

第八章 智能制造大数据技术应用实例

目录

8.1 高炉炼铁基本原理	1
8.1.1 高炉炼铁的背景与生产机理	1
8.1.2 高炉操作对高炉状态的影响	2
8.2 高炉参数的相关性分析	3
8.2.1 高炉操作参数间的相关性分析	3
8.2.2 高炉操作参数与状态参数间的相关性分析	5
8.2.3 高炉多参数的自相关性分析	5
8.3 高炉数据预处理	6
8.3.1 高炉数据问题分析	6
8.3.2 高炉数据去噪处理	7
8.3.3 高炉数据规范化	9
8.4 高炉操作参数状态聚类	11
8.5 高炉炉况分类	15
8.6 高炉煤气利用率拟合与预测	16
8.6.1 高炉煤气利用率线性回归分析	17
8.6.2 高炉煤气利用率非线性回归分析	19
8.7 本章小结	24
参考文献	25

第八章 智能制造大数据技术应用实例

随着信息化与工业化的深度融合，智能制造技术渗透到了工业企业产业链的各个环节。在工业生产过程中，工业设备运行所产生、采集和处理的数据量十分庞大、数据类型多种多样。然而，工业中大多数数据仍未得到充分地利用，如何合理有效地运用这些生产数据对提升工业生产过程运行效率十分重要，而智能制造大数据技术的发展和运用，能使工业系统具备描述、诊断、预测、决策、控制等智能化功能。

针对大数据技术在工业场景的应用，本节以钢铁产业中代表高炉生产过程指标的煤气利用率为例，运用本书所介绍的工业大数据分析方法对实际工业生产数据进行分析。本章首先介绍了高炉炼铁的背景与基本原理，然后将相关性分析、数据预处理、聚类分析、分类分析以及回归预测等方法分别应用于高炉生产数据的分析中，并给出了详细的分析结果。

8.1 高炉炼铁基本原理

在钢铁工业生产过程中，高炉冶炼是其中最重要的环节。为了较好地利用高炉生产中的大量数据，熟悉高炉生产的基本原理以及明确高炉不同操作对高炉状态的影响十分关键。

8.1.1 高炉炼铁的背景与生产机理

钢铁产业是国民经济的重要支柱产业，是衡量一个国家工业水平和综合国力的重要指标，对国民经济的可持续发展有着不可忽视的影响。高炉炼铁是钢铁生产过程中的上游工序，是通过高炉中的一系列复杂反应，将铁矿石冶炼，还原成生铁的过程。高炉炼铁的能耗和排放量占整个流程的主要部分，约占吨铁可比能耗的 77%，是生铁冶炼过程最消耗能源的工序。其中 CO_2 的排放占到整个钢铁工序排放量的 90% 以上，并且炼铁过程能耗占钢铁工业总能耗的比例正在持续升高。随着国家对环境、能源、安全等问题的日渐关注，对于高炉炼铁这种传统行业也提出了更高的要求。根据工业和信息化部关于印发《2020 年工业节能监察重点工作计划》（工信部节函〔2020〕1 号）的通知，重点指出要在钢铁、化工等行业，开展能耗限额标准执行情况专项监察。因此，钢铁企业降耗对我国节能减排事业具有深远意义，其中，高炉生产过程的深入分析、高炉操作的合理调节是实现高炉节能减排的重要突破口。

如图 8.1 所示，高炉炼铁的主体设备是高炉，它是一个密闭的逆流式热交换竖炉，用高强度钢板作炉壳，壳内砌耐火砖内衬。在炉内堆积成料柱状的炉料，受逆流而上的高温还原气流的作用，不断地被加热、分解、还原、软化、熔融、滴落，并最终形成渣铁融体而分离。在生产过程中，炉长需要根据不同的高炉冶炼指标去判断冶

炼过程的好坏，并通过调节不同的高炉操作调节炉内的氧化还原反应，将高炉控制在期望的煤气流状态与热状态下，从而保证炉况保持在一定范围内的稳定状态下。但是此类调控方法主要依托于现场工人的人工经验，对高炉内部的特性很难描述清楚。

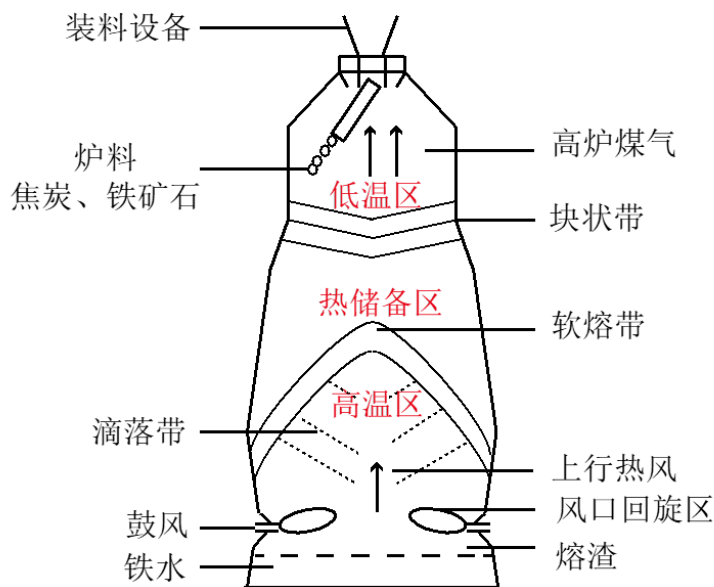


图 8.1 高炉冶炼原理图

高炉炉况运行情况的好坏决定了铁水的出炉质量，因此需要对高炉炉顶布料与炉底送风配合控制，以保持炉况的稳定。高炉煤气流分布与高炉布料和送风操作密切相关：煤气流的分布关系到炉内温度分布、软熔带结构、炉况顺行、煤气的利用状况和高炉长寿，最终影响到高炉冶炼指标。因此，需要明确高炉操作参数与高炉状态参数之间的关系，建立合理的模型对煤气流分布进行调整与控制，并对不同高炉炉况进行相应的分类分析，针对炉况进行特定的调整。

8.1.2 高炉操作对高炉状态的影响

高炉操作主要分为布料操作与送风操作，两种操作从不同角度对高炉状态产生影响。高炉顶部布料主要是将铁矿石、焦炭等一圈一圈从炉顶溜槽布下，在自上而下的过程中与底部煤气流在高温下，经过一系列的反应最终被还原为液态的铁水以及炉渣由炉底铁口和渣口排出，废弃气（烟道气）则从炉顶排出。布料操作通过改变高炉内固体原料分布、改变高炉内部煤气流发展与分布、调控高炉内部燃料比、调整热状态和物理化学反应来影响煤气利用率。布料操作的调节主要是改变布料矩阵，其中包括

改变布料溜槽的角度、布料圈数、料批重量等参数，主要是影响软融带和块状带的煤气流分布。

高炉送风操作主要通过向炉底风口鼓入热风并吹动焦炭，使其回旋燃烧形成初始煤气流。送风主要对高炉内部煤气流初始分布、高炉底部热能的调节、高炉内部压力环境、煤气上升动能和高炉冶炼强度等进行调节。是实现高炉稳定运行的重要调节制度。常见的高炉送风操作参数有风量、风压、富氧等。

在冶炼过程中，由于影响高炉冶炼的过程参数较多，很难确定需要调节的操作参数，而且高炉的生铁冶炼过程是在密闭状态下进行的。因此，过程参数大多不能直接观测，只能间接测量过程的输入输出变量。例如，虽然反应过程中的具体参数难以测量，但能够通过采集炉顶气体的排放含量来推测还原反应的进行程度。高炉煤气利用率能够实时反映煤气利用程度、煤气流分布状态、能源消耗程度以及铁水的质量和产量，是反映高炉整体状态的重要参数。实际情况下，常将高炉煤气利用率（GUR）的大小作为反映高炉冶炼效果的指标之一。在一定范围内，炉煤气利用率越高，代表还原效率越高、气体分布越合理、高质量和低的高炉消耗；而煤气利用率越低，则表示能量利用率低，冶炼状态越差。高炉煤气利用率是二氧化碳含量与一氧化碳和二氧化碳总含量之比。计算公式如下：

$$\eta_{CO} = \frac{V_{CO_2}}{V_{CO_2} + V_{CO}} \quad (8.1)$$

其中， V_{CO} 为一氧化碳含量， V_{CO_2} 为二氧化碳含量。在此本章选择影响高炉煤气利用率的数据作为操作参数，使用煤气利用率作为状态参数。后面章节首先对这些数据进行分析处理，主要包括数据的相关性分析、去噪处理以及归一化。接着，为了明确高炉冶炼状态，往往需要对不同炉况进行分类分析，并在不同炉况下建立合理的模型对煤气利用率进行预测，从而针对炉况进行特定的调整。本章后续的内容主要包括了对高炉炉况的分类、聚类以及煤气利用率的预测。

