

基于遗传规划的多源信号交易策略组合优化

文丹艳¹, 马超群², 曹建平³

(1. 南京理工大学经济管理学院, 江苏 南京 210094; 2. 湖南大学工商管理学院, 湖南 长沙 410082;
3. 国防科技大学系统工程学院, 湖南 长沙 410073)

摘要: 将财经新闻事件转化为交易信号融入到策略组合中, 提出基于遗传规划的多源信号股票交易策略组合优化模型. 模型综合财经事件信息以及移动平均、动量等传统技术交易策略, 利用遗传规划对多源信号交易策略组合进行整体优化. 采用中国股市数据实验发现, 典型财经新闻事件可以有效转化为交易信号并与技术交易策略融合; 融合新闻事件信号与传统技术策略可以有效提升股票自动交易的收益率; 基于遗传规划的组合优化模型可以生成有效的交易策略, 其中新闻与指数移动平均组合的策略可以产生较高的收益率并具有较强的稳健性.

关键词: 交易策略; 新闻事件; 技术指标; 股票收益率; 遗传规划

中图分类号: F832.5

文献标识码: A

文章编号: 1000-5781(2020)05-0642-15

doi: 10.13383/j.cnki.jse.2020.05.007

Genetic programming based trading strategy optimization with multi-sourced signals

Wen Danyan¹, Ma Chaoqun², Cao Jianping³

(1. School of Economics & Management, Nanjing University of Science & Technology, Nanjing 210094, China;
2. Business School, Hunan University, Changsha 410082, China;
3. System Engineering College, National University of Defense Technology, Changsha 410073, China)

Abstract: This paper integrates the financial news events as signals into trading strategies and proposes a blended trading strategy model based on genetic programming. Such a model is the exploitation of news events in traditional strategies like moving averages and momentum. Experiments on the data from news web-sites and China stock market show, that the typical financial news events can be effectively transferred into signals for trading strategies, that the integrated trading strategy will prominently improve on traditional trading strategies, while the proposed model based on genetic programming can produce effective trading strategies for application, and that the news with exponential moving average strategy will reap higher returns and are more robust than other combination strategies.

Key words: trading strategy; news event; technical indicator; stock return; genetic programming

1 引言

随着自动交易的广泛盛行, 处理多源的金融市场信息在量化投资、风险管理及资产组合等领域的关键

收稿日期: 2019-03-24; 修订日期: 2019-11-28.

基金项目: 国家自然科学基金资助项目(71901122; 71521061; 71501095); 中央高校基本科研业务费专项基金资助项目(3091-9013204).

性地位日益凸显。一方面, 利用技术交易策略, 尤其是移动平均、动量等策略预测市场趋势可以获得超额收益, 其表现甚至可以和宏观经济变量相媲美^[1]。技术交易策略成为应用于量化投资组合最主要的信息变量^[2]。另一方面, 随着大数据、人工智能等信息技术的兴起与运用, 新闻事件信息、网民群体信息也逐渐被学者们所采用, 成为自动交易系统决策的重要辅助信息。

新闻信息是影响股价的重要因素。金融学界关于新闻消息对股市影响的分析由来已久。早在1966年, Merrill 就率先探索了利空新闻对股票价格的影响^[3]。随着互联网新闻的兴起, 近年来金融学界关于财经新闻与股市关系的研究大多从新闻或消息的外在特征以及新闻事件类型两个角度展开^[4-7]。新闻消息的外在特征包括新闻的数量、公告发布量、新闻关注度、新闻事件中网民群体等一些可以量化的指标^[8]。Ewalds 等^[9]研究了盈余公告与交易量之间的关系, 发现在此类消息发布日附近交易量会上升。Yuan^[10]研究了关注度和股票交易之间的关系, 认为如果市场广泛关注将导致投资者在高位时急剧的抛出股票, 这种激进的抛售将会对价格产生负面作用, 降低市场收益率。刘锋等^[11]针对中国金融市场也做了类似研究, 分析了投资者关注与“媒体关注”的相互关系。卫强等^[12]利用股票搜索量作为投资者关注的重要代理变量, 预测了股价走势进而设计了交易策略。王建新等^[13]从行为金融角度出发, 研究了股票收益率和媒体关注的关系, 发现“媒体效应”主要来自于媒体关注度的未预期部分。

新闻事件对股市影响的研究也非常多, 将新闻事件分为不同类别并分析其对股市的影响是新闻与股市关系研究的另一个主流。Warner 等^[14]研究了管理层变动对股市的影响。Rosen 等^[15]观察了收购消息发布之后的价格动量, 发现价格动量和起初的市场反应以及收购消息的传播有关。此外, 相关研究常以一些突发事件为例^[16,17]。Burch 等^[18]以“911”事件为例, 研究了股票价格对突发事件如何做出反应; 赵静梅等^[19]采用事件研究法系统地分析了天灾(四川“5.12”地震、青海“4.14”地震)和人祸(西藏“3.14”事件、新疆“7.5”事件)对股价的冲击效应。值得注意的是, 随着 Web 2.0 技术的深度发展、智能便携设备的普及以及社交网络的兴起, 许多学者从新闻角度观察市场情绪, 研究新闻、情绪与金融市场之间的关系^[15,20-23]。

以往研究主要聚焦于新闻对股价或者股票收益率的影响, 很少落脚到交易策略层面, 而且单一的事件分析缺乏普适性, 难以适用于算法交易系统。上市公司哪些类型的新闻可以作为算法交易的依据? 如何利用新闻信息帮助实现股票的自动交易? 新闻公开后的哪一段时间内收益率会比较高? 种种关于股市交易的问题既是金融理论研究的关键, 也对日渐兴起的算法交易系统具有重要作用, 是学术界和实业界共同面临的重要课题。分析新闻事件对股价的影响, 构建考虑新闻事件的交易策略优化组合有两大难点。首先, 传统的技术交易策略与新闻事件属于不同层面, 如何将复杂的新闻消息嵌入到具体的交易策略中是本文面临的第一个挑战。第二, 如何将财经事件信号与传统的交易策略相互融合, 通过计算机智能地选择合适的策略组合完成自动交易。针对上述难点, 本文参考相关遗传网络规划等交易策略优化求解方法^[24,25], 将新闻事件转化为“买入”或“卖出”交易信号融入到交易策略中, 并基于遗传规划(genetic programming, GP) 算法设计了一种多源信号的交易策略组合优化模型, 以更精准地选择股票交易的时间, 获取更高的收益率。

本文首先介绍交易策略组合的信号来源, 重点提出将多类财经新闻事件转化为交易信号的方法。然后基于 GP 算法提出了自动交易策略组合优化模型, 建立了算法适应度函数并分析了适应度函数的上、下限, 在此基础上提出评估策略优化组合的方法, 给出了模型优化算法的具体过程。最后, 采用中国股市数据从模型的有效性、效率以及稳健性角度验证了组合优化模型。实验发现, 财经新闻事件可以有效地转化为交易信号并与技术策略信号融合; 经过融合后的交易策略可以有效提升股票自动交易的收益率; 本文提出的组合优化模型可以生成有效的交易策略, 其中新闻信号与指数移动平均策略组合可以产生更高的收益率并具有较好的稳健性。

2 多源交易策略信号

2.1 技术交易信号

参考相关文献^[1,26], 本文采用四种常见的技术交易策略: 简单移动平均策略(simple moving average,

SMA)、基于交易量的移动平均策略(volume-based moving average, VMA)、指数移动平均策略(exponential moving average, EMA)和动量交易策略(momentum, MOM). 这些交易策略通过比较过去短期和长期趋势判断最佳的买入和卖出时机, 当短期趋势大于长期趋势时, 则持有(表示为1); 反之, 则不持有(表示为0). 具体计算公式和买卖信号产生规则见表1.

