(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 110135057 B
(45)授权公告日 2021.03.02

(21)申请号 201910397710.5

(22)申请日 2019.05.14

(65)同一申请的已公布的文献号
申请公布号 CN 110135057 A

(43)申请公布日 2019.08.16

(73)专利权人 北京工业大学
地址 100124 北京市朝阳区平乐园100号

(72)发明人 乔俊飞 郭子豪 汤健

(74)专利代理机构 北京思海天达知识产权代理有限公司 11203
代理人 刘萍

(51)Int.Cl.
G06F 30/27(2020.01)

(56)对比文件
CN 109190660 A,2019.01.11
CN 108549792 A,2018.09.18
CN 107944173 A,2018.04.20
CN 103366100 A,2013.10.23
CN 108090317 A,2018.05.29
CN 109583115 A,2019.04.05
CN 108062566 A,2018.05.22
CN 109492319 A,2019.03.19
US 2015278703 A1,2015.10.01
US 9875440 B1,2018.01.23
WO 2005008572 A1,2005.01.27
汤健等.基于选择性集成核学习算法的固废焚烧过程二英排放浓度软测量.《化工学报》.2019,第70卷(第2期),
汤健等.基于遗传算法-偏最小二乘进行谱特征选择的磨机负荷软测量方法.《第二十九届中国控制会议论文集》.2010,
简葳玓.软测量模型的变量选择方法研究.《中国优秀硕士学位论文全文数据库 工程科技I辑》.2017,(第8期),
Yujing Sun et al..Correlation Feature Selection and Mutual Information Theory Based Quantitative Research on Meteorological Impact Factors of Module Temperature for Solar Photovoltaic Systems.《energies》.2016,第10卷(第1期),

审查员 杨杭

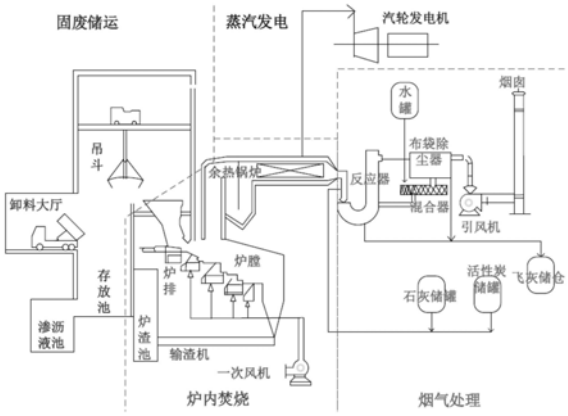
权利要求书6页 说明书16页 附图9页

(54)发明名称

基于多层特征选择的固废焚烧过程二噁英排放浓度软测量方法

(57)摘要

基于多层特征选择的固废焚烧过程二噁英排放浓度软测量方法属于软测量领域。本文提出基于多层特征选择的MSWI过程DXN排放浓度软测量方法。首先,从单特征与DXN相关性视角,结合相关系数和互信息构建综合评价值指标,实现MSWI多个子系统过程变量的第1层特征选择;接着,从多特征冗余性和特征选择鲁棒性视角,多次运行基于GA-PLS的特征选择算法,实现第2层特征选择;最后,结合上层选择特征的统计频次、模型预测性能及机理知识进行第3层特征选择,构建得到DXN排放浓度软测量模型。结合某焚烧厂的多年DXN检测数据验证了所提方法的有效性。



CN 110135057 B

1.基于多层特征选择的固废焚烧过程二噁英排放浓度软测量方法，其特征在于：

结合焚烧工艺将基于炉排炉的城市固废焚烧MSWI过程分为6个子系统：燃烧处理工程、锅炉设备工程、尾气处理工程、蒸汽发电工程、烟囱排放工程、公用传输工程；

软测量模型的输入数据$X\in R^{N\times P}$包括N个样本即为行和P个变量即为列，其源于MSWI流程的不同子系统；

此处，将来自第ith个子系统的输入数据表示为$\boldsymbol{X}_i\in R^{N\times P_i}$，即存在如下关系，

$$\boldsymbol{X}=[\boldsymbol{X}_1,\cdots,\boldsymbol{X}_i,\cdots,\boldsymbol{X}_I]=\{\boldsymbol{X}_i\}_{i=1}^{I} \quad (1)$$

$$P=P_1+\cdots+P_i+\cdots+P_I=\sum_{i=1}^{I}P_i \quad (2)$$

其中，I表示子系统个数，P_i表示第ith个子系统包含的输入特征个数；

相应的，输出数据$\boldsymbol{y}=\{y_n\}_{n=1}^{N}\in\boldsymbol{R}^{N\times 1}$包括N个样本，其来源于采用离线直接检测法得到排放物二噁英DXN检测样本；

过程变量以秒为单位在DCS系统采集与存储，DXN排放浓度以月/季为周期离线直接化验获得，故存在N＜＜P；

将X_i改写为如下形式，

$$\begin{aligned}\mathbf{X}_i&=[\{(x_n^1)_i\}_{n=1}^{N},\cdots,\{(x_n^{P_i})_i\}_{n=1}^{N},\cdots,\{(x_n^{P_i})_i\}_{n=1}^{N}]\\&=[(\mathbf{x}^1)_i,\cdots,(\mathbf{x}^{P_i})_i,\cdots,(\mathbf{x}^{P_i})_i]\\&=\{(\mathbf{x}^{p_i})_i\}_{p_i=1}^{P_i}\end{aligned} \quad (3)$$

其中，$(\mathbf{x}^{p_i})_i$表示第ith个子系统的第$p_i$th个输入特征，$\mathbf{x}^{p_i}=\left\{\mathbf{x}_n^{p_i}\right\}_{n=1}^{N}$表示列向量；

提出基于多层特征选择的MSWI过程DXN排放浓度软测量策略：先介绍以下术语：$(\mathbf{X}_{\text{corr}}^{\text{sel}})_i$和$(\mathbf{X}_{\text{mi}}^{\text{sel}})_i$表示针对第ith个子系统的输入特征采用相关系数和互信息度量所选择的候选特征集合，$(\mathbf{X}_{\text{1st}}^{\text{sel}})_i$表示对基于相关系数法和互信息法所选择的候选特征集合采用综合评价值度量所选择的对第ith个子系统的第1层特征，$\mathbf{X}_{\text{1st}}^{\text{sel}}$表示串行组合全部子系统的第1层特征所得到的基于单特征相关性的第1层特征，$(\mathbf{X}_{\text{2nd}}^{\text{sel}})_j$表示运行第jth次GA-PLS算法所选择的基于多特征冗余性的第2层特征，$f_{\text{num}}^{p_{\text{1st}}^{\text{sel}}}$表示第1层特征中第$p_{\text{1st}}^{\text{sel}}$th个特征被选择的次数，$\mathbf{X}_{\text{3rd}}^{\text{sel}}$表示依据特征选择阈值$\theta_{3\text{rd}}$和先验知识从$\mathbf{X}_{\text{1st}}^{\text{sel}}$中所选择的第3层特征，$M_{\text{para}}$表示软测量模型的参数，$\hat{\boldsymbol{y}}$表示预测值；

1.1)首先，计算不同原始输入特征与DXN排放浓度间的原始相关系数；此处以第ith个子系统的第pth个输入特征$(\mathbf{x}^{p_i})_i=\{(x_n^{p_i})_i\}_{n=1}^{N}$为例进行描述，如下，

$$(\xi_{\text{corr\_ori}}^{p_i})_i=\frac{\sum_{n=1}^{N}\left[\left((x_n^{p_i})_i-\bar{x}_{p_i}\right)(y_n-\bar{y})\right]}{\sqrt{\sum_{n=1}^{N}\left((x_n^{p_i})_i-\bar{x}_{p_i}\right)^2}\sqrt{\sum_{n=1}^{N}(y_n-\bar{y})^2}} \quad (4)$$

其中，$\bar{x}_{p_i}$，$\bar{y}$分别表示第ith个子系统的第pth个输入特征及DXN排放浓度N个建模样本的平均值；将原始相关系数$(\xi_{\text{corr_ori}}^{p_i})_i$进行如下预处理，

$$(\xi_{\text{corr}}^{p_i})_i=\left|(\xi_{\text{corr\_ori}}^{p_i})_i\right| \quad (5)$$

其中，|·|表示取绝对值；

重复上述过程，获得全部原始输入特征的相关系数并记为$\{\xi_{\text{corr}}^{p_i}\}_{p_i=1}^{P_i}$；设定第ith个子系统权重因子$f_i^{\text{corr}}$，将基于相关系数选择输入特征的阈值$\theta_i^{\text{corr}}$采用如下公式计算，

$$\theta_i^{\text{corr}}=f_i^{\text{corr}}\cdot\frac{1}{p_i}\sum_{p_i=1}^{P_i}(\xi_{\text{corr}}^{p_i})_i \quad (6)$$

其中，最大$(f_i^{\text{corr}})_{\max}$和最小值$(f_i^{\text{corr}})_{\min}$采用如下公式计算，

$$\begin{cases}(f_i^{\text{corr}})_{\max}=\dfrac{\max\left((\xi_{\text{corr}}^{1})_i,\cdots,(\xi_{\text{corr}}^{p_i})_i,\cdots,(\xi_{\text{corr}}^{P_i})_i\right)}{\dfrac{1}{p_i}\sum_{p_i=1}^{P_i}(\xi_{\text{corr}}^{p_i})_i}\\ (f_i^{\text{corr}})_{\min}=\dfrac{\min\left((\xi_{\text{corr}}^{1})_i,\cdots,(\xi_{\text{corr}}^{p_i})_i,\cdots,(\xi_{\text{corr}}^{P_i})_i\right)}{\dfrac{1}{p_i}\sum_{p_i=1}^{P_i}(\xi_{\text{corr}}^{p_i})_i}\end{cases} \quad (7)$$

其中，max(·)和min(·)分别表示取最大和最小值的函数；

以θ_i^{corr}作为阈值，第ith个子系统的第$p_i$th输入特征的选择准则如下所示，

$$\alpha_i^{p_i}=\begin{cases}1,\ \text{if}\quad (\xi_{\text{corr}}^{p_i})_i\geq\theta_i^{\text{corr}}\\ 0,\ \text{else}\ (\xi_{\text{corr}}^{p_i})_i<\theta_i^{\text{corr}}\end{cases} \quad (8)$$

选择其中$\alpha_i^{p_i}=1$的特征$(\mathbf{x}^{p_i})_i$作为基于相关系数选择的候选特征并将其标记为$(\mathbf{x}^{(p_i)_{\text{corr}}^{\text{sel}}})_i$；对第ith个子系统的全部原始输入特征执行上述过程，并将所选择的候选特征标记为，

$$(\mathbf{X}_{\text{corr}}^{\text{sel}})_i=[(\mathbf{x}^{1})_i,\cdots,(\mathbf{x}^{(p_i)_{\text{corr}}^{\text{sel}}})_i,\cdots,(\mathbf{x}^{(P_i)_{\text{corr}}^{\text{sel}}})_i] \quad (9)$$

其中，$(P_i)_{\text{corr}}^{\text{sel}}$表示基于相关系数选择的第ith个子系统的过程变量个数；

对全部子系统重复上述过程，基于相关系数度量选择的特征标记为$\{(\mathbf{X}_{\text{corr}}^{\text{sel}})_i\}_{i=1}^{I}$；

1.2)基于互信息的单特征相关性度量

首先，计算不同原始输入特征与DXN排放浓度间的互信息值；以第ith个子系统的第pth个输入特征$(\mathbf{x}^{p_i})_i$为例，如下：

$$(\xi_{\text{mi}}^{p_i})_i=\sum_{n=1}^{N}\sum_{n=1}^{N}\left\{p_{\text{rob}}\left((x_n^{p_i})_i,y_n\right)\log\left(\frac{p_{\text{rob}}\left((x_n^{p_i})_i,y_n\right)}{p_{\text{rob}}\left((x_n^{p_i})_i\right)p_{\text{rob}}(y_n)}\right)\right\} \quad (10)$$

其中，$p_{\text{rob}}\left((x_n^{p_i})_i,y_n\right)$表示联合概率密度，$p_{\text{rob}}\left((x_n^{p_i})_i\right)$和$p_{\text{rob}}(y_n)$表示边际概率密度；

重复上述过程，获得全部原始输入特征的互信息值并记为$\{\xi_{\text{mi}}^{p_i}\}_{p_i=1}^{P_i}$；设定第ith个子系统的权重因子$f_i^{\text{mi}}$，将基于互信息选择输入特征的阈值$\theta_i^{\text{mi}}$采用如下公式计算，

$$\theta_i^{\text{mi}}=f_i^{\text{mi}}\cdot\frac{1}{p_i}\sum_{p_i=1}^{P_i}(\xi_{\text{mi}}^{p_i})_i \quad (11)$$

其中，f_i^{mi}的最大$(f_i^{\text{mi}})_{\max}$和最小值$(f_i^{\text{mi}})_{\min}$采用如下公式计算，

$$\begin{cases}(f_i^{\mathrm{mi}})_{\max}=\dfrac{\max\left((\xi_{\mathrm{mi}}^{1})_i,\cdots,(\xi_{\mathrm{mi}}^{p_i})_i,\cdots,(\xi_{\mathrm{mi}}^{P_i})_i\right)}{\dfrac{1}{p_i}\sum_{p_i=1}^{P_i}(\xi_{\mathrm{mi}}^{p_i})_i}\\(f_i^{\mathrm{mi}})_{\min}=\dfrac{\min\left((\xi_{\mathrm{mi}}^{1})_i,\cdots,(\xi_{\mathrm{mi}}^{p_i})_i,\cdots,(\xi_{\mathrm{mi}}^{P_i})_i\right)}{\dfrac{1}{p_i}\sum_{p_i=1}^{P_i}(\xi_{\mathrm{mi}}^{p_i})_i}\end{cases}\quad(12)$$

以θ_i^{mi}作为阈值，第ith个系统的第pith输入特征的选择准则如下所示，

$$\beta_i^{p_i}=\begin{cases}1,\ \text{if}\quad(\xi_{\mathrm{mi}}^{p_i})_i\geq\theta_i^{\mathrm{mi}}\\0,\ \text{else}\ (\xi_{\mathrm{mi}}^{p_i})_i<\theta_i^{\mathrm{mi}}\end{cases}\quad(13)$$

选择其中$\beta_i^{p_i}=1$的特征$(\mathbf{x}^{p_i})_i$作为基于互信息选择的候选特征并将其表标记为$(\mathbf{x}^{(p_i)_{\mathrm{mi}}^{\mathrm{sel}}})_i$；对第ith个子系统的全部输入特征执行上述过程，并将所选择的候选特征标记为：

$$(\mathbf{X}_{\mathrm{mi}}^{\mathrm{sel}})_i=[(\mathbf{x}^{1})_i,\cdots,(\mathbf{x}^{(p_i)_{\mathrm{mi}}^{\mathrm{sel}}})_i,\cdots,(\mathbf{x}^{(P_i)_{\mathrm{mi}}^{\mathrm{sel}}})_i]\quad(14)$$

其中，$(P_i)_{\mathrm{mi}}^{\mathrm{sel}}$表示基于互信息选择的第ith个子系统的全部特征的个数；

对全部子系统重复上述过程，基于互信息度量选择的特征可标记为$\{(\mathbf{X}_{\mathrm{mi}}^{\mathrm{sel}})_i\}_{i=1}^{I}$；

1.3)基于综合评价值的单特征相关性度量

以第ith个子系统为例，同时考虑具有相关系数和互信息贡献度的输入特征在$(\mathbf{X}_{\mathrm{mi}}^{\mathrm{sel}})_i$和$(\mathbf{X}_{\mathrm{corr}}^{\mathrm{sel}})_i$中得到候选特征集合，其策略为：

$$(\mathbf{X}_{\mathrm{corr\_mi}}^{\mathrm{sel}})_i=(\mathbf{X}_{\mathrm{mi}}^{\mathrm{sel}})_i\cap(\mathbf{X}_{\mathrm{corr}}^{\mathrm{sel}})_i=[(\mathbf{x}^{1})_i,\cdots,(\mathbf{x}^{(p_i)_{\mathrm{corr\_mi}}^{\mathrm{sel}}})_i,\cdots,(\mathbf{x}^{(P_i)_{\mathrm{corr\_mi}}^{\mathrm{sel}}})_i]\quad(15)$$

