

# 基于自适应网络模糊推理系统的择时研究

## ——拨开云雾见月明

罗军 金融工程 分析师

电话: 020-87555888-655

eMail: lj33@gf.com.cn

SAC执业证书编号: S0260511010004

胡海涛 金融工程 分析师

电话: 020-87555888-406

eMail: hht@gf.com.cn

SAC执业证书编号: S0260511020010

### 模糊推理机制

传统逻辑推理的前提条件及结论都是通过具有明确边界的集合来判断和定义的, 例如“如果某人 A 是高个子, 那么其子女也是高个子”这样一条推理, 传统方式可定义“高个子”集合为身高大于等于 1.8 米, 在此意义下, 一个人是不是高个子或者说是否属于“高个子”集合是非常明确的, 是或不是必选其一, 不存在模棱两可的情况, 然而这并不能表达人类语言意义上的高个子的含义, 因为 1.7999 米将不被认为是高个子。应用基于模糊集合的模糊推理机制这一问题将迎刃而解, 在模糊集合概念下某人对“高个子”集合存在一个隶属度 (如 0.8), 隶属度也可以看成是前提条件的满足程度。因此模糊推理机制对于解决带有不确定性的问题具有独到之处, 可以应用于股票价格波动择时研究中。

### ANFIS 模型及实证效果

ANFIS 模型对每一个输入变量分别赋予多个模糊集合并计算其满足程度 (隶属度), 形成多条件模糊规则并通过 T 范式算子 (如代数积算子) 计算模糊规则的激励强度, 再采用函数式推理方式计算每一条规则的推理结果并由激励强度加权得到最后输出。

本文采用滚动预测方式, 对每一组训练样本, 分别优化由不同隶属函数、不同迭代参数改动量 (Step Size) 组成的共计 21 个 ANFIS 的模型库, 取训练检验样本均方误差最小的 9 个来代入新的输入变量值, 取 9 个模型输出的均值为最后输出。

以 1996 年 1 月 5 日至 2002 年 12 月 27 日 347 个周线数据为训练样本, 2003 年 1 月 3 日至 2011 年 6 月 24 日 428 个数据为检验样本。由 12 个行情或技术指标+5 个宏观变量指标构成 17 个基础输入, 再经过滞后期算子得到 51 个衍生变量, 总共 68 个变量, 拟合目标按涨跌分两类记为 1 或-1, 分别取两输入和三输入进行了 52394 次不同输入组合的寻优计算, 给出了 5 个最优的输入组合, 其中样本外预测准确率最高者为“RSI、收盘滞后 2 期、成交额滞后 3 期”, 准确率达到 57.24%, 考虑单边 0.5% 手续费后模拟收益率表现超过指数, 年化信息比率为 1.02, 无交易费下年化信息比率为 5.80。

### 自定义优化目标之 ANFIS 模型改进

对 ANFIS 模型预测输出按正负号记为涨或跌并与实际涨跌比较计算准确率, 应用遗传算法优化模型参数最大化准确率目标函数, 得到的改进模型 A 在同样的样本外区间预测准确率达到 57.94%, 略高于最佳输入情形的 57.24%, 但考虑手续费后的模型收益表现略不及指数且较最佳输入情形逊色。

改进模型 B 以经交易次数调整后的择时准确率为最大化目标函数, 改进模型 B 交易次数由改进模型 A 的 121 次下降至 119 次, 准确率达到 60.51%, 模拟收益表现强于指数且强于最佳输入情形, 无交易费及有交易费下年化信息比率分别为 6.13、2.24。

## 目录索引

一、人工智能与量化投资 .....	4
(一) 人工智能之量化投资回顾 .....	4
(二) 自适应网络模糊推理系统 (ANFIS) .....	5
二、模糊逻辑与模糊推理系统 .....	5
(一) 模糊集与隶属函数 .....	5
(二) 模糊规则与模糊推理系统 .....	7
三、自适应网络模糊推理模型 .....	8
(一) ANFSI 模型结构 .....	8
(二) ANFIS 择时实证分析 .....	10
四、自定义优化目标的 ANFIS 模型 .....	15
(一) 两个改进的 ANFIS 择时模型 .....	15
(二) 遗传算法 .....	21
五、总结 .....	23
(一) 研究意义和创新点 .....	23
(二) 模型的不足 .....	23
(三) 后续研究方向 .....	23

## 图表索引

图 1: 年轻、中年、年老模糊集合 .....	5
图 2: 三角高斯钟形隶属函数例图 .....	6
图 3: 模糊推理系统 .....	7
图 4: 三类模糊推理系统 .....	8
图 5: 两输入单输出的一阶 Sugeno 推理 ANFIS 模型 .....	9
图 6: 不同两、三输入组合下模型误差 .....	11
图 7: ANFIS 模型择时流程图 .....	12
图 8: 最佳输入变量下样本外滚动择时结果 .....	13
图 9: 五种不同输入下模拟交易累计收益率 (无交易费用) .....	13
图 10: 五种不同输入下模拟交易累计收益率 (0.5%单边交易费用) .....	13
图 11: 最佳输入变量下模拟交易净值相对指数变化 .....	14
图 12: ANFIS 样本外预测涨跌准确率比较图 .....	15
图 13: 两种自定义优化目标函数下的 ANFIS 择时模型 .....	16
图 14: ANFIS 模型改进 A 样本外预测结果 .....	17
图 15: 改进模型 A 与最佳输入模型净值比较 (考虑手续费) .....	18
图 16: 改进模型 A 与最佳输入模型净值之比 (考虑手续费) .....	18
图 17: 改进模型 B 样本外预测结果 .....	20
图 18: 改进模型 B 与最佳输入模型净值比较 (考虑手续费) .....	20
图 19: 改进模型 B 与最佳输入模型净值之比 (考虑手续费) .....	21
表 1: 基础输入变量 .....	10

表 2: 训练误差较小的前五名输入变量组合.....	11
表 3: ANFIS 模型样本外择时准确率.....	12
表 4: 不同输入变量组合下模拟交易信息比率.....	12
表 5: ANFIS 样本外预测涨跌结构.....	14
表 6: 最佳输入与改进模型 B 模拟交易信息比率比较.....	19

## 一、人工智能与量化投资

### （一）人工智能之量化投资回顾

众所周知，股票投资分析方法分为两种，一是基本面分析法；二是技术分析法，自股票交易现象出现以来，始终是两种分析方法并存之局面。而具体从事投资工作的人，其中部分信赖基本面分析，部分将技术分析奉若神明，近百年以来，两种投资理念孰优孰劣之辩屡见不鲜，经常见到的是，基本面分析人士认为技术分析无异于算卦，他们认为左右价格波动的是标的物背后的宏观形势、行业背景以及供需关系的变化等因素；技术分析人士则批评基本面分析方法，称研判宏观形势等基本面因素异常困难，其结果一方面需要相当长时间来验证，另一方面预测准确率也乏善陈可，且价格之短期波动受非常多的噪音因素影响，难以用基本面分析法研判。在此我们无意评价两种投资分析方法孰优孰劣，既然两者皆流行于世必然有其存在的道理，而且近二十年来，技术分析人士引入人工智能方法，将复杂的数学模型应用于分析价格的波动方向上，取得了巨大的成功（例如大家耳熟能详的大奖章基金掌门人西蒙斯）。

技术分析始于十九世纪初查尔斯·道创立的道氏理论，道氏将价格波动趋势分为长期趋势、中期趋势、短期趋势三种，并认为价格反映一切市场行为，价格随趋势变化，历史会不断重演等。

提及技术分析不得不说到波浪理论，艾略特受潮汐起落变化的自然现象启发，认为价格波动遵循着5浪推动向上、3浪调整向下的运行方式，由于波浪理论在研判价格波动方向及波动幅度方面有独到之处因而广泛流行。

另外，技术分析中形态分析也是常见的一种分析方法，例如双顶底、头肩顶底、三角形、奔走形、缺口、趋势线等等，以及随后大量技术指标的出现，如：MACD、RSI、KDJ等等。

但是技术分析方法之弊病也是显而易见，例如，随着年月的推进，历史数据逐渐形成海量级别，而这些数据每一个交易日仍在增多，特别是日内高频数据更甚，手工方式逐个研读每个交易品种的历史走势图是不现实的，从这一角度来说引进人工智能模型进行程序化分析是大势所趋；另一方面，批评技术分析人士常说技术分析“讲起来无敌、用起来无力”，我们认为部分是因为技术分析中存在主观臆断的成分，例如背驰形态分析中，张三认为背驰但李四不一定认为背驰，如何确认背驰，如何触发买卖操作，以及整套操作体系的成功概率、盈利情况等等之分析研究都需要借用量化分析甚至人工智能工具。

人工智能（Artificial Intelligence, AI）综合了数学、计算机科学、信息论、控制论等等领域的知识，研究领域包括机器学习、语音识别、机器翻译、图像识别、自然语言处理等等。

择时问题本质上是一种分类问题，即如何根据历史数据对下一个观察周期的价格波动方向或者幅度进行分类。国内常见之于研究报告的人工智能模型包括ANN、SVM、HMM等。人工神经网络（ANN）模型以历史数据为输入，以对应下一个观察周期涨跌的分类结果为输出，不断调整模型内部参数达到最小化模型输出与实际输出之间的差别，再引入新的输入达到预测的目的。支持向量机（SVM）模型通过核函数将低维空间的输入变量映射到高维空间，分析其在高维空间的类别聚集特征，通过对比实际输出调整核函数参数进而达到预测分类的目的。隐马尔科夫模型（HMM）在描述推动价格波动背后的隐藏因素的能力上面颇为突出，通过观察输入变量来刻画隐含因素的动态变化过程，进而达到预测分类的目的。上述模型已成功用于A股择时，在此不再赘述。

## （二）自适应网络模糊推理系统（ANFIS）

ANFIS，英文全称为Adaptive-Network-Based Fuzzy Inference System，以模糊推理系统为主体，通过建立多条模糊规则来描述专家知识与人类逻辑推理的过程，对不确定、不精确系统具体很好的处理能力。

传统逻辑是波尔逻辑，在逻辑推理中对所描述的事物都可以用（0，1）来分类标识，例如，如果X是偶数，则X可以被2整除，其中X要么偶数，要么非偶数，不存在模棱两可的情况。但是人类在对事物进行语言描述时并不是严格的波尔逻辑，非黑即白的分类划分是办不到的，例如，如果温度是非常冷，那么关掉空调，冷暖是人的感知，因人而异，不存在严格的划分标准，因此，对于这一类事物的推理传统逻辑是不合适的。

对于股票价格波动方向的研判上也有类似的情况，例如，如果T-5日MACD值高，并且T日MACD值低，那么未来股价将下跌，同样的高或低是一种语言描述，无法精确分类，建立在波尔逻辑基础上的传统推理是不适宜的。

