IJCSN - International Journal of Computer Science and Network, Volume 8, Issue 3, June 2019
ISSN (Online) : 2277-5420
www.IJCSN.org
Impact Factor: 1.5

Machine Learning Approach to House Price Prediction with Ensemble Model

[1] V Manju; [2] Stephie Sara Philipose; [3] Dr. Radhakrishnan B

[1] CSE department, APJ Abdul Kalam Technological university, BMCE, Sasthamcotta, Kerala, India

[2] CSE department, APJ Abdul Kalam Technological university, BMCE, Sasthamcotta, Kerala, India

[3]CSE department, University of Kerala, BMCE, Sasthamcotta, Kerala, India

Abstract - House is a place of residence, accommodation or shelter, determining the price of house is challenging, as it is reliant on a number of features. In this paper we propose a model to predict the house price considering the environmental and housing characteristics. The proposed approach is designed to estimate the price of house and land using a neural network ensemble model. The neural network ensemble model is an architecture designed to process the data parallel with two different mechanisms namely recurrent, to estimate time series data and cascading to estimate the difference or deviation in the data with respect to time. The proposed system takes data like house data, location data, house and location demand etc. The system finds or extracts time related, time and location based price distribution from the input data, during the training process, and test data or evaluation data is evaluated against the trained model. The proposed system offers better accuracy which is more than the current systems.

Keywords - *NJOP, Residual Neural Network, Recurrent Neural Network, Meta-learner, Ensemble learning*

1. Introduction

House is a place of residence, accommodation or shelter, determining the price of house is challenging, as it is reliant on a number of features. House price prediction can help the developer determine the selling price of a house and can help the customer to arrange the right time to purchase a house. In housing market, the initial prices of buildings are an important factor in selling or buying houses and land .various factors of the house has to be considered in determining the right price of land and house. In general the factors affecting the price are both the residential factors and the geographic location of the land.

There is a need to determine the land or house price such that the buyers or sellers can ensure that they are buying or selling their property at the right price. The residential facilities as well as the geographic location influence the initial price, which means that if the property is near to recreational centres or public places the price gradually increases.

The objective of this paper is to develop an ensemble model with neural networks as the base predictors. Ensemble learning model is constructed by training base predictors to solve the same problem. The base predictors used are the Recurrent neural network and the Residual neural network. Neural networks are used because of their ability to handle large amount of data and to implicitly detect the complex nonlinear relationship between the dependent and independent variables. This model also increases the prediction accuracy and also produces more stable result than any individual base predictors.

2. Related Work

Muhammad Fahmi Mukhlishin et al[1] compared three different methods to find the most appropriate method for the prediction of house price i.e., Fuzzy Logic, Artificial Neural Network, K Nearest Neighbor. Variables used in these methods are NJOP of land, the locations, the age, NJOP of Building, and the valuable location of the land. The performance of each method is measured by comparing the actual output with the predicted output using MAPE formula. The experiment result shows that the fuzzy method is superior to neural network and K-nearest neighbour.

Hakan Kusan et al[2] employed a new grading system, the Fuzzy Logic System for price prediction. It uses various parameters as input such as the plan of city, nearness to public facilities, other environmental factors for the model generation. Sugeno type fuzzy inference system has been used for model construction. RMSE(Root Mean Square Error) is used to express the errors in the training and testing processes in the FL

model. It is concluded that the predicted unit price is very much close to the real house price.
Gang-Zhi Fan et al[3] uses an alternative statistical pattern recognition tool, the decision tree for examining the relationship between the housing price and housing characteristics, determinants of housing price and predicting house price. Decision tree provides an effective approach for identifying the determinants of public housing resale price.

Neelam Shinde et al[4] predicts the price of houses using various machine learning techniques such as, logistic regression, support vector regression, lasso regression and decision tree. The algorithms were compared based on parameters such as MAE, MSE, RMSE and accuracy. It is found that the decision tree with the highest accuracy of 84.64%, logistic regression with 72.81%, support vector regression with 67.81% and lasso with the least, 60.32%.

