

分类号 _____
UDC _____

密级 _____
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 分解-集成框架下大豆期货价格的预测方法研究

研究生姓名: 何林芸

指导教师姓名、职称: 孙景云、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: 2023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： 何林芸 签字日期： 2023年5月30日

导师签名： 孙景云 签字日期： 2023年5月30日

导师(校外)签名： 严雪林 签字日期： 2023年5月30日

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： 何林芸 签字日期： 2023年5月30日

导师签名： 孙景云 签字日期： 2023年5月30日

导师(校外)签名： 严雪林 签字日期： 2023年5月30日

Research on Forecasting Method of Soybean Futures Price under Decomposition Integration Framework

Candidate : He Linyun

Supervisor: Sun Jingyun

摘要

农业是国民经济的基础，而大豆在我国农业中占据着重要的位置，大豆市场的波动影响着我国农业的发展乃至社会经济的稳定；期货在经济发展过程中能够有效地规避风险，提高市场的透明度。因此，对大豆期货价格的研究具有重要意义。

本文将数据分解方法与机器学习方法相结合，基于“分解-集成”框架下的大豆期货价格预测分为以下两个部分：

第一部分是基于分解去噪框架的大豆期货价格预测研究，对大豆期货价格进行分解去噪，并引入与大豆紧密相关的豆粕和玉米的期货价格信息对大豆期货价格进行预测。期货价格序列中含有噪声，这可能会掩盖豆粕期货价格与玉米期货价格对大豆期货价格的影响，因此在本文中对引入的豆粕期货价格与玉米期货价格进行去噪处理。结果表明，分解去噪可以有效地提升预测效果，而引入豆粕期货和玉米期货的信息能进一步提升模型的预测性能。

第二部分是基于“分解-重构-集成”框架的大豆期货价格预测研究，主要研究不同的重构方法对大豆期货价格预测的影响，并在此基础上引入其它期货市场的信息。本文的重构方法包括考虑复杂性、周期性和与原始序列的相关性三种数据特征的单特征重构与多特征重构。本文引入豆粕期货与玉米期货的价格信息，对豆粕期货价格与玉米期货价格进行分解重构，并对豆粕期货价格和玉米期货价格的重构序列与大豆期货重构序列进行相关性分析，从而筛选出预测大豆期货重构序列的输入变量。实证结果表明多特征重构方法优于单特征重构方法，引入其它期货市场信息能够有效提升模型的预测性能。

本文的研究层层深入，所提出的预测模型在大豆期货价格预测方面的良好表现为大豆期货价格预测提供了一种新的思路。

关键词：大豆期货 机器学习 分解集成 价格预测

Abstract

Agriculture is the foundation of the national economy, and soybeans occupy an important position in China's agriculture. The fluctuations in the soybean market affect the development of China's agriculture and even the stability of the social economy; Futures can effectively avoid risks and improve market transparency in the process of economic development. Therefore, the study of soybean futures prices is of great significance.

This thesis combines data decomposition methods with machine learning methods, and the soybean futures price prediction based on the "decomposition integration" framework is divided into the following two parts:

The first part is a study on soybean futures price prediction based on a decomposition denoising framework, which decomposes and denoises soybean futures prices and introduces futures price information of soybean meal and corn closely related to soybeans to predict soybean futures prices. The futures price sequence contains noise, which may mask the impact of soybean meal futures prices and corn futures prices on soybean futures prices. Therefore, in this thesis, the introduced soybean meal futures prices and corn futures prices are denoised. The results indicate that decomposition denoising can effectively improve the prediction performance, while introducing information from soybean meal futures and corn futures can further improve the predictive performance of the model.

The second part is a study on soybean futures price prediction based on the "decomposition reconstruction integration" framework, mainly studying the impact of different reconstruction methods on soybean futures price prediction, and introducing information from other futures markets on this basis. The reconstruction method in this thesis includes single feature reconstruction and multi feature reconstruction considering three data features: complexity, periodicity, and correlation with the original sequence. This thesis introduces the price information of soybean meal futures and corn futures, decomposes and reconstructs the prices of soybean meal futures and

corn futures, and analyzes the correlation between the reconstructed sequence of soybean meal futures prices and corn futures prices and the reconstructed sequence of soybean futures, thereby screening the input variables for predicting the reconstructed sequence of soybean futures. Empirical evidence shows that the multi feature reconstruction method is superior to the single feature reconstruction method, and introducing other futures market information can effectively improve the predictive performance of the model.

The research in this thesis is in-depth layer by layer, and the good performance of the proposed prediction model in predicting soybean futures prices provides a new approach for predicting soybean futures prices.

Keywords: Soybean futures; Machine learning; Decomposition integration; Price forecasting

目 录

1 引言	1
1.1 研究背景	1
1.2 研究意义	2
1.3 国内外研究现状	3
1.4 研究内容与论文框架	6
1.4.1 研究内容	6
1.4.2 论文框架	7
1.5 本文创新点	9
2 理论方法介绍	10
2.1 分解方法	10
2.2 重构方法	11
2.2.1 样本熵	11
2.2.2 近似熵	11
2.2.3 LZ 复杂度	12
2.2.4 综合贡献指数	13
2.2.5 快速傅立叶变换	14
2.3 预测方法	14
2.3.1 神经网络	14
2.3.2 支持向量回归	15
2.3.3 极限学习机	16
2.4 优化方法	17
2.5 本章小结	19
3 基于分解去噪框架的大豆期货价格预测	20
3.1 数据来源及模型评价指标	20
3.1.1 数据来源与描述统计	20
3.1.2 模型评价指标	21
3.2 基于单一模型的大豆期货价格预测	22
3.2.1 SSA-BP 模型预测	22
3.2.2 SSA-SVR 模型预测	23
3.2.3 SSA-ELM 模型预测	23
3.3 基于分解去噪的大豆期货价格预测	24
3.3.1 模型构建	24
3.3.2 模型预测	25
3.4 本章小结	30
4 基于“分解-重构-集成”框架的大豆期货价格预测	31
4.1 不同重构方法下的大豆期货价格预测	31

4.1.1 模型构建.....	31
4.1.2 模型预测.....	33
4.2 引入外生信息的大豆期货价格预测.....	38
4.2.1 模型构建.....	38
4.2.2 模型预测.....	40
4.3 本章小结.....	46
5 总结与展望	47
5.1 总结.....	47
5.2 展望.....	47
参考文献	49
后 记	53

1 引言

1.1 研究背景

期货交易是商品交易发展的产物，其诞生与演变经历了漫长的过程。1848年，世界上第一家期货交易所--芝加哥期货交易所诞生，开始交易玉米和小麦等农产品期货，为现代期货市场奠定了基础。中国期货交易的历史是从1840年后开始的，在那个艰难动荡的时期中国许多有志青年开始向西方学习，其中就包括开展期货交易。但是在纷乱的局势之下，中国期货的发展遭到了重重阻碍，在抗战胜利之后解放战争之前，中国期货市场经历了短暂的发展时期，而后直至我国实行改革开放，市场经济得以发展，期货市场才开始焕发活力。1990年，我国第一个商品期货市场--中国郑州粮食批发市场开业，标志着我国期货市场的初步形成。

期货是在现货的基础上发展出来的，与现货相对应。现货交易是人们常说的“一手交钱一手交货”的交易模式，而期货则不同，期货是一种在交易所内买卖标准化期货合约的交易方式。具体来讲，期货交易是在当下进行买卖，却在将来进行实物的交易，这种未来的实物交易被称为交割。期货可以理解作为一种约定和购买实物的权利。以大豆期货为例，大豆的价格会随着供需关系与市场状况波动，对于卖出大豆的人来说，若能在较高价位出售大豆则会获取更高额的回报，若在低价售出则可能会使出售方产生亏损，但对于大豆买方来讲则是低价购入更佳。大豆价格波动会导致买卖双方的损益出现相反的情况，这就需要一种可以平衡买卖双方收益和损失的约定，也就是期货。买方购买期货，购买的是一种权利，这种权力要求卖方在合约中约定的时间提供约定的实物，若卖方接受买方提出的交易价格，则该合约成立，并且在合约交割时间之前买方有权力对购买的期货进行转让和交易。期货交割时间比较灵活，一般以周、月、季度和年为单位。不同的期货种类有不同的主力合约，主力合约是指每日成交量和持仓量最大的合约，研究期货价格一般对该期货的主力合约进行研究。

我国作为农业大国和人口大国，农产品一直在我国经济中占据重要位置。作为四大农作物之一的大豆，一直以来都需要依靠进口来满足人们的需求，这导致我国大豆价格与国际大豆市场紧密相关。在 2018 年的中美贸易战中我国大豆进口量出现了七年来首次下降，而后我国出台许多政策振兴国产大豆，促进大豆行业的良性发展。2019 年，中央一号文件提出实施大豆振兴计划。据国家统计局数据和中国海关总署数据，2020 年，全国大豆产量约为 1960 万吨，同比增长 8.3%。2021 年农业农村部发布意见稳定大豆生产，2021 年我国大豆的进口量小幅下降，降至 9652 万吨，同比减少 3.8%。2022 年，全国大豆产量约为 2029 万吨，同比增长 23.7%，中国大豆进口量 9108 万吨，同比减少 5.6%。2022 年发布的中央一号文件指出，要大力实施大豆产能提升工程，促进大豆期货市场的发展。由于我国大豆对于国际市场的高度依赖性以及不断攀升的产量，大豆市场具有很大的波动风险，合理利用大豆期货的功能成为重点，因此对大豆期货的研究极具价值。

1.2 研究意义

从理论上讲，期货市场和其他金融市场一样，受到国家政策、宏观经济、工业经济和投资者心理等复杂因素的影响。此外，商品期货市场还将受到商品品种产业链上下游各种因素的影响，因此商品期货价格变化复杂，这对预测商品期货价格走势是一个巨大的挑战。目前对大豆期货价格的预测研究，多是直接根据影响因素进行预测，或只利用大豆期货价格原始数据进行预测，较少地考虑分解方法与机器学习方法相结合以及引入其它期货市场信息。因此，本文既考虑到机器学习算法在预测非平稳时序数列的优越性，引入智能算法对其中参数进行优化，又将分解方法和其它期货市场信息引入到预测模型当中，对大豆期货价格预测体系的建设来说，是具有推动和引导作用的。

从现实来看，作为世界公认的农业大国，中国农产品在中国经济中占据重要位置，农产品期货则是农产品产业链的重要组成部分。除此之外，农产品期货在我国期货市场上占有较大的交易量，价格波动影响深远。从微观角度来讲，买卖双方可以通过期货套期保值规避风险，同时农产品期货能为农产品市

场提供预警依据，农产品相关公司可以根据农产品期货信息规划生产经营。从宏观角度来看，农产品期货能够在一定程度上稳定农业市场，从而有助于稳定经济，同时为政府制定宏观经济政策提供依据。因此农产品价格预测模型是有现实意义的研究课题，而大豆作为我国具有代表性的大宗农产品之一，对其建立预测模型是有现实意义的。

1.3 国内外研究现状

时间序列预测一直以来都是研究热点，通过文献梳理，相关的预测模型大致可分为三种主要类型：传统统计模型、机器学习模型和分解集成模型。传统统计模型广泛地用于时间序列预测。滕永平等^[1]介绍了 ARMA 模型在期货价格预测中的作用，并通过对具体的大豆期货价格进行预测，认为 ARMA 模型在期货价格短期预测方面有一定的准确性。陈林等^[2]将 ARIMA 方法运用于期货价格预测，表明 ARIMA 在期货价格的短期预测中非常有效。刘峰等^[3]将对白菜月价格用 ARIMA 方法进行预测也得到了良好的预测效果。李苏等^[4]用 ARMA 模型对通过季节模型提取的猪肉价格趋势循环序列进行预测，预测效果理想。丁慧娟^[5]用 ARIMA 模型和灰色模型对遵义猪肉价格进行预测，并比较了两种模型的预测效果，实证结果表明 ARIMA 模型的短期预测效果较好，而灰色模型更适用于长期预测。郭静利等^[6]利用 SARIMA 模型对国际稻米价格进行预测，价格显示 SARIMA 模型较好地预测了国际稻米价格趋势。邓美玲等^[7]用残差自回归模型对国际石油价格进行预测，结果显示，残差自回归模型优于其它预测模型。何文琪等^[8]将 ARIMA 模型与线性回归模型相组合的组合模型对煤炭价格进行了较好的拟合。Ge 等^[9]分别建立了以时间为自变量的一元非线性回归模型和以生产消费、进出口量为自变量的多元线性回归模型来预测玉米价格，实证结果表明提出模型能够在一定程度上预测玉米价格，但精度较低，只能应用于粗略预测。

传统统计模型建立在统计理论上，因此在建立传统统计模型之前，必须满足某些假设条件，这将限制传统统计模型在实际问题中的应用。而价格时间序列多是不稳定非线性序列，传统的统计模型并不总能取得良好的预测效果。

而机器学习模型则不需要满足统计假设，也不需要为预测模型提供明确的公式，它们已广泛应用于许多领域。段青玲等^[10]采用遗传算法对支持向量回归模型的参数进行优化，构建 GA-SVR 模型预测水产品价格，提升了预测效果。郑薇等^[11]用混合灰狼算法优化的支持向量回归（HGWO-SVR）对农产品价格进行预测，实证结果显示相较于基准模型，优化后的 SVR 模型有更好的预测性能。Yu 等^[12]用遗传算法（GA）优化的 BP 神经网络对农产品价格进行预测，实证结果显示相比于基准模型，该机器学习模型对农产品价格有更好的预测。范俊明等^[13]通过长短期记忆网络（LSTM）对大豆期货价格进行预测，结果显示 LSTM 模型在拟合优度指标上有良好的表现。卢超凡等^[14]用典型相关分析与主成分分析选择对苹果价格有较大影响的因素，然后将选择的变量作为长短期记忆网络（LSTM）模型的输入进行价格预测，得到了很好的预测效果。张旻等^[15]用人工蜂群算法优化的极限学习机预测模型（ABC-ELM）对上证综指收盘价格进行预测，得到了较好的预测效果。沈虹等^[16]用长短期记忆网络（LSTM）模型对 6 种有色金属期货价格分别进行长短期预测，都取得了很好的预测效果。陈俊华等^[17]通过深度信念网络（DBN）对原油期货市场价格进行了非常有效的预测。沈欣怡等^[18]采用多种机器学习方法对铜期货价格进行预测，都得到了较优的预测效果，这表明机器学习模型在期货市场价格预测中具有良好的适用性。

第三类方法是分解集成模型。分解集成模型表示将分解方法与预测方法相结合的预测模型。目前，分解方法大多用于数据去噪预测和数据分解重构集成预测，这种分解集成的思想可以简化复杂的问题，降低建模难度，提高预测精度。Zhao^[19]在分析影响农产品期货价的基本因素的基础上，使用小波分析方法对数据进行平滑处理，然后建立模型来处理信号分解后的信息，最终得到了很好的预测效果。Wang 等^[20]用奇异谱分解（SSA）、经验模式分解（EMD）和变分模式分解（VMD）对序列进行去噪，然后用 ARIMA 模型、支持向量回归（SVR）、递归神经网络（RNN）、选通递归神经网络（GRU）和长短期记忆神经网络（LSTM）五种预测模型预测去噪序列，生成了不同的预测子模型，再将预测子模型组合得到最终预测效果，实证表明分解方法和预测方法的结合得到了更高精度的预测。曹霜等^[21]用小波分解对农产品价格进行分解，然后用

