The Mechanism of the Role of Big Data Knowledge Management in the Development of Enterprise Innovation

Guangyu Yan[1], Rui Ma*[2]
Department of Scientific Research, Chengdu Polytechnic, Chengdu, 610041, China[1]
Health and Rehabilitation College, Chengdu University of Traditional Chinese Medicine, Chengdu, 611137, China[2]

***Abstract*—The effectiveness and efficiency of enterprise knowledge management depends on the effectiveness and efficiency of the enterprise's implementation of knowledge management. Big data technology can collect, analyse and apply the massive amount of data in an organisation to support the implementation of knowledge management. Therefore, exploring the role of big-data knowledge management in the development of enterprise innovation will help enterprises to better implement knowledge management. Based on this, the study aims to propose a model for predicting big data knowledge management and enterprise innovation development for high-tech enterprises in China. The study firstly used Principal Component Analysis (PCA) to decrease the dimensionality of the model, and then used the particle swarm algorithm to optimize BP neural network (PSO-BP). Network (PSO-BP) was used to evaluate enterprise knowledge management and enterprise innovation development. The results of the study show that the absolute values of the relative errors of the pre-processed model do not exceed the 5% threshold, and only the relative errors of some indicators are relatively large, such as X5 and X7, with values of 4.5% and -3.8%, indicating that the model has a good performance in predicting the innovation effect of enterprises.**

Keywords*—Big data knowledge management; BP neural network algorithm; enterprise innovation development; principal component analysis; particle swarm optimization algorithm; correlation analysis

I. INTRODUCTION

With the advent of the information age, the amount of data and information has increased exponentially, and how to effectively manage big data knowledge has become an important research direction for the current development of the country and enterprises [1,2]. At the same time, managing big data can strengthen enterprises' cognition of data-related knowledge, find their own positioning, better access the dividends of the times, and enhance their attention to big data capabilities. At the same time the rational use of big data can also enable the whole enterprise to enhance the innovation ability of products and improve the performance of the enterprise [3]. However, the current research on big data is at the initial stage, and many of the constructed algorithm models cannot comprehensively analyze the characteristics of big data knowledge management, for example, the commonly used back propagation neural network (BPNN) algorithm cannot analyze larger and broader data, while the slow computing speed also causes problems for data analysis of big data knowledge management. Although the particle swarm optimization (PSO) algorithm is simple and easy to implement with few parameters, the poor performance and troublesome network parameters also make the algorithm unable to better solve the problems arising in big data knowledge management. Meanwhile, many current studies on big data knowledge management are still in the dialectical analysis stage [4-5]. Based on this, this experiment will study big data knowledge management by using principal component analysis (PCA) to organize and analyze the data, then using PSO algorithm to determine the weights of the factors influencing knowledge management on the development of enterprise innovation (EnIn), and finally using BPNN algorithm to predict the data set and determine the role of big data knowledge management on the development of EnIn and the feasibility of its development. This research is divided into four parts, the first part is to explain the current research status of big data knowledge management at home and abroad, the second part establishes a new optimization algorithm model by analyzing the indicators of big data characteristics, the third part is to analyze the performance of the optimization algorithm and the data processing results, and the fourth part is to conclude the whole article.

II. RELATED WORK

Goncharenko et al. aimed to study the functional support of organisational and economic mechanisms of innovation and integration potential management in enterprises. The study examines the organisational and economic mechanisms of innovation and integration potential management in enterprises. The study showed that organisational and economic mechanisms are the basis for the management of innovation and integration potential of enterprises and that their functional support is divided into four areas: internal management support, market support, policy support and social support [6]. Zadorozhnyi et al. aimed to explore the determinants of innovation affecting enterprises and assessed them on the basis of actual data from financial statements. It was found that the financial statements of firms can provide useful information reflecting the innovation of the firm, which includes investment, profit and asset structure. In addition, financial statements can increase the transparency of financial statements by providing information on details such as technology development, organisation and management of the firm [7]. Katsarski discussed the relationship between integrated business management and water ecosystems, summarised the importance of integrated business management and its role in water ecosystem conservation, and made recommendations for future research and practice [8]. Vasyltsiv et al. aimed to explore the creativity, information and knowledge determinants of economic growth in the EU region in the context of smart development strategies. The study provides an integrated analysis of smart development strategy research and economic growth. The findings show that economic growth in the EU region depends mainly on policy creativity, effective transmission of information and the driving force of knowledge. In addition, factors such as technological innovation and innovative social networks also have a significant impact on economic growth in the EU region [9]. Straková et al. examine trends in organisational and managerial structures, exploring many of the key issues involved in practice and research, such as organisational structure, management theory, organisational culture and leadership styles. The article also details the changes in organisational and management structures and how the changes affect organisational performance. Finally, the article offers suggestions for addressing these issues to help managers better manage organisational and management structures [10].

