

引用格式: 毛远宏, 孙琛琛, 徐鲁豫, 等. 基于深度学习的时间序列预测方法综述[J]. 微电子学与计算机, 2023, 40(4): 8-17. [MAO Y H, SUN C C, XU L Y, et al. A survey of time series forecasting methods based on deep learning[J]. Microelectronics & Computer, 2023, 40(4): 8-17.] DOI: 10.19304/J.ISSN1000-7180.2022.0725

基于深度学习的时间序列预测方法综述

毛远宏¹, 孙琛琛², 徐鲁豫³, 刘 曦¹, 柴 波¹, 贺鹏超¹

(1 西安微电子技术研究所, 陕西 西安 710054;

2 天津理工大学 计算机科学与工程学院, 天津 300384;

3 中国航天电子技术研究院, 北京 100094)

摘 要: 时间序列预测通过分析时间序列找到其内在规律性对未来发展进行预测, 其研究有着重要的学术意义和应用价值. 特别随着传感器和网络技术的发展, 如何基于大量历史时序数据进行更加精准和高效的预测分析成为了需要解决的迫切问题. 时间序列预测任务充分借鉴了深度学习的技术研究成果, 在近年来取得了快速发展. 本文分析了时间序列预测技术的研究现状, 论述了时间序列预测所涉及到的深度学习方法的相关理论和方法, 包括卷积神经网络、循环神经网络、注意力机制和图神经网络等方法在时间预测领域的应用, 归纳总结近年来基于深度学习的时间序列研究成果, 比较了基于各种深度学习时间序列方法的优缺点, 在此基础上对基于深度学习时间序列预测方法的发展进行了展望.

关键词: 时间序列预测; 深度学习; 卷积神经网络; 循环神经网络; 注意力机制; 图神经网络

中图分类号: TP391.41

文献标识码: A

文章编号: 1000-7180(2023)04-0008-10

A survey of time series forecasting methods based on deep learning

MAO Yuanhong¹, SUN Chenchen², XU Luyu³, LIU Xi¹, CHAI Bo¹, HE Pengchao¹

(1 Xi'an Microelectronics Technology Institute, Xi'an 710054, Shaanxi, China;

2 School of Computer Science and Engineering, Tianjin University of Technology, Tianjin 300384, China;

3 China Academy of Aerospace Electronics Technology, Beijing 100094, China)

Abstract: Time series forecasting finds its internal regularity by analyzing time series to forecasts its future. Its research has important academic and application. Especially with the development of sensor and network technology, how to make more accurate prediction and analysis based on a large number of historical time series data has become an urgent problem to be solved. At present, time series forecasting methods fully use the research results of deep learning, and have made rapid development in recent years. This paper analyzes the research status of time series forecasting technology, discusses the relevant theories and methods of deep learning methods involved in time series forecasting of time overview, including the application of convolutional neural network, recurrent neural network, attention mechanism, graph neural network and other methods in the field of time forecasting, and summarizes the research achievements of time series based on deep learning in recent years. The advantages and disadvantages of various time series methods based on deep learning are compared. Finally, this paper forecasts the development trend of time series prediction methods based on deep learning.

Key words: Time series forecasting; Deep learning; Convolutional neural network; Recurrent neural network; Attention mechanism; Graph neural network

1 引言

时间序列是按照时间维度生成的一组或者多组随机变量,通常以相对固定的采样频率对于变化过程进行记录.对于时间序列来说,过去的数据潜在包含了现在或者将来数据发展变化的规律,它们之间存在着相关性,因此时间序列数据预测研究就是依据已有的时间序列数据,找出未来时刻或时间段内的某个或者某些随机变量随时间变化的趋势和规律.

时间序列数据在金融、气象、农业、工业及医疗等领域存在广泛的应用^[1-3].近年来,随着传感器和网络技术的高速发展,使得生成并积累大量的时间序列数据成为了可能.对于时间序列数据领域来说,通常涉及到分类、异常检测和预测等多个研究问题,其中时间序列预测任务是其研究的重点.与普通的回归分析预测模型不同,时间序列模型更依赖于数据在时间维度上的先后顺序.时间序列预测包括连续型预测(数值预测或范围估计)与离散型预测(事件预测)等,其核心就是从过去时序数据中挖掘出规律,并利用其对将来的发展趋势做出估计.由于时间序列预测研究能够分析其蕴含的规律来预测未来发展趋势,为各行业的决策过程提供指导,因此具有很高的学术意义和应用价值.

对于时间序列预测问题来说,随着时序数据积累量增加和数据维度增长,其研究方法也不断得到改进.从最初的数学统计方法进展到机器学习的方法,再发展到深度学习的方法.由于自然语言处理本质上也就有时序性要求,随着深度学习在自然语言处理任务上取得的巨大成果,因此被广泛使用的深度学习技术也被借鉴到时间序列研究领域.基于深度学习的时间序列预测方法得到了快速的发展,因此有必要对其发展的历史和现状进行分析和讨论.

本文首先对于时间序列预测问题进行了讨论,然后对于其研究方法进行了梳理,针对基于深度学习的研究方法按照其不同的实现网络类型分别进行了分析和讨论,在文章最后对于时间序列预测技术的发展进行了进一步展望.

1.1 问题定义

时间序列预测问题希望通过目前既有的信息来预测未来的目标值 \hat{y}_{t+1} ,也就是说利用 t 时刻之前的历史数据预测 t 时刻之后的数据.其数学表述如公式(1)所示:

$$\hat{y}_{t+1} = f(y_{t-k:t}, x_{t-k:t}, s) \quad (1)$$

式中, $y_{t-k:t}$ 表示了时间序列从 $t-k$ 时刻到 t 时刻已经得到的观测值; $x_{t-k:t}$ 表示从 $t-k$ 时刻到 t 时刻的协变量输入; s 表示观测过程保持不变的静态因素,例如为被观测对象的型号和类别等; $f(\cdot)$ 是模型学习的预测函数.需要说明,式中 y 和 x 可以是单变量,也可以是多维向量.

