

1. 特征工程

1.1 特征构建

构建了众多常用的技术指标，包括：

- Average True Range (ATR)
- Ease of Movement (EMV)
- Force Index (FI)
- Money Flow Index (MFI)
- Standard Deviation (std)
- Relative Strength Index (RSI)
- Stochastic Oscillator (KDJ)
- Ultimate Oscillator (UOS)
- Average Directional Index (ADX)

以日内波幅作为 target，计算各个特征与它的相关系数：

Table 1 部分技术指标与日内波幅的相关系数

ATR	EMV	EMV_ma	FI	MFI	Std	RSI
0.5419	-0.38	-0.301	-0.1682	-0.1467	0.43474	-0.1996
KDJ_k	KDJ_d	KDJ_j	UOS	ADX_+di	ADX_-di	ADX
-0.1272	-0.10728	-0.07414	0.178836	0.02433	0.11329	-0.16211

下图为各个特征之间的相关系数热图矩阵：

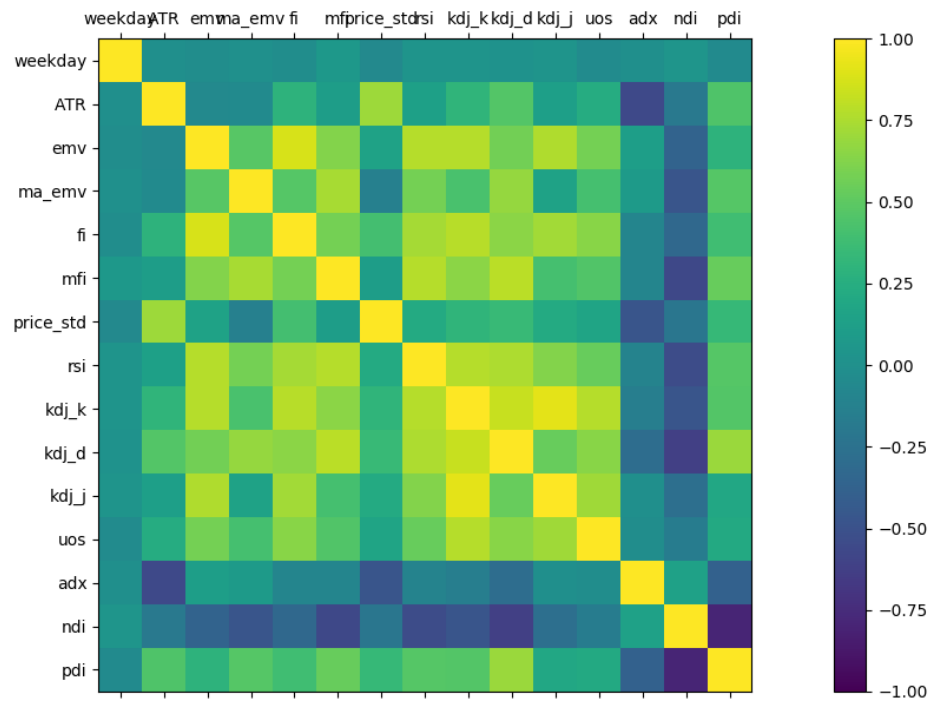


Figure 1 特征的相关系数矩阵热图

另外，本文尝试性地构建了一些新特征。

1) True Range Day (TRd)

据观察，日内上午和下午通常具有不同的走势，因此分别计算日内上午(TRm)和下午的波幅(TRa)。与波幅的相关系数：TRm: 0.765, TRa: 0.774。

2) Volume Weighted ATR (VWATR)

计算了每日交易量与波幅的相关系数，为 0.5763。交易量与波幅相关性较高，因此考虑结合交易量与价格构造新的特征。尝试构造交易量加权的 ATR 指标，设N为窗口大小，则第t日的指标值为：

$$VWATR_t = \frac{\sum_{k=1}^N ATR_{t-k} \times Vol_{t-k}}{\sum_{k=1}^N Vol_{t-k}}$$

