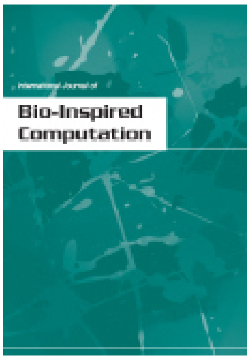
INDERSCIENCE PUBLISHERS
Linking academia, business and industry through research



International Journal of Bio-Inspired Computation

ISSN online: 1758-0374 - ISSN print: 1758-0366
https://www.inderscience.com/ijbic

Methanol price prediction method based on multimodal fusion by using CNN-GRU and attention mechanism

Shuang Luo, Xuhui Zhu, Zhiwei Ni, Pingfan Xia, Liping Ni

DOI: 10.1504/IJBIC.2024.10065641

Article History:

Received:	06 July 2022
Last revised:	02 January 2023
Accepted:	28 January 2023
Published online:	03 January 2025

Methanol price prediction method based on multimodal fusion by using CNN-GRU and attention mechanism

Shuang Luo, Xuhui Zhu*, Zhiwei Ni, Pingfan Xia and Liping Ni

School of Management,
Hefei University of Technology,
Hefei 230009, Anhui, China
and
Intelligent Interconnected Systems Laboratory of Anhui Province,
Hefei University of Technology,
Hefei 230009, Anhui, China
and
Intelligent Decision-Making and Information System Technology,
Engineering Research Center of Ministry of Education,
Hefei 230009, Anhui, China
and
Key Laboratory of Process Optimization and Intelligent Decision-Making,
Ministry of Education,
Hefei University of Technology,
Hefei 230009, Anhui, China
Email: hapluo@163.com
Email: zhuxuhui@hfut.edu.cn
Email: zhiwein@163.com
Email: xiapingfan@mail.hfut.edu.cn
Email: niliping@hfut.edu.cn
*Corresponding author

Abstract: Considering that text is important to methanol price prediction, the text and the quantity information need to be fused for achieving high precision prediction. Hence, we put forward a novel methanol price forecasting approach ground on 'text-quantity' multimodal fusion using CNN-GRU-attention mechanism network, named MFCGAM, which merges quantity information and text information obtained from 'research report', 'information' and 'investor comments'. Firstly, Word2Vec model is applied to process text, and the 'text-quantity' dual channel based on CNN and GRU is established to extract text and quantity features respectively. Secondly, attention mechanism is employed to get 'text-quantity' fused characteristics, which are used to predict methanol price. The experimental outcomes of three real datasets show that MFCGAM model obtains superior performance than other traditional models. Additionally, predictive ability of models can be improved by adding texts, and it is found that the results of short-term prediction are better than that those of long-term forecasting when using texts. It provides a very useful predictive tool for smart scheduling of coking intelligent plants.

Keywords: methanal price prediction; multimodal fusion; attention mechanism; gated recurrent unit; GRU; convolutional neural network; CNN.

Reference to this paper should be made as follows: Luo, S., Zhu, X., Ni, Z., Xia, P. and Ni, L. (2025) 'Methanol price prediction method based on multimodal fusion by using CNN-GRU and attention mechanism', *Int. J. Bio-Inspired Computation*, Vol. 25, No. 1, pp.43–55.

Biographical notes: Shuang Luo is currently pursuing her Master's in Management Science and Engineering from the Hefei University of Technology. Her research interests include deep learning and price forecast.

Xuhui Zhu is a Lecturer in the Hefei University of Technology, and received his PhD degree from HFUT. His research interests include evolution computation, deep learning and smart manufacturing.

Zhiwei Ni is a Professor and PhD supervisor in the Hefei University of Technology (HFUT). From 2010 to 2022, he has been the Director of the Institution for Intelligent Management in HFUT. His research interests include artificial intelligence, machine learning and cloud computing.

Pingfan Xia is currently pursuing her PhD in Management Science and Engineering from the Hefei University of Technology. Her research interests include machine learning, evolution computation and risk prediction.

Liping Ni is an Associate Professor in the Hefei University of Technology, and received her PhD degree from HFUT. Her research interests include deep learning and text mining.

1 Introduction

Methanol is a momentous industrial chemical, which can be used to produce various kinds of olefins, acetic acid and plastic products, etc. In addition, it is also a high-quality clean fuel, which is mainly produced from natural gas and coke. With the accelerated pace of digital transformation, coking industry emerges the tendency of constructing intelligent coking plants. Intelligent planning and scheduling is a very critical part during the building of smart coking factories, and how to accurately predict price becomes the primary problem to be solved for carrying out intelligent planning and scheduling. Methanol is one of the major products in coking industry, and its price is affected by various information from multiple channels. We aim to deal with the methanol futures price forecasting problem by fusing multi-source heterogeneous data, and achieve high-precision prediction of methanol futures prices by advancing the fusion of multimodal data features, which can contribute to the smart scheduling of coking plant.

When forecasting price, quantity information is the most frequently utilised data type. Mardianto et al. (2020) forecasted rice prices based on the daily prices, sales volume, import and export quantities of rice. Xiao et al. (2019) used 27 stock trading indicators, containing daily opening price, daily high price, daily closing price, and daily trading volume, to predict next-day prices for three stocks. All of the quantity information mentioned above is usually publicly available, easily accessible, data-rich and easy to handle. Therefore, quantity information has turned into the most commonly used data in the field of price forecasting.

Benefit from the maturity of social networks, text has become a prevailing type of information, such as stockholders' comments and news events reported in the media. Many studies have demonstrated that texts such as news and social comments indeed play a positive role in the prediction tasks. Vanstone et al. (2019) transformed news on Twitter into sentiment indicator with the help of text mining techniques, and proved that the addition of textual factors made the prediction of stock prices more precise. Huang and Liu (2020) built a stock price forecasting model based on the logistic regression, and confirmed that the accuracy of prediction model was significantly improved with the adjunction of investor sentiment. Yet, it is also found that the forecasting results are unstable and inaccurate sometimes, if we solely rely on texts to forecast without considering quantitative data (Selimi and Besimi, 2019). It is better to combine multiple variables. Therefore, price forecasting tasks based on multimodal information fusion has been launched. Cheng et al. (2021) proposed a multimodal graph neural network (MAGNN) to fuse two modal data, stock historical price and news events, which complete financial time-series forecasting tasks excellently. Li et al. (2018) put forward a crude oil price prediction model applying convolutional neural network (CNN) and LDA topic methods, which combined topic features with financial features to fulfil highly accurate price prediction.

For multimodal fusion prediction, pre-processing of text is a very momentous part. Text cannot be directly input into the model, but generally need to be transformed into structured data that is friendly to neural networks firstly. Currently, the ordinary way is to turn text into quantitative metrics, which can be grouped into three kinds: The first category is latent Dirichlet allocation (LDA) topic modelling (Xiao et al., 2018), which classifies texts into several topics and belongs to an unsupervised Bayesian model. Loginova et al. (2021) convert online texts into topic sentiment features for predicting the price gain of Bitcoin through LDA technique. The second category is sentiment analysis (Bing, 2021). It can generate emotion value for texts, which will be regarded as a new attribute and train with other attributes. Zhao et al. (2020) conducted sentiment analyses of the oil market. They divided web texts into four sentimental types, and corroborated that the inclusion of sentiment indicators could raise forecasting accuracy. The third category is word vector technique (Bojanowski et al., 2017), which can represent words into a series of vectors with specific meanings. Word2vec is the mainstream method of text vectorisation (Mikolov et al., 2013; Kang et al., 2017). Kilimci (2019) merged Word2Vec technique with CNN and LSTM to pre-train and encode social texts on Twitter, and won higher classification precision in sentiment analysis task.

