

控制与决策

Control and Decision

一种基于深度学习的时间序列预测方法

鹿天柱, 钱晓超, 何舒, 谭振宁, 刘飞

引用本文:

鹿天柱, 钱晓超, 何舒, 等. 一种基于深度学习的时间序列预测方法[J]. *控制与决策*, 2021, 36(3): 645–652.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2019.0809>

您可能感兴趣的其他文章

Articles you may be interested in

融合稀疏编码与深度学习的草图特征表示

A feature representation of sketch based on fusion of sparse coding and deep learning

控制与决策. 2021, 36(3): 699–704 <https://doi.org/10.13195/j.kzyjc.2019.0941>

基于KPCA和G-G聚类的多元时间序列模糊分段

Fuzzy segmentation of multivariate time series with KPCA and G-G clustering

控制与决策. 2021, 36(1): 115–124 <https://doi.org/10.13195/j.kzyjc.2019.0849>

结合注意力机制的循环神经网络复述识别模型

Recurrent neural networks based paraphrase identification model combined with attention mechanism

控制与决策. 2021, 36(1): 152–158 <https://doi.org/10.13195/j.kzyjc.2019.0638>

基于改进堆叠自动编码器的循环冷却水系统工艺介质温度预测控制方法

Predictive control method of process medium temperature in circulating cooling water system based on improved stacked auto encoders

控制与决策. 2020, 35(12): 2835–2844 <https://doi.org/10.13195/j.kzyjc.2019.0694>

考虑时间序列的动态大群体应急决策方法

Dynamic large group emergency decision-making method considering time series

控制与决策. 2020, 35(11): 2609–2618 <https://doi.org/10.13195/j.kzyjc.2019.0088>

一种基于深度学习的时间序列预测方法

鹿天柱¹, 钱晓超², 何舒², 谭振宁¹, 刘飞^{1†}

(1. 华南理工大学软件学院, 广州 510006; 2. 上海机电工程研究所, 上海 201108)

摘要: 时间序列是一种广泛存在于现实各领域之中的海量高维数据, 时间序列预测是该领域的一个研究重点. 传统的时间序列预测方法仅仅从时间的维度对时间序列进行分析, 忽略了外界影响因素对时间序列可能产生的影响. 针对传统时间序列预测方法存在的问题, 提出一种基于深度学习的时间序列预测模型 DAFDC-RNN (dual-stage attention and full dimension convolution based recurrent neural network). 该模型引入目标注意力机制来学习输入特征与被预测特征之间的相关性, 引入全维度卷积机制来学习输入特征之间的相关性, 并引入时间注意力 (temporal attention) 机制来学习时间序列的长期时间依赖性. 在实验部分首先确定模型的超参数, 然后对模型部件的有效性进行验证, 最后通过对比实验验证了所提出的 DAFDC-RNN 模型在大特征量数据集上具有最佳的预测效果.

关键词: 时间序列; 预测分析; 深度学习; 目标注意力; 全维度卷积; 时间注意力

中图分类号: TP273

文献标志码: A

DOI: 10.13195/j.kzyj.2019.0809

开放科学(资源服务)标识码(OSID):



引用格式: 鹿天柱, 钱晓超, 何舒, 等. 一种基于深度学习的时间序列预测方法[J]. 控制与决策, 2021, 36(3): 645-652.

A time series prediction method based on deep learning

LU Tian-zhu¹, QIAN Xiao-chao², HE Shu², TAN Zhen-ning¹, LIU Fei^{1†}

(1. College of Software Engineering, South China University of Technology, Guangzhou 510006, China; 2. Shanghai Institute of Mechanical and Electrical Engineering, Shanghai 201108, China)

Abstract: Time series is a kind of high-dimensional data that exists widely in various fields in reality, and time series prediction is a research focus in research activities related to time series. The traditional time series prediction methods only analyze the time series from the time dimension, ignoring the influence of external influence factors on the time series. Aiming at the problems existing in traditional methods, this paper proposes a time series prediction model based on deep learning named dual-stage attention and full dimension convolution based recurrent neural network (DAFDC-RNN). The model introduces a target attention mechanism to learn the correlation between the input features and the predicted features, introduces a full dimension convolution mechanism to learn the correlation among input features, and introduces a temporal attention mechanism to learn the long-term temporal dependencies of time series. In the experimental part, we firstly determine the hyperparameters of the model, and then verify the validity of the model's components. Finally, the comparative experiments show that the proposed DAFDC-RNN model has the best prediction effect on the dataset with large features.

Keywords: time series; predictive analysis; deep learning; target attention; full dimension convolution; temporal attention

0 引言

为了解某种事物的发展规律, 人们通常会以一定的采样频率对该事物的某项指标进行观测, 这种按照时间先后顺序所产生的观测值集合就是时间序列. 对时间序列的历史观测值进行分析, 并推测出未来时间节点上值的过程称为时间序列预测. 时间序

列预测已经应用于众多领域之中, 具有重要的现实意义. 例如, 在生态环境领域, 可通过对日降水量进行预测来指导农业生产^[1]; 在医疗领域, 可通过对疾病发病率进行预测来预防疾病传播^[2]; 在经济领域, 可通过对房价进行预测来稳定国民经济和维护社会稳定^[3].

收稿日期: 2019-06-07; 修回日期: 2019-08-24.

基金项目: 国家重点研发计划项目(2018YFC0830900); 装备预研领域基金项目(61400010205); 上海航天科技创新基金项目.

责任编辑: 侯忠生.

†通讯作者. E-mail: feiliu@scut.edu.cn.

