

集成主角及其在故障检测中的应用

甘良志^{1†}, 张士诚², 李旭超¹

(1. 江苏师范大学 电气工程及自动化学院, 江苏 徐州 221116; 2. 江苏师范大学 数学科学学院, 江苏 徐州 221116)

摘要: 为了解决稀疏化造成的核主角不稳定问题, 提出了集成主角方法. 集成主角求再生核希尔伯特空间的多组近似基, 将核主角问题极值向量的解空间限定在近似基张成的空间求核主角, 然后集成特征值. 利用集成主角(ensemble principal angle, EPA)可以对复杂环境下的多变量工业过程进行在线故障检测. 最后本文通过在Tennessee Eastman数据集上的实验, 对集成主角在故障检测中的应用进行了说明.

关键词: 主角; 集成学习; 故障检测; 无监督学习

中图分类号: TP273 **文献标识码:** A

Ensemble principal angle and its application to fault detection

GAN Liang-zhi^{1†}, ZHANG Shi-cheng², LI Xu-chao¹

(1. School of Electrical Engineering and Automation, Jiangsu Normal University, Xuzhou Jiangsu 221116, China;

2. School of Mathematical Sciences, Jiangsu Normal University, Xuzhou Jiangsu 221116, China)

Abstract: Ensemble principal angle is proposed to deal with the instability of a sparse kernel principal angle. Groups of approximate basis are found in the Reproducing Kernel Hilbert Space. Eigenvectors for the principal angle problem are limited to the spaces spanned by the approximate basis. Eigenvalues in different subspace are integrated to make up for the sparsity. Ensemble principal angle (EPA) can be applied to online multivariable process for fault detection in complicated conditions. An example is given to illustrate the application in fault detection by performing experiments on the tennessee eastman data set.

Key words: principal angle; ensemble learning; fault detection; unsupervised learning

1 引言(Introduction)

统计学习理论为机器学习提供了重要的理论基础和实际方法: 它明确指出了结构风险与经验风险的关系, 并且给出了控制结构风险的途径; 同时, 它还引入核函数方法构造了支持向量机以解决非线性问题^[1]. 由于核函数方法解决非线性问题方便有效, 一时之间核主分量分析(kernel principal component analysis, KPCA)^[2]、核典型相关分析(kernel canonical correlation analysis, KCCA)^[3]、核独立分量分析(kernel independent component analysis, KICA)^[4]等纷纷涌现. Wolf L和Shashua A适时提出了核主角方法(kernel principal angle, KPA)^[5]. 这些核学习方法需要解决两个问题: 第一, 避免样本增加引起的核矩阵维数灾难; 第二, 预防过学习. 过学习问题通常有两种解决方法, 即正则化方法和稀疏化方法. 正则化方法无助于核矩阵维数的缩减, 所以稀疏化方法备受关注^[6-8]. 由于稀疏化方法很多时候牺牲了学习机的稳定性^[9], 所以本文提出集成主角(ensemble principal angle, EPA).

计算机通过执行一定的程序获得按照某种标准评价的性能改进. 监督学习通过外部教师为学习机提供学习目标. 无监督学习的目的是发现数据中的重要模式和特征. 较早的集成学习方法用在监督学习中^[10], 通过奖优惩劣提高学习机的性能. 无监督学习由于无法判断单次学习的效果, 很难实现集成, 集成学习的研究成果集中在监督学习方面. 集成主角是一种非监督学习方法, 它在确保简单学习器多样性的前提下, 尽力保障每个简单学习器性能. 集成学习理论证明, 如果能够确保简单学习器的多样性和有效性, 集成学习机的性能就能得到提高^[11]. EPA借鉴核学习方法, 在无限维特征空间中求样本集的若干组近似基, 然后将核主角问题极值向量的解空间限定在近似基张成的空间求核主角, 最后平均集成.

2 核主角(Kernel principal angle)

文献 [12]首先介绍了计算子空间夹角的方法, 为解决非线性问题, Wolf等人^[5]在此基础上提出了核主

角的计算方法. 考虑两个相同维数的零均值的随机变量 X 和 Y . $\{\mathbf{a}_1; \dots; \mathbf{a}_N\}$, $\{\mathbf{b}_1; \dots; \mathbf{b}_M\}$ 是随机变量 X 和随机变量 Y 的两个样本集. 用它们构造两个矩阵 $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_N]$ 和 $\mathbf{B} = [\mathbf{b}_1, \dots, \mathbf{b}_M]$. 两个空间的主角被定义为 θ , 满足^[5]

$$\begin{aligned} \cos \theta &= \max_{\mathbf{u} \in U_A} \max_{\mathbf{v} \in U_B} \mathbf{u}^T \mathbf{v}, \\ \text{s.t. } \mathbf{u}^T \mathbf{u} &= \mathbf{v}^T \mathbf{v} = 1. \end{aligned} \quad (1)$$

$\cos \theta$ 的值通常被称为矩阵对 (A, B) 的典型相关值 (canonical correlations); u, v 被称作极值向量.

