Proposed study on evaluating and forecasting the resident property value based on specific determinants by case base reasoning and artificial neural network approach

Sinan Adnan Diwan
Computer Science and Information Technology College, University of Wasit, Iraq

ABSTRACT
Real estate forecasting has become an integral part of the larger process of business planning and strategic management in real estate sector. This study covers residential estate markets and concentrates on property types, while previous studies that have considered country wide house price indices. There is a gap identified in the literature which need to study correlations between property types within a region or a city and whether they will provide diversification benefits for real estate investors such as risk reduction per unit of returns. This aim of the paper is to propose and develop a computer assisted real estate property price forecasting model. This proposed framework will examine the current uses of artificial intelligence, particularly combining case base reasoning and artificial neural network, in the business-forecasting field and considers suitable applications in real estate. The methodology consists of five phases: 1) Data gathering b) Data cleaning c) ANN Training d) CBR (Case Base Reasoning) similarity retrieval e) Result retrieval. This research will investigate the influence of residential real estate property characteristics on property values (prices) in global context, it revealed a high positive linear correlation between property characteristics and the property market values; an indication that these characteristics reasonably predict property market values. The results of the study will enable Real Estate Professionals to make fair estimates of the market values of residential real estate properties given the features/characteristics of such housing units.

1. INTRODUCTION
The Real estate is a term defined as land, or above the air or below the ground and any type of buildings or constructions on top of it. It is always referred as reality in many articles and newspaper. Real estate covers residential housing areas which includes apartments or flats or condominiums or individual houses, commercial office buildings, trading areas which includes hotels, restaurants, theatres, retail markets, entertainment centers, hyper malls and supermarkets, industrial areas which includes factories, government sites and buildings. Involvement of real estate is more on purchase of property, selling the property, land development, residential plots and buildings or non-residential sites. The important players in the real estate market are landlords who owns the property, developers, construction builders, real estate agents, property evaluators, tenants and buyers who buy the property. The significance of real estate is an asset class and it is difficult to overstate, there may be fall or rise in the market, which is due to country economic growth or political situations. In the global market, US have a total value of asset investment was about $50 trillion at the end of the year 2018 and out of that more than $25 trillion was invested in residential

Copyright © 2020 Institute of Advanced Engineering and Science. All rights reserved.
properties. In one of the real estate survey conducted by US real estate agencies, comparing to the US stock market capitalization in the neighboring countries were $20 trillion [1]. Recent history states that the real estate price fluctuations may be in bubble or burst mode, but it has the very potential to cause disasters on the financial sectors or with remaining part of the economy. Some of the impacts are due to influences in which real estate is considered as one of the easiest assets to burrow via mortgage based on household perspective. Survey conducted based on global market states that, 75% of global populations are not happy with the current market due to expensive living cost, properties are over-priced and lack of government benefits, more technology based industries are growing around the world. The satisfaction rate around the globe real estate climate have dropped from second quarter of 2017 due to cost factors, increase in fuel and gold prices, governments investments towards military sectors, higher taxes and government strategic plans towards people on buying affordable houses. Researchers Case, Quigley and Shiller [2] have made a study to show that how the real estate price variations have created a significant impact on the aggregate consumption in the US markets compare to stock market and current volatility in residential market, impact is more similar in the countries like Australia, China, Japan and India. The study conducted by Reinhart and Rogoff [3] have stated that real estate pricing factors will create a huge impact on the nations growth, from economic perspective, real estate values is not least important compared to understanding the price dynamics on assets such as stocks, bond, commodities and currencies. In specific, Asian countries stands on the certain threshold of creating a secondary mortgage market to assemble the capital financial investment for the main market, therefore, there is a need to assist real estate professionals with detailed information’s on the residential property markets features which can influence the buyers or sellers. There were several research studies conducted on money investment and capital share markets, but real estate is still considered as one of the best investment around the global market.

