

# 基于 LSTM 循环神经网络的税收预测

文 豪 陈 昊

湖北大学计算机与信息工程学院 武汉 430062

(716149013@qq.com)

**摘 要** 分析历史税收数据之间的隐藏关系,利用数学模型来预测未来的税收收入是税收预测的研究重点。在此,提出了一种结合小波变换的长短期记忆(LSTM)循环神经网络的税收预测模型。在数据预处理上结合小波变换来去除税收数据中的噪声,提高模型的泛化能力。LSTM 神经网络通过加入隐藏神经单元和门控单元能够更好地学习到历史税收数据之间的相关关系,并进一步提取有效的输入序列间的状态新息,而且解决了循环神经网络的长期依赖问题。实验结果表明,基于 LSTM 神经网络的编码器-解码器结构能够增强税收预测的时间步长,在中长期的税收预测中相比单步滑动窗口的 LSTM 神经网络模型以及基于差分微分方程的灰色模型和基于回归的自回归移动平均模型(ARIMA),在预测精度上有明显提升。

**关键词:** 税收预测;长短期记忆网络;小波变换;编码器-解码器

**中图法分类号** TP183

## Tax Prediction Based on LSTM Recurrent Neural Network

WEN Hao and CHEN Hao

School of Computer Science & Information Engineering, Hubei University, Wuhan 430062, China

**Abstract** Analyzing the hidden relationship between historical tax data and using mathematical models to predict future tax revenue is the focus of tax forecast research. A tax prediction model of long short-term memory (LSTM) recurrent neural network combined with wavelet transform is proposed in this paper. Combining wavelet transform on data preprocessing to remove noise from tax data and improve the generalization ability of the model. The LSTM neural network can better learn the correlation between historical tax data by adding hidden neural units and gated units, and further extract valid state innovations between input sequences, and overcome the long-term dependency problem of recurrent neural networks. Experimental results show that the encoder-decoder structure based on the LSTM neural network can enhance the time step of tax prediction. Compared with the single-step sliding window LSTM neural network model and the gray model based on difference differential equations in the long-term tax prediction, the model and the regression-based autoregressive moving average model (ARIMA) significantly improve the prediction accuracy.

**Keywords** Tax forecasting, Long-short term memory network, Wavelet transform, Encoder-decoder

## 1 引言

税收是国家重要的财政收入来源,税收收入的稳定对于国家的经济、政治、文化有着重大影响。由于税收收入受到多种因素影响,使得税收收入的管控难度较大,税收收入的征收风险也相应增大。通过对未来税收收入进行相关分析预测,能发掘并避免征收过程中可能出现的风险,即时采取相关的管理手段来保证未来税收质量和税收趋势的稳定性。同时,税收预测也为决策部门的相关政策进行了前瞻性的预测,分析某项政策的实施可能带来的经济收益和社会收益,以及可能带来的征收风险,将其作为税收决策工作和计划编制中的参考,做出有利于促进纳税人企业发展和创造更多社会经济效益的改革方案。

税收预测的研究方法,主要基于回归预测和时间序列预

测。基于差分微分回归的灰色系统理论能对事物发展规律做长期性的模糊描述,在宏观经济预测<sup>[1-2]</sup>、农业产量预测<sup>[3]</sup>等领域有广泛应用。基于时间序列分析的 ARIMA 模型,自提出以来也被广泛应用于时间金融数据预测中。Maldonado-Molina<sup>[4]</sup>等用差分整合移动平均自回归模型(ARIMA)来分析酒精税收入;Lin<sup>[5]</sup>利用 ARIMA 模型对海关税收进行预测分析。

近年来,随着深度学习算法的兴起,神经网络在时间序列处理中的作用也凸显出来<sup>[6]</sup>,经典的时间序列模型无法在数据中捕获复杂的模式,而多元技术难以扩展到大型问题规模,基于深度神经网络的模型可以学习复杂的模式和依赖关系<sup>[7]</sup>。传统神经网络虽然对输入数据的平稳性要求较低,但由于其结构特性,无法提取时间序列数据中的上下文状态信息。相比传统神经网络,循环神经网络因为其强大的计算能

基金项目:国家自然科学基金面上项目(61977021);贵州税务大数据汇聚整理平台项目(182001022)

This work was supported by the General Program of National Natural Science Foundation of China (61977021) and Guizhou Tax Big Data Consolidation Platform Project (182001022).

通信作者:陈昊(ch@hubu.edu.cn)

力并且对时变数据处理任务保留过去输入的有效性<sup>[8]</sup>以及解决蕴含时间序列信息的问题具有天然优势<sup>[9]</sup>而被广泛应用于机器翻译<sup>[10]</sup>、自然语言处理<sup>[11-12]</sup>和金融时间预测<sup>[13]</sup>中。基于循环神经网络的长短期记忆网络(LSTM)模型,通过门控方式来动态控制权重,更好地提取时间序列中的上下文状态信息,并有效地解决了循环神经网络的长期依赖问题,Zhen等<sup>[14]</sup>将隐藏状态张量化,只需扩增张量而无需增加参数大小即可扩充 LSTM 的网络容量。LSTM 与其他算法的组合应用也非常广泛,Jia 等<sup>[15]</sup>通过双注意力机制的 LSTM 模型来对农产品价格进行预测;Aniekan 等<sup>[16]</sup>在 LSTM 神经网络的基础上引入堆栈编码器、卷积神经网络及小波变换来进行单步时间序列数据的预测。循环神经网络模型在时间序列数据预测中的研究也日益兴起,但是在税收预测上还鲜少运用。