为了求解方便, 极值问题(1)重写为

$$\begin{aligned} \cos \theta &= \max_{\mathbf{x}, \mathbf{y}} \mathbf{y}^T \mathbf{B}^T \mathbf{A} \mathbf{x}, \\ \text{s.t. } \|\mathbf{A} \mathbf{x}\|_2^2 &= 1, \|\mathbf{B} \mathbf{y}\|_2^2 = 1. \end{aligned} \quad (2)$$

问题(1)几何意义明确, 问题(2)是典型的等式约束优化问题. 用Lagrangian方法求解问题(2)时构造函数: $L = \mathbf{y}^T \mathbf{B}^T \mathbf{A} \mathbf{x} - \alpha(\|\mathbf{A} \mathbf{x}\|_2^2 - 1) - \beta(\|\mathbf{B} \mathbf{y}\|_2^2 - 1)$, 对变量求导并令其等于零, 可得

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{y}} &= \mathbf{B}^T \mathbf{A} \mathbf{x} - 2\beta \mathbf{B}^T \mathbf{B} \mathbf{y} = 0, \\ \frac{\partial L}{\partial \mathbf{x}} &= \mathbf{A}^T \mathbf{B} \mathbf{y} - 2\alpha \mathbf{A}^T \mathbf{A} \mathbf{x} = 0, \end{aligned}$$

其中 α 和 β 具有惟一性, 所以得到以下表达式:

$$\mathbf{x} = \alpha (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{B} \mathbf{y}, \quad (3)$$

$$\mathbf{y} = \beta (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \mathbf{A} \mathbf{x}. \quad (4)$$

表达式(3)-(4)互相代入得到式(5)-(6):

$$(\mathbf{B}^T \mathbf{B})^{-1} (\mathbf{B}^T \mathbf{A}) (\mathbf{A}^T \mathbf{A})^{-1} (\mathbf{A}^T \mathbf{B}) \mathbf{y} = \lambda^2 \mathbf{y}, \quad (5)$$

$$(\mathbf{A}^T \mathbf{A})^{-1} (\mathbf{A}^T \mathbf{B}) (\mathbf{B}^T \mathbf{B})^{-1} (\mathbf{B}^T \mathbf{A}) \mathbf{x} = \lambda^2 \mathbf{x}. \quad (6)$$

式(5)-(6)合并后成为广义特征值问题:

$$\begin{bmatrix} 0 & \mathbf{B}^T \mathbf{A} \\ \mathbf{A}^T \mathbf{B} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ \mathbf{x} \end{bmatrix} = \lambda \begin{bmatrix} \mathbf{B}^T \mathbf{B} & 0 \\ 0 & \mathbf{A}^T \mathbf{A} \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ \mathbf{x} \end{bmatrix}. \quad (7)$$

方程(7)是计算主角的标准方程. 主角几何意义明确, 计算方便, 但是在工程应用中长期受到限制. 从主角的几何意义不难知道, 主角是由两个样本张成空间的最小夹角. 当样本 $\{\mathbf{a}_1; \dots; \mathbf{a}_N\}$ 和 $\{\mathbf{b}_1; \dots; \mathbf{b}_M\}$ 数量之和超过样本 \mathbf{a}_i (或 \mathbf{b}_i) 的维数时, 空间

$$U_A = \text{span}\{\mathbf{a}_1; \dots; \mathbf{a}_N\},$$

$$U_B = \text{span}\{\mathbf{b}_1; \dots; \mathbf{b}_M\}$$

必然重叠, 最终最小夹角为零, 主角计算没有意义. 所以最初的主角只是应用于图像识别等高维数、小样本学习领域.

