

Fuzzy Logic Application for House Price Prediction

Abdul G. Sarip¹ and Muhammad Burhan Hafez²

¹Faculty of Built Environment, University of Malaya, 50603 Kuala Lumpur, Malaysia

²Faculty of Computer Science and Information Technology,
University of Malaya, 50603 Kuala Lumpur, Malaysia

¹ag.sarip@gmail.com, ²burhan.hafez@gmail.com

Abstract

Various methods have been used previously to estimate housing sales prices to model the underlying non-linearity relation between housing attributes and the price of property. More advanced non-linear modelling techniques such as Artificial Neural Networks (ANN), Fuzzy Inference System (FIS) and Fuzzy Logic (FL) emerged recently to model the nonlinear relation between the independent variables and the price function. A new structured model for house prices prediction based on Fuzzy Logic is proposed. A fuzzy logic based regression model has proved to be effective to address many prediction problems used in business forecasting, marketing and insurance. This paper highlights the development of a theoretical formulation for sales price prediction through the utilisation of a fuzzy regression model by applying fuzzy logic and fuzzy inference system techniques. The results show favourable outputs which indicate superior prediction function when compared with ANN and FIS as well as indicate the fuzzy functional relationship between dependent and independent variables.

Keywords: property price prediction, fuzzy logic, fuzzy regression, artificial neural networks

1.0 INTRODUCTION

Real estate consists of the land and physical building on it or any other improvements above or under the land (Graaskamp, 1981). Real estate is a composition of attributes that can be categorised as either static or dynamic (Lawson, 2008). These attributes can be either in the forms of quantitative or qualitative data. In order to provide a quantitative measure of the value of real estate, valuation or appraisal is therefore required to determine the benefits and liabilities accruing from the ownership of real estate (Pagourtzi et al., 2003).

The most common method used in the valuation of residential property is Comparison approach. The objective of comparison is to compare sales of similar properties (also known as comparable) in order to estimate the market value of a subject property. This is done by way of analysis of the attributes of comparable that were transacted recently. Any differences between the subject and comparable property attributes will be adjusted by using a comparison unit that should become the basis for comparison. In addition to time and location, factors such as physical characteristics (i.e. lot size, square footage, age, bedrooms, bathrooms, swimming pool, etc.), quality, neighbourhood and other indicators are usually considered in the comparable analysis (Sarip, 2006).

The advent of latest technology in computing hardware and software as well as advanced data processing techniques undeniably has provided an influence on the valuation process. This has led to the use of more analytical applications, i.e. machine learning which include techniques such as Artificial Neural Networks (ANN), Fuzzy Logic (FL) and Expert System (ES) (Zurada et al., 2006). Furthermore these techniques when employed in combinations with methodologies such as Automated Valuation Model (AVM) and Computer Assisted Mass Appraisal (CAMA) have shown variable degrees of prediction accuracy and error. The underlying purpose of integrating new technology into the valuation process is to provide fast and improved decisions (Sarip, 2006). One such technology is Fuzzy Logic.

2.0 LITERATURE REVIEW

Multiple Regression Analysis (MRA) is a statistical method that predicts value of a dependent variable such as house price through the values of the independent variables such as property characteristics. This classical method has been widely used in prediction applications despite its weakness to clarify the relationship between independent and the dependent variables in the presence of outliers, nonlinearity, and non-normality (Brunson et al., 1994; Do & Grudnitski, 1992). Similarly, Artificial Neural Networks (ANN) model is also subjected to several shortcomings such as over-fitting problem when the relation between each of the input and the network's output is monotonic (Daniels & Kamp, 1999). The application of Fuzzy Logic (FL) to predict house price has begun to show success in modeling complex and nonlinear behaviour of input and output variables. According to Zurada et al. (2006), FL is designed to mimic the function of a human brain in making decision by considering all quantitative and qualitative data in making the best possible decision (output) given the input factors. The potential strength of FL is when the use of less quantifiable data could improve the prediction of selling prices (Zurada et al., 2006).

