

基于时间序列的非周期预测模型

曹建文¹, 委兴宝^{2,3}, 杨裔^{2,4}, 李彩虹^{2,3}, 赵文清^{2,3}

1. 兰州工业学院计算机与人工智能学院, 甘肃 兰州 730050;
2. 兰州大学信息科学与工程学院, 甘肃 兰州 730000;
3. 旅游信息融合处理与数据权属保护文化和旅游部重点实验室, 甘肃 兰州 730000;
4. 甘肃省人工智能与算力技术重点实验室, 甘肃 兰州 730000

摘要

在实际应用中, 纯周期性的数据相对罕见, 大多数数据往往表现出非周期性特征, 难以通过简单的周期性变化进行预测或描述。而单一神经网络在处理非周期性时间序列时往往面临过拟合、长时依赖捕获困难、非线性关系捕获有限等问题。为了有效地对非周期性时间序列进行预测, 基于 Informer 模型提出了 ILTNet 模型。ILTNet 模型结合线性预测 (AR 模型) 和非线性预测 (Informer 模型与循环跳跃组件), 能有效捕获长期依赖关系。实验证明, 与 LSTNet、Informer、AR 以及 GRU 模型相比, ILTNet 模型在非周期时间序列预测上表现出显著优势。例如, 在 Exchange Rate 数据集上, ILNet 模型相对于 LSTNet 模型, 在步长为 96 和 128 时将 RSE 分别降低了 0.0333 和 0.0277, 相对于 Informer 模型在所有步长下, RSE 均有显著降低, 尤其是在步长为 96 时将 RSE 降低了 0.2877。

关键词

多变量; 时间序列预测; 卷积神经网络; LSTNet; Informer

中图分类号: TP302.7

文献标志码: A

doi: 10.11959/j.issn.2096-0271.2025010

Model for non-periodic time series prediction

CAO Jianwen¹, WEI Xingbao^{2,3}, YANG Yi^{2,4}, LI Caihong^{2,3}, ZHAO Wenqing^{2,3}

1. School of Computer Science and Artificial Intelligence, Lanzhou Institute of Technology, Lanzhou 730050, China
2. School of Information Science and Engineering, Lanzhou University, Lanzhou 730000, China
3. Key Laboratory of Tourism Information Fusion Processing and Data Ownership Protection, Ministry of Culture and Tourism, Lanzhou 730000, China
4. Gansu Provincial Key Laboratory of Artificial Intelligence and Computing Technology, Lanzhou 730000, China

Abstract

In practical applications, purely periodic data is relatively rare, and most data often exhibit non-periodic characteristics, making it challenging to predict or describe through simple periodic changes. Single neural network frequently encounters issues such as overfitting, difficulties in capturing long-term dependencies, and limited ability to capture nonlinear relationships when dealing with non-periodic time series. To effectively forecast non-periodic time series, the ILTNet model was proposed based on the Informer model. The ILTNet model integrated linear prediction (AR model) and nonlinear prediction (Informer model and recurrent skip components), enabling effective capture of long-term dependencies. Experiments show that ILTNet model demonstrates significant advantages in non-periodic time series forecasting compared to the LSTNet, Informer, AR, and GRU models. For example, on the Exchange Rate dataset, ILTNet model reduces RSE by 0.0333 and 0.0277 compared to LSTNet model at horizons of 96 and 128, respectively.

Compared to Informer model, ILTNet model achieves significant RSE reductions at all horizons, particularly reducing RSE by 0.2877 at a horizon of 96.

Key words
multivariate, time series forecasting, convolutional neural network, LSTNet, Informer

0 引言

通过分析丰富的历史时间序列数据，人们可以深入洞察未来发展趋势。根据实际需求，预测时间跨度可以划分为短期、中期和长期预测。其中，长期多变量时间序列预测具有重要的现实应用意义。它不仅有助于提前识别趋势变化，还能在变革到来之前提供采取应对措施的宝贵机会，从而在复杂多变的环境中保持竞争优势。这种方法不仅在经济领域^[1]具备显著价值，还被广泛应用于交通^[2]、光能^[3]、气象学^[4]等众多领域。笔者将进一步探讨这一关键方法的实际应用和潜在影响，为决策者和研究者提供宝贵的见解。

多变量时间序列预测方式目前主要被分为3类：统计模型、机器学习模型以及深度神经网络。其中自回归模型（autoregressive model，AR模型）^[5]与自回归移动平均模型（autoregressive integrated moving average model，ARIMA）^[6]是两个被广泛使用的统计模型。自回归模型用同一变量历史的表现情况来预测该变量未来的表现情况，并假设它们为线性关系，而ARIMA将时间序列建模为历史数据和白噪声的线性组合，但这些模型只能应用于内部具备线性关系的序列，并且其并没有考虑变量之间的依赖关系，所以在很多实际应用中不能使用。机器学习模型如支持向量机（support vector machines，SVM）^[7]、随机森林（random forest，

RF）^[8]、极端梯度提升（extreme gradient boosting，XGBoost）^[9]等，以数据驱动的方式学习序列的特征，建模历史数据和未来数据之间的映射关系，虽然适用范围更广，精度也有所提升，但是特征选择往往非常耗时。近年来，神经网络能够提取非线性关系，能够从输入信号中学习复杂的模式等特性，以神经网络为代表的深度学习算法成为预测领域的研究热点。在研究将深度神经网络应用于时间序列预测时，尽管这些模型通常表现优异，但它们在处理非周期性时间序列方面存在一些问题。例如，卷积神经网络（convolution neural network，CNN）和递归神经网络（recurrent neural network，RNN）在提取特征和处理时间依赖性方面表现出色，但在处理非周期性模式时往往表现较差^[10]。其他混合模型如TCN-Linear（temporal convolutional network - linear）尝试通过结合时间卷积网络和线性组件来增强在混沌时间序列上的表现，但即便是这些先进的模型，在处理非周期性场景时也有其局限性^[11]。另外，Lai等^[12]提出了一种新的深度学习框架，通过CNN提取短期特征和RNN中的门控循环单元（gated recurrent unit，GRU）学习时间依赖性，引入线性分量AR模型，简单地组合起来就能得到很好的预测效果，但是在非周期模式上表现较差。多变量时间序列数据集的预测误差很大程度上取决于元数据集的时间周期模式和相互依赖性，通常采用各种隐式或显式的方法提取多变量时间序列之间的相互依

赖关系。Zhou 等^[13]在 Transformer 模型的基础上提出了一种优化模型 Informer 模型，继承了 Transformer 模型的 Self-Attention 机制与编码器译码器构造，并引入 ProbSparse Self-Attention 机制，提升了预测精确度，同时有效降低了计算复杂度，它的编码器和译码器都是一系列循环神经网络，输入序列和输出序列的长度可以不同。尽管 Informer 模型能够有效地实现长距离时间序列预测，但在非周期性时间序列预测方面仍存在不足，这导致其在某些实际应用中的预测精度相对较低，需要进一步改进。

