使用卷积神经网络和 Transformer 进行金融时间序列预测

曾臻、拉奇尼特·考尔、苏切塔·西达甘加帕、萨巴·拉希米、
塔克·巴尔奇、曼努埃拉·维洛索
摩根大通人工智能研究，纽约，纽约州，美国

arXiv:2304.04912v1 [cs.LG] 11 Apr 2023

摘要

时间序列预测在各个领域的决策过程中都十分重要。尤其是股票价格等金融时间序列，由于难以对数据点之间的短期和长期时序依赖关系进行建模，因此很难预测。卷积神经网络（CNN）善于捕捉局部模式，适用于建模短期依赖关系。然而，由于感受野有限，卷积神经网络无法学习长期依赖关系。另一方面，Transformer 能够学习全局上下文和长期依赖关系。本文提出结合卷积神经网络和 Transformer 的优势，对时间序列中的短期和长期依赖关系进行建模，并预测未来价格会上涨、下跌还是保持不变（持平）。实验结果表明，与常用的统计方法和深度学习方法相比，所提出的方法在预测标普 500 指数成分股的日内价格变化方面表现出色。

引言

时间序列预测具有挑战性，尤其是在金融行业（Pedersen 2019）。这需要通过统计方法理解历史数据中复杂的线性和非线性交互，以预测未来。在金融行业，预测的常见应用包括预测在市场上交易的公司股票价格会上涨还是下跌，或呈现正向还是负向的价格变化。传统的统计方法通常采用线性回归、指数平滑法（Holt 2004；Winters 1960；Gardner Jr 和 McKenzie 1985）以及自回归模型（Makridakis、Spiliotis 和 Assimakopoulos 2020）。随着深度学习的发展，近期的研究大量投入于集成模型和序列到序列建模，例如循环神经网络（RNN）和长短期记忆网络（LSTM）（Hochreiter 和 Schmidhuber 1997）。在计算机视觉领域，卷积神经网络（卷积神经网络）（Ren 等，2015；Ronneberger、Fischer 和 Brox，2015）在学习局部模式方面表现突出，适用于对短期依赖关系建模，但由于感受野有限，不适合对长期依赖关系建模。最近，Transformer（Vaswani 等，2017）在自然语言……中取得了巨大成功。

版权所有 ©2023，美国人工智能促进协会（www.aaai.org）。保留所有权利。

ural 语言处理（NLP）（Devlin 等，2018；Brown 等，2020；Smith 等，2022）领域，与 LSTM 相比，在长期依赖关系建模方面取得了更优异的性能。

我们的贡献如下：如图 1 所示，我们利用卷积神经网络和 Transformer 的优势，对金融时间序列中的短期和长期依赖关系进行建模。在实验中，我们展示了所提出方法在标普 500 成分股日内股价预测方面的优势，其表现优于包括自回归积分滑动平均（ARIMA）和指数移动平均（EMA）在内的统计方法，以及一种先进的基于深度学习的自回归模型 DeepAR（Salinas 等，2020）。

相关工作

时间序列预测

文献中的典型预测技术利用统计工具，例如指数平滑法（ETS）（Holt 2004；Winters 1960；Gardner Jr and McKenzie 1985）和自回归积分滑动平均（ARIMA）（Makridakis, Spiliotis, and Assimakopoulos 2020），对数值时间序列数据进行一步超前预测。随后将这些预测递归地馈入未来输入，以获得多步预测。多时域预测方法，例如（Taieb, Sorjamaa, and Bontempi 2010；Marcellino, Stock, and Watson 2006），则直接同时生成对多个预先定义的未来时间步的预测。

