基于数据驱动的多工况过程控制性能评估研究

林晓钟，谢 磊

（浙江大学 智能与系统控制研究所 工业控制技术国家重点实验室，310027 杭州）

摘 要：为了解决如何选择合适的基准数据与实时数据进行控制性能评估问题，提出一种基于多变量分类的方法.用一种综合相似因子来衡量当前实时数据和某一工况下数据的相似性，确定实时数据所属的工况，而后再进行控制性能评估，通过仿真结果验证了新方法的有效性.

关键词：多工况；相似因子；性能评估

中图分类号：TP14 **文献标志码**：A **文章编号**：0367－6234(2012)11－0081－06

The performance assessment of multiple mode MIMO industrial process based on Data-driven controller

LIN Xiao-zhong, XIE Lei

(Institute of Cyber-Systems& Control, State Key Laboratory of Industrial Control Technology, Zhejiang University, 310027 Hangzhou, China)

Abstract: To choose the suitable benchmark data for data-driven controller performance monitoring from data based on kinds of multiple mode in the history database, a new method based on multiple variables classification is proposed. First the method provides a combined similarity factor to measure the similarity between real-time acquisition data and data from some mode in the history database, and then it determines the mode that real-time data belongs to, finally it assesses the controller performance. Simulation study demonstrates the efficiency of the proposed approach.

Key words: multiple mode, similarity factor, performance assessment

随着现代工业及科学技术的迅速发展，现代化的流程工业呈现出规模大、结构复杂、生产单元之间强耦合等特点.在复杂工业过程中，控制回路数目众多，这些控制回路大多在运行初期具有良好的性能，在运行一段时间后，受原料性质、对象特征、优化目标变化以及维护不利等各种因素的影响，控制回路性能将会变差，导致产品的质量和工厂的效益受到影响，所以工业界对控制系统性能要求的提高促进了控制系统性能评估这一领域的发展.最初，很多研究主要聚焦在基于最小方差基准(MVC)的SISO系统的性能监视方面[1－2]研究了MIMO线性反馈控制系统的性能评价，并对控制变量协方差矩阵的理论下界评价进行了定量分析.Huang[3]研究了MIMO前馈反馈控制器的性能评价，通过对操作数据处理得到前馈反馈最小方差从而作为性能监视的基准.大多数基于MVC基准和LQG基准的性能评估方法离不开过程的传递函数模型，而实际过程中过程的模型又不容易精确地获得，使得这些方法的应用受到了限制.

针对基于过程传递函数模型来进行性能评估的不足，近年来学者们提出直接利用过程输出的数据来进行性能评估的方法. 基于数据驱动方法最大的优势在于不需要过程知识的精确信息且输出数据较容易获得，从数据中挖掘出过程的信息，因此该类方法具有很强的实用性.Jie Yu等提出了基于数据驱动的性能评估的基准[4－6]，把一段理想的历史输出数据作为参考数据基准，通过分

收稿日期：2012－02－13.
基金项目：国家自然科学基金资助项目(61134007).
作者简介：林晓钟(1986—)男，硕士研究生；
谢 磊(1979—)男，副教授，博士生导师.
通信作者：林晓钟，shangdonghappy@126.com.

析所监视时段数据和基准时段数据的广义特征值,提出相应的性能优/劣的特征向量. 进一步利用统计推断方法得出特征值在相应特征方向上的置信区间,以及在优/劣子空间下的性能指标,从而评价和监视控制性能的优劣. Xuemin Tian[7] 等提出了一种基于 2 - 范数的协方差性能基准,通过将当前操作工况数据与基准数据进行比较,从而监控当前控制器的性能变化情况. Gudi[8] 等提出了一种基于对应分析的统计基准上的控制器性能评估基准的思想.

然而,以上基于数据驱动的性能评估方法都有一个前提假设:基准数据都来自单一的稳定的工况,而实际生产过程极其复杂,往往不是运行在单一的工况,生产负荷、产品特性、原料组分等的改变,都会导致工况的改变. 所以必须考虑到工况变化引起的基准调整问题,即在多工况下历史数据库中有多组可选的数据作为基准,如何选择恰当的基准是较为关键的问题,其关系到性能评估结果的正确与否. 本文提出在进行性能评估前,通过分类的方法先确定实时数据所属的工况,从历史数据库中选择恰当工况的数据作为基准,再进行性能评估. 采用结合 PCA 相似因子与几何距离相似因子的多变量分类的方法来衡量两组数据矩阵之间的相似性,从而得到较为准确的分类结果,避免了由于基准数据选择的不恰当影响了控制器性能评估的结果. 仿真例子证实了该方法的有效性.

