

引入小波分解的神经网络预测

——以玉米大宗商品价格序列为例

霍永良

(广州华商学院 数据科学学院, 广州 511300)

摘要:采用世界银行发布的连续 752 期月度玉米国际价格,将其视为离散价格时间序列,运用小波理论中的 Mallat 算法,把价格序列分解为若干高频分量和一个低频分量,然后将各分量导进循环神经网络,再累加所得的各个分量预测值,作为最终预测价格。实验表明,引入小波分解的神经网络模型,在玉米价格时间序列中,比较灵活地捕捉到高频和低频信号,并准确拟合和预测这些部分的数值,结果表明该方法对于价格频繁剧烈波动的场景具有实用意义。

关键词:小波分解;神经网络;时间序列

中图分类号:TP183 **文献标志码:**A **文章编号:**1671-1807(2023)14-0259-06

玉米、水稻、小麦和大豆并称为中国四大粮食作物,其中玉米产量的大约六成用作饲料谷物,其余的用作食品用麦片、甜味剂、工业用酒精等。随着经济水平发展和生活品质提升,全球市场对肉、蛋、奶的需求都在增加,致使玉米的市场用途越来越广泛。在国际贸易中,中国是其中一个主要玉米进口国,而出口国多分布在美洲和欧洲;在当前复杂动荡多变的国际局势之下,影响价格走势的因素进一步增多凸显,作为大宗农产品重要品种之一的玉米,与其他大宗商品共同构成相互关联的多边网络。加强研究大宗农产品的价格预测,直接关系到国家粮食安全战略。要增强中国农产品进出口的主动权,加大贸易谈判的筹码,其中必不可少的一个环节是追踪、预测包括玉米在内的大宗农产品的国际价格。

世界银行每月发布《大宗商品市场展望》报告(*World Bank Commodity Markets Outlook*),给出全球主要大宗商品的月度价格;现采集玉米从 1960 年 1 月到 2022 年 8 月共 752 期数据,作为数据集,时间跨度为 62 年。

1 理论方法

传统的时间序列分析,通常把数据信号分解为季节性、周期性、随机性和趋势 4 个组成部分,然后使用回归方法进行预测。表面上这种思路考虑到

数据对象本身对现实世界的映射,但有两种情况值得注意:一是并非所有应用场景都在上述 4 种维度上有明显表征;②现实世界的复杂性决定了这四种维度无法完整囊括所有场景特性。

从数理工具的角度看,需要存在工具去分解呈现时间序列的数值信号,这里说的“分解”显然不仅仅针对上面提及的 4 个传统角度,该工具应具备一定的自适应能力,探求包含且不仅包含这 4 个方面特征的能力。正值新一轮人工智能发展高潮的当下,作为分类器和函数逼近工具的神经网络,在多个实践领域都表现出强大的“类人脑”能力;带有强烈预测需求的时间序列场景也逐渐引入神经网络的运用,产生了大量有别于传统回归方法的新思路。

从“可解释性人工智能”的角度看,时间序列信号的传统处理方式偏重于可解释、可理解的方向,一定程度上牺牲了部分难以直白解释的信号特征;神经网络方案则倾向于内核黑盒子的自适应、自处理手法,由此不可避免地存在随机性和盲目性,从而难以在应用场合针对具体事物作出事理性的解释。

在上述两者之间事实上需要建立合理的数理弥合机制,既充分发挥两方面优势,又能提供应对各自缺陷的方案。

自一百多年前 Alfred Haar 发表关于正交函数

收稿日期:2023-03-27

基金项目:广州华商学院 2022 年度青年学术科研项目(2022HSXS080)

