

# 基于组合模型的燃气轮机 NO<sub>x</sub> 排放影响因素研究<sup>①</sup>



石翠翠, 刘媛华

(上海理工大学 管理学院, 上海 200093)

通信作者: 刘媛华, E-mail: liuyuanhua@usst.edu.cn

**摘要:** 为了减少电厂污染物的排放, 如何准确估计燃气轮机氮氧化物 (NO<sub>x</sub>) 排放值并识别其关键影响因素, 对有效采取优化设计是至关重要的. 由于燃气轮机的运作过程存在变工况等情况, 单一模型的准确度与泛化性能难以达到工业应用的要求. 将偏最小二乘法 (PLS) 和互信息 (MI) 组合建模保证了 NO<sub>x</sub> 特征变量选取的有效性与精确性. 利用 PLS 确定影响燃气轮机 NO<sub>x</sub> 的特征变量数目, 避免了选择变量的主观因素并降低维数. 再用互信息 (MI) 选择出最优的特征变量, 通过不同的预测模型进行仿真分析, 并把单一和组合特征选择进行对比. 结果表明, 对燃气轮机 NO<sub>x</sub> 排放影响因素的研究中, PLS-MI 组合模型选取的特征变量更具代表性, 并能够保证预测模型的泛化精度, 降低模型复杂度, 为电厂优化控制提供了理论依据, 具有一定的应用前景.

**关键词:** 偏最小二乘法; 互信息; 氮氧化物; 特征提取

引用格式: 石翠翠, 刘媛华. 基于组合模型的燃气轮机 NO<sub>x</sub> 排放影响因素研究. 计算机系统应用, 2022, 31(6): 354-360. <http://www.c-s-a.org.cn/1003-3254/8527.html>

## Research on Influencing Factors of NO<sub>x</sub> Emission from Gas Turbine Based on Combined Model

SHI Cui-Cui, LIU Yuan-Hua

(Business School, University of Shanghai for Science and Technology, Shanghai 200093, China)

**Abstract:** For the reduction of pollutant emissions from power plants, the ways to accurately estimate the emission value of nitrogen oxide (NO<sub>x</sub>) from gas turbines and identify its key influencing factors are crucial for effective optimization design. The accuracy and generalization performance of a single model can hardly meet the requirements of industrial applications because of variable operating conditions in the gas turbine operation process. The combined modeling of the partial least square (PLS) method and mutual information (MI) ensures the effectiveness and accuracy of the feature variable selection of NO<sub>x</sub>. Specifically, PLS is employed to determine the number of feature variables affecting the NO<sub>x</sub> of gas turbines, which can avoid subjective factors of variable selection and reduce dimensions. Then, mutual information (MI) is applied to select optimal feature variables. Different prediction models are used for simulation analysis, and single and combined feature selection is compared. The results show that in this study, the combined model PLS-MI can select more representative feature variables and can ensure the generalization accuracy of the prediction model, reducing the model complexity and providing a theoretical basis for optimal control of power plants with certain application prospects.

**Key words:** partial least square method; mutual information; nitrogen oxide; feature extraction

燃气轮机的运行是一个多变量影响的复杂系统, 运行产生的烟气包含大量污染物, 是造成环境污染主

要的来源之一<sup>[1]</sup>. 氮氧化物 (NO<sub>x</sub>) 是烟气污染物中最常见的一种, 容易引发化学烟雾、温室效应等环境问题.

① 基金项目: 上海市软科学重点研究项目 (19692104000)

收稿时间: 2021-09-08; 修改时间: 2021-10-14; 采用时间: 2021-10-15; csa 在线出版时间: 2022-05-26

选择性催化还原 (SCR) 是燃气轮机运行中的一项重要技术, 由于 SCR 脱硝系统入口 NO<sub>x</sub> 浓度测量具有一定的滞后特性, 而且受不同因素的影响, 燃气轮机排放的氮氧化物浓度会有较大的波动<sup>[2]</sup>. 如果能够提前准确预测 NO<sub>x</sub> 排放的入口浓度及确定关键特征变量, 能够为建立精确的预测模型, 实现燃烧控制工艺参数优化效率打下基础.

