



北京理工大学  
BEIJING INSTITUTE OF TECHNOLOGY



# 基于隐式马尔科夫模型的证券指数 评估与预测

——本科毕业设计答辩

答辩人：江子昊

导师：马宏宾

## 论文摘要

## 主题

基于金融市场股票数据，对证券指数**隐状态**进行分析，并**预测**发展趋势

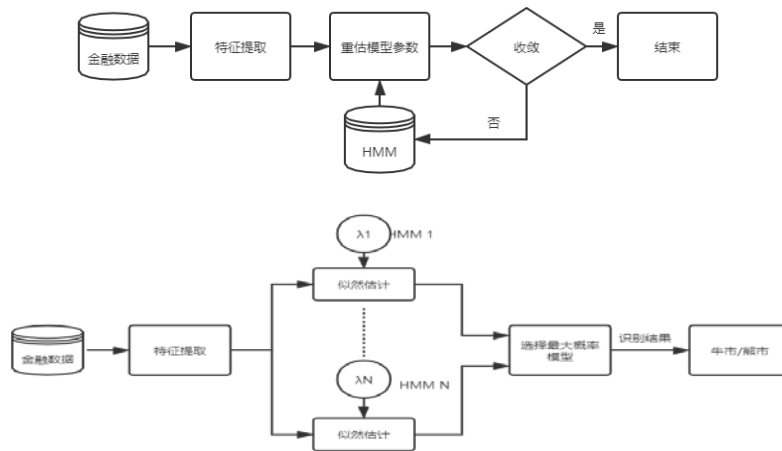
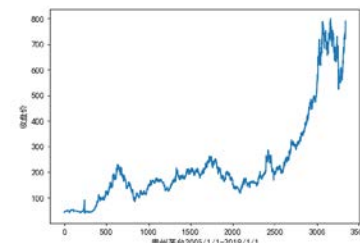
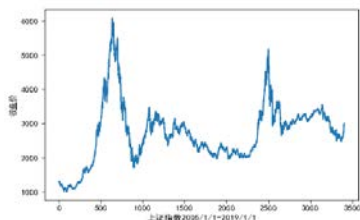
## 研究背景