作者简介:霍永良(1978—),男,广东佛山人,广州华商学院数据科学学院,助教,硕士,研究方向为社会智能计算。

系统(orthogonal function system)以来,多个领域已经见证了小波理论的强大力量,尽管其系统工程理论和应用在20世纪80年代初才正式拉开帷幕^[1-3]。小波理论同时具有理论深刻和应用广泛的双重意义,无论对数学还是对工程应用都具有重要价值。过去近40年已涌现出大量文献和实践,把小波理论运用到图像视频处理、生物医学工程、数字信号处理、工程故障诊断等多个领域;该理论以其独特的自适应、多尺度解释能力,成为不容忽视的强大数学工具^[4]。

1.1 小波变换

在许多应用领域中,工程上需要使用傅立叶变换提取给定信号 $f(t)$ 在特定局部时间的频率(ω)成分,其突出缺陷在于时域上没有分辨率,即无法从中获悉某种频率成分发生在哪些时间内,因此仅对平稳信号有效^[5]。

$$f(\omega) = \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt \quad (1)$$

式中: $f(x)$ 为信号函数。

实践中的信号几乎都是非平稳信号,频域特征随时间变化,短时傅里叶变换由此应运而生:若要定位频率则需保留频率 ω 变量,若要定位时间则需保留时间 t 变量,短时傅里叶变换定义如下。

$$G_f(\omega, \tau) = \int_{-\infty}^{\infty} f(t) g^*(t - \tau) e^{-i\omega t} dt \quad (2)$$

式中: $*$ 为复共轭; $g(t)$ 为以时间 τ 为中心的滑动窗口函数;基函数 $e^{-i\omega t}$ 则起频限作用; $G_f(\omega, \tau)$ 为信号 $f(t)$ 在时刻 τ 附近、频率为 ω 的信号成分。一旦确定窗口函数 $g(t)$,时频窗口的时宽和频宽不变,意味着分辨率也固定下来。由于频率跟周期成反比,高频信号需要窄的时间窗口,低频信号需要宽的时间窗口,短时傅里叶变换的瓶颈由此产生。20世纪80年代初法国工程师 Morlet 创造性地提出了小波理论,取得突破性成果^[6]。

首先,小波是函数空间 $L^2(R)$ 中满足下面条件的一个函数或信号 $\varphi(t)$,小波函数的可容许条件的公式为

$$C_\varphi = \int \frac{|\hat{\varphi}(\omega)|^2}{|\omega|} d\omega < \infty \quad (3)$$

式中: $\varphi(t)$ 为基小波或母小波函数; $\hat{\varphi}(t)$ 为相应的傅里叶变换。

把函数 $\varphi(t)$ 进行伸缩和平移,可得函数

$$\varphi_{a,b}(t) = \frac{1}{\sqrt{a}} \varphi\left(\frac{t-b}{a}\right) \quad (4)$$

式中: $a, b \in R$, 且 $a > 0$; a 为伸缩因子(尺度参数); b

为平移因子(位移参数)。相应的连续小波变换定义为

$$WT_f(a, b) = \frac{1}{\sqrt{a}} \int f(t) \varphi^*\left(\frac{t-b}{a}\right) dt \quad (5)$$

式中:WT 为小波变换(wavelet transform);连续信号 $f(t)$ 的参数 a 和 b 可连续取值。

式(5)可表示为通用的内积形式 $WT_f(a, b) = \langle f(t), \varphi_{a,b}(t) \rangle$ 。实际应用中的信号往往是离散的且无法连续监测,因此需引入离散小波变换。通常取 $a = a_0^j, b = k b_0 a$; $j = 0, 1, 2, \dots; k \in Z; a_0 > 0$ 且 $a_0 \neq 1, b_0 \neq 0$ 。那么母小波函数 $\varphi_{a,b}(t)$ 变换为

$$\varphi_{j,k}(t) = a_0^{-j/2} \varphi(t - k a_0^j b_0) = a_0^{-j/2} \varphi(a_0^{-j} t - k b_0) \quad (6)$$