In a sense, the word vector technique can be considered as one part in sentiment analysis. However, the word vector technique eliminates the step of classifying and quantifying text. Thus, it can retain the original information of text to the greatest extent, which allows prediction models to mine features autonomously. Therefore, for the sake of maximise the advantages of deep learning methods in autonomous learning, this paper chooses the word vector technique to process texts, for excavating the deep features of texts.

In terms of prediction methods, deep learning methods become the mainstream ways for multimodal data fusion forecast (Ramachandram and Taylor, 2017). Deep learning methods have more powerful autonomous learning and feature representation capabilities, which can better cope with the challenges posed by multi-source multimodal heterogeneous data (Bayoudh et al., 2021; Nie et al., 2020). Huang et al. (2019) extracted the 'image-text' multimodal features by building deep multimodal attention model, and effectively improved the classification precision of social media data. Kang and Kang (2017) fused the 'text-quantity-image' multimodal characteristics of Chicago population, and achieved high-precision forecasting of criminal behaviour. To overcome the shortcomings of single model, some researchers have tried to integrate multiple neural networks. Lin et al. (2020) put forward a SCONV model, which use convolutional LSTM network (ConvLSTM) to train sentiment indicators and prices, and realised excellent performance on small sample datasets. Feng et al. (2019) set up a deep model based on 3D-CNN and BiLSTM methods for fusing diagnostic images and numerical data of Alzheimer's disease (AD), and increased the recognition and classification accuracy of AD cases.

From the perspective of research objects, the price prediction of oil, natural gas and agricultural products has been widely concerned. However, as an important energy product, the price forecast of methanol has received little attention. Although multimodal fusion is widely applied in the emotion classification, the current studies about price prediction have some deficiencies, which are as follows: first, in the selection of text, it is usually to choose one kind of text for prediction, such as news or comments. But few literatures have explored the influence of different type's text on predicting. Second, in terms of processing text, most of studies prefer quantifying texts before forecasting. Yet, it may destroy the information integrity of texts to a certain extent. Finally, the prediction network is relatively ordinary, machine learning methods and single deep learning models are widespread. However, it is a complex prediction task when multimodal data are introduced. Hence, it is necessary to establish multi-network fusion forecast model.

Therefore, in this paper, we extend text type and collect three kinds of text, including 'research reports', 'information' and 'investor comments'. 'Research reports' are the analysis of current market by professional institutions. 'Information' is important news and events about methanol industry. And 'investor comment' is the social comments of investors on the forum, which can reflect investor's sentiment. We use word vector technique to handle texts and propose a multimodal data fusion method employing CNN, GRU network and attention mechanism for methanol price prediction, abbreviated as MFCGAM. Furthermore, we investigate the performance of three types of texts on long-term, medium-term and short-term prediction tasks, and enhance forecast precision of MFCGAM model by finding the optimal time windows for them.

To sum up, the main research work and innovations of this paper are summarised as below:

1 We explore the influences of different types of texts interrelated to methanol futures prices on predictive results.

2 Word vector technique is employed to avoid quantifying texts in advance and retain more original features.

3 A dual-channel feature extraction module is built, and the attention mechanism is used to fuse textual features and numerical features.

4 We investigate the different effects of texts on the long-term, medium-term and short-term forecasts of methanol price, the optimal prediction window for the three texts are found.

The structure of rest contents is as follows: Section 2 introduces several methods used in MFCGAM model, and expounds the ideas and forecasting steps of MFCGAM model. Section 3 analyses the experimental results in detail. And the final conclusions are displayed in Section 4.

2 The construction of prediction model

2.1 The word vector techniques

Word vector technique, also called word embedding technique, is the most fundamental measure in natural language processing. It is used for transforming texts into vectors with specific semantic meanings that can be understood by computers. One-hot representation (Xiong et al., 2020), Word2Vec (Bojanowski et al., 2017; Kang et al., 2017), Glove (Sumarsono, 2020), etc. are some common word embedding methods. One-hot representation is the most traditional way for word vector representation, which uses the combination of 0 and 1 to represent all words. Its principle is simple, but the word vector obtained by it is high-dimensional and sparse and it is difficult to discover the association between words.

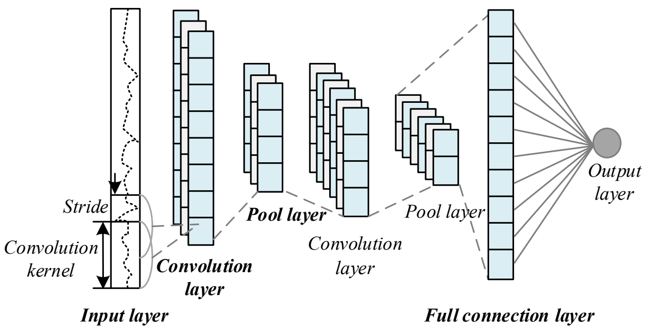
However, the emergence of Word2Vec technology has alleviated the above problems. Word2Vec is introduced by Google in 2013, which is an efficient tool for characterising words as real-valued vectors (Mikolov et al., 2013). Word2Vec method has made many improvements to the traditional one-hot representation coding means. On the one hand, in the representation of words, the word vector generated by Word2Vec no longer contain only 0 or 1, but is a dense vector with fixed dimensions. It greatly reduces the costs of computer storage and computational resources in processing natural language. On the other hand, the word vector trained by Word2Vec can give full play to the role of contextual information. Through mapping words with similar meanings to the vector space with close distance, it is able to achieve the determination of near-sense words.

Therefore, Word2Vec becomes an effective tool for word vectors.

2.2 Convolutional neural networks

As one of the most representative deep learning methods, CNN has powerful representational capability. It is broadly employed in classification and time-series forecasting tasks, since it can efficiently extract features from text, numerical and other modal data (Wang et al., 2018; Rezaei et al., 2020; Wang, 2022). The convolution layer, pooling layer and full connection layer are the three structural components and functional modules of CNN. Figure 1 reveals the internal construction of CNN.

Figure 1 The structure of CNN (see online version for colours)



The first step is the convolutional operation implemented by the convolutional layer, which excavates the short-time features of sequence. Next, the features will be aggregated by the pooling layer, which retains the important characteristics. Pooling layer can reduce the complexity of model by cutting down feature's dimensionality. Finally, the fully connected layer realises predicting by the nonlinear activation function. A complete CNN network may be composed of multiple convolutional and pooling layers (Zeiler and Fergus, 2013). The convolution and pooling process is shown in equation (1) and equation (2).

$$C = \mathrm{f}(X \otimes \boldsymbol{W} + b) \tag{1}$$

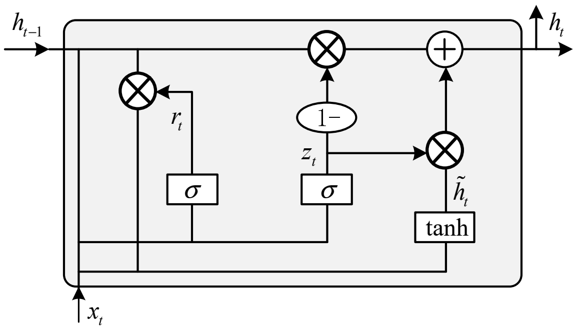
$$P = \mathrm{pool}(C) \tag{2}$$

where C refers to the characteristic combination obtained through convolution operation; f represents the nonlinear activation function; $\otimes$ denotes the convolution operation; X is the input data; $\boldsymbol{W}$ is the weight vector trained for convolutional kernel; b means the bias unit; P refers to the characteristic matrix aggregated by pooling layer; pool(C) is the pooling operation, and we adopt the maximum pooling rule.

