A hybrid attention and time series network for enterprise sales forecasting under digital management and edge computing

Xi Zhang and Taesun Kim[*]

Abstract

Enterprises have both new opportunities and new challenges as a result of the rapid advancements in information technology that have accompanied the age of economic globalization. With the growth of internet of Things devices, data sizes have significantly increased. Further, the traditional cloud platform has been enriched with edge computing so that the huge data can be processed where it is collected. Therefore, businesses must adapt to new size requirements and rising standards for technical content. Forecasting corporate sales has emerged as a hot topic in the field of digital management. To successfully direct the future production and existence of enterprises, time series forecasting is of utmost importance and value. This is because it makes use of already-existing data to get the best predicting result. This work proposes a combination of enterprise sales forecasting from the perspective of digital management and neural networks, and proposes a network HATT-CNN-BiLSTM model for enterprise sales forecasting. First, this work combines multi-scale CNN (MSCNN) with improved BiLSTM (IBiLSTM) model. The MSCNN is utilized to extract spatial features with different scale, and it is often impossible to effectively explore the rules of time series features, and the processing of time series data is the strength of the LSTM network. Moreover, the IBiLSTM model can explore time series features in both directions, and therefore more useful information can be obtained. The MSCNN-IBiLSTM model, which is composed of MSCNN and IBiLSTM, can take advantage of strengths and avoid weaknesses, and give full play to the roles of the two models in different fields. Second, this work proposes a hybrid attention mechanism that combines self-attention, channel attention, and spatial attention. It enhances features extracted by MSCNN-IBiLSTM through a hybrid attention to build HATT-MSCNN-IBiLSTM network, which can extract more discriminative features. Third, this work conducts comprehensive and systematic experiments on HATT- MSCNN-IBiLSTM to verify feasibility of the proposed method. The proposed model is implemented over an edge computing platform that increases the model training speed and improve the response time.

Keywords Enterprise sales forecast, Hybrid attention, Time series network, Digital management, Cloud-edge model

Introduction

All parts of social life have been impacted by the proliferation of computer technology and network communication technologies, which have transformed people's professional and personal lives to unprecedented extents. Modern information technology is playing an increasingly crucial role in business management. It is becoming increasingly common for businesses to use information technology into the production and operation management process, and this trend opens up new avenues for advancing enterprise management theory, methodologies, and tools. Due to the current state of overproduction brought on by the growth of social economy, industry competitiveness is getting increasingly harsh. In

*Correspondence:
Taesun Kim
taesunkim@163.com; taesunk@hanyang.ac.kr
College of Design, Hanyang University, ERICA Campus, Ansan-si, Republic of Korea

order to obtain better corporate benefits, enterprise managers have to start paying attention to how to improve employee work, reduce costs, and improve corporate benefits and competitiveness. More and more enterprise managers realize that the application of digital management is one of the most significant ways to address these problems [1–5].

Enterprise information system is the comprehensive application for information technology in production as well as operation management, and it is a tool to realize transformation of traditional industries and promote the efficiency and effect of enterprise management. It has become an important development trend of enterprise management reform to realize the transformation of enterprise production and management through the construction of information system. Enterprise informatization construction has been carried out in an all-round way in the business world, but the paradox of benefit of informatization construction still plagues enterprises and their managers who are or plan to carry out informatization construction. In the technical aspects of enterprise informatization construction and information system construction, a relatively mature theoretical system with information engineering as the core has been formed. It includes the analysis, development, implementation and maintenance of information systems involving the entire life cycle of the methodology of the information system.

The theoretical research of enterprise information system should be traced back to the source, starting from enterprise management, and taking enterprise management reform under modern information technology as its theoretical basis from the perspective of management technology and management tools. Enterprise information system is a significant technical means for enterprise management. The demand for enterprise management reform drives the construction of information system, and enterprise information system serves enterprise management. The theory of information system construction is subordinate to the theory of enterprise management. Therefore, only by studying the construction of enterprise information system from the level of enterprise management theory, can we find the best theoretical guidance for the construction of information system [6–10].

Modern information technology has made possible for researchers to focus on the role of information technology in enterprise management, as well as how information technology may be used to improve enterprise management. These studies are mainly practice-oriented, and many related studies are empirical generalizations or specific methods, basically focusing on applied research, which has played a certain guiding role in the management practice of enterprises. However, there is a lack of general and systematic theoretical refinement. Enterprise digital management is an innovation of modern enterprise management theory and its application. The introduction of the enterprise digital management model provides a broader space for traditional management to change, and also provides a way for many management innovation theories to be realized. Information technology has penetrated into every corner of management, and management is full of digital flavor from head to toe. Digitalization softens the hardness of management, liberates the elements of management that were originally restricted by time and space factors, and further adheres the original relatively isolated relationship between management elements. The synergy between the management elements has been brought into play more effectively. As a non-linear open complex system, the management system has caused a sudden change in traditional management after the fluctuation of factors such as hardness and adhesion between elements. The management after the mutation, that is, digital management, has realize a fundamental change from inorganic management to organic management, showing the characteristics that traditional management does not have [11–15].

Because digital management has brought many positive effects to enterprises, it is a very important subject to forecast and evaluate enterprise sales from the perspective of digital management, which is also a subject closely related to time series. With development for science, people's life has many possibilities and many unknowns. Whether it is an individual or an enterprise, they all hope to have a certain forecast or control over the future, so as to limit the losses to a controllable range as much as possible. Or the relevant personnel can take action in advance to obtain benefits based on the predicted trend. In the academic field, the research on time series has never stopped. At present, time series has become a kind of data science, guides people's production and life through the data generated by the prediction method, so as to carry out reasonable resource allocation and scheduling. Sales forecasting is very important for enterprises. A reasonable sales forecast can affect the production, procurement, distribution and inventory of enterprises, and optimize supply chain management for enterprises [16–20].

This work proposes a combination of enterprise sales forecasting from the perspective of digital management and neural networks, and proposes a network HATT-CNN-BiLSTM model for enterprise sales forecasting. First, this work combines MSCNN with IBiLSTM models. The MSCNN model is mainly used to extract spatial features with different scale, while IBiLSTM can explore time series features in both directions, and more useful information can be obtained. The MSCNN-IBiLSTM,

which is composed of MSCNN and IBiLSTM, can take advantage of strengths and avoid weaknesses, and give full play to the roles of the two models in different fields. Second, this work proposes a hybrid attention mechanism that combines self-attention, channel attention, and spatial attention mechanisms. It improves features extracted by the MSCNN-IBiLSTM model through a hybrid attention module to build HATT- MSCNN-IBiL-STM network, which can extract more discriminative features. Third, this work conducts comprehensive and systematic experiments on HATT- MSCNN-IBiLSTM to verify the feasibility and superiority. The proposed model is implemented over an edge computing platform that increases the model training speed and improve the response time. The following are the major contributions of the research work completed in this paper.

- We combine MSCNN and IBiLSTM models so that the MSCNN model is mainly used to extract spatial features with different scale, while the IBiLSTM approach can explore time series features in both directions.
- We propose a hybrid attention mechanism that combines self-attention, channel attention, and spatial attention mechanisms.
- We conduct comprehensive and systematic experiments over an edge-cloud computing platform that increases the model training speed and improve the response time.

The rest of the manuscript is organized in the following manner. We offer a brief review of the existing works in Related work section. The proposed model along with the machine learning model and its implementation over an edge-cloud infrastructure is deliberated in The proposed method section. Detail experimental parameters and the achieve results are explained in Experiments and discussion section. Comparison of the proposed model with other closest rivals is also discussed in this section. Finally, Conclusions and future work section concludes this paper and offers few directions for further research.

Related work

In [21], the authors started from the development method, and divided the digital mode of enterprises into three categories. The first type was the self-develop-ment mode. This mode was for enterprises to conduct research on their own actual needs and used a variety of systems for information operation. The second type of model was combination of management consulting and self-development. Enterprises in need seek consulting from enterprises with advanced experience. On this basis, enterprises completed the design, construction and debugging of digital systems. The third type of mode was the overall introduction mode. Enterprises selected companies with relevant mature systems and independently choose to introduce software products as a whole according to the digital solutions they provide. The classification method proposed by authors in [22] was different from other scholars. Starting from the development model of digital management, it was classified into three types: (i) digital development of enterprise internal management, (ii) portal website, and (iii) e-commerce construction. In [23], the authors summarized four kinds of enterprise digitalization strategies and strategies from the perspective of the combination of production, education and research. The first was the enterprise-industry interaction strategy, the second was the challenge-response model, the third was the flying geese model, and the fourth was the regional interaction model.

In [24], the authors have divided the enterprise digital application modes into three types from a practical point. The first point type was the digitalization of each enterprise was in a stand-alone mode and a stand-alone application. The second was the chain type. Moreover, the digital technology and information technology were applied to various lines within the enterprise, but they were not related to each other and are not integrated. The third type was the mesh type, which realized integration and networking on basis of the chain type. When discussing the status quo of enterprise informatization and digitization, the authors in [25] investigated the current status of informatization and digitization development of construction enterprises. The results found that at this stage, construction enterprises had insufficient awareness of the difference between digitalization and informatization, insufficient investment funds, and lack of reasonable planning. On this basis, it was difficult for enterprises to achieve the purpose of implementing digital management.

