Systems and Soft Computing 6 (2024) 200107

Contents lists available at ScienceDirect

Systems and Soft Computing

journal homepage: www.journals.elsevier.com/soft-computing-letters

Big data processing and analysis platform based on deep neural network model

Sheng Huang

Equipment Department, Shanghai Institute of Tourism, Shanghai 201418, China

ARTICLE INFO

Keywords:
Big data processing
Analytics platform
Deep neural network
Stock prediction

ABSTRACT

Users are increasingly turning to big data processing systems to extract valuable information from massive datasets as the field of big data grows. Data analytics platforms are used by e-commerce enterprises to improve product suggestions and model business processes. In order to meet the needs of large-scale data center operation and maintenance management, Internet companies often use Flink to process log data. This paper takes the big data processing and analysis platforms built by Internet financial companies and large banks as examples, and implants a stock prediction model based on Deep Neural Network (DNN). In this context, this paper completes the following work: 1) The research status of big data processing and analysis platforms at home and abroad is introduced. 2) Drawing on the modular design idea, the commercial bank big data platform is designed and the functions of each sub-module are introduced. Then the basic principle and structure of Convolutional Neural Networks (CNN) are expounded. 3) The optimal parameters of Convolutional Neural Networks are selected through experiments, and then the trained model is used for experiments. It can be seen that the stock prediction model proposed in this article has a higher prediction accuracy compared to existing models, which also verifies the validity of the proposed model. Input the data and compare the obtained results with the actual results, and finally show that the model in this paper has a good performance on stock prediction.

1. Introduction

Many new computer applications have emerged in recent years, and the amount of data created by people's everyday activities is becoming ever larger. People's behavioral characteristics are stored in the disk in the form of images, sounds, etc., which is one of the conditions for the formation of this era of big data. People gradually realize that information data is growing at an unprecedented rate, and the data generated by our behaviors are also stored inadvertently, forming our habits [1]. When dealing with very complex tasks, it is particularly important to choose a suitable processing platform. Currently, the more popular massive data processing platforms are Hadoop and Spark [2]. As early as a few years ago, Hadoop and Spark have been used in the industry to some extent. Both Facebook and google have used Hadoop technology to build recommendation systems. Hadoop's excellent processing capability for massive data has led many data companies to start researching and applying it. In addition to the aforementioned massive data processing platforms, data processing models are also constantly evolving. For example, in recent years, deep learning has set off a wave of learning booms in industry and academia [3]. With the deepening of deep learning research, people's recognition of the feature extraction ability of deep learning has also resulted in a series of famous applications based on deep learning. For example, the AlphaGo go robot of the google team has continuously challenged the top human players and achieved world-renowned achievements. Not only that, in the field of artificial intelligence, predecessors are always trying to use more effective and advanced learning methods to change the intelligence level of computers. For example, the team in the United States uses deep reinforcement learning algorithms to improve AI (Artificial Intelligence), so that its AI algorithm can continuously strengthen the automatic learning of Atari's games, and the algorithm can update its own level, so that it can get high scores in the games developed by Atari [4]. These seemingly insignificant little advancements in life actually represent a big step forward in the field of AI. Data processing technology is also developing as the amount of data increases. In practice, the amount of data that needs to be processed is too much, such as video image information in monitoring systems, and game image information in high-speed rendering of video games. All are dozens of frames or even hundreds of frames per second [5]. In this context, if there is a strategy that can reduce the training pressure of the model for a large amount of data and

E-mail address: shengshengking@sitsh.edu.cn.

https://doi.org/10.1016/j.sasc.2024.200107
Received 21 November 2023; Received in revised form 26 April 2024; Accepted 20 May 2024
Available online 21 May 2024

