参赛队员姓名：黄淇谦、刘思源

中学：北京一零一中学

省份：北京

国家/地区：中国

指导教师姓名：魏云捷

指导教师单位：中国科学院数学与系统科学研究院

论文题目：A new LASSO-BiLSTM-based ensemble learning approach for exchange rate forecasting

论文题目：**A new LASSO-BiLSTM-based ensemble learning approach for exchange rate forecasting**

作者：Siyuan Liu， Qiqian Huang

论文摘要 ： Foreign Exchange Rate affects many countries' economic status and development. Therefore, it is essential to find the factors affecting the exchange rate price and make reasonable predictions. This paper proposes the novel LASSO-BiLSTM-based ensemble learning method by integrating LASSO and BiLSTM to predict the USD/CNY exchange rate. This paper, 29 market-level variables and macroeconomic variables are selected to predict the monthly United States Dollar to Chinese Yuan (USD/CNY) exchange rate. The LASSO-BiLSTM-based ensemble learning approach contains two parts. First, the LASSO method is utilized to select the six most correlated variables with the exchange rate for future prediction; Second, the BiLSTM is employed to forecast exchange with the six chosen variables. Moreover, to test the effectiveness of BiLSTM, a comparison with four deep learning algorithms, extreme learning machine (ELM), kernel extreme learning machine (KELM), long short-term memory (LSTM), and support vector regression (SVR), are conducted. Root mean squared error (RMSE) and mean absolute error (MAE) are employed for analysis to demonstrate the accuracy of the prediction. This paper found that the proposed LASSO-BiLSTM-based ensemble learning method provides the best predictive result overall.

关键词 ： exchange rate forecasting, BiLSTM, LASSO, macroeconomic factors, market-based variables

目录 ：

论文正文：

1. Introduction

Foreign exchange rates are always characterized by high volatility and uncertainty since it is influenced by numerous unstable factors, such as economic factors and geopolitical events. However, forecasting the foreign exchange rates is essential for the government as it relates to the country's financial security and may be used to improve the accuracy and foresight of policy making. Moreover, for enterprises and individuals, accurate foreign exchange rates may help them to hedge risks and gain considerable investment income. Thus, various models are adopted to precisely capture the complexity and nonlinearity of the foreign exchange rates. This research aims to predict the exchange rate of USD/CNY by employing new combinations of variables and using different neural network models, improving the original methodology.

These methods can be divided into three categories: 1) econometric and statistic methods, including ARIMA, GARCH, VAR, etc. Econometric and statistical methods always have the stationarity assumption and perform better with stationary data. However, the predictions would be far less than ideal with the nonstationary data. 2) artificial intelligence model, including ANN, SVR, NN, etc. Artificial intelligence models overcome the shortages of econometric and statistic methods, but also have parameters optimal and overfitting problems. 3) Hybrid approach, which can combine parametric and nonparametric techniques. Ince & Trafalis divided the prediction process into two stages, using ARIMA to determine the correct input in the first stage, and combining ANN and SVR to forecast the exchange rate in the second stage (Ince & Trafailis, 2006). Henríquez & Kristjanpoller (2019) firstly used ICA to extract the noise of time series and then used NN in forecasting. Other hybrid models such as ANN-GJR (Baffour et al., 2019), and GRU-LSTM (Islam & Hossain, 2020) were used in recent years. Whereas existing literature mainly adopted a data-driven framework. The long-term result may not be captured accurately without comprehensively considering the influence of correlated variables.

Traditionally, macroeconomic and market-level variables are used to forecast the exchange rate. Since macroeconomic variables and the exchange rate are strongly and directly correlated, the predictions were made in various countries. Nor & Masron (2020) analyzed the situation in Somalia and illustrated that shocks and macroeconomic factors lead the high volatility. Kim & Park concluded that specific macroeconomic variables affect short-term and long-term fluctuations of the exchange rate in the US, respectively. Moreover, the relationships between macroeconomic variables and the exchange rate vary in different countries.

Market-level variables may be divided into different categories. For instance, for the commodity market, the cross-correlations between the exchange rate and commodity prices were analyzed (Cashin et al., 2004) (Lu et al., 2017). Some pieces of literature may focus on specific commodities and takes the financial market into account, studying correlations across energy, gold, the stock market, and the exchange rate (Sujit & Kumar, 2011) (Jain & Biswal, 2016) (Singhal et al., 2019). Others may use google trends to predict the exchange rates since it represents public opinions to a certain degree and may be more accurate in predicting the actual directions of change in the nominal exchange rate (Bulut, 2018).

However, existing literature mainly analyzed the relationship between macroeconomic and market-level variables, and few use both to predict the exchange rate. Moreover, among these variables, some may have the highest relevance to the exchange rate (Wei et al., 2018). Thus, filtering out the most necessary variables and excluding variables with weak relevance may reduce noise, providing a more accurate result.

Based on the literature, we proposed a novel LASSO-BiLSTM-based ensemble learning approach for exchange rate forecasting. First, the LASSO is used to filter the macroeconomic variables and market-based variables; Second, the BiLSTM is utilized to forecast the exchange rate with the chosen variables.

