

DOI: 10.16652/j.issn.1004-373x.2021.14.001

引用格式:沐年国,姚洪刚.基于注意力机制的循环神经网络对金融时间序列的应用[J].现代电子技术,2021,44(14):1-5.

基于注意力机制的循环神经网络对 金融时间序列的应用

沐年国, 姚洪刚

(上海理工大学, 上海 200093)

摘 要: 金融时间序列由于高噪声性以及序列间的相关性,导致传统模型的预测精度和泛化能力往往较低。为了克服这一问题,提出一种基于注意力机制的循环神经网络预测模型。利用金融时间序列的技术指标作为特征序列,通过GRU得到隐藏状态,再利用注意力机制将其重构,并与目标序列一起作为新的GRU的输入,提高目标序列的预测效果。利用上证综指数据,分别使用加入注意力机制的门控循环网络与标准门控循环网络进行预测。在6个输入维度的情况下,基于注意力机制的GRU模型平均绝对百分比误差为0.76%,低于标准GRU模型的0.90%;在48个输入维度的情况下,基于注意力机制的GRU模型平均绝对百分比误差为0.73%,低于标准GRU模型的1.61%。结果表明,加入注意力机制后门控循环网络的预测效果得到提升,并且在特征序列的输入维度增大时,其预测效果提升更为明显。

关键词: 循环神经网络; 金融时间序列; 注意力机制; GRU模型; 预测模型; 隐藏状态重构; 实证分析

中图分类号: TN711-34

文献标识码: A

文章编号: 1004-373X(2021)14-0001-05

Application of attention mechanism based recurrent neural network in financial time series

MU Nianguo, YAO Honggang

(University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: Due to the high noise of financial time series and the correlation between the series, the prediction accuracy and generalization ability of the traditional models are often low. In order to overcome these problems, a recurrent neural network prediction model based on the attention mechanism is proposed in this paper. By taking the technical index of financial time series as the feature sequence, the hidden state is obtained by the gated recurrent unit (GRU). The obtained hidden state is reconstructed with the attention mechanism. Together with the target sequence, it is used as the input of a new GRU to improve the prediction effect of the target sequence. With the SSE composite index data, the GRU with the attention mechanism and the standard GRU are used to make predictions. In the case of 6 input dimensions, the average absolute percentage error of the GRU model based on the attention mechanism is 0.76%, which is lower than 0.90% of the standard GRU model. In the case of 48 input dimensions, the average absolute percentage error of the GRU model based on the attention mechanism is 0.73%, which is lower than 1.61% of the standard GRU model. The results show that the prediction effect of the gated recurrent network with attention mechanism is improved, and when the input dimension of the feature sequence increases, the improvement of the prediction effect is more obvious.

Keywords: recurrent neural network; financial time series; attention mechanism; GRU model; forecasting model; hidden state reconstruction; empirical analysis

0 引 言

传统时间序列模型受限于固定的模型框架,无法对复杂的金融时间序列做出准确的预测,而基于深度学习

的方法从数据本身出发能更好地应对“非理想化”的时间序列。近年来,具有非线性处理能力的方法,如支持向量机^[1]、BP神经网络^[2]、小波神经网络^[3]以及循环神经网络^[4]被应用于分析时间序列,也有经典计量模型与机器学习方法的组合模型^[5-6],取得了不错的效果。

目前而言,RNN(循环神经网络)及其变种LSTM(长

收稿日期:2020-11-24

修回日期:2021-01-13

基金项目:国家自然科学基金青年科学基金项目(11701370)

短期记忆网络)由于其关注历史信息的网络结构而广泛被应用。LSTM可与进化算法结合以提升预测能力^[7-8]。Althelaya等使用双向LSTM以及堆叠LSTM提高性能^[9]。基于金融时间序列高噪声的特点,也有在使用LSTM预测之前,对原始序列做相关处理:Li和Tam对原始序列进行小波去噪^[10];Bao等在此基础上加入堆叠自编码器^[11];Singh等利用双向二维主成分分析对原始序列降维提取特征^[12];Zhang等在LSTM单元内的遗忘门以及输入门后将状态向量分解为多频率信息,最后在输出门前做聚合重构^[13]。

可以看出,一般通过优化网络结构或者对原始数据进行前期处理两个角度来提升预测能力。然而,对于特征序列对目标序列在时间维度上的影响的关注却很少,本文利用被广泛应用于情感分析^[14]和机器翻译^[15]的注意力机制,对只有特征序列输入的RNN的每一时间步的隐藏状态通过注意力机制进行重构,并与目标序列共同作为新的RNN输入进行预测。结果表明,加入注意力机制的RNN具有较好的预测性能,尤其是在特征维度增加的情况下,其预测效果相较于标准RNN提升更为明显。

