



计算机工程与应用  
Computer Engineering and Applications  
ISSN 1002-8331, CN 11-2127/TP

## 《计算机工程与应用》网络首发论文

题目: 基于深度迁移学习的多尺度股票预测  
作者: 程孟菲, 高淑萍  
网络首发日期: 2022-03-15  
引用格式: 程孟菲, 高淑萍. 基于深度迁移学习的多尺度股票预测[J/OL]. 计算机工程与应用. <https://kns.cnki.net/kcms/detail/11.2127.TP.20220314.1331.008.html>



**网络首发:** 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

**出版确认:** 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

# 基于深度迁移学习的多尺度股票预测

程孟菲，高淑萍

西安电子科技大学 数学与统计学院，西安 710126

**摘要：**股票市场不仅是上市公司的重要融资渠道，也是重要的投资市场，股票预测一直受到人们的关注。为了充分利用来自不同股票价格的信息，提高股票的预测效果，本文提出一种多尺度股票价格预测模型 TL-EMD-LSTM-MA(TELM)。TELM 模型首先通过经验模态分解将收盘价分解为多个时间尺度分量，不同时间尺度分量震荡频率不同，反映了不同的周期性信息；然后根据分量的震荡频率选择不同方法进行预测，高频分量利用深度迁移学习的方法训练堆叠 LSTM，低频分量利用移动平均法进行预测；最后将所有分量的预测值相加作为收盘价的最终预测输出。通过深度迁移学习训练的堆叠 LSTM，包含来自不同股票的信息，具备更多行业或市场的知识，能有效降低预测误差。利用移动平均法预测低频分量，更有效捕获股票的总体趋势。对中国 A 股市场内 500 支股票以及上证指数、深证成指等指数进行预测，结果表明，与其它模型相比，TELM 预测误差最低，拟合优度最高。根据 TELM 预测的股票收盘价模拟股票交易过程，结果表明 TELM 投资风险低、收益高。

**关键词：**股票预测；深度迁移学习；经验模态分解；长短期记忆网络；移动平均方法

**文献标志码：**A **中图分类号：**TP183 **doi：**10.3778/j.issn.1002-8331.2112-0501

## Multi-Scale Stock Prediction Based on Deep Transfer Learning

CHENG Mengfei, GAO Shuping

School of Mathematics and Statistics, Xidian University, Xi'an 710126, China

**Abstract:** The stock market is not only an important financing channel for listed companies, but also an important investment market. Stock prediction have always attracted people's attention. In order to make full use of the information from different stock data and improve the prediction performance of stocks, a multi-scale stock price prediction model TL-EMD-LSTM-MA (TELM) is proposed. Firstly, TELM decomposes the closing price into multiple time-scale components through empirical mode decomposition. Different time-scale components have different oscillation frequencies, reflecting different periodic information. Secondly, different components are predicted by different method according to oscillation frequencies. High-frequency component are predicted by the stacked LSTM which is trained by deep transfer learning method. Low-frequency component by the moving average method. Finally, the predicted closing price is calculated by summing the predicted values of all components. The stacked LSTM trained by deep transfer learning contains information from different stocks and has more industry or market knowledge, which can effectively reduce the predictive error. Using the moving average method to predict the low-frequency component can more effectively capture the general trend of stocks. 500 stocks in China's A-share market, Shanghai stock index, SZSE component index, and other stock indexes are predicted. The results show that the error of TELM is the lowest and the goodness of fit is the highest compared with other models. By joining in the stock predicted closing price of TELM, the stock trading process is simulated. The results show that TELM has low investment risk and high return.

**Key words:** stock prediction; deep transfer learning; empirical mode decomposition; long short-term memory; moving average

**基金项目：**国家自然科学基金(91338115)；高等学校学科创新引智基地“111”计划(B08038)。

**作者简介：**程孟菲（1995-），女，硕士。研究方向为数据融合、大数据分析与管理、时间序列预测；高淑萍（1963-），通讯作者，女，博士，教授。研究方向为多目标优化理论及应用，大数据分析与管理，E-mail: gaosp@mail.xidian.edu.cn。

股票是整个国民经济的重要组成部分,与国家经济发展息息相关。股票的选择和投资策略是股票交易的核心,两者都需要分析未来股票走势。股市每天都会产生大量的数据,对中小投资者来说,很难从大量的数据中分析股票的未來趋势并进行决策。

股票数据具有非线性、不稳定的特征,数据量大,同时具有很高的噪声<sup>[1]</sup>。股票价格的变动受到经济、政策、市场和投资者心理等多重因素的影响,增加了股票预测的难度。随机游走假设的提出,为能否利用历史信息预测信息提出挑战<sup>[2]</sup>。随机游走假设认为,市场上股票的价格已经反映了所有的信息,所以余下的价格波动都是源于无法预测和捕捉的随机事件,因此股票的价格波动是一种随机游走的形态。随着技术的进步,对股票预测问题的研究有了很大的发展。很多学者通过对非线性、不稳定的股票数据进行平稳化或降噪处理,预测股票未来走势,如娄磊等人<sup>[3]</sup>利用小波对股票进行降噪,贺毅岳<sup>[4]</sup>通过分解对数据进行平稳化处理。

股票价格预测模型可分为传统模型、机器学习模型、深度学习模型。传统回归模型如多项式回归、线性回归、差分移动自回归模型(Autoregressive Integrated Moving Average, ARIMA)<sup>[5]</sup>等,但是难以捕获股票历史数据和未来价格之间的关系,预测误差大。机器学习模型可以捕获人难以发现的内在关系,股票预测常用的机器学习模型有支持向量机<sup>[6]</sup>、随机森林<sup>[7]</sup>、K 临近值<sup>[8]</sup>等。深度学习是机器学习的进一步发展,常见的深度学习模型有卷积神经网络(Convolutional Neural Networks, CNN)、循环神经网络(Recurrent Neural Net-work, RNN)、长短期记忆网络(Long Short Term Memory Network, LSTM)。其中由于 LSTM 解决了模型训练过程中 RNN 的梯度消失以及梯度爆炸的缺陷,被广泛用于股票预测问题的研究中<sup>[9, 10]</sup>。Zhao 等人<sup>[11]</sup>重新定义股票趋势预测的时间权重 LSTM 模型, Mootha 等人<sup>[12]</sup>使用双向 LSTM 对印度股票进行预测, Althelaya 等人<sup>[13]</sup>采用双向 LSTM 和堆叠 LSTM 对股票进行预测, Qian 等人<sup>[14]</sup>在引入静态分析,验证 LSTM 的预测效果与数据的非平稳性之间的关系。Li 等人<sup>[15]</sup>提出自适应的梯度下降算法训练 CNN-LSTM 模型,预测效果优于传统的梯度下降算法。Nguyen 等人<sup>[16]</sup>提出 DTRSI 模型,利用深度迁移学习训练 LSTM 网络,有效提高了股票的预测效果。经验模态分解(Empirical Mode Decomposition, EMD)将时间序列数据分解为一组与不同时间尺度相关的分量和趋势,即本征模函数(Intrinsic

Mode Function, IMF)和残差(Residual, RES),其中 IMF 反映了在不同时间尺度下数据的震荡情况,RES 则反映了市场趋势。张倩玉<sup>[17]</sup>利用 EMD 提出 EMD-LSTM 模型,表明 EMD 能有效提高不稳定、高噪声数据的预测效果。

