

基于门控权重单元的多变量时间序列预测

张冬梅¹, 李金平¹, 李江^{2†}, 余想¹, 宋凯旋¹

(1. 中国地质大学 计算机学院, 湖北 武汉 430074;

2. 湖北省自然资源厅 信息中心, 湖北 武汉 430071)

摘要:多变量时间序列各变量间依赖性较强, 数据变化趋势不明显, 预测难度高. 传统研究采用带门控机制的循环神经网络及变体进行预测, 但序列间存在相互依赖关系, 突变数据建模预测不精确. 基于信息熵, 本文提出一种新的改进门控权重单元, 利用信息熵技术量化数据序列的变化程度, 动态调整权重矩阵刻画数据的变化趋势. 基于 4 个公开数据集分别进行实验, 实验结果表明新模型比传统循环神经网络模型具有更好的预测性能.

关键词:多变量时间序列; 门控机制; 循环神经网络; 门控权重单元; 信息熵

中图分类号:TP391

文献标志码:A

Multivariate Time Series Prediction Based on Gating Weight Unit

ZHANG Dongmei¹, LI Jinping¹, LI Jiang^{2†}, YU Xiang¹, SONG Kaixuan¹

(1. School of Computer Science, China University of Geosciences, Wuhan 430074, China)

2. Information Center, Department of Natural Resources of Hubei Province, Wuhan 430071, China)

Abstract: There is strong dependence among the variables of multivariate time series, which makes the data trend unobvious and the prediction difficult. Traditionally, recurrent neural network with gating mechanisms and its variants are used for prediction. But the interdependence between sequences makes the prediction result of mutation data not accurate. Based on information entropy, a new modified gating weight unit is presented. The change degree of data is quantified by using information entropy to dynamically adjust the weight matrix and describe the trend of data. The experiment is conducted with four public data sets. The experimental results show that the proposed model has better prediction performance than the traditional recurrent neural network.

Key word: multivariate time series; gated mechanism; recurrent neural network; gating weight unit; information entropy

多变量时间序列预测根据历史数据对未来进行预测, 是近年来研究的热点问题. 多变量序列间存在的相互依赖关系难以学习、特别是突变数据的变化

趋势挖掘困难, 导致该类问题的预测精度不高^[1]. 早期研究主要是通过建立加权滑动平均模型、差分整合移动平均自回归模型等模型进行预测, 简单直观,

* 收稿日期: 2020-10-25

基金项目: 国家自然科学基金联合基金重点支持项目 (U1911205), Key Projects Supported by National Nature Science Foundation of China Joint Fund (U1911205); 国家重大科技专项 (2016ZX05014-003-003), National Science and Technology Major Project of China (2016ZX05014-003-003)

作者简介: 张冬梅 (1972—), 女, 湖北武汉人, 中国地质大学教授, 博士生导师

† 通信联系人, E-mail: johnlee1124@126.com

但极易损失变量间的动态依赖信息.近年基于浅层机器学习的非参数模型是多变量时间序列预测问题的主流方法,可有效克服序列非线性和变量依赖性强等问题,如支持向量回归(Support Vector Regression,SVR)、高斯过程回归(Gaussian Process Regression,GPR)等具有更好的性能,但浅层机器学习方法易过拟合,影响预测精度^[2].

早期网络模型不能解决长期依赖问题,存在梯度消失和梯度爆炸现象^[3-4].目前多利用深度学习中的循环神经网络(Recurrent Neural Network,RNN)模型解决多变量时序预测问题^[5].Graves在RNN单元中引入记忆单元和门控机制,提出长短期记忆网络(Long Short-Term Memory,LSTM)^[6],解决RNN的长期依赖问题.目前对于循环神经网络的改进多基于LSTM门控机制的扩展,如Zhou等提出的最小门控单元(Minimal Gated Unit,MGU)^[7]、Gers等提出的窥视孔连接^[8]、Cho等提出的门控循环单元(Gate Recurrent Unit,GRU)^[9]、刘颀曦等提出的混合门单元^[10].各类LSTM变体虽然对LSTM的结构进行了改进优化,但没有根本解决门控组件难以学习突变数据变化趋势的问题,对变化剧烈的时序数据预测效果较差^[11-12].

信息熵(Information Entropy)技术通过概率分布计算刻画数据分布变化,Haghizadeh等使用信息熵度量灾害突发信息,在洪涝灾害预防方面取得较好

效果^[13].Naghibi等使用信息熵模型预测地下水变化趋势^[14].信息熵对数据变化敏感,可较好度量时序数据的变化程度.针对非稳态数据预测问题,本文对网络单元的门控机制进行扩展,提出一种新的E-LSTM单元(Entropy-LSTM,E-LSTM).与同类研究工作不同的是通过引入信息熵技术量化数据变化程度,设置两个动态调整权重矩阵代替传统遗忘门权重矩阵,刻画数据变化趋势.新单元通过信息熵动态调整权重矩阵,克服了传统LSTM单元难以挖掘突变数据段变化趋势的问题,有效提升了模型的预测精度.

