House Price Estimation in Hanoi using Artificial Neural Network and Support Vector Machine: in Considering Effects of Status and House Quality

Quang-Thanh BUI, Nhu-Hiep DO, Huu-Phe HOANG, Vietnam

Key words: House price, Ha Noi, Neural network, Support vector machine, Ensemble

SUMMARY

House pricing is considered to be a complex social - economic process that is difficult to model with relevant accuracy. Based on Status Quality Trade Off theory, this paper aims to employ regression models, namely Artificial Neural Network (ANN) and Support vector machine (SVM) and Ensemble techniques, in estimating the sale prices of residential properties. Ha noi was selected as a case study in which 1000 locations of houses and influencing factors were collected and used to train and to validate the models. Outputs of the models were further analyzed in considering the effects of status of the house (intangible) and quality of the houses (tangible). The results show that the forecasting methods based on ANN and SVM are feasible and effective.
1. INTRODUCTION

The housing market has always played crucial role in any nation’s economy as a large part of total gross domestic product (Plakandaras, Gupta, Gogas, & Papadimitriou, 2015) and been considered as key element contributing to the national growth. Its stability is seen as an important index that any fluctuation in the market reflects on a significant effects on every family (Selim, 2009) as well as the whole nation. It is, therefore, essential to have accurate forecasting models, that measure the changes of house prices, to support decision makers to implement the macroscopic regulation under the market economy (Wang, Wen, Zhang, & Wang, 2014). But, most fundamentally, a selection of base theory will reflect the performance of forecasting models.

In shaping the structure of housing market, several theories have been proposed, aiming at discussing the relationship between location and physical quality of houses as in (Alonso, 1964; Fujita, 1989; Kim, Pagliara, & Preston, 2005; Smith, 1987). Recently, (Huu Phe & Wakely, 2000) explained the dynamics of urban transformation and development, or in particular the dynamics of choice of residential houses. This Status Quality Trade Off theory (SQTO) reflects the tradeoff between house status and dwelling quality that significantly form the variation of house prices. SQTO seeks to explain location-based choices of house by considering dynamic process of physical characteristics of houses (house quality such house area, number of floors, number of rooms...) and house status (as consequence of urban development process that form the social desirability of housing such as distance from ritual center, feeling of safety...).

Related studies

Price prediction mapping seeks to understand the influential factors controlling the changes of price over a given area. Driving forces of houses are mainly depended on physical/virtual distances to status poles and quality of the houses (Huu Phe & Wakely, 2000). Traditionally, there are several detection techniques that map study area with varying prices depending on location of the houses. Of all, hedonic regression and geographically weighted regression are two common approaches in
dealing with spatial data and with house prices in particular as in works of (Gollini, Lu, Charlton, Brunsdon, & Harris, 2015; Lu, Harris, Charlton, & Brunsdon, 2014) (Chen, Clapp, & Tirtiroglu, 2011; Dorsey, Hu, Mayer, & Wang, 2010; Liao & Wang, 2012; Selim, 2009). Several automated valuation model used to predict values of house prices that based on traditional statistics approach were also proposed by (Vo, Shi, & Szajman, 2014). Although have been used extensively over the past several decades, hedonic regression received lot of criticisms on model assumption and estimation (Fan, Ong, & Koh, 2006; Malpezzi, 2002) and on solving non-linear problem (Selim, 2009) global regression and local clustering.

However, the driving forces for house prices are complex and inter-related processes that traditional statistical approach fail to address. Recent researches have proved non-linear machine learning and the use of fuzzy logics (Kuşan, Aytekin, & Özdemir, 2010) can be a suitable solution to predict the variation of house prices. (Chiarazzo, Caggiani, Marinelli, & Ottomanelli, 2014; Steven P. Peterson, 2009; Vo et al., 2014; Wong, So, & Hung, 2002) employed Neural network for property prediction, (Gu, Zhu, & Jiang, 2011; Wang et al., 2014) used Support vector machine in combination with optimization models such as Generic algorithm and Particle swarm optimization. (Selim, 2009) made a comparison between hedonic regression and artificial neural network. The non-linear data mining models improve the prediction model, however its application property market are still rare.

