

Method for Meteorological Early Warning of Precipitation-Induced Landslides Based on Deep Neural Network

Lu Huang¹ · Lu-yang Xiang²

Published online: 3 January 2018

© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract The meteorological early warning model of precipitation-induced landslides is a significant yet challenging task, due to the complexity and uncertainty of various influence factors. Generally, the existing machine learning methods have the drawbacks of poor learning ability and weak capability of feature extraction. Inspired by deep learning technology, we propose a deep belief network (DBN) approach with Softmax classifier and Dropout mechanism for meteorological early warning of precipitation-induced landslides to overcome these problems. With the powerful nonlinear mapping ability of DBN when training a large number of sample data, we use the greedy unsupervised learning algorithm of DBN to extract the intrinsic characteristics of landslide factors. Then, to further improve prediction accuracy of landslides, the Softmax classifier is added to the top layer of DBN neural network. Moreover, the Dropout mechanism is introduced in the training process to reduce the prediction error caused by the over-fitting phenomena. Taking Wenchuan earthquake affected area for example, after analysis of the factors influencing landslide disasters, the meteorological early warning model of landslides based on Dropout DBN-Softmax is established. Compared with the existing BP neural network algorithm and BP algorithm based on Particle Swarm Optimizer (PSO-BP) algorithm, the experimental results show that the new approach proposed has the advantages of higher accuracy and better technological performances than the former algorithms.

Keywords Deep neural network · Landslide · Meteorological early warning · Softmax · Wenchuan

✉ Lu Huang
huanglu213@163.com

¹ Hubei Surveying Mapping Engineering Institute, Hubei Bureau of Surveying, Mapping and Geoinformation, Wuhan 430074, Hubei, China

² Engineering Department, The 709th Research Institute of China Shipbuilding Industry Corporation, Wuhan 430074, Hubei, China

1 Introduction

The occurrence and development of landslide disasters is an extremely complex system with nonlinear dynamic process [1]. Since the variety, complexity and uncertainty of the geological conditions and inducing factors, the prediction of landslides is still a hot topic in the world. Rainfall is one of the most important influence factors to regional landslide. Therefore, it is very necessary to get down to the research on the method for meteorological early warning and forecast of landslides by taking the rainfall as the triggering factor. It is the major scientific and technological problem of landslide hazard predicting and preventing.

The former research concentrates much more on the probability and mathematical statistics or shallow neural network. Such as approach of antecedent precipitation for rainfall-triggering landslide forecast [2,3], Bayes statistical inference model [4], Logistic regression model [5], probability quantization model of hazard factors [6], early warning analysis model of normalized equation [7], forecasting index method of disaster [8], rainfall level index method [9], meteorological and geological environment coupling model [10], BP model [11] and so on. Since these models established in the case of limited training samples and computational units, the utilization of data sample is not high enough. Besides, there exists great subjectivity in the analysis process, which lead the result that the complex nonlinear relationships between landslide and its influencing factors cannot be fully exploit. Therefore, these models still have many disadvantages and not ideal for prediction and poor generalization ability.

Deep learning is a new research direction in the field of artificial intelligence, which has now been applied in many fields, including big data mining [12,13], digital image processing [14,15], speech recognition [16], human activity recognition [17,18], character recognition [19], etc. Deep belief network (DBN) is one of the typical deep learning methods articulated by Hinton. It consists of Restricted Boltzmann Machine (RBM) stacked in series, and it has the ability to extract features from a large number of samples, which can greatly improve the classification accuracy and prediction precision. Currently, DBN has been successfully used in various kinds of prediction and classification problems. But so far, it has not been applied in the landslide forecast area yet.

In this paper, we extend a method of the meteorological early warning of precipitation-induced landslides based on Dropout DBN-Softmax. Firstly, there is a brief introduction toward the natural condition and geological environment condition of the study area in Wenchuan earthquake affected area. And the data preparation is described, including the samples of 5000 history landslide disaster data from 2009 to 2013. Selection and classification of the influence factors, the data quantization and preprocessing are also described. And then the adopted methodology is particularly introduced. A method of landslide hazard prediction based on DBN is proposed. Softmax classifier is introduced to further improve prediction accuracy of landslide based on DBN. The Dropout mechanism is also applied in the training process to reduce the prediction error caused by over-fitting. The meteorological early warning model based on Dropout DBN-Softmax is established. Furthermore, to verify the performance of the method presented, a series simulation experiments are presented in this paper. The experimental results show that, compared with the traditional BP algorithm and PSO-BP algorithm, the method presented has the advantages of higher accuracy and better technological performances.

The rest of the paper is organized in the following manner. Section 2 introduces the study area, and articulates the experimental data preparation process. Section 3 illustrates the adopted methodology in details. Section 4 presents a series of experiments to demonstrate the



Fig. 1 Administrative zoning map of the study area

effectiveness and performance of the proposed method. Finally, the summary and conclusions are given in the last section.

2 Study Area and Data Preparation

2.1 Study Area

The study area is the hard-hit areas in 5 · 12 Wenchuan earthquake in 2008. It is mainly in the northwest of Sichuan Province, in the border area of Gansu and Shaanxi Province. It lies between $29^{\circ}13'$ and $34^{\circ}10'N$ latitude and $101^{\circ}58'$ and $106^{\circ}55'E$ longitude, covering a total area of about $14.66 \times 10^4 \text{ km}^2$. It has a tropical monsoon climate with great annual rainfall and uneven rainfall distribution. The Wenchuan earthquake affected area locates at the alp-gorge zone, which has varied types of landform and big topographic relief. The variation range of elevation is about 297–6676 m, and its average slope is 23.52° . Within the geographical area of Longmen Mountain fault zone, there has the complexity of structure and the assemblage of rock and stratum, which leads to the high susceptibility zone of geological disasters such as landslides [20]. From the administrative district divisions, the study area belongs to Sichuan, Gansu and Shanxi province, including 62 counties (cities or districts) as shown in Fig. 1.