increase the training speed without greatly reducing the performance index, then the strategy will become very meaningful. In any field, the processing models for large amounts of data include simple and fast processing models similar to decision trees, and complex models with many parameters that are difficult to train, such as deep convolutional network models [6]. In the process of processing data for these complex models, if certain a priori processing can be performed according to the characteristics of different data sets to make the processing of the model faster and more convenient, then these strategies have a huge impact on improving the usability of the model. effect. For example, after the simple stacking of DBN (Deep belief networks) model layers, if traditional backpropagation training is used, problems such as gradient disappearance will occur, making training difficult. However, experts proposed a new training strategy, first using unlabeled data to train the RBM (Restricted Boltzmann Machine) network one by one in a layer-by-layer greedy pre-training manner. After training, the network is a simple feature extractor, and then the label data to be trained is imported for fine-tuning to obtain an accurate classification model [7]. From this example, we can see how important the training strategy is to the model. If we add some prior knowledge before training, the model will often get better training results. In machine learning training tasks, model training is often accompanied by many strategies. In these training strategies, some preprocessing can often reduce the data dimension and enhance the model's expressive ability. The models used for training range from simple to complex. If we can perform some processing in advance to reduce the cost of training or improve the effect of training, this strategy will also be considered effective. If some excellent models are appropriate training strategies, they will not be able to give full play to their advanced nature [8]. Therefore, in addition to exploring the model, exploring the training strategy for the model also has far-reaching significance. If you can design some good enough strategies for some data scenarios to reduce unnecessary training, it will also be a great improvement for the model itself. The application of big data technology in life has been very common. When the scale of data is very small, we may be satisfied with some existing algorithms with high accuracy. However, in the context of rapidly changing application scenarios, the amount of data is growing extremely fast. It is very likely that what people pursue at this time is not how to better predict the results of the data, but to quickly obtain a rough result. In the following context, the speed of prediction is often better than the accuracy, because getting results quickly and compensating accordingly is exactly what decision makers in many companies want to do [9]. For example, a game company may release a certain activity to retain some players for later operation before a large number of game users are lost. Therefore, the necessity of quickly obtaining a training effect that is not the optimal result becomes very important, and the means to train the game AI and obtain it quickly can use the training experience related to machine learning. Big data is a revolution that will alter our daily routines, places of employment, and way of thinking. Big data has emerged as a major technical shift in the ICT (information and communications technology) sector, after mobile Internet and cloud computing [10]. China has seen tremendous growth in the use of big data across a wide range of industries since the dawn of the "big data" era, including retail, logistics, medical, the arts, and transportation. Many big data applications may be found in the financial sector, including banking, insurance, securities, trust, and more [11]. It is in this context that this paper designs a financial big data processing and analysis platform based on a DNN model. The platform aims to solve the information island phenomenon caused by the lack of information sharing among various research institutes and different databases. Commonly used financial mathematical models are made possible by integrating significant commercial databases with financial fundamental data and research outcomes. This results in a unified administration and standardization of this data. In the module, a stock forecasting model based on DNN is designed and two simulated stocks are predicted to rise and fall.

In order to effectively predict the stock index and improve the profitability of stock market investments by reducing losses. On the basis of fully considering the nonlinear and non-stationary characteristics of the stock index time series, this article introduces the DNN decomposition method, which decomposes the stock index time series into different scale intrinsic mode functions with a single feature compared to the original signal; Then, the neural network performs dimensionality reduction on it to obtain the most informative features; Finally, the features that can greatly reflect the characteristics of the stock index are input into the neural network for prediction, in order to obtain the prediction results.

2. Related work

Big data is becoming more important to the domestic banking sector. The financial sector, according to recent studies, has the most potential benefit from big data. Banking sector research on big data platforms may be classified into two categories: businesses and banks. It is more reliable to predict corporate loan risk through big data analysis of the industry chain. Label the enterprise from the bank's point of view, and reduce the risk for the banking industry by analyzing the basic information, ownership structure, patent data industry competition and other information of the enterprise [12]. Banks may become more intelligent through gaining a better understanding of their customers' demands, implementing timely and accurate marketing, and providing a tailored user experience. These enterprises and banks rely on big data platforms to analyze user behavior and market environment, so as to provide consulting and services for users' investment and financial decision-making. In addition, other industries in the financial field also have excellent big data analysis applications. The golden era of financial big data development might be considered to be upon us right now, according to certain experts. Big data has ushered in a new era of disruption in the banking industry. Big data finance is attracting conventional financial institutions as well as newer Internet start-ups [13]. Data persistence is the foundation of the big data platform. It is responsible for data storage and is a prerequisite for subsequent data processing and analysis. Currently, according to the data types it can handle, it is divided into relational database solutions represented by DB2 (DB2 Universal Database) and unstructured solutions represented by HBase (Hadoop Database). With the development of technology, Oracle has developed the Oracle RAC parallel database solution, which runs on multiple servers. Compared with earlier databases, different database instances do not share CPU (Central Processing Unit) and memory, thereby achieving horizontal expansion of computing resources. Although scalability is much improved compared to earlier days, IO (Input/Output) is still shared between different instances. In the direction of performance evaluation of big data processing platforms, researchers have done a lot of work and designed a variety of evaluation benchmarks [14]. A metric termed basic operations per second has been developed to quantify the operational capability of large-scale analytical processing systems, and reference [15] introduced it to evaluate data centers. Aiming at the difference of computing power and load diversity of cluster nodes, a scheduling algorithm using resource partitioning is proposed. Financial forecasting may be done in a variety of ways. Qualitative forecasting is a type of financial forecasting that relies on intuition and subjectivity. Experts use historical and current facts, to predict the future development patterns and rules of financial operations based on their knowledge of the subject matter [16]. It takes reasonable problem assumptions as the premise, uses probability and statistics methods to conduct theoretical derivation and obtains the final prediction model. Time series forecasting, regression forecasting, probability forecasting, and combination forecasting are some of the most common methods in this category [17]. It is the process of finding fascinating and important information from enormous amounts of data as well as the prospective, inherent. When it comes to data mining, it is not a single subject but rather a collection of connected ideas and methods from a wide range of fields [18]. It is becoming more difficult to handle a big