为了实现多变量长距离预测，并解决 Informer 模型对于非周期序列预测精度不足的问题，提出了一种基于 Informer 的长期时间序列网络（Informer long-term time-series network, ILTNet）模型。该模型使用没有池化层的卷积神经网络提取时间维度上的短期模式以及变量间的局部依赖关系，基于编译码结构捕捉长期依赖的优势，结合循环跳跃组件准确捕捉时序数据中蕴藏的周期性，使用线性自回归模型提取时间序列数据的线性趋势，最后并行输出。该模型在周期的电能数据集以及非周期的汇率数据集上进行了验证。

1 模型构建与相关理论

在实际应用中，时间序列数据呈现出多样的特征，有些具有明显的周期性，而有些则表现为非周期性。模型构建的过程中，为了更好地应对这种多样性，将预测任务分解为周期性预测任务和非周期性预测任务。循环跳跃组件预测周期性部分，自回归模型和 Informer 模型分别预测非周期性的线性部分和非线性部分。模型预测

窗口的大小固定，使用滚动预测，在 t 时刻输入数据 $X^t = \{x_1^t, x_2^t, \dots, x_L^t\}$ ，其中 L 为输入长度。先用卷积神经网络进行卷积，再使用编译码结构和循环跳跃组件捕捉长期依赖关系，将该非线性神经网络部分与自回归线性模型并行，使模型对时间序列具有更强的鲁棒性，输出 $Y^t = \{y_1^t, y_2^t, \dots, y_L^t\}$ ， $x_i^t, y_i^t \in R^n$ ， n 为变量维度，为了预测 j 时刻后 h 时间的 X^{j+h} ，假设 $\{X^1, X^2, \dots, X^j\}$ 是已知的。模型总体构建情况如图 1 所示。

1.1 Informer 模型

对于非周期性的变化，传统的神经网络很难进行精确预测。为了有效捕获输出和输入之间精确的长期依赖关系，从而使模型拥有优秀的预测能力，可使用 Informer 模型^[13]机制。Informer 模型采用 Transformer 模型^[14]的编译码结构，使用自注意力蒸馏机制，提取主要注意力，大幅降低了网络的规模，并引入概率注意力机制来预测长序列，有效地降低了运算量，缩短了运行时间。

传统的 Transformer 模型计算 Self-Attention 值的方法是根据输入信息 Q (query)、键 K (key) 和值 V (value)，

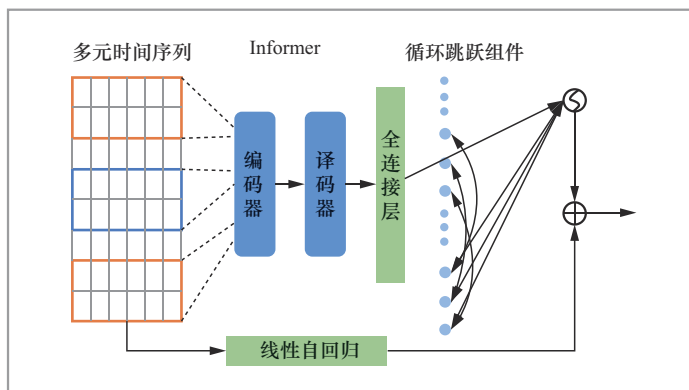


图1 模型结构

将 $\mathbf{Q}$ 和每个 $\mathbf{K}$ 进行相似度计算得到权重, 通常使用相似度函数, 例如点积、拼接、感知机等, 再使用一个 Softmax 函数对这些权重进行归一化, 最后将权重和相应的键值进行加权求和得到最后的 Attention, 计算式可表述为式 (1), 其中 d 是输入的维数。

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V} \quad (1)$$

第 i 个信息 q_i 的注意力概率形式如下。

$$A(q_i, \mathbf{K}, \mathbf{V}) = \sum_j \frac{k(q_i, k_j)}{\sum_l k(q_i, k_l)} v_j = \mathbb{E}_p(k_j | q_i) [v_j] \quad (2)$$

原自注意力机制存在稀疏性, 较少的点积贡献了绝大部分的注意力得分, 其余部分贡献极小, 可以忽略不计。利用 KL 散度 (Kullback-Leibler divergence) 区分稀疏性, 其中第 i 个 query 的稀疏性评估方法可由下述计算式计算得到。

$$M(q_i, \mathbf{K}) = \ln \sum_{j=1}^{L_K} e^{\frac{q_i k_j^T}{\sqrt{d}}} - \frac{1}{L_K} \sum_{j=1}^{L_K} \frac{q_i k_j^T}{\sqrt{d}} \quad (3)$$

ProbSparse 自注意力计算式如 (4) 所示, 它只包含稀疏测度 $M(w, \mathbf{K})$ 下的主要 u 信息, $u = c \cdot \ln L_Q$, c 为常数采样因子, 每个头产生不同的稀疏查询键对, 避免了严重的信息损失。

$$A(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\bar{\mathbf{Q}}\mathbf{K}^T}{\sqrt{d}}\right)\mathbf{V} \quad (4)$$

其中, $\bar{\mathbf{Q}}$ 与 $\mathbf{Q}$ 是大小相同的稀疏矩阵。

Informer 模型中的稀疏注意力机制通过选择性地关注少数关键位置, 将计算复杂度和内存使用量降低到 $O(L \log L)$, 从而显著提高计算和内存效率, 减少冗余计算, 增强模型的可扩展性, 并有效捕捉长程依赖关系和保留全局信息, 使模型能够高效

处理长序列数据, 并保持对整体数据模式的理解。

另外, Informer 模型的编码器允许处理更长的序列输入, 捕获长序列输入之间的鲁棒的远距离依赖关系。由于 ProbSparse 自注意机制特征映射会存在值的冗余组合, 因此, 这里利用蒸馏操作对具有优势特征赋予更高权重, 并在下一层生成有重点的自我注意力的特征映射。从 i 到 $i+1$ 层的蒸馏操作的过程如下。

$$\mathbf{X}_{j+1}' = \text{MaxPool}\left(\text{ELU}\left(\text{Conv1d}\left([\mathbf{X}_j']_{\text{AB}}\right)\right)\right) \quad (5)$$

其中, $[\mathbf{X}_j']_{\text{AB}}$ 代表注意力块。使用 ELU 激活函数^[15]在时间序列上执行一维卷积。在堆叠一层后, 在其半片中加入最大池化层和下采样 $\mathbf{X}'$, 从而降低整个内存使用量。为了增强蒸馏操作的鲁棒性, 用一半的输入构建主堆栈的副本, 并通过每次减少一层来逐步减少自注意蒸馏层的数量, 这样它们的输出维数是对齐的。将所有堆栈的输出连接起来, 得到编码器的最终隐藏表示。

Informer 模型的译码器在一次前向计算过程中, 能够预测长序列的输出, 该模型由多头注意力层和掩码多头概率自注意力层组成, 在计算过程中, 采用迭代式预测策略, 输出多步长预测结果。

