# Research on the design and implementation of economic data analysis system based on big data background

ZiXin Yin*

GuanghuaQidi, Shanghai, 200082, China

Corresponding author Email 13764100936@163.com

**Abstract.** This paper mainly introduces the process of analyzing regional economy with academic data. First, we set the perspective of academic data to analyze the regional economy, and selected the scholar activity and the diversity of scholar cooperation to analyze the economic situation; at the same time, we constructed the indicators of the scholar activity and the diversity of scholar cooperation. Secondly, we analyzed the correlation between academic data and regional economic conditions from the set research perspectives by using Pearson's coefficient and Spearman's coefficient; moreover, we constructed a network of collaborators for the relationship between scholars' collaborations. This paper is based on the new big data processing technology, under the support of the big data background of the economic data analysis of the combination of research, under the support of certain technologies, combined with the technology of neural networks on the economic research and design, in the fusion of the research of the aspects of a certain breakthrough.

**Keywords:** Big data; economic data analysis; system implementation; network neural

## 1 Introductory

According to the definition of KPIs of major economic indicators, based on the full amount of economic data obtained, through economic statistics, economic modeling and mining analysis of economic big data, it obtains the dynamic trend of various indicators from massive information, effectively analyzes the relationship between various influential elements of the associated economy; forms a big data analysis model of economic resources (including land, labor, energy, capital, science and technology, etc.); accurately controls the input and output of the economy; forecasts and warns the trend of key indicators in the economy, such as investment, environmental damage, excessive consumption, etc., to effectively prevent systemic risks [1]. It can form a big data analysis model of economic resources (including land, labor, energy, capital, science and technology, etc.), accurately control the input and output of the economy, and enhance the comprehensive international competitiveness; and forecast and warn the trend of key indicators in the economic field, such as overheating of investment, environmental damage, and overconsumption, so as to effectively guard against systemic risks [2] [3].



Y. Jiao et al. (eds.), *Proceedings of the 3rd International Conference on Internet Finance and Digital Economy (ICIFDE 2023)*, Atlantis Highlights in Economics, Business and Management 1,
https://doi.org/10.2991/978-94-6463-270-5_41

# 2 Correlation analysis

## 2.1 Relationship between scholarly activity and regional economic conditions

The Pearson coefficients and Spearman coefficients between the number of papers and economic indicators are shown in Table 1, where the highest value in each row is emphasized in bold. Based on the results of the 10-year correlation analysis, it can be found that the number of papers published has a highly positive linear correlation with GDP [5] [6], a relatively strong positive linear correlation with the proportion of tertiary industry, a relatively weak positive linear correlation with GDP per capita, and no obvious linear correlation with the proportion of primary and secondary industries [4].

**Table 1.** Pearson coefficient correlation between the total number of published papers and the regional economic index by region over the years

| | Pearson Correlations | | | | |
|---|---|---|---|---|---|
| | GDP-NP | AGDP-NP | Primary-NP | Secondary-NP | Thirdy-NP |
| 2008 | **.753**** | .324** | -.248** | -.077 | 283** |
| 2009 | **.742**** | .319** | -.246** | -.091 | .397** |
| 2010 | **.757**** | .330** | -.251** | -.121* | .411** |
| 2011 | **.758**** | .299** | -.252** | -.130* | .420** |
| 2012 | **.748**** | .294** | -.252** | -.158** | .433** |
| 2013 | **.749**** | .283** | -.245** | -.162** | .426** |
| 2014 | **.745**** | .302** | -.251** | -.184** | .436** |
| 2015 | **.756**** | .315** | -.249** | -.186** | .434** |
| 2016 | **.755**** | .342** | -.249** | -.187** | .441** |
| 2017 | **.767**** | .366** | .010 | -.196** | .007 |

There is a positive correlation between the number of papers and the tertiary sector, and the results are in line with the development of the real society [7].

# 3 Theoretical Foundations of Deep Learning

## 3.1 Forward algorithm.

The neural network structure consists of an input layer, a hidden layer and an output layer, whose basic unit is the neuron [8]. The neural network currently used is mainly

a feed-forward neural network, which means that the output of the neurons in the previous layer is used as the input of the neurons in the next layer. At the same time, it contains multiple hidden layers, and the more layers there are, the more expressive the model is and the higher the complexity [9]. The number of hidden layers of the neural network determines the size of the depth of the network, and the layers are fully connected to each other, and the structure of the neural network is shown in Figure 1.

