

LSTM 深度学习模型进行日内股指预测报告 V3.0

(对应代码: LSTM_V5.0)

一、V2.0 的策略测试结果

根据模型 LSTM_V4.0 进行策略编写后, 没有起到好的预测效果。主要不足有:

- (1) 预测结果有太多的大涨大跌趋势, 导致频繁开仓。
- (2) 预测趋势准确率较低。从某天交易图示来看, 11 笔交易中, 趋势预测准确的只有 5 笔, 未过半。

二、修改一

1. 修改内容:

针对策略实现显示出来的种种不足, 需要重新对模型进行修改和优化。

(1) 增大数据量。之前的预测模型的训练集只取了 10000 个序列, 这对于深度网络需要的数据量远远不足。考虑将 IC、IH 合约的 bar 数据也进行处理, 应用到模型中来。因为 tick 数据缺少最高价最低价等指标, 后续指标难以添加, 先不予考虑。目前的数据集增大到 90000 个序列。先用它们来训练, 效果不理想时继续增大数据量。

(2) 更改时间序列的长度。原来的时间序列为 5 分钟, 相关性还不是很很高。考虑到特征中有 26 分钟移动平均线这一特征, 使得开盘后 26 分钟只收集数据, 不做预测。同时因为第一分钟接收到的数据有误, 予以删除, 因此将时间序列调整为 25 分钟, 既保证不会有数据浪费, 又能使预测时参考数据大大增加, 提高预测准确率。

(3) 将数据集打乱重排再进行训练, 能在一定程度上提高准确率。

2. 修改后的结果:

增大数据量, 拉长序列长度后, 训练时间有了成倍的增加。训练完之后, 准确率有了极大的提高。

(1) 后一分钟涨跌分类准确率, 84.79%:

```
Iter 850000, Minibatch Loss= 0.573726, Training Accuracy= 0.86400
Iter 875000, Minibatch Loss= 0.605627, Training Accuracy= 0.85000
Iter 900000, Minibatch Loss= 0.624047, Training Accuracy= 0.84800
Iter 925000, Minibatch Loss= 0.631733, Training Accuracy= 0.84600
Iter 950000, Minibatch Loss= 0.630488, Training Accuracy= 0.84400
Iter 975000, Minibatch Loss= 0.579400, Training Accuracy= 0.86000
Optimization Finished!
Model saved.

Model restored.
('Testing Accuracy:', 0.84793115)
```

(2) 后两分钟涨跌分类准确率, 69.12%:

```
Iter 850000, Minibatch Loss= 1.057175, Training Accuracy= 0.68000
Iter 875000, Minibatch Loss= 0.999548, Training Accuracy= 0.70800
Iter 900000, Minibatch Loss= 1.013177, Training Accuracy= 0.70000
Iter 925000, Minibatch Loss= 1.028159, Training Accuracy= 0.68600
Iter 950000, Minibatch Loss= 1.000371, Training Accuracy= 0.71200
Iter 975000, Minibatch Loss= 1.003331, Training Accuracy= 0.70400
Optimization Finished!
Model saved.

Model restored.
('Testing Accuracy:', 0.69120592)
```

(3) 后三分钟涨跌分类准确率，58.4%：

```
Iter 850000, Minibatch Loss= 1.139645, Training Accuracy= 0.58200
Iter 875000, Minibatch Loss= 1.080260, Training Accuracy= 0.63400
Iter 900000, Minibatch Loss= 1.161784, Training Accuracy= 0.58000
Iter 925000, Minibatch Loss= 1.114544, Training Accuracy= 0.62200
Iter 950000, Minibatch Loss= 1.071264, Training Accuracy= 0.64200
Iter 975000, Minibatch Loss= 1.098840, Training Accuracy= 0.62800
Optimization Finished!
Model saved.

Model restored.
('Testing Accuracy:', 0.58405048)
```

(4) 后四分钟涨跌分类准确率，54.6%：

```
Iter 850000, Minibatch Loss= 1.240263, Training Accuracy= 0.53600
Iter 875000, Minibatch Loss= 1.249962, Training Accuracy= 0.53400
Iter 900000, Minibatch Loss= 1.231785, Training Accuracy= 0.55600
Iter 925000, Minibatch Loss= 1.232758, Training Accuracy= 0.54000
Iter 950000, Minibatch Loss= 1.233661, Training Accuracy= 0.54800
Iter 975000, Minibatch Loss= 1.255963, Training Accuracy= 0.52800
Optimization Finished!
Model saved.

Model restored.
('Testing Accuracy:', 0.54603326)
```

