



中国科技期刊卓越行动计划项目入选期刊

控制与决策

CONTROL AND DECISION

基于时间动态因果图的复杂工业过程故障预测方法

王姝, 魏楠, 孟思彤, 王福利

引用本文:

王姝, 魏楠, 孟思彤, 王福利. 基于时间动态因果图的复杂工业过程故障预测方法[J]. *控制与决策*, 2024, 39(7): 2242–2250.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2022.1811>

您可能感兴趣的其他文章

Articles you may be interested in

基于广义主成分分析的重构故障子空间建模方法

Reconstructed fault subspace modelling method based on generalized principal component analysis

控制与决策. 2021, 36(4): 808–814 <https://doi.org/10.13195/j.kzyjc.2019.0818>

基于交叉熵的改进NPE间歇过程故障检测算法

Improved NPE batch process fault detection algorithm based on cross entropy

控制与决策. 2021, 36(2): 411–417 <https://doi.org/10.13195/j.kzyjc.2019.0725>

基于局部信息增量与MPLS的质量相关故障检测方法

Quality-related fault detection method based on local information increment and MPLS

控制与决策. 2021, 36(7): 1647–1654 <https://doi.org/10.13195/j.kzyjc.2019.1402>

超启发式交叉熵算法求解模糊分布式流水线绿色调度问题

Hyper-heuristic cross-entropy algorithm for green distributed permutation flow-shop scheduling problem with fuzzy processing time

控制与决策. 2021, 36(6): 1387–1396 <https://doi.org/10.13195/j.kzyjc.2019.1681>

基于核典型相关性-熵成分分析的工业过程质量监测方法

A quality monitoring method for industrial process based on kernel canonical correlation-entropy component analysis

控制与决策. 2021, 36(12): 2999–3006 <https://doi.org/10.13195/j.kzyjc.2020.0548>

基于时间动态因果图的复杂工业过程故障预测方法

王 姝^{1†}, 魏 楠¹, 孟思彤¹, 王福利^{1,2}

- 东北大学 信息科学与工程学院, 沈阳 110819;
- 东北大学 过程工业自动化教育部国家重点实验室, 沈阳 110819)

摘要: 故障预测技术有力地保证了生产过程的平稳有序和人员安全. 但在实际操作过程中, 过程数据的定性与定量信息并存, 模型较为复杂. 此外, 在生产过程中, 利用在线收集的数据进行故障预测时存在时序延迟问题. 对此, 建立一种基于时间动态因果图(TDCD)的故障预测模型. 在模型建立过程中, 提出参数的延迟时间间隔学习算法, 即移动搜索最大信息系数(MIC)算法, 充分考虑了时序方面的延迟问题. 在推理过程中, 加入趋势分析和延时信息排序以优化推理过程, 减少因延迟时间造成的故障误报率. 最后, 使用某浮选过程因果图网络进行算法验证, 并将所提出的策略应用于湿法冶金浸出过程, 与单值/多值不确定动态因果图进行对比, 以表明故障预测策略的先进性和有效性.

关键词: 湿法冶金; 故障预测; 时间动态因果图; 延迟时间学习; 异常度函数; 趋势分析

中图分类号: TP206.3 文献标志码: A

DOI: 10.13195/j.kzyjc.2022.1811

引用格式: 王姝, 魏楠, 孟思彤, 等. 基于时间动态因果图的复杂工业过程故障预测方法[J]. 控制与决策, 2024, 39(7): 2242-2250.

Fault prediction method for complex industrial process based on time dynamic causality diagram

WANG Shu^{1†}, WEI Nan¹, MENG Si-tong¹, WANG Fu-li^{1,2}

- College of Information Science and Engineering, Northeastern University, Shenyang 110819, China;
- State Key Laboratory of Process Industry Automation of Ministry of Education, Northeastern University, Shenyang 110819, China)

Abstract: Fault prediction technology effectively guarantees the smooth and orderly production process and the safety of personnel. However, in actual operation, qualitative and quantitative informations of process data coexist, and the model is complex. In addition, in the production process, there is a timing delay problem when using online collected data for fault prediction. This paper establishes, verifies and applies a fault prediction model based on time dynamic causality diagram(TDCD). In the process of model building, a parameter delay time interval learning algorithm is proposed, that is, the mobile search maximum maximal information coefficient(MIC) algorithm, which fully considers the timing delay problem. In the reasoning process, trend analysis and delay information sorting are added to optimize the reasoning process and reduce the false alarm rate caused by delay time. Finally, the algorithm is validated by using a causal graph network for a flotation process. The proposed strategy is applied to the hydrometallurgical leaching process, and compared with the single-valued/multi-valued uncertain dynamic causality diagram, which shows the advancement and effectiveness of the fault prediction strategy.

Keywords: hydrometallurgy; fault prediction; time dynamic causality diagram; delay time learning; anomaly function; trend analysis

0 引言

复杂系统中的运行状态监测和故障诊断对于减少故障排除时间和维护成本, 提高系统可用性具有重要意义. 在线收集的数据存在时序信息: 一方面可为了解复杂系统提供额外帮助; 另一方面, 时序信息的

复杂性、高维性、延时性对故障预测和推理过程造成很大干扰. 因此, 需要灵活的表示和推理形式来解释现实应用中的系统动态和内在的不确定性^[1].

为了分析系统的动态行为, 通常选择隐马尔可夫模型(HMM)来表示个体的组件状态和系统随时间的

收稿日期: 2022-10-19; 录用日期: 2023-06-04.

基金项目: 国家重点研发计划项目(2021YFF0602404).

[†]通讯作者. E-mail: wangshu@ise.neu.edu.cn.

内部依赖关系^[2-3]. 然而, HMM的初始假设只适用于一步预测推理, 后续虽有研究人员试图引入具有高阶依赖的模型^[4], 即多步时序的故障预测推理, 但这显著地增加了计算成本^[5].

基于格兰杰 (Granger) 因果关系的时间网络模型可以描述变量之间的依赖性和交互作用, 从而对时间行为进行建模与分析, 用于过程领域的故障诊断^[6]. 然而, Granger 因果关系检验只适用于线性稳态的时间序列, 不适用于非平稳或非线性时间序列. 此外, 它仅适用于确定变量之间的成对关系, 不适用于多变量之间的因果关系情况^[7].

