## 研报复现3：遗传规划&遗传算法

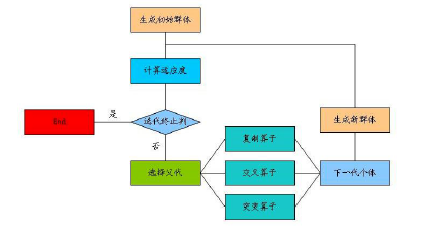
### 一、基于遗传规划的智能交易策略方法

#### **1.策略概述**

遗传规划由达尔文的进化论演变而来，是对遗传算法的推广和更一般的形式。遗传规划所要解决的问题是找到一个函数表达式，不同于线性回归或者是遗传算法，遗传规划并不用初始假定一个函数表达式，从而求取系数，而是自行创造。

具体流程为：先生成函数体群体，每一个个体为一个函数表达式，先初始化群体，之后进入循环迭代，计算个体的适应度，判断迭代终止条件，若不满足则暗适应度比值的选择概率在上一代群体中随机选择个体，通过复制交叉突变等操作生成新个体，再循环操作直到终止迭代。

###### 图1：遗传规划流程图



遗传规划生成可以用树的方式来表达，其中符号为树节点，自变量和常数为叶节点。

#### **2.交易策略**

（1）策略假设：适应度选取策略累计收益率/回撤

该比值越大的个体适应度越高，交易信号为个体函数表达式的值，每个交易日从9:30开始，当交易信号值大于0时，开仓做多，小于0，开仓做空。

（2）符号选取：

模拟中选择了常规的加减乘除（其中除法考虑分子为0情况，采用自定义的protected\_div），同时加入了sin，cos，tan，acos，以及自定义的protected\_pow（防止乘方结果过大，采用了一定的输入限制）。

（3）python实现：

采用基于deap库上的geppy进行实现。

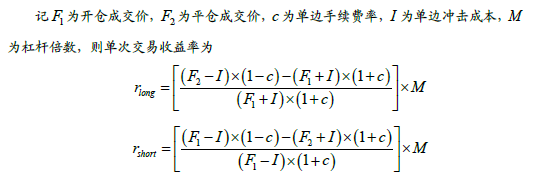
#### **3.实证分析**

**（1）数据选取**：

股指期货合约加权平均（IF8888）在2010年4月16日到2012年8月24日的5分钟K线。其中2010年和2011年为训练集，2012年为测试集。评价指标考察：累计收益，多仓空仓胜率、持有期、最大回撤、年化收益、年华标准差、夏普比率等。