表1 技术指标信号介绍

Table 1 Introduction of technical indicators

指标	公式	买卖信号
SMA	$SMA_{it,m} = \frac{P_{i,t-m+1} + P_{i,t-m+2} + \dots + P_{i,t}}{m}, m = s, l$	$SMA_{it,s} \geq SMA_{it,l} \quad SMA_{i,t} = 1$ $SMA_{it,s} < SMA_{it,l} \quad SMA_{i,t} = 0$
VMA	$VMA_{it,m} = SMA_{it,m}^{OBV}; OBV = \sum_{k=1}^t Vol_k D_k, m = s, l$	$VMA_{it,s} \geq VMA_{it,l} \quad VMA_{i,t} = 1$ $VMA_{it,s} < VMA_{it,l} \quad VMA_{i,t} = 0$
EMA	$EMA_{it,m} = \frac{P_{i,t}}{m} + (1 - 1/m)EMA_{i-1,m}, m = s, l$	$EMA_{it,s} \geq EMA_{it,l} \quad EMA_{i,t} = 1$ $EMA_{it,s} < EMA_{it,l} \quad EMA_{i,t} = 0$
MOM	$MOM_{it,s} = P_{i,t}; MOM_{it,l} = P_{t-l}$	$MOM_{it,s} \geq MOM_{it,l} \quad MOM_{i,t} = 1$ $MOM_{it,s} < MOM_{it,l} \quad MOM_{i,t} = 0$

注: $P_{i,t}$ 表示股票 i 在 t 时刻的收盘价, s, l 分别表示过去短期和长期滞后项. D_k 为示性函数, 当 $P_k - P_{k-1} \geq 0$ 时,

取值1, 反之取-1. SMA、EMA、VMA 信号取 $s = 5, 10, 20; l = 60, 200$; MOM 信号取 s 为当期股价, $l = 60, 200$.

2.2 新闻事件信号

本文选取的典型财经新闻事件依据以下两个原则: 1) 必须和股价波动紧密相关. 影响股票价格的因素很多, 与股价波动的关系也很复杂, 应选择对股价有直接影响的事件. 2) 必须有一定的出现频率. 如果出现次数少于一定频次, 难以总结其对股价的影响规律, 不宜作为自动交易程序的输入信号.

通过对文献[9, 27–29]的分析以及大规模财经领域语料的统计分析, 本文总结出以下与股市波动相关的14类典型的频繁事件: 1) 股价上涨; 2) 股价下跌; 3) 公司业绩上扬; 4) 公司业绩下滑; 5) 公司声誉提升; 6) 公司声誉受损; 7) 公司利润上升; 8) 公司利润下滑; 9) 公司负债良好; 10) 公司负债堪忧; 11) 公司高层变动; 12) 公司合作; 13) 公司业务扩张; 14) 公司并购. 为便于计算机处理, 本文将同类财经事件视为没有差别的独立个体. 另外, 本文仅考虑交易日的新闻事件对股价的影响, 对于新闻不区分具体时间段.

为了将财经新闻事件转化为系统信号, 本文将上述各类事件与股价分别进行统计关系检验, 通过检验后再综合为系统信号输入到模型中, 此处需要说明的是, 考虑到财经新闻事件对股价的影响是复杂而微妙的, 涉及媒体影响力、事件发生的先后顺序等具体问题无法逐一严格分析, 综合考虑模型复杂度和模型应用场景, 本文仅使用统计方法分析代表变量之间的相关关系. 本文的基本思路为首先对上述14种事件预定义不同的影响力 I_p (predefined impact), 研究预定义影响力和平均收益率 R 之间的关系, 进而分析新闻事件与股票价格涨跌之间的统计关系, 评估新闻事件对股价的影响.

定义某一支股票的平均收益率为 $r_i = (P_{i,n} - P_i)/P_i$, 其中 P_i 代表事件发生前一天的收盘价, $P_{i,n}$ 表示股票 i 经过 $n(n > 0)$ 天后的收盘价. 如果同一支股票的 N 个同类事件在某一段时间内发生, 取各事件收益率的平均值为股票 i 的收益率,

$$R_i = \frac{1}{N} \sum_{k=1}^N r_{i,k}, \quad (1)$$

其中 $r_{i,k}$ 表示事件 k 影响股票 i 所产生的收益率. 针对每类事件分析其预定义影响力 I_p 的正负与股价波动方向(上涨或下跌)之间的关系, 统计方向预测“正确”的比例并计算相关性的显著程度. 当某支股票在事件发生后股价上扬(下跌), 若该事件的预定义影响 I_p 为正(负)就算一次“正确”预测, 否则为“错误”预测.

3 基于遗传规划的自动交易模型

自动交易模型大致可以分为基于机器学习和基于启发式的两种算法. 由于机器学习的算法多为“黑

箱”模型,难以解读其输出结果的经济意义,本文拟采用启发式算法.遗传规划(genetic programming, GP)和遗传算法(genetic algorithm, GA)一样,都属于启发式算法^[30,31].这两个在名称和思想上都类似的算法有两个主要区别,一是GA算法使用等长的信号作为输入,而GP算法则可以使用不等长的信号作为输入,在信号融合方面更加多样灵活.二是GP算法通过构建决策树来建立输入,便于将新闻事件信号融入进交易策略中.决策树结构可以反映在不同的情况下采取不同交易策略的原因,对于深入理解策略之间的关系、探索新闻与股价的影响机理具有重要意义.因此,本文采用GP算法进行交易策略组合优化并计算相应的择时.

本文GP算法的个体为交易策略组合,以编码的格式出现.算法从一个随机产生的包含若干个体的初始种群开始,模拟一个自然进化过程来改进个体后代的适应度,提升种群质量至无法再优化为止.在该进化过程中,个体不断地根据其适应度值进行“进化”.“进化”的方式有两种,一种是交叉,即不同个体的基因(即策略/信号)重新排列组合成新的个体;另一种是变异,即以一个设定的概率轻微地改变个体的部分基因.

3.1 适应度函数与效率评估

交易策略的适应度通过数据集上的收益率来计算.本文使用多支股票的平均收益率作为适应度函数

$$\begin{aligned} \text{Fitness}(X) &= \left(\sum_{i=1}^N \sum_{j=1}^D C_{ij}(X) R_{ij} \right) \left(\sum_{i=1}^N \sum_{j=1}^D C_{ij}(X) \right)^{-1} \\ &= \left(\sum_{i=1}^N \sum_{j=1}^D C_{ij}(X) \frac{P_{ij}^n - P_{ij}}{P_{ij}} \right) \left(\sum_{i=1}^N \sum_{j=1}^D C_{ij}(X) \right)^{-1}, \end{aligned} \quad (2)$$

其中 n 取 1, 2, 5, 10. N 和 D 分别表示股票的总量和交易日总数. R_{ij} 表示股票 i 在 j 日的收益率. $C_{ij}(X)$ 为控制函数,表示股票 i 在第 j 日是否被选中进行交易, X 为GP算法要计算的交易策略组合.

为准确评估自动交易的效率,本文提出使用收益比率来衡量优化组合的效率.适应度函数存在理论上的取值范围,该值可以反映优化组合的效率.关于取值范围的分析为在股票交易时间区间内所有交易次数中选取最好的“卖出”结果(为正)和最好的“买入”结果(为负).本文将该问题转化为一个从排列组合中寻找最好的 N 个组合与最差的 N 个组合的问题.由于分别为一个正的最大值,一个负的最大值,所以恰好是 $\text{Fitness}(X)$ 的上限和下限.对收益率 R_{ij} 排序,若有 q_s 频次的交易被选中为“卖出”, q_b 频次的交易被选中为“买入”,则 $\text{Fitness}(X)$ 的范围为

$$\frac{1}{q_s} \sum_{i=1}^N \sum_{j=1}^D R_{ij}^{(s)} \leq \text{Fitness}(X) \leq \frac{1}{q_b} \sum_{i=1}^N \sum_{j=1}^D R_{ij}^{(b)}, \quad (3)$$

其中 $\max R_{ij}^{(s)} < R_{ij} \setminus R_{ij}^{(s)}, i = 1, 2, \dots, N, j = 1, 2, \dots, D; \min R_{ij}^{(b)} > R_{ij} \setminus R_{ij}^{(b)}, i = 1, 2, \dots, N, j = 1, 2, \dots, D$. 据此,可以计算策略组合相对于理论值的效率.“买入”和“卖出”的效率分别为

$$\gamma_{\text{buy}} = \frac{1}{q_b} \text{Fitness}(X) \left(\sum_{i=1}^N \sum_{j=1}^D R_{ij}^{(b)} \right)^{-1}, \quad \gamma_{\text{sell}} = \frac{1}{q_s} \text{Fitness}(X) \left(\sum_{i=1}^N \sum_{j=1}^D R_{ij}^{(s)} \right)^{-1}. \quad (4)$$

3.2 树状图规则

GP算法的交易策略组合可以用树状结构表示,结果输出为布尔值的交易信号.当式(2)中 $X \geq 1$ 时,返回值为“真”,表示产生交易信号, $C_{ij}(X) = 1$; 若 $X = 0$,返回值为“假”,表示没有产生交易信号, $C_{ij}(X) = 0$. 交易策略组合至少包含一个技术信号或一个新闻变量,大多数时候是技术信号与新闻事件变量的综合,这些变量通过逻辑运算符“& (AND, 与)”和“| (OR, 或)”组合在一起.

