

隐马尔可夫模型的贝叶斯优化及量化择时研究

摘要

本文以隐马尔可夫模型原理为基础，选取 HS300 指数的历史数据建立模型，选择基金的各项指标数据作为观测值进行训练。训练过程中，我们选择了多个状态数进行比较，采用 Baum-Welch 算法进行参数估计。同时，为了保证转移概率矩阵的时效性，我们对模型进行了基于贝叶斯先验估计的参数优化，使得经过新数据的参与，转移概率矩阵逐步接近时下最优情况。对该单特征模型，我们运用交叉验证的思想，在评估集上进行策略回测，确定其收益能力、风险指数、交易成本等系数，并通过层次分析法确定各系数的权重，加权得到该单特征模型的综合评分。通过比较各指标得分以及其它系数完成特征筛选，选取多个有效指标组成特征集，在测试集上完成多特征的综合模型的回测及分析，发现综合模型可以得到长期稳定的超额收益，且有较强的抗风险能力及合适的交易成本。

关键词：隐马尔可夫模型、贝叶斯先验、约束优化、层次分析法、交叉验证、特征筛选

目录

第一章	绪论
1.1	研究背景及意义
1.2	国内外研究现状
1.3	研究内容安排
1.3.1	研究的主要内容
1.3.2	研究内容安排
第二章	隐马尔可夫模型理论综述
2.1	隐马尔可夫模型的基本原理
2.2	隐马尔可夫模型的三个基本问题
2.2.1	问题综述
2.2.2	相关算法原理
2.3	高斯隐马尔可夫模型的参数估计
第三章	贝叶斯先验估计及优化算法
3.1	贝叶斯估计原理综述
3.2	贝叶斯估计与优化算法
3.2.1	贝叶斯先验假设
3.2.2	目标似然函数构造
3.2.3	约束优化算法设计
3.3	优化效果检验
第四章	单特征模型训练及回测结果
4.1	模型训练结果及分析
4.2	交易策略设计及实现
4.3	基本回测结果
4.4	加权评分结果
4.4.1	层次分析法原理及权重设计
4.4.2	系数归一化与加权评分结果
第五章	多特征综合模型回测及结果分析
5.1	综合模型在评估集的回测结果
5.2	综合模型在测试集的表现
第六章	结论与展望
6.1	基本结论与分析
6.2	研究展望与优化空间
	参考文献
	谢辞

第一章 绪论

自 1960 年世界上第一支量化投资基金产生，量化投资方式开始逐渐进入投资领域。随着经济体制和金融市场日趋成熟完善，近年来量化投资在我国得到了迅速的发展。近几年，伴随着大数据及统计学习模型在金融领域的大量应用，我国的量化投资市场也在逐渐向成熟化发展。

在量化研究领域，择时是一种非常重要的策略，该策略认为只要选择合适的时机进行交易，无论选取何种股票，都可以获取收益。本文选择沪深 300 指数作为研究对象，运用马尔可夫模型设计合理的交易策略并进行回测，目标得到针对股票交易的合理择时策略。

1.1 研究背景及意义

近年来，随着我国经济体制日趋完善，金融市场逐渐占据了重要地位，成为了融资、投资的重要渠道。作为金融市场的重要组成部分，股票市场的变动反映着社会各行业的形势变化，是国家经济的晴雨表，也影响着家庭的财务状况，股票交易已经成为了一种重要的投资手段。然而股票市场的特性决定了其通常是复杂且易于波动的，投资者很难得到一套系统的方法指导股票投资，因此，对于股票价格的走势预测已经成为目前金融行业的重大挑战之一。由于未来股票价格的不确定性，其通常可以用某个概率分布来描述，因此统计学方法是最基本的研究股票价格变化的手段。近年来，随着大数据的发展和统计学习方法的大面积应用，其在股票市场也取得了突出的表现，其中，隐马尔可夫模型因其优秀的刻画市场状态的能力更在研究和投资中得到了广泛的应用。

隐马尔可夫模型（HMM）最早在上世纪七十年代由 Baum 等人提出，自 Hamilton（1989）首次在经济领域的研究中引入隐马尔可夫的模型后，其开始在经济金融领域得到广泛应用。然而，现存的研究多数直接使用携带较多噪声的日内开盘价、最高价、最低价、收盘价作为模型的输入特征进行训练，没有在特征选择上进行深入研究与探讨，更缺乏策略层面的实践与验证。

为了在复杂多变的国内市场中研发相对可靠的量化择时策略，取代主观性较强的人工交易方式，本研究利用 HMM 特点对市场状态进行预测，将多个指标观测值（包括收益率、交易量、波动率等）分别作为特征值进行单特征模型训练。为了选取对于马尔可夫模型有效的指标，对于单个特征，我们采用交叉验证的方法，在评估集上进行回测，判断指标有效性并进行打分与筛选。对于选出的指标集，组成多特征的模型进一步训练得到综合模型，预测交易日的市场状态，并设计合理的交易策略。同时，为了保证隐马尔可夫模型各参数的时效性，我们运用了贝叶斯先验估计与约束优化的方法，希望模型参数随着新数据点的加入不断迭代优化，趋向近期最优值。

1.2 国内外研究现状

自 Baum 等人在上世纪七十年代首次提出隐马尔可夫模型，其开始在各个领域得到广泛应用。

1989 年，Hamilton 首次在经济领域的研究中引入隐马尔可夫模型，其将自回归的参数视为离散状态马尔可夫过程的结果，用于估计战后美国经济衰退与国民生产总值下降的关系。自此，隐马尔可夫模型开始被应用于经济金融领域，不断得到完善与改进。

（一）在特征选择方面：Hassan 和 Nath（2005 年）在预测美国一些航空公司的股价时采用了隐马尔可夫模型：在观测值的选择上，他们使用了由当日开盘价、最高价、最低价、收盘价组成的向量作为模型的输入特征，进行观测和训练，并在对比预测时，以该向量出现的概率作为模式识别的依据。

此后的实证分析中，研究者通常使用开盘价、最高价、最低价、收盘价这四个观测值组成的向量作为输入特征进行训练与分析。然而，考虑到国内市场的多变情况，这一观测结果往往包含较大噪声，因而难以用来准确描述市场真实状态。傅中杰、吴清强（2018 年）在已有模型基础上，对不同市场指标值进行回测与分析，选取多个有效指标组合成为综合模型，得到了具有较高累积收益率与较低风险的 HMM 择时策略。

本文在他们研究的基础上，进一步采用交叉验证的思想对候选特征进行筛选，希望得到在真实市场中稳定可靠的择时策略。

（二）在模型优化方面：Scott（2002 年）使用贝叶斯先验和 MCMC 方法对隐马尔可夫模型进行参数估计，其使用基于前向后向算法的 Gibbs 采样以及对数似然估计的方法对隐马尔可夫模型进行拟合与训练。

吴漫君（2011 年）分别使用离散和连续模型对股市的高频数据进行分析并预测股价，发现连续 HMM 模型的预测准确性显著高于离散 HMM 模型。

商晔（2011 年）引入了时域上相关的 HMM 模型，并且将遗传算法应用于 HMM 模型的参数训练中，实证表明在预测准确性上有一定程度的改进。

Tomas A. Dean 等人（2014 年）引入了近似贝叶斯计算（ABC）对 HMM 模型进行参数估计，并推导出了接近标准极大似然估计的一致性和渐进正态的结果。并在 2015 年使用 MCMC 方法与近似贝叶斯计算对模型进行参数估计。

Peter Nystrop 等人（2017 年）提出了一种自适应估计方法，使隐马尔可夫模型的参数具有时变性，可以更好的模拟出市场结果。

本文研究中，选择了较长时间区域作为训练集，因此认为数据基本涵盖了所有可能出现的状态，从而选择固定发射函数的参数，随着新数据的加入，不断优化转移概率矩阵，使其满足时效性。

1.3 研究内容安排

1.3.1 研究的主要内容

本文对经典隐马尔可夫模型进行了贝叶斯优化，在此基础上，针对股市不同指标数据设计了基于过去状态盈利能力的择时策略并在评估集上完成了回测，通过层次分析法进行评分并对指标集做了筛选，最终得到了有较高收益能力与较低风险的综合择时模型。

隐马尔可夫模型由三大经典问题组成：

一、计算观测序列概率。本文使用了经典的前向后向算法进行计算。

二、解码问题，即已知模型参数和观察序列，求出使观察序列出现概率最高的隐状态序列。一般有基于统计的近似算法和基于动态规划的 Viterbi 算法两种计算方法。本文多次采样对比了两种算法的表现，发现其解码效果相似，近似算法的计算复杂度较低。因此在本文研究中主要采用近似算法对隐马尔可夫模型进行解码。

三、参数估计问题。考虑到连续模型有较好的预测准确性，我们假设每个隐状态都服从某个正态分布。在训练集上，我们采用经典的 Baum-Welch 算法对参数进行估计，得到相应的参数集 $(b = (\mu_i, \sigma_i)^n, A = (a_{ij})^{n \times n})$ 。其中 b 对应各正态分布的参数， A 对应转移概率矩阵。针对国内股票市场不稳定多噪声的特点，希望通过贝叶斯优化使得各参数符合当前最优；

