Advances in Applied Mathematics 应用数学进展, 2023, 12(9), 3920-3932
Published Online September 2023 in Hans. https://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2023.129384

Hans汉斯

基于S-V-PSAL混合模型的股票预测研究

孙钰华*，吕卫东#，杜潇鉴

兰州交通大学数理学院，甘肃 兰州

收稿日期：2023年8月9日；录用日期：2023年9月3日；发布日期：2023年9月11日

摘　要

由于股票价格数据具有非平稳、非线性、复杂性高等特点，欲对其进行预测就存在一定的困难，提出了一种基于S-V-PSAL混合模型的预测方法。首先使用奇异谱分析(SSA)对股票历史数据进行一次分解，得到趋势项和噪声项。对于较平稳的趋势项，使用支持向量回归(SVR)模型进行预测；对复杂度依旧很高的噪声项序列，利用变分模态分解(VMD)再次分解，并使用长短期记忆网络和注意力机制(ALSTM)对得到的模态函数(IMFs)和残差序列(res)进行预测。最后将各预测结果重构得到最终结果。文中使用亿纬锂能股票的历史数据对提出的模型进行检验，通过三种评价指标，可以表明提出的模型比其他对比模型得到的预测效果更好，有更高的准确性。

关键词

股票价格预测，奇异谱分析，变分模态分解，支持向量回归，长短期记忆网络

Research on Stock Prediction Based on S-V-PSAL Mixed Model

Yuhua Sun*, Weidong Lyu#, Xiaojian Du

School of Mathematics and Physics, Lanzhou Jiaotong University, Lanzhou Gansu

Received: Aug. 9th, 2023; accepted: Sep. 3rd, 2023; published: Sep. 11th, 2023

Abstract

It is difficult to forecast stock price data because of its non-stationary, nonlinear and complex characteristics. A prediction method based on S-V-PSAL mixed model is proposed. First, the historical stock data is decomposed by singular spectrum analysis (SSA), and the trend term and

*第一作者。
#通讯作者。

文章引用: 孙钰华, 吕卫东, 杜潇鉴. 基于 S-V-PSAL 混合模型的股票预测研究[J]. 应用数学进展, 2023, 12(9): 3920-3932.
DOI: 10.12677/aam.2023.129384

noise term are obtained. For stable trend items, support vector regression (SVR) model is used to predict them. The noise sequence is decomposed again by variational mode decomposition (VMD), and the modal function (IMFs) and residual sequence (res) are predicted by long short-term memory network and attention mechanism (ALSTM). Finally, the forecast results are reconstructed to get the final result. In this paper, the historical data of EVE Energy stock is used to test the proposed model. Through three evaluation indexes, it can be shown that the proposed model has better prediction effect and higher accuracy than other comparison models.

Keywords

Stock Price Prediction, Singular Spectrum Analysis, Variational Mode Decomposition, Support Vector Regression, Long Short-Term Memory

1. 引言

金融市场是经济活动的重要组成成分，对国家、对企业以及对于个人都有着重要的意义。股票市场对于实体经济的影响和作用都是相互的，作为金融市场的一个重要的构成部分，股票市场从它一出现就引起了人们的关注，并与很多人的生活紧密相关。近些年来，我国经济发展态势较好，人们的生活水平也有显著提升，与此同时带来的是手里资金的闲置，简单的银行储存已不能满足人们的需求。随着投资和金融意识的增长，大家的投资需求更多了，与此同时市场上可供选择的投资类型也更多了，股票市场就是其中之一。高风险和高回报是股票市场的特性，长期以来人们被这特性不断吸引到其中，股票市场也逐渐成为大众投资的重要手段之一。

因此，股票价格预测一直是金融市场领域的研究热点。股票市场中的数据包含以下几个特点：第一是数据量非常大，股市每天的交易都会产生大量的数据，用好这些数据，并且找出规律性非常重要；第二是数据间关系的错综复杂，股市中包含相关和不相关的，也包含线性的和非线性的关系；第三是数据的动态性，股市中的数据时刻都在发生变化。一些研究者在对股票进行分析和预测的时候选择建立统计计量模型，由于计量模型对数据有整体性、平稳性、低噪声等非常严格的要求，加上假定限制过多，这对于现实中的股票市场数据往往是很难达到的，所以该模型用于实际的预测中很难得到想要的预测效果。近来更多的国内外学者在对股票进行预测研究时，往往使用机器学习的方法，如逻辑回归、支持向量机和循环神经网络等。

近年来，随着预测要求的精度越来越高且单一浅层的机器学习模型对于股票的波动性、记忆性等特性无法很好的克服等问题的出现，产生了将分解算法与时序预测模型相结合的组合模型。因为股票数据中包含各种影响分量，它们在相互作用时会遮挡我们要识别的规律或让规律扭曲，因此有理由将股票的过程分解为单个分量并单独分析每个分量的规律，从而进一步提高股票预测的精度。常见的分解算法有奇异谱分析(SSA)、经验模态分解(EMD)、集合经验模态分解(CEEMDAN)、变分模态分解(VMD)等，他们被广泛应用在金融时序数据分解中。比如，刘遵雄等通过基于 SSA 的广义神经预测模型对同方股份的收盘价进行预测，得到了较高的预测精度[1]。Cao 等将 CEEMDAN 和 LSTM 相结合，对全球主要的股市指数的收盘价进行预测，取得了较好的预测结果[2]。Niu 等使用变分模态分解对伦敦富时指数和纳斯达克指数进行分解，并使用基于注意力机制的门控循环网络进行预测，提高了预测的准确性[3]。程文辉等

通过 VMD 对原始金融序列一次分解后又使用 EEMD 对残差项进行二次分解，之后使用融合 FM 和 LSTM 的方法进行预测，得到了较高的预测精度[4]。

因此，本文基于将分解算法和机器学习模型相结合的思想，提出了一种基于二次分解与粒子群优化算法优化的支持向量回归和基于注意力机制的长短期记忆网络的混合模型，即 S-V-PSAL 混合模型，用来进行股票价格的预测。在使用亿纬锂能股票历史数据进行的实证分析中，通过与其他模型进行预测效果的对比，验证了该模型的有效性。

