

基于局部线性嵌入的制造过程多重共线性参数特征选择

胡胜^{1,2*} 高冰冰¹ 张溪¹ 刘登基¹

1.西安工程大学机电工程学院,西安,710048

2.湖北工业大学现代制造质量工程湖北省重点实验室,武汉,430068

摘要:针对制造过程中参数众多易引发多重共线性,致使质量指标预测不准确的问题,提出了一种基于局部线性嵌入(LLE)的制造过程多重共线性参数特征选择方法。首先诊断制造过程参数的多重共线性问题,再用最小绝对收缩和选择算子(LASSO)回归方法将其消除;然后用LLE算法对LASSO回归后的参数做特征选择,获得彼此独立的特征空间,并将其输入到鲸鱼优化支持向量机模型(WOA-SVM)中验证所提算法的参数特征选择效果;最后通过案例分析验证了所提方法的有效性。结果显示,与原始数据相比,采用所提出的方法能够在更低的特征空间维度下获取更精确的预测效果,相关系数值高达0.9702,特征选择的准确率增加了24.989%。

关键词:制造过程;多重共线性;局部线性嵌入;特征选择

中图分类号:TH165.4

DOI:10.3969/j.issn.1004-132X.2025.06.011

开放科学(资源服务)标识码(OSID):



Multicollinearity Parameter Feature Selection for Manufacturing Processes Based on LLEs

HU Sheng^{1,2*} GAO Bingbing¹ ZHANG Xi¹ LIU Dengji¹

1.School of Mechanical Engineering and Electrical Engineering, Xi'an Polytechnic University, Xi'an, 710048

2.Hubei Key Laboratory of Modern Manufacturing Quality Engineering, Hubei University of Technology, Wuhan, 430068

Abstract: In manufacturing processes, a large number of parameters were easily caused to have multicollinearity, which led to problems such as inaccurate prediction of quality indicators. To address these problems, a feature selection method for multicollinear parameters in the manufacturing processes was proposed based on LLE. Firstly, the multicollinearity of the manufacturing process parameters was diagnosed, and then the multicollinearity was eliminated by using the least absolute shrinkage and selection operator(LASSO) regression. Secondly, the LLE algorithm was used to perform feature selection on the parameters after LASSO regression to obtain independent feature spaces, and they were input into the whale optimization algorithm-support vector machine(WOA-SVM) model to verify the parameter feature selection effectiveness of the proposed algorithm. Finally, the effectiveness of the proposed method was verified through case analysis. The results show that compared with the original data, the proposed method may obtain more accurate prediction results under a lower-dimensional feature space, the correlation coefficient value is up to 0.9702, and the accuracy of feature selection increases by 24.989%.

Key words: manufacturing process; multicollinearity; local linear embedding(LLE); feature selection

0 引言

制造过程复杂庞大,工序繁琐,每个工序在生产中会产生海量参数。以弱刚性铝合金零件加工

为例,切削深度、进给量、主轴转速等是关键工艺参数,但这些参数并非完全独立,存在一定的相关性,若对其独立分析或优化,不考虑参数间的相互作用,就会出现多重共线性问题。多重共线性的存在不仅使产品稳定性受到影响,还会导致产品质量的下降和加工效率的降低^[1]。制造过程中参数的多重共线性会使参数指标结果失准。近年来国内外学者对制造过程的多重共线性选择展开了

收稿日期:2024-05-14

基金项目:国家自然科学基金(72001166);教育部人文社科基金(24YJC630073);现代制造质量工程湖北省重点实验室开放基金(KFJJ-2024007);陕西省自然科学基金基础研究计划(2025JC-YBMS-824,2022JQ-721)

一系列研究。KATRUTSA 等^[2]针对特征间线性相关性导致多重共线性问题,介绍了一种基于二次规划的特征选择方法。王珂瑶等^[3]针对属性变量存在严重多重共线性问题,提出了一种修正的马氏距离判别法,以选择预测效果最好的变量维度。JING 等^[4]为了估计具有测量误差和多重共线性的参数,研究了存在多重共线性和测量误差情况下获得参数估计量的新方法。JING 等^[5]针对中介轴承故障信号传输路径复杂的问题,提出了一种基于局部线性嵌入(local linear embedding, LLE)算法的故障特征提取方法,该方法利用 LLE 对多传感器的高维故障特征进行约简和融合,构建故障样本。XIANG 等^[6]针对工业生产过程中采集真实数据的复杂性从而增加基于数据的知识发现的难度,提出了一种集成学习驱动的稳定特征选择方法,以提高特征选择的稳定性和准确性。晶圆良率是衡量半导体产品质量的关键指标,许鸿伟等^[7]针对晶圆良率的数据间关系复杂等特点,对晶圆电性测试参数间多重共线性关系进行主成分分析,获取了预测模型的输入变量。阴艳超等^[8]针对生产过程中多工艺参数时序耦合导致生产质量难以准确预测的问题,提出了基于深度时间卷积神经网络与迁移学习的生产质量快速高效预测方法。为了选择相关参数的正确值以提高加工表面质量,WIBOWO 等^[9]针对二阶多项式回归(second-order polynomial regression, SOPR)的开发没有考虑多重共线性的存在,可能导致不适当预测的问题,使用内核主成分回归方法克服了 SOPR 的弱点。钟杰等^[10]提出了一种基于相关性参数选择与卷积神经网络的异常检测方法,利用最大信息系数和 Pearson 相关系数法挖掘参数之间的相关性,建立相关性非参集合。HE 等^[11]针对具有不确定参数和测量噪声的线性系统,提出了一种新的基于区间观测器的补偿器设计方法。

综上所述,制造过程参数的多重共线性诊断多依据单一数据分布特点判断,客观性不足,解决多重共线性的方法主要是手动移除共线性变量或增加样本量,易产生较大误差。本文提出一种基于 LLE 算法的制造过程参数特征选择框架,集成多种智能算法优势,通过多阶段或并行处理提高特征选择的鲁棒性。先运用容许值、方差膨胀因子与特征根判断法诊断参数的多重共线性,再用最小绝对收缩和选择算子(least absolute selection and shrinkage operator, LASSO)回归模型处理共线性参数,以明晰参数对产品质量的影响

