

# 数据挖掘技术在量化交易指标选取中的应用

罗忠沂, 纪汉霖

上海理工大学管理学院, 上海

收稿日期: 2024年10月13日; 录用日期: 2024年12月2日; 发布日期: 2024年12月9日

## 摘要

股票市场的技术分析法是近年来量化投资研究的新热点, 数据挖掘作为一门交叉学科, 能够大幅度提高经济学中预测类研究的效果。本文系统性地运用数据挖掘技术, 挖掘影响股票涨跌的关键指标, 提升量化投资中的股票收益预测模块。本文设置特定观察期与交易策略, 基于2021年1月至2022年1月A股市场的17项因子, 采用数据挖掘中关联规则FP-Growth算法, 使用Python编程语言对上市公司股票数据进行关联分析。在支持度阈值为30%、置信度阈值为80%条件下, 挖掘出35条关联规则。实证结果显示, K线形态为阳线与短长下影、每股收益高、市盈率低、市净率低、市销率低、营业总收入同比增长率高这六大特征对股票未来收益率的影响较大。同时验证了投资策略的有效性和可行性, 为股票分析以及股票投资中的特征指标的关注提供有意义的参考依据。

## 关键词

关联规则挖掘, 股票收益, FP-Growth算法

# Research on Selecting Quantitative Trading Indicators Based on Data Mining

Zhongyi Luo, Hanlin Ji

Business School, University of Shanghai for Science and Technology, Shanghai

Received: Oct. 13<sup>th</sup>, 2024; accepted: Dec. 2<sup>nd</sup>, 2024; published: Dec. 9<sup>th</sup>, 2024

## Abstract

The technical analysis method of the stock market has become a new hot topic in quantitative investment research in recent years. As an interdisciplinary field, data mining can greatly improve the effectiveness of predictive research in economics. This article systematically applies data mining techniques to explore key indicators that affect stock price fluctuations and improve the stock return prediction module in quantitative investment. This article sets specific observation periods

and trading strategies, based on 17 factors in the A-share market from January 2021 to January 2022. The FP-Growth algorithm of association rules in data mining is used to conduct association analysis on the stock data of listed companies using Python programming language. Under the conditions of a support threshold of 30% and a confidence threshold of 80%, 35 association rules were mined. The empirical results show that the K-line pattern is characterized by a positive line and a short shadow, high earnings per share, low P/E ratio, low P/B ratio, low P/S ratio, and high year-on-year growth rate of total operating revenue, which have a significant impact on the future return of stocks. At the same time, the effectiveness and feasibility of investment strategies were verified, providing meaningful reference for stock analysis and the attention to characteristic indicators in stock investment.

## Keywords

Association Rule Mining, Stock Returns, FP-Growth Algorithm

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

股票交易是金融市场的重要成分, 股票的价格受到企业的经营状况、市场环境等综合因素的复杂影响。如何在股票市场中获得超额收益始终是金融研究的热点问题之一, 量化交易投资方法是当前股票交易分析的基本方法。数据挖掘结合了人工智能、机器学习、模式识别、数据库等多种技术, 致力于从海量数据中挖掘数据中隐含的、潜在的规律和信息, 使用数据挖掘模型预测股票价格信息逐渐与量化分析技术结合, 成为挖掘股票背后有价值的主要方法。本文主要针对数据挖掘在股票分析中的应用, 挖掘与股票收益高关联度较高的指标, 为股票收益提供更加可靠的理论与实证依据。

## 2. 文献综述

大量的实证和理论表明数据挖掘技术在股票收益率预测中起到显著的作用, 许多学者应用数据挖掘技术构建模型, 挖掘股票市场背后潜在的规则, 充分借助 SAS、SPSS、Python、MATLAB 等软件来实现对数据的处理, 取得了良好的效果(靳雨佳, 2018) [1]。Chong (2017)将数据挖掘与股票市场分析和预测结合, 预测韩国股票的走势, 选取 5 分钟日内交易数据, 取得了较好的结果[2]。兰秋军等(2005)采用了深度学习的方法, 随机抽取 45 只股票, 研究其弱有效性, 进而实证了我国沪深 A 股具有可预测性, 通过历史价格数据挖掘, 可以获得短期超额收益[3]。

