HYBRID GENETIC-BASED SUPPORT VECTOR REGRESSION WITH FENG SHUI THEORY FOR APPRAISING REAL ESTATE PRICE

Chih-Hung Wu[a], Chi-Hua Li[b], I-Ching Fang[c], Chin-Chia Hsu[d], Wei-Ting Lin[e], Chia-Hsiang Wu[f]

[a,b,c,e,f] Department of Digital Content and Technology, [d] Department of Information Management
[a,b,c,e,f] National Taichung University; Taichung, Taiwan, R.O.C. [d] National Taiwan University, Taipei, Taiwan, R.O.C.
E-mail: chwu@ntcu.edu.tw; bit096105@ms3.ntcu.edu.tw, bit096103@ms3.ntcu.edu.tw, d93725007@ntu.edu.tw, bit096101@ms3.ntcu.edu.tw, chia0210@gmail.com

Abstract

In this paper, we proposed a novel house prediction model that integrated hybrid genetic-based support vector regression (HGA-SVR) model and feng shui theories for developing a high accuracy appraising real estate price system in Taiwan. In Taiwan, feng shui theory applies in choosing good days, divination and house selection. From the past researches, many factors might affect the real estate price which are the announced land values, the building room age, building total number of floor, the transportation condition and surrounding environment of house etc. However, few studies have been considered the feng shui effects in appraising real estate price. Therefore, the present study pioneers in applying Feng Shui theories for developing a high accuracy real estate price prediction system with back-propagation neural network(BPN), fuzzy neural network (FNN) and Hybrid Genetic-based SVR (HGA-SVR) to compare.

Our results obtained from the comparison between two house price models with various artificial neural network models. By comparing the accuracy with various network architectures, the result demonstrates that HGA-SVR is the best network architecture and the feng shui model has a better performance in BPN, FNN and HGA-SVR. Our house price prediction system discovers some real estate price much higher than the reasonable prices. This result shows these unreasonable price needs adjusting to become more reasonable to conform the housing market.
Keywords: Hybrid Genetic-based Support Vector Regression, feng shui theory, house price prediction

1. Introduction - The relationship between feng shui and the appraisal of real estate

In Chinese culture, feng shui plays a very important role in buying and selling houses. It determines the values and beliefs that dictate expectations as well as responses to housing prices [1]. The transaction mode of between seller and buyer is an important research subject at housing market [2]. There is a phenomenon of Taiwan's property market that has been a decrease in Pre-Construction Real Estate and increase in older or brand new houses recently. [3]. Housing Demand Survey of the Fourth Quarter 2007, 5, 4). The housing market in Taiwan is very much driven by the local market economy. Houses are expensive commodities for Taiwanese, whereas about 87.83% of total households own houses [4].

Actually, geomantic taboos also provide the rules of traditional building [5]. The traditional housing market in Taiwan, it has been limited customers to choose their special needs. Standardized building and design let owners often require to remodel their new houses before moving in, spending more time and money for remodeling.

How to provide customers a simple and easy system that can estimate houses with Feng Shui is a develop importance of property market in the future [6]. Besides, numerous people thinking about feng Shui at buying step is another bigger character of Taiwan housing market [1]. Dr. Tsai thinks feng Shui has its valuable to research, whether from nationality and architecture or not. Geomantic sciences have been verified by architecture physics and structure lately. So, Feng Shui certainly has a relation between our live and Chinese building.

In addition, the one of techniques of predicting house price: SVM. Support vector machines (SVMs) have been successfully applied to a number of applications such as including handwriting recognition, particle identification (e.g., muons), digital images identification (e.g., face identification), text categorization, bioinformatics (e.g., gene expression), function approximation and regression, and database marketing and so on. Although SVMs have become

978-0-7695-3580-7/09 $25.00 © 2009 IEEE
DOI 10.1109/ACIIDS.2009.41

more widely employed to forecast time-series data [7-9] and to reconstruct dynamically chaotic systems [10-13], a highly effective model can only be built after the parameters of SVMs are carefully determined [14]. By using and improving SVM, we can use the novel method which is the Hybrid Genetic-based SVR (HGA-SVR) to obtain the better performance in predicting house price.

2. Literature Review

2.1. Feng shui theory application in buying house

Feng shui is a part of Chinese traditional culture. There is a term, Feng shui, which first appears in the book, translated literally as Burial Book, which is attributed to Guo Pu. The theoretical foundation of Feng shui lies in the cosmogonic concepts of "ch'i" [15]. We are able to have good interaction between buildings and environment through the chi which is around houses. It will be an ideal house which can gather wind and water [5].

