

Tennessee-Eastman过程的学习型案例推理故障诊断方法

严爱军^{1,2,3,4†}, 王英杰^{1,3}, 王殿辉^{1,5}

(1. 北京工业大学 信息学部自动化学院, 北京 100124; 2. 计算智能与智能系统北京市重点实验室, 北京 100124;
3. 数字社区教育部工程研究中心, 北京 100124; 4. 城市轨道交通北京实验室, 北京 100124;
5. 拉筹伯大学 计算机科学与计算机工程系, 墨尔本VIC 3086)

摘要: 为了提高 Tennessee-Eastman (TE) 过程的故障诊断准确率, 本文研究一种学习型伪度量 (learning pseudo metric, LPM) 代替距离度量的案例检索方法, 并建立了 TE 过程的案例推理 (case-based reasoning, CBR) 故障诊断模型. 首先建立 LPM 度量准则并对 LPM 模型进行训练, 其次度量目标案例与每一个源案例的相似度, 从中检索与目标案例相似的同类案例, 再采用多数重用原则从同类案例中决策出目标案例的解, 最后通过 TE 过程的运行数据对该方法的性能进行测试, 并与典型的 CBR 和 BP (back-propagation) 神经网络和支持向量机等方法进行对比, 表明本文方法能有效提高故障诊断准确率, 在实际化工过程中具有一定的推广应用价值.

关键词: TE 过程; 故障诊断; 案例推理; 学习型伪度量; 案例检索

中图分类号: TP273 文献标识码: A

Fault diagnosis method using learning case-based reasoning for Tennessee-Eastman process

YAN Ai-jun^{1,2,3,4†}, WANG Ying-jie^{1,3}, WANG Dian-hui^{1,5}

(1. School of Automation, Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China;
2. Beijing Key Laboratory of Computational Intelligence & Intelligent System, Beijing 100124, China;
3. Engineering Research Center of Digital Community, Ministry of Education, Beijing 100124, China;
4. Beijing Laboratory for Urban Mass Transit, Beijing 100124, China;
5. Department of Computer Science and Computer Engineering, La Trobe University, Melbourne, VIC 3086, Australia)

Abstract: To diagnose the fault in the Tennessee-Eastman (TE) process more accurately, a learning pseudo metric (LPM)-based case retrieval method is proposed to replace distance measure retrieval method and a case-based reasoning (CBR) fault diagnosis model of TE process is established. Firstly, the LPM metrics are established to train the LPM model. Then, the similarity between the target case and each source case is measured to find the same type of cases as the target case. Next, the solution of the target case is obtained based on the majority of reuse principle. Finally, the running data of TE process are used to carry out a performance test and a comparison experiment. The results show that the proposed LPM-based CBR method is superior to traditional CBR, back-propagation (BP) neural network and support vector machine method and significantly improves the accuracy of the fault diagnosis. It has a promotional value for fault diagnosis in the actual chemical process.

Key words: TE process; fault diagnosis; case-based reasoning; learning pseudo metric; case retrieval

1 引言(Introduction)

化工生产过程是经过化学反应将原料转变成产品的工艺过程, 作为过程工业的一部分, 在国民经济中具有重要地位^[1]. 在生产过程中, 一旦发生故障而又不能及时排除时, 就会影响生产的进行, 并威胁到人员和设备的安全, 所以, 需要准确及时地识别故障类

型并进行相应的处理. 然而, 由于化工过程机理复杂, 很难建立准确的机理诊断模型, 因此, 化工过程故障诊断方法的研究具有重要的现实意义^[2].

TE(Tennessee-Eastman)过程是美国Eastman化学公司的Downs & Vogel于1993年建立的一个实际化工过程的模型^[3], 为研究该类过程的故障诊断技术提供

收稿日期: 2016-09-24; 录用日期: 2017-05-25.

†通信作者. E-mail: yanaijun@bjut.edu.cn; Tel: +86 10-67396376.

本文责任编辑: 陈增强.

国家自然科学基金项目(61374143), 北京市自然科学基金项目(4152010)资助.

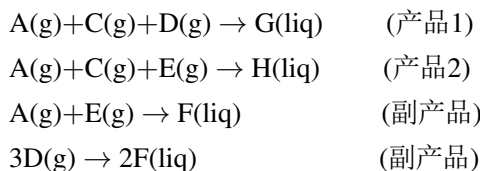
Supported by National Natural Science Foundation of China (61374143) and Beijing Natural Science Foundation of China (4152010).