传统时间序列预测方法利用统计学知识对时间序列表现出的发展过程、方向和趋势进行建模,并以此模型为研究对象进行延伸和外推,最终达到预测的目的.常用的传统时间序列预测模型主要包括AR模型、MA模型、ARMA模型和ARIMA模型.然而,传统时间序列预测方法没有考虑外界影响因素和被预测对象间的因果关系,而是将所有影响因素综合作用的结果归结到时间上,仅从时间的维度对被预测对象建模.因此,当某些影响因素发生较大变化时,传统时间序列预测方法的预测精度会大大降低.

现代时间序列预测方法主要利用机器学习和深度学习技术对时间序列建模,并利用模型进行预测.通过将时间序列的影响因素特征化,现代时间序列预测方法可以捕获影响因素和被预测对象间的因果关系,克服了传统时间序列预测方法的弊端,大大提高了预测精度.在时间序列预测问题中常用的机器学习方法有支持向量机^[4-5]、贝叶斯网络^[6-7]、随机森林等^[8-9].

相比于传统的机器学习方法,深度学习方法无需人工提取特征,且在面对大量数据样本时,往往表现出更好的性能.因此,越来越多的研究者利用深度学习方法来解决时间序列预测问题,如BP神经网络^[10-11]、长段时记忆网络(LSTM)^[12-13]、卷积神经网络(CNN)^[14-15]等.同时,研究者们通过将一些新的理论和方法与已有神经网络模型进行结合来实现更好的预测效果,其中注意力机制使用最为广泛.例如,文献^[16]使用一种基于注意力机制的LSTM进行股价预测,大大提高了模型的预测效果.文献^[17]提出了一种基于双阶段注意力机制的循环神经网络模型DA-RNN,该模型基于编码器-解码器结构,在编码器中引入输入注意力(input attention)机制来自适应地选择关联度更高的特征,在解码器中引入时间注意力(temporal attention)机制来捕获时间序列的长期时间依赖.实验结果表明了DA-RNN具有优秀的预测能力,但DA-RNN模型仍存在3点不足:1)输入注意力机制中未考虑输入特征与被预测特征之间的相关性;2)未考虑输入特征之间的相关性;3)问题定义不符合实际情况.具体而言,假设时间步长为 T ,DA-RNN需要由1至 T 时刻的输入特征组成的时间序列以及1至 $T-1$ 时刻被预测特征组成的时间序列来预测出 T 时刻被预测特征的值.DA-RNN的问题定义存在的缺陷是:在实际时间序列预测应用场景下, T 时刻的输入特征值往往是获取不到的,而在DA-RNN中,该值是模型输入的一部分.

针对DA-RNN中存在的3个问题,本文提出一种基于DA-RNN的改进模型DAFDC-RNN(dual-stage attention and full dimension convolution based recurrent neural network).DAFDC-RNN综合考虑了输入特征与被预测特征之间的相关性、被预测特征之间的相关性以及时间序列的长期时间依赖性,同时修正了DA-RNN的问题定义中与实际应用场景不符的部分.实验部分首先确定模型的最佳超参数组合,接着对模型部件的有效性进行验证,最后通过对比实验对模型的预测性能进行验证.

1 DAFDC-RNN模型的构建

1.1 问题定义及符号表示

DAFDC-RNN模型的输入有1至 $T-1$ 时刻输入特征组成的多变量时间序列 X 和1至 $T-1$ 时刻被预测特征组成的单变量时间序列 Y . $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_t, \dots, \mathbf{x}_{T-1}) = (\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^k, \dots, \mathbf{x}^n)^T \in R^{n \times (T-1)}$, $\mathbf{Y} = (y_1, y_2, \dots, y_t, \dots, y_{T-1})^T \in R^{(T-1) \times 1}$, T 表示时间序列的步长, n 表示输入特征的特征数量; $\mathbf{x}_t \in R^{n \times 1}$,表示 n 个特征在 t 时刻的特征值集合; $\mathbf{x}^k \in R^{(T-1) \times 1}$,表示1至 $T-1$ 时刻,由第 k 个特征组成的时间序列; $y_t \in R$,表示被预测特征在 t 时刻的特征值.

DAFDC-RNN模型的目标是预测出 T 时刻被预测特征的特征值 y_T ,即需要学习一个非线性关系 F ,使得 F 满足如下公式:

$$y_T = F(\mathbf{X}, \mathbf{Y}). \quad (1)$$

1.2 模型描述

DAFDC-RNN模型使用编码器-解码器结构,在编码器中引入目标注意力(target attention)机制和全维度卷积(full dimension convolution, FDC)机制,在模型解码器中引入时间注意力(temporal attention)机制.其中:target attention用于学习输入特征与被预测特征之间的相关性,FDC用于学习输入特征之间的相关性,temporal attention用于学习时间序列的长期时间依赖性.解码器的输出是模型的预测目标.DAFDC-RNN模型的结构如图1所示.

