Comparison of Expert Algorithms with Machine Learning Models for Real Estate Appraisal

Bogdan Trawiński, Zbigniew Telec, Jacek Krasnoborski, Mateusz Piwowarczyk, Michał Talaga
Faculty of Computer Science and Management
Wrocław University of Science and Technology
Wrocław, Poland
{bogdan.trawinski, zbigniew.telec}@pwr.edu.pl

Tedeusz Lasota, Edward Sawiłow
Department of Spatial Management
Wrocław University of Environmental and Life Sciences
Wrocław, Poland
{tadeusz.lasota, edward.sawilow}@up.wroc.pl

***Abstract*—Machine learning models require numerous training examples to provide reliable predictions of real estate prices. Expert algorithms could be applied wherever only several training samples are available. The accuracy of two expert algorithms based on the sales comparison approach was experimentally examined using real-world data derived from a cadastral system and registry of real estate transactions. The performance of the algorithms was compared with three data driven regression models for property valuation. Statistical analysis of the obtained results was conducted.**

Keywords—machine learning, regression models, expert property valuation algorithms, real estate appraisal

I. INTRODUCTION

Real estate appraisers are licensed professionals who estimate values of properties using standardized procedures, their expert knowledge of the real estate market, and appropriate amount of complete and accurate data [1], [2]. Their work is more often supported by computer systems called Automated Valuation Models (*AVM*) and Computer Assisted Mass Appraisal (*CAMA*) which employ various methods and models for single and mass appraisal of properties [3], [4], [5]. A broad spectrum of approaches to property valuation have been developed for many years. They ranged from multiple regression [6] through neural networks [7] and decision trees [8] to fuzzy systems [9], and hybrid approaches [10, 11].

Various cutting edge machine learning techniques as ensemble methods [12], [13], [14], hybrid methods including evolutionary fuzzy systems [15] and evolving fuzzy systems [16] were explored by the authors and utilized to construct valuation models. We worked out methods and algorithms for building data driven regression models based on ensembles of genetic fuzzy systems [17], [18], evolving fuzzy models [19], [20]. We applied ensembles of genetic fuzzy systems and neural networks to predict from a data stream of real estate sales transactions [21], [22]. We worked also on the preparation of data, namely on merging different areas of an urban municipality into uniform zones to obtain homogenous areas comprising greater number of samples to build more reliable valuation models [23], [24].

In this paper we propose two expert algorithms designed for a computerized system to aid in real estate appraisal. Both algorithms are based on the sales comparison approach and differ in the way of selecting similar properties to calculate an estimated price. The first one takes into account a number of the latest transactions of similar properties within a given market area. The second one, in turn, computes the estimated price as an average of values of a number of the nearest properties which belong to the same similarity classes. The performance of the algorithms was experimentally investigated using real-world data derived from a cadastral system and registry of real estate transactions. Moreover, the accuracy of both algorithms was compared with three data driven regression models for property valuation created in the *WEKA* data mining system [25].

II. EXPERT ALGORITHMS FOR REAL ESTATE APPRAISAL

Two expert algorithms for property valuation are proposed in the paper. Both are based on sales comparison approach which assumes that only similar properties to the appraised one are considered to predict its price. In order to obtain an estimated price of a given property an average price of a number of similar properties is calculated. The experts chose a few key features of a real estate to select a subset of similar properties. Then, they partitioned the values of key features by range. The ranges for each feature were contiguous but not overlapping and therefore they could determine classes of similarity. The list of similarity classes defined for residential properties located in a given Polish municipality is shown in Table I. The list was worked out based on the dataset of sales transactions available and four key features were taken into account: usable area of a flat (*Area*), year of a building construction (*Year*), number of storeys in a building (*Storeys*), number of rooms in a flat including a kitchen (*Rooms*). Finally, the following similarity criterion (*SC*) can be formulated: two properties are similar if the values of all their key features fall to the same similarity classes.

The first expert algorithm is called ***N-Latest Transactions in an Area*** – *LTA*. It estimates the predicted price as an average of values of the N-latest transactions of similar properties within the same area. The pseudocode of the *LTA* algorithm is given in Table II.

