•信息工程•

DOI:10.15961/j.jsuese.202101248

本刊网刊

基于数据标签的智能电网监控与异常检测

管　黄[1], 谢小川[2*], 胡　琳[3], 尚　鹏[1], 黎　明[2]

(1.国网山东省电力公司, 山东 济南 250001;
2.四川师范大学 计算机科学学院, 四川 成都 610101;
3.国网山东省电力公司建设公司, 山东 济南 250001)

摘　要: 智能电网是电网的智能化系统, 是以输电网、各级电网协调发展为基础的通信信息支撑平台, 是包括输变电、配电与电力调度的各电压等级的信息化、自动化和互动化等为特征的高度一体化系统。智能电网在电网监控的数据传输协议、计算与处理效率、各种信息与网络攻击和数据异常检测等方面仍存在不足。因此, 本文利用物联网和数据标签技术, 提出一种基于数据标签的智能电网监控架构和异常数据检测算法。首先, 针对智能电网数据标签监控与异常检测的框架, 设计了智能电网的监控数据标签与异常检测框架、数据标签化方法和监控大数据任务划分方法; 然后, 对异常检测流程、稀疏化与精简算法和检测算法进行设计, 提出一种基于数据标签的数据精简和异常检测算法; 最后, 对实验和仿真进行设置, 对时序数据维度数、异常数据量的算法准确率和召回率与参比算法进行了仿真与对比实验, 并对不同测试数据量的本文算法与参比算法运行时间进行了对比实验。结果表明: 本文设计的数据标签智能电网监控与异常数据检测算法与参比算法相比, 当时序数据维度数递增时, 其异常数据检测的准确率大于80%, 召回率高于82%; 当异常数据量增加时, 本文算法的异常数据检测准确率和召回率较优越; 比较不同测试数据量的运行时间发现, 本文算法比参比算法的运行时间少2.0～3.0 s。

关键词: 异常检测; 监控; 智能电网; 数据标签

中图分类号: TP277; TP301.6　　**文献标志码**: A　　**文章编号**: 2096-3246(2023)03-0243-12

Outlier Data Detection and Monitoring of Smart Electric Grid Based on Data-tag

GUAN Yi[1], *XIE Xiaochuan*[2*], *HU Lin*[3], *SHANG Peng*[1], *LI Ming*[2]

(1.State Grid Shandong Electric Power Co., Ji'nan 250001, China;
2.College of Computer Sci., Sichuan Normal Univ., Chengdu 610101, China;
3.State Grid Shandong Electric Power Co. Construction Co., Ji'nan 250001, China)

Abstract: Smart grid, the intelligent system for the power grid, can be defined as a communication and information support platform based on the coordinated development of transmission networks and distributed power grid. It is also a highly integrated system characterized by informatization, automation, and interaction of all voltage levels, including power transmission, transformation, distribution, and dispatch. However, the smart grid is facing the following technical challenges, including the data transmission protocol for power grid monitoring, computing/processing efficiency, as well as detections for information, network attacks, and data anomalies. To this end, a novel data tag-based intelligent power grid monitoring architecture and anomaly detection algorithm is proposed using the Internet of Things and data-tagging technologies. The monitoring framework and anomaly detection are firstly designed based on data tagging for the intelligent power grid, including data labeling method, and

收稿日期:2021 – 12 – 22
基金项目:国网山东省电力公司科技项目(2018A–111); 国家自然科学基金项目(61701331)
作者简介:管　黄(1979—), 女, 高级工程师, 硕士. 研究方向: 调度自动化、电力监控系统网络安全技术和电网异常检测等. E-mail: zylunw@163.com
* **通信作者**:谢小川, 讲师, E-mail: xiexch2008@163.com
网络出版时间:2022 – 08 – 19 13 : 36 : 13　　**网络出版地址**:https://kns.cnki.net/kcms/detail/51.1773.TB.20220819.1017.002.html

task categories for monitoring-related big data. Subsequently, the anomaly detection paradigm is designed to support the monitoring framework, in which the sparsification/simplification and detection algorithms are proposed based on data labels. To validate the proposed framework, the experimental and simulation configurations are provided to conduct the result comparison over selective baselines, in terms of the accuracy and recall under different sequential data dimensions, anomaly data amounts, as well as running time for different test data amounts. Experimental results demonstrate that, compared to the baselines, the proposed approach achieves over 80% accuracy and 82% recall with the increase of the sequential data dimensions. In addition, a similar performance improvement in terms of accuracy and recall can also be obtained for different anomaly data amounts. As to the running time, the proposed approach also harvests a reduction of 2.0～3.0 seconds for different test data amounts.

Key words: outlier data detection; monitoring; smart electric grid; data-tag

近年来，智能电网（smart electrical grid, SEG）快速发展，电网承载的业务数据快速增长，业务数据中蕴含大量高价值信息和异常数据未得到挖掘和利用，存在数据量大而信息匮乏的现象。大数据技术的发展和应用，一方面，使得SEG能快速、高效地处理和挖掘业务数据中的高价值信息和异常数据；另一方面，由于SEG数据具有多源异构性、高维度和先验性等，对SEG进行监控时，须挖掘和提取高质量的SEG运行数据，并整合SEG外部环境数据进行有效的监控和异常数据检测，以实现由单一监视向大数据监控发展[1−2]。对SEG基于数据标签的监控和异常数据检测进行研究，旨在减少电网异常发生和提高SEG运行效率。

SEG系统中，由于设备故障和性能下降导致用户数据异常，在设备工况检测、故障监测、设备监控等多方面需要实时数据监控与分析。另外，因SEG数据源类型多、数据量大、数据复杂度高、数据维度和实时性高，使得异常数据检测时开销大、容易漏判错判，进而导致检测的准确性和效率等受到影响，难以满足大区域电网对异常检测精度、效率和实时性等要求。

综上，针对SEG监控存在的数据传输协议的不一致性、业务数据定义的不规范性、数据计算与处理的效率和实时性不足等问题，利用数据标签对数据进行抽象与规范，将各种数据传输协议定义的数据源统一以数据标签化进行规范，提升数据多源汇聚效率和规范性，并对数据标签进行全使用周期的监控；同时，对数据标签的异常检测流程、稀疏化与精简算法和检测算法等进行设计。

1 相关工作

关于SEG监控，研究者们已在提高电网智能化、数据实时性、数据利用率和用户满意度等方面取得较好效果，但仍然存在一些问题和不足[3−4]，主要为：1）传输协议不一致。电网数据产生量巨大，所使用的通信协议有IEC608705−101协议、IEC60870−5−104协议、IEC61850协议和循环远动协议等，这就使得在分层分类存储数据时不规范，无法提高监控主站效率和高速数据互联互通等[5−7]。2）业务数据不一致。缺乏面向全电网的业务数据规范标准，使得全电网在建设和管理数据粒度上存在差异；并且在多源数据描述和表达上，亦具有数据异构性和多样性，严重影响对全电网数据需求的规范描述与表达[8−10]。3）数据处理和计算效率不高。电网监控数据分析未从业务应用角度挖掘和处理累积的各种数据，使得业务数据的处理和计算无法满足毫秒级、微秒级的实时性要求[11−13]。4）电网监控异常数据检测存在不足。SEG监控的异常检测在检测架构[14]、实时数据检测[15]、电网内部的异常行为检测[16]和异常数据检测[17]、电网外部的信息入侵检测[18]和虚假攻击检测[19]等方面存在问题，使得异常数据检测[20]和异常流量数据检测[21]在实时性和准确性等方面仍存在不足之处。

