

基于×××方法对炼钢过程中脱氧合金化配料方案优化的研究

摘要

炼钢过程中的脱氧合金化是钢铁冶炼中的重要工艺环节，通过在钢液中加入合金使得不同钢种中所含的合金元素达到国家标准，从而保证成品钢一定的物理性能。本文通过建立数学模型，对脱氧合金化过程中主要元素的收得率进行预测，实现了钢水脱氧合金化的成本优化并给出相应的合金配料方案。

针对问题一，计算历史收得率并分析主要影响因素。首先根据脱氧合金化的原理得出历史收得率的计算公式，将数据导入 scripy 第三方集成库进行计算。剔除计算后出现空值、0 值和无穷大的数据后作为基础数据集。利用杜宾-瓦特森检验自变量之间的自相关性，得到 D-W 值为 1.78，说明变量之间不具有自相关性。最后通过结合信息增益和 Pearson 系数两种方法检验自变量和因变量之间的相关性，结果显示，**锰硅合金 FeMn68Si18(合格块)、钢水净重和转炉终点温度对 Mn 的得出率影响最大，其值分别为 3.556、3.015 和 2.377，碳化硅、钢水净重和锰硅合金 FeMn68Si18(合格块)对 C 的得出率影响最大，其值分别是 0.813, 0.541 和 0.428。**

针对问题二，分别建立 C 和 Mn 的得出率预测模型尽可能地提高准确率。利用线性回归建立 C 和 Mn 的得出率模型，首先，根据数据完整性分别选择 611、193 条数据建立 C 和 Mn 作为样本，其中 70%用于训练，30%用于检验。通过单标签线性回归模型得到 C 的预测模型。由于 Mn 的数据集中缺乏 193-611 的转炉终点数据，因此采用 193 条 Mn 和 C 的数据建立双标签输出模型得到 Mn 的预测模型，结果显示，C 的预测准确率为 ，Mn 的预测准确率为 。

针对问题三，建立数学模型实现钢水脱氧合金化成本优化计算并给出合金配料方案。首先根据国家标准对数据进行预选择，筛除元素含量不符合国家标准的炉次数据；接着利用合金物成本和合金收得率构建双目标函数关系；利用问题二中的方法构建 C、Mn、S、P、Si 五种元素的预测回归模型，结合附件二中的数据计算的各个炉次所加合金的成本，最后构造成本最低函数并计算得出最佳合金配料方案。

针对问题四，根据

1 问题的重述

1.1 背景知识

钢铁产业是国民经济支柱产业之一，脱氧合金化作为钢冶炼过程的最后一项操作，对于钢铁的质量具有重要影响。在转炉吹炼过程中，由于不断向炉中吹氧导致的钢水中残留一定数量的氧。因此需要通过加入一定质量的脱氧合金剂，使钢水达到合乎规定的脱氧程度。脱氧的同时，还要使钢水中硅、锰及其他元素的含量符合规程要求，以达到合金化的目的。传统的炼钢过程控制中，合金加入量的确定通常按照不同元素的固定收得率和经验值计算得到，由于误差较大，造成一定的浪费，降低了脱氧合金化效果。因此，如何通过历史数据对脱氧合金化缓解建立数学模型，在线预测并优化投入合金的种类和数量，在保证钢水质量的同时最大限度地降低合金钢的生产成本是各大钢铁企业提高竞争力所要解决的重要问题。目前，使用较多的加料计算方法有参考炉次法，回归分析和神经网络等。

1.2 问题提出

问题一 根据炼钢历史数据计算 C、Mn 两种元素的历史收得率并分析影响其收得率的主要因素。

问题二 在问题一的基础上，建立数学模型，对 C、Mn 两种元素收得率进行预测，并进一步改进模型和算法，尽可能提高这两种元素收得率的预测准确率。

问题三 不同合金料的价格不同，其选择直接影响钢水脱氧合金化的成本，根据问题二中合金收得率的预测结果以及附件 2，建立数学模型，实现钢水脱氧合金化成本计算，并给出合金配料方案。

问题四 根据研究结果，给炼钢厂领导写一封建议信（一页以内）。

其中，附件 1 给出了炼钢历史数据，包括炉号、钢种、钢号、转炉终点温度、C、Mn、S、P、Si 的转炉终点含量以及连铸正样、钢水净重、加入合金料的质量以及硅钙碳脱氧剂的加入量。附件 2 给出了向钢水中加入的合金料的自身成分。

2 问题的分析

2.1 对问题一的分析

问题一要求我们根据附件 1 计算 C、Mn 两种元素的历史收得率并分析影响其收得率的主要因素。收得率的计算公式可以根据原理获得。由于数据集存在缺失，因此分别选择 C 和 Mn 的不缺失数据计算收得率。

首先利用 SPSS 软件对附件 1 数据进行预处理，将处理后的数据利用公式进行计算。剔除计算后出现空值、0 值和无穷大的数据后作为基础数据集。

考虑到数据集炉号是连续的，可能是在同一或不同钢炉中依据连续生产的，即具有时间序列的特征，因此需要在考虑变量相关性之前需要先确定自变量之间的自相关性。因此利用杜宾-瓦特森检验自变量之间的自相关性。要判断影响历史收得率的主要因素，可以通过相关性的确定来确定主要影响因素。我们选择综合信息增益和 Pearson 系数检验两种方法判断自变量（炼钢数据集中的多个特征）和因变量（合金的收得率）之间的相关性，并根据结果得出对其影响较大的三个因素。

