

中国研究生创新实践系列大赛  
“华为杯”第十七届中国研究生  
数学建模竞赛

学 校 北京邮电大学

---

参赛队号 20100130029

---

队员姓名	1. 唐麒淳
	2. 段祥卿
	3. 戴维

---

中国研究生创新实践系列大赛  
“华为杯”第十七届中国研究生  
数学建模竞赛

题    目    全降低汽油精制过程中的辛烷值损失模型

---

摘        要：

摘要文字，请删除我巴拉巴拉

问题一：唐麒淳

问题二：唐麒淳

问题三：唐麒淳

问题四：唐麒淳

问题五：唐麒淳

关键字： 分布转换   基于模型的特征筛选   贝叶斯优化   5折交叉验证   TPE 算法

## 目录

<b>1. 问题重述</b>	<b>4</b>
1.1 问题背景	4
1.2 问题提出	4
<b>2. 模型假设</b>	<b>4</b>
<b>3. 符号说明</b>	<b>4</b>
<b>4. 问题分析与求解</b>	<b>5</b>
4.1 问题一：数据处理	5
4.1.1 问题分析	5
4.1.2 对含空值变量的分析	5
4.1.3 剔除不在操作范围内的样本	6
4.1.4 用 $3\sigma$ 准则删除异常样本	7
4.1.5 对工业数据中的空值变量进行处理	7
4.2 问题二：寻找建模主要变量	7
4.2.1 问题分析	7
4.2.2 场景建模或者其他	7
4.2.3 特征设计及其他	7
4.3 问题三：建立辛烷值（RON）损失预测模型	7
4.3.1 问题分析	7
4.3.2 场景建模或者其他	8
4.3.3 特征设计及其他	8
4.4 问题四：主要变量操作方案的优化	8
4.4.1 问题分析	8
4.4.2 场景建模或者其他	9
4.4.3 特征设计及其他	9
4.5 问题五：模型的可视化展示	9
4.5.1 问题分析	9
4.5.2 场景建模或者其他	9
4.5.3 特征设计及其他	9
<b>5. 模型评价</b>	<b>9</b>
5.1 模型的优点	9

5.2 模型的缺点.....	10
<b>6. 加图的方法，记得删除 .....</b>	<b>10</b>
<b>参考文献.....</b>	<b>11</b>
<b>参考文献.....</b>	<b>11</b>
<b>附录 A 我的 Python 源程序.....</b>	<b>12</b>

## 1. 问题重述

## 1.1 问题背景

巴拉巴拉一堆话

## 1.2 问题提出

### 问题 1：巴拉巴拉几个字

[illegible]

### 问题 2:

### 问题 3:

#### 问题 4:

### 问题 5:

## 2. 模型假设

1. 假设一
2. 假设二

### 3. 符号说明

表1 论文中用到的符号定义

符号	意义
D	木条宽度 (cm)
N	第 n 根木条
T	木条根数
H	桌子高度 (cm)

## 4. 问题分析与求解

### 4.1 问题一：数据处理

#### 4.1.1 问题分析

由于（套话）原因，采集到的原始数据存在一些异常，包括（套话）

问题一的目标是参考近 4 年的工业数据（325 个数据样本数据.xlsx），依“样本确定方法”对 285 号和 313 号数据样本进行预处理，并加入到原工业数据中。

