

用于多元时间序列预测的自适应频域模型

王晓迪^{1,3} 刘鑫^{2,3} 于晓^{2,3}

1 山东财经大学财政税务学院 济南 250014
2 山东财经大学计算机科学与技术学院 济南 250014
3 山东省数字媒体技术重点实验室 济南 250014

摘要 近年来,学术和工业领域对时间序列数据的研究热潮不断增长,但其中蕴含的频率信息仍缺乏有效的建模。研究发现,时间序列预测依赖于不同的频率模式,为未来的趋势预测提供有用的线索:短期的序列预测更多依赖于高频分量,而长期预测则更多关注低频数据。为更好地挖掘时间序列的多频模式,提出了一个多特征自适应频域预测模型 MAFD。该模型分为两个阶段:在第一阶段中,模型通过 XGBoost 算法对输入向量进行重要性度量,选择高重要性特征;在第二阶段,模型将时间序列的频率特征提取和目标序列的频域建模集成到一起,并根据时间序列对频率模式的依赖特点构建一个端到端的预测网络。MAFD 的创新性体现在预测网络能够根据输入序列的动态演变自动关注不同的频率分量,从而揭示时间序列的多频模式,强化模型的学习能力。采用 4 种不同领域的数据集对模型进行了性能验证,实验结果表明,与现有经典的预测模型相比,MAFD 具有更高的准确性和更小的滞后性。

关键词: 时间序列预测;XGBoost;多频模式;深度学习;自适应建模

中图法分类号 TP391

Adaptive Frequency Domain Model for Multivariate Time Series Forecasting

WANG Xiao-di^{1,3}, LIU Xin^{2,3} and YU Xiao^{2,3}

1 School of Public Finance and Taxation, Shandong University of Finance and Economics, Jinan 250014, China
2 School of Computer Science and Technology, Shandong University of Finance and Economics, Jinan 250014, China
3 Shandong Key Laboratory of Digital Media Technology, Jinan 250014, China

Abstract In recent years, the research enthusiasm for time series data in academic and industrial fields has been increasing, but the frequency information contained in it still lacks effective modeling. The studies found that time series forecasting relies on different frequency patterns, providing useful clues for future trend forecasting: short-term series forecasting relies more on high-frequency components, while long-term forecasting focuses more on low-frequency data. In order to better mine the multi-frequency mode of time series, this paper proposes a multi-feature adaptive frequency domain prediction model (MAFD). MAFD is composed of two stages. In the first stage, it uses XGBoost algorithm to measure the importance of the input vector and selects high-importance features. In the second stage, the model integrates the frequency feature extraction of the time series and the frequency domain modeling of the target sequence, and builds an end-to-end prediction network based on the dependence of the time series on the frequency mode. The innovation of MAFD is reflected in the predictive network's ability to automatically focus on different frequency components according to the dynamic evolution of the input sequence, thereby revealing the multi-frequency pattern of the time series and strengthening the learning ability of the model. This work uses 4 datasets from different fields to verify the performance of the model. The experimental results show that compared with the existing classic prediction models, MAFD has higher accuracy and smaller lag.

Keywords Time series prediction, XGBoost, Multi-frequency pattern, Deep learning, Adaptive modeling

1 引言

时间序列是按时间索引顺序排列的一系列数据点^[1],其分析方法可分为时域分析和频域分析两大类。时域分析方法将时间序列数据视为有序点序列并分析它们之间的相关性,例如隐马尔可夫模型^[2]和自回归积分滑动平均 (ARIMA) 模型^[3]。频域分析方法使用变换算法 (例如离散傅里叶变换^[4]

和经验模态 (EMD) 分解^[5-6]) 将时间序列变换为频谱,并将其作为分析原始序列的特征。频域预测方法能够引导研究者深入到时间序列的内部,挖掘时间序列的确定性特征及运动规律。然而,目前的频域分析方法主要存在两个问题:1) 主要为单特征建模,对已有数据利用不充分,忽略了相关特征的作用;2) 频域特征提取和预测是分离的,这种松弛连接的参数无法进行全局优化,导致频域信息之间的相互作用及时间序列

基金项目:国家自然科学基金 (62072274)

This work was supported by the National Natural Science Foundation of China (62072274).

通信作者:王晓迪 (20049447@sdufe.edu.cn)

的频率依赖模式难以利用。为解决上述问题,本文提出一种多特征自适应频域预测模型(Multi-features Adaptive Frequency Domain Prediction Model,MAFD),以准确挖掘时间序列的频域模式。本文的主要贡献如下:

- (1)针对问题 1),本文利用 XGBoost 对时序特征进行重要性度量,选择多维有效特征,避免因输入特征太少产生较大的预测误差,或因输入特征太多使模型训练时间过长等问题。
- (2)针对问题 2),将时频转换的思想无缝嵌入到深度学习框架中,并根据时间序列的频率依赖特点提出一种端到端的时间序列频域预测网络。预测网络具有追踪频域信息的动态模式、捕捉重要分量、挖掘频域信息模式依赖性的能力。
- (3)将 XGBoost 和神经网络软连接,构建多特征自适应频域预测模型,并在个不同领域时间序列数据集上对 MAFD 进行评估。实验结果表明,与现有的经典预测模型相比,MAFD 通过对时间序列的多频模式建模,获得了更具竞争力的性能。

