

评估文本挖掘和预测技术分析 金融时间序列

阿里·拉什加里*

经济系
堪萨斯州立大学,曼哈顿,堪萨斯州
2023 年 1 月

抽象的

预测金融时间序列 (FTS)是金融和经济学中预测金融市场走势的重要领域。本文研究了文本挖掘和技术分析在预测金融时间序列方面的准确性。它重点关注大流行期间的标准普尔500股票市场指数,该指数跟踪美国最大的上市公司的表现。该研究比较了预测标准普尔 500 指数未来价格的两种方法:文本挖掘(使用 NLP 技术从财经新闻中提取有意义的见解)和技术分析(使用历史价格和交易量数据进行预测)。该研究探讨了这两种方法的优点和局限性,并分析了它们在预测 S&P500 方面的表现。FinBERT 模型在 S&P500 价格方面优于其他模型

预测(其较低的 RMSE 值证明了这一点)有可能彻底改变利用财经新闻数据进行的金融分析和预测。

关键词: ARIMA、BERT、FinBERT、预测金融时间序列、GARCH、LSTM、
技术分析、文本挖掘

JEL分类: G4、C8

*感谢Yoonjin Lee教授

介绍

预测是从地理到商业等各个学术领域的有用工具。

预测可以帮助地理学家预测天气模式、自然灾害和土地利用的变化,从而更好地进行城市规划和备灾[1]。它可用于预测城市地区的能源消耗模式 and 需求,从而为适当的照明计划或交通能源的设计提供信息[2-4]。

预测在数学和物理学中对于理解和预测复杂现象至关重要。预测在物理学中用于建模和预测量子力学和流体动力学等物理系统的行为。通过使光子带隙光纤的末端变细,研究人员创建了具有高传输效率和中等线中心精度的充气光子微电池。这些电池可以连接起来,尽管效率略有下降。

这一进步有可能用于监测和预测 PMC 性能随时间的变化,从而有可能提高激光微/纳米加工工艺的效率和准确性 [5-8]。

医生可以利用预测来改善 PCOS 和 T2DM 的预防和管理。预测可以根据生活方式、遗传和其他相关因素帮助预测个体患这些疾病的风险。然后,医生可以使用这些信息为患者制定个性化的预防和管理计划。例如,如果一个人患 T2DM 的风险较高,医生可能会建议改变特定的生活方式和/或服用药物来预防或延缓疾病的发作。同样,如果患者患有 PCOS,则可以使用预测来预测患者对各种治疗的反应,并帮助医生相应地制定治疗计划 [9-10]。

使用计算方法进行预测在工程中至关重要,因为它可以实现更好的决策和有效地利用资源。金融时间序列预测 (FTS) 是金融和经济学的一个重要领域,它可以预测金融市场的潜在风险、市场趋势和投资时机。由于金融环境的不确定性和嘈杂性,这个主题已经被广泛使用了几十年。为了分析和预测金融市场行为,通常使用基本面分析和技术分析。

在引入自然语言处理 (NLP) 模型之前,基本面和技术分析被广泛用于分析和预测金融市场行为。基本面分析需要检查财务指标和指标,以确定公司的财务健康状况和增长潜力,而技术分析则利用历史价格和数量数据来识别趋势并预测未来市场行为。两种方法都有优点和缺点,可以一起或单独使用,具体取决于投资者的目标和投资风格。

另一方面,预测金融市场行为很困难,并且受到经济指标、地缘政治事件和投资者情绪等多种因素的影响。因此,投资者在做出投资决策时,应考虑多种信息来源和分析 (图1)。

近年来大数据和人工智能的快速普及,导致金融数据量不断增加,数据之间的关联模式更加复杂,使得金融时间序列数据的预测变得更加困难。需要不受统计假设约束的非参数方法,因此机器学习算法已成为专家关注的焦点。文本挖掘利用自然语言处理技术对海量文本数据进行分析,已发展成为时间序列预测的有力工具,成为传统统计方法的补充。文本挖掘分析的主要好处是它能够快速

