



ELSEVIER

Contents lists available at ScienceDirect

Engineering

journal homepage: www.elsevier.com/locate/eng



Research
Intelligent Manufacturing—Article

因果模型启发的复杂工业过程软传感器自动化特征选择方法

孙衍宁^a, 秦威^{a,b,*}, 胡锦涛^a, 许鸿伟^a, 孙兆辉^a

^a School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

^b Institute of Artificial Intelligence, Shanghai Jiao Tong University, Shanghai 200240, China

ARTICLE INFO

Article history:

Received 9 January 2022

Revised 20 April 2022

Accepted 8 June 2022

Available online 20 August 2022

关键词

大数据分析

机器学习

质量预测

软传感器

智能制造

摘要

关键性能指标(KPI)的软测量在复杂工业过程决策中起着至关重要的作用。许多研究人员已经使用先进的机器学习(ML)或深度学习(DL)模型开发出了数据驱动的软传感器。其中,特征选择是一个关键的问题,因为一个原始的工业数据集通常是高维的,并不是所有的特征都有利于软传感器的建立。一种完善的特征选择方法不应该依赖于超参数和后续的ML或DL模型。换言之,这种特征选择方法应该能够自动地选择一个特征子集进行软传感器建模,选择的每个特征对工业KPI都有独特的因果影响。因此,本研究提出了一种受因果模型启发的自动特征选择方法,用于工业KPI的软测量。首先,受非线性因果模型的启发,本研究将数据驱动的软传感器与信息论相结合,以量化原始工业数据集中每个特征和KPI之间的因果效应。然后,提出了一种新的特征选择方法,即自动选择具有非零因果效应的特征来构造特征子集。最后,利用所构造的子集,通过AdaBoost集成策略开发KPI的软传感器。两个实际工业应用中的实验证实了该方法的有效性。在未来,该方法也可以应用于其他工业过程,以帮助开发更先进的数据驱动的软传感器。

© 2022 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. 引言

在过去的十年里,凭借工业物联网(IIoT)、大数据分析(BDA)和人工智能(AI)技术的优势,数据智能在复杂工业过程中发展迅速[1–3]。在此背景下,为了更好地利用各种材料和资源创造产品和服务,亟需在产品质量、生产效率、能源消耗和污染排放等方面进一步改进复杂工业过程的性能。然而,这些关键性能指标(KPI)通常不能通过现有的传感器[4–5]进行在线测量和分析。离线实验分析引入了高延迟,使得及时改善工业生产具有挑

战性[6–7]。因此,软传感器技术在工业KPI的在线测量方面引起了广泛的关注。

软传感器技术的目的是通过构建以易测变量为输入、以KPI为输出的数学模型来描述系统的输入-输出行为。该技术大致可以分为两类:第一性原理(白盒)模型和数据驱动(黑盒)模型[8–9]。前一类模型代表了实际系统的因果关系,只有通过物理或化学知识的先验理解才能很好地工作[10–11]。因此,仅关注关联关系而不反映实际因果关系的数据驱动模型,已成为工业KPI软传感器的主流[12]。

* Corresponding author.

E-mail address: wqin@sjtu.edu.cn (W. Qin).

例如, 浅层机器学习 (ML) 模型, 如偏最小二乘 (PLS) 模型、支持向量回归 (SVR) 模型及其扩展形式, 已被用于从复杂工业过程的历史数据中学习质量特征。参考文献[13]中提出了一种优化的稀疏 PLS (OSPLS) 模型来估计工业间歇过程的产品质量, 并提出一种鲁棒多输出最小二乘 SVR (M-LS-SVR) 方法, 用于高炉炼铁过程质量指标的在线估计和控制[14]。此外, 深度学习 (DL) 模型已被广泛研究, 以捕获复杂工业数据中的非线性特征。针对多源异构数据, Ren 等[15]提出了一种结合宽深 (WD) 模型和长短期记忆 (LSTM) 网络的宽深序列 (WDS) 模型, 从非时序变量和时序特征中提取质量相关信息。Yuan 等[16]开发了一种采样间隔注意 LSTM (SIA-LSTM) 模型, 以应对关键质量变量的时间序列与不规则采样数据的软测量建模。Ou 等[17]提出了一种带有质量驱动正则化的堆叠自动编码器 (SAE), 用于工业加氢裂化过程中的质量预测。为了减少 DL 模型的信息丢失和泛化退化, Yuan 等[18]提出了一种基于分层数据增强的 SAE (LWDA-SAE), 用于加氢裂化过程中沸点的软测量。考虑 DL 中输入特征选择的问题, Wang 等[19]提出了一种具有进化特征选择的多目标进化非线性集成学习模型 (MOENE-EFS), 用于高炉炼铁中的硅预测。此外, 在参考文献[20]中提出了一种深度概率迁移学习 (DPTL) 框架, 用于解决多相流过程中的分布差异和数据缺失问题。虽然上述方法在一些工业应用中取得了良好效果, 但仍存在一些研究缺陷。

一方面, 特征选择仍然是一个关键问题, 因为一个原始的工业数据集通常是高维的, 并不是所有的特征都有利于软传感器的建立。数据驱动的软传感器建模对于识别工业数据中的模式, 从而确定工业 KPI 与其相关特征 (变量) 之间的定量关系至关重要。选择一个紧凑且信息丰富的特征子集可以大大降低模型的复杂性, 并帮助我们充分理解复杂工业过程[21–23]的工艺机理。如果所选的特征是 KPI 的因果变量, 那么数据驱动的软传感器无疑将更具可解释性和稳定性。否则, 盲目改进数据驱动模型将引入复杂的模型结构和难以调优的超参数, 这违背了奥卡姆剃刀原理和工业领域的可靠性要求[24]。换句话说, 特征选择方法最好能自动选择一个特征子集用于软传感器建模, 选择的每个特征对工业 KPI 都有独特的因果影响。

另一方面, 实际的工业数据难以获取, 而且成本昂贵, 特别是对于离散工业, 这阻碍了数据驱动的软传感器在工业中的应用。根据生产行为, 复杂工业过程可以分为流程工业和离散工业[25]。流程工业的生产行为要么是连续的, 如化学过程、发电和炼铁; 要么发生在一批难以区

分的材料上, 如食品加工、造纸和注塑。离散工业的生产行为是材料的物理或机械过程, 如发动机装配、半导体制造[26]和家用电器制造, 其中所使用的材料通常是其他工业过程的产品[27–28]。与流程工业相比, 离散工业通常具有更大的规模、更强的动态性和更不清晰的机理。数据的收集几乎完全依赖于从业者的经验, 因此原始的工业数据更加具有非线性、不完备和不确定性的特点。在这种情况下, 具有参数较少、鲁棒性良好和可解释性优势的集成 ML 更适用于机理较弱的复杂工业过程[29]。

综上, 本研究主要关注以下两个科学问题: ①如何量化原始工业数据集中的每个特征和 KPI 之间的因果效应? ②如何自动选择用于软传感器建模的特征子集? 因果模型如格兰杰因果关系、条件独立性检验和结构方程[30–32]已被广泛应用于金融[33]、气候[34]和工业[35–37]的研究。然而, 目前还没有在数据驱动软传感器中将因果效应和特征选择进行集成。主要挑战如下: 由于数据充分性假设, 条件独立性测试将导致 KPI 软测量的信息丢失; 此外, 格兰杰因果关系模型和结构方程模型依赖于数据生成机制的假设。综上所述, 本研究的主要工作内容如下:

(1) 受后非线性因果模型的启发, 将其与信息论相结合, 以量化原始工业数据集中每个特征和 KPI 之间的因果效应。这可以避免数据生成机制的假设, 并为理解复杂工业过程提供有用的见解。

(2) 提出一种新的特征选择方法, 通过自动选择具有非零因果效应的特征来构造特征子集, 从而减少信息丢失, 提高软传感器模型的可解释性, 有助于提高准确性和鲁棒性。

(3) 利用所构建的子集, 通过 AdaBoost 集成策略来开发 KPI 的软传感器。还介绍了两个实际的复杂工业过程: 富士康科技集团的注塑成型过程和广西玉柴集团的柴油发动机装配过程。这两种工业应用中的实验证实了该方法的有效性。

本文其余部分的结构如下: 第 2 节描述了相关工作; 第 3 节和第 4 节提供了对所提方法的详细描述; 第 5 节对两个实际的复杂工业过程进行了实验研究; 最后, 第 6 节总结了本文结论。

2. 相关工作

本节回顾了特征选择和因果发现方法, 这将激发本研究的问题构想和基本思路。

2.1. 特征选择

正如在行业中流传的那样，数据和特征决定了ML的上限，而模型和算法只是接近这个上限。特征选择是指从给定的候选特征集[38]中选择一个特征子集作为ML的输入，主要涉及两方面动机。首先，即使没有先验知识或领域知识，特征选择也有利于充分理解数据，并提供对问题的感性见解[39]。其次，特征选择直接实现了特征降维，有效地降低了ML模型的复杂性[40]。一般来说，特征选择涉及两个关键方面：子集搜索策略和子集评价标准。

