



Predicting housing price in China based on long short-term memory incorporating modified genetic algorithm

Rui Liu^{1,2} · Lu Liu²

Published online: 19 February 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

Predicting the future trend and fluctuation of housing price is an important research problem of housing market. The machine learning approach is rarely used in existing studies, while the traditional prediction models have strict requirements on input variables and are weak in solving nonlinear problem. To overcome the problems of traditional models, a long short-term memory (LSTM) approach is proposed to predict the housing price of a city by using historical data. The proposed LSTM incorporates a modified genetic algorithm with multi-level probability crossover to select appropriate features and the optimal hyper-parameters. The data of housing price and related features of Shenzhen, China, from year 2010 to 2017 have been used to test the performance of the model. The results indicate that the proposed method has good performance in modeling housing price and is obviously outperforms other algorithms including back propagation neural network, support vector regression and different evolution LTSM. Therefore, this proposed model can be used efficiently for predicting housing price and thus can be a good tool for policy makers and investors to monitor the housing market.

Keywords LSTM · Genetic algorithm · Housing price · Predict · Optimization · Multi-level probability

1 Introduction

One major research problem of housing market is how to predict the future trend and fluctuation of housing price (Ghysels et al. 2013; Guirguis et al. 2005). Only based on such knowledge, policymakers can react instantly and appropriately to reduce the boom and bust of housing market (Claessens 2015). In the past few decades, China has experienced a rapid growth of its housing market, and now, the sky-rocketing housing price in major cities of China has raised the concerns of both policymakers and residents (Jia et al. 2017).

Unlike western countries which have free market, China's housing market has several unique characteristics. First, the urban land market was established in 1988, when

the first land auction also sparked the light of housing market. Since then, the local government intervenes a lot in land leasing and makes it an important source for extra-budgetary revenue (Huang and Du 2017). Therefore, besides a trading product, land performs more as an important tool to stimulate local economics and contributes a lot to the high housing price. Second, the housing price keeps up growing in an extremely long period of time. The commodification of housing in China began in 1990s. In 1998, the central government abolished the welfare-oriented distribution of public housing, which stimulated the housing market for the first time. Since then, the housing in China keeps up with a soaring price (Fang et al. 2015; Wei et al. 2016). Meanwhile, strong optimistic exists in the housing market, because the housing price raised so much when compared to the housing income and price of other commodities. Wei et al. (2016) has indicated that the ownership of houses becomes the major measure of one's wealth. Meanwhile, since house is deemed as a necessity of marriage in China, it becomes a vital factor in the competition for marriage partners, which further stimulates the housing price to grow. Third, the government intervention is an important element in China's housing market. To cope with the booming market, the government has enacted

Communicated by V. Loia.

✉ Lu Liu
luliu.wh@gmail.com

¹ Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen 518000, China

² Center for Assessment and Development Research of Real Estate, Shenzhen 518000, China

various policies to keep the stability of the market (Jia et al. 2017). These policies include restricting home-purchase, tightening mortgage rates and rising property taxes. However, all these policies failed to keep the housing price from growing too fast in the long run. Zhou (2018) has proved that the high sentiment negatively impacts the effectiveness of tightening policies. Especially, even if the price is temporarily suppressed by the central policy, a following retaliatory price rebound would happen when related local policy is published.

Because of these special characteristics, predicting the trend and fluctuation of future housing price in China is needed, but extremely difficult. To find an appropriate solution, we first review several existing models of the housing market, as discussed below.

The past studies proved that, at least partially, the time path of housing price can be predicted. In the existing literatures, various methodologies have been adopted to forecast the future trend of housing price. Regression approach has been most usually used. For instance, Pain and Westaway (1997) have developed an error correction model to estimate the housing price in UK, and Malpezzi (1999) have also specified an error correction model for house prices. Crawford and Fratantoni (2003) have compared the performance of regime switching, autoregressive integrated moving average model (ARIMA) and generalized autoregressive conditional heteroskedastic (GARCH) model in predicting housing price; results show that ARIMA models perform better in out-of-sample forecasting. Guirguis et al. (2005) have put varying parameters to several of the existing autoregressive models and forecasted the housing price in US. These models, however, have strict requirements for the input datasets. If the sample size is small, or when the data are non-stationarity, the model cannot be correctly established.

Grey model contributes another solution in modeling housing price. Yang and Xing (2006) have utilized a Grey–Markov model to predict the housing price index, which has achieved a satisfying result. Wang et al. (2013) have used a gray system, which successfully predicts the slow-down rise of housing price index in China. However, though the construction of Grey model only needs a few samples, it is unable to simulate the complex up and downs of housing price; therefore, it can only be used to predict a monotonously increasing or decreasing trend (Wu and Chen 2005).