#### 4.1.2 对含空值变量的分析

处理 1：对时序上基本平稳的含空值变量做均值填充

我们在 313 样本的数据中，发现了 2 个时序上基本平稳但含有空值的变量

TODO: 两图并列

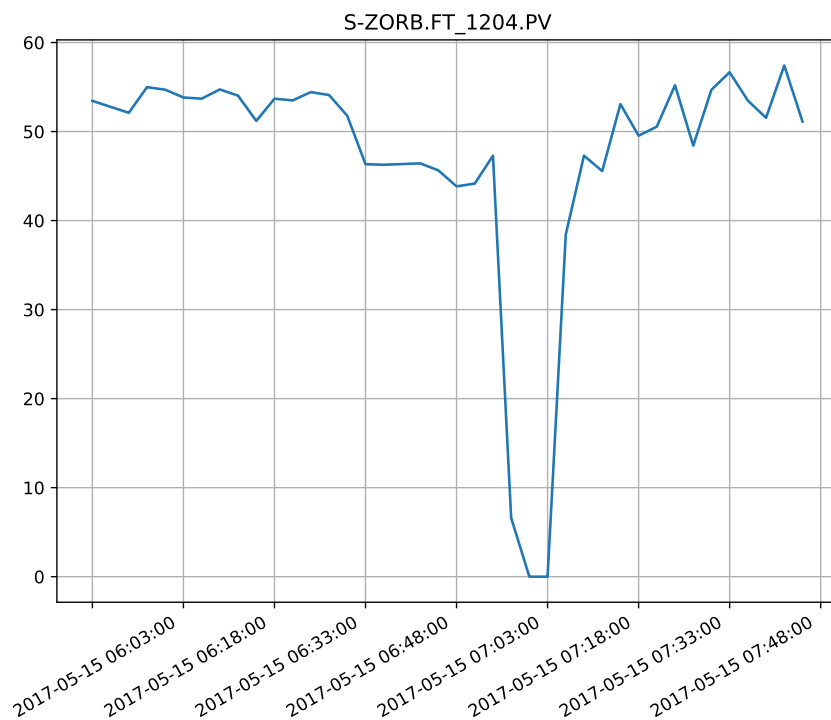


图 1 S-ZORB.FT\_1204.PV

对于这两个变量的空值用 313 样本数据的其余采样相应变量的均值来填充

处理 2：对时序上杂乱或呈周期趋势的含空值变量做删除操作

TODO: 三图并列

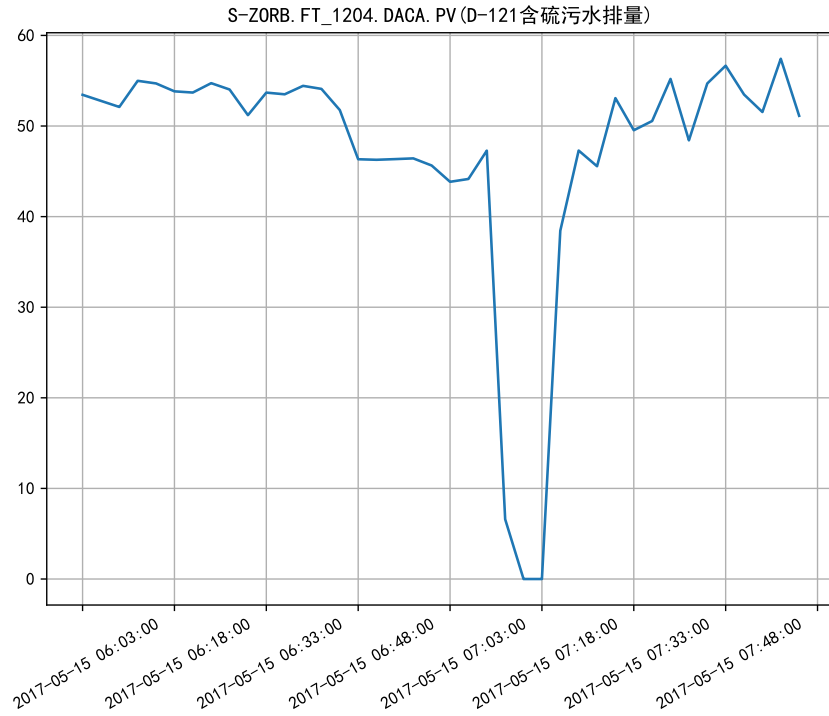


图 2 D-121 含硫污水排量

#### 4.1.3 剔除不在操作范围内的样本

考虑到“325 个样本数据.xlsx”中的很多样本已经超出了“354 个操作变量信息.xlsx”所规定的范围，我们进行了简单的处理，用“325 个样本数据”中每个操作变量的最大与最小值来扩充“354 个操作变量信息”中操作变量的范围，并用扩充后的操作变量范围代替原操作变量范围。

经过计算，285 样本数据中所有的采样数据都在操作变量范围内，而 313 样本数据仅有 1 个采样数据的所有变量在操作变量范围内。经过综合考虑，我们决定用以下公式来计算一个采样样本超出操作变量范围的程度：

$$InvalidDegree = \sum_j^M \frac{Exceed_j}{Upper_j - Lower_j} \quad (1)$$

其中， $j$  表示某采样样本的第  $j$  个操作变量， $M$  表示共有  $M$  个操作变量， $Upper_j$  表示操作变量  $j$  的上界， $Lower_j$  表示操作变量  $j$  的下界， $Exceed_j$  表示操作变量  $j$  超出范围的大小。

通过上述公式，我们对 313 样本的采样数据进行了计算：











## 参考文献

- [1] Zheng L, Wang S, Tian L, et al., Query-adaptive late fusion for image search and person re-identification, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 1741-1750.
- [2] Arandjelović R, Zisserman A, Three things everyone should know to improve object retrieval, Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, IEEE, 2012: 2911-2918.
- [3] Lowe D G. Distinctive image features from scale-invariant keypoints, International journal of computer vision, 2004, 60(2): 91-110.
- [4] Philbin J, Chum O, Isard M, et al. Lost in quantization: Improving particular object retrieval in large scale image databases, Computer Vision and Pattern Recognition, 2008. CVPR 2008, IEEE Conference on, IEEE, 2008: 1-8.

## 参考文献

- [1] liuhaiyang2013latex 刘海洋.  $\text{\LaTeX}$  入门[J]. 电子工业出版社, 北京, 2013.
- [2] mathematical-modeling 全国大学生数学建模竞赛论文格式规范 (2020 年 8 月 25 日修改).
- [3] <https://www.latexstudio.net>

## 附录 A 我的 Python 源程序

```
while ~isempty(V)
    [tmpd,j]=min(W(i,V));tmpj=V(j);
    for k=2:ndd
        [tmp1,jj]=min(dd(1,k)+W(dd(2,k),V));
        tmp2=V(jj);tt(k-1,:)= [tmp1,tmp2,jj];
    end
    tmp=[tmpd,tmpj,j;tt];[tmp3,tmp4]=min(tmp(:,1));
    if tmp3==tmpd, ss(1:2,kk)=[i;tmp(tmp4,2)];
    else,tmp5=find(ss(:,tmp4)~=0);tmp6=length(tmp5);
    if dd(2,tmp4)==ss(tmp6,tmp4)
        ss(1:tmp6+1,kk)=[ss(tmp5,tmp4);tmp(tmp4,2)];
    else, ss(1:3,kk)=[i;dd(2,tmp4);tmp(tmp4,2)];
    end;end
    dd=[dd,[tmp3;tmp(tmp4,2)]];V(tmp(tmp4,3))=[];
    [mdd,ndd]=size(dd);kk=kk+1;
end; S=ss; D=dd(1,:);
```