1.2.1 目标注意力机制

某个特定时刻被预测特征的特征值是所有输入特征综合作用的结果,即目标特征与输入特征之间具有一定的相关性,并且通常情况下不同的输入特征与被预测特征之间的相关性大小也是不同的.例如,假设一个多变量时间序列记录了某城市的空气质量情况,输入特征包括空气中的含氧量、温度、气压、湿

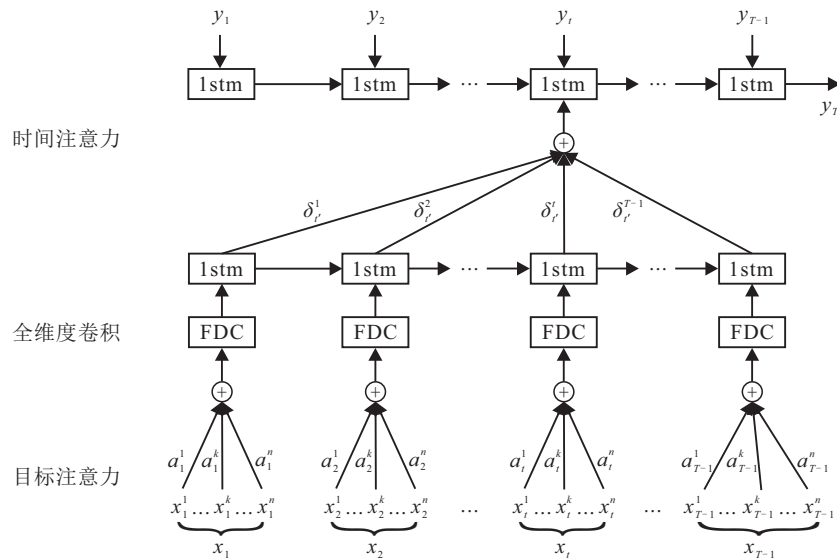


图1 DAFDC-RNN模型结构

度、PM2.5浓度、SO2浓度等,被预测特征是空气质量指数(AQI)。通常情况下,含氧量、温度、气压和湿度对空气质量指数的影响相对较小,而PM2.5浓度、SO2浓度对空气质量指数的影响相对较大。DAFDC-RNN模型通过引入target attention机制来学习输入特征与被预测特征间的相关性。

为了量化 t 时刻模型输入中的第 k 个特征与被预测特征的相关性大小,首先为第 k 个特征定义上下文向量 $\mathbf{context}_t^k$,有

$$\mathbf{context}_t^k = \tanh(\mathbf{W}_l[\mathbf{h}_{t-1}; \mathbf{s}_{t-1}] + \mathbf{U}_l \mathbf{x}^k + \mathbf{b}_l). \quad (2)$$

其中: \mathbf{W}_l 、 \mathbf{U}_l 和 \mathbf{b}_l 表示模型需要学习的参数, \mathbf{b}_l 表示偏置项, $\mathbf{W}_l \in R^{(T-1) \times 2m}$, $\mathbf{U}_l \in R^{(T-1) \times (T-1)}$, $\mathbf{b}_l \in R^{(T-1) \times 1}$, m 表示编码器隐藏层的大小; $[\mathbf{h}_{t-1}; \mathbf{s}_{t-1}]$ 表示target attention处理 \mathbf{x}_t 时编码器的状态。由于 $[\mathbf{h}_{t-1}; \mathbf{s}_{t-1}]$ 与 \mathbf{x}^k 以线性相加的方式进行关联, $\mathbf{context}_t^k$ 表示的是模型输入中第 k 个特征在编码器当前状态下的上下文表示。

为了将 $\mathbf{context}_t^k$ 与被预测特征组成的时间序列 \mathbf{Y} 之间建立关联,引入注意力机制中的点积评分函数

$$u_t^k = \mathbf{Y}^T \cdot \mathbf{context}_t^k, \quad (3)$$

其中 u_t^k 表示在编码器当前状态下,模型输入中第 k 个特征与被预测特征之间的相关性度量。最后引入Softmax函数将 u_t^k 转化为权重 a_t^k ,有

$$a_t^k = \frac{\exp(u_t^k)}{\sum_{i=1}^n \exp(u_t^i)}. \quad (4)$$

根据模型为不同特征分配的权重,可以得到 t 时刻target attention的输出 $\tilde{\mathbf{x}}_t$,有

$$\tilde{\mathbf{x}}_t = (\alpha_t^1 x_t^1, \alpha_t^2 x_t^2, \dots, \alpha_t^n x_t^n)^T. \quad (5)$$

为了描述方便,记 $c_t^k = \alpha_t^k x_t^k$ ($k = 1, 2, \dots, n$),此时 $\tilde{\mathbf{x}}_t$ 的表达式可进一步简化为

$$\tilde{\mathbf{x}}_t = (c_t^1, c_t^2, \dots, c_t^n)^T. \quad (6)$$

1.2.2 全维度卷积机制

对于一个多变量时间序列而言,不同输入特征之间可能存在着相关性。以1.2.1节提到的记录某城市空气质量的时间序列为例,温度、气压、湿度和PM2.5浓度都是输入特征,且温度、气压和湿度都会对PM2.5浓度产生影响,因此该时间序列的输入特征之间存在着相关性。时间序列数据可以看作是在固定时间间隔采样的一维网格数据,可以利用卷积自主学习时间序列数据之间的相互关系。然而,对于传统的带有卷积层、池化层和全连接层的CNN模型,需要手动设置卷积核的大小和步长、池化窗口的大小和步长、全连接层的神经元数量等参数,这些参数的设置大多依靠个人经验,随机性比较强,并且设置的参数是否可以充分学习特征之间的相关性,往往无法得到很好的解释。本文针对多变量时间序列在特定时刻的特征值集合,提出一种全维度卷积机制,该机制不仅能够克服传统CNN模型需要手动设置大量参数的弊端,同时可以充分学习输入特征之间的相关性。

全维度卷积(FDC)共分为3层,分别是卷积层、池化层和Concat层。卷积层共包括 n 个卷积核,维度分别为 $1, 2, \dots, n-1, n$,对应的 n 个特征图维度分别为 $n, n-1, \dots, 2, 1$ 。使用大小不同的卷积核是为了充分提取特征之间的相关性。池化层分别对 n 个特征图进行平均池化(average pooling),且池化窗口大小与特征图的维度一致,即任意一个特征图经过池化层处

理后的结果都是一个实数. Concat层将池化层产生的 n 个实数依次拼接, 拼接的结果就是 FDC 的最终输出. FDC 中使用 Concat 层代替传统 CNN 模型中的全连接层, 不仅仅是为了避免手动设置神经元数量等参数, 更是为了防止全连接层破坏输入特征之间的相关性. 下面对卷积层、池化层和 Concat 层中关键步骤的公式进行说明.