程度;接着采用 LLE 算法对 LASSO 回归后的参数进行特征选择,获取更独立的特征空间;最后将所选参数输入鲸鱼优化支持向量机(whale optimization algorithm-support vector machine, WOA-SVM)模型预测产品质量,并通过案例分析验证该方法的可行性。

1 制造过程多重共线性参数特征选择框架

1.1 多重共线性参数特征选择问题描述

在制造过程中产生的多重共线性参数的特征选择有以下两个问题需要解决:①如何诊断并消除参数间多重共线性;②如何对数据进行特征选择,提高预测模型精度。如图 1 所示。

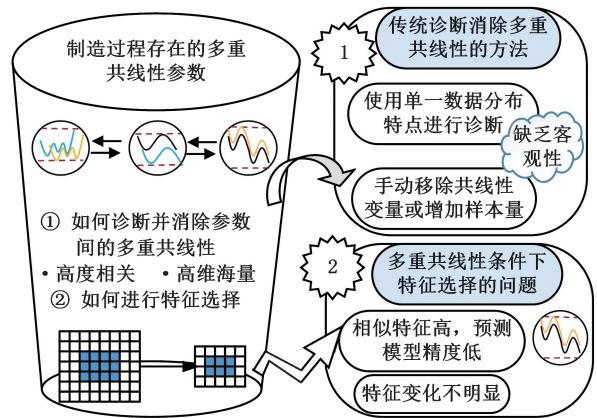


图 1 制造过程多重共线性参数特征选择问题
Fig.1 Multi-collinearity parameter feature selection problem in manufacturing process

1.2 多重共线性参数特征选择框架

制造过程多重共线性参数即制造过程中可以直接影响制造产品质量、性能和稳定性的相关参数,对其进行多重共线性的消除及选择可以获得更优的产品质量。基于此,文中提出的基于局部线性嵌入的制造过程多重共线性参数特征选择框架如图 2 所示。具体内容有以下两个流程:首先收集制造过程产生的海量参数,采用方差膨胀因子法和特征根判断法诊断制造过程涉及的参数是否存在多重共线性,并利用 LASSO 回归对多重共线性参数进行处理,以提高回归模型的稳定性和预测能力;然后采用 LLE 算法对 LASSO 回归后的参数进行特征选择,并将其输入到 WOA-SVM 中对选择后的参数进行预测分析,验证所选方法的特征选择效果。

2 制造过程参数的多重共线性消除

2.1 多重共线性参数诊断

为了诊断出制造过程参数变量间多重共线

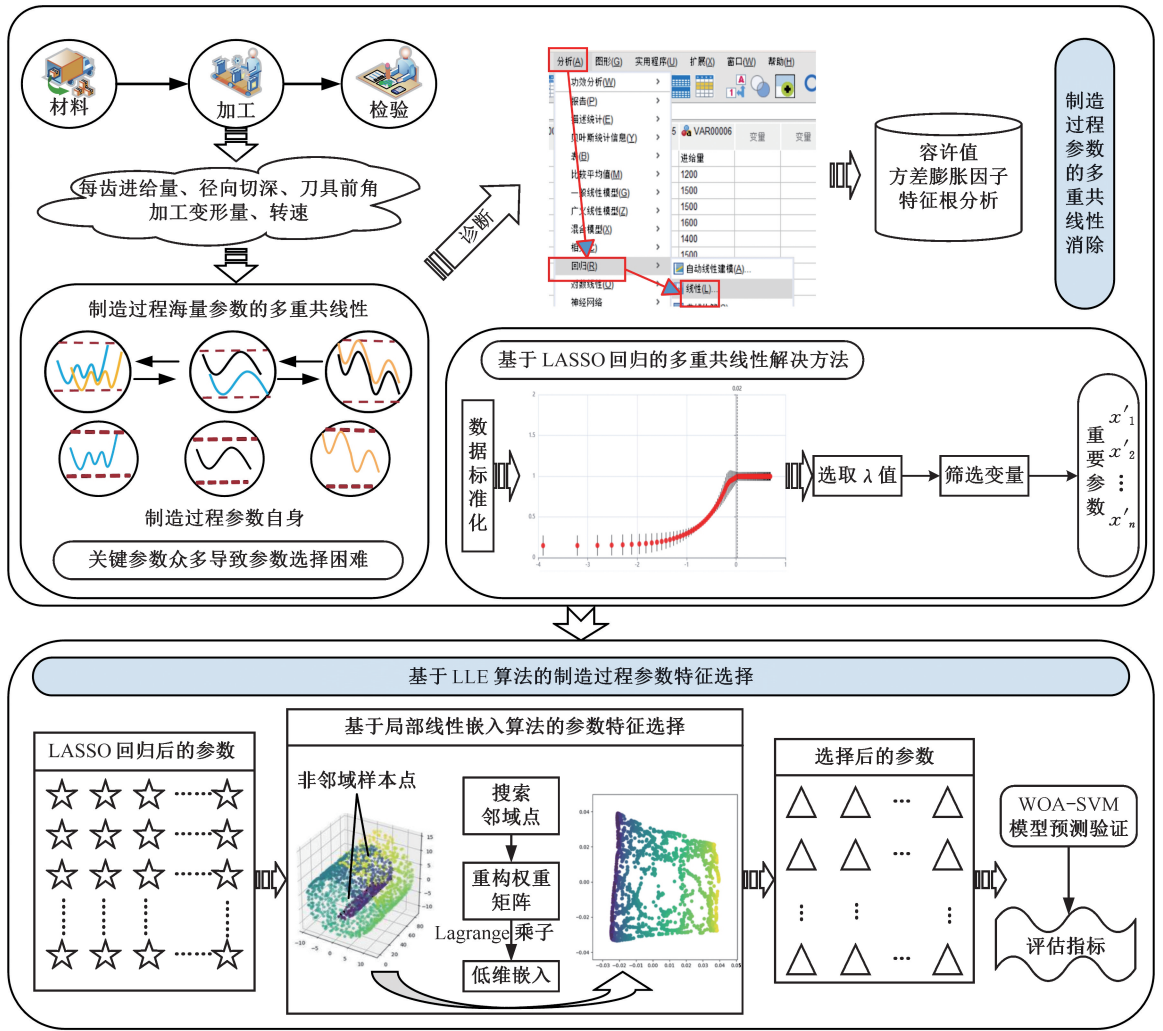


图 2 制造过程的多重共线性参数特征选择框架

Fig.2 Multi-collinearity parameter feature selection framework for the manufacturing process