1.2 单步预测和多步预测

对于单步时间序列来说,使用历史观测值 $y_{t-k:t}$ 和协变量 $x_{t-k:t}$ 作为输入变量,来预测下一个时间步的观测值 \hat{y}_{t+1} ,此种预测过程被称为单步预测.

然而在很多预测应用中,需要预测在未来一段时间内观测序列变化的趋势,比如某一区域在未来一段时间的用电量,元器件在一段时间的退化趋势,股票价格在未来时间段中的变化趋势等,这都需要对未来多个时间步的时间序列进行预测.通常使用历史观测值 $y_{t-k:t}$ 和 $x_{t-k:t}$ 作为输入,来预测未来 n 个时间步的预测值 $\hat{y}_{t+1:t+n}$.这种时间序列预测被称为多步预测.

1.3 多步迭代预测和多步序列预测

对于多步预测来说,采用迭代预测方式和一次生成多步的两种方式对于未来一段时间进行预测.

对于迭代预测方式的多步预测来说,图1中左边虚框部分表示了时间序列预测使用真实值 y 来进行训练的过程.如图1所示,预测模型在测试过程中将 t 时刻的预测值 \hat{y}_{t+1} 作为下一次预测值 \hat{y}_{t+2} 的输入,依次进行递归输入,最终输出未来迭代多步的预测值.理论上通过此种迭代的方式可以实现单步预测模型向多步预测模型的转化.然而由于每步预测值相对于真实值都会产生一定的误差,我们将预测值作为下一个时刻的输入时,模型经过长时间迭代预测会导致误差累积,因此过长的预测步长会导致预测精度变差,而无法正常使用.

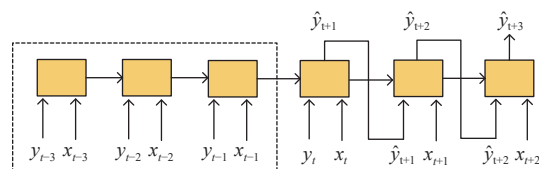


图1 迭代预测的多步预测示意图

Fig. 1 Iterative multi-step forecasting

另外一种预测方式是直接输出未来多个时间步的预测结果.对于预测模型来说,这是一个利用历史序列预测未来序列的过程(seq2seq问题).如图2所示,在预测的过程中,输入之前一段时间历史观测值和协变量,直接得到后续 n 步的预测值,此种预测可

以有效缓解迭代预测方法存在的误差累计问题,但是一般来说,这种方法受限于模型的计算能力,直接输出的步数也有限制。

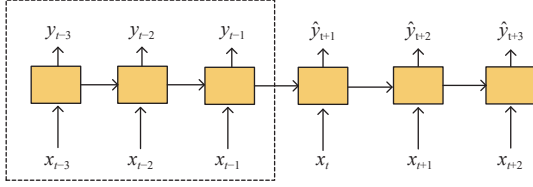


图2 seq2seq 多步预测示意图

Fig. 2 Seq2seq multi-step forecasting

1.4 时间序列预测的优化目标

针对不同场景,时间序列预测目标可以分成三种不同的类型:点估计、概率预测和区间预测。

点估计方法认为时间序列预测模型输出的预测值是一个固定的值。根据具体的问题不同,输出的预测值可以是离散输出的分类任务,例如预测未来设备的退化程度可以分为正常、一般和严重等几个类别,这需要网络预测出未来一段时间中退化类别的发生概率。输出的预测也可以是回归任务,例如预测未来设备的退化数值,这样可以定量地给出距离退化门限的差值。在深度学习中其优化目标,对于分类预测和回归预测可以分别用下面的公式(2)和公式(3)来表示。

$$Loss_{cls} = -\frac{1}{T} \sum_{i=1}^T y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \quad (2)$$

$$Loss_{reg} = \frac{1}{T} \sum_{i=1}^T (y_i - \hat{y}_i)^2 \quad (3)$$

概率预测方法认为时间序列本身就是一组随机过程,因此更倾向于估计出时间序列在未来分布,认为这样更加符合实际情况。在此种情况下,预测的目标主要获取到已知分布的参数,例如高斯分布(公式(4))或者负二项分布(公式(5))。

$$\begin{aligned} Loss_G(z|\mu, \sigma) &= (2\pi\sigma^2)^{-\frac{1}{2}} \exp(-(z-\mu)^2/(2\sigma^2)) \\ \mu(\mathbf{h}_{i,t}) &= \mathbf{w}_\mu^T \mathbf{h}_{i,t} + b_\mu \\ \sigma(\mathbf{h}_{i,t}) &= \log(1 + \exp(\mathbf{w}_\sigma^T \mathbf{h}_{i,t} + b_\sigma)) \end{aligned} \quad (4)$$

$$Loss_{NB}(z|\mu, \alpha) = \frac{\Gamma\left(z + \frac{1}{\alpha}\right)}{\Gamma(z+1)\Gamma\left(\frac{1}{\alpha}\right)} \left(\frac{1}{1+\alpha\mu}\right)^{\frac{1}{\alpha}} \left(\frac{\alpha\mu}{1+\alpha\mu}\right)^z$$

$$\begin{aligned} \mu(\mathbf{h}_{i,t}) &= \log(1 + \exp(\mathbf{w}_\mu^T \mathbf{h}_{i,t} + b_\mu)) \\ \alpha(\mathbf{h}_{i,t}) &= \log(1 + \exp(\mathbf{w}_\alpha^T \mathbf{h}_{i,t} + b_\alpha)) \end{aligned} \quad (5)$$