同时,因为择时策略总是基于某状态过去一段时间的表现,希望在优化过程中固定各发射概率的参数不变,仅改变转移概率矩阵。本文假设转移概率矩阵的每行均服从 Dirichlet 分布,初始参数 $\Delta = (\delta_{ij})^{n \times n}$ 使用在训练集上由 Baum-Welch 算法得到的参数,随着新数据不断加入,利用约束优化方法使似然函数增大,转移概率矩阵逐渐逼近局部最优值。

为了有效观察策略在国内股票市场上的表现,本文选择了沪深 300 指数的日频数据作为观测和回测对象,并计算得到了多种常用指标,包括波动率、收益率、交易量等。为了选择更合适的特征集并进行有效回测,本文采用了交叉验证的思想,将全部数据集分为训练集(2006 年 1 月 1 日至 2015 年 1 月 1 日)、评估集(2015 年 1 月 1 日至 2018 年 1 月 1 日)、测试集(2018 年 1 月 1 日至 2019 年 4 月 10 日),并在训练集上进行 HMM 模型的训练与参数估计,在评估集上进行参数优化与单特征模型回测,并筛选得到综合模型的指标,最后在测试集上验证综合择时模型的表现。

在设计择时策略时,考虑到国内市场的的不稳定性以及高噪声特点,本文择时策略不涉及股价预测,而是选择观察某状态在过去一段时间(一个月即 20 个交易日)的盈利能力。如果某状态在过去一段时间至少出现过一次且收益为正,则标记为盈利状态,否则标记为非盈利状态。对于次日股价,利用隐马尔可夫模型进行状态预测,得到一个离散的概率向量,若次日处于盈利状态的概率大于 60%,则认为次日应当持有股票,否则不持有。以此为依据设计择时策略,并在评估集上对不同单指标分别进行回测。对于回测曲线,分别计算其常用评价指标,包括收益率、最大回撤、交易频率、夏普比率等。

在状态数选择上,观察到不同指标的最优状态数有较大差别,其回测表现对状态数选择的依赖程度也不尽相同,因此本文每次选取区间 $[3, 9]$ 内的每个整数作为状态数,分别进行训练与回测,取其中表现最优者参与指标筛选。

本文利用层次分析法确定各评价指标的重要性即权重,首先得到基于经验的比较权重矩阵,以此为依据计算出各评价指标的权重,其次对各评价指标进行归一化处理,并确定符号(收益率等指标取正号,最大回撤等取负号),最后加权求和,得到某指标的回测评分。本文依据回测评分与其他评价指标对各指标进行了筛选,得到了几个表现较好的指标(包括单日收益率、真实波幅、人气指标等),对其进行组合形成特征集。将它们组成的多维向量作为输入对综合隐马尔可夫模型进行训练与回测,观察其在评估集上的表现,确定状态数范围,最后在测试集上进行回测,得到其实证表现并进行了分析与评价。

1.3.2 研究内容安排

本文在经典隐马尔可夫模型的基础上应用了贝叶斯优化进行参数估计,并对单个指标分别进行了回测与筛选,得到的指标组合成为综合隐马尔可夫模型的输入特征,并最后进行了回测与分析。

论文主要分为六个部分:

第一章,绪论。介绍了本文的研究背景及意义,国内外研究现状以及本文研究的主要内容及行文安排。

第二章,隐马尔可夫模型理论综述。主要从数学角度介绍了隐马尔可夫模型的基本原理及定义,然后提出了隐马尔可夫模型的三个基本问题(计算观测序列概率、解码问题、参数估计)并给出了相应经典算法及其数学推导,最后假设观测序列服从混合高斯分布,各隐状态服从多元高斯分布,给出了连续隐马尔可夫模型的参数估计方法。

第三章,贝叶斯先验估计及优化算法。首先简单介绍了贝叶斯先验估计的原理,然后阐述了利用贝叶斯估计与约束优化方法通过迭代使转移概率矩阵逼近最优值的过程,包括贝叶斯先验假设、目标似然函数的构造以及优化算法的设计,最后对优化效果进行了检验与分析。

第四章,单特征模型训练及回测结果。本章首先对模型训练结果进行了简单的分析,然

后介绍了择时策略的设计与实现，包括基本设计思想和市场情况考虑，接下来展示了各指标的基本回测结果（用多种评价指标表示）。对于各指标进行加权评分，介绍了层次分析法原理，比较权重矩阵与权重向量，并最后展示了各测试指标加权评分的结果。

第五章，多特征综合模型回测及结果分析。本章介绍了特征筛选规则的制订标准及筛选结果，分别展示了综合模型在评估集与测试集上的表现并进行了简单分析。

第六章，结论与展望。本章总结了回测模型的实证表现结果并进行了简单的分析，最后对后续研究进行了展望，说明了优化空间。

第二章 隐马尔可夫模型理论综述

隐马尔可夫模型（HMM）是由状态序列和观测序列组成的双重随机过程。其中，状态序列是由一个隐藏的马尔可夫链随机生成的，对应每个状态又生成了一个观测值，这样就组成了一个观测随机序列。

2.1 隐马尔可夫模型的基本原理

隐马尔可夫模型由初始概率分布、状态转移概率矩阵、以及观测概率分布确定。隐马尔可夫模型的形式定义如下：

设 Q 是有可能状态的集合， V 是有可能观测的集合。

$$Q = \{q_1, q_2, \dots, q_N\}, \quad V = \{v_1, v_2, \dots, v_M\}$$

其中， N 是可能的状态数， M 是肯的观测数。

假设观测序列长度为 T ， I 是这段时间的状态序列， O 是对应的观测序列。

$$I = \{i_1, i_2, \dots, i_T\}, \quad O = \{o_1, o_2, \dots, o_T\}$$

A 表示状态转移概率矩阵：

$$A = [a_{ij}]_{N \times N}$$

其中，

$$a_{ij} = P(i_{t+1} = q_j | i_t = q_i), \quad i = 1, 2, \dots, N; j = 1, 2, \dots, N$$

表示在时刻 t 处于状态 q_i 的条件下在时刻 $t+1$ 转移到状态 q_j 的概率。

B 是由隐藏状态生成观测序列时的概率矩阵：

$$B = [b_j(k)]_{N \times M}$$

其中，

$$b_j(k) = P(o_t = v_k | i_t = q_j), \quad k = 1, 2, \dots, M; j = 1, 2, \dots, N$$

是在时刻 t 处于状态 q_j 的条件下生成观测 v_k 的概率。

π 是初始状态概率向量：

$$\pi = (\pi_i)_N$$

其中，

$$\pi_i = P(i_1 = q_i), \quad i = 1, 2, \dots, N$$

是时刻 $t=1$ 时处于状态 q_i 的概率。

隐马尔可夫模型由初始状态转移向量 π 、转移概率矩阵 A 、以及观测概率矩阵 B 决定。因此，隐马尔可夫模型 λ 可以唯一地用三元符号表示，即

$$\lambda = (A, B, \pi)$$

由定义可知，隐马尔可夫模型做了两个基本的假设：

(1) 齐次马尔可夫性假设。即假设隐藏的马尔可夫链在任意时刻的状态只与其前一刻的状态有关，与其他时刻的状态及观测无关，也与时刻 t 无关，也就是：

$$P(i_t | i_{t-1}, o_{t-1}, \dots, i_1, o_1) = P(i_t | i_{t-1}), \quad t = 1, 2, \dots, T$$

(2) 观测独立性假设。即假设任意时刻的观测只依赖于该时刻的马尔可夫链隐状态，与其他观测及状态均无关，也就是：

$$P(o_t | i_T, o_T, \dots, i_{t+1}, o_{t+1}, i_{t-1}, o_{t-1}, \dots, i_1, o_1) = P(i_t | i_{t-1}), \quad t = 1, 2, \dots, T$$

2.2 隐马尔可夫模型的三个基本问题

2.2.1 问题综述

隐马尔可夫模型有三个基本问题：

(1) 概率计算问题。给定隐马尔可夫模型相关参数 $\lambda = (A, B, \pi)$ 和某观测序列 $O = \{o_1, o_2, \dots, o_T\}$ ，计算在该模型下观测序列 O 出现的概率 $P(O | \lambda)$ 。

(2) 学习问题。已知观测序列，对隐马尔可夫模型的参数进行估计，使得在该模型下观测序列概率 $P(O | \lambda)$ 最大，也就是用极大似然估计的方法估计参数。

(3) 预测问题（也称解码问题）。已知模型和观测序列，求对于给定的观测序列，使其条件概率 $P(O | \lambda)$ 最大的状态序列 $I = (i_1, i_2, \dots, i_T)$ 。

下面逐一介绍这三种问题的解法及原理。

2.2.2 相关算法原理

(1) 概率计算问题：

概率计算问题一般有两种解法：直接算法以及前向后向（forward-backward）算法。其中，直接计算方法仅在理论上可行，但由于其计算复杂度太高，一般不能得到有效应用。本文对这两种算法分别进行了介绍，首先介绍直接算法的原理：

1) 直接算法

对于给定的观测序列 $O = \{o_1, o_2, \dots, o_T\}$ ，通过列举所有可能的长度为 T 的隐藏状态序列，分别求联合概率 $P(O, I | \lambda)$ ，最后对所有可能的隐藏状态序列 $I = (i_1, i_2, \dots, i_T)$ 求和，得到 $P(O | \lambda)$ 。