了一个实验平台. 目前, 化工过程的故障诊断方法主要有模型驱动法^[4]和数据驱动法. 由于化工过程往往具有慢时变、分布参数、非线性和强耦合特性, 难以得到精确的数学模型, 使得模型驱动方法的性能不尽理想. 数据驱动法中比较有代表性的如多变量统计^[5-6]、神经网络^[7]、支持向量机^[8]等. 其中, 多变量统计方法在应用时有许多的假设条件, 实际上限制了它在实际过程中的应用. 虽然神经网络、支持向量机等方法不需设置过多的前提条件, 但本身也存在一些问题, 比如神经网络的结构设计问题、支持向量机在具有大规模训练样本时效率很低等, 因此, 这些诊断方法在通用性上很难分出孰优孰劣, 需要进一步研究新的故障诊断方法. 近年来, 案例推理(case-based reasoning, CBR)作为人工智能领域一种较新的问题求解与机器学习方法, 在工业过程故障诊断领域的应用得到了广泛关注^[9-12]. 然而, CBR在推理求解的过程中, 还有一些问题尚未完全解决, 比如, 案例检索过程中, 经常使用的是基于距离的相似性度量方法, 存在两个问题, 一是如何确定特征属性的权重^[13], 即使权重的分配是合理正确的, 但又不得不面临距离陷阱问题, 即距离最近的两个案例并不一定最相似^[14]. 因此, 应用CBR进行故障诊断时, 需要重点研究案例检索过程的相似性度量方法.

基于上述, 本文利用学习型伪度量 (learning pseudo metric, LPM)^[15]代替距离度量检索方法并建立一种基于LPM案例检索的CBR故障诊断模型. 首先通过TE仿真模型构造故障诊断案例库, 然后, 根据LPM度量准则对LPM模型进行训练, 基于此模型实现案例检索功能, 再以K近邻法(K-nearest neighbor, KNN)^[16]实现源案例的结论重用, 之后存储目标案例并得出故障诊断结论, 最后通过实验验证了方法的有效性.

2 TE过程故障描述(Faults of TE process)

TE过程有5个主要单元: 反应器、冷凝器、压缩机、分离器、汽提塔; 8种成分: A, B, C, D, E, F, G, H. 主要反应过程如下:



其中: A, C, D与E为生产所需的主要原料, B为惰性组分, 经过工艺过程所产生的G和H为目标液体产物, F为副产品.

根据产品G和H的混合比不同, 将该过程划分为6种操作模式, 如表1所示. 本文仅参考模式1, 即基本工况模式.

TE过程模型中预设了21种过程故障, 文献[17]列出了这些故障的原因及类型, 即IDV(1)–IDV(21). 其

中, 故障号IDV(1)–IDV(15)和IDV(21)是16种已知故障, IDV(16)–IDV(20)是5种未知故障. 更进一步, IDV(1)–IDV(7)等7种故障的类型是变量的阶跃变化; IDV(8)–IDV(12)等5种故障的类型是变量的随机变化; IDV(13)是缓慢漂移; IDV(14)、IDV(15)与IDV(21)等3个故障表示阀门卡死.

表1 TE过程操作模式

Table 1 Operation modes of TE process

模式号	G/H质量比	模式名称
1	50/50	7038 kg·h ⁻¹
2	10/90	G为1408 kg·h ⁻¹ , H为12699 kg·h ⁻¹
3	90/10	G为10000 kg·h ⁻¹ , H为1111 kg·h ⁻¹
4	50/50	最大生产率
5	10/90	最大生产率
6	90/10	最大生产率

3 基于LPM的CBR故障诊断模型(LPM-based CBR fault diagnosis model)

针对TE过程21种故障的诊断问题, 本节建立一种基于LPM案例检索的CBR诊断模型, 并讨论了算法实现, 最后给出算法步骤.