To explore the strategy of obtaining sustainable competitive advantages in emerging wine producing regions in southern Sweden, Kompaniets introduced the concept of this advantage and explored various strategies. This includes resource base, technology base, customer base, market base, and organizational base, which can be transformed into a sustainable competitive advantage for the region by implementing specific strategic measures [11]. Chatterjee et al. aimed to explore knowledge sharing for product and process innovation in international markets in order to better knowledge sharing and thus more benefits of innovation [12]. Lopes et al. aimed to explore how competitiveness management, knowledge management and corporate education are implemented in Brazilian companies, collecting data from three Brazilian companies. The study found that competitiveness management and knowledge management contributed to the efficiency of the enterprise, while corporate education helped to improve the skills and knowledge of employees [13]. Babgohari et al. aimed to explore the relationship between knowledge management competencies, entrepreneurial creativity, entrepreneurial passion and corporate performance, processing and analysing survey data from 385 of these companies. It verified that there was a prominent positive relationship between knowledge management capabilities and entrepreneurial creativity, entrepreneurial passion and firm performance, with the dual power of the firm playing a mediating role [14]. Wang et al. aimed to explore the relationship between market orientation and service innovation, and the study used a quantitative research approach to analyse data from three major Chinese airlines. The conclusion was that there was a remarkable positive relationship between market orientation and service innovation, while knowledge sharing contributed moderately to this relationship. The results provide valuable management guidance for the company's market orientation and service innovation activities [15].

It can be seen through the research of scholars at home and abroad that there are more studies on the relationship between enterprise knowledge management and EnIn, but most of them stay in the perspective of empirical analysis and do not adopt algorithmic models to further dissect them. Based on this, the research is mainly built on the PCA-PSO-BPNN algorithm to design and study the role mechanism of big data knowledge management in the development of EnIn, and then analyse the specific mechanism of big-data knowledge management on the development of EnIn.

III. Building a Machine Learning-based Enterprise Innovation Model in a Big Data Environment

This chapter mainly provides an overview of the characteristics of big data knowledge management and data management process, discusses some data knowledge management metrics and pre-processing work, then establishes a big data knowledge management algorithm model by BP neural network, and finally improves and optimizes the BP neural network algorithm model by combining with PSO neural network.

A. Establishment of Big Data Knowledge Management in Enterprise Innovation Index System and Pre-Processing Work

Big data knowledge management mainly involves various aspects such as data collection, data cleaning, data analysis, knowledge discovery and knowledge application [16]. When establishing an EnIn model, the EnIn index system should first be established to detect the innovation level of the enterprise, and the technology of big data knowledge management can be used to realize data collection, data cleaning and data analysis, so as to establish an EnIn index system. In establishing the EnIn model, machine learning technology can be used to fully explore the innovation elements in the EnIn index system, so as to establish a machine learning-based EnIn model. According to the EnIn index system, a classification model can be constructed using supervised learning methods in combination with machine learning techniques to detect the innovation level of the enterprise. The research has initially constructed the big data capability index system and the EnIn performance system (Table I).

TABLE I. BIG DATA CAPABILITY INDICATORS AND ENTERPRISE INNOVATION PERFORMANCE INDICATOR SYSTEM

Main target	Subgoal	Numbering	Index
Metrics for big data capabilities	Access to resources Ability to access resources	X_1	Ability to access internal and external data sources on a continuous, real-time basis to support the company's business
		X_2	Possess and master the technical equipment required for big data analysis
	Analysis of integration Integration capabilities	X_3	Ability to process large amounts of data and obtain valuable information
		X_4	Ability to analyse a wide range of structured and unstructured data in real time
	Insightful anticipation Predictability	X_5	The company is able to achieve real-time market insights based on big data and thus identify new business opportunities opportunities
		X_6	Ability to forecast consumer behaviour and corporate opinion based on big data
	Knowledge Discover	X_7	There is a dedicated process for obtaining relevant information from external sources
		X_8	Access to timely and relevant information on customers and suppliers
	Knowledge Integration	X_9	Ability to effectively integrate internally created knowledge with externally acquired knowledge
		X_{10}	Ability to effectively integrate knowledge belonging to different technologies or application areas
	Knowledge Applications	X_{11}	Regularly evaluate and adjust our forecasts based on new knowledge of market trends and technology Long-term forecasts
		X_{12}	regularly evaluate our investment and resource allocation decisions in the light of new knowledge
A measure of corporate innovation and development	Patent results	A_1	The research uses the R&D input ratio to total assets to express R&D intensity
	R&D intensity	A_2	The research uses the amount of utility model, invention and design patents granted to measure the number of patent achievements in the innovation development of enterprises

As shown in Table I, the study divided the big data capability indicators into eight categories, based on which the study further divided them into six indicators of enterprise access to big data capability and two indicators of EnIn performance. The six indicators of enterprise access to big data capability were, in order, enterprise resource access capability, enterprise analysis and integration capability, enterprise insight and prediction capability, enterprise knowledge discovery capability, enterprise knowledge integration capability and enterprise knowledge application capability; the innovation development of enterprises is divided into two major indicators, namely patent achievement and R&D intensity. Based on this, the study selected the financial data of all A-share listed companies in the high-tech industry in Shanghai and Shenzhen from 2013 to 2020. The sample was selected as follows: firstly, A-share listed companies in the financial and insurance industries were excluded; secondly, there were and are only A-share shares in circulation; thirdly, listed companies with missing relevant data were excluded, and the final sample size was 1344; fourthly, R&D expenditure data were mainly collected manually by reviewing the annual reports of listed companies. In view of the differences between the old and new standards on the accounting treatment of R&D expenditures, we specifically divided the time periods from 2013 to 2016 and 2017 to 2020 for collection respectively, and other research data were obtained from the CSMAR database.

