

分类号_____

密级 公开

UDC _____

编号 10741

兰州财经大学

LANZHOU UNIVERSITY OF FINANCE AND ECONOMICS

硕士学位论文

(专业学位)

论文题目 基于“分解-重构-集成”范式的
预测方法研究及其应用

研究生姓名: 赵盼盼

指导教师姓名、职称: 孙景云 副教授

学科、专业名称: 统计学、应用统计硕士

研究方向: 大数据分析

提交日期: 2022年5月30日

独创性声明

本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文作者签名: 赵盼盼 签字日期: 2022.5.30

导师签名: 孙景云 签字日期: 2022.5.30

关于论文使用授权的说明

本人完全了解学校关于保留、使用学位论文的各项规定，同意（选择“同意”/“不同意”）以下事项：

1. 学校有权保留本论文的复印件和磁盘，允许论文被查阅和借阅，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文；
2. 学校有权将本人的学位论文提交至清华大学“中国学术期刊（光盘版）电子杂志社”用于出版和编入 CNKI《中国知识资源总库》或其他同类数据库，传播本学位论文的全部或部分内容。

学位论文作者签名: 赵盼盼 签字日期: 2022.5.30

导师签名: 孙景云 签字日期: 2022.5.30

Research and application of forecasting method based on "decomposition- reconstruction-ensemble" paradigm

Candidate: Panpan Zhao

Supervisor: Jingyun Sun

摘要

理想状态下的数据一般具有线性、平稳性及复杂性低等特点，数据的准确预测，不仅可以为投资者提供决策支持，也可以让政府制定相关政策有参考依据。然而，现实中的数据由于外部各种因素的影响而呈现出复杂性高的特性。此外，再加上突发事件的影响，使数据的预测变得越来越困难，因此寻找一种可靠且有效的方法来预测数据至关重要。

当前金融数据预测的方法主要有传统的计量经济方法、人工智能方法和分解集成方法。传统的计量经济方法在非线性、非平稳性及一些复杂数据的处理中无能为力。人工智能方法通过让机器模拟人类智力来解决一些比较复杂的任务，从而获得更准确的预测，但人工智能方法存在参数敏感、容易陷入局部极小值、过度拟合等问题。分解集成方法是当前研究的主流方法，它将复杂的数据分解为易于描述的简单分量，从而降低建模的难度，提高模型的预测性能，达到“化繁为简，各个击破”的目的。但是，现有的分解集成方法，当分量过多对各个分量分别进行预测时，会增加计算成本，最终在结果集成时可能会出现误差累加的问题。

针对以上问题，本文基于“先分解后集成”的思想，从数据的分解、分量的重构和预测方法的优化三个方面出发，构建基于“分解-重构-集成”范式的预测方法，在此基础上开展相关实证研究：基于集合经验模态分解、重构和粒子群优化最小二乘支持向量机的美元兑人民币汇率预测；基于二次分解（CEEMDAN-CEEMDAN）、重构、混沌麻雀搜索算法优化 KELM 的比特币价格预测；基于二次分解（ICEEMDAN-EMD）、二次重构及混沌麻雀搜索算法优化 KELM 的原油期货价格预测。利用误差评价指标和 DM 检验，与其他多种模型的预测结果对比，实证结果表明，本文构建的模型预测结果更好，精度更高。

关键词： 二次分解；排列熵；相空间重构；智能优化算法

Abstract

The data in the ideal state generally has the characteristics of linearity, stability and low complexity. The accurate prediction of the data can not only provide decision support for investors, but also provide a reference basis for the government to formulate relevant policies. However, the real data show the characteristics of high complexity due to the influence of various external factors. In addition, coupled with the impact of emergencies, the prediction of data becomes more and more difficult. Therefore, it is very important to find a reliable and effective method to predict data.

At present, the methods of financial data prediction mainly include traditional econometric methods, artificial intelligence methods and decomposition integration methods. Traditional econometric forecasting methods are powerless in the processing of nonlinearity, nonstationarity and some complex data. Artificial intelligence methods solve some complex tasks by making machines simulate human intelligence, so as to obtain more accurate prediction. However, artificial intelligence methods are sensitive to parameters, easy to fall into local minimum, over fitting and so on. Decomposition integration method is the mainstream method in current research. It decomposes complex data into simple components that are easy to describe, so as to reduce the difficulty of modeling, improve the prediction performance of the model, and achieve the purpose of "turning complexity into simplicity and breaking each one". However, in the existing decomposition integration methods, when there are too many components to predict each component separately, it will increase the calculation cost, and finally there may be the problem of error accumulation in the result integration.

Aiming at the above problems, based on the idea of "decomposition first and then integration", this paper constructs a prediction method based on the "decomposition-reconstruction-integration" paradigm from three aspects: data decomposition, component reconstruction and prediction method optimization. On this basis, relevant empirical research is carried out: Based on set empirical mode decomposition Reconstruction and particle swarm optimization of least squares support vector machine to predict the exchange rate of US dollar against RMB; Based on quadratic

decomposition (CEEMDAN-CEEMDAN), reconstruction and chaotic sparrow search algorithm, kelm's bitcoin price prediction is optimized; KELM's crude oil futures price forecast is optimized based on ICEEMDAN-EMD, quadratic reconstruction and chaotic sparrow search algorithm. Using the error evaluation index and DM test, compared with the prediction results of other models, the empirical results show that the prediction results of the model constructed in this paper are better and more accurat.

Key words: Quadratic decomposition; Permutation entropy; Phase space reconstruction; Intelligent optimization algorithm

目 录

1 引言	1
1.1 研究背景及意义	1
1.2 研究内容与技术路线	2
1.2.1 研究内容	2
1.2.2 技术路线	3
1.3 本文的创新点	5
2 文献综述	6
2.1 传统计量经济方法下的金融预测	6
2.2 人工智能方法下的金融预测	7
2.3 分解集成方法下的金融预测	8
3 研究方法	10
3.1 分解方法	10
3.1.1 EEMD 分解方法	10
3.1.2 CEEMDAN 分解方法	11
3.1.3 ICEEMDAN 方法	12
3.2 重构方法	13
3.2.1 样本熵	13
3.2.2 排列熵	14
3.3 优化方法	15
3.3.1 粒子群算法	15
3.3.2 混沌麻雀搜索算法	16
3.4 输入变量的确定方法——相空间重构	18
3.5 预测方法	20
3.5.1 最小二乘支持向量机模型	20
3.5.2 极限学习机	21
3.5.3 核极限学习机	22
3.6 评价指标	22
3.7 本章小结	23
4 基于分解重构集成的美元兑人民币汇率预测	24

4.1 基于 EEMD-SE-PSO-LSSVM 模型的框架构建	24
4.2 数据来源.....	25
4.3 美元兑人民币汇率的 EEMD 分解和重构	26
4.4 预测结果及评价.....	28
4.5 本章小结.....	30
5 基于二次分解集成的比特币价格预测	31
5.1 CEEMDAN-PE-CEEMDAN-PSR-CSSA-KELM 模型的框架构建.....	31
5.2 数据来源.....	33
5.3 CEEMDAN 分解及重构结果	33
5.4 CEEMDAN 二次分解结果	35
5.5 相空间重构的结果.....	36
5.6 模型比较分析.....	37
5.7 本章小结.....	42
6 基于二次分解重构集成的原油期货价格预测	43
6.1 ICEEMDAN-PE-EMD-PSR-CSSA-KELM 模型框架的构建	43
6.2 数据来源	45
6.3 ICEEMDAN 分解及重构结果	46
6.4 EMD 二次分解及重构结果	48
6.5 模型比较分析.....	49
6.6 本章小结.....	57
7 结论与展望	58
7.1 结论.....	58
7.2 展望	59
参考文献	60
硕士期间完成的工作.....	66
致谢	67

1 引言

1.1 研究背景及意义

预测是指根据事物过去和现在的发展规律对事物进行估计。近年来,越来越多的专家学者开始关注数据预测,数据预测研究涉及多个领域,譬如金融、能源、航空航天、经济、环境等。由于各领域的数据都有其独特的特点,因此对数据进行预测和挖掘其潜在的价值至关重要。

理想状态下的数据一般具有线性、平稳性及复杂性低等特点,但现实中的数据由于受到外部因素的影响,呈现出复杂性高的特点。此外,加上突发事件的影响,数据的预测变得越来越困难。现有的金融数据预测方法主要有传统的计量经济方法、人工智能方法和分解集成方法。传统的计量经济方法在非线性、非平稳性及一些复杂数据处理中无能为力,然后一些学者引入人工智能方法,发现通过机器模拟人类智力,可解决一些比较复杂的任务,得到更准确的预测。然而,人工智能方法具有对参数敏感、过度拟合、容易陷入局部极小值等问题。特别是处于大数据时代,数据多样繁杂,简单的预测方法已不能满足我们的需求,对数据进行预测和建模来说,这是一个极具挑战意义的难题。

随着研究的深入,针对传统的计量经济方法和单一的人工智能方法的缺点,一些学者在预测模型中加入了分解方法和智能优化算法,通过充分利用各模型和算法的优势,来提高模型的预测能力,这种混合预测方法就是分解集成模型。分解集成模型是当前研究的主流方法,它通过将数据分解为易于描述的简单分量,从而降低建模难度,达到“化繁为简,各个击破”的目的。

但是,现有的分解集成模型也有其缺点,当分量过多对各个分量分别进行预测时,会增加计算成本,最终在结果集成时可能会出现误差累加的情况。为解决这一问题,本文在传统分解集成模型的基础上进行改进,构建基于“分解-重构-集成”范式的预测方法,在此基础上开展相关的研究和应用,创新相关的预测理论和方法,提高模型的预测性能。

本文的研究具有一定的理论意义和现实意义。理论意义在于:本文在吸收前人研究成果的基础上,梳理了金融预测领域已有的一些研究成果,分析了该领域

前沿的分解集成方法，在此基础上使用“分解-重构-集成”的新思路；现实意义在于：其一，提高预测的准确性，分析其波动规律，有利于政府提前掌握金融市场动态，对相关预警和决策政策起到指导作用，并调整相关宏观政策。其二，对于相关企业来说，准确把握价格波动，特别是方向上的变动，可以帮助他们做出合理的交易决策，避免价格波动的风险，使企业收益最大化。

1.2 研究内容与技术路线

1.2.1 研究内容

本文主要研究金融数据的预测，通过将原始数据分解、模态分量重构、智能优化算法优化预测模型，对“分解集成”范式进行创新，使用“分解-重构-集成”范式。与当前现有的预测方法相比较，本文构建的模型降低了预测误差，提高了预测精度。本文研究的主要内容如下：

(1) 基于分解重构集成的美元兑人民币汇率预测

在预测美元兑人民币汇率时，建立了基于集合经验模态（EEMD）分解、样本熵（SE）和粒子群（PSO）优化最小二乘支持向量机（LSSVM）的组合预测方法，即 EEMD-SE-PSO-LSSVM。

(2) 基于二次分解集成的比特币价格预测

在预测比特币价格时，建立了基于 CEEMDAN-CEEMDAN 二次分解、排列熵（PE）重构、相空间重构（PSR）以及混沌麻雀搜索算法（CSSA）优化 KELM 的组合预测方法，即 CEEMDAN-PE-CEEMDAN-PSR-CSSA-KELM。

(3) 基于二次分解重构集成的原油期货价格预测

在预测原油期货价格时，建立了基于 ICEEMDAN-EMD 二次分解、排列熵（PE）二次重构及混沌麻雀搜索算法（CSSA）优化 KELM 的原油期货价格预测，即 ICEEMDAN-PE-EMD-PSR-CSSA-KELM。本文对 Brent 原油和 WTI 原油做了实证分析，值得注意的是，在预测过程中，不仅对 2 种原油期货分别做了 1 步预测，而且做了 2 步预测和 3 步预测，以此来衡量本文构建的模型在短期及中期的预测能力。

1.2.2 技术路线

本文主要对特定金融领域时序数据进行分析 and 预测建模，通过梳理金融领域现有的预测方法，我们可以得出以下结论：分解集成方法是当前数据预测的主流；数据的分解方法和预测方法对最终结果有着重要的影响。由此，本文针对特定金融领域的的数据，通过将原始数据分解，分析分量的内部特征，对不同特性的分量进行处理并构建预测方法，同时，对预测方法进行优化，最终对“分解-集成”范式进行新的创新思路：(1) 对原始数据分解后，加入重构过程，以此来减少运算时间，提高预测精度。(2) 利用智能优化算法对现有的预测方法进行优化，以提高模型的整体预测精度。

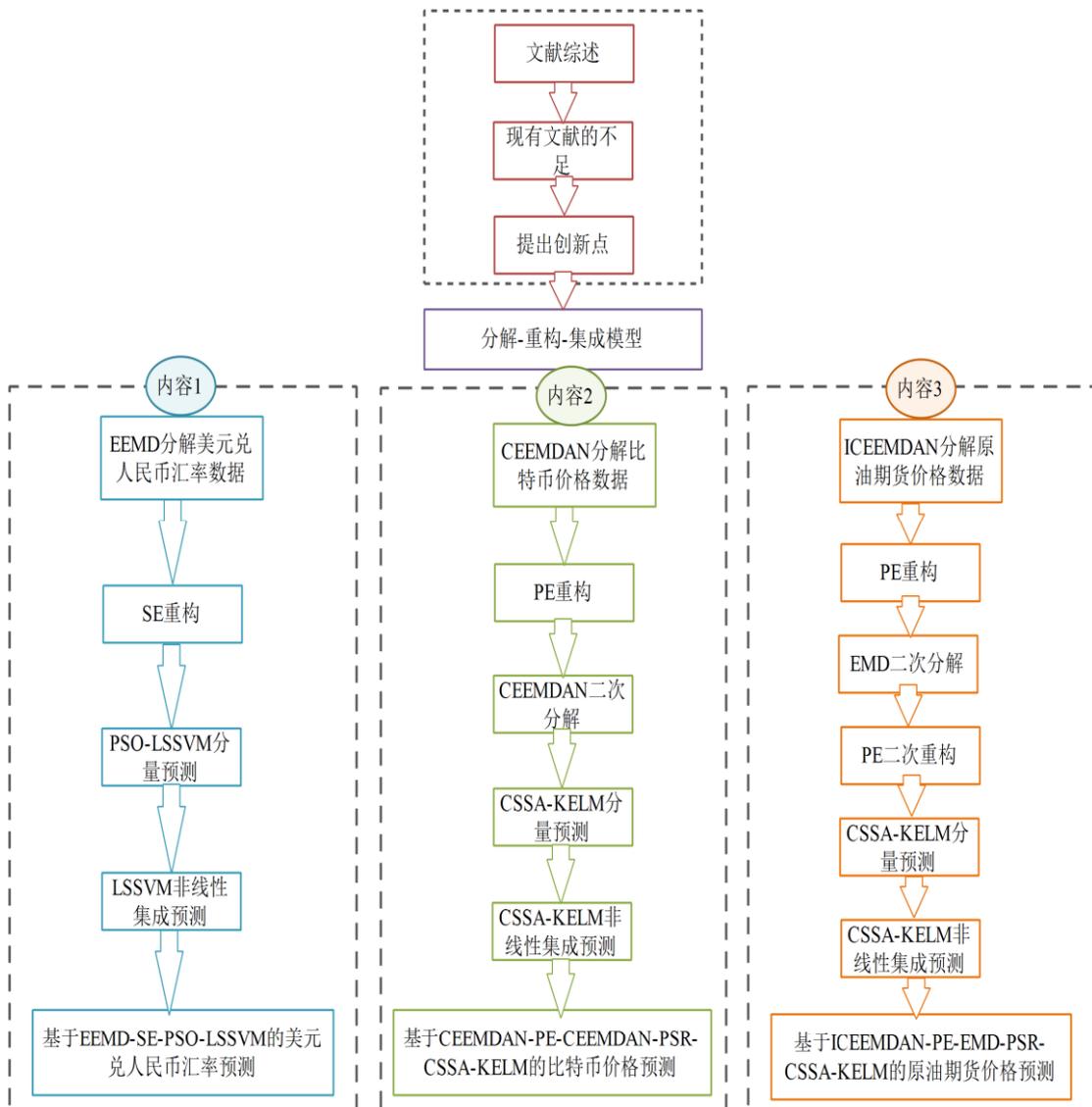


图 1.1 本文的技术路线图

本文基于上述思路，引入重构方法、智能优化算法，研究基于“分解-重构-集成”范式的预测方法，将其应用于美元兑人民币汇率预测、比特币价格预测、原油期货价格预测。本文的技术路线如图 1.1 所示。

全文共包括 7 章，框架结构如图 1.2 所示，结构安排如下：

第 1 章为引言。

第 2 章为文献综述。

第 3 章介绍了本文所用到的研究方法。

第 4 章为基于分解重构集成的美元兑人民币汇率预测的实证研究。

第 5 章为基于二次分解集成的比特币价格预测的实证研究。

第 6 章为基于二次分解重构集成的原油期货价格预测的实证研究。

第 7 章为结论与展望。主要描述本文的结论、不足之处及未来的研究方向。

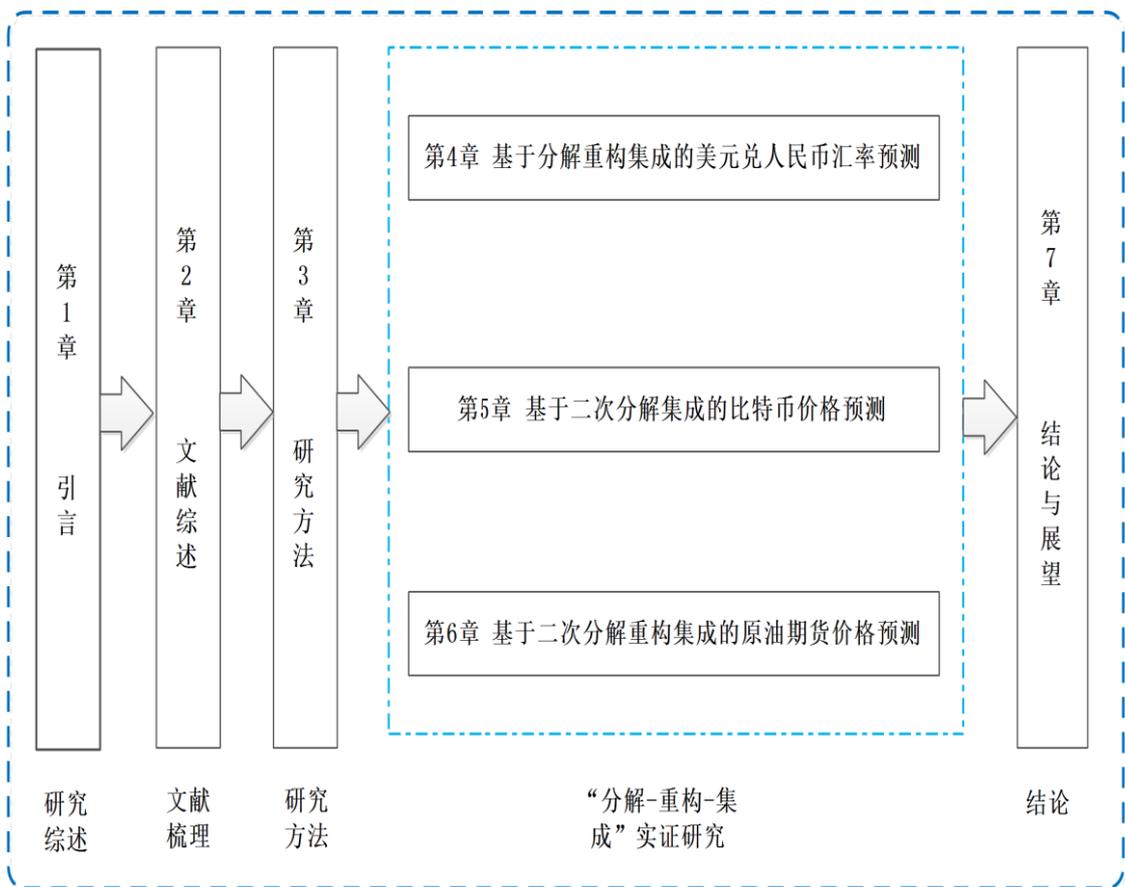


图 1.2 本文的框架结构

1.3 本文的创新点

本论文的创新点主要体现在以下四个方面：

第一，考虑到现有分解方法不能完全提取分解后的高频信号，故通过分解一次分解产生的高频信号（即二次分解）来降低信号的复杂性；

第二，在分量重构阶段，通过计算熵值（本文用到的熵值是 SE 和 PE）来减少子序列的个数，降低计算的复杂度；

第三，考虑到人工智能预测方法（本文用到的人工智能方法是 LSSVM 和 KELM）的参数对预测结果的影响，本文使用智能优化算法（本文用到的智能算法是 PSO 和 CSSA）对人工智能预测方法的参数进行优化；

第四，本文采用相空间重构（PSR）来确定输入变量的最佳输入维数，以此避免利用主观经验选择输入维数，进一步提高模型的预测性能。

2 文献综述

2.1 传统计量经济方法下的金融预测

传统的计量经济方法，是在统计学的基础上，遵循一定的经济计量规律，以相关经济变量为指标，代入数学方程式中计算的一种方法。常见的传统计量经济方法主要有 GARCH 模型、ARIMA 模型、误差修正模型（ECM）、残差自回归模型、随机游走模型（RW）、VAR 模型等。

基于现有的传统计量经济方法，国内外学者对上述方法在金融领域进行了实际的预测及应用。譬如，惠晓峰等运用 GARCH 模型，对汇率体制改革后的人民币兑美元汇率进行建模和预测，分别采用一步向前预测的滚动算法和递归算法，取得了不错的预测效果^[1]。戴晓枫等使用人民币兑美元的日汇率值进行实证研究，建立 ARIMA 模型和 EGARCH 模型进行预测，结果表明，EGARCH 模型的预测结果较 ARIMA 模型理想，可以较好的描述汇率的变动趋势^[2]。Alessandro 等使用 ECM 模型对石油价格进行了预测分析，将 ECM 的预测性能与未采用任何协整关系的单模型进行比较^[3]。邓美玲等使用残差自回归模型对国际石油价格进行短期预测，结果显示，与 ARIMA 模型及 GARCH 模型相比较，残差自回归模型优于其他模型^[4]。Atilim 等使用 RW 模型对石油价格进行了预测^[5]。侯璐利用 ARIMA 模型对石油价格进行了短期分析预测^[6]。秦鹏等使用广义指数预报因子模型对石油价格进行预测，实证结果显示，该模型较其他模型预测精度更高^[7]。Hassan 等使用 ARIMA-GARCH 模型对原油现货价格进行了预测研究，与传统的条件方差模型相比，发现可以更好地捕获石油收益的波动性^[8]。Aijun 等使用 GARCH 模型对油价收益率波动进行建模和预测，结果表明，GARCH 模型提高了油价收益波动率预测的准确性^[9]。Rout 等使用 ARMA-DE 预测模型对汇率进行预测，结果显示该模型具有较好的短期和长期预测潜力^[10]。韩晓茹等利用汇率改革之后的人民币兑美元汇率建立 ARIMA(14, 1, 0)模型，残差检验证明该模型是合理的^[11]。孙鹏等以美元兑人民币的中间价作为研究样本，将数据一阶差分后建立 ARIMA 模型，对美元兑人民币汇率进行短期预测^[12]。Bakar 等使用 ARIMA 模型来预测比特币的价格^[13]。Demir 等利用向量自回归模型发现经济政策不确定性指数对比特币收益具有预测能力^[14]。