3.1 传统CBR故障诊断模型(Traditional CBR fault diagnosis model)

设CBR故障诊断模型为 $\Phi: F \rightarrow C$, F 是特征空间, C 是故障类别集合, 根据Aamodt & Plaza提出的CBR模型^[18], 可得如图1所示的诊断模型 Φ . 案例库中存储的是 p 个源案例记录 $C_k (k = 1, 2, \dots, p)$, 包括问题描述 X_k 及相应的故障类别 Y_k . 令目标案例的问题描述为 X_{p+1} , 为了得到它的故障类别 Y_{p+1} , 依次进行如下步骤:

1) 案例检索. 采用欧氏距离计算 X_{p+1} 与 X_k 的相似度 sim_k :

$$\text{sim}_k = 1 - \sqrt{\sum_{i=1}^n \omega_i (x_{i,p+1} - x_{i,k})^2}, \quad k = 1, 2, \dots, p, \quad (1)$$

其中: $x_{i,k}$ 和 $x_{i,p+1}$ 分别是 X_k 和 X_{p+1} 中第 i 个过程变量的归一化值; $\omega_i (i = 1, 2, \dots, n)$ 是第 i 个过程变量的权重, 表示各变量的重要程度. 通过式(1)的计算, 共得到 p 个相似度, 分别是 $\text{sim}_1, \dots, \text{sim}_p$, 并将其按照降序排列, 在此基础上, 采用KNN策略^[16]得到前 K 个源案例对应的故障类别;

2) 案例重用. 根据多数重用原则($K > 1$ 且为奇数), 将检索出的 K 个源案例对应的故障类别进行统计, 占多数的那一种类别作为 X_{p+1} 的建议故障类别 \hat{Y}_{p+1} ;

3) 案例修正. 对此建议故障类别 \hat{Y}_{p+1} 进行评价, 若评价为失败则需对诊断结果进行修正, 以获得确认

的最终故障类别 Y_{p+1} ;

4) 案例存储. 形成一个新的案例记录 $C_{p+1} : \langle X_{p+1}; Y_{p+1} \rangle$, 将其存储于案例库中, 供后续的诊断求解, 至此, 源案例总数由 p 变为 $p + 1$, 从而实现CBR故障诊断模型的增量式学习过程.

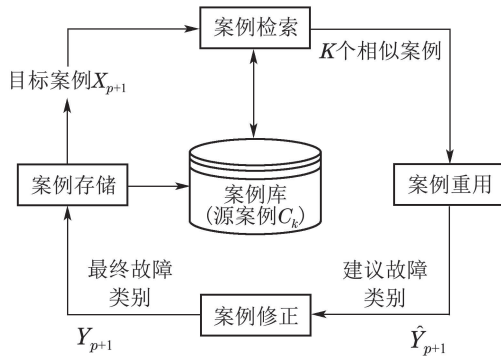


图1 传统CBR故障诊断模型

Fig. 1 Traditional CBR fault diagnosis model

图1所示传统CBR故障诊断模型在案例重用阶段可以得到目标案例 X_{p+1} 的建议故障类别 \hat{Y}_{p+1} , 这个结果是否正确, 受式(1)所示的基于欧氏距离的相似性度量方法的影响, 具体来讲, 案例检索环节中, 基于欧氏距离的相似性度量策略遇到的困难主要有权重的分配及距离的陷阱2个主要问题. 迄今为止, 仍然没有一种变量的权重分配方法可以准确衡量出各变量在问题求解过程中的重要程度, 而距离陷阱问题的存在也会限制检索后案例重用的效果, 从而制约诊断模型的诊断性能. 因此, 这2个困难使得CBR在问题求解过程中的适应性、可靠性、准确性等性能仍未彻底解决, 为了避免权重分配及距离陷阱问题, 下文研究一种新的相似性度量方法.

3.2 基于LPM的CBR模型(LPM-based CBR model)

在图1的传统CBR诊断模型基础上, 重新设计了一种新的诊断模型, 如图2所示.

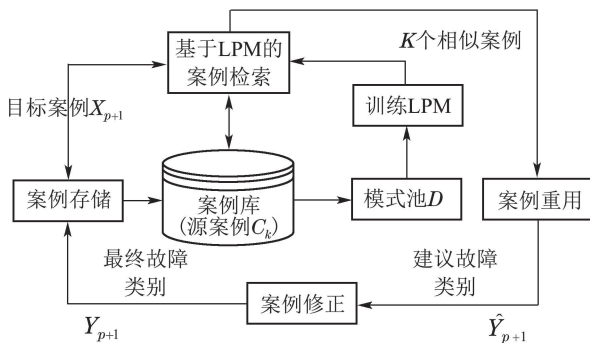


图2 基于LPM的CBR故障诊断模型

Fig. 2 CBR fault diagnosis model based on LPM

图2中的主要功能是: 首先存储源案例 $C_k (k = 1, 2, \dots, p)$ 至案例库, 并用这些源案例构造出训练样

本, 形成一个模式池 D , 然后根据LPM度量准则对LPM模型进行离线训练, 接着对出现的目标案例 X_{p+1} 进行基于LPM的案例检索, 度量出目标案例与每一个源案例的相似性, 得到与目标案例相似的同类案例(K 个), 最后采用多数重用原则从 K 个同类案例中得到目标案例的建议故障类别 \hat{Y}_{p+1} , 后续的案例修正和案例存储两个环节与图1的描述一样.