基于物理分析的影响 NO<sub>x</sub> 排放的关键运行参数, 往往是针对某一具体型号或工况, 适用范围差. 近年来, 基于数据驱动建模的方法因其具有良好的预测稳定性和非线性处理能力, 在电厂中得到了广泛应用. 吕游等<sup>[3]</sup> 首先利用偏最小二乘法 (PLS) 对变量进行降维以及消除彼此间的相关性, 然后建立 NO<sub>x</sub> 排放的最小二乘支持向量机 (LS-SVM) 模型, 经过与其他模型对比分析, 验证了 PLS 特征提取能够降低模型的复杂性. Li 等<sup>[4]</sup> 建立基于改进的粒子群算法优化支持向量机 (SVM) 的 NO<sub>x</sub> 预测模型, 以此降低预测误差和提高预测结果的稳定性. 赵刚等<sup>[5]</sup> 在保证燃气轮机运行稳定的情况下利用 BP 神经网络模型对影响燃气轮机 NO<sub>x</sub> 排放的变量因素进行敏感性分析, 确定它们的影响大小及正负相关性. 伦智梅等<sup>[6]</sup> 建立神经网络回归算法模型 (RNN), 在此基础上利用平均影响值 (MIV) 算法定量分析各输入变量对发动机 NO<sub>x</sub> 的影响大小. 朱钰森等<sup>[7]</sup> 为了解决锅炉非线性、多工况、多耦合等特点, 提出了一种多模型方法预测 NO<sub>x</sub> 的排放量, 根据输入变量对输出变量的影响程度来评价样本之间的差异性, 并把计算出的平均影响值 (MIV) 作为与多模型连接的权系数, 克服了由于特征变量波动较大使得预测模型不稳定的缺陷. 于静等<sup>[8]</sup> 为了避免由于变量数目的原因造成预测模型精度降低, 采用互信息 (MI) 筛选出 5 个特征变量作为预测模型的输入, 降低模型训练的复杂性.

综上所述, 除了预测模型的选择之外, 特征变量的选择也将直接影响预测模型的精确度. 由实验确定的 NO<sub>x</sub> 排放规律的泛化性能较差, 不能很好地预测特定燃气轮机的 NO<sub>x</sub> 排放以及影响 NO<sub>x</sub> 排放的特征变量. 因此, 本文首先利用 PLS 确定应选择影响燃气轮机 NO<sub>x</sub> 排放的特征变量个数, 克服了由于经验和机理分析选取的主观性, 再利用 MI 选择相关性较大的变量, 确定预测模型的基础数据集. 为了验证 PLS-MI 组合特征选择模型的有效性, 将其与常用的特征选择方法进行对比分析.

## 1 氮氧化物与 SCR 脱硝系统

目前, 电站选择降低 NO<sub>x</sub> 排放主要有两种方式, 分别为燃烧优化控制和 SCR 脱硝处理<sup>[9]</sup>. 但无论哪种减排方式都需要对燃气轮机运行过程中 NO<sub>x</sub> 的排放量进行准确监测. SCR 脱硝系统的工作原理是在催化剂的作用下, 通过还原剂氨将 NO<sub>x</sub> 转化为氮气和水, 如图 1 所示. 在适当的温度和催化剂条件下, SCR 脱硝系统效率主要由氨的量决定, 发生的化学反应如式 (1)~式 (4)<sup>[10]</sup>. 从反应原理可以看出若喷氨量不足会降低 SCR 系统的脱硝效率, 导致 NO<sub>x</sub> 浓度排放超标; 若喷氨量过多, 会提高 SCR 系统的脱硝效率, 但是会造成还原剂氨的浪费和增加成本, 甚至会造成二次污染<sup>[11]</sup>. 为了减少 NO<sub>x</sub> 的排放, 避免氨流量过多导致资源浪费以及氨流量过少造成排放的 NO<sub>x</sub> 浓度超标, 必须实时测量和监控 SCR 脱硝系统的入口 NO<sub>x</sub> 排放浓度, 并对燃气轮机的运行实施优化控制<sup>[12]</sup>.

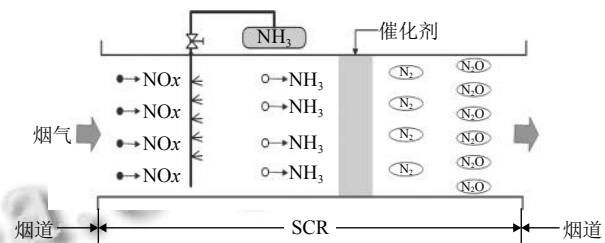
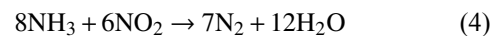
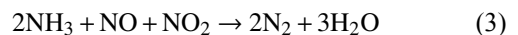
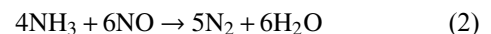
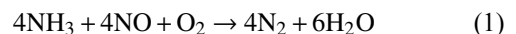


图 1 SCR 系统反应示意图

燃气轮机难以通过机理建模的方式实现对 NO<sub>x</sub> 排放浓度的精确描述. 相比于机理建模, 数据建模不需要深入了解对象机理特性. 它主要是以大数据为驱动, 统计学原理为基础, 利用先进人工智能算法获得对象的动态特性<sup>[13]</sup>. BP 神经网络是非线性过程建模的常用数据驱动模型, 若没有对数据集进行特征选择, 容易带来模型过拟合及建模时间变长等问题, 而且基于 BP 神经网络等方法建立 SCR 脱硝系统的入口 NO<sub>x</sub> 浓度预测模型参数的设置过多, 准确性受噪音影响较大, 这些模型对于小样本数据集的训练具有泛化性差和过拟合的缺陷.

