

底流干砂质量预测

底流干砂质量预测模块主要通过对深锥浓密机、底流管道、排尾管道以及水泥管道的传感器检测值进行分析，结合数据分析技术以及时序序列预测算法，实现对浓密机底流干砂质量的预测，从而实现对水泥管道的流量以及浓度的调节。

注：本文只说明其中一台浓密机和一台所属的底流泵的干砂质量预测算法，其他浓密机和其所属底流泵的预测算法

数据预处理

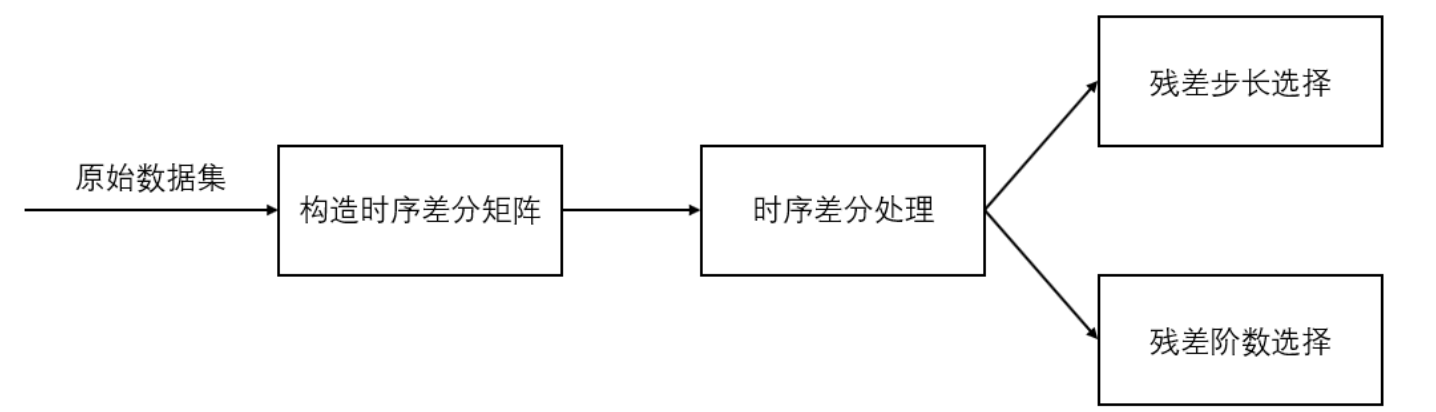
对获取到的数据集进行预处理，剔除与后续研究无关的属性，如日期、时间、机械功率等，重新整理后的数据集包含以下变量。

系统编号	检测项名称	说明
1	泥层高度	浓密机内部泥层高度
2	1电机耙架扭矩	深锥浓密机内部耙架扭矩
3	2电机耙架扭矩	深锥浓密机内部耙架扭矩
4	选厂进料浓度	充填系统进料浓度
5	絮凝剂添加量	进入浓密机的絮凝剂添加量
7	浓密机转速	浓密机内部搅拌耙架转速
11	底流浓度	浓密机底部料浆浓度
16	选厂进料流量	充填系统进料流量
17	浓密机压力	浓密机内部泥层压力
29	1号系统充填瞬时量	一般使用1号系统充填
47	1号系统充填累积量	1号充填系统充填总量
52	2号系统水泥瞬时量	1号充填系统凝结剂的消耗量
54	2号系统充填瞬时量	备用系统充填流量
57	2号系统充填浓度	备用系统充填浓度

特征工程

基于差分矩阵的特征构造

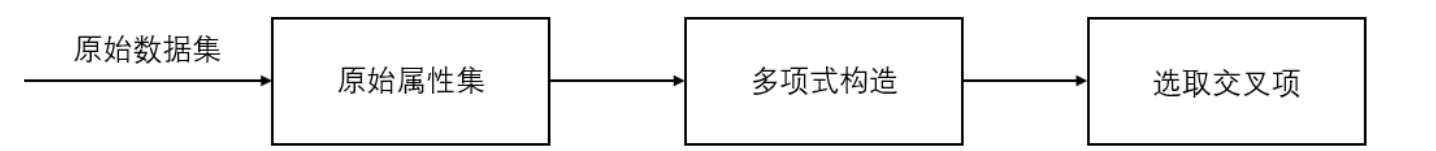
本模块采用的数据集为传感器的时序记录数据，对于时序数据来说其残差是重要的参考指标，选择使用时序差分矩阵进行特征构造。对于残差的计算方法来说有两个不同的选择，第一是残差步长的确定，第二个是残差阶数的选择。在构造时序差分矩阵时，若差分阶数过高，会导致属性值变得稀疏，影响后续多项式特征的构造；若选择的矩阵过大，则会显著增加算法训练时的计算量，延长算法的训练时间。所以最终选定差分矩阵中步长为2，阶数为1，对在数据预处理阶段选定的关键属性进行差分特征构造，将结果作为特征集的一部分。



