Urban Residential Land Price Appraisal via Quantifying Impact Factors Based on Deep Belief Networks

Hua Ai
School of Literature
Neijiang Normal University
Neijiang, P.R. China
1259883898@qq.com

Qiang Liu
School of Resources and Environment
University of Electronic Science and Technology of China
Chengdu, P.R. China
liuqiang6904@uestc.edu.cn

Yuxin Jiang
School of Economics
Hefei University of *Technology*
Hefei, P.R. China
446990843@qq.com

Ran Yang
School of Resources and Environment
University of Electronic Science and Technology of China
Chengdu, P.R. China
liuqiang_em@sina.com

Abstract—**The relationship between urban residential land prices and the explanatory variables is highly complex. This causes that it is difficult to quantify the impact factors for appraising urban residential land prices. This paper explores the use of Deep Belief Networks for quantifying impact factors of urban residential land prices. The proposed approach applies grid cells to express the samples finely 458 features are extracted as input from collected raw data of 37 important impact factors, residential land prices are divided into 9 levels as output, and both BBRBM and GBRBM are utilized to form the network. A deep belief network model is finally employed to appraise urban residential land prices via their impact factors, and the average accuracy can achieve about 90%.**

Keywords—urban residential land price; Deep Belief Network; restricted Boltzmann machine; grid cell

I. INTRODUCTION

Land prices are the core mechanism in the land market. At the macro level, the land price mechanism is an important means by which the state controls the supply and demand in the land market to realize the effective allocation of land resources. Further, at the micro level, analyzing the factors affecting the price performance of urban land has important reference value in guiding the compilation of urban planning, real estate development and the allocation of land resources [1].

Many studies have been conducted on land price. Shan Y.H. and Nie J.C build GWR model to explore the regional differences of residential land price influenced by some factors in Wuhan city [1]. Zhu C.G et al. build the Hedonic price model which reflect the urban residential land price and the influence factors and analyze the influencing pattern of the factors based on the residential monitoring price on the secondary market of urban land [3]. Zhang L.P. et al. explore to establish a forecast GM(1,1) model of integrated land price in Xinjiang and analysis its Influential factors based on grey prediction theory and grey association analysis theory [4]. Lin F uses Markov Chain prediction model to predict the trend of Land Price [5]. Zhu C.J. et al. take the grid technology, genetic neural network and other research methods to evaluate the city residential land price while combining theory and practice [6]. However, existing research has the following deficiencies: (1) most of studies pay attention to analysis model, so there are relatively fewer studies on prediction model for land price.(2) while there are many factors affect the land price, current research only takes into consideration a small portion of these factors. (3) existing research mainly focuses on some simple linear model such as GWR model, Hedonic model and so on [2, 7]. And artificial neural network used to predict land price have only one hidden layer and the network is too shallow.

In recent years, due to its powerful feature extraction and classification deep learning algorithm has shown great potential in application areas such as pattern classification, computer vision, natural language processing and speech recognition. DBN (deep belief network) is one of the most widely used deep learning algorithms [8, 9, 10, 11, 12]. A DBN consists of several layers of controlling restricted Boltzmann machine (RBM). It then performs supervised learning using backpropagation after unsupervised learning. However, the application of DBN to analysis of land price is still rare, which motivated us to apply DBN to land price classification.

Motivated by the great potential of deep learning, this paper proposes a deep learning novel approach that appraises urban residential land prices anywhere in Shenzhen by quantifying import impact factors.

DBN is applicable for a large amount of input data for training, thus, in this paper, we put forward a DBN-based quantifying model for impact factors of urban residential land prices which consisted of hundreds of input features and hundreds of thousands of samples. We converted all vector spatial data into raster data as large input and out data for this model. Firstly, we simulate the practical urban residential land price distribution in Shenzhen by the IDW interpolation method which uses urban residential land price monitoring points in Shenzhen. Then hundreds of features are extracted for every sample and the dataset is divided into 9 classes based on the land price. Finally, the DBN-based quantification model of land price impact factors was implemented.

The rest of this paper is organized as follows: In Section II, a DBN model for urban residential land price appraisal in Shenzhen is presented. Section III concludes some conclusions.