已知状态序列 $O = \{o_1, o_2, \dots, o_T\}$ 在 HMM 模型 $\lambda = (A, B, \pi)$ 发生的概率为

$$P(I | \lambda) = \pi_{i_1} a_{i_1 i_2} a_{i_2 i_3} \dots a_{i_{T-1} i_T}$$

对于已知的状态序列，观测序列 $O = \{o_1, o_2, \dots, o_T\}$ 出现的概率是

$$P(O | I, \lambda) = b_{i_1}(o_1) b_{i_2}(o_2) \dots b_{i_T}(o_T)$$

则可以得到联合概率分布

$$\begin{aligned} P(O, I | \lambda) &= P(O | I, \lambda) P(I | \lambda) \\ &= \pi_{i_1} b_{i_1}(o_1) a_{i_1 i_2} b_{i_2}(o_2) a_{i_2 i_3} \dots a_{i_{T-1} i_T} b_{i_T}(o_T) \end{aligned}$$

对其求和得到

$$\begin{aligned} P(O | \lambda) &= \sum_I P(O | I, \lambda) P(I | \lambda) \\ &= \sum_{i_1, i_2, \dots, i_T} \pi_{i_1} b_{i_1}(o_1) a_{i_1 i_2} b_{i_2}(o_2) a_{i_2 i_3} \dots a_{i_{T-1} i_T} b_{i_T}(o_T) \end{aligned}$$

但是，直接算法的计算复杂度达到了 $O(TN^T)$ 阶，因此其虽然在概念上成立，但在实际应用中一般是不可行的。

2) 前向后向算法

一、前向算法

首先定义前向概率：给定隐马尔可夫模型 λ ，到时刻 t 部分观测序列为 o_1, o_2, \dots, o_t 且当

前状态为 q_i 的概率称为前向概率，记作

$$\alpha_t(i) = P(o_1, o_2, \dots, o_t, i_t = q_i | \lambda)$$

通过递推可以得到前向概率及观测序列概率。

算法步骤：

初值：

$$\begin{aligned}\alpha_1(i) &= P(o_1, i_1 = q_i | \lambda) \\ &= P(o_1 | i_1 = q_i, \lambda) P(i_1 = q_i | \lambda) \\ &= \pi_i b_i(o_1), \quad (i = 1, 2, \dots, N)\end{aligned}$$

递推：对 $t = 1, 2, \dots, T-1$,

$$\begin{aligned}\alpha_{t+1}(i) &= P(o_1, o_2, \dots, o_{t+1}, i_{t+1} = q_i | \lambda) \\ &= \sum_{j=1}^N P(o_1, o_2, \dots, o_t, i_t = q_j | \lambda) P(o_{t+1}, i_{t+1} = q_i | i_t = q_j, \lambda) \\ &= \sum_{j=1}^N \alpha_t(j) P(i_{t+1} = q_i | i_t = q_j, \lambda) P(o_{t+1} | i_{t+1} = q_i, \lambda) \\ &= [\sum_{j=1}^N \alpha_t(j) a_{ji}] b_i(o_{t+1}), \quad (i = 1, 2, \dots, N)\end{aligned}$$

终止：

$$\begin{aligned}P(O | \lambda) &= \sum_{i=1}^N P(o_1, o_2, \dots, o_T, i_T = q_i | \lambda) \\ &= \sum_{i=1}^N \alpha_T(i)\end{aligned}$$

前向算法高效的关键在于其局部计算前向概率，然后利用路径结构将前向概率递推到全局，也就是每一次计算直接引用前一个时刻的计算结果，避免重复计算。前向算法的计算复杂度是 $O(N^2T)$ 阶的。

二、后向算法

定义后向概率：给定隐马尔可夫模型 λ ，在时刻 t 部分观测序列为且当前状态为 q_i 的条件下，从 $t+1$ 到 T 的部分观测序列为 $o_{t+1}, o_{t+2}, \dots, o_T$ 的概率称为后向概率，记作

$$\beta_t(i) = P(o_{t+1}, o_{t+2}, \dots, o_T | i_t = q_i, \lambda)$$

与计算前向概率相似，通过递推可以得到后向概率及观测序列概率。

算法步骤：

初值：

$$\beta_T(i) = 1, \quad i = 1, 2, \dots, N$$

递推：对 $t = T-1, T-2, \dots, 1$,

$$\begin{aligned}\beta_t(i) &= P(o_{t+1}, o_{t+2}, \dots, o_T | i_t = q_i, \lambda) \\ &= \sum_{j=1}^N P(o_{t+1}, o_{t+2}, \dots, o_T, i_{t+1} = q_j | i_t = q_i, \lambda) \\ &= \sum_{j=1}^N P(o_{t+2}, \dots, o_T | i_t = q_i, i_{t+1} = q_j, o_{t+1}, \lambda) P(o_{t+1}, i_{t+1} = q_j | i_t = q_i, \lambda) \\ &= \sum_{j=1}^N P(o_{t+2}, \dots, o_T | i_{t+1} = q_j, \lambda) P(i_{t+1} = q_j | i_t = q_i, \lambda) P(o_{t+1} | i_{t+1} = q_j, \lambda) \\ &= \sum_{j=1}^N \beta_{t+1}(j) a_{ij} b_j(o_{t+1})\end{aligned}$$

终止：

$$\begin{aligned}
P(O|\lambda) &= \sum_{i=1}^N P(o_1, o_2, \dots, o_T, i_1 = q_i | \lambda) \\
&= \sum_{i=1}^N P(o_1 | i_1 = q_i, o_2, o_3, \dots, o_T, \lambda) P(i_1 = q_i, o_2, \dots, o_T | \lambda) \\
&= \sum_{i=1}^N P(o_1 | i_1 = q_i, \lambda) P(o_2, \dots, o_T | i_1 = q_i, \lambda) P(i_1 = q_i | \lambda) \\
&= \sum_{i=1}^N b_i(o_1) \beta_i(i) \pi_i
\end{aligned}$$

与前向算法相同，后向算法的计算复杂度也是 $O(N^2T)$ 阶的。

(2) 学习问题：

1) 监督学习方法

假设给定训练数据的观测序列与状态序列均为已知，则可以利用极大似然的方法估计隐马尔可夫模型的参数。

转移概率矩阵的估计：

设样本中由状态 i 转移到状态 j 的频数为 A_{ij} ，则

$$a_{ij} = \frac{A_{ij}}{\sum_{j=1}^N A_{ij}}, \quad i = 1, 2, \dots, N; j = 1, 2, \dots, N$$

观测概率的估计：

设样本中状态为 j 且观测值为 k 的频数是 B_{jk} ，则

$$b_j(k) = \frac{B_{jk}}{\sum_{k=1}^M B_{jk}}, \quad j = 1, 2, \dots, N; k = 1, 2, \dots, M$$

2) Baum-Welch 算法

对于给定的观测数据 $O = \{o_1, o_2, \dots, o_T\}$ ，假设不知道其对应的状态序列 $I = \{i_1, i_2, \dots, i_T\}$ ，那么隐马尔可夫模型事实上可以看做是一个含有隐变量的概率模型

$$P(O|\lambda) = \sum_I P(O|I, \lambda) P(I|\lambda)$$

其参数学习可以由 EM 算法实现。

首先确定对数似然函数，完全数据 $(O, I) = (o_1, o_2, \dots, o_T, i_1, i_2, \dots, i_T)$ 的对数似然函数是 $\log P(O, I|\lambda)$ 。

然后完成 EM 算法的 E 步：求 Q 函数 $Q(\lambda, \bar{\lambda})$

$$Q(\lambda, \bar{\lambda}) = \sum_I \log P(O, I|\lambda) P(O, I|\bar{\lambda})$$

其中， $\bar{\lambda}$ 是隐马尔可夫模型参数的当前估计值， λ 是需要最优化的参数。由于

$$P(O, I|\lambda) = \pi_{i_1} b_{i_1}(o_1) a_{i_1 i_2} b_{i_2}(o_2) \dots a_{i_{T-1} i_T} b_{i_T}(o_T)$$

得到

$$\begin{aligned}
Q(\lambda, \bar{\lambda}) &= \sum_I \log \pi_{i_1} P(O, I | \bar{\lambda}) \\
&\quad + \sum_I \left(\sum_{t=1}^{T-1} \log a_{i_t i_{t+1}} \right) P(O, I | \bar{\lambda}) \\
&\quad + \sum_I \left(\sum_{t=1}^T \log b_{i_t}(o_t) \right) P(O, I | \bar{\lambda})
\end{aligned}$$

接下来完成 EM 算法的 E 步：求极大化 Q 函数的模型参数 A, B, π 对各项分别进行极大化：

(1)

$$\sum_I \log \pi_{i_1} P(O, I | \bar{\lambda}) = \sum_{i=1}^N \log \pi_i P(O, i_1 = i | \bar{\lambda})$$

由约束条件：

$$\sum_{i=1}^N \pi_i = 1$$

利用拉格朗日乘子法，得到函数：

$$\sum_{i=1}^N \log \pi_i P(O, i_1 = i | \bar{\lambda}) + \gamma \left(\sum_{i=1}^N \pi_i - 1 \right)$$