In recent years, the technique of machine learning is developing fast, because of its great advantage compared with traditional methods (Al-Janabi 2017, 2018; Al-Janabi

and Al-Shourbaji 2017; Al-Janabi et al. 2015), and various related methods have been used to model housing price. Because the impact factors of housing price have very high dimensions and are usually nonlinear, these machine learning methods are expected to achieve better results than the above traditional methods. The applications are two-fold. On the one hand, many studies focus on predicting or evaluating the single unit house in the city by using machine learning. For example, Park and Bae (2015) have compared various machine learning methods of predicting the housing price in Fairfax country, Virginia, and demonstrate that RIPPER algorithm outperforms other models. Selim (2009) has used artificial neural network (ANN) to examine the determinants of housing price in Turkey. Compared to the traditional hedonic-based regression model, these machine learning approaches have been proved to provide better results in accuracy (Lim-sombunchai et al. 2004). On the other hand, the macro-level housing price index can be also modeled and predicted by these methods. Wang et al. (2014) has adopted support vector machine (SVM) to forecast the housing price index in Chongqing, China, with using warm optimization (PSO) to determine the parameters; the results show that PSO-SVM has better performance compared to grid and genetic algorithm. Such approach, which includes optimizing methodology to do parameter optimization and feature selecting, can usually improve the performance of the original method and has been widely adopted in various applications (Abualigah and Khader 2017; Abualigah et al. 2017a, b, 2018a, b).

To sum up, traditional methodologies have several limitations relating to the fundamental model assumptions and estimations. Moreover, though in recent years, some attempts have been made to model housing market by machine learning methods, few studies have utilized and compared various machine learning methods in modeling housing price for a city or a country, especially for China, where the government behavior and financial policy have great impact on the housing market. To fill this research gap, we propose a model to predict the housing price index at city level in China based on long short-term memory (LSTM). This method has not yet been used in the previous studies, but is expected to achieve good results in forecasting housing price, because its advantage in predicting time series with long time lags between important events. Moreover, to achieve better results, a modified genetic algorithm with multi-level probability crossover is adopted to implement feature selection and optimize the hyper-parameters for the model. The real housing price data and

the related features of Shenzhen, China, from the year 2012 to the year 2017 are used to test the performance of the model. By comparing the results of the proposed model with BPNN, SVR and DELSTM, it is proved that the proposed LSTM approach achieves the best results, whose RMSE is 41, MAE is 40, and MAPE is 0.06.

2 Datasets

We choose the housing market in Shenzhen, China, as the object city. To predict the future trend and fluctuation of the newly built commercial housing in Shenzhen, eight features are selected based on two criteria. First, the features are proved to influence the future housing price. Second, the monthly time series data for the feature can be obtained from available data source.

These features can be divided into three dimensions. The first dimension is about the residential land. In China, the local governments intervenes the land market by monopolizing their right to supply land and lease them to developers (Huang and Du 2017). Therefore, the newly released residential land supply (NewhouseS) can indirectly affect the supply side of housing market. On the other hand, the floor area under construction (AreaCons) can reflect the developer’s response to the market driven by investment incentives and space demand driven by residents (Zhou 2018) and thus can also be an indicator of the housing price.

The second dimension contains several basic economic features. Price of the newly built commercial housing (PriceNewHouse) is our prediction target, and its historical time series can greatly affect the future housing price (Huang et al. 2008) and can thus be put into the model as a feature. Another indicator is Completed investment in Fixed Assets (CIFAsseets), which can reveal the attitude of investors, as well as the housing demand of common residents. Therefore, CIFAsseets works as an important indicator of housing market by showing a picture of the demand side. Moreover, the increase in Consumer Price Index (MIncreCPI) can influence the housing market in China, because when the CPI goes up fast, people are very prone to invest on houses to maintain the value of their asset (Wei et al. 2016).

The third dimension is about the states’ financial policies that would influence the housing market. One of the main policies concerning government intervention should be the adjustment of medium- and long-term loan interest (LInterestRate). A benign interest rate environment can

lead to the boom of housing market (Demary 2010), and a tight interest policy is usually taken to suppress the rocketing up price. Meanwhile, in China, all employees are required to contribute a proportion of their salaries to housing provident fund (HPF) (Yeung and Howes 2006), and as a return, the HPF loan for a house has lower interest rate than commercial loan. Therefore, the interest rate of HPF also affects the housing market and should be added to the model as a feature. Moreover, the monthly net increase in RMB loans (RMBloan) of the city can be a useful indicator, because it reflects people’s optimism toward the market. The fast increase in the RMB loans reveals that people are passionate in putting their money to the housing market.

We extracted 664 records for all these eight features, spanning period from December 2010 to October 2017. For each individual feature, 83 monthly records are obtained from various data source. The detailed information of these features is listed in Table 1.

3 Algorithm

3.1 LSTM

Long short-term memory (LSTM) is one of the recurrent neural networks (RNN), whose nodes contain self-loop connections (Evermann et al. 2017; Gensler et al. 2017; Hochreiter and Schmidhuber 1997).

Different from classical RNN, the LSTM introduces the concept of memory cell which contains several control structures of information flow: input gate, forget gate and output gate. The classical RNN has problems of vanishing/exploding gradient and long-term dependencies, while the LSTM can make up these disadvantages by using its special structure.

