

申请上海交通大学硕士学位论文

基于量化交易下个股选择系统建模与应用

论文作者 吴鸣远

学 号 1130339048

指导教师 张忠能

专 业 计算机科学与技术

答辩日期 2016 年 2 月 15 日

Submitted in total fulfilment of the requirements for the degree of Master  
in Computer Science and Engineering

# Modeling and System Design of the Stock Selection and Timing based on the Data Mining

MINGYUAN WU

Supervisor

ZHONGNENG ZHANG

SCHOOL OF ELECTRONIC INFORMATION AND ELETRICAL ENGINEERING  
SHANGHAI JIAO TONG UNIVERSITY  
SHANGHAI, P.R.CHINA

15th Feb. 2016

# 基于量化交易下个股选择系统建模与应用

## 摘 要

随着计算机技术与互联网在中国的快速发展，国内金融市场的不断开放，越来越多的人开始将资产投入到金融产品上。另一方面，计算机与互联网的快速发展带来了信息时代的大爆炸，人们每天都面临着海量数据，这直接导致了信息的低能效吸收，并产生碎片化时间。为此，国内许多机构都开始针对证券市场进行量化分析，旨在通过信息挖掘产生更大投资价值。但目前情况是：国内多数为券商、投行等机构在从事相关研究工作；学术领域从事股市技术分析尚处于初级阶段，对技术指标所带来的概率统计、指标选股等，也都存在许多空白。

在此背景下，本文基于国内外学术、商业上研究成果，对一些著名证券市场技术指标，包括指数平滑移动平均线 (MACD, Moving Average Convergence / Divergence)、相对强弱指标 (RSI, Relative Strength Index)、布林线 (BOLL)、威廉指标 (W %R, Williams Larry %R)、随机指标 (KDJ) 等进行了量化研究，考察了这些指标在传统意义上的买卖点有效性特征，并根据这些指标向用户推荐符合指标的股票代码，以供用户作为参考，直接从近 3000 支股票中选出目标股票。

为了量化实验的延伸性，本文通过设计、构建量化选股系统服务平台，实现了对最新以及历史上不同时间段的选股操作，同时构建了一套量化指标分析平台，以分析不同时间段，不同股票池，不同技术指标的效率情况。系统平台服务器建构过程中，使用了

Python 进行数据抓取，并采用 NodeJS 进行服务器的搭建，充分利用 NodeJS 事件驱动，非阻塞 I/O 等轻量特性。此外，在实验过程中，由于数据规模等问题，系统还采用了多进程、算法优化等方法，提高系统效率。

最后，系统成功地通过了正确性、完备性等数据对比测试，检验了系统的可靠性；考察了单技术指标与多技术指标组合时的有效性。同时，文章也针对已经存在的两个国内主要信息平台进行了对比，分析了文中系统与二者之间的创新点与不足之处。其中，摆脱对操作系统的依赖、实现跨平台操作为本系统的一大创新，在量化技术指标有效性分析上，也对其他两个平台进行了补充与完善，更侧重于学术研究价值。

**关键词：** 量化选股、技术指标、数据库、算法、量化分析

# **Modeling and System Design of the Stock Selection and Timing based on the Data Mining**

## **ABSTRACT**

With the rapid development of computer and the Internet technology, more and more people are beginning to invest their assets into financial products. On the other hand, the development of the Internet brought the era of information explosion, people face huge amounts of data every day, directly resulting in low efficiency absorption and fragments of time. For this reason, many enterprises are beginning to embark on stock quantitative analysis, digging valuable information from the markets through technical methods. However, few academic institutes are working on this field. Therefore, it is significant and meaningful to combine the finance and computer science together and do some research in quantitative analysis and stock selection.

In this context, some famous stock market oscillators, including the Moving Average Convergence Divergence (MACD), Relative Strength Index(RSI), Bollinger Bands (BOLL), Williams Larry %R (W %R), Stochastic (KDJ) are analyzed for the academic study and the characteristics are summarized for the system designing and implementation.

In order to quantify experimental extension, a quantitative stock selection system is designed and implemented, operating on historical and nowadays stock pools according to different oscillators and periods. During the implementation, NodeJS server is made full use to utilize its event-

driven, non-blocking I/O and lightweight characteristics. Besides, due to the scale of the data, multiprocessing and algorithm improvement are also adopted in order to improve the efficiency of the system.

Finally, the system designed and implemented successfully passed the correctness and completeness test by comparing with the commercial platform. The single and combined oscillators' efficiency has also been tested and 2 samples are given as examples. Meanwhile, comparison among the existing platforms is also made, which show the cross-platform system design as one of the biggest creativities. The function of academic analysis in the system is also a made-up of the existing platforms, which has a more meaningful academical value.

**KEY WORDS:** quantitative stock selection, oscillator, database, quantitative analysis

# 目 录

摘要	i
ABSTRACT	iii
目录	v
表格索引	ix
插图索引	xi
第一章 绪论	1
1.1 论文研究背景与意义	1
1.1.1 研究背景	1
1.1.2 研究意义	2
1.2 国内外研究现状	3
1.2.1 国内研究现状	3
1.2.2 国外研究现状	8
1.2.3 综合述评	8
1.3 论文行文框架与章节安排	9
1.3.1 研究思路	9
1.3.2 章节安排	9
第二章 量化选股技术指标理论	10
2.1 指数平滑移动平均值 (MACD) 指数	10
2.1.1 定义及计算公式	10
2.1.2 指数研判	10
2.2 随机指标 (KDJ)	11

2.2.1	概念定义及计算公式	11
2.2.2	随机指标研判	12
2.3	相对强弱指标 (RSI)	12
2.3.1	概念定义与计算公式	12
2.3.2	相对强弱指标研判	13
2.4	威廉指标 (Williams R)	14
2.4.1	概念定义及计算公式	14
2.4.2	威廉指标研判	14
2.5	布林 (BOLL) 指标	15
2.5.1	概念定义及计算公式	15
2.5.2	布林线研判	16
2.6	本章小节	16
<b>第三章</b>	<b>量化个股选择系统分析与建模</b>	<b>17</b>
3.1	问题的提出	17
3.2	系统功能需求分析	17
3.2.1	系统功能分析总概	17
3.2.2	数据处理模块	19
3.2.3	量化选股模块	19
3.2.4	技术指标量化分析模块	20
3.2.5	个股信息展示模块	22
3.3	系统平台流程分析与建模	22
3.3.1	时序图	23
3.3.2	活动图	24
3.4	本章小结	25
<b>第四章</b>	<b>量化个股选择系统平台实现</b>	<b>26</b>
4.1	系统平台实现环境与技术	26
4.2	系统平台架构设计	27



4.3	数据抓取与数据库建立	28
4.3.1	股价原始数据抓取	28
4.3.2	建立数据库	31
4.3.3	数据更新及清洗	33
4.4	基于 NodeJS 下的 MVC 服务器架构	35
4.4.1	数据接口，即模型层 (Model)	36
4.4.2	控制层 (Control) 实现	37
4.4.3	启动程序 app.js	42
4.4.4	前台可视化实现，即视图层，View	43
4.5	本章小结	49
<b>第五章</b>	<b>量化选股系统平台测试</b>	<b>50</b>
5.1	个股信息展示功能模块测试	50
5.2	量化选股功能模块测试	51
5.2.1	案例 I: MACD-KDJ 组合	51
5.2.2	案例 II: RSI-WR 组合	53
5.3	技术指标分析模块测试	54
5.3.1	单技术指标检验测试	55
5.3.2	技术指标组合分析	58
5.4	不同量化选股平台对比分析	60
5.5	本章小结	61
<b>第六章</b>	<b>总结与展望</b>	<b>62</b>
6.1	论文总结	62
6.2	研究展望	63
<b>附录 A</b>	<b>个股信息展示正确性测试案例 II</b>	<b>64</b>
<b>附录 B</b>	<b>量化个股选择系统分析与建模</b>	<b>65</b>
<b>参考文献</b>		<b>68</b>

致谢	71
攻读学位期间发表的学术论文目录	72

## 表格索引

2-1 表：RSI 指标市场信号 . . . . .	14
5-1 表：测试样例 I 信息提取 . . . . .	50
5-2 表：测试样例 II 信息提取 . . . . .	51
5-3 表：震荡下 MACD 技术指标相关参数跌幅数据 . . . . .	55
5-4 表：KDJ 技术指标相关参数跌幅数据 . . . . .	57
5-5 表：KDJ-MACD 技术指标组合相关参数跌幅数据 . . . . .	58
5-6 表：三个同类信息平台对比分析 . . . . .	60
B-1 表：牛市下 MACD 技术指标相关参数跌幅数据 . . . . .	65
B-2 表：熊市下 MACD 技术指标相关参数跌幅数据 . . . . .	66

## 插图索引

1-1 图：2014 年中期新开户 A 股市场用户分布图 . . . . .	2
1-2 图：量化交易所涉及的五个概念领域 . . . . .	4
1-3 图：大智慧量化选股设置界面 . . . . .	5
1-4 图：大智慧量化选股结果 . . . . .	6
1-5 图：同花顺技术指标展示 . . . . .	7
1-6 图：同花顺量化技术选股设置 . . . . .	7
3-1 图：量化个股选择系统用例图 . . . . .	18
3-2 图：数据处理模块 . . . . .	19
3-3 图：量化技术指标选股模块 . . . . .	20
3-4 图：量化技术指标分析模块 . . . . .	21
3-5 图：个股信息展示模块 . . . . .	22
3-6 图：系统平台工作时序图 . . . . .	23
3-7 图：系统平台工作活动图 . . . . .	24
4-1 图：系统平台架构图 . . . . .	27
4-2 图：新浪财经股份信息版块 . . . . .	28
4-3 图：股价信息抓取过程 . . . . .	29
4-4 图：多进程抓取股票数据流程图 . . . . .	30
4-5 图：数据库表格信息 . . . . .	31
4-6 图：原始数据清洗 . . . . .	32
4-7 图：多进程抓取股票数据流程图 . . . . .	33
4-8 图：数据更新算法优化比较 . . . . .	34
4-9 图：系统文件结构 . . . . .	35
4-10 图：服务器响应流程 . . . . .	36

4-11 图：系统模型层 . . . . .	37
4-12 图：系统控制层 . . . . .	37
4-13 图：量化选股算法比较 . . . . .	40
4-14 图：并序与顺序算法比较结果 . . . . .	41
4-15 图：视图结构图 . . . . .	44
4-16 图：量化选股平台主页 . . . . .	46
4-17 图：股价展示页面 . . . . .	46
4-18 图：量化技术指标分析页面 . . . . .	47
4-19 图：量化技术指标分析数据来源 . . . . .	48
4-20 图：量化选股页面 . . . . .	48
5-1 图：个股信息展示页面 I . . . . .	51
5-2 图：个股信息展示页面 II . . . . .	52
5-3 图：MACD-KDJ 组合量化选股测试 . . . . .	52
5-4 图：MACD-KDJ 组合量化选股对比测试 . . . . .	53
5-5 图：RSI-WR 组合量化选股测试 . . . . .	53
5-6 图：MACD-KDJ 组合量化选股对比测试 . . . . .	54
5-7 图：震荡市下 MACD 单指标结果分布 . . . . .	56
5-8 图：震荡市下 MACD 单指饼状分析图 . . . . .	56
5-9 图：震荡市下 KDJ 单指标结果分布 . . . . .	57
5-10 图：震荡市下 KDJ 单指饼状分析图 . . . . .	58
5-11 图：震荡市下 KDJ、MACD 组合结果分布 . . . . .	59
5-12 图：震荡市下 KDJ、MACD 组合饼状分析图 . . . . .	59
A-1 图：同花顺中成交量与主要技术指标信息 . . . . .	64
A-2 图：本文系统中成交量与主要技术指标信息 . . . . .	64
B-1 图：牛市下 MACD 单指标结果分布 . . . . .	65
B-2 图：牛市下 MACD 单指饼状分析图 . . . . .	66
B-3 熊市下 MACD 单指标结果分布 . . . . .	67
B-4 熊市下 MACD 单指饼状分析图 . . . . .	67

# 第一章 绪论

## 1.1 论文研究背景与意义

### 1.1.1 研究背景

随着大数据概念的不断普及和深化与计算机性能、数据挖掘、互联网的不断发展,以及市场规范化程度的不断提高,越来越多的领域正在为机器所取代。在金融方面 [1],计算机技术也在不断地渗透到证券市值估值、风险价值评估与管理 and 金融衍生品等各个领域。

根据证券发展史,人类历史上第一支现代意义股票于 1606 年由荷兰东印度公司 (the Amsterdam market for Dutch East India Company) 发行 [2]。从 1606 年至今 2015 年四百多年间,证券市场得到了极大的发展,世界各地陆续兴起不同功能方向的证券交易所,其中包括 1792 年成立的纽约证券交易所、1801 年创立的英国伦敦证券交易所、1891 年中国香港成立的香港证券交易所等。中国大陆 [3] 现代第一个交易所是上海证券交易所,于 1990 年在上海创立。

随着证券市场不断地发展与扩大,与证券相关的投资理论与方法也层出不穷,如针对宏观、行业、公司进行研究的基本面分析方法;以价格变化趋势为基础的股价指数指标技术分析;以不同投资策略模型,如资本资产定价 (Capital Asset Pricing Model, CAPM) 为基础 [4],依赖计算机和互联网以实现统计套利<sup>1</sup>、算法交易<sup>2</sup>、高频交易<sup>3</sup>等的现代量化交易体系,其兴起于 20 世纪 80 年代左右,时间最晚,迄今也不过只有 30 余年的历史,发展却最为迅猛 [7]。

<sup>1</sup>统计套利 (Statistical Arbitrage) [5]: 是通过对历史数据进行统计、分析,计算统计相关变量的概率分布,并纳入基本面的数据进行计算分析而得到的一种套利模式。与无风险套利相比,统计套利增加了一定量的风险,但得到的套利机会,则是无风险套利的几倍。

<sup>2</sup>算法交易 (Algorithmic Trading): 通过计算机平台,输入算法策略为基础的交易指令,执行预定策略,包括交易时间前后、价格高低、交易量大小。一般情况下,算法交易由“机器人”发出指令,无人工干预。算法交易目前广泛应用于养老基金、投资银行、共同基金等买方投资者,以把大单交易分割为许多小单来对冲市场风险。

<sup>3</sup>高频交易 (High Frequency Trading) [6]: 是通过使用计算机,实现人类无法实现的极短时间的高频率交易。例如,某支股票出现了买入与卖出之间极小的偏差,让计算机有机会在极短时间内“低买高卖”。其对计算机技术及网络要求非常高,一些企业甚至将公司服务器搬至证券交易所附近。

本文则通过对证券投资技术研究，结合量化交易思想与理论，依赖计算机技术，以中国主板（A 股）市场个股股池数据为基础、股票技术指标为辅助，建立数据模型，检验市场主流技术指标在中国 A 股市场的适用程度并建立起个股推荐系统。

1.1.2 研究意义

结合目前中国市场发展两大背景：资本资产证券化<sup>4</sup>与互联网金融发展。[9] 随着中国资本市场的不断开放和互联网的不断发展，中国金融市场正在逐步走向繁荣，如艾瑞咨询集团报告中提到，“2013 年中国个人可投资资产高于 130 万元高净值投资者数量由 2012 年提升至 1202 万人，同比增长 17.2%，而未来同样会保持年均 15% 以上增速” [10]。同样，另一份报告显示，中国 2014 年网络经济规模已经达到 8706.2 亿元人民币，相比去年增长 47.0%，预计 2015 年突破万亿 [11]。在证券投资领域，《证券日报》于今年 7 月份报告中提到，2015 年上半年，A 股新增股民近 2000 万，净增长超过四成。另一方面，在 2014 年中期后新开户股民中，有 5.8% 的人是文盲，67.6% 股民持初中及以下学历，本科及本科以上仅约 6% 人数，其分布如图 1-1 所示。

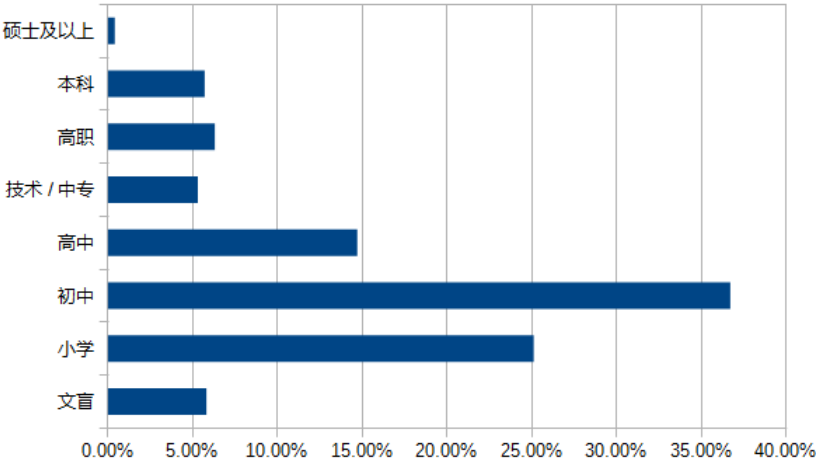


图 1-1 2014 年中期新开户 A 股市场用户分布图  
Fig 1-1 New Stock User Distribution in 2014

<sup>4</sup>资产证券化（Asset Securitization）[8]：指将具有流动性差、但是有预期收入的资产，在金融资本市场以发行证券方式进行出售以获取融资的一种方式，其目的是极大得提高资产的流动性。

因此，在互联网与金融正在迅速发展的时代背景下，利用计算机技术，对国内国外市场尚未探索的领域进行相关研究具有重要意义。本文通过研究业已存在的技术指标和文本挖掘，搭建相关数据模型，并构建分析与推荐系统，检验了美欧等提出的技术指标在中国主板市场上的适应程度，旨在为机构、券商、大中小型投资者及散户提供一些可作参考的数据和结论。

## 1.2 国内外研究现状

目前为止，关于量化选股作为投资组合相关学术研究在国外金融领域有许多成果，并且在应用领域也已经日臻成熟，有数据统计，在美国，有超过 80% 的交易都是计算机量化进行 [12]。国内基于金融市场起步较晚等原因，学术上研究较少，研究者主要集中在众多机构研究所，其研究根据具体市场出发，具有较大参考意义。

### 1.2.1 国内研究现状

量化交易发展目前在中国市场上尚处于初级阶段，学术与商业机构对于证券市场历史数据的研究和挖掘也于近几年成为研究热点。在国内资本市场不断开放与发展的背景下，越来越多的高校、机构开始涉足量化交易领域。

知网上，共计有 66,693 篇文献涉及量化交易，然而涉及量化交易的计算机学科领域文献只有 2781 篇，在学术上的研究较少，国内主要的研究者都在公募、私募、证券公司研究所等机构从事研究，研究者一般从市场实际出发，他们在研究成果也通常以研报的形式出现在各大研究平台，包括万得、同花顺等。

如图 1-2 所示，在研究内容上，量化交易涉及的几个领域：自动化交易、量化投资<sup>5</sup>、程序化交易<sup>6</sup>、算法交易以及高频交易，然而，受制于国内资本市场开放特征，上述几大领域并没有充分向投资者开放<sup>7</sup>。

<sup>5</sup>量化投资 (Quantitative Investment)：一般概指通过概率，微积分等数学工具去研究金融市场各种资产价格的结构性的原因来决定的投资。

<sup>6</sup>程序化交易 (Program Trading)：即利用程序 (program) 进行交易，具体交易仓位、时机、止损止盈、获利标准包含在程序中或独立于程序之外，程序本身仅为执行方式，与人工交易对应。

<sup>7</sup>中金所 9 月 2 日公布政策新规：机构与个人单日开仓交易量不得超过 10 手；提高合约品种保证金各 10%；平仓手续费由 0.015% 增至 0.23%；实名制，这在实践上基本宣告高频交易、程序化交易在中国已经不可行或者尚且未到开放的时机。



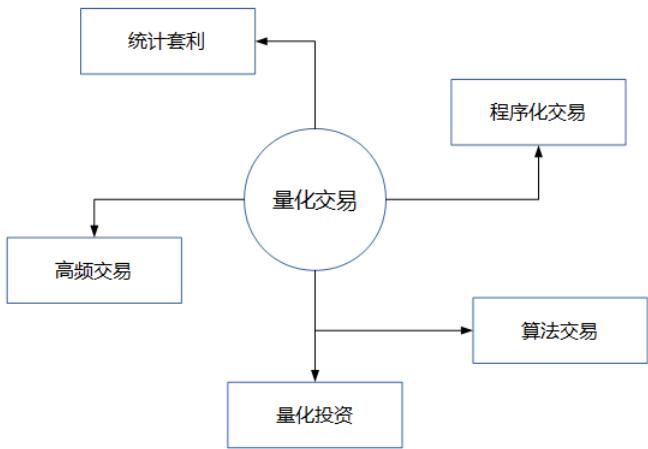


图 1-2 量化交易所涉及五个概念领域

Fig 1-2 Five Main Area in Quantitative Trading

目前国内主要量化集中在量化投资部分，根据凯纳陈曦 [13] 报告，“量化投资”已经成为 2014 年中国券商、基金公司、私募等最热门关键词，其中主要涉及量化对冲产品和量化 CTA 产品。前者通过利用股票市场和沪深 300 的股指期货，反向交易实现多空的对冲，套期保值，获取利润，例如：如果沪深 300 指数上涨 2%，股票投资组合上涨 3.0%，则对冲组合收益为 3%-2%=1%；如果沪深 300 下跌 1%，而股票投资组合下跌 0.5%，则投资组合整体收益为 1-0.5=0.5%。因此，只要投资组合相对沪深 300 指数强势就能获得绝对收益；后者为量化 CTA 产品，主要包括期货合约的量化交易，目前国内四大期货交易所所有 40 多品种，而期货 CTA 则是交易期货品种量化择时或者统计套利来获取利润。