令倒数为零：

$$\begin{aligned}
\frac{\partial}{\partial \pi_i} \left[\sum_{i=1}^N \log \pi_i P(O, i_1 = i | \bar{\lambda}) + \gamma \left(\sum_{i=1}^N \pi_i - 1 \right) \right] &= 0 \\
\Rightarrow P(O, i_1 = i | \bar{\lambda}) + \gamma \pi_i &= 0 \\
\Rightarrow \gamma &= -P(O | \bar{\lambda})
\end{aligned}$$

代入得到：

$$\pi_i = \frac{P(O, i_1 = i | \bar{\lambda})}{P(O | \bar{\lambda})}$$

(2)

$$\sum_I \left(\sum_{t=1}^{T-1} \log a_{i_t i_{t+1}} \right) P(O, I | \bar{\lambda}) = \sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^{T-1} \log a_{ij} P(O, i_t = i, i_{t+1} = j | \bar{\lambda})$$

由约束条件：

$$\sum_{i=1}^N \pi_i = 1$$

利用拉格朗日乘子法，得到函数：

$$\sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^{T-1} \log a_{ij} P(O, i_t = i, i_{t+1} = j | \bar{\lambda}) + \gamma \left(\sum_{i=1}^N \pi_i - 1 \right)$$

令倒数为零：

$$\begin{aligned}
\frac{\partial}{\partial a_{ij}} \left[\sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^{T-1} \log a_{ij} P(O, i_t = i, i_{t+1} = j | \bar{\lambda}) + \gamma \left(\sum_{i=1}^N \pi_i - 1 \right) \right] &= 0 \\
\Rightarrow \sum_{t=1}^{T-1} P(O, i_t = i, i_{t+1} = j | \bar{\lambda}) + \gamma a_{ij} &= 0 \\
\Rightarrow \gamma &= -\sum_{t=1}^{T-1} P(O, i_t = i | \bar{\lambda})
\end{aligned}$$

代入得到：

$$a_{ij} = \frac{\sum_{t=1}^{T-1} P(O, i_t = i, i_{t+1} = j | \bar{\lambda})}{\sum_{t=1}^{T-1} P(O, i_t = i | \bar{\lambda})}$$

(3)

$$\sum_I (\sum_{t=1}^T \log b_{i_t}(o_t)) P(O, I | \bar{\lambda}) = \sum_{j=1}^N \sum_{t=1}^T \log b_j(o_t) P(O, i_t = j | \bar{\lambda})$$

同样使用拉格朗日乘子法，得到：

$$b_j(k) = \frac{\sum_{t=1}^T P(O, i_t = j | \bar{\lambda}) I(o_t = v_k)}{\sum_{t=1}^T P(O, i_t = j | \bar{\lambda})}$$

对于以上出现的条件概率，不妨令：

$$\gamma_t(i) = P(i_t = q_i | O, \lambda) = \frac{P(i_t = q_i, O | \lambda)}{P(O | \lambda)}$$

$$\xi_t(i, j) = P(i_t = q_i, i_{t+1} = q_j | O, \lambda) = \frac{P(i_t = q_i, i_{t+1} = q_j, O | \lambda)}{P(O | \lambda)}$$

其中， $\gamma_t(i)$ 和 $\xi_t(i, j)$ 均可通过前向后向算法得到：

由

$$\alpha_t(i) \beta_t(i) = P(i_t = q_i, O | \lambda)$$

得到：

$$\gamma_t(i) = \frac{P(i_t = q_i, O | \lambda)}{P(O | \lambda)} = \frac{\alpha_t(i) \beta_t(i)}{\sum_{j=1}^N \alpha_t(j) \beta_t(j)}$$

又由定义可知

$$P(i_t = q_i, i_{t+1} = q_j, O | \lambda) = \alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)$$

于是

$$\xi_t(i, j) = \frac{\alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^N \sum_{j=1}^N \alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}$$

Baum-Welch 算法步骤：

初值：

$$\text{选取 } \lambda^{(0)} = (A^{(0)}, B^{(0)}, \pi^{(0)})$$

递推：对 $n = 1, 2, \dots$

$$a_{ij}^{(n+1)} = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(j)}$$

$$b_j(k)^{(n+1)} = \frac{\sum_{t=1, o_t=v_k}^T \gamma_t(j)}{\sum_{t=1}^T \gamma_t(j)}$$

$$\pi_i^{(n+1)} = \gamma_1(i)$$

终止：

当迭代次数超过阈值或参数达到收敛时终止。

(4) 预测问题：

隐马尔可夫的预测（解码）问题通常采用两种解法：近似算法和 Viterbi 算法。其中，近似算法的计算简单，但难以保证预测序列整体的最大发生概率，特别是当存在不互通的状态时。Viterbi 算法利用了动态规划的思想求最优路径（概率最大路径），有效的解决了近似算法的不足之处，但相对而言计算更加复杂。

1) 近似算法

已知时刻 t 观测处于状态 q_i 的概率 $\gamma_t(i)$ 是

$$\gamma_t(i) = \frac{P(i_t = q_i, O | \lambda)}{P(O | \lambda)} = \frac{\alpha_t(i)\beta_t(i)}{\sum_{j=1}^N \alpha_t(j)\beta_t(j)}$$

则在每一时刻最有可能的状态是

$$i_t^* = \arg \max_{1 \leq i \leq N} [\gamma_t(i)], \quad t = 1, 2, \dots, T$$

2) Viterbi 算法

定义两个变量 δ 和 ψ ， $\delta_t(i)$ 为时刻 t 状态为 q_i 的所有单个路径中概率最大值：

$$\delta_t(i) = \max_{i_1, i_2, \dots, i_{t-1}} P(i_t = q_i, i_{t-1}, \dots, i_1, o_t, \dots, o_1 | \lambda)$$

由定义可得：

$$\begin{aligned} \delta_{t+1}(i) &= \max_{i_1, i_2, \dots, i_t} P(i_{t+1} = q_i, i_t, \dots, i_1, o_{t+1}, \dots, o_1 | \lambda) \\ &= \max_{1 \leq j \leq N} [\delta_t(j) a_{ji}] b_i(o_{t+1}) \end{aligned}$$

$\psi_t(i)$ 为所有满足条件的单个路径中概率最大的路径的第 $t-1$ 个节点。

$$\psi_t(i) = \arg \max_{1 \leq j \leq N} [\delta_{t-1}(j) a_{ji}]$$

得到 Viterbi 算法的步骤：

初值：

$$\begin{aligned} \delta_1(i) &= \pi_i b_i(o_1), \quad i = 1, 2, \dots, N \\ \psi_1(i) &= 0, \quad i = 1, 2, \dots, N \end{aligned}$$

递推：对 $t = 2, 3, \dots, T$

$$\begin{aligned} \delta_t(i) &= \max_{1 \leq j \leq N} [\delta_{t-1}(j) a_{ji}] b_i(o_t), \quad i = 1, 2, \dots, N \\ \psi_t(i) &= \arg \max_{1 \leq j \leq N} [\delta_{t-1}(j) a_{ji}], \quad i = 1, 2, \dots, N \end{aligned}$$

终止：

$$\begin{aligned} P^* &= \max_{1 \leq i \leq N} \delta_T(i) \\ i_T^* &= \arg \max_{1 \leq i \leq N} \delta_T(i) \end{aligned}$$

最优路径：

$$i_t^* = \psi_{t+1}(i_{t+1}^*), \quad I^* = (i_1^*, i_2^*, \dots, i_T^*)$$

2.3 高斯隐马尔可夫模型的参数估计

考虑到股票的相关数据是连续变动的量，且其能够用混合高斯分布进行较好的拟合，在训练隐马尔可夫模型时，我们认为，假设每个隐状态下观测数据的发射概率服从高斯分布是一种较为合理的做法。本节主要介绍当发射概率服从高斯分布时，使用 Baum-Welch 算法进行参数估计的一些技术细节。

对于单特征隐马尔可夫模型，高斯分布的参数包括：

$$\mu = [\mu_1, \mu_2, \dots, \mu_N], \quad \sigma = [\sigma_1, \sigma_2, \dots, \sigma_N]$$

在设置初始值 $\mu^{(0)}$, $\sigma^{(0)}$ 时，我们使用聚类方法得到一组较好的估计。

时刻 t 观测序列处于状态 k 的概率密度函数：

$$b_k(o_t) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left\{-\frac{(o_t - \mu_k)^2}{2\sigma_k^2}\right\}$$

则有

$$\log b_k(o_t) = -\log \sqrt{2\pi} - \log \sigma_k - \frac{(o_t - \mu_k)^2}{2\sigma_k^2}$$

上节中 Q 函数的第三项可以写为：

$$\begin{aligned} & \sum_{I=1}^T \left(\sum_{t=1}^T \log b_{i_t}(o_t) \right) P(O, I | \bar{\lambda}) \\ &= \sum_{k=1}^N \sum_{t=1}^T \log b_k(o_t) P(O, i_t = k | \bar{\lambda}) \\ &= \sum_{k=1}^N \sum_{t=1}^T \left(-\log \sqrt{2\pi} - \log \sigma_k - \frac{(o_t - \mu_k)^2}{2\sigma_k^2} \right) P(O, i_t = k | \bar{\lambda}) \end{aligned}$$