性,选取容许值、方差膨胀因子以及特征根作为分析诊断指标。容许值是量化自变量之间共线性程度的一个重要指标,通过容许值分析,可以识别出哪些是对共线性影响较大的变量,从而有针对性地进行处理。方差膨胀因子能够量化解变量之间的共线性程度,通过方差膨胀因子分析,可以确定哪些解释变量受共线性的影响较大,进而优化回归模型结构,提高回归模型的预测精度。特征根分析可以识别哪些可能导致多重共线性的变量组合,通过特征根分析,可以揭示变量之间的潜在关联和相互作用,为进一步优化模型提供支持。

2.1.1 容许值 T

容许值 T 的定义为

$$T_{i'} = 1 - R_{i'}^2 \quad i' = 1, 2, \dots, m \quad (1)$$

其中, $R_{i'}$ 为自变量 $\mathbf{X}_{i'}$ 与其他 $m - 1$ 个自变量间的相关系数。如果自变量 $\mathbf{X}_{i'}$ 与其他 $m - 1$ 个自变量间存在严重共线性(即 $R_{i'}^2 \approx 1$),则 $T_{i'} \approx 0$,反之,则 $T_{i'} \approx 1$ 。容许值小于 0.2 可以认为是多重共线性存在的标志。当容许值小于 0.1 时,说明

制造过程参数变量的多重共线性很严重^[12-13]。

2.1.2 方差膨胀因子

方差膨胀因子 V 即为容许值的倒数。方差膨胀因子是指因自变量之间的共线性而导致的回归系数估计方差增加的相对度量。第 i' 个回归系数的方差膨胀因子可表示为

$$V_{i'} = \frac{1}{1 - R_{i'}^2} = \frac{1}{T_{i'}} \quad i' = 1, 2, \dots, m \quad (2)$$

$V_{i'}$ 越大, $R_{i'}^2$ 越接近于 1,说明自变量 $\mathbf{X}_{i'}$ 与其他自变量间的共线性越强。通常情况下, $V \geq 5$ 时说明参数自变量间存在严重的多重共线性^[12-13]。

2.1.3 特征根分析

1) 条件指数。条件指数公式为

$$\eta_k = \sqrt{\lambda_{\max} / \lambda_k} \quad (3)$$

即为 $\mathbf{X}^T \mathbf{X}$ 阵中最大特征值与其他各特征值之比的算术平方根。如果 λ_k 接近于 0,则条件指数 η_k 将很大;条件指数越大,共线性越强。通常情况下,若最大条件指数大于等于 10 且各变量的方差分解比例中的最大值大于 0.5,可认为自变量间存

在共线性。在应用中,如果 $\eta_k \geq 30$,则认为自变量之间存在较强的共线关系^[12]。

2) 方差比例。方差比例是通过分解回归系数方差在不同主成分上的贡献比例,识别哪些参数变量组合可能因共线性导致回归系数不稳定。设自变量的观测值构成的设计矩阵为 \mathbf{X} ,先将 \mathbf{X} 单位化,再对其进行奇异值分解:

$$\mathbf{X}_{n \times m} = \mathbf{U}_{n \times m} \mathbf{D}_{n \times m} \mathbf{V}_{n \times m} \quad (4)$$

式中: $\mathbf{D}_{n \times m}$ 为对角阵,其对角线上的元素为 $\mathbf{X}^T \mathbf{X}$ 的特征根; $\mathbf{V}_{n \times m}$ 为 $\mathbf{X}^T \mathbf{X}$ 的特征向量组成的正交矩阵; $\mathbf{U}_{n \times m}$ 为由 \mathbf{X} 、 \mathbf{D} 、 \mathbf{V}^T 确定的另一个正交矩阵。

回归系数 b_k 的方差为

$$\text{Var}(b_k) = \sigma^2 \sum_{i'=1}^m v_{i'k}^2 / u_{i'k}^2 \quad i', k = 1, 2, \dots, m \quad (5)$$

式中: $v_{i'k}$ 为 $\mathbf{V}_{n \times m}$ 中第 (k, i') 个元素; $u_{i'k}$ 为 $\mathbf{D}_{n \times m}$ 中对角线上第 (i', k) 个元素; σ^2 为随机误差项的方差。

设 $\varphi_{i'k} = v_{i'k}^2 / u_{i'k}^2$, $\varphi_k = \sum_{i'=1}^m \varphi_{i'k}$, 则第 (i', k) 个方差比例定义为

$$\Pi_{i'k} = \varphi_{i'k} / \varphi_k \quad i', k = 1, 2, \dots, m \quad (6)$$

理论上一般认为,在参数自变量标准化情况下,采用回归系数方差比例方法诊断制造过程参数变量多重共线性的诊断原则为:① 相关的条件指标较高,一般指大于 10;② 2 个或多个估计回归系数有较大的方差比例,一般大于 0.5^[12]。

上述方法表明,容许值、方差膨胀因子与特征根分析这三种多重共线性的诊断方法能够有效地诊断出参数间的多重共线性,发现参数变量之间的线性相关关系。因此,需要在后文采取相应的解决方法以避免模型估计失真、减少计算误差。

2.2 基于 LASSO 回归的参数多重共线性消除

本文选取 LASSO 回归方法解决参数多重共线性问题。LASSO 回归是一种替代最小二乘法的压缩估计方法,即为最小绝对值选择与选择算子回归方法^[14]。在 LASSO 回归中,模型的目标函数被修改为最小化残差平方和加上一个与模型参数绝对值相关的正则化项,这个正则化项通常是一个常数乘以所有模型参数绝对值的总和。假设制造过程参数 $X_i = \{x_{i1}, x_{i2}, \dots, x_{im_*}\}$, 其中, m_* 表示制造过程参数的数量, Y_i 为第 i 个过程参数对应的因变量,考虑线性回归模型:

$$\left. \begin{aligned} Y_i &= \alpha_i + \sum_{j=1}^{m_*} \beta_j x_{ij} + \epsilon_i \\ \epsilon_i &\sim N(0, \sigma^2) \end{aligned} \right\} \quad (7)$$