分析大量文本数据并识别数据中的趋势和模式,可用于识别文本数据中表达的情绪与股市走势之间的关系[11-18]。

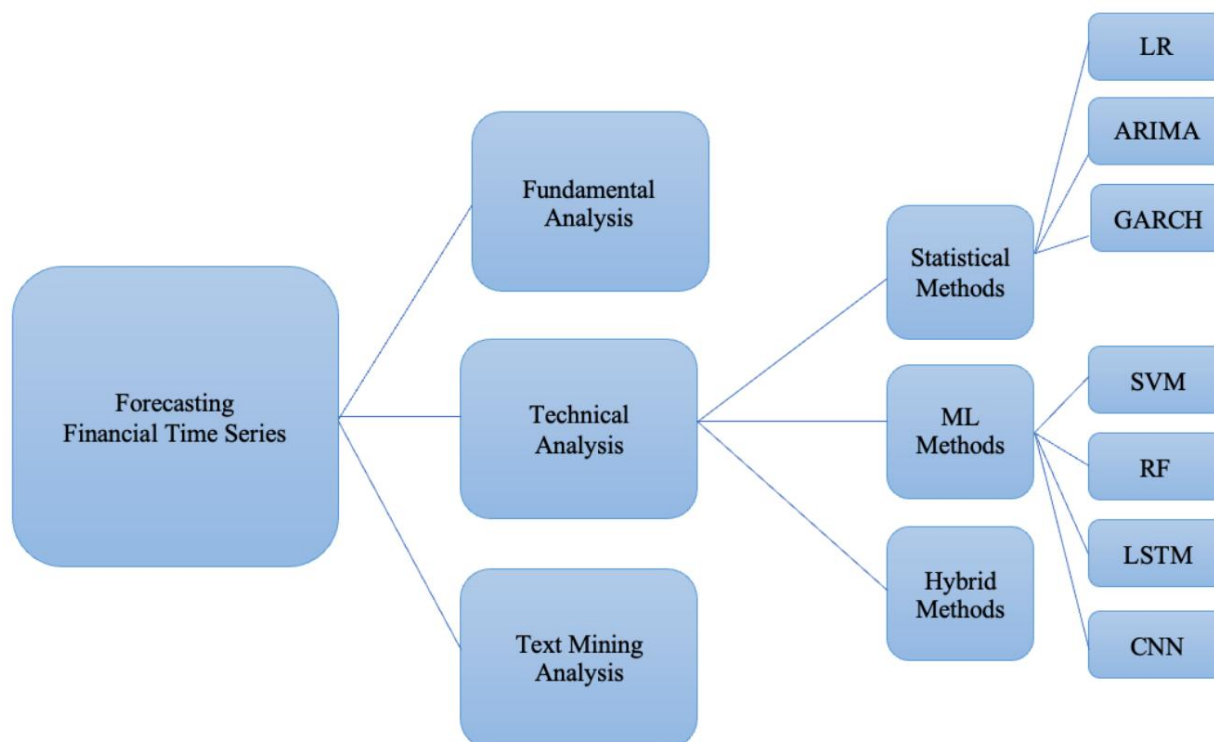


图 1:FTS

本文研究了使用财经新闻和历史价格数据预测 S&P500 的不同技术方法和文本挖掘的准确性。与之前的类似研究相比,所提出的文本挖掘方法显示出一些改进,结果可以帮助投资者在面对不可预测的事件时做出更好的决策。

通过比较这两种不同的方法,我们可以确定哪种方法在解决现实问题(历史价格或新闻)时更可靠。这种比较可以帮助投资者在预测不可预测事件的未来价值时决定使用哪种方法。

最终,价格还是新闻是否能更好地预测未来的问题是一个复杂的问题,答案可能取决于具体情况。然而,本文的研究结果可以为了解这些不同方法的优缺点提供有价值的见解,并帮助投资者做出更明智的决策。

以下是本文的结构。相关工作将在下面的部分中描述。拟议的方法、评估指标和数据集均在第 3 节中介绍。

实验设置和结果如第 4 节所示。第 5 节最终揭示了我们的结论和未来的工作。

文学评论

词嵌入是自然语言处理 (NLP) 中的一种技术,用于将单词表示为数值向量。词嵌入的目标是以可用作机器学习模型输入的方式捕获词的含义。

文本语料库中的每个单词都表示为高维向量,其中每个维度对应于该单词的特定语义或句法属性。向量的值是从训练数据中学习的,向量旨在捕获语料库中单词之间的关系。例如,语义相似的单词预计具有相似的向量,而语义不相似的单词预计具有不相似的向量。

自然语言处理 (NLP) 中常用的词嵌入技术有多种类型。一些最受欢迎的包括:

Word2Vec:由 Tomas Mikolov 等人于 2013 年发布。[19]

GloVe:由 Jeffrey Pennington 等人于 2014 年出版。[20]

FastText:由 Piotr Bojanowski 等人于 2016 年发布。[21]

ELMo:由 Matthew Peters 等人于 2018 年出版。[22]

BERT:由 Jacob Devlin 等人于 2018 年发布。[23]

GPT:由 Radford 等人于 2018 年发布。[24]

Atkins、Niranjan 和 Gerding (2018)的研究为金融新闻相对于资产或指数收盘价的预测能力提供了宝贵的见解。作者使用机器学习模型来分析新闻源并预测资产价格走势和资产波动性走势的方向。他们的研究表明,与资产或指数的收盘价相比,新闻衍生信息更能预测市场波动,平均方向预测准确度为 56 [12]。