本文主要贡献有2个方面:

- 1)融合信息熵和循环神经网络技术进行时序序列预测;
- 2)引入信息熵动态调整权重矩阵扩展门控机制,对时序数据中的突变特征提出新的量化方法和技术,相较于传统LSTM预测精度更高.

1 模型

1.1 RNN 预测模型

基于RNN时序预测模型框架,针对不同问题模型中的RNN单元采用不同的改进策略,如神经网络模型中的RNN-unit可以用简单循环神经网络(Simple-Recurrent Neural Network,Simple-RNN)^[15]、LSTM单元等进行替换,整个RNN模型框架如图1

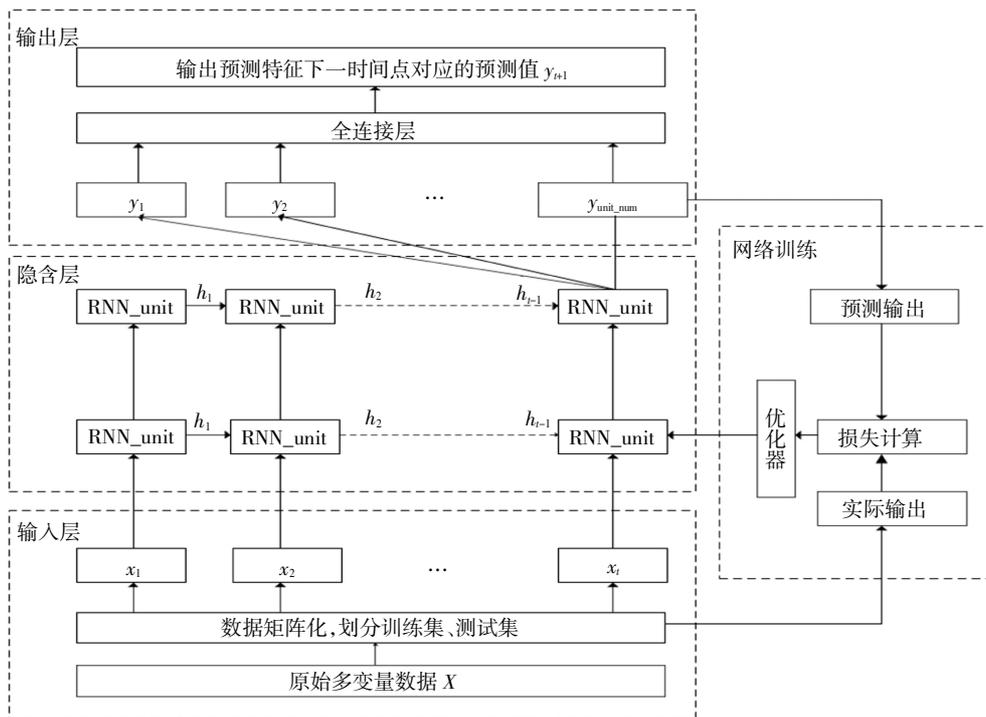


图1 基于RNN的时序预测模型

Fig.1 Time series prediction model based on RNN

所示. t 时刻隐藏单元 RNN-unit 接收网络前一时刻的隐藏层输出 h_{t-1} 和当前输入层输入的 x_t , 按隐含层更新方法计算当前时刻的最后一层隐藏层输出 y_{t+1} .

1.2 Simple-RNN 模型

循环神经网络是内部存在自连接的神经网络, 通过保持数据中的变量依赖关系处理具有长程相关性的时序序列. 由 Jordan 和 Elman 提出的循环神经网络框架 (称为简单循环网络) 是目前广泛流行的 RNN 基础版本.

Simple-RNN 单元隐含态更新规则为:

$$h_t = \tanh(W * [h_{t-1}, x_t] + b) \quad (1)$$

式中: h_{t-1} 和 h_t 分别表示 $t-1$ 时刻和 t 时刻的输出, W 为权重调节矩阵. Simple-RNN 单元如图 2 所示.

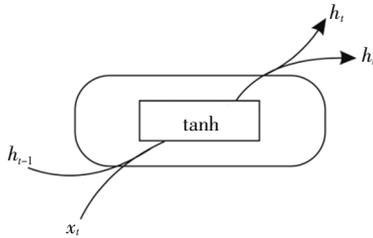


图 2 Simple-RNN 单元

Fig.2 Simple Recurrent Neural Network unit

Simple-RNN 结构过于简单, 当模型深度较大时无法获取所需的长程信息, 出现“梯度消失”、“梯度爆炸”等现象.

1.3 LSTM 模型

LSTM 是一种特殊的 RNN 单元, 引入记忆单元和门控机制的概念. 记忆单元存储状态信息, 门控机制用于控制何时及如何更新记忆单元的状态.

