【作品名称】**基于中文文本的投资者情绪因子有效性检验**

【关键字】异质因子、中文文本分类、BERT、投资者情绪

【作品简介】

本文旨在研究量化投资中异质因子的有效性。我们相信金融文本能够反映市场的动向，影响价格走势，从而具有预测未来价格的能力。行为金融学指出人具有一定的认知偏差，是有限理性的，因此散户投资人在股吧上发帖分析市场行情，或是评论他人操作，亦或是仅仅表达个人情绪，其言论能够反映价格及情绪的走势或者影响他人投资，进而对未来价格造成一定影响，因此本文将从中提取散户投资者情绪因子并检验其有效性。此外，机构分析师的分析与评级往往是投资风向标，且机构主导了大部分成交量，因此机构在对于市场的影响中占较大权重，本文将从分析师研报中提取机构投资者情绪因子。最后，本文将在Baker and Wurgler (2006) 所提出的市场情绪指数基础上对上述两类情绪因子进行相关性、有效性和稳健性检验，以剥离市场对情绪的影响。

关于因子挖掘和检验的步骤，首先，我们将从东方财富各大期货吧上使用爬虫获取发帖内容、发帖时间、发帖人、发帖人影响力、发帖人吧龄、发帖人关注数等原始数据，从wind上下载期货分时行情数据并通过相应数学变换构建因子。其次，我们将提取的发帖数据进行发帖人分类，采用机器学习模型将营销号、分析师在股吧上的发帖内容清洗出去，仅对散户投资者的发帖内容采用深度学习模型提取情绪值。然后，我们将在分钟、小时、日度等不同的时间频度上，采用事件研究法、时间序列回归、截面回归等方法检验因子有效性。

在进行初步的文献后，我们认为该作品的创新之处在于：（1）研究相对小众的文本类异质因子；（2）采用最前沿模型进行中文文本分词；（3）同时包括散户投资者/机构投资者与市场三种情绪因子，能够剥离三者对价格的预测力。

注：作品简介篇幅**不得超过A4纸1页**，可附图不超过1张。