1 基于协方差的数据驱动型的性能评估方法

文献[5]等提出了基于数据驱动的基准. 即选取一段控制器性能较为满意的历史数据集Ⅰ作为基准数据,将需要评估的实时数据作为数据集Ⅱ,定义基于数据驱动的控制器性能评估基准如下:

$$\eta = \frac{E(y_{II}^{\mathrm{T}} y_{II})}{E(y_I^{\mathrm{T}} y_I)} = \frac{|\operatorname{cov}(y_{II})|}{|\operatorname{cov}(y_I)|}. \tag{1}$$

从几何意义上说,$|\operatorname{cov}(y_I)|$ 和 $|\operatorname{cov}(y_{II})|$ 分别表征由数据集 y_I 和 y_{II} 的协方差所张成的超曲面的体积;从代数角度,$|\operatorname{cov}(y_I)|$ 和 $|\operatorname{cov}(y_{II})|$ 综合考虑了协方差阵整体的因素,而并非单纯地考虑其对角线元素. 由分析的结论可知:当 η 越接近于0时,控制器性能越好;当 η 越大时,控制器性能越差. 通过求解如下优化问题寻找 η 最大的方向:

$$\lambda(\boldsymbol{p}) = \underset{p}{\operatorname{argmax}} \frac{\boldsymbol{p}^{\mathrm{T}} \operatorname{cov}(y_{II}) \boldsymbol{p}}{\boldsymbol{p}^{\mathrm{T}} \operatorname{cov}(y_I) \boldsymbol{p}}. \tag{2}$$

对式(2)求解得

$$\operatorname{cov}(y_{\mathrm{II}}) \boldsymbol{P} = \boldsymbol{\lambda} \operatorname{cov}(y_I) \boldsymbol{p}. \tag{3}$$

式中:λ 为广义特征值;$\boldsymbol{p}$ 为对应的特征向量. 其矩阵形式为

$$\operatorname{cov}(y_{\mathrm{II}}) \boldsymbol{P} = \operatorname{cov}(y_I) \boldsymbol{P} \boldsymbol{\Lambda}. \tag{4}$$

式中:$\boldsymbol{P} = [\boldsymbol{p}_1, \boldsymbol{p}_2, \cdots \boldsymbol{p}_q]$,$\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \lambda_2, \cdots \lambda_q)$. 最大的广义特征值 $\lambda_{\max}$ 所对应的特征向量 $\boldsymbol{p}_{\max}$ 为控制性能变得最差所投影的方向. 求上式的行列式得

$$|\operatorname{cov}(y_{II})| \cdot |\boldsymbol{P}| = |\operatorname{cov}(y_I)| \cdot |\boldsymbol{P}| \cdot |\boldsymbol{\Lambda}|. \tag{5}$$

从而得到表征总体性能的指标 I_V 如下:

$$I_v = \frac{|\operatorname{cov}(y_I)|}{|\operatorname{cov}(y_I)|} = |\boldsymbol{\Lambda}| = \prod_{i=1}^{q} \lambda_i. \tag{6}$$

由性能指标(6) 可知:$I_V > 1$ 时,监控的控制器性能比历史基准性能要差;$I_V < 1$ 时,监控的控制器性能比历史基准性能要好;$I_V \approx 1$ 时,监控的控制器性能与历史基准性能较为接近,较小的偏差可认为是统计误差. 只有明显偏离 1 时,才认为性能发生较为显著的变化.

