

基于多粒度多尺度深度时空模型的长期序列预测方法

陈加毫, 谢良, 廖思灏, 吴雨琛, 徐海蛟

引用本文

陈加毫, 谢良, 廖思灏, 吴雨琛, 徐海蛟. [基于多粒度多尺度深度时空模型的长期序列预测方法](#) [J]. 计算机科学, 2025, 52(2): 91-98.

CHEN Jiahao, XIE Liang, LIAO Sihao, WU Yuchen, XU Haijiao. [Long-term Series Forecasting Method Based on Multi-granularity Multi-scale Deep Spatio-Temporal Modeling](#) [J]. Computer Science, 2025, 52(2): 91-98.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于季节分解的混合神经网络的时间序列预测](#)

Time Series Prediction of Hybrid Neural Networks Based on Seasonal Decomposition

计算机科学, 2024, 51(11A): 231200008-7. <https://doi.org/10.11896/jsjcx.231200008>

[基于双域稀疏Transformer的变电站设备故障预警方法](#)

Substation Equipment Malfunction Alarm Algorithm Based on Dual-domain Sparse Transformer

计算机科学, 2024, 51(5): 62-69. <https://doi.org/10.11896/jsjcx.230300001>

[基于动态选择预测器的深度强化学习投资组合模型](#)

Deep Reinforcement Learning Portfolio Model Based on Dynamic Selectors

计算机科学, 2024, 51(4): 344-352. <https://doi.org/10.11896/jsjcx.230100048>

[基于核技巧改进的Informer模型的长序列时间序列预测方法](#)

Prediction Method of Long Series Time Series Based on Improved Informer Model with Kernel Technique

计算机科学, 2023, 50(11A): 221100186-6. <https://doi.org/10.11896/jsjcx.221100186>

[基于spike-and-slab先验的贝叶斯时间序列模型](#)

Bayesian Time-series Model Based on spike-and-slab Prior

计算机科学, 2023, 50(11A): 221200131-6. <https://doi.org/10.11896/jsjcx.221200131>

基于多粒度多尺度深度时空模型的长期序列预测方法

陈加毫¹ 谢良¹ 廖思灏¹ 吴雨琛¹ 徐海蛟²

¹ 武汉理工大学理学院 武汉 430070

² 广东第二师范学院计算机学院 广州 510303

(2248919270@qq.com)

摘要 时间序列建模一直是金融和交通等多个领域研究的热点,时空模型因能更全面捕捉时序数据的复杂关联和趋势,受到研究者的广泛关注。近年来,基于时空模型的长期序列预测取得显著成果,但现有方法受到多粒度或多尺度研究的限制,无法充分挖掘数据的时空信息。为解决这一问题,提出了一种多粒度多尺度深度时空模型(MMDSTM)。该模型首先通过分解初始数据获取季节、周期和粒度序列;然后,利用基于多尺度等距卷积生成尺度预测,利用基于注意力的时空特征层生成多粒度预测;最后,通过多层次融合合并多粒度与多尺度预测的预测结果。在实验中,MMDSTM相比近期的新方法在股票、交通和电池数据集上MSE指标分别下降了6.2%,21.5%和1%。多粒度和多尺度的引入显著提升了时间序列预测性能。

关键词: 多粒度学习;多尺度学习;时间序列预测;金融市场;交通流速度

中图分类号 TP391

Long-term Series Forecasting Method Based on Multi-granularity Multi-scale Deep Spatio-Temporal Modeling

CHEN Jiahao¹, XIE Liang¹, LIAO Sihao¹, WU Yuchen¹ and XU Haijiao²

¹ College of Science, Wuhan University of Technology, Wuhan 430070, China

² School of Computer Science, Guangdong University of Education, Guangzhou 510303, China

Abstract Time series modeling has been the focus of research in a number of fields, including finance and transportation, and spatio-temporal models have received a lot of attention from researchers because of their ability to capture the complex associations and trends in time-series data more comprehensively. In recent years, long-term series forecasting based on spatio-temporal modeling has achieved remarkable results, but the existing methods are limited by multi-granularity or multi-scale studies, which cannot fully mine the spatio-temporal information of the data. To overcome this problem, a multi-granularity multi-scale deep spatio-temporal model(MMDSTM) is proposed. The model first obtains seasonal, periodic and granularity sequences by decomposing the initial data. Then, the multi-scale isometric convolution generates scale predictions, while attention-based spatio-temporal feature layers generates multi-granularity predictions. Finally, the prediction results of multi-granularity and multi-scale predictions are merged by multi-level fusion. In experiments, MMDSTM's MSE metric decreases by 6.2%, 21.5% and 1% compared to other methods on stock, traffic and battery datasets, and the introduction of multi-granularity and multi-scale significantly improves the time series forecasting performance.

Keywords Multi-granularity learning, Multi-scale learning, Time series forecasting, Financial markets, Traffic flow speed

1 引言

时间序列预测已经成为学术界和工业界关注的焦点,涵盖了金融市场^[1-3]、交通运输^[4-5]等多个领域。现代社会利用各种传感器如摄像头、雷达等记录数据。例如,交通监控系统

通过实时监测车流量、车速和拥堵情况,优化信号灯控制,提高交通效率;在经济金融领域,高频交易传感器捕捉微小股价波动,帮助投资者即时决策,实现更精准的交易策略。然而时间序列数据通常受到多种因素的影响,包括季节性、趋势、周期性等,使得数据具有复杂的结构,从而增加了预测的难度。

到稿日期:2024-04-17 返修日期:2024-10-20

基金项目:广东省自然科学基金(2020A1515011208);广州市基础研究教计划基础与应用基础研究项目(202102080353);广东省普通高校自然科学类特色创新项目(2019KTSCX117)

This work was supported by the Natural Science Foundation of Guangdong Province, China(2020A1515011208), Basic and Applied Basic Research Project of Guangzhou Basic Research Teaching Program(202102080353) and Characteristic Innovation Project of Natural Science in General Colleges and Universities in Guangdong Province(2019KTSCX117).