2.2 Selection and Classification of Influence Factors

Landslide is the result of both disaster pregnant environment factors and disaster inducing factors. Accordingly, we selected the influence factors of typical characteristics of landslides

Table 1 Selection and classification of influence factors

Influence factors	Divisions
Geomorphological type	Plain, hill, highland region, river valley basin, lake, mountain, low mountain
Rock type and soil mass	Igneous, sedimentary, metamorphic, and others
Elevation (m)	< 500, [500, 1000), [1000, 1500), [1500, 2000), [2000, 2500), [2500, 3000), > 3000
Slope (°)	[0, 5), [5, 10), [10, 20), [20, 30), [30, 40), [40, 50), [50, 60)
Seismic intensity	I, II, III, IV, V, VI, VII, VIII, IX, X, XI
Hydrogeological type	Pore phreatic water in loose rock mass, carbonate karst fissure of water, intrusion, rock crevice water of magmatic rock, fault pore-fracture water, karst fracture water, fracture water, and others
Average annual rainfall after earthquake (mm)	[0, 50), [50, 60), [60, 70), [70, 80), [80, 90), [90, 100), [100, 1100), [1100, 1200), [1200, 1300), [1300, 1400), [1400, 1500), [1500, 1600), [1600, 1700), > 1700
Daily rainfall (mm)	Actual rainfall value
Cumulative precipitation in previous 7 days (mm)	Actual rainfall value

from the both aspects. With the development characteristics of landslides in the study area, considering the availability and practicality of data samples, there were 9 typical factors chosen as influence factors of landslides, including geomorphological type, rock type and soil mass, elevation, slope, seismic intensity, hydrogeological type, average annual rainfall after earthquake, daily rainfall and cumulative precipitation in previous 7 days (Table 1). It was laid the foundation to further study of meteorological early warning model of landslides in Wenchuan earthquake affected area.

2.3 Data Selection

In this paper, the sample data of landslide disasters was obtained from geological environment monitoring institution of China. The rainfall data was from China Meteorological Administration. We used 5000 history landslide disaster data from 2009 to 2013 in Wenchuan earthquake affected area as our experimental samples. All samples were divided into two parts, 80% of them were used as the training samples, the remaining 20% were as the testing samples; that is, 4000 items were selected as the training samples and other 1000 items were as the testing samples. The model of meteorological early warning of landslides based on DBN can be created with the training samples. And then the performance of the model can be tested and evaluated by using the testing samples.

2.4 Data Quantization and Preprocessing

For the disaster pregnant environment factors, *CF* values of each factors can be calculated using certainty factor (*CF*) method [21]. Firstly, each layer of disaster pregnant environment factors was divided into different subsets according to certain classification rules. Then with the spatial analysis technology of geographic information system (GIS), each data layer of disaster pregnant environment factor was overlaid with landslide data layers. Moreover, the *CF* value of each subset was obtained respectively by use of certainty factor method.

The CF expression can be represented as follows:

$$CF = \begin{cases} \frac{PP_a - PP_s}{PP_a(1 - PP_s)}, & PP_a \geq PP_s \\ \frac{PP_a - PP_s}{PP_s(1 - PP_a)}, & PP_a \leq PP_s \end{cases} \quad (1)$$

where CF is the certainty coefficient, its value range is $[-1, 1]$; PP_a is the conditional probability of landslide disaster events in set a ; PP_s is the prior probability of landslide disaster events in the whole study area.

Using the CF probability model, we can get the deterministic coefficient of the classification of each disaster pregnant environmental factor. The results are shown in Table 2.

For the disaster inducing factors, the rainfalls during main flood season (from May till September each year) were chosen as the meteorological factors. For daily rainfall and cumulative precipitation in previous 7 days, their actual rainfall values were as the quantitative values.

Based on the index system of influencing factors in the study area, the CF values of the disaster pregnant environment factors were calculated, and the normalized process to the factors can be made by using the expression as follows. The sample data can be mapped to $[0, 1]$.

$$y = \frac{x - MinValue}{MaxValue - MinValue} \quad (2)$$

where x is the value before conversion and y is the value after conversion; $MaxValue$ and $MinValue$ represent the maximum and minimum value of the samples, respectively.

The disaster pregnant environment factors and disaster inducing factors were taken as input for training and simulation of DBN network model.

At present, there has not yet been a unified prediction evaluation criteria of landslide hazard in academic circles. Generally speaking, landslide activity intensity mainly includes the frequency, scale and movement speed of the landslides. In this paper, according to the disaster degrees and scales of landslide in the data samples, the level of landslides was divided into 4 types in Wenchuan earthquake affected area: minor landslide, medium-sized landslide, large-scale landslide and huge landslide (Table 3).

3 Adopted Methodology

3.1 Deep Belief Network (DBN)

Deep Belief Network was a probabilistic generative model proposed by Professor Hinton in 2006. It is a deep network structure stacked by a series of restricted Boltzmann machines (RBM) [22]. In this paper, we used the greedy unsupervised learning algorithm of DBN for data pre-training to obtain a better distributed representation of the input data. BP algorithm was used to adjust the parameters of the whole network to get the optimal result. The algorithm presented in the paper effectively overcomes the shortcoming of unsuitable for multi-layer network in the traditional neural network training method. The model structural diagram of DBN can be constructed as follows.

As shown in Fig. 2, the training process of DBN can be divided into two processes: pre-training and fine-tuning.