2.1 Fuzzy Logic

Fuzzy Logic (FL) was established by Dr. Lotfi Zadeh of University of California Berkeley in the 1960s. FL was introduced as a technique to model the improbability of normal spoken and written language. This includes linguistic functions that are qualitative in nature such as near, far; or hot, cold; or tall, short. According to Steele (1990), FL is based on the "degree of truth" instead of the conventional (Boolean) logic that specify function either "true or false" with assigned values of 1 or 0. FL is a superset of Boolean logic that has been extended to handle the concept of partial truth - truth values between "completely true" and "completely false" (Sun & Qiu, 2010). FL combines real values [0, 1] and logic operations – all things in between and allows values to be more accurate.

The underlying concept of FL is a linguistic variable, that is, a variable whose values are qualitative (words) rather than quantitative (numbers). FL may be viewed as a methodology for computing with words rather than numbers. Although words are inherently less precise than numbers, their use is closer to human intuition which then can exploit the tolerance for imprecision (Rubens, 2006). Another concept in FL that plays a central role in many of its applications, is that of a fuzzy if-then rule or simply, fuzzy rule (Singh & Singh, 2013). Rule-based systems have a long history of use in Artificial Intelligence (AI), but what is missing in such systems is a mechanism that deals with fuzzy consequents and fuzzy antecedents. In FL, this mechanism is provided by the calculus of fuzzy rules. The calculus of fuzzy rules serves as a basis for what might be called the Fuzzy Dependency and Command Language (FDCL). For example, FL includes 0 and 1 as extreme cases of truth (or "the state of matters") and also includes the various states of truth in between so that, for example, the result of a comparison between two things could be not "cold" or "hot" but "2.5" of variation" (Csepedia.com, 2014). FL is observed to be useful for approximate reasoning that allows for prediction with estimated values under incomplete information (Steele, 1990).

2.2 Fuzzy Sets

A set is a collection of things that belong to a specific group. It is specifying a common property among "things" and then gathers all the "things" that have this common property. Any "things" either belongs to or does not belong to that set. In describing the set, Figure 1 illustrates the degree of membership of the 'tall' set, and a person to either belongs in it or does not belong in it.

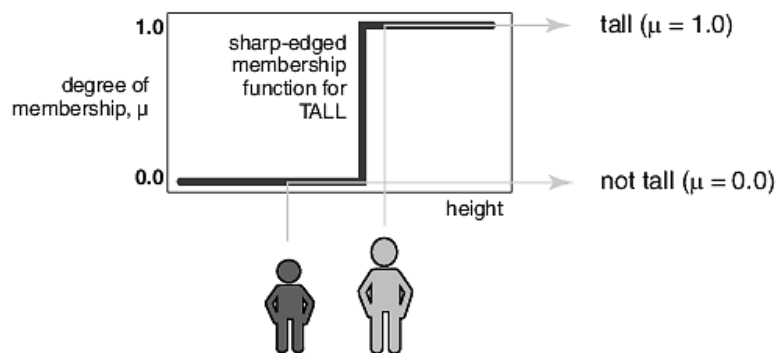


Figure 1: Normal data set representation
Source: Mathworks.com (2014)

The concept of Boolean logic (1.0, 0.0) and a subset (tall, not tall) has been shown to have a strong relationship as indicated in Figure 1. The same relationship can be used to explain the relationship between fuzzy logic and fuzzy subset theory. FL defined a fuzzy subset TALL, which should answer the question "to what degree is person A tall" or simply put, "person A is taller than person B by x inches". In this case, TALL is a "linguistic variable" that expresses the "tallness" category. Similar to Figure 1, the fuzzy set approach to the set of tall men provides a much better representation of the tallness of a person (Borah et al. 2013) as illustrated in Figure 2.