分别对 σ, μ 求导并令倒数为零，可以得到：

$$\begin{aligned} \mu_k &= \frac{\sum_{t=1}^T o_t P(O, i_t = k | \bar{\lambda})}{\sum_{t=1}^T P(O, i_t = k | \bar{\lambda})} = \frac{\sum_{t=1}^T o_t \gamma_t(k)}{\sum_{t=1}^T \gamma_t(k)} \\ \sigma_k^2 &= \frac{\sum_{t=1}^T (o_t - \mu_k)^2 P(O, i_t = k | \bar{\lambda})}{\sum_{t=1}^T P(O, i_t = k | \bar{\lambda})} = \frac{\sum_{t=1}^T (o_t - \mu_k)^2 \gamma_t(k)}{\sum_{t=1}^T \gamma_t(k)} \end{aligned}$$

对于多特征隐马尔可夫模型，设特征数为 m ，多元高斯分布的参数包括：

$$\begin{aligned} \mu &= [\mu_1, \mu_2, \dots, \mu_N], \quad (\mu_i = [\mu_{i1}, \mu_{i2}, \dots, \mu_{im}]) \\ \Sigma &= [\Sigma_1, \Sigma_2, \dots, \Sigma_N] \end{aligned}$$

时刻 t 观测序列处于状态 k 的概率密度函数：

$$b_k(o_t) = \frac{1}{(2\pi)^{D/2} |\Sigma_k|^{1/2}} \exp\left\{-\frac{1}{2}(o_t - \mu_k)^T \Sigma_k^{-1} (o_t - \mu_k)\right\}$$

与单特征模型类似，得到：

$$\mu_k = \frac{\sum_{t=1}^T o_t \gamma_t(k)}{\sum_{t=1}^T \gamma_t(k)}, \quad \Sigma_k = \frac{\sum_{t=1}^T (o_t - \mu_k)^T (o_t - \mu_k) \gamma_t(k)}{\sum_{t=1}^T \gamma_t(k)}$$

第三章 贝叶斯先验估计及凸优化算法

3.1 贝叶斯估计原理综述

在统计学领域存在两大经典学派：经典学派与贝叶斯学派。前者认为与随机变量分布有关的参数是一组确定的常数，可以通过观察总体信息与样本信息得到；而贝叶斯学派则认为，这组参数 θ 也是一个随机变量，其分布可以用 $\pi(\theta)$ 表示，而研究对象 $f(x; \theta)$ 的分布可以看做条件分布，用 $f(x|\theta)$ 表示。

在对参数进行统计推断时，贝叶斯学派不仅利用了总体信息和样本信息，而且利用了关于参数 θ 的信息 $\pi(\theta)$ ，称为先验信息。

以下简单介绍这三种信息：

(1) 总体信息。总体信息是有关总体分布的信息，例如“总体是正太分布的”。总体信息十分重要，在进行假设时，需要根据参数的具体需要进行分布假设。

(2) 样本信息。样本信息是抽取样本取得观测值得到的信息。它作为总体的代表，含有统计推断的重要内容。

(3) 先验信息。在实际问题中，在抽取样本值之前，人们对参数已经有了一些了解和估计，这些信息统称为先验信息。

贝叶斯学派认为，建立统计推断的最终效果，是把对参数 θ 的认识由 $\pi(\theta)$ 调整到

$$h(\theta | x_1, x_2, \dots, x_n) = \frac{\pi(\theta) \prod_{i=1}^n f(x_i; \theta)}{\int_{\theta} \pi(\theta) \prod_{i=1}^n f(x_i; \theta) d\theta}$$

并且对参数的统计推断应该建立于后验分布上。

3.2 贝叶斯估计与优化算法

本文对股票数据进行了多次隐马尔可夫训练，并对由 Baum-Welch 算法进行参数估计得到的参数 $\lambda = (A, B, \pi)$ 进行了分析。观察发现，由于选取的训练数据集较长，参数 B 中基本覆盖了所有可能发生的观测值，在训练过程中引入新数据不会造成参数 B 发生剧烈变化。对于初始分布的参数 π ，因为测试数据与训练数据是连续发生的，在完成对后一交易日状态的估计时，不涉及到该参数。然而，对于转移概率矩阵 A ，本文认为仅仅采用 Baum-Welch 算法得到的转移概率矩阵缺乏时效性特点，无法有效的代表当时的状态转移情况。考虑到国内市场市场状态的频繁转换与其本身的不稳定性，本文利用贝叶斯估计，使得转移概率矩阵随着新数据的加入不断完成自我优化。

进行贝叶斯估计时，需要特别考虑以下几点要求：

- (1) 引入新数据进行贝叶斯估计时，不应应对原有数据进行重复训练。
- (2) 为了确保转移概率矩阵的鲁棒性，其应当对异常数据具有一定的承受能力。
- (3) 算法应具有较低复杂度。

3.2.1 贝叶斯先验假设

针对股票数据的具体问题，首先给出几个关于贝叶斯估计的基本假设：

(1) 选取的训练数据集足够长，使得参数 B 基本覆盖了所有可能发生的状态，因此在贝叶斯估计时只需完成对参数 A 的优化。

(2) 状态转移矩阵 A 具有时效性特点，与当前市场状态密切相关。

(3) 状态转移矩阵 A 的每一行均服从 Dirichlet 分布，其相关先验信息即 Dirichlet 分布的参数 $\delta_i = [\delta_{i1}, \delta_{i2}, \dots, \delta_{iN}]$ ， $i = 1, 2, \dots, N$ 由过去训练数据得到。

(4) 在历史过程中进行参数估计得到的相关参数 $\lambda_t = (A_t, B, \lambda)$ 总是符合当时最优的。

(5) 考虑到股票数据的连续性以及高斯分布假设，认为任意两状态之间总是互通的，即 $a_{ij} > 0, (\forall i, j)$ 。

(6) 当前数据更重要，随时间距离增大，数据的重要性减弱。

3.2.2 目标似然函数构造

在引入新数据点 y 后，目标似然函数可以写为

$$P(y, A | \Delta = (\delta_{ij})_{N \times N}, l, B = (\mu, \Sigma))$$

其中， y 表示新加入的观测值， A 是统计推断的目标——转移概率矩阵， Δ 代表转移概率矩阵 Dirichlet 先验分布假设的相关参数， l 表示由过去数据进行预测（解码）得到的前一交易日的状态分布密度函数向量， $l = [l_1, l_2, \dots, l_N]$ 。

目标似然函数可以写为：

$$\begin{aligned} P(y, A | \Delta = (\delta_{ij})_{N \times N}, l, B = (\mu, \Sigma)) \\ &= P(A | B, l, \Delta) P(y | A, B, l, \Delta) \\ &= P(A | \Delta) P(y | A, B, l) \end{aligned}$$

其中认为 A 仅与先验参数 Δ 相关，与 B, l 不存在直接关系； y 发生的概率则与之前的参数估计结果 Δ 无关，取决于 A, B, l 。

进一步考虑数据点 y 对应的隐状态 s ：

$$\begin{aligned} P(A | \Delta) P(y | A, B, l) \\ &= P(A | \Delta) \sum_{i=1}^N P(y, s=i | A, B, l) \\ &= P(A | \Delta) \sum_{i=1}^N P(y | s=i, A, B, l) P(s=i | A, B, l) \\ &= P(A | \Delta) \sum_{i=1}^N P(y | s=i, B) P(s=i | A, l) \end{aligned}$$

$$\text{Dirichlet 分布下的概率密度函数 } Dir(a_i | \delta_i) = \frac{\Gamma(\sum_j \delta_{ij})}{\prod_j \Gamma(\delta_{ij})} \prod_j a_{ij}^{\delta_{ij}-1} = \frac{1}{B(\delta_i)} \prod_j a_{ij}^{\delta_{ij}-1}。$$

记在参数 B 下， y 对于状态 i 的概率密度函数为 $b_i(y)$ 。

不妨将概率函数统一写为概率密度函数，

$$\begin{aligned} p(A | \Delta) p(y | A, B, l) \\ &= p(A | \Delta) \sum_{i=1}^N p(y | s=i, B) p(s=i | A, l) \\ &= p(A | \Delta) \sum_{i=1}^N \left(b_i(y) \sum_{j=1}^N l_j a_{ji} \right) \\ &\propto \prod_{i,j=1}^N a_{ij}^{\delta_{ij}-1} \left(\sum_{i,j=1}^N l_i a_{ij} b_j(y) \right) \\ &\triangleq f(A) \end{aligned}$$

可以得到对数似然函数

$$\log f(A) = \sum_{i,j=1}^N (\delta_{ij} - 1) \log a_{ij} + \log \left(\sum_{i,j=1}^N l_i a_{ij} b_j(y) \right)$$

3.2.3 约束优化算法设计

以下利用拉格朗日乘数法完成约束优化，首先写出目标函数

$$\begin{aligned} & \max \{ \log f(A) \} \\ &= \min \{ -\log f(A) \} \\ &= \min \left\{ -\sum_{i,j=1}^N (\delta_{ij} - 1) \log a_{ij} - \log \left(\sum_{i,j=1}^N l_i a_{ij} b_j(y) \right) \right\} \end{aligned}$$

