

2014 年 03 月 20 日

金融工程

研究所

证券分析师: 王丹 S0350512060001
021-68591560 wangd01@ghzq.com.cn联系人: 黄韡力 S0350113090002
021-68591558 huangwl@ghzq.com.cn

CTA 交易策略系列之一:

基于贝叶斯网络概率推断的交易策略

报告要点:

- **贝叶斯网络:** 贝叶斯网络是机器学习的一个领域,是目前不确定知识表达和推理领域最有效的理论模型之一。通过对历史数据的整理和归纳,贝叶斯网络可以对未来的不确定性事件进行概率推断。利用贝叶斯网络对开盘价和收盘价的价差进行预测,从而将预测结果转化为投资决策。
- **测试结果:** 选取 2011.11.1-2013.12.31 为测试窗口,该策略的年化收益率为 98.22%,夏普比率为 3.7,最大回撤为 9.35%,预测准确率为 56.88%,盈亏比 1.83;如果增加 50 个 BP 冲击成本,策略的年化收益率下降为 66.81%,夏普比率为 1.86,最大回撤为 13.49%。
- **风险提示**
 - 1) 策略模型对其他资产不一定存在适用性;
 - 2) 策略的历史收益表现不代表未来收益表现。

目 录

1、 绪论	3
2、 贝叶斯网络简介	3
3、 策略模型	6
4、 策略回测结果和分析	8
5、 总结和未来展望	12

1、绪论

对于大部分金融资产来讲，其基本的价格决定机理在于供求关系的变化。但是，引导资产价格变化的供求关系变动的因素非常复杂，所以我们在市场上会看到一些出现定价错误的金融资产。通过研究资产供求关系的变化，可以在一定程度上为判断未来资产价格走势提供依据。历史价格和成交量等代表供需关系的数据，很多时候被用来作为判断资产供求关系的关键变量。

贝叶斯网络 (Bayesian Network)，也称为信度网络，目前不确定知识表达和推理领域最有效的理论模型之一。本文将探索将贝叶斯网络理论，结合相关变量，应用到交易当中。文中的策略模型不适用于有中断行情的资产，例如因停牌等原因造成交易中止的股票。

2、贝叶斯网络简介

2.1 贝叶斯网络结构

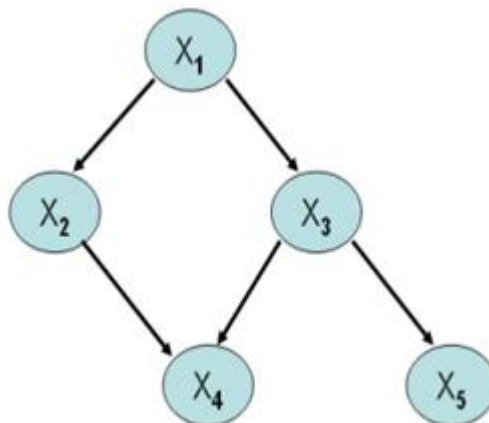
条件概率是指事件 A 在另外一个事件 B 已经发生条件下的发生概率。通常，事件 A 在事件 B 发生的条件下的概率，与事件 B 在事件 A 的条件下的概率是不一样的，然而，这两者是有确定的关系。贝叶斯定理是关于随机事件 X 和 Y 的条件概率和边缘概率：

$$P(B|A) = \frac{P(B)P(A|B)}{P(A)}$$

如果变量 X 的值依赖于变量 Y，那么变量 X 与变量 Y 的关系可以表示为： $Y \rightarrow X$ ，其中 Y 可以称为 X 的父节点，X 称为 Y 的子节点。如果 X 有多个父节点，那么这些父节点的集合可以表示为： $\text{Set}\{X \text{ 的父节点}\} = \{Y_1, Y_2, Y_3, \dots\}$ 。

贝叶斯网络和条件概率有着密不可分的关系。贝叶斯网络是一种概率网络，利用图形化的方式进行决策分析，是一种将因果关系和概率知识相结合的信息表示框架。贝叶斯网络是有向无环图(Directed Acyclic Graph, DAG)，由代表变量节点及连接这些节点有向边构成。构成一个完整的贝叶斯网络有两个要素：节点和有向边，其中，节点代表随机变量，节点间的有向边代表了节点间的互相关系，用条件概率进行表达关系强度，没有父节点的用先验概率进行信息表达。因此，对于具有不确定性的事件，我们可以依赖贝叶斯网络对事件做出有条件的推断。下图是一个简单的贝叶斯网络拓扑图例。