In [26], the researchers proposed an economic benefit analysis and calculation formula based on the characteristics of the enterprise and the net value method. This helped enterprises understand the results of informatization and digitization, and provided a basis for enterprises to implement informatization and digitization. Similarly, the authors in [27] proposed that digitalization was the key means for enterprises to gain a competitive advantage, and discussed the digital development scheme of construction enterprises in the research. On the basis of analyzing the existing problems of enterprise informatization and digital management, it proposed a framework of enterprise digitalization. In [28], the authors believed that digital management was a revolutionary change to the management system, and cited a large number of local cases to introduce digital management methods in

different industries and the embodiment of digital management in different value chains. In addition, it also introduced the emerging digital management system and information technology, the future of digital management and so on. Similarly, the authors in [29] believed that digital enterprises would be the standard enterprise survival form in the twenty-first century. A digital enterprise was a new type of enterprise that uses digital technology to change the strategic choices of enterprises and greatly expand the scope of choices. A digital enterprise was a new type of enterprise that realized comprehensive informatization and digitization of internal, external and entire business processes through comprehensive application for digital technology in enterprise.

As a discipline, forecasting originated from the rapid development for science. In increasingly fierce market competition, in order to gain a favorable position, enterprises must collect market information and master the laws of market changes. Engineers and scholars applied time series forecasting methods to various fields, and actively explored new concepts and methods to make them play a greater role in practical applications. The support vector regression algorithm (SVR) could be used to transform a nonlinear regression problem into a linear regression problem in a high-dimensional feature space, according to [30]. Because of this, it was applicable to a wide range of time series forecasting issues. The price of crude oil was predicted by authors in [31]. Over time, the price of crude oil had a significant impact on national budgets and plans. The XGBoost algorithm was used to make crude oil price predictions and to examine its various characteristics and properties. The XGBoost algorithm was also performed admirably in the tests.

In order to improve sales forecasting and supply chain management, the authors in [32] provided a hybrid learning method that combined seasonal mode and support vector regression analysis. With the help of this model, five computer products were forecasted and compared to the ARIMA model, GA-SVR. Experimental findings revealed that the model was more accurate than other models and was superior in terms of accuracy. In [33], the authors used the LSTM model to predict the sales of dishes, and the results showed that the model could fit well for time series data, and the accuracy was higher than other autoregressive models. Moreover, authors in [34] also used the LSTM approach to predict photovoltaic output power, which could capture abstract concepts in photovoltaic power generation sequences. The LSTM model could simulate the change of photovoltaic output power with time. Compared with LSTMs of five different architectures, it was found that the LSTM with time step had a smaller prediction error. Similarly, in [35] the authors studied the demand forecast of fashion clothing,

analyzed historical sales data, and applied time series method to forecast total sales. It also used the Obermeyer method and the concept of ABC product classification on the basis of total sales forecast to forecast the specific quantity demanded by each category.

In [36], the authors introduced the PERT model and a time series model according to the actual characteristics of sales. Based on the respective advantages of the two models, a new comprehensive model was established, which realized an organic combination of sales contingency and inevitability prediction. This effectively simulated the characteristics of product sales, and overcame the subjective judgment of historical data in the previous forecasting process with a fully quantified time series model. Furthermore, in [37] the authors subdivided product demand into deterministic, random, and seasonal types, and proposed the use of grey control theory to establish a forecasting model for deterministic and random clothing products. This avoided the blindness of management based only on experience in the past, and obtained better prediction results. In [38], the authors forecasted the sales of clothing product types that were sensitive to seasonal effects and had a long sales cycle, and used seasonal factors to process the data. It combined the least squares method to conduct statistical analysis of time and sales volume, and established the functional relationship between time and sales volume. This method had better accuracy for predicting the future sales demand of garment enterprises.

The proposed method

This work proposes a HATT-MSCNN-IBiLSTM architecture for predicting corporate sales from a digital management perspective. This method combines MSCNN and IBiLSRM to fuse the advantages of both networks for data processing. Then, hybrid attention mechanism is proposed to further process the features to promote prediction performance.

Multi-scale convolution neural network (CNN)

The CNN model is actually a feedforward network which is based on convolution calculation. Moreover, the CNN has the ability to express features, and can perform translation-invariant classification of input features according to each level of its network. CNN is actually a feedforward neural network with deep structure based on convolution calculation, and CNN has the ability of feature expression, which can perform translation-invariant classification of input features according to each level of its network.

The convolutional layer is the key to understanding the convolutional neural network. A large number of trainable parameters are concentrated in the convolution

kernels, and the network is modified by the loss during the learning process. It uses a back-propagation algorithm along the gradient, and the parameters are continuously moved in the direction that minimizes the loss. By continuously inputting data and labels, the network is trained to obtain appropriate weight parameters, which is the process of allowing the network to acquire knowledge. The design of the convolution operation contains two core ideas, parameter sharing and local sparse computation. Both ideas contain the original intention of reducing memory space and speeding up computing. While achieving suitable accuracy improvement effect with less computation, it also reduces the problems caused by network overfitting. The convolution operation is given by the following Eq. (1):

$$x_j^l = f\left(\sum_{i,j} x_i^{l-1} * k_{ij}^l + b_j^l\right) \tag{1}$$

Where the variable x is used to characterize a feature, k is convolutional kernel, and b is the bias.

After the convolutional layer extracts the features of input, it is input to the next pooling layer. The main function of the pooling layer is feature selection and information filtering, so as to achieve the effect of reducing dimension, reducing the amount of network parameters, reducing the probability of overfitting and improving the speed of network calculation. The working principle of the pooling layer is to use a certain value to replace a certain area in the feature map, and the pooling layer has no weight coefficient. The pooling operation is given by Eq. (2):

$$x_j^l = f\left(down\left(x_j^{l-1}\right)\right) \tag{2}$$

Where x is used to characterize a feature.

Both convolutional layers and pooling layers are computationally processed for local features. The fully connected layer, on the other hand, recombines the previously calculated local features into complete information through the weight matrix. In order to minimize input data loss, the fully connected layer's primary job is to re-fit local features in a nonlinear fashion. There can be one or more completely connected layers overlay prior to the network output layer in order to better tackle nonlinear issues.

The output layer is to directly input result of previous FC layer into softmax function in the output layer to classify the input image. The softmax function is often used in multi-classification tasks. The principle of this function is to calculates the probability that the output of the previous multiple neurons belongs to the class. The Softmax operation is mathematically illustrated as given in Eq. (3):

$$s_i = e^i / \sum_j e^j \tag{3}$$

Where e^i is the ith element of the one-dimensional vector output.

From the perspective of learning feature information in the local receptive field, if the scale and span of the convolution kernel are too small, the data feature resolution will be good, but the low-frequency features existing in the signal cannot be learned well. Conversely, convolution kernels with larger scales use larger strides and can learn information over longer time horizons. That is, the low-frequency characteristics in the signal cannot reflect the high-frequency characteristics. To combine the advantages of different scales, this work proposes a multi-scale feature learning strategy. The one-dimensional multi-scale convolutional layer is demonstrated in Fig. 1.

The one-dimensional multi-scale convolution layer (1D-MSCL), in fact, uses one-dimensional convolution kernels of different scales to perform multi-scale feature extraction on the original features to improve the
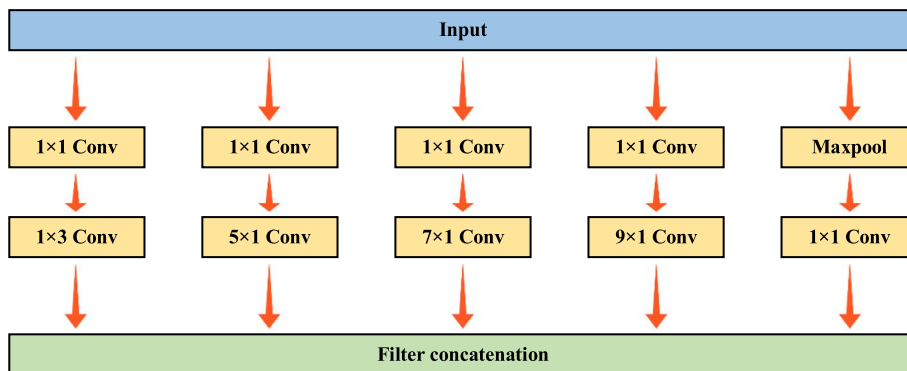

Fig. 1 The one-dimensional multi-scale convolutional layer

robustness and discrimination of features. Based on this multi-scale convolutional layer, multi-scale convolutional network designed in this work is demonstrated in Fig. 2.

Improved bi-directional long short-term memory

The CNNs have no memory and process each input individually without saving any state between the input and the input. This also means that they cannot be connected before and after input and output. For such a network, to process time series of data, the entire sequence needs to be presented to network. But the biggest problem with this is the huge amount of parameters and too much calculation. When mapping between input and output sequences, recurrent neural networks have the advantage of using contextual information retained by hidden states. Standard RNN architectures can only access a small number of possible contexts. There's problem since the effect of a particular input on the hidden layer, and hence on the network output, declines exponentially over time. So that the information of the previous sequence obtained by the later time step is very limited, which will cause the model to fail to train.

The LSTM model is one of the variants of standard RNN that can pass information from the past across multiple time steps. It runs parallel to the multiprocessing time series and is designed to filter information with a select-through mechanism. The LSTM introduces gated units and cell states, as illustrated in Fig. 3, one of the time step units of LSTM model.