volume of data using solely human resources as the quantity and frequency of financial operations increase. Because of this, financial forecasts have increasingly turned to this technique of data mining as well. Financial forecasting industries such as stocks, bankruptcy assessment, exchange rates, and futures all employ data mining techniques. This may be done using data mining techniques without any assumptions or minimal assumptions, and the results can be used to anticipate future trends in financial activity [19]. ANN (Artificial Neural Network), SVM (Support Vector Machine), GA (Genetic Algorithm), and HMM (Hidden Markov Model) are now the most widely used data mining algorithms in financial forecasting. An example of a neural network is one that learns from training samples and generates a mapping between input and output via the computation of many neurons. Many studies have shown that neural networks are capable of accurately modeling and forecasting time series. Using an extended regression neural network technique, reference [20] developed a financial forecasting model that is both faster to use than a typical back-projection neural network and more accurate in its predictions. It is possible to use the BPNN model to accurately forecast binary, multivariate, and continuous data because of its ability to capture nonlinear properties. The accuracy of the forecast, on the other hand, is dependent on the network parameters that are chosen. Overfitting is possible if the training process is managed or the parameters are set incorrectly. Multi-swarm genetic algorithm and neural network are proposed in reference [21], which may take into account both global and local optimization solutions. SVM can accurately forecast five financial time series, according to a study in [22]. According to testing findings, five data samples from the SVM outperform the BPNN when it comes to assessment metrics like MSE (mean-square error) and MAE(Mean Absolute Error). According to Reference [23], a wavelet SVM was developed using the wavelet kernel function instead of the gaussian kernel function to forecast the volatility of two simulated stocks. This exhibited superior prediction performance than the gaussian kernel.

3. Method

3.1. System architecture

In this article, the CNN method was used to decompose the stock index data, extract the data information with the strongest correlation with the stock index, and input the extracted features into BPNN for prediction. The model structure diagram is shown in Fig. 1.

Drawing on the modular design idea, the functional and non-functional requirements of the big data platform of small and medium-sized commercial banks are analyzed in detail, and the big data platform system is divided into 5 sub-modules, as shown in Fig. 2.

The whole big data platform includes 5 sub-modules: 1) Data ETL (Extract-Transform-Load) module. It is mainly responsible for data processing operations such as data collection, preprocessing, loading, cleaning, etc. involved in the entry of source data into the big data platform. 2) Data persistence module. Responsible for storing the data results generated by each processing stage of the data ETL module, and providing basic query and calculation to the outside world. 3) Data service module. Provide data support services for peripheral applications, including data query, analysis, and mining, and provide data push and subscription services for peripheral applications such as performance appraisal, risk control, stock forecasting, and precision marketing. 4) Data management module. Responsible for the management of data in the big data platform, including data quality, life cycle, data standards, access rights, and data governance management. 5) Develop operation and maintenance monitoring module. Responsible for the monitoring of server hardware resources and background services related to the big data platform, the monitoring and scheduling of data ETL task batches, and provide ETL development functions for the scientific and technological personnel of the industry. At the same time, in order to reduce the difficulty of operation and maintenance, it also supports the automatic function management of table structure changes issued by higher-level organizations.