In recent years, Artificial Intelligence(AI) have emerged fasters in every domain and there was rapid increased in computing power to make faster predictions on business growth. Artificial Intelligence have the ability to perform intelligent functions as similar to the human brains. In specific, logic and reasoning are part of the AI engine which can learn and improve over time in making self-decisions. The application of AI is mainly used on computing and robotic domains, whereas they are formed as integral part of image processing and speech recognition software’s. Currently AI applications are extending to various social science domains which includes business domain. AI approaches are based on statistical methods and computation intelligence. As part of AI tools, Artificial Neural Networks(ANN) and Genetic algorithms are very popular in the field of market study, research and forecasting. Expert systems applications are part of AI which is widely used in healthcare, military and business. Researcher DeLurgio [4] have clearly stated difference between Conventional program system(CPS) and Expert system(ES). Author states that CPS involves more on researcher perspective, who wants in creating a system which deals with more impacts and handle difficult tasks without concerns to whether these are similar to those used by humans i.e. it does not give more importance on how the job to be completed as long as it completes the job. Comparing to CPS. the ES tries to gain an understanding of how humans solve problems and then uses the computer to explain and predict their behaviour. In current industry trends, many systems contain both the elements. Therefore, many systems have certain aspects of expert systems but often rely on certain champing abilities of a CPS. Therefore, ES systems have become more popular or trend in real estate application, while combining hybrid methods are becoming prominent [5]. The emergence of expert component or partially expert component is very important for educators in all domains. The strength of ES have the capability to provide expert advice to the non-experts, assist the people to solve problem who are not experience in specific domain (i.e. non-expert) [6]. For educators or industrialist, solving a problem is very beneficial component in the expert system. Therefore, well framed ES software can be very valuable teaching and training tool for the industry.

There were several research reviews carried out for past years on artificial intelligence in various domain, but only very limited research works have been conducted to demonstrate the capabilities of neural networks to evaluate the prices in residential market and many of such carried works are based on single city datasets. Current research conducted is based on real estate property determinants using residential housing data. The proposed study on forecasting the real estate price in residential properties by combination of Artificial Neural Network (ANN) and Case Base Reasoning (CBR) techniques.

The major research challenge demonstrated in this research study is on unpredictability of commercial real estate markets. This unpredictability has resulted in rising and falling values over the time and, in parallel, it provides a better understanding of the dynamics of property markets, their drivers, economic and other linkages which could deliver more practical development decisions. When there is an over- development or the oversupply of space may be considered as a prime cause of market ‘boom/bust’ cycles. Current real estate data have unique challenges in forecasting settings. First, the predictive results must be taken into account with a high transaction costs, which in real estate can be approximately 6 percent, or even higher, of the property value. In adding to statistical significance, the estimated coefficient need to be
large enough to cover the additional costs. Second, the availability of real estate data is relatively short in its time duration and it is observed at a frequency of monthly or quarterly. The main challenge is the lack of consistency between the market forecasts applied in property assessments and the evident unpredictability in property markets causes the problem of potential valuation inaccuracies and presents the challenge of seeking improved forecasting methods for assessment. Therefore, questions arise in this study to find out the solution for forecasting the price evaluation are as follows:

a) Which property characteristics are important in the determination of the market value of a residential real estate?

b) What basis do property professionals and developers currently formulate and apply property market forecasts in their project viability studies and valuations?

c) Do varying office rent growth forecasts have a significant impact on the results of building valuation assessments?

d) Can the artificial neural network and case base reasoning be a practical model for the valuation of residential housing units in global property market given property characteristics/features?

The purpose of the study is to gain maximum returns for the investors and owners of the property. Based on the amount of money involved in real estate, there is a need of accurate prediction of returns and associated risks. The primary objective of this research is to identify property features/characteristics that have significant influence on property market values using case base reasoning and artificial neural network. The main work is to determine if the artificial neural network and CBR could be a practical model for real estate valuation in property market, and to contribute to the development of knowledge in the field of real estate finance. The study is restricted to residential real property only and uses property transaction data from different geographical location around the globe. Expectedly, the study should be of practical relevance to practicing real estate practitioners as well as academics. Aligned with the research problem and questions, the research detailed in this work possesses few objectives, which are as follows:

a) Design a combined classifier engine module [ANN and CBR] for price evaluation and prediction.

b) Representation of input by identifying features, including what specific determinants should be included, organization of the case base (including indexing attributes and index structure) and retrieval of cases that is potentially relevant to a new case.

c) Develop a pilot prototype to evaluate and forecast the price of real estate market.

