

文章编号: 2096-1618(2023)01-0028-09

基于信号分解降噪的 CNN-BiLSTM 金融市场趋势预测

戴宇睿, 安俊秀, 李焯炜

(成都信息工程大学软件工程学院, 四川 成都 610225)

摘要:随着金融时间序列数据日趋复杂,如何捕捉金融数据未来多天的趋势变化成了难题。针对该问题提出了基于信号分解降噪和注意力机制的 CNN-BiLSTM 金融市场趋势预测模型 (attention-based DWT-VMD-CBiLSTM)。首先利用离散小波变换 (DWT) 对原始金融股指序列进行降噪处理,然后利用变分模态分解 (VMD) 对降噪后的数据进一步分解为若干子序列。再结合多元基本面特征,利用基于注意力机制的 CBiLSTM 网络模型对各子序列进行多步预测,最后将各预测结果相加得到最终结果,实现较为长期的趋势预测。为证明所提出的模型性能,在不同金融股指数据集上与不同模型进行了实验比较。结果表明,提出的模型预测精度优于其他方法,在平均绝对误差 (MAE) 和平均百分比误差 (MAPE) 上分别达到 12.28、0.39 和 80.27、0.71,在可决系数 (R^2) 和可释方差值 (EVS) 上达到 72%、74% 和 79%、69% 的拟合度。

关键词:金融时间序列预测;离散小波变换;变分模态分解;CNN;双向 LSTM

中图分类号:TP391.1

文献标志码:A

doi:10.16836/j.cnki.jcui.2023.01.005

0 引言

近年来由于互联网技术的迅速发展,金融数据获取速度和渠道也越来越多,传统的金融行业和互联网技术通过融合,使互联网金融^[1]快速发展。因此,如何科学有效地对市场进行预测成为了高关注度、高价值的工作。

在传统的时间序列预测问题中,有很多常用的算法,如差分整合移动平均自回归 (autoregressive integrated moving average model, ARIMA)、广义自回归条件异方差 (generalized autoRegressive conditional heteroskedasticity, GARCH) 等。利用这些算法针对时序预测已经有了很多研究成果,如 Rangel-Gonzalez 等^[2]使用 ARIMA 模型对墨西哥证券交易所数据进行预测研究,取得了一定的研究成果。然而,随着机器学习的发展,传统时序预测算法在金融数据上稍显不足。于是,研究人员转而寻求传统算法和机器学习算法的结合,或纯粹的机器学习算法进行研究。像特征加权的 SVM^[3]、XGBoost 模型短期股票预测^[4]等方法的提出,都拓展和丰富了时间序列预测的研究。

随着深度学习的发展,时序预测方法也产生了改变。与传统机器学习相比,深度学习可以提取数据的深度时序特征,并取得更好的性能。其在时序预测问题上也可作为有效工具^[5],并且利用深度学习对金融市场预测的可行性也得到了研究验证^[6]。如 Deng

等^[7]使用经过 BP 网络优化的 ARIMA-LSTM 模型对医院门诊流量进行预测;庄家懿等^[8]使用 XGBoost 结合 CNN 与 LSTM 对短期电力负荷进行预测;Niu Z 等^[9]使用基于注意力机制的门控循环单元来进行风电预测;黄婕等^[10]采用 RNN 和 CNN 来对 PM2.5 浓度序列进行预测;杨青等^[11]使用 LSTM 对全球股指进行预测研究,并且和 ARIMA、SVR、MLP 等模型进行对比,证明了深度学习模型在金融时序预测上的优势。

1 相关工作

深度学习的强大特征学习和挖掘能力,赋予了金融时间序列预测更多的可能性。随着深度学习在自然语言处理中将卷积神经网络 (convolutional neural networks, CNN)、循环神经网络 (recurrent neural network, RNN) 以及注意力机制 (attention mechanism, AM) 相结合用于文本分析^[12],双向长短时记忆网络 (bidirectional long-short term memory, BiLSTM) 应用在感知上下文信息^[13],时间序列预测领域也出现了类似的研究。如林靖皓等^[14]利用基于自注意力机制的双向门控循环单元以及卷积神经网络,对芒果产量进行预测;Lu 等^[15]结合 CNN 以及 BiLSTM 和注意力机制对股票价格进行预测。

与此同时,为了化解金融时序数据的复杂波动和噪声这一难题,达到简化时序数据的目的,许多学者将模态分解等信号处理方法引入了研究。周怡娜等^[16]证明了使用变分模态分解 (variational mode decomposi-

tion, VMD) 能有效对混沌信号序列进行去噪, 并且 VMD 与经验模态分解 (empirical mode decomposition, EMD) 相比能取得更好的预测效果^[17-19]。Liu 等^[20] 和 Li 等^[21] 将 VMD 与单独的 RNN 模型结合, 对金融数据进行往后一天的短期预测, 达到不错的效果。

虽然时间序列预测有了以上优秀的研究方法, 但大部分文献都只针对极短的时间或是往后一天的预测, 力求预测结果的准确。而实际在金融市场中, 只对后一天的预测往往并不能够给出太有价值的参考, 金融市场的从业者更希望看到一个多步的区间趋势预测。同时, VMD 算法对于序列中高频高噪的部分分解效果不佳, 不能完全消除短期扰动带来的影响, 提前对高频部分做出一些预处理也许是一种不错的办法。

根据以上思路, 本文提出了一种融合离散小波变换 (discrete wavelet transform, DWT)、VMD 以及加入注意力机制的金融市场趋势预测模型 (attention-based DWT-VMD-CBiLSTM), 该模型综合了深度学习模型和信号学分解降噪的方法, 利用 DWT 在 VMD 分解前进行降噪, 去除一部分高频噪声, 降低 VMD 对高频处理不佳的影响。之后, 再进行 VMD 分解, 把金融时间序列数据的趋势特征抽取构建出来, 将短期高频波动和低频长期趋势特征分离, 让加入了 Attention 机制的 CNN-BiLSTM 网络能够更精确地捕捉时序趋势变化特点, 从而准确地预测未来一段时间的金融市场趋势变化。通过 CBiLSTM 与注意力机制的组合, 对分解后的子序列进行训练分析, 利用一维卷积捕捉时序局部特征, 利用双向长短时记忆网络来提取前后时序特征, 外加注意力机制将重要时序信息加权, 最终根据历史时间步长数据来对未来一段时间的变化趋势进行预测。之后, 将预测结果相加, 得到较为长期的金融股指趋势变化。实验证明预测结果有一定的准确性和泛化性。