式中: $f(t)$ 为信号; $\varphi(t)$ 为母小波函数; $a_0^{-j/2}$ 为不同的信号频率成分;参数 j, k 为尺度 a 和位移 b 作离散化处理时的辅助参数。

相应地,离散小波变换定义为

$$\begin{aligned} WT_t(j, k) &= \int f(t) \varphi_{j,k}^*(t) dt = \frac{1}{\sqrt{a}} \int f(t) \varphi^*\left(\frac{t-b}{a}\right) dt \\ &= a_0^{-j/2} \int f(t) \varphi^*(a_0^{-j} t - k b_0) dt \end{aligned} \quad (7)$$

式中: $f(t)$ 为信号函数; $\varphi(t)$ 为母小波函数。从上述演算可知,通过调整参数 j ,得到信号 $f(t)$ 不同的频率成分 a_0^{-j} ,相应就可确定时域上的取样步长为 $b_0 a_0^j$,如此定义的取样步长自然也是可调整的。高频局部信号对应较小的 j ,取样步长就相对小;低频局部信号对应较大的 j ,取样步长就相对大,即离散小波变换实现了采样窗口大小固定、形状伸缩平移可变的时频局部化功能。

1.2 小波分解

从理论上讲,小波定义在函数空间 $L^2(R)$ 中,该空间由平方可积的函数组成,意味着小波分析中的信号都是平方可积,那么其信号能量必须是有限的。正如前文所述,小波变换通过尺度参数和位移参数,在一个时间序列的整个频谱上分析出高频的细节分量和低频的逼近分量;也就是说信号函数 $f(t)$ 被表示为一个逐级逼近的极限,整个过程包含一系列逐级逼近子空间 $\{V_j, j \in Z\}$ 的子过程,每一个逼近过程都具有不同的分辨率和尺度。根据 Stephane Mallat 和 Yves Meyer 在20世纪80年代末提出的多分辨分析理论^[7],存在一个函数 $\varphi(t) \in V_0$,其对应闭子空间 $\{\varphi(t-n)\}_{n \in Z}$ 形成零尺度空间 V_0 的正交基,该函数称为尺度函数。一系列具有不同位移参数的同一尺度函数 $\varphi(t)$ 组成了闭子空间

$\{V_j, j \in Z\}$ 的集合。小波变换的多种离散形式中, 使用比较普遍的一种形式是取式(7)中的 $a_0 = 2$ 以及 $b_0 = 1$ (称为二元离散化)^[8]。那么小波函数 $\varphi(t)$ 和尺度函数 $\varphi(t)$ 之间的内在关系可表示为如下的双尺度方程^[9]。

尺度方程定义为

$$\varphi(t) = \sqrt{2} \sum_{n=-\infty}^{+\infty} h(n) \varphi(2t - n), n \in Z \quad (8)$$

小波方程定义为

$$\varphi(t) = \sqrt{2} \sum_{n=-\infty}^{+\infty} g(n) \varphi(2t - n), n \in Z \quad (9)$$

式中: $\varphi(t)$ 为尺度函数; n 为形成零尺度空间 V_0 的正交基的参数; 展开系数 $h(n)$ 为低通滤波系数, 内积形式的计算公式为

$$h(n) = \langle \varphi(t), \varphi_{-1,n}(t) \rangle \quad (10)$$

展开系数 $g(n)$ 为高通滤波系数, 内积形式的计算公式为

$$g(n) = \langle \varphi(t), \varphi_{-1,n}(t) \rangle \quad (11)$$

尺度方程 $\varphi(t)$ 经过伸缩和平移后, 可得

$$\begin{aligned} \varphi(2^{-j}t - k) &= \sqrt{2} \sum_{n=-\infty}^{+\infty} h(n) \varphi(2^{-j+1}t - 2k - n), \\ n &\in Z, j \in Z, k \in Z \end{aligned} \quad (12)$$