综上所述, 准确的测量燃气轮机 NO<sub>x</sub> 排放以控制

氨注入量是需要解决的关键问题. 我国电厂主要是通过燃气轮机后接入了 SCR 脱硝反应器系统, 与 NH<sub>3</sub> 反应生成氮和水, 减少 NO<sub>x</sub> 排放, 但 SCR 系统的工作温度必须满足一定的范围和 NH<sub>3</sub> 适量使用. 另一种可行的方式是通过调整控制相关变量来保证燃气轮机健康运行的情况下, 对影响 NO<sub>x</sub> 排放的变量进行敏感性分析, 确定它们对燃气轮机 NO<sub>x</sub> 排放的影响大小以及正负相关性, 这一方法没有附加产物, 更加安全有效, 具有良好的应用前景.

## 2 燃气轮机 NO<sub>x</sub> 特征选择的模型理论

### 2.1 偏最小二乘法

偏最小二乘法 (partial least squares, PLS) 是一种多元统计数据分析方法, 在成分提取中, 通过将高维空间相关变量投影到一个低维空间变量<sup>[14]</sup>. 通过信息综合与筛选技术, 既可以从原始数据集中选择适当数量的可以较好地概括自变量系统  $x$  中的信息的潜在变量, 也可以消除变量间的共线问题, 尽可能准确的解释因变量  $y$  的变化<sup>[15]</sup>.

令  $Rdx$  表示第  $h$  个成分  $t_h$  对  $x$  的解释能力, 自变量个数为  $p$ , 第  $i$  个自变量  $x_i$  与第  $h$  个成分  $t_h$  的相关系数为  $r(x_i; t_h)$ , 如式 (5):

$$Rdx = \frac{1}{p} \sum_{i=1}^p r^2(x_i; t_h) \quad (5)$$

用  $m$  表示利用交叉有效性提取的主要成分个数, 则  $m$  个成分对  $x$  的累计解释能力  $CRdx$  为:

$$CRdx = \sum_{h=1}^m \left[ \frac{1}{p} \sum_{i=1}^p r^2(x_i; t_h) \right] \quad (6)$$

令  $q$  表示因变量个数,  $Rdy$  表示第  $h$  个成分  $t_h$  对  $y$  的解释能力:

$$Rdy = \frac{1}{q} \sum_{j=1}^q r^2(y_j; t_h) \quad (7)$$

则  $m$  个成分对  $y$  的累计解释能力  $CRdy$  为:

$$CRdy = \sum_{h=1}^m \left[ \frac{1}{q} \sum_{j=1}^q r^2(y_j; t_h) \right] \quad (8)$$

PLS 中的变量投影重要性 (variable importance in projection, VIP) 通过相关自变量综合的主成分解释特征变量  $x_i$  ( $i=1, 2, \dots, n$ ) 对自变量  $y$  的重要性, 并根据 VIP 值筛选影响较大的特征变量, 可以克服特征变量

$x_i$  间的多重相关性, 计算简单, 应用广泛. 它每个变量  $x_i$  的 VIP 值是通过 PLS 加载权重的平方和 ( $\omega_{ih}$ ) 计算得到, 加权的是每个组和中解释的平方和的数量, 重要性指标的计算公式为:

$$VIP_i = \sqrt{\frac{p}{CRdy} \sum_{h=1}^m Rdy \omega_{ih}^2} \quad (9)$$

其中,  $\omega_{ih}$  是轴  $\omega_h$  的第  $i$  个分量.