图 1 贝叶斯网络拓扑图例



2.2 贝叶斯网络构建与训练

对于一个未知的贝叶斯网络，通常有三种方法可以对其结构进行推断：（1）专家知识；（2）从数据集推断；（3）前面两种方法的结合。在许多研究中，构建贝叶斯网络首先由专家知识直接给出条件概率表（CPT），再根据 CPT 估计贝叶斯网络结构。但是，在很多情况下，对于特定贝叶斯网络结构，专家知识无法用于估算其中的条件概率。此时，如果有足够的数据集，可以网络参数学习的算法，根据给定的数据集，对网络结构进行概率推断。

目前，关于贝叶斯网络参数学习的方法有很多种。本文主要探讨 K2 算法在贝叶斯网络推断中的应用。K2 算法由 Gregory F.Cooper 和 Edward Herskovits 在 1991 年和 1992 年的两篇文章中首次提出，该算法的基本原理是通过对不同的贝叶斯网络结构进行打分，从而对贝叶斯网络结构进行选择 and 推断。在 K2 算法中，假设有一个节点没有父节点，如果该节点新增加的父节点可以增加该网络结构的评分，那么该新增节点则作为父节点加入网络中。换言之，在网络推断过程中，K2 按顺序逐一考察节点变量，确定父节点，然后添加相应的边和节点。当额外增加的父节点不能增加评分，则停止增加该节点的父节点。

评分函数定义如下：

$$g(i, \pi_i) = \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)!} \prod_{k=1}^{r_i} N_{ijk}!$$