(5) 后五分钟涨跌分类准确率，49.88%：

```
Iter 850000, Minibatch Loss= 1.309224, Training Accuracy= 0.50200
Iter 875000, Minibatch Loss= 1.283159, Training Accuracy= 0.52800
Iter 900000, Minibatch Loss= 1.342526, Training Accuracy= 0.49800
Iter 925000, Minibatch Loss= 1.278099, Training Accuracy= 0.53000
Iter 950000, Minibatch Loss= 1.328031, Training Accuracy= 0.47600
Iter 975000, Minibatch Loss= 1.345263, Training Accuracy= 0.49400
Optimization Finished!
Model saved.

Model restored.
('Testing Accuracy:', 0.49888012)
```

三、修改一后的策略测试结果

(1) 实时分类结果与得分显示：

inited	trading	pos	pos_rec	count	count_bar	real_class	predict_1	predict_2	predict_3	predict_4	predict_5	score	acc
True	True	0	0	42	42	3	3	3	5	5	5	3.8	6

其中，real_class 是当前时刻的实际分类，predict_1 至 predict_5 为预测的后 1-5 分钟的分类。acc 是今天所有分钟的 predict_1 与 real_class 的比值，即后一分钟预测成功率。（显示不合常理，可能代码有误，仍需修改）

(2) 模型分析

修改后的模型在测试时性能有了明显改善，但是在实际应用中还是效果不好。之前是太多的出现大涨大跌的预测，现在是预测结果不灵敏，经常连续几分钟始终保持同样的预测结果。怀疑是 IC 和 IH 的涨跌幅与 IF 的涨跌幅区间不同，导致分类阈值设置的不准确有关。考虑更多的使用 IF 合约，加上前几年的和最近时间的。

四、修改二

修改一呈现出的种种不足，分析最主要的原因是分类阈值的问题。同时，从以下几个方面对模型进行了较大的改动。

(1) 继续增大数据量。用上了数据源中的所有分钟数据，从 14 年到 16 年 11 月，合约包括 IC、IF、IH 三种，经过清洗后的数据量为 33 万条。

(2) 分类改变。之前是将三种合约合起来分类，由于三种合约涨跌幅差异比较大，导致分类效果很差。现在使用归一化后的结果进行分类，并且将三种合约区分开，避免合约不同带来的差异影响。同时，细化分类，由 5 类变为 7 类。统计每种合约的区间分布，使三种合约的分类比保持基本相同。如下图：

```
IF合约:
count      146399.000000
mean        0.462578
std          0.092574
min          0.000000
25%          0.432531
50%          0.461256
75%          0.489928
max          1.000000
Name: RaiseDown_1, dtype: float64
涨幅3: 1849 0.0126298676904
涨幅2: 4056 0.0277051072753
涨幅1: 18540 0.126640209291
平稳: 98495 0.672784650168
跌幅1: 17003 0.116141503699
跌幅2: 4666 0.031871802403
跌幅3: 1790 0.0122268594731
```

```
IC合约:
count      92931.000000
mean        0.443013
std          0.124230
min          0.000000
25%          0.375561
50%          0.439235
75%          0.505320
max          1.000000
Name: RaiseDown_1, dtype: float64
涨幅3: 1207 0.0129881309789
涨幅2: 2838 0.0305387868419
涨幅1: 11337 0.121993737289
平稳: 62569 0.673284479883
跌幅1: 11079 0.11921748394
跌幅2: 2870 0.0308831283425
跌幅3: 1031 0.0110942527251
```

```
IH合约:
count      95090.000000
mean        0.408681
std          0.071421
min          0.000000
25%          0.393536
50%          0.406805
75%          0.419971
max          1.000000
Name: RaiseDown_1, dtype: float64
涨幅3: 数量: 975 比例: 0.0102534441056
涨幅2: 数量: 3084 比例: 0.0324324324324
涨幅1: 数量: 11522 比例: 0.121169418446
平稳: 数量: 63855 比例: 0.671521716269
跌幅1: 数量: 11419 比例: 0.120086234094
跌幅2: 数量: 3219 比例: 0.0338521400778
跌幅3: 数量: 1016 比例: 0.0106846145757
```