3.3 算法实现(Implementation of the algorithm)

下面介绍图2所示各部分的算法实现.

1) 构建案例库. 对于具有数值特征的源案例来说, 一种比较简单的案例表示方法是属性特征值描述法^[19], 它将问题描述与相应的解表示为一个特征向量的形式. 每个源案例均具有41个过程变量, 则第 k 个源案例 C_k 可表示为如下的二元组形式:

$$C_k : \langle X_k; Y_k \rangle, k = 1, 2, \dots, p, \quad (2)$$

其中: p 是源案例总数; Y_k 是故障类别; X_k 是 C_k 的问题描述, 可表示为

$$X_k = (x_{1,k}, \dots, x_{i,k}, \dots, x_{41,k}), \quad (3)$$

其中 $x_{i,k}$ 是 X_k 中第 i 个过程变量的归一化值.

2) 构建模式池. 设 $C = \{C_1, C_2, \dots, C_p\}$ 是一个数据集, C_k 是 C 的一个子集, 表示一类实值数据. M 是将 C_k 映射至它的特征空间 F_k 的一个 n 维提取器^[15], 即

$$F_k = \{x = M(e) \in \mathbb{R}^n : e \in C_k\}, \quad k = 1, \dots, p, \quad (4)$$

其中 x 是点集 χ 中的元素.

定义模式池 D 如下^[15]:

$$D = \{(x, y) \mapsto \delta_{ij} : (x, y) \in F_i \times F_j\}, \quad i, j = 1, \dots, p \quad (5)$$

其中: x 和 y 均是点集 χ 中的元素; δ_{ij} 表示狄利克雷符号函数, 即 $i = j$ 时 $\delta_{ij} = 0$, 否则 $\delta_{ij} = 1$; “ \times ”表示笛卡尔积.

根据式(5)的定义, 可以从案例库中存储的源案例构建出模式池 D .

3) 训练LPM模型.

命题 1 函数 f 是点集 χ 上的一个伪度量^[15]:

$$f(x, y) = \begin{cases} 1, & x \in [a_i] \text{ 且 } y \in [a_j], i \neq j, \\ 0, & x \in [a_i] \text{ 且 } y \in [a_j], i = j, \end{cases} \quad (6)$$

其中: $[a]$ 表示一个等价类, 由 χ 中等于 a 的元素组成.

学习型伪度量通过学习技术, 比如BP神经网络去实现式(6)中的函数 f . 对于故障诊断等分类问题来说, 采用一个标准的BP网络(具有一个隐含层)进行处理时, 网络的输出 $f(x, y)$ 正好等于0或1几乎是不可能的, 因而, 采用以下度量准则判断LPM的性能^[15].

- A1) $Y_{NN}(x, y) < \varepsilon_1$, 当 x 和 y 属于同一故障类别;
 A2) $Y_{NN}(x, y) \geq \varepsilon_2$, 当 x 和 y 属于不同故障类别;
 A3) $|Y_{NN}(x, y) - Y_{NN}(y, x)| \leq \varepsilon_3$, 对任意的 x 和 y ;
 A4) $Y_{NN}(x, z) \leq Y_{NN}(x, y) + Y_{NN}(y, z)$, 对任意的 z, x 和 y 属于不同故障类别.

其中: x, y, z 表示特征向量; $Y_{NN}(x, y)$ 是BP网络的输出, 表示 x, y 之间的相似程度; $\varepsilon_1, \varepsilon_2, \varepsilon_3$ 是常数, 一般来说 $\varepsilon_1 = \varepsilon_3$, 取值为 $0.2 \sim 0.3$, ε_2 取值为 $0.7 \sim 0.8$ [15].

根据上面得到的模式池 D , 可以将其划分为训练集和测试集, 训练LPM模型的终止条件可以在训练集和测试集上观察是否以较高比例满足上述度量准则A1)–A4), 比如80%, 此时, 就可以将 $Y_{NN}(x, y)$ 用于案例检索过程; 若达不到比例要求, 将该种故障判定为新的故障模式.