There were several studies investigated on the application of neural networks to real estate valuation, still there were gaps on forecasting the accurate cost estimation. Research studies carried out by Borst [7], have used ANNs to certain sets of residential housing areas in New England, author have used several datasets for training, the forecasting classification accuracy was 70%. Research team of Tay and Ho [8] have investigated on the residential flats in Singapore and their reasons of price fluctuations; another researcher Evans [9], et all have applied AI techniques on residential datasets from California and UK, observed the price variations and factors. Wilson, et.al [10] have used neural network classifier to forecast the future trends in UK housing market based on national residential time series datasets. Author Pacharavanich and team have investigated on the valuating the price factors in Bangkok town house, authors have applied basic CBR model to evaluate the price estimation. Research team lead by Lokshina have used ANN and ANFIS model to measure the real estate prices, authors have created price models using ANN and ANFIS and compared which model have highest accuracy, ANN have achieved 85% of accuracy in cost estimation [11, 12].

There were numerous of machine learning methods and approaches are applicable for property appraisal and valuation and it is reported in the literature’s. Researchers have designed intelligent systems for commercial property price predictions. Intelligent systems were designed using hybrid Multi-Layer Perceptron[MLP] ANN and rule-based system, which achieves an accuracy of 80% [13, 14]. The study conducted by Guan et al [15] have explained on design and development of Adaptive Neuro-Fuzzy Inference System [ANFIS] approach which is useful to determine the residential property prices. Team of researchers have applied multi-resolution approach in china real estate market datasets to determine the value of real estimate prices, authors have applied three set of theories which includes a) Unascertained principles and theory b) Principle Component Analysis (PCA) c) Ant Colony Optimization (ACO) based ANN approach. The results achieved through these methods were accurate and meet the actual market requirements.

The research carried out based on previous CBR (Case Base Reasoning) approaches which has been reported in the literatures are applied in the field of legal, healthcare and finance not on the real estate market. For instance, strength of CBR based legal reasoning have more advances in modelling the cases based on previous case outcome and generate new decisions. Some of the models have generated legal arguments which needs to simulate human adversarial discourse. These attempts of developing intelligent systems on law domain were boosted by HYPO systems [16]. HYPO is a CBR system which creates legal arguments from case base on the trade secret law domain. Researcher Gardener [17] in his doctorate dissertation have
taken a problem of identifying the hard and easy questions in the context on analysing the facts from the school exams on contract law. Another team of researchers Branting and Lester [18] have designed the concept of case reuse for document drafting, in this approach, authors have demonstrated the rhetorical structures of documents in self-explanatory formats. Daniels and Rissland [19] have developed hybrid CBR and Information Retrieval (IR) system, where the CBR components is playing a key role in query processing and feed into IR system for retrieving the results. Most of the research carried out by ANN model are applied in medical and other domains which includes electricity consumptions, house pricing predictions [20-25] and CBR applied in finance datasets. By combining, two machine learning approaches for specific domains, there were very limited existing literatures.

Based on all review works, still there is need of accurate price evaluation forecasting systems for real estate property valuation. Therefore, the proposed research stems from the combination of ANN and CBR in the implementation of a decision support system in real estate property valuation to improve the accuracy of predication and decision making.

In this paper, as mentioned above, the main objective is to study and forecast the real estate prices in residential property areas around the globe by having specific determinants. The rest of the paper is organized as follows: in Section 2, proposed methods which explains the research contribution; In Section 3, proposed research methods to predict the real estate price based on combination of ANN and CBR approach is explained in several stages. In Section 4, results are discussed. Finally, paper is concluded with Section 5.

2. PROPOSED METHOD

The proposed approach in this study will be a single system and act as combination approach which includes ANN and CBR framework which strengthens the instance based learning methods. The experienced-based problem solving capability of CBR systems and its viable explanation mechanism is combined with the strong interpolative capability of ANN will form a Combined Intelligent System (CIS), will act as decision support system for the real estate property price evaluation in global market. Our proposed study is an initial study, it innovates the combination of ANN and CBR in to a single system framework leveraging the strengths of the two instances based learning techniques. Therefore, first attempt of hybridizing the two approaches into a real time system is for achieving an improved decision support in the real estate domain. To implement this, data from selected input variables of new cases are transformed through a pre-processing procedure into numerical data which must be in suitable format for ANN computation, then the result of the ANN is passed to the CBR component. Thereafter, the CBR component seeks for existing past cases that are sufficiently similar to the input case whose solution and explanation can be adapted to the new context. Hence, this study introduces the novel hybridization of ANN and CBR decision support in real property valuation for improved performance relative to the application of a solitary ANN or CBR approach in real estate price evaluation and forecasting.