本文余下部分的结构安排如下：第 1 节介绍奇异谱分析、变分模态分解、支持向量回归和长短期记忆网络模型的基本理论；第 2 节介绍所提出的 S-V-PSAL 混合模型的整体结构与具体流程；第 3 节进行实证分析；第 4 节中对本文进行总结及未来展望。

2. 相关工作

文章中提出的 S-V-PSAL 混合模型，综合了奇异谱分析和变分模态分解两次分解以及支持向量回归模型和长短期记忆网络模型。本节先分别对它们进行介绍，以便在前人的基础上提出混合模型。

2.1. 奇异谱分析

奇异谱分析(SSA)是一种处理非线性时序数据的方法，包括统计学和信号处理中的多种方法。SSA 先前一直被用于数字信号处理，在 1978 年被 Colebrook 应用于海洋浮游生物研究中，随后 Hassani 将其引入社会科学领域，对美国每月意外死亡人数进行预测[5]，现在 SSA 被广泛应用于气候、环境、社会科学以及金融等多个领域。SSA 的主要目的是将原始时间序列分解成不同的成分序列之和，包括长期趋势、噪声项，从而对原始数据进行分析或去噪，继而用于其他任务。

SSA 由分解与重建这两个阶段组成，在两个阶段又都包括两个独立的步骤，第一个阶段是将原始序列延时排列，将单变量时间序列增广为多变量序列也即矩阵形式，对其进行奇异值分解。第二个阶段根据特征值对序列分组，形成新的时间序列[6]。

2.2. 变分模态分解

变分模态分解(VMD)由 Dragomiretskiy 在 2014 年提出，是一种在模态变分和信号处理问题中的自适应、完全非递归的方法，可以根据输入信号的特性确定模态分解的数目 K，并匹配每种模态的最佳中心频率和有限带宽[7]。VMD 的目标是将一个原始输入信号分解成一组离散的具有特殊稀疏性的子信号[8]，也即模态。在保证分解序列各个模态的带宽之和最小的前提下，最终获得变分问题的最优解。

2.3. 支持向量回归

支持向量机(SVM)由 Vapnik 提出，主要应用于模式识别领域，后来发展到回归领域[9]。支持向量回归(SVR)就是其在回归领域的应用，这是一种利用核函数检测高维特征空间中最优回归超平面的回归方法。通过输入样本的特征向量，求出输入特征和输出值之间的回归关系，并根据此关系来对未来值进行预测[10]。

2.4. 长短期记忆网络

长短期记忆网络(LSTM)在 1997 年由 Hochreiter 提出，是一种改进后的循环神经网络[11]。LSTM 中存在门控单元，可以帮助系统存储更多信息。单元格通过打开和关闭门来决定是否删除或存储信息。单元由四个主要元件组成：输入门、具有自重复连接的神经元、遗忘门和输出门。遗忘门是允许细胞记住或忘记其先前状态的元素。LSTM 的具有自重复连接的神经元结构示意图如图 1 所示。

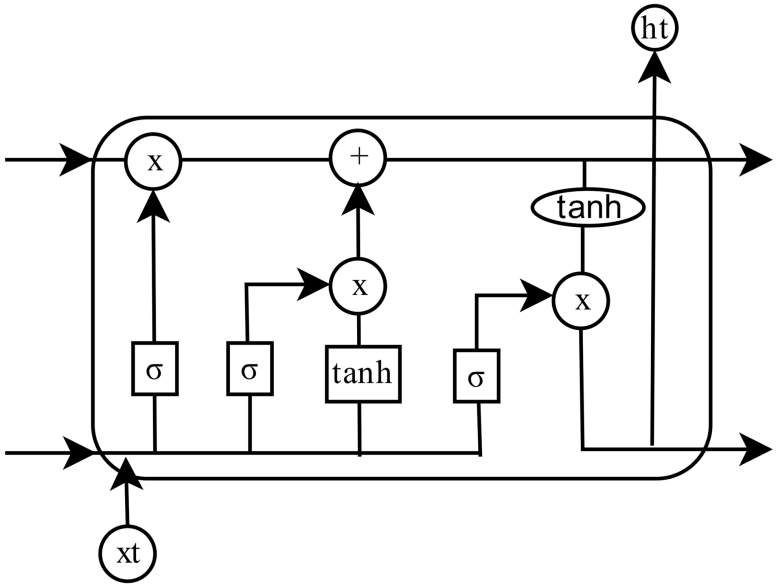

Figure 1. Unit structure of LSTM
图 1. LSTM 的单元结构

首先“遗忘门”决定从单元状态转储哪些信息，这个功能需要 h_{t-1}(从前一个隐藏层输出)和 x_t (当前输入)共同决定，并通过式(1)输出[0, 1]中的数字，其中 1 表示“完全保持”，0 表示“完全转储”。LSTM 决定在单元状态中的存储时有两个步骤。首先，由“输入门”来决定保留的信息，如公式(2)。然后，用 tanh 层对保留的消息生成向量 $\tilde{C}_t$ 来更新单元状态，如公式(3)。公式(4)将单元状态 C_{t-1} 更新为 C_t。最后“输出门”决定 LSTM 输出的内容。LSTM 首先运行一个 sigmoid 层，即公式(5)，它决定要输出哪个单元状态。并用 tanh 层把输出取值固定在−1 和 1 之间，并乘以 sigmoid 层的输出，如公式(6)，以此来输出想要的部分。

$$f_t = \sigma\left(W^f x_t + U^f h_{t-1}\right) \tag{1}$$

$$i_t = \sigma\left(W^i x_t + U^i h_{t-1}\right) \tag{2}$$

$$\tilde{C}_t = \tanh\left(W^n x_t + U^n h_{t-1}\right) \tag{3}$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{4}$$

$$o_t = \sigma\left(W^o x_t + U^o h_{t-1}\right) \tag{5}$$

$$h_t = o_t * \tanh\left(C_t\right) \tag{6}$$

3. 基于 S-V-PSAL 混合模型的股票预测

本文提出了一种基于 S-V-PSAL 混合模型的预测方法来进行股票价格预测。该混合模型首先利用奇异谱分析(SSA)将股票价格数据进行第一次分解，得到长期趋势子序列和噪声子序列，接着用变分模态分解方法(VMD)对噪声子序列进行第二次分解，得到一系列频率不同的子序列 IMFs，之后使用支持向量回归模型(SVR)对长期趋势子序列进行预测，并且使用粒子群优化算法(PSO)来确定 SVR 的最佳参数，包括核函数系数、惩罚因子和 epsilon。对于 VMD 分解之后的子序列，使用基于注意力机制的长短时记忆网络模型(ALSTM)来预测。最后重构各序列得到最终结果。预测的具体流程图如图 2 所示：