4) 基于LPM的案例检索. 将目标案例的问题描述 X_{p+1} 与源案例的问题描述 $X_k (k = 1, 2, \dots, p)$ 组成 p 个输入对, 即

$$D_k : \langle X_{p+1}; X_k \rangle, k = 1, 2, \dots, p, \quad (7)$$

再利用LPM模型输出 p 个 $Y_{NN}(X_{p+1}, X_k)$, 根据上述的度量准则A1), 可以统计出与 X_{p+1} 属于同一类别的源案例个数, 假设为 K 个.

5) 案例重用. 根据多数重用原则, 将检索出的 K 个源案例对应的故障类别进行统计, 占多数的那一类类别作为 X_{p+1} 的建议故障类别 \hat{Y}_{p+1} .

6) 案例修正. 对此建议的故障类别 \hat{Y}_{p+1} 进行评价, 若评价为失败则需对诊断结果进行修正, 以获得确认的正确故障类别 Y_{p+1} .

7) 案例存储. 形成一个新的案例记录 $C_{p+1} : \langle X_{p+1}; Y_{p+1} \rangle$, 将其存储于案例库中, 供后续的诊断求解, 至此, 源案例总数由 p 变为 $p + 1$.

3.4 算法步骤(Algorithm steps)

综上所述, TE过程故障的学习型伪度量诊断为2个阶段, 第1阶段是训练LPM度量模型, 第2阶段进行TE过程的故障诊断, 算法步骤如下:

阶段 1 训练LPM模型.

步骤 1 选取TE仿真模型产生的第一类故障作为案例库;

步骤 2 数据归一化, 设定BP的参数;

步骤 3 开始十折交叉实验;

步骤 4 由案例库分别构建训练集、测试集;

步骤 5 利用训练集训练BP网络;

步骤 6 将训练集、测试集输入训练后的网络, 检验模型是否满足度量准则A1)–A4), 若满足则进行步骤7, 不满足则返回步骤5;

步骤 7 判断十折交叉实验是否结束, 结束后进入阶段2, 否则继续进行交叉实验;

阶段 2 故障诊断.

步骤 8 对新出现的过程变量 $x_1 \sim x_{41}$ 的数值归一化处理形成目标案例 X_{p+1} ;

步骤 9 构建目标案例与案例库的模式池 D ;

步骤 10 将 D 输入LPM模型, 度量目标案例与案例库中每条案例的相似性, 得到模型的输出矩阵;

步骤 11 从输出矩阵中提取出所有与目标案例同类别的 K 个源案例;

步骤 12 统计 K 个源案例的故障类别, 选出个数最多的一种故障类别作为目标案例 X_{p+1} 的建议类别 \hat{Y}_{p+1} ;

步骤 13 修正故障类别 \hat{Y}_{p+1} , 得到正确的故障类别 Y_{p+1} ;

步骤 14 将新案例 $C_{p+1} : \langle X_{p+1}; Y_{p+1} \rangle$ 存储于案例库中, 若有新的过程变量出现, 返回该阶段的步骤1.

4 实验与应用(Experiment and application)

为了方便查看, 约定如下缩写词, BP表示BP神经网络算法; SVM表示支持向量机算法; 采用KNN检索策略的CBR算法记为KNNCBR; 采用LPM检索策略的CBR算法记为LPMCBR.

4.1 实验设计(Experimental design)

在MATLAB环境下, TE模型中设置的仿真时间为16 h, 每3 min采样一次. 整个分析过程所需要的数据包括20组带不同干扰的测试数据, 即故障的种类是IDV(1)–IDV(20). 在这20组数据中, 前8个小时处于稳定无故障的工况, 在第9个小时引入各自不同的故障.

各种方法中的参数设置如下: BP算法中, 采用3层网络结构, 隐层神经元个数为10, 激发函数使用Sigmoid型函数, 训练函数选择Trainrp, 训练次数为1200, 目标误差值为 10^{-5} , 学习速率和收敛误差分别为0.1和0.01; SVM算法中, 惩罚因子取1, 损失因子是0.1, 核函数取高斯径向基函数; KNNCBR算法中 K 取5; LPMCBR算法中, $\varepsilon_1 = \varepsilon_3 = 0.3$, $\varepsilon_2 = 0.7$, 终止LPM模型的条件是满足第3.3节中度量准则A1)–A4)的比例为80%, 其他参数设置与BP算法相同.