Souma、Vodenska 和 Aoyama (2019)的研究探索了深度学习在财经新闻中进行情绪分析的用途,重点是根据新闻文章发布后的股价回报来定义极性。作者报告说,他们的方法结合了循环神经网络和长短期记忆单元,在分别选择正面和负面分数最高的新闻作为正面和负面新闻时,显示出更高的预测准确性。他们提出了未来研究的几种途径,包括探索定义极性的不同方法和使用不同的深度学习方法。

总的来说,这项研究的结果对金融预测领域具有重要意义,并展示了深度学习方法在增强情绪分析方面的潜力[16]。

郭和塔克菲尔德 (2020) 的论文研究了基于新闻的机器学习和深度学习方法在预测股指和个股方面的有效性。

该研究使用通过自然语言处理处理的新闻数据来比较机器学习和深度学习方法的性能。结果表明,深度学习在预测股指和个股方面均优于机器学习,股指提高了 4.5 个百分点,个股提高了 3 个百分点。

作者讨论了这些发现的含义,并提出了未来的研究领域,例如使用其他类型的数据和优化模型以获得更强的泛化能力。这项研究

强调了深度学习在预测股市趋势方面的潜力,并强调了数据处理和模型优化在获得准确结果方面的重要性[13]。

本研究使用 FinBERT (Araci 2019) 来提高文本挖掘方法的准确性。FinBERT 是由 Bloomberg 研究团队创建的预训练语言模型。它接受了大量金融文档的训练,例如新闻文章、SEC 文件和财报电话会议记录,以提高对金融术语、实体识别和情绪分析的理解 [25]。

这项研究的独特之处在于它考虑了视频新闻情绪,而这在以前的研究中很大程度上被忽略了。此外,比较以前的模型,例如统计方法和机器学习方法,看看哪种方法在现实问题中具有最高的准确性。

方法论和数据集

投资者不断寻求新的创新方法来预测股票和指数等金融资产的未來价格。标准普尔 500 指数是追踪美国 500 家最大上市公司业绩的股票市场指数,由于其对整体经济的影响而受到特别关注。

我们比较两种流行的预测标普500指数未来收益率的方法:首先是技术分析,包括自回归综合移动平均线 (ARIMA)、广义自回归条件异方差 (GARCH)等统计模型,以及机器学习方法,长短期记忆 (LSTM)。其次,文本挖掘分析,它利用 Transformers 的双向编码器表示 (BERT) 模型以及 FinBERT (专为金融文本分析而设计的 BERT 专用版本)。

技术分析和文本挖掘分析都使用不同的技术来分析和预测标普500指数的未来回报率。技术分析采用统计方法和机器学习模型来分析历史数据,而文本挖掘分析则使用自然语言处理和机器学习技术来分析财经新闻。

阿里玛

自回归综合移动平均线是用于时间序列分析和预测的统计模型类。它是计量经济学领域广泛使用的方法,用于建模和预测股票价格、GDP 和通货膨胀率等经济数据。ARIMA 模型假设时间序列数据可以描述为自回归 (AR)、移动平均 (MA) 和差分项的组合。ARIMA(p,d,q) 模型的基本数学方程:

$$y(t) = c + (1 - \alpha_1)y(t-1) + \dots + (1 - \alpha_p)y(t-p) - (1 - \beta_1)\epsilon(t-1) - \dots - (1 - \beta_q)\epsilon(t-q) + \epsilon(t)$$

在哪里:

- $y(t)$ 是时间序列在时间 t 的值

- c 是常数或截距项

- $(1 - \alpha_1)$ 至 $(1 - \alpha_p)$ 是模型的自回归 (AR) 系数,表示 $y(t)$ 与其前 p 个值 ($y(t-1)$ 至 $y(t-p)$) 之间的线性关系 - $(1 - \beta_1)$ 到 $(1 - \beta_q)$ 是模型的移动平均 (MA) 系数,表示误差之间的线性关系

项 ϵ_t 及其 q 个先前值 (ϵ_{t-1} 到 ϵ_{t-q})

- ϵ_t 是时间 t 时的误差项或白噪声

- d 是用于使时间序列平稳的差分程度,这对于 ARIMA 建模通常是必需的。

伽马奇

一类称为广义自回归条件异方差模型的统计模型用于分析和预测时间序列数据的波动性。这些模型广泛应用于金融、经济和其他理解和预测波动性至关重要的领域。

GARCH 模型背后的思想是时间序列的方差可以随时间变化,并且方差取决于该序列的先前值。GARCH 模型可以对序列随时间变化的波动性进行建模,这在许多现实世界的时间序列 (例如股票价格、汇率和商品价格)中非常重要。

基本的 GARCH 模型可以在数学上表示为:

$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2 + \dots + \alpha_p \epsilon_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \beta_2 \sigma_{t-2}^2 + \dots + \beta_q \sigma_{t-q}^2 \quad (2)$$

$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2 + \dots + \alpha_p \epsilon_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \beta_2 \sigma_{t-2}^2 + \dots + \beta_q \sigma_{t-q}^2 \quad (3)$$