2.1.1. 子集搜索策略

取一个具有 M 个输入特征的候选输入特征集 F ，其中： $F = \{X_1, X_2, \dots, X_M\}$ ， X_i (i 是候选特征的数量， $i = 1, 2, \dots, M$)表示候选输入特征，有 2^M 个候选子集 S ， $S \subseteq F$ 。该子集搜索策略的目标是从 F 中选择一个最优的特征子集 S [41]。公式(1)表明，前向搜索策略首先用一个空集初始化 S 。然后，根据子集评价准则，从 F 中选择一个特征，并且在每次迭代时添加到 S 中，直到达到停止阈值。

$$S = S \cup \{EC(S \cup X_i, Y) > ST, X_i \in F \setminus S\} \quad (1)$$

式中， $EC(\cdot)$ 为评价标准； ST 为停止阈值； Y 为输出特征。

另一种策略被称为背向搜索，如公式(2)所示。首先初始化 $S = F$ 。然后，在每次迭代中从 S 中删除一个特征，直到达到停止阈值。

$$S = S \setminus \{EC(S \setminus X_i, Y) < ST, X_i \in S\} \quad (2)$$

这两种策略都是贪婪的，因为只实现了局部最优性。此外，难以确定最佳的具有良好的可解释性和理论基础的评价标准和停止阈值。

2.1.2. 子集评估标准

许多评价标准被用来判断是否在每次迭代中保留候选特征，如关联程度和发散程度，以及ML模型的性能。方差 σ^2 度量特征发散的度，而不考虑输入和输出特征之间的关联[42]。皮尔逊相关系数(PCC)选择与目标最相关的输入特征，但只关注线性关联[43]。最大信息系数

(MIC)检测两个变量之间的非线性关联[44]，但需要更多的样本，总关联很容易被低估。基于上述准则的特征选择不依赖于ML模型，也被称为一种过滤方法。

ML模型的参数，如决策树的信息增益和回归系数[45]也可以作为子集评价标准，用来衡量特征的重要性或权重。这种方法被称为嵌入法，它依赖于一个计算成本高昂的ML训练过程，并且本质上是基于关联的。除了将性能最大化作为评价标准外，包装法还将ML与优化算法如遗传[38]、进化[19]和粒子群算法[39]相结合，以自动选择最优特征组合。包装法也带来了高昂的计算成本，而且很容易导致过拟合，特别是在工业应用中。

2.2. 因果发现

发现因果关系是科学研究和技术进步的一项基本任务。该方法严格区分因果变量，比关联关系更有效地揭示其机制和指导决策。在不考虑时间滞后效应的情况下，因果发现方法主要依赖于条件独立检验和结构方程模型从观测数据[46]中学习因果效应。

2.2.1. 条件独立性测试

给定一组三元变量 $\{X, Y, Z\}$ ，具体的因果结构可以通过变量之间的条件独立性来检验。如图1所示，如果三元变量之间的关系是 X, Y, Z 相互独立，则因果结构必须是马尔可夫等价类[图1(a)]。如果 X 和 Z 本身是独立的，但一旦引入 Y 就不是独立的，则因果结构必然是V型结构[图1(b)]。在此基础上，适用范围广泛的Peter-Clark(PC)和归纳因果关系(IC)算法通过因果骨架和因果方向的两阶段过程来学习因果结构[47-48]。

2.2.2. 结构方程模型

数据生成机制的假设描述了结果变量是如何由原因变量和因果机制决定的，包括线性非高斯无环模型(LiN-GAM)[49]、加性噪声模型(ANM)[50]、信息几何因果推断(IGCI)[32]和后非线性模型(PNM)[51]。作为最一般的模型，如图2所示，PNM包含了原因 X 的非线性影

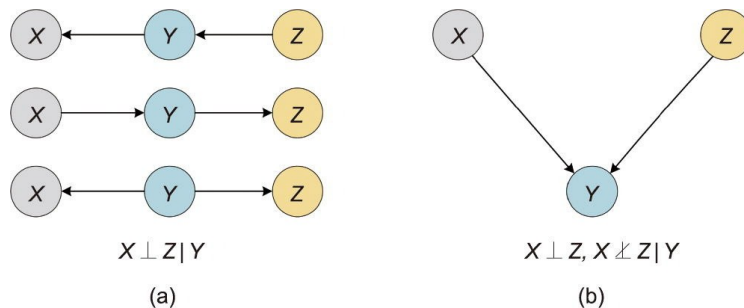


图1. 三元变量之间的因果关系。(a) 马尔可夫等价类；(b) V型结构。

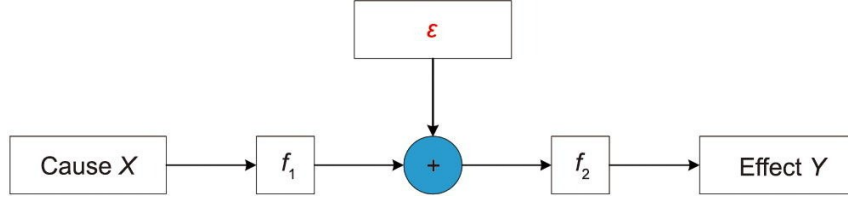


图2. 后非线性因果模型。 f_1 、 f_2 ：非线性函数； ε ：噪声或干扰。

响 f_1 、噪声或干扰 ε ，以及结果 Y 的测量失真 f_2 。该公式如下所示。

$$Y=f_2(f_1(X)+\varepsilon) \quad (3)$$

式中， $\varepsilon \perp X$ ； f_1 和 f_2 是非线性函数； f_2 应该是可逆的。

由于数据充分性和条件独立性检验的局限性，从PC和IC算法中得到的因果结构并不等同于实际的物理对象。基于这种因果结构的特征选择会导致显著的信息损失，从而无法获得ML的最佳特征组合。相比之下，结构方程模型中的PNM可以更有效地结合因果发现和特征选择。

对于特征选择问题，嵌入法和包装法依赖于ML和计算成本高昂的训练过程。它们的性能直接受所选的ML模型的影响。过滤法，如基于方差的方法、基于PCC的方法和基于MIC的方法，不依赖于ML模型，并通过预先手动设置停止阈值来选择特征的子集。一个典型的停止阈值包括特定数量的选定特征，如特定的方差值、PCC值或MIC值。显然，很难确定一个具有良好的可解释性和理论基础的停止阈值。因果发现通过量化原始工业数据集中的每个特征和KPI之间的因果效应，进而自动地选择特征用于数据驱动的软传感器建模，为解决这个问题带来了新的思路。下一节将详细介绍所提出的方法。

3. 受因果模型启发的特征选择

3.1. PNM与信息理论

给定一组原因变量 $\{X_1, X_2, \dots, X_k\}$ （其中 k 是变量的数量）和结果变量 Y ，公式（3）中的PNM可以被扩展到公式（4）。

$$Y=f_2(f_1(X_1, X_2, \dots, X_k)+\varepsilon_k) \quad (4)$$

为了发现另一个变量 X_{k+1} 和 Y 之间的因果关系，公式（4）被进一步扩展到公式（5）。

$$Y=f_2(f_1(X_1, X_2, \dots, X_k, X_{k+1})+\varepsilon_{k+1}) \quad (5)$$

如果 X_{k+1} 减少了噪声项，则它包含了 Y 的因果信息。因此， X_{k+1} 对 Y 的因果效应可以用公式（6）来量化。

$$CE_{X_{k+1} \rightarrow Y} = \frac{1}{2} \log \frac{\sigma^2(\varepsilon_k)}{\sigma^2(\varepsilon_{k+1})} \quad (6)$$

式中，CE是因果效应。

问题是需要建立和依赖这两个回归模型，这两个模型具有较高的计算复杂度并影响精度。此外，对PNM中数据生成机制的假设也有待改进。

本研究通过信息论定义因果效应，以此来解决这些问题。在信息论中，采用Shannon熵来测量离散随机变量 X 中的不确定性和平均信息，如下：

$$H(X) = -\sum_x P(x) \log P(x) \quad (7)$$

式中， $H(\cdot)$ 为Shannon熵； $P(x)$ 表示概率质量函数； x 为 X 的观测值。

两个离散随机变量 X 和 Y 的总不确定性可以通过联合熵计算如下：

$$H(X, Y) = -\sum_{x, y} P(x, y) \log P(x, y) \quad (8)$$

式中， y 是 Y 的观测值。

如果给定 X ，则通过考虑 X 中的信息，可以减少 Y 中的不确定性，那么， Y 中的剩余不确定性可以用条件熵计算如下：

$$H(Y|X) = H(X, Y) - H(X) \quad (9)$$

通过将公式（7）和公式（8）代入公式（9），条件熵可以用 X 和 Y 的概率来表示。信息论通过考虑不确定性而不是方差来扩展PNM [30]。换句话说，可以通过测量 X_{k+1} 降低 Y 的不确定性的程度来量化因果效应。如图3所示，给定一组原因变量 $\{X_1, X_2, \dots, X_k\}$ ， Y 中的剩余不确定度可以用下式计算：

$$H(Y|X_1, X_2, \dots, X_k) = H(X_1, X_2, \dots, X_k, Y) - H(X_1, X_2, \dots, X_k) \quad (10)$$