Three gates in the LSTM approach are: (i) forget gate, (ii) input gate, and (iii) output gate. The forget gate in LSTM is to decide which information needs to be deleted in the current state as Eq. (4). The input gate is to decide what information to add to the memory state as Eq. (5). The internal memory unit will choose to add a part of the candidate memory state information to obtain new memory state information as Eqs. (6) and (7). The output gate is based on input and memory cell to determine what to output as Eqs. (8) and (9).

$$f_t = \sigma\left(W_f x_t + U_f x_{t-1} + b_f\right) \tag{4}$$

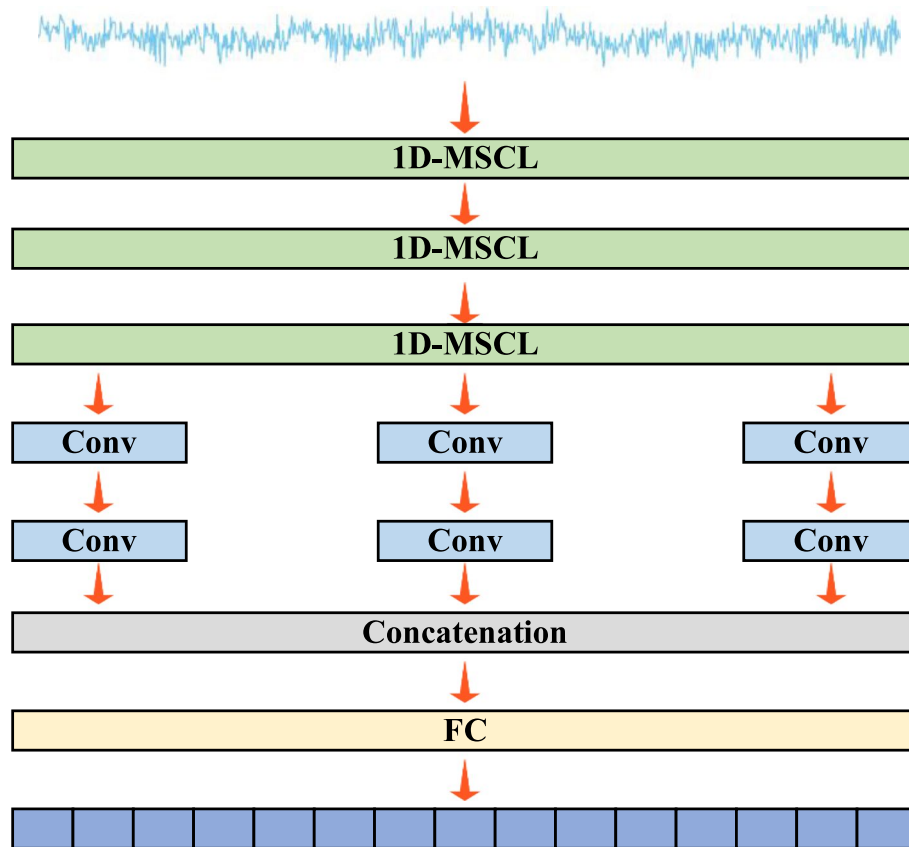$$i_t = \sigma\left(W_i x_t + U_i x_{t-1} + b_i\right) \tag{5}$$

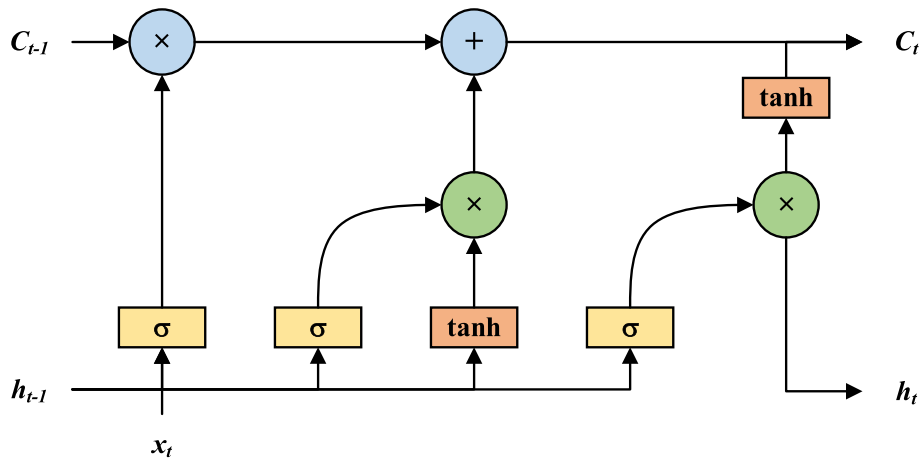**Fig. 2** The MSCNN pipeline approach

Fig. 3 Illustrations for the LSTM model

$$\tilde{c}_t = \tanh(W_c x_t + U_c x_{t-1} + b_c) \tag{6}$$

$$c_t = f_t * c_{t-1} + i_{t-1} * \tilde{c}_t \tag{7}$$

$$o_t = \sigma(W_o x_t + U_o x_{t-1} + b_o) \tag{8}$$

$$h_i = o_t * \tanh(c_t) \tag{9}$$

Where W_f, W_i, W_c, W_o are the weight matrices of forget gate, input gate, update unit state and output gate. b_f, b_i, b_c, b_o are the bias of the forget gate, input gate, update unit state and output gate.

Although, the LSTM model can perform forward data feature extraction on time series data. However, time-series data is not only related to previous data, but also related to subsequent data to some extent. As a deformation structure, BiLSTM includes the forward propagation layer as well as the backward propagation layer of LSTM. Figure 4 is a schematic diagram of the BiLSTM approach.

The problems of manual parameter adjustment and slow convergence of BiLSTM network are improved in this work. In this work, PSO is used to optimize its hyper-parameters, and the hyper-parameters to be optimized are mapped to particles. Each particle shares the individual extremum and compares the global extremum, and continuously updates the position and velocity for iterative optimization. The inertia weight has a positive correlation with the particle's global search capacity and a negative correlation with the particle's local search ability When using a typical PSO, the inertia weight has a predetermined value, which reduces the capacity of the particles to find the global or local optimum. Consequently, this study optimizes the inertia weight factor and employs a nonlinear inertia weight to enhance performance of the particle swarm optimization and build the IBiLSTM model.

$$w = (w_{max} - w_{min})^{t/t_{max}} \tag{10}$$

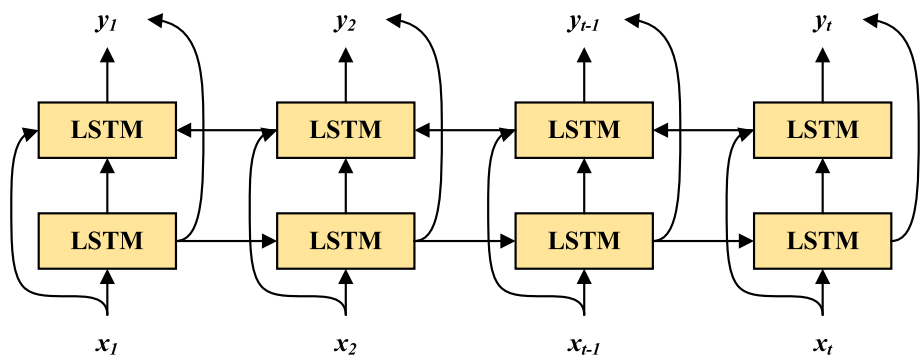Where w_{max} is the maximum inertia weight, and w_{min} is the minimum inertia weight.

Fig. 4 Illustration of the BiLSTM approach

The hybrid attention mechanism

The attention mechanism largely borrows from the human visual attention mechanism in principle. The core of the attention mechanism is weight distribution. In terms of mathematics and programming, the attention mechanism can be understood as the weighted summation of sequence information. Formally, the attention mechanism can be understood as a key-value query. In physical sense, the attention mechanism can be understood as similarity calculation. For neural network models, the importance of information is reflected by weights.

The self-attention mechanism is to obtain Query, Key and Value from the input sequence through three sets of linear changes. Then find the similarity of each Query to each Key as a score. It gets the weight of the input sequence after going through softmax. Finally, the attention weights are injected into Value, and weighted summation is performed to obtain the output sequence. The advantage for the self-attention mechanism is it can be processed in parallel and can directly see the information of the entire input sequence. The calculation process is:

$$Q = X * W^Q \tag{11}$$

$$K = X * W^K \tag{12}$$

$$V = X * W^V \tag{13}$$

$$Y = softmax\left(Q * K^T / \sqrt{d_K}\right)V \tag{14}$$

Where X is input, and W^Q, W^K, W^V are used to characterize the parameter matrix.

A variant is multi-head self-attention mechanism, and advantage of multi-head self-attention is that different heads can focus on different information. The specific process is to obtain multiple sets of Query, Key and Value through multiple linear transformations of the input sequence. The intermediate process is the same as the self-attention mechanism. After obtaining multiple sets of output sequences, the output sequences at the same time point are spliced by channel. If the connected dimension does not match the input dimension of the next layer, another linear layer can be passed to transform the dimension to the input dimension required by the next layer. The middle groups indicate how many heads there are. The multi-head self-attention is demonstrated in Fig. 5.