LSTM 中含有三类门控单元, 分别为输入门、遗忘门和输出门, 如图 3 所示.

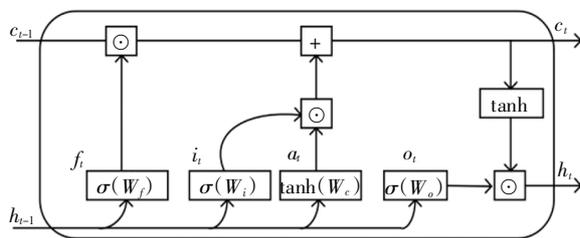


图 3 LSTM 单元

Fig.3 Long Short-Term Memory unit

隐含层状态在 t 时刻的传递过程为:

$$f_t = \sigma(W_f * [h_{t-1}, x_t] + b_f) \quad (2)$$

$$o_t = \sigma(W_o * [h_{t-1}, x_t] + b_o) \quad (3)$$

$$i_t = \sigma(W_i * [h_{t-1}, x_t] + b_i) \quad (4)$$

$$a_t = \tanh(W_c * [h_{t-1}, x_t] + b_c) \quad (5)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot a_t \quad (6)$$

$$h_t = o_t \odot \tanh(c_t) \quad (7)$$

式中: i_t, f_t, o_t 及 W_i, W_f, W_o 分别表示输入门、遗忘门和输出门 t 时刻的计算结果和权重矩阵, c_t 为记忆单元状态, h_{t-1} 和 h_t 分别表示 $t-1$ 时刻和 t 时刻的输出, $*$ 代表矩阵相乘, \odot 表示元素对应相乘, 单元间通过控制单元状态 c_t 和隐含层状态 h_t 进行信息传递.

1.4 E-LSTM 模型

研究表明 LSTM 内部的门控组件在训练过程中随机性较强, 难以精确学习数据变化趋势, 导致存在突变的数据段预测效果较差.

信息熵通过概率描述事件分布衡量不同概率事件包含的信息量. E-LSTM 通过在门控机制中加入与信息熵动态关联的权重矩阵, 使网络单元在调节权重参数时能自适应动态挖掘异常数据的变化程度特征.

1.4.1 信息熵

随机变量 X 的信息熵 H 定义如下:

$$H(X) = E[-\ln(p(X))] \quad (8)$$

式中: X 为 $\{x_1, x_2, \dots, x_n\}$, E 是期望, $p(X)$ 是 X 的概率质量函数.

当样本有限时, 随机变量 X 的信息熵 H 定义如下:

$$H(X) = - \sum_{i=1}^n P(x_i) \log_b P(x_i) \quad (9)$$

计算信息熵时, 已知某种序列值域 $Z = [c_{\min}, c_{\max}]$ 中包含 N 个数据, 按升序将 Z 划分为 10 等份, 记第 l 个区间为 $Z_l, l \in [1, 10]$. 假设序列落在区间 Z_l 中的数据点的个数为 num_Z , 则序列信息熵计算概率为 $\frac{\text{num}_Z}{N}$.

1.4.2 E-LSTM 模型

在 LSTM 门控组件中, 遗忘门对模型性能影响最大^[12]. 当前门控机制研究多通过减少门控单元加快模型训练速度, 难以挖掘出数据变化趋势. 本文设计新的 E-LSTM 单元, 引入基于信息熵值的改进权重矩阵代替传统遗忘门矩阵. 新单元加强了遗忘门的记忆能力, 根据信息熵量化数据特征, 自适应动态调整特征权重.

E-LSTM 模型将原始遗忘门权重矩阵设计为两个新的遗忘门矩阵 W_{f1}, W_{f2} , 分别与已激活信息熵 $\sigma(E)$ 和 $1-\sigma(E)$ 对应元素相乘得到临时门控输出 $f1, f2$, 如图 4 所示.

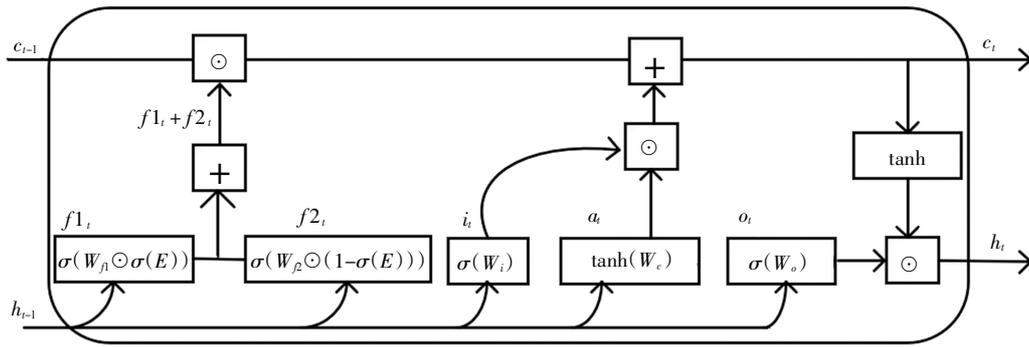


图 4 E-LSTM 单元

Fig.4 Entropy-Long Short-Term Memory unit

按时间步计算, 临时门控输出结果 $f1_t$ 和 $f2_t$ 相加得到整体遗忘门 f_i , E-LSTM 使用整体遗忘门代替传统 LSTM 遗忘门参与计算, 其余更新状态的方式与 LSTM 保持相同. 对应门控计算流程为:

$$f1_t = \sigma(W_{f1} * [h_{t-1}, x_t] \odot \sigma(E) + b_f) \quad (10)$$

$$f2_t = \sigma(W_{f2} * [h_{t-1}, x_t] \odot (1 - \sigma(E)) + b_f) \quad (11)$$

$$f_i = f1_t + f2_t \quad (12)$$

式中: $f1_t$ 、 $f2_t$ 分别代表两个临时遗忘门; W_{f1} 、 W_{f2} 分别代表两个临时遗忘门矩阵; E 为输入数据的信息熵矩阵.

1.4.3 E-LSTM 模型反向传播

E-LSTM 的工作方式与 LSTM 基本相同, 区别是在反向传播时根据序列的熵值自适应动态调整权重.

根据 E-LSTM 向前传播规则以及公式(12)将整体遗忘门 f_i 表示为两个临时遗忘门 $f1_t$ 和 $f2_t$, 忽略偏置项, 使用 I_t 表示输入矩阵, E-LSTM 正向传播矩阵 z_t 为:

$$z_t = W * I_t =$$

$$\begin{bmatrix} w_c & u_c \\ w_i & u_i \\ w_{f1} \odot \sigma(E) & u_{f1} \odot \sigma(E) \\ w_{f2} \odot (1 - \sigma(E)) & u_{f2} \odot (1 - \sigma(E)) \\ w_o & u_o \end{bmatrix} * \begin{bmatrix} x_t \\ h_{t-1} \end{bmatrix} \quad (13)$$

式中: w_d, u_d 是权重 W 矩阵分块表示后的结果, $d \in \{i, f1, f2, c, o\}$.

设定损失函数为 L , 已知 $\delta h_t = \frac{\partial L}{\partial h_t}$, δ 为偏导, 推得:

$$\delta W_t = \delta z_t * (I_t)^{-1} \quad (14)$$

将时间 T 内的所有 δW_t 累加, 得到权重矩阵梯度 δW .

$$\delta W = \sum_{t=1}^T \delta W_t \quad (15)$$

δW 中整体遗忘门权重矩阵梯度由 $\delta f1_t$ 、 $\delta f2_t$ 共同确定, 其中 $\delta f1_t$ 与 $\sigma(E)$ 呈正相关, $\delta f2_t$ 与 $\sigma(E)$ 呈负相关. E-LSTM 分别通过 $\delta f1_t$ 、 $\delta f2_t$ 调整遗忘门权重矩阵权值, 刻画信息变化程度.

2 实验与分析

本节结合 4 个多变量时序数据集评估新提出的 E-LSTM 模型, 并将其性能与其他模型 (Simple-RNN、LSTM、GRU) 进行比较, 所有 RNN 模型均基于 TensorFlow 框架实现.

2.1 度量指标

本文选择平均绝对误差 MAE 和平均绝对百分比误差 MAPE 作为评估模型性能指标. 并使用均方根误差 RMSE 作为模型训练的损失函数, 训练时依据各批量 RMSE 值进行反向传播操作. 度量指标定义如下:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_{\text{predict}} - y_{\text{actual}}| \quad (16)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_{\text{predict}} - y_{\text{actual}})^2}{n}} \quad (17)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_{\text{predict}} - y_{\text{actual}}|}{|y_{\text{actual}}|} \quad (18)$$

式中: n 是样本总数, y_{predict} 是预测值, y_{actual} 是真实值.

2.2 数据说明

实验使用的 4 个多变量时序数据集为 3 个股市数据集和 1 个空气湿度指数预测数据集, 划分前 60% 为训练集, 后 40% 为测试集.

股票数据集: 3 个股票公开数据集收录于 Kaggle Datasets 机器学习库, 分别记作“BAC”、“GS”、“C”. 每个数据集有 2 517 条记录, 包含 5 个有效特征 (收盘

价、开盘价、最低价、最高价、成交量)。

空气湿度指数预测数据集:收录于 UCI 机器学习库,主要记录在雅典沿海地区的海面湿度数据,数据集记作“USV”。该数据集有 1 672 条记录,包含 4 个有效特征(设备号、湿度、温度、报告时间)。数据缺失时,选择用对应属性平均值补齐。

2.3 参数调优

选择 LSTM 网络作为基准结构。循环神经网络训练均以小批量进行,批次大小为 128,所有模型训练 2 000 代。构建 E-LSTM 预测模型前,预先调整网络模型的隐含层层数和各层神经元数量。其中隐含层层数的候选集为 {1,2,3}, 神经元数目候选集为 {16,32,64,96,128}。