当进一步给出 X_{k+1} ， Y 中的剩余不确定性可以表示为：

$$H(Y|X_1, X_2, \dots, X_k, X_{k+1}) = H(X_1, X_2, \dots, X_k, X_{k+1}, Y) - H(X_1, X_2, \dots, X_k, X_{k+1}) \quad (11)$$

因此，得到 X_{k+1} 对 Y 的因果效应如下：

$$CE_{X_{k+1} \rightarrow Y} = H(Y|X_0, X_1, \dots, X_k) - H(Y|X_0, X_1, \dots, X_k, X_{k+1}) \quad (12)$$

公式（12）仅依靠信息论来实现无回归模型的因果效应量化。此外，数据离散化是计算连续随机变量熵的一种重要的数据预处理技术。在本研究中，将一种基于直方图

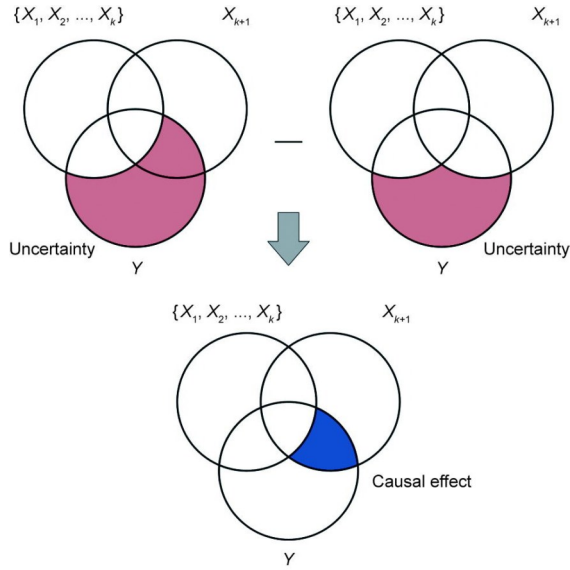


图3. 改进的因果效应 Venn 图。其中第一个红色阴影表示，当给定一组原因变量 $\{X_1, X_2, \dots, X_k\}$ 时， Y 中的剩余不确定性；第二个红色阴影表示，当进一步给出 X_{k+1} 时， Y 中的剩余不确定性；蓝色阴影表示 X_{k+1} 对 Y 的因果效应。

的方法应用于离散化数据，并得到最优的箱数 n_h ：

$$n_h = \max \left(\frac{R}{2 \cdot \text{IQR} \cdot n^{\frac{1}{3}}}, \log_2(n+1) \right) \quad (13)$$

式中， R 为数据范围； IQR 为四分位数范围； n 为样本数。

3.2. 基于因果效应的自动特征选择

本研究提出了一种新的特征选择策略，该策略以前向搜索策略作为子集搜索策略和以公式 (12) 中的因果效应作为子集评价标准。其形式表达式如下：

$$S = S \cup \{CE_{X_i \rightarrow Y} \neq 0, X_i \in F \setminus S\} \quad (14)$$

与公式 (1) 和公式 (2) 相比，公式 (14) 中所示的特征选择方法只需要按照特定的顺序遍历所有候选输入特征 X_i ，不需要设置停止阈值，并自动选择具有非零因果效应的输入特征组合。在实际执行过程中，根据每个候选输入特征 X_i 与输出特征 Y 之间的互信息来确定遍历顺序。算法 1 给出了基于因果效应的自动特征选择算法的伪代码。该方法的详细实现过程如图 4 所示。

4. 基于决策树 AdaBoost 的软传感器建模

本研究以决策树为基本学习器，采用 AdaBoost 集成 ML 算法对工业 KPI 进行软传感器建模。需要指出的是，该模型并不是为了优于所有现有模型而设计的，而是强调因果特征的重要性。一旦提取了因果信息，数据驱动的软传感器可以达到令人满意的准确性和可解释性。在未来，

Algorithm 1. Causal effect-based automatic feature-selection algorithm.

Input: Dataset $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$ (where N is the number of dataset) with candidate input feature set $F = \{X_1, X_2, \dots, X_M\}$ and output feature Y

Output: Casual feature set S

- 1: Initialize: $S = \emptyset$; dataset discretization processing using Eq. (13)
- 2: **for** $i = 1$ to M **do**
- 3: Calculate the causal effect $CE_{X_i \rightarrow Y}$ of X_i on Y using Eq. (12)
- 4: **if** $CE_{X_i \rightarrow Y} \neq 0$, **then** $S = S \cup \{X_i\}$
- 5: **end for**
- 6: **return** S

将研究更先进的 ML 或 DL 模型。

4.1. 决策树回归

决策树回归主要是一种分类和回归树 (CART) 算法，可以解决分类或回归问题。取训练数据集 $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$ (N 是样本数量)。当应用 CART 求解回归问题时，基于二分递归切分的思想，利用平方误差最小化准则选择最优分割变量 j 和分割点 s ，即求解以下公式：

$$\min_{j,s} \left[\min_{c_1} \sum_{x_i \in R_1(j,s)} (y_i - c_1)^2 + \min_{c_2} \sum_{x_i \in R_2(j,s)} (y_i - c_2)^2 \right] \quad (15)$$

式中， c_1 和 c_2 是输出值； R_1 和 R_2 是输入空间中的两个区域。

然后，通过变量 j 和点 s 将输入空间划分为 R_1 和 R_2 。从这个节点中生成两个子节点，分别包含 N_1 和 N_2 样本。

$$R_1(j,s) = \{x | x^{(j)} \leq s\}, R_2(j,s) = \{x | x^{(j)} > s\} \quad (16)$$

这两个区域的最优输出值 \hat{c}_1 、 \hat{c}_2 被进一步确定如下：

$$\hat{c}_1 = \frac{1}{N_1} \sum_{x_i \in R_1(j,s)} y_i, \hat{c}_2 = \frac{1}{N_2} \sum_{x_i \in R_2(j,s)} y_i \quad (17)$$

让这个过程的循环，直到满足结束条件；最后，将输入空间划分为 W 个区域 (R_1, R_2, \dots, R_W) 以生成决策树：

$$f(x) = \sum_{w=1}^W \hat{c}_w I(x \in R_w) \quad (18)$$

式中， $I(\cdot)$ 为指示函数； w 是区域数量。如果 $x \in R_w$ ，则 $I = 1$ ；否则， $I = 0$ 。

生成回归树后，从底部到根节点对回归树进行修剪。对于每个修剪情况，生成一个子树，从而形成一个子树序列 $f_1(x), f_2(x), \dots, f_n(x)$ 。接下来，在独立验证数据集上使用交叉验证方法，比较每个子树相对于验证集的平方误差，并选择最优决策树 $f_\alpha(x)$ (α 是序列的数量， $\alpha = 1, 2, \dots, n$)。

4.2. AdaBoost 集成学习

如算法 2 所示，给定 $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$ 作为训练集， $W_t = \{w_t(1), w_t(2), \dots, w_t(N)\}$ (t 为迭代次数，

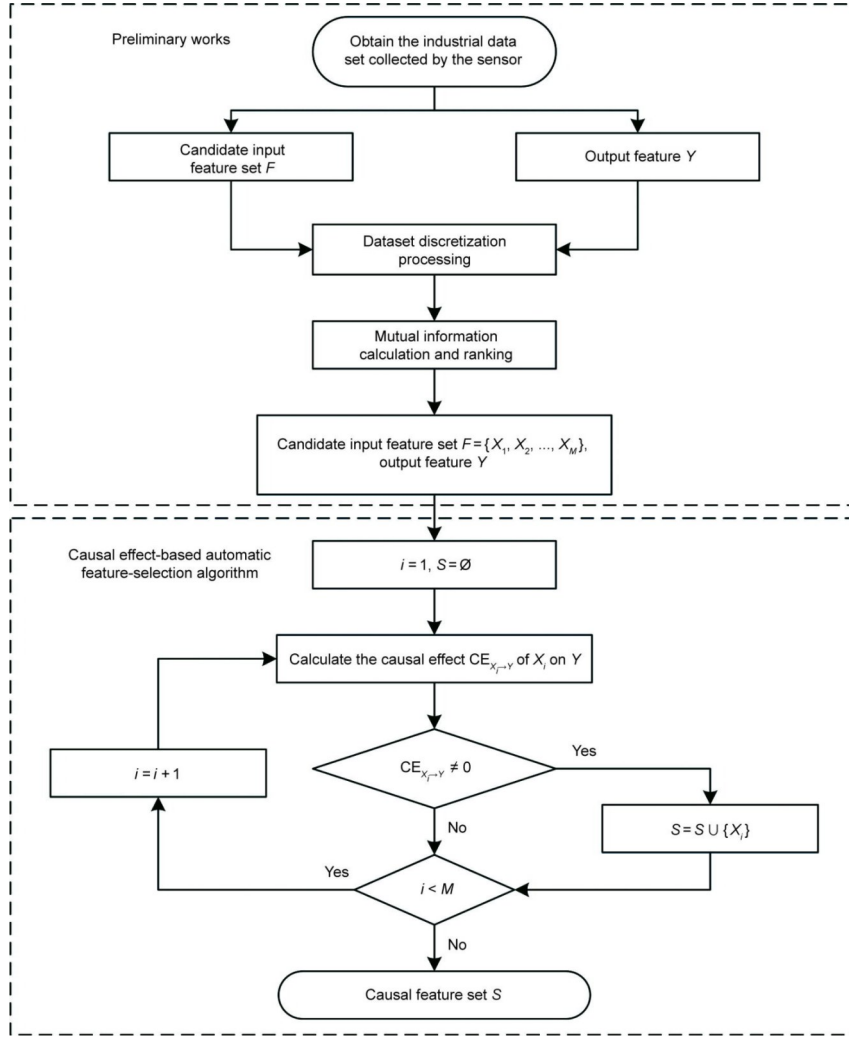


图4. 本文所提方法的流程图。

$t = 1, 2, \dots, T$, 其中 T 为总迭代次数) 表示第 t 次提升迭代时在 D 上的权值分布。在以后的迭代中, 将通过增加性能较差的样本的权重和减少性能较好的样本的权重来更新权重分布。度量性能的平均损失函数由下式计算:

