FULL PAPER PROCEEDING
Multidisciplinary Studies
Available online at www.academicfora.com

Full Paper Proceeding BESSH-2016, Vol. 117- Issue.2, 18-27

ISBN 978-969-670-260-4

BESSH-16

Combine Grey Relational Analysis and Weighted Synthesis for Housing Price Prediction

Weipeng Tan [1*], Tsung-Nan Chou [2]

[1, 2] *Department of Finance, Chaoyang University of Technology, Taiwan*

Abstract

The objective of this paper is to evaluate the performance of the grey relational analysis in the forecast of housing price for the real estate market of Taiwan. An instance-based approach which used k-nearest neighbor classifier was also applied for performance comparison. The grey relational analysis was modified to calculate the weighted synthesis of the top ten matching instances through various weighting strategies. The experimental results in this paper concluded that the grey relational analysis outperformed the instance-based approach in terms of the mean absolute error and root mean square error. In addition, the synthesis strategy with descending weights performed better than the averaging weights during the integration process of matching instances. The result also suggested that the performance was slightly decreased if the top ten matching instances was reduced to five instances. The grey relational analysis integrated with the weighted synthesis model can assist both buyers and owners in identifying opportunities and estimating the potential risks in a worsening real estate market.

Keywords— Housing Price Prediction, Grey Relational Analysis, Instance-Based Approach

Introduction

Money, is one of the most glamorous thing in this world. Everyone knows it, likes it with heartfelt desire, and hopes to make big money. Then, compared to gambling, efficient investment can provide a safer and faster shortcut to help us earn more money in short time. Since the investment will be full of considerable risks, how to predict the market trend and future investment performance precisely become the best way to reduce the risk. Real estate is one of the investment instruments in the financial markets and its long-term investment risk might be smaller to other investment products. It is because the real estate has a different feature where the homebuyers can fill their shelter needs and also preserve the appreciation of investment. That's why so many people are willing to invest in real estate. Additionally, a person's life cannot break off relations with the house. If a person wants to have his own house, it requires decades of effort. So, how to predict the future value of real estate and possess a house is a worth exploring questions.

By using the data of the Federal Reserve Bank of Dallas, we can find out the Real House Price Index (RHPI) for different countries. Through the analysis of the data, we can discover that many house price index fluctuations have some similarities even though the price index of the countries are different from each other. This issue deserves further exploration to investigate the relationship between house price fluctuations and nations. Perhaps the house price of a country will be affected by other countries, or vice versa. According to the survey from the Federal Reserve Bank between 2000 and 2015, we have found that the world real estate experienced a golden boom from 2000 to 2007. However, the world housing price began to fall dramatically after 2007. At the same time, the United States subprime mortgage crisis broke out, and the next year the world's largest financial holding institution Lehman Brothers (Lehman Brothers Holdings Inc.) declared bankruptcy. So what does it mean? Is there any relationship between those events caused the world's real estate prices fell sharply? The above question motivated this study to search for a solution to predict the future value of the real estate and decrease the default risk of the investors.

The research scope and applied variables of this study referred to some previous articles. Ge (2004) focus on Hong Kong real estate market and develops a forecasting model of residential property prices by using an artificial neural network approach. According to Lynn Wu et al. (2009) the future of prediction. By using google searches

*All correspondence related to this article should be directed to Tan Wei Peng from Department of Finance, Chaoyang University of Technology, Taiwan
Email: tanweipeng_2006@hotmail.com

foreshadow housing prices and Quantities. Gupta et al. (2012) study of the Southern California Market. Author using time-series models as a method to examine the time-series relationship between house prices. As suggested by Ng (2015) presents a mobile application that can generate predictions for future housing prices in London. Gaussian Processes (GP) for regression had been chosen as a model due to its flexible and probabilistic approach to learning and model selection. These references inspired us to shape the scope of this study.