2 相关工作

时间序列预测是通过分析数据本身和其相关特征来预测未来数据,被广泛应用于临床医学^[7]、金融预测^[8]、气象预测^[9-10]、人类行为预测^[11]等领域。与其他预测模型不同,时间序列受自然因素、人为因素的影响,通常具有强烈的波动性和非稳定性,这给未来趋势的预测增加了难度。因此,研究时间序列的有效预测方法可以更好地利用现有信息进行分析 and 决策,有助于把握数据运动规律^[12],降低各种不稳定因素产生的负面影响。性能极佳的时间序列预测模型弥足珍贵。

时间序列虽然呈现复杂特性,但是数据流变化具有较强的自相关性、长期记忆性,并且和其他时间序列有着较强的相关性,这为时间序列的多特征预测奠定了夯实的基础。

在时序分析与预测领域,各种各样的统计模型和神经网络模型都已经应用其中,并取得一些进展。自回归移动平均模型(ARMA)^[13]采用线性回归估计模型参数进行预测。但是,时间序列通常是高度非线性和非平稳的,这限制了 AR-MA 的实际应用。随着机器学习技术的发展,研究者根据大量的历史数据建立非线性预测模型,例如梯度增强决策树(GBDT)^[14]和卷积神经网络(CNN)^[15]。二者通过对数据的迭代训练逐渐逼近真实数据,可以获得比传统的统计模型更准确的预测结果。然而,这种预测方式并没有体现出时序数据的数据依赖特点。1986 年,Jordan 等人提出了在处理时间序列预测方面具有独特优势的循环神经网络(RNN)^[16]。RNN 可以更好地表达时间序列的依赖特点,预测序列的未来趋势并拟合未来数据,但当其隐藏层较多时容易产生梯度消失和梯度爆炸问题。1997 年,长短期记忆神经网络(LSTM)^[17]应运而生。LSTM 是 RNN 的一个优秀变种模型,通过增加细胞状态的连接更好地利用时间序列上下文的历史信息,在时序分析中发挥较强的适应性^[18]。但是,LSTM 的不足之处在于没有揭示时间序列的多频特征,无法对时序数据的频域建模^[19]。为此,Mouatadid 团队将频域变换算法用作 LSTM 前的数据预处理工具进行频域预测,这种松弛连接的方式虽然能提高原始神经网络模型的性能,但无法将参数进行全局优

化^[20]。2017 年,Zhang 等利用傅里叶变换将 LSTM 存储单元的隐藏状态分解为多个频率成分,并以潜在交易模式的特定频率为基础模拟股票开盘价的波动。但是,傅里叶变换处理非平稳时间序列时往往产生时频局部化的矛盾^[21]。与傅里叶变换相比,小波变换通过引入可变的尺度因子和平移因子,对时间序列进行多尺度多方面的细化,能自动适应时间序列分析的要求,解决了傅里叶变换的困难问题^[22]。因此,本文将小波变换嵌入到长短期神经网络中搭建端到端的神经网络。而时间序列预测依赖于不同的频率成分,将小波变换简单嵌入到 LSTM 中无法满足这种需求。为此,本文在小波变换和 LSTM 的基础上引入了卷积层和自适应层,提出并设计了一种新颖的频域预测网络。卷积层的优势在于捕捉相邻输入的交互用于自动学习特征表示,提高模型的学习能力;而自适应层的作用是根据数据的动态演变,将学习到的权重向量自适应地赋予频率分量,突出不同的频率分量对预测目标的影响。

与此同时,随着大数据时代的到来,时间序列的时域预测逐渐从单特征预测发展为多特征预测,而近年来提出的端到端的频域预测模型仍为单特征预测,这极大地限制了预测性能。然而,将时间序列全部特征输入会增加模型复杂度,大量不相关和冗余特征也会影响模型的性能,如何对输入特征进行合理选择和降维也成为重要的研究课题。本文在对目前已有的特征选择模型充分研究的基础上,引入极端梯度提升(XGBoost)^[23]进行特征选择,该策略已经在电力负荷预测^[24]、信号处理^[25]等领域得到成功应用。与其他选择模型相比,XGBoost 在时序特征提取方面的主要优势在于能够自动利用 CPU 的多线程并行地计算增益,按照特征粒度寻找最佳分割点,达到特征重要性度量的目的^[26]。

综上所述,本文将 XGBoost 和预测网络软连接,提出一个多特征自适应频域预测模型 MAFD。首先,MAFD 利用 XGBoost 算法进行特征性重要度量,提取特征子序列;然后,运用频域预测网络挖掘序列的数据变化趋势,对特征的频域建立自适应模型;最后,将时序数据集输入 MAFD 模型中,复杂任务的实验结果表明,MAFD 可以跟踪数据频域趋势生成可靠的预测结果。

3 MAFD 模型

3.1 问题阐述

时间序列的单特征预测是基于前 T 时刻的历史观测数据 $y_1, y_2, \dots, y_T$ 生成 N 步预测值 $\tilde{y}_{T+N}$ 。但是,观测值不仅与自身历史数据有关,还受外源特征序列的影响,这就产生了时间序列的多特征预测。

图 1 为利用外源特征序列和观测序列进行时间序列预测的多特征建模方法。其中图 1(a)和图 1(b)将外源特征数据和观测数据输入模型生成预测值。二者不同的是图 1(b)考虑了数据间的时间动态,利用数据的时间关系建模。图 1(c)为目前常用的时序建模方法^[27],该方法利用外源特征序列之间的相关性和时间特性建模,获得了更精确的预测值。时间序列步预测的定义如下:

$$\tilde{y}_{T+N} = F(y_1, y_2, \dots, y_T, X_1, X_2, \dots, X_T)$$

(1)