The role of the channel attention is to focus on channel dimension of feature map to obtain a set of attention weight distributions about the channel. The feature map on each channel dimension is equivalent to obtaining a feature on the original feature map, and the channel attention can help network to extract meaningful features. The upper layer feature map is utilized as input feature map, and adaptive average pooling and adaptive max pooling are performed based on the spatial dimension to compress the spatial information and obtain two feature channels. Then, through two 1×1 convolutions of shared parameters, dimensionality reduction and dimensionality increase are performed. The purpose of dimensionality reduction is reducing parameters. Then, the outputs of the two branches are added element-wise and Sigmoid
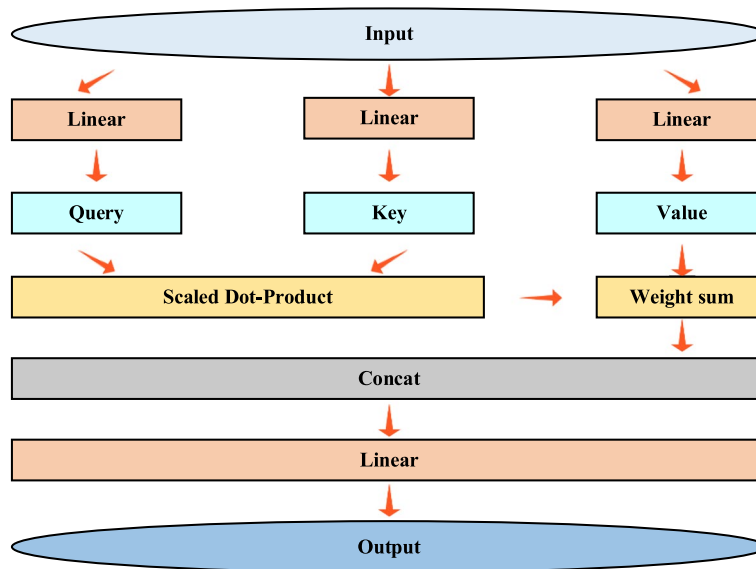

Fig. 5 The multi-head self-attention module

is performed to generate channel attention feature weight score. Finally, the channel attention weight score is injected into the feature map of the original input, that is, the feature map with channel attention is obtained. The channel attention module is demonstrated in Fig. 6.

The role of the spatial attention is to focus on the spatial dimension of feature map to obtain a set of spatial attention weight distributions. After the data is filtered through the convolutional layer, each pixel in the extracted feature map represents a certain feature of certain region in the upper-layer feature map. The spatial attention mechanism indicates that the neural network should pay attention to the features of a certain region of the data. The neural network is able to maintain the essential information when the spatial attention mechanism transforms the original input into another space. Channel attention module's feature map is used as the input feature map for spatial attention module's feature map. Global average pooling and global maximum pooling are used to compress channel information based on channel dimension. Splicing together the two feature maps is done depending on the channel dimension and then a convolution is used to reduce the channel dimension. It then generates its weight score by using the Sigmoid function. Finally, the feature map with spatial attention is obtained by injecting the weight score for spatial attention into the original input feature map. The spatial attention module is demonstrated in Fig. 7.

The hybrid attention mechanism proposed in this work is to fuse multi-head self-attention, channel attention and spatial attention to enrich the diversity of features. The hybrid attention mechanism is demonstrated in Fig. 8.

HATT-CNN-BiLSTM for enterprise sales forecast

Although, the CNN model has shown great advantages in extracting the characteristics of data, however it only
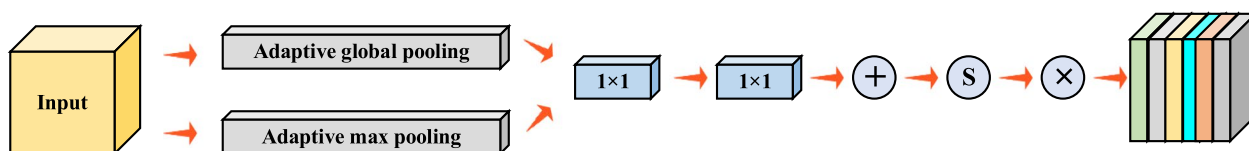

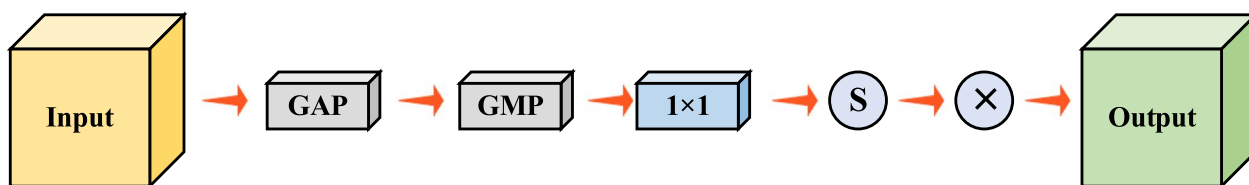**Fig. 6** The channel attention module
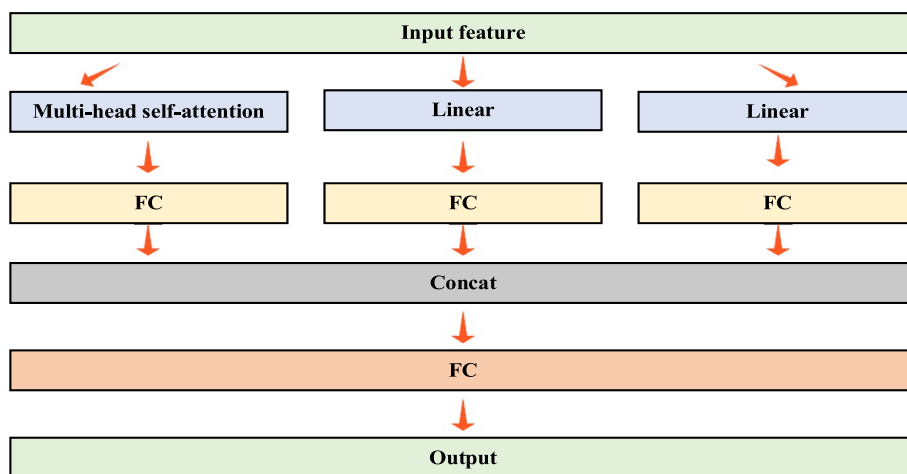

Fig. 7 The spatial attention module

Fig. 8 The hybrid attention pipeline approach

considers the correlation of adjacent vectors of data, and does not consider the time series between continuous data. Although, the BiLSTM approach can handle time-series-related features, it is not as comprehensive as CNN for feature extraction. This work uses CNN to extract high-dimensional spatial features, reduces the feature dimension of data, and realizes feature extraction of early weak features with insignificant features. The BiLSTM model is selected to process the feature sequence output by the CNN part, and mine the time series information in the data. This work integrates CNN and BiLSTM to obtain a CNN-BiLSTM hybrid model, uses CNN to extract unique features, and uses BiLSTM to solve the problem of timing characteristics. Finally, this work introduces an attention to further process features to improve robustness and discrimination of features. The proposed HATT-MSCNN-IBiLSTM deep learning model for enterprise sales forecasting is demonstrated in the following Fig. 9.

HATT-MSCNN-IBiLSTM model first extracts the corporate sales features of historical information through CNN. A bidirectional LSTM network is then used to train and predict on the corporate sales dataset. Attention mechanism is introduced in process of network training, and different weights are assigned to each sales feature, so that the network can learn feature information more effectively.

The edge computing model

According to Fig. 10, the suggested model is composed of three distinct layers: the cloud layer, the edge layer, and the local layer. Through a variety of IoT devices and sensors, the local layer is in charge of gathering crucial data (connected to businesses). After data has been gathered, it is processed (and/or stored) at the edge layer using small-scale datacenters, or edge clouds. Aggregation that might be accomplished by deleting duplicate data could be a part of the edge level processing. The filtered data can then be sent to the distant cloud layer for further processing, including resource management, storage, and monitoring. The cloud network architecture may experience severe delays if the
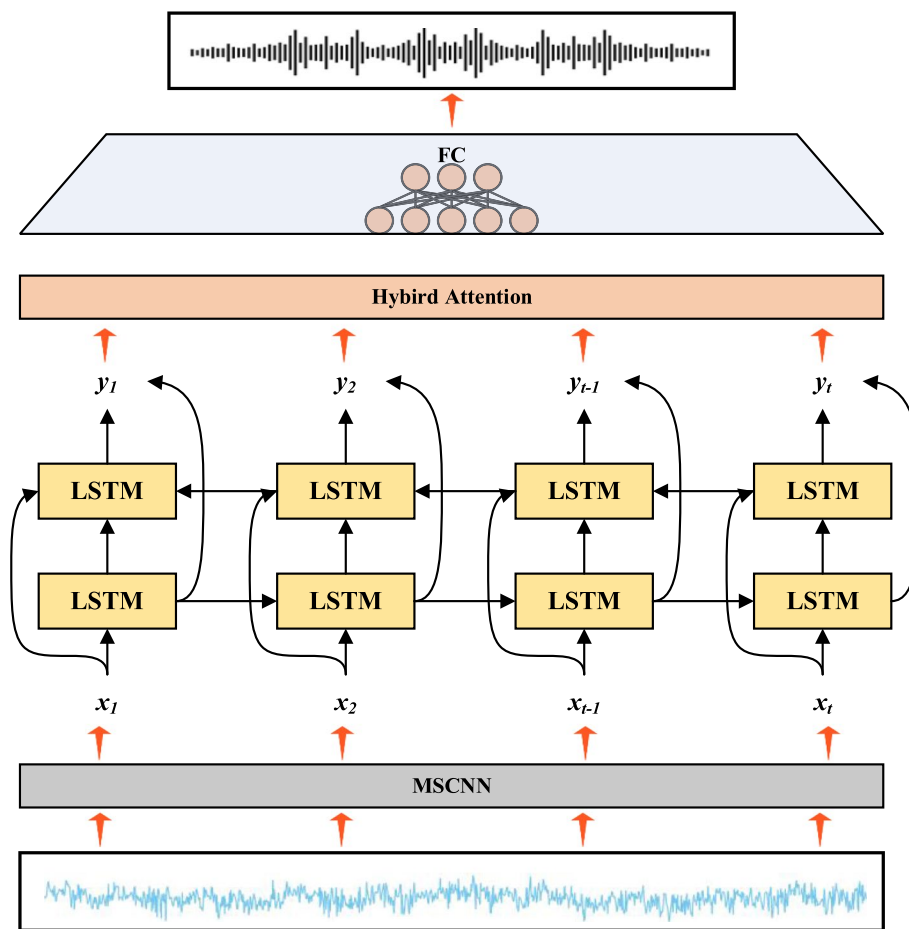