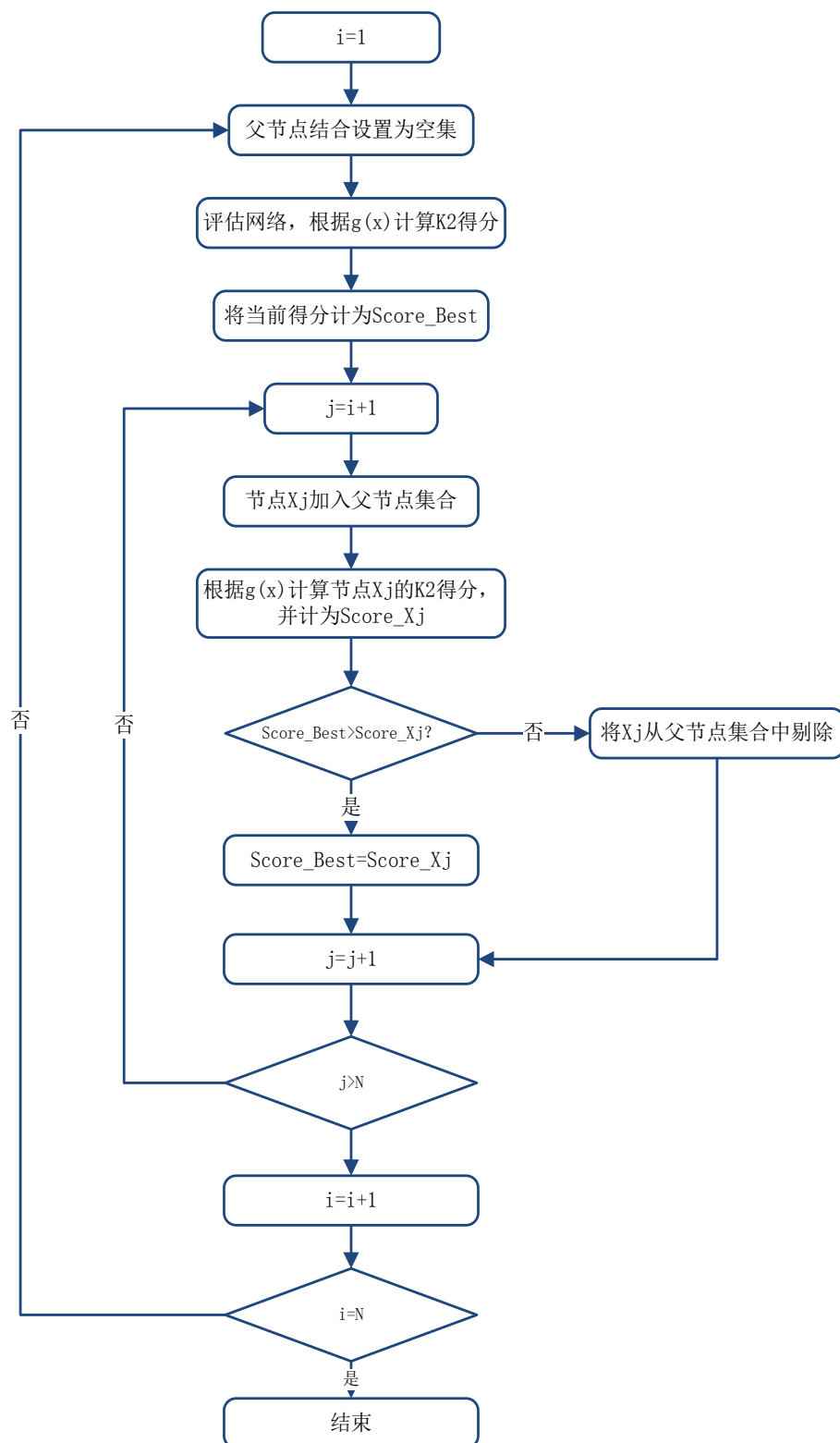
$$N_{ij} = \sum_{k=1}^{r_i} N_{ijk}$$

N_{ijk} ：样本数据中，当第 i 个节点（变量）的父节点为 j 的时候，该变量为 k 的数据样本数量；

r_i : 第 i 个节点 (变量) 拥有的变量值上限。

详细的 K2 算法计算流程如图 2 所示。

图 2 K2 算法计算流程



资料来源：国海证券研究所

3、策略模型

3.1 变量选择

本文的目的在于根据已有信息，预测未来 K 线的变化。选择的变量数量越多，回溯的期数越长，引入的历史信息就会越多。但是我们认为指标的数量与预测精度并不成正比，这是因为引入越多的变量指标，可能会增加越多的噪声信号，从而影响判断精度。

表 1 自定义指标库

序号	名称	描述
1	收盘价	/
2	成交量	/
3	委比	(委买数-委卖数)/(委买数+委卖数)
4	K 线变化幅度	收盘价-开盘价
5	开盘价	/

资料来源：国海证券研究所

除了上述 6 组指标，还可以引入其他的指标，如持仓量、最高价、最低价、KDJ、MACD 等。在完成指标选择之后，我们需要对指标进行处理，通过函数将其转化为状态变量。