The major contributions in this paper are as follows. First, compared with the previous study, it is the first time that the market-based variables and macroeconomic variables have been selected to predict the exchange rate of USD/CNY and this new method may take market sentiment, uncertainty, public opinions, and other explicit relationships between macroeconomic variables and exchange rate into account. Second, Lasso filters out six variables with the highest relevance before the models are utilized since excessive variables may lead to overfitting. Third, to the best of our knowledge, BiLSTM is used to predict the exchange rate for the first time. And the empirical study shows that the proposed LASSO-BiLSTM approach can outperform the benchmarks.

The rest of this study is organized as follows. Section 2 introduces the related models, including LASSO, SVR, ELM, LSTM, BiLSTM, and KELM. Section 3 proposes 29 initially chosen variables in four different categories, six selected variables with the most significant relevance, and the adopted evaluation criteria. Finally, the performances of five neural network models and the other five LASSO-combining models are compared horizontally in Section 4.

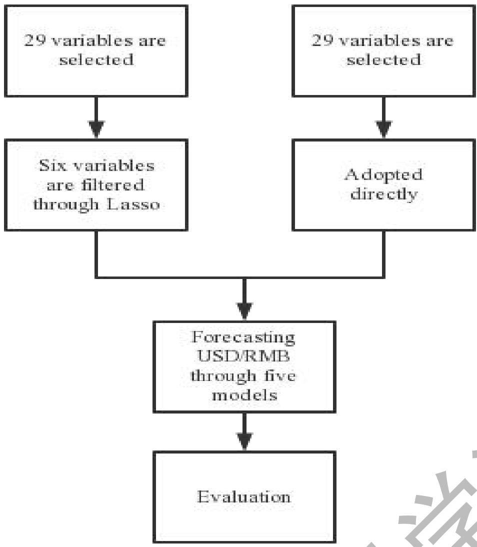

Figure 1. A LASSO-BiLSTM-based ensemble learning approach

2. Methodology

In this research, we employed several novel approaches for dollar against RMB forecasting. The least absolute shrinkage and selection operator (LASSO) method is used to select several optimal variables out of 29. Bidirectional long short-term memory (BiLSTM) model, extreme learning machine (ELM), kernel extreme learning machine (KELM), long short-term memory (LSTM), and support vector regression (SVR) models will be used in forecasting dollar against RMB, and its accuracy will be compared amongst these models.

2.1 Least absolute shrinkage and selection operator

The LASSO is a powerful method for variable selection and constructing regression. Tibshirani (1996) invented this method, which is used to enhance the testing accuracy of the model used afterwa. The basic idea of LASSO is that it adds a penalty to the model's residual sum of squares (RSS) (Tibshirani, 1996). It will accelerate the penalty if the RSS exceeds the penalty term; therefore, a regularization process will be applied to shrink some of its coefficients to zero so the penalty will nrdot be too high (Tibshirani, 1997). Thus, only several features are left, and feature selection is achieved.

The LASSO method can bring us several advantages. First, the LASSO method can increase the interpretability of the regression model due to the shrinkage of several redundant or least correlated variables. Researchers can then interpret which explanatory variable has the more significant effect on the response variable, eliminating irrelevant variables that are not correlated. Second, this process reduces overfitting. Overfitting usually occurs when the training model has too many explanatory variables. Although the model will work well in training, it does not work well in testing data. The LASSO feature selection method reduces overfitting, where a more accurate prediction of testing data will be achieved. Last, a trade-off between bias and variance will be found. This is because when the penalty term increases, bias increases and variance decreases, thus making the model less complex and reaching a balance between bias and variance.

In a linear model, we assume the formula will be

$$Y = X\beta + \varepsilon$$

with response variable Y and explanatory variable X. X can also be defined as an X_{n*k} matrix and Y can be defined as a Y_{n*1} matrix in vector form.

The LASSO minimizes the RSS with an upper bound to the sum of the absolute values of the model parameters. The upper bound is denoted as t.

$$minimize \sum_{i=1}^{n}\left(y_i - \sum_{j} x_{ij}\beta_j\right)^2 \qquad subject\ to \sum_{j=1}^{k}\left|\beta_j\right| < t$$

Which also means in vector form, the optimization will be

$$minimize \left(\frac{|Y - X\beta|_2^2}{n}\right) \qquad subject\ to \sum_{j=1}^{k}\left|\beta_j\right| < t$$

A penalty term λ is added before the sum of model parameter's absolute value. The larger the penalty term, the more shrinkage the model will be. Therefore, the parameter estimation will be

$$\widehat{\beta}(\lambda) = \underset{\beta}{\operatorname{argmin}} \left(\frac{|Y - X\beta|_2^2}{n} + \lambda\sum_{j=1}^{k}\left|\beta_j\right|\right)$$

For some values of j, $\beta_j(\lambda) = 0$. In this circumstance, these features with coefficient = 0 will be excluded from the model. Note that the penalty value is changed by researchers. Therefore, the LASSO method of feature selection is a useful method for feature selection in machine learning.

2.2 Long short-term memory and bidirectional long short-term memory

The long short-term memory (LSTM) and bidirectional long short-term memory (BiLSTM) are modeling techniques that can predict a variable based on existing variables. BiLSTM is a bidirectional LSTM approach that can detect both positive and negative cues. Both methods are instrumental, where many explanatory variables need to be explained.

