Vol. 4, No. 8 Aug 2013 ISSN 2079-8407
Journal of Emerging Trends in Computing and Information Sciences

http://www.cisjournal.org

A Neural-CBR System for Real Property Valuation

[1] Adebola G. Musa, [2] Olawande Daramola, [3] Alfred Owoloko, [4] Oludayo Olugbara

[1] Dept. of Software Engineering, Tshwane University of Technology, South Africa
[2] Dept of Computer Science, Covenant University, Nigeria
[3] Dept of Mathematics, Covenant University, Nigeria
[4] Dept of Information Technology, Durban Univ. of Tech, South Africa

[1] adebola.musa@gmail.com, [2] olawande.daramola@covenantuniversity.edu.ng,
[3] alfred_owoloko@yahoo.co.uk, [4] oludayoo@dut.ac.za

ABSTRACT

In recent times, the application of artificial intelligence (AI) techniques for real property valuation has been on the increase. Some expert systems that leveraged on machine intelligence concepts include rule-based reasoning, case-based reasoning and artificial neural networks. These approaches have proved reliable thus far and in certain cases outperformed the use of statistical predictive models such as hedonic regression, logistic regression, and discriminant analysis. However, individual artificial intelligence approaches have their inherent limitations. These limitations hamper the quality of decision support they proffer when used alone for real property valuation. In this paper, we present a Neural-CBR system for real property valuation, which is based on a hybrid architecture that combines Artificial Neural Networks and Case-Based Reasoning techniques. An evaluation of the system was conducted and the experimental results revealed that the system has higher satisfactory level of performance when compared with individual Artificial Neural Network and Case-Based Reasoning systems.

Keywords: *Case-based reasoning, artificial neural networks, real property valuation, intelligent system*

1. INTRODUCTION

In recent years, the application of artificial intelligence (AI) approaches for real property valuation has been on the increase. Expert systems that leverage machine intelligence concepts such as rule-based reasoning [39, 29], case-based reasoning (CBR) [30], and artificial neural networks (ANN) [10, 15, 27, 43, 45] have been used. These AI approaches have been found to outperform traditional statistical approaches such as hedonic regression, logistic regression, and discriminant analysis, and very capable to complement the decision making process [37].

However, these AI approaches have their individual strengths and weaknesses, which inherently affect the quality of performance when used alone for real estate valuation. For example, in order to engage a rule-based expert system approach, optimal weights must be assigned to individual property attributes that are to be used for composing rules of the rule-base, by using standardized regression coefficients. However, a major challenge of a rule-based system is that these optimal weights derived from regression are not generalized, but rather are location dependent, therefore, the rules and weights must be updated regularly in order to sustain the relevance of the system, which is usually a very demanding task [12]. Data mining offers an alternative approach to developing intelligent systems for real property valuation but their viability is only guaranteed when there is a large pool of transaction data to work with, which may not exist or may be unreliable in some locations [26].

The use of Artificial Neural Networks (ANN) for the appraisal of real estate property is particularly prevalent [44, 2]. ANN could be defined as a group of simple interconnected units, called neurons that function together in parallel for the purpose of performing a common task. It is a model of computation that emulates the operational principles of the biological nervous systems by providing a mathematical equivalent of the combination of neurons connected in a network. The neurons of an ANN are linked with each other through connections. Each connection is assigned a weight that controls the flow of information among the neurons. Whenever data is fed into a neuron through the connections, it is summed up first and then gets transformed by an activation function. The outputs of this activation function are then sent to other neurons (for feed forward networks) or back to itself as input (for recurrent networks) [35]. ANN has very strong adaptive learning ability from which it derives its strong interpolative capability. This makes it very suitable for prediction, especially in instances of noisy data or incomplete data, which many other alternative prediction models are not able to handle [44]. However, ANN has very weak explanation mechanism, which makes it difficult to understand the reasoning behind its conclusions [2]. This is a major limitation particularly in the real estate domain where it is essential to have a strong rationale for making investment decisions.

CBR is an approach that entails the use of the experience gained in previous problem episodes to arrive at a solution for a new problem [1, 21]. It is a machine learning paradigm that closely models the human reasoning process. Solving a problem using CBR involves a number of processes: (1) case matching and retrieval of a relevant case using defined similarity metrics; (2) case adaptation for reuse; (3) case revision for appropriateness and; (4) case retention in the case base [2, 1]. The nature of CBR, which relates every new episode to similar past episodes, makes it very suitable for building intelligent systems with effective explanation mechanism. It has proved useful for real property valuation in some locations where transaction data is not readily available or is unreliable. The past experiences of experts have been used as a basis to implement a CBR system for decision support purposes [32]. CBR systems have very strong explanation mechanism because of the existence of sufficiently similar previous cases that provides good rationale for new solutions obtained. However, the disadvantage of CBR is that the quality of its solutions depends solely on the existence of good cases that are relevant to solving the new case at hand. This brings the tendency to overly rely on previous experiences without validating them in a new situation.

Our approach in this work, innovates the combination of ANN and CBR in a single system framework leveraging the strengths of the two instance-based learning techniques. The experienced-based problem solving capability of CBR systems and its viable explanation mechanism is combined with the strong interpolative capability of ANN in producing a Hybrid Intelligent System (HIS) for qualitative decision support for real property valuation. This, to the best of our knowledge represents a first attempt at hybridizing these two approaches in a practical scenario for improved decision support in the real estate domain. To achieve this, data from selected input variables of new cases are transformed via a pre-processing procedure into numerical data that are suitable for ANN computation, and 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 work 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.

The remaining part of the paper is described as follows: In section 2 we give an overview of related work, while Section 3 discusses the hybrid architecture of the Neural-CBR system. Section 4 is a case study report of the application of the Neural-CBR system to business data of properties sales obtained from a Nigerian company (Dan Odiete and Co. Ltd. based in Benin City, Nigeria). The paper is concluded in section 5 with a brief note.