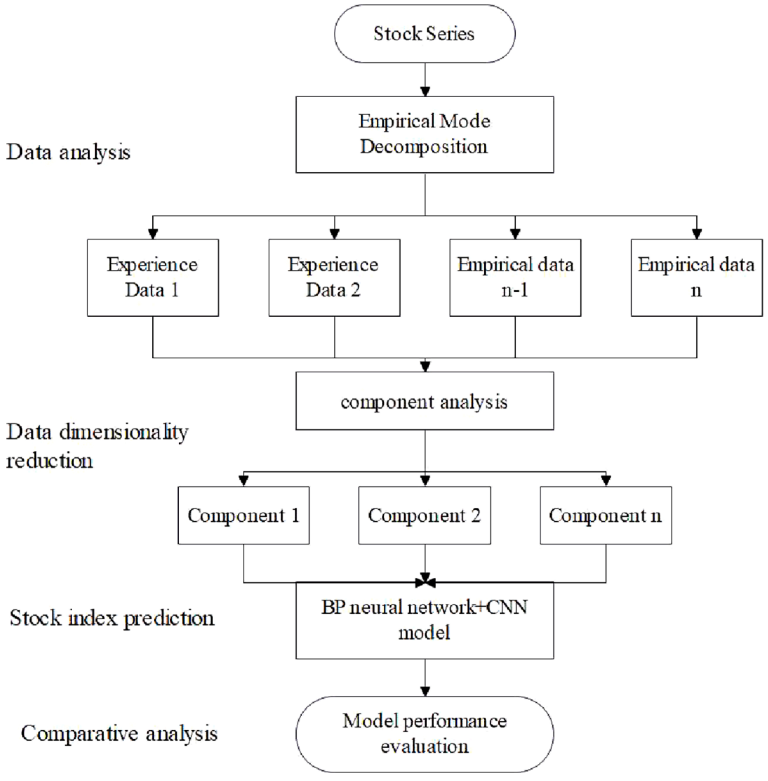

Fig. 1. The model structure diagram is shown.

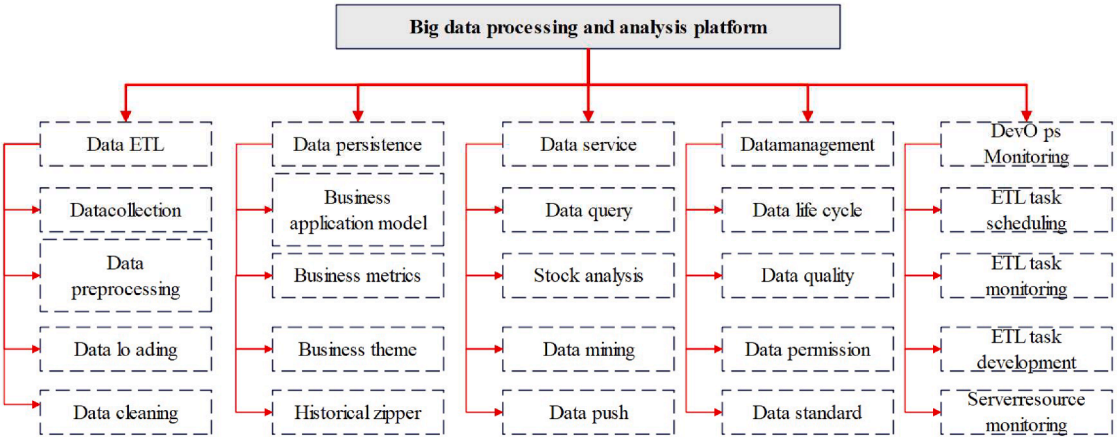

Fig. 2. Functional architecture of big data platform.

3.2. Stock index prediction model based on CNN

Financial models are built on the basis of time series data. An important aspect in making sound monetary choices is understanding how market rules change over time. Predictive models may be used to show the correlation between input and output data. Forecasting financial time series involves feeding past financial data into a forecasting model, which then uses that data as input and uses the model's output to make predictions about the future. The stock index serves as a financial market barometer and a key sign for assessing an industry's health and the efficiency of its operations. As a result, the stock index is an important study subject.

3.2.1. CNN-Based stock index prediction model

Given that CNNs' basic building blocks are convolutional layers (also known as convolution kernel layers) and downsampling layers (also known as downsampling drops), it's easy to see how the structure of a CNN can be influenced by its number of convolutional layers, convolution kernel sizes, and downsampling drops. For financial time series data, variables that occurred in the past have a less impact on current values. This minimizes the computational complexity of the network by simplifying the CNN structure, specifically the size of the convolution kernel. When it comes to image identification, this is by far the most significant departure from the CNN framework. There are two convolutional and downsampling layers in the CNN, and the model structure will be complicated and inconvenient for training if the size of each kernel and the rate of downsampling are variable. Due to these considerations, in this article, we set the two downsampling rates to be equal. Three convolutional and downsampling layers and a single-layer perceptron make up the basic CNN prediction model. Using a fully-connected technique, a single-layer perceptron is created from the output of the second downsampling.