在数据挖掘课题中, 关联规则占据举足轻重的地位。其核心功能在于通过深入剖析不同特征指标之间的内在联系, 从而精准捕捉特征之间的关联程度。值得注意的是, 关联规则在股票数据分析领域的应用尤为突出, 它能有效地帮助投资者深入挖掘与股票相关的数据, 为投资决策提供更为全面和精准的参考(文娇, 2016) [4]。郑海蓉(2009)基于 ID3 模型的改进算法对股票的财务指标进行分析, 而后通过关联规则分析挖掘出股票之间的规律[5]。在证券领域, Agrawal (1993)收集客户的交易数据, 构建项集间的关联规则, 揭示其中的隐含关系[6]。王赟(2018)应用灰色关联分析法分析因子和股票收益率的相关性, 并验证相关程度稳定性, 以此进行有效因子的选择与构建投资组合, 同时验证了模型的有效性[7]。

在特征变量的选择过程中, 不同学者采用了多样化的构造方式, 但整体上均聚焦于个股市场表现和财务特征这两个关键维度。黄海波(2022)利用关联规则分析股票相关数据时, 选取上市公司营业总收入的

同比增长率、总资产净利率、负债合计、总资产周转率以及经营活动产生的现金流等指标,旨在揭示这些财务指标与股价波动之间的潜在关系[8]。在深入探讨公司财务指标与股价之间的关系时,唐东升(2012)的研究揭示了流动比率和每股收益与股价之间的紧密关联性。这一发现为投资者和企业管理者提供了重要的决策参考[9]。杨艳林(2016)以电力行业上市公司为具体案例,深入剖析了公司财务指标对股价的影响机制。研究发现盈利能力指标和成长能力指标对股价具有显著的正面效应,偿还能力和营运能力则产生相反的效果[10]。闻岳春(2013)等学者研究了创业板高管离职事件对股价的影响,发现高管离职事件发生前后的15个交易日内,股价受到的影响最为显著[11]。

通过对于上述学者的理论和实证分析的梳理可以看出,现有文献多数是对整个股票的走势进行分析,没有涉及股票成交量放大时的预测研究。在特征变量的选择方面,大多文献仅考虑了财务指标变量,并未考虑到个股本身的形态与大盘情况对股票收益的影响。财务指标能够提供公司基本面的深入了解,然而股票价格的变动不仅仅受到公司基本面的影响,还受到市场情绪、技术分析、宏观经济等多种因素的影响。在模型选择方面,多数文章只采用关联分析 Apriori 算法。

不同于前人以整个股票走势为主的研究,本文从股票成交量放大出发,综合考虑个股的形态、大盘走势与财务特征,从三个方面对股票进行分析和预测,更全面地理解股票价格的波动,利用关联规则的 FP-Growth 挖掘与股票收益高关联度较高的指标,为投资者在股价的波动过程中的分析提供了一种新的角度。

### 3. 研究设计

为利用历史市场数据和模拟交易来评估和验证投资策略的有效性和可行性,本文设定的量化交易策略:设置特定观察期,在观察期内股价先出现大幅度下跌,而后小幅度回升,同时伴随在股价达到底部时成交量放量,分析放量当日过后股票收益所能达到的收益情况。

确认回测策略后,首先获取历史数据,进而进行数据预处理,为后续分析做准备。在实证部分运用 FP-Growth 算法对处理后的数据集进行关联性分析,得到满足支持度与置信度的指标,为股票收益提供参考。