2. RELATED WORK

A number of machine learning methods and techniques that are applicable to property appraisal and valuation have been reported in literature. Wilson et al. [44] reported the implementation of an intelligent system for valuation of residential property. The intelligent system was built using a hybrid of Multi-Layer Perceptron (MLP) ANN and rule-based expert system. The study by Guan et al [17] describes the design and implementation of an Adaptive Neuro-Fuzzy Inference System-based (ANFIS) approach to estimate prices for residential properties. The paper represents a first attempt to evaluate the feasibility and effectiveness of ANFIS in assessing real estate values. In [22], a multi-resolution approach was used to determine real estate price in the Chinese real estate market applying three theories: (1) Unascertained theory, (2) Principal Component Analysis - PCA and (3) Ant Colony Optimization ACO-based ANN. The result forecasted is in good agreement with the actual values, and have been very accurate and meet the actual needs. Lin and Chen [23] applied Back Propagation Neural Networks (BPN) and Support Vector Regression (SVR) to property valuation in Taiwan. The results of the BPN & SVR were compared. It was found that SVR with trial-and-error method performed the best with MAPE = 4.466% and R2 = 0.8540. That is, stepwise regression is efficient but not the best variable selection method with both BPN and SVR. Also ANN was used as a valuation technique in [10, 15, 27, 43, 45, 44, 8, 9, 33].

Most of the existing CBR systems that have been reported in literature find application in the fields of medicine, law, planning and design. These include CHEF [18], PESUADER [41], CABOT [11], GINA [14]), and CYRUS [20]. Relatively few CBR systems have been reported to have application in the real estate domain. However, PROFIT [6], is a Fuzzy CBR (FCBR) system for residential property valuation. It is an advanced prototype system developed to estimate residential property values for real estate transactions that was based on the use of CBR techniques with Fuzzy predicates. PROFIT has been successfully tested on thousands of real estate transactions. Also, Pacharavanich et al. [32] reported the application of a CBR tool for the valuation of residential property in Bangkok, also an evaluation of the CBR tool was conducted. Juan et al. [19] developed a “pre-sale housing”-based decision support system for the Taiwan real estate market using a hybrid of CBR and Genetic Algorithm (GA). Based on the customer’s needs, CBR is used to retrieve relevant housing layout. Out of the retrieved cases, nearest neighbour method was used to calculate similarity of cases. GA was then applied to optimize cost and housing conditions. Hybrid CBR systems are those that combine other forms of knowledge and reasoning methods with CBR. Examples include Fuzzy-CBR [7], rule-based and case-based [36, 16], combining case-based and model-based [34], case-based and inductive learning [13, 5, 25, 3]. Thus far, relatively few instances of Neural-CBR hybrids have been reported in the literature with no report of its application to the real estate domain. Although, Taffasse [42] discussed the prospects of Neural-CBR approach to real property appraisal, the paper did not report any implementation experience to practically validate the propositions made. Specifically, a combination of CBR and a Radial basis ANN in the implementation of a Sales-Advisory system was reported in [28].

In [24] an ongoing work on the development of hybrid Neural-CBR classifiers for building on-line communities was reported. The objective of the work is to identify communities of use in the context of an organized group of people. The process involves mining users bookmark files in order to identify communities that share the same information interests. An intelligent agent is used to observe user behavior in order to learn the user bookmark classification strategy before hybrid neural case-based reasoning component is used as incremental classifier. Also, Bajo et al. [4] reported the implementation of a Case-Based Planner for Monitoring Patients (CBPMP) system. It is an autonomous deliberative case-based planner designed to plan the nurses' working time dynamically, to maintain the standard working reports about the nurses' activities, and to guarantee that the patients assigned to the nurses are given the right care. The CBPMP was integrated with a Routing Problems with Time Windows (RPTW) neural network component in order to realize an intelligent environment for monitoring patients' health care in execution time in hospital environments. Hence, the contribution of this work stems from the novel hybridization of MLP-ANN and CBR in the implementation of a Neural-CBR decision support system in real estate property valuation.

3. THE HYBRID NEURAL-CBR SYSTEM

The Neural-CBR system is a hybrid modular architecture of two components with five user interfaces. The two components are the ANN component and the CBR component (see Figure 1).

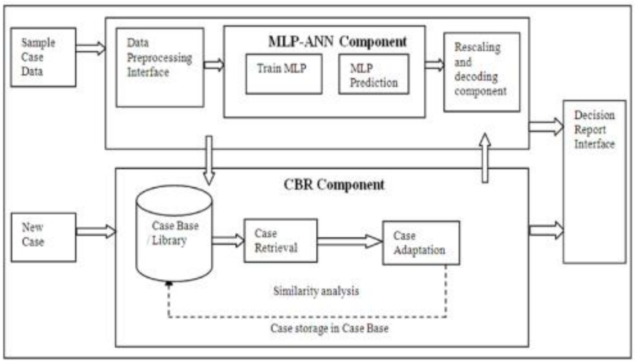

Fig 1: A Schematic View of the hybrid Neural-CBR Architecture

3.1 The Multi-Layer Perceptron 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:

$$\text{Sum} = w_0 + \sum_{j=1}^{n} w_i \; x_i \qquad (1)$$

Where w_0 is the bias weight, w_i and x_i are the weight and input vectors respectively. The activation at each neuron is given by the sigmoid function:

$$f(a) = \frac{1}{(1+e^{-sum})} \qquad (2)$$

The sigmoid function is continuous and differentiable in the interval [0, 1], and is one of mostly used activation function for the MLP. The MLP is trained using the back propagation algorithm [35], 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.

In our Neural-CBR system, 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 Neural-CBR system.

3.2 The 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. The typical structure of a case consists of the following:

- Case_id – which is an auto-generated primary key of the table;
- Case_simscore – which is the computed similarity score of a case in the case base relative to a particular case instance;
- Case_attributesset – input values of individual attributes stored as a string separated by delimiters;
- Case_sellingprice – the predicted valuation of a case instance;
- Case_weightSet – the set of weights associated with each attribute variable such that w_i represents the weight of the i^{th} attribute.