$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2 + \dots + \alpha_p \epsilon_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \beta_2 \sigma_{t-2}^2 + \dots + \beta_q \sigma_{t-q}^2 \quad (4)$$

其中 ϵ_t 是在时间 t 观察到的回报, μ 是回报系列的平均值, σ_t^2 是误差项, σ_t 是时间 t 时的条件标准差, z_t 是均值为零、方差为 1 的标准化误差项。参数 α_0 、 α_1 和 β_1 是非负的, 分别反映波动性的持续性、过去平方误差对当前波动性的影响以及方差波动性的影响。

过去 σ_{t-1}^2 是有条件的 σ_t^2 在当前的 σ_t^2 的 GARCH 模型是使用最大似然估计来估计的, 其中涉及找到使观察数据的可能性最大化的模型参数值。

长短期记忆网络

长短期记忆 (LSTM) 模型是一种循环神经网络 (RNN), 可以分析和预测具有长期依赖性的序列数据。Hochreiter 和 Schmidhuber 于 1997 年推出了 LSTM, 从此成为自然语言处理、语音识别和时间序列分析的流行工具。

LSTM 能够长期选择性地记住和忘记信息, 这使得它们非常适合对具有复杂模式和依赖关系的数据序列进行建模和预测。LSTM 通过使用存储单元来实现这一点, 这使它们能够保留先前输入的长期记忆, 并根据新输入有选择地更新该记忆。

这使得 LSTM 在语言翻译、情感分析和股票价格预测等应用中特别有用, 在这些应用中, 理解和建模顺序数据中的复杂依赖关系至关重要。长短期记忆 (LSTM) 网络的数学模型可以表示如下

$$= (\sigma(x_t + \frac{1}{2} \tanh(x_t)) - 1 + \frac{1}{2}) (5)$$

$$= (\sigma(x_t + \frac{1}{2} \tanh(x_t)) - 1 + \frac{1}{2}) (6)$$

$$= (\sigma(x_t + \frac{1}{2} \tanh(x_t)) - 1 + \frac{1}{2}) (7)$$

$$= (\sigma(x_t + \frac{1}{2} \tanh(x_t)) - 1 + \frac{1}{2}) (8)$$

其中 h_t 表示 LSTM 在 t 时刻的隐藏状态, c_t 表示单元状态, x_t 表示 t 时刻的输入, y_t 代表 t 时刻的输出。函数 σ 和 g 分别表示 sigmoid 函数和激活函数, f 是忘记门函数。

LSTM 模型具有三个主要组成部分:输入门、遗忘门和输出门。

输入门控制进入细胞状态的信息流,而遗忘门则选择性地丢弃不再相关的信息。输出门控制从单元状态到隐藏状态的信息流。

LSTM 网络使用随时间反向传播 (BPTT) 进行训练,其中涉及在每个时间步计算损失函数相对于模型参数的梯度,并将其随时间向后传播。这使得网络能够学习顺序数据中的复杂模式和依赖性,使 LSTM 成为自然语言处理、语音识别和时间序列分析等广泛应用的强大工具。

伯特

Transformers 的双向编码器表示是一种深度学习模型,已通过 Google AI 语言进行预训练。它通过在文本分类、问答和文本生成等各种任务上实现尖端性能,改变了自然语言处理任务。与按顺序处理文本的传统语言模型不同, BERT 是一种基于转换器的模型,可以考虑前面和后面的单词来确定单词或句子的上下文。这使得 BERT 能够捕获复杂的语言现象,例如词义消歧和共指消解,从而显着提高准确性和性能。

BERT 模型架构基于 Transformer 架构,该架构由一系列编码器和解码器层组成,它们以基于注意力的方式并行处理输入序列。 BERT 使用双向 Transformer 编码器在大型文本数据语料库上预训练模型,然后使用掩码语言建模目标来预测句子中缺失的单词。预训练的 BERT 模型在微调过程中可以作为各种自然语言处理任务的起点,使其能够用相对较少的数据实现出色的性能。

BERT 是一个复杂的模型,涉及多层神经网络,但其核心使用基于 Transformer 的架构,可以用数学方式表示如下:

给定一个输入序列 X 的 token, BERT 首先应用嵌入层将每个 token 转换为向量表示。然后,嵌入的令牌被输入一系列变压器编码器层,这些编码器层以并行、基于注意力的方式处理序列。最终的输出

转换器层是一系列上下文感知的标记表示,可用于下游任务,例如文本分类或问题回答。

芬伯特

它是一种基于 BERT 的特定领域语言模型,专门用于分析金融文本数据。FinBERT 已经过大量金融文本数据的预训练,使其能够理解金融语言和术语的微妙之处。它已被用于从情绪分析到财务文档分类和财务预测的各种领域。FinBERT 在各种金融文本数据集上展示了尖端性能,使其成为金融分析师和研究人员的宝贵工具。