2.2 对问题二的分析

问题二要求我们建立数学模型对 C、Mn 两种元素的收得率进行预测并进一步改进模型和算法提高预测准确率。通过问题一中所得到的主要影响因素建立第一个回归模型，其中用于训练的数据为炼钢数据集中 C 和 Mn 均不缺失的数据。接着利用第一个回归模型计算得到炼钢数据集中缺少转炉终点 Mn 数据的 Mn 得出率得到数据集的扩大。根据扩大的数据集建立第二个回归模型计算同时缺少连铸正样 C 和 Mn 数据的元素收得率，进而获得所有数据的元素收得率。

在分析准确率时我们发现数据存在一定的误差，通过对已有模型的检查我们发现在建立第二个回归模型时利用扩大的数据集造成了误差的积累。因此，在建立第二个回归模型时改用炼钢数据集中 C 和 Mn 均不缺失的数据进行训练，获得元素收得率较准确的预测模型。

2.3 对问题三的分析

问题三要求我们通过建立数学模型实现钢水脱氧合金化的成本优化计算并给出配料方案。要实现成本的最小化需要使得各合金成分的价格与对应使用量的积之和最小。另一方面，元素收得率表示钢水中吸收的元素占合金中元素的比例，元素收得率越大即表明合金的利用率越高。另外，不同元素收得率会对应不同的合金投入量，在元素收得率准确预测的基础上通过可以选择得到最优化的合金配料方案，实现钢水脱氧合金化成本优化计算。因此可以将问题转化为成本最小，元素的收得率最大，合金的元素含量符合国家标准的三目标优化问题。

2.4 对问题四的分析

问题四要求我们根据研究结果给炼钢厂领导写一封信。由于收信人是炼钢厂

的领导，所以给他提出的建议应该是我们通过研究所得出的在炼钢脱氧合金化过程中的最佳合金配料方案，使他能够最大程度地减小成本增大利润。此外，应该注意由于炼钢厂领导虽然具有丰富的炼钢经验但是不一定具有相关的数学和计算机基础，因此在向他阐述建议时需要注意语言要通俗易懂。

3 模型的假设

- 3.1 假设钢包中钢液混合充分；
- 3.2 假设合金料称量系统准确；
- 3.3 假设钢水成分检验系统准确；
- 3.4 假设同种钢种精炼渣成分变化范围不大；
- 3.5 假设处理过程中空气不会对钢液产生二次氧化，即钢液不在大气中吸氮和氧；
- 3.6 假设合金、脱氧剂有充足时间与钢液充分混合；
- 3.7 烧损掉的合金全部与钢液中溶解氧反应生成氧化物

4 符号说明

序号	符号	符号说明
1	η	
2	C	
3	w	
4		
5		
6		

5 模型一的建立与求解

5.1.对问题一的分析

问题一要求我们根据附件 1 计算 C、Mn 两种元素的历史收得率并分析影响其收得率的主要因素。收得率的计算公式可以根据原理获得。由于数据集存在缺失，因此分别选择 C 和 Mn 的不缺失数据计算收得率。

首先利用 SPSS 软件对附件 1 数据进行预处理，将处理后的数据导入 scripty 第三方集成库进行计算。剔除计算后出现空值、0 值和无穷大的数据后作为基础数据集。

考虑到数据集炉号是连续的，可能是在同一或不同钢炉中依据连续生产的，即具有时间序列的特征，因此需要在考虑变量相关性之前需要先确定自变量之间的自相关性。因此利用杜宾-瓦特森检验自变量之间的自相关性。要判断影响历史收得率的主要因素，可以通过相关性的确定来确定主要影响因素。我们选择信息增益和 Pearson 系数两种方法检验自变量和因变量之间的相关性，并根据结果得出对其影响较大的三个因素。具体流程如图 1 所示。

5.2 模型的建立

5.2.1 数据检查与预处理

附件 1 给出了炼钢历史数据，包括炉号、钢种、钢号、转炉终点温度、C、Mn、S、P、Si 的转炉终点含量以及连铸正样、钢水净重、加入合金料的质量以及硅钙碳脱氧剂的加入量。

首先利用 SPSS 对附件 1 的数据进行探索性分析（详情见附录），发现存在缺失，对此本文利用排除个案法，对原始数据进行分析，剔除计算后出现空值、0 值和无穷大的数据后作为基础数据集。

5.2.2 合金收得率的计算

根据合金收得率的定义，合金收得率指脱氧合金化时被钢水吸收的合金元素的质量与加入该元素的总质量之比，即：

$$\text{元素收得率} = \frac{\text{钢中合金元素的质量}}{\text{合金元素加入总量}} \times 100\%$$

根据（1）式，结合附件 1 中所给的数据，可得以下公式：

$$\eta_B = \frac{w[m]_x - w[m]_c}{\sum_{i=1}^{i=s} g_i * w[m]_i * G}$$

式中：

$w[m]_x$ 为终点钢液中 m 元素质量分数；

$w[m]_c$ 为初始钢液中 m 元素质量分数；

$w[m]i$ 为 gi 合金中 m 元素质量分数;

gi 为合金加入量;