Wolf等人给出的KPA算法^[5]解决了以上问题. KPA通过非线性映射将样本映射到RKHS空间中: $\mathbf{a}_k \rightarrow \varphi(\mathbf{a}_k)$. 在无限维的空间中, 自然可以避免

$$\varphi(U_A) = \text{span}\{\varphi(\mathbf{a}_1); \dots; \varphi(\mathbf{a}_N)\}$$

与 $\varphi(U_B) = \text{span}\{\varphi(\mathbf{b}_1); \dots; \varphi(\mathbf{b}_M)\}$ 相交的问题. 并且, 只要在特征空间中找到满足条件 $k(\mathbf{a}_i, \mathbf{a}_j) = \varphi(\mathbf{a}_i)^T \varphi(\mathbf{a}_j)$ 的核函数, 不需要显式描述 $\varphi(\mathbf{a}_k)$, 而是利用核函数在训练过程中获得核矩阵. 由于支持向量机的普及, 找到合适的核函数并不困难. 将无限维空间的样本向量表示为

$$\Phi_A = [\varphi(\mathbf{a}_1); \dots; \varphi(\mathbf{a}_N)],$$

$$\Phi_B = [\varphi(\mathbf{b}_1); \dots; \varphi(\mathbf{b}_M)].$$

极值问题(2)可以写为^[5]:

$$\begin{aligned} \cos \theta &= \max_{\mathbf{x}, \mathbf{y}} \mathbf{y}^T \Phi_B^T \Phi_A \mathbf{x}, \\ \text{s.t. } \|\Phi_A \mathbf{x}\|_2 &= 1, \|\Phi_B \mathbf{y}\|_2 = 1, \end{aligned} \quad (8)$$

式(8)在特征空间描述了主角计算问题. $\Phi_A \mathbf{x}$ 和 $\Phi_B \mathbf{y}$ 分别是空间 $\text{span}\{\varphi(\mathbf{a}_1); \dots; \varphi(\mathbf{a}_M)\}$ 和空间 $\text{span}\{\varphi(\mathbf{b}_1); \dots; \varphi(\mathbf{b}_N)\}$ 中的向量, 其单位化后的内积, 也就是它们的夹角余弦值. 单位向量的夹角余弦越趋于1, 这两个向量越是近似相等, 也就是说两个空间十分接近. 所以主角描述了两个空间的相似程度. 与式(1)到式(7)的变换类似, 极大值问题(8)转换为广义特征值问题:

$$\begin{bmatrix} 0 & \Phi_B^T \Phi_A \\ \Phi_A^T \Phi_B & 0 \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ \mathbf{x} \end{bmatrix} = \lambda \begin{bmatrix} \Phi_B^T \Phi_B & 0 \\ 0 & \Phi_A^T \Phi_A \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ \mathbf{x} \end{bmatrix}. \quad (9)$$

利用核函数可以方便地计算出方程(9)中的 $\Phi_A^T \Phi_B$, $\Phi_B^T \Phi_A$, $\Phi_A^T \Phi_A$ 和 $\Phi_B^T \Phi_B$, 它们构成的方阵的维数为 $M + N$. 其中 M 和 N 分别是两个数据集的样本数, 与特征空间的维数无关. 所以方程(9)特征矩阵的规模是由样本数决定的, 显然样本数太大, 求解就会困难. 另一方面, 也容易出现过学习: 本来差异很大的两个数据集, 计算获得的主角却比较小, 错误地显示两个数据集是高度相似的. 为了避免这种情况, 本文需要发展稀疏核主角 (sparse kernel principal angle, SKPA) 和EPA.

3 集成主角 (Ensemble principal angle)

3.1 稀疏极值向量 (Sparse extreme vector)

极值问题(8)清楚地揭示出极值向量是样本的线性组合. 文献 [13]指出, 如果能够求出样本 $\{\varphi(\mathbf{a}_i)\}_{i=1}^N$ 和 $\{\varphi(\mathbf{b}_i)\}_{i=1}^M$ 的基, 用这两组基来描述极值向量应该可以避免过学习; 但是在无限维特征空间中 $\{\varphi(\mathbf{a}_i)\}_{i=1}^N$ 和 $\{\varphi(\mathbf{b}_i)\}_{i=1}^M$ 的基不可能稀疏, 所以通过求近似基获得稀疏性.

求样本集 $\{\varphi(\mathbf{x}_i)\}_{i=1}^N$ 的近似基^[13]:

Step 1 建立集合 \mathbf{X}_I 和 \mathbf{X}_D (初始时为空集), 随机从 $\{\varphi(\mathbf{x}_i)\}_{i=1}^N$ 抽取一个样本放入 \mathbf{X}_I .

