

分类号 _____
UDC _____

学校代码 10590 _____
密 级 _____

深圳大学硕士学位论文

基于深度学习的金融时间序列趋势预测算法研究

徐子超

学 位 类 别	工程硕士专业学位
专 业 名 称	计算机技术
学院（系、所）	计算机与软件学院
指 导 教 师	陈剑勇 教授

深圳大学学位论文原创性声明和使用授权说明

原创性声明

本人郑重声明： 所提交的学位论文_____是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不含任何其他个人或集体已经发表或撰写过的作品或成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本声明的法律结果由本人承担。

论文作者签名：

日期： 年 月 日

学位论文使用授权说明

(必须装订在印刷本首页)

本学位论文作者完全了解深圳大学关于收集、保存、使用学位论文的规定，即：研究生在校攻读学位期间论文工作的知识产权单位属深圳大学。学校有权保留学位论文并向国家主管部门或其他机构送交论文的电子版和纸质版，允许论文被查阅和借阅。本人授权深圳大学可以将学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文。

(涉密学位论文在解密后适用本授权书)

论文作者签名：

导师签名：

日期： 年 月 日

日期： 年 月 日

摘 要

对金融时间序列未来趋势的预测，不管是对国家经济调控，还是对个人投资者以及机构投资者都有着重要的意义，他们可以基于对金融时间序列未来趋势的预测和判断，来进行相关决策，达到规避风险、获取超额收益的目的。如果个人或机构投资者等市场参与者能够准确预测市场行为，这将使他们能够持续获得比市场行情更高的风险调整后回报。

近年来，随着计算机性能的不断提升，深度学习等人工智能技术得到了大量的研究和应用，并且深度学习模型在金融时间序列预测领域的探索也越来越多。但是，由于金融时间序列是非平稳、高波动的序列，并且影响其变动的因素非常多，利用传统的统计学方法或者单一的深度学习方法来进行金融时间序列趋势预测面临着巨大困难。因此，如何处理好金融时间序列这类高波动非平稳数据，成为了深度学习模型完成预测任务的关键。

本文针对上述问题，做了以下研究工作：

(1) 本文提出了一个新的基于注意力机制的深度模态分解编码解码模型（AED-DEMD）。该模型由本文设计的基于注意力机制的编解码模型 AED 和深度模态分解模块 DEMD 组成，其中 DEMD 通过独特的堆栈结构和残差连接能够针对高波动的金融时间序列进行细致的降噪处理，减少了序列数据中存在的高频波动分量，降低了序列的样本熵值，从而提高了金融时间序列趋势的可预测性。

(2) 针对 (1) 中 DEMD 需要人工选择、泛化能力不足的问题，提出了一种新的管道结构 EMD-OPS。EMD-OPS 通过分离式端口来实现时间序列数据的分解与合成的功能。其中，最优表现选择模块（OPS）通过内置的模态选择模块、回测交易模块以及评估模块，可以实现自适应地以收益为导向的自主最优表现模态组合选择。通过实验证明，这种管道结构明显有助于提升模型的泛化能力，解决了 DEMD 需要人工选择而导致的泛化能力不足的问题。最后，结合 (1) 中的 AED，提出了 EMD-AED-OPS，实验结果表明，其各项性能有所提升。

(3) 本文针对多步金融时间序列趋势预测问题，提出了一种新的基于双端延拓模

态分解的最优表现选择模型 Opemod。首先, 本文在 EMD 的基础上提出了一种双端延拓模态分解算法 TEEMD, 其通过在时间序列的两端分别延拓的办法解决了传统 EMD 在分解时间序列过程中存在明显的端点效应问题, 减小了时间序列分解的误差, 提高了分解的准确性。其次, 继承了(2)中设计的最优表现选择模型 OPS, 并结合提出的 TEEMD 设计了 TEEMD-OPS 管道结构, 这个结构可用于时间序列数据的分解与自适应合成。最后, 结合(1)中提出的 AED, 构建了 Opemod。大量实验结果表明, Opemod 在多数指标上能够取得较好的性能。

关键词: 深度学习; 金融时间序列; 预测; 经验模态分解; 长短期记忆网络

目 录

摘 要	I
Abstract.....	III
第 1 章 绪论	1
1.1 研究背景	1
1.2 研究现状	2
1.3 研究内容与论文结构	3
1.3.1 主要研究内容	3
1.3.2 论文组织结构	4
第 2 章 相关理论基础	6
2.1 长短期记忆网络	6
2.2 经验模态分解	7
2.3 样本熵值理论	8
2.4 数据集与预处理	9
2.5 多步预测方法	11
2.6 交易策略和评价指标	11
2.7 本章小结	13
第 3 章 基于注意力机制的深度模态分解编码解码模型 AED-DEMD	14
3.1 引言	14
3.2 模型介绍	15
3.2.1 AED-DEMD 总体架构	15
3.2.2 动态分解 EMD 层	16
3.2.3 深度经验模态分解 DEMD	17
3.2.4 基于多头注意力机制的编码解码模型 AED	19
3.3 实验结果与分析	21
3.3.1 实验设置	21
3.3.2 实验结果	22
3.3.3 DEMD 结果	26
3.3.4 样本熵与能量值结果展示与分析	28

3.4 本章小结	33
第 4 章 基于注意力机制的自适应最优表现选择模型	35
4.1 引言	35
4.2 模型介绍	36
4.2.1 EMD-AED-OPS 模型总体架构	36
4.2.2 EMD-OPS 自适应管道模型	36
4.2.3 AED 模块	39
4.3 实验结果与分析	39
4.3.1 实验设置	39
4.3.2 交易结果与讨论	40
4.4 本章小结	43
第 5 章 基于双端延拓模态分解优化的最优表现选择模型 Opemod	45
5.1 引言	45
5.2 模型介绍	45
5.2.1 Opemod 总体架构	45
5.2.2 TEEMD 算法	46
5.2.3 AED 模型	48
5.2.4 最优表现选择模型 OPS	49
5.3 实验结果与分析	49
5.3.1 实验设置	49
5.3.2 TEEMD 结果展示与讨论	51
5.3.3 交易结果与讨论	52
5.3.4 稳定性测试	55
5.4 本章小结	58
第 6 章 总结与展望	59
5.1 工作总结	59
5.2 后续展望	60
参考文献	61
致 谢	66
攻读硕士学位期间的研究成果	67

第 1 章 绪论

1.1 研究背景

对金融时间序列未来趋势的预测，不管是对国家经济调控，还是个人投资者以及机构投资者都有着重要的意义，他们可以通过对金融时间序列未来趋势的判断和预测，来进行相关决策，来达到规避风险、获取超额收益的目的。量化投资是当前比较流行的一种方法，它是指凭借现代计算机强大的计算能力，再利用相关的数学或者统计等量化模型，发掘历史金融时间序列数据中存在的规律，以此计算出对应的未来趋势。在此基础上，进行投资决策，并获得收益的一种投资方式^[1]。在过去的几十年里，人们对股票市场的兴趣呈指数级增长。因此，每天都有价值数十亿美元的资产在证券交易所交易也就不足为奇了。投资者们在资本市场上十分活跃，希望在其投资期限内实现盈利。如果个人或机构投资者等市场参与者能够准确预测市场行为，这将使他们能够持续获得比市场行情更高的风险调整后回报。

欧美发达国家的资本市场拥有悠久的历史 and 完善的制度，世界上有很多知名的金融机构采取了量化的方式来进行投资，这样可以将交易策略规范化、模型化，能够更加有效地进行风险管理。截止至2022年1月，全球对冲基金管理的资产首次突破4万亿美元，其中量化基金占到30%以上。与欧美发达国家相比，我国的量化投资行业虽然起步较晚，但是经过十多年的高速发展，已经形成了相当的规模。从2019年开始，我国量化私募行业规模高速增长，在2020年达到了8850亿元，在2021年上半年突破万亿大关。截至2021年末，量化私募基金规模已经达到了1.5万亿元。

量化交易是一个复合学科，需要使用到许多的统计学以及计算机专业技术。例如，可以利用统计方法来分析历史金融时间序列数据序列的趋势性与周期性规律，以此来对未来进行合理的预测；利用新颖的数据挖掘技术可以在资本市场的海量数据中提取大量的数据特征，挖掘它们的相关性，从而寻找可能有效的规律；随着计算机的计算性能的不断升级，人工智能和深度学习等技术得到了长足的发展，有许多的方法和模型可以帮助人们更加高效地处理市场中的各种信息特征，并且人们可以利用它们对市场行情进行预测。

然而，影响股票市场的因素非常多，各种金融时间序列数据也随着交易过程不断地波动变化，因此这是一个非常复杂的市场。影响金融时间序列数据变化的因素不但包括

微观方面的，例如公司的经营以及财务状况、投资者的情绪等。也有宏观方面的，例如通货膨胀、国家宏观经济政策、国际形势、自然灾害、战争等。因此，金融时间序列数据看似简单，只是一系列时间序列数据，但其中包含了大量的非平稳和不可控因素。而这些非平稳的混沌信号数据如果不能得到很好的处理，将会对人们使用的方法或者模型产生巨大的干扰，甚至使得该系统得到完全错误的结论和判断。综上所述，针对金融时间序列的趋势预测任务面临着巨大的挑战，如何针对这类数据进行合理地处理？如何有效地提取出市场中真正有用的、对未来趋势产生影响的信息，以协助深度学习预测模型来进行预测，已经成为了当前工作的主要难题所在。

1.2 研究现状

多步股价预测对于股票市场的趋势投资至关重要，因为可以通过适当地确定投资资产的未来价格变化来降低决策风险。然而，股票价格序列是非线性、非平稳、混沌的信号，它们可能受到许多因素的影响，如利率、通货膨胀率、交易者的期望、灾难、政治事件和经济环境^[2]。如果噪声和异常情况不能被有效地检测到，那么多步预测的风险将会非常高，所以对未来的股票价格的预测很有挑战性。在使用传统统计学和机器学习方法来分析过去的股票趋势和预测未来的股票价格方面已经有了许多工作。在时间序列预测任务中使用的传统统计方法有很多，包括差分整合移动自回归平均模型（ARIMA）^[3]用于非平稳数据序列预测，简单指数平滑法（SES）^[4]用于预测时间序列，以及霍尔特阻尼指数平滑法^[5]。此外，还有许多的传统的机器学习技术已经用于时间序列预测，如支持向量回归（SVR）^[6]和各种混合方法^[7-8]。随着深度学习方法的在各个领域^[9-14]取得了众多成果，现在也有许多深度学习模型被广泛应用于时间序列预测任务中^[15-21]。LeCun等人^[22]的研究指出，只要给定相应的数据，就可以利用 LSTM 找到股票趋势中包含的特征。有许多方法^[23,24,25]已经表明，深度学习方法优于使用手工特征的传统机器学习方法。如递归神经网络（RNNs）用于预测股票价格^[26,27]。长短期记忆网络（LSTM）^[28]被提出来解决长时间依赖遗忘的问题，并且一些使用 LSTM 的股票序列预测方法被提出^[29,30]。^[26]将 autoencoders 和 LSTM 结合在一起，预测全球各个不同市场的指数，并在各种发展中市场和发达市场中显示了其预测能力。Jing N 等人^[31]利用 CNN 建立一个情绪分类器，并将其与 LSTM 相结合来预测股票未来的价格。

然而，传统的 LSTM 预测一步股价的能力有限，因为短期股价受到大量噪声的影响，是嘈杂和不稳定的，仅靠 LSTM 难以预测。相比之下，LSTM 更适合于多步预测，因为

它具有存储单元并且可以保持序列的模式。但是，多步股价预测的主要难点在于它需要对整个序列进行预测，而后面预测的价格都是基于之前的预测价格。因此，随着迭代过程的进行，预测的误差将逐步累积。Mid-LSTM^[30]被提出用来预测中期股票价格。它使用隐马尔可夫模型（HMM）^[32]去推导交易量与股票价格之间的隐藏状态，同时利用 Correlator 计算大盘指数与个股股价之间的皮尔逊相关系数，最后用这两者的输出去优化 LSTM 的预测结果。然而，HMM 是无后效性的，即当前状态仅与前一个状态有关，而与更前面的状态无关，因此 Mid-LSTM 并不适用于短期趋势的预测并且很难关注到整个时间窗口上的模式变化，并且量价状态变化一定程度上依赖人为经验判断。为了避免人为因素，Oreshkin B N 等人提出了 N-BEATS^[33]，它基于后向和前向残差连接以及一个非常深的全连接层堆栈设计的，用于解决时间序列的点预测问题，不过目前只能用于单变量时间序列预测。尽管各种新颖的模型的体系结构中的细节各不相同，但它们表明深度学习方法在股票市场中寻找非线性模式是有效的。然而，如 Rojat T 等人^[34]所述，深度学习方法的主要缺点是缺乏可解释性，而时间序列预测在现实世界中的应用，如自动驾驶领域和金融领域，都是需要比较强的可解释性的领域。与此同时，Makridakis S 等人^[35]的研究显示，传统的统计方法在可解释性和稳定性方面仍然具有明显的优势。Smyl S 等人^[36]向 M4 时间序列预测竞赛提交了一种创新的混合方法，该方法将指数平滑（ES）公式与递归神经网络预测模型混合，并且取得了更好的效果。

1.3 研究内容与论文结构

1.3.1 主要研究内容

本文的研究内容是面向以深度学习模型为基础的金融时间序列短期趋势预测问题研究和优化。本文通过对金融时间序列特性进行深入分析，并对深度学习模型在股票价格预测上存在的问题进行研究，提出了三种基于统计和深度学习的混合模型来进行金融时间序列短期趋势预测任务。同时，本文进行了大量的实验分析，与现有的研究进行了对比，分析了本文所提出的方案的优劣，并对模型的原理进行了详细的阐述和讨论。本文的主要研究内容如下：

（1）本文提出了一个新的基于注意力机制的深度模态分解编码解码模型 AED-DEMD。该模型通过其独特的堆栈结构针对高波动的金融时间序列进行细致的降噪处理，减少了金融时间序列数据中存在的高频波动分量，降低了金融时间序列数据的样本熵值，

从而提高了时间序列趋势的可预测性。首先,本文利用经验模态分解(EMD)^[37]来对金融时间序列进行分解处理,去除其中最高频不规则分量,降低其对序列趋势的影响。并且,针对 EMD 存在的模态混叠问题^[38],提出了深度模态分解(DEMD),能够在降低模态混叠效应的影响的同时,尽可能的保留原始数据中的趋势信息。接着,设计实现了一种基于注意力机制(Attention)^[39]的编码解码模型 AED,该模型的编码器和解码器由 LSTM 组成,AED 能够利用 Attention 在整个时间窗口上可学习地关注不同的模式,增强 LSTM 对不同模式的记忆,从而使模型更加准确的预测时间序列的趋势,同时,AED 利用残差连接来链接 Encoder 和 Attention 模块,减小搜索范围,提高收敛速度。并且,将提出的 DEMD 与 AED 进行结合,形成 AED-DEMD 模型。其中,DEMD 用来处理信息,形成原始序列的趋势信息,而 AED 用来处理得到的趋势信息,以预测未来趋势,使得模型达到更优的效果。

(2) 针对(1)中 DEMD 需要人工选择、泛化能力不足的问题,提出了一种新的管道结构 EMD-OPS。EMD-OPS 通过分离式端口来实现时间序列数据的模态分解与自适应合成的功能。其中,OPS 通过内置的模态选择模块、回测交易模块以及评估模块,可以实现自适应的、以收益为导向的自主最优表现模态组合选择。通过实验证明,这种方式明显有助于提升模型的泛化能力,解决了 DEMD 需要人工选择而导致的泛化能力不足的问题。最后,结合(1)中的 AED,提出了 EMD-AED-OPS,实验结果表明,其各项性能有所提升。

(3) 本文针对多步金融时间序列趋势预测问题,提出了一种新的基于双端延拓模态分解优化的最优表现选择模型 Opemod。首先,在 EMD 的基础上提出了一种新的双端延拓模态分解算法 TEEMD,其通过在金融时间序列两端分别延拓的办法有效抑制了传统 EMD 分解时间序列过程中存在明显的端点效应,减小了时间序列分解的误差,提高了分解的准确性。其次,继承了(2)中设计的最优表现选择模型 OPS,结合提出的 TEEMD 设计了 TEEMD-OPS 管道结构,该结构可用于时间序列数据的模态分解与自适应合成。最后,结合(1)中提出的 AED,构建了 Opemod。大量实验表明,Opemod 在多数指标上取得了较好的性能。

1.3.2 论文组织结构

本文的组织结构如下:

第1章介绍了论文的研究背景,概括了当前深度学习模型在金融时间序列预测上的

研究现状以及当前针对高波动性的金融时间序列数据降噪的一些解决方案。另外阐述了本文的主要研究内容以及文章的组织结构。

第2章介绍了相关理论基础。主要对数据预处理和多步预测方法进行了表述,并介绍了经验模态分解算法的优势和局限性。对本文中使用到的所有实验数据集、数据封装方法、评价指标等做了介绍,同时,定义了一种简单朴素的自动交易策略,能够帮助本文中所有的预测模型完成自动交易任务,以显式地展示和对比各大模型的收益率表现。

第3章提出了基于注意力机制的编码解码模型AED-DEMD。首先对AED-DEMD中的基础背景做了阐述,然后对AED-DEMD进行了详细介绍,包括AED-DEMD的整体架构,模型的各个组成部分等。最后,进行了相应的仿真实验,并对DEMD在时间序列降噪上的性能进行了详细的展示与分析。

第4章提出了基于注意力机制的最优表现选择模型EMD-AED-OPS。首先对第3章中AED-DEMD模型的缺点和局限性进行了分析,然后介绍了新的EMD-AED-OPS模型的架构。最后,进行了相应的仿真实验,并对实验结果进行了详细的分析。

第5章提出了新颖的基于双端延拓模态分解优化的最优表现选择模型Opemod。首先对EMD存在的缺点进行了阐述,接着对Opemod模型的架构进行了介绍,并着重介绍了所提出的双端延拓模态分解算法TEEMD。通过仿真实验,验证了TEEMD算法能够有效抑制EMD中存在的端点效应问题。最后,对大量的实验结果进行了展示与分析。

第6章对本文进行总结,并对未来的工作进行了展望。

第 2 章 相关理论基础

2.1 长短期记忆网络

长短期记忆网络^[28]是循环神经网络 (Recurrent Neural Network, RNN) 的一种, LSTM 在 RNN 的基础上, 解决了传统 RNN 梯度消失^[40]的问题, 并且具有更好的学习长期依赖的能力。LSTM 很适合用于处理有时间依赖性的时间序列数据。这么多年来, LSTM 已经在很多领域进行了探索和研究, 例如自然语言处理、机器翻译、用电量预测、天气预测等, 并且在这些领域得到了广泛的应用。

LSTM 通过在隐藏层的各个神经单元中增加记忆单元, 记忆单元能够记住通过隐藏层的信息。LSTM 的结构如图 2.1 所示。从图 2.1 中可以看到, 通过三个可被控制的门结构, 即输入门 (input gate)、遗忘门 (forget gate)、输出门 (output gate), 信息在不同的门之间传递, 并以可学习的方式决定记忆和遗忘的程度, 学习长期依赖的能力得到了有效的提升。

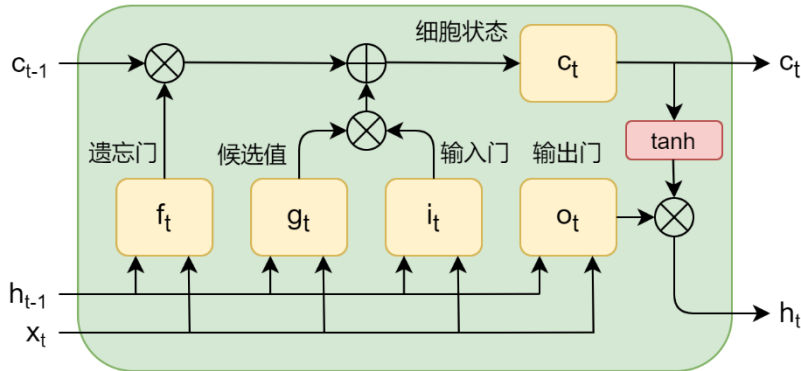


图 2.1 LSTM 结构图^[30]

在 LSTM 中, 首先需要决定从所有信息中“遗忘”掉哪些信息, 这个功能由“遗忘门”来完成。在单元状态 c_{t-1} 上, 对于前一时刻的隐藏层状态 h_{t-1} 和当前时刻获取到的信息 x_t , “遗忘门”会输出一个在 0 到 1 之间的数字 f_t , 该数字代表在 t 时刻应该保留信息的程度。其计算公式如下:

$$f_t = \sigma(W_{if}x_t + b_{if} + W_{hf}h_{t-1} + b_{hf})^{[30]} \quad (2.1)$$

接着, LSTM 需要决定存储哪些新的信息。这个步骤由“输入门”的 sigmoid 和 \tanh 层协作完成, “输入门”的输出值 i_t 和临时细胞状态 g_t 的计算如公式 (2.2) 和 (2.3) 所示:

$$i_t = \sigma(W_{ii}x_t + b_{ii} + W_{hi}h_{t-1} + b_{hi})^{[30]} \quad (2.2)$$

$$g_t = \tanh(W_{ig}x_t + b_{ig} + W_{hg}h_{t-1} + b_{hg})^{[30]} \quad (2.3)$$

紧接着，根据 t 时刻遗忘门的输出 f_t ， $t-1$ 时刻的细胞状态 c_{t-1} ，“输入门”的输出值 i_t 和临时细胞状态 g_t ，计算出当前时刻的细胞状态 c_t 。

$$c_t = f_t * c_{t-1} + i_t * g_t^{[30]} \quad (2.4)$$

最后，在“输出门”的作用下，LSTM 根据当前 t 时刻的细胞状态 c_t ， $t-1$ 时刻的隐藏层状态 h_{t-1} 和当前 t 时刻的信息 x_t ，计算出“输出门”的输出值和当前 t 时刻的隐藏状态 h_t 。

$$o_t = \sigma(W_{io}x_t + b_{io} + W_{ho}h_{t-1} + b_{ho})^{[30]} \quad (2.5)$$

$$h_t = o_t * \tanh(c_t)^{[30]} \quad (2.6)$$

综上所述，通过在时间维度上不断迭代上述过程，LSTM 可以得到与输入时间序列维度相同的隐藏层序列数据 $\{h_0, h_1, \dots, h_t\}$ 。在上述的各个计算公式当中， \tanh 是激活函数， σ 代表 sigmoid 激活函数， W 代表各层的权重， b 代表偏置项。