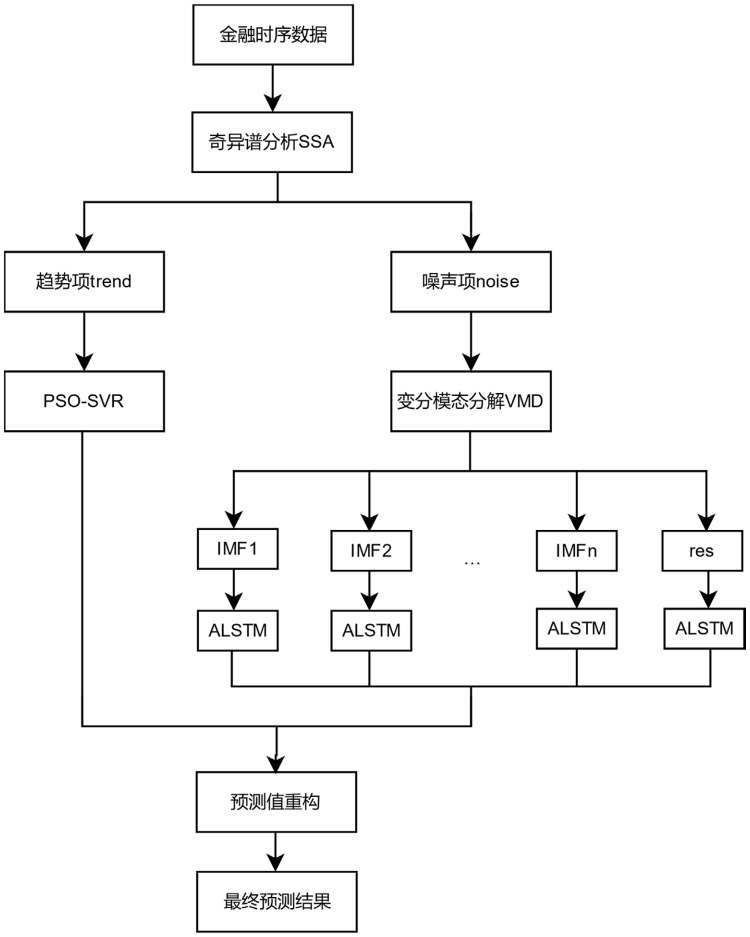

Figure 2. Financial time series data forecasting flow chart based on S-V-PSAL mixed model
图 2. 基于 S-V-PSAL 混合模型的金融时序数据预测流程图

4. 实证分析

4.1. 数据来源

为了验证使用混合模型预测的有效性，使用 python 从 baostock 网站获取股票历史数据，进行实证分析。实例中使用的数据为 2020 年 1 月 2 日~2022 年 12 月 30 日这三年的亿纬锂能股票(股票代码为 SZ300014)的历史数据，共 728 条交易日数据，每条数据包含的特征有 open、high、low、close 等。部分数据展示如下表 1 所示。图 3 展示的是原始股票数据的收盘价序列。

Table 1. Partial raw data
表 1. 部分原始数据

date	open	high	low	close	volume	psTTM
2020/1/2	26.61	26.82	25.83	26.43	22,840,092	8.26
2020/1/3	25.94	26.24	24.89	25.19	40,936,496	7.87
2020/1/6	25.41	27.62	25.40	27.19	47,402,752	8.50
.	.	.	.	.	.	.

Continued

2022/12/28	90.30	90.43	88.08	88.85	15,375,506	6.10
2022/12/29	88.61	91.86	87.83	89.91	13,363,723	6.17
2022/12/30	90.80	91.42	87.87	87.90	11,786,755	6.04

Figure 3. Closing price of stock
图 3. 股票收盘价

由于数据之间有不同的量纲，为了实验后续的顺利进行，需要对原始数据进行归一化处理，并对处理后的数据以 8:2 进行划分，前 578 条数据为训练集，后 145 条数据为测试集。

4.2. 模型评价指标

下面使用平均绝对误差(MAE)、均方误差(MSE)和决定系数(R^2)这三种在回归问题中常用的评价指标，来对预测结果进行评价，以更直观、定量的比较所提出的 S-V-PSAL 混合模型与对比模型的对股票价格的预测性能。这三种评价指标的计算公式如下式(7) (8) (9)所示：

$$\mathrm{MAE}=\frac{1}{S}\sum_{s=1}^{S}\left|\widehat{y_s}-y_s\right| \tag{7}$$

$$\mathrm{MSE}=\frac{1}{S}\sum_{s=1}^{S}\left(\widehat{y_s}-y_s\right)^2 \tag{8}$$

$$R^2=\frac{\sum_{s=1}^{S}\left(\widehat{y_s}-\overline{y_s}\right)^2}{\sum_{s=1}^{S}\left(y_s-\overline{y_s}\right)^2} \tag{9}$$

其中，S 是数据长度，y 是数据的真实值，$\hat{y}$ 是预测值，$\bar{y}$ 是真实值的平均值。作为评价指标，MAE、MSE 的值越小，表明模型预测效果越准确。R^2 的值越靠近 1，表明模型的拟合程度越好。

4.3. 实验结果分析

4.3.1. 二次分解

首先对数据进行一次奇异谱分析，将亿纬锂能股票的收盘价序列以 20 为窗口长度构造轨迹矩阵，继续对轨迹矩阵进行奇异值分解，得到的奇异值越大表示包含的有用信息更多，绘制奇异值占比图如图 4

所示，从图中可以看出前两阶包含了 90%以上的信息，所以将前两阶作为趋势序列，将三阶以后的部分作为噪声序列，由此将原始数据分解为 trend 序列和 noise 序列，如图 5 所示。

