Acta Technica 62 (2017), No. 7A, 277–286

Mining environment assessment based on softmax convolutional neural network of cost

Zhiwei Cao[1,*], Xibing Li[1]

Abstract. In order to improve the accuracy of evaluation system for mining radiation environment based on circular economy theory, a method for constructing evaluation system of mining radiation environment based on Softmax convolutional neural network of cost is proposed. Besides, in order to improve the classification performance of convolutional network, in this Thesis, detailed theoretical analysis is carried out for convolutional neural network model; and through a large number of contrast experiments, factors that influence the performance of convolutional network are found out. By combing theoretical analysis and contrast experiment, in this Thesis, a deep convolutional network with 8 convolution layers is designed, and the classification accuracy in CIFAR-10 dataset reached 88.1%. According to experimental results, the evaluation system for mining radiation environment based on Softmax convolutional neural network of cost can effectively improve the effect of radiation pollution prediction.

Key words. Softmax cost, Convolutional neural network, Mining environment, Radiation.

1. Introduction

Environment refers to the space surrounding crowds and the totality of various natural factors and social factors which can directly and indirectly influence human life and development in it. The totality of natural factors is called natural environment, and the totality of social factors is called social environment. Environmental impact refers to possible changes of environmental conditions or formation of new environmental conditions caused by development behaviors. Environmental impact can be harmful or favorable. Environmental impact assessment, also called prospective environment assessment, refers to assessment of impact of proposed development on environment; namely, during resource exploitation and utilization as well

[1]School of Resources and Safety Engineering, Central South University, Changsha Hunan, 410083, China

*. Corresponding author

as engineering construction design, carry out prediction and assessment of project impact on environment during construction and after putting into use; in addition, compare various alternative solutions, and put forward mitigation measures so as to reduce the adverse impact on environment to the minimum degree. The purpose is to control and reduce environmental pollution brought by new projects and extension projects on the precondition of guaranteeing constant growth of national economy. The implementation of environmental impact assessment system changed traditional economic development mode that only current and direct economic benefit is valued; and after it, economic benefit and environmental benefit are combined to realize harmonious development of economy and environment. The process of environmental impact assessment is also the process of knowing the correlative dependence and interaction relationship between ecological environment and human's economic activity.

In *Environmental Protection Law of the People's Republic of China* (provisional) published in 1979, it is specified that at the time of implementing new projects, reconstruction projects, and extension projects, an environmental impact report must be submitted. This system requires that environmental impact assessment shall be made for industrial regions, residential areas, public utilities, greenbelts, and other places according to regional environment features (namely, meteorological condition, geographical condition, mining conditions, and ecological condition) so as to provide scientific basis for overall planning, reasonable arrangement as well as prevention and treatment of pollution and other public hazards. In accordance with the above Law, national planning department, economy department, and environment treatment department define environmental impact assessment as an important constituent part of feasibility study of development and construction projects. China is a developing country which undertakes tasks of both developing economy and protecting environment at present; by considering national conditions, China takes environment protection as a basic national policy and the realization of sustainable development as a great strategy during the process of comprehensively promoting modernization construction; during nearly 20 years of carrying out large-scale pollution prevention and ecological environmental protection nationwide, environmental impact assessment system played a great role in coordinating economic development and environmental protection in China. With the deepening of environmental impact assessment research and practice, the concept of environmental impact assessment has been extended, and the theory system, method system, and management mechanism of environmental impact assessment have been gradually improved and developed; in addition, assessment objects have been expanded from industrial pollution items to resource development and construction projects as well as regional development projects which have adverse effects on ecological environment, and some new environmental impact assessment modes also appeared.

In this Thesis, in order to improve the accuracy of evaluation system for mining radiation environment based on circular economy theory, a method for constructing evaluation system of mining radiation environment based on Softmax convolutional neural network of cost is proposed. Besides, in order to improve the classification performance of convolutional network, a deep convolutional network with 8 convolution layers is designed. According to experimental results, the evaluation system for mining radiation environment based on Softmax convolutional neural network of cost can effectively improve radiation pollution prediction.