Through literature review, we understood that there have been no researches trying to explore the SQTO theory in combination with non-linear data mining models, even though these models have been employed in several researches. This paper focused on a systematic comparison of artificial neural network and support vector machine and ensemble models in predicting the house prices in Ha Noi, Viet Nam. Based on SQTO, we selected appropriate variables and those variables were used to run three models. For accuracies assessment, we calculated root mean squared error (RMSE) and mean absolute error (MAE). Data were collected during 2014 survey in 1000 locations in Ha Noi. Analysis of house pricing data and modeling were carried out using QGIS 2.10 Pisa and Weka 3.8.1 softwares.

The rest of the paper is organized as follows: the next section provides description of the study area, sampling strategies and selection of variables and introduces research methodology; the third part describes the designation of experiments; and the fourth includes results and conclusion; final remarks are concluded in the last section.

House Price Estimation in Hanoi Using Artificial Neural Network and Support Vector Machine: in Considering Effects of Status and House Quality (8751)
Quang Thanh Bui (Viet Nam)

FIG Working Week 2017
Surveying the world of tomorrow - From digitalisation to augmented reality
Helsinki, Finland, May 29–June 2, 2017
2. DATA AND METHODS

2.1 Description of the study area

Housing in Hanoi City has transformed from a social service to a goods in the market following the introduction of the Doi Moi (renovation) policy, which started in 1986. As a result, the segmentation of the nascent market in the capital city of 7 millions residents poses complex problems for an equitable provision of housing environment to different income groups. A model of urban housing based on the interaction between tangible and intangible attributes (SQTO) seems to be able to achieve important results in price estimation using hedonic price index techniques with the emphasis placed on status – related elements. A survey of 1,000 households within the core city of Hanoi has initially showed the potentials of using SQTO as a tool for successful analysis of the house price dynamics and its spatial manifestation.

Throughout several stages of development, the housing markets of Hanoi have undergone a dynamic movement that change itself from a social service to a commodity in market-oriented economy. Under the prism od SQTO, social aspect of housing market is the emphasis on physical factors such as size, number of floors, building materials..etc, while the market characteristics include intangible factors such such vision, housing neighborhoods, asseccibility..etc. The values used are largely concentrated in the housing sector with relatively low utilities, while the exchange value expressed common in higher utility segment.
As results, there are many manners for provision of housing, of which commercial factors are considered to be crucial elements. In recent years, The house market of Viet nam is motivated by integrated movement of social driving forces and market driving forces. It is essential to have a model to predict house price variation across the study area. The study area was defined in old districts of Ha noi, in which 1000 samples were randomly collected using systematic sample grid of 400m x 400m with 245 variables (Figure 1).

2.2 Research Methodology

2.2.1 Artificial Neural Network

ANN uses a model derived from biological brain and is extensively applied to solve classification and regression problem. ANNs can be applied to problems where the input data are well-defined and the distribution of input dataset is not dependent on pre-assumptions (Gardner & Dorling, 1998). ANN maps sets of input data onto a set of appropriate outputs (house price, in this case study).
The structure of MLP is shown in (Figure 2). It consists of three main components with each layer fully connected to the next one, namely input layers, hidden layers and output layer. The first are influential factors or input data to the next step. Hidden layers in the second components act as processing element with a nonlinear activation function. The background classification of input data to classes is a complex iterative processes in which weights of all parameters are adjusted. The classified results are binary inferring fired or non-fired (based on predefined threshold). MLP utilizes back-propagation learning technique that input are propagated through hidden layers and outputs are compared to pre-defined values to decide whether the iteration stops.