图1树状图展示交易策略组合案例.该规则表示为 $\text{EMA} \& (\text{SMA} | \text{MOM})$, 其产生交易信号的规则是,指数移动平均策略(EMA)产生一个信号,同时简单移动平均策略(SMA)和动量交易策略(MOM)至少产生一个“真”的交易信号.计算 X 时,用乘法规则“ $\times$ ”计算 AND, 用加法规则“ $+$ ”计算 OR, 则图1的规则可以表示为 $X = \text{EMA}(\text{SMA} + \text{MOM})$.

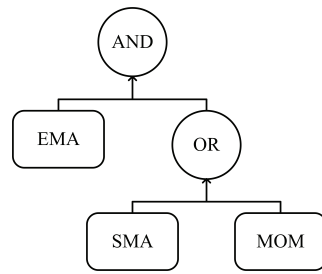


图 1 GP 算法的树状图规则

Fig. 1 Rule of genetic programming

3.3 交叉和变异

GP 算法使用到“交叉”和“变异”两类进化规则。“交叉”包括了选择多种交易策略组合以及决定随机“交叉”的位置。例如，每棵决策树产生一个“交叉”的位置，两个交易策略组合的“交叉”点分裂出的子树相互交换，由此产生新的交易策略并入到新的种群里。分裂出的子树应尽量不重复，否则无法产生有意义的新个体。交叉和变异示例如图 2 所示。图 2 的左边部分展示的是交叉案例，交易策略组合 EMA & (SMA | MOM)，VMA + SMA 经过交叉可以产生 EMA & SMA 和 VMA | (SMA | MOM)。

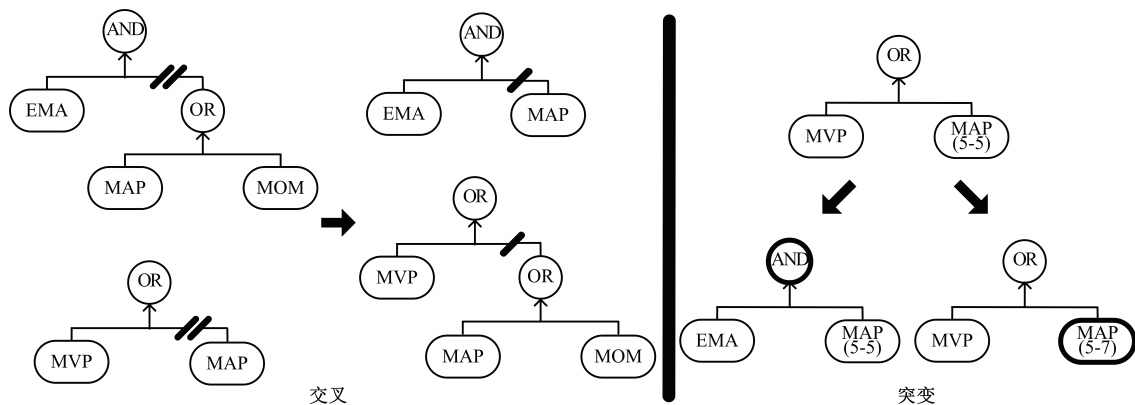


图 2 交叉和变异示例

Fig. 2 Crossover and mutation

“变异”直接对个体操作形成新的个体。为确保“变异”后的策略有意义，本文以两种突变方式改变个体。一种是以一定的概率改变个体的符号。如图 2(右)，将原策略组合中的“AND”改为“OR”。另一种是以一定概率改变随机选择的技术策略的某些参数。如图 2(右)，改变简单移动平均策略(SMA)的滞后天数，从 SMA(20~60)改为 SMA(20~61)。

3.4 算法过程

本文使用 GP 算法搜索优化的交易策略组合并计算相应的最佳择时。由于最佳的交易时间 $t_{v'}$ 与所选择的交易策略 v' 相对应，本文在计算群体 π 适应度 $\text{Fitness}(\pi)$ 时，同时寻找该策略最佳适应天数，即该种策略适应度最高时的交易时间。算法具体步骤如下：

步骤 1 初始种群生成：根据种群规模 σ 产生随机的初始种群 π ，初始化适应度 σ_{old} 为负无穷， σ_{new} 为初始种群的适应度，尝试次数 b 设置为 0。

步骤 2 交叉：随机选择 ρ 个父节点 $\theta = (\pi, \sigma_{\text{new}}, \rho)$ ，通过交叉 $\text{crossover}(\theta)$ 步骤进行优化获得新个体 v' 。

步骤 3 变异：按照 μ_1, μ_2 的概率对 v 进行变异 $\text{mutation}(v, \mu_1, \mu_2)$ ，获得新个体 v' 。

步骤 4 计算适应度和最佳适应天数: 计算新生成个体 v' 的适应度 $\text{Fitness}(v')$, 同时计算相应的最佳适应天数 $t_{v'}$.

步骤 5 选择: 取适应度值较高的新个体 v' 加入到新的种群 π' 中.

步骤 6 终止: 当群体适应度无法继续改进时, 算法会终止, 否则重复 步骤 2~ 步骤 5. 可以使用的终止条件有两种: 1) 给定最大尝试次数 M , 当累计尝试的次数超过最大尝试次数 M 时跳出循环. 2) 给定问题下界的计算方法, 当进化中达到要求的偏差 ϵ 时, 算法终止.

3.5 模型应用框架

图 3 展示了本文提出的算法交易整体框架. 该框架的信号源共分为两部分, 分别为新闻信号和技术指标信号. 基于遗传规划的组合优化算法将上述两方面的信号综合在一起, 通过优化生成最终的交易策略. 为验证组合优化算法的有效性, 运用后续时间窗口产生的交易数据对筛选出的交易策略进行验证. 在实际应用中, 可以设置为时间区间固定的滑动窗口, 以相同的时间长度向后滑动, 使用历史交易数据和新闻数据筛选优化策略, 在新的时间区间上进行测试.

本文主要聚焦框架的前一部分, 阐述如何运用遗传规划算法优化策略, 验证新闻信号和技术交易策略信号的有效性, 进而分析综合策略的有效性. 为证明遗传规划算法的有效性, 引入不同的数据集验证优化组合在新的数据集上的有效性.

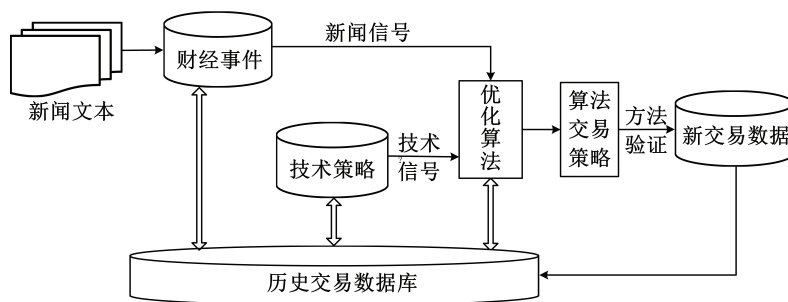


图 3 算法交易整体框架

Fig. 3 The framework of algorithm trading

4 实验与结论

本节通过实验验证多源信号交易策略组合优化模型. 4.1 节首先介绍要使用的数据集. 4.2 节~4.4 节从有效性角度对模型进行验证, 首先计算单独使用技术策略信号或新闻事件信号所产生的收益率, 验证不同源信号对于股票交易的有效性, 然后分析信号叠加后的效应, 证明综合新闻信号与技术策略是有效的. 4.5 节从择时角度分析了如何优化策略组合提升收益率. 4.6 节分析了策略组合的效率. 4.7 节展示了通过 GP 优化算法生成的混合交易策略的有效性. 4.8 节采用更多数据集验证了优化模型表现的稳健性.