其中，$\cap$表示取交集；$\mathbf{x}_i^{(p_i)_{\mathrm{corr_mi}}^{\mathrm{sel}}}$表示第ith个子系统的第$(p_i)_{\mathrm{corr_mi}}^{\mathrm{sel}}$th个候选特征，其对应的相关系数值与互信息值为$(\xi_{\mathrm{corr}}^{(p_i)_{\mathrm{corr_mi}}^{\mathrm{sel}}})_i$和$(\xi_{\mathrm{mi}}^{(p_i)_{\mathrm{corr_mi}}^{\mathrm{sel}}})_i$；

为消除不同输入特征的相关系数值和互信息值的大小导致的差异性，按如下公式进行标准化处理；

$$(\zeta_{\mathrm{corr\_norm}}^{(p_i)_{\mathrm{corr\_mi}}^{\mathrm{sel}}})_i=\frac{(\zeta_{\mathrm{corr}}^{(p_i)_{\mathrm{corr\_mi}}^{\mathrm{sel}}})_i}{\sum_{(p_i)_{\mathrm{corr\_mi}}^{\mathrm{sel}}=1}^{(P_i)_{\mathrm{corr\_mi}}^{\mathrm{sel}}}(\zeta_{\mathrm{corr}}^{(p_i)_{\mathrm{corr\_mi}}^{\mathrm{sel}}})_i}\quad(16)$$

$$(\zeta_{\mathrm{mi\_norm}}^{(p_i)_{\mathrm{corr\_mi}}^{\mathrm{sel}}})_i=\frac{(\zeta_{\mathrm{mi}}^{(p_i)_{\mathrm{corr\_mi}}^{\mathrm{sel}}})_i}{\sum_{(p_i)_{\mathrm{corr\_mi}}^{\mathrm{sel}}=1}^{(P_i)_{\mathrm{corr\_mi}}^{\mathrm{sel}}}(\zeta_{\mathrm{mi}}^{(p_i)_{\mathrm{corr\_mi}}^{\mathrm{sel}}})_i}\quad(17)$$

其中，$(\zeta_{\mathrm{corr_norm}}^{p_{\mathrm{corr_mi}}^{\mathrm{sel}}})_i$和$(\zeta_{\mathrm{mi_norm}}^{p_{\mathrm{corr_mi}}^{\mathrm{sel}}})_i$表示第ith个子系统的第$p_{\mathrm{corr_mi}}^{\mathrm{sel}}$th个标准化的相关系数值和互信息值；

新定义一种候选输入特征的综合评价值$\zeta_i^{(p_i)_{\mathrm{corr_mi}}^{\mathrm{sel}}}$，其表示形式为，

$$\zeta_{\mathrm{corr\_mi}}^{(p_i)_{\mathrm{corr\_mi}}^{\mathrm{sel}}}=k_i^{\mathrm{corr}}\cdot\zeta_{\mathrm{corr\_norm}}^{(p_i)_{\mathrm{corr\_mi}}^{\mathrm{sel}}}+k_i^{\mathrm{mi}}\cdot\zeta_{\mathrm{mi\_norm}}^{(p_i)_{\mathrm{corr\_mi}}^{\mathrm{sel}}}\quad(18)$$

其中，k_i^{corr}和k_i^{mi}表示比例系数，均取值为0.5，其满足$k_i^{\mathrm{corr}}+k_i^{\mathrm{mi}}=1$；

重复上述过程，获得全部候选输入特征的综合评价值并记为 $\{\zeta_{\text{corr\_mi}}^{(p_i)_{\text{corr\_mi}}^{\text{sel}}}\}_{(p_i)_{\text{corr\_mi}}^{\text{sel}}=1}^{(P_i)_{\text{corr\_mi}}^{\text{sel}}}$；

设定第ith个子系统的权重因子$f_i^{\text{corr_mi}}$，将基于综合评价值选择输入特征的阈值θ_i^{1stsel}采用下式计算，

$$\theta_i^{\text{1stsel}} = f_i^{\text{corr\_mi}} \cdot \frac{1}{(P_i)_{\text{corr\_mi}}^{\text{sel}}} \sum_{(p_i)_{\text{corr\_mi}}^{\text{sel}}=1}^{(P_i)_{\text{corr\_mi}}^{\text{sel}}} (\zeta_{\text{corr\_mi}}^{(p_i)_{\text{corr\_mi}}^{\text{sel}}})_i \qquad (19)$$

其中，$f_i^{\text{corr_mi}}$的最大$(f_i^{\text{corr_mi}})_{\max}$和最小值$(f_i^{\text{corr_mi}})_{\min}$采用如下公式计算，

$$\begin{cases} (f_i^{\text{corr\_mi}})_{\max} = \dfrac{\max\left((\zeta_{\text{corr\_mi}}^{1})_i,\cdots,(\zeta_{\text{corr\_mi}}^{(p_i)_{\text{corr\_mi}}^{\text{sel}}})_i,\cdots,(\zeta_{\text{corr\_mi}}^{(P_i)_{\text{corr\_mi}}^{\text{sel}}})_i\right)}{\dfrac{1}{(P_i)_{\text{corr\_mi}}^{\text{sel}}} \sum\limits_{(p_i)_{\text{corr\_mi}}^{\text{sel}}=1}^{(P_i)_{\text{corr\_mi}}^{\text{sel}}} (\zeta_{\text{corr\_mi}}^{(p_i)_{\text{corr\_mi}}^{\text{sel}}})_i} \\ (f_i^{\text{corr\_mi}})_{\min} = \dfrac{\min\left((\zeta_{\text{corr\_mi}}^{1})_i,\cdots,(\zeta_{\text{corr\_mi}}^{(p_i)_{\text{corr\_mi}}^{\text{sel}}})_i,\cdots,(\zeta_{\text{corr\_mi}}^{(P_i)_{\text{corr\_mi}}^{\text{sel}}})_i\right)}{\dfrac{1}{(P_i)_{\text{corr\_mi}}^{\text{sel}}} \sum\limits_{(p_i)_{\text{corr\_mi}}^{\text{sel}}=1}^{(P_i)_{\text{corr\_mi}}^{\text{sel}}} (\zeta_{\text{corr\_mi}}^{(p_i)_{\text{corr\_mi}}^{\text{sel}}})_i} \end{cases} \qquad (20)$$

以θ_i^{1stsel}作为阈值，以第ith个子系统的第$(p_i)_{\text{corr_mi}}^{\text{sel}}$th个候选输入特征为例，按如下规则进行选择，

$$\gamma^{(p_i)_{\text{corr\_mi}}^{\text{sel}}} = \begin{cases} 1, \text{ if } \zeta_{\text{corr\_mi}}^{(p_i)_{\text{corr\_mi}}^{\text{sel}}} \geq \theta_i^{\text{1stsel}} \\ 0, \text{ else } \zeta_{\text{corr\_mi}}^{(p_i)_{\text{corr\_mi}}^{\text{sel}}} < \theta_i^{\text{1stsel}} \end{cases} \qquad (21)$$

对全部的原始候选输入特征执行上述过程，选择其中$\gamma^{(p_i)_{\text{corr_mi}}^{\text{sel}}=1}$的变量作为基于综合评价值选择的输入特征，并标记为，

$$(\mathbf{X}_{\text{1st}}^{\text{sel}})_i = [(\mathbf{x}^1)_i,\cdots,(\mathbf{x}^{p_i^{\text{sel}}})_i,\cdots,(\mathbf{x}^{P_i^{\text{sel}}})_i] \qquad (22)$$

重复上述过程完成对全部子系统第1层特征的选择，并串行排列得到基于单特征相关性的第一层特征$\mathbf{X}_{\text{1st}}^{\text{sel}}$，

$$\mathbf{X}_{\text{1st}}^{\text{sel}} = [(\mathbf{X}_{\text{1st}}^{\text{sel}})_1,\cdots,(\mathbf{X}_{\text{1st}}^{\text{sel}})_i,\cdots,(\mathbf{X}_{\text{1st}}^{\text{sel}})_I] = [\mathbf{x}^{1_{\text{1st}}^{\text{sel}}},\cdots,\mathbf{x}^{p_{\text{1st}}^{\text{sel}}},\cdots,\mathbf{x}^{P_{\text{1st}}^{\text{sel}}}] \qquad (23)$$

其中，$\mathbf{x}^{p_{\text{1st}}^{\text{sel}}}$表示第1层特征选择集合中的第$p_{\text{1st}}^{\text{sel}}$th个特征，$P_{\text{1st}}^{\text{sel}} = \sum_{i=1}^{I} P_i^{\text{sel}}$表示全部第1层特征的数量；

2)基于多特征冗余性的第2层特征选择

采用基于GA-PLS的特征选择算法同时考虑多个特征间的冗余性进行第2层特征选择；采用如下所示的第2层特征选择策略；

上述策略的输入为第1层选择特征$\mathbf{X}_{\text{1st}}^{\text{sel}}$；运行第jth次GA-PLS的输出为第2层选择特征$(\mathbf{X}_{\text{2nd}}^{\text{sel}})_j$；最终输出为运行J次GA-PLS后每个第1层输入特征的被选择次数，选择预测误差小于J_{sel}次均值的第2层特征进行统计处理，其中第$p_{\text{1st}}^{\text{sel}}$th个特征的选择次数为$f_{\text{num}}^{p_{\text{1st}}^{\text{sel}}}$，相应的全部$P_{\text{1st}}^{\text{sel}}$个第1层特征为$\{f_{\text{num}}^{p_{\text{1st}}^{\text{sel}}}\}_{p_{\text{1st}}^{\text{sel}}=1}^{P_{\text{1st}}^{\text{sel}}}$；J为GA-PLS算法的运行次数，$J_{\text{sel}}$为GA-PLS模型预测误差小于J次运行均值的数量；

上述第2层特征选择的步骤如下：

第1步:设定GA-PLS运行次数J,以及GA-PLS算法参数:初始种群数量、最大遗传代数、变异概率、交叉方式、PLS算法潜在变量数量;设定j=1,启动第2层的特征选择过程,开始运行;

第2步:判断是否达到运行次数J,若满足,则转到第11步,否则,转到第3步;

第3步:采用二进制方式对特征进行编码,其中染色体的长度为输入特征个数,1表示特征被选中,0表示特征未被选中;

第4步:采用随机方式对种群初始化;

第5步:对种群进行适应度评价,采用留一法交叉验证法计算均方根验证误差RMSECV,值越小表明适应度越好;

第6步:判断是否达到最大遗传代数的终止条件,如不满足,转第7步,否则转第9步;

第7步:进行选择、交叉和变异遗传操作,其中:选择操作采用精英替代策略即采用适应度好的个体替换适应度较差的个体,交叉操作采用单点交叉,变异操作采用单点变异;

第8步:获得新种群,转到执行第5步;

第9步:获得第jth次运行GA-PLS算法的最佳个体,进一步解码得到所选择的第2层特征,并将其记为$(\mathbf{X}_{2\mathrm{nd}}^{\mathrm{sel}})_j$;

第10步:令j=j+1,转到第2步;

第11步:计算全部J次运行得到的预测模型的均方根误差RMSE的平均值,将大于此平均值的GA-PLS模型的数量标记为J_{sel};对J_{sel}次所选择的第2层特征进行处理,统计$P_{1\mathrm{st}}^{\mathrm{sel}}$个第1层特征的被选择次数,如下所示,

$$\{(\mathbf{X}_{2\mathrm{nd}}^{\mathrm{sel}})_j\}_{j=1}^{J_{\mathrm{sel}}} \Rightarrow \{f_{\mathrm{num}}^{1_{1\mathrm{st}}^{\mathrm{sel}}},\cdots,f_{\mathrm{num}}^{p_{1\mathrm{st}}^{\mathrm{sel}}},\cdots,f_{\mathrm{num}}^{P_{1\mathrm{st}}^{\mathrm{sel}}}\}=\{f_{\mathrm{num}}^{p_{1\mathrm{st}}^{\mathrm{sel}}}\}_{p_{1\mathrm{st}}^{\mathrm{sel}}=1}^{P_{1\mathrm{st}}^{\mathrm{sel}}},\quad 1\le f_{\mathrm{num}}^{p_{1\mathrm{st}}^{\mathrm{sel}}}\le J_{\mathrm{sel}} \tag{24}$$

其中,$f_{\mathrm{num}}^{p_{1\mathrm{st}}^{\mathrm{sel}}}$为第$p_{1\mathrm{st}}^{\mathrm{sel}}$th个第1层特征的被选择次数;

3)基于模型预测性能的第3层特征选择与建模

基于上述步骤得到的全部$P_{1\mathrm{st}}^{\mathrm{sel}}$个第1层特征的被选择次数为$\{f_{\mathrm{num}}^{p_{1\mathrm{st}}^{\mathrm{sel}}}\}_{p_{1\mathrm{st}}^{\mathrm{sel}}=1}^{P_{1\mathrm{st}}^{\mathrm{sel}}}$,设定比例系数$f_{\mathrm{DXN}}^{\mathrm{RMSE}}$,确定用于第3层特征选择的阈值下限$\theta_{\mathrm{DXN}}^{\mathrm{downlimit}}$,采用如下公式计算;

$$\theta_{\mathrm{DXN}}^{\mathrm{downlimit}}=floor\left(f_{\mathrm{DXN}}^{\mathrm{RMSE}}\cdot\frac{1}{P_{1\mathrm{st}}^{\mathrm{sel}}}\sum_{p_{1\mathrm{st}}^{\mathrm{sel}}=1}^{P_{1\mathrm{st}}^{\mathrm{sel}}}f_{\mathrm{num}}^{p_{1\mathrm{st}}^{\mathrm{sel}}}\right) \tag{25}$$

其中,floor(•)表示取整函数;$f_{\mathrm{DXN}}^{\mathrm{RMSE}}$的最大值$(f_{\mathrm{DXN}}^{\mathrm{RMSE}})_{\max}$和最小值$(f_{\mathrm{DXN}}^{\mathrm{RMSE}})_{\min}$采用如下公式计算,

$$\begin{cases}(f_{\mathrm{DXN}}^{\mathrm{RMSE}})_{\max}=\dfrac{\max\left(f_{\mathrm{num}}^{1_{1\mathrm{st}}^{\mathrm{sel}}},\cdots,f_{\mathrm{num}}^{p_{1\mathrm{st}}^{\mathrm{sel}}},\cdots,f_{\mathrm{num}}^{P_{1\mathrm{st}}^{\mathrm{sel}}}\right)}{\dfrac{1}{P_{1\mathrm{st}}^{\mathrm{sel}}}\displaystyle\sum_{p_{1\mathrm{st}}^{\mathrm{sel}}=1}^{P_{1\mathrm{st}}^{\mathrm{sel}}}f_{\mathrm{num}}^{p_{1\mathrm{st}}^{\mathrm{sel}}}}\\ (f_{\mathrm{DXN}}^{\mathrm{RMSE}})_{\min}=\dfrac{\min\left(f_{\mathrm{num}}^{1_{1\mathrm{st}}^{\mathrm{sel}}},\cdots,f_{\mathrm{num}}^{p_{1\mathrm{st}}^{\mathrm{sel}}},\cdots,f_{\mathrm{num}}^{P_{1\mathrm{st}}^{\mathrm{sel}}}\right)}{\dfrac{1}{P_{1\mathrm{st}}^{\mathrm{sel}}}\displaystyle\sum_{p_{1\mathrm{st}}^{\mathrm{sel}}=1}^{P_{1\mathrm{st}}^{\mathrm{sel}}}f_{\mathrm{num}}^{p_{1\mathrm{st}}^{\mathrm{sel}}}}\end{cases} \tag{26}$$

第3层特征选择的阈值上限$\theta_{\mathrm{DXN}}^{\mathrm{uplimit}}$取为全部$P_{1\mathrm{st}}^{\mathrm{sel}}$个第1层特征被选择次数的最大值,

$$\theta_{\mathrm{DXN}}^{\mathrm{uplimit}}=\max\left(f_{\mathrm{num}}^{1_{\mathrm{st}}^{\mathrm{sel}}},\cdots,f_{\mathrm{num}}^{p_{1\mathrm{st}}^{\mathrm{sel}}},\cdots,f_{\mathrm{num}}^{P_{1\mathrm{st}}^{\mathrm{sel}}}\right)\qquad(27)$$