3.2.2. Input sample selection sliding window technique

According to this approach, the value of financial time series may be anticipated using past data that has been intercepted through this window of time, and this value is the forecast value. Sliding windows, landmark windows, and tilted windows are the three most common ways of window interception. But the window slides with time, allowing data to be captured. The width of the sliding window, n, is the single parameter of the sliding window. For the purposes of this work, the first n-1 financial data and the nth financial data are utilized as input samples.

3.2.3. CNN model structure parameter relationship

The length of the basic input for the model in this article is represented as 1, the convolutional kernel of the system model in this article is represented as C, and the downsampling of the system model is set as s. When the convolutional and downsampling layers of the system are 2 and remain unchanged, there is the following relationship:

$$\frac{\left[\frac{l-c+1}{s}-c+1\right]}{s}=N \tag{1}$$

If the system does not have a two-layer system model and only has one layer, the above parameters can be re expressed as:

$$\frac{l-c+1}{s}=N \tag{2}$$

where N is a positive integer and s is not equal to 1.

It is dependent on the properties of the financial time series data and the duration of the input sample how long a value at a given instant may have an influence.

3.2.4. CNN model algorithm implementation

Basic parameters such as how many convolutional layers and how much downsampling are included in the model structure initialization process are determined. The convolution kernel connection weight, bias value, parameter settings of the single-layer perceptron, etc. are all specified via model parameter initialization. between (−1, 1) is randomly given to convolution kernel weights and the weights of the single-layer perceptron connections. Bias is set to 0. Gradient application and parameter update are all included as part of the model's training procedure. Once you've gone through a set amount of iterations, you're done training. A random selection of training samples is used in each iteration to limit the influence of the training samples on prediction outcomes. Validity of the model is verified by conducting a test set, which results in the appropriate evaluation index values.

4. Experiment and analysis

4.1. Data sample source and preprocessing

1) The experimental data adopts the highest price of a country's stock market from January 1, 2000 to December 30, 2010. RMSE, R-square, the actual output sequence are all used to evaluate the accuracy of the prediction findings. The RMSE is a measure of the forecast's total error. Anticipated output sequence correlation coefficient reveals if the predicted price trend is in agreement with actual price trend. Price forecasting accuracy improves as correlation coefficients go closer to 1. The greater the ability to forecast is, the closer the R-squared index is to 1. A stock index's other parameters will stay unaltered while discussing the impact of a single parameter on its forecast outcome.

2) Data preprocessing. Financial time series data is characterized by non-stationarity and noise, which may lead to data loss and abnormalities if a sufficient quantity of data is collected. Prior to model training and prediction, data has to be preprocessed because of these considerations. According to the two primary types of data preparation procedures, there are many: data cleaning and transformation activities. The primary goal of data cleaning is to eliminate errors, outliers, and other undesirable elements from the raw data. Normalization, dimension reduction, and transformation are the most common data transformation processes. An important part of this paper's data preparation is removing missing values and outliers. handling of missing values and outliers: More often than not, outliers are seen in numbers that are less than 0. Filling and replacing the current value with that of the preceding instant is the third strategy used to deal with missing data or outliers in this study. Financial time series data must be denoised before they may be smoothed to improve prediction accuracy. Smoothing the data is the best approach to eliminate noise. Smoothing and denoising of financial time series may be performed using a variety of methods, including kernel smoothing, which has been utilized for many years. The data in this research has been cleaned up using the kernel smoothing approach, which is presented in this paper. There are several ways to do this, but a kernel function may be used as a first step. Filtering noise can be expressed as:

$$m_h(x) = \frac{\sum_{i=1}^{T} K_h(x - X_t) Y_t}{\sum_{i=1}^{T} K_h(x - X_t)}, \; K_h(x) = \frac{1}{h\sqrt{2\pi}} e^{-\frac{x^2}{2h^2}} \tag{3}$$

where $K_h(x)$ is the kernel function, Y_t is the financial data of the time series, X_t represents the coordinates of the sample center point Y_t, x represents the distance between the center point and its nearby observation sample points, h is the bandwidth of the kernel smoothing, $m_h(x)$ is the corresponding series value of the processed financial time series data Y_t.

In order to have a clear picture of what is going on, the data must be constrained to a certain range. In most cases, the data is restricted to the intervals [−1, 1] or [0, 1]. Data reduction to [0, 1] yields better predictions than data reduction to the range of [−1, 1], according to this paper's comparisons. So this paper chooses to reduce the data to the range of [0, 1] and normalize it. The formula is:

$$y_t^{'} = \frac{y_t - \min\{Y_t\}}{\max\{Y_t\} - \min\{Y_t\}} \tag{4}$$

where $\{Y_t\}$ is the financial data of the time series, y_t is the series value at time t, $\max\{Y_t\}$ is the maximum value of the series, and $\min\{Y_t\}$ is the minimum value of the series.

