

独立分量分析和流形学习在 VSC-HVDC 系统故障诊断中的应用

李志雄, 严新平

(武汉理工大学能源与动力工程学院, 430063, 武汉)

摘要: 提出一种基于独立分量分析(ICA)和局部线性嵌入流形学习算法(LLE)的新型高压直流输电(VSC-HVDC)系统故障诊断方法. 由于随机噪声的干扰, 单个传感器测得的系统故障信号无法直接用于故障检测, 故使用快速 ICA 对多通道传感器测得的直流电压和电流信号进行盲源分离处理以恢复去噪的系统故障源信号; 然后利用 LLE 挖掘潜藏于恢复信号中的子流形, 提取故障敏感特征; 最后将 LLE 提取的故障特征量作为支持向量机(SVM)的输入, 建立系统故障诊断模型. 通过对系统交流相对相故障、交流相对地故障以及复合故障等仿真信号进行分析, 表明所提出的 ICA-LLE 方法能够有效地提取故障关键特征, 并在 3 维空间将故障特征隔离, 从而得到满意的 SVM 故障识别效果, 且 SVM 分类精度比只使用 LLE 提高了近 20%.

关键词: 输电; 故障诊断; 独立分量分析; 流形学习

中图分类号: TP273 **文献标志码:** A **文章编号:** 0253-987X(2011)02-0044-05

Independent Component Analysis and Manifold Learning with Applications to Fault Diagnosis of VSC-HVDC Systems

LI Zhixiong YAN Xinping

(School of Energy and Power Engineering, Wuhan University of Technology, Wuhan 430063, China)

Abstract: A fault diagnosis scheme for voltage source converter-high voltage direct current transmission (VSC-HVDC) system is proposed based on the independent component analysis (ICA) and the locally linear embedding (LLE) in the present work. The measured signals in the VSC-HVDC can not be used directly to detect system fault due to the heavy inference noise. The FastICA is hence employed to eliminate the disturbed noise and recover the fault sources from the measured DC line voltage and current observation signals. Then the LLE algorithm is applied to extract distinct characteristics hiding in the recovered fault signals. To enhance the fault pattern recognition, the support vector machine (SVM) is adopted to learn the relationship between the fault features and the system operation conditions. The ability of the proposed ICA-LLE method to detect VSC-HVDC system fault is evaluated with the simulated data. The analysis results demonstrate the feasibility and effectiveness. The distinguished features of the fault signals, such as AC line-to-line fault and AC line-to-ground fault, and the compound faults, can be extracted efficiently and then isolated in the 3-D feature space correctly. The classification rate of the SVM with the proposed scheme is increased by 20% compared with LLE.

Keywords: transmission; fault diagnosis; independent component analysis; manifold learning

新型直流输电系统(VSC-HVDC)在分布式发电、远距离负荷供电等诸多领域具有广阔的应用前景^[1-2],对促进国民经济的发展具有重要作用^[2]。然而,一旦系统发生故障,将会造成巨大的经济损失,甚至导致严重的安全事故。因此,进行早期故障隐患的监测与诊断是保证整个系统安全运行的重要手段^[3-7]。

目前,国内外对 VSC-HVDC 系统故障诊断的专门研究还处于试验阶段,相关文献报道较少^[2]。文献[3-5]应用人工神经网络诊断 HVDC 系统故障,取得了可喜的成绩,然而神经网络学习过程需要较多训练样本数,这在工程实践中往往难以满足。为此,文献[6-7]将支持向量机(SVM)用于高压输电系统的故障诊断中,利用 SVM 在较少样本条件下较好的模型泛化性能来抑制过学习问题。需要考虑的是,测量到的系统信号不可避免地会受到各种随机噪声的干扰,这对故障特征的准确提取有很大影响,但目前大多数研究成果未讨论这个问题。