约束条件包括：

(1)

$$0 < a_{ij} < 1, \quad (\forall i, j)$$

(2)

$$\sum_j a_{ij} = 1, \quad (\forall i)$$

整理得到：

$$\begin{aligned} & \min \left\{ -\sum_{i,j=1}^N (\delta_{ij} - 1) \log a_{ij} - \log \left(\sum_{i,j=1}^N l_i a_{ij} b_j(y) \right) \right\} \\ & \quad s.t. \begin{cases} \sum_j a_{ij} - 1 = 0, & (\forall i) \\ -a_{ij} + \varepsilon \leq 0, & (\forall i, j) \end{cases} \end{aligned}$$

其中 ε 是一个极小常数，使约束优化可以达到边界条件。

引入拉格朗日乘子 λ_i, μ_{ij} 和松弛变量 η_{ij} ，写出拉格朗日函数：

$$\begin{aligned} L(A, \lambda_i, \mu_{ij}, \eta_{ij}) &= -\sum_{i,j=1}^N (\delta_{ij} - 1) \log a_{ij} - \log \left(\sum_{i,j=1}^N l_i a_{ij} b_j(y) \right) \\ &\quad + \sum_{i=1}^N \lambda_i \left(\sum_{j=1}^N a_{ij} - 1 \right) + \sum_{i,j=0}^N \mu_{ij} (-a_{ij} + \varepsilon + \eta_{ij}^2) \end{aligned}$$

对各变量分别求偏导得到，

$$\begin{cases} \frac{\partial L}{\partial a_{ij}} = -\frac{1}{a_{ij}} (\delta_{ij} - 1) - \frac{l_i b_j(y)}{\sum_{i,j=1}^N l_i a_{ij} b_j(y)} + \lambda_i - \mu_{ij} = 0 \\ \frac{\partial L}{\partial \mu_{ij}} = -a_{ij} + \varepsilon + \eta_{ij}^2 = 0 \\ \frac{\partial L}{\partial \eta_{ij}} = 2\eta_{ij} \mu_{ij} = 0 \\ \mu_{ij} \geq 0, \quad (\forall i, j) \end{cases}$$

化简得到不等式约束优化的 KKT 条件，

$$\begin{cases} -\frac{1}{a_{ij}}(\delta_{ij}-1) - \frac{l_i b_j(y)}{\sum_{i,j=1}^N l_i a_{ij} b_j(y)} + \lambda_i - \mu_{ij} = 0 \\ (-a_{ij} + \varepsilon)\mu_{ij} = 0 \\ \mu_{ij} \geq 0 \end{cases} \quad (\forall i, j)$$

当不等式约束不起作用即 $\mu_{ij} = 0$ 时，有

$$a_{ij} = (1 - \delta_{ij}) / (1 - 1/N - \frac{l_i \sum_j b_j(y) / N - l_i b_j(y)}{\sum_{i,j} l_i a_{ij} b_j(y)})$$

通过迭代可以求出贝叶斯估计的数值解。

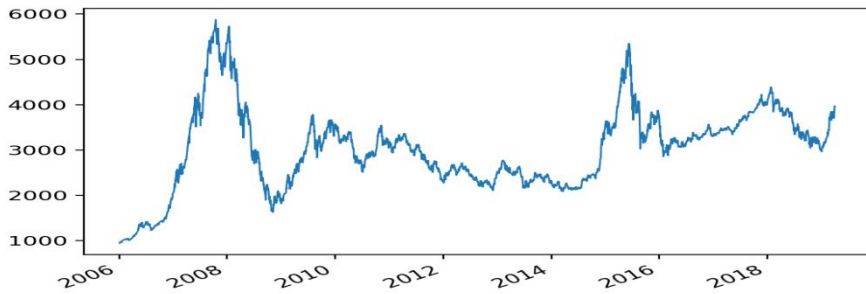
记拉格朗日约束优化得到的转移概率矩阵为 A' ，为了增强优化算法的鲁棒性，降低异常值的影响，每次优化后，我们令

$$A_{t+1} = (1 - \alpha)A_t + \alpha A'_t, \quad (0 \leq \alpha \leq 1)$$

其中 α 是一个可调节的参数， α 越小，模型对异常值的耐受性越高，但是优化能力也随之减弱， $\alpha=0$ 时模型无贝叶斯优化效果。

3.3 优化效果检验

贝叶斯估计的目的是使转移概率矩阵具有时效性，因此，在检验优化效果时，可以选取市场整体状态发生明显变化的前后时段进行测试与观察。为了便于观察，我们选择了与市场状态具有高度相关性的每日收益率指标作为模型的输入数据。



观察发现，沪深 300 指数在 2007 年到 2009 年经历了一次由牛市到熊市的剧烈转变，可以粗略的认为 2008 年 1 月 1 日为转变发生的时刻。我们将每日收益率指标按区间进行划分，2007 年 1 月 1 日至 2008 年 1 月 1 日为训练集，我们在这段时间序列上对隐马尔可夫模型进行训练，并通过 Baum-Welch 算法得到对应的参数估计。2008 年 1 月 1 日至 2009 年 1 月 1 日为优化测试集，观察优化前后的转移概率矩阵与相关参数：

模型各状态对应高斯分布的参数

$$\mu = [0.0095 \quad 0.0120 \quad -0.0119], \quad \sigma = [0.00029 \quad 0.00019 \quad 0.00089]$$

对应的三个状态不妨记作，小涨、大涨、下跌。

在训练集（2007 年 1 月 1 日至 2008 年 1 月 1 日）上完成训练后，状态转移矩阵的估计

$$A_0 = \begin{bmatrix} 0.334 & 0.376 & 0.290 \\ 0.340 & 0.392 & 0.268 \\ 0.333 & 0.352 & 0.315 \end{bmatrix}$$

在优化测试集上进行优化后，得到

$$A' = \begin{bmatrix} 0.320 & 0.342 & 0.338 \\ 0.326 & 0.332 & 0.342 \\ 0.329 & 0.331 & 0.340 \end{bmatrix}$$

观察发现，经过优化，市场状态保持上涨的平均概率由原先的 0.721 降低到了 0.660，降低了约 8.46%；市场状态保持下跌的概率由原先的 0.315 增加到了 0.340，增加了约 7.94%；市场状态由上涨转移到下跌的平均概率由原先的 0.279 增加到了 0.340，增加了约 21.86%；市场状态由下跌转移到上涨的平均概率由原先的 0.3425 降低到了 0.330，降低了约 3.65%。以上数据均符合市场状态由牛市转变为熊市的预期结果。

为了更好的检验贝叶斯方法对于市场状态在牛市、熊市、震荡市之间相互转换的优化效果，我们定义隐马尔可夫模型的适应度函数

$$fitness(\lambda) = \frac{\sum_{t=1}^T \log P(o_t | \lambda)}{T}$$

分别计算优化前后在转移概率矩阵下，模型在测试集上的适应度。特别地，观察适应度函数的相对值可以判断参数选择的优劣，但一般认为适应度函数本身绝对值不具有讨论价值。

编号	市场状态转换	时间			fitness		fitness 变化率
		开始	转换	结束	优化前	优化后	
1	牛市->熊市	2007/1/1	2008/1/1	2009/1/1	1.89	1.93	2.1%
2	牛市->震荡	2008/10/1	2009/8/1	2010/8/1	2.508	2.509	0.04%
3	牛市->熊市	2015/1/1	2015/6/1	2015/10/1	1.685	1.837	9.02%
4	熊市->震荡	2015/5/1	2015/8/1	2016/4/1	2.239	2.238	-0.04%

观察适应度函数在优化前后的值，可以发现，对沪深 300 指数上的每日收益率进行贝叶斯估计的参数优化，对于牛市熊市相互转换的情况能够大幅提高模型适应度，使转移概率矩阵迅速逼近当前最优值，对于震荡市和其他市场状态之间的转换，贝叶斯估计的优化效果不明显。

第四章 单特征模型训练及回测结果

傅中杰等人（2018）在利用 HMM 模型设计量化择时策略时，引入了特征筛选这一方法，选择股票指标取代股价原始数据代入隐马尔可夫模型进行训练，完成策略回测并对各指标进行筛选，最后选择表现良好的指标组成综合模型的输入。本文在他们研究的基础上引入了贝叶斯优化方法与交叉验证的思想，对比了贝叶斯优化前后的回测结果并进行了分析。

本章主要介绍了使用沪深 300 指数的单个指标作为训练对象，并利用隐马尔可夫模型建立择时策略的回测并观察其结果，同时，对于贝叶斯优化前后的择时模型，我们比较了二者的回测结果，发现贝叶斯优化方法在实证分析中表现出了良好的性质。在下一章节中将介绍综合模型及其在不同投资标上进行回测的结果。

为了使模型更具有现实可操作性且能够保持较稳定的盈利能力，本文使用了交叉验证的思想，将全部数据空间（2006 年 1 月 1 日至 2019 年 4 月 10 日）划分为训练集（2006 年 1 月 1 日至 2015 年 1 月 1 日）、评估集（2015 年 1 月 1 日至 2018 年 1 月 1 日）和测试集（2018 年 1 月 1 日至 2019 年 4 月 10 日）。本章单特征模型在训练集上进行训练，并且在评估集上完成回测。