统计套利与算法交易上，国内研究平台近年也已经有一个长足发展。通联量化平台 [14]，即为一个提供在线程序化交易的大数据测试平台，用户可以撰写量化策略，可进行 alpha 回溯，测试策略。

量化选股上，大智慧信息平台与同花顺分别提供了相关量化选股系统平台。

大智慧

大智慧是一个证券行情显示、分析、外汇及期货的证券信息平台，根据付费等级不同，提供不同金融服务的本地系统平台，主要存在 Windows PC 与

Mac 两个版本。这里，文章仅针对免费信息平台进行讨论。

作为一个综合性证券信息平台，大智慧不仅提供了行情功能，还提供了主题版块功能、财务信息模块、指标系统与主力监控功能。结合本篇文章，这里仅针对行情展示功能、指标系统与量化选股功能进行介绍。

个股信息展示上，除了股价 K 线图与成交量外，大智慧仅提供了一个栏目作为技术指标的展示模板，候选的技术指标因子共计 12 个，包括常用的 MA、MACD、KDJ、KD、WR、RSI 等，不易作对比操作，且用户交互并不友好。

选股上，大智慧提供了不同模式的条件选股，包括技术指标、条件选股、模式匹配等，其中，技术指标如图 1-3 所示，在设置基本变量（参数、时间段等）后，执行选股后，会产生一系列选股结果，如图中右侧所示。

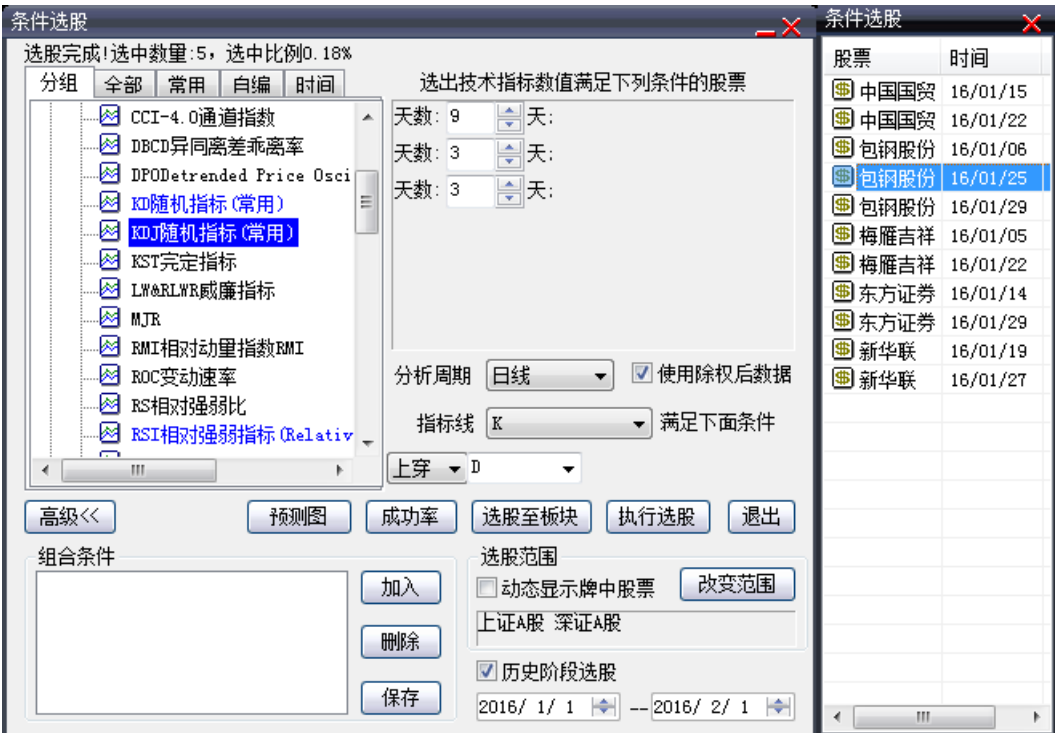


图 1-3 大智慧量化选股设置界面

Fig 1-3 DZH Quantitative Stock Selection Settings

单击选定日期，则会弹出如图 1-4 所示日期股价信息数据。此外，大智慧还提供历史成功率计算页面，如图 1-4 中所示，其中 (a) 为其量化数据信息，(b) 中所示为其图形分布。

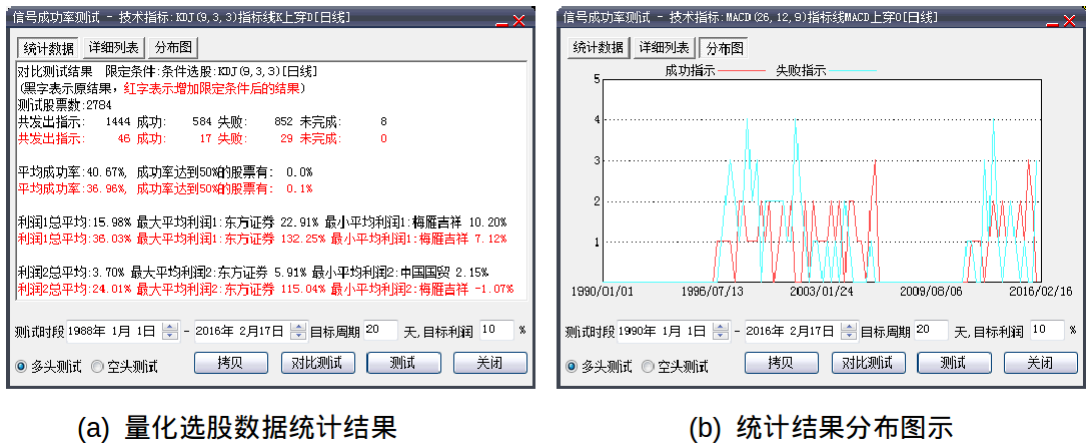


图 1-4 大智慧量化选股结果

Fig 1-4 DZH Quantitative Stock Selection Result

总结大智慧证券信息平台中不足如下：

- 1. 技术指标无法对比参照，组合对比无法实现；
- 2. 量化技术指标选股结果标的丢失，即使在测算成功率中涵盖了全体股票池，其结果仍然只有部分标的；
- 3. 技术指标分析宏观，没有对全体股价进行检验分析，没有对标的未来数日涨跌幅进行统计分析；
- 4. 系统离线，安装操作复杂繁冗，无法跨平台操作，功能庞杂，功能不精确。

同花顺

与大智慧类似，同花顺也是一款成熟的证券信息平台，为用户提供了更为复杂的、多功能系统。这里，文章仅介绍与本文研究相关的一些属性。较之于大智慧，同花顺在个股 K 线展示上，允许用户对指标组合进行设定，以进行多指标组合参照对比分析。如图 1-5 所示，即为三个技术指标 RSI、KDJ、MACD 组合对比分析图。

同时，同花顺也为用户提供量化技术指标选股等功能模块。如图 1-6 所示：用户可以根据自己需求，选定参数，对全体股票池进行选取。

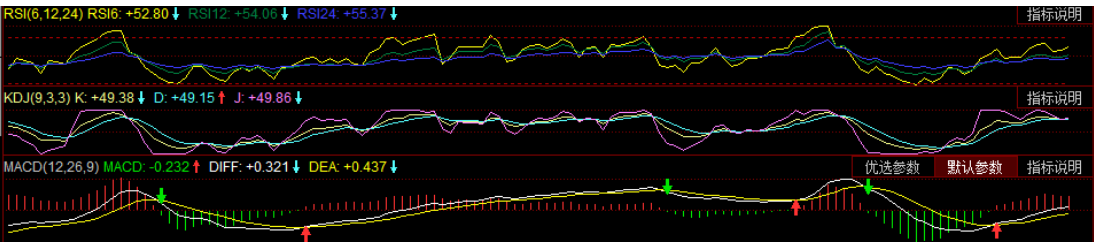


图 1-5 同花顺技术指标展示

Fig 1-5 THS Oscillator Board

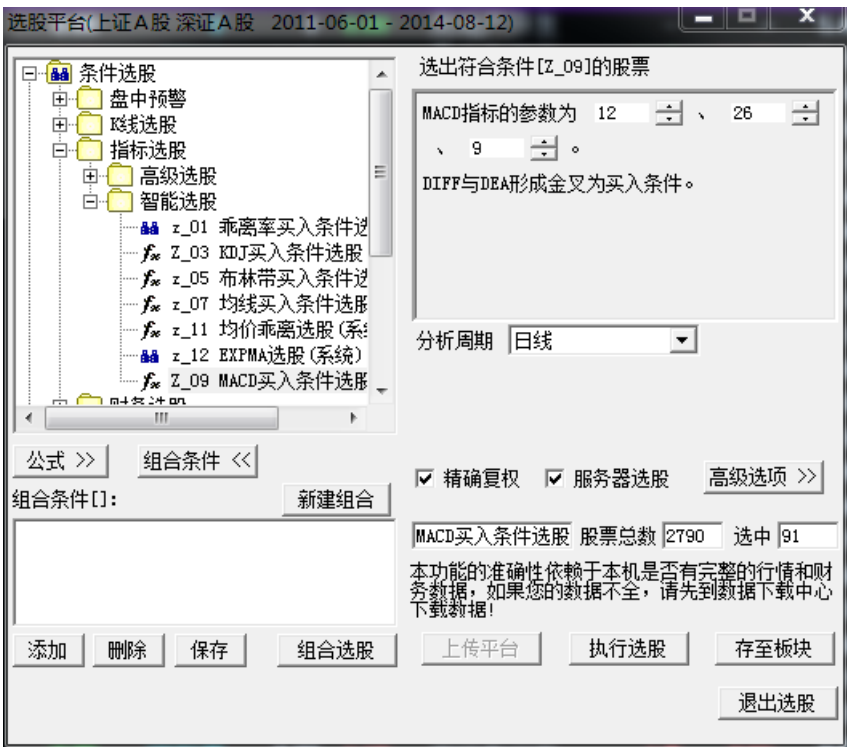


图 1-6 同花顺量化技术选股设置

Fig 1-6 THS Quantitative Stock Selection Settings

然而，同花顺选股系统中也存在一些不足之处，除了无法实现跨平台，对操作系统依赖之外，还存在如下问题：

1. 与大智慧平台不同，同花顺无法对历史选股进行到点还原；
2. 在自定义执行技术指标选股上，不同时段推荐股票数量相同，且不提供推

荐股票满足条件的日期及股价信息数据;

### 1.2.2 国外研究现状

国外量化交易研究起步于 20 世纪 80 年代, 已有 30 余年, 金融市场的成熟度也高于国内。著名量化投资基金有: 西蒙斯 (James Simons) 于 1982 年创立的文艺复兴科技公司 (Renaissance Technology) [15], 大卫 (David Shaw) 创立的对冲基金 D.E. Shaw, 并于 1986 年加入摩根斯坦利旗下 APT 量化交易部门。

而国外在量化交易领域的研究则可以直接追溯至 1952 年 Markowitz 提出的投资组合相关理论, 首次对证券风险因素进行了正规阐述, 为现在证券组合理论的开端 [16]; Sharpe、Litner、Mossin 分别于 1964 年、1965 年和 1966 年研究了资产价格均衡结构, 得出了资本资产定价模型 (Capital Asset Pricing Model, CAPM) [17]; 1965 年, Samuelson 提出了有效市场假说 (Efficient Market Hypothesis, EMH) [18]; 1976 年, Ross 建立了套利定价理论 (Arbitrage Pricing Theory, APT) [19], 多因子选股即为 APT 经典代表; Asness 于 1997 年提出, 市场股票估值应该关注公司基本面, 如现金流、偿债能力、财务水平等数据; 20 世纪 80 年代, 行为金融学诞生 [20]; 20 世纪 90 年代, 计算技术与量化交易进一步融合, 如非线性领域对遗传算法、决策树、神经网络等模型的构建, 进行量化选股。

此外, 由于国外二级市场与我国主板市场差异性等因素, 量化交易体系成熟, 并侧重于高频交易、策略投资等开放市场模型上, 因此将不作为重点在此讨论。

### 1.2.3 综合述评

通过对国内外量化交易发展研究, 可总结出目前国内量化交易发展与发达国家, 尤其是美国之间的距离。但随着中国资本市场的快速发展, 互联网对金融领域地不断渗透, 量化交易所蕴涵的巨大潜力, 也必然将得到释放。学术上对这一领域的研究, 也显示出其巨大的学术价值与意义。

基本国内外研究现状, 本文将对 A 股市场上的股票进行量化技术指标研究。并基于目前已经存在的系统平台, 通过其存在的一些不足, 研究、设计并实现互联网平台上的量化技术选股系统, 进行跨平台实践。

## 1.3 论文行文框架与章节安排

### 1.3.1 研究思路

本文主要研究了技术指标量化选股的应用，数据源为主板（A 股）市场 2784 支股票的市价日数据及交易量。第一步是通过爬虫获取原始数据源并进行预处理、存储；第二步是建立模型，产生后台数据；第三步是搭建选股系统，给出推荐股票池，第四步为分析技术指标并给出实践分析结果。

本研究旨在建立起一个开放、开源研究平台，并对量化研究人员和投资者提供帮助。

### 1.3.2 章节安排

本论文一共分为六个章节，第一章节为绪论，主要介绍了课题研究的背景和意义，国内、国外主要研究成果，对比了国内主流平台之间的优点与不足之处，研究思路与章节安排；第二章量化选股技术指标理论，详细阐述了量化技术指标中的各个技术指标，如 MACD，KDJ，WR，BOLL 等，并详细介绍了各个技术指标的计算公式。

第三章是系统设计部分，包括四个模块，即数据处理模块、个股选择模块、技术指标分析模块、个股信息展示模块，涉及到系统平台的用例图、活动图与时序图等，为第四章中实现部分提供设计框架；第五章为量化技术指标分析测试部分，这一章基于论文搭建的系统平台，对比国内成熟商业系统结果进行对比测试，分析了个股信息平台、选股结果、单个技术指标与多个技术指标组合的有效性检验正确性进行检验；最后一章为本文总结，对整个研究进行总结，得出技术指标可行性，并对下一步研究作出展望。

## 第二章 量化选股技术指标理论

### 2.1 指数平滑移动平均值 (MACD) 指数

#### 2.1.1 定义及计算公式

指数平滑移动平均值 (后文简称 MACD, Moving Average Convergence / Divergence) 由 Gerald Appel 于 20 世纪 70 年代发明 [21], 该指数由两条移动平均线 (DIF 与 DEA) 差值而成 [2], 移动平均线的计算公式如下:

$$MACD = 2 \times (DIF - DEA) \quad (2-1)$$

其中, 差离值 (DIF 或 DIFF, Differenece) 为 12 日移动平均值 (EMA, exponential moving average) 与 26 日移动平均值之差, 即:

$$EMA_{t+1}^n = \frac{2}{n+1} \times P_{t+1} + \frac{n-1}{n+1} \times EMA_t \quad (2-2)$$

$$DIF_t = EMA_t^{12} - EMA_t^{26} \quad (2-3)$$

其中,  $t$  为计算时间, 本文中,  $t$  为日为单位;  $n$  为 EMA 计算周期。因此, 在持续的上涨中, DIF 值会越来越大; 在跌势中, DIF 则越来越小。根据差离值 DIF 作出 9 日移动平滑, 则得到慢速线 (DEA, Difference Exponential Average), 如下公式所示:

$$DEA_{t+1} = \frac{4}{5} \times DEA_t + \frac{1}{5} \times DIF_{t+1} \quad (2-4)$$

至此, 我们即可得到 MACD 指标中的三个重要参数 MACD 柱值、DIFF 与 DEA。

#### 2.1.2 指数研判

MACD 指数指标的研判主要从 DIFF、DEA 和柱值形态展开 [22]。本文中, 主要考虑其位置与交叉情况对股票未来发展趋势的研判。



DIFF 与 DEA 金叉、死叉情况:

1. DIFF 与 DEA 均为大于 0, 即在横轴线上时, 个股属于多头市场, DIFF 向上突破 DEA, 为买入信号, 这种交叉即为 MACD 指标中“黄金交叉”的一种。
2. DIFF 与 DEA 均为小于 0, 即在横轴线下时, 个股属于空头市场, 但 DIFF 由下向上突破 DEA, 视为跌势已尽, 市场即将转强, 为买入信号, 这种交叉为 MACD 指标中第二种“黄金交叉”形式。
3. DIFF 与 DEA 均为小于 0, 即在横轴线下时, 个股属空空头市场, DIFF 向下跌破 DEA, 为卖出信号, 这种交叉即为 MACD 指标中“死亡交叉”的一种。
4. DIFF 与 DEA 均为大于 0, 即在横轴线上时, 个股属于多头市场, 但 DIFF 由上向下突破 DEA, 视为强势已尽, 市场即将转弱, 为卖出信号, 这种交叉为 MACD 指标中第二种“死亡交叉”形式。

## 2.2 随机指标 (KDJ)

### 2.2.1 概念定义及计算公式

随机指标 (KDJ) 由 George C. Lane 在 20 世纪 50 年代发展出来 [24], 是一个量能指标, 最早起源并应用于期货市场, 用以表示股票交易过程中出现的强弱势和超买超卖现象。KDJ 指标最早以 K、D 指标的作为应用, 而 K、D 指标是在 RSV (未成熟随机值, Raw Stochastic Value) 基础上建立起来。但是, K、D 仅判断股票超买超卖状态, KDJ 指标则结合移动平均线相关因素, 形成了较准确的买卖信号。本文中采用 K 线、D 线与 J 线三条主线, 研究股票最高价、最低价和收盘价之间的关系, 同时也涉及了动量、强弱和移动平均线等因素。为计算 KDJ 值, 首先需要计算出相应的 RSV 值 [25], 如下所示:

$$RSV(n) = \frac{c_t - l_t}{h_n - l_n} \times 1000 \quad (2-5)$$

其中,  $n$  为时间跨度, 即随机值周期, 一般通用周期为 9 天,  $t$  为当日日期。因此,  $c_t$  为当日收盘价,  $l_t$  为当日最低价,  $h_n$  为  $n$  天内最高收盘价,  $l_n$  为  $n$  日内最低收盘价。在计算完 RSV 值后, 计算相对应的 K、D、J 值, 计算公式如下:



$$K_{t+1} = \frac{2}{3} + \frac{1}{3} \times RSV_n \quad (2-6)$$

$$D_{t+1} = \frac{2}{3} \times D_t + \frac{1}{3} \times K_{t+1} \quad (2-7)$$

$$J = 3K - 2D = K + 2(K - D) \quad (2-8)$$

K,D,J 初始值为 100，其中，K 值为 D 值基础，J 值为 K、D 值辅助，共同研判市场行情。

### 2.2.2 随机指标研判

KDJ 指主要通过 K、D、J 三条曲线构成图形关系来分析个股的超买和超卖现象，走势背离与 K 线、D 线、J 线交叉等，以此来预测个股股价短、中、长期走势。其中，主要有以下一些预判标准 [26]:

- 1 K、D、J 值永远在 [0, 100] 区间，K、D、J 大于 80 为超买现象，K、D、J 小于 20 为超卖现象；K、D、J 在 20 到 80 之间时，为徘徊区间，建议观望；
- 2 KDJ 曲线黄金交叉：上涨行情中，K 值小于 D 值，D 值小于 J 值，并且由下向上突破，形成交叉时，为买进信号；
- 3 KDJ 曲线死亡交叉：相反，下跌行情中，K 值大于 D 值，D 值大于 J 值，并且由上向下突破，形成交叉时，为卖出信号。[3]

KDJ 具有直观、迅速、快捷研判行情的性质，对市场反应敏捷，被广泛用作股市中短期趋势指标分析，是本文研究的一个重要指标。

## 2.3 相对强弱指标 (RSI)

### 2.3.1 概念定义与计算公式

相对强弱指标 (RSI, Relative Strength Index) 由 J. Welles Wilder [27] 于 20 世纪 90 年代提出，是一个量能指标，用来测算市场价格的变化及其速度。根据 J. Welles Wilder 理论，RSI 值在任何市场价格的上漲或下跌中，都限定在 [0,

100] 之间变动，其数值变动反映了标的供求之间的关系，当适合短线投资者使用。

$$RSI_t^n = \frac{RS_t^n}{1 + RS_t^n}, \text{ for } RS_t^n = \frac{AG_t^n}{AL_t^n} \quad (2-9)$$

其中，n 代表指标周期，t 为当日日期，RS 为相对强度，为上升强度（AG）与下降强度（AL）的比值，计算方式如下：

$$AG_{t+1}^n = \frac{AG_t^n \times 13 + G_{t+1}^n}{14}, \text{ for } AG_1^n = \sum_{t=1}^{14} G_t^n \quad (2-10)$$

$$AL_{t+1}^n = \frac{AL_t^n \times 13 + L_{t+1}^n}{14}, \text{ for } AL_1^n = \sum_{t=1}^{14} L_t^n \quad (2-11)$$

经过 J. Welles Wilder 测试，给出一个最佳周期值，即 n=14，亦为本文中所用数据参数。上述公式中，AG 为 n 天（这里为 14）收盘价增加值的累计，同样，AL 为 n 天收盘价下降的累计结果。初始值计算如下：

1. 获取最初 14 天累计上升值和，除以 14，为  $AG_1$
2. 获取最初 14 天累计下降值和，除以 14，为  $AL_1$
3. 让  $AL_1$  除以  $AG_1$ ，得到  $RS_1$  值
4. 计算  $RS_1$  与 1 和，并除以  $RSI_1$
5. 100 乘以 4) 中结果，即得到 RSI 初始值