8.2 高炉参数的相关性分析

高炉冶炼过程是一个多变量的复杂工业过程，高炉数据大部分来源于现场不同传感器的测量值，少部分为人工记录值，过程数据众多且复杂，数据关系也不明确。因此在使用这些数据前，首先需要先对这些数据进行相关性分析。

8.2.1 高炉操作参数间的相关性分析

由于影响高炉煤气利用率的高炉操作参数众多，若将其全部作为模型的输入将会造成数据的冗余，因此需要对数据进行相关性分析。在此采用皮尔逊相关性分析法对

影响煤气利用率的主要参数进行相关性计算。皮尔逊的定义及计算公式参见第二章 2.2.3 节。

通常， C_p （皮尔逊相关系数）的绝对值越大，两个时间序列之间的相关性越大。 C_p 为负值时，两个时间序列为负相关，其为正值时，两个时间序列为正相关。相关性的计算结果如表 8.1 所示，所用数据均为某钢铁厂内体积为 2800 m³ 的高炉连续五个月的现场数据。用于皮尔逊相关性分析的高炉操作参数分别为冷风流量（ X_1 ）、冷风压力（ X_2 ）、热风压力（ X_3 ）、富氧流量（ X_4 ）、富氧压力（ X_5 ）、喷煤量（ X_6 ）、边缘矿焦比 3（ X_7 ）、边缘矿焦比 4（ X_8 ）、中心焦比 7（ X_9 ）以及中心焦比 11（ X_{10} ）。

表 8.1 高炉多操作参数皮尔逊相关性分析结果

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
X_1	1.00	-0.31	-0.33	0.13	-0.27	-0.14	0.25	0.25	0.01	-0.14
X_2	-0.31	1.00	0.99	-0.19	0.97	0.03	-0.002	-0.008	-0.008	0.09
X_3	-0.33	0.99	1.00	-0.02	0.97	0.03	-0.02	-0.03	-0.08	0.11
X_4	0.13	-0.19	-0.21	1.00	-0.08	0.04	-0.11	-0.12	0.08	-0.14
X_5	-0.28	0.97	0.97	-0.08	1.00	0.04	-0.01	-0.02	-0.08	0.08
X_6	-0.14	0.03	0.03	0.04	-0.04	1.00	-0.04	-0.04	0.03	0.03
X_7	0.23	0.01	-0.005	-0.1	0.003	-0.06	1.00	0.93	-0.18	0.20
X_8	0.23	0.01	-0.01	-0.12	-0.001	-0.04	0.93	1.00	-0.21	0.22
X_9	0.02	-0.08	-0.08	0.09	-0.08	0.03	-0.18	-0.21	1.00	-0.69
X_{10}	-0.15	0.11	0.12	-0.15	0.10	0.05	0.220	0.22	-0.69	1.00

由皮尔逊相关性计算结果可知，冷风压力（ X_2 ）、热风压力（ X_3 ）以及富氧压力（ X_5 ）相关性高；边缘矿焦比 3（ X_7 ）与边缘矿焦比 4（ X_8 ）相关性高；中心焦比 7（ X_9 ）与中心焦比 11（ X_{10} ）相关性高。在此，可以从上述 7 个参数中选取热风压力（ X_3 ）以及边缘矿焦比 3（ X_7 ）、中心焦比 11（ X_{10} ），结合冷风流量（ X_1 ）、富氧流量（ X_4 ）、

喷煤量 (X_6)，用于后续实验，以去除冗余变量，减少模型复杂度。

8.2.2 高炉操作参数与状态参数间的相关性分析

为了选取影响高炉煤气利用率大小的关键操作参数，在此，把 8.2.1 节中得到的高炉参数与状态参数煤气利用率做进一步相关性分析，相应结果如下所示：

表 8.2 高炉操作参数与状态参数皮尔逊相关性分析结果

	X_1	X_3	X_4	X_6	X_7	X_{10}
煤气利用率	0.42	0.16	0.04	-0.05	0.12	-0.12

通过状态参数的相关性分析，可以挑选出影响高炉煤气利用率的各个关键操作参数。如表 8.2 所示，在此可以选用冷风流量 (X_1)、热风压力 (X_3)、边缘矿焦比 3 (X_7) 与中心焦比 11 (X_{10}) 这 4 个具有较大相关性的参数作为后续实验所需的操作参数。

8.2.3 高炉多参数的自相关性分析

高炉炼铁生产状态参数具有一定的自相关性，即前一段时间的系统状态会对之后的状态产生明显的影响，因此有必要对这种自相关性进行分析，自相关函数的计算公式详见 2.3.3 节。

煤气利用率自相关系数计算结果如图 8.2 所示，从图中可以发现随着时滞时间的延长，煤气利用率的 ACF（自相关系数）值从 1 开始逐步下降，取阈值水平为 0.6 所对应的前 7 个历史煤气利用率历史数据作为预测模型的部分输入，从而提高预测模型的预测效果。

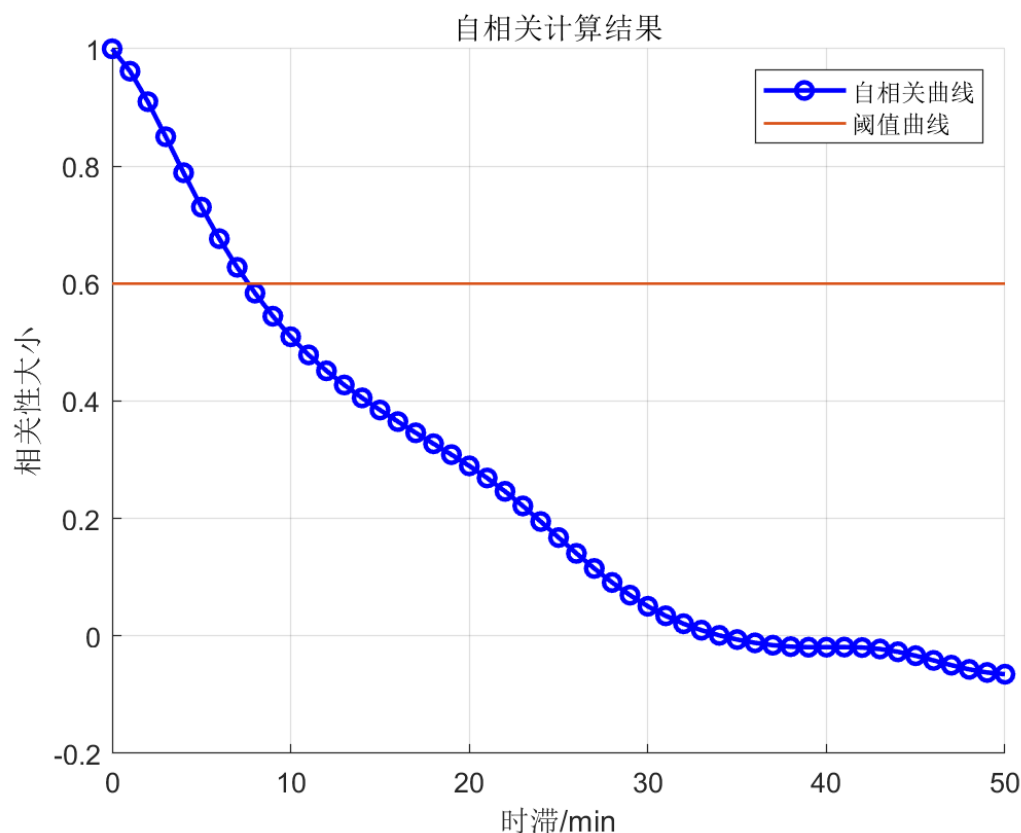


图 8.2 高炉煤气利用率自相关分析结果

8.3 高炉数据预处理

高炉是一个典型的“黑箱系统”，其生铁冶炼过程是在密闭状态下进行的，过程参数大多不能直接观测，只能间接测量过程的输入输出变量。并且冶炼过程涉及多种物理、化学变化，包括“固-液-气”之间的转化以及能量的流动，是一个非线性、大耦合、多变量的高难度冶金过程，参数测量和数据采集容易受到过程状态和环境因素影响。因此，需要对高炉现场数据进行质量分析和预处理，使得处理后的数据符合后续建模要求。

8.3.1 高炉数据问题分析

高炉炼铁生产过程中产生的数据不仅具有高维度、大体量、高价值的大数据共有特征，而且具有其自身的典型数据特点。由于生产过程的复杂性，高炉冶炼过程数据往往存在以下问题：

1) 高炉炼铁属于连续性生产过程，其物质流、能量流、信息流呈有序化流动状态，一定时间段内的工序操作对设备的影响具有持续性，因此所产生的数据也具有一

定的时序性特点，而对这种时间序列的合理分析是建立准确高炉模型的关键步骤。

2) 高炉炼铁是一个极为复杂的生产工艺，各种传感装置长期处于高温、腐蚀等恶劣的生产环境，传感装置本身所具有的敏感性极易受到外部条件的干扰，因此会带来大量数据缺失、测量不准等问题。同时，大多数测量装置也不能做到实时测量。这些问题给高炉数据的处理筛选带来一定困难。