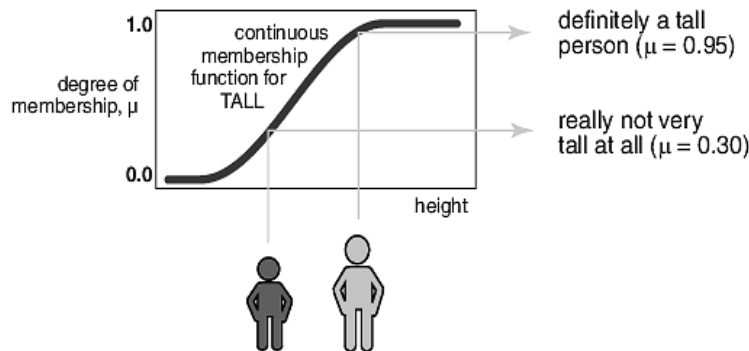


Figure 2: Fuzzy data set representation

2.3 Logic Operations

Zadeh (1965) has given a clear example on how logic operations work. He expressed a statement like "A is SHORT" means in fuzzy logic, and how to interpret a statement like :

A is SHORT and B is LONG or (not C is MEDIUM)

The standard definitions in fuzzy logic are:

$$\begin{aligned} \text{truth (not A)} &= 1.0 - \text{truth [A]} \\ \text{truth (A and B)} &= \text{minimum [truth(A), truth(B)]} \\ \text{truth (A or B)} &= \text{maximum [truth(A), truth(B)]} \end{aligned}$$

2.4 Fuzzy Logic System

A fuzzy logic system is the nonlinear mapping that turns data from fuzzy crisp input into non fuzzy value output. It consists of four main stages: (1) fuzzifier (input), (2) rules, (3) inference engine (process), and (4) defuzzifier (output) as illustrated in Figure 3 (Yuan & Herbert, 2012).

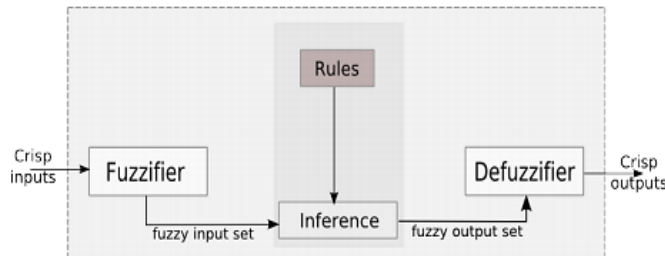


Figure 3: A Fuzzy Logic system

The process of fuzzy logic is explained in Figure 4 (www.cs.bilkent.edu.tr, 2010). Firstly, a crisp set of input data are gathered and converted to a fuzzy set using fuzzy linguistic variables, fuzzy linguistic terms and membership functions. This step is known as fuzzification. Thereafter, an inference is made based on a set of rules. Lastly, the resulting fuzzy output is mapped to a crisp output using the membership functions, in the defuzzification step.

Step	Stage	Process
1	Fuzzifier	Define the linguistic variables and terms
2	Fuzzifier	Construct the membership functions
3	Rules	Construct the rule base
4	Inference	Convert crisp input data into fuzzy values
5	Inference	Evaluate the rules in the rule base
6	Inference	Combine the result in each rule
7	Defuzzification	Convert the output data to non-fuzzy values

Figure 4: Fuzzy Logic algorithm
Source: www.cs.bilkent.edu.tr (2010)

2.5 Regression Model and Fuzzy Regression

Regression is a least square method of an explanation of a causal relation between variables. It shows the changes (variance) in independent variables (e.g. size, location, age, bedroom, quality, etc.) which will then be reflected upon the changes (variance) in a dependent variable (e.g. price). If the independent variables satisfactorily explain the changes in the dependent variable, the model is deemed as ‘best fit’ and can be used for modeling function such as prediction. Figure 5 illustrates a basic regression model between dependent variable Y and independent variables X.