II. Proposed Method

A. DBN-Based Model for Land Price Appraisal

In this paper, we propose a DBN-based land price appraisal (DBLPA) mode. The DBLPA model used in this work go through two training phases: (1) unsupervised pre-learning. The visible layer consists of 458 visible units. The data array at each grid cell with the combination of a total of 458 input grid maps overlapped spatially together is fed into the 458 units of the visible layer of the deep belief network. Our DBNBLP model has 5 hidden layers each of which is an RBM. And the nodes of these hidden layers are 400, 200, 100, 50 and 25 respectively. Takayoshi Y. et al. proposed that the type of RBM could be selected according to the distribution of the visible unit [10]. Benefited from this idea, we determined whether to use BBRBM or GBRBM in each hidden layer depended on the distribution of the visible unit. If the distribution of visible units is closer to the Gaussian mixture distribution than the pseudo binary distribution, GBRBM should be chosen, and vice versa. Consequently, GBRBM is chosen for the first two layers while BBRBM is selected for the last three layers. An example of the distribution of visible unit is shown in Figure 1. For each GBRBM, to make the input data uniform with no condition variations and the distribution smoother, each node of the visible units is normalized with the zero-mean and the unit variance. And for each BBRBM, to speed up convergence and avoid an error, each node of the visible units is scaled to the interval [0, 1]. The contrastive divergence (CD) approximation with one-step Gibbs sampler is used to update w, bv and bh of RBM. And we also use momentum to increase the speed of learning and weight-decay to prevent overfitting and make CD learning a better approximation to maximum likelihood [11]. The structure of our DBN model for pre-learning is illustrated in Figure 2. (2) Fine-tuning of the DBN parameters after layer-wise learning. After pre-training, the output layer is added to the DBN. This output layer has 9 nodes and is a linear layer (softmax will be achieved later), as illustrated in Figure 1. The input of the first hidden layer will be normalized with the zero-mean and the unit variance.

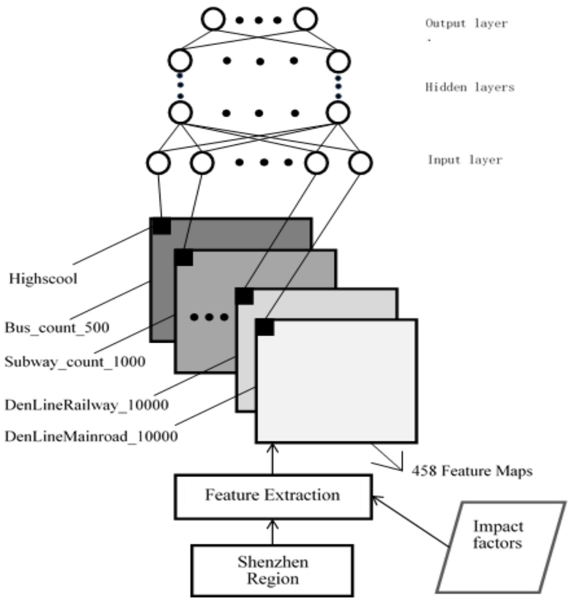


Fig. 1. An intuitive expression of features as the input to the model.

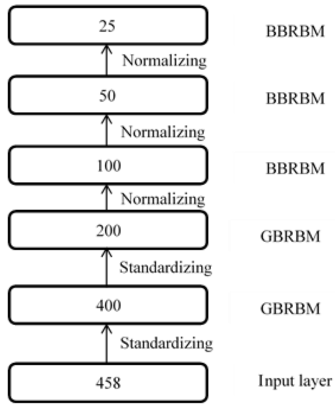


Fig. 2. Unsupervised pre-learning of the DBN.

B. Training Procedure

We conduct our experiment on the Intel Core E5-1630V3 3.70 GHz Ubuntu 16.04 machine with 16 GB of memory. Python 3.4 is used in this research. Tensorflow1.2.0 and tensorlayer1.7.2 are used to implement DBN. We divide the whole data set into two parts: training and testing sets. There are 72884 samples in the training set and 13500 samples in the testing set, respectively.