FinBERT模型架构与BERT模型架构类似,但在金融文本数据上进行了微调,以实现更高的准确性和性能。它对输入文本数据进行编码,并使用基于深度变压器的神经网络提取上下文表示。

之后,上下文表示可用于各种金融文本分析任务,例如情感分析或文档分类。FinBERT 可以显着减少金融文本分析所需的手动工作量,同时还提高结果的准确性。

数据和评估指标

在本研究中,我们将使用 Investing.com 的金融时间序列数据 (S&P500 价格),它是该指数的每日收盘价格。来自不同财经视频新闻 (CNBC、彭博社和雅虎财经) 的文本数据作为我们的主要数据集。该数据将包括2019年1月至2020年12月的两年财经视频新闻和市场数据,为我们在面对疫情时提供对市场趋势和行为的全面而有力的了解[11]。

均方根误差 (RMSE) 用于评估回归模型的准确性。它是数据集中预测值和实际值之间的平均平方差的平方根。下面是RMSE的计算公式:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}$$

在哪里

\hat{y}_i 是预测值,
 y_i 是实际值,并且
 n 是数据集中的观测值数量。

结果与讨论

本研究使用了六种不同的模型来预测 S&P500 的价格:ARIMA、GARCH、LSTM、BERT 和 FinBERT。数据集分为训练集和测试集,以确定每个模型的最佳配置。训练集包含前 75% 的数据,测试集包含剩余的 25%。对不同的模型参数和设置进行了实验,以优化并找到测试集上 RMSE 最低的配置。

比较每个模型的 RMSE 以评估其预测性能 [26-30]。

技术分析

在统计模型中,Python “统计模型”库用于使用三个常用标准找到返回数据的最佳 ARIMA 模型:赤池信息准则 (AIC)、贝叶斯信息准则 (BIC) 和均方误差 (MSE)。使用这些标准比较模型后,发现数据的最佳 ARIMA 模型具有 $p=4$ 、 $q=1$ 和 $d=0$ 。

这意味着数据的最佳拟合是由具有适量自回归和移动平均项但没有差分项的模型提供的。所选模型可用于预测和其他返回数据分析。使用测试数据,通过计算其均方根误差 (RMSE)来进一步评估所选的 $p=4$ 、 $q=1$ 和 $d=0$ 的 ARIMA 模型。最终的 RMSE 为 0.0103,表明该模型与数据拟合良好,能够提供准确的预测。

在 GARCH 模型选择中,Python “arch”库使估计和选择最佳 GARCH 模型变得简单。该库包含用于拟合 GARCH 模型、参数估计和预测生成的函数。它还具有计算 AIC、BIC 和 HQIC 标准的功能,可以更轻松地根据这些标准选择最佳模型。

当比较 GARCH(1,1) 和 ARIMA(4,0,1) 模型的 RMSE 值时,前者的 RMSE 值更高,为 0.5922,这表明 ARIMA 模型可能更适合该数据集。这可能是由于数据集的高度稳定 (图 2)性质,这可能限制了 GARCH 模型捕获波动模式的能力。最后,对于高度平稳的返回数据,ARIMA 模型在预测精度方面可能优于 GARCH 模型。

在机器学习模型中,LSTM 模型用于预测时间序列数据。Keras 库用于创建和训练模型,输入层由两个 LSTM 层和后面的密集输出层组成。之后,使用 “adam”优化器和 “mean”编译模型

平方误差损失函数。因此,该代码计算均方根误差 (RMSE) 以评估模型在测试集上的性能,RMSE 为 0.5646,表明预测准确性有希望。

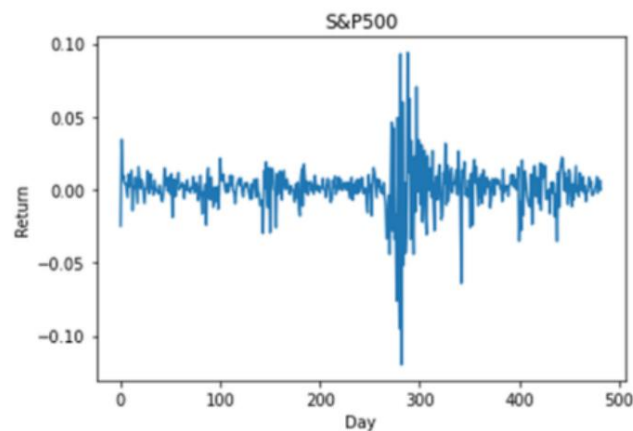


Figure 2: Daily return rate

该代码使用训练数据训练 LSTM 模型并保存训练历史记录。然后使用 matplotlib 使用历史记录绘制 100 个 epoch 过程中的训练和验证损失函数（图 3）。该图在 y 轴上显示损失值,在 x 轴上显示 epoch 数,训练和验证损失有单独的线。图上的图例将两条线分别标识为训练损失和测试损失。该图提供了模型在训练期间的性能的直观表示,可用于确定模型是否过度拟合或欠拟合数据。如果验证集的损失开始增加,而训练集的损失继续减少,这可能表明模型对训练数据过度拟合,并且不能很好地推广到新数据。