令 $m = 2k + n$, 由式(12)可得

$$\begin{aligned} \varphi(2^{-j}t - k) &= \sqrt{2} \sum_{m=-\infty}^{+\infty} h(m - 2k) \varphi(2^{-j+1}t - m), \\ m &\in Z, j \in Z, k \in Z \end{aligned} \quad (13)$$

基于 Mallat 算法, 通过 $h(n)$ 和 $g(n)$ 这样一对滤波器组, 可以把信号分解为低频的逼近分量和高频的细节分量, 实现诸如特征提取、噪声消除等的信号预处理功能, 具体为

$$\begin{cases} a_{j,k} = \sum_{m=-\infty}^{+\infty} h(m - 2k) a_{j-1,m} \\ d_{j,k} = \sum_{m=-\infty}^{+\infty} g(m - 2k) a_{j-1,m} \end{cases}, \quad m \in Z, j \in Z, k \in Z \quad (14)$$

式中: h 为低通滤波系数; g 为高通滤波系数; $a_{j,k}$ 代表信号的低频分量, 称为逼近系数, 通过把前一个尺度的逼近系数, 跟低通滤波器系数进行卷积运算获得; $d_{j,k}$ 代表信号的高频分量, 称为细节系数, 通过把前一个尺度的逼近系数跟高通滤波器系数进行卷积运算获得。图 1 展现了基于 Mallat 算法对原始信号进行两层小波分解的过程。

离散小波变换的一个重要应用是信号去噪, 其核心目标包括抑制噪声和信号重构两个方面, 一般来说包括三个步骤: 首先挑选一个母小波函数和分

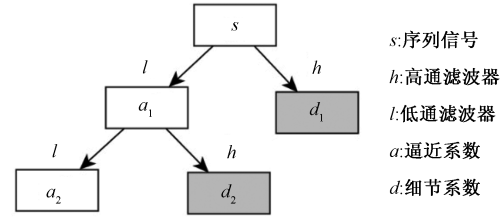


图 1 基于 Mallat 算法的小波分解

解的层数 j , 进行离散小波变换到 j 层; 然后对每一个分解层级分别选择一个阈值, 应用到对应的细节系数; 最后使用经过阈值处理过的每一层细节系数, 以及第 j 层原始的逼近系数重构去噪后的信号。例如如图 1 中, 信号重构过程表达为信号 $s = d_1 + a_1 = d_1 + d_2 + a_2$ 。

1.3 小波神经网络

神经网络是当前人工智能热潮中的重要推手, 无论作为函数逼近器还是分类器, 都表现出良好性能。从经典的 M-P 模型 (McCulloch-Pitts model)、多层感知机模型 (multi-layer perceptron, MLP)、BP 模型 (error back propagation model), 到卷积神经网络 (convolutional neural network, CNN), 及其后数不胜数的衍生优化模型, 都彰显着神经网络强大的挖掘潜力和广阔的应用前景。

已有文献介绍小波分析理论跟神经网络模型相结合, 进行时间序列分析^[10-11]。把小波分解引入神经网络、用于预测时间序列信号, 主要基于以下几点考虑。

1) 现实世界中的时间序列数据往往带有随机噪声, 例如大宗农产品的价格波动更是如此, 叠加政治、经济、科技、自然乃至军事等层面的大量周期性和突发性影响因素。然而传统的平均回归方法要求数据序列是稳定的, 且本质上只能捕捉线性关系。

2) 离散小波变换的分解和重构适合对时间序列信号做特征提取和噪声去除, 尤其是 Mallat 算法强于自适应获取序列信号的高频分量, 十分适合处理非线性价格波动的场景。

3) 小波变换的过程可视为小波基函数在时域内与原始信号做内积运算提取小波域特征, 神经网络则使用隐藏层权重和卷积核对原始输入进行内积运算, 因此两者具备融合、相互优化的基础。