4.2. The influence of model parameters on the prediction results

CNN-related parameters will be examined in this part to establish the best parameters and stock forecasting models for CNN-related forecasting outcomes. The model's capacity to extract features from the data is reflected in the number of convolutional and downsampling layers. The single-layer perceptron's excitation function in the CNN structure impacts the prediction outcomes in addition to the four criteria listed above. The excitation function is found to be a sigmoid function after an experimental comparison.

4.2.1. The effect of convolutional and downsampling layers on results

For each convolutional layer, two input sample lengths are tested. The 40- and 50-sample input samples are partitioned into 3 and 5 convolution kernels, respectively. The downsampling reduction is two when there are two convolutional layers and two downsampling layers, resulting in a total of five convolution kernels and ten downsampling layers. The experimental results are shown in Table 1.

4.2.2. Influence of convolution kernel size on prediction results

Finally, four different convolution kernel sizes were used in the prediction and comparison experiments: 15, 11, 7, and 3. Convolutional neural networks with two convolutional layers and downsampling layers predict stock prices for four different convolution kernel sizes, as shown in Fig. 3. The RMSE grows as the convolution kernel size decreases, but it begins to decrease when the convolution kernel size reaches 3. The correlation coefficient becomes closer to one as the size of the kernel gets smaller. It illustrates that the projected stock trend is closer to the genuine changing trend when the convolution kernel is smaller. As the convolution kernel size shrinks, the R-squared value becomes closer and closer to 1. It indicates that when the convolution kernel is smaller, the prediction accuracy is greater. It is clear that the short-term properties of the time series stock index contribute to an improved prediction impact when the convolution kernel is smaller.

4.2.3. The influence of the number of convolution kernels on the results

If there are four convolution kernels in the first layer of convolution, there are eight kernels in the second layer of convolution (4, 8). The input sample length is 60, and the downsampling reduction is 2. This section of the experiment is performed in this way. Following the findings of the previous experiments, the kernel size is set at 3. Experiments on four different convolution kernel sizes are conducted, and the results are compared for each of the four different convolution kernel sizes. The predicted effect is shown in Fig. 4.

Stock prediction indicators for each of the four experimental groups are shown in Fig. 4. A V-shaped change trend in prediction quality can be noticed in the image, which illustrates that the number of convolution kernels does not alter prediction quality monotonically. The RMSE of the prediction result is the lowest and the R-squared is the closest to 1 when the number of two convolution kernels is 5 and 10. However, it does not vary much from other situations in terms of correlation coefficient, demonstrating that a smaller or larger number of convolution kernels does not have much influence on the general trend of the forecast. The image also shows that the maximum error value is less than 0.1 when the number of convolution kernels is 5 and 10. Using the above-mentioned indications, this research concludes that using 5 and 10 convolution kernels for the CNN stock prediction model yields the best results and is the most logical option.

4.3. Optimal model performance test

Through the above experiments, the best parameters of the model are selected and then trained, and the trained model is used for stock simulation prediction. The results of the model output are compared with the actual results, and the results are shown in Table 2. It can be seen that the predicted results are very close to the actual results, indicating that the model has excellent performance in stock prediction.

Table 1
Prediction result indicators of different convolutional layers.

Sample length	Convolutional and downsampling layers	Convolution kernel size	RMSE	Correlation coefficient	R-square
40	1	3	4.9e-04	0.9335	1.2231
40	2	3	4.1e-04	0.9417	1.2307
50	1	5	7.5e-04	0.8962	1.2653
50	2	5	5.6e-04	0.9151	1.1662