#### 3.1. 数据选取与处理

1) 数据选取。本文考察上市公司股价收益相关的关键性指标。在此,围绕上市公司股票交易数据以及财务数据。选取2021年1月至2022年1月中国沪深A股市场所有上市公司的交易数据为研究样本,常见的用于量化选股指标可以从市场、规模、价值等方面考虑,一般有成交量、股票市值、涨跌幅等,特征指标的选择需要考虑相应的维度,同时还应具备有实际的经济意义。本文综合了多篇文献分析的结果,尽可能多地选择特征因子,力求找到最能够影响股票价格变动的特征指标,同时,为了剔除市场整体波动对个股价格变化的影响,因此特别选取大盘指标,具体见表1。财报数据大部分为年度公布,本文采用年度数据进行了季度填充。本文数据来源于东方财富 Choice 数据库。数据库中存在着一定比例的缺失值,使样本数据更具代表性和完整性,其处理过程分为两步:①若某只股票在第t月收益数据存在缺失,则剔除该股票在月份t上的所有数据;②若某只股票的指标存在缺失,则以0填充。

2) 数据处理。在已得到的数据中抽取底部放量的股票,为研究需要,本文对历史数据筛选设置了三个条件:①股价出现较大幅度的下跌,本文设置跌幅30%以上,此时市场对该股票出现了较大的抛售压力,股票可能已经逼近或达到了一种超卖状态,市场上出现供过于求现象。对于时间窗口的选择,为涵盖不同季度和重要时段,更好地体现市场季节性和周期性特点,选择120个交易日作为时间窗口。②当前股价距离前期最低点出现小幅度上涨,本文设置涨幅15%以下,这表明市场整体情绪逐渐乐观,投资者倾向于增加对低价股票的买入需求,从而推动股价上涨。③近期的最大成交量相对于120天的最大成交量1倍以上,表明股票有较强的上涨动能和市场关注度,股票可能还有较大的上涨空间。

因此抽取满足距近期最低价相对于 120 天最高价的跌幅在 30%以上, 近 20 天最低价的涨幅在 15%以下, 相对 120 天的最大成交量的 1 倍以上的股票, 之后观察股价后续上涨情况, 满足上述条件的某一交易日称之为放量当日。

为方便观察趋势, 选择两个月的交易时间跨度, 44 个交易日中最大盈利达到 5%以上赋值为  $y$ , 最大盈利没有达到 5%赋值  $n$ , 最终得到 6210 条股票数据。

考虑到上市公司数据属于连续性变量, 在实现关联规则时, 往往无法直接采用数据, 因此需要对收集的上市公司以及规则数目数据进行离散化处理。去除收集数据中的重复数据、噪声数据以及不完整数据等, 得到价值属性较高的数据, 计算各类平均值, 通过比较平均值与上市公司数据, 将大于平均值和小于平均值的上市公司数据进行文字转化, 为下一步进行准备。

### 3.2. 变量定义

针对与股票收益关键性指标挖掘关联问题, 各指标将作为关联规则挖掘中的关键变量。在原始数据中得出各项指标平均值后, 通过比较平均值与上市公司数据, 将大于平均值和小于平均值的上市公司数据进行字母转化, 按照指标属性分为个股形态、大盘指标与财务指标三类, 进而细分为二级共 17 个指标小类, 特征变量解释如表 1 所示。

**Table 1.** Definition of characteristic variables  
**表 1.** 特征变量定义

指标类型	指标细分	指标小类	指标意义
个股形态	X1-涨跌幅	X11 大跌	当日股价相对于前期最低点的涨幅
		X13 平稳	
		X14 大涨	
	X2-最大成交量	X21 成交量小	当日成交量/观察日最大成交量
		X22 成交量较小	
		X23 成交量较大	
X24 成交量大			
X3-阳线/阴线	X31 阴线	阳线: 当日收盘价高于开盘价	
	X32 阳线	阴线: 当日收盘价低于开盘价	
X4-实体大小	X41 小实体	当日收盘价与开盘价之差	
	X42 中等实体		
	X43 大实体		
X5-上影线比例	X51 短长上影	当日最高价与收盘价/开盘价之差	
	X52 中等长上影		
	X53 长上影		
X6-下影线比例	X61 短长下影	当日开盘价/收盘价与最低价之差	
	X62 中等长下影		
	X63 长下影		
大盘指标	Y1-大盘涨跌	Y11 大跌	当日大盘涨跌
		Y12 中跌	
		Y13 平稳	
		Y14 中等涨	
		Y15 大涨	
Y2-大盘成交量	Y21 大盘成交量小	当日大盘成交量	
	Y22 大盘成交量中等		
	T23 大盘成交量大		

