Prediction of the real estate industry economics based on LSTM model

Taoen Zhang*, Baojun Fang
United World College of Changshu China, Changshu, China

*Corresponding author: tezhang18@uwchina.org

Abstract. The economics of real estate industry has been seriously affected by the 2019-NcoV. Therefore, this study analyzed the changes of real estate industry and LSTM model was applied to predict the price of real estate industry. The error analysis was done to test the accuracy of model and the result indicated the model established in this paper has high accuracy and meaningful for practical application.

1. Introduction
Historical experience proves that the impact of the epidemic on the macro economy is short-term. In the early stage of the epidemic, due to the epidemic and market sentiment, the economic growth rate will drop sharply for a short time. However, with the gradual subside of the epidemic and the gradual stabilization of the market, coupled with reasonable improvement policies, the pent-up demand of consumption and investment will be released, bringing about a bottoming out of the economy. Therefore, the economic analysis part of this paper has the following important practical significance for China's economic recovery and revitalization. The real estate sector has been greatly affected by the epidemic, resulting in a sharp rise in unemployment and fluctuations in housing prices. That's despite an official survey showing China's unemployment rate at 6.0% in April. But more research suggests that unemployment in China has actually surged to around 20% between outbreaks [1]. The huge pressure of unemployment has seriously affected the living standards of unemployed families (and low-income families), and affected the possibility of economic recovery and social stability. The Suggestions on the improvement of the real estate industry in this paper will ensure the normal life of working families, improve labor productivity and social stability, and guarantee the economic revival of China

2. Methodology
LSTM (Long Short-Term Memory) RNN neural network, is a special RNN variant, it can learn long-term dependent information. The LSTM is deliberately designed to avoid long-term dependency problems [2]. All RNNS have a chained form of repeating neural network modules, as shown in Figure 2. In standard RNN, this repeating module has only a very simple structure, such as a tanh layer.
Figure 1. Structure of RNN

The LSTM is the same structure, but the repeating modules have a different structure. Unlike a single neural network layer, here are four small devices that interact in very specific ways \[3\]. Each black line in Figure 2 transmits an entire vector from the output of one node to the input of another. The pink circle represents the pointwise operation, such as the sum of vectors, while the yellow matrix represents the learned neural network layer. The key to the LSTM is the Cell, the horizontal line running through the Cell. Cells are like a conveyor belt. It runs directly along the chain with only a few linear interactions. Information is easy to keep constant on a horizontal line.

Figure 2. Structure of LSTM

The LSTM removes or adds information to the Cell by carefully designing gate structures. A door is a method (filter) for information selection to pass through. They consist of a Sigmoid neural network layer and a Pointwise multiplication operation. The Sigmoid layer prints values between 0 and 1, describing how much of each part can pass. Zero means no quantity allowed and one means any quantity allowed.

3. Computational procedure of LSTM model
The LSTM uses two gates to control the contents of the unit state Cell. One is the Forget Gate, which determines how much of the unit state of the last moment \( c_{t-1} \) is retained until the current moment. The other is the Input Gate, which determines how much of the network's input \( x_t \) is currently stored in the
unit state $c_t$. The LSTM uses the output gate to control the current output value $c_t$ of the unit state $h_t$ to the LSTM.

### 3.1. Forget gate

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (1)$$

Where $W_f$ is the weight matrix of oblivion gate, $[h_{t-1}, x_t]$ means to connect two vectors into a longer vector, $b_f$ is the bias term of oblivion gate, and $\sigma$ is the Sigmoid function. If the dimension of the input is $d_x$, the dimension of the hidden layer is $d_h$, and the dimension of the unit state is (usually $d_c = d_h$), then the dimension of the weight matrix of the forgotten gate $W_f$ is $d_c \times (d_h + d_x)$. In fact, the weight matrix $W_f$ is a Mosaic of two matrices: one is $W_{fh}$, it corresponds to the input item $h_{t-1}$, and its dimension is $d_c \times d_h$; One is $W_{fi}$, it corresponds to the input term $x_t$, and the dimension is $d_c \times d_x$.

In summary, the function of the forget gate is to control how much information from the memory cell of the previous moment can be accumulated into the memory cell of the current moment [4]. Its mathematical formula can be written as Equation (2), and it’s shown in Figure 3.

$$f_t = \text{sigmoid}(W_f x_t + W_f h_{t-1} + b_t) \quad (2)$$

![Figure 3. Forget gate](image)

### 3.2. Input gate

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (3)$$

Where $W_i$ is the weight matrix of the input gate, and $b_i$ is the bias term of the input gate. Figure 4 shows the calculation of the input gate [5]:

---

[1]: https://example.com
[2]: https://example.com
[3]: https://example.com
[4]: https://example.com
[5]: https://example.com
The next step is to calculate the state of the cell used to describe the current input \( \tilde{c}_t \), which is calculated based on the last output and this input:

\[
\tilde{c}_t = \tanh(W_c[h_{t-1}, x_t] + b_c)
\]

(4)

Now calculate the state of the cell \( c_t \) at the current moment. It is produced by multiplying the previous unit state \( c_{t-1} \) by the element \( f \) and the current input unit state \( \tilde{c}_t \) by the element by the input gate \( i \), and then adding them:

\[
c_t = f_t \circ c_{t-1} + i_t \circ \tilde{c}_t
\]

(5)
In this way, we combine the LSTM’s current and long-term memories to form a new unit state. Because of the control of the forget gate, it can store information from a long time ago, and thanks to the control of the input gate, it can keep irrelevant information from entering the memory.

3.3. **Output gate**

\[
o_t = \sigma(W_o[h_{t-1}, x_t] + b_o)
\]

(6)

The final output of the LSTM is determined by the output gate and the unit state.

\[
h_t = o_t \cdot \tanh(c_t)
\]

(7)
4. Application in prediction of real estate industry

4.1. Data processing
The LSTM neural network was trained by using the data of 40 months from January, 2017 to April, 2020 in real estate. The data of the first 32 months was the training sample, and the data of the last 8 months was the test data. And the data were normalized by Z-scores method [6], which is shown in Equation (8):

$$Z = \frac{x - \mu}{\sigma}$$ (8)

Where $\mu$ is the mean and $\sigma$ is the variance.

4.2. Solution and Results analysis
In order to get high precision of the neural network model, the latency period according $t$ in this paper, the neural network structure and the number of hidden layer nodes $n$ the adjustable parameter test, a total of 9 kinds of structures, shorthand for $(t_i, x_i)$, each run 30 times for each of the selected parameters, record every time error, and the minimum error of the network model parameters as the group's final model. The following Table I shows the training results of each network structure.

| STRUCTURES | (5,20) | (6,20) | (7,20) | (5,30) | (6,30) |
|------------|--------|--------|--------|--------|--------|
| ERROR      | 15.42% | 14.87% | 7.43%  | 18.77% | 18.19% |

It can be concluded from the above table that the number of delay periods of the neural network structure is 7, and the model error is the smallest when the number of hidden layer nodes is 20, only 7.43%. Therefore, the LSTM RNN network structure established in this paper is shown in Figure 8.

![Neural network in MATLAB](image)

Figure 8. Neural network in MATLAB

The error of model can be calculated by Equation (9);

$$\varepsilon = \frac{\sum_{i=1}^{8} |y'_i - y_i|}{\sum_{i=1}^{8} y_i} \times 100\%$$ (9)

Figure 9 compares the predicted and the real value, it can be seen that the error is quite small.
4.3. Prediction of the real estate industry in future

The LSTM RNN model and historical data trained above are used to predict the real estate economic situation in the remaining 8 months of 2020. The results are shown in Table II:

| Month | Prediction value |
|-------|------------------|
| 5     | 12318.51         |
| 6     | 16729.77         |
| 7     | 13225.07         |
| 8     | 18991.82         |
| 9     | 3901.918         |
| 10    | 6694.613         |
| 11    | 8017.34          |
| 12    | 7531.329         |

As can be seen from the prediction result, according to the forecast model, the sales volume of real estate in the latter months of 2020 fluctuates greatly, first increasing, then decreasing significantly, and finally recovering slightly.

5. Conclusions

LSTM model is employed in this study to predict the real estate industry, and high accuracy model was obtained to forecast the value of future months in 2020. The model established in this study is meaningful.

References

[1] A, Ibrahim Khalifa , et al. Polyacylated anthocyanins constructively network with catalytic dyad residues of 3CL pro of 2019-nCoV than monomeric anthocyanins: A structural-relationship activity study with 10 anthocyanins using in-silico approaches. Journal of Molecular Graphics and Modelling 100(2020).
[2] Zhang, Songyang, et al. Fusing Geometric Features for Skeleton-Based Action Recognition using Multilayer LSTM Networks. IEEE Transactions on Multimedia (2018):1-1.

[3] Peddinti, Vijayaditya, et al. Low Latency Acoustic Modeling Using Temporal Convolution and LSTMs. IEEE Signal Processing Letters 25.99(2018):373-377.

[4] Ghaeini, Reza, et al. DR-BiLSTM: Dependent Reading Bidirectional LSTM for Natural Language Inference. (2018).

[5] Yildirim, and zal. A novel wavelet sequences based on deep bidirectional LSTM network model for ECG signal classification. "Computers in Biology & Medicine (2018):189.

[6] Hancock, Arthur A., et al. "Data normalization before statistical analysis: keeping the horse before the cart." Trends in Pharmacological sciences 9.1(1988):29-32.