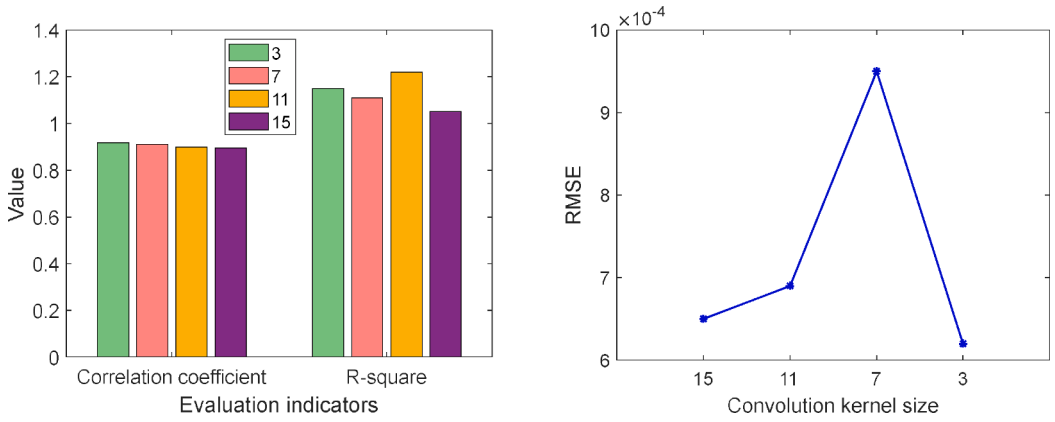

Fig. 3. Prediction outcome metrics for different convolution kernel sizes.

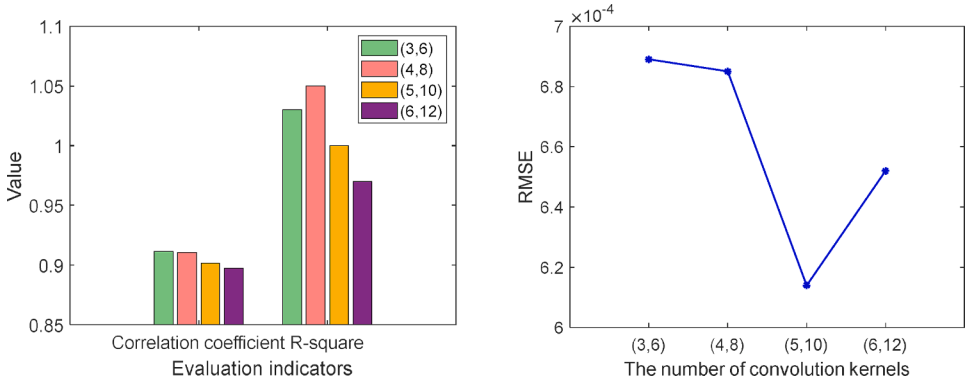

Fig. 4. Prediction outcome metrics for different number of convolution kernels.

Table 2
Experimental comparison between model output and actual result.

Number	1	2	3	4	5	6	7	8
Model output	0.875	0.868	0.723	0.734	0.899	0.918	0.871	0.739
Actual result	0.874	0.867	0.723	0.734	0.900	0.919	0.872	0.740

4.4. Test model comparison experiment

Compare the model of this article with references [10,11], and [12], obtain stock data from the network for the past three years, use the first two years of data as training data, and use the last year of data as prediction data. Compare the predicted data of the model with the actual data to obtain the accuracy of data prediction. Finally, the experimental results shown in Table 3 are obtained.

From the above comparative experimental results, it can be seen that the stock prediction model proposed in this article has a higher prediction accuracy compared to existing models, which also verifies the validity of the proposed model.

Table 3
Statistics of Stock Prediction Results by the Model.

Model	Prediction accuracy
The model of reference [10]	46.32 %
The model of reference [11]	35.32 %
The model of reference [12]	40.87 %
The model of this article	69.65 %

5. Conclusion

Financial sector data governance is the primary focus of the big data platform. As the financial sector continues to grow rapidly, the quantity of data it generates will similarly grow at an exponential rate over time. Therefore, through this platform, they can save a lot of time in data processing and management, and focus on other trend analysis of data. The age of big data has come of age in tandem with the fast growth of the Internet. With the fast advancement of machine learning, deep learning, and artificial intelligence, there will be more and more data governance methods and algorithms. The quality of data governance will also gradually increase, which will also increase enterprises' dependence and trust in data governance, and at the same time help enterprises to better process and analyze data. In this context, this paper completes the following work: 1) The research status of big data processing and analysis platforms at home and abroad is introduced. 2) Drawing on the modular design idea, the commercial bank big data platform is designed and the functions of each sub-module are introduced. Then the basic principle and structure of CNN are expounded. 3) The optimal parameters of CNN are selected through experiments, and then the trained model is used for experiments. Input the data and compare the obtained results with the actual results, and finally show that the model in this paper has a good performance on stock prediction. it can be seen that the stock prediction model proposed in this article has a higher prediction

accuracy compared to existing models, which also verifies the validity of the proposed model.