本文将收集整理的贵州省某市 2017 年 1 月至 2019 年 7 月的矿产品税收收入作为研究数据集。构建基于长短期记忆网络(LSTM)的编码器-解码器模型来进行税收预测的中长期预测,并在前期的数据处理中引入小波变换进行数据去噪。税收数据分为训练数据和测试数据两个部分,训练数据用来训练模型,测试数据用来评估模型的泛化能力,并利用单步滑动窗口的 LSTM 神经网络模型以及灰色模型和差分整合移动平均自回归(ARIMA)模型来进行预测对比研究。

## 2 问题描述及税收预测模型构建

### 2.1 问题描述

矿产品税收收入不仅受到销售单价、开采政策、销量等主观因素的影响,还受到开采季节、天气等客观因素的影响。因此,在实际的税收收入趋势中会存在较大的上下波动,表现为非平稳的时间序列。对于上下摆动的数据,灰色模型无法准确地预测其精度,而 ARIMA 模型需要在模型构建前期进行复杂的差分操作来消除其非平稳性。

基于微分方程的灰色模型虽然被广泛应用于时间序列预测场景中,但是要求输入的时间序列数据具有较好的平稳性,对于波动性较大的时间序列数据,预测精度也有较大的损失。相比灰色模型,基于时间序列分析的 ARIMA 模型适用于非平稳的具有小幅波动的时间序列数据,但是在前期的数据处理中,需要利用一阶或多阶差分来消除数据中的波动性和季节周期性,对数据的平稳化处理较为复杂。

LSTM 神经网络由于具有动态调整和记忆力学习的能力,无需在前期过多地分析原始数据中的隐藏关系,而是通过隐藏层的神经网络单元和单元之间的相关权重关系来自动调整和学习时间序列之间存在的隐藏关系,即使原始数据不具备明显的规律性。神经网络模型的难点在于:如何使用随机梯度下降的方法来求解最优超参数,以来获得最小的损失度和最小的误差;网络模型的容量达到多大才能更好地拟合数据,并且不会出现过拟合的问题。LSTM 由于引入动态自循环结构,很好地解决了超参数最优求解过程中存在的梯度消失和爆炸问题。分析模型训练结果来适当扩充网络的层次结构和隐藏神经元个数,利用 LSTM 构建编码器-解码器结构能更好地提取样本特征,增强模型的泛化能力。

### 2.2 长短期记忆(LSTM)神经网络

为了解决长期依赖的问题,Hochreiter 等<sup>[17]</sup>提出长短期记忆(Long-Short Time Memory, LSTM)网络,用于改进传

统的循环神经网络模型。LSTM 通过引入自循环来产生梯度长时间持续流动路径。对比传统的循环神经网络,LSTM 的一个关键扩展是使自循环的权重视上下文而定,而不是固定的。LSTM 的关键是“细胞”状态,即图 1 中上部横穿整个结构的直线。“细胞”状态像一个传送带,能将上下文相关信息传送到整个网络结构。LSTM 通过一种称为门的结构来删除或向“细胞”中添加信息。如图 1 所示,LSTM 神经网络的内部“细胞”主要由遗忘门(forget gate)、输入门(input gate)及输出门(output gate) 3 种门结构构成。

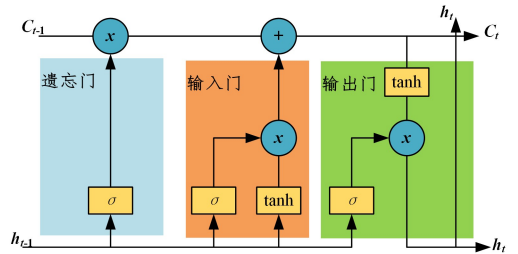


图 1 LSTM 循环神经网络“细胞”结构图

Fig. 1 Cell structure diagram of LSTM recurrent neural network

遗忘门是 LSTM 单元结构中的关键部分,决定了哪些信息是需要从“细胞”状态中丢弃,由 sigmoid 单元将权重设置为 0~1,并且其能够避免因梯度反向传播而导致梯度消失或梯度爆炸的问题。其表达式如式(1)所示。

$$f_t = \sigma(U_f x_t + U_f h_{t-1} + b_f) \quad (1)$$

其中, $x_t$  为输入向量, $b_f$  为偏置, $U_f$  为输入权重, $W_f$  为遗忘门权重, $h_{t-1}$  为前一隐藏层向量。

输入门决定了哪些新的信息需要被更新到“细胞”状态中,如式(2)所示。同时还需要通过一个 tanh 层来产生新的候选向量值,如式(3)所示。

$$i_t = \sigma(U_i x_t + W_i h_{t-1} + b_i) \quad (2)$$

$$\bar{C} = \tanh(U_c x_t + W_c h_{t-1} + b_c) \quad (3)$$

新的“细胞”状态由两部分组成,一部分为遗忘门所要移除的信息,另一部分为输入门和 tanh 层所要更新的信息,新的“细胞”状态表达式如式(4)所示。

$$C_t = f_t * C_{t-1} + i_t * \bar{C}_t \quad (4)$$

输出门控制“细胞”状态对当前输出值  $h_t$  的影响,即细胞状态中哪些信息会在时间步  $t$  输出。输入信息由两部分组成,一部分由 sigmoid 层来决定“细胞”状态中哪些信息需要输出,如式(5)所示,另一部分通过 tanh 层来决定当前的输出值。其具体表达式如式(6)所示。

