

DOI: 10.19783/j.cnki.pspc.200904

基于多源异构数据融合的综合管廊电力舱系统保护

王罡¹, 刘敬文², 李国鹏², 李志雷², 沈学良², 龚志丹³, 黄文林³, 赵辉⁴, 谷志成⁴

(1. 国网河北省电力有限公司, 河北 石家庄 050022; 2. 国网河北省电力有限公司雄安新区供电公司, 河北 雄安 071800; 3. 厦门亿立吉奥信息科技有限公司, 福建 厦门 361008; 4. 上海谷元电气科技有限公司, 上海 201708)

摘要: 针对综合管廊电力舱运维数据的利用和分析不足、未实现基于大数据的智能运维的问题, 提出了一种新的综合管廊电力舱状态分析判断方法, 旨在保护综合管廊电力舱系统, 提高安全性, 提升运维水平。首先, 将来自管廊电力舱的多个分布式数据源采用中间件技术进行数据层融合。其次, 将具有相似特性的数据分配到同一子空间, 并提出一种新的既能保持数据集全局结构又能保持数据集局部结构的局部全局投影方法提取特征。进而, 对各个子空间提取的特征进行融合, 利用支持向量数据描述方法构建分类模型。最后, 在综合管廊电力舱的运维数据上进行测试以证明所提出方法的有效性以及优越性。

关键词: 管廊电力舱; 多源异构; 数据融合; 状态分析判断; 特征提取

System protection of a pipe corridor power cabin based on multi-source heterogeneous data fusion

WANG Gang¹, LIU Jingwen², LI Guopeng², LI Zhilei², SHEN Xueliang², GONG Zhidan³, HUANG Wenlin³, ZHAO Hui⁴, GU Zhicheng⁴

(1. State Grid Hebei Electric Power Co., Ltd., Shijiazhuang 050022, China; 2. Xiong'an New District Power Supply Company, State Grid Hebei Electric Power Co., Ltd., Xiong'an 071800, China; 3. Xiamen Great Power Geo Information Technology Co., Ltd., Xiamen 361008, China; 4. Shanghai Guyuan Electric Technology Co., Ltd., Shanghai 201708, China)

Abstract: There is insufficient use and analysis of the operational and maintenance data of the pipe corridor power cabin. Also there is a failure to achieve intelligent operation and maintenance based on big data. Thus a new method for analyzing and judging the status of a pipe corridor power cabin is proposed to protect the integrated pipeline gallery power cabin system. This is intended to improve safety and enhance the level of operation and maintenance. First, the data from multiple distributed data sources is integrated using middleware technology for data layer fusion. Secondly, data with similar characteristics are allocated to the same subspace. Then, in each subspace, a novel Locality Global Projections (LGP) method that can maintain both the global structure of the dataset and the local structure of the dataset is proposed to extract features. The extracted features of each subspace are fused, and the Support Vector Data Description (SVDD) method is used to build a classification model. Finally, through testing on the operational and maintenance data of the power cabin of the integrated pipe corridor, the effectiveness and superiority over other methods are proved.

This work is supported by the Science and Technology Project of State Grid Corporation of China (No. KJGW2019-04-II).

Key words: pipe corridor power cabin; multi-source heterogeneous; data fusion; state analysis and judgment; feature extraction

0 引言

综合管廊电力舱运行可靠与安全是电网本质安全的关键。随着大数据智能感知的迅速发展, 综合

管廊电力舱运维过程中会产生指数级增长的海量异构大数据, 主要体现为数据来源多、数据体量大、增长快、数据类型异构多样、数据关联复杂、状态数据离散化、孤岛化严重、作业效率低下^[1]。受限于巡检技术自动化与智能化水平较低, 传统的电力系统保护与控制方法不足之处慢慢浮现, 运维数据的利用和分析不足, 缺乏大数据分析 with 处理相关技

基金项目: 国家电网公司科技项目资助(KJGW2019-04-II)
“城市电网综合管廊电力舱智能化运检关键技术研究”

术, 未能充分利用所获取的海量数据提升系统运行水平与效率^[2-3]。

中国电机工程学会发表的中国电力大数据发展白皮书指出:“中国的电力工业经过几十年的高速发展, 随着下一代智能化电力系统建设的全面展开, 电力大数据将贯穿未来电力工业生产管理等各个环节, 起到巨大的作用, 是中国电力工业在打造下一代电力工业系统过程中有效应对资源有限、环境压力等问题, 实现厚积薄发、绿色可持续发展的关键^[4]。大数据技术在电网中的应用, 将为系统运行状态的实时检测与准确预测、系统运行状态的安全评估、系统运行优化控制提供有力的技术支持^[5]。