总的来说，传统的计量经济方法主要采用函数形式来描述问题的定量特征，对符合其假设条件的数据，它具有较好的预测能力，但是它对研究样本有着较为严格的假设，在复杂系统的分析研究中，譬如非线性、非平稳性及复杂性数据处理中，难以做到运用数学公式对其进行量化处理。因此，在实际的预测过程中，不能仅仅依靠传统的计量经济方法，需要对传统的计量经济方法进行改进来提高预测精度。

2.2 人工智能方法下的金融预测

由于数据的非线性、非平稳性及复杂性高等特征，传统的计量经济方法的预测效果往往不理想。随着网络的发展，数据的获取更加便捷，计算机的运算能力不断增强，人工智能方法逐渐流行，可以较好的解决非线性数据的预测问题，因此人工智能方法被相应提出并得到广泛应用。

常见的人工智能方法主要有神经网络、支持向量机、Elman 动态递归神经网络、BP 神经网络、长短期记忆模型等等。在早期的研究中，主要以单一的人工智能方法为主，譬如，杨忻等运用神经网络对汇率进行预测，研究表明，与随机游走模型相比，神经网络用于汇率预测是有效的^[15]。文献^[16-17]通过神经网络对汇率进行预测，发现相比线性自回归和随机游走模型，神经网络具有更好的预测效果。Kulkarni 等使用多层前馈神经网络来预测原油现货价格，结果表明，该模型在短期内具有较好的预测效果^[18]。文献^[19-20]使用支持向量机来预测原油的价格，与 ARIMA 和 BP 相比较，实验结果表明，支持向量机的预测精度更优。杨新臣等利用支持向量机对日元、英镑和加元汇率数据进行回归和预测研究，实验结果表明，支持向量机具有较好的方向预测性，预测结果可作为投资决策的依据^[21]。慕晓茜等使用 SVR 模型预测美国原油期货价格，结果表明该模型能较好地拟合美原油期货价格的走势，且在预测短期价格方面取得理想的效果^[22]。何树红等使用 Elman 动态递归神经网络预测石油价格的走势，与 BP 神经网络相比，预测精度进一步提高^[23]。金艳凤利用附加动量的 BP 神经网络对人民币兑美元汇率进行预测，发现相比 BP 神经网络，附加动量的 BP 算法具有更优的预测效果^[24]。李秋敏等使用改进的 LSSVM 对汇率时间序列进行了预测，实证结果表明，改进模型的预测精度得到了明显的提升^[25]。Chiroma 等使用遗传算法

和神经网络的组合模型来预测原油价格，并将预测结果与 BP 对比，结果表明提出的预测模型在预测精度和计算效率方面都更好^[26]。Greaves 等通过构建一个比特币交易网络，利用传统的机器学习方法来预测比特币价格^[27]。范秋枫等使用量子粒子群算法优化模糊神经网络系统的参数，将其应用于原油价格预测，结果表明该模型在预测精度方面取得理想的效果^[28]。Indera 等通过构造基于多层感知器的非线性自回归模型来预测比特币的价格^[29]。McNally 等使用长短期记忆模型来预测比特币的价格，实证结果表明，长短期记忆模型的预测效果优于 ARIMA 模型^[30]。Liu 等通过构建影响比特币价格决定因素的特征系统，采用叠加去噪自动编码器（SDAE）的深度学习方法对比特币价格进行预测^[31]。

由上可知，人工智能方法虽然在复杂系统的预测中显示出其独特的优势，但它们同样存在着自身的缺陷，譬如具有参数敏感性、过度拟合、容易陷入局部最优等，这使得单一的人工智能预测方法出现预测性能不佳的问题。

2.3 分解集成方法下的金融预测

由于数据具有复杂的非线性特征，它既有某种确定性规律，同时又表现出随机性特征，传统的计量经济方法和单一的人工智能方法已经不能再满足预测精度方面的需求。因此，为了进一步提高预测性能，国内外众多学者开始运用分解集成的组合预测方法进行预测研究。譬如，汪寿阳等于 2005 年提出了 TEI@I 方法论，显著地提高了油价的预测性能^[32]。谢赤等通过对汇率序列进行分解，然后利用 Elman、SVM 预测，实证结果表明，组合预测方法较单一模型方法能够取得更好的效果^[33-34]。杨云飞等使用 EMD 和 SVM 的组合预测方法对原油现货价格进行预测，结果表明，与单一的 SVM 和 ANN 相比，该组合模型具有较高的预测精度^[35]。Rania 等使用小波分解和马尔可夫切换向量自回归对原油价格进行预测，并探讨了原油冲击对英国、法国和日本的股票市场收益的影响^[36]。李成等利用 EMD 分解对原油价格序列分解得到多个 IMF，对于每个 IMF 进行变参数的前向神经网络训练，将每个 IMF 下的预测结果集成，从而得到预测的原油价格。实证结果表明，相比已有的 EMD 和神经网络的预测方法，所提出的方法预测效果有一定的改善^[37]。戴伟通过使用 EEMD 和 EELM 的分解集成模型对油价进行预测，相较于单一预测模型，所提出的模型能够显著提升预测模型的预测性能

[38]。王岚使用 EMD 分解和 GA-KELM 对国际原油价格进行预测分析, 实证结果表明, 采用 EMD-GA-KELM 模型的预测效果要远远好于采用 GA-KELM 模型的预测效果[39]。Yu 等利用 EEMD 的数据分解功能简化了复杂数据, 并采用有效稳定的 EELM 来保证预测精度, 结果显示, 该模型在高精度、省时和鲁棒性方面显示出优越的性能[40]。He 等提出使用变分模态分解 (VMD) 进行序列分解, 采用均方误差和误差熵准则进行参数优化, 实证结果显示, 与基准模型相比, 提出的组合模型展示出优越的性能[41]。陈焕采用稀疏贝叶斯学习 (SBL) 模型分析原油价格的重大影响因素, 并提出使用自适应白噪声的完整经验模态分解 (CEEMDAN) 和灰狼优化算法 (GWO) 优化核极限学习机 (KELM) 参数的预测模型来预测原油价格[42]。高海翔使用一种基于 EEMD 分解、小波阈值去噪、fine-to-coarse 法重构和 LSTM 神经网络的组合预测方法来预测国际石油价格[43]。梁少东提出了基于带比例约束分解集成和基于偏差-方差-复杂性均衡分解的新的分解集成方法对国际油价进行预测, 实证比较结果表明, 所提出的模型具有较好的预测性能[44]。崔金鑫等使用 CEEMDAN 分解算法和极限学习机模型, 利用粒子群优化算法对机器学习模型进行参数寻优, 进一步提升了原油期货价格预测的精准性[45]。Gyamerah 使用 VMD 分解和广义加法的集成模型(GAM)来预测当日的比特币价格, 并将其与 EMD 和 GAM 集成在一起进行比较, 实证结果显示, VMD-GAM 模型的预测效果要优于 EMD-GAM 模型[46]。杨静凌等使用互补集合经验模态分解 (CEEMD) 和 Elman-Adaboost 神经网络的组合模型对 Brent 原油价格序列进行预测, 实证结果表明, 相比其他模型, 所提模型预测精度更高[47]。

综上所述, 国内外学者在金融预测领域中没有考虑分解后的高频序列, 这些高频数据具有较高的复杂度, 不利于预测数据未来的变化趋势。因此, 针对以上问题, 本文在已有研究的基础上, 对分解集成范式进行改进, 构建基于“分解-重构-集成”范式的预测方法。

3 研究方法

3.1 分解方法

3.1.1 EEMD 分解方法

经验模态分解 (EMD) 方法由 Huang 等^[48]于 1998 年提出, 它是一种自适应的数据处理方法, 主要适用于处理非线性、非平稳时间序列, 该方法可以将原始数据分解为多个 *IMF* 分量和一个残余项 (Res)。分解后的每个 *IMF* 分量一般满足以下两个条件:

(1) 对于分解得到的 *IMF* 分量, 极值点 (包括极大值和极小值) 的个数和零点的个数相等或最多相差一个。

(2) 极大值点和极小值点所确定的上下包络线的均值为零。

集合经验模态 (EEMD) 分解方法通过在数据中加入白噪声来改变信号中极大值和极小值的位置, 并利用噪声的特性有效抑制 EMD 分解方法产生的模态混叠问题^[49]。其原理是对加入高斯白噪声的信号进行多次 EMD 分解并进行算术平均, 利用高斯白噪声的随机性来降低相应 *IMF* 分量的混叠程度。该方法具体的分解步骤如下^[49]:

(1) 在原始信号 $x(t)$ 中加入高斯白噪声信号 $m(t)$, 得到新信号 $y(t)$:

$$y(t) = m(t) + x(t)$$

(2) 对信号 $y(t)$ 按照 EMD 分解方法进行分解, 得到 *IMF* 分量和残余项:

$$y(t) = \sum_{j=1}^n c_j(t) + r(t)$$

式中, $c_j(t)$ 为第 j 个 *IMF* 分量, n 为 *IMF* 分量的个数, $r(t)$ 为分解后的残余项。

(3) 重复步骤 (1) 和 (2), 每次加入不同的白噪声, 就得到相应的 *IMF* 分量和残余项, 当添加 M 次时, 就会得到 M 组分解结果。

(4) 利用高斯白噪声频谱的均值为零的原理, 对 M 组的各 *IMF* 分量和残差

项取平均值：

$$c_j = \frac{1}{M} \sum_{i=1}^M c_{ij}$$

$$r = \frac{1}{M} \sum_{i=1}^M r_i$$

其中， M 表示添加白噪声序列的数目。

3.1.2 CEEMDAN 分解方法

CEEMDAN 分解方法^[50]是在 EMD 分解方法^[48]和 EEMD 分解方法^[49]的基础上进一步发展和改进而来的。该方法一方面克服了 EMD 分解时引起的模态混叠问题，另一方面解决了 EEMD 在分解重构序列中残留噪声问题。

假设比特币的原始价格序列为 $x(n)$ ，第 $t(t=1,2,\dots,T)$ 次加入的白噪声序列记为 $\lambda^t(n)$ ，则第 t 次分解时，比特币价格序列可表示为 $x^t(n) = x(n) + \lambda^t(n)$ 。将第 j 个由 EMD 分解和 CEEMDAN 分解产生的模态分量分别用 $E_j(\cdot)$ 和 IMF_j 表示。

参照已有研究^[50]，CEEMDAN 分解算法的步骤如下：

步骤 1 比特币价格的原始序列采用 EMD 分解方法进行分解，获得第一个 IMF 分量以及余量：

$$IMF_1(n) = \frac{1}{T} \sum_{t=1}^T IMF_1^t(n) = \overline{IMF_1}(n)$$

$$r_1(n) = x(n) - \overline{IMF_1}(n)$$

步骤 2 进行 t 次实验，在每次实验的过程中，把正负成对的白噪声信号加入到 $r_1(n)$ 中，同时对新信号 $r_1(n) + \varepsilon_1 E_1(\lambda^t(n))$ 进行分解（其中 ε_1 为噪声标准差），当得到第一个 EMD 模态分量时停止分解。在此基础上，使用以下公式计算第二个 IMF 分量：

$$IMF_2(n) = \frac{1}{T} \sum_{t=1}^T E_1(r_1(n) + \varepsilon_1 E_1(\lambda^t(n)))$$

步骤 3 对其余每个阶段，即 $j=2,3,\dots,J$ ，计算第 j 个余量信号，其计算方法与步骤 2 相同。然后将正、负成对的白噪声信号添加到 $r_j(n)$ 中，可得到第 $j+1$ 个模态分量：

$$r_j(n) = r_{j-1}(n) - IMF_j(n)$$

$$IMF_{(j+1)} = \frac{1}{T} \sum_{t=1}^T E_1(r_j(n)) + \varepsilon_j E_j(\lambda^t(n))$$

步骤 4 重复步骤 3，直到剩余分量 $r_j(n)$ 无法再分解时为止，获得 J 个 IMF 分量。因此，最终的余量信号由以下公式表示：

$$R(n) = x(n) - \sum_{j=1}^J IMF_j$$

因此，原始比特币价格序列 $x(n)$ 最终被分解为：

$$x(n) = \sum_{j=1}^J IMF_j + R(n)$$

在本文 CEEMDAN 分解算法中，添加噪声的标准差设置为 0.2，最大迭代次数设置为 1000。

3.1.3 ICEEMDAN 方法

ICEEMDAN 方法是 2014 年 Colominas 等^[51]对 CEEMDAN 方法进行改进的一种新算法。CEEMDAN 方法在分解过程中加入的是高斯白噪声，而 ICEEMDAN 方法是在 EMD 分解过后加入高斯白噪声，它主要是计算每个模态分量信号加噪声的局部均值，并将分解得到的分量定义为残差信号与局部均值的差值。ICEEMDAN 分解方法改善了传统方法在分解重构的早期阶段容易出现虚假分量和模态混叠的情况，实验表明，ICEEMDAN 方法解决了 CEEMDAN 方法在本征模态函数中可能存在的虚假模式与残余噪声的问题，从而有效提高了降噪效果。

定义操作符 $E_k(\cdot)$ 表示 EMD 分解后的第 k 个 IMF 分量，定义操作符 $M(\cdot)$ 表示信号的局部均值。具体分解过程如下^[51]：

- (1) 向原始信号 x 中添加噪声分量：

$$x^{(i)} = x + \sigma_0 E_1(w^{(i)})$$

其中, i 是迭代次数 ($i=1,2,\dots,N$), $w^{(i)}$ 表示添加的第 i 个白噪声, σ_0 表示噪声标准差。得到第一个残差分量:

$$r_1 = \frac{1}{N} \sum_{i=1}^N M(x^{(i)})$$

(2) 计算第一模态分量 IMF_1 值 \tilde{d}_1 : $\tilde{d}_1 = x - r_1$;

(3) 计算第二模态分量 IMF_2 值 \tilde{d}_2 :

$$\tilde{d}_2 = r_1 - r_2 = r_1 - \frac{1}{N} \sum_{i=1}^N M(r_1 + \sigma_1 E_2(w^{(i)}))$$

(4) 计算第 k 个模态分量 IMF_k 的值 \tilde{d}_k : $\tilde{d}_k = r_{k-1} - r_k$

式中, $r_k = \frac{1}{N} \sum_{i=1}^N M(r_{k-1} + \sigma_{k-1} E_k(w^{(i)})), k=2,3,\dots,K$ 。

3.2 重构方法

3.2.1 样本熵

样本熵^[52]由 Richman 和 Moorman 提出, 它可以度量时间序列的复杂性^[53], 序列差异性越小, 其样本熵值越小; 反之, 序列差异性越大, 其样本熵值也越大。样本熵作为一种衡量数据复杂性的方法, 在时间序列预测方面得到了广泛应用。

为定义样本熵, 假设一个由 N 个数据组成的序列: $x = \{x(1), x(2), \dots, x(N)\}$, 则样本熵的计算步骤如下^[52]:

步骤 1 按照序号形成一组 m 维的向量组 $X = \{x_m(1), x_m(2), \dots, x_m(N-m+1)\}$, 其中, $x_m(i) = \{x(i), x(i+1), \dots, x(i+m-1)\}, 1 \leq i \leq N-m+1$ 。

步骤 2 计算向量 $x_m(i)$ 和向量 $x_m(j)$ 之间最大差值的绝对值 $d[x_m(i), x_m(j)]$:

$$d[x_m(i), x_m(j)] = \max_{k=0,1,\dots,m-1} (|x(i+k) - x(j+k)|)$$

步骤 3 对于给定的 $x_m(i)$, 计算 $x_m(i)$ 与其它向量 $x_m(j)$ 之间的距离小于等

于 r 的向量数目，这个数目用 B_i 表示，然后计算 B_i 与其它向量总数 $N-m+1$ 的比值 $B_i^m(r)$ ：

$$B_i^m(r) = \frac{B_i}{N-m+1}$$

步骤 4 计算 $N-m+1$ 个 $B_i^m(r)$ 的平均值 $B^m(r)$ ：

$$B^m(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} B_i^m(r)$$

步骤 5 将维数由 m 维增加为 $m+1$ 维，对 $m+1$ 维向量重复以上步骤，得到 $B_i^{m+1}(r)$ 。

步骤 6 该信号序列的样本熵值为：

$$SampEn(m, r, N) = -\ln \left[\frac{B^{(m+1)}(r)}{B^{(m)}(r)} \right]$$

式中称 m 为重构维数； r 为相似容限； N 为数据长度。由于参数 m 和 r 的取值会对结果产生影响，本文中选择 $m=2$ ， $r=0.15\sqrt{\text{var}(x)}$ （借鉴文献[54]，对参数这样设置）。

3.2.2 排列熵

排列熵^[55]由 Bandt 和 Pompe 提出，它主要用于检测时间序列的随机性，可以有效放大时间序列的微弱变化。它不仅具有算法简单、抗干扰能力强等优点，而且具有很好的鲁棒性^[56]。因此，通过利用排列熵可以来度量各子模态分量的复杂性。

PE 算法的具体步骤如下：

步骤 1 一个时间序列 $\{x(i), i=1, 2, \dots, N\}$ ，对其在相空间中重构后得到重构向量：

$$Y_s = [x(s), x(s+\tau), \dots, x(s+(m-1)\tau)] \quad (3.1)$$

公式(3.1)中， $s=1, 2, \dots, N-(m-1)\tau$ ， τ 代表延迟时间， m 代表嵌入维数（本文中 τ 和 m 由相空间重构函数确定）。

步骤 2 将得到的每个重构向量 Y_g 的分量从小到大排列，可对应得到一个符号序列。有些重构向量的符号序列可能会相同，我们将不同的符号序列向量记为：

$$Y(g) = (s_1, s_2, \dots, s_m), g = 1, 2, \dots, l \text{ 且 } l \leq m! \quad (3.2)$$

步骤 3 计算每种符号序列 $Y(g)$ 出现的概率 P_g ：

$$P_g = \frac{N(Y(g))}{N - (m-1)\tau} \quad (3.3)$$

公式(3.3)中， $N(Y(g))$ 为 $Y(g)$ 的个数，显然有 $\sum_{g=1}^l P_g = 1$ 。

步骤 4 将该时间序列的 PE 定义为：

$$H_p(m) = -\sum_{g=1}^l P_g \ln P_g \quad (3.4)$$

由公式(3.4)可看出，PE 作为一种度量数据复杂性的方法，其取值大小反映序列的复杂程度，PE 值越大，数据的复杂性越高；反之，PE 值越小，数据的复杂性越小。因此，可以计算时间序列分解后各个 *IMF* 分量的 PE 值，以此来评估每个子序列的复杂度。

3.3 优化方法

3.3.1 粒子群算法

PSO 算法是一种基于群体智能的进化算法，它主要通过群体和个体之间的协作和信息共享来寻找最优解^[57]。PSO 算法源于对鸟类觅食行为的研究，鸟类觅食时，每只鸟寻找食物最简单、最有效的方法是搜索离食物最近的鸟的周围区域。

假设在一个 D 维的搜索空间中，每个优化问题的可能解都被看作是空间中的一个“粒子”，种群 $x = (x_1, x_2, \dots, x_n)$ 由 n 个粒子组成，定义 $x_i = (x_{i1}, x_{i2}, \dots, x_{iD})^T$ 为粒子 i 当前的位置， $v_i = (v_{i1}, v_{i2}, \dots, v_{iD})^T$ 为粒子 i 当前的飞行速度， $p_i = (p_{i1}, p_{i2}, \dots, p_{iD})^T$ 为粒子 i 的个体极值， $p_g = (p_{g1}, p_{g2}, \dots, p_{gD})^T$ 为全局极值。

PSO 算法的具体步骤^[57]：

(1)初始化参数。在本文中，种群规模设置为 20；最大迭代次数为 50 次；学习因子 $c_1 = 1.7, c_2 = 1.8$ ；搜索变量的维度 $D = 2$ （本文优化了 LSSVM 的 2 个参数：正则化参数和核参数）。

(2)计算粒子的适应度值。本文采用 k 折交叉验证的方法对训练样本进行划分，以 k 次均方根误差 $RMSE_i$ 的平均值作为适应度函数，可有效解决模型的过拟合问题，根据样本数据的大小，取 k 的值为 5。则 $RMSE_{k-cv} = \frac{1}{5} \sum_{i=1}^5 RMSE_i$ ，其中

$$RMSE_i = \sqrt{\frac{\sum_{t=1}^{n/5} (y_t - \hat{y}_t)^2}{n/5}}$$

y_t, \hat{y}_t 分别为真实值和预测值， n 表示训练集样本容量的大小。

(3)根据适应度更新个体极值和群体极值。对当前每个微粒的适应度值和它最好的局部位置、全局位置相比较，选取两者之间更优的局部位置和全局位置。

(4)更新速度和位置。根据以下更新公式进行更新：

$$v_{id}^{k+1} = wv_{id}^k + c_1 r_1 (p_{id}^k - x_{id}^k) + c_2 r_2 (p_{gd}^k - x_{id}^k)$$

$$x_{id}^{k+1} = x_{id}^k + v_{id}^{k+1}$$

式中， w 是惯性权重； $d = 1, 2, \dots, D; i = 1, 2, \dots, n$ ； k 是当前迭代次数； v_{id} 是粒子的速度； x_{id} 是粒子的位置； r_1 和 r_2 是 0 和 1 之间的随机数； p_{id} 和 p_{gd} 分别表示个体极值最优和全局最优。

(5)判断是否满足最大迭代次数或预测精度。如果满足，则停止迭代，输出结果；否则，重复步骤(3)和步骤(4)，直到满足设定的最大迭代次数为止。

3.3.2 混沌麻雀搜索算法

麻雀搜索算法 (Sparrow Search Algorithm, SSA) 是 2020 年由薛建凯等^[58]提出的一种新型群体智能优化算法。麻雀作为一种群居的鸟类，在觅食的过程中，群体内部有着明显的分工，主要存在着 2 种行为：发现者负责寻找食物；加入者跟随发现者觅食，其中部分加入者作为侦察者，决定种群是否放弃食物。

假设在 D 维的搜索空间中有 n 只麻雀，则第 i 只麻雀在此空间中的位置为 $X_i = [x_{i1}, \dots, x_{id}, \dots, x_{iD}]$ ，其中 $i = 1, \dots, n$ ， x_{id} 是第 i 只麻雀在第 d 维空间中的位置。发现者位置更新如下：

$$x_{id}^{t+1} = \begin{cases} x_{id}^t \cdot \exp\left(-i/\alpha \cdot T\right), R_2 < ST \\ x_{id}^t + Q \cdot L, R_2 \geq ST \end{cases} \quad (3.5)$$

公式 (3.5) 中， t 表示当前的迭代次数； T 表示最大的迭代次数； α 是介于 $(0,1]$ 之间的随机数； Q 是服从 $N(0,1)$ 的随机变量； L 表示大小为 $1 \times d$ ，元素均为 1 的矩阵； $R_2 \in [0,1]$ 和 $ST \in [0.5,1]$ 分别表示预警值和安全值。当 $R_2 < ST$ 时，负责侦察的麻雀没有发现捕食者，发现者没有被捕食的危险；当 $R_2 \geq ST$ 时，负责侦察麻雀发现捕食者，并立即向正在觅食的麻雀发出信号，种群撤离危险区域。

加入者位置更新如下：

$$x_{id}^{t+1} = \begin{cases} Q \cdot \exp\left(\frac{(x\omega_d^t - x_{id}^t)}{i^2}\right), i > \frac{n}{2} \\ xb_d^{t+1} + |x_{id}^t - xb_d^{t+1}| A^+ \cdot L, otherwise \end{cases} \quad (3.6)$$