在工业故障诊断领域中, 贝叶斯网络 (BNs) 被广泛应用于不确定知识的表示和推理^[8-9]. 为了处理与时间相关的数据, BNs 已扩展到动态 BNs (DBNs) 模型^[10], 并衍生出一系列相关模型^[11-14]. DBNs 通过将网络展开为时间片, 然后对静态 BN 应用推理算法来实现精确推理. 然而, 这种方法的计算复杂性很高, 使得 DBNs 在大规模应用中难以实现在线精确推理.

针对 BN 网络面对环状因果图时遇到的困难, 有学者先后提出了动态因果图 (DCD)^[15] 及其改进的不确定动态因果图模型 (DUCG)^[16], 该方法被推广到工程、医疗等各个领域^[17-19]. S-DUCG 模型用于解决子变量为单赋值情形的动态不确定因果图, 引进了表示子变量与父变量之间因果关系的连接作用事件; M-DUCG 模型用于解决子变量为多赋值情形的动态不确定因果图, 它引进了表示子变量与父变量之间因果关系的随机作用事件.

在故障预测过程中, 许多因果关系之间存在时间延迟. 例如, 在复杂的工业过程中, 增加流量的异常不会立即导致液体溢出, 实际上, 液位会继续缓慢上升, 直到液体溢出. 因此, 分析因果关系需要考虑延迟时间信息, 即异步因果关系. 为了建立异步因果模型, 需要在多个时间点检查变量之间的关系. 本文针对在线收集的时间序列数据的故障时序延迟现象, 融合传统的时间贝叶斯网络和动态因果图结构, 提出一种时间动态因果图 (TDCD) 模型, 将时间信息放在连接事件上, 使得连接事件包括父节点到子节点之间的因果强度和时延信息.

为更好地计算时延信息, 本文提出一种移动搜索最大 MIC 获取延迟时间的方法, 这种方法弥补了时延信息导致的故障预测错漏等问题, 并给出相关事件的时延排序. 在工业过程中, 很多故障都是缓慢发生的, 因此在故障发生之前就存在亚健康状态. 为了得到先进的预测结果, 本文将趋势分析和异常度函数集

成到在线预测过程中, 使得故障值从离散的 0, 1 数值分布转为连续的异常度 (信任因子), 将故障的异常程度进行量化. 该方法可以在亚健康状态下开始预测推理, 从而可以提前得到结果, 操作者会有足够的时间进行相应决策. 故障推理过程利用预测过程得到的时序信息进行排序, 加权到后验概率排序中, 减少因时延信息导致的故障推理误报.

将本文方法应用于湿法冶金浸出过程的故障预测和推理, 仿真结果验证了移动搜索最大 MIC 在处理时延信息方面的有效性, 以及将趋势分析和时延加权后验概率信息集成到在线预测过程中对于提前推理和正确推理的重要意义.

1 TDCD 的定义与构建

传统的动态因果图 (DCD) 可以处理不确定的信息, 但不能处理时间信息. 为了获得更多的结果信息, 应该丰富 DCD 中包含的信息. 本文通过增加时间信息对 DCD 进行改进, 使连接事件包含更丰富的信息.

1.1 TDCD 定义

在时序动态因果图 (TDCD) 中, 基本事件的定义与传统的 DCD 相同, 但节点事件、连接事件的定义则不同. TDCD 定义如下: X 代表一个节点事件, 在 TDCD 中用一个圆形节点表示, 节点上的状态可能包含时间延迟信息, 在没有具体的时间间隔值的情况下, 时延信息只能由“长时间”“短时间”“中长时间”等模糊概念来定义. 连接事件变量 P 不仅包含两个变量之间的因果强度信息, 还包含父节点与子节点之间的时间延迟信息. 延迟信息表示为时间间隔 $[t_{\text{star}}, t_{\text{end}}]$, t_{star} 代表开始时间, 即最短延迟时间; t_{end} 代表结束时间, 即最长延迟时间. 每个时间节点的时间间隔数量和持续时间可以不同 (多粒度). 例如, 如果基本事件 B_1 的 1 状态立即导致 X_2 的 1 状态发生, 则连接事件表示为 p_{12}^1, p_{12}^2 只包含事件 B_1 与事件 X_2 之间因果关系的强度, 不包含时间间隔 (B_1^1 与 X_2^1 之间没有时间延迟). 如果事件 X_3 的异常有时间延迟导致 X_1 的异常发生, 则连接事件 P_{31} 表示为

$$P_{31} = \left(\begin{matrix} p_{31}^1 & p_{31}^2 \\ p_{31}^{12} & p_{31}^{22} \end{matrix} \right) : \left(\begin{matrix} 0.63 & \emptyset \\ \emptyset & 0.28 \end{matrix} \right) : \left(\begin{matrix} [6, 12] \\ [12, 23] \end{matrix} \right).$$

事件 X_3 引起 X_1 的延迟时间分为两段, $[6, 12]$ 之间延迟时间的概率为 0.63, $[12, 23]$ 之间延迟时间的概率为 0.28. TDCD 不仅关注节点变量变化的状态, 还关注节点变量变化的时间. TDCD 在因果关系中加入了时间信息, 表示从父节点状态变化到子节点状态变化的延迟时间.

1.2 TDCD模型的学习算法

时序动态因果图模型由结构和参数两部分组成. 建立TDCD结构需要确定每个节点之间的因果关系. TDCD的参数包括基本事件概率、事件之间的连接强度和延迟时间间隔. 本节主要分析TDCD中时间间隔的学习算法.

时间间隔表示当父节点异常发生时子节点异常发生的延迟. 一般地, 延迟时间仅由专家经验给出, 并不完全准确. 针对现有方法的缺陷, 本文使用一种通过历史故障数据获取延迟时间的新方法, 即移动搜索最大MIC(最大信息系数)算法^[20]. 该方法可以弥补专家经验知识的不足.

在复杂的工业过程中, 相关事件之间的因果关系大多数之间是存在延迟的. 例如, 在工业生产过程中, 流量变量 x_2 位于变量 x_1 的下游位置. 假设采集1~6个采样点期间一共有6组数据, 分别是(1.5, 3.5), (1.6, 3.6), (1.5, 3.4), (0.8, 3.6), ..., (0.7, 1.5), (0.6, 1.4), 根据机理知识, x_2 与 x_1 之间存在因果关系. 根据以上6组数据, 计算出最大信息系数 $MIC(x_1, x_2) = 0.4591$, 进而得出 x_2 与 x_1 并不高度相关的结论, 与已知不符. 这个错误的原因是当 x_1 的流量变小时, x_2 的流量变

化会有一些延迟.