4) 神经网络的结构决定了在每一神经层中进行并行计算, 其运算效率优于平均回归方法, 响应实际场景中缩短计算时耗的期望。

经典的全连接神经网络一般设置输入层、隐藏

层和输出层的三层结构,各层之间的节点没有直接连接^[12];对于自然语言文本、价格时间序列等数据,其前后样本带有确定性的先后顺序关联性质并没有体现在模型训练过程中,因此不适宜直接使用传统的神经网络模型。循环神经网络(recurrent neural network, RNN)则是一种专门处理序列数据的神经网络^[13],基本思路是循环神经网络隐藏层的值 $h(t)$ 综合取决于本轮输入 x 和上一轮隐藏层的值 $h(t-1)$,数学表达为

$$Z_t = g(DH_t) \quad (15)$$

式中: $H_t = f(WX_t + VH_{t-1})$; Z_t 为输出层的值向量; D 为从隐藏层到输出层的权重矩阵; H_t 为 t 时刻隐藏层的值向量; W 为从输入层到隐藏层的权重矩阵; V 为隐藏层上一轮的值作为本轮输入的权重矩阵; g 和 f 为常见激活函数。具体如图2所示。

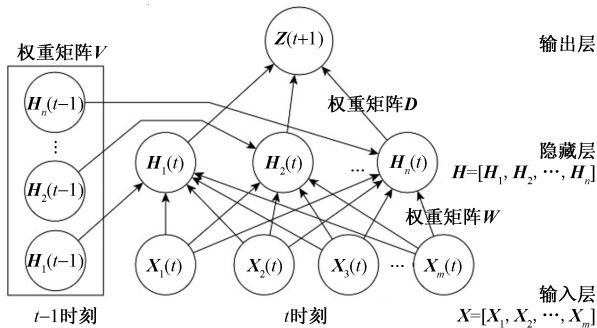


图2 循环神经网络

1.4 模型构造

基于上述离散小波变换和循环神经网络的理论,结合含有非周期性高频分量的价格序列的预测需求,构造以下小波神经网络。

1) 现有时刻 t 的离散序列信号 x_t , 内含 m 个离散值。使用 Mallat 算法对其进行小波分解, 生成细节系数序列和逼近系数组成的复合序列 $\{d_1^t, d_2^t, \dots, d_i^t, a_i^t\}$, 作为神经网络的输入。

2) 序列 x_t 分解为 $(j+1)$ 个分量序列, 每一个分量序列各自应用到如图3结构的神经网络。

图3神经网络的结构详情如下。

输入层含 m 个节点, 令 $i=1, 2, \dots, m$ 。

隐藏层含 n 个节点, 令 $j=1, 2, \dots, n$, 隐藏层节点为

$$h_j = a_j^{-1} \left(\sum_{i=1}^m w_{ij} x_i - b_j \right) + \sum_{j=1}^n v_j u_j$$

式中: 在第 j 个隐藏层节点; a_j 为尺度参数; b_j 为位移参数; v_j 为 RNN 循环层相对应节点的权重; u_j 为 RNN 循环层相对应节点的值, 即上一轮隐藏层相应

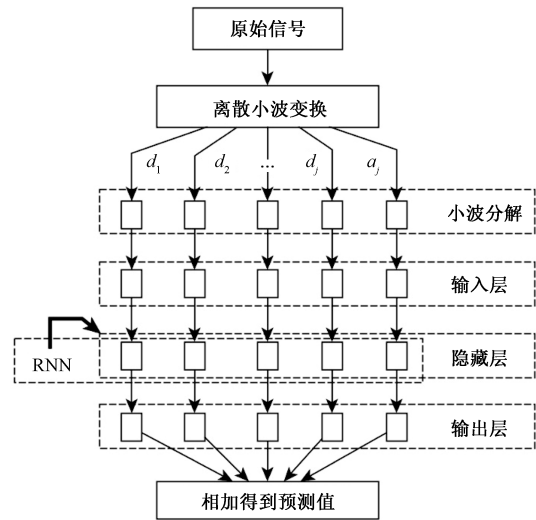


图3 引入小波分解的神经网络

节点的值; 隐藏层的输出值 $z_j = \varphi_{a,b}(h_j)$, 这里取 Morlet 母小波函数作为激活函数, 其定义为 $\varphi(t) = \cos(1.75t) \exp\left(-\frac{t^2}{2}\right)$ 。

