

# 大语言模型、文本情绪与金融市场\*

姜富伟 刘雨旻 孟令超

**摘要:**“人工智能+”行动是发展新质生产力的重要途径,其在金融领域的应用有助于金融强国建设。本文创新性地融合结构化金融市场数据和非结构化金融文本大数据,并结合中国特色金融市场的独特特征,训练了一个更适用于我国金融领域的中文金融大语言模型,并开展金融市场情绪测度和资产价格风险预测。研究发现,与传统字典法相比,使用中文金融大语言模型构建的大模型情绪在金融市场回报预测方面表现显著更佳。大模型情绪对很多宏观经济变量也有显著预测能力,能够捕捉非理性情绪冲击对宏观经济基本面的影响。大模型情绪在经济下行和极端风险事件期间的预测效果更强,契合了金融理论中非理性情绪对金融市场和宏观经济会产生非对称与非线性影响的结果。综上,本研究展现了“人工智能+”行动在我国金融领域应用落地的潜在技术路径和理论逻辑。

**关键词:**文本情绪 深度学习 大语言模型 资产定价

## 一、引言

人工智能是当前全球科技创新发展的最前沿,在我国经济与金融业数字化转型中扮演了重要角色。在此背景下,2024年国务院《政府工作报告》指出要“深化大数据、人工智能等研发应用,开展‘人工智能+’行动”,这是发展新质生产力的重要途径,党的二十届三中全会指出“要健全因地制宜发展新质生产力体制机制,健全促进实体经济和数字经济深度融合制度”。开展“人工智能+”行动,促进实体经济与数字经济相融合,为包括金融业在内的各行各业高质量发展提供了新方向。尤其是,以BERT、ChatGPT等为代表的人工智能大语言模型迅猛发展,体现出了广阔的应用前景。思考如何开展“人工智能+金融”垂直领域应用研究,对做好科技金融、数字金融等5篇大文章,发展新质生产力与推动金融强国建设都具有重要意义。为此,本文从金融市场非理性情绪测度的角度出发,结合中国特色金融市场的独特特征,构造了适用于金融领域的中文金融大语言模型,并探讨了“人工智能+”与金融应用相结合的路径和逻辑。

从历史上的1637年荷兰郁金香泡沫到近期的2020年新冠疫情爆发,一系列极端事件都表明金融市场受到投资者非理性情绪的严重影响(夏皮罗等,2022)。金融市场均衡与资产价格不仅取决于经济基本面等内在价值,还会受到投机情绪的影响,这在以个人投资者为主体的中国市场上表现尤为突出(陈彦斌,2005)。受非理性投机情绪驱动的噪音交易会使得金融资产价格偏离其内在价值,并导致金融泡沫的快速膨胀与惨烈破裂,影响金融稳定(游家兴、吴静,2012)。因此,情绪成为了金融学术和政策研究的重要话题。然而,情绪作为一种主观信念,难以使用传统的数据与方法加以准确测度,如何准确测度金融市场情绪因而成为了一个重要的基础性学术研究问题(贝克、沃格勒,2007)。

现有文献中的情绪测度主要分为基于市场交易指标构建和基于词典法构建两类。首先,部分市场交易指标与市场情绪变化紧密相关,可以从中提取情绪信息(贝克、沃格勒,2007)。该方法有数据可得性强的优势,但由于

收稿时间:2022-9-21;反馈外审意见时间:2022-12-19、2023-5-22;拟录用时间:2024-4-24。

\*本研究得到国家社科基金重大项目“三重压力下双支柱调控的政策效应评估与优化研究”(22&ZD063)、国家自然科学基金面上项目(72072193,71872195,72342019)、中央财经大学青年科研创新团队和科教融合研究生学术新星孵化计划(202303)的资助。感谢洪永森、周颖刚、王一鸣、张晓燕、杨之曙、吴卫星、尹志超、周国富、方颖、潘越、张学勇、李斌、姚加权、林建浩等专家学者的宝贵建议和意见。文责自负。孟令超为本文通讯作者。

市场交易指标还受其他因素影响,该方法难以准确反应情绪变动。其次,越来越多的文献尝试通过字典法从非结构化文本大数据中提取市场情绪信息,使用的文本数据包括媒体新闻(姜富伟等,2021b;汪昌云、武佳薇,2015;张琦等,2016;张成思、芦哲,2014;张晓燕、葛慧敏,2022)、上市公司报告(谢德仁、林乐,2015;曾庆生等,2018;胡楠等,2021)、社交媒体文本(张紫琼等,2010;姚加权等,2021)、货币政策报告(姜富伟等,2021a;林建浩等,2021)等。然而,理论表明,投资者情绪不仅能够对资产价格造成影响,还能通过资产价格影响实体经济波动,且彼此间不断复杂交互(贝克、沃格勒,2007;本哈比等,2016)。因此,投机情绪、金融市场与实体经济之间的关系存在复杂的非线性与非对称特征(夏皮罗等,2022)。现有文献采用的文本字典法是简单的线性方法,难以刻画非线性与非对称的市场情绪动态变化,并受字典设定质量影响较大,还忽略了上下文等其他维度的文本信息,因此亟待改进。

为解决上述问题,受到“人工智能+”行动启发,本文创新性地将人工智能大语言模型与中国特色金融市场独特特征相融合,训练了一个中文金融大语言模型,实现了对金融市场情绪的更准确测度。大语言模型是近年来自然语言处理与人工智能领域研究的最新突破,其本质上是一个规模极大的深度学习模型,通过数以亿计的参数全面训练并刻画文本特征,典型代表是谷歌公司推出的BERT模型和开放人工智能研究中心(OpenAI)的ChatGPT模型。由于大语言模型参数规模的庞大性,其通常也被简称为大模型。大模型在捕捉非线性关系中具有突出优势,与金融理论模型中情绪与资产价格之间的非线性与非对称关系十分契合。此外,我国资本市场具有鲜明的中国特色,起步晚、发展快,以个人投资者为主体,具有较强的投机性和不稳定性。为了更准确地测度我国金融市场的非理性投机情绪,本文创新性地引入了以BERT(Bidirectional Encoder Representations from Transformers)为代表的深度学习和大语言模型,将市场回报作为情绪标签变量引入到大模型训练中,创新融合结构化市场数据和非结构化文本数据,训练了一个更具有中国市场特色的中文金融大语言模型,构建能反映我国金融市场非线性与非对称特征的大模型情绪测度。

具体来说,本文利用2003~2021年间的五百多万条媒体新闻文本大数据,训练出本文提出的中文金融大语言模型,计算大模型情绪测度指标,开展金融市场资产价格风险预测实证检验。本文还使用姜富伟等(2021b)的金融文本情感字典计算字典法情绪,以作为对比。实证结果表明,基于中文金融大语言模型构建的大模型情绪指标在样本内与样本外预测检验中都表现出了远优于字典法情绪的预测能力,并且在资产配置中能够获得更好的收益。进一步分析发现,基于大模型构建的情绪指标还能准确预测宏观经济变量,且在极端事件期间对市场回报有着更强的预测能力。上述结果表明大模型情绪很好地契合了金融理论模型中非理性情绪对金融市场和宏观经济会产生非对称与非线性影响的理论预测。

本文主要的贡献在于:(1)本文训练了一个适用于我国金融市场情景的中文金融大语言模型,改进了文本大数据情绪测度在我国金融市场的应用效果。本文的中文金融大语言模型基于BERT模型构造,现有文献的探索多针对英文文本(黄等,2023;杨等,2020;刘等,2021)。少数在中文语境下应用BERT模型的文章关注点集中于提升情绪分类精确度(许雪晨、田侃,2021),缺乏对大模型情绪测度与我国金融市场特色的深入关联。本文有力补足了上述文献研究中的不足。(2)本文借助大语言模型的深度学习算法优势,识别我国金融市场复杂的非线性与非对称情绪特征,构造了更加符合金融理论的情绪变量。本文发现大模型情绪变量在极端事件期间可以更好地预测市场回报,并能够预测宏观经济状态,展现了大模型与金融理论的高度契合。(3)本文将柯等(2019)和周等(2024)的以金融市场回报为情绪标签的做法,与大语言模型的微调技术相结合,在充分利用大语言模型非线性优势的同时,进一步结合市场回报中蕴含的结构化情绪信息进行训练。总之,与现有文献使用词典法或人工标注训练大语言模型相比(黄等,2023;杨等,2020;刘等,2021;石等,2022),本文做法与经济金融理论模型更加契合,构造了一个与中国特色金融市场关联更加紧密的大模型情绪测度,为理解深度学习和人工智能大模型为何在金融应用上表现优越提供了新的经济学见解,探索了“人工智能+金融”行动的潜在技术路径和理论逻辑。

后文主要组成部分包括:第二部分是文献回顾,第三部分介绍了数据来源,第四部分是情绪测度,第五部分是实证分析,第六部分是进一步分析,第七部分是结论与政策建议。

## 二、文献回顾

### (一)情绪与金融市场理论

金融市场资产价格的变化与情绪的变动息息相关(席勒,2015)。情绪的本质是投资者的非理性预期,而当这一预期得到越来越多的投资者认可、传染和强化时,其对资产价格的影响就越强。德隆等(1990)提出的DSSW噪音交易者模型,论证了市场上噪音交易者受到非理性情绪影响,通过噪音交易对资产价格产生作用,而当情绪过度积极或者过度消极时,那么此类噪音交易需求会过于旺盛或者过度不足,从而进一步扭曲资产价格。帕洛米诺(1996)则进一步论证了在不完全竞争市场中,噪音交易者能够对资产价格具有更强的影响能力和解释效果。国内外有诸多实证文献也论证了情绪对资产价格具有显著预测效果(洛克伦、麦克唐纳,2011;黄等,2015;姜富伟等,2021b)。

情绪作为一种重要的非理性冲击,不光可以影响资产价格(贝克、沃格勒,2007),还能通过资产价格的信号渠道和经济周期中情绪与经济预期的自我实现机制(安吉利托斯、拉奥,2013;佩里、奎德里尼,2018)等多种渠道对宏观经济波动和经济基本面产生影响。投机情绪、金融市场与宏观经济三者之间的交互使得投资者情绪对资产价格的影响可以是非线性和非对称性的(本哈比等,2016)。

情绪对资产价格具有非线性影响。非理性投资者的情绪化预期具有时变性,从而导致资产价格扭曲和非理性波动(德隆等,1990)。对于中国股票市场而言,受到媒体影响产生情绪交易的庞大个人投资者群体所形成的非理性交易会进一步削弱理性投资者和基本面定价因素的市场定价能力,从而导致资产价格偏离宏观基本面价值。而当非理性交易者的噪音交易行为获得投资收益,理性投资者也会形成“噪音追逐”,导致资产价格进一步偏离基本面价值(门德尔、施莱弗,2012)。因此,非理性投资者和理性投资者所产生的情绪化交易耦合会进一步强化金融市场资产价格的非线性变化。

情绪对资产价格具有非对称影响。有诸多文献表明,情绪可以通过资产价格的信号渠道和经济周期中的预期自我实现机制,对宏观基本面理性变量产生影响(佩里、奎德里尼,2018)。比如,积极的情绪冲击表明繁荣的经济预期,资本需求将增加,从而推高资产价格,而资产价格的上升会向实体经济传递一个强劲的信号,激发实际产出和就业的增长(夏皮罗等,2022;本哈比等,2016)。因此,来自非理性基本面的情绪冲击,可以对经济基本面的宏观经济变量产生较强的影响,具有一定的解释能力。并且由于情绪与资产价格间存在“沉默的螺旋”,情绪在不同经济周期和金融市场状态下表现出更强的非对称影响,从而对资产价格具有更强的非对称影响(游家兴、吴静,2012)。特别是在经济下行阶段与极端事件期间(例如股灾、新冠疫情等),负面情绪能够对金融市场价格和波动具有更强烈的非对称影响能力(伯南克,2013)。由于上述信号渠道、预期自我实现机制等,基本面理性变量的微小变化可能会在情绪因素的放大下,导致资产价格极大偏离基本面价值,带来资产价格暴跌,甚至崩盘(本哈比等,2016)。

因此,从情绪与金融市场理论来看,情绪会对资产价格产生显著影响,总结如下:首先,情绪会对资产价格产生显著影响,从而可以使用情绪进行收益预测。其次,情绪可以通过信号渠道、预期自我实现等机制,对实体经济生产和投资等宏观经济变量产生影响。最后,情绪对资产价格影响具有明显的非对称与非线性特点。当经济处于下行期和极端事件发生时,负面情绪冲击能够对资产价格产生更强影响,且在价格的各种因素中占据主导。以BERT模型为代表的深度学习模型和大语言模型具有的非线性特征能够契合理论分析中情绪与资产回报之间的非对称关系。

### (二)文本情绪测度

在金融领域,以词典法为代表的文本情绪分析应用日趋丰富,利用文本分析方法提取情绪信息从而更好地研究其对资产价格的影响已经成为热点(泰特洛克,2007;唐国豪等,2016),因此词典法具有极强的应用价值(李,2010)。代表性英文词典是洛克伦和麦克唐纳(2011)提出的LM字典,这一字典在股票收益预测方面获得了广泛的应用。

由于中文金融语境的特殊性,近年来也有不少学者针对构建中文金融情感词典展开研究。例如,姜富伟



等(2021b)在LM字典的基础上,结合word2vec模型,构建了适用于金融新闻文本分析的中文金融情绪字典,在金融领域应用中取得了良好效果。此外,姚加权等(2021)基于词典重组和深度学习的新颖方法构建了一款优秀的分别适用于年报的正式用语情绪词典和适用于社交媒体的非正式用语情绪词典;游等(2018)和杜等(2022)也采取类似的方法构建了中文金融情绪词典;范小云等(2022b)基于新闻报道和社交媒体,构建了中文金融文本情绪词典;汪昌云和武佳薇(2015)基于新闻报道构建了中国财经媒体正负面词库;卞等(2021)基于年报文本等构建了中文金融情绪词典。上述文献都使用字典法度量了金融文本情绪。

尽管字典法具有直观易行的优点,但也存在诸多不足。首先,字典法迁移性较差,针对不同问题需要构建不同的词典,例如评级报告词典(潘怡麟等,2023)、央行沟通用语词典(林建浩等,2019;姜富伟等,2021a)、金融稳定报告词典(姜富伟等,2022;科雷亚等,2021)等。其次,情感词典需不断更新,特别是对于突发事件(如新冠疫情)和社会用语变化等,否则将影响分析质量。此外,字典法中字典词汇选取容易受到主观判断影响,人的有限认知也容易导致在构建词典时遗漏相关词汇(巴苏等,2022)。最后,字典法假设词语与情绪之间服从线性关系,忽略掉了文本中的其他维度信息,这使得字典法难以捕捉理论文献中情绪与资产回报间的非线性、非对称的特征(德隆等,1990;本哈比等,2016)。