将第3层特征选择的阈值记为$\theta_{\mathrm{DXN}}^{3\mathrm{rd}}$，其值在$\theta_{\mathrm{DXN}}^{\mathrm{downlimit}}$和$\theta_{\mathrm{DXN}}^{\mathrm{uplimit}}$之间；第3层特征的筛选机制为：

$$\mu^{p}=\begin{cases}1\ ,\ \mathrm{if}\quad f_{\mathrm{num}}^{p_{1\mathrm{st}}^{\mathrm{sel}}}\geq\theta_{\mathrm{DXN}}^{3\mathrm{rd}}\\ 0\ ,\ else\ f_{\mathrm{num}}^{p_{1\mathrm{st}}^{\mathrm{sel}}}<\theta_{\mathrm{DXN}}^{3\mathrm{rd}}\end{cases}\qquad(28)$$

其中，$f_{\mathrm{num}}^{p_{1\mathrm{st}}^{\mathrm{sel}}}$表示第$p_{1\mathrm{st}}^{\mathrm{sel}}\mathrm{th}$个第1层特征经J次GA-PLS算法被选择的次数；$\mu^{p}$表示第3层特征选择的阈值筛选标准；选择$\mu^{p}=1$的特征变量依次存入$\mathbf{X}_{3\mathrm{rd}}^{\mathrm{sel_temp}}$中；接着，以$\mathbf{X}_{3\mathrm{rd}}^{\mathrm{sel_temp}}$为输入构建基于PLS的DXN软测量模型，并计算RMSE；

在$\theta_{\mathrm{DXN}}^{\mathrm{downlimit}}$和$\theta_{\mathrm{DXN}}^{\mathrm{uplimit}}$之间逐个增加$\theta_{\mathrm{DXN}}^{3\mathrm{rd}}$值，构建基于PLS算法的DXN软测量模型，选择RMSE最小的作为基于数据驱动选择过程变量的基于PLS的DXN排放浓度软测量模型；

检查上述数据驱动软测量模型的输入中是否包括烟囱排放的CO浓度、HCL浓度、O_2浓度和NOx浓度，同时去除公用传输系统中的特征；若未包括，则将上述特征进行补选，进而获得第3层的选择特征$\mathbf{X}_{3\mathrm{rd}}^{\mathrm{sel}}$；构建基于PLS的DXN软测量模型。

2.根据权利要求1所述的方法，其特征在于：变量维数287维。

3.根据权利要求1所述的方法，其特征在于：f_i^{corr}、f_i^{mi}和$f_i^{\mathrm{corr_mi}}$均为0.8，k_i^{corr}和k_i^{mi}均取为0.5。

基于多层特征选择的固废焚烧过程二噁英排放浓度软测量方法

技术领域

[0001] 本发明属于软测量领域。

背景技术

[0002] 基于炉排炉的城市固废焚烧(MSWI)技术是目前应用最广泛的生活垃圾资源化处理手段[1,2]。截止2017年,我国大陆已有MSWI电厂303座,其中220座采用基于炉排炉的焚烧技术。目前引进的国外MSWI过程多处于手动控制运行状态,难以保持稳定运行与进行优化控制[3]。针对发展中国家,当前最为紧迫的问题是控制MSWI所造成的污染物排放[4,5],其中排放物二噁英(DXN)是目前世界上毒性最强的污染物[6],也是造成焚烧建厂存在“邻避效应”的主要原因。DXN是多氯代二苯并对二噁英(PCDDs)、多氯代二苯并呋喃(PCDFs)以及部分具有类二噁英性质的多氯联苯所组成的持久性有机污染物的统称,被称为“世纪之毒”[7],其在生物体内具有显著的积累和放大效应[8,9]。

[0003] 当前,MSWI焚烧企业的主要关注点是如何通过优化控制运行参数实现DXN排放的最小化[10]。为优化固废焚烧过程,实现DXN排放浓度的在线测量非常必要。现阶段主要检测方法包括离线直接检测法、指示物/关联物在线间接检测法和软测量方法,其中:第1种检测方法需要专门实验室和相应化验分析设备,滞后时间尺度为月/周;第2种检测方法分为在线采集烟气、检测指示物/关联物浓度和基于映射模型间接计算DXN排放浓度等3个步骤,需要昂贵复杂的在线化验分析设备,滞后时间尺度为天/小时;第3种检测方法不需要化验分析设备,滞后时间尺度为分钟/秒[11]。本文主要关注第3种方法的研究。

[0004] DXN排放浓度软测量的已有研究包括:依据机理和经验选择的输入特征,文献[12,13,14]采用数十年前欧美研究机构针对不同类型焚烧炉采集的小样本数据,基于线性回归、人工神经网络(ANN)、选择性集成(SEN)最小二乘-支持向量机(LS-SVM)等方法构建模型;文献[15]选用台湾某焚烧厂4年多的实际过程数据,综合相关性分析、主成分分析(PCA)和人工神经网络(ANN)等算法,从23个易检测过程变量中选择13个为输入构建DXN软测量模型,指出贡献率较大的输入特征为活性炭注入频率、烟囱排放HCL气体浓度和混合室温度;文献[16]以炉膛温度、锅炉出口烟温、烟气流量、SO_2浓度、HCl浓度及颗粒物浓度为输入变量构建基于支持向量机(SVM)的DXN排放浓度与毒性当量预测模型。实际MSWI过程的变量有数百维,这些变量在不同程度上均与DXN的生成、吸收与排放相关[17]。上述过程均未结合MSWI过程的多工序特性和变量间的共线性进行特征选择。此外,DXN软测量的标记样本难以获得,建模中应重点考虑小样本高维数据的特征选择问题。

[0005] 特征选择的本质就是去除原始数据中的“无关特征”与“冗余特征”,保留重要特征。从消除“无关特征”的视角,应考虑MWSI过程中的单个特征(自变量)和DXN排放浓度(因变量)间的相关程度。文献[18]对高维数据利用相关系数进行维数约简,缩短运算时间和建模复杂度。文献[19]提出基于相关系数的多目标半监督特征选择方法。但研究表明,基于相关系数的线性方法难以描述自变量与因变量间的复杂任意映射关系[20]。文献[21]指出互

信息对特征间的相关性具有良好的表征能力。文献[22]提出基于个体最佳互信息的特征选择方法。文献[23]提出基于条件互信息的特征选择方法，能够有效地对上一步所选择的特征进行评价。由此可知，相关系数与互信息均可以表征自变量和因变量间的相关性[24,25]；前者的着重点在线性关系，后者的着重点在非线性关系[26,27]。针对实际的复杂工业过程，自变量和因变量间的映射关系难以采用单一的线性或非线性进行统一表征。上述这些方法均未考虑如何进行特征的自适应选择。

[0006] 在获得与DXN具有较好相关性单输入特征的基础上，从消除“冗余特征”的视角，主要考虑MSWI过程众多过程变量间的冗余性。文献[28]采用相关系数表示已选特征与当前特征之间的冗余性。文献[29]提出PCA解决变量间的共线性问题，但所提取的潜在变量会破坏原始特征自身物理含义。文献[30]提出改进岭回归方法的回归系数为有偏估计量从而处理多重共线性问题。文献[31]验证了偏最小二乘(PLS)对输入特征间的多重共线性问题有良好的解释与分解能力。文献[32]提出了结合遗传算法(GA)全局优化搜索能力和PLS多重共线性处理能力的特征选择方法，即遗传-偏最小二乘算法(GA-PLS)。汤等人的研究表明，GA-PLS对高维谱数据具有良好的选择性[33]，但在面对小样本高维数据时，GA的随机性导致其每次特征选择的结果存在着差异性，有必要对多次选择的特征进行统计，以提高鲁棒性和可解释性。

[0007] 本文进行特征选择的目标是提高软测量模型的预测性能和可解释性。此外，上述特征选择过程主要从数据驱动视角出发，样本数量有限时可能存在偏差。依据已有的研究成果和先验知识，我们需要扩充机理含义明确的重要特征，使得软测量模型更具可解释性并且符合焚烧过程DXN排放特性，进而为后续的优化控制研究提供支撑。

[0008] 综上，本文提出基于多层特征选择的MSWI过程DXN排放浓度软测量方法。首先，从单特征与DXN相关性视角，结合相关系数和互信息构建综合评价值指标，实现MSWI多个子系统过程变量的第1层特征选择；接着，从多特征冗余性和特征选择鲁棒性视角，多次运行基于GA-PLS的特征选择算法，实现第2层特征选择；最后，结合上层选择特征的统计频次、模型预测性能及机理知识进行第3层特征选择，构建得到DXN排放浓度软测量模型。结合某焚烧厂的多年DXN检测数据验证了所提方法的有效性。

[0009] 国内某厂的炉排炉焚烧工艺流程如图1所示：

[0010] 由图1可知，MSW由专用的运输车收集后运至卸料大厅，倾倒至密封的存放池内；由人工操控的吊斗将MSW放入焚烧炉进料斗内，给料机将其推至炉排炉；在焚烧炉的炉排内依次经历干燥、点燃、燃烧和烧尽四个阶段，其中：燃尽后的残渣掉入水冷渣斗内，再由输渣机将其推入炉渣池内，收集后送至填埋场处理；焚烧产生的烟气由废热锅炉转换为高压蒸汽并推动汽轮机组发电，锅炉出口的待处理烟气进入脱酸反应器进行中和反应，并在反应器入口处添加石灰和活性炭以吸附其中的DXN和重金属，其中：飞灰进入飞灰储仓，烟气进入布袋除尘器；烟气在袋式除尘器中被除去烟气颗粒物、中和反应物和活性炭吸附物，处理后分为三个部分，其中：尾部飞灰进入灰仓后再运走进行无害化处理，部分烟灰混合物在混合器中加水后重新进入脱酸反应器，尾部烟气则由引风机经烟囱排入大气，排放的尾气中含有HCL、SO_2、NOx、HF和DXN等质。

[0011] 由上述过程产生的DXN形态包括焚烧灰、飞灰和排放气体3种，其中：焚烧灰量最大但DXN浓度较低、飞灰量稍小但DXN浓度较焚烧灰高，该两类需进行特殊处理；排放气体中的

DXN浓度为最高，包括垃圾不完全燃烧和新规合成反应生成(de novo syhthesis)两类[34]。为保证有毒有机物的有效分解，焚烧烟气应达到至少850℃并保持至少2秒。在烟气处理阶段，石灰和活性炭被喷射进入反应器用以移除酸性气体和吸附DXN以及某些重金属，再经袋式过滤器过滤后通过引风机排入烟囱；此外，该阶段存在的DXN记忆效应会导致排放浓度增加。通常，上述炉内焚烧和烟气处理阶段中与DXN产生和吸收相关的过程变量以秒为周期由现场分布式控制系统进行存储。排放烟气中的易检测气体(CO、HCL、SO2、NOx和HF等)浓度通过在线检测仪表实时检测。焚烧企业或环保部门通常以月或季为周期采用离线直接化验法对排放烟气进行DXN浓度检测。

[0012] 综上可知，DXN排放浓度软测量存在的难点包括：MSW的原始DXN含量未知、DXN生成和吸收阶段的机理复杂不清、烟气处理阶段DXN存在的记忆效应导致测量存在不确定性等。因此，非常有必要对MSWI过程的输入特征进行分区域的特征选择。

发明内容

[0013] 结合焚烧工艺将MSWI过程分为6个子系统：燃烧处理工程、锅炉设备工程、尾气处理工程、蒸汽发电工程、烟囱排放工程、公用传输工程。

[0014] 本文中，软测量模型的输入数据$X\in R^{N\times P}$包括N个样本(行)和P个变量(列)，其源于MSWI流程的不同子系统。此处，将来自第i个子系统的输入数据表示为$\boldsymbol{X}_i\in R^{N\times P_i}$，即存在如下关系，

[0015] $$\boldsymbol{X}=[\boldsymbol{X}_1,\cdots,\boldsymbol{X}_i,\cdots,\boldsymbol{X}_I]=\{\boldsymbol{X}_i\}_{i=1}^{I} \quad (1)$$

[0016] $$P=P_1+\cdots+P_i+\cdots+P_I=\sum_{i=1}^{I}P_i \quad (2)$$

[0017] 其中，I表示子系统个数，P_i表示第i个子系统包含的输入特征个数。

[0018] 相应的，输出数据$\boldsymbol{y}=\{y_n\}_{n=1}^{N}\in\boldsymbol{R}^{N\times 1}$包括N个样本(行)，其来源于采用离线直接检测法得到DXN检测样本。

[0019] 显然，模型的输入/输出数据在时间尺度上具有较大的差异性：过程变量以秒为单位在DCS系统采集与存储，DXN排放浓度以月/季为周期离线直接化验获得，故存在N＜＜P。

[0020] 为便于后文描述和理解，将X_i改写为如下形式，

[0021] $$\begin{aligned}\mathbf{X}_i&=[\{(x_n^1)_i\}_{n=1}^{N},\cdots,\{(x_n^{p_i})_i\}_{n=1}^{N},\cdots,\{(x_n^{P_i})_i\}_{n=1}^{N}]\\&=[(\mathbf{x}^1)_i,\cdots,(\mathbf{x}^{p_i})_i,\cdots,(\mathbf{x}^{P_i})_i]\\&=\{(\mathbf{x}^{p_i})_i\}_{p_i=1}^{P_i}\end{aligned} \quad (3)$$

[0022] 其中，$(\mathbf{x}^{p_i})_i$表示第i个子系统的第$p_i$th个输入特征，$\mathbf{x}^{p_i}=\{\mathbf{x}_n^{p_i}\}_{n=1}^{N}$表示列向量。

[0023] 本文提出基于多层特征选择的MSWI过程DXN排放浓度软测量策略，如图2所示。

[0024] 在图2中，$(\mathbf{X}_{\text{corr}}^{\text{sel}})_i$和$(\mathbf{X}_{\text{mi}}^{\text{sel}})_i$表示针对第i个子系统的输入特征采用相关系数和互信息度量所选择的候选特征集合，$(\mathbf{X}_{\text{lst}}^{\text{sel}})_i$表示对基于相关系数法和互信息法所选择的候选特征集合采用综合评价值度量所选择的对第i个子系统的第1层特征，$\mathbf{X}_{\text{lst}}^{\text{sel}}$表示串行组合全部子系统的第1层特征所得到的基于单特征相关性的第1层特征，$(\mathbf{X}_{\text{2nd}}^{\text{sel}})_j$表示运行第j次GA-PLS算法所选择的基于多特征冗余性的第2层特征，$f_{\text{num}}^{p_{\text{lst}}^{\text{sel}}}$表示第1层特征中第$p_{\text{lst}}^{\text{sel}}$th个特

征被选择的次数，$\mathbf{X}_{3rd}^{sel}$表示依据特征选择阈值θ_{3rd}和先验知识从$\mathbf{X}_{1st}^{sel}$中所选择的第3层特征，M_{para}表示软测量模型的参数，$\hat{\boldsymbol{y}}$表示预测值。

[0025] 算法实现

[0026] 基于单特征相关性的第1层特征选择

[0027] 基于相关系数的单特征相关性度量

[0028] 首先，计算不同原始输入特征与DXN排放浓度间的原始相关系数。此处以第ith个子系统的第pth个输入特征$(\mathbf{x}^{p_i})_i=\{(x_n^{p_i})_i\}_{n=1}^N$为例进行描述，如下，

[0029] $$(\xi_{corr\_ori}^{p_i})_i=\frac{\sum_{n=1}^{N}\left[\left((x_n^{p_i})_i-\bar{x}_{p_i}\right)(y_n-\bar{y})\right]}{\sqrt{\sum_{n=1}^{N}\left((x_n^{p_i})_i-\bar{x}_{p_i}\right)^2}\sqrt{\sum_{n=1}^{N}(y_n-\bar{y})^2}} \quad (4)$$

[0030] 其中，$\bar{x}_{p_i}$，$\bar{y}$分别表示第ith个子系统的第pth个输入特征及DXN排放浓度N个建模样本的平均值。将原始相关系数$(\xi_{corr_ori}^{p_i})_i$进行如下预处理，

[0031] $$(\xi_{corr}^{p_i})_i=\left|(\xi_{corr\_ori}^{p_i})_i\right| \quad (5)$$

[0032] 其中，|·|表示取绝对值。

[0033] 重复上述过程，获得全部原始输入特征的相关系数并记为$\{\xi_{corr}^{p_i}\}_{p_i=1}^{P_i}$。设定第ith个子系统的权重因子$f_i^{corr}$，将基于相关系数选择输入特征的阈值$\theta_i^{corr}$采用如下公式计算，

[0034] $$\theta_i^{corr}=f_i^{corr}\cdot\frac{1}{p_i}\sum_{p_i=1}^{P_i}(\xi_{corr}^{p_i})_i \quad (6)$$