TABLE I. CITERIA FOR SIMILARITY OF RESIDENTIAL PROPERTIES

Feature	Class	Values
Area	1-small	under 40 m2
	2-medium	40-60 m2
	3-large	over 60 m2
Year	1-very old	before 1918
	2-old	1919-1945
	3-medium	1946-1975
	4-new	1976-1995
	5-newest	after 1995
Storeys	1-low-rise	1-2
	2-mid-rise	3-5
	3-high-rise	more than 5
Rooms	1-small	1-2
	2-medium	3-4
	3-large	more than 4

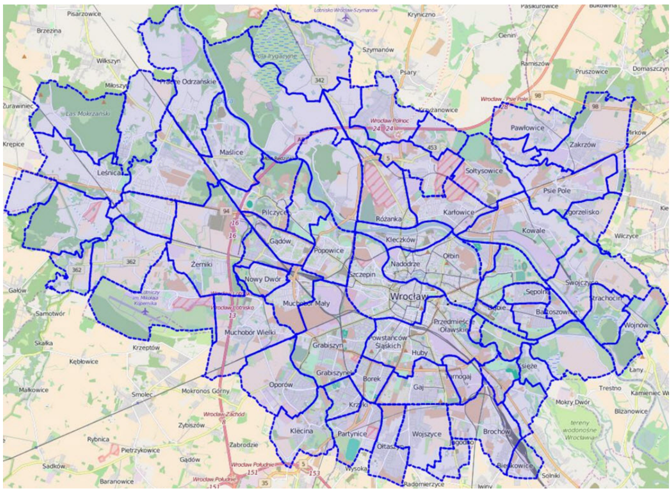

Fig. 1. Cadastral regions of an urban municipality as market areas

TABLE II. EXPERT ALGORITHM LTA: N-LATEST TRANSACTIONS IN AN AREA

Step	Action
1	Compute the average price per square metre for each region R_j: $AvgP_j$ *for j=1,2,..,NR*
2	Take next property X_i to estimate
3	Set the number *N* of properties needed to compute the estimate
4	Select all properties similar to X_i according to the similarity criterion *SC*
5	If the number of properties similar to X is less than *N*, then skip X_i and goto Step 2
6	Determine the region R_i where the appraised property is located
7	If the number of properties similar to X_i is greater or equal to *N*, then take the *N* most recent transactions and goto Step 11
8	Repeat Steps 9-10 until *N* similar transactions are selected
9	Take the next region R_k where the difference $\|avgP_k-avgP_i\|$ is the smallest
10	Take the *m* most recent transactions from R_k to complete the required number *N*
11	Compute the estimated price of X_i as the average price per square metre of *N* selected transactions

TABLE III. EXPERT ALGORITHM ENS: N-NEAREST SIMILAR PROPERTIES

Step	Action
1	Take next property X_i to estimate
2	Set the number *N* of properties needed to compute the estimate
3	Select all properties similar to X_i according to SC criteria
4	If the number of properties similar to X_i is less than *N*, then skip X_i and goto Step 1
5	Sort the similar properties by distance from X_i in ascending order
6	Take the *N* nearest similar properties
7	Compute the estimated price of X_i as the average price per square metre of *N* selected properties

The second expert algorithm is named the ***N-Nearest Similar Properties*** – *NSP*. It calculates the estimated price as an average of values of N-nearest properties in terms of the Euclidean distance falling to the same similarity classes as the appraised one does. The pseudocode of the *NSP* algorithm was placed in Table III. The area considered in the paper is a cadastral region of an urban municipality. Cadastral regions are delineated by the boundaries of residential quarters and housing estates as well as by main roads, rivers, railway lines, parks, forests, etc. In consequence, the cadastral regions could be regarded as areas with uniform real estate price levels. The area of an urban municipality partitioned into 69 cadastral regions is depicted in Fig. 1.

III. SETUP OF EVALUATION EXPERIMENT

The main aim of evaluation experiments was to examine the performance of expert algorithms compared to machine learning models created over the selected areas in an urban municipality. The evaluation experiment was done using real-world data about sales and purchase transactions acquired from a cadastral system and a public registry of real estate transactions maintained by an urban municipality. The dataset after cleansing comprised 12,439 records of sales transactions of residential premises accomplished during 16 years from 1998 to 2013. All prices were updated for the last day of the analyzed period based on a market trend of property price change. The test set was randomly drawn from the whole dataset using the uniform distribution and counted 3,731 transaction records, i.e. 30 percent of all transactions. The remaining 70 per cent of records constituted the dataset to determine similar properties to an appraised one and used to calculate its estimated price. This 70% dataset was used also as a training set for machine learning algorithms.

