

期货数据

模式识别
40240452

April 4, 2018

数据文件

- 网络学堂课程文件: futuresData.7z
- 密码: pattern
- 注: 仅限课堂内部使用

交易的简单解释

为了更好的理解数据表达的含义，本节对交易指令处理过程做简单的解释。

- 1 买卖双方 within 系统内申报价格
- 2 系统将同一合约的买卖申报指令排序，形成买单卖单列表。
主要按照价格优先原则，若价格相同则按时间顺序。
 - 对于买方来讲，价格高优先成交
 - 对于卖方来讲，价格低优先成交

买入价	买入量	卖出价	卖出量
6	2	5	3
5	3	6	1
4	2	7	2

交易的简单解释

- 3 若买入价 \geq 卖出价, 则开始交易。
- 4 (6|2) 价格比 (5|3) 高, 可以交易 2 个合约单位, 即累计成交数量增加 2, 累计成交金额增加 2* 成交价格。同时买单买单变为:

买入价	买入量	卖出价	卖出量
5	3	5	1
4	2	6	1
-	-	7	2

交易的简单解释

- 5 (5|3) 价格等于 (5|1)，可以交易 1 个合约单位，即累计成交数量增加 1，累计成交金额增加 $1 \times \text{成交价格}$ 。同时买单卖单变为：

买入价	买入量	卖出价	卖出量
5	2	6	1
4	2	7	2
-	-	-	-

交易的简单解释

- 6 此时买入价 < 卖出价，列表内买单卖单不继续进行交易，等待新买卖申报。剩余买单卖单根据优先顺序，分部称为买 1, 买 2, ..., 卖 1, 卖 2, ...

买入价	买入量	卖出价	卖出量
5 (买 1 价)	2 (买 1 量)	6 (卖 1 价)	1 (卖 1 量)
4 (买 2 价)	2 (买 2 量)	7 (卖 2 价)	2 (卖 2 量)
- (买 3 价)	- (买 3 量)	- (卖 3 价)	- (卖 3 量)

成交价格确定

- 取买入价、卖出价和前一成交价（已确定成交最后一笔交易的成交价）的中值。
- 这样可以保证交易价格不比申请时的买入价高，也不比申请时的卖出价低。
 - 对之前的例子假定最初时前一成交价为 6
 - 过程 4 中，前一成交价为 6，买入价为 6，卖出价为 5，则成交价格为 6
 - 过程 5 中，前一成交价变为 6，买入价为 5，卖出价为 5，则成交价格为 5

期货交易的时间

- 每周一到周五，国家法定假日及周六日期货市场不能交易。
- 不是 24 小时交易，中午和半夜需要休息。
- 这意味着重新开始交易，交易所会重置最高价、最低价。

涨停，跌停

- 每天的价格变化有一定限制，当变化幅度超过限制时，不允许超过变化幅度的交易，成为涨停和跌停。
- 当然，此时相反方向还可以交易。即涨停时可以卖出，跌停时可以买入。

数据说明

- 该数据为 2017 年 7-8 月的数种期货数据
- 频率：500ms (每 500ms 检查行情数据。若行情数据变化，将新行情写入文件)
- 文件名由 (大类)-(日期)-(day/night).log 组成，分别记载了对应种类和时间期货的行情记录。