Feng shui has attached much importance to research and discuss in the world recently. And it has been applied to many domains widely. For instance, house design by surname in Feng shui [16]; Integrating natural and cultural heritage: the advantage of Feng shui landscape resources [17]; Feng shui and its impacts on land and property developments [1]. And there are many scholars to bring out some studies continually about Feng shui. For example, Mak and Ng [18] bring up a point of Feng shui to check up quality of architecture design; Chiou and Krishnamurti [19] develop an interactive system which can assist us to measure size and located direction of traditional building; Hwangbo [20] thinks Feng shui theory can support architecture and planning design. At last, Feng shui certainly can construct our living environment and buildings [5].

2.2. Factors affecting of real estate price

From the past literatures, we discovered many factors affect the real estate price. Tam [1] pointed out many factors that are environmental setting, transport facilities, school network, shopping facilities, regional characteristics, including of three significant independent variables, Feng shui, accessibility, and building age. According to the study of Juan (2006), living room, dining room, kitchen, study room, master room, bedroom, bathroom, storeroom. Besides above factors, land size, age of the house, type of house, garages, amenities also affect the real estate price [21]. Based on the results of these studies, we choose some significant independent variables of the real estate price as our variables. The list of variables is shown in section 3.

2.3. Applications of computational intelligence techniques in the real estate price prediction

Din, A. a l., [22] point out that the pricing of real estate research has focused on real estate valuation and the construction of real estate indices using mostly hedonic pricing models [23-29]. Din, A. a l., [22] use geographic information system (GIS) and ANN to compare various real estate valuation models and the manner in which they take into account environmental variables.

There are many ways to predict real estate prices. Limsombunchai [21] compares the predictive power of the hedonic model with ANN model on house price prediction. Lai [30] applies the BPN and hedonic price method to predict real estate prices. A FNN prediction model based on hedonic price theory to estimate the appropriate price level for a new real estate is proposed. [31]

Several studies have proposed optimization methods which used a genetic algorithm for optimizing the SVR parameter values. To overcome the problem of SVR parameters, a GA-SVR has been proposed in an earlier paper [32] to take advantage of the GAs optimization technique. However, few studies have focused on concurrently optimizing the type of SVR kernel function and the parameters of SVR kernel function. The present study proposed a novel and specialized hybrid genetic algorithm, HGA-SVR, for optimizing all the SVR parameters simultaneously to predict house price.

3. Our proposed house price prediction model

3.1. Our proposed novel HGA-SVR model

In our proposed novel HGA-SVR model, the type of kernel and the parameter value of SVR are dynamically optimized by implementing the evolutionary process, and the SVR model then performs the prediction task using these optimal values. Our approach simultaneously determines the appropriate type of kernel function and optimal kernel parameter values for optimizing the SVR model to fit various datasets. The types of kernel function and optimal values of the SVR's parameters are determined by our proposed novel HGAs with a randomly generated initial population of chromosomes. The types of kernel function (Gaussian (RBF) kernel, polynomial kernel, and linear kernel) and all the values of the parameters are directly coded into the chromosomes with integers and real-valued numbers, respectively. The proposed model can implement either the roulette-wheel method or the tournament method for selecting chromosomes. Adewuya's crossover method and boundary mutation method were used to modify the chromosome. Only the one best chromosome in each generation survives to move on to the succeeding generation. Thus, we believe our proposed HGA-SVR model is able to handle huge data sets and can easily and efficiently be combined with the integer genetic algorithm and real-valued genetic algorithm for developing the hybrid genetic algorithm.

3.2. Research Architecture

In this study, we used both Model 1 without feng shui variables and Model 2 with feng shui variables which are from the perspective of Asian homebuyers to predict residential housing price and evaluate the forecasting accuracy of both models based on the twelve inputs.

3.3. House Price Prediction Model

The present study uses the Super PCNeuron 5.0 software and Matlab for the BPN, FNN and HGA-SVR model training and testing. After repeating the validation test in the various network architectures, we use the parameter setting of the minimum RMSE as the criterion to construct our network model. The parameter setting of this model is depicted in Table 1.