$$o_t = \sigma(U_o x_t + W_o h_{t-1} + b_o) \quad (5)$$

$$h_t = o_t * \tanh(C_t) \quad (6)$$

### 2.3 编码器-解码器模型

循环神经网络是由隐藏状态和一个与输入序列相对应的输出构成,通过训练预测序列中的下一个输出,循环神经网络可以学习序列上的概率分布,从而最大化该概率来输出最准确的下一个预测值。简单的循环神经网络模型只能处理固定长度的输入和输出序列,并且对于较长的输出序列预测,随着预测步长的增加,每一步预测的误差也会相应积累,导致预测序列随着步长的位移,相对于原始序列会有更大的偏差。而编码器-解码器模型能将变长序列编码为定长向量表示,并将给定的定长向量表示解码为变长序列<sup>[18]</sup>,能很好地解决单一

神经网络的多步预测误差积累问题。编码器-解码器模型由于对变长序列的处理能力,被广泛应用于语音识别、机器翻译等人工智能领域。文献[18]提出了一个用于机器翻译的编码器-解码器结构,如图2所示。

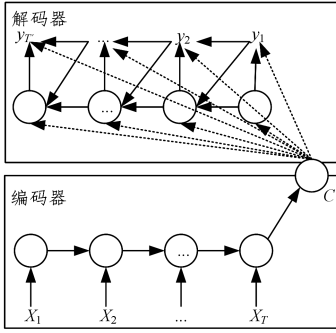


图2 编码器-解码器模型示意图

Fig.2 Schematic of encoder-decoder model

编码器-解码器结构用于将输入序列 $(x_1, x_2, \dots, x_T)$ 经过模型一系列的函数变换后输出 $(y_1, y_2, \dots, y_{T'})$ ,其中输入序列与输出序列的长度可以不等,可以通过训练来学习变长序列上的条件分布概率。编码器结构用于读取输入序列 $(x_1, x_2, \dots, x_T)$ 中的隐藏状态,其中隐藏状态随着读入的序列不断改变,其数学表达式如式(7)所示。

$$h_t = f(h_{t-1}, x_t) \quad (7)$$

上下文状态  $C$  是对整个输入序列中隐藏状态的总结,反映了整个输入序列的特征信息。

解码器结构用于输出不定长度的预测序列,解码器通过循环神经网络中的隐藏状态  $h_t$  来预测下一个输出  $y_t$ ,依次递归从而预测出整个输出序列。其中,编码器中的上下文状态  $C$  贯穿整个解码器状态,解码器中的隐藏状态  $h_t$  由上一个输入和隐藏状态及编码器中提取的上下文状态  $C$  经过函数转换而来,其具体的数学表达式如式(8)所示。

$$h_t = f(h_{t-1}, y_{t-1}, c) \quad (8)$$

相比编码器结构中的隐藏状态函数变换,解码器结构中的隐藏状态函数变换中多了上下文状态  $C$ ,从而将编码器的上下文状态向量传递至解码器部分,使得编码器-解码器结构能够紧密结合,特征向量能在整个模型结构中进行传递,从而进一步提高模型的泛化能力。

基于循环神经网络的编码器-解码器模型,其最终目的是,对于给定的输入序列 $(x_1, x_2, \dots, x_n)$ 和经过模型预测出的输出序列 $(y_1, y_2, \dots, y_n)$ ,通过时间序列数据的训练来最大化条件似然概率。

$$\max_{\theta} \frac{1}{N} \sum_{n=1}^N \log p_{\theta}(y_n | x_n) \quad (9)$$

式(9)中的  $\theta$  为编码器-解码器模型中的参数,在模型训练的过程中,经过基于梯度下降的算法来进行求解,以达到使得条件似然概率最大化的最优解。

## 2.4 数据预处理

为了使模型具有较低的损失度和较快的收敛速度,需要在模型进行训练之前对数据进行降噪和标准化处理。

### 2.4.1 小波去噪

贵州省某市 2017 年 1 月 1 日至 2019 年 7 月 22 日的矿产品税收收入趋势如图 3 所示。

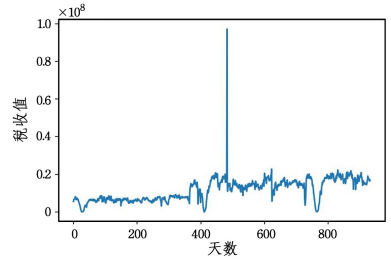


图3 矿产品税收收入趋势图

Fig.3 Trend of mineral products tax revenue

从图中可以看出,税收收入中存在一个明显的高波峰,并且在多个时间段有明显的随机波动,为了使得 LSTM 模型更好地学习序列数据中的隐藏状态和相关信息,在前期的数据处理中,采用小波变换来去除序列数据中的噪声,使得时间序列数据中的特征数据得到凸显,降低和简化神经网络模型的学习难度和加快模型的收敛速度,并降低模型的损失率。

小波变换是一种多尺度信号分析,其在时域和频域都具有伸缩性和平移性,并且具有局部表征特性。由于函数中的尖峰(序列数据中的异常波峰)具有严格的局部性,因此用具有局部特性的小波基函数通常比用余弦和正弦基函数要简单得多。时间序列数据经过小波变换后,被分解成了不同的频率分量,然后用与尺度相匹配的分辨率来研究每个分量<sup>[19]</sup>。小波滤波与真实时间序列的波动性和时间变化特性密切相关,并不受限于平稳性假设的限制。小波分析已被证明在分析、建模和预测金融工具(如股票和汇率)的行为方面有较好的表现形式<sup>[20-21]</sup>。