Table 2 CF values of disaster pregnant environment factors

Index	Disaster pregnant environment factor	Subset	CF values
x_1	Geomorphological type	Plain	0.145
		Hill	0.0450
		Highland region	-0.907
		River valley basin	-0.542
		Lake	0.807
		Mountain	-0.595
		Medium mountain	0.277
		Low mountain	0.430
x_2	Rock type and soil mass	Igneous	-0.563
		Sedimentary	0.344
		Metamorphic	-0.544
		Others	-0.227
x_3	Elevation	Interval value	
		< 500	0.260
		[500, 1000)	0.503
		[1000, 1500)	0.493
		[1500, 2000)	0.124
		[2000, 2500)	-0.363
		[2500, 3000)	-0.528
		> 3000	-0.940
x_4	Slope	Interval value	
		[0, 5)	0.029
		[5, 10)	0.104
		[10, 20)	0.228
		[20, 30)	0.270
		[30, 40)	0.023
		[40, 50)	-0.295
		[50, 60)	-0.498
x_5	Seismic intensity	VI	-0.674
		VII	0.197
		VIII	0.025
		IX	0.541
		X	0.707
		XI	0.604
x_6	Hydrogeological type	Pore phreatic water in loose rock mass	0.065
		Carbonate karst fissure of water	-0.168
		Intrusion	-0.705
		Rock crevice water of magmatic rock	0.198
		Karst fracture water	-0.026

Table 2 continued

Index	Disaster pregnant environment factor	Subset	CF values
x ₇	Average annual rainfall after earthquake	Interval value	
		[0, 500)	- 0.456
		[500, 600)	0.084
		[600, 700)	- 0.130
		[700, 800)	- 0.383
		[800, 900)	- 0.514
		[900, 1000)	0.134
		[1000, 1100)	0.564
		[1100, 1200)	0.467
		[1200, 1300)	0.067
		[1300, 1400)	0.342
		[1400, 1500)	- 0.716
		[1500, 1600)	- 0.636
		[1600, 1700)	- 0.406
> 1700	- 0.758		

Table 3 Coding table of meteorological warning level for landslides

Scale of landslide	Classification criteria	Coding
Minor landslide	< 0.1 million cubic metres	0 0 0 1
Medium-sized landslide	0.1 ~ 1 million cubic metres	0 0 1 0
Large-scale landslide	1 ~ 10 million cubic metres	0 1 0 0
Huge landslide	> 10 million cubic metres	1 0 0 0

3.1.1 Pre-training

Through the greedy unsupervised learning for RBM from low to high layer by layer, the weights and bias values of the network can be obtained by pre-training. The feature vectors are mapped into different feature spaces to retain characteristic information by greatest possibility and thus formed more conceptual features. A RBM unit consists of two layers, the visual layer (*V*) and the hidden layer (*H*).

As a system, RBM is a model based on energy. Between the input layer vector (*v*) and the output vector of the hidden layer (*h*), it can be described by the joint configuration energy function as follows:

$$\begin{aligned}
 E(v, h|\theta) &= -v^T W h - a^T v - b^T h \\
 &= - \sum_{i=1}^n \sum_{j=1}^m w_{ij} v_i h_j - \sum_{i=1}^n a_i v_i - \sum_{j=1}^m b_j h_j
 \end{aligned}
 \tag{3}$$

where *v_i* and *h_j* are for the state of the visual layer (*V*) and hidden layer (*H*) of each node; $\theta = \{w_{ij}, a_i, b_j\}$ represents the parameters to be optimized in which the three parameters determine the performance of RBM network during the whole RBM network training stage;

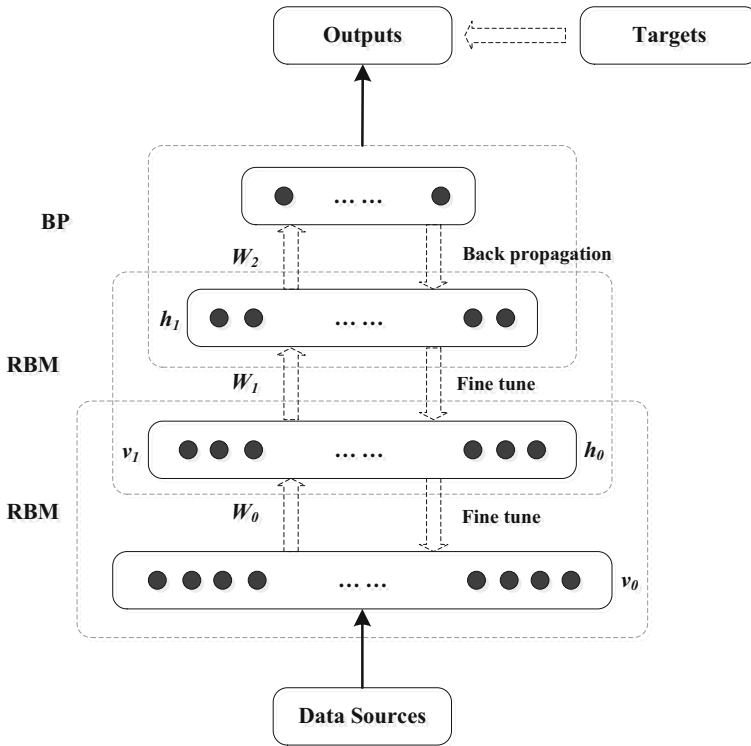


Fig. 2 The model structural diagram of DBN

w_{ij} is the connection weights between the visual layer node i and the hidden layer node j ; a_i and b_j are the bias of visual layer nodes and hidden layer nodes, respectively.

