

基于能源管理的新能源系统运行数据综合处理分析

侯嘉麒

(中国能源建设集团山西省电力勘测设计院有限公司, 山西 太原 030002)

摘要: 首先提出了基于能源管理的新能源系统运行数据综合处理方案, 分析了新能源运行管理的基本需求; 说明了数据接入网络的基本构成; 其次分析了数据的存储结构及新能源运行的大数据特征, 建立了基于能源管理的系统数据运行架构; 然后分析了平台设计的智能算法, 并提出数据分类的分步处理方法; 最后利用实际案例论证了所提出平台和数据方法处理的有效性。

关键词: 数据分析; 能源管理; 系统设计; 新能源系统

Comprehensive processing and analysis of new energy system operation data based on energy management

HOU Jiaqi

(China Energy Engineering Group Shanxi Electric Power Engineering Co., Ltd., Taiyuan 030002, China)

Abstract: In order to improve the management effect of new energy dispatching operation, a new energy dispatching operation analysis system based on data platform is proposed. This paper proposes a comprehensive processing scheme for new energy system operation data based on energy management. The basic requirements of new energy operation management are analyzed. The basic structure of the data access network is explained, and the storage structure of the data and the big data characteristics of the new energy operation are analyzed. The system data operation framework based on energy management is established, and the intelligent algorithm of platform design is analyzed. And a step-by-step processing method for data classification is proposed. Finally, the actual case is used to demonstrate the effectiveness of the platform and data processing methods proposed in this paper.

Key words: data analysis; energy management; system design; new energy system

0 引言

随着环境保护需求的紧迫性不断增加, 可再生能源在电网中的渗透率也逐渐增加, 并取代传统化石能源为主的发电形式^[1]。作为清洁发电的首要选择, 这给优化运行带来了新的挑战。在网络中的渗透率需要以更精准的数据模型进行分析归纳。并且随着可再生能源渗透率的增加, 微电网系统的灵活性有降低趋势, 需要提升对储能装

置的利用效率, 以平衡可再生能源给电网运行带来的波动^[2]。上述内容均需要在较长时间尺度和更精细的时间颗粒度方面进行数据分析。为此, 本文针对新能源机组运行的数据处理进行分析。

目前针对新能源运行数据分析, 文献[3]提出了基于数据平台的新能源运行调度分析系统; 文献[4]分析了极端天气事件对新能源发电和电网运行影响; 文献[5]分析了面向新能源多业务

场景的大数据平台构建关键技术；文献[6]提出了基于测试和运行数据的新能源快速调频参数动态修正算法；文献[7]分析了新能源数据全景化监测与综合分析管理。

由此可见，目前针对新能源运行数据的处理和分析仍存在进一步深化的必要。为解决上述问题，本文提出新能源运行数据处理方法，将全时段数据作为数据输入，得到整个时段的代表运行周期。当处理完全运行数据后，利用时间序列运行聚类算法进行排序，并将聚类结果的相似性进行对比，形成融合数据，具体为连续运行小时和周数。

1 新能源调度运行分析

1.1 新能源运行管理概述

可再生能源主要包括风能、光伏、生物质能、地热、水能、压缩空气储能以及核能等。可再生能源发电的组织形式需要对能源的利用效果和生形式进行评估，以提升对能源需求的针对性。针对参与能源使用的实施决策进行数据分析，确保与预测负荷需求相匹配大数据分析和可

再生能源运行数据相辅相成。根据大数据流分析，可以提升对可再生能源等形势的新能源机组运行质量管控的能力^[8]。

不同种类的负荷在容量、数量以及并网电压等级方面存在着显著的空间分布不均衡。考虑到风电、光伏和储能等新能源在电力系统中的消纳问题，通过合理优化新能源运行实现对不均衡分布的用户负荷进行有效协调。目前，全国范围内统一的协调调度模式在面对新能源以及负荷波动等情况下的响应速度和响应模式难以支撑运行信息的进一步发展。由于在滚动优化、实施优化阶段，海量的数据传输占据了主要的信息传输通道，造成相应的数据丢失和数据质量下降，难以满足协调优化调度的数据指标要求。