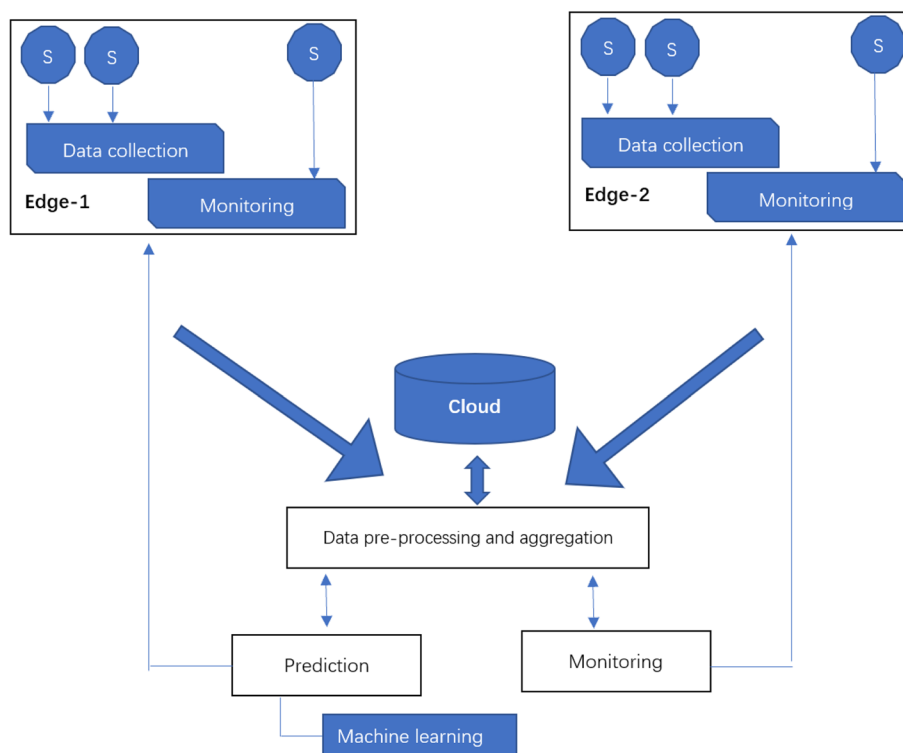**Fig. 9** The HATT-CNN-BiLSTM model and its architecture

Fig. 10 Framework for real-time prediction and monitoring of enterprises

data obtained at the local layer is sent straight to the cloud layer, albeit this is not always the case.

Additionally, real-time prediction may take place at the edge layer, but cloud-based prediction for monitoring services is also an option. Additionally, real-time predictions made on a remote cloud, such quickest or safest route calculation, may need considerable delays due to the network's capacity and quality. In such case, provided the necessary data is kept locally, the closest edge cloud can forecast traffic, distance, and road conditions. The data could not, however, be accessible or processed locally due to the edge cloud's limited capacity for storage and processing. There are three alternatives in such situation: (i) transport the necessary data from the cloud to the edge, analyze it, and then discard it; (ii) complete the prediction at the distant cloud; and (iii) train the prediction model at the remote cloud and forecast at edge layer (distributed fashion computation). Regarding (i) depending on the model being utilized, training may be costly. Regarding (ii), greater and longer reaction times would result from looking for the right data and then transferring it across the network interface. The model can be regularly trained using the suggested technique (iii), and only edge prediction would take place.

Experiments and discussion
Dataset and experimental details

This paper collects enterprise sales data from the perspective of digital management to construct the dataset required for the experiment. This paper uses enterprise sales data from three different industries, ESDA, ESDB, and ESDC. These datasets contain different sample compositions, and the specific information is demonstrated in Table 1. This work is based on the deep learning platform for experiments, and the specific experimental environment is demonstrated in Table 2. We have created one cloud server, one fog server, and multiple user nodes at the edge layer which are inter-connected through a network. The characteristics of the network and CPU power are illustrated in Table 3. We assume similar machines both in the cloud and edge. The hyper-parameters are set as follows: batch size is set to 128, time step is set to 20, training epoch is set to 100, learning rate is set to 0.001,

Table 1 Dataset information

Dataset	Training set	Test set	Total
ESDA	20,397	12,831	33,228
ESDB	49,812	27,469	77,281
ESDC	60,491	38,501	98,992

Table 2 Experimental environment

Item	Name
GPU	GTX 2060Ti
CPU	Intel Core i7
Deep learning framework	PyTorch1.7
Operation system	Ubuntu 18.04

Table 3 Cloud-edge infrastructure setup

Parameter	Cloud	Fog
Level	0	1
CPU (MIPS)	44,800	2800
RAM (GB)	40	4
Uplink BW (MB)	100	10,000
Downlink BW (MB)	10,000	10,000
CostPerMIPS	0.01	0.0

dropout is set to 0.3, optimizer is set to Adam. The evaluation indicators are RMSE as well as MAE, which are calculated as given by the following Eqs. (15) and (16), respectively:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \qquad (15)$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \qquad (16)$$

Analysis on training loss

The HATT-CNN-BiLSTM model built in this work is a deep learning network, so network training is an extremely important step. In order to evaluate the network training process of HATT-CNN-BiLSTM, this work analyzes training loss, data results are demonstrated in Fig. 11.

As the training goes deeper, the loss of HATT-CNN-BiLSTM on both datasets gradually decreases. When training reaches 60 epochs, the network loss is basically stable. The loss converges around 0.4 on the ESDA dataset and around 0.2 on the ESDB dataset and around 0.1 on the ESDC dataset. This may be due to the fact that the ESDC dataset is larger and the samples are more abundant, the network training is more thorough.

Method comparison

To verify the superiority of HATT-CNN-BiLSTM designed in this paper for enterprise sales forecasting,
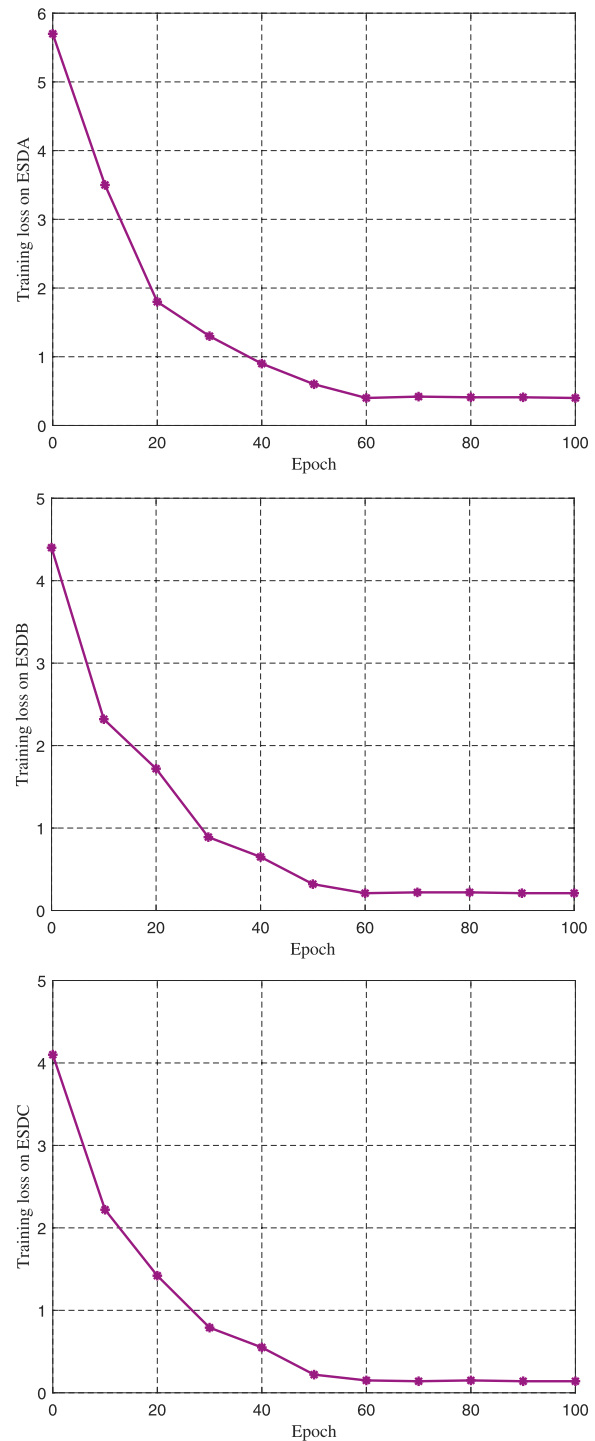