出于简化模型的目的，在本文中，我们使用的状态变量为二元变量。转换函数定义如下：

$$r_t = \begin{cases} 1, & x_t \leq x_{t-1} \\ 2, & x_{t-1} > x_t \end{cases}$$

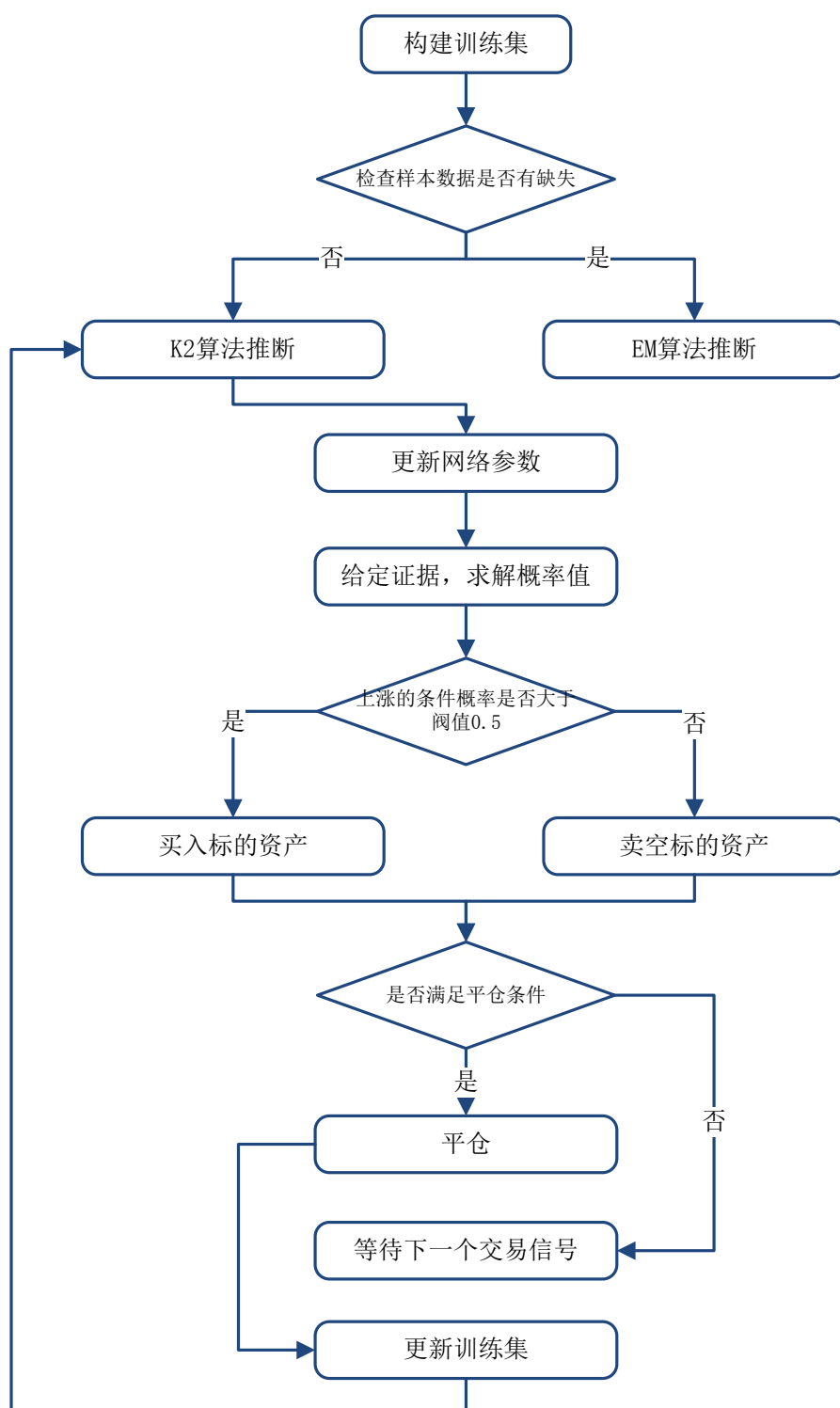
其中， r_t 为变量 x 在时间 t 和时间 $t-1$ 之间的变化率。通过上式，我们可以将所有变量转化为二元变量，且消除了单位的影响。

3.2 策略实施步骤

交易策略的基本原理是：首先利用已有的数据集对贝叶斯网络进行推断，然后将最新的变量数据代入网络中，对未来的 K 线走势进行预测，再根据预测的概率结果，产生相应的交易信息号。此外，为了充分利用历史经验，我们为贝叶

斯网络不断补充新知识，并实时更新结构。然而，必须要注意的是数据缺失会使 K2 算法失效。因此如果出现数据确实的情况，我们就需要使用替代算法，如最大期望（Expectation Maximization, EM）等，数据缺失的交易策略不在本篇报告论述的范围内，相关算法和策略将在后续报告中阐述。策略实施的详细步骤如图 3 所述。

图 3 策略实施步骤



资料来源：国海证券研究所

4、策略回测结果和分析

4.1 主要参数设定

在本文中，我们主要测试沪深 300 股指期货主力合约。图 4 是沪深 300 股指期货主力合约收盘价和开盘价的价差，即收盘价-收盘价。表 4 是自 2010 年 5 月 1 日起，截至 2014 年 1 月 31 日的价差统计数据。

我们策略的目的在于判断 K 线的收盘价和开盘价之间的相对位置，从而获取两个位置差之间的收益。从统计数据 and 图表可以看出，如果预测准确的话，在大部分情况下，该策略在扣除手续费和冲击成本后，都能稳定地获得收益。