The big data knowledge management in EnIn index system constructed by the study includes 14 indicators, and if all of them are incorporated into the BP neural network model, it will make the complexity and operation speed of the network increase significantly, the network performance decreases and the generalization ability of the neural network decreases [17]. Therefore, it is essential to comprehensively analyse the micro factors that affect the degree of innovation and scientific research of enterprises, and reduce the number of indicators while minimising information loss, which means reducing the dimensionality of evaluation indicators while ensuring the evaluation effect. PCA is a multivariate statistical correlation analysis of multivariate correlations, where a small number of main components (linear combination of the original variables) are used to account for changes in multiple variables, i.e. a small number of main components are deduced, thus keeping as much information as possible about the original variables and not correlating with each other, thus making the data more simplified. The specific working procedure is as follows, first setting the indicators for the evaluation of corporate innovation research as $x' = (x1', x2', \cdots, xn')$ and assigning an empirical weight to these parameters $W_x = (W_{x1}, W_{x2}, \cdots, W_{xn})'$, where $W_{x1} + W_{x2} + \cdots + W_{xn} = 1$. Then find out its covariance, i.e. , find out the eigenvalues of $V=\begin{Bmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ v_{n1} & v_{n2} & \cdots & v_{nn} \end{Bmatrix}$ V and arrange them in order of alphabetical size, i.e. $\lambda_1 > \lambda_2 > \cdots > \lambda_l, \lambda_{l+1} = \lambda_{l+2} = \cdots = \lambda_n = 0$, find out the cumulative contribution ratio and get the first m eigenvolume at $\sum_{j=1}^{m}(\lambda_j / \sum_{i=1}^{1}\lambda_i) \geq 90\%$ first and round off the others to find out the eigenvector corresponding to $\lambda_j(j=1,2,\cdots,m)$ $r_j(j=1,2,\cdots,m)$, find out its $x_1 = y_1{'}w_x, x_2 = y_2{'}w_x, \cdots, x_m = y_m{'}w_x$, so there is $x=(x_1, x_2, \cdots, x_m)'$. To better grasp the role of enterprise big data capability on the impact of R&D EnIn.

B. *Establishment of an Original Model of Corporate Innovation Based on the BPNN Algorithm*

Neural networks are a multidisciplinary intersection whose definition varies greatly across disciplines. Research has proposed one of the most widely used concepts to date, which can be used to simulate various informations in the brain by organising the network with a variety of intelligent behaviours such as self-organisation, self-learning and self-adaptation [18]. Its individual neuron model is Fig. 1.

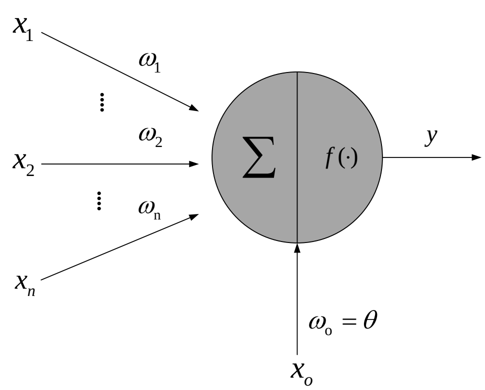


Fig. 1. M-P neuron model.

The specific neuron model listed in Fig.1 is also known as the M-P neuron model. $x_i(i=1,2,\cdots,n)$, i x_0 $\omega_i(i=1,2,\cdots,n)$ x_i θ $f(\cdot)$ y x_i x_0 The signal transmission procedure between x_0 and x_i is approximately as follows: first, the input signal from the n neuron connected to the current neuron x_i is received and the corresponding connection weighting ω_i completes the transmission of the full input signal, followed by the activation threshold x_0 to compare the activation threshold of with the full input signal received. threshold θ to compare the start threshold of x_0 with the total input signals received. The final output y is displayed in the Equation (1).

$$y = f\left(\sum_{i=1}^{n}\omega_i x_i - \theta\right) \tag{1}$$

In the M-P neuron model, a step function is a temporal function with a specific continuum that converts inputs into outputs, and its expression is shown in Equation (2).

$$f(x)=\begin{cases} 0 & x<0 \\ 1 & x\geq 0 \end{cases} \tag{2}$$

The Sigmoid function has the properties of a single increasing and inverse function, and its function image is shown in Fig. 2.

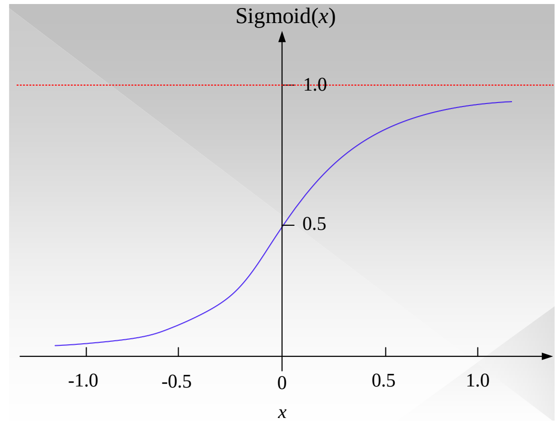


Fig. 2. Sigmoid function.

The M-P neuron model describes the neuron in terms of a logical function, which allows it to theoretically understand the information better. The input layer of the perceptron is in line with the M-P neuron as the output layer, and the specific model structure is Fig. 3.