其中, α_i 为制造过程参数的因变量, β_j 为制造过程参数的自变量。在一般的回归结构中,通常修改目标函数来引入正则化项,假设过程参数彼此

独立,或 Y_i 在过程参数给定的情况下独立,即 Y_i 关于 X_i 条件独立,同时假设样本参数 x_{ij} 是标准化的,也就是 $\frac{1}{n} \sum_j x_{ij} = 0$, $\frac{1}{n} \sum_j x_{ij}^2 = 1$, 则 LASSO 回归的目标函数为

$$\left. \begin{aligned} (\hat{\alpha}, \hat{\beta}) &= \arg \min_{\beta} \left\{ \sum_i (Y_i - \alpha_i - \sum_j \beta_j x_{ij})^2 \right\} \\ \text{s.t.} \quad &\sum_j |\beta_j| \leq t \end{aligned} \right\} \quad (8)$$

式(8)中, t 为一个正的调和参数,用于控制正则化的强度。此时对于一切 t ,有 α 的估计 $\hat{\alpha} = \bar{\beta}$ 。其中, $\bar{\beta}$ 为过程参数自变量的平均值。当 t 增大时,正则化项的权重增加,使得回归系数总体变小,若令 $t^{(0)} = \sum_j |\beta_j|$, $t \leq t^{(0)}$, 就会使一些回归系数缩小并趋于 0,一些系数甚至会达到 0,从而简化模型并减少过拟合。

3 基于 LLE 的制造过程参数特征选择

采用 2.2 节方法虽然消除了制造过程中参数的多重共线性,筛选了与产品质量指标显著性小的冗余过程参数,但还有参数彼此之间存在相关性,因此,需要对 LASSO 回归后参数的特征空间重新构建,以确保新空间中的特征彼此更加独立,增强预测模型的泛化能力。基于 LLE 算法的制造过程参数特征选择如图 3 所示。

3.1 制造过程参数特征选择

对于制造过程参数这种非线性高维数据, LLE 算法可以保持邻域内样本之间的线性关系,使得样本之间的映射坐标能够在低维空间中得以保持^[15], 因此,选取 LLE 算法对 LASSO 回归后的参数进行特征选择。

LLE 算法是给定一个过程参数样本集 X^* , $X^* = \{x_1, x_2, \dots, x_N\} \in \mathbf{R}_{D^* \times N}$, 其中, x_i 为是过程参数样本集的第 i ($i = 1, 2, \dots, N$) 个过程参数,过程参数样本集的特征维度为 D^* , $x_i = \{x_{i1}, x_{i2}, \dots, x_{iD^*}\} \in \mathbf{R}_{D^* \times 1}$, 算法的目标是从中挖掘出映射到低维空间的特征,其结果为 Y 。该目标的步骤如下:

1) 邻域的选取。对过程参数样本集 X^* 中的每个过程参数样本点 x_i 计算它与其他过程参数样本点之间的相似性,选择与之最相似的前一个点作为其近邻点。

2) 计算重构权重矩阵 \mathbf{W} 。找到给定的过程参数样本点 x_i 的邻域 x_j ($j = 1, 2, \dots, k$), 求它在邻域里与 k 近邻点的线性关系,并通过最小化重构误差函数:

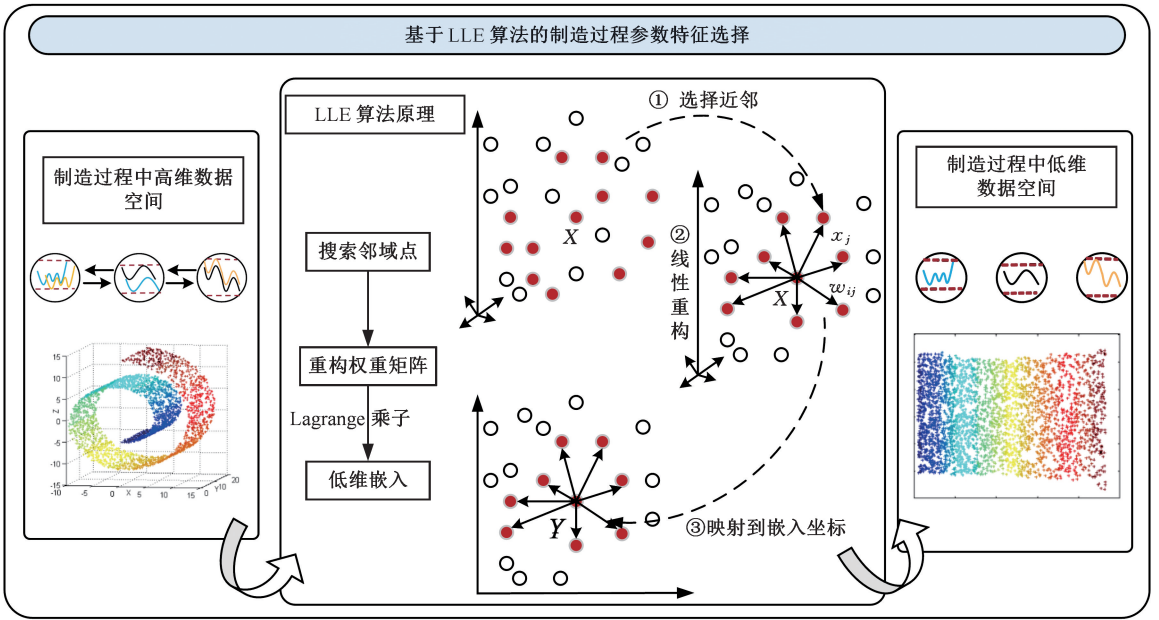


图 3 基于 LLE 的制造过程参数特征选择

Fig.3 Characteristic selection of manufacturing process parameters based on LLE

$$\left. \begin{aligned}
 & \arg \min_{\mathbf{W}} \left\| x_i - \sum_{j=1}^N w_{ij} x_j \right\|_2^2 = \\
 & \arg \min_{\mathbf{W}} \left\| \sum_{j=1}^N w_{ij} x_i - \sum_{j=1}^N w_{ij} x_j \right\|_2^2 = \\
 & \arg \min_{\mathbf{W}} \mathbf{W}_i^T (x_i - x_j) (x_i - x_j)^T \mathbf{W}_i; \\
 & \text{s.t. } \sum_{j=1}^N w_{ij} = 1
 \end{aligned} \right\} \quad (9)$$