为了测试LPMCBR方法的有效性, 下面首先对算法的稳定性和鲁棒性进行测试, 然后将其与BP, KNNCBR及SVM进行对比实验. 对比实验包含两部分: 故障检测和多故障诊断. 其中, 故障检测的实验数据由上述的仿真操作得到, 共有20类故障, 即IDV(1)–IDV(20), 多故障诊断的实验数据根据等量抽取的原则从数据中抽取并组合, 各案例库的构成、名称缩写、样本数、故障类别数等信息参见表2.

表 2 数据集基本信息

Table 2 Basic information of data set

案例库构成	缩写	样本数	类别数
IDV(0)–IDV(2)	A	300	3
IDV(0)–IDV(3)	B	300	4
IDV(0)–IDV(4)	C	300	5

4.2 性能测试(Performance testing)

4.2.1 稳定性分析(Stability analysis)

为了评估LPM模型的稳定性,以IDV(1)为例,进行十折交叉验证.首先,从模式池D分出的训练集、测试集中分别随机选取一定数目的输入对;然后,将它们输入LPM并输出结果;最后,验证度量准则A1)–A4)的满足率,如表3–4所示.从表3–4可见,对于故障IDV(1)对应的数据集,随着输入对个数的增加,训练集满足度量准则A1)–A4)的标准差分别为0.10, 0.13, 0.31, 1.29, 测试集满足度量准则A1)–A4)的标准差分别为1.12, 0.98, 1.24, 1.91, 表示训练集和测试集的满足率基本保持不变,说明LPM模型能够保持在合理的精度范围内,具有一定的稳定性.

表 3 IDV(1)训练集度量准则的满足率

Table 3 Measurement criteria satisfaction rate of IDV(1) training set

输入对数	训练集			
	A1)	A2)	A3)	A4)
1000	98.61	98.68	89.62	88.89
4000	98.85	98.39	90.46	89.26
8000	98.85	98.34	90.04	89.46
12000	98.84	98.6	90.41	87.05
16000	98.86	98.42	90.2	88.32
20000	98.7	98.54	90.25	86.24
平均值	98.79	98.5	90.16	88.2
标准差	0.1	0.13	0.31	1.29

表 4 IDV(1)测试集度量准则的满足率

Table 4 Measurement criteria satisfaction rate of IDV(1) test set

输入对数	训练集			
	A1)	A2)	A3)	A4)
1000	98.61	98.68	89.62	88.89
4000	98.85	98.39	90.46	89.26
8000	98.85	98.34	90.04	89.46
12000	98.84	98.6	90.41	87.05
16000	98.86	98.42	90.2	88.32
20000	98.7	98.54	90.25	86.24
平均值	98.79	98.5	90.16	88.2
标准差	0.1	0.13	0.31	1.29

4.2.2 鲁棒性分析(Robustness analysis)

如同文献[15]的分析过程,本节在引入输入噪音的条件下测试LPMCBR的鲁棒性.由于对全部故障种类所对应的数据集进行训练的时间消耗过高,因此选取故障原因不同但故障类型相同(IDV(1), IDV(2), IDV(3), IDV(5)),以及故障原因相同但故障类型不同(IDV(3)与IDV(9))时的5组故障作为典型故障案例,以此测试LPMCBR方法的性能.首先,每一折实验时产生一个随机向量(即噪音noise),服从(-1, 1)的均匀分布;然后,针对LPM模型的每一个输入向量(以s表示),分别以10个不同的干扰因子(以 $\lambda_i (i = 1, \dots, 10)$ 表示,从1%变化到10%)加入噪音,此时,输入向量变为 $[I + \lambda_i \times \text{diag}(\text{noise})]s$,其中I是适当维数的单位矩阵,diag(·)表示对角矩阵.在存在输入噪音的条件下,观察诊断准确率的变化,实验结果如图3所示.可以看出,LPMCBR的准确率在存在输入干扰的情况下波动不大,比较平稳,从而说明算法具有一定的鲁棒性.