2. Model analysis of convolutional neural network

2.1. Basic topological structure of network

The biggest difference between convolutional neural network and other neural network models is that convolutional neural network is connected with convolutional layer in front of input layer of neural network, and in such way, the convolutional layer is turned into data input of convolutional neural network. LeNet-5 is a classical convolutional neural network model developed by Yan Lecun for identifying handwritten characters; refer to Fig.1 for its structure chart.

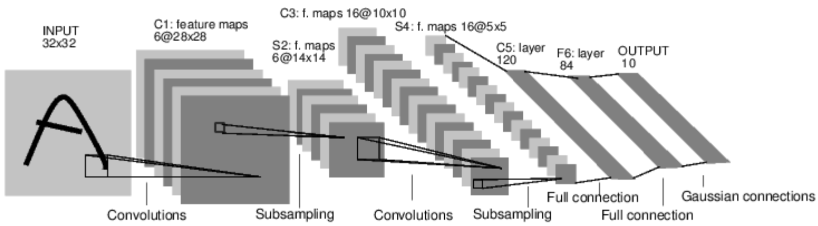

Fig. 1. LeNet-5 structure chart

There are 7 layers in LeNet-5 architecture, including 3 convolutional layers. The first convolutional layer is composed of 6 Feature Maps (FM), and therefore C1 includes 156 trainable parameters (6 5X5 kernels plus 6 deviation values) which are used for create 122304 (156* (28*28) -122, 304) connections. In C1 layer, the size of FM is 28 x 28; due to boundary conditions, in the second convolutional layer, C3 includes 1500 weight and 16 offset; in C3 layer, there are 1516 trainable parameters and 151600 connections in total. In the structure of LeNet-5, there also are two subsampling layers which are S2 and S4; S2 includes 6 feature maps, and S4 includes 16 feature maps. In S2 layer, there are 12 trainable parameters and 5880 connections; and in S4 layer, there are 32 trainable parameters and 156000 connections.

To summarize the network structure of LeNet-5, we can get that the basic structure of convolutional neural network can be divided into four parts which are input layer, convolutional layer, full connection layer, and output layer.

Input layer: convolutional input layer can directly act upon original input data; for image input, the input data is pixel value of such image.

Convolutional layer: the convolutional layer of convolutional neural network is also called feature extraction layer, and it includes two parts. The first part is the true convolutional layer of which the main function is to extract input data features. The features of input data extracted by each different convolution kernel are different; the convolutional layer having the largest amount of convolution kernels can extract the largest number of input data features. The second part is the pooling layer which is also called subsamping layer; the main function of this layer is to reduce data processing quantity on the basis of maintaining useful information so as to increase the speed of training network. Under normal conditions, convolutional neural network includes at least two convolutional layers (the true convolutional layer and down-sampling layer are collectively called convolutional layers hereby) which are convolutional layer-pooling layer-convolutional layer-pooling layer. With the increase of convolutional layer number, more abstract features can be extracted based on former convolutional layer.

Full connection layer: there can be multiple full connection layers; in fact, it refers to the hidden layer of multilayer perception. Usually, the neural nodes of latter layers are connected with all neural nodes of the former layer, but there is no connection between neuron nodes in the same layer. The neuron nodes on each layer carry out forward propagation by connecting online weight respectively, and the input of neuron nodes on next layer can be gotten through weighted combination.

Output layer: the number of neuron nodes on output layer is set according to specific application missions. In case of classification mission, the output layer of convolutional neural network is usually a classifier which is Softmax classifier in most cases.

2.2. Convolution and pooling

At the time of carrying out convolution, convolution kernel is often used to extracted features of images, in which convolution kernel is the most important. The design of convolution kernel usually involves the size, number, and stride of convolution kernels.