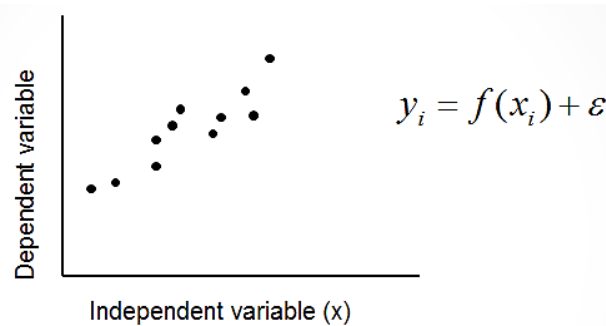


Figure 5: A basic regression model

In conventional statistical linear regression,

$$Y = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n + \varepsilon$$

where Y is the dependent (observed) response, $x_1 \dots x_n$ is the independent variables, n is the number of variables, $\beta_0 \dots \beta_n$ are the unknown parameters (coefficients), and ε is the error term.

Fuzzy Least-Square Regression (FLSR) is an extension of the use of the least-square method to solve for fuzzy regression. The FLSR model for property price prediction is

$$\tilde{Y}_i = \tilde{\beta}_0 + \tilde{\beta}_1 x_1 + \dots + \tilde{\beta}_n x_n + \tilde{\varepsilon}$$

where \tilde{Y}_i is the fuzzy response, $\tilde{\beta}_0 \dots \tilde{\beta}_n$ are the fuzzy parameters of the model, $x_1 \dots x_n$ are the input variables, and $\tilde{\varepsilon}$ is the fuzzy error term.

3.0 DATA PREPARATION

A data set of 350 terrace houses in the District of Petaling, Kuala Lumpur has been collected (October 2013 – March 2014) for the construction and testing of the FSLR model. Property specific data were acquired from the National Property Information Centre (NAPIC) of the Malaysia Department of Valuation & Property Services. Twelve (12) features were selected as the best representatives for each property as shown in Table 2. The input variables RC, QF and LO which are ordinal categorical have been converted to numerical forms.

Table 2: Datasets

#	Feature	Meaning	Sample Value
1	LA	Land area (sq.m)	153.29
2	BA	Built-up area (sq.m)	167.23
3	BE	# Bedrooms	4
4	BT	# Bathrooms	3
5	AG	Building age (years)	7
6	RC	Repair condition	Good
7	QF	Quality of finishes	Very good
	LO	Location - distance function to...	
8	LO1	Shops	Near
9	LO2	School	Near
10	LO3	Main roads and highways	Far
11	LO4	Parks and Green Areas	Adjacent
12	LO5	Industrial zone	Far

4.0 ANALYSIS AND EVALUATION

The dataset is divided into training set and testing set, with 90:10 ratios. The training set is used in the construction of the model, while the testing set is used to test the performance of the model. The evaluation criterion is based on Mean Absolute Error (MAE) on the test set, which is defined as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

where n is the size of the test set, y_i is the actual property's price, and \hat{y}_i is the estimated price. The value of the eight (8) input variables have been normalised to fall into the range [-1, 1] in order to prevent the variables from overweighting each other.

The result of the FLSR model which is illustrated in Figure 6 shows how this model captures the trend in the prices data in almost the whole range of the test set.