SVM 模型与 ARIMA 模型对分解序列进行预测，最后将预测分解序列进行集成，结果表明较于传统的时间序列预测模型，其提出的分解集成模型表现更优。王珏等^[22]将奇异谱分解方法与极限学习机算法相结合，对玉米，原油，黄金期货价格进行了预测，实验结果表明从水平预测精度与方向预测精度上来讲，该模型都能达到很好的预测效果。方雪清等^[23]将集合经验模式分解（EEMD）与长短期记忆神经网络（LSTM）相结合，构建了一个农产品价格短期预测模型，并且引入了 Fine-to-coarse 重构方法对分解序列进行重构，实证表明提出的 EEMD-LSTM 模型在农产品价格短期预测问题上具有一定的性能优势。Fang 等^[24]采用集合经验模式分解（EEMD）技术对六种不同类别的农产品期货价格进行分解，并用不同的机器学习方法预测分解分量再进行集成，结果表明 EEMD 组合模型的预测性能优于单个模型。Yu 等^[25]在分解基础上，对分解序列进行成分特征分析以探索隐藏的数据特征并由其数据特征选择预测方法，得到了很好的预测效果。Zhu 等^[26]提出了一种融合经验模态分解（EEMD）和最优组合预测模型（CFM）的混合预测方法，对原始序列进行分解后根据每个分解序列的综合贡献指数（CCI）对分解序列进行重构，再选择最优模型对重构序列进行预测再集成得到最终结果，实证表明该模型具有良好的预测性能。Niu 等^[27]建立了一个基于误差修正策略、分治策略的组合预测系统来预测碳价格序列；采用变分模式分解（VMD）算法对原始数据进行分解并利用样本熵（SE）来重构模式分解序列，实证表明提出方法优于考虑的其他基准方法。

通过对现有时间序列预测方法及相关研究的有序梳理可以看出基于“分解-集成”框架下的预测模型是时序预测研究的前沿技术。本文将分解方法与预测方法相结合，构建了分解去噪框架下的预测模型和基于“分解-重构-集成”框架下的预测模型，并引入了其它农产品期货市场的信息，将为预测研究提供新的创新思路。

1.4 研究内容与论文框架

1.4.1 研究内容

本文选用大豆期货价格作为研究对象，在研究过程当中阅读了大量国内外相关文献，研究了时间序列及农产品期货价格预测方法，并对时间序列预测的主流方法进行总结分析，在此基础上形成了本文的预测方法并对大豆期货价格进行建模预测分析。本文对大豆期货价格时间序列分别构建了单一模型、分解去噪模型、引入外生信息的分解去噪模型、不同重构方法下的“分解-重构-集成”预测模型以及引入其它期货市场信息的“分解-重构-集成”预测模型，各个模型层层递进。

具体来讲，本文基于“分解-重构”框架下有关大豆期货价格的预测分为两个部分。第一部分是基于分解去噪框架下的大豆期货价格预测研究。首先用麻雀搜索算法优化的神经网络（SSA-BP）、支持向量回归（SSA-SVR）和极限学习机（SSA-ELM）三种方法对大豆期货价格进行预测；然后通过使用改进自适应噪声完备集合经验模态分解算法（ICEEMDAN）将大豆期货收盘价格序列进行分解，成功得到多个分解序列，将复杂度最高的分解序列视为噪声序列并舍弃，其它分解序列加总形成大豆期货价格去噪序列。引入豆粕期货和玉米期货的信息，同样对这两种期货价格进行分解去噪并将其作为大豆去噪模型的输入变量，对大豆期货价格去噪序列进行预测得到大豆期货的预测结果。第二部分是基于“分解-重构-集成”框架下的大豆期货价格预测研究，主要研究不同的重构方法对大豆期货价格预测的影响，并在此基础上引入了豆粕期货价格信息和玉米期货价格信息。首先通过使用ICEEMDAN分解算法将大豆期货收盘价格序列进行分解，得到多个分解序列，并从周期性、复杂性和与原序列的相关性三个方面分析分解序列的特征，分别根据各个分解序列的特征进行单特征重构和多特征重构，然后对重构序列进行预测，最后对重构序列的预测结果进行非线性集成得到最终的结果。除此之外，在对分解序列使用不同方法重构的基础上，引入其它自变量--豆粕期货收盘价与玉米期货收盘价，首先通过使用

ICEEMDAN 分解算法将大豆期货收盘价格序列、豆粕期货收盘价格序列与玉米期货收盘价格序列分别进行分解，分别得到多个分解序列并对其进行重构得到重构序列，然后对豆粕期货与玉米期货的重构序列与大豆期货重构进行相关性分析，确定大豆期货重构序列的自变量并对其进行预测，最后将预测的大豆期货重构序列进行非线性集成得到大豆期货收盘价最终的预测效果。

1.4.2 论文框架

本文的主要研究内容分为五个部分，论文的框架如图 1.1 所示。

第一章：引言。本章介绍了大豆期货预测的研究背景与研究意义，并分别从传统的时间序列模型，机器学习模型，分解集成模型三个角度对国内外相关的研究文献及研究现状进行了阐述，说明了本文的研究框架，最后指出了本文的创新之处。

第二章：理论方法介绍。本章介绍了本文使用到的相关模型与方法的理论基础，包括数据分解方法、数据重构方法、数据预测方法与优化算法的原理。

第三章：基于分解去噪框架的大豆期货价格预测。本章构建了基于分解去噪框架下的预测模型。首先，利用改进的自适应噪声完备集合经验模态分解（ICEEMDAN）分解方法对原序列进行分解去噪，再用不同的机器学习方法对去噪序列进行预测，最后，引入其它期货市场信息，得到最终的预测结果。

第四章：基于“分解-重构-集成”框架的大豆期货预测。本章构建了基于“分解-重构-集成”框架下的预测模型。该章主要分为两个部分，第一部分研究不同的重构方法对大豆期货价格预测的影响。第二部分在第一部分的基础上，引入其它农产品期货价格的信息--豆粕期货收盘价与玉米期货收盘价。

第五章：总结与展望。本章总结了全文的研究成果，指出了研究工作的不足，并提出了一些改进建议，作为进一步研究的展望。

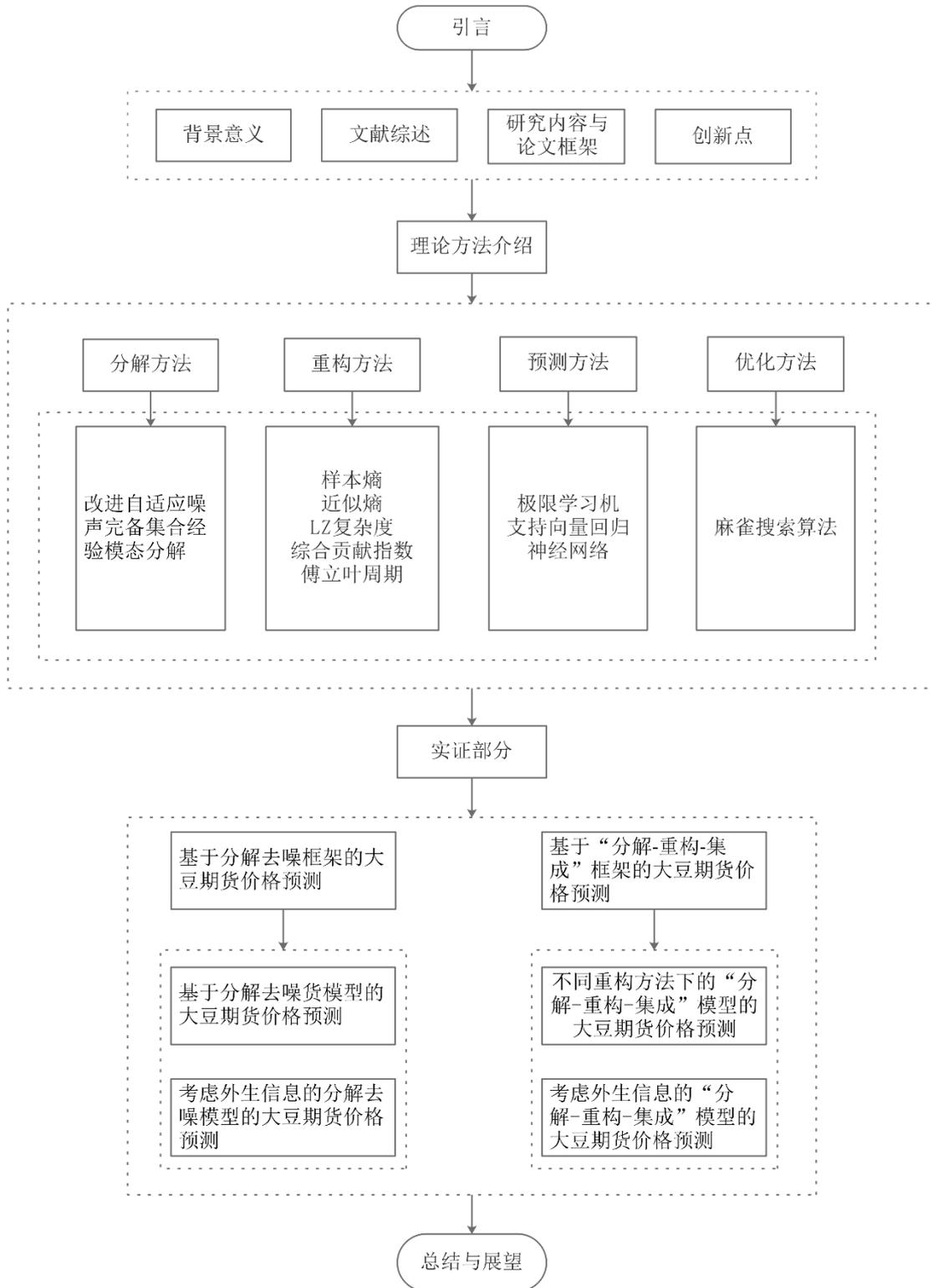


图 1.1 论文框架示意图

1.5 本文创新点

本文的创新点体现在以下三个方面：

(1) 将分解方法引入到大豆期货价格的预测当中，对大豆期货价格进行去噪处理，并引入了豆粕期货和玉米期货的信息。对豆粕期货价格和玉米期货价格也进行了去噪处理，并作为预测大豆期货去噪序列的输入变量，从而对大豆期货价格进行预测。

(2) 在“分解-重构-集成”的框架下探讨了基于三种数据特征（复杂性、周期性和相关性）的重构方法，并提出了将基于多种数据特征重构方法的“分解-重构-去噪”模型对大豆期货价格进行预测。

(3) 在基于多种数据特征重构方法的“分解-重构-去噪”模型的基础上，引入了豆粕期货信息和玉米期货信息。对豆粕期货价格和玉米期货价格都进行了分解重构，并对豆粕期货价格与玉米期货价格的重构序列和大豆期货价格重构序列进行相关性分析，根据两种规则筛选出预测大豆重构序列的输入变量。其中，规则一是选取与大豆期货重构序列相关性最强的豆粕期货重构序列和玉米期货重构序列作为其预测的输入变量，规则二则是筛选与大豆期货重构序列在统计意义上显著相关的豆粕期货重构序列和玉米重构序列作为其输入变量。

2 理论方法介绍

2.1 分解方法

本文的分解方法选用改进自适应噪声完备集合经验模态分解 (ICEEMDAN)。自适应噪声完备集合经验模态分解 (ICEEMDAN) 是EMD分解方法簇中的一种分解方法。EMD^[28]方法是一种基于数据自身的特征将原始信号分解为一系列不同特征尺度的分解序列的自适应的信号分解方法, 这些分解序列包含本征模态函数(IMF)以及残余项Res。EMD方法很适用于具有非线性波动的非平稳性时间序列。然而, EMD分解方法存在一定的缺陷, 如模态混叠和端点效应。经过改进的集合经验模态分解(EEMD)^[29]、完备集合经验模态分解(CEEMD)^[30]和自适应噪声完备集合经验模态分解(CEEMDAN)^[31]方法通过在原始信号中加入白噪声信号对EMD方法进行了步步改进。目前, 改进自适应噪声完备集合经验模态分解 (ICEEMDAN)^[32]是EMD分解方法簇中较优的方法, 该方法降低了IMF中的残余噪声, 还解决了分解阶段的虚假成分和模态混叠问题。ICEEMDAN的具体分解过程可以描述如下:

(1) $w^{(i)}$ 表示要添加的第 i 个白噪声, $E_k(w^{(i)})$ 表示对白噪声 $w^{(i)}$ 进行EMD分解后得到的第 k 个IMF分量, β_{k-1} 表示第 k 次添加的白噪声序列标准差。将第一个白噪声分量 $E_1(w^{(i)})$ 加到原始信号 x 上, 得到如下信号 $x^{(i)}$:

$$x^{(i)} = x + \beta_0 E_1(w^{(i)}) \quad (2.1)$$

(2) 用EMD分解得到 $x^{(i)}$ 的局部均值, 取其平均值得到第一个残差 $r_1 = \frac{1}{T} \sum_{i=1}^T M(x^{(i)})$ 。其中, $M(\cdot)$ 表示信号的局部均值, T 为添加的白噪声序列个数。然后, 原始序列 x 的第一个IMF值可以通过 $c_1 = x - r_1$ 来计算。

(3) 第二个模态分量值(IMF2)可通过以下公式计算:

$$c_2 = r_1 - r_2 \quad (2.2)$$

其中 $r_2 = \frac{1}{T} \sum_{i=1}^T M(r_1 + \beta_1 E_2(w^{(i)}))$ 。

(4) 同样, 根据 $c_k = r_{k-1} - r_k$ 计算第 k 个IMF值, 其中

$$r_k = \frac{1}{T} \sum_{i=1}^T M(r_{k-1} + \beta_{k-1} E_k(w^{(i)})) \quad (2.3)$$

2.2 重构方法

2.2.1 样本熵

样本熵(Sample Entropy, SE)^[33]是度量时间序列复杂程度的一种方法。样本熵值越大,说明该序列的复杂程度越高,反之亦然。本文使用样本熵来度量分解后各分解序列的复杂度,为序列的重构提供依据。具体的样本熵计算过程如下:

- (1) 对时间序列 $\{y_i\} i=1,2,\dots,N$, 取相邻 m 个 y_i 组成新的序列

$$Y_i = [y_i, y_{i+1}, \dots, y_{i+m-1}], i=1,2,\dots,N-m+1 \quad (2.4)$$

- (2) 按下式计算距离

$$D_m(Y_i, Y_j) = \max\{|y_{i+l} - y_{j+l}|\}, l=0,1,2,\dots,m-1. \quad (2.5)$$

其中 $j=1,2,\dots,N-m$ 且 $i \neq j$ 。

- (3) 给定阈值 r , 计算 Y_i 与 Y_j 的距离小于 r 的序列个数 B_i 在总序列个数中的占比:

$$B_i^m(r) = \frac{B_i}{N-m}. \quad (2.6)$$

- (4) 计算 $B_i^m(r)$ 的均值 $B^m(r)$

$$B^m(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} B_i^m(r) \quad (2.7)$$

将维数从 m 增加到 $m+1$, 重复上述步骤, 得到 $B^{m+1}(r)$

- (5) 计算样本熵值

$$SE(m, r) = \lim_{N \rightarrow \infty} \left\{ -\ln \frac{B^{m+1}(r)}{B^m(r)} \right\} \quad (2.8)$$

- (6) 当 N 为有限值时, 熵值的计算公式如下

$$SE(m, r, N) = -\ln \frac{B^{m+1}(r)}{B^m(r)} \quad (2.9)$$

2.2.2 近似熵

近似熵(Approximate Entropy, ApEn)概念最早是由 Steven M. Pincus^[34]提出

的，用于刻画数据的复杂性，具体的近似熵计算过程如下：

(1) 对于一个长度为 N 时间序列： $u(1), u(2), \dots, u(N)$ ， m 表示窗口长度，生成维数为 m 的一组向量： $x(1), x(2), \dots, x(N-m+1)$

其中， $x(i) = \{u(i), u(i+1), \dots, u(i+m-1)\}, i = 1 \rightarrow N-m+1$

(2) $x(i)$ 和 $x(j)$ 的距离 $d[x(i), x(j)]$ 为两者对应元素中差值最大的一个，即：

$$d[x(i), x(j)] = \max_{k=0 \rightarrow m-1} [|x(i+k) - x(j+k)|] \quad (2.10)$$

(3) 对每一个 i 值计算 $x(i)$ 与 $x(j), j = 1 \rightarrow N-m+1$ 间的距离

(4) 设定阈值 r ，对每一个 i 值计算距离 d 小于 r 的个数，计算该个数与 $N-m$ 的比值，记为 $C_i^m(r)$ ，即：

$$C_i^m(r) = \frac{1}{N-m} \{d[x(i), x(j)] \leq r\} \quad (2.11)$$

(5) 将 $C_i^m(r)$ 取对数，并求其对所有 i 的均值，记作：

$$\phi^m(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \ln C_i^m(r) \quad (2.12)$$