基准结构在设置超参数时采用“控制变量法”。固定其他参数,从候选集{16,32,64,96,128}中依次选择神经元数目。随着隐含层神经元数量的增加,模型预测性能逐步提升,并在神经元数量为 32 或 64 时达到候选集参数的最高精度,当神经元数量超过 64 时会出现过拟合导致精度下降。基准模型不同神经元节点数量的具体性能评价指标如表 1 所示。后续实验中,将隐含层神经元的数量设置为 64。

表 1 LSTM 不同神经元节点数量的指标值

Tab.1 The index value of different neuron nodes size in LSTM

Neuron size	RMSE	MAE	MAPE
16	3.934	2.611	0.050 2
32	1.199	0.917	0.017 9
64	1.218	0.887	0.017 4
96	1.514	1.073	0.021 0
128	1.835	0.884	0.017 0

注:灰色背景表示的数据是实验的最佳结果

确定神经元数目后更改隐含层层数,结果如表 2 所示。

表 2 LSTM 不同隐含层层数的精度变化

Tab.2 Accuracy of LSTM with different hidden layers

Layer size	RMSE	MAE	MAPE
1	1.378	0.952	0.018 7
2	1.218	0.887	0.017 4
3	1.662	1.290	0.025 4

注:灰色背景表示的数据是实验的最佳结果

实验结果表明,层数较少的深度学习模型不能满足复杂数据集的精度要求,而层数较多又会存在“长期依赖”问题,结果如图 5 所示。根据实验结果隐含层层数设为 2 时模型预测效果最优。

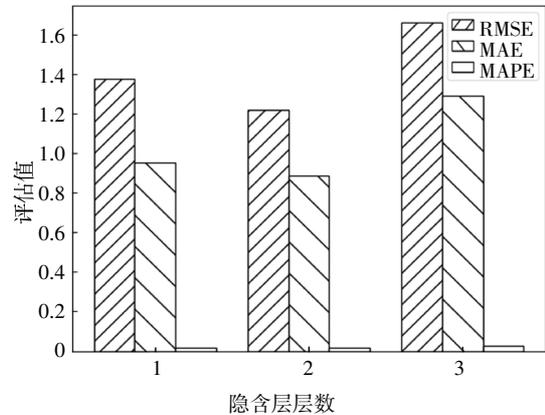


图 5 LSTM 模型中隐含层层数的影响

Fig.5 Influence of hidden layer of LSTM model

为公平起见,实验中所有深度学习模型都使用相同的隐藏层和神经元数量,最终设置 2 层隐含层单元,每层包含 64 个神经元,有且仅有 1 个神经元的全连接层作为输出层。

2.4 权重矩阵分析

为说明 E-LSTM 模型两个基于信息熵的权重矩阵的动态调整效果,对训练过程中 W_{f1} 和 W_{f2} 的变化进行实验。由于权重矩阵维度较高、不易衡量特征向量的变化趋势,使用频率直方图描述训练过程中权重矩阵的变化,其中横坐标为权重矩阵 W_{f1} 和 W_{f2} 的取值,纵坐标为权重矩阵值出现次数。以数据集“BAC”为例进行实验,结果如图 6 所示,图 6(a)与图 6(b),图 6(c)与图 6(d)分别为训练初期和结束训练时的 W_{f1} 和 W_{f2} 的频率分布直方图。对比图 6 (a)、图 6(c)发现 W_{f1} 早期部分权重值分布多集中于 0,后期部分权重系数上升至 1,这说明 W_{f1} 捕获到了数据的突变特征,改变了权重系数。对比图 6(b)、图 6(d)发现 W_{f2} 在训练过程中发生大幅度变化,0 附近的权重参数频率从 4 000 下降到 3 000,0.5 和 -0.5 附近的权重参数频率显著增加,说明 W_{f2} 权重变化与数据变化趋势相关。