连续小波变换即表示信号的函数  $f(t)$  在一组小波基函数上的展开,其数学定义表达式如式(10)所示。

$$c(s, \tau) = \int_{\mathbb{R}} f(t) \Psi_{s, \tau}^*(t) dt, s \in \mathbb{R}^+ - \{0\}, \tau \in \mathbb{R} \quad (10)$$

其中,  $*$  代表复共轭。小波函数是由单个母小波  $\Psi(t)$  通过尺度伸缩  $s$  和平移  $\tau$  派生出来的,其数学表示如式(11)所示,其中的  $\frac{1}{\sqrt{s}}$  是为了在不同尺度间规范化能量。

$$\Psi_{s, \tau}(t) = \frac{1}{\sqrt{s}} \Psi\left(\frac{t-\tau}{s}\right) \quad (11)$$

连续小波逆变换用于将小波系数  $c(s, \tau)$  合成有限能量的信号  $f(t)$ ,其数学定义如式(12)所示。

$$f(t) = \int_{\mathbb{R}^+} \int_{\mathbb{R}} c(s, \tau) \Psi_{s, \tau}(t) ds d\tau \quad (12)$$

式(12)中,函数  $\Psi(t)$  要成为小波函数必须满足两个重要的条件,分别为容许条件和正则。容许条件<sup>[22]</sup>要求小波具有带通频谱,因此小波必须是一个震荡波。正则条件则要求小波函数在时域和频域都具备平滑性和紧凑性。

连续小波变换通过连续地对母小波进行尺度伸缩、位移、与被分析的函数做内积运算,导致分解后的小波系数存在很大的冗余性,并且变换的最终结果不可能解析地计算出来,此外解法的复杂度相对较高,需要高效的数值解法。离散小波变换通过小波系数采样,并将时间序列分解成两组正交的分量,有效地减小了小波系数的冗余性,并且简化了小波分析的计算且不会损失精度。Mallat<sup>[23]</sup>提出了一种利用高通滤波器和低通滤波器对矢量进行卷积的高效离散小波变换及其逆变换方法。离散小波变换涉及两个小波函数,分别为母小波和

父小波(尺度函数)。式(13)表示的母小波描述了信号的高频部分,即信号的近似分量;式(14)表示的父小波描述了信号的低频部分,即信号的细节分量。变量  $j$  代表多分辨率分析的尺度数量。

$$\Psi_{j,k}(t) = 2^{-\frac{j}{2}} \Psi(2^{-j}t - k) \quad (13)$$

$$\varphi_{j,k} = 2^{-\frac{j}{2}} \varphi(2^{-j}t - k) \quad (14)$$

离散小波逆变换逼近原始信号  $f(t)$  的重构如式(15)所示。

$$f(t) = \sum_k c_{j,k} \Psi_{j,k}(t) + \sum_k d_{j,k} \varphi_{j,k}(t) + \sum_k d_{j-1,k} \varphi_{j-1,k}(t) + \sum_k d_{j-2,k} \varphi_{j-2,k}(t) + \dots + \sum_k d_{1,k} \varphi_{1,k}(t) \quad (15)$$

其中,  $c_{j,k}$  表示近似系数,  $d_{j,k}$  表示细节系数。

原始税收数据经离散小波变换后,噪声数据集中在细节分量中,通过设置阈值来对细节系数进行去噪,将去噪后的小波系数进行重构得到去噪后的税收数据。去噪后的数据如图4所示,数据变得更加平滑。

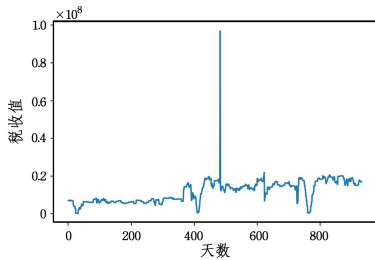


图4 小波去噪后的矿产品税收收入趋势

Fig. 4 Trend of mineral products tax revenue after wavelet denoising

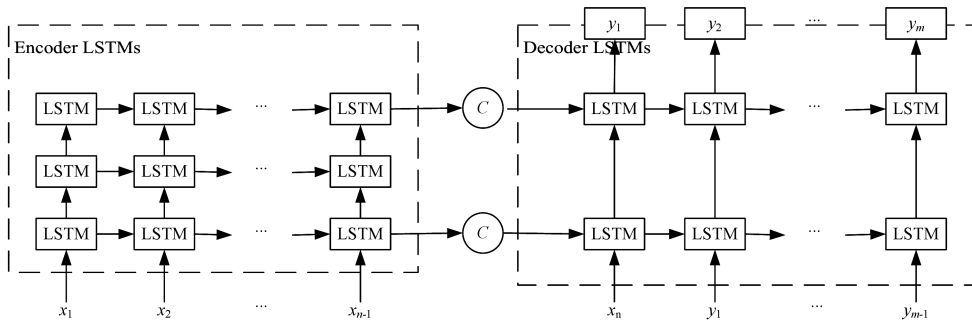


图5 LSTM-Encoder-Decoder 税收预测模型

Fig. 5 LSTM-Encoder-Decoder tax prediction model

矿产品税收预测模型利用历史税收数据来预测未来,因此解码器部分的第一个输入应该为历史序列数据的最后一天的矿产品税收值,因此第一个输入为整个输入序列  $(x_1, x_2, \dots, x_n)$  中的最后一个序列值  $x_n$ ,解码器的第一层 LSTM 神经网络的第一个隐藏状态表示如式(18)所示。