输出层的值为

$$y_{t+1} = \sum_{j=1}^n d_j z_j$$

式中: d_j 为隐藏层到输出层的权重。

上述神经网络的各层权重的初始值均随机生成, 使用随机梯度下降的方法更新权重。

3) 把上述 $j+1$ 个分量序列的神经网络输出值累加, 即为时刻 $t+1$ 的最终预测值。

2 研究实验

2.1 实验过程

采用玉米大宗商品价格连续 752 期月度数据作为数据集, 视为离散的时间序列信号。基于 Mallat 算法, 对该序列进行 5 层小波分解, 由此获取 5 组含有渐次升高的频率分量, 以及一组含有最低的频率分量。然后把 6 组信号分量序列进行归一化处理之后, 分别作为输入导入各自独立的循环神经网络中, 激活函数采用 Morlet 母小波函数。经实验验证, 循环神经网络在以下参数设置下取得较好拟合效果: 隐藏层设 10 个节点, 学习率取 0.005, 迭代次数为 20。最后把上述 6 个神经网络得出的分量预测值直接线性累加, 作为最终预测值。

2.2 实验结果

图4展示了小波分解的最后两个分量序列(d_5 , a_5)进行归一化处理之后, 经过神经网络处理得到的拟合值; 从中可见小波分解对序列信号局部时频实

现了不同分辨率的分析:使用大尺度宽时窗口识别信号的轮廓,使用小尺度短时窗口识别信号的细节,达到“既见森林、又见独木”的效果。

图5呈现了同期实际值和最终拟合值的比较,并且给出相应的误差比例(等于最终拟合值跟实际值的差值,取绝对值后除以实际值,再转换为百分比);可见拟合情况良好,玉米价格大幅波动时期的

误差比例大部分在0%~4%,普遍小于价格平稳的时期,如图5中所示的四个窗口期;相对地,玉米价格相对平稳的时期,误差比例普遍在2%~7%。这说明引入小波分解的神经网络预测模型体现出小波分析擅长捕捉高频分量的突出优点,该特性对于证券、期货、贵金属、大宗商品交易等价格频繁剧烈波动的场景具有重要意义。