### 2.2 互信息理论

信息熵是由美国科学家香农提出, 主要是用来度量随机变量的不确定性程度. 其值越大, 所蕴含的信息量越大<sup>[16]</sup>. 用  $H(X)$  表示信息源  $X = \{x_1, x_2, \dots, x_n\}$  对应的信息熵, 每个信息源  $x_i$  所对应的概率为  $p(x_i)$ , 则:

$$H(X) = - \sum_{i=1}^n p(x_i) \log_2 \frac{1}{p(x_i)} = - \sum_{i=1}^n p(x_i) \log_2 p(x_i) \quad (10)$$

在条件  $Y$  确定的情况下, 信息源  $X$  可以用条件熵来描述:

$$H(X|Y) = - \sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) \log_2 p(x_i|y_j) \quad (11)$$

$p(x_i, y_j)$  为二维向量  $(X, Y)$  的联合概率分布, 则联合熵定义为:

$$H(X, Y) = - \sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) \log_2 p(x_i, y_j) \quad (12)$$

互信息 (mutual information, MI) 用于分析每个可能的输入变量与目标输出变量之间的非线性关系, 从中选择最相关的变量作为模型的输入变量<sup>[17]</sup>. 对于给定目标变量  $Y$  后, 特征变量  $X$  的不确定性减少程度为:

$$\begin{aligned} I(X; Y) &= \iint_{YX} p_{XY}(x, y) \log \left( \frac{p_{XY}(x, y)}{p_X(x)p_Y(y)} \right) dx dy \\ &= \sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) \log_2 \frac{p(x_i, y_j)}{p(x_i)p(y_j)} \end{aligned} \quad (13)$$

其中,  $p(x_i)$  为  $X$  的边缘概率分布,  $p(y_j)$  为  $Y$  的边缘概率分布.

偏最小二乘法既能较好地概括自变量系统中的信息, 又能很好的解释因变量的变化并排除系统中的噪声干扰<sup>[18]</sup>. 互信息不仅能反应线性关系, 也能反应其非线性关系, 因此互信息在反应变量之间的相关性比相关系数更加全面<sup>[19]</sup>. 因此, 本文采用 PLS 特征提取和 MI 特征选择混合算法 (PLS-MI) 选取燃气轮机氮氧化



物的影响变量。

### 3 仿真结果分析

#### 3.1 数据获取与预处理

燃气轮机的运作过程是一个复杂的化学反应过程,国内外学者对基于数据的氮氧化物的特征变量均有所研究,从不同参数条件下进行了分析,发现控制燃料流量、燃料组分、环境温度、燃烧温度和燃气轮机的进出口压力差等都可以影响 NO<sub>x</sub> 的产生。燃气轮机 NO<sub>x</sub> 排放特性主要与燃烧室内燃烧模式及燃烧特性相关,影响因素主要有燃料流量、空气流量、燃气轮机负荷、燃烧室温度等。燃气轮机燃烧室温度主要由燃料流量和空气流量决定。

为了研究影响 NO<sub>x</sub> 排放情况的关键因素,文献 [5] 中选取电厂稳态数据的环境温度、天然气温度、透平出口温度、IGV 开度、值班火焰燃料质量流量、预混火焰燃料质量流量、燃烧室进口压力、燃烧室压降等 8 个变量作为模型的输入变量。本文基于试验过程中积累的先验知识,以冗余的选择原则,选取影响燃气轮机 NO<sub>x</sub> 排放水平的 19 个性能参数作为燃气轮机的影响因素,分别为:环境温度(AT)、环境压力(AP)、环境湿度(AH)、空气过滤器压差(AFDP)、燃气轮机排放压力(GTEP)、燃气轮机轴转矩(GTT)、燃气发生器转速(GGn)、GT 压缩机进出口空气温度差 T、涡轮进口温度(TIT)、涡轮后温度(TAT)、涡轮能量产率(TEY)、高压涡轮出口压力(P1)、燃气轮机废气压力(Pexh)、燃气轮机压缩机进出口空气压力差(P2)、涡轮出口高压(HP)温度(T)、压缩机排气压力(CDP)、燃料流量(mf)、涡轮注入控制(TIC)、一氧化碳(CO)等,在此基础上提取对燃气轮机 NO<sub>x</sub> 影响较大的特征变量。

偏最小二乘法的特征变量个数是通过残差阈值来确定。当特征变量个数增加时,方差解释信息的变化小于一定的阈值就停止分解,如图 2 的特征变量的方差解释。从图 2 中可以看出当特征变量个数为 7 时,因变量中约有 95% 的方差信息被解释。此时提取新的特征变量得到的方差解释信息的变化较小,因此剩余的残差可被认为是噪声干扰。若再增加特征变量个数,会引入噪声并且使模型的复杂度增加,从而降低模型的精确度。