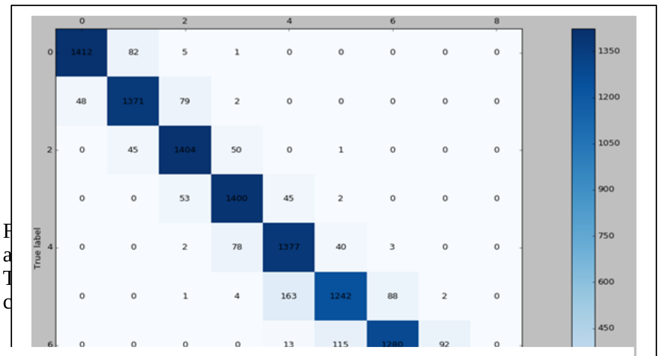
For pre-learning, as labels are unnecessary, all the samples are applied. Weights are initialized with small random values sampled from a truncated normal distribution with zero mean and standard deviation of 0.01. The hidden biases are initialized to 0 while the visible biases are initialized to 0.1. To speed up the pre-training of each RBM, we subdivided the data into mini-batches, each including 10 data vectors. The weights are updated after each mini-batch. According to Hinton, when using Gaussian visible units, the learning rate needs to be about one or two orders of magnitude smaller than when using binary visible units [11]. Therefore, the learning rate is set to 0.0001 for GBRBM while 0.001 for BBRBM. Each hidden layer is pre-trained for 100 passes through the entire training set. In addition, a coefficient of momentum times the previous update is added to each weight and bias. And weight cost times the value of the weight is subtracted to penalize large weights. The momentum is started with 0.5, after 5 epochs the momentum is increased to 0.9, and the weight cost is set to 0.0001.

C. Experimental Results

Our experimental results are satisfactory, and the accuracy and recall rate of each class in the testing set are presented in Table 1. The confusion matrix is shown in Figure 3. The loss function graph and the accuracy graph during the fine-tuning are shown in Figures 4 and 5, respectively. A comparison of the original distribution and the distribution generated by the DBN model is shown in Figure 6.

TABLE 1. THE ACCURACY AND RECALL RATE OF EACH CLASS IN THE TESTING SET.

Class	Accuracy	Recall Rate
Class 0	0.967	0.941
Class 1	0.915	0.914
Class 2	0.909	0.936
Class 3	0.912	0.933
Class 4	0.861	0.918
Class 5	0.887	0.828
Class 6	0.903	0.853
Class 7	0.861	0.949
Class 8	0.977	0.907
Mean	**0.909**	**0.909**



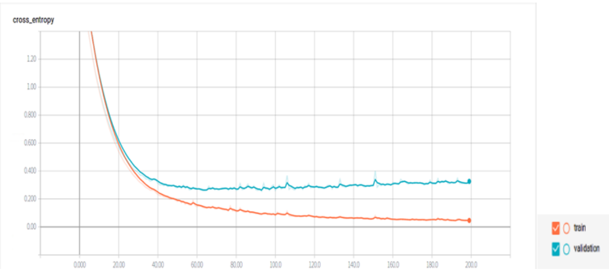


Fig. 4. Loss function during the fine-tuning.

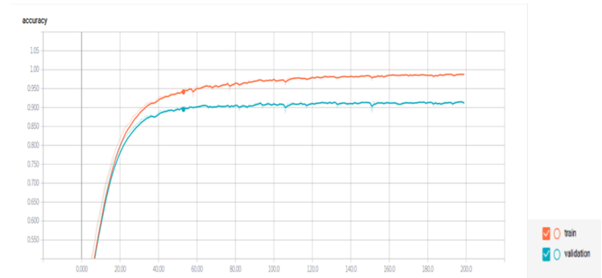
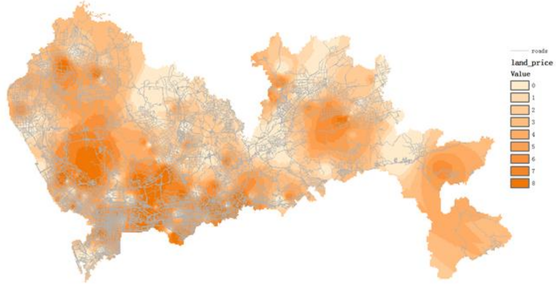
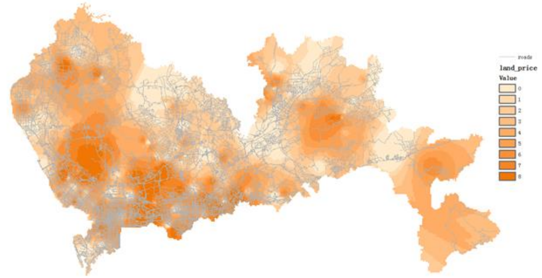


Fig. 5. Accuracy during the fine-tuning.