在译码器中, 在第一层译码器输入时将序列长度遮蔽一半, 大大减小了译码器内存开销, 缩短了计算时间, 在 t 时刻输入数据 $\mathbf{X}'$ 变换为矩阵, 也就是式 (6) 同时提出在译码结构中使用生成式结构, 能够一次生成全部预测序列, 缩短了预测译码耗时, 加快推理速度。

$$\mathbf{X}_{\text{feed_de}}' = \text{Concat}(\mathbf{X}_{\text{token}}', \mathbf{X}_0') \in \mathbb{R}^{(L_{\text{token}} + L_y) \times d_{\text{model}}} \quad (6)$$

其中 $\mathbf{X}_{\text{token}}' \in \mathbb{R}^{L_{\text{token}} \times d_{\text{model}}}$ 是起始字符, $\mathbf{X}_0' \in \mathbb{R}^{L_y \times d_{\text{model}}}$ 是目标序列的占位符, 可以将

标量设置为0。通过将屏蔽的点积设置为 $-\infty$ ，防止每个位置都参与到下一位置，避免了自回归，最终输出如下。

$$h_t^l = A(X^l, X^{l-1}) + b \quad (7)$$

Informer 的模型结构如图2所示。

1.2 循环跳跃组件

对于周期的数据，时序性的循环神经网络是一种很好的处理方式，但是对于周期较长的数据，现有的循环神经网络很难被捕捉，会产生梯度消失的问题。为了捕捉数据中的长期相关性，使用一个带有时间跳跃连接的GRU循环神经网络，这个SKIP组件^[12]就是在相邻周期的当前隐藏单元格和同一阶段的隐藏单元格之间添加跳转连接，延长数据的时间跨度，避免了可能出现的梯度消失问题。

长短期记忆(long short-term memory, LSTM)网络是在传统的循环神经网络上引入了门结构，使用输入门、遗忘门和输出门分别控制输入值、记忆值和输出值，可以解决长期依赖问题，而GRU^[16]在保留了LSTM网络的特性后，进行了进一步的简化，直接把某一时刻的长期记忆作为输出，只有更新门和重置门两种门操作，减少了运算量。GRU在 t 时刻前向传播计算的过程如下。

重置门 r_t ，根据当前输出需要用到的记忆部分如下。

$$r_t = \text{sigmoid}(W_r x_t + U_r h_{t-1}) \quad (8)$$

更新门 z_t ，表示以前的信息是否需要更新。

$$z_t = \text{sigmoid}(W_z x_t + U_z h_{t-1}) \quad (9)$$

候选输出 c_t ， t 时刻加入新信息，用 r_t

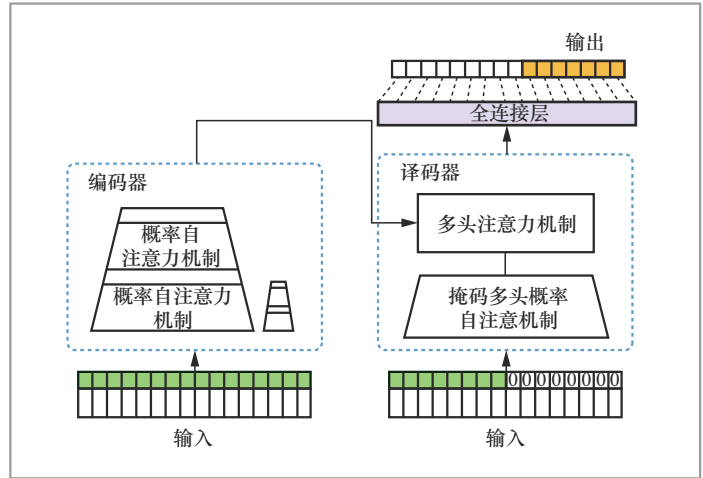


图2 Informer模型结构

来控制需要保留多少信息。

$$c_t = \tanh(W_h x_t + U(r_t \odot h_{t-1})) \quad (10)$$

在 t 时刻，忘记传递下来的 $t-1$ 时刻中的某些维度信息，并加入当前节点输入的某些维度信息，最终输出如下。

$$h_t = (1 - z_t) h_{t-1} + z_t c_t \quad (11)$$

其中， x_t 为 t 时刻的输入， W 是与当前输入 x_t 相关的权重矩阵， U 是与前一时刻隐藏状态 h_{t-1} 相关的权重矩阵， $\odot$ 表示矩阵点乘。

使用RELU函数作为隐藏层的更新激活函数。加入跳转连接后，整个循环跳跃更新如下。

$$r_t = \text{sigmoid}(x_t W_{xr} + h_{t-p} W_{hr} + b_r) \quad (12)$$

$$u_t = \text{sigmoid}(x_t W_{xu} + h_{t-p} W_{hu} + b_u) \quad (13)$$

$$c_t = \text{RELU}(x_t W_{xr} + r_t \odot (h_{t-p} W_{hr}) + b_c) \quad (14)$$

$$h_t = (1 - u_t) \odot h_{t-p} + u_t \odot c_t \quad (15)$$

其中, h_t 表示在 t 时刻的循环分量的隐藏状态, p 是跳过的隐藏单元数, p 值可根据数据集明显的周期确定, 能显著地提高模型的性能。 t 时刻的循环跳跃层输出为 h_t^R , 从时间 $t-p+1$ 到 t 时循环跳跃组件跳过量为 p , 可用 $h_{t-p+1}^S, h_{t-p+2}^S, \dots, h_t^S$ 表示, 全连接层输出式如下。

$$h_t^D = W^R h_t^R + \sum_{i=0}^{p-1} W^S h_{t-i}^S + b \quad (16)$$

1.3 自回归组件

Informer 模型部件主要处理非周期性非线性部分, 导致对线性部分的预测准确度下降。而线性模型对输入数据的线性变化非常敏感, 因此选择使用 AR 模型来处理线性分量。AR 模型^[17]的时间序列中存在线性关系, 它使用之前的观察值作为回归方程的输入, 以预测下一个时间步的值。

$$h_t^L = \sum_{k=0}^{q^{AR}-1} \varphi_k X_{t-k} + \varepsilon_t \quad (17)$$

其中, q^{AR} 为输入矩阵上窗口的大小, φ_k 为自相关系数, ε 为随机误差项, 均值为 0, 标准差为 δ 。由以上计算式可知, 自回归模型具有简单、易解释、计算效率高和擅长捕捉线性依赖关系的优势。