Fig. 11 Training loss of HATT-CNN-BiLSTM on ESDA and ESDB

this work compares it with other methods. The compared methods include CNN, LSTM, GRU, and BiLSTM. The compared indicators include the correct rate in training and RMSE and MAE for test. The comparison results are demonstrated in Fig. 12 and Table 4.
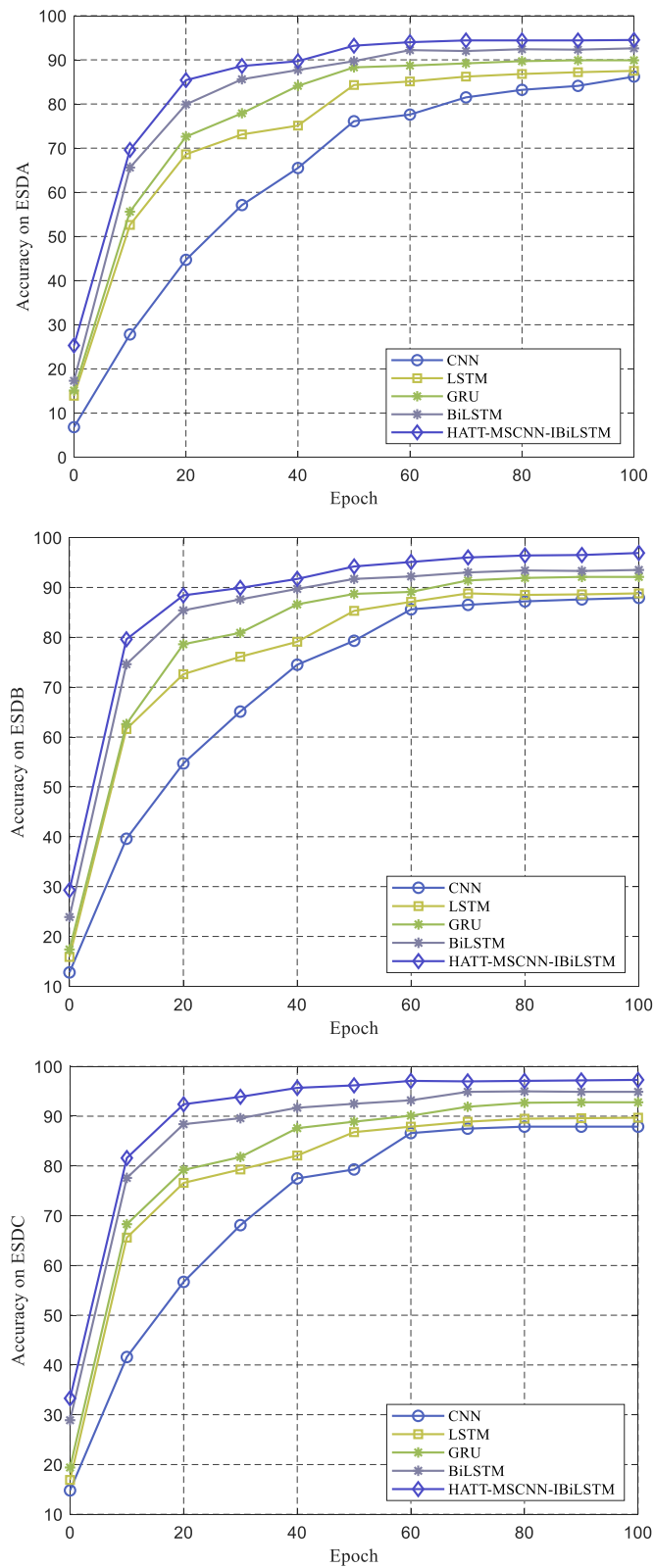
Fig. 12 Training accuracy for different methods

As can be concluded from Fig. 11, the accuracy of HATT-MSCNN-IBiLSTM is higher than other methods at any training epoch. In addition, its convergence speed is also the fastest. Compared with other methods, HATT-MSCNN-IBiLSTM achieves 8.3%, 7.0%, 4.6% and 1.9% accuracy improvement on ESDA dataset respectively after training. On the ESDB dataset, the accuracy improvements are 9.0%, 8.1%, 4.8%, and 3.4%. From Table 4, it can be concluded that the RMSE and MAE corresponding to the HATT-MSCNN-IBiLSTM method are the lowest. Compared to other methods, the RMSE of HATT-MSCNN-IBiLSTM on the ESDA dataset drops by 16.2, 11.9, 9.1 and 6.9. For MAE, the declines are 24.4, 17.8, 11.7 and 8.2. RMSE of HATT-MSCNN-IBiLSTM on the ESDB dataset drops by 18.3, 14.5, 11.6 and 9.8. For MAE, the declines are 21.3, 16.3, 12.6 and 9.5. These data support the feasibility and superiority of the HATT-MSCNN-IBiLSTM method designed in this work for enterprise sales forecasting from the perspective of digital management.

Analysis on MSCNN

The HATT-MSCNN-IBiLSTM model utilizes a multi-scale convolutional network to improve representation ability for features. To verify the feasibility of multi-scale strategy, this work compares the network performance with and without multi-scale convolution, as demonstrated in Fig. 13.

Compared with the model using single-scale features, the network performance will be improved accordingly after processing the features with multi-scale convolution. On the ESDA dataset, the RMSE and MAE of HATT-MSCNN-IBiLSTM drop by 3.7 and 3.8.

Analysis on IBiLSTM

The HATT-MSCNN-IBiLSTM model designed in this paper utilizes a nonlinear inertia weight to construct IBiLSTM. To verify the feasibility of this dynamic inertia weight strategy, this work compares the network performance with and without nonlinear inertia weight, as demonstrated in Fig. 14.

Compared with the model using traditional BiLSTM, the network performance will be improved accordingly after processing the features with improved BiLSTM model. On the ESDA dataset, the RMSE and MAE of HATT-MSCNN-IBiLSTM drop by 2.5 and 2.7. On the ESDB dataset, the RMSE and MAE of HATT-MSCNN-IBiLSTM drop by 3.3 and 3.4. On the ESDC dataset, the RMSE and MAE of HATT-MSCNN-IBiLSTM drop by 5.2 and 4.3. These comparative data support the superiority of the improved BiLSTM approach.

Analysis on hybrid attention

The HATT-MSCNN-IBiLSTM model designed in this paper utilizes a hybrid attention mechanism to enhance feature. To verify the feasibility of this hybrid attention strategy, this work compares the network performance with and without hybrid attention, as demonstrated in Fig. 15.

Compared with the model using traditional attention, the network performance will be improved accordingly after processing the features with hybrid attention. On the ESDA dataset, the RMSE and MAE of the HATT-MSCNN-IBiLSTM drop by 2.1 and 2.3. On the ESDB dataset, the RMSE and MAE of the HATT-MSCNN-IBiLSTM model drop by 2.6 and 2.5. On the ESDC dataset, the RMSE and MAE of the HATT-MSCNN-IBiLSTM model drop by 5.5 and 2.4. These comparative data support the superiority of the proposed hybrid attention mechanism.

Analysis on 1D-MSCL layer

In HATT-MSCNN-IBiLSTM network, one-dimensional multi-scale convolutional layers can be repeatedly embedded. To evaluate the impact of different