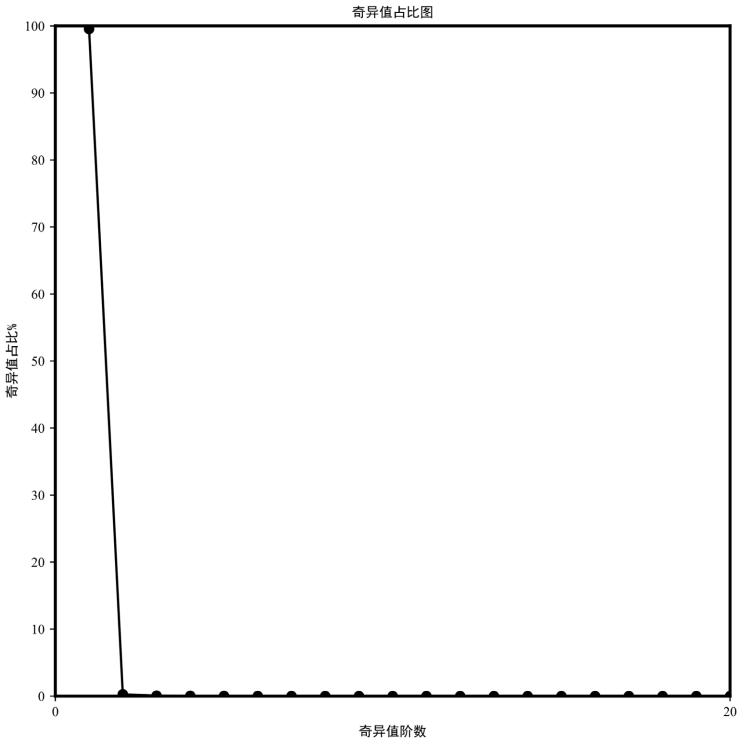

Figure 4. Proportion graph of singular values
图 4. 奇异值占比图

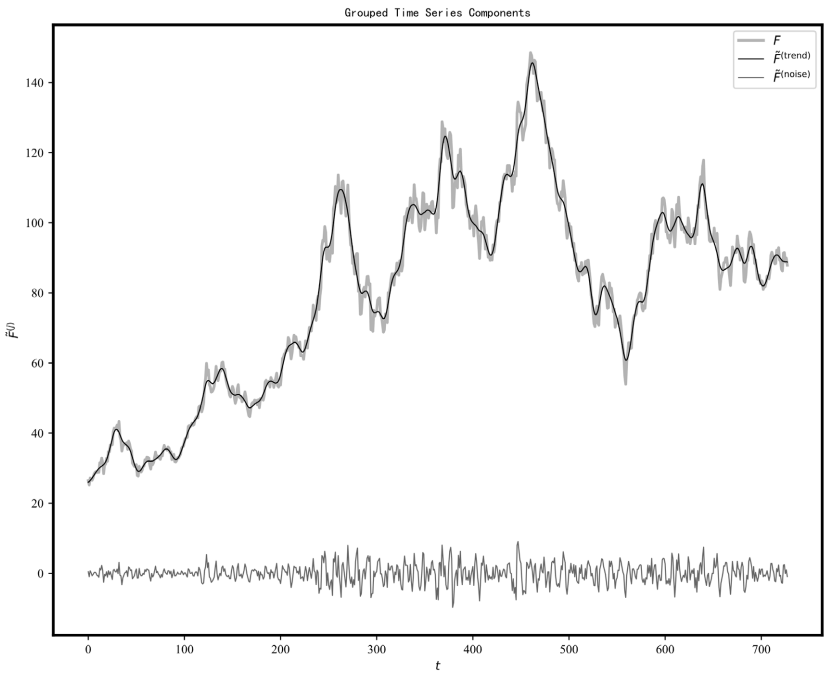

Figure 5. Results of singular spectrum analysis
图 5. 奇异谱分析结果

之后，对 SSA 得到的噪声序列继续进行变分模态分解。选择分解模态的数目为 3。noise 序列分解得到的模态 IMFs 如图 6 所示，残差序列如图 7 所示。