G 为钢包中钢水重量。

分析数据集中的数据可知,C 元素的连铸正样数据在炉号 7A06060 之后缺失,因此可准确计算的收得率数据为炉号 7A06060 之前的数据。同理, Mn 元素的转炉终点数据在炉号 7A06620 之后缺失, 因此可准确计算的收得率为炉号 7A06620 之前的数据。(计算结果见附录)

5.2.3 主要影响因素分析

(1) 数据标准化

由于所选数据的量纲和数值大小都不一致,数值的变化范围也不同,因此必须首先对所选数据进行标准化处理,如果有 n 个样本,个样本有 m 个指标,则每个变量可表示为 x_{ij} , 均值

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$$

标准方差

$$s_j = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}$$

则标准化后

$$x_{ij}^* = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (s_j \neq 0)$$

(2) 主要影响因素分析

①相关分析是描述两个变量间关系的密切程度,主要由相关系数值表示,当相关系数 r 的绝对值越接近于1,则表示两个变量间的相关性越显著。双变量系数测量的主要指标有卡方类测量、Spearman相关系数、pearson相关系数等,在进行两者间的相关性检验时pearson相关系数来判断,其公式为:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

Pearson简单相关系数检验统计量为:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

其中 t 统计量服从 $n-2$ 个自由度的 t 分布。

②信息增益中的衡量标准是看特征能够为分类系统带来多少信息，带来的信息越多，该特征越重要。对一个特征而言，系统有它和没它时信息量将发生变化，而前后信息量的差值就是这个特征给系统带来的信息量。所谓信息量，就是熵。

假如有变量 X ，其可能的取值有 n 种，每一种取到的概率为 P_i ，那么 X 的熵就定义为

$$H(X) = - \sum_{i=1}^n p(x_i) \log p(x_i).$$

也就是说 X 可能的变化越多， X 所携带的信息量越大，熵也就越大。对于文本分类或聚类而言，就是说文档属于哪个类别的变化越多，类别的信息量就越大。所以特征 T 给聚类 C 或分类 C 带来的信息增益为 $IG(T)=H(C)-H(C|T)$ 。

$H(C|T)$ 包含两种情况：一种是特征 T 出现，标记为 t ，一种是特征 T 不出现，标记为 t' 。所以 $H(C|T)=P(t)H(C|t)+P(t')H(C|t')$ ，再由熵的计算公式便可推得特征与类别的信息增益公式。

利用相关增益法和 Pealson 系数衡量影响其收得率的主要因素，并将求得的值赋予不同的权重，将相关增益法的结果赋予 0.3 权重，Pealson 系数赋予 0.7 权重，计算结果为影响程度 i ，将 i 进行排序后得出结果，其中 Mn 的影响结果如表所示。

影响 Mn 收得率的因素	信息增益	Pearson 简单相关系数	i
转炉终点温度	2.378	0.778	1.258
钢水净重	3.015	0.332	1.137
锰硅合金 FeMn68Si18	3.556	0	1.067
钒铁(FeV50-B)		1	0.7
连铸正样 Mn	1.386	0.039	0.443
转炉终点 Mn	0.956	0.177	0.411
硅锰面(硅锰渣)	0.705	0.008	0.217

锰硅合金 FeMn64Si27	0.248		0.074
--------------------	-------	--	-------

影响 C 收得率的因素	信息增益	Pearson 简单相关系数	i
碳化硅（55%）	0	0.813	0.5691
钢水净重	0.387	0.541	0.4948
转炉终点温度	0.302	0.358	0.3412
锰硅合金 FeMn68Si18	0.108	0.428	0.332
锰硅合金 FeMn64Si18	0.108	0.428	0.332
钒铁(FeV50-B)	1	-8.33e-16	0.3
钒铁(FeV50-A)	1	-8.33e-16	0.3
硅块（合格块）	1	-8.33e-16	0.3
硅钙碳脱氧剂	0.893	0.029	0.2882
锰硅合金 FeMn64Si27	0.722	0.059	0.2579
转炉终点 C	0.003	0.331	0.2326
硅铝合金 FeAl64Si25	6.25e-23	0.267	0.1869
硅铁(FeSi75-B)	4.70e-34	0.202	0.1414
连铸正样 C	8.59e-23	0.196	0.1372
硅锰面（硅锰渣）	0.001	0.123	0.0864

（3）得出主要影响因素

根据表 可知对 Mn 收得率影响最大的因素为转炉终点温度，其影响程度 i 为 1.258；其次是钢水净重，影响程度 i 为 1.137；接着是锰硅合金 FeMn68Si18，其影响程度 i 为 1.067。

