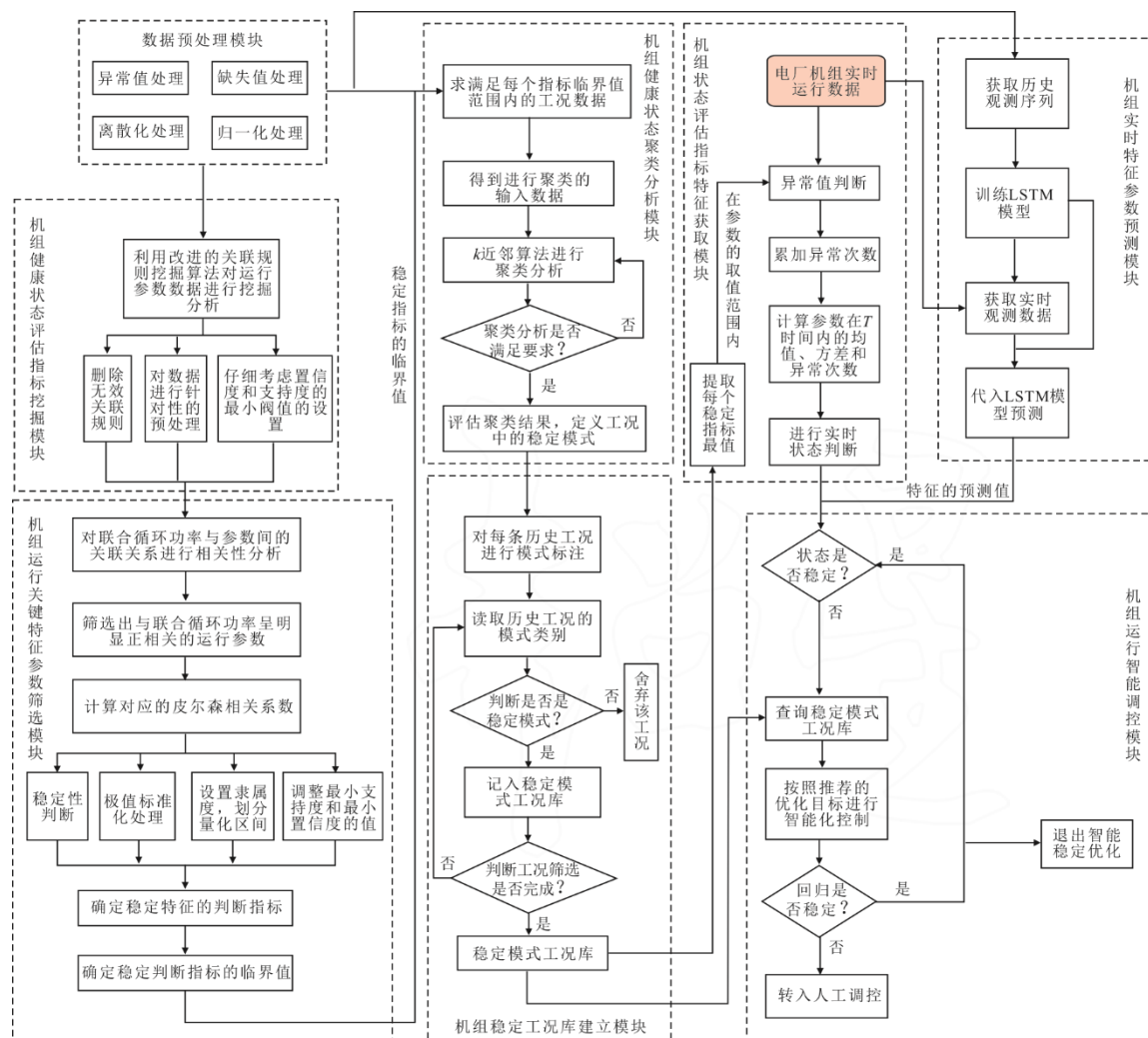


图 6 燃气蒸汽联合循环发电机组运行调控系统总体结构

Fig.6 Overall structure of operation control system of the gas-steam combined cycle unit

