Proceedings of the Fourth International Conference on Mathematical and Computational Applications
June 11-13, 2013. Manisa, Turkey, pp.293-301

COMPARISON OF SVR AND MRA METHODS IN REAL ESTATE VALUATION

Şükran Yalpir[1], Gülay Tezel[2] and Fatma Bünyan Ünel[3]

[1]Selcuk University, Computer Engineering, 42031, Konya, Turkey
[2]Selcuk University, Geomatics Engineering Department, 42031, Konya, Turkey
[3]Selcuk University,Guneysinir Vocational School of Higher Education, Map and Cadastro Programme, 42190, Guneysinir, Konya, Turkey
sarici@selcuk.edu.tr, gtezel@selcuk.edu.tr, fbunyan@selcuk.edu.tr

Abstract- The importance of real estate valuation increases in recent years. Real estate valuation, which is important for countries economy, is quite difficult. There is a lot of existing studies in literature to find proper approach in valuation. In these studies, different criteria are used according to application areas. Success levels show variations according to the used criteria and methods.

In this study, data set with physical and spatial criteria belong to 142 houses in Konya Karatay region, Dogus district was created by a questionnaire. This data set consists of 8 input variables (area of house, age, side, heating type, number of rooms, total floor quantity, floor situation of house and position index of houses) and an output variable (real estates market value). Support Vector Regression (SVR) and traditional Multiple Regression Analysis (MRA) methods with created data set were applied and the most successful result was reached via SVR. Therefore, R^2, AA%, RMSE, MSE, SD and ROC curve were selected as the success evaluation parameters.

Key Words- Real estate valuation, Support Vector Regression, Multiple Regression Analysis.

1. INTRODUCTION

As real estate valuation is one of the important factors in country's economics, today its popularity is increasing. Real estate valuation treatment is used in a lot of area such as taxing, hypothec treatments in banks, socialization, buying selling. In these areas that real estate valuation is needed, different approaches and estimation of the value is done. For this reason, each application results show differences for that real estate.

As valuation is a response that holds objective and subjective approaches, it is difficult to get real value. In the international valuation standards 2005 (IVS), issued by the International Valuation Standards Committee (IVSC), the market value is defined as the estimated amount of money for which a property should exchange on the date of valuation between a willing buyer and a willing seller in arm's-length transaction after proper marketing wherein the parties acted knowledgeably, prudently and without compulsion [1].

The influence factors of the real estate value are complex and changed according to regional conditions. Also the number of these factors was varying both regional and physical conditions. For this reason, it was important to determine an optimum number

of real estate valuation criteria and which criteria were the most effective according to the conditions of the region. These factors must be considered both the characteristic of the real estate market and the influence factors in the appraisal of the real estate.

In real estate valuation there is a lot of techniques in the literature. Among these techniques, Multi Regression Analysis (MRA), which is used nowadays along with comparison, income and cost techniques, is not enough. Mass appraisal models are commonly based on the sales comparison approach. MRA-based methods have been popular because of their established methodology, long history of application, and wide acceptance among both practitioners and academicians. The limitations of traditional linear MRA for assessing value of real estate property have been recognized for some time [2-3].

Recently, new searches come out with the effect of the fast developing computer technology. Artificial Intelligence (AI) based methods have been particularly proposed as an alternative instead of MRA for real estate valuation [4-10]. There are different methods based on AI techniques and heretic approaches applied in real estate valuation. Some of them are artificial neural networks (ANN) [11], and fuzzy logic [12], which are the member of AI, rough set (RS) theory, heretic approach [13] etc. At the same time, SVM [14] and SVR [15] algorithms based on the machine learning optimization are preferred in real estate valuation.

In this study, a data set was created about 142 houses market value in Konya Karatay region, Dogus district and belonging criteria of them. SVR was applied to this data set and the results were compared with MRA. The area of house, age, side, heating type, the number of rooms, total floor quantity, floor situation of house and index of house were selected as input variables and market value of house was assigned as output variable to realize SVR and MRA methods. Results of methods were compared statistically as model performances.