(6) 增加 m 维数，变为 $m+1$ ，重复步骤 2 至步骤 5，得到 $C_i^{m+1}(r)$ 和 $\phi_i^{m+1}(r)$

(7) 得到该时间序列的近似熵为：

$$ApEn(m, r) = \lim_{N \rightarrow \infty} [\phi_i^m(r) - \phi_i^{m+1}(r)] \quad (2.13)$$

2.2.3 LZ 复杂度

Lempel A 和 Ziv J^[35]于 1976 年提出了一种复杂度算法，称为 Lempel-Ziv 复杂度。Lempel-Ziv 复杂度算法简要描述如下^[36]：

(1) 对于时间序列 $S = (s_1, s_2, \dots, s_n)$ ($s_i, i \in 1, 2, \dots, n$)，令 S, Q 分别代表两个字符串， SQ 表示把 S, Q 两个字符串相加组成的总字符串， SQP 表示把 SQ 中最后一个字符删去所得的字符串。令 $V(SQP)$ 表示 SQP 的所有不同的子串的集合。

(2) 将 $c(n), S, Q$ 的初始化为 $c(n) = 1, S = s_1, Q = s_2$ ，因此 $SQP = s_1$ 。假定

$S = s_1, s_2, \dots, s_r, Q = s_{r+1}$, 若 $Q \in V(SQP)$, 则 S 不变, 将 Q 更新为 $Q = s_{r+1}s_{r+2}$, 再判断 Q 是否属于 $V(SQP)$, 重复步骤直到发现 $Q \notin V(SQP)$ 时为止。设此时 $Q = s_{r+1}, s_{r+2}, \dots, s_{r+i}$, 即表明 $s_{r+1}, s_{r+2}, \dots, s_{r+i}$ 不是 $s_1, s_2, \dots, s_r, s_{r+1}, \dots, s_{r+i-1}$ 的子串, 因此 $c(n)$ 值将加 1。然后将上述 Q 组合到 S 中, 使 S 更新为 $S = s_1, s_2, \dots, s_r, s_{r+1}, \dots, s_{r+i}$, 而取 Q 为 $Q = s_{r+i+1}$ 。

(3) 重复以上步骤, 直到 Q 取到最后一位为止, 并计算“归一化复杂度”

C_{LZN} :

$$\lim_{n \rightarrow \infty} c(n) = b(n) = \frac{n}{\log_2^n} \quad (2.14)$$

$$C_{LZN}(n) = \frac{c(n)}{b(n)} \quad (2.15)$$

2.2.4 综合贡献指数

为了衡量每个分解序列对原始序列的重要性, 本文引入了综合贡献指数 (CCI) [26]。假若一个长度为 N 的序列被分解成立多个分解序列, 第 p 个分解序列的 CCI 可以定义为:

$$CCI_p = \frac{1}{T} \sum_{t=1}^T \frac{|x(t) - \tilde{x}(t)|}{x(t)} + \left(1 - \frac{1}{T-1} \sum_{t=1}^{T-1} R(t) \right) \quad (2.16)$$

其中 $\frac{1}{T} \sum_{t=1}^T \frac{|x(t) - \tilde{x}(t)|}{x(t)}$ 表示去除第 p 个分解序列对原始序列的相对误差的影响,

$1 - \frac{1}{T-1} \sum_{t=1}^{T-1} R(t)$ 表示去除第 p 个分解序列对于原始序列的方向精度的影响。

$R(t)$ 是一个特征函数, 如果去掉第 p 个分解序列的序列具有与原始序列相同的趋势, 则 $R(t) = 1$, 否则 $R(t) = 0$ 。

$$\tilde{x}(t) = \sum_{k=1, k \neq p}^N g_t(k) + r(t) \quad (2.17)$$

$$R(t) = \begin{cases} 1, & \text{if } (x(t+1)-x(t)) \cdot (\tilde{x}(t+1)-x(t)) \geq 0 \\ 0, & \text{otherwise} \end{cases} \quad (2.18)$$

CCI_p 值越大, 第 p 个分解序列对原始序列越重要。

2.2.5 快速傅立叶变换

傅立叶分析是一种常用的信号分析方法。傅立叶转化是傅立叶分析的一个基本部分, 它可以对信号在时域和频域之间进行转换。对于一个时域信号, 经多个正弦信号叠加可分析正弦波的频率, 从而将时域变换为频域, 从而在频域中可分析出不同时域下的信号特征。在实际应用中, 需要对在时域和频域的信号信息进行离散化处理, 而离散周期信号的傅立叶变换符合此要求。为了确保其有限性, 仅在时间和频域的离散周期信号上采取离散傅里叶变换 (Discrete Fourier Transform, DFT) 方法, 对一个长度为 N 的信号, 具体表达式如下:

$$F(j) = \sum_{i=0}^{N-1} f(i) \cdot e^{-j \frac{2\pi}{N} ij}, j = 0, 1, \dots, N-1 \quad (2.19)$$

相应的离散逆傅立叶变换公式如下:

$$f(j) = \frac{1}{N} \sum_{j=0}^{N-1} f(j) \cdot e^{-j \frac{2\pi}{N} ij}, i = 0, 1, \dots, N-1 \quad (2.20)$$

式中, $f(i)$ 和 $F(j)$ 分别代表在时域和在频域的离散信号。然而, DFT 方法计算效率低, 常通过快速傅里叶变换(Fast Fourier Transform, FFT)^[37]来提升运算效率。在本文中, 将通过快速傅立叶变换计算出的数据周期简称为傅立叶周期。

2.3 预测方法

2.3.1 神经网络

BP 神经网络(Back-Propagation Neural networks, BPNN)由 Rumelhart 和

McClelland 等^[38]提出，被普遍应用于时间序列预测。BP 神经网络由输入层、隐含层及输出层的三个部分组成，通过反向传播误差来不断调整网络的权值与阈值，直至使网络的最后输出误差达到允许范围内。

BP 神经网络的结构如图 2.1 所示，表中 x_1, x_2, \dots, x_n 为 BP 神经网络输入， y_k 为神经网络预测值， ω_{ij} 与 ω_{jk} 分别表示输入层与隐含层、隐含层与输出层的连接权值， y_1, y_2, \dots, y_k 为神经网络预测输出，输入节点数为 n ，输出节点数为 K ，隐含层数目为 J 。

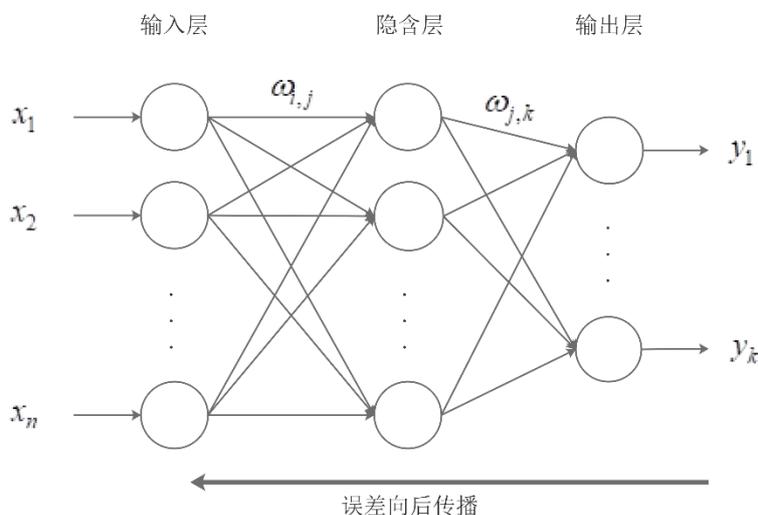


图 2.1 BP 网络的结构

2.3.2 支持向量回归

支持向量回归(Support Vector Regression, SVR)^[39]是将支持向量机的思想运用到回归问题之中，能够有效地解决非线性问题，通过引入核函数将非线性问题转换为高维空间中的线性问题。

传统的回归模型通常基于模型输出 $f(x)$ 与真实输出 y 之间的差别来计算损失，当且仅当 $f(x)$ 与真实输出 y 完全相同时，损失才为零。支持向量回归与之不同，它假设能容忍 $f(x)$ 与真实输出 y 之间最多有 ε 的偏差，即仅当 $|f(x) - y| \geq \varepsilon$ 时，才计算损失。SVR 相当于以 $f(x)$ 为中心，构建了一个宽度为 ε

的间隔带，若训练样本落在此间隔带内则被认为是预测正确的。SVR 的参数主要是选择基函数，如径向核函数。

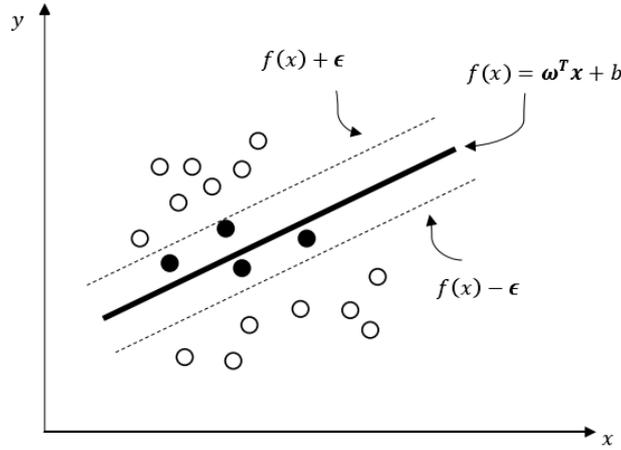


图 2.2 支持向量回归示意图

SVR 模型可以用如下公式来表示：

$$f(x) = w^T x + b = \sum_{i=1}^m (a_i^{\wedge} - a_i^{\vee}) x_i^T x + b \tag{2.21}$$

2.3.3 极限学习机

极限学习机(Extreme Learning Machine, ELM)是由 Huang 等^[40]提出的一种求解单隐层前馈神经网络的算法。该算法不需要设置输入层与隐含层的连接权值以及隐含层神经元的阈值，只需设置隐含层神经元的数目及隐含层神经元的激活函数即可。其基本原理如下：

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 的输出函数可以表示为：

$$f(x_i) = \sum_{l=1}^l \beta_l h_l(x_i) = h(x_i) \beta \tag{2.22}$$

其中， $h(x_i)$ 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 模型的训练目标为:

$$\begin{aligned} \min : l &= \frac{1}{2} \|\beta\|^2 + \frac{1}{2} C \sum_{i=1}^n \|\xi_i\|^2 \\ \text{s.t.} : h(x_i) \beta &= y_i^T - \xi_i^2 \end{aligned} \quad (2.23)$$

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 函数解决上述问题, 求极值, 可得:

$$\beta = H^T (HH^T + \frac{1}{C})^{-1} Y \quad (2.24)$$

其中, $H = [h^T(x_1), \dots, h^T(x_n)]^n$ 为隐含层输出矩阵。

因此 ELM 的输出可以表达为:

$$f(x) = h(x) \beta = h(x) H^T (HH^T + \frac{1}{C})^{-1} Y \quad (2.25)$$

2.4 优化方法

本文优化方法选择麻雀搜索算法(Sparrow Search Algorithm, SSA)。麻雀搜索算法是Xue和Shen^[41]于2020年提出的一种新型的群智能优化算法, 该算法模拟一群麻雀觅食的过程。在算法规则中, 将麻雀群体分为发现者和加入者, 另外还有以麻雀为对象的捕食者。根据麻雀个体所对应的不同适应度分为发现者和加入者, 发现者承担搜索食物丰富的区域, 为种群提供觅食区域和方向, 而加入者则通过发现者提供的信息来获取食物; 二者可以相互转化但二者数量占比不变。在觅食过程中, 部分麻雀发现捕食者时会发出警报, 称为警戒者, 当警报值大于安全值时, 发现者会带领加入者转移。其流程图如图2.3所示:

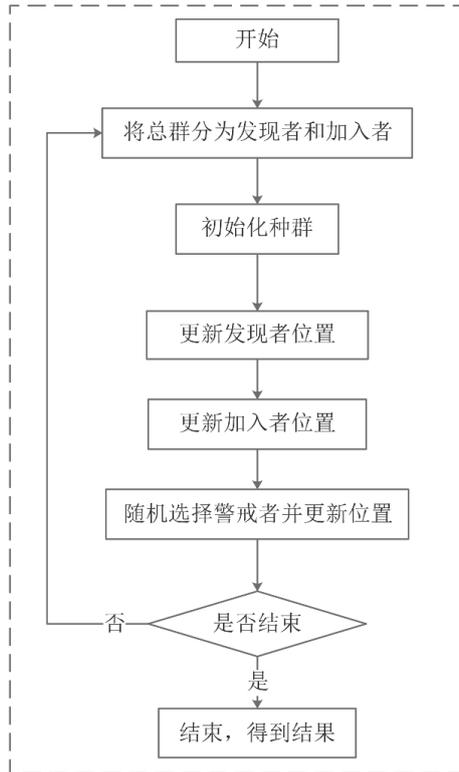


图2.3 麻雀搜索算法流程图

该算法比较新颖，具有寻优能力强，收敛速度快的优点，具体算法步骤如下：

(1) 初始化种群，迭代次数，初始化发现者和加入者比例。假设麻雀总数为 n 只，最大迭代次数为 $iterm_{max}$ ，搜索空间的维度为 d 。

(2) 根据 n 只麻雀的初始位置计算相应的适应度值，并对它们排序。

(3) 更新发现者位置：

$$X_{i,j}^{t+1} = \begin{cases} X_{i,j}^t \cdot \exp\left(-\frac{i}{\partial \cdot iterm_{max}}\right), & R_2 < ST \\ X_{i,j}^t + Q \cdot L, & R_2 \geq ST \end{cases}, \quad (2.26)$$

上式中， $X_{i,j}^t$ 表示在第 t 次迭代时第 i 只麻雀在第 j 维中的位置信息，这里 $j=1,2,\dots,d$ 。 $\partial \in (0,1]$ 和 $R_2 \in [0,1]$ 都是均匀分布的随机数， R_2 和 ST ($ST \in [0.5,1]$) 分别表示预警值和安全值。 Q 是服从标准正态分布的随机数。 L 表示元素均为1的一个 $1 \times d$ 的矩阵。

(4) 更新加入者位置：

$$X_{i,j}^{t+1} = \begin{cases} Q \cdot \exp\left(\frac{X_{worst}^t - X_{i,j}^t}{i^2}\right), i > n/2 \\ X_p^{t+1} + |X_{i,j}^t - X_p^{t+1}| \cdot A^+ \cdot L, \text{其他} \end{cases}, \quad (2.27)$$

其中, X_p^{t+1} 是 $t+1$ 时刻发现者的最佳位置, X_{worst}^t 则表示当前全局最差的位置。

A 表示 $1 \times d$ 的矩阵, 其中每个元素随机赋值为 1 或 -1, 并且 $A^+ = A^T (AA^T)^{-1}$ 。

(5) 更新警戒者位置:

$$X_{i,j}^{t+1} = \begin{cases} X_{best}^t + \beta \cdot |X_{i,j}^t - X_{best}^t|, f_i > f_g \\ X_{i,j}^t + K \cdot \left(\frac{|X_{i,j}^t - X_{best}^t|}{(f_i - f_w) + \varepsilon}\right), f_i = f_g \end{cases} \quad (2.28)$$

其中 X_{best}^t 是当前的全局最优位置。 β 作为步长控制参数, 是服从标准正态分布的随机数。 $K \in [-1, 1]$ 是一个均匀分布的随机数, f_i 则是当前麻雀个体的适应度值。 f_g 和 f_w 分别是当前全局最佳和最差的适应度值。 ε 为常数, 以避免分母出现零。