Table 4 Test RMSE and MAE for different methods

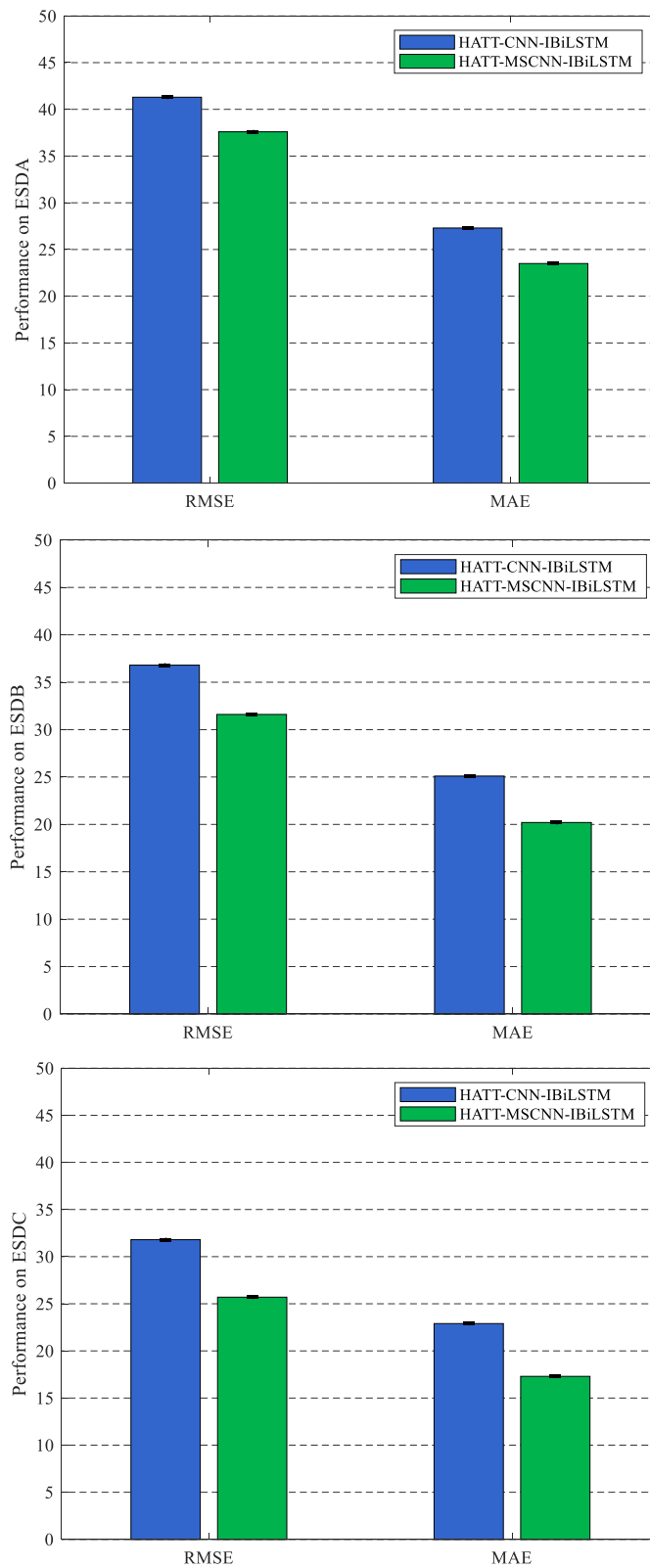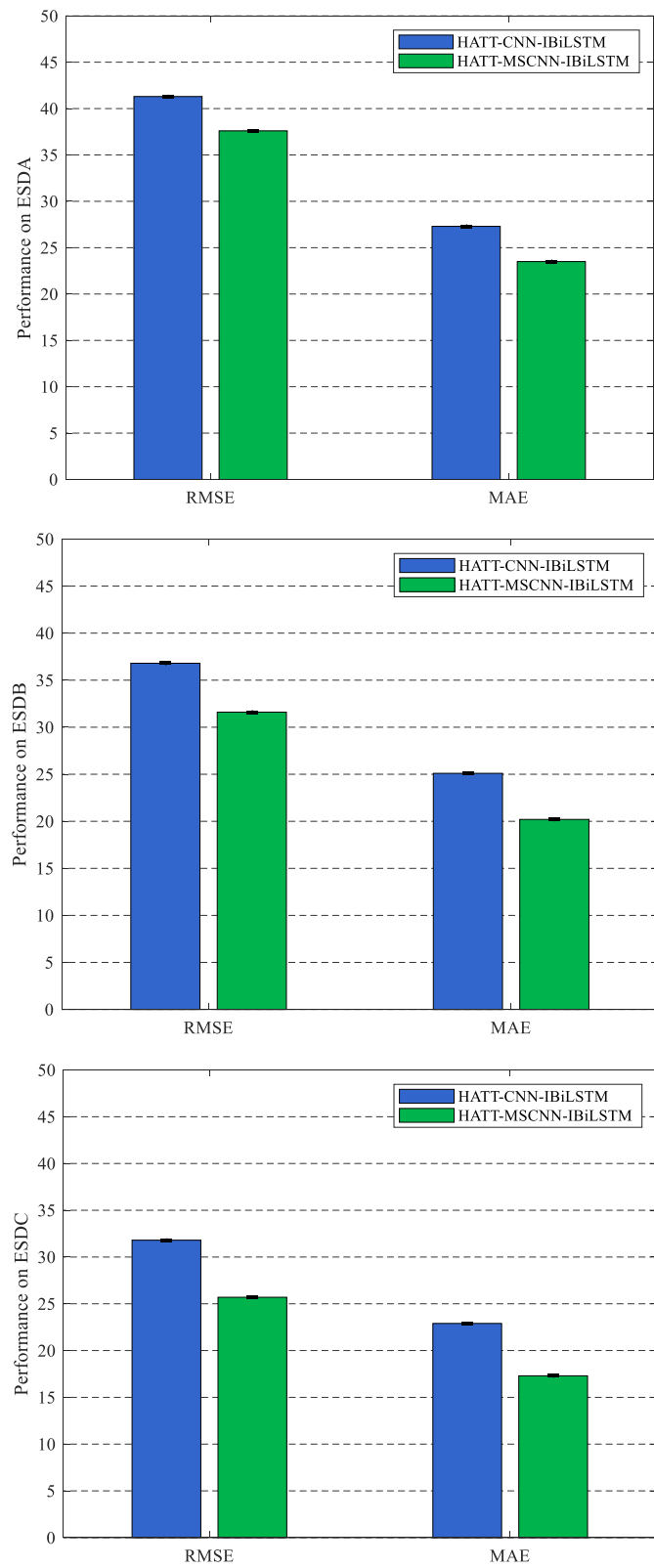| Method	ESDA		ESDB		ESDC	
	RMSE	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**
CNN	53.8 ± 3.6	47.9 ± 3.1	49.9 ± 3.8	41.5 ± 3.3	47.2 ± 2.9	39.6 ± 2.6
LSTM	49.5 ± 2.7	41.3 ± 2.5	46.1 ± 1.9	36.5 ± 2.7	43.5 ± 3.3	34.1 ± 3.0
GRU	46.7 ± 4.1	35.2 ± 2.7	43.2 ± 2.8	32.8 ± 5.6	39.2 ± 3.8	30.5 ± 2.8
DEEPAR	51.2 ± 2.6	44.6 ± 2.3	47.9 ± 2.2	38.7 ± 4.3	45.5 ± 3.2	37.2 ± 3.2
Nbeats	48.1 ± 2.9	37.9 ± 3.3	44.6 ± 2.5	33.9 ± 3.7	42.1 ± 2.8	32.3 ± 3.5
edRVFL	45.1 ± 3.6	32.6 ± 3.7	43.6 ± 3.5	32.1 ± 2.9	38.9 ± 4.4	28.2 ± 4.1
DESN	47.3 ± 1.9	37.1 ± 1.7	44.9 ± 5.1	35.5 ± 3.4	42.1 ± 4.1	32.6 ± 2.7
BiLSTM	44.5 ± 1.5	31.7 ± 1.8	41.4 ± 2.7	29.7 ± 2.4	36.8 ± 2.9	25.4 ± 2.2
HATT-MSCNN-IBiLSTM	37.6 ± 1.1	23.5 ± 1.3	31.6 ± 2.0	20.2 ± 1.6	25.7 ± 1.9	17.3 ± 0.8

Fig. 13 Analysis on MSCNN

Fig. 14 Analysis on IBiLSTM

Fig. 15 Analysis on hybrid attention

Table 5 Analysis on 1D-MSCL layer

1D-MSCL layer	ESDA		ESDB		ESDC	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
1	41.2	27.9	38.1	25.6	35.2	23.7
2	39.9	25.1	33.5	23.7	30.6	20.5
3	37.6	23.5	31.6	20.2	25.7	17.3
4	40.1	27.6	34.2	29.7	31.5	21.5
5	43.3	29.1	38.6	26.1	37.9	26.1

Table 6 Analysis on IBiLSTM layer

IBiLSTM layer	ESDA		ESDB		ESDC	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
1	37.6	23.5	31.6	20.2	25.7	17.3
2	39.2	27.1	33.5	25.9	32.1	23.3
3	42.9	30.5	37.1	27.4	35.8	25.7

Table 7 Analysis on IBiLSTM node

IBiLSTM node	ESDA		ESDB		ESDC	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
10	39.9	28.7	37.2	26.1	31.9	24.2
20	37.6	23.5	31.6	20.2	25.7	17.3
30	40.5	29.8	37.9	27.6	33.2	26.1
40	43.5	31.7	38.9	28.8	35.6	28.3

multi-scale convolutional layers on network performance, this work conducts comparative experiments for different 1D-MSCL layers, as demonstrated in Table 5.

As 1D-MSCL layers increases, RMSE and MAE of HATT-MSCNN-IBiLSTM show a trend of decreasing first and then increasing. When the parameter is set to 3, the lowest RMSE and MAE can be obtained. Therefore, in this work, the value of this parameter is 3.

Analysis on IBiLSTM layer and node

In HATT-MSCNN-IBiLSTM network, IBiLSTM is embedded. To evaluate the impact of different IBiLSTM layers on network performance, this work conducts comparative experiments for different IBiLSTM layers, as demonstrated in Table 6.

As iBiLSTM layers increases, the RMSE and MAE metrics values of the HATT-MSCNN-IBiLSTM show a gradually increasing trend. When this parameter is set to 1, then we overserved through numerical experiments that the lowest values for RMSE and MAE can be obtained. Therefore, in this work, the value of this parameter is set and always kept to 1.

Furthermore, in IBiLSTM model, the number of neuron nodes is also a variable. In this work, a comparative experiment is carried out to find the optimal number of nodes. The experimental data is illustrated in Table 7.

As the number of IBiLSTM nodes increases, the RMSE and MAE of HATT-MSCNN-IBiLSTM show a trend of decreasing first and then increasing. When the parameter is set to 20, the lowest RMSE and MAE can be obtained. Therefore, in this work, the value of this parameter is 20.

Analysis of the cloud-edge model

We modelled an edge cloud infrastructure made up of a distant cloud node, three edge devices, and the popular iFogSim simulator. With dedicated networks, the bandwidth between edge devices and the distant cloud

is 10 GB/s. Additionally, we assume that sensors and vehicles are linked to the edge cloud through a 1 GB/s link. We construct a small mobile edge cloud MEC) infrastructure based on these characteristics to test our suggested framework, and the results in terms of training and prediction timeframes are displayed in Fig. 16. According to Fig. 16, utilizing the edge/fog platform for training and prediction of the proposed mechanism performs better than using the cloud alone in both scenarios (edge + cloud) and edge only. It should be noted that lengthier response times for edge-only applications are caused by the requirement to transport data from the cloud to the edge. As a result, fluctuations in network traffic circumstances and data volume may occur.

