APPLIED GEOPHYSICS, Vol. 16, No. 3 (September 2019), P. 257—266, 11 Figures.
DOI:10.1007/s11770-019-0774-1

Anomaly detection of earthquake precursor data using long short-term memory networks*

Cai Yin[1,2], Mei-Ling Shyu[3], Tu Yue-Xuan[3], Teng Yun-Tian[1], and Hu Xing-Xing♦[1]

Abstract: Earthquake precursor data have been used as an important basis for earthquake prediction. In this study, a recurrent neural network (RNN) architecture with long short-term memory (LSTM) units is utilized to develop a predictive model for normal data. Furthermore, the prediction errors from the predictive models are used to indicate normal or abnormal behavior. An additional advantage of using the LSTM networks is that the earthquake precursor data can be directly fed into the network without any elaborate preprocessing as required by other approaches. Furthermore, no prior information on abnormal data is needed by these networks as they are trained only using normal data. Experiments using three groups of real data were conducted to compare the anomaly detection results of the proposed method with those of manual recognition. The comparison results indicated that the proposed LSTM network achieves promising results and is viable for detecting anomalies in earthquake precursor data.
Keywords: Earthquake precursor data, deep learning, LSTM-RNN, prediction model, anomaly detection

Introduction

China is the most intense and high frequency intraplate seismic activity region because of the interaction between the Indian and the Eurasian plates, combined with the westward subduction impact of the Pacific plate (Wu et al., 2013). Numerous strong earthquakes have been hitting mainland China since the last century. Such earthquakes are characterized by their shallow focal depths and wide distributions, and inflict great losses to lives and properties (Lu et al., 2018). Scientists in China have been investigating and conducting researches on earthquake prediction as early as the 1960s to effectively mitigate the disaster caused by earthquakes, and China has eventually become the pioneer in this field. After more than 50 years of development, although successful cases have been recorded, such as accurate prediction

Manuscript received by the Editor December 20, 2018; revised manuscript received July 6, 2019
*This work was supported by the Science for Earthquake Resilience of China (No. XH18027),Research and Development of Comprehensive Geophysical Field Observing Instrument in Mainland China (No. Y201703) and Research Fund Project of Shandong Earthquake Agency (Nos. JJ1505Y and JJ1602).
1. Institute of Geophysics, China Earthquake Administration, Beijing 100081, China.
2. Shandong Earthquake Agency, Jinan 250014, China.
3. Department of Electrical and Computer Engineering, University of Miami, Coral Gables, FL 33146, USA.
♦Corresponding author: Hu Xing-Xing (Email: huxx@cea-igp.ac.cn)

of the Haicheng earthquake in 1975 (Wang et al., 2006), the success rate is rare and a lot of advancement is needed. Generally, earthquake prediction still remains in the primary stage of obtaining facts based on observation and accumulating empirical knowledge. However, scientists have deepened their understanding of earthquake prediction and have concluded earthquake precursors as the most direct sign of earthquake occurrence. Thus, searching for precursors is important in earthquake prediction studies (Huang et al., 2017; Tsai et al., 2018).

By analyzing the observation data from different domains such as crustal deformation, geoelectricity and geomagnetism, and ground fluid, scientists revealed the relationship between the anomalies in precursor data and the occurrence of large earthquakes (Ouyang et al., 2009; Ye et al., 2015; Yao et al., 2012; Han et al., 2015; Shirokov et al., 2015). Earthquake precursor anomaly data can be divided into two main categories: tendency change and high-frequency mutation. Tendency turning refers to the abnormal change in tendency compared with the normal periodic variations. High-frequency disturbance refers to the abrupt change in observation data with high frequency and large amplitude, which commonly has irregular patterns. These anomalies usually overlap in actual observations, which greatly complicates their detection.