$$\bar{L}_t = \sum_{i=1}^N L_t(i) w_t(i) \quad (19)$$

式中, L_t 是一个范围为 0~1 的损失函数。三个候选人 L_t 见参考文献[52]; 本研究采用指数法, 如下:

$$L_t(i) = 1 - \exp\{-l_t(i) / \max(l_t(i))\}, i = 1, 2, \dots, N \quad (20)$$

式中, $l_t(i) = |f_t(x_i) - y_i|$, 是每个训练示例的损失。权值更新程序如下:

$$w_{t+1}(i) = w_t(i) \alpha_t^{1-L_t(i)} / Z_t \quad (21)$$

式中, $\alpha_t = \bar{L}_t / (1 - \bar{L}_t)$ 为权重更新参数; Z_t 是使 W_{t+1} 呈概率分布的归一化因子。最终的 AdaBoost 回归结果可以由下式获得:

$$f(x) = - \sum_{t=1}^T f_t(x) \log \alpha_t \quad (22)$$

Algorithm 2. AdaBoost ensemble ML for soft sensor modeling.

Input: Training set with N datasets $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$, the basic learning algorithm, and the total number of iterations T

Output: The soft sensing model $f(x)$

- 1: Initialize: $W_1 = \{w_1(1), w_1(2), \dots, w_1(N)\}$, $w_1(i) = 1/N, i = 1, 2, \dots, N$
- 2: **for** $t = 1$ to T **do**
- 3: Take a sample set R_t from D using distribution $W_t = \{w_t(1), w_t(2), \dots, w_t(N)\}$
- 4: Calculate the loss function $L_t(i)$ for each training sample
- 5: Calculate the average loss \bar{L}_t
- 6: Set α_t to update distribution W_t as $W_{t+1} = \{w_{t+1}(1), w_{t+1}(2), \dots, w_{t+1}(N)\}$
- 7: **end for**
- 8: **return** $f(x)$

5. 实验研究

在本节中, 通过对两个实际复杂工业过程的实验验证

了该方法。

5.1. 实验装置

理论推导结果表明，所提出的特征选择方法是一种过滤法。该方法以因果效应作为子集评价标准，并采用前向搜索策略自动选择特征子集来训练软传感器模型。该方法不需要设置停止阈值。在这个子集中的每个特征 X_i 对 Y 有独特的因果效应，显然其他过滤法缺乏这一优势。

特征选择的性能评价通常考虑两个方面：所选特征的数量和软传感器的性能。一般希望使用最少数量的输入特征来实现软传感器的最佳性能。众所周知，基于方差、基于PCC和基于MIC的方法是最简单、最有效的过滤式特征选择方法，具有良好的泛化性。因此，将这三种方法作为比较的基准。首先使用所提出的方法来确定所选特征的数量（标记为 K ）。然后，将三个基准的停止阈值设为 K 。最后，利用由上述四种方法得到的特征子集对基于AdaBoost决策树的软传感器模型进行训练，并比较软传感器的性能。在此过程中，将两个复杂工业过程的实验数据按60:40的比例随机分为两组，即以60%作为训练集，以40%作为测试集。均方根误差（RMSE）和决定系数 R^2 是两个广泛使用的性能评价指标，由公式（23）和公式（24）定义，并分别在本研究中被采用。最终，如果本文所提方法的RMSE和 R^2 优于三个基准的RMSE和 R^2 ，则可以验证所提方法的有效性。

$$\text{RMSE} = \sqrt{\sum_{i=1}^{N_T} (y_i - \hat{y}_i)^2 / N_T} \quad (23)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N_T} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N_T} (y_i - \bar{y}_i)^2} \quad (24)$$

式中， N_T 是测试集中的样本数量； y_i 为第 i 个样本的真实值； \hat{y}_i 是软传感器模型的估计值； \bar{y}_i 是所有估计值的平均值。

本研究中的所有代码都是用Python 3.7编写的。基于AdaBoost决策树的软传感器模型的四个最重要的超参数是每个决策树回归模型的最大深度和最小样本分割，以及AdaBoost集成学习的估计器数量和学习率。在两个实验中，通过在默认值附近进行上下微调，这四个参数分别被设置为10、5、20和1.3。所有其他超参数都使用默认值。硬件环境为Intel (R) Core (TM) i7-8700 中央处理器 (CPU) @3.20 GHz 32.00G 随机存取存储器 (RAM)。

5.2. 注塑工艺的实验研究

第一个复杂工业过程是来自中国富士康科技集团的注塑成型过程。该过程使用注塑机（图5）在高温下熔化塑料原料。然后在高速、高压下将塑料熔体注入模具。熔体在恒定压力下经历复杂的物理化学变化，形成塑料制品。通过这个过程的重复操作，可以生产出大量相同的产品。在此过程中，最终产品质量的测量有较大延迟，严重影响了确保质量稳定性的及时决策。因此，采用注塑成型过程来验证和应用所提出的方法。收集了16 600个生产批次的数据，包括86个候选输入特征，以产品尺寸作为KPI [53]。

基于第3节，量化了86个候选输入特征对注塑过程的产品尺寸（mm）的因果效应。如图6所示，发现只有

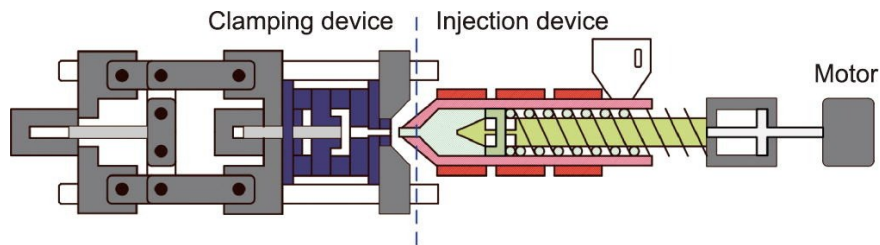


图5. 注塑机示意图。

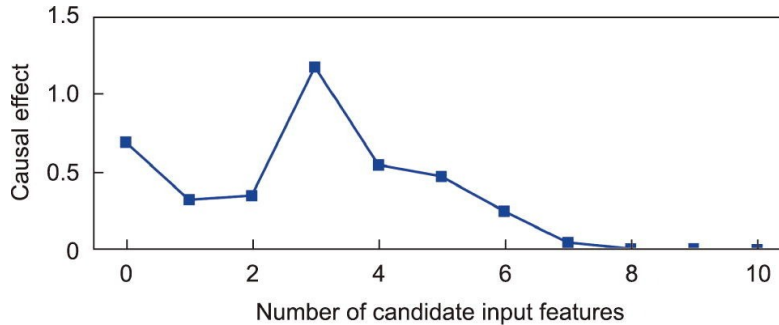


图6. 注塑过程中不同候选输入特征对产品尺寸的因果效应。

9个候选输入特征包含关于产品尺寸的因果信息。考虑这9个特征，其余的特征对其没有因果影响。因此，利用这9个特征作为软传感器模型的输入特征来估计产品尺寸的值。这9个特征是：瞬时流量 ($\text{m}^3 \cdot \text{s}^{-1}$)、循环时间 (s)、顶升时间 (s)、冷却后时间 (s)、模具温度 ($^{\circ}\text{C}$)、夹紧时间 (s)、喷射时间 (s)、夹紧压力 (Pa) 和开启时间 (s)。

表1显示了RMSE和 R^2 对软传感器模型在不同的特征选择方法下进行的分析。可以看到，基于因果效应的特征选择方法提供了最低的RMSE和最大的 R^2 。由于产品尺寸的因果信息提取准确，因此该方法优于三个基准方法。此外，还有效地消除了冗余的非因果信息。与基准方法相比，该方法不需要设置停止阈值，可以自然地避免信息丢失。

