

Analysis of Real Estate Predictions Based on Different Models

Yixu Li^{1, *}

¹Maynooth international engineering college, Fuzhou university, Fuzhou, China

*Corresponding author: mmoore53170@student.napavalley.edu

Abstract. As a matter of facts, real estate price prediction is meaningful for investors. This study conducts a comprehensive analysis of multiple real estate price prediction models, including genetic algorithms based on optimization algorithms, random forests based on machine learning, and recurrent neural networks (RNN) based on deep learning. The aim is to evaluate the strengths and weaknesses of these models in terms of predictive accuracy, explanatory power, and computational efficiency. Based on the analysis, genetic algorithms perform well in improving prediction accuracy and model convergence speed, but are sensitive to parameter selection and suffer from overfitting problems. Random forests improve prediction accuracy by integrating multiple decision trees, but are not interpretable enough and are susceptible to data noise. RNN has advantages in processing time series data, but requires a large amount of data and computing resources due to high computational complexity. Although each model has its limitations, it also has great potential. Through methods such as algorithm optimization, model integration, and feature engineering, the accuracy and reliability of real estate price prediction can be further improved. This study not only provides multi-angle model evaluation for real estate price prediction, but also provides development directions for future research.

Keywords: Model; real estate; prediction; analyze.

1. Introduction

With the continuous development of the global economy and the acceleration of urbanization, the real estate market has become an indispensable part of social and economic activities. Since real estate investment involves huge capital and a long return cycle, it is particularly important to accurately predict the trend of the real estate market. Real estate forecasts are not only related to the capital security of individual investors, but also affect the government's macro-control and urban planning, and are even closely related to the steady development of the national economy. Although research on real estate forecasting began in the mid-20th century, it is only in recent years that research in this field has developed rapidly with the rapid advancement of computing power and data analysis technology. Most early research relied on linear regression and time series analysis, but in recent years, machine learning and artificial intelligence methods have begun to emerge [1-3]. These advanced models, such as neural networks, random forests, and support vector machines, can more comprehensively consider the complexity of real estate prices that are affected by multiple factors. However, so far no model can be universally recognized as the "best" prediction model, and different models perform differently in different scenarios and data sets. Therefore, this study aims to systematically compare and evaluate the accuracy and reliability of these different types of prediction models in various practical application scenarios. Through such comparison, this article hopes to provide valuable reference information for investors, policymakers and academic researchers so that they can more accurately understand and predict the future trend of the real estate market [4-6].

In recent years, there has been a great deal of innovation and improvement in the field of predictive real estate analytics, particularly in model building and data analysis methods. Traditional forecasting models, such as linear regression and time series models (such as ARIMA), have always been the main tools for researchers and the industry, but these models often ignore the complexity and diversity of the real estate market [1]. Therefore, in recent years, machine learning and artificial intelligence technologies have been widely used in real estate prediction, such as neural networks [2]. These advanced models can more comprehensively handle multi-factor and non-linear relationships, improving the accuracy and reliability of predictions. Among them, the neural network model has

performed well in predicting housing prices in many cities in the United States due to its powerful data fitting capabilities. The random forest model predicts housing prices in many major cities in China, providing a more robust and robust prediction method. Support vector machine has been successfully applied in the prediction research of Australian house prices, demonstrating its advantages in processing small sample data [3, 4].

2. Basic Descriptions

The core purpose of real estate forecasting models is to predict future housing prices or market trends by analyzing past and existing data. These models are typically based on a range of factors that influence home prices and demand, such as economic indicators, location, property characteristics and market trends. In principle, these models can be divided into several categories: linear models, time series models, and machine learning models.