(6) 根据麻雀的更新的位置重新计算每个麻雀个体的适应度值。

(7) 根据适应度值及最大迭代次数判断是否满足停止条件, 满足则退出, 输出结果。否则, 重复执行步骤 2-6。

2.5 本章小结

本章主要介绍了本文所用到的研究方法。主要包括分解方法 (改进自适应噪声完备集合经验模态分解 ICEEMDAN)、重构方法 (样本熵 SE、近似熵 AE、LZ 复杂度、综合贡献指数 CCI 和快速傅立叶变换得到的傅立叶周期)、预测方法 (极限学习机 ELM、支持向量回归 SVR 和神经网络 BPNN) 和优化方法 (麻雀搜索算法 SSA)。

3 基于分解去噪框架的大豆期货价格预测

3.1 数据来源及模型评价指标

3.1.1 数据来源与描述统计

本文实验数据选取大豆期货主力合约的日收盘价价格，日期跨度为 2010-01-04 到 2022-04-08，数据去除非交易日，共 2970 条数据，数据来源为 Wind 数据库，本文根据所构建模型要求，首先要把所有大豆期货价格数据集划分为训练集和测试集，划分比例为 8: 2，训练集数据跨度为 2010-01-04 到 2019-10-16，共包含 2376 个数据，测试集数据跨度为 2019-10-16 到 2022-04-08，共包含 594 个数据。首先对大豆期货收盘价价格数据进行描述性统计分析，得到表 3.1。再绘制大豆期货价格时序图如图 3.1 所示。

表 3.1 大豆期货收盘价的统计分析

均值	中间值	最大值	最小值	标准差	偏度	峰度
4332.661	4322	6508	3129	707.9268	0.974438	3.662265

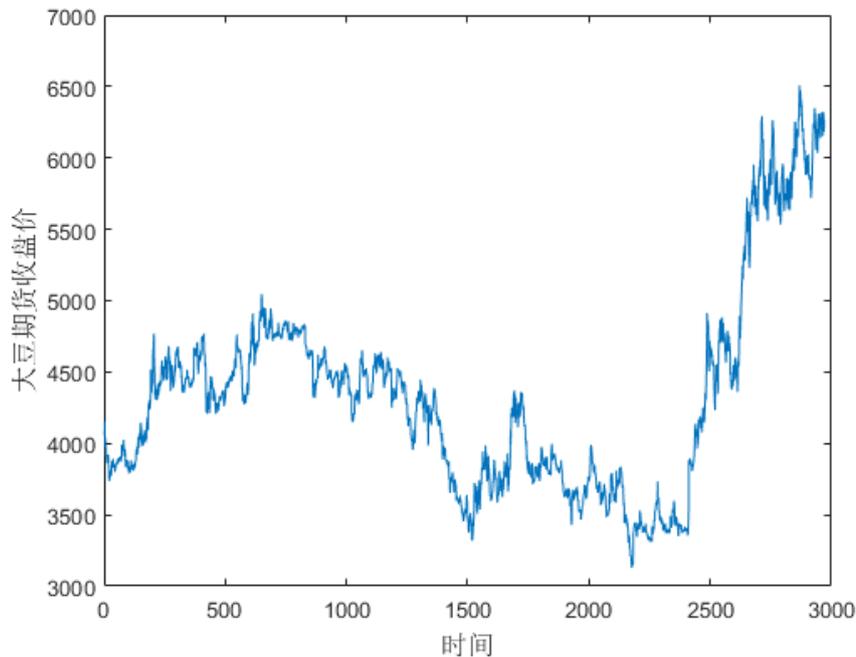


图 3.1 2010-01-04 到 2022-04-08 大豆期货价格趋势图

从表 3.1 的基本统计结果可以看出, 2010-01-04 到 2022-04-08 大豆期货价格序列的偏度值大于 0, 从偏度值可看出, 大豆期货价格序列具有明显的不对称性, 呈右偏分布; 峰度大于 3, 具有峰度, 不服从正态分布, 序列的分布与正态分布相比较为陡峭。从图 3.1 可以看出, 大豆期货价格序列具有随机波动和非平稳性的特点。

3.1.2 模型评价指标

本文用均方根误差(RMSE)、平均绝对百分比误差(MAPE)、平均绝对误差(MAE)和方向精度(DA)分别从水平预测精度和方向预测精度两个方面评估模型的预测效果, 计算方法如表 3.2 所示。

表 3.2 误差评估指标

评估指标	定义	公式
RMSE	均方根误差	$\sqrt{\frac{1}{N} \sum_{t=1}^N (y(t) - \hat{y}(t))^2}$
MAE	平均绝对误差	$\frac{1}{N} \sum_{t=1}^N \hat{y}(t) - y(t) $
MAPE	平均绝对百分比误差	$\frac{1}{N} \sum_{t=1}^N \left \frac{\hat{y}(t) - y(t)}{y(t)} \right \times 100\%$
DA	方向预测精度	$\frac{1}{N} \sum_{t=1}^N d(t)$

$$\text{其中, } d(t) = \begin{cases} 0 & \text{if } (y(t+1) - y(t))(\hat{y}(t+1) - y(t)) < 0 \\ 1 & \text{if } (y(t+1) - y(t))(\hat{y}(t+1) - y(t)) > 0 \end{cases}.$$

RMSE、MAE 和 MAPE 测量模型预测的水平精度, RMSE、MAE 和 MAPE 越小, 表明模型预测值与真实值差距越小, 预测效果越好。DA 衡量模型的方向预测精度, DA 值越高, 说明模型对真实值的趋势刻画更加准确。

除此之外, 本文引入改进率(IR)来测试不同模型之间预测能力的差异, 如表 3.3 所示。

表 3.3 改进率 (IR)

评估指标	定义	公式
IR_{RMSE}	均方根误差改进率	$-\frac{RMSE_A - RMSE_B}{RMSE_B} \times 100\%$
IR_{MAE}	平均绝对误差改进率	$-\frac{MAE_A - MAE_B}{MAE_B} \times 100\%$
IR_{MAPE}	平均绝对百分比误差改进率	$-\frac{MAPE_A - MAPE_B}{MAPE_B} \times 100\%$
IR_{DA}	方向预测精度改进率	$\frac{DA_A - DA_B}{DA_B} \times 100\%$

其中下标 B 表示基准方法 B，A 表示与基准方法对比的方法 A。当方法 A 优于方法 B 时，IR 的值为正；当方法 B 优于方法 A 时，IR 的值为负。

3.2 基于单一模型的大豆期货价格预测

3.2.1 SSA-BP 模型预测

本小节通过采用 MATLAB 搭建 SSA-BP 模型对大豆期货价格进行预测。进化次数设置为 30，种群规模设置为 20，时间窗口长度为 4，即以过去四天的收盘价来预测未来一天，来构建 SSA-BP 模型，对测试集预测效果如下表 3.4 所示。

表 3.4 SSA-BP 模型预测结果评价指标

预测方法	RMSE	MAE	MAPE(%)	DA(%)
SSA-BP	248.27	180.96	3.12	49.66

由表 3.4 可以看出，RMSE 和 MAE 两个绝对指标的值偏大，这两个指标和大豆收盘价的大小相关，而 MAPE 这个相对评价指标为 3.12%，这表示 SSA-BP 模型对大豆期货价格的预测效果一般，方向精度为 49.66%，从预测方向的角度看，SSA-BP 模型对大豆期货价格的预测效果也很一般，这表明对于大豆期货的预测可以进行更进一步的研究。

3.2.2 SSA-SVR 模型预测

本小节通过采用 MATLAB 搭建 SSA-SVR 模型对大豆期货价格进行预测。麻雀数量设置为 10，迭代次数为 100，滞后期为 4，来构建 SSA-SVR 模型，对测试集预测效果如下表 3.5 所示。

表 3.5 SSA-SVR 模型预测结果评价指标

预测方法	RMSE	MAE	MAPE(%)	DA(%)
SSA-SVR	1014.96	714.24	12.00	49.83

从表 3.5 的 SSA-SVR 预测的效果来看，SSA-SVR 这个方法对于对大豆期货价格原始序列的预测效果相较于 SSA-BP 的效果更差，RMSE 值与 MAE 值都偏大，MAPE 值高于 10%，表明该模型的预测效果较差，从预测的精度角度评价，该模型对原始序列的预测方向精度只有 49.83%，方向预测也不佳，对大豆期货价格的预测研究需要更加深入，提升预测效果。

3.2.3 SSA-ELM 模型预测

本节通过采用 MATLAB 搭建 SSA-ELM 模型对大豆期货价格进行预测。麻雀数量设置为 10，迭代次数为 100，滞后期为 4，来构建 SSA-ELM 模型，对测试集预测效果如下表 3.6 所示。

表 3.6 SSA-ELM 模型预测结果评价指标

预测方法	RMSE	MAE	MAPE(%)	DA(%)
SSA-ELM	91.17	69.51	1.28	47.81

从表 3.6 的 SSA-ELM 预测的效果来看，SSA-ELM 这个方法对于对大豆期货价格原始序列的预测效果相较于 SSA-BP 与 SSA-SVR 的效果有了很大提升，RMSE 值与 MAE 值都偏小，分别为 91.17 与 69.51，MAPE 值也较低，为 1.28%，这表明该模型的预测效果较好，但从预测结果的方向精度来看，DA 值仅为

47.81%，预测结果在方向精度指标的表现很差，需要进一步对数据进行处理分析，提升预测效果。

综上，从水平预测精度来讲，本节单一的机器学习（SSA-BP、SSA-SVR、SSA-ELM）模型对大豆期货价格的预测效果尚可，其中 SSA-ELM 的预测效果最佳；但从方向精度指标的表现来看，三种机器学习方法的预测效果不佳。因此，需要对模型进行进一步优化。

3.3 基于分解去噪的大豆期货价格预测

3.3.1 模型构建

由于单一的机器学习模型的拟合不能很好地处理复杂数据信息，易受噪声干扰，所以本文提出与 ICEEMDAN 分解方法相结合，对大豆期货价格进行去噪处理，构建去噪预测组合模型，并在分解去噪模型的基础上，引入其它农产品期货市场的信息对模型性能进行进一步的提升。

对原始数据进行 ICEEMDAN 分解，获得多个分解序列，这些分解序列中包含若干个高频分解序列和若干个中低频分解序列，一般中低频分解序列主要是有效信息，而高频分解序列则含有大量随机噪声。在本章中，通过直接去除最高频的分解序列的方法来消除原始序列中的噪声。而大豆期货价格也受到其它农产品期货价格波动的影响，因此本文在分解去噪模型的基础上，引入其它农产品期货的价格信息。豆粕作为大豆的下游产物，其期货价格与大豆期货价格紧密相连，而玉米与大豆的强替代关系也导致玉米期货价格与大豆期货价格息息相关，因此本文引入豆粕期货价格与玉米期货价格的信息。但因为期货价格序列中含有随机噪声，这可能会“掩盖”或者“削弱”豆粕期货价格与玉米期货价格对大豆期货价格的影响，因此在本文中也对豆粕期货价格与玉米期货价格进行去噪处理。

基于分解去噪框架下的大豆期货价格预测的流程图如图 3.2 所示，首先通过使用 ICEEMDAN 分解算法将大豆期货收盘价格序列进行分解，得到多个分解序列，将复杂度最高的序列视为噪声序列舍弃，其它分解序列线性加总形成大豆

期货去噪序列，然后将引入的农产品期货价格进行去噪处理，其得到的去噪序列作为预测大豆期货价格去噪序列的输入变量，其结果则为大豆期货价格的预测。本章节的基准模型为单一的机器学习预测模型（SSA-BP、SSA-SVR 和 SSA-ELM）与不引入其它信息的分解去噪模型（ICEEMDAN-SSA-BP、ICEEMDAN-SSA-SVR 和 ICEEMDAN-SSA-ELM）。

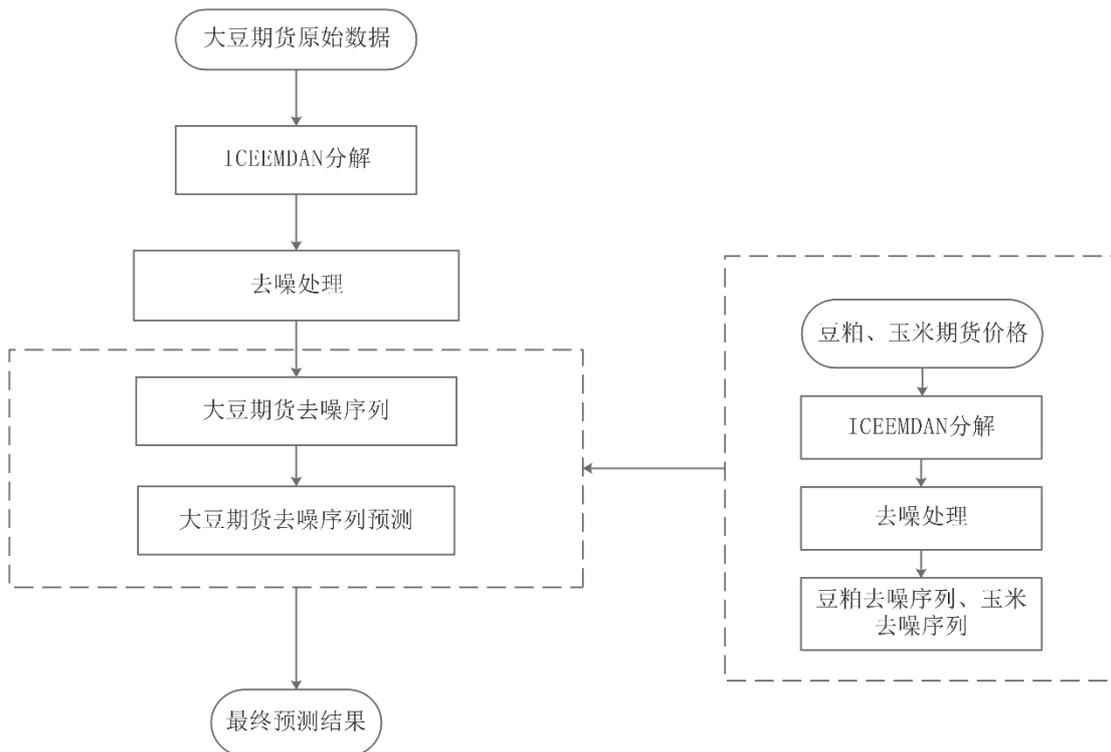


图 3.2 引入其它期货市场信息的分解去噪模型流程图

3.3.2 模型预测

本节主要通过 MATLAB 按照 ICEEMDAN 分解过程，分别对大豆期货价格、豆粕期货价格和玉米期货价格进行 ICEEMDAN 分解，三种期货的价格序列都被分解为 10 个分解序列，包含 9 个 IMF 分量和一个残差项 Res，三种期货价格各自的分解序列如图 3.3、图 3.4 和图 3.5 所示，从各自的分解序列的走势图中明显可以看出，越靠后的分解序列其走势越具有规律，越靠前的分解序列其波动性更强。因此，对分解的分解序列进行复杂性分析，三种期货价格分解序列的样