## 理论基础

## 核心内容

## 成果展示

## 论文总结



论文摘要

研究背景

理论基础

核心内容

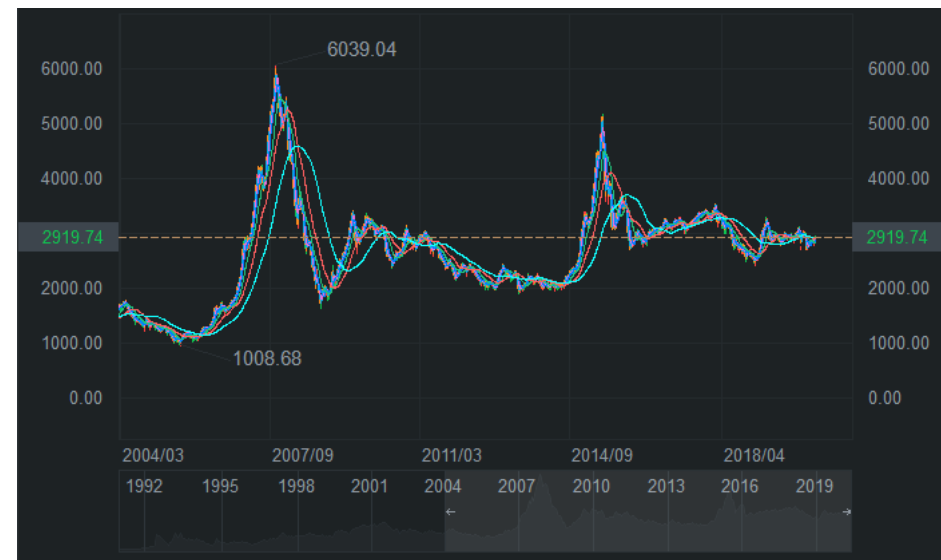
成果展示

论文总结

## 研究对象



a.道琼斯工业指数



b.上证指数

- 证券指数反应了经济形式与市场风险
- 本文以上证指数为研究对象

## 研究意义

论文摘要

研究背景

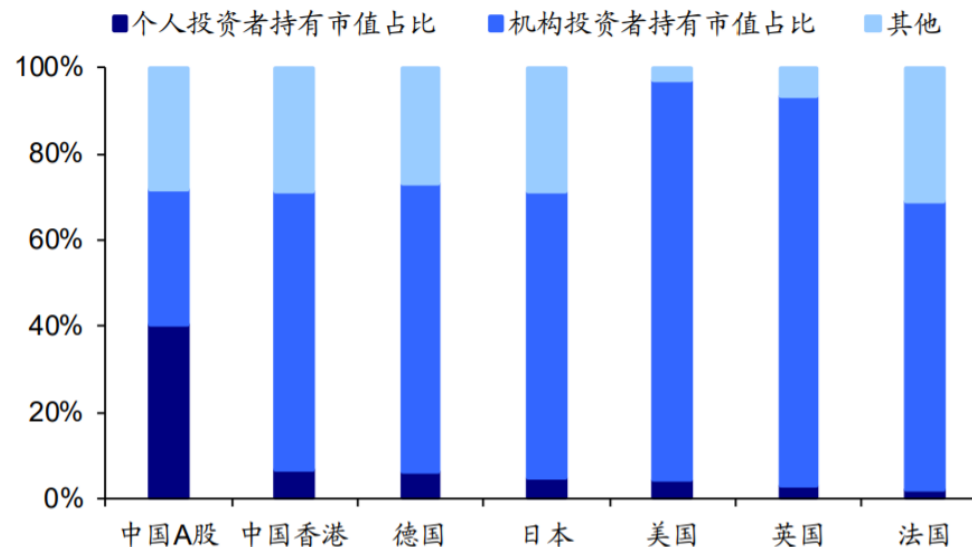
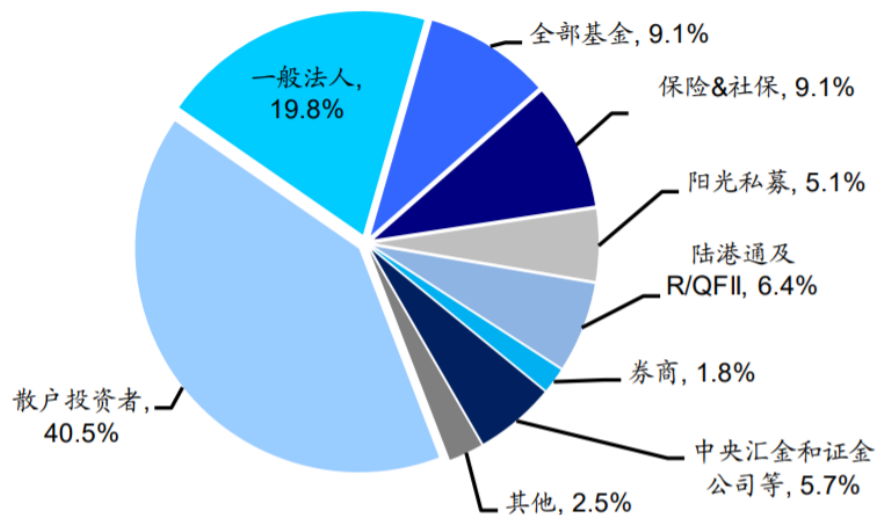
理论基础