从数据方面分析，根据数据平台可以提升区域内优化、区域之间协调的运行优化模式，有效提升数据传输效率，提升信息交互能力，避免不同区域以及区域内部调度优化运行数据的传输障碍。新能源数据交互基本结构如图1所示^[9]。

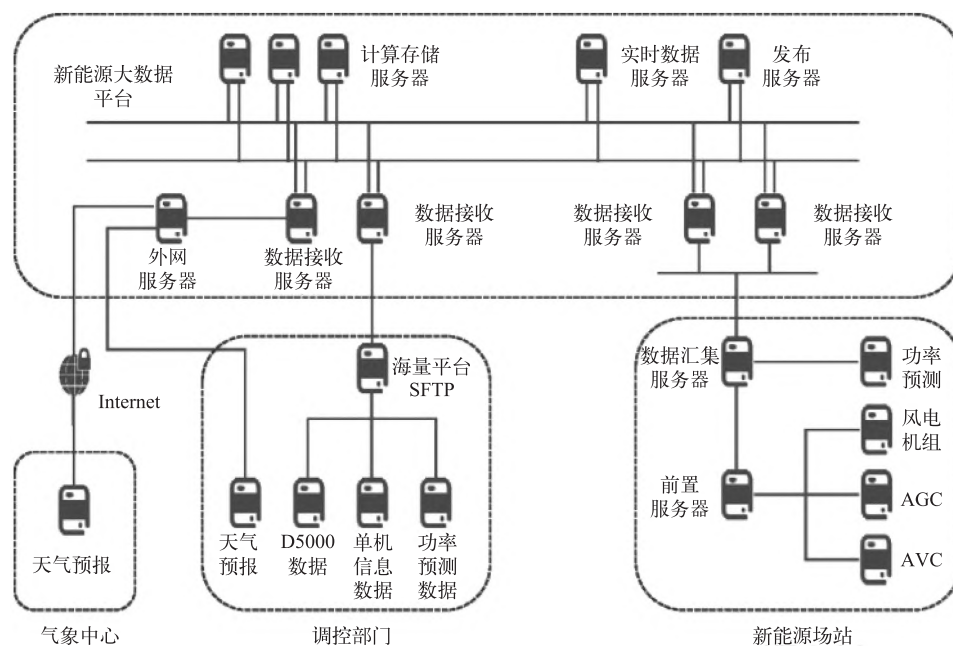


图1 智能调度系统

Fig. 1 Smart dispatch system

1.2 数据存储结构

合理的数据存储结构是对数据进一步分析和应用的前提。按照分层的结构设计相应的存储模式。每层与上层和下层通过边界接口进行交互，避免跨层信息错误流动。具体结构如图2所示，包括存储层、数据接入层和计算层。数据存储层根据名称节点进行Hadoop聚类。聚类服务器利用Linux系统并使用Hadoop HDFS文件系统进行数据存储。基于HDFS构建分布式数据库Hbase，用于海量数据监视、存储和管理^[10]。在数据接入层和计算层，数据可根据上层结构的应用需求，将数据写入HDFS文件操作。另外，数据也可存储在Hbase中。HDFS文件存储具有更高的灵活性和可塑性^[11]。数据格式可由用户定义。数据队列和分析功能可根据用户程序执行。Hbase也可允许用户自定义数据存储模式，提供更加多样化的数据请求接口。这种结构对于含有大量新能源运行信息的历史数据在线请求服务十分有益。相比于HDFS文件接入，本文提出的数据接入方式具有更低的接入延迟，对于直接存储在HDFS文件中的数据，Map Reduce可用于完成大数据分析。