1 方法与模型

1.1 GRU原理和结构介绍

循环神经网络的网络层接收当前时刻的输入 x_t 以及上一个时刻的网络状态 h_{t-1} ,得到当前时刻的网络状态 h_t 并作为下一时刻的输入。循环神经网络关注历史因素的影响使其适合处理时间序列问题,其工作原理如下:

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

$$o_t = g(h_t) \quad (2)$$

式(1)表示状态更新过程,式(2)表示预测输出过程。式中: x_t 为 t 时刻的特征输入; h_t 为 t 时刻的状态向量; o_t 为 t 时刻的输出。

标准的循环神经网络更新状态向量常常表示如下:

$$h_t = \sigma(w_s x_t + w_h h_{t-1} + b) \quad (3)$$

式中激活函数 σ 多采用tanh函数。

由于标准循环神经网络的结构过于简单,常常无法长期保存有效信息,即短时记忆问题。为了有效延长这种短时记忆,提出LSTM(长短期记忆网络)并取得了不错的效果,同时其简化的变种门控循环网络(Gated Recurrent Unit, GRU)也被广泛应用。GRU将LSTM的内部状态向量和输出向量合并为状态向量,同时门控数量也由3个(遗忘门、输入门、输出门)减少到2个(复位门、更新门)。虽然GRU是在LSTM的基础上做出简化,

但在大部分情况下两者的效果相差不大,更为重要的是GRU的参数相比于LSTM要少1/3,计算代价降低的同时减轻了过拟合的可能。

GRU的网络结构如图1所示,状态的更新规则如下:

$$r = \text{sigmoid}(w_r[h_{t-1}, x_t] + b_r) \quad (4)$$

$$\tilde{h}_t = \tanh(w_h[r * h_{t-1}, x_t] + b_h) \quad (5)$$

$$z = \text{sigmoid}(w_z[h_{t-1}, x_t] + b_z) \quad (6)$$

$$h_t = (1 - z) * h_{t-1} + z * \tilde{h}_t \quad (7)$$

其中,式(4)和式(5)表示复位门的过程,用于控制上一时刻的状态 h_{t-1} 进入GRU的量。 r 表示复位门门控向量,由当前输入 x_t 和上一时刻的状态 h_{t-1} 得到,由于激活函数使用了sigmoid函数, r 的取值范围在0和1之间,所以可以控制 h_{t-1} 的接收量并与 x_t 共同得到新输入 \tilde{h}_t 。式(6)和式(7)表示更新门的过程,用于控制新输入 \tilde{h}_t 与上一时刻状态 h_{t-1} 对当前时刻状态 h_t 的贡献量。 z 表示更新门门控向量,同样使用sigmoid函数作为激活函数,以 $1 - z$ 作为上一时刻状态 h_{t-1} 对 h_t 的贡献程度,以 z 作为新输入 \tilde{h}_t 对 h_t 的贡献程度。

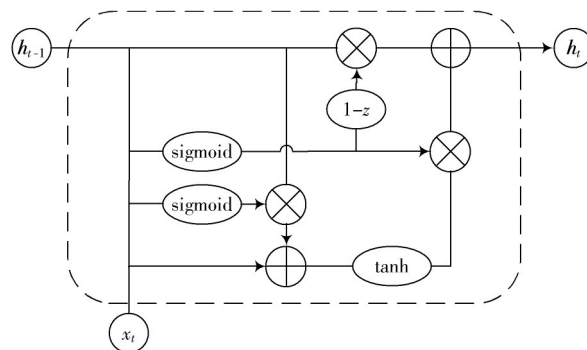


图1 GRU网络结构

1.2 注意力机制

注意力机制实现了每一时刻的输入对于当前时刻的输出的贡献差异,通过构造得分函数得到相应的权重,通过加权平均的方法重构新的输入用来预测当前输出。注意力机制多在编码-解码模型(Encoder-Decoder)中使用,对于循环神经网络而言,编码器所有时刻的隐藏状态对于某一时刻解码器的隐藏状态的贡献度应表现出差异,才能更有效地完成预测任务。这种贡献度的度量可以通过得分函数式(8)获得,具体形式多样,如式(9)、式(10)所示。得分函数需要归一化才能表现为权重,一般使用softmax函数式(11),最后将编码器的隐藏状态加权平均即可得到新的解码器输入式(12)。

$$e_t = \text{score}(s_{t-1}, h) \quad (8)$$

$$e_t = v^T \tanh(w_1 h + w_2 s_{t-1}) \quad (9)$$

$$e_t = s_{t-1}^T h \quad (10)$$

$$a_t = \text{softmax}(e_t) \quad (11)$$

$$c_t = \sum_{i=1}^T a_{ti} h_i \quad (12)$$

式中: s_{t-1} 表示解码器在 $t-1$ 时刻的隐藏状态; $h = (h_1, h_2, \dots, h_T)$ 表示编码器的各个时刻的隐藏状态; e_t 表示贡献度的得分; a_t 表示得到的权重; c_t 表示解码器的新输入。

