## 研报复现1：

## 十三：基于统计语言模型（SLM）的择时交易研究

### 一、策略概述

#### **1.N元模型**

##### **1.1 语言建模**

在给定自然语言L时，利用给定的语言样本估计P(s)的过程被称为语言建模。故可根据语料库中的样本估计出概率分布P，例如对给定句子S=w1w2...wl，计算P(s)公式如下：

即为第一个词出现的概率，为条件概率（已知第一个词w1的前提下w2出现的概率），依次计算可以获得概率。

但是，当词的数量逐渐增多，估算概率更为困难，故引入马尔科夫假设，即一个词的出现的概率仅仅与前n-1个词出现的概率相关，即有：

例如：若仅与前一日有关，则概率如下：

### 1.2 择时交易模型

首先将指数收盘价序列 转换成符号序列 ，转换规律如下：

若 则 ，否则 。（这里设置成01串方便后序二进制算出数组中存放元素的位置）

择时策略只需要判断下一个交易日的涨跌状况，故需要判断下一个交易日的出现涨跌的两个概率，举例如下：

设前五日涨跌情况为：涨涨涨跌跌，转换成01串为：11100，判断第六日的涨跌情况只需要比较111000和111001出现的概率。

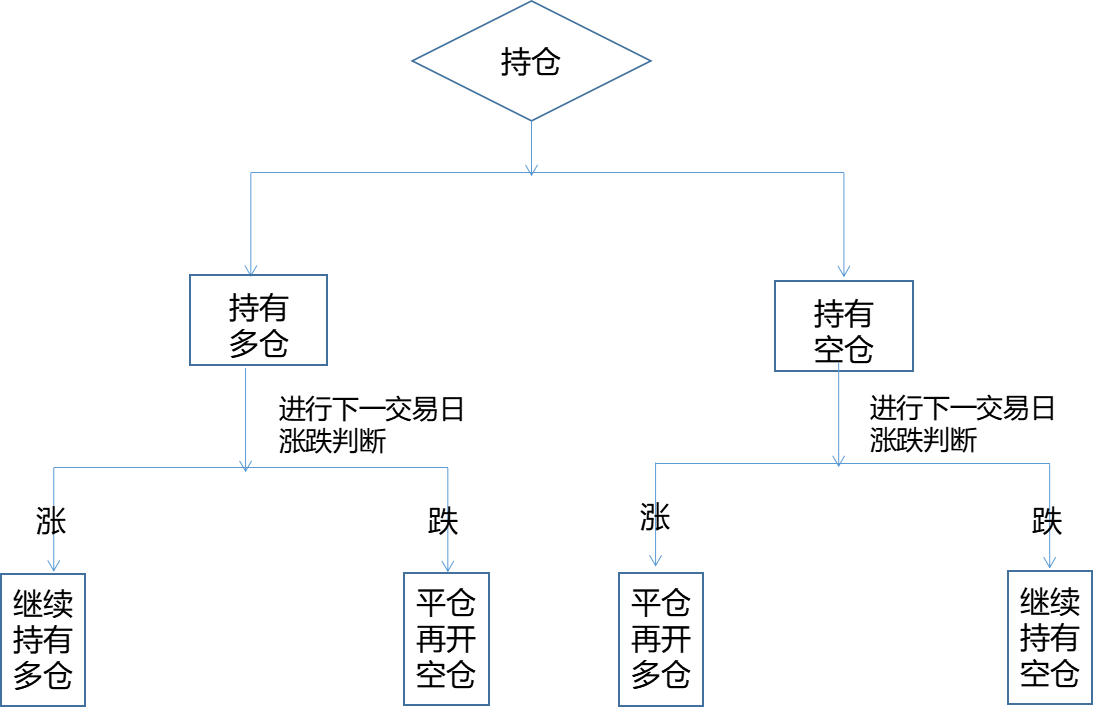
根据大数定律，频率近似为概率，故只需要统计语料库111000和111001分别出现的次数，若111000出现的次数多于111001出现的次数，则第六日涨跌情况为0，即会跌。

本文选取沪深300指数期货当月合约2014年4月16日到2016年12月31日的历史日线数据为样本内初始语料库，用来训练模型。同时不断扩充历史数据，依次扩充语料库。