续表

	Z1-市净率	Z11 市净率高 Z12 市净率低	年度每股股价/每股净资产
	Z2-市销率	Z21 市销率高 Z22 市销率低	年度总市值/主营业务收入
	Z3-市盈率	Z31 市盈率高 Z32 市盈率低	年度股票价格/每股收益
	Z4-总资产净利率	Z41 总资产净利率高 Z42 总资产净利率低	年度净利润与平均资产总额
财务指标	Z5-每股收益	Z51 每股收益高 Z52 每股收益低	年度税后利润/股本总数
	Z6-总资产周转率	Z61 总资产周转率高 Z62 总资产周转率低	年度销售收入净额/平均资产总额
	Z7-资产负债率	Z71 资产负债率高 Z72 资产负债率低	年度负债总额/资产总额
	Z8-净资产收益率	Z81 净资产收益率高 Z82 净资产收益率低	年度企业净利润/平均净资产
	Z9-营业总收入同比增长率	Z91 营业总收入同比增长率高 Z92 营业总收入同比增长率低	年度营业收入/上年同期营业收入

### 3.3. 模型构建

FP-Growth 算法作为关联规则挖掘领域经典算法, 其核心思想是将数据集按照项目出现频率的降序排序, 然后利用 FP 树结构来表示数据集, 实现频繁项集的收集只需要扫描两遍数据集, 大大提高了算法的效率, 最后形成简单关联规则。简单关联规则的一般表示形式是:

$$X \rightarrow Y$$

对  $X$  称为规则的前项, 是规则的条件部分;  $Y$  称为规则的后项, 是规则的结论部分。其中支持度 (support)、置信度 (confidence) 是两个重要的概念。支持度表示规则中项集出现的频率, 即项集在事务数据库中出现的次数占总事务的百分比, 其值越高, 表明规则在数据集中出现的频率越高。而置信度确定  $Y$  在包含  $X$  的事务中出现的频繁程度, 即包含  $X$  的事务中同时包含  $Y$  的比例, 其值越高, 表明如果事务包含  $X$ , 则该事务也很可能包含  $Y$  的概率相对较高。FP-Growth 算法包括产生频繁项集和依据频繁项集产生关联规则两个部分。

#### 1) 产生频繁项集

对包含项目  $A$  的项集  $C$ , 将支持度大于等于阈值的候选集作为频繁项集。则  $C(A)$  被称为频繁项集。包含 1 个项目的频繁项集为一项集, 记为  $L_1$ , 包含  $k$  个项目的频繁项集成为频繁  $k$  项集, 记为  $L_k$ 。

#### 2) 依据频繁项集产生简单关联规则

选择置信度大于最小置信度阈值的关联规则, 组成有效规则集合。对于每个频繁项集  $L$ , 计算  $L$  所有非空子集的置信度, 若  $CL' \rightarrow (L-L)$ , 大于最小置信度阈值, 则生成关联规则  $L' \rightarrow (L-L)$ 。

## 4. 实验与结果分析

本文运用 python 实现 FP-Growth 算法, 为了校验模型在近期样本是否有同样的效果, 有必要划分训练集和测试集。因此随机选取 70% 的数据样本放入测试集用于评估模型, 进行股票收益与各指标的简单关联规则挖掘, 剩余 30% 的数据样本划分到训练集用于验证模型, 观察规则的表现。