通信作者:谢良(whutxl@hotmail.com)

与此同时,多个时间序列之间存在一定的相关性,例如同一行业股票和相邻道路之间具有一定的空间相关性,这为深刻把握时间序列的演变提供了关键的分析视角。

传统的时间序列分析方法有 SVR^[6] 和 ARIMA^[7]。ARIMA 模型可以处理数据的趋势和周期性,但对于非线性关系的建模相对有限。为了捕捉非线性关系,多种自回归模型如 LRidge^[8] 和 LSVR^[9] 等被提出。然而,这些模型仍然基于对时间序列具有一定分布或函数形式的假设,因此在处理高度非线性的实时时间序列数据时存在一定的局限。近年来,深度学习在股票和交通预测领域取得了显著的成果。深度学习技术包括循环神经网络(RNN)^[10]、长短期记忆网络(LSTM)^[11] 和门控循环单元(GRU)^[12],它们在预测方面表现出色。

此外,股票和交通数据不仅包含时间特征,还由于它们之间复杂的相互关系而包含空间特征。一些时空深度学习技术逐渐被应用于时间序列预测,例如,卷积神经网络(CNN)^[13] 和图卷积网络(GCN)^[14] 通过学习时序中的空间关系来进行预测。为了满足中期和长期预测任务要求,STGCN^[15] (Spatio-Temporal Graph Convolutional Network)通过建模多尺度交通网络有效捕获全面的时空相关性。这些网络利用数据之间复杂的相互依赖关系来实现更精确的价格和交通速度预测。尽管上述方法通过准确预测在解决交通和股票投资问题上取得了一定成果,但值得强调的是,很多研究在很大程度上忽略了同时探索多粒度和多尺度的重要性。基于多粒度和多尺度的方法可以帮助揭示时序数据中的复杂结构和动态特性,从而提高对数据的理解和预测能力。

为了克服上述模型的局限,本文提出了多粒度多尺度深度时空模型(Multi-granularity Multi-scale Deep Spatio-Temporal Model, MMDSTM)来提高时间序列预测的性能。具体来说,对于多尺度,本文采用多尺度等距卷积网络来得到季节周期性预测;对于多粒度,本文构建了多个时空网络块。为了学习不同粒度数据的时空特征,时空网络块使用 GRU 捕获时间特征,并结合 Transformer 捕获空间特征,最后融合得到不同粒度的时空关系嵌入。每个块负责学习特定粒度级别的时空信息。该方法不仅可以提取时间信息,还能够有效地捕获动态空间信息,从而实现更有效的信息获取。为了充分利用各粒度数据的时空信息,减小有效性差异,对提取的时空隐藏状态进行融合,最终生成每个粒度预测结果。

本文工作的主要贡献包括:

1)相较于传统的图定义方法,本文采用多头自注意力机制,旨在计算不同粒度时序之间的动态空间相关性,以捕获时空特征。挖掘不同粒度的空间关系,能够深入捕捉数据的演变规律,提高预测精度。

2)在每个粒度进行时空学习后,通过融合多尺度等距卷积得到的预测结果生成多个粒度和尺度融合的多层次融合预测,最后平均得到最终预测。通过粒度和尺度的结合,使得模型能捕获到在不同运动趋势下所展现的季节周期性。

3)在英国 A1 道路交通数据集和中国 CSI300 股票市场数据集上进行了大量实验,结果表明,与其他模型相比,所提模型表现出更高的精度,在金融和交通数据上的出色性能

验证了其良好的通用性。

2 相关工作

2.1 时空预测模型

传统的时间序列预测方法,如 SVR 和 ARIMA,起源于 20 世纪 70 年代,适用于简单的时间序列,但难以处理时空相关性。随着深度学习的发展,递归神经网络(Recurrent Neural Networks, RNN)逐渐成为时间序列建模的默认选择。然而,当面对长序列时,它们容易遇到梯度消失的问题。长短期记忆网络(Long Short-Term Memory, LSTM)和门控循环单元(Gated Recurrent Unit, GRU)在一定程度上缓解了这一问题,但无法满足长期依赖性的需求。为了实现更准确的预测,涌现出了一些复杂的结构,如时间注意层(LSTNet-A)^[16] 和时间模式注意(TPA)^[17]。随后,基于自注意机制的 Transformer^[18] 被成功应用于序列建模。近几年,时空预测模型得到广泛应用,例如,MTGNN^[19] (Multivariate Time Series Forecasting with Graph Neural Networks)通过学习变量间的单向关系构图,来捕获时间序列间的时空依赖特征。尽管以上模型取得了成功,但这些方法主要关注单一粒度的预测,缺乏对不同粒度的广泛研究。

2.2 多粒度时间序列预测

近年来,多粒度时间序列预测成为时间序列研究领域备受关注的热点。国内外学者在这一领域进行了深入的研究和探索。近三年里,Hou 等(2021)引入了对比多粒度学习框架(CMLF)^[20],该框架以多粒度时间数据为基础,被应用于股票趋势预测任务。为了解决多粒度输入和单一粒度目标之间的差异,CMLF 采用了一种创新的对比双重目标学习机制。随后,Hou 等在 2022 年进一步扩展了这项工作,提出了多粒度残差学习框架(MRLF)^[21],旨在提升时间序列预测性能。然而,这些方法未明确考虑每个粒度数据之间的关系。实际上,数据之间存在动态相关性,充分利用这些潜在的依赖关系有助于提高模型的预测性能。此外,将多粒度和多尺度进行融合,可以更充分地挖掘信息,实现更准确的预测。

2.3 多尺度时间序列预测

在交通和股票数据的研究中,数据呈现出显著的季节性和周期性趋势。季节性趋势在股票市场中体现为每年在相同时间点出现相似的涨跌趋势,而在交通数据中可能表现为每天、每周或每年相同时间段内交通流量模式的周期性重复。为解决这一问题,近期的研究倾向于采用多尺度模型。Wang 等提出了 FAMC-Net^[22] 模型,该模型采用了特殊的频域奇偶校正注意(FPCA)机制,通过将时间序列转换为频域进行全局信息建模。此外,利用多尺度扩展卷积(MSDC)模块提取每个周期内的局部特征。为了将局部建模和全局建模结合成一个统一的模型,以达到高效率 and 可解释性,Wang 等提出了多尺度等距卷积网络(MICN)^[23],通过不同卷积核的多个分支对序列的不同潜在模式信息进行建模。每个分支使用基于下采样卷积的局部模块提取序列的局部特征,再利用基于等距卷积的全局模块建模全局相关性,最后采用 Merge 操作将多个分支中不同模式的信息融合。这些方法在多尺度卷积的应用上取得了显著成果,然而,它们仍然缺乏对多粒度的