本熵值如表 3.7 所示。

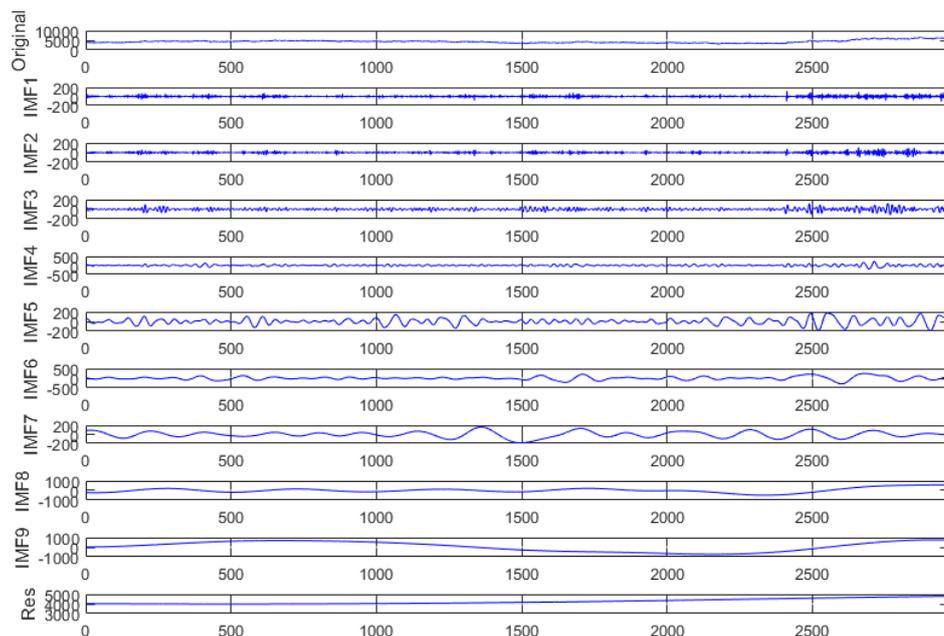


图 3.3 ICEEMDAN 分解大豆期货的结果

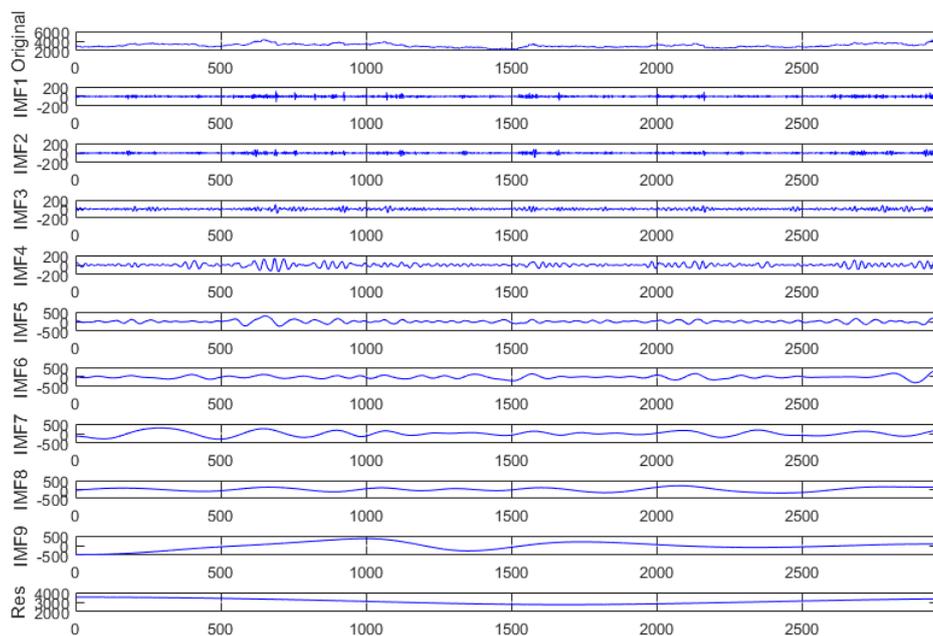


图 3.4 ICEEMDAN 分解豆粕期货的结果

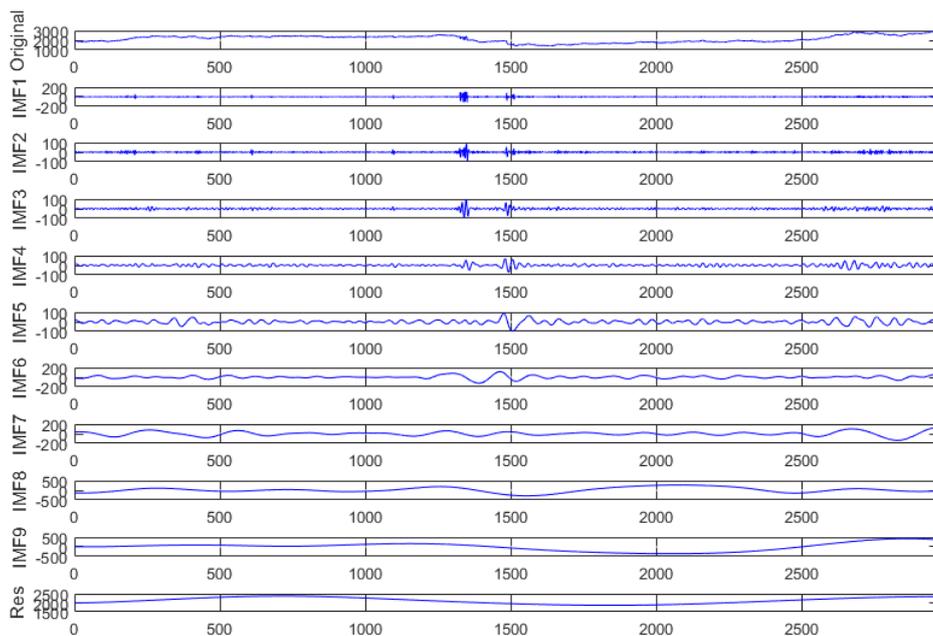


图 3.5 ICEEMDAN 分解玉米期货的结果

表 3.7 各个期货分解序列样本熵值

分解序列	大豆期货	豆粕期货	玉米期货
IMF1	1.6669	1.6404	1.0740
IMF2	1.0843	1.0311	0.7733
IMF3	0.6326	0.6434	0.5506
IMF4	0.5702	0.5051	0.5182
IMF5	0.4072	0.2955	0.3554
IMF6	0.1629	0.1712	0.1285
IMF7	0.0815	0.0645	0.0807
IMF8	0.0262	0.0414	0.0209
IMF9	0.0044	0.0129	0.0063
Res	0.0017	0.0033	0.0067

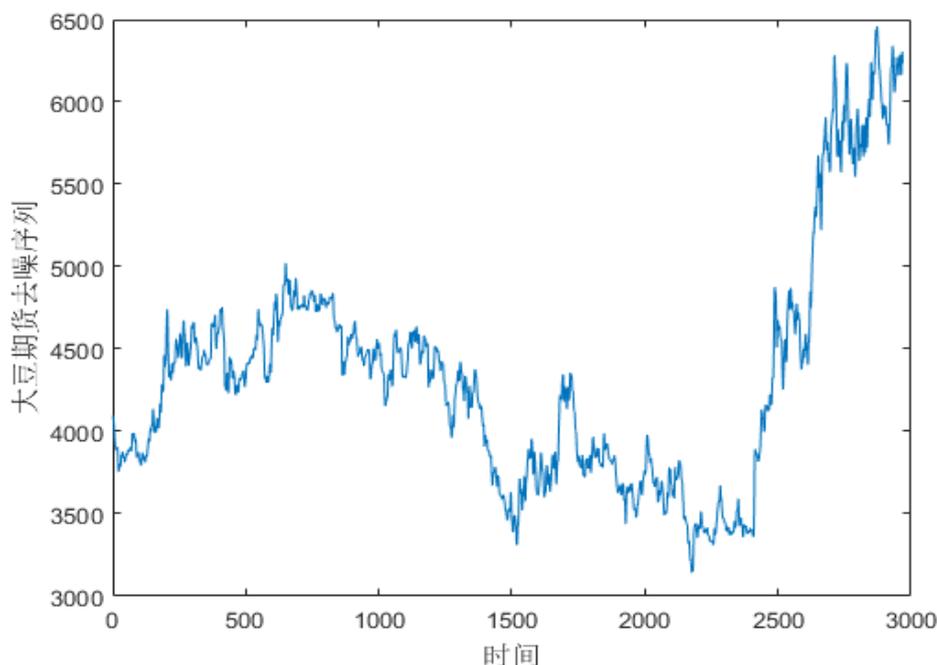


图 3.6 大豆期货 ICEEMDAN 分解去噪序列

从表 3.7 可以看出，大豆期货价格、豆粕期货价格与玉米期货价格的分解序列样本熵值均是 IMF1 最大，分别为 1.6669、1.6404 和 1.074，这表明三种期货价格分解后的 IMF1 的复杂度都很高，都将其视为噪声序列并剔除，然后将大豆期货价格、豆粕期货价格与玉米期货价格各自分解剩余的分解序列相加分别得到大豆期货去噪序列、豆粕期货去噪序列和玉米期货去噪序列，大豆期货去噪序列趋势图如图 3.6 所示。

在对大豆期货价格、豆粕期货价格与玉米期货价格进行去噪后，将豆粕期货去噪序列和玉米期货去噪序列作为预测大豆期货去噪序列的自变量，并使用 SSA-BP、SSA-SVR、SSA-ELM 三种预测方法对大豆期货去噪序列进行预测，参数设置与单一机器学习模型的参数设置相同，得到的预测结果与基准模型预测结果如表 3.8 所示，表中包含机器学习模型（SSA-BP、SSA-SVR、SSA-ELM）、分解去噪模型（ICEEMDAN-SSA-BP、ICEEMDAN-SSA-SVR 和 ICEEMDAN-SSA-ELM）以及本文提出的引入其它信息的分解去噪模型（ICEEMDAN-SSA-BP-M、ICEEMDAN-SSA-SVR-M 和 ICEEMDAN-SSA-ELM-M）。为进一步研究引入其它信息的分解去噪模型的预测效果，计算了改

进率 IR 对实证结果进行分析, 如表 3.9 所示, 表 3.8 和表 3.9 中的 ICE 均表示 ICEEMDAN 分解算法。

表 3.8 各模型预测效果

	RMSE	MAE	MAPE(%)	DA(%)
SSA-BP	248.27	180.96	3.12	49.66
SSA-SVR	1014.96	714.24	12.00	49.83
SSA-ELM	91.17	69.51	1.28	47.81
ICE-SSA-BP	235.74	163.54	2.76	65.99
ICE-SSA-SVR	863.84	595.83	9.96	64.98
ICE-SSA-ELM	46.90	34.78	0.61	76.77
ICE-SSA-BP-M	186.16	125.91	2.14	67.17
ICE-SSA-SVR-M	307.96	207.49	3.49	62.46
ICE-SSA-ELM-M	39.81	30.83	0.55	77.10

表 3.9 各模型预测改进率对比

	IR _{RMSE}	IR _{MAE}	IR _{MAPE}	IR _{DA}
ICE-SSA-BP→SSA-BP	5.05%	9.63%	11.28%	32.88%
ICE-SSA-SVR→SSA-SVR	14.89%	16.58%	16.98%	30.41%
ICE-SSA-ELM→SSA-ELM	48.56%	49.97%	52.06%	60.56%
ICE-SSA-BP-M→ICE-SSA-BP	21.03%	23.01%	22.75%	1.79%
ICE-SSA-SVR-M→ICE-SSA-SVR	64.35%	65.18%	64.92%	-3.89%
ICE-SSA-ELM-M→ICE-SSA-ELM	15.13%	11.34%	10.13%	0.44%

从表 3.8 和表 3.9 可以看出, 对于分解去噪模型, 相较于直接对原始序列进行预测的效果, SSA-BP、SSA-SVR、SSA-ELM 三种方法的预测效果均有了提升, 方向预测精度指标 DA 值的提升效果最为明显, ICE-SSA-ELM 的 DA 值表现最佳, 达到了 76.77%, 相较于 SSA-ELM 方法对大豆期货价格预测提升了 60.56%。而本文提出的引入豆粕期货与玉米期货信息的分解去噪模型较于分解去噪模型, 其预测效果得到了较大的提升, 尤其是水平评价指标 (RMSE、MAE 和 MAPE), 均提升了超过 10%, ICE-SSA-SVR-M 模型相对于 ICE-SSA-SVR 模型其水平预测指标甚至提高超过了 50%, 即使 ICE-SSA-SVR-M 模型的

方向预测精度不如 ICE-SSA-SVR 模型。

3.4 本章小结

本章建立了引入豆粕期货价格信息和玉米期货价格信息的分解去噪模型。实证结果表明通过分解去噪模型相较于直接对大豆期货价格序列进行预测的机器学习模型预测效果有很大提升，而通过对豆粕和玉米的期货价格进行去噪处理的方法引入豆粕价格信息和玉米价格信息在分解去噪模型的基础上进一步提升了预测效果。

4 基于“分解-重构-集成”框架的大豆期货价格预测

4.1 不同重构方法下的大豆期货价格预测

4.1.1 模型构建

“分解-重构-集成”的思想常用在价格预测中，而不同重构方法下的“分解-重构-集成”模型有不同的预测效果，因此本文研究了不同的重构方法对模型预测效果的影响。对于分解序列的重构，一般是将特征相似、波动趋势相似的分解序列进行重构，因此本文的重构方法主要考量了分解序列本身的特征，包括分解序列的复杂性、分解序列的周期性以及分解序列与原始序列的相关性三个方面的特征。本文对分解序列的重构方法分为两个部分，一部分是探讨单特征的分解序列重构方法对预测效果的影响，即分别利用复杂性特征、周期性特征与相关性特征对分解序列进行重构的“分解-重构-集成”模型的预测效果；另一部分则是综合考虑多个特征的多特征重构方法，即综合利用复杂性、周期性与相关性三个特征的信息对分解序列进行重构，研究了多特征重构方法与单特征重构方法的优劣。除此之外，为了使本文的研究结果更具说服力，本文选择了三个常用指标--样本熵 SE、近似熵 AE 和 LZ 复杂度来描述各个分解序列的复杂性特征，并分别运用到本文所研究的问题之中。

本节所构建的模型主要步骤包括四个步骤：数据分解，分解序列重构，重构分量预测和分量预测结果非线性集成。本小节构建的模型流程图如图 4.1 所示。

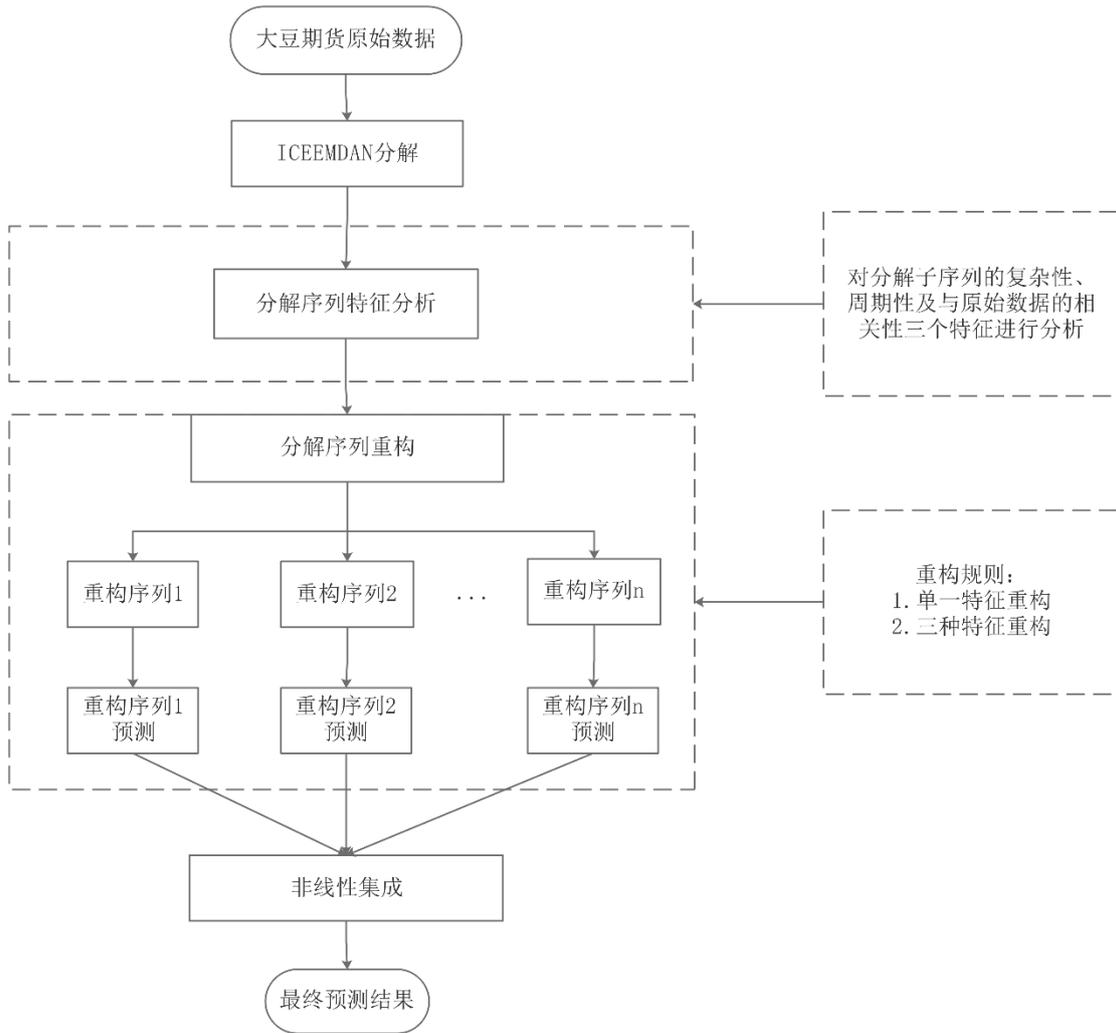


图 4.1 不同重构方法的“分解-重构-集成”预测模型流程图

表 4.1 不同重构方法

数据特征		特征指标
单一特征重构	复杂性	样本熵 近似熵 LZ 复杂度
	周期性	傅立叶周期
	相关性	综合贡献指数
	多个特征重构	复杂性、周期性和相关性
		样本熵、傅立叶周期和综合贡献指数 近似熵、傅立叶周期和综合贡献指数 LZ 复杂度、傅立叶周期和综合贡献指数