即窗口内的交易量加权的 ATR 均值。

与波幅的相关系数为 0.6598。

3) Volume Std (Volstd)

简单的窗口内交易量标准差，仿照窗口价格标准差计算得来。与波幅的相关系数为 0.2719。

4) Cumulative Absolute Change (CAC)

考虑到有时两个波幅相同的交易日，其日内 tick 级数据的走势特征可能迥然不同。为了综合考虑这种日内的波动形式对波幅的影响，尝试构建了日内累-计波动值特征。

累积价格波动值：

$$CAC_{price} = \sum_t abs(Price_t - Price_{t-1})$$

累积交易量波动值：

$$CAC_{vol} = \sum_t abs(Vol_t - Vol_{t-1})$$

其中 t 为日内的第 t 分钟交易的 tick。

与波幅的相关系数：CAC_Price: 0.5625, CAC_Vol: 0.4649。

1.2 特征选择

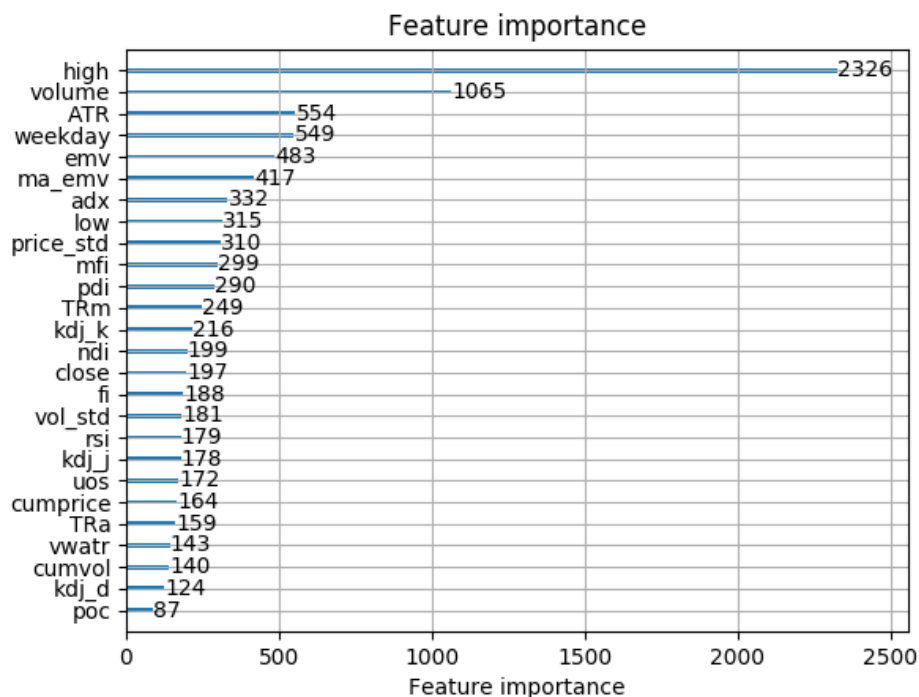


Figure 2 由 LightGBM 模型生成的特征重要度排名

根据 lightgbm 的 plot_importance 结果，选择了特征 high, volume, weekday, ATR & EMV。

2. 训练与预测

选择了 sid 为 000001 的股票从 2016 年 1 月 1 日至 6 月 1 日的 tick 和 bar 数据，使用 LSTM 进行建模训练并预测。

2.1 timesteps 的设置

Timesteps 作为 LSTM 区别于全连接 DNN 特有的参数，对模型的泛化能力具有重要的影响。尝试了多种不同的 timesteps 在 train 集上的拟合结果与 test 集上的预测表现，结果如下：