2.3 Gated recurrent unit

When processing long sequence data, some traditional recurrent neural networks are prone to occur gradient disappearance or gradient explosion. GRU is a gated recurrent network that has emerged in recent years to resolve such problems (Chung et al., 2014). With the 'memory' and 'forgetting' functions, GRU network can better learn and catch long-term dependencies in time-series data. GRU is similar to LSTM networks in function and they are both frequently applied in computer translation, speech recognition, price prediction, and other fields (Rajamani et al., 2021; Zhang et al., 2020; Wang and Li, 2019). However, GRU network has the simpler structure. It preserves the reset gate r of LSTM and sets the update gate z, which is the fusion of forgetting gate and input gate of LSTM. Hence, GRU method requires fewer parameters and can avoid the overfitting problem to a certain extent. Figure 2 displays the structural cell of GRU.

Figure 2 The basic composition of GRU network



First, when the sequence x_t is input into GRU unit, the state information of its preceding time enters into the present state are mainly controlled via the update gate z. This gate also decides the degree of the retained candidate state information at current moments. The value of the update gate z is directly proportional to the amount of state information transferred from previous moment to current time. This process is shown in equation (3).

$$z_t = \sigma\left(\boldsymbol{W}_z \cdot \left[h_{t-1}, x_t\right] + \boldsymbol{b}_z\right) \tag{3}$$

where x_t refers to the input of immediate hidden layer; h_{t-1} means the output generated by previous hidden layer; $\boldsymbol{W}_z$ is the weight matrix and $\boldsymbol{b}_z$ is the bias unit of update gate z; σ represents activation function.

Then, the deserted extent for status messages of preceding time is mainly regulated by the reset gate r. And the less information is forgotten, the larger the value of the reset gate. This step is expressed by equation (4).

$$r_t = \sigma\left(\boldsymbol{W}_r \cdot \left[h_{t-1}, x_t\right] + \boldsymbol{b}_r\right) \tag{4}$$

where $\boldsymbol{W}_r$ refers to the weight matrix of reset gate, and $\boldsymbol{b}_r$ is the bias unit of reset gate.

In the end, utilising equation (5) and equation (6), we can get the hidden status of candidate set and the vector output by network at present time.

$$\tilde{h}_t = \tanh\left(\boldsymbol{W}_c \cdot \left[r_t \times h_{t-1}, x_t\right] + \boldsymbol{b}_c\right) \tag{5}$$

$$h_t = z_t \times h_{t-1} + \left(1 - z_t\right) \times \tilde{h}_t \tag{6}$$

where $\boldsymbol{W}_c$ is the weight matrix of hidden status unit; $\boldsymbol{b}_c$ is the bias unit of hidden status unit and tanh refers to activation function.

2.4 Attention mechanism

Attention mechanism is originally used to handle machine translation work (Bahdanau et al., 2014). Nowadays, we can find it in various types of tasks such as image processing, speech recognition, and prediction (Niu et al., 2021; Hu and Li, 2021; Hu et al., 2019). The key of attention mechanism is to quickly and correctly pick the specifics that are more significant to target from miscellaneous information. In model training, this process is reflected as weight training.

In this work, first, text features extracted from raw text data are denoted as $\mathbf{T}$, and the numerical characteristics dug from original numerical data are recorded as $\mathbf{N}$. Equation (7) is utilised to fuse $\mathbf{T}$ and $\mathbf{N}$, and the preliminary fused feature $\boldsymbol{F}$ is got. Then, weights between 0 and 1 are produced for each input feature to mirror the significance of every fused-feature. And the weight values are obtained by equation (8), which is the principle of calculating the correlation between the features at prediction moment and the characteristics at each historical moment. Finally, equation (9) is applied to weight and sum the weight values A and merged-feature $\boldsymbol{F}_i$ at various historical time to gain the final fusion feature $\boldsymbol{F}_A$.

$$\boldsymbol{F} = \tanh\left(\mathbf{T}\boldsymbol{W}_a\mathbf{N}^{\mathrm{T}} + \boldsymbol{b}_{\mathrm{a}}\right) \tag{7}$$

$$A = \frac{\exp(\boldsymbol{F})}{\sum_{i=1}^{n}\exp\left(\boldsymbol{F}_i\right)} \tag{8}$$

$$\boldsymbol{F}_A = \sum_{i=1}^{n} A\boldsymbol{F}_i \tag{9}$$

where $\boldsymbol{b}_{\mathrm{a}}$ and $\boldsymbol{W}_{\mathrm{a}}$ represent the bias unit and the weight coefficient, respectively. As with other parameters in the network, they all need to be trained. $\boldsymbol{F}_A$ refers to the merged 'text-quantity' characteristic.

2.5 The MFCGAM prediction model

The construction of forecasting model in this study is as follows: Firstly, considering that text and quantity data belong to different modalities, a dual-channel feature extraction module is developed. One channel is used to mine text features, and the other channel is to capture numerical features. Secondly, for feature excavation methods, this work tries to combine three deep learning methods to increase forecasting accuracy, including CNN, LSTM and GRU. They were paired in pairs for experiment to find the optimal method combination. Finally, in terms of fusing multimodal data feature, taking into account the diverse impact of textual and numerical indicators on methanol prices, attention mechanism is introduced to further elevate the performance of prediction model. Through the experiment in Section 3.5, it is proved that CNN and GRU can perform the best, and the combination is used in this work. The structure of MFCGAM forecasting model is exhibited in Figure 3.

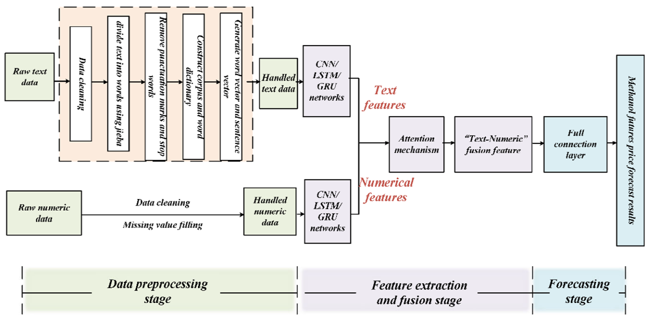
The prediction process includes three parts, which are data processing, feature extraction and fusion, and prediction.

1 The first stage is the data processing of raw text data and numerical data. For numerical data, the filling of missing values is the mainly work. Yet, the processing of raw text is more complicated. The cleaning of raw text, handling outliers and missing values are carried out first; second, the text is divided into words by using word splitting toolkit called 'jieba'; third, the punctuation marks and stop words are removed from the text; fourth, word vectors and sentence vectors can be trained. Finally, the pre-processed text data and numerical data can then be obtained.