公式 (3.6) 中， $A^+ = A^T (AA^T)^{-1}$ ， $x\omega_d^t$ 表示第 t 次迭代时麻雀在第 d 维空间的最坏位置； xb_d^{t+1} 表示第 $t+1$ 次迭代时麻雀在第 d 维空间的最优位置；当 $i > \frac{n}{2}$ 时，第 i 个加入者在当前位置不会获得食物，而是会在其他地方寻找食物；当 $i \leq \frac{n}{2}$ 时，第 i 个加入者会在当前的最优位置 xb 附近觅食。

当意识到危险时，麻雀种群会进行反捕食行为：

$$x_{id}^{t+1} = \begin{cases} xb_d^t + \beta(x_{id}^t - xb_d^t), f_i \neq f_g \\ x_{id}^t + K \left(\frac{x_{id}^t - x\omega_d^t}{|f_i - f_\omega| + e} \right), f_i = f_g \end{cases} \quad (3.7)$$

公式 (3.7) 中： β 代表步长； K 是 $[-1,1]$ 之间的随机数，代表麻雀的移动方向； e 是一个很小的常数； f_i 是第 i 只麻雀的适应度值， f_g 和 f_ω 分别表示麻雀的最优和最差适应度值。当 $f_i \neq f_g$ 时，表示麻雀远离种群中心，容易被捕食者攻击；当 $f_i = f_g$ 时，表示麻雀离种群中心较近，通过靠近其他麻雀来避免被捕食者攻击。

为了避免 SSA 陷入局部最优,混沌麻雀搜索算法 (CSSA) 在 SSA 的基础上加入了混沌扰动和高斯变异,以此来提高算法的搜索性能^[59]。当一次迭代完成后,重新计算每只麻雀的适应度值 f_i 和麻雀种群的平均适应度值 f_{avg} , 具体如下^[59]:

(1) 当 $f_i < f_{avg}$ 时,对个体 i 进行高斯变异。若变异后个体更优,则使用变异后的个体来代替变异前的个体,否则,使用变异前的个体。高斯变异公式如下:

$$x' = x(1 + N(0,1)) \quad (3.8)$$

其中, x 为原来的个体, $N(0,1)$ 是服从标准正态分布的随机数, x' 是高斯变异后的数值。

(2) 当 $f_i \geq f_{avg}$ 时,对个体 i 进行混沌扰动。若扰动后个体更优,则用扰动后的个体代替扰动前的个体,否则,使用扰动前的个体。混沌扰动的步骤如下^[59]:

$$\textcircled{1} \text{生成混沌变量 } Z_d: Z_{d+1} = (2Z_d) \bmod 1 + \text{rand}(0,1) * \frac{1}{NT}$$

其中, NT 是混沌序列中的粒子个数, $\text{rand}(0,1)$ 是随机数;

$$\textcircled{2} \text{将混沌变量放入待解问题的解空间: } \text{new}X_d = \min_d + (\max_d - \min_d) * Z_d$$

其中, \min_d 和 \max_d 是 d 维空间变量 $\text{new}X_d$ 的最小值和最大值;

$$\textcircled{3} \text{对个体进行混沌扰动: } \text{new}X' = (X' + \text{new}X) / 2$$

其中, X' 是混沌扰动的个体; $\text{new}X$ 是产生的混沌扰动量; $\text{new}X'$ 是混沌扰动后的个体。

3.4 输入变量的确定方法——相空间重构

在单一时序建立机器学习预测模型的过程中,时间序列中滞后变量的输入个数对模型的预测效果具有重要的影响。因此,通过采用相空间重构函数来确定滞后输入变量的个数,进而确定嵌入维数,其原理如下:

对于一个时间序列 $\{x_1, x_2, \dots, x_n\}$, 若已知延迟时间 t 和嵌入维数 h , 则重构的相空间可描述为:

$$Y_i = [x_i, x_{i+t}, \dots, x_{i+(h-1)t}], i = 1, 2, \dots, G$$

其中, G 是相空间的数据点数, 定义为: $G = n - (h-1)t$ 。在本文中, 采用 Kim 等^[60]提出的 $C-C$ 方法计算 h 值和 t 值, 计算步骤如下^[60]:

步骤 1 对于时间序列 $\{x_1, x_2, \dots, x_n\}$, 其关联积分定义为:

$$C_s^h(h, G, f, t) = \frac{2}{h(h-1)} \sum_{1 \leq i < j \leq G} \theta(f - d_{ij}) \quad (3.9)$$

其中, $f > 0, d_{ij} = \|Y_i - Y_j\|, \theta(z) = \begin{cases} 0, & z < 0 \\ 1, & z \geq 0 \end{cases}$ 。

步骤 2 定义检测统计量为:

$$S(h, G, f, t) = C_s^h(h, G, f, t) - C_s^h(1, G, f, t) \quad (3.10)$$

步骤 3 将时间序列 $\{x_1, x_2, \dots, x_n\}$ 划分为 k 个不相交的子序列, 此时检测统计量如下:

$$S(h, G, f, k) = \frac{1}{k} \sum_{s=1}^k \left[C_s^h\left(h, \frac{G}{k}, f, k\right) - C_s^h\left(1, \frac{G}{k}, f, k\right) \right] \quad (3.11)$$

当 $G \rightarrow \infty$ 时, 检测统计量为:

$$S(h, f, k) = \frac{1}{k} \sum_{s=1}^k \left[C_s^h(h, f, k) - C_s^h(1, f, k) \right], (h = 2, 3, \dots) \quad (3.12)$$

步骤 4 选择最大值和最小值对应的两个半径 f , 并将最大值和最小值之间的差值定义为:

$$\Delta S(h, k) = \max_f \{S(h, f, k)\} - \min_f \{S(h, f, k)\} \quad (3.13)$$

步骤 5 根据 $3-\sigma$ 准则和公式 (3.11) 可得:

$$\begin{aligned} \bar{S}(k) &= \frac{1}{16} \sum_{h=2}^5 \sum_{j=1}^4 S(h, f_j, k) \\ \Delta \bar{S}(k) &= \frac{1}{4} \sum_{h=2}^5 \Delta S(h, k) \\ S_{cor}(k) &= \Delta \bar{S}(k) + |\bar{S}(k)| \end{aligned} \quad (3.14)$$

在公式 (3.14) 中, 当 $\bar{S}(k)$ 的第一个零点或 $\Delta \bar{S}(k)$ 取第一个最小值时, 最优延迟时间可用 $C-C$ 方法求得, 用 $S_{cor}(k)$ 的全局最小值可求得时间窗口 w 的值, 嵌入维数 h 可由公式 $w = (h-1)t$ 得到。

3.5 预测方法

3.5.1 最小二乘支持向量机模型

最小二乘支持向量机 (LSSVM) 由 Suykens 等^[61]提出, 它是一种基于 SVM 的改进算法。与传统的 SVM 相比, LSSVM 在目标函数中引入了一个平方项, 通过等式约束代替不等式约束, 将二次规划问题转换为线性方程组求解, 它不需要逼近指定的精度, 因此计算的精度和速度极大提高。

LSSVM 的具体过程^[61]如下:

对于给定的一组向量, $\{(x_i, y_i), i=1, 2, \dots, n\}$, $x_i \in \mathbf{R}^n$ 为第 i 个样本的输入值, y_i 为第 i 个样本的输出值, n 为样本数。我们选择一个非线性映射 $\Phi(x)$, 通过将原始样本映射到高维空间, 来构造最优决策函数:

$$f(x) = w^T \Phi(x) + b$$

其中, w 为权值向量, b 为偏置量。

根据结构风险最小化原则, 通过引入松弛变量 e_i , 寻找使得 $J(w, e)$ 最小的 w 和 b 值, 可以得到如下 LSSVM 模型:

$$\min J(w, e) = \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{i=1}^n e_i^2$$

约束条件为:

$$y_i = w^T \Phi(x_i) + b + e_i, i=1, 2, \dots, n$$

式中, γ 为正则化参数, e_i 为误差。为了求解该优化问题, 引入了 *Lagrange* 函数:

$$L(w, b, e, \alpha) = \frac{1}{2} \|w\|^2 + \frac{\gamma}{2} \sum_{i=1}^n e_i^2 - \sum_{i=1}^n \alpha_i (w^T \Phi(x_i) + b + e_i - y_i)$$

其中, $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_n]^T$ 为 *Lagrange* 乘子。

对上式求其偏导, 可得 LSSVM 的决策函数为:

$$f(x) = \sum_{i=1}^n \alpha_i K(x, x_i) + b$$

式中, $K(x, x_i)$ 代表核函数。本文选取 RBF 函数作为核函数, 故

$K(x, x_i) = \exp\left(-\frac{x-x_i}{2\sigma^2}\right)$, 其中, σ 是其核密度。

由于正则化参数和核参数的选择决定着 LSSVM 模型的预测效果, 因此, 本文采用 PSO 算法对这两个参数进行优化。

3.5.2 极限学习机

极限学习机 (Extreme Learning Machine, ELM) 最早由 Huang 等^[62]提出, 是一种用于求解单隐含层前馈神经网络的算法。该算法只需要设置隐含层神经元的数目及激活函数^[62]。原理如下:

对于数据集 $D = (x_i, y_i)$, 其中, $x_i, y_i \in R^n$, $i = 1, 2, \dots, n$, x_i 为输入数据, y_i 为输出值, 则含有 L 个隐层节点的 ELM 的输出函数可表示为:

$$f(x_i) = \sum_{l=1}^L \beta_l h_l(x_i) = h(x_i) \beta \quad (3.15)$$

其中, $h(x_i) = [h_1(x_i), h_2(x_i), \dots, h_L(x_i)]$ 为激活函数, $\beta = [\beta_1, \beta_2, \dots, \beta_L]^T$ 为隐含层与输出层的权值向量。

ELM 模型的目标函数为:

$$\begin{aligned} \min : L &= \frac{1}{2} \|\beta\|^2 + \frac{1}{2} C \sum_{i=1}^n \|\xi_i\|^2 \\ \text{s.t.} : h(x_i) \beta &= y_i^T - \xi_i^2 \end{aligned} \quad (3.16)$$

其中, $i = 1, 2, \dots, n$, ξ_i 为误差, C 为正则化参数。

根据 KKT 条件, 用 Lagrange 函数求解上述问题, 求其极值得:

$$\beta = H^T (HH^T + \frac{1}{C} I)^{-1} Y \quad (3.17)$$

其中, I 为单位矩阵, $H = [h^T(x_1), \dots, h^T(x_n)]$ 为隐含层输出矩阵。因此 ELM 的输出可以表示为:

$$f(x) = h(x) \beta = h(x) H^T (HH^T + \frac{1}{C} I)^{-1} Y \quad (3.18)$$

3.5.3 核极限学习机

KELM 是由 Huang 等^[63]在 ELM 基础上的扩展，它主要是在 ELM 中引入核原理，也就是说，隐含层的随机映射被核函数所取代。该方法不仅避免了 ELM 隐含层参数随机设定导致稳定性差的问题，而且提高了模型的鲁棒性^[63]。与 ELM 算法相比，KELM 算法具有更强的问题解决能力和更好的泛化性能。

从公式(3.18)可知，有矩阵内积存在，因此可以使用满足条件的核函数来代替矩阵内积运算，也就是说，核矩阵 $H^T H$ 的元素可表示为：

$$H^T H(i, j) = K(x_i, x_j)$$

$$h(x)H^T = [K(x, x_1), \dots, K(x, x_n)]^T$$

其中， $K(x_i, x_j)$ 是核函数。最终得到 KELM 的输出形式为：

$$f(x) = [K(x, x_1), \dots, K(x, x_n)]^T \left(\frac{1}{C} I + K(x_i, x_j) \right)^{-1} Y \quad (3.19)$$

在本文中，选取 RBF 函数作为核函数，即 $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$ ，其中，核参数 $\gamma > 0$ ，是外生的。为了提高预测精度，采用 CSSA 优化算法来确定 KELM 最佳的核参数。

3.6 评价指标

本文采用 MAE、MAPE、RMSE 和 DS 作为评价准则。其中，MAE、MAPE 和 RMSE 的值越小，表明预测值越接近真实值，模型的预测精度越高；DS 值越大，表明预测结果的趋势判断越准确，模型在预测方向上的准确性越高。本文采用的评价指标的具体内容如表 3.1 所示。

为了对各预测模型的差异性进行统计推断和稳健性判断，采用 DM 统计量^[64]进行检验。DM 检验的原假设为两个比较模型的预测性能没有显著差异，如果原假设被拒绝，则意味着两个比较模型的性能存在着显著差异^[64]。DM 统计量定义如下：

$$S_{DM} = \frac{\bar{g}}{(\hat{V}_g / N)^{1/2}}$$

其中, $\bar{g} = \frac{1}{N} \sum_{i=1}^N g_i$, $g_i = (x_i - \hat{x}_{A,t})^2 - (x_i - \hat{x}_{B,t})^2$, $\hat{V}_{\bar{g}} = r_0 + 2 \sum_{i=1}^{\infty} r_i$, $r_i = \text{cov}(g_t, g_{t-i}) \circ \hat{x}_{A,t}$

和 $\hat{x}_{B,t}$ 分别是 t 时刻模型 A 和模型 B 的预测值。

表 3.1 评价指标

评价指标	定义	公式
MAE	平均绝对误差	$MAE = \frac{1}{n} \sum_{i=1}^n y_i - \hat{y}_i $
MAPE	平均绝对百分比误差	$MAPE = \frac{1}{n} \sum_{i=1}^n \left \frac{y_i - \hat{y}_i}{y_i} \right \times 100\%$
RMSE	均方根误差	$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}$
DS	方向统计量	$DS = \frac{100}{n} \sum_{i=1}^n d_i, d_i = \begin{cases} 1, (y_{t+1} - y_t)(\hat{y}_{t+1} - y_t) \geq 0 \\ 0, otherwise \end{cases}$

注: y_i 代表真实值, \hat{y}_i 代表预测值。

3.7 本章小结

本章主要介绍了本文所用到的研究方法。它主要包括几种分解方法(EMD 分解、CEEMDAN 分解和 ICEEMDAN 分解), 重构方法(SE 和 PE), 优化方法(PSO 和 CSSA), 输入变量的确定方法(相空间重构), 预测方法(LSSVM 和 KELM)和本文中所用到的评价指标(MAE、MAPE、RMSE、DS 及 DM 检验)。

4 基于分解重构集成的美元兑人民币汇率预测

本章以美元兑人民币的日度汇率数据为研究对象。4.1 节介绍了基于 EEMD-SE-PSO-LSSVM 模型的构建框架；4.2 节介绍了本章的数据来源；4.3 节展示了美元兑人民币汇率的 EEMD 分解和重构过程；4.4 节描述了模型的预测结果、指标对比结果、DM 检验结果及对模型的分析总结；4.5 节对本章节的内容进行了小结。

4.1 基于 EEMD-SE-PSO-LSSVM 模型的框架构建

为了降低原始序列分解后的复杂度,合理确定时序数据输入变量的个数以及优化预测模型的参数,本章构建了 EEMD-SE-PSO-LSSVM 模型,该模型的预测流程如图 4.1 所示。

模型的主要步骤概括如下:

步骤 1 分解。使用 EEMD 方法将原始汇率数据分解为 N 个相对简单且有意义的 *IMF* 分量和一个残余项。

步骤 2 模态分量重构。使用样本熵对得到的 N 个模态分量及原始序列分别计算 SE 值,将 SE 值较小的子序列进行重构。

步骤 3 重构项预测。采用 PSO-LSSVM 方法对得到的各个重构项分别进行预测。

步骤 4 集成预测。采用 LSSVM 对各重构项的预测结果进行非线性集成,得最终的预测结果。

步骤 5 预测效果分析。通过与其他模型的评价指标进行对比,进而分析本章 EEMD-SE-PSO-LSSVM 模型的预测效果。

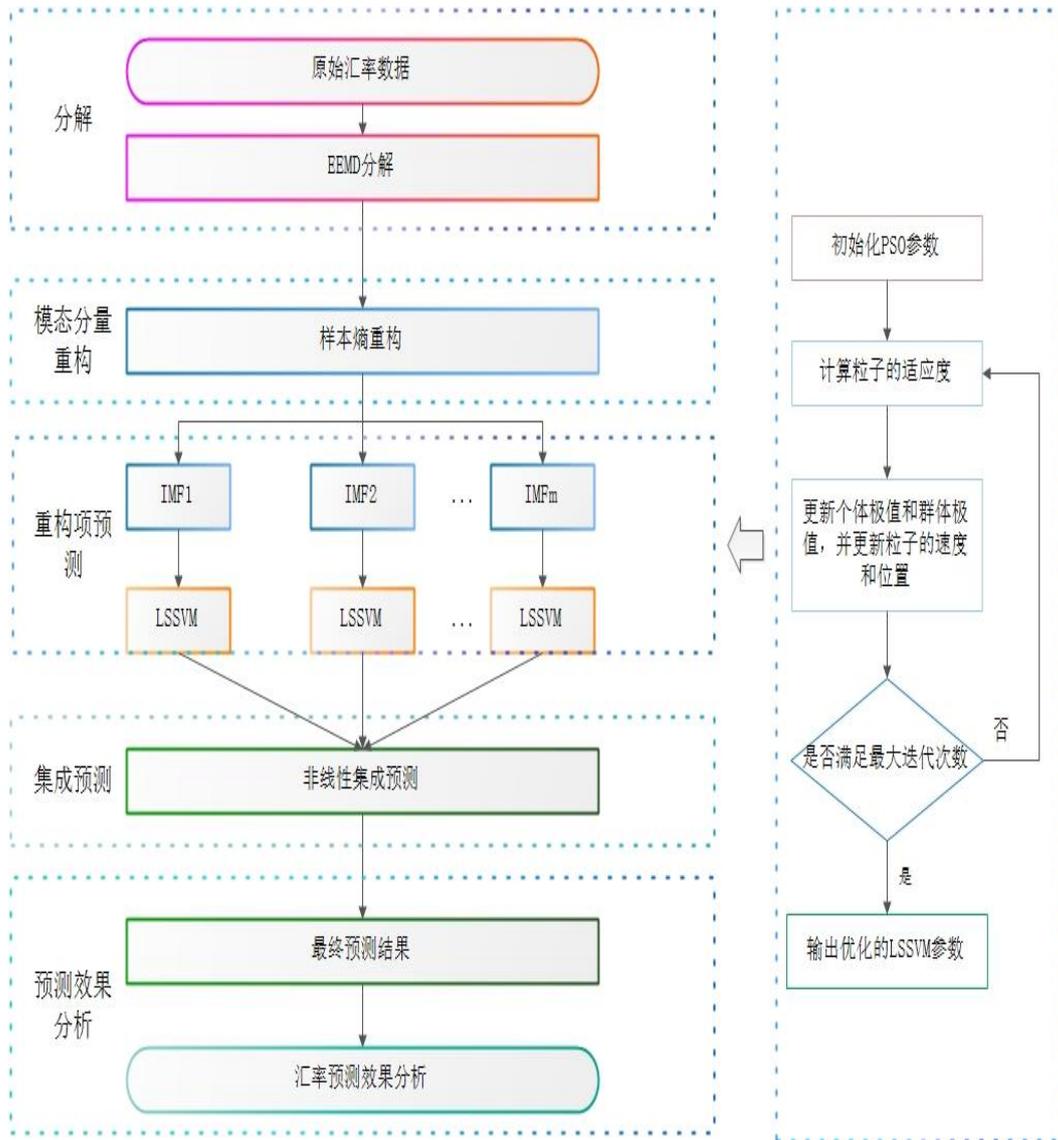


图 4.1 EEMD-SE-PSO-LSSVM 模型流程图

4.2 数据来源

在本章中，使用的美元兑人民币的汇率数据来源于 Wind 数据库。样本区间自 2015 年 8 月 11 日至 2020 年 8 月 10 日，共计 1218 个数据，具体如图 4.2 所示。将汇率数据划分为训练数据和测试数据，其中，自 2015 年 8 月 11 日至 2019 年 12 月 31 日（共 1072 个数据）为训练数据；自 2020 年 1 月 1 日至 2020 年 8 月 10 日（共 146 个数据）为测试数据。由于外汇市场在周末休市，周六周日汇率不变，所以本章不对周末的汇率做分析。本章节中数据的计算使用软件 MATLAB 2020a 完成。

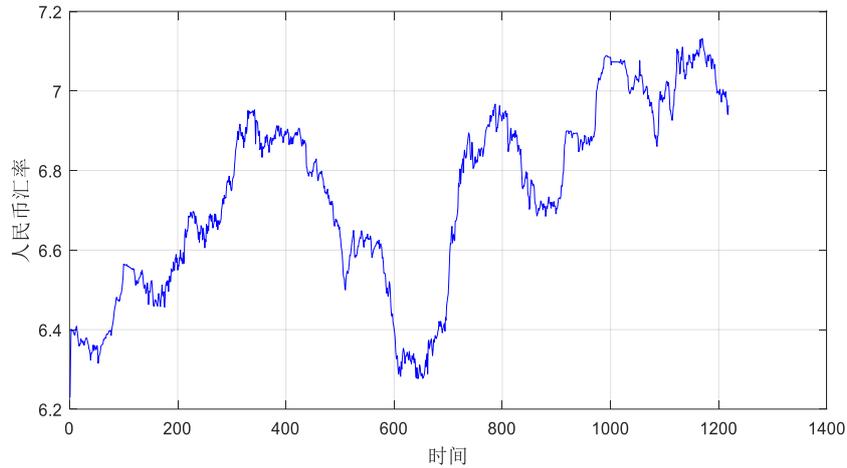


图 4.2 原始美元兑人民币汇率数据

4.3 美元兑人民币汇率的 EEMD 分解和重构

本节使用 EEMD 对原始美元兑人民币汇率序列进行分解，分解结果如图 4.3 所示。从图 4.3 中可以看出，原始的美元兑人民币汇率序列被分解为 10 个分量，依次是 $IMF1, IMF2, \dots, IMF9$ 和 Res ，各个 IMF 分量的频率从高到低依次排列，其中， $IMF1$ 的频率最高，该序列表现为无序，包含有大量噪声； $IMF2 - IMF9$ 的频率由高频逐渐向低频变化，序列中包含的噪声逐渐减少，复杂程度逐渐降低。

