

DOI: 10.19783/j.cnki.pspc.200636

基于海量数据的多维度负荷特性分析系统开发

曹华珍¹, 吴亚雄¹, 李浩¹, 高崇¹, 唐俊熙¹, 管维灵²

(1. 广东电网有限责任公司电网规划研究中心, 广东 广州 510030;

2. 苏州华天国科电力科技有限公司, 江苏 苏州 215000)

摘要: 随着智能电网的发展, 在已知用户负荷所属行业分类的基础上, 利用海量用户负荷数据进行负荷特性分析, 可以提取行业典型负荷特征, 为开展业扩报装工作及其他配电网规划工作提供指导, 同时提高配用电的智能化水平。在此背景下, 为了提高负荷特性分析工作效率, 开发了基于海量数据的多维度负荷特性分析系统。首先, 设计了多维度负荷特性分析的各个功能模块。然后, 对负荷特性分析的多个功能进行整合, 介绍了多维度用户负荷特性分析功能的实例。最后, 对该系统的负荷特性分析可视化界面进行了展示说明。总的来说, 所开发的多维度负荷特性分析系统以 Java 语言为基础, 采用微服务、微应用架构, 支持所有功能的灵活拓展以及各个功能块之间的快速数据传输交互, 可为基于海量数据的负荷特性分析提供便利。此外, 该系统底层算法能够同时支持多种语言进行开发, 对于发出的计算需求可迅速做出响应。

关键词: 多维度负荷特征库; 典型行业; 微服务架构设计; 功能模块化开发; 数据可视化

Development of a multi-dimensional load characteristic analysis system based on massive data

CAO Huazhen¹, WU Yaxiong¹, LI Hao¹, GAO Chong¹, TANG Junxi¹, GUAN Weiling²

(1. Power Grid Planning & Research Center, Guangdong Power Grid Co., Ltd., Guangzhou 510030, China;

2. Suzhou Huatian Power Technology Co., Ltd., Suzhou 215000, China)

Abstract: As part of the development of the smart grid, and based on the industry classification of known user loads, using massive user load data for load characteristics analysis can extract typical load characteristics of the industry and provide guidance for carrying out industry expansion installation and other distribution network planning. This can improve the intelligence level in power distribution. With this as background, to improve the efficiency of load characteristic analysis, a multi-dimensional load characteristic analysis simulation system based on massive data is designed. First, each functional module is designed for multi-dimensional load characteristics analysis. Secondly, we integrate multiple functions of load characteristic analysis in the system and describe its functions. Lastly, the interfaces of the developed system are introduced and illustrated. Overall, the system is developed based on the Java language, adopts micro-services and micro-application architecture, supports flexible expansion of all functions and fast data transmission and interaction between various functional blocks. It also facilitates the analysis of load characteristics based on massive data. In addition, the underlying algorithm can support multiple languages for development at the same time, and can quickly respond to computing needs.

This work is supported by the Science and Technology Project of China Southern Power Grid Co., Ltd. (No. GDKJXM20172939).

Key words: multi-dimensional load signature database; typical industries; micro-service architecture design; functional modular development; data visualization

0 引言

“十三五”期间, 随着社会经济持续发展, 广

东地区电网最大负荷持续增长, 峰谷差逐步增大, 电网在不同时期内的供需矛盾十分突出, 调峰难度持续加大, 对电力系统稳定性造成了潜在危害^[1-2], 也给电网规划建设、电力负荷预测及电力市场管理分析等工作带来了障碍^[3-4]。同时, 在产业转型升级有力推进、经济结构逐步调整的大背景下, 广东省

基金项目: 中国南方电网有限责任公司科技项目资助 (GDKJXM20172939)

能耗高、污染强的落后企业被大量淘汰，对电网的负荷特性带来了重大影响^[5-6]。因此，对广东省的负荷特性展开全面的研究分析对改善整体电网的规划建设，保证其安稳运行有着重要意义^[7]。

近年来，负荷特性的分析和研究愈加受到重视^[8]，各地电网公司和电力行业的研究者都从不同角度展开了负荷特性的研究。文献[9]中，南方电网云南电网公司采用抽样分析方法，对云南省不同产业中的负荷特征展开研究，并对负荷的发展趋势进行了展望。文献[10]中，国家电网湖南电网公司利用负荷数据，完成了包括年最大负荷在内的多项负荷指标的计算，并根据其电力负荷结构，分析了不同季节下负荷率的不同及其相关影响因素。文献[11]收集唐山中心城区真实的配网负荷数据，重点对居民小区负荷进行了分析，提出了入住率、需用系数等指标以对传统负荷指标体系进行补充，并完成了不同类型小区负荷特性模型的构造。文献[12]归纳了郑州电网中各个负荷特性指标，利用熵权法对灰色关联分析进行改进，并以此对引起负荷特性发生变化的相关因素进行量化，从而确定其中的主要影响因素。

总的来说，上述文献分别利用个案剔除或均值代替等方法对负荷数据进行了简单预处理，这样造成了负荷数据信息和资源的浪费，削弱了数据的客观性，甚至可能带来错误的结果。另一方面，传统负荷特性分析往往重点研究地区负荷总量，缺少对根据行业分类标准划分后的各行业负荷的分析和比较。除此之外，面对负荷变化的多样性以及负荷特性分析手段的与时俱进，现如今缺少一套能够与时俱进、便于拓展的负荷特性综合分析系统，为电网获取用户信息、进行合理规划提供便利。

针对上述问题，本文基于广东电网计量系统中的实际运行数据，采用线性插值法，对海量负荷特性数据中缺失的数据进行填充恢复。在此基础上，采用 Java 开发语言，开发了一套灵活性强、可拓展的多维度负荷特性分析系统。该系统以微服务、微应用为核心架构，实现对各行业负荷特性在不同时间尺度上进行分析研究，并通过对比、统计和总结，对电网负荷预测和规划建设提供参考。

1 多维度负荷特性分析模块开发设计

为实现基于行业划分的负荷特性分析，提取各个用户相关负荷特性指标，总结变化规律，并为后续负荷预测和电网规划等工作提出具体建议等功能，本文以 Java 语言为基础，采用微服务、微应用架构开发了一套多维度负荷特性分析系统。该系统

应该满足以下技术要求：支持所有功能的灵活拓展；各个功能模块之间的数据共享；拥有支持多种采集协议的数据采集接口；满足多种数据类型的储存要求，具有良好的安全防护功能；对于发出的需求可迅速做出响应；底层算法能够同时支持多种语言进行开发。

本节对多维度负荷特性分析系统开发方法进行了介绍，包括微服务架构设计、功能模块化开发、数据采集接口设计、数据的储存及安全、高效数据动态刷新技术和算法集成设计。

1.1 微服务架构设计

传统的单体应用架构会将所有的功能放入同一个 WAR/JAR 包中，几乎没有外部依赖，以实现集中式管理，降低开发难度。但同样的，这样将会导致功能过于集中，代码和数据中心化；复杂性高；形成技术债务积累；扩展能力受限；任何维护都需要对项目进行重构等多方面的缺点。

为此，本文选择使用微服务架构^[13]为核心展开软件开发，如图 1 所示。微服务架构将每一个独立的功能放在指定的服务中，各个服务高度自治，每个服务只关注一个特定的业务功能，业务清晰，易于开发维护^[14]。同时，多个服务可运行，无需等待其他服务的完成，达到平滑伸缩的效果，大大减小了单个微服务的启动工作时间。此外，对单个功能模块进行拓展时，不同于传统架构需要重构整个项目，微服务架构只需要重新对需要拓展修改的特定服务进行部署。