不同重构方法下的大豆期货价格“分解-重构-集成”预测模型步骤如下：

(1) 数据预处理。对大豆期货价格原始数据进行归一化处理，并划分为训练集和测试集，本文中划分比例为 8:2，训练集用于训练模型，测试集用于检验模型。

(2) 数据分解。对大豆期货价格原始数据采用 ICEEMDAN 分解方法进行分解，分解为有限个分解序列。

(3) 分解分量重构。首先，对分解序列的复杂性、周期性和相关性三个特征进行描述与分析。其中，数据复杂性特征由样本熵、LZ 复杂度、近似熵确定，分解序列与原始序列的相关性由综合贡献指数确定，每个分解序列的周期性由快速傅里叶变化所求解的傅立叶周期来确定。然后，对分解序列进行不同方法的重构，重构方法如表 4.1 所示，对于只考虑单个数据特征的重构方法，分别用复杂性指标（样本熵、近似熵和 LZ 复杂度）、相关性指标（综合贡献指数）、周期性指标（傅立叶周期）五个数据特征指标进行重构，对于考虑多个数据特征的分解序列重构，综合考虑复杂性、周期性和相关性三个特征，分别用样本熵、近似熵和 LZ 复杂度三个复杂性指标与傅立叶周期和综合贡献指数采用 K-Means 聚类方法对相似数据特征的分解序列进行重构。

(4) 重构分量预测。对于重构序列，采用 SSA-ELM 方法对其进行预测。

(5) 非线性集成。对重构序列的预测值进行非线性集成，本章选取 SSA-ELM 方法进行集成。

4.1.2 模型预测

本章主要通过 MATLAB 2019a 软件，按照 ICEEMDAN 分解过程，对大豆期货价格序列进行 ICEEMDAN 分解，分解为 10 个分解序列，包括 9 个 IMF 分量和一个残余项 Res，然后对分解后的分解序列进行重构，为探究不同重构方法对预测效果的影响，本文对分解后的分解序列进行特征分析并重构，包括分解序列的复杂性特征、分解序列的周期性特征以及分解序列与原始序列的相关性特征三个数据特征。分解序列重构包括两个部分，一部分是利用分解序列单一特征进行重构，即分别对分解序列的复杂性、周期性和相关性进行分析并重构；第二个部分则是综合考量分解序列复杂性、周期性和相关性三个特征进行

多特征重构，进而研究单独使用复杂性、周期性和相关性的属性进行分解序列重构与综合考虑三个方面的数据特性进行分解序列重构对于大豆期货价格预测的优劣。

表 4.2 大豆期货分解序列数据特征

序列	样本熵	近似熵	LZ 复杂度	傅立叶周期	综合贡献指数
IMF1	1.6669	1.6307	0.9298	3.2108	0.1330
IMF2	1.0843	1.1594	0.6844	7.4953	0.1358
IMF3	0.6326	0.6603	0.5522	15.7351	0.1806
IMF4	0.5702	0.6237	0.3304	36.5538	0.2400
IMF5	0.4072	0.4673	0.2266	81.9310	0.2939
IMF6	0.1629	0.1982	0.1322	148.5000	0.3319
IMF7	0.0815	0.0828	0.0802	182.7692	0.3829
IMF8	0.0262	0.0250	0.0425	475.2000	0.4632
IMF9	0.0044	0.0048	0.0142	2376.0000	0.5993
Res	0.0017	0.0006	0.0142	2376.0000	1.5173

ICEEMDAN 分解后各个分解序列的基本特征如表 4.2 所示，该表展现了各个分解序列的复杂性、周期性和与原始序列的相关性的指标，包含了分解序列样本熵 SE、近似熵 AE、LZ 复杂度、傅立叶周期以及综合贡献指数 CCI 一共五个特征指标的值。从表 4.2 中可看出，大豆期货价格序列被 ICEEMDAN 方法分解的 10 个分解序列的波动程度由高到低排列，呈现出不同的频率和特征。根据分解序列的各项特征值可发现如下规律：

(1) 分解序列的复杂性呈现出由高到低的趋势。波动频率越高的分解序列的非线性特征与无规律性特征更加明显，因此分解序列的复杂性越强。通过样本熵这一指标来看，ICEEMDAN 分解方法获得的前几个分量中，频率最高的分解序列样本熵值为 1.6669，样本熵值最低为残余项，仅有 0.0017，其间的样本熵值从高到低依次递减，分解序列 IMF1 和 IMF2 的样本熵值大于 1，样本熵值在 0.1-1 范围内有 4 个，在 0.1 以下的分解序列有 4 个；从近似熵指标来看，分解序列的近似熵特征的分布趋势与样本熵特征的分布趋势十分相近，在以 0.1 和 1 为界的取值区间有相同的分解序列个数；从 LZ 复杂度指标的表现来看，前 6 个

LZ 复杂度的值处于 0.1-1 范围，剩下的 4 个分解序列 LZ 复杂度值小于 0.1。

(2) 分解序列的周期呈现出由短到长的趋势。越靠前的分解序列的周期越短，越靠后的分解序列的周期越越长，这表明大豆期货价格种糅杂的短期信息、中长期信息与长期信息被分解方法提取出来。分解序列当中靠前的分解序列周期性较短，主要表现出大豆期货价格中短期和中期波动的信息，而靠后几个分解序列则反应的是长期的大豆期货价格的走向。

(3) 每个分解序列与原始序列之间的相关性呈现出由弱到强的趋势。综合贡献指数代表着分解序列与原始序列之间的相关性，将综合贡献指数与复杂性指标结合来看，复杂性越强的分解序列，其综合贡献指数越小。这表明综合贡献指数靠前的几个分解序列波动剧烈变化复杂，包含了较少的大豆期货价格序列信息；而综合贡献指数靠后的几个分解序列波动较小趋势明显，包含了更多的大豆期货价格序列信息。

根据各分解序列基本特征统计值及上述分析，对于只考虑单一特征的分解序列重构，采用 K-Means 算法对周期性和相关性特征值进行聚类，以达到相似分量重构的目的，设定聚类数为 3，对于基于复杂性特征的重构，指标值大于 1 的分解序列视为一类并线性加总，大于等于 0.1 并小于等于 1 的分解序列为一类加总，小于 0.1 的分解序列为一类并加总。对于考虑三个特征的分解序列重构，即复杂性指标、周期性指标和相关性指标三个数据基本特征的重构，也采用 K-Means 聚类算法进行聚类，聚类数也设定为 3。对 ICEEMDAN 方法分解的分解序列的 LZ 复杂度重构结果如表 4.3 所展示，其它重构方法的重构结果如表 4.4 所示。

表 4.3 LZ 复杂度重构结果

重构方法	重构序列 1	重构序列 2
LZ 复杂度	IMF1-IMF6	IMF7-IMF9、Res

从表 4.3 可以看出，以 LZ 复杂度为衡量指标，将 10 个分解序列重构了两个序列，IMF1-IMF6 重构为一项，IMF7-IMF9 和 Res 项重构为 1 项。其余重构方法均将分解序列重构为 3 项，具体如表 4.4 所示。表 4.4 中样本熵、近似熵、综

合贡献指数和傅立叶周期表示只考虑单个特征的重构；而样本熵聚类、近似熵聚类与 LZ 复杂度聚类则表示考虑了三个特征的重构方法，分别用样本熵、近似熵和 LZ 复杂度三个复杂性指标与综合贡献指数和周期值聚类对分解序列进行重构，例如样本熵聚类表示以样本熵、傅立叶周期和综合贡献指数三个指标聚类，近似熵聚类与 LZ 复杂度聚类同理。

表 4.4 不同重构方法重构结果

重构方法	重构序列 1	重构序列 2	重构序列 3
样本熵	IMF1-IMF2	IMF3-IMF6	IMF7-IMF9、Res
近似熵	IMF1-IMF2	IMF3-IMF6	IMF7-IMF9、Res
傅立叶周期	IMF1-IMF7	IMF8	IMF9、Res
综合贡献指数	IMF1-IMF5	IMF6-IMF9	Res
样本熵聚类	IMF1-IMF2	IMF3-IMF8	IMF9、Res
近似熵聚类	IMF1-IMF2	IMF3-IMF8	IMF9、Res
LZ 复杂度聚类	IMF1-IMF3	IMF4-IMF8	IMF9、Res

从表 4.4 可以看出，同一特征下采用不同特征指标的重构方法可能会得到相同的重构结果，只考虑一种特征的样本熵重构结果与近似熵重构结果相同，将 IMF1 和 IMF2 重构为重构序列 1，IMF3 至 IMF6 重构为重构序列 2，IMF7 至 IMF9 以及残差项重构为重构序列 3；考虑复杂性、周期性和相关性的样本熵聚类重构结果与近似熵聚类重构结果也相同，将 IMF1 和 IMF2 重构为重构序列 1，IMF3 至 IMF8 重构为重构序列 2，IMF9 和残差项重构为重构序列 3。其余的重构方法得到的重构结果则各不相同，其中只考虑综合贡献指数与傅立叶周期的重构结果与其它重构方法明显不同，这是因为重构时考虑了分解序列不同的特征，以综合贡献指数进行分解序列重构得到了 IMF1-IMF5、IMF6-IMF9 以及残差项三个重构序列，而以周期指标重构则得到了 IMF1-IMF7、IMF8、IMF9 和残差项三个重构序列，综合考虑 LZ 复杂度、周期和综合贡献指数三个特征指标的重构方法则将 IMF1 至 IMF3、IMF4 至 IMF8 以及 IMF9 与残差项分别重构为 3 个重构序列。综上，同一特征下采用不同特征指标的重构方法可能会得到相同的重构结果，而不同特征下的重构方法可能会使得重构结果大不相同。

在用不同的方法将分解序列重构后，对重构序列分别用 SSA-ELM 方法进行预测，再用 SSA-ELM 方法进行非线性集成，具体的预测效果如表 4.5 所示。为了进一步比较，得到了本文提出的考虑多特征的“分解-重构-集成”模型较于仅考虑单一特征的“分解-重构-集成”模型的 IR 值，具体结果如表 4.6 所示。表 4.6 中， P_S 、 P_A 和 P_L 表示重构方法为样本熵 SE 聚类、近似熵 AE 聚类和 LZ 复杂度聚类的预测模型，O 表示大豆期货价格序列的 SSA-ELM 预测模型，A、B、C、D 和 E 分别表示重构方法为样本熵、近似熵、LZ 复杂度、傅立叶周期和综合贡献指数的预测模型。

表 4.5 不同重构方法下的预测效果

重构方法	RMSE	MAE	MAPE(%)	DA(%)
LZ 复杂度	86.83	64.26	1.20	56.23
样本熵/近似熵	52.49	39.62	0.74	73.91
综合贡献指数	280.09	205.68	3.56	51.85
傅立叶周期	143.32	109.97	1.96	49.83
LZ 复杂度聚类	67.38	50.52	0.95	67.51
样本熵/近似熵聚类	67.58	35.82	0.68	78.96

表 4.6 多特征重构模型预测改进率对比

	IR _{RMSE}	IR _{MAE}	IR _{MAPE}	IR _{DA}
$P_S/P_A \rightarrow O$	25.87%	48.46%	47.25%	65.15%
$P_S/P_A \rightarrow A/B$	-28.74%	9.59%	8.60%	6.83%
$P_S/P_A \rightarrow D$	52.85%	67.43%	65.48%	58.45%
$P_S/P_A \rightarrow E$	83.29%	88.34%	87.27%	62.28%
$P_L \rightarrow O$	26.10%	27.31%	25.54%	41.20%
$P_L \rightarrow C$	22.40%	21.37%	20.46%	20.06%
$P_L \rightarrow D$	52.99%	54.06%	51.28%	35.47%
$P_L \rightarrow E$	83.34%	83.56%	82.03%	38.75%

从表 4.5 与表 4.6 可以得出以下结论：

(1) 从水平预测精度和方向预测精度两个方面来讲，仅仅考虑周期性指标傅

立叶周期与相关性指标综合贡献指数进行重构的预测模型的预测结果较差，其 MAPE 值分别为 1.96%和 3.56%，方向预测精度指标仅有 49.83%和 51.85%，用复杂性特征的指标进行重构的预测相对来讲更好，其中用样本熵或近似熵进行重构的预测模型表现优于用 LZ 复杂度进行重构的预测模型，使用样本熵或近似熵重构的模型 MAPE 值降到了 0.74%，方向精度 DA 提升到了 73.91%。这说明考虑不同特征的重构方法会“分解-重构-集成”模型的预测效果大不相同，选取不合适的特征进行重构可能会“适得其反”，在复杂性、周期性和相关性三个特征中，基于复杂性重构的模型有较好的预测效果。

(2) 考虑多特征重构的预测模型较于仅考虑单一特征重构的预测模型更优。虽然只考虑某一特征的重构方法的预测模型可能表现不佳，但是该特征指标可以用来优化其它特征重构的模型，即与其它特征相结合对分解序列进行重构，使得分解序列在重构时可以更加全面地利用分解序列的信息进行分解序列重构，从而提升模型的预测效果。如用分解序列复杂性信息对其重构时，考虑重构序列的周期性和相关性信息，可以提升模型的预测效果。如表 4.5 和表 4.6 所示，样本熵聚类重构和近似熵聚类重构的模型相较于样本熵重构和近似熵重构，水平预测精度指标 MAPE 提升了 8.6%，方向预测精度指标 DA 提升了 6.83%；LZ 复杂度聚类重构的模型相较于 LZ 复杂度重构，水平预测精度指标 MAPE 提升了 20.46%，方向预测精度指标 DA 提升了 20.06%

4.2 引入外生信息的大豆期货价格预测

4.2.1 模型构建

4.1 部分对于大豆期货价格的预测仅仅根据序列本身的信息进行预测，在本节中，认为其它农产品期货市场的信息对大豆期货价格的预测具有影响，本节选取豆粕期货日收盘价与玉米期货日收盘价作为外生变量，探讨加入豆粕期货价格与玉米期货价格后对于考虑多特征重构的“分解-重构-集成”预测模型是否有性能提升。许多学者对于加入影响因素对某价格序列的预测直接将某个影响因素的原始序列作为输入变量对其进行预测，而在本文当中，对外生变量豆

粕期货价格和玉米期货价格也用 ICEEMDAN 方法进行分解，并对豆粕期货价格和玉米期货价格的分解序列分别用多特征重构方法对其进行重构，再进一步研究豆粕期货和玉米期货的重构序列与大豆期货重构序列的相关性，选择豆粕期货价格和玉米期货价格的重构序列作为预测大豆期货重构序列的输入变量。考虑外生信息的大豆期货价格预测模型的框架图如图 4.2 所示。