According to the energy function, the joint probability distribution of (v, h) can be obtained when the parameters θ of is given.

$$P(v, h|\theta) = \frac{1}{Z(\theta)} \exp(-E(v, h|\theta)) \tag{4}$$

where $Z(\theta)$ is the normalization factor, it is also called partition function.

$$Z(\theta) = \sum_{v,h} \exp(-E(v, h|\theta)) \tag{5}$$

To obtain the marginal distributions of $P(v, h)$ to h , we can get the probability that RBM model assigned to visual node.

$$P(v|\theta) = \frac{1}{Z(\theta)} \sum_h \exp(-E(v, h|\theta)) \tag{6}$$

The purpose of using RBM is to obtain the parameter of θ and to fit training data to the best extent possible. In order to get the parameter of θ , we use the minimum logarithmic likelihood $P(v)$, which equivalents to maximize $\log P(v)$. Now considering that there are N samples, let $L(\theta) = \log P(v)$, the key step is to calculate the partial derivatives of each parameters of the model as follows:

$$\frac{\partial L(\theta)}{\partial \theta} = \sum_{n=1}^N \left(\left\langle \frac{\partial(-E(v^{(n)}, h|\theta))}{\partial \theta} \right\rangle_{P(h|v^{(n)}, \theta)} - \left\langle \frac{\partial(-E(v, h|\theta))}{\partial \theta} \right\rangle_{P(v, h|\theta)} \right) \tag{7}$$

where $\langle \cdot \rangle_P$ is the mathematic expectation about distribution P . $P(h|v^{(n)}, \theta)$ is the probability distribution of the hidden layers when the known training sample of visual layer units is $v^{(n)}$; $P(v, h|\theta)$ is the joint distribution of the visual layer units and the hidden layer units.

The “data” and “model” are respectively used to indicate $P(h|v^{(n)}, \theta)$ and $P(v, h|\theta)$. Then taking partial derivative of $L(\theta)$ to w_{ij} , a_i and b_j can be derived as follows:

$$\frac{\partial \log P(v|\theta)}{\partial W_{ij}} = \langle v_i h_j \rangle_{data} - \langle v_i h_j \rangle_{model} \tag{8}$$

$$\frac{\partial \log P(v|\theta)}{\partial a_i} = \langle v_i \rangle_{data} - \langle v_i \rangle_{model} \tag{9}$$

$$\frac{\partial \log P(v|\theta)}{\partial b_j} = \langle h_j \rangle_{data} - \langle h_j \rangle_{model} \tag{10}$$

where $\langle \cdot \rangle_{data}$ is the mathematic expectation of the data set and $\langle \cdot \rangle_{model}$ is the mathematic expectation defined in the model.

For the calculation of the gradient of $L(\theta)$ with respect to θ , the CD (Contrastive Divergence) algorithm is introduced in this paper. The steps can be summarized as follows.

- Let the initial state of the visual layer unit v_0 equals x_0 . The connection weights W needs to be initialized. The bias a and b of the visual and hidden layers should be small random numbers obeyed the Gaussian distribution. Then the maximum number of iterations for each layer of RBM is also needed to be specified.
- For all hidden layer units, $P(h_{1j} = 1|v_1)$ can be calculated according to the formula as follows:

$$P(h_{1j} = 1|v_1) = \sigma \left(b_j + \sum_i v_{1i} W_{ij} \right) \tag{11}$$

where $\sigma(x)$ is the sigmoid function.

- For all visual layer units, $P(v_{2i} = 1|h_1)$ can be calculated according to the formula as follows:

$$P(v_{2i} = 1|h_1) = \sigma \left(a_i + \sum_j h_{1j} W_{ij} \right) \tag{12}$$

where $v_{2i} \in \{0, 1\}$ can be extracted from conditional distribution $P(v_{2i} = 1|h_1)$.

- For all hidden layer units, $P(h_{2j} = 1|v_2)$ can be calculated according to the formula as follows:

$$P(h_{2j} = 1|v_2) = \sigma \left(b_j + \sum_i v_{2i} W_{ij} \right) \tag{13}$$

- Each parameter needs to be upgraded according to the following formulas:

$$W \leftarrow W + \varepsilon \left(P(h_1 = 1|v_1) v_1^T - P(h_2 = 1|v_2) v_2^T \right) \tag{14}$$

$$a \leftarrow a + \varepsilon (v_1 - v_2) \tag{15}$$

$$b \leftarrow b + \varepsilon (P(h_1 = 1|v_1) - P(h_2 = 1|v_2)) \tag{16}$$

3.1.2 Fine-Tuning

Fine-tuning is a supervised training process with labeled data, which use the BP neural network algorithm to optimize the parameters of each layer so that the predication performance of the network is better. After the process of pre-training, the initialization parameters of each RBM can be obtained. Then parameters of DBN must be to be tuned and trained, the network parameters of each layer can be optimized, bringing the performance of neural network forecasting more effective. By receiving the output feature vector of RBM with BP neural network algorithm as the input feature vector, the entity relationship classifier can be supervised trained with labeled data. What follows are the characteristics of RBM network learning can be combined and classified, and the error information can be transferred to all the RBM networks based on the error function. Then the parameters of the whole DBN network can be fine-tuned to ensure that the final results are the optimal parameters of the DBN network.

3.2 Softmax Regression

Softmax regression is a generalization of logistic regression, it is adapt to solve the nonlinear multiple classification problems. In this paper, the Softmax classifier was added to the top layer of DBN neural network. It combined the supervised Softmax regression model with the unsupervised learning parts of DBN neural network to make a better judgment performance.

Suppose the Softmax regression model is a training set consisting of m training samples $\{(x^{(1)}, y^{(1)}), \dots, (x^{(m)}, y^{(m)})\}$, any of the sample $x^{(i)} \in R^{n+1}$, its corresponding classification label $y^{(i)} \in \{1, 2, \dots, k\}$. The hypothesis function of Softmax regression is represented as follows.

$$h_{\theta}(x^{(i)}) = \begin{bmatrix} p(y^{(i)} = 1|x^{(i)}; \theta) \\ p(y^{(i)} = 2|x^{(i)}; \theta) \\ \vdots \\ p(y^{(i)} = k|x^{(i)}; \theta) \end{bmatrix} = \frac{1}{\sum_{j=1}^k e^{\theta_j^T x^{(i)}}} \begin{bmatrix} e^{\theta_1^T x^{(i)}} \\ e^{\theta_2^T x^{(i)}} \\ \vdots \\ e^{\theta_k^T x^{(i)}} \end{bmatrix} \tag{17}$$

where the hypothesis vector $h_{\theta}(x^{(i)})$ is used to calculate the probability value $p(y^{(i)} = j|x^{(i)}; \theta)$ of which classification result j the test sample $x^{(i)}$ belongs to. θ is the parameter vector of the model, which can be expressed as a matrix as follows.

