

基于LSTM和GRU的股价预测研究

王代颖

贵州大学经济学院, 贵州 贵阳

收稿日期: 2024年3月22日; 录用日期: 2024年4月23日; 发布日期: 2024年5月31日

摘要

随着计算机水平和数据库技术的迅猛发展, 以神经网络为基础的深度学习已经成为机器学习领域中最热门的研究方向, 人工神经网络、卷积神经网络、BP神经网络等先后被广泛应用于各领域中, 并取得了很好的效果。如今, 越来越多的金融学者也将目光投向了神经网络, 运用网络技术建立股票收盘价趋势预测模型, 取得了显著效果。本文基于用python构建了门控循环单元模型(GRU)和长短期时间记忆(LSTM)模型对光大银行收盘价进行预测, 并对两种算法进行比较。实证结果表明: GRU模型在MAE、RMSE、MAPE三个评价指标的预测精度均高于LSTM模型, 测试集的拟合也是最好的, 能够在一定程度上反应股票价格波动趋势, 为投资者进行投资决策提供了一个简单实用的方法。

关键词

门控循环单元模型, 长短期时间记忆, 趋势预测

A Study on Stock Price Prediction Based on LSTM and GRU

Daiying Wang

School of Economics, Guizhou University, Guiyang Guizhou

Received: Mar. 22nd, 2024; accepted: Apr. 23rd, 2024; published: May 31st, 2024

Abstract

With the rapid development of computer level and database technology, deep learning based on neural network has become the hottest research direction in the neighborhood of machine learning, and artificial neural network, convolutional neural network, BP neural network and so on have successively been widely used in various fields and achieved good results. Nowadays, more and more financial scholars also focus on neural networks, using network technology to establish

stock closing price trend prediction model, and have achieved significant results. This paper is based on the gated recurrent unit model (GRU) and long short-term time memory (LSTM) model constructed in python to predict the closing price of Everbright Bank, and compare the two algorithms. The empirical results show that the prediction accuracy of the GRU model in the three evaluation indicators of MAE, RMSE and MAPE is higher than that of the LSTM model, and the fit of the test set is also the best, which is able to respond to the trend of the stock price fluctuations to a certain extent, and provides a simple and practical for investors to make investment decisions.

Keywords

Gated Cyclic Cell Model, Long and Short-Term Time Memory, Trend Prediction

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

股票市场一直是人们广泛关注的话题之一。许多个人投资者、机构投资者和分析师希望能够准确预测股票价格的走势，以做出明智的投资决策。

然而，股票市场的价格变化受到诸多因素的影响，这使得股票价格的预测变得极为复杂和困难。传统的统计模型和方法往往无法充分捕捉到这些复杂的非线性关系和动态变化。随着计算机技术的发展，深度学习技术在股票价格预测领域取得了巨大进展。传统的机械学习方法如决策树、逻辑回归、随机森林和支持向量机(SVM)等在股票价格预测中被广泛应用。然而，现代机器学习中的神经网络模型在处理多种影响因素和复杂非线性问题时相对传统的机械学习方法具有优势。因此，越来越多的学者开始使用神经网络模型来预测股票价格。其中，基于长短期时间记忆(LSTM)和门控循环单元(GRU)的模型备受关注。LSTM 和 GRU 都能更好地捕捉时间序列数据中的长期依赖关系和非线性关系，避免了标准 RNN 中梯度爆炸和梯度消失的问题，这些模型通过学习历史股价的模式和趋势来预测未来价格的变化趋势，由于考虑多个时间步长的信息，利用之前的价格和相关指标来预测未来走势，因此提高了预测的准确性。

本研究旨在探讨利用 LSTM 和 GRU 模型进行股价预测的可行性和有效性。本文将使用光大银行 2015 年 1 月 5 日到 2023 年 6 月 30 日的历史数据，构建和训练 LSTM 和 GRU 模型，并通过对比实验结果和评估指标，评估它们在股价预测任务上的性能表现。通过对 LSTM 和 GRU 模型在股价预测中的研究，本文希望为金融市场的参与者和研究者提供有益的见解和指导，并促进深度学习在金融领域的应用和发展。