As the number of edge nodes rises, so does the network traffic at the edge layer, and as a result, the traffic at the fog server rises proportionately as well. For reliable network commu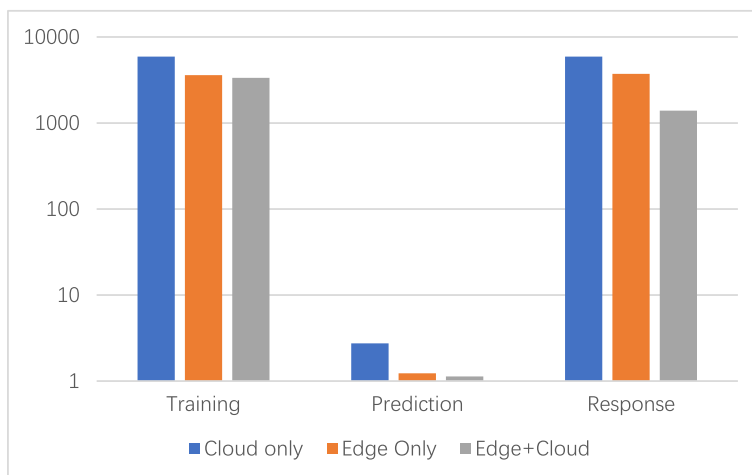nication, the packet loss rate must be as low as practical. We observed that the rate of packet loss increases as the number of users on the edge increases because more messages are exchanged as the user base expands. The packet loss rate increases as a result of the increased congestion between the cloud and edges.

Conclusions and future work

The differences in enterprise resources are narrowing day by day, and the impact of factors of production on enterprise benefits is decreasing. Enterprises are looking for new ways to gain competitive advantages. With the advent of the mobile Internet era, digital transformation is on the agenda, and digital management is related to an enterprise's business strategy and business model. It can enhance the competitiveness and efficiency of the enterprise, optimize the existing business of the enterprise, and help the enterprise to create new business opportunities. Enterprise management has

(a) Dataset 1

(b) Dataset 2

Fig. 16 Time complexity in terms of training, prediction and response time (measured in seconds) using two datasets

shifted from a traditional method to a digital one in the age of the network economy and modern information technology. Correct theoretical direction is required for the development of an enterprise information system, and modern information technology is required to actualize management innovation. From the perspective of digital management, forecasting enterprise sales has become an important topic. Sales forecasting and decision analysis are crucial for enterprises, and reasonable sales forecasts can optimize supply chain management for enterprises.

This work proposes a combination of enterprise sales forecasting from the perspective of digital management and neural networks, and proposes a network HATT-CNN-BiLSTM model for enterprise sales forecasting. First, this work combines MSCNN with IBiLSTM. MSCNN is utilized to extract spatial features with different scale, it is often impossible to effectively explore the rules of time series features, and the processing of time series data is the strength of the LSTM network. The IBiLSTM model can explore time series features in both directions, and more useful information can be obtained. The MSCNN-IBiLSTM, which is composed of MSCNN and IBiLSTM, it can take advantage of strengths and avoid weaknesses, and give full play to the roles of the two models in different fields. Second, this work proposes a hybrid attention mechanism that combines self-attention, channel attention, and spatial attention. It improves features extracted by MSCNN-IBiLSTM through a hybrid attention to build HATT- MSCNN-IBiLSTM network, which can extract more discriminative features. In the future, we will extend this work through implementing more robust activation functions and observed how the accuracy of the proposed model can be increased. Furthermore, we will work toward designing a real cloud-edge framework and implement the proposed model so that the obtained outcomes can be validated.

References

1. Jeon S, Son I, Han J (2020) Exploring the role of intrinsic motivation in ISSP compliance: enterprise digital rights management system case. Inf Technol People 34(2):599–616
2. Zhou D (2022). Application of Data Mining Technology in Enterprise Digital Human Resource Management. Security and Communication Networks 2022:9. Article ID 7611623. https://doi.org/10.1155/2022/7611623.
3. Ge J, Wang F, Sun H et al (2020) Research on the maturity of big data management capability of intelligent manufacturing enterprise. Syst Res Behav Sci 37(4):646–662
4. Pei J, Zhong K, Li J et al (2022) PAC: partial area clustering for re-adjusting the layout of traffic stations in city's public transport. IEEE Trans Intell Transp Syst. https://doi.org/10.1109/TITS.2022.3179024
5. Fang X, Zhang F, Ghoreishi M (2022). Evaluation Index System of Enterprise Digital Transformation Based on Balanced Scorecard. Forest Chemicals Review 11;324-342. http://forestchemicalsreview.com/index.php/JFCR/article/view/1128.
6. Miethlich B, Belotserkovich D, Abasova S et al (2021) The impact of COVID-19 on digital enterprise management. IEEE Eng Manag Rev 49(4):16–29
7. Yiping L (2021) Innovation strategy driven by enterprise digital transformation. Acad J Bus Manag 3(9):50–52
8. Li C (2020) E-commerce enterprise marketing management: realistic dilemma, main incentives and resolution strategies. Open J Soc Sci 8(07):245
9. Ruifeng JIN, Shijie Z (2021) Research on the trends and features of enterprise digital transformation: based on the WOS database. Can Soc Sci 17(4):55–60
10. Catuogno L, Galdi C, Riccio D (2020) An enterprise rights management system for on-the-field maintenance facilities. IEEE Access 8:95987–95996
11. Romprasert S, Trivedi A (2021) Sustainable economy on community enterprise and digital marketing. ABAC J 41(1):62–80
12. Pashkevych M, Bondarenko L, Makurin A et al (2020) Blockchain technology as an organization of accounting and management in a modern enterprise. Int J Manag 11
13. Xiumei Z, Xiaoyue L, Tiandong W (2022) The impact of enterprise digital transformation strategy and capability on product service system. Foreign Econ Manag 44(04):137–152
14. Pei J, Yu Z, Li J et al (2022) TKAGFL: a federated communication framework under data heterogeneity. IEEE Trans Netw Sci Eng. https://doi.org/10.1109/TNSE.2022.3177748
15. Banerjee RP (2020) Art and science of management in the digital era: Indian spiritual wisdom for managing sustainable global enterprise. Sage Publications Pvt. Limited
16. Lim B, Zohren S (2021) Time-series forecasting with deep learning: a survey. Phil Trans R Soc A 379(2194):20200209
17. Sezer OB, Gudelek MU, Ozbayoglu AM (2020) Financial time series forecasting with deep learning: a systematic literature review: 2005–2019. Appl Soft Comput 90:106181
18. Lara-Benítez P, Carranza-García M, Riquelme JC (2021) An experimental review on deep learning architectures for time series forecasting. Int J Neural Syst 31(03):2130001
19. Zhang G, Qiu H (2021) Competitive product identification and sales forecast based on consumer reviews. Math Probl Eng 2021
20. Tallón-Ballesteros AJ (2020) The design of ERP intelligent sales management system. Fuzzy Syst Data Min VI Proc FSDM 2020 331(2020):413

21. Liu Z, Duan M, Wu C (2021) Exploration and practices on enterprise network security system of Baotou steel. Sci Technol Baotou Steel Corp 47(4):86–88
22. Zhao T (2018) Application and extension of ERP system finance in engineering consulting industry. Account Learn 18:41–42
23. Xie K (2000) The information demand of and informationized investment patterns of China's businesses. Manage World 3:96–103
24. Ibrahim MH (2015) Secure and robust enterprise digital rights management protocol with efficient storage. Int J Inf 18(2):625–640
25. Zhang J, Xiang D, Guo J (2005) A survey and requirement analysis on application of information technologies in Chinese construction enterprise. Constr Technol 34(2):3–7
26. Wang Y, Zhang L, Liu J (2005) Economy benefits analysis for construction informatization. J Chongqing Univ 28(1):117–120
27. Zhu L, Ren H (2006) Informationization: the key to improvement of construction enterprises' core competence. J Chongqing Univ 12(6):42–47
28. Katuu S (2014) Enterprise content management and digital curation applications. In: Maturity model connections, p 1
29. Laudon KC, Laudon JP (2004) Management information systems: managing the digital firm. Pearson Educación
30. Kazem A, Sharifi E, Hussain FK et al (2013) Support vector regression with chaos-based firefly algorithm for stock market price forecasting. Appl Soft Comput 13(2):947–958
31. Gumus M, Kiran MS (2017) Crude oil price forecasting using XGBoost. In: International conference on computer science and engineering. IEEE, pp 1100–1103
32. Lu CJ (2014) Sales forecasting of computer products based on variable selection scheme and support vector regression. Neurocomputing 128:491–499
33. Liu X, Ichise R (2017) Food sales prediction with meteorological data—a case study of a Japanese chain supermarket. In: International conference on data mining and big data. Springer, Cham, pp 93–104
34. Abdel-Nasser M, Mahmoud K (2019) Accurate photovoltaic power forecasting models using deep LSTM-RNN. Neural Comput Appl 31(7):2727–2740
35. Wang C (2006) A method of forecasting the short-term requirement of fashion products in apparel industry. Shanghai Manag Sci 28(6):36–38
36. Fang Y, Chen S, Dai S (2006) Study on forecast method for garment sales based on time series and PERT. Res Technol 11:60–63
37. Yu L (2006) Gray forecast model on stochastic demand of clothing products. Melliand China 34(8):78–80
38. Xue M, Shen J, Yang Y (2004) New exploration on quantitatively forecasting method of clothing sales volume. J Donghua Univ 30(6):75–77

Publisher's Note