表1 RMSE和 R^2 软传感器模型在不同特征选择方法下的注塑过程

Methods	RMSE (mm)	R^2 (%)
Variance-based	0.031	65.1
PCC-based	0.031	65.2
MIC-based	0.027	73.2
Cause effect-based	0.023	80.4

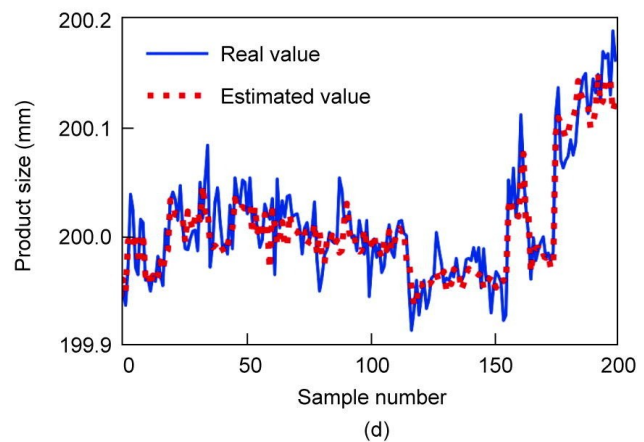
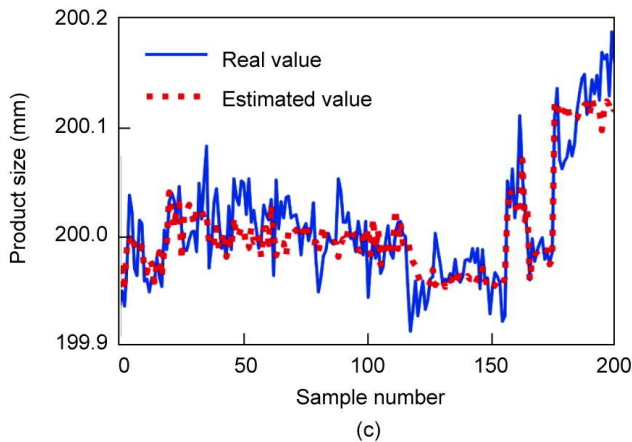
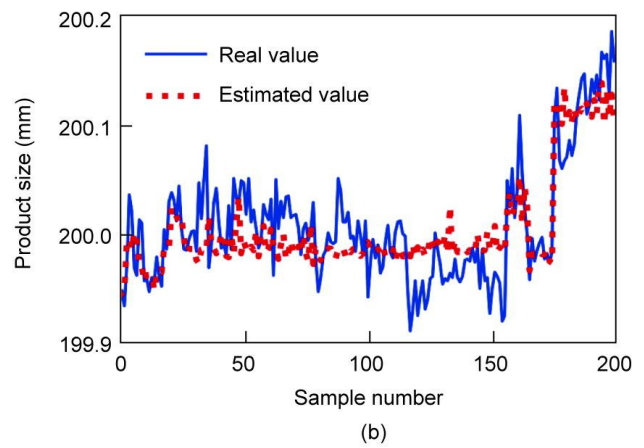
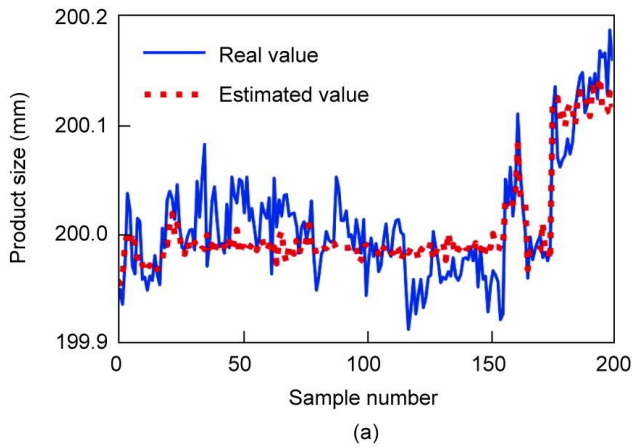


图7. 注塑过程中不同特征选择方法下产品尺寸的软传感器结果。(a) 方差；(b) PCC；(c) MIC；(d) 因果效应。

图7显示了不同的特征选择方法下产品尺寸的软传感器结果。可以看出，基于因果效应的方法比基于三个基准的方法能更有效地估计质量的轻微波动。图8显示了不同特征选择方法下软传感器结果的散点图和概率密度曲线。可以看出，基于因果效应的方法的估计值更接近真实值。此外，基于因果效应的方法的概率密度曲线比基准方法的概率密度曲线“更瘦”“更高”，这也证明了该方法具有更好的精度。

5.3. 柴油机装配工艺的试验研究

第二种复杂工业过程是广西玉柴集团有限公司的柴油机装配工艺。如图9所示，机器部件通过8条总装线组装成柴油机产品，包括主流水线、5条分装线、性能测试线、包装线。在相同的工况下，额定功率的一致性是最重要的KPI之一，但其检查需要耗时和高成本的台架测试。对1763个样本进行测试。对于每个样本，沿着装配过程[36–37]收集了39个过程变量的数据，并作为候选输入特征来验证和应用所提出的方法。