针对智能电网数据异常检测，国内外研究者进行了相应的研究，主要分为基于统计分析和基于机器学习算法的检测方法。其中，基于统计的异常检修方法，通过假设正常数据服从同一概率分布，异常数据的概率分布与正常数据概率分布存在差异而实现异常数据检测；该方法不能检测到局部异常数据，特别是小范围、单点单域的异常无法检测[22]。关于基于机器学习、深度学习方面的异常检测，Jiang等[23]进行了综述，并指出了各种机器学习、深度学习算法在异常数据检测方面的优缺点。Chahla等[24]将聚类与预测结合，提出了一种新的无监督深度学习检测功耗异常数据算法，并对算法进行实验验证；但算法运行时间复杂度未明显提高。在基于机器的异常检测方面，有研究者采用神经网络算法、Markov理论和时序等进行研究，如：Tsukada等[25]利用神经网络进行边缘设备的异常检测研究；Wu等[26]利用双隐马尔可夫模型对无线终端非信任环境下的异常检测进行研究；Cook等[27]综述了物联网时序异常数据检测的研究现状及未来潜在的机会等。如上所述，这些研究能有效应用于各种检测任务，但每种检测方法都有其优缺点，对SEG的异常检测提供了有益参考。

虽然研究者对SEG监控和异常检测进行了诸多

研究，但仍存在难以准确发现监控时的异常数据的问题，特别是在多维异构特性下，其异常数据挖掘难度更大。因此，本文利用数据标签实现SEG监控，在监控中利用异常数据检测算法实现SEG数据标签监控异常数据检测，并提出了相应的流程和算法。

2　监控架构

2.1　系统结构

在提出的SEG数据标签监控系统中，针对各种协议的原始数据输入，首先，对原始数据进行数据源命名、设备信息数据和数据结构等规格化处理；然后，按照规格化后的统一格式传输到数据标签化处理中。数据标签化处理过程中，对各类数据进行标签化处理，如监控运行数据、检修数据和用户数据等。将经过标签化处理的监控数据输入到后端进行数据集标签提取与分析，即进行数据标签语义提取、可交互监控数据标签集提取和数据文本语义提取等，这一过程在SEG监控大数据处理与分析中完成[28]。具体框架如图1所示。

如图1所示，经过标签提取后，将数据发布到计算分析部分，进行数据标签的关联性分析、数据稀疏化、精简关联数据集生成和异常检测等处理，得到SEG数据标签监控的正常数据和异常数据，以便进行各种实时处理和管理决策。

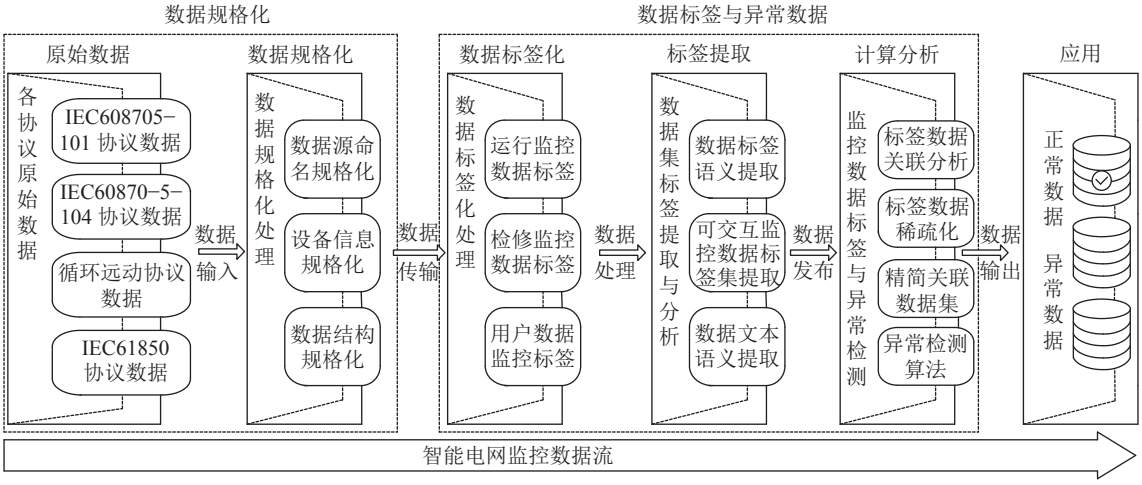


图 1　SEG监控数据标签与异常检测框架

Fig. 1　SEG monitoring data-tag and anomaly detection framework

原始数据来源于5个方面：1）电网原先的监控系统采集而得到的数据；2）现有监控的遥测、遥感、遥调和遥控数据；3）发电端输入电网的输变电在线采集、检测的实时数据，以及二次侧在线录波数据和各种占用及电网设备采集数据与运行数据；4）SEG各种子系统的设备检修、故障、缺陷和台账等基础管理数据；5）各种SEG相关辅助子系统的各种辅助数据，如气象、工作票、雷电监测、视频、地理信息系统等。依据数据源的差异和特征，对图1进行层次化和模块化抽象，得到SEG标签数据监控架构，如图2所示。

图2中，虚线框内的数据层，即图1的规格化数据接入部分，在此仅对其进行了细化，如输电线路组成，即各种输电线路传感器、无线感知设备、设备识别标签、摄像头等感知监控数据，经过规格化处理后，通过传输通道进入下一个数据处理层。在标签层，对规格化数据按照标签化处理，得到感知数据集的数值语义标签和时序文本语义标签；感知数据集的数值语义标签和时序文本语义标签在次层依据关联规则进行关联。服务层包括信息平台标签数据库、各种专

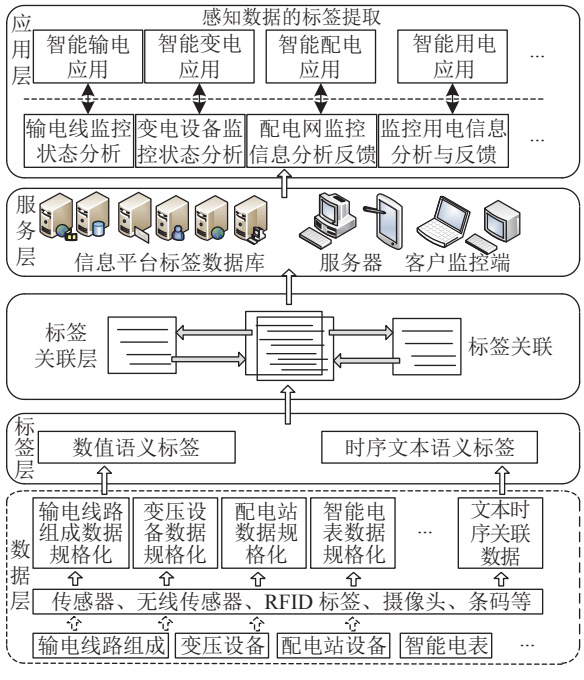


图 2　SEG数据标签监控架构

Fig. 2　SEG data-tag monitoring architecture

业服务器和客户监控终端等，在服务器中设置专用标签提取与分析算法，并将标签提取与分析分析结果输出到应用层。应用层主要进行各种监控数据状态分析及应用，包括监控数据标签提取后的呈现形式、表示方法和结果反馈等。

2.2 数据标签

数据标签采用非手工标签生成方式，即依据业务规则构建标签，主要包括属性标签、事实标签、模型标签和复合标签等。SEG的各种数据，需要建立数据标签化规则，即：对各类设备的标签，建立分类分级、应用场景、组织对象、供应商和所属员工等属性，以便进行标签对属性关联[29]。