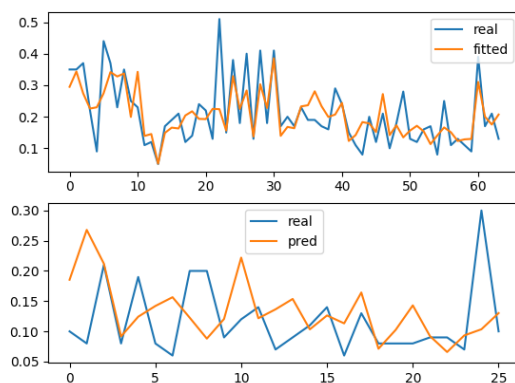


Figure 3 Timesteps=1 时的拟合结果

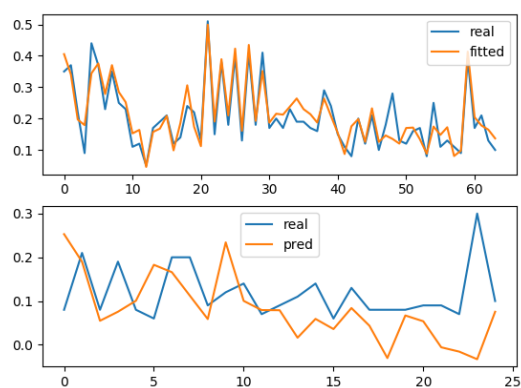


Figure 4 Timesteps=2 时的拟合结果

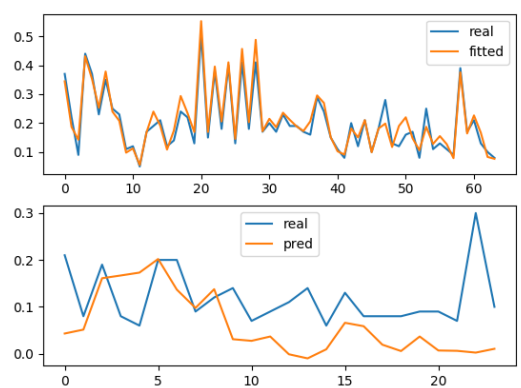


Figure 5 Timesteps=3 时的拟合结果

当 timesteps 取的越大，模型在训练集上的拟合越紧密，而在测试集上表现则越差，甚至会出现负值的预测结果。因此选择 timesteps=1 在这个问题下模型表现最好。

2.2 加入 target 的滞后值作为特征

检验如果加入 target 作为训练特征是否能提高泛化能力，即使用昨日的 target 作为今日的特征。

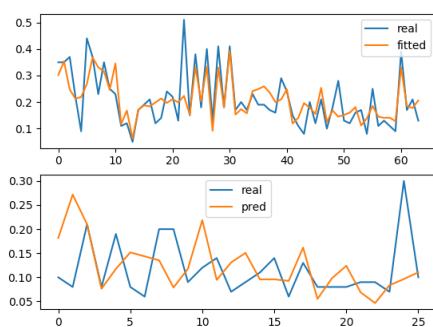


Figure 6 加入 target 滞后值

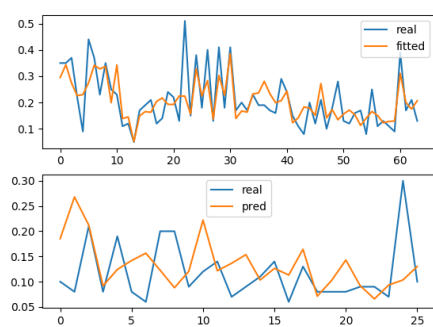


Figure 7 不加入 target 滞后值

二者相差不大，秉持奥卡姆剃刀原则，不加入 target 滞后值作为特征。

3. 如何利用

3.1 策略背景

对日内波动幅值进行预测本身并不能直接指导交易，因为这只是一个范围。在确定范围后，日内的走势可能在范围内波动，可能突破上幅值，也可能突破下幅值，这就需要日内的交易模型来面对各种信号做针对性的决策。之所以没有直接在日内 tick 级序列上建模并预测，是因为笔者经过多次实验后，发现日内的走势随机性极强，这就导致对日内序列建模所依赖的 feature 的发挥可能非常不稳定。另外日内 tick 经常会出现多达数分钟的价格静止的情况，这对 ARIMA, LSTM 等基于时间序列的算法十分不利。对日内价格或者收益率等一阶值直接建模，很多时候的确能抓住小的波幅，但如果在日间大的下跌趋势中多头而没有相应的判断机制，会在后续未能及时平仓而迅速输掉浮盈以至于亏损本金。基于上述思考，笔者认为如果引入日间的 bar 级数据以辅助日内的序列建模或者交易逻辑构建，会获得更好的效果。