3) 高炉炼铁过程中涉及相当多的物理化学反应，例如在布料过程中，炉料（矿石、熔剂、焦炭）按照确定的比例通过装料设备分批从炉顶装入炉内，高温热风从下部风口鼓入与焦炭反应生成高温还原性煤气，炉料在下降过程中被加热、还原、熔化、造渣，最后生成液态渣铁聚集在炉缸并周期性地从高炉排出；煤气流上升过程中，温度不断降低，成分逐渐变化，最后形成高炉煤气从炉顶排出。多种物理化学反应同时进行，交叉影响生产过程，使得高炉炼铁数据的复杂性以及耦合性显著提高。

高炉生产过程中不同参数的数据量级具有很大差别，并且由于现场环境影响，高炉数据包含多种噪声。因此，对高炉数据进行去噪是首要任务。

8.3.2 高炉数据去噪处理

由于现场采集到的数据包含多种噪声的影响，因此需要对各参数数据的有效信号进行提取，以减少噪声数据对后续研究的影响。

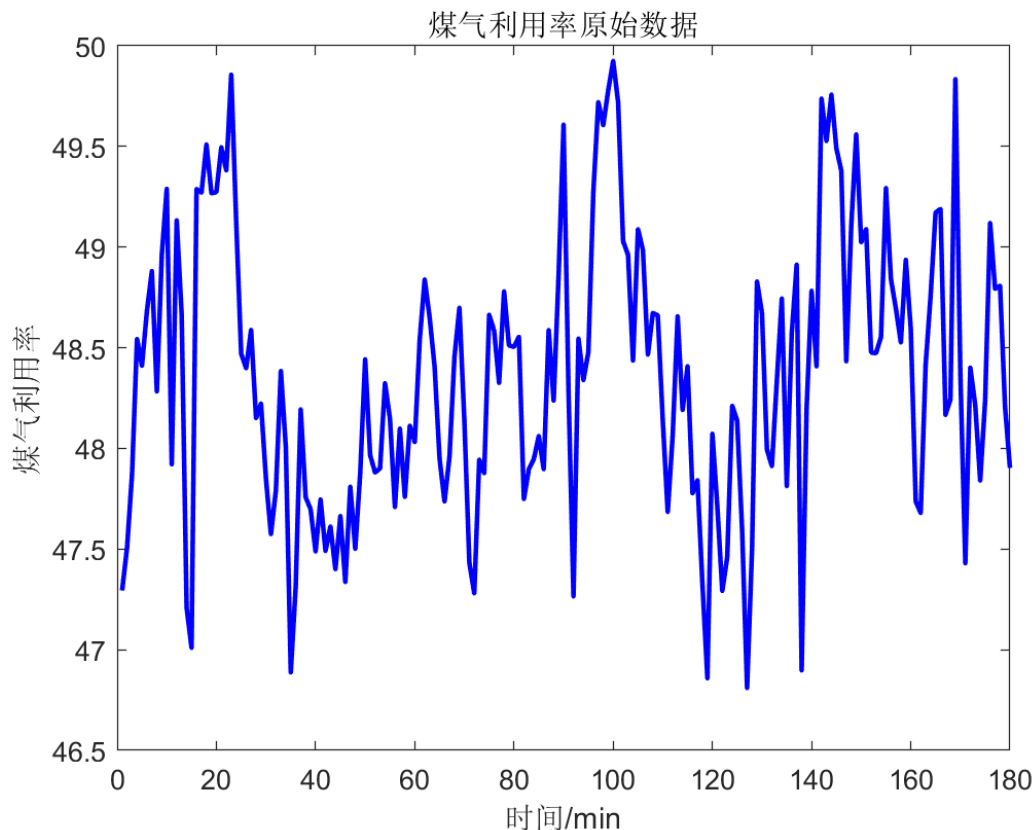


图 8.3 高炉煤煤气利用率原始图

图 8.3 给出了高炉煤气利用率的原始数据，可以发现具有明显噪声波动特征，这是由于高炉现场环境复杂，以及换炉和周期性布料等操作，会在高炉检测信息中引入大量噪声。高炉信号中的噪声不是平稳的白噪声，且存在多尖峰或突变，将对后续实验产生巨大影响。小波分析能够将信号中各种不同的频率成分分解到互不重叠的频带上，可以保留具有重要意义的信号奇异点，对于像高炉中的这种非平稳信号具有良好的噪声处理能力。因此，这里选择小波分析去噪方法对高炉现场数据进行预处理。

从高炉现场采集多个参数的数据样本序列，包括冷风流量 (X_1)、热风压力 (X_3)、边缘矿焦比 3 (X_7)、中心焦比 11 (X_{10}) 与煤气利用率，并且分别进行去噪处理。以煤气利用率为例，对数据进行小波分析去噪。小波分析的定义及计算公式参见第三章 3.2.3 节。

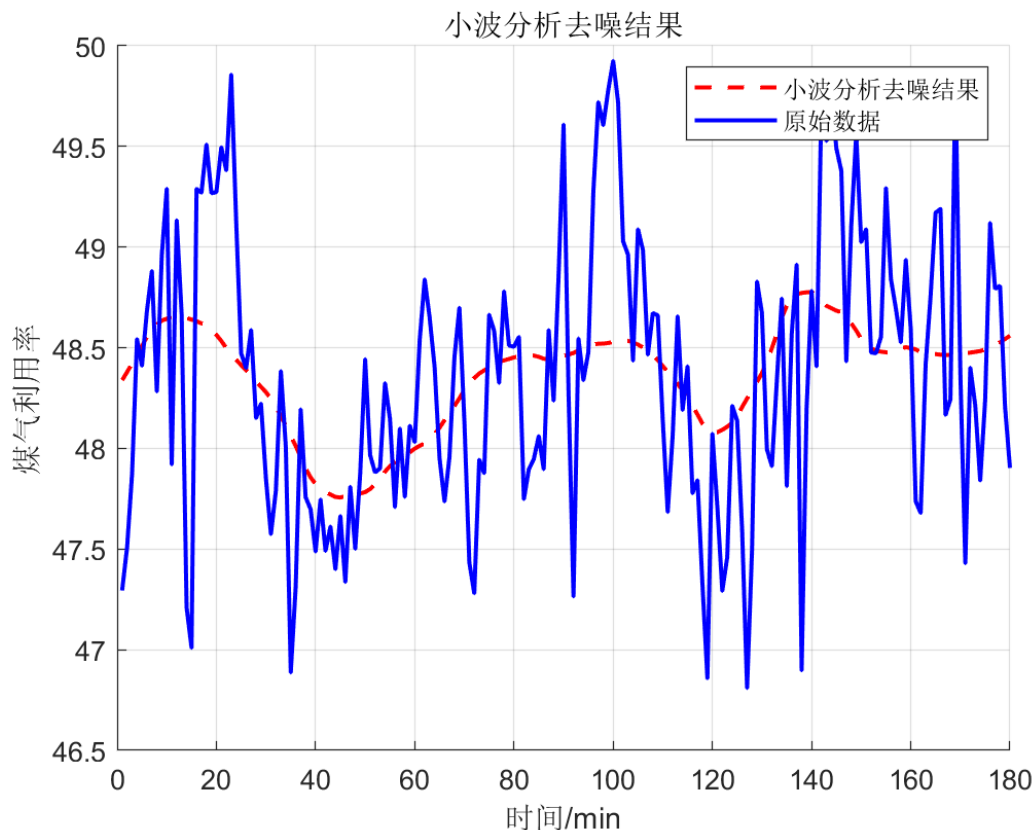


图 8.4 高炉煤气利用率的小波分析去噪结果

图 8.4 是高炉煤气利用率的小波分析去噪结果。由图可知，小波分析能够很有效地去除高炉现场数据的尖峰和毛刺，使得高炉煤气各参数曲线更加圆滑，，并保留数据原有的总体变化趋势。

8.3.3 高炉数据规范化

数据规范化处理是数据挖掘的一项基础工作。不同高炉数据往往具有不同的量纲。为了消除指标之间的量纲影响，需要进行数据规范化处理，以解决数据指标之间的可比性。原始数据经过数据规划化处理，各指标处于同一数量级，适合进行综合对比评价。在此采用最小最大（Min-Max）规范化方法对原始数据进行线性变换，使结果值映射到[0, 1]区间。规范化方法的原理以及基本步骤具体参见第三章 3.4.2 节。高炉布料与送风参数归一化后的部分数据如表 8.3 与表 8.4 所示：