The focus of our research work on Taiwan's real estate market is motivated for a number of reasons. Firstly, Taiwan is an island and a beautiful country attracted to resident owned destinations. Taiwan is one of the Asia's four dragons, and is famous for high-tech industry and technological innovation. However, Taiwan's economy is highly dependent on exports and therefore it is always vulnerable to unpredictable external factors. Whether the real estate industry of Taiwan will be affected by these factors or not, to forecast real estate prices will become the same concern to other countries. Secondly, the performance of Taiwan's financial sector is speedily thriving, although it is not well as those of Singapore and Hong Kong. This means that Taiwan's real estate industry has much space to develop and this makes more interested to know the situation of the house price in the future of Taiwan. Thirdly, real estate market is a very high volatility market. Now in Taiwan, the volatility of the real estate market is stable, occasional reversal but compared to other financial assets, Taiwan's real estate market can be described as "not fall myth". The Taiwan, Central Bank of China (CBC) President Perng also view that Taiwan's housing market has "stabilized" at 26 March 2015. So this is a good time to learn how to predict the real estate prices and maybe one day it will lead us to become a regal.

Invest a real estate is expensive, so investors will try to research the all information of the real estate during the process of investment. This can reduce the chance to buy low-quality housing and to avoid the phenomenon of investment losses. This paper focuses on the analysis of the factors affecting the price of real estate, and how to take advantage of the current situation with real estate factors to predict the future value of housing prices in the housing market. We use the global price index to compare with Taiwan price index and then find out the country which its real estate price index is the most similar to the index of Taiwan. This will help us to explore the correlation between them and find the impact of variables. The valuation of real estate investment is important for predicting house price and the variables for the valuation is the reason to affect the price of a house. So, how to determine the factor affecting the price of real estate is very important. The generated of the factor helps to analyze the price of real estate and also helps to in-depth study of real estate and its future trends. Hope this is a worthy of further exploration article, and provides a buyer or investor how to understand the price fluctuation through a variety of factors. Last, we primarily using grey relational analysis in this study to predict the price of real estate.

Literature Review

Asencio-Cort´es et al. (2015) present an efficient nearest neighbor (NN) approach, called PKK-PCP, and an application for the protein inter-residue contact prediction. The great strength of using this approach is its adaptability to that problem. Furthermore, their method improves considerably the efficiency with regard to other NN approaches and their NN-based method combines parallel execution with a k-d tree as search algorithm. The input data used in algorithm are based on structural features and physical-chemical properties of amino acids besides of Evo- lutionary information. Results obtained show better efficiency rates, in terms.

Park and Bae (2015) uses machine learning algorithms as a research methodology to develop a housing price prediction model. To improve the accuracy of housing price prediction, the study analyzes the housing data of 5359 townhouses in Fairfax County, Virginia, gathered by the Multiple Listing Service (MLS) of the Metropolitan Regional Information Systems (MRIS). They develop a housing price prediction model based on machine learning algorithms such as C4.5, RIPPER, Naïve Bayesian, and AdaBoost and compare their classification accuracy performance. They then propose an improved housing price prediction model to assist a house seller or a real estate agent, make better informed decisions based on house price valuation. The experiments demonstrate that the RIPPER algorithm, based on accuracy, consistently outperforms the other models in the performance of housing price prediction.

The internet is a resourceful virtual world where a variety of information is open to public access, which expedites knowledge explosion. While the openness of the internet facilitates knowledge, learning, it also carries some information people do not need. The identification of advertisement on the internet pages develops its necessity in the areas such as internet security, search engine, or even education (Chao and Yang, 2015). In Chao's paper, as an attempt to differentiate advertisements, some approaches are examined to predict the advertisements on internet pages. As a matter of fact, some promising results are presented in their paper.

With the rapid development of China's economy and the continuous expansion of the scale of commercial housing, the price of commercial housing has been undergoing a greater fluctuation, which has significantly affected economic and social development. Hu and Li (2013) using the method of grey relational analysis, this paper takes Harbin for example; studying the factors affecting the price fluctuation of commercial housing has important

theoretical and realistic significance. To analyze the various factors affecting the price of commercial housing. Finally, in view of the above factors, some appropriate suggestions for the healthy development of the commercial housing market in Harbin are put forward.

Chen et al. (2014) employs the Grey Relational Analysis (GRA) and Artificial Neural Network (ANN) to measure the impact of key elements on the forecasting performance of real estate investment trust (REIT) returns. They adjust the parameters from GRA and inserts the key elements into the fitted ANN model by comparing the learning effect of the Back-propagation Neural Network (BPN). Their study found that the ranking provided by the GRA is significant in correcting prediction errors using the learning outcome of the BPN. The neural network model proved to minimize the error function and was able to adjust weighted values in order to enhance prediction accuracy.

