Research on short-term forecasting Technology of Local economy based on Deep Learning

Zhongyao Wang[1], Dianya Pan[2], Jia Yang[3]

[1]Logistics School, Beijing Wuzi University, Beijing, 101100

[2]School of Management, Jilin University, Changchun, Jilin, 130000

[3]Xihua University, Chengdu, Sichuan, 610039

Keywords: Deep Learning; Economic Forecast; Deep confidence Network; Gross domestic Product

Abstract: The standard uses deep confidence network deep learning algorithm to study the short-term prediction of local economy. Firstly, the strong correlation index of local economic prediction and analysis is extracted, and the index data are obtained quickly and effectively through the Internet by crawler program. With the help of Python platform, the index is output automatically, and the output data are normalized and vectorized. Taking the local GDP as the prediction object, the short-term forecast of local economy is completed through DBN deep learning algorithm. In the process of DBN training, the weight parameters are cross-adjusted forward and backward, and the optimal parameters are obtained to determine the DBN structure. Taking Chongqing as an example, the simulation results show that the short-term prediction accuracy of local economy based on big data drive and DBN deep learning is high and the MSE is small. The experimental results show that the DBN deep learning algorithm has higher accuracy and smaller MSE in predicting local GDP compared with common economic prediction algorithms. The following research will further study DBN network scale, learning rate and other aspects, in order to improve the time performance of local economic short-term prediction.

1. Introduction

With the continuous development of big data technology, it is possible to apply the powerful analytical capability of big data technology to local economic statistics and analysis. Economic indicators of complex and difficult to statistics, by relying on nothing more than a year the government public basic statistical data for analysis of inefficient [1], because the government's statistics have certain hysteresis, moreover the government's statistics have much depend on artificial statistics and report statistics, data integrity and accuracy is poor. Economic development is vulnerable to the changes of social environment or the influence of uncertain factors, and manual statistical data alone cannot analyze the transient impact of micro-environment on local economy.

Therefore, local economic analysis based on big data-driven and big data technology brings new opportunities for local economic forecasting and analysis. The significance of short-term local economic forecast is to provide scientific support for policy makers' near-term policy planning. Big data drives provide data support for local economic prediction and analysis [2], and data analysis tools or methods are also needed for detailed analysis of big data. At present, using the algorithm of machine learning or deep learning has made some achievements in the field of economic forecasting, literature [3] and [4] are realized by using neural network algorithm for local economic forecast, the former is based on the least absolute shrinkage and selection algorithm optimization algorithm, which combines wavelet theory to optimize, has achieved good effect. However, the accuracy of the above method in practical application is not ideal. The key content of short-term economic forecasting is GDP forecasting, which directly affects the overall economic development trend and is also a key indicator of economic forecasting. Based on the big data environment, this paper adopts the Deep-belief-network (DBN) algorithm to complete the short-term GDP forecast of local economy, so as to further improve the accuracy of local economy forecast.

2. Acquisition of local economic indicators

GDP is selected as the object of economic forecast in order to reflect the economic development situation most directly. In order to establish the short-term prediction model of local economy, it is necessary to extract the categories and related indexes that affect local economy, comprehensively evaluate the factors that affect local economy, and try to extract the indexes that are highly correlated with local economy.

In this paper, a total of 15 predictive indicators from 5 categories are selected, as shown in Table 1. Economic and environmental indicators, import and export indicators and consumer indicators can be obtained through public announcement data and yearbooks on the government website. Investment and resource indicators use crawlers to dynamically obtain the latest data through the Internet [5]. On Python and other platforms, the Numpy toolkit is used to output the index table automatically, and the open source Tushare financial data API is called for data collection. After selecting the local economic forecast control sample, select the local economic short-term forecasting index, normalize all the forecasting indicators, and then generate the labeled eigenvector combined with the local GDP gross domestic product, and finally get the prediction results through algorithm learning.

3. Analysis of DBN algorithm

Set the visible layer of DBN as v=(v1,v2...,vm), the hidden layer as h=(h1,h2...,hn), and (v,h)∈ {0,1} m+n. A stable Restricted boltzmann machine (RBM) is determined by the energy parameter {w,c,b}. W is the weight between the two layers, and c and b represent their respective offset.