2.2.2 Support Vector Machine

Technically, Support vector machine (SVM) is used to find the optimal separating hyperplane which maximizes the gap or margin between classes of the training data. By other way of saying, the searching for optimal hyperplane leads to the solving of the objective function (Equations 1 and 2) using Lagrangian multipliers

\[
\text{Objective Function: } \min_{w,b,\xi_i} \left\{ \frac{||w||^2}{2} + C \sum_{i=1}^{n}(\xi_i) \right\}
\]

(1)

Linear Constraints: 
\[
y_i (w^T x_i + b) \geq 1 - \xi_i , \ \forall x_i \in D
\]
\[
\xi_i \geq 0 \ \forall x_i \in D
\]

(2)
Support Vector Machine can also be used as a regression method with the same principles, keep solving the optimization problem through minimizing error, individualizing hyperplane, keeping in mind that part of the error is tolerated

2.2.3 Stacking

Stacking (sometimes called stacked generalization) is one of the well-known ensemble techniques used to solve classification and regression problems. Usually it is used in conjunction with other classifiers to improve the performance of those techniques. The stacking training process starts by fitting all algorithms on input data, then a combiner algorithm is trained to make a final prediction using all the predictions of the other algorithms as additional inputs. The final result yields performance better than any single one of the trained models

2.2.4 Bagging

Bagging or Bootstrap aggregating is another resemble technique that is used to reduce variance and to avoid over-fitting. It generates additioinal data for training from original dataset by sampling with replacement that leads to inceasement of training data. Consequentely, the model decreases the variance, narrowly tuning the prediction to expected outcome.

2.2.5 Performance assessment

Two statistical measures, namely Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) were often used to evaluate the performance of house price prediction models (Park & Bae, 2015; Wu, Gyourko, & Deng, 2012), eventhough RMSE is found sensitive to large value and outliers. To overcome this sensiveness, we had screened to clean out outliers and standardized input data to bring them into comparable ranges.

3. EXPERIMENTAL DESIGN

3.1 House price influencing factor

There are two components in SQTO theory that supports shaping the house price index, namely tangible and intangible. Tangible variables are explained as house quality or as physical characteristics of the house. Intangible variables are identified as status or desirable marketisable chaacters of the location. Of all 245 variables (HH Phe, 2015) narrowed down to 30 significant attributes through stepwise regression technique. Table 1 show our choice of tangible and intangible

---

House Price Estimation in Hanoi Using Artificial Neural Network and Support Vector Machine: in Considering Effects of Status and House Quality (8751)

Quang Thanh Bui (Viet Nam)

FIG Working Week 2017
Surveying the world of tomorrow - From digitalisation to augmented reality
Helsinki, Finland, May 29–June 2, 2017
variables as concluded in (HH Phe, 2015). For data mining models, the selected variables in (Table 1) were standardized to the same range to avoid effect of high values.

**Error! Reference source not found.** shows a spatial interpolation of house price in VND per square meters though using kriging algorithm. Several picks had been detected as in old districts, Westlake surrounding area, and West part of the city. The three can be considered as significant status poles in considering distribution of house price across the city.