1.3 基于注意力机制的GRU

将金融时间序列分为特征序列与目标序列,设定时间窗口大小为 T ,以特征序列为输入,建立GRU网络作为编码器,将输出的所有时刻状态与解码器的每一时间步的状态使用注意力机制重构新的解码器输入分量,并与目标序列一起作为解码器GRU网络的输入。基于注意力机制的GRU结构如图2所示。

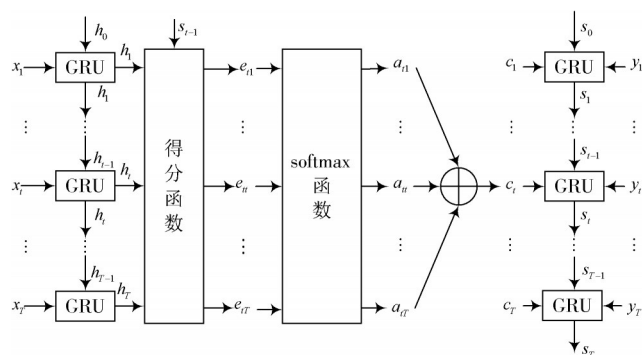


图2 基于注意力机制的GRU

2 实证分析

2.1 数据说明与预处理

本文的实证数据来自于上证指数,上证指数反映了上海证券交易所上市的所有股票价格的变动情况,其变动趋势在一定程度上能够反映中国宏观经济的走势。本文的预测目标是上证指数的收盘价,收盘价作为交易所每一个交易日的最后一笔交易价格,既是当日行情的标准,又是下一个交易日的开盘价的依据,具有重要的意义。为了更为准确地考量模型的预测效果,考虑不同维度的特征输入下模型的性能变化。因此,这里选择两个不同维度的特征输入,一个是6个输入维度的上证指数开盘价、最高价、最低价、交易量、成交额以及收盘价本身;另一个是在此基础上增加了与上证指数相关的7个综合指数(上证A指、上证B指、上证工业类指数、上证商业类指数、上证房地产指数、上证公用事业股指数、上证综合股指数)的开盘价、收盘价、最高价、最低价、交易量和成交额作为特征输入,即48个输入维度。

本文选取了2000年1月4日—2019年12月26日的

日交易数据,共4 843个样本。选择时间窗口大小为30个交易日,即以连续30个交易日的特征输入和收盘价数据预测下一个交易日的收盘价。因此共有4 813组数据作为模型的输入样本,并以最后900组数据作为模型的测试集,前3 913组数据作为模型的训练集。

为了克服序列之间不同量纲的影响,提升模型精度,并提高迭代求解的收敛速度,在训练模型之前选择以训练集的序列数据的最大值和最小值对所有的特征序列以及目标序列做归一化处理:

$$x' = \frac{x - \min(x_{\text{train}})}{\max(x_{\text{train}}) - \min(x_{\text{train}})} \quad (13)$$

式中: x 为所有原始数据; x_{train} 为训练集数据; x' 为所有归一化后的数据。

将归一化后的数据输入模型,模型变成对归一化后的收盘价的预测,因此需再将最终输出的结果进行反归一化处理,即得到最终的预测结果。

2.2 衡量指标

为了对比基于注意力机制的GRU与标准形式的GRU的预测效果,本文用绝对偏差和相对偏差两个衡量指标进行表示。绝对偏差的衡量指标选择测试集的收盘价预测值和真实值的均方误差(Mean Square Error, MSE),其计算公式为:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y'_i - y_i)^2 \quad (14)$$

相对偏差的衡量指标选择测试集的收盘价预测值和真实值的平均绝对百分比误差(Mean Absolute Percentage Error, MAPE),其计算公式为:

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y'_i - y_i}{y_i} \right| \quad (15)$$