虽然基于深度学习的模型提高了股票的预测效果,但是现有的股票预测方法主要利用了股票的历史数据而忽略了来自其它股票的有效信息,股票价格的涨跌受到其它股票的影响,某一支或多支股票价格的涨跌会影响相关行业內其它股票价格的变化<sup>[18]</sup>。为了进一步提高股票预测效果,本文提出了一种多尺度股票价格预测模型 TL-EMD-LSTM-MA(TELM),预测股票的收盘价。TELM 首先通过 EMD 将收盘价分解多个尺度分量,即多个 IMF 和 RES;然后,将多个 IMF 和 RES 分为高频部分和低频部分,针对高频部分,采用基于深度迁移学习的方法进行预测,即先利用市场內其它股票数据训练预训练模型,再利用待预测股票或指数的对应分量对模型进行微调 and 预测,针对低频部分采用移动平均法(Moving Average, MA)预测;最后,将所有分量的预测值相加作为最终的预测输出。为了验证 TELM 的性能,分别对单支股票和上证指数、深证成指、上证 50、中证 500、沪深 300 进行预测。其中单支股票包括中国 A 股市场 20 个行业的 500 支股票,并分别从行业、上市时间等因素对实验结果进行分析。实验结果表明 TELM 在单支股票和指数的预测上比 MA、LSTM、CNN-LSTM、EMD-LSTM、DTRSI 效果更好,降低了模型的训练规模,节约了时间成本,有较好的泛化能力。进一步通过模拟股票的交易过程,表明 TELM 预测的股票收盘价对股民的投资决策有重要的参考价值,提高了投资回报率,降低了损失。

## 1 相关理论知识

### 1.1 经验模态分解

任何非线性、非平稳的信号都可以通过经验模态分解<sup>[19]</sup>(EMD)将原始信号分解为多个本征模函数(IMF)和残差(RES)。EMD 在处理数据的过程中,不需要设定基函数,根据数据自身的尺度特征进行分解。IMF 必须满足两个条件:

- 1) 在时间范围内,局部极值点和过零点的数目必须相等或相差 1;
- 2) 在任意时刻,局部最大值的上包络线和局部最小值得下包络线均值必须为 0。

假设原始信号为  $S(t)$ ，EMD 的分解过程如下：

**Step1:** 确定原始序列  $S(t)$  的所有极大值和极小值；

**Step2:** 采用 3 次样条插值法，根据极大值与极小值构造  $S(t)$  的上下包络线；

**Step3:** 由上下包络线计算出  $S(t)$  的局部均值  $m_i(t)$ ，以及差值  $S(t)$  与  $m_i(t)$  差值：

$$h_{i1}(t) = S(t) - m_i(t) \quad (1)$$

**Step4:** 用  $h_{i1}(t)$  代替原始序列  $S(t)$ ，重复步骤 Step1-Step3，直到  $h_{i1}(t)$  和  $h_{i-1}(t)$  的差值小于阈值，

则第一个 IMF 分量即为  $h_{i1}(t)$ ，记作

$$IMF_1(t) = h_{i1}(t), \quad r_1(t) = S(t) - IMF_1(t),$$

$$S(t) = r_1(t).$$

**Step5:** 重复 Step1-Step4，直到  $r_n(t)$  不能再分解出 IMF 分量。其中  $S(t)$  的分解结果如下表示：

$$S(t) = \sum_{i=1}^n IMF_i(t) + RES \quad (2)$$

这里  $RES = r_n(t)$  为趋势项， $i$  越小， $IMF_i(t)$  单位时间内震荡幅度越大。

## 1.2 深度迁移学习

迁移学习把知识从一个或多个领域（源域）迁移到另一个领域（目标域），利用已知某领域的知识求解不同但相关领域的问题<sup>[20, 21]</sup>。已经在图像分类<sup>[22]</sup>、自然语言处理<sup>[23]</sup>等领域有很多应用，受到越来越多的关注。与机器学习、深度学习相比，迁移学习不再假设训练数据和测试数据必须是独立同分布，因此目标模型的训练不需要从头开始训练，降低了对训练数据的要求，加快训练速度。Chuanqi Tan 等人<sup>[24]</sup>给出了迁移学习和深度迁移学习的定义。

**定义1: (迁移学习)。** 给定一个基于数据  $D_t$  的学习任务  $T_t$ ，以及基于数据  $D_s$  的学习任务  $T_s$ 。发现  $D_s$  和  $T_s$  中的隐知识，并用来提高任务  $T_t$  的函数  $F_T(\cdot)$  效果的任务，称为迁移学习。其中  $D_t \neq D_s$ ， $T_t \neq T_s$ ，且  $D_s$  的数据量远大于  $D_t$ 。

**定义 2: (深度迁移学习)。** 给定一个  $\langle D_s, T_s, D_t, T_t, F_T(\cdot) \rangle$  的迁移学习任务，如果  $F_T(\cdot)$  是基于深度学习的非线性函数，则称该任务为深度迁移学习。

## 1.3 长短期记忆网络

为了解决 RNN 在训练过程中梯度消失和梯度爆炸的问题，Hochreiter 等人<sup>[25]</sup>提出了 LSTM 模型，解决了传统神经网络长期和短期依赖的问题。LSTM 模型使用长短期记忆细胞记录长短期信息，其结构单元

如图 1 所示，可以概括为三门两态：输入门  $i_t$ 、输出门  $o_t$  和遗忘门  $f_t$ ，以及单元状态  $c_t$  和隐藏状态  $h_t$ 。

输入门：用于记忆当前输入数据的信息，包括 sigmoid 层和 tanh 层。sigmoid 层决定那些信息将被更新，tanh 层用于创建新的变量  $\tilde{c}_t$ 。

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (3)$$

$$\tilde{c}_t = \tanh(W_c[h_{t-1}, x_t] + b_c) \quad (4)$$

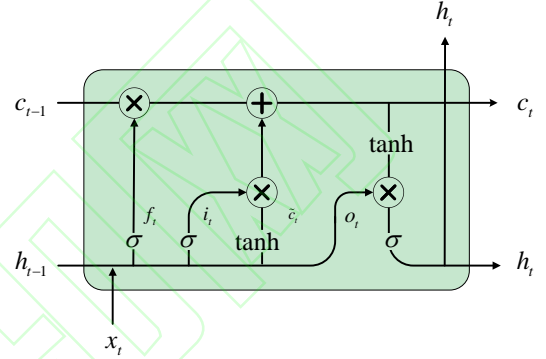


图1 LSTM结构

Fig.1 Structure of LSTM

遗忘门：用于选择过去的哪些信息将被遗忘，1 表示完全保留历史信息，0 表示完全遗忘历史信息。

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (5)$$

记忆单元：记忆单元用于整合过去和现在的信息。

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (6)$$

输出门：输出门决定输出的信息。输出门得到原始输出值  $o_t$ ，进一步得到的隐藏状态  $h_t$ ，并将  $h_t$  作为该时刻的输出。

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (7)$$

$$h_t = o_t \odot \tanh(c_t) \quad (8)$$

激活函数  $\sigma$  和  $\tanh$  计算公式为：

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (9)$$

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (10)$$

其中  $[h_{t-1}, x_t]$  表示将  $h_{t-1}$  和  $x_t$  进行拼接， $\odot$  表示矩阵的哈达玛积， $W_f$ 、 $W_i$ 、 $W_c$ 、 $W_o$  为权重矩阵， $b_f$ 、 $b_i$ 、 $b_c$ 、 $b_o$  为偏置，权重矩阵和偏置将在模型的训练过程中不断优化。