2 金融市场趋势预测模型

本文提出了一种基于 DWT-VMD-CBiLSTM 和注意力机制的金融市场趋势预测模型。提出的模型的整体流程框架如图 1 所示。首先, 利用 DWT 原始股指金融序列进行分解—降噪—重构, 提取出降噪后的数据。其次, 使用 VMD 对降噪序列继续进行分解, 得到多个子序列 IMF1, IMF2, IMF3, ..., IMF12。对于这些子序列, 为了更好地把握数据的趋势特征, 将 IMF9 ~ IMF12 这几个高频低振幅的序列抛弃, 只保留前 8 个序列用于训练预测, 进一步消除高频波动对预测结果的影响。接着, 对这 8 个子序列进行训练数据样本构造以及归一化等预处理操作。最后, 为每一个分解子序列构建

多层 Conv1D 和 BiLSTM 以及注意力层的组合模型, 分别进行训练预测, 将每一个子序列的预测结果相结合即可得到最终预测结果。

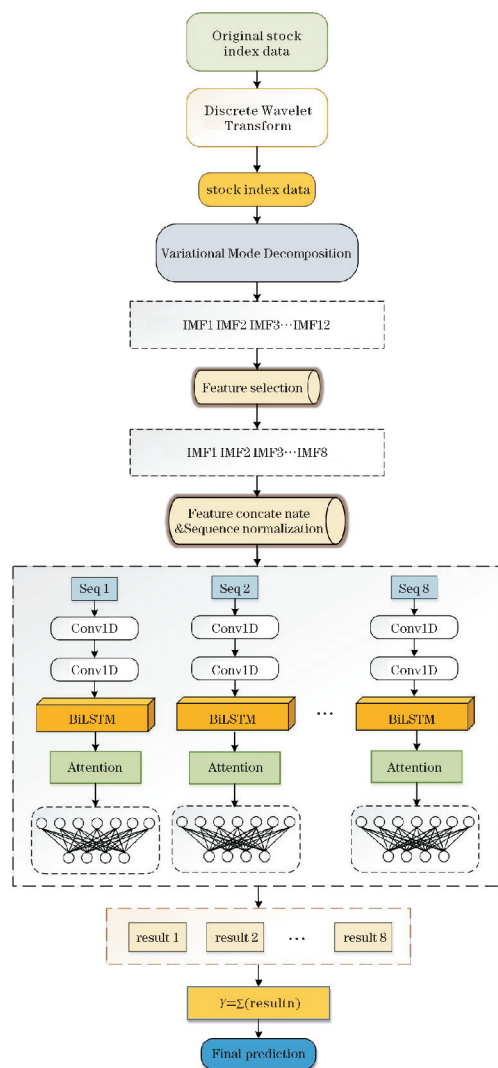


图 1 模型整体框架图

2.1 变分模态分解

在时间序列分析中利用模态分解可以将训练集各股指特征分解为多个模态, 不同模态反映了不同频度的信息特征。低频度序列反映长期趋势, 高频度序列反映短期波动。对这些经过“分类”的序列进行分析, 可以降低拟合的难度, 提高模型的准确度。本文使用的变分模态分解算法基于维纳滤波的思想, 以中心频率展开, 通过增高拉格朗日乘数法最小化带宽来达到模态分解的目的。克服了经验模态分解算法 EMD 的模态混叠和端点效应问题, 是一种自适应的信号分解方法。算法的流程如下:

(1) 假设一个时间序列 F 已通过希尔伯特变换处理为单边谱并且含有 k 个时间序列集合 $F_k(t)$, 其对应的频域信号为 $\hat{F}_k(w)$, 则其对应的带宽和频率中心的

计算公式如下:

$$w_k = \frac{1}{H_k} \int w |\hat{F}_k(w)|^2 dw$$

$$\Delta w_k = \frac{4\pi}{H_k} \int (w - w_k)^2 |\hat{F}_k(w)|^2 dw$$

$$H_k = \int |\hat{F}_k(w)|^2 dw$$

(2) 通过最小化带宽, 可以建立目标优化式:

$$\min \left\{ \sum_k \int (w - w_k)^2 |\hat{F}_k(w)|^2 dw \right\}$$

$$\sum_k \hat{F}_k(w) = F$$

(3) 通过增广的拉格朗日乘数法对上式进行优化可以得到:

$$L = \int \alpha \sum_k (w - w_k)^2 |\hat{F}_k(w)|^2 + \left[\sum_k \hat{F}_k(w) - F \right]^2 + \sigma(w) \left[\sum_k \hat{F}_k(w) - F \right] dw$$

(4) 接着计算找到对 $\hat{F}_k(w)$ 的鞍点, 通过交替乘法迭代:

$$\hat{F}_k^{n+1}(w) = \left(\hat{F}_k^n - \sum_{i \neq k} \hat{F}_i^n(w) + \frac{\sigma^n(w)}{2} \right) \frac{1}{1 + \alpha(w - w_k^n)^2}$$

以及建立中心频率的更新公式:

$$w_k^{n+1} = \frac{\int w |\hat{F}_k^{n+1}(w)|^2 dw}{H_k^{n+1}}$$

梯度上升法得到:

$$\sigma^{n+1}(w) = \sigma^n + \tau \left(\sum_k \hat{F}_k^{n+1}(w) - F_0 \right)$$

(5) 整个迭代条件的结束条件是:

$$\frac{\|\hat{F}_k^{n+1}(w) - \hat{F}_k^n(w)\|_2^2}{\|\hat{F}_k^n(w)\|_2^2} < \varepsilon$$

其中, 算法需要人为设置的参数是 k 、 α 和 τ 。 k 是需要分解的模态数量。 α 是低通滤波的带宽范围, 它能够迭代出占优势的成分, 作为中心频率来形成模态。最后通过傅里叶反变换恢复到时域就可以得到 k 个训练集。