Given a multivariate time series $x = \{x_1, x_2, \dots, x_n\}$, where x_1 is a univariate time series $\{x_{1,1}, x_{1,2}, \dots, x_{n,T}\}$ and T is the length of x_1 . The LSTM block can be represented by Eqs. (1)–(6). Figure 1 shows the structure of a LSTM block.

$$i_t = \sigma(W_i x_t + V_i h_{t-1} + B_i) \tag{1}$$

$$f_t = \sigma(W_f x_t + V_f h_{t-1} + B_f) \tag{2}$$

$$o_t = \sigma(W_o x_t + V_o h_{t-1} + B_o) \tag{3}$$

$$m_t = \tanh(W_m x_t + V_m h_{t-1} + B_m) \tag{4}$$

$$c_t = i_t * m_t + f_t * c_{t-1} \tag{5}$$

Table 1 List of features for the experiment

Category	Name of features	Description	Data source
Residential Land	NewhouseS	Newly released residential land supply (Floor area)	Urban planning land and resource commission of Shenzhen Municipality
	AreaCons	Floor area under construction	Shenzhen Statistics Bureau
Economics	PriceNewHouse	Price Index for newly built commercial housing	State Statistics Bureau
	CIFAssets	Completed Investment in Fixed Assets	Shenzhen Statistics Bureau
	MIncrCPI	Month-on-Month increase on Consumer price index of Shenzhen	State Statistics Bureau
Financial	LInterestRate	Medium- and long-term loan interest rate (5 years or above)	The People’s Bank of China
	HPFInterstRate	Interest rate of loans for individual housing provident fund (5 years or above)	The People’s Bank of China
	RMBLoan	Monthly Net increase of RMB loan for Shenzhen	The People’s Bank of China

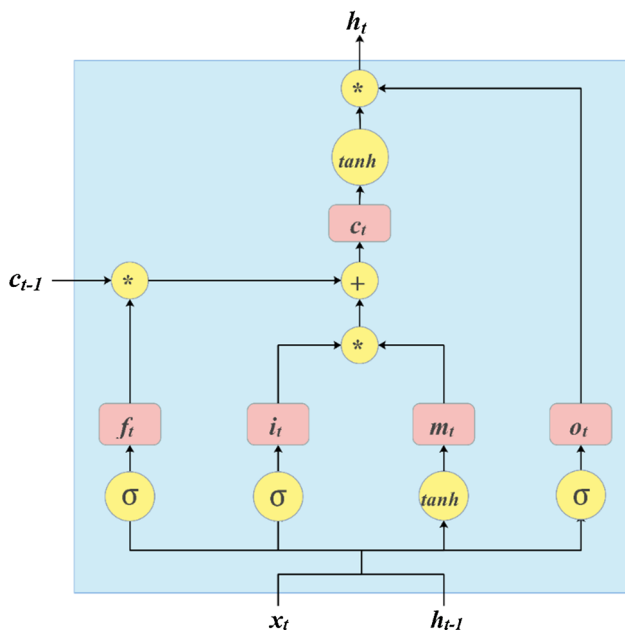


Fig. 1 A LSTM block

$$h_t = o_t * \tanh(c_t) \tag{6}$$

where x_t is a multidimensional input values at step t . $\sigma(\cdot)$ is the sigmoid function and $\tanh(\cdot)$ is the hyperbolic tangent function. i_t, f_t and o_t are the control structure of input gate, forget gate and output gate, respectively, at step t . c_t is the memory cell state, and h_t is the hidden state at step t . W and V are the weight parameters, and B is the bias parameters.

3.2 GA

GA is a metaheuristic that simulates species evolution strategy (Holland 1992), which is usually used to generate solutions in optimization and searching problems. It has simple operations and is very robust. In recent years, GA has been used widely in various applications such as production scheduling and controlling science. A classical GA mainly includes four steps, which are initialization, selection, crossover and mutation. Initialization generates some feasible solutions randomly targeting to an actual problem. Selection indicates the process of selecting individuals in the population based on its fitness (Goldberg 1989). Generally, individuals with higher fitness are more likely to be retained, while individuals with lower fitness are more likely to be eliminated. Crossover refers to exchanging part of two individuals to generate a new individual, which may significantly improve the searching performance. Mutation randomly rewrites part of the individual with some probability. It intends to increase the diversity of the population and thus to effectively reduce the probability of falling into a local optimum. Based on iterating these four steps, GA can help us to obtain the optimal result.

3.3 GA-LSTM

The hyper-parameters can greatly affect the performance of LSTM. The number of units in each hidden layer (NUHL) is an important hyper-parameter, and the determination of this NUHL is crucial in the whole process. Meanwhile,

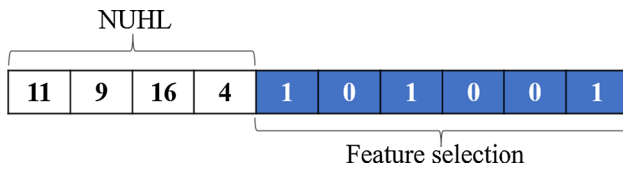


Fig. 2 An individual

feature selection can also greatly influence the performance of LSTM. Therefore, a modified GA with multi-level probability crossover is proposed to optimize the NUHL and do the feature selection. The GA-LSTM procedure is described as follows:

Step 1 Data preprocessing. The data set is first normalized to $[0, 1]$ by Eq. (7) and then divided into two subsets: the training set and the test set.

$$x = (x - x_{\min}) / (x_{\max} - x_{\min}) \tag{7}$$

Step 2 Initialization. The parameters of GA include population size; maximum generation, G_{\max} ; crossover factor, CF; level of number of hidden layers (NHL), C_h ; level of feature selection C_s ; mutation factors, MF. The parameters of LSTM include bounds of NUHL, training number, batch size and look back. An individual is generated based on these settings. As shown in Fig. 2, an individual is generated with 4 hidden layers and 6 features. The NUHL is generated randomly between the lower and upper bound of NUHL. The part of feature selection is randomly generated in $\{0, 1\}$, where 1 and 0 represents that the feature is selected and not selected, respectively. The initial population is generated accordingly.