## 1.4 移动平均方法

Joseph E. Granville 提出均线理论<sup>[26]</sup>，被用来对股市的趋势方向进行追踪，虽然有一定的滞后性，但能真实反应市场情况。均线是由移动平均法（MA）计算前  $N$  个数据的平均值。对一维向量  $x$ ，预测向量  $y$  的计算公式如下：





图2 TELM模型流程图

Fig.2 TELM Model's Flow Chart

TELM 模型分两步进行分别进行预训练模型的训练和单支股票或指数的模型训练与预测, 步骤如下:

**Step1: 预训练模型的训练。**预训练数据集的 50 支股票, 每支股票的 IMF 的个数  $I_n \geq 7$ , 即高频部分的分量包括  $IMF_1, IMF_2, \dots, IMF_5$ , 分别训练对应的预训练模型  $model_1, model_2, \dots, model_5$ 。针对  $IMF_i, i=1, 2, \dots, 5$ , 训练堆叠双层 LSTM 的深度学习模型  $model_i$ , 该模型结构如图 3 所示, 输入的数据包括预处理后的第  $j$  至第  $j+time\_step-1$  天的  $IMF_i$ 、开盘价 (close)、收盘价 (open)、最高价 (max)、最低价 (min)、成交量 (vol) 和成交额 (amount), 输出数据  $\hat{y}_j$  为  $IMF_i$  在第  $j+time\_step$  天的预测值,  $time\_step$  为时间步长。

**Step2: 单支股票或指数的预测。**对高频部分和低频部分采用不同的方法进行预测, 并将所有分量预测值相加。

**Step2.1: 高频部分的预测。**针对  $IMF_i, i=1, 2, \dots, h$ , 输入数据通过预处理后, 将其划分为训练集、验证集和测试集。利用训练集微调  $model_i$  部分参数, 并对验证集进行预测, 得到预测值  $\hat{y}^{(1)}$ ; 利用  $model_i$  直接对验证集

Fig.3 Two-layer LSTM Model's Flow Chart

进行预测, 得到验证集的预测值  $\hat{y}^{(2)}$ ; 通过分别比较并  $\hat{y}^{(1)}$ 、 $\hat{y}^{(2)}$  与测值真实值的误差进行模型选择, 判断  $model_i$  微调后预测误差是否变小, 如果  $\hat{y}^{(1)}$  与真实值的误差更小, 则利用微调后的模型对测试集进行预测, 否则直接使用预训练模型  $model_i$  进行测试集的预测。最终得到

$IMF_1, IMF_2, \dots, IMF_h$  的测试集预测值  $\hat{H}_1, \hat{H}_2, \dots, \hat{H}_h$ 。

**Step2.2: 低频部分的预测。**计算分量和  $IMF_l$ , 计算公式如下:

$$IMF_l = \sum_{i=h+1}^{I_n} IMF_i + RES \quad (14)$$

并利用移动平均法计算测试集的预测值  $\hat{L}$ , 即

$$\hat{L} = MA(IMF_l) \quad (15)$$

**Step2.3:** 最后将  $\hat{H}_1, \hat{H}_2, \dots, \hat{H}_h$  和  $\hat{L}$  的预测值进行相加, 结果作为最终的股票预测值  $\hat{Y}$ , 计算公式如下:

$$\hat{Y} = \sum_{i=1}^h \hat{H}_i + \hat{L} \quad (16)$$

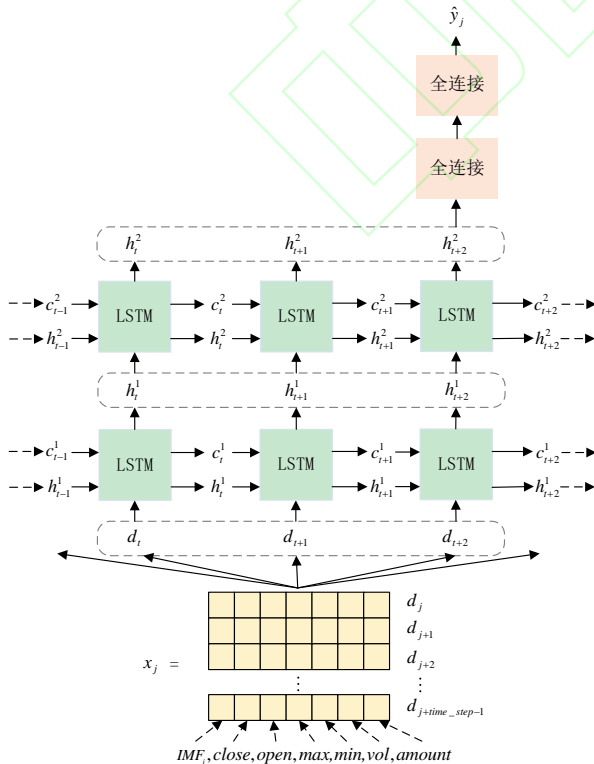


图3 堆叠LSTM流程图

### 3 实验结果及分析

本节分别对数据及模型参数、单支股票预测结果、指数预测结果、股票交易模拟进行分析。数据及模型参数是对数据和参数设置进行说明; 单支股票预测结果分析除了比较不同模型的性能外, 还从行业、IMF 个数、上市时间等方面分析预测误差; 指数预测结果分析旨在验证 TELM 是否能降低指数预测的误差, 提高预测效果; 股票交易模拟则是基于股票的预测收盘价模拟股票交易过程, 验证 TELM 在股市交易中是否能降低投资风险、提高投资回报率。

#### 3.1 数据及模型参数

##### 3.1.1 数据来源

实验所用股票数据来自财经数据接口包 Tushare (<https://www.tushare.pro>), 包含 20 个行业的 550 支股票以及上证指数、深证成指、上证 50、中证 500、沪深 300 从 2010 年 11 月 1 日至 2021 年 11 月 1 日的历史行情信息。预训练数据集包括来自酿酒、新能源车、有色金属、环保工程 4 个行业共 50 支股票数据, 用于 Step1 中预训练模型, 每支股票的上市时间都在 5 个月以上; 预测数据集包括 500 支股票和上证指数、

深证成指、上证 50、中证 500、沪深 300，单支股票或指数的数据被划分为训练集、验证集、测试集，分别占总数据的 80%、10%、10%，用于 Step2 中模型的训练、选择和预测，并验证模型的性能。

### 3.1.2 数据预处理

为了消除不同特征数据量纲的影响，在训练前，对数据进行 Min-Max 归一化，对归一化后的数据根据时间步长  $time\_step$  切分数据。

Min-Max 归一化的计算公式如下：

$$y_j = \frac{x_j - \min\{x_i\}}{\max\{x_i\} - \min\{x_i\}} \quad (17)$$

模型输出的预测值需要进行反归一化：

$$\tilde{x}_j = (\max\{x_i\} - \min\{x_i\}) \cdot \hat{y}_j + \min\{x_i\} \quad (18)$$

其中  $x_i$  为需要归一化的数据， $y_i$  为归一化后的数据， $\hat{y}_j$  为预测数据， $\tilde{x}_j$  为反归一化后的数据。

### 3.1.3 性能指标

采用均方根误差 (Root Mean Squared Error, RMSE)、平均绝对误差 (Mean Absolute Error, MAE) 和决定系数  $R^2$  评估模型的性能，其中  $R^2$  评价回归模型系数拟合优度， $R^2$  越高，表示可以被解释的程度越高，回归模型的效果越好。评价指标的计算公式如