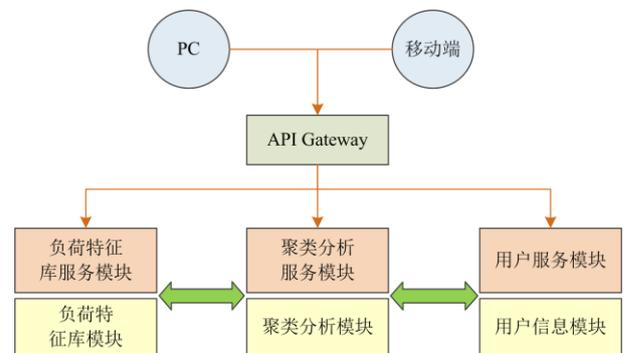


图 1 软件的微服务架构设计

Fig. 1 Software microservices architecture design

微服务架构的引入实现了系统的灵活拓展，减少了其中各个功能模块的开发周期，为系统功能的实时更新提供了极大的便利。

1.2 功能模块化开发

系统中各个功能采用模块化开发，如图 2 所示。其中，前端 HTML(Vue 或 React)以 ajax 请求方式通

过 Nginx 网关调用 Restful API 服务。

其中, API 网关通过 Zuul 构建, 可对外提供跨模块的 HTML 页面访问功能以及跨业务领域数据访问的功能, 引入 Nginx 则有助于服务的横向扩展。

各个功能模块以微服务的形式注入注册中心, 模块之间通过 API 网关进行数据交互, 使用 Nginx 网关为系统的负载均衡提供便利, 确保整个系统的长期稳定运行。不同功能模块利用相应的接口进行数据传输, 达到数据共享的目的。

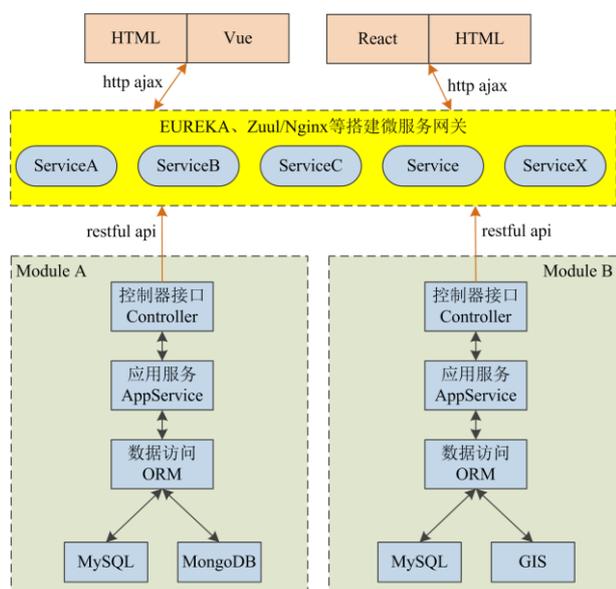


图2 模块化开发

Fig. 2 Modular development illustration

1.3 数据采集接口设计

数据采集接口方面, 其设计原理如图3所示。具体而言, 通过搭建数据采集管理平台, 对各个类型的接口进行管理以及对运行状态进行监控, 据此进一步实施接口的启停管理和测点的配置管理, 以满足支持多种数据类型的功能。本数据采集平台支持有业务系统数据库支撑的结构化与电子报表制成的半结构化等类型的数据源, 且同时支持定制 TCP 协议以及 IEC101/102/103/104 规约等多种采集协议。

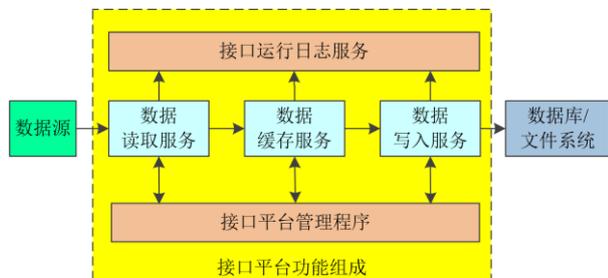


图3 数据采集接口设计原理

Fig. 3 Data acquisition interface design principle

此外, 本系统通过建立数据仓库、融合多源数据及利用数据挖掘形成相应的报表, 提供实时查询、CIM 解析等功能。

1.4 数据的储存及安全

为保证满足多类型数据储存的要求, 且具有大数据量、高容错率的性能, 系统平台构建了多种数据库以支持多种数据类型的存储, 具体包括列式库、NoSql 库以及内存库。同时, 平台支持数据库使用便捷可靠的扩展多节点副本集部署, 满足 PB 级别的海量数据存储, 其中主节点功能设定为完成所有节点的读写, 而其余的副本节点则负责完成数据同步备份的工作, 如图4所示。副本集的副本节点利用心跳机制对主节点状态进行实时监测, 判断其是否发生故障。在检测到故障发生时, 集群内将会发起主节点的选举机制, 选出一个新的主节点, 实现自动切换, 完成了高容错率的实现。

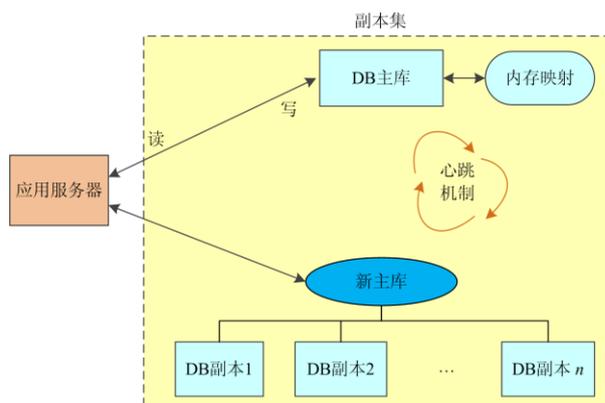


图4 副本集部署模式

Fig. 4 Replica set deployment mode

而安全主要由数据和系统两个方面组成。对于数据的安全, 本系统建立了长期有效的数据备份与安全管理机制, 实现实时数据保护、历史数据保护以及异地数据保护, 如图5所示。

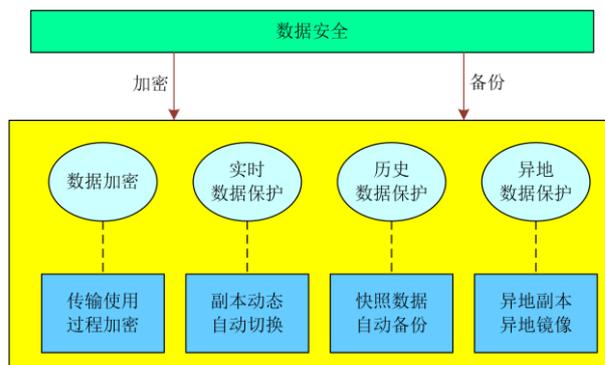


图5 数据安全保护

Fig. 5 Data security protection

在系统安全方面，系统根据冗余热备特性，保证了对外提供服务的应用稳定运行，并采用安全防护技术构成 Web 防护，对 XSS 跨站脚本攻击和 SQL 脚本注入等入侵进行防护。此外，通过为工业互联网平台提供一套认证授权系统，进一步加强保护。