Theoretically, convolution kernel number refers to the number of feature maps obtained through convolutional filtering in upper layer; as more feature maps are extracted, the feature expression space in the network will be larger; besides, learning ability will be stronger, and result identification at last will be harder. But in case the number of convolution kernels is too much, network complexity will be increased, and parameter number will be raised, which will increase computation complexity and the possibility of over-fitting; therefore, that the more convolution kernels are the better result will be is not always true; on the contrary, the number of convolution kernels shall be determined according to the size of specific dataset image.

Image convolution feature extraction: carry out convolution processing for a new image ($n_h \times n_w$) by setting a filter with convolution kernel size of $w \times w$ and stride of k pixel to get a feature map with size of $\frac{n_h - w + k}{k} \times \frac{n_w - w + k}{k}$, as shown in Fig.2. Generally speaking, the smaller the convolution kernel size is the higher the feature extraction quality will be; but it is still required to decide the specific size according to the size of input image.

Carry out convolution processing of input image neighborhood to get neighborhood feature map of the image; then use pooling technology through sub-sampling layer to carry out sub-sampling in small neighborhood so as to get new features. Through pooling of upper layer, in feature results, parameters can be reduced (feature dimension is reduced) and features can be reinforced, which can maintain certain

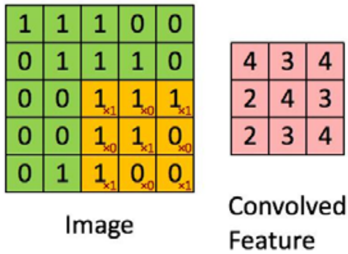

Fig. 2. Schematic diagram of image convolution

invariance (rotation, translation, expansion, and so on) of the last feature expression; therefore, pooling is in fact a process of dimensionality reduction. Those commonly used methods include mean-pooling and max-pooling.

According to correlation theory, errors in feature extraction mainly come from two aspects: (1) the estimated value variance increases due to limited neighborhood size; (2) the estimated mean value deviates due to parameter error in convolutional layer. Generally speaking, mean-pooling can decrease the first error and reserve more background information of images, while max-pooling can decrease the second error and reserve more texture information.

2.3. Activation function

The commonly used activation functions in neural network include Sigmoid function, Tanh function, ReLu function and others; the former two activation functions are most commonly used in traditional algorithms mentioned in this Thesis, while ReLu function is largely used in deep learning.

ReLu (rectified linear unit) function is a rectified linear unit proposed by Hinton; after being trained for several times by CNNs, it can be obviously seen that ReLu function is much quicker than traditional sigmoid function and tanh function.

Suppose the activation function of one neural unit is $h^{(i)}$, in which i refers to the number of hidden layer units, and $w^{(i)}$ refers to the weight of hidden units; then the expression of ReLu function is as follows:

$$h^{(i)} = \max((w^{(i)})^T x, 0) = \begin{cases} (w^{(i)})^T x & (w^{(i)})^T x > 0 \\ 0 & else \end{cases}$$

Its functional image is shown in Fig.3:

Since ReLu function has linear form and unsaturated form, and in consideration of unilateral suppression, relatively wide excitement boundary, and sparse activation, the using effect of ReLu function in convolutional neural network in better than that of sigmoid function and tanh function.

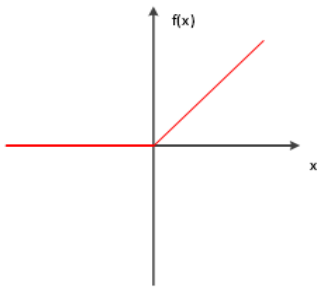

Fig. 3. Functional image of relu

2.4. Softmax classifier and cost function

When convolutional neural network is used for image classification, we connect a Softmax classifier to the back of the last full connection layer in neural network for prediction of image tags.

In softmax regression, the problem we solved is the multi-class problem (compared with binary classification problem solved in logistic regression); k different values (but not 2) can be taken for class tag y. Therefore, for training set $\{(x^{(1)}, y^{(1)}), ..., (x^{(m)}, y^{(m)})\}$, we have $y^{(i)} \in \{1, 2, ..., k\}$. (Attention: class subscript herein is started from 1 but not 0.)