由于特征值的真值 l_i 与其统计值 λ_i 的差 $l_i - \lambda_i$ 服从正态分布 $N(0, 2\lambda_i^2 (f_I^{(i)}/(m-1) + f_{II}^{(i)}/(n-1)))$,每一个特征值的真值 l_i 在取置信度为 $(1-\alpha)\times 100\%$ 时置信区间如下:

$$P\{L(\lambda_i) \leqslant l_i \leqslant U(\lambda_i)\} = 1 - \alpha. \tag{7}$$

对应于

$$P\left\{ -z_{\alpha/2} \leqslant \sqrt{2\left(\frac{f_I^{(i)}}{m-1} + \frac{f_{II}^{(i)}}{n-1}\right)}^{-1} \left(\frac{l_i - \lambda_i}{\lambda_i}\right) \leqslant z_{\alpha/2} \right\} = 1 - \alpha.$$

由此可得 l_i 的置信上限 $U(\lambda_i)$ 和置信下限 $L(\lambda_i)$

$$U(l_i) = \left[1 - z_{\alpha/2}\sqrt{2\left(\frac{f_I}{m-1} + \frac{f_{II}}{n-1}\right)}\right]^{-1} \cdot \lambda_i. \tag{8}$$

$$L(l_i) = \left[1 + z_{\alpha/2}\sqrt{2\left(\frac{f_I}{m-1} + \frac{f_{II}}{n-1}\right)}\right]^{-1} \cdot \lambda_i. \tag{9}$$

当 $L(\lambda_i) > 1$,可以判断总体特征值的真值大于 l,表明监控的控制器性能差于沿此特征值对应的特征向量方向上的基准性能;当 $U(\lambda_i) < 1$,可以判断总体特征值的真值小于 1,表明监控的控制器性能优于沿此特征值对应的特征向量方向上的基准性能;当 $L(\lambda_i) < 1 < U(\lambda_i)$,从统计角度可判断总体特征值真值等于 1,表明监控性能沿此特征值对应的特征向量方向上与基准性能差别不大.

进一步可以确定出由所有大于 1 的 w 个特征值 λ_i 及其所对应的 w 个特征向量,以及由该 w 个特征向量所构成的恶化性能子空间 P_w. 将基准数据集 y_I 和监控数据集 y_{II} 分别往子空间 P_w 上按下式进行投影得到恶化空间数据集 $z_I^{(w)}$ 和 $z_{\mathrm{II}}^{(w)}$

$$z_I^{(w)} = (P_w^{\mathrm{T}} P_w)^{-1} P_w^{\mathrm{T}} y_I, \quad z_{\prod}^{(w)} = (P_w^{\mathrm{T}} P_w)^{-1} P_w^{\mathrm{T}} y_{\mathrm{II}}. \tag{10}$$

同时可获得基于协方差数据驱动的恶化性能子空间的评估指标 I_w

$$I_w = \frac{|\operatorname{cov}(z_{II}^{(w)})|}{|\operatorname{cov}(z_I^{(w)})|} = |\Lambda_w| = \prod_{i=1}^{w}\lambda_i. \quad (11)$$

指标 I_w 给出了所监控的数据与历史基准数据在恶化性能子空间中投影方差的比值，其可作为沿恶化特征方向的控制器性能下降程度的度量指标来判断控制器的性能好坏.

2　正确选择性能基准数据的必要性

文献[5]等所给方法可直观有效地对控制器进行性能评估，但其只考虑评估基准是单一工况的情况，未考虑在多工况下评估基准的选择问题. 由于评估基准选择的不当有可能带来的对控制器性能评估的误判，如无法判断是控制器性能的真实下降还是由于基准选择不当造成的性能下降的结果而得出错误的结论. 以下这个 2 输入 2 输出的例子可以说明未考虑多工况评估基准而得到错误评估结果的问题.

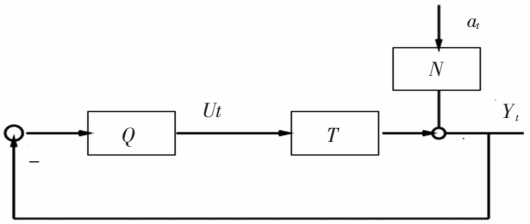


图 1　多变量反馈控制系统示意图

其中，$N = \begin{bmatrix} 0.6z^{-1} & 0 \\ 0 & 0.5z^{-1} \end{bmatrix}$，

白噪声协方差阵

$$\Sigma_a = \begin{bmatrix} 0.01 & 0 \\ 0 & 0.01 \end{bmatrix},$$

$$Q = \begin{bmatrix} 0.5 - 0.2z^{-1} & 0 \\ 0 & 0.25 - 0.2z^{-1} \end{bmatrix}.$$