2 The second stage is the data feature extraction and fusion, which relies on deep learning methods and extract characteristics from text and numerical data apart. Since the extracted text and quantity features are in different dimensions, we first utilise two fully connected layers to map them to the low-dimensional feature space of the same dimension; then fuse them with the help of attention mechanism to get the final 'text-quantity' features. CNN and GRU are two deep learning methods of MFCGAM model.

3 The third stage is the forecasting stage, where the 'text-quantity' fusion features are input and the methanol price prediction results are generated by fully connected layer.

Figure 3 Methanol futures price forecasting method named MFCGAM (see online version for colours)



3 Experimental description and analysis of results

3.1 Dataset description

The datasets in this paper covers two types of data related to methanol: numerical data and textual data. The numerical data are obtained from AI media Data Centre Website[1], under the 'basic database' – 'industry data' – 'energy sector' module. And the attributes in dataset contain futures closing price of methanol, futures settlement price of methanol, increase/decrease volume of methanol, empty volume of methanol, trading volume of methanol, export quantity of methanol, export amount of methanol, import quantity of methanol, import amount of methanol, futures closing price of coking coal, futures settlement price of coking coal, futures closing price of coke, futures settlement price of coke, futures closing price of crude oil, futures settlement price of crude oil, spot price of crude oil, inventory of crude oil, futures settlement price of fuel oil, futures settlement price of gasoline, and inventory of natural gas, for a total of 20 attributes. The textual data are gained from 'research reports', 'information' and 'stock bar' modules on Orient Wealth Website.[2] And we use crawler technology to get the title and main body of texts.

Owing to the dates of numerical data and text data are not exactly the same, we take the intersection of text data and numerical data by date, to ensure the completeness and meaningfulness of daily data. And three datasets final acquired are displayed in Table 1. The three data sets have different time spans and data volumes, with the most data from 'investor comments' and the least data from 'research reports'. In the experiment, the first 80% data volume of each dataset is treated as the training set, the last 20% is the test set, and the results in experiments are all from the test set. To reduce the effect of predictive randomness in this work, we take the mean values of running 20 times as the finally experimental results.

Table 1 Three datasets used in this paper

Datasets	*Time span*	*Data volume*
Research reports	2014/12/09-2021/11/10	412
Information	2015/03/06-2021/11/16	861
Investor comments	2014/12/06-2021/11/16	1,045

3.2 Data pre-processing

3.2.1 Data cleaning and data integration

There are some abnormal values in texts, which need to be eliminated. And for few missing values in the numerical data, this paper adopts the mean filling method to fill them (Raja and Thangavel, 2020). The average value of the five samples closest to the missing location is used as the replacement value.

Since the dates of textual and numerical data are not exactly the same, we take the intersection set for them by date to ensure the completeness and meaningfulness of daily data, and the three datasets obtained are exhibited in Table 1.

3.2.2 Chinese word separation and deactivation

First, we utilise 'jieba' to split sentences into words. 'jieba' is good at Chinese word segmentation and has great advantages in identifying texts like finance and national policies (Guoju, 2021). Then the deactivation form of Harbin Institute of Technology is employed to filter the punctuation marks and the words without semantic meaning.

3.2.3 Transformation of word vectors

A word dictionary will be constructed based on the separated words. Each word is numbered based on their term frequency. The more frequently the word appears, the smaller the number obtained. Next, the Word2vec model is applied to train the word embedding, which can represent texts as word vectors and obtain sentence vectors with the specific length. The specific parameter settings during word vector training are displayed in Table 2.

3.2.4 Data normalisation

For the collected data, each attribute has different magnitude units. If we directly use them, the training effect of model will be influenced. Data normalisation can map the data to the interval [0–1], which transform the data in different scales into data with the same scale. In this paper, we adopt Min-Max normalisation method to make data normalisation. And equation (10) shows the principle of it.

$$x_{\mathrm{n}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{10}$$

where x_{n} is the normalised data, $x_{\min}$ and $x_{\max}$ refer to the minimal value and maximal value of sample data.

3.3 Model evaluation index

The experimental environment of this paper is carried out under Windows 10 operating system, and other hardware configurations are AMD Ryzen 5 3500U processor, AMD Radeon (TM) Vega 8 Graphics integrated graphics card. All algorithms for the experiments are written in Python 3.7 language and Tensorflow2.0 Deep Learning Framework. In this paper, we take MSE as the loss function during model training, and equation (11) reveals the calculation process.

$$F_{\mathrm{loss}} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 \tag{11}$$

where y_i means the observed value; $\hat{y}_i$ means the forecasting value of methanol futures settlement price obtained by model.

For the evaluation of model's prediction effect, we chose three common error indicators in the regression task, namely, mean absolute error (MAE), root mean square error (RMSE,) and mean absolute percentage error (MAPE), calculated as displayed in equation (12), equation (13) and equation (14).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{y}_i - y_i\right| \tag{12}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2} \tag{13}$$

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \tag{14}$$

In addition, in order to more intuitively observe the effect of text on the forecasting precision, the percentage reduction of error before and after text addition is calculated for three error metrics, MAE, RMSE and MAPE. The specific calculation formula is exhibited in equation (15)–equation (17).

$$ERP_{MAE} = \left(A_i - A_j\right) / A_i * 100 \quad (15)$$

$$ERP_{RMSE} = \left(R_i - R_j\right) / R_i * 100 \quad (16)$$

$$ERP_{MAPE} = \left(P_i - P_j\right) / P_i * 100 \quad (17)$$

where ERP_{MAE}, ERP_{RMSE} and ERP_{MAPE} denote the percentage error reduction of the MAE, RMSE and MAPE metrics, respectively. A_i, R_i and P_i refer to the values of MAE, RMSE, and MAPE indicators without texts, separately. A_j, R_j and P_j represent the values of MAE, RMSE, and MAPE after the addition of texts, respectively. If ERP is positive, it means that prediction accuracy of model after adding texts is higher than that before, and the text has an improvement effect on forecasting precision.

Table 2 The parameter settings of MFCGAM model

Training models	*Network layers*	*Parameter names*	*Values/rules*
The Word2Vec training model		num_features	50
		min_word_count	3
		Context	4
		maxLen	15
Forecasting model	CNN layer	Filters	64
		kernel_size	1
		Activation	Sigmoid
		pool_size	5
	GRU layer	gru_units	64
		Activation	relu
	Dense layer	Number of output categories	16
		Activation	relu
	Attention layer	Dense	16
		Activation	Softmax
	Dropout layer	Dropout	0.01
	Dense layer	Number of output categories	1
		Activation	relu
		Epoch	100
		batch_size	16

Table 3 Loss function values of different methods combinations on three datasets

Model combinations	*Research reports*			*Information*			*Investor comments*		
	MAE	*RMSE*	*MAPE*	*MAE*	*RMSE*	*MAPE*	*MAE*	*RMSE*	*MAPE*
CNN+CNN	202.9613	248.5127	8.1716	201.7379	259.7566	7.4147	173.9823	229.9332	6.2154
CNN+LSTM	80.5128	102.3451	3.0488	57.769	72.4419	2.1941	49.8578	60.7484	1.8754
CNN+GRU(MFCGAM)	*61.4452*	*83.4694*	*2.4221*	*50.5339*	*64.1587*	*1.8659*	*41.5706*	*53.196*	*1.5318*
LSTM+CNN	193.4503	238.4599	7.6793	199.461	267.4252	7.5288	175.0003	218.9104	7.0792
LSTM+LSTM	82.0243	114.3692	3.2271	70.4599	94.5232	2.6062	51.7593	70.2199	1.9308
LSTM+GRU	84.7896	112.0672	3.354	125.9647	200.5227	4.0872	60.7814	77.9387	2.1211
GRU+CNN	204.8071	260.9747	7.9824	196.8975	267.1542	7.7691	188.2779	240.2531	7.2788
GRU+LSTM	83.4876	110.6608	3.3263	70.4386	98.8435	2.4799	52.0896	70.0139	1.9546
GRU+GRU	64.124	87.3414	2.5554	56.6364	72.9333	2.1201	49.7429	59.4668	1.8951