During each run of *LTA* and *NPS* algorithms 100 properties randomly drawn from the test set were valuated. The number of properties needed to compute the estimate was set to *N=11*. The accuracy measure *MAPE* was calculated for each run using Formula 1, where P_i^a and P_i^p denote the actual and predicted prices respectively and *n* stands for the number of transactions valuated in each run. Both algorithms were executed 150 times and produced 150 values of *MAPE* to enable us to compare the accuracy of both algorithms using statistical significance tests.

$$MAPE = \frac{1}{n} * \sum_{i=1}^{n} \left| \frac{P_i^p - P_i^a}{P_i^a} \right| * 100\% \quad (1)$$

The procedure of computing *MAPE* values for machine learning models (*ML* models) was designed to be similar to the experts algorithms as much as possible. The same 70% training set and 30% test sets were used in this part of the experiment. Exactly the same *150x100* properties as the ones drawn for *LTA* and *NPS* algorithms were applied to test the accuracy of machine learning models. The training set was divided into 69 subsets comprising records of sales transactions made within individual cadastral regions. Machine learning models were built only for those cadastral regions which counted at least 50 transactions. This cardinality of training sets might ensure that machine learning models could achieve an acceptable level of reliability.

Three following types of predictive models were built for each region over training set using machine learning algorithms included in the *WEKA* data mining software suite [25]:

M5P – Pruned Model Tree applies procedures for building *M5* model trees. The algorithm employs decision trees, however, instead of having values at tree's nodes, it comprises a multivariate linear regression model at each node.

MLR – Multilayer Perceptron is a feed-forward neural network employing the backpropagation learning algorithm. For regression problems, i.e. when classes are numeric, the output nodes become unthresholded linear units.

LR – Linear Regression is a standard statistical procedure to create a linear model which uses the least mean square method to adapt the parameters of the linear function.

Each algorithm used four input attributes selected by the experts as the key features of a real estate and available in the dataset employed to examine *LTA* and *NSP* algorithms, i.e.: *Area, Year, Storeys*, and *Rooms*. In turn, price per square metre (*Price*) constituted the output variable.

The main steps of the procedure to evaluate machine learning models are presented in Table IV.

TABLE IV. PROCEDURE TO EVALUATE MACHINE LEARNING MODELS

Step	Action
1	Take next property X_i to estimate
2	Determine the region R_i where the appraised property is located
3	If *ML* models were not generated for the region R_i, then skip X_i and goto Step 1
4	Compute the predicted prices of X_i as the output of *M5P*, *MLP*, and *LR* models built for the region R_i

TABLE V. AVERAGE RANK POSITIONS OF MODEL PERFORMANCE IN TERMS OF MAPE PRODUCED BY THE FRIEDMAN TEST

	1st	2nd	3rd	4th	5th
Algorithm	M5P	MLP	LR	LTA	NSP
Avg rank	1.17	2.58	3.01	3.33	4.92

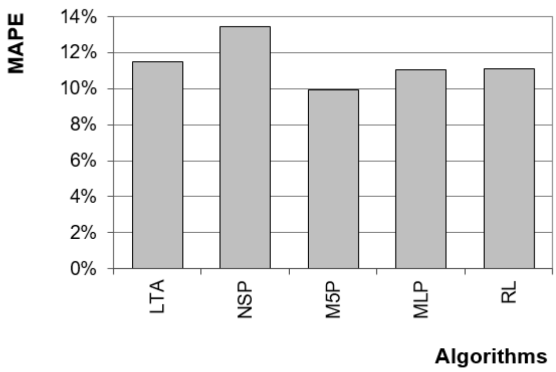

Fig. 2. Median of MAPE provided by expert algorithms and machine learning models for 150 sets of test residential premises

IV. ANALYSIS OF EXPERIMENTAL RESULTS

The *LTA* and *NSP* expert algorithms as well as *M5P*, *MLP*, and *LR* machine learning models provided 150 values of *MAPE* which were used as points of observation, data samples in statistical tests. The median of *MAPE* values obtained for individual algorithms and models is presented in Fig. 1. It is clearly seen that *M5P*, *MLP*, and *LR* models provided lower values of *MAPE*, i.e. 10.0%, 11.1%, 11.1% respectively, than *LTA* and *NSP* algorithms, i.e. 11.5% and 13.4% respectively.