3.2.1 Similarity Analysis

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

$$\text{SIM}(X, Y) = 1\text{- DIST}(X,Y) \qquad (3)$$

$$\text{DIST}(X, Y) = \sqrt{\textstyle\sum_i^n w_i (x_i^o - x_i^r)^2} \qquad (4)$$

Where X and Y are the new and stored case respectively with n number of attributes while x_i^o and x_i^r are the normalized values for the i^{th} attribute. A normalized weight w_i is assigned to each attribute based on contextual experts' knowledge of the location concerned. This calculation is repeated for every stored case in the case base. The cases with the highest similarity score (that is up to or above the similarity benchmark value) are picked as candidates for adaptation in providing explanation for the new case scenario. The algorithm of the Neural-CBR system is given in Figure 2.

```
Algorithm Neural-CBR
1. Input: case-base CB, probe-case Cnew;
2. Output: updated CB, property valuation Pvalue, Explanation Expl;
3. SimBenchmark←t (0.75 ≥ t ≤ 1.0); {Initialize minimum acceptable degree for similarity}
4. Preprocess(Cnew);
5. MLP-Train(train-data); {Train MLP-ANN on training data}
6. Pneural ←MLP-Predict(Cnew); {ANN prediction of Pvalue}
7. Simcases[ ]←ComputeSimscore(Cnew); {One pass computation of similarity of cases to Cnew}
8. If ThereExist (cases with similarity ≥ SimBenchmark) then {Sufficient similarity}
     Pcbr ← doCaseAdaptation(Simcases[ ], Cnew); {Adapt and reuse old similar cases for Cnew}
9.   Pvalue← Pcbr; {Case based prediction of Pvalue}
     Expl← getCBexplanation(Simcases[ ],Cnew); {Get explanation from CB for Cnew}
10.  CB← Savecase(CB, Cnew); {Retain Cnew}
11. else {Less sufficient enough}
12. If ThereExist (cases with similarity ≥ 0.5 < SimBenchmark) then
13.  Pcbr←doNeuro-CaseAdaptation(Simcases [ ], Cnew, Pneural); {Use Pneural in case adaptation}
14.  Pvalue← Pcbr; {Case-trained Neuro prediction of Pvalue}
     Expl←getSomeCBexplanation(Simcases[ ],Cnew); {Get some explanation from CB for Cnew}
15.  CB← Savecase(CB, Cnew); {Retain Cnew }
16. else {Not enough similarity}
17.  Pvalue←Pneural; {Take Pneural as final result}
18.  Pcbr← Pneural; {Adapt Pneural as solution for Cnew}
19.  Expl←getNewexplanation(Cnew); {Use Cnew for explanation}
20.  CB← Savecase(CB, Cnew); {Retain}
21. Return (CB, Pvalue, Expl)
```

Fig 2: Neural-CBR System's Operational Procedure

3.3 Algorithm of Neural-CBR System

In this section, we present the algorithm of the Neural-CBR systems that details how it reaches its conclusions (see Figure 2). Given an input C_{new} and the existence of the case base CB. A variable Sim Benchmark is set as the minimum acceptable value for sufficient similarity between cases. The algorithm selects the SimBenchmark value to be in the interval [0.75, 1.0], such that the value 0.75 implies a degree of similarity in the upper quartile while the value 1.0 denotes perfect similarity. Likewise, the value 0.5 connotes an average similarity score. Step 4, 5, 6 in the algorithm represents the pre-processing, ANN training and ANN prediction phases of the system's operation. In Step 7, a one pass scan of the case base (CB) was carried out to compute similarity between C_{new} and the existing cases in the case base using the Weighted Euclidean distance (see equations 1, 2). If cases with similarity score up to or above the SimBenchmark exist (i.e. similarity ≥ SimBenchmark) then case adaptation is conducted as follows:

i. Rank all cases found by their similarity score;
ii. Group the retrieved cases based on their solution values P_{cbr} (case selling price) into different categories, g_1, g_2, g_3..., g_k, with their corresponding P_{cbr} as P_1, P_2,.., P_k;
iii. Take count of the number of cases retrieved in each category g_i (i = 1...k) and store them as t_1, t_2...t_k);
iv. Choose category g_i with the highest frequency;
v. If there exist only one category g_i with highest frequency then take P_{cbr} of a case in g_i as P_{value} (i.e. final output) else, if there is more than one category g_i with highest frequency, such as g_{i1}, g_{i2}... g_{im} then compute average value P_{ave} of all P_{im} in g_i as P_{value} (final output) and as P_{cbr} for C_{new} .
vi. Next, use descriptions of case attributes (Case_attributesset) in category g_i as explanation for C_{new} (line 9-12).

If cases with similarity score of at least 0.5, but less than SimBenchmark exists then the Neural-CBR system uses the neural computed P_{neural} (ANN predicted output) in case adaptation thus:

1) Use retrieved cases to retrain ANN-MLP in order to determine P_{neural};
2) Assign P_{neural} as solution for P_{cbr} and also as P_{value} in this case;
3) Use descriptions of attributes of retrieved cases that are closest to corresponding C_{new} variables for explanation. Where no sufficiently similar case is found, then restate description attributes of C_{new} for explanation (see line 15-18) i.e. for all retrieved cases $C_r = \{c_1, c_2 \ldots c_n\}$ where f_i (i=1...15) is a specific attribute feature of $C_k \in C_r$, f_j a corresponding attribute feature of C_{new}, w_i, the weight associated with attribute feature f_i, compute $SIM(f_i, f_j) = |w_i f_i - w_i f_j|$ and rank cases where $SIM(f_i, f_j) \leq 0.5$. Pick as explanation for C_{new} highest ranked f_i in C_r, for each instance where $SIM(f_i, f_j) \leq 0.5$ is not found use the description of f_j in C_{new} as explanation for C_{new}. end for

If sufficiently similar cases do not exist then the system takes P_{neural} as P_{value}. However in this case it restates the features of C_{new} as explanation for output (line 20-21).