4.1 新闻、股价数据集

本文利用网络新闻爬行者, 采集了来自中国证券网、凤凰财经、和讯财经等 30 多家财经新闻网站(频道)的新闻报道共计 870 162 篇. 为了保证综合量化的有效性和公平性, 忽略每月仅有 1 篇以下新闻的网站. 由于每条新闻都有发布时间, 不在预处理过程区分“原创”和“转载”并默认所有网站的影响力都相同. 对上述新闻进行清洗、去重、分词和词性标注等预处理, 获得与股市相关的财经新闻共 15 217 篇, 时间覆盖 2015-10-01~2018-09-30. 利用 Chen 等^[32]提出的方法, 将非结构化的新闻数据抽取为结构化的新闻事件信息, 抽取具有代表性的“沪深 300”标的股票相关的新闻事件, 并参考韩永峰等^[33]提出的方法去掉冗余事件信息, 最终获得可以使用的时间跨度各为 1 年的 3 个新闻事件数据集, 包括 14 类新闻事件的数据集, 各类财经事件数量统计如表 2 所示.

表 2 新闻数据集事件统计

Table 2 Statistics of news events

序号	事件类型	数据集 1	数据集 2	数据集 3	序号	事件类型	数据集 1	数据集 2	数据集 3
1.	股价上涨	964	1 207	1 897	8.	公司利润下滑	89	79	82
2.	股价下跌	1 016	1 592	2 091	9.	公司负债良好	94	84	88
3.	公司业绩上扬	349	386	403	10.	公司负债堪忧	63	61	62
4.	公司业绩下滑	246	289	257	11.	公司高层变动	72	95	99
5.	公司声誉提升	207	198	205	12.	公司合作	64	65	68
6.	公司声誉受损	43	56	99	13.	公司业务扩张	74	107	121
7.	公司利润上升	174	166	133	14.	公司并购	36	30	93

本文使用 Wind 公司提供的股市交易数据, 时间区间与新闻数据集对应, 分别为 2015-10-01~2016-09-30, 2016-10-01~2017-09-30 和 2017-10-01~2018-09-30 的“沪深 300”股票¹数据集和所抽取的新闻事件集合进行股票交易实验, 数据集可满足此时间范围的所有收益率及交易策略的计算需求. 考虑到新闻事件信号需要经过一定时间的积累, 本文未将数据集分割进行测试. 基本思路为, 首先将数据集 1 作为训练数据集, 证明本文方法的可行性同时获得相应的策略组合结果; 然后将数据集 2 和数据集 3 作为测试集, 将该数据集 1 的结果运用到数据集 2 和数据集 3, 报告这些策略的稳健性分析结果.

4.2 新闻事件信号分析

表 3 列出了事件发生 x 天的收益率与事件关系的统计分析结果, $x = 0, 1, 2, 5, 10$. 其中 I_p 表示事件的预定义影响力, Freq 表示数据集中该类事件发生的频次. $d_x(\%)$ 表示预测准确的百分比, 即预定义影响力方向与股价波动方向相同的百分比率. p_x 表示 R_x 显著程度的双尾 t-检验值.

观察表 3 可以发现大部分股票收益率与预定义影响力方向一致. 即如果预定义影响力为正($I_p > 0$), 则收益率 R_0, R_1, R_2, R_5 和 R_{10} 大多数都为正; 如果预定义影响力为负($I_p < 0$), 则收益率 R_0, R_1, R_2, R_5 和 R_{10} 大多数为负. 有两类事件的方向是波动的, 即“公司高层变动”与“公司业务扩张”, 这两类事件的预定义影响都是轻微的正向(+1), 而实际情况是收益率有时轻微正向, 有时轻微负向. 由于这两类事件的样本数比较少, 可能受其他事件干扰, 且这种小的负面收益率在 95% 的置信水平上并不显著, 所以对整体的结果影响不大. 综上, 本文使用的预定义影响力 I_p 是基本可靠的.

观察表 3 可知, 考虑事件当日收益率 R_0 时表现最优的是“股价上涨”事件, 平均收益率为 1.63%, 有 83% 该类型事件发生后产生了正面收益. 另外, “公司声誉提升”等事件也都有 60% 以上的正确预测比率. 意味着上述类型新闻事件在报道当天会对价格产生一个向上的驱动. 当然, 在某些情况下, 与这些事件同现的其他事件也可能产生反方向的驱动力. 同理, 可以发现“股价下跌”、“公司声誉提升”等事件 I_p 较大且 d 较高, 说明上述信号都是比较强烈的信号, 容易主导价格波动. 考虑随后几日的收益率情况, 即 R_0, R_1, R_2, R_5 和 R_{10} , 大部分事件(如“公司业绩上扬”, “利润上升”和“利润下滑”等), 出现频率均比较高, 其收益率在经济和统计意义上均显著, 而且与预定义影响力方向相同. 由此可知, 一些财经新闻事件报道后, 其影响力不会即时发挥, 而是在随后的几个交易日中表现出来, 同时也表明中国股票市场不是完全有效市场.

表 4 报告了预定义影响力 I_p 和收益率 R_x 之间的皮尔逊相关(Person's correlation) 系数, Cor 表示相关性系数, p 表示显著程度. 观察该表可发现收益率 R_x 和预定义影响力 I_p 之间的相关系数均在 80% 左右, 表现为显著的正相关.

综合表 3 和表 4 可得出两个结论: 首先, 预定义事件影响力和股价波动紧密, 且影响力强度与股价波动幅度具有较强的线性相关性. 其次, 从新闻中获取的事件信息可以融入交易策略, 事件对股价的影响力是明显的. 综上所述, 将新闻事件作为股票交易信号从统计角度分析是可行的.

¹本文以 2016 年 6 月修订版的“沪深 300”股票为研究对象, 样本股交易数据来源于 Wind 数据库. 另外, 为便于统一标记, 时间节点选取整月份, 如 10-01, 09-30, 股票样本数据以实际交易日为准.

表 3 财经事件与股票收益率统计分析

Table 3 Statistics of financial events and stock returns

事件	I_p	Freq	R_0	d_0	p_0	R_1	d_1	p_1	R_2	d_2	p_2	R_5	d_5	p_5	R_{10}	d_{10}	p_{10}
公司业绩上扬	3	349	0.89%	62%	0.00	0.77%	58%	0.04	1.21%	57%	0.01	1.03%	54%	0.00	0.94%	52%	0.03
公司利润上升	3	174	0.39%	49%	0.49	0.40%	47%	0.16	0.57%	48%	0.67	0.64%	70%	0.02	0.78%	67%	0.00
公司并购	3	36	0.17%	53%	0.04	0.25%	53%	0.07	0.34%	52%	0.15	0.10%	52%	0.03	0.08%	55%	0.54
股价上涨	2	964	1.63%	83%	0.00	1.21%	80%	0.02	1.24%	73%	0.01	1.18%	71%	0.01	1.48%	69%	0.00
公司声誉提升	2	207	0.37%	61%	0.08	0.42%	57%	0.34	0.36%	55%	0.36	0.28%	58%	0.52	0.19%	46%	0.17
公司负债良好	1	94	0.15%	53%	0.19	0.37%	49%	0.45	0.04%	49%	0.04	0.26%	45%	0.07	0.22%	49%	0.07
公司合作	1	64	0.13%	47%	0.00	0.17%	49%	0.38	0.04%	49%	0.10	0.07%	55%	0.31	0.12%	49%	0.03
公司高层变动	1	72	-0.10%	49%	0.23	0.16%	59%	0.47	-0.10	49%	0.18	0.24%	49%	0.18	0.15%	56%	0.19
公司业务扩张	1	74	0.23%	49%	0.21	0.57%	48%	0.13	-0.47	49%	0.24	-0.15%	48%	0.03	0.45%	48%	0.23
公司负债堪忧	-1	63	-0.18%	55%	0.32	-0.25%	56%	0.85	-0.64%	56%	0.10	-0.18%	49%	0.22	-0.28%	58%	0.03
股价下跌	-2	1 016	-0.98%	81%	0.01	-1.19%	76%	0.00	-1.24%	68%	0.00	-0.86%	66%	0.04	-0.72%	66%	0.01
公司业绩下跌	-2	246	-0.53%	51%	0.09	-0.86%	45%	0.02	-0.78%	49%	0.28	-0.85%	49%	0.52	-0.72%	49%	0.44
公司声誉受损	-2	43	-0.24%	50%	0.09	-0.38%	49%	0.27	-0.19%	49%	0.04	-0.10%	48%	0.52	-0.15%	49%	0.03
公司利润下滑	-3	89	-1.45%	68%	0.28	-0.86%	63%	0.12	-0.69%	60%	0.10	-0.47%	62%	0.05	-0.43%	66%	0.01
	3	491		58%			56%			55%			55%			56%	

表 4 预定义影响力 I_p 和收益率 R_x 的关系

Table 4 Relationships of predefined impact I_p and return R_x

	R_0	R_1	R_2	R_5	R_{10}
Cor	0.809	0.862	0.823	0.804	0.787
p	0.000	0.000	0.001	0.000	0.000

4.3 多源信号作用分析

技术策略 表 5 报告了使用 SMA, VMA, EMA 和 MOM 四种技术交易策略的收益率. 其中“频次”指标为样本区间内发出买入信号或者卖出信号的次数, R_0, R_1, R_2, R_5 和 R_{10} 分别为交易信号发出当天, 以及持有 1 日, 2 日, 5 日和 10 日的平均收益率². 本文主要通过分析买入或者卖出交易信号发出后各平均收益率的大小判断交易策略的优劣.