**（2）交易假设**：

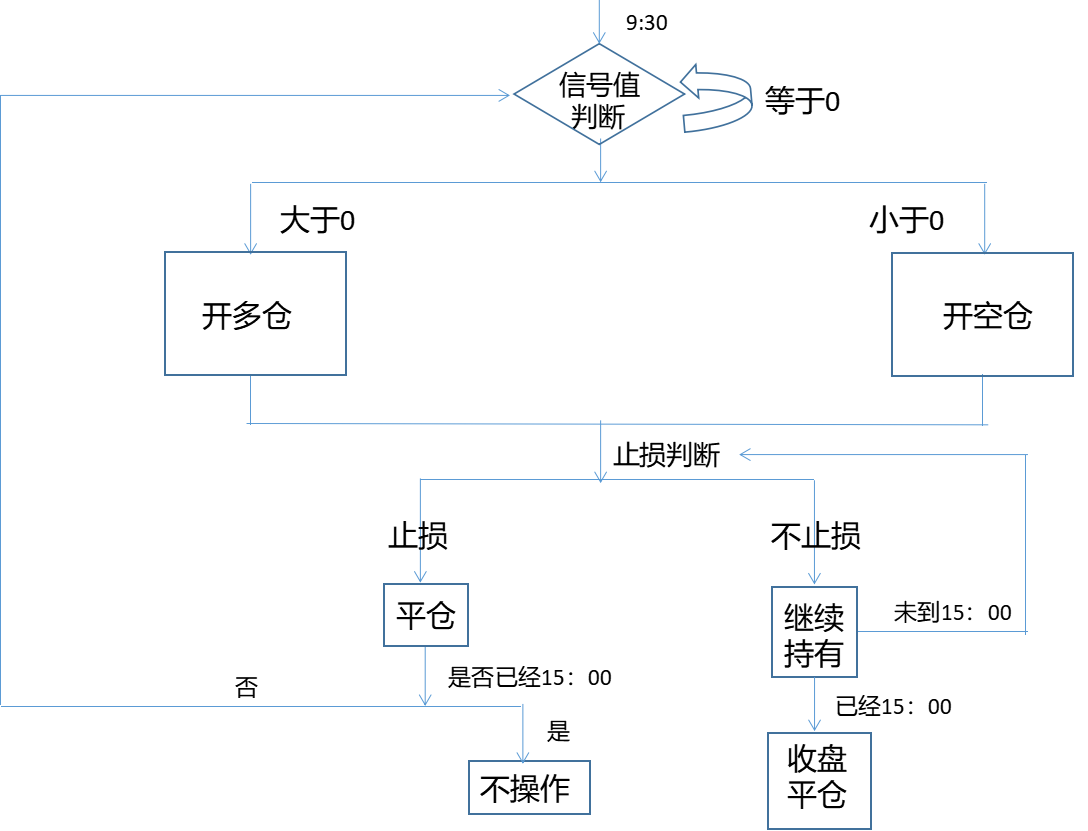
设置单边冲击成本0.4个指数点：



此外单边手续费万分之零点五，开仓价格为信号发生后第一根K线的开盘价格。

加入止损条件，设置固定止损幅度为0.5%，触及止损线时，平仓止损，再根据之后的交易信号值与0的关系开仓，若当日未发生止损，则收盘15:00强制平仓。

###### 图2：判断交易流程图



**（2）回测结果：**

回测的效果都不是很好，达不到研报中的收益率，测试了几个，其中收益率最高的函数表达式在2010-2012年度累计收益为221.87%，但主要是训练集贡献的收益率，在测试集上收益率不高。

具体如下：

**训练集：**

累计收益 203.92%
  
多仓次数 386
  
多仓成功次数 152
  
多仓胜率 39.38%
  
多仓平均持有期 21.99
  
空仓次数 425
  
空仓成功次数 189
  
空仓胜率 44.47%
  
空仓平均持有期 0.44
  
周期胜率 50.43%
  
最大回撤 9.42%
  
年化收益/最大回撤 9.74
  
年化收益 91.80%
  
年化标准差 18.53%
  
年化夏普 4.95

**测试集：**

累计收益 5.26%
  
多仓次数 120
  
多仓成功次数 46
  
多仓胜率 38.33%
  
多仓平均持有期 23.61
  
空仓次数 169
  
空仓成功次数 73
  
空仓胜率 43.20%
  
空仓平均持有期 0.43
  
周期胜率 50.29%
  
最大回撤 7.47%
  
年化收益/最大回撤 1.12
  
年化收益 8.36%
  
年化标准差 14.98%
  
年化夏普 0.56

### 二、基于遗传算法的期指日内交易系统

#### **1.策略概述**

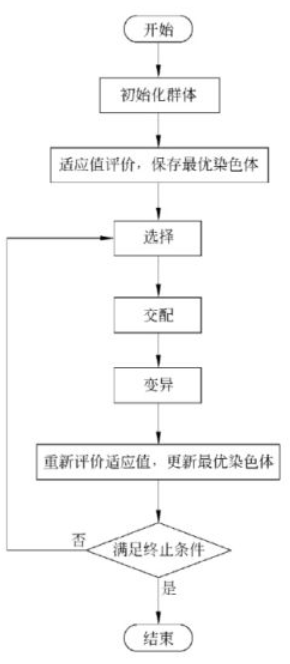
##### **1.1 日内策略**

日内策略是一种相对于隔夜交易策略而言风险较小的一种策略，其中日内突破反转模式应用较为广泛。其逻辑为，如果价格向上突破最高一个价位，进场做多，如果股指达到第二个价位并向下跌破第三个价位，则形成下反转，进场做空；下突破和上反转类似。

##### **1.2 遗传算法**

遗传算法是模仿自然界生物进化机制发展起来的随机全局搜索和优化方法，借鉴了达尔文的进化论和孟德尔的遗传学说。其本质是一种高效、并行、全局搜索的方法，能在搜索过程中自动获取和积累有关搜索空间的知识，并自适应地控制搜索过程以求得最佳解。