[0035] 其中，f_i^{corr}的最大$(f_i^{corr})_{max}$和最小值$(f_i^{corr})_{min}$采用如下公式计算，

[0036] $$\begin{cases}(f_i^{corr})_{max}=\dfrac{\max\left((\xi_{corr}^{1})_i,\cdots,(\xi_{corr}^{p_i})_i,\cdots,(\xi_{corr}^{P_i})_i\right)}{\frac{1}{p_i}\sum_{p_i=1}^{P_i}(\xi_{corr}^{p_i})_i}\\(f_i^{corr})_{min}=\dfrac{\min\left((\xi_{corr}^{1})_i,\cdots,(\xi_{corr}^{p_i})_i,\cdots,(\xi_{corr}^{P_i})_i\right)}{\frac{1}{p_i}\sum_{p_i=1}^{P_i}(\xi_{corr}^{p_i})_i}\end{cases} \quad (7)$$

[0037] 其中，max(·)和min(·)分别表示取最大和最小值的函数。

[0038] 以θ_i^{corr}作为阈值，第ith个子系统的第$p_i$th输入特征的选择准则如下所示，

[0039] $$\alpha_i^{p_i}=\begin{cases}1,\ \text{if}\quad (\xi_{corr}^{p_i})_i\geq\theta_i^{corr}\\0,\ \text{else}\ (\xi_{corr}^{p_i})_i<\theta_i^{corr}\end{cases} \quad (8)$$

[0040] 选择其中$\alpha_i^{p_i}=1$的特征$(\mathbf{x}^{p_i})_i$作为基于相关系数选择的候选特征并将其标记为$(\mathbf{x}^{(p_i)_{corr}^{sel}})_i$。对第ith个子系统的全部原始输入特征执行上述过程，并将所选择的候选特征标记为，

[0041] $$(\mathbf{X}_{corr}^{sel})_i=[(\mathbf{x}^{1})_i,\cdots,(\mathbf{x}^{(p_i)_{corr}^{sel}})_i,\cdots,(\mathbf{x}^{(P_i)_{corr}^{sel}})_i] \quad (9)$$

[0042] 其中，$(P_i)_{corr}^{sel}$表示基于相关系数选择的第ith个子系统的过程变量个数。

[0043] 对全部子系统重复上述过程，基于相关系数度量选择的特征标记为$\{(\mathbf{X}_{corr}^{sel})_i\}_{i=1}^{I}$。

[0044] 基于互信息的单特征相关性度量

[0045] 首先，计算不同原始输入特征与DXN排放浓度间的互信息值。以第ith个子系统的第pth个输入特征$(\mathbf{x}^{p_i})_i$为例，如下：

[0046] $$(\xi_{\mathrm{mi}}^{p_i})_i=\sum_{n=1}^{N}\sum_{n=1}^{N}\left\{p_{\mathrm{rob}}\left((x_n^{p_i})_i,y_n\right)\log\left(\frac{p_{\mathrm{rob}}\left((x_n^{p_i})_i,y_n\right)}{p_{\mathrm{rob}}\left((x_n^{p_i})_i\right)p_{\mathrm{rob}}\left(y_n\right)}\right)\right\} \quad (10)$$

[0047] 其中，$p_{\mathrm{rob}}\left((x_n^{p_i})_i,y_n\right)$表示联合概率密度，$p_{\mathrm{rob}}\left((x_n^{p_i})_i\right)$和$p_{\mathrm{rob}}(y_n)$表示边际概率密度。

[0048] 重复上述过程，获得全部原始输入特征的互信息值并记为$\{\xi_{\mathrm{mi}}^{p_i}\}_{p_i=1}^{P_i}$。设定第ith个子系统的权重因子$f_i^{\mathrm{mi}}$，，将基于互信息选择输入特征的阈值$\theta_i^{\mathrm{mi}}$采用如下公式计算，

[0049] $$\theta_i^{\mathrm{mi}}=f_i^{\mathrm{mi}}\cdot\frac{1}{p_i}\sum_{p_i=1}^{P_i}(\xi_{\mathrm{mi}}^{p_i})_i \quad (11)$$

[0050] 其中，f_i^{mi}的最大$(f_i^{\mathrm{mi}})_{\max}$和最小值$(f_i^{\mathrm{mi}})_{\min}$采用如下公式计算，

[0051] $$\begin{cases}(f_i^{\mathrm{mi}})_{\max}=\dfrac{\max\left((\xi_{\mathrm{mi}}^{1})_i,\cdots,(\xi_{\mathrm{mi}}^{p_i})_i,\cdots,(\xi_{\mathrm{mi}}^{P_i})_i\right)}{\dfrac{1}{p_i}\sum_{p_i=1}^{P_i}(\xi_{\mathrm{mi}}^{p_i})_i}\\ (f_i^{\mathrm{mi}})_{\min}=\dfrac{\min\left((\xi_{\mathrm{mi}}^{1})_i,\cdots,(\xi_{\mathrm{mi}}^{p_i})_i,\cdots,(\xi_{\mathrm{mi}}^{P_i})_i\right)}{\dfrac{1}{p_i}\sum_{p_i=1}^{P_i}(\xi_{\mathrm{mi}}^{p_i})_i}\end{cases} \quad (12)$$

[0052] 以θ_i^{mi}作为阈值，第ith个系统的第$p_i$th输入特征的选择准则如下所示，

[0053] $$\beta_i^{p_i}=\begin{cases}1,\ \text{if}\quad (\xi_{\mathrm{mi}}^{p_i})_i\geq\theta_i^{\mathrm{mi}}\\ 0,\ \text{else}\ (\xi_{\mathrm{mi}}^{p_i})_i<\theta_i^{\mathrm{mi}}\end{cases} \quad (13)$$

[0054] 选择其中$\beta_i^{p_i}=1$的特征$(\mathbf{x}^{p_i})_i$作为基于互信息选择的候选特征并将其表标记为$(\mathbf{x}^{(p_i)_{\mathrm{mi}}^{\mathrm{sel}}})_i$。对第ith个子系统的全部输入特征执行上述过程，并将所选择的候选特征标记为：

[0055] $$(\mathbf{X}_{\mathrm{mi}}^{\mathrm{sel}})_i=[(\mathbf{x}^{1})_i,\cdots,(\mathbf{x}^{(p_i)_{\mathrm{mi}}^{\mathrm{sel}}})_i,\cdots,(\mathbf{x}^{(P_i)_{\mathrm{mi}}^{\mathrm{sel}}})_i] \quad (14)$$

[0056] 其中，$(P_i)_{\mathrm{mi}}^{\mathrm{sel}}$表示基于互信息选择的第ith个子系统的全部特征的个数。

[0057] 对全部子系统重复上述过程，基于互信息度量选择的特征可标记为$\{(\mathbf{X}_{\mathrm{mi}}^{\mathrm{sel}})_i\}_{i=1}^{I}$。

[0058] 基于综合评价值的单特征相关性度量

[0059] 以第ith个子系统为例，同时考虑具有相关系数和互信息贡献度的输入特征在$(\mathbf{X}_{\mathrm{mi}}^{\mathrm{sel}})_i$和$(\mathbf{X}_{\mathrm{corr}}^{\mathrm{sel}})_i$中得到候选特征集合，其策略为：

[0060] $$\begin{aligned}(\mathbf{X}_{\mathrm{corr\_mi}}^{\mathrm{sel}})_i&=(\mathbf{X}_{\mathrm{mi}}^{\mathrm{sel}})_i\cap(\mathbf{X}_{\mathrm{corr}}^{\mathrm{sel}})_i\\&=[(\mathbf{x}^{1})_i,\cdots,(\mathbf{x}^{(p_i)_{\mathrm{corr\_mi}}^{\mathrm{sel}}})_i,\cdots,(\mathbf{x}^{(P_i)_{\mathrm{corr\_mi}}^{\mathrm{sel}}})_i]\end{aligned} \quad (15)$$

[0061] 其中，∩表示取交集。$\mathbf{x}_i^{(p_i)_{\mathrm{corr_mi}}^{\mathrm{sel}}}$表示第ith个子系统的第$(p_i)_{\mathrm{corr_mi}}^{\mathrm{sel}}$th个候选特征，其对应的相关系数值与互信息值为$(\xi_{\mathrm{corr}}^{(p_i)_{\mathrm{corr_mi}}^{\mathrm{sel}}})_i$和$(\xi_{\mathrm{mi}}^{(p_i)_{\mathrm{corr_mi}}^{\mathrm{sel}}})_i$。

[0062] 为消除不同输入特征的相关系数值和互信息值的大小导致的差异性，按如下公式进行标准化处理，

[0063] $$(\zeta_{\text{corr\_norm}}^{(p_i)_{\text{corr\_mi}}^{\text{sel}}})_i=\frac{(\zeta_{\text{corr}}^{(p_i)_{\text{corr\_mi}}^{\text{sel}}})_i}{\sum_{(p_i)_{\text{corr\_mi}}^{\text{sel}}=1}^{(P_i)_{\text{corr\_mi}}^{\text{sel}}}(\zeta_{\text{corr}}^{(p_i)_{\text{corr\_mi}}^{\text{sel}}})_i} \quad (16)$$

[0064] $$(\zeta_{\text{mi\_norm}}^{(p_i)_{\text{corr\_mi}}^{\text{sel}}})_i=\frac{(\zeta_{\text{mi}}^{(p_i)_{\text{corr\_mi}}^{\text{sel}}})_i}{\sum_{(p_i)_{\text{corr\_mi}}^{\text{sel}}=1}^{(P_i)_{\text{corr\_mi}}^{\text{sel}}}(\zeta_{\text{mi}}^{(p_i)_{\text{corr\_mi}}^{\text{sel}}})_i} \quad (17)$$

[0065] 其中，$(\zeta_{\text{corr_norm}}^{p_{\text{corr_mi}}^{\text{sel}}})_i$和$(\zeta_{\text{mi_norm}}^{p_{\text{corr_mi}}^{\text{sel}}})_i$表示第ith个子系统的第$p_{\text{corr_mi}}^{\text{sel}}$th个标准化的相关系数值和互信息值。

[0066] 本文新定义一种候选输入特征的综合评价值$\zeta_i^{(p_i)_{\text{corr_mi}}^{\text{sel}}}$，其表示形式为，

[0067] $$\zeta_{\text{corr\_mi}}^{(p_i)_{\text{corr\_mi}}^{\text{sel}}}=k_i^{\text{corr}}\cdot\zeta_{\text{corr\_norm}}^{(p_i)_{\text{corr\_mi}}^{\text{sel}}}+k_i^{\text{mi}}\cdot\zeta_{\text{mi\_norm}}^{(p_i)_{\text{corr\_mi}}^{\text{sel}}} \quad (18)$$

[0068] 其中，k_i^{corr}和k_i^{mi}表示比例系数(默认取值为0.5)，其满足$k_i^{\text{corr}}+k_i^{\text{mi}}=1$。

[0069] 重复上述过程，获得全部候选输入特征的综合评价值并记为$\{\zeta_{\text{corr_mi}}^{(p_i)_{\text{corr_mi}}^{\text{sel}}}\}_{(p_i)_{\text{corr_mi}}^{\text{sel}}=1}^{(P_i)_{\text{corr_mi}}^{\text{sel}}}$。

[0070] 设定第ith个子系统的权重因子$f_i^{\text{corr_mi}}$，将基于综合评价值选择输入特征的阈值θ_i^{lstsel}采用下式计算，

[0071] $$\theta_i^{\text{lstsel}}=f_i^{\text{corr\_mi}}\cdot\frac{1}{(P_i)_{\text{corr\_mi}}^{\text{sel}}}\sum_{(p_i)_{\text{corr\_mi}}^{\text{sel}}=1}^{(P_i)_{\text{corr\_mi}}^{\text{sel}}}(\zeta_{\text{corr\_mi}}^{(p_i)_{\text{corr\_mi}}^{\text{sel}}})_i \quad (19)$$

[0072] 其中，$f_i^{\text{corr_mi}}$的最大$(f_i^{\text{corr_mi}})_{\max}$和最小值$(f_i^{\text{corr_mi}})_{\min}$采用如下公式计算，

[0073] $$\begin{cases}(f_i^{\text{corr\_mi}})_{\max}=\dfrac{\max\left((\zeta_{\text{corr\_mi}}^{1})_i,\cdots,(\zeta_{\text{corr\_mi}}^{(p_i)_{\text{corr\_mi}}^{\text{sel}}})_i,\cdots,(\zeta_{\text{corr\_mi}}^{(P_i)_{\text{corr\_mi}}^{\text{sel}}})_i\right)}{\dfrac{1}{(P_i)_{\text{corr\_mi}}^{\text{sel}}}\sum_{(p_i)_{\text{corr\_mi}}^{\text{sel}}=1}^{(P_i)_{\text{corr\_mi}}^{\text{sel}}}(\zeta_{\text{corr\_mi}}^{(p_i)_{\text{corr\_mi}}^{\text{sel}}})_i}\\(f_i^{\text{corr\_mi}})_{\min}=\dfrac{\min\left((\zeta_{\text{corr\_mi}}^{1})_i,\cdots,(\zeta_{\text{corr\_mi}}^{(p_i)_{\text{corr\_mi}}^{\text{sel}}})_i,\cdots,(\zeta_{\text{corr\_mi}}^{(P_i)_{\text{corr\_mi}}^{\text{sel}}})_i\right)}{\dfrac{1}{(P_i)_{\text{corr\_mi}}^{\text{sel}}}\sum_{(p_i)_{\text{corr\_mi}}^{\text{sel}}=1}^{(P_i)_{\text{corr\_mi}}^{\text{sel}}}(\zeta_{\text{corr\_mi}}^{(p_i)_{\text{corr\_mi}}^{\text{sel}}})_i}\end{cases} \quad (20)$$

[0074] 以θ_i^{lstsel}作为阈值，以第ith个子系统的第$(p_i)_{\text{corr_mi}}^{\text{sel}}$th个候选输入特征为例，按如下规则进行选择，

[0075] $$\gamma^{(p_i)_{\text{corr\_mi}}^{\text{sel}}}=\begin{cases}1,\ \text{if}\quad \zeta_{\text{corr\_mi}}^{(p_i)_{\text{corr\_mi}}^{\text{sel}}}\geq\theta_i^{\text{lstsel}}\\0,\ \text{else}\ \zeta_{\text{corr\_mi}}^{(p_i)_{\text{corr\_mi}}^{\text{sel}}}<\theta_i^{\text{lstsel}}\end{cases} \quad (21)$$

[0076] 对全部的原始候选输入特征执行上述过程，选择其中$\gamma^{(p_i)_{\text{corr_mi}}^{\text{sel}}}=1$的变量作为基于综合评价值选择的输入特征，并标记为，

[0077] $$(\mathbf{X}_{\text{1st}}^{\text{sel}})_i=[(\mathbf{x}^1)_i,\cdots,(\mathbf{x}^{p_i^{\text{sel}}})_i,\cdots,(\mathbf{x}^{P_i^{\text{sel}}})_i] \quad (22)$$

[0078] 重复上述过程完成对全部子系统第1层特征的选择，并串行排列可得到基于单特征相关性的第一层特征$\mathbf{X}_{\text{1st}}^{\text{sel}}$，

[0079] $$\mathbf{X}_{\text{1st}}^{\text{sel}}=[(\mathbf{X}_{\text{1st}}^{\text{sel}})_1,\cdots,(\mathbf{X}_{\text{1st}}^{\text{sel}})_i,\cdots,(\mathbf{X}_{\text{1st}}^{\text{sel}})_I]=[\mathbf{x}^{1_{\text{1st}}^{\text{sel}}},\cdots,\mathbf{x}^{p_{\text{1st}}^{\text{sel}}},\cdots,\mathbf{x}^{P_{\text{1st}}^{\text{sel}}}] \quad (23)$$

[0080] 其中，$\mathbf{x}^{p_{\text{1st}}^{\text{sel}}}$表示第1层特征选择集合中的第$p_{\text{1st}}^{\text{sel}}$th个特征，$P_{\text{1st}}^{\text{sel}}=\sum_{i=1}^{I}P_i^{\text{sel}}$表示全部第1层

特征的数量。

[0081] 基于多特征冗余性的第2层特征选择

[0082] 上述第1层特征的选择过程仅考虑单输入特征与DXN排放浓度之间的相关性，未考虑多特征间存在的冗余性。此处采用基于GA-PLS的特征选择算法同时考虑多个特征间的冗余性进行第2层特征选择。考虑到DXN排放浓度建模的小样本特点和GA算法的随机性，此处采用如下所示的第2层特征选择策略。