2.2 离散小波变换

变分模态分解相当于做多次低通滤波, 对于高频部分的还原度并不是很好, 高频扰动容易夹杂在某些模态当中未分解干净。基于此, 本文寻求使用小波变换在模态分解之前先一步进行降噪。首先消除一部分高频分量, 避免分解后对实验带来干扰。小波分解具有精细的分解能力, 对高频成分比较友好, 可以很好地弥补 VMD 的劣势。通过小波基不断平移缩放来拟合原始信号, 使时间序列数据变得更加简单, 更容易拟合。采用离散小波变换^[22], 小波系数通过硬阈值进行去噪。时域中的表达式如下:

$$h(k) = \sum_n F(k) * g(2k - n)$$

$$l(k) = \sum_n F(k) * s(2k - n)$$

式中 $*$ 表示卷积, g 和 s 分别代表的是小波函数和尺度函数, 分别得出高频系数和低频系数以及小波重构的公式。

经过 DWT 和 VMD 的组合将会得到一系列具有明显趋势和规律特征子序列, 对于这些序列分别建立模型进行训练, 能得到更好的拟合效果。最终将这些结果相加, 能够得到比直接对原始高噪无规律数据预测更好的效果。

2.3 基于注意力机制的 CBiLSTM 模型

在深度学习时序预测中, 大多数研究都使用单独的 LSTM 网络进行预测, 忽略了时序局部特征以及在不同时间步上一些具有代表性重要信号的作用。同时, 由于金融时间序列往往具有一定程度的周期性, 一些特殊或典型的走势会出现在历史或未来的不同时间点重现^[23]。基于这些考虑, 本文在深度学习模型上采用了 CNN 来提取局部特征, 并且加入注意力机制, 判断关键信号点对未来时序走向的影响力大小。为了捕捉前后类似走势出现时两者之间的关联性, 引入 BiLSTM 来对前后两个方向的时间序列进行把控。

原始股指数据为 $T = [T_1, T_2, \dots, T_n]$, $T_i \in R^m$, 其中 n 为股指数据特征数, m 为样本总数。按照每一列特征进行信号分解降噪, 得到前 8 个子序列 $X^n = [x_1^n, x_2^n, \dots, x_{\text{imf}}^n]$, $x_{\text{imf}}^n \in R^m$, 其中 $\text{imf} = 1, 2, \dots, 8$, 为分解序列数, $n = 1, 2, \dots, 7$ 为得到的 7 个特征分解序列。7 个特征分别为开盘价、最高价、最低价、收盘价、每日交易量、每日成交额、涨跌幅。

在数据输入网络前, 将股指基本面特征在模态子序列上进行时序拼接。通过这一操作, 实现了对金融数据的多元特征分析, 相比大多数的单变量分析有更强的预测能力。拼接后的特征矩阵 X 为

$$X = \begin{pmatrix} x_1^1 & \cdots & x_1^n \\ \vdots & \ddots & \vdots \\ x_{\text{imf}}^1 & \cdots & x_{\text{imf}}^n \end{pmatrix} \in R^{\text{imf} \times n \times m}$$

针对每个 imf 子序列, 将每日特征、时间步长以及样本数构成三维向量 $R^{m \times s \times n}$ 输入多层 Conv1D 网络, 其中 s 为时间步长。通过一维卷积提取每个时间步长内的数据局部特征, 之后经过池化层, 将每个时间步的数据输送到 BiLSTM 网络进一步训练, 然后将数据通过注意力机制进行加权操作, 最后将结果通过全连接层处理后在输出层输出多步的预测结果。

实验采用的一维卷积 CNN 主要由 3 部分组成: 卷积层、池化层和全连接层。为捕捉不同的特征, CNN

结构中每个卷积层包含多个卷积核。对输入每一个时间步数据进行卷积运算后,提取出数据的特征,但此时特征维数非常高。为了解决这个问题,减少网络的训练成本,在卷积层之后又增加池化层,降低特征维数。卷积层的计算式为

$$y_t = \text{relu}(\mathbf{x}_t \cdot \mathbf{w}_t + b_t)$$

其中: y_t 为卷积后的输出值; relu 为激活函数; \mathbf{x}_t 为输入向量; \mathbf{w}_t 为权重; b_t 为偏置。由卷积结果得到的 y_t 接着输入 LSTM 网络。

LSTM 是一种特殊的循环神经网络,旨在解决 RNN 中存在的梯度爆炸和梯度消失问题^[24]。LSTM 网络引入门控机制来控制信息的遗忘与刷新,从而调整信息传递的路线。LSTM 单元主要由遗忘门(f_t)、输入门(i_t)和输出门(o_t) 3 部分组成,如图 2 所示。

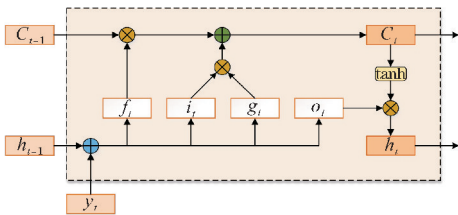


图 2 LSTM 单元结构

LSTM 中的各公式定义如下:

$$\begin{aligned} f_t &= \sigma(w_f \cdot [h_{t-1}, y_t] + b_f) \\ i_t &= \sigma(w_i \cdot [h_{t-1}, y_t] + b_i) \\ \tilde{c}_t &= \tanh(w_c \cdot [h_{t-1}, y_t] + b_c) \\ c_t &= c_{t-1} \cdot f_t + i_t \cdot \tilde{c}_t \\ o_t &= \sigma(w_o \cdot [h_{t-1}, y_t] + b_o) \\ h_t &= o_t \cdot \tanh(c_t) \end{aligned}$$

其中, f_t, i_t, o_t 为主要的 3 个门, σ 为激活函数,通常使用 Sigmoid 函数。 \tilde{c}_t 为当前时刻候选向量; c_t 为当前时间节点的状态向量; h_t 为输出给外部的状态。

在 LSTM 的基础上,BiLSTM 解决了 LSTM 只能提取历史时序数据而无法把握未来时序特征的问题,BiLSTM 的输出取决于前向时序节点和后向时序节点的双重影响。BiLSTM 的整体结构如图 3 所示。