下：

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (19)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (20)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (21)$$

$\hat{y}_i, y_i$  分别代表预测值和真实值， $\bar{y}$  表示  $y_i$  的均值。

### 3.1.4 参数设置

在 TELM 的训练和预测过程中涉及许多超参数，超参数的设置影响着最终的预测效果，超参数设置如表 1 至表 3 所示。

通过网格搜索算法<sup>[27]</sup>，寻找最优的双层 LSTM 模型的参数，参数设置如表 1 所示，其中  $out\_dim$  表示该层的输出参数， $activation$  表示该层的激活函数， $batch\_size$  表示每轮训练的样本个数， $time\_step$  表示时间步长， $optimizer$  表示模型训练使用的优化器，损失函数为 MSE。

表1 堆叠LSTM模型的参数

Table 1 Parameters of the stacked LSTM Network

层	类型	参数值	训练参数
0	Input	——	$batch\_size = 20$
1	LSTM	$out\_dim=30$	$time\_step = 5$
2	LSTM	$out\_dim=30$	$optimizer = Adam$
3	Flatten	——	$lr=0.001$
4	Dense	$out\_dim=10, activation="sigmoid"$	$decay=2 \times 10^{-9}$
5	Dense	$out\_dim=1, activation="sigmoid"$	$loss = MSE$

在 Step2.1 预测过程中涉及到模型的微调，微调模型哪些层的参数对模型的训练效果以及模型的训练的时间成本有很大的影响，为了确定最优的微调参数，分别选择不同层的参数进行对比实验，实验结果如表 2 所示。从表 2 中可以看出微调第 4、5 层的参数要优于微调其它层参数，在降低预测误差的同时，训练参数由 12201 降至 321，减少了 97.4%。因此在 Step2.1 预测高频部分的分量时，将第 4、5 层的参数设为可微调的层。

TELM 模型涉及到高频部分包含的分量数量  $h$ ，

$h$  的最大值设为 5， $h$  越大表明预训练模型的数量越多，训练时间越长。由于股票上市时间不同，以及每支股票数据的复杂程度不同，导致某些股票经过 EMD 分解后得到的 IMF 分量个数不同，因此必须确定  $h$  的最大值。 $h$  取不同值时的预测性能如表 3 所示， $h$  越大，预测误差越低，预测结果越好，在  $h=5$  时，预测误差最低，且拟合度最好。在预测数据集中 94% 的股票的 IMF 分量个数  $I_n$  大于等于 5。

表2 微调的层及其性能指标

Table 2 Fine-tuning Layers and Their Performance Indicators

微调 的层	参数量	RMSE	MAE	$R^2$
----------	-----	------	-----	-------

1 层	4,560	0.0496	0.0367	0.9515
2 层	7,320	0.0498	0.0369	0.9505
1, 2 层	11,880	0.0496	0.0366	0.9516
<b>4, 5 层</b>	<b>321</b>	<b>0.0495</b>	<b>0.0365</b>	<b>0.9553</b>

表3  $h$  取值及其性能指标

Table 3 Performance Indicators of Different Value of $h$				
$h$	RMSE	MAE	$R^2$	
1	0.0821	0.0602	0.8719	
2	0.0658	0.0499	0.9207	
3	0.0557	0.0423	0.9445	
4	0.0515	0.0386	0.9522	
<b>5</b>	<b>0.0495</b>	<b>0.0365</b>	<b>0.9553</b>	
6	0.0512	0.0376	0.9495	

### 3.2 单支股票预测结果分析

#### 3.2.1 不同模型的性能比较

为了验证 TELM 的性能,对预测数据集的 500 支股票通过不同的模型进行预测,包括 MA、LSTM、CNN-LSTM、EMD-LSTM 以及迁移学习模型 DTRSI 进行对比,不同模型的性能指标如表 4。从表 4 中可以看出 EMD-LSTM、DTRSI 以及 TELM 的预测结果

显示, RMSE 和 MAE 误差低于 LSTM 和 CNN-LSTM,  $R^2$  数值也远高于 LSTM 和 CNN-LSTM, EMD-LSTM 和 TELM 对股票数据进行了多尺度分解,说明多尺度分解能有效降低股票的预测误差,而 DTRSI 和 TELM 的预测效果则说明迁移学习同样提高股票预测的拟合效果。表 4 中股票个数一项可以看出,在 500 支股票中,TELM 在 428 支股票的预测任务中表现最优,占总量的 85%,说明 TELM 的泛化能力强,在股票的批量预测任务中具有优势。

表4 六种模型的预测性能指标

Table 4 Performance Indicators of Six Models

模型	RMSE	MAE	$R^2$	股票个数
<b>TELM</b>	<b>0.0686</b>	<b>0.0530</b>	<b>0.9014</b>	<b>428</b>
DTRSI	0.0945	0.0699	0.8227	34
EMD-LSTM	0.1007	0.0779	0.8096	35
LSTM	0.1433	0.1128	0.6400	0
MA	0.1624	0.1257	0.5320	0
CNN-LSTM	0.1711	0.1368	0.4496	3