综合管廊电力舱在运维过程中积累了大量的传统传感器数据、红外图像、音频、文本等多源异构信息, 可以从多角度全面准确地表征管廊电力舱状态, 克服传统单通道信息的片面性和不足。然而, 综合管廊电力舱在海量多源异构信息处理与融合, 进而分析判断综合管廊电力舱状态方面的措施落实不到位, 缺乏大数据驱动的综合管廊全生命周期足迹监测与风险控制的手段, 以保障管廊电力舱安全高效运行。

多源异构数据指的是结构、存取、形式不一样的多个数据源集成的数据, 往往存储在异构分布式数据库中^[6-7]。多源异构数据融合可以被划分为数据层融合、特征层融合、决策层融合三个层次。数据层融合直接面向原始数据, 是底层的融合。该层次的融合能够在减少数据量损失的同时提高数据信息的精度, 其包括模式融合和数据复制两类方法。模式融合典型代表方法包括联邦数据库和中间件融合方法。联邦数据库将所有数据库一一对应进行连接, 以实现每个数据库和其他各个数据库之间的相互操作。其中格式冲突是最需要解决的问题, 需要为每一个数据库提供转换规则。中间件融合方法通过每个数据源中的包装器将数据虚拟化为公共数据模式。数据复制方法典型代表为数据仓库^[8], 其将数据源中的数据复制到单一数据源中, 目的是数据能保证一致性, 提高数据信息共享率。

特征级融合常用方法包括多元统计方法、人工神经网络、特征压缩聚类法、卡尔曼滤波、多假设法等。在众多特征提取方法中, 多元统计分析方法由于能够有效处理多变量高度耦合的数据而得到了广泛的应用, 其主要通过多元投影策略对数据降维, 提取关键特征和有效信息, 进而将多元样本空间分解为不同的低维子空间。目前比较常用的多元统计分析方法有主元分析法(Principal Component Analysis, PCA)^[9-10]、独立元分析法(Independent

Component Analysis, ICA)^[11-13]、偏最小二乘法(Partial Least Squares, PLS)^[14-15]、费舍尔判别分析法(Fisher Discriminant Analysis, FDA)^[16]、典型相关性分析法(Canonical Correlation Analysis, CCA)^[17-18]等。为了解决非线性数据的特征提取问题, 典型的方法主要包括基于核的方法^[19]以及基于局部线性化的方法^[20]。考虑到数据之间的动态性, 自回归动态潜变量模型被广泛用于提取自关联关系以及互相关关系^[21]。对于具有多峰分布的数据集, 多模型特征提取方法^[22]以及全局特征提取方法^[23]被提出。为了解决数据高斯分布与非高斯分布混合的问题, 独立元分析-主成分分析方法(Independent Component Analysis-Principal Component Analysis, ICA-PCA)方法被提出^[24]。

除了多元统计分析方法, 流形学习方法也被广泛用于提取数据特征。最常用的流形学习方法包括邻域保持投影(Neighborhood Preserving Projections, NPE), 局部线性嵌入(Locally Linear Embedding, LLE), 局部保持投影(Locality Preserving Projections, LPP), 拉普拉斯特征映射(Laplacian Eigenmaps, LE)^[25-26]。Ma 等通过最小化近邻点间距离并最大化非近邻点间距离, 提出了局部非局部映射方法(Local and Nonlocal Embedding LNLE)^[27]。基于构建的双重权重矩阵以及增强目标函数, 增强邻域保持投影(Enhanced Neighborhood Preserving Embedding, ENPE)方法被提出^[28]。Zhang 等通过结合 PCA 以及 LPP 的优点, 提出了全局局部结构分析方法(Global Local Structure Analysis, GLSA)^[29]。基于全局连接图以及局部连接图中的邻域嵌入结构, 多流形投影(Multi-Manifold Projection, MMP)方法被提出^[30]。为了充分利用动态信息, Miao 等提出了时间邻域保持嵌入方法(Time Neighborhood Preserving Embedding, TNPE)^[31]。加权邻域保持嵌入以及支持向量数据描述(Weighted Neighborhood Preserving Embedding and Support Vector Data Description, WNPE-SVDD)方法被提出以解决有益信息被压制的问题^[32]。

为了保护综合管廊电力舱系统, 提高安全性, 需要对其运行状态进行评估。目前, 已有多种方法被提出并成功应用^[33-34]。本文研究大数据技术在综合管廊电力舱安全保护与控制中的运用, 在多源异构数据融合的基础上构建分类器判断分析综合管廊电力舱运行状态, 利用大数据技术建立数据驱动模型对综合管廊电力舱状态运行进行分析, 提高设备的可靠性和综合使用率。首先, 将来自多个分布式数据源的数据集, 采用基于 JSON 的中间件方法进行数据层融合。其次, 考虑到数据层融合之后的