6 模型二的建立与求解

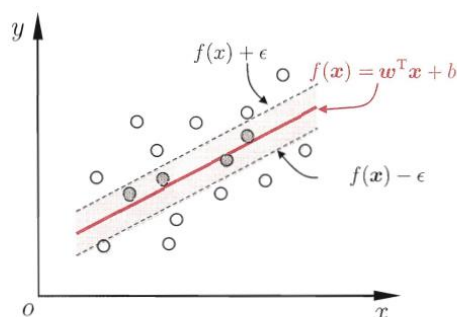
6.1 问题二的分析

问题二要求我们建立数学模型对 C、Mn 两种元素的收得率进行预测并进一步改进模型和算法提高预测准确率。通过建立 SVR、神经网络和线性回归模型，比较他们的均方误差，选择利用线性回归模型预测 C 和 Mn 的合金收得率。

6.2 模型的建立

6.2.1 SVR 模型的建立

给定训练样本 $D=\{(x_1,y_1),(x_2,y_2),\dots,(x_n,y_n)\}$, $y_i \in \mathbb{R}$, 我们希望学习到一个 $f(x)$ 使得其与 y 尽可能的接近, w, b 是待确定的参数。在这个模型中, 只有当 $f(x)$ 与 y 完全相同时, 损失才为零, 而支持向量回归假设我们能容忍的 $f(x)$ 与 y 之间最多有 ϵ 的偏差, 当且仅当 $f(x)$ 与 y 的差别绝对值大于 ϵ 时, 才计算损失, 此时相当于以 $f(x)$ 为中心, 构建一个宽度为 2ϵ 的间隔带, 若训练样本落入此间隔带, 则认为是被预测正确的。(间隔带两侧的松弛程度可有所不同)



支持向量回归示意图。红色显示出 ϵ -间隔带, 落入其中的样本不计算损失。

因此 SVR 问题可转化为 (下式左部是正则化项) :

$$\min_{w,b} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m \ell_{\epsilon}(f(x_i) - y_i)$$

ℓ 为损失函数

$$\ell_{\epsilon}(z) = \begin{cases} 0, & \text{if } |z| \leq \epsilon ; \\ |z| - \epsilon, & \text{otherwise.} \end{cases}$$

因此引入了松弛因子，重写第一个式子为：

$$\min_{\mathbf{w}, b, \xi_i, \hat{\xi}_i} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^m (\xi_i + \hat{\xi}_i)$$

$$\begin{aligned} \text{s.t. } & f(\mathbf{x}_i) - y_i \leq \epsilon + \xi_i , \\ & y_i - f(\mathbf{x}_i) \leq \epsilon + \hat{\xi}_i , \\ & \xi_i \geq 0, \hat{\xi}_i \geq 0 , \quad i = 1, 2, \dots, m. \end{aligned}$$

最后引入拉格朗日乘子，可得拉格朗日函数：

$$\begin{aligned} & L(\mathbf{w}, b, \alpha, \hat{\alpha}, \xi, \hat{\xi}, \mu, \hat{\mu}) \\ &= \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^m (\xi_i + \hat{\xi}_i) - \sum_{i=1}^m \mu_i \xi_i - \sum_{i=1}^m \hat{\mu}_i \hat{\xi}_i \\ &+ \sum_{i=1}^m \alpha_i (f(\mathbf{x}_i) - y_i - \epsilon - \xi_i) + \sum_{i=1}^m \hat{\alpha}_i (y_i - f(\mathbf{x}_i) - \epsilon - \hat{\xi}_i) . \end{aligned}$$

对四个遍历求偏导，令偏导数为零，可得

$$\mathbf{w} = \sum_{i=1}^m (\hat{\alpha}_i - \alpha_i) \mathbf{x}_i ,$$

$$0 = \sum_{i=1}^m (\hat{\alpha}_i - \alpha_i) ,$$

$$C = \alpha_i + \mu_i ,$$

$$C = \hat{\alpha}_i + \hat{\mu}_i .$$

把上边的式子带入，即可求得 SVR 的对偶问题

$$\begin{aligned} \max_{\alpha, \hat{\alpha}} \quad & \sum_{i=1}^m y_i (\hat{\alpha}_i - \alpha_i) - \epsilon (\hat{\alpha}_i + \alpha_i) \\ & - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m (\hat{\alpha}_i - \alpha_i) (\hat{\alpha}_j - \alpha_j) \mathbf{x}_i^T \mathbf{x}_j \\ \text{s.t.} \quad & \sum_{i=1}^m (\hat{\alpha}_i - \alpha_i) = 0 , \\ & 0 \leq \alpha_i, \hat{\alpha}_i \leq C . \end{aligned}$$

上边的过程需要满足 KKT 条件，即

$$\begin{cases} \alpha_i (f(\mathbf{x}_i) - y_i - \epsilon - \xi_i) = 0 , \\ \hat{\alpha}_i (y_i - f(\mathbf{x}_i) - \epsilon - \hat{\xi}_i) = 0 , \\ \alpha_i \hat{\alpha}_i = 0 , \xi_i \hat{\xi}_i = 0 , \\ (C - \alpha_i) \xi_i = 0 , (C - \hat{\alpha}_i) \hat{\xi}_i = 0 . \end{cases}$$

最后，可得 SVR 的解为

$$f(\mathbf{x}) = \sum_{i=1}^m (\hat{\alpha}_i - \alpha_i) \mathbf{x}_i^T \mathbf{x} + b .$$