系统 $x_i (i \in (1, \dots, n))$ 中有 n 个节点变量. 每个样本点对应的所有节点变量数据都是一组数据, 每个异常对应于一组异常数据集, 每组异常数据集中包含的组数不一定相同. 在计算一组异常数据集 D 中的MIC时, 对两个节点变量 x_i 和 x_j 的历史数据进行移动搜索. 在移动搜索过程中, 一次将节点变量移动一个单位长度. 单位长度为数据采样间隔, 记为 δ . 每次移动后, 异常数据集中两个节点变量的数据都会更新一次, 即得到一个新的数据集 D' . 根据新的数据集再次计算MIC. 与之前的MIC进行比较, 保存较大的MIC并记下对应的移动次数. 完成对一组异常数据集的搜索后, 就可以得到最大MIC所对应的移动次数 l , 因此, 两个节点变量之间的延迟时间可以计算为 $t = l \times \delta$, 算法流程如图1所示. 当两个节点变量中只剩下两组数据时停止移动, 从而移动搜索最大MIC过程结束. 经过移动搜索过程, 可以得到 $m = \max MIC(x_1, x_2)$ 和对应的移动次数 l . 本文将在第3节中验证所提出的算法, 并证明该移动搜索算法可以获得较高的延迟时间精度. 对于复杂系统中的不可测变量, 本文利用专家知识来确定时间区间.

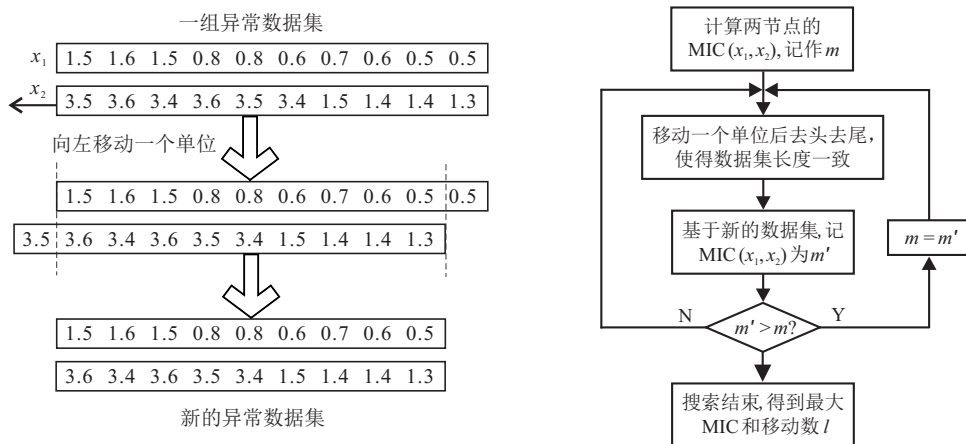


图1 移动搜索最大MIC流程

2 在线预测

在传统方法中, 对于可测量的节点变量, 需要预先设置阈值. 当实时数据超过阈值时, 变量异常. 然而, 由于客观事物的复杂性和人的认知能力的模糊性, 所定义的数值范围不够合理. 另一方面, 将阅读信息转换为纯定性信息会导致信息丢失. 针对传统方法的不足, 引入故障度函数来评价观测事件变量的异常程度. 每个测量值对应一个故障度值. 得到的结果将不再是纯粹的定性信息, 而是对异常程度的定量描述. 在推理计算中使用异常度值 d_E , 可以使预测结果

更加丰富和准确.

2.1 异常度函数

异常度函数 $D(B)$ 的计算方法如下:

$$D(B) = \begin{cases} 0, & |x - \mu_B| \leq D_B \times \mu_B; \\ \frac{|x - \mu_B| - D_B}{1 - D_B}, & D_B \times \mu_B < |x - \mu_B| < \mu_B; \\ 1, & |x - \mu_B| \geq \mu_B. \end{cases} \quad (1)$$

选择正常状态下变量历史值的平均值, 记为 μ 并

作为异常的判断标准. 根据偏离正常状态平均值 μ 的百分比, 设置异常阈值 D_B . 如果百分比在阈值内, 则为正常; 如果百分比大于 D_B 且小于 1, 则表示不健康; 如果百分比大于 1, 则为异常. 当检测到变量处于不健康状态或异常状态时, 作为 TDCD 推理的证据 E . 异常度值 d_E 分别作为其对应节点事件的排序概率的信任因子.

通过引入异常度函数, 将变量的实时状态分为正常、不健康、异常这 3 种.

2.2 趋势分析

当使用 TDCD 进行故障预测时, 异常的产生通常不是瞬时的, 而是缓慢的. 传统方法一般很难设定一个合适的阈值来判断变量是否异常. 如果阈值设置得太高, 则开始预测推理的时间就会太晚, 从而导致结果获得时间与实际发生时间之间的时间间隔太短, 异常的影响很难在短时间内补救; 如果阈值设置得太低, 则容易因为测量误差而导致误报. 针对这一问题, 本文提出对监测变量进行趋势分析, 可以提前开始预测推理的时间, 避免因测量误差引起的误报. 对于被监测变量, 通过最小二乘拟合得到未来变量达到异常范围的时间. 这个得到的时间作为推理过程中的补充证据, 可以提前进入预测推理过程, 从而提前得到预测结果.

对于被监控的事件变量, 当实时数据偏离均值 μ 的百分比达到阈值 D_B 时, 即变量不健康时, 开始趋势分析过程. 以实时数据开始偏离均值 μ 的时间点作为趋势分析数据集的起点, 当前时间点作为趋势分析数据集的终点, 趋势分析的数据集会随着时间的推移不断扩大. 基于不断更新的数据集, 采用最小二乘法进行拟合, 得到被监测变量的趋势分析结果, 即被监测变量未来会出现异常的时间.