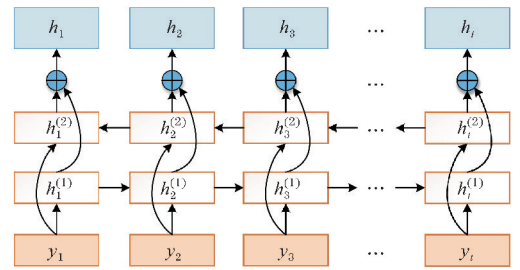


图 3 按时间展开的 BiLSTM 单元结构

BiLSTM 将 LSTM 一个放在正向的位置上,一个位于反向的位置上,使用正向的 LSTM 单元处理历史时序事件数据,反向 LSTM 单元生成未来时序事件来预测观测到的结果,当确定序列在两个方向上的趋势时,网络能得到更好的训练结果^[25]。BiLSTM 的输出值由正向和反向的两个 LSTM 单元来决定:

$$\begin{aligned} h_t^{(1)} &= \vec{h}_t = \text{LSTM}(h_{t-1}, y_t, c_{t-1}, b_t^{(1)}) \\ h_t^{(2)} &= \overleftarrow{h}_t = \text{LSTM}(h_{t+1}, y_t, c_{t+1}, b_t^{(2)}) \\ h_t &= h_t^{(1)} \oplus h_t^{(2)} \end{aligned}$$

由 BiLSTM 提取出的结果 h_t 进一步进行注意力计算,对不同隐层信息分配权重。注意力机制通过计算注意力概率分布,使用权重分配的方法选择性地关注一些比较重要的信息,减轻不重要信息的干扰,并利用权重分配这些信息的重要性。

注意力机制的计算流程如下所示。

$$\begin{aligned} u_t &= \tanh(\mathbf{W}_t \mathbf{h}_t + \mathbf{b}_t) \\ \alpha_t &= \frac{\exp(u_t^T v)}{\sum_{i=0}^L \exp(u_i^T v)} \\ s_t &= \sum_{i=0}^L \alpha_i h_i \end{aligned}$$

首先,利用注意力机制的权重矩阵 \mathbf{W}_t 和偏置矩阵 \mathbf{b}_t 来对股指特征向量 \mathbf{h}_t 进行线性变换得到隐层表示 u_t 。然后,通过 softmax 函数来计算每个输入向量和查询向量之间的相关性,也就是计算出注意力分布 α_t, v 为可学习的矩阵参数, L 为时间步长度,注意力分布 α_t 可以解释为在给定任务相关的查询 q 时,第 i 个输入向量受关注的程度。接着,将各值加权求和,得到最终的注意力值 s_t 。最后,将 s_t 送入全连接层,在输出层输出相应的预测值。整个网络的流程如图 4 所示。

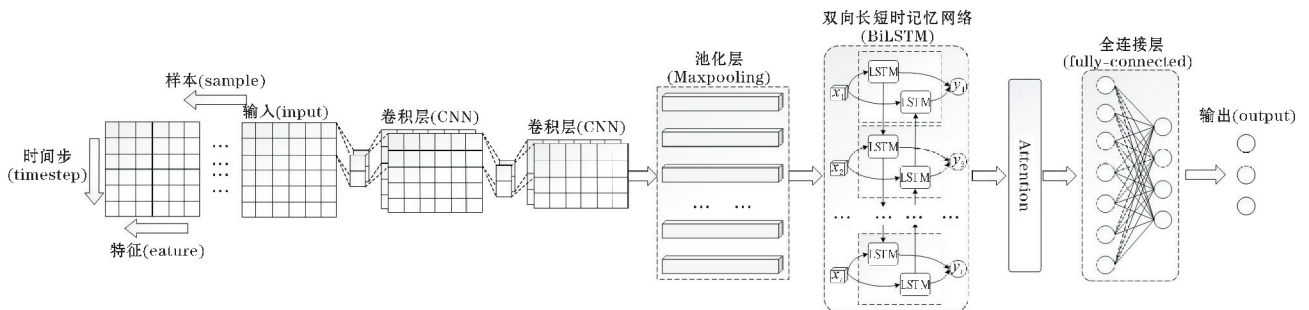


图 4 加入 attention 机制的 CBiLSTM 网络图

3 实验与分析

3.1 数据集

实验通过 BaoStock 提供的数据 API 接口,以爬虫的方式获得上证指数数据(上证综合指数)和深证成指数据(深证成份指数)。上证指数由上海证券交易所发布,以上交所挂牌上市的全部股票作为样本股,从总体上反映了上交所上市股票价格的变动情况。深证成指由深圳证券交易所发布,选择在深交所上市交易且满足入围标准的 A 股作为样本空间,选取一段时期平均总市值的比重、平均自由流通市值的比重和平均成交金额的比重综合排名在前 500 名的 A 股构成初始样本股。通过分析预测两大权威证券交易所发布的反映股市整体动向的指数数据,可以更加准确地判断市场未来的趋势变化。

上证指数数据涵盖范围为从 1996 年 1 月 1 日至 2021 年 11 月 8 日,共 6268 条数据,深证成指涵盖范围为 2001 年 1 月 1 日至 2021 年 11 月 12 日,共 5062 条数据。两个数据集内容涵盖了多个股指基本面数据。根据这些基本面数据可以构建多特征的时序样本,从而进行多因素的综合预测分析。实验模型主要根据一系列基本面特征和历史收盘价,来预测未来市场收盘价的变化趋势。

3.2 实验步骤

为了提高模型的预测效果,采用最大-最小归一化方法对分解后的子序列以及和基本面特征拼接构造的样本数据进行归一化处理。数据归一化后,数据的取值为 $[0,1]$ 。对于数据序列 $\{X_1, X_2, \dots, X_m\}$,最大-最小归一化后的序列为 $\{N_1, N_2, \dots, N_m\}$,最大-最小归一化计算公式如下:

$$N_i = \frac{X_i - \min_{1 \leq i \leq m} \{X_i\}}{\max_{1 \leq i \leq m} \{X_i\} - \min_{1 \leq i \leq m} \{X_i\}}$$

根据本文模型的训练方式,将整个数据集在进行小波变换和模态分解前按需求划分为训练集和测试集。在训练集中,实验需要将无标签的数据集划分为有标签的数据序列,以此来达到有监督的学习方式,从而进行样本采集。实验以过去历史 40 天作为一个时间步来预测未来 15 天的指数变化趋势,以此形式划分出训练数据和标签。划分时将 40 天的数据作为一个滑动窗口,以一天为步长往后滑动,每滑动一次将窗口以后的 15 天数据作为当前窗口数据的标签。过程如