核心内容

成果展示

论文总结

A股投资者自由流通市值占比（2018年中报）



- 对于个人，**规避投资风险，提高投资收益**
- 对于政府，**尽早发现其潜在的客观规律，从而及时做出判断和预警**

论文摘要

研究背景

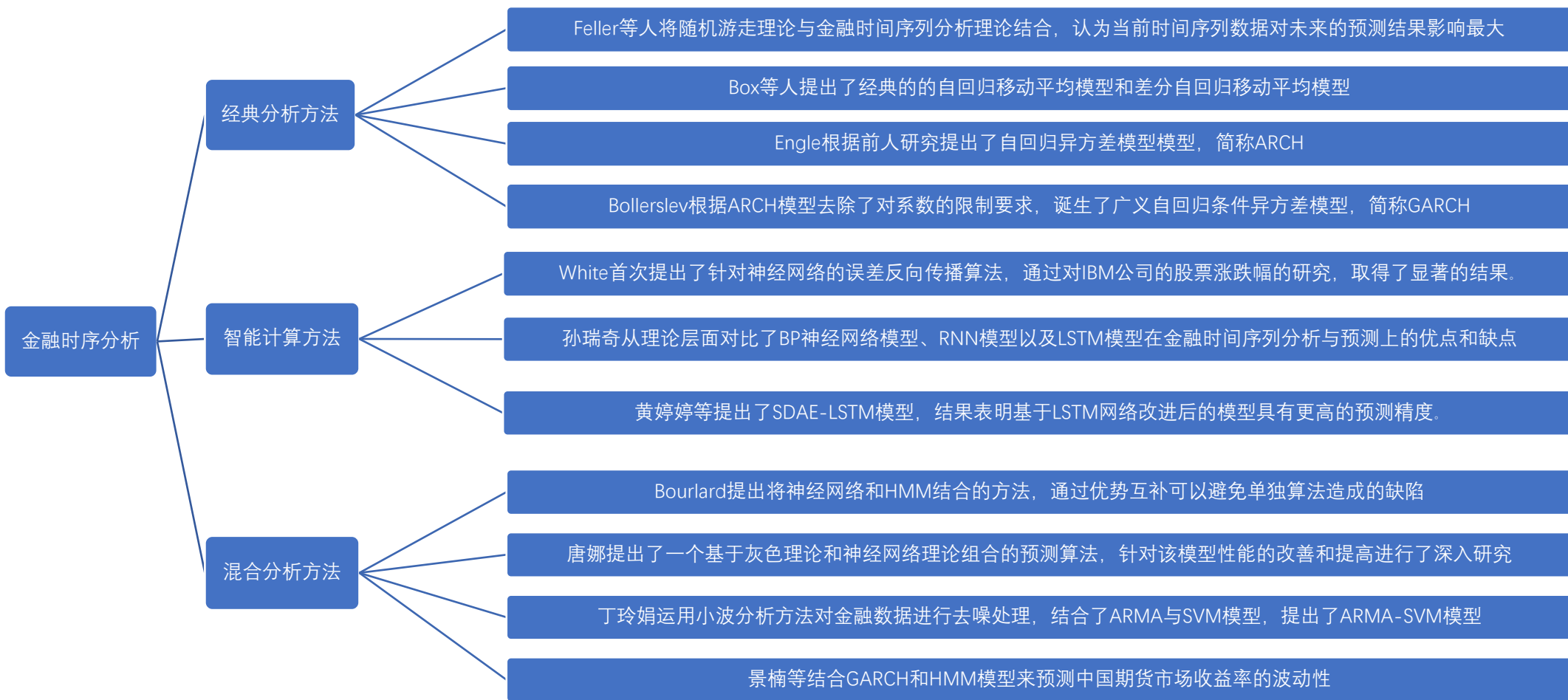
理论基础

核心内容

成果展示

论文总结

## 文献综述



## 相关理论基础

论文摘要

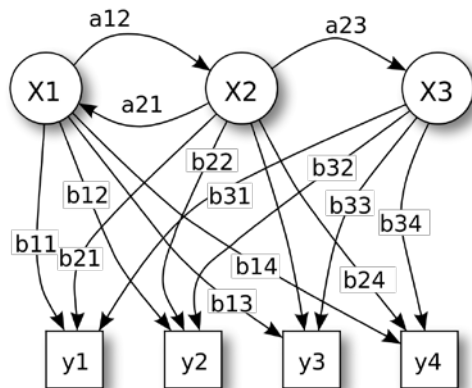
研究背景

理论基础

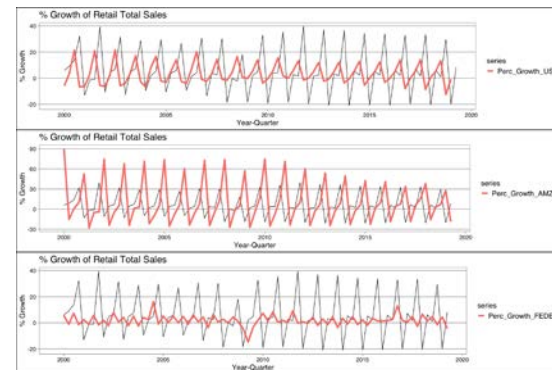
核心内容

成果展示

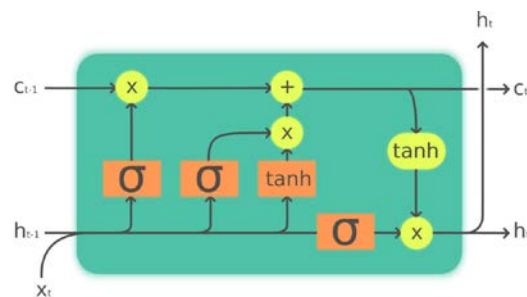
论文总结



a. 隐马尔科夫模型 (Hidden Markov Model)



b. 自回归滑动平均模型 (Autoregressive Moving Average Model)



Legend:

Layer

Pointwise op

Copy

c. 长短期记忆模型 (Long Short-Term Memory Model)

论文摘要

研究背景

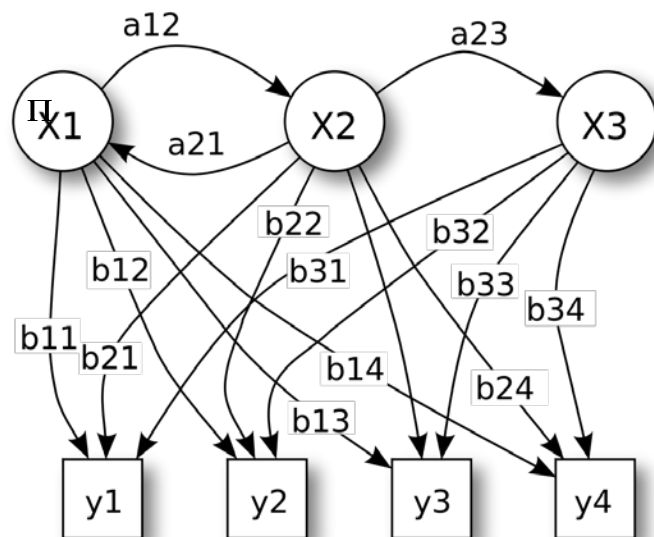
理论基础

核心内容

成果展示

论文总结

## 隐马尔科夫模型



隐马尔科夫过程

隐马尔科夫模型可以由一个五元组构成  
 $\lambda = (S, V, A, B, \Pi)$

- $S$ 表示模型中包含 $N$ 个隐藏状态的状态集合
- $V$ 表示模型中包含 $M$ 个观测结果的集合
- $A$ 为状态转移概率矩阵
- $B$ 为观测概率矩阵
- $\Pi$  为初始的状态概率分布

本研究中模型关键参数:

- 隐状态数目 $N$
- 状态转移概率矩阵 $A$

论文摘要

研究背景

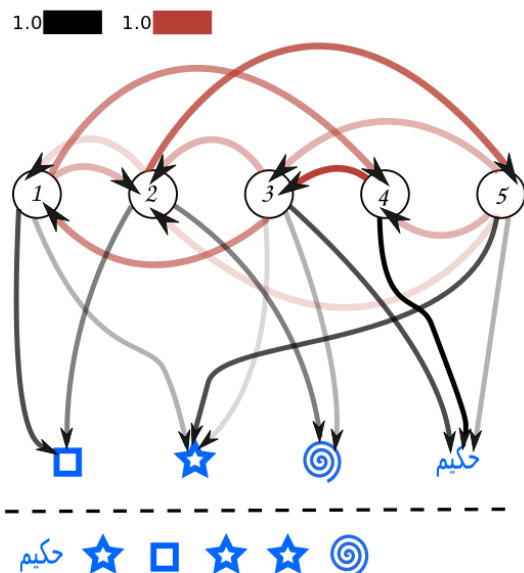
理论基础

核心内容

成果展示

论文总结

## Baum-Welch算法——学习问题



a. Baum-Welch算法应用场景

### The Baum-Welch algorithm

- **Initialization:** input parameters  $\lambda$ , the tolerance  $tol$ , and a real number  $\Delta$
- **Repeat:** until  $\Delta < tol$ 
  - Calculate  $P(O, \lambda)$  using forward algorithm
  - Calculate new parameters  $\lambda^*$  : for  $1 \leq i \leq N$ 
$$p_i^* = \gamma_1(i)$$
$$a_{ij}^* = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)}, 1 \leq j \leq N$$
$$b_{ij}^* = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)}, 1 \leq k \leq M$$
- Calculate  $\Delta = |P(O, \lambda^*) - P(O, \lambda)|$
- Update  $\lambda = \lambda^*$
- **Output:** parameters  $\lambda$ .

b. Baum-Welch算法伪代码

影响Baum-Welch算法主要因素:

- 初始参数  $\lambda$  的设置
- 迭代阈值  $\Delta$



论文摘要

研究背景

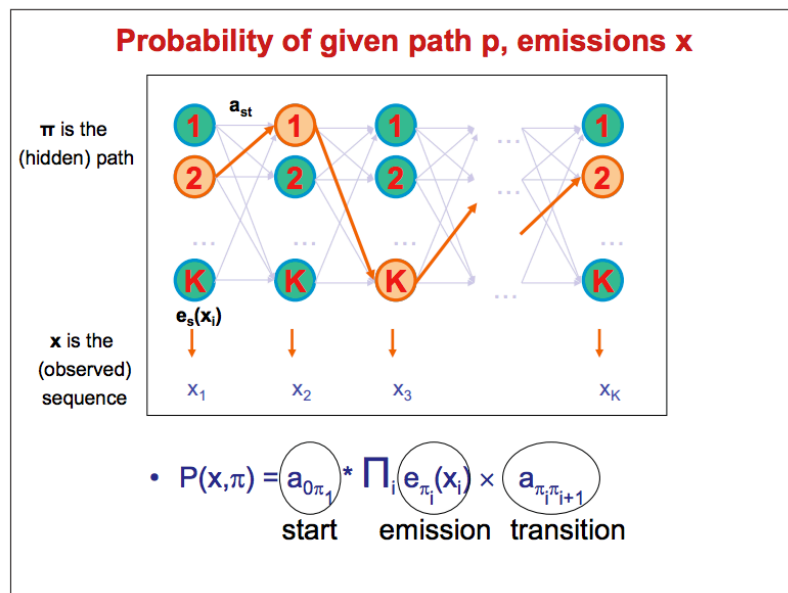
理论基础

核心内容

成果展示

论文总结

## Viterbi算法——解码问题



a. Viterbi算法应用场景

## The Viterbi algorithm

- **Initialization:**

$$\delta_1(j) = p_j b_j(O_1), j = 1, 2, \dots, N;$$

$$\phi_1(j) = 0$$
- **Recursion:** for  $2 \leq t \leq T$ , and  $1 \leq j \leq N$ 

$$\delta_t(j) = \max_i [\delta_{t-1}(i) a_{ij}] b_j(O_{t+1})$$

$$\phi_t(j) = \arg \max_i [\delta_{t-1}(i) a_{ij}]$$
- **Output:**

$$q_T^* = \arg \max_i [\delta_T(i)]$$

$$q_t^* = \phi_{t+1}^*(q_{t+1}^*), t = T-1, \dots, 1$$

b. Viterbi算法伪代码

先通过Baum-Welch评估参数，后用Viterbi算法解码观测值

论文摘要

研究背景

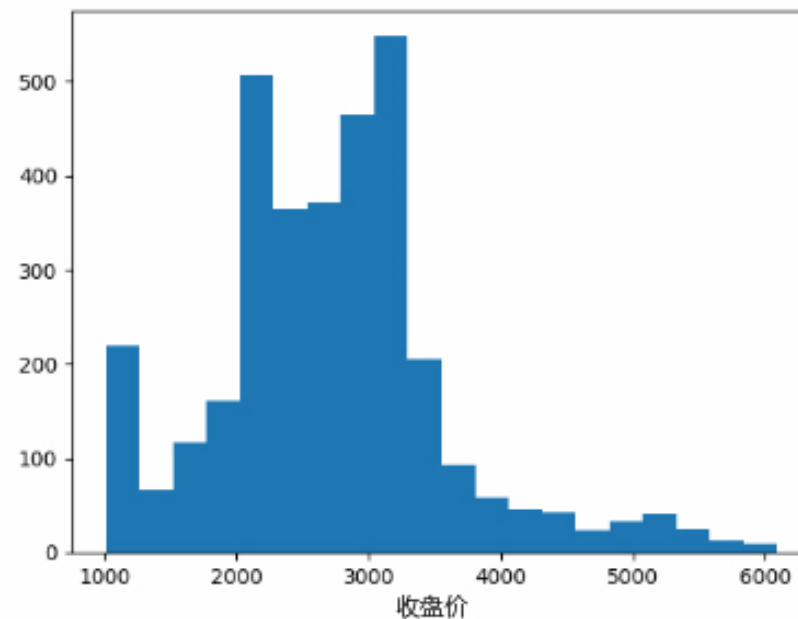
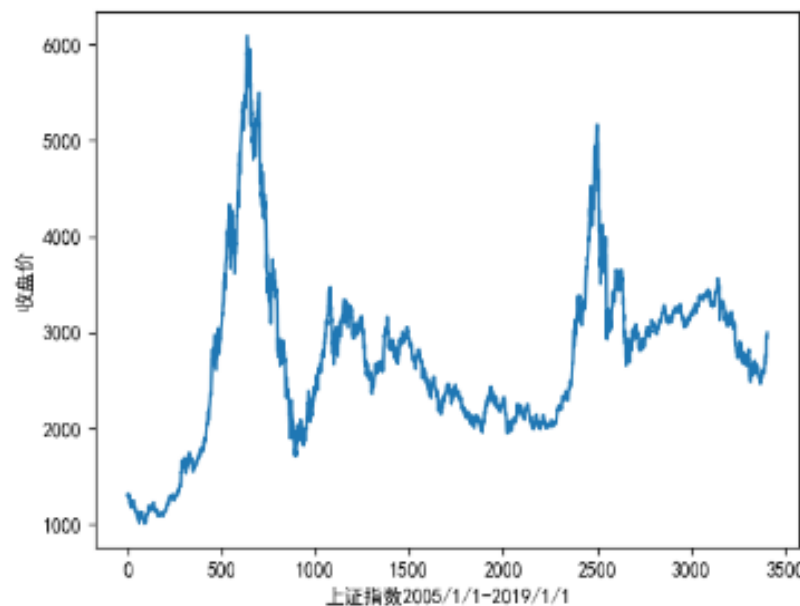
理论基础