多项式特征构造

多项式构造是通过原始属性的多项式组合产生新的特征，获得属性之间更高维度的相互关系，最终获得数据的非线性特征，提升机器学习模型的表达能力。

假设有属性 X ，属性 Y ，设定多项式次数为2，我们可以获得其多项式特征组合 $(1, X, Y, XX, XY, Y*Y)$ ，一般只需要特征间的交互项，对于该种情形即为 $(1, X, Y, X*Y)$ ，由此我们就从两个属性获取到了其非线性特征。当多项式构造所使用的属性和次数增大时，构造出的特征会以指数形式快速增加，引起特征空间的爆炸。因此在实际使用时，多项式次数不超过3，原始属性不超过20个较为实用。

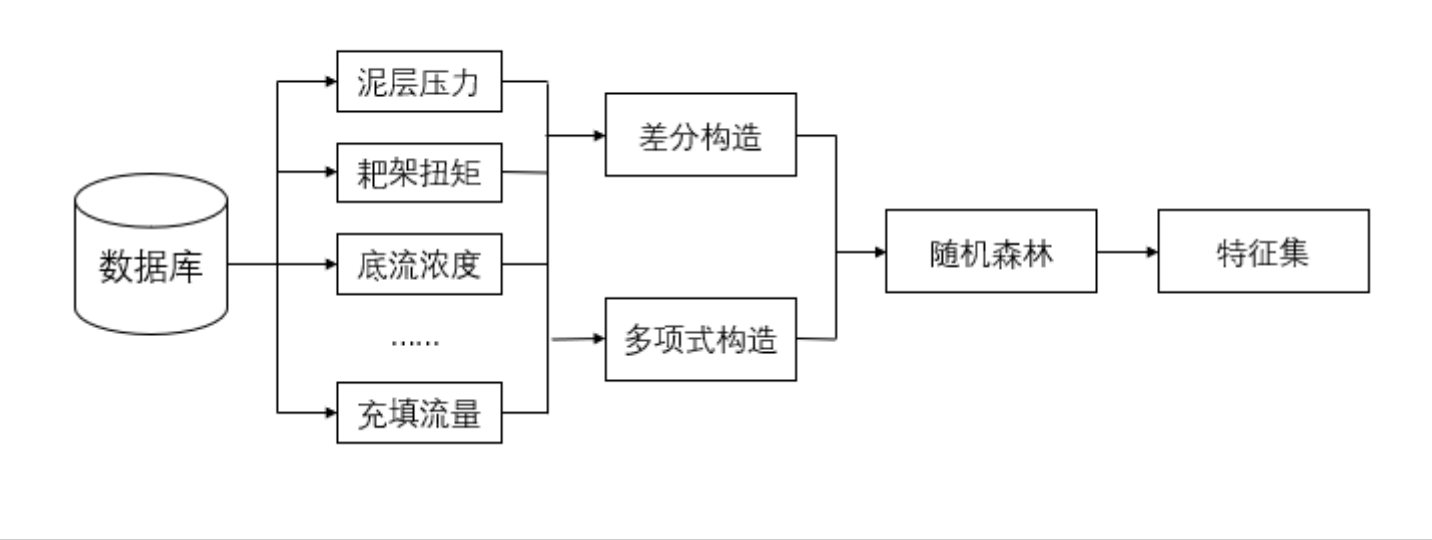


基于机器学习模型的特征选择

直接将特征投入算法训练，会导致严重的特征冗余，大大提升模型的计算量，并且可能降低模型的预测结果。因此使用特征选择方法对特征进行选择，其做法为选用机器学习模型对特征和目标变量进行建模，根据模型训练过程中依照各个特征的重要性进行特征权重系数计算，最终将输入模型的所有特征进行排序，根据具体需要进行特征的数量的选择。

本模块选用的机器学习模型为随机森林模型，随机森林在以及决策树为基础分类器的同时引入了随机属性选择的做法，该种做法使得随机森林模型可解释性较强，对于非线性关系建模较好，并且能够很好地体现各个特征的重要性差别。训练迭代次数为1000次随机森林算法，算法训练完毕并得到特征重要性后，计算所有属性的平均特征值重要性，然后选择特征重要性不小于均值的0.2倍的特征作为最终模型所用的特征。