$$h_{1,1} = f(c_1, x_n) \quad (18)$$

其中,  $h_{i,j}$  表示第  $i$  层的第  $j$  个隐藏状态,  $c_i$  表示解码器部分的第  $i$  个上下文特征状态向量。

解码器的第二层 LSTM 神经网络的第一个隐藏状态表示如式(19)所示。

$$h_{2,1} = f(c_2, o_1) \quad (19)$$

其中,  $o_t$  为第一层 LSTM 神经网络的第  $t$  个输出,其表达式如式(20)所示。

$$\begin{cases} o_1 = f(h_{1,1}, x_n) \\ o_t = f(h_{i,t-1}, y_{t-1}), t > 1 \end{cases} \quad (20)$$

矿产品税收预测序列是随着解码器的时间位移递归生成

## 2.4.2 数据标准化

由于税收数据的实际值较大,如果直接用于计算,则会导致函数永远无法收敛,因此有必要对税收值进行标准化处理。税收数据被处理为多个窗口序列  $\{x_1, x_2, x_3, \dots, x_m\}$ , 每个窗口序列的长度等长。以窗口序列的第一个值  $x_1$  为基准点,窗口序列中的每一个值经过式(16)转化为与第一个值差值的比例,这样凸显了每个窗口序列中的变化率,使得模型收敛得更加快速。其逆变换如式(17)所示。

$$n_i = \left(\frac{x_i}{x_1}\right) - 1 \quad (16)$$

$$x_i = (n_i + 1)x_1 \quad (17)$$

## 2.5 LSTM-Encoder-Decode 税收预测模型

LSTM 神经网络矿产品税收预测模型使用循环神经网络的编码器-解码器的序列到序列结构,模型的解码器部分用于提取历史税收数据中的特征向量,并将其沿着神经网络链路传输至模型解码器部分。模型解码器部分,通过对特征向量解码来预测未来时间段内的矿产品税收。历史税收数据为单维的时间序列金融数据,简单的单层编码器-解码器序列预测模型对单维时间序列的隐藏状态提取和特征状态向量的重构能力有限,而且单一地通过增加隐藏节点数来扩充网络的容量会导致模型的过拟合问题,导致对未来真实值的预测存在较大偏差。因此,本文构建了多层的编码器-解码器结构的基于 LSTM 神经网络模型来进行矿产品税收预测,其具体层次结构如图5所示。

的。每一个 LSTM 单元的输出由上一个单元隐藏层的输出和上一个单元输出经第一层 LSTM 层转换后的输出构成,其表达式如式(21)所示。

$$y_t = f(h_{2,t-1}, o_{t-1}) \quad (21)$$

由 LSTM 神经网络构成的编码器和解码器矿产品税收预测模型的目的是,通过历史矿产品税收数据的训练来最大化未来税收预测值  $y_t$  的条件似然概率。

## 3 实验结果分析与对比

本节利用构造的 LSTM-Encoder-Decoder 税收预测模型,对贵州省某市的矿产品税收收入进行预测,其中用于输入的数据分为经过小波降噪和未经降噪两类,来验证小波降噪在提高模型预测精度方面是否有积极作用,并采用单步滑动窗口 LSTM 预测模型以及 GM(1,1)模型和 ARIMA 模型做对比实验,进一步验证 LSTM-Encoder-Decoder 模型在税收预测中的预测能力。

### 3.1 实验流程

本文将贵州省某市的税收数据分为训练数据和测试数据,训练数据用来训练模型,测试数据用来验证模型的效果。

### 3.2 数据集

用于循环神经网络税收预测模型的样本数据为贵州省某市 2017 年 1 月 1 日至 2019 年 7 月 22 日的矿产品税收数据,样本数据被分为用于训练的训练数据,为 2017 年 1 月 1 日至 2019 年 6 月 13 日的税收数据,用于验证的测试数据为 2019 年 6 月 23 日至 2019 年 7 月 22 日的税收数据,并用此来评估模型的预测精度。

### 3.3 预测结果度量

税收预测模型的预测精度表现将基于平均百分比误差 (MAPE) 和皮尔逊相关系数 (R) 来进行评估。MAPE 通过计算预测值和真实值之间的相对误差的平均值来衡量误差的大小,其定义如式(22)所示。皮尔逊相关系数用来描述两个变量之间的相关性,其定义如式(23)所示。

$$MAPE = \frac{\sum_{t=1}^N \left| \frac{y_t - \hat{y}_t}{y_t} \right|}{N} \quad (22)$$

$$R = \frac{\sum_{t=1}^N (y_t - \bar{y}_t)(\hat{y}_t - \bar{\hat{y}}_t)}{\sqrt{\sum_{t=1}^N (y_t - \bar{y}_t)^2 \sum_{t=1}^N (\hat{y}_t - \bar{\hat{y}}_t)^2}} \quad (23)$$

### 3.4 实验步骤与结果分析

将小波降噪的数据和未经降噪的训练数据经过数据标准化处理后来分别训练模型。模型训练次数为 500 次,每次抓取用于训练的批次大小 batch\_size 为 32。使用 adam 优化器<sup>[24]</sup>动态调整学习率,初始学习率设置为 0.01,为防止模型过拟合,模型的 dropout<sup>[25]</sup>设置为 0.2,损失函数使用均方对数损失 (MSLE) 函数。模型训练的损失度如图 6 所示。