区间预测主要是考虑在决策时就必然要考虑到最好的情况和最坏的情况可能带来的风险,因此就需

要使用区间预测来描述预测值可能的上下限。即使对于具有非固定方差或非正态分布的残差,基于分位数的回归模型也可以提供合理的预测区间。损失函数根据所选分位数 γ 值对高估和低估的预测值给予不同的惩罚值,公式(6)列出了相应的损失函数, γ 的取值在 0 和 1 之间。

$$L_\gamma(y, y^p) = \sum_{i=y_i < y_i^p} (\gamma - 1) |y_i - y_i^p| + \sum_{i=y_i > y_i^p} (\gamma) |y_i - y_i^p| \quad (6)$$

2 时间序列预测方法

本章首先论述了深度学习出现之前的时间序列分析方式,接着着重探讨了基于深度学习的预测方法。实际上这两大类方法的思路有相通之处,深度学习方法也常会借鉴之前的方法作为子处理流程。

2.1 传统时序建模方法

传统时间序列预测方法主要是在确定时间序列参数模型的基础上,求解出模型参数,并利用求解出的模型完成未来预测工作。典型的方法如下:

自回归综合移动平均模型 ARIMA(Auto-Regressive Integrated Moving Averages) 是一种被广泛使用的时间序列预测统计方法^[4]。在处理过程中,首先需要对观测值序列进行平稳性检测,接着对其进行白噪声检验,计算 ACF(自相关系数)、PACF(偏自相关系数)进行检验,进而进行预测。基于此方法还有很多改进算法,如 SARIMA 和 SARIMAX 等^[5]。

Holt-Winters 法扩展了传统的 Holt 方法来捕捉季节性^[6]。Holt-Winters 方法包括预测方程和设置三个平滑方程中参数 α 、 β 和 γ 分别用于水平、趋势和季节性分量预测。

对于处理简单的时序预测问题,传统时序模型可以发挥较大的优势。但是如果变量和维度过多,变化模式过于复杂,传统时序模型就显得力不从心了。

2.2 基于时间序列分解的分析法

时间序列分解法是在时间序列方面一直是非常有用的方法,此方法认为一个时间序列往往是以下多种变化形式的叠加或耦合:

- 长期趋势 (Secular Trend, T): 指在较长时期内持续发展变化的整体趋势或者状态。
- 季节变动 (Seasonal Variation, S): 指由于季节变化引发的时间序列规则变化。
- 循环波动 (Cyclical Variation, C): 指以若干年(或周期)为期限,不具严格规则的周期性连续变动。
- 不规则波动 (Irregular Variation, I): 指由于各种偶然因素对时间序列发展造成的影响。

根据不同的应用场景,将时间序列按照加法原则和乘法原则进行了解析,加法模型的形式如下公式(7)所示,加法模型中的四种成分之间是相互独立的,某种成分的变动并不影响其他成分的变动.各个成分都用绝对量表示,并且具有相同的量纲.

$$X_t = T_t + S_t + C_t + I_t \quad (7)$$

乘法模型的形式如公式(8)所示,乘法模型中四种成分之间保持着相互依存的关系,一般而言,长期趋势用绝对量表示,具有和时间序列本身相同的量纲,其他成分则用相对量表示.

$$X_t = T_t \times S_t \times C_t \times I_t \quad (8)$$

采用时间序列分解的典型方法是 Facebook 公开提供了开源代码的 Prophet 模型^[7],其将时间序列分解为趋势项 $g(t)$,季节性项(周或月) $s(t)$,节假日项 $h(t)$ 以及噪音项 ϵ_t 相加.

时间序列分解的方法更重要的是提供了一种对于时间序列的分析思路.在后续很多机器学习和深度学习的方法都按照此种分解思路对于时间序列预测问题进行设计和规划.

2.3 基于机器学习的方法

时间序列数据预测工作本质上与机器学习方法中的回归分析之间存在着紧密的联系.其机器学习算法按照实现方法,主要分为以下几类.

SVM 以统计学习理论为基础,具有很好的泛化能力.同时 SVM 通过核函数方法向高维空间映射时并不增加计算的复杂性,又可以有效地克服维数灾难问题.将 SVM 用于时间序列预测函数回归称为支持向量回归(Support vector Regression,SVR),它对非线性时间序列存在稳定的预测能力^[8].SVR 通过非线性函数变换,把 x 映射到高维特征空间,在此高维特征空间中找到能够准确地表明输出数据及输入数据之间的函数 f ,通过回归分析来对时间序列进行拟合和预估.

梯度渐进回归树(Gradient Boosting Regression Tree,GBRT)是一种将梯度下降方法引入解决回归问题的方法.GBRT 算法为了损失函数的最小值,利用其负梯度进行计算和迭代,最终得到最佳模型^[9].设训练样本序列为 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$,训练 GBRT 算法的目的在于找到一个函数 $f(x)$,使得该函数对应的损失函数期望最低.渐进回归算法实际上是利用多个基函数的累加展开式对该函数进行逼近的方法.GBRT 算法对异常值的鲁棒性较高,也被广泛运用于时间序列预测方面.

隐马尔可夫模型(Hidden Markov Model,HMM)

是一种统计学模型^[10],它可以为多变量时间序列预测建模提供了一个概率框架.HMM 是双随机过程,具有一定状态数的隐马尔可夫链和显示随机函数集.前者不可观测,但可以通过产生一系列观测结果的一组过程来估计.对于给定模型 λ 和观测序列 y ,计算 $P(y|\lambda)$ 得到观察序列 y 的发生概率,选择最能解释观测结果的状态序列.HMM 也是一种最简单的动态贝叶斯网络,动态贝叶斯网络框架内的其他模型泛化了基本 HMM,允许更多隐藏状态变量.