实验结果表明,基于信息熵的权重矩阵 W_{f1} 和 W_{f2} 在训练过程中能够识别出数据的变化趋势,并做相应的权重调整。

2.5 时间分析

为衡量新引入的门控权重矩阵对算法时间效率的影响,以数据集 C 为例,将 E-LSTM 与基准模型

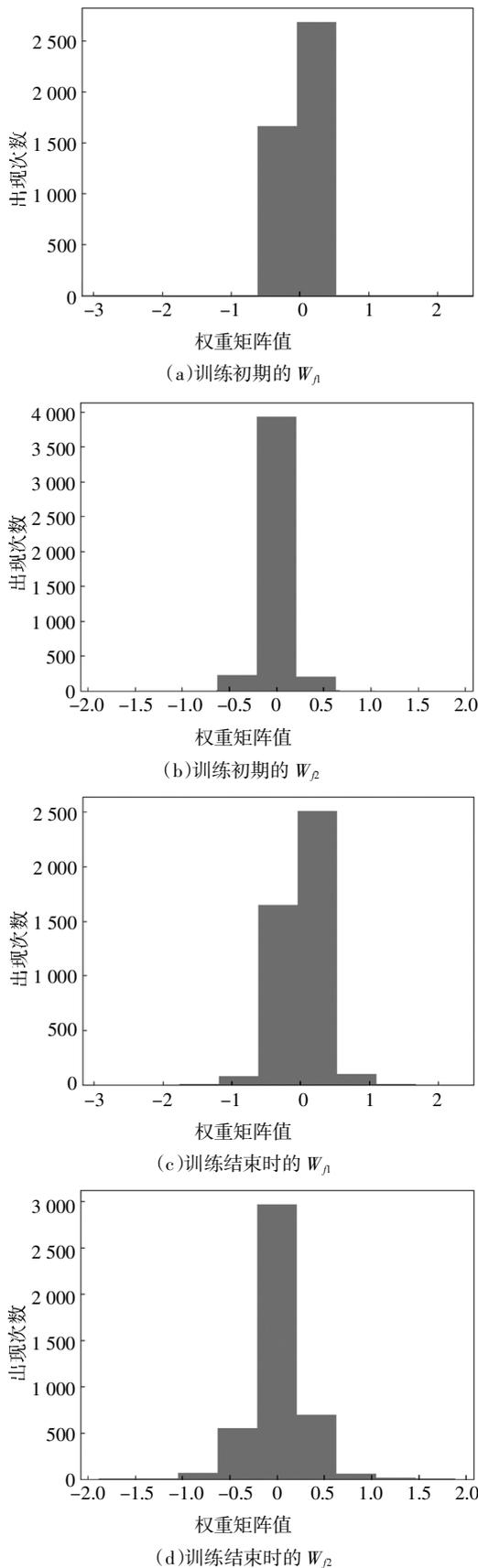


图6 不同训练时期权重矩阵 W_{f1} 和 W_{f2} 的动态变化
Fig.6 Dynamic change of weight matrix W_{f1} and W_{f2} in different periods

型 LSTM 开展时间开销对比分析. 设置时间消耗比 p 作为测试指标, 定义如下:

$$p = \frac{t_{E-LSTM} - t_{LSTM}}{|t_{LSTM}|} \quad (19)$$

式中: t_{E-LSTM} 是新模型消耗时间; t_{LSTM} 是基准结构消耗时间; p 为时间消耗比.

实验对比相同迭代代数和收敛条件下的时间开销, 结果分别如表 3 和表 4 所示. 设定迭代代数为 150 代, 基准模型时间开销为 10 808.98 ms, 改进模型为 11 802.453 ms, 时间消耗比为 0.091, 相同代数下改进模型的计算开销略大. 但由于增加了门控权重矩阵计算, 在相同收敛条件下, 改进模型的初始训练误差低于基准模型, 拥有更快的收敛速度. 基准模型收敛总体时间开销为 14 413.4 72 ms, 改进模型为 11 574.089 ms, 时间消耗比为 -0.196. 两类模型收敛曲线如图 7 所示. 综上所述新算法效率更高.

表 3 相同代数下性能比较

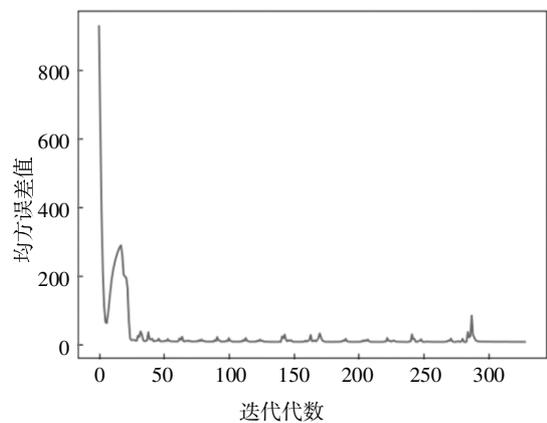
Tab.3 Performance comparison under the same algebra