1.5 高效数据动态刷新技术

如图 6 所示，设计了数据动态刷新流程，可提高组态页面响应时的网络传输速度，从而提升系统的数据加载性能。具体而言，对系统做出了以下三个方面的优化。

(1) 采用微软 SignalR 库基于 Web Socket 双工通信协议，由服务端向客户端推送变化数据，提高实时数据相应能力。

(2) 组态页面支持多个轮询间隔，避免每次都请求低频变化的数据。

(3) 服务端与页面建立对应关系，由服务端完成页面实时数据差异化比较，轮询请求数据时只返回增量差异数据。

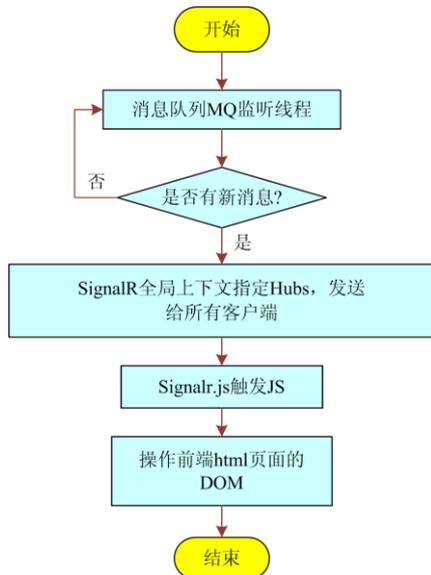


图 6 数据动态刷新流程图
Fig. 6 Dynamic data refresh flow chart

1.6 算法集成设计

同时，为支持包括 Java、Python、C#算法的多语言算法，本系统建立了算法集成平台，如图 7 所示。该平台架构提供标准的统一调用接口、http 请求以及数据交互功能，支持自动识别和动态加载算法库，并以微服务架构集成，利用 Zuul 网关代理拦截，实现算法访问权限控制。

在具体的底层算法 API 实现中，本系统考虑到数据结构规范、代码简洁直观等方面，目前选择使用 Python 语言进行开发编写。

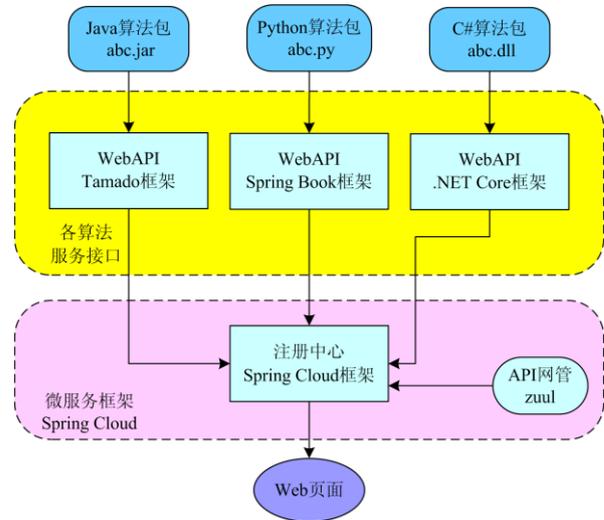


图 7 算法集成架构设计
Fig. 7 Algorithm integration architecture design

2 系统多维度用户负荷特性分析方法

近年来，负荷增长、产业转型和经济结构调整对电网的负荷特性造成了多方面的影响。面对现今电网负荷的复杂情况，本文拓展负荷特性研究思路，从多个角度充分挖掘负荷特征，并针对业扩报装、负荷预测等方面给出更准确和全面的指导。为此，本文借助上述仿真系统，建立单个用户的负荷特征库，并可将其扩展用于对各行业用户的负荷特性进行用电模式研究。过程如图 8 所示。

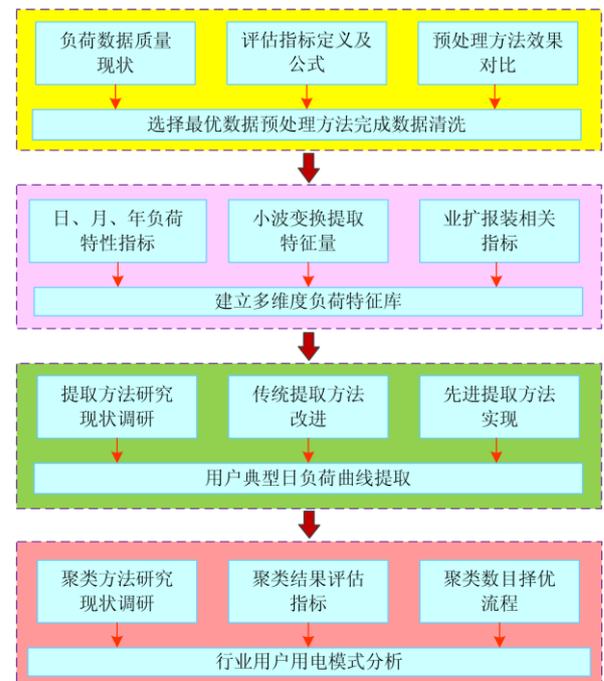


图 8 多维度负荷特性分析流程
Fig. 8 Multi-dimensional load characteristic analysis process

具体步骤如下:

(1) 调用数据预处理 API, 对从电网计量系统中获取的数据进行分析, 筛选出异常数据并对其进行相应的处理。

(2) 调用负荷特性指标计算 API, 根据各行业用户的典型负荷曲线, 提取出相应的特征指标。

(3) 调用日负荷曲线提取方法 API, 根据需求及应用场景选择方法, 提取单个用户的典型曲线。

(4) 调用聚类算法 API, 对行业的多个用户进行聚类, 分析其行业用户的典型用电模式。

应用该模块对广东省某区域的纺织业用户进行具体分析, 验证了本系统的合理性与有效性。

2.1 负荷特性数据预处理

数据预处理的方法多种多样, 考虑到个案剔除、均值代替等简单处理方法会造成负荷数据信息和资源的浪费, 甚至导致错误的结论, 本系统的数据预处理 API 采用了线性插值法。该方法利用同一天各时点前后具有较强的连续性和自相关性的数据恢复算法, 其假设少量连续几个负荷数据点呈现出线性的变化规律。计算缺失数据前后完整数值的平均值, 并以此补齐所缺少的负荷。根据数据发生缺失的具体位置的不同, 可以分为首末位缺失和中间缺失两种方式。缺失方式的不同, 采用的修补方法也存在一定的差异。对前者而言, 将选用离缺失值最近的完整数据作为修复结果, 计算公式为

$$a_1 = a_2 = \dots = a_s \quad (1)$$

$$a_N = a_{N-1} = \dots = a_e \quad (2)$$

式中: a_s 是日负荷曲线中离首位最近的非缺失值; a_e 是离末位最近的非缺失值; N 为单条负荷数据的数据维度。

至于后者, 若为单个缺失值, 则对前后完整数值取平均后作为其修复值; 若是连续多个点的缺失, 则对未缺失的前后数据点求线性表达式, 根据比例把所有的缺失数据点求出, 计算公式如式(3)。

$$a_x = a_m + \frac{a_n - a_m}{n - m}(x - m) \quad (3)$$

式中: a_x 为所求缺失数据; a_m 和 a_n 分别为前后最近的完整数值。

总的来说, 本系统的数据预处理 API 建立了数据预处理方法库。该方法库中的负荷数据预处理方法研究逻辑框图如图 9 所示。

基于图 9, 其中的数据恢复的误差评价方法使用绝对值百分比误差(MAPE), 如果是多点丢失则采用平均绝对值百分比误差, 其定义为

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{P'_i - P_i}{P_i} \right| \times 100\% \quad (4)$$