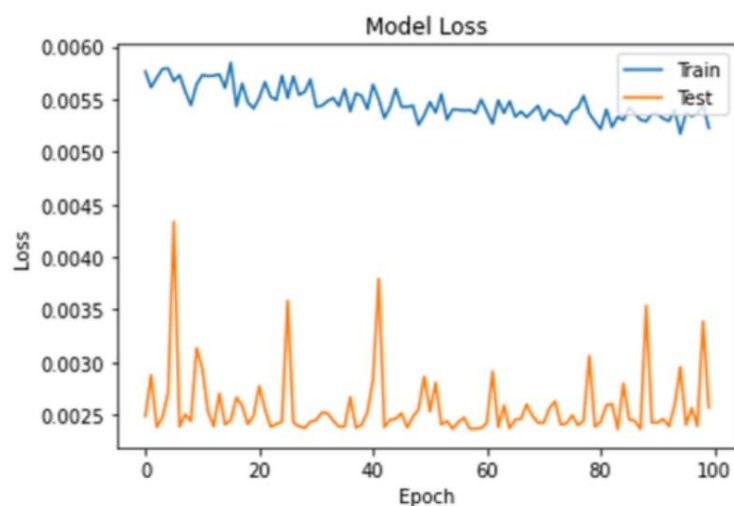


Figure 3: LSTM model

文本挖掘分析

BERT 转换器用于创建回归任务的深度学习模型。Hugging Face Transformers 库中的 AutoConfig 和 BertRegressor 函数分别用于配置和初始化模型。Adam 优化器以 LR 的学习率来优化模型参数,而均方误差损失函数用于计算预测输出和目标输出之间的损失。BERT 模型的 RMSE 值为 0.0120,表明该模型在预测目标变量方面已达到较高的准确度。然而,FinBERT 模型已经在大量金融文本语料库上进行了预训练,可以有效地学习金融数据中的潜在模式并做出准确的预测。

我们在金融新闻文章数据集上对 FinBERT 模型进行了微调,并获得了 0.010284 的 RMSE 值,这对于回归任务来说是一个很好的结果。FinBERT 模型与 BERT 模型的有效性可能会有所不同,具体取决于数据集和可用训练数据量。如果数据集相对较小,FinBERT 的性能可能会优于 BERT,因为它经过了金融文本的专门训练,因此可以更好地捕捉金融领域的特定领域语言和细微差别。因此,在为数据集选择预训练语言模型时对于特定任务,考虑任务的性质、数据集的特征以及可用模型的优点和局限性至关重要。为了确定哪个模型表现最好,可能需要试验不同的模型并针对特定任务对其进行微调。

结论和未来的工作

本研究的目标是使用文本挖掘和技术分析方法来预测标准普尔 500 指数的回报。我们使用了investing.com 和各种金融平台的数据集,其中包括两年的数据。使用词嵌入等技术,我们将文本数据转换为数值向量,然后将其输入回归模型。另一方面,技术分析涉及使用 ARIMA、GARCH 等传统方法以及 LSTM 模型等一些机器学习方法。仅使用一种指标来评估各种预测 S&P500 未来回报率的方法所做出的预测的准确性:均方根误差 (RMSE)。

我们发现 FinBERT 模型在预测 S&P500 价格方面优于 ARIMA 和 LSTM 模型,因为它的 RMSE 值较低。这表明 FinBERT 模型(而不是 ARIMA 或 LSTM 模型)可能是预测 S&P500 价格的更可靠工具。数据显示,在 COVID-19 大流行期间,财经新闻似乎比价格更有用。

最后,我们相信 FinBERT 模型有潜力改变金融分析和预测领域,带来新的见解和发现。然而,需要更多的研究来充分利用 FinBERT 模型的潜力并评估其在各种金融数据集上的有效性。

参考

1. He, Feng, Neda Mohamadzadeh, Mostafa Sadeghnejad, Ben Ingram 和 Yaser Ostovari.
“土壤颗粒的分形特征作为不同土地利用模式和坡向下土地退化的指标。”土地 12, 没有。 3 (2023) :1-14。
<https://doi.org/10.3390/land12030615>

2. 卡泽米德姆内, 玛丽亚姆。 “滨水城市地区的照明设计原则 (案例研究: 斯德哥尔摩皇家海港的 Kolkajen) 。” (2020) 。
<http://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1463604&dswid=3025>
URN: [瓮:nbn:se:kth:diva-280042](https://nbn-resolving.org/urn:nbn:se:kth:diva-280042)
OAI: oai:DiVA.org:kth-280042
DiVA, ID: diva2:1463604