SEG运行时各种监控感知数据和运行数据均为单一数值构成的集合，称集合为感知数据集。感知数据集为具有同一属性且在连续时间段内的数值序列，即：

$$pdsObj = \{p_1, p_2 \cdots, p_i, \cdots, p_n\} \tag{1}$$

式中，p_i为在某时刻感知数据集$pdsObj$设备v_i收集的值，$i = 1,2,\cdots,n$，n为感知层设备数量。对$pdsObj$的数据集赋予一个数据对象和数值λ，构成数据集的数值语义标签$numSL$，即：

$$numSL = \{pdsObj, \lambda\} \tag{2}$$

式中，λ为数据对象的影响因子。

感知数据文本与感知数据集语义相关联，感知数据文本$txtObj$是感知数据集对应标签的一份文本数据，即：

$$txtObj = \{w_1, w_2, \cdots, w_i, \cdots, w_n\} \tag{3}$$

式中，w_i为$txtObj$中设备v_i的权重。

主体词库$keyTwl$为主体词的集合，包括语料库中所有感知文本数据的语义标签，用以限定语料库的语义范围，即：

$$keyTwl = \{key_1, key_2, \cdots, key_i, \cdots, key_n\} \tag{4}$$

式中，key_i为主体词库中设备v_i的主体词。

主体词表示文本语义或具有一定辨识度的词或短语，每个主体词对应一个权值的概率，即：

$$Key = \{wv_1, wv_2, \cdots, wv_i, \cdots, wv_n\} \tag{5}$$

式中，wv_i为主体词key_i取权值w_i的概率。

属性词是经由分词标注处理后的词或短语，是实际语义的载体。属性词aW为一个与主体词关联的权值概率序列，即：

$$aW = \{pv_1, pv_2, \cdots, pv_i, \cdots, pv_n\} \tag{6}$$

式中，pv_i为属性词aW的主体词key_i的权值概率。

由此，原始感知数据集由原数值型数据转换为具有标签特性的感知数据文本$sdlObj$，且感知数据集与感知数据文本具有关联关系，即：

$$sdlObj = \{pdsObj, txtObj\} \tag{7}$$

对$sdlObj$添加时间分量，使得感知数据集与感知数据文本关联的同时，具有时序关联特性，即：

$$sdlTObj = \{sdlObj, timeObj\} \tag{8}$$

式中，$timeObj$为感知数据集的时序关联特征值序列，即：

$$timeObj = \{t_1, t_2, \cdots, t_s, \cdots, t_S\} \tag{9}$$

式中，t_s表示感知数据集$pdsObj$在时刻s的时序值。

由此，利用式(1)～(9)，即可建立SEG监控感知数据集的时序文本关联数据标签。

2.3 任务划分

SEG监控状态判断采用潮流计算模型，即用极坐标方式描述数据[30]。其描述的极坐标形式为：

$$\begin{cases} P_\psi - U_\psi \sum\limits_{\zeta \in E_\psi} U_\zeta (G_{\psi\zeta} \cos\ \theta_{\psi\zeta} + B_{\psi\zeta} \sin\ \theta_{\psi\zeta}) = 0, \\ Q_\psi - U_\psi \sum\limits_{\zeta \in E_\psi} U_\zeta (G_{\psi\zeta} \cos\ \theta_{\psi\zeta} - B_{\psi\zeta} \sin\ \theta_{\psi\zeta}) = 0 \end{cases} \tag{10}$$

式中：$\xi \in E_\psi$，E_ψ为和节点ψ相联的节点集合；U_ψ为检测节点ψ的电压幅值；$\theta_{\psi\zeta}$为节点ψ和ζ的相角；P_ψ、Q_ψ为节点ψ的有功和无功功率；$G_{\psi\zeta}$、$B_{\psi\zeta}$为节点ψ和ζ相联的自导电和自电纳。

利用式(10)对SEG节点ψ的运行状态进行判断时，依据P_ψ、Q_ψ、U_ψ、$\theta_{\psi\zeta}$中的任意两个参数值，以及约束条件，求解另外两个参数[31]。若式(10)的节点ψ数目较多，则式(10)的计算量巨大；为了在给定时间内计算得到SEG的状态值，利用潮流计算得到结果，从而满足电网需求。

依据数据标签的SEG监控和式(10)计算量，采用主从模式进行任务划分。假设节点ψ与ζ相联异常，将该异常进行细分，在MapReduce大数据处理中用Map代表若干细分，所有Map为串行潮流计算，即可得到节点ψ运行状态。Map等价于MapReduce中的子任务，即采用该方式对SEG数据标签监控进行任务划分，但SEG的节点ψ与节点ζ相联不一定导致异常，故需要对数据标签SEG监控的异常数据进行映射，并将各映射与异常数据检测算法进行归纳，以输出异常数据检测算法的结果。对图1的计算分析部分，进行基于MapReduce的SEG数据标签任务划分，得到任务划分模型如图3所示。

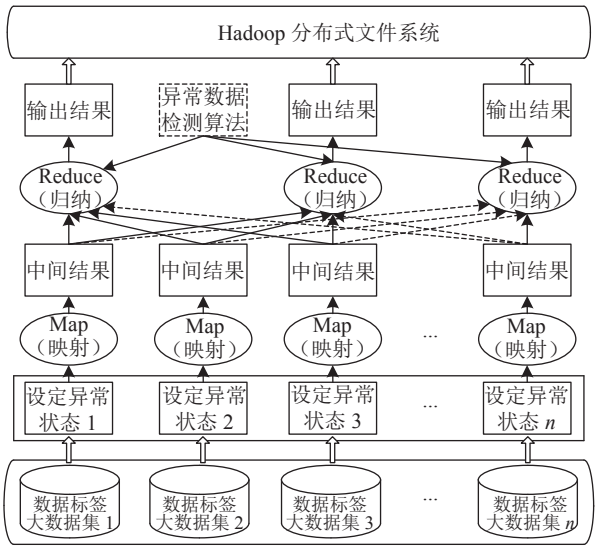


图 3　SEG监控大数据任务划分

Fig. 3　SEG monitoring big data task division

3　异常检测模型

3.1　检测流程

依据第2节论述和SEG数据特征, 以及SEG监控大数据系统对链路延迟、网络吞吐率和存储使用率等对时序的要求, 对数据标签增加时序关联, 以进行数据标签的异常检测[32]。

定义1　设SEG由多个设备节点构成的时序关联无权图$G(V,E)$, 其中: V表示SEG节点集合, $|V|=n$为设备节点数量; E为网络中边的集合。

根据**定义1**, 设$d_i(t)$为设备节点i在t时刻的时序关联数据, 即为数据关联时序, 则:

$$D(t)=\{d_1(t),d_2(t),\cdots,d_i(t),\cdots,d_n(t)\} \tag{11}$$

式中, $D(t)$为SEG中所有节点在t时刻的时序关联数据集。

由此, 将SEG数据标签监控的异常数据检测转化为数据标签化后的时序关联, 用给定时间序列对所有时刻的时序关联数据集D进行时序检测, 以检测是否存在时序异常的数据。因此, 对时序关联数据集D依据如图1所示的计算分析步骤, 即进行时序关联数据生成、时序关联分析、异常数据检测和数据输出, 细化流程如图4所示。

由图4可知, 异常检测时, 先进行时序关联, 并依据关联分析得到的属性进行时序关联稀疏矩阵和时序关联系数矩阵的密度矩阵计算。异常数据检测包括预处理、时序关联计算和异常检测等步骤。其中: 预处理先对SEG数据标签的时序关联数据进行时序对齐和时序缺失填充等处理; 时序关联计算是对是预处理后的数据进行标准和时序关联处理, 生成时序关联矩阵, 并依据矩阵中元素值建立时序关联图, 再对时序关联图按照关联性阈值划分时序关联团; 异常检测将检测结果输出为正常数据和异常数据。