2.4 基于深度学习的方法

随着深度学习方法在计算机视觉和自然语言处理方面取得的显著成就,深度学习的方法也逐渐引入到时间序列预测应用中.通过构建各种网络结构,深度神经网络能够更好的进行高维数据表征,从而减少了对手动特征工程和模型设计的需求,并通过定义损失函数,可以更加方便地进行端到端的训练.

深度神经网络通过多个非线性层来构建以往时间序列特征表示,来学习时间序列内部变化规律.在神经网络的学习过程中,可以认为将相关历史时序信息编码到隐变量 h_t 中,最终 \hat{y}_{t+1} 成为了隐变量 h_t 和当前输入 x_t 的输出,如公式(9)所示.

$$\begin{aligned} h_t &= g_{enc}(y_{t-k:t}, x_{t-k:t}, s) \\ \hat{y}_{t+1} &= f(h_t, x_t) \end{aligned} \quad (9)$$

针对不同的应用场景,在深度学习中隐藏层采用了不同的实现形式,从而采用了不同的神经网络架构来进行实现,通过调研对于基于深度学习的方法进行了分门别类.

2.4.1 基于卷积神经网络的方法

卷积神经网络(Convolutional Neural Network, CNN)在计算机视觉方面得到广泛应用,它能够提取跨空间维度不变的局部特征.同样它也可以在时间维度提取出局部特征,通过多层隐层逐渐在上一层进行汇总,从而获得过去时间段较长序列的隐藏信息,其原理如图3所示.

从图3中改进前的传统卷积连线可以看出,通过多层隐层汇聚,输入信息和最终输出预测结果的关联也非常有限.输出信息很难关联到之前更远的时间序列信息,因此导致传统卷积神经网络在时间序列预测能力表现有限.

为扩大隐藏层的感受范围,文献[11]是 DeepMind 提出了一种基于空洞因果卷积式的语音生产模型 WaveNet,后续专家们发现它也可被用于时间序列的预测,膨胀卷积可以解释为对于本层特征进行下采样的卷积,即降低分辨率以合并来自更久远的信息.因

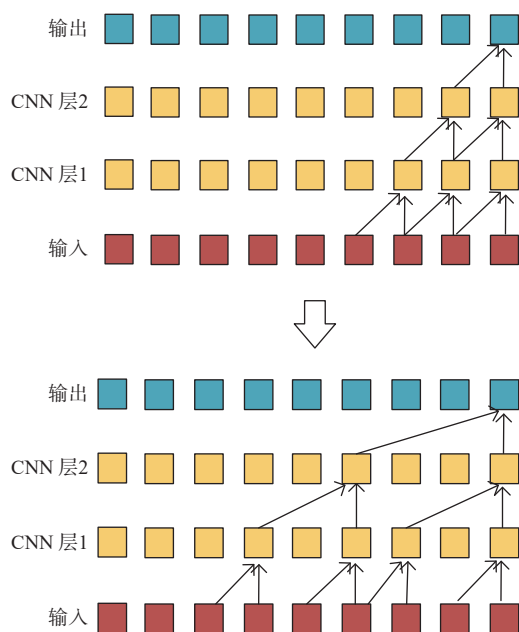


图3 基于 CNN 的改进前后时间序列预测原理示意图

Fig. 3 Time series forecasting based on CNN

此通过增加每一层的膨胀率,膨胀卷积可以逐渐聚合不同时间块的信息,从而有效地利用到时间序列中中长期的历史信息. 文献 [12] 基于此种思想在网络中构建了膨胀卷积的堆叠结构,能够允许汇聚更加久远的历史记录对于未来时序进行预测,如图 3 中改进后的卷积所示.

文献 [13-14] 提出的时间卷积网络 (TCN) 将序列看作一个一维对象帧,通过迭代多层卷积来捕捉长期关系. 时间卷积网络利用因果卷积和空洞卷积,以及残差卷积的跳层连接,它可以适应时序数据的时序性并可以提供更加宽广的时域感受野用于时序建模. 同时由于 TCN 网络结构允许并行计算输出,也提高了运行效率. 基于此思想,在时间序列预测领域还出现了一系列基于 TCN 网络的改进网络.

文献 [15] 提出的 SCINet 采用层次化卷积网络结构,以不同的时间分辨率迭代地提取和汇聚特征从而学习具有增强可预测性的有效表示. 整个 SCINet 的结构是一个二叉树,在每个 SCI-Block 中,时间序列被分为两个部分,随着二叉树深度的增加,更精细的时序信息也会被提取出来. 通过这样的方式,时间序列中短期和长期的依赖关系都能被网络学习到.

2.4.2 基于循环神经网络的方法

循环神经网络 (Recurrent Neural Network, RNN) 一直被用于序列建模,在各种自然语言处理任务上取得了很好的效果. 在其核心为 RNN 网络可以学习到预测前所有时间序列内部隐藏状态,作为过去信息的

特征表征,并结合当前的输入,给出下一步预测的结果,如图 4 所示. 在每个时间步,可以使用新的观测值不断地递归更新隐层状态. 因此在深度学习方法中基于 RNN 的网络架构也是最早被应用于时间预测场景.