2.2 经验模态分解

经验模态分解 EMD^[41,42] 是 Huang 等人提出的，也成为希尔伯特-黄经验模态分解。EMD 是一种基于傅里叶变换的、对非平稳信号进行自适应分解的方法。EMD 可以对序列数据中不同时间尺度的波动或趋势成分进行逐级迭代分解，产生若干个不同特征尺度的序列信号-本征模态分量(IMF)。EMD 对每个 IMF 作 Hilbert 变换，利用解析信号相位求导后可以得到有物理意义的瞬时频率和幅值，所以时间-瞬时频率-瞬时幅值特征分布也就被计算出来了。EMD 适合处理非平稳、高波动时间序列信号，因其能够采用瞬时频率来描述信号特征，所以它在故障诊断^[43]等方面有着明显的优势和潜力。然而，EMD 本身存在许多有待改进的缺陷^[31]，其中一个就是存在模态混叠^[44]问题。近年来，许多研究人员又提出了多种针对其模态混叠问题的改进办法^[44-52]，但在这些方法中引入了新的噪声信号。

EMD 的基本计算过程是用“筛”的方法，通过迭代过程，一次次地把一个复杂的序列信号分解为若干个 IMF 信号之和，对于输入序列信号数据 $x(t)$ ，EMD 计算步骤如下：

(1) 计算序列 $x(t)$ 所有的局部极大值点和极小值点；

(2)经三次样条插值算法,计算出 $x(t)$ 的上(极大值点)、下(极小值点)包络线,分别表示为 $u(t)$ 和 $v(t)$;

(3)计算上、下包络线的均值序列: $m(t) = \frac{u(t)+v(t)}{2}$;

(4)求 $x(t)$ 与 $m(t)$ 的差: $h(t) = x(t) - m(t)$;

(5)不断迭代上述步骤,直到 $h(t)$ 满足预设的条件。此时,计算得第 1 个 IMF 分量 $c_1(t)$, 它表示原始信号中频率最高的成分;

(6)用 $x(t)$ 减去 $c_1(t)$ 继续重复上述计算过程。最终可以计算得到所有的 IMF 以及一个残余分量,这些分量与原始信号满足可加性,即:

$$x(t) = \sum_{i=1}^n c_i(t) + r_n(t) \quad (2.7)$$

如上式,原始序列信号 $x(t)$ 被分解为 n 个 IMF 和一个残余分量,分量 $c_i(t)$ 代表原始序列信号中的不同频率成分,它们按照频率从高到低排列。

2.3 样本熵值理论

近似熵 (ApEn) 是一种用于量化时间序列波动的规律性和不可预测性的参数。样本熵 (SampEn) 与近似熵的物理意义相似,都是通过测量信号中产生新模式的概率大小来衡量该时间序列的复杂性,产生新模式的概率越大,该序列的复杂性就越大。与近似熵相比,样本熵具有两个特点:(1) 样本熵的计算不依赖数据长度;(2) 样本熵具有更好的一致性。其中,时间序列样本熵值的详细计算步骤如算法 2.1 所示。

样本熵实际反应的是一个时间序列的复杂程度和可被预测的概率。一个时间序列的样本熵值越低,其自我相似性就越高,也就说明其可被预测的概率越大;相反,一个时间序列的样本熵值越大,其自我相似度就越低,即该时间序列就越复杂,可被预测的概率就越低,可预测性就越低。因此,本文在针对高波动、非平稳的金融时间序列预测任务中,样本熵可以用来评估金融时间序列的复杂程度和可预测性,并针对降低时间序列样本熵值做了相应的研究。

算法 2.1 样本熵算法^[58]

输入: 原始时间序列 $x(N)$

参数: 时间序列长度 N , 比较向量长度 m , 相似度量 r

输出: $x(N)$ 的样本熵值 $SampEn(x, m, r)$

定义 $X_m = \begin{bmatrix} x(1) & \cdots & x(m) \\ \vdots & \ddots & \vdots \\ x(N-m) & \cdots & x(N-1) \end{bmatrix}$

for $i = 1$ to $N - m$

for $j = 1$ to $N - m$ ($j \neq i$)

 令 $d(i, j) = \max(|X_m(i, 1) - X_m(j, 1)|, |X_m(i, 2) - X_m(j, 2)|, \dots, |X_m(i, i + m - 1) - X_m(j, j + m - 1)|)$

 令 $D(i, j) = H(d(i, j))$

end for

 令 $B(i) = \frac{1}{N-m-1} \sum_{j=1}^{N-m+1} D(i, j)$

end for

令 $\phi_m = \frac{1}{N-m} \sum_{i=1}^{N-m} B(i)$

定义 $X_{m+1} = \begin{bmatrix} x(1) & \cdots & x(m+1) \\ \vdots & \ddots & \vdots \\ x(N-m) & \cdots & x(N) \end{bmatrix}$

for $i = 1$ to $N - m$

for $j = 1$ to $N - m$ ($j \neq i$)

 令 $d(i, j) = \max(|X_{m+1}(i, 1) - X_{m+1}(j, 1)|, |X_{m+1}(i, 2) - X_{m+1}(j, 2)|, \dots, |X_{m+1}(i, i + m) - X_{m+1}(j, j + m)|)$

 令 $D(i, j) = H(d(i, j))$

end for

 令 $B(i) = \frac{1}{N-m-1} \sum_{j=1}^{N-m+1} D(i, j)$

end for

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

得样本熵值 $SampEn(x, m, r) = \ln \phi_m - \ln \phi_{m+1}$

2.4 数据集与预处理

表 2.1 实验数据集

数据集	日期范围	训练集天数	测试集天数
SSE 50	2005.04.20-2021.07.30	3540	360
CSI 300	2005.04.20-2021.07.30	3540	360
DJI	2011.02.03-2022.01.20	2340	360

本文中，将基于上证 50ETF（SSE 50）、沪深 300 指数（CSI 300）以及道琼斯指数（DJI）三个数据集来进行所有算法的实验，之所以选择这三个数据集是因为上证 50 由上海证券交易所上市的最具有代表性的 50 只股票组成，沪深 300 指数是由上海证券交易所和深圳证券交易所两市上市的最具代表性的 300 只股票组成，而道琼斯指数是世界上最具有影响力的指数之一，它由纽约证券交易所最具有代表性的 65 只股票组成，所以这三个数据集大体上表示了中国和美国股市的整体行情，非常具有代表性。本文使用的每日数据是三维的，包括每日收盘价，大盘指数以及交易量。其中，除了以上三个直接预测标的，本文使用的指数以及交易量数据部分分别对应为上证指数、上证指数以及标普 500 指数。本文所使用的数据集详细信息如表 2.1 所示。

不同的股票和同一只股票的不同时期都有不同的范围和规模。数据归一化被定义为将在不同尺度上测量的值调整到统一尺度^[53]。为了检测多步序列趋势的模式，需要对股价数据进行归一化处理。

本文使用“max-min”规范化方法来预处理数据集，它保留了数据的原始模式，如下所示：

$$X(t) = \frac{x(t) - x_{min}}{x_{max} - x_{min}} \quad (2.8)$$

其中 $x(t)$ 和 $X(t)$ 分别表示原始数据和归一化后的数据。同时，在时间序列预测领域，因为数据有严格的时间先后关系，因此在实验过程中尤其要注意到数据泄露的问题。因此，本文的具体做法为：先将原始历史数据严格分割为训练集和测试集，然后在训练集上寻找最大值和最小值，并用它们对训练集数据进行整体的归一化操作。在此基础上，直接将训练集上的参数迁移到测试集上对测试集数据进行归一化操作。另外，在测试阶段，所有的数据处理都采用滑动窗口分段式处理，测试一段处理一段，以这种方式来严格防止数据隐性泄露。

如图 2.2 所示，本文使用滑动窗口来打包数据，并将窗口大小设置为 60，这样本文构建的网络就可以粗略地看到整个时间序列的趋势，希望它在观察历史数据的基础上，建立一种对于未来趋势的判断。根据滚动窗口的大小，将训练数据分为多个训练输入集（每个集合长度为 59）和训练输出集（每个集合长度为 59）。测试数据分为 72 个窗口的输入数据和输出数据。

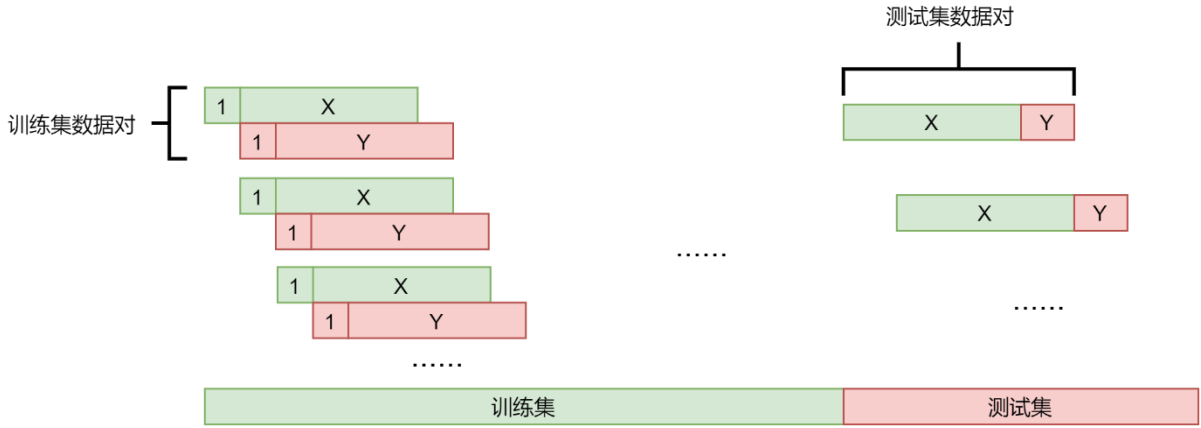


图 2.2 数据封装示意图

2.5 多步预测方法

本文的任务是股票价格的多步趋势预测（5 天）。与逐点预测处理不同，多步预测基于先前预测中预测的数据预测几天的价格，因此本文选择迭代预测方法。具体来说，首先用所有真实数据初始化窗口，将窗口向后滑动一步，并将第一个预测点添加到窗口的最后一个点。然后，新窗口被馈送到模型中以预测下一点。在本文的实验中，上述过程被迭代 5 次，因此可以获得未来 5 天的预测序列。

2.6 交易策略和评价指标

为了全面的评估模型的性能，本文使用了预测精度，收益率以及风险三大方面的一共六大指标，它们分别是对称平均绝对百分比误差（*sMAPE*），方向准确率（*DA*），累计收益率（*CR*），年化收益率（*AR*）、夏普比率（*SR*）以及最大回撤率（*MDD*）。

对称平均绝对百分比误差（*sMAPE*）定义为：

$$sMAPE = \frac{100\%}{n} \sum_{i=1}^n \frac{|\hat{y}_i - y_i|}{(|\hat{y}_i| + |y_i|)/2} \quad (2.9)$$

方向准确率（*DA*）用于评估所提出的模型，其定义为：

$$DA = \frac{T}{100} \sum_{t=1}^T d_t \quad (2.10)$$

其中， d_t 应有如下公式计算：

$$d_t = \begin{cases} 1, (\hat{x}_t - \hat{x}_{t-1})(x_t - x_{t-1}) \geq 0 \\ 0, (\hat{x}_t - \hat{x}_{t-1})(x_t - x_{t-1}) < 0 \end{cases} \quad (2.11)$$

其中 \hat{x}_t 对应于由预测模型所产生的预测值。此外，本文还设计了一个简单的交易策略，即根据模型预测的时间序列结果计算交易信号，按照交易信号进行仓位调整操作。本文利用持仓情况反映交易策略已作出买入、卖出或维持不变（保持）的决定。持仓情况变化的计算公式为：

$$position_t = \begin{cases} position_{t-1}, & position_{t-1} = 0 \text{ and } \hat{x}_t \leq \hat{x}_{t-1} \\ 1, & position_{t-1} = 0 \text{ and } \hat{x}_t > \hat{x}_{t-1} \\ position_{t-1}, & position_{t-1} = 1 \text{ and } \hat{x}_t \geq \hat{x}_{t-1} \\ 0, & position_{t-1} = 1 \text{ and } \hat{x}_t < \hat{x}_{t-1} \end{cases} \quad (2.12)$$

持仓情况变化详情载于表 2.2。

表 2.2 交易策略

上一天持仓情况	看涨	看跌
0	买入，持仓情况设为满仓	保持
1	保持	卖出，持仓情况设为空仓

为了直观地评估模型的有效性，本文基于测试集数据的交易结果计算累计回报率（ CR ）、年化回报率（ AR ）、夏普比率（ SR ）和最大回撤率（ MDD ）。首先，定义每日资产 $Asset$ 。初始化 $Asset_0 = 1.0$ ，而 $Asset_t$ ， $t = 1, \dots, T$ 是基于上述计算的持仓变化结果来计算的。累计回报定义为：

$$CR = \prod_{t=1}^{T-1} (R_t + 1) \quad (2.13)$$

其中：

$$R_t = \frac{Asset_t}{Asset_{t-1}} - 1, t = 1, \dots, T - 1 \quad (2.14)$$

R_t 代表每日收益率。年化收益率 AR 可由下式计算：

$$AR = 252 \times \frac{CR-1}{T} \times 100\% \quad (2.15)$$

其中 252 代表的是每年有 252 个交易日。为了全面考虑收益和风险，本文还引入了夏普比率（ SR ）来评估本文所有提出和对比的模型，其计算公式为：

$$SR = \frac{R - R_f}{\sigma} \quad (2.16)$$

其中收益率 R 减去无风险利率 R_f （设定为 1.5%），然后除以收益率的标准差 σ 。在本

文的实验中，最大回撤率（ MDD ）被用来描述在回测过程中的最大损失以衡量风险性，其可计算为：

$$MDD = -\max\left(\frac{Asset_i - Asset_j}{Asset_i}\right) \times 100\%, i < j \text{ and } i, j \in [0, T - 1] \quad (2.17)$$

2.7 本章小结

本章阐述了相关理论基础，接着介绍了本文使用的数据集、数据封装方法以及多步预测方法。最后介绍了交易策略和丰富全面的评价指标。

第3章 基于注意力机制的深度模态分解编码解码模型 AED-DEMD

3.1 引言

在资产定价中，基于观察相关变量找到解释股票价格的行为的方法是常见的。例如，在技术分析中，重点是根据分析过去的价格高低和可预测模式的信息，产生超过市场平均回报的超额回报。然而，金融时间序列和临床心血管以及其他生物序列的混沌行为^[55]一样，通常与某种确定性混沌相关^[54]。Richman 和 Moorman^[55]建议将样本熵(SampEn)作为一个指标，它反映了在一个信号中发现的规律性和可预测性的程度。SampEn 所表达的是“当模式长度增加时，时间序列中相似模式保持相似的可能性”^[56]。换句话说，某个时间序列的 SampEn 值越高，那么它的可预测性就越低。

由于信噪比较低，对金融时间序列的预测的成功率比较低^[57]，并且上面阐述的那些方法都没有考虑以及讨论如何解决信号中的噪声问题。由于存储单元没有无限的容量来存储和处理在信号中发现的所有模式，因此复杂的时间依赖性问题是 RNNs 系列模型最重要的挑战之一。噪声是一个重要的扰动源，可能会误导模型，从而导致从序列中检测的主要模式存在大量误差。然而，在存在噪声信号的时间序列中，如果噪声与该序列共享基本谐波的一些情况下，很难滤除噪声信号^[57]。此外，在某些情况下，噪声不是由任何外部信号源引入的，而是金融时间序列本身就飘忽不定。因此，对信号和噪声之间的区分不仅不合适，而且也是不期望的。因为在这种情况下，如果简单区分信号和噪声，并从中把噪声剔除，可能会从信号中去除重要信息。因此，与其把问题当作纯序列和随机噪声来处理，不如把信号看成一系列稳定和不规则的模态的复合序列^[58]，但如何把它们区分开来是一个非常困难和重要的问题。

由 Huang 等人^[41]提出的经验模态分解 (EMD) 是一种基于数据本身时间尺度特征来分解信号的傅立叶变换^[59,60]。它被开发为一种迭代算法，该算法将原始时间序列分解为一组正交信号，称为本征模式分量 (IMFs)。EMD 非常适用于非平稳序列，因为其基于仅基于数据自身局部时间尺度表示来分解信号，而无需预设函数或使用任何谐波。然而，EMD 存在模态混叠的问题，即其中不同的 IMF 中可能存在相似频率的信号成分^[60]。Chacon H D 和 M. E. Torres 等人^[58,61]提出了一些方法来处理这个问题，但这些方法将引入随机噪声并增加时间序列的复杂性，而在金融时间序列预测任务中，本文期望的是提出原始序列中的噪声成分，并且降低序列复杂性，因此这些方法与本文的目的截然相反而被避免使用。此外，这些方法大大增加了计算量。

本章节旨在预测未来多步的股票价格，基于上述背景和深入的分析，提出了一个独特的以注意力机制为基础的编码解码深度模态分解模型（AED-DEMD）。首先，本章提出了深度模态分解模型（DEMD）来去除时间序列中的高频不相关成分。DEMD 适用于复杂的金融时间序列，因为它消除了原始序列中的高频噪声信号，降低了序列样本熵值，提高了改序列的可预测性。深的堆栈结构是由两种 EMD 层堆叠而形成。这些层通过残差链路连接，这有利于在保留低频趋势信息的同时对高频分量进行深度过滤。其次，本章设计了一个新颖的基于 LSTM 和多头注意机制的编码解码模型（AED）来预测该序列的趋势。该编码解码模型应用两个 LSTM 层来构成 AED 的编码器和解码器。本章设计它们来挖掘序列中的信息和规则，并输出预测结果。此外，受到 LSTM 的长记忆力的不足的情况的启发，长期依赖学习的改进的注意机制^[34]在 AED 中被应用为内部连接器，注意力层使用动态生成的权重来聚合时间特征和模式，以增强全局的注意力，提高学习效果。注意力机制允许网络直接关注过去的重要模式，并捕捉序列中的多步趋势信息。最后，本章展示了序列样本熵的结果和能量值。实验结果证实，DEMD 可以利用这样的深层次堆栈结构和算法有效地降低金融时间序列的复杂性和模态混叠的影响，同时提高了金融时间序列的能量保留率及其可预测性。此外，本章显式地展示了所提出的 DEMD 的序列处理结果，并将所提出的 AED-DEMD 模型与基于上证 50 ETF、沪深 300 指数以及道琼斯指数的其他模型进行了实验效果比较。各种结果表明，AED-DEMD 模型能更准确地预测多步股价方向准确率，并显著提高了投资回报率。

本章节其余部分的结构如下。在第 2 节中，详细描述了本章所提出的 AED-DEMD 模型架构。实验配置细节、实验结果的展示与分析在第 3 节呈现。第 4 节对本章进行了总结。

3.2 模型介绍

3.2.1 AED-DEMD 总体架构

AED-DEMD 旨在通过最小化均方误差（MSE）损失函数来实现多步股价预测。本章的任务在于对股票价格的短期多步预测（5 天）。AED-DEMD 的整体架构如图 3.1 所示。首先，本章提出的 DEMD 它是一种数据处理方法，可用于分解原始信号 $x(t)$ （包括每日收盘价、市场指数和成交量），并去除其中高频波动干扰信号，同时保留较低频率的趋势信息。经 DEMD 处理后的序列数据被送入编码器，得到中间隐藏状态向量。然后，

使用多头注意力层来聚合该隐藏状态时间序列向量，注意力机制通过使用动态生成的权重允许网络在整个序列片段时间维度上直接关注过去时刻的重要状态和模式。然后利用 Add & LayerNorm 层将隐藏状态向量与多头注意层的输出相加，再进行层归一化处理，此时得到的向量作为 AED-DEMD 模型的解码器的输入。最后，由解码器解码获得预测信号 $y(t)$ 。在 AED-DEMD 中，编码器和解码器都使用 LSTM 来构成。

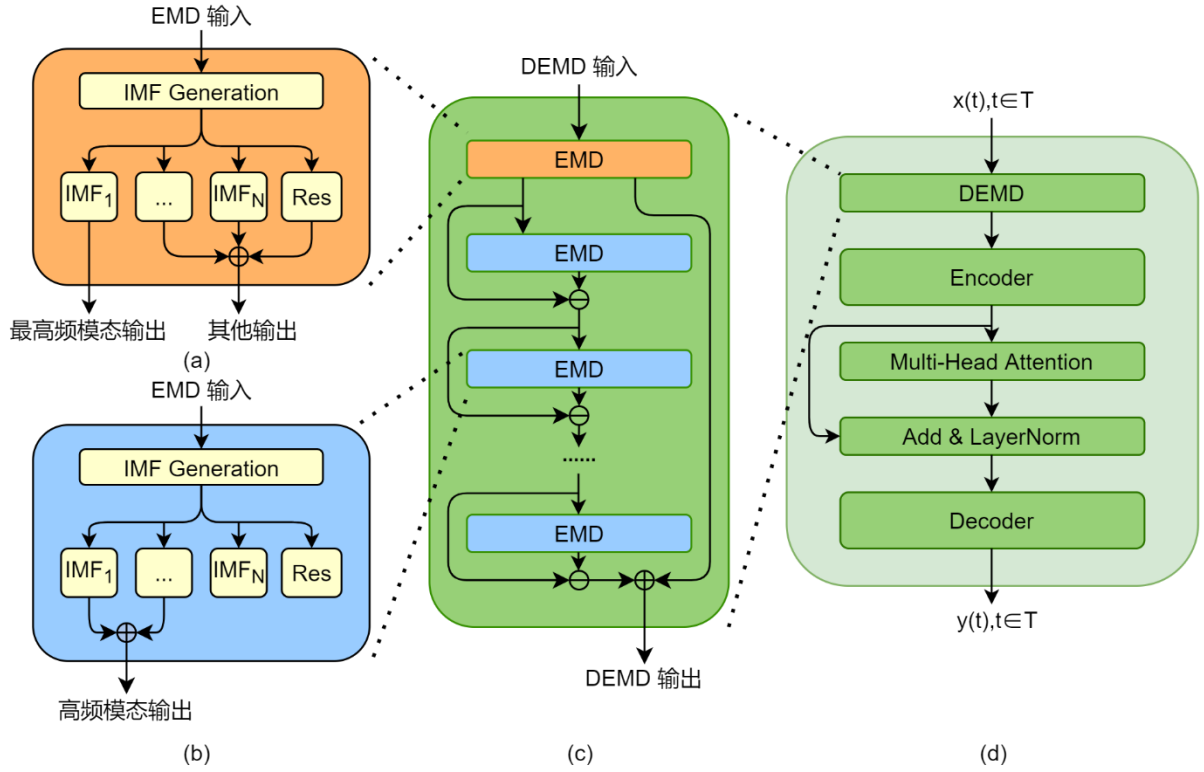


图 3.1 AED-DEMD 架构图

3.2.2 动态分解 EMD 层

如前所述，金融时间序列受到许多复杂因素影响，它们是非线性、非平稳的混沌信号，通常表现出不稳定的行为，这导致众多预测模型用于预测金融时间序列，如股票价格序列难以得到比较好的性能。为了处理这个问题，如何寻找、去除股票价格序列中的仅造成价格波动而不影响趋势的信号成分显得尤为关键。首先，通过改变输出设计了两个不同的 EMD 层来划分和处理由 IMF 迭代过程所获得的 IMF 分量。具体来说，如图 3.1 中的 (a) 和 (b) 所示，第一 EMD 层可以将原始序列信号分解为一组正交本征模态分量信号 (IMFs)，这些 IMFs 会按照频率从高到低进行排列。本章将这些 IMFs 信号被分成两个部分，第一部分是最高频率模态分量 IMF₁，第二部分是除了 IMF₁ 以外的所有剩余分量，通过加和的形式将这些剩余分量输出到 DEMD 模块中最后的缓存中等待输