被监控变量异常状态的边界是 $|x - \mu_B| \geq \mu_B$. 为了使预测结果更加准确, 设置异常状态边界的阈值 D_B . 将异常状态 $[(2-D_B) \times \mu_B, (2+D_B) \times \mu_B]$ 的边界代入最小二乘拟合法得到的拟合函数中, 结果 $[t_{bs}, t_{bl}]$ 预测监控变量在未来时间范围会出现异常. 其中: t_{bs} 表示监控变量出现异常的最短时间, t_{bl} 表示最长时间

2.3 TDCD 推理

TDCD 的推理计算分为两部分: 后验概率计算和时间间隔计算. 后验概率计算需要结合异常度值 d_E , 节点在证据 E 影响下的后验概率 P_D 如下式所示:

$$P_D(X|E) = d_E \times P(X|E). \quad (2)$$

对于时间间隔计算, 需要定义计算规则, 具体定

义如下.

定义 1 对于连接事件的乘积运算, 将对应的时间间隔相加. 时间区间的求和运算公式为

$$[t_{1star}, t_{1end}] + \dots + [t_{nstar}, t_{nend}] = [t_{1star} + \dots + t_{nstar}, t_{1end} + \dots + t_{nend}]. \quad (3)$$

例如, 计算 $P_{12}P_{23}$, 有

$$P_{12} = \begin{pmatrix} p_{12}^{11} \\ p_{12}^{12} \end{pmatrix} : \begin{pmatrix} 0.25 \\ 0.35 \end{pmatrix} : \begin{pmatrix} [8, 16] \\ [16, 28] \end{pmatrix},$$

$$P_{23} = \begin{pmatrix} p_{23}^{11} & p_{23}^{21} \\ p_{23}^{12} & p_{23}^{22} \end{pmatrix} : \begin{pmatrix} 0.51 & \emptyset \\ \emptyset & 0.69 \end{pmatrix} : \begin{pmatrix} [9, 17] \\ [17, 33] \end{pmatrix},$$

$$T(p_{12}^{11}p_{23}^{11}) = ([8, 16] + [9, 17]) = [17, 33].$$

定义 2 对于连接事件的或 \cup 操作, 通过联合操作计算对应的时间间隔. 时间区间的联合运算公式为

$$[t_{1star}, t_{1end}] \cup \dots \cup [t_{nstar}, t_{nend}] = [\min(t_{1star} \dots t_{nstar}), \max(t_{1end} \dots t_{nend})]. \quad (4)$$

例如, 计算 $P_{12}P_{23} \cup P_{11}P_{13}$, 有

$$P_{11} = \begin{pmatrix} p_{11}^{11} \\ p_{11}^{12} \end{pmatrix} : \begin{pmatrix} 0.21 \\ 0.41 \end{pmatrix} : \begin{pmatrix} [13, 20] \\ [20, 38] \end{pmatrix},$$

$$P_{13} = \begin{pmatrix} p_{13}^{11} & p_{13}^{21} \\ p_{13}^{12} & p_{13}^{22} \end{pmatrix} : \begin{pmatrix} 0.41 & \emptyset \\ \emptyset & 0.65 \end{pmatrix} : \begin{pmatrix} [8, 15] \\ [15, 25] \end{pmatrix},$$

$$T(p_{12}^{11}p_{23}^{11} \cup p_{11}^{11}p_{13}^{11}) = ([17, 33] \cup [21, 35]) = [17, 35].$$

2.4 在线预测过程

基于 TDCD 的故障预测流程如图 2 所示, 具体故障预测步骤如下.

step 1: 通过结构学习算法和参数学习算法得到 TDCD 模型. 根据传统方法可以得到 TDCD 的结构和概率参数. 特别地, 事件之间的时间间隔可以通过 1.2 节中描述的方法获得. 因此 TDCD 模型也可以确定.

step 2: 实时监控过程变量. 通过比较变量偏离正常均值的百分比与阈值, 判断变量处于异常状态. 如果变量处于不健康状态或异常状态, 则以此为证据, 进行趋势分析, 得到基本时间区间 $[t_{bs}, t_{bl}]$ 和异常度值 d_E .

step 3: 将证据 E 的基本时间间隔 $[t_{bs}, t_{bl}]$ 和异常度值 d_E 代入 TDCD 推理计算过程. 计算每个变量在证据影响下可能出现异常的后验概率和时间间隔, 并以时间间隔的中点进行延迟时间排序.

step 4: 比较每个变量的后验概率排序和延迟时间排序的加权结果. 设置关于加权后验概率的适当

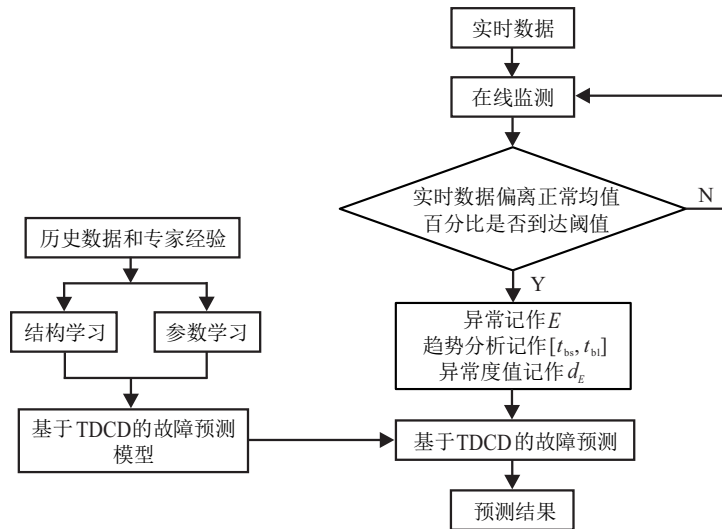


图2 故障预测流程

阈值,当节点的加权后验概率超过该阈值时,将该节点添加到预测结果中.

3 仿真实验与应用

对上述算法分别进行仿真实验.首先选择某矿物浮选过程来验证方法的有效性;然后以湿法冶金浸出工艺为背景,将本文提出的故障预测方法应用到实际工业中.

3.1 验证方法的有效性

本文使用某简化后的矿物浮选过程因果图来验证延迟时间学习算法,结构如表1所示.根据该因果图网络结构,生成1000组变量数据.根据实验可得

$$p_{21}^{11} = 0.7 : [24, 40], p_{12}^{11} = 0.5 : [18, 26], p_{34}^{11} = 0.6, p_{13}^{11} = 0.6 : [20, 31], p_{24}^{11} = 0.6 : [9, 12].$$

通过使用移动搜索最大MIC算法,得到延迟时间 t_i .