4) 机组健康状态聚类分析模块 基于确定的机组健康状态评估指标,结合实际生产经验和工况库中的数据分布,确定会导致运行异常的临界值的稳定判断指标,在多个临界值的限定范围内对预处理后的数据进行进一步筛选,求满足所有限制条件的数据,得到的筛选结果作为聚类的输入数据。

5) 机组稳定工况库建立模块 按照对聚类分析中数据状态的定义,完成对已有机组运行工况记录的类别标注,将稳态与非稳态类别标签分别设置为 0 和 1,从中提取稳定工况,建立稳定模式工况库。

6) 机组状态评估指标特征获取模块 通过分析机组运行状态下采集的实时数据的特点,确定进行实时状态判断的特征参数及其获取方法。以上述与联合循环功率呈明显正相关的运行状态评估参数的实时数据为基础,计算每个参数的均值、方差和异常值出现次数,把得到的结果作为稳态判断的

特征变量。

7) 机组实时特征参数预测模块 利用 LSTM 神经网络模型对机组健康状态特征获取模块中确定的特征值进行训练,预测参数随时间的变化趋势,以辅助状态判断。

8) 机组运行智能调控模块 在机组运行指标中的参数出现异常时,启动调控程序,程序会从稳定模式库中搜索调控目标,返回距离当前状态最近的点作为待选工况,实现参数智能调控。

### 3 应用实例及分析

本系统已在浙江杭州某联合循环发电厂投入运行,系统运行界面如图 7 所示。本文基于 Hadoop 平台的改进关联规则数据挖掘算法,选取浙江杭州某联合循环发电机组 1 年的电力生产历史运行参数进行分析处理。



图 7 系统运行界面  
Fig.7 The system operation interface

### 3.1 数据获取

依据联合循环功率的数学模型,在可视化、相关性分析的基础上,从电厂信息系统(SIS)中获取与数学模型紧密相关的可调整运行参数变量以及对应的性能指标数据,共计 800 万组数据,作为采用改进关联规则数据挖掘的机组运行参数优化的数据分析对象。部分参数见表 1。

表 1 部分机组运行参数数据  
Tab.1 Operation parameter data of some units

空气质 量流量/ (t·h <sup>-1</sup> )	燃气质 量流量/ (t·h <sup>-1</sup> )	压气机 进口温 度/℃	压气机进 口压力/kPa	压气机 出口温 度/℃	压气机出 口压力/ kPa
1 484.93	9.76	20.40	102.37	305.39	688.89
54.42	6.90	10.38	101.99	48.98	262.15
1 541.63	10.46	10.01	102.72	291.11	659.27
1 477.57	10.78	21.14	101.75	304.22	660.45
52.65	6.89	10.16	101.99	48.59	261.90
1 482.69	8.67	22.16	102.54	305.53	649.95
1 530.64	11.63	14.18	102.68	289.77	759.23
49.44	6.87	10.27	101.99	48.20	261.66
1 505.48	9.75	17.83	102.91	299.75	679.82
48.21	6.85	10.21	101.99	47.92	261.42

### 3.2 数据预处理

#### 1) 稳态数据筛选

机组处于非稳定状态下运行时,数据挖掘所确定的优化调控区间难以准确反映机组真实的运行状态,由此,本文选择机组在稳定运行状态下的数据进行分析。

根据式(1)进行稳定性判断:

$$\sum_{i=t-d}^t (P_{\text{main},i} - P_{\text{main}})^2 < \xi \quad (1)$$

式中:  $P_{\text{mian}}$  为运行参数  $P_{\text{mian},i}$  在该时间段内的均值。即式(1)表示运行参数  $P_{\text{mian},i}$  在时间段  $[t-d, t]$  内的方差小于阈值  $\xi$ 。

由式(1)得到稳定工况的时间区间  $[t-d, t]$  后,即可得到该稳定状态下的运行数据。表 2 为筛选后机组稳态运行部分数据。

表 2 筛选后机组稳态运行部分数据  
Tab.2 Operation data of the unit after steady state screening