出，与此同时，最高频率分量 IMF_1 被传递到下一层用于进一步处理。如 3.1 中的 (c) 所示，经第一层 EMD 输出的最高频分量 IMF_1 会被送到由第二类 EMD 组成的深度堆栈结构中进行深度分解和过滤，在这个结构中利用预设变量将最高频率成分 IMF_1 动态划分为高频波动成分和低频趋势信息，通过仅输出高频成分用以继续处理而保留趋势信息的方式完成序列数据处理任务。

3.2.3 深度经验模态分解 DEMD

本章提出的 DEMD 模型可用于分解原始金融时间序列 $x(t)$ ，去除其中的高频干扰信号，并重新组合较低频率的分量以保留其中趋势信息，为数据送入预测模型进行短期多步趋势预测任务做准备。如图 3.1 (c) 所示，本章提出的 DEMD 由上一小节提到的几个经过改造的两类 EMD 层组成，将这些 EMD 层堆叠起来，形成一个具有深度层次的整体结构，每个 EMD 层之间用残差的方式连接起来。其中，每个独立的 EMD 层可以将时间序列数据分解为一组 IMFs（每个分量用 C_i 标记），而在两类经改造的 EMD 层内部，数据得到了不同的操作。如前所述，原始金融时间序列数据由许多不同频率的分量组成，高频分量为波动干扰信号，低频分量为趋势信息。那么，因为高频率分量 C_1 是最复杂的分量并且其频率中最高，将通过第一层 EMD 层经过 IMF 迭代过程分解原始数据并输出，得到 C_1 分量、其他分量和参与信号数据。接着，将 C_1 输出给下一层继续处理，其他剩余的所有分量以及残余信号都被直接输出到缓存区等待处理。从第二层 EMD 层开始，通过与深度相关的参数，模型都会将上一层的输出结果动态地划分为高频波动分量和低频趋势信息，并将每一层的高频波动分量单独分离输出。然后，通过残差连接网络，可将这部分被输出的高频波动分量从本层输入的原始序列数据中过滤掉，从而保留其中的低频趋势信息。最后，将此时获得的残差结果输出给下一层继续做深度的处理。综上，通过这样设计的堆栈结构和残差连接，所提出的模型不但能在每一层中区分原始信号中高频波动成分和低频趋势信息，将波动成分从序列中过滤掉以降低序列复杂性。同时，也能通过这样多层次深度处理的方式，将混叠在相对低频成分中的模态一层层分离出来，从而得到更加准确的趋势信息。更详细地，本章基于以下假设和推导来设计此结构：

$$x(t) \xrightarrow{EMD} C_1(t) + C_2(t) + C_3(t) + \sum_{i=3}^n C_i(t) + res \quad (3.1)$$

首先，原始金融时间序列信号被分解为若干个 $C_i(t)$ 信号分量和残差成分，并且将这些

信号分量按照频率从高到低进行排列。然后,如前所述,模型单独提取出最高频率分量 $C_1(t)$, 并对其进行进一步深度的分解处理:

$$C_1(t) \xrightarrow{EMD} C_{11}(t) + C_{12}(t) + C_{13}(t) + \sum_{i=3}^n C_{1i}(t) + res_1 \quad (3.2)$$

由于 EMD 中存在的模态混叠问题,为降低其影响,去除了一些复杂的较高频成分,并让:

$$\begin{aligned} R_1(t) &= C_1(t) - \sum_{i=1}^k C_{1i}(t) \\ &= \sum_{i=k+1}^n C_{1i}(t) + res_1 \end{aligned} \quad (3.3)$$

之后,模型可以进行下一步,继续分解 $R_1(t)$, 并根据频率高低重新排列分解得到的信号分量:

$$\begin{aligned} R_1(t) &\xrightarrow{EMD} [C_{1(k+1)1}(t) + C_{1(k+1)2}(t) + \dots] + [C_{1(k+2)1}(t) + C_{1(k+2)2}(t) + \dots] + \dots \\ &= [C_{1(k+1)1}(t) + C_{1(k+2)1}(t) + \dots] + [C_{1(k+1)2}(t) + C_{1(k+2)2}(t) + \dots] + \dots \end{aligned} \quad (3.4)$$

同样地,模型重复上述步骤,继续消除一些高频成分,并得到:

$$\begin{aligned} R_2(t) &= R_1(t) - [C_{131}(t) + C_{141}(t) + \dots] \\ &= [C_{132}(t) + C_{141}(t) + \dots] + \dots \end{aligned} \quad (3.5)$$

请注意,为了在降低模态混叠效应的影响的同时,尽可能的保留更多的低频趋势信息,在 DEMD 模型中,随着深度的加深,过滤掉的高频模态分量个数会逐渐减少,因此在算法中,设置了一个控制变量 k 来实现这个操作。

本章设计的 DEMD 的堆栈结构可以多次重复上述步骤,最终获得 $R_n(t)$, 其中高频波动成分信号在迭代过程中被一层层过滤掉,而相对的低频趋势信息被保留。最后,得到的 $R_n(t)$ 与 DEMD 模型中的第一层 EMD 层输出时间序列数据 res 相加,得到最终的处理结果。上述整个过程的详细计算过程如算法 3.1 中展示。

算法 3.1 DEMD 算法

输入: 原始金融时间序列 $x(t)$

参数: 深度 D

输出: 处理后时间序列 $y(t)$

定义 $\tilde{x}(t) = x(t)$ 和控制变量 $d = 1$

while $d \leq D + 1$ **do**

if $d = 1$

 对 $\tilde{x}(t)$ 进行 IMF generation, 得到所有的本征模态函数 $C_{d,n}(t)$ 以及残余分量 $r_d(t)$

 令 $res = \sum_{k=2}^n C_{d,k}(t) + r_d(t), \tilde{x}(t) = C_{d,1}(t)$

else

 对 $\tilde{x}(t)$ 进行 IMF generation, 得到本层所有的本征模态分量 $C_{d,n}(t)$ 和 残余分量 $r_d(t)$

 令 $\tilde{x}(t) = \tilde{x}(t) - \sum_{k=1}^{D+2-d} C_{d,n}(t)$

$d = d + 1$

end while

 令 $y(t) = res + \tilde{x}(t)$

3.2.4 基于多头注意力机制的编码解码模型 AED

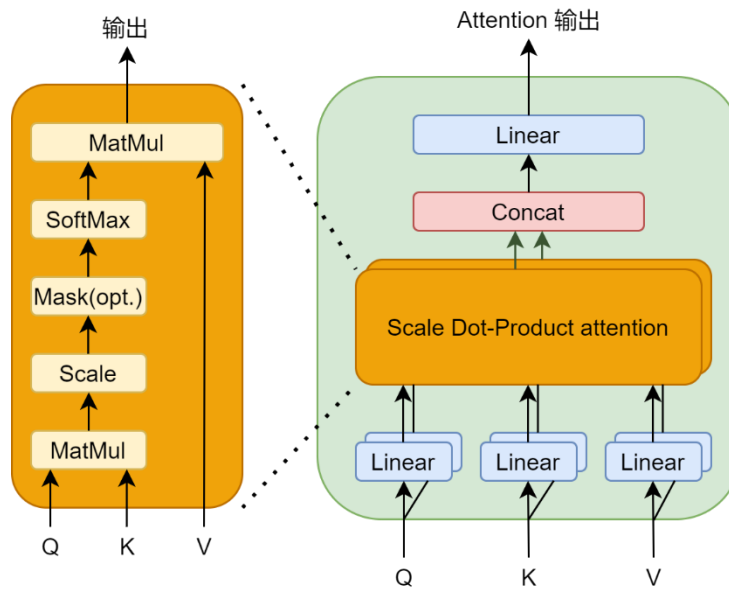


图 3.2 多头注意力层结构

在这项研究中，本章使用 LSTM 建立了一个用于金融时间序列预测的编码解

码模型 AED（如图 3.1（d）所示）。作为一种特殊类型的递归神经网络（RNNs），LSTM 被特别设计为避免序列数据的梯度爆炸和梯度消失问题。LSTM 由输入层、一个或多个隐藏层和输出层组成。隐藏层由存储单元组成，每个存储单元具有被称为 CEC^[62]的核心递归自连接线性单元，其具有输入门、忘记门和输出门以提供短期存储。

通常，短期股价趋势受许多因素的影响，并且与噪声或者异常事件高度相关，但是 LSTM 的输入门、遗忘门和输出门防止存储记忆单元中的内容受到异常(不相关的输入和输出)的干扰。因此，LSTM 适合于学习包含更长期模式的序列^[63]。但是在本章大量实验后发现，这种长期修正模式具有明显的时间滞后性，这对于在金融领域的应用来说是一大问题。因此，基于近年来注意力机制的发展，本章设计了基于注意力机制的编码解码模型。与其他方法不同的是，本章并不是将注意力层集成在编码层或者解码层以辅助完成多步预测任务，而是将注意力层作为中间转换层，使其能够处理编码层的完整输出结果。

细节上来看，本章引入了多头注意力机制^[39]，如图 3.2 所示，多头注意力是在单头注意力的基础上设计的，单头注意力可以描述为将查询（Q）和一组密钥-值（K-V）对映射到输出^[64]。多头注意力网络是几个单头注意力网络的级联。多头注意力的输出被计算为来自公式（3.6）的值的单头注意力加权求和的级联输出^[65]。在本章的 AED 模型中多头注意力网络使用其动态生成的权重聚合经编码器编码后的中间序列中的时间特征和模式，使其能够重新回顾整个编码结果，允许网络直接关注过去的重要模式并捕捉序列中的多步趋势，降低 LSTM 模型因其迭代结构而受到序列近期波动影响，最后产生错误判断或者预测仅是近期趋势的延伸。使用其动态生成的权重聚合经编码器编码后的中间序列中的时间特征和模式，使其能够重新回顾整个编码结果，允许网络直接关注过去的重要模式并捕捉序列中的多步趋势，降低 LSTM 模型因其迭代结构而受到序列近期波动影响，最后产生错误判断或者预测仅是近期趋势的延伸。

$$MultiHeadAttention(Q, K, V) = Concat(head_1, \dots, head_h)W^O \quad (3.6)$$

$$where head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \quad (3.7)$$

$$and Attention(QW_i^Q, KW_i^K, VW_i^V) = softmax\left(\frac{QW_i^Q(KW_i^K)^T}{\sqrt{d_k}}\right)VW_i^V \quad (3.8)$$

其中， $W_i^Q \in R^{d_{model} \times d_k}$ ， $W_i^K \in R^{d_{model} \times d_k}$ ， $W_i^V \in R^{d_{model} \times d_v}$ ， $W^O \in R^{hd_v \times d_{model}}$ 都是投影参数矩阵。

3.3 实验结果与分析

3.3.1 实验设置

为了考察算法性能，本章基于上证 50 ETF、沪深 300 指数以及道琼斯指数三个数据集进行大量实验。所有数据均已归一化，并且封装好。数据集具体信息以及封装的方法细节已在 2.3 节中展示。

然经过归一化处理以及窗口封装之后的所有数据片段馈送到 DEMD 中，以去除该系列的“无用”高频分量，接下来送入本章的 AED 预测模块进行短期趋势的预测。

图 3.1 中的 AED-DEMD 有五层，分别是 DEMD 层、编码层、多头注意力层、Add&LayerNorm 层、解码层，其中 LSTM 用于编码层和解码层。本章所使用的损失函数为均方误差 MSE，优化器为 Adam。实验中网络所使用的具体参数如表 3.1 所示。

表 3.1 AED-DEMD 参数配置

层	参数	值
DEMD	Depth D	3
	Input size	3
Encoder(LSTM)	Hidden size	60
	Head	2
Multi-Head Attention	Embedding dimension	60
	Input size	60
Decoder(LSMT)	Output size	3

本章的重点是股票价格的多步预测（5 天）的项目。与逐点预测处理不同，多步预测基于先前预测过程中预测的结果来预测未来几天的价格，因此本章选择迭代预测方法。具体来说，首先用所有真实数据初始化窗口，得到下一天的预测结果后，本章将窗口向后滑动一步，并将第一个预测点添加到窗口的最后一个点。然后，新窗口被馈送到模型中以预测下一点。在本章的实验中，上述过程被迭代 5 次，以获得未来 5 天的预测序列。

为了更好的验证模型的有效性，在本章实验中，不但将提出的模型的性能与传统的 LSTM、Mid-LSTM 和 N-BEATS 进行了比较。此外，本章还进行了消融实验。首先，本章将传统的 EMD 与 LSTM 相结合，并采取 EMD 分解信号后直接去除最

高频信号分量 IMF_1 的方式构建了 EMD-LSTM。其次，本章利用本章提出的 DEMD 取代 EMD，构建了 DEMD-LSTM。最后，用设计的 AED 来取代单一的 LSTM，并利用它们来搭建 AED-DEMD。

表 3.2 各神经网络训练超参数

超参数	LSTM	Mid-LSTM	EMD-LSTM	DEMD-LSTM	AED-DEMD
轮次	100	100	100	100	200
Batch 大小	64	64	64	64	64
学习率	0.0001	0.02	0.0001	0.0001	0.0001

由于 N-BEATS 与本章的体系结构有很大不同，本章使用的它的超参数可以在 N-BEATS^[24]原论文中找到。除 N-BEATS 外，用于实验的其它模型的超参数列于表 3.2。在实验中，本章计算了对称平均绝对百分比误差 ($sMAPE$) 和预测的方向准确度 (DA)。本章亦根据模型的预测结果进行自动交易，根据交易结果计算累计回报率 (CR)、年化回报率 (AR)、夏普比率 (SR) 以及最大回撤率 (MDD)，这些指标能够分别从时间序列预测准确性，收益性以及风险的角度对模型进行全面的评估。

3.3.2 实验结果

所有模型在三个数据集上的实验结果列于表 3.3-3.5 中。为了更加直观的看到各模型效果，图 3.3-3.6 展示的是基于所有模型在测试集上的预测结果进行自动交易所产生的每日累计回报收益情况。

表 3.3 上证 50 ETF 数据集实验结果

模型	$sMAPE$	DA	CR	AR	SR	MDD
LSTM	3.1614	49.03%	-8.73%	-6.19%	-0.4429	-25.5%
Mid-LSTM	9.1931	49.58%	4.84%	3.36%	0.2306	-15.51%
N-BEATS	2.1752	50.7%	-2.65%	-1.86%	-0.1104	-18.79%
EMD-LSTM	2.3806	49.86%	14.45%	9.91%	0.5258	-12.97%
DEMD-LSTM	2.3708	52.09%	16.39%	11.21%	0.5794	-14.27%
AED-DEMD	2.2098	53.48%	20.76%	14.12%	0.7452	-11.99%

表 3.3 展示的是所有对比模型和本章提出的模型在上证 50 ETF 数据集上的实验

结果汇总。从这些结果中，本章可以观察到，本章提出的 AED-DEMD 模型在 DA 、 CR 、 AR 、 SR 、 MDD 评估指标上都比其他模型有更好的表现。 DA 指标显示本章的 AED-DEMD 模型在判断趋势方向上的能力要优于其他模型，搭配自动交易策略，从而在测试周期上取得了更高的累计回报率和年化回报率。同时，从上表结果来看，基于 AED-DEMD 模型的最大回撤率较小，所以在投资过程中遭受的最大损失最小，风险最低。因此，平衡风险之后的收益情况，即夏普比率是最高的。此外，本章可以注意到，作为溶蚀实验的参照组， $sMAPE$ 显示 DEMD-LSTM 比 EMD-LSTM 有更高的预测精度，这说明 DEMD 在处理高频波动信号时比使用 EMD 分解再直接去除高频波动分量的方式更加细致，验证了本章的 DEMD 能够保留了更多的信息，从而取得更好的效果。同时从表中可以明显发现，AED-DEMD 的方向准确率高 DEMD-LSTM，而 DEMD-LSTM 的方向准确率又高于 EMD-LSTM，这说明 DEMD 在判断方向上的性能优于 EMD，而因为注意力机制的存在，使得在 LSTM 编码后，能够重新回顾整个时间序列片段，从而产生了对趋势更准确的判断。因为有更高的方向准确率，这让其获得了更高的累计回报率和年化回报率。

从表中也可以看到，在 $sMAPE$ 预测精度指标上 N-BEATS 取得的效果最好，AED-DEMD 因其损失了部分短期波动信息，导致在预测每日级别上的价格波动上表现稍差，单纯预测价格的绝对误差要大于 N-BEATS。

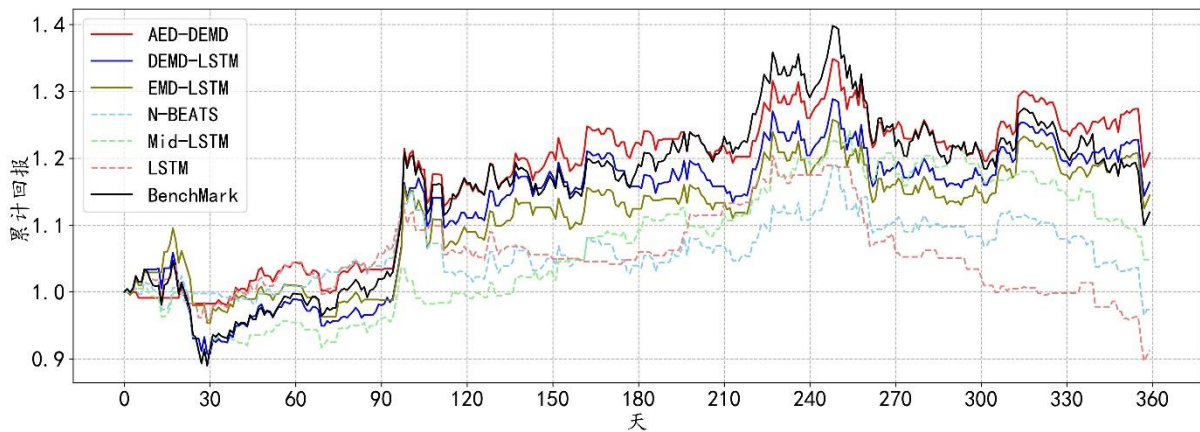


图 3.3 上证 50 ETF 数据集累计收益图

综上所述，在上证 50 ETF 数据集上的实验结果表明，本章提出的 DEMD 方法能明显提高金融时间序列短期趋势的预测方向准确率，并取得了最高的收益率。同时，所提出的 AED-DEMD 比 DEMD-LSTM 具有更好的性能，证明 AED 更善于处理趋势信息，因而具有比只有普通 LSTM 的模型具有更佳的效果。

各模型在上证 50 ETF 数据集上的累积回报情况如图 3.3 所示。为了更客观地反映市场情况和表现，本章还增加了上证 50 ETF 作为比较基准。在图 3.3 中，本章可以发现 DEMD-LSTM 的累积回报高于 EMD-LSTM 的累积回报，而 AED-DEMD 的累积回报最高。值得注意的是，DEMD-LSTM 模型在市场急剧下跌之前就已经预测出了纠正的趋势。因此，本章设定的自动交易机制在正确的时间进行出售操作，避免了巨大的损失。然后，本章在整个测试期间获得了超额回报。

表 3.4 沪深 300 指数数据集实验结果

模型	<i>sMAPE</i>	<i>DA</i>	<i>CR</i>	<i>AR</i>	<i>SR</i>	<i>MDD</i>
LSTM	3.1452	43.45%	-4.7%	-3.31%	-0.476	-8.77%
Mid-LSTM	2.9272	52.37%	26.98%	18.2%	0.9897	-16.08%
N-BEATS	2.6942	48.47%	10.4%	7.17%	0.5911	-7.71%
EMD-LSTM	2.3551	49.58%	19.44%	13.24%	1.0232	-7.19%
DEMD-LSTM	2.7372	52.65%	35.81%	23.9%	1.4123	-10.52%
AED-DEMD	2.3982	53.82%	37.06%	24.69%	1.2069	-14.34%

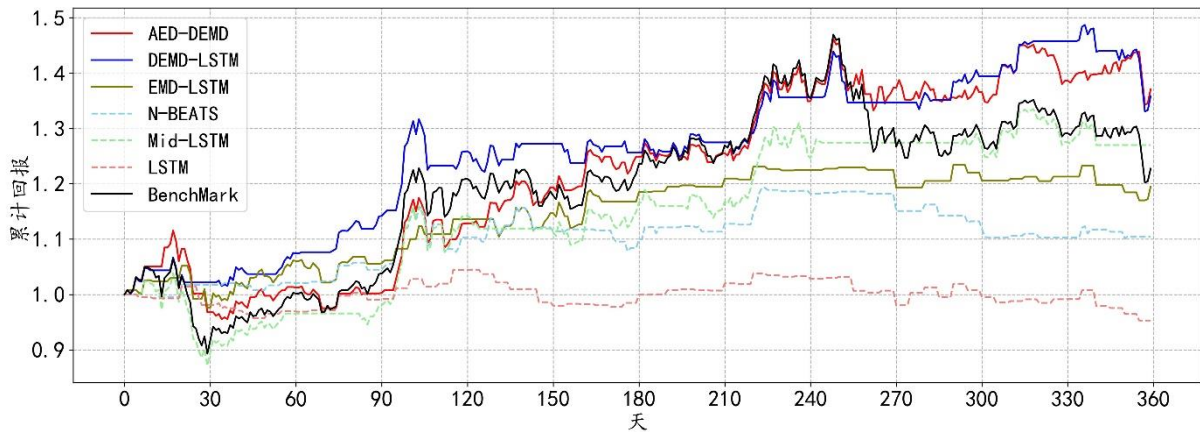


图 3.4 沪深 300 指数数据集累计收益图

表 3.4 展示的是所有模型在沪深 300 指数数据集上的实验结果汇总，图 3.4 展示的是各模型在沪深 300 指数上的累积回报情况。从中本章可以观察到，AED-DEMD 模型在关键指标方向准确率 *DA* 的表现是最好的，再次验证了本章针对高频模态分量的细致处理是有效的，同时在累计收益率和年化收益率上 AED-DEMD 也取得了最优的表现。值得注意的是，从沪深 300 数据集上的收益率指标测试结果来看，DEMD-LSTM 明显要高于 EMD-LSTM，这说明在大涨大跌行情来临时，对