The Artificial Neural Network -CBR system is a combined system of two components. The two components are the ANN component and the CBR component. It is illustrated in below Figure 1 of proposed methodology.

![Figure 1. Illustrates the proposed methodology for forecasting the price in real estate](image-url)
**ANN Component**: The multi-layer perceptron (MLP) ANN is a powerful neural network model that can be used for solving approximation, estimation, classification and prediction problems. Generally, it consists of an input layer, an output layer and one hidden layer. The hidden layer(s) and the output layer are the processing layers in the network where activation takes place. The knowledge of the network is encoded in the weights connecting the neurons. Each neuron in an inner layer acts as a linear combiner whose summation function is given as:

\[ \text{Sum} = W_0 + \sum_{i=1}^{n} w_i x_i \]

Where \( W_0 \) is the bias weight, \( w_i \) and \( x_i \) are the weight and input vectors respectively. The MLP is trained using the back propagation algorithm, which is a form of supervised learning, by presenting sample input-output pairs to the network. The error difference between the network’s output and the expected target output are fed back into the network for updating the weights connecting the hidden-output layers and the input-hidden layers. Next is symbolic data obtained from case instances are transformed into numerical data through pre-processing and fed as input into the MLP. The data pre-processing is therefore, a necessary precursor to the operations of the MLP-ANN component of the CBR system.

**CBR Component**: The set of input variable values and the predicted output obtained from the MLP component is passed to the CBR component. The CBR component has a case base indexed on unique case identity field (case_id) and the computed similarity score of each case.

a) Case_identify – which is able to be the auto-generated primary key of the table;
b) Case_similarity score – It is the computed similarity score of a case in the case base relative to a particular case instance;
c) Case_attribute set – input values of individual attributes which can be stored as a string separated by delimiters;
d) Case_sellingprice – the predicted valuation of a case instance related to selling price;
e) Case_weightSet – the set of weights associated with each attribute variable so that it represents the weight of the ith attribute.

Similarity analysis was computation is performed using the nearest neighbor algorithm. The similarity measure used was the inverse of weighted normalized Euclidian distance. A similarity score is computed by:

\[ \text{Similarity}(x, y) = 1 - \text{Dist}(x, y) \]

This calculation is repeated for every stored case in the case base. The cases with the highest similarity score are classified as candidates for adaptation in forecasting and predication of price in real estate market.

3. RESEARCH METHOD: COMBINED ANN CBR MODULE

A pilot prototype system is developed using data collected from a real estate data sets. The instance data will be used in the case study is for the certain past year period (for example: year 2017-2018). Data associated with nine core attribute features used in the appraisal of residential properties were extracted and used to train the neural network component to yield an estimate of the price of the property. Above figure illustrates the proposed prediction Architecture of ANN-CBR Module:

a) Data Gathering Module: The data will be collected from the real estate firm and feed as input to ANN CBR component. Input Determinants will be: Status of the Location, Year of the Property, Month Property was sold, Land Size, Neighborhood Gap, Accessibility, Bore hole, Electricity Supply, Condition and Usability of land.

b) Data Cleaning Module: The raw input data will be normalized based on min-max normalization to values between 0 and 1 using the data-preprocessing interface of the ANN-CBR system.

c) ANN training Module: The configuration of the MLP-ANN model used for our training instance is a 15-16-1 MLP in which the number of core attribute variables (15 of them) corresponds to the number of input neurons with one hidden layer containing 16 neurons and 1 neuron in the output layer, which returns as output the predicted sales price estimate. After the training experiment using back propagation, the MLP configuration was 15-16-1, with 1.0 learning rate and 0.005 threshold value. This was used for predicting the sale price of a property and the predicted result passed to the CBR component.