考虑。本研究致力于实现多粒度和多尺度综合考虑,以实现
对数据复杂性更全面的建模,提高预测性能。

3 模型设计

3.1 方法框架

图 1 展示了本文的整体结构。首先对输入进行两个
分解块的处理,以获取趋势周期序列、季节序列和多粒度
序列。对于趋势周期和季节序列,采用基于等距卷积的多
尺度方法进行预测;对于多粒度序列,则使用相应的基于
自注意力的多粒度预测。之后,将每个粒度的预测与尺度
预测相加实现多层次融合预测,再平均得到最终的预测

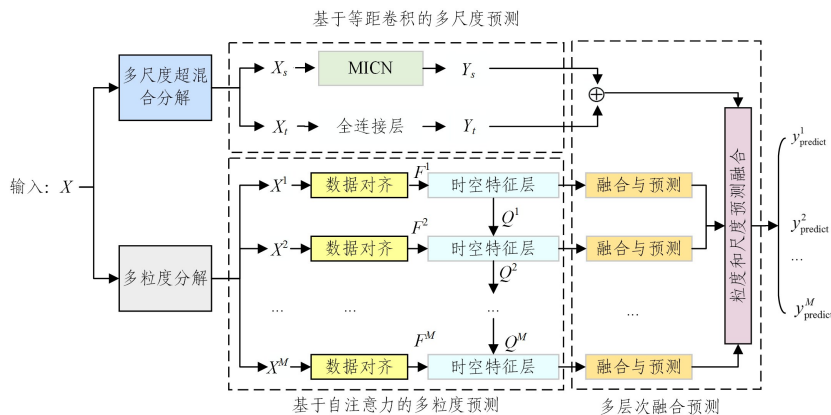


图 1 MMDSTM 整体结构图(电子版为彩图)

Fig. 1 Overall structure of MMDSTM

3.2 多粒度和多尺度分解层

为了在预测任务中获取多粒度、多尺度的季节和周期趋势
信息,本文引入了多粒度分解层和多尺度混合分解层。这
两个层的应用旨在实现对时序数据更为细致和全面的分解,
从而更有效地捕捉不同时间尺度下的季节性和周期趋势信息,
以及不同粒度下反映的数据运动趋势信息。

多粒度分解层:给定原始序列 $X = [x_1, x_2, \dots, x_T]$, 其时
间长度为 T , 时间间隔为 1, 粒度为 M 。那么最粗粒度 $X^1 =$
 $[x_1, x_{\lfloor T/2^s \rfloor + 1}, \dots, x_{\lfloor T/2^s \rfloor (s-1)}, x_T]$ 是将 T 分为 s
等份, 其中 s 是满足等式 $1 = \lfloor T/2^{M-1} s \rfloor$ 的最大整数; 第二粒度 $X^2 = [x_1,$
 $x_{\lfloor T/2^s \rfloor + 1}, \dots, x_{\lfloor T/2^s \rfloor (2s-1)}, x_T]$ 是将 T 分为 $2s$
等份; 第 m 个粒度 $X^m = [x_1, x_{\lfloor T/2^m s \rfloor + 1}, \dots, x_{\lfloor T/2^m s \rfloor (2^{m-1} s-1)}, x_T]$ 是将 T 分为
 $2^{(m-1)} s$ 等份, 以此类推。最细的粒度是被等分为 $2^{(M-1)} s$
份。例如, 假设原始序列为 $X = [x_1, x_2, \dots, x_{95}, x_{96}]$, 则粒度数
为 2, 然后计算得出 $s = 48$, 此处最粗粒度为 $X^1 = [x_1, x_3, x_5, \dots,$
 $x_{T_{95}}, x_{T_{96}}]$, 最细粒度为 $X^2 = [x_1, x_2, x_3, \dots, x_{T_{95}}, x_{T_{96}}]$ 。

多尺度混合分解层:该块使用了 $\text{AvgPool}(\cdot)$ 的几个不
同核, 可以有目的地分离几个不同的趋势周期和季节部分模
式。其次, 使用简单的均值运算来整合这些不同的模式, 因为
无法在了解每个模式的特征之前确定每个模式的权重。相应
地, 在特征表示完成后, 将此加权操作放在季节预测块的
Merge 部分。具体来说, 对于输入序列 $X \in R^{T \times D}$, 这个过程
如下:

$$X_t = \text{mean}(\text{AvgPool}(\text{Padding}(X))_{\text{kernel}_1}, \dots, \text{AvgPool}(\text{Padding}(X))_{\text{kernel}_n}) \quad (1)$$

结果。具体来说, 首先, 通过多粒度和多尺度分解层得到
不同粒度数据、季节趋势数据和周期性趋势数据; 其次, 每
个粒度数据先通过数据对齐, 再通过残差去除粒度间的信
息冗余; 之后将得到的特定粒度信息输入时空特征层以提
取出时空特征; 最后通过时空特征融合并预测出当前粒度
结果。同时, 季节趋势数据和周期性趋势数据在多尺度等
距卷积和线性层中输出季节和周期预测, 再相加得到尺度
预测结果。最后, 将每个粒度和尺度结果相加实现尺度和
粒度的多层次融合预测, 接着平均所有结果, 输出最终预
测。这个结构能够全面考虑尺度和不同粒度对时间序列
的影响, 提高预测的准确性和鲁棒性。

$$X_s = X - X_t \quad (2)$$

其中 $X_t, X_s \in R^{T \times D}$ 分别为趋势周期部分和季节部分。

3.3 基于自注意力的多粒度时空特征层

3.3.1 数据对齐

由于不同粒度的尺寸不一致 $\{X^1, X^2, \dots, X^M\}$, 因此我们
将其对齐到相同的空间, 以便后续的残差学习操作, 如图 1
的黄色部分所示。输入 $X^i = [x_1, x_2, \dots, x_{T^i}] \in R^{N \times T^i \times D}$, 这里以
第 i 个粒度为例, 它代表 T^i 时间步上的粒度为 i 的数据。首
先, 将 X^i 的第 2 维和第 3 维进行交换, 然后对其进行线性变
换, 使其与 X^i 的第 3 维保持一致的时间长度。最后, 通过对
第 2 维和第 3 维进行置换, 恢复张量的原始结构。特征对齐
的描述如下:

$$F^i = \text{Align}^i(X^i) \quad (3)$$

对齐后的特征表示为 $F^i \in R^{N \times K \times D}$, 其中 K 表示所有粒
度数据都对齐的时间长度。

3.3.2 时空特征层

为了学习每个粒度数据的时空特征, 本文构建了多个时
空网络块。这些时空网络块分别采用 GRU 捕获时间特征,
结合 Transformer 捕获空间特征, 最终融合得到不同粒度的
时空关系嵌入。每个块专注于学习特定粒度的时空信息, 这
种方法能够提取时间信息, 还能有效地捕捉动态空间信息, 实
现更为全面的信息获取。该模块的结构如图 2 所示。