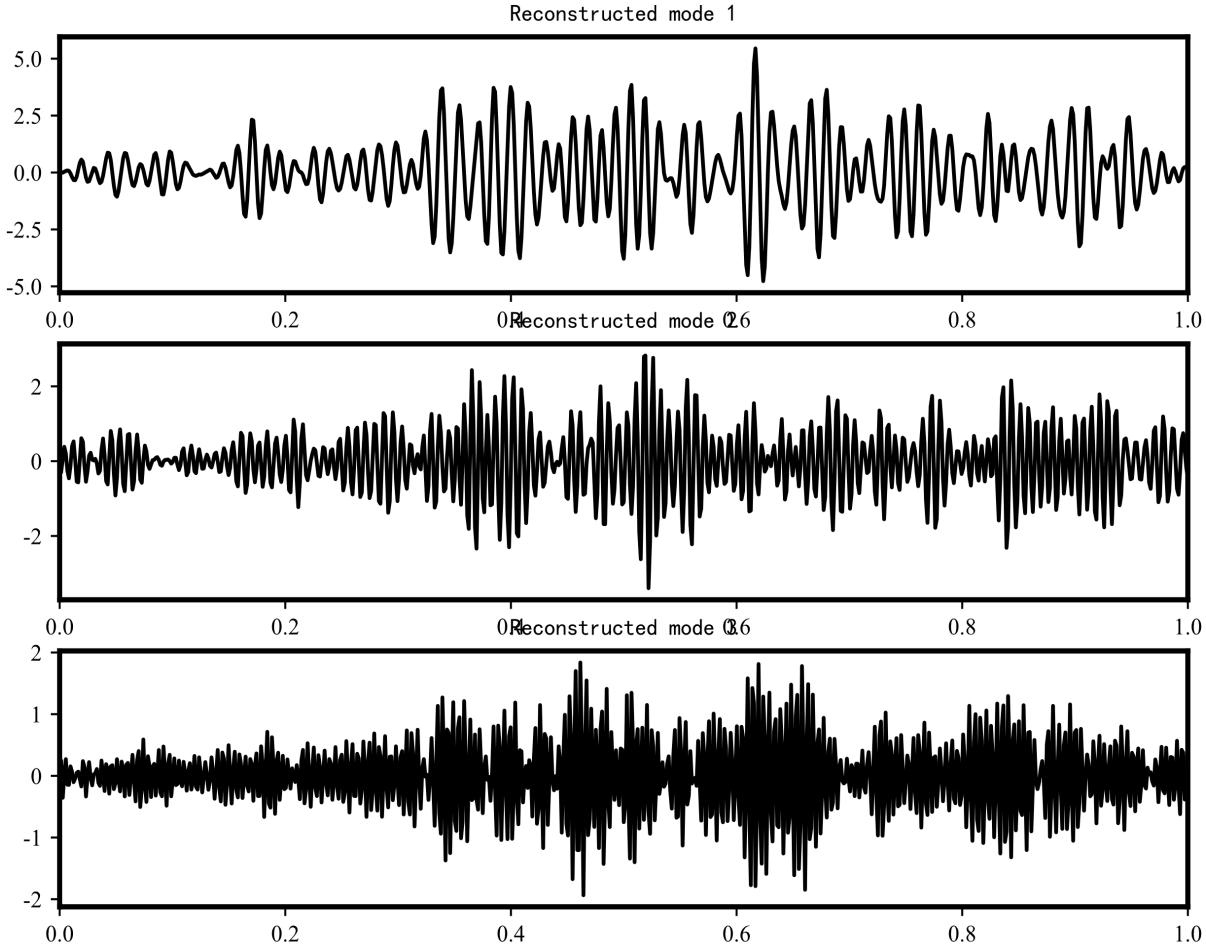

Figure 6. IMFs of VMD decomposition
图 6. VMD 分解后的 IMFs

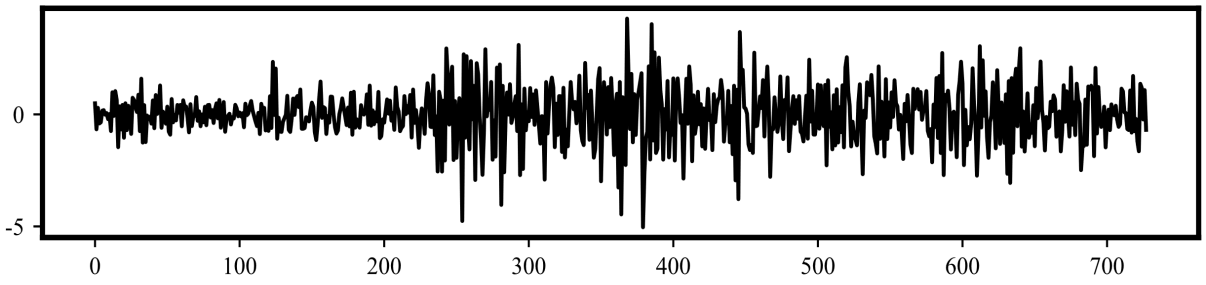

Figure 7. Residual sequence of VMD decomposition
图 7. VMD 分解后的残差序列

4.3.2. 预测结果

首先使用 PSO-SVR 对原始序列进行预测，结果如图 8 所示。

然后分别对 SSA 得到的 trend 序列和 noise 序列进行 PSO-SVR 预测，结果分别如图 9、图 10 所示。

对两段序列得到的预测序列相加，得到 SSA-PSO-SVR 模型预测结果，如图 11 所示。

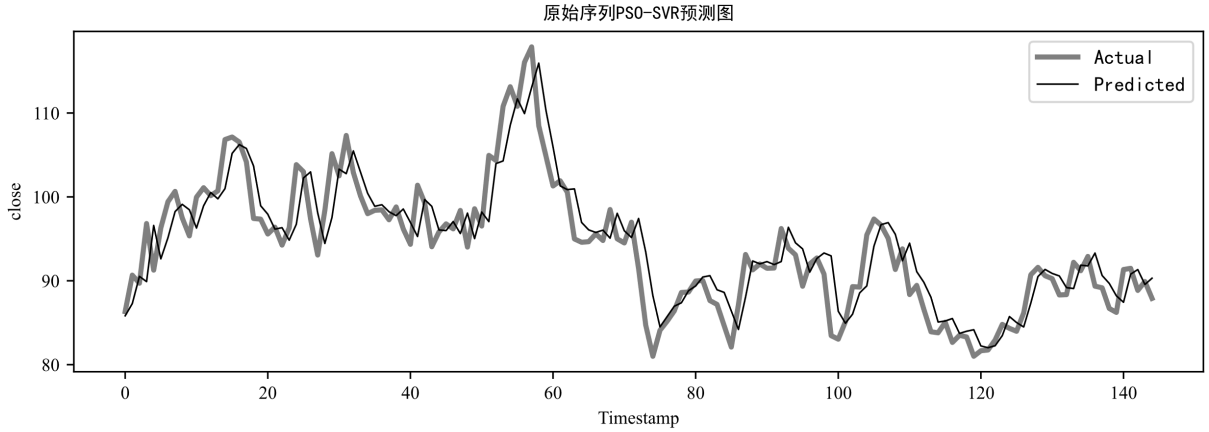