### 2.3.2 相对强弱指标研判

RSI 的变化区间一般在 [0, 100]，值在 [30, 70] 之间为正常变动，在 [80, 90] 时，视为市场已经达到超买状态，价格将面临回落调整风险；在 [10, 20] 时，认为市场已经达到超卖状态，至此市场价格有面临企稳回升机会。

国内 A 股市场属于单边做多股市：强弱指标值（RSI）一般分布在 [20, 80]，其数值及买卖信号如下所示：

表 2-1 RSI 指标市场信号

Table 2-1 RSI Oscillator Market Signal

RSI	市场状况	市场信号
80-100	极强	卖出
50-80	强	买入
20-50	弱	观望
0-20	极弱	买入

2.4 威廉指标（Williams R）

2.4.1 概念定义及计算公式

威廉指标（Williams %R, W %R）由拉里·威廉斯（Larry Williams）于 1973 年提出 [28]。是一个振荡指标，依据股价的摆动衡量个股价格是否处于超买或超卖现象。WR 指标度量多空双方竞争出的最高值和最低值距与每天收盘价的距离在一定时间内（n）的股价波动比例，用以提供股市趋势反转信号，是一个用来分析短期市场行情走势指标。W%R 指标与 KDJ 指标具有相似属性，都是基于未成熟随机值（RSV, Raw Stochastic Value），其计算函数以 W%R(N, DOWN, UP) 表示，公式如下：

$$W\%R(N, DOWN, UP) = \frac{HHV(HIGH, N) - CLOSE}{HHV(HIGH, N) - LLV(LOW, N)} \times 100 \quad (2-12)$$

其中，N 为计算周期，HHV (HIGH, N) 为 N 日内的收盘价最高值，CLOSE 为当日收盘价，LLV (LOW, N) 为 N 日内收盘价最低值。本文中，N 值取 6 与 10 两个日期。

数值初始化为上市当日最高价减收盘价比最高与最低价，初始值一般为 0。

2.4.2 威廉指标研判

威廉指标用以研判市场的超买和超卖状态，W%R 值变化区间范围在 [0, 100] 之间，当指标数值超过 80 时，市场为超卖现象；当指标数值低于 20 时，为市场超买现象 [29]。

个股在选择和研判的过程中，以下述原则作为买卖股票的信号触发点：

1. 威廉指标超卖区别为  $[80, 100]$ ，当指标数值超过 80 时，投资者不宜入市，但可重点关注，当指数跌破 80 时，可以考虑积极买入，为买入信号；
2. 指标超买区别为  $[0, 20]$ ，当指标数值突破 20 时，未来股市看跌可能性较大，当指标突破 20 时，投资者可以考虑卖出，视为卖出信号点。

## 2.5 布林 (BOLL) 指标

### 2.5.1 概念定义及计算公式

布林线指标 (BOLL, Bolinger Bands) [30]，是由约翰·布林先生在 19 世纪 80 年代发明的中长期证券市场分析工具。在统计原理基础上，他推算出证券价格标准差与价格变化的信赖区间，来股价波动范围和未来走势。布林线由上轨、中轨和正轨组成，其上下轨道随股价动态变化。

在相关技术指标中，布林指标计算最为复杂，它引进了统计学标准差，涉及到中轨线 (MMA)、上轨线 (UPPER) 和下轨线 (DOWN) 的计算。上轨线为标准中轨线加上两位标准差，下轨线为中轨线减去两位标准差，中轨线为证券市场交易值的  $N$  日移动平均值，布林线指标为 3 根轨道线的组合。布林线中的上轨与下轨体现了市场的波动性，当市场波动剧烈，上下轨道变宽，当市场波动变小，则轨宽收缩。

以日 BOLL 指标计算为例，其计算方法如下：

$$MA = \frac{\sum_{t=T-19}^T close\_price}{20}; \quad (2-13)$$

$$UPPER = MA + 2 \times STD(close\_price_{t=20}) \quad (2-14)$$

$$DOWN = MA - 2 \times STD(close\_price_{t=20}) \quad (2-15)$$

其中，MA 是移动平均值 (Moving Average)，UPPER 为布林线上轨线值，DOWN 为布林线下轨线值。

$t$  为当日时间点， $n$  为计算周期，30 日移动平均值即为 30 天内收盘价和与 30 之比，常用移动平均线 20、30、50、100 和 200 天。一条移动平均线表征着股份的变化趋势 [31]，当股份跌破移动平均线时，股份有继续下跌的趋势，当股份突破移动平均线时，股份有继续上升的趋势。

### 2.5.2 布林线研判

通常情况下，K 线都在布林线上下轨道之间 [22]，牛熊情况下也会有突破情况，根据 K 线与布林线上、中、下轨道之间的位置及上下轨道之间的喇叭口大小，可以给出相关中长期预测。

1. 当 K 线向上突破布林线中轨道线时，预示股市强势性开始显现，股份将上涨，为买入信号；
2. 当 K 线向上突破布林线上轨道线时，预示短期内将大涨，可短期持有或短线买入；如果布林线上、中、下轨线继续上涨，则大概率预示股市将保持强势特征，股价短期内还将继续上涨，直到 K 线运动方向开始回调；

## 2.6 本章小节

本章引入并重点介绍了证券市场中常用的五个技术指标：指数平滑移动均线（MACD，Moving Average Convergence / Divergence）、相对强弱指标（RSI，Relative Strength Index）、布林线（BOLL）、威廉指标（W %R，Williams Larry %R）、随机指标（KDJ），其原理与计算公式，研判方法与缺陷不足，为论文选股系统的设计与实现提供了理论基础。

不同的技术指标，由于其形成原理不同而具有不同特点。如，MACD 平滑均线作为改良的均线，经过平滑处理后，很好地滤去了短期股价波动带来的影响，因此比较适用于中长期股价研判；KDJ 可以直观、迅速、快捷研判行情的性质，对市场反应敏捷，被广泛用作股市中短期趋势指标分析；RSI 检测市场超买状态，从市场供求关系上反映股份未来的变化趋势，是一个中短线技术指标等等。

然而，各个技术指标其本身都具有一定的局限性，单个技术指标并不能最优地选出推荐股票，因此，其技术指标组合研究具有重要意义，下一章将针对如何利用技术指标进行系统与架构设计，并进行相关分析与说明。

## 第三章 量化个股选择系统分析与建模

个股选择系统是一个为用户提供选股操作的完整互联网产品，在构建过程，系统平台包括了数据挖掘与处理、数据库建立、后台服务器搭建与前台展示等部分。本章中将从选股问题的抛出、分析、建模再到模型优化，详细阐述选股平台实现的准备工作。

### 3.1 问题的提出

在第二章中，本文详细介绍了国内外市场上几个主要的短线、中长线选股技术指标，为投资者投融资提供参考指标。尽管在金融领域，不乏有学术期刊对技术指标进行回溯并检验其收益率情况，但鲜有人研究这些技术指标在统计学意义上，从短期到长期的一个表现情况，即这些技术指标在牛熊市表现情况对比，技术指标在不同周期中的表现对比，技术指标组合效益，技术指标对大小盘股效益影响等。

本文即在此背景下，结合大数据概念、互联网金融等交叉领域，提出构建量化分析与选股平台，为技术指标在中国 A 股市场上的综合应用提供学术研究价值。

### 3.2 系统功能需求分析

根据上一节中问题描述，系统功能需求分为四个主要模块：量化技术指标选股、量化技术指标分析、个股信息展示模块和数据处理模块。其中，前三个功能模块面向前台用户，第四个功能模块为系统后台定期自动化处理模块，以保证提供及时有效股票信息。

#### 3.2.1 系统功能分析总概

如图 3-1 所示，平台由两个角色组成：用户与系统。其中，用户角色有选股、分析与个股信息展示三个模块组成，分析与选股模块输入部分包括股票池、时间跨度、指标组合，个股信息展示输入则是时间跨度与股票关键字或代码。

分析模块输出涨跌幅散点分布图、细分饼状图与股票信息；选股模块则输出推荐股票代码信息，个股信息展示模块输出股票价格信息、技术指标信息，并且相关信息相互联动。

系统角色为平台提供每日股价信息的自动化处理机制，定期更新并处理数据，这个模块包括两个部分，即：

- 数据抓取与更新
- 数据指标计算

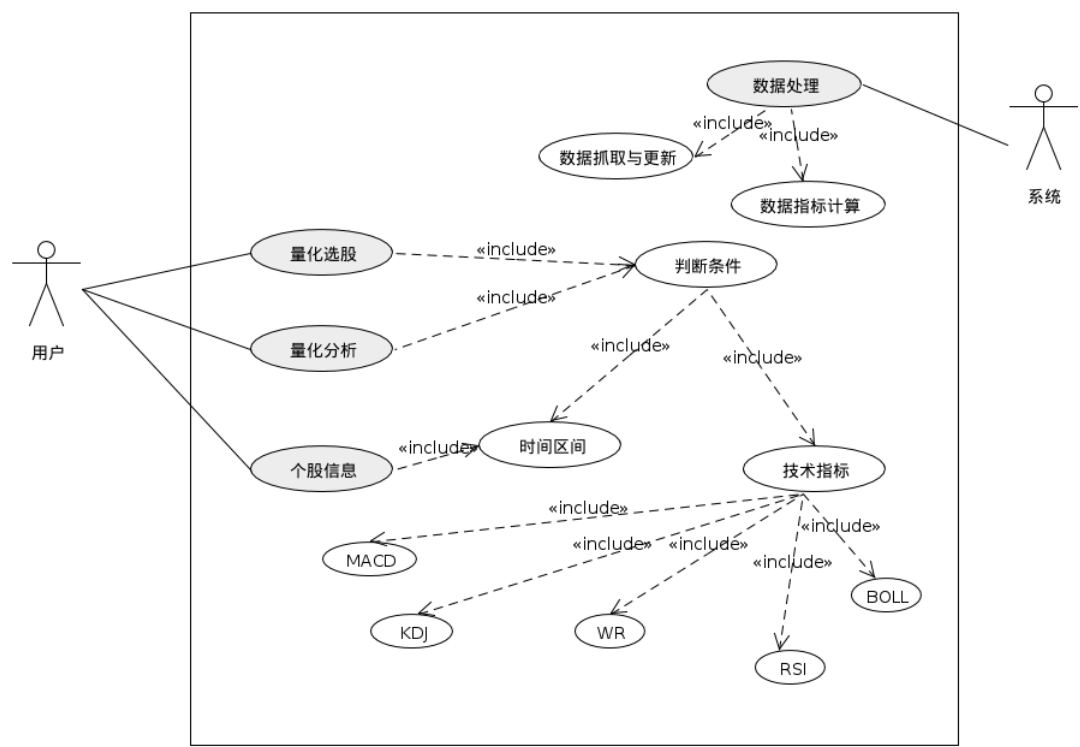


图 3-1 量化个股选择系统用例图

Fig 3-1 Stock Selection Use Case Graph

3.2.2 数据处理模块

量化选股平台系统数据处理模块为后台自动化实现部分，包括原始数据的抓取部分与量化技术指标进一步处理部分。如图 3-2 所示，原始数据抓取与更新部分抓取了所有股票代码的股价基本信息：开盘价、收盘价、最高价、最低价、成交量等日数据信息。同时，这一部分也包括了对原始数据的清洗工作。

数据据理部分则负责各项技术指标应用在股价数据后的数据值，如第二章中所述，文中涵盖了 MACD、KDJ、RSI、WR、BOLL 五个主要技术指标的参考值，为后续量化单技术指标及指标组合提供数据参考值。

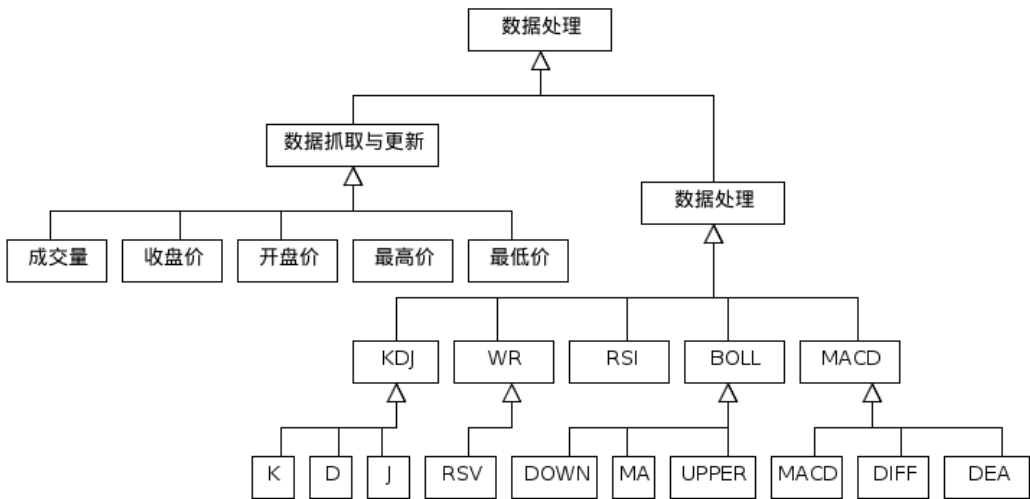


图 3-2 数据处理模块

Fig 3-2 Stock Information Handle

3.2.3 量化选股模块

技术指标量化选股通过技术指标出现的买入点来判断股票是否在一定时间跨度内的买入时机，即是否存在“黄金交叉”。根据万得数据所提供的最新的股票代码表，文章涵盖了 A 股上市公司共计 2781 支股票，包括了深圳证券交易所、上海证券交易所主板市场，以及创业板与中小板市场。

选股方面，平台针对时间跨度、股票池跨度与技术指标上三个维度设置了输入参数：



- 时间跨度：允许用户选择不同时间区间进行选股，并输出所有金叉股票及其时间点；
- 股票池：根据不同用户的不同需求，择股上允许用户选择全体股票或者相关版块股票池或者自定义股票池；
- 技术指标：允许用户选择单指标选股或者多指标组合选股，当用户没有选择股票池数据时，默认使用 MACD；

图 3-3 为量化个股选择系统选股模块功能图。系统通过输入时间区间、股票池与技术指标组合，来进行个股选择，输出为所有符合先前条件的股票代码、名称、多金叉点等推荐股票池，包括当日开盘价、收盘价与买入时间节点等数据信息。

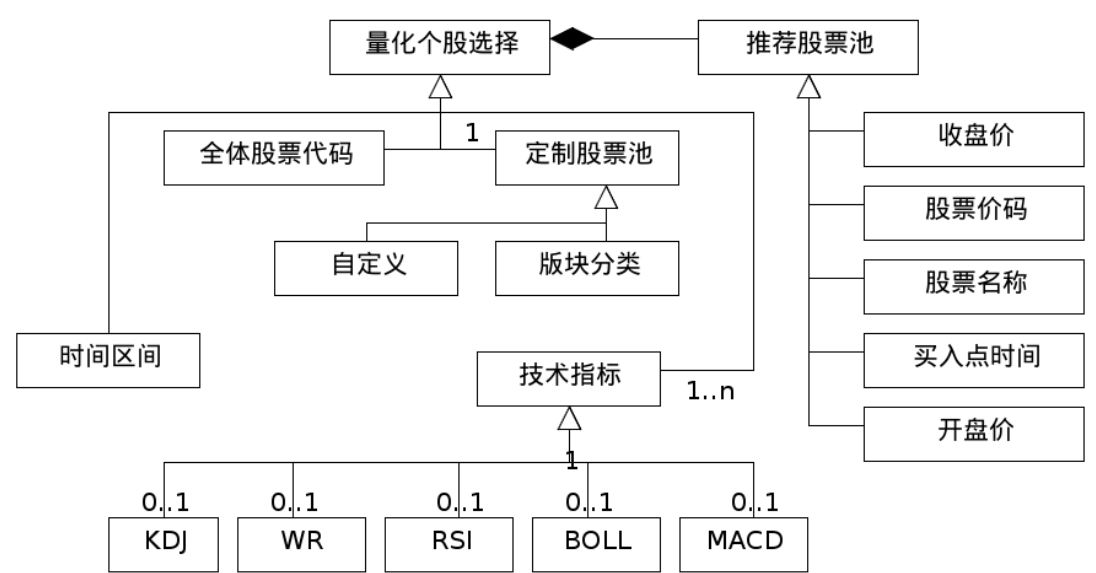


图 3-3 量化技术指标选股模块  
Fig 3-3 Stock Selection Model

3.2.4 技术指标量化分析模块

该模块通过判断指标出现的买入时机为参考，视买入点未来一段时间内的股票价格变动作为分析对象，进行定制化分析。旨在为用户提供一个可定制化，

对不同时间段、不同股票池以及不同技术指标不同组合有效性检验的平台。

如图 3-4 所示，量化个股分析模块主要提供了以下三类信息：

- 1. 目标股票池在买入点后整体股价变动散点分布情况，区间长度为 0 ~ 30 天；
- 2. 长、中、短目标天数股票涨跌比例分布饼状图；
- 3. 股票分布涨跌福数据表单列表；

功能一让用户可以直观地看到目标股票池的技术指标有效性，其涨跌分布强度与概率总览。功能二让用户了解到技术指标在不同时间跨度上的有效性，直接反应技术指标在不同情况下的有效性，辅助用户进行决策；功能三让用户可以对分析结果进行检验，每支股票所链接的股票信息将为用户提供佐证。

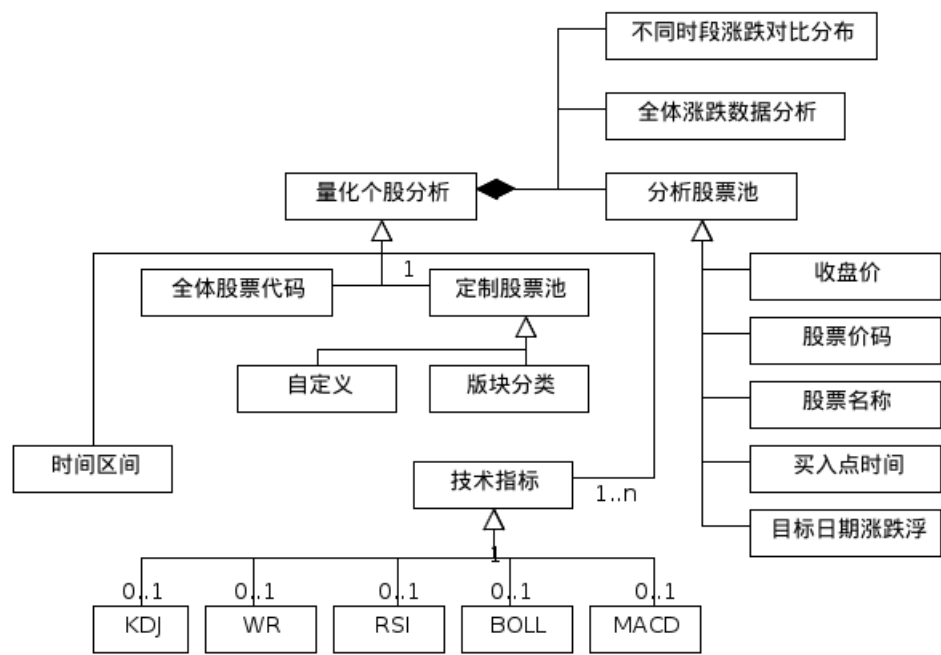


图 3-4 量化技术指标分析模块  
Fig 3-4 Stock Analysis Model

3.2.5 个股信息展示模块

股票信息展示模块为平台基本模块，主要为用户提供了及时的股票日数据信息，如图 3-5 所示，包括：

- 股票四个价格信息 K 线图：开盘价、收盘价、最高价、最低价
- 股票日数据成交量；
- 平均平滑指数（MACD）；
- 随机指标（KDJ）；
- 强弱指标（RSI）；
- 威廉指标（WR）；
- 布林线指标（BOLL），考虑到布林线中涉及到股票的 K 线图，因此将布林线展示在 K 线图上；

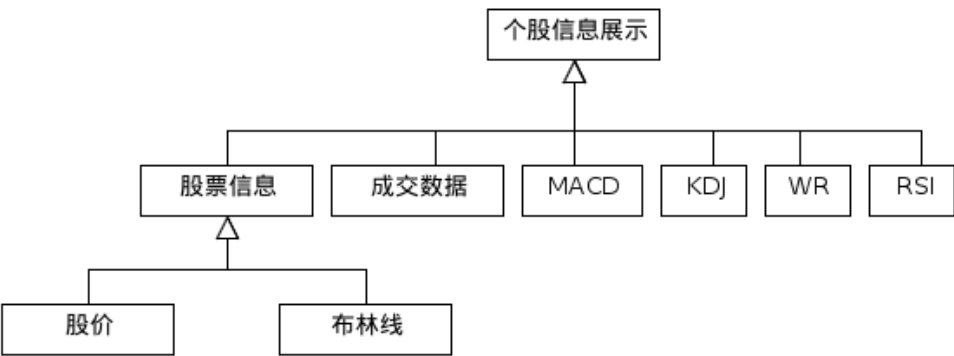


图 3-5 个股信息展示模块  
Fig 3-5 Stock Information Panel

3.3 系统平台流程分析与建模

根据上一节中所述，本文量化技术选股平台分为三个功能模块：个股信息、选股与分析。本节从系统的工作流程角度，进一步阐述平台的搭建。

3.3.1 时序图

如图 3-6 所示，系统根据用户、前后台与数据库共分为四个节点。系统首先为用户提供互联网接口，用户即通过浏览器对系统进行使用。在输入系统域名后，平台为用户提供搜查个股信息接口，用户输入个股股票代码或者股票名称，提交，即可获取个股信息。在进入个股信息平台后，即可自由选择量化个