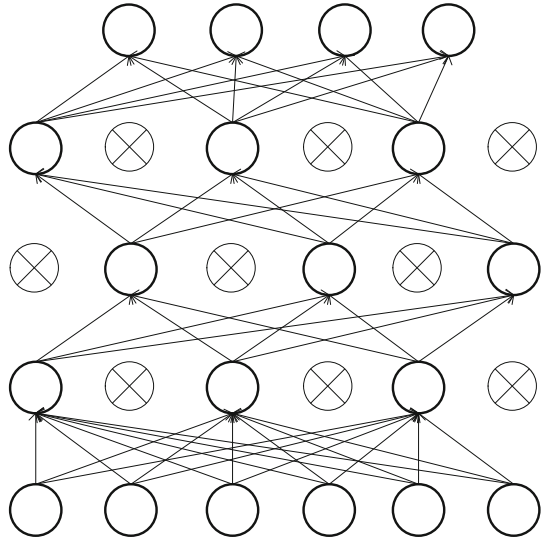
$$\theta = \begin{bmatrix} -\theta_1^T - \\ -\theta_2^T - \\ \vdots \\ -\theta_k^T - \end{bmatrix} \tag{18}$$

The cost function of model is defined as follows.

$$J(\theta) = -\frac{1}{m} \left[\sum_{i=1}^m \sum_{j=1}^k 1\{g^{(i)} = j\} \log \frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^k e^{\theta_l^T x^{(i)}}} \right] \tag{19}$$

where $1\{\cdot\}$ is indicator function. Its value rule is that $1\{\text{when the value of the expression is true}\} = 1$, otherwise $1\{\text{when the value of the expression is false}\} = 0$. Iterative optimal algorithm is usually used to solve the minimization problem. Through the derivation, we can get the cost function gradient formula is as follows.

Fig. 3 Deep neural network architecture of Dropout



$$\nabla_{\theta_j} J(\theta) = -\frac{1}{m} \sum_{i=1}^m \left[x^{(i)} (1\{y^{(i)} = j\} - p(y^{(i)} = j|x^{(i)}; \theta)) \right] \quad (20)$$

where $\nabla_{\theta_j} J(\theta)$ is a vector as the symbol for the partial derivative of $J(\theta)$ with respect to the l component θ_j .

3.3 Dropout Technology

Dropout is a random retreat mechanism to overcome the data problem of over-fitting, which is proposed by Hinton in 2012 [23]. Dropout mechanism is introduced in the DBN training process in this study. In the pre-training process of DBN network, random sampling of the hidden layer nodes weights with a certain probability, with the input and output nodes remaining constant. Such a thinner network is set each time, and some neurons are not involved in the forward propagation training process. In this paper, the random probability of Dropout was set at 50%. It can raise the generalization ability and improve the time-consuming problem of network training effectively so that the precision of forecasting can be improved distinctly. The deep neural network architecture of Dropout is shown in Fig. 3.

3.4 Model Setting

In this study, we construct the meteorological early warning model in Wenchuan earthquake affected area based on Dropout DBN-Softmax. As mentioned in Sect. 2.2, the inputs of the DBN neural network are the 9 factors which influence the landslides of Wenchuan earthquake affected area. Dropout mechanism is introduced into the pre-training process of DBN network raining process, random sampling of the hidden layer nodes weights with a certain probability. Trained by multi-layer RBM, we use the BP algorithm to optimize the parameters of each layer. At last, the outputs after Softmax classifier are the probability values of which kind the corresponding sample belongs to. The meteorological warning model of landslide disaster based on Dropout DBN-Softmax is shown in Fig. 4.

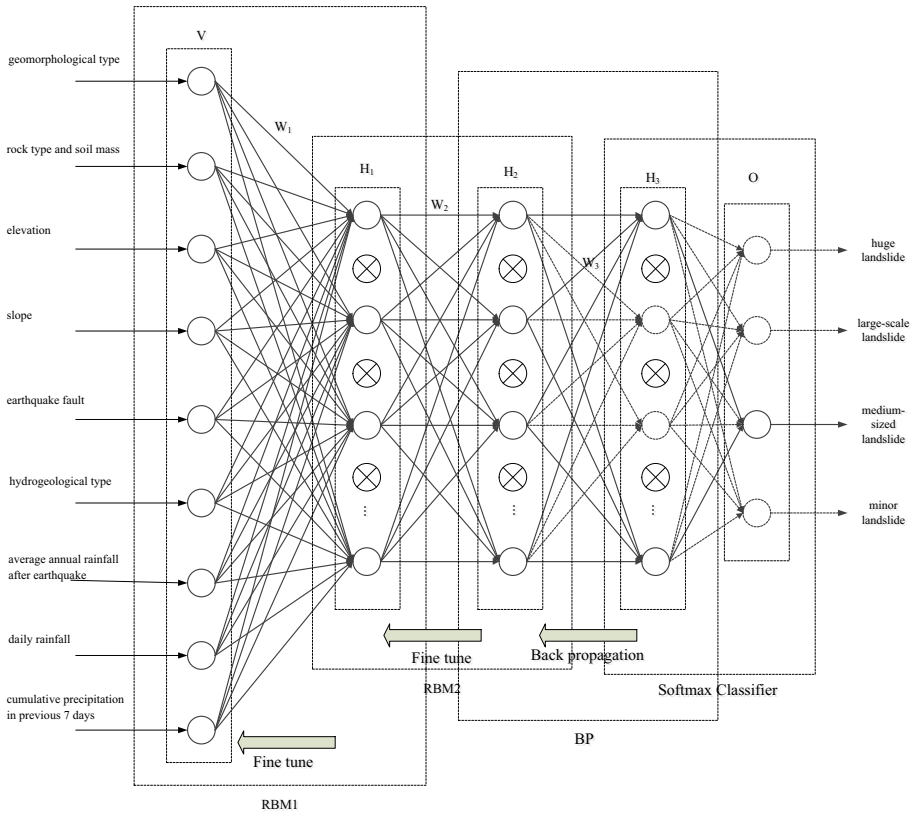


Fig. 4 The meteorological warning model of landslide disaster based on Dropout DBN-Softmax