表 8.3 送风操作相关参数规范化结果

序号	冷风流量	热风压力	煤气利用率
1	0.9046	0.7267	0.6294

2	0.6604	0.8495	0.2833
3	0.6992	0.7357	0.1815
4	0.5471	0.7527	0.1371
5	0.5039	0.6613	0.2091
6	0.4025	0.4991	0.3318
7	0.3590	0.7052	0.5577
8	0.9154	0.8674	0.6981
9	0.7698	0.8764	0.8091
10	0.8409	1	0.3479
11	0.8050	0.8656	0.7565
12	0.8035	0.7303	0.5988
13	0.3364	0	0.1088
14	0.5435	0.9319	0.0412
15	0.6736	0.6971	0.8034
16	1	0.6030	0.8785
17	0.8625	0.8898	0.8398
18	0.6797	0.9615	1
19	0.4768	0.8602	0.7535
20	0.3964	1	0.5334
21	0.2892	0.8396	0.5090
22	0.7829	0.6694	0.5044
23	0.8730	0.8271	0.3773
24	0.7723	0.9758	0
25	0	0.2975	0.2895

表 8.4 布料操作相关参数的规范化结果

序号	矿焦比 3	中心焦比 11	煤气利用率
1	0.7690	0.4122	0.6294
2	0.6519	0.3916	0.5577
3	0.5163	0.4006	0.5130
4	0.6578	0.3626	0.6047
5	0.7906	0.4005	0.6716
6	0.5968	0.4027	0.4699
7	0.6614	0.3815	0.6981
8	0.5257	0.3796	0.8091
9	0.8802	0	0.3479
10	0.6448	0.3925	0.7565
11	0.6423	0.3713	0.5988
12	0.8192	0.4484	0.1088
13	0.6612	0.4559	0.8034
14	0.5491	0.4274	0.8785
15	0.5026	0.4675	0.8398

16	0.2853	0.4213	1
17	0.3364	0.4642	0.7535
18	0	0.4169	0.5334
19	0.5010	0.5862	0.5090
20	0.5125	0.9444	0.5731
21	1	0.9049	0.3044
22	0.9953	0.9405	0.5044
23	0.7602	1	0.3773
24	0.5458	0.5344	0
25	0.4039	0.4667	0.1439

8.4 高炉操作参数状态聚类

煤气利用率的生成是通过下降的炉料和上升的煤气流相互接触实现的。因此，煤气利用率的高低主要受到炉顶的布料操作与炉底的送风操作的影响。布料操作通过改变高炉内固体原料分布、高炉内部煤气分布以及调控高炉内部燃料比、调整热状态和物理化学反应来影响煤气利用率。该操作对高炉炉况控制作用大，对高炉生产状态的调节具有长时间尺度的滞后作用，通常其滞后时间达到 6~7 个小时。高炉底部风口区送风操作可以对高炉内部煤气流初始分布、高炉底部热能的调节、高炉内部压力环境、煤气上升动能和高炉冶炼强度等调节迅速。该操作对高炉煤气流状态和热状态的控制速度快，对高炉生产状态的调节具有短时间尺度的滞后作用。基于操作参数预测煤气利用率时，如果不区分时间尺度很难得到较好的预测结果。因此，在进行高炉煤气利用率预测时，可以利用聚类方法分别处理布料与送风操作参数，为后续高炉煤气利用率的多尺度预测建模做准备。

选取 8.3 节对高炉操作参数规范化后的冷风流量 (X_1)、热风压力 (X_3)、边缘矿焦比 3 (X_7)、中心焦比 11 (X_{10}) 以及煤气利用率作为聚类算法的输入。

根据得到的规范化数据，在此采用基于 K-means 均值聚类的操作参数状态分类方法，将送风与布料操作参数进行聚类，该算法基本步骤及原理如第五章 5.2.1 所示。

K-means 聚类算法属于无监督的聚类算法，基于给定的现场采集到的数据点，按照数据点之间距离大小，将这些数据点划分为 k 个簇。每个簇内数据点之间的距离尽可能的小，而两个不同簇之间的距离尽可能的大，以实现将数据点划分的目标。

算法首先随机从数据集中选取 k 个点作为初始聚类中心， k 的大小与输入数据的维度有关。然后计算各个样本到聚类中心的距离，通过两者之间的距离大小把样本划分到离对应聚类中心最近的类。计算新形成的每个聚类的数据对象的平均值来得到新的聚类中心，调整到相邻两次的聚类中心没有任何变化为止。

聚类数的大小对聚类的结果存在较大影响。也就是说， k 的选择会较程度地影响分类效果。在聚类之前，需要预先设定分类数 k 的大小，但是对于数据特点不明显的数据集而言，很难确定将数据分成几类是最佳的。在此，利用肘部法确定最佳聚类数，相应方法的基本原理与实现步骤如第五章 5.2.1 所示。不同操作参数的最佳聚类数如图 8.4 与图 8.5 所示：

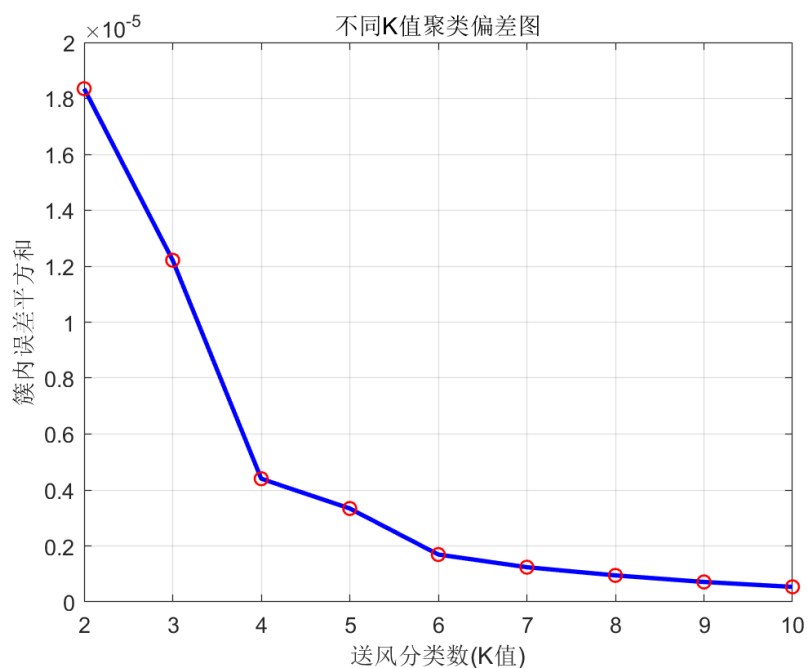


图 8.5 送风操作参数肘部法聚类结果

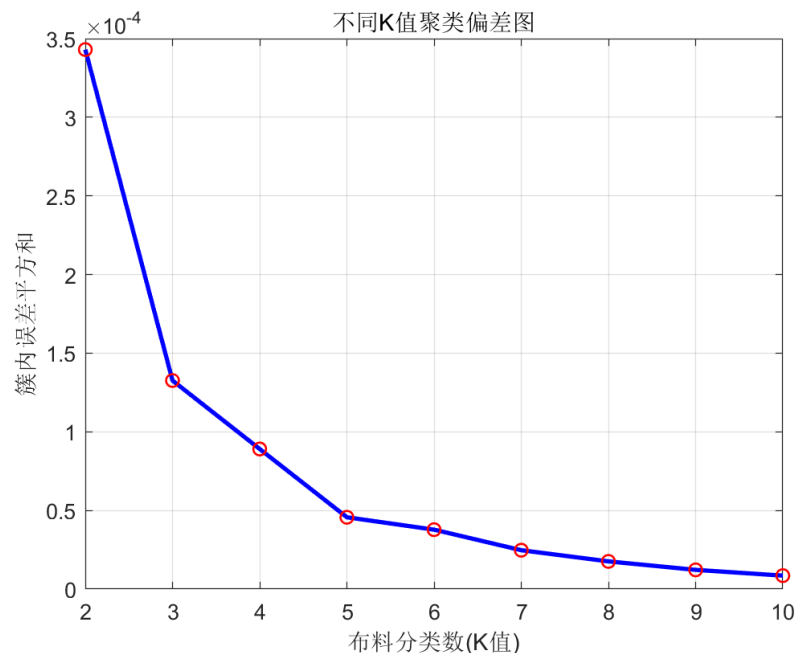


图 8.6 布料操作参数肘部法聚类结果

由图 8.5 可知，当 k 值为 4 时，聚类数所代表的点两边的斜率差距是最大的。因此，结合送风操作与布料操作聚类的实际需求后，采用的 $k=4$ 作为高炉送风与煤气利用率的类别数。如图 8.6 可知，当 k 值为 5 时，聚类数所代表的点两边的斜率差距是最大的，因此采用的 $k=5$ 作为高炉布料与煤气利用率的类别数。两种不同操作下所得到的聚类结果如图 8.7 与图 8.8 所示。