深度学习方法的引入有助于解决上述问题。由于具备非线性等独特的算法优势,深度学习在收益预测(李等,2019;弗兰克尔等,2022;范小云等,2022b;马甜等,2022)、风险识别(范小云等,2022a)、量化交易(李斌等,2019)等方面具有显著预测效果。同样,许多学者尝试借助深度学习方法改善文本分析效果。首先,深度学习算法可以引入语序等多维文本特征信息,提高文本分析应用精度。例如,姜富伟等(2021b)借助word2vec神经网络模型引入文本语序信息来改善词典构建质量。中文比英文语法更复杂,所以更有必要借助深度学习模型来进行文本情绪分析(洪永淼、汪寿阳,2021)。其次,深度学习包括有监督学习、无监督学习和强化学习,而其中有监督类的深度学习模型还有助于借助金融市场信息来改进金融领域文本分析效果。例如柯等(2019)和周等(2024)将市场回报作为情绪标签,并融入文本分析模型中,从而获得了对市场回报预测表现更好的情绪变量。深度学习中的有监督算法和非线性的算法特征不仅有助于改善文本分析效果,也有助于构建更加符合理论特征的文本变量(苏治等,2017)。

近年来,文本分析领域的深度学习与大语言模型层出不穷,其中最为突出的便是BERT模型。BERT模型是一种由谷歌(德夫林等,2018)提出的大型深度学习语言模型,在数以百万计的大规模语料上训练而来,并具有上亿的参数数量。因此,BERT模型可以更好地捕捉文本中的语义和句法特征。BERT模型的训练分为预训练与微调两个步骤,在得到预训练模型后,仅需针对下游任务进行微调,便可将BERT模型应用于不同的文本分析任务中。这一特性使得BERT模型在多种文本分析应用上都取得了优异表现(阿迪卡里等,2019;刘,2019)。因此,本文选择基于BERT模型训练中文金融大语言模型。

许多学者已将BERT模型应用至金融领域。黄等(2023)借助美国市场数据,发现BERT模型在多种金融文本上的情绪分类精度显著优于字典法与一些传统的机器学习模型。黄等(2023)的研究表明除去情绪分类外,BERT模型也有助于捕捉ESG等其他公司和金融市场上的重要信息,这体现出了BERT模型的丰富应用前景。同样,杨等(2020)和刘等(2021)等文献也发现BERT模型在金融文本情绪分析上相对于其他方法具有更好的表现。上述文献均针对英文金融文本进行研究,石等(2022)是少数检验了BERT模型在中文金融文本上分析效果的文章之一,发现BERT模型对个股股吧社交媒体投资者情绪的分类精度更高。上述文章往往只从情绪分类精度和回报预测精度简单对比BERT模型与其他方法,对BERT模型情绪变量与经济动态间的深层次关系以及BERT模型相对于其他方法优势的深入分析仍显不足,本文研究对这一文献空白进行了补足。

### 三、数据来源

#### (一)数据使用

本文主要使用3个新闻数据集。前两个数据集分别为证券关联新闻数据集和新闻基本信息数据集,时期为2003年1月1日~2021年12月31日,来源于国泰安数据库(CSMAR)。该数据库收集了1993年以来,我国宏

观经济、股票、债券、基金等综合金融新闻报道和各项主要信息披露、政策文件,包括基本的证券行业新闻和与之关联的报道评论文章等。由于数据集中2003年以前的数据缺失较多,本文选取2003年1月1日以后新闻基本信息库、新闻证券关联信息库的新闻数据作为研究样本。第三个数据集为2019~2020年的微博财经新闻,经过清洗后共30余万条有效数据。由于微博没有统一的财经新闻入口,本文首先人工列举出了“21世纪经济报道”、“东方财富网”、“央视财经”等21家具有微博账户的权威新闻媒体列表,然后爬取这些账户的微博内容作为微博财经新闻数据集。尽管该数据来自微博,但由于爬取账户均为权威新闻媒体,获取的财经新闻仍然是标准的正式新闻。

新闻基本信息数据集内容丰富,基本涵盖了市场上主流媒体的新闻报道,且样本连续不间断。基于这些特点,本文使用新闻基本信息数据集中的新闻语料计算新闻情绪,并合成情绪时间序列指数。微博财经新闻数据集与新闻基本信息数据集重合度较低。为此,本文使用微博财经新闻数据集作为中文金融大语言模型针对金融领域进行特质化预训练的语料,这可以避免预训练中出现过拟合问题。证券关联新闻数据集则是新闻基本信息数据集的子集,我们可以通过证券关联新闻数据集中的“新闻ID”字段获取新闻基本信息数据集中对应的新闻内容,并通过“证券代码”字段获取关联上市公司代码。这方便我们借助市场回报生成对应的情绪标签,将证券关联新闻数据集中每一个新闻数据与对应情绪标签应用于中文金融大语言模型微调阶段。

## (二)数据清洗

为便于后续分析,本文对3个文本数据集进行如下预处理与文本清洗。首先,去除网址链接、乱码等各类非中文字符。其次,为了减少噪音新闻的影响,本文删除了标题或内容短于5个字符的短新闻数据,因为通常来说过短的文本几乎无法表达有效信息。随后,删除重复新闻字段、无效新闻(如考试通知、新年贺词等)和回顾性历史新闻(如一周新闻回顾等)等无效样本。同时,本文从新闻基本信息集中剔除了证券关联新闻数据集、微博财经新闻数据集中共同包含的样本,这可以避免一个样本同时出现在训练集与预测集中,缓解过拟合问题。

经过上述预处理过程后,微博财经新闻数据集共计30万余条样本,新闻证券关联数据集共计100万余条样本,新闻基本信息数据集共550万余条样本。图1展示了样本期内新闻基本信息集各年新闻数据数量,2003年及以后每年数据最少10万余条以上,最多超过80万余条,考虑样本规模和质量,本文最后选择使用2003~2021年数据,进行后文数据分析。

## 四、情绪测度

由前文文献综述可知,BERT模型是一种典型的深度学习和大语言模型,在各种文本分析任务上取得了良好的表现。为此,本文选择基于BERT模型构造中文金融大语言模型。接下来,本文针对BERT模型的基本原理、基于BERT的中文金融大语言模型训练以及情绪测度进行逐一介绍。

### (一)BERT模型

本文应用了专业人工智能实验室开源的BERT基本模型框架(pytorch版本<sup>①</sup>),具体算法可以参照官方标准使用手册。参考德夫林等(2018),BERT模型框架如下所示:

#### 1. 预训练和微调

BERT模型的训练与应用包含预训练与微调两个步骤。在预训练步骤中,BERT模型在大型的无标注文本数据集中通过特定的任务进行无监督的预训练,使模型学习文本中的语义和句法特征,获得可进一步用于特定下游任务的预训练模型。在微调步骤中,BERT模型则针对具体的下游任务(如文本分类等)使用一些有标注的文本数据集进行有监督的训练。BERT预训练模型

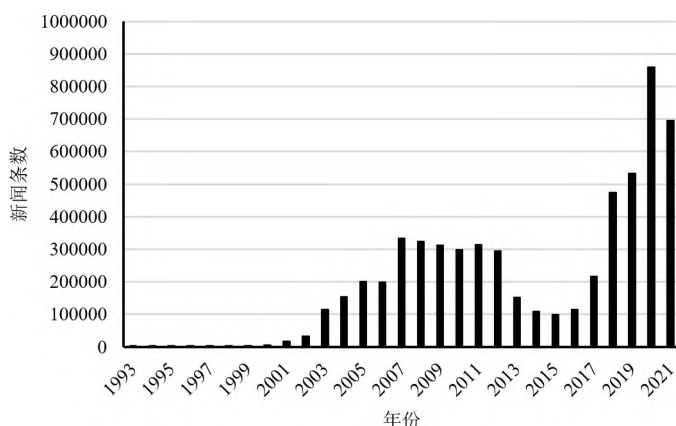


图1 各年份新闻数据总量

参数在微调步骤中会进行细微优化,从而使模型针对具体下游任务能够获得更优的表现。图2所展示的为下游任务是文本分类任务的BERT模型框架,左侧为BERT模型的预训练阶段,右侧为针对文本分类任务的微调阶段。

## 2. BERT模型预训练任务

在预训练阶段中,BERT模型包含遮罩语言模型(Masked Language Model,简称MLM测试)和下文预测任务(Next Sentence Prediction,简称NSP测试)两个预训练任务。MLM测试类似于填空题,通过覆盖某个字词,让模型通过上下文理解字词关系,填充空白词汇。NSP测试类似于成语接龙,让模型通过学习句子、段落之间的关系判断两个句子的先后顺序。上述两个预训练任务使得BERT模型可以捕捉文本的上下文关系与动态特征信息,从而更好地捕捉语义信息和句法特征,这是相对于字典法模型的显著优势。

### (二)中文金融大语言模型

#### 1. 整体方法

本文按照图3的整体思路训练中文金融大语言模型,并应用于金融领域的文本分析中。本文所使用的BERT基础模型由专业人工智能实验室训练而来,这是一个适用于通用领域文本的BERT预训练模型。为了提高该模型在金融领域文本上的表现,本文在此基础上借助微博财经新闻集,通过“进一步预训练”方法得到了更加适用于金融领域的中文金融预训练模型。随后,本文借助证券关联新闻集,并将市场回报作为情绪标签,从而在微调阶段将中文金融大语言模型针对金融市场上的情绪分类任务进行优化。最后,本文利用训练完成的中文金融大语言模型,输出对微博财经新闻集和证券关联新闻集取差集后的新闻基本信息集中新闻的情绪值,并在此基础上合成大模型情绪变量,进一步检验其与金融市场关联。

#### 2. 中文金融大语言模型的预训练

由于金融领域文本在词语选用与行文风格上与一般的通用领域文本存在较大差异,训练一个金融领域的中文金融大语言模型十分必要。将一个通用领域的BERT模型迁移至特定领域有两种主流方法。方法一是选取高质量的大型特定领域文本数据集,通过标准的BERT模型预训练任务从零构建一个特定领域的预训练模型。该方法与德夫林等(2018)中原始BERT模型的训练过程一致,但对算力要求极高,训练耗时极长,除去谷歌、百度等领先的人工智能实验室外难以实现。

方法二则是在现有的通用领域预训练模型基础上,通过“进一步预训练”(Further pre-training)引入特定领域的文本数据重新进行预训练任务,以达到增强模型在特定领域文本分析中应用效果的目标(古兰甘等,2020;李等,2020;阿拉奇,2019;韩、艾森斯坦,2019;洛格斯瓦兰等,2019;孙等,2019;贝尔塔吉等,2019;霍华德、鲁德,2018)。进一步预训练中的训练任务与基本的预训练相同,均为MLM测试与NSP

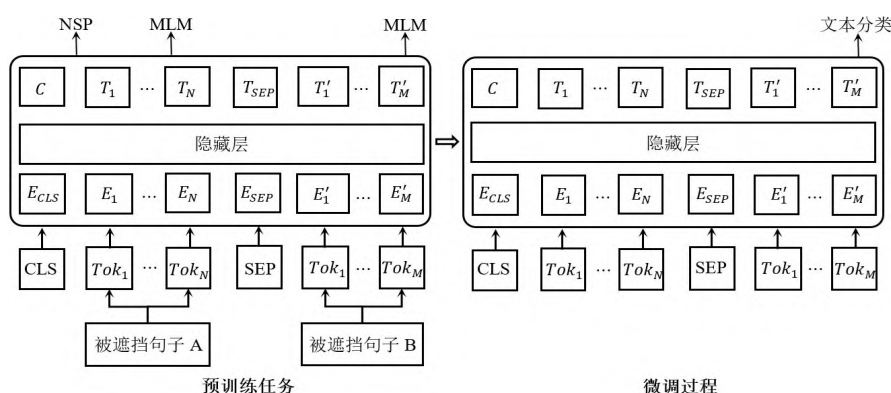


图2 BERT模型框架示意图

注:CLS是用于分类任务的标记符号;SEP是分句标记符号,用于断开语料;Tok是Token的缩写,代表BERT模型识别的文本分割单元; $Tok_i$ 是第*i*个文本分割单元; $E_{CLS}$ 、 $E_{CLS}$ 、 $E_{SEP}$ 、 $E_i$ 分别表示对应的词嵌入向量, $C$ 、 $T_{SEP}$ 、 $T_i$ 分别表示对应经过BERT模型神经网络计算得到的特征向量;箭头示意运算方向。

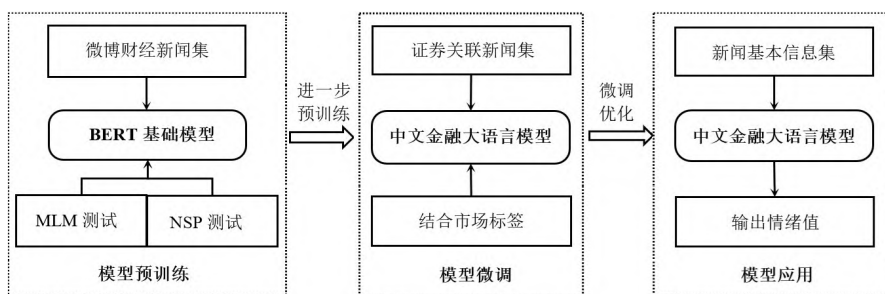


图3 中文金融大语言模型总体应用流程图



测试。进一步预训练方法在特定领域文本分析效果与计算可行性之间实现了较好的均衡,并得到了广泛的应用(孙等,2019;贝尔塔吉等,2019;霍华德、鲁德,2018)。

考虑到计算的可行性,本文选择进一步预训练方法来构建一个适用于金融文本的预训练模型,从而满足本文应用需求<sup>②</sup>。为避免引入样本外数据造成过拟合问题与影响后续应用,本文使用微博财经新闻数据集进行进一步预训练。

### 3. 中文金融大语言模型的微调

在得到适用于金融领域的中文金融预训练模型后,本文进一步针对情绪分类任务对该模型进行微调与优化。BERT模型的微调步骤依赖于有标签的文本数据,本文借助市场回报来生成情绪标签。本文的核心目标之一是构建一个更加符合理论特征的情绪变量。在金融理论中,情绪变量与资产价格之间紧密相关,情绪通常表现为对资产价格的非理性预期(夏皮罗等,2022;本哈比等,2016)。同时,观测到的资产价格与回报是在多种机制下受到情绪和其他因素影响的均衡结果(席勒,2015;贝克、沃格勒,2007),包含了与情绪相关的信息。因此,使用市场回报作为情绪标签有助于捕捉蕴含于价格回报中的情绪信息<sup>③</sup>。柯等(2019)、周等(2024)、弗兰克尔等(2022)、杰加迪什和吴(2013)都尝试将收益率引入文本分析模型,并在预测资产回报和解释市场异象上都得到了良好表现。

使用市场回报生成情绪标签的另一优势是可行性高。获取情绪标签的常见做法是进行人工情绪判定。尽管该方法直观明了,但在可行性与准确度上也存在诸多问题。首先,人工情绪判定工作量大、耗时长、成本高。目前主流的带有人工情绪标签的文本数据集多为通用领域文本,缺乏相应的高质量金融领域标注文本数据。其次,人工情绪判定也会受到主观偏差与情绪分歧偏差等影响(西加诺斯等,2017),不同个体对同一文本的情绪判定可能存在差异。因此人工情绪判定可能难以反映整体市场的情绪变化。使用市场回报生成情绪标签可以较好地处理上述问题。