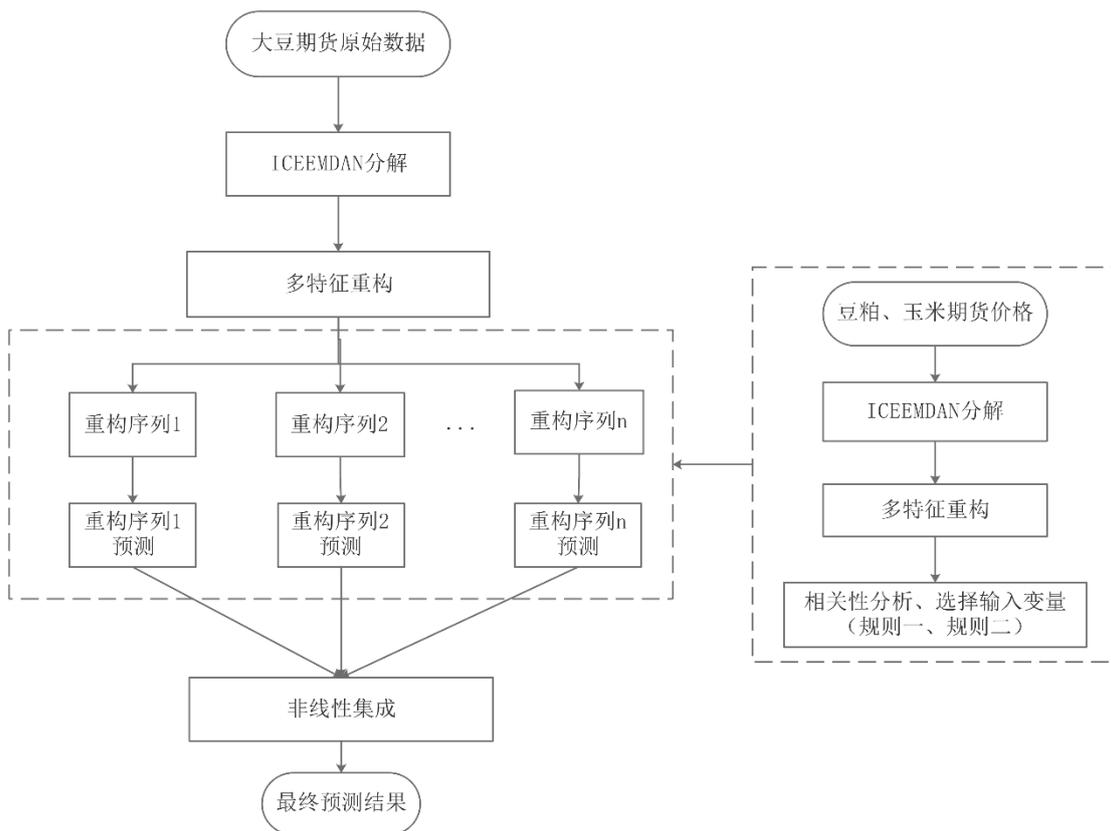


图 4.2 引入其它信息的“分解-重构-集成”模型流程图

考虑其它农产品期货市场的信息所构建的方法主要步骤包括五个步骤：大豆期货、豆粕期货和玉米期货数据分解，三种期货分解序列重构，玉米期货和豆粕期货重构序列与大豆期货重构序列的相关性分析，大豆期货重构分解序列预测和非线性集成预测结果。本章构建的方法具体步骤如下：

(1) 数据分解。对大豆期货价格、豆粕期货日收盘价和玉米期货日收盘价原始数据分别采用 ICEEMDAN 分解方法进行分解；

(2) 三种期货分解分量重构。本节以多特征重构的方法对三种期货分解序列

进行重构，即样本熵聚类重构、近似熵聚类重构和 LZ 复杂度聚类重构三种重构方法，重构规则与前文一致；

(3) 三种期货重构序列相关性分析。分析玉米期货和豆粕期货各自的重构序列与大豆期货重构序列的相关性；

(4) 大豆期货重构分量预测。对于大豆期货各个重构序列，根据第（3）步的结果，选择引入豆粕期货价格和玉米期货价格的重构序列作为大豆期货重构序列预测的自变量，用 SSA-ELM 对大豆期货重构序列其进行预测。本文采用两种规则选取大豆期货重构序列预测的自变量，第一种选取规则是选取与大豆期货重构序列相关性最强的豆粕期货价格和玉米期货价格的重构序列作为大豆期货重构序列预测的自变量，例如，选取豆粕期货价格重构序列 1、2、3 中与大豆期货价格重构序列 1 相关性最强的豆粕重构序列作为预测大豆期货价格重构序列 1 的输入变量，第二种规则是选取与大豆期货价格重构序列统计学意义上显著相关的豆粕期货价格和玉米期货价格的重构序列作为大豆期货重构序列预测的自变量；

(5) 非线性集成。对大豆期货重构序列的预测值进行非线性集成，本章选取 SSA-ELM 方法进行集成。

4.2.2 模型预测

本节是在上一节分解重构预测的基础上的进一步研究，引入了其它农产品期货市场的信息，大豆期货价格的分解重构过程与前文一致。对于玉米期货收盘价与豆粕期货收盘价进行 ICEEMDAN 分解后分别对豆粕期货价格和玉米期货价格的分解序列进行进一步的特征分析，结果如表 4.7 和表 4.8 所示。

从表 4.7 与表 4.8 可以看出，豆粕期货价格和玉米期货价格的分解序列特征同大豆期货价格的分解序列都呈现出相同的特征走势，代表复杂性特征的样本熵、近似熵、LZ 复杂度三个指标都呈现出逐步减小的趋势，这表明三种期货的分解序列越靠后其复杂性越低，而在周期性和与原始序列相关性这两个特征上的表现则是越靠后的分解序列其周期性越强，与原始序列的相关性越低。

表 4.7 豆粕分解序列特征

分解序列	样本熵	近似熵	LZ 复杂度	傅立叶周期	综合贡献指数
IMF1	1.6404	1.5951	0.9251	3.0423	0.1335
IMF2	1.0311	1.0798	0.7127	7.7902	0.1346
IMF3	0.6434	0.6526	0.5286	14.5767	0.1713
IMF4	0.5051	0.5897	0.3729	34.4348	0.2332
IMF5	0.2955	0.3382	0.203	72.0000	0.3427
IMF6	0.1712	0.1717	0.1227	148.5000	0.3832
IMF7	0.0645	0.0644	0.0566	339.4286	0.4515
IMF8	0.0414	0.0398	0.0472	475.2000	0.4472
IMF9	0.0129	0.013	0.0236	1188.0000	0.5004
Res	0.0033	0.0029	0.0142	2376.0000	1.5146

表 4.8 玉米分解序列特征

分解序列	样本熵	近似熵	LZ 复杂度	傅立叶周期	综合贡献指数
IMF1	1.074	1.2358	0.9487	3.2022	0.1233
IMF2	0.7733	0.8816	0.7033	8.4857	0.1278
IMF3	0.5506	0.5839	0.5569	11.5902	0.1539
IMF4	0.5182	0.5751	0.354	31.2632	0.2013
IMF5	0.3554	0.3904	0.2171	60.9231	0.3041
IMF6	0.1285	0.1614	0.118	169.7143	0.3354
IMF7	0.0807	0.0845	0.0661	237.6000	0.3971
IMF8	0.0209	0.0219	0.0472	792.0000	0.5123
IMF9	0.0063	0.0062	0.0142	2376.0000	0.542
Res	0.0067	0.0073	0.0189	2376.0000	1.5132

根据重构规则，豆粕期货价格和玉米期货价格分解序列的重构结果如表 4.9 所示。对于豆粕期货分解序列的重构，样本熵聚类和 LZ 复杂度聚类重构结果相同，IMF1-IMF3 重构为一类，IMF4-IMF9 重构为一类，残差项为一类。对于玉米期货分解序列的重构，样本熵聚类和近似熵聚类得到了相同的重构结果，IMF1-IMF4 重构为一类，IMF5-IMF8 重构为一类，IMF9 和残差项重构为一类。

表 4.9 不同重构方法的重构结果

	重构方法	重构序列 1	重构序列 2	重构序列 3
豆粕分解序列	样本熵聚类	IMF1-IMF3	IMF4-IMF9	Res
	LZ 复杂度聚类	IMF1-IMF3	IMF4-IMF9	Res
	近似熵聚类	IMF1-IMF4	IMF5-IMF9	Res
玉米分解序列	样本熵聚类	IMF1-IMF4	IMF5-IMF8	IMF9、Res
	LZ 复杂度聚类	IMF1-IMF3	IMF4-IMF8	IMF9、Res
	近似熵聚类	IMF1-IMF4	IMF5-IMF8	IMF9、Res

表 4.10 多种期货重构序列皮尔逊相关系数

重构方法	序列	大豆重构序列 1	大豆重构序列 2	大豆重构序列 3
样本熵聚类	豆粕重构序列 1	0.4629**	0.0497	0.0014
	豆粕重构序列 2	-0.0009	0.3991**	0.1243**
	豆粕重构序列 3	-0.0020	0.0033	0.6557**
	玉米重构序列 1	0.1801**	0.0311	0.0025
	玉米重构序列 2	-0.0039	0.1263**	-0.1985**
	玉米重构序列 3	-0.0005	0.1166**	0.9705**
近似熵聚类	豆粕重构序列 1	0.3135**	0.1405**	-0.0107
	豆粕重构序列 2	-0.0018	0.3842**	0.1273**
	豆粕重构序列 3	-0.0020	0.0033	0.6557**
	玉米重构序列 1	0.1801**	0.0311	0.0025
	玉米重构序列 2	-0.0039	0.1263**	-0.1985**
	玉米重构序列 3	-0.0005	0.1166**	0.9705**
LZ 复杂度聚类	豆粕重构序列 1	0.5208**	0.0166	0.0014
	豆粕重构序列 2	0.0087	0.4013**	0.1243**
	豆粕重构序列 3	-0.0002	0.0031	0.6557**
	玉米重构序列 1	0.2140**	0.0020	0.0006
	玉米重构序列 2	0.0024	0.1295**	-0.1975**
	玉米重构序列 3	-0.0002	0.1177**	0.9705**

对豆粕期货和玉米期货重构序列与大豆期货重构序列的相关性进行分析，即对大豆期货、豆粕期货和玉米期货的分解序列用某一复杂性指标、周期性指

标和相关性指标进行重构，然后对同一种多特征重构方法下的各个农产品期货重构序列的相关性进行分析，结果如表 4.10 所示。

从表 4.10 展现了不同多特征重构方法下各个农产品期货分解序列按某一种多特征重构方法进行重构的重构序列的皮尔逊相关系数，其中数字上角有**符号表示该数字代表的两重构序列的皮尔逊性系数相关性检验 P 值小于 0.05，表示两重构序列在统计学意义上显著相关。本文采取两种规则选取大豆期货重构序列预测输入自变量，第一种选取规则是选取与大豆期货重构序列相关性最强的豆粕期货价格和玉米期货价格的重构序列作为大豆期货重构序列预测的自变量，第二种规则是选取与大豆期货重构序列统计意义上显著相关的豆粕期货价格和玉米期货价格的重构序列作为大豆期货重构序列预测的自变量。

根据规则一，将豆粕期货与玉米期货的重构序列 1 作为大豆期货重构序列 1 序列预测的输入变量，将豆粕期货与玉米期货的重构序列 2 序列作为大豆期货重构序列 2 序列预测的输入变量，将豆粕期货与玉米期货的重构序列 3 序列作为大豆期货重构序列 3 序列预测的输入变量。根据规则二选取大豆期货重构序列预测输入自变量的结果如表 4.11 所示。

表 4.11 基于规则二的选择变量结果

重构方法	因变量	筛选的自变量
样本熵聚类	大豆重构序列 1	豆粕重构序列 1，玉米重构序列 1
	大豆重构序列 2	豆粕重构序列 2，玉米重构序列 2、3
	大豆重构序列 3	豆粕重构序列 2、3，玉米重构序列 2、3
近似熵聚类	大豆重构序列 1	豆粕重构序列 1，玉米重构序列 1
	大豆重构序列 2	豆粕重构序列 1、2，玉米重构序列 2、3
	大豆重构序列 3	豆粕重构序列 2、3，玉米重构序列 2、3
LZ 复杂度聚类	大豆重构序列 1	豆粕重构序列 1，玉米重构序列 1
	大豆重构序列 2	豆粕重构序列 2，玉米重构序列 2、3
	大豆重构序列 3	豆粕重构序列 2、3，玉米重构序列 2、3

分别根据规则一和规则二选择大豆期货价格重构序列的输入变量，加入其它期货市场信息后得到的预测效果如图 4.3、图 4.4 和表 4.12 所示。

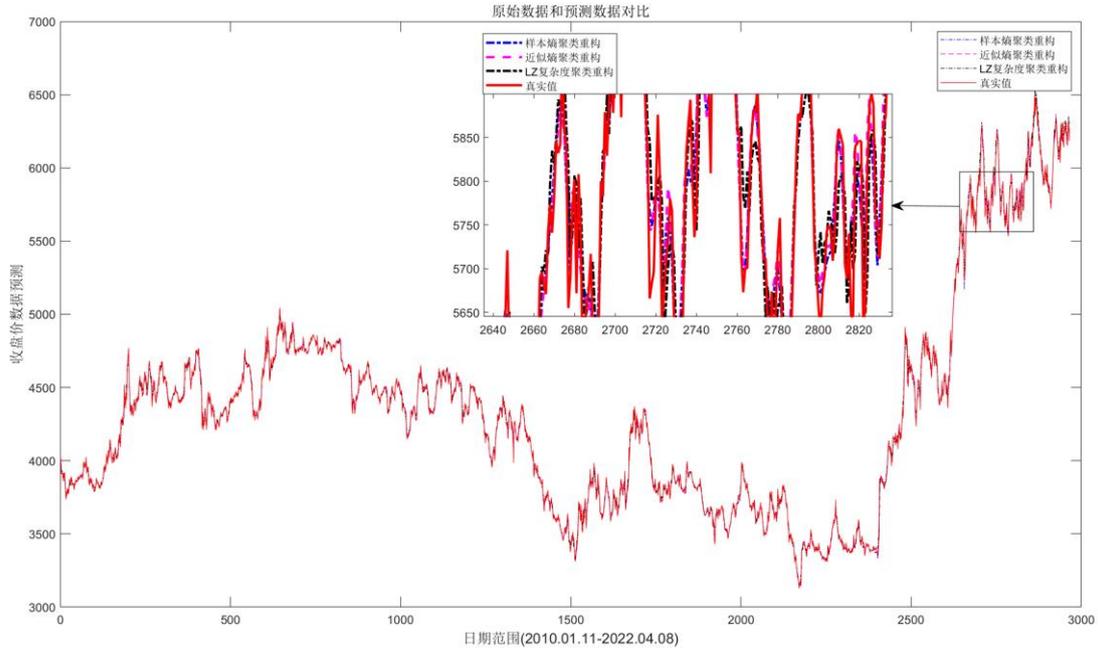


图 4.3 引入其它信息后不同重构方法的预测结果（规则一）

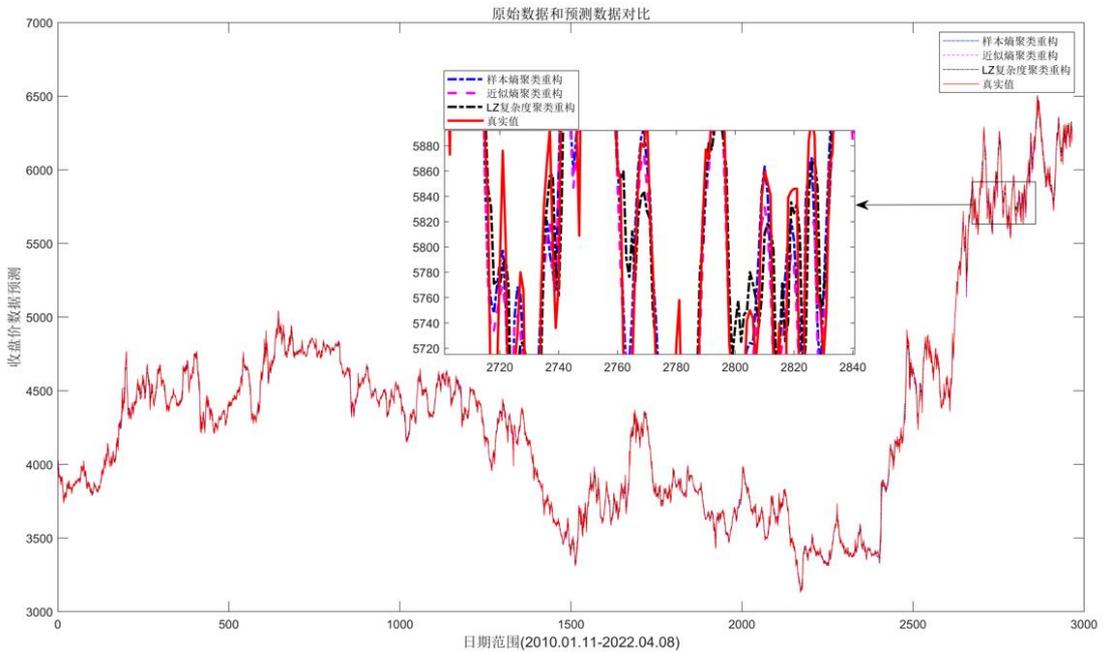


图 4.4 引入其它信息后不同重构方法的预测结果（规则二）

图 4.3 和图 4.4 展现了在两种输入变量筛选规则情况下，加入其它市场信息的考虑多个特征的不同重构方法的模型预测走势图，可以看出加入其它期货市场信息的模型预测十分贴合真实值。