[0083] 由图3可知：上述策略的输入为第1层选择特征$\mathbf{X}_{1st}^{sel}$；运行第jth次GA-PLS的输出为第2层选择特征$(\mathbf{X}_{2nd}^{sel})_j$；最终输出为运行J次GA-PLS后每个第1层输入特征的被选择次数，选择预测误差小于均值的J_{sel}次第2层特征进行统计处理，其中第p_{1st}^{sel}th个特征的选择次数为$f_{num}^{p_{1st}^{sel}}$，相应的全部P_{1st}^{sel}个第1层特征为$\{f_{num}^{p_{1st}^{sel}}\}_{p_{1st}^{sel}=1}^{P_{1st}^{sel}}$；J为GA-PLS算法的运行次数，一般取值为100次以上；J_{sel}为GA-PLS模型预测误差小于J次运行均值的数量。

[0084] 上述第2层特征选择的步骤如下：

[0085] 第1步：设定GA-PLS运行次数J，以及GA-PLS算法参数：初始种群数量、最大遗传代数、变异概率、交叉方式、PLS算法潜在变量(LV)数量，一般设定为6；设定j＝1，启动第2层的特征选择过程，开始运行。

[0086] 第2步：判断是否达到运行次数J，若满足，则转到第11步，否则，转到第3步；

[0087] 第3步：采用二进制方式对特征进行编码，其中染色体的长度为输入特征个数，1表示特征被选中，0表示特征未被选中；

[0088] 第4步：采用随机方式对种群初始化；

[0089] 第5步：对种群进行适应度评价，采用留一法交叉验证法计算均方根验证误差RMSECV，值越小表明适应度越好；

[0090] 第6步：判断是否达到最大遗传代数的终止条件，如不满足，转第7步，否则转第9步；

[0091] 第7步：进行选择、交叉和变异遗传操作，其中：选择操作采用精英替代策略即采用适应度好的个体替换适应度较差的个体，交叉操作采用单点交叉，变异操作采用单点变异；

[0092] 第8步：获得新种群，转到执行第5步；

[0093] 第9步：获得第jth次运行GA-PLS算法的最佳个体，进一步解码得到所选择的第2层特征，并将其记为$(\mathbf{X}_{2nd}^{sel})_j$；

[0094] 第10步：令j＝j+1，转到第2步；

[0095] 第11步：计算全部J次运行得到的预测模型的均方根误差(RMSE)的平均值，将大于此平均值的GA-PLS模型的数量标记为J_{sel}。对J_{sel}次所选择的第2层特征进行处理，统计P_{1st}^{sel}个第1层特征的被选择次数，如下所示，

[0096] $$\{(\mathbf{X}_{2nd}^{sel})_j\}_{j=1}^{J_{sel}} \Rightarrow \{f_{num}^{1_{1st}^{sel}},\cdots,f_{num}^{p_{1st}^{sel}},\cdots,f_{num}^{P_{1st}^{sel}}\} = \{f_{num}^{p_{1st}^{sel}}\}_{p_{1st}^{sel}=1}^{P_{1st}^{sel}}, \quad 1 \le f_{num}^{p_{1st}^{sel}} \le J_{sel} \quad (24)$$

[0097] 其中，$f_{num}^{p_{1st}^{sel}}$为第p_{1st}^{sel}th个第1层特征的被选择次数。

[0098] 基于模型预测性能的第3层特征选择与建模

[0099] 基于上述步骤得到的全部P_{1st}^{sel}个第1层特征的被选择次数为$\{f_{num}^{p_{1st}^{sel}}\}_{p_{1st}^{sel}=1}^{P_{1st}^{sel}}$，结合确定的比例系数$f_{DXN}^{RMSE}$(其默认值为1)，确定用于第3层特征选择的阈值下限$\theta_{DXN}^{downlimit}$，采用如下公式计

算，

[0100] $$\theta_{\mathrm{DXN}}^{\mathrm{downlimit}}=floor\left(f_{\mathrm{DXN}}^{\mathrm{RMSE}}\cdot\frac{1}{P_{1\mathrm{st}}^{\mathrm{sel}}}\sum_{p_{1\mathrm{st}}^{\mathrm{sel}}=1}^{P_{1\mathrm{st}}^{\mathrm{sel}}}f_{\mathrm{num}}^{p_{1\mathrm{st}}^{\mathrm{sel}}}\right)\quad(25)$$

[0101] 其中，floor(•)表示取整函数；$f_{\mathrm{DXN}}^{\mathrm{RMSE}}$值取1时表示阈值下限为全部第1层特征选择次数的均值，其最大值$(f_{\mathrm{DXN}}^{\mathrm{RMSE}})_{\max}$和最小值$(f_{\mathrm{DXN}}^{\mathrm{RMSE}})_{\min}$采用如下公式计算，

[0102] $$\begin{cases}(f_{\mathrm{DXN}}^{\mathrm{RMSE}})_{\max}=\dfrac{\max\left(f_{\mathrm{num}}^{1_{1\mathrm{st}}^{\mathrm{sel}}},\cdots,f_{\mathrm{num}}^{p_{1\mathrm{st}}^{\mathrm{sel}}},\cdots,f_{\mathrm{num}}^{P_{1\mathrm{st}}^{\mathrm{sel}}}\right)}{\dfrac{1}{P_{1\mathrm{st}}^{\mathrm{sel}}}\displaystyle\sum_{p_{1\mathrm{st}}^{\mathrm{sel}}=1}^{P_{1\mathrm{st}}^{\mathrm{sel}}}f_{\mathrm{num}}^{p_{1\mathrm{st}}^{\mathrm{sel}}}}\\(f_{\mathrm{DXN}}^{\mathrm{RMSE}})_{\min}=\dfrac{\min\left(f_{\mathrm{num}}^{1_{1\mathrm{st}}^{\mathrm{sel}}},\cdots,f_{\mathrm{num}}^{p_{1\mathrm{st}}^{\mathrm{sel}}},\cdots,f_{\mathrm{num}}^{P_{1\mathrm{st}}^{\mathrm{sel}}}\right)}{\dfrac{1}{P_{1\mathrm{st}}^{\mathrm{sel}}}\displaystyle\sum_{p_{1\mathrm{st}}^{\mathrm{sel}}=1}^{P_{1\mathrm{st}}^{\mathrm{sel}}}f_{\mathrm{num}}^{p_{1\mathrm{st}}^{\mathrm{sel}}}}\end{cases}\quad(26)$$

[0103] 第3层特征选择的阈值上限$\theta_{\mathrm{DXN}}^{\mathrm{uplimit}}$取为全部$P_{1\mathrm{st}}^{\mathrm{sel}}$个第1层特征被选择次数的最大值，

[0104] $$\theta_{\mathrm{DXN}}^{\mathrm{uplimit}}=\max\left(f_{\mathrm{num}}^{1_{1\mathrm{st}}^{\mathrm{sel}}},\cdots,f_{\mathrm{num}}^{p_{1\mathrm{st}}^{\mathrm{sel}}},\cdots,f_{\mathrm{num}}^{P_{1\mathrm{st}}^{\mathrm{sel}}}\right)\quad(27)$$

[0105] 将第3层特征选择的阈值记为$\theta_{\mathrm{DXN}}^{3\mathrm{rd}}$，其值在$\theta_{\mathrm{DXN}}^{\mathrm{downlimit}}$和$\theta_{\mathrm{DXN}}^{\mathrm{uplimit}}$之间。第3层特征的筛选机制为：

[0106] $$\mu^{p}=\begin{cases}1\ ,\ \mathrm{if}\quad f_{\mathrm{num}}^{p_{1\mathrm{st}}^{\mathrm{sel}}}\geq\theta_{\mathrm{DXN}}^{3\mathrm{rd}}\\0\ ,\ else\ f_{\mathrm{num}}^{p_{1\mathrm{st}}^{\mathrm{sel}}}<\theta_{\mathrm{DXN}}^{3\mathrm{rd}}\end{cases}\quad(28)$$

[0107] 其中，$f_{\mathrm{num}}^{p_{1\mathrm{st}}^{\mathrm{sel}}}$表示第$p_{1\mathrm{st}}^{\mathrm{sel}}$th个第1层特征经J次GA-PLS算法被选择的次数；μ^{p}表示第3层特征选择的阈值筛选标准。选择$\mu^{p}=1$的特征变量依次存入$\mathbf{X}_{3\mathrm{rd}}^{\mathrm{sel_temp}}$中；接着，以$\mathbf{X}_{3\mathrm{rd}}^{\mathrm{sel_temp}}$为输入构建基于PLS的DXN软测量模型，并计算RMSE。

[0108] 进一步，在$\theta_{\mathrm{DXN}}^{\mathrm{downlimit}}$和$\theta_{\mathrm{DXN}}^{\mathrm{uplimit}}$之间逐个增加$\theta_{\mathrm{DXN}}^{3\mathrm{rd}}$值，构建基于PLS算法的DXN软测量模型，选择RMSE最小的作为基于数据驱动选择过程变量的基于PLS的DXN排放浓度软测量模型。

[0109] 进一步，进一步，检查上述数据驱动软测量模型的输入中是否包括烟囱排放的CO浓度、HCL浓度、O_2浓度和NOx浓度，同时去除公用传输系统中的特征；若未包括，则将上述特征进行补选，进而获得第3层的选择特征$\mathbf{X}_{3\mathrm{rd}}^{\mathrm{sel}}$；进一步，构建基于数据驱动与机理结合选择过程变量的基于PLS的DXN软测量模型。。

[0110] 综上可知，本文所提多层特征选择的过程可表示如下，

[0111] $$\underset{\text{第1层特征选择}}{\boldsymbol{X}\xrightarrow{\text{子系统}}\{(\boldsymbol{X}_{1\mathrm{st}}^{\mathrm{sel}})_j\}_{j=1}^{J}\xrightarrow{\text{组合}}\boldsymbol{X}_{1\mathrm{st}}^{\mathrm{sel}}}\Rightarrow\underset{\text{第2层特征选择}}{\{(\boldsymbol{X}_{2\mathrm{nd}}^{\mathrm{sel}})_j\}_{j=1}^{J}}\Rightarrow\underset{\text{第3层特征选择}}{\boldsymbol{X}_{3\mathrm{rd}}^{\mathrm{sel\_temp}}\xrightarrow{\text{机理先验}}\boldsymbol{X}_{3\mathrm{rd}}^{\mathrm{sel}}}\quad(29)$$

附图说明

[0112] 图1基于炉排炉固废焚烧工艺流程

[0113] 图2软测量策略

[0114] 图3基于多特征冗余性的第2层特征选择策略图

[0115] 图4焚烧子系统所选择的过程变量的相关系数值、互信息值和综合指标评价值

[0116] 图5锅炉子系统所选择的过程变量的相关系数值、互信息值和综合指标评价值

[0117] 图6烟气处理子系统所选择的过程变量的相关系数值、互信息值和综合指标评价值

[0118] 图7蒸汽发电子系统所选择的过程变量的相关系数值、互信息值和综合指标评价值

[0119] 图8尾气排放子系统所选择的过程变量的相关系数值、互信息值和综合指标评价值

[0120] 图9公用子系统所选择的过程变量的相关系数值、互信息值和综合指标评价值

具体实施方式

[0121] 建模数据描述

[0122] 本文建模数据源于北京某基于炉排炉的MSWI焚烧企业，包括DXN排放浓度检测样本的数量为34个，变量维数为287维(包含了MSWI过程的全部过程变量)。可见，输入特征数量远远超过建模样本数量，进行维数约简非常有必要。本文中，将焚烧、锅炉、烟气处理、蒸汽发电、烟气排放和公用工程6个子系统标记为Incinerator、Boiler、Flue gas、Steam、Stack和Common。

[0123] 建模结果

[0124] 基于单特征相关性的特征选择结果

[0125] 针对不同的子系统，取相关系数和互信息的特征选择权重因子f_i^{corr}、f_i^{mi}和$f_i^{corr_mi}$均为0.8，k_i^{corr}和k_i^{mi}均取为0.5，不同的子系统所选择的过程变量的相关系数值、互信息值和综合指标评价值如图4-9所示。

[0126] 由图4-9可知，不同子系统过程变量的相关系数，互信息值和综合指标评价值的间存在差异，其最小值、均值和最大值的统计结果如表1所示。

[0127] 表1不同子系统过程变量的相关性度量结果统计

[0128]

序号	子系统	相关系数值			互信息值			综合指标评价值		
		最小	均值	最大	最小	均值	最大	最小	均值	最大
1	锅炉	0.06305	0.1743	0.3358	0.2596	0.5861	0.8025	0.09123	0.1250	0.1568
2	焚烧	0.006888	0.2098	0.6760	0.4680	0.7254	0.8665	0.01771	0.02380	0.03661
3	烟气排放	0.001346	0.2816	0.4948	0.6811	0.7401	0.8103	0.2329	0.2500	0.2827
4	尾气处理	0.03686	0.2448	0.4756	0.4885	0.7005	0.8103	0.05765	0.07142	0.09420
5	蒸汽发电	0.01507	0.2011	0.4970	0.3003	0.6125	0.7856	0.02457	0.03448	0.04523
6	公用工程	0.8848e-4	0.1630	0.5628	0.1928	0.6014	0.8511	0.01296	0.01960	0.03331

[0129] 由表1可知：(1)子系统过程变量相关系数值、互信息值和综合指标评价值平均值的最大值均源于烟气排放子系统，分别为0.2816、0.7401和0.2500；烟气排放子系统测量的是与DXN同时排放至大气中的气体，如烟囱排放HCL浓度、烟囱排放O2浓度、烟囱排放NOx浓度、烟囱排放CO浓度等，这与DXN的产生机理和文献中关于DXN排放检测的综述是相符的[11]；(2)子系统过程变量相关系数值、互信息值和综合指标评价值最大值的最大值分别源于焚烧子系统、烟气排放子系统和烟气排放子系统，分别为0.6760、0.8665和0.2827，是与DXN生成过程相关的系统；(3)子系统过程变量相关系数值、互信息值和综合指标评价值最

小值的最小值均源于公用工程子系统，从机理上讲，该子系统与DXN产生的物质流不具备直接的联系，但从单特征相关性的度量结果可知，其包含的部分过程变量与DXN间的相关系数值和互信息值还是较大的；(4)上述统计表明了DXN排放工业数据具有一定程度的可靠性，从单特征相关性的视角，排在前3的是与DXN生成、处理和排放相关的系统；但其他子系统的部分过程变量从数据视角也与DXN排放浓度的相关性较大，故需要结合机理知识进行最终的特征选择。

[0130] 进一步，基于综合指标评价值所选择的过程变量数量如表2所示。

[0131] 表2基于综合指标评价值所选择的过程变量数量

[0132]

序号	统计项目		锅炉	燃烧	烟气排放	尾气处理	蒸汽发电	公用工程	汇总
1	原始特征数量		14	79	6	20	53	115	287
2	相关性指标	相关系数	9	44	4	14	29	58	158
		互信息	12	77	6	19	44	90	248
		综合评价值	6	39	4	14	27	42	132
3	汇总序号		6	45	49	63	90	132	--

[0133] 结合图4-9和表2可知，基于相关系数和互信息选择的特征数量并不相同；基于综合评价值选择的特征变量为132个，数量最多的子系统为焚烧(39)和公用工程(42)。此外，分别从各个子系统进行过程变量的选择保证了每个子系统均能够为下步变量选择贡献特征，也便于后续对不同子系统进行独立分析。

[0134] 基于多特征冗余性的特征选择结果

[0135] 对上述过程所选择的132个基于单特征相关性的过程变量，采用GA-PLS算法确定最佳过程变量的组合，去除冗余特征。

[0136] GA-PLS所采用的运行参数为：种群数量20、最大遗传代数40、最大LV数量6、遗传变异率0.005、窗口宽度1、收敛百分比98％、变量初始化百分比30％。

[0137] 基于上述参数运行100次，所得预测模型的RMSE统计结果如表3所示。

[0138] 表3运行100次GA-PLS的RMSE统计结果

[0139]

	最大	均值	最小	备注
训练数据	0.005726	0.001359	4.3480e-8	
测试数据	0.03110	0.02571	0.01853	

[0140] 由表3可知，从预测性能的统计结果看，GA-PLS的运行结果具有较大的波动性，这与本论文所采用的建模样本数量小和GA算法自身具有随机性相关。对大于预测均值的GA-PLS算法所获得的预测模型进行统计，可得到用于特征选择频次统计的模型数量为49个。进一步，计算132个过程变量的被选择次数，如下表所示。