##### **1.3 交易策略**

在交易日15:00时进行下一个交易日的涨跌预测，在股指期货市场进行相应的头寸开仓，如果下一个交易日判断方向与当前持仓方向相同，那么不进行操作，如果持仓方向不同，那么先进行平仓，再开仓。

###### 图1：判断交易流程图



##### **1.4 模型阶数**

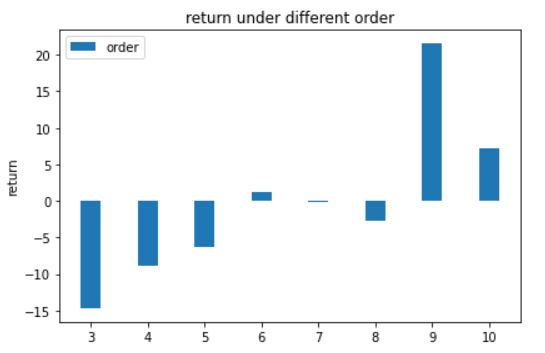
在进行下一交易判断时，考虑下一交易日受前几个交易日影响对模型的结果产生重要影响，故选择3-10个交易日分别进行样本内计算，确定原则以2016年1月1日到2018年1月1日的收益率最大化。

###### 

###### 表1：3-10个交易日下的收益

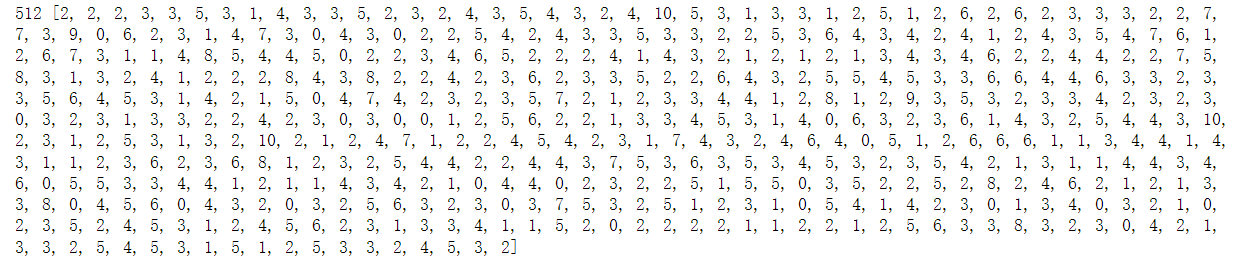
| 阶数 | **3** | 4 | 5 | 6 |
| --- | --- | --- | --- | --- |
| **收益** | -14.71% | -8.82% | -6.33% | 1.24% |
| **阶数** | **7** | **8** | **9** | **10** |
| **收益** | -0.19% | -2.76% | 21.58% | 7.29% |

###### 图2：3-10个交易日下的收益



故根据上述原则，选择阶数为9，即考虑9个交易日进行下一交易日的涨跌情况判断。

###### 图3：9阶模型下的样本出现次数



### 二、实证分析

**1.实证说明**

（1）数据选取，本实证选取沪深300指数期货当月合约2014年4月16日到2016年12月31日的历史日线数据为样本内初始语料库，2017年至2022年4月29日的历史日线数据为测试集。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。