特征工程流程图

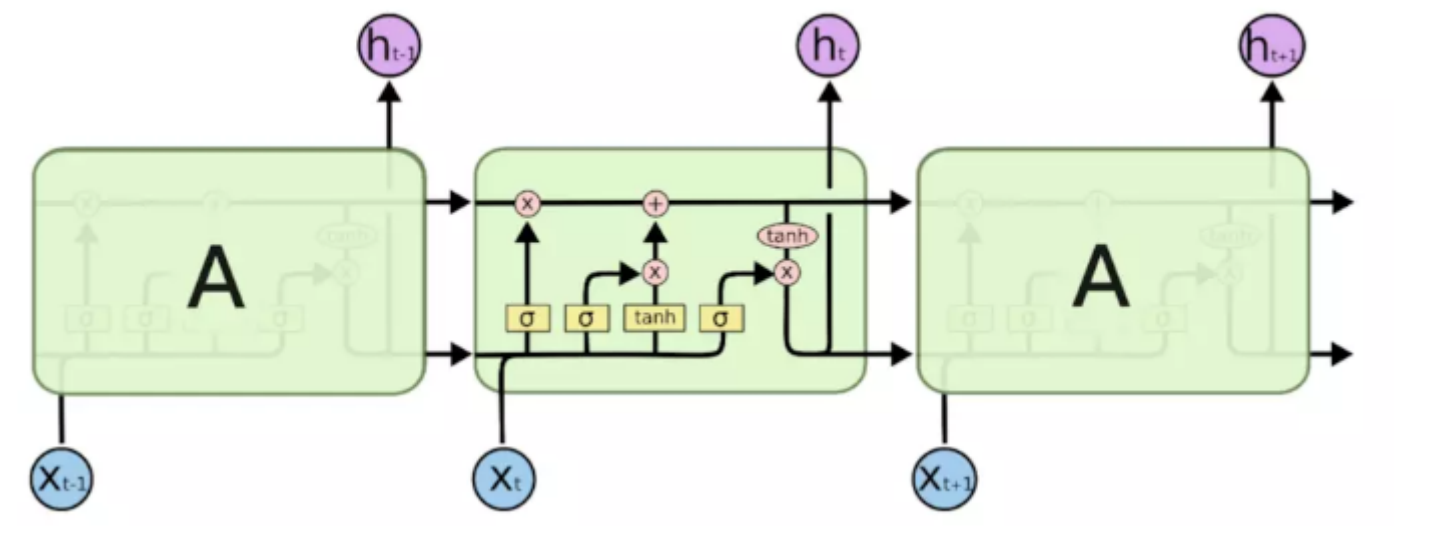


预测算法

模型选择

时间序列预测分析是利用过去一段时间内某事件时间的特征来预测未来一段时间内该事件的特征。时间序列模型是依赖于事件发生的先后顺序的，同样大小的值改变顺序后输入模型产生的结果是不同的。时间序列模型最常用最强大的工具就是递归神经网络（recurrent neural network, RNN），由于RNN模型如果实现长期记忆的话需要将当前的隐含态的计算与前n次的计算挂钩，导致计算量会呈指数式增长，模型训练的时间大幅增加，因此RNN模型一般直接用来进行长期记忆计算。LSTM（Long Short-Term Memory）模型是一种RNN的变型，优势在于解决RNN的梯度消失和梯度爆炸的问题，目前广泛应用于序列数据处理和预测。

LSTM的网络结构如下：



网络训练

选择10000余条浓密机历史运行数据进行网络训练，训练的设置如下：

- 按照训练集：测试集 = 3：1 的比例将数据进行划分
- 损失函数选择MAE和MSE损失

干砂质量预测

将测试集数据送入预测模型，得到底流浓度的预测结果，浓度与密度的映射关系如下：

密度(g/cm^3)	1.10	1.15	1.20	1.25	1.30	1.35	1.40	1.45
浓度(%)	14.1	20.3	25.9	31.1	35.9	40.3	44.4	48.3

密度(g/cm^3)	1.50	1.55	1.60	1.65	1.70	1.75	1.80	1.85	1.90
浓度(%)	51.9	55.2	58.3	61.3	64.1	66.7	69.1	71.5	73.7

定义以下几个变量，底流流量FT，底流浓度DT，底流密度CT，单位时间内底流干砂质量M,则有：

$$M = FT * DT * CT$$

M即为单位时间内底流干砂质量的预测结果。