The proposed method for meteorological early warning of landslide based on Dropout DBN-Softmax is as follows: (1) A large number of unlabeled data is used as input. The initialization parameters of the model are set to small random numbers which obeys the Gauss distribution. The CD algorithm is introduced to unsupervised pre-training process of the underlying RBM with pre-training samples. The Dropout method is used in the RBM hidden layer structure to unsupervised pre-train the network parameter W_0 . (2) The hidden layer of the first RBM is used as the input of the second RBM, with the second hidden layer to form the second RBM. Similarly, the Dropout mechanism is used in the second RBM hidden layer structure to train the network parameter W_1 . (3) Likewise, according to the training samples to unsupervised learn the RBM in the DBN layer by layer. The network parameter W of each RBM can be also trained and obtained. (4) Taking the trained network parameters as input of the Softmax classifier, and then training the network parameter of the Softmax classifier with labeled data. The cost function of the Softmax classifier can be calculated. (5) BP algorithm is used to train and learn the entire DBN network with labeled data. Then fine tune the parameters of each RBM and Softmax layer. In the Softmax layer, the parameter of the minimum value of the cost function in the Softmax classifier is obtained. In each RBM layer, the network parameters of each RBM in the whole DBN can be tuned to get the optimal parameters. (6) The well trained neural network was then taken as a prediction model to get the predict results of the landslide.

4 Experiments and Performance Assessments

In order to demonstrate the feasibility and assess the performance, a series of experiments were conducted and the detailed experimental data was given out. The testing server is composed of Intel Core I7-2600 at 3.4G HZ and 4GB of memory. The operating system of the experimental computer is Windows 10. We compiled with MATLAB R2015a.

In this experiment, the parameter setting of the meteorological warning model of landslide disaster based on Dropout DBN-Softmax is as follows:

- The visual note number of DBN is 9.
- The output note number of DBN is 4.
- Iteration number of each RBM layers is 10.
- Drop value is 0.5.
- The learning rate of BP neural network layer is 0.01.
- The accuracy of landslide disaster prediction is selected as the evaluation index of the prediction performance in the model.

Using the preprocessed training data set of landslides in Wenchuan earthquake affected area, the following experiments were carried out based on the Dropout DBN-Softmax model.

4.1 Different Number of Pre-training Samples and Different RBM Layers Effect

Under different number of pre-training samples and different RBM layers, some experiments were conducted to look in their relationships with the prediction accuracy of landslides. A series of experiments were carried out when the pre-training sample data were taken as 1000, 2000, 3000, 4000, 5000 for different network layers and the number of RBM network layers was taken as the different value of 1–8. The accuracy rate of the meteorological early warning of landslides is taken as the evaluation index. The experimental results are shown in Table 4.

It is found out in the experiments that the prediction accuracy of landslides is related to both the number of pre-training samples and the layers of RBM. After extensive experiments based on the experimental data, we plot the graph of the relation diagram among the number of pre-training samples, the layers of RBM and the prediction accuracy of landslides (Fig. 5).

As shown in Fig. 5, the trend of the x axis indicated the relationship between the number of pre-training samples and the accuracy of prediction when the RBM layer was constant. As the number of pre-training samples increased from 1000 to 5000, the prediction accuracy improved. It was consistent with the basic characteristics of deep learning. The results indicated that the prediction performance would be better with more sample data trained and more characteristics learned from DBN network. When the extracted features were good enough, the predicted results were gradually stabilized. The trend of the y axis showed the relationship between different layers of RBM and the accuracy of prediction when the pre-training sample number was given. The accuracy of the prediction of the landslide was obviously increased with the increase of the number of RBM. When the RBM layer reached 5, the trend became slow gradually.

4.2 Different RBM Layers and Different Nodes of RBM Effect

When the training samples kept constant as 5000, the influence of different RBM layers and nodes of RBM on prediction accuracy was studied experimentally. As shown in Table 5, as the RBM layers increased in the ranges from 1 to 8 and the nodes of each RBM changed, the prediction accuracy of landslides was a trend of rising with the rising of the RBM layers, and

Table 4 Table of pre-training sets and RBM layer parameters

RBM layers	Pre-training sets	Accuracy rate (%)	RBM layers	Pre-training sets	Accuracy rate (%)
1	1000	60.2	5	1000	70.5
1	2000	67.5	5	2000	76.5
1	3000	71.3	5	3000	84.2
1	4000	74.6	5	4000	89.5
1	5000	78.3	5	5000	92.5
2	1000	63.8	6	1000	73.5
2	2000	69.9	6	2000	81.5
2	3000	73.6	6	3000	86.4
2	4000	78	6	4000	90.6
2	5000	82.8	6	5000	92.6
3	1000	66.4	7	1000	75.1
3	2000	73.5	7	2000	84.1
3	3000	77.3	7	3000	88.5
3	4000	82.1	7	4000	90.8
3	5000	86.5	7	5000	92.7
4	1000	68.5	8	1000	77
4	2000	74.9	8	2000	87.5
4	3000	81.5	8	3000	90.8
4	4000	85.7	8	4000	91.8
4	5000	90.2	8	5000	92.7