Bowen Yang et al[5] built a stacking ensemble model with ET, RF,GDBT and XGB as base predictors. During the model construction the individual predictions are given as input to the ensemble predictor which provides a final prediction result. The result shows that it can improve the prediction accuracy and also can produce more stable results.

3. Design and Implementation

Generally Artificial Neural Networks are having high variance, frustrating when developing a new prediction model. To reduce variance, train multiple models instead of single model and combine their predictions. This is known as ensemble learning, which not only reduces variance but also result in better predictions than any single model

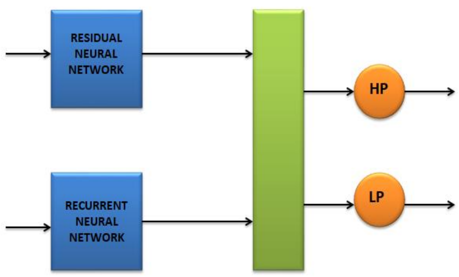

Fig.1. Proposed System

.
Artificial neural networks are nonlinear models which can learn complex nonlinear relationship in the data. The proposed system is shown in fig.1

3.1 Ensemble Learning

A new model can be trained to learn how to best combine the contributions from each sub-model. This approach is called stacked generalization, or stacking for short, and can result in better predictive performance than any single contributing model. In stacking, an algorithm takes the outputs of sub-models as input and attempts to learn how to best combine the input predictions to make a better output prediction. It may be helpful to think of the stacking procedure as having two levels: level 0 and level 1.

- Level 0: The level 0 input is the training dataset and the level 0 model learns to make predictions from this data.
- Level 1: The level 1 takes the output of the level 0 models as input and the single level 1 model, or meta-learner, learns to make predictions from this data.

When using neural networks as sub-models, it may be desirable to use a neural network as a meta-learner. Specifically, the sub-networks can be embedded in a larger multi-headed neural network that then learns how to best combine the predictions from each input sub-model. It allows the stacking ensemble to be treated as a single large model. The benefit of this approach is that the outputs of the sub-models are provided directly to the meta-learner. Further, it is also possible to update the weights of the sub-models in conjunction with the meta-learner model, if this is desirable.

The level 0 models are the Residual neural network and the recurrent neural network. The residual neural networks are used to reduce the under-fitting problem and the Recurrent Neural Networks are used to process time series data.

3.2 Residual Neural Network

In deep learning networks, a residual learning framework helps to preserve good results through a network with many layers. One problem commonly cited by professionals is that with deep networks composed of many dozens of layers, accuracy can become saturated, and some degradation can occur. Some talk about a different problem called "vanishing gradient" in which the gradient fluctuations become too small to be immediately useful.

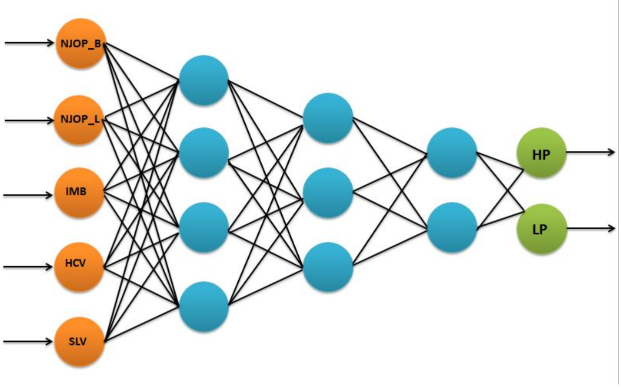

Fig.2 Residual Neural Network

The deep residual network deals with some of these problems by using residual blocks, which take advantage of residual mapping to preserve inputs. By utilizing deep residual learning frameworks, engineers can experiment with deeper networks that have specific training challenges. The residual neural network is shown in fig.2

3.3 Recurrent Neural Network

A recurrent neural network is a class of artificial neural network where the connections between nodes form a directed graph along a temporal sequence. This allows it to exhibit temporal dynamic behaviour. RNN can use their internal state(memory) to process sequence of inputs. RNN were designed to work with sequence prediction problems. The RNN utilizes a supervised machine learning technique called back propagation for training.