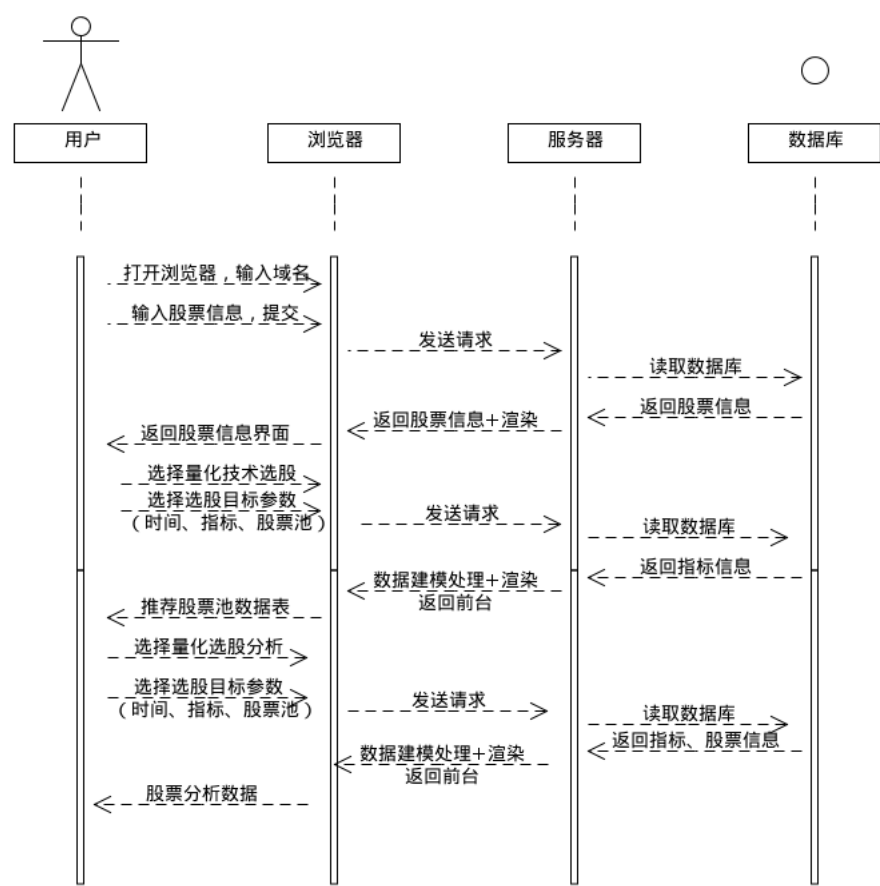


图 3-6 系统平台工作时序图

Fig 3-6 System Schedule Graph

股选选择还是量化技术指标分析。如在选择量化选股后，用户在设定完参数之后，系统服务器调用后台数据库信息，并进行处理，返回前台渲染并绘制表格，即展示给前台用户使用。

量化技术指标分析与量化选股流程类似，只是数据处理环节所涉及的算法

有所不同，最终展示在用户面前的数据信息也有所不同，关于此，章节 3.2 中有详细论述。

3.3.2 活动图

如图 3-7 所示，为量化选股平台活动流程图，平台以搜索界面为开始，引导用户搜索股票信息，首先向用户展现股价信息模块，然后在股价信息平台上，提供给用户量化选股模块、股价对比分析模块和股票信息搜索模块。

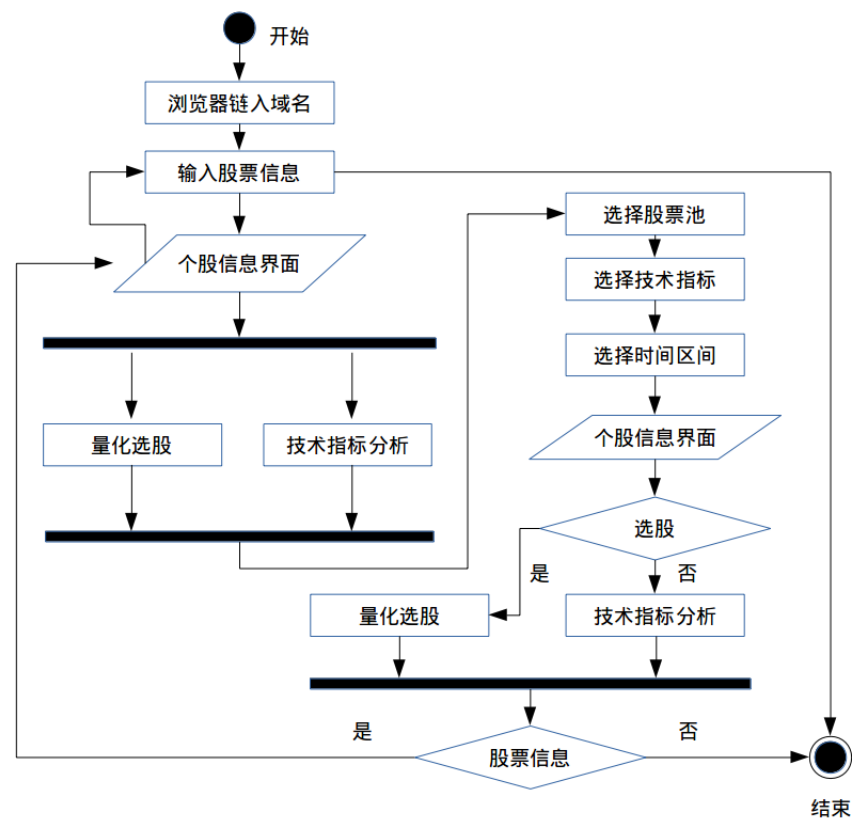


图 3-7 系统平台工作活动图

Fig 3-7 System Activity Graph

在用户获得选股结果，或者是技术分析结果后，允许用户直接点击股票名称，进入到股份展示模块，为及时体现并检测股价变化提供依据。

### 3.4 本章小结

本章首先提出了量化技术指标选股面临的一些问题，并对其解决方案作出分析；提出了构建四个功能模块的量化个股选择系统平台。四个功能模块分别为：数据处理模块、个股选择模块、技术指标分析模块与个股信息展示模块。

数据处理模块负责每日股票日数据的更新与技术指标的计算处理，数据库更新操作等；个股选择模块提供输入接口，即时间跨度、技术指标与股票池，输出为符合技术指标买入规则的推荐股票池；技术指标分析模块为解决技术指标及其组合有效性检验提供辅助工具；个股信息展示模块为用户提供最新的个股信息数据，包括股价信息及技术指标数值等。

## 第四章 量化个股选择系统平台实现

根据上一章中的系统需求分析，量化选股平台可以分为数据抓取、清理存储部分、MVC 架构服务器实现部分。其中，服务器架构部分都可细分为模型层（Model）、控制层（Control）和视图层（View）三个子部分。

在系统实现过程中，涉及到的技术有诸如 Python 数据挖掘、NodeJS 服务器搭建、HTML5、CSS、Javascript 前台实现等计算机技术支撑，这都将在本章中予以详细解释。

### 4.1 系统平台实现环境与技术

本文所涉及系统平台 Windows 操作系统进行研究开发，具体平台环境如下所示：

- 数据库：Sqlite 3.9.2
- 服务器：NodeJS v0.12.7
- 数据抓取：Python 2.7
- 网站框架：Express

涉及的主要技术有：网络数据抓取、数据清洗与存储、服务器搭建与算法优化、数据可视化等：

1. 网络数据抓取：主要使用 Python 包库所提供爬虫工具 Beautiful soup 进行相关公开金融数据 HTML 文本抓取、解析、存储等处理工作；
2. 数据清洗与存储：在数据抓取与实验过程中，存在一部分干扰数据，影响实验分析与结果，因此，在抓取与存储过程中进行了双重数据清洗工作，以保证数据的可靠性；
3. 服务器搭建与算法优化：实验涉及数据量大，需要对数据挖掘与分析算法进行进一步优化；

4. 可视化：采用轻量级、开源服务器 NodeJS 作为 WebWeb 服务器，引入百度开源项目 Echarts<sup>1</sup>，实现数据可视化。

4.2 系统平台架构设计

量化个股选择系统分为数据抓取、数据分析与存储部分、服务器部分与前台可视化部分。服务器框架采用 MVC 架构，分别模型层（Model）、控制层（Control）和视图层（View），结构与如图 4-1 所示：

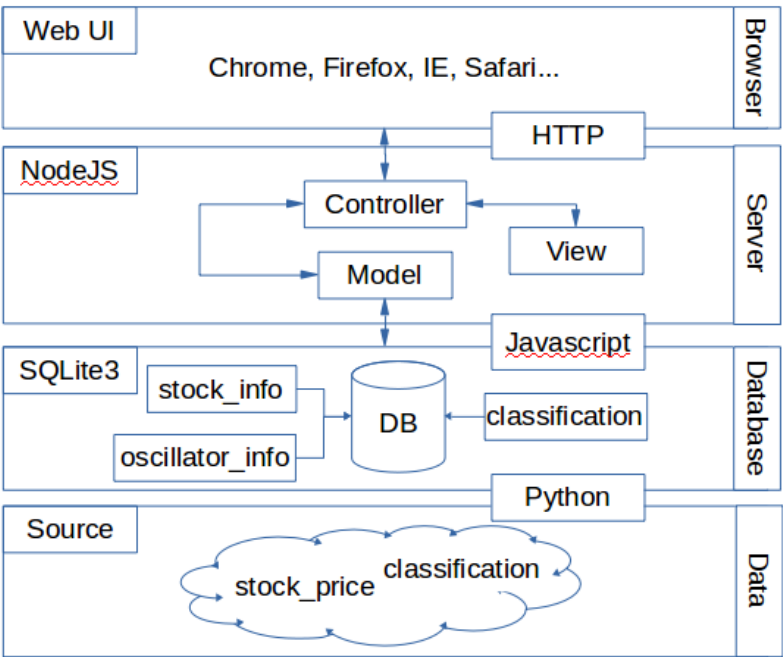


图 4-1 系统平台架构图  
Fig 4-1 System Structure Graph

1. 数据源与数据库部分：本文所使用数据源都为互联网上公开信息源，其中，数字信息源来自于新浪财经<sup>2</sup>，涉及目前主板市场上 2783 支股票，自

<sup>1</sup>ECharts, Enterprise Charts 为一款商业数据图表显示开源项目。最初，Echarts 是为了满足企业、公司等商业体系中的各种商业系统报表需求而产生，基于开源项目 Zrender，并为百度公司所优化，现被广大用户所采用。本文中数据的可视化手段即来自于开源项目 Echarts。

<sup>2</sup>新浪财经提供了日、分钟等数据源，通过网络可以进行爬去、解析，考虑到版权、侵权问题，这里声明文中所涉仅用于学术研究，不涉及任何商业利益。



2006 年至 2016 年 10 年间近 400 万条交易日数据，包括股票代码、股票名称、开盘价、收盘价、最高价、最低价、交易量、交易金额等；

- 2. 模型层 (Model): 后台与数据库交互处理部分，主要对控制层命令进行处理，与数据库交互、分析并优化算法，使使服务器可以快速响应；
- 3. 控制层 (Control): 前台服务器端，充分利用 NodeJS 事件驱动、异步和非阻塞 I/O 等特性，发挥其轻量和高效率的性质，处理与前端事件之间的连接；
- 4. 视图层 (View): 前台显示部分，采用 HandleBars 框架，为 HTML 模板进行可视化渲染；

4.3 数据抓取与数据库建立

4.3.1 股价原始数据抓取

在互联网金融发展背景下，越来越多金融数据在网上进行公开披露，如大盘指数、个股信息、公司财务等。为了进行科研实验，文章以非商业性质目的，抓取网上已经存在股票数据，进行项目研究。

如图 4-2 所示，为新浪财经网站，图中左侧为股份信息数据，右侧为其相对应的当日股份变动情况。股价信息具有一定商业价值，信息也相对滞后，这是文章仅以日为单位进行量化选股的主要原因。

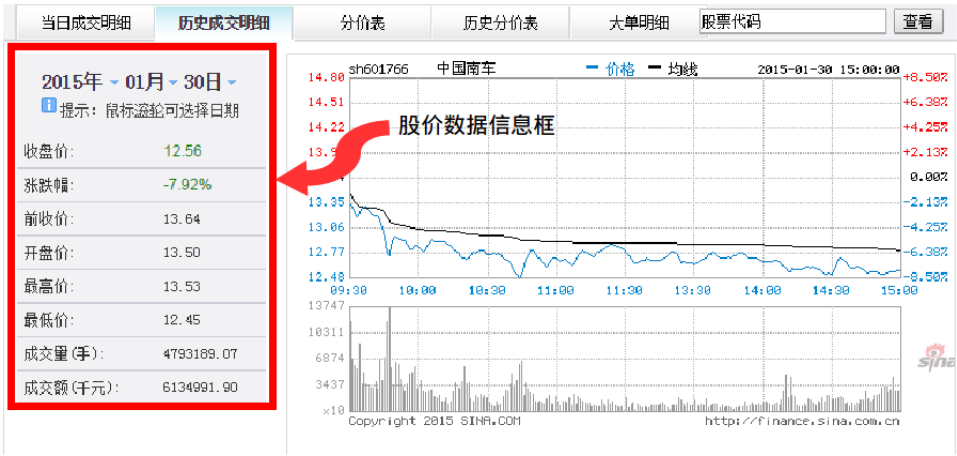


图 4-2 新浪财经股份信息版块  
Fig 4-2 Sina Stock Information Board

从网页面 HTML 源结构考察，如下源码模块所示，可发现股价信息版块具有固定范式。

```
1 <table>
2   <tr><td>收盘价:</td><td><h6> <!-- 位置序号 13 -->
3     <span style="color:#008000">12.56</span></h6></td></tr>
4   <tr><td>涨跌幅:</td><td><h6>
5     <span style="color:#008000">-7.92%</span></h6></td></tr>
6   <tr><td>前收价:</td><td>13.64</td></tr> <!-- 位置序号 17,28 -->
7   <tr><td>开盘价:</td><td>13.50</td></tr>
8   <tr><td>最高价:</td><td>13.53</td></tr>
9   <tr><td>最低价:</td><td>12.45</td></tr>
10  <tr><td>成交量(手):</td><td>4793189.07</td></tr>
11  <tr><td>成交额(千元):</td><td>6134991.90</td></tr>
12 </table>
```

开盘价、收盘价、最高最低价、成交量等数值都分布一个固定表格栏列表  $list = [<td>_1, <td>_2 \dots <td>_n]$  中，经计算，可知其数值分布在第 [14,16,18,20,22,24,26,28] 位置上，符合，Python Beatifulsoup 的抓取规范，因此，通过对相应位置的文本进行抓取，即可得到文章需要的一系列数据。

图 4-3 显示了 Python 在调用 Beatifulsoup 抓包进行抓取的过程：

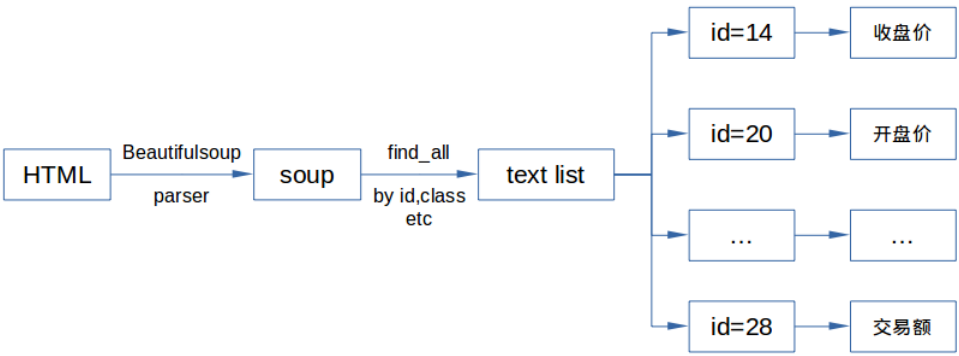


图 4-3 股价信息抓取过程  
Fig 4-3 Sina Stock Prices Crawling Process

其中涉及到核心程序源码如下所示：

```
1 ...
2 soup = BeautifulSoup(html , "html.parser") # soup为抓取包下的一个类
3 text = soup.find_all('td') # 其通过解析html结构文件，生成DOM结构类
```

```
4                                     # 方便开发者进行文本挖掘工作等;
5     for l in text:
6         if (i == 14):
7             close_price = l.text.encode('gbk','ignore');
8             ...
```

在程序进行数据抓取的过程中，为减少网络阻塞、延迟，提高抓取速度，程度模拟多进程处理模式，对抓取过程进行优化，如图 4-4 所示，为多进程抓取流程，程序优化实行了“抓取 ⇒ 锁死 ⇒ 结束 ⇒ 释放”方法，开启多个进程，对股票进行同时抓取。

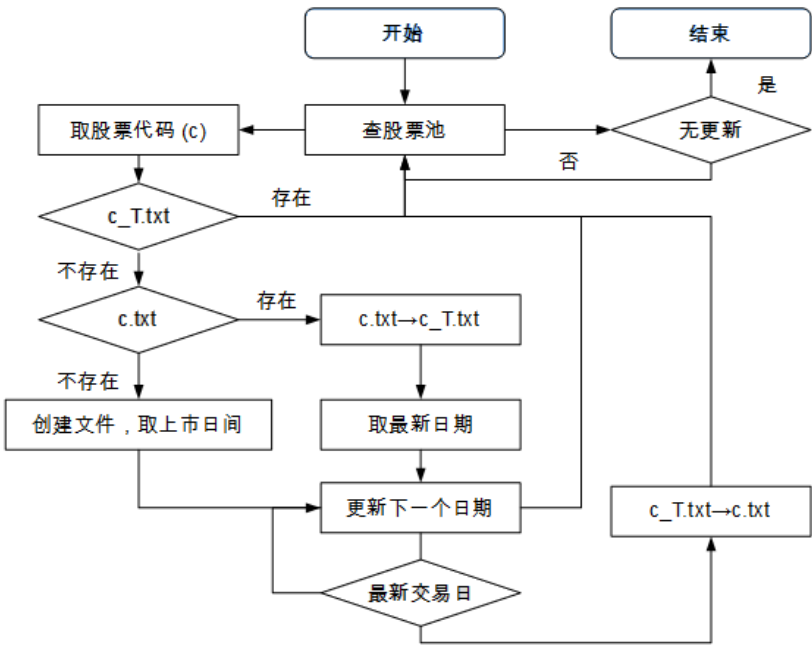


图 4-4 多进程抓取股票数据流程图

Fig 4-4 Stock Infomation Crawl Process Multi-Process

核心代码如下所示：

```
1 NEW_PATH = PATH + 'Folder/' + CODE + '_' + info[2]; # 股票代码+上市日期
2 if os.path.exists(NEW_PATH) or os.path.exists(NEW_PATH + '_p'):
3     continue; # 如果已经抓取或者正在抓取，则进行下一支股票的抓取工作
4 file = open(NEW_PATH + '_p', 'w'); # 否则新建进程，进行抓取
```

4.3.2 建立数据库

在选择数据库时，考虑到数据库必须具备诸如跨平台、轻型、开源和尽可能少地占有本地资源等特性，最终选择了 SQLite。SQLite 是一个开源的数据库，支持 Windows、Mac OS、Linux、Unix 等主要现代操作系统 [32]。同时，SQLite 也可以与许多程序语言相融合，如 Python、C++、Java 等，还具有 ODBC 接口，语法规则并继承支持 SQL 大量语句等。相较于 Mysql、PostgreSQL 两款开源数据库管理系统，它的处理速度更快。

本文主要涉及两张个数据库表，一个为原始数据信息表 stock\_info，一为指标数据信息表 oscillator\_info。其中，stock\_info.db 包含字段为股票代码号 (code)、数据日期 (date)、开盘价 (open\_price)、最高价 (high\_price)、收盘价 (close\_price)、最低价 (low\_price)、成交量 (volume) 和交易金额 (value)，oscillator\_info.db 包括所涉指标数据信息，表格信息如图 4-5 所示：

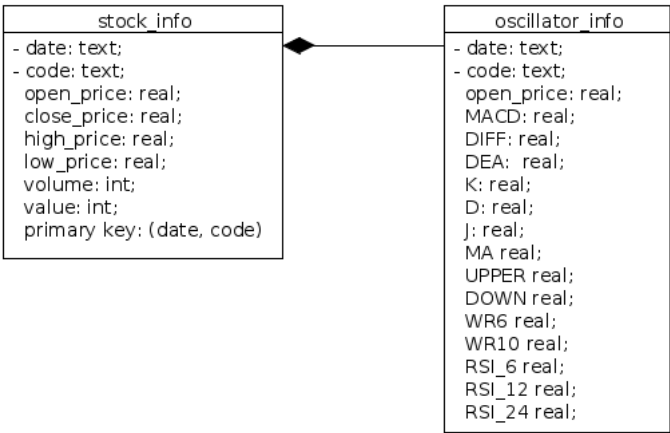


图 4-5 数据库表格信息  
Fig 4-5 Stock Information Database Table

创建表格 stock\_info 核心代码如下：

```
1 table_name = 'stock_info';
2 # create table command line.
3 cu.execute("create table " + table_name + " (code text , date text , open_price
    real , high_price real , close_price real , low_price real , volume integer ,
    value integer , primarykey (code, date))");
```

数据库建完成后，将原始股价数据存入数据库。与此同时，对股票数据进行第一次清洗，即对非开市日期，仍然有的股票信息进行清洗：对于获取的股票信息，解析股票信息数据，并将每支股票、每天所对应的股份存储至数据库中。

