

# 控制与决策

Control and Decision

基于TN-LP-LSTM-SVM混合模型的原油期货次日价格变化方向和波动率大小预测

赵戈雅, 薛明皋

引用本文:

赵戈雅, 薛明皋. 基于TN-LP-LSTM-SVM混合模型的原油期货次日价格变化方向和波动率大小预测[J]. 控制与决策, 2022, 37(10): 2627–2636.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2021.0761>

---

## 您可能感兴趣的其他文章

### Articles you may be interested in

[基于EEMD-LMD-LSTM-LEC深度学习模型的短时物流需求预测](#)

Short-term logistics demand forecasting based on EEMD-LMD-LSTM-LEC deep learning model

控制与决策. 2022, 37(10): 2513–2523 <https://doi.org/10.13195/j.kzyjc.2021.0413>

[基于近端强化学习的股价预测方法](#)

Method of stock prices forecast based on proximal reinforcement learning

控制与决策. 2021, 36(4): 967–973 <https://doi.org/10.13195/j.kzyjc.2019.1245>

[基于近端强化学习的股价预测方法](#)

Method of stock prices forecast based on proximal reinforcement learning

控制与决策. 2021, 36(4): 967–973 <https://doi.org/10.13195/j.kzyjc.2019.1245>

[基于卷积长短时记忆神经网络的城市轨道交通短时客流预测](#)

Metro short-term traffic flow prediction with ConvLSTM

控制与决策. 2021, 36(11): 2760–2770 <https://doi.org/10.13195/j.kzyjc.2020.0501>

[迁移学习引导的变源域长短时记忆网络建筑负荷预测](#)

Load forecasting of buildings using LSTM based on transfer learning with variable source domain

控制与决策. 2021, 36(10): 2328–2338 <https://doi.org/10.13195/j.kzyjc.2020.0215>

# 基于TN-LP-LSTM-SVM混合模型的原油期货次日价格 变化方向和波动率大小预测

赵戈雅, 薛明皋<sup>†</sup>

(华中科技大学 管理学院, 武汉 430074)

**摘要:** 原油价格受国际政治、经济、军事、外交以及其他复杂因素的影响, 这些因素的频繁变化使油价表现出随机波动, 给原油投资及交易决策带来困难, 准确预测油价已成为能源领域学术界的热点. 但是, 现有关于原油价格预测的文献大多数是预测原油价格的数值而不是变化方向, 而且不是同时预测原油价格和波动率, 因此无法给投资者充分的决策指导信息. 为了填补这一研究空白, 提出一种结合转移网络(TN)、链接预测(LP)、长短期记忆模型(LSTM)和支持向量机(SVM)的新的混合模型TN-LP-LSTM-SVM来更精确地预测WTI期货次日价格变化方向和波动率大小, 为投资者、能源相关企业和参与政策决策的政府人员提供有益的建议. 在不同的时间窗口下( $h \in [1, 50]$ 且 $h \in \mathbf{Z}^+$ )对TN-LP-LSTM-SVM与CNN-SVM、LSTM和SVM的预测精度作比较, 发现在进行中长期预测时( $h \geq 4$ ), TN-LP-LSTM-SVM总是稳健地优于CNN-SVM、LSTM和SVM.

**关键词:** 原油期货; 方向预测; 转移网络; 链接预测; 长短期记忆模型; 支持向量机

中图分类号: F416.22

文献标志码: A

DOI: 10.13195/j.kzyjc.2021.0761

**引用格式:** 赵戈雅, 薛明皋. 基于TN-LP-LSTM-SVM混合模型的原油期货次日价格变化方向和波动率大小预测[J]. 控制与决策, 2022, 37(10): 2627-2636.

## Predicting next-day price change direction and volatility size of crude oil futures based on a hybrid TN-LP-LSTM-SVM model

ZHAO Ge-ya, XUE Ming-gao<sup>†</sup>

(School of Management, Huazhong University of Science and Technology, Wuhan 430074, China)

**Abstract:** Crude oil prices are influenced by international political, economic, military, diplomatic and other complex factors, and the frequent changes in these factors cause oil prices to exhibit random fluctuations, making crude oil investment and trading decisions difficult. Therefore, predicting oil prices accurately has become a hot research topic in the academic field of energy. However, most of the existing literature on the crude oil price forecasting predicts the value of crude oil prices rather than the change direction, and does not predict crude oil prices and volatility simultaneously, thus can't give investors sufficient information to guide their decisions. To fill this research gap, this paper proposes a new hybrid TN-LP-LSTM-SVM model combining the transition network (TN), link prediction (LP), long short-term memory model (LSTM) and support vector machine (SVM) to predict the next-day price change direction and volatility size of WTI futures more accurately, providing useful advices for investors, energy-related companies, and government personnel involved in policy decisions. Comparing the prediction accuracy of the TN-LP-LSTM-SVM model with the CNN-SVM model, LSTM and SVM for different time windows ( $h \in [1, 50]$  and  $h \in \mathbf{Z}^+$ ), we find that the TN-LP-LSTM-SVM model always outperforms the CNN-SVM model, LSTM and SVM robustly for medium and long term predictions ( $h \geq 4$ ).

**Keywords:** crude oil futures; direction forecast; transition network; link prediction; long short-term memory model; support vector machine

## 0 引言

原油是社会经济发展中必不可少的基本能源、化学材料和战略资源, 其价格变化会严重影响一个

国家的经济发展、社会稳定和国家安全<sup>[1-2]</sup>. 同时, 由于原油市场供需、美元汇率、投机交易、地缘政治冲突、自然灾害等因素的综合影响, 国际原油价格在过

收稿日期: 2021-04-30; 录用日期: 2021-08-26.

基金项目: 国家自然科学基金面上项目(70871046, 71171091, 71471070).

责任编辑: 李勇建.

<sup>†</sup>通讯作者. E-mail: xmgwt@hust.edu.cn.

去10年间剧烈波动,WTI期货价格在每桶-38~114美元之间,具有极高的市场风险.因此,设计科学的方法来准确预测原油价格变化方向和波动率大小,对识别原油市场的极端风险并寻找获利机会具有重要意义.

关于原油价格和波动率预测的研究大多使用计量经济学模型、机器学习方法以及各种混合模型.

早期文献中计量经济学模型被广泛使用.文献[3-5]比较了几种常用的GARCH族模型(GARCH、EGARCH、GJR-GARCH、APARCH、IGARCH、FIGARCH等)对原油价格和波动率的样本外预测表现,发现能刻画波动率的长记忆和非对称性质的GARCH族模型(GJR-GARCH、APARCH、FIGARCH)具有更好的预测能力.而后非参数GARCH族模型<sup>[6]</sup>和混合记忆GARCH等模型<sup>[7]</sup>实现了比传统的GARCH族模型更好的预测表现.BEKK-GARCH<sup>[8-9]</sup>、DCC-GARCH<sup>[10]</sup>等多元GARCH模型被用于研究原油市场和其他金融市场的联动关系.除GARCH族模型外,随机过程<sup>[11]</sup>和异质自回归(HAR)模型<sup>[12]</sup>也被用于原油价格和波动率预测中.