基于上述考量,本文通过市场回报生成情绪标签,并应用至中文金融大语言模型的微调中。具体方法如下:使用数据预处理后的证券关联数据集,利用新闻关联证券的回报作为当日新闻情绪的市场标签测度。本文将新闻匹配对应证券的回报涨跌幅,若关联证券回报为正,市场情绪标签记为1,反之记为0。参考周等(2024)和柯等(2019)的设定,从市场反应角度,公司股票价格只有上涨和未上涨两种结果。因此,选择0-1标注能够完整有效地匹配市场情绪反应含义。

通过上述方法生成证券关联新闻数据集中每篇新闻的市场情绪标签后,本文在此基础上进行中文金融大语言模型的微调。在微调过程中,本文运用十折交叉验证方法划分训练集、验证集,实现模型参数的微调优化。通过上述步骤,本文得到了适用于金融领域情绪分析任务的中文金融大语言模型。

### 4. 中文金融大语言模型的应用

在该阶段,本文利用前文所得到的中文金融大语言模型来输出新闻基本信息集中每篇新闻的情绪值,具体步骤如下。首先,本文使用剔除证券关联新闻、微博财经新闻后的新闻基本信息集作为新闻语料,这样可以最大限度的保留新闻基本信息集的样本外特征,避免过度拟合问题;然后,本文将新闻基本信息集数据代入微调后的中文金融大语言模型,输出每篇新闻的情绪值,并用于后文计算大模型情绪变量。

#### (三)情绪变量合成

本文基于前文数据清洗处理后的新闻基本信息数据集的新闻内容文本,分别应用前文得到的中文金融大语言模型输出的情绪值来计算第 $t$ 交易日大模型情绪变量(记作 $S_t^{llm}$ ),以及应用姜富伟等(2021b)提出的金融情绪字典合成金融字典法情绪变量(记作 $S_t^{dic}$ )。具体情绪变量计算公式如下:

$$S_t^k = \frac{\sum_{i=1}^{N_t} A_i^k}{N_t}, k = llm \text{ or } dic$$
$$A_i^{llm} = \begin{cases} 0, negative \\ 1, positive \end{cases}; A_i^{dic} = \frac{\sum_{j=1}^{n_i} word_{ij}}{n_i}, word_{ij} = \begin{cases} -1, negative \\ 1, positive \end{cases} \quad (1)$$

其中: $N_t$ 表示在第 $t-1$ 交易日末至第 $t$ 交易日末所包含新闻的条数<sup>④</sup>;  $A_i^{llm}$ 为该日第 $i$ 篇新闻对应的中文金

融大语言模型输出的情绪值,取值为0表示市场情绪消极,取值为1表示市场情绪积极; $A_{it}^{dic}$ 表示对应时期内第*i*篇新闻通过金融字典法计算得到的情绪值,取值为-1~1之间的常数,越接近-1则表明情绪越消极,越接近1则表明情绪越积极; $n_i$ 表示第*i*篇新闻含有的情感词汇总量,

$word_{ij}$ 表示第*i*篇新闻中第*j*个情感词汇的情绪数值,取值为-1表示是消极情感词汇,取值为1表示是积极情感词汇。

在得到日度情绪变量后,本文通过月内取均值并做标准化处理的方法得到了月度大模型情绪变量与月度金融字典法情绪变量。为了进一步直观表现数据效果,本文将两个情绪变量序列同标准化后上证指数收益率的时序变化绘制于图4(具体变量定义与描述性统计信息详细内容参见《管理世界》网络发行版附录1和附录2)。

图4报告了两种情绪变量与上证指数收益率的时序变化情况。大模型情绪变量为正时表示市场情绪积极,为负时表示市场情绪消极。总体来看,上述两类情绪变量走势与市场回报走势密切相关。比如在2005年股权分置改革、2007年股灾、2015年股灾、2020年新冠疫情等时期,上述两类情绪变量都有效捕捉了市场变化趋势。这表明通过两种方法构建的情绪变量都较为有效地捕捉了市场情绪变化动态。

## 五、实证分析

本文选择上证指数收益率作为市场指数收益率的代理变量,并对大模型情绪变量与金融字典法情绪变量对市场回报的预测能力进行实证检验。相关金融市场数据与宏观经济数据选自CSMAR数据库和WIND数据库,并经过标准化处理,样本时期为2003年1月1日~2021年12月31日。

### (一)非参数检验

本文首先参照贝克和沃格勒(2007)、姜富伟等(2021b)中的方法对各情绪指标与市场回报间关系进行了非参数检验。具体来说,本文按照各期情绪指标是否大于全样本平均值与中位数将样本划分为情绪高涨时期与情绪低落时期,然后分别计算下月等权上证指数收益率的均值。结果如图5所示。

根据图5不难看出,在情绪高涨期间,上证指数的下一月平均收益率都显著的比情绪低落期间的收益率高,这表明了情绪对市场存在着显著影响,与姜富伟等(2021b)中的发现一致。同时,从依据不同指标分组的表现情况来看,使用大模型情绪变量划分情绪时期时,不同情绪期间的回报差异要强于使用金融字典法进行样本划分的结果。这一发现初步表明使用中文金融大语言模型方法可以更好地捕捉情绪与市场回报间的关系,后文将对这一点进行更加深入的分析。

### (二)样本内预测能力检验

参考姜等(2019)和姜富伟等(2021b),本文进行如下样本内预测能力检验。

$$r_t^m = \alpha + \beta \text{SENTIMENT}_{i,t-1} + \varepsilon_{i,t}, i=1,2 \quad (2)$$

其中: $r_t^m$ 分别为第*t*期等权与市值加权A股上证指数收益率, $\text{SENTIMENT}_{i,t-1}$ 是第*t*-1期第*i*个情绪指数, $i=1,2$ ,分别

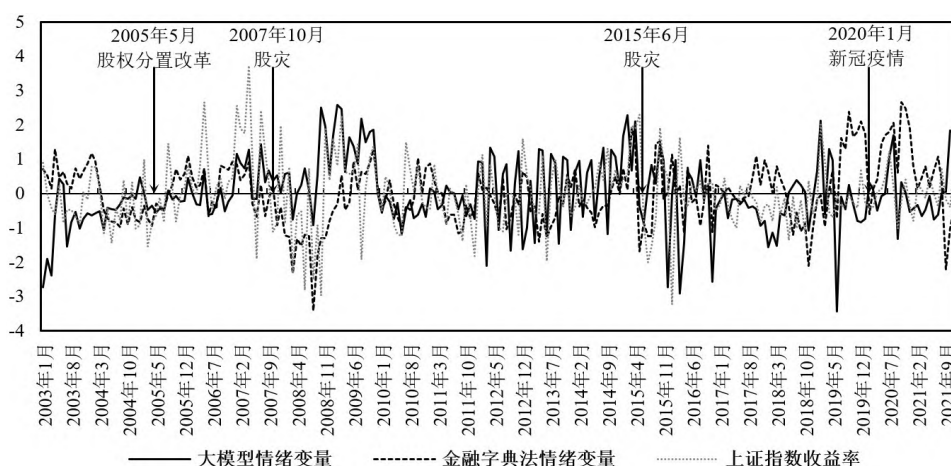


图4 情绪变量与上证指数收益率时序变化

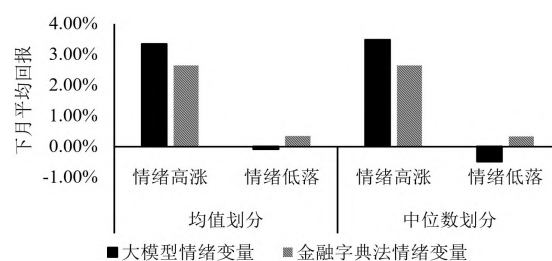


图5 情绪变量非参数检验



对应大模型情绪变量( $S_{lm}$ )、金融字典法情绪变量( $S_{dic}$ )。

为分析两类情绪变量在不同经济状况下的表现,参考文献拉帕奇等(2010)、黄等(2015)的做法,本文以宏观经济景气一致指数(记为 $Macro$ )代表宏观状况,并根据当期指数是否大于样本均值划分宏观基本面不同区间,并计算如下分区间 $R^2$ 。

$$R_c^2 = 1 - \frac{\sum_{t=1}^T I_t^c (\hat{\varepsilon}_{i,t})^2}{\sum_{t=1}^T I_t^c (R_t^m - \bar{R}^m)^2}, c = up, down \quad (3)$$

其中, $I_t^{up}(I_t^{down})$ 是一个示性函数,当第 $t$ 期是宏观经济基本面上行区间(下行区间)时为1否则为0, $R_{up}^2$ 和 $R_{down}^2$ 表示在宏观经济基本面上行和下行时期对应的子区间 $R_c^2$ , $\hat{\varepsilon}_{i,t}$ 表示根据式(2)样本内估计得到的残差项, $\bar{R}^m$ 是样本整体区间 $R_t^m$ 的均值, $T$ 是全样本区间的时期数量。与全样本的 $R^2$ 不同,分区间的 $R_c^2$ 可能取负值。

表1结果显示,大模型情绪变量与金融字典法情绪变量都能显著地正向预测市场回报,这再次强调了情绪是市场回报预测中不可忽略的重要决定因素。根据坎贝尔和汤普森(2008),如果月度预测性回归 $R^2$ 大于0.5%,则表明这一变量的预测效力在经济学意义上是显著的,两种情绪指标的预测 $R^2$ 均远超这一标准。从预测效果对比来看,大模型情绪变量对等权与市值加权市场回报的预测回归 $R^2$ 均明显高于金融字典法情绪变量。尤其在等权市场回报预测中,大模型情绪变量的预测回归 $R^2$ 是金融字典法情绪变量的两倍多。上述结果强有力地表明大模型情绪变量在预测市场回报上的突出表现与优势。

从不同经济状况下的预测表现来看,大模型情绪变量在经济上行与经济下行区间上都能获得不错的预测效果,且在经济下行区间的预测效力尤其强,而金融字典法情绪变量对市场回报的预测效力则主要来自于经济上行区间。从前文文献梳理来看,情绪在不同经济状况下都能显著影响市场回报,且在经济下行状态下对市场回报影响更强。表1中的实证结果表明使用中文金融大语言模型度量的文本情绪更加符合理论中情绪对市场回报存在非线性与非动态影响的特征。

部分市场交易数据也能反应市场情绪变化,本文还进一步对比了媒体新闻情绪与基于市场交易价格指标的情绪测度。实证结果表明,媒体新闻情绪在控制了基于市场交易价格指标的情绪测度后仍然有增量信息(限于篇幅,详细内容请参见《管理世界》网络发行版附录3)。

### (三)样本外预测能力检验

接下来,本文将检验两类情绪指标的样本外预测能力。参考韦尔奇和戈亚尔(2008)、黄等(2015)、姜富伟等(2021b)、姜富伟等(2011)等文献的常见做法,样本外预测值计算方法如下:

$$\hat{r}_t = \hat{\alpha}_{i,t-1} + \hat{\beta}_{i,t-1} SENTIMENT_{i,t-1}, i = 1, 2 \quad (4)$$

其中, $i=1,2$ 表示两个情绪变量, $\hat{\alpha}_{i,t-1}$ 和 $\hat{\beta}_{i,t-1}$ 是逐期滚动估计得到的方程(2)中各系数的最小二乘估计结果。假设样本期总长度为 $T$ 期,初始训练时期样本长度为 $P$ ,则样本外期间长度为 $T-P$ ,包含 $P+1, P+2, \dots, T$ 等时期样本。初始训练期数据用于估计第一个回归方程,并得到第一个样本外预测值。随时间推移,预测性回归方程中不断增添新样本,并进行重新估计,以此得到下一期的样本外预测值。

为了增强样本外预测检验的稳健性,结合我国股票市场的实际发展历程,本文根据样本是否包含股权分置改革、2008年国际金融危机、2015年股灾等重要股票市场事件,划分如下3种区间。第一种样本区间划分(区间一)下初始训练期为2003年1月~2005年12月,样本外预测期为2006年1月~2021年12月。第二种样本区间划分(区间二)下初始训练期为2003年1月~2009年12月,样本外预测期为2010年1月~2021年12月。第三种样本区间划分(区间三)下初始训练期为2003年1月~2015年12月,样本外预测期为2016年1月~2021年12月。

本文运用坎贝尔和汤普森(2008)提出的样本外 $R_{os}^2$ 统计量对样本外预测能力进行检验,具体的表达式是:

表1 月度情绪与整体股票市场关系

	(1)	(2)	(3)	(4)
	等权重	等权重	市值加权	市值加权
$S_{lm}$	0.330*** (5.26)		0.254*** (3.95)	
$S_{dic}$		0.254*** (3.28)		0.235*** (3.61)
Constant	-0.004 (-0.07)	-0.002 (-0.04)	-0.005 (-0.08)	-0.003 (-0.05)
$R^2$ (%)	10.95	4.55	6.49	5.47
$R_{up}^2$ (%)	5.88	6.22	3.55	8.25
$R_{down}^2$ (%)	17.19	2.51	9.99	2.18

注:括号中为 $t$ 值。\*\*\*、\*\*、\*分别表示在1%、5%、10%的水平上显著,下同。

$$R_{i,os}^2 = 1 - \frac{\sum_{t=p+1}^T (r_t^m - \hat{r}_t^m)^2}{\sum_{t=p+1}^T (r_t^m - \bar{r}_t^m)^2}, i=1,2 \quad (5)$$

其中,  $\hat{r}_t^m$  表示在上证指数预测方程下第  $i$  个情绪变量的样本外预测值,  $\bar{r}_t^m$  是历史均值预测基准, 取值为市场指数在第 1 期至  $t-1$  期的历史收益均值。

表 2 报告了不同样本外区间的预测结果。只要样本外  $R_{os}^2$  统计量大于 0, 当前预测指标就可以取得超过历史均值基准的样本外预测能力, 并在实际投资中具有应用价值。表 2 结果显示, 大模型情绪变量和金融字典法情绪变量均具有大于 0 的样本外  $R_{os}^2$ , 能够取得良好的样本外预测效果, 这一结果同姜富伟等 (2021b) 实证结果相印证。同时, 中文金融大语言模型的  $R_{os}^2$  远超金融字典法, 说明大模型情绪变量相对于金融字典法情绪变量而言具有更强的样本外预测精度。

#### (四) 资产配置检验

本文接下来进一步借助资产配置检验考察并对比两类情绪指标的实际应用价值。参照坎贝尔和汤普森 (2008)、德米格尔等 (2009) 以及拉帕奇和周 (2021), 本文检验了均值方差投资者分别利用大模型情绪变量和金融字典法情绪变量在市场组合与无风险资产间进行资产配置时所能获得的收益。根据均值方差理论, 投资者投资于市场组合的最优权重为

$$\omega_{t-1} = \frac{1}{\gamma} \frac{\hat{R}_t^m}{\hat{\sigma}_t^2} \quad (6)$$