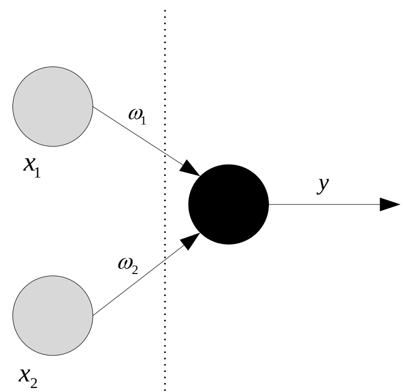


Fig. 3. Perceptron model with two input neurons.

During the training period of the perceptron, it is assumed that the weighting value at a node is $\omega_i(i=1,2,\cdots,n)$. The training sample set denotes (x,y), the actual output value is y' and the learning rate is η. Then the perceptron will modify the learning rules for the connection weighting ω as shown in Equations (3) and (4).

$$\omega_i \leftarrow \omega_i + \Delta\omega_i \tag{3}$$

$$\Delta\omega_i = \eta(y-y')x_i \tag{4}$$

From Equation (3) and Equation (4), it can be seen that there is no change in the connection weighting ω ,when the sensing machine correctly predicts the training example (x,y) , which is $y=y'$, otherwise the connection weights ω will be adjusted and modified accordingly according to the learning rules. The multi-layer neural network structure is exhibited in the form of Fig. 4.

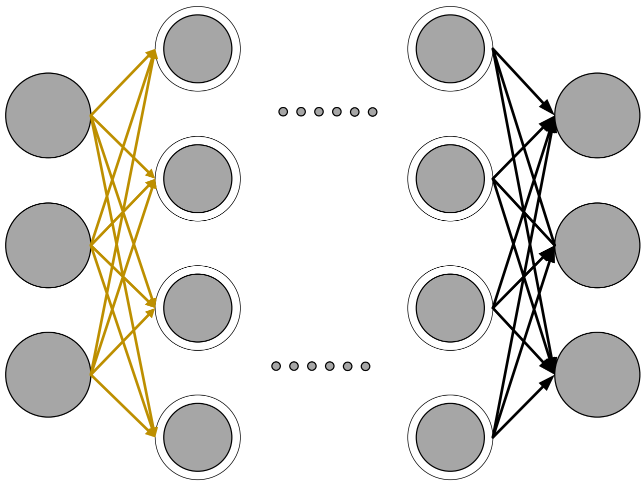


Fig. 4. Multilayer feedforward neural network model.

As shown in Fig. 4, the samples of the training set are fed from the input layer of the multilayer feedforward network, and then the output of the neuron in each layer is considered as its next input. The specific unfolded connection weight matrices W^1 and W^2 matrices are shown in Equations (5) and (6).

$$W^1 = \begin{bmatrix} \omega_{11}^1 & \omega_{12}^1 & \cdots & \omega_{1m}^1 \\ \omega_{21}^1 & \omega_{22}^1 & \cdots & \omega_{2m}^1 \\ \vdots & \vdots & \vdots & \vdots \\ \omega_{l1}^1 & \omega_{l2}^1 & \cdots & \omega_{lm}^1 \end{bmatrix} \quad (5)$$

$$W^2 = \begin{bmatrix} \omega_{11}^2 & \omega_{12}^2 & \cdots & \omega_{1l}^2 \\ \omega_{21}^2 & \omega_{22}^2 & \cdots & \omega_{2l}^2 \\ \vdots & \vdots & \vdots & \vdots \\ \omega_{n1}^2 & \omega_{n2}^2 & \cdots & \omega_{nl}^2 \end{bmatrix} \quad (6)$$

In Equations (5) and (6), m, l and n denote the neurons number in the input, hidden and output layer, respectively. The neuron activation threshold vectors in the hidden and output layers of the network are distributed as $\theta^1 = [\theta_1^1, \theta_2^1, \cdots, \theta_l^1]'$ and $\theta^2 = [\theta_1^2, \theta_2^2, \cdots, \theta_n^2]'$, and the output of the neuron O_j can be pushed out of the network's hidden layer, as Equation (7).

$$O_j = f\left(\sum_{i=1}^{m} \omega_{ji}^1 x_i - \theta_j^1\right) = f(net_j) \quad (7)$$

In Equation (7), $j = 1, 2, \cdots, l$ and $f(\bullet)$ are the functions of the activation functions expressed in the hidden layer, $net_j = \sum_{i=1}^{m} \omega_{ji}^1 x_i - \theta_j^1$. The resulting expression for the neuron z_k in the output layer is shown in Equation (8).

$$z_k = g\left(\sum_{j=1}^{l} \omega_{kj}^2 O_j - \theta_k^2\right) = g(net_k) \quad (8)$$

In Equation (8), $k = 1, 2, \cdots, n$, $g(\bullet)$ is the activation function of the output layer, and the error value E is derived by performing an error operation on the true output value and the desired output value as shown in Equation (9).

$$E = \frac{1}{2}\sum_{k=1}^{n}\left(y_k - g\left(\sum_{j=1}^{l} \omega_{kj}^2 f\left(\sum_{i=1}^{m} \omega_{ji}^1 x_i - \theta_j^1\right) - \theta_k^2\right)\right)^2 \quad (9)$$