Table 1. Parameters setting among various network architectures (BPN, FNN and HGA-SVR)

Network architecture	Parameter	Model 1	Model 2
BPN, FNN	Number of input variables:	12	16
	Number of neurons in the first hidden layer:	14	14
	Number of neurons in the second hidden layer:	14	14
	Number of output variables:	1	1
	Number of train examples:	133, 114	152
	Number of test examples:	57, 76	38
	Number of train cycles:	150000	
	Learn rate:	1.0	
	Learn rate reduced factor:	0.99	
	Learn rate minimum bound:	0.1	
	Momentum factor:	0.9	
	Momentum factor reduced factor:	0.99	
	Momentum factor minimum bound:	0.1	
HGA-SVR	Population Size:	20	
	Generations:	50-100	
	Gamma Range:	1	
	Sigma Range:	0 ~ 1000	
	Selection Method:	tournament	
	Mutation Method:	uniform	

3.4. Housing Data

The training and test example of 190 housing information in the North District and the West District of Taichung City is selected from one of public estate agent in June 2008. The housing data with 190 observations is split into training and test collections by a 6/4, 7/3 and 8/2 ratio. The West District has 110 observations and North District has 80 observations. The statistics of our sample distribution are shown in Table 2.

Table 2. Statistics of estate sample distribution

Ratio:	6:4		7:3		8:2	
District	Training data	Test data	Training data	Test data	Training data	Test data
West (N=110)	66	44	77	33	88	22
North (N= 80)	48	32	56	24	64	16
Total (N=190)	114	76	133	57	152	38

Note: N denotes the observations

3.5. Attributes of housing data

The attributes of housing data are summarized in Table 3. Different attributes have different code demand and can be directly coded into different numbers with integers, respectively, based on the attribute characteristic. When the housing data have these Feng shui taboos, the code of these attributes increases one value.

Table 3. Attributes and Recoding book of housing data

Attribute	Value	ReCode
Bedroom	Continuous value	--
Living room	Continuous value	--
Bathroom	Continuous value	--
House age	Continuous value	--
Storey	Continuous value	--
Land size	Continuous value	--
Exposure	East	1
	West	2
	South	3
	North	4
	Southeast	5
	Northeast	6
	Southwest	7
	Northwest	8
Management	All day	1
	Daytime	2
	Nothing	3
semidetached house	Yes	1
	No	2
Construction	SRC	1
	SC	2
	RC	3
Parking facilities	Plane	1
	Mechanical	2
	Nothing	3
Amenities around the house	School district	+1
	Park	+1
	Transport system	+1
	Major construction	+1
Feng shui taboo	Housing appearance	+1
	Door	+1

	Window	+1
	Toilet	+1
	Stair	+1
Housing price	Continuous value	--

4. Data Analysis

4.1. Evaluation

The measurement indicators which are RMSE, MAPE and FE be used to evaluate the accuracy of both models. These indicators are defined as follows:

1. Root mean square error (RMSE)

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(A_i - P_i\right)^2} \quad (1)$$

2. Mean absolute percentage error (MAPE)

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{A_i - P_i}{A_i}\right| \quad (2)$$

3. Forecasting error (FE)

$$FE = \left|\frac{\left(P_i - A_i\right)}{A_i} * 100\right| \quad (3)$$

where A_i denotes the actual housing price, P_i represents the estimated housing price, i denotes the ith observations, and n is the number of observations. The forecasting error results are classified into three major categories: (1) FE < 5% with higher predicted performance; (2) FE between 5% and 15 % which is a fuzzy predicted area and slightly unreliable; (3) FE > 15% with unsatisfying performance [33].

4.2. Accuracy comparison between model 1 and model 2 in various network architectures

This study adopts the minimum RMSE as the performance indicator to construct our model for the first time [30]. In order to find the minimum RMSE as the criterion to remove deviation, the number of training cycle is 50000, 100000 and 150000 iterations with various ratios (6:4, 7:3 and 8:2) be used in two models, respectively. The RMSE comparison among different train cycles are shown in Fig. 1.