4. THE CASE STUDY

A prototype system was developed using data collected from a real estate firm (Dan Odiete and Co. Ltd) in Nigeria [31]. The instance data used in the case study is for the periods of 2002 to 2005, a sample is shown in Table 1. Data associated with fifteen 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. The features, description and range of values obtained from the training set used are given in Table 2.

Table 1: Bedroom Bungalow BQ optional

Loc	Year Built	Month Sold	Land size	Bore-hole Facilities	Fenced round	BQ	Neighbourhood group	Accessibility	Master Room	No of Other Rooms	No. of convenience	Recreation space	CDU	Plumbing features	Sales (N)
0.7	1986	10	1000 sqm	A	A	A	B	A	3	-	1	150	4	10	8M
0.7	1991	13	1500 sqm	A	A	A	B	A	3	-	1	250	4	10	7.5M
0.7	2000	12	1000 sqm	A	A	B	C	A	2	1	1	150	5	9	5M
0.7	2000	10	1000 sqm	A	A	A	C	B	2	1	1	150	5	9	4.5M
0.7	1999	14	1000 sqm	A	A	B	C	B	3	-	1	150	5	9	4.0M
0.7	1980	12	1500 sqm	A	A	A	A	A	3	-	1	200	3	9	8M
0.7	2000	13	1000 sqm	A	A	A	C	A	3	-	1	150	4	9	4.5M
0.7	1989	11	1000 sqm	A	A	B	B	B	2	1	1	150	4	9	6.0M
0.7	1985	13	1500 sqm	B	A	A	A	A	3	-	1	200	4	10	6.50M
0.7	2001	8	1000 sqm	A	A	A	C	A	3	-	1	150	4	9	5.5M

Table 2: Description of features for real property valuation

Feature Code	Feature name	Description	Range of Values
F1	Loc	Location status (a measure of the relative pricing based on location)	0-1
F2	Year_built	Year property was built	1975-2005
F3	Month_sold	Month of the year property was sold	1-12
F4	Land size	Size of the land (in square feet)	200 – 2000
F5	Bore-hole	Availability of Bore-hole	A or B
F6	Fenced_round	Whether property is fenced	A or B
F7	BQ	Availability of Boys quarter	A or B
F8	Neighbourhood_gp	The kind of neighbourhood where property exist, either high class, medium or low class	A,B or C
F9	Accessibility	Accessibility of property (good or poor)	A or B
F10	Master's room	Number of master bed room unit	0-5
F11	Other rooms	Number of other rooms	0-3
F12	No. of convenience	Number of toilets/ bathrooms	0-2
F13	Recreation space	The measure of recreation space available	0-650
F14	CDU	Condition/ desirability/ usefulness	0-5
F15	Plumbing features	Number of plumbing features	5-20

4.1 Data-Preprocessing

The raw input data (see Table 1) were normalized based on min-max normalization [40] to values between 0 and 1 using the data-preprocessing interface of the Neural-CBR system. The rescaled values of the attribute features extracted from the data are as shown in Table 3. All the values are numerical except the following: borehole, BQ, fenced_round, neighbourhood_group and accessibility. These five were category inputs, therefore were represented as A (available) or B (not available), which are assigned vales 1 for A and 0 for B. In the case of neighbourhood_group, A (high), B (medium), or C (low) are assigned values 1, 0.5 and 0 respectively.

Table 3: Showing normalized feature values from Table 1

F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	Sales (₦)
0.7	0.293	0.4348	0.4444	1.000	1	1	1.0000	1.0000	0.8000	0	0.5	0.2308	0.800	0.533	0.7500
0.7	0.1613	0.5217	0.4444	1.000	1	1	1.0000	1.0000	0.800	0	0.3333	0.2304	0.6000	0.5333	0.5417
0.7	0.4839	0.8696	0.5556	1.0000	1	1	0.500	1.0000	0.8000	1	0.3333	0.3077	1.0000	0.533	0.3750
0.7	0.2258	0.3478	0.4444	0.0000	1	0	1.0000	0.0000	0.8000	1	0.333	0.3077	0.8000	0.4667	0.6250
0.7	0.6774	0.4783	0.1666	1.0000	1	0	0.0000	0.0000	0.8000	0	0.3333	0.1539	0.8000	0.5333	0.0833
0.7	0.7419	0.4348	0.4444	1.0000	1	0	0.000	1.0000	0.8000	0	0.3333	0.2308	0.8000	0.5333	0.1667
0.7	0.5161	0.6522	0.4444	1.0000	1	1	0.5000	1.0000	0.8000	0	0.3333	0.1530	0.8000	0.4667	0.3750
0.7	0.4516	0.0870	0.4444	0.0000	1	0	0.5000	1.0000	0.8000	1	0.3333	0.3077	1.000	0.4667	0.4167
0.7	0.2258	0.6522	0.5556	1.0000	1	1	1.0000	1.0000	0.8000	0	0.3333	0.2308	0.8000	0.5333	0.5417
0.7	0.7097	0.9130	0.4444	1.0000	1	1	0.5000	1.0000	0.8000	0	0.3333	0.1539	0.8000	0.2667	0.125

4.2 Training the MLP-ANN

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. The Neural-CBR system environment allows for the specification of a desired configuration for the MLP network, which is then dynamically created. The MLP was trained using the back propagation algorithm with three sets of data, the training set, the validation set (to verify correctness during training) and the test set. The summary of the ANN experiments that was conducted with the Neural-CBR system framework in order to arrive at the 15-16-1 MLP configuration and other optimal training parameters such as number of neurons in the hidden layer, learning rate and threshold are presented in Tables 4, 5 and 6. In each occasion recordings were taken and the average computed to determine the optimum value in each case.