Hochreiter and Schmidhuber (1997) invented the LSTM approach. The key idea behind the LSTM approach is to regulate the cell states by using three gates, the input, forget, and output gates. The forget gate determines whether the information should be kept or forgotten. The input gate determines how much information appears in the input values. The output gate controls information output from the cell state to the hidden state (Hochreiter & Schmidhuber, 1997). Figure 2 demonstrates the architecture of LSTM.

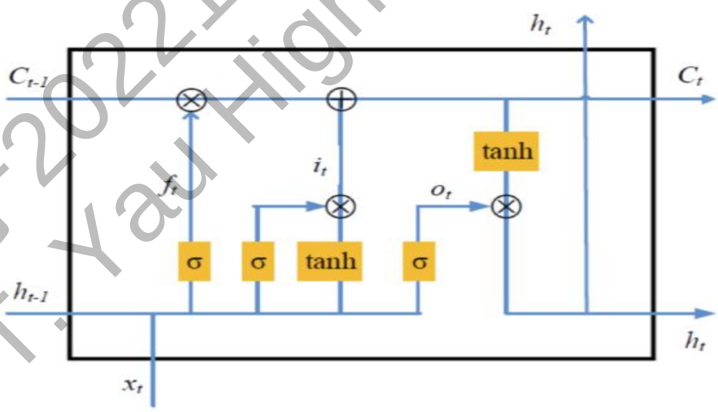

Figure 2. Graphical representation of LSTM

LSTM replaces the nodes of the recurring neural network (RNN) model with special cells, which solves RNN's vanishing gradient problem (Aslan et al., 2021). The LSTM calculated the hidden state h_t, for gates, transformation, and state update, at time t (Sherstinsky, 2020). The initial data sequence inputted into the LSTM is represented as $x_1, x_2, \ldots, x_n$. The LSTM model can be described as follows:

Gates:

$$f_t = \delta\left(W_f \cdot [h_{t-1}, x_t] + b_f\right)$$

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

$$o_t = \delta(W_o \cdot [h_{t-1}, x_t] + b_o)$$

Transformation:

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

State update:

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

$$h_t = o_t * \tanh\,(C_t)$$

where f_t, i_t, and o_t are three gates, δ represents the sigmoid function, C_t represents the cell state, and h_t represents the hidden state (Zhang et al., 2020). W and b are cell parameters. These structures work together to form a network, where inputs can pass by getting through these functions and outputted from the hidden state.

On the other hand, BiLSTM is a bidirectional LSTM approach, and it is more complicated than LSTM. BiLSTM contains six layers, input layer, embedding layer, BiLSTM layer, global pooling layer, concatenation layer, and output layer (Zhang et al., 2020). Figure 3 demonstrates the architecture of the BiLSTM model.

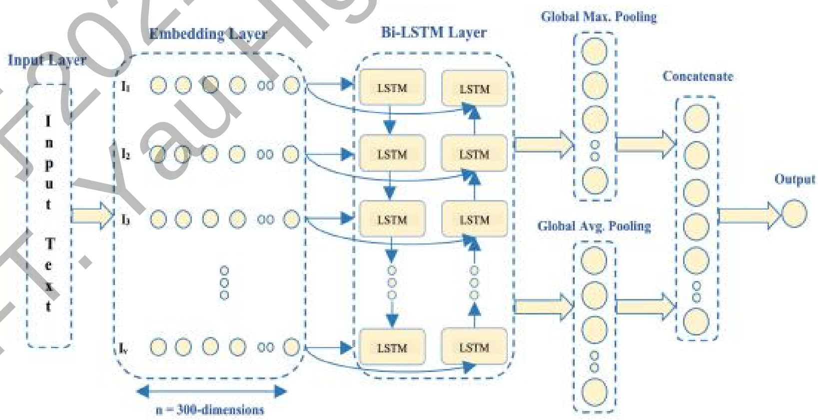

Figure 3. Graphical representation of BiLSTM

In the first layer, the input layer, let $D = d_1, d_2, d_3, \ldots, d_v$ be the total number of dimensions the data input. Then, $i_1, i_2, i_3, \ldots, i_n$ be the total unique indices in each dimension. The data is now carried as a sequence of unique indices that allows the embedding layer and BiLSTM layer to compute (Hameed & Garcia-Zapirain, 2020).

The input layer processes data into the machine and organizes it, preparing for later data processing and model running.

The embedding layer transformed the input into an embedding matrix with dimension $n*d$, where n depicts the number of indices in dimension d (Hameed & Garcia-Zapirain, 2020). The resulting embedding matrix X looks approximately like the following.