Step 2 对 $\{\varphi(\mathbf{x}_i)\}_{i=1}^N$ 中的每个 $\varphi(\mathbf{x}_k)$, 令

$$\bar{\varphi}(\mathbf{x}_k) = \varphi(\mathbf{x}_k) - \sum_{\varphi(\mathbf{x}_i) \in \mathbf{X}_I} \lambda_i \varphi(\mathbf{x}_i),$$

计算

$$\min \bar{\varphi}^T(\mathbf{x}_k) \bar{\varphi}(\mathbf{x}_k). \quad (10)$$

Step 3 如果 $\min \bar{\varphi}^T(\mathbf{x}_k) \bar{\varphi}(\mathbf{x}_k) \leq \varepsilon$, 认为 $\varphi(\mathbf{x}_k) \in \mathbf{X}_D$, 因为 \mathbf{X}_I 可以描述 $\varphi(\mathbf{x}_k)$; 否则认为 $\varphi(\mathbf{x}_k) \in \mathbf{X}_I$ 因为 \mathbf{X}_I 不能描述 $\varphi(\mathbf{x}_k)$, ε 是控制线性描述误差的参数。

Step 4 如果 $k = N$ 结束程序; 否则 $k = k + 1$, 返回 Step 2.

3.2 EPA算法(EPA algorithm)

极大值问题(8)的极值向量 $\Phi_A \mathbf{x}$ 和 $\Phi_B \mathbf{y}$ 分别是 Φ_A 和 Φ_B 的线性组合. 将 Φ_A 分解为 Φ_{AI} 和 Φ_{AD} (同理可以将 Φ_B 分解为 Φ_{BI} 和 Φ_{BD}), 如果认为 Φ_{AI} 和 Φ_{BI} 分别是 Φ_A 和 Φ_B 的一组基, 自然就可以将极大值问题(8)改写为

$$\begin{aligned} \cos \theta &= \max_{\mathbf{x}, \mathbf{y}} \mathbf{y}^T \Phi_{BI}^T \Phi_{AI} \mathbf{x}, \\ \text{s.t. } &\|\Phi_{AI} \mathbf{x}\|_2 = 1, \|\Phi_{BI} \mathbf{y}\|_2 = 1. \end{aligned} \quad (11)$$

极大值问题(11)等价于广义特征值问题

$$\begin{aligned} &\begin{bmatrix} 0 & \Phi_{BI}^T \Phi_{AI} \\ \Phi_{AI}^T \Phi_{BI} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ \mathbf{x} \end{bmatrix} = \\ &\lambda \begin{bmatrix} \Phi_{BI}^T \Phi_{BI} & 0 \\ 0 & \Phi_{AI}^T \Phi_{AI} \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ \mathbf{x} \end{bmatrix}, \end{aligned} \quad (12)$$

可以肯定, Φ_{AI} 和 Φ_{BI} 的列数分别小于 Φ_A 和 Φ_B 的列数, 特征矩阵的维数缩小了. $\Phi_{AI} \mathbf{x}$ 和 $\Phi_{BI} \mathbf{y}$ 都具有稀疏性, 所以方程(11)被称为 SKPA 问题^[13]. SKPA 最大的缺陷在于描述极值向量使用了近似基计算, 这可能造成精度降低和不稳定. 当核函数选择径向基高斯函数时, 特征空间实际是一个无限维数的空间, 为了获得稀疏性, 不得不采用截断误差参数, 因而引入了计算误差. 实验证明, 稀疏化确实引起主角数值的波动. 文献[9]从理论上分析了稀疏化可能引起的不稳定, 并且提出这种不稳定是必须付出的代价. 由此, 集成学习自然进入视野. 只要近似基能够对数据集做较高精度的描述, 就可以对 SKPA 集成. 因为从近似基算法可知, 每组近似基都可以以小于截尾误差的精度线性描述样本集中的每一个样本, 具备近似相等的描述全体样本的能力, 也就应当具有相同的权重, 所以采取平均集成是可行的. 在这一指导思想下, 有了 EPA 算法:

Step 1 设定集成数 Num、计算近似基的截尾误差参数 ε .

Step 2 调用近似基算法将 Φ_A 分裂为两个集合 \mathbf{A}_d 和 \mathbf{A}_i , 同样地将 Φ_B 分裂为两个集合 \mathbf{B}_d 和 \mathbf{B}_i .

Step 3 利用核函数和近似基计算矩阵 $\mathbf{B}_i^T \mathbf{A}_i$,

$\mathbf{B}_i^T \mathbf{B}_i$, 以及 $\mathbf{A}_i^T \mathbf{B}_i$ 和 $\mathbf{A}_i^T \mathbf{A}_i$, 并根据式(11)求解特征值问题, 其中最大的特征值即为最小核主角的余弦.

Step 4 返回 Step 2, 注意为了保持 \mathbf{A}_i 和 \mathbf{B}_i 的多样性, 后一次求近似基时, 优先从 \mathbf{A}_d 和 \mathbf{B}_d 中选择样本. 共重复 Num 次计算.

Step 5 求主角(最大的特征值)的 Num 次平均值即为集成主角余弦值.