Step 3 Selection. The fitness is calculated first. The selection is then done based on roulette wheel method.

Step 4 Crossover. First, a formula of multi-level probability is proposed to maintain the stability and enhance the diversity of the population, as shown in Eq. (8), where p presents for the initial probability. Based on this formula, two individuals are randomly selected. Several positions are then randomly selected for NUHL and feature selection separately. If reach the CF, exchanging is processed at these positions. This procedure is described in Algorithm 1. Moreover, Fig. 3 shows crossover of two individuals with $C_h = C_s = 1, 2, 3$.

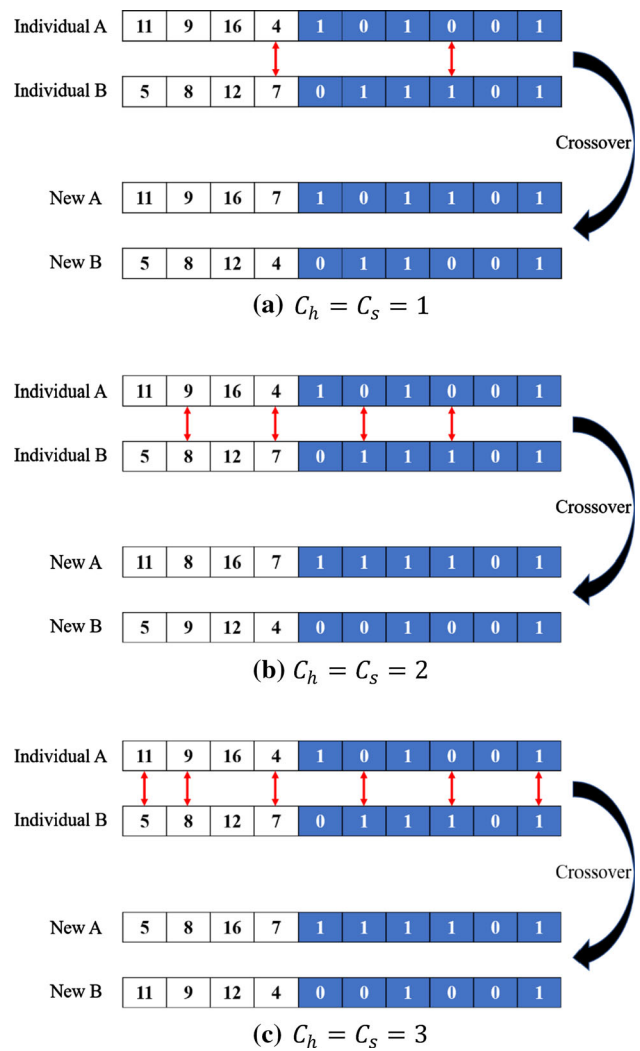


Fig. 3 Crossover

$$LP(l) = \begin{cases} p & l = 1 \\ (1 - p) \times (2/3) & l = 2 \\ (1 - p) \times (1/3) \times (2/3) & l = 3 \\ (1 - p) \times (1/3)^2 \times (2/3) & l = 4 \\ \dots & \dots \\ (1 - p) \times (1/3)^{n-3} \times (2/3) & l = n - 1 \\ (1 - p) \times (1/3)^{n-2} & l = n \end{cases} \tag{8}$$

Algorithm 1. Mechanism of crossover operation

1. Input initial probability p , level of number of NHL and feature selection: C_h and C_s ;
2. $r_h = rand(0,1)$;// A random float number between 0 and 1//
3. **For** $l_h = 1$ to C_h **do**
4. **If** $LP(l_h) \leq r_h \leq LP(l_h + 1)$ **then**
5. Number of NHL: crossover operation with l_h ;
6. **Break**
7. **Endif**
8. **Endfor**
9. $r_s = rand(0,1)$;// A random float number between 0 and 1//
10. **For** $l_s = 1$ to C_s **do**
11. **If** $LP(l_s) \leq r_s \leq LP(l_s + 1)$ **then**
12. Feature selection: crossover operation with l_s ;
13. **Break**
14. **Endif**
15. **Endfor**
16. **Return** new individuals

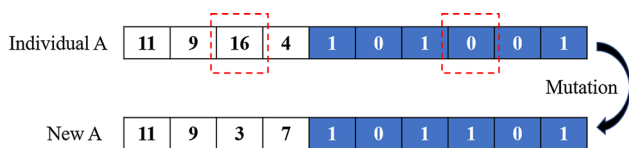


Fig. 4 Mutation

Step 5 Mutation. For an individual, a position is selected individually for NUHL and feature selection. Then values are generated randomly at these two positions if reach the MF. Figure 4 shows mutation operation of two individuals.