该方法在柴油机装配过程中得到了进一步的验证和应用。同样，39个候选输入特征对柴油机产品额定功率

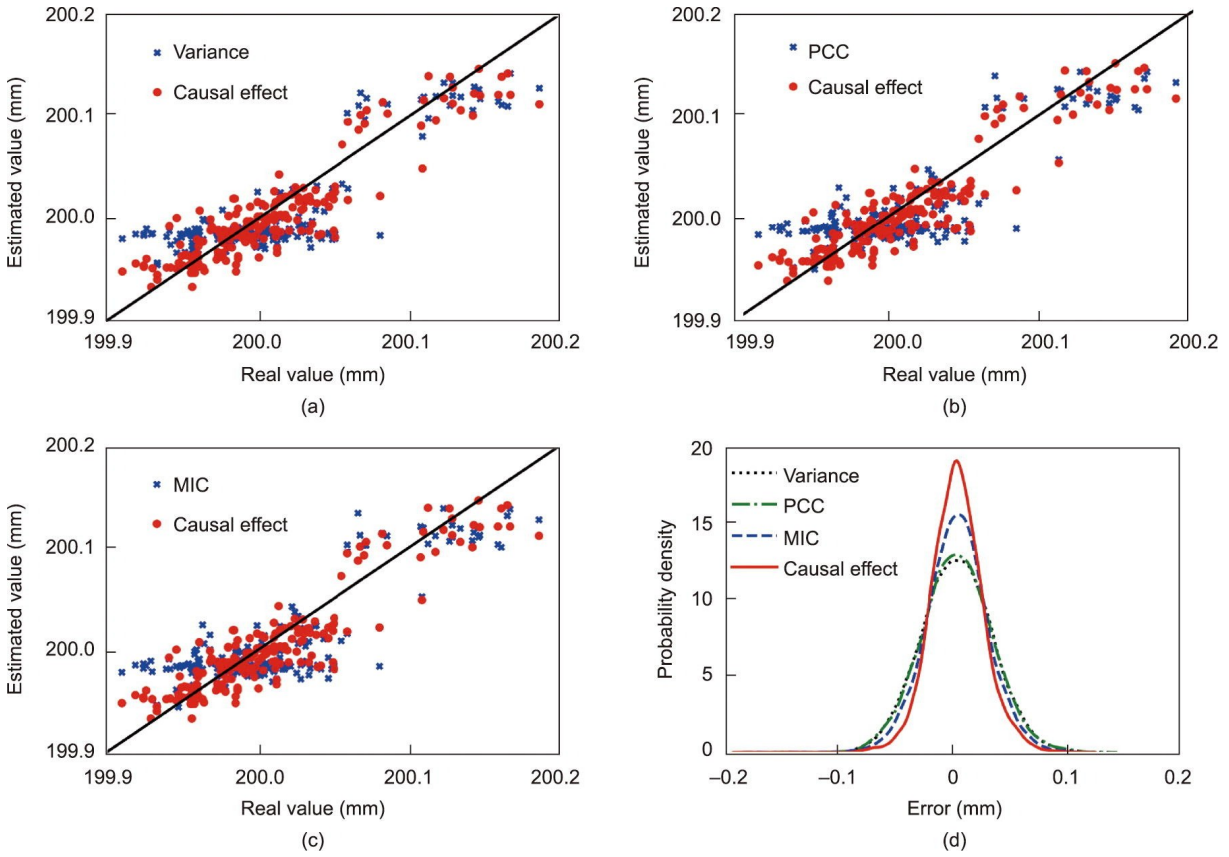


图8. 注塑过程中不同特征选择方法下软传感器结果的散点图和概率密度曲线。

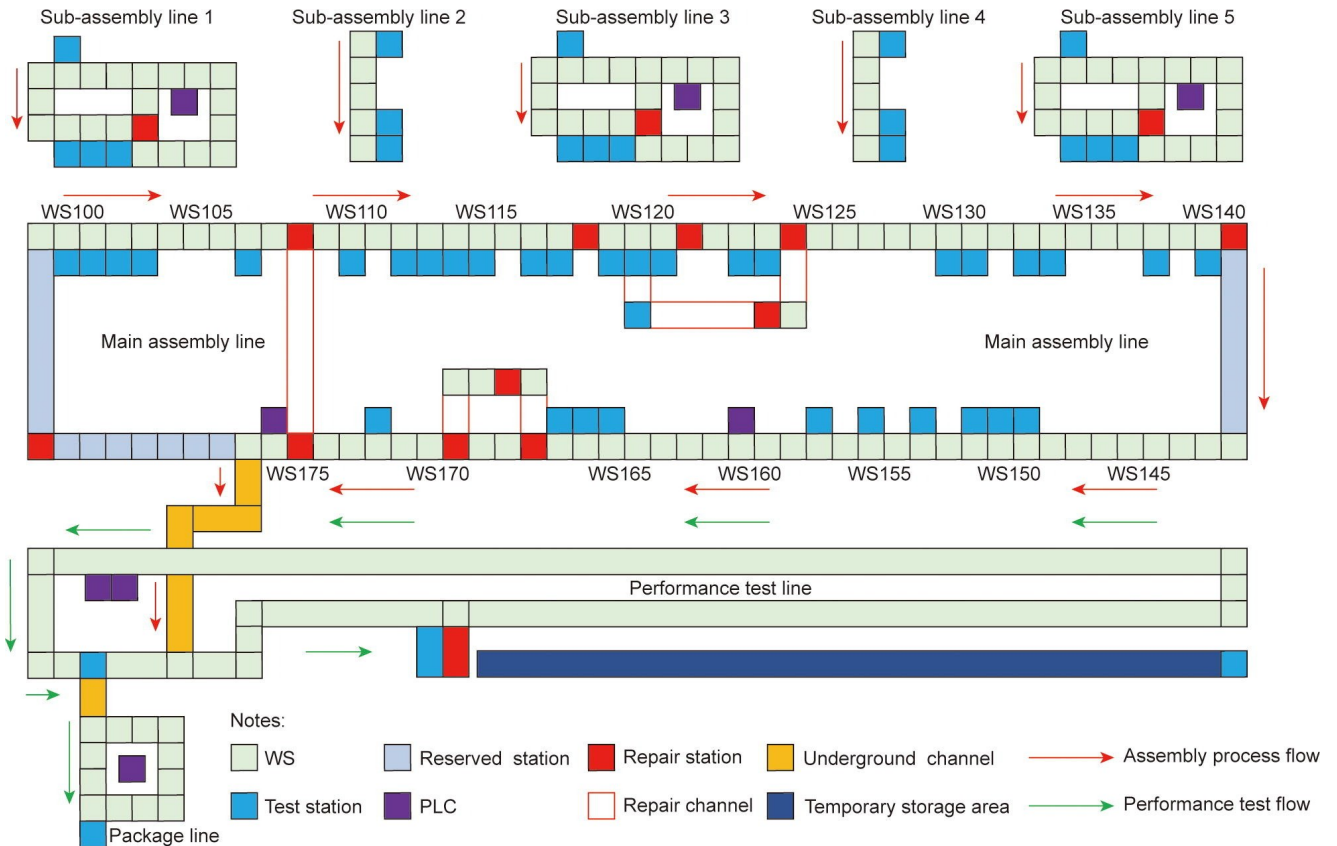


图9. 柴油机装配工艺图。PLC: 可编程逻辑控制器。

(kW) 的因果效应也被量化。如图 10 所示, 发现只有 6 个候选输入特征包含关于额定功率的因果信息, 而鉴于这 6 个特征, 其余的特征对柴油机产品额定功率没有因果影响。因此, 利用这 6 个特征作为软传感器模型的输入特征来估计额定功率的值。这 6 个特征包括: 每 100 公里的油耗 (L)、运行时间 (min)、油耗率 (%)、中冷器入口压力 (Pa)、中冷器入口温度 ($^{\circ}\text{C}$) 和轴向间隙 (mm)。

表 2 显示了 RMSE 和 R^2 对软传感器模型在不同的特征选择方法下进行的分析。同样, 可以看到, 基于因果效应的特征选择方法提供了最低的 RMSE 和最大的 R^2 。值得注

意的是, 这三个基准的 R^2 非常低, 说明很难用所选的变量来解释输出特征。图 11 显示了不同特征选择方法下额定功率的软传感器结果。可以看出, 基于因果效应的方法比三个基准能更准确地估计额定功率的值。图 12 显示了不同特征选择方法下软传感器结果的散点图和概率密度曲线。基于因果效应的方法的估计值比其他方法更接近实际额定功率。此外, 基于因果效应的方法的概率密度曲线“更瘦”“更高”, 再次证明了所提出的模型比三个基准的模型具有更好的精度。

根据这两个实验的结果, 可以得到以下几点见解。该

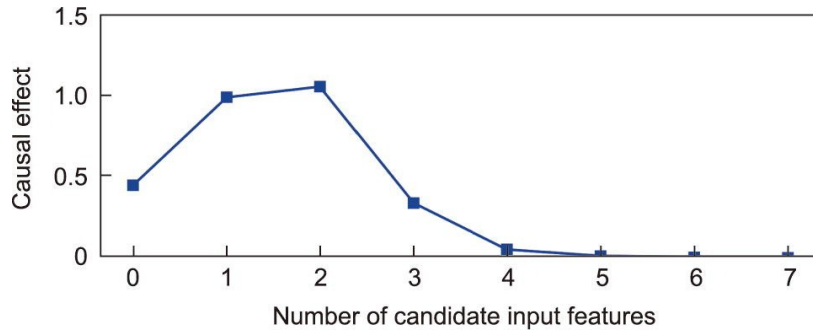


图 10. 不同特征对柴油机装配过程中额定功率的因果影响。

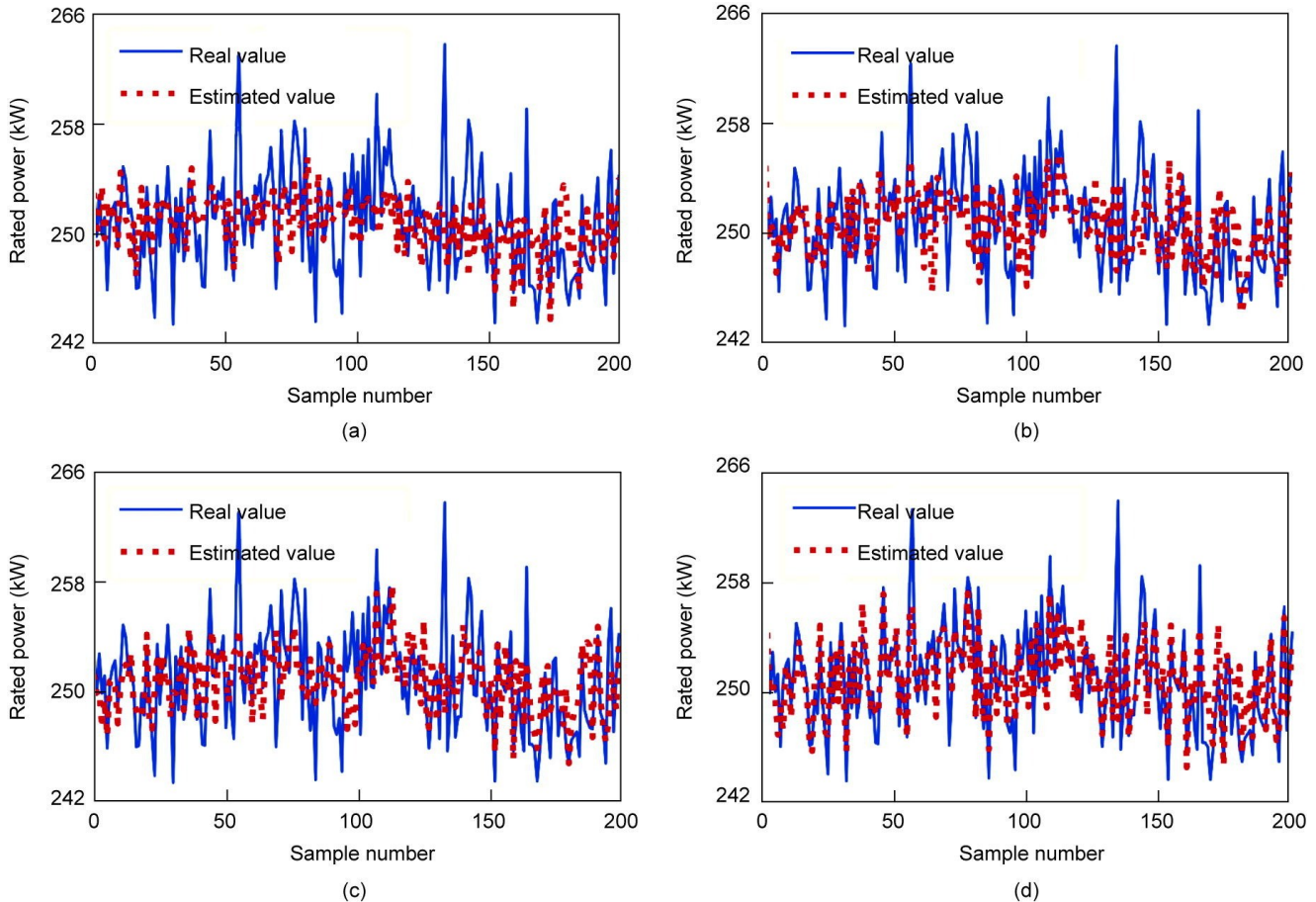


图 11. 柴油机装配过程中不同特征选择方法下产品尺寸的软传感器结果。(a) 方差; (b) PCC; (c) MIC; (d) 因果效应。

表2 RMSE和 R^2 软传感器模型在不同特征选择方法下的应用

Methods	RMSE (kW)	R^2 (%)
Variance-based	3.207	18.5
PCC-based	3.078	24.9
MIC-based	3.066	25.5
Cause effect-based	2.215	61.1