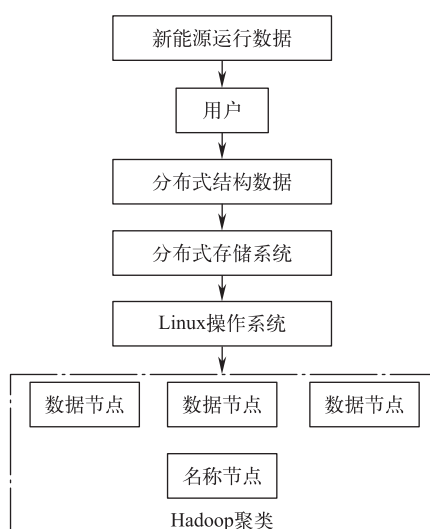


图2 数据存储结构

Fig. 2 Data storage structure

由于新能源运行数据具有实时性、可变性和多结构特性，因此在有限时间内对这类数据

进行分析首先需要梳理数据结构。新能源大数据分析可利用大数据流分析手段实现低延迟和高可靠性。

1.3 能源管理大数据特征分析

新能源运行数据具有大数据的特征，并且随着新能源机组数量的增加呈现出大量性。大量数据的处理需要分类。目前较常用的分类方法为K最邻近算法^[12]。同时，对于大量数据的处理方法还包括并行处理，可根据分布式决策数学习实现大数据流的挖掘。分布式电源运行大体量数据的处理方法如图3所示。

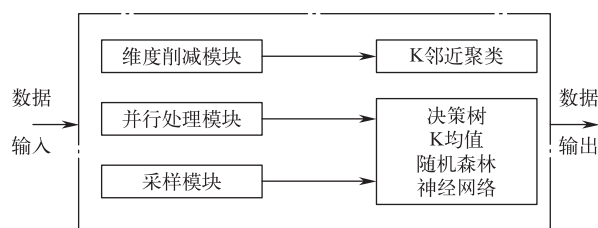


图3 大体量数据处理方法

Fig. 3 Large volume data processing methods

含有分布式电源运行的数据一般需要快速处理、实时分析^[13]。针对这一需求，相应的数据挖掘技术也不断被应用到这类数据处理中。包括SAMOA、PARMA等并行处理技术。对于分布式电源大数据高速性要求的处理流程如图4所示。

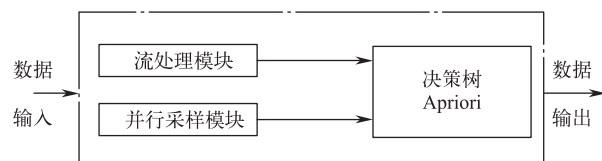


图4 处理流程

Fig. 4 Process flow

不同类型的分布式电源产生多样性数据，多样性数据的处理需要针对不同的数据结构和数据处理流程进行分类。根据大数据多样性的要求，不同的数据类型应对不同的数据处理方法，在处理时间和处理效果上均有所差别，以不同的方式进行呈现，最终得到具有多维度、全方位的运行数据。

2 基于数据融合的新能源运行数据监测系统

2.1 平台总体构架

平台总体架构如图5所示。分为边缘层、IaaS、PaaS和SaaS层。具体功能为数据采集、边缘计算、数据融合平台、大数据基础平台和应用服务。

2.2 平台智能算法

平台智能算法如图6所示。为了提升运算效率，降低系统冗余，平台针对数据质量提升和多

业务场景应用的需求，嵌入了大量数据挖掘及数据融合算法，包括聚类算法、线性回归算法和深度学习算法等。算法种类丰富，能够通过专用或组合的方法实现新能源异常数据检测及清洗、关键部件故障识别、光伏图像识别、功率预测和风电机组控制优化等功能应用，为新能源多业务场景的需求提供解决方案。