时间特征层:为了捕获时间序列中的关键信息并学习长
期依赖关系, 本文利用 GRU 对每个序列进行时间特征提取,
提取的隐藏状态用于后续空间特征提取和下一个粒度级别

预测。为了详细说明具体过程,本文将第 i 个粒度为例介绍该模块。

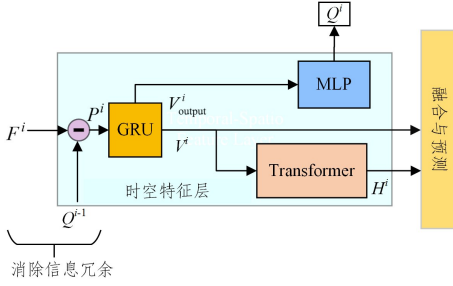


图2 多粒度时空特征提取层

Fig. 2 Multi-granularity spatio-temporal feature extraction layer

当接收到上层输出 P_i 后,将其输入 GRU 模型以捕获每个股票的隐藏特征 v^j ,并生成下一级粒度数据 v_{output}^j 的预测隐藏特征。其中 P_i 表示残差运算后得到的数据,具体过程将在下一小节中描述。具体公式如下:

$$v_{output}^j, v^j = GRU^j(P_i) \quad (4)$$

$$Q^{i+1} = MLP^i(v_{output}^j) \quad (5)$$

其中, $v^j \in R^{N \times D'}$ 表示 GRU 在第 i 个粒度提取的时间特征, D' 表示隐藏维度; $v_{output}^j \in R^{N \times K \times D}$ 粒度预测未来价值 $Q^{i+1} \in R^{N \times D'}$ 。对于预测,本文采用多层感知器方法。在这种情况下,时间特征层产生两个输出:一个用于后续下游的空间信息特征提取 v^j ,另一个 v_{output}^j 用于生成在下一层消除不同粒度之间信息冗余的 Q^{i+1} 。

自注意力空间特征层:对于空间特征的形成,GRU 捕获时间序列的数据特征后,将得到的隐藏状态特征视为包含序列相关信息的节点。本文通过 Transformer 进行卷积提取空间隐藏状态,这种机制保证了每个节点都能专注于学习当前粒度下的空间状态,从而更好地捕捉每粒度下的时序关系。

$$H^i = Transformer^i \\ = Decoder(MultiHeadAttention(Encode(Embedding(v^j)))) \quad (6)$$

其中,嵌入和编码层是 Transformer 标准形式,Multi-HeadAttention 是将每条道路或每只股票提取到的时间隐藏状态作为一个“词向量”,然后通过注意力机制计算出道路或股票之间的权重,这里多头数量为 8。解码层用一个线性层代替,式(6)中的 $H^i \in R^{N \times D'}$ 表示从中提取的空间特征。这一步的目的是更好地学习时序之间的空间相关性。

3.3.3 消除信息冗余

为了减轻不同粒度的数据冗余,本文采用了基于残差的方法。首先,从原始数据中提取粗粒度和细粒度数据。其次,观察到细粒度数据由粗粒度趋势主导,这给直接捕获细粒度趋势带来了挑战。然后,考虑到不同粒度的数据来自不同时间级别的统计测量,粗粒度数据包含细粒度数据分布的先验信息。在此基础上,利用给定的粗粒度数据可以预测可能的细粒度数据。最后,通过计算不同粒度数据之间的残差,消除冗余信息,获得特定的细粒度数据。对于第一层,输入粗粒度数据 $P^1 = F^1$,其余层使用残差。具体表达式如下:

$$P^i = \begin{cases} F^i, & i=1 \\ F^i - Q^i, & i \neq 1 \end{cases} \quad (7)$$

其中,去除冗余的输入 $P^i \in R^{N \times K \times D}$ 表示第 $i-1$ 级预测粒度数据与当前第 i 级粒度数据的残差。引入残差方法后,该模型能够更敏锐地感知数据中的微小趋势变化,从而更准确地学习和捕获其中的复杂相关性。

为了从粗粒度数据中捕获趋势信息,同时在残差结构中保留细粒度信息的独特特征,必须在粗粒度信息中最大程度地表示原始细粒度信息。将此约束表述为:

$$L_1 = \sum_{i=1, \dots, N} \|F^i - Q^i\|_F^2 \quad (8)$$

其中, $\|\cdot\|_F^2$ 为 Frobenius 范数。为了不影响细粒度信息的提取过程,在优化时,这里固定了细粒度信息的提取过程,只对 Q^i 进行优化。

3.4 基于等距卷积的多尺度预测

首先,本文采用和 FEDformer 一样的 3 部分嵌入来综合多源信息,提升了模型对复杂任务的表达能力;其次, MICN 结合多尺度分支结构和等距卷积,解决局部和全局相关性的问题,实现对时间序列信息的高效捕捉。MICN 模块的结构如图 3 所示。

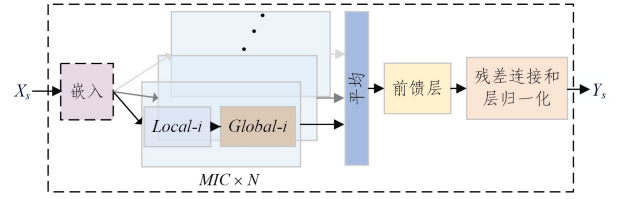


图3 MICN 整体结构图

Fig. 3 Overall structure of MICN

嵌入:采用 3 部分嵌入输入。过程如下:

$$X_s^{emb} = sum(TEF + PE + VE(X_s)) \quad (9)$$

其中, $X_s^{emb} \in R^{T \times D}$, TEF 表示时间特征编码(如 Hour Of Day, Day Of Week), PE 表示位置编码, VE 表示值嵌入。

多尺度等距卷积(MICN)层:该层包含多个分支,具有不同的尺度大小,用于模拟可能不同的时间模式。如图 1 所示, Local-Global 模块提取序列的局部特征和全局相关性。具体表达式如下:

$$Y_{s,t} = Y_{s,t-1} \quad (10)$$

$$Y_{s,t}^{local,i} = Conv1d(AvgPool(Padding(Y_{s,t-1})))_{kernel=i} \quad (11)$$