图 4 沪深 300 股指期货主力合约收盘价和开盘价的价差

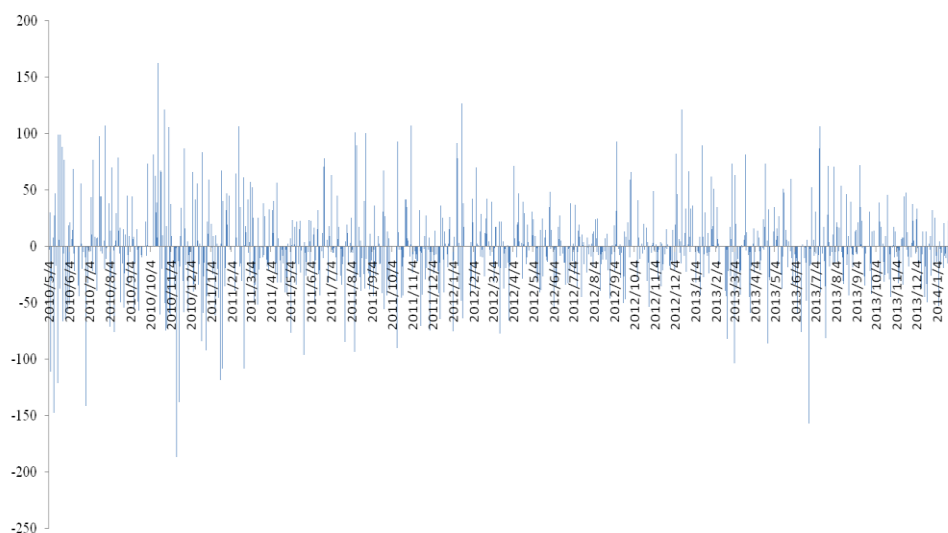


表 2 沪深 300 股指期货主力合约收盘价和开盘价价差分析

项目	次数
上涨次数	398
下跌次数	505
其他	6
最大价差距离	186.6
最小价差距离	0
平均价差距离	25.64
价差距离中位数	17

资料来源：国海证券研究所

其他测试参数如下表所示。

表 3 初始参数

参数	数值
初始训练时间	2010/5/1 – 2011/10/31
样本外测试时间	2011/11/01 – 2014/01/31
交易成本	单边 0.5%
冲击成本	单边 0.1%
保证金比例	12%
交易频率	每天一次
初始资金规模	100 万

资料来源：国海证券研究所

4.2 策略模拟结果

假如当前合约价格为 2700 元，则最多可以买 10 手合约（ $1000000 / (2700 * 300 * 12\%)$ ）；假如当前合约价格为 2200 元，则最多可以买 12 手合约。通常情况下，我们需要预留交易资金一倍的资金作为补充保证金，因此实际可开仓合约数量在 5-6 张。在测试中，我们每次总是开 5 手合约，并且当损失超过当日开仓合约金额所占用保证金的 5% 时，进行强制平仓。