其中:  $\gamma$  是投资者风险厌恶系数, 本文设置为 3、5、7<sup>⑤</sup>;  $\hat{\sigma}_t^2$  为样本外方差估计值, 本文假设为滚动 75 个月的历史波动率;  $\omega$  是资产的权重约束, 本文将其取值范围约束于 0 到 1.5 之间, 即允许最大为 50% 的杠杆并禁止卖空;  $\hat{R}_t^m$  是第  $t$  期市场指数预测方程下样本外期间的市场预测超额收益。因此, 投资者在第  $t$  期的组合收益率为:

$$R_t^{portfolio} = \omega_{t-1} R_t^m + R_t^f \quad (7)$$

其中:  $R_t^f$  为无风险利率, 其他符号定义与上文相同。因此, 在每一期投资者可以获得数值为  $R_t^{portfolio}$  的实现收益, 其确定性等价收益为:

$$CER_p = \hat{\mu}_p - \frac{1}{2} \gamma \hat{\sigma}_p^2 \quad (8)$$

同时, 为了进一步计算使用情绪预测进行资产配置带给投资者的确定性等价收益增值, 还需要计算基准资产配置的确定性等价收益。与前文一致, 本文将历史均值预测作为基准资产配置方式, 并计算其确定性等价收益。使用情绪指标进行配置的确定性等价收益与历史均值配置的确定性等价收益之差即为情绪指标在资产配置中带给投资者的确定性等价收益增值 (CER gain)。本文还计算了使用情绪变量进行资产配置时投资组合的夏普比率 (Sharpe Ratio)。为了使检验更加符合现实情形, 本文同时汇报了无交易费用 (OBP) 与存在 50BP 交易费用时的结果。

表 3 报告了资产配置检验结果。大模型情绪变量在各种设定下均具有为正的确定性等价收益增值, 这表明与历史均值基准相比, 基于大模型情绪变量预测回报并进行组合配置总能提高投资者的效用。金融字典法情绪变量尽管也能取得不错的资产配置表现, 但在不同时间区间和交易成本下, 其确定性等价收益增值和夏普比率均低于大模型情绪变量。从实际投资的角度来看, 基于中文金融大

表 2 样本外预测结果

	区间一 (%)	区间二 (%)	区间三 (%)
$S_{lm}$	9.97	9.42	16.72
$S_{dc}$	3.85	5.20	0.66

表 3 资产配置检验结果

情绪变量	无交易费用		交易费用 (50BP)	
	CER gain (%)	Sharpe Ratio	CER gain (%)	Sharpe Ratio
Panel A: $\gamma=3$				
区间一				
$S_{lm}$	0.8834	0.2919	0.5614	0.2517
$S_{dc}$	0.6410	0.2633	0.4178	0.2323
区间二				
$S_{lm}$	0.7973	0.2634	0.4644	0.2190
$S_{dc}$	0.5582	0.2309	0.3375	0.1983
区间三				
$S_{lm}$	0.7153	0.2069	0.4341	0.1586
$S_{dc}$	0.0390	0.1003	-0.1761	0.0652
Panel B: $\gamma=5$				
区间一				
$S_{lm}$	0.5770	0.3050	0.3378	0.2686
$S_{dc}$	0.3578	0.2557	0.1485	0.2226
区间二				
$S_{lm}$	0.4803	0.2702	0.2421	0.2301
$S_{dc}$	0.2852	0.2250	0.0775	0.1913
区间三				
$S_{lm}$	0.3931	0.2017	0.1826	0.1573
$S_{dc}$	-0.1475	0.1004	-0.3848	0.0551
Panel C: $\gamma=7$				
区间一				
$S_{lm}$	0.3092	0.3005	0.1281	0.2660
$S_{dc}$	0.1273	0.2436	-0.0532	0.2129
区间二				
$S_{lm}$	0.2128	0.2606	0.0382	0.2242
$S_{dc}$	0.0654	0.2174	-0.1155	0.1865
区间三				
$S_{lm}$	0.2020	0.1897	0.0352	0.1457
$S_{dc}$	-0.2643	0.0981	-0.4766	0.0524

语言模型构建的情绪变量也能带给投资者更多的增量效用。

### (五)稳健性分析

本文考虑以下三方面的稳健性检验:第一,替换噪音新闻剔除标准;第二,替换样本区间;第三,考虑新闻来源。上述稳健性结果与基准结果均保持一致(限于篇幅,详细内容请参见《管理世界》网络发行版附录4)。

前文检验了情绪指标对整体市场回报的预测能力,本文将进一步检验中文金融大语言模型所刻度的个股情绪能否预测个股回报。基于证券关联新闻集,本文分别使用中文金融大语言模型和金融词典计算了与个股对应的新闻情绪。实证检验结果表明,基于大模型个股情绪构建的多空投资组合有显著为正的收益,且不能被主流多因子定价模型解释。此外,法马-麦克贝斯(Fama-MacBeth)回归检验同样表明个股情绪能够有效预测个股回报。从两种方法的对比上看,大模型个股情绪指标对个股情绪的预测效果好于金融词典情绪,再次体现了中文金融大语言模型的价值(限于篇幅,详细内容请参见《管理世界》网络发行版附录3)。

## 六、进一步分析

从前文的诸多检验结果来看,基于中文金融大语言模型构建的文本情绪指标在预测效力与实际应用价值等诸多方面都显著优于使用词典法构建的文本情绪指标。本文认为中文金融大语言模型对情绪刻画的优异性主要体现在以下两方面:第一,与情绪经济学意义的适配性。从金融理论文献可知,市场情绪是一个复杂变量,与市场回报存在着非线性、非对称的动态关系。本文所采用的中文金融大语言模型是一种典型的深度学习方法,并以市场回报为标签进行训练,能够更好地捕捉文本情绪与市场回报间的复杂关系,准确刻画理论中的情绪特征。第二,文本分析方法的先进性。与简单的词典法相比,中文金融大语言模型对文本结构做出了更加复杂与现实的设定,能够捕捉到文本中的丰富信息。这使得中文金融大语言模型在各种类型文本中都能具有良好分析效果,并在新文本下具有广泛的适用性。

为证明上述观点,本文将通过分析情绪指标在不同经济状态下和极端事件期间的预测效力来验证基于中文金融大语言模型构建的情绪更加符合理论中情绪的特性,并通过对比在长文本和短文本下的情绪指标预测能力来证明大模型情绪变量在文本分析方法上的优越性。

### (一)经济变量预测分析

从本哈比等(2016)等文章的理论分析可知,情绪变量除去预测资产价格与回报外,还能对经济基本面产生影响。本文在该部分检验大模型情绪变量与金融字典法情绪变量能否体现对宏观经济存在预测能力的情绪理论特征。本文选取如下代表性经济变量( $Econ$ )进行检验,包括:采购经理人指数( $PMI$ )、消费者物价指数( $CPI$ )、广义货币供应量( $M2$ )、社会融资规模存量同比增速( $SF$ )、公共财政支出( $F_{out}$ )、固定资产投资完成额( $inv$ )、房地产开发投资完成额( $house$ )、市盈率( $PE$ )。本文还参考姜富伟等(2021b),引入市场换手率( $Turn-over$ )、平滑波动率( $LVOL$ )变量测度噪音交易者带来的非理性波动和市场风险的风险水平的影响。平滑波动率计算方法如下:

$$SVAR_t = \sum_{i=1}^{N_t} \hat{R}_{i,t}^2 \quad (9)$$

$$LVOL_t = \ln(\sqrt{SVAR_t}) \quad (10)$$

其中, $\hat{R}_{i,t}$ 是第 $t$ 月第 $i$ 个交易日市场超额收益, $N_t$ 为第 $t$ 月的交易日个数。 $LVOL_t$ 是对 $SVAR_t$ 取对数以后的平滑波动率。

本文通过如下回归方程实证检验两类情绪变量在不同经济运行时期对经济变量的预测情况。

$$Econ_t = \alpha_i + \beta_i SENTIMENT_{i,t-1} + \varphi_i Econ_{t-1} + \varepsilon_{i,t}, i=1,2 \quad (11)$$

其中, $Econ_t$ 为在第 $t$ 期的经济变量, $SENTIMENT_{i,t-1}$ 是第 $t-1$ 期第 $i$ 个情绪指数, $i=1,2$ ,分别对应大模型情绪变量、金融字典法情绪变量。该部分经济运行时期的划分方式与前文样本内检验部分一致。

表4展示了情绪变量对经济变量的预测能力。无论在经济上行期还是经济下行期,金融字典法情绪变量对经济变量的预测能力都较为有限,仅对少数变量具有预测能力。与之对比,大模型情绪变量对各宏观经济变量的预测效果更好。尤其是在经济下行时期,大模型情绪变量能够显著预测几乎所有的宏观变量。在经济



上行期,大模型情绪变量也能显著预测PMI、换手率等重要指标。这一结果表明大模型情绪变量与理论模型中情绪能够影响基本面的特征更加符合。此外,大模型情绪变量在经济下行期对宏观变量预测更好,这与前文所发现的在经济下行期情绪对市场回报预测表现更好的发现相呼应。从金融理论文献中可以看出,资产价格波动在下行期和极端事件期间主要由情绪冲击影响。通过资产价格对经济基本面的影响,情绪也能在经济下行期表现出对宏观经济指标更强的预测能力。

## (二)极端事件预测分析

现有理论文献表明,情绪对市场回报存在着非对称的非线性影响,这在极端事件期间具体体现为情绪冲击可能导致市场回报的暴涨暴跌。如果基于中文金融大语言模型构建的情绪指标更好地体现了情绪与市场回报间的非线性关系,那么大模型情绪变量应当能够更好地预测各种极端事件期间市场回报变化。

此外,从文本分析方法的角度上讲,由于极端事件大多数是未预期到的黑天鹅事件,新闻中的相关叙述在用词与行文上会出现较大变化。这对词典法在极端事件期间的适用性提出了极大的挑战。举例来说,在新冠疫情发生之前,大量与疫情相关的情感词语在各种情感词典中都是欠缺的,这严重影响了词典法情绪在疫情期间的分析效果。然而,作为一类深度学习方法,中文金融大语言模型在训练时并不依赖于人为给定一个特定的词典,可以更好地捕捉字词与文本情绪之间的全面关系。因此,中文金融大语言模型在极端事件期间可以表现出对“陌生词语”与“新词语”的更好的适应性,从而获得比传统词典法更好的预测表现。

为验证大模型情绪变量的确具有上述两类优势,本文选择2007年股灾与2020年新冠疫情作为极端事件的代表。2007年股灾指2007年10月我国股票市场出现的暴跌,该事件及其前后一段时间的市场变化反映了我国股票市场受到的国际金融危机与部分国内风险叠加的影响。2020年新冠疫情指2020年1月左右突然发生的新冠肺炎疫情,我国股票市场受到了此次疫情事件的剧烈影响。参考弗兰克尔等(2022)的设定,本文首先借助日度回归来检验情绪指标对极端市场变动的预测能力,并选取两次极端事件发生前后100个交易日的市场交易数据作为分析样本。具体来说,本文进行如下日度回归检验:

$$r_{it}^m = \alpha + \beta \text{SENTIMENT}_{i,t-1} + \varphi_j \text{Control}_{i,t-1} + \varepsilon_{i,t}, i=1,2 \quad (12)$$

其中,  $\text{Control}_{i,t-1}$  为一组控制变量,包括滞后一期的市场收益率和计算窗口为15天的市场滚动波动率。其他变量含义同前文保持一致。

除去上述极端事件期间的日度样本内回归预测效力检验外,本文也从样本外预测检验的角度探究了两类情绪指标在极端事件期间的实时预测能力。借助前文的月度样本外回归预测检验,本文计算了两次极端事件发生的前后5个月(基本对应前后一百个交易日)间各种预测方法的累计平方误差和,并将其走势绘制

表4 情绪变量与经济变量预测

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: 经济上行期大模型情绪变量										
变量	PMI	CPI	M2	SF	$F_{int}$	inv	house	PE	Turnover	LVOL
$S_{lin}$	0.15* (1.85)	0.02 (0.87)	0.001 (0.95)	0.05 (0.71)	0.04 (0.67)	-0.01 (-0.99)	-0.004 (-0.50)	0.05 (1.11)	0.16** (2.54)	0.04 (0.54)
Econ	0.68*** (9.47)	0.94*** (34.07)	1.01*** (884.61)	0.98*** (12.42)	0.74*** (11.98)	1.10*** (106.37)	1.11*** (113.79)	0.93*** (19.84)	0.68*** (10.98)	0.69*** (9.77)
Constant	-0.05 (-0.73)	0.03 (1.31)	0.01*** (10.23)	-0.08 (-1.03)	0.01 (0.24)	0.16*** (17.53)	0.17*** (19.45)	-0.05 (-1.20)	-0.01 (-0.17)	0.03 (0.38)
$R^2$ (%)	52.11	91.32	99.98	82.39	54.01	99.13	99.24	86.45	57.87	47.64
Panel B: 经济下行期大模型情绪变量										
变量	PMI	CPI	M2	SF	$F_{int}$	inv	house	PE	Turnover	LVOL
$S_{lin}$	0.06 (0.71)	-0.09** (-2.16)	0.01*** (3.35)	0.09* (1.70)	0.27*** (2.58)	0.14* (1.81)	0.13* (1.81)	0.13*** (3.66)	0.27*** (3.70)	-0.01 (-0.10)
Econ	0.52*** (5.69)	0.88*** (20.27)	1.01*** (430.32)	0.99*** (17.52)	0.64*** (5.32)	0.71*** (8.80)	0.78*** (10.30)	0.93*** (26.31)	0.64*** (8.82)	0.61*** (7.28)
Constant	0.04 (0.43)	-0.06 (-1.39)	0.02*** (10.55)	0.09 (1.64)	0.05 (0.51)	0.03 (0.37)	0.04 (0.52)	0.05 (1.30)	0.03 (0.42)	-0.05 (-0.62)
$R^2$ (%)	26.74	84.54	99.95	91.41	24.77	47.14	54.96	88.87	52.11	36.87
Panel C: 经济上行期金融字典法情绪变量										
变量	PMI	CPI	M2	SF	$F_{int}$	inv	house	PE	Turnover	LVOL
$S_{dic}$	0.01 (0.15)	0.004 (0.12)	0.0004 (0.30)	-0.001 (-0.01)	0.08 (1.30)	-0.01 (-1.10)	0.004 (0.40)	0.04 (0.96)	0.13** (2.23)	-0.17** (-2.15)
Econ	0.72*** (10.14)	0.95*** (29.03)	1.01*** (852.78)	0.97*** (12.29)	0.72*** (11.48)	1.10*** (104.50)	1.11*** (109.13)	0.96*** (22.10)	0.72*** (12.15)	0.62*** (7.88)
Constant	-0.03 (-0.44)	0.03 (1.31)	0.01*** (10.20)	-0.08 (-1.05)	0.01 (0.24)	0.16*** (17.51)	0.17*** (19.44)	-0.04 (-1.03)	-0.01 (-0.19)	0.03 (0.43)
$R^2$ (%)	50.51	91.26	99.98	82.12	54.47	99.13	99.23	86.40	57.36	49.45
Panel D: 经济下行期金融字典法情绪变量										
变量	PMI	CPI	M2	SF	$F_{int}$	inv	house	PE	Turnover	LVOL
$S_{dic}$	-0.03 (-0.38)	0.04 (0.89)	0.002 (1.03)	0.02 (0.29)	0.07 (0.77)	-0.04 (-0.56)	-0.05 (-0.69)	0.09** (2.33)	0.09 (1.13)	0.03 (0.34)
Econ	0.52*** (5.61)	0.90*** (21.45)	1.00*** (415.34)	0.96*** (15.80)	0.49*** (4.51)	0.68*** (8.50)	0.76*** (10.00)	0.92*** (24.89)	0.67*** (8.68)	0.61*** (7.22)
Constant	0.04 (0.42)	-0.06 (-1.37)	0.02*** (10.04)	0.09 (1.61)	0.04 (0.45)	0.03 (0.36)	0.04 (0.50)	0.05 (1.37)	0.03 (0.42)	-0.05 (-0.61)
$R^2$ (%)	26.44	83.89	99.95	90.58	19.50	45.35	53.51	87.94	45.67	36.94