2. 文献综述

传统股票价格预测方法包括 VAR、ARIMA 和 GARCH 模型等。赵宁[1] (2021)通过结合技术指标和 ARIMA 模型成功预测了万科公司股票收盘价走势，说明该方法适用于股票预测。王星慧等[2] (2023)使用 GARCH 模型平均(MA)方法，在收益序列中引入非线性分位数回归模型，并提出了固定形式和非固定形式的非线性分位数回归方法(QR + GARCH 族 MA 和 QRNN + GARCH 族 MA)来测度多期 VaR 和 CVaR 风险。研究结果发现，QRNN + GARCH 族 MA 方法在多期 VaR 和 CVaR 风险测度中表现出更高的准确性和稳健性，为 GARCH 模型测度 VAR 值提供了新思路。

然而，由于股票数据的非线性性和干扰因素的影响，传统线性模型在处理股票数据时存在局限性。因此，学者们尝试引入非线性模型，如支持向量机(SVM)和神经网络等方法，并成功应用于预测股票。在 1988 年，White 已经利用 BP 神经网络预测 IBM 股票，但由于 BP 神经网络的局限性(优化方法缺乏全局性、学习速度慢、难以保证最优解等问题)，其预测效果不理想。尽管 BP 神经网络存在局限性，且受主观因素的影响较大，但仍有许多学者采用其他方法进行股价预测。例如，张玉川和张作泉[3] (2007)使用多个技术指标作为数据支持进行股价预测，而张晨希等[4] (2006)在解决较大数据集求解二次优化时发现支持向量机存在一定的局限性。支持向量机在二分类问题上表现出色，但在多类别问题中仍有改进的空间。通过结合其他相关算法、改进 SVM 算法或应用集成学习和深度学习方法，可以进一步提高支持向量机在多类别问题中的预测性能。

近年来，随着深度学习的发展，神经网络模型在各个领域得到广泛应用。在处理时间序列数据的预测任务中，常用的神经网络模型包括 MLP 神经网络模型、CNN 模型、RNN 模型、LSTM 模型和 GRU 模型。乔若羽[5] (2019)发现，在处理时间序列任务时，相对于 MLP 和 CNN，递归神经网络具有更好的性能。RNN 广泛应用于序列数据分析和预测，但由于梯度消失和梯度爆炸问题，处理长期记忆任务时存在困难。因此，近年来广泛使用 LSTM 模型。李洁和林永峰[6] (2019)的研究表明，LSTM 递归神经网络能够有效处理和预测时间跨度较长的数据，在许多领域取得显著成果。Gers 等人在 1999 年引入了遗忘门机制，提出了三门控模型，包括遗忘门、输入门和输出门，这些门控机制帮助 LSTM 模型处理序列数据中的长期依赖关系。于孝建等[7] (2013)在 LSTM 模型中加入了市场情绪特征，进一步提高了 LSTM 模型的准确度和精准性，对 LSTM 模型预测股价变化提供了新思路。而其他人则在 LSTM 模型本身的基础上进行改进，丛敬奇等[8] (2023)构建了一种 CEEMD-CNN-LSTM 模型，将补集成经验模态分解(CEEMD)、卷积神经网络(CNN)与长短期记忆网络(LSTM)结合在一起，提高了股票预测的准确性，将神经网络学习的各模型的优点结合在一起，进一步预测股价。

3. 相关理论和模型介绍

3.1. 长短期时间记忆(LSTM)

LSTM (长短期记忆网络)是 RNN (循环神经网络)的一种变体，用于处理序列数据和时间序列数据的建模和预测任务。相对于传统的 RNN，LSTM 通过引入特殊的门控机制来解决梯度消失和梯度爆炸等问题，从而更有效地捕捉到长期依赖性。

LSTM 的核心思想是引入三个特殊的门控机制来控制信息的流动和记忆的选择性。这三个门控机制分别是遗忘门(f_t)、输入门(i_t)和输出门(o_t)。这些门控机制使得 LSTM 具备了记忆和选择性更新细胞状态的能力，使网络可以更好地学习和适应时间序列数据中的模式和趋势。通过这种特殊的门控机制，LSTM 可以在训练过程中自主学习和优化每个门的参数，以适应不同的序列数据。这使得 LSTM 在处理时间序列数据时表现出色，并被广泛应用于股价预测、自然语言处理、语音识别等领域的任务中。