Data mining methods are often implemented at advanced universities today for analyzing available data and extracting information and knowledge to support decision-making. Kabakchieva (2013) presents the initial results from a data mining research project implemented at a Bulgarian university, aimed at revealing the high potential of data mining applications for university management.

AL-Marwani (2014) aims to provide an approach to real estate residential modeling and forecasting covering property types' correlation. Previous studies have examined real estate price indices at the macro level (the general, all real estate house price indices). There has not been a study that examines real estate price forecasts by property types within a city. The contribution of this thesis is its focus on time series analysis as well as causal modeling within a city with the objective of providing a better understanding of the dynamics of real estate price changes.

Vijayarani and Muthulakshmi (2013) have analyzed the performance of Bayesian and Lazy classifiers for classifying the files, which are stored in the computer hard disk. There are two algorithms in Bayesian classifier namely BayesNet, and Naïve Bayes. In lazy classifier has three algorithms, namely IBL, IBK and Kstar. The performances of Bayesian and lazy classifiers are analyzed by applying various performance factors. From the experimental results, it is observed that the lazy classifier is more efficient than Bayesian classifier.

Wu and Brynjolfsson (2009) demonstrate a highly accurate but simple way to predict future business activities by using data from such search engines. Applying their methodology to predict housing trends, they find that their index of housing search terms can predict future quantities and prices in the housing market. During their sample period, each percentage rise in their housing search index predicts sales of 121,400 additional houses in the next quarter. This approach can be applied to other markets, transforming the future of prediction.

Method

The grey relational analysis is a method of forecasting and decision-making to analyze the status of the system. The system is a measurement method used to analyze the size of the degree of correlation in the data of a discrete sequence. The non-important factor will delete by in a range of standard values. Another prediction system, named K-nearest-neighbor classifier, is able to predict nominal classes and their class probabilities, as well as numeric class values. It is a kind of different search algorithms can be used to speed up the task of finding the nearest neighbors. The methodologies are briefly described later in this section. Then explain detail several important implementation notes of predictor. Finally, define the evaluation measures used for effectiveness and efficiency.

Grey relational analysis is one of the method of grey system theory. Grey System Theory was created in 1982 by Professor Deng Ju-Long. It is against the consulting inadequate and incomplete case with internal system models for process the system of relational analysis and model construction. It is using the methods of forecasting and decision-making to analyze the status of the system. Grey theory is widely used in various fields, it can focus on the uncertainly of things, multivariable input and incompleteness discrete data to do a systematic and efficient processing

Grey Relational Analysis is a measurement method used to analyze the size of the degree of correlation in the data of a discrete sequence. The non-important factor will delete by in a range of standard values. General abstract systems which are similar to social systems, economic systems, agricultural systems, ecological systems, education systems, etc., involve many factors. Some mutual reactions among the factors use to determine the development situation and tendency of the systems. We often want to know: Which factors among them are more important? Which factors have more effects on the future development of the systems than others? Which factors actually cause desirable changes in the systems so that these factors need to be strengthened and which factors hinder desirable development of the systems so that they need to be controlled?

All these problems are commonly studied in the analysis of systems. But there are exist many factors, such as the area planted, irrigation facilities, fertilizers, soil quality, seeds, labor availability, weather conditions, farming technologies, and related government policies. All of them are affecting the desirable outcome. In order to achieve the goal of less input and more output with as great economic, social, and ecological benefits as possible, we must make use of the theory of systems analysis.

Many methods in statistics, such as regression analysis, variance analysis, and principal component analysis, are all commonly used in the analysis of systems. However, these methods have the following pitfalls.

- A large amount of data is required. Otherwise, it would be difficult to draw statistical conclusions with reasonable confidence and reliability.
- It is required that all samples or populations satisfy certain typical probability distribution(s), that the relation between the main characteristic variable of the system and factor variables is roughly linear. These requirements are often difficult to satisfy in real-life practice.
- Heavy-duty computations are often needed.
- It often happens that quantitative conclusions may not agree with qualitative analysis results, causing misunderstandings about the systems under consideration.

The fundamental idea for the grey relational analysis is to measure the closeness of a relationship between data sequences by the similarity level of the geometrical curves formed by the sequences. In other words, the incidence degree of each attribute is calculated by comparing the geometrical similarity between the corresponding sequences. In grey relational analysis, the more similar in the geometrical shape of comparing curves means the higher grey incidence degree for the sequences. The grey relational analysis employed in this study is categorized as an absolute incidence approach because all the sequences in the training data are compared to a specified target sequence rather than being compared with each other in the relative incidence approach.