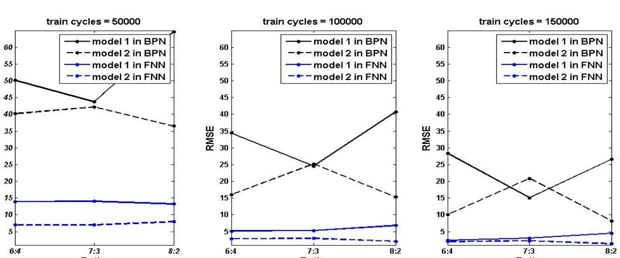

Figure 1. RMSE comparison among different train cycles

After removing the deviations which have over estimation price, we use the remaining data as training and test examples to get the better prediction performance. The accuracy comparison results are shown in Table 4. The best network architecture has the minimum MAPE, higher FE 5% and lower FE more than15%. Thus, the best network architecture is the HGA-SVR with 4.97% MAPE, 89.16% FE of less than 5%, 4.01% FE between 5% and 15% and 6.83% FE more than 15% in model 2 by a ratio 8/2. Another obvious result is the fact that the model 2 with Feng shui variables can acquire better accuracy. Finally, the performance comparison among different network architectures is compared in Fig. 2. It proves a point that our novel model with Feng shui variables can improve the accuracy of traditional model without Feng shui variables.

Table 4. Results in various network architectures

		Ratio	MAPE	FE less than 5%	FE 5%~15%	FE more than 15%
BPN	Model 1 (*N*=130)	7:3 (91:39)	12.5%	70.59%	9.80%	19.61%
	Model 2 (*N*=151)	8:2 (121:30)	9.86%	82.12%	4.64%	13.25%
FNN	Model 1 (*N*=117)	6:4 (70:47)	9.8%	68.38%	11.11%	20.51%
	Model 2 (*N*=153)	8:2 (122:31)	5.67%	84.97%	3.27%	11.76%
HGA-SVR	Model 1 (*N*=94)	7:3 (66:28)	8.99%	70.21%	14.89%	13.82%
	Model 2 (*N*=120)	7:3 (76:33)	4.79%	89.16%	4.01%	6.83%

Note: Model 1 includes basic/spatial environment variables
Model 2 includes basic/spatial environment/feng shui taboo variables
N denotes the observations without the deviations more than NT$300,000

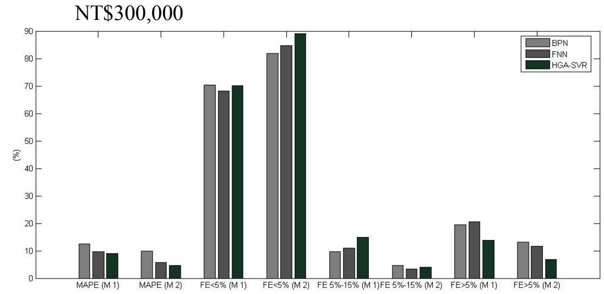

Figure 2. Performance comparison among different network architectures

Note: M 1 denotes model 1 and M 2 denotes model 2
Smaller MAPE is better; Bigger FE is better

4.3. Discussions

4.3.1. Discussion of variable importance

After comparing the results of various network architectures, we use the back-propagation neural network to observe the sensitivity analysis clearly among two models as shown in Table 5.

Table 5. Results of sensitivity analysis in various models of BPN

Model 1		Model 2	
Variables	Sensitivity analysis	Variables	Sensitivity analysis

Bedroom	-0.0325	Bedroom	-0.1291
Living room	-0.008	Living room	-0.0879
Bathroom	-0.0116	Bathroom	-0.0212
House age	-0.0293	House age	0.0048
Exposure	-0.0223	Exposure	-0.0234
Storey	-0.0056	Storey	0.0005
Management	-0.0655	Management	-0.0699
Semidetached house	-0.0128	Semidetached house	-0.0082
Construction	-0.0063	Construction	0.0135
Parking facilities	0.0312	Parking facilities	0.0361
Amenities around the house (school, park, transport system and major construction)	0.0348	Amenities around the house (school, park, transport system and major construction)	0.0252
Land size	**0.1893**	**Land size**	**0.2739**
		Toilet	*-0.0148*
		Door	*0.0152*
		Housing appearance	*-0.0012*
		Window	*-0.0072*

Note: Words in bold denote the most important variable in predicting house price
Words in italic denote the feng Shui variables

Both in the model 1 and model 2 in Table 5, land size has the maximum positive value denotes the maximum influence between variables and housing price. Conversely, when the weight is getting smaller, the variables have less influence. When the housing data have the Feng Shui taboos, the code of Feng Shui factor increases one value. Thus, we can find almost every Feng Shui variables have negative value which correspond our expectation in model 2. The importance sequence of four representative Feng Shui indexes: (a) Toilet (b) Window (c) Housing appearance (d) Door. Specifically, the Feng Shui variables require a lower weight to enhance housing price. According to the rule based evaluation approach above and the predicting result, the sensitivity analysis of Feng Shui variables show the close relation with housing price in Chinese building.