为了消除干扰噪声以及降低不敏感故障特征对 VSC-HVDC 系统故障诊断的影响,本文采用独立分量分析(ICA)对系统直流电压、电流等观测信号进行去噪,并分析了当观测信号偏离相互独立条件时 ICA 的去噪性能。应用局部线性嵌入算法(LLE)来提取最敏感的故障特征,结合支持向量机(SVM)建立系统故障诊断模型,实现了系统故障的智能识别。

1 快速独立分量分析(FastICA)

工程实践中所采集到的信号往往是源信号与噪声的混合,为了获得准确的源信号必须滤除噪声。独立分量分析(ICA)^[8-10]是近年来兴起的信号处理技术,其实质是在各源信号相互独立的假设下,对多路观测信号进行盲分离,挖掘出隐含在混合信号中的独立成分,从而达到排斥噪声干扰的目的。

ICA 基本模型可描述为^[8]

$$\mathbf{x} = \mathbf{A} \cdot \mathbf{s} \quad (1)$$

式中: $\mathbf{A}$ 为 $n \times n$ 的混合矩阵; $\mathbf{x}$ 为 $m \times 1$ 维观测信号矢量; $\mathbf{s}$ 为 n 个独立的源信号组成的列向量。ICA 的目的就是估算 $\mathbf{A}$ 的逆矩阵 $\mathbf{W}$ (即分离矩阵)来恢复独立源信号

$$\hat{\mathbf{s}} = \mathbf{W} \cdot \mathbf{x} \quad (2)$$

式中: $\hat{\mathbf{s}}$ 为 $\mathbf{s}$ 的估计。

文献[8]提出了快速固定点 ICA 算法,即 FastICA,其收敛速度是传统梯度下降算法的 10~100 倍。FastICA 采用负熵最大化判据,其定义如下

$$J(\mathbf{y}) = [E(g(\mathbf{y})) - E(g(\gamma))]^2 \quad (3)$$

式中: $\mathbf{y} = \mathbf{W}^T \mathbf{z}$, $\mathbf{z}$ 为白化后的观测列向量; γ 为零均值、单位方差的高斯变量; $g(\cdot)$ 为高斯函数。在 $\mathbf{W}$ 正交的约束条件下,求取式(3)极大值可得到估算 $\mathbf{W}$ 的迭代式为

$$\mathbf{W}(k) = E\{\mathbf{z}g(\mathbf{W}^T(k-1)\mathbf{z})^3\} - E\{g'(\mathbf{W}^T(k-1)\mathbf{z})\}\mathbf{W}(k-1) \quad (4)$$

式中: $g'(\cdot)$ 为 $g(\cdot)$ 的导数。

2 局部线性嵌入算法(LLE)

故障特征的有效提取是实现故障准确可靠诊断的核心,然而从故障设备采集到的信号往往呈现出复杂非线性,这给准确提取特征信息带来了极大困难。相对于传统的线性降维方法(如主成分分析和 Fisher 判别分析等)在处理非线性问题时的不足,流形学习能够发现非线性高维数据的本质结构,进而能够有效进行数据降维^[10]。局部线性嵌入(LLE)^[11]正是一种用于非线性降维的流形学习算法,它能够使降维后的数据保持原有拓扑结构,从而保留了原始数据的非线性特征。

给定观测数据集 $\mathbf{X} = \{\mathbf{x}_1 \ \mathbf{x}_2 \ \cdots \ \mathbf{x}_n\}$, $\mathbf{x}_i \in \mathbb{R}^p$ (p 为数据维数),LLE 的目标就是通过 $\mathbf{X}$ 来寻找未知映射 $f(\cdot)$ 和低维参数化流形 $\mathbf{T}$, $\mathbf{T} = \{\mathbf{t}_1 \ \mathbf{t}_2 \ \cdots \ \mathbf{t}_n\}$, $\mathbf{t}_i \in \mathbb{R}^q$ ($q \ll p$),使得对于任意的 $\mathbf{t}_i = \mathbf{T}$ 都可以用 $f(\mathbf{t}_i)$ 来估计 $\mathbf{X}$ 。LLE 算法实现步骤如下^[11]。