3. Kazemidemneh, Maryam 和 Mohammadjavad Mahdavinejad。 “利用空间句法技术来提高照明质量并改变城市空间的能源消耗模式。”欧洲可持续发展杂志 7, no. 2 (2018) :29-29。 <https://doi.org/10.14207/ejsd.2018.v7n2p29>

4. Lashgari, Ali, Hasan Hosseinzadeh, Mohammad Khalilzadeh, Bahar Milani, Amin Ahmadisharaf 和 Shima Rashidi。
“基于元启发式算法的台湾交通能源需求预测。”能源, A 部分:回收、利用和环境影响 44, no. 2 (2022) :
2782-2800。 <https://doi.org/10.1080/15567036.2022.2062072>

5. 薛连杰, 刘松, 杨航, Adam M. Summers, Derrek J. Wilson, Xinya Wang, Pingping Chen 等。 “揭示六方氮化硼中的超快光电离。” arXiv 电子印刷品 (2021):arXiv-2101。

[arXiv:2101.10429 \[物理.光学\] 或 arXiv:2101.10429v2 \[物理.光学\]](https://arxiv.org/abs/2101.10429)
<https://doi.org/10.48550/arXiv.2101.10429>

6. Hosseini-Zavareh, Sajed, Ryan Luder, Manasadevi Thirugnanasambandam, HW Kushan Weerasinghe, Brian R. Washburn 和 Kristan L. Corwin。 “短乙炔填充光子微电池的制造和表征。”应用光学 58, no. 11 (2019) :
2809-2816。

<https://doi.org/10.1364/AO.58.002809>

7. Luder, Ryan J., Sajed Hosseini-Zavareh, Chenchen Wang, Manasadevi Thirug-nanasambandam, Brian R. Washburn 和 Kristan L. Corwin。 “短乙炔填充光子带隙光纤电池符合实用的行业标准。”载于《CLEO:科学与创新》, 第 SM2H-6 页。光学出版集团, 2016。

https://doi.org/10.1364/CLEO_SI.2016.SM2H.6

8. 萨吉德,侯赛尼-扎瓦雷。 “充满乙炔的压力拓宽了短光子微电池。”博士讨论。 ,2018。
<https://krex.kstate.edu/bitstream/handle/2097/39423/SajedHosseiniZavareh2019.pdf?sequence=7>
9. Noormohammadi,Morvarid,Ghazaleh Eslamian,Shirin Malek,Nargeskhatoun Shoaibinobarian 和 Seyedeh Nooshan Mirmohammadali。 “生育饮食评分与多囊卵巢综合征之间的关联:病例对照研究。”国际妇女保健 43,第 1 期。 1-3 (2022) : 70-84。 <https://doi.org/10.1080/07399332.2021.1886298>
-
10. Jafari,Naser,Nargeskhatoun Shoaibinobarian,Azadeh Dehghani,Amirhosein Rad,Seyedeh Nooshan Mirmohammadali, Mohammad Javad Alaeian,Yasaman Hamed,Mohamad Zamani,Mohammad Ali Goudarzi 和 Omid Asbaghi。 “食用马齿苋对血糖控制氧化的影响系统评价和剂量反应荟萃分析。” (2023) 。
在 和 压力:
<https://doi.org/10.1002/fsn3.3311>
11. Alzazah,法腾、程晓春和高晓红。 “根据财经视频新闻网站的情绪预测市场走势。” 2022 年 IEEE 第 16 届语义计算国际会议 (ICSC),103-10。美国加利福尼亚州拉古纳山:IEEE,2022。doi :10.1109/ICSC52841.2022.00022。
-
- 12.阿特金斯、亚当、马赫桑·尼兰詹和恩里科·格尔丁。 “财经新闻比收盘价更能预测股市波动。”金融与数据科学杂志 4,no. 2 (2018 年 6 月) :120-37。 <https://doi.org/10.1016/j.jfds.2018.02.002>
-
13. 郭俊杰和布拉德福德·塔克菲尔德。 “基于新闻的机器学习和深度学习方法进行股票预测。”物理学杂志:会议系列 1642,编号。 1 (2020 年 9 月 1 日) :012014。
<https://doi.org/10.1088/1742-6596/1642/1/012014>
14. Kumbure,Mahinda Mailagaha,Christoph Lohrmann,Pasi Luukka 和 Jari Porras。 “股票市场预测的机器学习技术和数据:文献综述。”专家系统与应用程序 197 (2022 年 7 月) :116659。
[DOI:10.1016/j.eswa.2022.116659](https://doi.org/10.1016/j.eswa.2022.116659)
15. Lin,Wei-Chao,Chih-Fong Tsai 和 Hsuan Chen。 “影响基于文本挖掘的股票预测的因素:文本特征表示、机器学习模型和新闻平台。”应用软计算 130 (2022 年 11 月) :109673。<https://doi.org/10.1016/j.asoc.2022.109673>
-
16. Souma,Wataru,Irena Vodenska 和 Hideaki Aoyama。 “使用深度学习方法增强新闻情绪分析。”计算社会科学杂志2,no. 1 (2019 年 1 月) :33-46。 <https://doi.org/10.1007/s42001-019-00035-x>