神经网络最终的预测结果是将周期部分和非周期部分的输出分量进行积分。 t 时刻的最终预测方法如式 (18) 所示。

$$\hat{Y}_t = h_t^I + h_t^D + h_t^L \quad (18)$$

其中, h_t^I 是 Informer 模型的输出, h_t^D 是循环跳跃组件处理后的输出, 其计算如前文所述。

1.4 实验设计

基于所构建的混合模型, 预测模型的

实验流程如图 3 所示。首先, 通过规范化周期性特征和填补缺失值对数据进行标准化处理, 同时计算相关系数以筛选有用特征, 从而去除冗余信息并确保数据的质量。其次, 使用滑动窗口机制构建时间序列样本, 通过选择不同的步长构建多种样本, 旨在捕捉时间序列中的短期和长期依赖关系。之后, 在模型训练阶段, 将构建的时间序列样本输入提出的 ILTNet 模型中进行训练。最后, 采用相对绝对误差 (relative absolute error, RAE)、相对平方误差 (relative squared error, RSE) 等评估指标对模型进行评估, 并比较不同模型 (如 ILTNet、LSTNet、Informer、AR、GRU 模型等) 的预测结果。这种方法可以确定最优模型, 并全面评估其性能表现。

选择 LSTNet、Informer、AR 和 GRU 模型作为对比模型的原因主要有两个。首先, ILTNet 模型的设计集成了 LSTNet、Informer、AR 和 GRU 模型的关键特性和优势。选择这些模型作为对比, 可以直观地展示 ILTNet 模型相对于其组成部分的创新和性能提升。例如, ILTNet 模型的线性预测部分基于 AR 模型, 而非线性预测部分则融合了 Informer 和 GRU 模型的优势。这种对比不仅突出了 ILTNet 模型的综合能力, 还验证了其在长期依赖捕获、周期性特征处理和综合预测能力方面的改进效果。

上述 4 种模型在时间序列预测领域具有代表性和广泛应用性。LSTNet 模型结合了 CNN 和 LSTM 网络, 在处理长期依赖和周期性特征数据方面表现出色。Informer 模型以其稀疏注意力机制和高效处理长序列的能力著称, 特别适用于大规模数据和长序列预测。AR 模型作为传统的线性时间序列模型, 以其简单性和高效

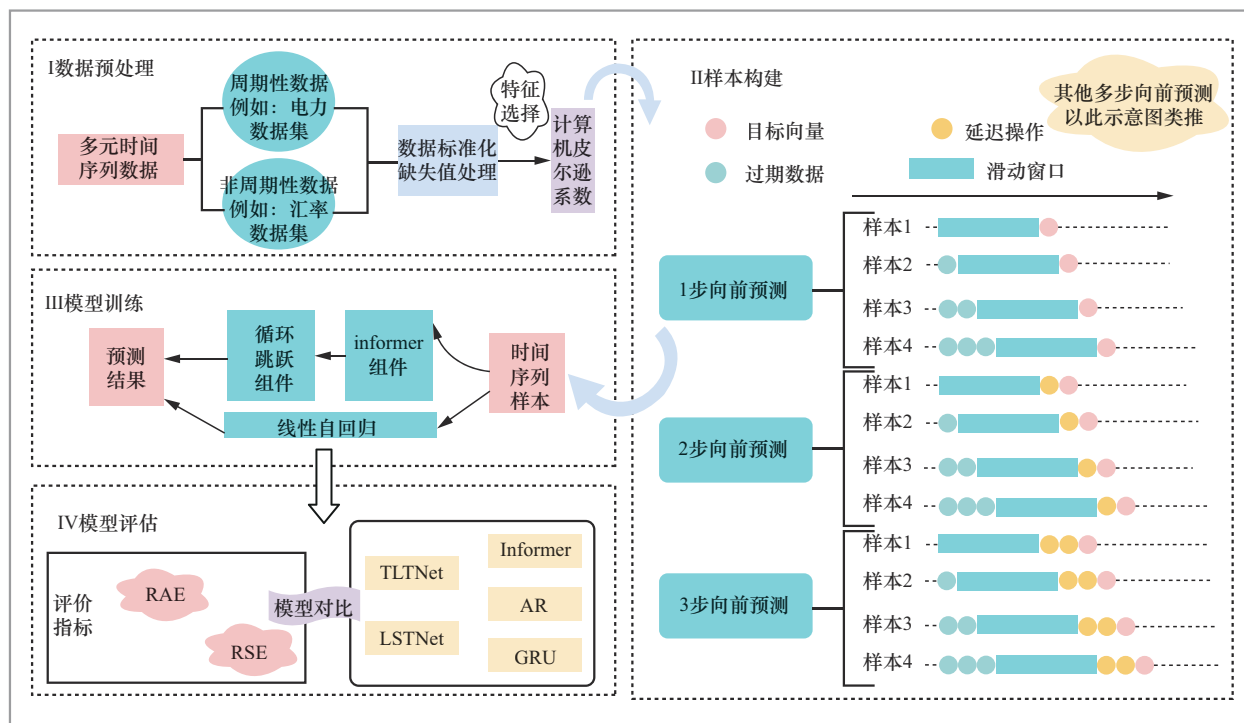


图3 模型预测的实验流程

性在处理线性关系时具有重要参考价值。GRU模型则通过其门控机制有效捕获时间序列中的依赖关系，特别适合非线性时间序列预测。选择这些模型作为对比基础，能够全面评估 ILTNet 模型在不同预测方法和策略下的性能，展示其相对于基础模型的优势和改进。

除了以上对比实验，为了保证对基于时间序列的周期性预测和非周期性预测的有效性，本文选择了两个具有不同时间分辨率和特征的数据集——Electricity 数据集（高频电力消费数据）和 Exchange Rate 数据集（低频汇率数据）。这两个数据集涵盖了周期性和非周期性的时间序列，分别代表了具有明显周期性特征的电力消耗和相对平稳的汇率变动，从而能够有效测试模型在不同时间序列特征下的预测能力。

1.5 模型复杂度分析

ILTNet 模型融合了 Informer 模型的稀疏注意力机制、GRU 模型的非线性处理能力以及 AR 模型的线性预测能力，这种综合设计虽然提升了模型的预测性能，但也增加了模型的复杂度。由于 ILTNet 模型结合了多种机制，模型的参数量较单一的 Informer、GRU 或 AR 模型要多，从而导致训练和推理过程中的计算开销增加。具体而言，ILTNet 模型的时间复杂度为 $O(n^2)$ （注意力机制）加上 $O(n)$ （循环神经网络），其中 n 是时间序列的长度，这意味着随着数据量和序列长度的增加，计算开销也会显著提高。

在训练过程中，由于 ILTNet 模型结构复杂、参数量大，其训练时间较长，特别是在处理大规模数据集时非常明显。为突破大规模数据集的计算瓶颈，可能需要采

用分布式计算或高性能计算资源来加速训练和推理。虽然 ILTNet 模型设计具备一定的可扩展性,通过调整网络层数或参数量可在一定程度上提高对大规模数据的适应能力,但这可能会影响模型的预测性能。因此,在实际应用中,需要在模型复杂度与预测精度之间找到平衡,同时考虑使用简化版模型或优化算法来提升效率。

2 实验与分析

2.1 实验环境

本文基于 Pytorch 框架进行实验。Pytorch 是一个 Python 优先的深度学习框架,所用库的版本号如下: Python 3.7.3; matplotlib 3.1.1; numpy 1.16.4; pandas 0.24.2; scikit_learn 0.21.2; torch == 1.10.0。