图 5 所示。

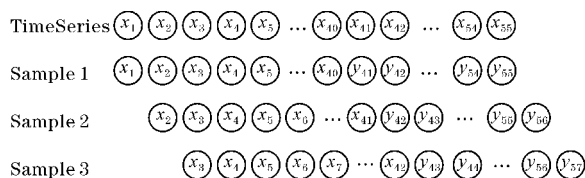


图 5 样本采集

对采集好的样本数据首先使用离散小波变换,将原始时序数据中的高频扰动去除,留下更加平滑和趋势明显的低频长期数据,这样的数据更有利于 VMD 分解。对上证指数和深证成指的股指原始数据进行小波变换后,前 100 个样本的对比如图 6 和 7 所示。

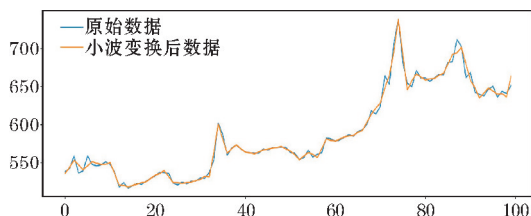


图 6 上证指数小波变换

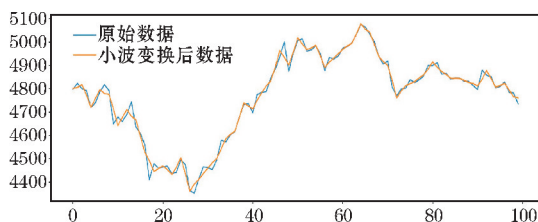


图 7 深证成指小波变换

经过去噪后的数据变化特征仍然比较复杂,将其进一步使用 VMD 进行分解。在进行 VMD 分解时需要确定惩罚因子 α 和分解模态数量 k 值等参数,而且其中起最重要作用的是模态分量数 k 值。当 k 值很小时,原始序列不能完全分解,分解误差大,预测精度低。当 k 值很大时,原始信号被过度分解。由于累积误差效应,模态过多会导致预测精度下降,增加计算量^[26]。因此,为使原序列被尽量分解的同时又消除过分解的影响,将 k 值设定为 12,分解为 12 个模态子序列,最终实际取前 8 个子序列,在获取到足够多分解信息的同时,抛弃掉过度分解的 4 个子序列,避免了结果受到杂乱高频信号波动的影响。上证指数和深证成指的 12 个 VMD 分解序列展示如图 8 和 9 所示。

从分解结果可以看出,从 VMD 分解得到的模态具有 3 个模式特征:周期性、突变性和倾向性^[27]。因此从时序分析的可行性角度来看,舍弃最后 4 个子序列,降低了突变性特征的干扰,尽最大可能地保留了周期性和倾向性,使预测性能有所提升。

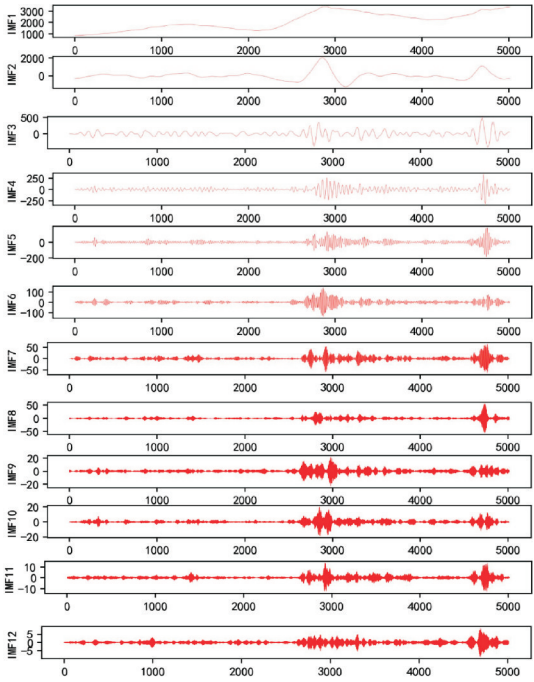


图 8 上证指数分解序列

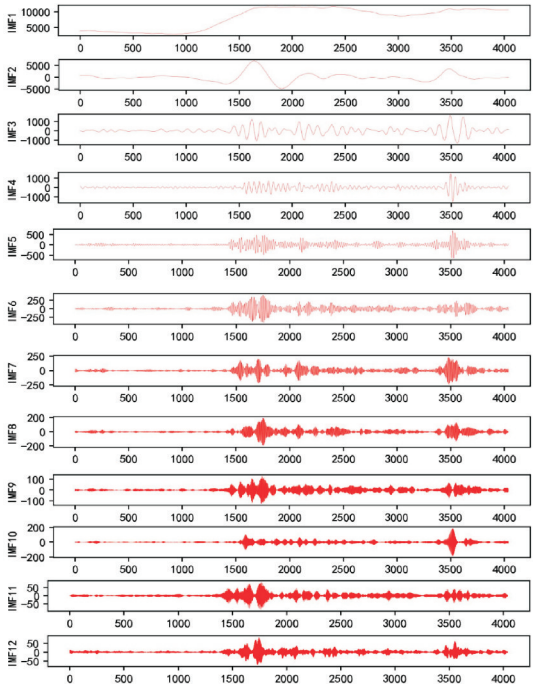


图 9 深证成指分解序列

将前 8 个序列分别和其他特征进行时序拼接后,以每批 40 个时间步,每个时间步 7 维特征数据的三维结构,输入深度学习模型进行训练预测,最终可以得到预测结果。

3.3 评价指标

实验使用 4 个指标来评价提出模型的预测效果。4 个指标分别为可释方差值 (EVS),用于解释回归模型的方差得分,取值为 $[0,1]$,越接近于 1 说明自变量