事实上，开盘区间突破策略是一种很常见的量化交易策略，一般是以今日开盘价加减一定比例的昨日振幅以确定上下轨。日内突破上轨时平空做多，突破下轨时则平多做空，包括 Dual Thrust, R-Breaker 等策略。这些传统策略对于区间的上下轨一般由一些简单的 bar 数据经过加减乘除获得，例如 Dual Thrust 策略的原理为：

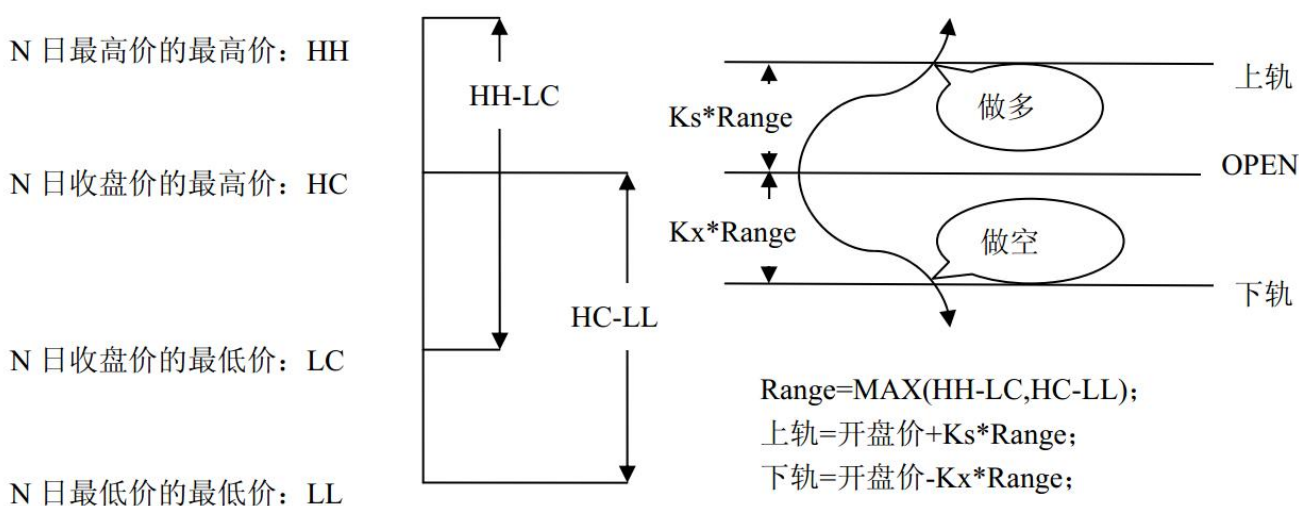


Figure 8 Dual Thrust 策略逻辑示意图^[1]

在日内波幅范围的确定上，该策略即采用前 N 日的 bar 级数据的简单加减确定当日的 range 上轨与下轨，并简单地在突破上轨时做多，击穿下轨时做空。图 9 是本文使用 Dual Thrust 所做的回测实验，发现该策略确实能一定程度跑赢 Buy & Hold 策略。