于方向的判断和预测非常重要，本章提出的模型具有更好的方向判断能力，能获取更高的超额收益。

表 3.5 道琼斯指数数据集实验结果

模型	<i>sMAPE</i>	<i>DA</i>	<i>CR</i>	<i>AR</i>	<i>SR</i>	<i>MDD</i>
LSTM	2.3374	46.52%	0.28%	0.2%	-0.1356	-4.39%
Mid-LSTM	4.9807	49.03%	0.18%	0.12%	-0.0392	-9.58%
N-BEATS	4.0631	52.09%	10.17%	7.01%	0.62	-6.09%
EMD-LSTM	6.5273	48.47%	6.06%	4.2%	0.5133	-4.39%
DEMD-LSTM	4.5239	50.14%	9.76%	6.74%	0.7726	-4.39%
AED-DEMD	6.0305	51.81%	18.83%	12.83%	1.1214	-7.22%

表 3.5 展示的是所有模型在道琼斯指数数据集上的实验结果汇总，图 3.5 展示的是各模型在道琼斯指数上的累计回报情况。值得注意的是，与前两个数据集上的结果不同，从表 3.5 中本章可以观察到，AED-DEMD 模型在关键指标方向准确率 *DA* 的表现并没有达到最优，相比较 N-BEATS 的 *DA* 少了 0.2% 左右，但是在收益率指标比如累计收益率、年化收益率上却达到了最好的效果。经过分析这是因为收益率指标与预测方向准确率 *DA* 指标只是统计上的正相关，而不是严格正相关。因为每日价格涨跌率并不相同，即使预测的方向准确率相同，也会出现收益率存在一定波动的情况。但是，本章可以注意的是，AED-DEMD 在美股道琼斯指数上的结果既没有达到最优，也没有跑赢指数本身，这个结果与上证 50 ETF、沪深 300 指数上的结论不同，因此说明 AED-DEMD 的泛化性能不够优秀。

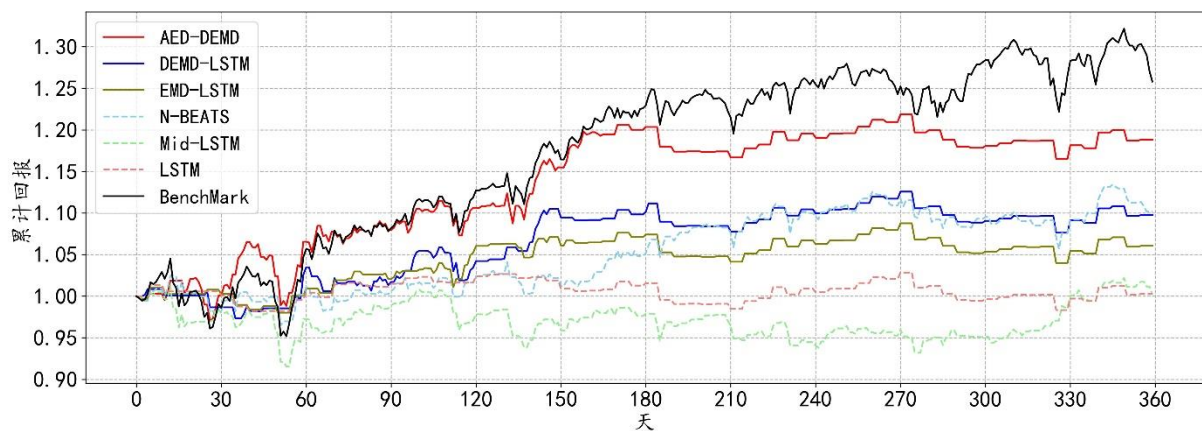


图 3.5 道琼斯指数数据集累计收益图

不过，从溶蚀实验结果来看，对比 EMD-LSTM 和 DEMD-LSTM，AED-DEMD

的关键性指标 DA 都有所提升，验证了本章针对高频模态分量的细致处理是有效的，使得 AED-DEMD 在判断行情走向上较为准确，同时在累计收益率、年化收益率上 AED-DEMD 也取得了最优的表现。

3.3.3 DEMD 结果

为了直观地显示 DEMD 的处理效果，在本小节，本章分别从上证 50 ETF，沪深 300 指数以及道琼斯指数三个数据集中抽取了各一个股票价格序列片段(360 天)，并用 DEMD 分别进行处理。

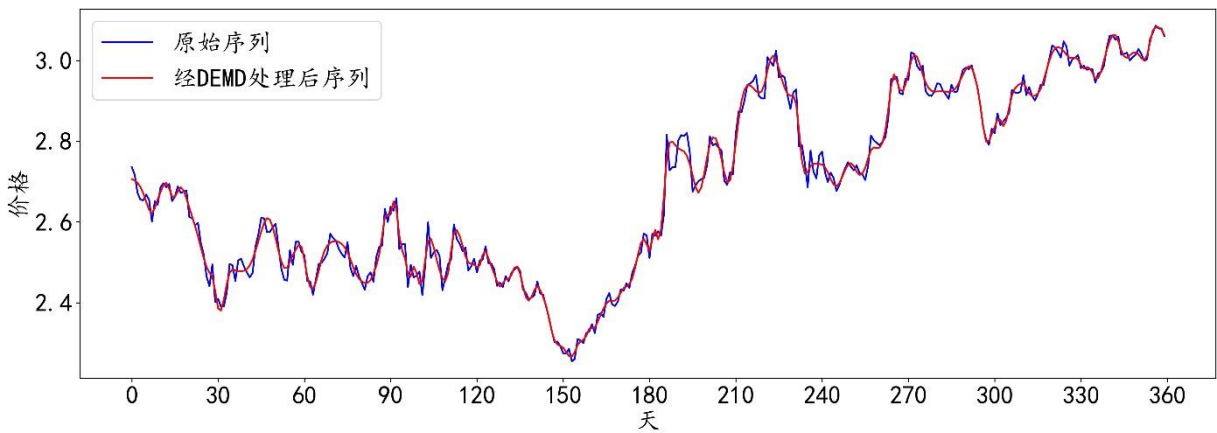


图 3.6 原始序列与 DEMD 处理后序列对比图（上证 50ETF 数据集）

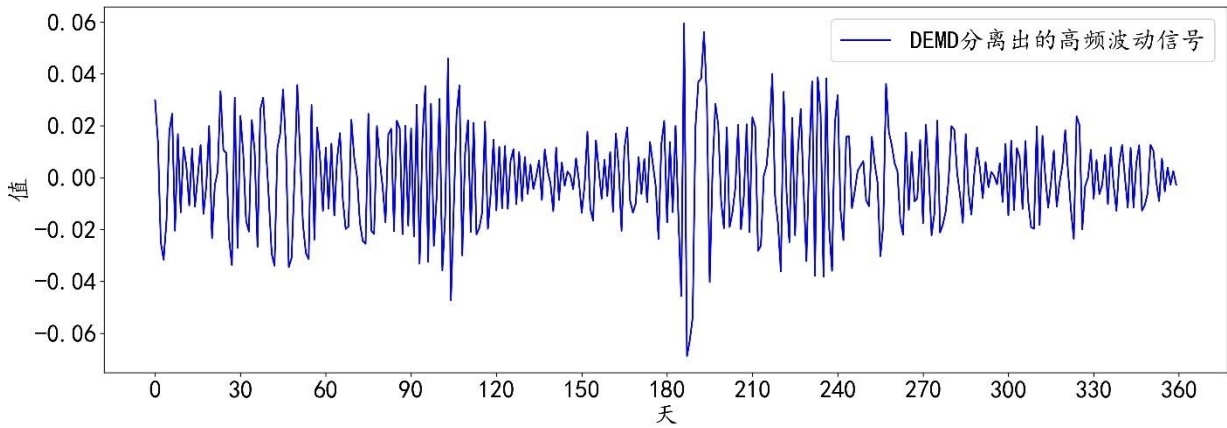


图 3.7 经 DEMD 过滤掉的高频不规则成分（上证 50ETF 数据集）

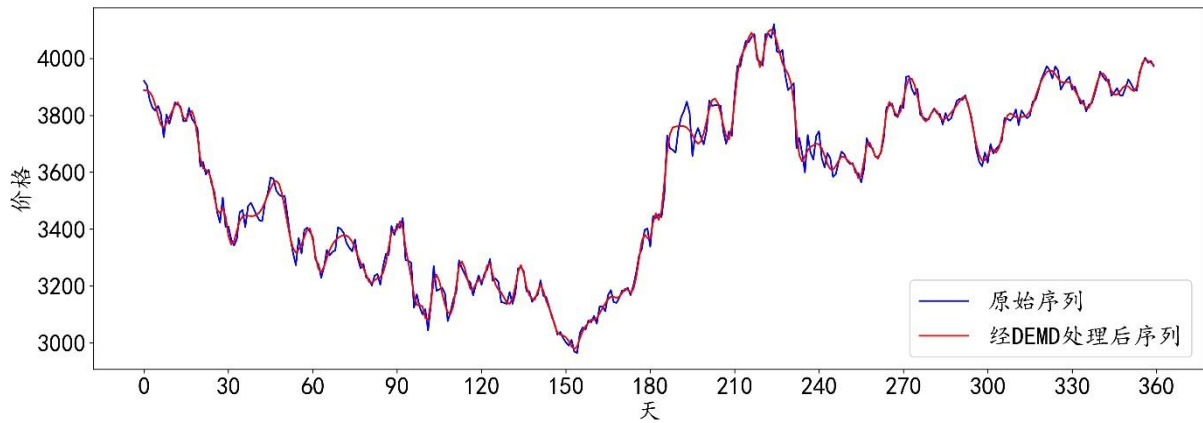


图 3.8 原始序列与 DEMD 处理后序列对比图（沪深 300 指数数据集）

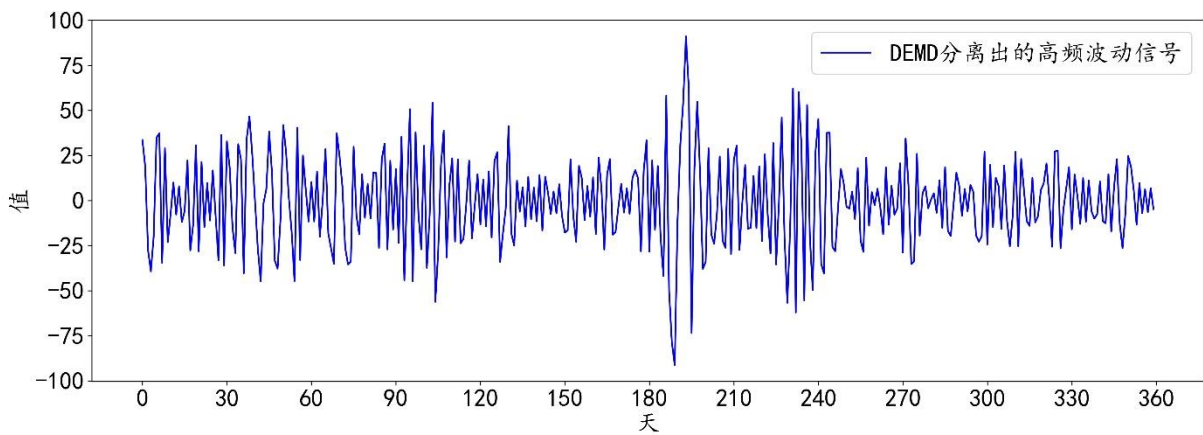


图 3.9 经 DEMD 过滤掉的高频不规则成分（沪深 300 指数数据集）

原始股票价格数据和由本章所提出的 DEMD 方法处理后的序列如图 3.6、3.8、3.10 所示，而从原始股票价格数据中过滤掉的高频波动模态数据如图 3.7、3.9、3.11 所示。

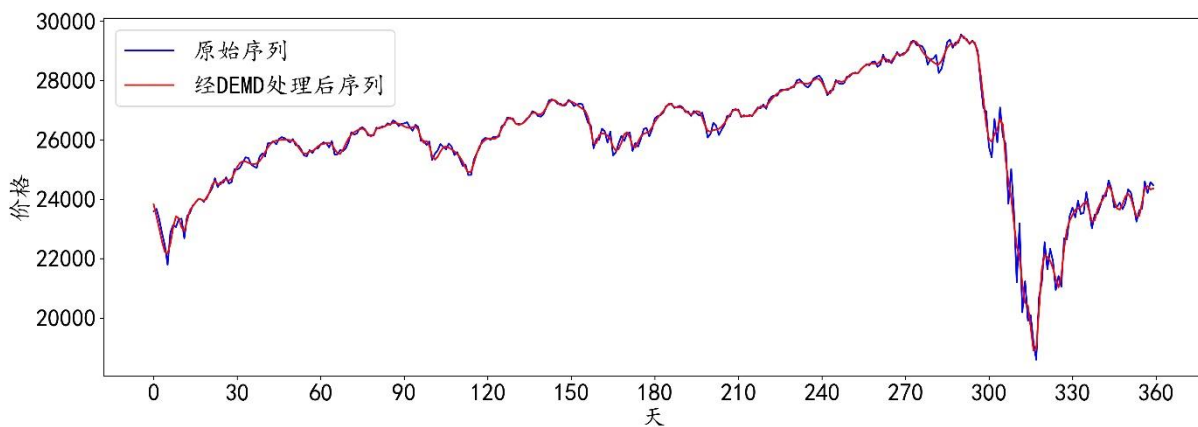


图 3.10 原始序列与 DEMD 处理后序列对比图（道琼斯指数数据集）

从图 3.6、3.8、3.10 中可以明显看出，DEMD 能够保证原始股票价格小级别短期

趋势信息不丢失。同时，从图 3.7、3.9、3.11 中可以明显观察到，使用 DEMD 去除的信号成分非常复杂，并且存在频率上的差异，这也直接说明了股票价格序列复杂度很高，其中高频模态分量中存在大量混叠成分。因此，本章的 DEMD 方法可以过滤掉其中不规则成分，从而降低了序列的复杂度，提高了股票价格序列数据的可预测性。

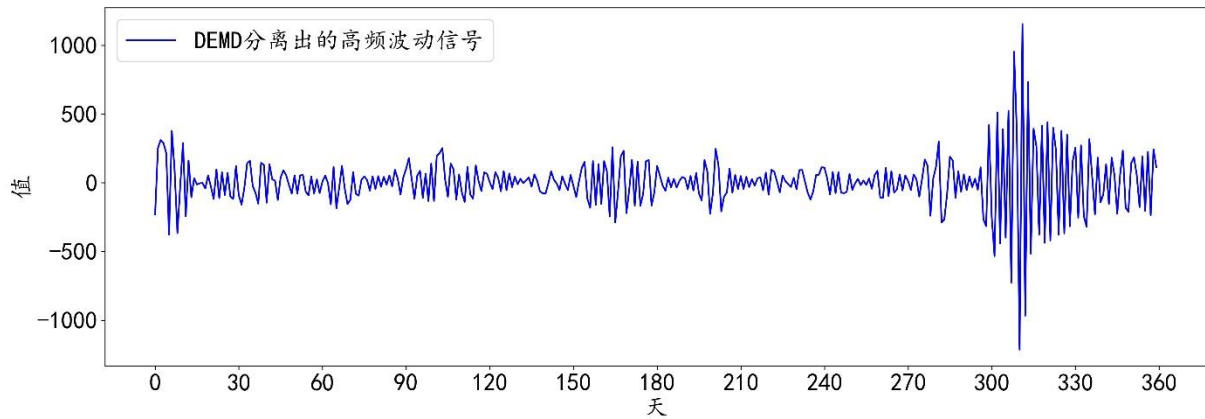


图 3.11 经 DEMD 过滤掉的高频不规则成分（道琼斯指数数据集）

3.3.4 样本熵与能量值结果展示与分析

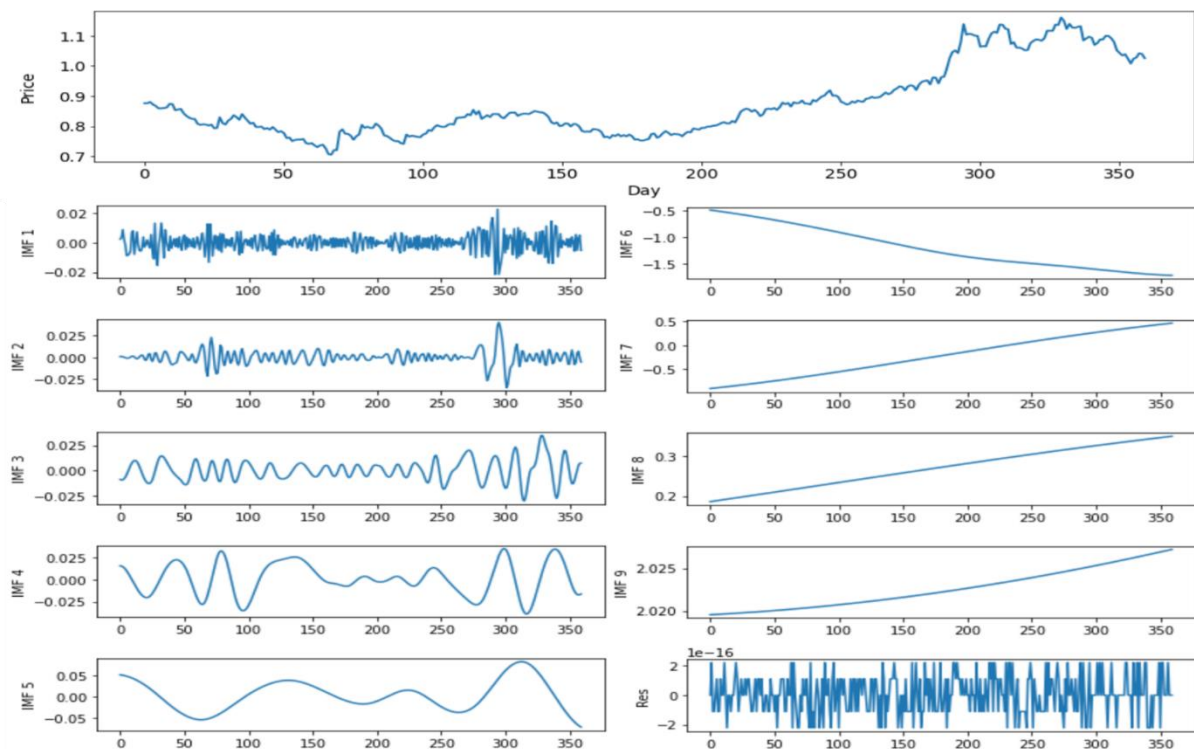


图 3.12 原始序列及其 IMFs 分量

本节使用上证 50 ETF、沪深 300 指数以及道琼斯指数日内价格序列数据进行序列的样本熵与能量值分析。首先，在图 3.12 中，本章展示了 EMD 对上证 50 ETF

价格数据序列的分解结果，最上方第一个子图展示的是原始数据(360 天)，通过 IMF 迭代分解得到的各个 IMF 按照频率从高到低依次展示在下方子图中，其中最后一个子图为残余分量 res 。

表 3.6 序列能量值对比表(上证 50ETF)

——	原始序列		EMD		DEMD	
	能量值	能量比例	能量值	能量比例	能量值	能量比例
信号成分						
x (t)	5.086843	100.0000%	5.086385	100.0000%	5.086507	100.0000%
C1	0.000459	0.0090%	0.000539	0.0106%	0.000539	0.0106%
C2	0.000539	0.0106%	0.001162	0.0229%	0.001162	0.0228%
C3	0.001162	0.0228%	0.003104	0.0610%	0.003104	0.0610%
C4	0.003104	0.0610%	0.006855	0.1348%	0.006854	0.1348%
C5	0.006855	0.1348%	0.325159	6.3927%	0.325203	6.3934%
C6	0.325159	6.3922%	0.087158	1.7136%	0.087161	1.7136%
C7	0.087158	1.7134%	0.107465	2.1128%	0.107473	2.1129%
C8	0.107465	2.1126%	4.554942	89.5517%	4.555011	89.5509%
C9	4.554942	89.5436%	——	——	——	——

显然，从图 3.12 中可以看出， IMF_1 具有最高的频率，其中信号成分最复杂，如果不对其进行处理，这将使得序列中存在大量的高频波动信号，干扰模型完成短期趋势预测。同时，随着分解过程的迭代， IMF_n 的频率越来越低，趋势信息也越来越明显。那么，去除序列中的干扰信号，并且尽可能保留有效的趋势信息是非常重要的。因此，本章对两种方法进行比较。一种是直接去除 IMF_1 （用 EMD 来表示），另一种是使用本章的 DEMD 进行逐层过滤（用 DEMD 来表示）。正如吴和黄^[66,67]所建议的，本章使用能量来表示序列中包含的能量信息。由于 IMFs 几乎是正交的，所以数据的总能量可以是：

$$\sum_{t=1}^T x(t)^2 = \sum_t |F_t|^2 = \sum_{i=1}^N E_n \quad (3.9)$$

$$E_n = \frac{1}{T} \sum_{t=1}^T |C_n(t)|^2 \quad (3.10)$$

表 3.7 序列能量值对比表(沪深 300 指数)

——	原始序列		EMD		DEMD	
信号成分	能量值	能量比例	能量值	能量比例	能量值	能量比例
$x(t)$	9906958.422	100.0000%	9906024.549	100.0000%	9906070.200	100.0000%
C_1	888.221593	0.0090%	1107.479184	0.0112%	1107.478988	0.0112%
C_2	1107.478988	0.0112%	2605.972375	0.0263%	2606.942937	0.0263%
C_3	2606.942937	0.0263%	8226.61563	0.0830%	8224.067736	0.0830%
C_4	8224.067736	0.0830%	16497.05302	0.1665%	16520.93496	0.1668%
C_5	16520.93496	0.1668%	233199.5186	2.3541%	233348.5748	2.3556%
C_6	233348.5748	2.3554%	247022.6558	2.4937%	246657.4601	2.4900%
C_7	246657.4601	2.4897%	123809.7796	1.2498%	123847.9102	1.2502%
C_8	123847.9102	1.2501%	9273555.475	93.6153%	9273756.831	93.6169%
C_9	9273756.831	93.6085%	——	——	——	——

本章分别对上证 50 ETF、沪深 300 指数以及道琼斯指数每日价格序列数据进行 EMD 和 DEMD 处理，然后对得到的每个分量（标记为 C_i ）进行能量值的计算。三个序列数据的 IMFs 的能量值计算结果如表 3.6、3.7、3.8 所示。首列表示原始序列的能量值结果，EMD 列表示的是对价格数据进行 EMD 分解后，去除直接最高频 IMF_1 之后的能量值结果，而 DEMD 列表示对价格数据进行 DEMD 算法处理之后的能量值结果。注意，因为去除了最高频信号，所以 EMD 列和 DEMD 列模态数量会少一个，在表中用“——”代替。

从表 3.6-3.8 中可以看出，大体上来说，频率越高的 IMF 中所含的能量值越小，其中 IMF_1 中的能量值最小。而横向对比 $x(t)$ 能量值来看，经过 DEMD 处理的序列比 EMD 能量值稍高，这也就说明 DEMD 层层过滤的方法比经过 EMD 分解并直接去除 IMF_1 的方法保留了更多的序列能量，即更多的序列信息。此外，本章计算了相对能量保持率 ($RERR$)其计算公式为:

$$RERR = 1 - \frac{E_{source} - E_{DEMD}}{E_{source} - E_{EMD}} \quad (3.11)$$

三个数据集上的 $RERR$ 的结果展示在表 3.9 中。可以看到在三个数据集上， $RERR$ 都是正数，说明 DEMD 方式相对 EMD 方式都保留了更多的序列能量。其中在道琼斯