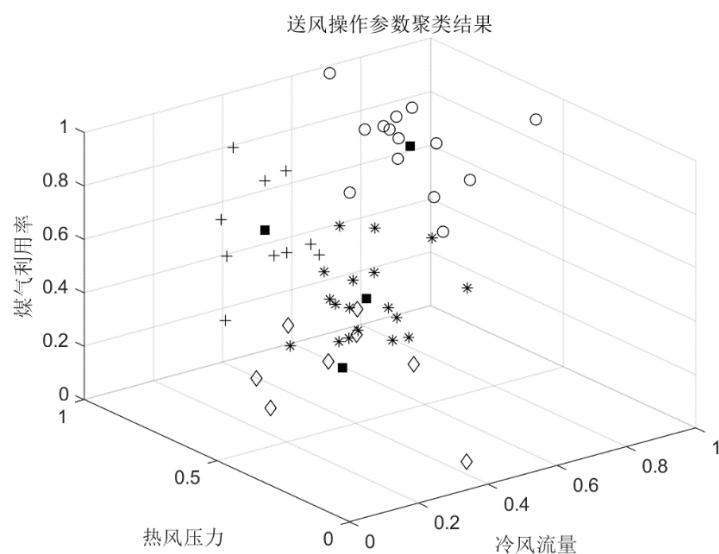


图 8.7 送风操作参数聚类结果

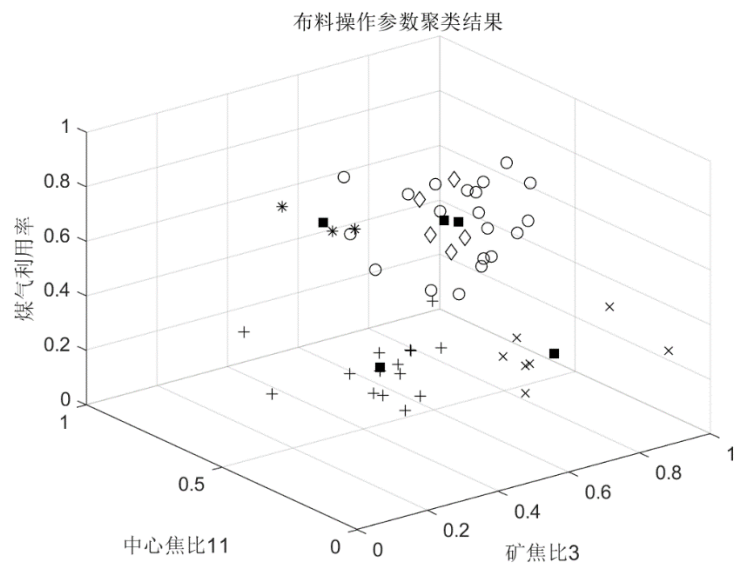


图 8.8 布料操作参数聚类结果

表 8.5 送风操作参数聚类结果

簇	聚类中心	送风操作参数聚类结果
1	(0.7941, 0.2987, 0.6616)	{ 1,9,12,13,15,16,20,21,22,23,24,25,26,36 }
2	(0.4056, 0.8080, 0.4178)	{ 6,17,31 ,32,43,44,47,48 }
3	(0.7982, 0.8483, 0.7579)	{ 2,3,4,5,11,14,18,33,34,35,37,38,39,41,42,45,49,50 }
4	(0.2803, 0.2379, 0.5604)	{ 7,8,10,19,27,28,29,30,40,46 }

表 8.6 布料操作参数聚类结果

簇	聚类中心	布料操作参数聚类结果
1	(0.8275, 0.6443, 0.6118)	{ 3,4,5,6,14,17,18 }
2	(0.4019, 0.9676, 0.3532)	{ 30,31,37 }
3	(0.9655, 0.4232, 0.4398)	{ 1,7,8,9,10,11,12,13,15,16,19,20,21,22,23,24,25,26,27,29 }
4	(0.9405, 0.1939, 0.4586)	{ 2,28,38,39,40,41,42,43,44,45,46,47,48,49,50 }

布料操作、送风操作与煤气利用率聚类后得到的各簇信息如表 8.5 与表 8.6 所示，表格最右侧聚类结果中的数值表示表 8.3 与表 8.4 中的样本序号。K-means 聚类算法能够对高炉在不同操作下的参数进行合理聚类。在高炉冶炼过程中，炉况是实时变化的，在短时间内，各参数的变化范围较小，而在长时间下，各参数的变化范围较大。根据聚类结果可得，高炉送风与布料操作分别与煤气利用率聚为多个不同类别，减少了由于炉况波动而导致的类间数据变化过大对炉况调节造成的影响。在区分类别的情况下，分段预测炉内状态参数往往能得到较好的结果。

8.5 高炉炉况分类

高炉炉况主要通过对不同高炉参数判断，通过在不同范围内对高炉操作参数数据进行区域划分，将其划分为正常状态与非正常状态，从而对其进行简单的炉况二分类。在此采用的数据集包括高炉 4 个参数，在规定范围内先将数据集打上标签。在此采用的训练集与测试集长度分别为 20 个、6 个。

表 8.7 与 8.8 是高炉炉况分类所需的数据训练集。为了简化问题，这里采用高炉操作参数原始数据的“0-1”值，来判断高炉炉况。表 8.9 是高炉分类所需测试机集以及相应的测试结果。

表 8.7 高炉分类训练数据集

训练集	X_7	X_{10}	X_1	X_3	高炉炉况
1	0	1	1	0	1
2	1	1	1	0	1
3	0	1	1	0	1
4	0	1	0	1	1
5	1	1	0	0	1
6	0	1	1	0	1
7	1	1	0	0	1
8	0	1	1	0	1
9	1	1	1	0	1
10	1	0	0	1	1

表 8.8 高炉分类训练数据集

训练集	X_7	X_{10}	X_1	X_3	高炉炉况
1	0	0	0	0	0
2	0	0	0	1	0
3	0	1	0	0	0

4	1	1	0	0	0
5	1	0	0	1	0
6	0	1	0	0	0
7	1	0	0	0	0
8	0	1	0	1	0
9	0	0	0	1	0
10	1	0	1	0	0

表 8.9 高炉分类测试数据集

测试集	X_7	X_{10}	X_1	X_3	高炉炉况（真实值）	高炉炉况（预测值）
1	0	0	0	0	0	0
2	1	1	0	0	1	1
3	0	1	0	0	0	0
4	1	1	1	0	1	1
5	1	0	0	1	0	0
6	0	1	0	1	1	0

在二元（0-1）分类问题中，当将其中一类（例如 0 类）标记为正类时，那么剩余的一类（例 1 类）则标记为负类。类推到多分类问题中，则是将每个类别自身标记为正类，而自身之外的类别标记为负类。针对正负类，TP、FP、TN、FN 四个概念经常出现在分类问题中。分类模型的评价指标如第六章 6.6.2 所示。

每种指标均有其本身针对的涵盖面。准确率是对样本分类整体精度的描述，能够直接反映总样本中被正确分类的样本的占比。精准率和召回率是对样本分类中细节精度的描述，而 F 度量则是对分类问题中细节的权衡考量。

表 8.10 高炉朴素贝叶斯分类器评价指标

准确率	精准率	召回率	F 度量
5/6	2/3	2/2	4/5

对选取的朴素贝叶斯分类器进行模型训练，并利用高炉分类测试数据集数据进行分类测试，分类结果如表 8.9 所示，分类评价指标如表 8.10 所示。由表所得，高炉朴素贝叶斯分类器的各项指标都比较高，反映了选取的高炉朴素贝叶斯分类器能满足高炉煤气利用率的基本分类要求，能够较好地对高炉的炉况进行二分类。

8.6 高炉煤气利用率拟合与预测

目前，基于高炉煤气利用率预测模型主要分为机理模型和数据驱动模型。高炉机理模型主要是研究高炉内部的化学反应，结合传热学、流体力学等学科来建立可以描述高炉内部反应的偏微分方程组并进行求解。方程组的建立主要是通过分析煤气流在

高炉内部的分部、煤气流指标含量以及煤气流和炉料直接的传热关系建立。然而，对于模型的分析大多从仿真以及理论分析的角度出发的，是一些理想化的模型，假设条件过多。而且高炉是一个极其复杂的逆流反应容器，在实际生产过程中，炉内会出现波动或者异常情况。另外，机理模型也需要大量的计算，难以实现在线求解，因此机理建模在实际生产过程中很难有有效的指导意义。

随着高炉中检测方法的不断提高和人工智能的不断发展，基于数据驱动的高炉建模也越来越受关注，由于检测方法的提高，高炉运行中大量的数据也被记录下来，比如，风量、风压、富氧率、喷煤、炉顶压力、煤气中各种成分的含量以及透气性指数（指高炉冶炼过程中风量与全压差（即热风压力和炉顶压力之差）之比，表示炉子接受风量的能力，是一个可以快速、直观、综合反映炉况的重要参数）等参数。另外还有布料的一些相关数据，比如，铁矿石和焦炭的成分等均被记录下来，这为数据驱动建模提供了基础。并且，虽然随着检测技术的发展，高炉内部的数据可以越来越完整的获取到，很便于对高炉进行数据驱动的建模。在此选择使数据驱动中的最小二乘法和支持向量回归法对高炉煤气利用率进行建模、预测。

