

笔试题答案

张林-香港科技大学-金融数学

问题一：

项目链接地址：<https://github.com/Meliere/Au-Prediction.git>

（‘WebAnalyze.py’ 以及 ‘MFIAnalyze.py’ 为 Python 代码，‘au.csv’ 为黄金主力合约日线数据，‘words.csv’ 为部分关键词）

（一） 开发环境：

Python 版本：python 3.5

使用 IDE：Spyder

（二） 整体思路：

预测价格的整体思路为两方面的分析：

1) 舆情数据分析：

利用 Python 爬取上期所黄金国际市场信息

<http://www.shfe.com.cn/bourseService/marketresearch/internationalmarket/>

与新浪财经黄金分析

http://roll.finance.sina.com.cn/finance/gjs/hjfx/index_1.shtml

网页上近期文章，对文章中内容进行关键字匹配，对不同的关键字给予不同的舆情分值（越正代表市场情绪越乐观，反之亦然）。

2) 对交易量指标进行分析：

市场交易量是反应市场情绪的重要因素，因此结合交易量的指标对价格走势进行综合分析。

根据个人经验，结合交易量的指标中，MFI 较为有效（在黄金期货主力合约中，MFI 能较大幅度准确追踪分钟级数据短期的趋势）。因此结合日线 MFI 指标，采用时间序列分析的方法进行短期预测。

根据两种方法返回的结果，结合实际情况对价格走势作出判断。

（三） 数据来源及理由：

黄金期货历史价格数据：港科大数学系万德数据终端下载，数据时间段为 2012 年 1 月 1 日至 2017 年 5 月 10 日；数据频率为日线数据，包括每日内最高价

（HPrice）、最低价（LPrice）、收盘价（CPrice）以及交易量（NTrade）。其中

2016 年 1 月至 2017 年 2 月 API 所搜集数据结果异常，数据质量不高，因此采用分钟级数据转换替代（先前从淘宝购得，只有 2016 年 1 月至 2017 年 3 月的数据）。

理由：对于五日区间价格的预测，以天为单位的数据较为合适，同时国内期货小时甚至分钟级别的数据起始时间均较晚，分钟单位数据跨越天数长度较小，故采用日线数据。

（四） 程序流程：

代码分为两部分：‘WebAnalyze.py’ 以及 ‘MFIAnalyze.py’。

1) ‘WebAnalyze.py’ 流程：

- i. 用 request 函数抓取网页数据，并用 beautifulsoup 函数进行整理。
- ii. 搜索原网页中各条文章的链接。
- iii. 对各个文章网页进行访问，获取文章原文于本文库中词汇进行比对，计算累积分值。

2) ‘MFIAnalyze.py’ 流程：

- i. 对日线数据计算 26 日 MFI 的值。
- ii. 生成 MFI 时间序列类，作图并作出（偏）自相关系数。
- iii. 根据图形定出大致阶数，确定参数，对指标进行大致的估计。

（五） 预测情况：

- 1) 采用“空头回补”、“黄金需求量减少”和“美元走强”等词汇对上述两个网页的各文章内容进行匹配搜索，每个词汇对应一点负面情绪；用“超卖”、“提振金价”等词汇进行匹配，每个词汇对应一点正面情绪，结果表明最终结果负面情绪居多，总分值约为 -22，但其值较前几日已有回升。
- 2) MFI 以连续多日处于较低水平，同时时间序列分析结果表明其较平稳具有二阶自回归效应，其系数约为 0.67 和 0.211，说明短期内大概率难以恢复正常值，但是其较上周极端低值已有回升。

结论：综合以上分析，黄金主力合约未来五日价格走势依然大概率下行，但下行速度可能放缓。

（六） 未完成部分及思路：

- 1) 在对网页信息的分析中，没有考虑网页上的时间的影响，可以考虑对时间尺度给予不同的权重分配（例如指数下降的权重）。
- 2) 在对市场新闻的分析中，对比词库的选择以及不同词汇的计分分值的确定较为困难且影响较大。对于这一部分，也许可以采用历史价格走势和历史新闻样本进行词汇与市场情绪关联的采样和分析，从而实现对于文本的简单挖掘。或许

也可以使用 Boosting 和卷积神经网络（模拟图片的分析方法对整篇文章进行分析）对不同词汇的排列组合、影响情况进行训练。

- 3) 不同网页的文章和报道存在转载和借鉴等方面的问题，可能导致同一信息源多次出现，造成搜索结果与实际的偏差；但转载和借鉴也对信息起到了过滤和筛选的作用，同时出现次数也可能对市场情绪有不同的影响。或许可以根据不同网页的历史情况做出分类和权威性（根据搜索量访问量，例如百度指数）和关联性（内容的相似度、文章引用的重合度）的打分，进一步精确网页信息的影响。
- 4) 传统技术指标、基本面指标与网页信息之间可能存在时间差（指标信息领先于网页上的文章），或许可以量化各网页信息反应的市场舆情之后分析其余各个指标的相关性，起到提前估计的作用。

问题二：

（一）必备物品：

- 1) 衣物鞋帽：冲锋衣裤、登山鞋、手套、帽子等。

根据个人体质、当时的季节气候和所登山的海拔高度决定衣物的厚度；雨季必带雨衣。

- 2) 装备：打火机、指北针、垃圾袋，瑞士军刀，登山杖、头灯或手电（保持足够备用电池）。
- 3) 通讯：手机、轻便的移动电源。
- 4) 食品：压缩饼干、能量棒、水等，根据面试官体能状况与同行人数，确定携带数量。
- 5) 药品：创可贴、绷带、云南白药、止痛药、藿香正气、以及根据面试官身体条件准备个人特药。
- 6) 日用品：纸巾、湿巾、免洗洗手液、驱蚊水

上述物品都较小，可整理放置于一个中等大小的背包中。

（二）在准备前需要同面试官充分沟通，了解面试官登山的主要意图和个人偏好，对上述物品进行适当选择。

（三）其他物品：

- 1) 若此次登山是一日内的短途出行，上述物品还可适当删减；若该出行计划为 2 日以上野外宿营，还须带帐篷、防潮垫、睡袋、炉头、气罐、套锅等。

- 2) 若此次登山性质为休闲娱乐为主，可以考虑准备相机，若是多人结伴出行，可以考虑准备一些方便携带的娱乐物品（如扑克）；若登山更偏向于探险性质，则要在应急物品上多下功夫，最好登山队伍中每人各备对讲机以便在没有信号的地方也能保持联系。
- 3) 在上述基础上同面试官确认是否有其他需要准备的个人特殊物品。