Step 6 Calculate fitness of the offspring population. If the iteration number reaches G_{max} , return the optimal individual; otherwise, $G = G + 1$, return to Step 3.

Step 7 The LSTM with the optimal individual is tested by the test data.

A flowchart showing GA-LSTM is shown in Fig. 5.

4 Experiment and discussion

To test the performance of the GA-LSTM in predicting housing price, results are also obtained by long short-term memory with differential evolution algorithm (DELSTM) (Peng et al. 2018), back propagation neural network (BPNN) and support vector regression (SVR) for compar-

ison. In this section, all algorithms are coded in Python (version 3.6). LSTM and BPNN are based on a Python deep learning library—Keras (version 2.2.2), and the backend is TensorFlow (version 1.11). Meanwhile, SVR is based on scikit-learn (version 0.19.1). All experiments are conducted on a personal computer with an Intel® Core i7-6700HQ, 2.6 GHz CPU, 8 GB RAM and Windows 10 Operational System.

4.1 Parameters setting

The values of corresponding parameters have a significant influence on the performance of GA-LSTM, DELSTM, BPNN and SVR. For the proposed GA-LSTM, the parameters of GA are set as follows: probability of crossover is 0.9, probability of mutation is 0.2, p is 0.6, level of number of NHL $C_h \leq \min\{\text{number of NHL}, 3\}$, feature selection $C_s \geq \min\{\text{number of features}, 3\}$, population size is 20, and iteration number is 10. Meanwhile, the parameters of LSTM are set as: the bound of NUHL is [5, 20], the training number is 100, the batch size is 5, and the number of hidden layer (NHL) is {1, 2,3, 4, 5, 10, 15, 20}. For the DELSTM. For BPNN, nine combinations: the NUHL is {5, 10, 20}, the training number is {250, 500, 1000}, are tested. For SVR, the parameters are default parameters in scikit-learn. In addition, the look back is set to 1 for the above algorithms.

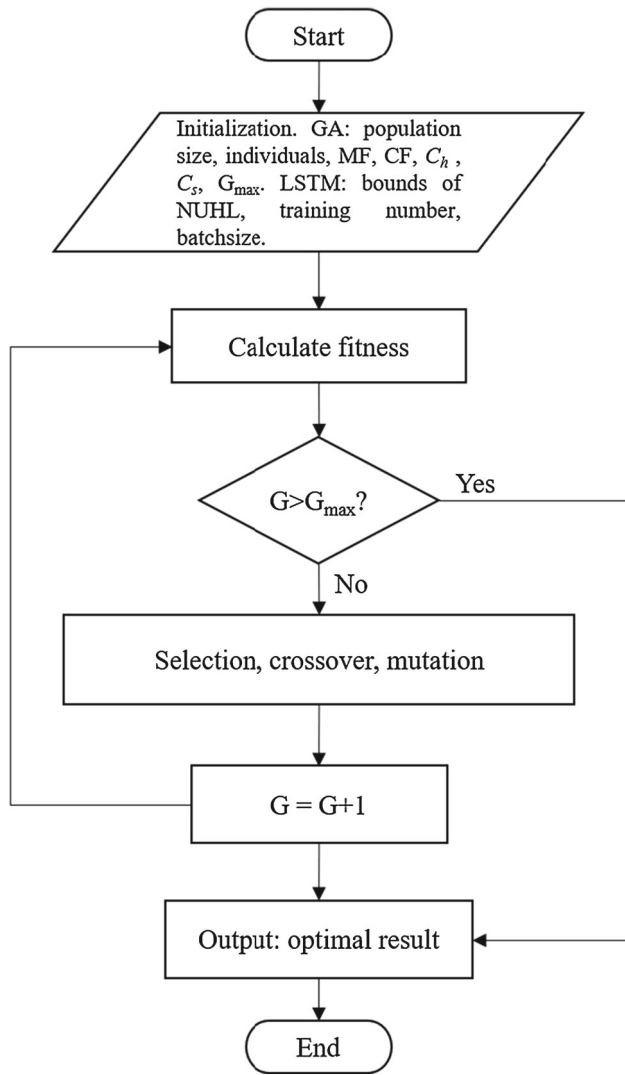


Fig. 5 The flowchart of GA-LSTM

4.2 Results of the experiment

In experiments, the first 90% of the dataset is set as the training data and the rest of the dataset is set as the test data. Each feature in the dataset is normalized by Eq. (7). Meanwhile, root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE) are adopted to evaluate the performance of the algorithms in experiments.

$$RMSE = \sqrt{\sum_{t=1}^T (\bar{y}_t - y_t)^2 / T} \tag{9}$$

$$MAE = \sqrt{\sum_{t=1}^T |\bar{y}_t - y_t| / T} \tag{10}$$

Table 2 Results of the GA-LSTM, DELSTM, BPNN and SVR

Algorithm	NHL	RMSE	MAE	MAPE (%)
GA-LSTM	1	959	855	1.57
	2	151	140	0.23
	3	41	40	0.06
	4	287	254	0.49
	5	377	323	0.66
	10	663	623	1.21
	15	27,340	27,334	50.12
20	26,869	26,870	49.25	
DELSTM		80	63	0.12
BPNN		1702	1648	2.42
SVR		1818	1284	2.35