2.2 数据集描述

笔者使用了两个公开的数据集,第一个是 UCI 机器学习与智能系统中心中的 Electricity 数据集:2012—2014 年,每 15 min 记录一次千瓦时用电量,一天呈现 96 个小节,所有时间点的凌晨 1:00 和 2:00 之间均为零,时间序列的长度为 26 304,包含 321 个变量。第二个是 Exchange Rate 数据集,包括澳大利亚、英国、加拿大、瑞士、中国、日本、新西兰、新加坡 8 个国家从 1990 年到 2016 年的每日汇率汇总,时间序列的长度为 7 588,包含 8 个变量。这两个数据集代表了两种不同的非周期性时间序列。一个是基于电力消费的高频数据,另一个是基于汇率的低频数据。它们涵盖了不同的时间分辨率和变量类型,能够较好地代表可能的非周期性

时间序列。

对两个不同类型的数据集,使用皮尔逊相关系数计算序列之间的特征相关性生成热力图,如图 4、图 5 所示。

2.3 对比实验和参数设置

为了进一步验证 ILTNet 模型的预测能力,从数据集的选取上考虑到数据集的周期性,Electricity 数据集具有一定的周期性,而 Exchange Rate 数据集具有非周期性。ILTNet 模型超参数进行网格搜索,采用 Adam 优化器进行优化。对比实验选用了以下 4 种对比模型,参数使用的是默认推荐参数。

(1) LSTNet 模型^[12]

其是一个多变量深度序列预测模型,使用 CNN 和 RNN 来提取变量之间的短期局部依赖模式,并发现时间序列趋势的长期模式,利用传统的自回归模型提取时间序列中的线性关系。

(2) Informer 模型^[13]

该模型是一种专为长序列时间序列预测设计的、基于注意力机制的编码器/译码器构造模型,解决了 Transformer 模型在应用于长序列时间序列预测时存在的一些严重的问题,如二次时间复杂度较高、较高的内存使用量和编译码器结构的固有限制等。

(3) AR 模型^[17]

AR 模型是一个基本的单变量时间序列建模、预测模型,是简单且具代表性的随机过程,其将当前时刻的输出变量当作历史时间段变量的线性加权和。

(4) GRU 模型^[16]

GRU 模型是 LSTM 网络的一种效果很好的变体,它较 LSTM 网络的结构更加简单,而且可以更好地解决长依赖问题。

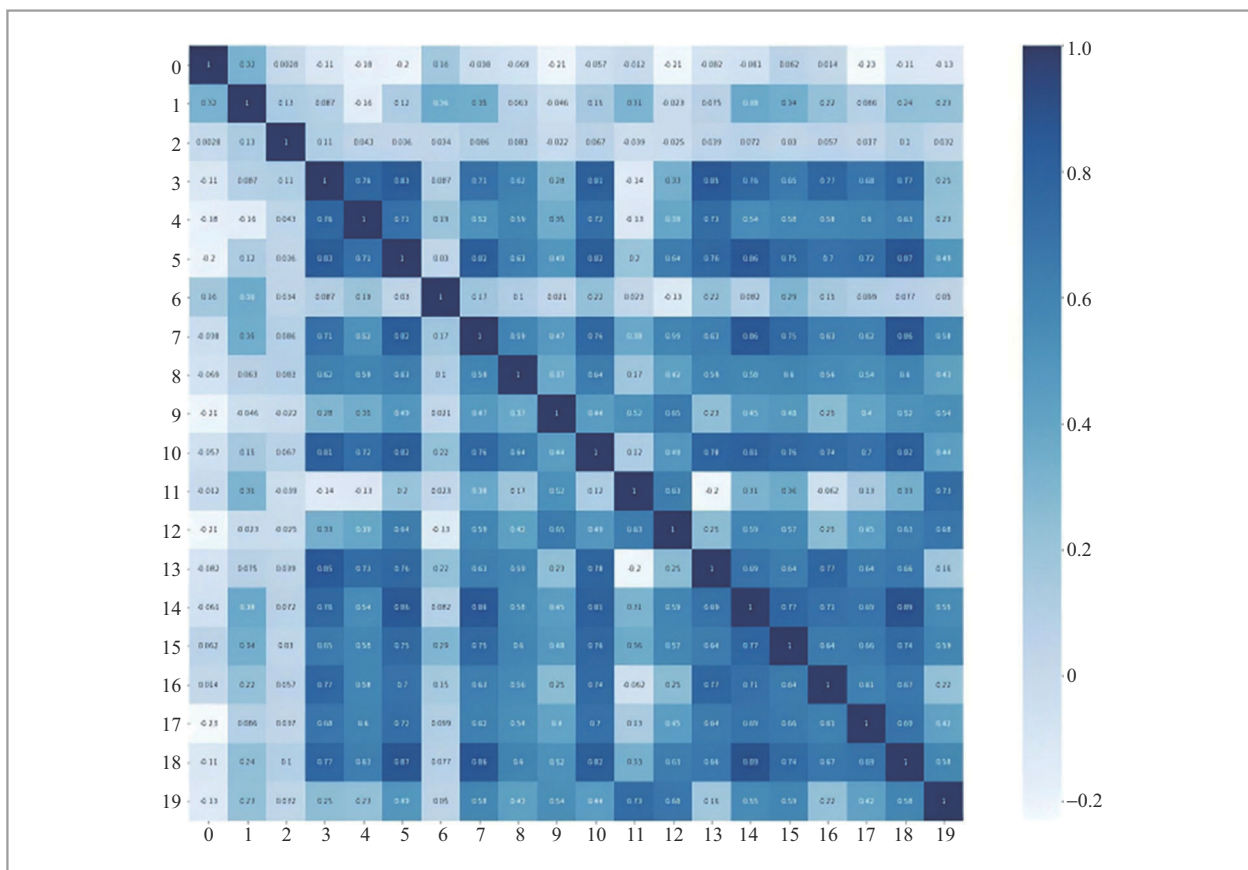


图4 Electricity数据集序列间相关性

2.4 评价指标

笔者使用了以下两种常用的指标来评价模型预测的效果。

(1) 相对平方根误差

$$RSE = \frac{\sqrt{\sum_{t \in \Omega_{\text{test}}} (Y_t - \hat{Y}_t)^2}}{\sqrt{\sum_{t \in \Omega_{\text{test}}} (Y_t - \text{mean}(Y))^2}} \quad (19)$$

(2) 相对绝对值误差

$$RAE = \frac{\sum_{t \in \Omega_{\text{test}}} |\hat{Y}_t - Y_t|}{\sum_{t \in \Omega_{\text{test}}} |Y_t - \text{mean}(Y)|} \quad (20)$$