为了验证 PLS 在选择 NO<sub>x</sub> 特征变量个数的有效性,选择特征提取中最常用的主成分分析(principal

component analysis, PCA) 和随机森林(random forest, RF) 方法作为对比<sup>[20,21]</sup>。利用主成分分析(PCA)对燃气轮机氮氧化物的影响因素个数进行分析,计算主成分因子的累计方差贡献率,如图 3 所示。为了减少信息的丢失,选择累计方差贡献率 95% 为阈值,把方差贡献率的累加值超过 95% 的因子数作为特征变量的选取数<sup>[20]</sup>。从图 3 中可以看出,当第 5 个特征因子出现的时候,其总体的累计方差贡献率为 95.06%,超过了 95%,则可以选取 5 个特征变量代替原变量作为预测模型的输入。在随机森林中,为了选择哪些特征更重要,需要考虑的是特征的累加重要性,通常以 95% 为阈值,当其特征重要性的累加值超过该阈值时,就选择其特征变量个数代替原来的数据集<sup>[21]</sup>。从图 4 可以看出,当特征变量个数为 6 时,特征变量的累加重要性达到了 95.35%,因此应该选取 6 个特征变量。

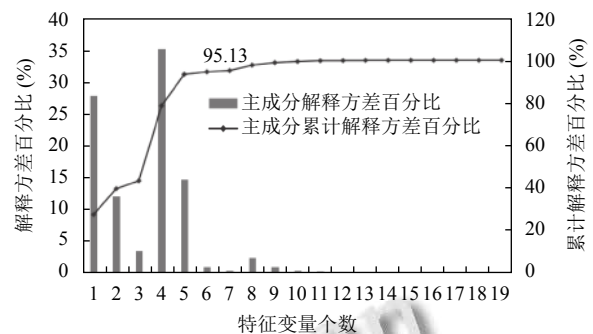


图 2 PLS 分析因变量方差解释与特征变量个数变化关系

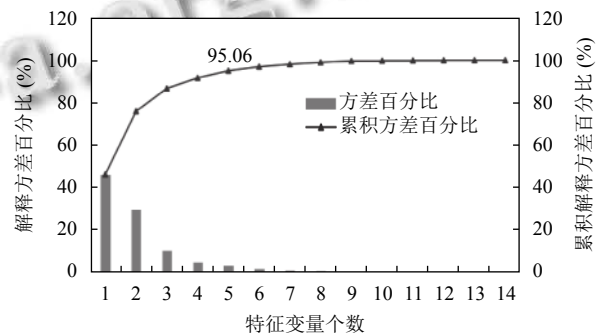


图 3 PCA 分析因变量方差解释与特征变量个数变化关系

综上,不同方法选择的特征变量的数目不同,利用 PLS、PCA、RF 应选择特征变量分别是 7 个、5 个和 6 个。为了更好地证明所选用模型在燃气轮机特征变量选择中具有较好的适用性,把单一模型 PLS、MI、RF 和组合模型 RF-MI、PCA-MI、PCA-RF、PCA-PLS、PLS-MI 的预测仿真结果对比分析。

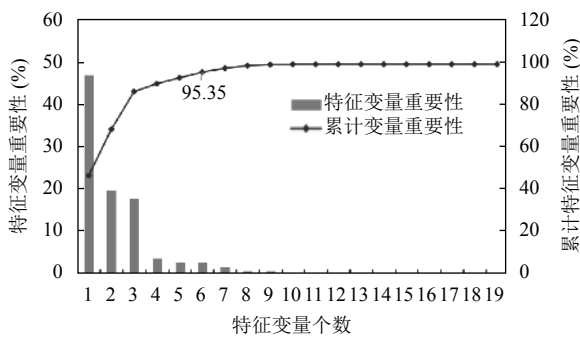


图4 RF分析特征变量重要性与特征变量个数之间的关系

### 3.2 不同变量选择方法的仿真结果

为了更好地验证预测模型的仿真效果,采用度量建立模型与样本数据拟合程度的均方根误差 (*RMSE*)、平均绝对值百分比误差 (*MAPE*)、平均绝对误差 (*MAE*) 3个评价指标,如式(14)–式(16)。

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t)^2} \quad (14)$$

$$MAPE = \frac{1}{N} \sum_{t=1}^N \frac{|y_t - \hat{y}_t|}{y_t} \quad (15)$$

$$MAE = \frac{1}{N} \sum_{t=1}^N |y_t - \hat{y}_t| \quad (16)$$

其中,  $y_t$  为第  $t$  时刻的  $\text{NO}_x$  排放量;  $\hat{y}_t$  为第  $t$  时刻的  $\text{NO}_x$  排放量的预测值;  $N$  为预测样本数目。 *RMSE* 是常见得衡量回归模型性能的评价指标, *RMSE* 指标越小,说明模型的预测精度越高,拟合程度越好,特征变量对目标变量的解释性越强。

