# 基于历史数据学习的证券上涨率预测方法

基于各平台公示证券编号、开盘价、收盘价、最高价、最低价、成交额、成交量等信息，我们可以应用常用股票指标计算公式来，使用适当多的指标来完成最终上涨概率的预测。本程序分为四个部分：第一，数据读取处理，即将平台提供的.day格式文件读取为matlab中含编号、开盘价等信息的矩阵；第二，指标值计算，即大量get前缀的函数，以及指标应用，即大量check前缀函数，基于指标值，计算某日股票上涨与否的预测值；第三，历史数据学习，对每一项指标，施加大量的历史数据训练，以得到预测正确次数与有效应用次数的比值；第五，轮盘竞争，以各指标预测正确率为轮盘比例，模拟竞争权值。第六，基于各项指标权值，计算最后的某股票上涨概率。

# 理论依据

## 核心思想

本方法基于股票指标计算公式及其应用，主要手段为在历史数据中找到指标的正确率，并通过轮盘竞争得到最终各指标权重wi.

在最终的预测公式里，上涨概率P=, 其中wi为第i个指标的权重，fi为该指标预测结果。

K为有效指标数，这里当某指标“无效”时，它将不参与该计算式，因此K中不包括该指标。同时,权重值会因此改变，但相对大小没有变化。

## 权重

在单个神经元P=（阈值为0）中，以每个指标函数为内部函数。

准确率反映对事物的正确解答，在一定实验条件下的多个测定值中，满足限定条件的测定值所占的比例，我们以这个比例为股票上涨率，即“满足上涨的测定值所占比例”，这里“测定值”为内部函数，由内部函数P=判定返回两个值（0或1），以1（上涨）为目标accepted value。

# 数据读取处理

## .day文件

Day格式文件即脱机个股资料、通信达数据，包含某支股票证券编号、开盘价、收盘价、最高价、最低价、成交额、成交量信息，这些信息即为此算法中所使用的股票数据信息。



图1.1 目标文件夹内的.day文件

## 1.2 函数readpath

函数 readpath,输入.day文件所在路径，返回**元胞数组U**，U的成分为各支股票的数据**信息矩阵M**（即证券编号、开盘价、收盘价、最高价、最低价、成交额、成交量信息）。

***示例：****（为便于调试，这里只使用3个.day文件测试）*

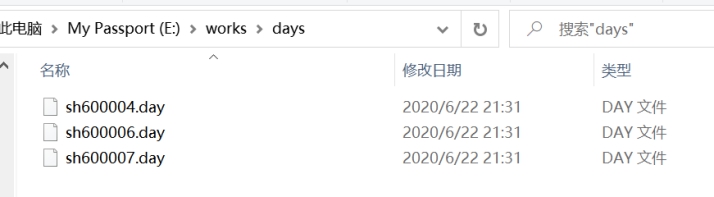
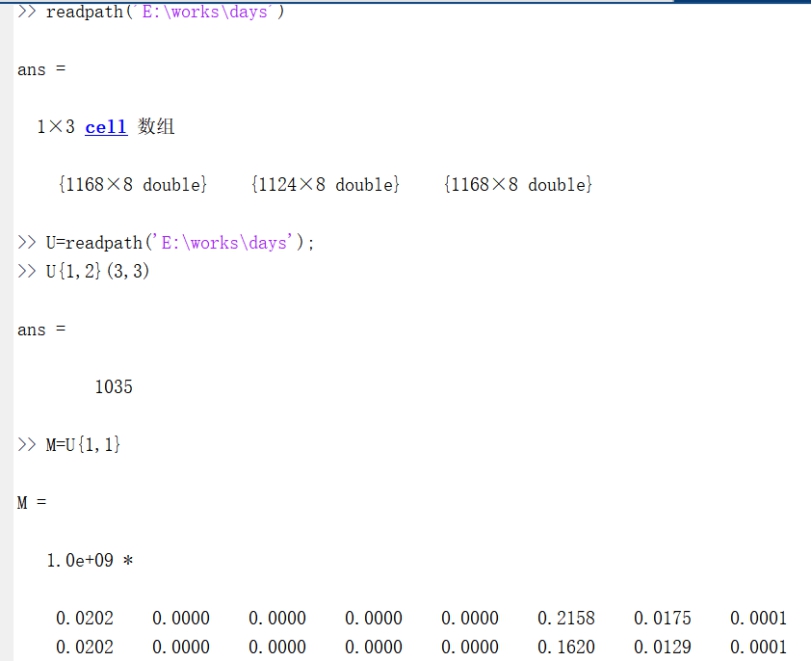


图1.2a 目标文件夹

*命令行窗口，见图1.2b：*

* *1035为sh600006股的第三日最高价*
* *M数据矩阵为sh60004（这里未使用format函数进行记数法处理，但是由结果可得double类型数据足够精确）*
* *最后显示sh600006的第三天日期（2015年9月8日）*