2. MATERIAL-METHOD

2.1. Support Vector Regression (SVR)

Support vector machines (SVMs) as a member of a family of generalized linear classifiers are a set of related supervised learning methods. In another terms, Support Vector Machine (SVM) is a classification and regression estimation tool that uses machine learning theory to maximize estimation accuracy while automatically avoiding over-fit to the data. Support Vector machines can be defined as systems which use hypothesis space of a linear functions in a high dimensional feature space, trained with a learning algorithm from optimization theory that implements a learning bias derived from statistical learning theory. The foundations of Support Vector Machines (SVM) have been developed by Vapnik [16-17]. SVMs were developed to solve the classification problem, but recently they have been extended to solve regression problems by the introduction of an alternative loss function [18].

The loss function must be modified to include a distance measure. In SVR, the regression can be linear and non linear. Linear models mainly consist of the following loss functions, e- intensive loss functions, quadratic and Huber loss function. Similarly to classification problems, a non-linear model is usually required to adequately model data. In the same manner as the non-linear SVR approach, a non-linear mapping can be

used to map the data into a high dimensional feature space where linear regression is performed. The kernel approach is again employed to address the curse of dimensionality. Dibike et al. (2001) [19] presented some results showing that Radial Basis Function (RBF) is the best kernel function to be used in SVR models [20]. In this study, Radial Basis Function (RBF) function was used to estimate market value for SVR.

2.2. Statistical Analysis of the Models' Performances

In order to determine the models' performances comparatively, the average approximation (AA) ratio (equation (1)), standard deviation (SD in equation 2), standard deviation percentage (equation 3), the root mean square error (RMSE in equation 4) and the coefficient of determination (R^2 in equation 5) were calculated.

$$AA\% = \frac{\sum_{i=1}^{n}\left(1 - \frac{|x_p - x_i|}{x_p}\right)}{n} * 100 \quad (1)$$

$$SD = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x_i})^2}{(n-1)}} \quad (2)$$

$$SD\% = 100 * \frac{SD_{\text{mod el}}}{SD_{output}} \quad (3)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_p - x_i)^2}{n}} \quad (4)$$

$$R^2 = 1 - \frac{\sum (x_p - x_i)^2}{\sum_{i}^{n}(x_i - \overline{x})^2} \quad (5)$$

x_p :Market value,
x_i :Value of the model,
i : {1,2,3….n},
n :Total number of the residential real-estates in the data set,
SD : Standard deviation of the approximations in the data set,
AA_i : Approximation value.

AA% shows that in what percentage the models predicting value closer to unit market value. The higher the AA%, the more successful the model is. The RMSE value represents the predicting error such that the model with less RMSE value is more successful in prediction than the model with higher RMSE. As SD is a statistical

measure of spread or variability, it is data-dependent and there is no criterion for it. SD% is the percentage response of model's SD which determines model SD is how close to output SD. R^2 is the power measurement the dependent variable of the equation obtained from the regression analysis

3. APPLICATION

In application, study area is decided as Konya city Karatay region Doğuş district (Figure 1-2). Doğuş district is of 800000.00m^2 area and its population is 8200. Total house quantity is 1671 and there are 802 lands in total. First development plan is done in 1972 and second development plan is applied in 1998. Now it is within the area of Urban Transformation Project. This district develops day by day and generally consists of middle class people. Buildings are generally multi floored and new but when we move to inner parts there are more single floor buildings.