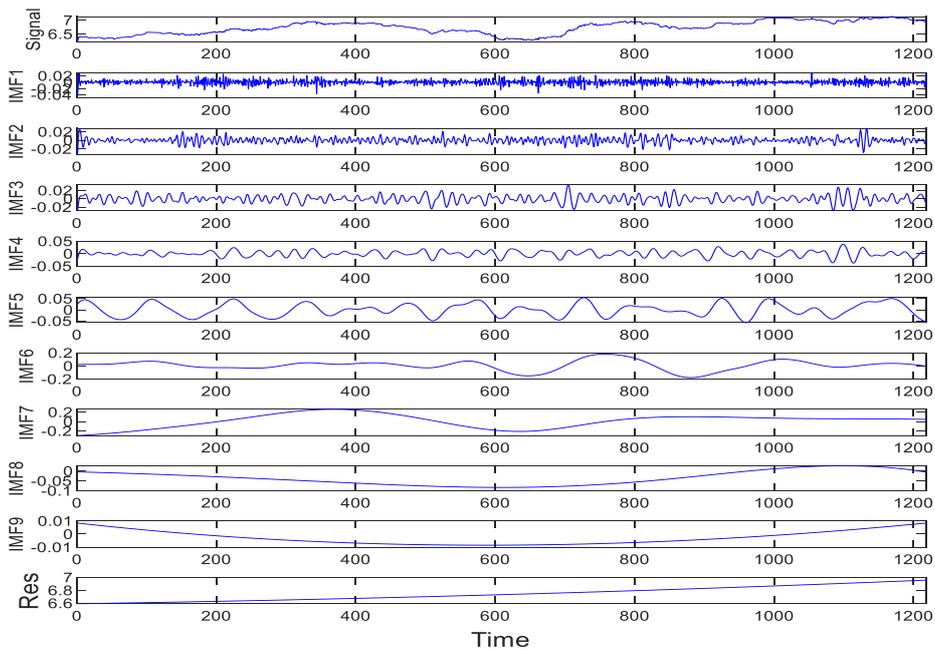


图 4.3 EEMD 分解原始汇率数据的结果

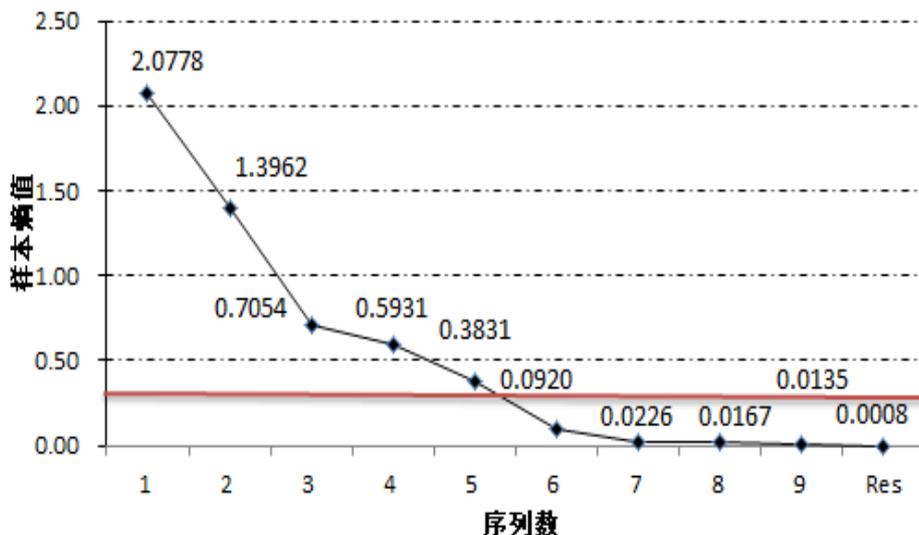


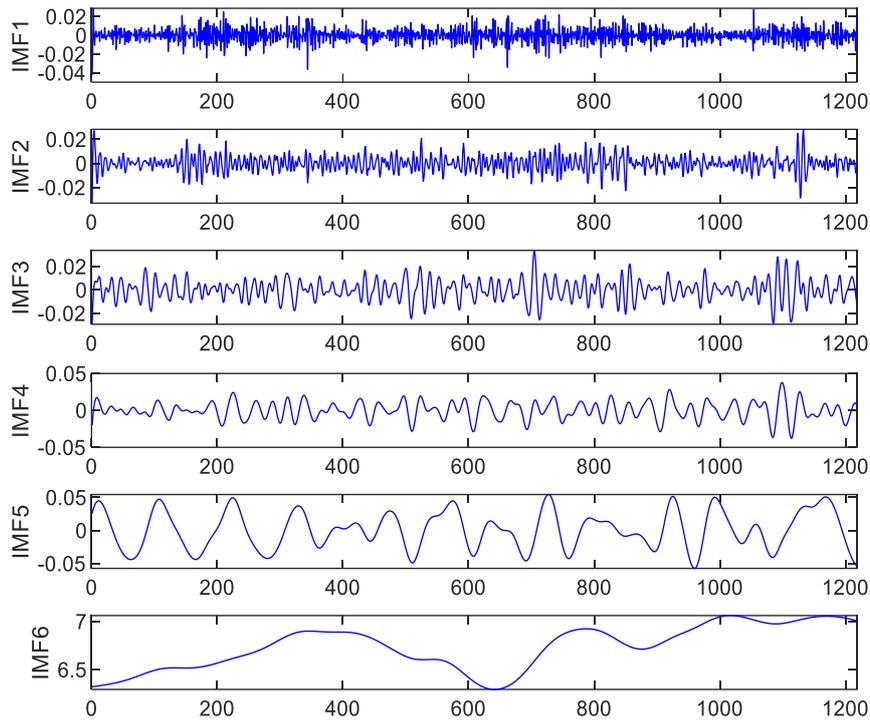
图 4.4 各子序列样本熵值

表 4.1 各 IMF 分量合并为新序列的结果

新序列	原序列
1	1
2	2
3	3
4	4
5	5
6	6,7,8,9,Res

从图 4.3 中可以看出，IMF 的分量比较多，如果使用 LSSVM 分别对每个 IMF 分量及 Res 进行建模，则每个模态分量的预测都会产生误差。模态分量越多，最终在结果集成时误差累计就会越大，因此，本章使用样本熵分别计算各个 IMF 分量及 Res 的 SE 值，以此来量化各个序列的复杂程度，并根据序列的复杂程度进行合并重构，各序列的 SE 值如图 4.4 所示。

各序列的重构规则如下：合并样本熵值小于原始序列的各序列；大于或等于原始序列的不进行合并。IMF1-IMF5 的 SE 值都大于原始序列（原始序列的 SE 值 0.1968），则对其不进行合并；IMF6-IMF9，Res 的 SE 值均小于原始序列，将其合并为一个序列。对 IMF 分量及 Res 重组后，结果如表 4.1 所示，最终重构后的子序列结果如图 4.5 所示。

图 4.5 样本熵重构后的 *IMF*

4.4 预测结果及评价

根据上一小节 4.3 所述，在预测重构后的各子序列时，通过相空间重构可以确定嵌入维数为 5，即用前 5 期的值作为输入来预测下一期的值。为了验证所构建方法的有效性，本章节选择了 LSSVM 和近年来文献中常使用的核极限学习机 (KELM) 对原始序列及重构后的各子序列进行预测，对各子序列的预测结果进行非线性集成时分别采用预测子序列时对应的方法进行集成。表 4.2 和表 4.3 中分别给出了本章模型 (EEMD-SE-PSO-LSSVM) 与其他 14 个对比模型的预测评价指标及 DM 检验结果。

从表 4.2 中可以看出，本章模型 (EEMD-SE-PSO-LSSVM) 在 MAPE, RMSE 和 DS 三个指标上均表现出最优的性能，从表 4.3 的 DM 检验中可以看出，对比模型的 p 值均接近于 0，这说明本章模型显著优于其他对比模型，进一步说明本章所构建模型的合理性及有效性。通过对表 4.2 中各预测模型指标的比较，可得如下结论：

表 4.2 各模型预测评价指标对比

模型	评价指标		
	MAPE	RMSE	DS
KELM	1.1602	0.0869	53.4247
LSSVM	3.4212	0.2498	53.4247
PSO-LSSVM	0.1769	0.0166	53.4247
EMD-KELM	0.9808	0.0782	53.4247
EEMD-KELM	0.8898	0.0663	52.7397
EMD-LSSVM	5.232	0.4187	53.4247
EEMD-LSSVM	1.1099	0.1054	68.4932
EMD-SE-KELM	0.9059	0.0732	51.3699
EEMD-SE-KELM	0.6657	0.0495	52.7397
EMD-SE-LSSVM	2.1306	0.1653	52.7397
EEMD-SE-LSSVM	1.3735	0.1024	52.7397
EMD-PSO-LSSVM	2.9479	0.224	54.1096
EEMD-PSO-LSSVM	1.8426	0.1468	55.4795
EMD-SE-PSO-LSSVM	0.3171	0.0301	65.0685
EEMD-SE-PSO-LSSVM	0.0846	0.0079	80.8219

(1) 采用分解方法的各预测模型（EMD-LSSVM，EEMD-LSSVM，EMD-KELM，EEMD-KELM）相比单一预测模型，我们可以发现，采用分解方法的预测模型总体上优于单一预测模型，而且采用 EEMD 分解方法的预测效果明显优于 EMD 方法，这主要是因为 EMD 受模态混叠现象的影响，EEMD 则避免了这一缺点，分解后获得的效果更好。

(2) 采用 PSO 优化后的 3 个模型，除了 EEMD-LSSVM 要优于 EEMD-PSO-LSSVM 之外，其他两个模型（PSO-LSSVM 和 EMD-PSO-LSSVM）的 MAPE，RMSE 和 DS 都优于优化前，优化后的水平精度和方向精度都有了明显的提高，这说明 PSO 可以比较理想地解决 LSSVM 模型参数的问题。

(3) 采用样本熵(SE)重构后的模型明显优于重构之前。重构后的 2 个 KELM 模型（EMD-SE-KELM 和 EEMD-SE-KELM），其 MAPE 值和 RMSE 值都优于重构前；重构后的 2 个 LSSVM 模型（EMD-SE-LSSVM 和 EEMD-SE-LSSVM），除了 EEMD-SE-LSSVM 不符合外，模型 EMD-SE-LSSVM 的 MAPE 值和 RMSE 值优于重构前。这主要是因为经过 SE 重构后，减少了模态的个数，从而降低了预测误差的累计。总的来说，采用 SE 重构对提高预测模型的精度是有效的。

(4) 本章构建的模型（EEMD-SE-PSO-LSSVM）一方面利用 SE 考虑了序列的复杂度，另一方面用 PSO 对参数进行了优化，与相应的单一模型对比，发现

经 SE 重构及 PSO 优化后，由于既对参数进行了优化，又通过 SE 减少了子序列个数，降低了预测误差累积，所以模型的预测效果最好，预测性能也有了很大的提升。综合上述结果，将原始的汇率数据经过有效的分解重构及优化后，发现可以得到更优的集成预测结果。

表 4.3 EEMD-SE-PSO-LSSVM 与对比模型的 DM 检验结果

目标模型	EEMD-SE-PSO-LSSVM	DM 值	P 值
对比模型	EEMD-SE-PSO-LSSVM	-5.0849	4.30E-07
	EEMD-PSO-LSSVM	-10.0617	6.55E-23
	EMD-PSO-LSSVM	-11.1229	2.05E-27
	EEMD-SE-LSSVM	-21.1527	1.02E-84
	EEMD-SE-KELM	-29.6247	1.91E-145
	EMD-SE-LSSVM	-13.486	1.08E-38
	EMD-SE-KELM	-25.7905	3.34E-117
	EEMD-LSSVM	-7.7231	2.42E-14
	EMD-LSSVM	-9.4104	2.48E-20
	EEMD-kELM	-25.9989	1.03E-118
	EMD-KELM	-25.9948	1.11E-118
	PSO-LSSVM	-14.6256	1.06E-44
	LSSVM	-27.9044	1.15E-132
	KELM	-33.6772	6.15E-176

4.5 本章小结

本章主要利用“分解-重构-集成”的思想，构建了一种基于样本熵重构和 PSO 优化的组合预测模型（EEMD-SE-PSO-LSSVM）。针对汇率数据的波动性、非线性、随机性等复杂的特点，首先采用 EEMD 方法对汇率数据序列进行了分解，获得了更好的降噪效果，利用 SE 对分解后的子序列进行重构减少了序列个数，然后利用 PSO 算法优化 LSSVM 的正则化参数和核参数，提高了模型的预测精度。实证结果表明，本章构建的 EEMD-SE-PSO-LSSVM 组合模型的预测效果明显优于其他的分解集成预测方法，其水平预测结果误差小，精度更高，方向预测性能好，能够有效的预测汇率数据。因此，该模型可以在实际汇率预测中作为一个可参考的预测工具，为投资者、外汇管理机构等未来的决策提供参考。

5 基于二次分解集成的比特币价格预测

本章以间隔 15min 的比特币价格数据的收盘价作为研究对象。5.1 节介绍了基于 CEEMDAN-PE-CEEMDAN-PSR-CSSA-KELM 模型的构建框架；5.2 节介绍了本章的数据来源；5.3 节展示了 CEEMDAN 分解及重构结果；5.4 节展示了 CEEMDAN 二次分解结果；5.5 节展示了相空间重构的结果；5.6 节对模型进行了分析总结；5.7 节对本章节的内容进行了小结。

5.1 CEEMDAN-PE-CEEMDAN-PSR-CSSA-KELM 模型的框架构建

本章基于“先分解后集成”的思想，首先将比特币价格数据用 CEEMDAN 方法分解为简单的 *IMF* 分量，以降低建模难度，然后通过计算各序列的 PE 值，将分解后的子序列根据 PE 值重构为高频项、低频项和趋势项。本章在一次分解的基础上使用了一种新的二次分解技术，即通过对一次分解产生的高频分量二次分解来降低高频分量的复杂度，提高预测精度。由于 KELM 模型的核参数是人为设定的，具有一定的主观性，因此本章采用 CSSA 算法优化 KELM 的核参数，以此来提高模型的预测性能，最终构建出一种基于分解算法、重构算法和优化算法相结合的组合预测新方法 (CEEMDAN-PE-CEEMDAN-PSR-CSSA-KELM)，该模型的流程构建如图 5.1 所示。

模型的主要步骤如下：

步骤 1 原始序列分解。使用 CEEMDAN 分解方法将原始比特币价格数据分解为有限多个相对简单且有意义的 *IMF* 分量。

步骤 2 模态分量重构。利用 PE 算法对得到的模态分量进行熵值计算，根据 PE 值将各子序列重构为高频项、低频项和趋势项。

步骤 3 重构项预测。对低频项和趋势项分别采用 CSSA-KELM 模型进行预测，得到低频分量的预测值和趋势分量的预测值；对高频项使用 CEEMDAN 方法进行二次分解，对分解后的各个分量分别使用 CSSA-KELM 模型预测，对预测

得到的分量非线性集成，得到高频分量预测值。

步骤 4 集成预测。采用 CSSA-KELM 方法对高频项、低频项和趋势项的预测结果非线性集成，得最终预测结果。

步骤 5 预测效果分析。通过对比本文模型与其他 11 种预测模型的 MAE、MAPE 和 RMSE 值，分析本文 CEEMDAN-PE-CEEMDAN-PSR-CSSA-KELM 模型的预测效果，用 DM 检验进一步证明该模型的有效性。

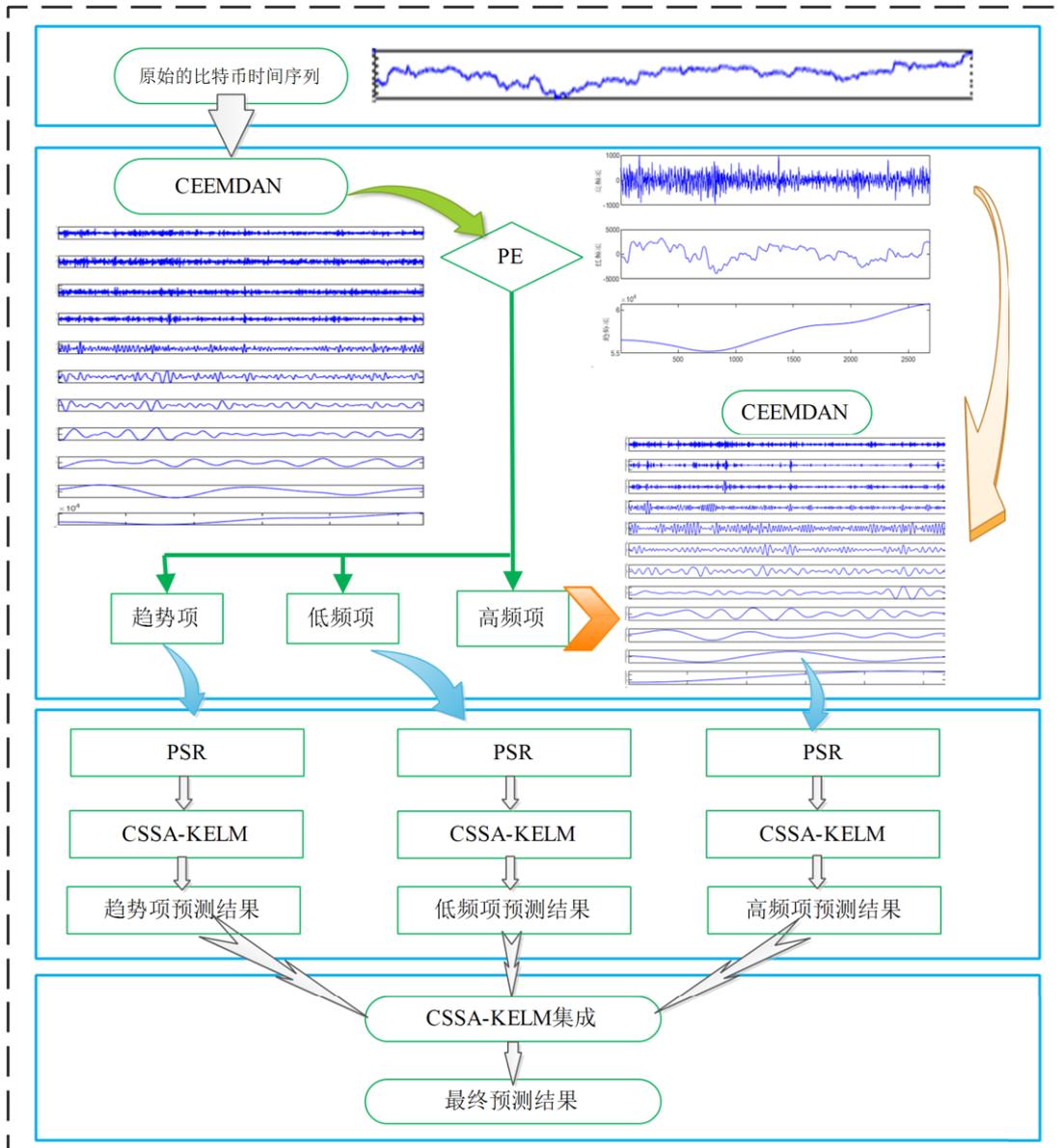


图 5.1 CEEMDAN-PE-CEEMDAN-PSR-CSSA-KELM 流程图

5.2 数据来源

本章选取比特币的币种为 bitstampUSD，数据来源于专业的比特币交易网站和平台 <https://bitcoincharts.com>，以收盘价数据作为研究对象，采样间隔为 15 分钟，样本区间自 2021 年 3 月 17 日 0:00 至 2021 年 4 月 14 日 0:00，总计 2689 个数据，具体如图 5.2 所示。将比特币数据划分为训练数据和测试数据，比特币数据的前 85% 为训练数据（共 2285 个），后 15% 为测试数据（共 404 个）。本章节中所有数据的计算使用 MATLAB 2020a 软件完成。

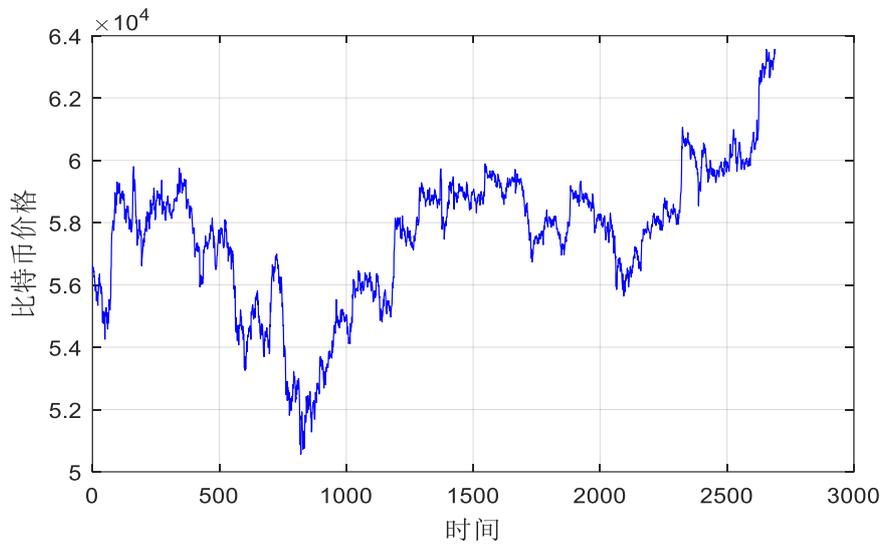


图 5.2 原始比特币价格数据（2021.3.17-2021.4.14）

5.3 CEEMDAN 分解及重构结果

按照 3.1.2 节 CEEMDAN 分解过程，使用 CEEMDAN 分解方法对原始比特币价格序列进行分解，分解结果如图 5.3 所示。图 5.3 中原始比特币价格序列分解为 10 个 *IMF* 分量和一个残余项 *Res*。

由图 5.3 可知，CEEMDAN 分解后 *IMF* 分量比较多，若对每个 *IMF* 分量及 *Res* 分别利用 CSSA-KELM 进行建模，每个模态分量的预测都会产生误差，模态分量越多，最后集成的累计误差就会越大，因此为了提高模型的预测精度，使用 PE 算法来评估各个分量的复杂性，并通过计算各个分量的 PE 值对其进行合并

重构，各分量的 PE 值如图 5.4 所示。重构各序列的规则如下：将 PE 值大于原始序列（其 PE 值为 2.7342）的各分量合并为一个序列，将小于原始序列的各分量合并为一个序列，趋势项不进行合并。最终将其重构为高频项，低频项，趋势项三个子序列，结果如表 5.1 所示，重构后的序列结果如图 5.5 所示。