利用 Matlab R2018a 的开发环境进行仿真实验,求出各个特征变量的互信息值、偏最小二乘法的重要性值、随机森林的重要性值,如表1所示。

从表1中可以看出,利用 PLS 选取的7个变量为  $X_1$ 、 $X_4$ 、 $X_9$ 、 $X_{11}$ 、 $X_{14}$ 、 $X_{16}$ 、 $X_{19}$ ,利用随机森林应该选取6个特征变量为  $X_1$ 、 $X_3$ 、 $X_4$ 、 $X_5$ 、 $X_{10}$ 、 $X_{19}$ ,利用主成分分析法应该选择5个特征变量。单一的 PLS、MI、RF 特征选择方法选择影响较大的变量,并进行预测仿真分析,组合模型是利用 PLS、PCA、RF 确定出应选择出的特征变量数目,然后再利用 MI 选择出相关性较强冗余性小的变量,代入 BP 神经网络的预测模型,RF-MI、PCA-MI、PCA-RF、PCA-PLS、PLS-MI 的预测仿真结果如表2所示。

表1  $\text{NO}_x$  排放量的多模型特征变量选择结果

序号	变量名称	PLS	RF	MI
$X_1$	环境温度AT (°C)	1.7502	0.47	0.4769
$X_2$	环境压力AP (kPa)	0.9795	0.01	0.4268
$X_3$	环境湿度AH	1.0283	0.03	0.4742
$X_4$	空气过滤器压差AFDP (kPa)	1.2457	0.18	0.4760
$X_5$	燃气轮机排放压力GTEP (kPa)	1.0286	0.03	0.4782
$X_6$	燃气轮机轴转矩GTT	0.6883	0	0.1761
$X_7$	燃气发生器转速GGn (rpm)	0.4376	0	0.2583
$X_8$	GT压缩机进出口温度差T (°C)	0.1971	0	0.3611
$X_9$	燃机涡轮进口温度TIT (°C)	1.1679	0.02	0.4609
$X_{10}$	燃机涡轮出口温度TAT (°C)	0.9009	0.04	0.3365
$X_{11}$	燃机涡轮能量产率TEY (%)	1.0557	0.01	0.4561
$X_{12}$	高压涡轮出口压力 (kPa)	0.5098	0	0.2051
$X_{13}$	GT废弃压力 (kPa)	1.0222	0	0.0957
$X_{14}$	压缩机进出口空气压力差 (kPa)	1.5827	0	0.3198
$X_{15}$	燃机涡轮出口高压温度 (°C)	0.5136	0	0.4122
$X_{16}$	压缩机排气压力 (kPa)	1.0614	0	0.4702
$X_{17}$	燃料流量 ( $\text{kg}^{-\text{s}}$ )	0.3500	0	0.3250
$X_{18}$	涡轮注入控制TIC (%)	0.3212	0	0.3225
$X_{19}$	一氧化碳CO	1.3748	0.2	0.3203

表2 BP神经网络预测模型的仿真结果

模型名称	MAE	RMSE	MAPE (%)	Time (s)
PLS	5.5863	6.3503	10.5215	0.80
MI	3.4847	4.4473	6.5172	2.59
RF	3.5400	4.7432	6.6844	0.26
RF-MI	3.3253	3.9877	6.2351	0.22
PCA-MI	3.6059	4.8541	6.6931	0.19
PCA-RF	3.5112	4.4394	6.5844	0.27
PCA-PLS	4.5603	5.1530	8.5644	0.16
<b>PLS-MI</b>	<b>2.8243</b>	<b>3.5061</b>	<b>5.2932</b>	<b>0.25</b>

为了防止 BP 神经网络预测模型存在偶然性问题,再利用遗传算法 (GA) 优化 BP 神经网络的 GABP 模型和天牛须搜索算法 (BAS) 优化 BP 神经网络的 BASBP 模型进行预测仿真实验,结果如表3和表4所示。

表3 GA-BP网络模型预测的仿真结果

模型名称	MAE	RMSE	MAPE (%)	Time (s)
PLS	4.3272	4.9725	8.0600	70.12
MI	3.2294	4.2357	6.0633	90.73
RF	3.6726	4.7448	6.9078	42.36
RF-MI	3.4319	4.1085	6.4124	24.73
PCA-MI	3.1628	4.3666	5.8932	100.65
PCA-RF	3.7715	4.6770	7.0682	27.61
PCA-PLS	3.4781	4.0168	6.4849	45.26
<b>PLS-MI</b>	<b>3.2213</b>	<b>4.0148</b>	<b>6.0387</b>	<b>25.91</b>