基于此模糊推理应运而生，Roger Jang（1993）在模糊数学以及神经网络理论的基础上提出了ANFIS模型，并成功应用于印刷文字识别、汽车MPG（每加仑燃油行驶英里数）预测、混沌时间序列预测、自适应噪声消除等领域。

本报告将全面阐述模糊逻辑以及ANFIS模型，并在该模型的基础上加入输出结果分类层，以改进模型预测成功率，并利用遗传算法来优化改进模型之参数，对改进模型加以检验。

## 二、模糊逻辑与模糊推理系统

### （一）模糊集与隶属函数

经典集合有精确的边界，某一元素是否属于该集合不存在模棱两可的结果，而模糊集合没有明确的边界，单个元素从“属于该集合”到“不属于该集合”的转变是平滑的。

设  $x$  是  $X$  的任一元素， $X$  上的模糊集合  $A$  定义为一组有序对：

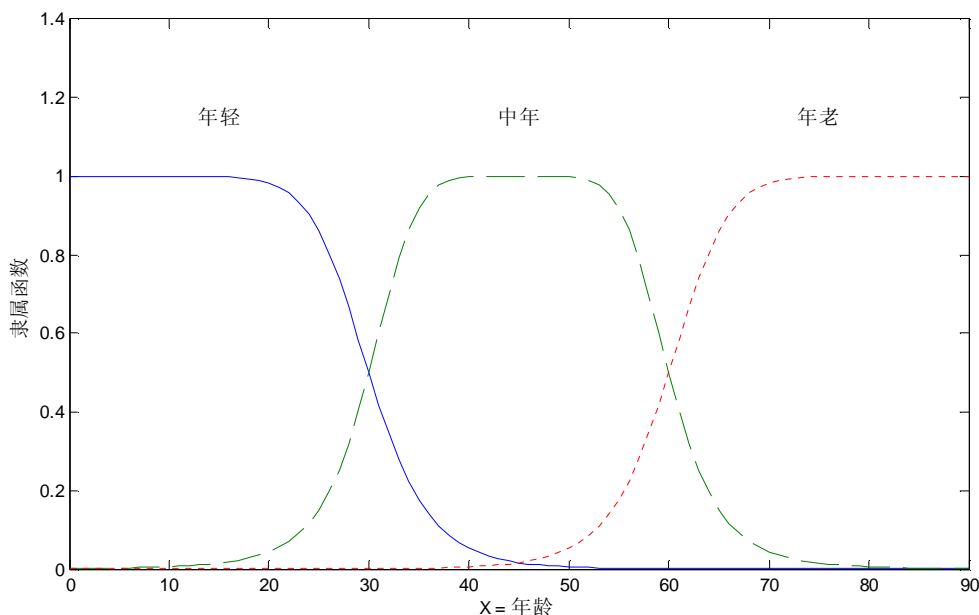
$$A = \{(x, m_A(x)) | x \in X\}$$

其中， $m_A(x)$  为模糊集合  $A$  的隶属函数，隶属函数将每个元素映射到0~1之间，显然模糊集合是传统集合的扩展，如果隶属函数取值仅为0或1，则模糊集合退化为传统集合。

例如对年龄的划分，可以分为年轻、中年、年老，传统集合可以将30、60岁作为界限，小于30为年轻，大于60岁为年老，之间为中年，按此标准29岁将被视为年轻，31岁被视为年中，而两者相差很小分类结果却完全不同，不尽合理。若采用模糊集合这一问题将迎刃而解，如图1，30岁对于“年轻”、“中年”两个模糊集来说隶属度均为0.5，29、31岁对于“年轻”模糊集的隶属度分别为0.5840和0.4188。通过隶属函数将某一元素转换为属于该模糊集合的隶属程度，从而达到对事物模糊分类，进而进行模糊推理。

图 1：年轻、中年、年老模糊集合





数据来源：广发证券发展研究中心

模糊集合完全由隶属函数定义，隶属函数的取值空间为 $[0, 1]$ ，常见的隶属函数有三角形、梯形、高斯型、广义高斯型、钟形等等，由于金融价格及其相关变量的概率分布基本属于高斯或类高斯型，因此我们在此仅介绍三种：三角形、高斯型、梯形。

(1) 三角形隶属函数

$$\text{triangle}(x, a, b, c) = \max\left(\min\left(\frac{x-a}{b-a}, \frac{c-x}{c-b}\right), 0\right), a < b < c$$

(2) 高斯型隶属函数

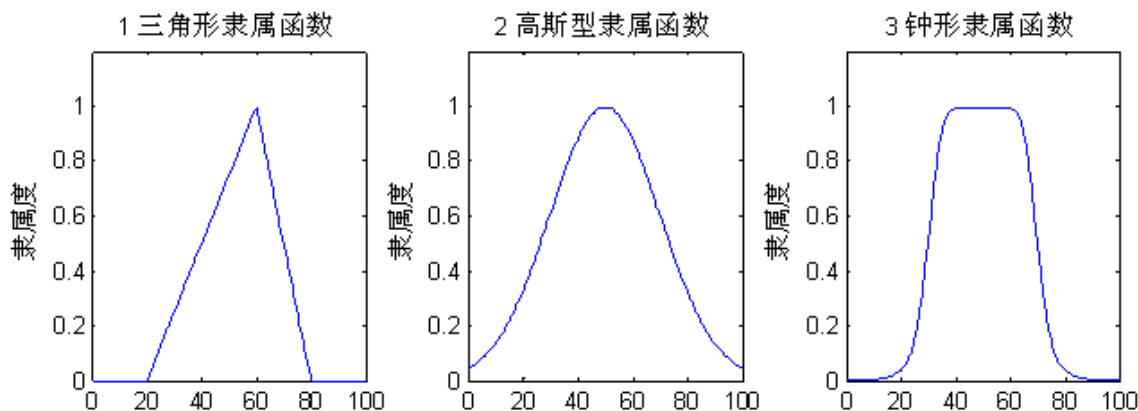
$$\text{gaussian}(x, c, s) = \exp\left(-\frac{1}{2}\left(\frac{x-c}{s}\right)^2\right)$$

(3) 钟形隶属函数

$$\text{bell}(x, a, b, c) = \frac{1}{1 + \left|\frac{x-c}{a}\right|^{2b}}, a \neq 0, b > 0$$

三类隶属函数的实例图如图2。

图 2：三角高斯钟形隶属函数例图



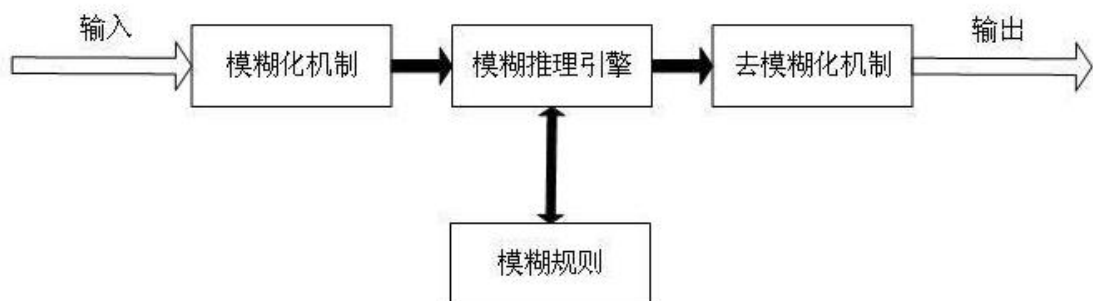
数据来源：广发证券发展研究中心

注：三角形（20，60，80）；高斯型（50，20）；钟形（20，40，50）。

## （二）模糊规则与模糊推理系统

模糊推理系统一般由四部分组成，分别是：模糊化机制、模糊规则、模糊推理引擎、去模糊化机制四部分组成，如图3。

图 3：模糊推理系统



数据来源：广发证券发展研究中心

模糊化机制将精确的变量取值通过隶属函数转换为隶属于该模糊集合的隶属程度，从而达到模糊化目的，前文我们已经介绍了隶属函数。

模糊规则，即模糊if-then规则，也称为模糊条件语句，它的形式为

If  $x$  是  $A$  then  $y$  是  $B$

其中  $A$  和  $B$  均为模糊集合，通常称“ $x$  是  $A$ ”为前提，“ $y$  是  $B$ ”为结论。实际人类语言表达中的例子如，如果路滑，则驾车危险；如果西红柿红了，则它熟了。实际应用中可以通过T范式算子将多个前提条件隶属度结合在一起，形成规则激励强度或称满足度，它表示模糊规则中前提条件部分的满足程度。T范式算子是一类满足有界性、单调性、交换性、结合性的二元函数，例如代数积算子、最小化算子，本文采用代数积算子。

模糊推理引擎利用模糊规则满足度或输出变量进行推理，推理结果可以为模糊的，也可以是精确的。若推理结果是模糊集，则再采用去模糊化机制得到最后的输出。

模糊推理系统主要有三种：语意式、函数式、Tsukamoto式，有两个输入变量的三类模糊推理系统的推理机制可参见图4。

### （1）语意式模糊推理系统（Mamdani）

语意式系统又称为Mamdani系统，系统输出为模糊集，需要去模糊化机制来得到精

确输出，可选（a）面积中心法、（b）面积等分法、（c）极大平均法、（d）极大最小法、（e）极大最大法五种方法其一来去模糊化，具体计算方法可参见相关文献，由于Mamdani系统需要去模糊化运算，对模型参数优化以及使用提出了更高要求。