式中: P' 为算法填充值; P 为真实数据值; n 为缺失的点数。通过比较多种数据丢失场景下, 各种数据预处理方法的 MAPE 指标的中位数可以得出算法整体恢复效果; 通过比较最大值、最小值之差或上下四分位数之差可以得出算法的稳定性。

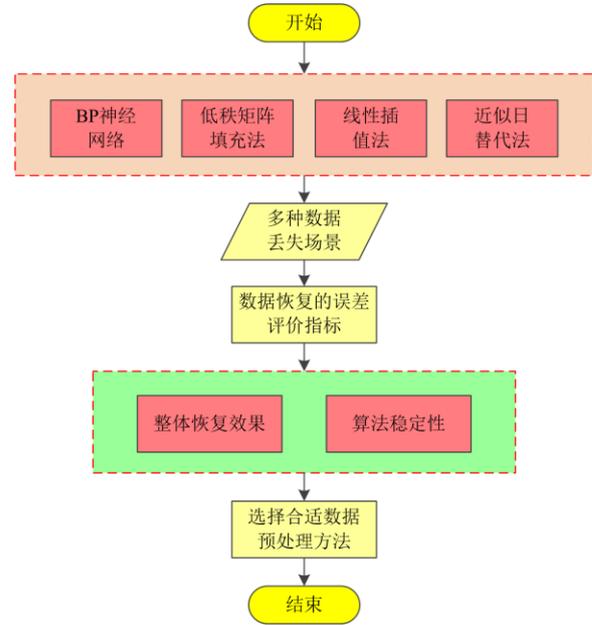


图 9 负荷数据预处理方法研究逻辑框图

Fig. 9 Logical block diagram of load data pre-processing method study

2.2 建立多维度负荷特征

一般来说, 常规负荷特性指标^[15-16]按照日、月、年三个时间跨度进行选取和计算, 对负荷的时序特征进行了描述。但上述的时域分析结果并不能很好地展现实际负荷的波动变化。为此, 本系统额外引入了小波分析方法展开负荷频域方面的分析研究。另外, 饱和负荷密度、实用系数及阶段系数这三种指标则反映了负荷在报装后的释放规律, 对指导业扩报装工作具有重要意义。对上述负荷特性指标进行选取和计算, 建立多维度负荷特征库, 各特性指标提取方法如下。

1) 日负荷特性指标提取方法

针对日负荷特性指标, 针对不同的报告期限(全年、夏季、冬季), 可以计算不同报告期限的日负荷率 $\gamma_{d,av}$ 、日最小负荷率 $\gamma_{d,min}$ 、日峰谷差率 ε_d 。计算公式分别为

$$\gamma_{d,av} = \frac{P_{av}}{P_{max}} \times 100\% \quad (5)$$

$$\gamma_{d,\min} = \frac{P_{\min}}{P_{\max}} \times 100\% \quad (6)$$

$$\varepsilon_d = \frac{\varepsilon_{\max}}{P_{\max}} \times 100\% \quad (7)$$

式中： P_{av} 、 P_{max} 、 P_{min} 分别为报告期间的日平均负荷、日最大负荷和日最小负荷。

2) 月负荷特性指标提取方法

针对月负荷特性指标，针对不同的报告期限(不同月份)，可以计算不同报告期限的月平均日负荷率 $\gamma_{m,\text{av}}$ 、月最小负荷率 $\gamma_{m,\min}$ 、月平均日峰谷差率 $\varepsilon_{m,\text{av}}$ 。计算公式分别为

$$\gamma_{m,\text{av}} = \frac{\sum_{i=1}^{N_{\text{month}}} \gamma_{m,i}}{N_{\text{month}}} \quad (8)$$

$$\gamma_{m,\min} = \min\{\gamma_{m,1}, \gamma_{m,2}, \dots, \gamma_{m,N_{\text{month}}}\} \quad (9)$$

$$\varepsilon_{m,\text{av}} = \frac{\sum_{i=1}^{N_{\text{month}}} \varepsilon_{m,i}}{N_{\text{month}}} \quad (10)$$

式中： $\gamma_{m,i}$ 为该市第 i 天的日负荷率； $\varepsilon_{m,i}$ 为该市第 i 天的日峰谷差率； N_{month} 为该月的天数。

3) 年负荷特性指标提取方法

针对年负荷特性指标，可以绘制年最大负荷曲线以及计算年最小负荷率、年最大峰谷差率、季不平衡系数、年平均日峰谷差率。年最大负荷曲线绘制方法为：将各月份负荷最大值按月份顺序用折线连接。其中，年最小负荷率 $\gamma_{y,\min}$ 、年最大峰谷差率 $\varepsilon_{y,\max}$ 、季不平衡系数 ρ 、年平均日峰谷差率 $\varepsilon_{y,\text{av}}$ 的计算如式(11)~式(14)。

$$\gamma_{y,\min} = \min\{\gamma_{y,1}, \gamma_{y,2}, \dots, \gamma_{y,i}, \dots, \gamma_{y,N_{\text{year}}}\} \quad (11)$$

$$\varepsilon_{y,\max} = \max\{\varepsilon_{y,1}, \varepsilon_{y,2}, \dots, \varepsilon_{y,i}, \dots, \varepsilon_{y,N_{\text{year}}}\} \quad (12)$$

$$\rho = \frac{P_{m,\text{av}}}{P_{m,\max}} \times 100\% \quad (13)$$

$$\varepsilon_{y,\text{av}} = \frac{\sum_{i=1}^{N_{\text{year}}} \varepsilon_{y,i}}{N_{\text{year}}} \quad (14)$$

式中： $\gamma_{y,i}$ 为该市第 i 天日负荷率； $\varepsilon_{y,i}$ 为该市第 i 天的日峰谷差率； N_{year} 为该年的天数； $P_{m,\text{av}}$ 为各月最大负荷平均值； $P_{m,\max}$ 为各月最大负荷的最大值。

4) 频域负荷特性指标提取方法

作为一种频域分析中的经典算法，小波变换广泛应用于电力系统分析等众多研究领域^[17]。通过对时间进行精确定位，小波变换分析能够真实地

反映电力负荷在频域维度的详细信息。故而，其对于提高负荷聚类及用户识别整体的准确度，更好地表现负荷波动特性有着极为突出的作用。选用 db3 小波对预处理后的用户负荷数据进行三层小波分解，各层系数的小波能量、均方根值、绝对模平均值和标准差，这些稳态指标的计算公式分别为

$$Ea_i = \sum_{j=1}^{N_1} |a_{i,j}|^2 \quad (15)$$

$$Ed_i = \sum_{j=1}^{N_2} |d_{i,j}|^2 \quad (16)$$

$$\text{rms}_a_i = \sqrt{\frac{1}{N_1} \sum_{j=1}^{N_1} |a_{i,j}|^2} \quad (17)$$

$$\text{rms}_d_i = \sqrt{\frac{1}{N_2} \sum_{j=1}^{N_2} |d_{i,j}|^2} \quad (18)$$

$$ua_i = \frac{1}{N_1} \sum_{j=1}^{N_1} |a_{i,j}| \quad (19)$$

$$ud_i = \frac{1}{N_2} \sum_{j=1}^{N_2} |d_{i,j}| \quad (20)$$

$$\text{std}_a_i = \sqrt{\frac{1}{N_1} \sum_{j=1}^{N_1} (a_{i,j} - \text{mean}_a_i)^2} \quad (21)$$

$$\text{std}_d_i = \sqrt{\frac{1}{N_2} \sum_{j=1}^{N_2} (d_{i,j} - \text{mean}_d_i)^2} \quad (22)$$