根据 t 时刻 FDC 的输入 $\tilde{\mathbf{x}}_t$ 和卷积层中第 p 个卷积核, 可以得到 t 时刻卷积层第 p 个特征图的第 i ($i = 1, 2, \dots, n-p$) 个值为

$$f_{(t,p)}^i = \gamma \left(\sum_{r=1}^p \text{kernel}(t, p, r) c_t^{i+r} + b_{(t,p)} \right). \quad (7)$$

其中: $\text{kernel}(t, p, r)$ 为 t 时刻卷积层第 p 个卷积核的第 r 个值; $b_{(t,p)}$ 为偏置项; c_t^{i+r} 为 $\tilde{\mathbf{x}}_t$ 中第 $i+r$ 个分量; γ 为激活函数, 这里选用 ReLU.

t 时刻卷积层第 k 个特征图经池化层处理后的结果为

$$f'_{(t,k)} = \frac{\sum_{i=1}^{n-k} f_{(t,k)}^i}{n-k}. \quad (8)$$

t 时刻池化层的输出分别为 $f'_{(t,1)}, f'_{(t,2)}, \dots, f'_{(t,n)}$. Concat 层将池化层的输出进行拼接后的结果为

$$\mathbf{j}_t = (f'_{(t,1)}, f'_{(t,2)}, \dots, f'_{(t,n)})^T. \quad (9)$$

1.2.3 时间注意力机制

随着输入序列长度的不断增加, 编码器-解码器结构的性能会迅速下降^[18]. 为了解决这一问题, DAFDC-RNN 模型在解码器中引入 temporal attention 机制, 自适应地选择相关的编码器隐藏层状态.

在任意时刻 t , 解码器会利用 temporal attention 计算出一个上下文向量 β_t , 有

$$g_t^z = \mathbf{v}_l^T \tanh(\mathbf{W}_l' [\mathbf{h}'_{t-1} \mathbf{s}'_{t-1}] + \mathbf{U}_l' \mathbf{h}_z + \mathbf{b}_l'), \quad (10)$$

$$\delta_t^z = \exp(g_t^z) / \sum_{i=1}^{T-1} \exp(g_t^i), \quad (11)$$

$$\beta_t = \sum_{z=1}^{T-1} \delta_t^z \mathbf{h}_z. \quad (12)$$

其中: \mathbf{v}_l 、 \mathbf{W}_l' 、 \mathbf{U}_l' 、 \mathbf{b}_l' 是模型需要学习的参数, \mathbf{b}_l' 是偏置项, $\mathbf{v}_l \in R^{q \times 1}$, $\mathbf{W}_l' \in R^{q \times 2n}$, $\mathbf{U}_l' \in R^{q \times m}$, $\mathbf{b}_l' \in R^{q \times 1}$, n 表示解码器隐藏层大小; $[\mathbf{h}'_{t-1} \mathbf{s}'_{t-1}]$ 表示解码器的当前状态, \mathbf{h}_z 表示 z 时刻编码器隐藏单元状态. 注意到 $[\mathbf{h}'_{t-1} \mathbf{s}'_{t-1}]$ 与 \mathbf{h}_z 以线性相加的方式关联, 因此 g_t^z 表示 \mathbf{h}_z 与当前时刻解码器状态的相关性度量. 式(11)中, g_t^z 通过 Softmax 函数转化为 δ_t^z , δ_t^z 是为

\mathbf{h}_z 分配的权重. 如式(12)所示, 将 1 至 $T-1$ 时刻编码器隐藏单元状态与隐藏单元对应的权重进行加权求和, 最终可以得到 t 时刻解码器的上下文向量 β_t .

1.2.4 模型输出

通过感知器将解码器的上下文向量 β_t 和被预测特征的特征值 y_t 建立关联, 可以得到 t 时刻解码器隐藏层的输入为

$$\tilde{y}_t = \tilde{\mathbf{w}}^T [\beta_t; y_t] + \tilde{b}. \quad (13)$$

其中: $\tilde{\mathbf{w}}$ 、 \tilde{b} 是模型需要学习的参数, \tilde{b} 是偏置项, $\tilde{\mathbf{w}} \in R^{(m+1) \times 1}$, $\tilde{b} \in R$. \tilde{y}_t 将作为解码器隐藏层的输入参与到隐藏层内部相关计算当中.

最后, 利用多层感知器将 $T-1$ 时刻解码器的上下文向量 β_{T-1} 和 $T-1$ 时刻解码器隐藏单元状态 \mathbf{h}'_{T-1} 建立关联, 可以得到模型的最终输出为

$$y_T = \mathbf{v}_y^T (\mathbf{W}_y [\mathbf{h}'_{T-1}; \beta_{T-1}] + \mathbf{b}_W) + b_v. \quad (14)$$

其中: \mathbf{v}_y 、 \mathbf{W}_y 、 \mathbf{b}_W 、 b_v 是模型需要学习的参数, \mathbf{b}_W 和 b_v 是偏置项, $\mathbf{W}_y \in R^{q \times (m+n)}$, $\mathbf{b}_W \in R^{q \times 1}$, $\mathbf{v}_y \in R^{q \times 1}$, $b_v \in R$.