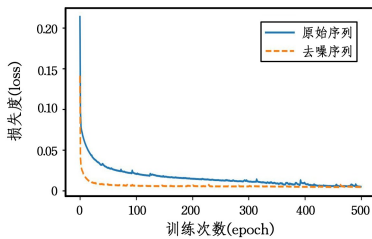


图 6 LSTM 循环神经网络模型的损失度

Fig. 6 LSTM recurrent neural network model loss

经过小波降噪的序列数据使得模型收敛速度更快,利用小波变换进行数据降噪提高了模型对数据的敏感程度。

采用小波降噪预处理和未采用的预测结果度量如表 1 所列。

表 1 小波降噪预处理的预测结果度量对比

Table 1 Comparison of prediction results of wavelet denoising preprocessing

模型名称	MAPE/%	R
降噪 LSTM-Encoder-Decoder 预测模型	6.46	0.59
未降噪 LSTM-Encoder-Decoder 预测模型	11.0	0.22

训练后模型的预测拟合曲线如图 7 所示,未对税收数据进行小波去噪处理的 LSTM 神经网络矿产品税收预测模型在第 7 天的矿产品税收预测与预期值的偏差较大,而对税收

数据进行小波降噪的矿产品税收预测模型的预测值与预期的矿产品税收值没有极端的过大偏差,相对百分比误差 MAPE 减少,线性相关性 R 增加,表明在前期数据中引入小波变换来进行数据降噪在一定程度上减少了原始时间序列中噪音数据对模型预测的干扰,相比未降噪的模型,预测精度提高了 3.54%,并且预测值与历史数据的相关性强度也较未进行数据降噪的模型有所提高。

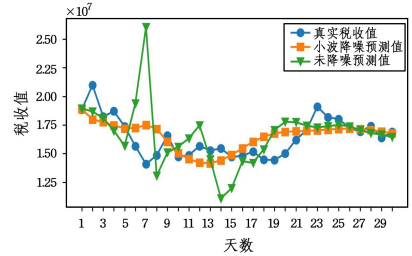


图 7 LSTM-Encoder-Decoder 税收预测模型的预测拟合曲线

Fig. 7 LSTM-Encoder-Decoder tax forecasting model predicts fitting curve

### 3.5 实验结果比对

未采用编码器-解码器结构的基于单纯的 LSTM 神经网络架构模型的矿产品税收预测模型,采用移动窗口的模式来进行税收预测,即每预测出下一个时刻  $t$  的输出,整个窗口向后移动一个时刻,并利用新的时间窗口序列来预测下一个  $t+1$  时刻的输出,并再次往后平移一个时刻,依次递归直到预测出整个输出序列。由于该 LSTM 神经网络预测模型会积累上一个时刻的预测误差,导致随着窗口向后移动,使得误差逐渐积累。

LSTM-Encoder-Decoder 税收预测模型与单步滑动窗口 LSTM 税收预测模型的预测结果度量如表 2 所列。

表 2 LSTM-Encoder-Decoder 与步滑动窗口 LSTM 预测结果的度量比对

Table 2 Measurement comparison between LSTM-Encoder-Decoder and step sliding window LSTM prediction results

模型名称	MAPE/%	R
降噪 LSTM-Encoder-Decoder 预测模型	6.46	0.59
降噪单步滑动窗口 LSTM 预测模型	7.91	0.50

LSTM-Encoder-Decoder 预测模型与单步滑动窗口 LSTM 预测模型的预测结果拟合比曲线结果如图 8 所示。

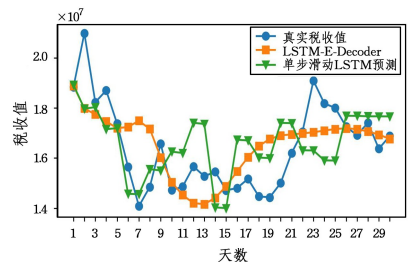


图 8 LSTM-Encoder-Decoder 与单步滑动窗口 LSTM 预测比对曲线图

Fig. 8 Comparison graph of LSTM-Encoder-Decoder and single-step sliding window LSTM prediction comparison

基于 LSTM 神经网络单步窗口移动进行递归预测模型

的相对百分比误差 MAPE 为 7.91%，线性相关性为 0.50，相比编码器-解码器预测模型 6.46% 的相对百分比误差，编码器-解码器预测模型的精度高出了 1.45%，并且线性相关性高出 9%。从图 8 中可以看出，虽然单步预测的预测精度在 1 到 7 天要好于基于多步的编码器-解码器预测，但是随着预测天数的增加，单步预测相比编码器-解码器结构，偏离真实税收值的程度相对较高。

灰色模型是一种针对“贫信息”或信息不完全的“小样本”系统进行研究的一门新兴边缘科学理论。灰色模型 GM(1,1) 通过微分方程求解预测方程系数来进行数据预测。ARIMA 模型是统计模型中常用来预测时间序列的模型，相比 ARMA 模型，ARIMA 模型多了一个差分操作。由于税收数据受到多种变量因素的影响，可能存在较大的波动，如季节性波动、周期性波动，也可能存在一定的趋势，大部分的税收数据都有递增的趋势。ARIMA 模型可以通过差分操作、滑动平均、多项式过滤等操作来消除原始税收数据的周期性、季节性以及趋势。通过模型参数的选定，构建 ARIMA(6,1,1) 模型来进行税收预测。