[0141] 表4基于多特征选择的过程变量被选择次数统计表

[0142]

序号	子系统	变量被选择次数	变量数量
1	锅炉	{12 7 12 7 22 8}	6
2	焚烧	{13 9 13 13 9 7 18 14 9 13 23 21 3 3 10 21 **33** 9 0 10 7 11 **29** 3 11 4 8 12 5 5 7 16 11 6 9 9 12 **28** 6}	39
3	烟气排放	{2 6 0 5}	4
4	尾气处理	{12 37 8 9 8 19 17 29 4 22 9 19 10 23}	14
5	蒸汽发电	{**37** 10 11 17 18 **27** **26** 23 20 16 8 20 11 11 15 13 11 11 18 18 14 23 13 32 18 **44** 10}	27
6	公用工程	{5 12 14 21 10 **48** 27 **26** **34** 10 14 **33** **26** 11 3 1 20 8 12 15 6 2 5 2 23 18 4 8 20 17 10 1 15 16 8 1 10 7 3 2 11 **32}**	42

[0143]

[0144] 由表4可知：(1)全部132个过程变量被选择的平均次数为13次，具有最大选择次数的过程变量源于公用工程子系统；(2)具有最大单特征相关性的烟气排放子系统的4个过程变量的被选择次数最大仅为6，可见进行多特征冗余性与单特征相关性的选择结果间存在差异性，同时，也说明GA-PLS算法所存在的随机性；(3)仅是基于数据驱动的特征变量选择还是存在缺陷的，需要机理知识的补充。

[0145] 基于模型预测性能的特征选择结果

[0146] 基于上述GA-PLS的运行结果，将特征选择阈值的范围设定为13-48。。

[0147] 按照特征选择阈值与预测性能间的关系，将阈值确定为18，则所选择的过程变量数量为39个，在各个子系统中所选择的变量如表5所示。

[0148] 表5基于模型预测性能选择的过程变量统计表

序号	子系统	变量被选择次数	变量数量
1	锅炉	{'反应器入口氧气浓度'}	1/6
2	焚烧	{'燃烧炉排右空气流量' '二次空预器出口温度' '干燥炉排入口空气温度' '燃烧炉排 2-2 左内温度' '燃烧炉排 2-2 右内温度 ' '二次风机出口空气压力' '燃烬炉排左侧速度'}	7/39
3	烟气排放	{--}	0/4
4	尾气处理	{'混合器水流量 A' '布袋差压 A' '烟道入口烟气流量' 'NID 入口 O2 浓度' ' 石灰储仓给料量' '尿素溶剂供应流量'}	6/14
5	蒸汽发电	{'省煤器出口压力' '凝汽器 A 侧循环水进口温度' '凝汽器 A 侧循环水出口温度' '凝汽器 B 侧循环水进口温度' '凝汽器 B 侧循环水出口温度' '凝汽器出口温度' '1#除氧器水位' '汽机轴向轴承副推力面金属温度' '发电机前轴承轴瓦温度' ' 汽机小齿轮后轴承温度' '汽机前轴承振动' '汽机后轴承振动' '发电机前轴承振动'}	13/27
6	公用工程	{'燃油罐油温 4' '定压补水罐压力' '仪用压缩空气母管流量' '1#汽包炉水' '2#汽包炉水电导率' '雨水泵前池液位' 'NID 系统补水箱液位' '1 段抽汽母管压力' '空预器减温减压器出口压力' '旁路减温减压器出口温度' '#1 发电机 B 相电流' '0#启动/备用变压器 6kv 侧电流'}	12/42

[0149]

[0150] 由表4可知，输入特征维数降为39，与DXN产生机理相关的特征为14个(焚烧7个，尾气处理6个，锅炉1个)。采用上述基于数据驱动选择的过程变量构建PLS模型。

[0151] 依据LV数量与预测性能RMSE间的关系，当LV数量为2，其训练和测试RMSE分别为0.01375和0.01929。不同LV提取的潜在变量贡献率如表6所示。

[0152] 依据DXN产生的机理可知，焚烧发电子系统和公用工程子系统与DXN排放浓度的相关性不大，烟气排放子系统与DXN相关。此处，结合机理增加烟气排放子系统的4个过程变量(烟囱排放HCL浓度 烟囱排放O_2浓度 烟囱排放NOx浓度 烟囱排放CO浓度)作为输入特征。

[0153] 采用上述基于数据驱动与机理结合选择的18个过程变量构建PLS模型。

[0154] 依据LV数量与预测性能RMSE间的关系，当LV数量为2时，其训练和测试RMSE分别为0.01638和0.02048。不同LV提取的变量，贡献率如表6所示。

[0155] 表6基于不同输入特征PLS模型的LV贡献率

LV #	数据驱动选择过程变量				数据驱动与机理结合方式选择过程变量			
	输入数据		输出数据		输入数据		输出数据	
	单个 LV	总计	单个 LV	总计	单个 LV	总计	单个 LV	总计
1	29.62	29.62	55.18	55.18	29.23	29.23	56.00	56.00
2	26.96	56.58	21.95	77.13	28.15	57.38	11.55	67.54
3	9.97	66.55	15.90	93.04	9.68	67.05	14.26	81.81
4	7.15	73.70	3.92	96.96	7.31	74.36	6.48	88.29
5	2.60	76.31	2.06	99.01	7.50	81.86	2.37	90.65
6	7.47	83.78	0.26	99.27	4.40	86.26	1.80	92.45
7	3.70	87.48	0.22	99.49	5.14	91.39	0.59	93.04
8	2.94	90.42	0.16	99.65	3.14	94.53	0.86	93.90
9	1.51	91.93	0.20	99.85	1.65	96.18	1.85	95.75
10	2.96	94.89	0.06	99.90	1.22	97.40	1.34	97.09

[0156]

[0157] 由表6可知，加入基于机理知识确定的过程变量后，LV在输入数据中的贡献率提高了2%，在输出数据中的贡献中降低了2%，可见去除和加入的过程变量对预测性能的影响是有限的。考虑到DXN建模数据预处理中是将24小时的数据进行均值化获得，对应的DXN检测值是连续采样8-9小时在离线化验1周获得的，在处理过程中难免会引入不确定因素。同时，此处以引入较小的预测误差为代价，引入部分机理相关的过程变量是适合的。具体的更深入的机理分析需要结合DXN排放过程的数值仿真研究深入进行。

[0158] 比较与讨论

[0159] 由上文可知，所提方法能够均衡的考虑相关系数与互信息度量的贡献度。采用PLS算法建立基于上述不同输入特征的软测量模型，统计结果如表7所示。

[0160] 表7基于不同输入特征的PLS模型统计结果

[0161]

序号	方法	特征选择系数 $(f_i^{corr}, f_i^{mi}\ f_i^{corr\_mi}), (k_i^{corr}, k_i^{mi})$	输入维数	RMSE		备注 LV数量，数据集
				训练	测试	
1	PLS	--	287	0.01720	0.02004	2，全流程
2	相关系数值PLS	(0.8,--,--),(1,--)	153	0.01612	0.02015	2，全流程
3	互信息值PLS	(--,0.8,--),(1,--)	235	0.01764	0.02055	2，全流程
4	综合评价值PLS	(0.8,0.8,0.8),(0.5,0.5)	98	0.01649	0.02070	2，全流程
5	本文PLS	(0.8,0.8,0.8),(0.5,0.5)	39	0.01375	0.01929	2，数据驱动,子系统
		(0.8,0.8,0.8),(0.5,0.5)	18	0.01638	0.02048	2,数据驱动+机理,子系统

[0162] 由上述结果可知：采用相同LV数量，基于不同输入特征的PLS建模方法在测试数据的预测性能相差不大，但在输入特征的维数约简上却差距明显。输入特征维数由高低分别为：原始特征287维、基于互信息235维、基于相关系数153维、基于综合评价值98维、基于本文数据驱动为39维、基于本文数据驱动与机理混合为18维；可见本文方法在特征数量上缩减了16倍。由此可见，本文所提方法对构建物理含义清晰、可解释的软测量模型是有效的。同时也表明，对工业过程数据的分析需要结合机理知识进行。

[0163] 本文在进行特征选择时，涉及到多个特征选择系数，这些系数对特征选择结果和模型预测性能的影响还需要进一步的深入分析。此外，本文所采用的建模方法为简单的线性模型，所选择的特征为混合的线性与非线性特征，因此在更为合理的建模策略的选择上也还有待于研究。工业过程数据的可靠性如何度量也是值得深入考虑的问题。针对机理知识明晰的输入特征，需要考虑在遗传算法的初始化中利用先验知识，以保证选择具有较强机理相关性的过程变量，如烟囱排放CO浓度等。

[0164] 本文针对城市固废焚烧的剧毒副产品DXN的生成与排放机理复杂不清、难以实时在线检测，用于DXN软测量的高维输入特征难以有效选择及建模样本有限等问题，提出了基于多层特征选择的MSWI过程DXN排放浓度软测量方法，主要贡献体现在：(1)定义综合评价值指标进行基于相关性的单特征选择与度量；(2)提出多次运行GA-PLS的面向多特征冗余性的特征选择方法；(3)基于模型预测性能，结合数据驱动和机理知识选择最终输入特征构建软测量模型。结合某焚烧厂的多年DXN检测数据验证了所提方法的有效性。

[0165] 参考文献

[0166] [1]Arafat H A,Jijakli K,Ahsan A.Environmental performance and energy recovery potential of five processes for municipal solid waste treatment[J].Journal of Cleaner Production,2015,105:233-240

[0167] [2]Zhou H,Meng A,Long Y Q,Li Q H,and Zhang Y G.A review of dioxin-related substances during municipal solid waste incineration[J].Waste Management,36:106-118,2015.

[0168] [3]Mukherjee A,Debnath B,Ghosh S K.A Review on Technologies of Removal of Dioxins and Furans from Incinerator Flue Gas[J].Procedia Environmental Sciences,2016,35:528-540.

[0169] [4]Yuanan H,Hefa C,Shu T.The growing importance of waste-to-energy (WTE)incineration in Chinas anthropogenic mercury emissions:Emission inventories and reduction strategies[J].Renewable and Sustainable Energy Reviews,2018,97:119-137.

[0170] [5]Huang T,Zhou L,Liu L,Xia M.Ultrasound-enhanced electrokinetic remediation for removal of Zn,Pb,Cu and Cd in municipal solid waste incineration fly ashes[J].Waste Management,2018,75:226-235.

[0171] [6]Jones P H,Degerlache J,Marti E,Mischer G,Scherrer M C,Bontinck M J,Niessen H J,The global exposure of man to dioxins-a perspective on industrial-waste incineration[J].Chemosphere,26(1993)1491-1497.

[0172] [7]Bai J,Sun X,Zhang C,Gong C,Hu J,Zhang J.Mechanism and kinetics study on the ozonolysis reaction of 2,3,7,8-TCDD in the atmosphere[J].Journal of Environmental Sciences,2014,26(1):181-188.

[0173] [8]俞明锋,付建英,詹明秀.生活废弃物焚烧处置烟气中二噁英排放特性研究[J].环境科学学报,2018,38(05):1983-1988.(Yu Ming-Feng,Fu Jian-Yinig,Zhan Ming-Xiu.The research of PCDD/Fs emission characteristics in flue gas from municipal solid waste incinerations[J].Acta Scientiae Circumstantiae,2018,38(05):1983-1988.)

[0174] [9]Gouin T,Daly T H L,Wania F,Mackay D,Jones K C.Variability of concentrations of polybrominated diphenyl ethers and polychlorinated biphenyls in air:implications for monitoring,modeling and control[J].Atmospheric Environment,2005,39(1):151-166.

[0175] [10]Zhang H J,Ni Y W,Chen J P,Zhang Q.Influence of variation in the operating conditions on PCDD/F distribution in a full-scale MSW incinerator [J].Chemosphere,2008,70(4):721-730.

[0176] [11]乔俊飞,郭子豪,汤健.面向城市固废焚烧过程的二噁英排放浓度检测方法综述[J].自动化学报(在审).(Qiao J F,Guo Z H,Tang J.Dioxin Emission Concentration Measurement Approaches for Municipal Solid Wastes Incineration Process:A Survey[J].Acta Automatica Sinica,in trial)

[0177] [12]Chang N B,Huang S H.Statistical modelling for the prediction and control of PCDDs and PCDFs emissions from municipal solid waste incinerators [J].Waste Management&Research,1995,13,379-400.

[0178] [13]Chang N B,Chen W C.Prediction of PCDDs/PCDFs emissions from

municipal incinerators by genetic programming and neural network modeling[J].Waste Management&Research,2000,18(4)41-351.

[0179] [14]汤健,乔俊飞.基于选择性集成核学习算法的固废焚烧过程二噁英排放浓度软测量[J],化工学报,2019,70(02):696-706.(Tang J,Qiao J F.Dioxin emission concentration soft measuring approach of municipal solid waste incineration based on selective ensemble kernel learning algorithm[J],Journal of Chemical Industry and Engineering(China),2019,70(02):696-706.)

[0180] [15]Bunsan S,Chen W Y,Chen H W,Chuang Y H,Grisdanurak N.Modeling the dioxin emission of a municipal solid waste incinerator using neural networks[J].Chemosphere,2013,92:258-264.

[0181] [16]肖晓东,卢加伟,海景,等.垃圾焚烧烟气中二噁英类浓度的支持向量回归预测[J].可再生能源,2017,35(8):1107-1114.(Xiao X D,Lu J W,Hai J.Prediction of dioxin emissions in flue gas from waste incineration based on support vector regression[J],Renewable Energy Resources,2017,35(8):1107-1114.)

[0182] [17]汤健,乔俊飞,郭子豪.基于潜在特征选择性集成建模的二噁英排放浓度软测量[J],自动化学报(在审).(Tang J,Qiao J F,Guo Z H.Soft Sensing of Dioxin Emission Concentration Based on Potential Characteristic Selective Integrated Modeling[J].Acta Automatica Sinica,in trial)

[0183] [18]Hasnat A,Molla A U.Feature selection in cancer microarray data using multi-objective genetic algorithm combined with correlation coefficient[A].2016International Conference on Emerging Technological Trends(ICETT)[C].2016:1-6.

[0184] [19]Coelho F,Braga AP,Verleysen M.Multi-Objective Semi-Supervised Feature Selection and Model Selection Based on Pearson's Correlation Coefficient[A].Iberoamerican Congress on Pattern Recognition.Springer[C],Berlin,Heidelberg,2010:509-516.

[0185] [20]Battiti R.Using mutual information for selecting features in supervised neural net learning[J].IEEE Transactions on Neural Networks,1994,5(4):537-550.

[0186] [21]Vergara J R,Estévez P A.A review of feature selection methods based on mutual information[J].Neural computing and applications,2014,24(1):175-186.

[0187] [22]Jain A K,Duin R P W,Mao J.Statistical pattern recognition:A review[J].IEEE Transactions on pattern analysis and machine intelligence,2000,22(1):4-37.

[0188] [23]Fleuret F.Fast binary feature selection with conditional mutual information[J].Journal of Machine Learning Research,2004,5:1531-1555.

[0189] [24]Coelho F,Braga AP,Verleysen M.Multi-Objective Semi-Supervised Feature Selection and Model Selection Based on Pearsons Correlation

Coefficient[J].Lecture Notes in Computer Science,2010,6419:509-516.

[0190] [25]Estévez P A,Tesmer M,Perez C A,Zurada J M.Normalized mutual information feature selection[J].IEEE Transactions on Neural Networks,2009,20(2):189-201.

[0191] [26]Amiri F,Yousefi M M R,Lucas C,Shakery A,Yazdani N.Mutual information-based feature selection for intrusion detection systems[J].Journal of Network and Computer Applications,34(2011)1184-1199.

[0192] [27]Mohammadi S,Mirvaziri H,Ghazizadehahsaee M.Multivariate correlation coefficient and mutual information-based feature selection in intrusion detection[J].Information Security Journal A Global Perspective,2017,26(5):229-239.

[0193] [28]Peng H,Long F,Ding C.Feature selection based on mutual information criteria of max-dependency,max-relevance,and min-redundancy[J].IEEE Transactions on pattern analysis and machine intelligence,2005,27(8):1226-1238.