仔细分析 EPA 算法, 可得出以下结论: 首先, KPA 描述极值向量依赖于全体样本, 不仅特征矩阵维数高、计算量大, 而且可能导致过学习; 其次, 稀疏化方法将极值向量限定在近似基张成的空间, 减小了计算量, 同时根据统计学习理论可知, 与参数成线性关系的函数集的 VC 维等于参数的个数, 所以稀疏化方法能够控学习能力, 预防过学习; 最后稀疏化方法可能遗漏数据集的某些特性(通常称作学习不足或欠学习), 可以通过反复多次的稀疏-集成加以弥补.

4 EPA应用于故障检测(Ensemble principal angle for fault detection)

故障是指为了保证系统的正常工作不能允许的状态偏离, 包括系统参数改变, 超范围的扰动, 执行或检测机构的非正常工作等. 故障检测是通过传感器检测的数据判断系统是否有故障发生. 故障检测尤其是实时的在线故障检测对安全高效生产具有重大意义^[14].

EPA 正好可以作为在线故障检测的工具. KPA 将故障检测范围从线性系统扩展到非线性系统, 摆脱了数据集数量与维数之间的束缚; EPA 则可以提高故障检测的稳定性和精度. TE 过程是一个多变量的复杂化工过程, Downs 和 Vogel 为了方便试验研究, 为该过程建立了等效模型, 并在网络公开, 作为故障检测的标准(benchmark)问题, 经常被用于测试故障检测方法的有效性^[8, 15].

研究人员模拟了 22 种不同工作状态下的生产过程, 每种工作状态运行两次, 运行过程中每 3 分钟对关键的 52 个过程变量采样一次, 每次的采样结果构成一个样本, 所以每种工作状态测得 2 个数据集, 将 22 种工况的数据按顺序编号, 第 0 组的两个数据集是正常工作下采集的, 将其中的训练数据集记作 train00, 测试数据集记作 test00. 其他 21 组数据都是在不同的故障状态下测得, 将它们编号为故障 01-21.

假设经过充分积累获得一个信息充分的无故障的数据集 $\mathbf{D}_0 = \{\mathbf{x}_i^0\}_{i=1}^N$, 用 EPA 检测在线采集的某数据是否为故障数据的流程如下.

EPA 在线故障检测流程:

Step 1 标准化数据集, 包括中心化、归一化. 对无故障数据集 $\mathbf{D}_0 = \{\mathbf{x}_i^0\}_{i=1}^N$ 每个样本, 先计算平均值 $\bar{\mathbf{X}} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i^0$ 和中心偏离值 $\tilde{\mathbf{x}}_i^0 = (\mathbf{x}_i^0 - \bar{\mathbf{X}})$, 再

作归一化处理,取每个过程变量的最大和最小中心偏离值,分别构成向量 \mathbf{X}_{\max} 和 \mathbf{X}_{\min} ,计算归一化值 $\mathbf{x}_i^c = \tilde{\mathbf{x}}_i^0 \div (\mathbf{X}_{\max} - \mathbf{X}_{\min})$.这里的运算符“ \div ”表示两个向量对应位置的元素相除,用 \mathbf{x}_{ci}^0 记录标准化样本数据.标准化处理后的无故障数据集记作 $\mathbf{D}_{c0} = \{\mathbf{x}_1^c, \dots, \mathbf{x}_N^c\}$.

Step 2 计算近似基:将标准化处理后的特征空间中无故障数据集记作 $\varphi(\mathbf{D}_{c0})$,求出 $\varphi(\mathbf{D}_{c0})$ 的 K 组近似基,并将其记作 $\mathbf{A}_i^1, \dots, \mathbf{A}_i^K$,这里 K 是设定的近似基的组数.

Step 3 在线数据处理:按照在线采样顺序获得现场数据集 $\mathbf{B} = \{b_{t-M+1}, b_{t-M+2}, \dots, b_t\}$,其中 M 是采样数据构建的样本长度.对在线获得现场数据集的每个数据进行中心化和归一化,取

$$b_i^c = (b_i - \bar{X}) \div (\mathbf{X}_{\max} - \mathbf{X}_{\min}),$$

其中: $i = t - M + 1, \dots, t$, \bar{X} , \mathbf{X}_{\max} 和 \mathbf{X}_{\min} 均为第1步中在无故障数据集上计算所得的向量,用 b_i^c 构造 $\mathbf{B}_i = [b_{t-M+1}^c, b_{t-M+2}^c, \dots, b_t^c]$.