(3) 更改涨跌幅计算方法。如计算第四分钟涨跌幅，之前是计算第四分钟相对于第一分钟的涨跌，现在是计算第四分钟相对于第三分钟的涨跌。这样能够更好的分析交易情况图，并且大大减少大涨大跌情况出现的频率。

(4) 更改训练集和测试集的比重。由 8:2 改为 7:3。后者是公认最好的训练测试集比重。

五、修改二后的策略测试结果

1、模型准确率（后 1 分钟忘记截图）

(1) 后两分钟涨跌分类准确率，67.3%：

```
Iter 500000, Minibatch Loss= 1.012756, Training Accuracy= 0.64740
Iter 1000000, Minibatch Loss= 0.981502, Training Accuracy= 0.66000
Iter 1500000, Minibatch Loss= 0.969737, Training Accuracy= 0.67060
Iter 2000000, Minibatch Loss= 0.900392, Training Accuracy= 0.68040
Iter 2500000, Minibatch Loss= 0.909509, Training Accuracy= 0.67700
Optimization Finished!
Model saved.
Model restored.
('Testing Accuracy:', 0.67268473)
```

(2) 后三分钟涨跌分类准确率，67.46%：

```
Iter 2600000, Minibatch Loss= 1.084730, Training Accuracy= 0.68060
Iter 2650000, Minibatch Loss= 1.069903, Training Accuracy= 0.68100
Iter 2700000, Minibatch Loss= 1.084430, Training Accuracy= 0.68060
Iter 2750000, Minibatch Loss= 1.086828, Training Accuracy= 0.67880
Iter 2800000, Minibatch Loss= 1.107975, Training Accuracy= 0.66760
Iter 2850000, Minibatch Loss= 1.101916, Training Accuracy= 0.67240
Iter 2900000, Minibatch Loss= 1.114731, Training Accuracy= 0.66300
Iter 2950000, Minibatch Loss= 1.096027, Training Accuracy= 0.67180
Optimization Finished!
Model saved.
Model restored.
('Testing Accuracy:', 0.67458671)
```

(3) 后四分钟涨跌分类准确率，67.1%：

```
Iter 1200000, Minibatch Loss= 1.084761, Training Accuracy= 0.67860
Iter 1250000, Minibatch Loss= 1.106140, Training Accuracy= 0.66500
Iter 1300000, Minibatch Loss= 1.114344, Training Accuracy= 0.66880
Iter 1350000, Minibatch Loss= 1.115861, Training Accuracy= 0.66200
Iter 1400000, Minibatch Loss= 1.102250, Training Accuracy= 0.67020
Iter 1450000, Minibatch Loss= 1.086806, Training Accuracy= 0.67240
Optimization Finished!
Model saved.
Model restored.
('Testing Accuracy:', 0.67105156)
```


(4) 后五分钟涨跌分类准确率，67.1%：

```
Iter 1200000, Minibatch Loss= 1.080665, Training Accuracy= 0.67460
Iter 1250000, Minibatch Loss= 1.094642, Training Accuracy= 0.67540
Iter 1300000, Minibatch Loss= 1.087011, Training Accuracy= 0.68180
Iter 1350000, Minibatch Loss= 1.118262, Training Accuracy= 0.66700
Iter 1400000, Minibatch Loss= 1.093449, Training Accuracy= 0.67420
Iter 1450000, Minibatch Loss= 1.092118, Training Accuracy= 0.67420
Optimization Finished!
Model saved.
Model restored.
('Testing Accuracy:', 0.67093211)
```

2、分类动态显示

inited	trading	pos	pos_rec	count	count_bar	real_class	predict_1	predict_2	predict_3	predict_4	predict_5	score	acc
True	True	1	1	97	97	7	5	7	3	2	1	4.3	0.527777...

从一下午的情况来看，预测的分类不再出现很多大涨大跌的情况，也不会一直保持平稳预测，更加贴近实际情况了。

但是，准确率还是远远不足。一分钟预测的准确率只能达到 50%多一些，其中平稳的预测占了大多数，难以实现稳定的盈利。

3、模型和策略分析

(1) 虽然前两分钟准确率比修改一降低了，但是 5 分钟之间的预测准确率相差很小，**对于更长时间的预测是一大利好。**

(2) 数据已经尽可能的利用了，改进空间不大。但预测准确率还不能使人满意，因此需要不断探索**更好的模型结构**，使准确率不断提高。

(3) 另外，为了更好的反映出预测是否有效，正在尝试在策略中添加代码，使其能够将**预测数据输出**，进行进一步的分析，并有针对性的改进。