其中, Y_t 、 $\hat{Y}_t$ 分别代表 t 时刻的真实值、预测值, $\text{mean}(Y)$ 代表实际值的平均值。相

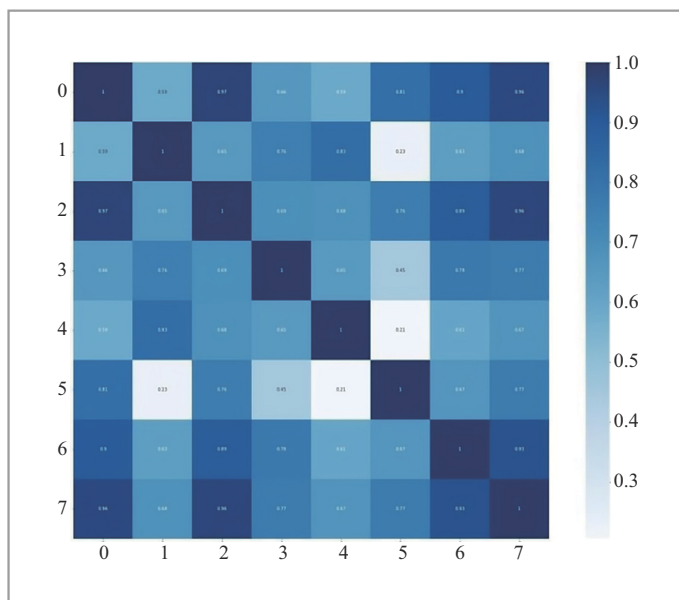


图5 Exchange Rate数据集序列间相关性

对平方根误差和相对绝对值误差越小表示模型越优。

RSE 和 RAE 是时间序列预测领域中被广泛使用的两个指标。RSE 衡量了预测值相对于真实值的平方误差在整个时间序列中的占比，而 RAE 则评估了预测误差的绝对值相对于真实值的平均值的比重。由于这两个指标能够直接反映模型的预测精度和误差分布，并在许多研究中被证明是有效和可靠的，因此常被用来评估预测模型的性能。

2.5 实验结果

5个模型在两个数据集上的预测结果见表1，每个数据指标对应的最好的结果用黑体突出显示。时间长度分别为 96、128、168、196，代表96小时、128小时、168小时、196小时的电力预测，以及96天、128天、168天、196天的汇率数据。一般时间长度越大，预测任务就越困难。用 RSE 度量时，ILTNet 模型均优于其他 4 个模型；用 RAE 度量时，ILTNet 模型仍然具有一定的优势，尤其是在非周期的 Exchange Rate 数据集

上也能表现出良好的预测情况。整体来说，ILTNet 模型在多变量预测任务中可以获得较高的整体预测精度。

在 Electricity 数据集上，ILTNet 模型相较于 LSTNet 模型，在不同步长下的性能表现如下。

- 当步长为 96 时，RSE 降低了 0.0464，但 RAE 增加了 0.0211。
- 当步长为 128 时，RSE 降低了 0.0202，RAE 降低了 0.0024。
- 当步长为 168 时，RSE 降低了 0.0022，但 RAE 增加了 0.0016。
- 当步长为 196 时，RSE 降低了 0.0038，但 RAE 增加了 0.0175。

相比于 Informer 模型，ILTNet 模型在以下方面表现出显著改进。

- 步长为 96 时，RSE 降低了 0.0673，RAE 降低了 0.0019。
- 步长为 128 时，RSE 降低了 0.0092，RAE 降低了 0.0215。
- 步长为 168 时，RSE 降低了 0.0206，RAE 降低了 0.0302。
- 步长为 196 时，RSE 降低了 0.0053，RAE 降低了 0.0048。

表1 多变量时序预测实验对比结果

方法	评价指标	Electricity 数据集				Exchange Rate 数据集			
		96	128	168	196	96	128	168	196
ILTNet 模型	RSE	0.1021	0.1589	0.1606	0.1798	0.0973	0.1299	0.111	0.1167
	RAE	0.0932	0.0770	0.0742	0.1014	0.0851	0.1163	0.1074	0.1033
LSTNet 模型	RSE	0.1485	0.1791	0.1628	0.1836	0.1306	0.1576	0.1869	0.184
	RAE	0.0721	0.0794	0.0726	0.0839	0.1198	0.148	0.1713	0.1639
Informer 模型	RSE	0.1694	0.1681	0.1812	0.1851	0.385	0.5587	0.518	0.5359
	RAE	0.0951	0.0985	0.1044	0.1062	0.3522	0.4648	0.433	0.4526
AR 模型	RSE	0.1661	0.1763	0.1786	0.1879	0.1087	0.1302	0.1374	0.1374
	RAE	0.0796	0.0802	0.0827	0.0883	0.0973	0.1086	0.1235	0.1235
GRU 模型	RSE	0.1743	0.1645	0.1712	0.1781	0.329	0.3286	0.3271	0.3264
	RAE	0.093	0.0933	0.0892	0.0975	0.3359	0.3356	0.3329	0.3316

综上所述,在周期数据集上,ILTNet模型相较于Informer模型在RSE和RAE误差方面均有显著降低,而与LSTNet模型相比,虽然在部分步长下RAE略有增加,但总体上RSE表现更优。

在Exchange rate数据集上,ILTNet模型表现出相对于LSTNet和Informer模型的显著优势。与LSTNet模型相比,ILTNet模型在步长为96和128时分别将RSE降低了0.0333和0.0277,同时将RAE降低了0.0347和0.0317。然而,在步长为168时,ILTNet的RSE和RAE分别增加了0.0759和0.0639。在步长为196时,RSE和RAE分别降低了0.0673和0.0606。相较于Informer模型,ILTNet模型在所有步长下均表现出显著的改进。

总的来看,在非周期性数据集上,ILTNet模型的表现全面优于对比的LSTNet模型和Informer模型,展示了其在处理不同步长数据时的稳健性和准确性。

在实验中选择不同步长(如96、128、168、196小时或天)进行预测是为了全面评估模型在不同时间尺度下的表现。较短的步长能够更有效地捕捉短期模式和趋势,而较长的步长则有助于评估模型在处理长期趋势和复杂依赖关系方面的能力。鉴于重点是评估模型的长期预测能力,因此最小步长设置为96小时。实验结果表明,对于两个数据集,随着步长的增加,各模型的RSE和RAE值普遍上升,预测精度逐渐降低。这一现象可以归因于较长步长下预测任务的复杂性增加,使模型在捕捉长期依赖和趋势时面临更多挑战。

为了直观地将本文算法和其他算法进行对比,笔者将数据进行了可视化。Electricity数据集的相对平方根误差和相对绝对值误差对比如图6所示。在RSE上,ILTNet模型效果优于其他4个模型,但是