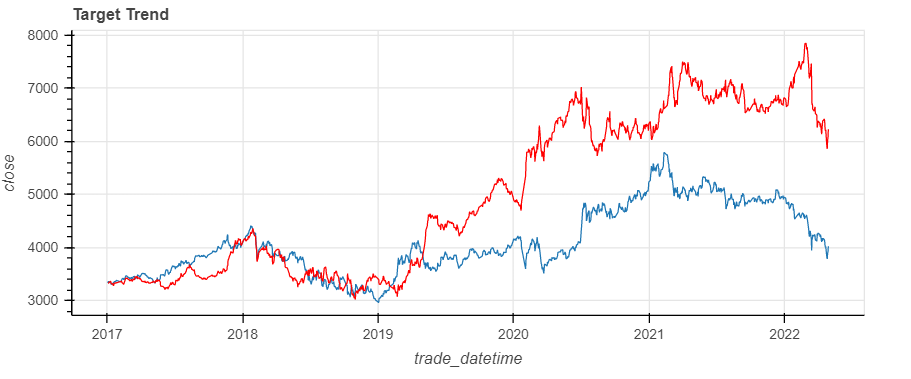
（2）交易假设：单边手续费为万分之一。

（3）回测结果：

累计收益 87.07%
  
多仓次数 317
  
多仓成功次数 173
  
多仓胜率 54.57%
  
多仓平均持有期 1.25
  
空仓次数 318
  
空仓成功次数 165
  
空仓胜率 51.89%
  
空仓平均持有期 0.79
  
周期胜率 52.52%
  
最大回撤 30.72%
  
年化收益/最大回撤 0.41
  
年化收益 12.49%
  
年化标准差 20.97%
  
年化夏普 0.60

1. 交易累计收益图示：

图4：指数期货与收益（红线为收益，蓝线为指数期货）



###### 

###### 图5：收益



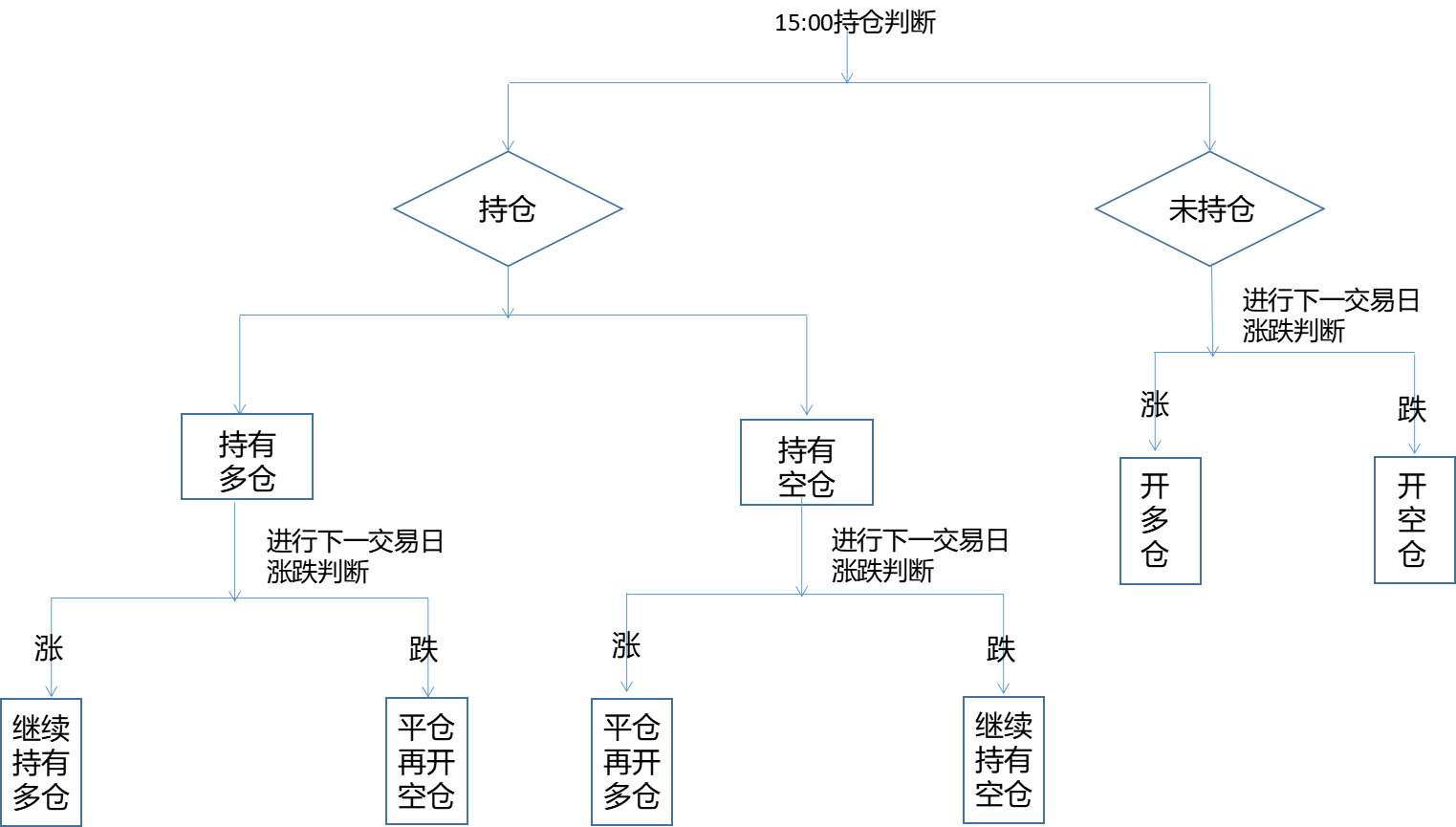
### 三、策略改进

#### **1.改进方案**

考虑止损情况，设置1%的止损幅度，在当日盘中波动较昨日收盘价边度幅度在不利方向超过或等于1%时，触发止损机制，强制平仓，否则继续持有到15点，再依据之前策略进行次日涨跌判断。

由于考虑到了日内平仓的可能性，故在15点时会有持有多仓，持有空仓，未持仓三种情况的出现。本文采用了10min的分钟数据进行日内波动止损机制的判断。

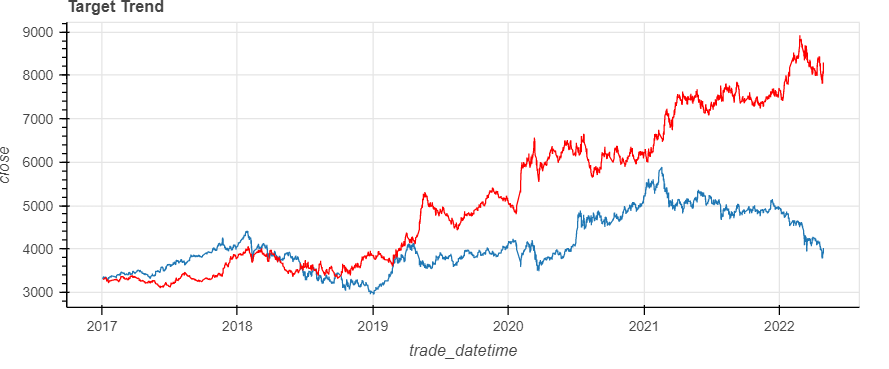
###### 图6：改进的判断交易流程图



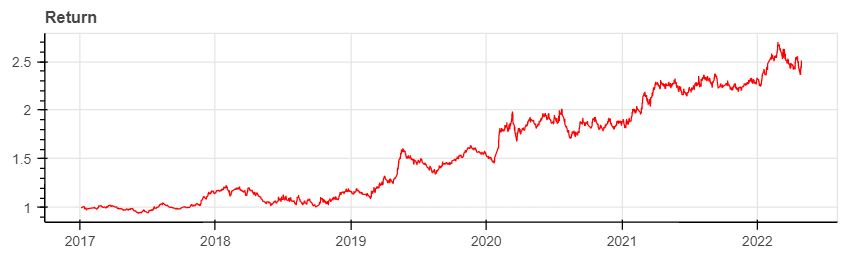