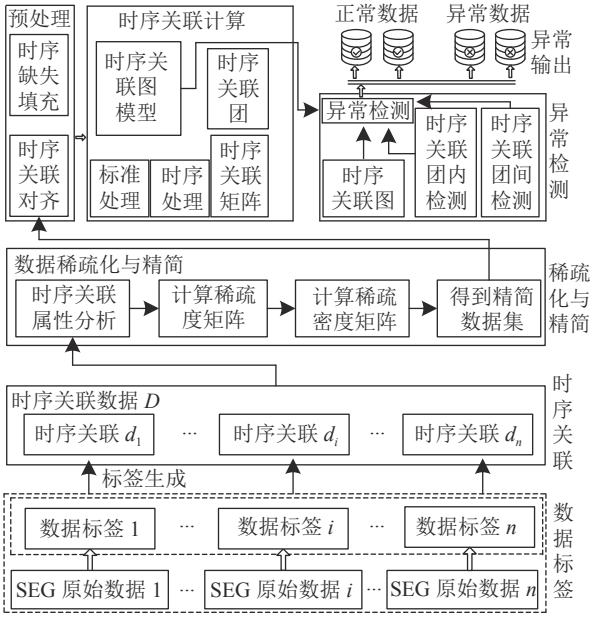


图 4　SEG数据标签异常检测流程图

Fig. 4　SEG data-tag anomaly detection flowchart

3.2　稀疏与精简算法

依据图1和4, 对得到的数据标签时序数据进行稀疏化处理, 得到精简数据集, 并设计稀疏和精简算法[33]。

定义2　设a为指定的一组属性, 则定义A为包含属性a的p维数据集, 即:

$$A=\{A_1^a,A_2^a,\cdots,A_j^a\cdots,A_p^a\} \tag{12}$$

定义3　设d_{ij}为大小为$p\times n$的时序关联数据集D的元素, d_{ij}为设备$v_i(i=1,2,\cdots,n)$在属性$A_j(j=1,2,\cdots,p)$上的取值, d_i为设备节点v_i收集的数据集合。

由**定义2**和**定义3**可知, 若给定A_j上对v_i的d_{ij}值, 得到d_{ij}稀疏度SMD_{ij}为:

$$SMD_{ij}=\sum_{x\in y(d_{ij})}\frac{(x-cv_i^j)^2}{k+1} \tag{13}$$

式中: x为d_{ij}的k近邻集中的设备节点的随机感知值; $y(d_{ij})$为d_{ij}的k近邻集, $y(d_{ij})=Knns(d_{ij})$, $|y(d_{ij})|=k+1$, $Knns(d_{ij})$为点v_i在维度为A_j上的k阶近邻集; cv_i^j为$y(d_{ij})$的中心值, 其表达式为:

$$cv_i^j=\sum_{x\in y(d_{ij})}\frac{x}{k+1} \tag{14}$$

若SMD_{ij}的值比较大, 则d_{ij}在稀疏区域下; 否则, d_{ij}在稠密区域中, 且位于时序关联数据集D的每个

维度。

又因k阶近邻使时序关联数据集为D的数据量急剧增加，为降低数据量，设计算法1生成精简时序关联数据集。

算法1　时序关联数据稀疏与精简算法

//输入：

//　　时序关联数据集：D

//　　数据维度：p

//　　稀疏度阈值：ξ

//输出：

//　　精简时序关联数据集：$Cdsv$，大小为$p\times n$

start:

　设置初值：$n\leftarrow|D|$;

　for $j\leftarrow$1 to p

　begin

　　$D_s\leftarrow$由小到大排列D;

　　for $i\leftarrow$1 to n

　　beginfor

　　　$y(d_{ij})\leftarrow Knns(d_{ij})$;

　　　$b_{ij}\leftarrow$ 按式(13)计算SMD_{ij};

　　endfor

　endfor

　for $i\leftarrow$1 to n

　beginfor

　　for $j\leftarrow$0 to p

　　beginfor

　　　if ($b_{ij}<\xi$)

　　　$c_{ij}\leftarrow$1;//稀疏密度矩阵元素赋值

　　　else $c_{ij}\leftarrow$0;

　　endfor

　endfor

　$Naa\leftarrow\phi$;//Naa为无关联属性集合

　for $i\leftarrow$1 to n

　beginfor

　　if$\left(\sum_{i=1}^{n}c_{ij}==n\right)$

　　$Naa\leftarrow Naa\cup D_s$;

　　else $Cdsv\leftarrow Cdsv\cup D_s$;

　endfor

　$Nadp\leftarrow\phi$;//Nadp无关联数据点集合

　for $j\leftarrow$1 to p

　beginfor

　　if$\left(\sum_{j=1}^{p}c_{ij}i\times j==p\right)$

　　$Nadp\leftarrow Nadp\cup D_s$;

　endfor

　return $Cdsv\leftarrow Cdsv-Nadp$;

endstart

算法1中：首先，对D中数据按照每个数据的维度值进行升序排列，并按照1维k近邻在排序后的数据集D_s中搜索，依据式(13)计算D中的每个数据的稀疏度SMD_{ij}，且将稀疏度SMD_{ij}存储于稀疏矩阵$\boldsymbol{B}$的元素b_{ij}中。然后，在给定稀疏度阈值ξ下，对$\boldsymbol{B}$进行稀疏区和稠密区的识别，得到稀疏密度矩阵$\boldsymbol{C}$，c_{ij}为$\boldsymbol{C}$的元素。若$b_{ij}<\xi$，则$c_{ij}=1$，即d_{ij}位于稠密区；否则，$c_{ij}=0$，即d_{ij}在稀疏区。最后，对$\boldsymbol{C}$中的值进行剪枝，得到精简数据集$Cdsv$。

3.3　检测算法

对$Cdsv$数据集进行逆向逐段聚集均值处理，以使数据集进一步减少而降低运算复杂度，且有利于对各时序数据集的属性进行提取[34]，设h_i为节点v_i在数据集$Cdsv$中对应的数据点。

定义4　设数据点h_i的最近邻节点h_m总数为K，$m=1,2,\cdots,K$。h_i的扩展邻域$Ekd(h_i)$包括3类节点的集合，与3类节点关系如下：

$$Ekd(h_i)=knn(h_i)\cup krnn(h_i)\cup ksnn(h_i) \tag{15}$$

式中，$Ekd(h_i)$为h_i所有近邻点的集合，$knn(h_i)$ 为h_i的k正则最近邻点的集合，$krnn(h_i)$ 为与h_i的k–共享最近邻点个数为0的点的集合，$ksnn(h_i)$为与h_i的k–共享最近邻点个数为1个或多个的点的集合。

定义5　数据点v_i的k近邻扩展可用带控制因子的高斯核函数$kEdf(h_i)$描述，为：

$$kEdf(h_i)=\tau\sum_{h_m\in Ekd(h_i)}\exp\left(\frac{-\lambda(h_i,h_m)}{\sigma}\right) \tag{16}$$

式中：$\tau\in(0,\ 1]$为常量控制因子；σ为点h_i与k近邻点之间的欧几里得距离的逆向逐段聚集均值，$\sigma=\frac{1}{n}\sum_{h_m\in Ekd(h_i)}Eud(h_i,knn(h_i))$；$\lambda(h_i,h_m)$为中间变量，可表示为：

$$\lambda(h_i,h_m)=\begin{cases}\min\limits_{h_m\in Ekd(h_i)}Eud(h_i,h_m),Rpaa(h_i)<Rpaa(h_m);\\ \max\limits_{h_m\in Ekd(h_i)}Eud(h_i,h_m),\text{其他}\end{cases} \tag{17}$$