4.3.2. Discussion of the outlier detection for property price

The focus of the outlier detection for property price is an important issue to detect the unreasonable house price and avoid the wrong result produced. When the prediction price has over estimation price, the house owners may need to revise their house price to conform house market. On the other hand, the difference between estimation price and prediction price will influence the prediction performance. Thus, the deviation of three hundred thousand dollars which has over estimation price have to be removed in advance before using various network architectures to predict the housing price. Since the number of deviations is too many, the part of deviations among model 2 in HGA-SVR is listed in Table 6.

Table 6. Unreasonable house price based on HGA-SVR model 2

Cases	House price	Prediction price	Deviations (over estimation price)
1	1650	1544.676	105.324
2	788	633.034	154.966
3	498	662.943	164.943
4	368	498.358	130.358
5	668	719.960	51.960
6	568	704.076	136.076
7	580	648.265	68.265
8	600	647.431	47.431
9	388	498.358	110.358
10	498	643.936	145.936

5. Conclusions and Economical Implications

By comparing the accuracy with various network architectures, the result demonstrates that HGA-SVR is the best network architecture and model 2 with Feng shui variables has a better performance in BPN, FNN and HGA-SVR. Furthermore, the prediction of real estate price system with feng shui factor has some advantages between sellers and buyers. First, we combine present search system of real estate and invite some professors of Feng Shui to analyze the houses and set up some rules for the system. It can provide some suggestions about house conditions for buyers. They can use the system to seek their own house. Second, when those buyers want to buy their exclusive houses, the system can offer some that are more adaptive price of house at once. And then they are able to choose one of them. In the future, as the database is created more complete, we can pick those bad houses which will harm people's health and make a more reasonable price. At last, through the database system, architects can refer to those layouts which people want to buy or high price. Many people in Taiwan just have one choice to buy their own house in their life. So, to buy an adaptive house is very important.

6. References

[1] C.M. Tam, T.Y.N. Tso, and K.C. Lam, "Feng Shui and Its Impacts on Land and Property Developments", *Journal of Urban Planning & Development*, 1999, 125(4), pp. 152, 12p.

[2] J. Barlow, P. Childerhouse, D. Gann, H.M. Séverine, M. Naim, & R. Ozaki, "Choice and delivery in housebuilding: lessons from Japan for UK housebuilders", *Building Research and Information*, 2003, 31(2), pp. 134-145.

[3] Institute for Physical Planning and Information, Housing Demand Survey of the Fourth Quarter 2007, 5, 4.Retrieved Jane 1, 2008, from http://www.ippi.org.tw/ (2008, Feb 26).

[4] Construction and Planning agency ministry of the Interior, Housing State Survey Retrieved Jane 1, 2008, from http://www.cpami.gov.tw/web/index.php?option=com_content&task=view&id=6976&Itemid=15 (2007, Dec 5).

[5] P.T. Han, "A study of taboos in Feng-Shui practice of home building", *Bulletin of architecture and city planning national Taiwan university volume*, 1987, number 1, research, pp.5-55.

[6] Y.K. Juan, and S.G. Shih, Y.H. Perng, "Decision support for housing customization: A hybrid approach using case-based reasoning and genetic algorithm", *Expert Systems with Applications*, 2006, 31(1), pp. 83-93.

[7] E. H. Tay, Francis, and L. Cao, "Application of support vector machines in financial time series forecasting", *OMEGA:The International Journal of Management Science*, 2001, 29(4), pp.309-317.

[8] L. Cao, "Support vector machines experts for time series forecasting", *Neurocomputing*, 2003, 51(1-4), pp.321-339.

[9] K. Kim, "Financial time series forecasting using support vector machines", *Neurocomputing*, 2003, 55(1), pp.307-319.

[10] K.-R. Müller, A. Smola, G. Rätsch, B. Schölkopf, J. Kohlmorgen, and V. Vapnik, "Predicting Time Series with Support Vector Machines", *Proceedings ICANN'97*, 1997, pp.999.

[11] S. Mukherjee, E. Osuna, and F. Girosi, "Nonlinear Prediction of Chaotic Time Series using a Support Vector Machine", *Proceedings NNSP'97*, 1997.