空气质 量流量/ (t·h <sup>-1</sup> )	燃气质 量流量/ (t·h <sup>-1</sup> )	压气机 进口 温度/℃	压气机 进口 压力/kPa	压气机 出口 温度/℃	压气机 出口 压力/kPa
2 342.71	49.26	5.17	15.80	379.74	1 595.95
2 342.67	48.26	4.68	15.79	373.84	1 596.51
2 343.09	49.26	5.14	15.80	379.54	1 595.92
2 338.55	48.26	4.72	15.79	373.53	1 596.56
2 340.64	48.27	4.76	15.79	373.57	1 596.60
2 342.83	49.26	5.11	15.80	379.34	1 595.90
2 353.01	48.60	6.72	15.80	379.45	1 601.39
2 341.40	48.27	4.80	15.79	373.46	1 596.65
2 340.46	49.26	5.08	15.80	379.28	1 595.87
2 350.58	48.61	6.74	15.80	379.52	1 601.44

#### 2) 标准化处理

对采集的运行数据进行稳态筛选后,通过极值标准化(式(2))处理,将原始数据缩放至  $[0, 1]$  区间,更有利于待优化参数量化区间的统一管理。

$$X_i = (X'_i - \min(X'_i)) / (\max(X'_i) - \min(X'_i)) \quad (2)$$

式中:  $X_i$  为标准化后的数据;  $\max(X'_i)$ 、 $\min(X'_i)$  分别为原始数据  $X'_i$  的上、下限值,可由数据库中 SOL 语句进行筛选得到。

标准化处理后的部分数据见表 3。

表 3 标准化处理后的部分数据  
Tab.3 Part of the data after standardization

空气质 量流量	燃气质 量流量	压气机进 口温度	压气机 进口压力	压气机 出口温度	压气机 出口压力
0.326 0	0.493 9	0.098 8	0.992 2	0.086 2	0.314 2
0.493 4	0.494 6	0.101 2	0.969 4	0.146 5	0.323 0
0.492 3	0.495 3	0.103 2	0.969 3	0.152 8	0.331 8
0.496 7	0.496 0	0.104 3	0.969 2	0.166 3	0.340 7
0.504 1	0.496 7	0.105 5	0.969 1	0.183 8	0.349 5
0.511 1	0.497 4	0.106 7	0.969 0	0.187 2	0.358 3
0.513 7	0.498 1	0.107 9	0.969 0	0.190 5	0.367 1
0.515 5	0.498 8	0.109 1	0.968 9	0.193 9	0.375 9
0.508 8	0.499 5	0.110 3	0.968 8	0.154 3	0.384 7
0.507 6	0.500 2	0.111 5	0.968 7	0.159 3	0.393 5

#### 3) 隶属度设置和量化区间的划分

采用区间划分的方法对数据进行分区,以提高算法的运算效率。采用隶属度函数进行量化区间的划分。当参数值在 0~0.2、0.2~0.4、0.4~0.6、0.6~0.8、0.8~1.0 之间分别为极低(VN)、低(N)、正常(NM)、高(P)、极高(VP)的量化区间。

将数据集中数据采用量化区间处理后作为量化区间关联规则挖掘的输入矢量,这样有利于提高运算效率,预防提前收敛。量化区间划分结果见表 4。

表 4 量化区间划分结果

Tab.4 The results of quantitative interval partition

空气质 量流量	天然气 质量流量	压气机进口 空气温度	压气机进口 空气压力	压气机出口 空气温度	压气机出口 空气压力
N	NM	VN	VP	VN	N
NM	NM	VN	VP	VN	N
NM	NM	VN	VP	VN	N
NM	NM	VN	VP	VN	N
NM	NM	VN	VP	VN	NM
NM	NM	VN	VP	VN	NM
NM	NM	VN	VP	VN	NM
P	NM	VN	VP	N	NM
VP	NM	VN	VP	NM	NM