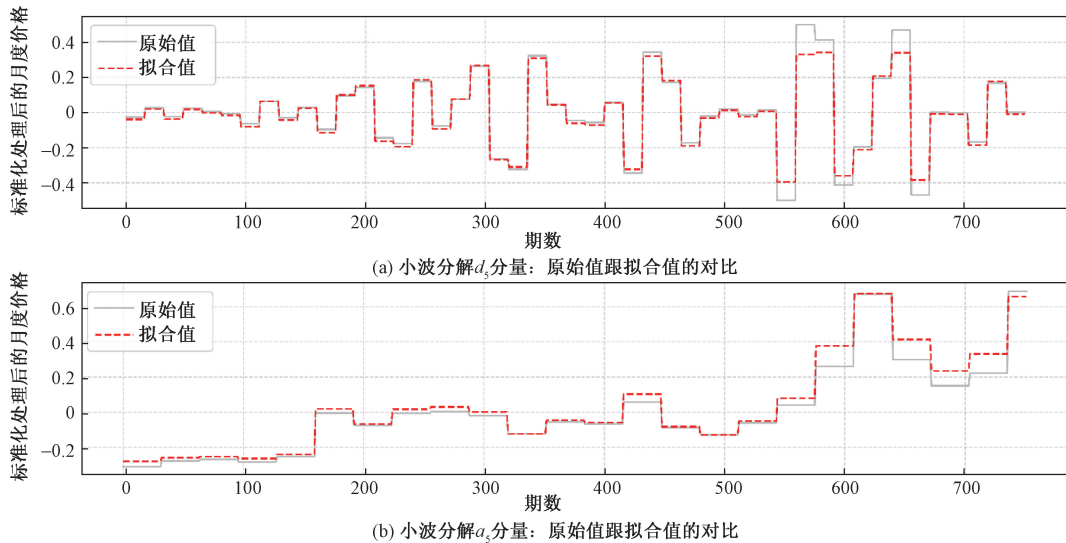


图4 小波分解得到 d_5 、 a_5 两个分量序列的数据拟合情况

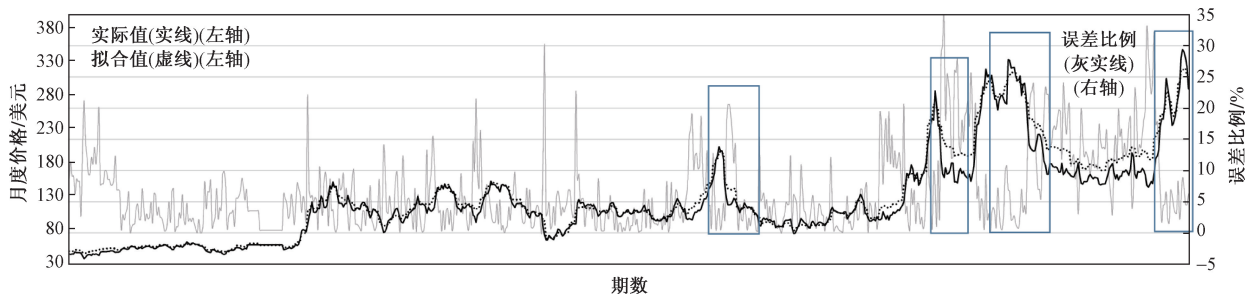


图5 实际值和最终拟合值的比较

时间序列预测模型的常见评估指标有均方根误差(root mean square error, RMSE)、平均绝对误差(mean absolute error, MAE)、平均绝对百分比误差(mean absolute percentage error, MAPE)、 R^2 (模型所解释的方差所占的比例)。在本实验中的值分别是:RMSE=13.52, MAE=8.75, MAPE=0.07, $R^2=0.95$ 。

应该指出,部分样本信号拟合效能有待改善。例如如图6中,实际值在横轴方向,拟合值在纵轴方向,由此生成相应的散点图。容易看出,全部散点都趋向于图中对角线,说明整体拟合效果良好;相对地,实际值在150~250的散点比较偏离对角线,偏离幅度大于前后两个范围的样本信号,而且这些样本大部分在对角线上方,说明这个范围内的拟合值误差较大,基本大于实际值。这种情况表明引入