Most precursor anomaly detection methods currently available are based on statistical knowledge, where a baseline and the changing rate of normal data are first calculated, using which any observation beyond a certain range of variations is considered abnormal (Huang et al., 2017). In addition, some techniques based on digital signal analysis are used to extract the abnormal data such as wavelet transform (Grossmann and Morlet, 1984; Zhang et al., 2016; Masci and Thomas, 2015), and empirical mode decomposition (Huang et al., 1998; Wang et al., 2017; Shah and Jin, 2018). These methods convert the observation data from the time domain to the frequency domain, and thus the anomaly can be analyzed and detected in detail. The above-mentioned detection techniques are applicable only to certain type of anomalies. However, in the actual observation data, different types of anomalies commonly overlap; thus, generalizing the current methods to detect all types of anomalies with relatively high accuracy is difficult (Yang et al., 2017). Therefore, a new approach that can identify most anomalies in a unified manner is required.

Deep learning techniques have become the most popular machine learning method. They can extract deep features through the self-learning process of data and achieve the best performance in both supervised and unsupervised learning areas (LeCun et al., 2015; Pouyanfar et al., 2018). Further, the recursive neural network (RNN) is a popular deep learning model that exhibits a dynamic temporal behavior for a time sequence by building connections between the current and previous states (Williams and Zipser, 1989; Bengio et al., 1994; Hochreiter et al., 2001; Gers et al., 2000; Greff et al., 2017). The long short-term memory (LSTM) network is an RNN comprising LSTM units and has been widely employed in numerous data analysis fields such as handwriting recognition, voice recognition, and machine translation (Soltau et al., 2016; Graves et al., 2013; Frinken and Uchida, 2015).

The earthquake precursor data is a collection of various continuously changing physical quantities belonging to typical time series data. Leveraging the capability of the LSTM networks to handle long-term dependencies is desired to capture deeper features that characterize the changing pattern of data. Such features are believed to improve the accuracy of automatic earthquake anomaly identification and bridge the gaps between deep learning techniques and earthquake anomaly detection.

However, the current anomaly detection methods pose three main problems: 1) they largely depend on manual recognition, whose judgement is subjective because different experts propose different outcomes for the same data; 2) they are inconsistent because their scope of application is only specific and they are inapplicable to process multiple types of data; 3) they cannot be employed for real-time detection because most circumstances necessitate great preprocessing works before the data can be analyzed. In addition, the outcome of these methods highly depends on the cumbersome parameters tuning, making them more time-consuming.

To address the aforementioned three problems, this study proposes an anomaly detection method for earthquake precursor data based on unsupervised learning. We first build an LSTM network model to learn the characteristics of normal data, and then use it to predict future trends and variations. The differences between the predictive and actual observation data serve as quantitative indicators of anomalies. The novelties of this method are the following: 1) the model simulates the manual identification process, and the training is unsupervised, which generates recognition results in a more unified manner; 2) this method is applicable to various types of observation data and does not require a specific knowledge of each data type; and 3) the raw data can be directly fed into the model without any preprocessing, making it applicable in real-time.

Earthquake precursor data anomaly detection using LSTM

In this study, a new approach to addressing the anomaly detection problem of earthquake precursor data is presented. First, the LSTM network is used to establish a data prediction model. Next, the process of using the model to detect anomaly data in the testing dataset is described. Finally, the details of the algorithm and the procedure are described.

LSTM Network

Compared with the conventional RNNs, the LSTM network replaces the neurons in the hidden layer with LSTM units to circumvent the disappearing gradient problem. An LSTM network comprises multiple hidden LSTM layers, where each hidden layer represents a time step and consists of multiple LSTM units, as shown in Figure 1. When expanded, it is equivalent to a feedforward neural network that shares the parameters between layers. The output of each layer is used as the input of the subsequent layer; thus, it can process time series data of an arbitrary length and extract more complex (abstract) features (Sak et al., 2014).

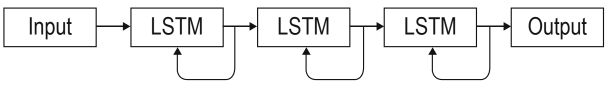

Fig. 1 The LSTM network architecture (Hermans and Schrauwen, 2013).

An LSTM unit commonly consists of a memory cell and several gates, where the cell stores the state information and the gates control when and how the states are updated. An LSTM structure is shown in Figure 2.