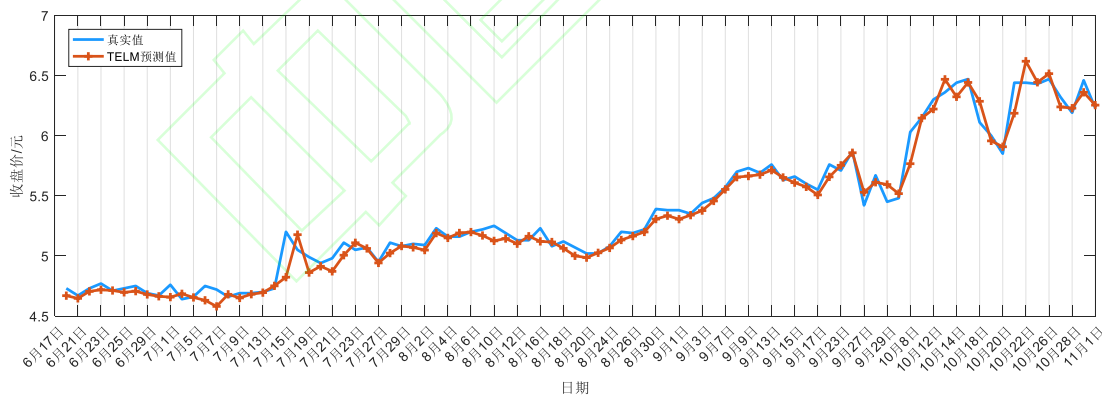


图4 三元股份收盘价预测曲线

Fig.4 Sanyuan Shares Closing Price Forecast Curve

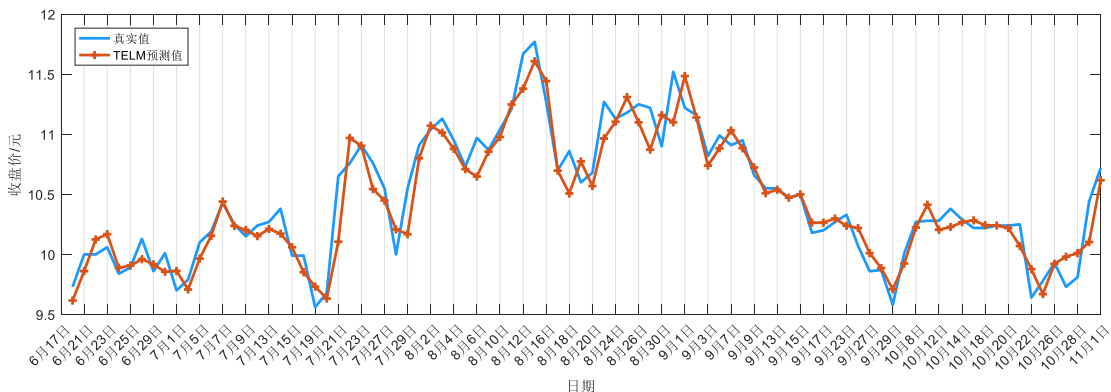




图5 圣龙股份收盘价预测曲线  
Fig.5 Shenglong Shares Closing Price Forecast Curve

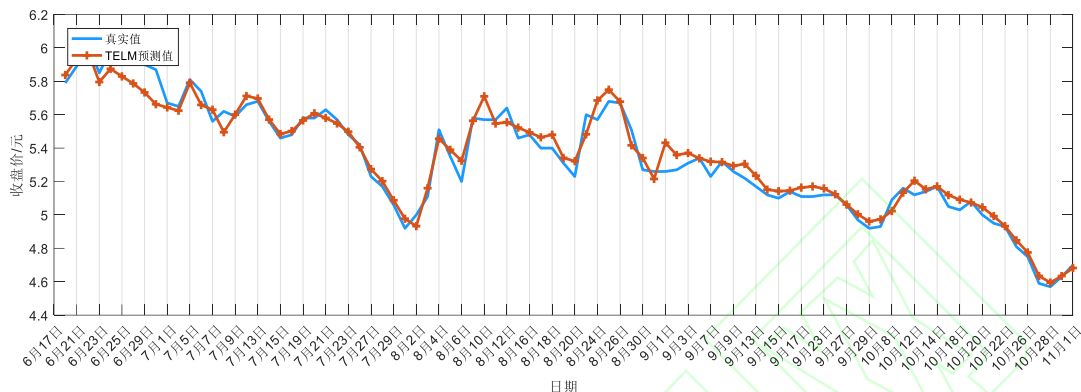


图6 西王食品收盘价预测曲线  
Fig.6 Xiwang Foods Closing Price Forecast Curve

图 4、图 5、图 6 分别为三元股份、圣龙股份、西王食品 3 支股票 2021 年 6 月 17 日至 2021 年 11 月 1 日的预测曲线图,从曲线中可以看出,图 4 总体呈现上涨趋势,图 5 股票总体先涨后跌,图 6 总体呈下跌趋势,从图 4 至图 6 可以看出 TELM 模型不仅能捕获股票未来趋势,而且在股票的总体走势变化较大的拐点预测准确,如图 4 中 8 月 20、10 月 8 日。提前预测出股票趋势变化幅度较大的点,为投资者的投资决策提出重要的依据,能有效提高股民的收益和降低投资风险。

### 3.2.2 行业因素、上市时间、IMF 数量分析

由于在预训练数据集包括酿酒、新能源车、有色金属、环保工程 4 个行业的股票数据,为了比较 TELM 模型的预测效果是否受行业因素的影响,本节对不同行业的预测性能指标进行对比,从表 5 中可以看出,酿酒、新能源车、有色金属、环保工程行业的 EMSE 误差为 0.044-0.060,除房地产和航运港口外,其它行业的股票预测误差大于 0.060;酿酒、新能源车、有色金属、环保工程行业的 MAE 误差则低于 0.039,其余行业的误差则高于 0.039;  $R^2$  的数值显示,除航运港口行业外,其它的行业的  $R^2$  数值都低于酿酒、新能源车、有色金属、环保工程。因此预训练数据集中是否包含该行业,对该行业的股票预测会有一定的影响。通过在预训练数据集中加入某行业的股票,能有效提高该行业内其它股票的预测效果。

由于上市时间不同,数据集的股票的数据量不同,为了分析股票的预测效果与股票数据量的关系,根据股票的数量大小划分为不同的区间,该区间内股

票的预测误差和拟合优度的平均值如表 6。从结果可以看出,当股票的数据量大于 500, RMSE、MAE 及  $R^2$  变化不大;当股票的数据量小于 500 时,性能指标较差。每年的交易日约为 250 天,所以上市两年以上的股票,比上市不足两年的股票的预测误差小、拟合优度高,预测效果更好。

表5 不同行业的性能指标

Table 5 Performance Indicators of Different Industrys

行业	RMSE	MAE	$R^2$
酿酒	0.0442	0.0332	0.9636
新能源车	0.0530	0.0389	0.9368
有色金属	0.0522	0.0382	0.9604
环保工程	0.0607	0.0448	0.9327
房地产	0.0562	0.0420	0.9345
农林牧渔	0.0652	0.0496	0.9136
银行	0.0638	0.0471	0.9280
汽车	0.0612	0.0463	0.9182
食品	0.0736	0.0563	0.8713
电子元件	0.0882	0.0691	0.8415
航运港口	0.0550	0.0411	0.9440
小家电	0.1310	0.1096	0.9090
基础建设	0.0803	0.0660	0.8300

表6 不同股票数据量区间的性能指标

Table 6 Performance Indicators of Different intervals of Stock Data Volume

股票数据量区间	RMSE	MAE	$R^2$
(0,500)	0.1688	0.1360	0.6319
(500,1000)	0.0555	0.0423	0.9393

(1000,2000)	0.0574	0.0436	0.9307
(2000,3000)	<b>0.0538</b>	<b>0.0407</b>	<b>0.9412</b>

利用 EMD 将股票数据分解为不同尺度分量，由于数据的复杂性不同，分解后的 IMF 数量不同，计算 IMF 数量为  $i$  的所有股票的 RMSE、MAE 及  $R^2$  均值， $i=1,2,\cdots,10$ ，图 7 中(a)、(b)分别为 RMSE、MAE

及  $R^2$  随 IMF 数量变化曲线图。从图中可以看出，当 IMF 数量大于等于 6 时，误差降低的速度和  $R^2$  增长速度变缓，说明当股票的 IMF 数量大于等于 6 时，IMF 分量数目对股票预测误差的影响较小，而 IMF 分量数量小于 6 时，股票的预测误差相对较高。因此 EMD 分解后 IMF 分量数目大于等于 6 的股票预测效果更好。