表 5.1 各 IMF 分量重组的结果

高频项	低频项	趋势项
IMF1+IMF2+IMF3+IMF4+IMF5	IMF6+IMF7+IMF8+IMF9+IMF10	Res

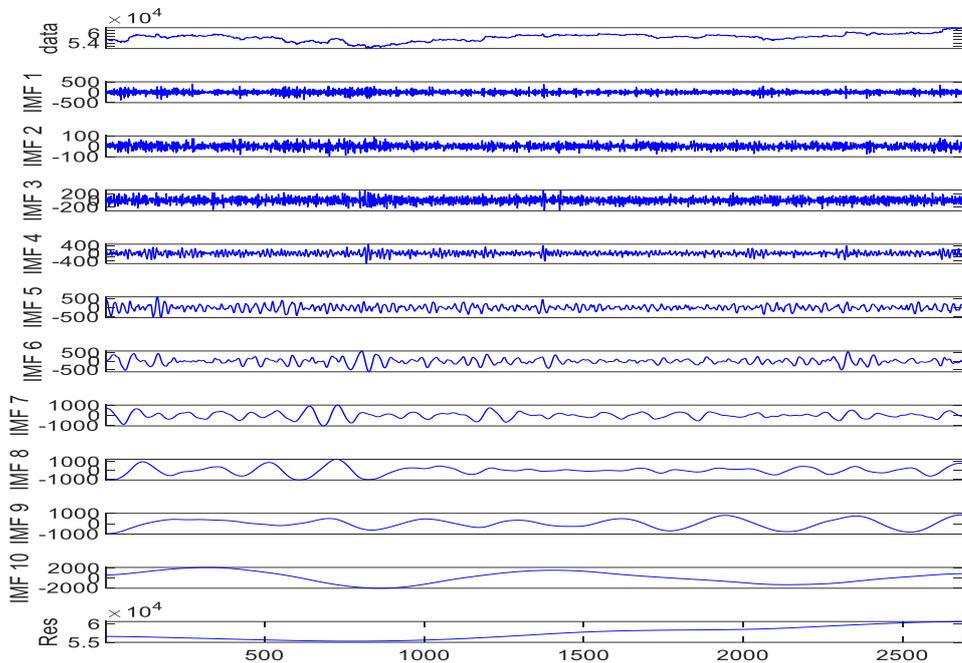


图 5.3 CEEMDAN 分解结果

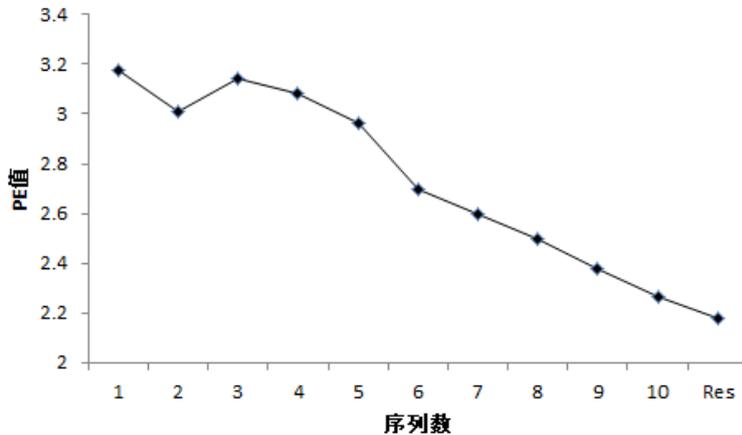


图 5.4 CEEMDAN 分解后各子序列的 PE 值

根据表 5.1 中各分量重组的结果，得到重构后的新序列如图 5.5 所示。

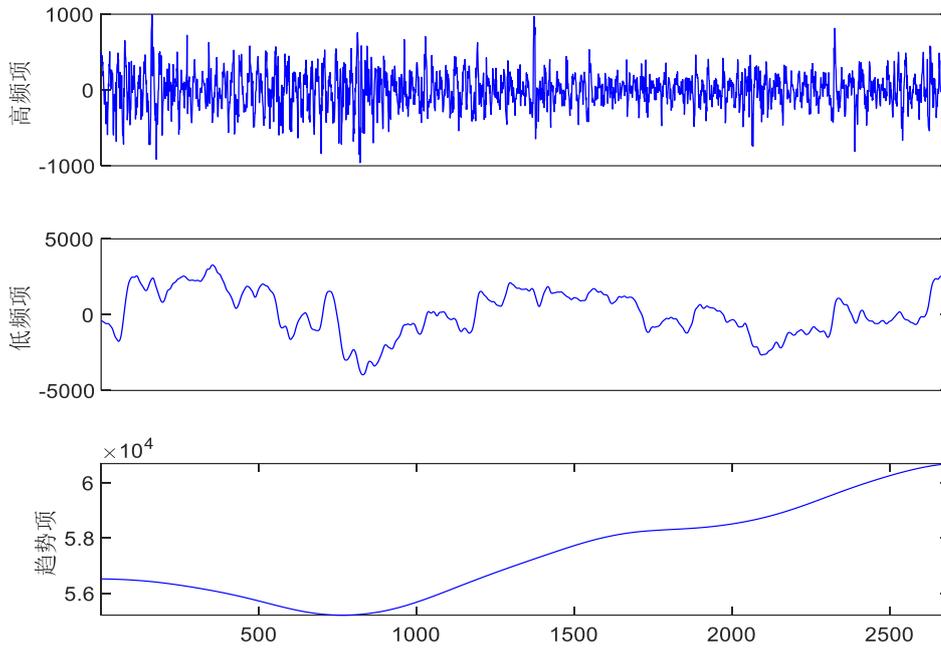


图 5.5 PE 重构后的结果

表 5.2 重构后各序列的 PE 值

高频项	低频项	趋势项
3.1462	2.7914	2.2534

5.4 CEEMDAN 二次分解结果

从表 5.2 重构后各序列的 PE 值和图 5.5 中 PE 重构后的结果可以看出，相对于低频项和趋势项，高频项波动程度大，具有较高的 PE 值，其对应的复杂度相对较高，因此本章中对高频项用 CEEMDAN 方法进行二次分解，分解结果如图 5.6 所示。从图 5.6 中我们可以看出，高频项被分解为 11 个 *IMF* 分量和 1 个 *Res*，通过使用 CSSA-KELM 模型对各分量分别进行预测，得到各个分量的预测结果，最后将各分量的预测结果用 CSSA-KELM 非线性集成即得最终的预测结果。

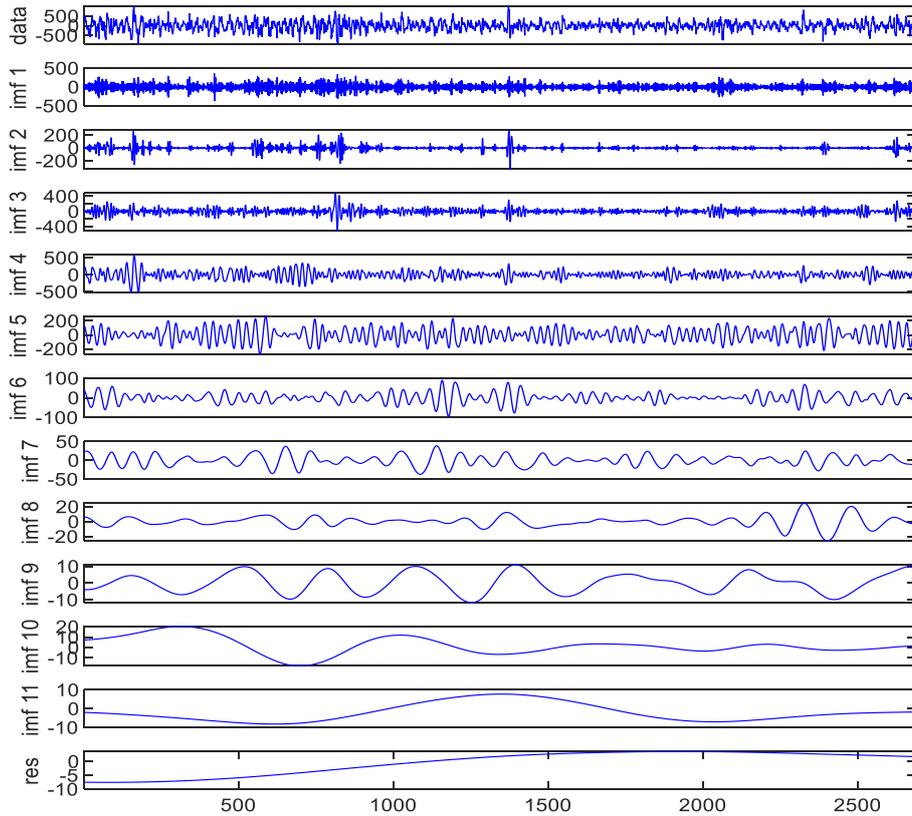


图 5.6 CEEMDAN 二次分解结果

5.5 相空间重构的结果

由于预测模型的输入维数与有效数据的提取有关，这会对比特币价格的预测结果产生影响，在现有文献中，我们发现时间序列的输入维数大多都是通过自相关函数或者依靠个人的主观经验确定的，而我们通过采用相空间重构（PSR）来确定预测模型的最佳输入维数，避免了主观选取，提高了模型的预测精度。本章所构建模型及各个比较模型的最佳输入维数如表 5.3~5.6 所示。

表 5.3 CEEMDAN 初次分解后 PSR 结果

序列数	1	2	3	4	5	6	7	8	9	10	Res
m	4	4	4	4	4	3	3	3	3	3	3
t	5	1	2	3	6	5	9	10	10	10	10
w	15	3	6	9	18	10	18	20	20	20	20

表 5.4 EEMD 分解后 PSR 结果

序列数	1	2	3	4	5	6	7	8	9	10	Res
m	4	4	4	4	3	3	3	3	3	3	2
t	5	2	4	8	9	10	10	10	10	10	2
w	15	6	12	24	18	20	20	20	20	20	2

表 5.5 重构后 PSR 结果

序列数	CEEMDAN			EEMD		
	1	2	3	1	2	3
m	4	3	3	4	3	2
t	5	10	10	7	10	2
w	15	20	20	21	20	2

表 5.6 CEEMDAN 二次分解后 PSR 结果

	1	2	3	4	5	6	7	8	9	10	11	Res
m	4	4	4	4	3	3	3	3	3	3	3	2
t	8	2	2	4	7	5	8	10	10	10	10	10
w	24	6	6	12	14	10	16	20	20	20	20	10

5.6 模型比较分析

为了验证构建的基于二次分解技术及 CSSA 优化 KELM 模型(即 CEEMDAN-PE-CEEMDAN-PSR-CSSA-KELM) 的有效性, 本章将从单一预测模型、分解算法、重构算法及优化算法四个方面进行比较分析, 各模型预测评价指标的对比如表 5.7 所示, 模型比较结果如图 5.7 所示, 可得结论如下:

(1) 就单一预测模型而言, 对于本章的比特币收盘价, 一方面, 从评价指标来看, 由表 7 中可以看出, ELM 预测模型的 MAE, MAPE 和 RMSE 分别为 1588.9495, 2.5375, 3469.1632; KELM 预测模型的 MAE, MAPE 和 RMSE 分别为 1557.8605, 2.5089, 2701.5835, 与 ELM 预测模型相比, 上述各个指标均下降, 这表明 KELM 预测模型的精度更高, 效果更好。另一方面, 从预测模型的结构上来说, ELM 预测模型中隐含层的参数是随机设定的, 当我们用核函数来代替隐含层的随机映射时, 不仅避免了 ELM 模型随机设定隐含层参数而带来的稳定性差的问题, 而且在很大程度上提升了模型的鲁棒性, 所以 KELM 模型优于 ELM 模型, 其预测性能更好。

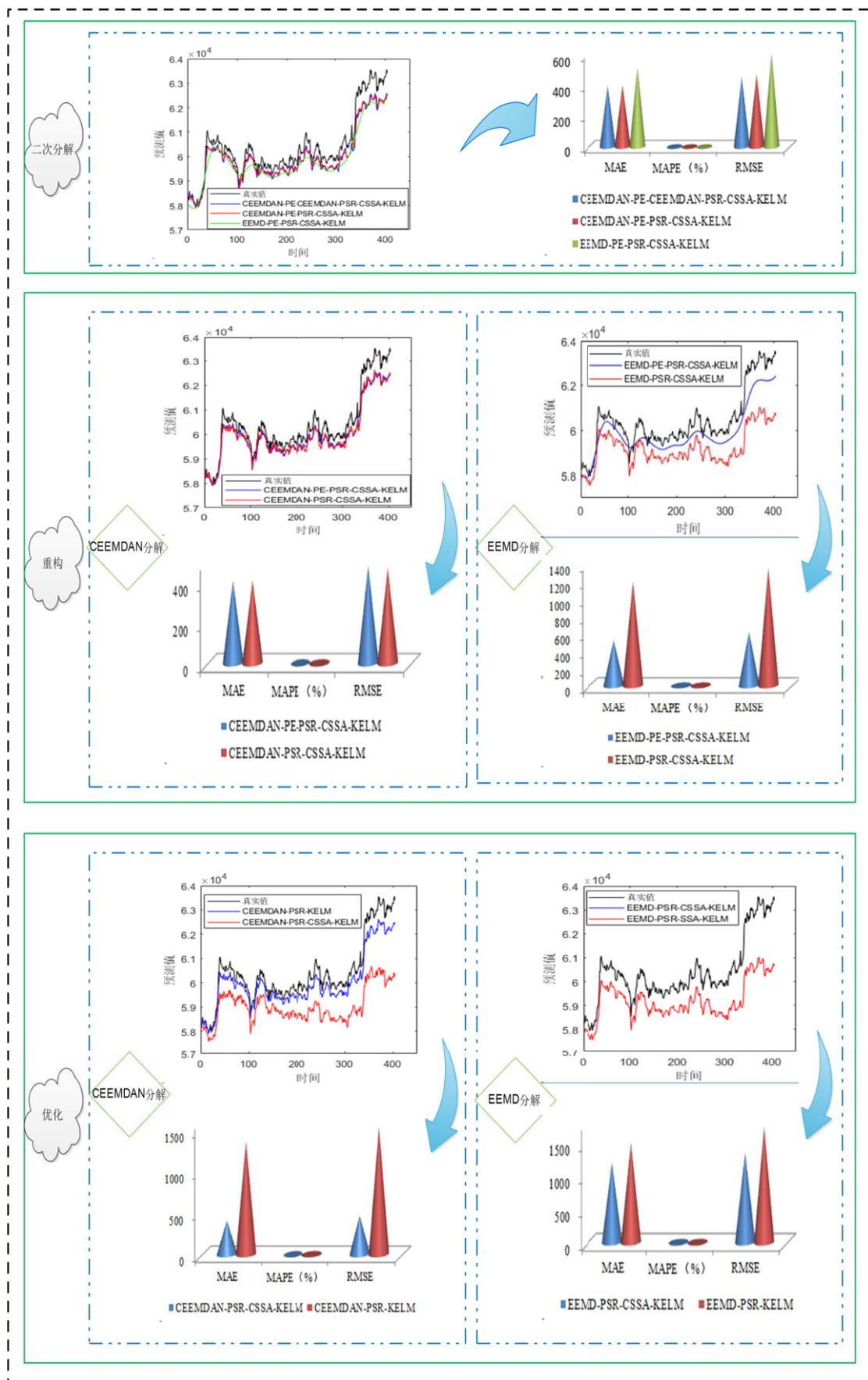


图 5.7 模型比较结果

表 5.7 各模型预测评价指标对比

	MAE	MAPE (%)	RMSE
CEEMDAN-PE-CEEMDAN-PSR-CSSA-KELM	400.5883	0.6562	459.9546
CEEMDAN-PE-PSR-CSSA-KELM	405.6745	0.6645	477.9594
CEEMDAN-PE-PSR-SSA-KELM	408.1547	0.6686	480.4997
EEMD-PE-PSR-CSSA-KELM	521.5323	0.8554	616.8227
EEMD-PE-PSR-SSA-KELM	530.0397	0.8689	608.8657
CEEMDAN-PSR-CSSA-KELM	406.6709	0.6663	465.7163
EEMD-PSR-CSSA-KELM	1195.2572	1.9604	1345.2625
EEMD-PSR-SSA-KELM	1195.5523	1.9609	1345.5661
CEEMDAN-PSR-KELM	1363.0394	2.2349	1538.5821
EEMD-PSR-KELM	1492.5593	2.4455	1731.0627
PSR-KELM	1557.8605	2.5089	2701.5835
PSR-ELM	1588.9495	2.5375	3469.1632

注：加粗的表示最优预测结果。

(2) 就分解算法而言，我们发现二次分解的效果要好于单次分解、单次分解的效果要好于未分解的效果，现从这两个方面进行比较。1) 二次分解相比单次分解：由表 7 中我们可以看出本文模型 CEEMDAN-PE-CEEMDAN-PSR-CSSA - KELM 的 MAE, MAPE, RMSE 分别为 400.5883, 0.6562, 459.9546, 相较于单次分解的模型其性能更优。譬如，与 EEMD-PE-PSR-CSSA-KELM 相比较，本文模型在 MAE、MAPE 和 RMSE 上分别降低了 23.2%、23.3%和 25.4%，从中我们可以看出，相比单次分解，二次分解模型预测效果更优，二次分解模型在提高预测精度方面是非常有效的。2) 单次分解相比未分解：由表 5.7 各模型预测评价指标对比中可以看出，相比用 PSR-KELM 直接预测，采用 CEEMDAN 分解方法后 MAE 降低了 12.5%，MAPE 降低了 10.9%，RMSE 降低了 43.1%；采用 EEMD 单次分解后，模型 EEMD-PSR-KELM 的 MAE 降低了 4.2%，MAPE 降低了 2.5%，RMSE 降低了 35.9%。从中我们可以看出，无论 CEEMDAN 分解还是 EEMD 分解，分解后的预测效果要显著优于未分解的，而且我们发现经过 CEEMDAN 分解后模型的预测效果明显优于 EEMD 分解后的预测效果。

(3) 就重构算法上来说，采用 PE 重构后的模型的预测效果明显优于未重构模型的预测效果。由表 5.7 各模型预测评价指标对比中我们可以看出，重构后的模型的 MAE, MAPE, RMSE 低于未重构的模型。譬如，CEEMDAN-PE-PSR - CSSA-KELM 模型的 MAE, MAPE, RMSE 分别为 405.6745, 0.6645 和 477.9594，

相较于未重构的 CEEMDAN -PSR-CSSA-KELM 模型，其上述指标分别降低了 0.25%，0.27%，2.6%，由此可看出：重构算法对提高模型的预测精度是有效的。

(4) 就优化算法上而言，我们比较了 2 种不同的优化方法来优化 KELM 的核参数，由表 5.7 各模型预测评价指标对比中我们可以看出，优化后的模型的预测效果要好于未优化的模型。譬如，相比未优化的(CEEMDAN-PSR-KELM)，采用 CSSA 优化后的模型（CEEMDAN-PSR-CSSA-KELM）的 MAE 降低了 70.1%，MAPE 降低了 70.2%，RMSE 降低了 69.7%，优化后模型的预测效果显著。另外，从图 5.7 可知，采用 CSSA 优化的模型 EEMD-PSR-CSSA-KELM 的预测效果要好于采用 SSA 优化的模型 EEMD-PSR-SSA-KELM 的预测效果。总之，采用 CSSA 优化方法可以更好的改善 KELM 的核参数，取得更好的预测效果。

综上所述，发现本章模型 CEEMDAN-PE-CEEMDAN-PSR-CSSA-KELM 在分解方法、重构方法和优化方法上预测精度都有所提升，而且我们发现，由于结合了分解方法和重构方法，在优化预测模型时提升效果更好，模型的整体预测精度也更好。

为从统计学意义上检验本章所构建的模型与对比模型预测性能的差异性，本章利用 DM 统计量对上面所提到的 11 个对比模型和本章所使用的模型进行检验，判断其是否具有显著差异，检验结果如表 5.8 所示。从表 5.8 的检验结果中我们可看出，各模型在显著性水平 1%、5%和 10%的水平下拒绝原假设，再次体现出本章模型预测性能的优势。

表 5.8 DM 检验结果

	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
M1	-3.0666 (***)	-1.2952 (*)	-3.4854 (***)	-11.2965 (***)	-15.4697 (***)	-18.6910 (***)	-18.6997 (***)	-19.3083 (***)	-17.5736 (***)	-9.1394 (***)	-7.9694 (***)
M2		1.8264 (*)	-21.702 (***)	-11.7788 (***)	-19.6094 (***)	-18.5065 (***)	-18.5154 (***)	-19.2148 (***)	-17.4815 (***)	-9.0190 (***)	-7.9578 (***)
M3			-2.1967 (**)	-10.2239 (***)	-13.7176 (***)	-18.5252 (***)	-18.5336 (***)	-19.2604 (***)	-17.4873 (***)	-9.0448 (***)	-7.9641 (***)
M4				-11.6265 (***)	-19.5486 (***)	-18.4994 (***)	-18.5083 (***)	-19.2102 (***)	-17.4773 (***)	-9.0027 (***)	-7.9566 (***)
M5					0.8163	-17.1411 (***)	-17.1513 (***)	-18.3056 (***)	-16.8257 (***)	-7.6869 (***)	-7.8702 (***)
M6						-18.1387 (***)	-18.1483 (***)	-18.9913 (***)	-17.2642 (***)	-8.0553 (***)	-7.8898 (***)
M7							-4.8020 (***)	-17.2378 (***)	-16.1020 (***)	17.7723 (***)	-7.2598 (***)
M8								-17.1923 (***)	-16.0958 (***)	17.7450 (***)	-7.2590 (***)
M9									-11.5420 (***)	28.6495 (***)	-6.9745 (***)
M10										22.7859 (***)	-6.6911 (***)
M11											-7.8679 (***)

注：M1 代表 CEEMDAN-PE-CEEMDAN-PSR-CSSA-KELM, M2 代表 CEEMDAN-PE -PSR-CSSA-KELM, M3 代表 CEEMDAN-PSR-CSSA-KELM, M4 代表 CEEMDAN- PE-PSR-SSA-KELM, M5 代表 EEMD-PE-PSR- CSSA-KELM, M6 代表 EEMD-PE-PSR -SSA-KELM, M7 代表 EEMD-PSR-CSSA-KELM, M8 代表 EEMD-PSR -SSA-KELM, M9 代表 CEEMDAN-PSR-KELM, M10 代表 EEMD-PSR-KELM, M11 代表 PSR-KELM, M12 代表 PSR-ELM。在每一行中,上面的数字是 DM 检验的值,下面圆括号中的星号(***)是 1%的水平下显著,(**)是 5%的水平下显著,(*)是 10%的水平下显著。

5.7 本章小结

本章基于“先分解后集成”的思想,构建出一种基于分解算法、重构算法和优化算法的组合预测新方法(CEEMDAN-PE-CEEMDAN-PSR-CSSA-KELM),为验证所构建模型的有效性,从单一预测模型、分解算法、重构算法及优化算法四个方面进行了对比试验的比较分析:(1)单一预测模型:与 ELM 模型对比,本章所用 KELM 模型预测性能更好,说明本章所用单模型具有稳定性;(2)分解算法:相比单次分解和未分解方法,本章对高频分量进行的二次分解可以更加有效的提高模型的预测精度;(3)重构算法:本章采用 PE 算法对分解后的各分量重构,减少了序列的个数,降低了计算的复杂度和误差累积,结果显示采用 PE 重构后的模型的预测效果明显优于未重构的模型的预测效果;(4)优化算法:相比未优化及 SSA 优化,本章使用 CSSA 优化 KELM 的核参数,可以明显的提高模型的预测精度。

总的来说,本章构建的 CEEMDAN-PE-CEEMDAN-PSR-CSSA-KELM 模型既利用 PE 考虑了模态序列的复杂度,将各模态序列重构为高频项,低频项和趋势项,又考虑了高频项波动性大,利用 CEEMDAN 对其二次分解降低了高频项的复杂性,还用 CSSA 对 KELM 的核参数进行了优化。本章利用 MAE、MAPE 和 RMSE 作为各个模型预测性能的评价指标,再利用 DM 进行稳健性检验,以证明模型预测性能的科学性。与相应的基准模型对比,我们可以发现,经过 CEEMDAN 分解、PE 重构及 CSSA 优化后,由于结合了 CEEMDAN 分解算法、PE 重构算法及 CSSA 优化算法各自独特的优势,本章模型显示出优良的预测效果,预测性能较其他对比模型也有了很大的提升。综上所述,将原始的比特币收盘价数据经过有效的分解重构及优化后,我们发现可以得到更优的预测结果。

6 基于二次分解重构集成的原油期货价格预测

本章以 Brent 原油期货和 WTI 原油期货的收盘价为研究对象。6.1 节介绍了基于 ICEEMDAN-PE-EMD-PSR-CSSA-KELM 模型的构建框架；6.2 节介绍了本章的数据来源；6.3 节展示了 ICEEMDAN 分解及重构结果；6.4 节展示了 EMD 二次分解及重构结果；6.5 节展示了对模型的分析总结；6.6 节对本章节的内容进行了小结。

6.1 ICEEMDAN-PE-EMD-PSR-CSSA-KELM 模型框架的构建

本章首先将原油期货价格数据用 ICEEMDAN 分解方法分解为简单的 *IMF* 分量，以降低建模难度，然后计算各 *IMF* 分量的 PE 值，将分解后的各 *IMF* 分量根据 PE 值重构为高频项、低频项和趋势项，对高频项进行 EMD 二次分解来降低高频信号的复杂度，有助于提高模型的预测精度，构建出一种基于二次分解重构集成的原油期货价格预测新方法（ICEEMDAN-PE-EMD-CSSA-KELM），该模型的流程构建如图 6.1 所示。