式中，$Rpaa(h_i)$为h_i的滑动平均分段聚合近似值，$Eud(h_i,h_m)$为h_i、h_m的欧几里得距离。

根据**定义4**和**定义5**，h_i密度与局部k近邻间的差异度$Dlknd(h_i)$为：

$$Dlknd(v_i)=\frac{\sum\limits_{h_m\in Ekd(h_i)}kEdf(h_m)}{kEdf(h_i)\cdot|Ekd(h_i)|} \tag{18}$$

定义1对$G(V,E)$进行时序关联初始化，即$G_\tau=(V,E)$；若$Dlknd(h_i)\geq\alpha$，α为设定的关联差异度值，则

对$G_r=(V,E)$的所有V顶点进行遍历; 对遍历所得到的图再进行处理, 其处理流程为:

1)设置时序关联差异度α, 用算法1对$G_r=(V,E)$每个顶点初始化, 得到稀疏度矩阵$\boldsymbol{B}$。

2)i从1到n, l从$i+1$到n, 若$Dlknd(h_i)\geq\alpha$, 则G_r中加入边$e_{il}=(v_i,v_l)$。

3)得到G_r即为时序关联图。

定义6　对时序关联图$G_r=(V,E)$, 假设V_{tag}为**定义1**确定的顶点集合, 即$V_{tag}=\{v_1,v_2,\cdots,v_n\}$, 当$n\geq 2$时, $dov(v_i)$为顶点的度, $woe(e_{il})$为边e_{il}的权值。若V_{tag}满足:

1)$\forall v_i\in V_{tag}$, 有$v_i\in V(G_r)$, $V(G_r)$为图G_r的顶点集合;

2)$\forall v_i\in V_{tag}$, 有$dov(v_i)\geq|V_{tag}|/2$;

3)给定α, $\forall v_i$、$v_l\in V_{tag}$, 有$woe(e_{il})\geq\alpha$;

4)V_{tag}为G_r上满足条件1)、条件2)的最大顶点集。

则称V_{tag}为时序关联图G_r上的一个时序关联团。

若时序关联图G_r可分解为若干时序关联团V_{tag}^q, 则所有时序关联团V_{tag}^q的顶点之和等于G_r的顶点数, 且$\forall V_{tag}^q$、$V_{tag}^g\in G_r$, $V_{tag}^q\cap V_{tag}^g=\phi$, q和g为关联团标号。

定义7　设V_{tag}^q、V_{tag}^g为时序关联图G_r的相邻两个时序关联团, V_{tag}^q、V_{tag}^g存在一定相互影响, 则V_{tag}^q、V_{tag}^g满足:

1)存在常量设置控制因子$\tau\in(0,1]$;

2)$Vic(V_{rtag}^q,V_{tag}^g)$为$V_{tag}^q$对$V_{tag}^g$的影响程度系数:

$$Vic(V_{tag}^q,V_{tag}^g)=\begin{cases}\tau^{g-q}\cdot V_{tag}^q, q\leq g;\\ 0, 其他\end{cases}\qquad(19)$$

时序关联图G_r的时序关联团V_{tag}^q、V_{tag}^g间的关联关系可用$Vic(V_{tag}^q,V_{tag}^g)$进行度量; 而V_{tag}^q内的关联用V_{tag}^q团内各个顶点间的边的权值表示。

依据**定义4**～**定义7**, 利用算法1得到的精简数据集$Cdsv$, 将$Cdsv$作为算法2输入, 得到异常数据集$OutlierDS$。在算法2中: 先将异常数据集$OutlierDS$设置为空集和时序关联图G_r; 然后, 利用算法1输出的精简数据集$Cdsv$进行时序关联图的时序关联团构建, 再进行时序关联团内和团间的异常数据检查; 最后, 输出异常数据集$OutlierDS$。其具体算法过程为:

算法2　异常检测算法

```
//输入:
//精简数据集: Cdsv
//数据维度: p
//差异阈值: α
//时序图: G_r
//常量控制因子: τ
//相关系数阈值: μ
//输出:
//异常数据集: OutlierDS
start:
    OutlierDS ← φ;
    初始化时序关联图G_r;
    for each h_i ∈ Cdsv do//对节点按定义4分类
    beginfor
        knn(h_i) ← {Ekd(h_i), h_i};
        krnn(h_i) ← {Ekd(h_i), h_i};
        ksnn(h_i) ← ∅;
        for each h_i ∈ knn(h_i) do
        beginfor
            krnn(h_i) ← {Ekd(h_i), h_i};
            ksnn(h_i) ← knn(h_i) ∪ krnn(h_i);
        endfor
    endfor
    for each h_i ∈ Cdsv do
    beginfor
      式(18)计算h_i的偏离度;
      Dlknd(h_i) ← sort(Dlknd(h_i),"由大到小");
      //排序
    endfor
    for V_tag^q ∈ G_r do
    beginfor
      if (Dlknd(V_tag^q) ≥ α)&&(dov(V_tag^q) == 0)
        &&(woe(V_tag^q) ≤ 2)
      beginif
        对V_tag^q进行单维度异常检测;
        异常数据加入OutlierDS中;
        continue;
      endif
      if ((V_tag^q是单点时序关联团)&&(Dlknd(V_tag^q) < α))
      then
        V ← V_tag^q;//将V_tag^q添加到G_r
      else
      beginelse
        if Vic(V_tag^q, V_tag^g) < μ then
        beginif
            |V(G_a)| ← max(woe(V_tag^q), woe(V_tag^g));
            //初始化图G_a, 并将其顶点赋值, 用以
            //记录相关团间的异常
            for v_i, v_l ∈ V_tag^q, V_tag^g&&e_il ∈ E(G_a) do
            //E(G_a)为图G_a的边
            beginfor
              if woe(e_il) < α
              beginif
                G_b ← e(v_i, v_l);//G_b记录异常边
```

$G_a \leftarrow v_i, if(dov(v_i)>0)$;
endif
if (G_a==二分图) then
$OutlierDS$ ←求解G_a最小覆盖;
else $OutlierDS$ ←求得G_a异常序列;
$OutlierDS \leftarrow V_{tag}^q, V_{tag}^g \backslash OutlierDS$;
//$V_{tag}^q, V_{tag}^g \backslash OutlierDS$表示时序相关团内
//异常数据点
endfor
endif
endelse
for $V_{tag}^q, V_{tag}^g \in G_r$
if $woe(e_{il}) < \alpha$
$G_r \leftarrow e(V_{tag}^q, V_{tag}^g)$;
endfor
$G \leftarrow G_r$图没有度的团;
if(G_r是二分图)
$OutlierDS$ ←求解G_r的最小覆盖;
else $OutlierDS$ ←求解得到异常数据集;
return $OutlierDS$;
endstart

由算法2可知，其时间消耗于精简时序数据集$Cdsv$的稀疏度计算、时序关联图构建和求解最小覆盖与计算异常时序关联数据。算法2中，求解最小覆盖时使用匈牙利算法，求解和计算异常时序关联数据时使用禁忌搜索算法，在此不再描述。

算法1的时间复杂度为$O(n\times p)$，最坏情况为$O(n^2)$。而算法2给定$Cdsv$数据集按照式（17）计算$Dlknd(v_i)$的时间复杂度为$O(n^2)$，计算时序关联图的时间复杂度为$O(n\times p)$，但求解最小覆盖与计算异常数据的最坏情况时间复杂度为$O(n^3)$。

4 实验与仿真

4.1 实验设置与实例

依据设计的框架，以电压为500～220、220～35 kV的主变电站为实验数据，进行近3个月的数据感知与存储，采集数据频率为每秒记录2次。利用所提出的算法1和算法2，对SEG数据标签异常监控与异常数据检测实验。