表 5 技术策略的收益率表现

Table 5 Returns of the technical strategies

技术策略	频次	R_0	R_1	R_2	R_5	R_{10}
Penal A: 买入信号						
SMA(20, 60)	877	0.82%	0.38%	0.81%	1.31%	1.10%
VMA(20, 60)	3 870	0.21%	0.16%	0.40%	0.32%	0.20%
EMA(20, 60)	788	0.69%	0.52%	1.93%	2.73%	3.09%
MOM(60)	2 762	1.72%	0.11%	0.12%	0.21%	-0.28%
Penal B: 卖出信号						
SMA(20, 60)	752	-0.81%	-0.25%	-0.47%	-0.87%	-1.99%
VMA(20, 60)	3 791	-1.18%	-0.08%	-0.13%	-0.33%	-0.22%
EMA(20, 60)	642	-1.21%	-0.08%	-0.48%	-0.44%	-0.09%
MOM(60)	2 604	-2.03%	0.03%	0.04%	-0.23%	-0.39%

观察表 5 可见, 指数移动平均策略(EMA)相比于其他三种策略表现更优. 买入信号发出当天, 指数移动平均策略(EMA)的平均收益率为 0.69%, 同时随着持有时间增加收益率呈现上升趋势(说明买入信号发出后持有股票将在接下来的一段时间内获得正收益), 基于价格的移动平均策略(SMA)表现次之(均大幅高于其

²篇幅限制, 本文只报告了 20 日为短期, 60 日为长期的结果. 另, 本文中 MOM 策略效果不佳的原因可能是中国股票市场的价格趋势呈现短期化.

他两种策略 VMA 和 MOM); 发出卖出信号后除了动量交易策略(MOM)表现不稳定之外, 其他三类策略的收益均为负, 表明发出卖出信号后股价出现下跌(如果此时不持有股票, 则可以避免由于股价下跌而带来的损失). 以上分析说明技术策略预测是有效的, 其中 EMA 表现最佳, SMA 次之.

新闻信号 如果某新闻事件发生后产生的收益率都为正, 且 R_0 在 0.2% 以上, 则输出买入信号. 如果某新闻事件发生后产生的收益率都为负, 且 R_0 在 -0.2% 以下, 则输出卖出信号. 参照表 3, 将新闻信号分为买入和卖出两类, 分别计算这两类信号产生的收益率. 表 6 展示了使用新闻事件信号进行交易所获得的平均收益率. 由表 6 可以看出新闻事件发出买入信号后, 产生的当日平均收益、次日平均收益率以及 2 日、5 日和 10 日的平均收益率都为正, 且由此产生的 R_x 值至少为 1.5%, 说明买入信号产生的收益率是可观的, 即买入信号是准确的. 同样地, 卖出信号发出后, 产生的当日平均收益率、次日平均收益率以及 2 日、5 日和 10 日的平均收益率都为负, 且 R_x 值都低于 -1.0%, 说明卖出信号也是准确的. 多频次的分析、综合可见, 本文抽取的新闻事件所产生的交易信号对股票交易而言是正确的, 会给股票的自动交易带来较好的收益.

表 6 新闻信号产生的收益率

Table 6 Returns of news signals

信号	频次	R_0	R_1	R_2	R_5	R_{10}
买入	1043	1.567%	1.625%	1.531%	1.783%	2.451%
卖出	729	-1.743%	-1.682%	-1.515%	-1.331%	-1.227%

比较独立使用新闻信号与独立使用技术策略所产生的不同情况. 对比表 5 和表 6, 从结果上看, 使用技术信号“买入”时并不都能获得正的收益率, 如动量信号(MOM)会在 10 日后消失. 而新闻信号产生的收益率都为正, 证明新闻信号产生的买入质量更优. 同样的情况也出现在卖出信号中, 新闻信号产生的最低收益率 -1.227%, 这明显高出大部分技术信号产生的收益率. 就中国市场而言, 新闻买入信号和卖出信号获得的收益率更高, 但是 EMA 与新闻信号相比在后期(5 日, 10 日)所产生的收益率有一定的优势. 下一步, 本文将通过新闻事件信号与技术策略信号的融合, 证明获得更高收益率的可行性.

4.4 信号叠加分析

将新闻事件“买入”的信号与技术策略中“买入”的信号组合, 将新闻“卖出”的信号与技术策略中“卖出”的信号组合, 只有两个信号同时满足时才会产生一个交易信号, 计算技术策略信号和新闻信号的组合策略最终可以获得的收益率. 表 7 展示的是新闻和技术策略信号叠加后获得的收益率³. 观察多次试验, 计算每一类组合所获得的平均收益率. “买入”和“卖出”时产生的收益率见表 7.

表 7 新闻与技术信号综合产生的收益率

Table 7 Returns of the combined strategies by using news and technical signals

Panel A: 买入信号						
信号	频次	R_0	R_1	R_2	R_5	R_{10}
News & SMA(20, 60)	763	1.587%	1.524%	1.445%	1.783%	1.668%
News & VMA(20, 60)	274	1.472%	1.269%	1.578%	1.346%	1.849%
News & EMA(20, 60)	379	1.842%	1.963%	2.037%	2.132%	3.442%
News & MOM(60)	472	1.743%	0.682%	0.515%	-0.331%	1.227%
Panel B: 卖出信号						
信号	频次	R_0	R_1	R_2	R_5	R_{10}
News & SMA(20, 60)	248	-1.567%	-1.625%	-1.531%	-1.783%	-1.451%
News & VMA(20, 60)	185	-1.378%	-1.276%	-1.672%	-1.563%	-1.604%
News & EMA(20, 60)	313	-1.673%	-1.894%	-1.932%	-2.023%	-2.374%
News & MOM(60)	328	-1.209%	-0.367%	-0.428%	0.374%	-1.459%

³篇幅限制, 本文只报告了 20 日为短期, 60 日为长期的结果.

观察表 7 可以发现,“买入”的大部分收益率为正,而“卖出”的大部分收益率为负,说明上述策略组合都产生了有效的交易信号. 新闻与技术信号的组合中,新闻与指数移动平均(News & EMA) 组合策略可以产生较高的收益率. 当出现“买入”信号时,其当日收益率、次日收益率及 2 日,5 日,10 日收益率都会有一定的涨幅. 而当出现“卖出”信号的时候,该组合策略的当日收益率、次日收益率以及 2 日,5 日,10 日的日收益率都会出现一定的跌幅,说明该组合所给出的“买入”和“卖出”信号是有效的,且在这几种组合中表现最优. 对比表 7 和表 5,表 6 的结果,发现此种组合产生的收益率比单独使用新闻或单独使用任何一种技术策略的效果都要好. 以 News & EMA 组合为例,该组合发出“买入”信号时,其最高收益率达 3.442%,超过新闻的最高收益率(2.451%)和单独使用技术策略的最高收益率(3.09%). 该组合发出“卖出”信号时,收益率最大为 -2.374%,也比单独使用新闻信号(-1.743%)和技术策略(-2.03%)高. 其它 4 类组合策略也都出现了类似的情况,说明将独立的技术策略和新闻事件信号综合起来决定股票交易可以产生比分开考虑更高的收益率. 下一步本文将基于多个技术信号、新闻信号自动产生更多更复杂的交易策略,并利用 GP 算法计算出最优的组合.