表1 浮选过程事件定义

符号	事件定义	数值
X_0	充气量	0: 正常; 1: 异常
X_1	给矿流量	0: 正常; 1: 异常
X_2	浮选柱液位	0: 正常; 1: 异常
X_3	pH值	0: 正常; 1: 异常
X_4	泡沫层厚度	0: 正常; 1: 异常
B_0	循环压力泵	0: 正常; 1: 异常
B_1	微泡发生器	0: 正常; 1: 异常
B_2	调浆水流量	0: 正常; 1: 偏大; 2: 偏小

延迟时间的精度定义为

$$t_{err} = \frac{t_i - t_{origi}}{t_{origi}}, \quad (5)$$

其中 t_{origi} 是对应的实际延迟时间. t_{err} 的平均误差

$avg(|t_{err}|) = 0.0327$,可以验证本文提出的算法得到的延迟时间是准确的.

基于TDCD模型,进行在线预测推理.在证据 $E = B_2^1$ 的假设下,预测 X_4 的未来状态.根据动态时间因果图的不变化割集的相关理论^[15],有

$$X_4^1 = p_{24}^{11}p_{12}^{11}p_{21}^{11}B_2^1 \cup p_{34}^{11}p_{13}^{11}p_{21}^{11}B_2^1.$$

根据推理规则,可以计算在证据 E 下的后验概率以及时间间隔如下:

$$\begin{aligned} P(X_4^1|B_2^1) &= p_{24}^{11}p_{12}^{11}p_{21}^{11} + p_{34}^{11}p_{13}^{11}p_{21}^{11} = \\ &0.6 \times 0.5 \times 0.7 + 0.6 \times 0.6 \times 0.7 = \\ &0.462, \\ T(X_4^1|B_2^1) &= T(p_{24}^{11}p_{12}^{11}p_{21}^{11}) \cup T(p_{34}^{11}p_{13}^{11}p_{21}^{11}) = \\ &([9, 12] + [18, 26] + [24, 40]) \cup \\ &([20, 31] + [24, 40]) = \\ &[51, 78] \cup [44, 71] = \\ &[44, 78]. \end{aligned}$$

即在证据 $E = B_2^1$ 的前提下, X_4 在未来的时间区间 $[44, 78]$ 内会出现异常的概率为0.462.为了评价预测结果中时间间隔的准确性,采用相对时间误差为标准衡量.为了衡量时间间隔预测准确率,取时间间隔的平均值作为预期时间,以相对时间误差为标准.预期时间 t_e 定义为两个节点之间时间间隔的平均值, $range(T)$ 表示时间间隔的范围,为两个节点之间时间间隔的最大值与最小值之差.相对时间误差计算如下:

$$RTE = \frac{|t_e - t_{orig}|}{range(T)}, \quad (6)$$

其中 t_{orig} 表示所预测事件发生时间的真实值. RTE 表示预测的时间间隔结果与原始实际值的偏差程度, 通过时间间隔的范围进行归一化. 在 1 000 次的仿真数据中, X_4 出现异常的次数为 530, 对应的 530 个 t_{orig} 均落在预测时间间隔 $[44, 78]$ 内, 说明预测结果的时间间隔范围准确. 再取预测时间间隔的平均值 $t_e = 61$ 为代表, 计算所有异常的 RTE 均值为 $avg(RTE) = 0.1527$. 由此可见, 在时间间隔范围标准化下的 RTE 比较小, 从而表明了预测结果的准确性.

3.2 湿法冶金浸出工艺的应用

本节以湿法冶金高铜浸出工艺过程为例, 实现基于 TDCD 的故障预测. 首先, 通过机理知识和专家经验建立 TDCD 模型^[21], 节点定义如表 2 所示, TDCD 如图 3 所示. 本次模拟使用的数据是从某高铜矿氰化浸出模拟平台采集的 1 000 组样本数据, 通过学习时间参数得到连接时间概率和时间间隔.

表 2 浸出过程节点定义

符号	事件定义	数值
		0: 正常
X_1	槽内氰离子浓度异常	1: 短时间后异常低 2: 长时间后异常低
		0: 正常
X_2	浸出率异常	1: 短时间后异常低 2: 长时间后异常低
X_3	浸出槽沉淀	0: 正常; 1: 过厚
X_4	浸出槽冒槽	0: 正常; 1: 冒槽
X_5	槽间排料管	0: 正常; 1: 堵塞
X_6	浸取液储槽冒槽	0: 正常; 1: 冒槽
B_1	来料矿浆浓度	0: 正常; 1: 偏高
B_2	来料矿浆流量	0: 正常; 1: 偏高
B_3	调浆水流量	0: 正常; 1: 偏低; 2: 偏高
B_4	氰化钠流量	0: 正常; 1: 偏低; 2: 偏高
B_5	空气输入流量	0: 正常; 1: 偏低; 2: 偏高

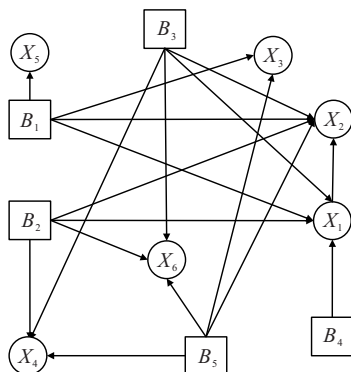


图 3 浸出过程的 TDCD

$$P_{22} = \begin{pmatrix} p_{22}^{11} \\ p_{22}^{12} \end{pmatrix} : \begin{pmatrix} 0.16 \\ 0.41 \end{pmatrix} : \begin{pmatrix} [9, 20] \\ [20, 40] \end{pmatrix},$$

$$P_{21} = \begin{pmatrix} p_{21}^{11} \\ p_{21}^{12} \end{pmatrix} : \begin{pmatrix} 0.17 \\ 0.46 \end{pmatrix} : \begin{pmatrix} [6, 12] \\ [12, 24] \end{pmatrix},$$

$$p_{24}^{11} = 0.58 : [12, 26],$$

$$p_{26}^{11} = 0.49 : [20, 40],$$

$$P_{12} = \begin{pmatrix} p_{12}^{11} & p_{12}^{21} \\ p_{12}^{12} & p_{12}^{22} \end{pmatrix} : \begin{pmatrix} 0.51 & \emptyset \\ \emptyset & 0.44 \end{pmatrix} : \begin{pmatrix} [9, 17] \\ [17, 33] \end{pmatrix}.$$