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,d} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,d} \\ x_{3,1} & x_{3,2} & \cdots & x_{3,d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n-1,1} & x_{n-1,2} & \cdots & x_{n-1,d} \\ x_{n,1} & x_{n,2} & \cdots & x_{n,d} \end{bmatrix}$$

The BiLSTM layer processes the pre-embedded data into the BiLSTM model. In this situation, BiLSTM is used to predict the hidden state $H = (h_1, h_2, \ldots, h_n)$, which is concatenated from both forward to back and back to forward. The hidden state in this paper will be the dollar against the RMB exchange rate. At time t, for the input x_t, the hidden unit vector function $\overrightarrow{h}$ is calculated based on the previous hidden state $\overrightarrow{h}_{t-1}$. At the same time, a backward hidden function $\overleftarrow{h}$ is calculated based on the future hidden state $\overleftarrow{h}_{t+1}$. For each t, the hidden states at both directions $\overrightarrow{h}$ and $\overleftarrow{h}$, are concatenated into a long vector (Chen et al., 2017). Mathematically, the hidden vector h is represented as follows:

$$h_t = \overrightarrow{h}_t \oplus \overleftarrow{h}_t, h_t \in R^{2L}$$

Where ⊕ represents the concatenation method and L means the size of each direction of LSTM. The output of BiLSTM can be summarized as a single strand $h_t = [\overrightarrow{h}_t, \overleftarrow{h}_t]$.

In the global pooling layer, the output h_t are passed into a global maximum pooling layer and global average pooling layer. A maximum value is retrieved in the global maximum pooling, and an average value is retrieved from the latter.

Then, the concatenation layer concatenated these two values, and the final value is shown in the output layer.

2.3 Support vector regression

Support vector regression (SVR) is a regression model adopted from the basis of support vectors. In a nutshell, SVR divides the dataset into different classes, and the response variable's prediction will be classified into one of those classes.

Considering a set of training dataset $\{(x_1, y_1), ..., (x_l, y_l)\} \subset X \times \mathbb{R}$, where X is a matrix for the input patterns. We define an ε-SV regression as finding a function that for each predicted point, has at most ε deviation, to the actual point (Cortes & Vapnik, 1995). This means this regression will not accept any deviations larger than ε.

We begin by assuming the function we want to achieve is a linear one. For the linear function f:

$$f(x) = \langle \omega, x \rangle + b, \omega \in X, b \in \mathbb{R}$$

With ω is a vector in X, and $\langle \omega, x \rangle$ represent the dot product of two vectors. To keep the regression flat, we need a smaller ω. Therefore, we tried to minimize $\langle \omega, \omega \rangle = \frac{1}{2}\|\omega\|^2$, in an optimization problem:

$$\text{minimize } \frac{1}{2}\|\omega\|^2$$
$$\text{S.T. } \begin{cases} y_i - \langle \omega, x_i \rangle - b \leq \varepsilon \\ \langle \omega, x_i \rangle + b - y_i \leq \varepsilon \end{cases}$$

This optimization problem will only exist is all of the points can lie between +ε and -ε. However, for some set of data, we cannot achieve that. Therefore, a slack variable, representing the difference between x_i to +ε or -ε is implemented as ξ_i, ξ_i^* (Wu et al., 2004). Therefore, we arrive a new optimization, which minimizes the $\langle \omega, \omega \rangle$ as well as ξ_i, ξ_i^*. The new equation will be:

$$\text{minimize } \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*)$$
$$\text{S.T. } \begin{cases} y_i - \langle \omega, x_i \rangle - b \leq \varepsilon + \xi_i \\ \langle \omega, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases}$$

where C is a constant that means the trade-off of flatness to the deviations it allows (Smola & Schölkopf, 2004). This equation is the final model proposed, effectively making valid predictions.

2.4 Extreme learning machine and Kernel extreme learning machine

Extreme learning machine (ELM) is another deep learning model we will compare with. This model was first introduced in 2006, aiming to improve the efficiency of the single-hidden-layer feedforward network (SLFNs) (Huang et al., 2006).

Overall, there are numerous benefits of ELM over other SLFN networks. First, the speed of ELM is extremely fast. It takes only seconds or faster than seconds to finish the model's learning phase (Huang et al., 2006). For complicated models, the ELM learning machine only spends 1.6 minutes on running, but similar algorithms, such as SVR, takes 12 hours (Huang et al., 2006). This makes the model run more efficient than before. Second, the model is more effective than other SLFN models.

Secondly, the ELM model's simplicity reduces multiple problems in other traditional gradient-based learning machines, such as overfitting and improper learning rate (Huang et al., 2006). The simple and straightforward ELM algorithm minimized these problems overall.

For N arbitrary distinct training samples of inputs (x_i, t_i), which x_i represents $[x_{i1}, x_{i2}, \dots, x_{in}]^T \in R^n$ and t_i represents $[t_{i1}, t_{i2}, \dots, t_{im}]^T \in R^m$. Therefore, the training sample $(x_i, t_i) \in R^n \times R^m, (i = 1, 2, 3, \dots, N)$, the output of SLFN for $\widetilde{N}$ hidden nodes can be represented as:

$$o_j = \sum_{i=1}^{\widetilde{N}} \beta_i f_i(x_j) = \sum_{i=1}^{\widetilde{N}} \beta_i f(a_i \cdot x_j + b_i)\, , j = 1, 2, \dots, N$$

Where o_j is the output in SLFN, for input $x_i, i = 1, 2, \dots, N$. a_i and b_i are the learning parameters for j[th] hidden node, and there are N hidden nodes in total. $\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]^T$ is a linking parameter connecting between the j[th] hidden node to the output o_j (Ding et al., 2014). $f(a_i \cdot x_j + b_i)$ is the activation function for the SLFN (Ding et al., 2014).