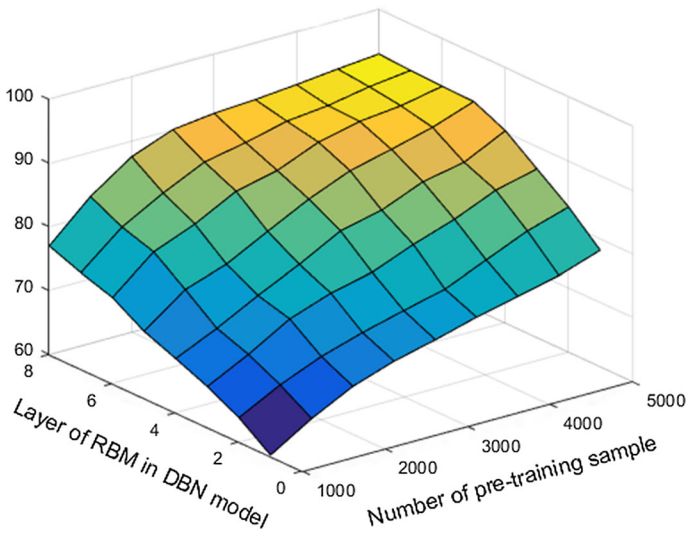


Fig. 5 The impact on prediction of landslides from different RBM layers and pre-training sets

Table 5 Prediction of landslides in the case of different hidden layers

Layers of RBM	Nodes of each RBM	Accuracy rate (%)	Layers of RBM	Nodes of each RBM	Accuracy rate (%)
1	20	78.3	5	20, 40, 40, 80, 40	92.5
2	20, 40	82.8	6	20, 40, 40, 80, 40, 20	92.6
3	20, 40, 40	86.5	7	20, 40, 40, 80, 40, 20, 20	92.7
4	20, 40, 40, 80	90.2	8	20, 40, 40, 80, 40, 20, 20, 20	92.7

then increases slowly. When the number of RBM layers was 5, the prediction accuracy of the landslide disaster was up to 92.5%. And later, it would not increase with the number of RBM layers and the nodes of each RBM increased. Therefore, when the number of RBM layers was 5, the prediction accuracy of Dropout DBN- Softmax model was obviously increased with the increase of the number of RBM. When the RBM layer reached 5, the trend became slow gradually. So the optimal structure of the Dropout DBN- Softmax prediction model could be determined.

4.3 Three Models Selected for Comparison

In order to evaluate the predictive effect of DBN, a comparison was made on traditional BP neural network model, BP neural networks learning algorithm based on Particle Swarm Optimizer (PSO-BP) model and the proposed Dropout DBN- Softmax model. When the training sets range from 1000 to 5000, the comparison of prediction results is shown in Table 6 and Fig. 6.

As shown in Table 6 and Fig. 6, when the training set is small, the average accuracy of the three models was basically equivalent. However, as the number of training samples increased, the average accuracy of the Dropout DBN-Softmax model was significantly higher than the BP and PSO-BP models. It was further verified the deep learning model was a kind of high efficient data processing method. The Dropout DBN-Softmax model could make full use of data information and could extract automatically the main factors affected the landslide disasters so as to improve the prediction accuracy effectively. Therefore, compared with the traditional shallow learning model (BP model and PSO-BP model), the deep learning method (Dropout DBN-Softmax model) was more applicable in a large number of sample data of landslide hazard prediction and training. With the most stable, the highest prediction accuracy and good scalability, it could better meet the actual needs.

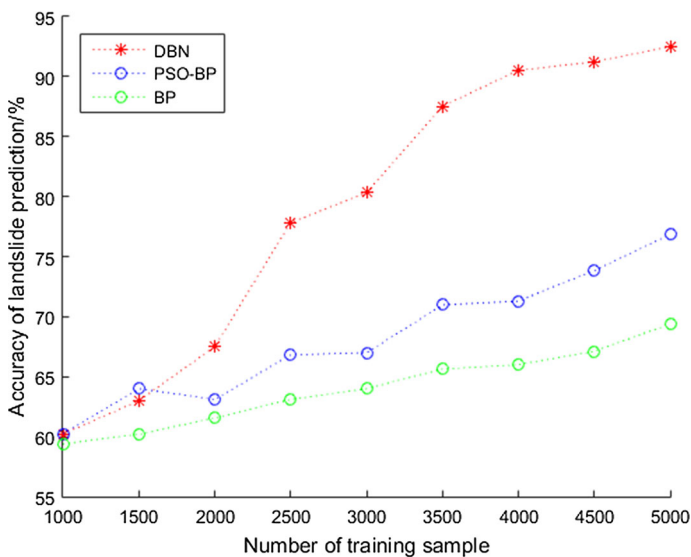
5 Summary and Conclusions

Considering the unsatisfactory results achieved with the probability and mathematical statistics and the shallow neural networks, we introduced a deep learning technique to improve the results of rainfall-triggering landslide forecast. The study has demonstrated that, at least for the study area, the new meteorological early warning method of precipitation-induced landslides based on Dropout DBN- Softmax is feasible and effective.

Deep Belief Network is a deep network structure stacked by a series of restricted Boltzmann machines (RBM). With the greedy unsupervised learning algorithm of DBN for data pre-training, a better distributed representation of the input data can be obtained. And then

Table 6 Prediction of landslide hazards in different training sets

Model	Training sets	Test sets	Accuracy rate (%)
BP	1000	1000	59.4
PSO-BP			60.23
Dropout DBN-Softmax			60.2
BP	2000	1000	61.56
PSO-BP			63.1
Dropout DBN-Softmax			67.5
BP	3000	1000	64
PSO-BP			67
Dropout DBN-Softmax			80.33
BP	4000	1000	66
PSO-BP			71.31
Dropout DBN-Softmax			90.52
BP	5000	1000	69.4
PSO-BP			76.84
Dropout DBN-Softmax			92.5

**Fig. 6** Predictions of BP, PSO-BP and Dropout DBN- Softmax models

using the BP neural network algorithm to optimize the parameters of each RBM layer, thus enabling the predictive performance of the network is optimal. The Softmax classifier is added to the top layer of DBN neural network, which combined the supervised Softmax regression model with the unsupervised learning parts of DBN neural network, to make a better judgment performance. Dropout mechanism is introduced in the DBN training process in this study to overcome the data problem of over-fitting.