The Kolmogorov-Smirnov and Shapiro-Wilk tests were employed to examine the assumption that data samples are normally distributed. Both tests revealed that the results produced by the *NSP* expert algorithm and all *ML* models are not from a normally distributed population. Only in the case of the *LTA* algorithm the null hypothesis that the data are normal was not rejected. In consequence, the nonparametric Friedman and Wilcoxon tests were used to analyze the experimental results. The Friedman test done in respect of *MSE* values for 150 result sets indicated significant differences between some models. Average rank positions of model performance in terms of *MAPE* produced by the Friedman test are presented in Table V, where the lower rank value the better model. In turn, the results of nonparametric Wilcoxon signed-rank test to pairwise comparison of the model performance are placed in Table VI. The zero hypothesis stated there were not significant differences in accuracy, in terms of *MAPE*, between individual pairs of models. In Table VI + denotes that the model in the row performed significantly better than, – significantly worse than the one in the corresponding column. The significance level of 0.05 was set for the null hypothesis rejection. The results indicate that there are statistically significant differences in accuracy between each pair of models and algorithms. The average rank in Table V shows that *M5P* is the most accurate.

TABLE VI. RESULTS OF THE WILCOXON TEST FOR PAIRWISE COMPARISON OF MODEL ACCURACY

	M5P	MLP	LR	LTA	NSP
M5P		+	+	+	+
MLP	–		+	+	+
LR	–	–		+	+
LTA	–	–	–		+
NSP	–	–	–	–	

V. Conclusions and Future Work

Two expert algorithms, based on the sales comparison approach, designed for a computerized system to assist with property valuation were proposed. The first algorithm processed a predefined number of the latest transactions of similar properties within the same cadastral region in which an appraised property was located. The second one, in turn, took a predefined number of the nearest similar properties to a valuated property to estimate its price.

The algorithms were experimentally compared in terms of accuracy with three machine learning models using real-world data. Mean absolute percentage error was employed as the performance measure and the statistical nonparametric Friedman and Wilcoxon tests were applied to analyze the results.

Following main conclusions can be drawn from our study. Statistically significant differences in accuracy occurred between each pair of models and algorithms. Machine learning models surpassed the expert algorithms. From among machine learning models pruned model trees – *M5P* revealed the best performance and linear regression – *LR* showed the least accuracy. The *LTA* expert algorithm outperformed the *NSP* ones. This is an interesting result because the routine professional appraisers' procedures try to find a number of similar properties located in the shortest distance from the appraised one. It turned out that selecting the latest transactions from among similar properties within a given market area provides better results. Nevertheless, all considered models and algorithms could be employed in a computerized system to support professional appraisers because the differences in accuracy among them were not high: the median of *MAPE* ranged from 10% to 13.4%. Moreover, the machine learning models could be built only for those market areas where an appropriate minimum number of training samples is available, e.g. 50. The expert algorithms, in turn, have not these limitations.

Further investigation is planned to explore the performance of expert algorithms depending on the number of similar properties took to calculate the estimated price of a property. Different zones within an urban municipality as market areas will be examined. Moreover, a few more automated valuation models, including ensemble learning ones, will be employed for the comparative analysis.

References

[1] E. Dudek-Dyduch and K. Rączka, "Information system for real estate valuation," Automatyka, vol. 15, no. 3, pp. 637-652, 2011.

[2] A. Barańska, "Real estate mass appraisal in selected countries – functioning systems and proposed solutions," Real Estate Management and Valuation, vol. 21, no. 3, pp. 35-42, 2013.

[3] E. Jahanshiri, T. Buyong, and A.R.M. Shariff, "A review of property mass valuation models," Pertanika J. Sci. & Technol., vol. 19 (S), pp. 23-30, 2011.

[4] W.J. McCluskey, M. McCord, P.T. Davis, M. Haran, and D. McIlhatton, "Prediction accuracy in mass appraisal: a comparison of modern approaches," Journal of Property Research, vol. 30, no. 4, pp. 239-265, 2013.

[5] M. d'Amato and T. Kauko (Eds.), "Advances in automated valuation modeling. AVM after the non-agency mortgage crisis," Studies in Systems, Decision and Control, vol. 86, Springer, 2017.

[6] J. Zurada, A.S. Levitan, and J. Guan, "A comparison of regression and artificial intelligence methods in a mass appraisal context," Journal of Real Estate Research, vol. 33, no. 3, pp. 349-388, 2011.

[7] S. Peterson and A.B. Flangan, "Neural network hedonic pricing models in mass real estate appraisal," Journal of Real Estate Research, vol. 31, no. 2, pp. 147-164, 2009.