The equation above can be represented concisely as:

$$H\beta = O$$

And H is the output matrix for the hidden layer, β is the linking matrix between hidden layer and outputs, and O is the output matrix for the SLFN (Ding et al., 2015). They can be represented mathematically as:

$$H=\begin{bmatrix} f(a_1\cdot x_1+b_1) & \cdots & f(a_i\cdot x_j+b_i) \\ \vdots & \ddots & \vdots \\ f(a_1\cdot x_N+b_1) & \cdots & f(a_i\cdot x_j+b_i) \end{bmatrix}_{N\times\tilde{N}},$$

$$\beta=\begin{bmatrix}\beta_1^T\\ \vdots \\ \beta_{\tilde{N}}^T\end{bmatrix}_{\tilde{N}\times m}, O=\begin{bmatrix}O_1^T\\ \vdots \\ O_N^T\end{bmatrix}_{N\times m}$$

For minimizing $\|O-T\|$ in ELM model, a_i and b_i are needed to be assigned randomly without considering the input data. Therefore, this model now became a linear one, using the Moore-Penrose (MP) generalization of H, denoted as $H^\dagger$ (Chen et al., 2016).

Therefore, the output weight β will be produced by the algorithm below after transformation, which finishes the ELM model:

$$\beta=H^\dagger T$$

Kernel extreme learning machine (KELM) is a Kernel approach to ELM. Multiple variations of KELM are used in previous research, such as multiple KELM (Liu et al., 2015), hybrid KELM (Lv et al., 2020). Variations of kernels also exist, such as the radial basis function (RBF) kernel (Iosifidis et al., 2015), exponential radial basis function (ERBF) kernel (Madheswaran & Dhas, 2015). In this research, the RBF kernel is used. This kernel is added to the ELM model making it more accurate in testing data. It is a popular kernel used in multiple areas. The RBF kernel can be defined as:

$$K_{RBF}(\mathbf{x},\mathbf{x'})=\exp\left[-\gamma\|\mathbf{x}\text{-}\mathbf{x'}\|^2\right]$$

where γ is the parameter that sets the spread of the kernel (Ding et al., 2021). By using such method, a more accurate prediction might be achieved.

3. Empirical Study

3.1 Data collection

29 variables ranging from various fields are selected, including: **(1)Macroeconomic variables:** Federal reserve asset (FRA), export value (EX), import value (IM), China money supply (CM2), China consumer price index (CCPI),

industrial value added (IVA), Producer Price Index (PPI), purchasing managers' index (PMI), America consumer price index(ACPI), dollar index (DXY), America money supply (AM2), Global economic policy uncertainty (GEPU), China economic policy uncertainty (CEPU), America economic policy uncertainty (AEPU), America Economic Policy Uncertainty: News Index (AEPU-NI). **(2) Financial assets:** West Texas Intermediate (WTI), future closing price of gold (FCPG), Dow Jones industrial average (DJIA), NASDAQ Composite Index (NASDAQ), The S&P 500 index (S&P), S&P 500 Volatility Index (VIX), Shenzhen composite index (SZCOMP), Shanghai composite index (SHCOMP). **(3) Sentimental index**: University of Michigan consumer sentiment index (UMCSI), average searching number in Google Trends(GT), average searching number in Baidu index(BI). **(4) Other composite indicators**: China composite leading indicator (CLI_China), global composite leading indicator (CLI_Global), America composite leading indicator (CLI_US).

Table 1: Selected variables within four categories

Category	Detailed variables
Macroeconomic variables	FRA, EX, IM, CM2, CCPI, IVA, PPI, PMI, ACPI, DXY, AM2, GEPU, CEPU, AEPU, AEPU-NI (15)
Financial assets	WTI, FCPG, DJIA, NASDAQ, S&P, VIX, SZCOMP, SHCOMP (8)
Sentimental index	UMCSI, GT, BI(3)
Other Composite indicators	CLI_China, CLI_Global, CLI_US(3)

In total, fifteen macroeconomic variables, eight financial asset variables, three sentimental variables, and three other composite indicators are chosen. The monthly data of GT and BI are obtained from google trends (https://trends.google.com/trends/) and the Baidu index (https://index.baidu.com/v2/index.html), respectively. UMCSI is obtained from Fred economic data (https://fred.stlouisfed.org/series/UMCSENT), CLI_China, CLI_Global, and CLI_US are obtained from OEDC (https://data.oecd.org/leadind/composite-leading-indicator-cli.htm), other variables are obtained from the Wind Database (http://www.wind.com.cn/).

The data were selected from January 2011 to June 2022, separated into an in-sample subset and an out-of-sample subset. The in-sample subset contains the former 80% of the data, while the out-of-sample subset contains the rest of the 20%.

Thus, 1) the in-sample subset is used for model training with the data from January 2011 to February 2020. 2) the out-sample subset if used for testing with the data from March 2020 to June 2022.