(1)构造近邻图。使用 k 近邻或 ε 邻域的方法计算每个样本点 $\mathbf{x}_i$ 的邻域。

(2)构建局部重构权值矩阵 $\mathbf{V}$ 。通过使如下损失函数最小化来计算 $\mathbf{V}$

$$\min_{\mathbf{V}} \sum_{i=1}^n \left| \mathbf{x}_i - \sum_{j=1}^k v_j^i \mathbf{x}_{ij} \right|^2 \quad (5)$$

式中: $\mathbf{x}_{ij}$ ($j=1,2,\dots,k$)为 $\mathbf{x}_i$ 的邻近点, v_j^i 为 $\mathbf{x}_i$ 与 $\mathbf{x}_{ij}$ 间的权值,且存在 $\sum_{j=1}^k v_j^i = 1$ 。式(5)不容易直接求解,因此引入局部协方差矩阵 $\mathbf{u}^i$ 来求解 $\mathbf{V}$

$$\mathbf{u}_{jl}^i = (\mathbf{x}_i - \mathbf{x}_{ij})^T (\mathbf{x}_i - \mathbf{x}_{il}) \quad (6)$$

采用拉格朗日乘子法,即可求出局部最优化重建权值矩阵为

$$v_j^i = \sum_{l=1}^k (\mathbf{u}^i)_{jl}^{-1} / \sum_{a=1}^k \sum_{b=1}^k (\mathbf{u}^i)_{ab}^{-1} \quad (7)$$

(3)将所有的样本点映射到低维空间中。低维流形空间 $\mathbf{T}$ 可以通过最小化如下损失函数来计算

$$\min_{\epsilon}(T) = \sum_{i=1}^n \sum_{j=1}^n m_j^i t_i^T t_j = \text{tr}(\mathbf{TMT}^T) \quad (8)$$

其中损失矩阵

$$\mathbf{M} = (\mathbf{I}_{n \times n} - \mathbf{V})^T (\mathbf{I}_{n \times n} - \mathbf{V}) \quad (9)$$

要使损失函数值达到最小，则取 $\mathbf{T}$ 为 $\mathbf{M}$ 的最小 q 个非零特征值所对应的特征向量. 这样，高维数据集 $\mathbf{X}$ 就被映射到了低维子流形空间 $\mathbf{T}$. 由于 LLE 约简了原始数据并保留了其本质结构，因此利用 SVM 可以快速准确地学习到低维空间所蕴含的故障特征信息，实现可靠的故障模式识别. 由于故障样本维数得到了约简，有利于实现样本集的可视化表达，因此使诊断过程直观易理解. 图 1 给出了 VSC-HVDC 系统故障诊断的流程结构.

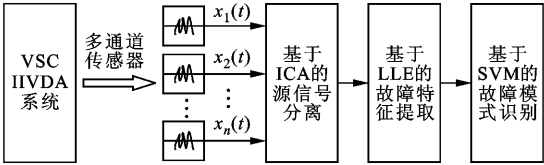


图 1 VSC-HVDC 系统故障诊断的流程结构

3 VSC-HVDC 系统的故障诊断

3.1 系统仿真模型

两端有源交流网络 VSC-HVDC 系统基本原理如图 2 所示. 该标准模型经过了大量的仿真研究，证实可以模拟实际 VSC-HVDC 系统的各种故障暂态响应，得到与实际相符的结果. 文献[1]较详细地介绍了其基本工作原理，这里不再赘述. 本文采用 Matlab/Simulink 仿真软件建立系统模型见图 2，电压为 230 kV，功率为 200 MW，频率为 50 Hz，传输距离为 100 km. 除系统正常运行情况外，本仿真模型同时对几种典型的故障进行了测试，包括整流侧交流相对相故障(LL)、逆变侧三相对地故障(TLG)、整流侧交流三相对地与逆变侧单相对地复合故障(TLG-LG)、直流线路短路与逆变侧三相对相复合故障(DS-LLL).