基于来料浆流 B_2 的异常进行仿真. 根据历史正态数据, 来料浆流的平均值 $\mu = 2 \times 10^4 \text{ kg/h}$, 阈值 $D_B = 5\%$. 来料浆流异常度函数定义为

$$D(x) = \begin{cases} 0, & |x - 20\,000| \leq 1\,000; \\ \frac{|x - 20\,000|}{20\,000} - 0.05, & 1\,000 < |x - 20\,000| < 20\,000; \\ 1, & |x - 20\,000| \geq 20\,000. \end{cases} \quad (7)$$

基于异常度函数 $D(B_2)$, 得到异常度值 d_E . 当监测到第 54 个采样点的流量偏离平均值 μ 的百分比达到阈值 D_B 时, 开始趋势分析过程. 其结果是未来在采样点 $[100, 105]$ 处将达到来料浆流的异常边界 $[3.9 \times 10^4, 4.1 \times 10^4] \text{ kg/h}$. 因此, 该时间间隔 $[100, 105] - [54, 54] = [46, 51]$ 被视为后续故障推理的基本时间间隔(故障预测和推理将同时进行).

根据故障预测计算方法, 筛选出与 B_2 相关的 4 个中间事件 X_1, X_2, X_4, X_6 , 利用 TDCD 的因果图化简和不交化割集方法, 计算在当前时间相关事件的后验概率(乘以异常度值 $d_E = 5\%$), 排序预测结果如图 4(a) 所示. 单从后验概率排序结果来看, 故障发生的先后顺序为 X_2, X_4, X_6, X_1 , 若以达到阈值 D_B 作为故障预测开始的标志, 则事件 X_1, X_2, X_4, X_6 将分别在未来的时间间隔 $[58, 73], [66, 108], [58, 75], [66, 91]$ 内发生故障, 如图 4(b) 所示. 根据延迟时间排序和后验概率排序结果的加权, 判断未来最有可能发生故障的事件为 X_4 , 虽然 X_2 发生故障的推理后验概率最大, 但发生的时间却是较晚的. 从因果图上可以看出, X_2 的后验概率不仅受 B_2 直接影响, 还受 X_1 间接影响. 根据 TDCD 因果图的相关理论, 给出 B_2 发生时 X_2 的后验概率和延时为

$$P(X_2^1 | B_2^1) = p_{21}^{11} p_{12}^{11} + p_{22}^{11} =$$

$$0.17 \times 0.51 + 0.16 = 0.2467,$$

$$T(X_2^1 | B_2^1) = T(p_{21}^{11} p_{12}^{11} \cup p_{22}^{11}) =$$

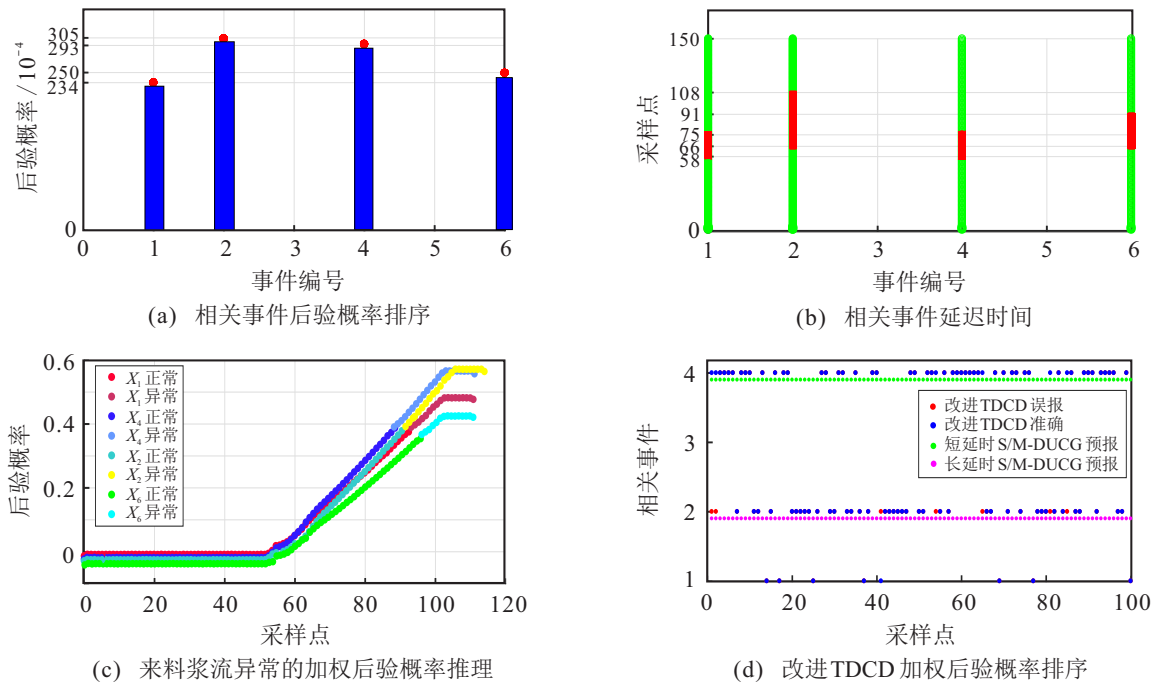


图4 来料浆流异常的改进TDCD预测及综合推理结果(图(c)中为了区分各推理预测曲线,纵坐标依次下移)

$$\begin{aligned}
 & [T(p_{21}^{11}) + T(p_{12}^{11})] \cup T(p_{22}^{11}) = \\
 & ([6, 12] + [9, 17]) \cup [9, 20] = \\
 & [15, 29] \cup [9, 20] = [9, 29]; \\
 & P(X_2^2 | B_2^1) = p_{21}^{12} p_{12}^{22} + p_{22}^{12} = \\
 & 0.46 \times 0.44 + 0.41 = 0.6124, \\
 & T(X_2^2 | B_2^1) = T(p_{21}^{12} p_{12}^{22} \cup p_{22}^{12}) = \\
 & [T(p_{21}^{12}) + T(p_{12}^{22})] \cup T(p_{22}^{12}) = \\
 & ([12, 24] + [17, 33]) \cup [20, 40] = \\
 & [29, 57] \cup [20, 40] = [20, 57].
 \end{aligned}$$

根据上述推导, X_2 的两种状态因为受 X_1 间接影响, 产生延迟时间的合并, 导致故障晚发生了 $[0, 9]$ 和 $[0, 17]$. 因此, 延迟时间的合并会导致故障较晚发生. 而普通的动态因果图(如 S/M-DUCG)^[17] 则只能根据后验概率排序结果推理出 X_2 最有可能发生故障, 忽略了延迟时间的影响, 从而造成故障误报.