LSTM 的结构如图 1 所示。

遗忘门(f_t): 遗忘门用于确定哪些信息应该被丢弃或保留。它接收来自前一个隐藏状态和当前输入的信息，并通过一个 sigmoid 函数输出一个介于 0 和 1 之间的值。这个值与每个细胞状态中的信息相关，接近 0 的值表示应该丢弃该信息，接近 1 的值表示应该保留该信息。

其公式是：

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

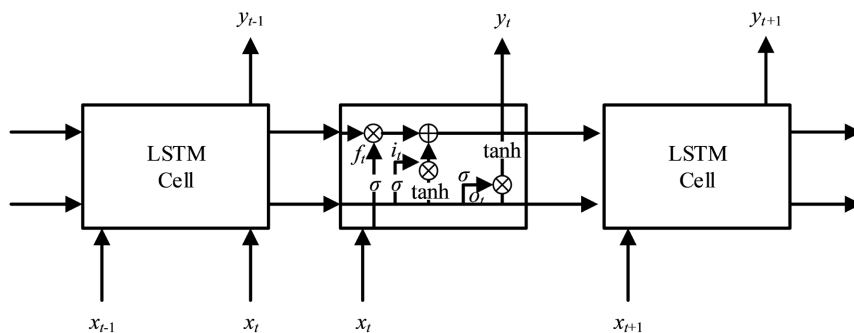


Figure 1. LSTM structure diagram
图 1. LSTM 结构图

遗忘门的输入信息由两部分组成，分别是 h_{t-1} ， x_t ，它将 h_{t-1} 与 x_t 先拼接，拼接以后再与 W_f 相乘，之后加上一个偏置 b_f ，最后经过一个 sigmoid 激活函数，这个 sigmoid 函数的作用是将线性变换后的 f_t 的取值变为为(0,1)，这个值决定了 C_{t-1} 的值表示记忆程度或遗忘程度。

输入门(i_t): 用于更新细胞状态，包含以下几个步骤：首先，输入门接收来自前一个隐藏状态和当前输入的信息，然后经过 sigmoid 函数进行处理。这个 sigmoid 函数的输出值被调整到 0 到 1 之间，用于确定哪些信息应该被更新。当输出接近 0 时，表示相应的信息是不重要的，而输出接近 1 时，表示该信息是重要的。

其次，前一个隐藏状态和当前输入的信息同时传递到 tanh 函数中，生成一个包含关于当前输入的新信息的候选值向量。tanh 函数的输出值在-1 到 1 之间。最后，通过将 tanh 函数的输出值与 sigmoid 函数的输出值相乘，决定哪些信息是重要的并需要被保留。如果 sigmoid 函数的输出接近 0，那么相应的值将被忽略，不会对细胞状态产生影响；反之，如果 sigmoid 函数的输出接近 1，那么相应的值将被保留。

这个机制允许 LSTM 模型根据输入和前一个隐藏状态的信息来更新细胞状态。通过使用输入门和 tanh 函数的组合，LSTM 可以选择性地记忆或遗忘特定的信息，从而实现了对重要上下文信息的捕捉和长期依赖的建模。

其公式是：

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

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

其中， i_t 决定着 \tilde{C}_t 的遗忘能力， \tilde{C}_t 是 LSTM 需要记忆的新的值。

将 C_{t-1} 遗忘剩余的部分，与新的要记忆的部分 \tilde{C}_t 相加，就得到要输出的 C_t ，其公式为：

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

输出门(o_t): 输出门用于确定下一个隐藏状态的值以及包含先前输入信息的隐藏状态。首先，前一个隐藏状态和当前输入经过 sigmoid 函数处理，计算出输出门的输出。接着，新计算得到的细胞状态经过 tanh 函数处理。最后，tanh 函数的输出与 sigmoid 函数的输出相乘，以确定隐藏状态所携带的信息。隐藏状态被用作当前细胞的输出，并将新的细胞状态和隐藏状态传递到下一个时间步长中。