Several seed words are selected for GT and BI. GT is used to represent public opinions in the US, while BI is used to represent public opinions in China. For BI, five seed words are selected: RMB to US dollar exchange rate, RMB exchange rate, RMB to US dollar, Dollar RMB exchange rate, and Dollar renminbi exchange rate trend (all in Chinese), and the average monthly search volume is calculated. For GT, five other seed words are selected: Dollar RMB exchange rate, RMB exchange rate, rate of RMB to USD, dollar to RMB, and dollar to yuan, and the average monthly search volume is used to represent the overall public opinions toward the exchange rate in different periods.

A logarithmic transformation is applied to the variables that contain data bigger than 100: FRA, EX, CM2, AM2, FCPG, DJIA, NASDAQ, S&P, SZCOMP, SHCOMP, BI, CLI_China, CLI_Global, CLI_US. The statistical properties of the exchange rate and selected variables are represented in Table 2. The data distributions of CLI_China and PPI are relatively asymmetric since the absolute skew are the greatest, while M2 and WTI tend to be symmetric distributions. For kurtosis, since the value of CEPU is greater than 3, the data distribution is mesokurtic, while the other variables are platykurtic distributed.

Table 2. Statistical properties of variables

Variables	**mean**	**Standard deviation**	**skewness**	**kurtosis**	**Jarque-bera**	**P-value**	**Observation**
FRA	10.41	0.09	0.28	2.59	2.22	0.00	109
EX	7.52	0.16	0.26	1.84	7.38	0.03	109
IM	7.34	0.13	0.31	1.79	8.45	0.14	109
M2	14.07	0.29	0.02	1.91	5.36	0.00	109
CPI	2.55	1.29	0.46	1.86	9.82	0.32	109
CPI_US	1.80	0.89	0.48	1.88	9.94	0.00	109
IVA	7.93	3.61	0.17	1.84	6.62	0.00	109

PPI	0.39	4.02	0.76	2.58	11.29	0.00	109
PMI	50.67	0.93	0.42	1.90	8.68	0.82	109
DXY	89.17	8.42	-0.27	2.55	2.25	0.00	109
WTI	71.55	22.82	0.01	2.46	1.28	0.00	109
COMEX	7.21	0.13	0.40	2.02	7.21	0.46	109
DJIA	9.78	0.27	-0.16	1.80	6.94	0.00	109
NASDAQ	8.46	0.37	-0.19	1.83	6.82	0.00	109
AM2	9.39	0.15	0.04	2.00	4.61	0.00	109
S&P	7.59	0.28	-0.20	1.78	7.43	0.00	109
VIX	16.20	4.82	0.35	1.97	7.04	0.21	109
UMCSI	87.13	10.83	-0.03	2.00	4.56	0.00	109
SZCOMP	9.19	0.17	0.47	1.98	8.77	0.53	109
SHCOMP	7.91	0.19	0.12	2.74	0.57	0.00	109
GEPU	5.03	0.31	0.33	2.30	4.31	0.00	109
CEPU	5.47	0.78	0.63	3.62	9.20	0.00	109
AEPU	4.81	0.26	0.49	1.81	10.70	0.04	109
AEPU-NI	4.92	0.31	0.29	1.91	6.98	0.00	109
CLI_China	99.92	0.95	0.80	2.84	11.87	0.00	109
CLI_US	100.00	0.61	0.53	2.00	9.69	0.00	109
CLI_Global	100.07	0.58	0.37	1.60	11.43	0.00	109
BI	8.33	0.64	0.21	2.23	3.45	0.00	109
GT	42.07	11.10	0.44	2.32	5.67	0.00	109

Lasso is used to filter variables. Former 80% of data (from January 2011 to February 2020) are used to select variables. As a penalty is added to the residual sum of square, some variables' coefficients shrink to zero, which can be illustrated in Figure 4. Eventually, six variables are filtered: FRA, DXY, S&P, GEPU, CEPU, and BI.

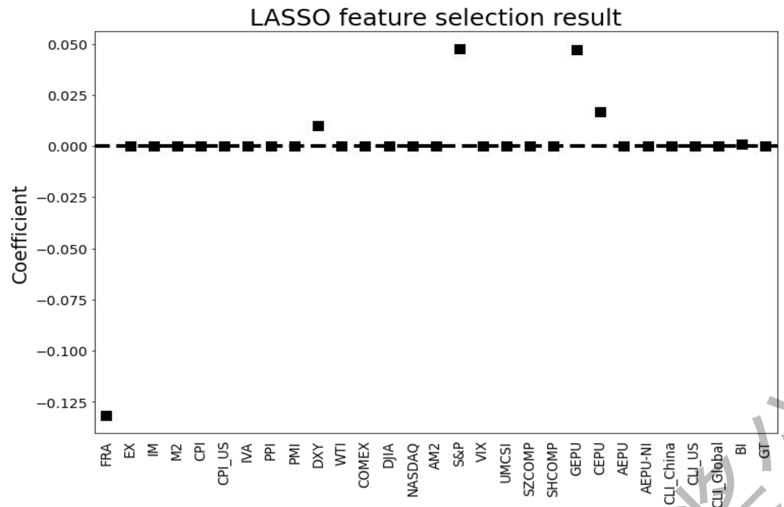

Figure 4. The coefficients of different variables for LASSO

3.2 Evaluation Criteria

Five machine learning approaches are used to make comparisons, including SVR, ELM, LSTM, BiLSTM, and KELM to find a model with strong predictive power.