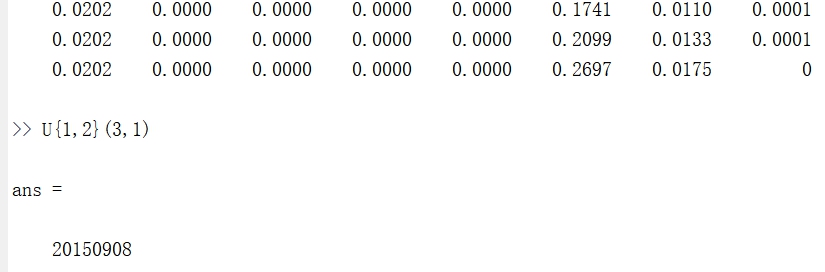


图1.2b readpath函数在matlab中的运行

# 指标值计算及应用

## 股票指标

股票指标依据一定的数理统计方法，运用一些复杂的计算公式，基于数据来论证股票趋向、买卖等。因此可分为两个步骤，其一为指标值计算，其二为指标值应用，得到某支股票是涨跌预测。

本算法应用指标有KDJ随机指标，MACD（平滑异同移动平均线）指标，TRIX三重指数平滑移动平均指标，CR中间意愿指标，BRAR人气意愿指标，RAI相对强弱指标，OBC能量潮指标，BIAS乖离率指标，CCI顺势指标，WR威廉指标共14项指标进行分析。

## get类函数

该类型函数用作计算各特征值，以KDJ为例，返回包含values的向量，注意这里KDJ实则包含K值、D值、J值、金叉判断值（J）等，同时每个向量的首项都为一个flag，当该特征无法计算时，通常是数据天数不同，flag为0，可以计算则为1，便于后续函数判定。

图2.2a getKDJ函数声明 图2.2b getKDJ函数返回向量R

## check类函数

该类函数输入向量为get函数返回的R，条件控制语句计算得到结果上涨（1），下跌（-1），特征无效（0）.注意，这里的无效与get函数中的无效不同，前者通常是数据天数不足以计算某指标值（它们通常需要当日前13天或者25天的股票数据），后者则是基于指标应用方法，如当K值位于【20,80】中时，我们无法用K指标来预测股票涨跌。

Check类函数返回值通常为1(上涨)，-1（下跌），0（无效），对于个别特征，如金叉（jincha），因为其含义为当股票满足金叉曲线时，则判断为上涨，checkjincha函数将只返回1（有金叉，上涨）或者0（无金叉，该特征无效）

# 历史数据学习

## 正确率

某指标的正确率定义为历史预测正确次数n特征有效总次数N。

其中n的计算方法为基于历史数据，如第x天，我们可以利用指标预测第x+1天该指标相对第x天是涨是跌，即计算check函数的返回值，同时第x+1天的实际涨跌情况也可以得到，将二者相乘，如果同号证明此次预测正确，异号则证明此次预测错误，这两种情况特征有效总次数都将累计一次（N=N+1）。第三种情况是结果为0，即该特征无效，我们并不把它算作特征有效次。其原因是在实践中，当某特征不适用该日股票，我们完全可以不使用该指标以规避错误，不影响其正确率。

## findprop函数

我们使用findprop函数来计算各指标正确率，它的输入变量为两个函数句柄以及数据对象M（单支股票的数据信息矩阵Mi），如图3.2，返回p即正确率。

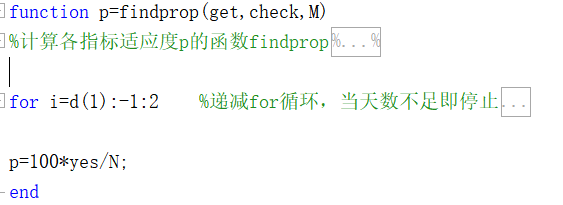


图 3.2 findprop函数的声明及返回值

get句柄和check句柄分别指向getXXX函数及checkXXX函数，此函数可以用于计算所有特征的正确率。

findprop函数中有两个控制语句，其一是检验R向量（getXXX返回值）首项flag是否为1，否则将终止天数循环（该for循环为逆向，即从已有的1168天股票日数据倒数第二天向前循环检测指标正确性），当数据不足以计算指标时，终止循环。其二是检验checkXXX返回值，即当check函数认为当日股票数据的指标不构成可以预测出涨跌条件时，n与N都不会变化(很明显，for循环不会停止)。

该函数为fitness函数的中间函数。

## fitness函数

该函数用于计算某特征在多支股票中的正确率，其输入变量为句柄getXXX、checkXXX,元胞数组U（readpath返回值），返回正确率w。



图3.3 fitness函数声明

# 轮盘竞争

## 轮盘竞争

## 函数RouletteWheelSelection

该函数输入变量为一个向量，【w1,w2,w3,…】，wi的含义为相对正确率（第i个指标的正确率pi 所有指标正确率和 。对每个指标编号index为1，2，3,…, i ,…,14（共14个指标），放入轮盘中随机抽取。最后输出一个值，该值为一个随机变量index,即代表一个指标被抽取