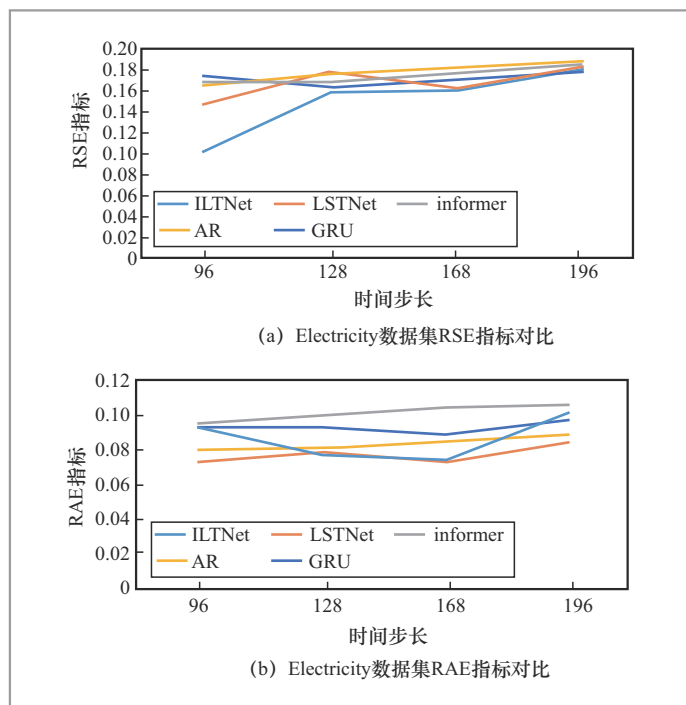


图6 Electricity数据集模型对比

在RAE上,ILTNet模型整体效果略低于LSTNet模型和AR模型,这是因为Electricity数据集具有较强的周期性,线性关系明显,随着模型复杂度的提升,使用ILTNet模型反而增加了冗余层的参数学习。Exchange Rate数据集的相对平方根误差和相对绝对值误差对比如图7所示,在RSE和RAE上,ILTNet模型都优于其他4个算法。对于周期性不明显的时间序列,ILTNet模型能敏锐地捕捉到长期依赖关系,以及提取到主要特征。

使用ILTNet模型在Electricity数据集和Exchange Rate数据集上进行预测,如图8所示,预测的时间序列(蓝色)和数据集上的真实数据(黄色)对比,ILTNet模型能较好地拟合出数据的趋势。

具体而言,ILTNet模型在不同数据集上的优越表现主要归因于其综合模型设计。该模型有效融合了Informer模型的稀疏注

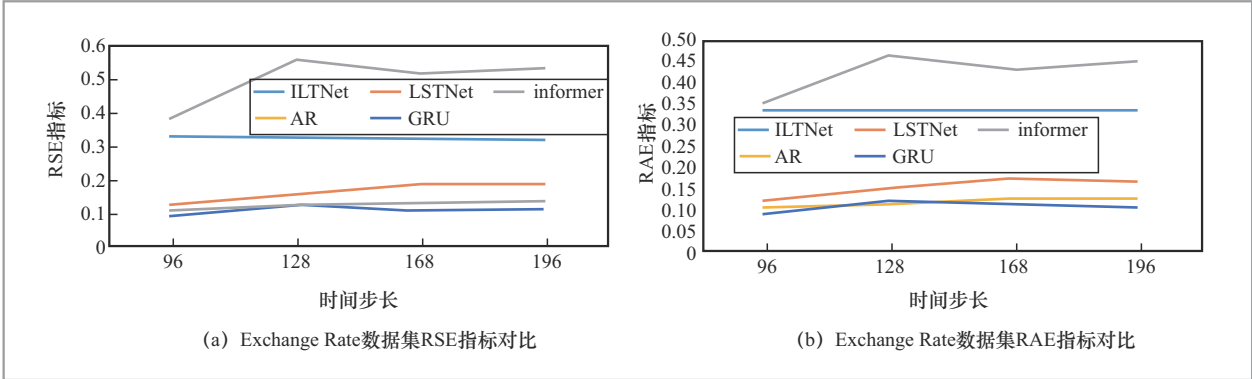


图7 Exchange Rate数据集模型对比

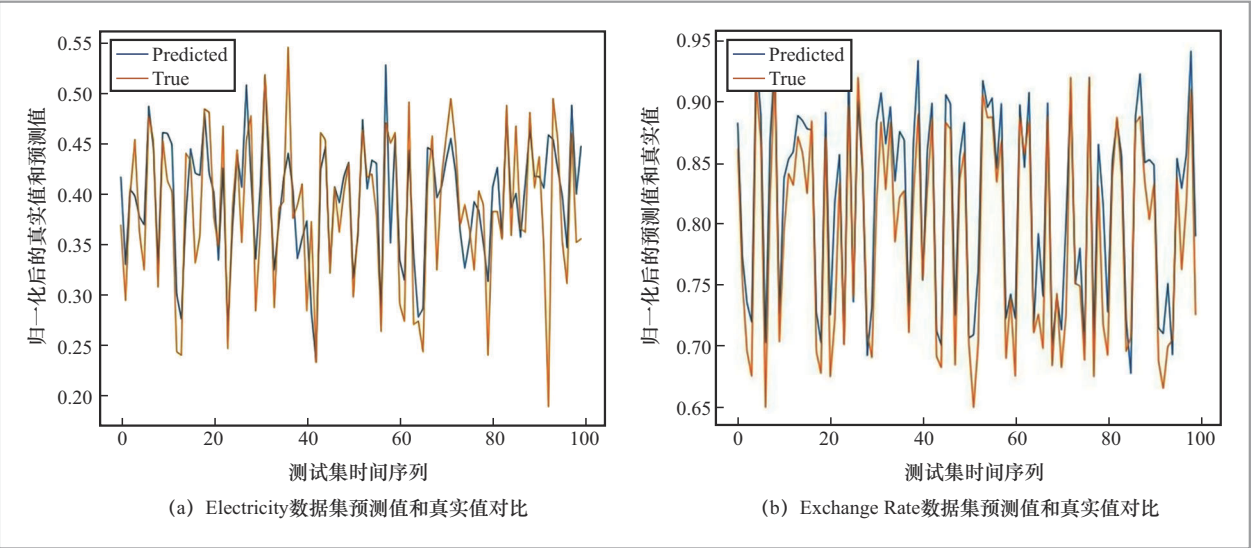


图8 预测数据和真实值对比

意力机制、跳接记忆循环神经网络的非线性处理能力与自回归模型的线性预测能力。这种综合设计形成了一个高效的预测框架，使模型能够充分发挥各组件的优势，从而显著提升了对于复杂时间依赖关系数据的整体预测性能。

此外，周期性与非周期性特征优化也是 ILTNet 模型优越表现的重要因素。在周期性数据集（如 Electricity 数据集）上，ILTNet 模型通过增强对周期性特征的

捕捉能力，显著降低了相对平方根误差。尽管在某些步长下，相对绝对值误差存在一定波动，但整体性能仍优于其他对比模型。在非周期性数据集，如 Exchange Rate 数据集上，ILTNet 展示了卓越的鲁棒性和准确性，尤其在较长时间步长下，能够有效降低 RSE 和 RAE，证明了其在长期依赖关系捕捉方面的显著优势。综合来看，ILTNet 模型在多变时间序列预测任务中的表现优异，得益于其综合性的模型