[0194] [29]汤健,田福庆,贾美英.基于频谱数据驱动的旋转机械设备负荷软测量[M].北京:国防工业出版社,2015.(Tang J,Tian F Q,Jia M Y.Soft Measurement of Rotating Machinery Equipment Load Based on Spectrum Data Drive[M].Beijing:National Defense Industry Press,2015.)

[0195] [30]Tihonov AN.Solution of incorrectly formulated problems and the regularization method[J].Soviet Math.,1963,4:1035-1038.

[0196] [31]Wold S,Ruhe A,Wold H,Dunn III W J.The collinearity problem in linear regression.The partial least squares(PLS)approach to generalized inverses[J].SIAM Journal on Scientific and Statistical Computing,1984,5(3):735-743.

[0197] [32]Leardi R,Boggia R,Terrile M.Genetic algorithms as a strategy for feature selection[J].Journal of chemometrics,1992,6(5):267-281.

[0198] [33]汤健,柴天佑,赵立杰,岳恒,郑秀萍,融合时频信息的磨矿过程磨机负荷软测量[J],控制理论与应用.2012,29(5):564-570.(TANG J,CHAI T Y,ZHAO L J,YUE H,ZHENG X P.Soft sensing mill load in grinding process by time/frequency information fusion[J].Control Theory and Applications,2012,29(5):564-570.)

[0199] [34]Bunsan S,Chen W Y,Chen H W,Chuang Y H,Grisdanurak N.Modeling the dioxin emission of a municipal solid waste incinerator using neural networks[J].Chemosphere,2013,92:258-264.

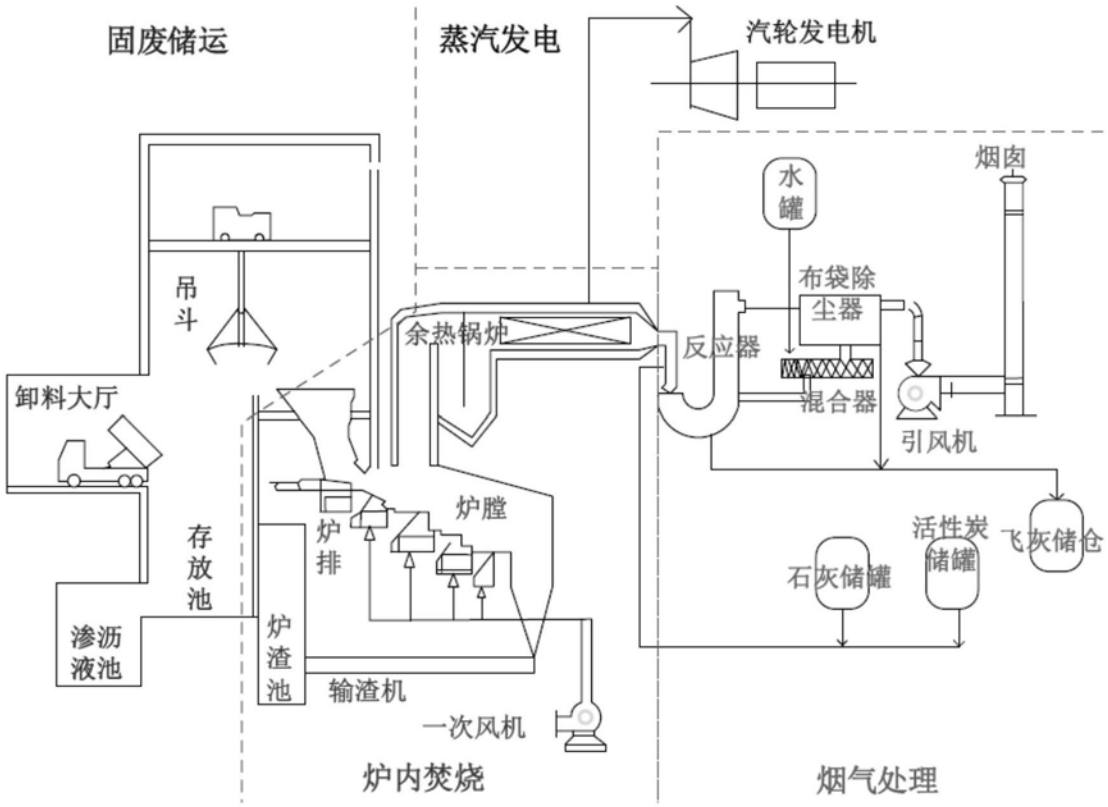

固废储运
蒸汽发电
汽轮发电机
烟囱
水罐
布袋除尘器
吊斗
余热锅炉
反应器
卸料大厅
混合器
引风机
炉膛
炉排
存放池
活性炭储罐
飞灰储仓
石灰储罐
渗沥液池
炉渣池
输渣机
一次风机
炉内焚烧
烟气处理

图1

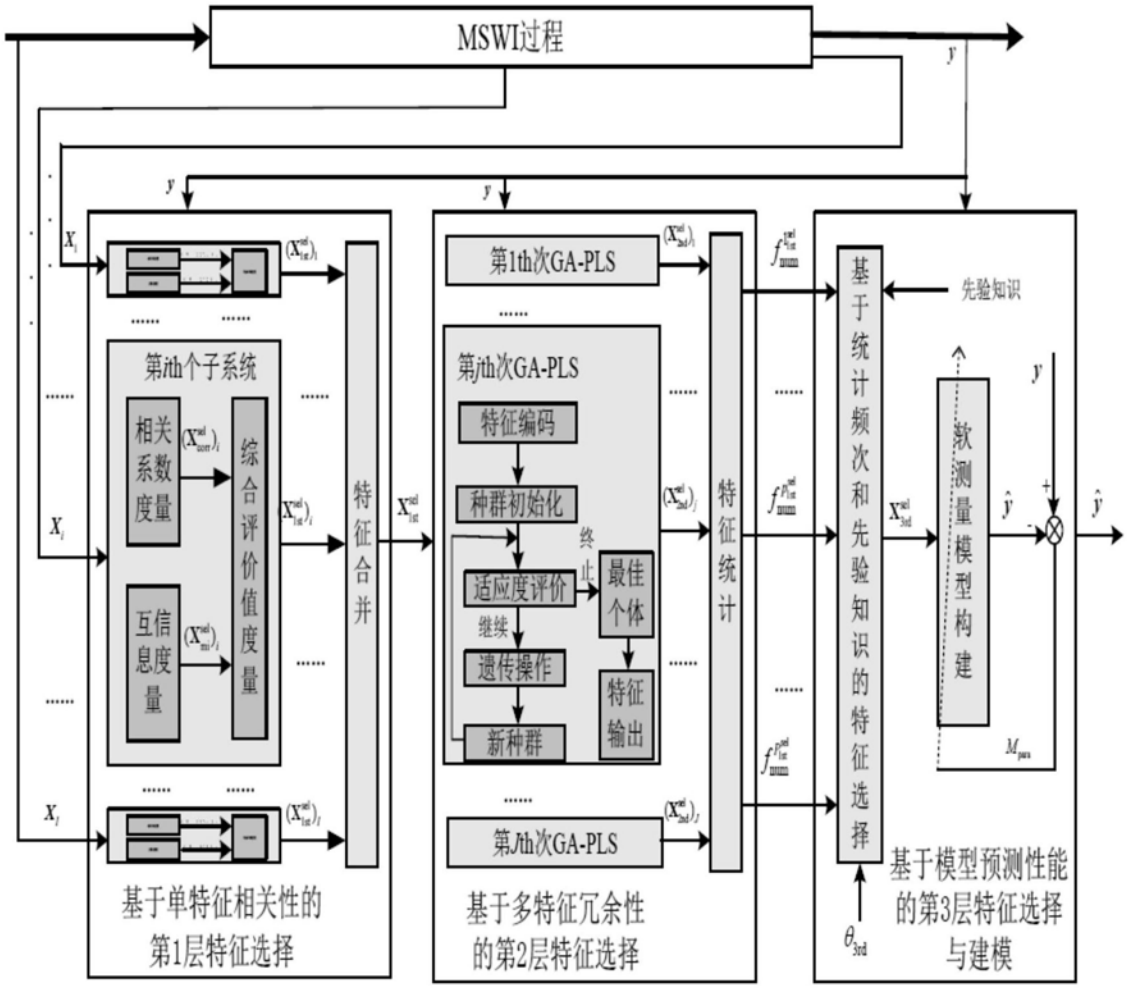

MSWI过程
y
第ith个子系统
相关系数度量
互信息度量
综合评价值度量
特征合并
基于单特征相关性的第1层特征选择
第1th次GA-PLS
第jth次GA-PLS
特征编码
种群初始化
适应度评价
终止
最佳个体
继续
遗传操作
新种群
特征输出
第Jth次GA-PLS
特征统计
基于多特征冗余性的第2层特征选择
基于统计频次和先验知识的特征选择
先验知识
软测量模型构建
基于模型预测性能的第3层特征选择与建模

图2

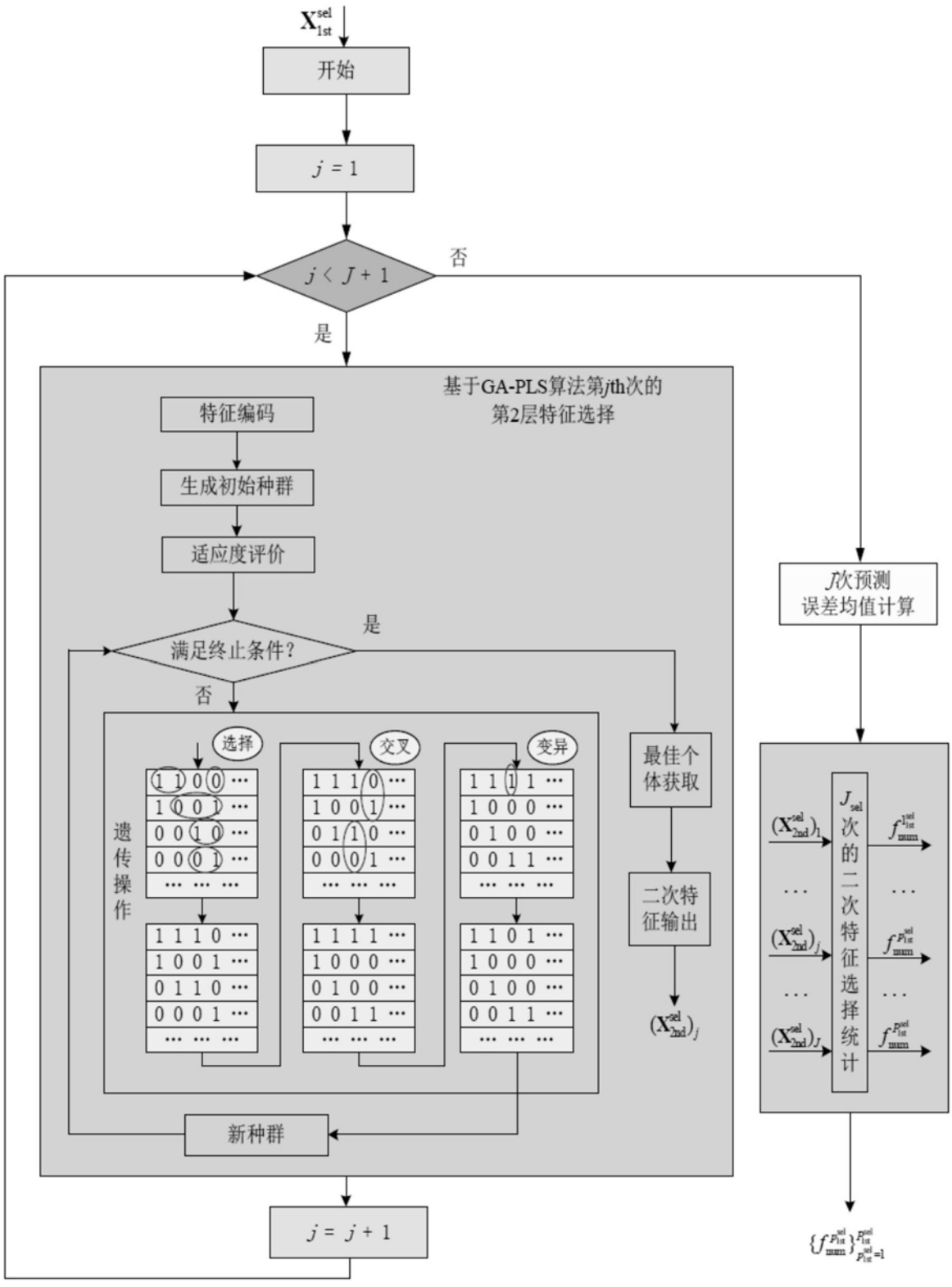

X_{1st}^{sel}
开始
$j = 1$
$j < J + 1$
否
是
基于GA-PLS算法第jth次的
第2层特征选择
特征编码
生成初始种群
适应度评价
满足终止条件?
是
否
选择
交叉
变异
遗传操作
最佳个体获取
二次特征输出
$(X_{2nd}^{sel})_j$
新种群
$j = j + 1$
J次预测
误差均值计算
J_{sel}次的二次特征选择统计
$(X_{2nd}^{sel})_1$
$(X_{2nd}^{sel})_j$
$(X_{2nd}^{sel})_J$
f_{num}^{1st}
$f_{num}^{P_{1st}^{sel}}$
$\{f_{num}^{P_{1st}^{sel}}\}_{P_{1st}^{sel}=1}^{P_{1st}^{sel}}$