算法	消耗时间(ms)	p
LSTM	10 808.98	0.091
E-LSTM	11 802.453	

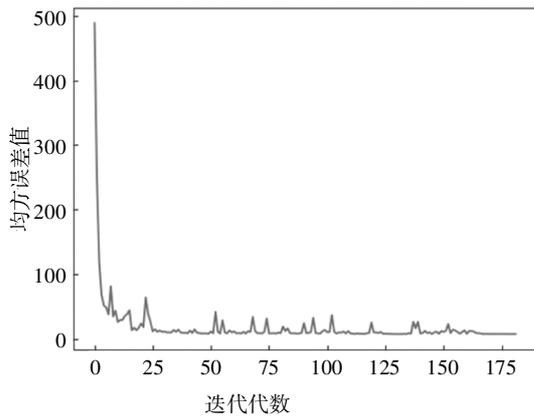
表 4 收敛条件下性能比较

Tab.4 Performance comparison under the same converger condition

算法	消耗时间(ms)	收敛时迭代代数	p
LSTM	14 413.472	340	-0.196
E-LSTM	11 574.089	175	



(a) LSTM 模型收敛曲线



(b)E-LSTM 模型收敛曲线

图 7 LSTM 和 E-LSTM 模型在数据集 C 上的收敛曲线

Fig.7 Convergence curves of LSTM and E-LSTM on dataset C

2.6 实验对比

根据调优设置参数进行训练,各模型在 4 个测试集上实验结果如表 5 所示.其中加粗部分表示传统循环神经网络(GRU,LSTM,Simple-RNN)的最优效果,灰色背景表示添加 E-LSTM 后所有模型的最优效果.实验得到以下结论:

传统循环神经网络方法中,Simple-RNN 结构简单,针对不同数据集效果差异大,数据集越复杂效果越差.LSTM 和 GRU 两者性能相当且稳定性均优于 Simple-RNN.

与传统循环神经网络方法相比,本文提出的 E-LSTM 模型在 4 个数据集上的 MAE、RMSE 和 MAPE 指标精度更高.

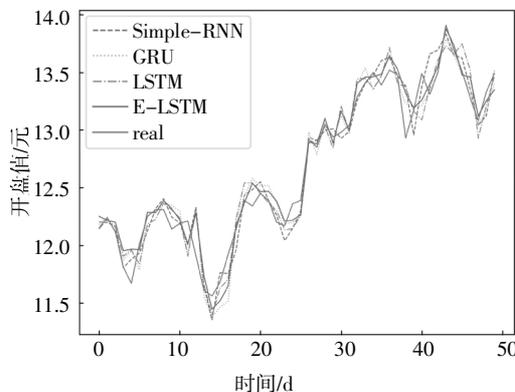
表 5 各类 RNN 模型性能比较

Tab.5 Performance comparison of various RNN models

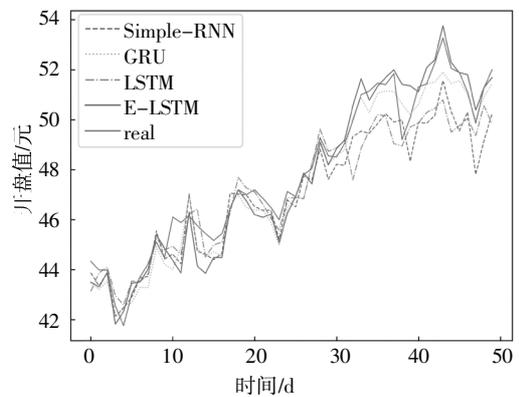
Dataset		BAC			C			GS			USV		
Parameter	Modal	RMSE	MAE	MAPE									
(2,64)	Simple-RNN	0.181	0.129	0.009 4	1.678	1.318	0.026 0	2.630	2.171	0.012 7	4.301	3.908	0.101 0
(2,64)	LSTM	0.182	0.138	0.009 6	1.218	0.887	0.017 4	1.523	1.094	0.006 8	4.118	3.469	0.091 0
(2,64)	GRU	0.206	0.143	0.010 6	1.307	0.933	0.018 4	1.481	1.075	0.006 6	4.566	3.973	0.103 1
(2,64)	E-LSTM	0.177	0.125	0.008 8	1.140	0.834	0.017 1	1.472	1.031	0.006 4	4.007	3.418	0.089 0

注:加粗数据表示在传统递归神经网络(Simple-RNN,LSTM,GRU)中的最佳结果,灰色背景表示添加 E-LSTM 后的最佳实验结果

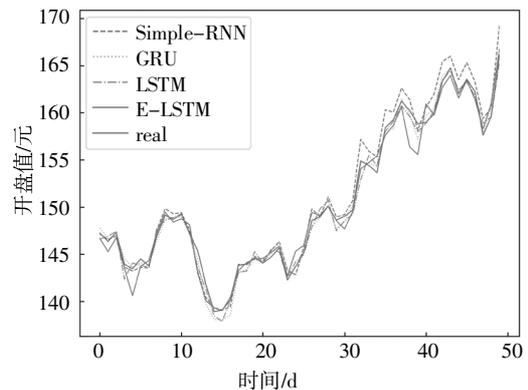
实验 4 个数据集预测部分曲线如图 8 所示.显然 E-LSTM 模型预测结果在所有数据集上与真实值最接近,预测精度最高.以数据集 C 为例,传统方法(GRU,LSTM,Simple-RNN)在突变数据段处(图 8(b))的预测值与真实值误差较大,由于基于信息熵的权重矩阵能够更好地量化数据的变化趋势,E-LSTM 模型在数据变化段的预测值与真实值基本一致,预测效果更优.