As for given test input x, we want to use hypothesis function to estimate probability value $p(y = j|x)$ of each class j. That is to say, we want to estimate the probability of each kind of classification result. Therefore, in our hypothesis function, a k-dimension vector (the sum of vector elements is 1) will be outputted to express the probability value of these k estimations. Specifically, the form of our hypothesis function $h_\theta(x)$ is shown in the following:

$$h_\theta(x^{(i)}) = \begin{bmatrix} p(y^{(i)} = 1|x^{(i)};\theta) \\ p(y^{(i)} = 2|x^{(i)};\theta) \\ ... \\ p(y^{(i)} = k|x^{(i)};\theta) \end{bmatrix} = \frac{1}{\sum_{j=1}^{k} e^{\theta_j^T x^{(i)}}} \begin{bmatrix} e^{\theta_1^T x^{(i)}} \\ e^{\theta_2^T x^{(i)}} \\ ... \\ e^{\theta_k^T x^{(i)}} \end{bmatrix}.$$

For convenience, we also use the symbol θ to represent all model parameters. In realizing Softmax regression, it will be very convenient to represent θ by $k \times (n+1)$ matrix which is obtained by listing $\theta_1, \theta_2, ..., \theta_k$ in lines, as shown in the following:

$$\theta = \begin{bmatrix} \theta_1^T \\ \theta_2^T \\ ... \\ \theta_k^T \end{bmatrix}.$$

According to the above equation, it can be obtained the probability that $x^{(i)}$

belongs to j is:

$$p(y = y^{(i)}|x^{(i)};\theta) = \frac{e^{\theta_j^T x^{(i)}}}{\sum\limits_{l=1}^{k} e^{\theta_l^T x^{(i)}}}\,.$$

When the conditional probability $p(y = y^{(i)}|x^{(i)};\theta)$ of class to which each sample belongs reaches the maximum, the identification rate of classifier is the highest, and it is equal to the following maximum likelihood function at this time:

$$L(\theta|x) = \prod_{i=1}^{m} p(y = y^{(i)}|x^{(i)};\theta)\,.$$

In order to reduce calculation quantity and to prevent overflow, take the logarithm of likelihood function and make proper transformation to get:

$$J(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{j=1}^{k} 1\{y^{(i)} = j\}\log\frac{e^{\theta_j^T x^{(i)}}}{\sum\limits_{l=1}^{k} e^{\theta_l^T x^{(i)}}}\right].$$

Where, 1{.} is called indicative function; its expressed functions: 1{ *true* } =1, and 1{false}= 0. At this time, the maximum likelihood function $L(\theta|x)$ is equal to minimum cost function $J(\theta)$, and therefore gradient descent algorithm is used to seek for the minimum value of $J(\theta)$ so as to determine parameter θ. The gradient of cost function $J(\theta)$ is as follows:

$$\nabla_{\theta_j} J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left[x^{(i)}(1\{y^{(i)} = j\} - p(y^{(i)} = j|x^{(i)};\theta))\right].$$

In practical application, in order to avoid over-fitting, we often add a regularization item $\frac{\lambda}{2}\sum\limits_{i=1}^{m}\sum\limits_{j=0}^{n}\theta_{ij}^2$ (**L2 regularization**) to the cost function, and then the cost function is turned into:

$$J(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{j=1}^{k} 1\{y^{(i)} = j\}\log\frac{e^{\theta_j^T x^{(i)}}}{\sum\limits_{l=1}^{k} e^{\theta_l^T x^{(i)}}}\right] + \frac{\lambda}{2}\sum_{i=1}^{m}\sum_{j=0}^{n}\theta_{ij}^2\,.$$

The second item in the above equation will punish excessively large parameter values, and it is also called **weight attenuation item**. Proper λ can reduce the magnitude order of weight so as to control the value of network parameters, which can prevent over-fitting to a certain degree.

