Applied Soft Computing 11 (2011) 443–448

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

# The mass appraisal of the real estate by computational intelligence

Vilius Kontrimas[a,*], Antanas Verikas[b,1]

[a] Department of Information Systems, Department of Applied Electronics, Kaunas University of Technology, Studentu g. 50, Kaunas, Lithuania
[b] Department of Applied Electronics, Kaunas University of Technology, Studentu g. 50, Kaunas, Lithuania

ARTICLE INFO

*Article history:*
Received 10 September 2007
Received in revised form 1 December 2008
Accepted 6 December 2009
Available online 16 December 2009

*Keywords:*
Ordinary least squares regression
Support vector regression
Multilayer perceptron
Committee
Self-organizing map
Mass appraisal of real estate

ABSTRACT

Mass appraisal is the systematic appraisal of groups of properties as of a given date using standardized procedures and statistical testing. Mass appraisal is commonly used to compute real estate tax. There are three traditional real estate valuation methods: the sales comparison approach, income approach, and the cost approach. Mass appraisal models are commonly based on the sales comparison approach. The ordinary least squares (OLS) linear regression is the classical method used to build models in this approach. The method is compared with computational intelligence approaches – support vector machine (SVM) regression, multilayer perceptron (MLP), and a committee of predictors in this paper. All the three predictors are used to build a weighted data-depended committee. A self-organizing map (SOM) generating clusters of value zones is used to obtain the data-dependent aggregation weights. The experimental investigations performed using data cordially provided by the Register center of Lithuania have shown very promising results. The performance of the computational intelligence-based techniques was considerably higher than that obtained using the official real estate models of the Register center. The performance of the committee using the weights based on zones obtained from the SOM was also higher than of that exploiting the real estate value zones provided by the Register center.



## 1. Introduction

There are mass and individual appraisals of real estate. The individual appraisal is such an appraisal, when value of the exact object is determined according to all its individual characteristics. Mass appraisal is the systematic appraisal of groups of properties as of a given date using standardized procedures and statistical testing [40]. This valuation method is applied to property objects with many similarities. Mass appraisal of real estate is commonly applied to compute real estate tax.

The purpose of mass valuation is to estimate the market value. It must be distinguished from the market price and other, non-market values [17]. According to the Lithuanian Republic normative documents, the market value is estimated as the money amount, for which a property can be exchanged on the valuation date between a willing buyer and a willing seller in arm's-length transaction after proper marketing, wherein the parties act knowledgeably, without compulsion and impact of other transactions and interests [27,28]. 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 [17]. The market price is formed when curves of supply and demand intersect, it is influenced by many objective and subjective factors. The market price equals to the market value very rarely, because the market of real estate is not an ideal market. The market price of real estate reflects many subjective factors, so a real estate assessor must find the most objective, suitable for all value.

There are three traditional real estate valuation methods: the sales comparison approach, income approach, and the cost approach [30,40]. According to the sales comparison approach, the value is determined by comparing the object with the other objects sold in the market. The value is adjusted according to differences, as real estate objects have the differences. A difference up to 30–35% between characteristics of various objects is acceptable [26]. This method is very suitable for clear land. Reflection of the market price, quick and simple computations is the main advantages of this approach. The income approach is based on the premise that the value is the present worth of future; the value is determined by discounting cache flows generated by the object. The approach is very suitable for objects generating incomes, for example, buildings with leased offices or flats, objects used for services or production. This approach is quite simple too and estimates the economic ben-

* Corresponding author.
*E-mail address:* kontrimv@takas.lt (V. Kontrimas).
1 Also: Intelligent Systems Laboratory, Halmstad University, Box 823, S-301 18 Halmstad, Sweden.

1568-4946/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.asoc.2009.12.003

efit from the object. In the cost approach case, the value of the object is determined by construction costs minus depreciation. This approach can be applied only to buildings, and it is very suitable for schools, objects of engineering infrastructure and similar, which do not generate incomes and there are only a few objects to compare with.

Mass appraisal models are commonly based on the sales comparison approach. The linear regression is the most popular technique used to build mass appraisal valuation models. However, other techniques such as neural networks, support vector machines (SVM), and committees of models can also be used.

Most studies on this topic compare the performance of linear regression and neural network-based models. One of the first well known studies was performed by Do and Grudnitski [37]. The authors used a one hidden layer perceptron trained by backpropagation. A very small 6.9% mean absolute error was achieved using sales data of individual houses in San Diego at 1991. The mean absolute error achieved with the linear regression on the same data was equal to 11.26%. Similar results were achieved in other studies by Borst [38], Tay and Ho [11], and Evans et al. [1]. However, Worzala et al. tested the previous studies with similar data and their results were not so promising [12]. For example, while trying to replicate the study of Do and Grudnitsky, for the neural networks they achieved only 10% and 13.1% mean absolute errors using different software packages, while the mean absolute error of linear regression was equal to 11.1%. The conclusion, therefore, was that neural networks must be used very carefully for the real estate valuation. There were many more studies by the other authors: for example, Amabile and Rosato [39], Rossini [32,33], Nguyen and Cripps [31], Ge et al. [21], Wilson et al. [16]. Their results show slight advantage of neural networks against the linear regression. It is obvious that both techniques may show advantage against each other depending on the quality and amount of the data, dependencies between the variables.