### （2）函数式模糊推理系统（Sugeno）

两输入变量的Sugeno系统有两条模糊规则：

如果  $x$  是  $A_1$  且  $y$  是  $B_1$ ，那么  $z_1 = a_1x + b_1y + c_1$

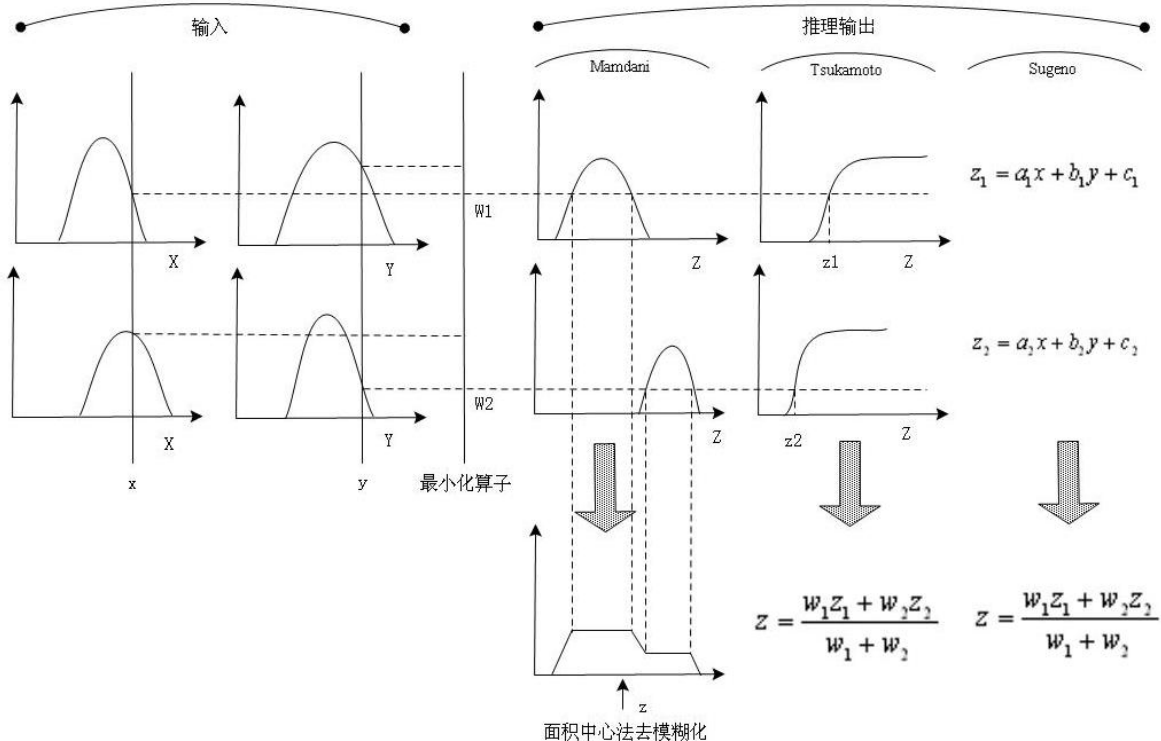
如果  $x$  是  $A_2$  且  $y$  是  $B_2$ ，那么  $z_2 = a_2x + b_2y + c_2$

两规则之输出再经过加权平均得到最后输出，加权重为两规则的激励强度。由于Sugeno系统输出结果为线性函数，因此无需去模糊化运算，不仅可以减小参数优化难度同时可以减少计算量，本文的ANFIS模型将采用Sugeno系统。

### （3）Tsukamoto模糊推理系统

Tsukamoto系统规则输出为激励强度的单调函数，然后再经过激励强度加权规则输出得到最后推理结果，在实际应用中不常用，不再赘述。

图 4：三类模糊推理系统



数据来源：广发证券发展研究中心

## 三、自适应网络模糊推理模型

### （一）ANFSI模型结构

ANFIS模型由Jang（1993）提出，我们以两个输入，每个输入有两个隶属函数，单个输出为例讲解模型，实际应用中输入变量个数以及每个变量的隶属函数可以有更多个，只要参数优化计算量在可承受范围内均可。两输入两隶属函数ANFIS模型结构图见图5。

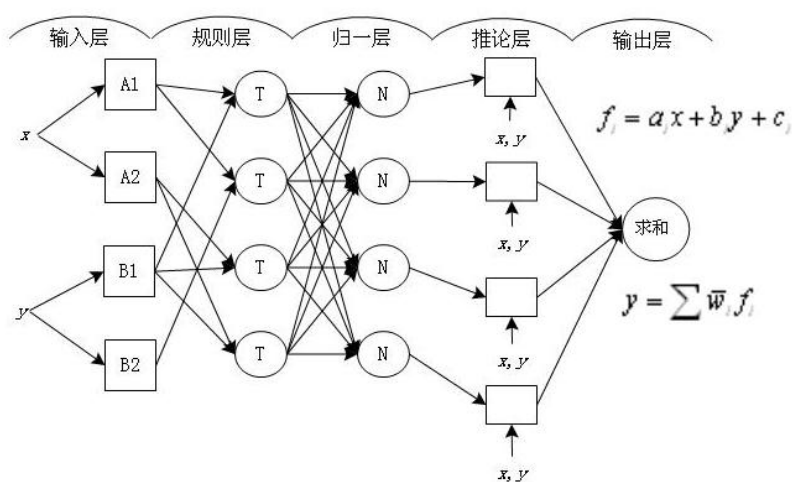


第一层：该层输出记为

$$O_{1,i,j} = m_{ij}(x_i), \quad i=1, \dots, N, \quad j=1, \dots, M$$

其中  $N$  为输入变量个数， $M$  为每个变量的隶属函数个数， $m_{ij}(x)$  为第  $i$  个输入的第  $j$  个隶属函数。由此可见，模型总共包含  $NM$  个隶属函数，假若隶属函数为高斯型，则每个隶属函数有2个参数，总共需要优化  $2NM$  个参数，若隶属函数为三角形或钟形，则需要  $3NM$  个参数，这部分参数称为前提参数。第一层的意义在于将精确的变量值转换为隶属不同模糊集的隶属度，正是所谓的模糊化操作。

图 5：两输入单输出的一阶 Sugeno 推理 ANFIS 模型



数据来源：广发证券发展研究中心

第二层：对每一个输入变量的模糊输出进行组合，得到  $M^N$  个模糊规则，采用代数积算子计算得到规则激励强度。

$$O_{2,p} = w_p = \prod_{i=1, j \in [1, \dots, M]} m_{ij}(x_i), \quad i=1, \dots, N$$

第三层：对第二层的规则激励强度归一化处理，

$$O_{3,p} = \bar{w}_p = \frac{w_p}{\sum_{i=1}^{M^N} w_p}, \quad p=1, \dots, M^N$$

第四层：根据Sugeno函数式推理方式计算每条推理结果，

$$O_{4,p} = f_p = \sum_{i=1}^N a_{pi} x_i + a_{p,N+1}$$

由此可见，单条推理有  $N+1$  个参数，总共有  $M^N$  条规则，参数总数为  $(N+1) \times M^N$  个，与第一层前提参数对应，称为结论参数。

第五层：综合第三层与第四层结果求和得到最后输出

$$O_5 = \sum_{p=1}^{M^N} \bar{w}_p f_p$$

ANFIS模型的参数优化以最小化模型输出与实际输出之间均方误差为目标，Jang（1993）提出了一种混合算法，对前提参数利用最速下降法优化，对结论参数由于其线性性质采用最小二乘法计算最佳参数。混合算法在给定初始参数值后迭代运算向最优解靠近，在此我们必须指出的是，迭代次数（epoch number）以及每次向负梯度方向靠拢的幅度大小（step size）对优化结果至关重要，若迭代次数太小则可能参数仍在收敛中，太大又需要太长时间来运算，且样本外拟合误差在超过一定迭代次数后反而会增大；step size设置的过大则可能跳过最优解，设置的过小则可能在迭代次数完成时仍进一步优化空间，因此如何设置step size和epoch number对模型优化及预测结果也很重要。

## （二）ANFIS择时实证分析

### （1）数据选取与输入变量的抉择

由于沪深300指数历史数据较少，本报告以上证指数为分析标的，同时由于A股交易初期涨跌停板制度不定，价格仅呈单边向上波动，同时1996年之前月度宏观经济数据的缺失，我们选择1996年1月5日至2011年6月24日共计775个周线样本数据。

一般而言，输入变量的选取对模型的预测效果影响颇大，这种想法再自然不过，找出最能表征股价涨跌变化之因素输入模型自然可以提高预测准确率。本文收集了12个行情或技术指标变量，5个宏观经济变量作为原始数据，参见表1，另外为了刻画变量的趋势变化特征，我们分别选取17个变量的滞后1期、滞后2期、滞后3期数据作为衍生变量待输入模型，总共可选取的输入变量为68个。

输出变量我们按涨跌分为两类并分别表示为1和-1。

最理想的输入变量选择方式是将所有可能的输入组合在一起，分别代入模型，并用训练样本来估计参数，用样本外数据来检验模型，得到最佳的输入变量，但是这样将耗费大量的时间，甚至是不可能的，退而求其次我们将不同输入代入模型，仅以训练样本优化参数，考察迭代1次后的样本内误差，选择误差最小者为最优输入，然后再进一步优化参数得到最终模型。

表 1：基础输入变量

变量名	变量说明	变量名	变量说明
周最高	指数周最高价	K	参数为 14 的随机指标 K 线
周最低	指数周最低价	D	参数为 14 的随机指标 D 线
收盘	指数周收盘价	RSI	参数为 14 的 RSI 指标
涨跌幅	指数周涨跌幅	M1	货币供应量 M1 同比增长
成交额(亿)	以亿为单位的周成交金额	M2	货币供应量 M2 同比增长
DIFF	MACD(26,12,9)DIFF	M1M2	M1-M2
DEA	MACD(26,12,9)DEA	CPI	居民消费价格指数
MACD	MACD(26,12,9)MACD	美元指数	美元指数
ema5	收盘价的 5 期 EMA 均线		

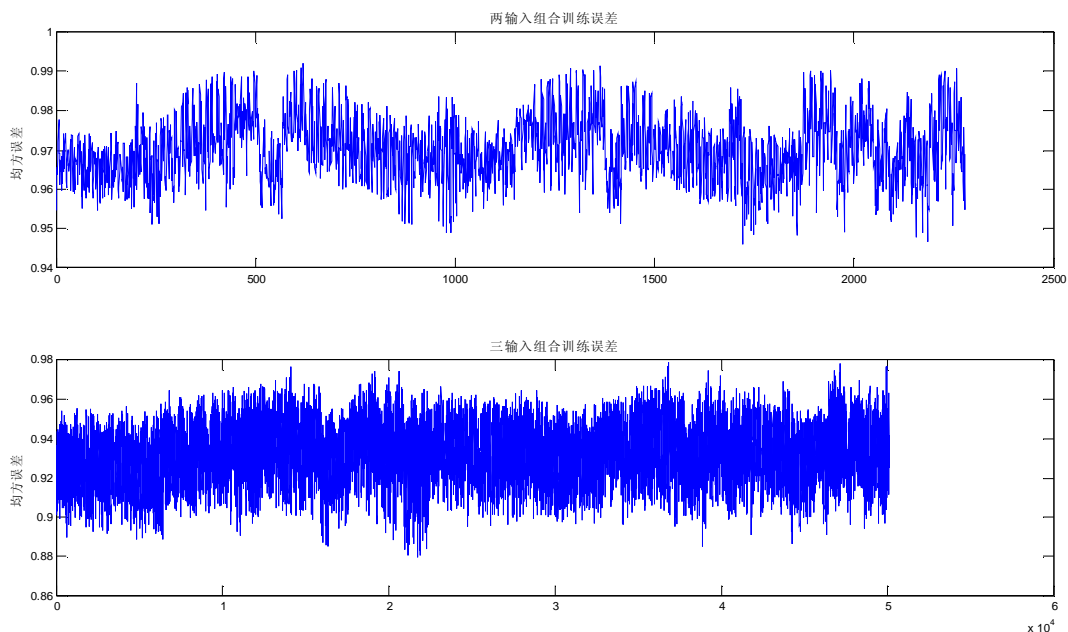
数据来源：广发证券发展研究中心

选择多少个变量输入模型也需要注意，输入变量数增多则模型总参数量会大大增加，这一方面需要更多样本训练模型，大大增加模型计算量耗费大量时间，另一方面模型参数的增多也会增加系统的不稳定性。具体的，若每个输入赋予2个隶属函数，高斯型隶属函数下2、3、4个输入时总参数量为20、44、96，三角形或钟形隶属函数下总参数量分别为24、50、104，鉴于此我们分别考虑输入变量数为2或3的情况。