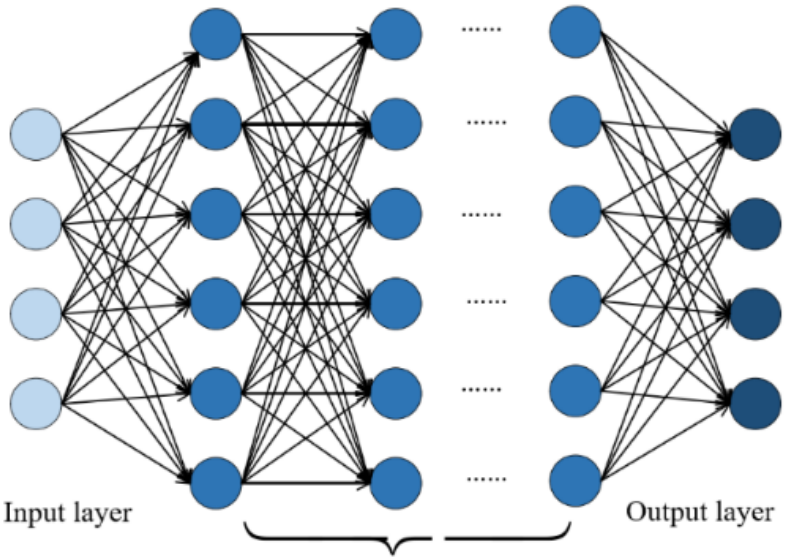


**Fig. 1.** Neural network architecture

In this case, the nonlinear mapping between neighboring layers of neurons is achieved through activation functions, and the neuron structure is shown in Figure 2.

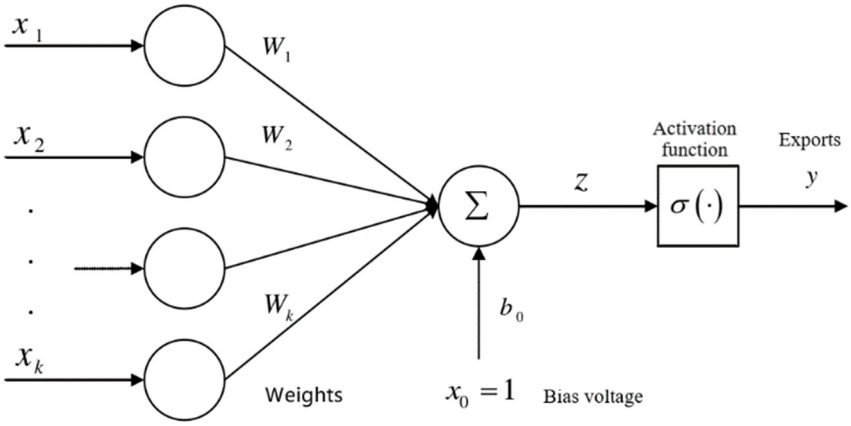


**Fig. 2.** Neuronal structure

The input features are weighted and summed into the neuron, and then the activation function is used to obtain the neuron output value:

$$\zeta = \sum w_i * x_i + b$$
$$x_{i+1} = \sigma(\zeta)$$

Where $w_i$ denotes the weight between neurons in layer i and neurons in layer $i+1$, $x_i$ denotes the output of layer i, b denotes the bias, $\zeta$ denotes the input of layer, $x_{i+1}$ denotes the output of layer i+1, and $\sigma$ denotes the activation function. The estimate of the output layer of the neural network is obtained by forward recursion.

### 3.2 Inversion algorithm

In order to make the output estimate of the training samples after passing through the neural network as close as possible to the desired value, a suitable weight matrix and bias are very important. Here, a suitable loss function J can be used to achieve convergence to constrain the output error value.The common iterative methods used for loss function optimization problems are gradient descent, Newton's method and simulated Newton's method. The output error is derived for the ith layer neural network parameter $w_i$, and the backward recursive derivation formula is simplified as follows:

$$\frac{\partial(J)}{\partial(w_i)} = \frac{\partial(J)}{\partial(w_{i+1})} * \frac{\partial(y_{i+1})}{\partial(\sigma_{i+1})} * \frac{\partial(\sigma_{i+1})}{\partial(w_i)}$$

Then update the weights and bias of the neural network:

$$w^{t+1} = w^t + \eta * \Delta w^{tl}$$

$$b^{t+1} = b^t + \eta * \Delta b^{tl}$$

Where $\eta, t, l$ denotes the learning rate of the neural network, the number of iterations and the number of neural network layers, respectively.

### 3.3 Activation function.

The Sigmoid activation function expression is:

$$\sigma(\zeta) = \frac{1}{1 + e^{-z}}$$

From Figure 3, it can be seen that as z takes larger and larger values, the function curve flattens and $\sigma(\zeta)$ it's getting smaller and smaller. Similarly, as $\zeta$ takes smaller and smaller values, $\sigma(\zeta)$ becomes smaller and smaller. In the inverse algorithm, the forward recursion is multiplied by $\sigma(\zeta)$ to obtain the changing value of the gradient, so

the sigmoid function causes $w$ and $b$ to be updated slowly. If placed in the middle layer of the neural network it may create a gradient vanishing problem, so it is generally placed in the final layer of the neural network.