### 4.1. 简单关联规则

将数据带入 FP-Growth 算法得到完整的频繁项集, 设置后项为股票收益率达到 5%, 筛选后的关联规则共 1241 条。由于本文研究的是有关股票收益率的关联规则, 所以只分析后项为收益率的关联规则, 在最小支持度 30%和最小置信度 80%条件下, 过滤掉重复和增益系数低的关联规则, 共获得有效关联规则 35 条。将挖掘出的关联规则按照置信度大小排序, 进行关联规则的解读。

1) 前项数为 1 的关联规则。结果按照置信度进行排序, 关联规则见表 2。

一项集出现的次数为 2849 次, 规则的支持度为 0.66, 说明在所有事务中放量当日营业总收入同比增长率高与股票收益率达到 5%同时出现的概率达到 66.25%。通过条件概率的计算, 在上市公司出现放量时, 上市公司财务指标中营业总收入同比增长率较高的条件下, 实现未来收益率 5%这一结果的概率达到 81.10%。一方面, 该规则从数据上更加有力地证明了营业总收入同比增长率高的公司在近期取得了良好的经营业绩增长, 公司有着强劲的盈利能力、市场需求或经营管理能力, 经营状况良好, 通常会支付更高的股息, 因此股票的收益率也可能较高。另一方面, 在上市公司股票放量当日可以将营业总收入同比增长率作为关注指标进而判断股票短期的收益情况。

**Table 2.** Association rules with a first item count of 1

**表 2.** 前项数为 1 的关联规则

序号	前项	实例	置信数	支持度
1	Z91	2849	0.8109	0.6625

2) 前项数为 2 的关联规则。结果按照置信度进行排序, 关联规则见表 3。

由于股票收益通常由多个因素相互作用, 共同决定, 上市公司股票放量后是否达到较高收益率更取决于各个影响因素对于股价作用力大小, 根据上表, 关联规则的分析如下。① 相较于上文的规则 1, 前项数为 2 的规则 1、规则 2、规则 3、规则 4、规则 6 与规则 8 的置信度更高, 均高于营业总收入同比增长率高的一项集关联规则, 说明放量当日当营业总收入同比增长率高分别与市盈率低、市净率低、市销率低、每股收益高、阳线、短长下影的组合时收益率高可能性更大。② 前项数为 2 的关联规则, 特征指标的呈现集中的分布状态, 主要集中在 K 线形态为阳线与短长下影、每股收益高、市盈率低、市净率低、市销率低、营业总收入同比增长率高这六大特征, 说明第一可以多关注这六项指标, 当上市公司放量后财务指标出现营业总收入同比增长率高与市净率低时, 进一步进行收益率分析。

**Table 3.** Association rules with a first item count of 2

**表 3.** 前项数为 2 的关联规则

序号	前项	实例	置信度	支持度	序号	前项	实例	置信度	支持度
1	Z91, Z51	2219	0.8267	0.5160	8	Z91, Z32	2370	0.8127	0.5511
2	Z91, X32	2068	0.8239	0.4809	9	Z51, Z12	2480	0.8107	0.5767
3	Z91, X6	2258	0.8213	0.5251	10	Z32, X42	1823	0.8048	0.4239
4	Z32, Z91	2343	0.8183	0.5448	11	Z91, Z62	1944	0.8006	0.4520
5	Z51, X21	2289	0.8175	0.5323	12	X32, Z12	2589	0.8005	0.6020
6	Z91, Z22	2421	0.8154	0.5630	13	Z32, Z12	3048	0.8004	0.7088
7	Z51, X61	2310	0.8142	0.5372					