指数价格数据上的 $RERR$ 最高，约为 76.77%。

表 3.8 序列能量值对比表（道琼斯指数）

—— 信号成分	原始序列		EMD		DEMD	
	能量值	能量比例	能量值	能量比例	能量值	能量比例
X(t)	366979105.9	100.0000%	366968076	100.0000%	366976543	100.0000%
C1	11029.95796	0.0030%	11176.0199	0.0030%	11199.17009	0.0031%
C2	11176.0199	0.0030%	26701.21275	0.0073%	26667.7531	0.0073%
C3	26701.21275	0.0073%	224534.9805	0.0612%	224320.7399	0.0611%
C4	224534.9805	0.0612%	398871.4653	0.1087%	398285.5559	0.1085%
C5	398871.4653	0.1087%	274055.156	0.0747%	273850.4432	0.0746%
C6	274055.156	0.0747%	422196.0347	0.1150%	422933.1927	0.1152%
C7	422196.0347	0.1150%	365610541.1	99.6301%	365619286.1	99.6302%
C8	365610541.1	99.6271%	——	——	——	——

表 3.9 三个数据集上的相对能量保留率结果

——	SSE50	CSI300	DJI
$RERR$	26.6376%	4.9251%	76.7703%

正如本章在第一部分中所提到的，样本熵（SampEn）用于衡量时间序列数据的复杂性和可预测性。因此本章对三个数据集中的每日价格序列分别进行了 SampEn 的计算。SampEn 计算结果如表 3.10、3.11 和 3.12 所示。与前面能量值计算实验中相似，Source 代表原始序列，EMD 代表经过 EMD 分解后直接去除 IMF_1 成分的序列，DEMD 代表经过 DEMD 处理后的序列。SampEn 行表示对应的数据处理后时间序列的样本熵值，SampEn prop.表示当前时间序列的样本熵与原始时间序列总样本熵的比率，而 Reduction prop.表示当前时间序列的样本熵相对于原始序列的减少比率。

表 3.10 样本熵结果(上证 50ETF)

——	Source	EMD	DEMD
SampEn	0.073716	0.058339	0.058341
SampEn prop.	100.0000%	79.1402%	79.1429%
Reduction prop.	——	20.8598%	20.8571%

上证 50 ETF 价格序列数据的样本熵计算结果如表 3.10 所示, 原始序列的 SampEn 为 0.073716, EMD 列的 SampEn 为 0.058339, 从表中可以注意到样本熵值降低了约 20%, 这显示了虽然 IMF₁ 中的能量值最低, 即所含信息量最低, 但是其样本熵很高, 即复杂度很高。通过直接去除 IMF₁ 的方法显著地减少了序列的 SampEn, 因此降低了复杂度。同时, 可以注意到 DEMD 方法处理后的序列样本熵与直接去除最高频分量 IMF₁ 后的序列非常接近, 因此说明本章的 DEMD 方法在处理原始价格序列时, 几乎没有增加序列的样本熵值, 但是提高了相对能量保留率。

表 3.11 样本熵结果(沪深 300 指数)

——	Source	EMD	DEMD
SampEn	3.349904	1.845185	1.762602
SampEn prop.	100.0000%	55.0822%	52.6165%
Reduction prop.	——	44.8598%	47.3835%

沪深 300 指数价格序列数据的样本熵计算结果如表 3.11 所示。从表中可以注意到直接去除 IMF₁ 的方式 (EMD 列) 样本熵值降低了约 45%, 而用 DEMD 方法处理后的序列样本熵值相对原始序列降低了超过 47%。这显示了在沪深 300 数据集上, DEMD 相对 EMD 再直接去除 IMF₁ 的方式降低了更多的样本熵, 同时提高了能量保留率, 再次印证了本章上面的结论。

表 3.12 样本熵结果(道琼斯指数)

——	Source	EMD	DEMD
SampEn	2.3892	1.794138	1.813266
SampEn prop.	100.0000%	75.0937%	75.8943%
Reduction prop.	——	24.9063%	24.1057%

道琼斯指数价格序列数据的样本熵计算结果如表 3.12 所示。表中结果与上证 50 ETF 数据集上的结果类似。通过 EMD 处理，并直接去除 IMF_1 后的序列样本熵约为原始价格序列的 75%，样本熵降低比率约为 24.90%，这个降低比率比本章的 DEMD 稍高，因为在道琼斯数据集上，DEMD 方式的能量保留率比直接去除 IMF_1 的方式高出了约 76.77%，因此包含有更多的能量，因为样本熵值稍高也属正常情况。

综合上面在三个数据集合上面针对能量值、能量相对保留率以及样本熵值的实验结果说明，整体来看频率越高的 IMF 信号成分能量值越低，其中最高频 IMF_1 能量值最低，但是样本熵值很高，它的存在增加了序列的复杂性，但仅提供很少的信息量。直接去除 IMF_1 能够显著降低序列的样本熵，而通过本章的 DEMD 方式相比前者能够显著提高序列中能量保留率，同时样本熵值不会有显著提高甚至有所降低。所以，DEMD 可以用来降低金融时间序列的样本熵值和复杂度，同时充分保留了更多的信息，增加了序列的可预测性。

3.4 本章小结

金融时间序列是非平稳、高波动的序列，纯粹的深度学习模型难以直接处理这类数据，导致预测结果不准确。如何对这类高波动的时间序列进行降噪或者过滤是一个亟待解决的问题。在本章中，本章提出了一个 AED-DEMD 混合模型来完成对金融时间序列短期趋势预测的任务。首先，提出了一种新颖的数据处理方法 DEMD，本章为它设计了一个独特的深层次的堆栈结构，利用这种结构它可以分解原始序列，并对其中最高频模态分量序列进行逐层分解，并且本章通过残差连接的方式将每一层串联起来，这样在分解的同时，可以逐层过滤掉上一层的输出序列中较高频率模态成分，从而达到深度降噪的目的。然后，本章设计了一个基于注意力机制的编解码预测模型，称为 AED，以捕获趋势序列中的大量长期依赖信息，这可以帮助本章处理趋势预测任务。最后，本章将 DEMD 与 AED 结合形成了 AED-DEMD。

本章在上证 50ETF、沪深 300 指数以及道琼斯指数三个标的上进行了大量的实验。实验结果表明，所提出的 AED-DEMD 在预测股票价格序列的方向准确率方面具有最优秀的性能。自动交易实验结果表明，基于 AED-DEMD 的预测价格可以显著提高投资回报，获得更高的收益率。同时，通过对特别设置的 EMD-LSTM 和 DEMD-LSTM 的对比实验来看，DEMD 的效果要优于 EMD，并且由于注意力机制的存在，使得 AED 能在整个时间片段上重新回顾编码器的预测结果，利用动态可学

习的参数进行再调整，改善了单独 LSTM 迭代结构的局限性。此外，通过对序列能量值和样本熵值实验结果与分析可以看到，DEMD 提高了能量保留率，同时经过 DEMD 处理后的序列样本熵值有明显减少，即 DEMD 能在几乎不增加甚至减少原始金融时间序列复杂度的基础上，保留更多的趋势信息。

第 4 章 基于注意力机制的自适应最优表现选择模型

4.1 引言

金融时间序列是非线性、非平稳、混沌的信号。它们可能受到许多因素的影响。如果金融时间序列中的噪声和异常情况不能被有效地检测到,那么多步预测的风险将会非常高,所以对未来的股票价格多步的预测很有挑战性。

在上一章节中,本章提出了 AED-DEMD 模型用于股票价格的多步短期趋势预测,并取得了相对于其他对比模型效果上的优化。但是,通过深入分析可以发现, AED-DEMD 模型的性能或者其可优化空间存在明显的上限以及局限性。从第三章 AED-DEMD 架构图以及 DEMD 算法细节中可以看到,在 DEMD 的第一层 EMD 进行分解时,上一章通过人工选择的方式手动挑选出其最高频模态分量 IMF_1 , 并对其进行一系列的深度处理,这是因为从能量值和样本熵计算结果来看, IMF_1 的能量值最低,但是其样本熵值却很高。这种人工选择的方式有其优点,但是也决定了 AED-DEMD 模型的局限性,因为该模型仅仅对 IMF_1 做进一步处理,而 IMF_2 - IMF_n 等模态分量都没有被处理到,但是从上一章节样本熵值分析来看,这些分量的样本熵值占比也很大,并且其中也可能存在大量混叠的模态。

再者,从实验结果来看, AED-DEMD 模型在上证 50 ETF、沪深 300 指数以及道琼斯指数三个不同的数据集上的表现差异性较大,同时可以注意到的是其在道琼斯指数数据集上的效果没有达到最优,其性能有待提升。综合对比来看,性能的差异性有可能来源于这种依赖分析和经验的人工选择。一方面,原始时间序列分离出来的所有模态分量中只有最高频分量经过模型处理,这也就导致了无论如何精细化设计模型结构,无论对最高频分量进行如何细致的优化处理,该模型上限也非常明显,其优化空间有限。另一方面,仍然存在大量的模态分量有待处理,其中包含大量的能量值以及样本熵值。因此,剩下所有的模态分量信号也应该被进一步筛选和处理。但如何确定这些模态分量哪些包含了未来趋势信息,而哪些仅仅是嘈杂的波动信号,以及如何筛选和利用这些信息是一个困难的问题!

基于上述存在的问题,本章节提出了一个独特的管道结构,它能够自适应地完成金融时间序列地分解与合成。并且基于上一章节的 AED 模型和这种管道结构,本章提出了新的 EMD-AED-OPS 模型。本章节其余部分的结构如下:在第 2 节中,本章详细描述了所提出的 EMD-AED-OPS 模型架构。实验配置细节、实验结果的展示与分析在第 3 节

呈现。第 4 节对本章进行了总结。

4.2 模型介绍

4.2.1 EMD-AED-OPS 模型总体架构

本章提出 EMD-AED-OPS 旨在进行多步短期（5 天）股价趋势预测。EMD-AED-OPS 的结构如图 4.1 所示。整个模型看起来像一个管道，其中有三部分组成，分别是经验模态分解模块 EMD、基于注意力机制的编解码模型 AED 和最优表现选择模型 OPS。首先，金融时间序列数据（每日收盘价、大盘指数以及交易量）被送入 EMD 中进行处理，通过 IMF 迭代结构计算，原始序列数据将被分解为若干个本征模态分量 IMFs，紧接着这些 IMFs 被送入 AED 中进行预测，AED 将输出与 IMFs 维度相同的若干个预测结果，这些预测结果将被送入 OPS 中，OPS 通过其遍历累加过程来完成时间序列的合成过程，OPS 中内置了自动交易模块，每个合成后的序列都将在自动交易模块中完成交易回测，并且其中预设的相关参数会将表现最好的模态组合记录下来，作为最终的最优表现结果输出。

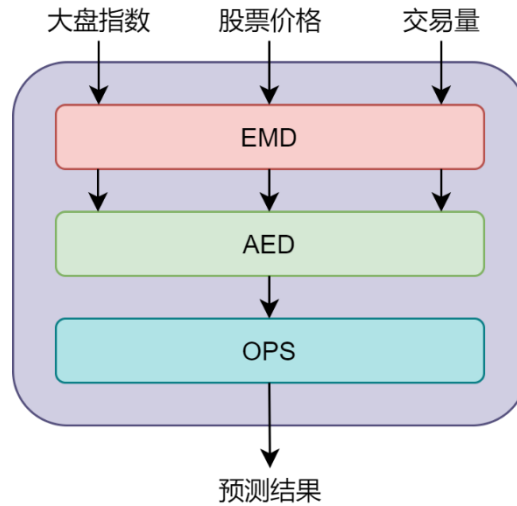


图 4.1 EMD-AED-OPS 结构图

4.2.2 EMD-OPS 自适应管道模型

如上所述，金融时间序列是一个复杂的非平稳序列，其中价格走势会受到很多因素影响。但从本质来说，它是一个多方买卖集合竞价的复合结果。以往的方法包括本文第三章中的 AED-DEMD 都将其作为一个单一序列进行处理，预测模型的输入和输

出都是单一序列，这一点与股票价格序列的复杂性相违背。在本章，本章设计了一个全新的管道模型，不像 DEMD 是一个整体性模块，本章将序列的分解过程与序列的合成过程分离为两个独立的模块，本章称它们为管道的两个端口。在本章的 EMD-OPS 管道中，EMD 是管道的分解端，OPS 为管道的合成端，它们分别完成原始时间序列的分解与合成。

首先，金融时间序列被送入到 EMD-OPS 管道的分解端，得到了若干个本征模态分量 $IMFs$ ，如上一章节所讨论的，高频的 IMF 本章称之为序列中的高频波动分量，而相对低频的 IMF 被称为低频趋势信息。接着，本章可以通过在管道中内置预测模型，分别完成对高频波动分量以及低频趋势的分离式预测，在得到的若干个预测结果里面，既包括波动的成分，也包括趋势性的成分。这些预测结果将被送到管道的合成端完成信号的合成，以形成最终的预测结果。

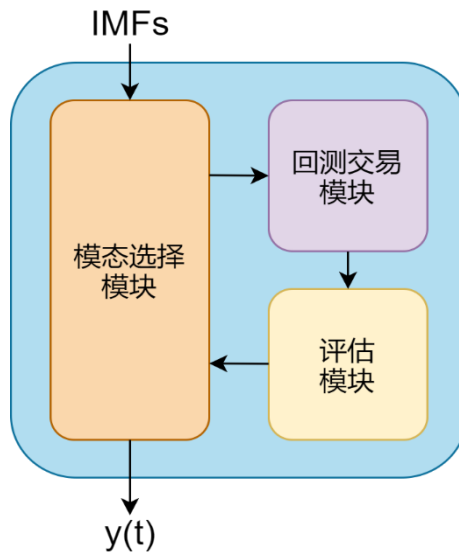


图 4.2 OPS 结构图

信号合成是另一个关键的问题。这里，OPS 是为了解决信号合成问题而设计的，它包含三个模块分别是模态选择模块（Mode Selection）、回测交易模块（Trading）和评估模块（Evaluation），其结构如图 4.2 所示。OPS 最大的亮点在于与 DEMD 中手工选择不同，它是一个可学习的自适应模型，它可以根据当前行情通过学习的方式自主地选择最优的模态组合。如上所述，可以通过 AED 模型获得若干个预测的 $IMFs$ 序列，在 OPS 模型中，内置了一个迭代累加计算器，它从频率最低的模态开始，按照频率从低到高的顺序依次累加，以得到一个合成后的序列，并且这些信号将在 OPS 模型中组合以获得复合序列。

算法 4.1 OPS 算法

输入: 模型预测得到的 $IMFs$

参数: $IMFs$ 中最大的组成部分数目 $maxshape$, 当前状态 $status$, 被选定的组成部分的个数 K 。

输出: 预测结果 $y(t)$

Let $y(t) = 0, x(t) = sum(IMFs)$

对 $x(t)$ 进行 IMF 迭代, 获得频率规整且从高到低分布的 $IMFs$

if $status = "train"$ **then**

Let $k = 1, maxincome = 0$

while $k < maxshape$ **do**

Let $i = 0$

while $i \leq k - 1$ **do**

Let $y(t) = y(t) + IMFs(maxshape - 1 - i)$

end while

对预测结果 $y(t)$ 进行自动交易, 并且计算收益率指标, 记录为 $income$

if $income > maxincome$ **then**

Let $maxincome = income$

记录当前选定的组成部分的个数 $K = k$

end if

Let $k = k + 1$

end while

else if $status = "test"$ **then**

for $k = 0$ to $K - 1$ **do**

Let $y(t) = y(t) + IMFs(maxshape - 1 - k)$

end for

end if

return: $y(t)$

当 OPS 在训练模式时, 模态选择模块输出训练集上的预测复合价格序列给回测交易模块, 回测交易模块会根据该预测的复合价格序列进行交易并且得到相关评价指标 (本章使用的是收益率), 评估器会根据收益率情况来进行评估, 通过参数 K 来记录最优模态组合, 并将它输出到模态选择模块中保存。当 OPS 在测试模式时, 在训练集上

学习到的参数 K 将直接被迁移到测试集上，从而使得累加器完成模态分量信号合成的工作，最终将得到的复合结果，即预测的价格趋势输出。OPS 的所有计算细节可以在算法 4.1 中找到。

4.2.3 AED 模块

AED 是本文第三章中设计的基于注意力机制的编码解码模型。AED 是 EMD-AED-OPS 的预测器，它需要分别对所有的 $IMFs$ 进行分别预测，本章设计这个模型的初衷是希望用波动来预测波动，用趋势来预测趋势。这里的数据流与第三章类似，如图 3.1 所示，这些 $IMFs$ （标记为 $IMFx$ ）被馈送到 AED 的编码器中，以获得中间隐藏状态向量。然后，多头注意力层重新回顾整个时间片段，通过使用动态生成的权重来聚合多步过去时间的隐藏状态向量，以加强对关键模式的学习。最后，通过解码器对所得向量进行解码处理，得到预测的 $IMFs$ 信号（标记为 $IMFy$ ）。在这里，本章的编码器和解码器都使用 LSTM。

4.3 实验结果与分析

4.3.1 实验设置

表 4.1 EMD-AED-OPS 中 AED 各层参数设置

层	参数	值
Encoder(LSTM)	Input size	3
	Hidden size	54
Multi-Head Attention	Head	2
	Embedding dimension	54
Decoder(LSMT)	Input size	54
	Output size	3

在本部分中，本章将基于上证 50 ETF（SSE 50）、沪深 300 指数（CSI 300）以及道琼斯指数（DJI）等数据集来验证本章的 EMD-AED-OPS 模型。

在 EMD-AED-OPS 中，只有 AED 需要进行训练。图一中的 AED 有 4 层，分别是编码层（LSTM），多头注意力层，Add&LayerNorm 层，以及解码层（LSTM）。本章所

使用的损失函数为均方误差 MSE，优化器为 Adam。实验中神经网络所使用的具体参数如表 4.1 所示。

为了更好的验证模型的有效性，在实验中，本章不但将本章的模型的性能与传统的 LSTM、Mid-LSTM 和 N-BEATS 进行了比较，本章还进行了消融实验。首先，使用 LSTM 和本章设计的 EMD-OPS 管道结合构建了 EMD-LSTM-OPS，再用本章的 AED 代替 LSTM 作为管道预测器构建了 EMD-LSTM-OPS，为了验证 EMD-OPS 管道的效果，这两个模型都被用来与本文第三章提出的 AED-DEMD 进行比较。本部分实验中模型所用到的超参数如表 4.2 所示。

表 4.2 EMD-LSTM-OPS 和 EMD-AED-OPS 的训练超参数

超参数	EMD-LSTM-OPS	EMD-AED-OPS
轮次	220	220
Batch 大小	64	64
学习率	0.0001	0.0001

4.3.2 交易结果与讨论

所有模型在三个数据集上的实验结果列于表 4.3、4.4、4.5 中。为了更加直观的看到各模型效果，图 4.3、4.4、4.5 展示的是基于所有模型在测试集上的预测结果进行自动交易所产生的每日累计回报收益情况。

表 4.3 上证 50 ETF 数据集实验结果

方法	<i>sMAPE</i>	<i>DA</i>	<i>CR</i>	<i>AR</i>	<i>SR</i>	<i>MDD</i>
LSTM	3.1614	49.03%	-8.73%	-6.19%	-0.4429	-25.5%
Mid-LSTM	9.1931	49.58%	4.84%	3.36%	0.2306	-15.51%
N-BEATS	2.1752	50.7%	-2.65%	-1.86%	-0.1104	-18.79%
AED-DEMD	2.2098	53.48%	20.76%	14.12%	0.7452	-11.99%
EMD-LSTM-OPS	2.8144	52.37%	18.89%	12.88%	0.7478	-12.25%
EMD-AED-OPS	3.0106	54.32%	37.11%	24.72%	1.4022	-7.75%

表 4.3 展示的是所有对比模型和本章提出的模型在上证 50 ETF 数据集上的实验结果汇总。从这些结果中可以观察到，本章提出的 EMD-AED-OPS 模型在

$sMAPE$ 、 DA 、 CR 、 AR 、 SR 、 MDD 评估指标上都比其他模型有更好的表现。 DA 指标显示本章的 EMD-AED-OPS 模型在判断趋势方向上的能力要优于其他模型，从而在测试周期上取得了更高的累计回报率和年化回报率。同时，从上表结果来看，基于 EMD-AED-OPS 模型的最大回撤率最小，风险最低。因此，平衡风险之后的收益情况，即夏普比率最高。此外，本章可以注意到，在关键指标 DA 以及累计收益率上，EMD-AED-OPS 比 EMD-LSTM-OPS 有更高，这说明 AED 依然有效，它比 LSTM 更加适合作为管道的预测器。同时，EMD-AED-OPS 的方向准确率和收益率都要高于 AED-DEMD，因此可以证实本章提出的 EMD-OPS 管道比 DEMD 更好。

然而，从表中也可以看到，在 $sMAPE$ 预测精度指标上 N-BEATS 取得的效果最好，不论 EMD-LSTM-OPS 还是 EMD-AED-OPS，因为它们是基于收益表现自适应地选择模态组合作为预测结果，因次会动态地丢弃部分短期波动信息，模型会自主认为这些信息不利于趋势预测，所以导致在预测每日级别上的价格波动上表现稍差，预测价格的绝对误差要大于 N-BEATS。

综上所述，在上证 50 ETF 数据集上的实验结果表明，本章提出的 EMD-OPS 管道方法是对 DEMD 的优化版，能明显提高金融时间序列短期趋势的预测方向准确率，并取得了最高的收益率。同时，所提出的 EMD-AED-OPS 比 EMD-AED-OPS 具有更好的性能，证明 AED 更善于处理趋势信息，因而具有比只有普通 LSTM 的模型具有更佳的效果。