The GIA approach starts with the generation of grey sequences to transform a normal data sequence into a non-dimensional sequence where all values in the sequence originally measured on different scales are adjusted to a common scale. After that, both the grey relational coefficient and the degree of grey incidence are calculated for individual elements of data sequence and the whole sequence respectively. Although some research work suggested different methods to calculate the grey relational coefficient, the results of these methods could lead to inconsistency in ranking order and require further synthesis for the outcomes. The processing of the grey relational analysis is briefed as the following steps.

Step 1: Transform original data sequences to grey generation sequences by different sequence operators which convert data according to the expectation of maximizing, minimizing and specific effects of attributes. These operators alter each value in sequence with the corresponding upper-bound, lower-bound and moderate-bound measurement in the data sequence.

Step 2: Calculate grey relational coefficient and degree of grey incidence. Assume that the following sequence x0 represents the characteristics of an instance in testing dataset.

$$x_0 = (x_0(1) \ldots\ldots x_0(n)) \in X_0$$

And the sequence xi is the relevant instance in training dataset.

$$x_i = (x_i(1) \ldots\ldots x_i(n)) \in X_i$$

$$o = 1,2,3,\cdots,m, \;\; i = 1,2,3,\cdots n$$

Suppose the $R\ (x_0, x_i)$ denotes the absolute degree of grey incidence for both x_0 and x_i sequences, and the grey relational coefficient for individual element of the sequence is denoted as $\gamma\ (x_0\ (k), x_i\ (k))$. For $\zeta\ ((0, 1)$, where ζ is called distinguishing coefficient and used to magnify the calculation for easier comparison between sequences. Deng (1998) defined the original grey relational coefficient as follows equations.

$$\Delta_{oi}(k) = \| x_0(k) - x_i(k) \|$$

$$\gamma(x_0(k), x_i(k)) = \frac{\Delta_{\min} + \zeta\Delta_{\max}}{\Delta_{oi}(k) + \zeta\Delta_{\max}}$$

The degree of grey incidence $R(x_0, x_i)$ can be calculated by the following average approach.

$$R(x_o, x_i) = \frac{1}{n}\sum_{k=1}^{n}\gamma(x_o(k), x_i(k))$$

Several researches such as Wu, Wen and Hsia, assigned the distinguishing coefficient ζ to 1, and modified the related equations of Deng's grey relational coefficient. Their modified methods combined with the Deng's method were implemented in this study, and represented as the model GIA (1) to GIA (4) respectively. Their modified formulas are given below.

Wu's method:

$$\overline{\Delta}_{0i} = \sqrt{\frac{1}{n}\sum_{k=1}^{n}[\Delta_{0i}(k)]^2}$$

$$\gamma(x_0(k), x_i(k)) = \frac{\Delta_{\min} + \Delta_{\max}}{\overline{\Delta}_{oi}(k) - \Delta_{\max}}$$

Wen's method:

$$\overline{\Delta}_{0i} = \left\{\frac{1}{n}\sum_{k=1}^{n}[\Delta_{0i}(k)]\right\}$$

$$\gamma(x_0(k), x_i(k)) = \frac{\Delta_{\min} + \Delta_{\max}}{\overline{\Delta}_{oi}(k) - \Delta_{\max}}$$

Hsia's method:

$$\overline{\Delta}_{0i} = \left\{\frac{1}{n}\sum_{k=1}^{n}[\Delta_{0i}(k)]\right\}$$

$$\gamma(x_0(k), x_i(k)) = \frac{\Delta_{\max} - \overline{\Delta}_{oi}(k)}{\Delta_{\max} - \Delta_{\min}}$$

Step 3: The ranking order of grey incidences for sequences is defined as the follows.

$$R(x_o, x_i) \geq R(x_o, x_j)$$

The model is rigid, simply cannot cater to the needs of humanity. So, in this study, we use the original formula reform to personal needs formula in this paper. The formula is as below:

$$\gamma(x_0(k), x_i(k)) = 1 - \frac{|x_i(k) - x_0(k)|}{\max[\max x_i(k) - x_0(k); x_0(k) - \min x_i(k)]}$$

IBk is a k-nearest-neighbor classifier that uses the same distance metric. Some classifiers, however, are able to predict nominal classes and their class probabilities, as well as numeric class values. Weka. classifiers. lazy. IBk is one of them. Can select appropriate value of K based on cross-validation and also can do the distance weighting. The number of training instances kept by the classifier can be restricted by setting the window size option. As new training instances are added, the older ones are removed to maintain the number of training instances at this size. It is a kind

of different search algorithms can be used to speed up the task of finding the nearest neighbors. And this K-nearest neighbor classifier, is the other method which use in this study.