4.5 交易策略优化与择时分析

按照算法 1 提出的 GP 算法的实验步骤,本文对交易策略组合进行优化并计算所能获得的最佳交易时间. 设置 GP 算法交叉的概率为 0.5,变异算子 μ_1 和 μ_2 的概率也都为 0.5,算法最多迭代 30 次,股票交易时间为 1 日~60 日. 然后通过算法 1 提供的方法,获得在该段时间内的最优交易策略. 讨论某个策略时,本文不仅会比较该组合策略和其他策略的收益率情况,同时也比较该方法与自身相比在何时会达到最优. 具体而言,在计算某个算法的收益率时,同时计算该算法在 1 日~60 日内每天的收益率,比较究竟哪一天会产生最高的收益率.

首先随机化初始种群,设置种群规模为 50. 计算种群中每个个体的买入信号或卖出信号,信号出现后,计算其在 x 日内的收益率, $x = 1, 2, \dots, 60$, 取最高收益率为该个体的收益率 R_x . 通过 GP 算法的计算,获得表现最好的 10 种策略组合,即 GP 算法获得的 10 个最优的策略组合,如表 8 所示.

表 8 策略组合优化及最佳交易时间计算
Table 8 Results of the optimized strategies and their corresponding trading time

买入信号									
序号	决策树	频次	R_x	x	序号	决策树	频次	R_x	x
1	EMA(19, 47)& News	64	3.667% (2.08)	6	6	SMA(13, 49)&(News VMA(15, 49))	93	2.231% (1.41)	4
2	VMA(20, 57)&(News VMA(20, 60))	108	2.988% (1.86)	8	7	VMA(20, 57)&(MOM(23) VMA(18, 62))	109	2.109% (1.40)	7
3	MOM(27)&(SMA(17, 48) News)	120	2.752% (1.90)	7	8	MOM(27)&(SMA(17, 48) News)	112	2.063% (1.34)	8
4	MOM(6) News	137	2.633% (1.82)	6	9	(MOM(6) News) & SMA(19, 53)	123	2.048% (1.37)	6
5	VMA(19, 51)& News SMA(20, 58)	142	2.587% (1.67)	8	10	VMA(19, 53)& News SMA(20, 58)	109	2.004% (1.39)	5
卖出信号									
序号	决策树	频次	R_x	x	序号	决策树	频次	R_x	x
1	EMA(20, 51)& News	121	-2.967% (1.91)	5	6	SMA(20, 51)& News	164	-1.765% (1.65)	6
2	EMA(20, 57) VMA(18, 49)	118	-2.764% (1.85)	6	7	VMA(20, 57) VMA(18, 47)	109	-1.664% (1.53)	7
3	EMA(27, 55)&(MOM(17) News)	141	-1.932% (1.57)	9	8	EMA(27, 55)& MOM(17)	88	-1.581% (1.60)	9

表 8 (续)
Table 8 (Cotinue)

买入信号									
序号	决策树	频次	R_x	x	序号	决策树	频次	R_x	x
4	MOM(24) News	94	-1.889% (1.61)	7	9	(MOM(24) News) & VMA(21, 53)	106	-1.503% (1.34)	9
5	VMA(19, 47) MOM(20)	89	-1.772% (1.63)	6	10	MOM(24, 27) News	128	-1.478% (1.28)	7

观察表 8 可见, 决策树表示 GP 算法中的策略组合; 频次表示所有实验中最佳策略使用该策略的次数; R_x 为收益率, 括号内为该策略对应的夏普比率, x 为达到该收益率的天数. 经 GP 算法优化后的 10 个买入、卖出交易策略组合中, 新闻因子分别出现了 9 次和 6 次, 新闻信号在最优策略组合中高频率的出现说明其在交易策略组合中的重要性. 从频次上看, 上述策略的频次都在 60 次以上, 证明上述策略是可靠的. 而且, 对比发现高收益率往往对应较高的夏普比率, 从多角度证明了本文所选策略的有效性. 从最佳的持有时间来看, 综合后的买入和卖出策略大部分在 5 个交易日之后, 10 个交易日日内达到最大值, 说明新闻信号产生的影响会经过一段时间的发酵.

将表 8 与表 5、表 6 和表 7 进行比较. 以表现最好的 EMA(19~47) & News 为例, 融合后的交易策略表明新闻信息增强了 EMA 的效果, 比单独使用 EMA 策略或 News 信号提升了收益. 而且此处的 EMA 中短期和长期时间参数分别为 19 和 47, 与原来的 20 和 60 不同, 说明经过 GP 算法的变异步骤优化了组合的效果, 获得了更高的收益率. 进一步分析发现新闻信号在买入和卖出策略扮演不同的角色. 新闻信号在买入策略中出现的次数和重要性(本文认为出现&即“与(AND)”的关系, 新闻显得更重要一些)比卖出策略中出现的多, 说明更加重要. 另外, MOM 策略效果不佳, 可能是由于为了和新闻信号组成更好的组合, 牺牲了部分时间, 但这也与韩豫峰等人对中国市场动量效应的观察基本相符^[34]. 综上, 新闻事件可以作为一种交易信号提升技术策略的效果, 为设计自动交易系统提供了一种思路. 由于本文获得的自动交易策略组合是通过启发式算法获得的, 因此上述结果对一般交易策略选择也具备一定程度的启发意义, 即可以通过新闻信号与技术策略结合的方法更好地设计交易策略, 提高其表现.

4.6 交易策略组合效率分析

图 4 是根据式(3)计算所得的适应度函数的取值范围, 由图 4 可知随着频次的增加, 该值呈下降趋势. 另外, 从时间角度看, 时间越长上限和下限的幅度越大. 根据式(4), $\gamma_{\text{buy}} = 3.667/56.1 = 6.5\%$, $\gamma_{\text{sell}} = 2.976/26.4 = 11.2\%$.

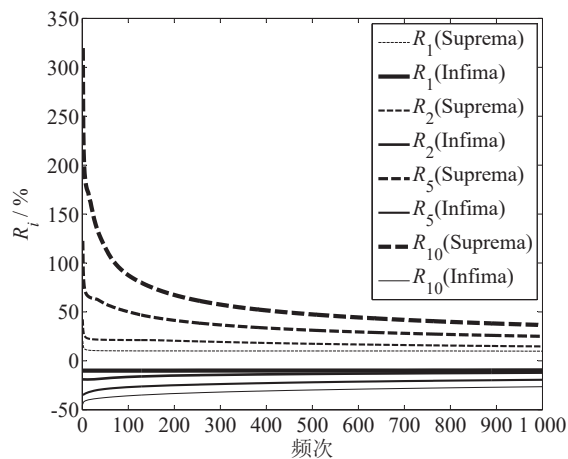


图 4 交易策略组合收益率的上、下限值

Fig. 4 The upper and lower return limit of combined trading strategy

本文的优化交易策略组合收益率仅仅占最大可能收益率的 10% 左右, 还远远没有达到取值范围的理论值. 由于上述交易策略组合收益率的上、下限值实际上取的是股票连续大涨(大跌)的情况, 即最好和最差的极限情况. 所以, 上述结果表明通过技术策略与新闻信号综合的策略组合对连续大涨(大跌)等极限情况的预测能力是有限的.

4.7 交易策略优化有效性分析

前文展示了技术信号、新闻信号以及混合信号的有效性, 本小节将展示通过 GP 优化算法生成的混合交易策略在新的时间区间上的有效性. 本文将数据集 1, 数据集 2 与数据集 3 合并, 通过设置滑动窗口构建策略生成与检验的数据集组合. 考虑到本文设计方法依赖于统计方法, 需要足够时长生成并检验策略, 本文将滑动窗口时间跨度设置为 1 年, 每次向前移动 1 个季度, 运用窗口内新闻数据集和交易数据集产生交易信号, 基于 GP 算法产生优化的交易策略, 选取其中 20 种最优的策略运用到下一个季度中, 计算在该季度的中表现最好的 10 组交易策略组合, 统计包含新闻信号(#News)的交易策略的个数, 计算这些策略平均收益率 R_x , 最好收益率 R_x^{Best} 和最差收益率 R_x^{Worst} , 验证上述策略的有效性.

表 9 列出了策略组合有效性分析结果.