核心内容

成果展示

论文总结

## 数据选取



上证指数收盘价及其统计分布



- 2005年1月1日到2019年1月1日内的数据作为训练数据
- 2019年1月1日到2019年2月1日的数据作为测试数据

## 数据选取

选取模型的输入时间序列，包括涨跌幅、成交量、收盘价，组成的观测数据如下：

	Ts_code	Trade_date	Pct_chg	vol	close
0	000001.SH	20181228	0.4353	135406843	2493.896
1	000001.SH	20181227	-0.6087	108799629	2483.086
2	000001.SH	20181226	-0.2605	140670560	2498.294
3	000001.SH	20181225	-0.878	97339852	2504.819
...	...	...	...	...	...
3398	000001.SH	20050107	0.4289	7725570	1244.746
3399	000001.SH	20050106	-0.999	7234683	1239.43
3400	000001.SH	20050105	0.7373	8940871	1251.937
3401	000001.SH	20050104	-1.873	7922254	1242.774

论文摘要

研究背景

理论基础

核心内容

成果展示

论文总结

论文摘要

研究背景

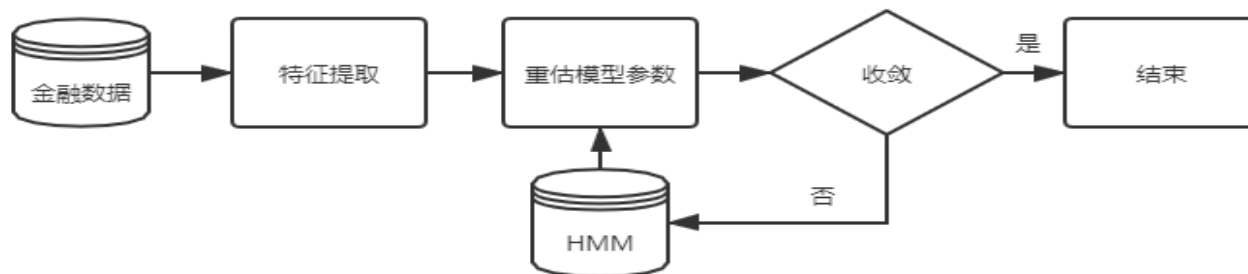
理论基础

核心内容

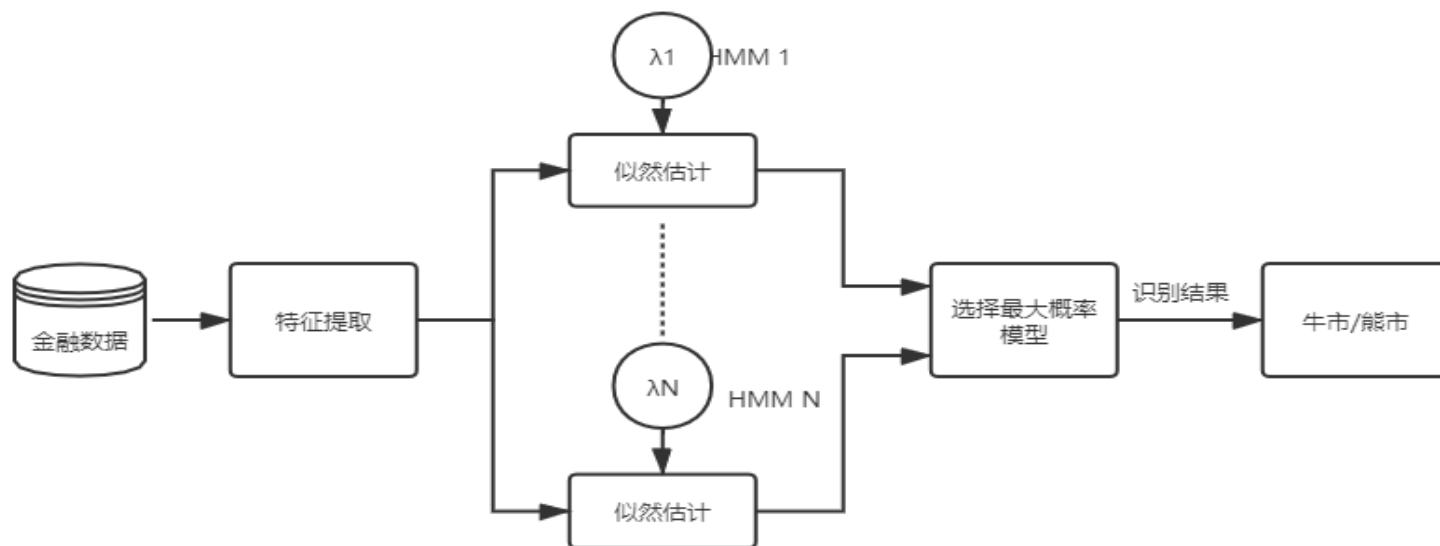
成果展示

论文总结

## Baum-Welch评估与Viterbi解码



a.模型建立过程



b.数据解码过程



论文摘要

研究背景

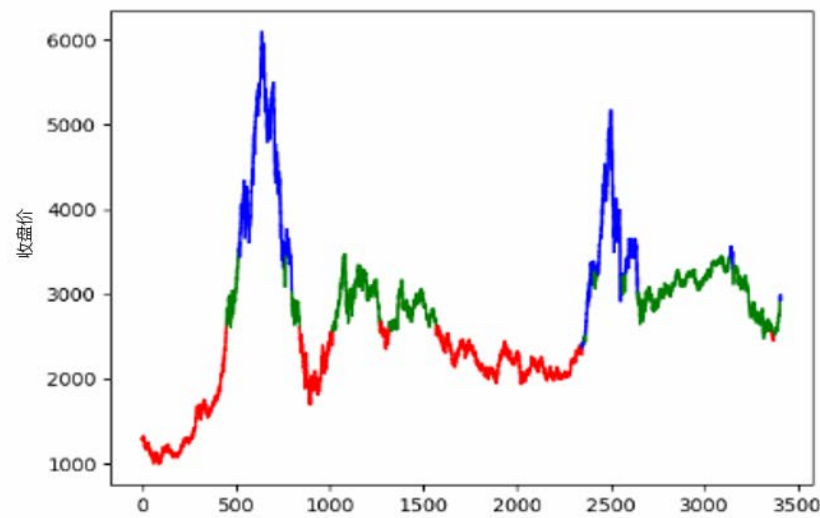
理论基础

核心内容

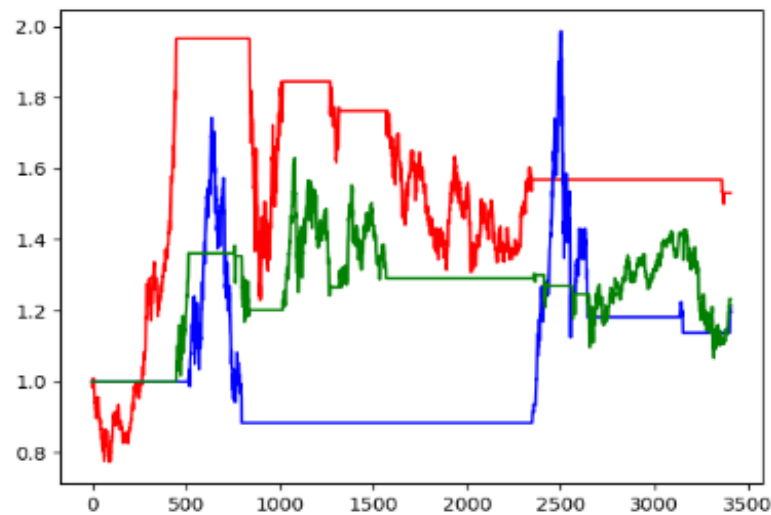
成果展示

论文总结

## 模型训练——三种隐状态



a.三种隐状态下HMM分类情况



b.三种隐状态各投资净收益

从图中收益不能明显判断哪一个种状态是明显的熊市或牛市，最终的投资收益不能保证，无法度量风险

论文摘要

研究背景

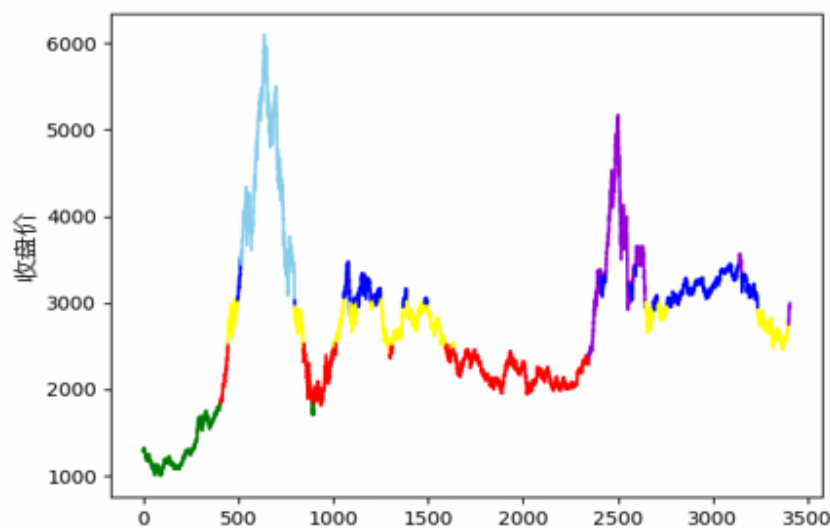
理论基础

核心内容

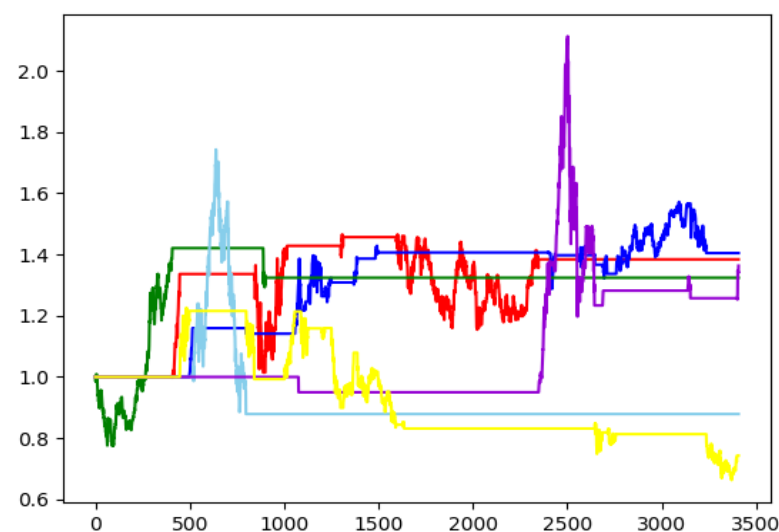
成果展示

论文总结

## 模型训练——六种隐状态



a.六种隐状态下HMM分类情况



b.六种隐状态各投资净收益

从HMM模型分类来看，上证指数被分成了低洼段（绿色），低位震荡状态（红色），中位震荡状态（黄色），高位震荡状态（蓝色），以及大幅牛熊状态（淡蓝色），最后是小幅牛熊状态（紫色）。从数据分析层面理解，不能简单的判断中国金融市场是否是牛熊市。