数据集维度较高,为了降低复杂度的同时提升精度,将具有相似特性的变量划分到同一子空间。此外,为了能够在特征提取过程中保持数据集的全局信息,本文提出了一种新的局部全局投影(Locality Global Projections, LGP)方法提取特征。进而,对各个子空间得到的特征进行融合,并利用支持向量数据描述(Support Vector Data Description, SVDD)算法构建分类模型,判断分析综合管廊电力舱运行状态。最后,通过在管廊电力舱的运维数据上进行测试,并与经典方法进行对比以说明本文提出方法的有效性。

本文创新点如下:(1)提出了利用 JSON 作为中间件实施数据层融合。(2)提出了一种新的既能保持数据集局部结构又能保持数据集全局结构的 LGP 方法提取数据特征。(3)在数据层融合的基础上,进一步利用子空间方法以及提出的 LGP 方法进行特征层融合。(4)在融合后特征的基础上构建分类器,并在实际管廊电力舱运维数据上进行了应用,能够准确判断评价运行状态,保障电力舱系统安全。

1 多源异构数据融合研究方法和分析

本文提出的多源异构数据融合方法按照数据层融合-特征层融合两层融合方式进行。

1.1 数据层融合

多源异构数据层融合的关键是以统一的数据格式描述各个异构数据源,屏蔽平台、系统、环境、表现形式等带来的异构问题。本文采用中间件整合方式,利用轻量级的 JSON 描述格式进行数据格式转换,为用户提供统一数据模式和通用接口,把不同格式、来源、性质的数据有机地集中起来,以消除底层数据源之间的异构问题,实现多个数据源的统一访问。

JSON 是一种轻量级的数据交换格式,具有语言独立性和平台无关性。不仅使人们容易阅读和编写,而且方便进行生成和解析。JSON 采用了完全独立于程序语言的文本格式,但是也使用了类 C(包括 C, C++, C#, Java, Java Script, Perl, Python 等)的特性。JSON 的种种特性使其成为理想的数据交换语言,作为通用数据格式以克服异构数据源之间的交互,大大提高数据交互处理的速度和效率。JSON 作为数据传输格式,异构数据源之间相互独立透明,无需了解其他数据源信息,可以灵活表示对象,能够表达各种类型的数据,其作为一种通用的结构化数据表示,能够有效实现异构数据源的格式转换并融合。

多源异构数据集中的每个异构数据源对应一个 JSON 数据模式文档,当数据源变化时,只需修

改对应的模式文件即可,能够实现数据源的“即插即用”。多源异构融合的关键主要包括语法异构和语义异构。语法异构较易,无需关注数据内容、语义,仅需根据数据结构实现各个数据结构之间的映射。本文使用 JSON Schema 来对数据源进行映射,将数据源元数据投影到 JSON 字符串,消除数据类型异构。JSON Schema 用于描述现有数据格式,还可以用作验证客户端提交的数据,进而确保数据有效、安全。本文利用关联数据 JSON-LD,解决多个数据源之间的语义异构,以确保语义一致,实现对数据源的语义描述。

综上,相比较传统的中间件 XML 格式而言,JSON 结构简单,没有标记化带来的数据量大的问题,且解析较方便。此外,JSON 支持多种环境、平台,在进行数据传输时快速便捷,适于数据交换。

1.2 特征层融合

特征层融合是中间层次的融合,其首先进行特征提取,然后对这些特征进行融合处理。数据层融合之后的数据集维度较高,不同变量的类型、数据特性可能不同,且往往存在冗余信息。为了去除冗余信息以及噪声,更加准确地提取特征,对数据层融合之后的数据集根据不同的数据类型或者数据特性划分子空间是十分有必要的。子空间构建的关键是保证同一个子空间中数据特性相似,不同子空间中数据特征具有差异。因此,将同一数据类型且具有相似数据特性的数据划分为同一子空间,进而各个子空间中采用特征提取方法提取特征,并对特征进行融合。典型的多元统计分析方法如 PCA 在特征提取过程中仅保持数据集的全局信息,局部信息被忽略。典型的流形学习方法如 NPE 在特征提取过程中仅保持数据集的局部信息,全局信息被忽略。

在各个子空间中,为了在提取特征的过程中既保持数据集的全局结构,又保持数据集的局部结构,本文提出了一种名为局部全局投影的 LGP 方法。假设子空间数据集可以表示为 $\mathbf{X}_c \in \mathbf{R}^{n \times m_c}$,其中 n 为样本个数, m_c 为观测变量个数,首先我们需要对样本数据进行归一化处理,即每个变量都减去各自的均值,除以各自的标准差。

在提出的 LGP 方法中,为了表征数据集局部结构,采用高斯核描述数据与其近邻的距离,且从原始空间到特征空间力求高斯核距离保持不变。换言之,原始数据集中的近邻关系得到保持。具体的方法是建立权重矩阵如式(1)。

$$W_{ij} = \begin{cases} \frac{\|x_i - x_j\|^2}{\sigma} & \text{如果 } x_j \text{ 是 } x_i \text{ 的近邻} \\ 0 & \text{否则} \end{cases} \quad (1)$$