式中： $i=1,2,3$ ； $a_{i,j}$ 为第 i 层近似系数的第 j 个模值； $d_{i,j}$ 为第 i 层细节系数的第 j 个模值； N_1 、 N_2 分别为近似系数和细节系数的数据长度； Ea_i 表示第 i 层近似系数的能量值； Ed_i 表示第 i 层细节系数的能量值； rms_a_i 表示第 i 层近似系数的均方根值； rms_d_i 表示第 i 层细节系数的均方根值； ua_i 表示第 i 层近似系数的绝对模平均值； ud_i 表示第 i 层细节系数的绝对模平均值； std_a_i 表示第 i 层近似系数的标准差； std_d_i 表示第 i 层细节系数的标准差； mean_a_i 表示第 i 层近似系数的平均值； mean_d_i 表示第 i 层细节系数的平均值。

5) 饱和和负荷密度、实用系数及阶段系数提取方法

为展现用户用电的真实情况，将用户负荷大小与报装容量大小的比值定义为实用系数。设有 N 个电力用户，则各指标计算公式分别为

$$\rho_i = \frac{P_{i,\max}}{S_i} \quad (23)$$

$$\eta_i = \frac{P_{ij_{\max}}}{P_{ibz}} \quad (24)$$

$$\gamma_{ij} = \frac{\eta_{ij}}{\eta_i} \quad (25)$$

式中: ρ_i 表示用户 i 的饱和负荷密度; η_i 表示用户 i 的实用系数; $P_{i_{\max}}$ 为用户 i 在终期年份的年最大负荷; S_i 为用户 i 的建筑面积; $P_{ij_{\max}}$ 为用户 i 在第 j 年的年最大负荷; P_{ibz} 表示用户 i 的报装容量; η_{ij} 表示用户 i 在第 j 年的阶段系数。

2.3 典型日负荷曲线提取方法

目前, 典型日负荷曲线提取方法没有统一的规定, 传统方法主要包括日负荷率选取法、日最大负荷选取法及固定日选取法^[18]。基于此, 本系统的日负荷曲线提取 API 建立了典型日负荷曲线提取方法库, 具体实现步骤描述如下。

首先, 对传统典型日负荷曲线提取方法进行了调研和总结。目前, 典型日负荷曲线提取方法没有统一的规定, 传统方法主要包括日负荷率选取法、日最大负荷选取法及固定日选取法。基于此, 针对电网负荷数据实际情况、电网规划工作实际需求及上述方法科学性上的不足, 舍去具有一定局限性的固定日选取法; 然后, 对日负荷率选取法进行改进, 纳入更多日负荷特性指标为依据对典型日负荷曲线进行选取, 改进算法称为日负荷特性指标选取法; 对日最大负荷选取法进行保留。除对传统方法的改进, 本系统采用数据挖掘技术, 提出基于时序降维的符号聚合近似方法^[19]和基于非参数核密度估计的典型日负荷曲线提取方法^[20], 对典型日负荷曲线进行更加科学合理的提取。最后, 本系统合理选定评估指标对典型日负荷曲线选取的合理性和表征性进行评价。总的来说, 本系统中的典型日负荷曲线提取方法研究逻辑框图如图 10 所示。

基于图 10, 为了验证上述单个用户日负荷曲线提取算法的有效性, 本系统选定使用关联度计算方法对提取效果进行量化分析, 并最终选取不同数量样本集的负荷数据集进行典型日负荷曲线提取实验。样本集中的每条负荷曲线的数据间隔都是 15 min, 将一天分为 96 个时间断面进行分析。上述关联度计算方法的基本思想是根据曲线间相似程度来判断关联程度。实际而言, 该方法是几种曲线间几何形状的分析比较, 即认为几何形状越接近, 则发展变化态势越接近, 关联程度越大。此方法可以用来比较几种预测模型对应的几条预测曲线与一条实际曲线的拟合程度, 关联度越大, 则说明对应的预测模型越优, 拟合误差也就越小。

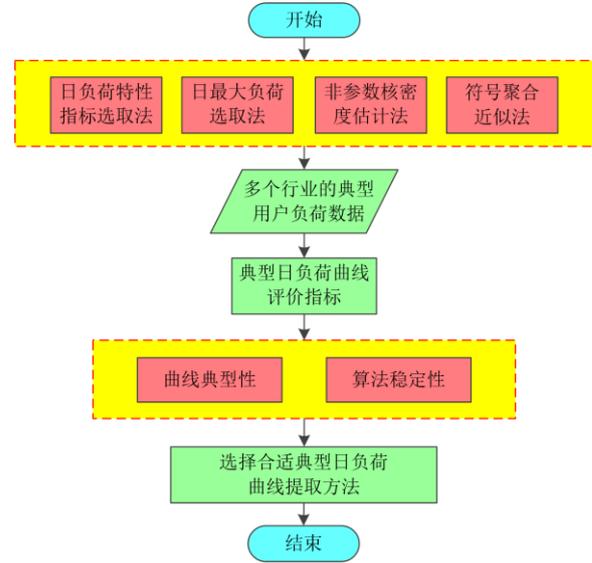


图 10 典型日负荷曲线提取方法研究逻辑框图

Fig. 10 Logical block diagram of typical daily load curve extraction method study

若某电力用户有 n 条日负荷曲线, 例如, $X_i = (x_1, x_2, \dots, x_{96})$ 为第 i 天的日负荷曲线, 提取得到的典型日负荷曲线为 $Y = (y_1, y_2, \dots, y_{96})$ 。则两个时间序列间的相关系数为 r_i , 计算公式为

$$r_i = \frac{\sum_{k=1}^{96} (x_k - \bar{x})(y_k - \bar{y})}{\sqrt{\sum_{k=1}^{96} (x_k - \bar{x})^2 (y_k - \bar{y})^2}} \quad (26)$$

式中, \bar{x} 与 \bar{y} 表示 X_i 与 Y 序列的均值, 负荷特性指标数据样本相关系数 r 是指标总体相关系数 ρ 的一致估计量。 r 取值越接近于 1, 则表明典型日负荷曲线 Y 与第 i 天的日负荷曲线 X_i 相关性越强。

综合用户各天日负荷曲线与典型日负荷曲线 Y 的相关系数, 可得出用户 n 天的日负荷曲线与参考曲线 Y 的关联度 r 为

$$r = \frac{1}{n} \sum_{i=1}^n r_i \quad (27)$$

基于此, 通过比较各种典型日负荷曲线提取方法的关联度 r 指标的中位数, 可以得出算法整体提取效果; 通过比较最大值、最小值之差或上下四分位数之差可以得出算法的稳定性。