在68个可能输入中选取两个可能的组合数为2278个，选择3个输入则有50116个可能

组合，分别赋以2个高斯型隶属函数计算初次迭代后误差，结果见图6。最佳输入抉择样本选取为1996年1月5日至2002年12月27日共计347个周线数据。

图 6：不同两、三输入组合下模型误差



数据来源：广发证券发展研究中心

从图6可以看出，三个输入变量情况下，模型训练误差较两输入情况有系统性下降，因此我们选择三输入来进行预报，同时上述实验训练误差最小的前5名输入组合见表2。

表 2：训练误差较小的前五名输入变量组合

序号	输入变量 1	输入变量 2	输入变量 3
1	RSI	M1M2 滞后 1 期	最高价滞后 2 期
2	RSI	收盘滞后 2 期	成交额滞后 3 期
3	RSI	最高滞后 1 期	成交额滞后 3 期
4	RSI	最高滞后 2 期	成交额滞后 3 期
5	RSI	最低滞后 2 期	成交额滞后 3 期

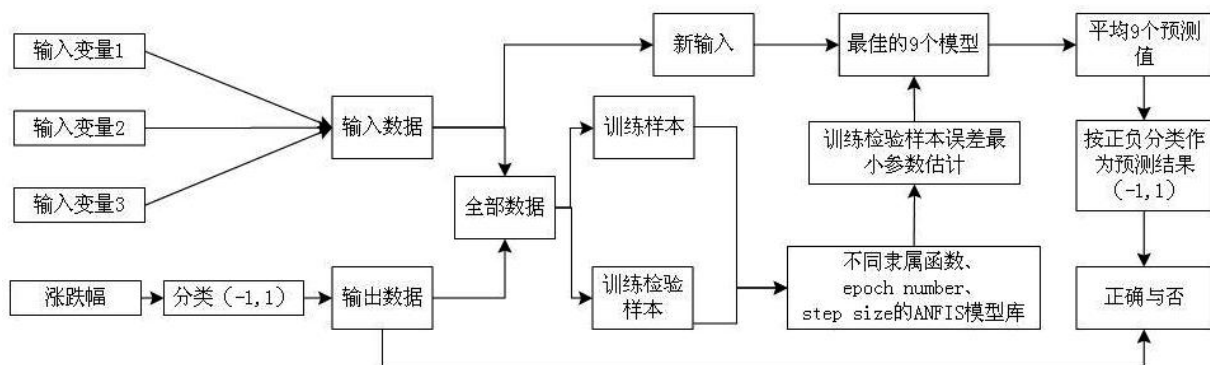
数据来源：广发证券发展研究中心

## （2）结果分析

选择了最佳输入变量以后，我们将数据代入如图7的预报模型，其中将参数优化样本分为两类，一是训练样本，根据前文提及的混合算法迭代优化参数；二是训练检验样本，每迭代优化参数一次，计算一次该样本下的误差，最后选取此误差最小的参数作为最终模型。我们总是选择最近的60个样本为训练检验样本，除此之外的所有样本作为训练样本。正如前文提及的隶属函数的选择、参数优化迭代次数、step size设置对模型优化结果是非常重要的，本报告选择三种备选隶属函数分别是三角形、高斯型、钟形，step size分别设置为0.01、0.05、0.1、0.2、0.3、0.4、0.5，迭代次数设置为50，根据我们的经验，大部分参数优化在30次迭代后训练检验样本误差就不再缩小了，因此为了节约计算时间，我们置为50。这样总共有21个ANFIS模型，训练完成后选择训练检验样本误差

最小的9个模型作为最佳模型组合，再将新的输入变量代入这9个模型，得到9个预测值，将其平均得到最后预测值，由于最后结果是浮点数值，我们根据其正负符号判定预测结果为涨或跌，并与真实价格涨跌进行比较得到预测准确率。

图 7: ANFIS 模型择时流程图



数据来源：广发证券发展研究中心

依此模型进行样本外检验，时间段为2003年1月3日至2011年6月24日，进行滚动择时预测，共计428个检验样本，于Intel Core2 2.4G CPU，2G内存的笔记本电脑中计算了大概8个小时，表3汇总了预测准确率结果，其中准确率最高的为（RSI、收盘价滞后2期，成交额滞后3期）输入时的57.42%，图8给出了整个样本外下的每一期择时结果（为图形简洁省略了其它4组输入变量情况）。

考虑择时策略的模拟交易收益情况，当预测下一期策略为上涨时保持仓位为100%，预测下一期策略为下跌时仓位下降至0%，即上涨满仓下跌空仓，交易费用方面分别考虑单边0.5%和0%两种情况，并假设期初净值为1，图9、图10分别给出了无交易费、有交易费两种情况下五种不同输入变量下的模拟交易结果。

表4给出了五种不同输入变量组合下，分有交易费和无交易费两种情况的模拟交易信息比率（模拟收益较指数基准的超额收益率的均值/超额收益率的标准差 $\times\sqrt{52}$ ），其中第二个输入变量组合表现最好，无交易费下信息比率为5.80，有交易费下信息比率为1.02。

图11给出了最佳输入变量（即第二个输入组合：RSI、收盘价滞后2期、成交额滞后2期）下模拟收益相对于指数涨跌幅下净值比例。

表 3: ANFIS 模型样本外择时准确率

输入变量	择时总次数	正确次数	准确率
RSI M1M2 滞后 1 期 最高滞后 2 期	428	213	49.77%
RSI 收盘价滞后 2 期 成交额滞后 3 期	428	245	57.24%
RSI 最高滞后 1 期 成交额滞后 3 期	428	234	54.67%
RSI 最高滞后 2 期 成交额滞后 3 期	428	240	56.07%
RSI 最低滞后 2 期 成交额滞后 3 期	428	238	55.61%

数据来源：广发证券发展研究中心

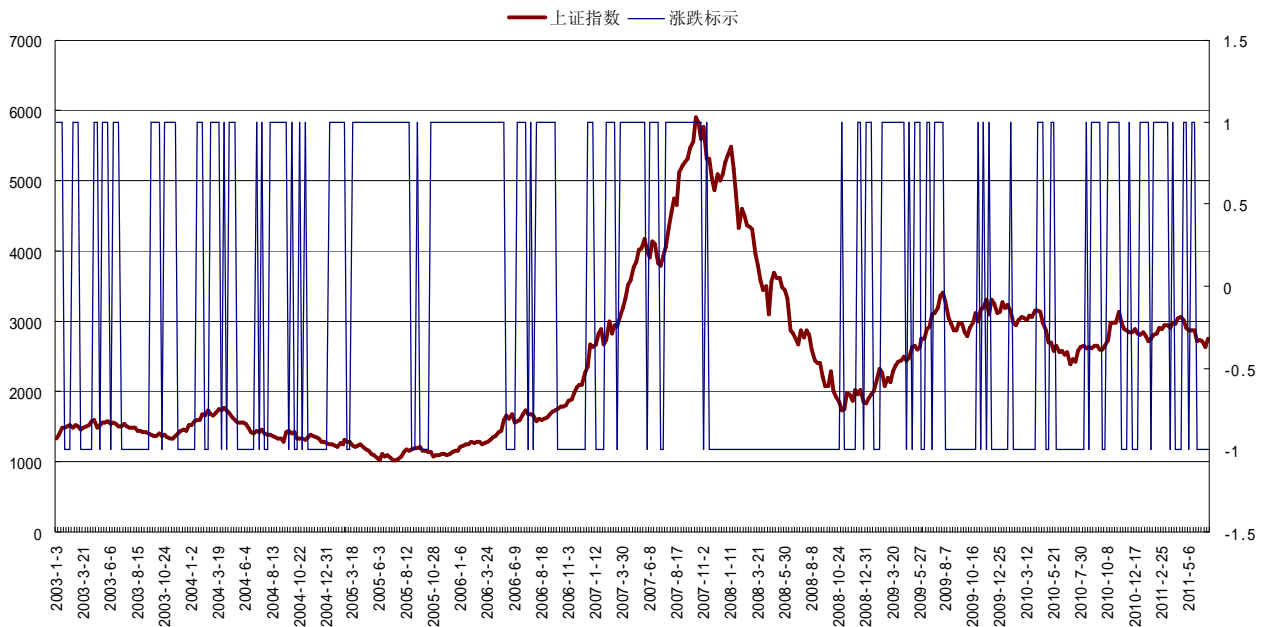
表 4: 不同输入变量组合下模拟交易信息比率

输入变量	无交易费	有交易费
RSI M1M2 滞后 1 期 最高滞后 2 期	-4.79	-5.87

RSI	收盘滞后 2 期	成交额滞后 3 期	5.80	1.02
RSI	最高滞后 1 期	成交额滞后 3 期	3.63	-2.75
RSI	最高滞后 2 期	成交额滞后 3 期	4.53	-2.51
RSI	最低滞后 2 期	成交额滞后 3 期	-0.01	-5.77