基于机器学习和深度学习的方法机器学习（ML）方法已表明，通过以无模型的方式处理模型中的高维和非线性特征交互，可以提升性能。这些方法包括基于树的算法、集成方法、神经网络、自回归以及循环神经网络（Hastie、Tibshirani 和 Friedman，2001）。近期更多研究将深度学习（DL）方法应用于数值时间序列数据（Bao、Yue 和 Rao，2017；Gensler 等，2016；Romeu 等，2015；Sagheer 和 Kotb，2019年；Sutskever、Vinyals 和 Le，2014）。深度学习能够自动完成特征提取过程，从而消除对领域专业知识的需求。

自 Transformer（Vaswani 等，2017）问世以来，它们已成为提升 NLP 应用性能的最先进模型，旨在改

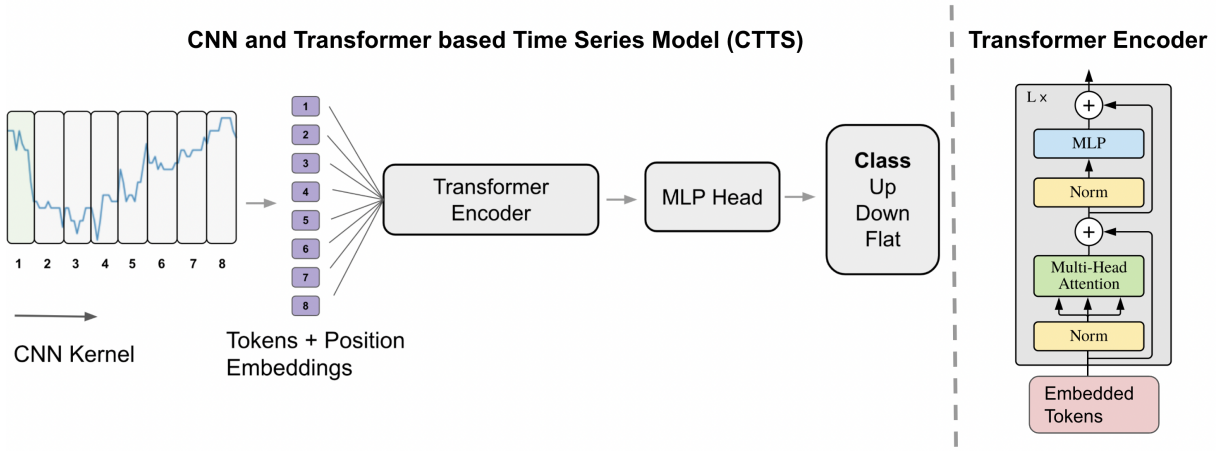


图1：所提出方法的概览。右侧为 Transformer 编码器架构的快速浏览（Dosovitskiy 等，2020）

善 NLP 应用的性能。常用方法是在大型数据集上进行预训练，然后在较小的特定任务数据集上微调（Devlin 等，2018）。Transformer 利用多头自注意力机制，取代了编码器-解码器架构中最常用的循环层。与按序处理输入数据的 RNN 和 LSTM 不同，Transformer 绕过循环结构，一次性接收所有输入；因此，Transformer 支持并行计算，从而缩短训练时间，也不会受到长期记忆依赖问题的影响。

本文其余部分组织如下。我们将讨论所提出的时间序列建模方法。此外，我们还将介绍基准模型、性能评估指标，并报告实验结果。最后，我们将总结本文并探讨未来的研究方向。

方法

我们提出的方法称为基于**C**NN 和**T**ransfo**r**mer 的**t**ime **s**eries 建模（CTTS），如概览图 1 所示。

预处理

我们使用标准最小-最大缩放，对 $[0, 1]$ 中的每个输入时间序列进行标准化。给定原始股票价格时间序列$\mathbf{x}$，标准化后的时间序列计算如下。

$$\mathbf{x}_{standardized} = \frac{\mathbf{x} - \min(\mathbf{x})}{\max(\mathbf{x}) - \min(\mathbf{x})}$$