Linear Model is one of the earliest and simplest predictive models. They usually use multiple linear regression analysis to predict a dependent variable (such as house price) by setting a series of independent variables (such as house area, year of construction, etc.). This model is easy to understand and explain, but it often ignores the interactions and nonlinear relationships of multiple factors in complex markets [7]. Time series model, such as the autoregressive integrated moving average model (ARIMA), is mainly used to analyze and predict time series data. It not only considers the relationship between housing prices and other variables, but also introduces time delay factors to capture market dynamics more accurately. Machine learning model is a more advanced and complex type of model, including neural networks, random forests, and support vector machines. These models can handle large amounts of high-dimensional, unstructured data and automatically learn complex patterns and nonlinear relationships in the data. For example, neural networks can capture very subtle patterns and regularities in data through multiple layers of forward and backpropagation algorithms. In addition to these basic models, there are also more complex models, such as hybrid models or ensemble models, which combine the advantages of multiple prediction models to improve prediction accuracy. In practical applications, which model to choose often depends on the specific data set, prediction goals and available resources.

Fig. 1 is a trend decomposition diagram of monthly real estate prices in Zhengzhou. The blue curve represents the original sequence of real estate prices, the red curve represents the trend sequence of real estate prices, and the green curve represents the decomposed random fluctuation sequence of real estate prices in Zhengzhou.

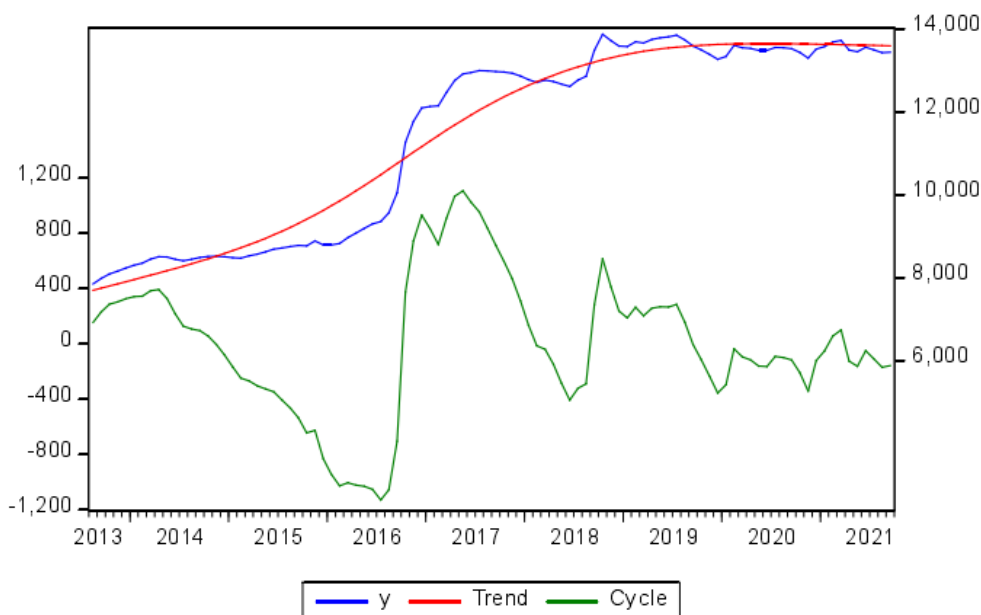


Fig. 1 Decomposition chart of monthly real estate price trends in Zhengzhou

3. Model Analysis

3.1. Optimization

In several studies, optimization algorithms such as genetic algorithms and neural network techniques are used as core parts of data analysis and predictive models to solve real-world problems. For example, Sommervoll and others were inspired by biological principles and tried to use genetic algorithms to conduct in-depth analysis of experimental data samples. They found that using genetic algorithms to search data can promote good clustering of samples, that is, data points can be grouped more effectively within the feature space, thereby improving the accuracy of classification or clustering tasks [7]. At the same time, Li and his team chose genetic algorithms to optimize the BP (back propagation) neural network model, and took China's housing prices as the research object. The addition of genetic algorithm not only speeds up the convergence of the model, but also effectively prevents the model from falling into a local optimal solution, which is a common problem with ordinary neural network algorithms [8]. This method enables the model to have higher prediction accuracy and provides a new solution path for similar prediction problems. Hu and others studied the housing price issues in major cities in Hainan Province. They used the TSA-BP model, an advanced model that integrates time series analysis (TSA) and BP neural networks [9]. Compared with the traditional BP model, the TSA-BP model shows more significant prediction effects. This discovery further confirms the effect of the combined use of optimization algorithms and neural networks, and also provides other researchers with new research directions and application frameworks in housing price prediction or similar fields [10].