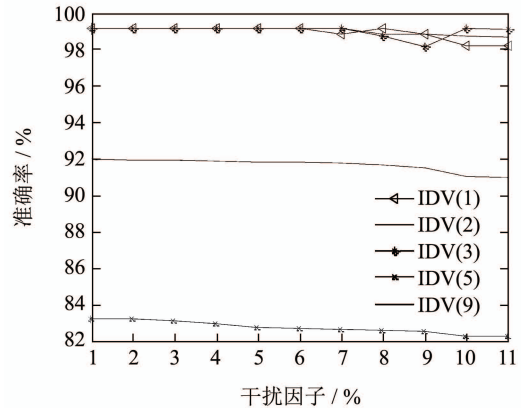


图 3 模型的鲁棒性测试

Fig. 3 Robustness testing of model

4.3 对比实验(Comparative experiment)

为了进一步验证LPM案例检索方法的有效性,与BP, KNNCBR及SVM进行了对比实验,对故障IDV(1)–IDV(20),这些方法的多故障诊断的准确率、故障检测准确率分别参见表5和表6.

表 5 不同方法的多故障诊断准确率

Table 5 Fault diagnosis accuracy of different methods

数据集名称	BP	SVM	KNNCBR	LPMCBR
A	45.67	84.33	86.67	93.33
B	25	71.13	60.79	80.05
C	20	54.49	54.17	57.84
平均值	30.22	69.98	67.21	77.07

从表5可以看出,LPMCBR方法在多故障诊断的准确率方面均高于其他方法.通过表6看出,LPM–CBR虽然对IDV(4), IDV(8)以及IDV(11)等3个故障的检测准确率不是最高,但从另外的17类故障的准确率

及统计平均值来看, LPMCIBR方法的故障检测准确率占有较大优势. 综合上述, 说明基于LPM的CBR诊断方法在TE过程的故障诊断中具有较好的综合性能.

表6 不同方法的故障检测准确率

Table 6 Fault detection accuracy of different methods

故障类别	BP	SVM	KNNCBR	LPMCIBR	%
IDV(1)	53.13	96.25	96.88	99.12	
IDV(2)	57.81	96.25	96.56	99.11	
IDV(3)	81.25	74.21	45.34	99.11	
IDV(4)	66.25	50.63	46.29	49.06	
IDV(5)	38.44	47.52	42.85	83.29	
IDV(6)	85.67	100	100	100	
IDV(7)	50	67.83	57.2	69.69	
IDV(8)	50	56.25	87.19	76.93	
IDV(9)	63.13	60.42	47.51	91.95	
IDV(10)	58.75	48.42	48.75	86.88	
IDV(11)	74.06	46.88	49.7	61.4	
IDV(12)	58.44	50.94	48.14	63.4	
IDV(13)	50.63	71.18	90.31	90.39	
IDV(14)	90	59.32	49.38	93.75	
IDV(15)	53.13	48.14	44.07	66.43	
IDV(16)	50	61.88	50	50.63	
IDV(17)	54.06	61.88	62.12	81.99	
IDV(18)	50.63	62.5	50.89	77.67	
IDV(19)	50.94	44.38	46.57	51.25	
IDV(20)	55.31	85.63	88.75	89.19	
平均值	59.58	64.53	62.43	79.06	

5 结论(Conclusions)

本文在传统CBR故障诊断模型基础上, 利用LPM形成一种新的案例检索方法, 并实现了基于LPM的CBR故障诊断模型. 通过实验表明本文方法具有一定的稳定性和鲁棒性, 对比实验证明在故障检测与诊断准确率方面优于其他故障诊断方法. 说明LPMCIBR故障诊断模型具有较好的求解质量, 可以提高CBR模型在检索过程中的学习性能. 然而, 本研究也有一定的局限性, 比如, 从机器学习的角度看, 建立一个良好的LPM模型并不简单. 研究中主要采用离线诊断的方式, 在计算时间上要求不高, 但随着数据量的增加, 训练LPM模型的时间会限制它的应用, 如何在现有基础上降低时间复杂度, 以更好地适应工业生产过程的故障诊断, 这是以后的研究重点和方向.

参考文献(References):

- [1] WU Q, ZHANG X, SHANG Z, et al. Political-economy based institutional industry complex and sustainable development: The case of the salt-chemical industry in Huai'an, China [J]. *Energy Policy*, 2015, 87(12): 39 – 47.
- [2] CHEN X, YAN X. Using improved self-organizing map for fault diagnosis in chemical industry process [J]. *Chemical Engineering Research & Design*, 2012, 90(12): 2262 – 2277.
- [3] DOWNS J J, VOGEL E F. A plant-wide industrial process control problem [J]. *Computers and Chemical Engineering*, 1993, 17(3): 245 – 255.