策略模拟的资产净值走势图如图 5 所示。

图 5 资产净值走势

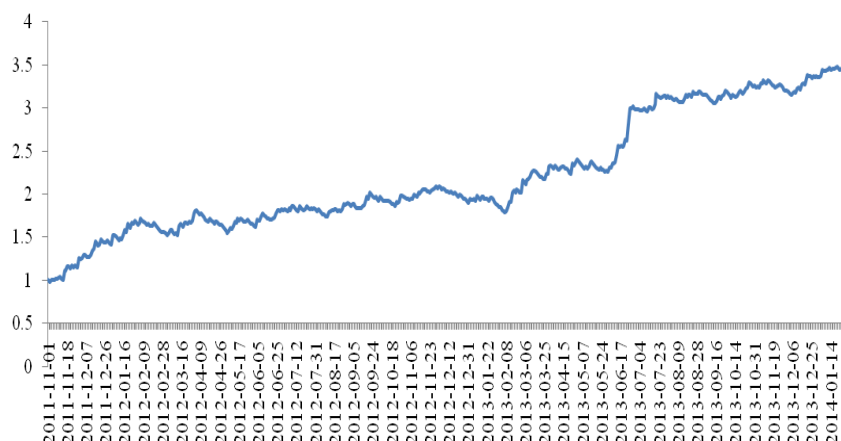


图6 模拟策略月度收益率

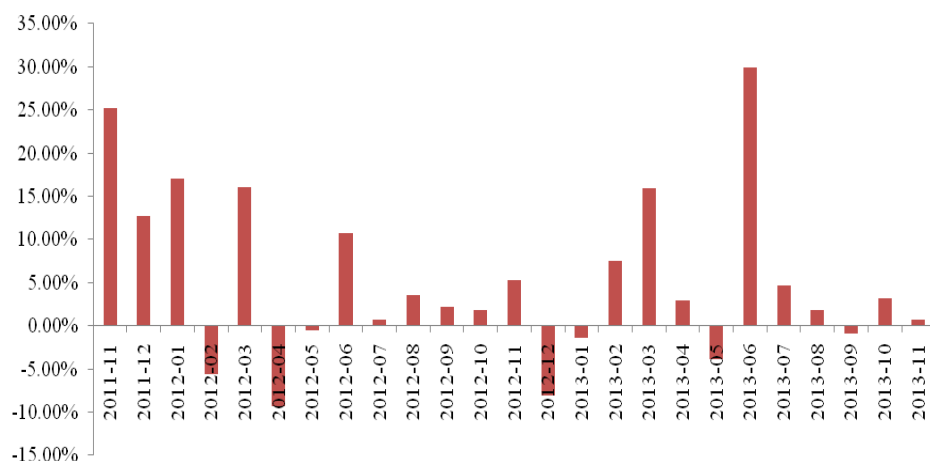


图7 模拟策略年度收益率

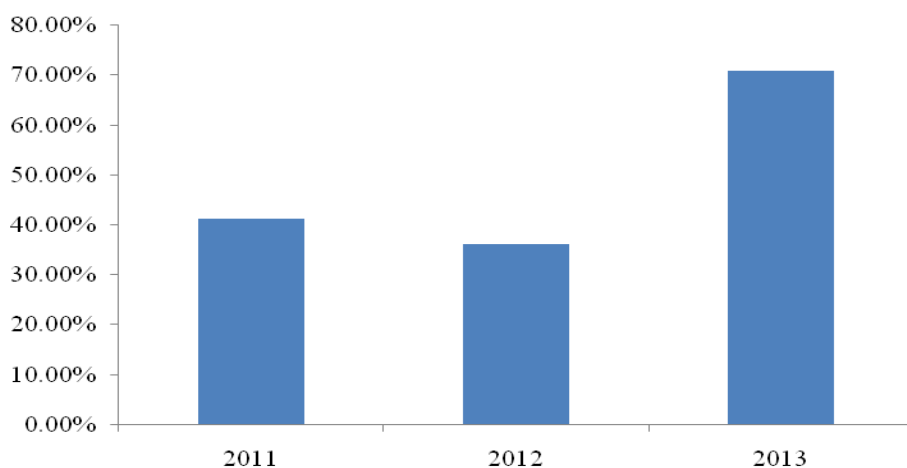


表4 模拟策略回测统计结果

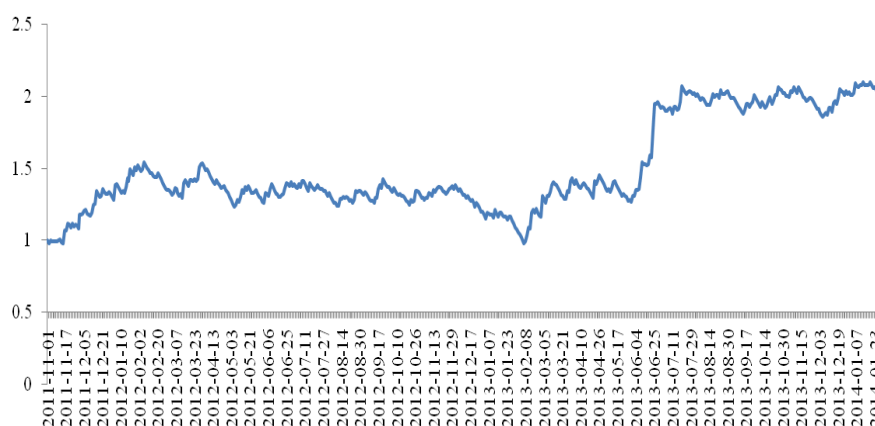
参数	测算结果
年化收益率	98.22%
年化波动率	25.57%
最大回撤	9.35%
夏普比率	3.704
预测准确度	56.88%
日交易胜率	45.32%
月度交易胜率	72%
年度交易胜率	100%
盈亏比	1.83