2.3 系统接口

该综合处理平台作为一个中央化的数据处理和管理系统，与各种其他能源系统进行数据交互和集成，以实现更高效的能源管理和运营。

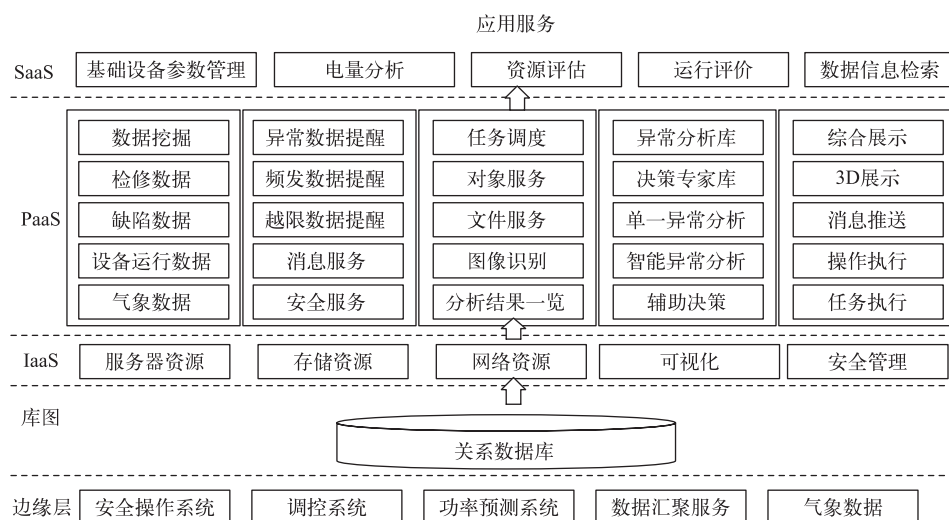


图5 平台总体架构

Fig. 5 Overall architecture of the platform

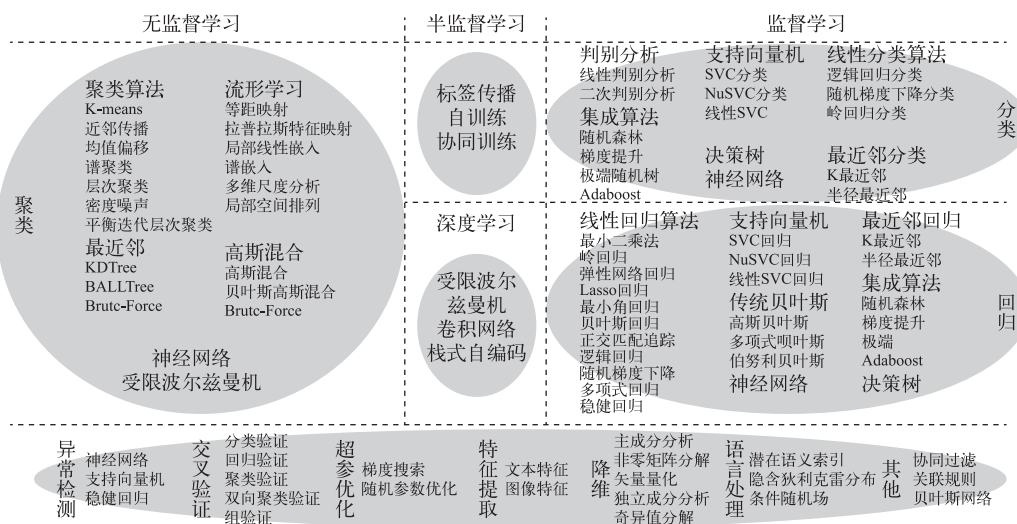


图6 平台智能算法

Fig. 6 Platform intelligent algorithm

与发电系统的接口：平台可以连接到不同类型的新能源发电设备，例如太阳能光伏、风能发电等，收集实时的发电数据和性能指标，以监测运行状态、发电量等信息。

与电网系统的接口：平台可以与电网系统进行数据交换，实时获取电网负荷和供应情况，从而进行优化调度，确保能源的稳定供应和分配。

与储能系统的接口：如果存在储能系统，平台可以连接到这些系统，监测储能装置的充放电状态，根据能源需求进行合理的能量储存和释放。

与用能系统的接口：平台可以与建筑、工业设施等用能系统进行连接，获取能耗数据，分析能源使用情况，提出节能建议，实现能源的合理利用。

与监测设备的接口：平台可以接入各类传感器和监测设备，收集环境数据、设备运行数据等，以支持对能源系统的全面监测和管理。

与管理决策系统的接口：平台所提供的数据和分析结果可以用于能源管理者的决策支持，帮助制定更合理的能源策略、维护计划和资源分配。

综上所述，此综合处理平台与其他能源系统之间的接口关系是相互连接、数据共享的关系，通过数据整合与分析，实现对整个新能源系统的集中监控、运营优化和管理决策支持。以控制中心为核心，形成基于多种数据结构的数据新能源运行平台^[14]，如图7所示。