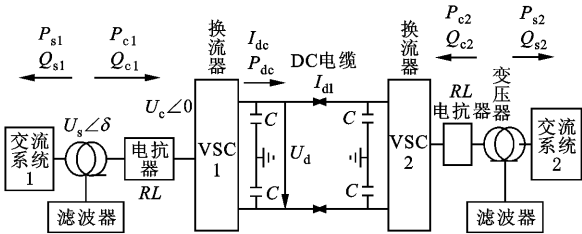


图 2 两端有源交流网络 VSC-HVDC 系统

在采集仿真信号过程中，为了模拟外界干扰噪声对系统响应以及传感器采集的影响，本文将脉冲信号和随机指数噪声混合到了系统交、直流电压、电流中. 图 3 是单个电测仪表采集到的系统直流电压观测信号波形. 从图中可见，不同故障的直流电压波形完全被噪声淹没，难以提取可靠的故障特征信息.

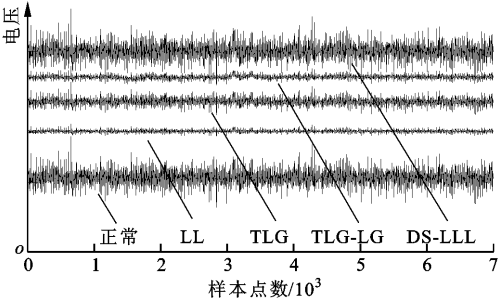


图 3 VSC-HVDC 系统不同故障下的直流电压原始信号

3.2 ICA 去噪处理

由于 VSC-HVDC 系统输电距离比较远，利用单个电测仪表只能获取一些局部测量数据，而且被噪声严重污染，难以判定故障性质. 鉴于 ICA 能够通过多通道信号进行数据融合处理，恢复故障源信号，所以本文在直流线路上等距布置电压/电流传感器各 10 个(间距为 10 km)，同步采集 10 通道直流电压和电流. 图 4 是利用 FastICA 对 10 通道传感器 LL 故障观测信号进行分离的结果(幅值已归一化处理). 其中，观测信号所包含的 HVDC 真实故障信号与干扰噪声的信噪比为 -19.823 dB，表明 HVDC 真实故障信号完全被脉冲冲击干扰和随机指数噪声所淹没. 从图 4 可见，FastICA 将观测信号分离成了 3 个独立源信号，包括 LL 故障信号、噪声信号以及脉冲冲击信号. 由此可见，即使观测信号信噪比较低，故障信号也能够被较好地分离出来. 但是，工程实践中系统发生的故障往往有一定的关联性，会对观测信号产生非线性干扰，导致观测信号中所蕴含的各源信号并不满足相互独立的要求，因此 ICA 的去噪性能会受到严重削弱. 为了定量描述故障关联干扰的影响，本文将干扰简化为观测信号的加性噪声，结合式(1)有

$$\mathbf{x}_1 = \mathbf{x} + \mathbf{d}_n = \mathbf{A}\mathbf{s} + \mathbf{d}_n \quad (10)$$

式中： $\mathbf{d}_n$ 为白噪声. 通过指标 p 可以评价非线性干扰对 ICA 去噪效果，定义该指标为

$$p = \frac{\text{估计故障信号的峭度}}{\text{真实故障信号的峭度}} \quad (11)$$

图 5 给出了 p 随加性噪声的变化曲线. 从图中

可知,当加性噪声低于 20 dB 时, p 急剧下降,达到 0 dB 时,估计故障信号已严重偏离真实信号, FastICA 分离失效。所以,在故障关联干扰条件下,加性噪声应该大于 20 dB, FastICA 才能实现故障源信号的有效分离。

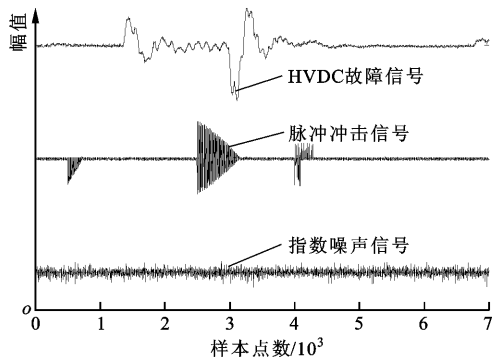


图 4 整流侧 LL 故障盲分离结果(信噪比为 -19.823 dB)