表 9 策略组合有效性分析
Table 9 Effectiveness Analysis of the optimized strategies

策略生成时间	策略测试时间	买入信号				卖出信号			
		#News	$\bar{R}_x$	R_x^{Best}	R_x^{Worst}	#News	$\bar{R}_x$	R_x^{Best}	R_x^{Worst}
2015-10-01~2016-09-30	2016-10-01~2016-12-31	5	2.247%	2.684%	1.304%	6	-1.727%	-2.309%	-1.074%
2016-01-01~2016-12-31	2017-01-01~2016-03-31	4	2.012%	2.453%	1.783%	5	-2.494%	-3.067%	-1.272%
2016-04-01~2017-03-31	2017-04-01~2017-06-30	4	1.747%	2.319%	1.048%	6	-1.972%	-2.470%	-1.346%
2016-07-01~2017-06-30	2017-07-01~2017-09-30	5	1.407%	2.821%	1.016%	5	-2.744%	-3.219%	-1.578%
2016-10-01~2017-09-30	2017-10-01~2017-12-31	3	1.367%	2.040%	1.002%	7	-2.479%	-2.382%	-1.434%
2017-01-01~2017-12-31	2018-01-01~2018-03-31	2	1.675%	1.984%	0.930%	4	-2.975%	-3.264%	-1.799%
2017-04-01~2018-03-31	2018-04-01~2019-06-30	4	2.024%	2.754%	1.469%	5	-2.863%	-3.119%	-1.882%
2017-07-01~2018-06-30	2018-07-01~2018-09-30	4	1.458%	2.609%	1.011%	6	-2.673%	-3.005%	-1.794%

表 9 表明, 在数据集覆盖的 8 个时间区间中, 由本文所提出的方法筛选所得的交易策略在新的一季度中表现都较好; 在买入策略中, 相关策略的平均收益率稳定在 2% 左右, 且最差的收益率接近 1%. 在卖出策略中, 相关策略的最好收益率达到 3% 左右, 比买入策略较好. 分析是由于负面新闻信号与股市共同作用导致卖出策略的表现较好. 而且在表现较好的策略中, 包含有新闻信号的策略超过了一半, 说明新闻信号的重要性. 由于相关策略都产生了较好的收益率, 说明本文所提出的策略优化模型是有效的.

4.8 交易策略优化稳定性分析

前文展示了交易策略组合优化的基本过程与结果, 本节使用相同的参数在数据集 2 和数据集 3 重复上述实验, 并报告实验的关键结果. 实验发现, 尽管不同类型的事件在上述两个时间区间内并没有保持一致的比例, 但数据集 2 和数据集 3 的预定义影响力 I_p 与收益率 R_x 都通过了统计检验, 表明利用上述事件研究市场收益率情况具有一定的普适性. 重复技术策略、新闻事件的作用分析, 表 10 展示了在上述两个数据集上观察到的决策树、频次、收益率(括号内为夏普比率)以及最佳天数 x . 篇幅所限, 本文只取收益率排名最高的前 5 种策略组合的结果.

由表 10 可以发现, 几乎表中所有的策略都需要新闻事件信号, 而且都使用了 & 运算, 表明新闻在策略组合优化中发挥了重要的作用. 排名靠前的策略夏普比率也相对较高, 进一步说明了新闻信号的重要性. EMA&News 的组合策略同样表现比较优秀, 说明此种策略组合的表现具有稳健性.

表 10 策略组合优化结果(数据集 2 和数据集 3)
Table 10 Results of the optimized strategies (Datasets 2 & 3)

买入信号									
数据集 2					数据集 3				
序号	决策树	频次	R_x	x	序号	决策树	频次	R_x	x
1	EMA(19, 54)& News	79	2.972% (1.97)	5	1	VMA(19, 51)& News EMA(20, 58)	105	2.861% (1.82)	5
2	VMA(20, 57)&(News VMA(20))	112	2.648% (1.83)	7	2	EMA(20, 57)&(MOM(23) VMA(18, 62))	92	2.649% (1.79)	6
3	MOM(27)&(SMA(17, 48) News)	108	2.341% (1.77)	7	3	MOM(27)&(SMA(17, 48) News)	106	2.303% (1.71)	7
4	SMA(13, 49)&(News VMA(15, 49))	121	2.335% (1.64)	9	4	(MOM(12) News) & SMA(19, 53)	114	2.188% (1.59)	5
5	EMA(19, 51)& News SMA(20, 58)	122	2.187% (1.60)	6	5	VMA(19, 53)& News SMA(20, 58)	121	2.014% (1.52)	6
卖出信号									
数据集 2					数据集 3				
序号	决策树	频次	R_x	x	序号	决策树	频次	R_x	x
1	VMA(19, 51)& News SMA(20, 58)	121	-3.017% (2.15)	6	1	VMA(21, 55)& News	135	-2.984% (2.09)	8
2	EMA(20, 57) VMA(19, 49)	118	-2.814% (2.07)	5	2	VMA(25, 46)&(MOM(21) News)	97	-2.754% (1.88)	6
3	EMA(27, 55)&(MOM(16) News)	141	-2.482% (1.92)	8	3	EMA(27, 55)& MOM(18)	92	-2.601% (1.75)	8
4	MOM(20) News	94	-1.909% (1.83)	7	4	VMA(20, 57) VMA(19, 47)	113	-2.413% (1.66)	7
5	EMA(27, 55)& MOM(17)	89	-1.892% (1.54)	7	5	MOM(24, 27) News	109	-2.048% (1.37)	9

比较表 10 和表 8 可知, 同样的算法在不同数据集中会产生一定差异. 前 5 名的策略组合在时间区间选择上稍微有一些变化, 但总体呈现出比较一致的结果. 在最佳交易时间上也与数据集 1 类似, 进一步证明了新闻事件对股市的影响会经历一段时间发酵. 上述结果表明本文提出的策略优化方法具有一定的鲁棒性.

5 结束语

本文基于 GP 算法提出了一种考虑新闻事件因子的股票交易策略优化模型, 通过中国股市数据分析了新闻与股价的影响关系并从有效性、效率以及稳健性三方面考察了策略组合优化模型. 实验表明, 本文给出的新闻预定义影响力和股票收益率呈现了较好的相关关系; 将新闻变量融入到技术策略里能产生比单独使用技术策略更高的收益率, 通过本文设计的算法不仅能优化交易策略组合, 而且可以计算出策略组合对应的最佳交易时间; 使用策略优化模型能产生更好的交易策略. 包含新闻信号的交易策略的最佳交易时间一般在两周(10 个交易日)内, 说明新闻所产生的效应也要经过一段时间的发酵, 具有一定的时效性. 上述结论同时在 3 个数据集上得到验证, 说明本文提出的交易策略组合优化模型具有一定的稳健性.

本文的实验结果是通过算法获得的交易策略组合, 主要适用于自动交易系统, 但对于现实系统中形成成熟的交易策略具有启发意义. 由于影响股价波动的因素多且复杂, 无法通过还原法对影响股价的每一个事件以及每一个因素都进行具体的分析. 本文只考虑了单个事件与收益率的统计关系, 在实践中运行自动交易时, 需根据个股具体情况作进一步改进. 此外, 实际交易系统非常复杂, 需要进一步检验所生成方法的其他指标, 以便于更进一步地检验本方法生成策略的有效性和稳健性.

参考文献:

- [1] Neely C J, Rapach D E, Tu J, et al. Forecasting the equity risk premium: The role of technical indicators. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2013, 60(7): 1772–1791.
- [2] Chincarini L B, Kim D. *Quantitative Equity Portfolio Management: An Active Approach to Portfolio Construction and Management*. New York: Business Expert Press, 2006: 134–209.
- [3] Merrill A. *Behavior of Prices on Wall Street*. New York: Analysis Press, 1966: 15–31.
- [4] Engelberg J E, Reed A V, Ringgenberg M C. How are shorts informed: Short sellers, news, and information processing. *Journal of Financial Economics*, 2012, 105(2): 260–278.
- [5] Hendershott T, Livdan D, Schürhoff N. Are institutions informed about news. *Journal of Financial Economics*, 2015, 117(2): 249–287.
- [6] Loughran T, McDonald B. Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, 2016, 54(4): 1187–1230.
- [7] 寇明婷, 杨海珍, 杨晓光. 中国股票市场的“塔西陀陷阱”效应研究. *系统工程学报*, 2018, 33(6): 801–811.
Kou M T, Yang H Z, Yang X G. Study of “Tacitus-trap” effect in Chinese stock market. *Journal of Systems Engineering*, 2018, 33(6): 801–811. (in Chinese)
- [8] 王 泰, 高 闯, 胡祥恩. 突发重大新闻事件中基于兴趣的网民活跃度模型. *系统科学与数学*, 2014, 34(3): 294–308.
Wang T, Gao C, Hu X E. Interest-based user activity model in breaking news events. *Journal of Systems Science and Mathematical*, 2014, 34(3): 294–308. (in Chinese)
- [9] Ewalds S G, Schauten M B J, Steenbeek O W. De informatiewaarde van kwartaalcijfers. *Maandblad voor Accountancy en Bedrijfseconomie*, 2000, 7(8): 333–341.
- [10] Yuan Y. Market-wide attention, trading, and stock returns. *Journal of Financial Economics*, 2015, 116(3): 548–564.
- [11] 刘 锋, 叶 强, 李一军. 媒体关注与投资者关注对股票收益的交互作用: 基于中国金融股的实证研究. *管理科学学报*, 2014, 17(1): 72–85.
Liu F, Ye Q, Li Y J. Impacts of interactions between news attention and investor attention on stock returns: Empirical investigation on financial shares in China. *Journal of Management Science in China*, 2014, 17(1): 72–85. (in Chinese)
- [12] 卫 强, 赵 羨, 张遵强, 等. 基于投资者关注的股价走势预测与交易策略设计—股票间交叉模式视角. *系统工程理论与实践*, 2016, 36(6): 1361–1371.
Wei Q, Zhao X, Zhang Z Q, et al. Stock price prediction and trading strategy design based on investor’ attention: Cross patterns perspective between stocks. *Systems Engineering: Theory & Practice*, 2016, 36(6): 1361–1371. (in Chinese)
- [13] 王建新, 饶育蕾, 彭叠峰. 什么导致了股票收益的“媒体效应”: 预期关注还是未预期关注. *系统工程理论与实践*, 2015, 35(1): 37–48.
Wang J X, Rao Y L, Peng D F. What drives the stock market “media coverage effect”: Expected media attention or unexpected or unexpected media attention. *Systems Engineering: Theory & Practice*, 2015, 35(1): 37–48. (in Chinese)
- [14] Warner J B, Watts R L, Wruck K H. Stock prices and top management changes. *Journal of Financial Economics*, 1988, 20(1): 461–492.
- [15] Rosen R J. Merger momentum and investor sentiment: The stock market reaction to merger announcements. *The Journal of Business*, 2006, 79(2): 987–1017.
- [16] 万 谍, 王军波, 杨晓光. 中国股市暴涨暴跌前有迹象吗. *系统工程学报*, 2016, 31(5): 643–656.
Wan D, Wang J B, Yang X G. Are there any clues before large price changes in Chinese stock markets. *Journal of Systems Engineering*, 2016, 31(5): 643–656. (in Chinese)
- [17] 张 磊, 王延章, 陈雪龙, 等. 面对突发事件应急决策的情景建模方法. *系统工程学报*, 2018, 33(1): 1–12.
Zhang L, Wang Y Z, Chen X L, et al. Scenario modelling method for unconventional emergency decision-making. *Journal of Systems Engineering*, 2018, 33(1): 1–12. (in Chinese)

- [18] Burch T R. Who moves markets in a sudden market-wide crisis: Evidence from Nine-Eleven. *Journal of Financial and Quantitative Analysis*, 2016, 51(2): 463–487.
- [19] 赵静梅, 申 宇, 吴风云. 天灾、人祸与股价: 基于地震、群体骚乱事件的研究. *管理科学学报*, 2014, 17(4): 19–33.
Zhao J M, Shen Y, Wu F Y. Natural disasters and social violence events and stock prices. *Journal of Management Science in China*, 2014, 17(4): 19–33. (in Chinese)
- [20] Schumaker R P, Zhang Y, Huang C N, et al. Evaluating sentiment in financial news articles. *Decision Support Systems*, 2012, 53(3): 458–464.
- [21] 闵 峰, 黄创霞, 文风华, 等. 宏观经济、投资者情绪和股票市场收益. *系统科学与数学*, 2017, 37(2): 370–382.
Min F, Huang C X, Wen F H, et al. Investor sentiment, macro economy and stock market returns. *Journal of Systems Science and Mathematical*, 2017, 37(2): 370–382. (in Chinese)
- [22] 刘 泉, 荣莉莉, 颜克胜. 考虑媒体效应的微博网络热点事件舆论演化模型. *系统工程学报*, 2018, 33(5): 577–588.
Liu Q, Rong L L, Yan K S. Public opinion evolution model of micro-blog network hot event with mass media impact. *Journal of Systems Engineering*, 2018, 33(5): 577–588. (in Chinese)
- [23] 王 轩, 杨海珍. 基于互联网搜索指数的多因素集成下人民币汇率预测. *系统工程学报*, 2017, 32(3): 360–369.
Wang X, Yang H Z. Research on the forecasting of RMB exchange rate based on multi factor integration and web-searching index. *Journal of Systems Engineering*, 2017, 32(3): 360–369. (in Chinese)
- [24] 李志生. 最优交易策略问题的动态求解方法. *中国管理科学*, 2008, 16(3): 102–108.
Li Z S. A dynamic programming approach for constructing optimal trading strategy. *Chinese Journal of Management Science*, 2008, 16(3): 102–108. (in Chinese)
- [25] 陈 艳, 王宣承. 基于变量选择和遗传网络规划的期货高频交易策略研究. *中国管理科学*, 2015, 23(10): 47–56.
Chen Y, Wang C. A study on high-frequency futures trading strategy based on variable selection and genetic network programming. *Chinese Journal of Management Science*, 2015, 23(10): 47–56. (in Chinese)
- [26] Nuij W, Milea V, Hogenboom F, et al. An automated framework for incorporating news into stock trading strategies. *IEEE Transactions on Knowledge & Data Engineering*, 2014, 26(4): 823–835.
- [27] Warner J B, Watts R L, Wruck K H. Stock prices and top management changes. *Journal of Financial Economics*, 1988, 20(1): 461–492.
- [28] Rosen R J. Merger momentum and investor sentiment: The stock market reaction to merger announcements. *The Journal of Business*, 2006, 79(2): 987–1017.
- [29] 张小涛, 熊 熊. 基于两类学习模型的多主体人工股票市场研究. *系统工程学报*, 2013, 28(6): 756–763.
Zhang X T, Xiong X. Artificial stock market of multi-agents based on two learning models. *Journal of Systems Engineering*, 2013, 28(6): 756–763. (in Chinese)
- [30] 贾凌云, 李冬妮, 田云娜. 基于混合蛙跳和遗传规划的跨单元调度方法. *自动化学报*, 2015, 41(5): 936–948.
Jia L Y, Li D N, Tian Y N. An intercell scheduling approach using shuffled frog leaping algorithm and genetic programming. *ACTA Automatica Sinica*, 2015, 41(5): 936–948. (in Chinese)
- [31] 黄安强, 李 梦, 杨丰梅. 基于改进遗传规划算法的非线性集成预测新方法. *系统科学与数学*, 2013, 33(11): 1332–1344.
Huang A Q, Li M, Yang F M. A new nonlinear ensemble forecasting method based on an improved genetic programming. *Journal of Systems Science and Mathematical*, 2013, 33(11): 1332–1344. (in Chinese)
- [32] Chen Z, Ji H. Language specific issue and feature exploration in Chinese event extraction. *Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics*, Boulder, Colorado, USA, 2009: 209–212.
- [33] 韩永峰, 许旭阳, 李弼程, 等. 基于事件抽取的网络新闻多文档自动摘要. *中文信息学报*, 2012, 26(1): 58–67.
Han Y F, Xu X Y, Li B C, et al. Web news multi-document summarization based on event extraction. *Journal of Chinese Information Processing*, 2012, 26(1): 58–67. (in Chinese)
- [34] 韩豫峰, 汪雄剑, 周国富, 等. 中国股票市场是否存在趋势. *金融研究*, 2014(3): 152–163.
Han Y F, Wang X J, Zhou G F, et al. Are there trends in Chinese stock market. *Journal of Financial Research*, 2014(3): 152–163. (in Chinese)

作者简介:

文丹艳(1988—), 女, 河北保定人, 博士, 讲师, 研究方向: 金融工程与风险管理, Email: wendy2018@njust.edu.cn;
马超群(1963—), 男, 湖南汨罗人, 博士, 教授, 研究方向: 金融工程与风险管理, Email: cqma1998@hnu.edu.cn;
曹建平(1987—), 男, 江苏东台人, 博士, 研究方向: 数据挖掘, Email: caojp_2005@126.com.