越能解释因变量;平均绝对误差 (MAE),用于反映预测值误差的实际情况;平均绝对百分比误差 (MAPE),用于反映真实值与预测值之间的误差百分比;可决系数 (R-Squared),用于反映模型的拟合程度,其值越接近 1 说明拟合程度越高。EVS、MAE、MAPE 和 R^2 的计算公式如下:

$$\begin{aligned} \text{EVS} &= 1 - \frac{\text{Var}\{y - \hat{y}\}}{\text{Var}\{y\}} \\ \text{MAE} &= \frac{1}{m} \sum_{i=1}^m |(y_i - \hat{y}_i)| \\ \text{MAPE} &= \frac{100\%}{m} \sum_{i=1}^m \left| \frac{y_i - \hat{y}_i}{y_i} \right| \\ R^2 &= 1 - \frac{\sum_{i=1}^m (y_i - \hat{y}_i)^2}{\sum_{i=1}^m (y_i - \bar{y})^2} \end{aligned}$$

其中, m 为测试集样本总量, y_i 为数据真实值, \hat{y}_i 为对测试集对应数据的预测值, Var 为计算方差值。

3.4 对比算法与实验环境

为了对比验证本文模型的效果,实验共选取和设计 3 类模型共 10 种方法作为对比实验。

(1)使用不带有注意力机制的深度学习模型。选取 BiLSTM、CNN+BiLSTM 这 2 种方法结合小波分解和 VMD 来进行实验。

(2)使用只带有 VMD 的多特征深度学习模型。选取 LSTM、BiLSTM、CNN+BiLSTM 这 3 种方法来分别进行实验,验证没有小波分解的情况下模型预测精度。

(3)使用基本的深度学习模型以及传统时序预测算法。选取 LSTM、BiLSTM、CNN + BiLSTM、ARIMA、SVM 这 5 种方法来进行实验,验证其与本文模型的预测精度差距。

本文所提出的 attention-based DWT-VMD-CBiLSTM 模型及对比算法模型都在 Python3.7 和 TensorFlow2.3.1 环境下运行。

3.5 实验结果与分析

通过在上证指数和深证成指两个金融数据集上的实验,将 3 类对比模型和本文模型共 11 种方法的预测结果以图示的方式展现。实验从上证指数训练集往后在测试集上预测 15 天的股指趋势结果如图 10 所示,每种方法的评价指标结果对比如表 1 所示。从深证成指训练集往后在测试集上预测 15 天的股指趋势结果如图 11 所示,每种方法的评价指标结果对比如表 2 所示。