4.1 模型训练结果及分析

首先进行训练及回测数据的准备。传统基于隐马尔可夫的量化模型往往选择日内开盘价、最高价、最低价、收盘价组成含有四个元素的向量，以此为输入对隐马尔可夫模型进行训练。由于日内开盘价、最高价、最低价、收盘价这四个指标基本上能够代表股票市场当天的信息，因此该模型是有效的。然而，我国股票市场具有其特殊的性质：高波动、频繁的状态转换以及高噪声。因此，对原始数据不加处理而作为输入难以最大化利用隐马尔可夫模型识别市场状态的能力。本文通过计算得到了一些股市的常用指标，经过处理的指标具有低噪声、较低波动的特点，能够较好的反映股票市场的真实状态。

在选择指标时，有以下几个基本原则：

- （1）指标可以被量化，且数据是连续的。
- （2）指标数据序列具有一定的波动性，且在短期会发生变化。
- （3）指标数据是有范围的，即未来的数据总是在过去发生过的。

按以上原则在常用股价指标中进行筛选，得到了 15 个指标，分别是：

每日收益率	20 日均线	20 日波动率	20 日乖离率	5 日均线
5 日波动率	5 日乖离率	14 日 ATR	长短期均线差值	MACD
6 日 RSI	12 日 RSI	BR	AR	交易量

可以看到，以上指标既包括长期指标（一个月左右），也包含短期指标（一天或一周），因此其可以基本代表所有股价信息。

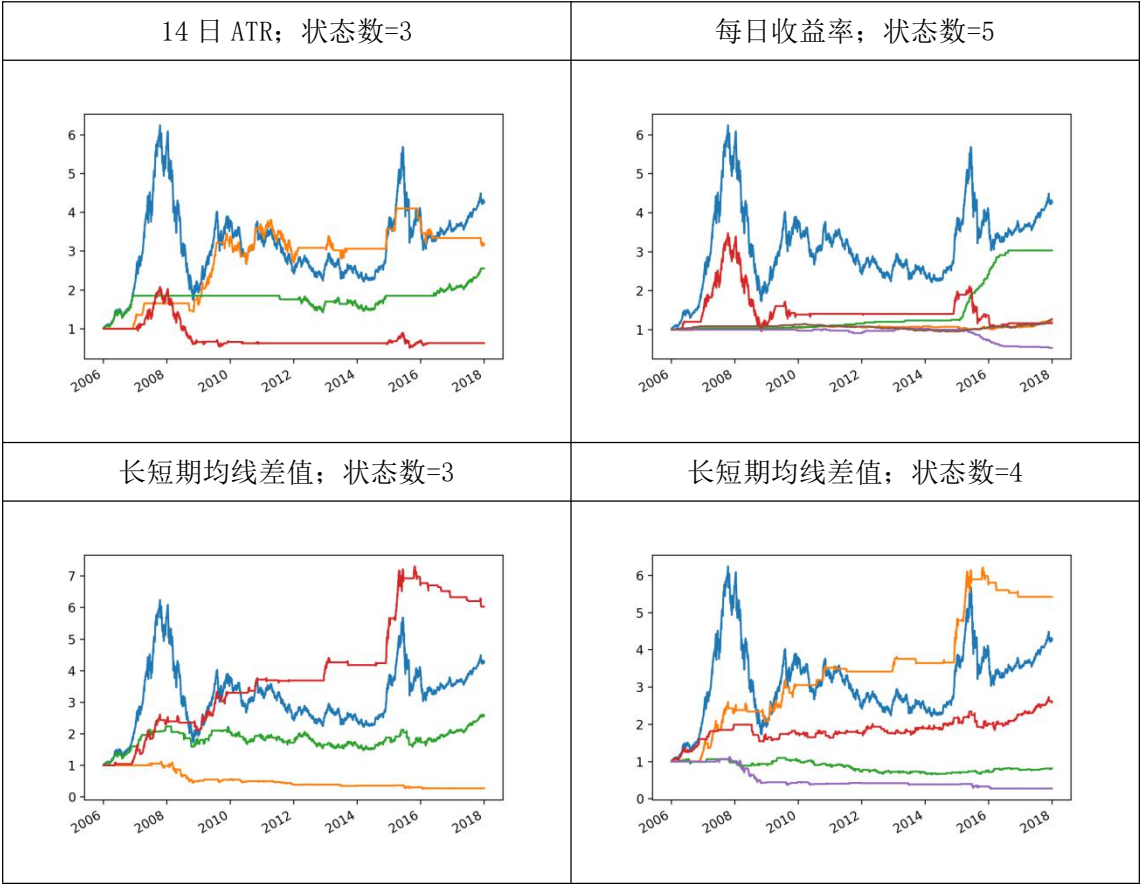
在未知其他信息的情况下，很难进行合适的状态数选择，因此对于状态数，我们选取了在区间 $[3, 9]$ 之内的所有整数分别进行训练，并观察训练结果。

对各项指标分别进行单特征模型训练：

图表中，蓝色曲线代表沪深 300 指数收盘价，其他颜色曲线分别表示经过解码得到的隐状态序列。在解码过程中，本文比较了 Viterbi 算法与近似算法，发现两者的计算结果没有明显差距，但近似算法的计算复杂度更低。因此，在本文研究中，使用近似算法完成对观测序列的解码。

图中彩色曲线分别表示坚持在某状态下持有股票，得到的历史净值曲线，在此过程中不考虑交易成本的影响。

观察图表可以发现，同一隐状态在不同时间阶段的盈利能力是不同的。指标数据在不同市场状态下具有不同的隐含意义，例如：较低的波动率在牛市代表稳定上涨的趋势，在震荡市则代表该时间段内股价变动不剧烈，缺少短线交易的盈利机会。因此，在设计择时策略时，需要观察某状态的近期盈利能力而非简单的对指标值进行预测与分析。



4.2 交易策略设计及实现

本文交易策略设计的基本思想是：预测下一交易日的市场状态，判断该状态是否具有盈利能力，若具有盈利能力，则在下一交易日持有该股票，否则不持有。

首先通过对某隐状态过去一段时间（记为验证时段）的表现进行分析，判断该隐状态在近期是否具有盈利能力。

判断标准：

- （1）在验证时段内，该隐状态至少出现过一次。
- （2）在验证时段内，该隐状态的累积收益率为正。

如果某状态满足以上两个条件，那么记这个状态为盈利状态。

在确定验证时段区间时，本文选择对过去 20 个交易日即一个月的情况进行判断与分析。实际应用中，如果希望提高择时策略的交易频率，或者增强其对市场状态变化的敏感度，可以适当减小验证时段的长度。

在预测下一交易日的市场状态时，传统方法一般选取发生概率最大的状态作为预测结果。然而，这种做法一定程度上浪费了发生概率向量所隐含的重要信息，即状态发生的可能性。为了强化状态发生的相对可能性对预测的影响，本文计算下一交易日市场状态为盈利状

态的概率，若该概率大于阈值，则承认下一交易日为盈利状态，否则不承认。

本文设定阈值为 0.6，实际应用中可以根据投资者的风险偏好调整该阈值，风险厌恶指数越高，阈值设定越大。

4.3 基本回测结果

对于单指标的择时模型，在评估集（2015 年 1 月 1 日至 2018 年 1 月 1 日）上完成策略回测，图表中展示了多个指标在不同状态数下的回测结果。

观察发现，同一指标在不同状态数下的回测表现差距较大，也就是说该隐马尔可夫的量化择时模型对状态数选择具有较强的依赖性。同时，不同指标的回测结果也有较大差距，这一点与预期相符。



我们在评估集上对回测结果的各项常用评价指标进行了计算，包括超额收益率，交易频率，最大回撤等。不同指标与状态数下择时策略的回测表现各不相同。

指标	状态数	开仓（次）	平仓（次）	交易频率（次/年）	胜率（%）	年化收益率（%）	年化超额收益率（%）	最大回撤（%）	夏普比率
14 日 ATR	3	32	31	10.76	62.50	13.00	11.04	16.58	0.65
14 日 ATR	4	41	41	14.00	51.22	6.95	3.78	20.68	0.26
14 日 ATR	8	46	46	15.71	56.52	12.79	10.77	8.05	0.64

BR	3	51	50	17.25	58.82	5.65	2.31	23.17	0.17
BR	7	76	76	25.96	51.32	6.45	3.20	17.84	0.29
BR	9	91	91	31.08	58.24	9.62	6.89	11.98	0.58
每日收益率	3	40	40	13.66	50.00	-4.65	-8.09	37.55	-0.43
每日收益率	5	23	23	7.86	69.57	17.54	17.01	3.55	1.64
每日收益率	6	19	19	6.49	52.63	0.85	-2.79	8.18	-0.36

观察发现，不同状态数下，同一指标的择时策略回测表现各不相同。同时，不同指标的回测表现各有其特色与优缺点，例如：状态数为3时，14日ATR能够得到较高的年化超额收益率和交易胜率、以及合适的交易频率，但是其最大回撤超过了15%，交易风险较高；状态数为9时，BR指标有较高的超额收益率和合理的最大回撤，但是其交易频率过高，容易带来额外的交易成本。