With the equation, the data from Taiwan Open Data (data.gov.tw) is used for getting the results. The data are adapted from March 2015 to September 2015, taking only the required about 74568 of data. Data are obtained by downloading from the website which updates every two weeks, the content of the release information was about two months ago. Each file contains 28 fields and the data is containing detail basic information about housing records. But the data have an error, so need to be treated before use. After screening the useless data and selection of variables, it is already standby further to analyst house price. The main point is to predict the house price of entire Taiwan.

The effectiveness and efficiency are assessed using several measures. Regarding the effectiveness are computed averages strategy and weight strategy, defined as shown in Eqs.1, Eqs.2 and Eqs.3 respectively. These measures are common indicator and widely used for comparison the prediction of different models.

RMSE is a frequently used measurement of the differences between values (population values and sample) predicted by a model or an estimator and the value actually observed.

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|f_i - y_i| = \frac{1}{n}\sum_{i=1}^{n}|e_i| \qquad (1)$$

Mean Absolute Error (MAE) is an average of the absolute errors $|e_i| = |f_i - y_i|$, f_i is the prediction and y_i is the true value. The MAE is a common measurement of forecast error in time series analysis.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{t=1}^{n}(f_i - y_i)^2}{n}} = \sqrt{\frac{\sum_{t=1}^{n}(x_{1,t} - x_{2,t})^2}{n}} \qquad (2)$$

RMSE also can call as RMSD, it is the square root of the variance, which is known as the standard error. Predicted values f_i for times t of a regression's dependent variable y_i is computed for n different predictions as the square root of the mean of the squares of the deviations. The formula change when measuring the average difference between two-time series $x_{1,t}$ and $x_{2,t}$.

$$\mathrm{RAE} = \frac{\sum_{j=1}^{n}|f_{(ij)} - y_j|}{\sum_{j=1}^{n}|y_{(ij)} - \bar{y}_j|} \qquad (3)$$

Relative absolute error is very similar with relative squared error in the sense that it is also relative to a simple predictor, which is just the average of the actual values. $f_{(ij)}$ is the value predicted by the individual program i for sample case j (out of n sample cases); y_j is the target value for sample case j; and $\bar{y}$ is the average of y_j.

Empirical Study

The empirical study begins with reduced-form model and Grey Relational Analysis with k-nearest-neighbor classifier can be used for forecasting housing prices. The selection of valid training input variables is the first step. The second step is to design the model of Grey Relational Analysis to perform pattern recognition and to generalize. After the model is tested, the predicted outcome may be validated. Then compare with the result of k-nearest-neighbor classifier to find out which will be better.

Research finds out that either five or seven variables, by using this formula have the same conclusion. The five variables are subject of the transaction, the area of transaction and the structure of the building. The structure of the building can be divided into three, which are living room, bathrooms and bedrooms. The total of all the above is five variables and the reason is wanting to know the prediction of the factors of houses will cause how much influence. While the reason to add two more variables is hoping to find out that the compartment of the house and is there anyone managed the house will affect the result. This result is telling us that the added variable is no effect on prediction. In this result, only using the data of Gao Xiong, which one of the city in Taiwan.

First, using the method of Grey Relational Analysis. In the beginning, do a comparison between test samples and training samples. They will find out the top ten that the training samples which are similar to the predict values.

Table 1:
The top ten result of Grey Relational Analysis

Rank	Degree	No.	Y (million)
1	1	546	20.78
2	0.9999404	6078	20.66
3	0.9998583	4101	15.39
4	0.9998352	2097	19.95
5	0.9998087	4744	16.50
6	0.9996824	3794	18.80
7	0.999512	3796	14.38
8	0.9994233	6218	19.50
9	0.9994127	73	20.00
10	0.9991302	3658	16.50

After that calculate the averages strategy and weight strategy of ten training samples. Then use Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Relative absolute error (RAE), three of them is a quantity used to measure how close forecast or predictions are to the eventual outcomes.