这$\mathbf{x}_{standardized}$ 随后被传入我们的模型，以学习紧接着的下一个时间步中变化的符号，该符号被定义为我们的预测目标。

预测

如图 1 所示，我们使用一维卷积神经网络卷积核沿时间序列进行卷积，将每个局部窗口映射为一个嵌入向量，我们称之为 token。每个 token 都包含时间序列的短期模式。然后，我们为 token 添加位置嵌入（Vaswani 等，2017），并将其输入 Transformer 模型，以学习这些 token 之间的长期依赖关系。Transformer 模型输出时间序列的潜在嵌入向量，随后将其输入一个末端带有 softmax 激活函数的多层感知机（MLP），以生成方向分类结果。输出为各类别（上涨、下跌、持平）的概率，其中 3 个类别的概率之和为 1。

实验

在我们的实验中，我们将CTTS——我们提出的方法——与4种方法进行了基准比较。1) DeepAR——一种最先进的基于自回归循环神经网络的时间序列预测方法；2) 自回归积分滑动平均（ARIMA）；3) 指数移动平均（EMA）；以及4)朴素的恒定类别预测。我们使用符号预测准确率和符号预测准确率的阈值版本（稍后讨论），利用模型预测概率对性能进行了基准评估。我们在一台配备 8 个 16GB NVIDIA T4 Tensor CoreGPU 的Linux 机器上运行了实验，并使用 Python 3.6 中的PyTorch v1.0.0深度学习平台。在所有模型中，我们都设置了固定的随机种子，以获得可复现的结果。

实验设置

数据我们使用了通过授权的彭博数据服务（blo 2022）获取的标准普尔 500 指数成分股日内股价。数据以 1 分钟为间隔采样，涵盖 2019年（52 周，每周有 5 个交易日）。对于每只股票，我们每周采样 7 条时间序列。第一至第三季度（第 1 至 39 周）的数据用于训练和验证。数据经过随机划分，其中 80% 用于训练，其余 20% 用于验证。

方法	二分类 ↑	二分类* ↑	三分类 ↑	三分类* ↑
EMA	53.2%	59.9%	39.5%	41.7%
ARIMA	50.9%	51.8%	37.5%	38.4%
DeepAR	51.1%	53.6%	37.4%	38.7%
CTTS	**56.7%**	**66.8%**	**44.1%**	**55.2%**

表 1：2019年最后一个季度的方向准确率汇总。我们提出的方法 CTTS 优于基线方法 DeepAR、ARIMA 和 EMA。2/三分类* 指经过阈值处理的方向准确率。真实值测试集中的符号分布如下：价格上涨| 下跌 | 保持平稳：37.1% | 36.5% | 26.4%。如果对所有时间序列都简单地预测多数类（上涨），方向准确率仅为 37.1%。

我们约有 507K 个训练样本和 117K 个验证样本。使用最后一个季度（第 40 至第 52 周）的数据进行测试，共计 209K 条时间序列。对于每条时间序列，使用前 80 个时间步（输入）来预测第 81 个时间步（目标）的价格变化方向。整体测试集性能以测试集上评估指标的汇总值表示。

CTTS 在我们的实验中，我们使用交叉熵损失和 AdamW（ Loshchilov and Hutter 2017）优化器，批量大小为 64，最大训练轮数为 100。我们使用的卷积神经网络卷积核大小为 16，步幅为 8。Transformer 的深度为 4，包含 4 个自注意力头，嵌入维度为 128，并采用 0.3 的丢弃率来防止过拟合。

基线方法

DeepAR DeepAR 预测算法是一种监督学习算法，它使用自回归循环神经网络预测一维时间序列。DeepAR 的优势在于可以在所有时间序列上联合训练单个模型。我们使用了 128 的批量大小、Adam 优化器和正态分布损失。模型最多训练 300 轮，并设置提前停止机制的耐心值为 15。基础学习率为 1e-3，并通过学习率调度器进行调整，衰减因子为 0.1，耐心值为 5。我们还使用了概率为 0.1 的丢弃。DeepAR 为每个时间序列生成多个（200 个）预测目标样本，并将三类的预测概率定义为每一类预测样本所占的比例。