机器学习可以快速且自动生成应用于大规模、多样化数据的分析模型.在预测原油价格的文献中,SVM<sup>[13]</sup>和人工神经网络(ANN)被广泛使用. Chiroma等<sup>[14]</sup>结合遗传算法(GA)优化模型参数以提高ANN的预测精度. Liu等<sup>[15]</sup>结合模糊神经网络(FNN)解决ANN的“黑盒”问题.混合ANN主要使用基于经验模态分解(EMD)<sup>[16-17]</sup>和小波变换(WT)<sup>[18]</sup>的分解集成学习范式.使用单个机器学习模型预测原油波动率的文献很少, Wang等<sup>[19]</sup>结合GA和分类器系统,预测向前一日的原油波动率是否高于给定阈值.

混合模型可以结合各种基础模型刻画数据的不同优势,实现更高的预测精度.在预测原油价格的文献中,混合模型的构建大多基于EEMD<sup>[20]</sup>、变分模态分解(VMD)<sup>[21]</sup>和互补集成经验模态分解(CEEMD)<sup>[22]</sup>的分解集成学习范式.此外, Safari等<sup>[23]</sup>将指数平滑(ES)、ARIMA和非线性自回归神经网络(Narxnet)进行线性组合,每种模型的时变权重由卡尔曼滤波器决定. Wang等<sup>[24]</sup>结合复杂网络分析、ANN和极限学习机(ELM)来优化建模并提升预测精度.在预测原油波动率的文献中,基于机器学习的混合模型很少,形式也较为简单, Kristjanpoller等<sup>[25]</sup>提出ANN-GARCH模型,相比经典的GARCH模型,在异方差调整均方误差(HMSE)方面,将预测精度提高

了30%.

回顾现有文献,可以看出混合模型能实现更高的预测精度,是原油期货价格与波动率预测研究的趋势,但仍存在如下不足:

1) 模型构建大多遵循分解集成方法的建模思路,但是当添加新数据时,同一时期的分解结果往往不同,这会导致这类方法失效.

2) 集中在值预测,关于方向预测的模型非常少.

3) 很少同时考虑原油价格和波动率来制定投资策略,然而波动率大小与价格变化方向的预测精度紧密相关.

4) 在比较模型的预测精度时仅考虑时间窗口长度取1(即向前一步预测)的情况,没有考虑中长期预测.

## 1 问题描述

2020年以来,随着新冠肺炎在全球范围蔓延,原油价格更难以预测,全球经济陷入衰退,各国对原油的需求大幅降低,WTI期货价格自1月初63.27美元/桶下跌至3月6日41.28美元/桶,跌幅达34.76%,市场预期OPEC+会减产以提高价格,却迎来增产与价格战,原油价格再次出现暴跌.4月20日,纽约商品交易所5月交割的WTI期货收盘价为-37.63美元/桶,日跌幅达305.97%.由此可见,原油市场的价格波动水平很高,直接影响大量相关的商品和服务,进而给经济和社会发展带来负面影响,因此预测原油价格变化方向和波动率大小非常重要.

本文基于转移网络(transition network, TN)、长短期记忆模型(long short-term memory, LSTM)、支持向量机(support-vector machine, SVM)和链接预测(link prediction, LP),提出一种新的混合模型TN-LP-LSTM-SVM,使用WTI期货、美元兑挪威克朗和美国大豆期货的历史收盘价、最高价和最低价来预测WTI期货次日价格变化模式(猛增、稳增、猛降或稳降),是关于原油期货价格方向预测的新的探索,可以辅助投资者根据自身的风险态度进行投资决策.

市场上存在许多中长期投资者,本文比较了在不同的时间窗口下( $h \in [1, 50]$ 且 $h \in \mathbf{Z}^+$ ), TN-LP-LSTM-SVM与CNN-SVM、LSTM和SVM的预测精度,投资者可以根据自身的预测期间选择最合适的模型进行投资决策.

## 2 TN-LP-LSTM-SVM混合模型的构建

如图1所示,本文使用网络拓扑特征分析、链接预测和LSTM提取9个特征值,输入到SVM中,得到最后的预测结果.特征提取过程如下:

1) 使用转移网络将WTI期货、美元兑挪威克朗和美国大豆期货这3种资产的价格序列转化为多个复杂网络, 设置滑动窗口长度  $s$  为30, 即基于每30日的价格序列构建一个复杂网络, 然后对每个网络提取两类特征, 即网络拓扑特征指标(4个)和链接预测结果(3个)。

2) 用LSTM分别预测WTI期货次日价格和波动率的数值(2个)。

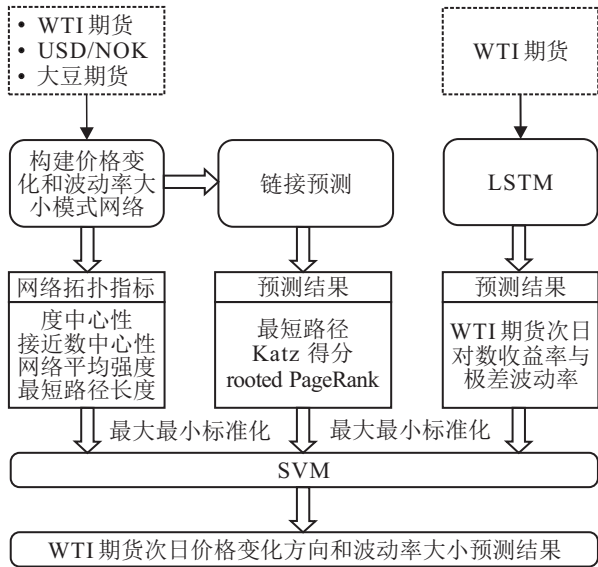


图1 TN-LP-LSTM-SVM的建模思路

### 2.1 构建价格变化方向与波动率大小组合模式网络

本文基于WTI期货、美元兑挪威克朗和美国大豆期货这3种资产的价格序列构建转移网络。

令时间窗口为  $h$ , 基于每日收盘价、最高价和最低价, 计算WTI期货、美元兑挪威克朗和美国大豆期货的每日对数收益率  $r_i(t)$ 、 $h$ 日对数收益率  $R_i(h, t)$ 、 $h$ 日极差波动率  $V_i(h, t)$  以及  $h$ 日极差波动率的均值  $\bar{V}_i(h)$ , 其中  $i$  表示资产类别, 计算方法依次为

$$r_i(t) = \ln \frac{\text{Close}_i(t)}{\text{Close}_i(t-1)}, \quad (1)$$

$$R_i(h, t) = \ln \frac{\text{Close}_i(t)}{\text{Close}_i(t-h)}, \quad (2)$$

$$V_i(h, t) = \left( \frac{1}{4 \ln 2} \right) (\ln H_{t-h,t}^i - \ln L_{t-h,t}^i)^2, \quad (3)$$

$$\bar{V}_i(h) = \frac{1}{T} \sum_{t=1}^T V_i(h, t). \quad (4)$$

其中:  $\text{Close}_i(t)$  是资产  $i$  第  $t$  日的收盘价,  $H_{t-h,t}^i$  和  $L_{t-h,t}^i$  分别是资产  $i$  第  $t-h$  日到第  $t$  日的最高价和最低价. 本文使用Parkinson<sup>[26]</sup>提出的经典的极差波动率的代理变量, Parkinson假定资产价格服从无漂移的几何布朗运动, 在理论上证明了这个波动率代理变量的有效性是传统的基于收盘价的度量方式的5.2倍。