其中 b 为

$$b = y_i + \epsilon - \sum_{i=1}^m (\hat{\alpha}_i - \alpha_i) \mathbf{x}_i^T \mathbf{x} .$$

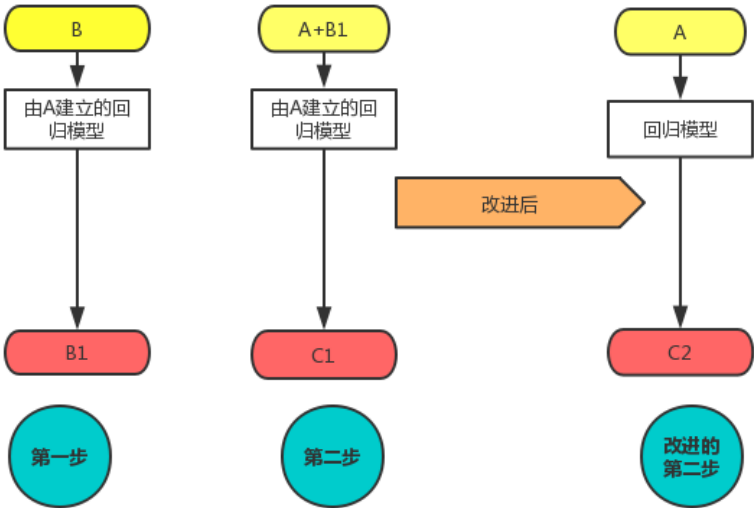
6.2.2 神经网络模型的建立

6.2.3 线性回归模型的建立

通过问题一中所得到的主要影响因素建立第一个回归模型，其中用于训练的数据为炼钢数据集中 C 和 Mn 均不缺失的数据。接着利用第一个回归模型计算得到炼钢数据集中缺少转炉终点 Mn 数据的 Mn 得出率得到数据集的扩大。根据扩大的数据集建立第二个回归模型计算同时缺少连铸正样 C 和 Mn 数据的元素收得率，进而获得所有数据的元素收得率。

在分析准确率时我们发现数据存在一定的误差，通过对已有模型的检查我们发现，在建立第二个回归模型时利用扩大的数据集造成了误差的积累。因此，在建立第二个回归模型时改用炼钢数据集中 C 和 Mn 均不缺失的数据进行训练，获得元素收得率较准确的预测模型。具体过程如图所示。

数据完整的Mn(A)
数据完整的C减数据完整的Mn (B)
其他 (C)



6.2 多元线性回归模型的建立

若因变量Y与解释变量X₁，X₂，X₃，X₄……具有线性关系，它们之间的线性回归模型可表示为：

$$Y=b_0+b_1X_1+b_2X_2+L+b_kX_k+\mu$$

其中μ为随机扰动项观测值。对于第i个观测值：

$$Y=b_0+b_1X_{1i}+b_2X_{2i}+L+b_kX_{ki}+\mu\quad i=1,2K\ n$$

即：

$$\begin{bmatrix} Y_1 \\ Y_2 \\ M \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{21} & L & x_{k1} \\ 1 & x_{12} & x_{22} & L & x_{k2} \\ L & L & L & L & L \\ 1 & x_{1n} & x_{2n} & L & x_{kn} \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ M \\ b_k \end{bmatrix} + \begin{bmatrix} \mu_1 \\ \mu_2 \\ M \\ \mu_n \end{bmatrix}$$

也即：Y=Xb+μ

假定：

$$E(\mu_i)=0$$

$$Var(\mu_i)=E(\mu_i^2)=\sigma_u^2 \quad i=1,2L,n$$

$$Cov(\mu_i,\mu_j)=0 \quad i\neq j, \quad i,j=1,2L,n$$

$$Cov(x_j,\mu_i)=0 \quad j=1,2L,k \quad i=1,2L,n$$

$$\mu \sim N(0,\sigma_u^2 J_n)$$

(1) 拟合方程的显著性检验

方差分析表：

离差名称	平方和	自由度	均方差
回归	RSS	K	RSS/k
残差	eSS	n-k-1	eSS/n-k-1
总离差	TSS	n-1	

检验：与解释变量 x_1, x_2, L, x_k 之间的线性关系是否显著。

$$H_0: b_1 = b_2 = L = b_k = 0$$

$$1. \quad H_1: b_i \text{不全为} 0 \quad (i=1, 2, L, k)$$

$$F = \frac{RSS/k}{ESS/n-k-1} \sim F(k, n-k-1)$$

$$2. \quad \text{或} F = \frac{R^2/k}{(1-R^2)/(n-k-1)} \quad k: \text{解释变量个数}$$

- 查表得： $F_\alpha(k, n-k-1)$
- 若 $F > F_\alpha$ ，拒绝 H_0 ，回归方程显著
 $F < F_\alpha$ ，接受 H_0 ，回归方程不显著