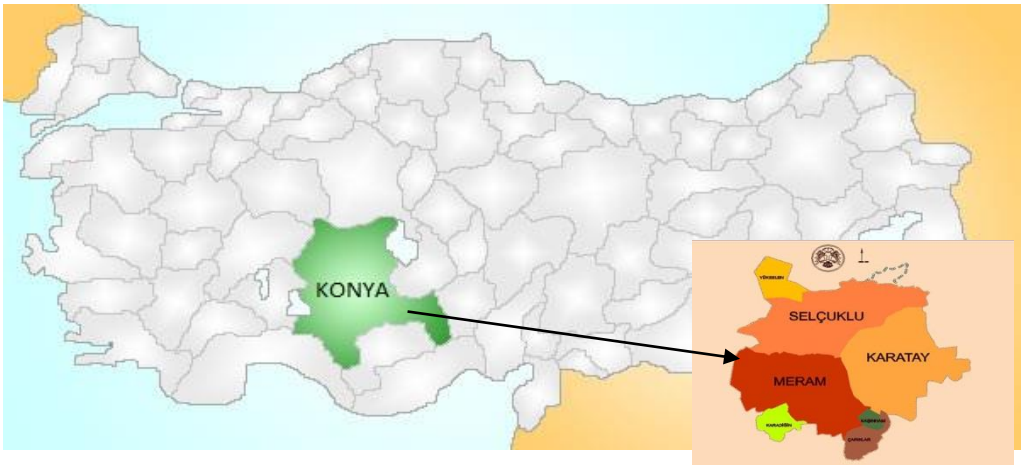

Figure1. General view of study area, Karatay

Figure2. Satellite view of Doğuş district

Seven criterias are determined in the area which effects the values of real estates and area study is done. These chosen criterias are about the houses' ;

- Residence area (usage area): It is the total residence area which is found by ignoring unused areas in the real estates main structure.
- Age: the year which passes starting from the completion of construction and getting residence license.
- Side: position of the building is determined according to the sun and north and west sides as 1 and south and east sides as 2 are assessed.
- Heating type: heating type is valued as 1, 2 and 3, according to coal, personal heating boiler and central heating system, respectively.
- Room quantity: it is the total room quantity of house
- Total floor quantity: total floor quantity of the building that the house is in.
- Floor situation of house: if the house in low floor 1, if it is on top floor 2, if it is in middle floor 3.
- Position index: This index is created by considering avenue, street and areas around the house.
- Market value: it is the TL equivalent of the real estate which is subject to buying-selling in market conditions.

Data set gaps are seen in Table1.

Table1. Maximum, Average, Minimum and standard deviation values of the housing criteria

NO	**Criterias**	**max**	**min**	**mean**	**SD**
X1	Residence Area/100	1,75	0,8	1,36	0,19
X2	Age	26	0	3,25	4,53
X3	Side	2	1	1,80	0,39
X4	Heating Type	4	1	2	0,85
X5	Room quantity	4	3	3,78	1,42
X6	Total floor quantity	8	1	4,17	1,37
X7	Floor situation of house	3	1	2,40	0,81
X8	Position index	8,75	0,09	3,06	2,70
Y	**Market Value /10000**	**14**	**5**	**9,69**	**2,12**

The data set was created with 8 criteria belonging 142 houses using questionnaire. After data set regulation, SVR and MRA, widely used in literature were applied using Orange program and SPSS program, respectively. It was seen that the most successful kernel function is RBF as a result of experiments with different kernel functions (linear, polynomial, radial basis function (RBF) and sigmoid) and different values of each parameters for SVR.

The achievement of SVR was compared with linear MRA using same data set. The produced (obtained) MRA model and its coefficients were presented in Equation (6). In the MRA method, the dependent variable (market value) is assumed to be a linear function of independent variables (residence area, age, heating type etc..) with added error introduced to account for all other factors.

$$\mathbf{Y}= -0{,}76708+0{,}854791*\mathbf{X1} -0{,}17716*\mathbf{X2} +0{,}999183*\mathbf{X3} -0{,}1859*\mathbf{X4} +2{,}11049*\mathbf{X5}-0{,}01236*\mathbf{X6}+0{,}125978*\mathbf{X7} +0{,}063622*\mathbf{X8} \quad (6)$$

4. RESULTS AND DISCUSSION

In this study, the estimated values by SVR and MRA methods were compared with the market values. Figure 3 a and b show the best curve fitting with zero intercept (y=ax line) used to define the level of accuracy between the collected market values and the models. The results for SVR and MRA mean how the model makes a prediction about new data. Both SVR and MRA models the slope of the best fits is close to 1 in data set whereas it is understood that MRA does not predict well the market value in the distribution point.

According to the distribution point of estimated values and market values, MRA model indicates the worse performance than SVR in term of slope and R^2, whereas, R^2 value of MRA has a very low. (Figure 3).