其中,过程参数样本 x_j 不是过程参数样本点 x_i 邻域内的近邻点时,对应的 w_{ij} 设为 0。 \mathbf{W} 是一个权重矩阵,其中每个元素都是 w_{ij} ,约束项用来满足式(8)的解为稀疏解,即部分系数趋近于零,用于 LLE 的特征选择。

3) 获得低维嵌入 \mathbf{Y} 。采用基于局部线性结构的方法,通过最小化目标函数来进行优化:

$$\left. \begin{aligned}
 & \arg \min \sum_{i=1}^N \left\| y_i - \sum_{j=1}^N w_{ij} x_j \right\|_2^2 = \sum_{i=1}^N \left\| \mathbf{Y}_i - \mathbf{Y} \mathbf{W}_i \right\|_2^2 = \\
 & \text{tr}(\mathbf{Y}(\mathbf{I} - \mathbf{W})(\mathbf{I} - \mathbf{W})^T \mathbf{Y}^T) = \text{tr}(\mathbf{Y} \mathbf{M} \mathbf{Y}^T) \\
 & \text{s.t. } \sum_{j=1}^N y_j = 0 \quad \frac{1}{N} \sum_{i=1}^N y_i y_i^T = \mathbf{I}
 \end{aligned} \right\} \quad (10)$$

其中, $\mathbf{I}_i, \mathbf{W}_i$ 分别为 \mathbf{I}, \mathbf{W} 的第 i 个列向量, $\mathbf{M} = (\mathbf{I} - \mathbf{W})(\mathbf{I} - \mathbf{W})^T$ 。引入拉格朗日乘子将约束融入目标函数,利用梯度共线条件寻找极值候选点可得到 \mathbf{Y} ,即 \mathbf{M} 的 k 个最小非零特征值所对应的特征向量, $\mathbf{Y} = (y_1, y_2, \dots, y_N)$ 。

通过 LLE 算法选择更加独立的特征空间之后,进行后续的验证分析。

3.2 参数特征选择效果评估指标

为了有效评估 LLE 算法的参数特征选择效

果,文中采用 WOA-SVM 预测模型对选择后的特征空间进行验证。所选评估指标包括均方误差 (MSE)、平均绝对误差 (MAE)、均方根误差 (RMSE)、平均绝对百分比误差 (MAPE) 及相关系数。其中, MSE, MAE, RMSE, MAPE 值 ($\sigma_{\text{MSE}}, \sigma_{\text{MAE}}, \sigma_{\text{RMSE}}, \sigma_{\text{MAPE}}$) 越小,表示预测的误差越小,相关系数 R 越大,表示预测效果越好。评估指标的公式如表 1 所示。

表 1 评估指标公式表

Tab.1 Formula table of evaluation indicators

评估指标	评估指标公式
均方误差	$\sigma_{\text{MSE}} = \frac{1}{l} \sum_{q=1}^l (y_q - \hat{y}_q)^2$
平均绝对误差	$\sigma_{\text{MAE}} = \frac{1}{l} \sum_{q=1}^l y_q - \hat{y}_q $
均方根误差	$\sigma_{\text{RMSE}} = \sqrt{\frac{1}{l} \sum_{q=1}^l (y_q - \hat{y}_q)^2}$
平均绝对百分比误差	$\sigma_{\text{MAPE}} = \frac{1}{l} \sum_{q=1}^l \left \frac{y_q - \hat{y}_q}{y_q} \right $
相关系数	$R = \frac{\sum_{q=1}^l (y_q - \bar{y}) \sum_{q=1}^l (\hat{y}_q - \frac{1}{l} \sum_{q=1}^l \hat{y}_q)}{\sqrt{\sum_{q=1}^l (y_q - \bar{y})^2 \sum_{q=1}^l (\hat{y}_q - \frac{1}{l} \sum_{q=1}^l \hat{y}_q)^2}}$

注: y_q 为测试集中第 q 个样本的真实值; \hat{y}_q 为测试集中第 q 个样本数据的预测值; \bar{y} 为 l 个测试样本质量指标的均值, $q \in [1, l], l$ 为测试集的样本数。

4 案例分析

为了验证所提出的基于 LLE 算法的制造过程多重共线性参数特征选择算法的有效性,选取

弱刚性铝合金零件加工过程的铣削参数作为实验数据的来源^[16]。该过程中若铣削加工参数(如主轴转速、每齿进给量等)特征选择不当,会使样本数据的特征数量多,预测模型过于复杂,容易出现过拟合现象,因此,需要进行合理的特征选择,降低数据的特征维度,进而提高质量指标的预测准确率。在弱刚性零件选定的前提下,铣削加工变形受铣削参数的影响,例如转速 $n(r/min)$ 、每齿进给量 $f_z(mm)$ 、径向切深 $a_p(mm)$ 、刀具前角 $\gamma_0(^{\circ})$ 。在铣削参数的特征选择过程中,需要对自变量进行多重共线性诊断,识别高度线性相关的冗余变量,并结合 LASSO 回归消除共线性的影响,从而进行参数的特征选择,以获得最佳的预测效果,提高预测模型的稳定性和加工效率。

4.1 铣削参数的多重共线性诊断

为对铣削参数与铣削加工变形的参数选择模型进行有效训练,通过文献[16]将转速 n 、每齿进给量 f_z 、径向切深 a_p 、刀具前角 γ_0 组成铣削参数,以及铣削加工变形量 y 组成100组训练样本,如表2所示。

表 2 样本数据

Tab.2 Sample data

样本	转速 $n / (r \cdot \min^{-1})$	每齿进给量 f_z / mm	径向切深 a_p / mm	刀具前角 $\gamma_0 / (^{\circ})$	加工变形量 y / mm
1	6100	0.133	0.5	10	0.3604
2	5600	0.143	0.46	15	0.3654
3	6900	0.128	0.54	10	0.2868
4	4100	0.101	0.41	5	0.208
⋮	⋮	⋮	⋮	⋮	⋮
49	5500	0.136	0.46	15	0.2782
50	7200	0.131	0.57	10	0.2544
51	4800	0.115	0.44	5	0.2434
⋮	⋮	⋮	⋮	⋮	⋮
98	6500	0.111	0.51	5	0.2866
99	4900	0.115	0.44	5	0.2444
100	7400	0.127	0.48	10	0.309