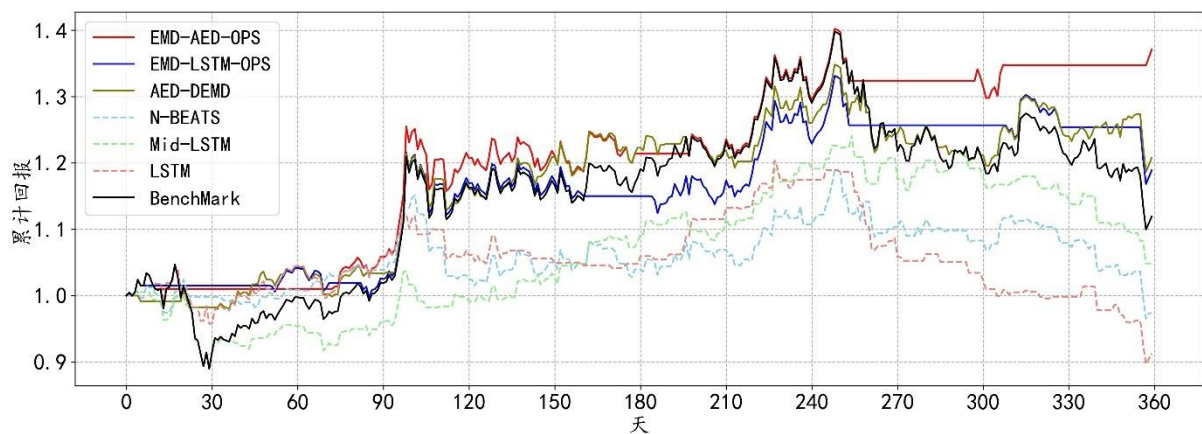


图 4.3 上证 50ETF 数据集累计收益图

各模型在上证 50 ETF 数据集上的累积回报情况如图 4.3 所示。为了更客观地反映市场情况和表现，本章还增加了上证 50 ETF 作为比较基准。在图 4.3 中，本章可以发现 EMD-AED-OPS 的累积回报高于 EMD-LSTM-OPS 的累积回报，并且 EMD-

AED-OPS 的累积回报高于 AED-DEMD。值得注意的是，DEMD-LSTM 模型在市场急剧下跌之前就已经预测出了纠正的趋势。因此，本章设定的自动交易机制在正确的时间进行出售操作，避免了巨大的损失。然后，本章在整个测试期间获得了超额回报。

表 4.4 沪深 300 指数数据集实验结果

方法	<i>sMAPE</i>	<i>DA</i>	<i>CR</i>	<i>AR</i>	<i>SR</i>	<i>MDD</i>
LSTM	3.1452	43.45%	-4.7%	-3.31%	-0.476	-8.77%
Mid-LSTM	2.9272	52.37%	26.98%	18.2%	0.9897	-16.08%
N-BEATS	2.6942	48.47%	10.4%	7.17%	0.5911	-7.71%
AED-DEMD	2.3982	53.82%	37.06%	23.9%	1.4123	-10.52%
EMD-LSTM-OPS	4.0558	53.20%	55.13%	35.98%	2.0353	-7.43%
EMD-AED-OPS	3.7241	55.71%	63.29%	40.95%	2.2733	-7.16%

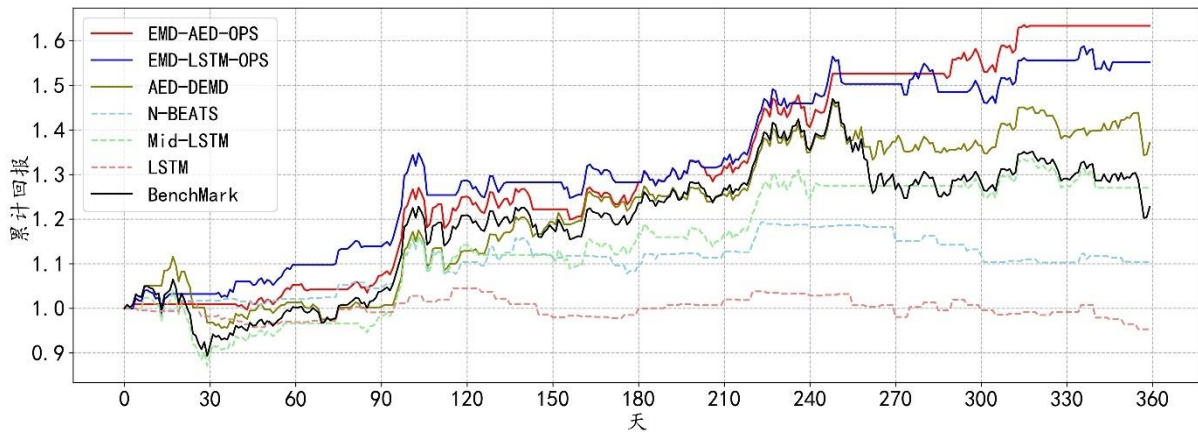


图 4.4 沪深 300 指数数据集累计收益图

表 4.4 展示的是所有模型在沪深 300 指数数据集上的实验结果汇总，图 4.4 展示的是各模型在沪深 300 指数上的累计回报情况。从中本章可以观察到，EMD-AED-OPS 模型在关键指标方向准确率 DA 的表现是最好的，再次验证了本章设计的自适应最优表现选择管道 EMD-OPS 的优越性能，同时在累计收益率和年化收益率上 EMD-AED-OPS 也取得了最优的表现。值得注意的是，从沪深 300 指数数据集上的收益率指标测试结果来看，相对于 AED-DEMD，EMD-AED-LSTM 明显要高于 AED-DEMD，这说明在大涨大跌行情来临时，对于方向的判断和预测非常重要，本章提出的模型具有更好的方向判断能力，能获取更高的超额收益。而从 $sMAPE$ 指

标来看，AED-DEMD 因其细致化的处理取得了最优的效果。

表 4.5 道琼斯指数数据集实验结果

方法	$sMAPE$	DA	CR	AR	SR	MDD
LSTM	2.3374	46.52%	0.28%	0.20%	-0.1356	-4.39%
Mid-LSTM	4.9807	49.03%	0.18%	0.12%	-0.0392	-9.58%
N-BEATS	4.0631	52.09%	10.17%	7.01%	0.6200	-6.09%
AED-DEMD	6.0305	51.81%	18.83%	12.83%	1.1214	-7.22%
EMD-LSTM-OPS	6.1332	53.20%	21.62%	14.68%	1.4316	-5.03%
EMD-AED-OPS	4.7609	53.48%	23.78%	16.11%	1.4800	-8.54%

表 4.5 展示的是所有模型在道琼斯指数数据集上的实验结果汇总，图 4.5 展示的是各模型在道琼斯指数上的累计回报情况。显而易见地，与前两个数据集上的结果相同，从表 4.5 中本章可以观察到，EMD-AED-OPS 模型在关键指标方向准确率、收益率指标比如累计收益率、年化收益率上都达到了最好的效果。这与第三章结论有所差异，第三章中 AED-DEMD 在上证 50 ETF 和沪深 300 指数表现最优，但是在道琼斯指数上没有达到最优，原因是因为 AED-DEMD 基于手工选择模态，泛化能力不足。但是，本章新颖的 EMD-OPS 管道结构解决了这个问题，以自适应可学习的模态选择算法代替了手工选择的方法，提升了模型的泛化能力，在中国和美国三个数据集上都达到了最优的效果。

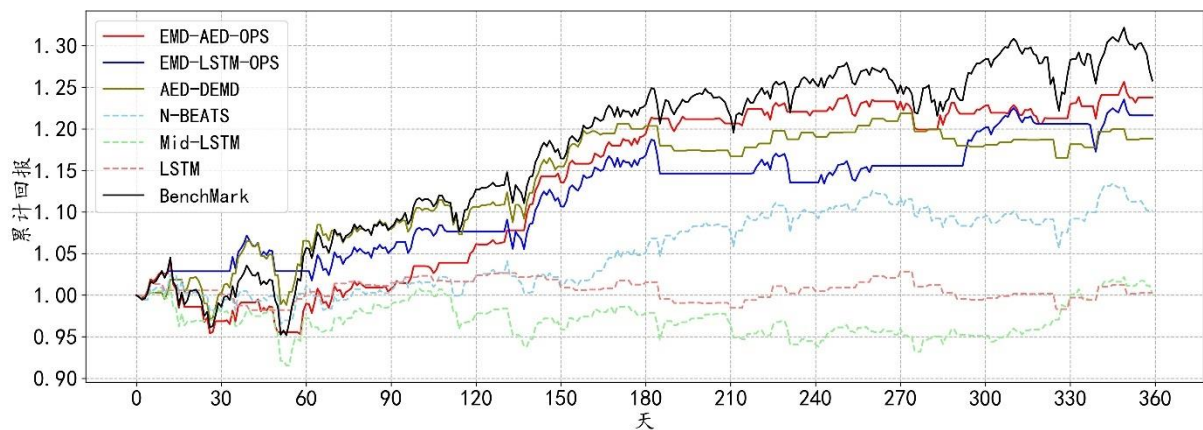


图 4.5 道琼斯指数数据集累计收益图

4.4 本章小结

为金融时间序列这类高波动的数据进行降噪或者过滤是一个亟待解决的问题。

上一章中提出的 AED-DEMD 因其需要手工选择最高频模态分量来进行处理，所以其存在性能上的局限性，并且存在泛化能力不足等问题。

在本章中，本章提出了一个 EMD-OPS 管道，它改变了传统数据处理的一体化流程，被设计为分离式端口的形式。其中 OPS 通过交易和选择表现最好的模态组合来完成模态选择和信号合成的过程，能够以收益指标为导向、自适应地学习最优模态组合。搭配第 3 章提出的 AED，构建了 EMD-AED-OPS 混合模型来完成对金融时间序列短期趋势预测的任务。基于上证 50 ETF、沪深 300 指数以及道琼斯指数的大量实验结果表明，所提出的 EMD-AED-OPS 在预测金融时间序列的方向准确率方面具有最好的性能，投资回报显著提高。同时，通过丰富的实验结果可以证明，本章提出的管道结构相比较 DEMD 泛化能力更强。

第 5 章 基于双端延拓模态分解优化的最优表现选择模型 Opemod

5.1 引言

正如前面所讨论的，金融时间序列因其受到的影响因素非常多，如何对其进行合理降噪和处理以适应预测任务是一个非常具有挑战性的工作。

在上一章节中，本章提出了 EMD-AED-OPS 模型用于股票价格的多步短期趋势预测。从实验结果来看，相对于其他对比模型，EMD-AED-OPS 取得了明显的优势。但是，通过深入分析发现，EMD 在进行模态分解时，需要求解上下包络线，在此过程中存在端点效应^[68-71]，即在某些情况下求解得到的包络线会在端点处发散，产生较大误差，这是由于在序列两端，无法同时存在局部极值点造成的。

基于上述存在的问题，本章节提出了一个基于双端延拓模态分解的最优表现选择模型 Opemod 用于金融时间序列的多步短期趋势预测。首先，本章基于 EMD 提出了双端延拓模态分解方法 TEEMD，它通过在时间序列两端进行不同的延拓、分解再截断的方法有效降低端点效应的影响。然后，本章将 TEEMD 与上一章中提出的 OPS 结合，构建了 TEEMD-OPS 管道，正如前面所讨论的，它能够自适应可学习的方式完成金融时间序列地分解与合成任务。最后，再结合 AED 预测模型，本章构建了全新的 Opemod 模型。本章节其余部分的结构如下：在第 2 节中，本章详细描述了所提出的 Opemod 模型架构。实验配置细节、实验结果的展示与分析在第 3 节呈现。第 4 节对本章进行了总结。

5.2 模型介绍

在本节将详细介绍 Opemod。它由三个部分组成：双端延拓模态分解（TEEMD）、基于注意力机制的编码解码模型（AED）以及最优表现选择模型（OPS）。

5.2.1 Opemod 总体架构

Opemod 旨在进行多步（5 天）股价预测。如前所述，本文认为金融时间序列如股票价格包含不同频率域的信息，高频信号代表波动成分，低频信号代表趋势信息。用高频信号预测高频信号，用低频信号预测低频信号是一种更合理的方法。Opemod 的结构如图 5.1 所示。首先，本章提出了 TEEMD。这是一种数据处理方法，可以更准确地将原始信号 $x(t)$ （包括每日收盘价、市场指数和成交量）分解为若干个模态分量 (IMFs)，并且按照频率从高到低排列。之后，这些 IMFs 被输入 AED 进行预测。AED

的输出是 IMF 的预测结果。最后，本章设计了以投资收益为导向的 OPS 来处理信号合成问题，从而得到最终的预测结果 $y(t)$ 。

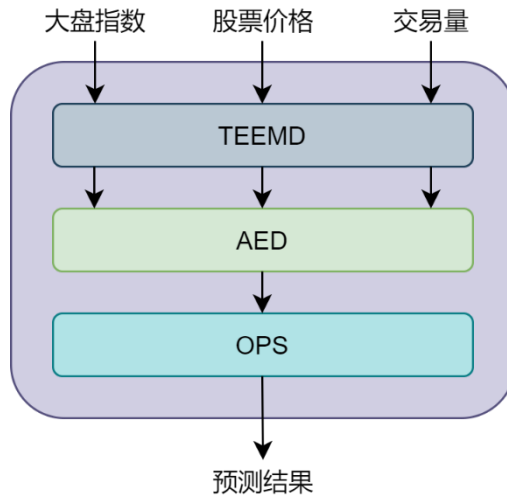


图 5.1 Opemod 架构图

5.2.2 TEEMD 算法

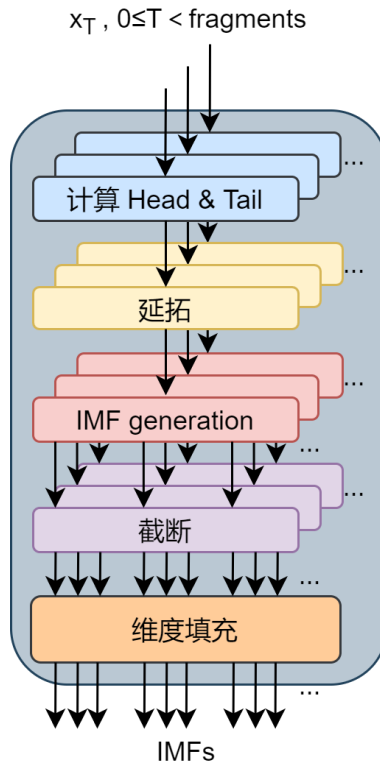


图 5.2 TEEMD 结构图

针对 EMD 中存在的端点效应问题，本章在 EMD 的基础上提出了 TEEMD，可将原始时间序列数据 $x(T)$ 更加准确地分解为若干个 IMF 的。这些 IMF 按频率从高到低的顺

序排列，每个 IMF 具有不同频率的模态特征。TEEMD 的结构如图 5.2 所示。首先，本章用历史已知的原始真实数据作为头部对 $x(T)$ 数据片段进行左侧延拓。然后，计算 $x(T)$ 的局部极大值序列和局部极小值序列，并根据这两对极值序列，基于统计公式分别计算出一对新的局部极大值和极小值点。接着，这两个极值点被当作尾部，用来对 $x(T)$ 做右侧延拓。经过双端延拓之后的 $x(T)$ 被送入 IMF 迭代过程，产生分解结果 $IMFs$ 之后，对每一个 IMF 进行截断操作，截断操作是延拓的反操作，即去掉延拓后 IMF 的头部和尾部，至此，本章即可获得正确维度的 $IMFs$ 。因为，TEEMD 对原始序列数据片段 $x(T)$ 的两端进行了延拓和截断操作，抑制了模态分解过程中因为端点缺少局部极值点而造成的端点效应，从而得到更准确的分解结果。最后，本章整合了所有序列数据片段的分解结果。在实验中，本章发现对各个时间序列片段进行分解得到的模态分量的个数存在差异，这种差异性体现在时间和特征两个维度上，因此为了适应深度预测模型执行预测任务，算法进行了填充操作。考虑到在这里时间序列必须满足可加性，TEEMD 算法使用 0 来填充缺失的维度。TEEMD 计算流程的所有细节可以在算法 5.1 中找到。

算法 5.1 TEEMD 算法

输入: 原始时间序列 $x(T), 0 \leq T < frags$

参数: 封装数据片数 $frags$, 预测序列长度 $seqlen$

输出: 原始数据的本征模态分量 $IMFs$

Let $i = 0, maxshape = 0, \alpha = 1.5$

while $i < frags$ **do**

if $i < seqlen$ **then**

 计算序列片段 $x(i)$ 的极大值序列 $peaks_{max}$

 计算序列片段 $x(i)$ 的极小值序列 $peaks_{min}$

 计算获得极大值点:

 Let $maxP = mean(peaks_{max}) + \alpha * std(peaks_{max})$ 用于对序列片段 $x(i)$ 的右侧延拓

 计算获得极小值点:

 Let $minP = mean(peaks_{min}) - \alpha * std(peaks_{min})$ 用于对序列片段 $x(i)$ 的右侧延拓

 将 $maxP$ 和 $minP$ 当成Tail, 对 $x(i)$ 进行右侧延拓, 得到 $x_{extension}(i)$

 对 $x_{extension}(i)$ 进行 IMF 迭代, 得到 $IMF_{extension}(i)$

 对 $IMF_{extension}(i)$ 进行截断处理, 去除掉右侧延拓Tail部分, 得到 $IMFs(i)$

else

获取 $x(i)$ 之前的长度为 $seqlen$ 的真实序列数据, 作为 $Head$

计算序列片段 $x(i)$ 的极大值序列 $peaks_{max}$

计算序列片段 $x(i)$ 的极小值序列 $peaks_{min}$

计算获得极大值点:

Let $maxP = mean(peaks_{max}) + \alpha * std(peaks_{max})$ 用于对序列片段 $x(i)$ 的右侧延拓

计算获得极小值点:

Let $minP = mean(peaks_{min}) - \alpha * std(peaks_{min})$ 用于对序列片段 $x(i)$ 的右侧延拓

将 $maxP$ 和 $minP$ 当成 $Tail$, 对 $x(i)$ 进行右侧延拓, 得到 $x_{extension}(i)$

对 $x_{extension}(i)$ 进行 IMF 迭代, 得到 $IMF_{extension}(i)$

对 $IMF_{extension}(i)$ 进行截断处理, 去除掉左侧延拓 $Head$ 和右侧延拓 $Tail$ 部分, 得到 $IMFs(i)$

end if

记录 $IMFs$ 的最大个数, 用来进行维度填充操作: $maxshape = \max(maxshape, shape(IMFs))$

令 $i = i + 1$

end while

while $i < frags$ **do**

使用 0 序列对 $IMFs(i)$ 进行时间维度和特征维度上的维度填充

令 $i = i + 1$

end while

return: $IMFs$

5.2.3 AED 模型

AED 是本文第三章中设计的基于注意力机制的编码解码模型, 其结构如图 5.3 所示。AED 是 Opemod 的预测器, 它需要分别对所有的 $IMFs$ 进行分别预测。对于高频的 IMF , AED 将会输出波动性预测结果, 对于低频的 IMF , AED 将会输出趋势性预测结果。这里的数据流与第三章类似, 所有的原始数据在 Opemod 的 TEEMD 端被分解为若干 $IMFs$ 信号序列, 这些 $IMFs$ (标记为 IMF_x) 被送入 AED 的编码器中, 得到中间隐藏状态向量。然后, 多头注意力层重新回顾整个时间片段, 通过使用动态生成的权重来聚合多步过去时间的隐藏状态向量, 以加强对关键模式的学习, 从而产生优化调整后的预测序列。最后, 通过 AED 的解码器对该向量进行解码处理, 得到若干个预测的

$IMFs$ 信号（标记为 IMF_y ），AED 的编码器和解码器都使用 LSTM 来搭建。

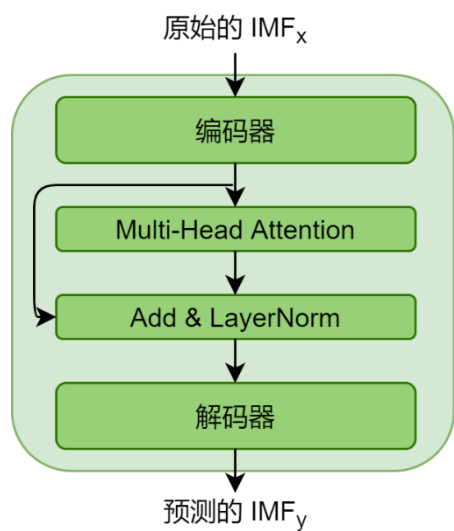


图 5.3 AED 结构图

5.2.4 最优表现选择模型 OPS

OPS 在本文第 4 章被提出，在本章，它是 Opemod 的合成端口。如上所述，本章可以通过 AED 模型获得预测的 $IMFs$ 信号序列，OPS 将根据实际的行情情况，自适应地选择表现最好的 IMF 组合，并将它们合成为复合序列，即最终的趋势预测结果。

5.3 实验结果与分析

5.3.1 实验设置

表 5.1 Opemod 中 AED 各层参数

层	参数	值
Encoder(LSTM)	Input size	3
	Hidden size	54
Multi-Head Attention	Head	2
	Embedding dimension	54
Decoder(LSTM)	Input size	54
	Output size	3

为了验证 Opemod 的有效性，本章分别在上证 50 ETF，沪深 300 指数以及道琼斯指

数进行了丰富的对比实验。三个数据集的详细情况以及训练集、测试集数据封装方法细节可以在 2.3 节找到。Opemod 中 AED 神经网络的各层参数如表 5.1 所示。本章所使用的损失函数为均方误差 MSE，优化器为 Adam。实验中 Opemod 所使用的具体超参数如表 5.2 所示。

表 5.2 Opemod 训练超参数

超参数	值
轮次	220
Batch 大小	64
学习率	0.00006

为了展示端点效应和 TEEMD 在解决端点效应上的实际性能，本章单独可视化了 IMF 迭代过程中，求解得到的时间序列上下包络线。同时，为了更好的验证模型的有效性，在实验中，本章不但将 Opemod 的性能与传统的 LSTM、Mid-LSTM 和 N-BEATS 进行了比较，还与本文前面两个章节所提出的模型 AED-DEMD，EMD-LSTM-OPS 以及 EMD-AED-OPS 进行比较。