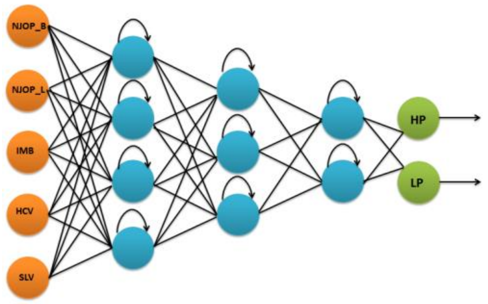

Fig.3 Recurrent Neural Network

Once the sub models/base predictors are trained, we can define our stacking model. The input layer of the sub-models can be taught of as inputs to this model. This means that 2 copies of any input should be provided to the model. The output of each model has to be merged, for which a hidden layer is defined. The hidden layer interprets this input to the meta-learner and the output layer makes its own prediction. Once the model is defined, the sub-models are not trainable, their weights will not be updated during training and only the weights of the new hidden and output layer will be updated.

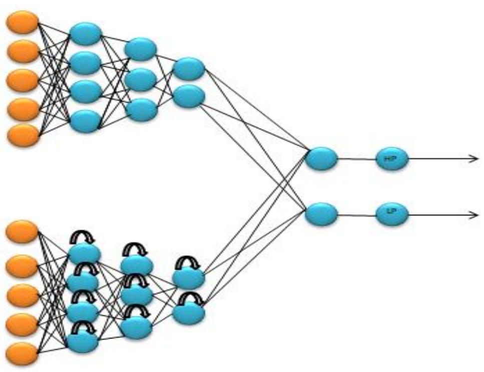

Fig.4. proposed model

4. Experimental Result and Analysis

We compare the predicted housing prices from the ensemble model to the actual ones to verify the accuracy or performance of the ensemble model. The computation results indicate that the ensemble model is able to provide the most accurate predictions.

Table 1: Prediction accuracy of different methods

Method	Accuracy
Ensemble method	93%
Decision Tree	84.64%
Linear Regression	90.1%

5. Conclusion

A base predictor would have its pros and cons, and it might not be able to work on all datasets with the universal superiority. By applying the ensemble techniques, we can strengthen the advantages of the underlying base predictors or models while suppressing the shortcomings of these base models. The ensemble model demonstrates its effectiveness in dealing with datasets with noise and over-fitting problems

References

[1] Muhammad Fahmi Mukhlishin, Ragil Saputra, Adi Wibowo, Predicting House Sale Price Using Fuzzy Logic, Artificial Neural Network and K-Nearest Neighbor.

[2] Hakan Kus_an *, Osman Aytekin, _Ilker Özdemir, The use of fuzzy logic in predicting house selling price.

[3] Gang-Zhi Fan, Seow Eng Ong and Hian Chye Koh, Determinants of House Price: A Decision Tree Approach.

[4] Neelam Shinde, Kiran Gawande, Valuation Of House Prices Using Predictive Techniques

[5] Bowen Yang, Buyang Cao, Ensemble Learning Based Housing Price Pre-diction Model.

Authors profile

V Manju received her B Tech (CSE) degree from university of Kerala in 2017. She is currently pursuing her Masters in Computer Science & Engineering from KTU. She has published a survey paper on data mining.

Stephy Philipose is working as an Assistant Professor in computer Science and Engineering Department. Her research interest focuses on Networking and Data security. She has published several papers on image processing

Dr. Radhakrishnan B is working as the head of CSE department. He has more than 14 years experience in teaching and has published papers on data mining and image processing. His research interests include image processing, data mining, image mining