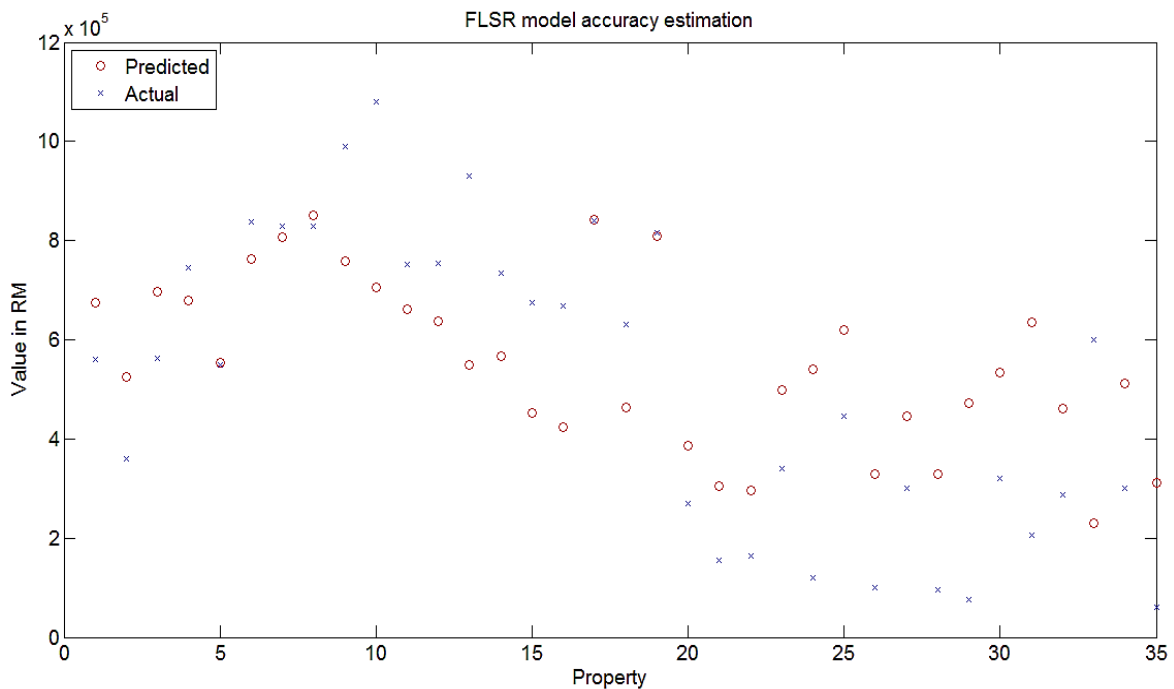


Figure 6: A Fuzzy Regression model

The results are then compared with ANN and FIS models, in respect to the Mean Absolute Error (MAE) through the application of the same data set to the different models. Figure 7 shows that a considerable reduction in MAE is achieved in FIS and FLSR where the error rate drops by more than 25000 compared to the MAE of ANN model. The maximum reduction is achieved in FLSR which reveals that FLSR model outperforms the other considered models.