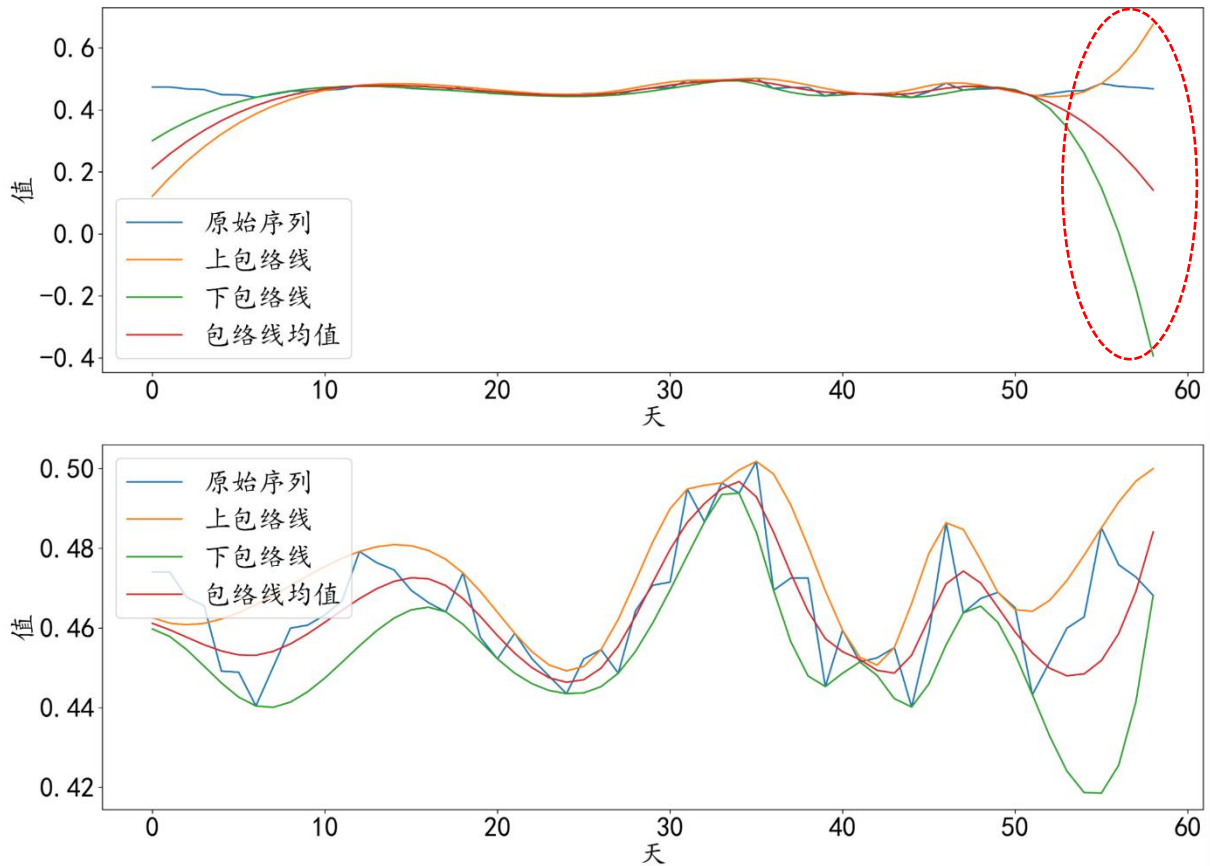


图 5.4 上证 50 ETF 数据集上 EMD 的包络线(上图)与 TEEMD 的包络线(下图)

5.3.2 TEEMD 结果展示与讨论

本章在上证 50 ETF、沪深 300 指数以及道琼斯指数上包络线求解过程可视化结果展示在图 5.4-5.6 中，这些结果是本章随机选择，具有一定的普遍性。从图 5.4-5.6 中可以很明显的看到原始的 EMD 在求解上下包络线时，在序列的端点处，出现了明显的发散现象，图中画红圈标记处包络线向一侧大幅度偏移，直接导致包络线均值出现较大误差。由于分解过程是一个迭代过程，这个误差会一直积累，从而影响整个分解进程。可想而知，因为端点效应的存在，分解得到的 IMF_s 中存在误差。然而，可以明显注意到，TEEMD 算法的计算结果十分优异，其上下包络线紧紧的围绕在真实价格序列两侧，端点处也不例外，包络线均值与真实数据亦步亦趋，可以说 TEEMD 抑制端点效应的效果非常明显。所以，通过 TEEMD 可以获得更加精确的 IMF_s。

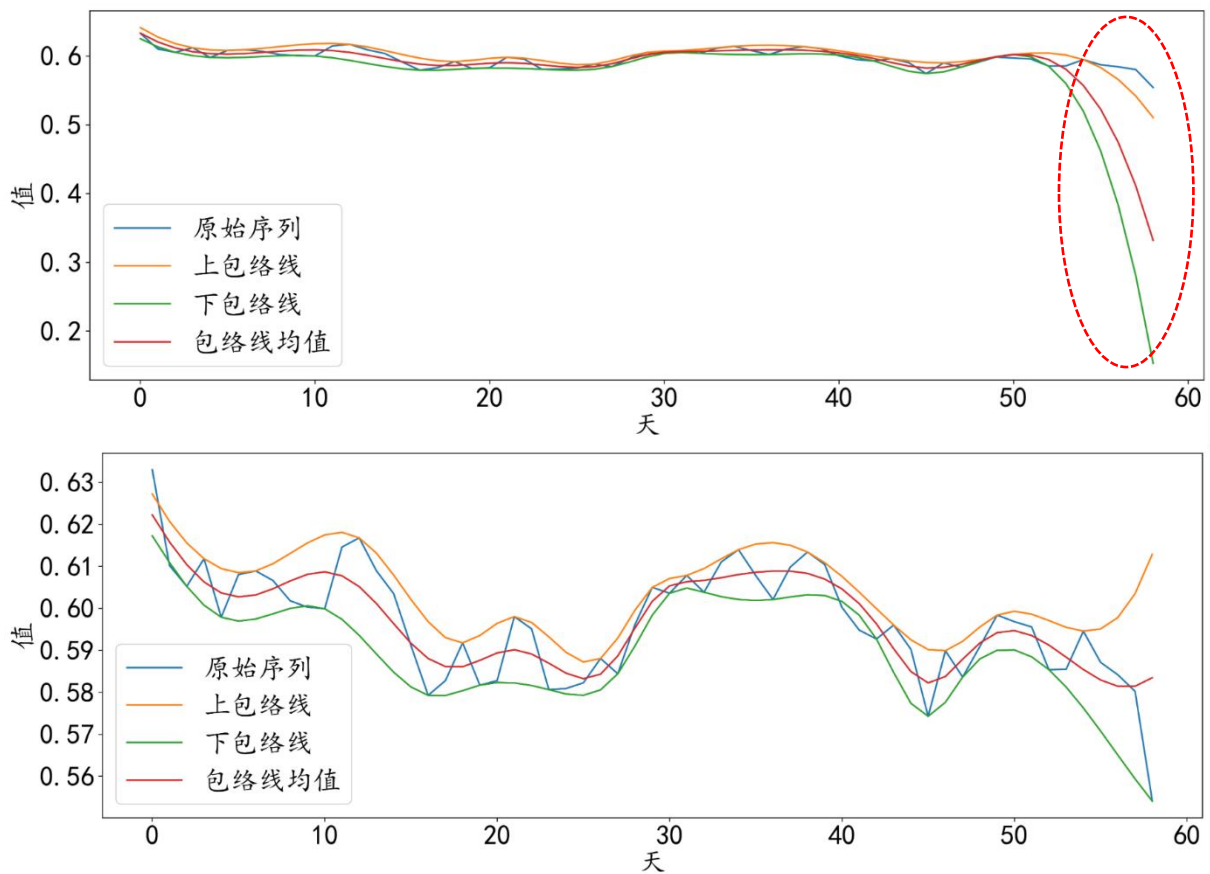
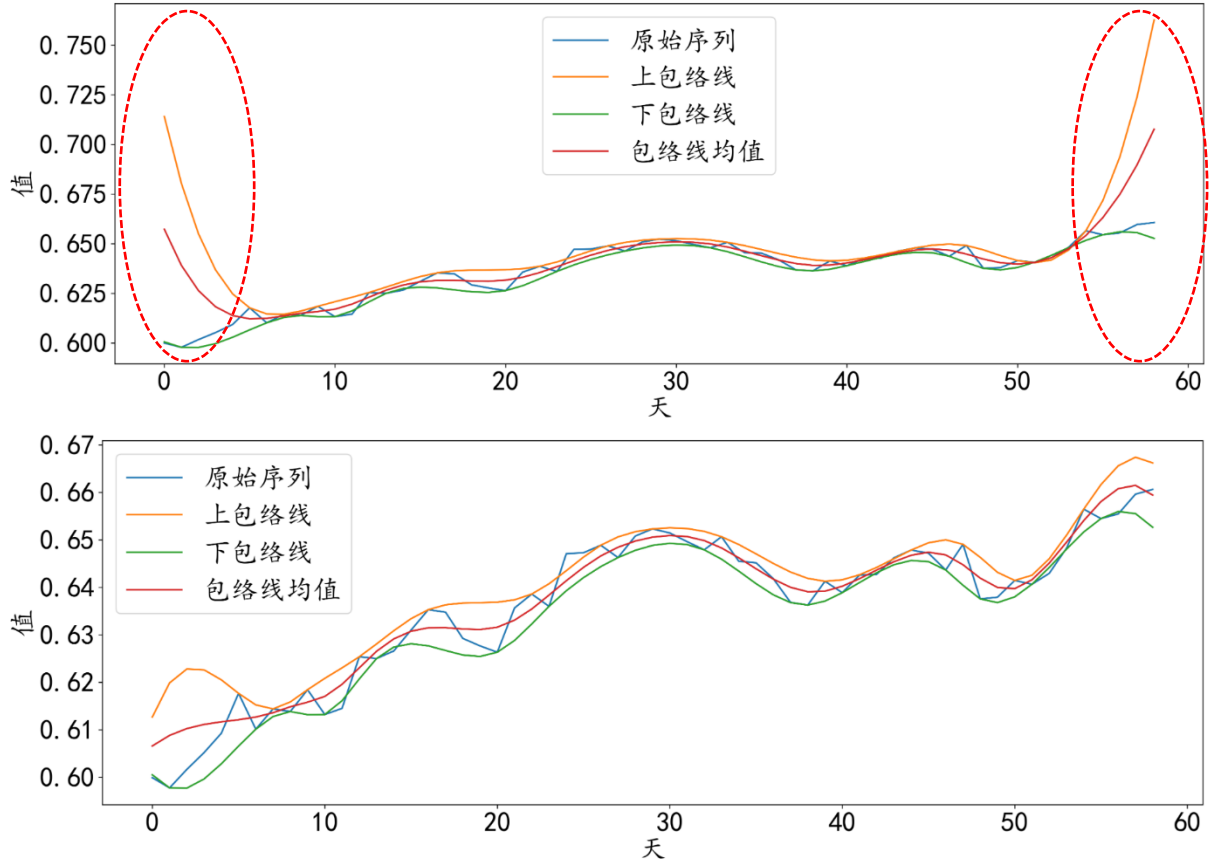


图 5.5 沪深 300 指数数据集上 EMD 的包络线(上图)与 TEEMD 的包络线(下图)



5.6 道琼斯指数数据集上 EMD 的包络线(上图)与 TEEMD 的包络线(下图)

5.3.3 交易结果与讨论

所有模型在上证 50 ETF、沪深 300 指数以及道琼斯指数三个数据集上的实验结果列于表 5.3-5.5 中。为了更加直观的看到各模型效果，图 5.7-5.9 展示的是基于所有模型在测试集上的预测结果进行自动交易所产生的每日累计回报收益情况。

表 5.3 上证 50ETF 数据集实验结果

方法	$sMAPE$	DA	CR	AR	SR	MDD
LSTM	3.1614	49.03%	-8.73%	-6.19%	-0.4429	-25.5%
Mid-LSTM	9.1931	49.58%	4.84%	3.36%	0.2306	-15.51%
N-BEATS	2.1752	50.7%	-2.65%	-1.86%	-0.1104	-18.79%
AED-DEMD	2.2098	53.48%	20.76%	14.12%	0.7452	-11.99%
EMD-LSTM-OPS	2.8144	52.37%	18.89%	12.88%	0.7478	-12.25%
EMD-AED-OPS	3.0106	54.32%	37.11%	24.72%	1.4022	-7.75%
Opemod	2.5657	55.43%	39.32%	26.13%	1.3872	-8.33%

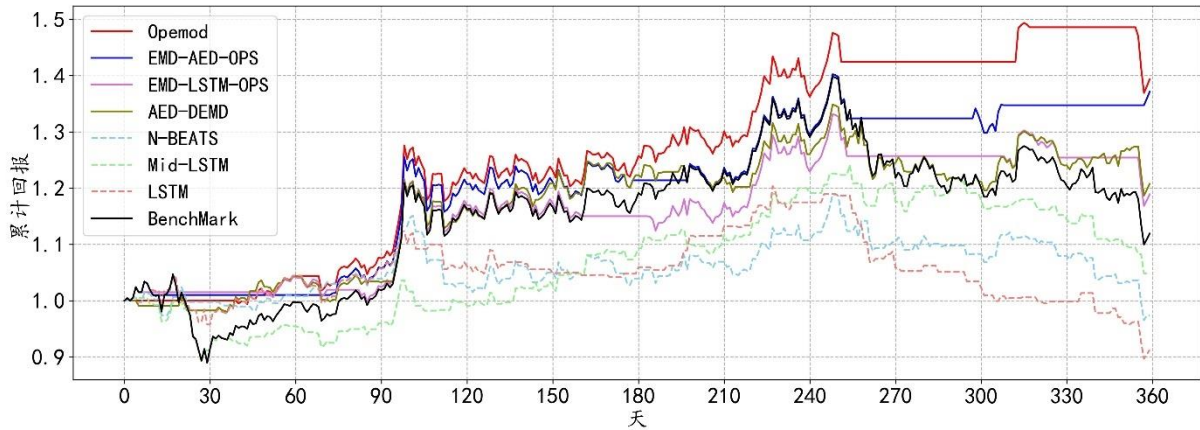


图 5.7 上证 50ETF 数据集累计收益图

表 5.3 展示的是所有模型在上证 50 ETF 上的实验结果汇总。首先，可以看到本章提出的 Opemod 在关键性指标方向准确率 DA 上达到了最优的效果，同时其累计收益率和年化收益率也是最高的。而在预测精度指标上，最好的是 N-BEATS 模型，但是其收益率却是负数，这也说明，在金融时间预测领域，预测价格的精度与收益率并没有直接的正相关关系，预测精度越高不代表能获取的收益率越高，而与收益率正相关的是预测方向的准确性。从表中还可以看到，EMD-LSTM-OPS 的方向准确率比 EMD-AED-OPS 低了 2% 左右，因此说明 AED 判断走势的能力要强于普通的 LSTM。Opemod 比 EMD-AED-OPS 的方向准确率高了约 1.1%，证明了 TEEMD 性能要优于 EMD，因为 TEEMD 解决了端点效应问题，分解获得的 $IMFs$ 更加准确，从而得到了更加准确的趋势判断。

图 5.7 显示了根据 Opemod 以及众多对比模型对上证 50ETF 的预测结果进行回测交易的收益率曲线图。根据此图可以直观的看出 Opemod 的性能是最好的，其收益率高达 39.32%。

表 5.4 展示的是所有模型在沪深 300 指数上的实验结果汇总。与上面的结果类似，Opemod 在关键性指标方向准确率 DA 、累计收益率以及年化收益率上取得了最优的效果，同时最大回撤与 EMD-AED-OPS 持平，为 -7.16%。值得注意的是，Opemod 的累计收益高达 63.38%，方向准确率也超出 EMD-AED-OPS 约 2.2%。而 AED-DEMD 因其细致化处理在 $sMAPE$ 指标上获得了最有表现。通过对比 EMD-LSTM-OPS、EMD-AED-OPS 以及 Opemod 可以看到，它们的方向准确率依次提高，而 $sMAPE$ 逐渐下降，这证明了本章的改进是有作用的。不论是预测趋势的方向，还是价格精度都相对有了一定的提升。

表 5.4 沪深 300 指数数据集实验结果

方法	$sMAPE$	DA	CR	AR	SR	MDD
LSTM	3.1452	43.45%	-4.7%	-3.31%	-0.476	-8.77%
Mid-LSTM	2.9272	52.37%	26.98%	18.2%	0.9897	-16.08%
N-BEATS	2.6942	48.47%	10.4%	7.17%	0.5911	-7.71%
AED-DEMD	2.3982	53.82%	37.06%	23.9%	1.4123	-10.52%
EMD-LSTM-OPS	4.0558	53.20%	55.13%	35.98%	2.0353	-7.43%
EMD-AED-OPS	3.7241	55.71%	63.29%	40.95%	2.2733	-7.16%
Opemod	2.8614	57.94%	63.38%	41.01%	2.1348	-7.16%

图 5.8 显示了根据 Opemod 以及众多对比模型对沪深 300 指数的预测结果进行回测交易的收益率曲线图。根据此图可以直观的看出 Opemod 的性能是最好的，搭配指标来看其收益率高达 63.38%。

表 5.5 展示的是所有模型在道琼斯指数上的实验结果汇总。Opemod 在关键性指标方向准确率 DA 、累计收益率以及年化收益率上取得了最优的效果。值得注意的是，Opemod 这次在 $sMAPE$ 指标上获得了最优表现。

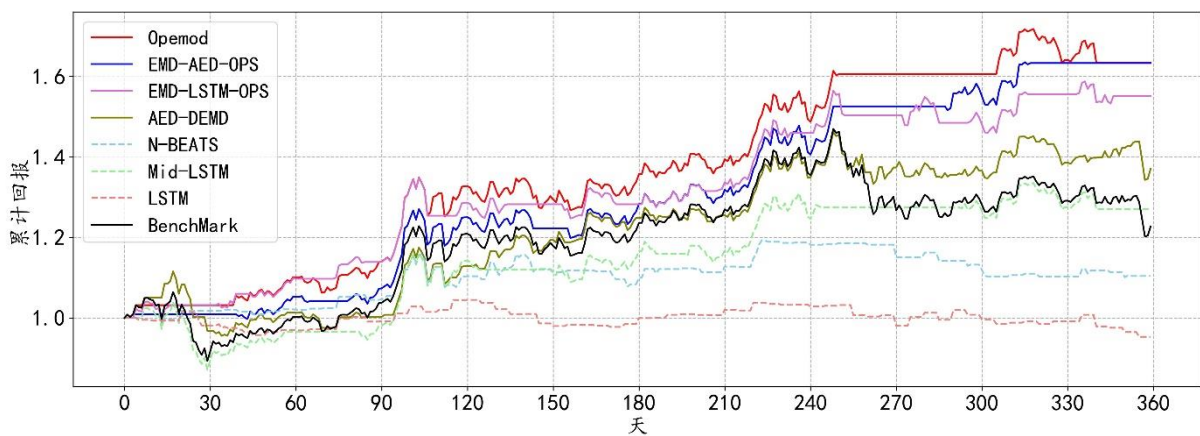


图 5.8 沪深 300 指数数据集累计收益图

相似地，通过对比 EMD-LSTM-OPS、EMD-AED-OPS 以及 Opemod 可以看到，它们的方向准确率 DA 依次提高，而 $sMAPE$ 逐渐下降，这再次证明了本章的改进是有作用的，不论是预测趋势的方向，还是价格精度都相对有了一定的提升。并且，因为 Opemod 集成了 OPS，其良好的泛化性能得到了保留，没有因为换成美股数据集而导致性能大幅度下降。

表 5.5 道琼斯指数数据集实验结果

方法	<i>sMAPE</i>	<i>DA</i>	<i>CR</i>	<i>AR</i>	<i>SR</i>	<i>MDD</i>
LSTM	2.3374	46.52%	0.28%	0.2%	-0.1356	-4.39%
Mid-LSTM	4.9807	49.03%	0.18%	0.12%	-0.0392	-9.58%
N-BEATS	4.0631	52.09%	10.17%	7.01%	0.62	-6.09%
AED-DEMD	6.0305	51.81%	18.83%	12.83%	1.1214	-7.22%
EMD-LSTM-OPS	6.1332	53.2%	21.62%	14.68%	1.4316	-5.03%
EMD-AED-OPS	4.7609	53.48%	23.78%	16.11%	1.48	-8.54%
Opemod	2.7137	57.66%	31.21%	20.94%	1.621	-8.93%

图 5.9 显示了根据 Opemod 以及众多对比模型对道琼斯指数的预测结果进行回测交易的收益率曲线图。可以明显看出，Opemod 地收益率最高，达到了 31.21%。同时这也是在道琼斯指数上实验，首次超越指数本身，并且比 EMD-AED-OPS 和 EMD-LSTM-OPS 各高了 7.5%和 9.6%。

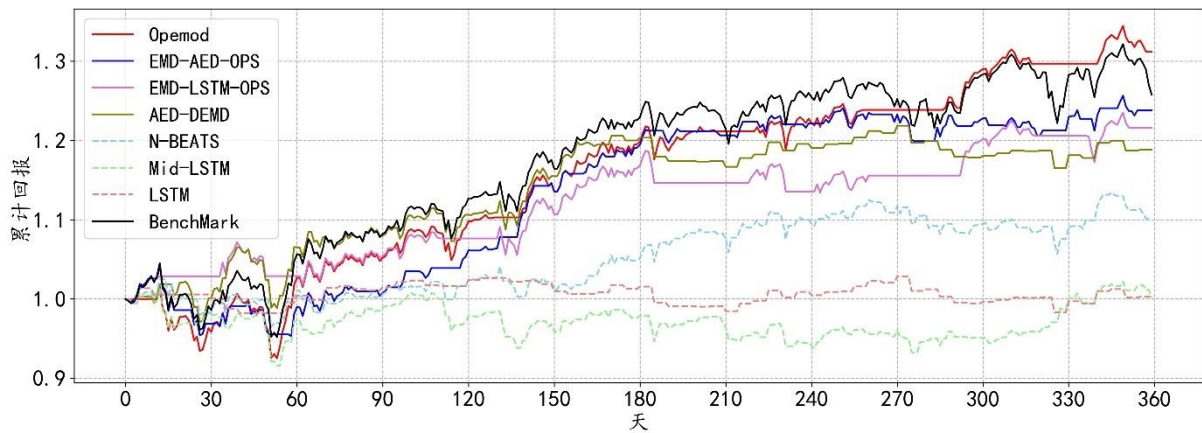


图 5.9 上证 50ETF 数据集累计收益图

5.3.4 稳定性测试

为了验证 Opemod 稳定性，以防展示模型效果是因为某种偶然性或者人工挑选而导致的最优结果，本章还进行了稳定性实验。稳定性实验就是本章对 Opemod 重复进行训练和测试，以获得模型的一般可重复性结果。

表 5.6 Opemod 在上证 50ETF 数据集上的稳定性测试结果

实验次数	<i>sMAPE</i>	<i>DA</i>	<i>CR</i>	<i>AR</i>	<i>SR</i>	<i>MDD</i>
1	2.3213	54.87%	50.93%	33.40%	1.861	-10.71%
2	2.7444	56.27%	42.16%	27.92%	1.5937	-7.75%
3	2.5093	51.25%	30.36%	20.39%	1.1201	-17.33%
4	4.0527	54.04%	25.52%	17.25%	0.8908	-15.28%
5	2.6443	56.55%	48.17%	31.68%	1.6426	-7.75%
6	4.2555	54.87%	32.83%	21.99%	1.2111	-9.24%
7	2.6413	57.38%	41.27%	27.36%	1.5553	-8.50%
8	2.775	54.87%	38.49%	25.60%	1.4238	-9.06%
9	3.4711	56.27%	53.63%	35.06%	1.8310	-7.75%
10	3.4667	56.55%	68.86%	44.30%	2.3027	-7.75%
平均值	3.0882	55.29%	43.22%	28.50%	1.5432	-10.11%
标准差	0.6418	1.67%	12.10%	7.56%	0.3876	3.25%

本章分别在上证 50ETF、沪深 300 指数以及道琼斯指数三个数据集上对 Opemod 重复进行了 10 次实验，所有结果分别展示在表 5.6-5.8 中。表 5.6 展示的是 Opemod 在上证 50 ETF 上的实验结果，本章分别计算了平均值和标准差，分别代表其平均性能和稳定性。从表中可以看到方向准确率 *DA* 的平均值为 55.29%，与上一小节的 55.43% 较为接近。平均累计收益率为 43.22%，比上一小节的 39.32% 还高出了 3.9%。而从其他几个指标来看，波动性差异都在可接受范围内。