考虑两种工况：

$$\text{工况一}, T_1 = \begin{bmatrix} \dfrac{z^{-3}}{1-0.9z^{-1}} & \dfrac{z^{-3}}{1-0.8z^{-1}} \\ \dfrac{z^{-3}}{1-0.7z^{-1}} & \dfrac{z^{-4}}{1-0.6z^{-1}} \end{bmatrix};$$

$$\text{工况二}, T_2 = \begin{bmatrix} \dfrac{-z^{-3}}{1-0.9z^{-1}} & \dfrac{z^{-3}}{1-0.8z^{-1}} \\ \dfrac{1.91z^{-3}}{1-0.7z^{-1}} & \dfrac{-z^{-4}}{1-0.6z^{-1}} \end{bmatrix}.$$

取两种工况下输出的 1000 组数据作为历史数据库中的两类基准数据，又假定实际采集的数据为工况二条件下白噪声协方差阵变为 $\Sigma_a = \begin{bmatrix} 0.001 & 0 \\ 0 & 0.01 \end{bmatrix}$ 下的数据，三组数据的二维联合分布情况如图 2 所示. 用文献[5]的方法将实际采集的数据分别与两种工况下基准数据进行评估，得总体特征值阵分别为 $\Lambda_1 = \begin{bmatrix} 5.252 & 0 \\ 0 & 1.2107 \end{bmatrix}$；$\Lambda_2 = \begin{bmatrix} 0.3959 & 0 \\ 0 & 0.3454 \end{bmatrix}$，由评估结果分析可知：若选取工况一的数据作评估基准，在第一个特征方向上，性能显著变差；若选取工况二的数据作评估基准，则在各方向上性能均得到改善. 而实际采集的数据是来自工况二的，由于白噪声的方差变小，性能得到改善是符合事实的.

故应采用工况二的数据作为评估基准才能得出正确的评估结果，倘若随意选择评估基准，则会得到与事实相反的评估结果，造成误判.

从该例不难说明，在用基于数据驱动的方法进行性能评估时，控制器性能基准的选取是十分重要的，若控制性能基准选取不当，将会在很大程度上影响评估结果的准确性.

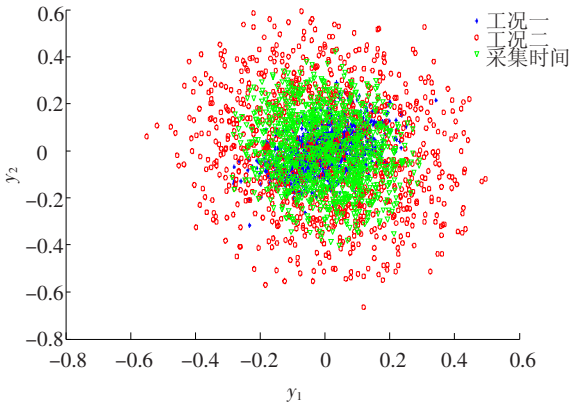


图 2　多变量系统输出变量三组数据二维联合分布

3　基于综合相似因子 SF 的多工况性能评估

上文已阐述了选择性能评估基准的必要性与重要性，针对多变量多工况的条件下，本文提出一种基于综合相似因子距离度量的方法，在进行评估之前，先对所采集的数据与历史数据库中各类工况下的基准数据分别进行匹配，找出所采集数据所属的工况，再进行性能评估，这样方能保证评估结果的正确.

3.1　PCA 相似因子

文献[9－10]给出了一种基于 PCA 相似因子的方法来衡量两类数据矩阵之间的相似性. 假设历史数据阵 H 和实时数据阵 S，均有 n 个变量（数据矩阵列数为 n），它们各自的 PCA 模型中均包含有 k 个主元，$k \leqslant n$，k 按如下原则选取：k 个主元的方差之和占到总方差的 95% 以上，则 H 与 S 相对应的主元子空间分别为 L 和 M，根据文献[9－10]，

PCA 相似因子按如下公式计算：

$$S_{PCA}^{\lambda} = \mathrm{trace}(\boldsymbol{R}^{\mathrm{T}}\boldsymbol{T}\boldsymbol{T}^{\mathrm{T}}\boldsymbol{R}) / \sum_{i=1}^{k} \lambda_i^l \lambda_i^m. \quad (12)$$