4.4 加权评分结果

4.4.1 层次分析法原理及权重设计

层次分析法（AHP）最早由美国运筹学家萨蒂于上世纪七十年代提出，用于解决美国工业部门电力分配问题。层次分析法应用了网络系统理论和多目标综合评价方法，将定性与定量结合起来，是一种层次权重决策分析方法。该方法利用决策者的经验各衡量目标的相对重要程度，并以此为依据给出每个决策方案每个标准的权重，定量的给出各方案的优劣次序，从而有效应用于难以定量解决的决策问题。

建立层次分析模型时，我们首先按标准设定模型的目标层、指标层、以及方案层。在本文中，目标是为隐马尔可夫模型选择合适的输入特征，对于判断指标，我们选择年化收益率、最大回撤、夏普比率、胜率、以及交易频率这五个常用评价指标，方案层表示各个股市指标（包括每日收益率、波动率等）进行排列组合的结果。

在合适的评价指标时，我们考虑到衡量投资表现的三个基本因素：

（1）卓越的收益能力。包括投资组合获取绝对收益与超额收益的能力，通常用年化收益率与年化超额收益率表示。

（2）较低的投资风险。优秀的策略应当能够规避市场环境带来的风险，这一抗风险能力一般用最大回撤衡量。

（3）合适的交易成本。从长期看，随着交易成本的累积，其可能会侵蚀部分利润，因此交易胜利与交易频率需要被纳入考虑范围。

本文根据以上五个评价指标（年化收益率、最大回撤、夏普比率、胜率、交易频率）对各市场指标作为输入的隐马尔可夫模型在评估集上进行回测的结果进行综合打分。根据经验，我们可以得到各评价指标的两两相对重要程度，即可得到层次分析法的比较矩阵 C 。

$$C = \begin{bmatrix} 1 & 6/5 & 3/2 & 2 & 3 \\ 5/6 & 1 & 5/4 & 5/3 & 5/2 \\ 2/3 & 4/5 & 1 & 4/3 & 2 \\ 1/2 & 3/5 & 3/4 & 1 & 3/2 \\ 1/3 & 2/5 & 1/2 & 2/3 & 1 \end{bmatrix}$$

对 C 进行特征向量分析，最大特征值对应的特征向量进行归一化即得到每个评价指标对应的权重，权重向量为： $w=[0.3 \quad 0.25 \quad 0.2 \quad 0.15 \quad 0.1]$ 。

4.4.2 系数归一化与加权评分结果

在加权求和之前，我们需要对各评价指标进行简单的预处理：包括反向指标取负号以及系数归一化。在本文研究涉及的五个评价指标（年化收益率、最大回撤、夏普比率、胜率、交易频率）中，最大回撤与交易频率是反向指标，需要取负号。夏普比率的取值范围在（-10，10）区间内，其他评价指标的取值范围在（0，100）区间内。对各指标进行最大最小标准化处理，使他们不带有额外的权重影响，并加权求和，得到各市场指标回测结果的评分。

下表展示了各市场指标在选取最优状态数时，在评估集（2015年1月1日至2018年1月1日）上的回测表现及得分：

指标	状态数	开仓 (次)	平仓 (次)	交易频率 (次/年)	胜率 (%)	年化收益率 (%)	年化超额收益率 (%)	最大回撤 (%)	夏普比率	评分
每日收益率	5	23	23	7.86	69.57	17.54	17.01	3.55	1.64	46.83
20日乖离率	7	54	53	18.27	50.00	15.18	13.86	10.62	1.04	28.38
5日波动率	7	57	56	19.30	49.12	13.27	11.39	9.02	0.98	26.81
20日均线	4	29	28	9.73	62.07	12.92	10.94	15.18	0.67	21.87
14日ATR	8	46	46	15.71	56.52	12.79	10.77	8.05	0.64	21.59
AR	6	66	65	22.37	50.00	13.06	11.12	11.33	0.72	20.82
长短期均线 差值	3	45	44	15.20	55.56	13.27	11.39	15.79	0.69	20.63
交易量	6	59	59	20.15	54.24	13.92	12.21	16.97	0.73	20.62
BR	9	91	91	31.08	58.24	9.62	6.89	11.98	0.58	17.11
5日均线	4	33	32	11.10	60.61	10.24	7.64	19.33	0.46	15.45
12日RSI	9	80	80	27.32	58.75	7.47	4.36	9.84	0.42	14.22
6日RSI	9	67	67	22.88	58.21	4.32	0.85	9.64	0.14	8.18
MACD	5	58	58	19.81	48.28	4.31	0.84	30.31	0.08	0.61
5日乖离率	3	24	24	8.20	41.67	0.60	-3.04	21.85	-0.15	-2.92
20日波动率	3	12	12	4.10	33.33	-2.01	-5.62	11.04	-0.69	-12.62

第五章 多特征综合模型回测及结果分析

5.1 综合模型在评估集的回测结果

在筛选市场指标时，我们选择回测结果评分超过 20 分且交易胜率不小于 50% 的市场指标，得到 7 个有效的市场指标，分别是：每日收益率、20 日乖离率、20 日均线、14 日 ATR、AR、长短期均线差值、交易量，组合成为多维特征向量，作为综合模型的输入。综合模型在评估集的回测结果如下：

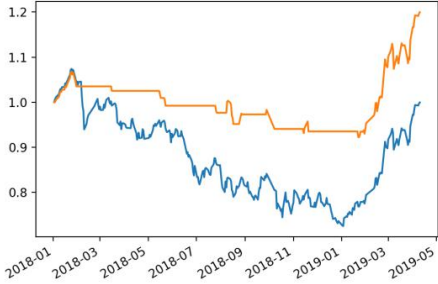
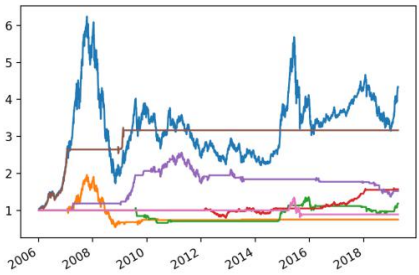
状态数	开仓 (次)	平仓 (次)	交易频率 (次/年)	胜率 (%)	年化收益率 (%)	年化超额收益率 (%)	最大回撤 (%)	夏普比率	评分
3	48	47	16.22	39.58	-1.03	-4.67	32.61	-0.24	-9.01
4	43	43	14.69	55.81	7.69	4.62	17.74	0.27	10.25
5	48	48	16.39	45.83	4.99	1.58	23.88	0.12	3.23
6	40	40	13.66	47.50	6.67	3.46	23.80	0.22	6.20
7	55	54	18.61	54.55	6.50	3.26	16.48	0.24	8.88
8	0	0	0.00	0.00	0.00	-3.65	0.00	/	/

观察发现，当状态数选择在 [4, 7] 之间时，综合模型可以得到较好的回测结果。

5.2 综合模型在测试集的表现

对于确定了有效市场指标的综合模型，我们在测试集上完成模型的回测，观察回测结果，发现其拥有较稳定的收益能力，较低的投资风险以及合适的交易成本。

图表展示了选择最优状态数（6）时的投资净值曲线以及相关评价指标。



状态数	开仓 (次)	平仓 (次)	交易频率 (次/年)	胜率 (%)	年化收益率 (%)	年化超额收益率 (%)	最大回撤 (%)	夏普比率	评分
6	18	17	14.20	50.00	15.92	16.23	13.63	1.00	27.35

第六章 结论与展望

PENDING...

参考文献

- [1] 傅中杰、吴清强. 基于隐马尔科夫模型的市场指数量化择时研究[J]. 厦门大学学报(自然科学版), 2018, 57(3).
- [2] Anastasios Petropoulos, Sotirios P. Chatzis, Stelios Xanthopoulos. A hidden Markov model with dependence jumps for predictive modeling of multidimensional time-series[J]. Information Sciences, 2017, 412(4):50-66.
- [3] 李少华. 基于隐马尔科夫模型的股指期货收益率波动预测[D]. 西安建筑科技大学硕士学位论文, 2018.
- [4] Thomas A. Dean, Sumeetpal S. Singh. Parameter Estimation for Hidden Markov Models with Intractable Likelihoods[J]. Scandinavian Journal of Statistics, 2014, 41(4):970-987.
- [5] Thomas A. Dean, Sumeetpal S. Singh. Parameter Estimation in Hidden Markov Models With Intractable Likelihoods Using Sequential Monte Carlo[J]. Journal of Computational and Graphical Statistics, 2015, 24(3):846-865
- [6] 吴漫君. 基于隐马尔科夫模型的股价走势预测[D]. 华南理工大学硕士学位论文, 2011.
- [7] Hamilton J D. A new approach to the economic analysis of nonstationary time series and the business cycle [J]. Journal of the Econometric Society, 1989, 57(21):357-384.
- [8] DE ANGELIS L, PAAS L J. A dynamic analysis of stock markets using a hidden Markov model[J]. Journal of Applied Statistics, 2013, 40(8):1682-1700.
- [9] SALHIK, DEACONU M, LEJAY A, et al. Regime switching model for financial data: empirical risk analysis[J]. Physica a- Statistical Mechanics and Its Applications, 2016, 461:148-157.
- [10] HASSAN R, NATH B. Stock market forecasting using hidden Markov model: a new approach[C]//5th International Conference on Intelligent Systems Design and Applications. Los Alamitos: IEEE, 2005:192-196.
- [11] 黄冉. 基于隐马尔科夫模型(HMM)的股票价格预测分析[D]. 青岛大学, 2015.