设计和对不同数据特征的优化能力。

然而,实验结果可能受到数据集特性影响。Electricity 数据集具有显著的周期性特征,其高频电力消耗数据展现出明显的周期性模式。这一特性使周期性数据的预测任务相对复杂,但也使 ILTNet 模型在捕捉这些周期性特征时表现优异。尽管在某些步长下相对绝对值误差有所波动,但整体的相对平方根误差依然较低,表明 ILTNet 模型能够有效处理周期性特征并进行准确预测。

相对而言,Exchange Rate 数据集具有平稳的非周期性特征,该低频汇率数据的波动较少,主要受长期趋势和随机变量的影响。ILTNet 模型在处理这种非周期性数据时展现出卓越的鲁棒性,特别是在较长时间步长下有效降低了 RSE 和 RAE,显示了其在捕捉长期依赖关系和复杂非周期性特征方面的显著优势。

因此,数据集特性对实验结果产生了显著影响。Electricity 数据集的周期性特征和 Exchange Rate 数据集的非周期性特征分别对模型的预测能力提出了不同的挑战。ILTNet 模型在这两种数据集上的优异表现证明了其在处理不同时间序列数据时的适应性和有效性。

3 结束语

针对多变量时间序列长期预测的不足之处,本文提出了一种基于 Informer 模型、跳接记忆循环神经网络以及自回归模型的 ILTNet 模型。在 Electricity 数据集和 Exchange Rate 数据集上,将 ILTNet 模型同 LSTNet、Informer、AR 以及 GRU 模型进行对比,在对比实验中均优于其他模型。对比单个模型,证明了将

LSTNet 模型与 Informer 模型有效结合,可以提升模型的性能。当预测长度增大时,模型的性能有所下降,这是长期时间序列预测很难避免的一个问题,为了提升模型的预测性能,下一步将考虑优化特征选择。此外,深度学习方法在对时间序列预测时,未充分考虑空间依赖性,未来可以考虑时域卷积网络(temporal convolutional network, TCN)^[18]提取时域特征,TCN 的卷积内核可以在一个层中共享,降低模型的使用内存,以改进当前算法,获得更佳的预测效果。此外,还需增加更多具有代表性的数据集,涵盖不同领域和时间尺度的非周期性时间序列,以全面验证所提模型的泛化能力和鲁棒性。

参考文献:

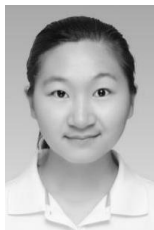
- [1] MAJUMDER M M R, HOSSAIN M I, HASAN M K. Indices prediction of Bangladeshi stock by using time series forecasting and performance analysis [C]//Proceedings of the 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE). Piscataway: IEEE Press, 2019: 1-5.
- [2] ZENG W B, WANG K T, ZHOU J H, et al. Traffic flow prediction based on hybrid deep learning models considering missing data and multiple factors[J]. Sustainability, 2023, 15(14): 11092.
- [3] ARAF I, ELKHADIRI H, ERRATTAHI R, et al. AMS solar energy prediction: a comparative study of regression models [C]//Proceedings of the 2019 Third International Conference on Intelligent Computing in Data Sciences (ICDS). Piscataway: IEEE Press, 2019: 1-5.

- [4] LALITHA C, RAVINDRAN D. Hybrid deep learning framework for weather forecast with rainfall prediction using weather bigdata analytics[J]. Multimedia Tools and Applications, 2024, 83(25): 66413–66432.
- [5] BOX G E P, JENKINS G M, REINSEL G C, et al. Time series analysis: forecasting and control[M]. Fifth edition. Hoboken, New Jersey: John Wiley & Sons, Inc., 2016.
- [6] ADHIKARI R, AGRAWAL R K. An introductory study on time series modeling and forecasting[EB]. arXiv preprint, 2013, arXiv: 1302.6613.
- [7] ŻBIKOWSKI K. Using volume weighted support vector machines with walk forward testing and feature selection for the purpose of creating stock trading strategy[J]. Expert Systems with Applications, 2015, 42(4): 1797–1805.
- [8] MA Y L, HAN R Z, FU X L. Stock prediction based on random forest and LSTM neural network[C]//Proceedings of the 2019 19th International Conference on Control, Automation and Systems (ICCAS). Piscataway: IEEE Press, 2019: 126–130.
- [9] PAN B Y. Application of XGBoost algorithm in hourly PM_{2.5} concentration prediction[J]. IOP Conference Series: Earth and Environmental Science, 2018, 113: 012127.
- [10] LIU X H, WANG W M. Deep time series forecasting models: a comprehensive survey[J]. Mathematics, 2024, 12(10): 1504.
- [11] WANG M J, QIN F T. A TCN-linear hybrid model for chaotic time series forecasting[J]. Entropy (Basel, Switzerland), 2024, 26(6): 467.
- [12] LAI G K, CHANG W C, YANG Y M, et al. Modeling long-and short-term temporal patterns with deep neural networks[C]//Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. New York: ACM, 2018: 95–104.
- [13] ZHOU H Y, ZHANG S H, PENG J Q, et al. Informer: beyond efficient transformer for long sequence time-series forecasting[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(12): 11106–11115.
- [14] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Advances in neural information processing systems. La Jolla: NIPS, 2017: 5998–6008.
- [15] CLEVERT D A, UNTERTHINER T, HOCHREITER S. Fast and accurate deep network learning by exponential linear units (ELUs)[EB]. arXiv preprint, 2015, arXiv: 1511.07289.
- [16] CHO K, VAN MERRIENBOER B, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[EB]. arXiv preprint, 2014, arXiv: 1406.1078.
- [17] PADHI D K, PADHY N, MISHRA J. Intraday stock prices forecasting using an autoregressive model[C]//Proceedings of the 2020 International Conference on Computer Science, Engineering and Applications (ICCSEA). Piscataway: IEEE Press, 2020: 1–6.
- [18] BAI S J, KOLTER J Z, KOLTUN V, et al. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling[EB]. arXiv preprint, 2018, arXiv: 1803.01271.

作者简介



曹建文（1972-），男，兰州工业学院计算机与人工智能学院副教授，主要研究方向为计算机技术。



委兴宝（1996-），女，兰州大学信息科学与工程学院硕士生，主要研究方向为时间序列分析、人工智能。



杨骞（1980-），男，博士，兰州大学信息科学与工程学院教授，主要研究方向为大数据分析、机器学习、人工智能。



李彩虹（1972-），女，博士，兰州大学信息科学与工程学院教授级高级实验师，主要研究方向为计算机科学与技术。



赵文清（1999-），女，兰州大学信息科学与工程学院硕士生，主要研究方向为旅游情感分析与人工智能。

收稿日期: 2024-06-19

通信作者: 委兴宝, 220220942340@lzu.edu.cn

基金项目: 科技创新2030—“新一代人工智能”重大项目(No.2021ZD0111405); 甘肃省2023年度省级重点人才项目

Foundation Items: Technological Innovation 2030—“New Generation Artificial Intelligence” Major Project (No.2021ZD0111405), Gansu Provincial Key Talent Project 2023