Table 4 Variance of loss function values got by diverse methods combinations on three datasets

Model combinations	*Research reports*			*Information*			*Investor comments*		
	MAE	*RMSE*	*MAPE*	*MAE*	*RMSE*	*MAPE*	*MAE*	*RMSE*	*MAPE*
CNN+CNN	233.6314	356.0359	0.3632	220.2118	245.1703	0.3592	268.159	335.6764	0.2712
CNN+LSTM	169.5077	333.2193	0.1906	304.908	272.7822	0.2587	210.5723	256.0139	0.3257
CNN+GRU(MFCGAM)	*93.0960*	*148.9000*	*0.1626*	*95.4935*	*94.1675*	*0.2010*	*198.0483*	*205.9433*	*0.1170*

Table 4 Variance of loss function values got by diverse methods combinations on three datasets (continued)

Model combinations	*Research reports*			*Information*			*Investor comments*		
	MAE	*RMSE*	*MAPE*	*MAE*	*RMSE*	*MAPE*	*MAE*	*RMSE*	*MAPE*
LSTM+CNN	499.6919	467.8007	0.6097	163.0167	269.8543	0.5905	238.5610	233.2170	0.3510
LSTM+LSTM	193.0960	331.6749	0.2028	211.7020	299.5117	0.4152	260.2244	404.4642	0.3739
LSTM+GRU	268.0169	333.056	0.3627	559.9502	579.6106	0.5277	263.6036	378.3034	0.3276
GRU+CNN	349.5658	488.5665	0.551	334.8329	486.4953	0.7097	198.0483	370.5265	0.5033
GRU+LSTM	112.5521	214.5633	0.2406	160.2552	256.9421	0.2793	261.4511	348.7809	0.3195
GRU+GRU	158.3152	229.7249	0.2313	192.2503	348.5505	0.2054	200.5345	203.6929	0.2809

Table 5 Mean errors gained by different models on three datasets

Datasets	*Models*	*No texts*			*Add texts*		
		MAE	*RMSE*	*MAPE*	*MAE*	*RMSE*	*MAPE*
Research reports	BPNN	366.0296	360.6692	10.5674	321.6663	310.2853	9.3608
	SVM	376.1925	412.2471	12.1094	348.9725	390.6425	10.9112
	LSTM	94.0858	119.3622	3.5121	86.6312	108.3777	3.1506
	GRU	90.6298	111.7545	3.2010	81.5115	103.5103	2.8646
	CNN+LSTM	92.3532	113.9099	3.4805	80.5128	100.3451	3.0488
	GRU+GRU	76.7808	98.5353	2.9679	64.1240	87.3414	2.5554
	N-ATCN	92.8895	114.5086	3.5146	81.0828	104.2740	3.1038
	TBF	98.9738	121.3631	3.7633	81.7754	100.6190	3.1844
	VADERLSTM	100.5116	125.1528	3.7807	88.4003	108.5105	3.2483
	MFCGAM	*71.0170*	*96.8325*	*2.8012*	*61.4452*	*83.4694*	*2.4221*
Information	BPNN	255.3723	311.5127	7.8738	230.9971	284.7971	7.2565
	SVM	281.5078	332.4682	8.7923	265.1378	317.4229	8.2059
	LSTM	66.7636	89.8787	2.2515	62.4456	83.0383	2.1021
	GRU	66.2763	86.8363	2.3535	61.6005	82.7609	2.2014
	CNN+LSTM	62.1048	82.4589	2.2495	56.7690	74.4419	2.1041
	GRU+GRU	61.8098	79.8026	2.2980	54.6364	73.9333	2.0201
	N-ATCN	64.6966	80.0403	2.3185	57.1294	73.4204	2.0927
	TBF	69.9983	91.3772	2.5246	58.0562	76.8633	2.2259
	VADERLSTM	78.3273	100.8772	2.7027	69.5657	90.4774	2.4049
	MFCGAM	*56.6269*	*73.4568*	*2.1749*	*50.5339*	*64.1587*	*1.8859*
Investor comments	BPNN	321.0394	351.4637	9.8625	316.5389	350.6993	9.5665
	SVM	299.9708	334.0483	9.5006	288.1593	324.3590	9.1471
	LSTM	61.3925	84.5379	2.0718	58.4554	78.1563	1.9913
	GRU	61.8002	79.2083	1.9937	57.6944	75.7251	1.8970
	CNN+LSTM	52.3364	69.0245	1.9335	48.8578	62.7484	1.8254
	GRU+GRU	55.6876	64.2366	2.0197	49.7429	60.4668	1.7951
	N-ATCN	55.1250	70.0517	1.8976	49.1159	65.7258	1.7634
	TBF	56.9610	76.4870	1.9448	55.6573	70.3214	1.9061
	VADERLSTM	85.5702	99.6373	2.9346	76.4356	92.1788	2.7456
	MFCGAM	*42.0163*	*53.1960*	*1.5738*	*41.5706*	*51.9400*	*1.5318*

Table 6 Error reduction percentage obtained by different models when adding three texts

Models	*MAE*			*RMSE*			*MAPE*		
	Research reports	*Information*	*Investor comments*	*Research reports*	*Information*	*Investor comments*	*Research reports*	*Information*	*Investor comments*
BPNN	12.1201	9.5450	1.4019	13.9696	8.5761	0.2175	11.4181	7.8399	3.0013
SVM	7.2357	5.8151	3.9375	5.2407	4.5253	2.9006	9.8948	6.6695	3.7208
LSTM	7.9232	6.4676	4.7841	9.2027	7.6107	7.5488	10.2930	6.6356	3.8855
GRU	10.0610	7.0550	6.6437	7.3771	4.6932	4.3975	10.5092	6.4627	4.8503
CNN-LSTM	12.8208	8.5916	6.6466	11.9084	9.7224	9.0926	12.4034	6.4637	5.5909
GRU-GRU	16.4843	11.6056	10.6751	11.3603	7.3548	5.8686	13.8987	12.0931	11.1205
N-ATCN	12.7105	11.6964	10.9009	8.9378	8.2707	6.1753	11.6884	9.7391	7.0721
TBF	17.3767	17.0606	2.2888	17.0926	15.8835	8.0610	15.3828	11.8316	1.9899
VADERLSTM	12.0497	11.1859	10.6750	13.2976	10.3094	7.4857	14.0820	11.0186	6.4404
MFCGAM	13.4782	10.7599	1.0608	13.8002	12.6579	2.3611	13.5335	13.2880	2.6687

Table 7 Error reduction percentage obtained under different time windows when adding three texts