2.4 基于行业划分的负荷聚类方法

根据 GB/T 4754-2017《国民经济行业分类》的标准, 利用多种聚类算法对各个行业中的多个用户负荷进行分析, 提取相应的典型负荷曲线。其中, 系统在聚类分析模块的 API 中封装了包括层次聚类算法^[21]、k-means 算法^[22-23]以及 Mini Batch k-means

算法等在内的多种聚类分析方法。本文基于 k-means 方法展开对负荷聚类及典型日负荷曲线提取模块的说明介绍。

k-means 作为聚类分析原型算法中的一种, 现阶段的研究以及应用已十分成熟。具体来看, 这类方法通常先进行初始化原型的工作, 并按照一定规律对所得结果进行多次迭代直至符合收敛条件。

k-means 算法的优点在于简单易操作, 所得结果大体上也可令人信服。然而, 它的缺点也十分突出, k-means 算法分类数目的设定主要取决于主观经验, 难以保证所选的聚类数目即为最优数目。为此, 本系统使用基于聚类有效性评价指标的改进 k-means 方法来处理负荷数据, 从而提取相关的典型日负荷曲线。

为解决上述问题, 本文将设定轮廓系数作为评价聚类结果有效程度的指标, 以判断分类数目的选取是否合理。评价聚类所得结果的优劣, 需要对类内内聚程度与类间的分离程度两个方面进行分析, 轮廓系数则将二者有效地结合在一起。可按式(28)计算得到总的轮廓系数。

$$S_{\text{total}} = \frac{\sum_{i=1}^n \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}}{n} \quad (28)$$

式中: $a(i)$ 是指向量 i 到它属于的簇中其他点的距离的平均值(即内聚度); 而 $b(i)$ 是指向量 i 到与其相邻最近的另一簇中各个点的平均距离(即分离度); n 为聚类对象的总个数。

由此可知, 总的轮廓系数的取值范围为-1 至 1, 且值越大代表内聚度与分离度综合得分越高, 即可认为聚类效果越好。

基于轮廓系数这一评价指标随分类数目变化而发生变化的趋势进行判断, 并依此选出最为合理的分类数目, 从而改善上述 k-means 的缺点。

对某行业多个用户负荷进行聚类分析, 以把握该行业用户整体的用电行为, 上述内容的具体步骤如图 11 所示。

按照上述步骤所得最优聚类数目下的聚类结果对该行业用户用电模式进行划分, 即可对该行业用户负荷划分成适宜的用电模式种类, 其聚类中心即为该用电模式下用户负荷的典型日负荷曲线。

3 系统负荷特性分析可视化界面开发

本文选用部分可视化组件展示多维度负荷特性分析的输出结果^[24-27]。为方便用户操作及阅读, 页面分为四个标签页, 负荷特征库页面展示用户的典

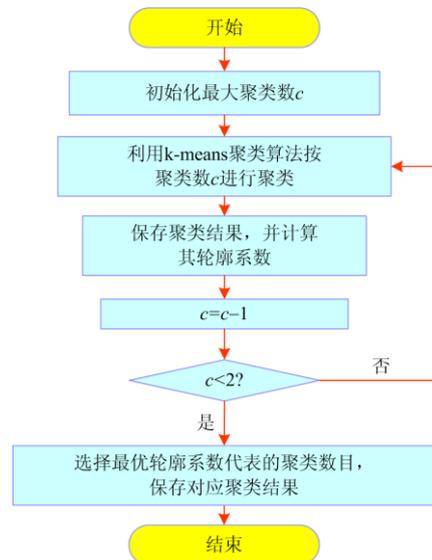


图 11 日负荷曲线聚类流程图

Fig. 11 Typical daily load curve extraction flow chart

型日负荷曲线及日月年等常规负荷特性指标; 负荷频域分析页面展示用户负荷小波变换后所得近似信号、细节信号及各层信号稳态特征量; 实用系数及阶段系数页面展示用户饱和负荷密度、实用系数及阶段系数指标; 聚类分析页面展示所选行业用电模式所得结果。

3.1 负荷特征库页面

本系统中开发的负荷特征库分析页面, 包括典型日负荷曲线和特性指标展示界面、月最大负荷和月最小负荷曲线界面以及年最大负荷曲线界面, 分别如图 12(a)一图 12(c)所示, 采用非参数核密度估计方法作为典型日负荷曲线提取方法, 统计时间为 2017-01-01 00: 00: 00 时刻。

该页面由典型日负荷曲线、日负荷特性指标、月最大负荷曲线、月负荷特性指标、年负荷特性指标 5 个板块构成。上述图表信息根据所选地区、时间跨度、用户所属行业、用户名称、典型日负荷曲线提取方法进行计算及直观展示。

其中, 所选地区、时间跨度及典型日负荷曲线提取方法通过下拉框操作进行选择, 用户所属行业及用户名称提供关键字搜索功能。典型日负荷曲线板块提供游标提示功能, 随鼠标移动显示具体时间及数据信息。

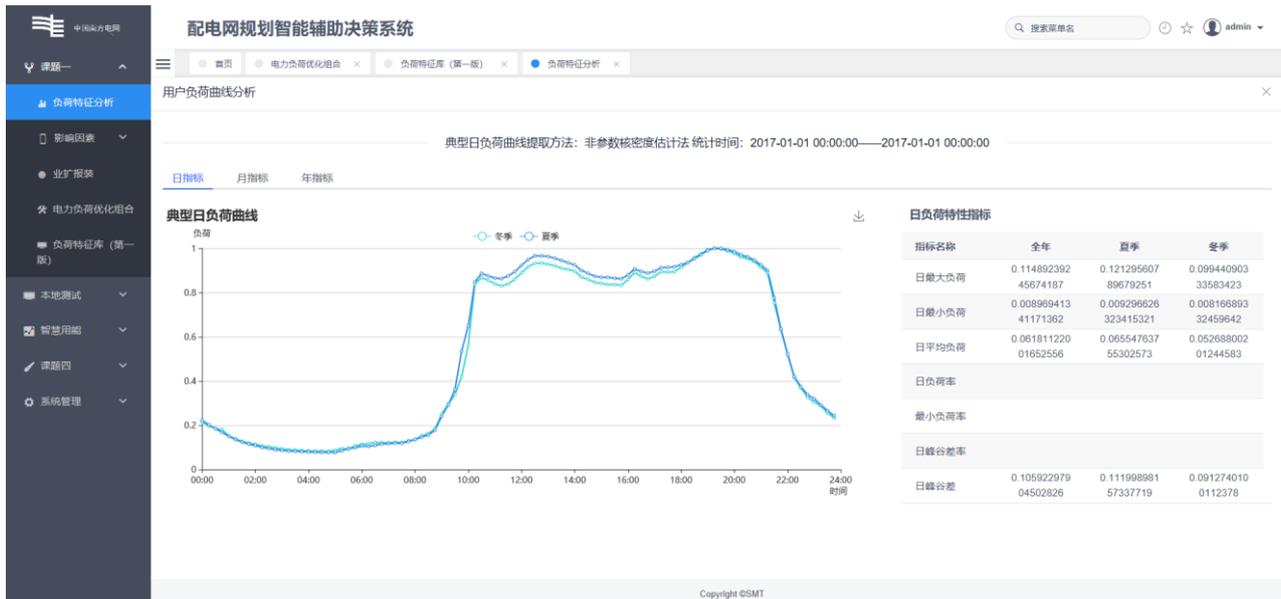
3.2 负荷频域特性分析页面

用户频域分析及特征系数界面展示如图 13 所示。其中, 负荷频域特性分析界面主要由图、表两个板块构成。负荷信号表展示了对日负荷数据进行

3 层小波变换后所得近似信号及细节信号的稳态特征量。负荷信号图对近似信号及细节信号进行绘制, 对多层信号的形状趋势进行直观展示。上述图表信息根据所选地区、时间跨度、用户所属行业、用户名称进行计算及直观展示。