其中, F 是本文要学习的非线性映射函数。

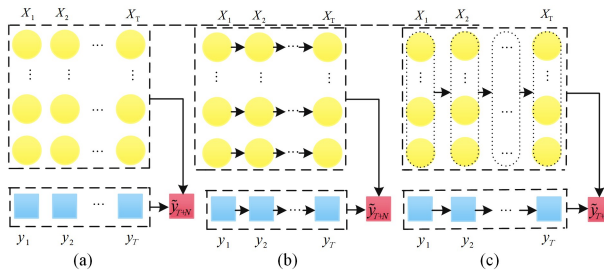


图1 时间序列建模方法

Fig. 1 Time series modeling method

MAFD 是基于图 1(c) 的频域建模方法, 其整体架构如图 2 所示。本文将历史观测数据 $y_1, y_2, \dots, y_T$ 看作特征数据一起输入 MAFD, 以期检验观测数据的自相关性。其中, $\{X^1, X^2, \dots, X^n\}$ 包含目标序列的 n 个相关特征, $\{\tilde{X}^1, \tilde{X}^2, \dots, \tilde{X}^m\}$ 为经 XGBoost 选择的标准化序列。

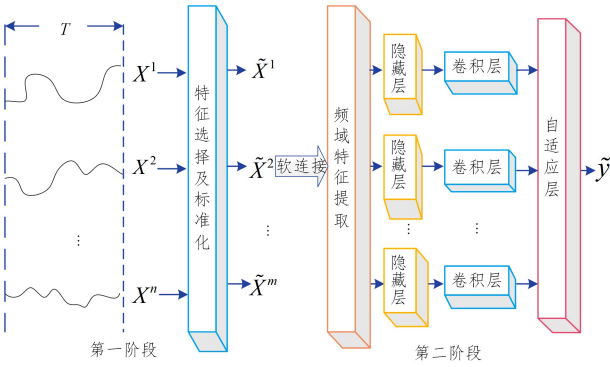


图2 MAFD 模型架构

Fig. 2 MAFD architecture

3.2 特征选择及标准化

特征选择本质上是根据实际问题的不同从原始序列集合中选取强有用且尽可能相互独立的特征子集, 从而减轻维度灾难, 提高模型性能^[28-29]。

MAFD 的第一阶段就是利用 XGBoost 进行特征选择, 去除冗余数据, 降低模型训练复杂度。XGBoost 是一种基于梯度增强决策树的改进算法, 可以有效地构造增强树且并行运算^[30]。式(2)为 XGBoost 的算法描述, 其核心是利用梯度增强构建增强树以智能地获得特征分数, 从而确定每个特征对目标序列的重要性^[31]。

$$w_X^*(K) = \frac{1}{M} \sum_{i=1}^M \tau^2(K_M) \quad (2)$$

其中, $w_X^*(K)$ 表示决策树得到的预测特征 X 的重要性分数, K_M 为第 M 棵决策树。选择的特征是在平方误差风险中提供最大估计改进 τ^2 的特征。

MAFD 的输入为 $X = \{X^1, X^2, \dots, X^n\}$, 代表具有 n 个特征的时间序列, 经 XGBoost 算法筛选后得到 $X = \{X^1, X^2, \dots, X^m\}$, $m \leq n$ 。随后, 模型利用式(3)对筛选特征 $X = \{X^1, X^2, \dots, X^m\}$ 进行标准化处理, 以加快模型的收敛速度。

$$\tilde{X} = \frac{X - \bar{X}}{\sqrt{\sum_i (X - \bar{X})^2}} \quad (3)$$

其中, $\tilde{X}$ 为标准化的特征序列, $\bar{X}$ 为频率分量均值。标准化特征 $\tilde{X} = \{\tilde{X}^1, \tilde{X}^2, \dots, \tilde{X}^m\}$ 将作为下一步预测网络的输入。

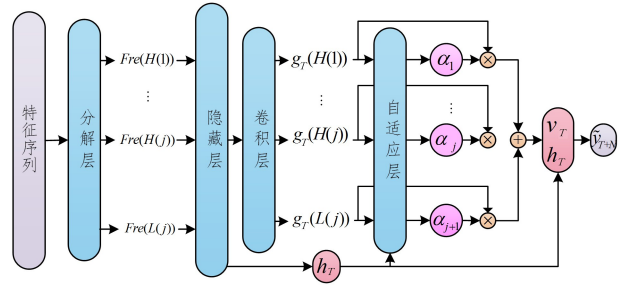


图3 自适应预测网络

Fig. 3 Adaptive prediction network

3.3 预测神经网络

MAFD 的第二阶段是基于时频转换思想改进的频域预测网络, 包括分解层、隐藏层、卷积层以及自适应层。分解层可以近似为在神经网络框架下实现小波变换处理时间序列。这个过程通常是在一些离散的时间点上进行伸缩操作和平移操作, 因此需要对伸缩因子 a 和平移因子 b ^[32] 作离散化处理。

$$a = a_0^p, b = qb_0 a_0^q, a_0 > 1, b_0 \in R \quad (4)$$

$$G(p, q) = a_0^{-\frac{p}{2}} \int_{-\infty}^{+\infty} S(t) \phi(a_0^{-p} t - qb_0) dt \quad (5)$$