最后通过基于 Hadoop 平台的数据挖掘算法得到优化求解量化区间, 根据求解量化区间, 利用隶属度函数对求解结果进行反量化区间化处理, 计算出真正的求解区间。

#### 4) 阈值设置

最小支持度和最小置信度是关联规则数据挖掘算法中最关键的 2 个阈值, 阈值设定直接决定算法的求解效率和求解结果的满意度。因此, 阈值的设置需要和电厂有经验的运行管理人员沟通, 获取更多的电厂参数调控的经验知识之后, 再根据对数据量化区间的整体估计情况和对参数调控的重要性进行。对于求解关联规则过多的情况, 本文在 MapReduce 框架最后的 Finalize 函数中添加兴趣度算法, 用于剔除无效的关联规则, 把用户不感兴趣的关联规则过滤掉, 有效克服了传统方法可能挖掘出过多无效规则的缺点。

### 3.3 基于关联规则数据挖掘算法分布式处理

本文采用基于 Hadoop<sup>[16]</sup>中 MapReduce<sup>[17]</sup>编程框架, 实现对电厂数据的分布并行化处理, 以提高数据处理的效率。Hadoop 是由 Apache 最早设计的大数据处理系统架构, 能够提高分布式处理的实现效率, 使得对大批量数据进行同步计算更加快速高效。MapReduce 分布式计算框架通常由 Map(映射)函数和 Reduce(化简)函数构成, 对于电厂获取的海量数据的深度挖掘, 若采取传统的基于小数据分析的数据挖掘平台, 则对大批量数据的处理效率会显著降低, 所以本文采用基于 Hadoop 中 MapReduce 编程框架, 实现对电厂海量数据的分布并行式处理, 以提高数据处理的效率, 节省时间。图 8 为基于关联规则数据挖掘算法分布式处理流程。

### 3.4 调控结果分析

选择当联合循环功率处于较优区间 (Tag33=

VP) 时对应的运行参数量化区间, 剔除一些无效的关联规则后, 剩余有效的关联规则, 表 5 为关联规则数据挖掘结果。

表 5 关联规则数据挖掘结果

Tab.5 The mining results of association rules

序号	关联规则挖掘结果
1	Tag33=VP $\Rightarrow$ Tag14=P $\wedge$ Tag20=VP
2	Tag33=VP $\Rightarrow$ Tag30=P $\wedge$ Tag31=N
3	Tag33=VP $\Rightarrow$ Tag28=N
4	Tag33=VP $\Rightarrow$ Tag27=N
5	Tag33=VP $\Rightarrow$ Tag12=VP Tag29=VP
6	Tag33=VP $\Rightarrow$ Tag29=VP
7	Tag33=VP $\Rightarrow$ Tag30=VP
8	Tag33=VP $\Rightarrow$ Tag14=P
9	Tag33=VP $\Rightarrow$ Tag21=VP
10	Tag33=VP $\Rightarrow$ Tag26=VP

通过将数据挖掘结果反量化区间处理, 如取第 1 条有效的强关联规则 Tag33=VP $\Rightarrow$ Tag14=P $\wedge$ Tag20=VP, 即 Tag33 $\in$ [0.8,1.0] $\Rightarrow$ Tag14 $\in$ [0.6,0.8] $\wedge$ Tag20 $\in$ [0.8,1.0], 得到结果: Tag33 $\in$ [363.924 7, 381.121 9] $\Rightarrow$ Tag14 $\in$ [881.243 1, 888.429 3] $\wedge$ Tag20 $\in$ [831.332 9, 838.295 4]。

上述关联规则表示在满负荷工况下, 联合循环机组的循环功率在[0.8, 1.0]量化区间内时, 运行机组的燃烧室进口烟气温度量区间设置在[0.6, 0.8], 再热器出口蒸汽温度设置在[0.8, 1.0]的情况满足支持度和置信度的要求。

结果表明: 在联合循环功率较高的区间时, 余热锅炉入口烟气温度的最优值区间为[881.243 1, 888.429 3] K, 再热器出口蒸汽温度的最优值区间为[831.332 9, 838.295 4] K。运行优化的参数值在最优区间内选取, 为了研究方便, 本文取区间的中心值作为最优值, 得到余热锅炉入口烟气温度最优值为 884.236 2 K, 再热器出口蒸汽温度的最优值为 834.814 2 K。