Table 4: Variation of neurons in the hidden layer of MLP

Parameters
Learning Rate - 0.35
Threshold - 0.005
Maximum epoch - 25,000
N - number of units in the input layer (15)

No. of Hidden neurons	Number of Epoch at Convergence					Average
N-1	11303	11691	11599	11148	12340	11616
N	11909	10853	12794	11690	12011	11851
N+1	11147	10652	12643	10791	11681	11383
N+2	13083	11241	11384	12239	11336	11857

Table 5: Variation of learning rate

Parameters
Threshold - 0.005
Maximum epoch - 30,000
Number of units in the input layer - N+1

Learning Rate	Number of Epoch in Each Experiment					Average
0.15	27477	27036	26371	28209	27094	27237
0.25	16593	15954	17070	16249	15674	16308
0.35	11528	11264	11741	10882	11822	11447
0.45	9563	9088	8748	9238	9677	9263
0.55	6936	7855	7193	7804	7594	7476
0.65	6065	5893	6121	5946	6360	6077
1.0	3935	4085	3782	4074	4197	4015
2.0	7194	8380	5561	6311	7291	6947

Table 6: Variation of threshold

Parameters
Learning Rate - 1.0
Maximum epoch - 30,000
Number of units in the input layer - N+1

Threshold	Number of Epoch in Each Experiment		Average
0.05	446	477	462
0.005	4288	4111	4200
0.0005	30000	30000	30000

After the training experiment, 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.

4.3 Implementation of the Neural-CBR System

The Borland C++ Builder version 6.0 software was used as the programming platform to realize the Neural-CBR system. The case base was implemented as SQL Server database table of records (cases) indexed on case_id and the computed weighted Case_simscore fields. The CBR component does similarity analysis and uses parameterized SQL statements to determine the best-case

Vol. 4, No. 8 Aug 2013 ISSN 2079-8407
Journal of Emerging Trends in Computing and Information Sciences

http://www.cisjournal.org

matches. The procedure employed by the Neural-CBR system to reach its final conclusion is based on the algorithm shown in figure 2. Figures 3-5 show

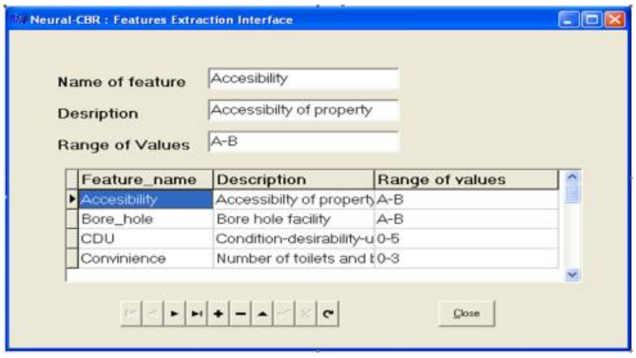

Fig 3: Feature Extraction Interface of the Neural-CBR System

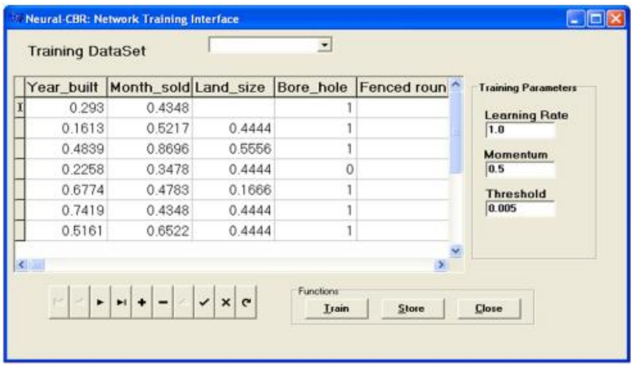

Fig 4: Network Training Interface of the Neural-CBR System

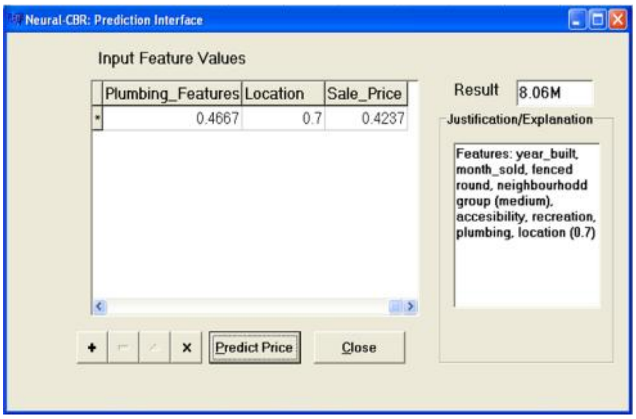

Fig 5: Prediction and Explanation Interface of the Neural-CBR System

4.4 Performance Evaluation of Neural-CBR System

Evaluation is the process of determining the appropriateness of a specific system relative to its functional requirements and objectives. 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. In our specific case, the output of the Neural-CBR system was validated by using a direct method [38]. In the direct method of expert system validation a human expert does a quantitative assessment of the expert system software by engaging it to perform a simple benchmark problem over a specified period. The expert then responds to a set of questions about the software based on past experience. The questions are quantitative and are based on a 0 (very false) to 5 (very true) numerical scales. Thereafter, a single numerical factor results called the Satisfaction Level that ranges from 0 (least satisfied user) to 5 (most satisfied user) is computed, which is used to rank the expert system software in terms of its likelihood to satisfy a prospective end user.