其中, $\phi(t)$ 为依赖于变换因子 a, b 的基函数, a_0 和 b_0 是固定化初值。 p, q 为离散化参数, $G(p, q)$ 为变换的系数。图 3 展示了图 2 预测阶段的具体实现过程, 标准化特征序列 $\tilde{X}^1, \tilde{X}^2, \dots, \tilde{X}^m$ 经分解层一次分解之后会产生一个低频信号 $Fre(L(1))$ 和高频信号 $Fre(H(1))$ 。低频信号 $Fre(L(1))$ 继续进行下一次分解得到 $Fre(H(2))$ 和 $Fre(L(2))$ 。特征序列经神经网络频域分解层完全分解后得到 $\{Fre(H(1)), Fre(H(2)), \dots, Fre(H(j)), Fre(L(j))\}$, 其中 j 代表分解次数, 共 $j+1$ 个频率分量, 且频率由高到低排列。

对于神经网络而言, 隐藏层越多, 模型的非线性拟合能力越强, 学习效果越好^[33]。但是由于模型训练需要消耗大量时间, 所以我们一般会选择效果较好且用时较少的方案。本文 MAFD 设置两层隐藏层便能用较少的时间得到很好的效果。第一层神经元个数为 128, 第二层为 64, 这是为了让数据流的体量缩小, 减少冗余数据的干扰。网络自适应的关键是在隐藏层之后加入卷积层和自适应层。卷积层能够利用卷积操作捕捉相邻输入的相互作用, 增强网络的横向泛化能力。而自适应层可以根据输入数据的动态变化, 将关注不同的频率分量, 突出不同的频率分量对预测目标的影响。如图 3 所示, 将 XGBoost 得到的特征序列输入预测网络, 并通过式(6)进行训练, 其目标函数为:

$$\zeta = \frac{1}{T} \sum_{i=1}^T (\tilde{y} - y)^2 \quad (6)$$

$$\theta - \eta \frac{\partial \zeta(\theta)}{\partial \theta} \rightarrow \theta \quad (7)$$

其中, $\tilde{y}$ 是模型得到的预测值, y 是实际值。式(7)为模型优化函数, θ 为模型的优化参数, η 为可调节的学习率。

各层频域信息经 LSTM 隐藏层处理得到一组隐藏状态 $\{h_1, h_2, \dots, h_T\}$ 。然后, 模型对隐藏状态的前 $T-1$ 时刻进行卷积操作, 产生 $g_T(j) = \{g_T(H(1)), \dots, g_T(H(j)), g_T(L(j))\}$, 表示 T 时刻第 j 层分解产生的高频和低频卷积信息。随后, 自适应层 h_T 根据 g_T 对频率分量进行动态加权:

$$f(g_T, h_T) = W_g h_T g_T + b_g \quad (8)$$

$$\alpha = \text{sigmoid}(f(g_T, h_T)) \quad (9)$$

其中, f 为自适应层的加权函数, g_T 为频率卷积信息, W 和 b 是模型学习的权重向量和偏差。利用权重 $\alpha_j = \{\alpha_1, \alpha_2, \dots, \alpha_{j+1}\}$ 对频率分量进行求和, 得到加权向量 v_T :

$$v_T = \sum_{j=1}^{j+1} \alpha_j g_T(j) \quad (10)$$

最后, 将 v_T 和 h_T 融合并进行线性变换, 得到最终预测。

$$\tilde{y}_{T+N} = W_y[v_T, h_T] + b_y \quad (11)$$

其中, W_y 和 b_y 是模型学习的权重向量和偏差。

预测网络由 4 层网络结构组成, 首先是两层 LSTM 隐藏层, 通过这两层 LSTM 层可以得到频域预测结果; 接着是一层卷积层, 卷积层可以捕捉相邻输入的相互作用, 自动学习特征表示, 得到卷积之后的特征向量; 最后是自适应层, 自适应层的本质是横向注意力机制, 传统的注意力机制是在时间维度上选择更相关的时间步长, 但当每个时间步存在多个变量时, 容易忽略对预测性能有重要贡献的变量; 而自适应层通过 CNN 滤波器从 LSTM 隐藏状态提取特征, 这使得模型能够跨多个时间步长选择相关信息, 为不同的频域特征进行加权得到最终的预测结果。

4 实验

4.1 数据集和评估指标

为了评估 MAFD 模型的有效性和泛化能力, 本文选择了 4 种不同类型的数据集进行评估。

(1) 沪深 300 股票数据集: 我们爬取并收集了沪深 300 近 7 年的股票日线数据。该数据集包括股票交易量、MACD 指标等 9 个相关特征, 以股票收盘价为预测目标。在本文实验中, 股票数据集被划分为训练集(60%)、验证集(20%)和测试集(20%)。

(2) 空气质量数据集: 该数据集为 2013—2018 年的每日空气质量数据。本文以每日的空气质量指数 AQI 为预测目标, 以相关污染物浓度为驱动序列进行预测。空气质量数据受环境的影响, 具有周期性和季节性, 因此实验将前 1300 个数据用于训练, 剩余 488 个数据用于测试。

(3) SML2010: 用于室内温度预测的公共数据集, 来源于 UCI 机器学习数据库。本文将室温作为目标系列, 并对缺失值较多的特征序列进行了处理, 最终选择 13 个相关特征序列进行实验。在本次实验中, 我们将前 3200 个数据点用作训练集, 将后 400 个数据点用作验证集, 剩余的 437 个数据点用作测试集。