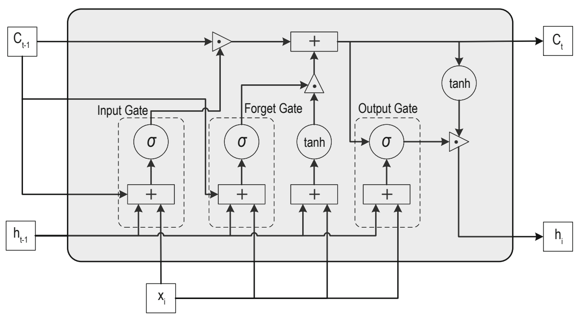

Fig. 2 Schematic of an LSTM structure (Greff et al., 2017).

Three types of gates are present in an LSTM unit, namely, input, output, and forget gates. The input gate determines the information to be updated, and the forget gate decides whether to retain or discard the memory of the previous state. Thus, updating of the cell state is achieved by the input and forget gates. The output gate determines the parts of the cell state to output. The equations below show the input and output, and the operation of the cell state and each gate (Graves et al., 2013).

$$\begin{aligned}
\tilde{c}_t &= \tanh(W_c \times (h_{t-1}, x_t) + b_c) && \text{(input)},\\
i_t &= \sigma(W_i \times (h_{t-1}, x_t) + b_i) && \text{(input gate)},\\
f_t &= \sigma(W_f \times (h_{t-1}, x_t) + b_f) && \text{(forget gate)},\\
o_t &= \sigma(W_o \times (h_{t-1}, x_t) + b_o) && \text{(output gate)},\\
c_t &= f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t && \text{(cell state)},\\
h_t &= o_t \cdot \tanh(c_t) && \text{(output)},
\end{aligned} \tag{1}$$

where i_t, f_t, o_t, c_t represent the states of the corresponding gates and cells at time t, x_t represents the input at time t, and $h_{t\text{-}1}$ and h_t represent the output at times t-1 and t, respectively, $\tilde{c}_t$ is a temporary state, W and b represent the weight matrix and bias vector, respectively, and σ denotes the sigmoid activation function.

Predication model

The predication model is built to enable the machine to learn normal characteristics of the data to predict future data trends and detect abnormal data. The LSTM is applied in this model, and the latter is trained using a large amount of normal data. The parameters of the model are continuously optimized to reduce the loss calculated from the loss function and achieve optimal learning results on the data features (Gupta et al., 2014).

The structure of the predication model is divided into three types of layers, namely, input, hidden, and output layers, as shown in Figure 3. The normal data are fed into the input layer as training sets. Moreover, multiple hidden layers exist between the input and output layers. The number of hidden layers and neurons in each layer are determined by the data set. The adjacent hidden layer adopts the full connection mode, and the dropout parameter is set to randomly suppress the output of some neurons to avoid overfitting in the training process, enabling complete training of all neurons. The size of the dropout parameter is inversely related to the amount of information in the dataset and directly related to the size of the network. The output layer is fully connected, and the number of its neurons equals the length of the output vector. In other words, each neuron in the output layer represents one step prediction output. Because the prediction model is mainly used for regression, linear activation is used in the output layer. The LSTM network is trained using the back propagation through time (BPTT) algorithm (Werbos, 1990). Finally, the mean squared error (MSE) is used as the loss function to characterize the prediction ability of the LSTM network model.

$$MSE = \frac{1}{n}\sum_{i}^{n}\left(y_i(\mathrm{t}) - \hat{y}_i(\mathrm{t})\right)^2, \qquad (2)$$

where y_i(t) is the predicted value of the model output at time t, $\hat{y}_i$(t) is the actual observed value at time t, and n is the total number of data samples.

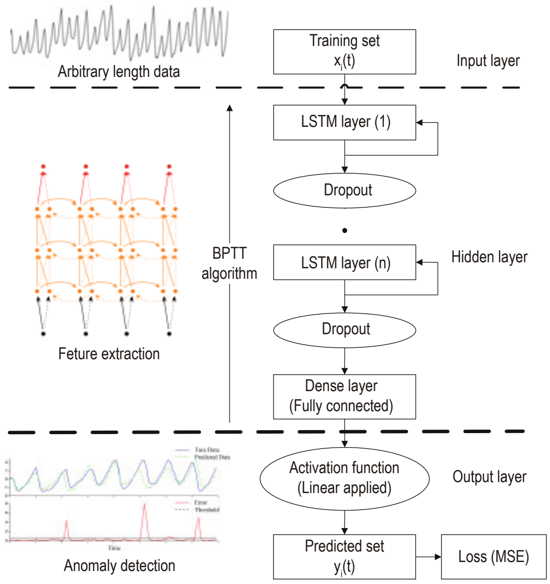

Fig. 3 Schematic of the LSTM prediction model.