4.4.1 Description of the Evaluation Experiment

The objective of the validation experiment was to determine the level of users' satisfaction with the hybrid Neural-CBR system relative to CBR alone and ANN alone systems. In order to do this, a real property valuation system was created that allows the user to alternate between three system modalities, Neural-CBR, CBR and ANN, such that, if one of the modality is activated the other two modalities are automatically disabled. A comparative assessment of the three system modalities was then undertaken using some selected human experts from the real estate industry for the evaluation. The participants in the experiment were persons with significant professional experience in the real property valuation. Fifteen (15) real estate experts drawn from two firms, Dan Odiete & Co. and Ajeb Associates (both located in Benin City, Nigeria) with varying professional experience ranging from 2 (lowest) to 15 (highest) years were selected to participate in the user-based system usability experiment. Each received a copy of a questionnaire instrument and had the system to be evaluated installed for them. The participants who were people with tangible knowledge of the use of software were given one-week training on how to make use the system prior to the commencement of the experiment. Details of how to switch modalities and to engage the system in specific modality for operations such as data preprocessing, training, and prediction were clearly outlined. In addition, the participants had the license to engage the system in as many trial sessions as deemed convenient commencing with the actual assessment. Moreover, one IT support staff was temporarily assigned to each company for the trial period of three weeks. The human experts were asked to give a rating of 0 to 5 of their assessment of the performance of the three (Neural-CBR, CBR, ANN) system modalities covering seven aspects and a total of sixteen questions. Each question has specific weight assigned to it, which had the mutual consent of human experts that participated in the experimental process. Table 7 shows a sample of a typical response to the questionnaire test instrument and how the satisfaction score for an evaluator is computed.

Table 7: A Typical System Evaluation Template

	Question	Assessment Value (0-5)	Weight	Value X Weight
	Correctness of Answer			
1	Is there enough information to evaluate the system?	4	(2)	8
2	Does the system reach the same conclusion similar to that of a human expert?	5	(2)	10
3	Does the system provide the right answer for the right reasons?	5	(2)	10
	Accuracy of Answer			
4	Is the system accurate in its answer(s)?	5	(2)	10
5	Is the answer complete? Does the user need to do additional work to get a usable result?	4	(2)	8
	Correctness of reasoning technique			
6	Does the answer change if new but irrelevant data is entered into the system?	5	(1)	5
7	Is the explanation given for the system's output acceptable in most cases?	4	(2)	8
	Sensitivity			
8	Does the answer change if relevant changes are made to the system data?	5	(1)	5
	Reliability			
9	Does the system crash or hang up in its host computer?	2	(1)	2
10	How well does the system handle instances of incomplete data or missing data?	5	(1)	5
	Confidence			
11	Are you comfortable using the system?	4	(1)	4
12	Does the conclusion of the system give adequate satisfaction?	4	(2)	8
13	Do you consider the system a credible means of decision support?	4	(2)	8
14	To what extent would you trust the output of the system?	4	(1)	4
	Limitations			
15	Can limitations of the system be detected at this point in time?	4	(1)	4
16	Can the system learn from increased data or experience?	2	(1)	2
	Result = Σ (weight x value) / Σ (weight)		24	101
			4.21	

4.5 Results and Discussion

At the end of the evaluation experiment, the mean satisfaction level as computed from the assessments of the fifteen real estate expert evaluators for the three systems are as shown in Table 8. The Neural-CBR system had a mean score of 3.83/5.0 (viz. 3.83 out of the possible maximum score of 5.0); CBR system had 3.64/5.0; and ANN system 3.75/5.0; all which are indicative of good performance.

Table 8: Result of Evaluation Experiment

Evaluator	Computed Satisfaction Level		
Variant types	**CBR**	**Neural-CBR**	**ANN**
1	3.9	4.19	4.1
2	3.60	3.7	3.60
3	3.70	3.9	3.93
4	3.42	3.62	3.60
5	3.25	3.6	3.52
6	3.60	3.80	3.70
7	3.78	3.82	3.78
8	3.60	3.89	3.80
9	3.71	3.80	3.70
10	3.56	3.76	3.56
11	3.61	3.74	3.61
12	3.70	3.80	3.82
13	3.60	3.90	3.80
14	3.72	3.80	3.70
15	3.80	4.10	4.0
Mean Satisfaction Level	**3.64**	**3.83**	**3.75**

However, in order to determine the best of the three systems; we compared the computed mean satisfaction level of the three systems to determine whether the differences in the mean values are statistically significant. To achieve this we formulated the following hypothesis:

Hypothesis one

Vol. 4, No. 8 Aug 2013 ISSN 2079-8407
Journal of Emerging Trends in Computing and Information Sciences

http://www.cisjournal.org

H_0: There is no significant difference in the mean satisfaction level of the CBR, Neural-CBR, and ANN systems, and hence the three systems are at par performance wise.

H_1: There is significant difference in the mean satisfaction level of the CBR, Neural-CBR, and ANN system, and the Neural-CBR systems is better than the other two systems.

Testing the hypothesis

In order to test the hypothesis that was formulated, the Analysis of Variance (ANOVA) statistic was employed to compare the three sets of data obtained from the evaluation experiment. The coefficient of variation (CV) (see Table 10) of the sets of data was computed in order to determine the system with the best rating distribution. This is given as:

$$C_v = \frac{\sigma}{\mu} \tag{5}$$

Where σ = standard deviation of the data distribution, μ = mean of the data distribution

At 1% significance level (i.e. $p < 0.01$), it was found that the mean satisfaction levels of the three systems (ANN, Neural-CBR and CBR) systems are significantly different because the p-value of 0.008305 from the ANOVA test is less than 0.01 (see Table 9). In addition, the coefficient of variation (CV) for the hybrid Neural-CBR system is the lowest when compared to the other two systems, which is indicative of a relatively better user rating. Therefore, H_0 is rejected and H_1 accepted, which states that there is significant difference in the mean satisfaction level of ANN, Neural-CBR and CBR systems and the Neural-CBR system is better than the other two systems.