(4) 家庭用电量数据集: 该数据集收集了 2006 年 12 月至 2010 年 11 月期间的 2075259 次测量数据。由于数据集过于庞大, 本文仅选用其中的 3 万条数据, 并将其分为训练集(80%)、验证集(10%)和测试集(10%)进行实验。

股票数据和空气质量数据具有强烈的波动性, 变化幅度较大, 预测极具挑战性, 而 SML2010 和家庭用电量数据受季节的影响较大, 具有一定的周期性和季节性。因此, 本文以数据较为丰富的股票数据集和 SML2010 数据集为代表展示可视化结果, 并利用这两个差异较大的数据集对模型的可行性进行分析。为了衡量各种方法对时间序列预测的有效性, 本文使用平均绝对百分比误差(MAPE)和均方根误差(RMSE)评估模型性能, 二者均为时间序列预测领域的常用评估指标。本文选取的评估值为模型运行 10 次的最佳实验结果。

4.2 特征选择分析

模型第一步利用 XGBoost 对上述 4 个数据集特征进行重要性度量。特征的重要性取决于随机噪声替换该特征时预测性能的变化程度^[34]。在沪深 300 股票数据集中, 除股票价格外, 股票每一时刻包括下列重要特征: 股票交易量、收益率、价格变动、5 日均线、10 日均线、MACD 指标、5 日均量、10 日均量以及涨跌幅。本文通过大量交叉实验发现, 对于股票数据集, 模型设置选择的特征数量为 2 时, 准确率最高。从图 4(a)中可以看出, 特征 ma5 和特征 MACD 的重要性较高, 说明这两个特征被随机噪声替换时预测误差较大。在股票市场中, 股票的最高价、最低价代表股票当日交易的最高价格和最低价格, 由于交易所对市场的保护, 若股票当日的最高价(最低价)超过(低于)开盘价的 10%, 就会出现涨(跌)停^[35]。股票价格的高低在一定程度上反映了市场中买卖双方的势力强弱, 而股市的开盘价和收盘价则暗含着股票市场的涨跌趋势: 当今日的开盘价大于昨日的收盘价则买方势力占优, 股票价格有上涨趋势; 相反, 股票价格有下跌趋势^[36]。总而言之, 股票的价格数据是股票市场受各种因素影响的最终体现。因此, 本文将 5 日均线和 MACD 作为已筛选特征, 同股票的 K 线指标(收盘价, 开盘价, 最高价, 最低价)输入预测网络。对于空气质量数据集和家庭用电量数据集, 经过重要性度量后的可视化结果如图 4(b)和图 4(d)所示, 由于这两个数据集相关特征序列较少, 本文将全部特征输入网络。与其他数据集相比, SML2010 的特征序列较多, 但大部分特征和预测目标的相关性较弱。经不断测试发现, 选择 7 维特征时模型的特征提取效率和预测准确率均达到最佳, 因此本文将图 4(c)中的室内温度(餐厅)、室外温度等 7 个特征作为该数据集的特征选择结果。

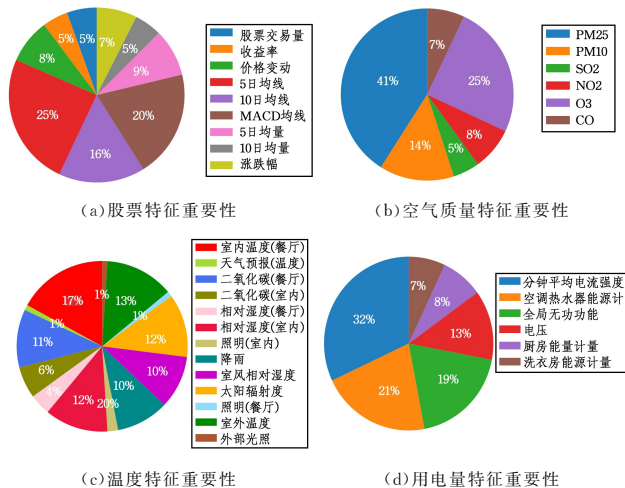


图 4 XGBoost 分析结果

Fig. 4 XGBoost analysis results

4.3 参数分析

本节将进行参数敏感性分析, 以评估和了解本文 MAFD 的优缺点。该分析固定了中间模型的体系结构, 通过改变频域分解层数和时间步长, 来观察整体性能的影响。频域分解层数的误差可视化结果如图 5 所示, 对于股票数据和空气质量数据, MAFD 神经网络模型在一层分解的预测结果误差较大。这是因为一层分解产生的两层频率分量并没有充分提取数据的频率信息。而从两层分解开始, 预测误差随着分解层

数的不断增加,也逐渐提高。这是由于较多的分解层数产生多层频率信息,导致在卷积和加权过程中产生较大的噪声。因此,本文对股票数据和空气质量采用两层分解。而对于相对平稳的室内温度数据和家庭用电量数据,二者具有较强的周期性和季节性,一层分解就可以充分挖掘内部信息,故本文采用一层分解。