ARIMA 自回归积分滑动平均（ARIMA）模型通过结合自回归模型、滑动平均模型和差分来捕捉数据中的自相关（Wilks 2011）。我们将连续值 ARIMA 预测值与输入数据中的最后一个已知价格进行比较，以确定预测符号，并将相应的预测概率定义为预测价格与最后一个已知价格之间绝对差值相对于过去 80 个数据点的标准差的百分比，上限为 1。

EMA 指数移动平均（EMA）是一种移动平均方法，它赋予最新的数据点更大的权重和重要性。我们使用了“估计”初始化方法，将初始值视为参数，并选择使误差平方和最小的值。与 ARIMA 类似，我们使用预测价格与最近已知价格之间的差值来定义预测方向及相应的预测概率。

恒定类别预测我们测试了三种朴素基线方法，即始终预测一个恒定的方向：预测价格始终上涨、下跌或保持不变。对于这些方法，预测概率被设置为获胜类别为 1，其余两个类别为 0。

指标

我们在三分类 [类别 1—价格上涨，类别 2—价格下跌，类别 3—价格保持不变] 和二分类 [类别 1—价格上涨或保持不变，类别 2—价格下跌]上评估了所有模型。我们使用所有测试样本的平均方向预测准确率来评估模型。准确率越高，预测就越准确。

此外，我们评估了经过阈值筛选的方向准确率版本，其中阈值定义为所有主导类别预测概率的第 75^{th} 百分位数。我们仅保留超过阈值的样本，并在这些保留样本上计算方向预测准确率。

结果与讨论

我们在表 1 中汇总了方向准确率的定量基准测试结果。如上所述，我们展示了二分类和三分类任务的方向准确率。请注意，在二分类任务中随机猜测的准确率为 50%，而在三分类任务中随机猜测的准确率为 33%。如表 1 的二分类和三分类列所示，CTTS 在这两种情况下均优于所有基线方法。这表明，将卷积神经网络与 Transformer 相结合有助于时间序列预测。

此外，我们评估了经过阈值筛选的方向准确率版本，其中阈值定义为所有主导类别预测概率的第 75 百分位数。我们仅保留超过阈值的样本，并在这些保留样本上计算方向预测准确率。如表 1 的二分类* 和三分类* 列所示，对预测概率进行阈值筛选后，所有方法的准确率均有所提高。此外，我们提出的方法 CTTS 的准确率提升幅度最大。这表明 CTTS 输出的类别概率具有可靠性。具体而言，CTTS 的高置信度预测通常是正确的，因此阈值筛选会滤除置信度较低的错误预测，从而提高方向准确率。

另一个亮点是，与其他基线方法相比，CTTS 在阈值化三分类* 准确率方面取得了显著优势。未来可以将这一优势用于交易决策。例如，当预测类别分别为上涨、下跌或持平时，可以直接决定买入、卖出或持有股票。如前文所述，CTTS 的预测概率具有可靠性，因此买入、卖出或持有的股票数量可以取决于】【。

预测概率越高，可考虑买入的股票数量就越多。

结论

本文探讨了金融领域中股票价格时间序列预测这一具有挑战性的问题。我们展示了卷积神经网络和 Transformer 结合建模时间序列中短期与长期依赖关系的能力。在针对 2019 年标普 500 成分股日内股价进行的实验中，我们证明了所提出的 CTTS 方法相较于 ARIMA、EMA 和 DeepAR 的优势，并展现了该方法未来应用于下游交易决策的潜力。