1.3 算法的工作过程

首先对输入的每个时间步 \mathbf{x}_t 引入目标注意力机制, 计算当前时间步的所有输入特征与被预测特征的相关性, 为不同的输入特征分配相应的权重 α_t^k , 从而得到 t 时间步目标注意力的输出 $\tilde{\mathbf{x}}_t$.

然后引入全维度卷积机制, 对 t 时间步学习输入特征 $\tilde{\mathbf{x}}_t$ 之间的相关性. 全维度卷积引入 n 个卷积核 (卷积核大小分别为 $1, 2, \dots, n$), 分别对 $\tilde{\mathbf{x}}_t$ 进行卷积运算得到 n 个特征图; 通过对 n 个特征图作平均池化 (池化窗口大小与特征图维度一致), 得到 n 个实数; 将 n 个实数一次拼接得到全维度卷积的输出 \mathbf{j}_t . 全维度卷积输出 \mathbf{j}_t 作为编码器 (encoder) 的 t 时间步输入, 计算得到编码器 t 时间步的隐层状态 \mathbf{h}_t .

解码器的任意时刻通过引入时间注意力机制, 自适应地选择相关的编码器隐藏层状态 \mathbf{h} , 得到任意时刻的上下文向量 β ; 随后通过感知器将 t 时间步上下文向量 β_t 和 t 时间步被预测特征的特征值 y_t 建立关联, 可以得到 t 时刻解码器隐藏层的输入 \tilde{y}_t ; 最后解码器依次通过 $t-1$ 时间步的上下文向量 β_{T-1} 和 $t-1$ 时间步解码器隐藏单元状态 \mathbf{h}'_{T-1} 建立关联, 可以得到模型的最终输出 y_T .

1.4 算法效率分析

DAFDC-RNN 模型主要由目标注意力、全维度卷积和时间注意力 3 种机制来计算学习输入特征与被预测特征、输入特征之间和输入时间步之间的相关

性.

在目标注意力机制每个时间步长,计算每个特征上下文向量的浮点运算次数(FLOPs)为 $(T-1)(2m+1)$,点积评分函数的FLOPs为 $(T-1)$.因此目标注意力机制的 $T(n) = (T-1)n((T-1)(2m+1)+(T-1))$,即 $T(n) = O((T-1)^2)$.在全维度卷积机制每个时间步长,卷积计算输出的FLOPs为 $\sum_{i=1}^n i(n-i+1)$,因此全维度卷积机制的 $T(n) = (T-1)\sum_{i=1}^n i(n-i+1)$,即 $T(n) = O(n^2)$.在时间注意力机制每个时间步长,计算上下文向量的FLOPs为 $q(2n+m+1)$,因此时间注意力机制的 $T(n) = (T-1)q(2n+m+1)$,即 $T(n) = O((T-1)q(n+m))$.因此,如果主要考虑上述3种机制对DAFDC-RNN模型进行时间效率分析,可得 $T(n) = O((T-1)^2) + O(n^2) + O(q(n+m))$,即 $T(n) = O((T-1)^2) + O(n^2)$.

2 实验分析

为了对DAFDC-RNN模型的预测性能进行验证,本研究选取3个具有显著性差异特征的公开数据集:NASDAQ100(样本数为40560,特征数为81)、

appliances energy prediction(简称AEP,样本数为19735,特征数为27)、Beijing PM2.5(样本数为43824,特征数为6)进行相关实验,同时选取RMSE、MSE、MAPE作为模型的性能评价指标.此外,本文采用箱线图分析检测异常值,并用异常点前后的观测平均值进行代替,采用0-1归一化法对时间序列进行归一化,采用均值填充的方法对缺失值进行处理.

2.1 模型超参数

DAFDC-RNN模型中共有3个超参数,分别是时间步长 T 、编码器隐藏单元大小和解码器隐藏单元大小.为了实验方便,本文假定编码器隐藏单元大小和解码器隐藏单元大小相等,记为 H .因此,最终模型超参数实验需要确定 T 和 H 两个超参数.

给定 T 的取值集合为 $\{4, 6, 8, 10, 12, 14\}$, H 的取值集合为 $\{16, 32, 64, 128, 256\}$,理论上单个数据集需要进行 $30(6 \times 5)$ 次实验才可以确定 T 和 H 的最佳组合.本研究对超参数实验过程进行简化:当对 T 的取值进行实验时,固定 H 的值为64;当对 H 的取值进行实验时,固定 T 的值为10.最终单个数据集需要进行 $11(6+5)$ 次实验就可以确定 T 和 H 的最佳组合.实验结果如图2所示.

