# 高炉打分系统模块

高炉打分系统包含数据预处理、特征选择、范围提取以及打分系统等四个模块，关联关系如下图所示：



## 数据预处理模块

|  |  |
| --- | --- |
| **项目** | **内容** |
| 描述 | 高炉稳定性评价模型的数据预处理模块包括三个步骤，首先由现场工业信息化平台实时采集传感器数据送入数据库保存，然后对数据异常进行清洗补充，完善数据集后对比不同采集项的统计规律进行时滞分析，将数据逐一对齐，形成按分钟、按批次、按日对齐数据集。 |
| 输入 | 1. 高炉现场直接检测的过程数据如：透气性指数、热风风压等。 2. 通过直接检测的数据计算得来的间接数据如：燃料比、日产量等， 3. 检化验数据如：各种化学元素含量 |
| 所需字段 | 1. 高炉现场直接检测的过程数据如：透气性指数、热风风压等。 2. 通过直接检测的数据计算得来的间接数据如：燃料比、日产量等， 3. 检化验数据如：各种化学元素含量 |
| 逻辑设计 | 生成数据预处理模块。主要有如下几个步骤：   1. 数据提取：首先，从现场工控系统利用opc协议提取过程操作数据存入分布式关系型数据库greenplum中，然后统计出日生产数据以及批次生产数据以供计算。。 2. 数据清洗：针对简单处理得到的不同类别产品的数据集合，进行缺失数据的补全，以及异常数据的剔除，从而提升数据质量，保证数据的有效性，并且在处理过程中尽可能让数据符合实际的生产规律变化情况，切忌因补全数据而出现与实际不符的情况。 3. 数据对齐：针对过程数据的多粒度（采样周期不同）特点，进行数据的时滞分析，将数据进行对齐，分为按分钟对齐的数据、按日对齐的数据以及按批次对齐的数据。 |
| 流程图 |  |
| 输出 | 对齐后的分钟数据集、日数据集、批次数据集。 |
| 用户界面 | 无 |

## 特征选择模块

|  |  |
| --- | --- |
| **项目** | **内容** |
| 描述 | 将上述整理好的数据集中所有的特征项作为输入，利用人工给予数据的标签训练随机森林模型，训练好后对所有特征项进行重要性排序，设定阈值，摘取高于阈值的特征，与人工经验相对照，最后挑取打分指标体系所需特征。 |
| 输入 | 1. 输入是所有可能的特征。接着的模块是确定每个指标的打分的范围。最后是基于这些打分。 |
| 逻辑设计 | 特征选择模块拥有以下步骤:   1. 将数据预处理模块得到的数据进一步处理。对该数据进行标准化归一化,消除因不同指标的不同量纲对数据分析的结果的影响,从而对下一步的机器学习算法做好数据准备。 2. 训练随机森林模型。使用Python的机器学习算法库Sci-kit Learn中的随机森林分类器(RandomForestClassifier)，喂入上一步中处理好的数据，训练得到所需的随机森林模型。 3. 指标选取。从训练得出的模型中获取特征重要性指数，对该指数进行排序。进一步设定一个阈值，选取高于阈值的特征作为下一步高炉打分的指标。 4. 参照人工经验调整选取的特征。调整随机森林模型的参数,并参考并对照人工给出的特征,敲定最终打分指标体系的特征。 |
| 流程图 |  |
| 输出 | 确定重要特征，作为打分指标 |
| 用户界面 | 无 |
|  |  |

## 范围提取模块

|  |  |
| --- | --- |
| **项目** | **内容** |
| 描述 | 高炉稳定性评价模型的范围提取模块，利用筛选出的特征构建二维空间进行聚类分析，按照二维特征所表征的信息分为不同的类别（一般为3类：较好、较差、正常），从每一类中挑选数据进行统计分析，针对某一参数提取出上下限。 |
| 输入 | 1. 特征选择模块的输出 |
| 逻辑设计 | 1. 确定由哪两个特征构成聚类的二维空间，二维空间的某一轴可以用合成参数。 2. 利用K-Means算法在构建好的二维空间中对标准化的数据进行聚类操作，依据特征所表达的信息分析某一类在高炉操作有哪些实际意义，并判断聚类的优劣。 3. 利用聚类之后的数据进行多目标提取，在每一类中的数据集中随机采样适当点，构成针对某一指标（如热风分压）的样本。 4. 对样本进行统计分析（如样本符合正态分布），获取这一指标的上下限及分布，这样就能得到这一参数的取值空间。 5. 基于以上的优化上下限，构建优化操作的星形图，以指导实际生产。 |
| 流程图 |  |
| 输出 | 依据指标划分出来的各参数的区间范围 |
| 用户界面 | 无 |

## 打分模块

|  |  |
| --- | --- |
| **项目** | **内容** |
| 描述 | 打分模块为每个区间范围设定相应的分值，采用相应的计算规则获得当前炉况的顺行指数，然后进行失分项分析，并判断参数的运行状态。 |
| 输入 | 1. 范围提取模块输出。 |
| 逻辑设计 | 1. 输入实时采集的数据，找到各参数对应的区间范围以及分值，将每个参数的分值进行求和计算，得出表明高炉顺行状态的最终分数。 2. 对分值低的参数进行失分项分析。 3. 利用构建的星形图，判断各参数的运行状态。 |
| 流程图 |  |
| 输出 | 表明高炉炉况的分值 |
| 用户界面 | 无 |
|  |  |