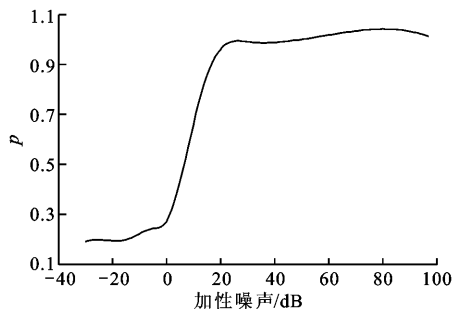


图 5 p 与加性噪声的关系

图 6 和图 7 为经过归一化处理后的系统各故障 ICA 恢复信号(加性噪声为 50 dB)。从图中可以看出,对于不同故障,直流电压、直流电流波形存在一定程度不同,因此本文将直流电压、直流电流作为故障分类的依据。

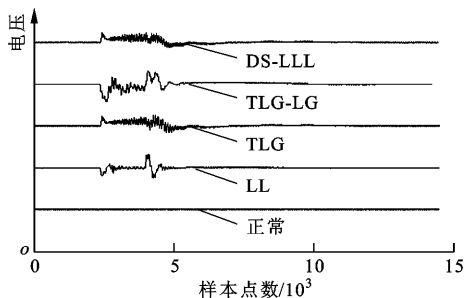


图 6 各故障直流电压 ICA 分离结果

3.3 LLE 特征提取与 SVM 故障识别

ICA 去噪后消除了干扰因素对观测信号的影响,有利于 LLE 挖掘隐藏在系统故障源信号中的本质特征。在 LLE 提取故障特征过程中,嵌入维数 d

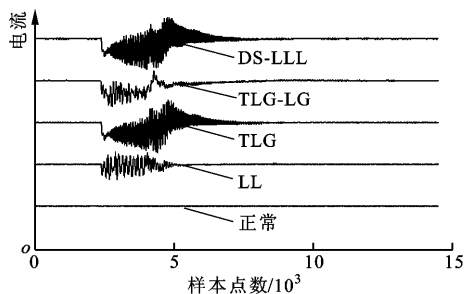


图 7 各故障直流电流的 ICA 分离结果

和邻域因子 k 的选择对提取效果十分重要,不合理的选择会造成关键流形特征的丢失。一般 d 和 k 的确定主要靠经验以及大量的试验。为克服经验设定对特征提取的影响,本文利用 Silhouette 指标来定量评价 LLE 降维的质量,从而确定合理的参数。

假设一个数据集可以划分成 N 个聚类 C_i ($i = 1, 2, \dots, N$), $a(x)$ 为样本 x ($x \in C_i$) 与类内所有其他样本的平均不相似度或距离, $d(x, C_j)$ 为样本 x 到另一个类 C_j 的所有样本的平均不相似度或距离,记 $b(x) = \min\{d(x, C_j)\}$, $j = 1, 2, \dots, N, j \neq i$, 则样本 x 的 Silhouette 指标为

$$S = \frac{b(x) - a(x)}{\max\{b(x), a(x)\}} \quad (12)$$

所有样本的 Silhouette 指标平均值反映了聚类结果的质量,值越大表示聚类质量越好,降维效果越能逼近原始数据的拓扑结构。如果 Silhouette 平均值超过 0.5,说明各聚类间有明显分界面,低于 0.5 则表示存在聚类重叠。