在高炉煤气调节以及故障检测中，不但需要对高炉某段时间内的数据进行趋势分析，也需要使用历史数据对煤气利用率进行预测，从而根据煤气利用率的变化对现有炉况状态进行相应的判断，进而通过改变相应的高炉操作使高炉保持在一个稳定的炉况范围内。针对此种情况，需要建立高炉煤气利用率的预测模型，这里分别采用最小二乘法和支持向量回归法，建立的高炉煤气利用率的线性和非线性两种预测模型。为了详细分析以上预测模型的效果，将采用第七章 7.2.2 节中的模型评价指标。

8.6.1 高炉煤气利用率线性回归分析

对于线性回归问题，在此采用最小二乘回归的方法进行拟合。拟合问题是指给定平面上 n 个点 $(x_i, y_i) (i = 1, \dots, n)$ 寻求一个函数（曲线） $y = f(x)$ ，使 $f(x)$ 在某种准则下与所有数据点最为接近。而问题首要关键的是确定 $f(x)$ 表达式的形式。

所用数据均为某钢铁厂内体积为 2800 m^3 的高炉连续五个月的现场数据。相应的处理后的数据集 (x_i, y_i) ，其中 x_i 、 y_i 分别为冷风流量（ X_1 ）与煤气利用率归一化后的值如表 8.11 所示。冷风流量（ X_1 ）与煤气利用率之间具有非线性关系，用二次函数形式描述如下：

$$f(x) = a_1 x^2 + a_2 x + a_3 \quad (8.2)$$

式中， a_1 、 a_2 与 a_3 为各项系数，模型构建的目标即是估计出这里的三个系数。

表 8.11 高炉煤气利用率原始数据集

x_i	0.4896	0.4994	0.5142	0.5341	0.5592	0.5886	0.6216	0.6571	0.6930	0.7283
-------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------

y_i	0.7994	0.7952	0.7896	0.7868	0.7888	0.7902	0.7887	0.7809	0.7708	0.7606
-------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------

这里采用最小二乘法估计未知系数 a_1 、 a_2 与 a_3 ，由于式 (8.2) 中存在一个二次项，因此，需要求出 x_i^2 并作为一个整体项，与 x_i 、 y_i 代入式 (8.2) 的最小二乘拟合。求解出的三个系数结果如表 8.12 所示。

表 8.12 二项式方程最小二乘拟合结果

参数	a_1	a_2	a_3
结果	-0.5551	0.5448	0.6601

将表 8.12 中求解得到的二次多项式系数代入 $f(x)$ 中，得到高炉煤气利用率的预测模型为

$$f(x) = -0.5551x^2 + 0.5448x + 0.6601 \quad (8.3)$$

将 x_i 的所有值代入式 (8.3) 求得对应的拟合值，高炉煤气利用率最小二乘回归拟合结果和真实值的比较如表 8.13 和图 8.9 所示。

表 8.13 高炉煤气利用率最小二乘拟合结果

x_i	0.4896	0.4994	0.5142	0.5341	0.5592	0.5886	0.6216	0.6571	0.6930	0.7283
y_i	0.7994	0.7952	0.7896	0.7868	0.7888	0.7902	0.7887	0.7809	0.7708	0.7606
拟合值	0.7938	0.7937	0.7935	0.7927	0.7912	0.7885	0.7843	0.7784	0.7711	0.7624

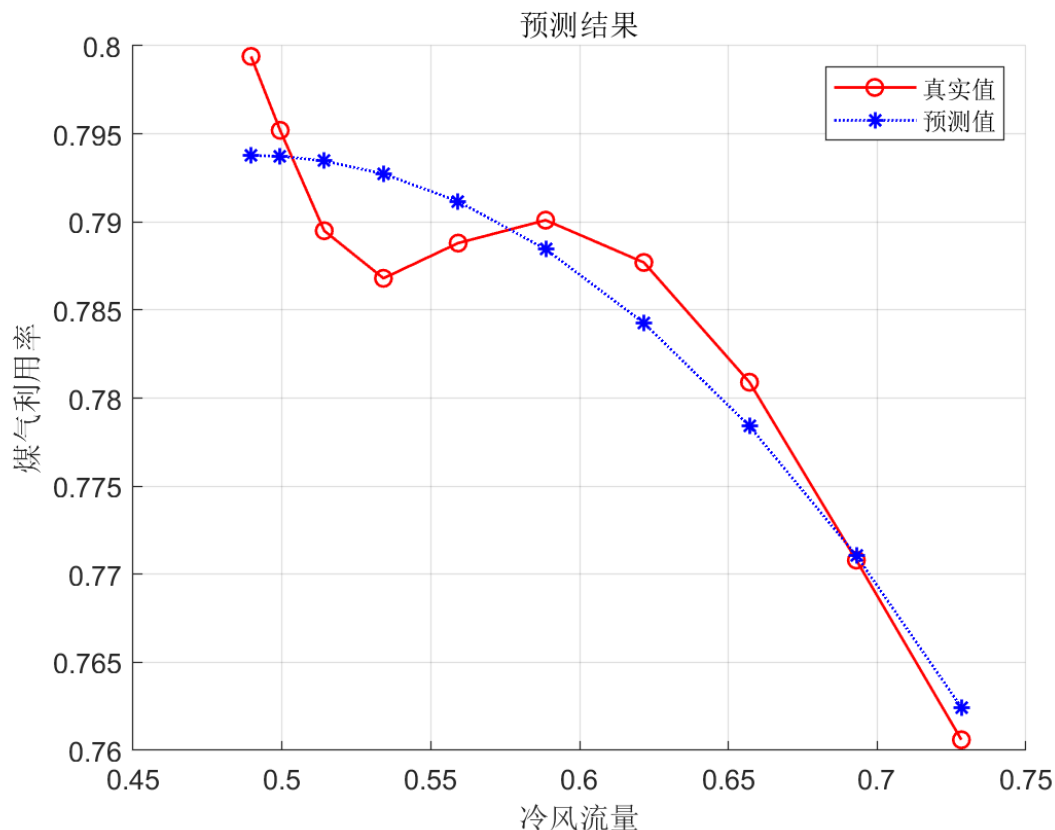


图 8.9 高炉煤气利用率最小二乘拟合图

由图 8.10 可知，该最小二乘模型能够较好的对煤气利用率进行回归拟合。最后，利用给出的指标进行模型性能计算，相应结果如表 8.14 所示：

表 8.14 高炉煤气利用率最小二乘拟合指标		
决定系数	标准差	均方根误差
89.41%	0.0034	0.0034

由此结果可知，该回归拟合模型大体上能够对一段时间内的煤气利用率进行拟合，基本能够达到预期效果。

8.6.2 高炉煤气利用率非线性回归分析

高炉冶炼过程中参数众多，反应复杂，多种参数相互影响，使得高炉炼铁过程呈现出大非线性特性。因此，线性回归方法不适用。需要针对其非线性特性来选择预测方法。并且高炉数据样本仅仅为小样本数据，没有太多的数据用于学习与训练，因此。

本节采用支持向量回归方法处理。

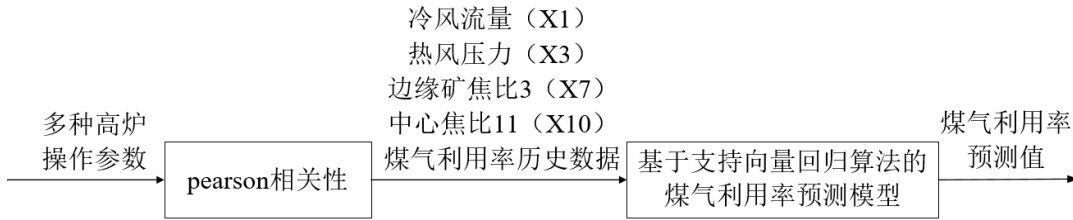


图 8.10 高炉煤气利用率预测方框图

图 8.10 是高炉煤气利用率整体预测方框图。基本思路为将经过相关性分析选择的多种高炉操作参数作为回归模型的输入，从而对未来时间的煤气利用率大小进行预测。在此，模型的输入为由相关性分析得出的对煤气利用率的变化具有较大相关性的高炉操作变量。根据 8.2 节的相关性分析结果，将利用冷风流量 (X_1)、热风压力 (X_3)、边缘矿焦比 3 (X_7)、中心焦比 11 (X_{10})，以及煤气利用率的 7 个历史值作为预测模型的输入，煤气利用率作为回归模型的输出，且用于模型训练的数据都经过 8.2 节中的去噪和规范化处理。在训练煤气利用率预测模型时，需要设置 3 个煤气利用率模型超参数： c ， ϵ 和 g ，如第七章 7.4.3 中节中所述。为了说明这 3 个参数对预测模型的影响，下面分别对这 3 个参数进行测试。在此所使用的数据均为某钢铁厂内体积为 2800 m^3 的高炉现场数据，模型的训练集数据为 400 个，测试集个数为 95 个。

c 的测试结果如表 8.15 与图 8.11 所示：根据预测结果和误差，当 c 值增大时，预测结果的绝对误差、均方根误差先逐渐减小然后增大，而决定系数（见第七章 7.4 节）是先增大后减小。决定系数作为模型的性能指标，可以代表回归直线对样本数据的拟合程度。决定系数的取值在 0 到 1 之间，大体上说明了回归模型所能解释的因变量占因变量总变化的比例。决定系数越高，越接近于 1，模型的拟合效果就越好，即模型解释因变量的能力越强。否则，越接近于 0，回归直线拟合效果越差。综合考虑可以得到 c 取值为 50 是合适的。