The link weights of the neurons in the hidden and output layers were optimally learned by calculating the error value of the network E and the partial derivative of the error value E with respect to the link weights ω_{kj}^2 is expressed as shown in Equation (10).

$$\frac{\partial E}{\partial \omega_{kj}^2} = \frac{\partial E}{\partial z_k}\frac{\partial z_k}{\partial \omega_{kj}^2} = -(y_k - z_k) g'(net_k) O_j = -\delta_k^2 O_j \quad (10)$$

In Equation (10), $\delta_k^2 = (y_k - z_k) g'(net_k)$, the connection weights of the neurons in the input and hidden layers are optimally investigated and the link weighting of the neurons in the input and hidden levels is optimally investigated and the error values are biased differential for the connection weights ω_{ji}^1 as shown in Equation (11).

$$\frac{\partial E}{\partial \omega_{ji}^1} = \sum_{k=1}^{n}\sum_{j=1}^{l} \frac{\partial E}{\partial z_k}\frac{\partial z_k}{\partial O_j}\frac{\partial O_j}{\partial \omega_{ji}^1} = -\delta_j^1 x_i \quad (11)$$

In Equation (11), $\delta_j^1 = \sum_{k=1}^{n} (y_k - z_k) g'(net_k) \omega_{kj}^2 f'(net_j) = f'(net_j) \sum_{k=1}^{n} \delta_k^2 \omega_{kj}^2$, the connection weights between its networks ω_{ji}^1 and ω_{kj}^2 modified by Equation (10) can be derived as shown in Equation (12).

$$\begin{cases} \omega_{ji}^1(t+1) = \omega_{ji}^1(t) + \Delta\omega_{ji}^1 = \omega_{ji}^1(t) - \eta^1 \dfrac{\partial E}{\partial \omega_{ji}^1} = \omega_{ji}^1(t) + \eta^1 \delta_j^1 x_i \\ \omega_{kj}^2(t+1) = \omega_{kj}^2(t) + \Delta\omega_{kj}^2 = \omega_{kj}^2(t) - \eta^2 \dfrac{\partial E}{\partial \omega_{kj}^2} = \omega_{kj}^2(t) + \eta^2 \delta_j^2 O_i \end{cases} \quad (12)$$

In Equation (12), the learning efficiencies in the hidden and output layers are η^1 and η^2. For the neuronal activation threshold of the network output layer θ_k^2, the partial differential representation of the computational error value E is shown in Equation (13).

$$\frac{\partial E}{\partial \theta_k^2} = \frac{\partial E}{\partial z_k}\frac{\partial z_k}{\partial \theta_k^2} = (y_k - z_k) g'(net_k) = \delta_k^2 \quad (13)$$

For the neuronal activation threshold of the hidden layer θ_j^1, the partial differential representation of the error value E is calculated as shown in Equation (14).

$$\frac{\partial E}{\partial \theta_j^1} = \sum_{k=1}^{n} \frac{\partial E}{\partial z_k}\frac{\partial z_k}{\partial O_j}\frac{\partial O_j}{\partial \theta_j^1} = \sum_{k=1}^{n} (y_k - z_k) g'(net_k) \omega_{kj}^2 f'(net_j) = \delta_j^1 \quad (14)$$

Thus, based on Equations (13) and (14), the modulation formulae for the neuronal activation thresholds θ_j^1 and θ_k^2 can be derived and expressed as shown in Equation (15).

$$\begin{cases} \theta_j^1(t+1) = \theta_j^1(t) + \Delta\theta_j^1 = \theta_j^1(t) + \eta^1 \dfrac{\partial E}{\partial \theta_j^1} = \theta_j^1(t) + \eta^1 \delta_j^1 \\ \theta_k^2(t+1) = \theta_k^2(t) + \Delta\theta_k^2 = \theta_k^2(t) + \eta^2 \dfrac{\partial E}{\partial \theta_k^2} = \theta_k^2(t) + \eta^2 \delta_k^2 \end{cases} \quad (15)$$

C. Establishment of an Innovation Model for PSO-based Optimization Neural Network Firms

BP neural networks can better reflect the complex non-linear relationships in the model, thus improving the prediction accuracy. However, BP neural networks also have their limitations. Firstly, BP neural networks are basically in a "black box" state during the solution process, and their results are difficult to be understood, and secondly, the neural networks have the phenomenon of overfitting. Based on big data technology, a sound knowledge management system can be established to collect and collate innovation indicators of enterprises and build an innovation model to measure the innovation capability of enterprises. Machine learning techniques, especially neural networks, can be effective in building an EnIn model, but it is difficult to find the optimal parameters due to the large number of parameters in a neural network model. For this reason, particle swarm algorithms (PSO) can be used to optimise neural network models to obtain the optimal parameters. Therefore, a corporate innovation model based on PSO optimised neural networks can be effectively constructed to measure the innovation capability of a company. The particle swarm algorithm model can be used to efficiently search for optimal parameters and to obtain a more accurate model of corporate innovation.