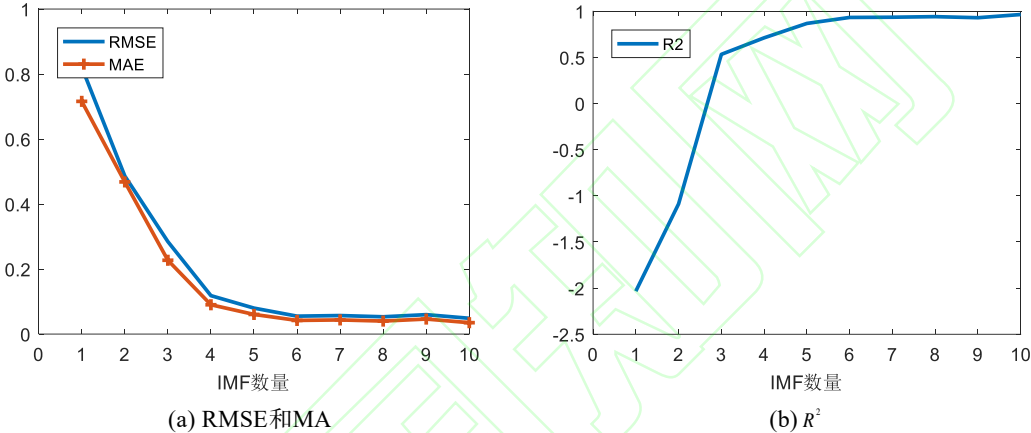


图7 RMSE、MA、 $R^2$  随IMF数量变化曲线图  
Fig.7 Change Curve With IMF of RMSE、MA、 $R^2$

3.3 指数预测结果分析

股票指数反映了股票市场的价格涨跌，投资者通过股票指数的变动情况预测股票市场未来的走势。本节对上证指数、深证成指、上证 50、中证 500、沪深 300 通过不同的模型对其进行预测，预测结果见表 7 至表 11。从表中可以看出，TELM 的预测效果最佳，其次是 EMD-LSTM 模型。其中 TELM 模型在上证 50 的预测中， $R^2$  值比 EMD-LSTM 提高了 5.6%，说明 TELM 能有效提高股票的指数的预测效果，更好地捕捉市场未来走势。

表7 上证指数的预测性能对比  
Table 7 Comparison of prediction performance of Shanghai Stock Index

模型	RMSE	MAE	$R^2$
<b>TELM</b>	<b>0.0487</b>	<b>0.0388</b>	<b>0.9317</b>
DTRSI	0.0913	0.0716	0.7600
EMD-LSTM	0.0519	0.0440	0.9224
LSTM	0.0937	0.0720	0.7474
MA	0.1263	0.0975	0.5410

CNN-LSTM	0.1211	0.0927	0.5780
----------	--------	--------	--------

表8 深证成指的预测性能对比  
Table 8 Comparison of prediction performance of SZSE Component Index

模型	RMSE	MAE	$R^2$
<b>TELM</b>	<b>0.0891</b>	<b>0.0704</b>	<b>0.8739</b>
DTRSI	0.1470	0.1168	0.6565
EMD-LSTM	0.0915	0.0726	0.8668
LSTM	0.1432	0.1104	0.6743
MA	0.2022	0.1515	0.3502
CNN-LSTM	0.1545	0.1168	0.6209

表9 上证50的预测性能对比  
Table 9 Comparison of prediction performance of SSE 50

模型	RMSE	MAE	$R^2$
<b>TELM</b>	<b>0.0464</b>	<b>0.0370</b>	<b>0.9627</b>
DTRSI	0.0834	0.0656	0.8794
EMD-LSTM	0.0713	0.0559	0.9119
LSTM	0.0792	0.0618	0.8915
MA	0.1050	0.0810	0.8091

CNN-LSTM 0.1066 0.0789 0.8033

表10 中证500的性能性能对比

Table 10 Comparison of prediction performance of CSI 500

模型	RMSE	MAE	$R^2$
<b>TELM</b>	<b>0.0433</b>	<b>0.0342</b>	<b>0.9662</b>
DTRSI	0.0686	0.0551	0.9151
EMD-LSTM	0.0472	0.0366	0.9598
LSTM	0.0705	0.0563	0.9103
MA	0.1041	0.0840	0.8041
CNN-LSTM	0.1011	0.0778	0.8155

表11 沪深300的预测性能对比

Table 11 Comparison of prediction performance of CSI 300

模型	RMSE	MAE	$R^2$
<b>TELM</b>	<b>0.0686</b>	<b>0.0542</b>	<b>0.9243</b>
DTRSI	0.1159	0.0915	0.7842
EMD-LSTM	0.0690	0.0575	0.9236
LSTM	0.1099	0.0863	0.8060
MA	0.1505	0.1105	0.6361
CNN-LSTM	0.1285	0.0937	0.7350

### 3.4 股票交易模拟

当选择一支股票进行投资时, 由于无法准确估计股票未来很长一段时间的涨跌情况, 难以确定股票的买入和卖出的时机。投资中收益越高意味着风险越高, 如何在一定风险的下获得最大收益是很多投资者关心的问题。本节模拟真实的股票交易过程, 利用股票的预测收盘价, 计算预测回报率, 提出一种股票投资策略, 获取在这种策略下股票的损失与收益情况。回报率反映了股票的盈利能力, 当回报率大于 0 时,

回报率越高说明收益越高, 当回报率小于 0 时, 回报率越高说明损失越小。 $T$  时的回报率  $R_T$  以及预测回报率  $RP_T$  的计算公式如下:

$$R_T = \frac{\text{porfit}_T - \text{profit}_{T_0}}{\text{profit}_{T_0}} \times 100\% \quad (22)$$

$$RP_T = \frac{\text{preporfit}_T - \text{profit}_{T_0}}{\text{profit}_{T_0}} \times 100\% \quad (23)$$

其中  $\text{porfit}_T$  表示  $T$  时的资产总值,  $\text{profit}_{T_0}$  表示初始时刻  $T_0$  的资产总值,  $\text{preporfit}_T$  表示根据  $T$  时预测收盘价  $\hat{Y}_T$  计算的买入或卖出后的资产总值。

假设初始时刻  $T_0$  时资产总值  $\text{profit}_{T_0}$  为 10000 元, 每次买入都是用目前的总资产全部买入该股票, 每次卖出都是将持有的股票全部卖出, 并考虑每次买入和卖出的交易手续费为千分之四。根据预测回报率  $RP_T$ 、预测收盘价  $\hat{Y}_T$  以及  $T-1$ 、 $T-2$  时的真实收盘价  $Y_{T-1}$ 、 $Y_{T-2}$ , 基于“低价买入、高价卖出”的原则, 制定的投资策略如下: 当未持有该股票时, 如果  $T-1$  时的收盘价  $Y_{T-1}$  低于  $T-2$  时的收盘价  $Y_{T-2}$ 、 $T$  时预测收盘价  $\hat{Y}_T$  比  $T-1$  时降低 0.5%, 即  $Y_{T-1} < Y_{T-2}$ 、 $\hat{Y}_T < Y_{T-1} \times (1-0.5\%)$ , 则全部买入; 当持有股票时, 如果  $T-1$  时的收盘价  $Y_{T-1}$  高于  $T-2$  时的收盘价  $Y_{T-2}$ 、预测收盘价  $\hat{Y}_T$  增长 1%, 即  $Y_{T-1} > Y_{T-2}$ 、 $\hat{Y}_T > Y_{T-1} \times (1+1\%)$ , 或者预测回报率  $RP_T$  大于 0 并且比  $T-1$  时的回报率  $R_{T-1}$  下降 50%, 即  $0 < RP_T < R_{T-1} \times 50\%$ , 则全部卖出。三元股份、圣龙股份、西王食品在 2021 年 6 月 17 日至 2021 年 11 月 1 日期间的收益情况如图 8 至图 10, 其中 Real 表示只根据已知的股票收盘价进行买卖, 没有考虑预测收盘价。