3.3 实用系数及阶段系数页面

本系统中, 该界面主要由实用系数与饱和负荷密度、用户阶段系数和行业阶段系数 3 个板块构成, 对指定用户负荷及所属行业的业扩报装相关指标以表格形式进行了直观展示, 如图 13 所示。



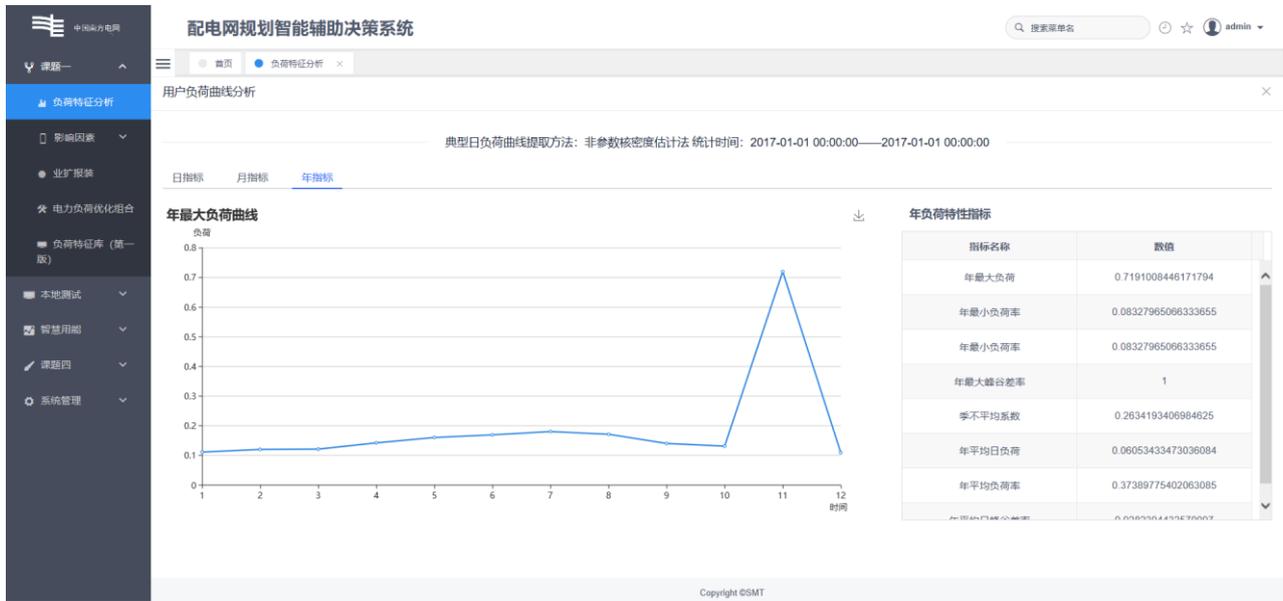
(a) 典型日负荷曲线展示界面



(b) 月最大负荷和月最小负荷曲线展示界面

3.4 聚类分析页面

本系统中, 该界面主要由行业典型日负荷曲线、指定类型负荷特征库 2 个板块构成, 行业典型日负荷曲线板块对行业内多个用户负荷聚类得到的若干用电模式进行绘图展示。通过鼠标触击用电模式图例可对需要分析的用户用电模式进行选择, 得到该类型用户负荷的特征库信息。指定类型负荷特征库的板块构成与上述负荷特征库页面一致。上述图表信息根据所选地区、时间跨度、用户所属行业、负荷聚类方法进行计算及直观展示。



(c) 年最大负荷曲线展示界面

图 12 负荷特征库页面界面展示

Fig. 12 Load feature library page interface display



图 13 用户频域分析及特征系数界面展示

Fig. 13 Load frequency domain analysis and eigenvalue interface display

4 结论

本文通过对现阶段负荷特性软件的需求进行详细分析，搭建了基于 Java 语言，以微服务、微应用为核心架构的多维度负荷特性分析系统。该系统微服务的架构为其带来了灵活可拓展的能力，而模块化开发的方式确保了各个功能模块之间的信息交

互，同时数据采集管理平台的搭建以及数据库采用多节点副本集部署使得系统的数据接口可支持多种数据采集协议，并保证了数据储存的安全性。在算法集成设计方面，系统构建了算法集成平台，提供了标准统一调用接口，支持对多种语言的自动识别及调用。目前，本系统已完成负荷相关因素关联性分析模块的开发，且留有负荷预测模块的相关接口，

