**沪深300多因子选股模型之季度轮动策略——基于LSTM预测模型**

模型：LSTM模型

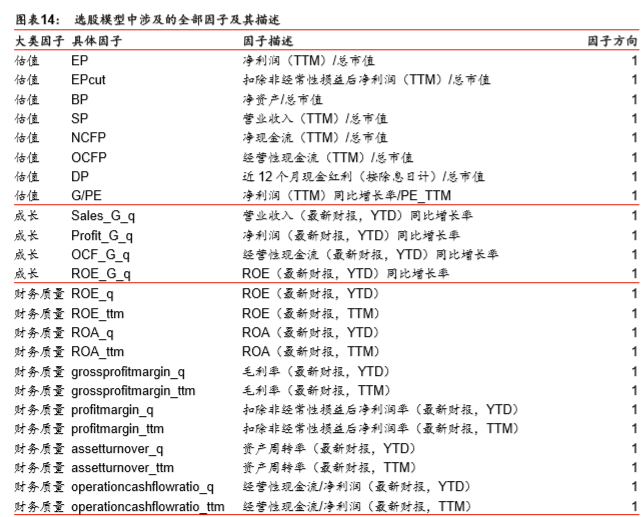
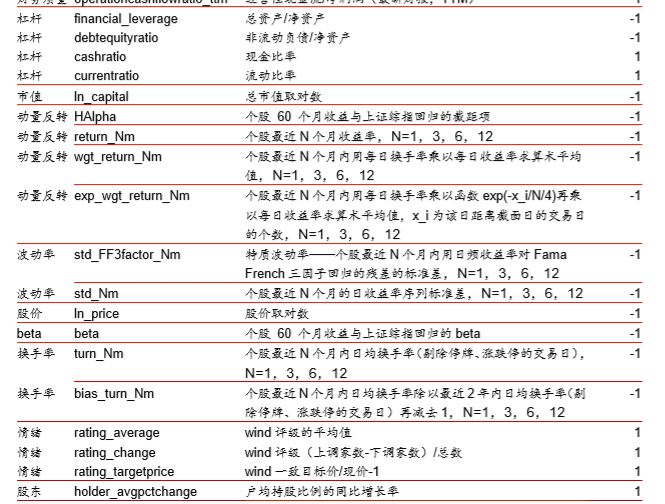
参考网站：https://www.sohu.com/a/143888793\_769402

**整个步骤：**

选择策略模型-确定因子-数据预处理-确定模型参数-建立模型得到预测结果-调整最优参数-依据模型结果对沪深300只股票进行预测-挑选预测期间收益率最好的10只股票-进行策略回测-回测结果评价。

**具体执行:**

1. 参数设定
   1. 回测标的：沪深300股票+每个行业随机挑选出20%的股票作为补充
   2. 回测时间：2009年1月1日至2016年1月1日
   3. 策略验证时间：2016年1月1日至2018年6月1日
   4. 因子库：企业基本面因子、市场表现因子（时间周期先按照日周期设置，针对具有季度属性特征的因子，则直接采取季度数据，如每股收益等基本面数据）

* 1. RNN 时间长度( steps)：36 个月，即每一个训练样本包含过去 36 个月的因子数据，依次从第一个月输入神经网络，并将返回值与下一个月因子同时循环输入神经网络，以此类推，直到得到第 36 个月的预测值。
  2. 因子降维（因子选择）：采取互信息方法进行因子降维。来源《华泰证券-华泰证券人工智能系列之十二：人工智能选股之特征选择》
  3. 因子预处理：来源：《华泰证券-华泰证券人工智能系列之十二：人工智能选股之特征选择》

1. 中位数去极值：设第 T 期某因子在所有个股上的暴露度序列为𝐷𝑖，𝐷𝑀为该序列中位 数，𝐷𝑀1为序列|𝐷𝑖 −𝐷𝑀|的中位数，则将序列𝐷𝑖中所有大于𝐷𝑀 +5𝐷𝑀1的数重设为 𝐷𝑀 +5𝐷𝑀1，将序列𝐷𝑖中所有小于𝐷𝑀 −5𝐷𝑀1的数重设为𝐷𝑀 −5𝐷𝑀1
2. 缺失值处理：得到新的因子暴露度序列后，将因子暴露度缺失的地方设为中信一级行业相同个股的平均值
3. 标准化：将中性化处理后的因子暴露度序列减去其现在的均值、除以其标准差，得到 一个新的近似服从 N(0, 1)分布的序列
4. 中性化处理：市值与行业中性化处理（其他因子暂时不考虑中性化处理）

<https://blog.csdn.net/xuxiatian/article/details/79731493>

<https://www.ricequant.com/community/topic/4309/>

* 1. batch size：1000，该参数属于 RNN 神经网络的系统参数，是BP算法中用来计算梯度的参数，即每次的训练，随机抽取训练样本中的 1000 个样本作为训练样本。
  2. 隐层神经元个数：400，2 层，该参数同样属于 RNN 神经网络的系统参数，是 输入样本与隐层细胞连接的“神经”个数，受限于电脑的性能，只能设定为三位数，隐层个数也仅为 2 层。
  3. 学习速率：0001，RNN 神经网络的系统参数，是模型训练时梯度下降的速度，过高容易导致梯度消失，过低则训练过慢。
  4. 交叉检验比例：30%，为了防止模型过拟合，选择样本中的70%作为训练集，用以训练模型参数，而剩余30%不参与训练，只作为测试集进行检验，如果训练集准确率与测试集准确率同时上升，则说明模型过拟合的可能较小。

1. 模型建立：

利用2009年1月1日-2012年1月1日的数据进行模型建立，然后利用模型结果进行近一季度的price预测，将price与2012年1月1日-2012年3月1日的价格进行精度对比。如果精度太差就调整参数，调整参数直至精度最优，记录此时得到的最优模型结果参数。