为折线图。不同方法对应的累计平方误差和在极端事件期间的走势变化可以反映不同方法的样本外预测能力强弱。

表5报告了2007年股灾事件与2020年新冠疫情期间的日度回归预测检验结果。无论是否引入控制变量,大模型情绪变量对市场回报都具有显著预测能力。而金融字典法情绪变量在纳入控制变量后不再具有对市场回报的预测能力。这一结果对比表明大模型情绪变量在极端事件期间表现出了对市场回报更强的预测能力,说明大模型情绪变量更加符合理论文献中表明的情绪所应具有的经济特征。

从两次极端事件期间的表现对比来看,金融字典法与中文金融大语言模型的差距在新冠疫情期间更加明显。在2007年股灾中,金融字典法情绪变量在单变量回归中表现出了对市场回报的预测能力。而在新冠疫情期间,无论是否包含控制变量,金融字典法情绪变量均不能显著预测市场回报。本文所使用的金融字典来自于姜富伟等(2021b),该文章基于2017年及之前的新闻样本构建词典。因此,该词典在2007年股灾样本中表现相对更好,在新冠疫情期间预测效力相对更差。基于中文金融大语言模型所构建的情绪变量表现出了更强的对预期外极端事件的预测能力,这证明引入更复杂的深度学习方法对于提高情绪变量样本外适用性极有帮助。

图6进一步绘制了基于两种情绪指标的样本外预测在极端事件期间的累计平方误差和变动趋势。与表5结果一致,在极端事件期内,大模型情绪变量的预测表现一直优于金融字典法情绪变量,持续具有更小的预测累计误差平方和。此外,大模型情绪变量相对于金融字典法情绪变量的优势在新冠疫情期间更加明显。这些结果再次表明大模型情绪变量与理论特征更加契合,且在样本外文本分析中更加适用。

### (三)文本特征预测分析

在该部分,本文聚焦于中文金融大语言模型在文本分析技术上的优越性。与传统的字典法等传统文本分析方法相比,中文金融大语言模型的一大突出优点是适用范围广,在不同类型文本上均能获得良好的分析表现。字典法受限于所使用的字典词汇,只能适用于某一特定语境文本。同时,在一些口语化风格明显、用语不规范或者长度过短的文本中,字典法也难以提取对应情绪关键词,表现较差。为此,本文选择从文本长度这一文本特征入手,探讨中文金融大语言模型在不同特征文本下的表现,为理解将中文金融大语言模型引入到金融领域文本分析的价值提供更多参考。

具体来说,本文将新闻标题与新闻正文作为短文本与长文本的代表,并利用两类文本分别计算文本情绪,然后进行预测能力对比。与完整的新闻正文相比,以新闻标题为代表的短文本中信息含量更少。在投资者有限注意力理论下,新闻标题会得到投资者的更多关注,从而对投资者的投资决策产生显著影响(游家兴、吴静,2012)。对短文本的过度关注会导致投资者忽视长文本的重要信息,从而难以准确定价。从这一角度来说,探讨新闻标题情绪对市场的影响也具有经济学价值。

在应用字典法分析新闻标题情绪时,非常容易遇到标题中不含情感关键词的情况。对于这类新闻标题,本文统一设定其在字典法下的情绪值为0。与之对比,在使用中文金融大语言模型进

表5 极端事件期间情绪变量与市场收益预测

变量	(1)	(2)	(3)	(4)
Panel A: 2007年股灾期间				
$S_{lm}$	0.014** (2.10)		0.012* (1.91)	
$S_{dic}$		0.259* (1.85)		0.227 (1.60)
Constant	Yes	Yes	Yes	Yes
Control	No	No	Yes	Yes
$R^2(\%)$	2.18	1.69	3.63	3.11
Panel B: 2020年新冠疫情期间				
$S_{lm}$	0.005* (1.87)		0.005* (1.89)	
$S_{dic}$		0.057 (1.43)		0.069 (1.57)
Constant	Yes	Yes	Yes	Yes
Control	No	No	Yes	Yes
$R^2(\%)$	1.74	1.03	1.83	1.30

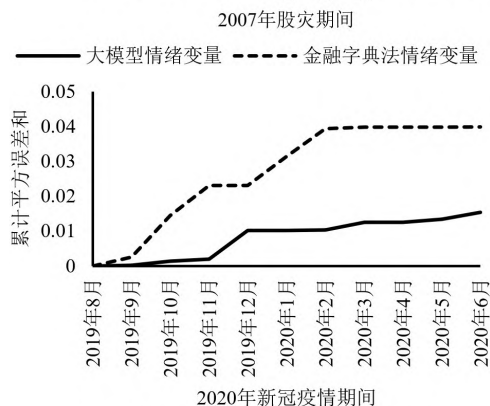
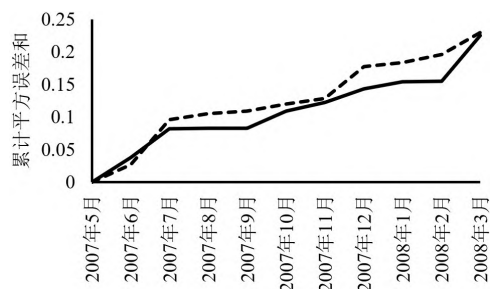


图6 极端事件期间累计平方误差和变动趋势

行情感分析时,任意新闻标题都可以输出相应的情绪值,这也表明了中文金融大语言模型在不同特征文本上的广泛适用性。在利用新闻标题计算文本情绪后,本文继续进行前文的月度样本内预测检验。

表6显示了月度新闻标题情绪与整体股票市场收益的关系。从样本内预测检验可以看出,基于新闻标题构建的大模型情绪变量与金融字典法情绪变量都仍然具有对市场回报显著的正向预测能力,与前文的基准检验结果一致。同时,上述两种情绪变量的预测性回归 $R^2$ 均超过了坎贝尔和汤普森(2008)中0.5%的标准,因此两种情绪变量的预测能力都具有显著的经济意义。这些结果说明新闻标题中的情绪信息也有助于预测市场回报。从与长文本新闻正文的检验结果对比来看,使用短文本新闻标题构建的情绪指标预测能力更弱。这表明长文本含有更加丰富的信息,短文本虽然具有情绪代表性,但是仍然损失了一些对收益预测有关键信息。从两种情绪指标的对比来看,基于中文金融大语言模型构建的情绪变量在预测市场回报上表现优于字典法,具有更高的预测性 $R^2$ 。这一结果初步表明了中文金融大语言模型在短文本上更强的适用性,这也是中文金融大语言模型价值的重要体现。

为了对中文金融大语言模型与字典法在新闻标题上的应用效果做出更加具体与直观的对比,本文参照前文样本外预测检验方法,利用两种情绪指标对市场回报进行样本外预测检验,并按照公式(6)进行资产配置。在这一部分,本文使用前文的样本外区间一来进行检验,即初始样本内区间为2003年1月~2005年12月,样本外预测区间为2006年1月~2021年12月。本文计算按照上述方法资产配置的简单累计超额收益率,并绘制图7。

在图7中,为了更好地进行对比,本文除了使用金融大模型和金融字典法计算的短文本新闻标题情绪变量以外,还使用前文基于长文本新闻正文计算的两类文本情绪变量进行资产配置检验,最后与利用历史均值进行基准预测的资产配置结果相对比,绘制了上述4类文本情绪资产配置组合与基于基准预测的资产配置组合的简单累计超额收益率时序变动图。可以明显看出,使用中文金融大语言模型分别基于长文本和短文本构建的两个情绪指标获得最佳的资产配置结果,分别能够获得约为400%与350%的累计超额收益率。尽管新闻标题较短,与新闻正文相比信息含量较低,但基于新闻标题构建的大模型情绪变量仍然获得了远超基于新闻正文构建的金融字典法情绪变量的资产配置收益结果。使用新闻标题构建的金融字典法情绪变量则只取得了略好于基准预测的资产配置结果。此外,使用中文金融大语言模型构建的情绪变量在股灾等极端事件期间都取得了较好的表现,而同期的市场表现都较差,这也突出了中文金融大语言模型的应用价值。综上,上述结果不仅表明中文金融大语言模型在情绪分析上的优势,还论证了中文金融大语言模型在不同特征的文本上都具有良好的表现与广泛的适用性,进一步拓展了人工智能大模型文本情绪分析技术的综合应用场景。

表6 月度新闻标题情绪与整体股票

变量	市场收益			
	(1) 等权重	(2) 等权重	(3) 市值加权	(4) 市值加权
$S_{lm}$	0.214*** (3.29)		0.200*** (3.07)	
$S_{dic}$		0.130** (1.96)		0.136** (2.05)
Constant	-0.004 (-0.05)	-0.003 (-0.05)	-0.004 (-0.07)	-0.004 (-0.06)
$R^2$ (%)	4.59	1.68	4.03	1.83

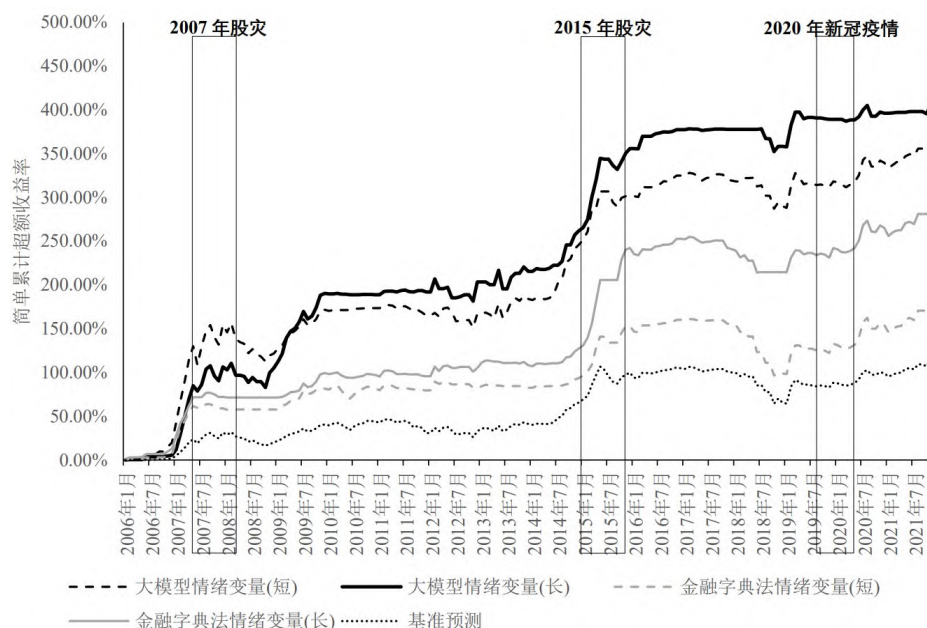


图7 文本情绪资产组合与基准预测资产组合的简单累计超额收益率时序变动



## 七、结论与政策建议

“人工智能+”行动是发展新质生产力的重要途径,其在金融领域的应用有助于金融强国建设。在此背景下,本文将人工智能大语言模型算法与中国特色金融市场应用相结合,融合结构化金融市场数据和非结构化金融文本大数据,构造了与我国市场关联紧密并更符合金融理论特征的中文金融大语言模型,并开展金融市场情绪测度和资产价格风险预测。为此,本文将BERT模型这一深度学习算法与大语言模型引入到情绪分析,基于五百多万条金融新闻文本语料大数据,并将我国股票市场回报作为训练标签,从而使得模型可以提取资产价格中隐含的情绪信息,训练了中文金融大语言模型。随后,本文计算了基于大模型的文本情绪,与现有文献中的基于金融情感字典法的文本情绪变量进行了对比。

实证结果表明,使用中文金融大语言模型构建的大模型情绪变量指标在样本内与样本外预测检验中均取得了明显优于金融字典法情绪变量的表现。从资产配置检验结果来看,基于大模型情绪变量的预测也能够取得更好的投资表现。以上结果论证了中文金融大语言模型在金融文本情绪分析中的应用价值。从进一步分析来看,借助中文金融大语言模型可以构造更加符合金融理论模型的非线性、非对称情绪指标。大模型情绪变量在不同经济阶段、极端事件期间的预测表现以及对宏观变量的预测能力均与金融理论相符。其次,中文金融大语言模型显著改进了金融文本分析的效果与精度。由于不需要预先设定字典,大模型情绪变量在突发事件文本以及各种不同特征文本上表现更佳,这体现出中文金融大语言模型在文本分析方法论上的先进性。

从应用前景来看,随着深度学习技术的进步、算力的提升和数据的累积,各类人工智能技术将在经济金融领域得到广泛应用。本文认为“人工智能+”行动的金融领域应用,需要以金融理论为指引,以中国特色金融市场特征事实为导向,以海量金融大数据为基础,融合结构化金融市场数据与非结构化金融文本数据,结合人工智能算法优势和算力资源,拓展“人工智能+金融”综合应用场景,实现市场导向、理论指引、数据驱动、技术支持的有机统一。本文对中文金融大语言模型在金融领域应用的探索仍然较为初步,仍有待完善和扩展。例如,可以将人工情绪标注与市场情绪标签融合,从而使大模型捕捉更全面的情绪信息。也可借助大模型构造复杂性等其他文本指标,探索金融文本多维丰富信息。还可对比ChatGPT等其它大语言模型。除去媒体新闻外,政策文本、论坛文本等也是金融市场上重要的文本,其中蕴含的信息、涉及的市场主体等都存在差异。在数据可得性不断提高的前提下,思考不同来源文本中情绪信息的异同与交互作用也是重要的研究方向。在得到更好的大模型情绪变量后,理论探讨情绪如何与市场风险、泡沫等相关联,以及如何采取政策更好地应对情绪冲击以实现金融稳定,也具有很强的现实意义。