The PSO is an evolutionary algorithm, which is a technique based on group intelligence and on the simulation of group forces. The basic idea of the PSO algorithm is that each particle has a current position and an optimal position, and the current position of the particle is influenced by two forces, one from the particle itself and the other from the global optimal position. The current position of the particle is influenced by both forces, while the force of the global optimum position is influenced by the global optimum position, which is the influence of the optimum positions of all the previous particles. The optimization steps of the PSO algorithm are: first, initialize the particle swarm, each particle has a position and a velocity; then, calculate the fitness function of each particle; then, calculate the optimum position of each particle, the global optimum position. Finally, the particle positions and velocities are updated and terminated when convergence to the optimal solution is achieved. The particles have both velocity and position properties and are adapted using independent search and position sharing, with the update rules shown in Equations (16) and (17).

$$v_i = v_i + c_1 rand(pbest_i - x_i) + c_2 rand(gbest_i - x_i) \quad (16)$$

$$x_i = x_i + v_i \quad (17)$$

In Equations (16) and (17), v_i represents the particle velocity, $rand$ represents the random number, x_i represents the current particle position, c_1 and c_2 represent the learning factors. The PSO is a commonly used parameter search method, which can find the best parameters in the training set, and it can cut down the optimization time and strengthen the accuracy of the search with fewer iterations compared with the traditional grid search method [19]. The steps of its optimisation algorithm are shown in Fig. 5.

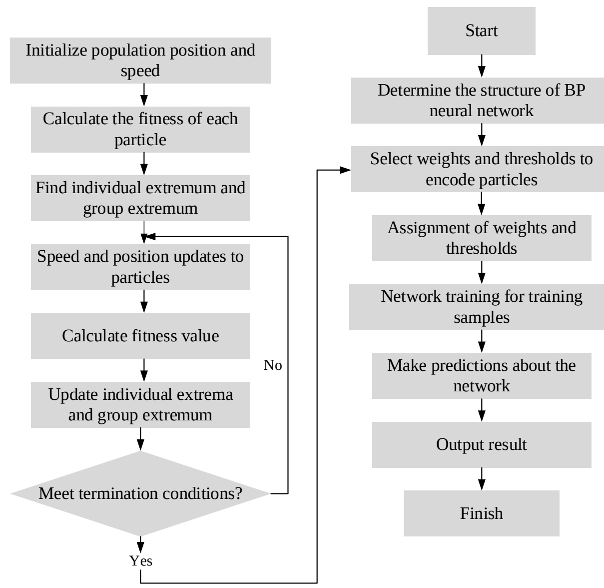


Fig. 5. Flow chart of PSO optimizing BP neural network model.

The node numbers in the input layer are first considered, as the system is based on the sample data of the enterprise big data knowledge evaluation index, and the node numbers in the input layer is decided by the amount of features extracted through PCA transformation [20]. The number of nodes in the output layer is decided by the size of the network connection power, i.e. the number of input nodes, output nodes and implied nodes, for both "yes" and "no" judgments. A particle is randomly generated with a position at (0,1) and the dimension of the velocity vector. The starting position of the particle is assumed to be pbest, and the optimal value of pbest is gbest. For each particle, if the exactness requirement is satisfied or the full evolution has reached the max iteration number (set to 2000), the algorithm ends and the current optimal individual in the full population is recorded, otherwise return to step 5.

IV. Analysis of the Effect of the Application of Machine Learning-based Enterprise Innovation Development Assessment Model in the Big Data Environment

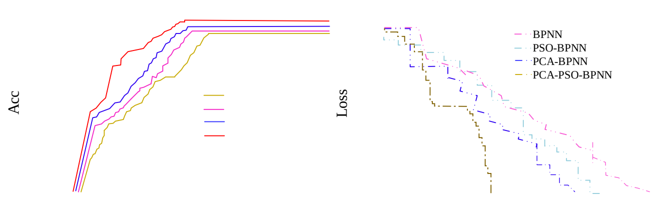
The study carried out a PCA using Matlab software-based on the above model and its theory. The data used for the study was derived from references, using a PCA of the 14 main use assessment indicator factors to derive the contribution of each major component, as shown in Fig. 6.

Fig. 6. PCA contribution rate statistics chart.

As can be clearly seen from Fig. 6, the cumulative contribution of the 20 indices when the number of principal components is 12 is a cumulative 89.365%. This method reduces the input nodes of the neural network from 14 to 12 and simplifies the input indices, causing a significant reduction in the size of the neural network and its reduction of 30%.

Fig. 7. Colour temperature diagram for the analysis of various factors related to the development of corporate innovation.

From Fig.7, we can see that among the 12 enterprise knowledge management and research innovation assessment indicators, the correlation coefficients are mostly around 0.5, for example, the correlation coefficients of X2, X8 are 0.519 and 0.553 in order, which have higher correlation, indicating that the previous PCA is effective, and the factors with poor correlation have been eliminated. The factors with high correlation are used as input data matrices, which lays the foundation for the next action of introducing the PSO-BPNN model.



(b) Comparison of loss values under different models of corporate credit risk

Fig. 8. Model training set accuracy curve and loss value curve.

The prediction error curves for the training samples are given in Fig. 8(b), with the number of training steps on the horizontal axis and the results of the errors on the vertical axis. Through 15 epochs of learning and training, the error results of the model proposed in the study were obtained close to 0. The other three lines in Fig. 8 are the model combining BP neural network and particle swarm algorithm, the model combining PCA method and BP neural network, and the model combining PCA method, particle swarm algorithm and BP neural network. Comparison shows that the PCA-PSO-BPNN has the highest training efficiency, reaching the target expectation after 15 steps. The training accuracy plots of the four models are given in Fig. 8(a), which clearly shows that the PCA-PSO-BPNN achieves 98% accuracy after 20 steps, which is significantly better than the other three models, so the model proposed in the study has good performance.