模型的主要步骤如下：

步骤 1 原始序列分解。使用 ICEEMDAN 分解方法将原始原油期货价格数据分解为有限多个相对简单且有意义的 *IMF* 分量。

步骤 2 模态分量重构。利用 PE 算法对得到的 *IMF* 分量进行熵值计算，根据 PE 值将各分量重构为高频项、低频项和趋势项。

步骤 3 重构项的预测。对低频项和趋势项分别采用 CSSA-KELM 模型进行预测，得到低频项的预测值和趋势项的预测值；对高频项使用 EMD 方法二次分解，对分解后的各个分量使用 PE 算法重构后，对其分别采用 CSSA-KELM 模型预测，对预测得到的结果进行线性集成，得到高频项的预测值。

步骤 4 集成预测。本章用到了 2 次集成预测方法，对高频项的预测结果采用了重构后各子序列预测值直接加和的线性集成方法；然后对高频项、低频项和趋势项的预测值采用 CSSA-KELM 进行非线性集成得最终预测值。具体做法如下：将高频项、低频项和趋势项的预测结果作为模型的输入变量，将原油期货的实际价格作为预测模型的输出变量；通过 CSSA-KELM 算法内部的学习，建立起

高频项、低频项和趋势项与实际原油期货价格之间的关系；最后输入高频项、低频项和趋势项的预测值，即得到原油期货价格的最终预测结果。

步骤 5 预测效果分析。本章选取 MAE、MAPE、RMSE 和 DS 作为预测精度衡量指标，通过与其它预测模型相比较，分析本章 ICEEMDAN-PE-EMD-PSR-CSSA-KELM 模型的预测效果，并使用 DM 检验进一步证明该模型的有效性。

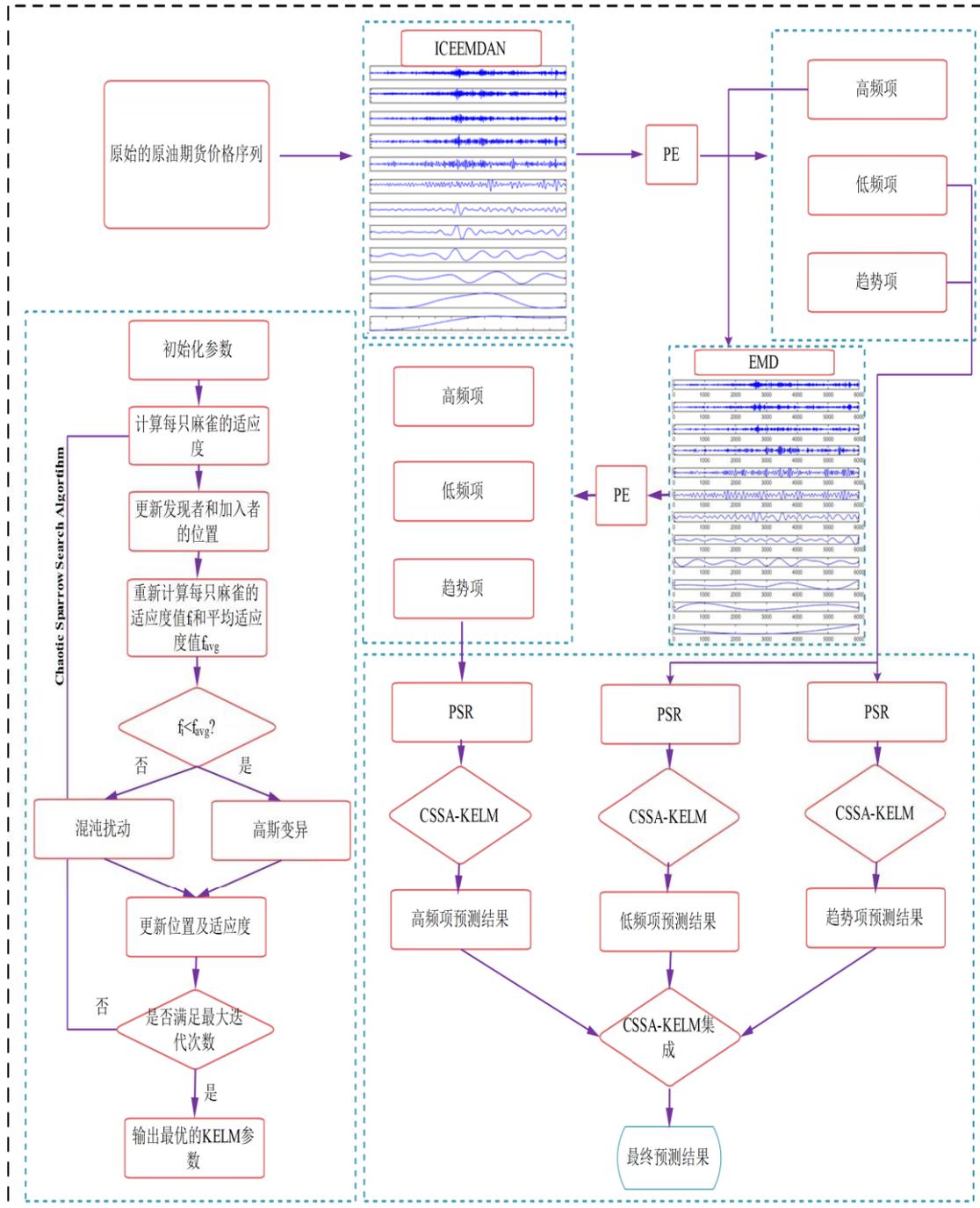


图 6.1 本文模型构建流程图

6.2 数据来源

本章选取 Brent 和 WTI 原油期货的收盘价作为样本数据，原油期货数据来源于 wind 数据库。Brent 的样本数据自 1997 年 12 月 30 日至 2021 年 6 月 1 日，总计 6000 个数据，WTI 的样本数据自 1994 年 6 月 1 日至 2021 年 6 月 1 日，总计 6818 个数据，具体如图 6.2 所示，本章使用 MATLAB 2020a 软件完成。

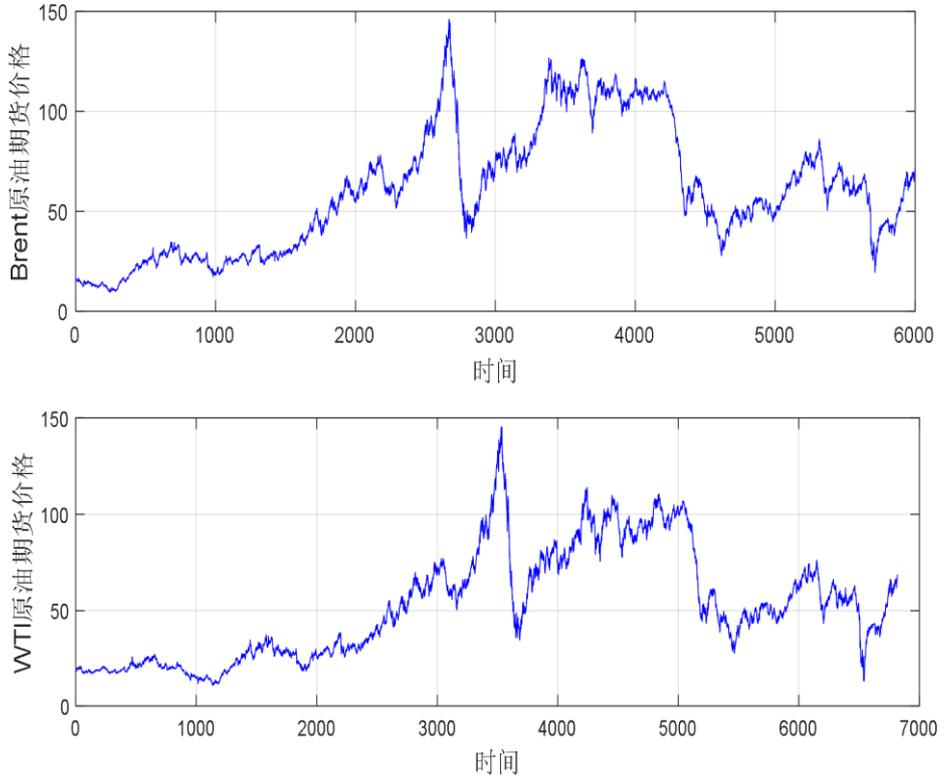


图 6.2 两种原油期货的日度价格序列

在本次实证中，Brent 原油期货自 1997 年 12 月 30 日到 2016 年 10 月 4 日（即前 80%为训练数据），共 4800 个数据作为训练样本，2016 年 10 月 5 日到 2021 年 6 月 1 日（即后 20%为测试数据），共 1200 个数据作为测试样本；WTI 原油期货自 1994 年 6 月 1 日到 2016 年 1 月 27 日（即前 80%），共 5455 个数据作为训练样本，2016 年 1 月 28 日到 2021 年 6 月 1 日（即后 20%为测试数据），共 1363 个数据作为测试样本。值得注意的是，在本次实证中，我们不仅进行了提前 1 步预测，而且还进行了提前 2 步和 3 步的预测，以此来评估模型的短期和中期预测能力。

对于一个时间序列 $x_t (t=1,2,\dots,T)$ ，提前 m 步预测的公式如下：

$$\hat{x}_{t+m} = F(x_t, x_{t-1}, \dots, x_{t-l}), t = 1, 2, \dots, T$$

其中， \hat{x}_{t+m} 是 t 时刻提前 m 步的预测值， x_t 是 t 时刻的真实值， l 代表滞后的期数。在本章中，滞后的期数 l 根据相空间重构函数来确定。

6.3 ICEEMDAN 分解及重构结果

按照 ICEEMDAN 分解过程，对原始的 Brent 原油期货和 WTI 原油期货价格序列进行 ICEEMDAN 分解，分解结果如图 6.3 所示。从图 6.3 中可看出 2 种原油期货价格序列分别被分解为 11 个 *IMF* 分量和一个残差项 *Res*。

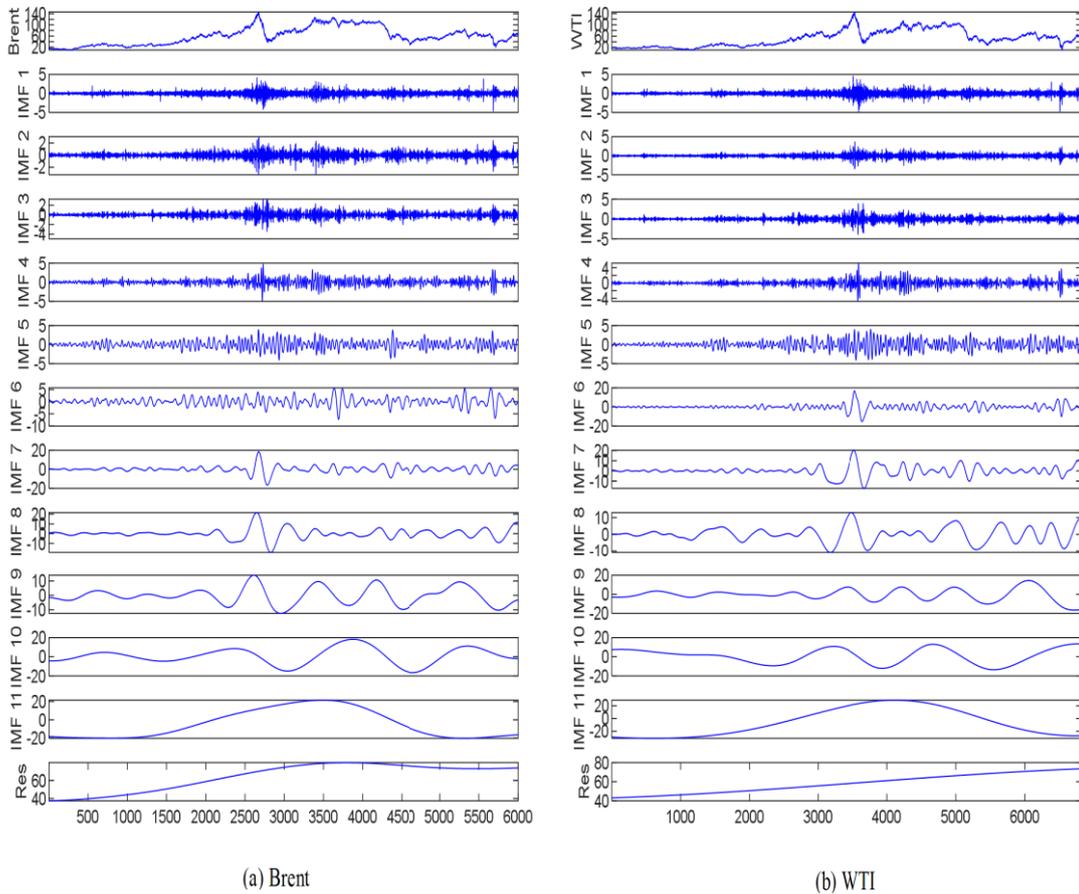


图 6.3 两种原油期货价格经 ICEEMDAN 分解的结果

由图 6.3 可知，ICEEMDAN 分解后 *IMF* 分量比较多，若对每个 *IMF* 分量及 *Res* 分别使用 CSSA-KELM 进行建模，每个 *IMF* 分量的预测都会产生误差，*IMF* 分量越多，最后集成时累计误差就会越大，因此为了提高模型的预测精度，

我们使用 PE 值来评估各个分量的复杂性，通过计算各个分量的 PE 值对其进行合并重构，各分量的 PE 值如图 6.4 所示。经过合并后，最终将其重构为高频项、低频项和趋势项三个序列，重构后的序列结果如图 6.5 所示，重构后各序列的 PE 值如表 6.1 所示。

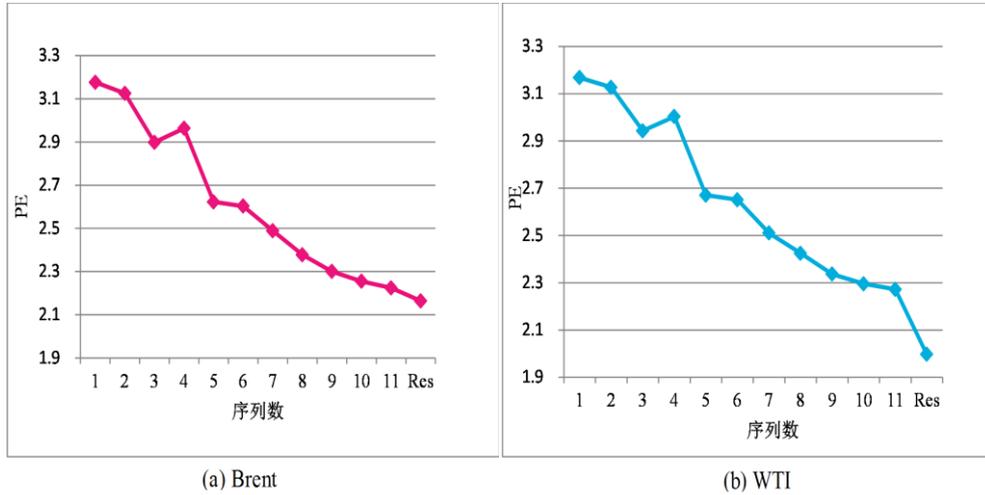


图 6.4 两种原油期货经 ICEEMDAN 分解后各子序列的 PE 值

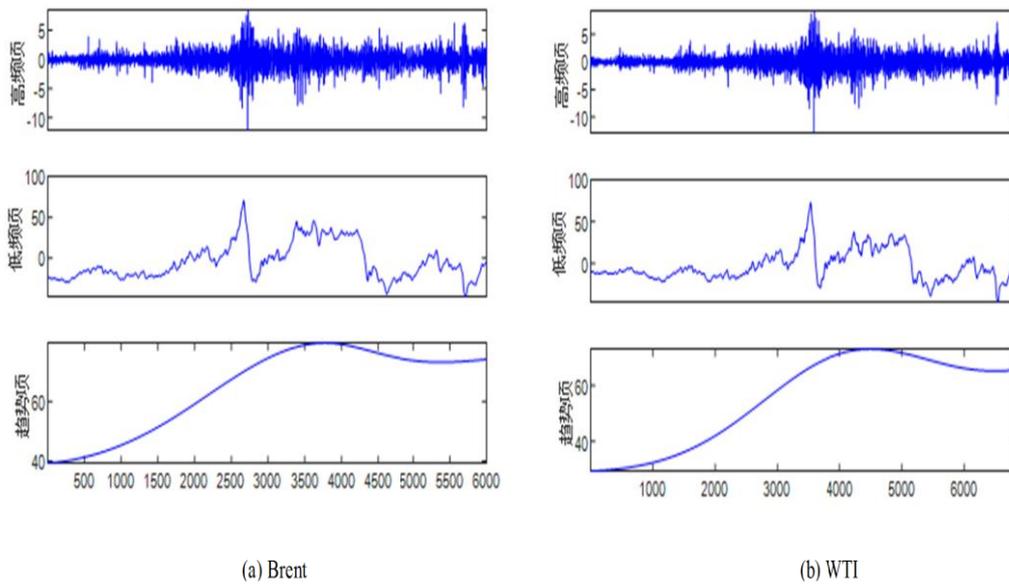


图 6.5 两种原油期货经 PE 重构后的结果

表 6.1 重构后各序列的 PE 值

序列	Brent			WTI		
	高频项	低频项	趋势项	高频项	低频项	趋势项
PE	3.16824	2.757274	2.26291	3.142932	2.765721	2.262687

6.4 EMD 二次分解及重构结果

从表 6.1 重构后各序列的 PE 值和图 6.5 中经 PE 重构后的两种原油期货价格子序列可看出, 相对于低频项和趋势项, 高频项波动程度大, 具有较高的 PE 值, 其对应的复杂度相对较高, 因此本章用 EMD 分解方法对高频项进行二次分解, 分解结果如图 6.6 所示。从图 6.6 中可看出, 高频项被分解为 11 个 imf 分量和 1 个残余项 res, 然后对二次分解后的各 imf 分量分别计算其 PE 值, PE 值的计算结果如图 6.7 所示, 将 PE 值高于原始高频项的子序列重构为一项, 对低于原始高频项的子序列重构为另一项, 趋势项为一项, 重构后的结果如图 6.8 所示。最终将其重构为 3 个模态分量, 对这 3 个模态分量分别使用 CSSA-KELM 模型预测, 将 3 个分量的预测结果采用加和的线性集成方法集成, 即得高频项预测结果。

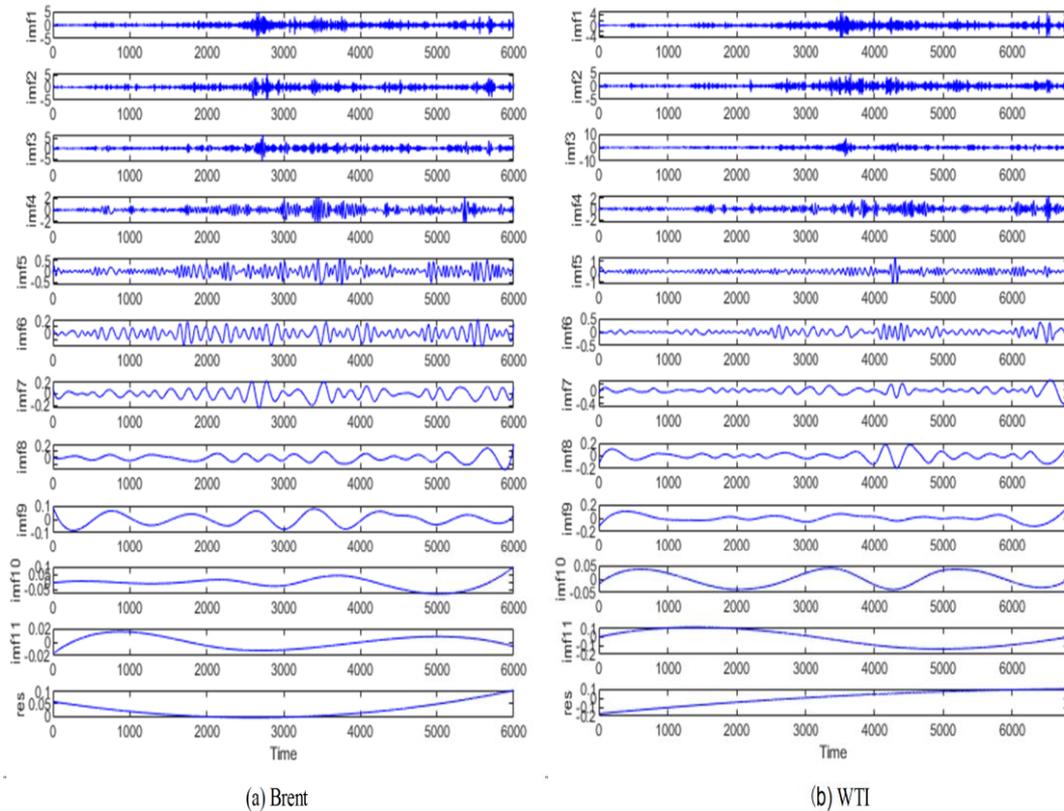


图 6.6 两种原油期货的高频项经 EMD 二次分解的结果

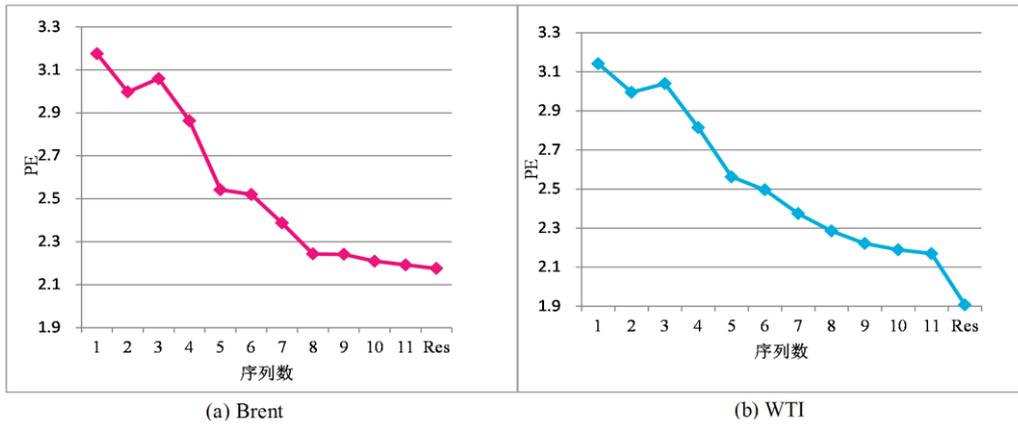


图 6.7 两种原油期货高频项经 EMD 二次分解后各序列的 PE 值

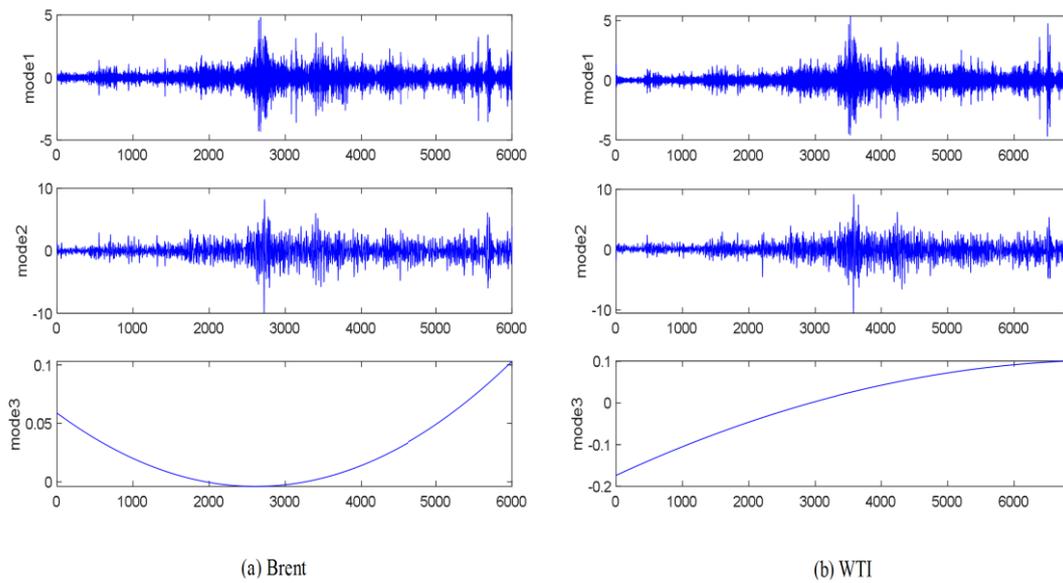


图 6.8 两种原油期货高频项二次重构的结果

6.5 模型比较分析

为验证本章所构建的“二次分解-重构-集成”模型的有效性，下面从单一预测模型、分解集成模型及采用智能算法优化参数的分解重构集成模型三个方面进行比较分析。不同模型的特征如表 6.2 所示，表 6.3 和表 6.4 分别为两种原油在 1 步预测、2 步预测和 3 步预测下 4 个评价指标（MAE、MAPE、RMSE 和 DS）的预测结果，图 6.9 和图 6.10 分别为 Brent 原油和 WTI 原油的预测结果，图 6.11 和图 6.12 分别为两种原油各模型评价指标的对比图。各模型的分析结果如下：