式中： $\mathbf{x}_i \in R^{m_c \times 1} (i=1,2,\dots,n)$ ； $\mathbf{x}_j \in R^{m_c \times 1} (j=1,2,\dots,n)$ 为 $\mathbf{X}_c \in R^{n \times m_c}$ 中的样本。

为了表征全局结构，LGP方法利用互信息描述整体数据集的关联关系，且从原始空间到特征空间力求互信息保持不变。换句话说，原始数据集中的变量关联关系得到保持。互信息的介绍如式(2)。

$$I(\mathbf{x}^i, \mathbf{x}^j) = \iint_{\mathbf{x}^i, \mathbf{x}^j} p(\mathbf{x}^i, \mathbf{x}^j) \log\left(\frac{p(\mathbf{x}^i, \mathbf{x}^j)}{p(\mathbf{x}^i)p(\mathbf{x}^j)}\right) d\mathbf{x}^i d\mathbf{x}^j \quad (2)$$

式中： $\mathbf{x}^i \in R^{n \times 1} (i=1,2,\dots,m_c)$ 和 $\mathbf{x}^j \in R^{n \times 1} (i=1,2,\dots,m_c)$ 为 $\mathbf{X}_c \in R^{n \times m_c}$ 中的变量； $I(\mathbf{x}^i, \mathbf{x}^j)$ 为变量 \mathbf{x}^i 与变量 \mathbf{x}^j 之间的互信息值， $p(\mathbf{x}^i, \mathbf{x}^j)$ 为联合概率密度函数； $p(\mathbf{x}^i)$ 为边缘概率密度函数。子空间 c 中数据集 $\mathbf{X}_c \in R^{n \times m_c}$ 的互信息矩阵表示为 $MI(\mathbf{X}_c)$ 。如果两个变量的关系近，其互信息值也较大。否则，互信息值较小。在提出的LGP方法中，子空间 c 中数据集的全局结构表示为互信息矩阵 $MI(\mathbf{X}_c)$ 。

LGP方法的目标函数构建如下：

$$J = J_g / J_l \quad (3)$$

$$J_g = \max A_c^T (MI(\mathbf{X}_c)) A_c \quad (4)$$

$$J_l = \min \frac{1}{2} \sum_{ij} (\mathbf{y}_i - \mathbf{y}_j)^2 W_{ij} = \min A_c^T \mathbf{X}^T (\mathbf{D} - \mathbf{W}) \mathbf{X} A_c$$

$$D_{ii} = \sum_j W_{ij} \quad (5)$$

目标函数 J 的约束如下：

$$A_c^T \mathbf{X}^T (\mathbf{D} - \mathbf{W}) \mathbf{X} A_c = I \quad (6)$$

基于拉普拉斯乘子，从原始数据集到特征数据集的投影矩阵 A_c 可以通过求解式(7)得到。

$$(MI(\mathbf{X}_c)) A_c = \lambda \mathbf{X}^T (\mathbf{D} - \mathbf{W}) \mathbf{X} A_c \quad (7)$$

提取的低维度特征 $\mathbf{Y}_c \in R^{n \times k_c}$ 确定如下：

$$\mathbf{Y}_c = \mathbf{X}_c A_c \quad (8)$$

相较于在构建目标函数的过程中仅保持数据集局部结构的基础流形学习算法LPP，所提出的LGP算法不仅能够保持数据集的局部结构而且能够保持数据集的全局结构。将各个子空间中的数据层融合后的数据用作LGP方法的输入，挖掘数据蕴含的有用信息，剔除噪声，提取特征。在子空间 c 中，针对数据集 $\mathbf{X}_c \in R^{n \times m_c}$ 提取的特征为 \mathbf{Y}_c 。对于各个子空间中的数据集 $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_C$ ，特征数据集 $\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_C$ 可以得到。最终，特征层的融合结果表示为

$$\mathbf{Y} = [\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_C] \in R^{n \times (k_1 + k_2 + \dots + k_C)} \quad (9)$$

\mathbf{Y} 的维度为 $k_1 + k_2 + \dots + k_C$ ，该维度小于 $m_1 +$

$m_2 + \dots + m_C$ 。综上，特征层不仅可以减小维数，更重要的是，它可以通过降维来消除噪声，这有利于挖掘数据中的模式。

2 基于融合特征的分类模型构建方法

在融合后特征集的基础上，为了判断综合管廊电力舱状态，本文构建支持向量数据描述SVDD分类模型以判断综合管廊运行在正常状态或异常状态。SVDD是一种数据描述方法，最早提出应用于二阶分类问题。SVDD将数据映射到较高维空间，然后，在较高维空间建立一个能够包含所有数据的超球体^[35-36]。最小化式(10)以构造超球体。

$$R^2 + G \sum_{i=1}^n \xi_i \quad (10)$$

$$\text{s.t. } \|\phi(\mathbf{x}_i) - \mathbf{a}\|^2 \leq R^2 + \xi_i, \xi_i \geq 0 \quad i=1,2,\dots,n$$