式中: n 为样本总数; y'_i 为收盘价的预测值; y_i 为收盘价的真实值。

2.3 网络结构设定

为了尽可能减少参数,降低模型复杂度,选择式(10)作为注意力机制的得分函数。另外,基于注意力机制的GRU模型的一些参数设定如下:编码器与解码器的GRU隐藏层神经元个数均为64;时间窗口大小为30;批处理大小为128(每训练128组样本迭代更新一次参数);训练次数为150次(将所有训练集样本完整训练150次);以归一化后的收盘价的预测值和真实值的均方误差MSE作为损失函数,使用Adam优化算法更新参数。

2.4 实证结果

以上证指数收盘价为预测目标,分别将6个输入维度的特征序列和48个输入维度的特征序列放入模型训练,将测试集数据输入训练好的模型,得到实验结果。

表1和表2分别给出了6个输入维度以及48个输入维度下标准形式的GRU与基于注意力机制的GRU在测试集上的预测效果。可以看出,无论是哪一种输入维度,基于注意力机制的GRU的预测效果均优于标准形式的GRU,这说明在GRU中加入注意力机制是有利于预测性能提升的。在输入维度为6时,两种模型在测试集上的MSE与MAPE相差不大,标准形式的GRU在MAPE上仅仅比基于注意力机制的GRU多0.14%。但在输入维度为48时,由于输入维度的增加,导致模型的参数增加,模型复杂度上升,对于信息的整合能力下降。

表1 6个输入维度的模型比较

衡量指标	GRU	基于注意力机制的GRU
MSE	1 041	963
MAPE /%	0.90	0.76

表2 48个输入维度的模型比较

衡量指标	GRU	基于注意力机制的GRU
MSE	3 893	920
MAPE /%	1.61	0.73

相较于6个输入维度,此时GRU的MAPE增加了0.71%,增加了近1倍。而基于注意力机制的GRU预测性能变化不大,相反MAPE降低了0.03%。相较于标准形式的GRU,可以认为注意力机制的引入增强了GRU应对更高维输入的能力。

图3~图6也分别给出了6个输入维度以及48个输入维度下标准形式的GRU与基于注意力机制的GRU在测试集上的实际预测结果。通过与原始数据的直观对比,可以看出在6个输入维度下两种模型都有着不错的预测效果,2个模型在输入维度变为48后才出现较为明显的预测差距。