<https://doi.org/10.1007/s42001-019-00035-x>

17. 唐亚娇、宋振宇、朱玉林、袁怀玉、侯茂章、季俊凯、唐成和李建强。“金融时间序列预测机器学习模型的调查。”神经计算 512 (2022 年 11 月) :363-80。 <https://doi.org/10.1016/j.neucom.2022.09.003>

-
18. Pour,Ehsan Sadeghi,Hossein Jafari,Ali Lashgari,Elaheh Rabiee 和 Amin Ahmadisharaf。
“利用 LSTM 和贝叶斯优化神经网络进行加密货币价格预测。”欧洲商业与管理研究杂志 7,no. 2 (2022) :
20-27。 <https://doi.org/10.24018/ejbmr.2022.7.2.1307>

-
19. Mikolov, T.,Sutskever, I.,Chen, K.,Corrado, GS,Dean, J.:单词和短语及其组合性的分布式表示。见:Burges, CJC,Bottou, L.,Welling, M.,Ghahramani, Z.,Weinberger, KQ (编辑)《神经信息处理系统进展》26,第 3111-3119 页。Curran Associates, Inc. (2013) <https://doi.org/10.48550/arXiv.1310.4546>

-
20. 杰弗里·彭宁顿、理查德·索切尔和克里斯托弗·D·曼宁。“Glove:用于单词表示的全局向量。”2014 年自然语言处理经验方法 (EMNLP) 会议记录,第 1532-1543 页。2014.10.3115 /v1/D14-1162

-
21. Joulin,Armand,Edouard Grave,Piotr Bojanowski,Matthijs Douze,Herve Jegou 和 Tomas Mikolov。“快速文本。zip:压缩文本分类模型。”arXiv 预印本 arXiv:1612.03651 (2016)。

<https://doi.org/10.48550/arXiv.1612.03651>

-
22. 彼得斯、马修·E、塞巴斯蒂安·鲁德和诺亚·A·史密斯。“调还是不调?使预训练的表示适应不同的任务。”arXiv 预印本 arXiv:1903.05987 (2019)。 <https://doi.org/10.48550/arXiv.1903.05987>

-
23. Devlin,Jacob,Ming-Wei Chang,Kenton Lee 和 Kristina Toutanova。“Bert:用于语言理解的深度双向转换器的预训练。”arXiv 预印本 arXiv:1810.04805 (2018)。

<https://doi.org/10.48550/arXiv.1810.04805>

-
24. 雷德福、亚历克·卡蒂克·纳拉辛汉、蒂姆·萨利曼和伊利亚·苏茨克弗。“改善语言通过生成预训练来理解。”(2018)。 https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf

-
- 25.阿拉奇,多古。“Finbert:使用预先训练的语言模型进行金融情绪分析。”arXiv 预印本 arXiv:1908.10063 (2019)。 <https://doi.org/10.48550/arXiv.1908.10063>

26. Noroozi, Mohammad, Hamed Mohammadi, Emad Efatinasab, Ali Lashgari, Mahdijeh Eslami 和 Baseem Khan。 “黄金搜索优化算法。” IEEE Access 10 (2022): 37515-37532。
[10.1109/访问.2022.3162853](https://doi.org/10.1109/ACCESS.2022.3162853)
-
27. Hajipour, Vahid, Seyed Taghi Akhavan Niaki, Majid Akhgar 和 Mehdi Ansari。 “具有可追溯性的医疗保健供应链网络设计:一种新颖的算法。” 计算机工业工程 161 (2021): 107661。 <https://doi.org/10.1016/j.cie.2021.107661>
-
28. 安萨里、迈赫迪、胡安·S·博雷罗和莱昂纳多·洛萨诺。 “多重涟漪效应中断下的鲁棒最小成本流问题。” INFORMS 计算杂志, 在线发布: 2022 年 10 月 17 日。
<https://doi.org/10.1287/ijoc.2022.1243>
-
29. R. Nematirad 和 A. Pahwa, “考虑特征选择的使用人工神经网络进行太阳辐射预测”, 2022 年 IEEE 堪萨斯电力和能源会议 (KPEC), 美国堪萨斯州曼哈顿, 2022 年, 第 1-4 页 [10.1109/KPEC54747.2022.9814765](https://doi.org/10.1109/KPEC54747.2022.9814765)
-
30. Goharipour, Hamed 和 Seyyed Soroosh Firoozabadi。 “评估货币汇率变化对德黑兰空气污染的影响: 部门审查。” 欧洲商业与管理研究杂志 7, 编号。 3 (2022) :12-19。 <https://doi.org/10.24018/ejbmr.2022.7.3.1411>
-