1. 策略步骤：
   1. 首先，利用2009年1月1日-2012年1月1日的数据进行模型建立，然后利用模型预测出沪深300股票池中从2012年1月1日-2012年3月1日的price数据，再利用聚类模型（如K均值聚类）对所有预测的价格数据进行聚类分析，将价格数据分为几个类，每个类都代表一种业绩层次，从而挑选表现最好的A类股票池（此处将业绩表现最好的一类股票假设命名为A类股票池），进行买入操作。
   2. 到2012年3月1日，得到2012年1月1日-2012年3月1日的真实价格数据，再利用2009年1月1日-2012年3月1日的真实价格数据进行model模型结果参数更新，并进行2012年3月1日-2012年6月1日的沪深300价格预测，利用聚类模型（K均值聚类）对所有预测数据进行聚类分析，从而挑选表现最好的A类股票池，进行调仓，持有一个季度。
   3. 到2012年6月1日，重复上述操作
   4. 到2012年9月1日，重复上述操作
   5. ……
   6. Loop操作，这样子不断的季度循环，循环到2016年1月1日
2. 买入卖出操作：
   1. 买入：

4.1.1 对于第一季度的开仓操作，在model得到A类股票池之后，依据马科维兹投资组合最优化理论寻找最优投资比例，进行0.9仓位系数的开仓操作。

4.1.2 针对每个季度调仓，针对更新的A类股票池，仍然依据马科维兹投资组合最优化理论进行最优投资比例调整。

4.1.3 根据仓位设置规则进行仓位系数调整。

4.1.4 开盘价买入。

4.1.5 不同风格买入要求见5大点。

* 1. 卖出：

4.2.1 每个季度之后，依据更新的股票结果，若原先的股票没有更新的股票池中，则全部清仓，如果没有持有，则开始建仓。以上，仍然依据马科维兹投资组合最优化理论寻找最优投资比例。依据仓位设置规则乘以仓位系数进行仓位调整进行投资。

4.2.2 收盘价卖出

* 1. 仓位设置：

将预测的A类股票池按照每一个季度的最优比例构造成一个虚拟的投资组合，然后根据该投资组合在该季度前的业绩表现应用CAPM模型测算其阿尔法。若其α大于10%，则给予仓位系数0.9；若α大于5%小于8%，给予仓位系数0.8；若大于0小于5%，给予仓位系数0.7；若大于-5%小于0，给予仓位系数0.6，若小于-5%,给予仓位系数0.5。

* 1. 套保设置：

依据10只股票的投资组合贝塔值，卖出对应仓位的沪深300指数期货，进行套保。每一季度将所有期货头寸全部清仓。

* 1. 剩余资金配置：

每个季度的仓位剩余资金用于投资3个月国债。

* 1. 止损设置

若单只股票买入亏损超过10%，则进行该股票清仓，同时利用马科维兹最优化理论进行仓位分配更新，重新在当时进行仓位更新操作，同样的期货也进行相应的套保更新。

* 1. 系统暴跌应急机制

若当日整个资金仓位亏损超过3%,全部清仓，并空仓1个月，一个月后重新启动交易机制，但是策略轮转周期仍然以季度为周期。即固定3、6、9、12月进行股票池更新操作。

1. 风格轮转因子

针对风格轮动，通过制定了一个比较简单的指标：成交量指标，来度量整个市场的状态，以判断市场牛熊属性。

将每个月的成交量作为一个因子，若上一个月的沪深300月成交量大于过去一年沪深300的月平均成交量的1.2倍范围区间，则判断为牛市，采取季初直接全部买入策略，即买入设置为策略触发时一次性买入，股票仓位分配比例仍然服从马科维兹最优化理论，资金仓位服从4.3要求。

将每个月的成交量作为一个因子，若上一个月的沪深300月成交量小于过去一年沪深300的月平均成交量0.8倍范围，则判断为熊市，采取定投策略，即买入设置为每日按照一定比例买入，股票仓位分配比例仍然服从马科维兹最优化理论，资金仓位服从4.3要求，设置套保。

将每个月的成交量作为一个因子，若上一个月的沪深300月成交量在过去一年沪深300的月平均成交量0.8-1.2倍范围之间，则判断为震荡市场，采取定投策略，即买入设置为每日按照一定比例买入，股票仓位分配比例仍然服从马科维兹最优化理论，设置套保，资金仓位服从4.3要求。

1. 最终模型参数调整：

在2016年1月1日，进行资金曲线拟画，算出整个投资策略的阿尔法。同时，设立对照组，进行参数修改，比如针对时间参数，将一个季度换成半年；针对投资比例参数，设置为等额投资；针对仓位调整参数，设置为满仓投资。在所有参数组中，挑出整体阿尔法最好的一个最为最终策略。

1. 结果回测：

模拟本金：10000000RMB，将上述最终策略应用到2016年1月1日至2018年6月1日的时间周期中，从而作为最终的回测结果进行提交。

1. 最后结果的评价指标：

最大回撤率、年化收益率、超额收益率、换手率、夏普比例

1. 交易成本：

五项：股票：卖出收取的印花税：每交易1000块钱收取1块钱

佣金：成交金额的万分之五（即每交易1万块钱则收费5元钱）

过户费：成交金额的万分之0.2（沪市收，深市不收）

期货：交易手续费：万分之0.23

交割手续费：万分之一

国债：暂不考虑

其他：用于应付比赛的前期任务：

1. 因子选择上，选择7个因子即可
2. 建立模型
3. 购买策略上，等额购买，持有期一个季度。
4. 不考虑套保
5. 不考虑投资国债
6. 可能难点在于：规则的繁琐，在建立循环框架上会存在困难