Anomaly detection

The basis of the proposed detection method is to simulate the process of manual detection, where the changes in normal data follow certain dynamic patterns, and the anomaly is identified when the observation data deviate from a certain range of variations in the normal behavior. Thus, an anomaly detection method is proposed based on the deviation of the observation data from the predictive value obtained by the prediction model. The prediction error is used as a quantitative measure of the probability of an anomaly, i.e., the prediction error is calculated by the difference between the actual observed value at time t and the predictive value at time t-1 obtained from the predictive model. In addition, this prediction error can be interpreted as an offset from the normal changing trend.

To make the prediction error an effective indicator of anomaly, the following two requirements must be satisfied: 1) sufficient amount of training data must be available to enable the features learned by the prediction model to effectively represent the normal changing characteristics; and 2) the anomaly must have a greater prediction error than the normal data to enable setting a threshold of prediction error to distinguish the normal and abnormal data.

The anomaly detection process is divided into two steps. First, the prediction model on normal data is trained, and the anomaly threshold is determined by the training error, i.e., choosing the highest training error as the threshold value. Next, a point-to-point prediction is performed on the prediction data sets using our prediction model. Finally, an anomaly is identified when the prediction error is greater than the threshold value.

Algorithm steps

First, each dataset is divided into three subsets: 1) training set N_n, which only includes normal data; 2) validation set V_n, which is a subset of N_n and is used to validate the convergence of the training process while preventing overfitting; and 3) testing set T_n, which includes both normal and abnormal data and is used to test the performance of the model.

Then, the anomaly detection algorithm is implemented as follows.

1. Prediction model training: according to the characteristics of the training set, the hyperparameters of the model are initialized, including the number of hidden layers, number of neurons in each layer, dropout, and learning rate. Next, Bayesian optimization is used to find the optimal values of the hyperparameters. In general, the single-step prediction can detect abnormal data changes in time, and the prediction result is acceptable. Therefore, the prediction step size is set to 1 by default.

2. Prediction model saving: randomly select 5 % of the training set as the validation set and track the validation errors. Save the prediction model and validation errors when the model reaches the optimal point, while avoiding overfitting.

3. Anomaly threshold value setting: set the anomaly threshold value based on the validation errors such that the abnormal data are identified with the fewest false positives possible. The anomaly threshold is set to be twice the average of the validation errors based on empirical studies.

4. Anomaly detection: feed the testing dataset as the input to the prediction model and detect the abnormal data. The prediction model continuously predicts the value at the next time unit based on the input data. The

abnormal data is identified when the prediction error exceeds the anomaly threshold value, and this error serves as a quantitative measure for the anomaly data. The anomaly detection algorithm is illustrated in Figure 4.

As discussed in the section “Anomaly detection”, the recognition accuracy greatly depends on the ability of the model to learn the characteristics of normal data behavior. However, the actual changes and variations of the observation data are diverse. Thus, a manual refining of the hyperparameters mentioned in step 1 is performed to achieve relatively higher prediction errors of the identified anomaly data in the testing set. In this way, the anomaly threshold value can be increased to minimize the amount of normal data detected as abnormal.

