随机森林

随机森林是决策树模型的集成算法，它的强实用性以及高预测精度已在大量实证分析中得到论证。它将n棵决策树的预测结果结合起来以做出比单棵决策树更为精确的预测，（这里要说一下什么是随机森林，可以参考一下其他的论文是如何进行论述的），随机森林在构造单棵决策树时除了在选择样本是使用随机抽样，在构建单棵决策树的输入特征时也采用随机抽样的方式从原有特征空间中选择子特征空间，此做法有效提高了模型的稳定性。随机森林回归模型的预测结果是由众多弱评估器的预测结果进行平均而得，即众多决策树预测结果的平均值作为整体模型的回归值。（这里可能还是要扩展一下，可以借助原始的paper来找公式）

随机森林是决策树模型的集成算法，它的实用性已经预测准确性已在大量实证分析中得到论证。接下来是来阐述一下什么是决策树。

随机森林回归模型的构建

随机森林回归模型的构建方法是：首先，从N个交易日时间序列样本集中，按7:3比例有序导出70%的样本量作为模型的训练集D，30%样本量作为模型的测试集；其次，使用bootstrap重抽样方法从训练集D中抽取同样数量的子样本集，重复此步骤n次生成子样本集Di（i=1，2，3，。。。n）；接着，以随机抽样的方式对上述生成的各个子样本集进行特征抽样，抽样的特征数量以随机方式获取并保持在70%以上，各个样本集根据特征抽样的结果进行训练子集数据的重新生成；再次，根据抽样的各子样本集构建决策树，并对n棵决策树的结果取算术平均值作为模型的模型的预测结果；最后，对上述模型进行调优，选出预测效果较高精度的回归模型。

对被预测变量和筛选出的预测变量进行标准化处理，使用上述相同的方式构建随机森林回归模型，形成对照模型，对比模型回归结果选取被预测变量的预测模型。归一化处理的方式为：

其中，为第i个特征变量的第j个样本值，为第i个特征变量的样本均值，为第i个特征变量的样本标准差，为第i个特征变量的第j个经过标准化后的样本值；为被预测变量的第j个样本值，为被预测变量的样本均值，为被预测变量的样本标准差，为被预测变量的第j个经过标准化后的样本值。

模型评估指标

本文采用的模型评估指标为平均绝对百分比误差（MAPE）和R2。设n为样本数，和分别表示为被预测变量的样本值和预测值，其计算方式为：

其中，为被预测变量的样本均值。

回归模型回测检验方法

依据上述对照模型预测结果，选定甲醇期货价格预测模型。根据甲醇期货价格回归模型预测结果构建回测交易规则，根据回测交易结果检验预测模型的有效性。本文选用的回归交易策略为经过修改的Aberration策略，并使用夏普比率来确定确定交易策略的相关参数。

Aberration交易策略分为两个部分，该策略的第一部分是相关参数的计算：

其中)为上轨线在t时刻的向上突破值，)为下轨线在t时刻的向下突破值，)为甲醇期货收盘价在t时刻的移动时长为N的移动平均值，)为甲醇期货收盘价在t时刻移动时长为N的移动标准差，和为阈值。

第二部分是交易规则，Aberration交易策略是建立在布林格通道Bollinger Bands理论的基础之上，Bollinger Bands理论是由John Bollinger在20世纪60年代提出，该理论指出金融市场产品价格将主要在由移动平均线以及在移动平均值基础上增加和减少一个单位的移动标准差形成的上轨和下轨线之间进行波动，当产品价格触及上轨或者下轨线时，价格将有很大可能性重新回到移动平均线附近。Keith Fitschen根据此理论进行了大量实验后发现布林格通道的上轨和下轨线其实是价格突破信号，即价格靠近上轨或者下轨线附近的时候将有很大可能性将会超过上轨值或者跌破下轨值。因此本位采用的交易策略为，当价格超过上轨线进行买多操作，并在价格回到移动平均线下方进行平仓操作；在价格跌破下轨线时进行买空操作，并在价格回到移动平均线上方时进行平仓操作。

夏普比率是能够同时对收益和风险进行考虑的经典指标之一，能够很好地评估回测交易效果。本文将根据使用夏普比率对Aberration交易策略的参数、、进行调整，根据调整后的交易策略结果评估基于基本面特征的甲醇期货价格预测模型的有效性。

Dual Thrust交易策略

第二部分为策略构建和交易逻辑。

依据上述对照模型预测结果，选取甲醇期货价格预测模型。根据甲醇期货价格回归模型预测结果构建回测交易规则，根据回测交易结果检验选取模型的实际价值。本文构建的回测交易规则是使用KDJ技术指标改进后的Dual Thrust交易策略。

Dual Thrust交易策略分为两个部分，第一部分为相关参数的计算：

其中，High为前N个交易日中的最高交易价，LC为前N个交易日收盘价中的最低价，HC为前N个交易日收盘价中的最高价，LL为前N个交易日最低价中的最低价，Open为当日开盘价，和为设定的阈值，预设值均设为0.7。

第二部分为：策略构建和交易逻辑。当大于时，空头相对容易被触发，当小于时，多头相对容易被触发。本文将KDJ技术指标引入到Dual-Thrust交易策略中，当K、D、J均大于50时，市场处于利多市场，当K、D、J均小于50时，市场为利空市场，根据市场动态地调整和以寻求更好的交易策略效果。当价格向上突破上轨时，如果当时持有空仓，则先平仓，再开多仓；如果没有持仓，则直接开多仓。当价格向下突破下轨时，如果当时持有多仓，则先平仓，再开空仓；如果没有仓位，则直接开空仓。若价格在下轨和上轨之间则保持原有仓位。

（这里的kdj的引入可能还是要思考一下如何处理为好，因为目前仅仅包含了两种情况，还是不太好，还是需要进行处理一下）

这里要思考一下策略的事情是否要借助KDJ 如果是借助了KDJ的话，那么如何将策略设置合理一些 这些是目前要考虑的主要事项，这部分的工作留到今天晚上 这里可以借助一下那片论文中的KDJ的策略，看能不能结合一下

将Dual-Thrust交易策略中的K1和K2初始值设为0.7，加入KDJ技术指标可以对交易市场进行有效跟踪，动态调整K1和K2的值，寻求投资获利。KDJ指标可以有效地对市场趋势进行跟踪，当K、D、J均大于50时，交易市场为利多市场，可将K1设置为0.6；当K、D、J均小于50时，市场为利空市场，将K2设置为0.6.

交易回测规则

主要是要设定阈值，大于某个阈值之后就做多，小于某个阈值之后就做空

那么这个时候该怎么处理呢