参考文献

2022年。彭博市场数据。https://www.bloomberg.com/professional/product/market-data/。访问日期：2022-08-16。Bao, W.; Yue, J.; and Rao, Y. 2017. 使用堆叠式自编码器和长短期记忆网络进行金融时间序列预测的深度学习框架。《*PloS one*》，12(7): e0180944。Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J. D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; 等。2020. 语言模型是少样本学习者。《神经信息处理系统进展》，33: 1877–1901。Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2018. Bert：用于语言理解的深度双向 Transformer 预训练。*arXiv* 预印本 *arXiv:1810.04805*。Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; 等。2020. 一张图像胜过 16×16 个词：用于大规模图像识别的 Transformer。*arXiv* 预印本 *arXiv:2010.11929*。Gardner Jr, E. S.; and McKenzie, E. 1985. 时间序列趋势预测。《管理科学》，31(10): 1237–1246。Gensler, A.; Henze, J.; Sick, B.; and Raabe, N. 2016. 用于太阳能发电预测的深度学习——一种采用自编码器和 LSTM 神经网络的方法。载于*2016 年 IEEE* 国际系统、人类与控制论会议（*SMC*），002858–002865。IEEE。Hastie, T.; Tibshirani, R.; and Friedman, J. 2001. 统计学习要素。Springer 统计学系列。美国纽约州纽约市。Hochreiter, S.; and Schmidhuber, J. 1997. 长短期记忆。《神经计算》，9(8): 1735–1780。Holt, C. C. 2004. 利用指数加权移动平均预测季节性和趋势。《国际预测期刊》，20(1): 5–10。Loshchilov, I.; and Hutter, F. 2017. 解耦权重衰减正则化。*arXiv* 预印本 *arXiv:1711.05101*。Makridakis, S.; Spiliotis, E.; and Assimakopoulos, V. 2020. M4 竞赛：10 万个时间序列和 61 种预测方法。《国际预测期刊》，36(1): 54–74。Marcellino, M.; Stock, J. H.; and Watson, M. W. 2006. 用于预测宏观经济时间序列的直接与迭代多步 AR 方法比较。《计量经济学杂志》，135(1-2): 499–526。Pedersen, L. H. 2019. 《高效的低效：聪明资金如何投资以及市场价格如何形成》。普林斯顿大学出版社。Ren, S.; He, K.; Girshick, R.; and Sun, J. 2015. Faster R-CNN：迈向基于区域提议网络的实时目标检测。《神经信息处理系统进展》，28。Romeu, P.; Zamora-Martínez, F.; Botella-Rocamora, P.; and Pardo, J. 2015. 用于短期时间序列预测的堆叠式去噪自编码器。载于《人工神经网络》，463–486。Springer。Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-Net：用于生物医学图像分割的卷积网络。载于《医学图像计算与计算机辅助干预国际会议》，234–241。Springer。Sagheer, A.; and Kotb, M. 2019. 使用深度 LSTM 循环网络预测石油产量时间序列。《神经计算》，323: 203–213。Salinas, D.; Flunkert, V.; Gasthaus, J.; and Januschowski, T. 2020. DeepAR：使用自回归循环网络进行概率预测。《国际预测期刊》，36(3): 1181–1191。Smith, S.; Patwary, M.; Norick, B.; LeGresley, P.; Rajbhandari, S.; Casper, J.; Liu, Z.; Prabhumoye, S.; Zerveas, G.; Korthikanti, V.; 等。2022. 使用 DeepSpeed 和 Megatron 训练 Megatron-Turing NLG 530B：大型生成式语言模型。*arXiv* 预印本 *arXiv:2201.11990*。Sutskever, I.; Vinyals, O.; and Le, Q. V. 2014. 使用神经网络进行序列到序列学习。《神经信息处理系统进展》，27。Taieb, S. B.; Sorjamaa, A.; and Bontempi, G. 2010. 用于多步超前时间序列预测的多输出建模。《神经计算》，73(10-12): 1950–1957。Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. 注意力机制就是你所需要的一切。《神经信息处理系统进展》，30。Wilks, D. S. 2011. 《大气科学中的统计方法》，第 100 卷。Academic Press。Winters, P. R. 1960. 利用指数加权移动平均预测销售额。《管理科学》，6(3): 324–342。