表 6.2 不同模型的特征

	采用分解方 法	采用 CSSA 优化方法	采用二次分 解方法
PSR-ELM(Model 1)			
PSR-KELM(Model 2)			
EMD-PE-PSR-KELM(Model 3)	√		
ICEEMDAN-PE-PSR-KELM(Model 4)	√		
EMD-PE-PSR-CSSA-KELM(Model 5)	√	√	
ICEEMDAN-PE-PSR-CSSA-KELM(Model 6)	√	√	
ICEEMDAN-PE-EMD-PSR-CSSA-KELM(Proposed)	√	√	√

表 6.3 Brent 原油各模型预测性能对比

	评价准则	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Proposed
1-step ahead	MAE	2.1646	1.5601	1.5257	1.3476	0.8268	1.0598	0.7484
	MAPE(%)	3.9268	3.0685	2.9822	2.7888	1.5853	2.5563	1.5416
	RMSE	2.6628	2.0543	1.9786	1.8341	1.1795	1.4253	1.1333
2-step ahead	DS	48.1605	53.0936	52.9264	57.3579	61.2876	58.8562	74.8328
	MAE	2.4323	1.6578	1.4628	1.4507	1.0503	1.1915	0.9517
	MAPE(%)	4.3970	3.1976	2.8132	2.9072	2.0157	2.3469	1.9688
	RMSE	3.1194	2.2651	1.9688	2.0048	1.4747	1.7122	1.4342
3-step ahead	DS	47.4080	52.9264	54.2642	58.0268	60.1171	58.6957	62.7090
	MAE	2.5245	1.8424	1.6075	1.5603	1.1484	1.3710	1.2511
	MAPE(%)	4.5518	3.5553	3.0885	3.1068	2.2113	2.6815	2.4760
	RMSE	3.3084	2.5395	2.1833	2.1877	1.6076	1.9413	1.8193
	DS	54.4314	53.0936	54.7659	58.1940	62.2074	57.8595	55.4348

表 6.4 WTI 原油各模型预测性能对比

	评价准则	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Proposed
1-step ahead	MAE	1.7742	1.4483	1.1763	1.0303	0.9435	0.9984	0.7951
	MAPE(%)	3.8138	3.2277	2.6829	2.5676	2.2498	2.4562	2.0162
	RMSE	2.2421	1.9415	1.6198	1.5789	1.4317	1.5002	1.2421
2-step ahead	DS	48.2708	52.0971	54.9669	56.4386	62.1781	56.5857	73.2892
	MAE	2.1393	1.6160	1.4858	1.3792	1.1828	1.2420	1.0293
	MAPE(%)	4.4696	3.5685	3.3856	3.2097	2.7725	2.9534	2.4940
	RMSE	2.6419	2.1748	2.0206	1.9153	1.7363	1.7982	1.5162
3-step ahead	DS	46.8727	52.0971	54.4518	57.5423	59.1611	57.1744	62.3252
	MAE	2.0647	1.8151	1.6484	1.4969	1.3113	1.4072	1.2823
	MAPE(%)	4.4387	4.0098	3.7357	3.4623	3.0491	3.2826	3.0054
	RMSE	2.5564	2.4465	2.2383	2.0896	1.8816	1.9922	1.8202
	DS	47.5350	52.0971	53.9367	57.3216	58.7196	55.7027	54.2311

(1) 就单一预测模型而言，对于本章的原油期货价格，从表 6.3 和表 6.4 中

可以看出，不管是 Brent 还是 WTI 原油期货，与 ELM 相比，KELM 不仅在 1 步预测上表现良好，在 2 步预测和 3 步预测上也表现出了不错的效果。以 WTI 原油期货价格的 1 步预测为例，在水平精度上，ELM 和 KELM 的 MAE 分别为 1.7742 和 1.4483，MAPE 分别为 3.8138%和 3.2277%，RMSE 分别为 2.2421 和 1.9415；在方向精度上，ELM 的 DS 为 48.2708，KELM 的 DS 为 52.0971，KELM 不仅在水平精度上表现良好，在方向精度上表现也不错。同时，在更高步的预测中，KELM 在水平和方向精度上的表现仍不错，以 WTI 原油 3 步预测为例，在水平精度上，ELM 和 KELM 的 MAE 分别为 2.0647 和 1.8151，MAPE 分别为 4.4387%和 4.0098%，RMSE 分别为 2.5564 和 2.4465；在方向精度上，ELM 的 DS 为 47.5350，KELM 的 DS 为 52.0971，结果表明，KELM 的预测效果明显优于 ELM。

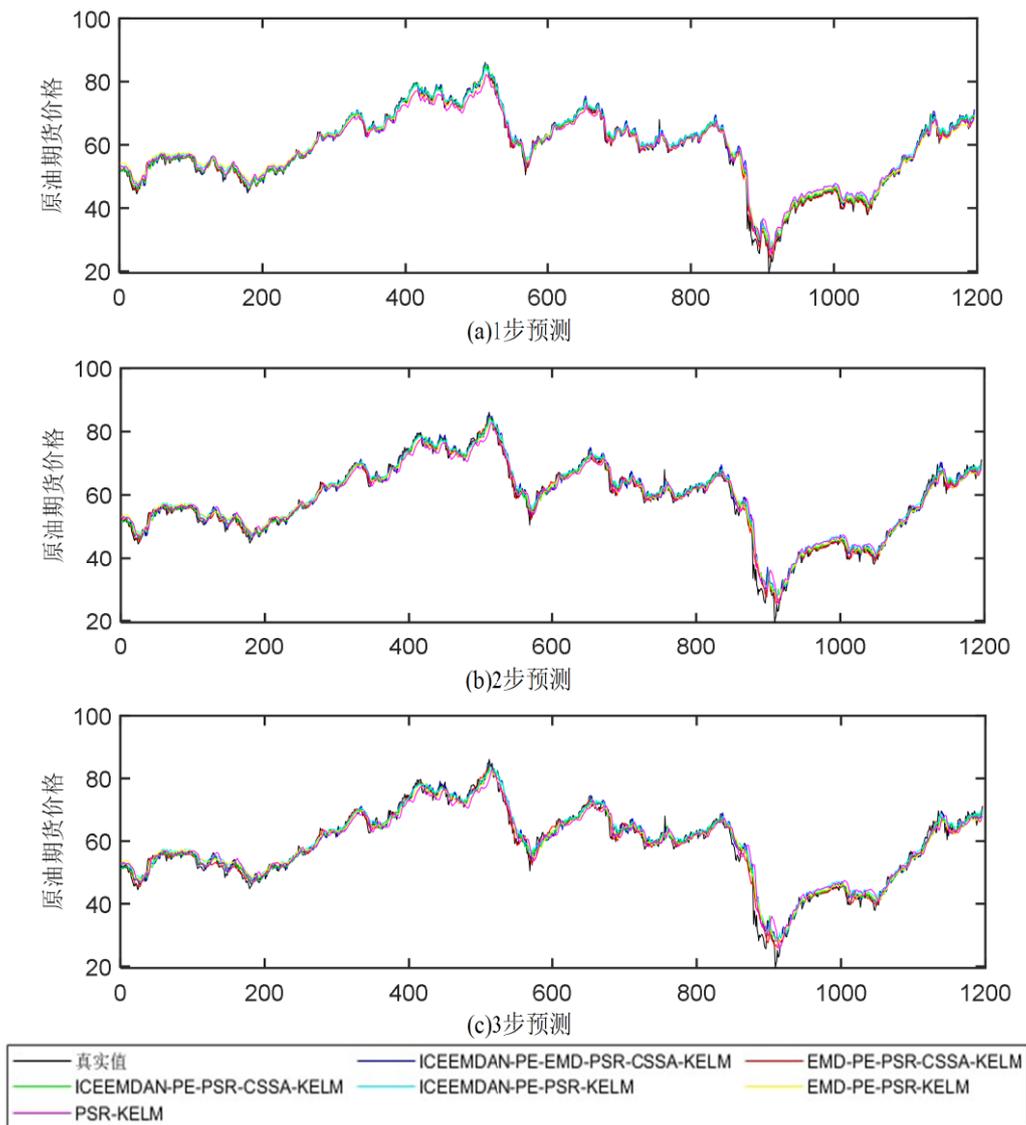


图 6.9 Brent 原油预测结果

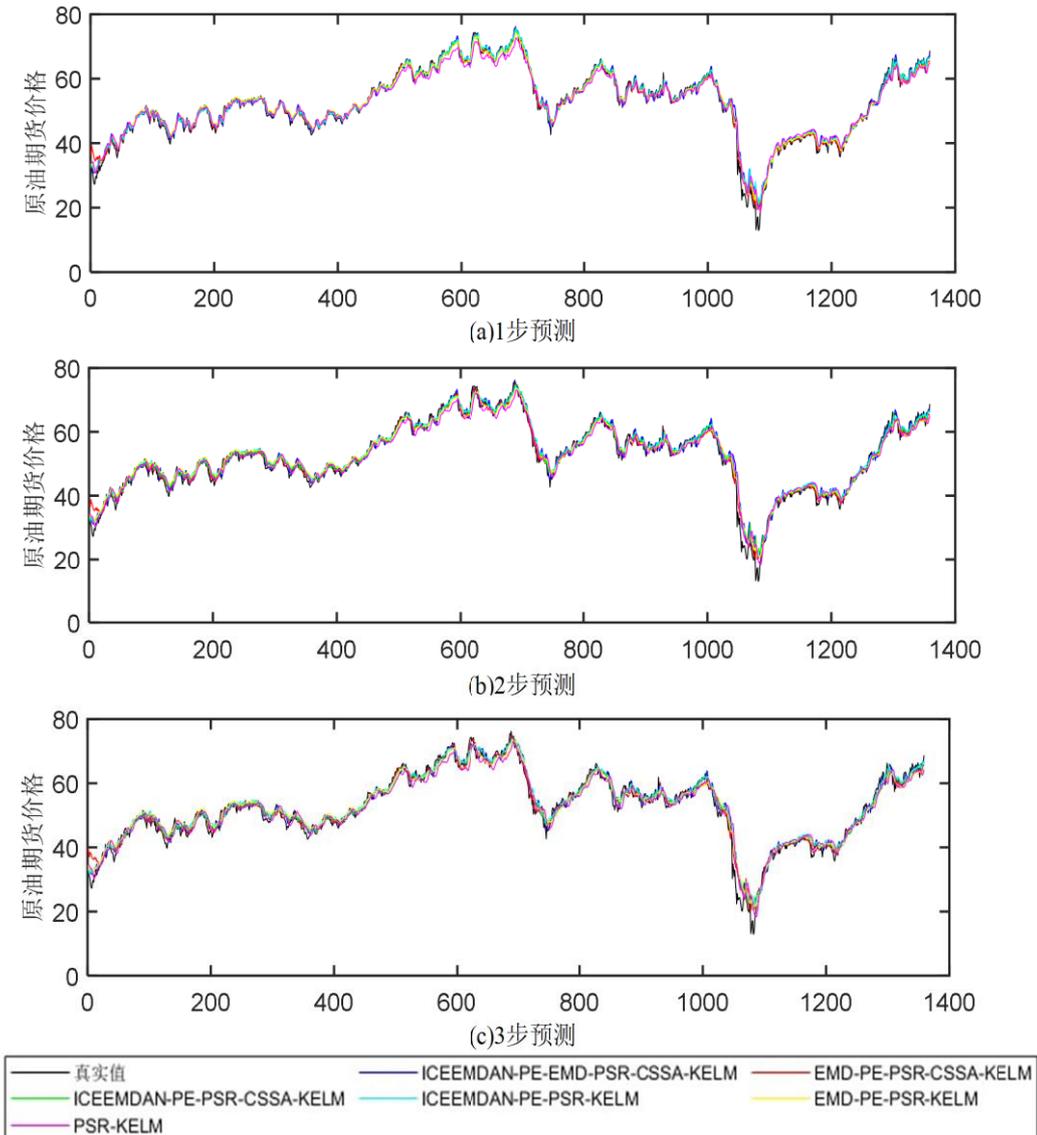


图 6.10 WTI 原油预测结果

(2) 为了提高对复杂序列的预测精度，我们采用 EMD 和 ICEEMDAN 两种不同的数据分解方法对原油期货价格序列进行分解，同时为了验证分解方法对预测模型的有效性，我们对 KELM 预测模型在不同分解方法下的预测性能进行了比较。从表 6.3 和表 6.4 中我们可以看出，相比未采用分解方法的模型及采用单次分解方法的模型来说，二次分解集成方法可有效降低预测的难度，提高模型的预测精度。以 Brent 原油期货价格的 1 步预测为例，与 PSR-KELM 模型相比，二次分解模型（ICEEMDAN-PE-EMD-PSR-CSSA-KELM）的评价指标 MAE、MAPE 和 RMSE 分别下降了 52%、49.8%和 44.8%，方向预测精度提高了 41%。同样，与单一分解模型 ICEEMDAN-PE-PSR-CSSA-KELM 相比，MAE、MAPE

和 RMSE 指数分别下降了 29.4%、39.7%和 20.5%，DS 值上升了 27.2%。从中可以看出，二次分解相比单次分解，提升水平更大，二次分解模型在整体预测方面表现最好。然而，在 3 步预测中，二次分解模型的表现不如单次分解模型，这表明，在我们的研究应用中，二次分解模型更适用于 1 步和 2 步预测。

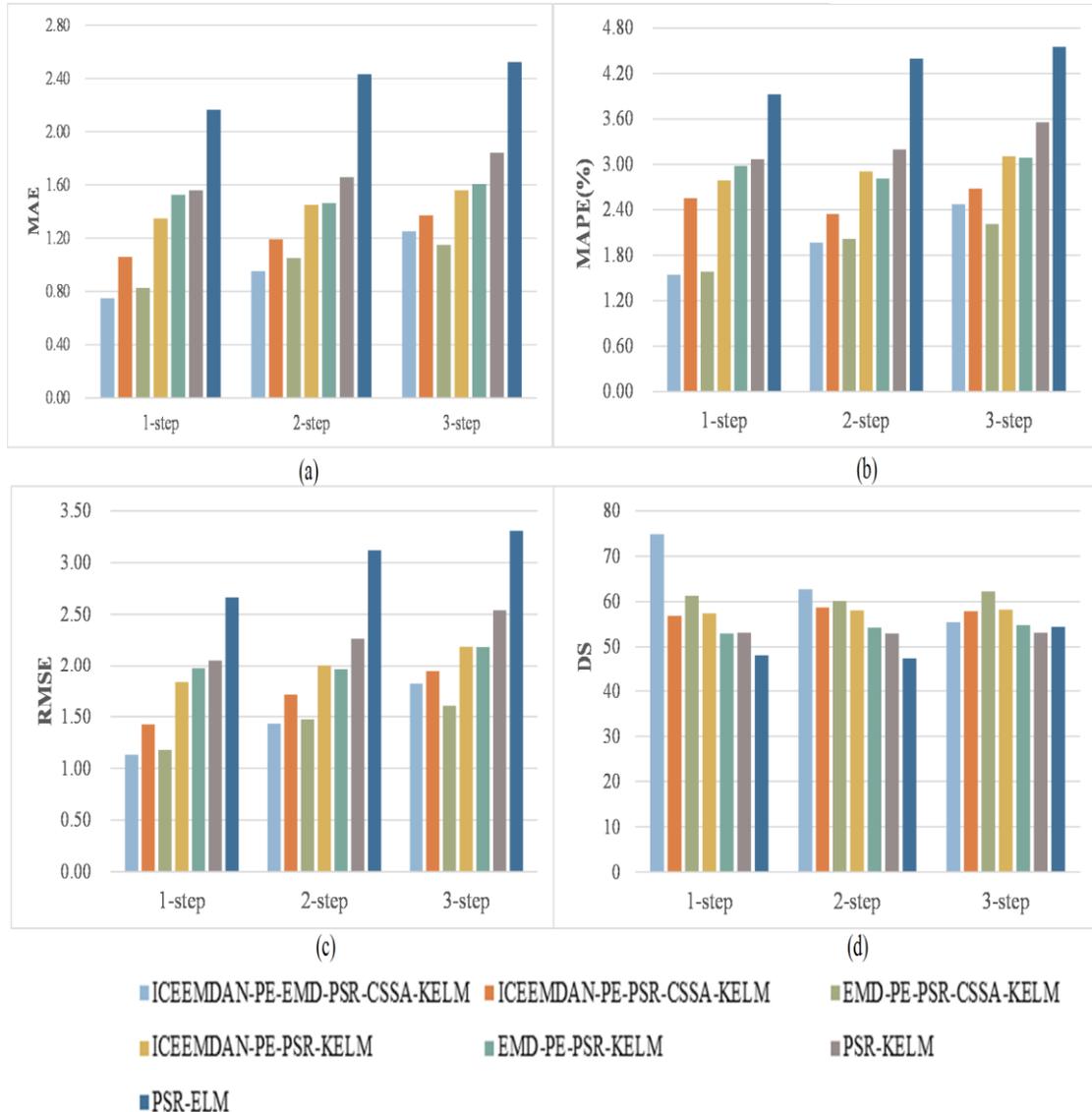


图 6.11 Brent 原油各模型评价指标

(3) 由于 KELM 模型易受核参数的影响，在本章中我们使用 CSSA 来优化 KELM 的核参数，通过比较使用 CSSA 优化核参数的模型（ICEEMDAN-PE-PSR-CSSA-KELM 和 EMD-PE-PSR-CSSA-KELM）及未优化的模型（ICEEMDAN-PE-PSR-KELM 和 EMD-PE-PSR-KELM）的预测效果，结果发现，使用 CSSA 优化后模型的预测效果有所提升。以 Brent 原油期货价格的 1 步预测为例，EMD-PE-PSR-CSSA-KELM 模型的水平预测精度 MAE、MAPE 和 RMSE 分别比未优

化模型降低了 45.8%、46.8%和 40.4%，相应的方向预测精度提高了 15.8%。同样，对于 ICEEMDAN-PE-PSR-CSSA-KELM 模型，MAE、MAPE 和 RMSE 分别比未优化模型 ICEEMDAN-PE-PSR-KELM 降低了 21.4%、8.3%和 22.3%，方向预测精度提高了 2.6%。

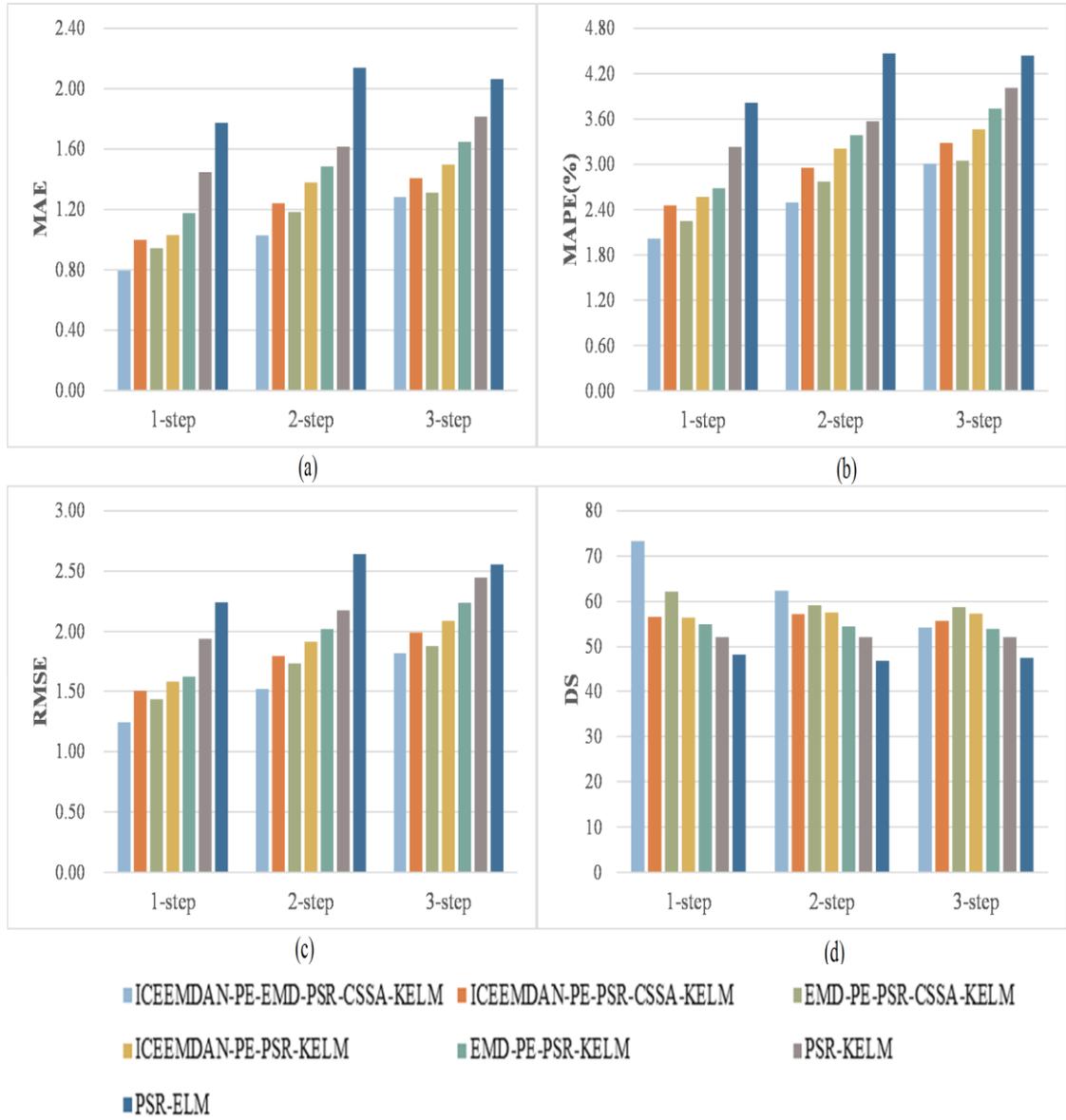


图 6.12 WTI 原油各模型评价指标

为从统计学意义上检验本章所构建模型与对比模型预测性能的差异性，进行了 DM 检验，检验结果如表 6.5-表 6.10 所示。这些结果证实，在 1 步预测的 90% 置信水平下，本章所构建的模型优于所有基准模型；对于多步预测，所构建的模型比大多数模型的预测效果好，这进一步反映出了该模型优越的预测性能。