根据  $h$  日对数收益率  $R_i(h, t)$  和极差波动率  $V_i(h, t)$ , 使用下式将第  $i$  种资产第  $t$  日的价格变化模式分为以下4种:

$$\begin{cases} P_1, R_i(h, t) \geq 0 \text{ 且 } V_i(h, t) \geq \bar{V}_i(h) \text{ (猛增);} \\ P_2, R_i(h, t) \geq 0 \text{ 且 } V_i(h, t) < \bar{V}_i(h) \text{ (稳增);} \\ P_3, R_i(h, t) < 0 \text{ 且 } V_i(h, t) < \bar{V}_i(h) \text{ (稳降);} \\ P_4, R_i(h, t) < 0 \text{ 且 } V_i(h, t) \geq \bar{V}_i(h) \text{ (猛降).} \end{cases} \quad (5)$$

通过组合这3种资产的每日价格变化模式可以获得对应的组合模式, 数量最多为  $4^3 = 64$  种. 构建转移网络时以每日组合模式作为网络的节点, 按时间顺序确定网络的边和权重, 其中权重为遵循组合模式的时间顺序的实证转移频次。

### 2.2 网络拓扑结构指标的计算

基于上述转移网络, 提取出4个网络拓扑结构指标: 网络度中心性、网络接近数中心性、网络强度和网络平均最短路径长度。

网络度中心性衡量实际连接数与最大可能的连接数之比, 即网络的边密度. 在无向网络中, 节点的度定义为与节点关联的边的数量, 在有向网络中, 节点的度分为入度和出度. 在构建转移网络时, 本文遵循组合模式观测值的时间顺序, 所以除了第1个节点和最后1个节点, 其他节点的入度和出度均相同, 因此使用平均入度中心性代表网络度中心性, 计算方法如下:

$$\overline{\text{DC}}_{in} = \frac{1}{N(N-1)} \sum_{j=1}^N \sum_{i=1}^N \mathbf{1}_{[(i,j) \in E]}. \quad (6)$$

其中:  $N$  是网络的节点数;  $E$  是边的集合;  $\mathbf{1}_K$  为示性函数, 当逻辑表达式  $K$  为真时返回1, 否则返回0。

网络接近数中心性. 接近数衡量节点传播信息的速度, 节点  $j$  的接近数中心性是网络中的其他节点到节点  $j$  的平均最短距离的倒数乘以可到达节点  $j$  的节点比例, 计算方法如下:

$$C_j = \frac{n-1}{N-1} \frac{n-1}{\sum_{i=1}^{n-1} d_{ij}}. \quad (7)$$

其中:  $d_{ij}$  是从节点  $i$  到节点  $j$  的最短距离,  $n$  是网络中可到达节点  $j$  的节点数(包括节点  $j$ ). 网络接近数中心性是所有节点接近数中心性的均值, 即

$$C = \frac{1}{N} \sum_{j=1}^N \frac{n-1}{N-1} \frac{n-1}{\sum_{i=1}^{n-1} d_{ij}}. \quad (8)$$

网络强度. 节点强度被定义为连接到该节点的连接权重的总和. 在有向网络中, 节点的强度分为入

强度和出强度. 在本文构建的转移网络中, 除了第1个节点和最后1个节点外, 其他节点的入强度和出强度均相同, 因此使用平均出强度代表网络强度, 计算方法如下:

$$S_{\text{out}} = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N w_{ij}, \quad (9)$$

其中  $w_{ij}$  是从节点  $i$  到节点  $j$  的边的权重.

网络平均最短路径长度. 它是网络中所有最短路径的均值, 在加权网络中使用加权最短路径  $l_{ij}$  来计算, 即

$$\overline{\text{SP}} = \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{j=1}^N l_{ij}. \quad (10)$$

### 2.3 链接预测

链接预测基于节点“接近性”的指标, 从网络拓扑结构中提取有关节点未来交互的信息<sup>[27]</sup>. 该方法基于输入图  $G$  的拓扑结构给节点对  $\langle x, y \rangle$  分配链接权重  $\text{score}(x, y)$ , 并生成降序排序列表, 选出排序靠前的链接作为未来可能出现的链接. 现有的链接预测方法可以分为两类: 基于节点邻居的方法和基于节点之间路径的方法. 本文构建有向加权网络, 所以选择最短路径、Katz得分和rooted PageRank这3种基于节点之间路径的方法.

在计算最短路径时, 使用Dijkstra算法获取网络中最后一个节点到其他节点的加权最短路径长度, python networkx包里的single\_source\_dijkstra函数可以方便求得, 此处不再赘述.

Katz得分由Leo Katz提出, 计算这个指标需考虑节点间的所有路径, 对短路径赋予较大的权重, 对长路径赋予较小的权重. 已知邻接矩阵  $A$ , Katz得分矩阵的定义如下:

$$S^{\text{Katz}} = \alpha A + \alpha^2 A^2 + \alpha^3 A^3 + \dots, \quad (11)$$

其中  $\alpha \in (0, 1)$  是控制路径权重的调节参数. 当网络节点数量很大时, Katz得分矩阵在有限时间内难以计算, 但是当  $\alpha$  小于邻接矩阵  $A$  最大特征值的倒数时,  $S^{\text{Katz}}$  可以收敛为

$$S^{\text{Katz}} = (\mathbf{I} - \alpha A)^{-1} - \mathbf{I}. \quad (12)$$

本文使用式(12)来计算转移网络中节点之间的Katz得分方阵  $S^{\text{Katz}}$ , 然后从最后1个节点对应的一行找到最大值, 进而从最大值对应的列得到次日组合模式节点的预测结果.

PageRank是谷歌搜索引擎使用的一种网页排名算法, 类似地, 在链接预测中, rooted PageRank计算从某个节点出发的随机游走粒子到达网络中其他节点

的稳定概率向量, 并对其排序. 初始时刻在节点  $x$  处的粒子在  $t+1$  时刻到达网络中其他节点的概率向量为

$$\vec{q}_x(t+1) = c\vec{q}_x(t) + (1-c)\vec{e}_x. \quad (13)$$

其中:  $c$  表示每一步随机游走粒子返回节点  $x$  的概率,  $\vec{e}_x$  表示初始状态(节点  $x$  位置上对应的元素为1, 其他元素全为0). 式(13)的稳态解

$$\vec{q}_x = (1-c)(\mathbf{I} - c\mathbf{P}^T)^{-1}\vec{e}_x, \quad (14)$$

其中  $\mathbf{P}$  为网络的马尔可夫概率转移矩阵. 本文使用式(14)计算从最后1个组合模式节点到其他节点的稳定概率向量, 找到向量中的最大值, 进而从最大值对应的行得到次日组合模式节点的预测结果.