Step 4 计算主角:对每个 $\mathbf{A}_i^k (k=1, \dots, K)$,求

$$\begin{bmatrix} 0 & \mathbf{B}_i^T \mathbf{A}_i^k \\ \mathbf{A}_i^{kT} \mathbf{B}_i & 0 \end{bmatrix} \begin{bmatrix} y \\ x \end{bmatrix} = \lambda \begin{bmatrix} \mathbf{B}_i^T \mathbf{B}_i & 0 \\ 0 & \mathbf{A}_i^{kT} \mathbf{A}_i^k \end{bmatrix} \begin{bmatrix} y \\ x \end{bmatrix}$$

的特征值,并取其最大值.

Step 5 故障判断:对 K 个最大的特征值求平均值,并判断这个平均值是否小于控制限,由此判断故障是否发生.

利用Tennessee Eastman数据集可以设计实验验证EPA具有比SKPA更好的稳定性和更强的泛化能力.

实验1 EPA的稳定性检验.实验原理:因为集成EPA的真值是无法获得的,只能观察逐渐逼近过程检验EPA的稳定性.所以在设计实验时,从train00独立产生两组近似基,其中一组有300个近似基,称为参照组,用于计算test00的样本与该组每个近似基的余弦值并计算平均值作为EPA的真值,详细算法参见3.2节的EPA算法;另一组有150个近似基,称为检验组,从检验组分别抽取1~150个近似基计算主角平均值,并与300个近似基计算的EPA对比.如果随着检验组近似基数量的增加其主角平均值趋近于参照组EPA,则可以证明EPA稳定.在图1中以虚线描绘的是参照组EPA,横轴是test00的500个样本,实线是部分检验组SKPA或平均值,然后通过检验组不同数目SKPA平均值与参照组EPA的对比来得到结论.子图1(a)和子图1(b)显示了基于检验组的不同近似基计算所得的SKPA,并同EPA对比. SKPA与EPA的显著差异说明SKPA不稳定.子图1(c)和子图1(d)显示随着

近似基数量的增加,检验组中多个SKPA的平均值趋向与EPA重合,说明EPA算法稳定.图1中的波谷意味着系统运行至此受到干扰,所以与正常数据集的相似度降低了,余弦值变小了.

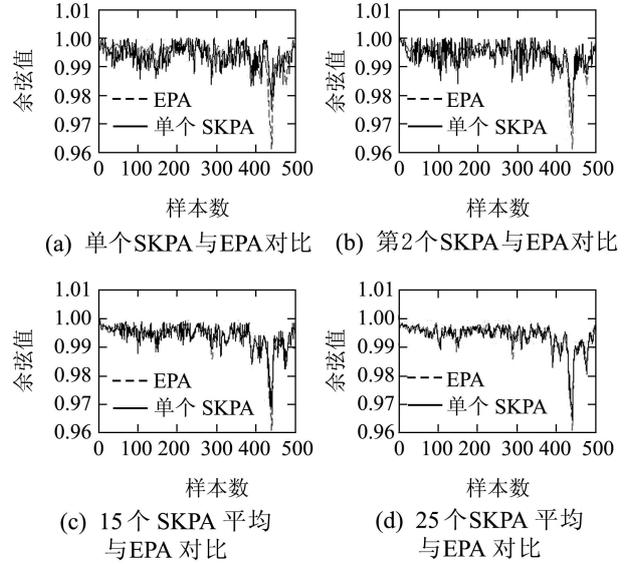


图1 EPA收敛

Fig. 1 Convergence of EPA

图2通过统计图1中虚线与实线的差值减小过程,更直观地验证EPA收敛.图2横轴是用于集成的SKPA的数量,纵轴是标准偏差

$$\sigma_K = \sqrt{\frac{1}{500} [(\cos \theta_i)_K - (\cos \theta_i)_{EPA}]^2},$$

其中: K 是用于集成的近似基的数量,从1~150变化, $(\cos \theta_i)_K$ 是test00的第 i 个样本与检验组的 K 个近似基的余弦平均值, $(\cos \theta_i)_{EPA}$ 是test00的第 i 个样本与参照组的全部近似基的余弦集成值.随着近似基数目增加标准差越来越小,所以认为EPA收敛.