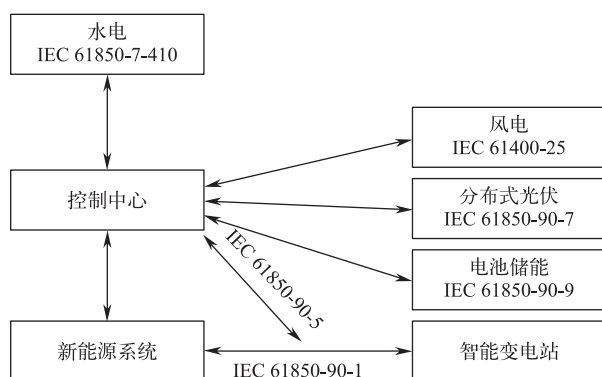


图7 接口示意

Fig. 7 Interface diagram

3 平台分步数据分类处理方法

两阶段数据处理方法用于将全部运行数据转化为代表性运行时段数据，从而对含有大规模可再生能源的电网规划和调度进行分析。全负荷信息以及规划时期内气象条件均属于全数据信息范围。在数据处理过程中，代表性运行周数可从全数据中抽取并作为第一阶段样本，代表性运行小时作为第二阶段样本。这种方法利用了原始数据相应工作周数之间的相似性以及对应时段的相似性，可以有效减少数据维度，提高原始数据之间的相关性。另外，与大部分变形运行周期抽取方法不同的是，本文所提出的方法能够按照运行和规划时间序列存储相应的数据。本文提出的方法流程如图8所示。

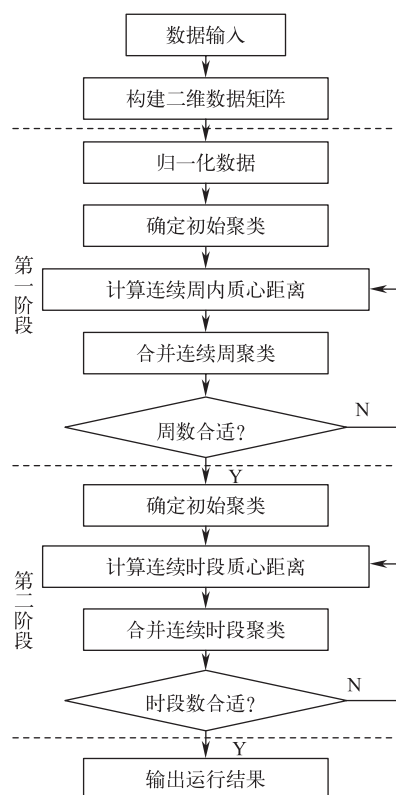


图8 数据分类流程

Fig. 8 Data classification process

对于体现分布式电源运行的数据，在类型和时间尺度方面的差异性，需要进行数据预处理。所有运行数据根据不同的时间尺度范围进行分类，形成二维数据矩阵，作为数据处理的输

入。其中维度之一为时间，另一维度为不同区域和数据类型。例如。对于小时段需求，光照辐射强度和风速信息作为不同数据位置，年数据矩阵有 $6 \times 8\ 760$ 个元素。数据处理方法利用聚类分析，不同类型的数据对应不同的单元，不同位置的数据具有明显的差异性。因此所有数据需要在数据预处理阶段进行规划，保证在数据聚类分析中计算的距离对不同的数据集有相同的权重。

数据处理将原始数据中相似度较高的数据融合形成以运行周数为时间尺度的数据。根据时序运行数据聚类的方法，对层次聚类技术进行相应的修改。在运行周期内，每个运行周作为初始聚类。然后，利用所提出的聚类方法，将连续运行的周数聚合成一类，形成所有数据时间范围序列化集合，从而降低运行点数。以加权欧几里得距离作为判断连续聚类差异性的标准。非连续聚类之间的距离视为 ∞ 。在所提出的距离计算方法中，两种聚类之间元素数用于计算调合平均数，从而避免在单独聚类中存在较多运行周数，同时简化聚类中元素较少的类^[15]。

$$D_w(X, Y) = \frac{2}{\frac{1}{|X|} + \frac{1}{|Y|}} \|C_X - C_Y\|^2 \quad (1)$$