其公式为：

$$O_t = \sigma(W_o * [h_{t-1}, x_t] + b_o)$$

$$h_t = O_t * \tanh(C_t)$$

3.2. 门控制循环单元(GRU)

GRU (门控循环单元)是一种常用的网络结构,与长短期时间记忆(LSTM)功能几乎相似。它将遗忘门和输入门合并为一个更新门,同时混合了细胞状态和隐藏状态,并进行了其他一些改动。相较于 LSTM 的三个控制门,GRU 只有两个门结构,分别是重置门和更新门。重置门控制是否重置之前的状态,即控制遗忘程度,用来决定多大程度上忘记之前的信息。更新门表示多大程度上使用候选信息来更新当前的隐藏状态,以捕捉输入的新信息。GRU 的结构如图 2 所示,它在结构上相对简单,但仍然具备强大的表达能力。通过合并门结构和隐藏状态的信息流动,GRU 可以在处理序列数据时适应长期依赖关系,并且相对于 LSTM 有更少的参数需要学习。

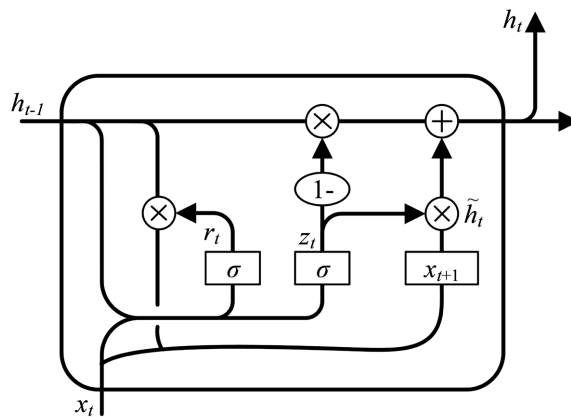


Figure 2. GRU structure diagram
图 2. GRU 结构图

其公式是:

$$z_t = \sigma(W_z * [h_{t-1}, x_t])$$

$$r_t = \sigma(W_r * [h_{t-1}, x_t])$$

$$\tilde{h}_t = \tanh(W * [r_t * h_{t-1}, x_t])$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

图中的 z_t 和 r_t 分别表示更新门和重置门。更新门确定前一时刻的状态信息对当前状态的影响程度。较大的更新门值表示前一时刻的状态信息会更多地传递到当前状态中。而重置门用于决定前一时刻的状态中有多少信息被写入当前的候选集中。较小的重置门值表示前一时刻的状态信息会较少地写入到当前状态中。

GRU 的基本原理是:首先,引入 x_t 和 h_{t-1} 生成两个门结构,用重置门通过与上一时刻的状态相乘,控制是否要重置以及需要重置的程度;然后,当前输入的 x 会与重置门的输出进行拼接,并经过一个经过 \tanh 函数激活的网络变换,生成候选信息的隐含变量 \tilde{h}_t ;最后,将上一时刻的 h_{t-1} 和候选信息 \tilde{h}_t 做一个线性组合,两者的权重之和为 1,候选信息的权重即为更新门的输出,表示需要更新信息的大小。

LSTM 和 GRU 在功能上有一些相似之处,但也存在一些不同之处。LSTM 的遗忘门与 GRU 的重置门功能类似,它们都控制前一时刻状态的遗忘程度或重置程度。而 LSTM 的输入门则与 GRU 的更新门功能相似,它们都决定多大程度上使用候选信息来更新当前状态。不过,LSTM 相对于 GRU 有一个额外

的控制门，即输出门，用于控制当前状态的输出。

相对于 LSTM 而言，GRU 模型具有更少的参数，更容易训练，并且结构相对简单。这也是本文选择这两个模型进行评价和预测比较的原因之一。使用 GRU 模型可以在减少计算负担的同时，实现相当的性能和准确度。这使得 GRU 成为一个在许多序列建模任务中广泛应用的选择。