###### 图3：遗传算法流程图

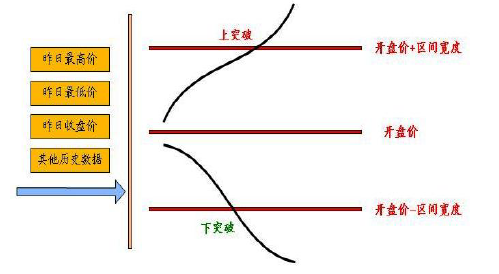


具体流程和上述遗传规划类似，区别在于，遗传算法是进行参数的优化，而遗传规划实际上相当于符号回归。

##### **1.3 基于遗传算法的日内策略**

通过昨日最高价、昨日最低价、昨日收盘价、当日最高价、当日最低价等交易日信息，进行回归，建立回归模型来预测下一个交易日的波动的高低，借鉴开盘突破方法，在当日开盘价的基础上加上一个区间宽度得到上突破价，减去一个区间宽度得到下突破价，当股指突破上突破价时做多，跌破下突破价时做空。

###### 图4：开盘区间突破模式



故如何寻找到一个合适的区间宽度值，便对收益率产生重要影响。这里假设区间宽度值与交易日信息符合线性回归模型，即有如下公式：

其中B为区间宽度值， 为输入变量， 为线性系数，而求取最佳系数的时候，便采用遗传算法，将个体适应度设置为样本内交易的累计收益率。

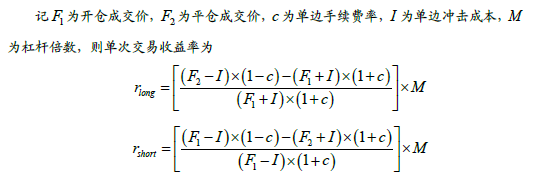
#### **2.实证分析**

**（1）数据选取**：

股指期货合约加权平均（IF8888）在2010年4月16日到2013年2月1日的2分钟K线为样本集。2011年12月31日前数据为训练集，2011年12月31日到2013年2月1日的数据为测试集。评价指标考察：累计收益，多仓空仓胜率、持有期、最大回撤、年化收益、年华标准差、夏普比率等。