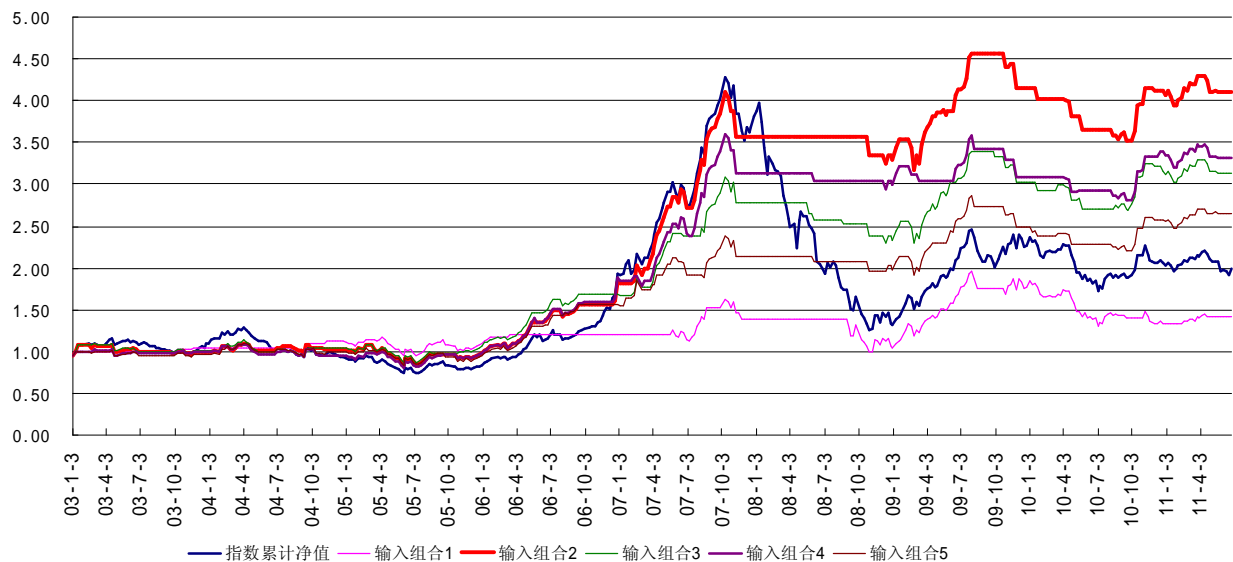
数据来源：广发证券发展研究中心

图 8：最佳输入变量下样本外滚动择时结果



数据来源：广发证券发展研究中心

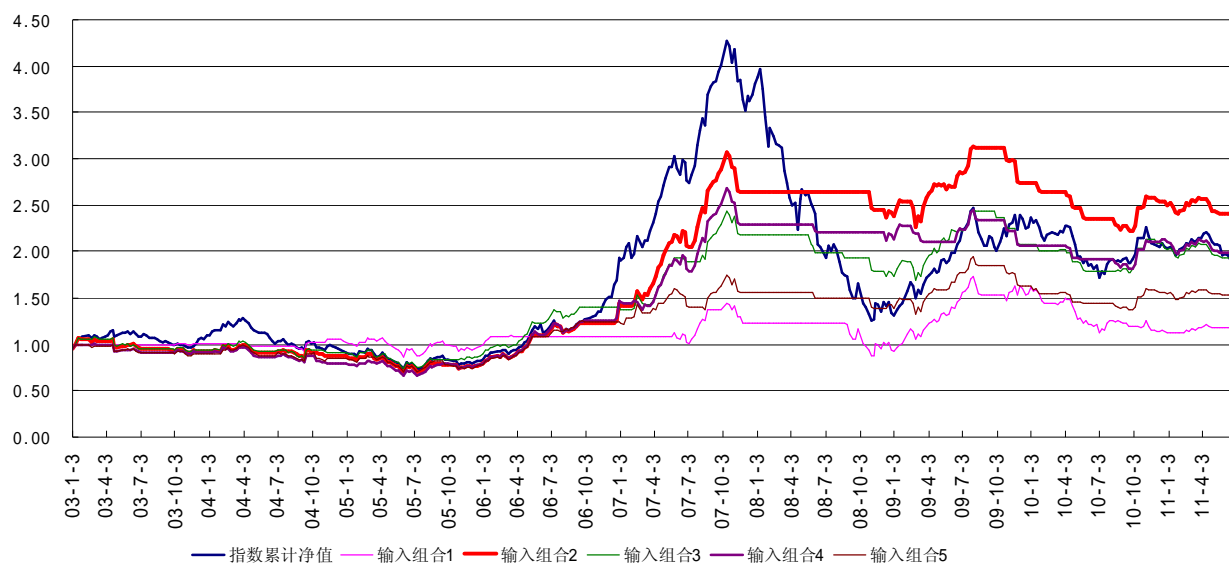
图 9：五种不同输入下模拟交易累计收益率（无交易费用）



数据来源：广发证券发展研究中心

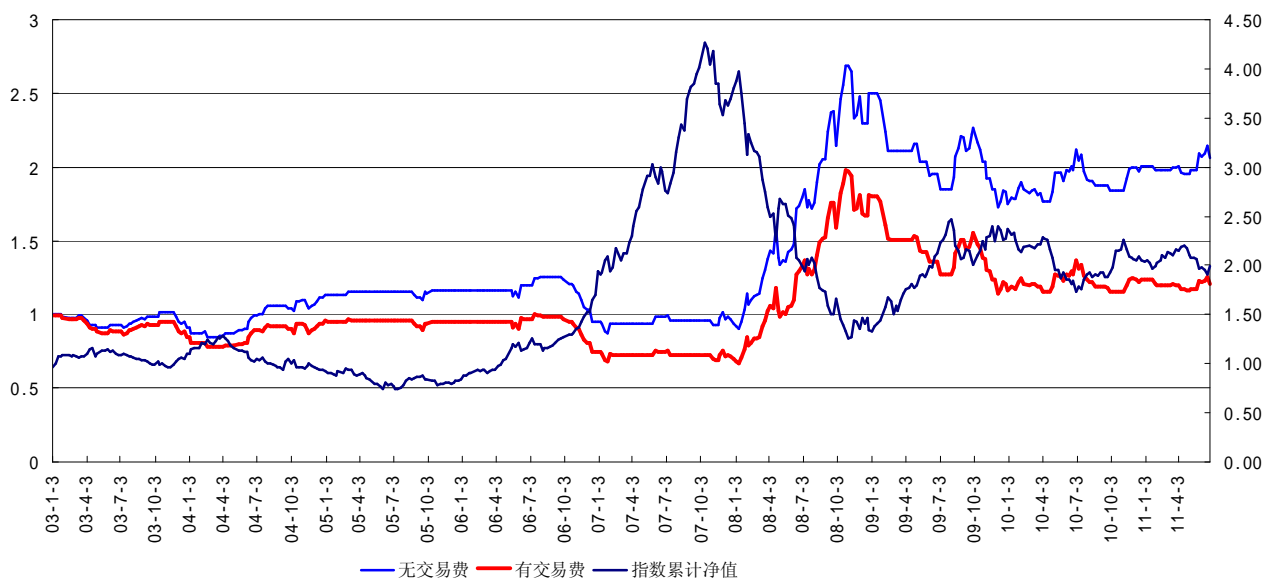
图 10：五种不同输入下模拟交易累计收益率（0.5%单边交易费用）





数据来源：广发证券发展研究中心

图 11：最佳输入变量下模拟交易净值相对指数变化



数据来源：广发证券发展研究中心

分类考察四种不同输入下样本外预测结果的结构性，从表5和图12的结果来看，无论是预测上涨还是预测下跌，大部分都超过了50%，且一个显著的特点是预测上涨的准确率要高于预测下跌的准确率。

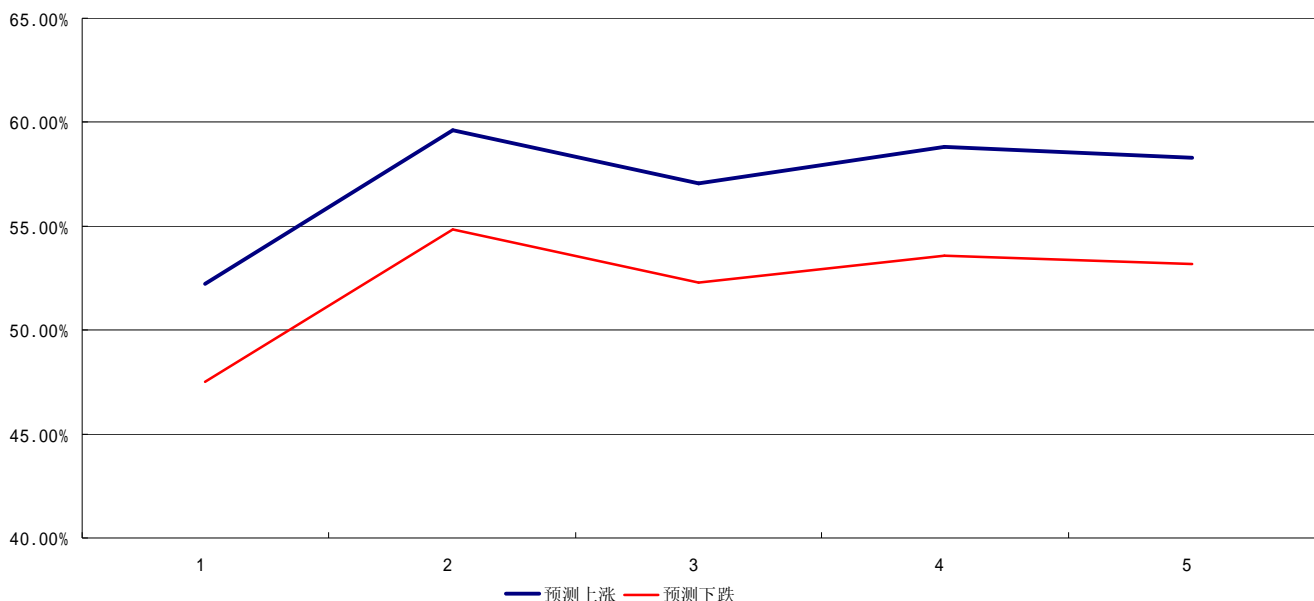
表 5：ANFIS 样本外预测涨跌结构

	预测次数	正确次数	正确率
	RSI M1M2 滞后 1 期 最高滞后 2 期		
预测上涨	203	106	52.22%
预测下跌	225	107	47.56%
	RSI 收盘滞后 2 期 成交额滞后 3 期		

预测上涨	213	127	59.62%
预测下跌	215	118	54.88%
RSI 最高滞后 1 期 成交额滞后 3 期			
预测上涨	212	121	57.08%
预测下跌	216	113	52.31%
RSI 最高滞后 2 期 成交额滞后 3 期			
预测上涨	204	120	58.82%
预测下跌	224	120	53.57%
RSI 最低滞后 2 期 成交额滞后 3 期			
预测上涨	206	120	58.25%
预测下跌	222	118	53.15%

数据来源：广发证券发展研究中心

图 12: ANFIS 样本外预测涨跌准确率比较图



数据来源：广发证券发展研究中心

## 四、自定义优化目标的ANFIS模型

### (一) 两个改进的ANFIS择时模型

前文所阐述的ANFIS模型的优化目标是模型输出与实际输出之间的均方误差，通过调整参数减小预报涨跌与实际涨跌的均方误差意义上的差异度来获得最佳预测模型，但我们可以通过设置更加直接的优化目标，来获得更好的模型，比如直接将择时预测准确率为最大化目标函数，通过调整参数，找出最佳预测模型，再将样本外数据输入模型检验效果。

基于上述理念，我们提出改进的ANFIS模型A，不以实际输出与预测输出之均方误差为目标函数，直接以样本内预测准确率为最大化目标函数来优化参数，目标函数计算流程见图13，模型具体算法如下：

记  $X$  为  $N \times M$  矩阵是训练样本，其中  $N$  为样本长度， $M$  为输入变量个数， $I_{N \times 1}$  为对应的下一期涨跌标志，取值为1或-1表示上涨或下跌。