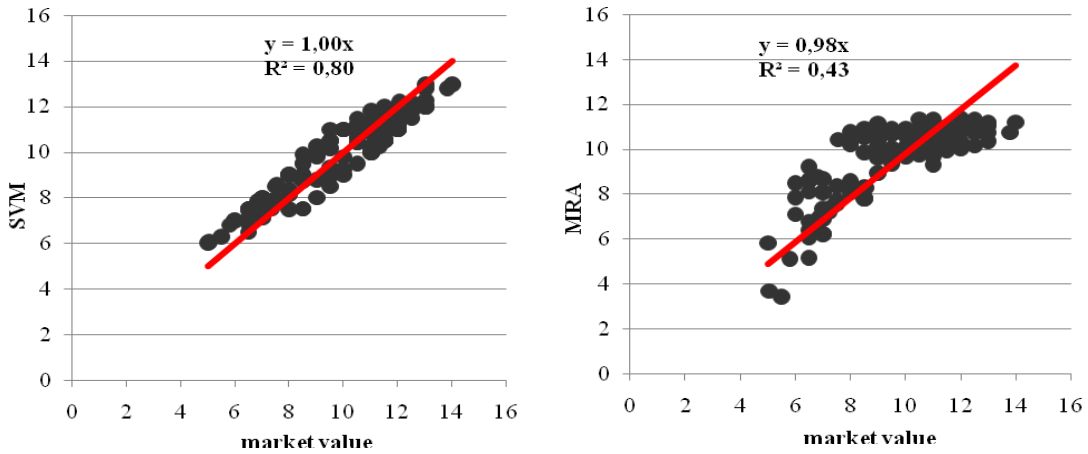

Figure 3. The regression line between the unit market values and SVR, MRA values (a) SVR, (b) MRA.

It was taught that graphical demonstration was insufficient to show the performance of models. The slope and R^2 of regression fits give the idea about the performance of the models but the suggestion of the most appropriate model according to slope and R^2 is impossible and the discussion depending upon them has poor certainty. Therefore, statistical calculations (i.e., AA%, RMSE, SD and SD%) were also performed and these results were showed in Table 2. The achievement and predicting error value (RMSE) are inversely proportional. Obtaining less RMSE with model

indicates that the performance is higher. RMSE value obtained with SVR is less then the RMSE value of MRA.

In addition, AA% shows that in what percentage the models predicting value closer to unit market value. The model with higher the AA%, is the more successful. As SD is a statistical measure of spread or variability, it is data-dependent and there is no criterion for it. In this study, the approximity of the model SD to real data SD is considered rather than the smaller SD. The degree of approximity is quantified by SD percentage. The SD value for market value was calculated as 2,12 (Table 1).

Consequently, when the estimated value and market value was compared, getting less RMSE value and higher %AA value with SVR model than MRA were an important indicator for an achievement of SVR (Table2). In statistical assessment, it was proved in all analysis that SVM method was more successful than MRA method.

Finally, the Receiver Operating Characteristic curve (ROC curve) on accuracy data was applied in this study. Measurement of SVR success defined with the area under the ROC curve was obtained as 0.9965. The area under the curve shows the superiority of one to another. If result is 1, it is perfect estimation; if result is 0.5, at that time it is not better than random model (SAS Institute Inc., 2010).

Table 2. The comparison of real-estate valuation models

	SVR	**MRA**
R^2	0,80	0,43
y=a*x	1,001*x	0,9815*x
AA (%)	92,48	88,85
RMSE	0,766	1,278
SD	1,719	1,680
SD%	81,10	79,26

5. CONCLUSIONS

Differences of country economical conditions, changes in human needs sourced by different life styles, socio-cultural structure and a lot of factor makes it difficult to create a mathematical model, to get a single real result and to use of common criteria in real estate valuation. For this reason, it is seen that there is an individual and different criteria and new approaches are used in each study. The selection of right approaches and criteria vary according to study areas. Also the sufficiency of the approaches according to usage area can be discussed.

In the study, area of the house, age, side, heating type, room quantity, total floor quantity, floor position of house and position index were utilized as input variables and market value of real estate was used as the output variable. The success of SVR method in estimating market value of real estates was investigated with these input and output variables. When R^2, the best curve and statistical analysis were compared, the market value and predicted value with SVR were closer than MRA. As a result, it was seen that SVR method will be enough in estimating market value [21].