按照上述步骤, 针对数据挖掘的所有有效的关联规则, 得到特定工况下机组运行参数的一组最优值。采用传统方法和改进关联规则方法确定的部分可控运行参数优化前后对比如图 9 所示, 图 9 中 1~10 分别代表余热锅炉出口烟气温度(K)、余热锅炉出口烟气压力(MPa)、再热器出口蒸汽温度(K)、再热器出口蒸汽压力(MPa)、低压缸排气温度(K)、高压缸排气压力(MPa)、高压缸排气温度(K)、低压缸进口蒸汽温度(K)、低压缸进口蒸汽压力(kPa)和燃烧室进口烟气温度(K)。



由此可得,并非所有运行参数与联合循环功率都存在强关联,通过改进后的关联规则数据挖掘算法求取的参数最优设置值,可以看到原始设置参数普遍低于优化值,这表明机组运行过程中将参数值控制在安全范围内以减少安全事故的发生,但该设置并未达到最优值,无法保证机组运行功率达到最高,使得发电效率有所限制。

本文采用改进关联规则数据挖掘算法确定的可控运行参数优化的目标值,通过在线性能监测系统对优化前后的性能指标进行模拟仿真试验,可得出机组联合循环功率提升了 0.412 5%,针对该电厂机组实际运行情况,年增加电能产量 4 806 MW·h,燃气轮机发电电价 0.68 元/(kW·h),如果每年每台机组运行 3 000 h,电厂每年可增加收益 326 万元。