式中：

$$\boldsymbol{R} = \boldsymbol{L}\boldsymbol{\Lambda}_l;\ \boldsymbol{T} = \boldsymbol{M}\boldsymbol{\Lambda}_m;$$

$$\boldsymbol{\Lambda} = \begin{bmatrix} \sqrt{\lambda_1} & 0 & 0 & 0 \\ 0 & \sqrt{\lambda_1} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \sqrt{\lambda_1} \end{bmatrix}.$$

特征值 λ_i^l 和 λ_i^m 分别对应于 L 和 M 的第 i 个主元成分，且 $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_k$，$0 \leqslant S_{PCA}^{\lambda} \leqslant 1$，且两组数据越相似，$S_{PCA}^{\lambda}$ 越大.

通过式(12)得到的 PCA 相似因子可以衡量两数据矩阵的相似程度，从几何意义上说，S_{PCA}^{λ} 表征的是两组数据矩阵在空间中分布形状的相似程度，却未考虑两组数据之间的几何距离因素，因此还需引入表征几何距离的度量因子才能得到更加准确的分类结果.

3.2 几何距离相似因子

若两组数据矩阵在空间分布上具有类似的几何形状，却相互隔着一定的空间距离，此时 PCA 相似因子对区分两组数据矩阵几乎不起什么作用，引入几何距离相似因子可以将该条件下的数据阵区分开来.

定义历史数据阵和实时数据阵的中心向量 $\overline{\boldsymbol{X}_H}$，$\overline{\boldsymbol{X}_S}$ 分别为各自的变量采样均值构成的向量，即

$$\overline{\boldsymbol{X}_H} = \frac{1}{m_H}\sum_{i=1}^{m_H} \boldsymbol{X}_i, \boldsymbol{X}_i \in \boldsymbol{H}, \overline{\boldsymbol{X}_S} = \frac{1}{m_S}\sum_{i=1}^{m_S} \boldsymbol{X}_i, \boldsymbol{X}_i \in \boldsymbol{S}. \quad (13)$$

m_H 和 m_S 分别为数据阵 $\boldsymbol{H}$ 和 $\boldsymbol{S}$ 的采样数，$\boldsymbol{X}_i$ 为数据阵的第 i 次采样值. 两组中心向量之间的马氏距离定义为

$$\boldsymbol{\Phi} = \sqrt{(\boldsymbol{X}_H - \boldsymbol{X}_S)^{\mathrm{T}} \boldsymbol{\Sigma}_S^{*\,-1} (\boldsymbol{X}_H - \boldsymbol{X}_S)}. \quad (14)$$

式中：$\boldsymbol{\Sigma}_S^{*\,-1}$ 为数据 S 协方差阵 $\boldsymbol{\Sigma}_S^{*}$ 的伪逆阵，该伪逆阵能够通过奇异值分解而得，用于计算伪逆阵的奇异值的个数为主元数目 k，则几何距离因子由下式计算而得

$$S_{dist} = \sqrt{\frac{2}{\pi}} \int_{\Phi}^{\infty} \mathrm{e}^{-z^2/2} \mathrm{d}z = 2 \times \left(1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\Phi} \mathrm{e}^{-z^2/2} \mathrm{d}z\right). \quad (15)$$

几何距离因子很好地补充了单用 PCA 因子来衡量两类数据相似性的不足之处.

3.3 综合相似因子 SF

在实际应用中，必须综合考虑这两类度量因子的因素，一种简单而有效的方法是将两类因子进行加权组合构成综合相似因子

$$SF = \alpha S_{PCA}^{\lambda} + (1 - \alpha) S_{dist}. \quad (16)$$

其中：$0 \leqslant \alpha \leqslant 1$，若无任何先验知识的情况下，$\alpha$ 可取 0.5. 实际应用中的具体步骤如下：计算实时采集的数据阵与历史数据库中的每一类工况下的数据阵的 SF 值，选取 SF 最大值所对应的那一类工况作为实时采集数据所属的工况，而后再按照基于数据驱动的方法进行性能评估.