6. REFERENCES

1 International Valuation Standards Seventh Edition, Appraisal Inst, 2005.
2. A. Q. Do, G. Grudnitski, A Neural Network Approach to Residential Property Appraisal, The Real Estate Appraiser, 58:3, 38–45, 1992.
3. J. Mark, and M. Goldberg, Multiple Regression Analysis and Mass Assessment: A Review of the Issues, Appraisal Journal, 56:1, 89-109, 1988.
4. J.Guan, and A. S. Levitan, Artificial Neural Network Based Assessment of Residential Real Estate Property Prices: A Case Study, Accounting Forum, 20:3/4, 311-26, 1997.
5. J.Guan, J. Zurada and A. S. Levitan, An Adaptive Neuro-Fuzzy Inference System Based Approach to Real Estate Property Assessment, Journal of Real Estate Research, 30:4, 395-420, 2008.
6. D.Krol, , T. Lasota, W. Nalepa, B. Trawinski, Fuzzy System Model to Assist with Real Estate Appraisals, Lecture Notes in Computer Science, 4570, 260-69, 2007.
7. S. McGreal, A. Adair, D. McBurney and D. Patterson, Neural Networks: The Prediction of Residential Values, Journal of Property Valuation and Investment, 16, 57-70, 1998.
8. S. Peterson, A. B. Flanagan, Neural Network Hedonic Pricing Models in Mass Real Estate Appraisal, Journal of Real Estate Research, 31:2, 147-64, 2009.
9. W. Z. Taffese, Case-Based Reasoning and Neural Networks for Real Estate Valuation, Proceedings of the 25th Conference on Proceedings of the 25th IASTED International Multi-Conference: Artificial Intelligence and Applications, Innsbruck, Austria, 84-89, 2007.
10. E. Worzala, M. Lenk and A. Silva, An Exploration of Neural Networks and Its Application to Real Estate Valuation, Journal of Real Estate Research, 10, 185-202, 1995.
11. G. Özkan, Ş. Yalpır, O. Oygunol, An Investigation on The Price Estimation of Residable Real-Estates By Using Artificial Neural Network and Regression Methods, XIIth Applied Stochastic Models and Data Analysis International conference(ASMDA), Chania, Crete, Greece, 2007.
12. Ş. Yalpır, G. Özkan, The Usage of Artificial Intelligence in Determining The Residential Real-Estate Prices in Urban Areas and The Comparison of Valuation Methods, FIG,2008.
13. M. d'Amato, Rough Set Theory as Property Valuation Methodology: Whole Story, Mass Appraisal Methods: An International Perspective for Property Valuers, ISBN: 978-1-405-18097-9. 220-258, 2008.
14. J. Zurada, A. S. Levitan, J. Guan, A Comparison of Regression and Artificial Intelligence Methods in a Mass Appraisal Context, journal of Real Estate Research. 33(3), 349-388, 2011.
15. H. Y. LIN, K. CHEN, Soft computing algorithms in Price of Taiwan Real Estates, WSEAS Transactions on Systems, ISSN: 1109-2777, 10(10), 2011.
16. V. N. Vapnik, The Nature of Statistical Learning Theory, NY: Springer-Verlag. 1995.
17. V. N. Vapnik, Statistical Learning Theory. New York: John Wiley and Sons, 1998.

18. V. Vapnik, S. Golowich, and A. Smola. Support vector method for function approximation, regression estimation, and signal processing. In M. Mozer, M. Jordan, and T. Petsche, editors, Advances in Neural Information Processing Systems 9, 281–287, Cambridge, MA, MIT Press, 1997.
19. Y.B. Dibike, S.Velickov, D.P Sololatine, M.B. Abbott, Model induction with support vecture machine: Introdaction and application. ASCE Journal of Computing in Civil Engineering. 15, 208-216, 2001.
20. V. Jakkula, Tutorial on Support Vector Machine (SVM), School of EECS, Washington State University, Pullman 99164.
21. SAS Institute Inc.,2010. JMP_ 9 modeling and multivariate methods. pp 272–1135 281, Cary, NC: *SAS Institute Inc.*