具体的故障事件发生先后顺序及其确切时间点需要进行实时数据的故障推理得到, 该推理的时间间隔以 $[46, 51]$ 的基本时间间隔作为参考. 基于第 1~第 110 个样本点的实时数据集进行故障推理过程, 滑动窗口的范围为 $[46, 51]$, 每次滑动经历一组实时采样数据, 得到的后验概率的结果如图 4(c) 所示. 设置 0.4 为异常阈值, 根据趋势分析的结果, 得到 X_4 的故障将在第 88 个采样点最先发生, 其次为 X_2 、 X_1 、 X_6 . 由此,

进一步验证了改进 TDCD 基于后验概率排序和延迟时间排序结合的故障推理方法的合理性和有效性.

图 4(d) 给出了 4 个中间事件在单/多值不确定动态因果图 S/M-DUCG 和本文提出的改进 TDCD 算法在来料浆流 B_2 的异常下, 对应的 100 次实验故障顺序推理误报率, X_2 两种状态下的延时可分为短延时 $[9, 20]$ 和长延时 $[20, 40]$ 两种.

通过移动搜索最大 MIC 算法, 改进的 TDCD 可以实现故障顺序的长延时异常零误报, 能够很好地捕捉到短延时异常, 误报率为 7% 左右. 经过分析, 其中有 4 个点的误报是由于随机的延迟时间相同或接近所导致. S/M-DUCG 算法对延时故障无法做出灵敏的反应, 只有当故障预测推理的加权后验概率排序与原后验概率排序一致时, S/M-DUCG 算法才可以做出正确的判断. 对于短延时异常, S/M-DUCG 算法推理结果均为 X_4 ; 对于长延时异常, S/M-DUCG 算法推理结果均为 X_2 .

图 5 给出了在采用趋势分析与异常度函数相结合的改进方法后, 改进的 TDCD 方法与 S/M-DUCG 方法对基于来料浆流 B_2 异常的相关事件加权后验概率的预测结果对比. 由于采用了异常度函数, 改进方法可以在第 54 个采样点得到初步的预测结果, 通过最小二乘拟合的趋势分析过程, 分别在第 88、90、94、97 时刻提前获得相关变量 X_4 、 X_2 、 X_1 、 X_6 明确的推理结果; 而 S/M-DUCG 因缺少相关的趋势分析过程,

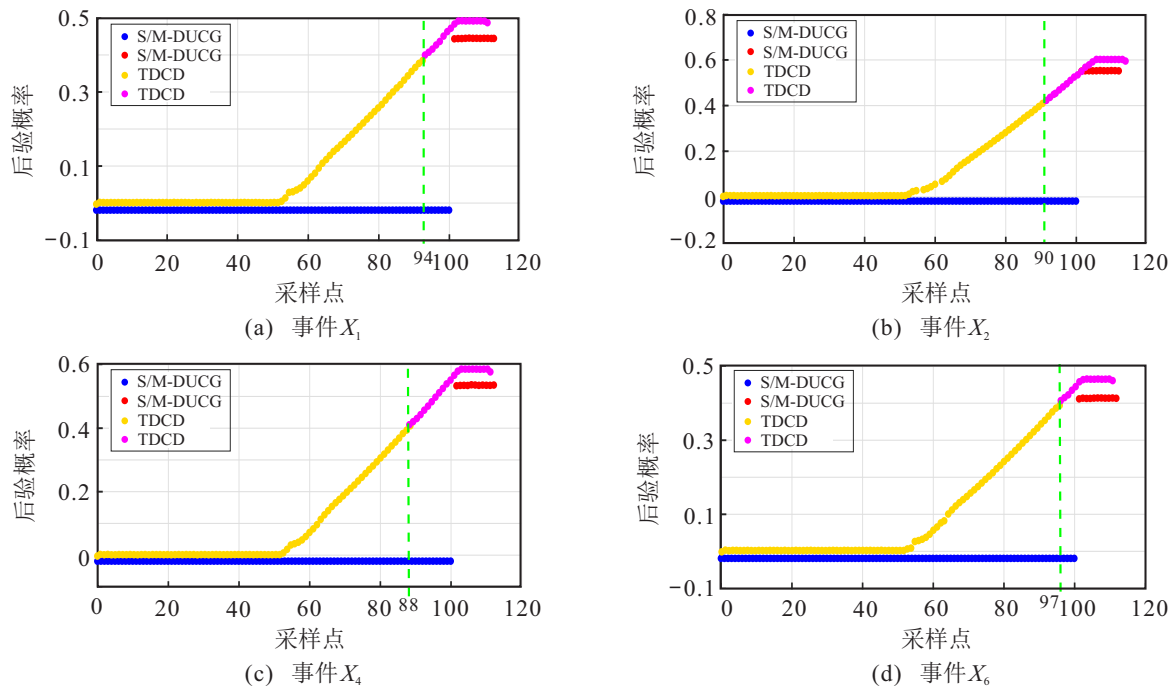


图5 相关事件后验概率预测结果对比(S/M-DUCG和改进TDCD)

只能在实际异常发生的第100个样本点后得到推理结果(推理结果不包括延时过程的加权)。当故障没有达到阈值时,改进TDCD的异常度函数能够及时反映出系统处于亚健康状态;而S/M-DUCG仍将亚健康状态视为正常状态进行推理。采用趋势分析与异常度函数相结合的改进方法可以及早获得预测结果及其故障发生的延时间隔,而提前获取的预测延时间隔将提供推理所需的延时排序信息,为利用实时数据进行时延加权后验推理做好准备。