第1步 对于第  $i$  个样本，将其带入ANFIS模型，计算输出值为  $O_i$ 。

第2步 若  $O_i > 0$ ，则预测涨跌标志  $F_i = 1$ ，否则为  $F_i = -1$ 。

第3步 若  $F_i = I_i$ ，则  $E_i = 1$ ，否则  $E_i = 0$ 。

第4步 对每一个样本，重复第1至3步，并计算此ANFIS模型训练样本择时准确率  

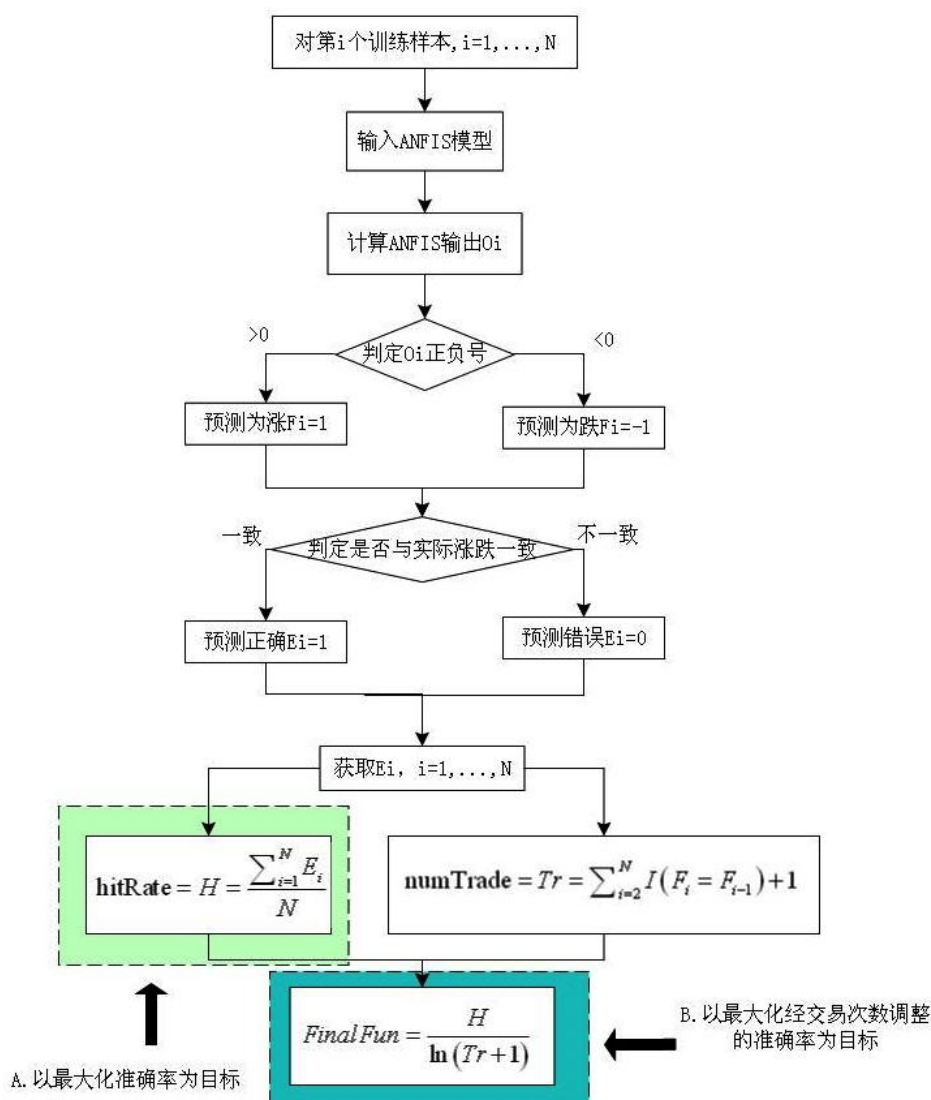
$$\text{hitRate} = H = \sum_{i=1}^N E_i / N。$$

第5步 调整ANFIS参数重新第1至4步直至达到某一优化迭代终止条件，（优化条件可以包括目标函数  $H$  变化幅度小于某一设置基准、迭代次数达到某一预设值等等）。

上述以最大化准确率为目标的ANFIS择时模型的参数优化方法我们采用一种非导数优化方法——遗传算法，其通过模仿物种的自然选择进化繁衍来更新最优参数群体，通过若干次进化找到最佳参数以达到最大化（或最小化）目标函数，具体介绍请参加下一小节内容。另外为便于模型简便以及考虑到高斯隶属函数的适用性，我们此处的ANFIS模型仅取高斯型隶属函数一种，每个输入变量同样赋予2个隶属函数。

为了便于与前文的ANFIS模型进行比较，我们同样选取1996年1月5日至2002年12月27日共计347个周线数据为训练样本，2003年1月3日至2011年6月24日共计428个轴线数据为检验样本，需要注意的是，我们此时不采用滚动预测方式，模型仅训练一次，然后将其余428个样本输入模型进行检验。同时由于改进后的模型与前述ANFIS模型优化目标不同，自然适合改进模型的输入变量也与之前的最佳输入存在差异，由于遗传算法优化模型需要耗费大量时间，我们略去最佳输入寻优过程，经过若干次测试之后，我们发现用如下三个变量输入模型得到的结果较为稳定，它们是涨跌幅、滞后1期的涨跌幅、滞后2期的涨跌幅。

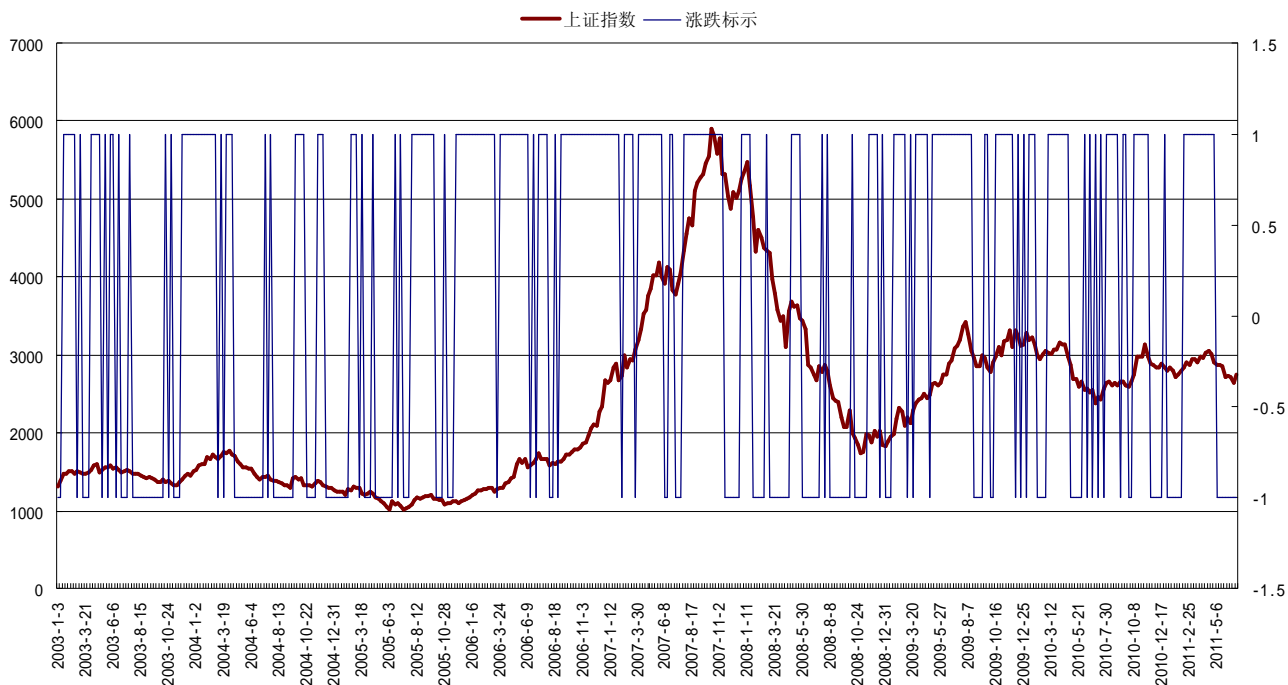
图 13: 两种自定义优化目标函数下的 ANFIS 择时模型



数据来源：广发证券发展研究中心

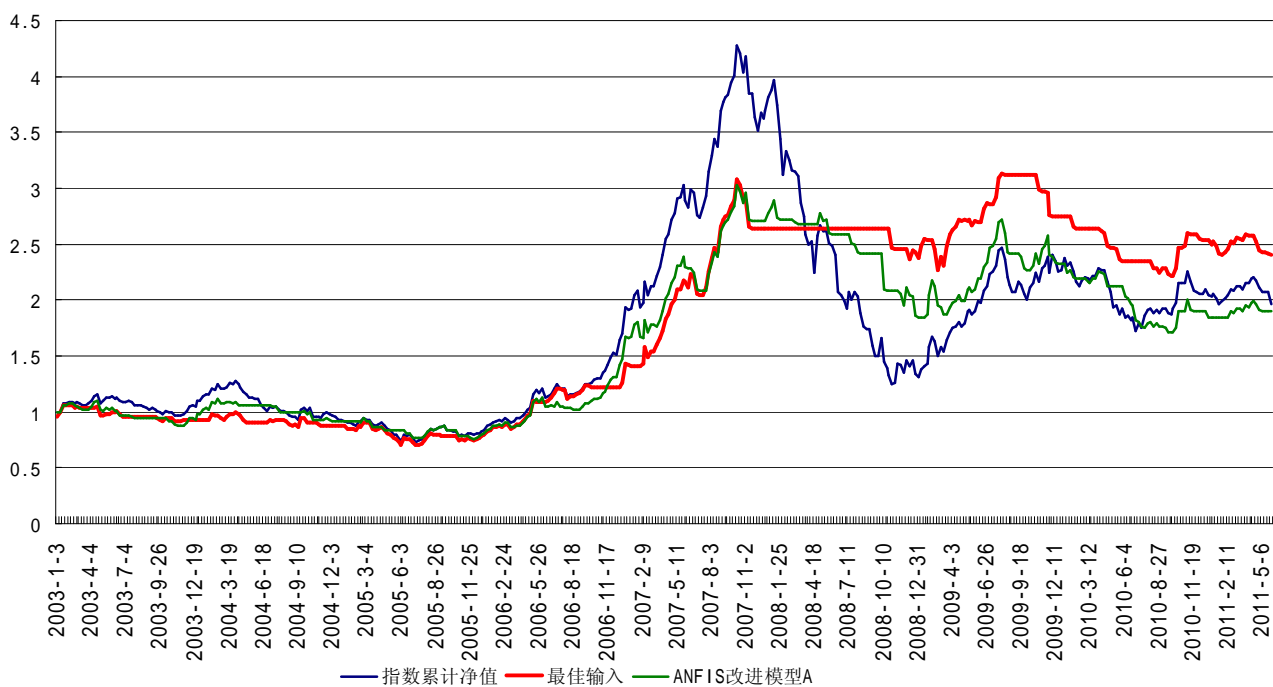
改进后的模型样本外预测准确率达到57.94%，高于前文最佳输入变量（第2组）的57.24%准确率，但从图16、图17的结果来看改进后模型不及前文最佳输入下模型，在考虑了0.5%的单边手续费后，改进模型没有超越指数，对此我们认为这可能是由于交易次数过多所致，也即买卖信号太多，买卖信号的增加会增加手续费支出，复利计算下可导致最后收益相差较大。经计算，前文最佳输入下模型的样本外检验交易次数为106次，改进后模型交易次数为121次，增加了15次。

