杨林彬 2016年9月9日星期五

这周主要是把一个暑假的编码结果进行了规范化整理，通过整理做了一些分类优化，也发现了暑假编码中存在的一些问题。目前除了香港联合交易所，所有已有的UK已经整理好。

优化：

1、 以上海金属期货交易所为例，其产品都是金属期货。那么文档就按照不同的品种来分类，每一个文档记录着对应种类的金属期货的10年的UK编码。

问题：

1、 暑假的期权编码我们是按照自己的一套方案来编码的

(1、一个月的上证50ETF期权按照其权利金的不同有许多支。2、期权在市场上的波动非常频繁。因此我们考虑采用这样的方案:按照月份来给期权编码，并把“交收价格”就看做是权利金，每个月份期权代码最小的期权的权利金令为0，作为基准。其他的高于基准的分别令为1、2、3，类似地，低于基准的权利金就令为-1，-2，-3。随着期权价格的变化，上下浮动的价格序号也会做出相应的调整)

因此在UKID的设置上，和其他产品会稍有不同，其他产品的UKID是直接去取UK码的后8位。而期权的UKID，是按照数据在Excel中的排序直接转化为8位64进制码的。

2、 香港联合交易所的数据暑假做得不够好，准备重新做一次（数据比较乱）。

这周收到张涵雅发来的数据，还没有来得及去处理(那个数据还需要分类)。

杨林彬2016年9月16日星期五

本周主要的做得是学习了一下sqlite数据库，这是一种轻量级别的数据库。安装也非常简单，我用python写了一个脚本，把深圳交易所的UK数据导入了数据库中并保存为DB文件。公司那边要求把提供三种下载数据：1、Excel 2、CSV 3、sqlite。现在数据整理工作全部完成了，考虑到编写wiki是一个较为费时的工作，我想的是先不着急，每周都带着做一点。这两周一直在做完善暑假成果的工作，UK项目没有向前推进。下周我需要转变一下思路，制定好计划了。

这周再学习之余，我还看了一个机器学习的算法，KNN邻居算法，目前算法已经跑通，感觉非常有意思，我还会坚持看下去。

张涵雅那边给出了筛选出来的数据，我负责把这些数据转化为UK。我拿到的数据是原始股票代码，我一开始是想着一个个校对，这显然是不现实的。从数据特点分析，我发现这是股票和基金，目前我们的Excel数据库中还没有这一项。因此我想的是先把哪个txt原始股票代码文件做成Excel数据库的样子，再把原来的UK Excel数据库中加上原始代码这一列，还是勇python写脚本同时处理这两个文件，实现自动对应，这样应该就能大大提高工作的效率。

杨林彬2016年9月24日星期六

这周主要是完成了和张涵雅同学那边的一点衔接上的工作，目前我已经收到那边传过来的上海和深圳的股票代码，我需要把它们翻译成UK。技术上就是把股票代码和我建立的数据库中的数据进行匹配，这一过程还是利用python脚本进行处理。

除此之外本周还把香港联合交易所的期货做了一点，香港联合交易所期货分类较为复杂，需要整理。（如下所示）

下周的打算是把期权的编码方式总结好，看看有没有可能有更好的编码方案了。等各种产品的编码方案成熟后，就可以着手项目专利的书写了

杨林彬2016年10月1日星期六

这周校庆，实验室的工作确实没有做了多少。所以我就讲一讲我国庆的计划吧。

1、上周四和龙哥交流了一下，UK项目需要做一个实时刷新的应用，这样的话就需要好好想想应该如何实现了。我还没有思路，因为想着没有一个稳定的数据源，不同交易所的情况又有不同。

2、希望自己把数据库全部做完（sqlite、Mysql），sqlite是由于wiki的需要。但是wiki现在出了问题，而且DB文件无法提交，这些正在和龙哥商量解决中。

3、我目前在跟进学习coursera的Machine Learning课程，每周都会在自己的github上上传学习日志，大家可以去我的github上查看（如果大家感兴趣）

杨林彬2016年10月15日星期六

本周因为个人原因实验室工作有点脱节。不过那就想一下下一周的工作安排吧：

1、廖总下一周要来实验室视察，我需要把最近的工作情况整理一下。

2、下周争取把数据库做完，不然应用开发没办法开头。这里会涉及到数据库的结构设计，数据导入等等。

3、本周希望把落下的Machine Learning课程给补上