如图5所示，在主变电站的500～220 kV主变压器、220～35 kV变压器、电容器组、主变压器抵抗、站用变和各种母线、刀闸等，使用标签和相应的传感器进行监控，采集到变电站监控数据、调度信息、雷电信息和输电监控等数据，将数据迁移到大数据系统中。其中，数据表2 000多张，数据10亿多条，非结构化数据100万个文件以上，总数据量达到30 TB。

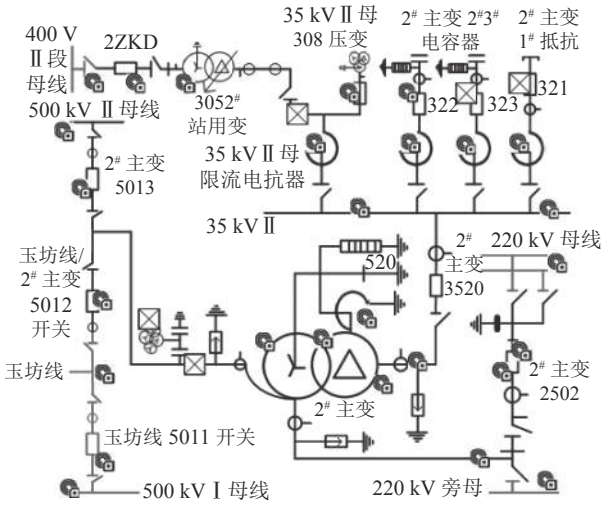


图 5 主变电站监控智能分析结构

Fig. 5 Main substation monitoring intelligent analysis structure

首先，对采集到的原始数据进行标签化处理；然后，依据采集时间进行时序关联；最后，利用本文算法进行异常数据检测。得到异常数据后，对异常数据依据数据标签进行反析定位，得到2#号主变压器保护RC12屏的35 kV电压偏移异常后，须立即进行应急处理。

上级调度依据监控智能分析得到的异常数据，立即通知变电站所，对2#号主变压器异常信息的位置、信息归类归档，现场进行查验核对，检查2#号主变压器保护RC12屏35 kV电压偏移、35 kV母线电压，并做好现场防护与保护。经过细致查验，未发现现场异常，汇报上级调度，申请试分合322、323开关。现场断开322开关，异常信号仍然存在；拉开323开关，异常信号消失；合上322开关后正常，无异常信号。

处理完现场异常立即汇报上级调度，即将电容器组323有接地的情况予以记录并汇报，上级调度将异常汇报给相关决策者，决策者授权于上级调度，并下达对2#主变压器的3#电容器组进行检修的工作票。

4.2 算法仿真与分析

采用Matlab软件进行算法仿真，并结合Spark和Storm大数据分析系统，利用第4.1节采集到的30 TB数据，进行仿真与分析。

待检测数据均为正常数据，通过算法1和算法2检测后，得到异常数据[35]。设定正常数据经检测成为异常数据称为$Nd2od$，异常数据经检测成为异常数据称为$Od2od$，正常数据经检测为正常数据称为$Nd2nd$，异常数据检测为正常数据称为$Od2nd$，算法的准确率Dda和召回率Drr的指标分别表示为：

$$Dda = \frac{Nd2nd}{Nd2nd + Od2nd} \times 100\% \quad (20)$$

$$Drr=\frac{Nd2nd}{Nd2nd+Nd2od}\times 100\% \qquad (21)$$

实验选取30 000个时刻、每个时刻96列数据。首先，对数据进行标签化处理；然后，进行时序关联处理；最后，用本文算法和参比算法，进行实验与仿真对比。其中，参比算法为独立同分布分类异常检测[36]、统计聚类异常检测[37]和机器学习异常检测[38]的算法，亦是异常检测领域较为先进的算法。实验时，对不同时序关联数据量、维度总数和异常数据量下中的算法性能进行对比实验，且设置*Dda*和*Drr*的基准值均为0.8，得到本文算法与参比算法在异常数据检测性能方面的对比曲线。不同数据量下的*Dda*和*Drr*如图6所示。

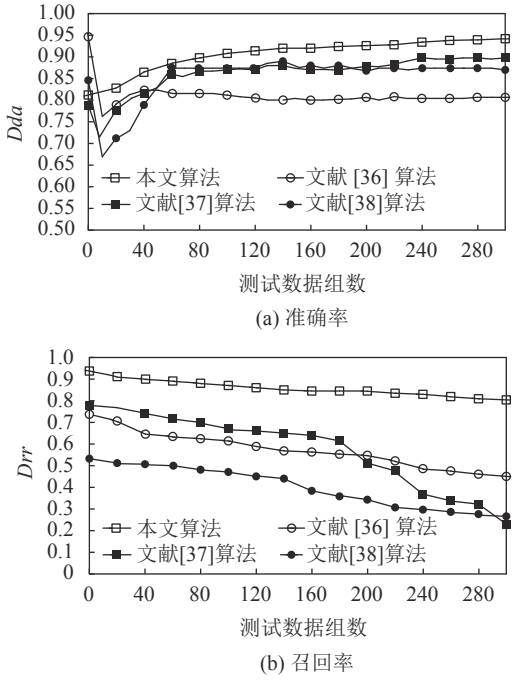


图 6　不同实验数据量下的算法准确率和召回率对比曲线

Fig. 6　Algorithm accuracy and recall comparison curves under different experimental data volume

由图6(a)可以看出：随着测试数据量的不断递增，当数据量在0到20组之间时，文献[36]、文献[37]和文献[38]的算法准确率呈现急剧下降的趋势；然后，文献[38]算法准确率呈现震荡上升，文献[37]算法准确率在上升到一定值后，缓慢下降而趋于平稳。但本文算法，随着数据量递增，其准确率呈现缓慢递增而逐步趋于平稳。由图6(b)可知：当数据量增加时，3个参比算法召回率伴以波动递减；本文算法的召回率总趋势为递减，但递减幅度较参比算法小，且曲线呈缓慢下降，其召回率在数值上维持在大于80%。

图7为数据量相同时，在不同数据的维度*p*下，本文算法与参比算法准确率和召回率曲线的比较。图7(a)中：随着数据维度*p*从3到45不断递增，文献[37]算法准确率在波动中平缓下降，而文献[36]算法准确率在50%到68%之间波动且下降梯度较大；本文算法在维度数较低时，准确率为93%，当维度数增加时，起初呈现下降趋势，但当维度数达到9以后，呈现平稳上升趋势，其上升幅度在5%以内；文献[38]算法表现较为平稳，但其准确率不高，仅在75%到78%之间。图7(b)中，本文算法的召回率在85%到90%之间，且随着维度数的增加而呈现缓慢上升趋势，上升幅度在4%以内；参比算法在数据量增加时，召回率总的趋势为下降。