图 14：ANFIS 模型改进 A 样本外预测结果



数据来源：广发证券发展研究中心

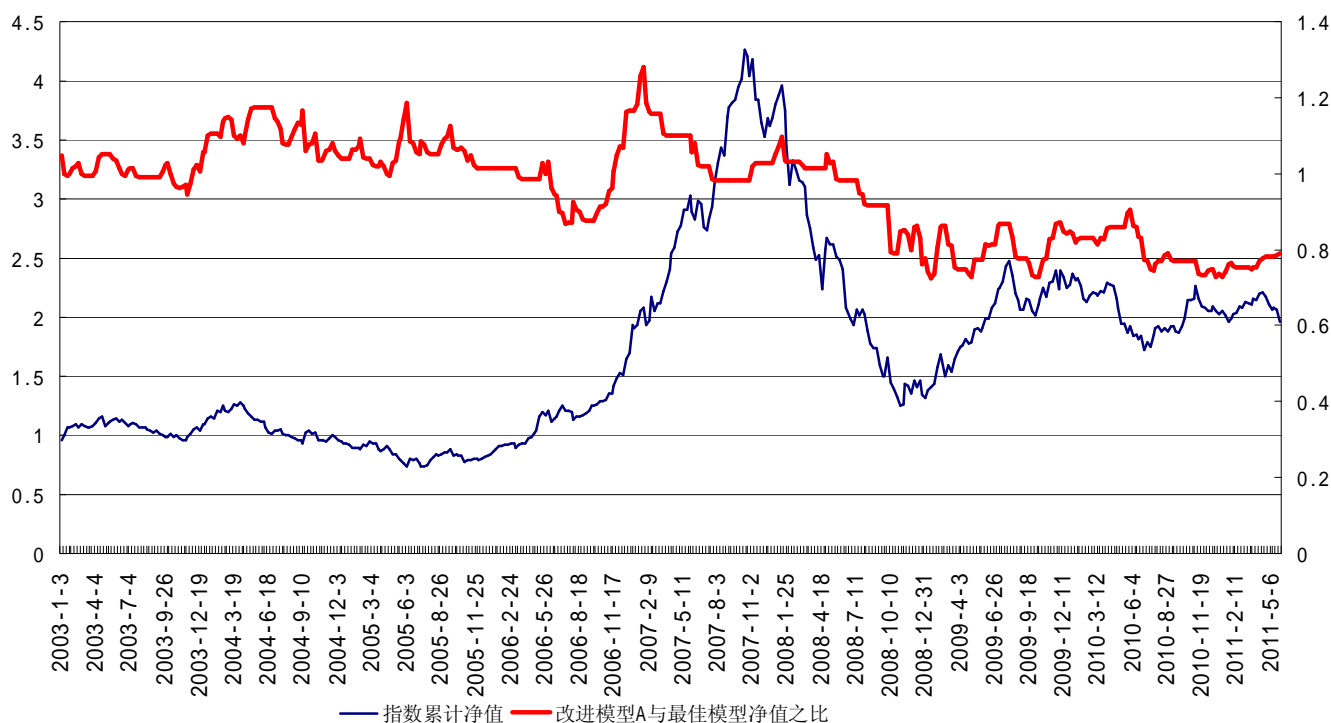
图 15: 改进模型 A 与最佳输入模型净值比较 (考虑手续费)



数据来源：广发证券发展研究中心

图 16: 改进模型 A 与最佳输入模型净值之比 (考虑手续费)





数据来源：广发证券发展研究中心

为了约束交易次数，我们进而提出改进ANFIS模型B，在图13中，改进模型A的优化目标函数为最大化准确率，在改进模型B中我们将目标函数改为经交易次数调整后的准确率函数，具体的，交易次数定义为  $\text{numTrade} = Tr = \sum_{i=2}^N I(F_i = F_{i-1}) + 1$ ，其中  $I(x)$  取值0或1，当  $x$  为真时  $I(x) = 0$ ，当  $x$  为假时  $I(x) = 1$ ，最后计算最终的目标函数：

$$\text{FinalFun} = \frac{H}{\ln(Tr + 1)}$$

其中对交易次数取对数是为了缩小准确率  $H$  与  $Tr$  量级差距带来的影响，将取对数后的交易次数置于分母是为了使得目标函数为交易次数的单调递减函数，对交易次数加1再取对数是为了防止交易次数为  $Tr = 1$  时分母为零。

图17、18、19给出了改进模型B的样本外检验结果，与改进模型A相比交易次数下降了2次，为119次，模型择时准确率达到60.51%，较改进模型A、最佳输入下的模型均有较大幅度提升，从收益来看，考虑手续费后改进模型B的收益超越了指数收益和最佳输入模型收益，净值与最佳输入下模型净值之比基本保持在1以上。

表6的结果显示，改进模型B较最佳输入下模拟交易信息比率有所提高，无交易费下由5.80提高至6.13，有交易费下由1.02提高至2.24。

表 6：最佳输入与改进模型 B 模拟交易信息比率比较

模型	无交易费	有交易费
最佳输入 (RSI 收盘滞后 2 期 成交额滞后 3 期)	5.80	1.02
改进模型 B	6.13	2.24

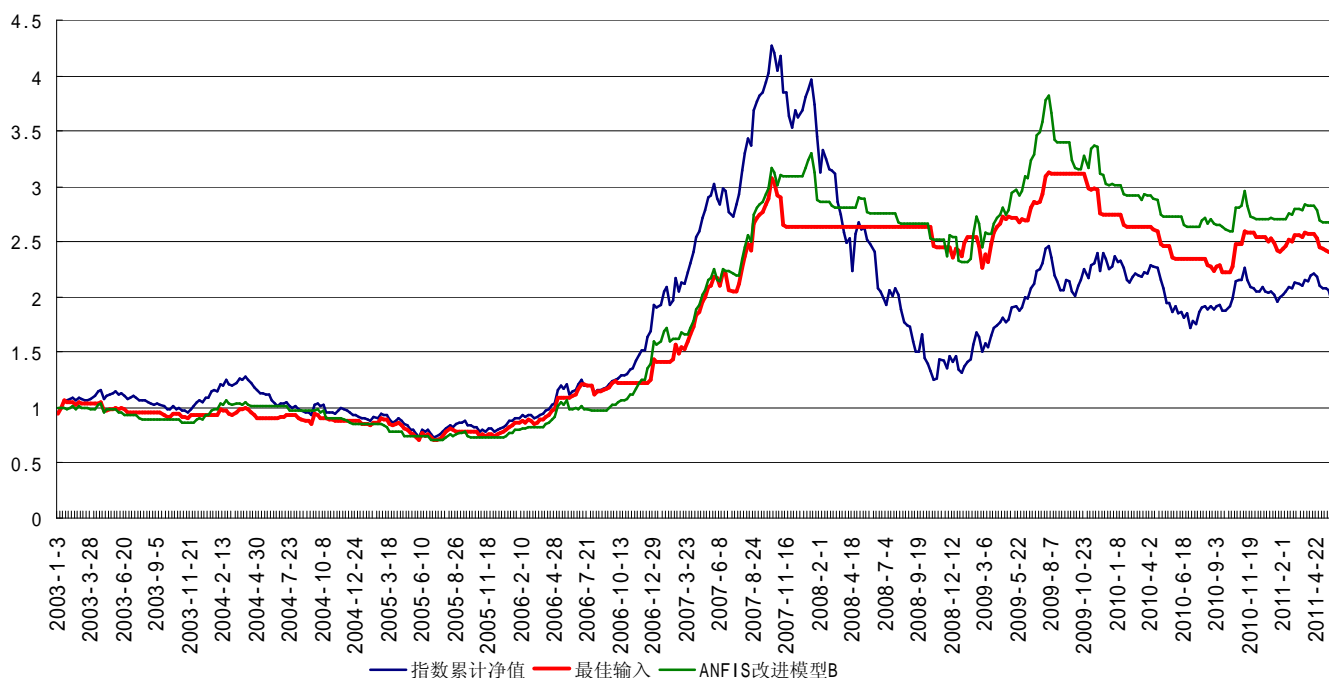
数据来源：广发证券发展研究中心

图 17: 改进模型 B 样本外预测结果



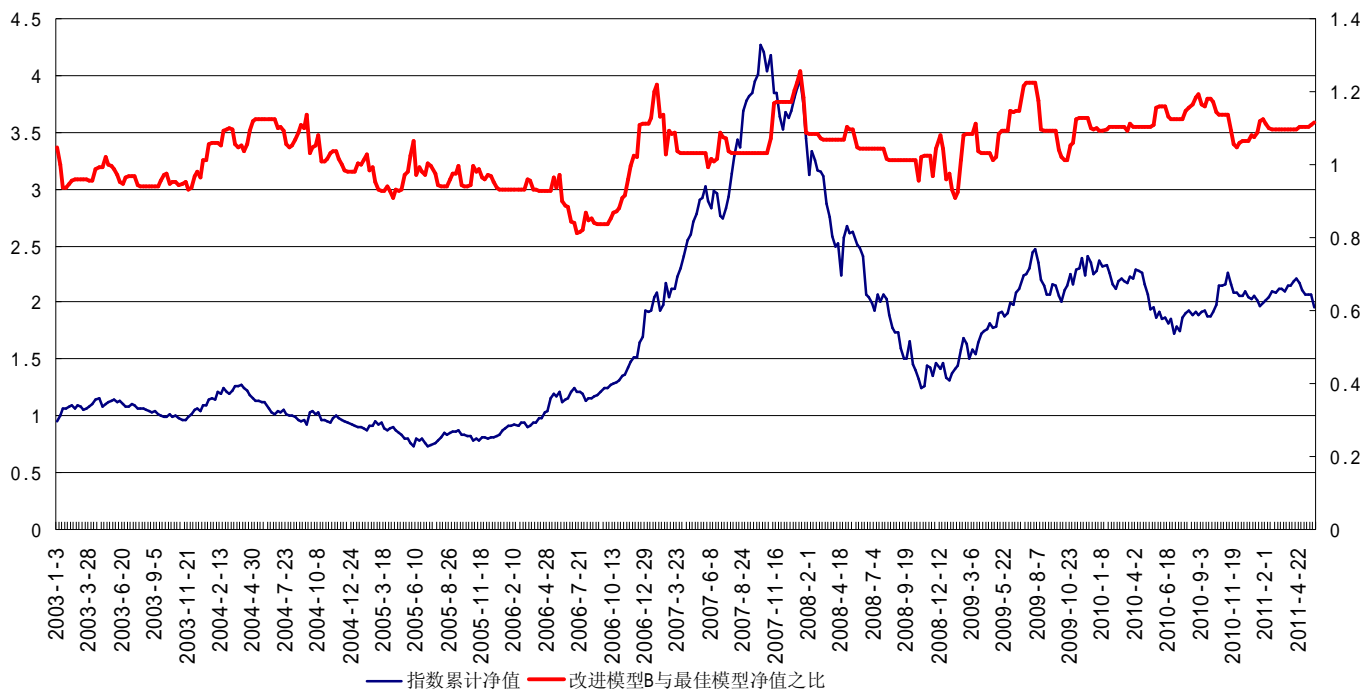
数据来源：广发证券发展研究中心

图 18: 改进模型 B 与最佳输入模型净值比较 (考虑手续费)