图3

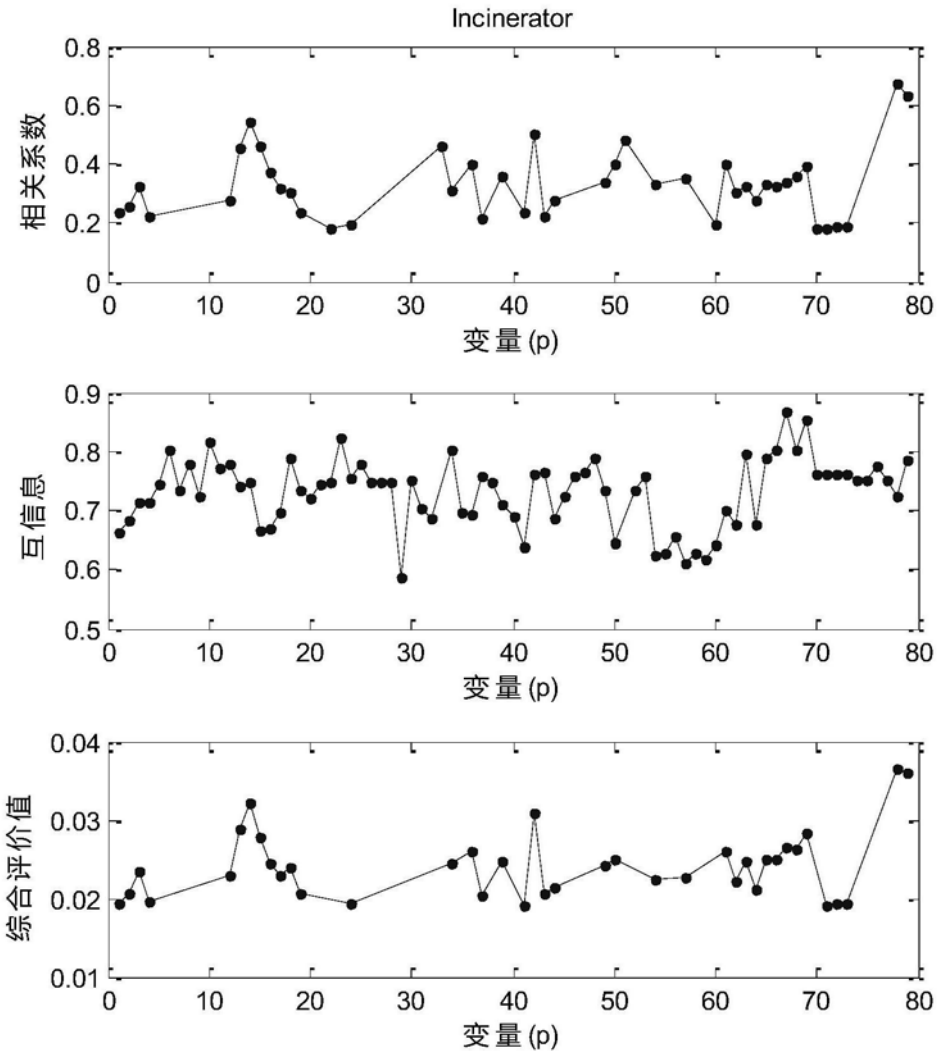


图4

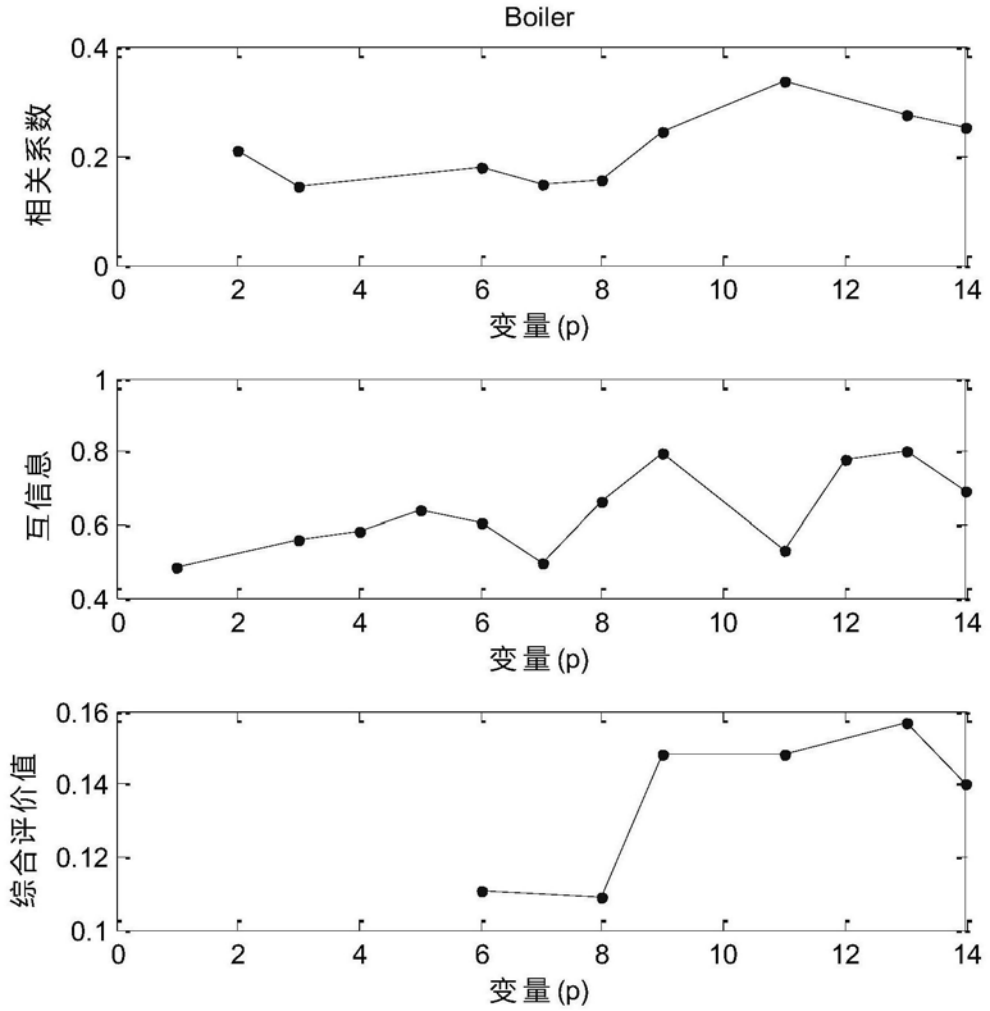

Boiler
相关系数
0.4
0.2
0
0
2
4
6
8
10
12
14
变量(p)
互信息
1
0.8
0.6
0.4
0
2
4
6
8
10
12
14
变量(p)
综合评价值
0.16
0.14
0.12
0.1
0
2
4
6
8
10
12
14
变量(p)

图5

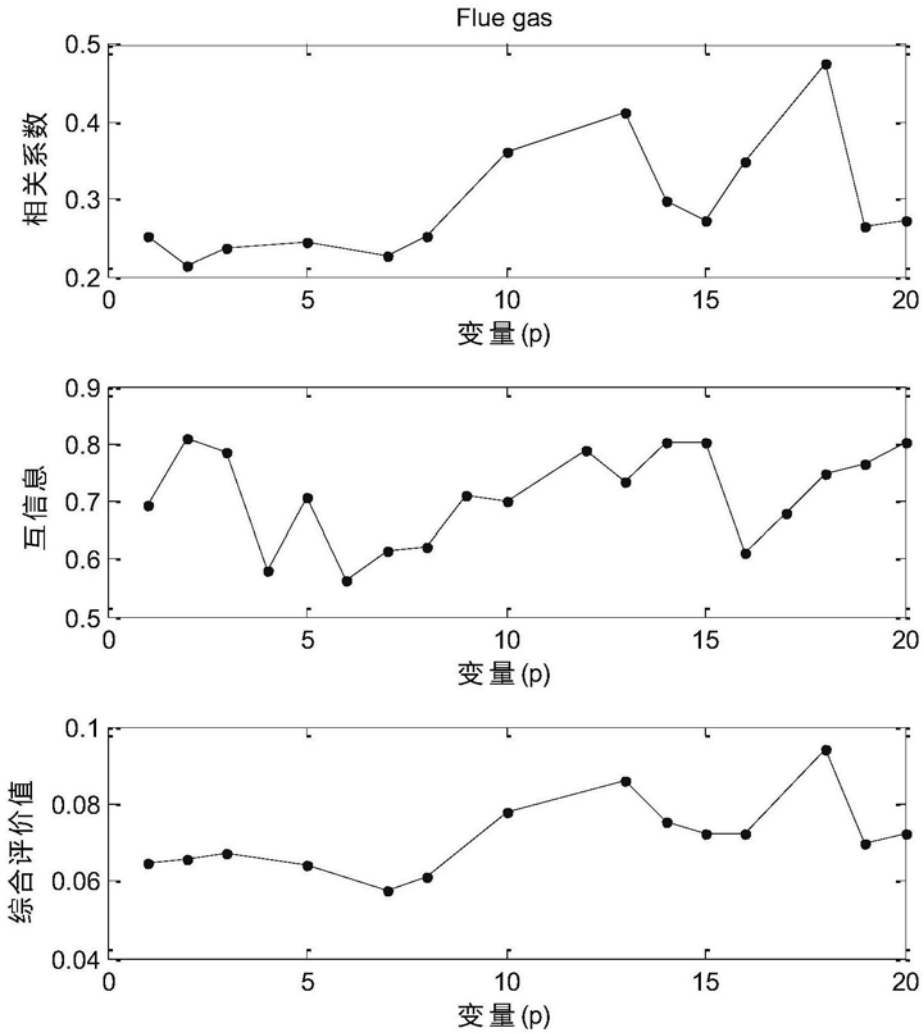

Flue gas
相关系数
互信息
综合评价值
变量(p)
0.5
0.4
0.3
0.2
0.9
0.8
0.7
0.6
0.1
0.08
0.06
0.04
0
5
10
15
20

图6

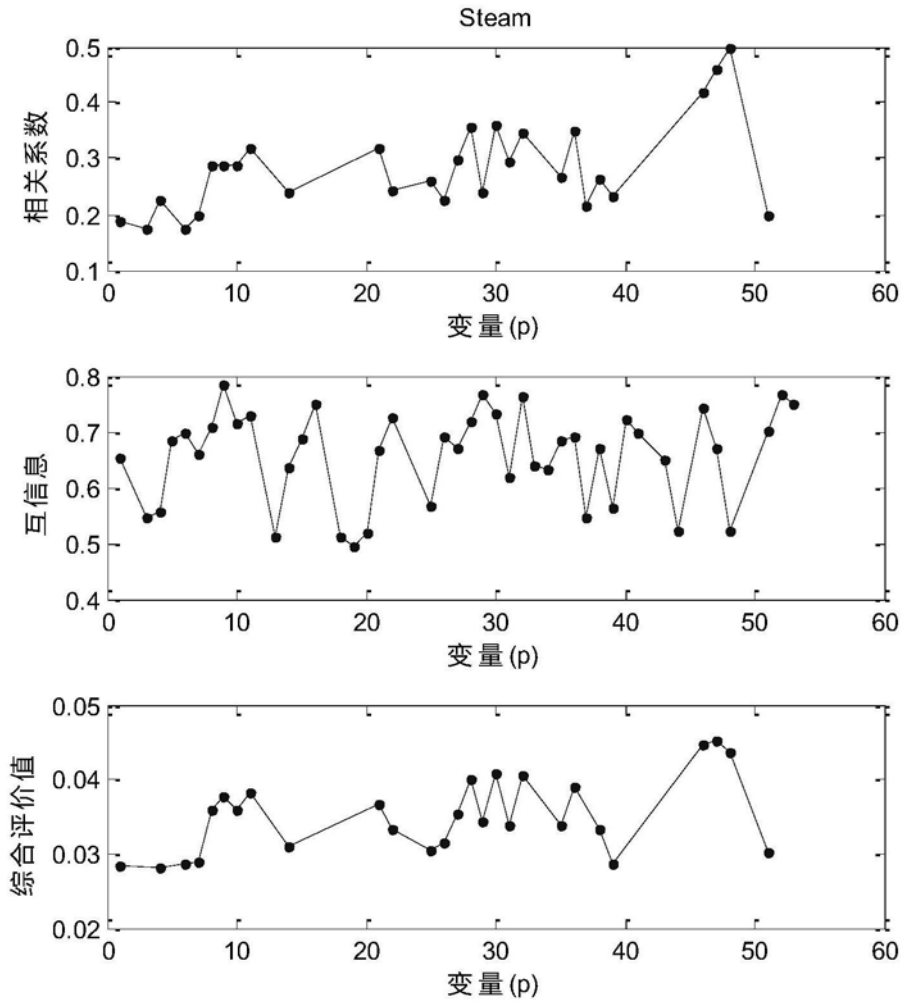

Steam
相关系数
0.5
0.4
0.3
0.2
0.1
0
10
20
30
40
50
60
变量(p)
互信息
0.8
0.7
0.6
0.5
0.4
0
10
20
30
40
50
60
变量(p)
综合评价值
0.05
0.04
0.03
0.02
0
10
20
30
40
50
60
变量(p)

图7

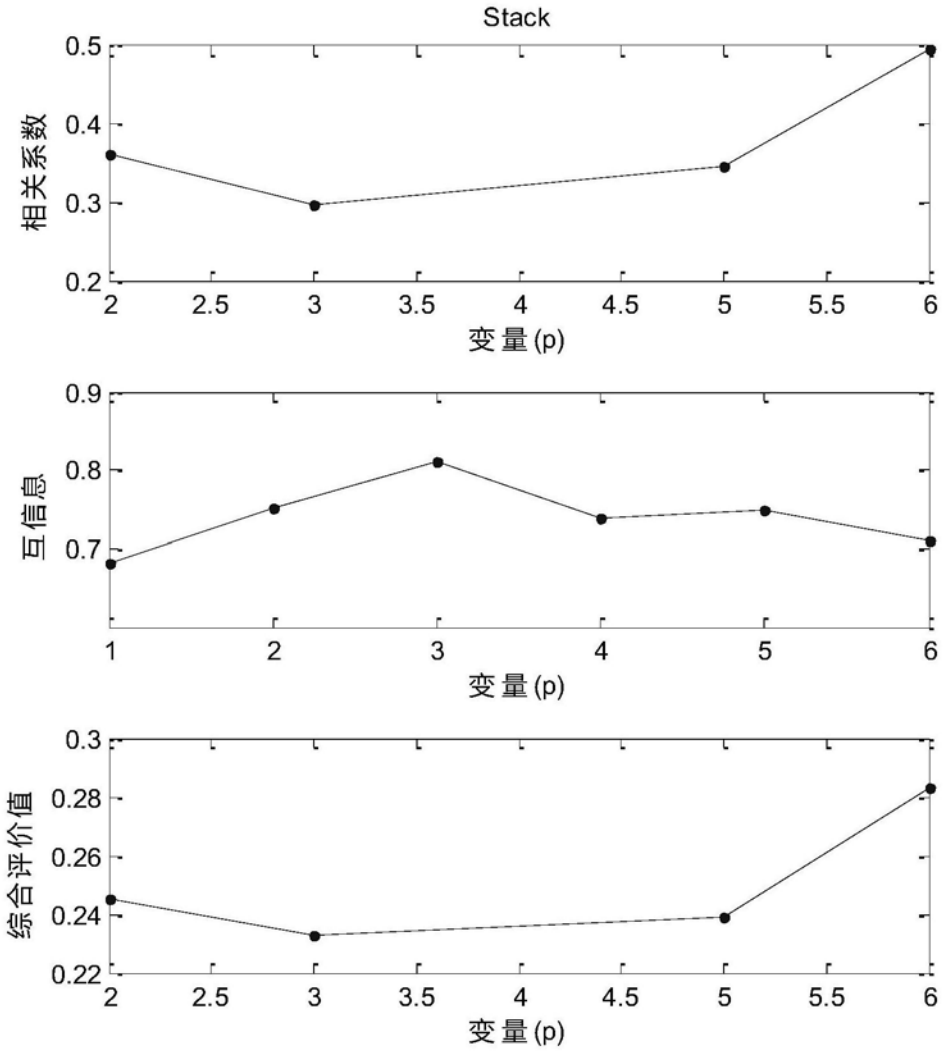

Stack
相关系数
0.5
0.4
0.3
0.2
2
2.5
3
3.5
4
4.5
5
5.5
6
变量(p)
互信息
0.9
0.8
0.7
1
2
3
4
5
6
变量(p)
综合评价值
0.3
0.28
0.26
0.24
0.22
2
2.5
3
3.5
4
4.5
5
5.5
6
变量(p)

图8

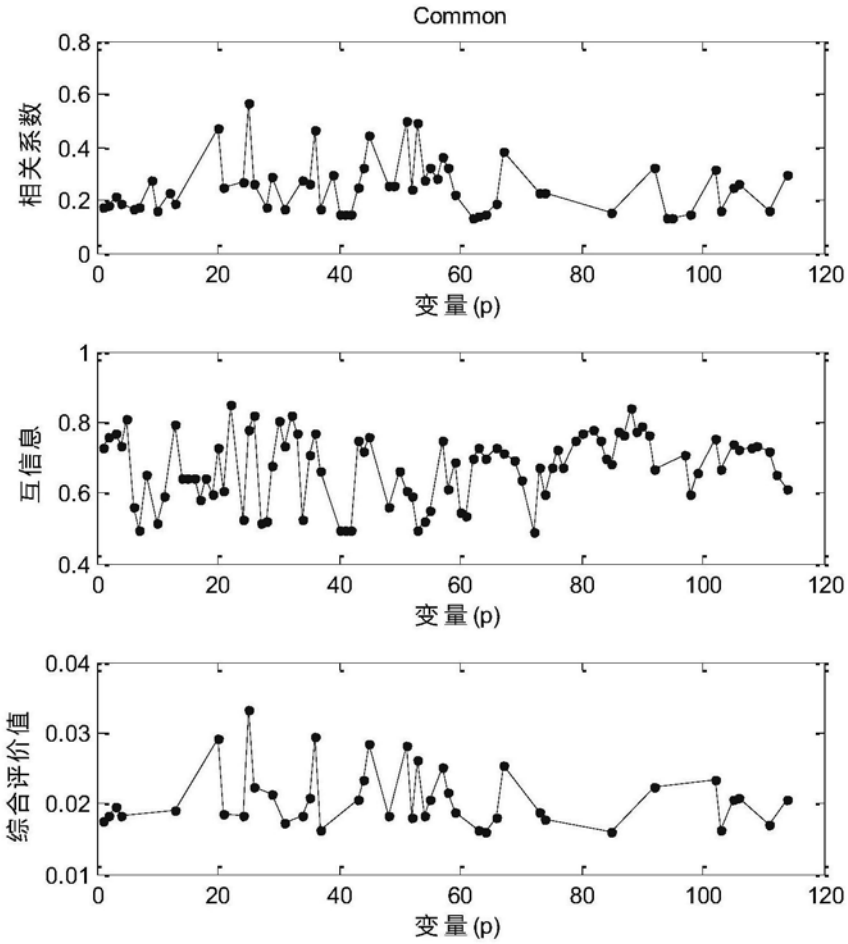

Common
相关系数
0.8
0.6
0.4
0.2
0
0
20
40
60
80
100
120
变量(p)
互信息
1
0.8
0.6
0.4
0
20
40
60
80
100
120
变量(p)
综合评价值
0.04
0.03
0.02
0.01
0
20
40
60
80
100
120
变量(p)

图9