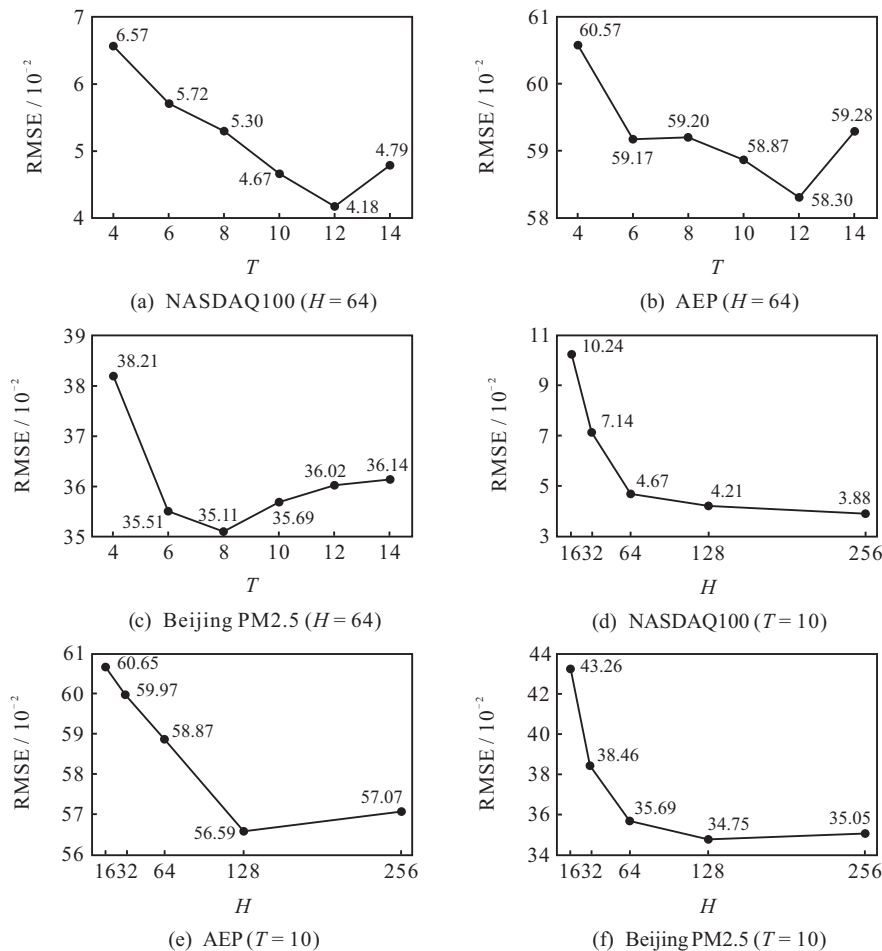


图2 DAFDC-RNN模型超参数实验结果

由图2(a)~图2(c)可知,对于时间步长 T ,在一定范围内随着 T 的增加,模型性能表现出逐渐上升的趋势,当到达一定值之后,随着 T 的增加,模型性能逐渐下降.这是由于在一定范围内随着 T 的增加,编码器可以对更多的输入数据进行编码,解码器可利用的输入信息也越来越多,模型的性能自然表现出上升的趋势.但是当 T 值超出一定范围之后,解码器无法完全捕获时间序列的长期时间依赖,导致模型性能逐渐下降.由图2(d)~图2(f)可知,对于隐藏单元大小 H ,随着 H 的增加,模型性能表现出上升的总体趋势.同时注意到,在AEP数据集和Beijing PM2.5数据集中, H 为256时的模型性能比 H 为128时的模型性能要差.这是由于AEP数据集和Beijing PM2.5数据集特征数量相对较少,而隐藏单元维度过大,最终导致过拟合.

由图2可知,对于NASDAQ100数据集,当 $T =$

12, $H = 256$ 时模型性能最佳;对于AEP数据集,当 $T = 12, H = 128$ 时模型性能最佳;对于Beijing PM2.5数据集,当 $T = 8, H = 128$ 时模型性能最佳.

2.2 模型部件验证

由上述可知,DAFDC-RNN模型有target attention、FDC和temporal attention三个核心部件.为了验证3个部件在模型中的有效性和作用,本节设计了如下4种部件组合方式:

- 1) target attention+temporal attention;
- 2) target attention+FDC;
- 3) temporal attention+FDC;
- 4) target attention+FDC+temporal attention.

将以上4种部件组合方式分别记为model I、model II、model III和model IV,分别在3个数据集上进行实验,实验结果如表1所示.

表1 模型部件验证实验结果

组合	NASDAQ100			AEP			Beijing PM2.5		
	RMSE	MSE	MAPE	RMSE	MSE	MAPE	RMSE	MSE	MAPE
model I	0.0429	0.0306	0.0123	0.5876	0.2815	1.8048	0.1938	0.1193	1.9655
model II	0.0465	0.0329	0.0132	0.6053	0.2785	1.8563	0.1987	0.1311	1.9807
model III	0.0800	0.0728	0.0287	0.6246	0.3036	1.6828	0.2021	0.1308	2.1035
model IV	0.0373	0.0276	0.0110	0.5753	0.2649	1.5888	0.1963	0.1242	1.9736

由表1可知,model IV在NASDAQ100和AEP数据集共6个评价指标上均取得了最佳效果.通过对比model I和model IV的部件组合方式,可以看出FDC在NASDAQ100和AEP数据集上均能提升模型性能.同理,通过对比model II和model IV以及model III和model IV的部件组合方式,可以看出temporal attention和target attention在NASDAQ100和AEP数据集上能够提升模型性能.然而,在Beijing PM2.5数据集中,model I的预测精度要高于model IV,这是由于Beijing PM2.5数据集特征较少,特征间相关性程度比较低,而FDC充分学习了输入特征之间的相关性.也就是说,FDC适合处理特征间相关性程度较高的时间序列,在特征较少或特征间相关性程度较低的时间序列中使用FDC反而会降低预测精度,因此是一种类似于过拟合的现象.同时,注意到在Beijing PM2.5数据集中,model IV的预测精度依旧高于model II和model III,表明temporal attention和target attention在Beijing PM2.5数据集中依旧可以提高模型的性能.

2.3 对比实验

为了进一步验证DAFDC-RNN模型的预测能力,本节选用以下3种对比模型与DAFDC-RNN模型的预测效果进行对比:1) ARIMA; 2) Encoder-Decoder (也称为Seq2seq); 3) DA-RNN.