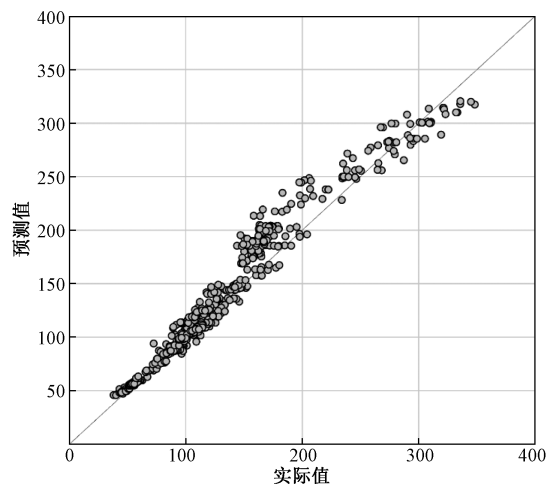


图6 实际值跟拟合值的比较

小波分解的神经网络,应用于包含多段高频分量的序列信号,对于低频和高频信号的拟合效果较好,高低频之间分量的拟合程度需要后续研究算法的优化改进。

3 小结

基于玉米对中国粮食安全的重要作用,选取玉米的国际大宗商品价格作为研究对象,共计连续752期的月度价格时间序列。研究方法方面,简要回顾小波分析和循环神经网络的基本理论和特点后,提出引入小波分解构造小波神经网络预测玉米价格时间序列。该方法的优点主要是融合小波分解优异的多尺度解释能力和神经网络良好的非线性逼近能力。拟合结果表明,引入小波分解的神经网络适用于价格时间序列的预测场景,神经网络的拟合值较好地接近实际值。

后续可进一步开展的研究如下:文中提及高频跟低频之间的数据拟合未如高频、低频分量的预测效果,需要进一步探索合适的分解层数、合适的母小波函数。另外,由于梯度下降算法容易陷入局部最优,可考虑把经典时间序列分析中的自回归特性,融入小波分析的信号分解和重构功能,并增加动量因子进入小波神经网络建模过程,获取更合适的学习速率,以此提高神经网络的学习效率。

参考文献

[1] FIRDOUS A S, AZHAR Y T. Wavelettransforms-kith

and kin[M]. Boca Raton: CRC Press, 2023.

[2] BULNES F. Recent advances in wavelet transforms and their applications[M]. London: Intech Open, 2022.

[3] BHATNAGAR N. Introduction to wavelet transforms [M]. Boca Raton: CRC Press, 2023.

[4] GAO R X, YAN R Q. Wavelets: theory and application for manufacturing[M]. New York: Springer, 2011.

[5] INGRID DAUBECHIES. 小波十讲[M]. 贾洪峰. 译. 北京: 人民邮电出版社, 2019.

[6] 冉启文. 小波变换与分数傅里叶变换理论及应用[M]. 哈尔滨: 哈尔滨工业大学出版社, 2001.

[7] 唐向宏, 李齐良. 时频分析与小波变换(第二版)[M]. 北京: 科学出版社, 2016.

[8] 李业丽, 曾庆涛. 软计算原理与实现[M]. 北京: 电子工业出版社, 2020.

[9] 郭业才. 深度学习与信号处理[M]. 北京: 机械工业出版社, 2022.

[10] CHENG R, HU H P, TAN X H, et al. Initialization by a novel clustering for wavelet neural network as time series predictor[J]. Computational Intelligence and Neuroscience, 2015(12): 1-9.

[11] MONEM A M, LAYLA A A. Forecasting wind speed using the proposed wavelet neural network[J]. Discrete Dynamics in Nature and Society, 2023(8): 1-10.

[12] 杨金显, 刘鹏威. 基于小波神经网络修正自适应滤波的钻具姿态解算研究[J]. 制造业自动化, 2023, 45(1): 156-160.

[13] 尚荣华, 焦李成, 刘芳. 智能算法导论[M]. 北京: 清华大学出版社, 2021.

Neural Network Prediction with Wavelet Decomposition:

A case on maize commodities price sequence

HUO Yongliang

(School of Data Science, Guangzhou Huashang College, Guangzhou 511300, China)

Abstract: Dataset of 752 consecutive monthly international prices of maize released by the World Bank, were regarded as discrete price time series, and the wavelet decomposition method based on Mallat's algorithm was used to split this sequence signal into several high-frequency components and one low-frequency component, each one was then imported into the recurrent neural network, after that the predicted values of each component from the networks were accumulated as the final predicted price. Research experiments show that the neural network model using wavelet decomposition can flexibly capture high-frequency and low-frequency signals in the price time series of maize, and accurately fit and predict the values of those parts, which tells that the method has practical application significance for scenarios with frequent and violent price fluctuations.

Keywords: wavelet decomposition; neural networks; time series