对样本数据通过 SPSS 进行数据分析,以及多重共线性诊断,结果如表3所示。

表 3 共线性诊断表

Tab.3 The collinearity diagnostic table

自变量	共线性统计		特征值	条件指标	方差比例最大值
	容许值	VIF			
(常量)	0	0	0.093	1.000	0
转速	0.139	7.209	0.535	7.061	0.66
每齿进给量	0.084	11.905	0.573	16.952	0.89
径向切深	0.145	6.886	-0.241	58.737	0.92
刀具前角	0.101	9.993	-0.079	78.593	0.79

由表3可以看出:样本变量的方差膨胀因子(VIF)值超过5,其容许值均小于0.2,每齿进给量变量的VIF值甚至为11.905,其容许值约为

0.084,说明参数变量之间存在严重的多重共线性,它们的相关性极强,在回归模型中不显著,这样会导致回归系数估计值偏离真实值,标准误差增大,使得显著性检验失败。转速、每齿进给量、径向切深和刀具前角的特征值约等于0,说明存在比较严重的共线性;每齿进给量、径向切深和刀具前角的条件指标大于10,说明存在比较严重的共线性;转速、每齿进给量、径向切深、刀具前角的方差比例五个主成分中方差比例最大值均超过0.5,说明存在比较严重的共线性。

样本数据存在多重共线性可能会对回归模型的稳定性产生负面影响。当存在多重共线性时,自变量的参数估计值可能会变得不稳定,对样本的微小变化非常敏感,这意味着在不同的样本或模型设定下,参数的估计值可能会有很大的变化,同时也会导致参数估计值的方差增大,从而使得变量的显著性检验失效,因此,后续将解决样本数据的多重共线性问题。

4.2 铣削参数的多重共线性处理

考虑到铣削参数对铣削加工变形的影响,文中采取 LASSO 回归方法解决样本数据的多重共线性问题。首先将样本数据进行标准化处理,然后利用 SPSSPRO 软件对样本数据进行 LASSO 回归分析,LASSO 回归交叉验证图见图4。

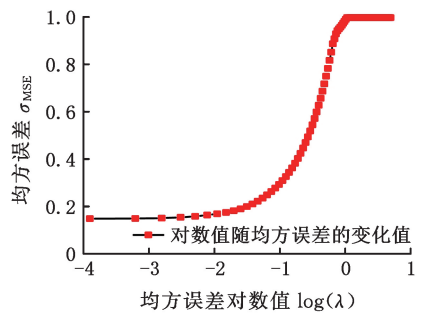


图 4 LASSO 回归交叉验证图

Fig.4 LASSO regression cross-validation figure

图4以可视化的形式展示了使用交叉验证选择 λ 值的情况,其中 λ 为正则化参数,选择的标准是使得模型的均方误差最小。因此,为使得均方误差最小,确定 $\lambda = 0.02, \log(\lambda) = -3.902$ 。

通过模型系数表确定筛选变量的情况,结果如表4所示。

当模型中标准化系数变量为0时,这个变量会被排除出模型,其中截距表示加工变形量的预测值。表4结果显示:截距、转速、每齿进给量、径向切深、刀具前角的标准化系数都不为0,因此没有变量被剔除。标准化系数提供了自变量对因变量影响力的相对度量,减弱了单位的影响,而非标准化系数则直接反映了自变量和因变量之间的原

表 4 模型系数表

Tab.4 Model coefficient table

变量名	标准化系数	非标准化系数	R^2
截距	0.051	0.094	0.598
转速	0.555	0.273	
每齿进给量	0.629	2.332	
径向切深	-0.249	-0.240	
刀具前角	-0.114	-0.002	

始关系,因此,LASSO 回归减弱了径向切深和刀具前角对加工变形量的影响,转速和每齿进给量对加工变形量有显著作用。

4.3 基于 LLE 的制造过程铣削参数特征选择

对 4.2 节 LASSO 回归后的铣削参数进行特征选择,采用 LLE 算法选择低维空间中正交性强的方向作为新特征,生成更独立的低维特征,将高维铣削参数映射到 2 维空间,直观可视化参数间的潜在关系。图 5 为 LLE 算法降维后的散点图,由图 5a 可以明显看出数据分布呈负相关,与回归后的数据分布相似,说明在降维过程中丢失的信息较少,降维效果良好;由图 5b 所示数据分布看不出数据呈何种分布状态,因此,基于 3.1 节设定近邻数 $K=10$,选择后的铣削参数特征空间为 2 维空间,如表 5 所示。

4.4 结果分析

为了说明选取 LLE 算法进行参数特征选择的有效性,将其与 LASSO 回归方法进行对比。文中选取 WOA-SVM 模型通过预测回归分析对参数选择效果进行验证,在 SVM 中,RBF 核函数

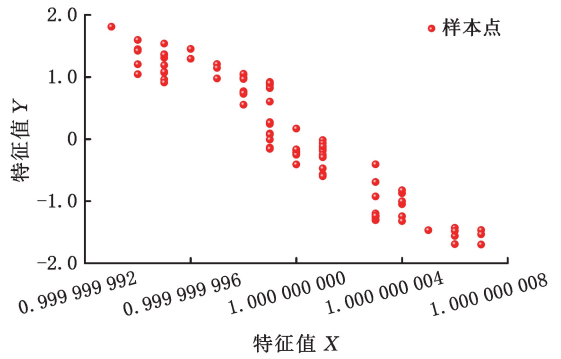
表 5 由 LLE 算法选择的低维特征空间

Tab.5 Low-dimensional feature space selected by the LLE algorithm

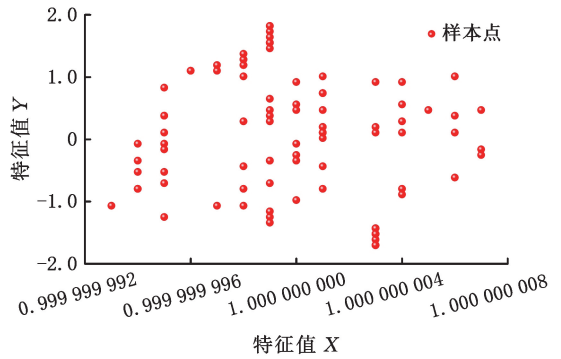
空间维数	特征数量										
	1	2	3	...	49	50	51	...	98	99	100
1	0.9999	0.9200	1.0000	...	0.9999	1.0000	0.9999	...	0.9999	1.0000	0.9999
2	0.9200	-0.1785	-1.5597	...	0.0738	-1.0500	1.4517	...	-0.0090	0.1695	1.1470