#### **回测结果：**

累计收益 151.08%
  
多仓次数 442
  
多仓成功次数 203
  
多仓胜率 45.93%
  
多仓平均持有期 30.77
  
空仓次数 406
  
空仓成功次数 174
  
空仓胜率 42.86%
  
空仓平均持有期 24.74
  
周期胜率 49.79%
  
最大回撤 18.19%
  
年化收益/最大回撤 1.04
  
年化收益 18.89%
  
年化标准差 16.90%
  
年化夏普 1.12

###### 图7：指数期货与收益（红线为收益，蓝线为指数期货）



###### 图8：收益



### 四、总结

在设置止损幅度1%之后，策略表现明显好于未加止损机制，分析其原因，主要在于，若对下一交易日的预测错误，原策略只能在下一交易日的15点才有可能弥补错误造成的损失，而加入止损机制，若盘中已造成一定损失，可立即止损平仓。当面临预测错误且向不利的方向进行跳空开仓的时候，止损策略更加有效。

原研报采用SLM的策略对上证指数进行回测，以1995年到2004年为初始语料库，2005年到2009年为样本内来寻找最佳阶数，根据历史收益最大化的远侧，确定模型阶数为6。在未设置止损机制时，2005年到2013年年化收益率达到了80.3%，设置止损机制后年化收益达到了75%。猜测原因有：SLM对上证指数效果更好，2005年-2013年指数走势整体较好，计算收益的时候考虑了其他研报没给出的参数或者条件。