3.4 多工况基准数据的性能评估

针对历史数据库中有多种工况下的历史数据，将实时采集的数据阵与历史数据库中的每一类工况的基准数据一一比对，找出相似程度最高的那一类工况，从而将该实时采集的数据判定为那一类工况，而后再进行性能评估.

综上所述，多工况过程基准数据的性能评估过程为

1)收集所有可能工况的历史基准数据集.

2)分别计算实时采集的数据阵与每一类工况基准数据阵的 S_{PCA}^{λ} 因子与 S_{dist} 因子.

3)根据已有的先验知识，选择一定的权重 α，计算 SF 综合相似因子.

4)选取 SF 最大值所对应的那一类工况作为实时采集数据所属的工况.

5)将实时采集数据与其所属工况的历史基准数据进行基于数据驱动型的性能评估.

4 仿真分析

4.1 文中第三部分的仿真例子

第三部分所给的例子三组输出数据如图 3 所示. 将 SF 综合相似因子应用到第三部分所给的例子中，将实时采集的数据分别与工况一(Ⅰ)和工况二(Ⅱ)下的数据用 SF 综合相似因子进行衡量，得到表 1. 由表中结果可以看出，实时采集的数据与工况二下的数据 SF 因子最大，故实时数据应被归类与工况二下的数据，应选取历史数据库中工况二下的数据作为性能评估基准，再进行性能评估. 该例子证实了用 SF 综合相似因子来衡量数据矩阵之间相似程度的有效性.

表 1 采集数据阵与两类工况下数据阵的相似性比较

实时采集与工况	S_{PCA}^{λ}	S_{dist}	$SF = 0.5S_{PCA}^{\lambda} + 0.5S_{dist}$
Ⅰ	0.862	0.997 0	0.930 0
Ⅱ	1.000	0.997 3	0.998 6

4.2 一个 4 输入 4 输出的仿真例子

由 McNabb 和 Qin[11] 提供的 4 输入 4 输出的仿真例子如下：

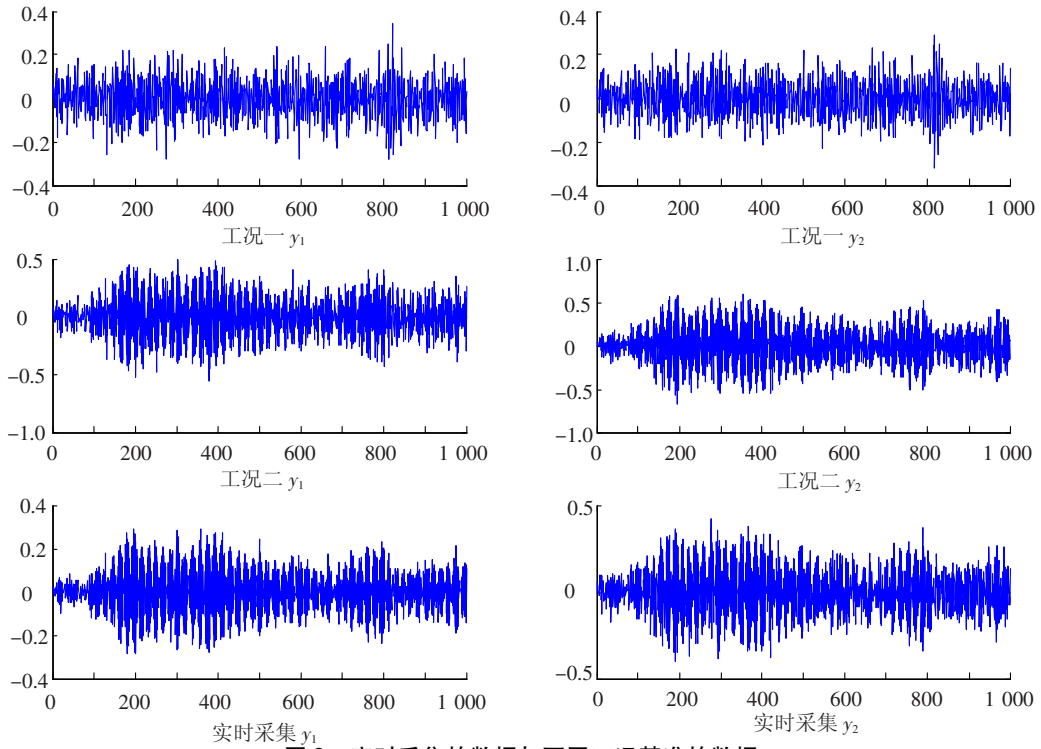


图 3　实时采集的数据与不同工况基准的数据

$$N=\begin{bmatrix}A&0&0&0\\0&A&0&0\\0&0&A&0\\0&0&0&A\end{bmatrix},$$

式中：$A=\dfrac{1-0.187\,5z^{-1}}{1-0.987\,5z^{-1}}$.