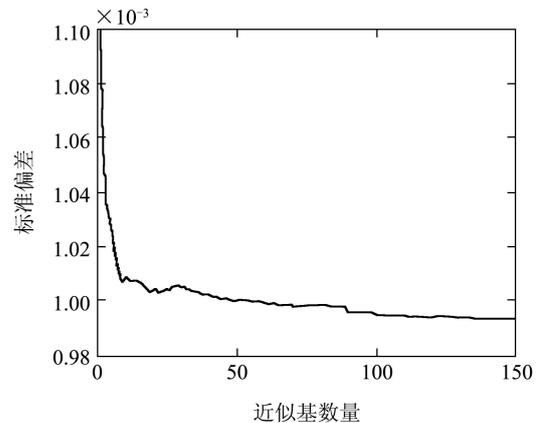


图2 EPA的标准偏差

Fig. 2 Standard deviation of EPA

实验2 EPA的学习精度与泛化能力检验.首先通过分析test00中的样本与train00数据集的相似程

度(用EPA描述), 确定一个控制限. 然后计算现场采样数据(来自test01–test21数据集)与train00数据集的主角余弦集成值, 最后根据控制限进行故障判断. 由于现场采样数据不在test00和train00中, 所以如果EPA能够正确做出检测结论, 就说明它具有良好的泛化能力.

图3显示了集成主角控制限的确定方法. 对于测试数据集test00中的每一个向量, 计算其与训练数据集train00的集成主角; 根据精度要求, 当要求误报警的频率小于1.0%时, 控制限取值为 $D = 0.985$.

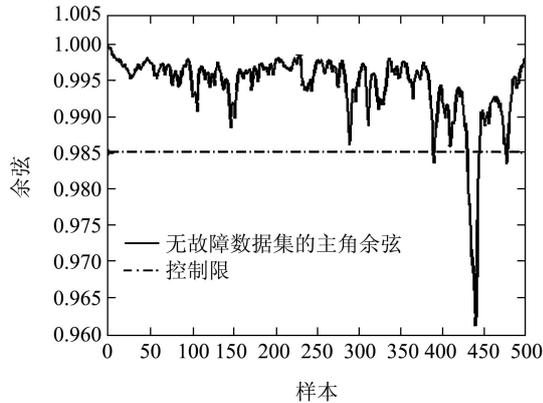
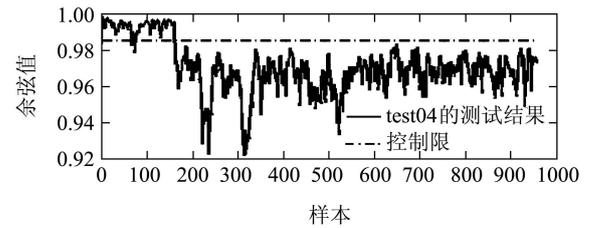


图 3 控制限的确定

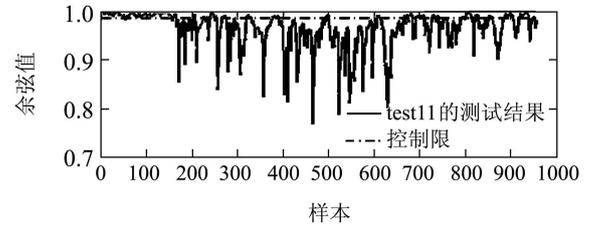
Fig. 3 Decide the limitation

图4显示了故障04和故障11的模拟检测过程. 在确定控制限取 $D = 0.985$ 后, 对于从test04或是test11中选取的待测数据, 计算其与无故障训练数据集train00主角的余弦集成值. 余弦值较大, 笔者认为测试数据与train00接近, 判断为正常. 否则判断为故障

数据.



(a) 故障04的检测结果



(b) 故障11的检测结果

图 4 故障检测过程

Fig. 4 Fault detection process

表1将EPA检测的结果列出, 与文献 [8, 15]的检测结果对比. 表中SKPA的检测结果来自文献 [8], KPCA和IKPCA(改进核主分量分析)的检测结果直接摘自文献 [15], EPA数据是本文提出的检测方法的统计结果. 表1明确显示, 基于EPA故障检测的结果精度全部高于KPCA方法的检测结果: 用EPA与IKPCA、SKPA对7组相同的故障数据进行检测, EPA在6组数据上获得了更精确的结果. 这表明, 集成学习在提高非监督学习机的学习精度和泛化能力.

表 1 故障检测错误率对比

Table 1 Comparison of error ratios

| 统计量 | 测试数据集 | | | | | | |
|---------------|-------|------|------|------|------|------|------|
| | 03 | 04 | 09 | 11 | 13 | 15 | 21 |
| SPE(KPCA) | 1.00 | 1.00 | 1.00 | 0.99 | 0.97 | 0.99 | 1.00 |
| T^2 (KPCA) | 0.97 | 0.98 | 0.96 | 0.70 | 0.21 | 0.91 | 0.72 |
| SPE(IKPCA) | 0.99 | 1.00 | 1.00 | 0.51 | 0.18 | 0.86 | 0.70 |
| T^2 (IKPCA) | 0.75 | 0.83 | 0.83 | 0.12 | 0.15 | 0.74 | 0.33 |
| COSINE(SKPA) | 0.90 | 0.01 | 0.93 | 0.11 | 0.04 | 0.40 | 0.34 |
| COSINE(EPA) | 0.86 | 0.01 | 0.72 | 0.17 | 0.03 | 0.38 | 0.28 |