Figure 8. Prediction of original sequence PSO-SVR
图 8. 原始序列 PSO-SVR 预测图

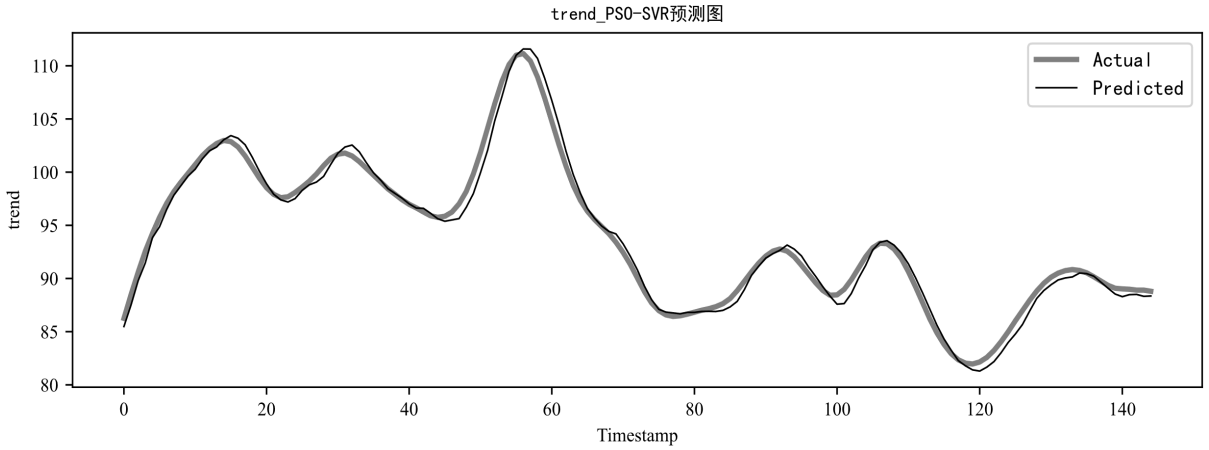

Figure 9. Prediction of trend_PSO-SVR
图 9. trend_PSO-SVR 预测图

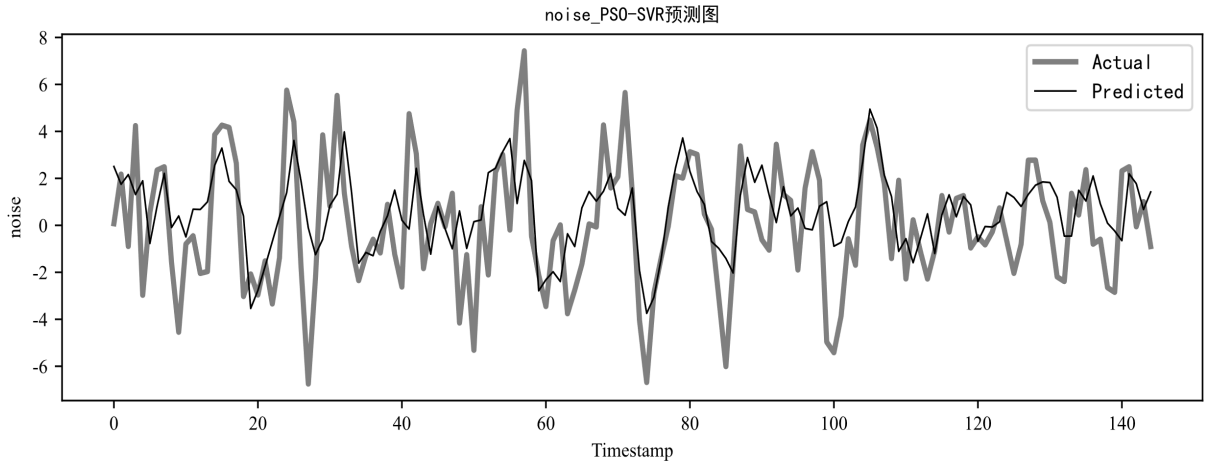

Figure 10. Prediction of noise_PSO-SVR
图 10. noise_PSO-SVR 预测图

之后对 VMD 得到的子序列 IMF1、IMF2、IMF3 和残差序列使用 ALSTM 进行预测。结果如图 12、图 13、图 14、图 15 所示。在表 2 中列出了各 IMFs 预测结果的评价指标数值。