数据来源：广发证券发展研究中心

图 19: 改进模型 B 与最佳输入模型净值之比 (考虑手续费)



数据来源：广发证券发展研究中心

## (二) 遗传算法

遗传算法于1975年由密西根大学John Holland提出，是一种建立在自然选择和进化进程概念基础上的非导数随机优化方法。

遗传算法相比于基于导数的优化方法（比如最速下降法，前文中ANFIS模型我们正是利用了最速下降法、最小二乘法并行的一种混合算法来优化参数）有如下特点：

直观意义：参数优化过程类似于物种的进化繁衍，在初始化一个参数群体后通过交换变异等遗传算子进化参数群体，若干次进化后得到最优解。

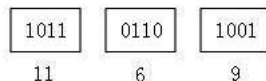
灵活性：传统模型将模型输出与实际输出之均方误差为最小化目标，但遗传算法可以优化任意目标函数，目标函数甚至可以是择时准确率、交易策略收益率、收益稳定性等等，可以对买卖信号触发模型参数以及资金管理参数（如仓位、加减仓参数等资金管理范畴的参数）进行整体优化。

速度慢：遗传算法通过对目标函数重复求值寻优，如果以策略收益为目标函数则需要对样本进行重复计算买卖信号并基于此信号模拟交易，目标函数中就包含很多层循环计算，导致速度较慢，但这样复杂的目标函数是基于导数的优化方法所无法解决的。

随机性：虽然适应度高的个体有更大的机会参与交换计算，但这种机会是带有随机性，因此整个参数种群的更新是具有随机性的，也正是这种随机性才保证了更可能找到最优解而不是局域解。

遗传算法主要包含五个部分：编码方案、适应度计算、父代选择、交换算子、变异算子。

编码方案：编码将参数空间中的点（个体）转换成位串来表示。例如三位空间中的点（11, 6, 9）进行二进制编码得到：



对于浮点数、负数等可对方案进行一定调整进行编码，编码方案提供了一种参数空间中的个体转化为遗传框架下的一种方式，对遗传算法的性能起到决定性作用。

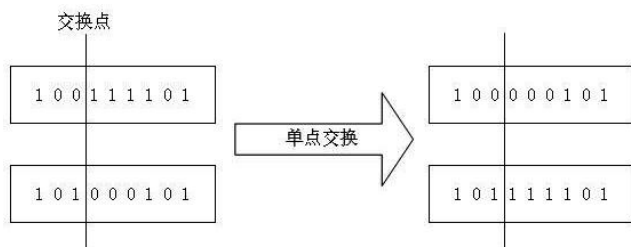
**适应度计算：**得到初始化参数群体后，对于群体中的个体（参数空间中的一个点）计算其适应度值，表示个体在当前环境下的适应度，适应度高的个体能够更有机会生存下去。对于最大化问题，每个个体的适应度通常是该个体的目标函数数值，对于负数可以采用某种单调变换转为正数。

**选择算子：**得到每个个体的适应度后，就要决定选择哪些个体来参与下一代生成，通常适应度高的个体应该更有机会参与，也就是优良的品种应该更能参与物种更新以便遗传其优良性，一般使用与个体适应度值成正比的选择概率来随机选择个体，如选择概率为：

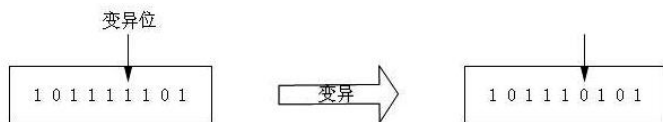
$$p_i = f_i / \sum_{i=1}^n f_i, \quad i=1, \dots, n$$

其中  $p_i$  为选择概率， $f_i$  为个体适应度， $n$  为群体大小。

**交换算子：**选择了参与生成下一代的个体后，利用交换算子生成新个体，并希望新个体能够保存上一代的优良特征。首先根据一定的交换率来对个体配对，若采用单点交换算子，那么对个体编码随机设置一个交换点，两个父代在交换点处互换染色体得到新一代，交换算子类似于自然界中物种交配现象，父代将自身染色体中的某一段遗传给下一代。



**变异算子：**对由交换算子得到的新一代依非常小的变异概率对某一位进行突变，变异算子的作用一方面在于当父代不包含目标函数最优解信息时，则任意次群体更新后可能都不会产生满意结果，此时变异算子将引入新信息来更新群体，另一方面变异算子也会防止群体落入局部最优解区域内。



综上所述，初始化群体后，对其中个体依次计算适应度值，依适应度选择个体参与下一代生成，利用交换算子生成新一代，利用变异算子对新生代变异得到新群体。

最后是初始群体的生成问题，本文自定义ANFIS择时模型参数初始化采用了如下方法：

给定某个输入变量的样本  $x_i$ ， $i=1, \dots, n$ ，记  $nmf$  为ANFIS模型为该输入变量赋予

的隶属函数个数， $x^l = \min_{1 \leq i \leq n} (x_i)$ ， $x^n = \max_{1 \leq i \leq n} (x_i)$ ，则该输入在每个高斯隶属函数标准

差  $S$  参数为:

$$S = \frac{x^n - x^l}{2\sqrt{2\ln(2)}(nmf - 1)}$$

另记  $x'_1, \dots, x'_2, \dots, x'_{nmf}$  为对  $[x^l, x^n]$  区间进行  $nmf$  等分后的子区间端点序列, 则

$$m_m = x'_m, \quad m = 1, \dots, nmf$$

其中  $m_m$  为第  $m$  个高斯隶属函数之均值参数。

## 五、总结

### (一) 研究意义和创新点

(1) 人工智能方法应用于股票价格波动方向预测领域屡见不鲜, 比如人工神经网络、支持向量机模型、隐马尔科夫模型等, 本文基于ANFIS模型应用模糊推理机制来寻求判定价格涨跌之方法。

(2) 布尔逻辑推理机制建立在传统集合方法上, 而传统集合具有明确的边界定义, 但实际中事物是否属于某一类具有模糊性, 模糊数学应运而生, ANFIS模型正是基于模糊数学来进行模糊推理, 适合于股价波动预测等不确定性、混沌性非常强的问题处理。

(3) 在模型输入变量选择上, 本文尝试了由宏观指标和技术指标组成的17个基础输入变量外加51个衍生变量共68个输入变量, 并取其不同组合尝试了5万多次选择了最优输入。

(4) 本文提出构建自定义优化目标的改进ANFIS模型, 以经交易次数调整的准确率为最大化目标得到的模型在准确率以及模拟收益率方面均有所提高。

### (二) 模型的不足

本报告提出的模型虽然取得了较好的预测结果, 但是仍然存在以下不足之处: 模型给出的买卖信号比较多, 特别是当价格呈窄幅波动时, 买卖信号会频繁发出; 模型输出为浮点数, 本文以其正负号为判断涨跌依据, 如何对输出结果更好的分类转换为预测涨跌结果有待进一步研究。

### (三) 后续研究方向

后续研究方向主要围绕以下几个方面: 对模型输出设置动态阈值再进行涨跌分类预测; 将模型应用于日线甚至日内高频数据探究适用性; 在自定义优化目标上再进行研究改进, 在保持准确率以及收益率的基础上减小交易次数。



## 广发金融工程研究小组

罗军，首席分析师，华南理工大学理学硕士，2010年新财富最佳分析师评选入围，2009年进入广发证券发展研究中心。

胡海涛，分析师，华南理工大学理学硕士，2010年新财富最佳分析师评选入围（团队），2010年进入广发证券发展研究中心。

安宁宁，研究助理，暨南大学经济学硕士，2011年进入广发证券发展研究中心，电话：0755-23948352，Email: [ann@gf.com.cn](mailto:ann@gf.com.cn)。

蓝昭钦，研究助理，中山大学数学硕士，2010年新财富最佳分析师评选入围（团队），2010年进入广发证券发展研究中心。

李明，研究助理，伦敦城市大学卡斯商学院计量金融硕士，2010年新财富最佳分析师评选入围（团队），2010年进入广发证券发展研究中心。

史庆盛，研究助理，华南理工大学管理学硕士，2011年进入广发证券发展研究中心。

## 相关研究报告

基于修正 TD 指标的指数择时研究	胡海涛	2010-07-15
探寻西蒙斯投资之道：基于 HMM 模型的周择时策略研究	罗军	2010-06-17

	广州	深圳	北京	上海
地址	广州市天河北路 183 号 大都会广场 5 楼	深圳市民田路华融大厦 2501 室	北京市月坛北街 2 号月坛大 厦 18 层 1808 室	上海市浦东南路 528 号 证券大厦北塔 17 楼
邮政编码	510075	518026	100045	200120
客服邮箱	<a href="mailto:gfyf@gf.com.cn">gfyf@gf.com.cn</a>			
服务热线	020-87555888-8612			

注：广发证券股份有限公司具备证券投资咨询业务资格。本报告只发送给广发证券重点客户，不对外公开发布。

## 免责声明

本报告所载资料的来源及观点的出处皆被广发证券股份有限公司认为可靠，但广发证券不对其准确性或完整性做出任何保证。报告内容仅供参考，报告中的信息或所表达观点不构成所涉证券买卖的出价或询价。广发证券不对因使用本报告的内容而引致的损失承担任何责任，除非法律法规有明确规定。客户不应以本报告取代其独立判断或仅根据本报告做出决策。

广发证券可发出其它与本报告所载信息不一致及有不同结论的报告。本报告反映研究人员的不同观点、见解及分析方法，并不代表广发证券或其附属机构的立场。报告所载资料、意见及推测仅反映研究人员于发出本报告当日的判断，可随时更改且不予通告。

本报告旨在发送给广发证券的特定客户及其它专业人士。未经广发证券事先书面许可，不得更改或以任何方式传送、复印或印刷本报告。