Time windows	*Research reports*			*Information*			*Investor comments*		
	MAE	*RMSE*	*MAPE*	*MAE*	*RMSE*	*MAPE*	*MAE*	*RMSE*	*MAPE*
3	10.792	14.83144	6.4366	3.1034	11.1452	4.0094	2.8678	3.8168	4.3065
5	13.4783	13.8002	13.5347	*10.7599*	*12.6579*	*14.2043*	3.7626	9.6053	3.7953
10	*22.0605*	*15.77318*	*16.2716*	1.1159	6.6872	0.2785	*10.2346*	*8.8225*	*10.951*
15	13.1875	11.0871	10.5321	2.6527	5.1364	2.2767	6.994	1.9355	7.2811
20	7.1714	1.8287	9.4783	10.0022	3.3259	10.7956	0.1371	3.5776	–0.9899
25	12.0743	4.9749	11.3827	7.8813	9.1449	9.3347	*–6.981*	*–16.1799*	*–1.6042*

Table 8 T-test results of MFCGAM method paired with other models on three datasets

Models	*Research reports*			*Information*			*Investor comments*		
	P-values (MAE)	*P-values (RMSE)*	*P-values (MAPE)*	*P-values (MAE)*	*P-values (RMSE)*	*P-values (MAPE)*	*P-values (MAE)*	*P-values (RMSE)*	*P-values (MAPE)*
BPNN → MFCGAM	1.43E-37	1.32E-40	3.15E-36	1.21E-25	1.11E-29	1.22E-23	1.42E-37	1.32E-40	3.15E-36
SVM → MFCGAM	3.88E-43	2.18E-44	4.97E-40	9.92E-41	4.28E-42	6.39E-36	3.88E-43	2.18E-44	4.97E-40
LSTM → MFCGAM	0.0067	0.0007	0.0429	0.0194	0.0005	0.0339	0.0068	0.0007	0.0429
GRU → MFCGAM	0.0027	3.49E-05	0.0246	0.0100	0.0020	0.0389	0.0027	3.49E-05	0.0246
CNN_LSTM → MFCGAM	0.0170	0.0105	0.0124	0.01413	0.0013	0.0140	0.0129	0.0105	0.0122
GRU_GRU → MFCGAM	0.0433	0.0286	0.04929	0.0433	0.0044	0.0495	0.0351	0.0286	0.0493
N-ATCN → MFCGAM	5.31E-07	1.43E-07	8.84E-08	0.0202	0.0098	0.0067	0.0184	0.0275	0.0270
TBF → MFCGAM	1.30E-06	7.11E-08	1.56E-07	0.0134	0.0168	0.0173	0.0233	0.0082	0.0061
VADERLSTM → MFCGAM	1.72E-08	5.65E-10	8.42E-09	0.0041	0.0007	0.0126	2.09E-05	1.11E-05	3.52E-05

Note: If $p < 0.05$, the result means significant.

3.4 Experimental parameter setting

The experimental parameters in this paper involve two models: the Word2Vec word vector training model and the predictive model based on deep learning networks. All parameter settings of MFCGAM model are listed in Table 2. 'num_features' is the dimensionality of generated word vectors. 'min_word_count' refers to the minimum frequency of words, and the words below this frequency do not train. 'context' specifies the size of the word context window; 'maxLen' assigns a uniform length for the sequences input to prediction network. The sentence vectors less than this length are complemented by 0, while sentence vectors longer than this length are cut off.

For raising training speed and effect of model, this paper employs Adam optimisation algorithm. The algorithm can set the learning rate adaptively by calculating the first-order moment estimates and second-order moment estimates of the gradients. By virtue of its faster convergence speed and excellent performance in dealing with non-stationary sequences, Adam optimisation algorithm is extensively applied in optimising deep learning model training (Kingma and Ba, 2014).

3.5 The presentation and analysis of experimental results

3.5.1 The determination of prediction model

Since a dual-channel feature extraction module is developed in our prediction model, which methods are the best for 'text-quantity' features extraction is the first matter that should to be settled. Therefore, we first conduct experiments to determine forecasting model by arranging the combination of CNN, LSTM, and GRU methods, and testing them on three datasets. The combination with the lowest error and the lowest variance is our final prediction model, which will be used in subsequent experiments. The experimental results on three datasets are shown in Table 3, and Table 4.

In the column of 'model combinations', 'CNN+LSTM' means that CNN is used to extract text features and LSTM is used for mining numerical characteristics, and the meanings of other 8 combinations are the similar as above. From Table 3 and Table 4, it can be seen that 'CNN+GRU' model wins the smallest error and variance on three indicators of MAE, RMSE and MAPE. Therefore, we decide to use 'CNN+GRU' combination as the prediction model in this paper. In other words, CNN method is selected to excavate textual features from three text datasets, and GRU method is utilised to extract numerical characteristics from methanol prices. We abbreviate this forecasting model as MFCGAM.

Furthermore, it may be worth thinking about that why 'CNN+GRU' can achieve the best performance. A sentence is composed of multiple words. For text data, the spatial relevance between words and sentences is more critical than temporal relevance. And CNN is just a method better at extracting spatial features. Although both LSTM and GRU are RNN, the internal structure of GRU is simpler than that of LSTM. The complex gating mechanism of LSTM easily leads to over-fitting problem, which is often caused by the use of overly complex models in a small amount of data. In our work, the data volume of dataset is not too large, and the number of attributes is moderate, so the effect of GRU training is better than that of LSTM training.

3.5.2 The effect of text attributes on prediction results

To verify that the inclusion of texts is helpful for forecasting methanol futures price, in this section we select the three most prominent performing models from Section 3.5.1 (including CNN+LSTM, GRU+GRU, and MFCGAM), some basic machine learning models (including BPNN and SVM), some single deep learning methods (including GRU and LSTM), and three recent excellent algorithms (including ATCN, TBF and VADER) to carry out comparative experiment.

Table 5 displays mean errors gained by different models on three datasets. As can be found from Table 5, regardless of the type of text added, ten methods reveal reductions on the values of three error indicators (MAE, RMSE, and MAPE) after adding text attributes. It implies that considering the text attribute in prediction model can help improve models' forecasting capability. From the error of each model, two machine learning models, BPNN and SVM, acquire the largest error, and MFCGAM model win the smallest error. It illustrates that simple machine learning methods are limited in multimodal price prediction, and MFCGAM proposed in this paper possesses the supreme prediction accuracy.

For N-ATCN (Hongli et al., 2021), TBF (Li et al., 2022) and VADERLSTM (Zhao et al., 2020; Yan et al., 2020) models in Table 5, they all use sentiment analysis technology to quantify texts. However, we directly input the word vector into prediction network. It can be seen that the errors of them are higher than MFCGAM model. Word vector is another form of text, which is unquantised and contains rich original text information, while the information in emotional value obtained by sentiment analysis is relatively single. Neural network is good at autonomously learning. Thus, unquantised text can help model learn more useful information and improve the performance of model.

In order to enhance the comparability among models, the percentage reduction of errors is calculated according to equation (15)–equation (17), which is shown in Table 6. By this way, we can more intuitively observe the help of texts on decreasing prediction error. We can observe that all models achieve the smallest error reduction percentage on 'investor comments' dataset, followed by 'information' dataset, and the largest is 'research reports' dataset. Research reports are the analysis of current market by professional institutions. The information is some important news and events about methanol industry. And investor comments are the social discussions of investors on the forum, which mainly reflect investors' sentiments. The experimental results in Table 6 make known that research reports of professional institutions have greater contributions to the increase of price prediction accuracy, while the help of investor sentiments to heighten forecasting precision is not so obvious.