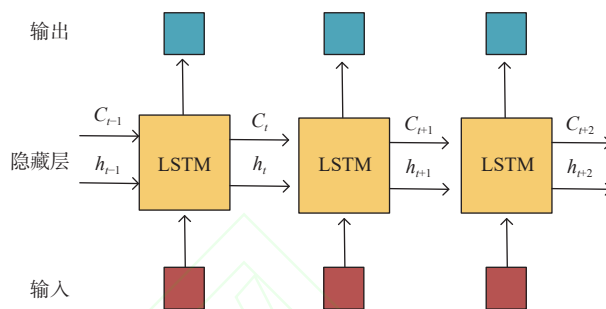


图4 基于 LSTM 的时间序列预测原理示意图

Fig. 4 Time series forecasting based on LSTM

早期的 RNN 变体在学习数据中的长期串行依赖关系时,可能会受到梯度爆炸和消失问题的限制. 解决方法之一是采用长短期记忆 (Long Short-term Memory, LSTM) 网络^[16]. 长短期记忆网络的设计灵感来自于计算机的逻辑控制门. 采用各种门控制记忆元,例如使用输出门来控制从单元中输出序列;使用输入门决定何时将数据读入单元;使用遗忘门来管理重置单元的内容. LSTM 通过训练得到的权重可以决定什么时候记忆或忽略隐状态中的输入. 因此, LSTM 和其变体 GRU 单元成为了基于 RNN 时间序列预测的重要组成部分.

DeepAR^[17] 网络采用 LSTM 模型解决时间序列预测问题. 在训练阶段,每个时刻传入上一时刻的真实值和协变量输入,经过 LSTM 单元后预测下一个时刻的值. 在预测阶段,将模型在上一个时刻的预测值作为输入,来替代训练过程中上一个时刻真实值的输入.

文献 [18] 提出了以循环网络为基础的 Deep State Space 模型. 该模型基于状态空间转换的思路基础上. 预估模型建模了连续两个隐状态的关系,实现了从当前的隐状态到当前时刻预估结果的预测. 相比之前的方法,在预测过程中不需要输入上一个时刻的真实值或预测值,由隐状态建立连续两个时刻的联系,解决了训练和预测不一致的问题.

考虑到训练模型时往往需要通过滑动窗口的形式产生训练集样本,使得数据进行了大量复制,占用了大量内存资源. 针对长期预测预测的不足,文献 [19] 提出了交叉式训练方法,在不复制数据的情况下,达

到滑动窗相似的训练效果,并很好能够很好地兼顾短期和长期预测。

文献[20]提出了一种能结合指数平滑和循环神经网络的混合模型,此模型兼容了上述各自的建模优势,有助于提升时间序列预测性能,取得了良好效果。

2.4.3 基于自注意力网络的方法

自注意力网络最早产生于自然语言处理任务,Transformer 是 2017 年论文《Attention is All You Need》提出的自注意力模型^[21],这篇论文针对机器翻译场景做了实验,实现了当时最好的效果,并且由于编码器部分采用并行计算方式,训练和推理时间被大大缩短,为序列化信息处理开拓了新的思路。目前在 NLP 各业务主流的语言模型如 GPT,BERT 等,都是

基于 Transformer 模型。由于与自然语言处理任务的相似性,因此 Transformer 的方法也很快能应用到时间序列预测任务上。

在 RNN 中,分析时间序列需要按照依次迭代的顺序处理从 $t-n$ 到 t 的所有信息,而当两个时间点相距较远时,他们之间隐层联系通常较差,且由于其顺序性处理效率也较低。基于注意力机制 (Attention Mechanism) 的模型,可以实现输入时间序列中每个单元之间的两两关联,对于下层特征中通过关联权重加权的表征,提供给上一层进行处理。自注意力机制可以更好地实现时序的上下文信息交互。如图 5 所示。

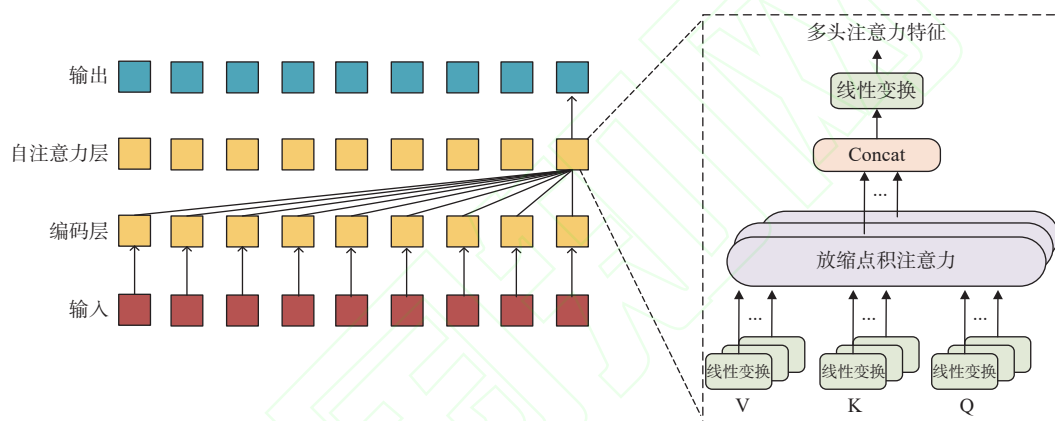


图 5 基于 Transformer 的时间序列预测原理示意图

Fig. 5 Time series forecasting based on Transformer

文献[22]采用了和 GPT 相似的 Transformer 结构尝试了时间序列预测任务,取得较好效果。Transformer 模型具有提高预测能力的潜力。但同时 Transformer 也存在诸如计算量大、内存占用高和编码器-解码器架构等限制,使其无法直接应用于更长时间序列预测问题。文献[23]通过在自注意力层中使用因果卷积生成查询 (Queries) 和键 (Key) 来提出卷积自注意力。它在自注意力模型中引入了 Log 稀疏掩码,将模型的计算复杂度从 $O(L^2)$ 降低到 $O(L \log L)$ 。

Informer^[24] 根据查询 (Queries) 和键 (Key) 的相似性,选择 $O(\log L)$ 占主导地位的查询,从而在计算复杂度上实现与 LogTrans 类似的改进。它还通过生成式解码器来直接产生长期预测序列,从而避免在使用单步前向预测进行长期预测时出现累积误差。