3. Mining environment assessment model based on SOFTMAX convolutional neural network of cost

Evaluation model for current mining radiation situation has 3 structure layers: input layer, hidden layer, and output layer. Evaluation items selected for this evaluation include 238U, 232Th, 226Ra, and others. The neuron number of input layer in the neural network is decided by evaluation factors, and therefore the neuron number of input layer is determined as 3, respectively corresponding 238U, 232Th, and 226Ra. The network output shall be classification result; in this research, the expected output results of Class I, Class II, and Class III mining area environments are respectively (1, 0, 0), (0, 1, 0), and (0, 0, 1); therefore, the neuron number of output layer is 3.

The function of hidden nodes is to extract and store inherent laws of samples; each hidden node has several weight values, and each weight is a parameter of enhanced network mapping capacity. According to calculation formula of hidden nodes, the number of hidden layer nodes is 2; take 3-3-3 as network structure to carry out cut-and-trial; and the final network structure is determined as 3-4-3.

In this research, self-compiled plug-in platform of neural network is adopted to carry out training; environmental standards for mining areas are taken as training sample input; expected error is 0.001; additional momentum algorithm for self-adaptive learning rate is adopted. Refer to Table 1 for network training sample. The relationship between network error and iteration time is shown in Table 2 and Fig.4.

Table 1. Network training sample

238U	232Th	226Ra	Sample output
1.5×104	59.0	0.02	(1,0,0)
1.5×105	5.9×102	0.20	(0,1,0)
4.4×106	1.8×104	5.90	(0,0,1)

Table 2. Table for network error during learning process

Learning times	0	10 000	20 000	30 000	40 000	48 766
Network error	1.030	0.634	0.245	0.012	0.003	0.002

It can be seen from Fig.1 that when the iteration time is less than 25 000, the network errors of test set will sharply decline, but with the implementation of training, iteration time will increase and the decline trend of network errors will slow down. And when iteration time is 48 766, the network error is 0.0019 (which meets requirements for network errors), and network training is stopped; if training is continued, over-fitting may occur; the appearance of over-fitting is because of network model deliberately conforming to individual sample; at this time, a relatively better fitting of samples joining in learning can be obtained through the network; but for samples not involved in the training, fitting deviation will increase; therefore,

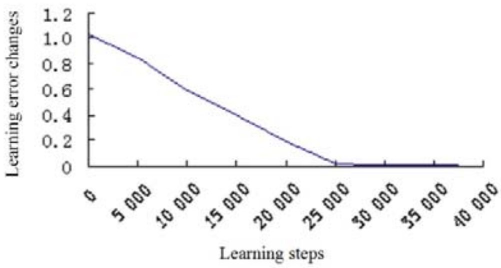

Fig. 4. Relationship between network error and iteration time

whether the number of errors in training set continues declining or not, iteration shall be stopped anyway. By considering network errors of both training set and test set, it is appropriate to set iteration time as 48 766 times.

For test set, the comparison between network output and ideal output (actual value) is shown in Table 3. And according to Table 5, the network output is very close to the actual value, which reflects that the network has relatively good simulation effect.

Table 3. Comparison between network output and ideal output

Sample output 1	Sample output 2	Sample output 3	Predicted output 1	Predicted output 2	Predicted output 3
1	0	0	0.976	0.035	0.002
0	1	0	0.023	0.962	0.013
0	0	1	0	0.017	0.989
1	0	0	0.976	0.035	0.002

Original data and information preparation related to radiation dose in mining areas is substituted into learned and trained algorithm in this Thesis to carry out simulating calculation; the related results are shown in Table 6. It can be learned from Table 6 that the radiation environment quality of all Ji'an sections in mining area is Class I; namely, it is lower than the radionuclide concentration in inhabitants' environment; and therefore the mining area is normal, which is in agreement with evaluation result shown in the test report for radionuclide content in mining areas of Ji'an City given by supervision and administration station for radiation environment in Jilin province.