研究问题时将 BP 神经网络、GABP、BASBP 网络预测模型分别在进行特征变量处理和不同特征变量选择方法下进行实验并对预测结果进行了统计,通过

表2-表4中的评价指标分析,利用随机森林和主成分分析时,选择的特征变量较PLS的较小,不具有代表性,缺失了一部分信息.在实际问题中,每一类方法都有一定的局限性.采用PLS-MI组合的特征选择优于单一的PLS、MI特征选择,说明PLS和MI在提取特征时只估计单个特征的得分,未评估特征子集的性能,导致提取特征变量时丢失大量有用信息,降低模型的预测性能.把PLS和MI模型组合对影响燃气轮机氮氧化物的特征变量选取实践中,通过不同的仿真模型证明了PLS-MI组合模型较单一的PLS、MI、RF和组合的PCA-MI、RF-MI、PCA-RF、PCA-PLS特征变量选择方法提取的特征变量更具有代表性且能提高模型预测性能.

表4 BAS-BP网络模型预测的仿真结果

模型名称	MAE	RMSE	MAPE (%)	Time (s)
PLS	3.7571	4.4766	6.9781	10.9
MI	2.8246	3.6135	5.1980	16.15
RF	2.9034	3.5240	5.3925	11.83
RF-MI	2.7667	3.4750	5.1377	10.46
PCA-MI	3.0985	3.9049	5.7010	16.84
PCA-RF	2.6252	3.4513	4.9030	16.14
PCA-PLS	3.2574	3.8827	6.0563	11.16
<b>PLS-MI</b>	<b>2.4881</b>	<b>3.1411</b>	<b>4.6508</b>	<b>14.00</b>

### 3.3 灵敏度分析

首先利用PLS-MI的组合特征选择方法选择出燃气轮机NO<sub>x</sub>的影响因素,根据表1选择出MI值较大的特征变量为: $X_1$ 、 $X_5$ 、 $X_4$ 、 $X_3$ 、 $X_{16}$ 、 $X_9$ 、 $X_{11}$ .

特征选择的目的是准确预测燃气轮机NO<sub>x</sub>的排放,则需要对影响NO<sub>x</sub>排放的变量进行敏感性分析,确定特征变量对NO<sub>x</sub>的影响大小以及正负相关性,如图5所示.选取的7个变量中, $X_1$ 、 $X_4$ 、 $X_3$ 为环境变量不可调,调整SCR系统的入口NO<sub>x</sub>浓度应着眼于 $X_5$ 、 $X_{16}$ 、 $X_9$ 和 $X_{11}$ 的4个独立变量进行调整.

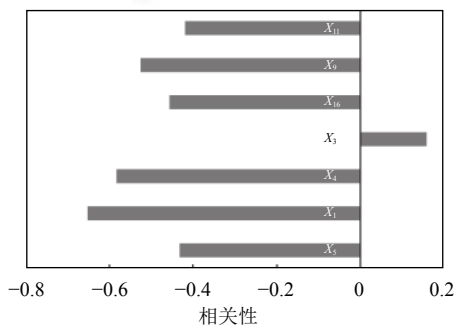


图5 相关性分析

## 4 结论

在对燃气轮机运行的系统分析研究前,需要对工艺流程中的各个变量进行评估,选择数据价值高的变量.目前研究常用的方法大多都是以多元线性回归作为基础进行分析,以特征与目标的相关性作为变量选择的条件,但是在研究中并没有考虑多重线性的影响.

通过研究发现,在对燃气轮机NO<sub>x</sub>数据分析和特征提取的过程中,为了得到更多的有用信息,都是尽可能多的选择有价值的特征.但当收集的样本数据集变量过多时会具有高度非线性,导致建立的预测模型会耗费大量的时间,也易出现过拟合的现象,特征变量选择在建模过程中成为了一个非常重要的部分.本文在选择燃气轮机NO<sub>x</sub>排放量的影响因素时,兼得了单一模型稳定的泛化性能与多模型较好的跟踪能力,克服了由于存在非线性、多工况、输出变量波动较大而使得建模效果不稳定的难题.