[8] E.A. Antipov and E.B. Pokryshevskaya, "Mass appraisal of residential apartments: an application of random forest for valuation and a CART-based approach for model diagnostics," Expert Systems with Applications, vol. 39, pp. 1772-1778, 2012.

[9] H. Kusan, O. Aytekin, and I. Özdemir, "The use of fuzzy logic in predicting house selling price," Expert Systems with Applications, vol. 37, no. 3, 1808-1813, 2010.

[10] V. Kontrimas and A. Verikas, "The mass appraisal of the real estate by computational intelligence," Applied Soft Computing, vol. 11, no. 1, pp. 443-448, 2011.

[11] A.G. Musa, O. Daramola, A. Owoloko, and O. Olugbara, "A neural-CBR system for real property valuation," Journal of Emerging Trends in Computing and Information Sciences, vol. 4, no. 8, pp. 611-622, 2013.

[12] M. Woźniak, M. Graña, and E. Corchado, "A survey of multiple classifier systems as hybrid systems," Information Fusion, vol. 16, pp. 3-17, 2014.

[13] B. Krawczyk, M. Woźniak, and B. Cyganek, "Clustering-based ensembles for one-class classification," Information Sciences, vol. 264, pp. 182-195, 2014.

[14] R. Burduk and K. Walkowiak, "Static classifier selection with interval weights of base classifiers," in: Nguyen, N.T., Trawiński, B., and Kosala R. (eds.) ACIIDS 2015, Part I. LNCS, vol. 9011, pp. 494-502, Springer, 2015.

[15] A. Fernández, V. López, M. José del Jesus, and F. Herrera, "Revisiting evolutionary fuzzy systems: taxonomy, applications, new trends and challenges," Knowledge-Based Systems, vol. 80, pp. 109-121, 2015.

[16] E. Lughofer, C. Cernuda, S. Kindermann and M. Pratama, "Generalized smart evolving fuzzy systems," Evolving Systems, March 2015.

[17] T. Lasota, Z. Telec, B. Trawiński, K. Trawiński, "Exploration of bagging ensembles comprising genetic fuzzy models to assist with real estate appraisals," in H. Yin and E. Corchado (Eds.): IDEAL 2009, LNCS 5788, pp. 554-561, Springer, 2009.

[18] M. Krzystanek, T. Lasota, Z. Telec, B. Trawiński, "Analysis of bagging ensembles of fuzzy models for premises valuation," in N.T. Nguyen et al. (Eds.): ACIIDS 2010, LNAI 5991, pp. 330-339, Springer, 2010.

[19] T. Lasota, Z. Telec, B. Trawiński, and K. Trawiński, "Investigation of the eTS evolving fuzzy systems applied to real estate appraisal," Journal of Multiple-Valued Logic and Soft Computing, vol. 17, no. 2-3, pp. 229-253, 2011.

[20] E. Lughofer, B. Trawiński, K. Trawiński, O. Kempa and T. Lasota, "On employing fuzzy modeling algorithms for the valuation of residential premises," Information Sciences, vol. 181, pp. 5123-5142, 2011.

[21] B. Trawiński, "Evolutionary fuzzy system ensemble approach to model real estate market based on data stream exploration," Journal of Universal Computer Science, vol. 19, no. 4, pp. 539-562, 2013.

[22] Z. Telec, B. Trawiński, T. Lasota, G. Trawiński, "Evaluation of neural network ensemble approach to predict from a data stream," in D. Hwang et al. (Eds.): ICCCI 2014, LNAI 8733, pp. 472-482, Springer, 2014.

[23] T. Lasota, E. Sawiłow, B. Trawiński, M. Roman, P. Matczuk, and P. Popowicz, "A method for merging similar zones to improve intelligent models for real estate appraisal," in: Nguyen, N.T., et al (Eds.) ACIIDS 2015, Part I. LNCS 9011, pp. 472-483, Springer, 2015.

[24] T. Lasota, E. Sawiłow, Z. Telec, B. Trawiński, M. Roman, P. Matczuk, and P. Popowicz, "Enhancing intelligent property valuation models by merging similar cadastral regions of a municipality," in: M. Nunez et al. (Eds.): ICCCI 2015, Part II, LNAI 9330, pp. 566-577, Springer 2015.

[25] I.H. Witten, E. Frank, M.A. Hall and C.J. Pal, "Data mining: practical machine learning tools and techniques," 4th Edition, Morgan Kaufmann, 2016.