The purpose of this study is to explore the usefulness of the most prominent computational intelligence techniques for mass appraisal. The linear regression which is the most popular technique in various studies, a multilayer perceptron (MLP), a support vector machine, and a committee of the models are the techniques used in the investigation. The number of unacceptable valuations is the main parameter of usefulness of a mass appraisal technique. Valuation differing from real value of an object more than certain percentage, 20% in Lithuania [28], is called an unacceptable valuation. There are two main reasons of using the linear regression instead of some higher order regression. Firstly, the linear regression is still used in most mass appraisal systems. Thus, old mass appraisal models can be included into a committee and, therefore, a much easier way of moving from legacy appraisal systems to ones proposed in this work is created. Secondly, Worzala has demonstrated that the linear regression is sometimes more accurate than non-linear computational intelligence techniques [12]. Our results also confirmed this fact. SVM and committees of models have shown excellent performance in the recently studied task of detecting fictious real estate transactions [46–48]. We present a way of obtaining data-dependent weights for aggregating the models into a committee. The weights are based on real estate value zones obtained from the SOM. We demonstrate that committees based on value zones generated by SOM are more accurate than those exploiting the real estate value zones provided by the Register center. The finding means that the time consuming and expensive step of establishing the real estate value zones by an expert can be avoided.

The remainder of the paper is organized as follows. In the next section, a brief description of the techniques used to solve the task is given. The results of the experimental tests are presented in Section 3. The conclusions of the study are given in Section 4.

## 2. Methods

### 2.1. Linear regression

The standard regression model is given by:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{1}$$

where $\mathbf{y}$ is a $n \times 1$ vector of dependent variable values with $n$ being the number of observations, $\mathbf{X}$ is a $n \times m$ matrix containing values of independent variables, $\boldsymbol{\beta}$ is a $m \times 1$ vector of regression coefficients, $\boldsymbol{\varepsilon}$ is a $n \times 1$ vector of true errors with the standard deviation $\sigma$, and $m$ is the number of independent variables. The meaning of $\mathbf{X}$ and $\mathbf{y}$ stated here is kept through entire paper. The estimate $\mathbf{b}$ of $\boldsymbol{\beta}$ is obtained as a solution to:

$$\min_{b} Q_{OLS}(\boldsymbol{b}) \tag{2}$$

where $Q_{OLS} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} e_i^2$ in the ordinary least squares (OLS) case, where $e_i$ is the estimation of the true error, and $\hat{y}_i$ is an estimate of $y_i$.

### 2.2. Support vector machine for regression

The support vector method can be applied to the case of regression, maintaining all the main features that characterize the maximal margin algorithm developed for classification: a non-linear function is learned by a linear learning machine in a kernel-induced feature space, while the capacity of the system is controlled by a parameter that does not depend on the dimensionality of the space [20].

The standard form of the support vector regression provided by Vapnik [50] is:

$$\min_{w,b,\xi,\xi^*} = \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\sum_{i=1}^{n}\xi_i + C\sum_{i=1}^{n}\xi_i^*,$$

subject to:

$$\begin{aligned} & y_i - \boldsymbol{w}^T\phi(\boldsymbol{x}_i) - b \le \varepsilon + \xi_i; \\ & -y_i + \boldsymbol{w}^T\phi(\boldsymbol{x}_i) + b \le \varepsilon + \xi_i^*; \;\; \xi_i, \xi_i^* \ge 0, \;\; i = 1, \ldots, n \end{aligned} \tag{3}$$

where $\mathbf{w}$ stands for the weights vector, $\phi(\mathbf{x}_i)$ is an image of $\mathbf{x}$ in the feature space, $y_i$ is the target, $n$ is the number of learning data points, $\xi_i, \xi_i^*$ are the slack variables, which measure the cost of the errors on the training points, $b$ is the threshold, the regularization constant $C$ controls the trade-off between the norm and the losses, and $\varepsilon$ is the threshold for ignoring the errors.

The dual form is:

$$\min_{\alpha,\alpha^*} = \frac{1}{2}(\alpha - \alpha^*)^T(\alpha - \alpha^*)\kappa(x_i, x_j) + \varepsilon\sum_{i=1}^{l}(\alpha + \alpha^*) + \sum_{i=1}^{l} y_i(\alpha - \alpha^*),$$

subject to:

$$\sum_{i=1}^{l}(\alpha - \alpha^*) = 0; \quad \alpha_i \ge 0; \;\; \alpha_i^* \le C; \;\; i = 1, \ldots, n \tag{4}$$

where $\kappa(x_i, x_j)$ is the kernel (we used the polynomial kernel in this study) and the parameters $\alpha_i^*$ and $\alpha_i$ are found during the optimiza-