4. 实证研究

4.1. 数据选取

本文使用了东方财富的数据作为数据来源，其中选择了一只随机抽取的 A 股股票，即光大银行(股票代码 601818)。数据集包含了 10 列数据，包括日期(trade_date)、开盘价(open)、最高价(high)、最低价(low)、收盘价(close)、前一天的成交价(pre_close)、涨跌值(change)、收益率百分比变化(pct_chg)、成交量(volume)和成交金额(amount)。这些数据记录了从 2015 年 1 月 5 日至 2023 年 6 月 30 日的股票数据，共计 2054 条数据。在本文中，将光大银行的股票数据集的前 80% 作为训练集用于模型训练，剩下的 20% 则用作测试集来验证模型的泛化能力。

在构建预测模型之前，对原始数据进行了预处理。首先，对数据按照时间进行排序。由于各个指标变量之间存在量纲上的差异，需要对它们进行统一的量纲处理。本文选择了归一化的方法，采用了 min-max 标准化，将指标数据的值映射到区间[0, 1]之间。这样可以消除不同指标之间的尺度差异，为后续的模型建立提供一致的数据输入。转化函数如下：

$$x_i = \frac{X_i - \min(X_i)}{\max(X_i) - \min(X_i)}$$

其中， $\max(X_i)$ 为样本的最大值， $\min(X_i)$ 为样本的最小值， X_i 为需要被标准化处理的原始值。在进行归一化处理后，为了确保模型预测结果与原始数据保持相同的量纲，需要进行反归一化处理，具体的反归一化转换函数如下：

$$Y_i = y_i * (\max(X_i) - \min(X_i)) + \min(X_i)$$

其中， y_i 是模型输出的原始数据， Y_i 是原始数据进行反归一化后的数据。

4.2. 模型评估标准

为了对训练后的模型在测试集上进行性能评估，本文选择了三个常用的深度学习评估指标，即平均绝对误差(MAE)、均方根误差(RMSE)和平均绝对百分比误差(MAPE)。通过使用这三个评估指标，可以全面评估模型在测试集上的性能表现，了解预测结果的准确性和拟合效果。这些指标的值越小，表示模型在测试集上的性能越好。

平均绝对误差计算公式为：

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

均方根误差计算公式为：

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

平均绝对百分比误差计算公式为：

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

其中， y_i 表示真实值， \hat{y}_i 表示预测值， n 表示值的个数。

4.3. 模型训练及结果对比分析

4.3.1. LSTM 模型

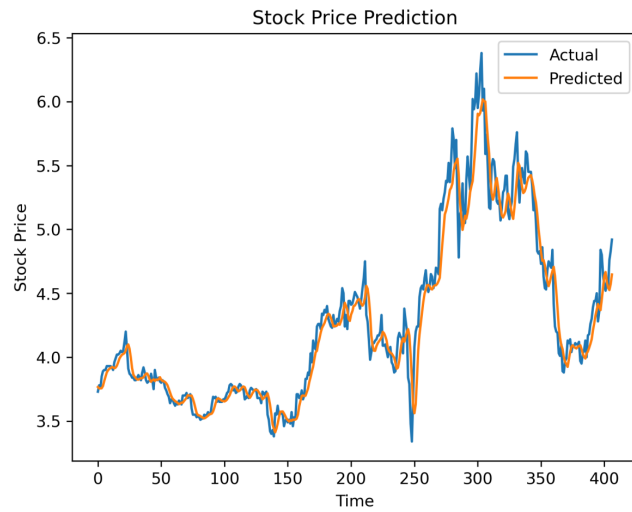


Figure 3. LSTM fitting results

图 3. LSTM 拟合结果

该模型下平均绝对误差(MAE)为 0.12046552099194613，均方根误差(RMSE)为 0.1772226347676657，平均绝对百分比误差(MAPE)为 2.6276476071219093。