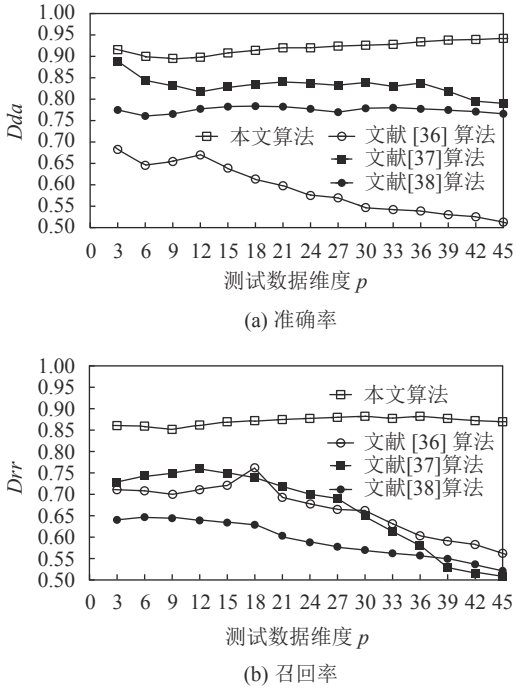


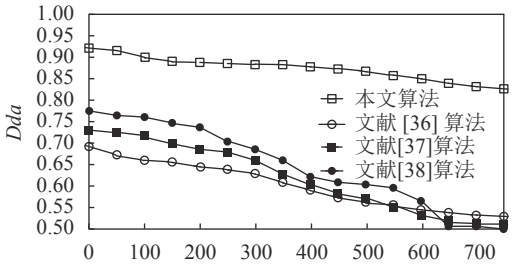
图 7　不同时序数据维度数下的算法准确率和召回率对比曲线

Fig. 7　Algorithm accuracy and recall comparison curves under different time series data dimensions

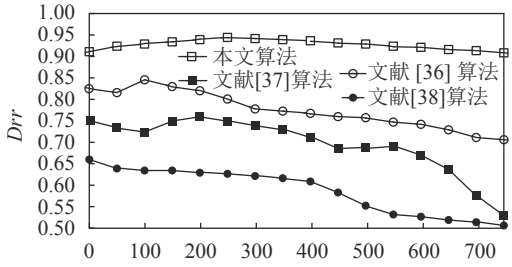
图8为不同测试数据异常数据量下，本文算法和参比算法性能变化曲线。仿真采用的数据集为750组，数据集的维度设置为30。由图8可知：当异常数据量的增加时，本文算法准确率总的趋势表现为递减，但其准确度在仿真时均保持在82%以上；参比算法的准确率递减下降，且下降趋势较为明显。本文算法的召回率先是缓慢增加，然后缓慢下降，其值维持在90%以上；参比算法召回率在数据量增加时，总的趋势呈波动下降。

综上，本文算法与参比算法相比，在准确率和召回率上均具有一定优越性，其原因是本文算法先按照算法1对原始数据进行了稀疏化和精简处理，使得在执行算法2进行异常数据检测时的运算复杂度明显下降。

采用不同数据量的测试数据集进行异常数据检测时，本文算法和参比算法运行时间比较结果如图9所示。



(a) 准确率



(b) 召回率

图 8 不同异常数据量的算法准确率和召回率曲线

Fig. 8 Algorithm precision and recall curves under different abnormal data volume

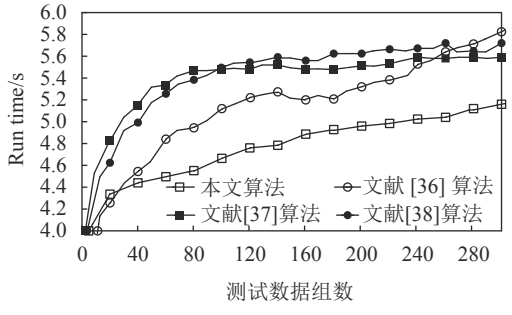


图 9 测试数据集的算法运行时间比较曲线

Fig. 9 Algorithm runtime comparison curves for the test dataset

由图9可知：随着测试数据集数量的增加，本文算法和参比算法的运行时间均增加。在数据集较少时，文献[37]算法运行时间呈线性递增趋势；当数据量达到80以后，其运行时间增长平缓。其他参比算法运行时间呈现波动。本文算法的总体运行时间比参比算法少2.0～3.0 s，这是因为在测试时先按照算法1对数据进行处理，使得算法2的数据量比参比算法少，故其运行时间增加量较少。

通过实验和仿真对比分析可知，本文算法在不同数据源下，准确率维持在80%以上，召回率维持在85%以上。因此，本文算法有效性得到提高；同时，本文算法在相同数据量下的运行时间也较低，与参比算法比较，本文算法执行效率较高。

5 结论与展望

针对智能电网监控在局部传输协议不一致、效率较低和异常数据检测存在不足等问题，提出一种基于数据标签的SEG监控架构和异常数据检测算法。首先，对SEG数据标签监控与异常检测的框架、大数据监控分析架构、数据标签化和系统任务划分等进行设计；然后，提出系统异常检测流程、数据标签稀疏化与精简算法和异常数据检测算法；最后，对所设计的SEG数据标签监控架构进行电网实例分析，同时对提出的算法与参比算法在准确率、召回率和运行时间等方面进行对比仿真实验。可知：本文所设计的SEG数据标签监控系统能有效分析数据异常；与参比算法相比，本文算法在异常数据检测的准确率、召回率和运行时间等方面，有一定的改进和提高。

当然，本文所提出的SEG数据标签监控与异常检测算法仍存在优化空间，提出的算法仍然需要进一步提高准确率和召回率。下一步，将进一步对SEG数据标签监控体系结构进行改进；并对异常检测算法进行深入研究，提出更加有效的算法，以便更好地为电网服务。

参考文献：

[1] Leng Xiwu,Chen Guoping,Bai Jingjie,et al.General design of smart grid monitoring operation big data analysis system[J]. Automation of Electric Power Systems,2018,42(12):160–166.[冷喜武,陈国平,白静洁,等.智能电网监控运行大数据分析系统总体设计[J].电力系统自动化,2018,42(12):160–166.]

[2] Lu Renjie,Liu Ning,Li Di,et al.Intelligent monitoring analysis of power grid monitoring information based on big data mining[J].Journal of Physics(Conference Series),2021,1992(3):032132.

[3] Kadam A P,Ankaliki S G.Robust smart grid monitoring network based on direct sequence spread spectrum intelligence[J].SN Computer Science,2020,1(2):1–9.

[4] Majeed Butt O,Zulqarnain M,Majeed Butt T.Recent advancement in smart grid technology:Future prospects in the electrical power network[J].Ain Shams Engineering Journal,2021,12(1):687–695.

[5] He Zhengyou,Xiang Yueping,Liao Kai,et al.Demand,form and key technologies of integrated development of energy–transport–information networks[J].Automation of Electric Power Systems,2021,45(16):73–86.[何正友,向悦萍,廖凯,等.能源–交通–信息三网融合发展的需求、形态及关键技术[J].电力系统自动化,2021,45(16):73–86.]

[6] Cui Jindong,Wang Shengwen,Xin Yechun.Research on technical framework of smart grid data management from consortium blockchain perspective[J].Proceedings of the

CSEE,2020,40(3):836–848.[崔金栋,王胜文,辛业春.区块联盟链视角下智能电网数据管理技术框架研究[J].中国电机工程学报,2020,40(3):836–848.]

[7] Chen Haibiao,Huang Shengyong,Cai Jierui.Trust evaluation protocol for cross-layer routing based on smart grid[J].Computer Science,2021,48(Supp1):491–497.[陈海彪,黄声勇,蔡洁锐.一个基于智能电网的跨层路由的信任评估协议[J].计算机科学,2021,48(增刊1):491–497.]

[8] Chen Xiaoyan,Liang Wei,Zhou Xinlian,et al.An efficient transmission algorithm for power grid data suitable for autonomous multi-robot systems[J].Information Sciences,2021,572:543–557.

[9] Panda D K,Das S.Smart grid architecture model for control,optimization and data analytics of future power networks with more renewable energy[J].Journal of Cleaner Production,2021,301:126877.