表 8.15 煤气利用率预测模型 C_{SVR} 不同时的标准差、均方根误差与决定系数

c	20	30	40	50	60	70
标准差	0.0077	0.0073	0.0069	0.0066	0.0066	0.0067
均方根误差	0.0078	0.0074	0.0071	0.0069	0.0070	0.0072
决定系数	98.89%	98.99%	99.08%	99.12%	99.11%	99.05%

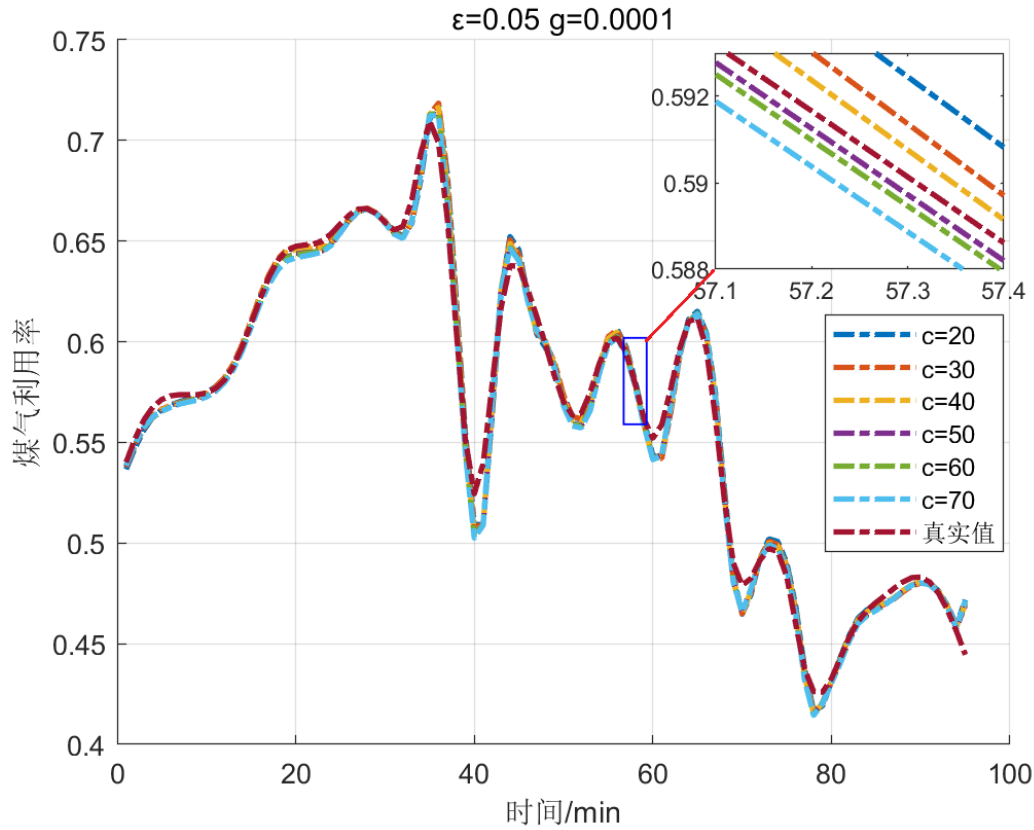


图 8.11 高炉煤气利用率预测模型 c 不同时预测结果

对于参数 ϵ ，其预测结果及误差计算如表 8.16 与图 8.12 所示。根据预测结果和误差，当 ϵ 值增大时，预测结果的绝对误差、均方根误差先逐渐减小然后增大，决定系数逐渐增大然后减小。综合考虑可以得到 ϵ 取值为 0.05 是合适的。

表 8.16 煤气利用率预测模型 ϵ 不同时的标准差、均方根误差与决定系数

ϵ	0.0001	0.001	0.01	0.05	0.1	1
标准差	0.0317	0.00180	0.0083	0.0066	0.0085	0.0279
均方根误差	0.0330	0.00187	0.0088	0.0069	0.0098	0.0289
决定系数	80.30%	93.58%	98.58%	99.12%	98.23%	84.82%

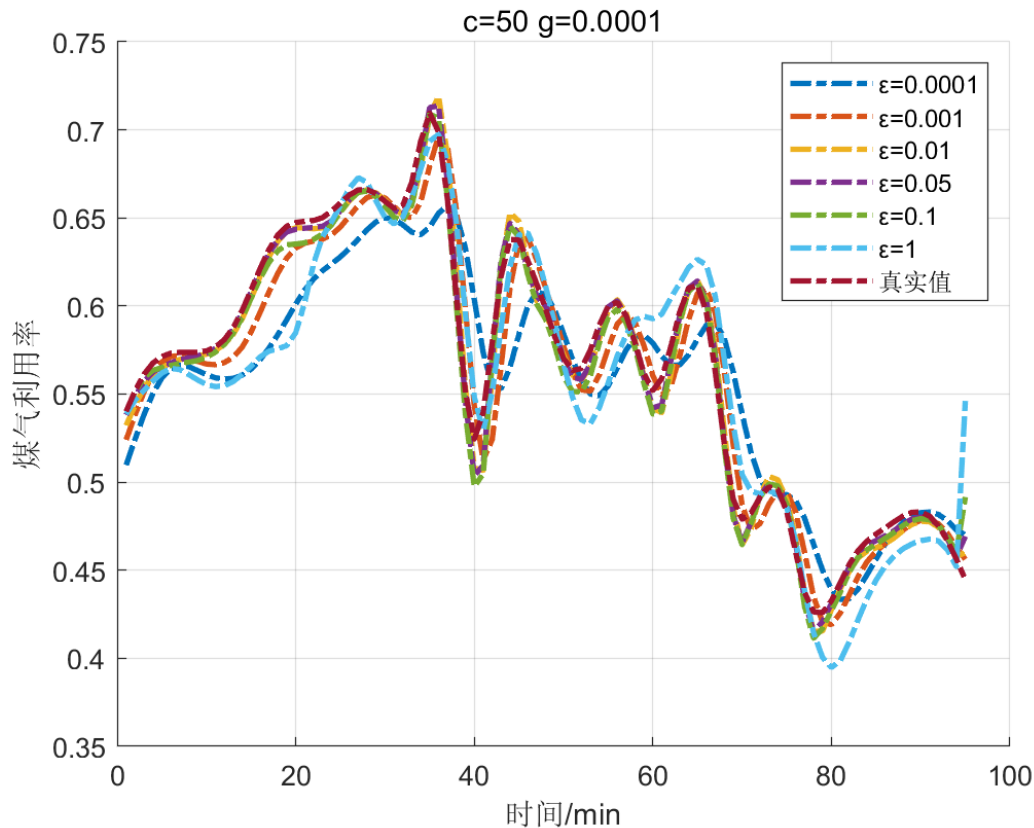


图 8.12 高炉煤气利用率预测模型 ϵ 不同时预测结果

参数 g 的预测结果及误差计算如表 8.17 与图 8.13 所示。根据预测结果和误差，当 g 值增大时，预测结果的绝对误差、均方根误差先逐渐减小然后增大，决定系数逐渐增大然后减小。综合考虑可以得到 g 取值为 0.001 是合适的。

表 8.17 煤气利用率预测模型 g 不同时的标准差、均方根误差与决定系数

g	0.0001	0.001	0.01	0.1
标准差	0.0066	0.0068	0.0089	0.0319
均方根误差	0.0069	99.10%	0.0089	0.0526
决定系数	99.12%	99.10%	98.54%	61.54%