**（2）交易假设**：

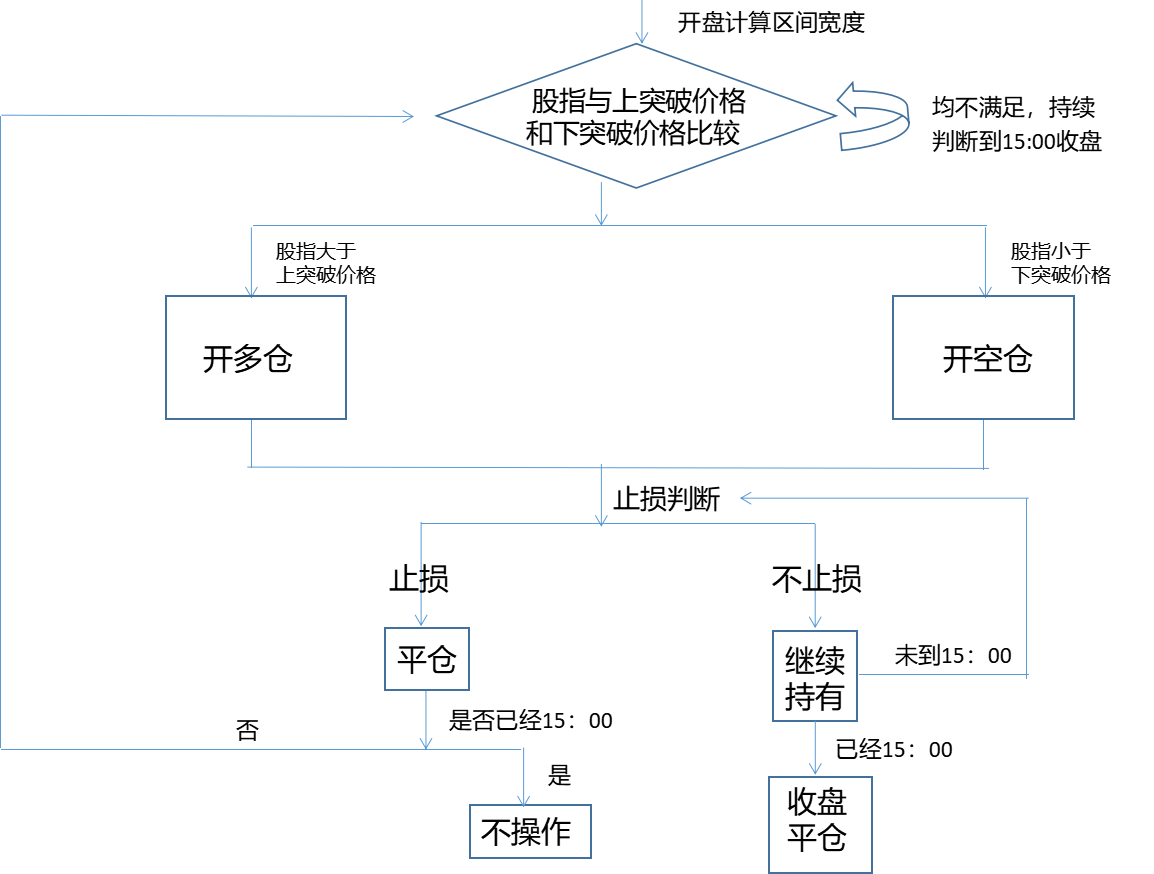
设置单边冲击成本0.4个指数点：



此外单边手续费万分之零点三，开仓价格为信号发生后第一根K线的开盘价格。

加入止损条件，设置固定止损幅度为千分之五，触及止损线时，平仓止损，再根据之后的价格比较开仓，若当日未发生止损，则收盘15:00强制平仓。

###### 图5：判断交易流程图



**（3）区间宽度**：

区间宽度B的线性函数的自变量选取：昨日开盘价，昨日最高价，昨日最低价，昨日收盘价，前日收盘价，真实高价（最高价和T-1日收盘价的最高值），真实低价（最低价和T-1日收盘价的最低值），真实区间（真实高价-真实低价），真实区间10日均值。

上突破价位 = 当日开盘价 + B

下突破价位 = 当日开盘价 - B

为了防止区间过宽，参数的取值范围为-2到2之间。

通过遗传算法计算出的区间宽度B若是负数，那么将当日不进行开仓，为方便将区间宽度设成10000，来满足当日无法开盘的假设。

**（4）python实现：**

采用deap库进行实现。

**（5）回测结果：**

回测的效果同样也不是很好，达不到研报中的收益率，测试了几个，其中收益率最高的参数集合在训练集累计收益为59.56%，但主要是训练集贡献的收益率，在测试集上收益率不高。

参数集取值为：

| ind = [0.6064531271027143, 1.1469476180530012, 1.2453036834458149,  -1.657479142147305, -1.2606158379235208, 0.3857420334185466,  -0.5111631710731213, 1.4795654027014211, -0.9825084296122464] |
| --- |

具体如下：

**训练集：**

累计收益 59.56%
  
多仓次数 34
  
多仓成功次数 25
  
多仓胜率 73.53%
  
多仓平均持有期 59.18
  
空仓次数 42
  
空仓成功次数 24
  
空仓胜率 57.14%
  
空仓平均持有期 55.05
  
周期胜率 52.32%
  
最大回撤 3.37%
  
年化收益/最大回撤 9.60
  
年化收益 32.32%
  
年化标准差 7.88%
  
年化夏普 4.10

**全样本集：**

累计收益 78.30%
  
多仓次数 43
  
多仓成功次数 31
  
多仓胜率 72.09%
  
多仓平均持有期 57.42
  
空仓次数 58
  
空仓成功次数 32
  
空仓胜率 55.17%
  
空仓平均持有期 49.14
  
周期胜率 52.53%
  
最大回撤 3.37%
  
年化收益/最大回撤 6.92
  
年化收益 23.32%
  
年化标准差 6.56%
  
年化夏普 3.55

### 三、总结

总体来说，遗传规划和遗传算法在样本集上的表现都不是很好，每一次跑出来的结果也都不一样，很难保证在测试集上仍表现良好。