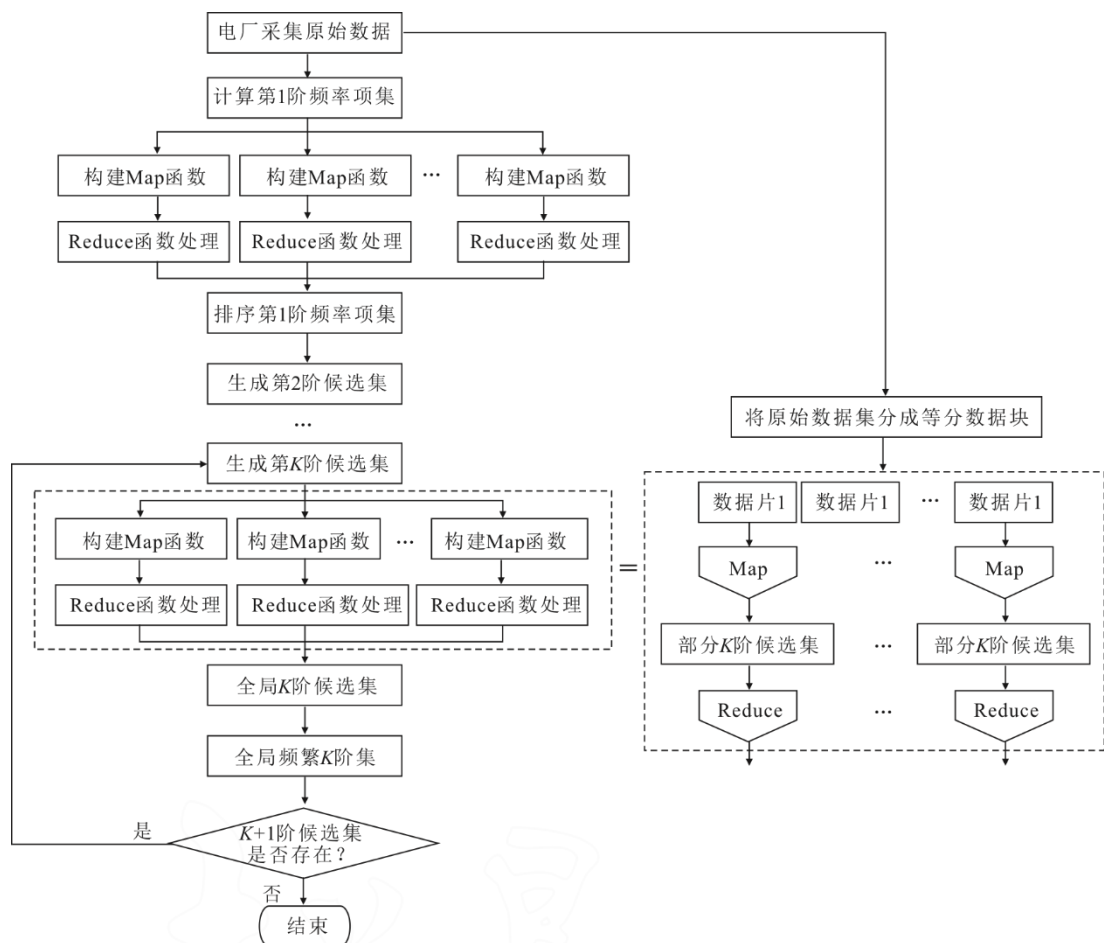


图 8 关联规则数据挖掘算法分布式处理流程

Fig.8 The distributed processing of association rule mining algorithm

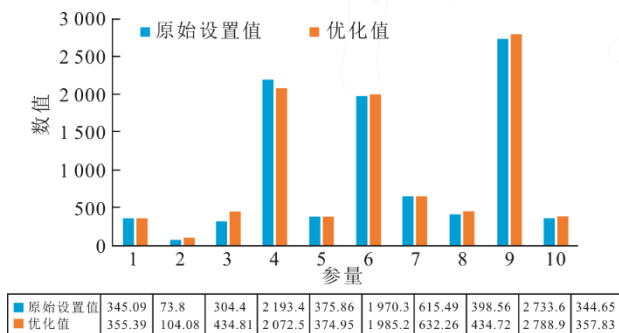


图 9 运行参数优化前后对比

Fig.9 The operation parameters before and after the optimization

## 4 结 论

1) 提出了一种基于数据挖掘的电厂机组运行优化及调控系统,该系统结构框架包括数据预处理、健康状态评估指标挖掘、运行关键特征参数筛选、健康状态聚类分析、稳定工况库建立、状态评估指标特征获取、实时特征参数预测和智能调控等模块,用于电厂机组运行过程中关键参数的提取、特征挖掘、后处理及优化调控,解决了人工调控的诸多问题。

2) 针对传统关联规则数据挖掘算法在电厂机



组优化中的问题,本文提出了改进的关联规则数据挖掘算法,通过 MapReduce 框架分布并行式处理数据,并在生产实践中应用,提高了电厂机组运行参数优化过程的效率。

3) 采用改进关联规则挖掘算法确定的可控运行参数优化的目标值,通过在线性能监测系统对优化前后的性能指标进行模拟仿真试验,得出机组联合循环功率提升了 0.412 5%,对于该联合循环机组运行的情况下,年增加电能产量 4 806 MW·h,年增加收益约 326 万元。