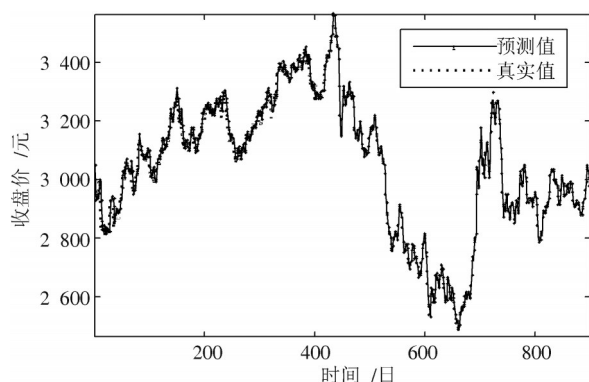


图3 6个输入维度的GRU预测结果

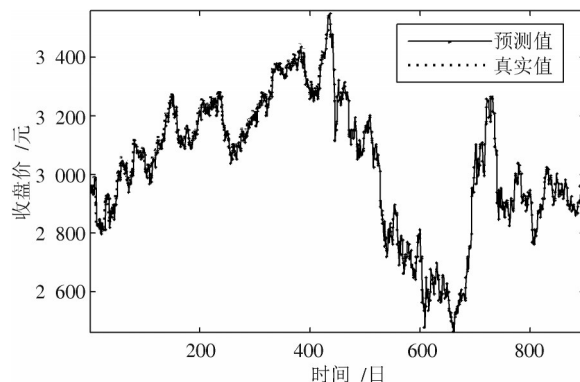


图4 6个输入维度的基于注意力机制的GRU预测结果

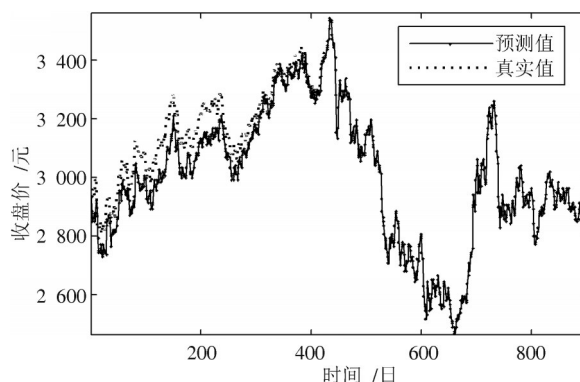


图5 48个输入维度的GRU预测结果

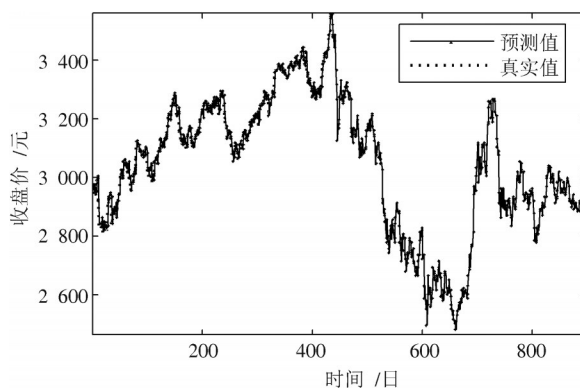


图6 48个输入维度的基于注意力机制的GRU预测结果

3 结 语

本文通过在循环神经网络中加入注意力机制,探寻其在金融时间序列方面的应用。以上证指数为例,建立基于注意力机制的GRU,并与标准形式的GRU进行对比,得出基于注意力机制的GRU可以提高预测金融时间序列的能力的结论。同时,考虑不同序列输入维度对模型的影响,发现随着输入维度的增加,标准形式的GRU预测能力下降明显,而基于注意力机制的GRU预测性能变化不大。可以认为加入注意力机制后,GRU对于信息的整合能力有所提升。因此,在循环神经网络

中加入注意力机制,对于金融时间序列的预测任务来说,不仅可以提升预测性能,还能在保证预测效果的同时应对更多维度输入的情况。

参 考 文 献

- [1] 陈亚男,薛雷.基于Bagging-SVM的股票趋势预测技术[J].电子测量技术,2019,42(14):58-62.
- [2] 邓烜堃,万良,黄娜娜.基于DAE-BP神经网络的股票预测研究[J].计算机工程与应用,2019,55(3):126-132.
- [3] 任水利,雷蕾,甘旭升,等.基于粗糙集与小波网络集成的股价走势预测研究[J].系统科学与数学,2017,37(11):2208-2221.
- [4] 王子玥.循环神经网络股票预测[J].电脑知识与技术,2018,14(22):171-172.
- [5] 尚卫平,戴昱.基于平滑ARIMA-LS-SVM组合模型的股价预测研究[J].区域金融研究,2018(5):17-23.
- [6] 陈小玲.基于ARIMA模型与神经网络模型的股价预测[J].经济数学,2017,34(4):30-34.
- [7] 宋刚,张云峰,包芳勋,等.基于粒子群优化LSTM的股票预测模型[J].北京航空航天大学学报,2019,45(12):2533-2542.
- [8] LIN M, CHEN C. Short-term prediction of stock market price based on GA optimization LSTM neurons [C]// Proceedings of the 2018 2nd International Conference on Deep Learning Technologies. [S.l.: s.n.], 2018: 66-70.
- [9] ALTHELAYA K A, EL - ALFY E S M, MOHAMMED S. Evaluation of bidirectional LSTM for short-and long-term stock market prediction [C]// 2018 9th International Conference on Information and Communication Systems (ICICS). [S. l.]: IEEE, 2018: 151-156.
- [10] LI Z, TAM V. Combining the real-time wavelet denoising and long-short-term-memory neural network for predicting stock indexes [C]// 2017 IEEE Symposium Series on Computational Intelligence (SSCI). [S.l.]: IEEE, 2017: 1-8.
- [11] BAO W, YUE J, RAO Y L. A deep learning framework for financial time series using stacked autoencoders and long-short term memory [J]. Public library of science, 2017, 12(7): 1-24.
- [12] SINGH R, SRIVASTAVA S. Stock prediction using deep learning [J]. Multimedia tools and applications, 2017, 76(18): 18569-18584.
- [13] ZHANG L, 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. [S.l.]: ACM, 2017: 2141-2149.
- [14] MA Yukun, PENG Haiyun, CAMBRIA E. Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM [C]// Thirty-second AAAI conference on artificial intelligence. New Orleans: AAAI, 2018: 22-31.
- [15] TANG G, MÜLLER M, RIOS A, et al. Why self-attention? a targeted evaluation of neural machine translation architectures [C]// Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels, Belgium: Association for Computational Linguistics, 2018: 4262-4273.

作者简介:沐年国(1971—),男,江苏人,博士,副教授,研究方向为计算机应用、金融计量统计。

姚洪刚(1996—),男,安徽人,硕士,研究方向为计算机应用、应用统计。