Algorithm steps

procedure anomaly_detect (raw_data)
model, detector ←CREATE ()
N_n, V_n, T_n ←SPLIT(raw_data)
for *n* epochs and batch_size **do**
  model. TRAIN (N_n, V_n)
end for
$N\_Prediction_n$ ← ***model.*** PREDICT (N_n)
N_error_n ← Compute the Loss Function using (2.2)
thresholds ← ***detector.***SET_THRESHOLDS ($2\times\frac{1}{k}\sum_{n=1}^{k}N\_error_n$)
detector.APPLY(thresholds)
  for each T_n **do**
    $T\_Prediction_n$ ← ***model.*** PREDICT ( )
    T_error_n ← Compute the Loss Function using Equation (2)
      if (T_error_n thresholds) **then**
        T_n: is normal point
      else
        A_n ← T_n: is anomalous
      return $\{A_n\}$
  end for
end procedure

Fig. 4 Pseudo code of the proposed LSTM anomaly detection algorithm.

Experiments

To ensure that the proposed method can be generalized to handle different types of data, three datasets are selected from three different disciplines which include typical earthquake precursor anomaly. In addition, the advantages and robustness of our proposed method are demonstrated by comparing our detection results with those of manual recognition.

Datasets

Gravity dataset

Gravity observation is an important observation data in the discipline of deformation. The dataset was selected from China’s earthquake precursor dataset–the November 23, 2013 Laizhou Ms4.6 earthquake. An anomalous data set from China’s earthquake precursors–the daily average gravity observations from Tai’an, Shandong seismic station between 1/1/2011 and 12/31/2013–which contains 1095 samples. As shown in the upper image in Figure 5, the daily mean value exhibits an overall increasing trend. To identify the abnormal data better, the data are generally processed as first-order difference to eliminate the increasing trend, as shown in the bottom image of Figure 5. In the figure, the red box area represents the range of manually determined anomalies. This anomaly can be described as follows. Before the earthquake occurred, the increasing rate of the gravity observation data slowed down on 8/21/2013, the trend decreased on 2013/9/11, and the decline rate slowed down on 10/3/2013, showing a fluctuating decline. The normal increasing rate resumed on 11/15/2013, and an earthquake occurred a week later.

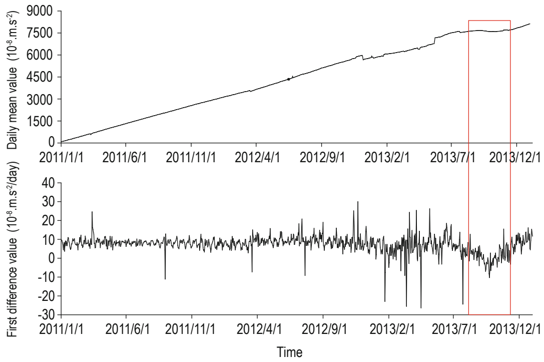

Fig.5 Gravity observation data for 2011–2013 in Tai'an Station.

Georesistivity dataset

Georesistivity is a representative observation method in the geoelectricity discipline. Currently, a georesistivity network of more than 80 stations has been built across China, which has accumulated rich observation and earthquake data. The observation data of most georesistivity stations show annual changes, i.e., the data are roughly periodic around a year, and the annual variation is generally 2–3 %. Because the changes in magnitude, shape, and phase of a periodic signal are anomalies themselves, identifying the anomaly from the observation data superimposed with the annual variation component and background noise is more challenging. Consider the 5/12/2018 Wenchuan Mw7.9 earthquake as an example. The daily mean value of N 58°E track obtained during October 2002–June 2008 in Chengdu station is selected. The manual detection methods can only intuitively predict a declining trend in georesistivity data from 2005 to 2007, when the previous annual changes are less apparent; this pattern resumes its appearance about 2 months before the earthquake. The red box in Figure 6 indicates the manually determined anomaly data; however, determining the exact time range and variation of this anomaly before an earthquake

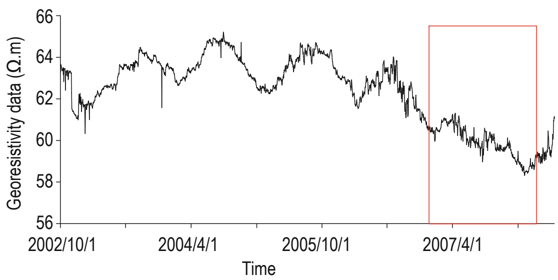

Fig. 6 Georesistivity observation data of October 2002–June 2008 in Chengdu station.

is difficult.