Table 9: ANOVA Table for Mean Satisfaction Level Comparison

Anova: Comparison of the Mean of the three system							
SUMMARY							
Groups	*Count*	*Sum*	*Average*	*Variance*			
CBR	15	54.55	3.636667	0.024667			
Neural-CBR	15	57.42	3.828	0.024746			
ANN	15	56.22	3.748	0.027803			
ANOVA							
Source of Variation	*SS*	*Df*	*MS*	*F*	*P-value*	*F crit*	
Between Groups	0.277018	2	0.138509	5.381408	**0.008305**	5.149139	
Within Groups	1.081013	42	0.025738				
Total	1.358031	44					

Table 10: Coefficient of Variation for three Systems

Systems	CBR	Neural-CBR	ANN
Mean	3.636667	3.828	3.748
Standard deviation	0.157056	0.157307706	0.166742
Coefficient of Variation	0.043187	0.041093967	0.044488

Generally the observations from the case study reveal potential benefits of the novel hybridization of ANN and CBR. First, the output of the system shows the interpolative power of ANN especially in instances where other predictive models may be deficient. Second, the results from the system reveal how well the ANN component could make up for the limitation of the CBR component in instances when there is lack of sufficiently similar old cases in the case base for a new case. At the same time the Neural-CBR system leverages the existence of a case base, in providing justifiable explanation for results instead of being a black box like a typical ANN. Additionally, the provisioning of rich GUI interfaces for preprocessing and training of ANN enables real-time acquisition of expert knowledge in the process of solving a problem such as being able to assign weights to specific real property attribute variables, which makes the system very adaptive and suitable as a decision support tool.

5. CONCLUSION

In this work, a novel hybridization of ANN and CBR techniques for real estate property valuation has been demonstrated. A prototype Neural-CBR system was developed and evaluated in a case study to confirm the viability of the concept. The result obtained revealed that the Neural-CBR combination offers more acceptable level of usability and performance satisfaction relative to solitary ANN systems and CBR systems. Also, the system showed significant strength in key areas of weaknesses usually associated with solitary ANN and CBR intelligent systems, which gives merit to the hybridization. In future work, we intend to investigate the applicability of hybrid Neural-CBR systems to other business application domains such as education, health, and finance where the potential of Neural-CBR is yet to be fully explored.

REFERENCES

[1] Aamodt A and Plaza “Case-based reasoning: foundational issues, methodological variations, and system approaches”. AI Communications. 7(1):39-59. (1994)

[2] Almond N, Lewis O, Jenkins D, Gronow S and Ware A “Intelligent systems for the valuation of residential property”. RICS Cutting Edge, Conference, Dublin 5-6 September. (1997)

[3] Armengol E. and Plaza E. “Integrating induction in a case-based reasoned”. Proceedings Second

European Workshop on Case-Based Reasoning: 243-252. (1994)

[4] Bajo J, de Paz JF , de Paz Y , Corchado JM "Integrating case-based planning and RPTW neural networks to construct an intelligent environment for health care", Expert Systems with Applications 36(3): 5844-5858. (2009)

[5] Bamberger, S and Goos K "Integration of case-based reasoning and inductive learning methods", Proceedings First European Workshop on Case-Based Reasoning, ed. Richter, Wess, Althoff, and Maurer, 2:296-300. (1993)

[6] Bonissone P and Cheetham W "Fuzzy Case-Based Reasoning for Residential Property Valuation". Handbook on Fuzzy Computing (G 15.1), Oxford University Press. (1998)

[7] Bonissone P, Mantaras R "Fuzzy Case-Based Reasoning Systems". Handbook of Fuzzy Computing Section F4.3, Ruspini, Bonissone, Pedrycz (Eds.), Institute of Physics Publishers. (1998)

[8] Borst RA "Artificial Neural Networks: The Next Modeling/Calibration Technology for the Assessment community?" Property Tax Journal, 10(1):69-94. (1991)

[9] Borst RA "A Method for the Valuation of Residential Properties using Artificial Neural Networks in Conjunction with Geographical Information Systems". IAAO Conference, Dublin. (1994)

[10] Borst R "Artificial neural networks in mass appraisal". Journal of Property Tax Assessment & Administration 1(2):5-15. (1995)

[11] Callan JP, Fawcett TE, Rissland EL "CABOT: An adaptive approach to case-based Search", Proceedings IJCA. (1991)

[12] Churbuck DC "Learning by example". Forbes, 6/8/92, 149(2):130-132. (1992)

[13] Connolly D and Christey S "Learning representation by integrating case-based and inductive learning", Proceedings AAAI Case-Based Reasoning Workshop: 157, Washington, DC. (1993)

[14] De Jong, KA and Schultz AC "Using experience-based learning in game playing". Proceedings of Fifth International Conference on Machine Learning: 284-290, Ann Arbor, MI, Morgan Kaufmann. (1988)

[15] Evans A, James H and Collins A "Artificial Neural Networks: an Application to Residential Valuation in the UK". Journal of Property Valuation & Investment 11:195-204. (1993)

[16] Gayer, G, Gilboa, I. and Lieberman O "Rule-Based and Case-Based Reasoning in Housing Prices". The B.E. Journal of Theoretical Economics, Vol. 7, Issue 1 (2007)

[17] Guan, J, Zurada, J and Levitan, A S "An Adaptive Neuro-Fuzzy Inference System Based Approach to Real Estate Property Assessment". Journal of Real Estate Research". Vol. 30, No. 4. (2008)

[18] Hammond K "Learning to anticipate and avoid problems through the explanation of failures". Proceedings of Fifth International Conference on Artificial Intelligence:556-560, Philadelphia, PA., Morgan Kaufmann. (1986)

[19] Juan, Y, Shih, Shen and Perng, Y "Decision support for housing customization: A hybrid approach using case-based reasoning and genetic algorithm". Expert Systems with Applications, Elsevier, Vol. 31, Issue 1, pp 83–93 (2006)

[20] Koldoner J "Maintaining organization in a dynamic long-term memory". Cognitive Science, Vol. 7(4): 281-328. (1983)

[21] Koldoner J "An introduction to case-based reasoning". Artificial Intelligence Review 6(1):3-34. (1992)