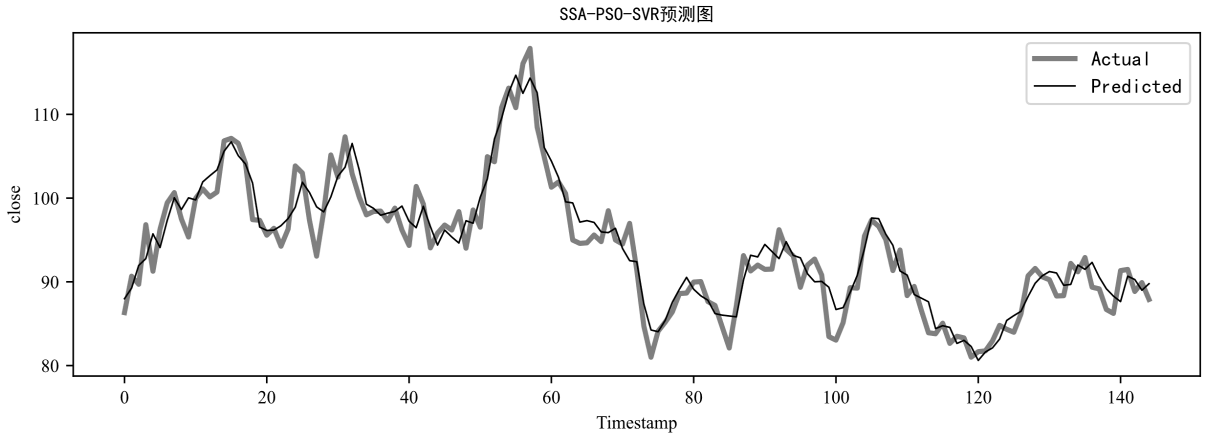

Figure 11. Prediction of SSA-PSO-SVR
图 11. SSA-PSO-SVR 预测图

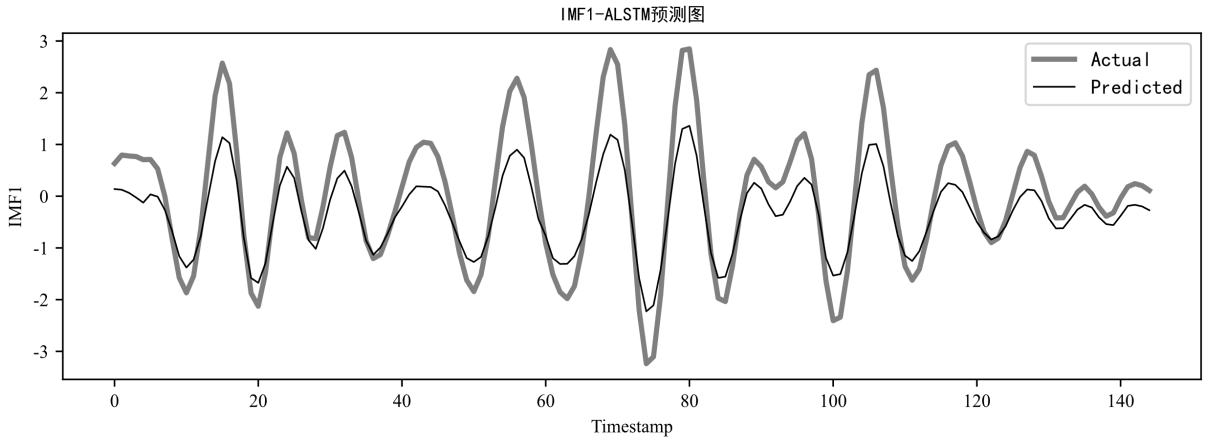

Figure 12. Prediction of IMF1-ALSTM
图 12. IMF1-ALSTM 预测图

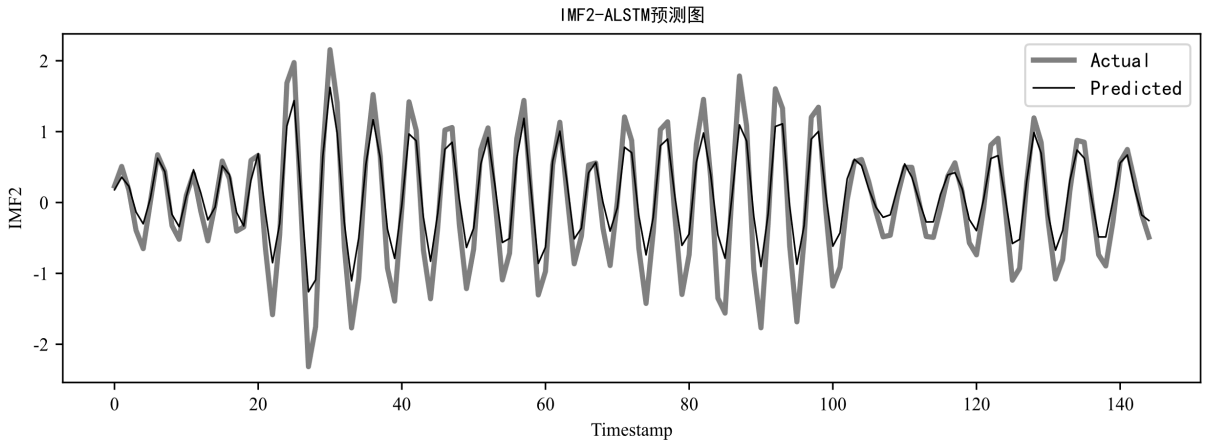

Figure 13. Prediction of IMF2-ALSTM
图 13. IMF2-ALSTM 预测图

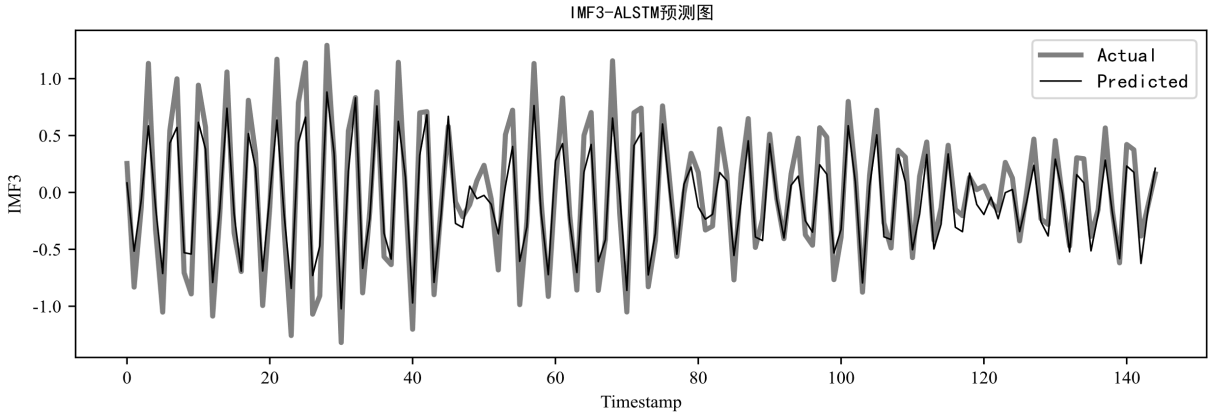

Figure 14. Prediction of IMF3-ALSTM
图 14. IMF3-ALSTM 预测图

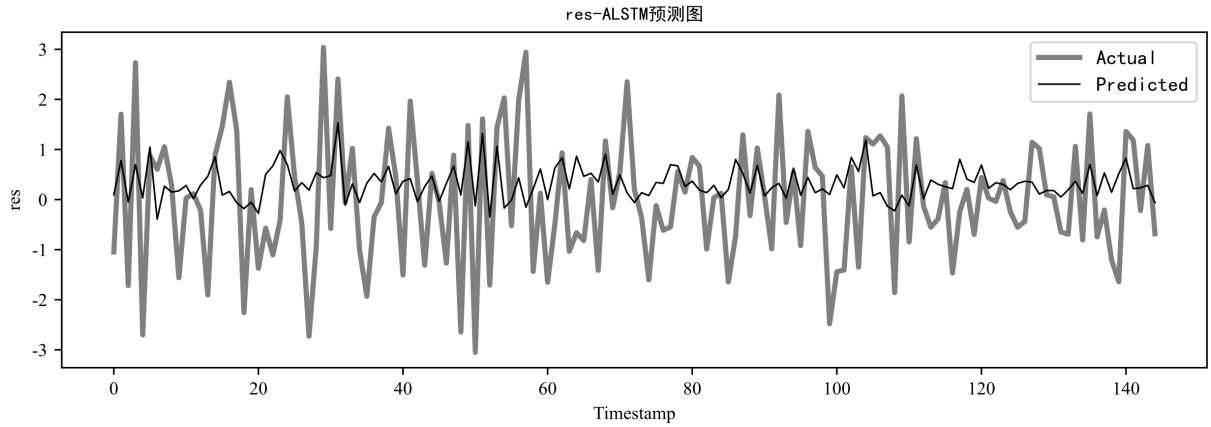

Figure 15. Prediction of res-ALSTM
图 15. res-ALSTM 预测图

Table 2. Evaluation index of each IMFs prediction result
表 2. 各 IMFs 预测结果的评价指标

imfs	MAE	MSE	R^2
IMF1	0.59	0.51	0.685
IMF2	0.30	0.14	0.824
IMF3	0.19	0.05	0.857
res	0.97	1.48	0.059