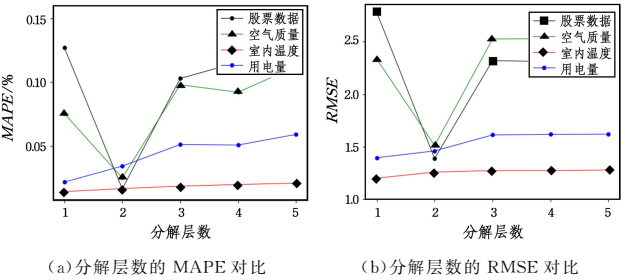


图 5 频域分解对比图

Fig. 5 Frequency decomposition comparison chart

在时序预测中,时间步长代表不同的时间粒度,例如每日、每周,它在某种程度上反映了神经网络需要记忆的信息量。当利用数据自相关模式进行预测时,该参数尤为重要。目前大部分预测均为一步预测,即预测下一天的数据。这种预测方法为短期预测,时间粒度较小,准确性较高。当时间步长增大时,时间粒度随之增大,更容易受到外界不稳定因素的干扰,产生较大的预测误差。本文比较了 1 步、3 步、5 步和 7 步预测的准确性。其中,1 步预测意味着第二天的趋势,这是一个短期预测,而 3 步预测表明半周趋势。由于股票等时间序列一周仅能收集 5 个工作日数据,因此 5 步预测和 7 步通常意味着下周的趋势,也是常用的长期预测步长,极具挑战性。通过图 6 的误差曲线不难看出,在各个数据集中,短期预测的误差都比长期预测小。这是预测领域中的普遍现象,因为长期预测数据间的相关性相对较弱,而且时间步长较大,不稳定因素增多,导致误差变大。为取得更好的预测效果,本文时间步长均为 1 步,进行短期预测。

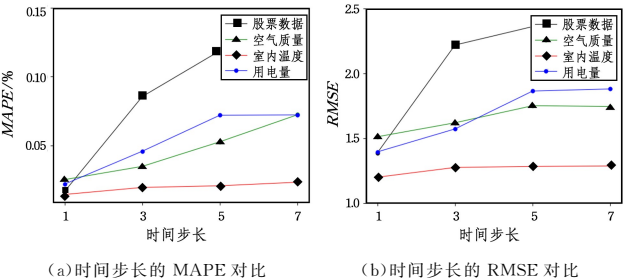


图 6 时间步对比图

Fig. 6 Time step comparison chart

通过上述两组实验可以看出,频域分解层数和时间步长是影响 MAFD 预测结果的关键参数。

4.4 基线比较

为进一步验证模型的性能,本文将 MAFD 模型与以下基线模型进行比较:自回归移动平均模型 ARMA^[13],卷积神经网络 CNN^[15],循环神经网络 RNN^[16]。ARMA 考虑了时间序列的依存性和随机波动的干扰性,是解决时间序列问题的经典统计方法;CNN 通过卷积池化操作得到原始数据的有效表征,是时序预测领域常用的神经网络模型;RNN 利用内部的记忆性,能够捕捉到数据的时间依赖特征,已在时间序列预

测领域得到广泛应用。

图 7 是上述模型预测值与实际值的折线图,从图中可以看出上述 4 种模型均能拟合出序列的未来趋势,但拟合程度相差较大。本文 MAFD 的预测曲线相比其他 3 种模型精度最高,与实际曲线的变化趋势基本一致;并且,值得注意的是,MAFD 在波峰和波谷处的拟合曲线相比其他模型偏差更小,这凸显了其预测非平稳序列的优异性能。

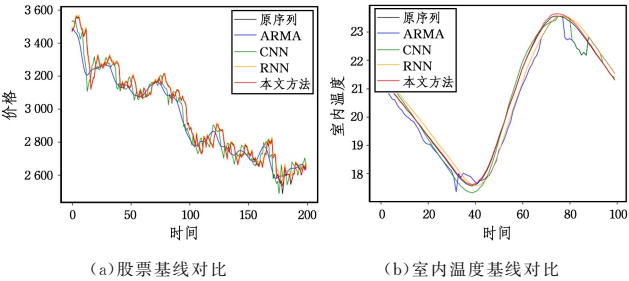


图 7 MAFD 与典型深度网络的基线对比图(电子版为彩色)

Fig. 7 Baseline comparison chart between MAFD and classic deep networks

从图 8 误差分析中也可以发现,ARMA 的预测误差高于基于神经网络的预测方法。这是因为 ARMA 缺少激活函数的非线性拟合功能,更倾向于简单的重复历史价格的轨迹。并且由于以下原因,本文提出的 MAFD 比另外两种神经网络模型取得了更佳的实验结果:1)CNN 和 RNN 仅考虑了时序特征,忽略了序列内部隐藏频域特征,而 MAFD 的分解层可以很好地深入到时间序列内部挖掘序列的频域特征;2)CNN 和 RNN 忽略了序列间的相互关系,无法捕捉输入特征之间的依赖性,而 MAFD 通过引入卷积层和自适应层,能够模拟时间序列的相互依赖性,对具有不同特性的复杂时间序列进行动态建模。

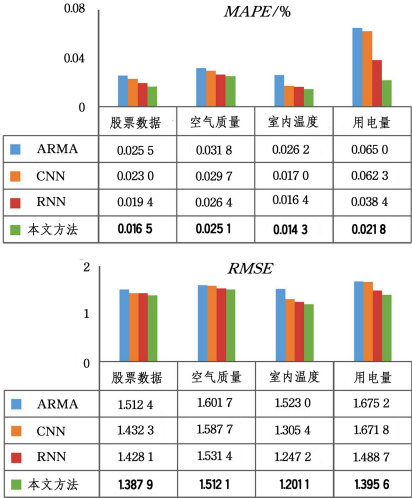


图 8 MAFD 基线误差分析

Fig. 8 MAFD baseline error analysis

4.5 模型解释性分析

目前端到端的频域预测模型仅利用目标序列进行时间序列预测,而本文通过 XGBoost 算法进行特征选择,对时间序列进行多特征预测。为验证多特征预测的必要性,本实验将进行 MAFD 模型的单特征和多特征预测对比。