Moreover, the meteorological early warning model based on Dropout DBN-Softmax was constructed. Using 9 factors influenced the landslides of the study area as the inputs of the DBN neural network. Dropout mechanism is introduced in the pre-training process to sample the hidden layer nodes weights with a probability of 50%. Trained by multi-layer RBM, we use the BP algorithm to optimize the parameters of each layer. Eventually, the outputs after Softmax classifier are the probability values of which kind the corresponding sample belongs to.

Finally, on the basis of several sets of experiments, the proposed method of precipitation-induced landslides based on Dropout DBN-Softmax is proved to be feasible with satisfactory result. Compared with the traditional BP and PSO-BP model, the new method proposed makes better prediction accuracy. Experiments show that as pre-training set rises, the prediction accuracy is increased. Therefore, the method proposed in this paper applies to training of a large number of data samples. And it has good scalability and great accuracy to meet the actual demand.

In the future research work, it needs to research other deep learning models in order to find a more suitable data model for landslide hazard prediction in our next study.

Acknowledgements The research was supported by the Project in 2014 by National Natural Science Foundation of China (Authorized Number: 41401449) and the University science and technology Project (No. J16LH05).

References

1. Gao ZK, Jin ND (2012) A directed weighted complex network for characterizing chaotic dynamics from time series. *Nonlin Anal Real World Appl* 13(2):947–952
2. Alvioli M, Baum RL (2016) Parallelization of the TRIGRS model for rainfall-induced landslides using the message passing interface. *Environ Modell Softw* 81:122–135
3. Lee S, Ryu JH, Won JS, Park HJ (2004) Determination and application of the weights for landslide susceptibility mapping using an artificial neural network. *Eng Geol* 71(3/4):289–302
4. Pham BT, Pradhan B, Bui DT, Prakash I, Dholakia MB (2016) A comparative study of different machine learning methods for landslide susceptibility assessment: a case study of Uttarakhand area (India). *Environ Modell Softw* 84:240–250
5. Colkesen I, Sahin EK, Kavzoglu T (2016) Susceptibility mapping of shallow landslides using kernel-based Gaussian process, support vector machines and logistic regression. *J Afr Earth Sci* 118:53–64
6. Kornejady A, Ownegh M, Bahremand A (2017) Landslide susceptibility assessment using maximum entropy model with two different data sampling methods. *Catena* 152:144–162
7. Gerolymos N, Gazetas G (2007) A model for grain-crushing-induced landslides—application to Nikawa, Kobe 1995. *Soil Dyn Earthq Eng* 27(9):803–817
8. Chen YC, Chang KT, Chiu YJ, Lau SM, Lee HY (2013) Quantifying rainfall controls on catchment-scale landslide erosion in Taiwan. *Earth Surf Process Landf* 38(4):372–382
9. Akcali E, Arman H (2013) Landslide early warning system suggestion based on landslide—rainfall threshold: Trabzon Province. *Teknik dergi* 24(1):6307–6332
10. Chen XL, Liu CG, Yu L, Lin CX (2014) Critical acceleration as a criterion in seismic landslide susceptibility assessment. *Geomorphology* 217:15–22
11. Pistocchi A, Luzi L, Napolitano P (2002) The use of predictive modeling techniques for optimal exploitation of spatial databases: a case study in landslide hazard mapping with expert system-like methods. *Environ Geol* 41(7):765–775
12. Soowoong K, Bogun P, Bong SS, Seungjoon Y (2016) Deep belief network based statistical feature learning for fingerprint liveness detection. *Pattern Recognit Lett* 77:58–65
13. Guillermo LG, Lucas CU, Mónica GL, Pablo MG (2016) Deep learning for plant identification using vein morphological patterns. *Comput Electron Agric* 127:418–424
14. Geng J, Wang HY, Fan JC, Ma SR (2017) Deep supervised and contractive neural network for SAR image classification. *IEEE Trans Geosci Remote Sens* 55(4):2442–2459

15. Gao ZK, Cai Q, Yang YX, Dang WD, Zhang SS (2016) Multiscale limited penetrable horizontal visibility graph for analyzing nonlinear time series. *Sci Rep* 6:35622
16. Hinton GE, Deng L, Yu D, Dahl E, Mohamed A, Jaitly N, Senior A, Vanhoucke V, Nguyen P, Sainath TN, Kingsbury B (2012) Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. *IEEE Signal Process Mag* 29(6):82–97
17. Ronao CA, Cho SB (2016) Human activity recognition with smartphone sensors using deep learning neural networks. *Expert Syst Appl* 59:235–244
18. Gao ZK, Cai Q, Yang YX, Dong N, Zhang SS (2017) Visibility graph from adaptive optimal kernel time-frequency representation for classification of epileptiform EEG. *Int J Neural Syst* 27(4):1750005
19. Shim H, Lee S (2015) Multi-channel electromyography pattern classification using deep belief networks for enhanced user experience. *J Central South Univ* 22(5):1801–1808
20. Papa JP, Scheirer W, Cox DD (2016) Fine-tuning deep belief networks using harmony search. *Appl Soft Comput* 46:875–885
21. Hinton GE, Osindero S, Teh YW (2006) A fast learning algorithm for deep belief nets. *Neural Computat* 18(7):1527–1554
22. Shen FR, Chao J, Zhao JX (2015) Forecasting exchange rate using deep belief networks and conjugate gradient method. *Neurocomputing* 167:243–253
23. Wu HB, Gu XD (2015) Towards dropout training for convolutional neural networks. *Neural Netw* 71:1–10