将经过 LASSO 回归筛选的关键参数与经 LLE 算法选择获取的特征空间分别输入到采用 RBF 核函数的 WOA-SVM 模型中进行预测验证,从 100 组数据中随机选取 70 组作为训练集,其余 30 组作为测试集,初始种群为 30,迭代次数为 100,WOA 搜寻的最优惩罚因子和 γ 范围设为 0.1~100 和 0.1~50,优化变量的下限和上限分别为 0.01 和 100,可得 WOA 寻优后的最佳惩罚因子值为 36.7769,最优 γ 值为 4.5944。WOA 算法适应度与迭代次数的曲线图见图 6。

由图 6 可知,迭代 100 次可满足 WOA 算法的收敛。样本数据在适应度为 0.0309 开始收敛,样本数据+LASSO 回归在适应度为 0.0155 开始收敛,LASSO 回归+LLE 在适应度为 0.0064 开



(a)LLE 的 2 维结果图



(b)LLE 的 3 维结果图

图 5 LLE 算法的降维结果图

Fig.5 Dimensionality reduction result plot for the LLE algorithm

的主要参数是惩罚因子 C 和 γ ,其中惩罚因子 C 在 SVM 中起到权衡的作用,用于控制模型对误差的宽容度, γ 决定了数据映射到新的特征空间后的分布。

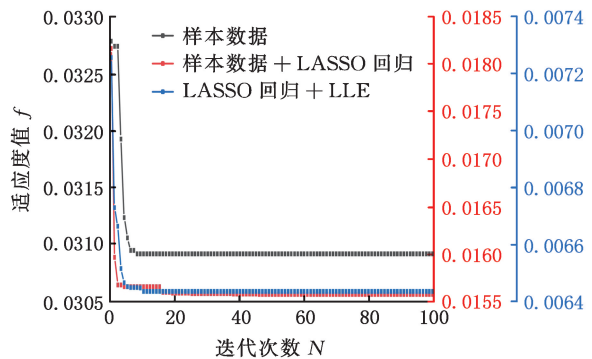


图 6 WOA 的适应度曲线

Fig.6 Fitness curves of the WOA

始收敛,由 MSE,MAE, RMSE, MAPE 评估指标越小预测效果越好可知,在对样本数据进行 LLE 算法处理后,WOA 算法收敛效果更好。

取 2 次预测结果的平均值进行对比分析,并以原始样本数据的铣削参数空间预测效果作为对照。每种方法输入到预测模型的最优核函数参数与结果对比如图 7 和表 6 所示。

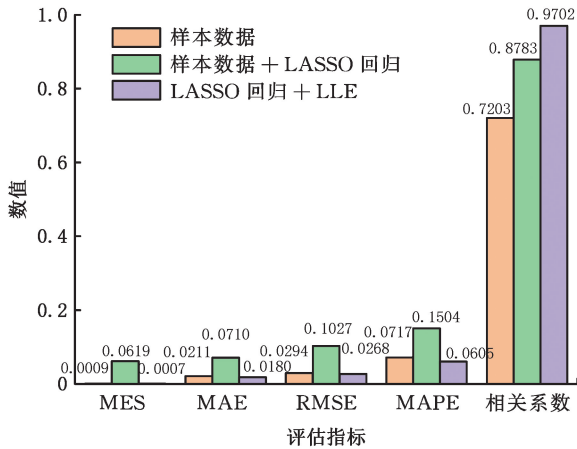


图 7 WOA-SVM 模型预测结果对比图

Fig.7 Comparison plot of the prediction results of the WOA-SVM model

表 6 不同参数特征选择方法的预测效果对比

Tab.6 Comparison of prediction performance of feature selection methods with different parameters

所选方法	数据维度	模型参数		运算时间/s	评估指标				
		C	γ		MSE	MAE	RMSE	MAPE	相关系数
样本数据	100×5	71.6520	3.6451	0.11	0.000 86	0.021 11	0.029 44	0.071 69	0.720 32
样本数据 + LASSO 回归	100×5	33.6299	4.5944	0.10	0.061 86	0.071 03	0.102 65	0.150 37	0.878 31
LASSO 回归 + LLE	100×2	36.7769	3.9265	0.08	0.000 72	0.018 03	0.026 82	0.060 50	0.970 21

5 结语

针对制造过程涉及的参数因高度相关而产生多重共线性问题所导致的参数选择结果不稳定和预测结果误差较大等问题,本文进行了过程参数多重共线性的诊断及消除,并建立了制造过程参数特征选择模型,得出以下结论:

1)通过 SPSS 分析对制造过程中的过程参数进行多重共线性诊断,综合容许值、方差膨胀因子和特征根分析三种诊断指标判断参数变量间的多重共线性,并通过 LASSO 回归解决过程参数间的多重共线性,筛选出对产品质量影响较大的过程参数,降低参数特征选择的不稳定性。

2)制造过程参数之间相互关联,通过 LLE 算法对 LASSO 回归后的过程参数进行特征选择,运算时间相比于原样本数据缩短了 0.03 s,特征选择的准确率增加了 24.989%,减小了因过程参数间相互关联而导致的预测结果误差较大等问题,从而提高了预测效果。

由图 7 和表 6 可知,对铣削参数进行参数特征选择的预测效果是优于跳过铣削参数筛选直接对过程参数进行参数特征选择的预测效果的,即在对具有多重共线性的制造过程参数数据进行参数特征选择前,进行 LASSO 回归的筛选是有必要的。对于 LASSO 回归筛选后的关键参数,采用 LLE 算法参数特征选择后的预测效果中的各项评价指标优于 LASSO 回归和原始样本数据,即 σ_{MSE} 减小了约 16.28%, σ_{MAE} 减小了约 14.59%, σ_{RMSE} 减小了约 9.24%, σ_{MAPE} 减小了约 15.61%,均小于原始样本数据的指标数值;相关系数高于原始样本数据的指标数值,高达 0.9702,即特征选择的准确率增加了 24.989%;与其相对应的运算时间为 0.08 s,相比于原样本数据缩短了 0.03 s。综上所述,文中提出的采用 LLE 算法对弱刚性零件进行铣削参数特征选择,可以在更短的时间内实现以更低的特征空间维度获取更优的参数特征选择效果。

参考文献:

[1] SIDDIQUEE M N, SIVARAMAKRISHNAN K, WU Y, et al. A Statistical Approach Dealing with Multicollinearity among Predictors in Microfluidic Reactor Operation to Control Liquidphase Oxidation Selectivity[J]. Reaction Chemistry & Engineering, 2018, 3(6):972-990.