The top ten of the training samples which are similar to the predicate value are 20.78, 20.66, 15.39, 19.95, 16.5, 18.80, 14.38, 19. 20.000 and 16.50 (million). After that calculate the averages strategy of this top ten and separately using 1, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1 with the top ten values to calculate the weighting strategy. So, we can calculate that the result of an averages strategy of MAE in grey relational analysis in top ten is 2695596 and the weighting strategy of MAE is 2664574. The result of an averages strategy of RMSE is 5220676 and the weighting strategy of RMSE is 5255079. Then the result of an averages strategy of RAE is 54.59% and the weighting strategy of RAE is 53.96%.

Table 2:
The MAE, RMSE and RAE of top ten result

Grey Relational Analysis (Top 10)		
	Averages strategy	Weighting strategy
MAE	2695596	2664574
RMSE	5220676	5255079
RAE	54.5911%	53.9629%

Second, use the method of k-nearest-neighbor classifier. Using the same test samples and training samples to get another result. The result of an averages strategy of MAE in k-nearest-neighbor classifier is 3183646. The result of an averages strategy of RMSE is 7941508. Then the result of an averages strategy of RAE is 64.881%.

Table 3:
The MAE, RMSE and RAE of k-nearest-neighbor classifier

K-nearest-neighbor Classifier (IBk)	
MAE	3183646
RMSE	7941508
RAE	64.881%

Third, compare the differences of the result of predicting the value of Grey Relational Analysis in top one until the top ten. Figure 1-3 is the calculation of averages strategy and weight strategy on top one until top ten. By using the result of calculation, we can find out by using how many predictive value the prediction error value is smaller.