Water level dataset

Numerous observations indicate that the prominent anomaly data of ground fluid before strong earthquakes are in the medium and short-medium terms. The most representative ground fluid data are water level and water radon observations. Because the dynamic variation characteristics of the groundwater level data are complex, and a single well hole shows different reflection effects for different earthquakes, detecting an anomaly directly from the raw observation data is generally very difficult. We consider the water level observation data obtained from the Gejiu station as an example. The average water level of each month during May 1982–May 2008 is selected, which includes a total of 313 data samples, as shown in Figure 7 The figure clearly indicates that the most significant feature of the 26-year water level observation data is its annual periodic changes, but the annual variation is inconsistent and mostly irregular. Thus, intuitively determining an abnormal change in the data from this image is impossible.

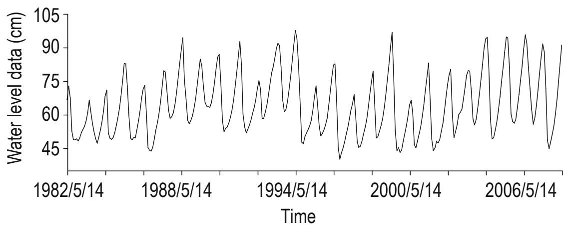

Fig.7 Water level observation data of May 1982–May 2008 in Gejiu station.

Results and analyses

A detailed discussion of the anomaly detection results of the aforementioned three datasets using the proposed LSTM network model is presented in this section. As mentioned in section

"Algorithm steps" , for each dataset, the part of the data that changes normally is first selected as the training data, and the rest is used as the testing data. When the data exhibit apparent periodic changes, the integrity of the periods is preserved to the maximum extent possible in the segmentation process to improve the ability of the model to learn the overall trend of the data. Simultaneously, zero-mean normalization is performed on all the data to increase the speed of convergence of model training. The model structure and the corresponding parameters used in the three datasets are provided in Table 1.

Table 1 Details of the prediction models used for each dataset.

Dataset		Network Architecture	Optimizer	Lookback	Batch Size	MSE (Training)	Threshold
Gravity	Daily Mean value	Recurrent: {60} Dropout: 0.1 Recurrent: {10} Dropout: 0.1 Dense: {1} Linear Activation	Learning Rate: 0.05 Decay: 0.99	40	3	0.001	0.002
	First Difference Value	Recurrent: {80} Dropout: 0.2 Recurrent: {20} Dropout: 0.2 Dense: {1} Linear Activation	Learning Rate: 0.03 Decay: 0.99	30	4	2	4
Georesistivity		Recurrent: {64} Dropout: 0.05 Recurrent: {256} Dropout: 0.05 Recurrent: {100} Dropout: 0.05 Dense: {1} Linear Activation	Learning Rate: 0.01 Decay: 0.99	200	50	0.1	0.2
Water Level		Recurrent: {100} Dropout: 0.2 Recurrent: {30} Dropout: 0.2 Dense: {1} Linear Activation	Learning Rate: 0.1 Decay: 0.99	50	8	0.25	0.5

Gravity dataset

The gravity observation dataset indicates that the overall data change is more stable before October 2012, and the anomalies mainly occur after this period. Therefore, the data obtained in January 2011–June 2012 are used as the training dataset, with the model parameters listed in Table 1. The anomaly threshold value is set as 0.002 because an average training error of 0.001 is obtained. Then, the model is used to predict the results for the testing dataset, which are shown in Figure 8.

In the upper plot of Figure 8, the solid blue line is the raw observation data which correspond to the testing dataset, and the green solid line is the point-by-point prediction result from the prediction model. The lower plot in Figure 8 shows the mean square error between the actual observation and the predicted value, and the black solid line is the anomaly threshold value. Similar to the manual detection results, the predictive model can accurately identify the abnormal trend of the observed data after 8/21/2013. In addition, the prediction error gradually declines when the observed data begin to resume their previous changing trend. However, the difference from the manual detection is that the prediction error shows an increasing trend since 12/12/2012, when there is a sudden drop. When the data suddenly rises on 5/26/2013, the error shows a significant high value jump, and then stays above the anomaly threshold value until the data resume their normal behavior. Therefore, the proposed model can detect anomalies that are not easily identified by humans, detect abnormal data in advance, and quantitatively describe the abnormality.