需要注意的是,Encoder-Decoder和DA-RNN模型与DAFDC-RNN模型一样有时间步长 T 和隐藏单元大小 H 两个超参数.为了保证实验的公平性,同样需要确定Encoder-Decoder和DA-RNN模型两个对比模型的最佳超参数组合,实验结果如图3所示.

由图3可知:对于NASDAQ100数据集,当 $T = 10, H = 128$ 时,Encoder-Decoder和DA-RNN模型效果都是最佳.对于AEP数据集,当 $T = 10, H = 64$ 时,Encoder-Decoder模型效果最佳;当 $T = 10, H = 128$ 时,DA-RNN模型效果最佳.对于Beijing PM2.5数据集,当 $T = 10, H = 64$ 时,Encoder-Decoder模型效果最佳;当 $T = 8, H = 128$ 时,DA-RNN模型效果最佳.

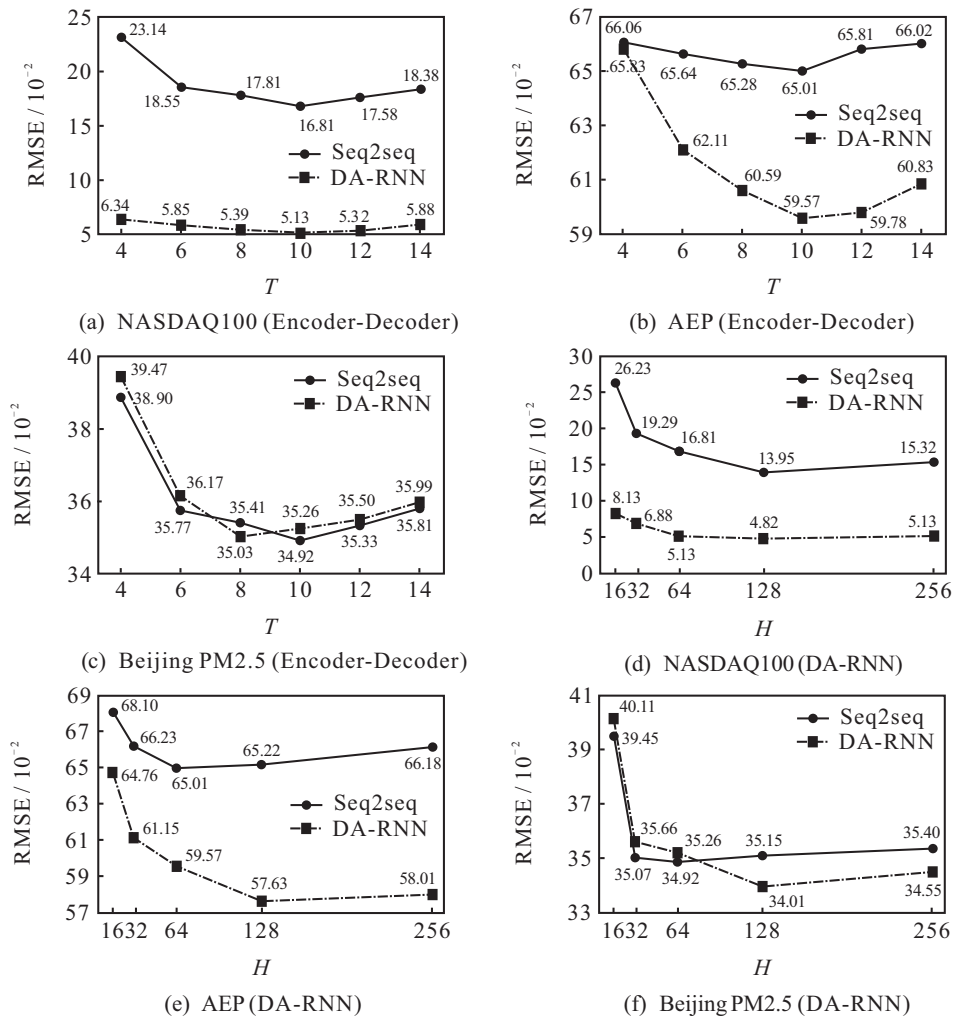


图3 Encoder-Decoder和DA-RNN模型的超参数实验结果

表2 对比实验结果

模型	NASDAQ100			AEP			Beijing PM2.5		
	RMSE	MSE	MAPE	RMSE	MSE	MAPE	RMSE	MSE	MAPE
ARIMA	2.3321	2.1563	1.9711	3.7432	2.4518	4.8931	3.3134	3.2259	3.9014
Encoder-Decoder	0.1310	0.1257	0.0496	0.6463	0.3625	2.2837	0.2314	0.1502	2.3497
DA-RNN	0.0452	0.0371	0.0148	0.5805	0.2967	1.7712	0.1943	0.1205	2.0078
DAFDC-RNN	0.0373	0.0276	0.0110	0.5753	0.2649	1.5888	0.1963	0.1242	1.9736

最终,所有模型在测试集上的实验结果如表2所示.在该实验中,将测试集和非测试集分别设为85%和15%.由表2可知,DAFDC-RNN模型在NASDAQ100和AEP数据集共6个评价指标上取得了最佳效果;在Beijing PM2.5数据集上只有MAPE指标取得最佳效果,在RMSE和MSE上的效果略差于DA-RNN,具体原因在2.2节已经提到,即由于Beijing PM2.5数据集特征数量比较少,而FDC充分学习了输入特征之间的相关性,在特征较少或者特征间相关性程度比较低的时间序列中使用FDC反而会降低预测精度,因此是一种类似于过拟合的现象.