### 2.4 LSTM

实证研究证明在时间序列的值预测方面, 相比传统的GARCH族模型, LSTM的耗时更短, 多数情况下精度更高且能以较高的预测精度向前预测许多步.

LSTM通过建立门机制从单元状态移除旧信息和添加新信息, 其结构如图2所示. 其中:  $x_t$  表示输入,  $C_t$  表示单元状态,  $h_t$  表示隐藏层的输出. 下面简要说明LSTM中的各个步骤:

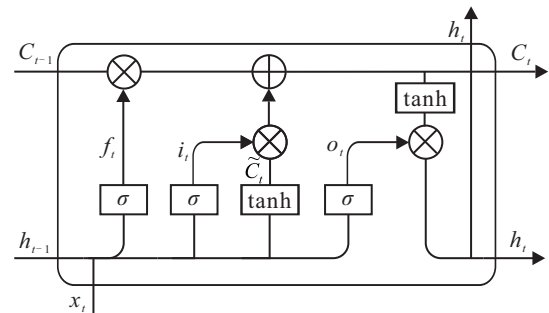


图2 LSTM的结构

1) “遗忘门”决定从单元状态中丢弃哪些信息, 它基于  $h_{t-1}$  和  $x_t$  输出一个0-1之间的函数值向量  $f_t$ , 1表示完全保留  $C_{t-1}$  中对应元素的信息, 0表示完全丢弃, 计算如下所示:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f). \quad (15)$$

2) “输入门”决定在单元状态中存储哪些新信息. tanh层创建一个新候选值向量  $\tilde{C}_t$ , 即

$$\tilde{C}_t = \tanh(W_C[h_{t-1}, x_t] + b_C); \quad (16)$$

然后“输入门”决定将更新哪些值, 即

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i); \quad (17)$$

最后按下式更新单元状态(其中  $\odot$  表示Hadamard积):

$$C_t = C_{t-1} \odot f_t + \tilde{C}_t \odot i_t. \quad (18)$$

3) “输出门”决定隐藏层的输出,使用 tanh 层处理更新后的单元状态,并按照下式计算“输出门”:

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o); \quad (19)$$

最后得到隐藏层输出

$$h_t = o_t \odot \tanh(C_t). \quad (20)$$

### 2.5 SVM

当前标准的SVM由Cortes等<sup>[28]</sup>提出,适用于小样本、非线性和高维模式识别问题,克服了“维数灾难”和“过度学习”。本文基于预测当日的组合模式将训练集划分为训练子集,进而构建不同的分类子模型(最多64个),用于构建每个子模型的样本量比较小。所以最终选择使用SVM而不是其他更复杂的机器学习分类器。

## 3 实证结果与分析

### 3.1 数据收集

从现有文献可以总结出,WTI期货价格与农产品期货价格和汇率有非常密切的关系。一方面,农业是能源密集型行业,油价与农产品生产成本直接相关,而且基于农产品的生物燃料是原油的替代品,本文选取美国大豆期货作为农产品期货的代表。另一方面,油价通过影响国际收支进而影响汇率,挪威是欧洲最大的产油国,挪威克朗是与油价相关性最强

的货币之一,故选取美元兑挪威克朗作为汇率的代表。本文利用WTI期货、美元兑挪威克朗和美国大豆期货的历史价格序列进行建模。

混合模型的构建需要大量的数据,本文首先从 <https://cn.investing.com> 上爬取到2020-11-12为止的WTI期货、美元兑挪威克朗和美国大豆期货的所有每日收盘价、最高价和最低价,然后筛选出这3种资产的价格记录同时存在的日期,最终选取了1990-01-02~2020-11-12总共7750条每日记录。

训练LSTM需要大量样本,一个滑动窗口长度的数据量远远不够,需要在已有样本的基础上,往前收集所有可获取的WTI期货数据,最终补充了1983-03-31~1989-12-29总共1695条每日记录。

### 3.2 结果与分析

WTI期货、美元兑挪威克朗和美国大豆期货的对数收益率的描述统计和交叉相关矩阵分别如表1和表2所示。表2中“0.038”表示WTI和滞后1期的美元兑挪威克朗的对数收益率的交叉相关系数。

表1 3种资产对数收益率的描述统计 %

资产	均值	标准差	偏度	超额峰度	最小值	最大值
WTI	0.008	2.577	-0.854	20.748	-40.048	22.394
USD/NOK	0.004	0.742	0.343	4.528	-5.625	7.177
Soybean	0.009	1.516	-0.932	18.494	-23.411	20.321

表2 3种资产对数收益率的交叉相关矩阵

资产	Lag = 0			Lag = 1			Lag = 2			Lag = 3			Lag = 4		
	WTI	USD/NOK	Soybean	WTI	USD/NOK	Soybean	WTI	USD/NOK	Soybean	WTI	USD/NOK	Soybean	WTI	USD/NOK	Soybean
WTI	1.000	-0.245	0.146	-0.015	<b>0.038</b>	0.004	-0.048	-0.022	0.000	-0.008	-0.019	0.001	0.040	-0.036	0.001
USD/NOK	-0.245	1.000	-0.168	-0.016	-0.004	-0.022	-0.018	-0.003	-0.010	-0.024	-0.017	-0.009	-0.012	0.009	-0.002
Soybean	0.146	-0.168	1.000	-0.027	0.009	0.001	0.009	-0.007	-0.009	-0.002	-0.004	0.011	0.009	-0.004	-0.005
资产	Lag = 5			Lag = 6			Lag = 7			Lag = 8			Lag = 9		
	WTI	USD/NOK	Soybean	WTI	USD/NOK	Soybean	WTI	USD/NOK	Soybean	WTI	USD/NOK	Soybean	WTI	USD/NOK	Soybean
WTI	-0.027	0.031	-0.033	-0.035	-0.017	0.014	0.012	-0.001	0.012	0.008	-0.020	0.004	-0.007	0.005	-0.003
USD/NOK	-0.006	-0.017	0.010	0.020	-0.025	0.006	0.011	0.003	-0.016	-0.006	-0.007	-0.011	-0.006	-0.007	0.011
Soybean	0.003	0.011	-0.037	-0.011	0.019	0.001	-0.008	0.004	-0.009	-0.001	0.014	0.027	-0.013	0.009	0.020

由表1可以看出,这3种资产的对数收益率具有尖峰厚尾特征且是有偏的,在这种情况下,机器学习方法往往优于计量经济学方法。由表2可以看出,美元兑挪威克朗和美国大豆期货的对数收益率与WTI期货的对数收益率在同期相关系数最大(分别为-0.245和0.146),验证了前文构造组合模式的合理性。从交叉相关矩阵的第1行可以看出,3种资产的对

数收益率的滞后值都有助于预测WTI期货的对数收益率。

构建混合模型的第1步是使用转移网络将3种资产的价格序列转换为多个复杂网络,设置滑动窗口长度  $s = 30$ ,以时间窗口长度取6的情况为例进行说明。如图3所示,第1日的组合模式为  $P_1P_4P_2$ ,第2日的组合模式为  $P_4P_4P_2$ ,所以生成一条从节点  $P_1P_4P_2$

到节点  $P_4P_4P_2$  的权重为1的有向边,如果节点之间的转换频次为  $w$ ,则有向边的权重为  $w$ ,按此规则共生成了7715个组合模式网络.第1个网络如图4所示.