Table 3 Solutions of the GA-LSTM with different NHL

Algorithm	NHL	NUHL	Feature Selection
GA-LSTM	1	[15]	[1, 1, 0, 0, 0, 0, 1, 0]
	2	[5, 16]	[1, 0, 1, 0, 0, 0, 0, 0]
	3	[14, 10, 13]	[1, 0, 0, 0, 0, 0, 0, 0]
	4	[14, 17, 13, 16]	[1, 0, 1, 0, 0, 0, 0, 0]
	5	[18, 5, 19, 12, 15]	[1, 0, 1, 0, 0, 0, 1, 0]
	10	[9, 12, 18, 11, 8, 10, 18, 19, 17, 18]	[1, 0, 0, 0, 0, 0, 0, 0, 0]
	15	[7, 17, 18, 15, 13, 7, 12, 10, 8, 17, 10, 18, 8, 8, 6]	[1, 1, 1, 1, 1, 1, 1, 1]
	20	[10, 10, 13, 10, 5, 9, 15, 7, 8, 10, 13, 19, 17, 11, 8, 11, 6, 5, 19, 5]	[1, 0, 0, 1, 1, 1, 1, 1]

$$MAPE = \sqrt{\sum_{t=1}^T |\bar{y}_t - y_t| / y_t / T} \tag{11}$$

T represents for the length of test data; \bar{y}_t and y_t refer to the forecasting value and real value of the test data, respectively.

The results of the GA-LSTM, DELSTM, BPNN and SVR are presented as follows in Tables 2, 3, 4 and Fig. 3. To be noticed that, we choose the best result of BPNN by testing various combination of parameters, as discussed in Sect. 4.1.

By scrutinizing the result, the following conclusions can be obtained:

Table 4 Results of the proposed GA-LSTM and the basic GA-LSTM

Algorithm	NHL	RMSE	MAE	MAPE (%)
GA-LSTM	1	959	855	1.57
	2	140	140	0.23
	3	41	40	0.06
	4	287	254	0.49
	5	377	323	0.66
	10	663	623	1.21
	15	27,340	27,334	50.12
	20	26,869	26,870	49.25
BGA-LSTM	1	408	397	0.66
	2	433	401	0.79
	3	198	173	0.35
	4	1199	1108	1.95
	5	729	704	1.32
	10	820	780	1.48
	15	26,877	26,871	49.27
	20	25,800	25,789	47.29

BGA-LSTM is the basic GA-LSTM with single-point crossover

- (a) As shown in Table 2 and Fig. 6, these four machine learning methods can successfully establish housing price models with acceptable results. However, the GA-LSTM performs better than DELSTM, BPNN and SVR. For the RMSE, the best result of GA-LSTM is 41, while the result of DELSTM, BPNN and SVR is 80, 1702 and 1818, respectively. Meanwhile, for the MAPE, the best result of GA-LSTM is 0.06%, while for BPNN and SVR the best results are 2.42% and 2.35%, respectively. For the MAE, GA-LSTM is also best. Moreover, when the NHL is less than 10, the results of GA-LSTM are always better than both BPNN and SVR.
- (b) With increasing NHL, the performance of GA-LSTM is first getting better and then getting worse. In the numerical examples, the results of GA-LSTM are the best when the NHL equals 3 by examining RMSE, MAE and MAPE.
- (c) In GA-LSTM, only a few features are appropriately selected from the original eight, to achieve the best result. As shown in Table 3, no more than 3 features are selected when NHL is no more than 10.
- (d) The proposed GA-LSTM performs better than the basic GA-LSTM with single-point crossover. As shown in Table 4, the MAPE of BGA-LSTM is much larger than that of GA-LSTM.

4.3 Discussion

The above results indicate that the proposed GA-LSTM approach can successfully predict the housing price of a city in China. Compared to the traditional methods, this approach has several advantages. First, it can achieve a good result because of a better feature selection process. It is known that the housing price can be affected by many factors, and the establishment of traditional models is usually bothered by the selection of variables, since these variables are usually complex, inconsistent through time and are not integrated from one another. On the comparison, the proposed method can automatically and dynamically select appropriate features by adopting a modified GA, with no need to consider the problems traditional models always need to face. Moreover, by only adopting a few number of features (only eight in our study), and with limited samples, a satisfying result can be achieved.

However, the proposed model has its limitations. First, the GA-LSTM approach can achieve a good result, but is very time-consuming. Therefore, the efficiency of the model needs to be improved. Second, when the dataset is small, the performance of the model is likely to be weakened. Third, the housing price is modeled only for one city in this study, and more cities should be included to test the availability of this model. Fourth, in this study, only eight features concerning residential land, housing economics and loan interest are considered in the model. In China, policy is such an important factor of the housing market, and a better result can be expected when more policy-related features are put into the model.

5 Conclusion

In this study, LSTM incorporating a modified GA is proposed for predicting the future trend and fluctuation of housing price of cities in China. Eight features that may influence the housing market are considered and have been used for training our model. Because China's housing market has many unique characters and is largely affected by the policy, the housing price in China is extremely hard to be accurately modeled and predicted. However, the results in this manuscript indicate that machine learning methods have good performance in modeling housing price of a city, even with limited features and data. Particularly, the proposed GA-LSTM obviously outperforms DELSTM, BPNN, SVR and basic GA-LSTM. Therefore, this GA-LSTM can be used as an efficient tool to assist policy makers as well as investors in monitoring and forecasting the dynamics of the housing market.