5 结论(Conclusions)

为了解决SKPA不稳定问题, 获得在线的复杂工业生产过程故障检测有效方法, 本文提出了集成主角算法. 该方法充分利用核函数的特点, 在特征空间中利用线性组合这一技巧, 用比较少的样本以较小的精度损失描述原本由全体样本构成的特征空间. 由近似描述引起的精度降低和不稳定性通过集

成方法得到了解决, 具有更高的学习精度和较强的泛化能力. 该方法在标准测试数据集(TE数据集)在线故障检测实验中, 展现出较高的检测精度, 因此认为是一种优秀的故障检测方法. 特别是通过与SKPA方法的比较, 说明本文提出的核心方法—集成主角能有效提高学习的精度和算法的稳定性. 未对非监督集成学习作深入的理论探讨, 是论文的不

足之处.

参考文献(References):

- [1] VAPNIK V N. *Statistical Learning Theory* [M]. New York: John Wiley and Sons, 1998.
- [2] SCHOLKOPF B, SMOLA A, MULLER K R. Nonlinear component analysis as a kernel eigenvalue problem [J]. *Neural Computation*, 1998, 10(5): 1299 – 1319.
- [3] AKAHO S. A kernel method for canonical correlation analysis [C] // *Proceedings of the International Meeting of the Psychometric Society (IMPS2001)*. Berlin: Springer, 2001.
- [4] BACH F R, JORDAN M I. Kernel independent component analysis [J]. *Journal of Machine Learning Research*, 2002, 3(7): 1 – 48.
- [5] WOLF L, SHASHUA A. Learning over sets using kernel principal angles [J]. *Journal of Machine Learning Research*, 2003, 6(4): 931 – 937.
- [6] HINTON G E, SALAKUTDINOV R R. Reducing the dimensionality of data with neural networks [J]. *Science*, 2006, 313(5786): 504 – 507.
- [7] d' ASPREMONT A, GHAOUI L EI, JORDAN M, et al. A direct formulation for sparse PCA using semidefinite programming [J]. *SIAM Review*, 2007, 49(3): 434 – 448.
- [8] GAN L Z, LIU H K, SHEN X F. Sparse kernel principal angles for online process monitoring [J]. *Journal of Computational Information System*, 2010, 6(5): 1601 – 1608.
- [9] XU H, MANNOR S. Sparse algorithms are not stable: a no-free-lunch theorem [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(1): 187 – 193.
- [10] OPITZ D, MACLIN R. Popular ensemble methods: an empirical study [J]. *Journal of Artificial Intelligence Research*, 1999, 11(1): 169 – 198.
- [11] HANSEN L K, SALAMON P. Neural network ensembles [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1990, 12(10): 993 – 1001.
- [12] BJORK A, GOLUB G H. Numerical methods for computing angles between linear subspaces [J]. *Mathematics of Computation*, 1973, 27(123): 579 – 594.
- [13] 甘良志. 核学习算法与集成方法研究 [D]. 杭州: 浙江大学, 2010. (GAN Liangzhi. *Kernel learning algorithms and ensemble methods* [D]. Hangzhou: Zhejiang University, 2010.)
- [14] 薄翠梅, 柏杨进, 杨海荣, 等. 多切面分类改进独立成份与支持向量机集成故障诊断方法 [J]. *控制理论与应用*, 2012, 29(2): 229 – 234. (BO Cuimei, BAI Yangjin, YANG Hairong, et al. Multi-section classification improving integrated fault diagnosis method based on independent component analysis and support-vector-machines [J]. *Control Theory & Applications*, 2012, 29(2): 229 – 234.)
- [15] GE Z, YANG C, SONG Z. Improved kernel PCA-based monitoring approach for nonlinear processes [J]. *Chemical Engineering Science*, 2009, 64(9): 2245 – 2255.

作者简介:

甘良志 (1974–), 男, 博士, 副教授, 主要研究方向为机器学习,

E-mail: lzh_box@163.com;

张士诚 (1975–), 男, 博士, 副教授, 主要研究方向为信息几何,

E-mail: zhangshicheng@jsnu.edu.cn;

李旭超 (1975–), 男, 博士, 副教授, 主要研究方向为模式识别,

E-mail: bsx7096@sohu.com.