日期	WTI	USD/NOK	Soybean	组合模式
1	$P_1$	$P_4$	$P_2$	$P_1P_4P_2$ → 第1个滑动窗口 (网络拓扑指标、 链接预测结果、 LSTM预测结果)
2	$P_4$	$P_4$	$P_2$	$P_4P_4P_2$
3	$P_3$	$P_3$	$P_2$	$P_3P_3P_2$
4	$P_3$	$P_3$	$P_3$	$P_3P_3P_3$ → 第2个滑动窗口 (网络拓扑指标、 链接预测结果、 LSTM预测结果)
⋮	⋮	⋮	⋮	⋮
30	$P_3$	$P_3$	$P_2$	$P_3P_3P_2$
31	$P_3$	$P_3$	$P_2$	$P_3P_3P_2$
⋮	⋮	⋮	⋮	⋮
7715	$P_4$	$P_4$	$P_2$	$P_4P_4P_2$ → 第7715个滑动窗口 (网络拓扑指标、 链接预测结果、 LSTM预测结果)
⋮	⋮	⋮	⋮	⋮
7744	$P_1$	$P_4$	$P_1$	$P_1P_4P_1$

图3 组合模式的构建和滑动窗口的移动过程

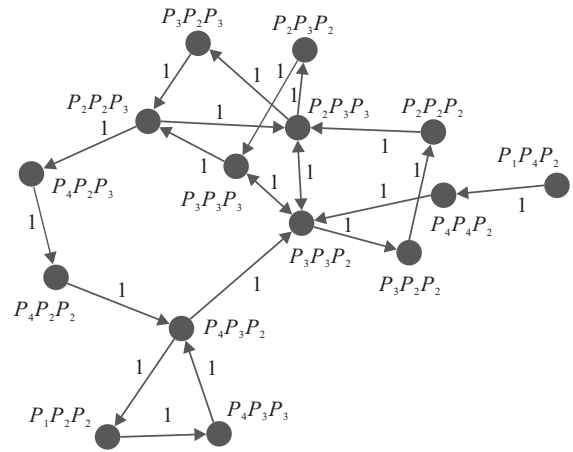


图4 组合模式网络示例

针对每个网络,分别使用式(6)、(8)~(10)计算网络度中心性、网络接近数中心性、网络平均强度和网络平均最短路径长度.图5描绘了3种资产的价格曲线以及上述4个网络拓扑特征指标的演变曲线,并标注出了A、B、C、D四个时点.可以看出,在这四个时点,

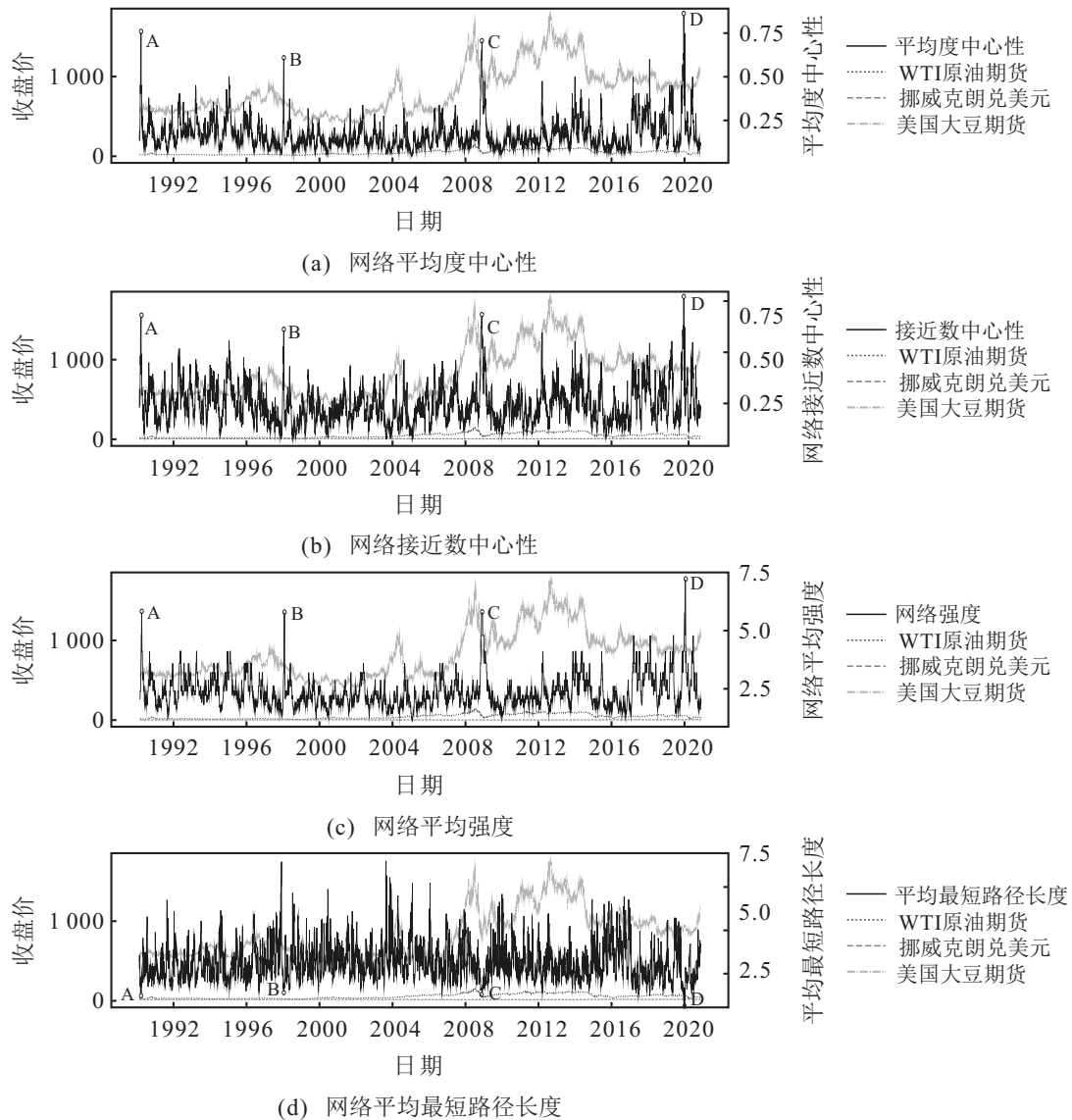


图5 网络拓扑特征值随时间的演变

网络平均度中心性、网络接近数中心性和网络强度达到较大值,平均最短路径达到较小值.对照价格曲线,当3种资产的价格剧烈波动时,网络中的节点数较少,相应地网络节点之间的加权最短路径较短,因此网络接近数中心性较大,网络平均最短路径较小.网络节点不断在几种组合模式之间切换,网络的边密度和有向边的权重较大,相应地网络平均度中心性和网络强度较大.