In the future study, a better housing price model can be possibly obtained through two ways. First, the policies

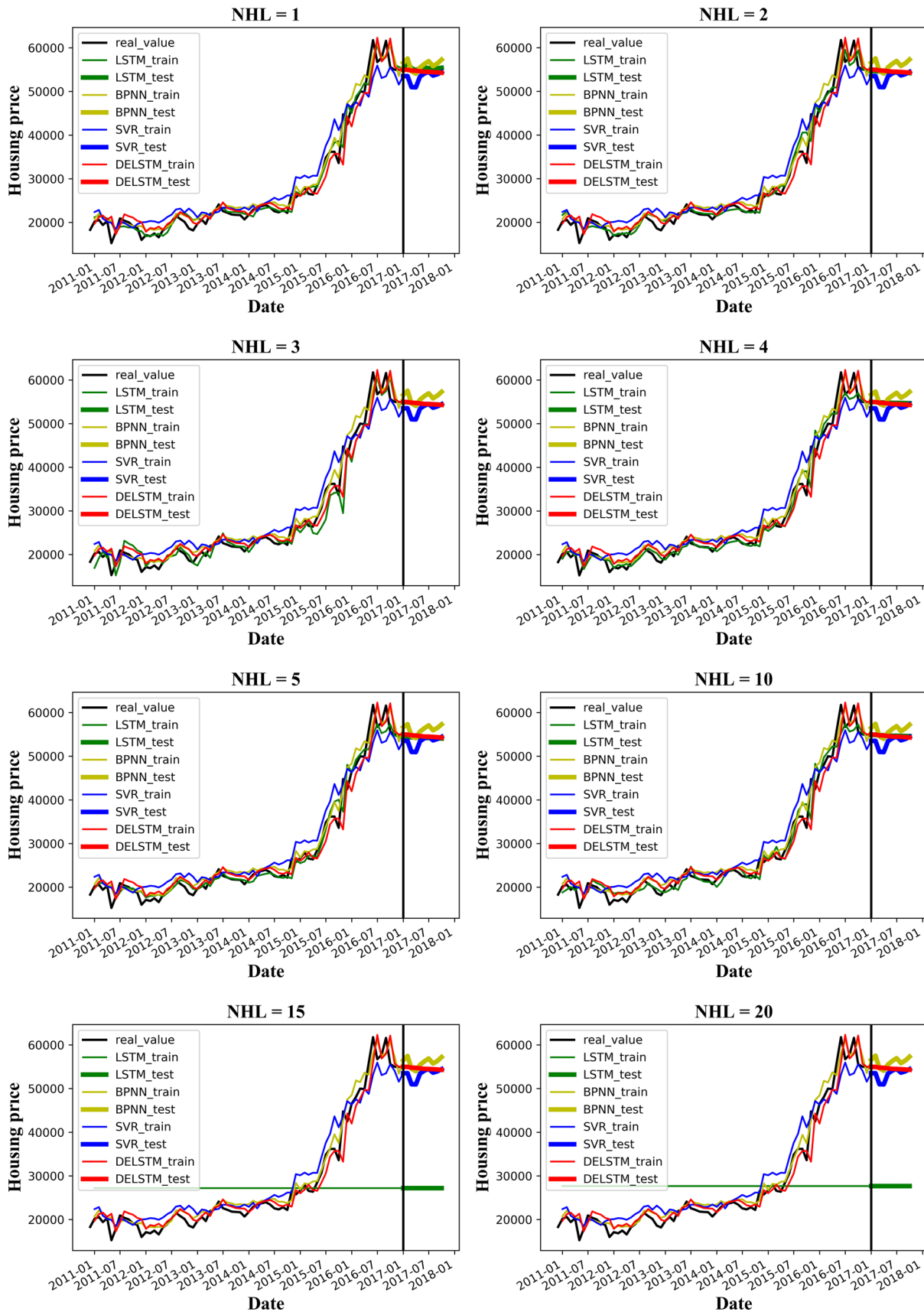


Fig. 6 Fitting curves of the housing price with different NHL

concerning the housing price can be classified and quantified to construct new features for the model. Second, the hyper-parameters can be further optimized by trying various algorithms.