[12] D. Mattera, and S. Haykin, "Support vector machines for dynamic reconstruction of a chaotic system", In B. Schölkopf, C. J. C. Burges, and A. J. Smola, editors, *Advances in Kernel Methods --- Support Vector Learning*, Cambridge, MA, MIT Press, 1999, pp. 211-242.

[13] A. Kulkarni, V.K. Jayaraman, and B.D. Kulkarni, "Control of chaotic dynamical systems using support vector machines", *Physics Letters A*, 2003, 317(5), pp.429-435.

[14] K. Duan, S.S. Keerthi, and A.N. Poo, "Evaluation of Simple Performance Measures for Tuning SVM Hyperparameters", *Neurocomputing*, 2003, 51(1-4), pp.41-59.

[15] A.B. Hwangbo, "An alternative tradition in architecture: conceptions in feng shui and its continuous tradition", *Journal of Architectural and Planning Research*, 2002, 19(2), pp. 110-130.

[16] S.J. Lu, "House design by surname in Feng Shui", *The Journal of Architecture*, 2000, **5**, pp. 355-367.

[17] L.M. Marafa, "Integrating natural and cultural heritage: the advantage of feng shui landscape resources", *International Journal of Heritage Studies*, 2003, 9(4), pp. 307-323.

[18] M.Y. Mak, & S.T. Ng, "The art and science of Feng Shui: a study on Architects' perception", *Building and Environment*, 2005, 40, pp. 425-432.

[19] S.C. Chiou, & R. Krishnamurti, "Unraveling feng-shui", *Environment and Planning B: Planning and Design*, 1997, 24, pp. 549-572.

[20] A.B. Hwangbo, "A new millennium and feng shui", *The Journal of Architecture*, 1999, 4, pp. 191-198.

[21] V. Limsombunchai, C. Gan, & M. Lee, "House Price Predication: Hedonic Price Model vs. Artificial Neural Network", *American Journal of Applied Science*, 2004, 3:pp. 193-201.

[22] A. Din, M. Hoesli, and A. Bender, "Environmental variables and real estate prices", *Urban Studies*, 2001, Vol. 38, pp. 1989-2000.

[23] J.W. HOAG, "Towards indices of real estate value and return", *The Journal of Finance*, 1980, 35, pp. 569–80.

[24] M. MILES, R. COLE, and D. GUILKEY, "A different look at commercial real estate returns", *Journal of the American Real Estate and Urban Economics Association*, 1990, 28, pp. 403–430.

[25] R.B. WEBB, M. MILES, and D. GUILKEY, "Transactions-driven commercial real estate returns: the panacea to asset allocation models", *Journal of the American Real Estate and Urban Economics Association*, 1992, 20, pp. 325–357.

[26] J.D. FISHER, D.M. GELTNER, and R.B. WEBB, "Value indices of commercial real estate: a comparison of index construction methods", *Journal of Real Estate Finance and Economics*, 1994, 9, pp. 137–164.

[27] J.R. KNIGHT, J. DOMBROW, and C.F. SIRMANS, "A varying parameter approach to constructing house price indices", *Real Estate Economics*, 1995, 23, pp. 187–205.

[28] R.A. MEESE, and N.E. WALLACE, "The construction of residential housing price indices: a comparison of repeat-sales, hedonic-regression, and hybrid approaches", *Journal of Real Estate Finance and Economics*, 1997, 14, pp. 51–73.

[29] K.A. KIEL, and J.E. ZABEL, "Evaluating the usefulness of the American housing survey for creating house price indices", *Journal of Real Estate Finance and Economics*, 1997, 14, pp. 189–202.

[30] P.Y. Lai, "Applying the Artificial Neural Network in Computer-assisted Mass Appraisal", *Journal of housing studies*, 2007, 16(2), pp. 43-65.

[31] Liu, J.-G. Zhang, X.-L. Wu, W.-P., "Application of Fuzzy Neural Network for Real Estate Prediction", *Lecture Notes IN Computer Science*, 2006, Vol. 3973, pp. 1187-1191.

[32] C.C. Hsu, C.H. Wu, S.J. Chen, and K.L. Peng, , "Dynamically Optimizing Parameters in Support Vector Regression: An Application of Electricity Load Forecasting", *Hawaii International Conference on System Science (HICSS39)*, January 4-7 (2006).

[33] N. Nguyen, & A. Cripps, "Predicting Housing Value: A Comparison of Multiple Regression Analysis and Artificial Neural Network", *The Journal of Real Estate Research*, 2001, 22(3):pp. 313-336.