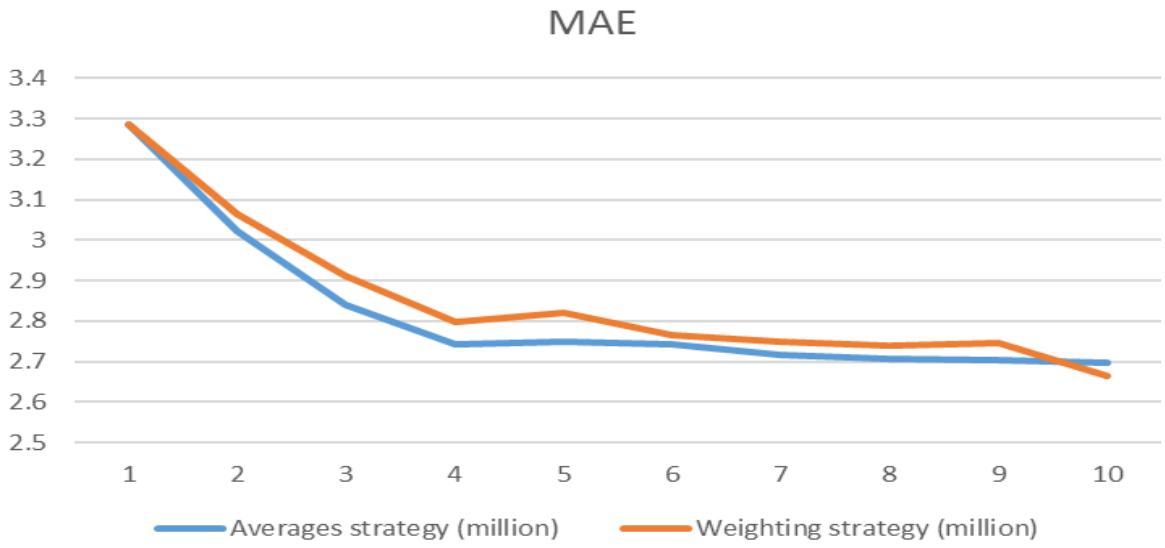

Figure 1: The MAE of the top one to top ten result in Grey Relational Analysis

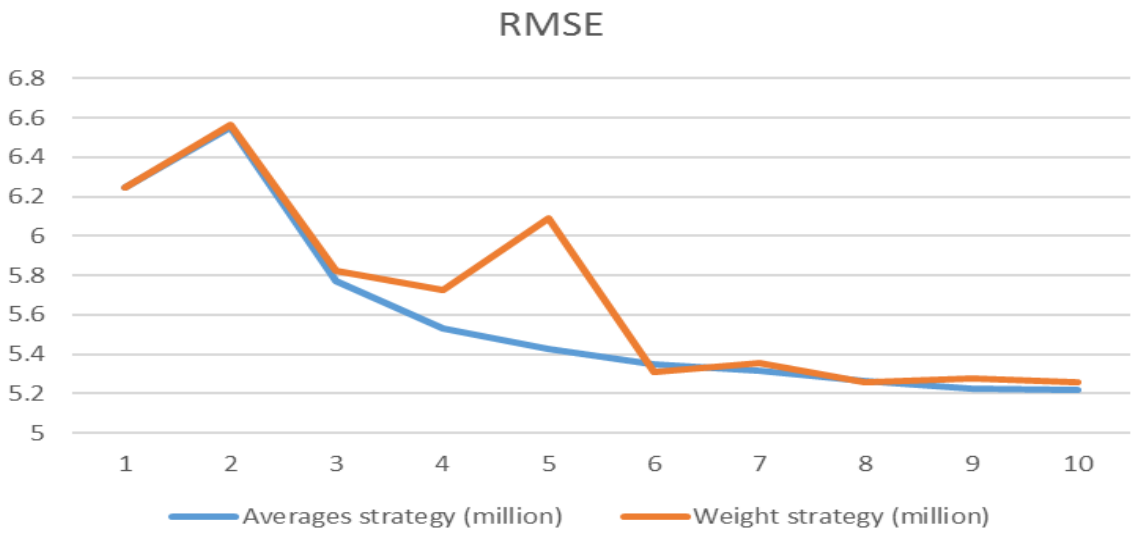

Figure 2: The RMSE of the top one to top ten result in Grey Relational Analysis

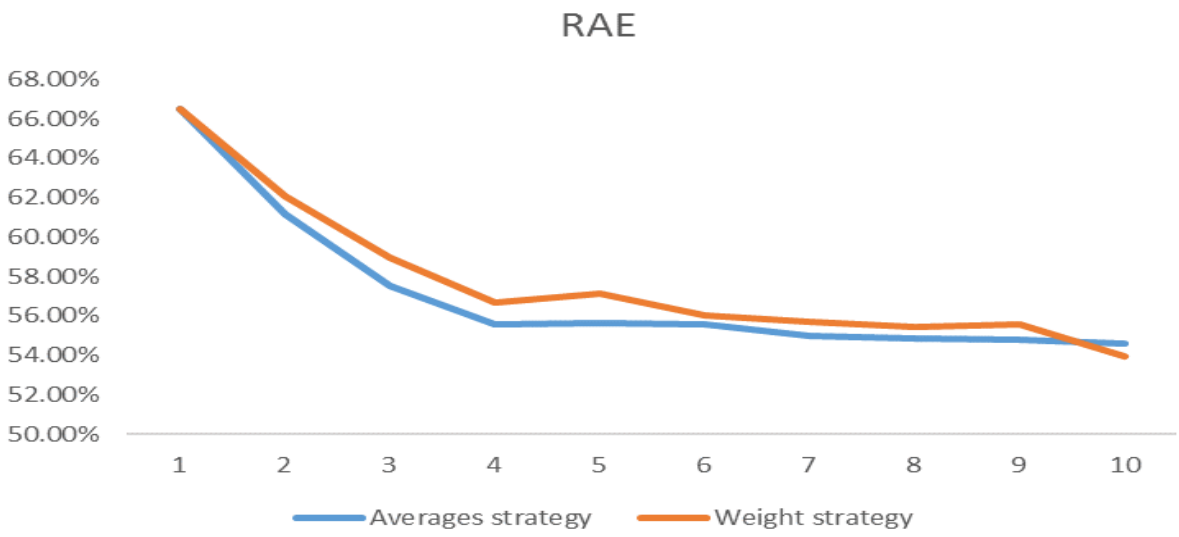

Figure 3: The RAE of the top one to top ten result in Grey Relational Analysis

The result of comparing two methods is the calculation of MAE, RMSE and RAE of k-nearest-neighbor classifier is higher than the calculation of Grey Relational Analysis. Besides that, compares the two results of Grey Relational Analysis, the calculation of MAE, RMSE and RAE in averages strategy and weighting strategy is significantly different. Through the figure, we find that the ten result of the averages strategy much better than weighting strategy.