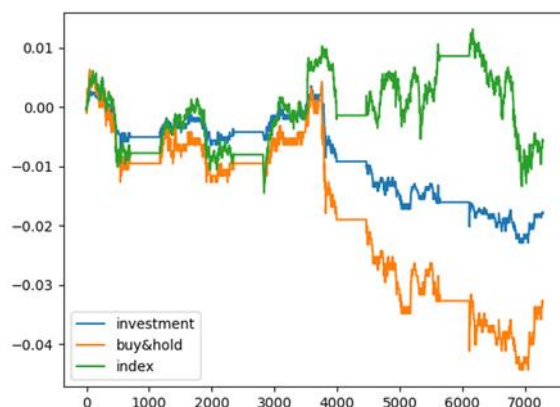


Figure 9 基于 Dual Thrust 的简单日内高抛低吸策略

3.2 简单交易思路

在确定日内波动的上下轨道之后，日内价突破上下轨的基本情况大致可分为四种：

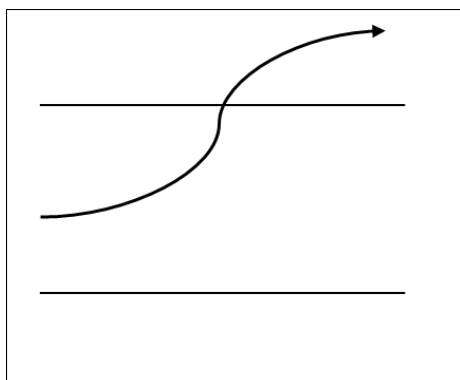


Figure 10 形态 1：向上突破

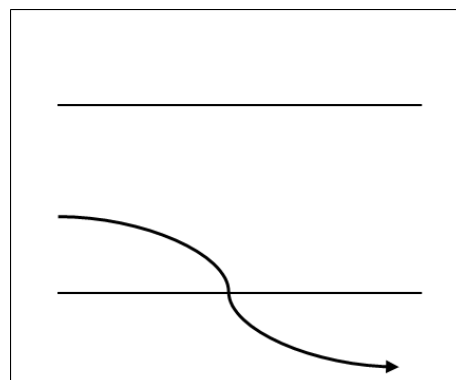


Figure 11 形态 2：向下突破

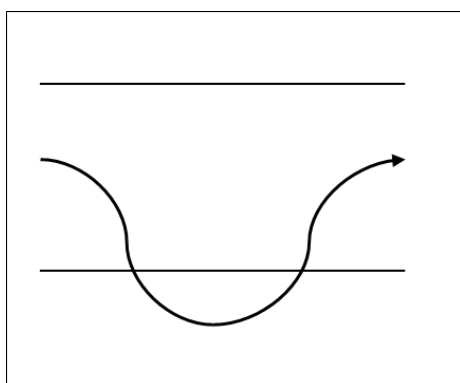


Figure 12 形态 3：触底反转

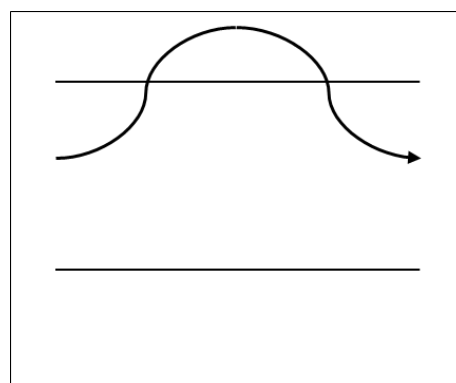


Figure 13 形态 4：顶部反转

在根据上下 range 这一指标构建交易逻辑时，不得不考虑在价格突破 range 时的趋势判断问题。例如同样是突破上轨，形态 1 和形态 4 是十分不同的。很简单的逻辑，如果在突破上轨时多头，那么在形态 1 下会盈利而在形态 4 下会亏损，这时一个好的判断方法是度量在突破点的趋势动量，只有该动量指标大于阈值才会触发形态 1 的信号而多头；而如果经判断是形态 4，则最好的选择则是空头。

关于动量指标的度量和交易逻辑的构建，本文未作深入探讨，还有待后续的研究。

参考文献

[1] 量化投资 Quant. 详解程序化交易 Dual Thrust 策略[EB/OL].

<https://xueqiu.com/5256769224/32429363?from=timeline&isappinstalled=0>.2014-10-24.