3 结论

本文利用深度学习技术对时间序列的相关特性进行建模,提出了一种称为DAFDC-RNN的时间序列预测模型,并通过相关实验对DAFDC-RNN的预测能力进行了验证.实验结果表明,DAFDC-RNN在大特征量数据集下具有更加优秀的预测效果.该模型有望应用于现实各领域中的时序数据预测任务,具有一定的应用前景和推广价值.

参考文献(References)

[1] Das M, Ghosh S K. BESTED: An exponentially smoothed spatial bayesian analysis model for spatio-temporal

- prediction of daily precipitation[C]. International Conference on Advances in Geographic Information Systems. New York: ACM, 2017: 1-4.
- [2] El-Shafeiy E A, El-Desouky A I, El-ghamrawy S M. Prediction of liver diseases based on machine learning technique for big data[C]. International Conference on Advanced Machine Learning Technologies and Applications. Cairo: Springer, 2018: 362-374.
- [3] Wang J, Wun K Y, Kin L F. Gaussian process kernels for noisy time series: Application to housing price prediction[C]. International Conference on Neural Information Processing. Siem Reap: Springer, 2018: 78-89.
- [4] Chen T T, Lee S J. A weighted LS-SVM based learning system for time series forecasting[J]. Information Sciences, 2015, 299: 99-116.
- [5] Wang Y, Li L, Xu X. A piecewise hybrid of ARIMA and SVMs for short-term traffic flow prediction[C]. International Conference on Neural Information Processing. Guangzhou: Springer, 2017: 493-502.
- [6] Das M, Ghosh S K. semBnet: A semantic Bayesian network for multivariate prediction of meteorological time series data[J]. Pattern Recognition Letters, 2017, 93: 192-201.
- [7] 朱原媛, 杨有龙, 张恒伟. 基于贝叶斯网络的混沌时间序列预测[J]. 计算机工程与应用, 2012, 48(13): 100-104.
(Zhu Y Y, Yang Y L, Zhang H W. Prediction of chaotic time series based on Bayesian network[J]. Computer Engineering and Application, 2012, 48(13): 100-104.)
- [8] Lin L, Wang F, Xie X, et al. Random forests-based extreme learning machine ensemble for multi-regime time series prediction[J]. Expert Systems with Applications, 2017, 83: 164-176.
- [9] 李婉华, 陈宏, 郭昆. 基于随机森林算法的用电负荷预测研究[J]. 计算机工程与应用, 2016, 52(23): 236-243.
(Li W H, Chen H, Guo K. Research on electricity load forecasting based on stochastic forest algorithms[J]. Computer Engineering and Application, 2016, 52(23): 236-243.)
- [10] Ma L, Tian F. Pneumonia incidence rate predictive model of nonlinear time series based on dynamic learning rate BP neural network[C]. Fuzzy Information and Engineering. Huludao: Springer, 2010: 739-749.
- [11] 季芳, 熊坚, 杨文臣. 基于粒子群优化BP网络的山区高速公路交通事故预测模型[J]. 电子科技, 2018, 31(10): 68-72.
(Ji F, Xiong J, Yang W C. Traffic accident prediction model of mountain expressway based on particle swarm optimization BP network[J]. Electronic Technology, 2018, 31(10): 68-72.)
- [12] Namini S S, Tavakoli N, Namin A S. A Comparison of ARIMA and LSTM in Forecasting Time Series[C]. International Conference on Machine Learning and Applications. Orlando: IEEE, 2018: 1394-1401.
- [13] Liu X, Liu Q, Zou Y, et al. A self-organizing LSTM-based approach to PM2.5 forecast[C]. International Conference on Cloud Computing and Security. Haikou: Springer, 2018: 683-693.
- [14] Koprinska I, Wu D, Wang Z. Convolutional neural networks for energy time series forecasting[C]. International Joint Conference on Neural Networks. Rio: IEEE, 2018: 1-8.
- [15] Binkowski M, Marti G, Donnat P. Autoregressive convolutional neural networks for asynchronous time series[C]. International Conference on Machine Learning. Stockholm: ACM, 2018: 579-588.
- [16] Cheng L C, Huang Y H, Wu M E. Applied attention-based LSTM neural networks in stock prediction[C]. International Conference on Big Data. Seattle: IEEE, 2018: 4716-4718.
- [17] Qin Y, Song D, Chen H. A dual-stage attention-based recurrent neural network for time series prediction[C]. International Joint Conference on Artificial Intelligence. Mecbourne: AAAI Press, 2017: 2627-2633.
- [18] Cho K, Merrienboer B V, Bahdanau D, et al. On the properties of neural machine translation: Encoder-decoder approaches[C]. Conference on Empirical Methods in Natural Language Processing. Doha: ACL, 2014: 103-111.

作者简介

鹿天柱(1993—), 男, 硕士, 从事深度学习和时间序列分析的研究, E-mail: scut_lu@163.com;

钱晓超(1986—), 男, 工程师, 博士, 从事体系建模与评估的研究, E-mail: everqxc@hotmail.com;

何舒(1992—), 女, 助理工程师, 硕士, 从事装备体系设计与评估的研究, E-mail: 15921393507@163.com;

谭振宁(1994—), 男, 硕士, 从事深度学习和数据挖掘的研究, E-mail: 18676643646@163.com;

刘飞(1976—), 男, 副教授, 博士生导师, 从事建模与仿真、人工智能等研究, E-mail: feiliu@scut.edu.cn.

(责任编辑: 齐 霖)