The probability of the visible layer relative to a single hidden neuron is:

$$P(v,h) = \frac{1}{Z}\mathrm{e}^{-E(v,h)} \tag{1}$$

It conforms to Bernoulli distribution, and is expressed as:

$$E(v,h) = -\sum_{j=1}^{m} b_j v_j - \sum_{i=1}^{n} c_i h_i - \sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij} v_j h_i \tag{2}$$

In the equation, P(v,h) only one neuron is considered for the solution, so the probability of n hidden neurons acting on the visible layer is:

$$P(v) = \sum_{h} P(v,h) = \frac{1}{Z}\sum_{h}\mathrm{e}^{-E(v,h)} \tag{3}$$

Probability of m neurons acting on the hidden layer in the visible layer:

$$P(h) = \sum_{v} P(v,h) = \frac{1}{Z}\sum_{v}\mathrm{e}^{-E(v,h)} \tag{4}$$

The probability that the i hidden neuron will be triggered is targeted at M visible neurons:

$$P(h_i = 1 \mid v) = \sigma(c_i + \sum_{j=1}^{m} w_{ij} v_j) \tag{5}$$

For n hidden neurons, the probability of the J visible unit being triggered is:

$$P(v_j = 1 \mid h) = \sigma(b_j + \sum_{j=1}^{n} w_{ji} v_i) \tag{6}$$

Where, machine probability function:

$$\sigma(x) = \mathrm{sigmoid}(x) = \frac{1}{1+\mathrm{e}^{-x}} \tag{7}$$

For N input samples,v =(v_0,v_1... v_N), and v_0,v_1...v_N is independently and identically distributed.

$$P(v) = \prod_{t=0}^{N} P(v_t) \tag{8}$$

The likelihood estimate of sample set V is denoted as

$$L(\theta) = \prod_{t=0}^{N} P(v_t \mid \theta) \tag{9}$$

Can be converted to solve for the maximum value of lnL

$$\hat{\theta} = \mathrm{argmax}_{\theta} L(\theta) = \mathrm{argmax}_{\theta} \sum_{t=0}^{N} \ln P(v_t \mid \theta) \tag{10}$$

$$\theta^* = \theta + \eta \frac{\partial \ln P(v)}{\partial \theta} \tag{11}$$

For a single sample V0 =(v01,v02... v0m), logarithmic solution is performed for the sample , and:

$$\ln P(v_0) = \ln \frac{1}{Z} \sum_h \mathrm{e}^{-E(v_0,h)} = \ln \sum_h \mathrm{e}^{-E(v_0,h)} - \ln \sum_{v,h} \mathrm{e}^{-E(v,h)} \tag{12}$$

$$\frac{\partial \ln P(v_0)}{\partial \theta} = -\sum_h P(h \mid v_0) \frac{\partial E(v_0,h)}{\partial \theta} + \sum_{v,h} P(v,h) \frac{\partial E(v,h)}{\partial \theta} \tag{13}$$

Conditional probability meets:

$$P(v,h) = P(h \mid v) P(v) \tag{14}$$

4. The instance simulation

In order to verify the performance of DBN deep learning algorithm for short-term prediction of local economy in big data environment, Matlab was used for example simulation. Taking Chongqing city as an example, the yearbook was combined to obtain a total of 10 years of economic data from 2009 to 2018, and the remaining index data were obtained through a crawler. The obtained data were processed by Python, normalized and vectorized to generate data samples, which were divided into 10 groups according to the year.

4.1. The effect of DBN network size on the performance of short-term economic forecasting

15 economic indicators were selected as data samples, local GDP values were selected as prediction objects, and THE GDP value after 2018 was predicted by DBN deep learning algorithm. By taking 7 sets of economic data samples from 2009 to 2016 as input, the GDP values of 2017, 2018 and 2019 were respectively predicted and compared with the actual GDP values of the 3 annual GDP values of 19,500.27 billion yuan, 2,036.319 billion yuan and 2,360.577 billion yuan. The threshold of forecast accuracy was set as 90%.

4.1.1 Hide the impact of layers on economic forecasts

The influence of hidden layers on Meansquareerror (MSE) and training time is shown in figure 1. With the increase of the number of L_h hiding layers, the economic forecast MSE decreases. L_h number decreases from 1 to 2, and MSE rapidly decreases from more than 0.3 to less than 0.1, with a large decrease. In the interval of MSE decreases slowly. When Lh exceeds 4, MSE tends to be stable. When DBN scale is 1 and 2, although MSE is large, it has poor fitness for economic prediction, so too small DBN scale is not suitable for short-term prediction of local economy.

With the increase of L_h, the training time increases, especially when L_h, is above 4, the training time increases rapidly. When L_h, is 6, the training time reaches 1300s;While L_h, is in the interval [1,4], the training time changes little. The reason is that THE increase of L_h, increases the time for pre-training, but the time for reverse fine-tuning the weight decreases. The combination of the two causes a small change in the training time. Moreover, the training time did not increase significantly, so there was a balance between MSE and forecast time. When L_h was set as 4, it could better meet the demand of local economic forecast.