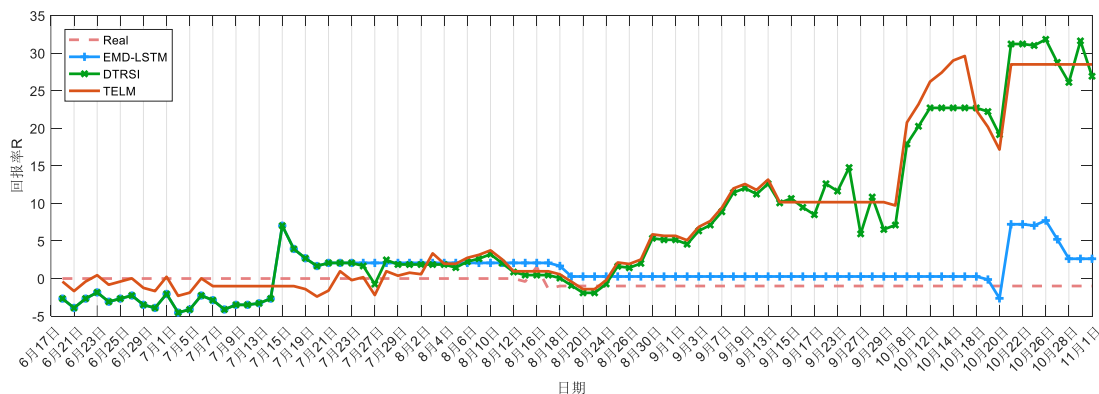


图8 三元股份回报率曲线

Fig.8 Sanyuan Shares Return Rate Curve

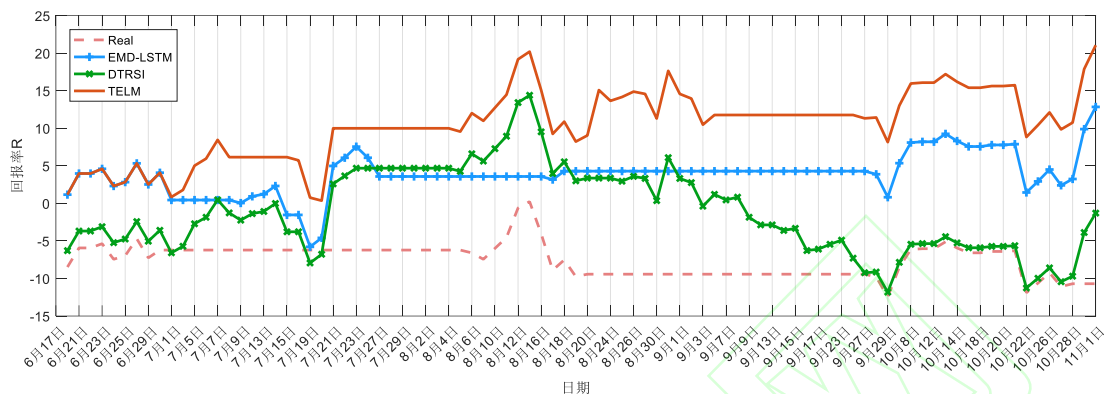


图9 圣龙股份回报率曲线

Fig.9 Shenglong Shares Return Rate Curve

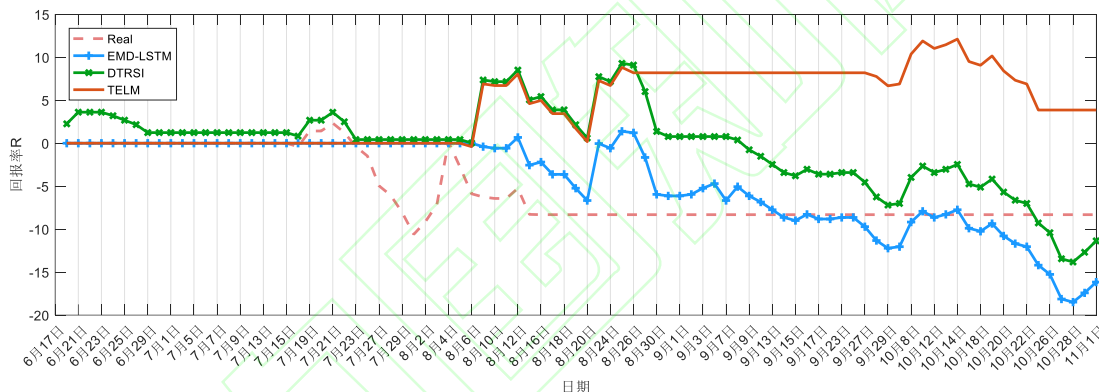


图10 西王食品回报率曲线

Fig.10 Xiwang Foods Return Rate Curve

从图 8 至图 10 中不同趋势股票的回报率曲线图可以看出,通过参考股票收盘价的预测值进行股票交易,能有效提高回报率,降低投资风险,而没有考虑预测的股票收盘价的 Real 曲线则收益率低,容易造成经济损失。对比三元股份的回报率曲线图 8 和价格曲线图 4 可以看出,三元股份估价总体是上升趋势,且增速缓慢,TELM 和 DTRSI 都取得较好的收益率;对比圣龙股份的回报率曲线图 9 和收盘价曲线图 5,可以发现,圣龙股份的股票收盘价在此期间总体是先涨后跌,且在前期上涨阶段价格涨跌波动幅度大,TELM 在上涨阶段获得了最大回报率,在后期下跌阶段及时卖出,未造成严重的经济损失;图 10 和图 6 分别为西王食品的回报率曲线和收盘价曲线,西王食品的收盘价总体呈现下跌趋势,极易造成经济损失,图 10 的回报率曲线可以看出,TELM 在合适的时间卖出,保证了收益,而 DTRSI、EMD-LSTM 未能及时止损,造成了严重的经济损失。由此可见,参考 TELM 预测的收盘价进行投资决策比 DTRSI、EMD-

LSTM 预测的股票收盘价更能有效降低股民的投资风险、提高投资收益。不同模型的回报率如表 12 所示,3 个模型中 TELM 的平均回报率、最低回报率、最高回报率均为最大值。

表 12 统计了 3 个模型的最终遭受了经济损失的股票占比及其回报率均值和盈利的股票占比及其回报率均值。损失股票占比越高、回报率均值越低说明风险越大,反之,盈利股票占比越高、回报率均值越高,说明该模型的收益越高。3 个模型相比,TELM 在损失股票占比中最低、盈利股票占比中最高,不同模型的损失回报率均值大小关系为  $DTRSI < TELM < EMD-LSTM$ ,盈利回报率均值大小关系为  $DTRSI > TELM > EMD-LSTM$ ,即 DTRSI 具有高风险、高收益,EMD-LSTM 具有低风险、低收益,TELM 的风险和收益均为中间水平。TELM、DTRSI、EMD-LSTM 盈利回报率均值与损失回报率均值的差分别为 1.04%、0.17%、0.42%,表明 TELM 与其它 2 个模型相比,盈利回报率均值远高于其损失回报率均值,说



明 TELM 在降低了投资风险的前提下,能更大程度提高收益。因此,将 TELM 预测的收盘价用于股票交易

中更能控制投资风险、提高投资回报率。

表12 三种模型的回报率统计  
Table 12 Return Rate Statistics of Three Models

模型	R均值	最低R	最高R	损失股票占比	损失R均值	盈利股票占比	盈利R均值
TELM	1.37%	-40.80%	108.82%	42.55%	-10.48%	50.68%	11.52%
DTRSI	-0.31%	-42.69%	80.32%	51.49%	-12.17%	48.24%	12.34%
EMD-LSTM	0.42%	-45.19%	96.63%	44.72%	-10.46%	46.88%	10.88%