本文研究同样具有一定的政策意义。“人工智能+”行动是发展新质生产力的重要途径,本文研究对人工智能在金融领域的应用有一定政策借鉴意义。第一,人工智能算法在金融领域的应用有着极大潜力,政策可在一定程度上予以推动。本文以测度金融市场情绪为切入点,构造了中文金融大语言模型,获得了预测效果更好和与理论特征更加匹配的大模型情绪变量,这是人工智能在金融领域应用潜力的体现。考虑到情绪对金融市场的显著影响,这一大模型情绪变量有被应用至政策制定、风险监控等领域的潜力。金融市场是一个数据丰富的行业,参与者众多,市场的运行时刻积累着海量数据,这为人工智能应用也都提供了大数据基础。因此,政策可在一定程度上推动人工智能与金融领域的融合,助力金融强国建设。

第二,开展“人工智能+”的金融垂直应用时需注意结合中国特色金融市场的独特特征,在金融理论与实践的指引下提高金融大模型的可解释性和安全性。金融市场交易机制、数据特征、参与者等均具有独特性,不同于通用人工智能应用场景。本文模型的良好表现在一定程度源于模型训练算法与金融理论的高度契合,将金融市场结构化信息应用到了非结构化人工智能大模型的训练中。这表明基于中国特色金融市场具体问题,并与金融理论和实践密切结合,才能充分挖掘“人工智能+”在金融强国和打造新质生产力上的潜力。

第三,积极借助人工智能大模型等开展金融市场非理性投机情绪监控工作。情绪是金融泡沫、金融风险、投机行为等的重要驱动因素。然而,情绪和预期等作为主观信念存在测度难题,如果没有一种良好的实时监控方法,那么就会阻碍有关部门及时地追踪并应对市场情绪变化,无法及时遏制市场中的负面因素。本文的

研究展示了人工智能手段在金融市场情绪风险监控上的应用空间。随着金融市场上各类大数据不断积累和人工智能大模型算法的不断发展,基于人工智能手段的情绪风险监控可以有益地补充传统的基于市场交易数据的监控手段,为金融政策制定和风险防范提供一定支撑<sup>⑥</sup>。

(作者单位:姜富伟,厦门大学经济学院、厦门大学王亚南经济研究院;刘雨旻,中央财经大学金融学院;孟令超,对外经济贸易大学中国金融学院)

#### 注释

①抱抱脸是一家专门开发深度学习应用和工具的公司,该公司的代表产品是其自然语言处理应用构建的Transformers库以及允许用户共享深度学习模型和数据集的平台。本文应用了抱抱脸官方人工智能实验室针对中文语料的开源BERT预训练模型:<https://huggingface.co/bert-base-chinese>。

②因此,本文得到的中文金融大语言模型实际上是在专业实验室的大规模通用领域语料与金融特色的微博财经新闻上进行的预训练,可以同时满足使用高质量大规模数据集进一步预训练要求,与满足引入金融领域文本特征的应用要求。

③另外一种利用资产回报信息的方式是直接利用机器学习模型和回报数据生成一个综合指数,并预测回报。本文则首先生成情绪变量,然后进行回报预测,这有着更强的理论基础(投资者情绪与资产回报间的关系被大量文献证明),更有潜力取得稳定的预测表现。此外,本文做法还能够将文本分析与资产回报相结合,同时利用新闻文本与资产回报中的丰富信息。

④比如:当第 $t$ 交易日为周一时,此处设定是指周一运算使用的数据包含上周的周六周日和周一这3天的新闻数据;当第 $t$ 交易日为周三等非周一情形,此处设定是指使用周三当天的新闻数据。

⑤风险厌恶系数的具体系数没有共识。但从资产定价实证文献来看,风险厌恶系数取值基本在10以内(姜等,2019;黄等,2015;拉帕奇、周,2021;德米格尔等,2009;拉帕奇等,2010;姜富伟等,2021b)。同时,主流资产定价模型中得到的风险厌恶系数也一般在10以内(科切拉科塔,1996;坎贝尔、科克伦,1999;班萨尔、亚伦,2004;巴罗,2006;马洛伊等,2009;班萨尔等,2016)。为此,本文将风险厌恶系数设置为3、5、7。

⑥中外文人名(机构名)对照:夏皮罗(Shapiro);贝克(Baker);沃格勒(Wurgler);本哈比(Benhbib);黄(Huang);杨(Yang);刘(Liu);柯(Ke);周(Zhou);石(Shi);席勒(Shiller);德隆(De Long);帕洛米诺(Palomino);洛克伦(Loughran);麦克唐纳(McDonald);安吉利托斯(Angelatos);拉奥(La'O);佩里(Petri);奎德里尼(Quadrini);门德尔(Mendel);施莱弗(Shleifer);伯南克(Bernanke);泰特洛克(Tetlock);李(Li);游(You);杜(Du);卞(Bian);科雷亚(Correa);巴苏(Basu);弗兰克尔(Frankel);德夫林(Devlin);阿迪卡里(Adhikari);古兰甘(Gururangan);李(Lee);阿拉奇(Araci);韩(Han);艾森斯坦(Eisenstein);洛格斯瓦兰(Logeswaran);孙(Sun);贝尔塔吉(Beltagy);霍华德(Howard);鲁德(Ruder);杰加迪什(Jegadeesh);吴(Wu);西加诺斯(Siganos);姜(Jiang);拉帕奇(Rapach);坎贝尔(Campbell);汤普森(Thompson);韦尔奇(Welch);戈亚尔(Goyal);德米格尔(DeMiguel);抱抱脸(Hugging Face);科切拉科塔(Kocherlakota);科克伦(Cochrane);班萨尔(Bansal);亚伦(Yaron);巴罗(Barro);马洛伊(Malloy)。

#### 参考文献

- (1)陈彦斌:《情绪波动和资产价格波动》,《经济研究》,2005年第3期。
- (2)范小云、王业东、王道平:《基于新闻大数据与机器学习的中国银行业系统性风险研究》,《世界经济》,2022年第4期。
- (3)范小云、王业东、王道平、郭文璇、胡煊翊:《不同来源金融文本信息含量的异质性分析——基于混合式文本情绪测度方法》,《管理世界》,2022年第10期。
- (4)洪永森、汪寿阳:《大数据如何改变经济学研究范式?》,《管理世界》,2021年第10期。
- (5)胡楠、薛付婧、王昊楠:《管理者短视主义影响企业长期投资吗?——基于文本分析和机器学习》,《管理世界》,2021年第5期。
- (6)姜富伟、胡逸驰、黄楠:《央行货币政策报告文本信息、宏观经济与股票市场》,《金融研究》,2021年第6期。
- (7)姜富伟、李梦如、孟令超:《调预期、防风险:央行金融稳定沟通效果研究》,工作论文,2022年。
- (8)姜富伟、孟令超、唐国豪:《媒体文本情绪与股票回报预测》,《经济学(季刊)》,2021年第4期。
- (9)姜富伟、涂俊、David E. Rapach、Jack K. Strauss、周国富:《中国股票市场可预测性的实证研究》,《金融研究》,2011年第9期。
- (10)李斌、邵新月、李玥阳:《机器学习驱动的基本面量化投资研究》,《中国工业经济》,2019年第8期。
- (11)林建浩、陈良源、罗子豪、张一帆:《央行沟通有助于改善宏观经济预测吗?——基于文本数据的高维稀疏建模》,《经济研究》,2021年第3期。
- (12)林建浩、陈良源、宋登辉:《如何测度央行行长的口头沟通信息——一种基于监督学习的文本分析方法》,《统计研究》,2019年第8期。
- (13)马甜、姜富伟、唐国豪:《深度学习与中国股票市场因子投资——基于生成式对抗网络方法》,《经济学(季刊)》,2022年第3期。
- (14)潘怡麟、张舒怡、朱凯:《文过饰非还是秉笔直书:中国债券评级报告文本信息的价值相关性》,《南开管理评论》,2023年第4期。
- (15)苏治、卢曼、李德轩:《深度学习的金融实证应用:动态、贡献与展望》,《金融研究》,2017年第5期。
- (16)唐国豪、姜富伟、张定胜:《金融市场文本情绪研究进展》,《经济学动态》,2016年第11期。
- (17)汪昌云、武佳薇:《媒体语气、投资者情绪与IPO定价》,《金融研究》,2015年第9期。
- (18)谢德仁、林乐:《管理层语调能预示公司未来业绩吗?——基于我国上市公司年度业绩说明会的文本分析》,《会计研究》,2015年第2期。
- (19)许雪晨、田侃:《一种基于金融文本情感分析的股票指数预测新方法》,《数量经济技术经济研究》,2021年第12期。
- (20)姚加权、冯绪、王赞钧、纪荣嵘、张维:《语调、情绪及市场影响:基于金融情绪词典》,《管理科学学报》,2021年第5期。

- (21) 游家兴、吴静:《沉默的螺旋:媒体情绪与资产误定价》,《经济研究》,2012年第7期。
- (22) 曾庆生、周波、张程、陈信元:《年报语调与内部人交易:“表里如一”还是“口是心非”?》,《管理世界》,2018年第9期。
- (23) 张成思、芦哲:《媒体舆论、公众预期与通货膨胀》,《金融研究》,2014年第1期。
- (24) 张琦、步丹璐、郁智:《媒体关注、报道情绪与政府“三公”预算抑制》,《经济研究》,2016年第5期。
- (25) 张晓燕、葛慧敏:《新闻语调与中国股票市场收益率》,《经济管理学报》,2022年第1期。
- (26) 张紫琼、叶强、李一军:《互联网商品评论情感分析研究综述》,《管理科学学报》,2010年第6期。
- (27) Adhikari, A., Ram, A., Tang, R. and Lin, J., 2019, “Docbert: Bert for Document Classification”, ArXiv Preprint, No.1904.08398.
- (28) Angeletos, G. and La'O, J., 2013, “Sentiments”, *Econometrica*, vol.81, pp.739~779.
- (29) Araci, D., 2019, “Finbert: Financial Sentiment Analysis with Pre-trained Language Models”, ArXiv Preprint, No.1908.10063.
- (30) Baker, M. and Wurgler, J., 2007, “Investor Sentiment in the Stock Market”, *Journal of Economic Perspectives*, vol.21, pp.129~152.
- (31) Bansal, R., Kiku, D. and Yaron, A., 2016, “Risks for the Long Run: Estimation with Time Aggregation”, *Journal of Monetary Economics*, vol.82, pp.52~69.
- (32) Bansal, R. and Yaron, A., 2004, “Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles”, *Journal of Finance*, vol.59, pp.1481~1509.
- (33) Barro, R. J., 2006, “Rare Disasters and Asset Markets in the Twentieth Century”, *Quarterly Journal of Economics*, vol.121, pp.823~866.
- (34) Basu, S., Ma, X. and Briscoe-Tran, H., 2022, “Measuring Multidimensional Investment Opportunity Sets with 10-K Text”, *The Accounting Review*, vol.97, pp.51~73.
- (35) Beltagy, I., Lo, K. and Cohan, A., 2019, “SciBERT: A Pretrained Language Model for Scientific Text”, ArXiv Preprint, No.1903.10676.
- (36) Benhabib, J., Liu, X. and Wang, P., 2016, “Sentiments, Financial Markets, and Macroeconomic Fluctuations”, *Journal of Financial Economics*, vol.120, pp.420~443.
- (37) Bernanke, B., 2013, “What Should Economists and Policymakers Learn from the Financial Crisis”, Public Lecture, <https://www.federalreserve.gov/newsevents/speech/bernanke20130325a.htm>.
- (38) Bian, S., Jia, D., Li, F. and Yan, Z., 2021, “A New Chinese Financial Sentiment Dictionary for Textual Analysis in Accounting and Finance”, SSRN Working Paper, No.3446388.
- (39) Campbell, J. Y. and Cochrane, J. H., 1999, “By Force of Habit: A Consumption-based Explanation of Aggregate Stock Market Behavior”, *Journal of Political Economy*, vol.107, pp.205~251.
- (40) Campbell, J. Y. and Thompson, S. B., 2008, “Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?”, *Review of Financial Studies*, vol.21, pp.1509~1531.
- (41) Correa, R., Garud, K., Londono, J. M. and Mislant, N., 2021, “Sentiment in Central Banks’ Financial Stability Reports”, *Review of Finance*, vol.25, pp.85~120.
- (42) De Long, J. B., Shleifer, A., Summers, L. H. and Waldmann, R. J., 1990, “Noise Trader Risk in Financial Markets”, *Journal of Political Economy*, vol.98, pp.703~738.
- (43) DeMiguel, V., Garlappi, L. and Uppal, R., 2009, “Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy?”, *Review of Financial Studies*, vol.22, pp.1915~1953.
- (44) Devlin, J., Chang, M., Lee, K. and Toutanova, K., 2018, “Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding”, ArXiv Preprint, No.1810.04805.
- (45) Du, Z., Huang, A. G., Wermers, R. and Wu, W., 2022, “Language and Domain Specificity: A Chinese Financial Sentiment Dictionary”, *Review of Finance*, vol.26, pp.673~719.
- (46) Frankel, R., Jennings, J. and Lee, J., 2022, “Disclosure Sentiment: Machine Learning vs. Dictionary Methods”, *Management Science*, vol.68, pp.5514~5532.
- (47) Gururangan, S., Marasović, A., Swayamdipta, S., Lo, K., Beltagy, I., Downey, D. and Smith, N. A., 2020, “Don’t Stop Pretraining: Adapt Language Models to Domains and Tasks”, ArXiv Preprint, No.2004.10964.
- (48) Han, X. and Eisenstein, J., 2019, “Unsupervised Domain Adaptation of Contextualized Embeddings for Sequence Labeling”, ArXiv Preprint, No.1904.02817.
- (49) Howard, J. and Ruder, S., 2018, “Universal Language Model Fine-tuning for Text Classification”, ArXiv Preprint, No.1801.06146.
- (50) Huang, A. H., Wang, H. and Yang, Y., 2023, “FinBERT: A Large Language Model for Extracting Information from Financial Text”, *Contemporary Accounting Research*, vol.40, pp.806~841.
- (51) Huang, D., Jiang, F., Tu, J. and Zhou, G., 2015, “Investor Sentiment Aligned: A Powerful Predictor of Stock Returns”, *Review of Financial Studies*, vol.28, pp.791~837.
- (52) Jegadeesh, N. and Wu, D., 2013, “Word Power: A New Approach for Content Analysis”, *Journal of Financial Economics*, vol.110, pp.712~729.
- (53) Jiang, F., Lee, J., Martin, X. and Zhou, G., 2019, “Manager Sentiment and Stock Returns”, *Journal of Financial Economics*, vol.132, pp.126~149.
- (54) Ke, Z. T., Kelly, B. T. and Xiu, D., 2019, “Predicting Returns with Text Data”, NBER Working Paper, No.26186.
- (55) Kocherlakota, N. R., 1996, “The Equity Premium: It’s Still a Puzzle”, *Journal of Economic Literature*, vol.34, pp.42~71.