其中, $Y_{s,t-1}$ 表示 $(l-1)$ -MIC 层的输出, $Y_{s,0} = X_s^{emb}$, $i \in \{T/4, T/8, \dots\}$ 表示不同分支对应的卷积尺寸。Conv1d 的卷积核为 $kernel = i$, 输出为局部特征 $Y_{s,t}^{local,i} \in R^{i \times D}$ 。

全局模块采用等距卷积的一种变体,核大小为 S ,使其能够对局部模块输出的全局相关性进行建模。使用大的卷积核,等距卷积有效地融合了局部特征信息,具体表达式如下:

$$Y_{s,t}^{',i} = Norm(Y_{s,t}^{local,i} + Dropout(Tanh(IsometricConv(Y_{s,t}^{local,i})))) \quad (12)$$

$$Y_{s,t}^{global,i} = Norm(Y_{s,t-1}^{local,i} + Dropout(Tanh(Conv1dTranspose(Y_{s,t}^{',i})))_{kernel=i}) \quad (13)$$

其中, $Y_{s,t}^{local,i} \in R^{i \times D}$ 表示全局特征信息, $Y_{s,t-1}$ 为 $(l-1)$ -MIC 层输出, $Y_{s,t}^{global,i} \in R^{i \times D}$ 表示分支的结果。这里使用 Conv2d 来合并不同权重的不同分支结果,过程如下:

$$Y_{s,l}^{\text{merge}} = \left(\text{Conv2} \left(Y_{s,l}^{\text{global},i}, i \in \left[\frac{I}{4}, \frac{I}{8}, \dots \right] \right) \right) \quad (14)$$

$$\bar{Y}_{s,l} = \text{Norm}(Y_{s,l}^{\text{merge}} + \text{FeedForward}(Y_{s,l}^{\text{merge}})) \quad (15)$$

其中, $\bar{Y}_{s,l} \in R^{T \times D}$ 表示第 l -MIC 层的结果。为了得到季节部分的最终预测,使用投影和截断运算。

$$Y_s = \text{Truncate}(\text{Projection}(\bar{Y}_{s,N})) \quad (16)$$

其中, $\bar{Y}_{s,N} \in R^{T \times D}$ 表示第 N 个 MIC 层的输出, $Y_s \in R^{T_p \times D}$ 表示季节部分的最终预测。

全连接层:对于趋势预测块,本文采用回归方法(全连接层)得到趋势预测结果 $Y_t \in R^{T_p \times D}$ 。

3.5 多层次融合预测

在粒度学习得到结果后,将其与多尺度等距卷积得到的预测结果相加,形成每个粒度和尺度融合的多层次预测;然后,对预测进行平均,得到最终预测。在粒度预测中融合尺度预测,可以提升模型对不同数据模式的适应性和泛化能力。

多粒度融合预测:为了充分利用每个粒度的时空信息,本文将捕获的时空特征拼接起来,然后传递给 MLP,从而生成每个粒度下的预测。

$$Y^i = \text{Softmax}(\text{MLP}(H^i \oplus V^i)) \quad (17)$$

其中, $Y^i \in R^{T_p \times D}$ 表示第 i 个粒度层第 T_p 个时间点的值预测。

粒度和尺度预测融合:将多个粒度预测与尺度等距卷积预测结果和周期预测结果相加,得到粒度与尺度的多层次融合预测结果 $y_{\text{predict}}^i \in R^{T_p \times D}$ 。最后,对每个预测结果求和平均得到最终预测 y_{predict} 。具体表达式如下:

$$y_{\text{predict}}^i = Y^i \oplus (Y_s \oplus Y_t) \quad (18)$$

$$y_{\text{predict}} = \text{mean}(y_{\text{predict}}^i) \quad (19)$$

其中, $y_{\text{predict}} \in R^{T_p \times D}$ 表示一个长序列预测。

3.6 模型优化

本文将预测任务的损失与跨粒度残差学习过程中的重构损失 L_1 结合起来,得到如下损失函数:

$$L = \lambda L_1 + L_2(y_{\text{predict}}, y_r) \quad (20)$$

其中, λ 是平衡损失的超参数, L_2 为均方误差损失函数, y_r 代表真实值。最后利用 Adam 算法来执行反向传播以及更新模型参数。

4 实验结果与分析

4.1 数据集描述

本文实验使用沪深 300 指数股票数据集¹⁾、英国 A1 高速公路交通数据集²⁾和 FC^[24] 电池数据集。股票数据集涵盖了从 2010 年 9 月 28 日到 2023 年 5 月 9 日的股票数据,其中包含 3064 个交易日,输入为每天的股票数据。交通数据集包含了从 2023 年 1 月 1 日到 2023 年 10 月 31 日这 304 天的交通数据,其中输入为每 15 min 的速度与流量数据。电池数据集输入为 1020 h 的电池状态数据。对于股票数据集,排除了 2010 年 9 月 28 日之后上市的股票。然后,从这些选定的股票中进行时间调整。如果某些股票在某些日子缺少数据,则使用前一天的数据来填补当天的缺失数据。对于交通数据

集,同样剔除了中间缺失超过 4 h 的道路,某些道路缺失数据也使用前一天的数据来填补,然后将 15 min 的数据聚合成 1 h 数据,这里聚合数据采用均值。最后,本文分别选取沪深 300 指数和交通数据中的 183 只股票、14 条道路和电池数据进行训练和测试。所有数据集均按照 7:1:2 的比例划分为训练集、验证集和测试集。股票训练集、验证集和测试集分别为 2144 天、307 天和 613 天,交通训练集、验证集和测试集分别为 5107 h、730 h 和 1459 h,电池训练集、验证集和测试集分别为 714 h、102 h 和 204 h。

输入数据特征:股票数据集统计了 6 个特征和 4 个技术指标,其有效性已在大量关注股市预测的研究^[25]中得到证实,具体如表 1 所列。最后,将其规范化并用作输入变量。交通数据包包含车辆速度(V_t)和车流量(F_t),这两个指标被广泛用于交通预测。同样将这两个特征规范化作为变量输入。电池数据集包含电压、电流等 4 种常用指标。

表 1 股票和交通输入

Table 1 Stock and traffic inputs

动量因子	Momentum
平滑异同平均	MACD
相对强弱指数	RSI
移动平均	EMA
开盘价	Z_{open}
收盘价	Z_{close}
最高价	Z_{high}
最低价	Z_{low}
交易量	Z_{volume}
调整后收盘价	$Z_{\text{adj_close}}$
车辆速度	V_t
车流量	F_t