	SVR 模型	神经网络模型	线性回归模型
均方误差	0.0328	0.0308	0.0291

根据三个模型的均方误差可得，线性回归模型的均方误差最小，因此选择线性回归模型作为预测得出率的模型。

具体过程为：通过问题一中所得到的主要影响因素建立第一个回归模型，其中用于训练的数据为炼钢数据集中 C 和 Mn 均不缺失的数据。接着利用第一个回归模型计算得到炼钢数据集中缺少转炉终点 Mn 数据的 Mn 得出率得到数据集的扩

大。根据扩大的数据集建立第二个回归模型计算同时缺少连铸正样 C 和 Mn 数据的元素收得率，进而获得所有数据的元素收得率。

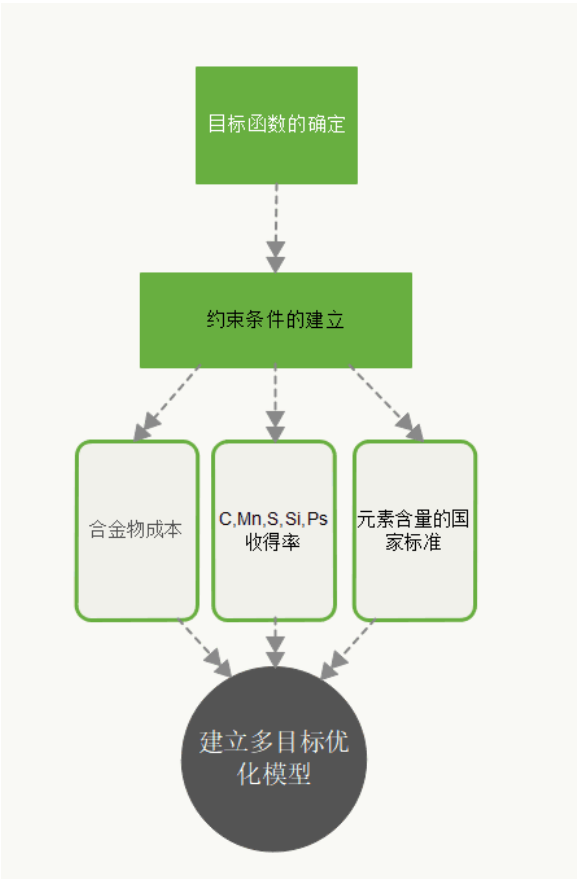
在分析准确率时我们发现数据存在一定的误差，通过对已有模型的检查我们发现在建立第二个回归模型时利用扩大的数据集造成了误差的积累。因此，在建立第二个回归模型时改用炼钢数据集中 C 和 Mn 均不缺失的数据进行训练，获得元素收得率较准确的预测模型。

7 模型三的建立与求解

7.1 对问题三的分析

问题三要求通过选择不同的合金料，然后根据问题 2 中的合金收得率的预测结果以及附件，建立数学模型，实现钢水脱氧合金化成本优化计算，并给出合金配料方案，我们将问题转化为成本最小，元素的收得率最大，合金的元素含量符合国家标准的三目标优化问题。

为了方便讨论我们按钢号对数据集进行分割，将“HRB400B”，“HRB400D”，“HRB500D”，“20MnKA”，“HRB500B”，“20MnKB”，“Q345B”，“Q235A”，“Q235”当成 A, B, C, D, E, F, G, H, I 共 9 中模型，为了使得问题的分析适当的简化以及保证模型的合理性，下面我们根据每种的钢铁进行分类讨论。具体过程如图所示。



7.2 模型三的建立

7.2.1 数据预选择

根据国家标准首先对附件 1 的数据进行预选择，从数据中筛选出各元素均符合国家标准的数据作为备选数据集。国家标准如图所示。

钢号	标准	元素含量 (%)				
		C	Si	Mn	P	S
HRB400B	内控区间	0.19-0.25	0.50-0.65	1.30-1.60	≤0.045	≤0.045
	目标值	0.23	0.55	1.45	0.04	0.04

7.2.2 构建目标函数关系

要实现脱氧合金化成本的优化计算，应满足两个条件即合金收得率更高，合金物的成本更低。对于炼钢厂而言，合金物成本和合金收得率应当是实现合金化成本优化时最重要的两个考虑因素。因此，我们合金化成本优化问题转化为双目标优化问题，双目标函数表述如下：

$$\begin{cases} \max \sum_{i=1}^n \eta_{ij} \\ \min \sum_{i=1}^n c_{im} \end{cases}$$

7.2.3 元素的预测模型

根据问题二建立 C、Mn 元素的预测模型方法，进一步建立 C、Mn、S、Si、P 五种元素的预测模型，具体方法如下：

首先根据不缺失数据的 C、Si 数据建立三输出回归模型预测 Mn、S、P 的收得率，过程同问题二的建模过程，所建立的回归模型为：

$$y = 7.50450039e-04 \times x_1 + -4.07878801e-03 \times x_2 + 3.09150606e-02 \times x_3 + -1.38777878e-17 \times x_4 + -4.16333634e-17 \times x_5 + 5.73682840e-03 \times x_6 + 1.38777878e-17 \times x_7 + -3.41841063e-02 \times x_8 + 1.04083409e-16 \times x_9 + 1.58233241e-02 \times x_{10} + -6.63972779e-03 \times x_{11} + -1.21212616e-01 \times x_{12} + -1.26816367e-01 \times x_{13} + -2.20332798e-03 \times x_{14} + -4.77029340e-03 \times x_{15} + 0.888329157662$$