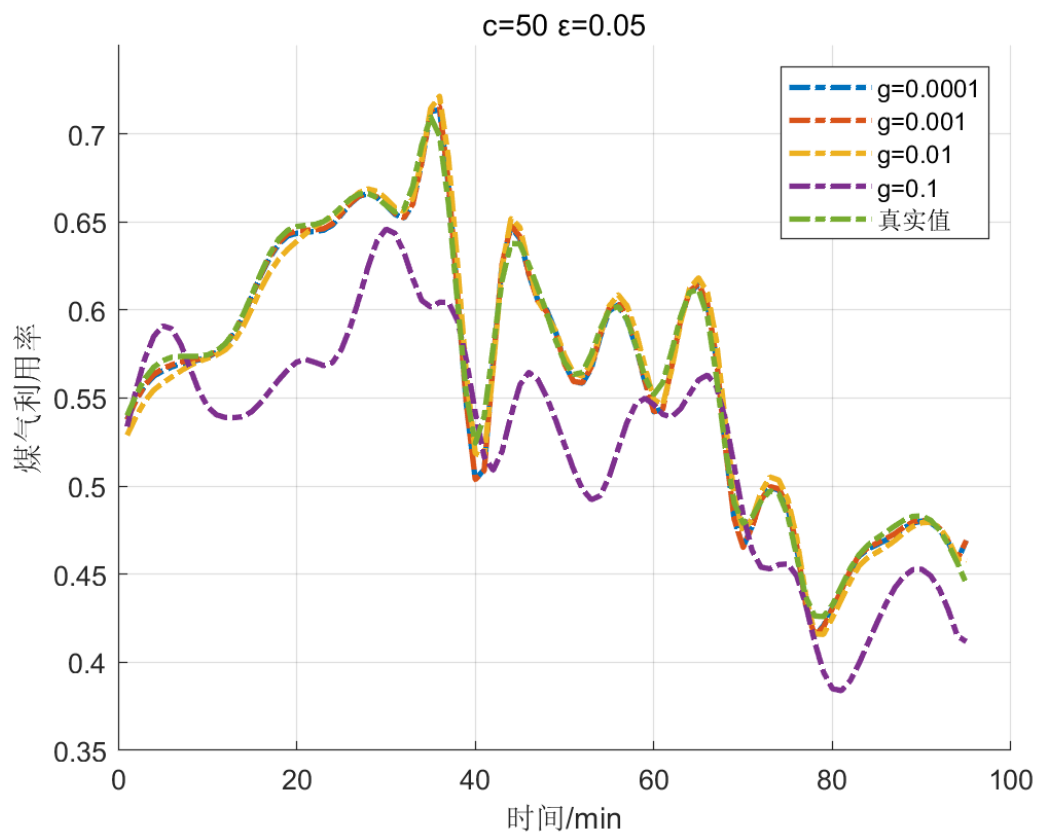


图 8.13 高炉煤气利用率预测

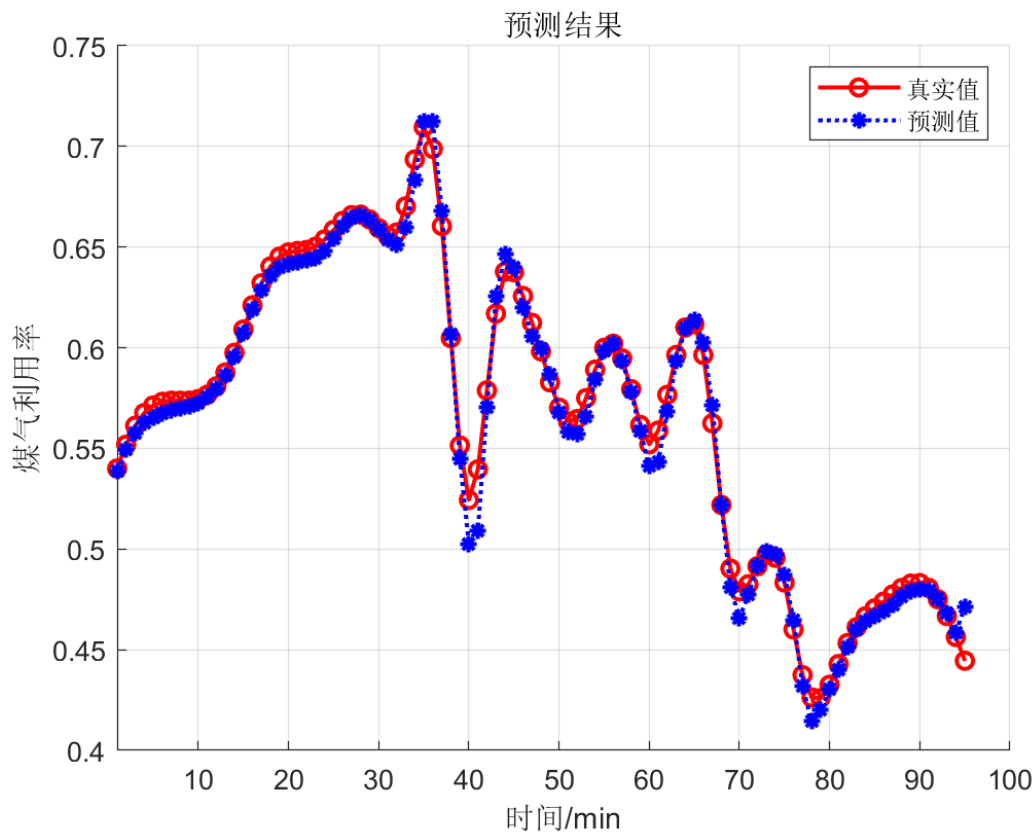


图 8.14 高炉煤气利用率支持向量回归预测图

图 8.14 表示经过支持向量回归模型预测得到的高炉煤气利用率预测结果。将得到的模型预测值与真实值带入模型评价指标中求解，得到回归模型的预测能力的评价指标值如表 8.18 所示：

表 8.18 高炉煤气利用率支持向量回归预测指标		
决定系数	标准差	均方根误差
99.12%	0.0066	0.0069

标准差与均方根误差越高，表明模型性能越差。反之，标准差与均方根误差越低，表明模型性能越好。由此结果可知，该回归模型的评价指标都具有较好的数值，故该回归模型能够较好地对未来一段时间的煤气利用率进行预测。

8.7 本章小结

智能制造与大数据技术为工业企业生产的安全、高效与稳定运行提供了坚实的

基础。本章首先介绍了高炉冶炼的基本背景以及原理，针对高炉数据存在的噪声问题进行了去噪处理。对高炉冶炼过程中存在的三个典型问题，从不同角度对问题进行描述，选择相关变量，并收集相关数据，针对提出的问题来选择对应的方法，包括 K-means 聚类算法、贝叶斯分类、最小二乘线性回归以及支持向量回归等，通过建立相应模型，最终解决相应的问题。

聚类分析：为满足高炉现场布料问题，通过利用K-means聚类算法对高炉不同布料操作参数进行距离聚类。包括聚类中簇的选择以及聚类结果的检测。

数据分类：高炉煤气利用率总是在一定范围内发生波动，利用贝叶斯分类模型，对实际例子进行分析，完成相关的高炉煤气利用率的二分类问题，并对模型的适用性以及分类性能进行了检验。

线性回归：针对高炉操作变量与状态变量之间的线性关系，利用最小二乘线性回归模型，对实际例子进行分析，完成相关参数估计，并对模型的适用性以及拟合性能进行了检验。

非线性回归：针对高炉操作变量与状态变量之间的非线性关系，利用支持向量回归这种非线性模型，对实际例子进行分析，完成相关超参数求解，并对模型的适用性以及拟合性能进行了检验。

参考文献

- [1] 杨天钧, 张建良, 刘征建, 等. 低碳炼铁 势在必行[J]. 炼铁, 2021, 40(04):1-11.
- [2] 周传典.高炉炼铁生产技术手册. 北京: 冶金工业出版社, 2002.
- [3] 吴敏, 曹卫华, 陈鑫. 复杂冶金过程智能控制, 科学出版社, 2016.
- [4] 吴敏, 王昌军, 安剑奇, 等. 基于料面温度场的高炉煤气流分布识别方法[J]. 信息与控制, 2011, 40(01): 78-82.
- [5] 吴敏, 聂卓赟, 曹卫华, 等. 面向高炉布料操作优化的在线信息检测方法及其应用(上)[J]. 冶金自动化, 2008(03): 6-9.
- [6] 吴敏, 聂卓赟, 曹卫华, 等. 面向高炉布料操作优化的在线信息检测方法及其应用(下)[J]. 冶金自动化, 2008(04): 5-8+56.
- [7] 曲飞, 吴敏, 曹卫华, 等. 基于支持向量机的高炉炉况诊断方法[J]. 钢铁, 2007

(10): 17-19.

- [8] 安剑奇, 陈易斐, 吴敏. 基于改进支持向量机的高炉一氧化碳利用率预测方法[J]. 化工学报, 2015, 66(01): 206-214.
- [9] 陈少飞, 刘小杰, 李宏扬, 等. 高炉炼铁数据缺失处理研究初探[J]. 中国冶金, 2021, 31(02): 17-23.
- [10] 李浩然, 邱彤. 基于因果分析的烧结生产状态预测模型[J]. 化工学报, 2021, 72(03): 1438-1446.
- [11] 周志华. 机器学习. 北京: 清华大学出版社, 2016.
- [12] 蒙哥马利. 线性回归分析导论 (美). 北京: 机械工业出版社, 2016, 4.
- [13] An J Q, Shen X L, Wu M, et al. A multi-time-scale fusion prediction model for the gas utilization rate in a blast furnace[J]. *Control Engineering Practice*, 2019, 92: 104120.