Meanwhile, three methods are used to evaluate the level of forecasting accuracy, including MAE, RMSE. Here are the following formulas:

$$\text{MAE}=\frac{\sum_{t=1}^{n}|\widehat{Y_t}-Y_t|}{n}$$

$$\text{RMSE}=\sqrt{\frac{1}{n}\sum_{t=1}^{n}(Y_t-\widehat{Y}_t)^2}$$

where Y_t and $\widehat{Y_t}$ stand for the actual value and forecast value at time t respectively, while n represents the number of observations. For MAE and RMSE, the smaller calculated values represent smaller errors, which means there are less deviation between the actual value and the forecast value.

4. Performance comparison

This section discusses the forecasting accuracy of five machine learning approaches and LASSO-combining approaches. Table 3 illustrates the forecasting deviation by utilizing 29 related variables. The statistical results illustrate that BiLSTM and KELM have relatively outstanding performances, while BiLSTM outperforms all the neural network models.

Table 3. Evaluated performances of neural network models

	MAE	RMSE
SVR	54.50	55.52
ELM	56.91	58.80
LSTM	36.70	41.80
BiLSTM	**18.59**	**22.36**
KELM	23.41	24.59

Comparing the results shown in Table 3 and Table 4, which give the forecasting deviation by using six variables with the highest relevance, the errors have significantly shrunk. ELM's performance improved the most after utilizing LASSO, which decreases 27.81 and 28.84 under MAE and RMSE criteria, respectively. This improvement illustrates two important conclusions: 1) Utilizing the LASSO method of feature selection does reduce overfitting and thus gives a more promising result. 2) ELM is especially sensitive to redundant variables under this research.

Table 4. Evaluated performances of Lasso-combining models

	MAE	RMSE
LASSO-SVR	37.84	39.53
LASSO-ELM	29.10	29.96
LASSO-LSTM	22.28	25.08
LASSO-BiLSTM	**8.553**	**10.56**
LASSO-KELM	16.52	17.47

Table 4 shows that BiLSTM and KELM combining models with stronger forecasting abilities can still be concluded, which corresponds to Table 3 and thus proves the models' accuracy. However, as LASSO is used to select variables, the deviation gap between LASSO-BiLSTM and LASSO-KELM increases, which leads LASSO-BiLSTM to be more outperformed than the other models.

5. Conclusion

In this paper, the novel LASSO-BiLSTM-based ensemble learning method is proposed to predict the USD/CNY exchange rate. The framework of the proposed approach includes three steps. Firstly, 29 related variables from four different categories are selected. Secondly, LASSO feature selecting approach is used to choose six variables with the greatest relevance. Lastly, five deep learning methods

(SVR, ELM, LSTM, KELM, BiLSTM) that employ 29 related variables and LASSO-combining models (LASSO-SVR, LASSO-ELM, LASSO-LSTM, LASSO-KELM) are adopted to forecast the exchange rate of USD/CNY, indicating that LASSO-BiLSTM has the best performance and LASSO plays an important role in forecasting accurately.

参考文献：

Aslan, M. F., Unlersen, M. F., Sabanci, K., & Durdu, A. (2021). CNN-based transfer learning–BiLSTM network: A novel approach for COVID-19 infection detection. *Applied Soft Computing*, *98*, 106912.

Baffour, A. A., Feng, J., & Taylor, E. K. (2019). A hybrid artificial neural network-GJR modeling approach to forecasting currency exchange rate volatility. Neurocomputing, 365, 285-301.

Bulut, L. (2018). Google Trends and the forecasting performance of exchange rate models. Journal of Forecasting, 37(3), 303-315.

Cashin, P., Céspedes, L. F., & Sahay, R. (2004). Commodity currencies and the real exchange rate. Journal of Development Economics, 75(1), 239-268.

Chen, H. L., Wang, G., Ma, C., Cai, Z. N., Liu, W. B., & Wang, S. J. (2016). An efficient hybrid kernel extreme learning machine approach for early diagnosis of Parkinson′ s disease. *Neurocomputing*, *184*, 131-144.

Chen, T., Xu, R., He, Y., & Wang, X. (2017). Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN. *Expert Systems with Applications*, *72*, 221-230.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, *20*(3), 273-297.

Ding, S., Xu, X., & Nie, R. (2014). Extreme learning machine and its applications. *Neural Computing and Applications*, *25*(3), 549-556.

Ding, S., Zhao, H., Zhang, Y., Xu, X., & Nie, R. (2015). Extreme learning machine: algorithm, theory and applications. *Artificial Intelligence Review*, *44*(1), 103-115.

Ding, X., Liu, J., Yang, F., & Cao, J. (2021). Random radial basis function kernel-based support vector machine. *Journal of the Franklin Institute*, *358*(18), 10121-10140.

Hameed, Z., & Garcia-Zapirain, B. (2020). Sentiment classification using a single-layered BiLSTM model. *IEEE Access*, *8*, 73992-74001.

Henríquez, J., & Kristjanpoller, W. (2019). A combined Independent Component Analysis–Neural Network model for forecasting exchange rate variation. Applied Soft Computing, 83, 105654.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735-1780.