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Description</th>
<th>Data type</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPRICVND</td>
<td>Price of house in Millions of Vietnamese Dong</td>
<td>Ratio</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tangible Independent variables</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIRCON</td>
<td>Air-Conditioner (Yes, No)</td>
</tr>
<tr>
<td>GFA</td>
<td>Total floor area (incl. mezzanine) (m²)</td>
</tr>
<tr>
<td>PLOTAREA</td>
<td>Total plot area (m²)</td>
</tr>
<tr>
<td>SHOPFRNT</td>
<td>Shop Front (Yes, No)</td>
</tr>
<tr>
<td>PLUMBING</td>
<td>Plumbing Quality (Good, Other)</td>
</tr>
<tr>
<td>HOUSEGRADE</td>
<td>Permanent, Other</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Intangible Independent variables</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAR</td>
<td>Car ownership (Yes, No)</td>
</tr>
<tr>
<td>CENTDISR</td>
<td>Measured distance to Centre District</td>
</tr>
<tr>
<td>DISCENDI</td>
<td>Perceived travel time to the Centre District</td>
</tr>
<tr>
<td>EDYEARS</td>
<td>Time in education of the interviewee (years)</td>
</tr>
<tr>
<td>OCCUP_PRIVBIZ</td>
<td>Private Business owner (1=Yes, 0=No)</td>
</tr>
<tr>
<td>SCHOOQLT</td>
<td>School Quality (Good, Other)</td>
</tr>
<tr>
<td>STRTYPE</td>
<td>Type of street (Business, Residential)</td>
</tr>
<tr>
<td>BACTULIEM*</td>
<td>Located in Bac Tu Liem District</td>
</tr>
<tr>
<td>BADINH*</td>
<td>Located in Ba Dinh District</td>
</tr>
<tr>
<td>CAUGIAY*</td>
<td>Located in Cau Giay District</td>
</tr>
<tr>
<td>DONGANH*</td>
<td>Located in Dong Anh District</td>
</tr>
<tr>
<td>DONGDA*</td>
<td>Located in Dong Da District</td>
</tr>
<tr>
<td>GIALAM*</td>
<td>Located in Gia Lam District</td>
</tr>
<tr>
<td>HADONG*</td>
<td>Located in Ha Dong District</td>
</tr>
<tr>
<td>HAIBATRUNG*</td>
<td>Located in Hai Ba Trung District</td>
</tr>
<tr>
<td>HOANGMAI*</td>
<td>Located in Hoang Mai District</td>
</tr>
<tr>
<td>HOANKIEM*</td>
<td>Located in Hoan Kiem District</td>
</tr>
<tr>
<td>LONGBIEN*</td>
<td>Located in Long Bien District</td>
</tr>
<tr>
<td>NAMTULIEM*</td>
<td>Located in Nam Tu Liem District</td>
</tr>
</tbody>
</table>

House Price Estimation in Hanoi Using Artificial Neural Network and Support Vector Machine: in Considering Effects of Status and House Quality (8751)
Quang Thanh Bui (Viet Nam)

FIG Working Week 2017
Surveying the world of tomorrow - From digitalisation to augmented reality
Helsinki, Finland, May 29–June 2, 2017
3.2 Analysis procedure

This section describes the results of two proposed models for house pricing modeling. In this research, the data were processed and visualized using QGIS. The Neural Network and Support Vector Machine, Ensemble models were implemented by using Weka data mining software. To determine the performance of each classifier, we explored stratified 10-fold cross-validation performance test. Training data of 1000 records were randomly split into ten folds, each had the same proportions of 100 records. The learning procedure is executed on different training sets as for 10 times. Errors of the 10 estimates are calculated by averaging.

4. RESULTS AND DISCUSSION

4.1 Spatial distribution of house prices

The SQTO suggests that the urban residential location follow certain organising regularities expressed in the concentric rings around the social status poles. These status poles represent distinctive, widely recognised poles of social values. At the preparation of the survey, the status poles of Hanoi City were preliminarily proposed as i) the Restored Sword Lake area (Hồ Gươm), ii) the West Lake area (Hồ Tây) and iii) the South West area (around Trung Hoa Nhan Chinh New Urban Area).

The role of the Restored Sword area (Hồ Gươm) as the most important city centre, is a historical fact, widely recognised. This is a conventional downtown area, a combined status pole including, but not limiting to, familiar aspects such as the level of wealth (especially real estate wealth), the market place prestige, the sophistication of an elite city lifestyle, the level of education attainment, etc. This area surrounds the famous “36 Old Streets” quarter of Hanoi, the written historical and literary records about which went back as far as the XIV century (HH Phe, 2015).

The West Lake area (Hồ Tây) is emerging as the most important food, fashion and creative quarter of Hanoi City. Traditionally favoured by the educated elite, the nouveau riche, expats and yuppies, who live in often rented private villas with superb views of the lake and mild microclimate due to
the vicinity of the largest body of water in town, the area is shrouded in a somewhat mystic fog of colourful historical events, both real and imagined. This is truly a status pole for very well-defined groups of the educated elite, the nouveau riche, expats and the yuppies.