在接下来的数据更新部分，将进一步介绍抓取部分的清洗工作。

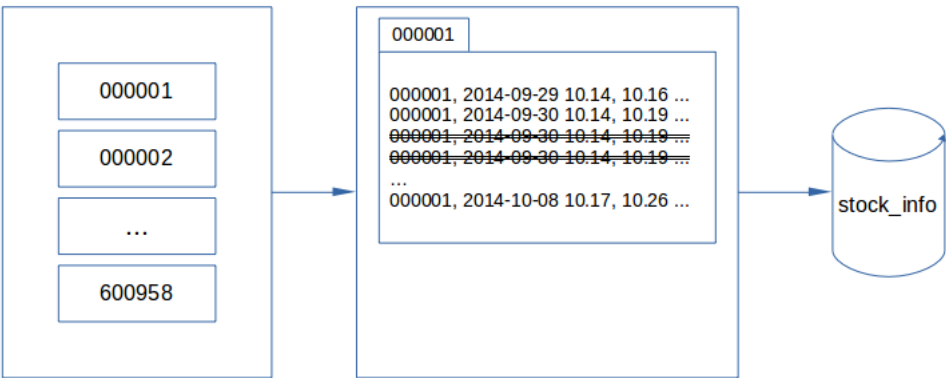


图 4-6 原始数据清洗

Fig 4-6 Original Stock Information Clean up

数据表 stock\_info 存储之后，根据清洗后股价信息数据，计算各项计算指标数值，其计算公式在第二章中已经介绍，本章节中仅给出部分指标的计算核心代码。

1. MACD, DIFF、DEA 计算过程:

```
1 EMA12 = 2*value/13+11*float(initial_EMA12)/13; # 12日均值
2 EMA26 = 2*value/27+25*float(initial_EMA26)/27; # 26日均值
3 DIFF = EMA12-EMA26; # 快均值线减去慢均值线
4 DEA = initial_DEA*8/10.0+DIFF/5.0; # 慢平滑移动平均线
5 MACD = 2*(DIFF-DEA); # 柱值
```

2. KDJ 计算过程:

```
1 RSV = 100 if (HHV - LLV ==0 ) else (value - LLV)*100/(HHV-LLV)
2 K = 2* initial_K /3.0 + RSV/3.0; # RSV 未成熟随机值
3 D = 2* initial_D /3.0 + K/3.0;
4 J = 3.0*K - 2.0*D;
```

### 3. 布林线计算过程:

```

1      MA = sum(value_list)/len( value_list ); # 均值线
2      UPPER = MA+2*np.std(value_list) # 上轨线
3      DOWN = MA-2*np.std(value_list) # 下轨线

```

最终，将技术指标数据信息存储入 SQLite3 数据库中。

```

1      cu.execute(" insert into oscillator_info (code, date, "
2                "MACD, DIFF, DEA, K, D, J, MA, UPPER, DOWN,"
3                "WR6, WR10, RSI_6, RSI_12, RSI_24) values"
4                "(?,?,?,?,?,?,?,?,?,?,?,?,?,?) ", information );

```

#### 4.3.3 数据更新及清洗

##### 数据更新

在原始数据库搭建后，为保证数据及时性，需要每日对股票信息进行更新，更新部分程序直接与数据库交互。其操作流程如图 4-7 所示：

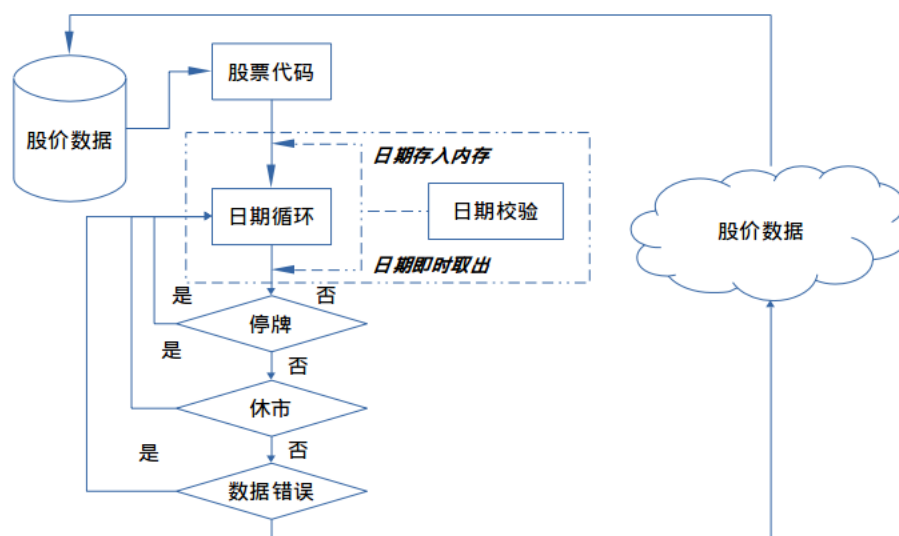


图 4-7 多进程抓取股票数据流程图

Fig 4-7 Stock Infomation Crawl Process Multi-Process

与原始数据抓取方法相似，二者都采用多进程操作处理，减少网络延迟，

不同之处在于，数据更新部分直接将网络获取数据存储至数据库中，而非中转至硬盘上。

### 算法优化与比较

在这里，针对股票当期数据是否已经存入数据库，进行了一次算法的优化与改进。初始检验将某支股票所有已存在日期存入内存进行验重；改进后算法采取逐个日期与数据库循环检验。

图 4-8 给出了两种算法随机更新 25 支股票后所给出的时间比较结果。

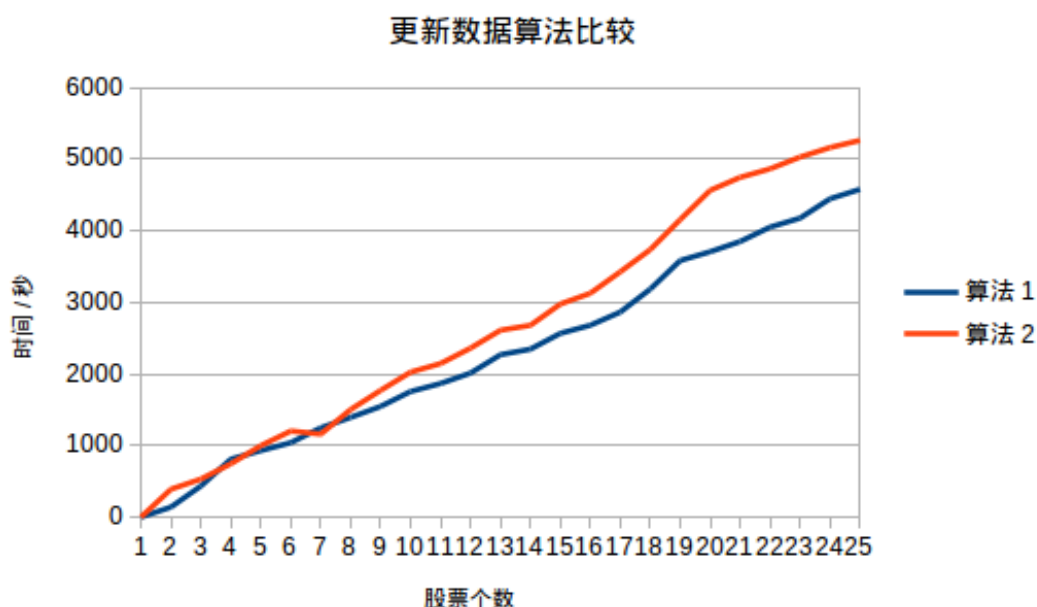


图 4-8 数据更新算法优化比较

Fig 4-8 Stock Information Updating Algorithm Comparison

从图中，可以得出结论：数据存入内存后验重效率要低于循环检验。有两点原因可以解释：1. 多进程前提下，数据存入内存占用空间大；2. 即使存入内存中，在比较验重时，仍然需要对数据进行查找，与直接查询数据库并无差异。

### 异常处理

在数据更新与抓取过程中，经常会遇到诸如网络丢包与中断等问题，为了保证数据的完整性，程序在运行时采取异常处理机制，并将错误股票代码保存

在硬盘上，在每日数据更新后，进行后期循环再处理，对异常行为进行补充。

```
1      except:
2          print 'HTTP error'
3          file . write (' HTTP error '+ dateToStr(i_DATE)+' ' + CODE)
```

4.4 基于 NodeJS 下的 MVC 服务器架构

Node.js 是一个基于 Chrome V8 引擎，具有无阻塞 (Non-Blocking) 和事件驱动 (Event-Driven) 等的特色，用于快速搭建快速响应、易扩展的网络应用。得益于事件驱动，非阻塞 I/O 模型，NodeJS 平台上搭建的服务器非常适合在分布式设备中实现数据密集型实时应用。本文即通过使用 Express 框架实现了基于 NodeJS 的 MVC 服务器架构。

图 4-9 描述了整个系统的文件结构：

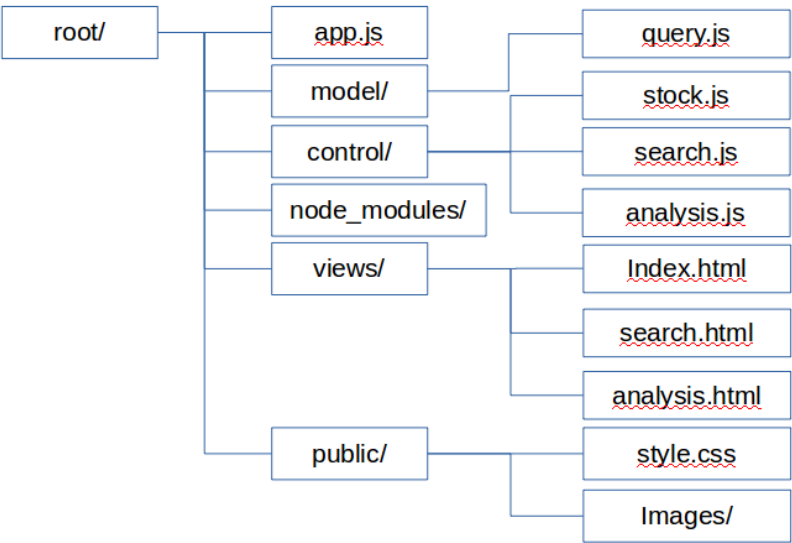


图 4-9 系统文件结构  
Fig 4-9 File System

系统主要由运行程序 app.js，模型层（model）、控制层（control）、模块包（node\_modules）、视图层（views）、公共文件（public）组成，其中，模型层、控制层与视图层响应了 MVC 结构架构。

各个部分主要功能有：



- `app.js`: 启动文件, 负责与控制层进行连接, 并对视图层进行渲染;
- `model`: 模型层文件, 负责访问并操作数据库;
- `node_modules`: 第三方模块包库, 提供模块化接口, 如 `express`;
- `views`: 视图层, 前台界面显示部分, 为主程序提供渲染框架模板;
- `public`: 公共文件夹, `css` 文件、图像文件都在此文件夹中;

图 4-10 显示了服务器运行期间, 各个模块的运行流程图: 用户通过 HTTP 向服务器发起请求响应, 主程序 `App.js` 接受来自 HTTP 的访问请求, 解析请求并通过模型层访问数据库, 结果传递给控制层, 并通过 `App.js` 渲染并返回给浏览器。

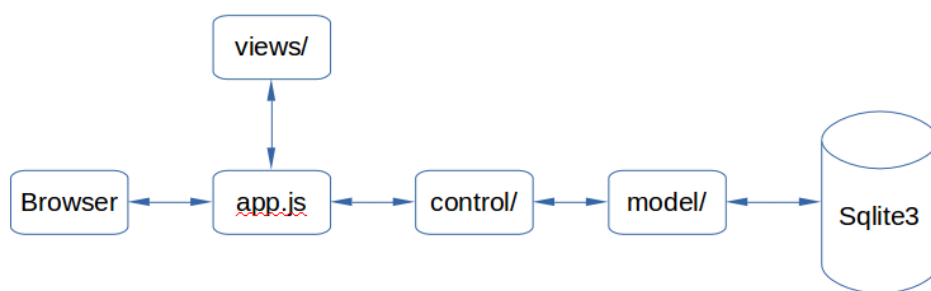


图 4-10 服务器响应流程

Fig 4-10 Web Server Response Progress

#### 4.4.1 数据接口, 即模型层 (Model)

作为服务器底端, 模型层连接起服务器控制层与数据库, 如图 4-11 所示:

控制层向模型层发出增、删、改、查指令后, 模型层执行数据库命令, 并向控制反馈操作结果。

由于 NodeJS 采用了异步处理方式, 因此, 在构建模型层时, 需要引入回调函数 (Callback), 来调用数据库操作所返回的结果。

如下述代码所示, 通过访问本地数据库 `stock_info`, 系统返回给控制层当前查询指令所得到的所有数据结果。

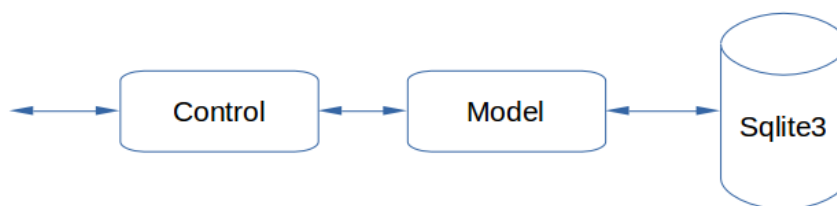


图 4-11 系统模型层

Fig 4-11 Part Model of the System

```

1  "use strict ";
2  var sqlite3 = require('sqlite3').verbose();
3  var database = 'D:/SJTU/MasterDegree/databases/stock_info.db';
4  exports.connect = function(request, callback){
5      var db = new sqlite3.Database(database);
6
7      console.log("readAllRows lorem");
8      db.all(request, function(err, rows) {
9          callback(rows);
10         db.close();
11     });
12 }

```

#### 4.4.2 控制层 (Control) 实现

控制层实现运行程序与模型层的连接，如图 4-12 所示，整个服务器的主要逻辑判断部分都在控制层部分。如第二章节中对系统功能分析的论述，控制层分为三个子模块：股价数据展示、量选选股模块、量化技术指标分析模块。

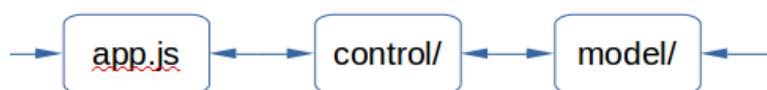


图 4-12 系统控制层

Fig 4-12 Part Control of the System

## 股价展示模块

股价展示模块允许通过部分字段对股票进行搜索，控制层通过对万得提供的股票代码、名称对照表进行检索，向前台提供相对应的股票信息，如果被检索的信息与目标投票数据双方一方包含另一方，则视为检索目标。检索实现过程如下所示：

```

1      var code_file = 'stock_name_code.txt'; // 所有股票信息
2      ... // code_list [0] = 000001 平安银行 e.g.
3      for (var i = 0; i < code_list.length-1; i++){
4          var code_info = code_list[i].split(' ');
5          if (code_info[0].indexOf(code_pre)> -1 ||
6              code_pre.indexOf(code_info[0])> -1){
7              code = code_info[0];
8              name = code_info[1]; // 如果股票代码一方包含另一方
9              break;
10         }
11         if (code_info[1].indexOf(code_pre)> -1 ||
12             code_pre.indexOf(code_info[1])> -1){
13             name = code_info[1];
14             code = code_info[0]; // 如果股票名称一方包含另一方
15             break;
16         }
17     }

```

除了股票信息外，系统还接收股票信息时间跨度，以缩小数据规模，提高效率，默认情况下搜索时间为 2015 年 1 月 1 日至今：

```

1      var From = req.query.from, to = req.query.to;
2      if (typeof(to)=='undefined' || to=='') to = '2018-01-04';
3      if (typeof(From)=='undefined' || From=='') From = '2015-01-01';
4      var time_condition = 'and date > "'+From+'" and date < "' + to + '"';

```

接下来，通过异步嵌套访问数据库的 stock\_info 与 oscillator\_info 数据表，从后台调用股价信息，并返回给前台渲染展示，调用 model 引擎、数据处理核心如下所示：

```

1      var engine = require('./model/stock_info'); // 调用 model 层引擎；
2      var stock = new engine(request_stock, function(rows){
3          var osEngine = require('./model/stock_info'); // 再次调用；
4          var osInfo = new osEngine(request_oscillator, function(rows){

```

```

5           for (var i = 0; i < rows.length; i++){
6               date_list [i] = rows[i]['date'];
7               ...
8               close_price [i] = rows[i]['close_price'];
9           } // rows<-stock_info
10
11          for (var i = 0; i < rows_.length; i++){
12              MACD[i] = rows_[i]['MACD'];
13              MA[i] = rows_[i]['MA'];
14              ...
15              WR10[i] = rows_[i]['WR10'];
16          } // rows_<- oscillator_info

```

最后，通过渲染并返回给前端展示页面，这部分将在运行程序 `app.js` 中详细介绍：

```

1      res.render('selection',{ title:'量化选股', stocks:target_stock , From:From, to:to
      , amount:target_stock.length , message:messages});

```

### 量化选股模块

量化选股模块为系统的核心部分，主要实现针对不同技术指标——单指标或者组合指标，在一定时间区间内符合金叉的所有股票列表。控制层从前台接收两类参数：技术指标组合与时间区间。

接收参与的时间处理上，与股价展示模块相似，这里不再赘述，技术指标组合，正常情况下为接收参数，特殊情况，如没有技术指标参数传回，则默认以 **MACD** 单技术指标为核准。

```

1      var macd = req.query.macd==='true';
2      var kdj = req.query.kdj==='true';
3      var rsi = req.query.rsi==='true';
4      var wr = req.query.wr==='true';
5      var boll_up = req.query.boll_up==='true';
6      if (!(macd || kdj || rsi || wr || boll_up) ) macd=true;

```

选股算法上，有两种方案，方案一，顺序对指标进行检验，将数据存入内存，剔除法，依次去除不符合金叉数据，方案依据，对金叉数据节点比较容易控制，缺陷：操作次数多，效率低下。方案二，非剔除法，并行处理技术指标，

仅对金叉数据进行操作，方案依据，减少重复计算开支，提高效率，缺陷：增加内存支出。

如图 4-13 所示，图（a）为顺序判断指标是否为黄金交叉，图（b）为并序计算，

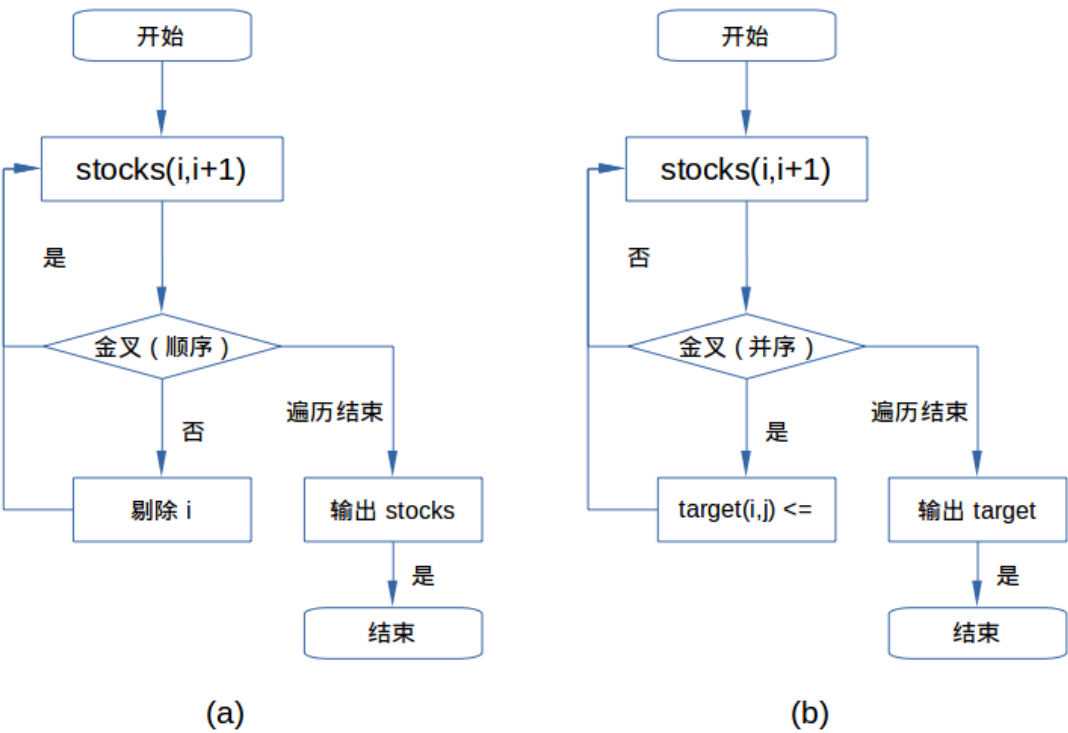


图 4-13 量化选股算法比较

Fig 4-13 Algorithm of Selection Stocks Comparison

其核心代码如下：

```
1      if (macd && !check_macd(stock_info)) {
2          stock_info . splice (i-1,1);
3          i=i-1;
4      }
5      if (kdj && !check_kdj(stock_info)) {
6          stock_info . splice (i-1,1);
7          i=i-1;
8      } // (a) 顺序计算
```

```

9      if (check_macd(stock_info) && check_kdj(stock_info)
10         && check_rsi(stock_info) && check_wr(stock_info)
11         && check_boll(stock_info)) {
12         target_stock[j++] = stock_info[i-1];
13         target_stock[j++] = stock_info[i];
14         i = i + 1;
15     } // (b) 并序计算

```

通过测试，结果证实并序计算与顺序计算相比，在周期 15 天内，顺序效率要高于并序，但是当周期超过 15 天，顺序效率大幅落后于并序计算的结果，如图 4-14 所示：