在A、B、C、D四个时点,分别发生了4个让WTI期货产生剧烈波动的事件.在时点A,1990-08-02,海湾战争爆发,伊拉克入侵科威特,受到国际制裁,原油供应中断,原油价格从7月底的每桶21美元上涨到10月中旬的每桶46美元.在时点B,1998年亚洲金融危机爆发,亚太地区的原油消耗量自1982年以来首次下降,较低的消耗量和较高的OPEC产量共同导致原油价格下跌.在时点C,2008年全球金融危机爆发,原油价格由2008年6月的最高价165.48美元/桶跌至2009年1月的最低价51.06美元/桶,跌幅达70%.在时点D,由于沙特阿拉伯无法与俄罗斯就石油减产达成一致,报复性增加石油开采量,引发油价大幅下跌,WTI期货更是在2020-04-20录得有史以来的首次负油价.

除了从网络中提取拓扑特征指标外,本文使用基于路径的链接预测方法进一步挖掘网络的拓扑结构信息,此处以图4中的第1个组合模式网络为例说明链接预测方法的实现过程.

1) 基于最短路径:调用python相关包计算最后1个节点 $P_3P_3P_2$ 到网络中所有节点的最短路径长度,结果发现节点 $P_3P_3P_2$ 处存在自环,如表3所示,预测第31日的节点为 $P_3P_3P_2$ .

表3 节点 $P_3P_3P_2$ 到网络中所有节点的最短路径

节点	最短路径长度	最短路径
332	0	332
333	1	332, 333
233	1	332, 233
322	1	332, 322
223	2	332, 333, 223
323	2	332, 233, 323
232	2	332, 233, 232
222	2	332, 322, 222
423	3	332, 333, 223, 423
422	4	332, 333, 223, 423, 422
432	5	332, 333, 223, 423, 422, 432
122	6	332, 333, 223, 423, 422, 432, 122
433	7	332, 333, 223, 423, 422, 432, 122, 433

2) 基于Katz得分:调节参数 $\alpha = 0.15$ ,使用式(12)计算节点之间的Katz得分方阵,为节约空间显示如下:

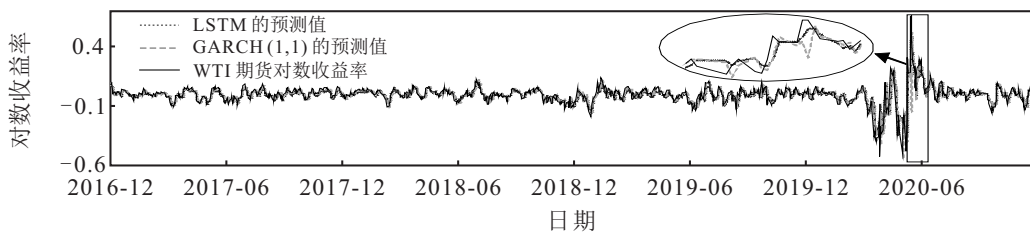
$$S^{Katz} = \begin{bmatrix} 0 & 0.15 & 0.045 & 0.008 & \dots & 0.001 \\ 0 & 0 & 0.303 & 0.055 & \dots & 0.009 \\ 0 & 0 & 1.022 & 0.367 & \dots & 0.063 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0.065 & 0.017 & \dots & 0.178 \end{bmatrix}. \quad (21)$$

节点 $P_3P_3P_2$ 对应方阵第3行,该行最大值1.022位于第3列,预测第31日的节点为 $P_3P_3P_2$ .

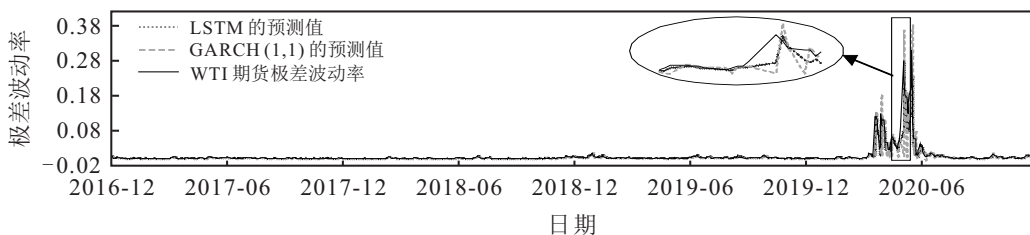
3) 基于rooted PageRank:使用式(14)计算从节点 $P_3P_3P_2$ 到其他节点的稳定概率向量,即

$$\vec{q}_{332} = [0, 0, 0.893, 0.032, \dots, 0.033, 0.004]^T. \quad (22)$$

最大值位于第3行,预测第31日的节点为 $P_3P_3P_2$ .



(a) WTI期货对数收益率的预测结果



(b) WTI期货极差波动率的预测结果

图6 LSTM和GARCH(1,1)对WTI期货对数收益率和极差波动率的预测结果比较

由图3可知,第31日对应的组合模式实际为 $P_3P_3P_2$ ,上述3种链接预测方法的结果都是正确的。

如3.1节所述,本文补充了样本数以拟合LSTM.由于LSTM具有比较好的可扩展性,为了节省运行时间,每拟合一个LSTM,就向前预测200步,然后将对应的200个 $R(h,t)$ 和 $V(h,t)$ 的实际值加入训练样本,继续拟合下一个LSTM,以此类推.以时间窗口 $h=6$ 为例,总共得到了7714个预测结果.图6和表4比较了LSTM和GARCH(1,1)在最后1000个数据上的预测表现,其中LSTM是向前200步预测,GARCH(1,1)是向前1步预测,每步预测的均值方程中ARIMA模型的阶数由AIC决定.

表4 LSTM和GARCH(1,1)对WTI期货对数收益率和极差波动率的预测结果比较

评价指标	WTI期货对数收益率		WTI期货极差波动率	
	LSTM	GARCH(1,1)	LSTM	GARCH(1,1)
MSE	<b>2.103e-3</b>	2.862e-3	<b>1.404e-4</b>	2.631e-4
MAE	<b>2.586e-2</b>	2.644e-2	<b>2.573e-3</b>	2.877e-3
MAPE	<b>2.862</b>	3.428	<b>3.872e-1</b>	4.929e-1
DM-test	$pvalue = 0.1149$		$pvalue = 0.01971$	

注:Diebold-Mariano检验的 $H_1$ : LSTM > GARCH(1,1)

图6矩形区域表示一个月的曲线段,通过放大可以看出,在剧烈波动阶段,LSTM的预测值较GARCH(1,1)更接近真实值,在平稳阶段,二者的预测表现基本一致.由表4可知,对于MSE、MAE、MAPE这3个常见的预测误差评价指标,LSTM均优于GARCH(1,1).DM检验结果表明,预测WTI期货对数收益率时,LSTM与GARCH(1,1)模型的预测精度没有显著的差异,而预测WTI期货极差波动率时,在5%的显著性水平上,LSTM的预测精度显著高于GARCH(1,1).由此可见,在构建混合模型时LSTM是更优的选择.