另一方面,学者们在时间序列建模中探索频域自注意力机制。Autoformer^[25] 设计了一种短期趋势分解架构,其自相关机制用作注意力模块。不同于之前注

意力机制,它测量输入信号之间的时延相似性,并聚合前 k 个相似的子序列以产生具有 $O(L \log L)$ 复杂度的输出。论文 FEDformer^[26] 通过傅里叶变换和小波变换在频域中应用注意力操作。它通过随机选择固定大小的频率子集来实现线性复杂度。

2.4.4 基于图神经网络的方法

多元时间序列预测中其变量相互依赖,但是现有方法无法完全利用变量对之间的潜在空间依赖性。近年来图神经网络 (Graph Neural Network, GNN) 在处理关系依赖方面表现出了很强的能力。一般先通过图结构学习层定义信息传播的图结构,在其上再进行卷积操作,通过以捕获时间序列内的空间和时间依赖性,对于多变量的时间序列进行预测,如图 6 所示。

文献[27]提出了为多元时间序列数据设计的通用图神经网络框架。通过图学习模块自动提取变量之间关系,将变量附带整合到其中。在网络中图卷积模块与时间卷积模块相互交错,分别捕获空间和时间

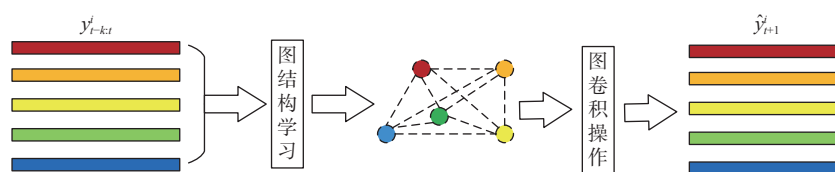


图 6 基于 GNN 的时间序列预测原理示意图

Fig. 6 Time series forecasting based on GNN

相关性. 使用混合跳跃传播层捕获时间序列内的空间和时间依赖性.

文献 [28] 在多变量时间序列预测中使用图神经网络, 在处理过程中将采用 N 个节点的全连接图, 按照内部相关性映射为 K 个节点连接图, 减少了图神经网络的节点规模, 将图模型中复杂程度由 $O(N^2)$ 转化为 $O(NK)$, 实现精度和计算复杂度之间的平衡.

文献 [29] 提出了一种新颖的图神经网络架构 Graph WaveNet, 用于时空图建模. 通过引入了自适应的图建模方法并通过节点嵌入来学习, 可以精确地捕获数据中隐藏的空间依赖性, 同时空洞因果一维卷积模型来替代 RNN 模型, 更容易提取长期时序关系.

2.4.5 基于残差的全连接网络方法

文献 [30] 提出 Nbeats 模型, 该模型内部结构中并没有 RNN、CNN 或 Attention, 网络全部为全连接组

成, 在时间序列预测方面也取得很好的效果.

Nbeats 的核心思路是通过多个全连接卷积块 (FC Blocks) 串联, 每个 FC Block 学习输入时间序列的一部分信息, 进入下一个 FC Block 的输入会去掉之前 Block 已经学到的信息, 后续的 Block 只关注学习于之前 Block 未学到的信息, 类似于 GBDT 不断拟合的思路. 在预测过程中将综合各个 Block 的预估结果得到最终预估结果. 通过这种方式, 模型每层更加关注前层无法正确拟合的残差, 也将时间序列进行了从趋势到细节的逐层分解.

Nbeats 的具体模型结构如图 7 所示, 在具体的实施过程中, 整个模型包括多个 Stack, 每个 Stack 包括多个 FC Block, 每个 FC Block 为最基础结构模块, 其由多个全连接层组成.

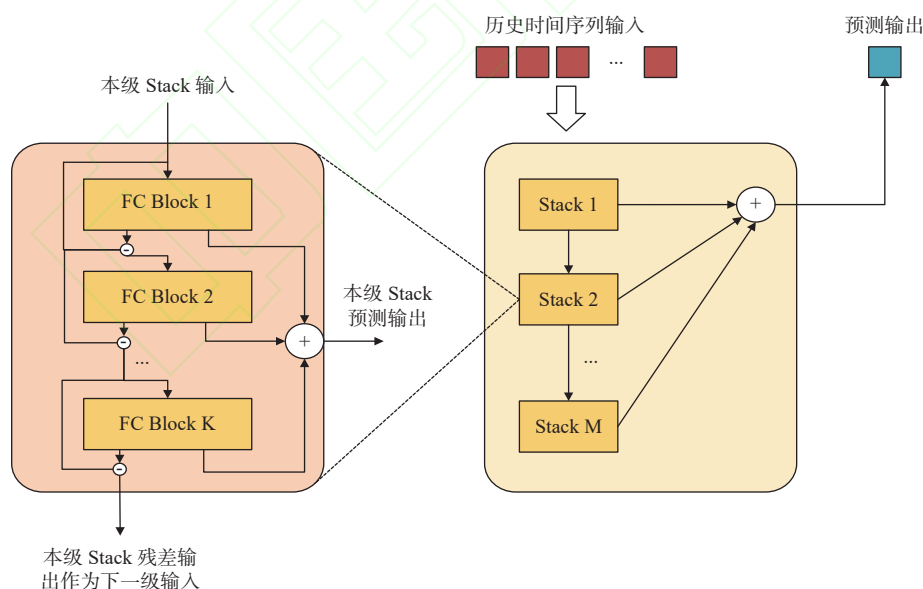


图 7 基于残差的全连接网络时间序列预测原理示意图

Fig. 7 Time series forecasting based on full residual network

为了进一步考虑到时间序列预测问题中诸如日期信息、节日信息、属性信息等外部特征. NBEATSx^[31] 又提出了可以引入外部特征的 Nbeatsx 网络, 和初版 Nbeats 的主要区别是引入了外部特征 x . 模型的主体

结构和 Nbeats 基本一致, 每个 Block 除了输出序列外, 还会输入外部特征, 二者一起通过全连接层参与到时间序列预测.