式中， C_X 和 C_Y 分别为 X 和 Y 类的质心； $|X|$ 和 $|Y|$ 为相应的元素。

连接准则包括单连接、全连接和平均连接等。本文采用质心连接方法。这种连接准则比较每类数据的中心，并将连续最近数据类合并。数据 X 的质心计算式为

$$C_{Xi} = \frac{1}{|X|} \sum X_{ki} \quad (2)$$

得到该阶段的聚类结果后，每类数据质心距离最近的单周作为该类的代表周数。这种方法避免了不同代表周数运行时段之间波动性带来的平均差异。每次运行第一阶段数据处理结果，则运行周数减1。如果 M 和 N 分别为初始运行周数和削减后运行周数，则聚类次数可由 $M - N$ 求出。每类数据中的元素数量可根据相应周数的比重进行权重分配。

数据处理将属于相同运行周数内的连续运行小时数据进一步合并，从而减少数据维度。这种数据处理方法也采用时序运行聚类方法。在代表周数内的运行小时根据第一阶段的数据处理方法得到，并作为初始聚类。距离较短的时段和连续时序时间合并为一类。隶属于不同运行周数的运行时间不可进一步合并。第二阶段的加权欧几里得距离计算方法为

$$D_H(X, Y) = \frac{2\sqrt{w_i}}{\frac{1}{|X|} + \frac{1}{|Y|}} \|C_X - C_Y\|^2 \quad (3)$$

在计算距离过程中，代表周数的方均根权重作为衡量权重较大周数连续小时距离的指标。利用这种方法，能够有效减少典型运行周数内的小时数，从而在数据信息含量方面有较大提升。链接准则仍然采用志兴链接。与第一阶段不同的是，第二阶段直接利用每个聚类的质心作为代表小时数。

在数据预处理阶段中，对全数据进行规划处理。数据处理结果需要根据实际名称进行存储。因此，这类数据代表不同运行模型的物理特性。这类数据处理结果根据时间顺序进行排列，不仅包括不同类型的数据，而且包括选择的周数和运行小时数的权重。数据处理结果为二维数据矩阵。例如，如果两个不同位置的三种数据削减为672个运行点，则数据矩阵含有 8×672 个元素。前六行代表两种位置的三种类型数据。第七行代表相应周数的权重，第八行代表相应时段的权重。

4 算例分析

针对某地区典型多能源电网进行应用分析，读取该地区新能源运行管理数据库。数据库中包含各监测点电压、电流、频率、有功、无功、视在功率、三相不平衡和谐波等。除了实时数据外，也包含最大值、最小值、平均值以及95%概率值等各类统计数据。历史数据采样间隔为5 min，气象监测部门提供的数据精确到1 h。某典型日内各机组的出力情况如图9所示。