其中

x1 = 转炉终点温度

x2 = 转炉终点 C

- X3 = 钢水净重
 X4 = 钒铁 (FeV50-A)
 X5 = 钒铁 (FeV50-B)
 X6 = 硅铝合金 FeAl30Si25
 X7 = 硅铝锰合金球
 X8 = 硅锰面（硅锰渣）
 X9 = 硅铁(合格块)
 X10 = 硅铁 FeSi75-B
 X11 = 石油焦增碳剂
 X12 = 锰硅合金 FeMn64Si27(合格块)
 X13 = FeMn68Si18(合格块)
 X14 = 碳化硅 (55%)
 X15 = 硅钙碳脱氧剂

根据已测得的五种元素的收得率建立五输出回归模型预测五种元素的收得率，过程同问题二的建模过程，所建立的回归模型为：

$$\begin{cases} y_1 = k_1 x_1 + k_2 x_2 + \dots + k_{15} x_{15} + b_1 \\ y_2 = g_1 x_1 + g_2 x_2 + \dots + g_{15} x_{15} + b_2 \end{cases}$$

其中，

序号	x_i	k_i	g_i
1	转炉终点温度	2.45e-03	-8.31e-02
2	转炉终点 C	3.36e-02	-2.78e-17
3	钢水净重	2.78e-17	-1.61e-02
4	钒铁 (FeV50-A)	1.73e-17	-8.39e-03
5	钒铁 (FeV50-B)	-3.47e-18	3.94e-02
6	硅铝合金 FeAl30Si25	-6.48e-02	-3.80e-02
7	硅铝锰合金球	-4.70e-02	-7.31e-03

8	硅锰面（硅锰渣）	-5.49e-03	7.50e-04
9	硅铁(合格块)	-4.08e-03	3.09e-02
10	硅铁 FeSi75-B	-2.08e-17	-4.16e-17
11	石油焦增碳剂	5.74e-03	2.08e-17
12	锰硅合金 FeMn64Si27(合	-3.42e-02	1.04e-16
13	FeMn68Si18(合格块)	1.58e-02	-6.64e-03
14	碳化硅(55%)	-1.21e-01	-1.27e-01
15	硅钙碳脱氧剂	-2.20e-03	-4.77e-03

7.2.4 成本计算

$$C(x_1, x_2 \dots x_{16}) = x_1 a_1 + x_2 a_2 + x_3 a_3 + \dots x_{16} a_{16}$$

其中，C 表示合金料的成本；

x_1 、 x_2 ... x_{16} 分别对应合金料的质量；

a_1 、 a_2 ... a_{16} 表示各种合金料对应的价格。

7.2.5 构建成本最低函数并计算

$$T = \eta + \frac{1}{C}$$

其中， η 表示收得率，C 表示合金料的成本。

要实现脱氧合金化成本的优化要求脱氧合金化过程中加入的合金成本最小，即要求目标函数 T 取最小值。通过对目标函数求值并对所求值筛选排序后即可获得脱氧合金化优化后的合金配料方案。所求目标函数的值见附件。钢号 HRB400B 的最佳合金配料方案如图 7-10 所示。

8 问题四的分析与回答

8.1 问题四的分析

问题四要求我们根据研究结果给炼钢厂领导写一封信。由于收信人是炼钢厂的领导，所以给他提出的建议应该是我们通过研究所得出的在炼钢脱氧合金化过程中的最佳合金配料方案，使他能够最大程度地减小成本增大利润。此外，应该注意由于炼钢厂领导虽然具有丰富的炼钢经验但是不一定具有相关的数学和计算机基础，因此在向他阐述建议时需要注意语言要通俗易懂。

8.2 问题四的回答

尊敬的厂长：

您好！

您身在此中，必然深谙，随着工业的发展和科技的进步，对钢材的质量和钢铁工业自动化也提出了更高的要求。作为钢铁生产的重要控制环节——炉外精炼，合理的成分控制是保证产品质量、降低生产成本和劳动强度的有效手段。为了实现钢铁行业中高附加值钢种产量的不断提升，如何通过历史数据对脱氧合金化环节建立数学模型，在线预测并优化投入合金的种类及数量，在保证钢水质量的同时最大限度地降低合金钢的生产成本，已然成为各大钢铁企业提高竞争力所要解决的重要问题。而屋漏偏风连夜雨，自从铁矿石协调机制被打破，铁矿石价格居高不下，我国钢铁业已然进入薄利期。是否能从生产低端产品的“产量型”向生产高端产品的“品质型”、“质量型”转变已成为我国钢铁企业能否继续发展的决定因素。应运而生的智能化控制代替人工经验控制能够有效降低劳动强度、生产成本，稳定并提高产品质量，进而推动生产、资源和设备管理的变革并被越来越多的钢铁企业所接受。