#### 4 总结

为了充分利用股市内不同股票的价格波动之间的关联性,提高股票的预测效果,本文提出一种基于深度迁移学习的多尺度股票预测模型 TELM,用于预测单支股票或指数的收盘价。通过 EMD 将收盘价数据分解为多尺度分量 IMF 和 RES,从非线性、非平稳的股票数据中捕获股票价格与时间的关系,并利用深度迁移学习的方法训练预测模型,从市场内其它股票的历史数据中提取信息,提高股票的预测效果。实验对 500 支股票和上证指数、深证成指、上证 50、中证 500、沪深 300 进行预测,并根据预测结果模拟股票交易过程。实验结果表明,与 MA、LSTM、CNN-LSTM、EMD-LSTM 和 DTRSI 相比,TELM 模型能更好的预测股票未来趋势,预测误差更低、拟合效果更好。从行业因素、上市时间、IMF 分量个数的角度进行分析发现,通过在预训练数据集中加入某行业的部分股票数据,能有效提升该行业内其它股票的预测效果,并且通过经验模态分解后 IMF 个数越多,股票的预测效果越好。模拟股票交易过程的实验表明,通过预测股票收盘价能有效提高股票的投资回报率,对股民的投资决策有重要的指导意义。

股票的波动还受经济、政策等因素的影响,如何利用这些因素提升股票的预测效果,是未来可以进一步研究的内容。

#### 参考文献:

- [1] SONG F M, Y. L H. Volatility Decomposition of an Empirical Study on China Stock Market[J]. Collected Essays on Finance and Economics, 2004(04): 23-8.
- [2] SHARMA A, BHURIYA D, SINGH U. Survey of stock market prediction using machine learning approach[C]//2017 International conference of Electronics, Communication and Aerospace Technology (ICECA). 2017: 506-9.
- [3] 娄磊, 刘璐, 刘先俊, 等. 基于小波去噪的 ARIMA-LSTM 混合模型及对股票价格指数的预测[J]. 长春理工大学学报(自然科学版), 2021, 44(02): 119-23.
- [4] LOU L, LIU L, LIU J X, et al. ARIMA-LSTM Hybrid Model Based on Wavelet Denoising and Prediction of Stock Price Index[J]. Journal of Changchun University of Science and Technology(Natural Science Edition), 2021, 44(02): 119-23.
- [5] 贺毅岳, 高妮, 王峰虎, 等. EMD 分解下基于 SVR 的股票价格集成预测[J]. 西北大学学报(自然科学版), 2019, 49(03): 329-36.
- [6] HE Y Y, GAO N, WANG F H, et al. Research on integrated forecasting of stock price based on EMD and support vector regression[J]. Journal of Northwest University(Natural Science Edition), 2019, 49(03): 329-36.
- [7] ZHANG X, ZHAI R, GAO W. Analysis and Economic Prediction of Stock Index Based on Index Tracking and ARIMA Model[J]. Journal of Physics: Conference Series, 2021, 1903(1): 012-5.
- [8] MALLQUI D C A, FERNANDES R A S. Predicting the direction, maximum, minimum and closing prices of daily Bitcoin exchange rate using machine learning techniques[J]. Applied Soft Computing, 2019, 75:596-606.
- [9] RIBEIRO M H D M, DOS SANTOS COELHO L. Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series[J]. Applied Soft Computing, 2020, 86:105837.
- [10] DASH R, DASH P K. A hybrid stock trading framework integrating technical analysis with machine learning techniques[J]. The Journal of Finance and Data Science, 2016, 2(1): 42-57.
- [11] YAN X, CAI Q, ZHANG S, et al. Exploring Machine Learning in Stock Prediction Using LSTM, Binary Tree, and Linear Regression Algorithms[J]. International Core Journal of Engineering, 2021, 7(3): 373-7.
- [12] ALTHELAYA K A, EL-ALFY E, MOHAMMED S. Stock Market Forecast Using Multivariate Analysis with Bidirectional and Stacked (LSTM, GRU)[C]//2018 21st Saudi Computer Society National Computer Conference (NCC).

- 2018: 1-7.
- [11] ZHAO Z, RAO R, TU S, et al. Time-Weighted LSTM Model with Redefined Labeling for Stock Trend Prediction[C]//2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI). 2017: 1210-7.
- [12] MOOTHA S, SRIDHAR S, SEETHARAMAN R, et al. Stock Price Prediction using Bi-Directional LSTM based Sequence to Sequence Modeling and Multitask Learning[C]//2020 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON). New York City, USA: IEEE, 2020: 78-86.
- [13] ALTHELAYA K A, EL-ALFY E M, MOHAMMED S. Evaluation of bidirectional LSTM for short-and long-term stock market prediction[C]//2018 9th International Conference on Information and Communication Systems (ICICS). 2018: 151-6.
- [14] QIAN F, CHEN X. Stock Prediction Based on LSTM under Different Stability[C]//2019 IEEE 4th International Conference on Cloud Computing and Big Data Analysis (ICCCBDA). 2019: 483-6.
- [15] LI B, LI L-F. Stock Prediction Based on Adaptive Gradient Descent Deep Learning[M]. Springer International Publishing, 2021: 51-62.
- [16] NGUYEN T T, YOON S. A Novel Approach to Short-Term Stock Price Movement Prediction using Transfer Learning[J]. Applied Sciences, 2019, 9(22): 4745.
- [17] 张倩玉, 严冬梅, 韩佳彤. 结合深度学习和分解算法的股票价格预测研究[J]. 计算机工程与应用, 2021, 57(05): 56-64.
- ZHANG Q Y, YAN D M, HAN J T. Research on Stock Price Prediction Combined with Deep Learning and Decomposition Algorithm[J]. Computer Engineering and Applications, 2021, 57(05): 56-64.
- [18] LIU F, LI X, Wang L. Exploring Cluster Stocks based on deep learning for Stock Prediction[C]//2019 12th International Symposium on Computational Intelligence and Design (ISCID). 2019: 107-10.
- [19] 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): 903-95.
- [20] KUMAR P, HATI A S. Transfer learning-based deep CNN model for multiple faults detection in SCIM[J]. Neural Computing and Applications, 2021, 1-12.
- [21] PAN S J, YANG Q. A Survey on Transfer Learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10): 1345-59.
- [22] HASSAN D S U, MUZAFFER Ö, BURAK Ç A, et al. A Transfer-Learning Approach for Accelerated MRI Using Deep Neural Networks[J]. Magnetic resonance in medicine, 2020, 84(2): 663-85.
- [23] WANG W, WANG Z, ZHOU Z, et al. Anomaly Detection of Industrial Control Systems Based on Transfer Learning[J]. Tsinghua Science and Technology, 2021, 26(06): 821-32.
- [24] TAN C, SUN F, KONG T, et al. A Survey on Deep Transfer Learning[C]//Artificial Neural Networks and Machine Learning – ICANN 2018. Cham: Springer International Publishing, 2018: 270-9.
- [25] S.HOCHREITER, J.SCHMIDHUBER. Long shart-term memory[J]. Neural Computation, 1997, 9(8): 1735-80.
- [26] CHIARELLA C, HE X-Z, HOMMES C. A dynamic analysis of moving average rules[J]. Journal of Economic Dynamics and Control, 2006, 30(9-10): 1729-53.
- [27] 曲健, 陈红岩, 刘文贞等. 基于改进网格搜索法的支持向量机在气体定量分析中的应用[J]. 传感技术学报, 2015, 28(05): 774-778.
- QU J, CHEN Y H, LIU W Z, et al. Application of Support Vector Machine Based on Improved Grid Search in Quantitative Analysis of Gas[J]. Chinese Journal of Sensors and Actuators, 2015, 28(05): 774-778.