资料来源：国海证券研究所

日交易胜率低于价格变动方向预测准确度的原因主要是部分交易日的点差过小，获利被交易成本全部侵蚀掉，所以导致交易胜率低于价格变动方向预测的准确度。总体来看，这是一个正期望的策略，具有一定的实用性。然而，当将止损条件放松至 10% 时，策略的盈利能力将急剧下降，因此，该策略模型需要通过强止损来保证一个比较稳定的盈利能力。

策略模拟的资产净值走势图如图 8 所示。

图 8 资产净值走势



资料来源：国海证券研究所

从图 8 可以看出，当增加冲击成本以后，许多微小的获利空间被冲击成本严重侵蚀。

图 9 模拟策略月度收益率

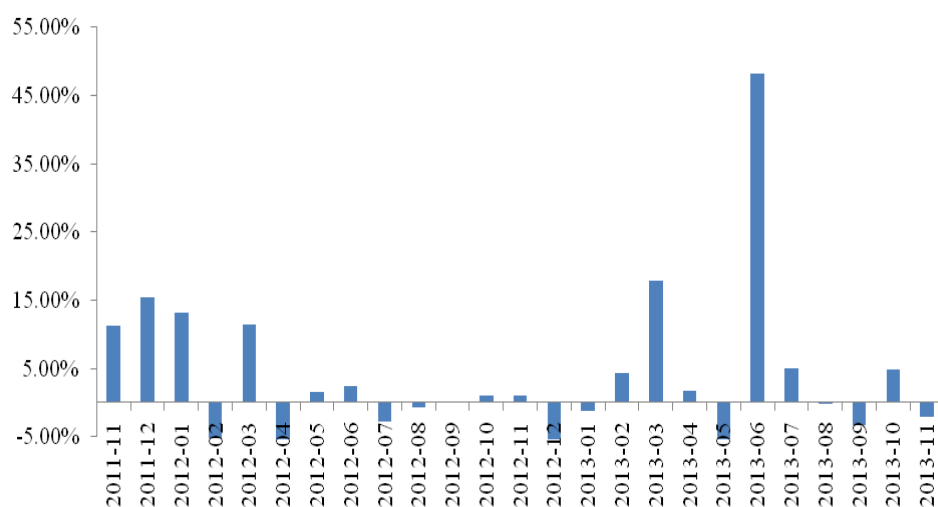


图 10 模拟策略年度收益率

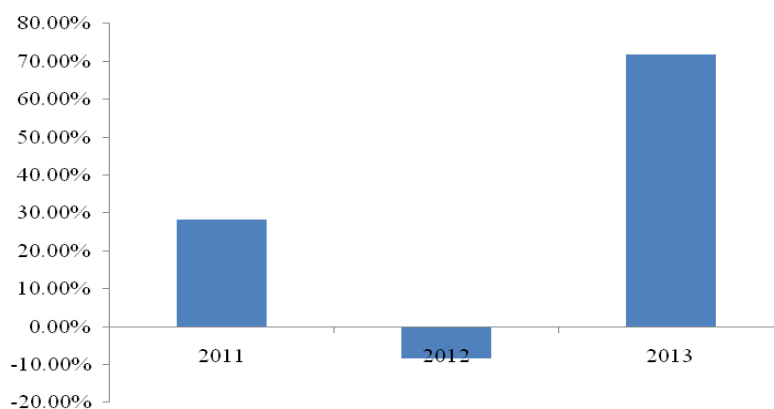


表 5 模拟策略回测统计结果

参数	测算结果
年化收益率	66.81%
年化波动率	34.11%
最大回撤	13.41%
夏普比率	1.87
日交易胜率	39.08%
月度交易胜率	60%
年度交易胜率	66.67%
盈亏比	1.8

资料来源：国海证券研究所

5、总结和未来展望

贝叶斯网络是机器学习中的一种方法。本文利用贝叶斯网络整合历史信息，对未来资产价格走势进行预测，并构建相应的交易策略。历史回测结果表明，贝叶斯网络如果设计得当，是可以作为交易策略应用在实际操作中的。