3.5.3 The analysis of time window experimental results

To investigate the influence of different time windows on prediction results, we conduct time window experiments on three datasets for MFCGAM model in this section. Since the futures market does not open on weekends, we regard five days as a week and 20 days as a month. The predicting tasks can be classified into short-term, medium-term and long-term prediction. The time windows corresponding to three types are 3 and 5, 10 and 15, 20 and 25, respectively.

The error reduction percentages are calculated, which exhibit in Table 7. Three texts win the highest percentage of error reduction when time windows are set to 5 or 10. And 5 and 10 are the best prediction time window for three texts. Moreover, it is worth noting that for 'investor comments' dataset, when time window is taken as 25, the error reduction percentages are negative. It means that the prediction error not only does not decrease but also increase. Investor comments are highly subjective, which mainly related to investors' sentiment and opinions in current moment, while the stock futures market is changing fast. So, if we use investor comments data from too long ago to predict, it may interfere with the forecasting results. We can consider that it is not so easy to achieve exact forecast of methanol futures price in the long term.

3.5.4 Significance analysis of prediction results

T-test method is a statistical method applicable to small sample data, which is used for verifying whether there is a significant difference between two groups of data. Given that the results in this paper are taken from the average values of 20 runs results, this section conducts t-test. We pair the nine comparison models in Table 5 with MFCGAM method for t-test and set the significance level at 0.05. In other words, if the p-values of three error indicators are less than 0.05, then the results of t-test are significant, which means that the results of 20 runs of two models are prominently discrepant. And Table 8 lists the results of t-test. As can be seen, whether for the 'research report', the 'Information', or the 'Investor comments', the P-values of MFCGAM and the other nine models on MAE, RMSE, and MAPE indicators are less than 0.05, which indicates the t-test results are all significant. From this, we can further hold that the prediction results of MFCGAM model are noticeably diverse from those of the other nine comparison models. The conclusion that the MFCGAM model has the highest prediction accuracy among the models compared in this paper holds true.

4 Conclusions

In order to predict methanol futures prices accurately, a novel methanol price forecasting method ground on multimodal fusion using CNN, GRU and attention mechanism is put forward, which is abbreviated as MFCGAM. First, the quantity and the text features related to methanol are comprehensively extracted using CNN and GRU. Then, the two types of features are integrated utilising attention mechanism for achieving the 'text-quantity' fusion features, which contributes to gain good forecasting results. Through experimenting on three real datasets, it can be concluded that CNN outperforms LSTM and GRU in mining text features. Compared with ordinary machine learning methods and deep learning networks, the proposed MFCGAM can effectively improve the predictive ability of methanol futures price. By contrasting three types of texts, we find that the inclusion of 'research report' is the most helpful in reducing forecasting error. Meanwhile, the results of time window experiments on three datasets show that the performances of the short-term and medium-term prediction are evidently better than the long-term forecasting. On the whole, it provides a novel approach for achieving more accurate methanol price forecasting, and contributes to the implementation of intelligent scheduling in smart plants. In future research, we will consider employing efficient data pre-processing methods to reduce the noise information in raw data, so as to further improve the predictive capability of model in long-term forecasting tasks.

Acknowledgements

This work was supported by the Fundamental Research Funds for the Central Universities of China Grant No. PA2024GDSK0105, No. PA2024GDGP0031, No. PA2023IISL0093 and No. PA2023IISL0092, the National Nature Science Foundation of China under Grant No. 91546108, No. 72071060, No. 72171068 and No. 72171073, the Anhui Provincial Natural Science Foundation under Grant No. 1908085QG298, the Anhui Provincial Science and Technology Major Projects No. 201903a05020020, and the Open Research Fund Program of Key Laboratory of Process Optimisation and Intelligent Decision-making (Hefei University of Technology), Ministry of Education.

References

Bahdanau, D., Cho, K. and Bengio, Y. (2014) 'Neural machine translation by jointly learning to align and translate', *Computer Science*, DOI: 10.48550/arXiv.1409.0473.

Bayoudh, K., Knani, R. and Hamdaoui, F. (2021) 'A survey on deep multimodal learning for computer vision: advances, trends, applications, and datasets', *The Visual Computer*, DOI: 10.1007/s00371-021-02166-7.

Bing, L. (2021) 'Opinion mining and sentiment analysis', *Data-Centric Systems and Applications*, pp.459–526, Springer, Berlin, Heidelberg, DOI: 10.1007/978-3-642-19460-3_11.

Bojanowski, P., Grave, E. and Joulin, A. (2017) 'Enriching word vectors with subword information', *Transactions of the Association for Computational Linguistics*, Vol. 5, pp.135–146, DOI: 10.48550/arXiv.1607.0460.

Cheng, D.W., Yang, F.Z., Xiang, S. et al. (2021) 'Financial time series forecasting with multi-modality graph neural network', *Pattern Recognition*, Vol. 121, DOI: 10.1016/j.patcog.2021.108218.

Chung, J., Gulcehre, C. and Cho, K.H. (2014) 'Empirical evaluation of gated recurrent neural networks on sequence modeling', Eprint Arxiv, DOI: https://arxiv.org/abs/1412.3555.

Feng, C., Elazab, A. and Yang, P. (2019) 'Deep learning framework for Alzheimer's disease diagnosis via 3D-CNN and FSBi-LSTM', *IEEE Access*, Vol. 7, pp.63605–63618, DOI: 10.1109/ACCESS.2019.2913847.

Guoju, S.H. (2021) 'Research on Chinese word segmentation technology based on Python', *Wireless Internet Technology*, Vol. 18, No. 23, pp.110–111.

Hongli, T., Shuo, J. and Huiqiang, Y. (2021) 'A stock forecasting method based on TCN and news emotion', *Computer Application*, Vol. 48, No. 6, pp.17–21, DOI: 10.3969 /j.issn.1672-9528.2021.06.003.

Hu, H. and Li, Q. (2021) 'Parallel deep learning algorithms with hybrid attention mechanism for image segmentation of lung tumors', *IEEE Transactions on Industrial Informatics*, Vol. 17, No. 4, pp.2880–2889, DOI: 10.1109/TII .2020.3022912.

Hu, T.T., Feng, Y.Q. and Shen, L.J. (2019) 'The salient feature selection by attention mechanism based LSTM in speech emotion recognition', *Technical Acoustics*, Vol. 38, pp.414–421, DOI: 10.16300/j.cnki.1000-3630.2019.04.010.

Huang, F., Zhang, X. and Zhao, Z. (2019) 'Image-text sentiment analysis via deep multimodal attentive fusion', *Knowledge-Based Systems*, Vol. 167, pp.26–37, DOI: 10.1016/j.knosys.2019.01.019.

Huang, J. and Liu, J. (2020) 'Using social media mining technology to improve stock price forecast accuracy', *Journal of Forecasting*, Vol. 39, No. 1, pp.104–116, DOI: 10.1002/for.2616.

Kang, H.W. and Kang, H.B. (2017) 'Prediction of crime occurrence from multi-modal data using deep learning', *Plos One*, Vol. 12, No. 4, p.e0176244, DOI: 10.1371/journal .pone.0176244.