3) 前项数大于 2 的关联规则。结果按照置信度进行排序, 关联规则见表 4。

根据上表, 关联规则的分析如下。① 相较于前项数为 1 的规则, 前项数为 3 的结果中, 含有营业总收入增长率高的结果如规则 1、规则 2、规则 3、规则 4、规则 5、规则 6、规则 7、规则 9, 这些结论的置信度均高于前项数为 1 结论的置信度, 说明当放量当日营业总收入同比增长率高与上文提到的六大特征组合出现时股票放量后短期的收益率高的概率更大, 这与上文前项数为 2 的统计分析中的结论基本保持一致。② 前项数为 3 的关联规则, 特征指标的也同样呈现集中的分布状态, 主要集中在 K 线形态为阳线与短长下影、每股收益高、市盈率低、市净率低、市销率低、营业总收入同比增长率高这六大特征, 说明第一可以多关注这六项指标, 当上市公司放量后财务指标出现营业总收入同比增长率高与市净率低时, 进一步进行收益率分析。这也与上文前项数为 2 的统计分析中的结论基本保持一致。

**Table 4.** Association rules with 2 or more preceding terms

**表 4.** 前项数为 2 以上的关联规则

序号	前项	实例	置信度	支持度	序号	前项	实例	置信度	支持度
1	Z91, Z51	2219	0.8267	0.5160	8	Z91, Z32	2370	0.8127	0.5511
2	Z91, X32	2068	0.8239	0.4809	9	Z51, Z12	2480	0.8107	0.5767
3	Z91, X6	2258	0.8213	0.5251	10	Z32, X42	1823	0.8048	0.4239
4	Z32, Z91	2343	0.8183	0.5448	11	Z91, Z62	1944	0.8006	0.4520
5	Z51, X21	2289	0.8175	0.5323	12	X32, Z12	2589	0.8005	0.6020
6	Z91, Z22	2421	0.8154	0.5630	13	Z32, Z12	3048	0.8004	0.7088
7	Z51, X61	2310	0.8142	0.5372					

## 4.2. 验证过程

上述结果可以发现, 关联规则模型对股票收益有一定的预测能力。为了进一步验证本文模型挖掘的量化投资能力, 本文分别将 30% 的数据作为验证集, 观察上述规则在验证集中的表现。

具体而言, 在验证集的数据集中抓取上文的规则, 得到验证集中前项的实例数与规则的实例数之比, 计算出规则出现的概率, 具体结果见表 5、表 6、表 7。

验证结果显示, 验证集中规则出现的概率 80% 左右, 模型表现良好, 因此验证结果证实了关联规则在股票的指标挖掘中的出色挖掘能力。

**Table 5.** Verification results with a first item count of 1

**表 5.** 前项数为 1 的验证结果

规则序号	规则实例	前项实例	概率
1	658	842	0.7814

**Table 6.** Verification results with a first item count of 2

**表 6.** 前项数为 2 的验证结果

规则序号	规则实例	前项实例	概率	规则序号	规则实例	前项实例	概率
1	516	650	0.7938	8	483	601	0.8036
2	348	430	0.8093	9	715	914	0.7822
3	483	601	0.8036	10	435	560	0.7767

续表

4	562	711	0.7904	11	440	572	0.7692
5	662	835	0.7928	12	703	901	0.7802
6	562	714	0.7871	13	877	1135	0.7726
7	638	814	0.7837				

**Table 7.** Verification results with a leading number of 2 or more**表 7.** 前项数为 2 以上的验证结果

规则序号	规则实例	前项实例	概率	规则序号	规则实例	前项实例	概率
1	460	573	0.8027	12	519	655	0.7923
2	430	572	0.7517	13	479	592	0.8091
3	320	396	0.8080	14	608	761	0.7989
4	431	533	0.8080	15	220	284	0.7746
5	463	575	0.8052	16	193	246	0.7845
6	333	399	0.8345	17	620	764	0.8115
7	333	395	0.8430	18	590	755	0.7814
8	520	646	0.8049	19	476	581	0.8192
9	469	585	0.8017	20	280	346	0.8092
10	508	625	0.8128	21	610	767	0.7953
11	494	627	0.7878				