Then compares the result of predicting the value of Grey Relational Analysis in top one until the top ten. Using the top ten that the training samples which is similar to the predicate value is more accurate. It is because no matter the result of MAE, RMSE and RAE in averages strategy or weighting strategy, the result of top ten is smaller than others.

Conclusions

Conclusion is the formula of Grey Relational Analysis confirm can be used to predict the house price. No matter what variables, it can help us to find out ten similar house price, then analyze computing the rules of housing pricing followed. By comparing the two methods, we can prove that the prediction of Grey Relational Analysis is better than k-nearest-neighbor classifier. Because no matter MAE, RMSE and RAE of Grey Relational Analysis, the result of calculation of averages strategy and weighting strategy is smaller than the result of k-nearest-neighbor classifier. After that compare between the averages strategy and weighting strategy of Grey Relational Analysis in top ten, the compote of MAE and RAE of weighting strategy in is smaller than averages strategy, but when calculate in RMSE, it is contrary, the RMSE weighting strategy is higher than averages strategy. This result occurs because weight strategy calculation needs to add square and square root, so it will lead the decreasing weights strategy and obtain lower results. Finally, compare the result between the top one to the top ten which are using Grey Relational Analysis. The result of top ten is smaller than the result of other nine results. This means that after the predictions of Grey Relational Analysis, it is better than averages strategy by using the top ten of the training samples which are similar to the predicate value to calculate the value of MAE, RMSE and RAE, the prediction error value is smaller and this means that the predicted value is the nearest.

Personal thinking that the next three years in Taiwan, house price fluctuations will not have improved significantly. This is because Taiwan is newly developed the taxation policy of housing and land. Little effect on long-term owner-occupied, but there is a big impact for investors. The difference in this research is the method and the newest data. In this research, it still has a limitation of the study. The limit is the time intervals of data must be select in one year. We recommendations for future research is adjusting the parameters of annuals. It can resolve the restrictions of the study.

References

Onieva, E., Santos, I., Osaba, E., Quintián, H., & Corchado, E. (Eds.). (2015). *Hybrid Artificial Intelligent Systems: 10th International Conference, HAIS 2015, Bilbao, Spain, June 22-24, 2015, Proceedings* (Vol. 9121). Springer.

Chen, J. H., Chang, T. T., Ho, C. R., & Diaz, J. F. (2014). Grey Relational Analysis and Neural Network Forecasting of REIT returns. *Quantitative Finance*, *14*(11), 2033-2044.

Gupta, R., & Miller, S. M. (2012). The time-series properties of house prices: A case study of the Southern California market. *The Journal of Real Estate Finance and Economics*, *44*(3), 339-361.

Al-Marwani, H. A. (2014). *An approach to modeling and forecasting real estate residential property market* (Doctoral dissertation).

Hu, Y., & Li, P. (2013). A Grey Relational Analysis on Factors Affecting the Price of Commercial Housing. *Proceedings ICCREM*, 929-936.

Janet, X., & Lam, K. C. (2012). Building a House Prices Forecasting Model in Hong Kong. *Construction Economics and Building*, *2*(2), 57-70.

Zhang, R., Zhang, Z., Liu, K., & Zhang, J. (Eds.). (2014). *LISS 2013: Proceedings of 3rd International Conference on Logistics, Informatics and Service Science*. Springer.

Kabakchieva, D. (2013). Predicting student performance by using data mining methods for classification. *Cybernetics and information technologies,13*(1), 61-72.

Ng, A., & Deisenroth, M. (2015). Machine Learning for a London Housing Price Prediction Mobile Application.

Park, B., & Bae, J. K. (2015). Using machine learning algorithms for housing price prediction: The case of Fairfax county, Virginia housing data. *Expert Systems with Applications*, *42*(6), 2928-2934.

Vijayarani, S., & Muthulakshmi, M. (2013). Comparative analysis of bayes and lazy classification algorithms. *International Journal of Advanced Research in Computer and Communication Engineering*, *2*(8), 3118-3124.

Wu, L., & Brynjolfsson, E. (2013). The future of prediction: How Google searches foreshadow housing prices and sales. *Available at SSRN 2022293*.

Witten, I. H., & Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.

Xin, J. G., & Runeson, G. (2004). Modeling Property Prices Using Neural Network Model for Hong Kong. *International Real Estate Review*, *7*(1), 121-138.