### 参考文献

- [1] 申鹏飞. 基于数据仓库的汽轮机性能分析预测系统的设计与实现[D]. 北京: 华北电力大学, 2005: 26-39.  
SHEN Pengfei. Design and implementation of steam turbine performance analysis and prediction system based on data warehouse[D]. Beijing: North China Electric Power University, 2005: 26-39.
- [2] OGILVIE T, SWIDENBANK E, HOGG B W. Use of data mining techniques in the performance monitoring and optimisation of a thermal power plant[C]//Knowledge Discovery & Data Mining. IET, 1998: 7.
- [3] KUSIAK A, SHAH S. Data-mining-based system for prediction of water chemistry faults[J]. IEEE Transactions on Industrial Electronics, 2006, 53(2): 593-603.
- [4] KUSIAK A, BURNS A, MILSTER F. Optimizing combustion efficiency of a circulating fluidized boiler: a data mining approach[J]. International Journal of Knowledge-Based and Intelligent Engineering Systems, 2005, 9(4): 263-274.
- [5] 何怀昌, 李晓东, 罗肖, 等. 基于大数据的燃气-蒸汽联合循环机组性能分析及诊断系统[J]. 热力发电, 2020, 49(11): 14-19.  
HE Huaichang, LI Xiaodong, LUO Xiao, et al. Performance analysis and diagnosis system of gas-steam combined cycle unit based on big data[J]. Thermal Power Generation, 2020, 49(11): 14-19.
- [6] 王秋平, 陈志强, 魏浩. 基于数据挖掘的电站运行参数目标值优化[J]. 电力科学与工程, 2015, 31(7): 19-24.  
WANG Qiuping, CHEN Zhiqiang, WEI Hao. The summary of optimal operation parameters in power station based on the data mining[J]. Electric Power Science and Engineering, 2015, 31(7): 19-24.
- [7] 郭华新, 高明. 基于数据挖掘的煤质特性与电厂辅机能耗的关联分析[J]. 科技信息, 2012(2): 147-149.  
GUO Huaxin, GAO Ming. Correlation analysis of coal quality characteristics and power plant auxiliary energy consumption based on data mining[J]. Science & Technology Information, 2012(2): 147-149.
- [8] TARIQ I, SINDHU M A, ABBASI R A, et al. Resolving cross-site scripting attacks through genetic algorithm and reinforcement learning[J]. Expert Systems with Applications, 2020, 168(1): 114386.
- [9] 朱清智, 董泽, 马宁. 基于即时学习算法的短期负荷预测方法[J]. 电力系统保护与控制, 2020, 48(7): 92-98.  
ZHU Qingzhi, DONG Ze, MA Ning. Forecasting of short-term power based on just-in-time learning[J]. Power System Protection and Control, 2020, 48(7): 92-98.
- [10] MALCHI S K, KALLAM S, AL-TURJMAN F, et al. A trust-based fuzzy neural network for smart data fusion in internet of things[J]. Computers & Electrical Engineering, 2021, 89: 106901.
- [11] 汪洋子, 陈茜. 计及冷、热、电联产的气电互联网络最优调度[J]. 电力系统保护与控制, 2019, 47(3): 168-175.  
WANG Yangzi, CHEN Qian. Optimal dispatch for the integrated electrical and natural gas network with combined cooling, heat and power plant[J]. Power System Protection and Control, 2019, 47(3): 168-175.
- [12] SHARMILA S, VIJAYARANI S. Association rule mining using fuzzy logic and whale optimization algorithm[J]. Soft Computing, 2021, 25(1): 1-16.
- [13] 高旭, 刘宏君, 杜丽艳, 等. 基于综合评价识别法的智能变电站虚回路在线状态监测技术研究[J]. 电力系统保护与控制, 2019, 47(3): 182-187.  
GAO Xu, LIU Hongjun, DU Liyan, et al. Study on the online status monitoring technology of intelligent substation virtual circuit based on comprehensive evaluation and recognition method[J]. Power System Protection and Control, 2019, 47(3): 182-187.
- [14] 张苏闽, 邓敏强, 朱静, 等. 基于大数据挖掘的火电厂辅机状态评估系统[J]. 信息化研究, 2020, 46(1): 64-69.  
ZHANG Sumin, DENG Minqiang, ZHU Jing, et al. Power plant auxiliary state assessment system based on big data mining[J]. Informatization Research, 2020, 46(1): 64-69.
- [15] LIU D R, HSU Y K, CHEN H Y, et al. Air pollution prediction based on factory-aware attentional LSTM neural network[J]. Computing: Archives for Informatics and Numerical Computation, 2021, 103(1): 75-98.
- [16] AHMED N, BARCZAK ANDRE L C, SUSNJAK T, et al. A comprehensive performance analysis of apache Hadoop and apache spark for large scale data sets using HiBench[J]. Journal of Big Data, 2020, 7(1): 388-392.
- [17] MALEKI N, RAHMANI A M, CONTI M. SPO: a secure and performance-aware optimization for MapReduce scheduling[J]. Journal of Network and Computer Applications, 2020, 176(4): 102944.

(责任编辑 杜亚勤)