Kang, L., Song, G. and Peiyuan, Q. (2017) 'Road2Vec: measuring traffic interactions in urban road system from massive travel routes', *International Journal of Geo-Information*, Vol. 6, No. 11, p.321, DOI: 10.3390/ijgi6110321.

Kilimci, Z.H. (2019) 'Financial sentiment analysis with deep ensemble models (DEMs) for stock market prediction', *Journal of the Faculty of Engineering and Architecture of Gazi University*, Vol. 35, No. 2, pp.635–650, DOI: 10.17341/gazimmfd.501551.

Kingma, D. and Ba, J. (2014) 'Adam: a method for stochastic optimization', *Computer Science*, DOI: arxiv.org/abs/ 1412.6980.

Li, J., Li, G. and Liu, M. (2022) 'A novel text-based framework for forecasting agricultural futures using massive online news headlines', *International Journal of Forecasting*, Vol. 38, No. 1, pp.35–50, DOI: 10.1016/j.ijforecast.2020.02.002.

Li, X., Shang, W. and Wang, S. (2018) 'Text-based crude oil price forecasting: a deep learning approach', *International Journal of Forecasting*, Vol. 35, No. 4, pp.1548–1560, DOI: 10.1016 /j.ijforecast.2018.07.006.

Loginova, E., Tsang, W.K. and Heijningen, G.V. (2021) 'Forecasting directional bitcoin price returns using aspect-based sentiment analysis on online text data', *Machine Learning*, pp.1–24, DOI: 10.1007/s10994-021-06095-3.

Mardianto, I., Gunawan, M.I. and Sugiarto, D. (2020) 'Comparison of rice price forecasting using the ARIMA method on Amazon Forecast and Sagemaker', *Rekayasa Sistem Dan Teknologi Informasi*, Vol. 4, No. 3, pp.537–543, DOI: 10.29207/resti.v4i3.1902.

Mikolov, T., Chen, K. and Corrado, G. (2013) 'Efficient estimation of word representations in vector space', *Computer Science*, DOI: 10.48550/arXiv.1301.3781.

Nie, W.Z., Yan, Y. and Song, D. (2020) 'Multi-modal feature fusion based on multi-layers LSTM for video emotion recognition', *Multimedia Tools and Applications*, Vol. 80, No. 11, pp.16205–16214, DOI: 10.1007/s11042-020-08796-8.

Niu, H., Xu, K. and Liu, C. (2021) 'A decomposition-ensemble model with regrouping method and attention-based gated recurrent unit network for energy price prediction', *Energy*, Vol. 231, p.120941, DOI: 10.1016/j.energy.2021.120941.

Peiguang, L., Jiaqian, Z. and Yulian, W. (2020) 'SCONV: a financial market trend forecast method based on emotional analysis', *Journal of Computer Research and Development*, Vol. 57, No. 8, pp.1769–1778, DOI: 10.7544/issn1000-1239.2020.20200494.

Raja, P. and Thangavel, K. (2020) 'Missing value imputation using unsupervised machine learning techniques', *Soft Computing*, Vol. 24, pp.4361–4392, DOI: 10.1007/s00500-019-04199-6.

Rajamani, S.T., Rajamani, K.T. and Mallol-Ragolta, A. (2021) 'A novel attention-based gated recurrent unit and its efficacy in speech emotion recognition', *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), ELECTR NETWORK*, pp.6294–6298, DOI: 10.1109/ICASSP39728.2021.9414489.

Ramachandram, D. and Taylor, G.W. (2017) 'Deep multimodal learning: a survey on recent advances and trends', *IEEE Signal Processing Magazine*, Vol. 34, pp.96–108.

Rezaei, H., Faaljou, H. and Mansourfar, G. (2020) 'Stock price prediction using deep learning and frequency decomposition', *Expert Systems with Applications*, Vol. 169, p.114332, DOI: 10.1016/j.eswa.2020.114332.

Selimi, M. and Besimi, A. (2019) 'A proposed model for stock price prediction based on financial news', *ENTRENOVA Conference Proceedings*, Rovinj, Croatia, DOI: 10.2139/ ssrn.3490159.

Sumarsono, A. (2020) 'Application of RXD algorithm to word vector representation for keyword identification', *2020 10th Annual Computing and Communication Workshop and Conference (CCWC)*, Univ. Nevada, Las Vegas, CA, DOI: 10.1109/CCWC47524.2020.9031139.

Vanstone, B.J., Gepp, A. and Harris, G. (2019) 'Do news and sentiment play a role in stock price prediction', *Applied Intelligence*, Vol. 49, pp.3815–3820, DOI: 10.1007/s10489-019-01458-9.

Wang, R. and Li, C. (2019) 'Deep learning method based on gated recurrent unit and variational mode decomposition for short-term wind power interval prediction, *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 31, No. 10, pp.3814–3827, DOI: 10.1109/TNNLS.2019.2946414.

Wang, Y., Chen, Q. and Gan, D. (2018) 'Deep learning-based socio-demographic information identification from smart meter data', *IEEE Transactions on Smart Grid*, Vol. 10, No. 3, pp.2593–2602, DOI: 10.1109/TSG.2018.2805723.

Wang, M. (2022) 'Prediction of the technology company's stock price through the deep learning method', *Open Journal of Modelling and Simulation*, Vol. 10, No. 4, pp.428–440, DOI:10.4236/ojmsi.2022.104024.

Xiao, C., Xia, W. and Jiang, J. (2019) 'Stock price forecast based on combined model of ARI-MA-LS-SVM, *Neural Computing and Applications*, Vol. 32, No. 10, pp.5379–5388, DOI: 10.1007/s00521-019-04698-5.

Xiao, Y., Liu, L. and Xu, M. (2018) 'GLDA-FP: Gaussian LDA model for forward prediction', *7th International Congress on Big Data (BigData Congress) Held as Part of the Services Conference Federation (SCF)*, DOI: 10.1007/978-3-319-94301-5_10.

Xiong, Y., Chen, S. and Qin, H. (2020) 'Distributed representation and one-hot representation fusion with gated network for clinical semantic textual similarity', *BMC Medical Informatics and Decision Making*, Vol. 20, p.72, DOI: 10.1186/s12911-020-1045-z.

Yan, H., Jian, N. and Liu, W. (2020) 'A hybrid deep learning approach by integrating LSTM-ANN networks with GARCH model for copper price volatility prediction', *Physica A*, Vol. 557, No. 1, p.124907, DOI: 10.1016/j.physa.2020.124907.

Zeiler, M. and Fergus, R. (2013) 'Visualizing and understanding convolutional neural networks', *Lecture Notes in Computer Science*, Vol. 8689, pp.818–833, DOI: 10.1007/978-3-319-10590-53.

Zhang, B.A., Xiong, D.Y. and Xie, J. (2020) 'Neural machine translation with GRU-gated attention model', *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 31, No. 11, pp.4688–4698, DOI: 10.1109/TNNLS.2019.2957276.

Zhao, L.T., Zeng, G.R. and Wang, W.J. (2020) 'Forecasting oil price using web-based sentiment analysis', *Energies*, Vol. 12, No. 22, p.4291, DOI: 10.3390/en12224291.

Notes

1 Website of numerical data source: AI media data centre, https://data.iimedia.cn/page-category.jsp.

2 Website of text data source: http://guba.eastmoney.com.