基于 LSTM 编码器-解码器的矿产品税收预测模型与 GM(1,1) 模型、ARIMA(6,1,1) 模型的预测结果度量对比如表 3 所列。

表 3 LSTM-Encoder-Decoder, GM(1,1), ARIMA(1,1,6) 预测结果度量对比

Table 3 Prediction result measurement comparison of STM-Encoder-Decoder, GM(1,1), ARIMA(1,1,6)

模型名称	MAPE/%	R
降噪 LSTM-Encoder-Decoder 预测模型	6.46%	0.59
GM(1,1) 预测模型	23.94	-0.06
ARIMA(6,1,1) 预测模型	10.12	0.10

从实验结果中看出，GM(1,1) 的线性相关性小于 0，表明预测值与预期值为负相关，说明模型的预测值与实际预测值的相关性很弱，不适合矿产品税收的长期预测。3 种模型的预测拟合对比曲线如图 9 所示，相比 GM(1,1) 模型，ARIMA 模型的预测值总体偏向于预期税收的中值，相比 LSTM 模型无法反映税收的应有趋势性，只能反映未来税收的平均税收值的上下摆动准线。

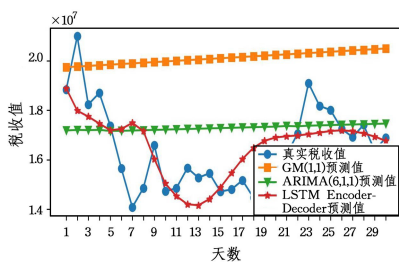


图 9 LSTM-Encoder-Decoder, GM(1,1), ARIMA(6,1,1) 预测曲线对比

Fig. 9 Prediction curve comparison of LSTM-Encoder-Decoder, GM(1,1), ARIMA(6,1,1)

**结束语** 在前期的数据预处理中使用小波变换进行降噪处理后的 LSTM-Encoder-Decoder 预测模型和 LSTM 神经网络单步预测模型相比其他模型的预测 MAPE 都要低，表明小波降噪在前期的数据预处理中能有效地过滤掉原始矿产品税收

数据中的噪声，降低模型的损失度，增强 LSTM 神经网络模型的泛化能力。基于灰色系统理论的 GM(1,1) 预测模型，对于数据波动较频繁的时间序列数据，预测精度和线性相关性都不及其他预测模型，其适用范围倾向于比较平稳的时间序列数据。基于差分回归的 ARIMA 模型，其预测精度虽然近似达到 90%，但是综合其线性相关性 R，其未来税收的预测无法较好地反映与实际预期值之间的关联性，只能总体反映未来税收预测的中间范围。结合 MAPE 和 R 的综合度量，深度学习中 LSTM 神经网络对于时间序列中隐藏的上下文关系，特别是非平稳时间序列，具有较好的学习能力，而相对于单步预测的 LSTM 神经网络，使用编码器-解码器模型的 LSTM 神经网络的预测精度更高，预测值与预期值的相关性更强。

本文提出的基于 LSTM 神经网络的矿产品税收预测模型，在中长期的预测结果的平均绝对百分比误差虽然较低，但是在个别时间节点存在误差较大的情况。针对单一维度的时间序列数据，LSTM 神经网络模型只能提取有限的隐藏信息，由于矿产品税收受到多种客观因素的影响，如天气、征收率、市场供需等客观因素的影响，要想进一步地提高矿产品税收预测的精度，需要后期在相关领域采集影响因子数据。通过数学建模量化这些影响因子，作为时间序列数据的其他维度信息，构造更高维度的初始输入序列，使得模型能提取更多影响未来税收的隐藏状态，以此在原有的精度之上得到进一步的提升。

## 参考文献

- [1] WANG Y, LI Y, WANG L L, et al. Software stage effort prediction based on analogy and grey model [J]. Computer Science, 2018, 45(S2): 480-487.
- [2] ZHAO Z, WANG J Z, ZHAO J, et al. Using a grey model optimized by differential evolution algorithm to forecast the per capita annual net income of rural households in China [J]. Omega-International Journal of Management Science, 2012, 40(5): 525-532.
- [3] XIANG C S, ZHANG L F. Grain yield prediction model based on grey theory and markov [J]. Computer Science, 2013, 40(2): 245-248.
- [4] MALDONADO-MOLINA M M, WAGENAAR A C. Effects of Alcohol Taxes on Alcohol-Related Mortality in Florida: Time-Series Analyses From 1969 to 2004 [J]. Alcoholism-Clinical And Experimental Research, 2010, 34(11): 1915-1921.
- [5] LIN J L. Application of time series model in customs revenue forecasting [J]. Statistics and consulting, 2008, 2008(5): 46-47.
- [6] HAVIV D, RIVKIND A, BARAK O. Understanding and Controlling Memory in Recurrent Neural Networks [C] // Thirty-sixth International Conference on Machine Learning. 2019: 2663-2671.
- [7] WANG Y Y, SMOLA A, MADDIX D C, et al. Deep Factors for Forecasting [C] // Thirty-sixth International Conference on Machine Learning. 2019.
- [8] CHARLES A, YIN D, ROZELL C. Distributed Sequence Memory of Multidimensional Inputs in Recurrent Networks [J]. Journal of Machine Learning Research, 2017, 18(7): 1-37.
- [9] GERS F A, ECK D, SCHMIDHUBER J. Applying LSTM to