式中： \mathbf{a} 为超球体的中心； R 为超球体的半径； G 权衡体积与误差； ξ_i 为松弛变量。

通过引入核函数，式(10)中的问题转换为最大化下式(11)。

$$\sum_{i=1}^n \alpha_i K(\mathbf{y}_i, \mathbf{y}_j) - \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j K(\mathbf{y}_i, \mathbf{y}_j) \quad (11)$$

$$\text{s.t. } 0 \leq \alpha_i \leq G, \sum_{i=1}^n \alpha_i = 1$$

其中： $\mathbf{y}_i (i=1,2,\dots,n)$ 和 $\mathbf{y}_j (j=1,2,\dots,n)$ 为 $\mathbf{Y} \in R^{n \times (k_1 + k_2 + \dots + k_C)}$ 中的样本； α_i 为拉普拉斯乘子。

超球体的中心 \mathbf{a} 及半径 R 如下：

$$\mathbf{a} = \sum_{i=1}^n \alpha_i \phi(\mathbf{y}_i) \quad (12)$$

$$R^2 = K(\mathbf{y}_k, \mathbf{y}_k) - 2 \sum_{i=1}^n \alpha_i K(\mathbf{y}_k, \mathbf{y}_i) + \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j K(\mathbf{y}_i, \mathbf{y}_j) \quad (13)$$

其中， \mathbf{y}_k 为支持向量。

对于数据层融合后的在线测试数据 \mathbf{x}_t ，通过子空间划分及特征提取得到其低维特征 \mathbf{y}_t 。监测指标SVDD按照下式定义：

$$SVDD = \frac{\|\mathbf{y}_t - \mathbf{a}\|^2}{R^2} = \frac{K(\mathbf{y}_t, \mathbf{y}_t) - 2 \sum_{i=1}^n \alpha_i K(\mathbf{y}_t, \mathbf{y}_i) + \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j K(\mathbf{y}_i, \mathbf{y}_j)}{R^2} \quad (14)$$

根据式(14)，监测指标SVDD的控制线为1。若当前在线测试数据的监测指标值大于1，则认为当前是异常状态。若当前在线测试数据的监测指标值

小于 1, 则认为当前是正常状态。

3 研究结果

本章将在综合管廊电力舱运维数据上验证提出方法的有效性。综合管廊电力舱关键运维数据指标包括甲烷、氯气、一氧化碳、二氧化碳、可燃气体、光栅温度、二氧化硫、湿度、氧气、有毒气体、温度、水位等 12 项。每项运行数据存储在数据库中, 例如, 甲烷测量数据存储如表 1 所示。

表 1 甲烷测量数据存储表

Table 1 Methane measurement data storage table					
序号	名称	数据类型	长度	小数位	允许空值
1	ID	varchar(36)	36	0	N
2	DeviceID	varchar(36)	36	0	N
3	UpdateTime	datetime	23	0	Y
4	Value	float	11	2	N
5	State	int	10	0	N
6	DeviceStatus	int	10	0	N

对于 12 项指标, 首先进行数据层融合, 采用 varchar 类型的 ID 号作为不同表融合的主键, 依据 datetime 类型的 UpdateTime 进行不同时刻对齐。将 12 张表进行数据层融合后, 对表中的 Value, State, DeviceStatus 这 36 项指标, 利用所提出的 LGP 方法在每个子空间中提取特征, 进行特征级融合, 并建立 SVDD 分类模型, 称之为子空间局部全局投影方法(SLGP-SVDD)方法, 进而判断分析综合管廊电力舱状态。采集正常运行状态下的 4 000 个数据作为训练数据集, 利用提出的 LGP 方法在每个子空间提取特征, 将各个子空间的特征进行融合。测试数据集设置如下: 综合管廊电力舱首先运行在正常状态, 在第 200 个数据点, 甲烷含量突然升高 0.2 ppm, 氯气含量突然升高 1 ppm。

为了说明本文提出的 SLGP-SVDD 方法的有效性及其优越性, 利用广泛使用的 PCA 方法在每个子空间中提取特征并进行融合, 建立 SVDD 分类模型, 称之为子空间主成分分析-支持向量数据描述方法(Subspace PCA-SVDD, SPCA-SVDD)。此外, 为了说明子空间划分的有效性, 利用 PCA 方法对不划分子空间的数据直接提取特征, 建立 SVDD 分类模型, 称之为主成分分析-支持向量数据描述方法(PCA-SVDD)。通过计算误报率和漏报率来说明三种方法 PCA-SVDD、SPCA-SVDD、SLGP-SVDD 的性能。误报率为测试数据 1 至数据 200 的区间, 将正常数据误判为故障数据的比例。漏报率为数据 201 至数据 1 000 的区间, 将故障数据误判为正常数据的比例。