图 8 是利用不同 d 和 k 对 500 个样本(每种故障 100 个样本)进行非线性降维后的 Silhouette 指标平均值曲线。从图中可见:①在嵌入维数为 2、3 和 4 的条件下, k 值在 1~50 间均能够取得较好的聚类效果($S > 0.5$);② $d=2$ 的效果相对最差;③ $k=35$ 时对应 Silhouette 指标曲线的最大值 0.553 1 ($d=3$) 或 0.564 9 ($d=4$)。考虑到 $d=3$ 时 Silhouette 指标最大值与 $d=4$ 的相差不大,但 $d=3$ 时能够将降维特征空间直观地表现在 3 维图像中,所以本文选取邻域因子 $k=35$, 嵌入维数 $d=3$ 。

图 9 为 VSC-HVDC 系统各故障直流电压经过 LLE 提取后的故障特征空间分布。从图中可见,系统不同故障的特征信息具有较明显的分界面, LLE 能够将正常、LL 故障和 TLG-LG 故障有效隔离,但是对 TLG 与 DS-LLL 两类故障的识别效果欠佳,这主要是由于两者电压波形相似度较高(见图 6),导致 LLE 学习到的嵌入子流形也十分相似,相似的

结果也存在于系统各故障直流电流特征分布当中。因此,本文采用 SVM 来进一步学习故障特征与系统运行状态之间的对应关系,以建立系统故障诊断模型。

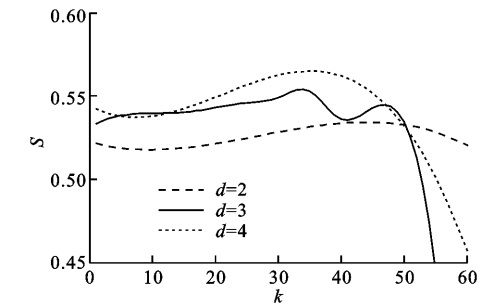


图 8 Silhouette 指标与领域因子 k 的关系

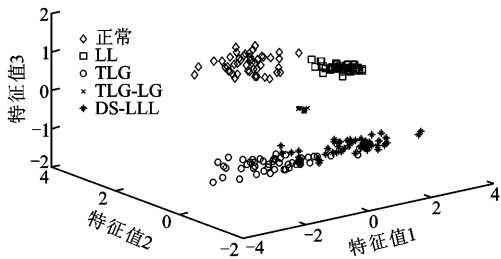


图 9 故障特征空间分布

在模式识别时,SVM 的输入为系统直流电压和电流信号经过 LLE 处理得到的 6 个子流形特征值,模型的核函数选用 RBF 径向基核.每种故障用 50 个样本来训练 SVM 模型,另外 50 个样本用来测试.图 10 为 SVM 模型的训练结果可视化图形,从中可以看出,SVM 准确地将训练样本集分成了 5 个聚类。

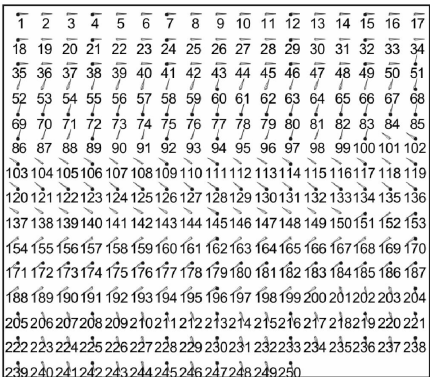


图 10 故障模式识别 SVM 训练结果

表 1 为 SVM 模型测试结果,测试中对比了 LLE-SVM 与 ICA-LLE-SVM 的诊断结果.从表 1

中可知,去噪处理能够明显地提高故障诊断识别精度,这是由于 ICA 滤去了噪声源,使得盲分离后的信号更具有征兆代表性,而对直流电压和电流波形特征的融合识别精度明显高于单独处理,这说明多源信息融合更能反映系统故障特征,提高诊断的可靠性。

表 1 SVM 诊断模型测试结果

诊断方法	输入特征 向量个数	训练 精度/%	测试 精度/%
LLE-SVM	3	60.0	60.0
	6	80.0	80.0
ICA-LLE-SVM	3	83.2	83.6
	6	100.0	100.0