[22] Li W and Shi H "Applying Unascertained Theory, Principal Component Analysis and ACO-based Artificial Neural Networks for Real Estate Price Determination". Journal of Software, Vol. 6, No. 9, pp. 1672-1679. (2011)

[23] Lin, H and Chen, K "Predicting Price of Taiwan Real Estates By Neural Networks and Support Vector Regression". In Proc of the 15th WSEAS international Conference. (2010)

[24] Malek M and Kanawati R "A Cooperating Hybrid Neural-CBR Classifiers for Building On-line Communities". www.aic.nrl.navy.mil/papers/2001/AIC-01-003/ws5/ws5toc8.PDF. (2009)

[25] Manago M, Althoff K., Auriol E, Traphoner R, Wess S, Conruyt N. and Maurer F. "Induction and reasoning from cases". Proceedings First European Workshop on Case-Based Reasoning, ed. Richter, Wess, Althoff and Maurer 2:313-318. (1993)

[26] McCluskey W and Anand S "The application of intelligent hybrid techniques for the mass appraisal

of residential properties". Journal of Property Investment and Finance 17(3):218-238. (1999)

[27] McCluskey W, Dyson K, McFall D & Anand S "Mass Appraisal for Property Taxation: An Artificial Intelligence Approach". Land Economic Review 2(1):25-32. (1996)

[28] Murray-Smith R and Thakar S "Combining case-based reasoning with neural networks". AAAI Workshop on AI in Service and support, Washington. (1993)

[29] Nawawi AH, Jenkins D and Gronow S "Expert system development for the mass appraisal of commercial property in Malaysia". Journal of the Society of Surveying Technicians 18(8): 66-72. (1997)

[30] O'Roarty B, Patterson D, McGreal WS, Adair AS. "A case based reasoning approach to the selection of comparable evidence for retail rent determination". Expert System with Application 12(4): 417-428. (1997)

[31] Owoloko EA "Neural Networks: A data mining technique for property appraisal". M.Sc. Thesis (Unpublished), University of Benin, Nigeria. (2005)

[32] Pacharavanich P and Wongpinunwatana N "The Development of a Case-Based Reasoning System as a Tool for Residential Valuation in Bangkok". Proc. of the 6th Annual Pacific-Rim Real Estate Society Conference:1-14, Sydney. (2000)

[33] Peterson, S and Flanagan, A B "Neural Network Hedonic pricing models in mass real estate appraisal". Journal of Real Estate Research, Vol. 31, No. 2, pp 147-164. (2009)

[34] Portinale L, Torasso P, Ortalda C and Giardino A "Using case-based reasoning to focus model-based diagnostic problem solving". Proceedings First European Workshop on Case-Based Reasoning, ed. Richter,Wess, Althoff and Maurer 2:335-340. (1993)

[35] Rich E and Knight K "Artificial Intelligence 2nd ed". McGraw-Hill, Inc. (1991)

[36] Rissland E and Skalak D "CABARET: Rule interpretation in a hybrid architecture", Int. J. of Man-Machine Studies, 34(6): 839-887. (1991)

[37] Rossini PA, "Accuracy Issues for Automated and Artificial Intelligent Residential Valuation Systems", International Real Estate Society Conference, Kuala Lumpur, January 26-30. (1999)

[38] Salim, MD, Villavicencio A and Timmerman MA "A method for evaluating expert system shells for classroom instruction". Journal of Industrial Technology, Vol. 19, No. 1, pp. 1-11. (2002)

[39] Scott I and Gronow S "Valuation expertise: its nature and application". Journal of Valuation 8(4):362-375. (1989)

[40] Shalabi LA, Shaaban Z and Kasasbeh B "Data Mining: A Preprocessing Engine". Journal of Computer Science. 2(9): 735-739. (2006)

[41] Sycara KP "Using case-based reasoning for plan adaptation and repair". Proceedings of DARPA workshop on case-based reasoning. San Mateo, Calif. Morgan Kaufmann. (1988)

[42] Taffese WZ "Case-based reasoning and neural networks for real estate valuation". Proceedings of the 25th IASTED International Multi-Conference: artificial intelligence and applications, Innsbruck, Austria: 98-104. (2007)

[43] Tay D and Ho D "Intelligent Mass Appraisal". Journal of Property Tax Assessment & Administration 1(1): 5-25. (1994)

[44] Wilson ID, Paris SD, Ware JA and Jenkins DH "Residential property price time series forecasting with neural networks". Knowledge Based Systems 15(5–6): 335–341. (2002)

[45] Worzala E, Lenk, M., & Silva A "An Exploration of neural networks and its application to real estate valuation". Journal of Real Estate Research 10(2):185-201.

AUTHOR PROFILES

Adebola G. Musa is a doctoral candidate at Tshwane University of Technology, South Africa. He earns MS and BS Computer Science at Covenant University and University of Ilorin, Nigeria respectively. His research interests include the Semantic Web, Artificial Intelligence and Neural Networks, e-Government. He has to his credit several research articles in conference proceedings and refereed journals. He is a member of IITPSA.

Vol. 4, No. 8 Aug 2013 ISSN 2079-8407
Journal of Emerging Trends in Computing and Information Sciences

http://www.cisjournal.org

Dr. Olawande Daramola is a Senior Faculty at Department of Computer and Information Sciences, Covenant University Nigeria. His research interests include the Semantic Web, Intelligent Systems, and Automated Software Engineering. He is widely published in top range journals and conferences. He is a member of ACM, IEEE, and the Nigerian Computer Society.

Owoloko E. Alfred is a doctoral candidate at Covenant University, Ota, Nigeria. He earned his MS and BS from the University of Benin, Nigeria. His research areas include stochastic differential equation to valuation, application of Neural Networks to solving real life problems, mathematical statistics and operation research. He has to his credit several research articles in refereed journals.

Olugbara O.O. is a Professor of Computer Science and Information Technology at Durban University of Technology, Durban, South Africa.