图 1、图 2、图 3 分别给出了 PCA-SVDD、

SPCA-SVDD、SLGP-SVDD 方法对于测试数据集的分析判断结果。从这三张图中可以看出, 三种方法对于正常过程均能给出较好的分析判断结果, 且三种方法均能判断出异常状态的发生, 但是三种方法对于异常状态的判断能力不同。表 2 和表 3 分别给出了三种方法的误报率及漏报率。

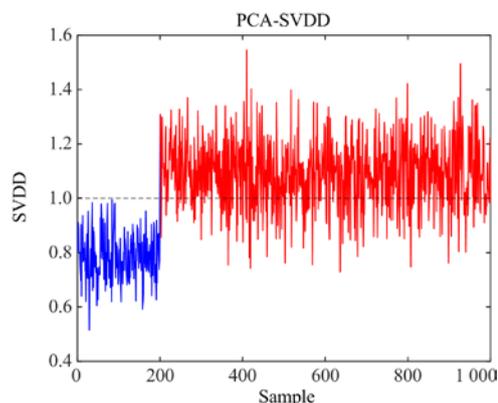


图 1 PCA-SVDD 方法对于测试数据集的结果

Fig. 1 Result of the PCA-SVDD method for the testing dataset

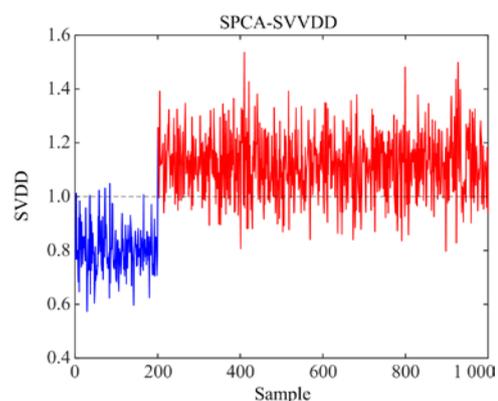


图 2 SPCA-SVDD 方法对于测试数据集的结果

Fig. 2 Result of the SPCA-SVDD method for the testing dataset

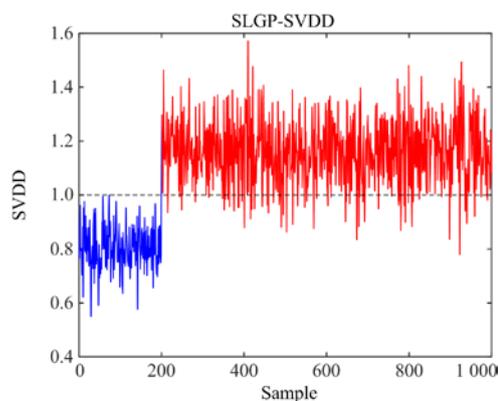


图 3 SLGP-SVDD 方法对于测试数据集的结果

Fig. 3 Result of the SLGP-SVDD method for the testing dataset

表 2 三种方法的误报率

Table 2 False alarm rates of three methods

PCA-SVDD/%	SPCA-SVDD/%	SLGP-SVDD/%
0.5	3.5	0.5

表 3 三种方法的漏报率

Table 3 Missed alarm rates of three methods

PCA-SVDD/%	SPCA-SVDD/%	SLGP-SVDD/%
23.875	16.25	8.25

在图 1 中, 超过 20% 的异常数据被判为正常数据, PCA-SVDD 方法的漏报率为 23.875%。相较于 PCA-SVDD 方法, 通过子空间划分, 在各个子空间中采用 PCA 提取特征的 SPCA-SVDD 方法的漏报率降到了 20% 以下, 说明了将具有相似特性的数据划分为子空间的策略是有效可行的。SPCA-SVDD 与 SLGP-SVDD 方法的不同之处在于各个子空间中的特征提取方法不同。相较于 SPCA-SVDD 方法, SLGP-SVDD 方法的漏报率降到了 10% 以下, 说明了本文提出的 SLGP-SVDD 方法的优越性。

4 结论

在该论文中, 提出了一种基于多源异构数据融合的综合管廊电力舱系统保护方法。该方法包括数据层融合以及特征层融合。在数据层融合方面, 采用 JSON 作为中间件以集成融合多个分布式数据源。在特征层融合方面, 首先将具有相似特性的数据划分为子空间以降低维度、提高精度, 进而利用提出的既能保持数据集局部结构又能保持数据集全局结构的 LGP 方法在每个子空间中提取特征, 最后将各个子空间中提取的特征进行融合。在融合后特征的基础上, 通过构建 SVDD 分类器, 分析判断综合管廊电力舱的状态。最后在综合管廊电力舱的运维数据上进行测试, 说明所提出的子空间构建方法以及提出的 LGP 方法的有效性以及优越性。