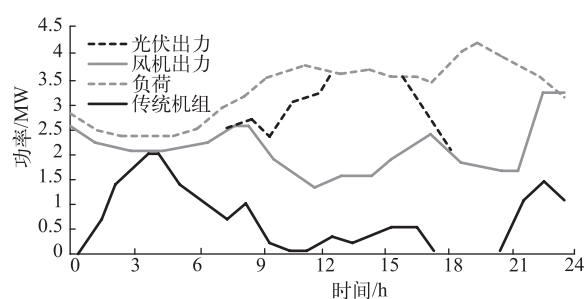


图9 典型日机组出力

Fig. 9 Typical daily unit output

根据负荷需求的变动, 光伏、风机和传统机组的出力呈现波动趋势, 在新能源发电上网过程中, 优先保证风机和光伏的上网发电。

另外, 读取夏季一周内典型负荷曲线, 结合该地区温度变化和相关系数表格, 得到曲线如图10所示。

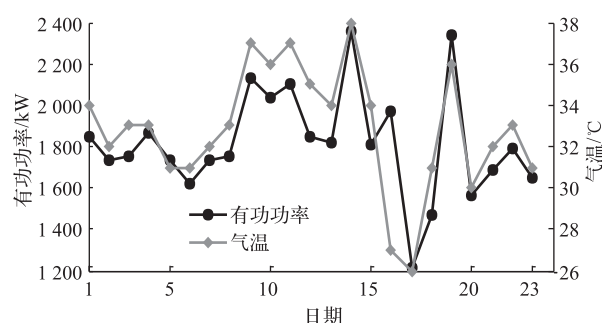


图10 功率与气温关系

Fig. 10 Relationship between power and temperature

日最大负荷与气温相关系数见下表。由下表可知, 夏季时日最大负荷与最高气温的线性相关度较强, 并且为正相关, 即最高气温越高, 有功消耗越大。

表 日最大负荷与气温相关系数

Tab. Correlation coefficient between daily maximum load and temperature

相关系数	夏季日最大负荷
日最高气温	0.784 6
日平均气温	0.531 9
日最低气温	0.392 2

5 结束语

根据本文提出的系统及数据处理方法, 在典型系统内进行测试, 通过仿真分析, 说明本文提

出的平台系统能够有效解决电网数据分析问题, 有效提升了地区新能源在跨区以及区域内部运行的质量和效率。

针对实时数据和历史数据的处理, 可以利用本文提出的平台进行选择分析, 提高数据分析的针对性和有效性。

参考文献

- [1] 潘广旭, 裴丽伟, 李英杰, 等. 可再生能源冷热电分布式供能系统智慧高效运行[J]. 山东电力技术, 2022, 49(8): 15-21.
- [2] 李付存, 张岩, 于丹文, 等. 高比例新能源接入下谐波电能计量方法研究[J]. 山东电力技术, 2022, 49(3): 52-58.
- [3] 刘显茁, 王皓怀, 田伟达, 等. 基于数据平台的新能源运行调度分析系统[J]. 机械与电子, 2023, 41(3): 13-17, 22.
- [4] 鞠冠章, 王靖然, 崔琛, 等. 极端天气事件对新能源发电和电网运行影响研究[J]. 智慧电力, 2022, 50(11): 77-83.
- [5] 王阔, 刘辉. 面向新能源多业务场景的大数据平台构建关键技术与应用[J]. 全球能源互联网, 2022, 5(2): 157-165.
- [6] 柯贤波, 程林, 牛拴保, 等. 基于测试和运行数据的新能源快速调频参数动态修正算法[J]. 控制理论与应用, 2022, 39(3): 546-552.
- [7] 王鹤, 郑涛, 杨宇峰, 等. 新能源数据全景化监测与综合分析管理技术[J]. 电子技术与软件工程, 2018(24): 129.
- [8] 彭昕, 张曦, 陆建宇, 等. 区域电网新能源运行管理机制设计[J]. 西北水电, 2019(4): 1-4.
- [9] 王阔, 刘辉. 面向新能源多业务场景的大数据平台构建关键技术与应用[J]. 全球能源互联网, 2022, 5(2): 157-165.
- [10] 明哲, 余芸, 甘杉. 电力业务流转数据库中分布式数据一致性算法[J]. 电子设计工程, 2023, 31(1): 120-123, 128.
- [11] 占永红, 王东, 吴丽莎, 等. 基于HDFS的电力设备监测数据存储设计研究[J]. 能源与环保,

- 2017, 39(12): 210-213.
- [12] 顾杨青, 兴胜利, 何平, 等. 基于大数据云平台的电力能源大数据采集与应用探讨[J]. 网络安全和信息化, 2023(1): 60-63.
- [13] 蒋荣, 陈可, 陈学先, 等. 能源大数据中心数据模型研究[J]. 能源与环保, 2023, 45(2): 243-249.
- [14] 杨颖, 余跃, 华钧, 等. 新能源大数据分析平台的实践[J]. 煤气与热力, 2021, 41(10): 24-26,

45-46.

作者简介

侯嘉麒（1989—），男，硕士，工程师，主要研究方向：光伏，风电，储能等新能源电站项目设计应用。

（收稿日期：2023-05-16）