Genetic algorithms have demonstrated significant advantages in data analysis and predictive model building. By combining with neural network technology, it can not only improve the prediction accuracy and convergence speed of the model, but also avoid some common computing traps, such as local optimal solutions. These studies not only achieved breakthroughs in their respective fields, but also provided valuable references for interdisciplinary research and applications. This joint application of genetic algorithms and neural networks provides a powerful tool for solving more advanced and complex problems in the future.

3.2. Random Forest

In the complex interweaving of housing price forecasting and macroeconomic analysis, methods such as random forest machine learning have been applied and verified, each with its own advantages, disadvantages and scope of application. For example, Gupta used random forest, an advanced machine learning method, to conduct an eye-catching analysis on the historical housing price data of US states and the District of Columbia [10]. His research shows that macroeconomic uncertainty has a simultaneous impact on housing price changes. This finding advances, in part, our understanding of the interconnections between the housing market and the larger economic system. At the same time, Dou and Wang used time series as their main analysis tool, and further used the seasonal index method to adjust the impact of seasonal factors, thereby optimizing the forecasting ability of the ARIMA model [11]. This method shows great power in predicting fluctuations in housing prices and can relatively completely capture the changing patterns of housing prices, so it has a high reference value. Tian started from another angle. She mainly used the gray-Markov model to conduct in-depth analysis of housing prices in Taiyuan City [12]. In order to improve prediction accuracy, she conducted two model optimizations in her research. This multiple optimization approach increases the robustness and accuracy of the model, thereby providing more reliable analysis results.

These studies not only provide different angles and methods to analyze the complex relationship between housing prices and macroeconomics, but also break through the limitations of a single model or method. Whether using random forest based on machine learning, using statistical time series and seasonal index methods, or integrating gray theory and Markov chain prediction models, these studies have made great achievements in housing price prediction and economic model construction. significant achievements

3.3. RNN Model

Serrano's research uses recurrent neural networks (RNN) for price prediction and focuses on the three fields of real estate, stocks and financial markets [12]. This study demonstrates the excellent accuracy of RNN models in processing time series data, indicating that advanced neural network models have great potential for complex, dynamic data analysis. Differently, ZonaKostic introduced image feature technology as a new perspective for real estate price prediction in his research. By quantifying the interior characteristics of a home, imaging technology can significantly improve the accuracy of predictive models [13]. This research not only broadens the multi-dimensional vision of housing price prediction, but also breaks through the limitations of traditional data analysis methods in capturing micro-features. Eren et al. explored the applicability of the prediction model from the other direction. They used the gray prediction model to focus on the prediction of Turkish real estate prices in the short term. The study found that this model performs particularly well in short-term prediction, indicating that the gray prediction model is a more appropriate tool in certain specific situations [14].

Based on the above research status, it can be seen that prediction models and methods have diverse application potential in different situations and data types. From time series analysis with advanced neural networks, to microscopic feature quantification with imaging techniques, to short-term analysis with gray predictive models, these different approaches each have their own advantages and present different challenges.

4. Limitations and Prospects

In real estate predictive analysis, different models have their own advantages and limitations. First of all, from the perspective of optimization algorithms, such as genetic algorithms, although it can well improve the traditional BP neural network model and improve prediction accuracy and convergence speed, it also has the risk of over-fitting and is very sensitive to the selection of algorithm parameters. The problem. Secondly, as an ensemble learning method, random forest performs well in handling multi-dimensional features and avoiding overfitting. However, it is sensitive to noise and outliers in the data, and the model is relatively poorly interpretable, which becomes a problem in an area of extremely high economic complexity such as the real estate market. Recurrent neural network (RNN) has obvious advantages in processing time series data, and is especially suitable for long-term trend analysis of real estate prices. However, the RNN model has high computational complexity and requires a large amount of data and computing resources, which limits its application in small-scale or short-term prediction. Faced with these challenges and limitations, future research can be expanded in multiple directions. For optimization algorithms, more experiments and parameter adjustment studies will help find more robust and efficient model combinations. Random forests can further improve their predictive accuracy and explanatory power through feature engineering or integration with other models. For RNN models, using more advanced algorithm architectures, such as long short-term memory (LSTM) or attention mechanisms, will solve some existing problems.