参考文献

- [1] 叶健锋. 智能电网设备状态监测评估及预测应用研究[J]. 电子制作, 2018, 10: 64-65.
YE Jianfeng. Research on application of monitoring, evaluation and prediction of smart grid equipment[J]. Practical Electronics, 2018, 10: 64-65.
- [2] 王金丽, 盛万兴, 王金宇, 等. 中低压配电网统一数据采集与监控系统设计和实现[J]. 电力系统自动化, 2012, 36(18): 72-76.
WANG Jinli, SHENG Wanxing, WANG Jinyu, et al. Design and implementation of a centralized data acquisition and supervisory system for medium-low voltage distribution network[J]. Automation of Electric Power Systems, 2012, 36(18): 72-76.
- [3] 姚成, 黄国方, 仲雅霓, 等. 基于站控层 GOOSE 的站域控制实现方案[J]. 电力系统自动化, 2012, 36(18): 159-160.
YAO Cheng, HUANG Guofang, ZHONG Yani, et al. A realization scheme of substation area control system based on GOOSE message in station level[J]. Automation of Electric Power Systems, 2012, 36(18): 159-160.
- [4] 宋亚奇, 周国亮, 朱永利. 智能电网大数据处理技术现状与挑战[J]. 电网技术, 2013, 37(4): 927-935.
SONG Yaqi, ZHOU Guoliang, ZHU Yongli. Present status and challenges of big data processing in smart grid[J]. Power System Technology, 2013, 37(4): 927-935.
- [5] 张沛, 吴潇雨, 和敬涵. 大数据技术在主动配电网中的应用综述[J]. 电力建设, 2015, 36(1): 52-59.
ZHANG Pei, WU Xiaoyu, HE Jinghan. Review on big data technology applied in active distribution network[J]. Electric Power Construction, 2015, 36(1): 52-59.
- [6] 屈子程, 高亮, 康保林, 等. 基于多源数据的电力系统故障全信息诊断模型[J]. 电力系统保护与控制, 2019, 47(22): 61-66.
QU Zicheng, GAO Liang, KANG Baolin, et al. A power system fault full information diagnosis model based on multi-source data[J]. Power System Protection and Control, 2019, 47(22): 61-66.
- [7] MING H, XIA B N, LEE K Y, et al. Prediction and assessment of demand response potential with coupon incentives in highly renewable power systems[J]. Protection and Control of Modern Power Systems, 2020, 5(2): 124-137. DOI: 10.1186/s41601-020-00155-x.
- [8] KALLE T, MARGUS K, TANEL T. Data integration patterns for data warehouse automation[J]. New Trends in Database and Information Systems II, 2015, 312: 41-55.
- [9] 陈勇, 李鹏, 张忠军, 等. 基于 PCA-GA-LSSVM 的输电线路覆冰负荷在线预测模型[J]. 电力系统保护与控制, 2019, 47(10): 111-119.
CHEN Yong, LI Peng, ZHANG Zhongjun, et al. Online prediction model for power transmission line icing load based on PCA-GA-LSSVM[J]. Power System Protection and Control, 2019, 47(10): 111-119.
- [10] LV Z M, JIANG Q C, YAN X F. Batch process monitoring based on multisubspace multiway principal component analysis and time-series Bayesian inference[J]. Industrial & Engineering Chemistry Research, 2014, 53(15): 6457-6466.
- [11] 田凤兰, 张恩泽, 潘思蓉, 等. 基于特征量优选与 ICA-SVM 的变压器故障诊断模型[J]. 电力系统保护与控制, 2019, 47(17): 164-170.
TIAN Fenglan, ZHANG Enze, PAN Sirong, et al. Fault diagnosis model of power transformers based on feature quantity optimization and ICA-SVM[J]. Power System Protection and Control, 2019, 47(17): 164-170.
- [12] JIANG Q C, YAN X F. Non-Gaussian chemical process