CBR Module: The case base was implemented with SQL Server database table of records (cases) indexed on case_identity and the computed weighted Case_similarity score fields. The CBR component does similarity analysis and uses parameterized SQL statements to determine the best-case matches. The procedure employed by the ANN-CBR system to reach its final conclusion.

*Proposed study on evaluating and forecasting the resident property value based on... (Sinan Adnan Diwan)
Validation Module: Validation of an expert system is conducted by determining whether the system's outcome is consistent with the conclusions of the human experts. Validation focuses on evaluating the outcomes rather than the process by which the outcomes are determined.

4. RESULTS AND DISCUSSION

The ANN and CBR was developed to fulfil the study’s objective, models were trained, cross validated and tested with 500 residential property datasets of real estate transactions from data. world [2017-2018]. There were initial 15 variables in the sample of data, only nine were used as input to the network which are as follows: Status of the Location, Year of the Property, Month Property was sold, Land Size, Neighborhood Gap, Accessibility, Bore hole, Electricity Supply, Condition and Usability of land. Based on the model validation, once the input is feed in to the model, forecast of real estate price are calculated for the year of 2017-2018. Regression outputs are calculated based on Mean Square Error (MSE), Mean Absolute Error (MAE) and Correlation and it is shown in the below Table 1 as an initial result.

<table>
<thead>
<tr>
<th>Training Data(%)</th>
<th>Testing Data(%)</th>
<th>$R^2$</th>
<th>$r^2$</th>
<th>Mean Square Error</th>
<th>Mean Absolute Error</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>85</td>
<td>15</td>
<td>0.971</td>
<td>0.992</td>
<td>0.851</td>
<td>0.651</td>
<td>0.997</td>
</tr>
<tr>
<td>75</td>
<td>20</td>
<td>0.992</td>
<td>0.921</td>
<td>0.921</td>
<td>0.756</td>
<td>0.996</td>
</tr>
<tr>
<td>70</td>
<td>30</td>
<td>0.991</td>
<td>0.995</td>
<td>0.922</td>
<td>0.762</td>
<td>0.992</td>
</tr>
<tr>
<td>90</td>
<td>10</td>
<td>0.959</td>
<td>0.974</td>
<td>3.872</td>
<td>1.421</td>
<td>0.985</td>
</tr>
</tbody>
</table>

The initial results of regression were compared with the real data to evaluate the performance of ANN and CBR models. The above results have shown that 80% of the data is selected for training and 20% of the data for testing with highest r value of 0.995 and lowest mean square error value of 0.851. The value of $R^2$ have shown that training and testing datasets are having closer value of 99.2% as of in actual values. The validations were verified by measuring the mean absolute error of the forecasting results. Past studies have observed that ANN and CBR have achieved within 10% is very good, while achieve 20%, it is considered as average. Current mean absolute error is within 10% and considered as good indicator. The training datasets in the model have been observed and shown that classifier network have performed well within the correlation value, $r$ which is between 0.99 and 0.98 and 0.97 and 0.99 for testing datasets. From the observation, highest correlation coefficient achieved for the network was with five hidden neurons with the value of 0.997 and mean square error value of 0.851 respectively. Results were achieved were at initial with training and testing of minimum datasets.

5. CONCLUSION

This research is fundamental and it is based on descriptive and explanatory research elements on real estate market by identifying the specific determinants, but it is specifically focused on predictive research by constructing a hybrid framework model. Research study has employed Artificial Neural Network and Case Base Reasoning as a hybrid model to forecast the house pricing model for real estate agents, professionals and builders. The analysis at initial level have shown that hybrid network model is a good alternative model compared to other machine learning prediction approaches for forecasting the house prices. The initial model has the capability to take nine important features to train in the input layer with optimal weights and produce forecast outcome with accuracy of 99%. it has the ability to map the difficult non-linear relationship between the house price which indicates the fluctuation in price and factors [i.e. determinants] influencing the price of the house. Compare to CBR, the ANN models have limitations such as limited output data for verification, but as a hybrid model, CBR have adjusted to produce accurate results. One of the challenges in this research, collecting house price index data, some of the public access datasets have only data for maximum 5 years. In future research work, more experimental studies will have conducted with higher volume of data with more dependent variables which will be added in the training of the forecast model and achieve higher accuracy.

REFERENCES