## 参考文献

- 唐振浩,柴向颖,曹生现,等.考虑时延特征的燃煤锅炉NO<sub>x</sub>排放深度学习建模.中国电机工程学报,2020,40(20):6633-6643.
- 董泽,闫来清.SCR脱硝系统NO<sub>x</sub>排放浓度建模与仿真.系统仿真学报,2020,32(2):172-181.
- 吕游,刘吉臻,杨婷婷,等.基于PLS特征提取和LS-SVM结合的NO<sub>x</sub>排放特性建模.仪器仪表学报,2013,34(11):2418-2424.
- Li QW, Zhou KY, Yao GH. Combustion optimization model for NO<sub>x</sub> reduction with an improved particle swarm optimization. Journal of Shanghai Jiaotong University (Science), 2016, 21(5): 569-575. [doi: 10.1007/s12204-016-1764-6]
- 赵刚,朱华昕,李苏辉,等.基于数据和神经网络的燃气轮机NO<sub>x</sub>排放预测与优化.动力工程学报,2021,41(1):22-27.
- 伦智梅,张振京,宋业栋,等.基于回归神经网络的柴油机NO<sub>x</sub>排放预测及影响分析.山东理工大学学报(自然科学版),2021,35(3):81-86.
- 朱钰森,金晓明,张泉灵.基于聚类与加权连接的锅炉NO<sub>x</sub>排放量多模型建模.控制工程,2019,26(4):688-693.
- 于静,金秀章,刘岳.基于结构改进RBF神经网络的NO<sub>x</sub>预测模型比较.控制工程. <https://doi.org/10.14107/j.cnki.kzgc.20210150>. (2021-06-05)[2021-07-05].
- 余印振,韩哲哲,许传龙.基于深度卷积神经网络和支持向量机的NO<sub>x</sub>浓度预测.中国电机工程学报. <https://doi.org/10.13334/j.0258-8013.pcsee.210077>. (2021-04-01)[2021-06-23].

- 10 许壮, 康英伟. 基于随机森林的火电机组 SCR 脱硝反应器建模. 动力工程学报, 2020, 40(6): 486–491, 501.
- 11 Yang WJ, Ren JN, Zhang HW, *et al.* Single-atom iron as a promising low-temperature catalyst for selective catalytic reduction of NO<sub>x</sub> with NH<sub>3</sub>: A theoretical prediction. Fuel, 2021, 302: 121041. [doi: [10.1016/j.fuel.2021.121041](https://doi.org/10.1016/j.fuel.2021.121041)]
- 12 Yan LQ, Dong Z, Jia H, *et al.* Dynamic inferential NO<sub>x</sub> emission prediction model with delay estimation for SCR de-NO<sub>x</sub> process in coal-fired power plants. Royal Society Open Science, 2020, 7(2): 191647. [doi: [10.1098/rsos.191647](https://doi.org/10.1098/rsos.191647)]
- 13 Lv Y, Yang TT, Liu JZ. An adaptive least squares support vector machine model with a novel update for NO<sub>x</sub> emission prediction. Chemometrics & Intelligent Laboratory Systems, 2015, 145: 103–113.
- 14 Dong Z, Ma N, Li CQ. NO<sub>x</sub> emission model for coal-fired boilers using partial least squares and extreme learning machine. Journal of Southeast University (English Edition), 2019, 35(2): 179–184.
- 15 Li Z, Lee YS, Chen JH, *et al.* Developing variable moving window PLS models: Using case of NO<sub>x</sub> emission prediction of coal-fired power plants. Fuel, 2021, 296: 120441. [doi: [10.1016/j.fuel.2021.120441](https://doi.org/10.1016/j.fuel.2021.120441)]
- 16 吴雨, 刘媛华. 基于最近最远邻和互信息的特征选择方法. 计算机应用研究, 2017, 34(12): 3713–3716. [doi: [10.3969/j.issn.1001-3695.2017.12.044](https://doi.org/10.3969/j.issn.1001-3695.2017.12.044)]
- 17 杨志淳, 靖晓平, 乐健, 等. 基于 MI-PSO-BP 算法的配电设备状态实时评估方法. 电力自动化设备, 2019, 39(12): 108–113.
- 18 Zhang C, Peng T, Zhou JZ, *et al.* An improved autoencoder and partial least squares regression-based extreme learning machine model for pump turbine characteristics. Applied Sciences, 2019, 9(19): 3987. [doi: [10.3390/app9193987](https://doi.org/10.3390/app9193987)]
- 19 Gao LY, Wu WG. Relevance assignment feature selection method based on mutual information for machine learning. Knowledge-Based Systems, 2020, 209: 106439. [doi: [10.1016/j.knosys.2020.106439](https://doi.org/10.1016/j.knosys.2020.106439)]
- 20 李郅琴, 杜建强, 聂斌, 等. 特征选择方法综述. 计算机工程与应用, 2019, 55(24): 10–19. [doi: [10.3778/j.issn.1002-8331.1909-0066](https://doi.org/10.3778/j.issn.1002-8331.1909-0066)]
- 21 Simsekler MCE, Alhashmi NH, Azar E, *et al.* Exploring drivers of patient satisfaction using a random forest algorithm. BMC Medical Informatics and Decision Making, 2021, 21(1): 157. [doi: [10.1186/s12911-021-01519-5](https://doi.org/10.1186/s12911-021-01519-5)]