[2] KATRUTSA A, STRIJOV V. Comprehensive Study of Feature Selection Methods to Solve Multicollinearity Problem According to Evaluation Criteria[J]. Expert Systems with Applications, 2017, 76 1-11.

[3] 王珂瑶, 王惠文, 赵青, 等. 一种修正的马氏距离判别法[J]. 北京航空航天大学学报, 2022, 48(5): 824-830.

WANG Keyao, WANG Huiwen, ZHAO Qing, et al. A Modified Mahalanobis Distance Discriminant Method[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022,48(5):824-830.

[4] JING G, YUNFENG Z. Parameter Estimation Approaches to Tackling Measurement Error and Multicollinearity in Ordinal Probit Models[J]. Communi-

- cations in Statistics Theory and Methods, 2020, 49 (16):3835-3859.
- [5] JING T, YUWEI Z, FENGLING Z, et al. A Novel Intelligent Method for Intershaft Bearingfault Diagnosis Based on Hierarchical Permutation Entropy and LLE-RF[J]. Journal of Vibration and Control, 2023, 29(23/24):5357-5372.
- [6] XIANG F, ZHAO Y, ZHANG M, et al. Ensemble Learning Based Stability Improvement Method for Feature Selection towards Performance Prediction [J]. Journal of Manufacturing Systems, 2024, 74 55-67.
- [7] 许鸿伟,张洁,吕佑龙,等. 基于改进的连续型深度信念网络的晶圆良率预测方法[J]. 计算机集成制造系统, 2020, 26(9):2388-2395.
XU Hongwei, ZHANG Jie, LYU Youlong, et al. Wafer Yield Prediction Method Based on Improved Continuoustype Deep Belief Networks[J]. Computer Integrated Manufacturing System, 2020, 26(9): 2388-2395.
- [8] 阴艳超,施成娟,邹朝普,等. 基于深度时间卷积神经网络与迁移学习的流程制造工艺过程质量时序关联预测[J]. 中国机械工程, 2023, 34(14):1659-1671.
YIN Yanchao, SHI Chengjuan, ZOU Chaopu, et al. Process Manufacturing Process Quality Timing Correlation Prediction Based on Deep Temporal Convolutional Neural Network and Transfer Learning[J]. China Mechanical Engineering, 2023, 34 (14):1659-1671.
- [9] WIBOWO A. Hybrid Kernel Principal Component Regression and Penalty Strategy of Multiple Adaptive Genetic Algorithms for Estimating Optimum Parameters in Abrasive Waterjet Machining[J]. Applied Soft Computing, 2018, 62 1102-1112.
- [10] 钟杰,罗冲,张恒,等. 基于相关性参数选择的飞行数据异常检测[J]. 北京航空航天大学学报, 2024, 50(5):1738-1745.
ZHONG Jie, LUO Chong, ZHANG Heng, et al. Abnormal Detection in Flight Data Based on Correlation Parameter Selection[J]. Journal of Beijing University of Aeronautics and Astronautics, 2024, 50(5):1738-1745.
- [11] HE Zhongwei. Interval Observer-based Compensators Design for Linear Systems with Uncertain Time-varying Parameters[J]. Transactions of the Institute of Measurement and Control, 2023, 45 (16):3247-3253.
- [12] 周菲. Logistic 回归多重共线性的诊断与改进及其在医学中的应用[D]. 兰州:兰州大学, 2011.
ZHOU Fei. Logistic Diagnosis and Improvement of Regression Multicollinearity and Its Application in Medicine[D]. Lanzhou:Lanzhou University, 2011.
- [13] WIBOWO A, DESA I M. Kernel Based Regression and Genetic Algorithms for Estimating Cutting Conditions of Surface Roughness in End Milling Machining Process[J]. Expert Systems with Applications, 2012, 39(14):11634-11641.
- [14] 刘明,杜建强,李郅琴,等. 融合 LASSO 的近似马尔科夫毯特征选择方法[J]. 计算机工程与应用, 2024, 60(8):121-130.
LIU Ming, DU Jianqiang, LI Zhiqin, et al. Fuses LASSO's Approximate Markov Blanket Feature Selection Method[J]. Computer Engineering and Applications, 2024, 60(8):121-130.
- [15] 张璐瑶,季伟东,程昊. 基于 LLE 降维思想的自然计算方法[J]. 系统仿真学报, 2020, 32(10):1943-1955.
ZHANG Luyao, JI Weidong, CHENG Hao. A Natural Computational Method Based on the Idea of LLE Dimension Reduction[J]. Journal of System Simulation, 2020,32(10):1943-1955.
- [16] 王全. 弱刚性铝合金零件加工参数优化研究[D]. 兰州:兰州理工大学, 2022.
WANG Quan. Study on Machining Parameter Optimization of Weak Rigid Aluminum Alloy Parts [D]. Lanzhou:Lanzhou University of Technology, 2022.

(编辑 王艳丽)

作者简介:胡胜*,男,1988年生,博士、副教授。研究方向为智能制造质量控制、质量管理与质量工程。Email: husheng@xpu.edu.cn。

本文引用格式:

胡胜,高冰冰,张溪,等.基于局部线性嵌入的制造过程多重共线性参数特征选择[J]. 中国机械工程,2025,36(6):1238-1246.

HU Sheng, GAO Bingbing, ZHANG Xi, et al. Multicollinearity Parameter Feature Selection for Manufacturing Processes Based on LLEs[J]. China Mechanical Engineering, 2025, 36(6):1238-1246.