数据说明

- 每个 log 文件内有许多记录，有效样例记录示例如下：

```
2017-08-08 09:00:00.688 [KeyedThreadPool-3-1] INFO - Quote[lastPrice=40950000,highestPrice=41080000,lowestPrice=40560000,volume=971204,turnover=39657599940000,bidPrice1=40950000,bidVolume1=326,askPrice1=40960000,askVolume1=70,instrumentID=B2,exchangeID=]
2017-08-08 09:00:00.694 [KeyedThreadPool-64-1] INFO - Quote[lastPrice=39480000,highestPrice=39740000,lowestPrice=39160000,volume=1626802,turnover=64236528480000,bidPrice1=39480000,bidVolume1=162,askPrice1=39520000,askVolume1=104,instrumentID=B3,exchangeID=]
2017-08-08 09:00:01.183 [KeyedThreadPool-3-1] INFO - Quote[lastPrice=40950000,highestPrice=41080000,lowestPrice=40560000,volume=971458,turnover=39668003520000,bidPrice1=40950000,bidVolume1=329,askPrice1=40960000,askVolume1=6,instrumentID=B2,exchangeID=]
2017-08-08 09:00:01.184 [KeyedThreadPool-64-1] INFO - Quote[lastPrice=39490000,highestPrice=39740000,lowestPrice=39160000,volume=1627528,turnover=64265204380000,bidPrice1=39490000,bidVolume1=189,askPrice1=39510000,askVolume1=156,instrumentID=B3,exchangeID=]
2017-08-08 09:00:01.679 [KeyedThreadPool-3-1] INFO - Quote[lastPrice=40980000,highestPrice=41080000,lowestPrice=40560000,volume=971776,turnover=39681029120000,bidPrice1=40960000,bidVolume1=6,askPrice1=40980000,askVolume1=192,instrumentID=B2,exchangeID=]
```

数据说明

- 记录开头为该条记录的时间，格式如样例所示。
- instrumentID: 合约 ID，代表这是哪一个合约
- Price 为对应的价格
 - highestPrice: 最高价
 - lastPrice: 最新成交价
 - lowestPrice: 最低价
- turnover: 累计成交金额
- volume: 累计成交数量
- 买方出价 (bid)/卖方报价 (ask)
 - askPrice1: 卖 1 价
 - askVolumn1: 买 1 量
 - bidPrice1: 买 1 价
 - bidVolumn1: 买 1 量

根据过去一段时间（长度建议在 10 几秒到几分钟，最长不超过十分钟）的行情，预测未来 10 秒内价格变化的方向。在实验进行前，需要对数据进行一点的标注。

- 不限定标注方式，建议使用自动标记方式来对数据进行标注。
- 对如下合约进行处理

合约

A1,A3,B2,B3

- 例如分为 3 类, 未来 10 秒价格 / 当前价格

$$\begin{array}{ll} > 1+\theta_1 & \Rightarrow \text{上涨} \\ 1-\theta_1 \sim 1+\theta_1 & \Rightarrow \text{稳定不变} \\ < 1-\theta_1 & \Rightarrow \text{下跌} \end{array}$$

■ 例如分为 5 类, 未来 10 秒价格 / 当前价格

$> 1+\theta_2$	\Rightarrow 强烈上涨
$1+\theta_1 \sim 1+\theta_2$	\Rightarrow 一般上涨
$1-\theta_1 \sim 1+\theta_1$	\Rightarrow 价格方向不变
$1-\theta_2 \sim 1-\theta_1$	\Rightarrow 一般下跌
$< 1-\theta_2$	\Rightarrow 强烈下跌

- 两个示例中的‘未来 10 秒价格’是一个模糊的说法，可以有很多不同的定义。
- 使用的阈值 (θ_1, θ_2) 也可以自行选择，比如 0.2%, 0.3%, 0.4% 等.
- 这些期货价格变化都很快，有时 10 秒至 20 秒就能产生 0.5% 甚至更大的变化

工具

- 语言不限: C/C++, Matlab, R, Python,
- 可视化: matplotlib, plotly, ...
- 数据整理: re, numpy, pandas, ...
- 机器学习:
 - scikit-learn
 - Torch/PyTorch
 - Tensorflow
 - Keras, Mxnet, ...
 - Matlab Statistics and Machine Learning Toolbox

作业内容

根据过去一段时间（长度建议在 10 几秒到几分钟，最长不超过十分钟）的行情，预测未来 10 秒内价格变化的方向。

- 如前所述，不限定标注方式，进行数据预处理
- 根据自己的分类标签，对价格变化方向进行预测
- 预测方法（任选）：
 - 简单的线性分类器
 - 线性回归，曲线回归，逻辑回归
 - 隐 markov 模型 HMM