CRediT authorship contribution statement

Sheng Huang: Writing – review & editing, Writing – original draft, Visualization, Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] R. Beakta, Big data and hadoop: a review paper, Int. J. Comput. Sci. Inf. Technol. 2 (2) (2015) 13–15.
[2] W. Zhang, J. Chang, Z. Lei, et al., MET-COFEA: a liquid chromatography/mass spectrometry data processing platform for metabolite compound feature extraction and annotation, Anal. Chem. 86 (13) (2014) 6245–6253.
[3] Z. Zhang, A. Zhang, G. Xiao, Improved protein hydrogen/deuterium exchange mass spectrometry platform with fully automated data processing, Anal. Chem. 84 (11) (2012) 4942–4949.
[4] Y. Nakamura, T. Higaki, F. Tatsugami, et al., Possibility of deep learning in medical imaging focusing improvement of computed tomography image quality, J. Comput. Assist. Tomogr. 44 (2) (2020) 161–167.
[5] H.L. Ciallella, H. Zhu, Advancing computational toxicology in the big data era by artificial intelligence: data-driven and mechanism-driven modeling for chemical toxicity, Chem. Res. Toxicol. 32 (4) (2019) 536–547.
[6] U.R. Acharya, S.L. Oh, Y. Hagiwara, et al., A deep convolutional neural network model to classify heartbeats, Comput. Biol. Med. 89 (2017) 389–396.
[7] X. Zhang, H. Dai, Significant wave height prediction with the CRBM-DBN model, J. Atmos. Oceanic Technol. 36 (3) (2019) 333–351.
[8] X. Yuan, C. Ou, Y. Wang, et al., A novel semi-supervised pre-training strategy for deep networks and its application for quality variable prediction in industrial processes, Chem. Eng. Sci. 217 (2020) 115509.
[9] R.K. Yadav, PSO-GA based hybrid with Adam Optimization for ANN training with application in Medical Diagnosis, Cogn. Syst. Res. 64 (2020) 191–199.
[10] S. Gössling, Technology, ICT and tourism: from big data to the big picture, J. Sustain. Tourism 29 (5) (2020) 849–858.
[11] M. Bennett, The financial industry business ontology: best practice for big data, J. Bank. Regul. 14 (3) (2013) 255–268.
[12] B. Fang, P. Zhang, Big data in finance. Big Data Concepts, Theories, and Applications, Springer, Cham, 2016, pp. 391–412.
[13] J. Lawler, A. Joseph, Big data analytics methodology in the financial industry, Inf. Syst. Educ. J. 15 (4) (2017) 38.
[14] C.T. Yang, J.C. Liu, S.T. Chen, et al., Implementation of a big data accessing and processing platform for medical records in cloud, J. Med. Syst. 41 (10) (2017) 1–28.
[15] D. Singh, C.K Reddy, A survey on platforms for big data analytics, J. Big Data 2 (1) (2015) 1–20.
[16] L. Di Persio, O. Honchar, Recurrent neural networks approach to the financial forecast of Google assets, Int. J. Math. Comput. Simul. 11 (2017) 7–13.
[17] M. Arvan, B. Fahimnia, M. Reisi, et al., Integrating human judgement into quantitative forecasting methods: a review, Omega (Westport) 86 (2019) 237–252.
[18] L.B. Tang, L.X. Tang, H.Y Sheng, Forecasting volatility based on wavelet support vector machine, Expert Syst. Appl. 36 (2) (2009) 2901–2909.
[19] A. Alyaseen, A. Poddar, N. Kumar, P. Sihag, D. Lee, T. Singh, Assessing the compressive and splitting tensile strength of self-compacting recycled coarse aggregate concrete using machine learning and statistical techniques, Mater. Today Commun. 38 (2024) 107970.
[20] A. Alyaseen, A. Poddar, N. Kumar, S. Tajjour, C.V.S.R. Prasad, H. Alahmad, P. Sihag, High-performance self-compacting concrete with recycled coarse aggregate: soft-computing analysis of compressive strength, J. Build. Eng. 77 (2023) 107527.
[21] A. Alyaseen, A. Poddar, H. Alahmad, N. Kumar, P. Sihag, High-performance self-compacting concrete with recycled coarse aggregate: comprehensive systematic review on mix design parameters, J. Struct. Integr. Maint. 8 (3) (2023) 161–178.
[22] V. Kashyap, A. Alyaseen, A. Poddar, Supervised and unsupervised machine learning techniques for predicting mechanical properties of coconut fiber reinforced concrete, Asian J. Civ. Eng. (2024) 1–21.
[23] A. Alyaseen, C.V.S.R. Prasad, A. Poddar, N. Kumar, R.R. Mostafa, F. Almohammed, et al., Application of soft computing techniques for the prediction of splitting tensile strength in bacterial concrete, J. Struct. Integr. Maint. 8 (1) (2023) 26–35.