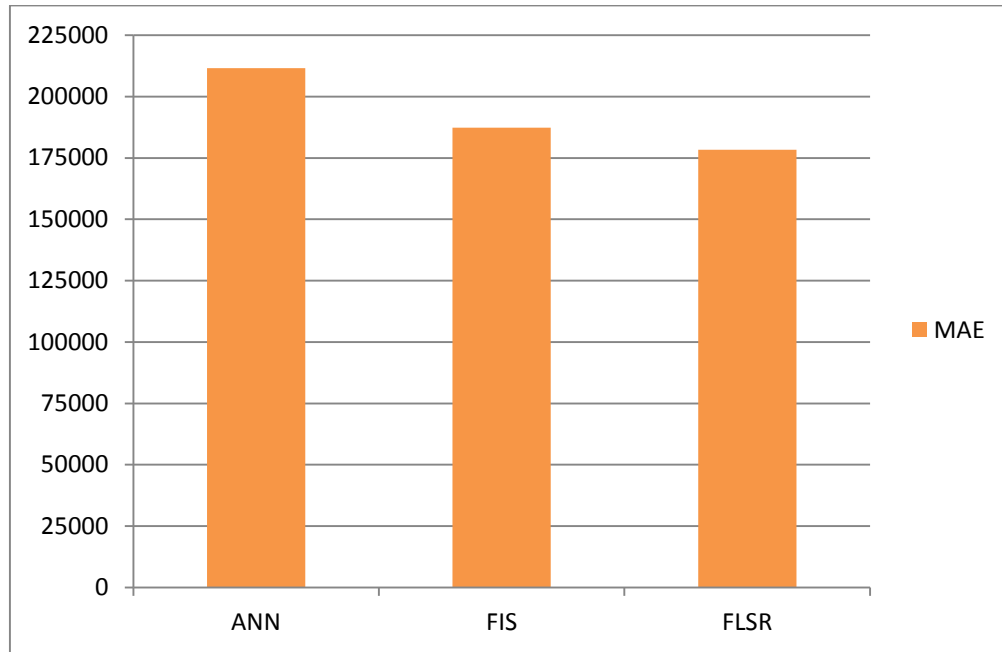


Figure 7: Results comparison

4.0 CONCLUSION

This paper attempts to show a study that has explored the performance of nonlinear models in predicting house prices for a selected Malaysian urban area. The main aim of the study was to investigate the possibility to adopt FL model and assess its performance in predicting house prices. The lower MAE for all three models i.e. FL, ANN and FIS have indicated that the values can be accepted. However, the lower values can be attributed to the distribution of the input samples and lack of structural information in Malaysian datasets. The experimental evaluation supports the earlier claim mentioned in this study that the FL-based approach to address the problem of predicting real estate prices is more accurate in modeling the relationship between the predictors (property features) and the dependent variable (property price) than other approaches, i.e. ANN, ANFIS. In order to obtain better prediction models, further study is required with the use of more sales data that incorporates the social and financial aspects into the nonlinear models.

5.0 REFERENCES

- Borah, K.K., Souvnik, R., & Harinarayana, T. (2013). Optimisation in site selection of wind turbine for energy using fuzzy logic system and GIS-A case study of Gujarat. *Open Journal of Optimization*, 2, 116-122.
- Brunson, A., Buttimer, R. J., & Rutherford, R. (1994). *Neural networks, nonlinear specifications, and industrial property values*. Working Paper Series, University of Texas at Arlington.
- Csepedia.com (2014). *Encyclopedia of computer and internet*. Retrieved August 14, 2015 from, www.facebook.com/Csepedia
- Daniels, H., & Kamp, B. (1999). Application of MLP networks to bond rating and house pricing. *Neural Computing & Application*, 8, 226-234.

- Do, Q., & Grudnitski, G. (1992). A neural network approach to residential property appraisal. *The Real Estate Appraiser*, 58(3), 38-45.
- Graaskamp, J.A. (1981). *Fundamentals of real estate development (Development component series)*. US: Urban Land Institute.
- Lawson, J.W.W. (2008). *Theory of real estate valuation*. (Unpublished Masters thesis). RMIT University, Australia.
- Mathworks (2014). www.mathworks.com. *Defining fuzzy sets*. Retrieved from <http://www.calvin.edu/~pribeiro/othrlnks/Fuzzy/fuzzysets.htm>, 18/1/2014.
- Pagourtzi, E., Assimakopoulos, V., Hatzichristos, T., & French, N. (2003). Real estate appraisal: A review of valuation methods. *Journal of Property Investment & Finance*, 21(4), 383 – 401.
- Rubens, N.O. (2006). The application of fuzzy logic to the construction of the ranking function of information retrieval systems. *Computer Modelling and New Technologies*, 10(1), 20-27.
- Sarip, A.G. (2006). Geo-information neural system for single-property valuation. *Journal of Valuation and Property Services*, 6(1), 41-57.
- Singh, K., & Singh, I. A. (2013). Synergistic suitability of transfusion of TQM-TPM for Indian manufacturing industries using fuzzy-based model simulation. *International Journal of Business Continuity and Risk Management*, 4(1), 36-53.
- Steele G.L. (1990). *CommonLisp: The language*. US: Thinking Machines Inc. Digital Press.
- Sun, B., & Qiu, Y. (2010) *Fuzzy expert system for flexible pavements crack performance prediction*. Proceedings of the 2010 International Conference of Logistics Engineering and Management (ICLEM), held in Chengdu, China, October 8-10, 2010: 2415-2421.
- www.cs.bilkent.edu.tr (April 8, 2010). *A short fuzzy logic tutorial*. Retrieved from http://cs.bilkent.edu.tr/~zeynep/files/short_fuzzy_logic_tutorial.pdf
- Yuan, B., & Herbert, J. (2012). Fuzzy CARA - A fuzzy-based context reasoning system for pervasive healthcare. *Procedia Computer Science*, 10, 357–365.
- Zadeh, L. (1965). Fuzzy sets. *Information and Control*, 8, 338–353.
- Zurada, J. M., Levitan, A. S., & Jian, G. (2006). Non-conventional approaches to property value assessment. *Journal of Applied Business Research*, 22(3), 1-14.