4. Conclusion

In order to improve the accuracy of evaluation system for mining radiation environment based on circular economy theory, a method for constructing evaluation system of mining radiation environment based on Softmax convolutional neural network of cost is proposed. Through a large number of contrast experiments, factors that influence the performance of convolutional network are found out. In combination with theoretical analysis and contrast experiment, according to experimental results, it is indicated that the evaluation system for mining radiation environment based on Softmax convolutional neural network of cost can effectively improve the effect of radiation pollution prediction.

References

[1] M. Oquab, L. Bottou, I. Laptev, et al.: *Is object localization for free?-weakly-supervised learning with convolutional neural networks*[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2015), 685–694.

[2] G. E. Hinton, S. Osindero, Y. W. Teh: *A fast learning algorithm for deep belief nets*[J]. Neural computation, *18* (2006), No. 7, 1527–1554.

[3] Y. Chen, H. Jiang, C. Li, et al.: *Deep feature extraction and classification of hyperspectral images based on convolutional neural networks*[J]. IEEE Transactions on Geoscience and Remote Sensing, *54* (2016), No. 10, 6232–6251.

[4] C. A. Ronao, S. B. Cho: *Human activity recognition with smartphone sensors using deep learning neural networks*[J]. Expert Systems with Applications, *59* (2016), 235–244.

[5] Y. Tan, P. Tang, Y. Zhou, et al.: *Photograph aesthetical evaluation and classification with deep convolutional neural networks*[J]. Neurocomputing, *228* (2017), 165–175.

[6] F. Zhang, B. Du, L. Zhang: *A multi-task convolutional neural network for mega-city analysis using very high resolution satellite imagery and geospatial data*[J]. (2017), arXiv preprint arXiv:1702.07985.

[7] M. Jaderberg, K. Simonyan, A. Vedaldi, et al.: *Synthetic data and artificial neural networks for natural scene text recognition*[J]. arXiv preprint arXiv:1406.2227. (2014).

[8] J. Kleesiek, G. Urban, A. Hubert, et al.: *Deep MRI brain extraction: a 3D convolutional neural network for skull stripping*[J]. NeuroImage, *129* (2016), 460–469.

[9] X. Shi, Y. Hu, Y. Zhang, et al.: *Multiple Disease Risk Assessment With Uniform Model Based on Medical Clinical Notes*[J]. IEEE Access, *4* (2016), 7074–7083.

[10] M. Veta, P. J. Van Diest, S. M. Willems, et al.: *Assessment of algorithms for mitosis detection in breast cancer histopathology images*[J]. Medical image analysis, *20* (2015), No. 1, 237–248.

[11] C. Lu, Z. Wang, B. Zhou: *Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification*[J]. Advanced Engineering Informatics, *32* (2017), 139–151.

[12] Y. Yao, H. Liang, X. Li, et al.: *Sensing Urban Land-Use Patterns By Integrating Google Tensorflow And Scene-Classification Models*[J]. (2017), arXiv preprint arXiv:1708.01580.

[13] E. W. Chan, Y. Y. Zhang, and K. E. Uhrich: *Amphiphilic Macromolecule Self-Assembled Monolayers Suppress Smooth Muscle Cell Proliferation*, Bioconjugate Chemistry, *26* (2015), No. 7, 1359–1369.

[14] Y. Y. Zhang, E. Mintzer, and K. E. Uhrich: *Synthesis and Characterization of PEGylated Bolaamphiphiles with Enhanced Retention in Liposomes*, Journal of Colloid and Interface Science, *482* (2016), 19–26.

[15] Y. Y. Zhang, A. Algburi, N. Wang, V. Kholodovych, D. O. Oh, M. Chikindas, and K. E. Uhrich: *Self-assembled Cationic Amphiphiles as Antimicrobial Peptides Mimics: Role of Hydrophobicity, Linkage Type, and Assembly State*, Nanomedicine: Nanotechnology, Biology and Medicine, *13* (2017), No. 2, 343–352.

Received May 7, 2017