Huang, G. B., Zhu, Q. Y., & Siew, C. K. (2006). Extreme learning machine: theory and applications. *Neurocomputing*, *70*(1-3), 489-501.

Iosifidis, A., Tefas, A., & Pitas, I. (2015). On the kernel extreme learning machine classifier. *Pattern Recognition Letters*, *54*, 11-17.

Islam, M. S., & Hossain, E. (2020). Foreign exchange currency rate prediction using a GRU-LSTM hybrid network. Soft Computing Letters, 100009.

Jain, A., & Biswal, P. C. (2016). Dynamic linkages among oil price, gold price, exchange rate, and the stock market in India. Resources Policy, 49, 179-185.

Kim, Y., & Park, C. (2020). Are exchange rates disconnected from macroeconomic variables? Evidence from the factor approach. Empirical Economics, 58(4), 1713-1747.

Liu, X., Wang, L., Huang, G. B., Zhang, J., & Yin, J. (2015). Multiple kernel extreme learning machine. *Neurocomputing*, *149*, 253-264.

Lu, X., Li, J., Zhou, Y., & Qian, Y. (2017). Cross-correlations between RMB exchange rate and international commodity markets. *Physica A: Statistical Mechanics and its Applications*, 486, 168-182.

Lv, L., Wang, W., Zhang, Z., & Liu, X. (2020). A novel intrusion detection system based on an optimal hybrid kernel extreme learning machine. *Knowledge-based systems*, *195*, 105648.

Madheswaran, M., & Dhas, D. A. S. (2015). Classification of brain MRI images using support vector machine with various Kernels. *Biomedical Research*, *26*(3), 505-513.

Sherstinsky, A. (2020). Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Physica D: Nonlinear Phenomena*, *404*, 132306.

Singhal, S., Choudhary, S., & Biswal, P. C. (2019). Return and volatility linkages among International crude oil price, gold price, exchange rate and stock markets: Evidence from Mexico. Resources Policy, 60, 255-261.

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing*, *14*(3), 199-222.

Sujit, K. S., & Kumar, B. R. (2011). Study on the dynamic relationship among gold price, oil price, exchange rate and stock market returns. International journal of applied business and economic research, 9(2), 145-165.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267-288.

Tibshirani, R. (1997). The lasso method for variable selection in the Cox model. *Statistics in medicine*, *16*(4), 385-395.

Wei, Y., Sun, S., Lai, K. K., & Abbas, G. (2018). A KELM-Based Ensemble Learning Approach for Exchange Rate Forecasting. Journal of Systems Science and Information, 6(4), 289-301.

Wu, C. H., Ho, J. M., & Lee, D. T. (2004). Travel-time prediction with support vector regression. *IEEE transactions on intelligent transportation systems*, *5*(4), 276-281.

Zhang, Y., Wang, J., & Zhang, X. (2018, June). Ynu-hpcc at semeval-2018 task 1: Bilstm with attention based sentiment analysis for affect in tweets. In *Proceedings of The 12th International Workshop on Semantic Evaluation* (pp. 273-278).

Acknowledgement

1. 论文的选题来源、研究背景

The news from The Economists provides us a window to learn more thoroughly what is happening worldwide. We are usually amazed by the critical analysis: double-sworded effects brought by the increasing interest rate, possible approaches to get through the winter without wrecking Europe's electricity market, and the precise correlation between an increase in gas prices and the drop in consumption. For this time, not only do we want to perceive information from professional researchers, but we also want to take a further step, trying to be an analyst by raising our own methodology. Inspired by our instructor, the foreign exchange rate becomes our research object.

2. 每一个队员在论文撰写中承担的工作以及贡献

In our team, Siyuan Liu is responsible for providing overall idea, constructing 6 models, writing the methodology part of the essay, and finding initial variables. Qiqian Huang is responsible for providing the overall idea, providing economical interpretations for the result, writing introduction and literature review.

3. 指导老师与学生的关系，在论文写作过程中所起的作用，及指导是否有偿

We would like to thank Yunjie Wei for providing many academic guidance and advices for our work. Her guidance is unpaid, and her effort and patience sincerely affecting us. Only because of her, we made this work possible.

4. 团队成员和指导老师的简历

Team member

Siyuan Liu and Qiqian Huang are two High School grade 11 students from Beijing 101 High School International Department.

Adviser: Yunjie Wei

April 2022 to Now: Associate Professor, Academy of Mathematics and Systems Science, Chinese Academy of Sciences

July 2017-April 2022: Assistant Professor, Academy of Mathematics and Systems Science, Chinese Academy of Sciences

2018.08.02-2018.10.01: Postdoctoral Fellow, City University of Hong Kong，Department of Information Systems (Research Fellow)

2019.07.02-2019.08.30: Postdoctoral Fellow, City University of Hong Kong，Department of Economics and Finance (Research Fellow)

2015.09-2018.02: City University of Hong Kong，College of Business, Ph.D.

2012.09-2017.06: Academy of Mathematics and Systems Science, University of Chinese Academy of Sciences, Ph.D.

2008.09-2012.06: School of Economy & Trade, Hunan University