其中白噪声协方差阵

$$\boldsymbol{\Sigma}_a=\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix},\quad \boldsymbol{Q}=\begin{bmatrix}\dfrac{17.136-16.32z^{-1}}{20-20z^{-1}}&0&0&0\\0&\dfrac{10.625-10z^{-1}}{16-16z^{-1}}&0&0\\0&0&\dfrac{0.710\,24-0.526\,2z^{-1}}{2.86-2.86z^{-1}}&0\\0&0&0&\dfrac{2.22-1.85z^{-1}}{5-5z^{-1}}\end{bmatrix}.$$

考虑两种工况：

工况一，

$$\boldsymbol{T}_1=\begin{bmatrix}\dfrac{0.05z^{-3}}{1-0.95z^{-1}}&0&\dfrac{0.7z^{-3}}{1-0.3z^{-1}}&0\\\dfrac{0.029\,66z^{-3}}{1-1.627z^{-1}+0.706z^{-2}}&\dfrac{0.062\,7z^{-6}}{1-0.937z^{-1}}&0&0\\0&\dfrac{0.235z^{-5}}{1-0.765z^{-1}}&\dfrac{0.5z^{-2}}{1-z^{-1}+0.25z^{-2}}&0\\\dfrac{0.5z^{-5}-0.4875z^{-6}}{1-1.395z^{-1}+0.455z^{-2}}&0&0&\dfrac{0.2z^{-6}}{1-0.8z^{-1}}\end{bmatrix};$$

工况二，

$$\boldsymbol{T}_2=\begin{bmatrix}\dfrac{0.05z^{-3}}{1-0.95z^{-1}}&0&\dfrac{0.7z^{-3}}{1-0.3z^{-1}}&0\\\dfrac{0.029\,66z^{-3}}{1-1.627z^{-1}+0.706z^{-2}}&\dfrac{0.062\,7z^{-6}}{1-0.937z^{-1}}&0&0\\0&\dfrac{0.235z^{-5}}{1-0.765z^{-1}}&\dfrac{2.83z^{-2}}{1-z^{-1}+0.25z^{-2}}&0\\\dfrac{0.5z^{-5}-0.487\,5z^{-6}}{1-1.395z^{-1}+0.455z^{-2}}&0&0&\dfrac{0.2z^{-6}}{1-0.8z^{-1}}\end{bmatrix}.$$

取两种工况下输出的 1 000 组数据作为历史数据库中的两类基准数据，假定实际采集的数据为工况一条件下白噪声协方差阵变为 $\boldsymbol{\Sigma}_a = \begin{bmatrix} 10 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 10 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ 下的数据. 三组数据中回路一与回路二的二维联合分布情况如图 4 所示. 用文献[5]的方法将实际采集的数据分别与两种工况下的基准数据进行评估，得总体特征值阵分别为

$$\boldsymbol{\Lambda}_1 = \begin{bmatrix} 138.647\,1 & 0 & 0 & 0 \\ 0 & 15.355\,4 & 0 & 0 \\ 0 & 0 & 1.964\,7 & 0 \\ 0 & 0 & 0 & 0.122\,6 \end{bmatrix},$$

$$\boldsymbol{\Lambda}_2 = \begin{bmatrix} 33.38 & 0 & 0 & 0 \\ 0 & 0.665\,8 & 0 & 0 \\ 0 & 0 & 0.142\,9 & 0 \\ 0 & 0 & 0 & 0.000\,3 \end{bmatrix},$$

由评估结果分析可知：若选取工况一的数据作评估基准，则实时采集的数据整体性能上是显著变差的；若选取工况二的数据作评估基准，则在整体性能上是得到改进的. 而实际采集的数据是来自工况一的，由于白噪声的方差变大，控制性能恶化是符合事实的.