表 6.5 Brent 原油 1 步预测 DM 检验结果

	Model 6	Model 5	Model 4	Model 3	Model 2	Model 1
Proposed	7.2305 (***)	1.1973 (*)	12.9170 (***)	15.6828 (***)	12.3918 (***)	16.7391 (***)
Model 6		6.5763 (***)	13.6636 (***)	16.0776 (***)	12.9978 (***)	17.8240 (***)
Model 5			12.7117 (***)	15.7339 (***)	11.6811 (***)	15.7039 (***)
Model 4				5.5299 (***)	5.7756 (***)	10.8899 (***)
Model 3					2.1216 (**)	9.8158 (***)
Model 2						9.8427 (***)

表 6.6 Brent 原油 2 步预测 DM 检验结果

	Model 6	Model 5	Model 4	Model 3	Model 2	Model 1
Proposed	7.5204 (***)	0.9211	10.7699 (***)	11.3827 (***)	9.5528 (***)	13.9830 (***)
Model 6		6.7099 (***)	9.8745 (***)	12.7854 (***)	10.3047 (***)	13.9133 (***)
Model 5			10.7969 (***)	9.2184 (***)	8.4682 (***)	12.5495 (***)
Model 4				1.4264	5.4544 (***)	9.9148 (***)
Model 3					5.7301 (***)	10.4319 (***)
Model 2						7.8843 (***)

表 6.7 Brent 原油 3 步预测 DM 检验结果

	Model 6	Model 5	Model 4	Model 3	Model 2	Model 1
Proposed	5.1869 (***)	4.1151 (***)	10.4334 (***)	9.7278 (***)	8.8179 (***)	13.9413 (***)
Model 6		7.1651 (***)	8.8643 (***)	11.6868 (***)	9.7723 (***)	15.2968 (***)
Model 5			9.5255 (***)	8.5803 (***)	8.1910 (***)	13.4796 (***)
Model 4				0.1609	5.5029 (***)	11.0599 (***)
Model 3					5.7995 (***)	10.9705 (***)

Model 2 7.5928
(***)

表 6.8 WTI 原油 1 步预测 DM 检验结果

	Model 6	Model 5	Model 4	Model 3	Model 2	Model 1
Proposed	5.7689 (***)	3.4577 (***)	6.2119 (***)	7.7565 (***)	11.2240 (***)	14.6210 (***)
Model 6		1.6119	4.7471 (***)	4.8310 (***)	9.5035 (***)	12.8602 (***)
Model 5			2.9094 (***)	4.1120 (***)	11.13303 (***)	13.0702 (***)
Model 4				1.1238	9.5808 (***)	11.4329 (***)
Model 3					10.4610 (***)	11.8812 (***)
Model 2						5.6651 (***)

表 6.9 WTI 原油 2 步预测 DM 检验结果

	Model 6	Model 5	Model 4	Model 3	Model 2	Model 1
Proposed	5.6382 (***)	3.4862 (***)	8.5834 (***)	10.9592 (***)	10.2484 (***)	15.7032 (***)
Model 6		1.3587	4.5786 (***)	7.1359 (***)	8.3555 (***)	13.4503 (***)
Model 5			4.2867 (***)	7.2565 (***)	7.9293 (***)	12.4870 (***)
Model 4				5.4212 (***)	6.8226 (***)	11.7539 (***)
Model 3					4.2912 (***)	10.6785 (***)
Model 2						10.6999 (***)

表 6.10 WTI 原油 3 步预测 DM 检验结果

	Model 6	Model 5	Model 4	Model 3	Model 2	Model 1
Proposed	5.1303 (***)	1.1115	7.8967 (***)	11.0868 (***)	10.0764 (***)	11.5689 (***)
Model 6		2.1218 (**)	5.8777 (***)	7.6167 (***)	8.37550 (***)	10.8488 (***)
Model 5			3.9356 (***)	9.4730 (***)	9.1796 (***)	9.4782 (***)
Model 4				6.1601 (***)	7.2441 (***)	7.6164 (***)
Model 3					4.7374 (***)	5.4217 (***)
Model 2						2.2614 (**)

6.6 本章小结

原油在全球经济发展过程中起着重要的作用,很多投资者和决策者十分关注原油价格的波动及走势,针对原油期货价格数据的非线性、非平稳性等特征,本章构建出一种基于二次分解重构集成技术及 CSSA 优化 KELM 的原油期货价格组合预测新方法 (ICEEMDAN-PE-EMD-PSR-CSSA-KELM)。

为验证本章所构建模型的有效性,通过对预测结果分析发现:(1)与 ELM 模型相比, KELM 模型具有更好的稳定性和突出的预测优势;(2)分解-重构-集成框架下的预测模型比单一预测模型精度更高;(3)使用 CSSA 优化 KELM 的核心参数时,预测能力也得到了提高,特别是对高频分量的二次分解可以有效降低预测数据的复杂度,从而有效提高预测精度。

总体来说,本章构建的 ICEEMDAN-PE-EMD-PSR-CSSA-KELM 模型既利用 PE 方法度量了子模态序列的复杂度,将各模态序列重构为高频项,低频项和趋势项,又考虑了高频项波动性大、复杂度高的特点,利用 EMD 对其二次分解降低了高频项的复杂性,还用 CSSA 对 KELM 的核参数进行了优化。与相应的基准模型对比,由于结合了 ICEEMDAN 分解算法、PE 重构算法及 CSSA 优化算法各自独特的优势,本章构建的模型表现良好。

7 结论与展望

7.1 结论

本文主要在基于分解重构集成框架下对金融领域的时间序列进行预测,包括美元兑人民币汇率预测、比特币价格预测和原油期货价格预测三个方面的内容。本文主要针对原始数据分解方法、分量重构方法和智能优化算法优化相关参数,构建了基于分解重构集成的美元兑人民币汇率预测、基于二次分解集成的比特币价格预测和基于二次分解重构集成的原油期货价格预测,具体研究的内容总结如下:

(1) 针对汇率数据的随机性、非线性、波动性等复杂特点,构建了一种汇率预测的新方法(EEMD-SE-PSO-LSSVM)。首先使用 EEMD 分解方法将原始美元兑人民币汇率序列分解为一系列的子序列;然后,通过 SE 量化各子序列的复杂度,将 SE 值较小的子序列进行重构;接着,通过相空间重构确定嵌入维数,然后通过 PSO-LSSVM 预测模型对各子序列分别进行了预测;最后,对各子序列的预测结果进行非线性集成,得到汇率的最终预测值。通过对本文模型和其他多种模型的预测结果对比,实证结果表明采用本文构建的模型预测结果更准确、预测精度更高。

(2) 基于“先分解后集成”的思想,构建了一种比特币价格预测的新方法(CEEMDAN-PE-CEEMDAN-PSR-CSSA-KELM)。首先,采用 CEEMDAN 分解方法将原始比特币价格序列分解为一系列的子序列;然后,通过 PE 算法将各子序列重构为高频项、低频项和趋势项;接着,对低频项和趋势项分别使用 CSSA 优化的 KELM 模型进行预测,得到低频项和趋势项的预测值,对高频项用 CEEMDAN 分解,将分解后的各子序列分别用 CSSA 优化的 KELM 模型预测,对高频项各分量的预测结果使用 CSSA-KELM 方法进行非线性集成,得到高频项的预测值;最后,对高频项、低频项和趋势项的预测结果使用 CSSA-KELM 模型进行非线性集成,得到比特币价格的最终预测值。为验证所构建模型(CEEMDAN-PE-CEEMDAN-PSR-CSSA-KELM)的有效性,通过与其它 11 种模型对比,发现本文构建的模型预测效果更好,DM 检验也进一步证明了该模型的稳健性。

(3) 基于“先分解后集成”的思想, 构建了一种原油期货价格预测的新方法 (ICEEMDAN-PE-EMD-PSR-CSSA-KELM)。首先, 使用 ICEEMDAN 分解方法将原始原油期货价格序列分解为一系列的子序列; 然后, 通过 PE 算法将各子序列重构为高频项、低频项和趋势项; 接着, 对低频项和趋势项分别使用 CSSA 优化的 KELM 进行预测, 得到低频项和趋势项的预测值。采用 EMD 分解方法对高频项进行二次分解, 将分解后的各子序列使用 PE 算法重构为 3 个分量, 对其分别用 CSSA 优化的 KELM 模型预测, 对各分量的预测结果使用线性集成的方法得到高频项的预测值; 最后, 对高频项、低频项和趋势项的预测结果使用 CSSA-KELM 方法进行非线性集成, 得到原油期货价格的最终预测值。为验证所构建模型的有效性, 本文对 Brent 原油期货和 WTI 原油期货价格进行了实证预测。实证结果表明, 本文构建的模型提高了原油价格预测的准确性, 相比其它基准模型, 本文的模型预测效果更好。

7.2 展望

尽管本文在分解重构集成范式下对金融相关领域的的数据做了一定的研究和取得了一些成果, 但本文也存在着不足, 不足之处可归纳如下:

(1) 本文构建的方法只利用时间序列数据进行预测, 缺乏对因变量影响因素的定性分析, 没有考虑到现实因素或突发事件对因变量的影响, 没有对其进行深入的分析 and 建模, 预测结果可能缺乏很好的解释性, 未来的研究中可把这些影响因素加入到预测方法中, 以提高模型的稳健性。

(2) 本文构建的模型研究内容较多, 为体现不同模型的应用场景, 并没有使用同一个数据应用于所有的模型, 在未来的研究中应尽可能多的使用同一个数据。

参考文献

- [1] 惠晓峰,柳鸿生,胡伟.基于时间序列 GARCH 模型的人民币汇率预测[J].金融研究,2003(5): 99-105.
- [2] 戴晓枫,肖庆宪.时间序列分析方法及人民币汇率预测的应用研究[J].上海理工大学学报,2005(4): 341-344.
- [3] Alessandro Lanza, Matteo Manera, Massimo Giovannini. Modeling and forecasting cointegrated relationships among heavy oil and product prices[J]. Energy Economics, 2005, 27(6):831-848.
- [4] 邓美玲,李小明,胡荣兴.国际石油价格之残差自回归模型短期预测[J].统计与决策,2008(22):146-147.
- [5] Atilim Murat, Ekin Tokat.Forecasting oil price movements with crack spread futures[J].Energy economics,2009,31(1):85-90.
- [6] 侯璐.基于 ARIMA 模型的石油价格短期分析预测[D],2009.
- [7] 秦鹏,缪柏其.基于广义指数预报因子的石油价格预测模型[J].系统工程理论与实践,2010,30(8):1389-1395.
- [8] Hassan Mohammadi, Lixian Su. International evidence on crude oil price dynamics: Applications of ARIMA-GARCH models[J]. Energy Economics, 2010,32(5):1001-1008.
- [9] Aijun Hou, Sandy Suardi.A nonparametric GARCH model of crude oil price return volatility[J].Energy economics,2012,34(2):618-626.
- [10] Rout M, Majhi B, Majhi R, et al. Forecasting of currency exchange rates using an adaptive ARMA model with differential based evolution[J]. Journal of King Saud University Computer & Information Sciences, 2013, 26(1):7-18.
- [11] 韩晓茹,林晓瑾. 基于 ARIMA 模型的人民币汇率预测与分析[J]. 佛山科学技术学院学报(自然科学版),2016(4): 8-12.
- [12] 孙鹏,程春梅.基于 ARIMA 模型的汇率预测研究——以美元人民币汇率为例 [J].辽宁工业大学学报(社会科学版),2016,18(06):20-23.
- [13] Abu Bakar N, Rosbi S. Autoregressive integrated moving average (ARIMA) model for forecasting cryptocurrency exchange rate in high volatility environment: A new

- insight of bitcoin transaction[J]. *International Journal of Advanced Engineering Research and Science*, 2017, 4(11): 130-137.
- [14] Demir E, Gozgor G, Lau C K M, et al. Does economic policy uncertainty predict the Bitcoin returns? An empirical investigation[J]. *Finance Research Letters*, 2018, 26: 145-149.
- [15] 杨焮,马洪波. 人工神经网络在中长期汇率预测中的应用[J]. *系统工程*, 1999(01): 3-5.
- [16] Chakradhara Panda, V. Narasimhan. Forecasting exchange rate better with artificial neural network[J]. *Journal of Policy Modeling*, 2007, 29(2): 227-236.
- [17] Svitlana Galeshchuk. Neural networks performance in exchange rate prediction[J]. *Neurocomputing*, 2016, 172: 446-452.
- [18] Kulkarni S, Haidar I. Forecasting Model for Crude Oil Price Using Artificial Neural Networks and Commodity Futures Prices[J]. *Computer Science*, 2009, 103-108.
- [19] Xie W, Yu L, Xu S, et al. A New Method for Crude Oil Price Forecasting Based on Support Vector Machines[J]. *Lecture Notes in Computational Science.*, 2006, 3994: 441-451.
- [20] Khashman A, Nwulu N I. Intelligent prediction of crude oil price using Support Vector Machines[C]. *2011 IEEE 9th International Symposium on Applied Machine Intelligence and Informatics*, 2011, 165-169.
- [21] 杨新臣, 吴仰儒. 基于支持向量机的非线性汇率预测分析[J]. *统计与决策*, 2010(18): 13-16.
- [22] 慕晓茜, 何佳, 倪旭敏, 段雨墨, 陆昊, 张作泉. 基于 SVR 的石油期货价格短期预测[J]. *科学技术与工程*, 2010, 10(18): 4585-4589.
- [23] 何树红, 杨博, 戴明爽. 基于动态递归神经网络的石油价格预测[J]. *云南民族大学学报: 自然科学版*, 2013, 22(1): 31-35.
- [24] 金艳凤. 基于 BP 神经网络的汇率预测模型研究(D). 武汉: 武汉理工大学, 2013.
- [25] 李秋敏, 田益祥, 张高勋. 基于 GMDH 的 LSSVM 预测模型及其实证[J]. *数学的实践与认识*, 2013, 43(6): 62-68.
- [26] Chiroma H, Abdulkareem S, Herawan T. Evolutionary Neural Network model

- for West Texas Intermediate crude oil price prediction[J]. *Applied Energy*, 2015, 142:266-273.
- [27] Greaves A, Au B. Using the bitcoin transaction graph to predict the price of bitcoin[J]. *Quoted*, 2015, 3: 22.
- [28] 范秋枫,王涛,张智峰. 量子粒子群智能算法在国际布伦特原油价格预测中的应用[J].*模糊系统与数学*,2017,31(04):84-90.
- [29] Indera N I, Yassin I M, Zabidi A, et al. Non-linear autoregressive with exogeneous input (NARX) Bitcoin price prediction model using PSO-optimized parameters and moving average technical indicators[J]. *Journal of fundamental and applied sciences*, 2017, 9(3S): 791-808.
- [30] McNally S, Roche J, Caton S. Predicting the price of bitcoin using machine learning[C]//2018 26th euromicro international conference on parallel, distributed and network-based processing (PDP). IEEE, 2018: 339-343.
- [31] Liu M, Li G, Li J, et al. Forecasting the price of Bitcoin using deep learning[J]. *Finance research letters*, 2021, 40: 101755.
- [32] Wang S Y, Yu L, Lai K K. Crude Oil Price Forecasting With TEI@I[J]. *Journal of Systems Science & Complexity*, 2005, 18(2):145-166.
- [33] 谢赤,郑林林,孙柏,等. 基于 EMD 和 Elman 网络的人民币汇率时间序列预测 [J]. *湖南大学学报(自然科学版)*,2006,36(6):89-92.
- [34] 谢赤,黄曦,孙柏,等. 基于小波分析与支持向量机的人民币汇率预测[J]. *湘潭大学学报(哲学社会科学版)*,2009,33(05):82-87.
- [35] 杨云飞,鲍玉昆,胡忠义等.基于 EMD 和 SVMs 的原油价格预测方法[J].*管理学报*,2010,07(12):1884-1889.
- [36] Rania Jammazi, Chaker Aloui. Wavelet decomposition and regime shifts: Assessing the effects of crude oil shocks on stock market returns[J].*Energy Policy*,2010,38(3):1415-1435.
- [37] 李成,周恒.原油价格改进型神经网络预测方法[J].*统计与决策*,2013(8):67-69
- [38] 戴伟.基于 EEMD 优化的国际油价分解集成预测模型研究[D],2015.
- [39] 王岚.基于 EMD_GA_KELM 的国际原油价格预测模型研究[D].重庆工商大

- 学,2015.
- [40] Yu L, D Wei, Ling T. A novel decomposition ensemble model with extended extreme learning machine for crude oil price forecasting[J]. *Engineering applications of artificial intelligence*, 2016,47:110-121.
- [41] He K, Chen Y, Tso G K F. Forecasting exchange rate using variational mode decomposition and entropy theory[J]. *Physica A: statistical mechanics and its applications*, 2018, 510: 15-25.
- [42] 陈焕. 基于 CEEMDAN-GWO-KELM 模型的国际原油价格预测研究[D].西南财经大学,2019.
- [43] 高海翔. 基于分解-深度集成学习算法的国际石油价格预测研究[D].北京化工大学,2020.
- [44] 梁少东. 基于分解集成模型优化的能源预测研究[D].北京化工大学,2020.
- [45] 崔金鑫,邹辉文.原油期货价格预测模型 CEEMDAN-PSO-ELM[J].*计算机系统应用*,2020,29(02):28-39.
- [46] Gyamerah S A. On forecasting the intraday Bitcoin price using ensemble of variational mode decomposition and generalized additive model[J]. *Journal of King Saud University-Computer and Information Sciences*, 2020.
- [47] 杨静凌, 唐国强,张建文.基于 CEEMD-Elman-Adaboost 组合模型的国际原油价格预测研究[J].*重庆理工大学学报(自然科学)*,2021,35(03):260-267.
- [48] Huang N E, Shen Z, Long S R, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis[J]. *Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences*, 1998, 454(1971): 903-995.
- [49] Wu Z H, Huang N E. Ensemble empirical mode decomposition: A noise-assisted data analysis method [J].*Advances in Adaptive Data Analysis*, 2009, 1(1): 1-41.
- [50] Torres M E, Colominas M A, Schlotthauer G, et al. A complete ensemble empirical mode decomposition with adaptive noise[C]//2011 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, 2011: 4144-4147.
- [51] Colominas M A, Schlotthauer G, Torres M E. Improved complete ensemble EMD: A suitable tool for biomedical signal processing[J]. *Biomedical Signal Processing*

- & Control, 2014, 14(1):19-29.
- [52] Richman J S, Moorman J R. Physiological time-series analysis using approximate entropy and sample entropy.[J]. American journal of physiology. Heart and circulatory physiology, 2000, 278(6):2039-2049.
- [53] Achmad Widodo, Min-Chan Shim, Wahyu Caesarendra, Bo-Suk Yang. Intelligent prognostics for battery health monitoring based on sample entropy[J]. Expert Systems With Applications,2011,38(9) : 11763-11769.
- [54] 曾发林,蔡嘉伟,孙苏民.基于 CEEMD 样本熵和 GA-BP 的排气噪声声品质预测[J].郑州大学学报(工学版),2020,41(06):19-25.
- [55] Bandt C,Pompe B. Permutation entropy: a natural complexity measure for time series[J].Physical Review Letters, The American Physiological Society,2002,88(17):174102.
- [56] Zunino L,Rosso O A,Soriano M C. Characterizing the hyper chaotic dynamics of a semiconductor laser subject to optical feedback via permutation entropy[J]. IEEE journal of selected topics in quantum electronics,2011,17(5) : 1250-1257.
- [57] Eberhart R,Kennedy J. A new optimizer using particle swarm theory[C]. Proceeding of the IEEE 6th International Symposium on Micro Machine and Human Science, 1995: 39-43.
- [58] Xue J, Shen B. A novel swarm intelligence optimization approach: sparrow search algorithm[J]. Systems Science & Control Engineering, 2020, 8(1): 22-34.
- [59] 吕鑫,慕晓冬,张钧,王震.混沌麻雀搜索优化算法[J].北京航空航天大学学报,2021,47(08):1712-1720.
- [60] Kim H S, Eykholt R,Salas J D,et al. Nonlinear dynamics,delay times,and embedding windows.Physica D:Nonlinear Phenomena.1999,127(1-2):48-60.
- [61] Suykens J A K, Vandewalle J. Least squares support vector machine classifiers[J]. Neural Processing Letters, 1999, 9(3):293-300.
- [62] Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: Theory and applications[J]. Neurocomputing, 2006, 70(1-3):489-501.
- [63] Huang G B, Wang D H, Extreme Y Lan learning machines: a survey[J].

International Journal of Machine Learning andCybernetics,2011,2(2):107-122.

- [64] Diebold F X, Mariano R S. Comparing predictive accuracy[J]. Journal of Business & Economic Statistics, 2002, 20(1):134-144.

硕士期间完成的工作

一、研究成果

- 1、孙景云,赵盼盼,丁毅. 基于样本熵重构和 PSO 优化算法的人民币汇率预测[J]. 数学的实践与认识,2021,(19):70-83.
- 2、Sun Jingyun, **Zhao Panpan**, Sun Shaolong. A new secondary decomposition-reconstruction-ensemble approach for crude oil price forecasting[J]. Resources Policy,2022,77.

二、获奖情况

- 1、2019 年 11 月获得兰州财经大学研究生三等奖学金。
- 2、2021 年 11 月获得第七届全国大学生统计建模大赛三等奖(研究生组)。

三、参与项目情况

- 1、2020 年参与兰州财经大学高等教育教学改革研究重点项目（项目编号：No.LJZ202008）。
- 2、2021 年参与甘肃省自然科学基金项目“市场波动率时变特性下最优动态配对交易模型的研究”（项目批准号：21JR1RA280）。

致谢

“逝者如斯夫，不舍昼夜”。不知不觉间，三年的研究生生涯即将结束，在这短暂的三年中，我收获了许多，也成长了许多。在此感谢曾经帮助过我的人，请接受我诚挚的谢意。

本论文是在导师孙景云的指导下完成的，从开题到论文写作，再到论文修改，孙老师都耐心指导，我每上一个台阶，都离不开孙老师的悉心教导，在此，对孙老师致以真挚的谢意，祝愿恩师绿水青山春常驻，桃李芬芳满天涯。感谢西安交通大学的孙少龙老师，在一些关键技术和框架的梳理方面给了我前所未有的帮助。

此外，还要特别感谢我的家人朋友，感谢他们对我的信任、理解与支持，他们的支持与关怀是我不断奋斗的动力源泉。

最后，向所有关心、爱护和支持我的人表示由衷的感谢。感谢百忙之中评阅论文以及毕业答辩的老师，感谢兰州财经大学。

赵盼盼

2022年5月15日