论文摘要

研究背景

理论基础

核心内容

成果展示

论文总结

## 模型训练——最优参数

初始概率分布:  $\Pi = (0, 0, 0, 0, 1, 0)$

状态转移概率矩阵:

$$A = \begin{pmatrix} 99.27\% & 0.10\% & 0.52\% & 0 & 0.10\% & 0 \\ 0 & 96.61\% & 0 & 3.39\% & 0 & 0 \\ 0.66\% & 0.26\% & 98.02\% & 1.06\% & 0 & 0 \\ 0 & 1.03\% & 1.35\% & 97.43\% & 0 & 0.19\% \\ 0.48\% & 0 & 0 & 0 & 99.52\% & 0 \\ 0 & 0 & 0 & 0.48\% & 0 & 99.52\% \end{pmatrix}$$

由状态转移概率矩阵可知，矩阵对角线的取值都近似为1，代表着六种状态都难以从本身的状态转移成另一种状态。

## 论文总结

[illegible]



论文摘要

研究背景

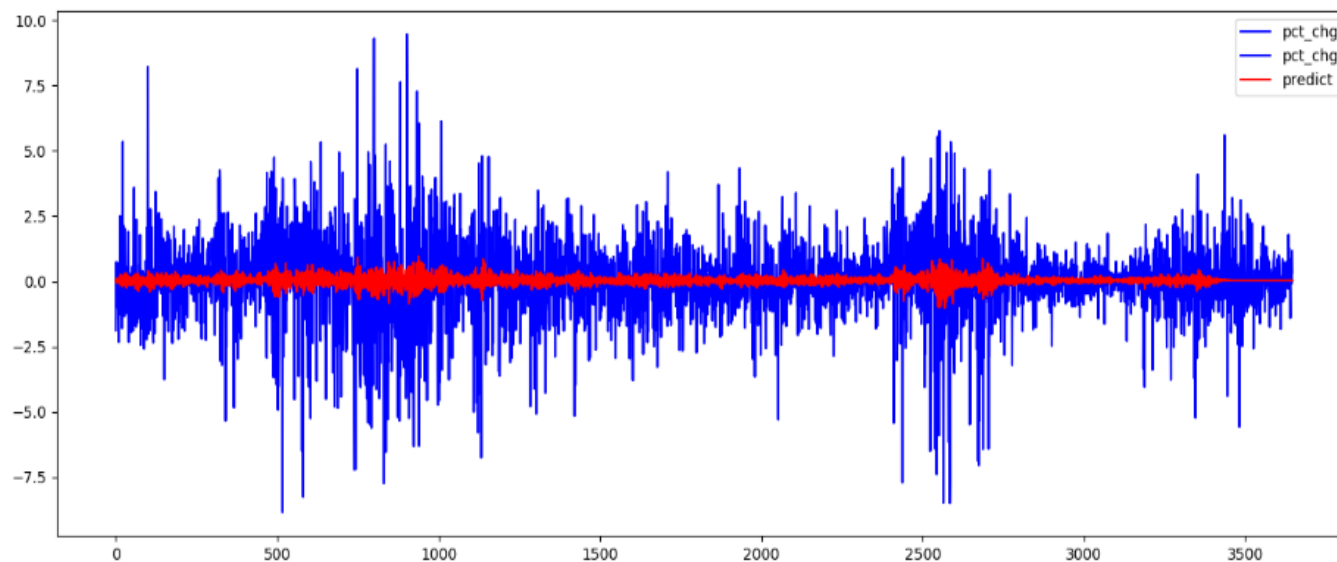
理论基础

核心内容

成果展示

论文总结

## 对比分析——ARMA模型



ARMA模型拟合结果

- ARMA模型严重低估了波动率，所以采取归零判决，最终模型回测准确性50.12%
- HMM模型相比于ARMA模型更适用于具有非线性数据特征的金融市场

论文摘要

研究背景

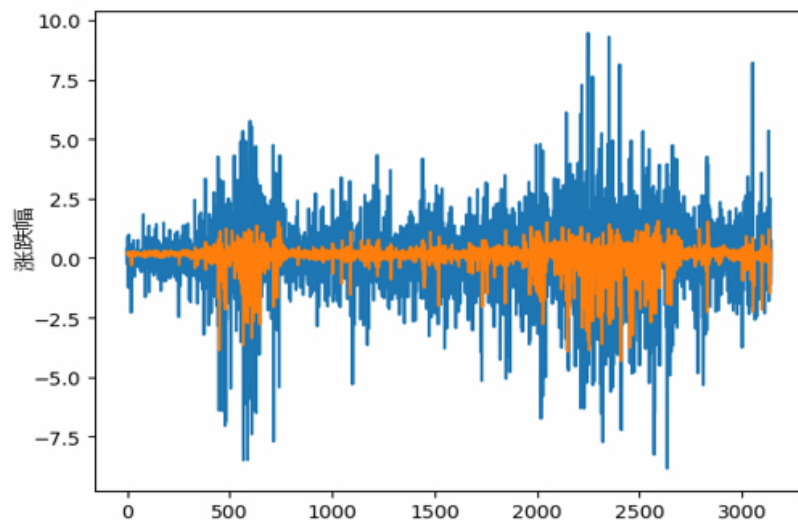
理论基础

核心内容

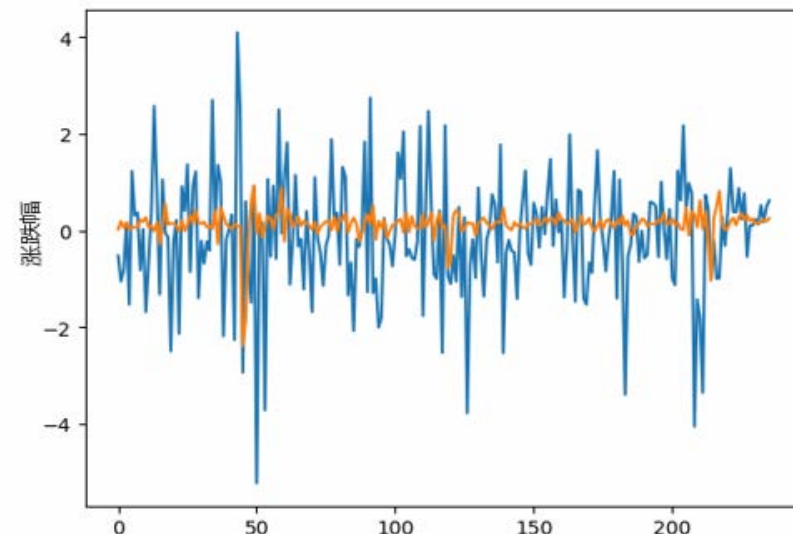
成果展示

论文总结

## 对比分析——LSTM模型



a. ARMA模型训练结果



b. ARMA模型预测结果

- LSTM需要大量算力，结果同样采取归零判决，最终模型回测准确性52.89%%
- HMM模型多输入时无需大量算力

论文摘要

研究背景

理论基础

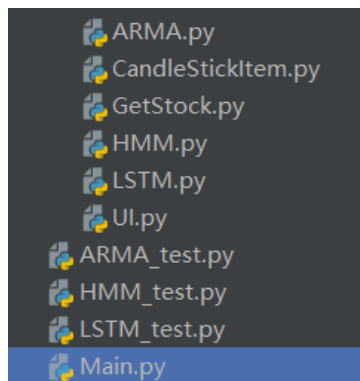
核心内容

成果展示

论文总结



软件设计工具



```
1 import numpy
2 import matplotlib.pyplot as plt
3 from keras.layers import Dense
4 from keras.layers import LSTM
5 import pandas as pd
6 import os
7 from keras.models import Sequential, load_model
8 import tushare as ts
9 from sklearn.preprocessing import MinMaxScaler
10
11 x = '000001' #股票代码
12 ts.set_token('4f4d3e053667fa5bb1c7568860126a6c40b4dfdb05356ba3b066bbe9')
13 pro = ts.pro_api()
14 train_df = pro.index_daily(ts_code=x+'.SH', start_date='20050101', end_date='20190101')
15 test_df = pro.index_daily(ts_code=x+'.SH', start_date='20190101', end_date='20190201')
16
17 train_pct_chg = train_df.pct_chg.values
18 test_pct_chg = test_df.pct_chg.values
19 train_pct_chg = numpy.reshape(train_pct_chg, (train_pct_chg.shape[0], 1))
20 test_pct_chg = numpy.reshape(test_pct_chg, (test_pct_chg.shape[0], 1))