超参数设置:在本文方法中,隐藏层尺寸大小从 [64, 96, 128] 中选择 64,批量大小从 [16, 32, 64, 128] 中选择 32,学习率从 $[10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}]$ 选择 10^{-1} ,学习速率权重衰减从 $[10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}]$ 选择 10^{-1} ,预测长度从 [24, 48, 96, 192] 选择 24。

4.2 对比方法

为了验证本文提出的方法 MDSTCF 的计算性能,本文将与其结果与其他方法进行了比较,包括 GRU^[12], STGCN^[15], TRANSFORMER^[18], MGTNN^[19], MICN^[23], StockMixer^[26]。每种方法执行 10 次,以减少随机性,从而获得更稳健的结果。

GRU:该方法可以有效地捕捉时间序列中的长期和短期依赖关系,因而能够确定过去一段时间内数据变动对未来的影响。对于该方法,隐藏层的维数设置为 64,预测长度 24 为长期序列预测结果。

Transformer:该模型采用编码器-解码器结构和自关注机制,有效捕捉历史序列中的时间特征,从而提高数据的预测精度。特别是在捕获长期依赖关系方面,Transformer 展现出良好的预测能力。编码器层和多头数量为 8。

MGTNN:该方法通过图学习模块自动提取变量之间的关系。此外,混合跳传播层和扩展初始层可以捕获时间序列

¹⁾ www.baostock.com

²⁾ http://tris.highwaysengland.co.uk/detail/trafficflowdata

中的时空依赖关系。最后,利用卷积层输出长期序列预测。 $L2$ 正则化惩罚从 $[10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}]$ 中选为 10^{-2} 。在每个图卷积模块后应用 LayerNorm。将 mixhop 传播层深度设置为 2,混合跳传播层的保留比设置为 0.05,图学习层激活函数的饱和率设为 3,节点内嵌维数为 40。

STGCN:时空图卷积网络,旨在有效地捕捉交通流的时间动态和空间依赖性。该架构通过结合图卷积层和卷积序列学习层,以纯卷积结构从交通网络中提取时空特征。ST-Conv 块的 3 层通道分别为 16,32,64,时间卷积核大小都设置为 3,初始学习率设置为 0.001,每 5 次迭代后衰减率为 0.7。

MICN:多尺度等距卷积网络方法,采用多尺度分支结构,以不同的卷积核分别对时间序列中不同潜在模式进行建模。该方法有效结合了局部特征提取和全局相关性建模,能够捕捉时间序列的整体视图。MICN 输入序列长度为 96,等距卷积块为两层卷积核,大小分别 12 和 16,嵌入层维度设置为 512。

StockMixer:该方法通过指标混合、时间混合和行业混合 3 个步骤实现预测,其中特别设计时间混合机制以实现多尺度时间信息交换。与复杂架构相比,StockMixer 不仅易于优化,还能有效捕获时序数据中复杂的运动规律。模型损失因子 α 为 0.1,多尺度因子设置为 3。

4.3 评价指标

为了评估模型的预测效果,本文采用 3 种评价指标,即平均绝对误差 (Mean Absolute Error, MAE)、均方误差 (Mean Square Error, MSE) 和平均绝对百分比误差 (Mean Absolute Percentage Error, MAPE)。MAE 表示真实值与预测值之间误差的平均绝对值,它衡量了预测值和实际值之间误差的平均绝对差,该指标越小越好,因为它表示预测误差的平均幅度。MSE 表示真实值与预测值之间误差的平均平方值,它衡量了预测值和实际值之间误差的平均平方差异。MSE 越小越好,因为它对较大误差的惩罚更强烈,更关注较大误差的影响。MAPE 表示真实值与预测值之间百分比误差的平均绝对值,其衡量了预测值和实际值之间误差的平均百分比差异,通过实际值的比例进行了归一化。MAPE 越小越好,因为它表示相对于实际值的误差的平均比例。

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (21)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (22)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (23)$$

其中, y_i 和 \hat{y}_i 分别表示真实值与预测值。

4.4 对比实验

表 2 的结果清晰地表明,在英国高速公路 A1 交通数据集上,本文方法相对于传统的时间序列方法 (GRU 和 Transformer) 表现出较大的性能提升。与 STGCN 和 MTGNN 时空模型相比,本文方法在 MSE 上分别降低了 22.7% 和 20.3%,在 MAE 上分别降低了 18.1% 和 22.0%,在 MAPE 上分别降低了 10.8% 和 11.4%。这些结果表明,融合了季节周期信息的多粒度时空预测方法在性能上优于现有的时空模

型。与近两年发表的多尺度模型 MICN 和 StockMixer 相比,本文提出的方法在处理时序数据时表现出更强的多尺度特征捕捉能力和预测精度。具体来说,相对于 MICN 和 StockMixer,本文模型的 MSE 分别降低了 6.1% 和 24.6%。在其他指标上,本文模型在交通数据集上与 MICN 预测结果相近,这可能是因为 MICN 对交通数据集具有更强的适应性。但在股票数据集上,本文提出的模型相对于 MICN 有一定的提升,其中 MSE, MAE 和 MAPE 分别降低了 38.6%, 19.6% 和 4.1%。同时,在该数据集上相比最新方法 StockMixer,本文模型 3 种指标分别降低了 4.4%, 9.5% 和 2.5%。在表 3 所列的两组电池数据集上,本文模型相对于其他方法具有较好的性能。例如,相对于近期的方法 MICN 和 StockMixer,本文模型的 MSE 分别降低了 1% 和 22.9%, MAE 分别降低了 8.3% 和 19.6%, MAPE 分别降低了 0.1% 和 2.6%。在这些数据集上的结果表明,本文方法具有更好的泛化性能,证明了其具有更强的适用性和稳健性。

表 2 MMDSTM 与对比方法在两个数据集上的性能比较

Table 2 Performance comparison of MMDSTM and contrastive methods on two datasets

methods	A1			CSI300		
	MSE	MAE	MAPE/%	MSE	MAE	MAPE/%
GRU	0.425	0.351	15.419	0.271	0.347	17.753
Transformer	0.952	0.745	18.403	1.509	0.967	19.457
STGCN	0.471	0.423	17.179	0.448	0.397	16.304
MTGNN	0.456	0.445	17.293	0.424	0.395	16.282
MICN	0.387	0.348	15.314	0.390	0.381	16.104
StockMixer	0.482	0.477	17.458	0.2502	0.338	15.429
MMDSTM	0.363	0.347	15.328	0.239	0.306	15.032