采用本文的提出的 SF 综合相似因子对所采集的数据进行归类，将实时采集的数据分别与工况一(Ⅰ)和工况二(Ⅱ)下的数据用 SF 综合相似因子进行衡量，得到 2 表.

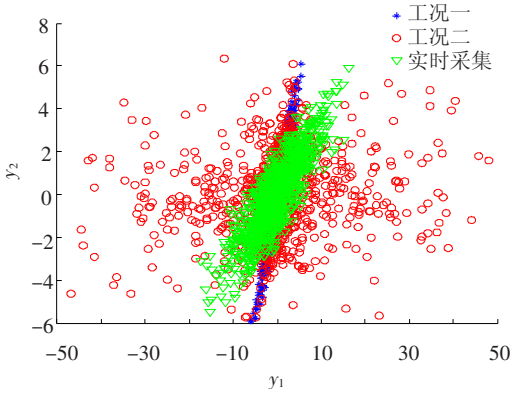


图 4 文献[11]例子回路一与回路二输出变量三组数据二维联合分布图

表 2 文献[11]例子中采集数据阵与两类工况下数据阵的相似性比较

实时采集与工况	S^{λ}_{PCA}	S_{dist}	$SF = 0.5S^{\lambda}_{PCA} + 0.5S_{dist}$
Ⅰ	0.943 8	0.863 3	0.903 5
Ⅱ	0.726 9	0.907 7	0.817 3

由表中结果可以看出，实时采集的数据与工况二下的数据 SF 因子最大，故实时数据应被归类与工况二下的数据，应选取历史数据库中工况二下的数据作为性能评估基准，再进行性能评估，这样方能得到正确的性能评估的结果.

5 结 论

本文提出一种针对多变量多工况条件下基于数据驱动的性能评估的方法，用仿真例子阐述了在进行性能评估前选取历史数据库中恰当的基准数据的重要性和必要性，采用综合 PCA 相似因子与几何距离因子的 SF 综合相似因子对所采集的数据与历史库中各工况下的数据进行正确的归类，从而才能保证性能评估结果的正确性，仿真例子证实了该方法的有效性.

参考文献：

[1] DUGARD L, GOODWIN G C, XIANYA X. The role of the interactor matrix in multivariable stochastic adaptive control[J]. Automatica, 1984, 20(5): 701 – 709.

[2] HARRIS T J, BOUDREAU F, MACGREGOR J F. Performance assessment of multivariable feedback controllers[J]. Automatica, 1996, 32(11):1505 – 1518.

[3] HUANG B, SHAH S L, KWOK E K. Good, bad or optimal? Performance assessment of multivariable processes[J]. Automatica, 1997, 33(6): 1175 – 1183.

[4] QIN S J, YU J. Recent developments in multivariable controller performance monitoring [J]. Journal of Process Control, 2007, 17(3): 221 – 227.

[5] YU H, QIN S J. Statistical MIMO controller performance monitoring. Part I: Data-driven covariance benchmark[J]. Journal of Process Control, 2008, 18(3 – 4): 277 – 296.

[6] YU H, QIN S J. Statistical MIMO controller performance monitoring. Part II: Performance diagnosis[J]. Journal of Process Control, 2008, 18(3 – 4): 297 – 319.

[7] TIAN X M, CHEN G Q, CHEN S. A data-based approach for multivariate model predictive control performance monitoring[J]. Neurocomputing, 2011. 74(4): 588 – 597.

[8] PUSHA S, GUDI R, NORONHA S. Polar classification with correspondence analysis for fault isolation [J]. Journal of Process Control, 2009, 19(4): 656 – 663.

[9] SINGHAL A, SEBORG D E. Pattern matching in historical batch data using PCA[J]. IEEE Control Systems Magazine, 2002, 22(5): 53 – 63.

[10] SINGHAL A, SEBORG D E. Pattern matching in multivariate time series databases using a moving-window approach[J]. Industrial & Engineering Chemistry Research, 2002, 41(16): 3822 – 3838.

[11] McNabb C A, Qin S J. Fault diagnosis in the feedback-invariant subspace of closed-loop systems[J]. Industrial & Engineering Chemistry Research, 2005, 44(8): 2359 – 2368.

（编辑 苗秀芝）