4 结论

为了处理与时间延迟相关的因果关系,本文提出了时间动态因果图的概念。TDCD结合专家知识和定量数据分析来预测哪些事件将是故障以及何时发生。为了构建TDCD模型并实现故障预测,本文提出了延迟时间区间学习算法和基于TDCD的故障预测方案。利用某矿物浮选过程的因果图网络结构,通过相对时间误差RTE的数值,验证了移动搜索最大MIC算法的合理性。为改进故障预测方法,在推理计算过程中引入异常度函数和趋势分析。以故障程度函数作为信任因子,作用于每个相关事件的排序概率;趋势分析用于分析证据的异常趋势,提前预测证据将出现异常的时间间隔,并将其用于每个相关事件的基本推理时间和后续推理预报过程的时延加权后验概率排序。以湿法冶金高铜浸出工艺为例,实施趋势分析、时延加权后验概率推理和基于TDCD的故障预测过程,并与单值/多值混合不确定动态因果图S/M-

DUCG故障预测做对比,表明了所提出方法的合理性与可行性。

参考文献(References)

- [1] Roychowdhury S, Diligenti M, Gori M. Regularizing deep networks with prior knowledge: A constraint-based approach[J]. Knowledge-Based Systems, 2021, 222: 106989.
- [2] Gámiz M L, Linnios N, del Carmen Segovia-García M. Hidden Markov models in reliability and maintenance[J]. European Journal of Operational Research, 2023, 304(3): 1242-1255.
- [3] Luan W P, Yang F, Zhao B C, et al. Industrial load disaggregation based on hidden Markov models[J]. Electric Power Systems Research, 2022, 210: 108086.
- [4] Perikos I, Kardakis S, Hatzilygeroudis I. Sentiment analysis using novel and interpretable architectures of hidden Markov models[J]. Knowledge-Based Systems, 2021, 229: 107332.
- [5] Mor B, Garhwal S, Kumar A. A systematic review of hidden Markov models and their applications[J]. Archives of Computational Methods in Engineering, 2021, 28(3): 1429-1448.
- [6] Chen H S, Yan Z B, Yao Y, et al. Systematic procedure for granger-causality-based root cause diagnosis of chemical process faults[J]. Industrial & Engineering Chemistry Research, 2018, 57(29): 9500-9512.
- [7] Shojaie A, Fox E B. Granger causality: A review and recent advances[J]. Annual Review of Statistics and Its

- Application, 2022, 9: 289-319.
- [8] Zhang C H, Fang W, Zhao B P, et al. Study on fault diagnosis method and application of automobile power supply based on fault tree-Bayesian network[J]. Security and Communication Networks, 2022, 2022: 1-10.
- [9] Li T T, Zhou Y Z, Zhao Y, et al. A hierarchical object oriented Bayesian network-based fault diagnosis method for building energy systems[J]. Applied Energy, 2022, 306: 118088.
- [10] Sheidaei A, Foroushani A, Gohari K, et al. A novel dynamic Bayesian network approach for data mining and survival data analysis[J]. BMC Medical Informatics and Decision Making, 2022, 22(1): 1-15.
- [11] Hanif A, Ali S, Ahmed A. A framework for fault diagnosis using continuous Bayesian network and causal inference[C]. IEEE the 19th International Conference on Industrial Informatics. Palma de Mallorca, 2021: 1-8.
- [12] Yan S, Gu Z, Park J H, et al. Synchronization of delayed fuzzy neural networks with probabilistic communication delay and its application to image encryption[J]. IEEE Transactions on Fuzzy Systems, 2022, 31(3): 930-940.
- [13] Zhang J Y, Hu C L, Zhang Q Q. Constructing a gene regulatory network based on a nonhomogeneous dynamic Bayesian network[J]. Electronics, 2022, 11(18): 2936.
- [14] Mahdi S K, Marco G. Non-homogeneous dynamic Bayesian networks with edge-wise sequentially coupled parameters[J]. Bioinformatics, 2020, 36(4): 1198-1207.
- [15] Shi Q S, Liang X Y. Dynamic causality diagram in fault diagnosis[C]. 2009 International Joint Conference on Computational Sciences and Optimization. Sanya, 2009: 225-227.
- [16] Gu Y K, Zhang M, Zhou X Q. Fault diagnosis of gearbox based on improved DUCG with combination weighting method[J]. IEEE Access, 2019, 7: 92955-92967.
- [17] Zhu Y J, Guo W, Liu H C. Knowledge representation and reasoning with an extended dynamic uncertain causality graph under the pythagorean uncertain linguistic environment[J]. Applied Sciences, 2022, 12(9): 4670.
- [18] Qi B, Zhang L G, Liang J G, et al. Combinatorial techniques for fault diagnosis in nuclear power plants based on Bayesian neural network and simplified Bayesian network-artificial neural network[J]. Frontiers in Energy Research, 2022, 10: 920194.
- [19] Bu X, Nie H, Zhang Z, et al. An industrial fault diagnostic system based on a cubic dynamic uncertain causality graph[J]. Sensors, 2022, 22(11): 4118.
- [20] Sharmin S, Shoyaib M, Ali A A, et al. Simultaneous feature selection and discretization based on mutual information[J]. Pattern Recognition, 2019, 91: 162-174.
- [21] Liu T, Yuan Q Y, Wang L N, et al. Data-based compensation method for optimal operation setting of gold cyanide leaching process[J]. Engineering Letters, 2023, 31(1): 35-42.

作者简介

王姝(1980—),女,副教授,博士,从事智能控制理论及应用、复杂工业过程监测与故障诊断、故障预测与智能维护、大数据分析挖掘、图像识别与智能学习等研究, E-mail: wangshu@ise.neu.edu.cn;

魏楠(1999—),男,硕士生,从事深度学习、复杂工业过程监测与故障诊断等研究, E-mail: 167486084@qq.com;

孟思彤(1996—),女,硕士,从事大数据分析挖掘、复杂工业过程监测与故障诊断等研究, E-mail: mengsitong123456789@qq.com;

王福利(1960—),男,教授,博士生导师,从事复杂工业过程智能控制、过程监测与故障诊断、过程建模与优化、软测量、过程运行状态评价、大数据分析、深度学习、异常工况识别与自愈控制等研究, E-mail: wangfuli@mail.neu.edu.cn.