表 3 MMDSTM 与对比方法在两组电池数据集上的性能比较

Table 3 Performance comparison of MMDSTM and contrastive methods on two battery datasets

methods	AFC1			CSIFC2		
	MSE	MAE	MAPE/%	MSE	MAE	MAPE/%
GRU	0.525	0.497	16.327	0.496	0.477	16.735
Transformer	0.646	0.691	18.283	0.674	0.620	18.026
STGCN	0.550	0.472	16.254	0.502	0.453	16.074
MTGNN	0.529	0.490	16.171	0.473	0.436	15.992
MICN	0.384	0.357	15.529	0.352	0.316	15.482
StockMixer	0.493	0.476	16.568	0.487	0.429	16.159
MMDSTM	0.380	0.324	15.142	0.368	0.311	15.107

4.5 消融实验

本文进行了消融实验,旨在探究模型的各个组成部分对整体性能的影响,深入了解每个特定组件的独特贡献和重要性,从而加深对模型内部机制的理解。通过仔细分析每个组成部分的单独影响,可以更全面地评估模型性能和基本特征。以下是 MMDSTM 的一些变体。

1) MMDSTM(G): 表示只有粒度预测,即粒度的时空特征提取及预测。

2) MMDSTM(G+S/M+T): 表示粒度预测和尺度趋势预测,但是 MICN 预测中去掉嵌入层。

3) MMDSTM(G+T): 表示粒度预测和周期趋势预测,但是尺度预测中去掉 MICN 层。

4) MMDSTM(G+S): 表示粒度预测和尺度预测,但是

尺度预测中去掉线性回归层。

表 4 中的结果明确表明,删除特定模型组件会导致性能下降,尤其是仅使用粒度预测的效果最差。在两个数据集上,A1 数据集的 MSE 下降了 62%,CSI 300 数据集的 MSE 下降了 52%,其他两个指标也有显著下降。在引入多尺度变体后,MAE 分数在两个数据集上有了明显提升,特别是在 MMDSTM(G+S/M+T)变体中,两个数据集上的 MSE 分数仅下降了 6.1%和 11%,表明加入多尺度对整体性能有积极的影响。

表 4 MMDSTM 变体在两个数据集上的得分

Table 4 MMDSTM variants' scores on two datasets

methods	A1			CSI300		
	MSE	MAE	MAPE/%	MSE	MAE	MAPE/%
MMDSTM(G)	0.592	0.549	17.542	0.457	0.4682	17.152
MMDSTM (G+S/M+T)	0.500	0.427	17.220	0.368	0.3972	16.071
MMDSTM(G+T)	0.540	0.469	17.490	0.438	0.4272	16.485
MMDSTM(G+S)	0.376	0.400	16.863	0.265	0.3642	15.930

4.6 参数实验

通过参数分析来选择模型的最优参数,主要从 3 个方面来进行参数选择:不同粒度(G1,G2,G3 分别代表从 1~3 粒大小)、不同尺度(C1,C2,C3,C4 分别代表从 1~4 种不同卷积的尺度)和输出序列长度(H1,H2,H3,H4 表示输出序列长度大小为 24,48,96 和 192)。

不同粒度:在实验中,首先重点分析不同粒度的准确性和整体性能,主要通过 MAE 和 MSE 分数来衡量模型的参数性能。从图 4 中可以明显看出,双粒度模型的结果相对来说是最优的。最初,本文考虑从单粒度扩展至多粒度,但随着粒度的增加,模型的时空复杂度也显著增加,因此展示了 3 个粒度的结果。在实验中,单粒度测试效果较差,三粒度训练效果较好,但其测试效果未达到预期,这是过拟合导致的。通过综合实验数据和时空复杂度得出,双粒度模型是最合适的选择。

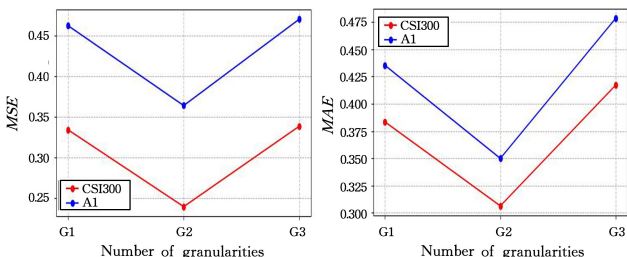


图 4 不同粒度在两个数据集上的 MSE 和 MAE 得分

Fig. 4 MSE and MAE scores of models with different granularities on two datasets

不同尺度:通过调整等距卷积的层数和相应卷积核的尺寸,深入研究了不同尺度对模型性能的影响。图 5 的结果表明,在不同尺度中,尺度为 2 的卷积表现出最佳性能。随着尺度的增加,模型性能逐渐下降,这表明过大的局部卷积核可能会对信息提取产生负面影响。基于实验数据的综合分析,从图 5 中可以看出,尺度为 2 的配置可使模型在性能和信息提取效果上达到最佳,表明该配置能够更有效地提取特征,实现较优的性能。

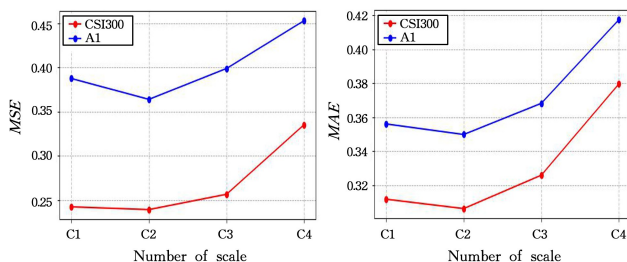


图 5 不同尺度在两个数据集上的 MSE 和 MAE 得分

Fig. 5 MSE and MAE scores of models with different scales on two datasets

输出序列长度:为了深入研究输出序列长度对长时间序列预测的影响,本文通过调整不同的输出长度来寻找最合适的预测长度。如图 6 所示,当输入序列长度为 24 时,模型表现相对较好;而随着输出长度的增加,模型性能先保持稳定然后逐渐下降。实验结果表明,在处理较长的预测任务时,所提模型能够保持相对稳定的性能,反映了模型对于更长时间序列的良好适应性。