As can be seen in Table II, when the four algorithms were tested on the test and training sets, the RMSE values of the PCA-PSO-BPNN algorithm model for both the test and training sets were much lower than those of the other three algorithms, as can be seen from the enterprise R&D intensity dataset which has an RMSE value of 3.56 for the training set and 3.18 for the test set, and a value of 0.9982 for the R2 training set and 0.9915, which is higher than the other three algorithms. Meanwhile, from the data of corporate patent licensing, the RMSE values of the PCA-PSO-BPNN algorithm for the test set and training set are much lower than the other three algorithms, with 3.09 for the test set, 3.18 for the training set, 0.9451 for the R2 training set, and 0.9569 for the test set, which are also higher than the other three algorithm models. This shows that the PCA-PSO-BPNN algorithm model has better test results, higher test accuracy and more stable model.

The relative error histogram in Fig. 9 clearly displays that the absolute value of the relative error of each indicator does not exceed the 5% threshold, and only the relative errors of individual indicators have relatively large problems, such as X5 and X7 with the values of 4.5% and -3.8%. But the relative errors of the rest of indicators remain at correspondingly low levels. Better predictive capability for corporate innovation R&D and has excellent practicality. Finally, the study compared the training accuracy analysis of the PCA-PSO-BPNN algorithm with the conventional BP algorithm, PCA-BPNN and PSO-BPNN algorithms under different data set capacity sizes, and the results are shown in Fig. 10.

TABLE II. RMSE AND R2 FOR FOUR MODELS ON X1 AND X2 DATASETS

Datasets	**Algorithm**	**Training**		**Testing**	
		RMSE	R^2	RMSE	R^2
A_1	BPNN	30.85	0.8545	32.11	0.8351
	PSO-BPNN	25.66	0.8965	26.78	0.8647
	PCA-BPNN	15.98	0.9014	16.85	0.8994
	PCA-PSO-BPNN	3.56	0.9982	3.18	0.9915
A_2	BPNN	31.49	0.8214	30.98	0.8211
	PSO-BPNN	26.77	0.8851	27.54	0.8913
	PCA-BPNN	16.56	0.9052	16.99	0.9115
	PCA-PSO-BPNN	3.18	0.9451	3.09	0.9569

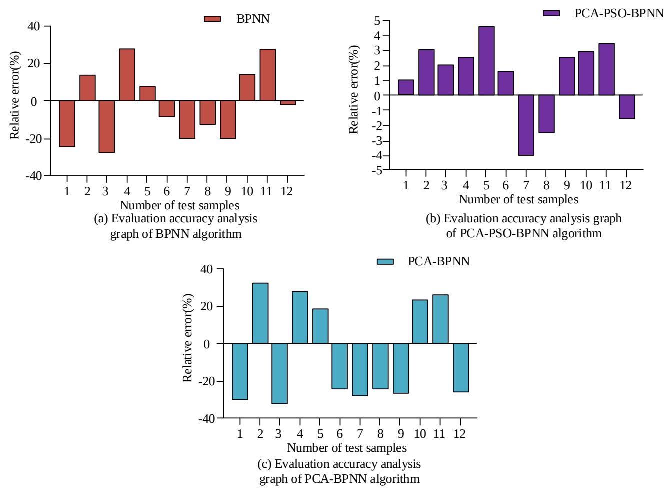


Fig. 9. Relative error in predicting corporate knowledge management and corporate innovation indicators based on the PCA-PSO-BPNN.

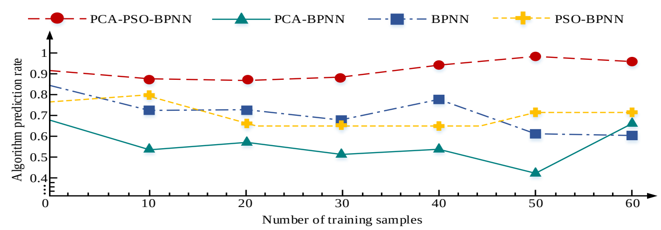


Fig. 10. Comparison of recognition accuracy of different algorithms trained on different datasets.

From Fig. 10, the study of the PCA-BPNN algorithm optimised by the PSO has the highest prediction accuracy among the four algorithms that conducted the experiment, and its average recognition rate can reach 94.21%; the recognition rate of the traditional BP algorithm model is slightly lower than that of the PCA-PSO-BPNN algorithm model, with an average recognition rate of 74.52%; the PSO-BPNN model The average recognition rate was 72.74%; the PCA-BPNN model had the lowest average recognition rate of 53.1%. The experimental results verify that the PCA-BPNN model optimised by the particle swarm algorithm is the best among the four models.