GAGA^[32] 将 Nbeats 扩展到了时空预测领域, 能

够处理存在空间关系的多个时间序列的建模.在先前的模型增加了 Time Gate 和 Graph Gate 等网络结构,嵌套到先前的网络中.

Nbeats 相比其他时间序列预测模型,独创了一种全部为全连接的主体网络架构,通过对于时间序列进行由粗到细的序列信息分解,不断进行细化,实现准确的时间序列预测.

3 进一步研究方向

虽然基于深度学习的时间序列预测方法取得很大发展,但未来的预测应用场景也对其提出了更高要求.这也为基于深度学习时间序列预测技术进一步研究指明了方向,主要体现在以下几点:

- 针对不同的使用领域,在深度学习模型中需要充分整合各自领域知识,从而获得更高的精度和效率;

- 目前时间序列的预测主要是通过序列内部的相关性分析实现,在未来应用中需要关注于因果性推断,其预测的可解释性要进一步说明.

- 在实际预测的过程中获得的数据通常存在异常点,在预测过程中需要忽略时间序列中的异常点,能够减小预测方法的误差,从而实现更好的鲁棒性;

- 目前深度学习方法主要是通过线下收集数据集进行批处理训练,在实际使用过程中需要通过增量学习的方法对时间序列数据进行在线分析预测.

上述方向的进一步研究可以使得基于深度学习的时间序列预测方法更加安全可靠、高效灵活地解决实际问题,让其能够更加快速落地于实际应用.

4 结束语

随着应用场景和技术的发展,时间序列预测方法从基于传统统计学方法逐渐发展到机器学习方法,直至目前以深度学习为代表的方法,追求如何实现对于时序序列预测更加精准和快速.随着深度学习在序列问题上的深入研究,也将不断推动着时间序列预测技术的发展.文中按照实现的网络类型,对于基于深度学习的时间序列方法进行了分门别类的详尽描述,梳理出发展思路,在此基础上对基于深度学习时间序列预测方法的进一步发展进行了展望.

参考文献:

[1] 次必聪, 张品一. 基于ARIMA-LSTM模型的金融时间序列预测[J]. 统计与决策, 2022, 38(11): 145-149. DOI: 10.13546/j.cnki.tjyjc.2022.11.029.

CI B C, ZHANG P Y. Financial time series forecast based on ARIMA-LSTM model[J]. Statistics & Decision, 2022, 38(11): 145-149. DOI: 10.13546/j.cnki.tjyjc.2022.11.029.

[2] 车畅畅, 王华伟, 倪晓梅, 等. 基于1D-CNN和Bi-LSTM的航空发动机剩余寿命预测[J]. 机械工程学报, 2021, 57(14): 304-312. DOI: 10.3901/JME.2021.14.304.

CHE C C, WANG H W, NI X M, et al. Residual life prediction of aeroengine based on 1D-CNN and Bi-LSTM[J]. Journal of Mechanical Engineering, 2021, 57(14): 304-312. DOI: 10.3901/JME.2021.14.304.

[3] 罗晓萌, 李建斌, 胡鹏. 基于时间序列预测的电子商务库存优化策略[J]. 系统工程, 2014, 32(6): 91-98.

LUO X M, LI J B, HU P. E-commerce inventory optimization strategy based on time series forecasting[J]. Systems Engineering, 2014, 32(6): 91-98.

[4] ZHANG G P. Time series forecasting using a hybrid AR-IMA and neural network model[J]. Neurocomputing, 2003, 50: 159-175. DOI: 10.1016/S0925-2312(01)00702-0.

[5] 李少雄, 李本光. 基于SARIMA模型和X-12-AR-IMA季节调整方法预测的比较[J]. 统计与决策, 2018, 34(18): 39-42. DOI: 10.13546/j.cnki.tjyjc.2018.18.008.

LI S X, LI B G. Comparison of prediction method based on SARIMA model and X-12-ARIMA seasonal adjustment method[J]. Statistics & Decision, 2018, 34(18): 39-42. DOI: 10.13546/j.cnki.tjyjc.2018.18.008.

[6] CHATFIELD C. The holt-winters forecasting procedure[J]. Journal of the Royal Statistical Society, 1978, 27(3): 264-279. DOI: 10.2307/2347162.

[7] TAYLOR S J, LETHAM B. Forecasting at Scale[J]. PeerJ Preprints, 2017, 5: e3190v2. DOI: 10.7287/peerj.preprints.3190v2.

[8] 陈荣, 梁昌勇, 谢福伟. 基于SVR的非线性时间序列预测方法应用综述[J]. 合肥工业大学学报(自然科学版), 2013, 36(3): 369-374. DOI: 10.3969/j.issn.1003-5060.2013.03.025.

CHEN R, LIANG C Y, XIE F W. Application of non-linear time series forecasting methods based on support vector regression[J]. Journal of Hefei University of Technology (Natural Science), 2013, 36(3): 369-374. DOI: 10.3969/j.issn.1003-5060.2013.03.025.

[9] ELSAYED S, THYSSENS D, RASHED A, et al. Do we really need deep learning models for time series forecasting?[Z]. arXiv: 2101.02118, 2021. DOI: 10.48550/arXiv.2101.02118.