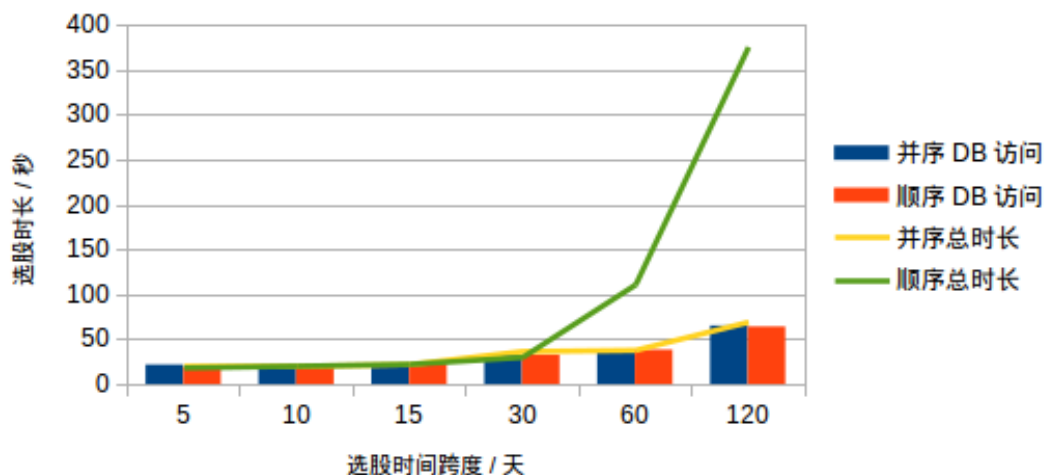


图 4-14 并序与顺序算法比较结果

Fig 4-14 serial and Parallel Algorithm Comparison Result

从图中可以得出结论，在数据读取上，二者基本没有差异，但在计算方法上，由于并序计算量少于顺序，因而随着数据量的增大，其优势也逐渐显现出来。

#### 量化技术指标分析模块

基于量化选股模块中的分析结论，并序计算效率在通常情况下要远远高于顺序计算。因此，分析模块继承选股模块，对所选金叉股票进行分析平台搭建。

不同于常用量化交易中以基准线为参照, 本文通过计算、统计各指标金叉出现时未来特定交易日的股价变动状况, 来研判技术指标的有效性。

与量化选股相同, 分析模块输入部分也分为时间区间与指标参数。不同于选股模块, 分析模块除了展示出所有出现金叉时间节点外, 还通过 Javascript 在前台绘制出了涨跌幅分布区间, 利于分析指标的有效性并适当规避风险。

统计部分, 后台统计了金叉出现后 30 天内的所有涨跌幅情况, 如下所示:

```

1      function analysis (i,j, stock_info ,count){
2          var target_stock = stock_info [i];
3          if (i+count > stock_info .length-1){
4              target_stock ['close_price_'+count.toString ()] = -1;
5              target_stock ['close_price_'+count.toString ()+' _rate '] = -101;
6              return target_stock ;
7          } # 如果未来 count 天超出选定日期, 则取默认值并退出;
8          target_stock ['close_price_'+count.toString ()] =
9              stock_info [i+count]['close_price '];
10         target_stock ['close_price_'+count.toString ()+' _rate '] =
11             (( stock_info [i+count]['close_price '] - stock_info [i]['close_price '])
12              / stock_info [i]['close_price '] * 100.0).toFixed (2);
13         return target_stock ; # 计算未来 count 日期的股价变动情况;
14     }
15
16     analysis_list = []
17     for (var i = 0; i < 30; i++){
18         analysis_list .push(i+1);
19     }
20     for (var i = 1; i < stock_info .length; i++){
21         if ( check_oscillators ()) {
22             for (var k = 0; k < analysis_list .length; k++)
23                 target_stock [j] = analysis (i,j, stock_info , analysis_list [k]);
24             j = j+1;
25             i = i + 1;
26         } # 对每个金叉股票进行统计计算, 计算对象为未来0~30天的数据情况;
27     }

```

#### 4.4.3 启动程序 app.js

下面部分代码为 express 创建 web 服务器过程和路由功能, require() 命令调动包库, 第 4 行 app.set() 设置 views 目录, 即设置服务器前台框架渲染的目

录地址；第 5 行为框架结构名称，即 **HandleBar** 框架。这在下一节会详细介绍。  
`app.use()` 设置公共目录地址。

```
1  var express = require ('express');
2  var app = express ();
3
4  app.set (' views', path.join (__dirname, 'views'));
5  app.set (' view engine', 'hbs');
6
7  app.engine ('html', hbs.__express);
8
9  app.get ('/', function (req, res) {
10      var output = new searchEngine(iconv.encode(keyword,'utf-8'),
11          history, period, resource);
12      cept = output.getConcept();
13      stocks = output.getStocks ();
14      res.render('index',{ title : title , concept:cept, stocks: stocks })
15  });
16
17  app.use(express.static (__dirname + '/public'));
18  app.listen (80);
```

通过在命令行执行下述命令，即可运行服务器。

```
1  supervisor app.js
```

#### 4.4.4 前台可视化实现，即视图层，View

如前文中所述，选股系统对照 MVC 框架，分为控制层、视图层和模型层，在这里，视图层（Views）即为 `views` 文件夹里的文件，则为前台 HTML 文件提供了一系列模板。前台还采用了 NodeJS-Express 支持的 **HandleBars**（HBS）模型框架，降低开发成本。

##### 前台技术介绍

前台视图层结构如图 4-15 所示，渲染过程中，控制器先行遍历 `header.html`, `layout.html` 与 `footer.html`，如果文件存在，则通过 `layout.html` 中 `{{body}}` 部分，引用目标文件，即 `search.html`, `index..html` 等模板文件。

网站框架结构 `layout.html` 如下所示：



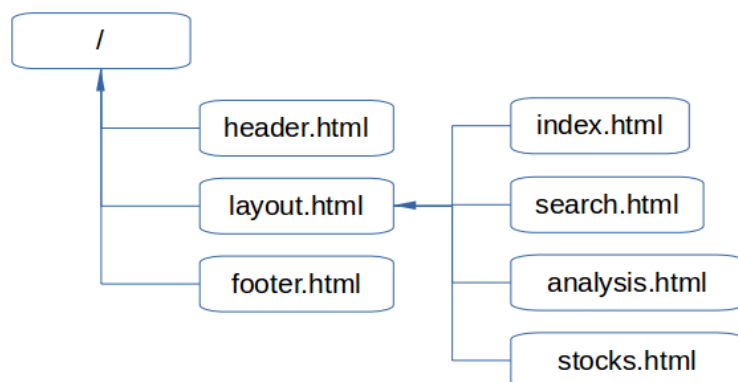


图 4-15 视图结构图

Fig 4-15 View Part Structure

```

1  <head>
2  < title >{{ title }}</ title >
3  </head>
4
5  <body>
6      {{{header}}}
7      {{{body}}}
8      {{{ footer }}}
9  </body>

```

{{header}} 如下代码所示，由每个不同功能目标页面提供模块信息：

```

1  {{#each menus}}
2  <div class='menus'>
3      <a class="{{{menu}}}" href="#">{{{name}}}</a>|
4  </div>
5
6  <div id='search'>
7      <form action="search" name="myform">
8          <a href="http://localhost">
9              
10         </a>
11     </form>
12 </div>

```

其中，*#each menus* 循环访问由后台提供的字典类型 *menus*，如下所示，为一 *menus* 样例：

```

1      menus = {
2          {'name':'量化选股','menu':'menu hover'},
3          {'name':'技术分析','menu':'menu'},
4          {'name':'股价展示','menu':'menu'}
5      }

```

文中提到，在数据图表渲染上，使用百度开源工具包 Echarts，其模板及使用方法如下所示：

```

1  <script type="text/javascript">
2      require.config({ // 配置文件来源目录;
3          paths: { echarts: "http://echarts.baidu.com/build/dist" });
4      require (
5          [ 'echarts',
6            'echarts/chart/bar' ], // 调用所需图表类型，有 bar, line, k等;
7          function (ec) {
8              // option 为数据内容，类型是字典型；数据内容、属性皆在此配置;
9              option['series']=[];
10             for (var i = 0; i < Option[0]['series'].length; i++){
11                 option['series'][i] = {
12                     name: Option[j]['legend'][i].parseInt((Option[j][k],
13                     itemStyle:{normal:{color:Option[j]['series'][i]['color']}}},)}
14                 // 找到 html 中展示图表位置，并渲染绘制
15                 myChart[j.toString()] =
16                     ec.init (document.getElementById('main'+j.toString()));
17                 myChart[j.toString()].setOption(option);
18             }
19         </script>

```

### 前台页面展示效果

页面展示部分主要一个主页，三个功能页，它们分别是：平台主页，个股股价展示页、量化个股选择页、量化技术指标分析页面。

如图 4-16 所示，系统平台主页面包括平台 logo、搜索框、功能模块三个子部分。其中，第二子部分提供个股信息展示搜索接口，允许使用关键字搜索个股股价信息；第三子部分提供功能链接，允许直接进入量化选股模块与量化技术指标分析模块。

图 4-17 为股价展示页面。页面共计 4 个子部分组成，即搜索框、功能菜单、条件框和股价信息显示框：



图 4-16 量化选股平台主页

Fig 4-16 QA System Index

- 1. 搜索框：提供搜索股价基本信息的接口，与主页搜索框功能相同；
- 2. 功能栏：提供三个功能模块的链接地址，允许直接进入选股与分析模块；
- 3. 条件栏：提供搜索的边界条件接口，侧栏为股价信息部分仅提供时间与股票窗口条件窗口；
- 4. 内容块：显示股价基本信息，包括股价信息（开盘价、收盘价、最高最低价等）与相关技术指标（MACD、KDJ、RSI、WR、BOLL 线等）；

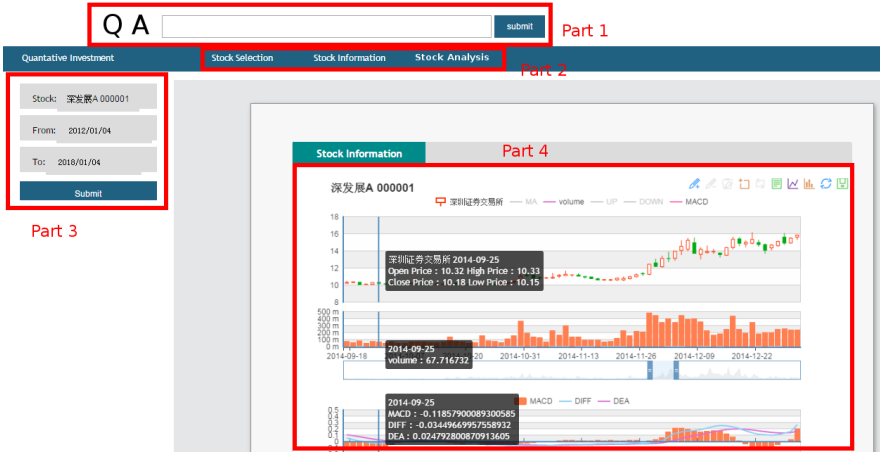


图 4-17 股价展示页面

Fig 4-17 Stock Information

图 4-18 为量化技术指标分析界面，界面共由四个部分组成：搜索框、功能框、条件框、内容块，其中内容部分又分为股价趋势分布、涨跌对比饼图与原始买点数据。

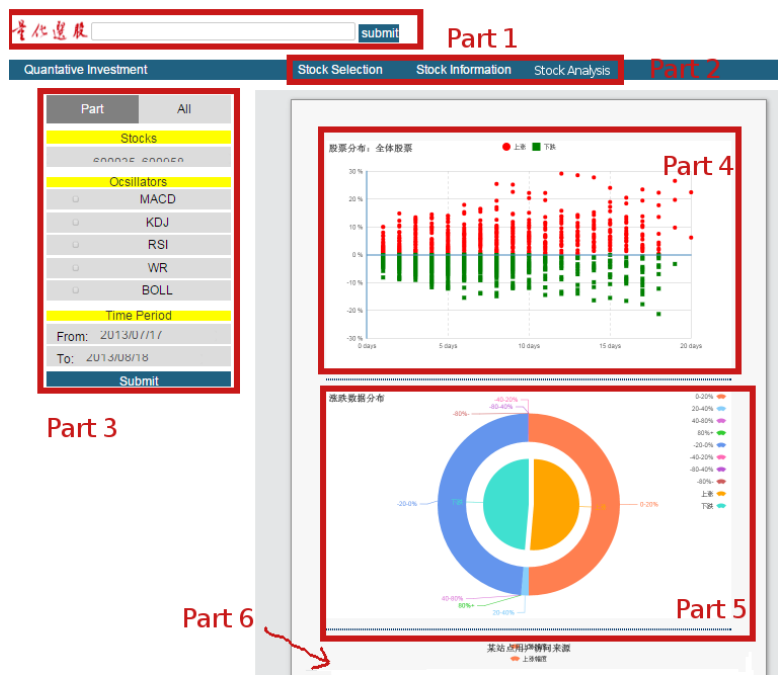


图 4-18 量化持械技术指标分析页面

Fig 4-18 Stock AAnalysis

1. 同股价展示模块；
2. 同股价展示模块；
3. 条件栏：提供搜索的边界条件接口，侧栏为指标分析部分提供
  - 股票池选项，即自定义股票池或者全体股票池；
  - 技术指标复选框；
  - 分析时间区间；
4. 内容块：
  - 股价趋势分布图：为散点图，通过数据统计与 Echarts 模块包，渲染出股票在出现金叉点后未来 30 天内的股价分布；
  - 涨跌对比饼状图：提供股价涨跌对比分析明细图，饼图分为内圈与外圈两部分，其中内圈对比在时间区间内的涨与跌对比关系；外圈展示涨跌对比明细；

– 原始数据表单：提供上述两点图形的原始数据来源，作为参考，其格式如图 4-19 所示：

Code	Stock Name	Buying Signal	Close Price	3 days	6 days	9 days	12 days	15 days	20 days	30 days
000014	沙河股份	2013-08-06	8.96	-1.45	2.57	-101	-101	-101	-101	-101
000021	深科技	2013-08-01	5.39	3.53	2.04	2.23	-101	-101	-101	-101
000024	招商地产	2013-08-01	27.46	-101	-101	-101	-101	-101	-101	-101
000033	xST新都	2013-08-06	3.75	1.07	2.67	-101	-101	-101	-101	-101
000043	中航地产	2013-08-01	6.29	-0.32	5.25	10.97	-101	-101	-101	-101

图 4-19 量化技术指标分析数据来源

Fig 4-19 Stock AAnalysis Data Source

图 4-20 为量化选股页面，除了搜索与功能链接，选股界面提供：

1. 条件框：与量化分析相同；
2. 内容块：量化选股结果，显示出股票代码、股票名称、买入信号点等：

Part

All

Stocks

600035, 600958

Oscillators

☐ MACD

☐ KDJ

☐ RSI

☐ WR

☐ BOLL

Time Period

From: 2014/01/04

To: 2014/01/10

Submit

Part 1

Code	Stock Name	Buying Signal	Open Price	Close Price
000038	深大通	2014-01-07	12.2	11.93
000048	康达尔	2014-01-08	10.84	11.83
000597	东北制药	2014-01-06	7.03	7.3
000721	西安饮食	2014-01-07	6.7	6.67
002076	雪莱特	2014-01-07	8.24	8.34
002104	恒宝股份	2014-01-07	15.6	16.82
002120	新海股份	2014-01-06	8.16	8.32
002208	合肥城建	2014-01-06	6.66	6.63
002242	九阳股份	2014-01-08	11.31	11.1
002253	川大智胜	2014-01-07	26.5	26.59
002268	卫士通	2014-01-08	27.9	30
002298	鑫龙电器	2014-01-06	7.43	7.21
002370	亚太药业	2014-01-08	17.82	17.62
002414	高德红外	2014-01-08	23.2	24.05

Part 2

图 4-20 量化选股页面示页面

Fig 4-20 Stock Selection

## 4.5 本章小结

本章详细阐述了量化选股系统的实现过程，包括后台数据抓取与数据库建立、NodeJS 服务器搭建以及 MVC 模型架构的实现。在系统搭建的过程中，多次涉及到多进程处理、内存算法优化等问题，并提出项目优化方案，提高平台性能。

在数据抓取过程中，通过引入多线程等方法，平行地抓取网络数据，既充分利用了网络带宽，又缩小了 IO 与内存、CPU 之间的运行速度差；清洗过程采用双通道清洗，即抓取过程清洗与数据库冗余清洗。

在 Web 服务器方面，采用了目前市场上新起、流行、开源的服务器 NodeJS，并使用 Express 网站构架，通过异步处理机制，实现轻便、快捷的服务器响应。

## 第五章 量化选股系统平台测试

章节三与四设计并实现了量化个股选择系统，即量化技术指标选股、个股信息展示与量化技术指标分析平台。测试环境如章节四中所述，本章将对系统平台进行测试，并比较不同平台之间的优缺点与差异。其中，个股信息展示模块通过与同花顺 PC 端平台进行比较检验；量化选股与技术指标分析通过抽样分析进行检验。

### 5.1 个股信息展示功能模块测试

本节主要针对个股信息展示功能正确性进行测试，由于技术指标的计算根据计算日期前后相互依赖，并且指标信息又依赖于股价信息，因此，指标信息正确，股价信息即正确，相反，股价信息的正确性是保证指标正确的前提。所以，本功能模块仅需要针对最新的数据信息进行对比测试，即可判断信息的正确性。

如图 5-1 所示，为对比测试结果，其中，图中（a）为系统股价展示平台，图中展示了国泰君安在 2016 年 1 月 22 日的股价信息，图中（b）为同花顺股价信息。从图中，提取出有效信息如表 5-1 所示：

表 5-1 测试样例 I 信息提取  
Table 5-1 Testing Sample I

	本文系统	同花顺
MACD	-0.32268087	-0.323
DEA	-1.06327055	-1.06
DIFF	-1.22461098	-1.22

对比 MACD 技术指标数据信息，可知股价信息在前期数据清洗与处理上是与同花顺一致，在假定同花顺信息正确的基础上，可以得出股价信息在数据清洗与处理上的的正确性。

图 5-2 为个股信息展示平台在 KDJ、RSI 和 WR 技术指标的参数数值。表 5-2 中为图中提取出的有效信息，对比即可验证技术指标的正确性。



图 5-1 个股信息展示页面 I

Fig 5-1 Stock Info Board I

表 5-2 测试样例 II 信息提取

Table 5-2 Testing Sample II					
	本文系统	同花顺		本文系统	同花顺
K	20.76722541	20.77	RSI 6	47.7784832	47.78
D	30.37408862	30.37	RSI 12	57.0416849	57.04
J	1.553499017	1.56	RSI 24	57.7768072	57.78
WR 6	68.783068783	68.78	WR 12	86.342592592	86.34

篇幅限制，更多测试结果详见附录 A。

5.2 量化选股功能模块测试

本小节通过测试两个不同时间段，不同技术指标组合的选股策略，对量化技术指标选股功能模块进行正确性测试，在推荐的股票池中，通过随机抽样方式对结果进行结果检验。

5.2.1 案例 I：MACD-KDJ 组合

案例 I 选取量化技术指标组合：MACD+KDJ，时间范围为 2015 年 12 月 1 日 -2015 年 12 月 8 日，目标股票群体为全体股票池。选取结果如图 5-3 所示：



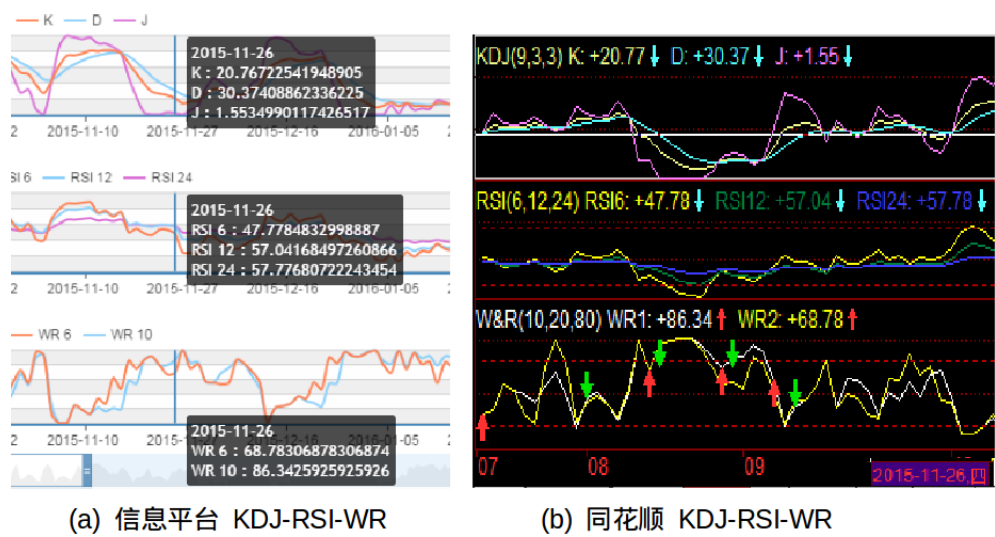


图 5-2 个股信息展示页面 II

Fig 5-2 Stock Info Board II

图 5-3 给出了选取结果按股票代码排序后的结果，对此，我们随机抽取四支股票，分别为：奥飞动漫、黑牛食品、山东海化、石化机械。比照同花顺中的技术指标，进行选股的正确性判断，判断结果如图 5-4 所示：

由图中可以看出，四支股票给出的日期都是 MACD 与 KDJ 同时出现金叉后的第一天。对比分析检验了 MACD、KDJ 指标选股的正确性。

Code	Stock Name	Buying Signal	Open Price	Close Price
000532	力合股份	2015-12-07	22.05	24.59
000790	华神集团	2015-12-07	15.29	15.97
000822	山东海化	2015-12-04	7.08	7.3
000852	石化机械	2015-12-07	20.88	21.4
002086	东方海洋	2015-12-03	26.11	28.95
002133	广宇集团	2015-12-03	8.16	9.01
002292	奥飞动漫	2015-12-07	37.08	40.49
002328	新朋股份	2015-12-03	11.65	12.58
002387	黑牛食品	2015-12-03	17.19	18.99