- (56) Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H. and Kang, J., 2020, "BioBERT: A Pre-trained Biomedical Language Representation Model for Biomedical Text Mining", *Bioinformatics*, vol.36, pp.1234~1240.
- (57) Li, F., 2010, "Textual Analysis of Corporate Disclosures: A Survey of the Literature", *Journal of Accounting Literature*, vol.29, pp.143~165.
- (58) Li, J., Chen, Y., Shen, Y., Wang, J. and Huang, Z., 2019, "Measuring China's Stock Market Sentiment", SSRN Working Paper, No.3377684.
- (59) Liu, Y., 2019, "Fine-tune BERT for Extractive Summarization", ArXiv Preprint, No.1903.10318.
- (60) Liu, Z., Huang, D., Huang, K., Li, Z. and Zhao, J., 2021, "Finbert: A Pre-trained Financial Language Representation Model for Financial Text Mining", Proceedings of the Twenty-Ninth International Conference on Artificial Intelligence.
- (61) Logeswaran, L., Chang, M., Lee, K., Toutanova, K., Devlin, J. and Lee, H., 2019, "Zero-shot Entity Linking by Reading Entity Descriptions", ArXiv Preprint, No.1906.07348.
- (62) Loughran, T. and McDonald, B., 2011, "When is a Liability not a Liability? Textual Analysis, Dictionaries, and 10-Ks", *Journal of Finance*, vol.66, pp.35~65.
- (63) Malloy, C. J., Moskowitz, T. J. and Vissing-Jørgensen, A., 2009, "Long-run Stockholder Consumption Risk and Asset Returns", *Journal of Finance*, vol.64, pp.2427~2479.
- (64) Mendel, B. and Shleifer, A., 2012, "Chasing Noise", *Journal of Financial Economics*, vol.104, pp.303~320.
- (65) Palomino, F., 1996, "Noise Trading in Small Markets", *Journal of Finance*, vol.51, pp.1537~1550.
- (66) Perri, F. and Quadrini, V., 2018, "International Recessions", *American Economic Review*, vol.108, pp.935~984.
- (67) Rapach, D. E., Strauss, J. K. and Zhou, G., 2010, "Out-of-sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy", *Review of Financial Studies*, vol.23, pp.821~862.
- (68) Rapach, D. E. and Zhou, G., 2021, "Asset Pricing: Time-Series Predictability", SSRN Working Paper, No.3941499.
- (69) Shapiro, A. H., Sudhof, M. and Wilson, D. J., 2022, "Measuring News Sentiment", *Journal of Econometrics*, vol.228, pp.221~243.
- (70) Shi, Y., An, Y., Zhu, X. and Jiang, F., 2022, "Better to Hear All Parties: Understanding the Impact of Homophily in Online Financial Discussion", *Electronic Commerce Research and Applications*, vol.54, pp.1~18.
- (71) Shiller, R. J., 2015, *Irrational Exuberance*, Princeton University Press.
- (72) Siganos, A., Vagenas-Nanos, E. and Verwijmeren, P., 2017, "Divergence of Sentiment and Stock Market Trading", *Journal of Banking & Finance*, vol.78, pp.130~141.
- (73) Sun, C., Qiu, X., Xu, Y. and Huang, X., 2019, "How to Fine-tune Bert for Text Classification?", China National Conference on Chinese Computational Linguistics.
- (74) Tetlock, P. C., 2007, "Giving Content to Investor Sentiment: The Role of Media in the Stock Market", *Journal of Finance*, vol.62, pp.1139~1168.
- (75) Welch, I. and Goyal, A., 2008, "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction", *Review of Financial Studies*, vol.21, pp.1455~1508.
- (76) Yang, Y., Uy, M. C. S. and Huang, A., 2020, "Finbert: A Pretrained Language Model for Financial Communications", ArXiv Preprint, No.2006.08097.
- (77) You, J., Zhang, B. and Zhang, L., 2018, "Who Captures the Power of the Pen?", *Review of Financial Studies*, vol.31, pp.43~96.
- (78) Zhou, Y., Fan, J. and Xue, L., 2024, "How Much Can Machines Learn Finance from Chinese Text Data?", *Management Science*, Forthcoming.

## Large Language Model and Textual Sentiment Analysis in Chinese Stock Markets

Jiang Fuwei<sup>a,b</sup>, Liu Yumin<sup>c</sup> and Meng Lingchao<sup>d</sup>

(a. School of Economics, Xiamen University; b. The Wang Yanan Institute for Studies in Economics, Xiamen University; c. School of Finance, Central University of Finance and Economics; d. China School of Banking and Finance, University of International Business and Economics)

**Abstract:** The AI Plus initiative is an important way to develop new quality productive forces, and its implementation in the financial industry contributes to building a financial powerhouse. This paper trains a Chinese financial large language model more suitable for Chinese stock markets by innovatively combining structured financial data and unstructured financial textual big data and taking into account the unique characteristics of the Chinese financial market. This paper then measures the financial market sentiment and predicts the financial market risks using the trained large language model. Empirical results show that the constructed large model sentiment variable using the Chinese financial large language model outperforms traditional dictionary methods in predicting asset prices. The large model sentiment variable can significantly predict macroeconomic variables across different economic conditions, capturing the impact of irrational shocks on macroeconomic fundamentals. The predictive power of the large model sentiment variable on market returns is stronger during economic downturns and extreme risk events, aligning with the theory that irrational sentiment has asymmetric and nonlinear effects on financial markets and the macroeconomy. Overall, our paper illustrates the potential pathways and the theoretical logic for integrating the AI Plus initiative into the financial industry.

**Keywords:** textual sentiment; deep learning; large language model; asset pricing

# Large Language Model and Textual Sentiment Analysis in Chinese Stock Markets

Jiang Fuwei<sup>a,b</sup>, Liu Yumin<sup>c</sup> and Meng Lingchao<sup>d</sup>

(a. School of Economics, Xiamen University; b. The Wang Yanan Institute for Studies in Economics, Xiamen University; c. School of Finance, Central University of Finance and Economics; d. China School of Banking and Finance, University of International Business and Economics)

**Summary:** The 2024 China government work report highlights the need to "step up R&D and application of big data and AI, launch an AI Plus initiative, and build digital industry clusters with international competitiveness." In particular, the rapid development of large language models reflects a broad application prospect. Therefore, from the perspective of measuring irrational sentiment in the financial market, our paper trains a Chinese financial large language model suitable for Chinese stock markets by innovatively combining structured financial data and unstructured financial textual big data and taking into account the unique characteristics of the Chinese financial market. Then, we utilize this model for financial market sentiment measurement and asset price prediction.

This paper constructs the Chinese financial large language model based on the BERT (Bidirectional Encoder Representations from Transformers) model. In model training, this paper uses market returns as the sentiment label, which allows the constructed model to learn financial information. This paper uses media news big data to train the Chinese financial large language model, to calculate the large model sentiment variables, and to test the predictive ability of this sentiment variable on market returns and risks. We also calculate sentiment using a financial sentiment dictionary for comparison. Empirical results show that the large model sentiment variable significantly outperforms the dictionary sentiment in both in-sample and out-of-sample predictive tests, and achieves better performance in asset allocation. Further analysis reveals that the large language model sentiment variable can also predict macroeconomic variables and has stronger predictive power during extreme events and economic downside. These results suggest that the large model sentiment variable fits well with the investor sentiment theory.

Our paper has the following policy implications. First, this study fully demonstrates the value and potential of AI models in solving important financial problems, and therefore, relevant policies should be introduced to further promote the application of AI Plus Finance. Second, the applying experience of our model shows that to develop vertical financial applications of AI Plus initiative, the key is to consider the specific problems of the financial market with unique Chinese characteristics, financial theory and practice. Third, our paper also fully demonstrates the application of AI tools in irrational market sentiment monitoring, and supports for policymaking.

Our paper has the following contributions. First, we train a Chinese financial large language model suitable for Chinese stock market, which improves the effectiveness of the sentiment analysis in Chinese stock market. We construct the model based on BERT model. Existing literature primarily focuses on English texts and lacks an in-depth analysis of the relationship between sentiment and Chinese stock market. Our paper fills this gap in the literature. Second, we take advantage of the large language model to identify the complex nonlinear and asymmetric characteristics of sentiment in Chinese financial market, and construct a sentiment variable more consistent with the financial theory, which is supported by the empirical results. Third, we use market returns as sentiment labels and apply them to the fine-tuning of large language model, which can further incorporate the sentiment information embedded in the market returns for training. Compared with the existing practice of using a dictionary and artificial sentiment labeling to train a large language model, our approach is more closely related to the financial market, which shows the logic of combining financial large language models with financial problems and the application of AI Plus Finance.

**Keywords:** textual sentiment; deep learning; large language model; asset pricing

**JEL Classification:** C53, G11, G12

《大语言模型、文本情绪与金融市场》附录

附录1 变量定义说明

附表1报告了本文主要变量的变量名称、变量说明、数据频率、数据时期和数据来源。

附表1 变量定义说明表

变量名	变量说明	数据时期	数据频率	数据来源
大模型情绪变量	使用中文金融大语言模型并通过公式计算得到的情绪变量。	2003~2021	月度	计算
金融字典法情绪变量	使用中文金融情绪字典并通过公式计算得到的情绪变量。	2003~2021	月度	计算
市场收益率	使用上证指数等权重收益率。	2003~2021	月度	CSMAR
宏观经济景气一致指数(Macro)	由国家统计局发布,反映当前中国宏观经济的基本运行形势。	2003~2021	月度	CSMAR
采购经理指数(PMI)	由国家统计局发布,可以衡量反映经济的变化趋势。	2005~2021	月度	CSMAR
消费者物价指数(CPI)	由国家统计局发布,反映消费品价格变动趋势。	2003~2021	月度	CSMAR
广义货币供应量(M2)	由中国人民银行发布,是指流通于银行体系之外的现金加上企业存款、居民储蓄存款以及其他存款,反映金融系统资金供应情况。(亿元)	2003~2021	月度	Wind
全社会融资规模存量同比增速(SF)	由中国人民银行发布,反映实体经济从金融体系获得的资金使用情况。(%)	2015~2021	月度	CSMAR
全国一般公共预算支出(F <sub>out</sub> )	由国家统计局发布,反映国家对一般公共预算收入的支出情况。(亿元)	2003~2021	月度	Wind
全国固定资产投资额(inv)	由国家统计局发布,反映完成建造和购置固定资产的货币工作量以及与此有关的费用。(亿元)	2003~2021	月度	Wind
全国房地产开发投资额(house)	由国家统计局发布,反映完成建造房地产的货币工作量以及与此有关的费用。(亿元)	2003~2021	月度	Wind
市场市盈率(PE)	反映股票市场整体市盈率。	2007~2021	月度	CSMAR
市场换手率(Turnover)	按流通股计算,反映市场整体换手率。	2003~2021	月度	CSMAR
市场平滑波动率(LVOL)	取对数计算,反映市场整体风险波动情况。	2003~2021	月度	计算

附录2 变量描述性统计说明

附表2报告了本文主要经济变量的描述性统计情况。

附表2 变量描述性统计表

变量名	样本量	均值	标准差	最小值	最大值
大模型情绪变量	228	0.000	1.000	-3.440	2.594
金融字典法情绪变量	228	0.000	1.000	-6.201	2.669
市场收益率	228	0.000	1.000	-3.237	3.723
Macro	228	99.530	4.797	76.390	117.590
PMI	204	51.590	2.765	35.700	59.200
CPI	228	102.515	1.882	98.200	108.700
M2	228	1030000	666000	190000	2380000
SF	73	12.393	1.683	10.000	16.411
F <sub>out</sub>	223	10589.807	7686.111	1097.100	37742.000
inv	209	185000	172000	1936.440	636000
house	209	38251.120	35652.943	598.260	148000
PE	180	34.675	13.443	17.911	83.553
Turnover	228	55.421	24.399	12.097	139.464
LVOL	228	-2.640	0.445	-3.779	-1.192

附录3 与市场交易价值指标对比

本文选取了文献中与情绪相关的市场交易指标以及能够显著预测市场回报的市场价值指标,并与媒体新闻文本情绪进行对比。文献表明许多市场交易指标能够反应投资者情绪变化,参照易志高和茅宁(2009),本文选择了基金折价率、IPO个数、IPO首日收益、新增投资者开户数与消费者信心指数6个指标,并使用主成分分析法构造了基于市场指标的投资者情绪指数。此外,大量市场价值指标所构成的技术指标也能够预测市场回报,本文选择了上一期收盘价、上一期换手率、移动平均指标、动量指标等几类指标。其中,移动平均指标是一种用于捕捉股价趋势的交易策略,当短期股价移动平均高于长期股价移动平均时为买入信号,否则为卖出信号。参照现实投资实践,本文基于1个月的时间窗口计算短期价格移动平均,基于12个月的时间窗口计算长期价格移动平均。动量指标是另一种常见的交易策略,当当期股价高于历史股价时为买入信号,否则为卖出信号。本文使用1个月前的股价作为历史股价,并与当期股价对比以计算动量指标。

随后,本文通过以下回归检验市场交易指标对市场回报的预测能力。

$$r_t^m = \alpha + \gamma MktVar_{t-1} + \varepsilon_t$$

其中, $r_t^m$ 为第 $t$ 期市值加权A股上证指数收益率, $MktVar_{t-1}$ 是第 $t-1$ 期的市场交易指标,具体变量展示如上。上式中 $\gamma$ 系数的大小和显著性体现了市场交易指标对市场回报的预测能力大小。

本文进一步通过如下回归对比市场交易指标和文本情绪指标对市场回报的预测能力。

$$r_t^m = \alpha + \gamma MktVar_{t-1} + \beta SENTIMENT_{i,t-1} + \varepsilon_{i,t}, i=1,2$$



其中,  $SENTIMENT_{i,t-1}$  是第  $t-1$  期第  $i$  个情绪指数,  $i=1, 2$ , 分别对应大模型情绪变量 ( $S_{lm}$ )、金融字典法情绪变量 ( $S_{dic}$ )。