V. Conclusion

With the development of Internet technology and big data technology, knowledge management has become one of the important factors for enterprise innovation and development. In this study, we first introduce the improved PSO-optimized BP neural network (PSO-BP) for evaluating the degree of the role of big data knowledge management in EnIn development, based on which PCA is conducted to reduce the dimensionality of the model and the weights of PSO-optimized BP neural network are used. The test results verified that the cumulative contribution value of the 14 indicators reached 89.365% cumulatively when the master-formation score was 12, which simplified the input indicators and made the neural network much smaller, and its reduction reached 30%. Secondly, the color temperature plot of Spearman rank correlation test shows that its correlation is high among 12 enterprise knowledge management and innovation evaluation indexes, which indicates that the previous PCA is effective, and the factors with poor correlation are eliminated and entered into the established model as input data. It can be concluded that the model is effective with RMSE and R2 values of 3.56 and 0.9982 for the X1 training set and 3.18 and 0.9915 for the X2 test set, respectively. The results show that PSO-BP neural network can effectively reduce the dimensionality of the input indicators, simplify the neural network, and achieve a significant reduction in size. The high correlation between 12 corporate knowledge management and innovation assessment indicators indicates the importance of these factors in promoting corporate innovation. Although this study has achieved considerable results but there are still many problems, firstly, the study for enterprise big data knowledge management only evaluates enterprise innovation development without considering other factors and the feasibility of the algorithm will be considered subsequently for market conditions, regulatory environment and other factors. In addition, for this data study only one set of data was used for testing, and more data will be used to test the generality and accuracy of the algorithm in the follow-up.

Acknowledgment

The research is supported by Chengdu Sino-Thai Vocational Education Research Center, Research Platform of Chengdu Polytechnic (No. 21kypt02).

References

[1] Webster, K. N. (2023). Low-rank kernel approximation of Lyapunov functions using neural networks. Journal of Computational Dynamics, 10(1), 152-174.

[2] Pavl, S. D., Craus, M. (2023). Reaction-diffusion model applied to enhancing U-Net accuracy for semantic image segmentation. Discrete and Continuous Dynamical Systems - S, 16(1), 54-74.

[3] Maraqa, M. R., Omari, G., Jarrah, M. A. (2021). The impact of knowledge management infrastructure on the innovation process and products. The mediating role of Management Science Letters, 261-270.

[4] Peng, L., Li, Z. (2021). Psychological contract, organizational commitment, and knowledge innovation: A perspective of open innovation. Problems and Perspectives in Management, 19(2), 418-432.

[5] Yuen, Y. Y., Xiang, P. N. (2021). Enhancing innovation performance of small and medium enterprises in Malaysia. Management Science Letters, 887-894.

[6] Goncharenko, I., Krakhmalova, N. (2021). Functional support of the organizational and economic mechanism management of innovation and integration potential enterprises. Management, 32(2), 119-127.

[7] Zadorozhnyi, Z. M., Ometsinska, I., Muravskyi, V. (2021). Determinants of firm's innovation: increasing the transparency of financial statements. Marketing and Management of Innovations, 5(2), 74-86.

[8] Katsarski, N. (2021). Integrated business management in relation to water ecosystem. Knowledge International Journal, 38.3(2020), 595-599.

[9] Vasyltsiv, T., Levytska, O., Lupak, R., Gudzovata, O., Mulska, O. (2021). Understanding creative, information and knowledge determinants of the economic growth. Management Science Letters, 1295-1308.

[10] Straková, J., Váchal, J., Kollmann, J., Talí, M. (2021). Development trends in organizational and management structures. Problems and Perspectives in Management, 19(2), 495-506.

[11] Kompaniets, O. R. (2022). Sustainable competitive advantages for a nascent wine country: an example from southern Sweden. Competitiveness Review: An International Business Journal, 32(3), 376-390.

[12] Chatterjee, S., Chaudhuri, R., Vrontis, D. (2022). Knowledge sharing in international markets for product and process innovation: moderating role of firm's. International Marketing Review, 39(3), 706-733.

[13] Lopes, F. J., Abreu, V., Kumasaka, R. S., Rosini, A. M. (2021). Competency management, knowledge management and corporative education: a study on brazilian companies. Journal on Innovation and Sustainability RISUS, 11(4), 147-158.

[14] Babgohari, A. Z., Mokhtarzadedeh, N. G., Jafarpanah, I. (2022). Knowledge management capability, entrepreneurial creativity, entrepreneurial intensity and firm performance: the mediating role of ambidexterity. British Food Journal, 124(7), 2179-2208.

[15] Wang, Z., Ling, K. C., Li, H. G. (2021). The Impact of Knowledge Sharing on the Relationship Between Market Orientation and Service Innovation. Journal of Knowledge Management, 17(2), 130-154.

[16] Takehara, H., Suto, M. (2022). Impact of corporate social responsibility intensity on firm-specific risk and innovation: evidence from Japan. Social Responsibility Journal, 18(3), 484-500.

[17] Udagawa, H., Okano, T., Saito, T. (2023). Permutation binary neural networks: Analysis of periodic orbits and its applications. Discrete and Continuous Dynamical Systems - B, 28(1), 748- 764.

[18] Yang, N., Cheng, Y., Qu, H. (2023). Quantitative Determination of Mannitol in Cordyceps Sinensis Using Near Infrared Spectroscopy and Artificial Neural Networks. Chinese Journal of Analytical Chemistry, 31(6), 664-668.

[19] Ponta, L., Puliga, G., Manzini, R. (2021). A measure of innovation performance: the Innovation Patent Index. Management Decision, 59(13), 73-98.

[20] Birkhoff, D. C, Dalen, A., Schijven, M. P. (2021). A Review on the Current Applications of Artificial Intelligence in the Operating Room. Surgical Innovation, 28(5), 611-619.