图 5-3 MACD-KDJ 组合量化选股测试

Fig 5-3 MACD-KDJ Combination Stock Selection Testing

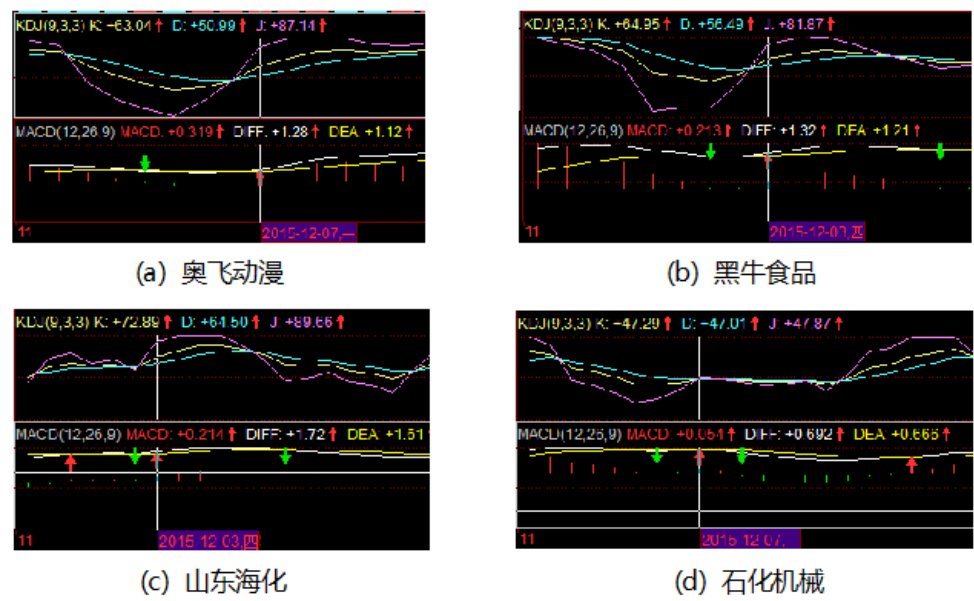


图 5-4 MACD-KDJ 组合量化选股对比测试  
Fig 5-4 MACD-KDJ Combination Stock Selection Testing

5.2.2 案例 II：RSI-WR 组合

案例 II 选取量化技术指标组合：WR+RSI，选股时间范围为 2014 年 6 月 4 日至 2014 年 6 月 11 日，目标股票群体为全体股票池。选取结果如图 5-5 所示：

002065	东华软件	2014-06-10	19.21	20.24
002083	孚日股份	2014-06-10	4.05	4.08
002158	汉钟精机	2014-06-09	15.54	15.88
002214	大立科技	2014-06-05	17.96	18.6
002218	拓日新能	2014-06-10	9.58	9.74
002234	民和股份	2014-06-06	7.49	7.62
002243	通产丽星	2014-06-09	6.41	6.51
002244	滨江集团	2014-06-05	5.96	6.01
002372	伟星新材	2014-06-05	11.13	11.27

图 5-5 RSI-WR 组合量化选股测试  
Fig 5-5 RSI-WR Combination Stock Selection Testing

对此，我们随机抽取四支股票，分别为：东华软件、大立科技、民和股份、

通产丽星。比照同花顺中的技术指标，进行选股的正确性判断，判断结果如图 5-6 所示：

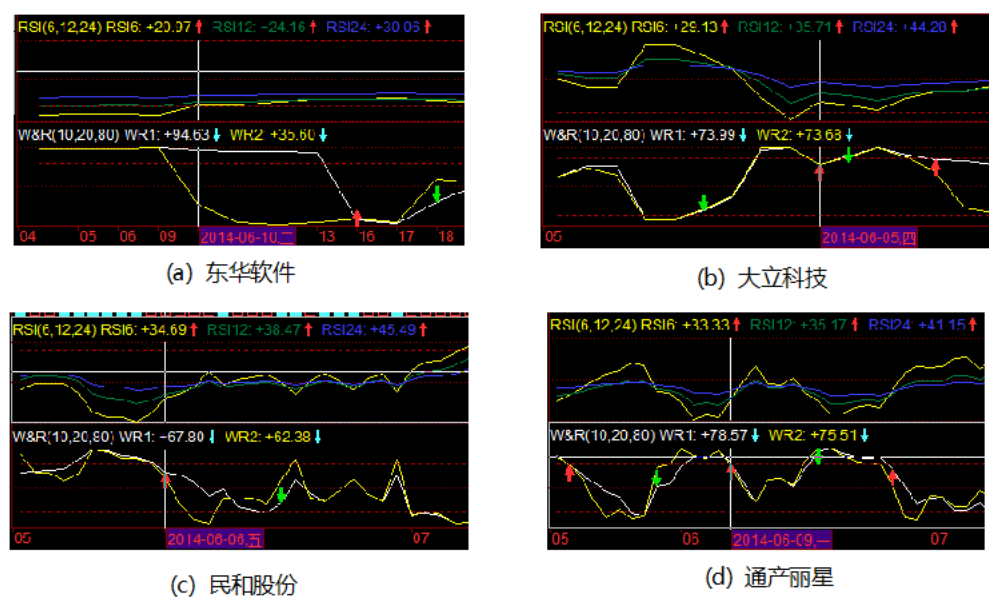


图 5-6 MACD-KDJ 组合量化选股对比测试  
Fig 5-6 MACD-KDJ Combination Stock Selection Testing

由图 5-6 中可以得出，四支股票给出的日期都是 RSI 与 WR 同时符合买入点信号后的第一天，也同时对应了同花顺所指出的买入信号点。

### 5.3 技术指标分析模块测试

本节旨在对量化技术指标分析模块进行测试，并对单个与多个技术指标提出案例分析。测试对象为全体股票池，变量为持仓长度 ( $\alpha$ )、市场情况 ( $\beta$ ) 与技术指标 ( $\gamma$ )，即  $(\alpha, \beta, \gamma)$ 。

为便于分析，将持仓长度划分为短、中、长三个跨度，其取值分别为： $short = [3, 9], middle = [10, 16], long = [17, 23]$ ， $\beta$  为市场情况，分为牛、熊与震荡市场，牛市取值为  $\beta_0 = [2014 - 7 - 23, 2015 - 06 - 12]$ ，熊市取  $\beta_1 = [2007 - 10 - 31, 2008 - 11 - 11]$ ，震荡市场取值为  $\beta_3 = [2013 - 07 - 17, 2014 - 07 - 18]$ ， $\gamma \in (MACD, KDJ, RSI, BOLL, WR)$ 。

测试上，由于分析实验是基于量选指标选股结果统计产生，因此文章仅针

对分析结果进行随机抽样验证其正确性，方法同上节所述，不再赘述，并将重点集中在技术指标选股有效性分析上。

鉴于篇幅限制，文章仅给出震荡市场行情下的分析结果，案例在牛熊市下的表现情况详见附录 B。

5.3.1 单技术指标检验测试

MACD 有效性检验

本节针对震荡行情下技术指标进行检验，即  $\alpha = short, middle, long, \beta = 2$  下进行有效性检验。

实验数据及图表结果如下所示，从表 5-3 中可以看出，在震荡市场行情下，个股涨跌幅分布主要集中在 0-20% 区间，且在此区间的上涨概率低于下跌概率；在 20% 以上变化幅度中，上涨概率显著大于下跌概率，反映了上节中得出结论，即 MACD 指标抗跌性属性比较显著。

表 5-3 震荡下 MACD 技术指标相关参数跌幅数据  
Table 5-3 MACD Examination under Shaking Market

	short		middle		long	
	0+	0-	1+	1-	2+	2-
0-20%	23429	27517	21147	26186	18763	25125
20-40%	956	244	2009	831	2598	1356
40-80%	123	202	386	337	700	444
>80%	20	1	71	7	106	6

从图 5-7 分布图与 5-8 饼状图中，可以得出，MACD 在震荡行情中，指标整体有效性较差，仅有 46% 左右概率上涨。但是，指标有更大概率获得超过 100% 甚至 200% 涨幅，即存在个股在 30 个交易日内，股价翻倍情况。

综合上面三个市场行情下的表现，可以看出，MACD 技术指标受市场行情影响大，在牛市与熊市中，无法判断其指标有效性与实用性，而在震荡行情的检验中，单个 MACD 技术指标并不能给出大于 50% 概念的股价上涨趋势研判，这与 MACD 指标本身属性制约有关，即指标相对滞后，反应迟钝。

另外，在震荡行情中，MACD 金叉点上上涨概率没有随着时间变化发生大幅度变化，因此可以初步得出结果，即 MACD 表现同时适用于长中短期市场。

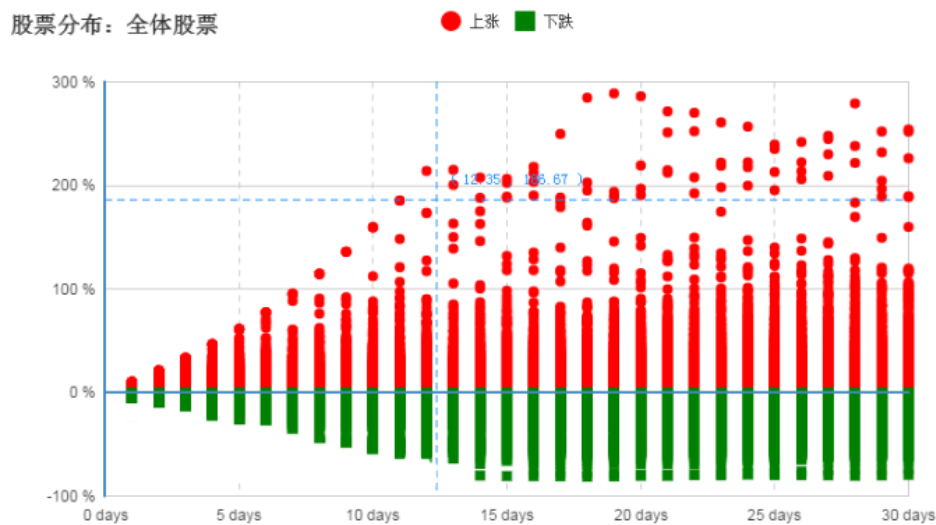


图 5-7 震荡市下 MACD 单指标结果分布  
Fig 5-7 MACD Examination under Shaking Market

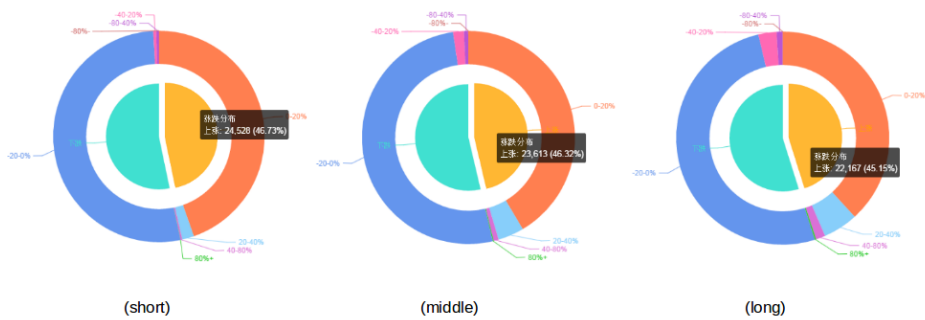


图 5-8 震荡市下 MACD 单指饼状分析图  
Fig 5-8 MACD Examination under Shaking Market Pie Graph

KDJ 有效性检验

本小节针对震荡行情下的 KDJ 技术指标进行了有效性检验，即  $\alpha = short, middle, long$ ， $\beta = 2$ ， $\gamma = KDJ$  的指标有效性检验。

如表 5-4 所示，可以看出，在震荡行情中，KDJ 在长、中、短持仓周期中表现并不是非常出色，甚至在某些情况下出现了跌幅占比大于涨幅的情况，如在短线中，变化浮动范围在 40%-80% 内的频次上涨频次即小于下跌频次。说明 KDJ 在震荡行情中，很难具有稳定研判未来股价趋势走势的属性。

图 5-9 为技术指标 KDJ 在出现金叉后 30 个交易日内的未来股价涨跌分布

表 5-4 KDJ 技术指标相关参数跌幅数据

Table 5-4 KDJ Examination under Shaking Market

	short		middle		long	
	0+	0-	1+	1-	2+	2-
0-20%	170291	163231	162781	152968	147636	145016
20-40%	3840	1122	10016	3261	15162	5572
40-80%	571	759	1695	1620	3132	2371
>80%	47	0	273	8	476	5

情况。从图中可以看出，股价在未来趋势中，上升空间仍然要大于下探空间。

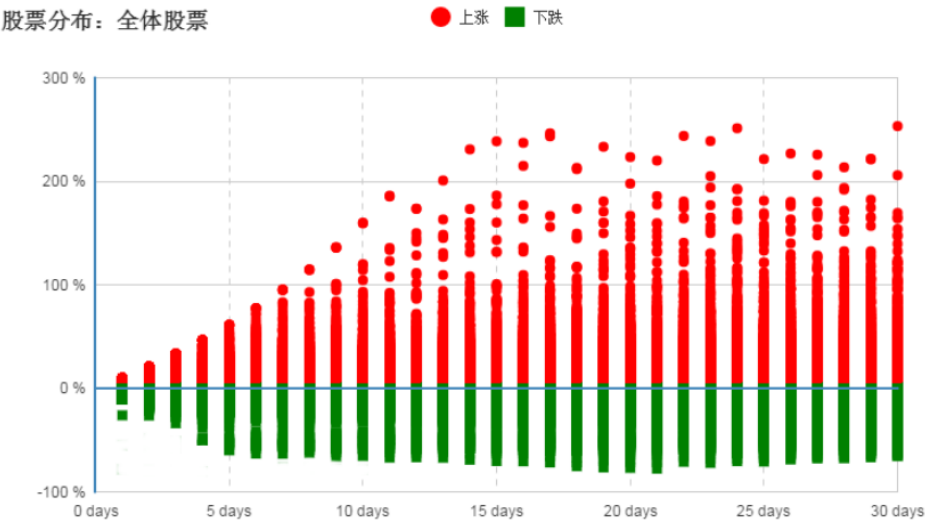


图 5-9 震荡市下 KDJ 单指标结果分布

Fig 5-9 KDJ Examination under Shaking Market

图 5-10 反应了在不同持仓时间内，KDJ 技术指标选股所体现的股价变化情况。图中股价涨跌对比中，长、中、短的上漲比例分别为：52.54%，51.42、%，52.10%、变化幅度不显著，说明 KDJ 在震荡市场中，也不能很好地针对未来股价进行大概率地预测。

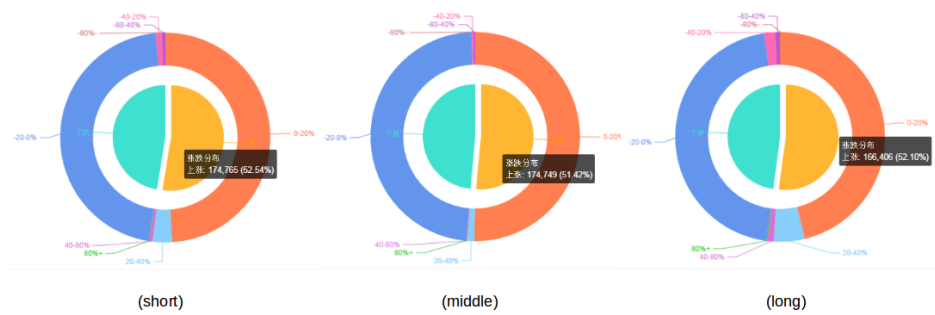


图 5-10 震荡市下 KDJ 单指饼状分析图

Fig 5-10 KDJ Examnation under Shaking Market Pie Graph

5.3.2 技术指标组合分析

上一节中，论文主要针对单个技术指标进行了量化分析。这一节中，将继续在系统平台上进行测试，检验技术指标组合在市场中的历史回溯情况。与上一章一致，本节将针对震荡市场进行统计回测。

使用到了如下参数： $\alpha = short, middle, long$ ， $\beta = 2$ ， $\gamma = KDJ, MACD$ ，即组合指标 MACD 与 KDJ 在震荡行情市场中的有效性检验。

如表 5-5 中所示，在涨跌幅度大于 20% 的分布区间中，组合上涨的频次要显著大于下跌的频次。

表 5-5 KDJ-MACD 技术指标组合相关参数跌幅数据

Table 5-5 KDJ-MACD Examnation under Shaking Market

	short		middle		long	
	0+	0-	1+	1-	2+	2-
0-20%	6021	5951	5296	5913	4515	5818
20-40%	260	40	507	151	611	274
40-80%	22	33	96	63	201	84
>80%	4	0	16	0	20	0

图 5-11 中给出了震荡市下 KDJ、MACD 组合饼状图分析结果。通过与图 5-10、图 5-12 对比，可以发现，符合组合条件的股票频次整体有大显著下降，KDJ-MACD 组合较之于 MACD 饼状图分析结果，技术指标整体上涨频次提升并超过了 50%，KDJ-MACD 组合较之于 KDJ 分析结果，可以发现，组合

在中短期内的有效性有明显提升。

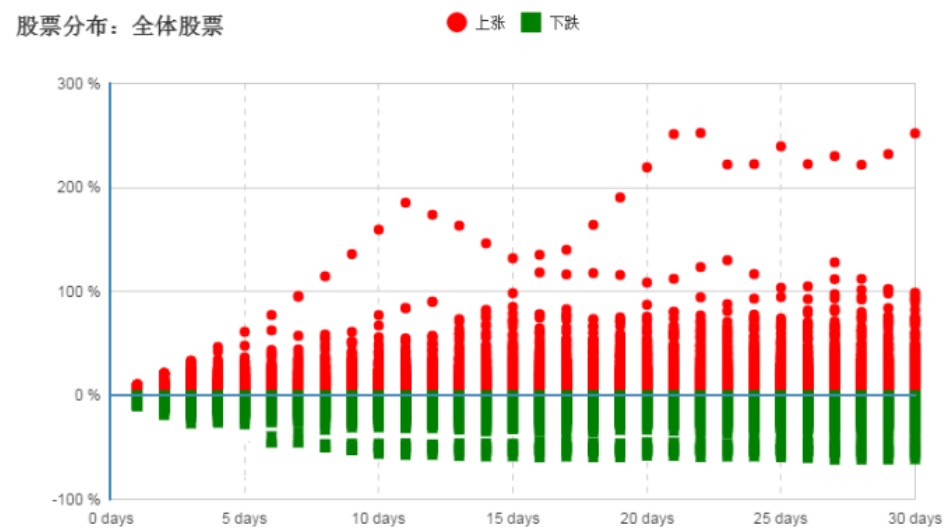


图 5-11 震荡市下 KDJ、MACD 组合结果分布  
Fig 5-11 KDJ-MACD Examnination under Shaking Market

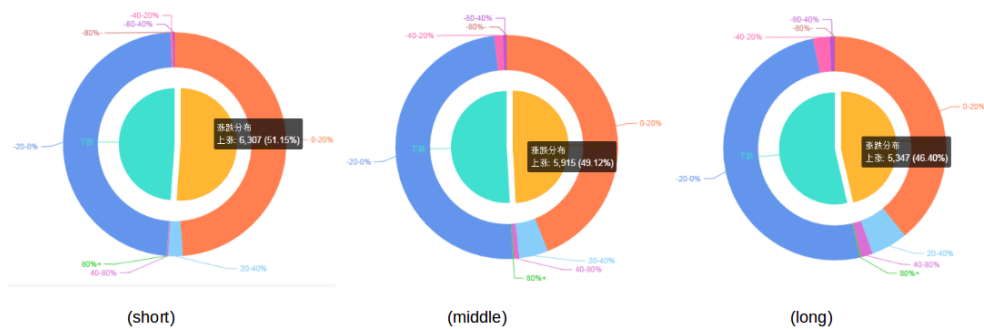


图 5-12 震荡市下 KDJ、MACD 组合饼状分析图  
Fig 5-12 KDJ-MACD Examnination under Shaking Market Pie Graph

总结来看，技术指标的组合对单个指标的缺点与不足都起到了弥补作用，在文中实践的案例中，KDJ-MACD 较之于 KDJ 或者 MACD 都有了很大的提升。

但是，就单纯结果分析来看，本实验中技术指标的组合并没有提供一个大概率上涨或者下跌趋势预判结果，即本次实验中的技术指标组合并没有非常高的有效性。

其中原因，一方面可能与技术指标本身设计有关系；另一方面，指标在量



化实践过程中参数选取的重要性，也可能影响技术指标在量化过程中的选取结果；还有一个原因是，中国主板市场尚未成熟，政府政策相互之间的干预会引起市场不定期扰动，影响技术指标本身的可操作性与有效性。

5.4 不同量化选股平台对比分析

本文量化选股系统平台与第一章中介绍的两个系统平台最大的区别在于平台的本地化问题上。大智慧与同花顺至此仍然依赖操作系统，而本课题实验则是创新性地将选股与分析工作交至后台服务器，将量化工具通过计算机手段实现互联网化，摆脱了系统对不同平台之间的依赖性问题。

此外，本文系统在交互上较之到大智慧与同花顺，也更加友好与方便，操作一目了然，并且结果准备，较之与两个平台的繁冗操作过程以及不完备、不精确的选股结果有了很大提升。表 5-6 中给出了三个量化技术指标选股的优缺点详细信息对比。