在未来，可以从以下几方面进行研究：（1）变量刻画。对变量进行更精确的划分，例如区分不同涨跌幅的状态变量；（2）适用于有停牌事件的股票的交易策略和相应算法；（3）引入其他指标，尝试在更高交易频率上的应用。

【分析师承诺】

王丹，本人具有中国证券业协会授予的证券投资咨询执业资格并注册为证券分析师，以勤勉的职业态度，独立、客观地出具本报告。本报告清晰准确地反映了本人的研究观点。本人不曾因，不因，也将不会因本报告中的具体推荐意见或观点而直接或间接收到任何形式的补偿。

【国海证券投资评级标准】

行业投资评级

推荐：行业基本面向好，行业指数领先沪深 300 指数；

中性：行业基本面稳定，行业指数跟随沪深 300 指数；

回避：行业基本面向淡，行业指数落后沪深 300 指数。

股票投资评级

买入：相对沪深 300 指数涨幅 20%以上；

增持：相对沪深 300 指数涨幅介于 10%～20%之间；

中性：相对沪深 300 指数涨幅介于-10%～10%之间；

卖出：相对沪深 300 指数跌幅 10%以上。

【免责声明】

本公司具有中国证监会许可的证券投资咨询业务资格。本报告中的信息均来源于公开资料，本公司对这些信息的准确性和完整性不作任何保证。报告中的内容和意见仅供参考，并不构成对所述证券买卖的出价和征价。本公司及其公司员工对使用本报告及其内容所引发的任何直接或间接损失概不负责。本公司或关联机构可能会持有报告中所提到的公司所发行的证券头寸并进行交易，还可能为这些公司提供或争取提供投资银行、财务顾问或者金融产品等服务。

【风险提示】

市场有风险，投资需谨慎。投资者不应将本报告为作出投资决策的惟一参考因素，亦不应认为本报告可以取代自己的判断。在决定投资前，如有需要，投资者务必向本公司或其他专业人士咨询并谨慎决策。在任何情况下，本报告中的信息或所表述的意见均不构成对任何人的投资建议。投资者务必注意，其据此做出的任何投资决策与本公司、本公司员工或者关联机构无关。

若本公司以外的其他机构（以下简称“该机构”）发送本报告，则由该机构独自为此发送行为负责。通过此途径获得本报告的投资者应自行联系该机构以要求获悉更详细信息。本报告不构成本公司向该机构之客户提供的投资建议，本公司、本公司员工或者关联机构亦不为该机构之客户因使用本报告或报告所载内容引起的任何损失承担任何责任。

【郑重声明】

本报告版权归国海证券所有。未经本公司的明确书面特别授权或协议约定，除法律规定的情况外，任何人不得对本报告的任何内容进行发布、复制、编辑、改编、转载、播放、展示或以其他方式非法使用本报告的部分或者全部内容，否则均构成对本公司版权的侵害，本公司有权依法追究其法律责任。

机构销售部人员联系方式

北京区：北京市海淀区西直门外大街 168 号腾达大厦 1509 室

姓 名	固 话	手 机	邮 箱
白 丹	010-88576898-803	18801090832	baid@ghzq.com.cn
顾科嘉	010-88576833-810	13466400185	gukj@ghzq.com.cn

上海区：上海市浦东新区世纪大道 787 号 4 楼

姓 名	固 话	手 机	邮 箱
陈 蓉	021-20281099	13801605631	chenr@ghzq.com.cn
胡 超	021-20281096	13761102952	huc@ghzq.com.cn
梁柳欣	021-20281099	13524364843	lianglx@ghzq.com.cn
刘云志	021-68591565	13918449204	liuyz01@ghzq.com.cn
颜 彦	0771-5539092	15978141626	yany@ghzq.com.cn

广深区：深圳市福田区竹子林四路光大银行大厦 3 楼

姓 名	固 话	手 机	邮 箱
徐伟平	0755-83705955	15012898363	xuwp@ghzq.com.cn
马 鑫	0755-83702072	18899776696	max01@ghzq.com.cn
周振宁	0755-82835810	15220268966	zhouzn@ghzq.com.cn
李 慧	0755-83704850	13823570518	lih03@ghzq.com.cn
张玉霞	0755-83707413	18520881700	zhangyx02@ghzq.com.cn
李 芬	0755-83700919	13570866113	lif@ghzq.com.cn