4 结 论

本文研究了基于 ICA 与 LLE 的 VSC-HVDC 系统故障诊断方法,利用多通道盲信号分离来有效提取系统故障特征,并应用 SVM 实现故障模式的准确辨识,数字仿真结果验证了本文算法的有效性.主要结论有:①加性噪声大于 20 dB 时,FastICA 能够有效滤除干扰信号;②LLE 能够有效提取重要故障特征,且 ICA-LLE 故障检测比 LLE 更准确可靠;③多源特征融合能提高 SVM 模型故障分类精度。

参考文献:

[1] WEIMERS L. HVDC light: a new technology for a better environment[J]. IEEE Power Engineering Review, 1998, 18 (8): 19-20.

[2] 颜秉勇,刘喜梅,田作华,等. 协同滤波器和支 持向量机在 HVDC 系统故障诊断中的应用[J]. 中国电机工程学报, 2008, 28(1): 23-29.

YAN Bingyong, LIU Ximei, TIAN Zuohua, et al. Application of consensus filter and SVM in fault diagnosis for HVDC systems [J]. Proceedings of the CSEE, 2008, 28(1): 23-29.

[3] NARENDRA K G, SOOD V K, KHORASANI K, et al. Application of a radial basis function (RBF) neural network for fault diagnosis in a HVDC system[J]. IEEE Trans on Power Systems, 1998, 13(1): 177-183.

[4] LAI L L, NDEH-CHE F, CHARI T. HVDC system fault diagnosis with neural networks[C]// Fifth European Conference on Power Electronics and Applications. Brighton, UK: [s. n.], 1993: 145-150.

[5] 张家奇, 吴坚, 于飞. 基于神经网络的 HVDC 系统故

- 障诊断[J]. 高电压技术, 2006, 32(5): 65-68.
- ZHANG Jiaqi, WU Jian, YU Fei. Neural network based fault diagnosis in HVDC system [J]. High Voltage Engineering, 2006, 32(5): 65-68.
- [6] 于飞, 李升娟, 吕冬梅, 等. 支持向量机在高压直流输电系统故障诊断中的应用[J]. 仪器仪表学报, 2006, 27(6): 1734-1736.
- YU Fei, LI Shengjuan, LÜ Dongmei, et al. Application of support vector machine for fault diagnosis in a HVDC system [J]. Chinese Journal of Scientific Instrument, 2006, 27(6): 1734-1736.
- [7] 张建明, 曾建武, 谢磊, 等. 基于粗糙集的支持向量机故障诊断[J]. 清华大学学报: 自然科学版, 2007, 47(S2): 1174-1177.
- ZHANG Jianming, ZENG Jianwu, XIE Lei, et al. Application of support vector machine for fault diagnosis in a HVDC system [J]. Journal of Tsinghua University: Science & Technology, 2007, 47(S2): 1174-1177.
- [8] HYVÄRINEN A. Fast and robust fixed-point algorithm for independent component analysis [J]. IEEE Trans on Neural Network, 1999, 10(3): 626-634.
- [9] 司文荣, 李军浩, 梁永春, 等. 采用独立成分分析的局部放电脉冲双通道提取技术研究[J]. 西安交通大学学报, 2008, 42(10): 1264-1269.
- SI Wenrong, LI Junhao, LIANG Yongchun, et al. Pulse extraction for partial discharge with double channels with independent component analysis [J]. Journal of Xi'an Jiaotong University, 2008, 42(10): 1264-1269.
- [10] 高军峰, 郑崇勋, 王沛. 基于独立成分分析和流形学习的眼电伪差去除[J]. 西安交通大学学报, 2010, 44(2): 113-118.
- GAO Junfeng, ZHENG Chongxun, WANG Pei. Real-time removal of ocular artifacts from EEG signals using ICA and manifold algorithm [J]. Journal of Xi'an Jiaotong University, 2010, 44(2): 113-118.
- [11] ROWEIS S, SAUL L. Nonlinear dimensionality reduction by locally linear embedding [J]. Science, 2000, 290(5500): 2323-2326.
- (编辑 杜秀杰)