我们此番建立的合金化模型的主要功能即是在保证钢水质量的同时，根据来钢成分和目标成分预报合金加入量，控制合金微调，从而优化合金的配料，简洁地解决这个老大难问题。而最小成本模型则是根据炼钢厂现有合金及其所对应的价格选择合金加入种类以获得最小成本，指导炼钢厂的原材料选择。在锰系合金价格急剧上涨的现如今，这无疑具有极强的现实意义和极大的可操作性。充分应用这种智能化控制必能在保证产品质量的同时，更大地提高产品的竞争力，裨补缺漏，有所广益，在优胜劣汰的社会主义市场经济下更好地立于不败之林，从而使企业获得更为长远的效益。

与此同时，我们并不能因为优化配料模型洋洋自得，坐井观天，安于一隅，

而应从动力学等其他方面全面入手,综合提升企业收益。在原料的优化选择方面,我们可以选择碳化硅。因为碳化硅具有更强的脱氧能力,同时产生 CO 气体,给化学反应提供动力学条件,起搅拌作用,极大地提高了化学反应速率,同时促进化学反应正向进行程度。此外,选择价格较为低廉的高碳锰铁替代碳锰铁来进行合金化也不失为一个好办法。

常言道:“磨刀不误砍柴工”。除了原料方面的优化,我们还可以从工艺流程和工业技术方面入手,引进先进工艺,提高工人工作效率和化学反应程度。转炉出钢与 RH 过程脱氧造渣工艺就颇有优势。“工欲善其事必先利其器”,炼钢转炉煤气回收自动控制系统具有控制逻辑完善、检测和执行设备配置合理、上位监控系统智能监控的优点。转炉煤气回收自动控制系统体现了预测技术与自动控制技术的融合,促进转炉煤气回收、利用的稳定运行。既延长了工业产业链,扩大了产业效益,又符合国家低碳节能,环保无污染的要求,可谓一举多得。

在合金优化配料模型的研究中,我们通过回归分析,找出了影响碳、锰元素收得率的主要因素。而炼钢 (steel making) 本就是在控制碳含量 (一般小于 2%) 的情况下,消除 P、S、O、N 等有害元素,保留或增加 Si、Mn、Ni、Cr 等有益元素并调整元素之间的比例,从而获得最佳性能。掌握了它们的主要影响因素,我们就能更好地随心控制各种元素的收得,从而在炼钢过程中如虎添翼。

祝您生意兴隆!

9 误差分析与改善

6.1 误差分析

在仪器采集测量数据时产生的误差,对采集到的数据进行录入时存在的录入错误,以及在进行数据检查与预处理的时候,存在离群点的忽略而造成的数据不全,以及处理数据时保留的小数位数的差别。

6.2 误差的改善

定期升级仪器设备,提高仪器的测量精度。录入采集到的数据时,尽量选择细心并且责任心较强的员工进行此项工作,并指派专人进行检查。在进行数据检查时,及时发现错误,对录入有误的数据进行修正补全,保证数据的完整,处理数据时尽量多保留几位小数,提高数据的精确度。

10 模型的评价与推广

7.1 模型的评价

(1) 从定性和定量的两个角度分析任务的规律,通过图像直观清晰地观察规律,在给出影响因子对规律地影响程度

(2) 利用标准化和归一化的思想, 将数据中的特征对目标值的影响转化到同一尺度上进行评价。在训练模型的时候充分的利用特征之间的关系和现实中的特征的实际意义来进行数据的预测、以及特征的选择问题。

(3) 使用易实现、鲁棒性好的线性回归模型, 可以较为准确的预测各种元素的收得率, 同时模型训练的时间较少。

(4) 采用 One-Hot 编码的方式将钢种进行编码, 充分的利用数据。并查阅相关的文献得到温度对元素的收得率的影响, 从而可以更加准确的预测模型

(5) 关于模型的训练方向, 本模型采用十倍交叉验证的方法, 在尽可能地利用数据的同时有效的避免了模型过拟合问题。

缺点:

(1) 使用模型相对简单, 没有考虑到特征之间更为复杂的关系

7.2 模型的推广

合金化模型可以包括三个功能, 一是根据来钢成分和目标成分预报合金加入量, 控制合金微调; 二是根据合金加入量预报钢水成分, 以便减少工人采样次数, 降低工人劳动强度; 三是计算合金加入后对钢水温度的影响。它具有广泛的应用意义。利用它的主要控制功能综合其他方面的建模与控制不仅可以按机械性能的要求确定最佳控制成分、使用线性规划法确定最佳补加合金方案, 还可以控制渣碱度、控制搅拌气体流量、控制废炉气排量等。除此之外, 合金化模型还可以用于预测各合金元素对钢中沉淀物形成的影响。对于炼钢有着更好的科学建设意义。

参考文献

- [1] 杨凌志. eAF-LF 炼钢工序终点成分控制研究[D]. 北京科技大学, 2015.
- [2] 胡井涛. LF 精炼脱氧合金化模型开发与在线应用[D]. 东北大学, 2011.
- [3] 徐俏. 基于多目标微粒群的转炉合金加入量优化研究[D]. 大连理工大学, 2010.
- [4] 韩敏, 徐俏, 赵耀, 林东, 杨溪林. 基于收得率预测模型的转炉炼钢合金加入量计算[J]. 炼钢, 2010, 26(01): 44-47.

附录