方法是有效和通用的，有助于理解复杂的工业过程。在实际的工业应用中，该方法可以从原始的工业数据集中选择一个紧凑且信息丰富的特征子集。例如，在注塑过程中的86个候选输入特征中，只有9个候选输入特征包含关于产品尺寸的因果信息。与这9个特征相比，其他特征对于软传感器建模是无用或冗余的。通过两种方式可以进一步提高软传感器的性能。一是开发一个更先进的数据驱动模型，该模型可以更充分地拟合所选特征的数据分布。根据经验，当选择相同的输入特征时，现有的数据驱动模型的性能是相似的。因此，本文只介绍了一个用于软传感器建模的AdaBoost集成学习模型，而对不同模型比较则超出了本文的研究范围。另一种方法是通过第一性原理，对工业过程有更深入的了解，从而获得更全面和足够的数

据，以帮助训练更好的数据驱动模型。换句话说，虽然正在开发数据驱动的方法，但对复杂工业过程第一性原理的研究不应被忽视。

6. 结论

本研究提出了一种基于因果效应的特征选择方法，用于开发复杂工业过程中KPI的软传感器。将PNM与信息论相结合，提出了一种因果效应量化方法来提取KPI的因果信息。该方法可以为KPI的软测量提供有用见解，并有助于提高ML的准确性和可解释性。此外，采用AdaBoost集成的决策树回归进行软传感器建模，几乎不需要对参数进行微调即可获得优异的性能。对实际工业过程的实验研究证实了该方法的有效性和应用前景。

然而，PNM是一个非时序因果模型，因此本文没有考虑因果关系的时间滞后效应。如果将该方法应用于时间序列数据，则需要首先估计因果延迟。这是在未来的工作中可能会讨论的另一个主题。此外，本研究侧重于基于因果效应的特征选择方法，而对下游ML模型的研究较薄

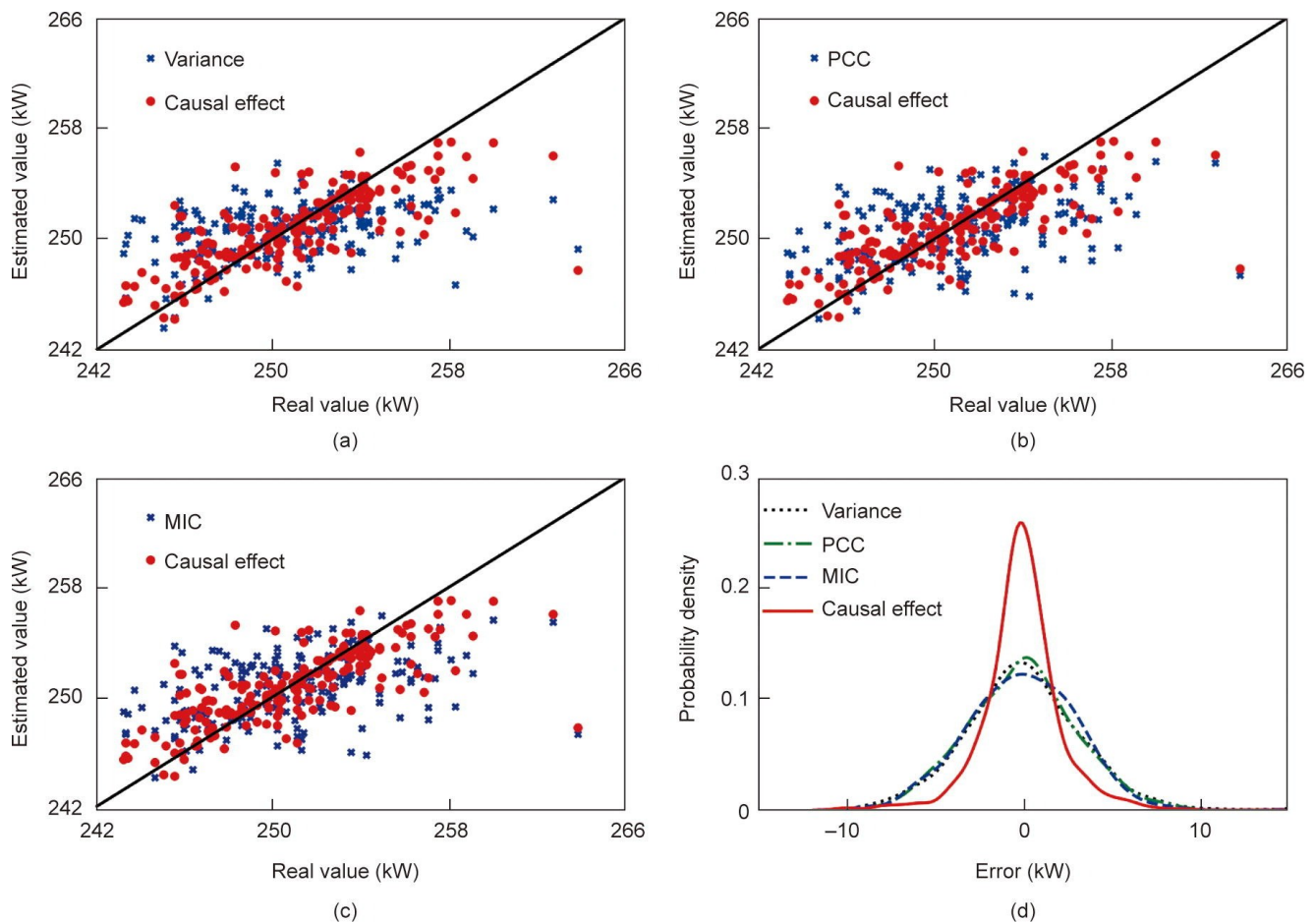


图12. 柴油机装配过程中不同特征选择方法下软传感器结果的散点图和概率密度曲线。

弱。正如前面提到的，在包含因果信息的相同输入特征下，利用先进的ML模型对软测量结果的改进是有限的。因此，特别是对于复杂的工业场景，未来的工作将集中在基于不确定性量化理论的KPI模型的区间估计和风险评估上。

致谢

感谢教育部-中国移动科研基金项目(MCM20180703)和国家重点研发计划项目(2020YFB1711100)的资金支持，感谢富士康科技集团提供注塑工艺的历史数据集，以及感谢编辑和审稿人的宝贵意见。

Compliance with ethics guidelines

Yan-Ning Sun, Wei Qin, Jin-Hua Hu, Hong-Wei Xu, and Poly Z. H. Sun declare that they have no conflict of interest or financial conflicts to disclose.