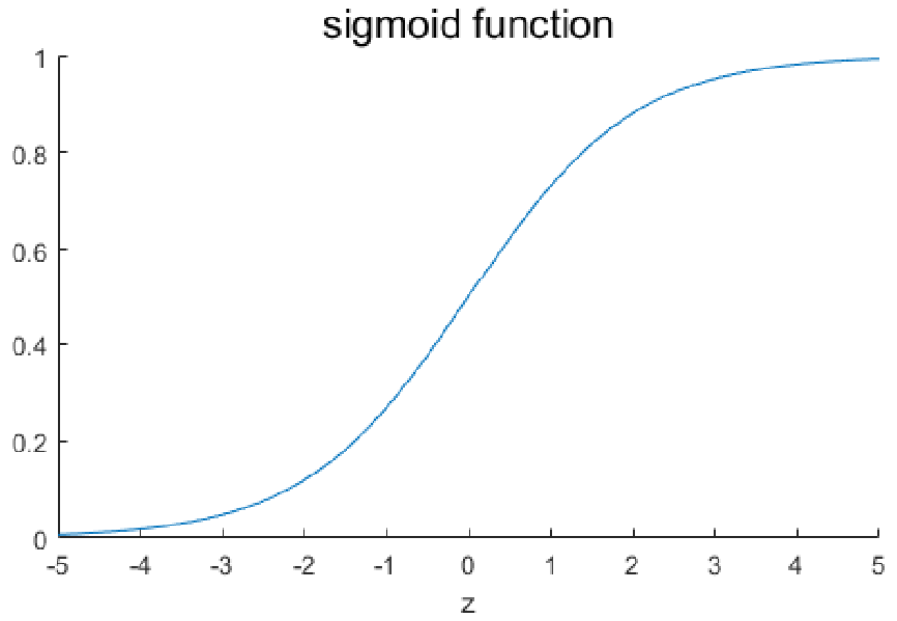


**Fig. 3.** sigmoid function

In addition, gradient vanishing or gradient explosion may also occur in the back-propagation algorithm for deep networks, and gradient explosion generally occurs when the initialization weights are too large, which can be solved by optimizing the initialization parameters of the network model [10]. Gradient disappearance is often due to the use of an inappropriate loss function, and the relu activation function can alleviate the problem to a certain extent, it is always 1 in the positive half-axis $\sigma(\zeta)$, is not easy to be affected, but in the negative half-axis $\sigma(\zeta)$ is always 0, which may lead to neuron inactivation, Leaky ReLU can solve the problem of neuron "death". The Relu activation function is commonly used in the intermediate hidden layer and its expression is:

$$\sigma(\zeta) = \max(0, \zeta)$$

## 4 Conclusion

This paper researches and analyzes the economic development based on the supportive background of big data, and analyzes and explores the economic development through a multi-dimensional study. In the research background of big data to repair the

economic development was integrated and cross-study. Academic data can give new insights into the state of macroeconomic development from an academic point of view and provide a new direction of thinking in observing the state of economic development.

## Reference


1. Qin Xiaopeng. Analysis of digital economy development based on big data background[J]. Modern Business,2023(07):47-50. DOI:10.14097/j.cnki.5392/2023.07.018.
2. Feng Yu. Analysis of the role of internal audit in enterprise economic management under the background of big data[J]. Old Brand Marketing,2022(24):100-102.
3. Xue Ru. Design of economic analysis system under the background of big data[J]. Wireless Internet Technology,2022,19(20):68-70.
4. YU Qiuju,SHEN Yuanyuan. Empirical analysis of China's digital economy and e-commerce based on big data[J]. Journal of Shangqiu Normal College,2022,38(06):16-19.
5. Yu Xulei. Analysis of Macroeconomic Statistics Problems and Development Suggestions in the Context of Big Data[J]. Business Culture,2022(12):38-40.
6. Jue Zhao. Design and implementation of regional economic census data visualization and analysis system [D]. University of Electronic Science and Technology,2015.
7. Xu Aiting. A statistical study on the scale of informal sector production and its impact [M]. Zhejiang Gongshang University Press:, 201402.251.
8. Aftab U A,Wen-Hui Y,En-Rong Y. Correlations between leaf economics, hydraulic, and shade-tolerance traits among co-occurring individual trees[J]. Acta Oecologica,2021,110.
9. Sébastien J,Mario L,Valérie M. Editorial, International Economics[J]. International Economics,2021,165(prepublish).
10. J. D W,James A,Janice A, et al. Impact of tariff structure on the economics of behind-the-meter solar microgrids[J]. Cleaner Engineering and Technology,2021,2.