附表 3 情绪指标与市场交易相关指标对市场情绪预测能力对比

市场交易相关指标	单变量回归	双变量回归 ( $S_{lm}$ )		双变量回归 ( $S_{dic}$ )	
	$\gamma$	$\gamma$	$\beta$	$\gamma$	$\beta$
基金折价率	-0.071 (-1.06)	-0.096 (-1.49)	0.272*** (4.15)	-0.069 (-1.07)	0.235*** (3.60)
交易量	0.152** (2.32)	0.102 (1.54)	0.232*** (3.53)	0.130** (2.01)	0.222*** (3.41)
IPO 个数	-0.079 (-1.18)	-0.049 (-0.75)	0.248*** (3.83)	-0.126* (-1.91)	0.258*** (3.92)
IPO 首日收益	-0.042 (-0.62)	-0.071 (-1.09)	0.271*** (4.11)	-0.088 (-1.33)	0.252*** (3.80)
新增投资者开户数	0.050 (0.75)	-0.016 (-0.25)	0.254*** (3.87)	0.041 (0.63)	0.233*** (3.58)
消费者信心指数	-0.033 (-0.50)	-0.073 (-1.12)	0.262*** (3.98)	-0.084 (-1.27)	0.252*** (3.79)
投资者情绪指数	-0.076 (-1.13)	0.047 (0.70)	0.262*** (4.00)	-0.124* (-1.88)	0.259*** (3.91)
收盘价	-0.099 (-1.48)	-0.134** (-2.08)	0.272*** (4.22)	-0.109* (-1.68)	0.239*** (3.69)
换手率	0.107 (1.61)	0.047 (0.70)	0.243*** (3.65)	0.090 (1.39)	0.228*** (3.50)
移动平均	0.138** (2.10)	0.125* (1.94)	0.247*** (3.86)	0.078 (1.16)	0.212*** (3.13)
动量	0.123* (1.86)	0.089 (1.38)	0.242*** (3.73)	0.078 (1.17)	0.219*** (3.29)

从实证结果上来看,交易量、移动平均、动量等市场交易价值指标的确对市场回报具有显著预测效果。在控制了市场价值指标后,基于媒体新闻的文本情绪指标对市场回报的预测系数仍然显著,与基准回归结果类似。因此,实证分析结果表明基于媒体新闻的文本情绪指标具有对市场回报的增量预测信息。

附录 4 稳健性分析结果

我们主要从以下方面进行稳健性分析:首先,为剔除噪音新闻,本文在数据清洗过程中删除了标题或内容短于 5 个字符的短新闻数据。为避免噪音新闻剔除标准设定对文章结果的影响,本文分别使用删除标题少于 4 字符或内容少于 10 字符的短新闻数据(后文记为标准 1)、删除标题少于 7 字符或内容短于 15 字符的短新闻数据(后文记为标准 2)重新清洗研究样本。此外,过短新闻不一定是噪音新闻,本文还使用未经剔除的全部新闻数据(后文记为不剔除)作为样本。附表 4 中相应结果显示,在使用上述数据清洗标准得到的新样本下,大模型情绪变量的样本外预测表现仍明显优于使用姜富伟等(2021)词典构建的金融字典法情绪变量,说明噪音新闻清洗标准不影响本文的主要结果。然后,本文还延长了研究样本以进一步论证本文结论的稳健性。从图 1 来看,2001 年与 2002 年新闻条数尽管少于后续年份,但仍显著多于 2000 年及之前样本。为此,本文在 2001 年 1 月~2021 年 12 月的延长样本期下重新检验了大模型情绪变量与金融字典法情绪变量的样本外预测能力,附表 4 中的相应结果显示中文金融大语言模型仍然具有更佳的样本外预测能力。最后,本文还考虑了媒体来源,一方面,本文从国家新闻出版署中国记者网官网获取官方处罚新闻媒体的信息,从总样本中删除来自上述被处罚媒体的数据,并剔除无具体来源新闻数据重新进行样本外检验(后文记为标准 1),另一方面,本文使用来源于 20 家权威媒体的新闻数据进行样本外检验(后文记为标准 2),比如上海证券报、证券日报、人民日报等。附表 4 相应结果表明中文金融大语言模型仍然具有更强的样本外预测能力,上述结果论证了本文结果并不会受到不良媒体、非官方权威媒体信息影响,本文结果仍然保持稳健。

附表 4 报告了本文稳健性分析部分的结果,上述结果与正文结果保持一致。

附表 4 稳健性分析结果

	变量	区间一		区间二		区间三	
		等权重	市值加权	等权重	市值加权	等权重	市值加权
替换噪音新闻 剔除标准	大模型情绪变量(标准 1)	9.65	6.71	9.02	7.47	15.92	10.09
	金融字典法情绪变量(标准 1)	3.94	5.64	5.13	6.41	0.26	1.52
	大模型情绪变量(标准 2)	9.57	6.65	8.87	7.32	15.47	9.60
	金融字典法情绪变量(标准 2)	3.79	5.52	5.08	6.39	0.30	1.56
	大模型情绪变量(不剔除)	10.10	7.28	10.52	9.30	8.15	5.15
替换样本区间	金融字典法情绪变量(不剔除)	3.76	5.21	4.76	5.67	-5.06	-6.19
	大模型情绪变量(延长期间)	8.36	6.37	8.01	6.26	5.70	3.75
	金融字典法情绪变量(延长期间)	1.59	2.54	2.65	3.93	-2.53	-2.71
考虑媒体来源	大模型情绪变量(标准 1)	9.90	6.93	9.27	7.73	16.26	10.63
	金融字典法情绪变量(标准 1)	3.77	5.58	5.21	6.67	0.65	2.38
	大模型情绪变量(标准 2)	9.37	6.53	8.96	7.57	15.60	10.25
	金融字典法情绪变量(标准 2)	3.96	5.59	5.12	6.56	0.06	1.43

注:单位:%。

## 附录5 个股情绪与个股回报预测

本文在该附录检验中文金融大语言模型所刻度的个股情绪能否预测个股回报。本文基于前文证券关联新闻数据集,分别应用前文得到的中文金融大语言模型输出证券关联新闻的情绪值来计算第 $t$ 月第 $i$ 公司关联新闻的大模型情绪变量(记作 $S_{it}^{lm}$ ),以及应用姜富伟等(2021)提出的金融情绪字典合成金融字典法情绪变量(记作 $S_{it}^{dic}$ )。参考葛兰和艾恩(2016)的做法,假设第 $t$ 月股票 $i$ 有 $M$ 个新闻,则当月个股新闻情绪变量 $S_{it}^k$ 的计算公式如下:

$$S_{it}^k = \frac{\sum_{m=1}^M I_{it}^m \times A_{it}^k}{M}, k = llm, dic$$

其中: $A_{it}^k$ 是根据正式文(1)应用中文金融大语言模型和金融情绪字典计算的每个证券关联新闻的情绪值, $I_{it}^m$ 是对第 $t$ 月第 $q$ 日发布新闻的时效权重,本文根据当月交易日天数均分为4个时效区间,按照时间由远及近设定4个时效区间上新闻的权重为1、2、3、4。 $I_{it}^m$ 时效权重意为根据当月该新闻发布时间所处当月4个不同时效区间进行时效加权,新闻发布时间越靠近月末(月初),对个股月末收益影响越大(越小),因此权重越大(越小),这样可以根据新闻时效性捕捉相关信息,从而提高新闻文本情绪变量对个股资产价格的预测效果。

在得到个股层面的情绪测度后,本文在此基础上构造做多做空投资组合,以此检验个股情绪是否能够捕捉股票的定价差异。在样本期内的每个月初,我们使用上月个股情绪变量将股票按照个股情绪变量的30和70分位数分为3个投资组合,第一个组合包括情绪变量最低的股票(小于新闻媒体情绪变量30分位数),第三个组合则包含最高的股票(大于新闻媒体情绪变量的70分位数)。随后,我们构造了多空组合,即买入高情绪组并卖出低情绪组。针对上述各投资组合,我们计算了市值加权平均超额收益率和等权平均超额收益率。此外,我们还使用法马-弗伦奇(1993)三因子模型(FF3)和刘-斯坦博-袁(2019)中国版三因子模型(CH3)检验情绪投资组合是否具备显著的不能被常见定价因子所解释的Alpha收益。检验公式如下:

$$r_{it} - r_{ft} = \alpha_{i,t}^{FF3} + \beta_{1,i,t} MKT_t + \beta_{2,i,t} SMB_t + \beta_{3,i,t} HML_t + \varepsilon_{i,t}$$

$$r_{it} - r_{ft} = \alpha_{i,t}^{CH3} + \beta_{1,i,t} CHMKT_t + \beta_{2,i,t} CHSMB_t + \beta_{3,i,t} CHVMC_t + \varepsilon_{i,t}$$

其中: $i$ 代表证券, $t$ 代表月份, $r_{it}$ 是股票收益率, $r_{ft}$ 是无风险利率, $MKT_t$ , $SMB_t$ 与 $HML_t$ 分别是法马-弗伦奇(1993)三因子模型中的市场因子、市值因子与价值因子, $CHMKT_t$ , $CHSMB_t$ 与 $CHVMC_t$ 分别是刘-斯坦博-袁(2019)中国版三因子模型中的市场因子、市值因子与价值因子。具体实证结果如附表5所示。

附表5的Panel A和Panel B分别报告了基于大模型情绪变量和金融字典法情绪变量的组合分析结果。从结果上来看,无论是等权重还是市值加权组合,基于两类情绪变量构造的投资组合的回报都随着情绪增大而出现单调递增,多空投资组合则有着显著为正的收益,表明文本情绪指标很好地捕捉了资产价格中的情绪定价信息。从多因子定价模型的检验结果来看,基于两种情绪指标构造的多空投资组合都有着显著区别于0的超额收益,表明情绪变量中含有不能被常见定价因子所解释的额外定价信息。整体上来看,上述结果表明情绪能够很好地捕捉到个股回报的横截面差异。此外,尽管两类情绪指标都能产生良好的投资表现,但从多空投资组合的回报大小和相应 $t$ 统计量大小上来看,大模型情绪变量有着更好的表现,体现出了本文所构造的中文金融大语言模型在刻度情绪上的优越性。

附表5 月度新闻个股情绪投资组合分析

分组	等权重			市值加权		
	RET-RF	$\alpha^{FF3}$	$\alpha^{CH3}$	RET-RF	$\alpha^{FF3}$	$\alpha^{CH3}$
Panel A: 大模型个股情绪						
1	0.0145* (1.8187)	0.0016 (1.4055)	0.0036*** (2.9971)	0.0056 (0.8513)	-0.0014 (-1.1746)	-0.0027** (-2.5190)
2	0.0252*** (3.0192)	0.0123*** (7.2594)	0.0150*** (9.9740)	0.0128* (1.7488)	0.0056*** (4.1022)	0.0033** (2.1635)
3	0.0369*** (3.8569)	0.0143*** (6.6269)	0.0188*** (9.0487)	0.0236*** (2.7227)	0.0090*** (4.6800)	0.0090*** (5.6726)
High-Low	0.0137*** (6.4263)	0.0128*** (7.2940)	0.0145*** (7.8721)	0.0124*** (5.2407)	0.0103*** (5.3595)	0.0121*** (6.2503)
Panel B: 金融字典法个股情绪						
1	0.0203** (2.3471)	0.0060*** (3.1431)	0.0102*** (4.9981)	0.0078 (1.0570)	-0.0014 (-0.9168)	-0.0001 (-0.0017)
2	0.0201** (2.4301)	0.0072*** (4.9502)	0.0094*** (6.9407)	0.0103 (1.4418)	0.0036** (2.4751)	0.0018 (1.5275)
3	0.0245*** (3.1398)	0.0124*** (8.5288)	0.0140*** (8.8496)	0.0124* (1.8337)	0.0053*** (4.3335)	0.0024* (1.8491)
High-Low	0.0042** (1.9949)	0.0064*** (3.4844)	0.0038* (1.6632)	0.0047** (2.4617)	0.0067*** (3.8922)	0.0024 (1.1475)

注:括号中为 $t$ 值,使用纽威和韦斯特(1987)进行调整,\*\*\*,\*\*, \*分别表示在1%,5%和10%水平上显著。

本文还进一步通过法马和麦克贝斯(1973)回归检验个股情绪对个股回报的解释能力。法马和麦克贝斯(1973)回归包含两个阶段。首先,我们将个股的超额回报与其相应的情绪值和公司特征进行逐月横截面回归。然后,我们检验情绪与其他变量系数的时间序列均值及其显著性。月度截面回归公式如下:

$$r_{it+1} - r_{ft+1} = \alpha_k + \beta_k S_{it}^k + \theta_k X_{it} + \varepsilon_{it+1}, k = llm, dic$$

其中, $X_{it}$ 为一组控制变量,包括市值、账面市值比、ROE、总资产增长率等。

附表6汇报了情绪变量与个股回报间的法马和麦克贝斯(1973)回归检验结果。第(1)列与第(3)列为单变量回归检验,两类情绪变量与个股回报之间均有显著为正的回归系数,与前文的组合分析结果相一致。第(2)列与第(4)列则分别汇报了引入控制变量后两类情绪变量与个股回报之间的关系,此时情绪变量仍然具有显著为正的回归系数。整体上来看,法马和麦克贝斯(1973)回归也表明新闻媒体个股情绪对个股回报也有良好的预测能力<sup>①</sup>。

附表6 月度新闻个股情绪回归分析				
变量	(1)	(2)	(3)	(4)
截距	1.412 <sup>*</sup> (1.725)	14.234 <sup>***</sup> (3.706)	2.047 <sup>**</sup> (2.394)	15.965 <sup>***</sup> (4.099)
大模型情绪变量	0.500 <sup>***</sup> (6.508)	0.397 <sup>***</sup> (6.562)		
金融字典法情绪变量			2.118 <sup>**</sup> (2.417)	3.477 <sup>***</sup> (4.498)
市值		-0.814 <sup>***</sup> (-3.722)		-0.902 <sup>***</sup> (-4.104)
账面市值比		-0.723 <sup>**</sup> (-2.210)		-0.025 <sup>**</sup> (-2.110)
ROE		-0.023 <sup>***</sup> (-2.211)		-0.659 <sup>*</sup> (-1.876)
总资产增长率		0.325 (1.580)		0.344 <sup>*</sup> (1.663)

注:括号中为*t*值,使用纽威和韦斯特(1987)进行调整,  
\*\*\*, \*\*, \*分别表示在1%, 5%和10%水平上显著。

注释

①中外文人名(机构名)对照:葛兰(Gulen);艾恩(Ion);法马(Fama);弗伦奇(French);刘(Liu);斯坦博(Stambaugh);袁(Yuan);纽威(Newey);韦斯特(West);麦克贝斯(MacBeth)。

参考文献

(1)姜富伟、孟令超、唐国豪:《媒体文本情绪与股票回报预测》,《经济学(季刊)》,2021年第4期。  
(2)易志高、茅宁:《中国股市投资者情绪测量研究:CICSI的构建》,《金融研究》,2009年第11期。  
(3)Fama, E. F. and French, K. R., 1993, “Common Risk Factors in the Returns on Stocks and Bonds”, *Journal of Financial Economics*, vol.33, pp.3~56.  
(4)Fama, E. F. and MacBeth, J. D., 1973, “Risk, Return, And Equilibrium: Empirical Tests”, *Journal of Political Economy*, vol.81, pp.607~636.  
(5)Gulen, H. and Ion, M., 2016, “Policy Uncertainty And Corporate Investment”, *Review of Financial Studies*, vol.29, pp.523~564.  
(6)Liu, J., Stambaugh, R. F. and Yuan, Y., 2019, “Size and Value in China”, *Journal of Financial Economics*, vol.134, pp.48~69.  
(7)Newey, W. K. and West, K. D., 1987, “A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation”, *Econometrica*, vol.55, pp.703~708.