图 9 为上述模型基于单特征和多特征的预测曲线拟合图,从图中可以看出,多特征模型的预测结果更接近原序列。

这是因为单特征预测只关注了目标序列的信息,而忽略了外源驱动序列的作用;不仅如此,外源驱动序列之间存在相互依赖关系,针对序列间的依赖性建模,有助于提高整体的预测准确度。从图 10 的误差分析可以看出,MAFD 的多特征预测误差略小于单特征预测,表明多特征预测能提高模型的性能,但效果并不显著。

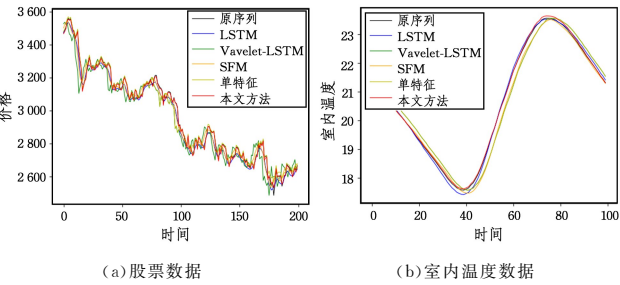


图 9 MAFD 解释性能对比图(电子版为彩色)

Fig. 9 Comparison chart of MAFD interpretation performance

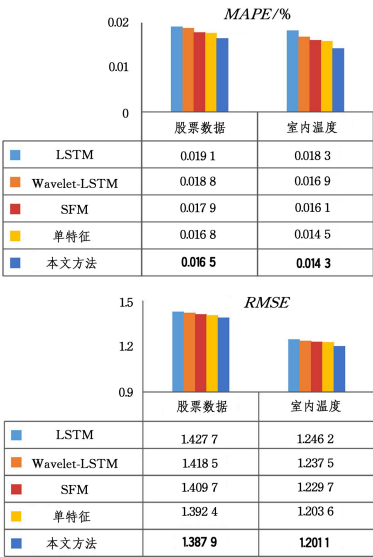


图 10 MAFD 解释性误差分析

Fig. 10 MAFD interpretive error analysis

为验证自适应频率域建模的可行性,本文将 MAFD 模型与长短期神经网络 LSTM^[17]、基于小波分解的 LSTM 神经网络(Wavelet-LSTM)^[20]、状态频率神经网络 SFM^[21] 进行比较。长短期神经网络是经典的时域预测网络,极具代表性;Wavelet-LSTM 为近期提出的利用小波变换和长短期神经网络对数据的频率信息建模的有效方法;SFM 是近期提出的端到端的单特征预测模型,以潜在在交易模式的特定频率为基础模拟股票价格的波动。与基线比较不同,本实验更侧重于评估时域与频域的预测结果。从图 9 中可以看出,频域预测方法的拟合效果优于时域预测,并且本文方法的曲线精度最高且滞后最小,与真实序列基本重合,表明频域特征可以用作神经网络输入和建模的强大表示。图 10 也给出了误差结果视图,视图显示具有相同输入的 MAFD 单特征预测误差小于 SFM,且相差相对较大,这说明自适应频域建模能较好地提高预测效果,且在 MAFD 中起关键作用。

结束语 时间序列的频率信息反映了时间序列以不同规律运动而产生的多频模式,发现数据的频率模式为未来的趋势预测提供了有用的线索。为挖掘时间序列的多频率模式,

本文设计了一种多特征自适应频域预测模型。该模型能够根据输入序列的动态演变,自动关注不同的频率分量,揭示时间序列的多频模式。本文利用 4 种不同领域的数据集对模型的性能进行了多方面的交叉验证,实验结果表明,MAFD 与其他典型方法相比具有更强的预测性和更小的滞后性。

但本文方法仍有不足之处,比如只考虑了时间序列的时频信息,还未挖掘时间序列的空间信息等其他影响预测性能的因素。针对上述不足,后续研究工作将进一步利用时频信息、空间信息等变量进行时间序列的时空预测,以期在更多的数据集上达到较好的预测效果。

参 考 文 献

[1] KAYACAN E,ULUTAS B,KAYNAK O. Grey system theory-based models in time series prediction[J]. Expert Systems with Applications,2010,37(2):1784-1789.

[2] KROGH A,LARSSON B,VON H G, et al. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes[J]. Journal of Molecular Biology, 2001,305(3):567-580.

[3] SINGH S N,MOHAPATRA A. Repeated wavelet transform based ARIMA model for very short-term wind speed forecasting [J]. Renewable Energy,2019,136:758-768.

[4] WEINSTEIN S,EBERT P. Data transmission by frequency-division multiplexing using the discrete Fourier transform[J]. IEEE Transactions on Communication Technology,1971,19(5):628-634.

[5] DAMERVAL C,MEIGNEN S,PERRIER V. A fast algorithm for bidimensional EMD [J]. IEEE Signal Processing Letters, 2005,12(10):701-704.

[6] XUAN Z Y,YANG G X. Application of EMD in the Atmosphere Time Series Prediction[J]. Acta Automatica Sinica,2008, 34(1):97-101.

[7] AMOR L B,LAHYANI I,JMAIEL M. Recursive and Rolling Windows for Medical Time Series Forecasting: A Comparative Study[C]// Recursive and Rolling Windows for Medical Time Series Forecasting: A Comparative Study. IEEE Computer Society,2016.

