

需提交的资料包括以下三部分，请分别压缩后在“会员中心”统一提交：

1、论文正文（pdf 格式），并压缩成“**论文正文.zip**”，提交时间：**2016 年 5 月 15 日 24:00** 前；

2、论文正文（doc 格式）、源数据（组委会提供的源数据外）、过程数据、程序及模型文件，压缩成“**附件资料.zip**”，提交时间：**2016 年 5 月 16 日 24:00** 前；

3、参赛承诺书打印后签名，并扫描生成 pdf 文件，压缩成“**参赛承诺书.zip**”，提交时间：**2016 年 5 月 15 日 24:00** 前。

所选题目： .....

综合评定成绩： .....

评委评语：

评委签名:

## 通过机器学习优化多因子选股模型

**摘 要:** 简要描述论文摘要

**关键词:** 提取论文关键词

# The thesis title

**Abstract:** A brief description of the abstract

**Key words:** The keyword extraction

## 目 录

1. 挖掘目标.....	错误!未定义书签。
2. 分析方法与过程.....	错误!未定义书签。
2.1. 总体流程 .....	5
2.2. 具体步骤 .....	5
2.3. 结果分析 .....	6
3. 结论 .....	6
4. 参考文献.....	6

## 1 问题重述

### 1.1 问题背景

优化收益和控制风险是投资永恒的主题。1992 年, Fama 通过分析美国市场几十年的数据发现, 美国股市绝大部分可以被市值、估值以及市场收益 3 个因子解释。Fama 的工作开启了通过因子化分析股市获取超额收益的先河, 此后学术界及业界不断地寻找其他能获取超额收益的因子及其组合和风险控制的方式。

在我国, 基于财务因子及长周期的量价因子为主要因子的传统多因子模型在 A 股市场曾经获得过较为稳健的超额收益, 但是由于 A 股市场存在明显的风格切换(比如 2017 年下半年从传统的小市值风格切换到只有极少数大市值股票上涨, 而绝大部分股票下跌的风格), 传统多因子模型的稳定性及有效性受到了较大的考验。

随着硬件的发展, 各种机器学习算法的应用范围不断扩大, 在很多问题上都展现了强大的求解能力, 得到了很多有现实意义的成果(如人脸识别, 自然语言处理, 搜索引擎等等)。在量化交易的问题中, 机器学习算法能够通过对因子的非线性表达, 捕捉到更加精细的市场信号, 获取较为稳健的超额收益。

### 1.2 要解决的问题

### 1.3 我们的工作

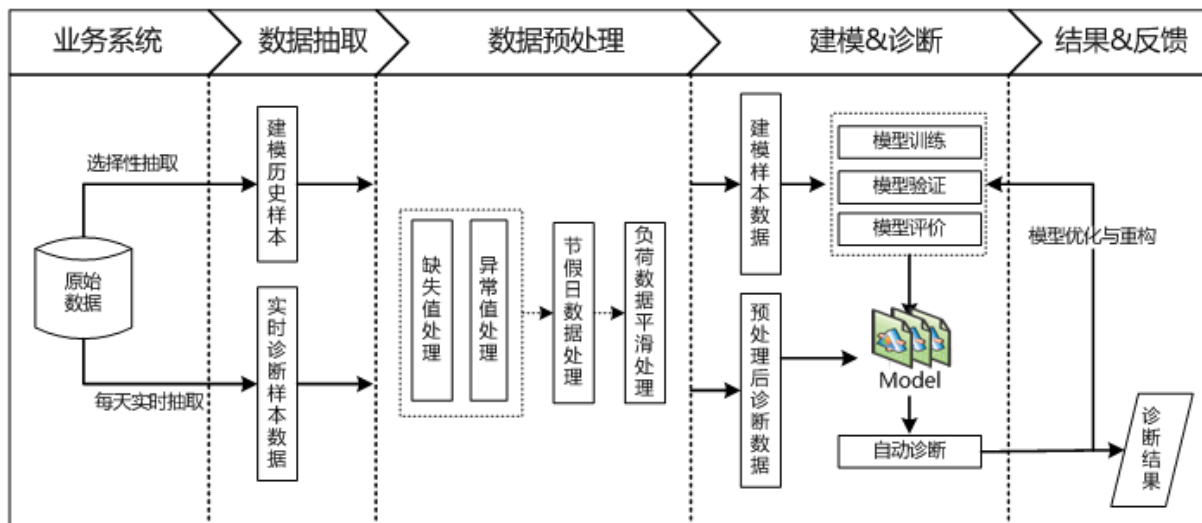
## 2 符号说明

## 3 问题分析

### 3.1 总体流程

用一个总体流程图描述建模方法及过程，并对各部分进行简要说明。

例如：



本用例主要包括如下步骤：

步骤一：...

步骤二：...

步骤三：...

### 3.2 具体步骤

结合总体流程图，对每一步骤做详细说明。

例如：

对本例自适应窃漏电诊断建模法进行详细说明。

**步骤 1：样本数据抽取** 由错误!未找到引用源。知，窃漏电评价影响因子主要有电量类指标、负荷类指标、线损类指标和报警类指标等，故进行窃漏电诊断时需根据建模要求有选择性地从营销系统及计量系统中抽取电量、负荷、报警及线损等数据。

为了尽可能全面覆盖各种窃漏电方式，建模样本包含近三年来所有的窃漏电用户及部分正常用户。

**步骤 2：样本数据预处理** 数据预处理主要包括缺失值处理、异常值处理、节假日数据处理、

负荷数据平滑处理等。

◇ 缺失值处理

在原始计量数据，特别是实时负荷数据抽取过程中，发现存在缺失的现象，为确保建模数据的有效性，将这些值抛弃掉，计算平均值时，不列入其中。

◇ 异常值处理

在原始样本数据中，存在大量异常值的情况，如-10000，以及一些超出指标阈值范围的数据，对于这类异常数据，当成缺失值处理。

◇ 节假日数据修正

节假日的用电量及实时负荷数据同工作日比起来，会明显偏低，为了确保不同日期类型的计量数据在一定时间周期内具有可比性和连贯性，需要对这些特殊类型日期的数据进行修正处理，处理规则是：去除星期六星期的数据。

.....

### 3.3 结果分析

对数据挖掘建模过程中产生的图表结果进行解释分析。

## 4 结论

结合研究目标和实现效果，对本次研究下一个结论性的评语。

## 5 参考文献

列举在本次研究中所参考的文献。

例如：

[1]刘涛, 张良均. 大规模智能用电系统海量数据处理与数据挖掘技术研究及应用[Z]. 博士后重大专项