References

- [1] Gao L, Shen W, Li X. New trends in intelligent manufacturing. *Engineering* 2019;5(4):619–20.
- [2] Wang J, Zheng P, Lv Y, Bao J, Zhang J. Fog-IBDIS: industrial big data integration and sharing with fog computing for manufacturing systems. *Engineering* 2019;5(4):662–70.
- [3] Wang J, Ma Y, Zhang L, Gao RX, Wu D. Deep learning for smart manufacturing: methods and applications. *J Manuf Syst* 2018;48:144–56.
- [4] Yuan X, Gu Y, Wang Y, Yang C, Gui W. A deep supervised learning framework for data-driven soft sensor modeling of industrial processes. *IEEE Trans Neural Netw Learn Syst* 2020;31(11):4737–46.
- [5] Liu C, Wang K, Wang Y, Yuan X. Learning deep multimanifold structure feature representation for quality prediction with an industrial application. *IEEE Trans Ind Inform* 2022;18(9):5849–58.
- [6] Ren L, Meng Z, Wang X, Zhang L, Yang LT. A data-driven approach of product quality prediction for complex production systems. *IEEE Trans Ind Inform* 2021;17(9):6457–65.
- [7] Geng Z, Dong J, Chen J, Han Y. A new self-organizing extreme learning machine soft sensor model and its applications in complicated chemical processes. *Eng Appl Artif Intell* 2017;62:38–50.
- [8] Shi J, Zhou S. Quality control and improvement for multistage systems: a survey. *IIE Trans* 2009;41(9):744–53.
- [9] Schrangl P, Tkachenko P, del Re L. Iterative model identification of nonlinear systems of unknown structure: systematic data-based modeling utilizing designs of experiments. *IEEE Control Syst Mag* 2020;40(3):26–48.
- [10] Mao J, Chen D, Zhang L. Mechanical assembly quality prediction method based on state space model. *Int J Adv Manuf Technol* 2016;86:107–16.
- [11] Zhou X, Zhang Y, Mao T, Zhou H. Monitoring and dynamic control of quality stability for injection molding process. *J Mater Process Technol* 2017; 249: 358–66.
- [12] Sun Q, Ge Z. Deep learning for industrial KPI prediction: when ensemble learning meets semi-supervised data. *IEEE Trans Ind Inform* 2021;17(1):260–9.
- [13] Jiang Q, Yan X, Yi H, Gao F. Data-driven batch-end quality modeling and monitoring based on optimized sparse partial least squares. *IEEE Trans Ind Electron* 2020;67(5):4098–107.
- [14] Zhou P, Guo D, Wang H, Chai T. Data-driven robust M-LS-SVR-based NARX modeling for estimation and control of molten iron quality indices in blast furnace ironmaking. *IEEE Trans Neural Netw Learn Syst* 2018;29(9):4007–21.
- [15] Ren L, Meng Z, Wang X, Lu R, Yang LT. A wide-deep-sequence model-based quality prediction method in industrial process analysis. *IEEE Trans Neural Netw Learn Syst* 2020;31(9):3721–31.
- [16] Yuan X, Jia Z, Li L, Wang K, Ye L, Wang Y, et al. A SIA-LSTM based virtual metrology for quality variables in irregular sampled time sequence of industrial processes. *Chem Eng Sci* 2022;249:117299.
- [17] Ou C, Zhu H, Shardt YAW, Ye L, Yuan X, Wang Y, et al. Quality-driven regularization for deep learning networks and its application to industrial soft sensors. *IEEE Trans Neural Netw Learn Syst*. In press.
- [18] Yuan X, Ou C, Wang Y, Yang C, Gui W. A layer-wise data augmentation strategy for deep learning networks and its soft sensor application in an industrial hydrocracking process. *IEEE Trans Neural Netw Learn Syst* 2021; 32(8):3296–305.
- [19] Wang X, Hu T, Tang L. A multiobjective evolutionary nonlinear ensemble learning with evolutionary feature selection for silicon prediction in blast furnace. *IEEE Trans Neural Netw Learn Syst* 2022;33(5):2080–93.
- [20] Chai Z, Zhao C, Huang B, Chen H. A deep probabilistic transfer learning framework for soft sensor modeling with missing data. *IEEE Trans Neural Netw Learn Syst* 2022;33(12):7598–609.
- [21] Peng H, Fan Y. Feature selection by optimizing a lower bound of conditional mutual information. *Inf Sci* 2017;418–419:652–67.
- [22] Wang J, Xu C, Zhang J, Zhong R. Big data analytics for intelligent manufacturing systems: a review. *J Manuf Syst* 2022;62:738–52.
- [23] Lee DH, Yang JK, Lee CH, Kim KJ. A data-driven approach to selection of critical process steps in the semiconductor manufacturing process considering missing and imbalanced data. *J Manuf Syst* 2019;52:146–56.
- [24] Sun S, Hu X, Liu Y. An imbalanced data learning method for tool breakage detection based on generative adversarial networks. *J Intell Manuf* 2021;2021: 1–15.
- [25] Peršič N, Dušak V. Conceptual modelling of continuous discrete production systems. In: *Proceedings of the 6th EUROSIM Conference on Modelling and Simulation*; 2007 Sep 9–13; Ljubljana, Slovenia. EUROSIM; 2007. p. 1–7.
- [26] Xu HW, Qin W, Lv YL, Zhang J. Data-driven adaptive virtual metrology for yield prediction in multi-batch wafers. *IEEE Trans Ind Inform* 2022; 18(12): 9008–16.
- [27] Diaz CJL, Ocampo-Martinez C. Energy efficiency in discrete-manufacturing systems: insights, trends, and control strategies. *J Manuf Syst* 2019;52:131–45.
- [28] Thiede S, Turetskyy A, Kwade A, Kara S, Herrmann C. Data mining in battery production chains towards multi-criterial quality prediction. *CIRP Ann* 2019; 68(1):463–6.
- [29] Finkeldey F, Volke J, Zarges JC, Heim HP, Wiederkehr P. Learning quality characteristics for plastic injection molding processes using a combination of simulated and measured data. *J Manuf Process* 2020;60:134–43.
- [30] Keskin Z, Aste T. Information-theoretic measures for nonlinear causality detection: application to social media sentiment and cryptocurrency prices. *R Soc Open Sci* 2020;7(9):200863.
- [31] Spirtes P, Zhang K. Causal discovery and inference: concepts and recent methodological advances. *Appl Inform* 2016;3:3.
- [32] Janzing D, Mooij J, Zhang K, Lemeire J, Zscheischler J, Daniušis P, et al. Information-geometric approach to inferring causal directions. *Artif Intell* 2012; 182–183:1–31.
- [33] Xu L. Machine learning and causal analyses for modeling financial and economic data. *Appl Inform* 2018;5:11.
- [34] Nowack P, Runge J, Eyring V, Haigh JD. Causal networks for climate model evaluation and constrained projections. *Nat Commun* 2020;11:1415.
- [35] Sun Y, Qin W, Zhuang Z, Xu H. An adaptive fault detection and root-cause analysis scheme for complex industrial processes using moving window KPCA and information geometric causal inference. *J Intell Manuf* 2021;32(7):2007–21.
- [36] Sun Y, Qin W, Zhuang Z. Nonparametric-copula-entropy and network deconvolution method for causal discovery in complex manufacturing systems. *J Intell Manuf* 2022;33(6):1699–713.
- [37] Sun Y, Qin W, Zhuang Z. Quality consistency analysis for complex assembly process based on Bayesian networks. *Procedia Manuf* 2020;51:577–83.
- [38] Xu H, Zhang J, Lv Y, Zheng P. Hybrid feature selection for wafer acceptance test parameters in semiconductor manufacturing. *IEEE Access* 2020;8:17320–30.
- [39] Qin W, Zhuang Z, Guo L, Sun Y. A hybrid multi-class imbalanced learning method for predicting the quality level of diesel engines. *J Manuf Syst* 2022;62: 846–56.
- [40] Cai J, Luo J, Wang S, Yang S. Feature selection in machine learning: a new perspective. *Neurocomputing* 2018;300:70–9.
- [41] Han M, Ren W. Global mutual information-based feature selection approach

- using single-objective and multi-objective optimization. *Neurocomputing* 2015; 168:47–54.
- [42] Han Y, Yu L. A variance reduction framework for stable feature selection. *Stat Anal Data Min: ASA Data Sci J* 2012;5(5):428–45.
- [43] Sun YN, Zhuang ZL, Xu HW, Qin W, Feng MJ. Data-driven modeling and analysis based on complex network for multimode recognition of industrial processes. *J Manuf Syst* 2022;62:915–24.
- [44] Reshef DN, Reshef YA, Finucane HK, Grossman SR, McVean G, Turnbaugh PJ, et al. Detecting novel associations in large data sets. *Science* 2011; 334(6062):1518–24.
- [45] Mokhtia M, Eftekhari M, Saberi-Movahed F. Feature selection based on regularization of sparsity based regression models by hesitant fuzzy correlation. *Appl Soft Comput* 2020;91:106255.
- [46] Cai RC, Chen W, Zhang K, Hao ZF. A survey on non-temporal series observational data based causal discovery. *Chin J Comput* 2017;40(6):1470–90. Chinese.
- [47] Glymour C, Zhang K, Spirtes P. Review of causal discovery methods based on graphical models. *Front Genet* 2019;10:524.
- [48] You D, Li R, Liang S, Sun M, Ou X, Yuan F, et al. Online causal feature selection for streaming features. *IEEE Trans Neural Netw Learn Syst*. In press.
- [49] Shimizu S, Hoyer PO, Hyvärinen A, Kerminen A. A linear non-Gaussian acyclic model for causal discovery. *J Mach Learn Res* 2006;7:2003–30.
- [50] Janzing D, Peters J, Mooij J, Schölkopf B. Identifying confounders using additive noise models. 2012. arXiv:1205.2640.
- [51] Zhang K, Hyvärinen A. Nonlinear functional causal models for distinguishing cause from effect. In: Wiedermann W, von Eye A, editors. *Statistics and causality: methods for applied empirical research*. Wiley;2016. p. 185–201.
- [52] Drucker H. Improving regressors using boosting techniques. In: *Proceedings of the 14th International Conference on Machine Learning (ICML)*; 1997 Jul 8–12; Nashville, TN, USA. San Francisco: Morgan Kaufmann Publishers Inc.; 1997. p.107–15.
- [53] Sun YN, Chen Y, Wang WY, Xu HW, Qin W. Modelling and prediction of injection molding process using copula entropy and multi-output SVR. In: *Proceedings of 2021 IEEE 17th International Conference on Automation Science and Engineering (CASE)*; 2021 Aug 23–27; Lyon, France. IEEE; 2021. p. 1677–82.