表 5-6 三个同类信息平台对比分析  
Table 5-6 Three Different Platform Comparison Analysis

	文中系统	同花顺	大智慧
操作系统	All	Windows, Mac	Windows, Mac
使用方式	在线	本地下载	本地下载
网络依赖	是	是	是
技术指标	5	12	20+
历史展现	是	否	否
个股信息	是	是	是
指标组合信息	是	是	否
指标量化分析	完备	不完备	不完备
分析展示	散点图、饼图、统计表	统计表	最优最差、拆线图

在指标的有效性分析上，二者系统都没有一个给出一个详细的分析报告，即买入点未来股价变化，以及涨跌幅分布，这是本文选股系统所特别提出并实现的功能模块。

然而，与商业化多年的综合性平台相比，本文中系统也有许多不足之处，

如指标系统不全，指标参数个性化定制待实现等优化，在未来的研究中都可以进一步实现。

## 5.5 本章小结

本章节测试了量化交易选股平台系统三个功能模块的正确性，针对单个技术指标与多个技术指标进行了量化分析，并与同类系统平台进行了比较分析。

结果显示，在基于商业平台数据正确与完备性的假设前提下，系统平台通过了测试，数据在清洗与完备性上通过了检验。在不同系统间比较分析时发现，由于商业综合系统发展时间长，其功能与指标群等较之于本文系统更为完备，但二者系统中都存在着一些不足之处，如都未能实现网络化，平台对操作系统依赖仍然十分严重；选股结果上，如同花顺，未能再现历史选股结果信息；大智慧未能实现多指标组合对比功能等。这在本文系统中都创新性地得到了解决。

与二者系统对比另一个不同点在于，本文系统更侧重于选股结果及其结果对指标分析的考察，侧重于学术研究，而另外两个系统平台则更侧重于推荐股票池、数据信息平台等功能，更加侧重于商业用途。

## 第六章 总结与展望

### 6.1 论文总结

本课题结合时代大背景，即数据指数爆发式增长、全球一体化进行不断加速、资本市场不断开放、计算机技术、互联网技术进一步深化普及到生活各个角落。其中，大数据、互联网、大金融等共同因素促进了互联网金融的蓬勃发展，如何更好得促进计算机融入金融领域，更好得为金融业服务，为学术领域、商业领域提供数据实证上的支撑，即引出了本论文的题目。

在阅读大量金融工程知识、文献的基础上，论文方向最终细化至对金融市场中所涉及到的技术指标的量化上。试图通过计算机技术手段，实证、检验证券二级市场股价价格变化与技术指标之间的关联程度与置信度。所构建的系统具有跨度时间长、数据内容全面、可拓展性强，并基于网络平台，摆脱了系统对底层操作系统依赖等特点。

整个论文课题展开可以分为四个阶段：前期金融工程基础、量化交易相关知识积累；项目数据建模；原始数据的爬去与预处理；量选选股系统搭建。涉及到了计算机领域诸如 UML 建模等、SQLite 数据库优化与操作、Python 数据分析与处理、NodeJS 服务器搭建、HTML5 技术引进等。在分析模型上，抛弃了传统回溯检验模式，而是通过对 10 年 2783 支股票，近 400 百万条数据的分析，所使用量化数据规模以 GB 为单位，实验比较具有代表性。另外，与传统的选股系统相比，本文创新性地实现了跨平台系统搭建，将互联网平台建设与创新相结合，使平台适用于各类操作系统上。

但是，论文在实践过程中，也存在一些问题与不足。如，在数据抓取时，没有仔细估计数据规模，导致最初一段时间都为单进程/线程操作，滞后的网络与 CPU 内存速度不相匹配，影响了研究进度；量化指标采用了常用的 5 个指标，没有将成交量、均线金叉、BOLL 线喇叭口等指标考虑进去。金融量化领域仍有相当大的可操作空间，公司机构在这方面的研究领先于学术机构。无论对公司机构、还是对高校学术机构，金融与计算机的结合都存在着巨大潜力。

## 6.2 研究展望

目前，互联网、计算机正不断地渗入金融领域，这不仅仅有百度、阿里和腾讯（BAT, Baidu、Alibaba、Tencent）的角逐，也有来自金融领域内部的变革，如平安银行旗下的陆金所。

处在变革浪潮时期的上海、中国，金融市场正在蓬勃发展着，未来数十年甚至是二十年间，都将不断发生变革与繁荣，而作为支撑的计算机技术，则是金融变革必不可少的工具之一。小额微贷、P2P、移动金融互联网服务、移动支付等，都为计算机从业、研究人员带来了机遇与挑战，正确地把握未来发展趋势，将理论放在实践中，才能更好地体现计算科学与技术专业的价值，为人们生活带来更多便捷。

尽管论文研究了相关股市技术指标的量化工作，但仍未能全面检验资本市场的发展规律，甚至是通过某些参数的改变，便会产生不同的结果，这些工作都需要进一步得到检验和认证，如进一步扩充量化技术指标群，用户技术指标参数自定义等，这在未来的量化技术指标选股上仍然存在很大研究空间。

最后，在系统的实现上，可以通过尝试引入分布式系统，以有效地提高系统平台的运行效率，这在未来的研究中可以进一步展开。

附录 A 个股信息展示正确性测试案例 II

本案例采集的是东方证券 2015 年 8 月 19 日至 2016 年 1 月 22 日间的个股信息，通过对比个股信息与技术指标值可以通过数据的正确性检验。

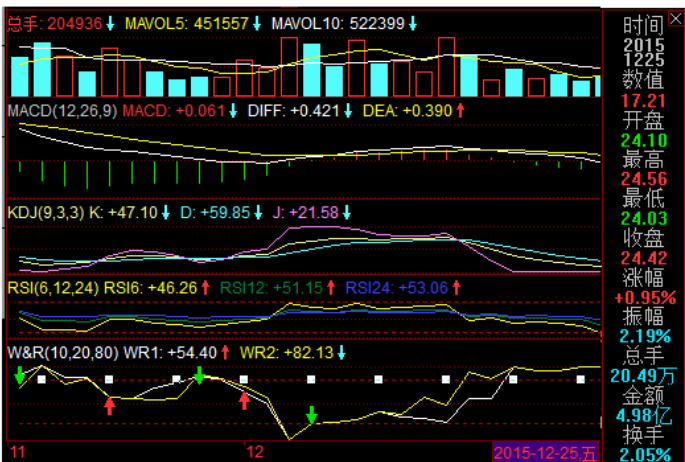


图 A-1 同花顺中成交量与主要技术指标信息

Fig A-1 THS Volume and main Oscillators Information

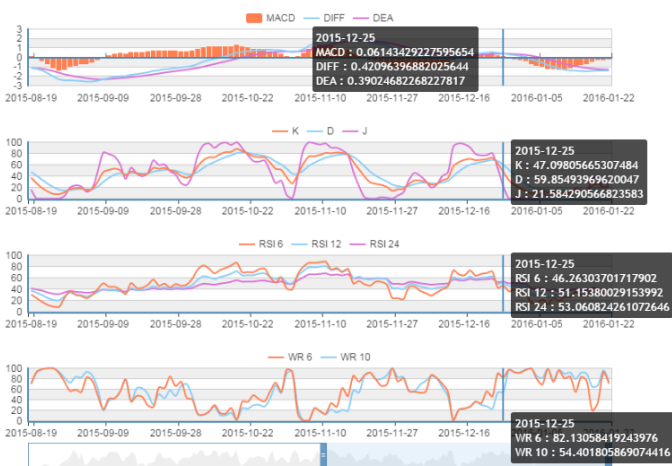


图 A-2 本文系统中成交量与主要技术指标信息

Fig A-2 Volume and main Oscillators Information in the Context

附录 B 量化个股选择系统分析与建模

牛市行情

本节通过选择市场、持仓区间，对技术指标 MACD 有效性进行检测。即在参数： $\alpha = short, middle, long$ ,  $\beta = 0$  下进行有效性检验，结果如下所示：

表 B-1 牛市下 MACD 技术指标相关参数跌幅数据

Table B-1 MACD Oscillator Bull Market Behavior

	short		middle		long	
	0+	0-	1+	1-	2+	2-
0-20%	53182	36536	47822	29143	193433	121678
20-40%	4868	552	10747	1074	44422	4680
40-80%	838	260	3280	414	15604	1521
>80%	72	0	434	0	2815	0

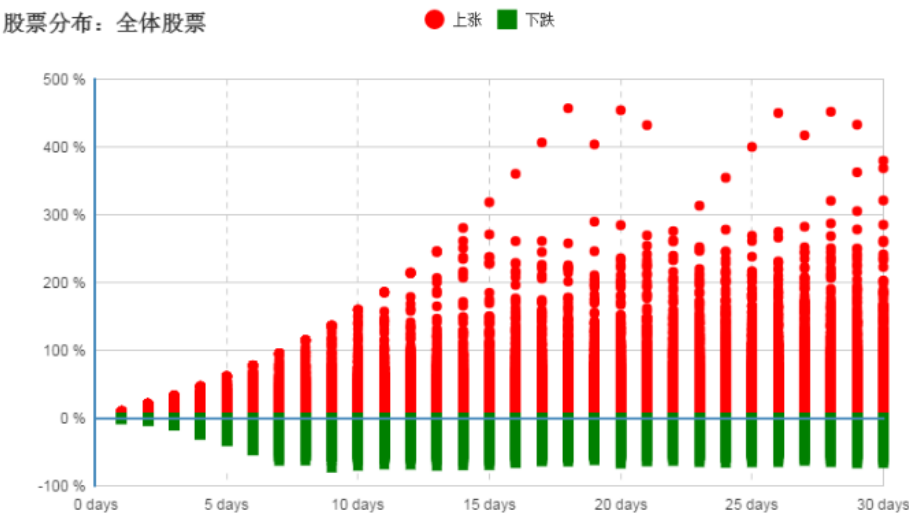


图 B-1 牛市下 MACD 单指标结果分布  
Fig B-1 MACD Examination under Bull Market

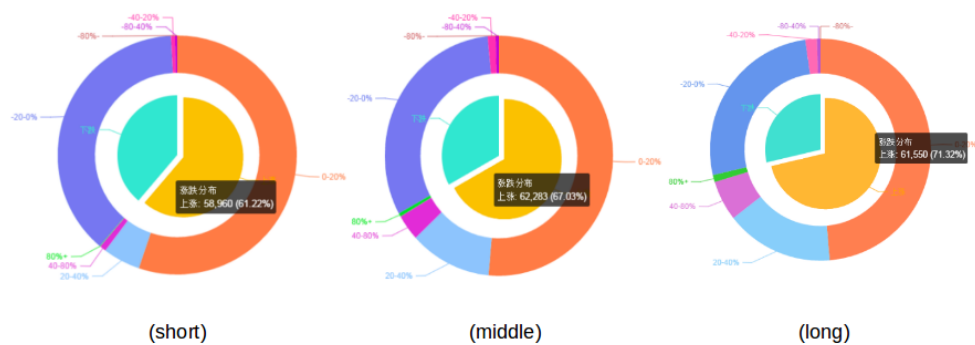


图 B-2 牛市下 MACD 单指饼状分析图

Fig B-2 MACD Examination under Bull Market Pie Graph

其中，表 B 为指标在长、中、短期内上涨与下跌在不同幅度范围内的数据；图 B 为指标在金叉点未来 30 天内的具体涨跌分布情况；图 B 为指标在长、中、短期内涨跌幅分布饼状图。结合表格与数据图，我们可以得出如下结论：

- 金叉出现后，后市有大概率看涨趋势，且时间越长，上涨概率越大；
- 有较大概率出现涨幅超过 80% 的情况，没有出现低于 80% 情况发生；
- MACD 金叉不能保证跌幅小于 40%-80% 跌幅情况发生；

熊市行情

本节检验了熊市行情下，不同持仓区间，技术指标 MACD 的表现情况，结果如下所示：

表 B-2 熊市下 MACD 技术指标相关参数跌幅数据

Table B-2 MACD Examination under Bear Market

	short		middle		long	
	0+	0-	1+	1-	2+	2-
0-20%	2311	4282	1860	3485	1188	3214
20-40%	118	449	292	1356	296	2085
40-80%	5	36	49	115	51	294
>80%	0	0	0	0	0	0

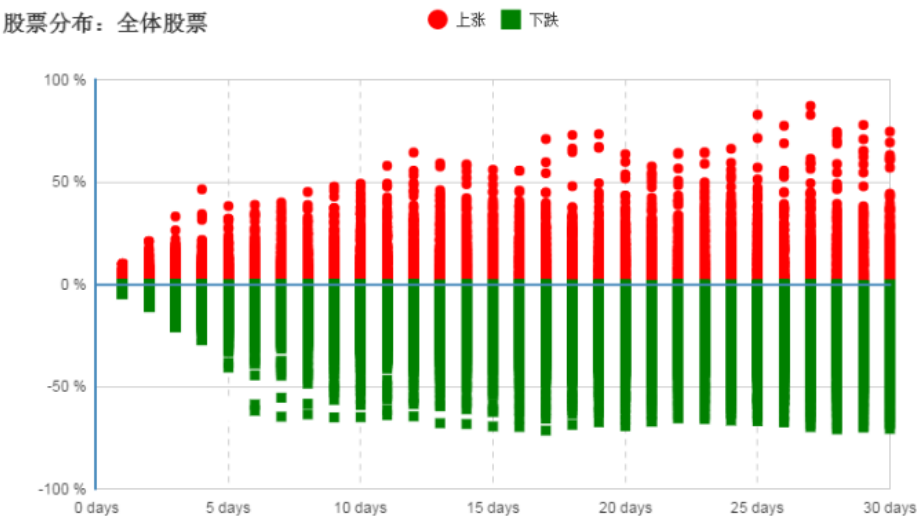


图 B-3 熊市下 MACD 单指标结果分布

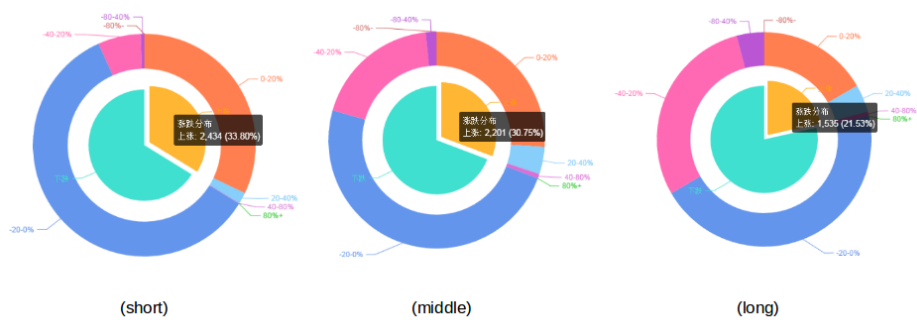


图 B-4 熊市下 MACD 单指饼状分析图

结合表格 B-2 与图 B-3 与图 B-4 可以看出：

- 在熊市中，MACD 获利概率小，且时间越长，有效性越差；
- 与牛市不同，在长、中、短期都表现出相对抗跌属性，即随着时间推移，大幅度下跌概率没有显著增长；



## 参考文献

- [1] ZENIOS S A. High-performance computing in finance: The last 10 years and the next[J]. Parallel Computing, 1999, 25(13):2149–2175.
- [2] PETRAM L O. The world’ s first stock exchange: how the Amsterdam market for Dutch East India Company shares became a modern securities market, 1602-1700[J]. 2011.
- [3] Shanghai Stock Market, <http://www.csai.cn/touzi/573583.html>, visited on: 2015-5-10[M].[S.l.]: [s.n.] , 2015-5-10. <http://www.csai.cn/touzi/573583.html>.
- [4] MODEL A P, DIVERSIFICATION I, ROLL’ s C. Investment Analysis & Portfolio Management[J]. 1987.
- [5] 宋亚洲. 白银统计套利及风险控制 [J]. 经营管理者, 2015, 17:046.
- [6] 郭朋. 国外高频交易的发展现状及启示 [J][J]. 证券市场导报, 2012, 7:56–61.
- [7] DEFUSCO R A, MCLEAVEY D W, PINTO J, et al. Quantitative investment analysis[M].[S.l.]: John Wiley & Sons, 2015.
- [8] 辉金波. 资产证券化在我国保险市场的运用 [J]. 河北企业, 2011.
- [9] 海通证券. 证券行业互联网化研究报告 [R].[S.l.]: 海通证券, 艾瑞咨询, 2014.
- [10] 艾瑞咨询. 2015 中国互联网演艺平台发展研究报告 [R].[S.l.]: 艾瑞咨询, 2015.
- [11] 艾瑞咨询. 2015 中国互联网演艺平台发展研究报告 [J]. 艾瑞咨询, 2015.
- [12] 马家俊. 美国量化交易已有超过三十年历史 [J]. 新浪财经, 2014.
- [13] 陈曦. 2014 年中国量化投资基金总结 [J]. 凯纳, 2014.

- [14] 通联量化平台, <http://www.datayes.com/>, visited on: 2015-9-10[M].[S.l.]: [s.n.], 2015-9-10. <http://www.datayes.com/#/home>.
- [15] ZUCKERMAN G. renaissance' s Man: James Simons does the Math on fund[J]. The Wall Street Journal, 2005, 1.
- [16] MARKOWITZ H. Portfolio selection\*[J]. The journal of finance, 1952, 7(1):77–91.
- [17] FAMA E F, FRENCH K R. The capital asset pricing model: Theory and evidence[J]. Journal of Economic Perspectives, 2004, 18:25–46.
- [18] MIRZA H H. Efficient Market Hypothesis (EMH)[J].
- [19] HUBERMAN G. Arbitrage pricing theory[R].[S.l.]: Staff Report, Federal Reserve Bank of New York, 2005.
- [20] 饶育蕾, 刘达锋. 行为金融学 [M]. Vol. 5.[S.l.]: 上海财经大学出版社, 2003.
- [21] CHONG T T L, NG W K. Technical analysis and the London stock exchange: testing the MACD and RSI rules using the FT30[J]. Applied Economics Letters, 2008, 15(14):1111–1114.
- [22] 廖海燕. 零基础学炒股从入门到精通 [J]. 人民邮电出版社, 2015:212–213.
- [23] FERNÁNDEZ-BLANCO P, BODAS-SAGI D J, SOLTERO F J, et al. Technical market indicators optimization using evolutionary algorithms[C]//Proceedings of the 10th annual conference companion on Genetic and evolutionary computation. .[S.l.]: [s.n.], 2008:1851–1858.
- [24] WANG P, HUANG H P. Discussion on RSI and KDJ expert system of securities software[C]//Computer Sciences and Applications (CSA), 2013 International Conference on. .[S.l.]: [s.n.], 2013:63–66.
- [25] ZHEN X N, PENG Y L. Forecasting the Variation Tendency of Forestry Sustainable Management Status with MACD and KDJ Methods [J][J]. Journal of Southwest Forestry College, 2003, 3:007.

- [26] MURPHY J J. Technical analysis of the financial markets: A comprehensive guide to trading methods and applications[M].[S.l.]: Penguin, 1999.
- [27] WILDER JR J W. The Relative Strength Index “[J]. J. of Technical Analysis of Stocks and Commodities, 1986, 4:343–346.
- [28] WILLIAMS L R. How I Made One Million Dollars Last Year Trading Commodities[M].[S.l.]: Windsor Books, 1979.
- [29] 柯原. 股票投资中技术指标 KDJ 的综合运用研究 [J]. 福建行政学院福建经济管理干部学院学报, 2005:212–213.
- [30] 高老庄. 布林线的使用方法及技巧 [J]. 股市动态分析, 2009(21):44–44.
- [31] KANNAN K S, SEKAR P S, SATHIK M M, et al. Financial stock market forecast using data mining techniques[C]//Proceedings of the International Multiconference of Engineers and computer scientists. .[S.l.]: [s.n.] , 2010, 1.
- [32] 武路. 移动应用实时统计系统设计与实现 [D].[S.l.]: 西安电子科技大学, 2014.

## 致 谢

在整个论文研究期间，感谢我的导师张忠能老师，给了我非常耐心的指导和帮助。张老师从论文选题开始，就为我提供了各种方向意见，也向我传授了论文中技术观点表达的一些方式方法技巧，极大地开拓了我的思路，其严谨的学术态度，广博的知识以及为师之道给我留下了深刻的印象，感谢张老师在我学习期间对我的指导和帮助。

同时感谢学校其他老师、朋友，以及在论文给我提供软硬件支持的工作单位！

非常感谢我们实验室的同学，张可、熊治钢、刁晓天，感谢他们在平时的学习、生活中给我提供支持与帮助，我们一起合作，完成实验室中的相关课题与任务。

感谢论文中所涉及到的学者，不站在巨人的肩膀上，我根本不可以完成课题的研究。

最后，特别感谢阍庭，感谢在我硕士就读期间对我的支持和鼓励，让我可以在学习、研究这条路上一往无前。

由于学术水平，论文中难免有不足之处，恳请各位老师批评、指正！谢谢！

## 攻读学位期间发表的学术论文目录

- [1] Mingyuan WU, Xiaotian DIAO, Technical Analysis of Three Stock Oscillators Testing MACD, RSI and KDJ Rules in SH & SZ Stock Markets[C]. IEEE International Conference on Computer Science and Network Technology, 2015

## 上海交通大学 学位论文原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

学位论文作者签名：吴明远

日期：2016 年 2 月 18 日

## 上海交通大学 学位论文版权使用授权书

本学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权上海交通大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

本学位论文属于

保 密 □，在 \_\_\_\_\_ 年解密后适用本授权书。  
不保密 ☒。

(请在以上方框内打√)

学位论文作者签名: 吴鸣远

日 期: 2016 年 2 月 18 日

指导教师签名: 胡胜

日 期: 2016 年 2 月 18 日