- time series predictable through time-window approach [C] // Proceedings of the 2001 International Conference on Artificial Neural Networks. London: Springer-Verlag, 2001: 669-676.
- [10] SUTSKEVER I, VINYALS O, LE Q V. Sequence to sequence learning with neural networks [C] // Neural Information Processing Systems 2014. Advances in neural information processing systems, 2014: 3104-3112.
- [11] SONG K T, TAN X, QIN T, et al. Tie-Yan Liu. MASS: Masked Sequence to Sequence Pre-training for Language Generation [C] // Thirty-sixth International Conference on Machine Learning. 2019.
- [12] TAY Y, PHAN M C, TUAN L A, et al. Learning to Rank Question Answer Pairs with Holographic Dual LSTM Architecture [C] // The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, 2018. ACM, 2018: 325-324.
- [13] GAO T W, CHAI Y T, LIU Y. Applying long short term memory neural networks for predicting stock closing price [C] // 2017 8th IEEE International Conference on Software Engineering and Service Science. IEEE, 2017: 575-578.
- [14] HE Z, GAO S B, XIAO L, et al. Wider and Deeper, Cheaper and Faster: Tensorized LSTMs for Sequence Learning [C] // Advances in Neural Information Processing Systems 30 (NIPS 2017). Curran Associates Inc, 2017: 1-11.
- [15] JIA L, ZHENG C J. Short-term Forecasting Model of Agricultural Product Price Index Based on LSTM-DA Neural Network [J]. Computer Science, 2019, 46(S2): 62-65, 71.
- [16] ESSIEN A, GIANNETTI C. A Deep Learning Framework for Univariate Time Series Prediction Using Convolutional LSTM Stacked Autoencoders [C] // 2019 IEEE International Symposium on Innovations in Intelligent Systems and Applications. IEEE, 2019: 1-6.
- [17] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural Computation, 1997, 9(8): 1735-1780.
- [18] CHO K, BAHNANAU D, BOUGARES F, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation [J]. EMNLP, 2014: 1724-1734.
- [19] POPOOLA A, AHMAD K. Testing the suitability of wavelet preprocessing for TSK fuzzy models [C] // ICFS 2006: 2006 IEEE International Conference on Fuzzy Systems. IEEE, 2006: 1305-1309.
- [20] ZHAO A, ZHANG D, SHI J Q. Forecasting and Analysis of EUR/USD Exchange Rate Moving Direction with Support Vector Machine [C] // 2018 IEEE 8th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems. IEEE, 2018: 1484-1489.
- [21] ZHANG S X, CONG X R. The Application of Wavelet Analysis in Financial Multiple Change Points Time Series [C] // 2018 5th International Conference on Industrial Economics System and Industrial Security Engineering. IEEE, 2018: 1-6.
- [22] SIRCAR R. An introduction to wavelets and other filtering methods in finance and economics [M]. Utah: Academic Press, 2002: 359.
- [23] MALLAT S G. A Theory of Multiresolution Signal Decomposition [J]. IEEE, 1989, 11(7): 581-767.
- [24] KINGMA D P, BA J. Adam: A Method for Stochastic Optimization [C] // Proceedings of the 3rd International Conference on Learning Representations, 2015.
- [25] SRIVASTAVA N, HINTON G E, KRIZHEVSKY A, et al. Dropout: a simple way to prevent neural networks from overfitting [J]. Journal of Machine Learning Research, 2014, 15(6): 1929-1958.



**WEN Hao**, born in 1994, postgraduate. His main research interests include machine learning and tax informatization.



**CHEN Hao**, born in 1977, Ph.D, professor. His main research interests include uncertain artificial intelligence and so on.

(上接第 429 页)

- [14] WANG Y, ZHANG L, TAN J, et al. HydraDB: a resilient RDMA-driven key-value middleware for in-memory cluster computing [C] // 2015 SC-International Conference for High Performance Computing, Networking, Storage and Analysis. IEEE, 2017: 22.
- [15] WANG X, WU Y, JIANG X H, et al. Incremental Parallel Fast Clustering Based on DBSCAN Algorithm Under Large-scale Data Sets [J]. Computer Applications and Software, 2018, 35(4): 269-280.
- [16] APILETTI D, GARZA P, PULVIRENTI F. A Review of Scalable Approaches for Frequent Itemset Mining [C] // East European Conference on Advances in Databases and Information Systems. Springer International Publishing, 2015: 243-247.
- [17] WANG B, ZHANG C, SONG L, et al. Design and optimization of DBSCAN Algorithm based on CUDA [J]. Computer Science, 2015, 40(5): 553-556.
- [18] CHEN R, ZHANG Y, ZHANG J, et al. Design and Optimizations of the MD5 Crypt Cracking Algorithm Based on CUDA [C] // International Conference on Cloud Computing. Springer International Publishing, 2014: 155-164.
- [19] NING J F. DBSCAN Text Clustering Algorithm Based on Spark Framework [J]. Journal of Shantou University (Natural Science Edition), 2018, 33(2): 73-80.
- [20] PENG X Y, YANG Y B, WANG C D, et al. An Efficient Parallel Nonlinear Clustering Algorithm Using MapReduce [C] // 2016 IEEE International Parallel and Distributed Processing Symposium Workshops. IEEE, 2016: 1473-1476.



**DENG Ding-sheng**, born in 1978, associate professor. His main research interests include algorithm analysis and design and so on.