The South West area (around Trung Hoa Nhan Chinh New Urban Area) has the distinction of a trigger for a massive urbanisation drive towards the West of Hanoi City, where a novel, if somewhat still unstable, high-rise lifestyle, is being formed, leaving behind many doubts and phobias related to the communal housing stock, the products of the first wave of industrialised, or system construction, in the early 1970s, following the East European precast technology. The area was deliberately targeting the yuppies group (Huu Phe, 2002).

The answers to the interview questions helped identify the status poles, in a surprising accordance with the assumed poles suggested at the beginning of the survey. Using the 3D kriging technique, the poles were discerned clearly, both in terms of housing unit price and in terms of level of utility, expressed in price per sq. m. (Figure 3)

![Figure 3: Kriging of house price in VND per Square meters of floor space (HH Phe, 2015)](image-url)
4.2 House price modelling

Selection of parameters is key step to ensure the best performance of the models. Parameters were optimized through trial-and-error process, in which algorithm was tuned by alternatively changing the parameter values. We decided to use default parameter values for simple calculation. For each algorithm, the best performed parameters were finally selected for further analysis. In this experiment, Neural Network, Support Vector Machine, Ensemble techniques (Stacking, Bagging) were used via following steps:

Step 1. Division of standardized dataset (Training, Validation, Test)
Step 2. Selection of regression techniques (Neural Network, SVM, Ensemble).
Step 3. Parameter tuning
Step 4. Model training with 10-fold-validation
Step 5. Model evaluation by test set
Step 6. Measurement of errors from test set

![Graph showing Mean Absolute Error](image)

Neural Network SVM Bagging Stacking
5. CONCLUSION

Through examining the case study in Ha Noi, we found that SQTO theory can fully be used to explain the structure of urban residents. In addition to the systematic prediction for the presence of different poles in Hanoi, the estimation of hidden (implicit) value of tangible and intangible attributes based on data mining algorithm has proved to be very useful in identifying the driving forces of house price formation across the market segments. The result also describes a strong movement of city structure from uni-polar to multiple-polar structure and it can be considered as preliminary foundation for an urban decision support system.

ACKNOWLEDGEMENT

The first author’s attendance at the Conference is supported by the National Foundation for Science and Technology Development (NAFOSTED), Vietnam
REFERENCES


House Price Estimation in Hanoi Using Artificial Neural Network and Support Vector Machine: in Considering Effects of Status and House Quality (8751)
Quang Thanh Bui (Viet Nam)

FIG Working Week 2017
Surveying the world of tomorrow - From digitalisation to augmented reality
Helsinki, Finland, May 29–June 2, 2017


---

**House Price Estimation in Hanoi Using Artificial Neural Network and Support Vector Machine: in Considering Effects of Status and House Quality (8751)**

Quang Thanh Bui (Viet Nam)

FIG Working Week 2017

Surveying the world of tomorrow - From digitalisation to augmented reality

Helsinki, Finland, May 29–June 2, 2017
CONTACTS

Quang-Thanh BUI
Institution: VNU University of Science
Address: 334 Nguyen Trai
City: Ha Noi
COUNTRY: Vietnam
Tel. +84 4 38582420
Email: qthanh.bui@gmail.com, thanhbq@vnu.edu.vn
Web site: hus.vnu.edu.vn

Nhu-Hiep DO
Institution: Hanoi University of Natural Resources and Environment
Address: 41A Phu Dien
City: Ha Noi
COUNTRY: Vietnam
Tel. +84 4 38370598
Email: dnhiep@hunre.edu.vn
Web site: hunre.edu.vn

Huu-Phe HOANG
Organization: Vinaconex R&D
Address: 3rd Floor, VIMECO building, Phạm Hùng Rd
City: Ha Noi
COUNTRY: Vietnam
Tel. +84 4 7848665
Email: hoanghuuphe@googlemail.com
Web site: vinaconex.com.vn

House Price Estimation in Hanoi Using Artificial Neural Network and Support Vector Machine: in Considering Effects of Status and House Quality (8751)
Quang Thanh Bui (Viet Nam)

FIG Working Week 2017
Surveying the world of tomorrow - From digitalisation to augmented reality
Helsinki, Finland, May 29–June 2, 2017