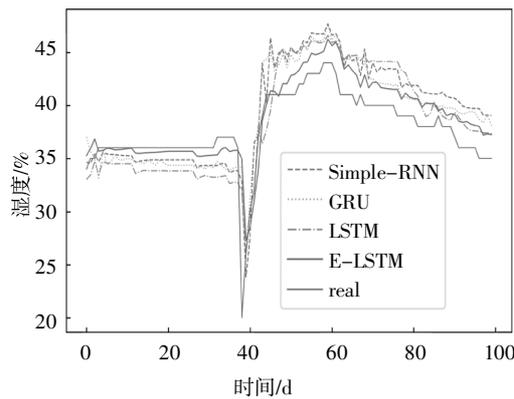
(a)BAC 数据集的“开盘值”预测曲线



(b)C 数据集的“开盘值”预测曲线



(c)GS 数据集的“开盘值”预测曲线



(d)USV 数据集的“湿度”预测曲线

图 8 E-LSTM 和各类预测模型的预测曲线

Fig.8 Prediction curves of E-LSTM and various prediction models

3 结论

本文提出一种基于信息熵改进门控权重单元的时间序列预测方法,并与 Simple-RNN、LSTM、GRU 等传统预测模型进行对比分析.研究证实 E-LSTM 可根据输入数据的特征动态调整新权重矩阵的权重系数,更好地刻画数据的变化趋势,实验结果表明,本文提出的 E-LSTM 模型对局部突变数据更为敏感,预测精度更高.

在后续的研究中将进一步结合多重分形等技术改进模型的门控权重,细粒度刻画数据多尺度的变化趋势,提升模型的预测效率.

参考文献

- [1] BOX G, JENKINS G M, REINSEL G C, *et al.* Time series analysis: forecasting and control, 5th edition [J]. Journal of the Operational Research Society, 2015, 22(2): 199—201.
- [2] BENGIO Y, LECUN Y. Scaling learning algorithms toward AI [M]// Large-Scale Kernel Machines. The MIT Press, 2007, 34(5): 1—41.
- [3] BENGIO Y, SIMARD P, FRASCONI P. Learning long-term dependencies with gradient descent is difficult [J]. IEEE Transactions on Neural Networks, 1994, 5(2): 157—166.
- [4] TSIRONI E, BARROS P, WEBER C, *et al.* An analysis of convolutional long-short term memory recurrent neural networks for gesture recognition [J]. Neurocomputing, 2017, 268: 78—86.
- [5] 许淼, 刘宏飞, 初凯. 基于 AM-LSTM 模型的共享单车时空需求预测 [J]. 湖南大学学报(自然科学版), 2020, 47(12): 77—85.
XU M, LIU H F, CHU K. Spatiotemporal demand prediction of bike-sharing based on AM-LSTM model [J]. Journal of Hunan University (Natural Sciences), 2020, 47(12): 77—85. (In Chinese)
- [6] GRAVES A. Long short-term memory [M]. Berlin: Springer, 2012: 1735—1780.
- [7] ZHOU G B, WU J X, ZHANG C L, *et al.* Minimal gated unit for recurrent neural networks [J]. Int Journal of Automation and Computing, 2016, 13 (3): 226—234.
- [8] GERS F A, SCHMIDHUBER J. Recurrent nets that time and count [C]//Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. Neural Computing: New Challenges and Perspectives for the New Millennium. Como, Italy: IEEE, 2000: 189—194.
- [9] CHO K, VAN MERRIENBOER B, BAHDANAU D, *et al.* On the properties of neural machine translation: encoder-decoder approaches [C]//arXiv Preprint arXiv: 1409.1259, 2014.
- [10] 刘颖羲, 陈松灿. 基于混合门单元的非平稳时间序列预测 [J]. 计算机研究与发展, 2019, 56(8): 1642—1651.
LIU J X, CHEN S C. Non-stationary multivariate time series prediction with MIX gated unit [J]. Journal of Computer Research and Development, 2019, 56(8): 1642—1651. (In Chinese)
- [11] JOZEFOWICZ R, ZAREMBA W, SUTSKEVER I. An empirical exploration of recurrent network architectures [C]//Proceedings of the 32nd International Conference on Machine Learning. Lille, France: JMLR, 2015: 2342—2350.
- [12] YANG Y. Hybrid method for short-term time series forecasting based on EEMD [J]. IEEE Access, 2020, 8: 61915—61928.
- [13] HAGHIZADEH A, SIAHKAMARI S, HAGHIABI A H, *et al.* Forecasting flood-prone areas using Shannon's entropy model [J]. Journal of Earth System Science, 2017, 126(3): 1—11.
- [14] NAGHIBI S A, POURGHASEMI H R, POURTAGHI Z S, *et al.* Groundwater qanat potential mapping using frequency ratio and Shannon's entropy models in the Moghan watershed, Iran [J]. Earth Science Informatics, 2015, 8(1): 171—186.
- [15] JEFFREY L. ELMAN. Distributed representations, simple recurrent networks, and grammatical structure [J]. Machine Learning, 1991, 7 (2/3): 195—225.