最后，将 PSO-SVR 预测的 trend 序列的预测结果和 ALSTM 预测的 IMFs 和 res 序列的预测结果重构，得到的最终预测结果如图 16 所示。

为了比较所提出的 S-V-PSAL 混合模型与 PSO-SVR 模型、SSA-PSO-SVR 模型的预测结果，在表 3 中列出了各个模型的评价指标，可以看出未经过分解的 PSO-SVR 模型的 MAE、MSE 最大，R^2 最小，经过一次分解的 SSA-PSO-SVR 模型各评价指标均居中，预测效果相比未经过分解的模型较好，而所提出的经过二次分解的 S-V-PSAL 混合模型的 MAE、MSE 均最小，R^2 最大，表示所提出的混合模型的预测效果最好。

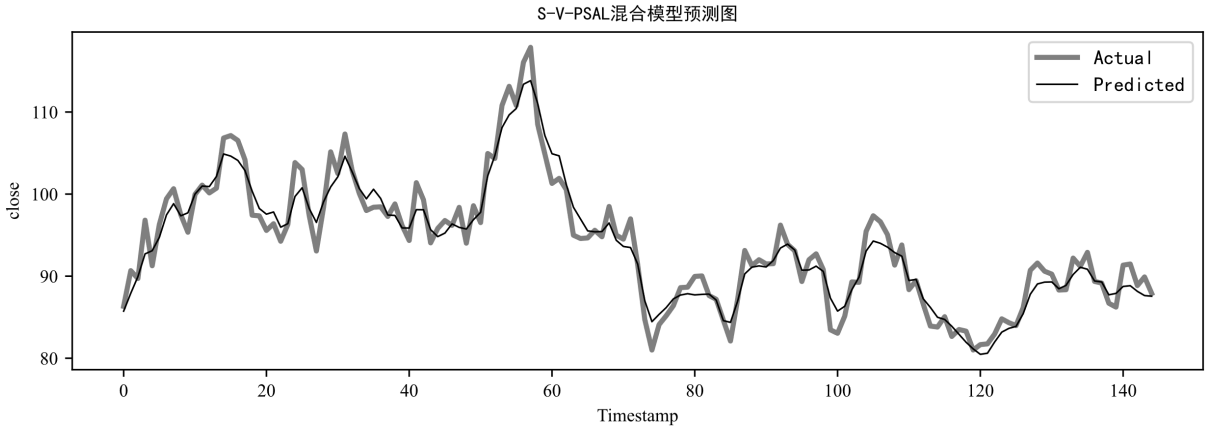

Figure 16. Prediction of S-V-PSAL mixed model
图 16. S-V-PSAL 混合模型预测图

Table 3. Evaluation index of each model
表 3. 各个模型评价指标

Model	MAE	MSE	R^2
PSO-SVR	2.57	11.23	0.759
SSA-PSO-SVR	1.89	5.46	0.895
S-V-PSAL	1.48	3.37	0.934

5. 结语

股票价格时间序列数据具有非线性、非平稳、复杂性高等特点，经过分解，能够一定程度的降低序列的复杂度。本文提出了基于 S-V-PSAL 混合模型的股票价格预测方法，经过奇异谱分析和变分模态分解二次分解，发挥分解算法的优势，有效降低了数据的复杂度。使用 PSO-SVR 和 ALSTM 对分解后的子序列进行预测，再将预测结果重构得到最终预测结果。使用亿纬锂能的股票历史数据进行验证，与 PSO-SVR、SSA-PSO-SVR 相比，S-V-PSAL 混合模型的预测效果较好。

虽然从评价指标看出 S-V-PSAL 混合模型具有较好的预测效果，但是仍需要对其进一步的改进。本文中对于 VMD 和 ALSTM 的参数设置，受人为影响，通过不断地调参来选定，后续考虑使用优化算法或者样本熵的方法来设置参数，使模型效果更优。

基金项目

国家自然科学基金(11961039)。

参考文献

[1] 刘遵雄, 周天清. 基于奇异谱分析的 GRNN 模型在金融时间序列中的应用[J]. 华东交通大学学报, 2011, 28(2): 29-34.

[2] Cao, J., Li, Z. and Li, J. (2019) Financial time series forecasting model based on CEEMDAN and LSTM. *Physica A: Statistical Mechanics and Its Applications*, **519**, 127-139. https://doi.org/10.1016/j.physa.2018.11.061

[3] Niu, H.L. and Xu, K.L. (2020) A Hybrid Model Combining Variational Mode Decomposition and an Attention-GRU Network for Stock Price Index Forecasting. *Mathematical Biosciences and Engineering*, **17**, 7151-7166. https://doi.org/10.3934/mbe.2020367

[4] 程文辉, 车文刚. 基于二次分解与 LSTM 的金融时间序列预测算法研究[J]. 重庆邮电大学学报(自然科学版),

2022, 34(4): 638-645.

[5] Hassani, H. (2007) Singular Spectrum Analysis: Methodology and Comparison. *Journal of Data Science*, **5**, 239-257. https://doi.org/10.6339/JDS.2007.05(2).396

[6] 张一, 惠晓峰. 基于奇异谱分析的汇率预测研究[J]. 统计与决策, 2012(6): 29-31.

[7] Dragomiretskiy, K. and Zosso, D. (2014) Variational Mode Decomposition. *IEEE Transactions on Signal Processing*, **62**, 531-544. https://doi.org/10.1109/TSP.2013.2288675

[8] 王秀杰, 王玲, 滕振敏, 等. 基于 VMD-PSO-LSTM 模型的日径流多步预测研究[J/OL]. 水利水运工程学报: 1-10. http://kns.cnki.net/kcms/detail/32.1613.TV.20230313.1130.002.html, 2023-09-06.

[9] 李坤, 谭梦羽. 基于小波支持向量机回归的股票预测[J]. 统计与决策, 2014(6): 32-36.

[10] 付占局. 基于文本情感分析及 GA-SVR 模型的股价预测研究[D]: [硕士学位论文]. 北京: 北京交通大学, 2021.

[11] 邓德军, 徐洪珍, 韦诗玥. E-V-ALSTM 模型的股价预测[J]. 计算机工程与应用, 2023, 59(6): 101-112.