[8] ZHONG X,DAVID E. Forecasting daily stock market return using dimensionality reduction[J]. Expert Systems with Applications,2017,67:126-139.

[9] QING X Y,YU G N. Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM[J]. Energy,2018,148: 461-468.

[10] WANG D C,LU Y Y,ZOU Y J. Multi-output Intuitionistic Fuzzy Least Squares Support Vector Regression Algorithm[J]. Computer Science,2019,46(5):163-168.

[11] TANG J,LIU F,ZHANG W, et al. Lane-changes prediction based on adaptive fuzzy neural network [J]. Expert Systems with Applications,2018,91:452-463.

[12] HU Y,FENG B,ZHANG X, et al. Stock trading rule discovery with an evolutionary trend following model[J]. Expert Systems with Applications,2015,42(1):212-222.

[13] ROUNAGHI M M,ZADEH F N. Investigation of market efficiency and financial stability between S&P 500 and London Stock Exchange: Monthly and yearly Forecasting of Time Series

Stock Returns using ARMA model[J]. *Physica A: Statistical Mechanics and its Applications*, 2016, 456: 10-21.

[14] JIDONG W, PENG L, RAN R, et al. A Short-Term Photovoltaic Power Prediction Model Based on the Gradient Boost Decision Tree[J]. *Applied Sciences*, 2018, 8(5): 689.

[15] LV M Q, HONG Z X, CHEN T M. Traffic Flow Forecasting Method Combining Spatio-Temporal Correlations and Social Events[J]. *Computer Science*, 2021, 48(2): 264-270.

[16] CONNOR J T, MARTIN R D, ATLAS L E. Recurrent neural networks and robust time series prediction[J]. *IEEE Transactions on Neural Networks*, 1994, 5(2): 240-254.

[17] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. *Neural Computation*, 1997, 9(8): 1735-1780.

[18] PETERSEN N C, RODRIGUES F, PEREIRA F C. Multi-output bus travel time prediction with convolutional LSTM neural network[J]. *Expert Systems with Applications*, 2019, 120: 426-435.

[19] HU H, GUO J Q. State-Frequency Memory Recurrent Neural Networks[C]// *International Conference on Machine Learning*. New York: ACM, 2017: 1568-1577.

[20] MOUATADID S, ADAMOWSKI J F, TIWARI M K, et al. Coupling the maximum overlap discrete wavelet transform and long short-term memory networks for irrigation flow forecasting[J]. *Agricultural Water Management*, 2019, 219: 72-85.

[21] ZHANG L H, AGGARWAL C, QI G J. Stock price prediction via discovering multi-frequency trading patterns[C]// *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM, 2017: 2141-2149.

[22] OUYANG L, ZHU F, XIONG G, et al. Short-term traffic flow forecasting based on wavelet transform and neural network [C]// *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*. NJ: IEEE, 2017: 1-6.

[23] CHEN T Q, GUESTIN C. Xgboost: A scalable tree boosting system[C]// *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM, 2016: 785-794.

[24] ZHENG H, YUAN J, CHEN L. Short-term load forecasting using EMD-LSTM neural networks with a XGBoost algorithm for feature importance evaluation[J]. *Energies*, 2017, 10(8): 1168.

[25] HSIEH C P, CHEN Y T, BEH W K, et al. Feature Selection Framework for XGBoost based on Electrodermal Activity in Stress Detection[C]// *2019 IEEE International Workshop on Signal Processing Systems (SiPS)*. NJ: IEEE, 2019: 330-335.

[26] KE J, ZHENG H, YANG H, et al. Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach [J]. *Transportation Research*, 2017, 85c(dec): 591-608.

[27] 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*. New York: ACM, 2018.

[28] CHU B, LI Z S, ZHANG M L, et al. Improvement Research on Feature Selection Algorithm Based on Forest Optimization[J]. *Journal of Software*, 2018, 29(9): 2547-2558.

[29] DASH M, LIU H. Feature selection for classification[J]. *Intelligent Data Analysis*, 1997, 1(3): 131-156.

[30] MA X, SHA J, WANG D, et al. Study on a prediction of P2P network loan default based on the machine learning LightGBM and XGboost algorithms according to different high dimensional data cleaning[J]. *Electronic Commerce Research and Applications*, 2018, 31: 24-39.

[31] TORLAY L, PERRONE B M, THOMAS E. Machine learning-XGBoost analysis of language networks to classify patients with epilepsy[J]. *Brain Informatics*, 2017, 4(3): 159-169.

[32] DAUBECHIES I. The wavelet transform, time frequency localization and signal analysis[J]. *IEEE Transactions on Information Theory*, 1990, 36(5): 961-1005.

[33] HUANG G B. Learning capability and storage capacity of two-hidden-layer feed forward networks[J]. *IEEE Transactions on Neural Networks*, 2003, 14(2): 274-281.

[34] LI Y F, XIE M G, THONG N. A study of project selection and feature weighting for analogy based software cost estimation [J]. *Journal of Systems and Software*, 2009, 82(2): 241-252.

[35] FAMA E F. Random walks in stock market prices[J]. *Financial Analysts Journal*, 1995, 51(1): 75-80.

[36] AL-SHUBIRI F N. Analysis the determinants of market stock price movements: An empirical study of Jordanian commercial banks[J]. *International Journal of Business and Management*, 2010, 5(10): 137.



WANG Xiao-di, born in 1981, master. His main research interests include artificial intelligence, data mining and so on.