## 5. 结论与建议

股票收益是投资者们关注的重要问题之一, 如何能够获得可观的超额收益受到市场参与者的青睐。本文选取 2021 年 1 月至 2022 年 1 月中国沪深 A 股市场所有上市公司的交易数据, 设置特定观察期, 在观察期内股价先出现大幅度下跌, 而后小幅度回升, 同时伴随在股价达到底部时成交量放量, 从该策略设定出发, 从个股的形态、大盘走势与财务特征三方面的指标变量入手, 共选取 17 个特征指标, 运用 python 实现挖掘过程, 利用 FP-Growth 算法挖掘股票收益率高与各个指标的有效关联规则, 表明投资策略的有效性和可行性。通过对实证部分的结果进行整理分析, 可以得到如下结论:

1) 过滤掉重复和增益系数低的关联规则, 提炼有效关联规则 35 条, 包括 1 条前项数为 1 的关联规则、13 条前项数为 2 的关联规则、21 条前项数大于 2 的关联规则。

2) 从关联性分析结果看, 得到的结果中特征指标主要集中在 K 线形态为阳线与短长下影、每股收益高、市盈率低、市净率低、市销率低、营业总收入同比增长率高这六大特征, 易获得高收益的关联特征多数为 6 个指标的组合。

关于结合本文研究结论和我国股票市场的发展状况, 本文提出以下建议: 首先, 从大数据的视角能够发现大量股票数据中隐含的信息价值, 数据挖掘驱动的基本面量化投资策略在挖掘指标能力上表现良好, 可以作为股票投资者决策方法和理论依据。其次, 在做出量化投资策略时, 可以关注以上特征组合, 为获得超额收益作为参考。同时股票放量后获得高收益还需要综合考虑多方面因素, 如对公司的盈利能力、行业前景、市场竞争等方面进行全面分析和评估。综合考虑其他因素的基础上, 根据本文基于现有文献发现股票放量后获得预期收益的指标, 进而为量化投资提供了新的思路与理论依据。



## 参考文献

- [1] 靳雨佳. 基于数据挖掘技术的股票市场分析 with 预测[D]: [硕士学位论文]. 长春: 吉林财经大学, 2018.
- [2] Chong, E., Han, C. and Park, F.C. (2017) Deep Learning Networks for Stock Market Analysis and Prediction: Methodology, Data Representations, and Case Studies. *Expert Systems with Applications*, **83**, 187-205. <https://doi.org/10.1016/j.eswa.2017.04.030>
- [3] 兰秋军, 马超群, 甘国君, 等. 中国股市弱有效吗?——来自数据挖掘的实证研究[J]. *中国管理科学*, 2005, 13(4): 17-23.
- [4] 文娇. 关联规则挖掘算法在证券交易中的应用[D]: [硕士学位论文]. 长沙: 湖南大学, 2016.
- [5] 郑海荣. 数据挖掘技术在股票分析与预测中的应用[D]: [硕士学位论文]. 哈尔滨: 哈尔滨工业大学, 2009.
- [6] Agrawal, R., Imieliński, T. and Swami, A. (1993) Mining Association Rules between Sets of Items in Large Databases. *ACM SIGMOD Record*, **22**, 207-216. <https://doi.org/10.1145/170036.170072>
- [7] 王赟. 基于灰色关联分析的多因子选股模型研究[D]: [硕士学位论文]. 北京: 北京交通大学, 2018.
- [8] 黄海波. 数据挖掘技术在股票数据分析中的应用[J]. *商业文化*, 2022(13): 92-94.
- [9] 唐东升. 会计信息对股价影响的实证分析[J]. *统计与决策*, 2012(6): 172-173.
- [10] 杨艳林. 上市公司财务指标对股价影响的实证研究——以电力行业为例[J]. *中外企业家*, 2016(6): 19-20.
- [11] 闻岳春, 段弈冰. 创业板高管离职事件对股价影响的实证分析[J]. *证券市场导报*, 2013(7): 31-35.