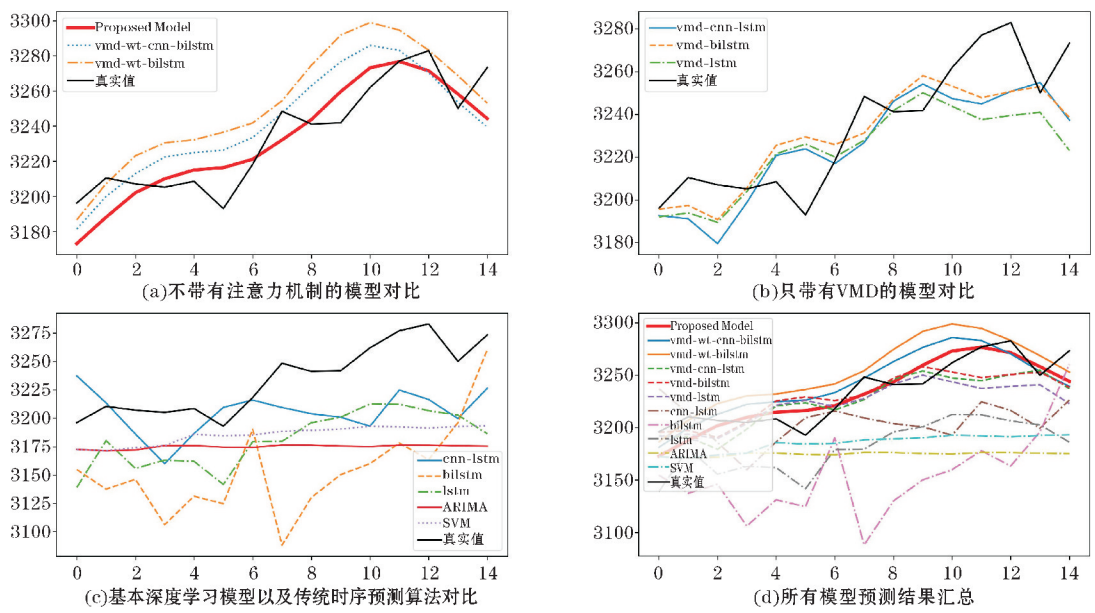


图 10 上证指数 15 天趋势预测结果

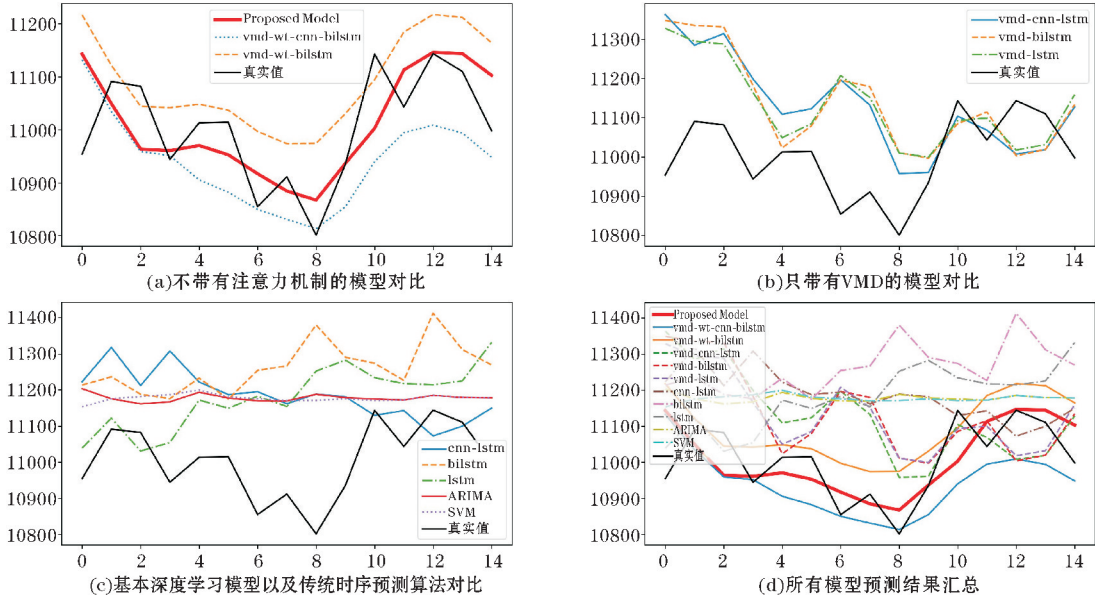


图 11 深证成指 15 天趋势预测结果

表 1 上证指数各模型预测结果指标对比

算法	MAE	R^2	EVS	MAPE
本文模型	12.82	0.72	0.74	0.39
DWT-VMD-CBiLSTM	16.93	0.56	0.57	0.52
DWT-VMD-BiLSTM	17.32	0.52	0.55	0.53
VMD-CBiLSTM	18.76	0.20	0.39	0.58
VMD-BiLSTM	17.78	0.21	0.32	0.55
VMD-LSTM	23.19	-0.26	0.15	0.72
CBiLSTM	33.97	-1.63	-1.22	1.05
BiLSTM	71.24	-3.44	-0.04	2.26
LSTM	46.89	-4.06	0.007	1.47
ARIMA	59.46	-69.03	-32.28	1.87
SVM	48.99	-49.69	-8.64	1.53

表 2 深证成指各模型预测结果指标对比

算法	MAE	R^2	EVS	MAPE
本文模型	80.27	0.79	0.69	0.71
DWT-VMD-CBiLSTM	100.81	0.64	0.76	0.91
DWT-VMD-BiLSTM	97.89	0.51	0.64	0.87
VMD-CBiLSTM	125.21	0.36	0.58	1.11
VMD-BiLSTM	132.08	0.52	0.44	1.17
VMD-LSTM	126.26	0.57	0.36	1.12
CBiLSTM	203.10	-11.03	-8.65	1.18
BiLSTM	241.56	-17.16	-11.85	2.14
LSTM	214.26	-8.05	-7.69	1.91
ARIMA	257.54	-85.24	-58.07	2.32
SVM	257.83	-98.36	-67.12	2.30

通过图 10、图 11 及表 1、表 2 可以看到在三类模型中,带有小波变换和 VMD 分解的深度学习组合模型具有最好的预测效果,而只带有 VMD 分解的深度学习组合模型效果会较差一点,没有经过小波降噪的预测曲线出现了短期的杂乱波动,对趋势的把握效果较差,但大致上的走势依然可以拟合。至于单独的深度学习模型和传统预测算法则较难预测未来的曲线走势,虽然也有着和真实数据相似的波动出现,但从图像效果对比上看远差于另外两类模型。

在上证指数的预测结果图表中可以看到,本文模型在 MAE、 R^2 、EVS 以及 MAPE 这 4 个指标上均优于其他模型,其在 R^2 和 EVS 上获得 0.72 和 0.74 的分数来看,均高于类似模型结果的 0.5 左右拟合度,以及只使用了 VMD 算法模型的 0.3 左右拟合度,充分说明了本文模型在一定程度上相比其他模型能更好地把握未来数据变化趋势。

在深证成指数数据集的预测结果中可以看到,本文模型在 MAE、 R^2 以及 MAPE 这 3 个指标上均优于其他模型,并且与上证指数实验不同的是,本文模型和其他模型相比在 MAE 与 MAPE 这两个误差指标上有着比较明显的优势,说明在某些复杂波动情况下,本文模型能够比其他模型更好地缩减误差,把握曲线走势。

最后,在单独使用深度学习模型和传统预测算法上的对比可看到预测效果不太理想。表中在 R^2 和 EVS 两个指标上出现了很多负值,说明拟合效果不佳。但从图像上来观察,也不是完全无法拟合,在上证指数中 3 个深度学习模型在误差较大的情况下,尤其是 CNN+BiLSTM 的组合依然大致地捕捉到了未来曲线走势,这也在一定程度上体现了组合模型比单一类别的网络有更大优势。

综上所述,本文模型能够更好地预测拟合未来的股指走势,能够为金融市场参与者提供一个比较好的参考意见。

4 结束语

根据股票价格数据的时序特征,提出了一种基于注意力机制的 DWT-VMD-CBiLSTM 模型来预测未来一段时间的股指序列。该方法以开盘价、最高价、最低价、收盘价、成交量、交易额、涨跌幅这些数据作为输入,充分利用了股票数据的时间序列特征。DWT 和 VMD 算法用来平稳和简单化数据序列,进而利用 CNN 提取输入数据的局部特征,利用 BiLSTM 对前后时间序列数据的特征进行学习和预测。此外,在深度学习模型上添加注意力机制可以用来捕捉不同时间的特征状态对结果的影响,提高了算法的预测精度。实验结果表明,本文的 attention-based DWT-VMD-CBiLSTM 模

型优于其他使用或没有使用 DWT 与 VMD 算法的模型,有着更好的预测精度。

本文模型在综合性的股市指数数据上有着不错的效果,未来的工作将会把目光放到金融市场消息面的新闻文本和股民评论情感倾向上,这些因素也在一定程度上反映了金融市场的波动情况。同时,在时序预测问题中应用 Seq2Seq 预测模型也是未来在多步时序预测研究中的一个不错的方向。

参考文献:

- [1] 赵洪科. 面向互联网金融微观对象的数据挖掘方法及应用研究[D]. 合肥: 中国科学技术大学, 2018.
- [2] Rangel-Gonzalez J A, Frausto-Solis J, González-Barbosa J J, et al. Comparative study of ARIMA methods for forecasting time series of the mexican stock exchange[J]. Studies in Computational Intelligence, 2018, 749: 475-485.
- [3] Chen Y, Hao Y. A Feature Weighted Support Vector Machine and K-Nearest Neighbor Algorithm for Stock Market Indices Prediction[J]. Expert Systems with Applications, 2017, 80(SEP.): 340-355.
- [4] 伯毅. 基于 XGBoost 模型的短期股票预测[D]. 哈尔滨: 哈尔滨工业大学, 2018.
- [5] Prieto A, Prieto B, Ortigosa E M, et al. Neural networks: An overview of early research, current frameworks and new challenges[J]. Neurocomputing, 2016, 214: 242-268.
- [6] Park, Frank, C, et al. Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies[J]. Expert Systems with Application, 2017, 83: 187-205.
- [7] Deng Y, Fan H, Wu S. A hybrid ARIMA-LSTM model optimized by BP in the forecast of outpatient visits[J]. Journal of Ambient Intelligence and Humanized Computing, 2020.
- [8] 庄家懿, 杨国华, 郑豪丰, 等. 基于多模型融合的 CNN-LSTM-XGBoost 短期电力负荷预测方法[J]. 中国电力, 2021, 54(5): 46-55.
- [9] Niu Z, Yu Z, Tang W, et al. Wind power forecasting using attention-based gated recurrent unit network[J]. Energy, 2020, 196: 117081.
- [10] 黄婕, 张丰, 杜震洪, 等. 基于 RNN-CNN 集成深度学习模型的 PM_{2.5} 小时浓度预测[J]. 浙江大学学报(理学版), 2019, 46(3): 370-379.
- [11] 杨青, 王晨蔚. 基于深度学习 LSTM 神经网络的全球股票指数预测研究[J]. 统计研究, 2019, 36(3): 65-77.
- [12] 程艳, 尧磊波, 张光河, 等. 基于注意力机制的多通道 CNN 和 BiGRU 的文本情感倾向性分析[J]. 计算机研究与发展, 2020, 57(12): 2583-2595.

- [13] 李卫疆,漆芳,余正涛. 基于多通道特征和自注意力的情感分类方法[J]. 软件学报,2021,32(9):2783-2800.
- [14] 林靖皓,秦亮曦,苏永秀,等. 基于自注意力机制的双向门控循环单元和卷积神经网络的芒果产量预测[J]. 计算机应用,2020,40(S1):51-55.
- [15] Lu W, Li J, Wang J, et al. A CNN-BiLSTM-AM method for stock price prediction[J]. Neural Computing and Applications,2020,33(10):1-13.
- [16] 周怡娜,路敬祎,张勇,等. VMD-SG-WT 去噪法及其在混沌去噪中的应用[J]. 东北石油大学学报,2020,44(4):113-120.
- [17] Xue Y J, Cao J X, Wang D X, et al. Application of the Variational-Mode Decomposition for Seismic Time-frequency Analysis[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,2016,9(8):1-11.
- [18] Zhang Y, Chen B, Pan G, et al. A novel hybrid model based on VMD-WT and PCA-BP-RBF neural network for short-term wind speed forecasting [J]. Energy Conversion and Management,2019,195(SEP.):180-197.
- [19] Niu H, Xu K, Wang W. A hybrid stock price index forecasting model based on variational mode decomposition and LSTM network [J]. Applied Intelligence,2020,50.
- [20] Liu Y, Yang C, Huang K, et al. Non-ferrous metals price forecasting based on variational mode decomposition and LSTM network-ScienceDirect [J]. Knowledge-Based Systems,2020,188:105006.
- [21] Li Y, Wang S, Wei Y, et al. A New Hybrid VMD-ICSS-BiGRU Approach for Gold Futures Price Forecasting and Algorithmic Trading [J]. IEEE Transactions on Computational Social Systems, 2021(99):1-12.
- [22] 黄擎,曾向阳. 小波分解和改进 CNN 相融合的水声目标识别方法[J/OL]. 哈尔滨工程大学学报,2022(2):1-7[2021-11-23]. <http://kns.cnki.net/kcms/detail/23.1390.U.20211026.1112.002.html>.
- [23] 李晓寒,贾华丁,程雪,等. 基于改进遗传算法和图神经网络的股市波动预测方法[J/OL]. 计算机应用:1-13[2021-11-22]. <http://kns.cnki.net/kcms/detail/51.1307.TP.20210816.1658.012.html>.
- [24] 王艺霏,于雷,滕飞,等. 基于长-短时序特征融合的资源负载预测模型[J/OL]. 计算机应用:1-9[2021-11-21]. <http://kns.cnki.net/kcms/detail/51.1307.TP.20210705.2001.017.html>.
- [25] Wang F K, Mamo T, Cheng X B. Bi-directional long short-term memory recurrent neural network with attention for stack voltage degradation from proton exchange membrane fuel cells[J]. Journal of Power Sources,2020,461:228170.
- [26] Shi G, Qin C, Tao J, et al. A VMD-EWT-LSTM-based multi-step prediction approach for shield tunneling machine cutterhead torque[J]. Knowledge-Based Systems,2021,228.
- [27] Bahdanau D, Cho K, Bengio Y. Neural Machine Translation by Jointly Learning to Align and Translate[J]. Computer Science,2014.

CNN-BiLSTM Financial Market Trend Prediction based on Signal Decomposition and Noise Reduction

DAI Yurui, AN Junxiu, LI Zhuowei

(College of Software Engineering, Chengdu University of Information Technology, Chengdu 610225, China)

Abstract: With the increasing complexity of financial time series data, it becomes a challenge to capture the trend changes of financial data for multiple days in the future. To address this problem, a financial market trend prediction model (attention-based DWT-VMD-CBiLSTM) based on CNN+BiLSTM with signal decomposition noise reduction and added attention mechanism is proposed. Firstly, the original financial stock index series are noise-reduced by using discrete wavelet transform (DWT), and then the noise-reduced data are further decomposed into several sub-series by using variational modal decomposition (VMD). Based on the first two steps combined with multivariate features, a CNN+BiLSTM network model based on the attention mechanism is used to make multi-step predictions for each sub-series, and finally the prediction results are summed to obtain the final results to achieve longer-term trend prediction. To demonstrate the performance of the proposed model, experimental comparisons are conducted with different models on different financial stock index datasets. The results show that the prediction accuracy of the proposed model outperforms other methods, reaching 12.28, 0.39 and 80.27, 0.71 in mean absolute error (MAE) and mean percentage error (MAPE), respectively, and 72%, 74% and 79%, 69% in the coefficient of resolvability (R^2) and interpretable variance value (EVS) of the fit.

Keywords: financial time series prediction; discrete wavelet transform (DWT); variational modal decomposition (VMD); convolutional neural network; bidirectional long-short memory network