(a)



(b)

Fig. 6. A comparison of (a) the original distribution of urban land prices and (b) the distribution of urban land prices generated by DBN model. Different colors represent different land price levels.

III. CONCLUSION

In this paper, we propose a deep learning scheme based on DBN for urban residential land price appraisal via quantifying multiple complex impact factors. A core idea to process all data into grid for acquiring fine and abundant samples is proposed. First, urban residential land price monitoring points and 37 kinds of POI data are as important impact factors collected. Second, we apply the IDW interpolation method to convert urban residential land price monitoring points into a raster, which simulates the practical urban residential land price distribution in Shenzhen. Each grid cell in the raster corresponds to the output nodes as labels to the DBN model. Third, we produce 458 rasters from the 37 kinds of POI data using spatial analysis tools such as Point Density or Buffer. All 458 cell values at the same position form an input to the DBN model. Finally, the DBN structure is constructed for the urban residential land price appraisal experiments. Experimental results show that our DBN model has great potential to provide support and help for the future urban residential land price analysis and research.

ACKNOWLEDGMENT

The Project Supported by the Open Fund of Key Laboratory of Urban Land Resources Monitoring and Simulation, Ministry of Land and Resources (No. KF-2016-02-007).

REFERENCES

[1] Shan Y.H., Nie J.C., "Influencing Factors of Residential Land Price in Wuhan Based on GWR Model," Journal of Huazhong Agricultural University, vol. 5, pp. 111-118, 2014.

[2] Liu L, Liu H.Y, "Economic Analysis of the Relationship between Land Price and House Price," Quantitative & Technica Economics, vol. 7, pp. 27-30, 2003.

[3] Zhu C.G., Tang Y, Wu., "A Study on the Influence Factors of the Residential Land Price Based on Hedonic Price Model: Taking Nanjing

City as An Example," AREAL RESEARCH AND DEVELOPMENT, vol. 33(3), pp. 156-160, 2014.

[4] Zhang L.P., Zheng Y.L., Zheng Y.J., "The Integrated Land Price Forecast and Influential Factors Analysis in Xinjiang Based on the Gray System Theory," MATHEMATICS IN PRACTICE AND THEORY, vol. 43(2), pp. 81-89, 2013.

[5] Lin F., "The Analysis and Application of Markov Chain in Trend Prediction of Land Price," Chinese Operations Research Association Ninth Symposium Proceedings, pp. 379-384, 2008.

[6] Zhu C.J., Zhu H.Y., Lin Y., He J.M., Wu Q., Tang Y., "Study on the City Residential Land Price Evaluation Based on Genetic Neural Network: Taking Nanjing as An Example," GUANGDONG LAND SCIENCE, vol. 13(5), pp. 40-44, 2014.

[7] Shi J. G., Li J., "M. A study on interpolation techniques in the Digital Land Price Model - the case of Shanghai inner belt-way district," CHINA CIVIL ENGINEERING JOURNAL, vol. 40(1), pp. 90-94, 2007.

[8] Fischer A., Igel C., "An Introduction to Restricted Boltzmann Machines," Lecture Notes in Computer Science, vol. 7441, pp. 14-36, 2012.

[9] Takayoshi Y., Masayuki T., Eiji Y., Yuji Y., Hironobu F., "To be Bernoulli or to be Gaussian, for a Restricted Boltzmann Machine," International Conference on Pattern Recognition, pp. 1520-1525, 2014.

[10] Hinton G.E., "A Practical Guide to Training Restricted Boltzmann Machines," Lecture Notes in Computer Science, vol. 7700, pp. 599-619, 2012.

[11] Szegedy C., Vanhoucke V., Ioffe S., Shlens J., Wojna Z., "Rethinking the Inception Architecture for Computer Vision," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) , pp. 2818-2826, 2016.

[12] Yuancheng Li, Xiangqian Nie, Rong Huang, "Web spam classification method based on deep belief networks," Expert Systems With Applications, vol. 96, pp. 261–270, 2018.