[10] Liu Lei,Tan Yanghong,Jin Jiayao,et al.Key node identification of power communication network[J].Proceedings of the CSU–EPSA,2020,32(2):28–34.[刘垒,谭阳红,金家瑶,等.电力通信网的关键节点辨识[J].电力系统及其自动化学报,2020,32(2):28–34.]

[11] Xu Ning,Wang Yanqin,Dong Zhen,et al.Research on distribution system big data preprocessing technology based on apache spark[J].Journal of North China Electric Power University(Natural Science Edition),2021,48(2):40–46.[徐宁,王艳芹,董祯,等.基于Apache Spark的配电网大数据预处理技术研究[J].华北电力大学学报(自然科学版),2021,48(2):40–46.]

[12] Li Jia,Xu Shengchao.Smart power system big data processing platform in cloud environments[J].Computer Engineering and Design,2018,39(10):3073–3079.[李佳,徐胜超.基于云计算的智能电网大数据处理平台[J].计算机工程与设计,2018,39(10):3073–3079.]

[13] Liu Yonghui,Zhang Xian,Sun Hongyan,et al.Discussion on application of big data in electricity market in background of energy internet[J].Automation of Electric Power Systems,2021,45(11):1–10.[刘永辉,张显,孙鸿雁,等.能源互联网背景下电力市场大数据应用探讨[J].电力系统自动化,2021,45(11):1–10.]

[14] Zheng Fengming,Li Shufang,Guo Zhimin,et al.Anomaly detection in smart grid based on encoder-decoder framework with recurrent neural network[J].The Journal of China Universities of Posts and Telecommunications,2017,24(6):67–73.

[15] Moghaddass R,Wang Jianhui.A hierarchical framework for smart grid anomaly detection using large-scale smart meter data[J].IEEE Transactions on Smart Grid,2018,9(6):5820–5830.

[16] Bao Haiyong,Lu Rongxing,Li Beibei,et al.BLITHE:Behavior rule-based insider threat detection for smart grid[J].IEEE Internet of Things Journal,2016,3(2):190–205.

[17] Liu Shengyuan,Zhao Yuxuan,Lin Zhenzhi,et al.Data-driven event detection of power systems based on unequal-interval reduction of PMU data and local outlier factor[J].IEEE Transactions on Smart Grid,2020,11(2):1630–1643.

[18] Xia Zhuoqun,Tan Jingjing,Gu Ke,et al.Detection resource allocation scheme for two-layer cooperative IDSs in smart grids[J].Journal of Parallel and Distributed Computing,2021,147:236–247.

[19] Xia Zhuoqun,Long Gaohang,Yin Bo.Confidence-aware collaborative detection mechanism for false data attacks in smart grids[J].Soft Computing,2021,25(7):5607–5618.

[20] Zhou Boyang,Guo Zhimin,Wang Yansong,et al.An anomaly traffic detection method using multi-resolution low rank model for wireless access network of electric power grids[J].Acta Electronica Sinica,2020,48(8):1552–1557.[周伯阳,郭志民,王延松,等.基于多尺度低秩模型的电力无线接入网异常流量检测方法[J].电子学报,2020,48(8):1552–1557.]

[21] Yang Ting,Hou Yucheng,Zhao Liyuan,et al.Abnormal traffic detection method of substation communication network based on time–frequency domain mixed features[J].Automation of Electric Power Systems,2020,44(16):79–86.[杨挺,侯昱丞,赵黎媛,等.基于时–频域混合特征的变电站通信网异常流量检测方法[J].电力系统自动化,2020,44(16):79–86.]

[22] Shi Yucheng,He Weiguo,Zhao Jian,et al.Expected output calculation based on inverse distance weighting and its application in anomaly detection of distributed photovoltaic power stations[J].Journal of Cleaner Production,2020,253:119965.

[23] Jiang Jinfang,Han Guangjie,Liu Li,et al.Outlier detection approaches based on machine learning in the Internet-of-things[J].IEEE Wireless Communications,2020,27(3):53–59.

[24] Chahla C,Snoussi H,Merghem L,et al.A deep learning approach for anomaly detection and prediction in power consumption data[J].Energy Efficiency,2020,13(8):1633–1651.

[25] Tsukada M,Kondo M,Matsutani H.A neural network-based on-device learning anomaly detector for edge devices[J].IEEE Transactions on Computers,2020,69(7):1027–1044.

[26] Wu Kehe,Li Jiawei,Zhang Bo.Abnormal detection of wireless power terminals in untrusted environment based on double hidden Markov model[J].IEEE Access,2021,9(11):18682–18691.

[27] Cook A A,Mısırlı G,Fan Zhong.Anomaly detection for IoT time-series data:A survey[J].IEEE Internet of Things Journal,2020,7(7):6481–6494.

[28] Zhang Yang,Sheng Hao,Wu Yubin,et al.Multiplex labeling graph for near-online tracking in crowded scenes[J].IEEE

Internet of Things Journal,2020,7(9):7892–7902.

[29] Li Ding,Dick S.Residential household non-intrusive load monitoring via graph-based multi-label semi-supervised learning[J].IEEE Transactions on Smart Grid,2019,10(4):4615–4627.

[30] Xiao Qing,Zhou Shaowu.Probabilistic power flow computation using quadrature rules based on discrete Fourier transformation matrix[J].International Journal of Electrical Power & Energy Systems,2019,104:472–480.

[31] Zhang Ningyu,Zhang Ke,Li Qun,et al.Optimal power flow calculation of power system containing TCPST based on improved MCCIPM[J].Electric Power Engineering Technology,2021,40(5):144–150.[张宁宇,张恪,李群,等.基于改进MCCIPM的含TCPST电力系统最优潮流计算[J].电力工程技术,2021,40(5):144–150.]

[32] Pei Maolin,Huang Yangjie,Zhao Wei,et al.A review on outlier detection techniques for smart meter data analytic[J].Electrical Measurement & Instrumentation,2018,55(23):129–135.[裴茂林,黄洋界,赵伟,等.智能电能表异常测量数据诊断方法综述[J].电测与仪表,2018,55(23):129–135.]

[33] Peng Huili,Jin Kaizhong,Fu Congcong,et al.Private time series pattern mining with sequential lattice[J].Acta Electronica Sinica,2020,48(1):153–163.[彭慧丽,金凯忠,付聪聪,等.基于序列格的隐私时序模式挖掘方法[J].电子学报,2020,48(1):153–163.]

[34] Thuy H T T,Anh D T,Chau V T N.Efficient segmentation-based methods for anomaly detection in static and streaming time series under dynamic time warping[J].Journal of Intelligent Information Systems,2021,56(1):121–146.

[35] Zhou Yanjun,Ren Huorong,Li Zhiwu,et al.An anomaly detection framework for time series data:An interval-based approach[J].Knowledge-Based Systems,2021,228:107153.

[36] Pang Guansong,Cao Longbing,Chen Ling.Homophily outlier detection in non-IID categorical data[J].Data Mining and Knowledge Discovery,2021,35(4):1163–1224.

[37] Krleža D,Vrdoljak B,Brčić M.Statistical hierarchical clustering algorithm for outlier detection in evolving data streams[J].Machine Learning,2021,110(1):139–184.

[38] Edje A E,Abd Latiff S M,Chan H W.Enhanced non-parametric sequence-based learning algorithm for outlier detection in the Internet of Things[J].Neural Processing Letters,2021,53(3):1889–1919.

（编辑　吴芝明）

引用格式：Guan Yi,Xie Xiaochuan,Hu Lin,et al.Outlier data detection and monitoring of smart electric grid based on data-tag[J].Advanced Engineering Sciences,2023,55(3):243–254.[管荑,谢小川,胡琳,等.基于数据标签的智能电网监控与异常检测[J].工程科学与技术,2023,55(3):243–254.]