5. Conclusion

In real estate prediction analysis, a variety of models are applied, including optimization algorithms such as genetic algorithms, ensemble learning methods such as random forests, and deep learning-based recurrent neural networks (RNN). Each of these models has advantages and disadvantages, but the common goal is to improve prediction accuracy and model robustness. The genetic algorithm optimizes the prediction model through the principles of natural selection and evolution, improving the accuracy and convergence speed of the BP neural network in housing price prediction. However, it is sensitive to parameter selection and suffers from overfitting problems. Random forests improve prediction performance by integrating multiple decision trees, but are poorly interpretable and suffer from data noise. RNN is particularly suitable for processing time series data

of housing prices, but it has high computational complexity and requires a large amount of data and computing resources. Each model has limitations and technical challenges in its application. Future research directions include algorithm optimization, model integration, and feature engineering to overcome the limitations of a single model. For example, genetic algorithms can be combined with other optimization methods to reduce overfitting and improve model generalization. Random forests can improve their prediction accuracy and explanatory power through more advanced feature selection and engineering methods. RNN can also improve its performance in real estate prediction by introducing more advanced algorithm architectures, such as long short-term memory (LSTM) networks. Different models provide a wealth of tools and methods in real estate prediction analysis, but they also face challenges in many aspects such as data quality, model interpretability, and computational efficiency. With the further development of big data and computing technology, it is expected to solve these problems and further improve the accuracy and reliability of real estate forecasting. Therefore, the comprehensive application and continuous optimization of multiple models and methods will be an important development trend in future real estate forecast analysis.

References

- [1] Peng H, Li J, Wang Z, et al. Lifelong Property Price Prediction: A Case Study for the Toronto Real Estate Market. *IEEE Transactions on Knowledge and Data Engineering*, 2021.
- [2] Fatemeh M. Real-estate price prediction with deep neural network and principal component analysis. *Data Analysis*, 2022.
- [3] Ma H, Chen M, Zhang J. The Prediction of Real estate Price Index based on Improved Neural Network Algorithm. *Computer Science and Technology*, 2015.
- [4] Meng Y, Sun S. Real estate price forecasting based on ARIMA time series model -- A case study of Shenyang City. *Neijiang Science and Technology*, 2021.
- [5] Sommervoll D E, Sommervoll V. Learning from man or machine: Spatial aggregation and house price prediction. *CLTS Working Papers*, 2019.
- [6] Li C, Li X, Zhang K. House price prediction analysis based on BP neural network improved by genetic algorithm. *Computer Technology and Development*, 2018, 28(8): 5.
- [7] Hu Z, Qin Y, Han Y. Research on commercial housing price based on TSA-BP. *Journal of Hebei United University*, 2022, 3: 44.
- [8] Gupta R, Marfatia H A, Pierdzioch C, et al. Machine Learning Predictions of Housing Market Synchronization across US States: The Role of Uncertainty. *The Journal of Real Estate Finance and Economics*, 2021.
- [9] Dou N, Wang J. Research on forecasting method of housing price based on volatility decomposition. *China Market*, 2016, 9: 2.
- [10] Tian H CPI prediction based on twice improved grey weighted Markov model. *Journal of North University of China*, 2015, 2: 113-117.
- [11] Huang J, Zhang X. Xiamen housing price forecasting based on gray-Markov model. *Fujian Architecture*, 2012, 1: 3.
- [12] Shi S, Jou J B, Tripe D. Can interest rates really control house prices? Effectiveness and implications for macroprudential policy. *Journal of Banking & Finance*, 2014, 47(10): 15-28.
- [13] Knoll K, Schularick M, Steger T. No Price Like Home: Global House Prices, 1870-2012. *Social Science Electronic Publishing*, 2013.
- [14] Miraç E, Ali K Ç, İbrahim H. A Genetic Algorithm-Based Multivariate Grey Model in Housing Demand Forecast in Turkey. *Modelling*, 2016.