- [4] TIAN W D, SUN S L, GUO Q J. Fault detection and diagnosis for distillation column using two-tier model [J]. *Canadian Journal of Chemical Engineering*, 2013, 91(10): 1671 – 1685.
- [5] CHIANG L H, JIANG B, ZHU X, et al. Diagnosis of multiple and unknown faults using the causal map and multivariate statistics [J]. *Journal of Process Control*, 2015, 28(7): 27 – 39.
- [6] JIANG Q, HUANG B, YAN X. GMM and optimal principal components-based Bayesian method for multimode fault diagnosis [J]. *Computers & Chemical Engineering*, 2016, 84(1): 338 – 349.
- [7] LÁZARO J M B D, MORENO A P, SANTIAGO O L, et al. Optimizing kernel methods to reduce dimensionality in fault diagnosis of industrial systems [J]. *Computers & Industrial Engineering*, 2015, 87(9): 140 – 149.
- [8] BEI W, YAN X, JIANG Q. Loading-based principal component selection for PCA integrated with support vector data description [J]. *Industrial & Engineering Chemistry Research*, 2015, 54(5): 1615 – 1627.
- [9] SHEN B. An improved case-based reasoning for fault diagnosis of CNC machine tool [J]. *Journal of Information & Computational Science*, 2015, 12(8): 2939 – 2949.
- [10] REYES E R, NEGRY S, ROBLES G C, et al. Improvement of online adaptation knowledge acquisition and reuse in case-based reasoning: Application to process engineering design [J]. *Engineering Applications of Artificial Intelligence*, 2015, 41(3): 1 – 16.
- [11] MA G, JIANG L, XU G, et al. A model of intelligent fault diagnosis of power equipment based on CBR [J]. *Mathematical Problems in Engineering*, 2015, 2015(4): 1 – 9.
- [12] YAN A, WANG W, ZHANG C, et al. A fault prediction method that uses improved case-based reasoning to continuously predict the status of a shaft furnace [J]. *Information Sciences*, 2014, 259(2): 269 – 281
- [13] HAN M, CAO Z. An improved case-based reasoning method and its application in endpoint prediction of basic oxygen furnace [J]. *Neurocomputing*, 2015, 149(pc): 1245 – 1252.
- [14] SHEN Yuanyuan, YAN Yan, WANG Hangzi. Recent advances on supervised distance metric learning algorithms [J]. *Acta Automatica Sinica*, 2014, 40(12): 2673 – 2686.
- [15] WANG D H, MA X H, KIM Y S. Learning pseudo metric for intelligent multimedia data classification and retrieval. *Journal of Intelligent Manufacturing* [J]. 2005, 16(6): 575 – 586.
- [16] WANG Guozhu, LIU Jianchang, LI Yuan, et al. Fault diagnosis of industrial processes based on weighted k -nearest neighbor reconstruction analysis [J]. *Control Theory & Applications*, 2015, 32(7): 873 – 880.
(王国柱, 刘建昌, 李元, 等. 加权 k 最近邻重构分析的工业过程故障诊断 [J]. *控制理论与应用*, 2015, 32(7): 873 – 880.)
- [17] SHEN Y, DING S X, HAGHANI A, et al. A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process [J]. *Journal of Process Control*, 2012, 22(9): 1567 – 1581.
- [18] AAMODT A, PLAZA E. Case-based reasoning: foundational issues, methodological variations, and system approaches [J]. *AI Communications*, 1994, 7(1): 39 – 59.
- [19] BERGMANN R, KOLODNER J, PLAZA E. Representation in case-based reasoning [J]. *Knowledge Engineering Review*, 2005, 20(3): 209 – 213.

作者简介:

严爱军 (1970–), 男, 博士, 教授, 博士生导师, 目前研究方向为工业过程智能自动化技术及应用, E-mail: yanaijun@bjut.edu.cn;

王英杰 (1992–), 女, 硕士研究生, 目前研究方向为过程优化控制算法和系统开发, E-mail: wangyingjie@emails.bjut.edu.cn;

王殿辉 (1962–), 男, 博士, 教授, 博士生导师, 目前研究方向为数据挖掘及计算智能技术在生物信息学、复杂系统建模与控制中的应用, E-mail: dh.wang@latrobe.edu.au.