- monitoring with adaptively weighted independent component analysis and its applications[J]. *Journal of Process Control*, 2013, 23(9): 1320-1331.
- [13] 方静, 彭小圣, 刘泰蔚, 等. 电力设备状态监测大数据发展综述[J]. *电力系统保护与控制*, 2020, 48(23): 177-186.
FANG Jing, PENG Xiaosheng, LIU Taiwei, et al. Development trend and application prospects of big data-based condition monitoring of power apparatus[J]. *Power System Protection and Control*, 2020, 48(23): 177-186.
- [14] ZHANG Y W, DU W Y, FAN Y P, et al. Process fault detection using directional kernel partial least squares[J]. *Industrial & Engineering Chemistry Research*, 2015, 54(9): 2509-2518.
- [15] LI G, QIN S J, ZHOU D H. Geometric properties of partial least squares for process monitoring[J]. *Automatica*, 2010, 46(1): 204-210.
- [16] CHIANG L H, RUSSELL E L, BRAATZ R D. Fault diagnosis in chemical processes using fisher discriminant analysis, discriminant partial least squares, and principal component analysis[J]. *Chemometrics and Intelligent Laboratory Systems*, 2000, 50(2): 243-252.
- [17] CHEN Z. Canonical correlation analysis-based fault detection methods with application to alumina evaporation process[J]. *Control Engineer Practice*, 2016, 46: 51-58.
- [18] YANG J, SONG B, TAN S, et al. Concurrent monitoring of global-local performance indicators for large-scale process[J]. *Journal of the Taiwan Institute of Chemical Engineers*, 2019, 102: 9-16.
- [19] YUAN X F, WANG Y L, YANG C H, et al. Weighted linear dynamic system for feature representation and soft sensor application in nonlinear dynamic industrial processes[J]. *IEEE Transactions on Industrial Electronics*, 2018, 65: 1508-1517.
- [20] WANG L, SHI H B. Multivariate statistical process monitoring using an improved independent component analysis[J]. *Chemical Engineering Research and Design*, 2010, 88: 403-414.
- [21] ZHOU L, ZHENG J Q, GE Z Q, et al. Multimode process monitoring based on switching autoregressive dynamic latent variable model[J]. *IEEE Transactions on Industrial Electronics*, 2018, 65: 8184-8194.
- [22] SONG B, TAN S, SHI H B. Key principal components with recursive local outlier factor for multimode chemical process monitoring[J]. *Journal of Process Control*, 2016, 47: 136-149.
- [23] SONG B, TAN S, SHI H B. Time-space locality preserving coordination for multimode process monitoring[J]. *Chemometrics Intelligent Laboratory Systems*, 2016, 151: 190-200.
- [24] GE Z Q, SONG Z H. Process monitoring based on independent component analysis-principal component analysis (ICA-PCA) and similarity factors[J]. *Industrial & Engineering Chemistry Research*, 2007, 46: 2054-2063.
- [25] BELKIN M, NIYOGI P. Laplacian eigenmaps for dimensionality reduction and data representation[J]. *Neural Computation*, 2003, 15: 1373-1396.
- [26] SAUL L K, ROWEIS S T. Think globally, fit locally: un-supervised learning of low dimensional manifolds[J]. *Journal of Machine Learning Research*, 2003, 4: 119-155.
- [27] MA Y X, SONG B, SHI H B, et al. Fault detection via local and nonlocal embedding[J]. *Chemical Engineering Research and Design*, 2015, 94: 538-548.
- [28] SONG B, TAN S, SHI H B. Process monitoring via enhanced neighborhood preserving embedding[J]. *Control Engineering Practice*, 2016, 50: 48-56.
- [29] ZHANG M, GE Z Q, SONG Z H, et al. Global local structure analysis model and its application for fault detection and identification[J]. *Chemical Engineering Research and Design*, 2011, 50: 6837-6848.
- [30] TONG C D, YAN X F. Statistical process monitoring based on a multi-manifold projection algorithm[J]. *Chemometrics Intelligent Laboratory Systems*, 2014, 130: 20-28.
- [31] MIAO A M, GE Z Q, SONG Z H, et al. Time neighborhood preserving embedding model and its application for fault detection[J]. *Industrial & Engineering Chemistry Research*, 2013, 52: 13717-13729.
- [32] JIANG Q C, YAN X F. Probabilistic weighted NPE-SVDD for chemical process monitoring[J]. *Control Engineering Practice*, 2014, 28: 74-89.
- [33] HE J H, CHEN K A, LI M, et al. Review of protection and fault handling for a flexible DC grid[J]. *Protection and Control of Modern Power Systems*, 2020, 5(2): 151-165. DOI: 10.1186/s41601-020-00157-9.
- [34] XU Y. A review of cyber security risks of power systems: from static to dynamic false data attacks[J]. *Protection and Control of Modern Power Systems*, 2020, 5(3): 190-201. DOI: 10.1186/s41601-020-00164-w.
- [35] YIN G, ZHANG Y T, LI Z N, et al. Online fault diagnosis method based on incremental support vector data description and extreme learning machine with incremental output structure[J]. *Neurocomputing*, 2014, 128: 224-231.
- [36] YOON S, MACGREGOR J F. Fault diagnosis with multivariate statistical models part I: using steady state fault signatures[J]. *Journal of Process Control*, 2001, 11(4): 387-400.

收稿日期: 2020-07-29; 修回日期: 2020-09-25

作者简介:

王 罡(1973—), 男, 博士, 高级工程师, 研究方向为电力系统及其自动化、管廊综合监控技术。E-mail: wangg@he.sgcc.com.cn

(编辑 葛艳娜)