Acknowledgements This work was supported by the Key Laboratory of Urban Land Resources Monitoring and Simulation Ministry of Land and Resource of China (Grant No. KF-2018-03-022).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

- Abualigah LM, Khader AT (2017) Unsupervised text feature selection technique based on hybrid particle swarm optimization algorithm with genetic operators for the text clustering. *J Supercomput* 73(11):4773–4795
- Abualigah LM, Khader AT, Al-Betar MA, Gandomi AH (2017a) A novel hybridization strategy for krill herd algorithm applied to clustering techniques. *Appl Soft Comput* 60:423–425
- Abualigah LM, Khader AT, Hanandeh ES (2017b) A new feature selection method to improve the document clustering using particle swarm optimization algorithm. *J Comput Sci* 25:456–466
- Abualigah LM, Khader AT, Hanandeh ES (2018a) A combination of objective functions and hybrid Krill herd algorithm for text document clustering analysis. *Eng Appl Artif Intell* 73:111–125
- Abualigah LM, Khader AT, Hanandeh ES (2018b) Hybrid clustering analysis using improved krill herd algorithm. *Appl Intell* 48(11):4047–4071
- Al-Janabi S (2017) Pragmatic miner to risk analysis for intrusion detection (PMRA-ID). In: International conference on soft computing in data science. Springer, Singapore, pp 263–277
- Al-Janabi S (2018) Smart system to create an optimal higher education environment using IDA and IOTs. *Int J Comput Appl* 1–16. <https://doi.org/10.1080/1206212X.2018.1512460>
- Al-Janabi S, Al-Shourbaji I (2017) Assessing the suitability of soft computing approaches for forest fires prediction. *Appl Comput Inform* 14:214–224
- Al-Janabi S, Rawat S, Patel A, Al-Shourbaji I (2015) Design and evaluation of a hybrid system for detection and prediction of faults in electrical transformers. *Int J Electr Power Energy Syst* 67:324–335
- Claessens S (2015) Important questions for housing markets and housing policy. *J Money Credit Bank* 47:419–421
- Crawford GW, Fratantoni AMC (2003) Assessing the Forecasting Performance of Regime-Switching, ARIMA and GARCH Models of House Prices. *Real Estate Econ* 31:223–243
- Demary M (2010) The interplay between output, inflation, interest rates and house prices: international evidence. *J Prop Res* 27:1–17
- Evermann J, Rehse JR, Fettke P (2017) Predicting process behaviour using deep learning. *Decision Support Syst* 100:129–140
- Fang H, Gu Q, Xiong W, Zhou LA (2015) Demystifying the Chinese Housing Boom. *Nber Chapters* 30:105–166
- Gensler A, Henze J, Sick B, Raabe N (2017) Deep learning for solar power forecasting—an approach using AutoEncoder and LSTM neural networks. In: IEEE international conference on systems, man, and cybernetics, pp 002858–002865
- Ghysels E, Plazzi A, Valkanov R, Torous W (2013) Chapter 9—forecasting real estate prices. In: Elliott G, Timmermann A (eds) *Handbook of economic forecasting*. Elsevier, Amsterdam, pp 509–580
- Goldberg DE (1989) Genetic algorithms in search. *Optim Mach Learn* xiii:2104–2116
- Guirguis HS, Giannikos CI, Anderson RI (2005) The US housing market: asset pricing forecasts using time varying coefficients. *J Real Estate Finance Econ* 30:33–53
- Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Comput* 9:1735–1780
- Holland JH (1992) *Adaptation in natural and artificial systems*. MIT Press, Cambridge
- Huang Z, Du X (2017) Government intervention and land misallocation: evidence from China. *Cities* 60:323–332
- Huang ZH, Wu CF, Du XU (2008) Empirical study on interaction between house price, interest rate and macro-economy in china. *China Land Sci* 22:38–44
- Jia S, Wang Y, Fan G-Z (2017) Home-purchase limits and housing prices: evidence from China. *J Real Estate Finance Econ* 56:386–409
- Limsombunchai V, Gan C, Lee M (2004) House price prediction: hedonic price model vs. artificial neural network. *Am J Appl Sci* 1(3):193–201
- Malpezzi S (1999) A simple error correction model of house prices. *J Hous Econ* 8:27–62
- Pain N, Westaway P (1997) Modelling structural change in the UK housing market: a comparison of alternative house price models. *Econ Model* 14:587–610
- Park B, Bae JK (2015) *Using machine learning algorithms for housing price prediction: the case of Fairfax County, Virginia housing data*. Pergamon Press Inc, Oxford
- Peng L, Liu S, Liu R, Lin W (2018) Effective long short-term memory with differential evolution algorithm for electricity price prediction. *Energy* 161:1301–1314
- Selim H (2009) Determinants of house prices in Turkey: hedonic regression versus artificial neural network. *Expert Syst Appl* 36:2843–2852
- Wang S, Shi H, Zhang J (2013) Commercial housing price index forecast based on the gray model. In: International conference on construction and real estate management, pp 1070–1077
- Wang X, Wen J, Zhang Y, Wang Y (2014) Real estate price forecasting based on SVM optimized by PSO. *Opt Int J Light Electron Opt* 125:1439–1443
- Wei SJ, Zhang X, Liu Y (2016) Home ownership as status competition: some theory and evidence. *J Dev Econ* 127:169–186
- Wu WY, Chen SP (2005) A prediction method using the grey model GMC(1, n) combined with the grey relational analysis: a case study on Internet access population forecast. *Appl Math Comput* 169:198–217
- Yang N, Xing LC (2006) Application of Grey-Markov model on the prediction of housing price index. *Stat Inf Forum* 5:011
- Yeung CW, Howes R (2006) The role of the housing provident fund in financing affordable housing development in China. *Habitat Int* 30:343–356
- Zhou Z (2018) Housing market sentiment and intervention effectiveness: evidence from China. *Emerg Mark Rev* 35:91–110

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.