表 5.7 Opemod 在沪深 300 指数数据集上的稳定性测试结果

实验次数	<i>sMAPE</i>	<i>DA</i>	<i>CR</i>	<i>AR</i>	<i>SR</i>	<i>MDD</i>
1	2.8153	54.04%	31.83%	21.34%	1.2088	-11.55%
2	4.1372	54.87%	66.87%	43.11%	2.3161	-7.16%
3	2.3381	53.20%	39.62%	26.32%	1.4604	-7.77%
4	2.9211	55.71%	53.28%	34.85%	1.9731	-7.54%
5	2.7907	56.82%	50.26%	32.98%	1.761	-7.77%
6	3.7386	55.71%	45.81%	30.21%	1.7176	-8.68%
7	2.6319	56.27%	55.35%	36.12%	1.9989	-9.13%

8	4.5051	58.22%	59.13%	38.43%	2.0201	-7.16%
9	2.7554	56.55%	56.86%	37.04%	2.0614	-7.77%
10	3.7884	55.43%	67.24%	43.33%	2.3408	-7.60%
平均值	3.2422	55.68%	52.63%	34.37%	1.8858	-8.21%
标准差	0.6969	1.36%	10.67%	6.61%	0.3383	1.26%

表 5.7 展示的是 Opemod 在沪深 300 指数上的实验结果。从表中可以看到方向准确率 DA 的平均值为 55.68%，比上一小节的 57.94% 差了约 2.3%。平均累计收益率为 52.63%，比上一小节的 63.38% 减少了约 10.7%，可以注意到累计收益率的标准差是 10.67%，由此可见，Opemod 存在一定波动性，63.38% 的收益几乎处于波动范围的上界。但是经过稳定性测试可以看到，其方向准确率的标准差只有 1.36%，因此说明 Opemod 的稳定性良好，其平均收益率 52.63% 仍然是一个非常优异的结果。

表 5.8 Opemod 在道琼斯指数数据集上的稳定性测试结果

实验次数	$sMAPE$	DA	CR	AR	SR	MDD
1	2.7137	57.66%	31.21%	20.94%	1.621	-8.93%
2	3.0287	55.71%	37.54%	24.99%	1.9067	-8.10%
3	2.2267	57.66%	37.58%	25.02%	2.1532	-3.87%
4	1.7685	53.48%	18.05%	12.32%	1.1685	-8.44%
5	2.2539	57.38%	40.75%	27.03%	2.2539	-3.89%
6	3.4169	53.48%	20.63%	14.03%	1.0285	-10.66%
7	5.2413	55.71%	29.90%	29.90%	1.4339	-8.93%
8	1.7264	53.20%	21.30%	14.47%	1.3022	-9.35%
9	1.6272	55.43%	28.84%	19.41%	1.3486	-8.10%
10	2.0813	54.60%	26.19%	17.68%	1.3185	-8.93%
平均值	2.6085	55.43%	29.20%	20.58%	1.5535	-7.92%
标准差	1.0377	1.65%	7.40%	5.71%	0.3972	2.14%

表 5.8 展示的是 Opemod 在道琼斯指数上的实验结果。从表中可以看到方向准确率 DA 的平均值为 55.43%，比上一小节的 57.66% 差了约 2.2%。平均累计收益率为 29.20%，

比上一小节的 31.21% 减少了约 2%，仍然远胜于其他模型。并且，其方向准确率的标准差只有 1.65%，因此说明 Opemod 的稳定性良好，其余各项指标波动性差异也都在可接受范围之内。

5.4 本章小结

EMD 在进行模态分解过程中存在端点效应，导致其在计算上下包络线时两侧端点附近会出现发散现象，从而带来较大误差。在本章中，本章在 EMD 的基础上提出了一个基于双端延拓的模态分解算法 TEEMD，并结合第 3 章提出的 AED，以及第 4 章的 OPS 管道模型，构建了全新的 Opemod 用于金融时间序列短期趋势预测。实验结果表明，TEEMD 能够有效地抑制端点效应，降低分解误差。使用 TEEMD 能够获得更加准确地分解结果。同时本章进行了大量的对比实验，表明 Opemod 在三个数据集上都拥有当前较优的性能以及良好的泛化能力，这说明 Opemod 中的 OPS 确实可以自适应不同市场，自主选择最优表现模态组合。此外，稳定性实验结果表明，Opemod 性能优异，波动性在可接受范围内，稳定性良好。

第 6 章 总结与展望

6.1 工作总结

本文主要对基于深度学习模型的金融时间序列趋势预测方法进行研究。针对金融时间序列的高波动、高噪声以及非平稳的特性,通过大量研究和分析提出了几种解决方案,并取得了良好的表现,各方面性能都有所提升。

本文的主要创新点如下:

(1) 本文提出了一个新的 AED-DEMD 模型。该模型通过独特的堆栈结构能够针对高波动的金融时间序列进行细致的降噪处理,减少了原始时间序列数据中存在的高频波动分量,降低了原始时间序列的样本熵值,从而提高了序列趋势的可预测性。首先,利用经验模态分解(EMD)来对金融时间序列进行分解处理,去除其中高频不规则分量,降低其对序列趋势的影响。并且,针对 EMD 存在的模态混叠问题,提出了深度模态分解(DEMD),能够在降低模态混叠效应的影响的同时,尽可能地保留原始数据中的趋势信息。设计实现了一种基于注意力机制(Attention)的编码解码模型 AED,该模型的编码器和解码器由 LSTM 组成,AED 能够利用 Attention 在整个时间窗口上可学习地关注不同的模式,强化模型对不同模式的记忆,从而使模型更加准确的预测趋势,同时,利用残差连接来连接 Encoder 和 Attention 模块,减小搜索范围,提高收敛速度。并且,将提出的 DEMD 与 AED 进行结合,构建了 AED-DEMD 模型。其中,DEMD 用来处理信息、形成趋势;AED 用来处理得到的趋势信息,以预测未来趋势,使得模型达到更优的效果。

(2) 针对(1)中 DEMD 需要人工选择,泛化能力不足的问题,提出了一种新的管道结构 EMD-OPS。EMD-OPS 通过分离式端口来实现时间序列数据的分解与合成的功能。同时,OPS 通过内置的模态选择模块、回测交易模块以及评估模块,可以实现自适应地以收益为导向的自主最优表现模态组合选择。通过实验证明,这种方式明显有助于提升模型的泛化能力,解决了 DEMD 需要人工选择而导致的泛化能力不足的问题。最后,结合(1)中的 AED,提出了 EMD-AED-OPS,实验结果表明,其各项性能有所提升。

(3) 本文针对多步金融时间序列趋势预测问题,提出了一种新的深度学习模型 Opemod。首先,在 EMD 的基础上提出了一种新的模态分解算法 TEEMD,通过在序列两端分别延拓的办法抑制了传统 EMD 在分解序列过程中存在明显的端点效应,减小了

序列分解的误差，提高了分解的准确性。其次，继承了（2）中设计的最优表现选择模型 OPS，结合提出的 TEEMD 设计了 TEEMD-OPS 管道结构，这个结构可用于序列数据的分解与合成。最后，结合（1）中提出的 AED，构建了 Opemod，这是一个完整的集金融时间序列自适应降噪与趋势预测于一身的深度学习模型。大量实验表明，Opemod 的各方面性能达到当前最优。

6.2 后续展望

当前，深度学习应用在投资领域还处于起步阶段，仍然有大量的问题值得探讨。比如说交易策略的问题，本文中使用的交易策略是最简单朴素的交易策略，而在实际应用中，好的预测模型搭配与之相适应的交易策略才能取得更好的收益。所以，关于交易策略的部分也是一个值得研究的问题。同时，通过本文实验结果可以看出，对于波动性较强的上证 50 ETF 以及沪深 300 指数，本章的模型表现要好于道琼斯指数这类波动性相对较弱的趋势性单边市场，这里存在的市场差异性问题的有待进一步研究。

参考文献

- [1] 陈健, 宋文达. 量化投资的特点、策略和发展研究[J]. 时代金融, 2016, 639(29): 246-248.
- [2] Chang P C, Fan C Y, Liu C H. Integrating a Piecewise Linear Representation Method and a Neural Network Model for Stock Trading Points Prediction[J]. IEEE Transactions on Systems Man & Cybernetics Part C, 2008, 39(1): 80-92.
- [3] Hyndman R J, Khandakar Y. Automatic Time Series Forecasting: The forecast Package for R[J]. Journal of statistical software, 2008, 26.
- [4] Gardner E S. Exponential smoothing: The state of the art[J]. Journal of Forecasting, 1985, 4(1): 1-28.
- [5] Taylor J W. Invited Comments on "Exponential Smoothing: The State of the Art - Part II" by E.S. Gardner, Jr. 2006.
- [6] Ortiz-Garcia E G, Salcedo-Sanz S, Perez-Bellido A M, et al. Short - term wind speed prediction in wind farms based on banks of support vector machines[J]. Wind Energy, 2011, 14(2): 193-207.
- [7] Haque A U, Nehrir M H, Mandal P. A Hybrid Intelligent Model for Deterministic and Quantile Regression Approach for Probabilistic Wind Power Forecasting[J]. IEEE Transactions on Power Systems, 2014, 29(4): 1663-1672.
- [8] Chen W, Zhang H, Mehlawat M K, et al. Mean-variance portfolio optimization using machine learning-based stock price prediction[J]. Applied Soft Computing, 2021, 100(1): 106943.
- [9] Hyndman R J, Khandakar Y. Automatic Time Series Forecasting: The forecast Package for R[J]. Journal of statistical software, 2008, 26.
- [10] Gardner E S . Exponential smoothing: The state of the art[J]. Journal of Forecasting, 1985, 4(1): 1-28.
- [11] Taylor J W. Invited Comments on "Exponential Smoothing: The State of the Art - Part II" by E.S. Gardner, Jr. 2006.
- [12] Ortiz-Garcia E G, Salcedo-Sanz S, Perez-Bellido A M, et al. Short - term wind speed prediction in wind farms based on banks of support vector machines[J]. Wind Energy, 2011, 14(2): 193-207.
- [13] Haque A U, Nehrir M H, Mandal P. A Hybrid Intelligent Model for Deterministic and Quantile Regression Approach for Probabilistic Wind Power Forecasting[J]. IEEE Transactions on Power Systems, 2014, 29(4): 1663-1672.
- [14] Chen W, Zhang H, Mehlawat M K, et al. Mean-variance portfolio optimization using machine learning-based stock price prediction[J]. Applied Soft Computing, 2021, 100(1): 106943.

- [15] Chou J, Nguyen T. Forward Forecast of Stock Price Using Sliding-Window Metaheuristic-Optimized Machine-Learning Regression[J]. IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, 2018.
- [16] He Q Q, Pang C I, Si Y W. Transfer Learning for Financial Time Series Forecasting[J]. 2019.
- [17] Shi L, Teng Z, Wang L, et al. DeepClue: Visual Interpretation of Text-Based Deep Stock Prediction[J]. IEEE Transactions on Knowledge & Data Engineering, 2019, PP(6): 1-1.
- [18] Guo T, N Jamet, Betrix V, et al. ESG2Risk: A Deep Learning Framework from ESG News to Stock Volatility Prediction[J]. SSRN Electronic Journal, 2020.
- [19] Wang W, Li W, Zhang N, et al. Portfolio formation with preselection using deep learning from long-term financial data[J]. Expert Systems with Application, 2020, 143(Apr.): 113042.1-113042.17.
- [20] WangLi-Xin. Fast Training Algorithms for Deep Convolutional Fuzzy Systems With Application to Stock Index Prediction[J]. IEEE Transactions on Fuzzy Systems, 2020.
- [21] V Satone, Desai D, Mehta D. Fund2Vec: Mutual Funds Similarity using Graph Learning[J]. Papers, 2021.
- [22] Lecun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015, 521(7553): 436.
- [23] Hu Z, Liu W, Bian J, et al. Listening to Chaotic Whispers: A Deep Learning Framework for News-oriented Stock Trend Prediction[J]. Papers, 2019.
- [24] Bao W, Yue J, Rao Y, et al. A deep learning framework for financial time series using stacked autoencoders and long-short term memory[J]. PLoS ONE, 2017, 12(7): e0180944.
- [25] Hxa B, Lei C A, Zl A, et al. Stock movement prediction via gated recurrent unit network based on reinforcement learning with incorporated attention mechanisms. 2022.
- [26] X Li, Li Y, Zhan Y, et al. Optimistic Bull or Pessimistic Bear: Adaptive Deep Reinforcement Learning for Stock Portfolio Allocation[J]. Papers, 2019.
- [27] Xiong Z, Liu X Y, Shan Z, et al. Practical Deep Reinforcement Learning Approach for Stock Trading[J]. Papers, 2018.
- [28] Hochreiter, Sepp, Schmidhuber, et al. Long short-term memory.[J]. Neural Computation, 1997.
- [29] Selvin S, Vinayakumar R, Gopalakrishnan E A, et al. Stock price prediction using LSTM, RNN and CNN-sliding window model[C]. International Conference on Advances in Computing. IEEE, 2017.
- [30] Li X, Li Y, Liu X Y, et al. Risk Management via Anomaly Circumvent: Mnemonic Deep Learning for Midterm Stock Prediction[J]. Papers, 2019.
- [31] Jing N, Wu Z, Wang H. A Hybrid Model Integrating Deep Learning with Investor Sentiment Analysis

- for Stock Price Prediction[J]. Expert Systems with Applications, 2021, 178(3): 115019.
- [32] Seymore K, McCallum A, Rosenfeld R. Learning Hidden Markov Model Structure for Information Extraction[J]. proc aaai'99 workshop machine learning for information extraction, 1999.
- [33] Oreshkin B N, Dudek G, Peka P, et al. N-BEATS neural network for mid-term electricity load forecasting[J]. Applied Energy, 2021, 293(1): 116918.
- [34] Rojat T, Puget R, Filliat D, et al. Explainable Artificial Intelligence (XAI) on TimeSeries Data: A Survey[J]. 2021.
- [35] Makridakis S, Spiliotis E, Assimakopoulos V, et al. The M4 Competition: 100,000 time series and 61 forecasting methods[J]. International Journal of Forecasting, 2020, 36(1): 54-74.
- [36] Smyl S. A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting[J]. International journal of forecasting, 2020, 36(1): 75-85.
- [37] Huang N E, Wu M, Long S R, et al. A confidence limit for the empirical mode decomposition and Hilbert spectral analysis[J]. Proceedings Mathematical Physical & Engineering Sciences, 2003, 459(2037): 2317-2345.
- [38] 赵进平. 异常事件对 EMD 方法的影响及其解决方法研究[J]. 中国海洋大学学报(自然科学版), 2001, 31(006): 805-814.
- [39] Vaswani A, Shazeer N, Parmar N, et al. Attention Is All You Need[J]. arXiv, 2017.
- [40] Lipton Z C, Kale D C, Elkan C, et al. Learning to Diagnose with LSTM Recurrent Neural Networks[J]. Computer Science, 2015.
- [41] Huang N E. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis[J]. Proceedings of Royal Society of London, 1998, 454.
- [42] Huang N E, Shen Z, Long S R, et al. A new view of nonlinear waves: the Hilbert spectrum[J]. Annual Review of Fluid Mechanics, 1999, 31: 417-457.
- [43] 程军圣, 于德介, 杨宇. EMD 方法在转子局部碰摩故障诊断中的应用[J]. 振动、测试与诊断, 2006.
- [44] 赵进平. 异常事件对 EMD 方法的影响及其解决方法研究[J]. 中国海洋大学学报(自然科学版), 2001, 31(006): 805-814.
- [45] 胡重庆, 李艾华. EMD 间歇信号的检测和提取方法[J]. 数据采集与处理, 2008, 23(1): 4.
- [46] A Z Y, A L Y, B C Q, et al. A method to eliminate riding waves appearing in the empirical AM/FM demodulation[J]. Digital Signal Processing, 2008, 18(4): 488-504.
- [47] Li H, Yang L, Huang D. The study of the intermittency test filtering character of Hilbert-Huang

- p transform[J]. Mathematics & Computers in Simulation, 2005, 70(1): 22-32.
- [48] 秦品乐, 林焰, 陈明. 基于平移不变小波阈值算法的经验模态分解方法[J]. 仪器仪表学报, 2008, 29(12): 5.
- [49] Deering R, Kaiser J F. The use of a masking signal to improve empirical mode decomposition[C]. IEEE International Conference on Acoustics. IEEE, 2005.
- [50] Pham H, Xie Q, Dai Z, et al. Meta Pseudo Labels[J]. 2020.
- [51] Edunov S, Ott M, Auli M, et al. Understanding Back-Translation at Scale[C]. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. 2018.
- [52] Fan D P, Ji G P, Zhou T, et al. PraNet: Parallel Reverse Attention Network for Polyp Segmentation[J]. 2020.
- [53] Hafezi R, Shahrabi J, Hadavandi E. A bat-neural network multi-agent system (BNNMAS) for stock price prediction: Case study of DAX stock price[J]. Applied Soft Computing, 2015, 29: 196-210.
- [54] Pincus S. Approximate entropy as a measure of system complexity[J]. Proc Natl Acad, 1991, 88.
- [55] Richman, Joshua S, Moorman, et al. Physiological time-series analysis using approximate entropy and sample entropy.[J]. American Journal of Physiology: Heart & Circulatory Physiology, 2000.
- [56] Looney D, Hemakom A, Mandic D. Intrinsic multi-scale analysis: a multi-variate empirical mode decomposition framework[J]. Proc Math Phys Eng Sci, 2015, 471(2173): 20140709.
- [57] G. G.-R.Wu, T. C.-T. Hou, J.-L. Lin, Can economic news predict taiwan stock market returns?. Asia Pacific management review 24 (1). 2019: 54-59.
- [58] Chacon H D, Kesici E, Najafirad P. Improving Financial Time Series Prediction Accuracy Using Ensemble Empirical Mode Decomposition and Recurrent Neural Networks[J]. IEEE Access, 2020, PP(99): 1-1.
- [59] 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 Mathematical Physical & Engineering Sciences, 1998, 454(1971): 903-995.
- [60] Cao J, Li Z, Li J. Financial time series forecasting model based on CEEMDAN and LSTM[J]. Physica A: Statistical Mechanics and its Applications, 2019, 519: 127-139.
- [61] M. E. Torres, M. A. Colominas, G. Schlotthauer, et al. A complete ensemble empirical mode decomposition with adaptive noise[C]. IEEE International Conference on Acoustics. IEEE, 2011.
- [62] Gersfelig A, Schraudolph Nicol N, Schmidhuber Jürgen. Learning precise timing with lstm recurrent

- networks[J]. The Journal of Machine Learning Research, 2003.
- [63] Wu Z, Huang N E. Statistical significance test of intrinsic mode functions[M]. 2014.
- [64] Bahdanau D, Cho K , Bengio Y. Neural Machine Translation by Jointly Learning to Align and Translate[J]. Computer Science, 2014.
- [65] Zhang Y, Gong Y, Zhu H, et al. Multi-head enhanced self-attention network for novelty detection[J]. Pattern Recognition, 2020, 107: 107486.
- [66] Wu Z, Huang N E, A study of the characteristics of white noise using the empirical mode decomposition method[J]. Proceedings Mathematical Physical & Engineering Sciences, 2004, 460(2046): 1597-1611.
- [67] He Z, Wang Q, Shen Y, et al. Multivariate Gray Model-Based BEMD for Hyperspectral Image Classification[J]. IEEE Transactions on Instrumentation and Measurement, 2013, 62(5): 889-904.
- [68] Zhi H, Yi S, Qiang W. Boundary extension for Hilbert–Huang transform inspired by gray prediction model[J]. Signal Processing, 2012, 92(3): 685-697.
- [69] Yang J, Shi G, Zhou T, et al. Waveform extension method based on similarity sequential detection for the end effects reduction of EMD[J]. Journal of Vibration and Shock, 2018, 37(18): 121-125.
- [70] Wang Y, Lei W, Wang S. Challenges in Applying the Empirical Mode Decomposition Based Hybrid Algorithm For Forecasting Renewable Wind/Solar in Practical Cases[C]. Power & Energy Society General Meeting. IEEE, 2016.
- [71] Safari N, Price G, Chung C Y. Analysis of Empirical Mode Decomposition-based Load and Renewable Time Series Forecasting[J]. 2020.

致 谢

时光如白驹过隙，时事如白云苍狗。很快就要和研究生岁月告别了。从 2019 年 9 月份开始读研，至今已经度过近三个春秋。在读研的三年里，有苦有乐，有坚持有气馁，既开阔了视野，也为未来的选择踌躇过。虽然大部分时间都很忙碌，却让我受益匪浅。即将毕业之际，在这里向这三年里帮助和关心过我的老师和同学、家人及朋友说一句感谢。

首先，我要感谢我的导师陈剑勇教授。陈剑勇老师待人和善，对待科研问题十分严谨客观，在我的整个读研期间给予了我很大的帮助和指导，并在学术上给了本章很大的自由。在读研期间我们能够自由地选择感兴趣的研究方向，同时陈剑勇老师教会了我们如何去分析问题、文献调研、解决问题，让我意识到了自己的不足，少走了很多弯路。在每周小组会议上，陈剑勇老师都能与我们进行高效的交流，帮助我分析当前遇到的种种问题，指出我没有考虑到的地方，总是在我失败受挫时给予我很大的信心。在完成科研论文以及毕业论文方面，陈剑勇老师也给予了我大量的指导和支持，最终得以完成论文的撰写。在这里由衷地感谢陈剑勇老师的帮助。

我还要感谢实验室的师兄师弟们。在读研期间，给予了我很大的宽慰和快乐。特别是诸事不顺、身处逆境时，师兄师姐们能在我身边支持和鼓励我，为我指明方向。我还要感谢曹远强、陈雷、郭焱平、郑发安、朱汉聪以及我的室友李彪等同学，我们在生活和学习上互帮互助，让研究生生涯中遇到的问题都能得到顺利解决，让我感受到同学情谊的温暖，衷心祝愿你们未来会越来越好。

我还要感谢我的父母，他们给予了我生活上无限的关爱，并支持我的学业，让我得以安心读研。我还要感谢许多默默关心和支持我的同学和朋友，希望你们一切顺利，未来可期。

最后，也十分感谢审阅本文的各位专家和教授。

我将带着人生中这一段难忘的时光，勇敢再出发！

攻读硕士学位期间的研究成果

- [1] **Zichao Xu**, Hongying Zheng, Jianyong Chen. Attention Based Encoder and Decoder Model with Deep Empirical Mode Decomposition for Financial Time Series Prediction. Data&Knowledge Engineering. 2022. (在审)
- [2] **Zichao Xu**, Hongying Zheng, Jianyong Chen. Opemod: An Optimal Performing Selection Model for Prediction of Non-Stationary Financial Time Series. ICANN. 2022. (在审)