4.3.2. GRU 模型

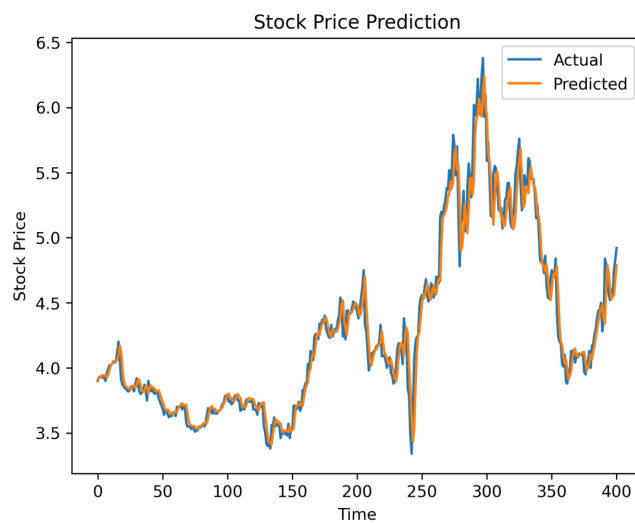


Figure 4. GRU fitting results

图 4. GRU 拟合结果

该模型下平均绝对误差(MAE)为 0.08697603444506105,均方根误差(RMSE)为 0.13067548475327212,平均绝对百分比误差(MAPE)为 1.9226223886380536。图 3、图 4 表示股票真实值与预测值的对比折线(蓝色线条为股票真实开盘价,红色线条为模型预测价格)。

通过股票真实值与预测值的对比折线图可知,拟合效果优劣为: LSTM < GRU。

通过查看模型的误差值来看,误差由大到小为: GRU < LSTM。

通过对模型的结果分析,可明显得出的三个模型的拟合效果优劣为: LSTM < GRU,即 GRU 模型的拟合效果更好。

5. 结论

本文使用了 LSTM 和 GRU 模型对股票的收盘价进行了预测研究。这两个模型都可以学习历史数据中包含的信息,但它们对历史数据的学习效果有所不同。根据结果显示,GRU 模型在对股票价格的拟合效果上相对于 LSTM 模型表现更优。这表明 GRU 模型能更准确地预测股票的收盘价,并提供更高的预测精度。可能是因为 GRU 模型在处理时间序列数据时更灵活且适应性更强,能更好地捕捉和利用时间序列的模式和关联性。

这些结果为投资者提供一定的依据和参考,使他们能更明智地运用这些模型来预测股票价格。然而,股票市场是复杂且多变的,单一的预测模型可能无法涵盖所有情况。因此,投资者在做出股票投资决策时仍需综合考虑其他因素和指标,以及行业和市场的整体状况,以提高决策的准确性。

参考文献

- [1] 赵宁. 基于 ARIMA 模型对万科股票股价的短期预测[J]. 金融客, 2023(1): 34-36.
- [2] 王星惠, 耿文静, 许启发. 基于 QRNN+GARCH 族 MA 方法的多期 VaR 和 CVaR 度量研究[J/OL]. 数理统计与管理: 1-21, 2023-07-22. <https://doi.org/10.13860/j.cnki.slj.20230628-002>
- [3] 张玉川, 张作泉. 支持向量机在股票价格预测中的应用[J]. 北京交通大学学报, 2007(6): 73-76.
- [4] 张晨希, 张燕平, 张迎春, 等. 基于支持向量机的股票预测[J]. 计算机技术与发展, 2006(6): 35-37.
- [5] 乔若羽. 基于神经网络的股票预测模型[J]. 运筹与管理, 2019, 28(10): 132-140.
- [6] 李洁, 林永峰, 陈亮, 等. 基于深度经验模态分解的金融市场时序预测[J]. 计算机工程与设计, 2019, 40(12): 3613-3619. <https://doi.org/10.16208/j.issn1000-7024.2019.12.041>
- [7] 于孝建, 刘国鹏, 刘建林, 等. 基于 LSTM 网络和文本情感分析的股票指数预测[J/OL]. 中国管理科学: 1-12, 2023-07-23. <https://doi.org/10.16381/j.cnki.issn1003-207x.2021.0084>
- [8] 丛敬奇, 成鹏飞, 赵振军. 基于 CEEMD-CNN-LSTM 的股票指数集成预测模型[J/OL]. 系统工程: 1-16, 2023-07-23. <http://kns.cnki.net/kcms/detail/43.1115.N.20230613.1454.002.html>, 2023-10-06.