如何利用负荷特性分析以及相关因素关联性分析模块所得的量化分析结果, 进一步完成对地区负荷的预测工作, 将是本文下一步研发工作的重点。

参考文献

- [1] 唐良艳. 电力系统负荷特性分析与负荷预测研究[D]. 广州: 华南理工大学, 2010.
TANG Liangyan. Load characteristics analysis and forecasting research of power system[D]. Guangzhou: South China University of Technology, 2010.
- [2] 蔡秋娜, 李嘉龙, 王一, 等. 广东电网负荷特性[J]. 广东电力, 2014, 27(12): 70-75.
CAI Qiuna, LI Jialong, WANG Yi, et al. Load characteristics of Guangdong power grid[J]. Guangdong Electric Power, 2014, 27(12): 70-75.
- [3] 姚星安, 曾智健, 杨威, 等. 广东电力市场结算机制设计与实践[J]. 电力系统保护与控制, 2020, 48(2): 76-85.
YAO Xing'an, ZENG Zhijian, YANG Wei, et al. Electricity market settlement mechanism design and practice in Guangdong[J]. Power System Protection and Control, 2020, 48(2): 76-85.
- [4] 程乐峰. 电力市场多群体策略博弈的长期演化稳定均衡理论研究[D]. 广州: 华南理工大学, 2019.
CHENG Lefeng. Theoretical investigation on long-term evolutionarily stable equilibrium of multi-population strategic games in electricity market[D]. Guangzhou: South China University of Technology, 2019.
- [5] 黄鸣宇, 祁升龙, 芦翔, 等. 面向配网保护的集分联合馈线自动化控制方法[J]. 河南师范大学学报: 自然科学版, 2020, 48(5): 49-54.
HUANG Mingyu, QI Shenglong, LU Xiang, et al. Combined centralized and distributed control method for distribution network protection[J]. Journal of Henan Normal University: Natural Science Edition, 2020, 48(5): 49-54.
- [6] 陈泽雄, 高军伟, 林亚培, 等. 园区综合能源微网动态优化调度算法[J]. 广东电力, 2020, 33(10): 64-74.
CHEN Zexiong, GAO Junwei, LIN Yapei, et al. Mixed integer linear programming algorithm for dynamic optimal scheduling of integrated energy campus microgrid[J]. Guangdong Electric Power, 2020, 33(10): 64-74.
- [7] 郭国太. 居民区用电负荷特性研究与模型预测[J]. 电气技术, 2020, 21(1): 41-46, 53.
GUO Guotai. Study on electricity load characteristic and model prediction in residential areas[J]. Electrical Engineering, 2020, 21(1): 41-46, 53.
- [8] 程乐峰, 余涛, 张孝顺, 等. 机器学习在能源与电力系统领域的应用和展望[J]. 电力系统自动化, 2019, 43(1): 15-43.
CHENG Lefeng, YU Tao, ZHANG Xiaoshun, et al. Machine learning for energy and electric power systems: State of the art and prospects[J]. Automation of Electric Power Systems, 2019, 43(1): 15-43.
- [9] 严洪丽, 张秀钊. 云南分产业典型负荷特性分析[J]. 云南电力技术, 2016, 44(5): 107-108, 114.
YAN Hongli, ZHANG Xiuzhao. Analysis on industries load characteristic of Yunnan province[J]. Yunnan Electric Power, 2016, 44(5): 107-108, 114.
- [10] 李光辉, 陈浩, 陈远扬, 等. 湖南电网负荷率特性分析[J]. 湖南电力, 2016, 36(5): 58-60.
LI Guanghui, CHEN Hao, CHEN Yuanyang, et al. Analysis of the load rate characteristics of Hunan power grid[J]. Hunan Electric Power, 2016, 36(5): 58-60.
- [11] 云飞. 唐山市电力用户负荷特性分析及应用[D]. 保定: 华北电力大学, 2018.
YUN Fei. Characteristics analysis and application for power consumer in cities of Tangshan[D]. Baoding: North China Electric Power University, 2018.
- [12] 康田园, 尹淑萍, 王现法, 等. 大型城市电网负荷特性及其影响因素分析[J]. 电测与仪表, 2016, 53(6): 51-56.
KANG Tianyuan, YIN Shuping, WANG Xianfa, et al. Load characteristics of large city power grid and analysis on its influencing factors[J]. Electrical Measurement & Instrumentation, 2016, 53(6): 51-56.
- [13] 陈建海, 陈淼, 浦云明. 基于微服务架构 B/S 系统的性能分析[J]. 计算机系统应用, 2020, 29(2): 233-237.
CHEN Jianhai, CHEN Miao, PU Yunming. B/S system performance analysis based on microservices architecture[J]. Computer Systems & Applications, 2020, 29(2): 233-237.
- [14] 郝小明, 张建君. 基于微服务架构的终端服务平台设计[J]. 数字技术与应用, 2020, 38(6): 124-126.
QIE Xiaoming, ZHANG Jianjun. Design of terminal service platform based on microservice architecture[J]. Digital Technology & Application, 2020, 38(6): 124-126.
- [15] 海世芳, 孙士云, 范小红, 等. 计及频率特性的感应电动机参数变化对负荷特性的影响[J]. 电力系统保护与控制, 2019, 47(8): 17-25.
HAI Shifang, SUN Shiyun, FAN Xiaohong, et al. Influence of induction motor parameter change considering frequency characteristics on load characteristics[J]. Power System Protection and Control, 2019, 47(8): 17-25.
- [16] 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.

- [17] ZHAO W J, SHANG L Q, SUN J F. Power quality disturbance classification based on time-frequency domain multifeature and decision tree[J]. Protection and Control of Modern Power Systems, 2019, 4(4): 337-342. DOI: 10.1186/s41601-019-0139-z.
- [18] 徐邦恩, 蔺红. 基于改进模糊聚类的典型日负荷曲线选取方法[J]. 电测与仪表, 2019, 56(4): 21-26.
XU Bang'en, LIN Hong. Selection method of typical daily load curve based on improved fuzzy clustering[J]. Electrical Measurement & Instrumentation, 2019, 56(4): 21-26.
- [19] 王潇笛, 刘俊勇, 刘友波, 等. 采用自适应分段聚合近似的典型负荷曲线形态聚类算法[J]. 电力系统自动化, 2019, 43(1): 110-121.
WANG Xiaodi, LIU Junyong, LIU Youbo, et al. Shape clustering algorithm of typical load curves based on adaptive piecewise aggregate approximation[J]. Automation of Electric Power Systems, 2019, 43(1): 110-121.
- [20] 高亚静, 孙永健, 杨文海, 等. 基于非参数核密度估计和改进谱多流形聚类的负荷曲线分类研究[J]. 电网技术, 2018, 42(5): 1605-1612.
GAO Yajing, SUN Yongjian, YANG Wenhai, et al. Study on load curve's classification based on nonparametric kernel density estimation and improved spectral multi-manifold clustering[J]. Power System Technology, 2018, 42(5): 1605-1612.
- [21] TSO W W, DEMIRHAN C D, HEUBERGER C F, et al. A hierarchical clustering decomposition algorithm for optimizing renewable power systems with storage[J]. Applied Energy, 2020, 270.
- [22] 刘季昂, 刘友波, 程明畅, 等. 基于分位数半径动态 K-means 的分布式负荷聚类算法[J]. 电力系统保护与控制, 2019, 47(24): 15-22.
LIU Ji'ang, LIU Youbo, CHENG Mingchang, et al. A distributed load clustering algorithm based on quantile radius dynamic K-means[J]. Power System Protection and Control, 2019, 47(24): 15-22.
- [23] 吴亚雄, 高崇, 曹华珍, 等. 基于灰狼优化聚类算法的日负荷曲线聚类分析[J]. 电力系统保护与控制, 2020, 48(6): 68-76.
WU Yaxiong, GAO Chong, CAO Huazhen, et al. Clustering analysis of daily load curves based on GWO algorithm[J]. Power System Protection and Control, 2020, 48(6): 68-76.
- [24] WANG Z Q, WANG H Y, LIN Z J, et al. Modeling of regional electrical heating load characteristics considering user response behavior difference[J]. International Journal of Electrical Power & Energy Systems, 2020, 123.
- [25] XIE Y, ZHANG H, LI C, et al. Development approach of a programmable and open software package for power system frequency response calculation[J]. Protection Control of Modern Power Systems, 2017, 2(2): 189-198. DOI: 10.1186/s41601-017-0045-1.
- [26] 程乐峰, 余涛, 张孝顺, 等. 信息-物理-社会融合的智慧能源调度机器人及其知识自动化: 框架、技术与挑战[J]. 中国电机工程学报, 2018, 38(1): 25-40.
CHENG Lefeng, YU Tao, ZHANG Xiaoshun, et al. Cyber-physical-social systems based smart energy robotic dispatcher and its knowledge automation: framework, techniques and challenges[J]. Proceedings of the CSEE, 2018, 38(1): 25-40.
- [27] 魏培文, 杨磊. 探析物联网技术在智能电网中的应用[J]. 河南师范大学学报: 自然科学版, 2013, 41(6): 138-140, 145.
WEI Peiwen, YANG Lei. Analysis of the internet of things technology in the application of the smart grid[J]. Journal of Henan Normal University: Natural Science Edition, 2013, 41(6): 138-140, 145.

收稿日期: 2020-04-16; 修回日期: 2020-07-27

作者简介:

曹华珍(1974—), 女, 硕士, 高级工程师, 主要从事电力系统规划与运行研究; E-mail: 729875469@qq.com

吴亚雄(1990—), 男, 硕士, 工程师, 主要从事电力系统规划与运行研究; E-mail: 258538790@qq.com

李浩(1991—), 男, 硕士, 助理工程师, 主要从事电力系统规划与运行研究。E-mail: 467588975@qq.com

(编辑 魏小丽)