21
22 #归一化后数据有问题
23 scaler = MinMaxScaler(feature_range=(0, 1))
24 trainlist = train_pct_chg
```

```
1 import tushare,matplotlib,datetime,pyqtgraph,pandas,numpy,math
2 from Script import GetStock,CandleStickItem,ARMA
3
4 matplotlib.use('Qt5Agg')
5 from PyQt5 import QtCore,QtWidgets,QtGui
6 from dateutil.parser import parse
7 global stock_total,ShenZZS,ShangZZS,yesterday
8
9 tushare.set_token('4f4d3e053667fa5bb1c7568860126a6c40b4dfdb05356ba3b066bbe9')
10 yesterday = (datetime.date.today() + datetime.timedelta(days=-1)).strftime('%Y-%m-%d')
11 (stock_total,ShenZZS,ShangZZS) = GetStock.get_stock_total()
12 class MainWindows(QtWidgets.QWidget):
13     def __init__(self):
14
15         super(MainWindows,self).__init__()
16         #主窗口的初始化
17         self.setFixedSize(1500, 1000)
18         self.setGeometry(100, 50, 1500, 1000)
19         self.setWindowTitle("北理金融分析平台")
20         self.setWindowIcon(QtGui.QIcon("./Img/BIT_icon.png"))
21
22         #groupbox1的初始化
23         self.groupbox1 = QtWidgets.QGroupBox('',self)
```

部分代码展示



# 成果展示

论文摘要

研究背景

理论基础

核心内容

成果展示

论文总结



以浦发银行为例，ARMA策略的收益结果

	累计净值	年化收益率(%)	夏普率	最大回撤
策略指数	1.41	29.51	98.335	0.132
基准指数	0.94	-4.368	172.49	0.321



以贵州茅台为例，LSTM策略的收益结果

	累计净值	年化收益率(%)	夏普率	最大回撤
策略指数	5.36	212.664	29.563	0.186
基准指数	3.29	111.812	45.698	0.335

论文摘要

研究背景

理论基础

核心内容

成功展示

论文总结

## 总结

- 本文认为我国的金融市场主要表现为“剧烈牛熊市”、“高位震荡”、“中位震荡”、“低位震荡”等几种状态，以此为参考状态能更有效的判断我国金融市场当前的情况。
- 本文通过HMM模型、ARMA模型与LSTM模型进行比较研究，分析了三种模型之间的优点和缺点
- 本研究设计一款UI界面软件，能给个人用户提供参考策略

## 不足之处

- 目前学者都在研究基于HMM的混合模型，而本研究只采用了经典的HMM模型，而经典HMM模型还存在这许多缺点与不足
- ARMA模型作为线性模型不太适合应用在非线性的金融市场，采用改进的GARCH模型做横向对比更有研究意义
- LSTM模型建立过程中受到计算机性能的限制，只采用了一维数据输入，如果能增加到多维数据，模型效果或许能更明显



北京理工大学  
BEIJING INSTITUTE OF TECHNOLOGY



# 致谢

答辩人：江子昊

导师：马宏宾