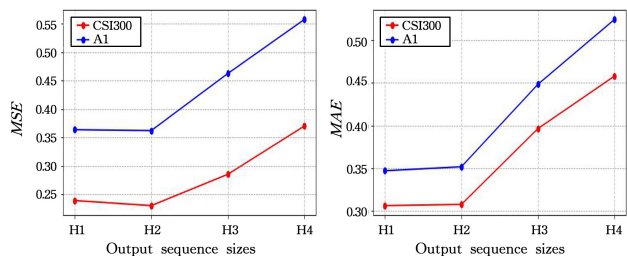


图 6 不同输出序列长度在两个数据集上的 MSE 和 MAE 得分

Fig. 6 MSE and MAE scores of models with different output sequence lengths on two datasets

结束语 本文通过引入多尺度多粒度深度时空模型(MMDSTM),成功突破了现有时间序列预测方法在多粒度和多尺度研究方面的局限。所提出的模型融合了 Transformer 的强大序列建模能力、MICN 的多尺度特征捕捉优势,以及 Attention 机制的高效特征选择能力,并进行了针对实际应用创新。例如,引入了多粒度特征提取机制和多尺度预测方法,以更好地捕捉股市和交通等时序数据中的趋势和季节周期信息;另外,通过融合多粒度和多尺度信息,本文模型在股票、交通和电池预测的实际应用中的预测性能和鲁棒性得到提升,从而能更准确地捕捉市场波动、交通模式和电池寿命变化。实验结果表明,所提方法在多个数据集上均取得了较好的性能提升,证明了这些改进的有效性。

未来将考虑从两个方面进一步优化 MMDSTM。首先,进一步研究多粒度和多尺度之间的联系,提高交通和股票预测的准确性;其次,通过考虑文本内容(如新闻)来优化模型。

参考文献

- [1] SALINAS D, FLUNKERT V, GASTHAUS J, et al. DeepAR: Probabilistic forecasting with autoregressive recurrent networks [J]. International Journal of Forecasting, 2020, 36(3): 1181-1191.
- [2] BAHDANAU D, CHO K, BENGIO Y. Neural machine transla-

- tion by jointly learning to align and translate[J]. arXiv:1409.0473,2014.
- [3] BELGHZI M I, BARATIN A, RAJESHWAR S, et al. Mutual information neural estimation[C]// International Conference on Machine Learning. PMLR, 2018:531-540.
- [4] BINKOWSKI M, MARTI G, DONNAT P. Autoregressive convolutional neural networks for asynchronous time series[C]// International Conference on Machine Learning. PMLR, 2018:580-589.
- [5] BOX G. BOX and JENKINS: time series analysis, forecasting and control[M]// A Very British Affair: Six Britons and the Development of Time Series Analysis During the 20th Century. London: Palgrave Macmillan UK, 2013:161-215.
- [6] VAN GESTEL T, SUYKENS J A K, BAESTAENS D E, et al. Financial time series prediction using least squares support vector machines within the evidence framework[J]. IEEE Transactions on Neural Networks, 2001, 12(4):809-821.
- [7] BOX G E P, PIERCE D A. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models[J]. Journal of the American statistical Association, 1970, 65(332):1509-1526.
- [8] YOO J, SOUN Y, PARK Y, et al. Accurate multivariate stock movement prediction via data-axis transformer with multi-level contexts[C]// Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. 2021:2037-2045.
- [9] DING Q, WU S, SUN H, et al. Hierarchical Multi-Scale Gaussian Transformer for Stock Movement Prediction[C]// IJCAI. 2020:4640-4646.
- [10] HAN M, XU M. Laplacian echo state network for multivariate time series prediction[J]. IEEE Transactions on Neural Networks and Learning Systems, 2017, 29(1):238-244.
- [11] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8):1735-1780.
- [12] VARGAS M R, DE LIMA B S L P, EVSUKOFFA G. Deep learning for stock market prediction from financial news articles[C]// 2017 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications(CIVEMSA). IEEE, 2017:60-65.
- [13] NIEPERT M, AHMED M, KUTZKOV K. Learning convolutional neural networks for graphs[C]// International Conference on Machine Learning. PMLR, 2016:2014-2023.
- [14] CHEN W, JIANG M, ZHANG W G, et al. A novel graph convolutional feature based convolutional neural network for stock trend prediction[J]. Information Sciences, 2021, 556:67-94.
- [15] YU B, YIN H, ZHU Z. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting[J]. arXiv:1709.04875, 2017.
- [16] LAI G, CHANG W C, YANG Y, et al. Modeling long-and short-term temporal patterns with deep neural networks[C]// The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. 2018:95-104.
- [17] SHIN S Y, SUN F K, LEE H. Temporal pattern attention for multivariate time series forecasting [J]. Machine Learning, 2019, 108:1421-1441.
- [18] VASWANI A, SHAZEERN, PARMAR N, et al. Attention is all you need[C]// Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017:6000-6010.
- [19] WU Z, PAN S, LONG G, et al. Connecting the dots: Multivariate time series forecasting with graph neural networks[C]// Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2020:753-763.
- [20] HOU M, XU C, LIU Y, et al. Stock trend prediction with multi-granularity data: A contrastive learning approach with adaptive fusion[C]// Proceedings of the 30th ACM International Conference on Information & Knowledge Management. 2021:700-709.
- [21] HOU M, XU C, LI Z, et al. Multi-Granularity Residual Learning with Confidence Estimation for Time Series Prediction[C]// Proceedings of the ACM Web Conference 2022. 2022:112-121.
- [22] WANG M, WANG H, ZHANG F. FAMC-Net: Frequency Domain Parity Correction Attention and Multi-Scale Dilated Convolution for Time Series Forecasting[C]// Proceedings of the 32nd ACM International Conference on Information and Knowledge Management. 2023:2554-2563.
- [23] WANG H, PENG J, HUANG F, et al. Micn: Multi-scale local and global context modeling for long-term series forecasting [C]// The Eleventh International Conference on Learning Representations. 2022.
- [24] JOUIN M, GOURIVEAU R, HISSEL D, et al. Prognostics of PEM fuel cell in a particle filtering framework[J]. International Journal of Hydrogen Energy, 2014, 39(1):481-494.
- [25] CHUNG H, SHIN K. Genetic algorithm-optimized multi-channel convolutional neural network for stock market prediction [J]. Neural Computing and Applications, 2020, 32(12):7897-7914.
- [26] FAN J, SHEN Y. StockMixer: A Simple Yet Strong MLP-Based Architecture for Stock Price Forecasting[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2024:8389-8397.



CHEN Jiahao, born in 1997. His main research interests include deep learning and data mining.



XIE Liang, born in 1987, Ph. D, associate professor. His main research interests include multimedia retrieval and machine learning.