在构建混合模型的最后1步,本文根据预测当日的组合模式训练不同的分类子模型,将前面得到的9个特征值输入不同的SVM分类器,其中数值型特征(网络拓扑指标、LSTM值预测)需要进行最大最小标准化.在构建每个分类子模型时,使用网格搜索算法对SVM分类器的主要参数(核函数的类别和惩罚参数 $C$ )进行调优.在对测试集数据进行预测时,首先根据预测当日的组合模式找到对应的分类子模型,然后将所有的特征值输入该模型,得到预测结果.以时间窗口 $h=6$ 为例,总共得到7715个组合模式网络,由于最后1个网络的次日组合模式未知,所以模型可用的数据为7714条.使用常用比例4:1划分出6171条

训练集数据和1543条测试集数据.训练集被划分为64个训练子集,组合模式“232”对应的训练子集数据量最大(486条),组合模式“431”对应的训练子集数据量最小(6条).由此可见,每个训练子集的数据量都比较小,所以在构建混合模型的最后1步时,选择了SVM分类器而不是其他更复杂的分类器以防止过拟合问题.

为了验证TN-LP-LSTM-SVM的有效性,本文比较了在不同的时间窗口下,TN-LP-LSTM-SVM、CNN-SVM、LSTM和SVM的预测精度.其中CNN-SVM由Niu等<sup>[29]</sup>提出,通过用SVM替换CNN的最后1个输出层来构建,使用CNN自动从原始图像中提取特征,再将特征输入到SVM中,提高了手写体识别问题的分类精度.由于本文的数据为时间序列,构建CNN-SVM时,使用一维卷积,通道数为6(3种资产的收益率和极差波动率).输入到CNN-SVM和LSTM的时间序列的长度与TN-LP-LSTM-SVM的滑动时间窗口长度相同,SVM的输入为预测当日3种资产的收益率和极差波动率.比较结果如图7和表5所示,可以总结出以下几点:

1)从整体的趋势来看,随着时间窗口变长,这4种模型的预测精度变高,因为时间窗口越长,预测WTI期货价格变化方向和波动率大小可用的有效信息越多.

2) TN-LP-LSTM-SVM的预测精度始终高于SVM,表明混合模型中提取的特征稳健地提高了SVM的预测精度.

3)当 $h=1$ 时,CNN-SVM的预测精度最高;当 $h \in [2,3]$ 且 $h \in \mathbf{Z}^+$ 时,LSTM的预测精度最高;当 $h \geq 4$ 时,TN-LP-LSTM-SVM稳健地优于其他3种模型,而且这种优势比之前的劣势更显著.原因如下:当时间窗口很短时,时间序列的非线性特征特别强烈,变化规律难以捕捉,CNN-SVM和LSTM略优于本文提出的混合模型;当时间窗口变长,时间序列的非线性、不确定性和非平稳性特征减弱,可以通过复杂网络分析、链接预测、LSTM值预测等方法探索变化规律,本文提出的混合模型优于其他3种模型.

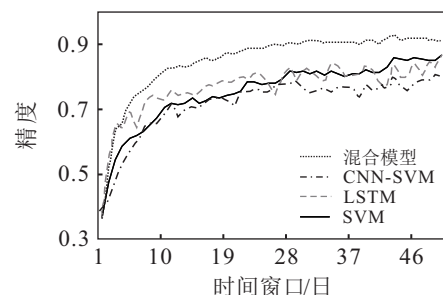


图7 TN-LP-LSTM-SVM与其他3种模型的预测精度

表5 不同的时间窗口下TN-LP-LSTM-SVM与CNN-SVM、LSTM和SVM的预测精度比较

时间窗口	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
TN-LP-LSTM-SVM	0.362	0.509	<b>0.635</b>	<b>0.664</b>	<b>0.719</b>	<b>0.748</b>	<b>0.757</b>	<b>0.779</b>	<b>0.797</b>	<b>0.821</b>	<b>0.825</b>	<b>0.833</b>	<b>0.830</b>	<b>0.837</b>	<b>0.853</b>	<b>0.858</b>	<b>0.853</b>
CNN-SVM	<b>0.396</b>	0.421	0.488	0.532	0.574	0.604	0.642	0.659	0.662	0.685	0.717	0.683	0.703	0.709	0.712	0.726	0.736
LSTM	0.378	<b>0.551</b>	<b>0.662</b>	0.650	0.702	0.645	0.716	0.747	0.741	0.736	0.763	0.748	0.756	0.754	0.750	0.782	0.776
SVM	0.369	0.473	0.544	0.586	0.611	0.618	0.634	0.653	0.674	0.705	0.720	0.716	0.721	0.734	0.720	0.727	0.738
时间窗口	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
TN-LP-LSTM-SVM	<b>0.862</b>	<b>0.876</b>	<b>0.871</b>	<b>0.880</b>	<b>0.889</b>	<b>0.886</b>	<b>0.888</b>	<b>0.896</b>	<b>0.903</b>	<b>0.898</b>	<b>0.910</b>	<b>0.901</b>	<b>0.909</b>	<b>0.911</b>	<b>0.913</b>	<b>0.905</b>	<b>0.909</b>
CNN-SVM	0.738	0.730	0.715	0.759	0.762	0.765	0.759	0.762	0.775	0.780	0.782	0.791	0.768	0.756	0.761	0.769	0.755
LSTM	0.792	0.789	0.793	0.807	0.803	0.817	0.809	0.785	0.751	0.804	0.834	0.793	0.828	0.827	0.808	0.795	0.847
SVM	0.734	0.743	0.748	0.755	0.787	0.786	0.778	0.781	0.779	0.794	0.817	0.814	0.817	0.812	0.817	0.806	0.817
时间窗口	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	
TN-LP-LSTM-SVM	<b>0.911</b>	<b>0.908</b>	<b>0.910</b>	<b>0.901</b>	<b>0.915</b>	<b>0.914</b>	<b>0.902</b>	<b>0.921</b>	<b>0.930</b>	<b>0.913</b>	<b>0.920</b>	<b>0.918</b>	<b>0.923</b>	<b>0.917</b>	<b>0.915</b>	<b>0.913</b>	
CNN-SVM	0.773	0.772	0.770	0.740	0.779	0.761	0.772	0.767	0.800	0.776	0.761	0.780	0.795	0.795	0.808	0.801	
LSTM	0.845	0.811	0.813	0.810	0.839	0.826	0.788	0.783	0.842	0.806	0.800	0.847	0.811	0.849	0.838	0.872	
SVM	0.811	0.803	0.812	0.812	0.820	0.812	0.816	0.830	0.859	0.854	0.853	0.858	0.857	0.851	0.850	0.867	