[10] ZAHARI A, JAAFAR J. A novel approach of hidden

- Markov model for time series forecasting[C]// Proceedings of the 9th International Conference on Ubiquitous Information Management and Communication. Bali, Indonesia: ACM, 2015: 91. DOI: 10.1145/2701126.2701179.
- [11] VAN DEN OORD A, DIELEMAN S, ZEN H G, et al. WaveNet: A generative model for raw audio[C]//The 9th ISCA Speech Synthesis Workshop. Sunnyvale, CA, USA: ISCA, 2016: 125.
- [12] BOROVYKH A, BOHTE S, OOSTERLEE C W. Conditional time series forecasting with convolutional neural networks[Z]. arXiv: 1703.04691, 2017. DOI: 10.48550/arXiv.1703.04691.
- [13] BOROVYKH A, BOHTE S, OOSTERLEE C W. Dilated convolutional neural networks for time series forecasting[J]. Journal of Computational Finance, 2019, 22 (4): 73-101. DOI: 10.21314/JCF.2019.358.
- [14] CHEN Y T, KANG Y F, CHEN Y X, et al. Probabilistic forecasting with temporal convolutional neural network[J]. Neurocomputing, 2020, 399: 491-501. DOI: 10.1016/j.neucom.2020.03.011.
- [15] LIU M H, ZENG A L, CHEN M X, et al. SCINet: Time series modeling and Forecasting with sample convolution and interaction[Z]. arXiv: 2106.09305, 2021. DOI: 10.48550/arXiv.2106.09305.
- [16] HOSSAIN M S, MAHMOOD H. Short-term load forecasting using an LSTM neural network[C]// 2020 IEEE Power and Energy Conference at Illinois (PECI). Champaign: IEEE, 2020. DOI: 10.1109/PECI48348.2020.9064654.
- [17] 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. DOI: 10.1016/j.ijforecast.2019.07.001.
- [18] RANGAPURAM S S, SEEGER M W, GASTHAUS J, et al. Deep state space models for time series forecasting[C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montréal, Canada: Curran Associates Inc., 2018: 7796-7805.
- [19] WEN R F, TORKKOLA K, NARAYANASWAMY B, et al. A multi-horizon quantile recurrent forecaster[Z]. arXiv: 1711.11053, 2017. DOI: 10.48550/arXiv.1711.11053. DOI: 10.48550/arXiv.1711.11053.
- [20] SMYL S. A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting[J]. International Journal of Forecasting, 2020, 36 (1): 75-85. DOI: 10.1016/j.ijforecast.2019.03.017.
- [21] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, CA, USA: Curran Associates Inc., 2017: 6000-6010.
- [22] WU N, GREEN B, BEN X, et al. Deep transformer models for time series forecasting: the influenza prevalence case[Z]. arXiv: 2001.08317, 2020. DOI: 10.48550/arXiv.2001.08317.
- [23] LI S Y, JIN X Y, XUAN Y, et al. Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting[C]//Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver, BC, Canada: Curran Associates Inc., 2019: 471.
- [24] ZHOU H Y, ZHANG S H, PENG J Q, et al. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting[C]//Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Virtual Event: AAAI Press, 2021: 11106-11115.
- [25] WU H X, XU J H, WANG J M, et al. Autoformer: decomposition transformers with auto-correlation for long-term series forecasting[C]//Proceedings of the 35th International Conference on Neural Information Processing Systems. Virtual Event, Curran Associates, Inc., 2021, 34: 22419-22430.
- [26] ZHOU T, MA Z Q, WEN Q S, et al. FEDformer: Frequency enhanced decomposed transformer for long-term series forecasting[C]//International Conference on Machine Learning. Baltimore: PMLR, 2022: 27268-27286.
- [27] WU Z H, PAN S R, LONG G D, 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. Virtual Event, CA, USA: ACM, 2020: 753-763.
- [28] SATORRAS V G, RANGAPURAM S S, JANUSCHOWSKI T. Multivariate time series forecasting with latent graph inference[Z]. arXiv: 2203.03423, 2022. DOI: 10.48550/arXiv.2203.03423.
- [29] WU Z H, PAN S R, LONG G D, et al. Graph WaveNet for deep spatial-temporal graph modeling[C]//Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence. Macao, China: IJCAI. org, 2019: 1907-1913.
- [30] ORESHKIN B N, CARPOV D, CHAPADOS N, et al. N-BEATS: Neural basis expansion analysis for interpretable time series forecasting[C]//8th International Conference on Learning Representations. Addis Ababa, Ethiopia: OpenReview. net, 2020.
- [31] OLIVARES K G, CHALLU C, MARCJASZ G, et al. Neural basis expansion analysis with exogenous variables: Forecasting electricity prices with NBEATSx[Z].

arXiv: 2104.05522, 2021. DOI: 10.48550/arXiv.2104.05522.

- [32] ORESHKIN B N, AMINI A, COYLE L, et al. FC-GAGA: Fully connected gated graph architecture for spatio-temporal traffic forecasting[C]//Thirty-Fifth AAAI Conference on Artificial Intelligence. Virtual Event, AAAI Press, 2021: 9233-9241.

作者简介:

毛远宏 男,(1980-),博士,高级工程师.研究方向为深度学习,时间序列,计算机视觉.

孙琛琛 男,(1987-),博士,副教授.研究方向为时间序列分析,大数据分析、数据库技术.

徐鲁豫 男,(1983-),硕士,高级工程师.研究方向为计算机技术与管理.

刘 曦(通讯作者) 男,(1978-),硕士,研究员.研究方向宇航计算机. E-mail: liuxi_771@qq.com.

柴 波 男,(1963-),研究员.研究方向为专用计算机设计、计算机退化分析.

贺鹏超 男,(1988-),硕士,工程师.主要研究方向为图像处理、SAR 成像、人工智能.