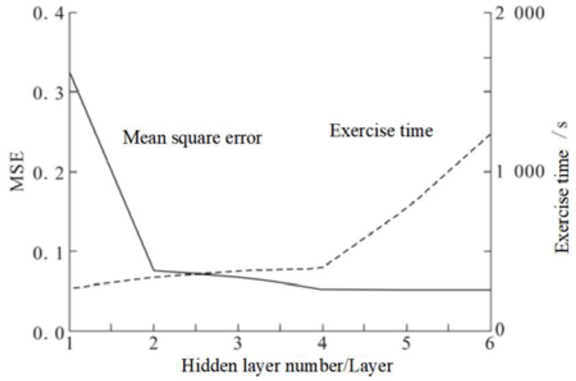

Fig.1 MSE and training time of different hidden layers

4.1.2 The influence of the number of hidden layer neurons on DBN prediction performance

The number of neurons in the hidden layer was differentiated to set Nh, and the prediction accuracy threshold was also set at 90% to verify the sensitivity of MSE to Nh. The Nh of each layer was initially set as [2,3,4,5], with a total of 4 layers . The remaining 3 groups of samples were used to test the model, and the actual values obtained in Chongqing, China statistical yearbook, were compared to solve the mean square error under different size structures. The results are shown in Table 2.When Nh changes, the mean square error of the three groups of samples changes little. The MEAN square error does not always decrease with the increase of the amount of neuron data in the hidden layer. When Nh is 16, the mean square error predicted by the three groups of samples is the lowest, and when the number of neurons is 20, the prediction error increases. Therefore, setting Nh equal to 16 is more suitable for the economic forecast of Chongqing.

4.2. The effect of DBN network size on the performance of short-term economic forecasting

In order to further verify the algorithm in the local economy short-term forecast to measure the performance, this paper adopts the algorithm and other four commonly used economic prediction algorithms to compare and analyze the total economic GDP in 2017, 2018 and 2019.DBN network Nh is 16, Lh is 4 layers. The average prediction accuracy and average MSE of the three years were calculated by simulation, and the results were shown in Table 1.

Tab.1 Prediction mean square error of different network scales

Group number	Number of neurons hidden layer	MSE		
		Max	Min	Average
1	8	0.0837	0.0707	0.0727
	12	0.0611	0.0511	0.0527
	16	0.0563	0.0475	0.0501
	20	0.0579	0.0517	0.0554
2	8	0.0789	0.0716	0.0729
	12	0.0631	0.0561	0.0595
	16	0.0579	0.0532	0.0555
	20	0.0668	0.0611	0.0629
3	8	0.0797	0.0707	0.0747
	12	0.0611	0.0549	0.0583
	16	0.0542	0.0499	0.0524
	20	0.0641	0.0588	0.0611

Tab.2 Economic prediction accuracy and MSE of different algorithms

Algorithm	Accuracy rate	MSE
PLSA algorithm	73.291	0.2111
Decision tree algorithm	72.237	0.1892

Logical regression algorithm	80.795	0.1233
Neural network algorithm	87.121	0.1247
Algorithm of this paper	94.355	0.0523

This algorithm in the local economy short-term prediction accuracy obvious advantages, reached 94.355%, low prediction accuracy of SVR and decision tree algorithm, neural network and logistic regression algorithm accuracy, although more than 80%, but these two kinds of algorithm of MSE than the 0.1, shows that these two algorithms for short-term forecasting performance is not stable, local economy MSE algorithm in this paper is 0.0523, the optimal performance.

5. Conclusion

Under big data environment convenient access to local economic data samples combined with DBN depth learning algorithm to achieve short term forecast for local economy. Local GDP is used as predictive object and local economy short term forecast is completed by DBN depth learning algorithm. During DBN training process, we adjust weight parameter forward and backward cross adjustment, obtain optimal parameter and determine DBN structure. Experimental results show that compared with common economic prediction algorithms DBN depth learning algorithm has higher prediction accuracy than usual algorithm and MSE is smaller. Further research will further research from DBN network scale learning rate etc. so as to improve the time performance of local economy short term forecast. It can also change annual forecast to quarterly or monthly forecast further refine local economy short term forecast granularity improve local economy short term forecast applicability under big data drive.

References

[1] Zhang Tao, LIU Kuanbin. Application of "Big Data" in macroeconomic Forecasting and Analysis [J]. Abstract of Social Sciences, 2018 (8) : 47-49

[2] Niu Wensheng, Dong Xiaobo, Gao Jiali.Big Data Application in Macroeconomic Analysis [J]. Economic Forum, 2015, 575 (6):94-96.

[3] RENUKADEVI T,KARUNAKARAN S. Optimizing deep belief network parameters using grasshopper algorithm for liver disease classification[J]. International Journal of Imaging Systems and Technology,2020,30 (1):168-184

[4] CHEN Y ,ZHAO X ,JIA X . Spectral-spatial classification of hyperspectral data based on deep belief network [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2015,8 (6):2381-2392.

[5] SHAO H, JIANG H, ZHANG H, et al. Rolling bearing fault feature learning using improved convolutional deep belief network with compressed sensing[J] . Mechanical Systems and Signal Processing, 2018,100 (1): 743-765.