表 4.12 不同预测模型效果比较

		RMSE	MAE	MAPE(%)	DA(%)
不引入其它信息	样本熵/近似熵聚类重构	67.58	35.82	0.68	78.96
	LZ 复杂度聚类重构	67.38	50.52	0.95	67.51
引入其它信息 (规则一)	样本熵聚类重构	40.32	30.08	0.57	81.48
	近似熵聚类重构	40.33	29.65	0.57	81.14
	LZ 复杂度聚类重构	59.06	44.28	0.85	70.54
引入其它信息 (规则二)	样本熵聚类重构	41.30	30.94	0.59	80.30
	近似熵聚类重构	43.50	32.60	0.61	79.97
	LZ 复杂度聚类重构	57.50	43.82	0.84	70.20

从表 4.12 可以看出, 引入豆粕期货与玉米期货的信息对于大豆期货价格的预测取得了更好的预测效果, 相较于只考虑大豆期货价格本身的信息, 从水平预测精度和方向预测精度来讲都有了效果提升。为进一步研究加入其它信息后的预测效果, 计算了改进率 IR 对预测结果进行分析, 结果如表 4.13 所示。在表 4.13 中, P_{MS1} 、 P_{MA1} 和 P_{ML1} 表示样本熵 SE 聚类重构、近似熵聚类重构和 LZ 复杂度聚类重构的预测模型加入用规则一筛选输入变量, P_{MS2} 、 P_{MA2} 和 P_{ML2} 则表示用规则一筛选输入变量的样本熵聚类重构、近似熵聚类重构和 LZ 复杂度聚类重构的预测模型, P_S 、 P_A 和 P_L 表示重构方法为样本熵 SE 聚类、近似熵 AE 聚类和 LZ 复杂度聚类的预测模型。

表 4.13 引入其它信息的模型预测改进率对比

	IR _{RMSE}	IR _{MAE}	IR _{MAPE}	IR _{DA}
$P_{MS1} \rightarrow P_S$	40.35%	16.02%	14.95%	3.20%
$P_{MA1} \rightarrow P_A$	40.32%	17.24%	16.07%	2.77%
$P_{ML1} \rightarrow P_L$	12.34%	12.36%	11.08%	4.49%
$P_{MS2} \rightarrow P_S$	38.89%	13.62%	12.70%	1.71%
$P_{MA2} \rightarrow P_A$	35.64%	9.01%	9.02%	1.28%
$P_{ML2} \rightarrow P_L$	14.65%	13.26%	11.72%	3.99%

从表 4.13 可以看出, 相较于不考虑其它外部信息, 将豆粕期货和玉米期货

信息通过规则一和规则二引入基于多特征重构的“分解-重构-集成”预测模型中，会使得模型在水平预测和方向预测表现更优，尤其是水平预测精度的提升，MAPE 值大多提升了 10%以上，方向预测精度则提升了 1%-4%。

4.3 本章小结

本章分为两个部分，一部分探讨了“分解-重构-集成”框架下，不同的重构方法对模型预测效果的影响，不同的重构方法分为只考虑单个特征的单特征重构方法与考虑多个特征的多特征重构方法，本文探讨的数据特征包括三种：复杂性、周期性和相关性，结果显示相对于只考虑周期性或相关性，仅考虑复杂性的重构方法更优，多特征重构方法优于单特征重构，不适当地使用数据特征对分解序列进行重构可能会降低模型性能。另一部分则是在基于多种数据特征重构方法的“分解-重构-去噪”模型的基础上，引入了豆粕期货和玉米期货的信息。对大豆期货价格、豆粕期货价格和玉米期货价格都进行了分解重构，并根据两种规则选择大豆期货重构序列的输入变量。其中，规则一是选取与大豆重构序列相关性最强的豆粕重构序列和玉米重构序列作为其输入变量，规则二则是筛选与大豆重构序列显著相关的豆粕重构序列和玉米重构序列作为其输入变量。结果显示基于两种规则下引入豆粕期货和玉米期货的信息，模型的预测性能都得到了提升。

5 总结与展望

5.1 总结

本篇论文将分解方法与机器学习方法相组合，基于分解去噪模型和“分解-重构-集成”模型对大豆期货的收盘价格进行了预测，通过本文的研究可以得到以下三个结论：

(1) 相较于 SSA-BP 模型和 SSA-SVR 模型，SSA-ELM 模型对大豆期货价格的预测更优；对大豆期货价格序列通过 ICEEMDAN 算法进行去噪处理后进行预测优于直接对大豆期货价格原序列进行预测。

(2) 在“分解-重构-集成”框架下，不同的重构方法会影响模型预测的效果。实证结果显示，对于单特征重构，考虑复杂性特征的重构方法优于考虑周期性特征和与原序列相关性特征的重构方法；综合考虑周期性、复杂性和与原序列的相关性三个数据特征的多特征重构方法优于仅考虑一个数据特征的单特征重构方法；不适当地使用数据特征对分解序列进行重构可能会降低模型性能。

(3) 将豆粕期货信息与玉米期货信息引入分解去噪模型和“分解-重构-集成”模型可以进一步提升模型预测效果。在分解去噪模型中，通过对豆粕和玉米的期货价格进行去噪处理的方法引入豆粕价格信息和玉米价格信息。在“分解-重构-去噪”模型中，在基于多特征重构方法的“分解-重构-去噪”模型的基础上，对大豆期货价格、豆粕期货价格和玉米期货价格都进行分解重构，然后对大豆期货价格重构序列与豆粕期货价格重构序列、玉米期货价格重构序列进行相关性研究，通过选择大豆期货重构序列的输入变量将玉米期货信息和豆粕期货信息引入到模型当中。实证结果显示，本文引入其它期货市场信息会提升大豆期货价格的预测效果。

5.2 展望

尽管本文的研究达到了预期效果，但由于本人能力所限，本文的研究还存在一些不足，不足之处如下所示：

(1) 在预测方法方面，可以考虑更加强大的深度学习算法，例如优化的 LSTM 算法。

(2) 本文针对大豆期货价格的预测，虽然引入了其它期货市场的信息，但引入的外部信息较少，可以考虑加入与政治事件、宏观经济、微观经济等相关的其他信息。

(3) 本文所构建的预测模型只在在大豆期货的预测中取得了一定的成果，可以考虑采用该模型对黄金期货价格、汇率等其他金融时间序列进行预测。

参考文献

- [1] 滕永平, 周婷婷. 大豆期货价格预测实证研究[J]. 沈阳工业大学学报(社会科学版), 2017, 10(04): 331-334.
- [2] 陈林, 黄章树. 基于 ARIMA 模型的期货价格分析与预测[J]. 福州大学学报(哲学社会科学版), 2010, 24(03): 32-37.
- [3] 刘峰, 王儒敬, 李传席. ARIMA 模型在农产品价格预测中的应用[J]. 计算机工程与应用, 2009, 45(25): 238-239+248.
- [4] 李苏, 宝哲. 我国猪肉价格波动特征及预测研究[J]. 价格理论与实践, 2020(06): 80-83+153.
- [5] 丁慧娟, 张金磊, 陈建中, 李均涛, 崔鹏. ARIMA 模型和灰色模型在农产品价格预测中的应用比较[J]. 安徽农业科学, 2018, 46(24): 191-194.
- [6] 郭静利, 董渤. 基于 SARIMA 模型的国际稻米价格预测[J]. 价格理论与实践, 2019, No. 415(01): 79-82.
- [7] 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, No.274(22): 146-147.
- [8] 何文琪. 基于 ARIMA 和线性回归组合模型的煤炭价格预测[J]. 中国管理信息化, 2021, 24(01): 187-188.
- [9] Ge Y, Wu H. Prediction of corn price fluctuation based on multiple linear regression analysis model under big data[J]. Neural Computing and Applications, 2020, 32: 16843-16855.
- [10] 段青玲, 张磊, 魏芳芳, 肖晓琰, 王亮. 基于时间序列 GA-SVR 的水产品价格预测模型及验证[J]. 农业工程学报, 2017, 33(01): 308-314.
- [11] 郑薇, 王灿强, 李维德. 基于季节指数调整与 HGWO-SVR 算法的农产品价格预测模型[J]. 统计与决策, 2018(19): 33-37.
- [12] Yu Y, Zhou H, Fu J. Research on agricultural product price forecasting model based on improved BP neural network[J]. Journal of Ambient Intelligence and Humanized Computing, 2018: 1-6.

- [13] 范俊明, 刘洪久, 胡彦蓉. 基于 LSTM 深度学习的大豆期货价格预测[J]. 价格月刊, 2021, No.525(02): 7-15.
- [14] 卢超凡, 史世凯, 王鲁. 基于改进 LSTM 的苹果价格预测模型研究[J]. 山东农业大学学报(自然科学版), 2022, 53(03): 491-496.
- [15] 张旸, 颜七笙. 基于 ABC-ELM 的上证综指收盘价格预测模型[J]. 计算机仿真, 2020, 37(05): 154-160.
- [16] 沈虹, 李旭, 潘琪. 基于深度学习长短期记忆神经网络的有色金属期货市场预测研究[J]. 南京理工大学学报, 2021, 45(03): 366-374.
- [17] 陈俊华, 郝彦惠, 郑丁文等. 基于 DBN 深度学习的期货市场价格预测建模与决策[J]. 计算机科学, 2018, 45(S1): 75-78+84.
- [18] 沈欣宜, 李旭, 沈虹. 基于机器学习的铜期货价格预测分析[J]. 扬州大学学报(自然科学版), 2021, 24(05): 1-7.
- [19] Zhao H. Futures price prediction of agricultural products based on machine learning[J]. Neural Computing and Applications, 2021, 33: 837-850.
- [20] Wang J, Wang Z, Li X, et al. Artificial bee colony-based combination approach to forecasting agricultural commodity prices[J]. International Journal of Forecasting, 2022, 38(1): 21-34.
- [21] 曹霜, 何玉成. 基于小波分解的 SVM-ARIMA 农产品价格预测模型[J]. 统计与决策, 2015(13): 92-95.
- [22] 王珏, 齐琛, 李明芳. 基于 SSA-ELM 的大宗商品价格预测研究[J]. 系统工程理论与实践, 2017, 37(08): 2004-2014.
- [23] 方雪清, 吴春胤, 俞守华, 张大斌, 欧阳庆. 基于 EEMD-LSTM 的农产品价格短期预测模型研究[J]. 中国管理科学, 2021, 29(11): 68-77.
- [24] Fang Y, Guan B, Wu S, et al. Optimal forecast combination based on ensemble empirical mode decomposition for agricultural commodity futures prices[J]. Journal of Forecasting, 2020, 39(6): 877-886.
- [25] Yu L, Ma Y. A Data-Trait-Driven Rolling Decomposition-Ensemble Model for Gasoline Consumption Forecasting[J]. Energies, 2021, 14(15): 4604.

- [26] Zhu J, Liu J, Wu P, et al. A novel decomposition-ensemble approach to crude oil price forecasting with evolution clustering and combined model[J]. *International Journal of Machine Learning and Cybernetics*, 2019, 10: 3349-3362.
- [27] Niu X, Wang J, Zhang L. Carbon price forecasting system based on error correction and divide-conquer strategies[J]. *Applied Soft Computing*, 2022, 118: 107935.
- [28] 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.
- [29] 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.
- [30] Yeh J R, Shieh J S, Huang N E. Complementary ensemble empirical mode decomposition: A novel noise enhanced data analysis method[J]. *Advances in Adaptive Data Analysis*, 2010, 2(2) : 135-156.
- [31] Torres M E, Colomians M A, Schlotthuer G, et al. Complete ensemble empirical mode decomposition with adaptive noise[J]. *Brain Research Bulletin*, 2011, 125(3) : 4144-4147.
- [32] Niu H L, Xu K L, Liu C. A decomposition-ensemble model with regrouping method and attention-based gated recurrent unit network for energy price prediction[J]. *Energy*, 2021, 231: 120941.
- [33] 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.
- [34] Pincus S M, Huang W M. Approximate entropy: statistical properties and applications[J]. *Communications in Statistics-Theory and Methods*, 1992,

21(11) : 3061-3077.

- [35] Lempel A, Ziv J. On the complexity of finite sequences[J]. IEEE Transactions on information theory, 1976, 22(1) : 75-81.
- [36] 解幸幸, 李舒, 张春利, 李建康. Lempel-Ziv 复杂度在非线性检测中的应用研究[J]. 复杂系统与复杂性科学, 2005(03) : 61-66.
- [37] Kiran Kumar G, Parimalasundar E, Elangovan D, et al. Fault investigation in cascaded H-bridge multilevel inverter through fast fourier transform and artificial neural network approach[J]. Energies, 2020, 13(6) : 1299.
- [38] Rumelhart D E, Hinton G E, Williams R J. Learning epresentations by back-propagating errors[J]. Nature, 1986, 323(6088) : 533-536.
- [39] J. A. K. Suykens, J. Vandewalle. Least Squares Support Vector Machine Classifiers. [J]. Neural Processing Letters, 1999, 9(3) : 293-300.
- [40] Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: theory and applications[J]. Neurocomputing, 2006, 70(1-3) : 489-501.
- [41] Xue J, Shen B. A novel swarm intelligence optimization approach: sparrow search algorithm[J]. Systems Science & Control Engineering, 2020, 8(1) : 22-34.

后 记

研究生求学的时光转瞬即逝，仿佛昨日才拿到兰财的录取通知书，在研究生生涯的尾端清晰地感受到了当初收到通知书那一刻开心而又忐忑的心情，仔细回忆了入学前对于研究生生活的规划与目标，如今三年过去，我的研究生之路也算是不虚此行。在这三年中，无论是在学识方面还是在性格、思想和个人认知等方面，我都收获颇多，其中也不乏些许遗憾，但这些遗憾也点缀了我的人生。凡是过往皆为经历，我在这些经历中感受积极或是消极的情绪，感受作为一个人的鲜活。在兰财的三年，我喜欢上了兰州这个城市，遇到了许多可爱的人，给予了我温暖和帮助。

首先我想要感谢的就是我的导师孙景云教授，感谢您三年来对我的教导和宽容。孙景云老师对我的教导不仅仅是在学业上不厌其烦的指导，孙老师作为一个优秀的人，言传身教，他勤勉严谨、极具探索精神并对工作和生活抱有源源不断积极向上的态度等等优秀的品性使我也在不断地成长，我很幸运能够成为孙老师的学生。

然后，我想感谢的是我可爱的朋友们，感谢我的几位舍友，我的师门以及我的两位师姐，同你们愉快地共度了一段时光，感谢你们给予我的温暖与包容。

最后，我想感谢我其它的一些朋友与我的亲人，无论是在物质上或是精神上，你们一直对我给予了支持。

虽然我的研究生生涯即将结束，我的学生时代即将成为过往，但我的人生之路仍在继续，希望在以后的人生道路中能多多体验不一样的风景，永葆一颗鲜活跳动的心。