#### 4 结 论

本文提出了一种新的混合模型,该模型在大于3天的中长期预测中,稳健地领先于CNN-SVM、LSTM和SVM。基于本文提出的混合模型,投资者可以根据自身的风险态度进行投资决策。本文结合当下流行的方法(复杂网络分析、链接预测、深度神经网络和机器学习)对WTI期货价格方向预测问题进行了探索,未来的研究可以从以下两个方向展开。

本文在建立转移网络后,通过网络拓扑特征分析和链接预测方法提取能帮助预测WTI期货价格变化方向和波动率大小的信息。在未来的研究中,可以尝试使用图嵌入算法将节点映射到向量空间并保留有关节点关系的信息,整合所有节点的向量可以得到机器学习分类器所需的特征向量。图嵌入算法降维时信息损失较少,得到的混合模型可能会实现较高的预测精度。

与股票指数等金融资产相比,WTI期货具有不平稳、高波动的特征,在未来的研究中,如果在模型中加入跳跃和结构断点,可能会实现更高的预测精度。

#### 参考文献(References)

[1] Hamilton J D. Oil and the macroeconomy since world war II[J]. *Journal of Political Economy*, 1983, 91(2): 228-248.  
 [2] Hendrix C S. Oil prices and interstate conflict[J]. *Conflict Management and Peace Science*, 2017, 34(6): 575-596.  
 [3] Sadorsky P. Modeling and forecasting petroleum futures

volatility[J]. *Energy Economics*, 2006, 28(4): 467-488.  
 [4] Kang S H, Kang S M, Yoon S M. Forecasting volatility of crude oil markets[J]. *Energy Economics*, 2009, 31(1): 119-125.  
 [5] Mohammadi H, Su L X. International evidence on crude oil price dynamics: Applications of ARIMA-GARCH models[J]. *Energy Economics*, 2010, 32(5): 1001-1008.  
 [6] Hou A J, Suardi S. A nonparametric GARCH model of crude oil price return volatility[J]. *Energy Economics*, 2012, 34(2): 618-626.  
 [7] Klein T, Walther T. Oil price volatility forecast with mixture memory GARCH[J]. *Energy Economics*, 2016, 58: 46-58.  
 [8] Wen F H, Xiao J H, Huang C X, et al. Interaction between oil and US dollar exchange rate: Nonlinear causality, time-varying influence and structural breaks in volatility[J]. *Applied Economics*, 2018, 50(3): 319-334.  
 [9] Saghaian S, Nemati M, Walters C, et al. Asymmetric price volatility transmission between US biofuel, corn, and oil markets[J]. *Journal of Agricultural and Resource Economics*, 2018, 43(1): 46-60.  
 [10] Singhal S, Ghosh S. Returns and volatility linkages between international crude oil price, metal and other stock indices in India: Evidence from VAR-DCC-GARCH models[J]. *Resources Policy*, 2016, 50: 276-288.  
 [11] Askari H, Krichene N. Oil price dynamics (2002 ~ 2006)[J]. *Energy Economics*, 2008, 30(5): 2134-2153.  
 [12] Wen F H, Gong X, Cai S H. Forecasting the volatility of

- crude oil futures using HAR-type models with structural breaks[J]. *Energy Economics*, 2016, 59: 400-413.
- [13] Xie W, Yu L A, Xu S Y, et al. A new method for crude oil price forecasting based on support vector machines[C]. *International Conference on Computational Science*. Berlin, Heidelberg: Springer, 2006: 444-451.
- [14] Chiroma H, Abdulkareem S, Herawan T. Evolutionary neural network model for West Texas Intermediate crude oil price prediction[J]. *Applied Energy*, 2015, 142: 266-273.
- [15] Liu J L, Bai Y, Li B. A new approach to forecast crude oil price based on fuzzy neural network[C]. *The 4th International Conference on Fuzzy Systems and Knowledge Discovery*. Haikou, 2007: 273-277.
- [16] Yu L A, Wang S Y, Lai K K. Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm[J]. *Energy Economics*, 2008, 30(5): 2623-2635.
- [17] Wu Y X, Wu Q B, Zhu J Q. Improved EEMD-based crude oil price forecasting using LSTM networks[J]. *Physica A: Statistical Mechanics and its Applications*, 2019, 516: 114-124.
- [18] Jammazi R, Aloui C. Crude oil price forecasting: Experimental evidence from wavelet decomposition and neural network modeling[J]. *Energy Economics*, 2012, 34(3): 828-841.
- [19] Wang F, Lin J Q, Wei L J. An AI model for oil volatility forecasting[J]. *IEEE Intelligent Systems*, 2020, 35(3): 62-70.
- [20] Zhang J L, Zhang Y J, Zhang L. A novel hybrid method for crude oil price forecasting[J]. *Energy Economics*, 2015, 49: 649-659.
- [21] Li J C, Zhu S W, Wu Q Q. Monthly crude oil spot price forecasting using variational mode decomposition[J]. *Energy Economics*, 2019, 83: 240-253.
- [22] Abdollahi H. A novel hybrid model for forecasting crude oil price based on time series decomposition[J]. *Applied Energy*, 2020, 267: 115035.
- [23] Safari A, Davallou M. Oil price forecasting using a hybrid model[J]. *Energy*, 2018, 148: 49-58.
- [24] Wang M G, Zhao L F, Du R J, et al. A novel hybrid method of forecasting crude oil prices using complex network science and artificial intelligence algorithms[J]. *Applied Energy*, 2018, 220: 480-495.
- [25] Kristjanpoller W, Minutolo M C. Forecasting volatility of oil price using an artificial neural network-GARCH model[J]. *Expert Systems with Applications*, 2016, 65: 233-241.
- [26] Parkinson M. The extreme value method for estimating the variance of the rate of return[J]. *The Journal of Business*, 1980, 53(1): 61.
- [27] Liben-Nowell D, Kleinberg J. The link-prediction problem for social networks[J]. *Journal of the American Society for Information Science and Technology*, 2007, 58(7): 1019-1031.
- [28] Cortes C, Vapnik V. Support-vector networks[J]. *Machine Learning*, 1995, 20(3): 273-297.
- [29] Niu X X, Suen C Y. A novel hybrid CNN-SVM classifier for recognizing handwritten digits[J]. *Pattern Recognition*, 2012, 45(4): 1318-1325.

### 作者简介

赵戈雅(1994—),女,博士生,从事环境金融、机器学习的研究, E-mail: 861215403@qq.com;

薛明皋(1965—),男,教授,博士生导师,从事资产定价、环境金融等研究, E-mail: xmgwt@hust.edu.cn.

(责任编辑:孙艺红)