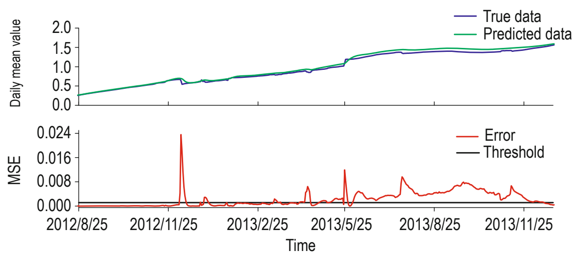

Fig. 8 Gravity data anomaly detection results.

The prediction results for the first-order differential data are shown in Figure 9. Moreover, the predictive model can accurately identify the abnormal changes between August 2013–October 2013 with a significant long-term variation in the errors. In addition, the error curve indicates a significant increase in high frequency disturbance from 1/20/2013, and the amplitudes of errors are much larger than the previous ones. These abnormal changes extracted by the proposed model are not identified by manual detection or are difficult to be described solidly and confidently by experts.

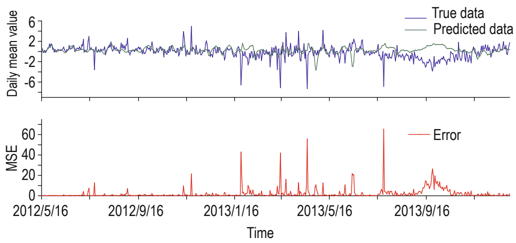

Fig.9 Gravity first-order differential data anomaly detection results.

Georesistivity dataset

Because the georesistivity data shows an annual changing pattern, the data obtained during 1/1/2014–9/30/2015 are selected as the training data to ensure that two complete annual change cycles are present in the training process. The anomaly threshold is set as 0.2 because the average training error is 0.1. The prediction results of the testing dataset are shown in Figure 10.

The plot of the prediction errors shows that since June 2006, the observed data exhibits apparent fluctuations and abnormal changes, and some prediction errors exceed the anomaly threshold value. Since March 2017, the predicted value begins increasing beyond the observed value because of the disappearance of the normal annual change; thus, the prediction error gradually increases and exceeds the anomaly threshold value. Further, an accelerated decline of the observed data is seen during November 2007–May 2008, and such trend did not appear in the training dataset. Therefore, the prediction model does not make a corresponding trend prediction, and the prediction error increases rapidly. From March 2008, the data resumes the normal rising behavior; thus, the error begins decreasing, and gradually reaches below the anomaly threshold value. On 5/12/2018, the Wenchuan Mw7.9 earthquake occurred at a distance 35 km from the Chengdu observation center.

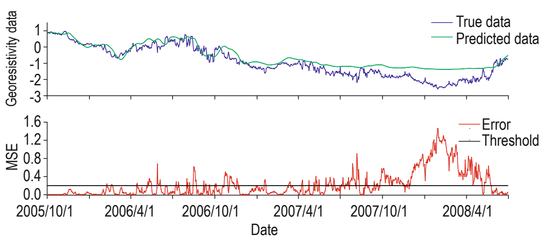

Fig.10 Georesistivity data anomaly detection results.

Water level dataset

Two devastating earthquakes of magnitude 6 or above occurred in the 500 km range of the Yunnan Gejiu station since June 2000, namely, Yunnan Dayao M6.2 earthquake on 7/21/2003 and Yunnan Pu'er M6.4 earthquake on 6/3/2007. To detect anomalies in the groundwater level data before the occurrence of these two earthquakes, a total of 217 data samples obtained in May 1982–May 2000 are used as the training sets, whose model parameters provided in Table 1. Therefore, the anomaly threshold value is set as 0.5, and the prediction results of the testing dataset are shown in Figure 11.

The results of the prediction errors clearly indicate an apparent increase in the errors during the three periods July–September 2002, June–October 2005, and July–October 2007, which values are well beyond the anomaly threshold value. An analysis and comparison of the predicted values and actual observations revealed that the three anomalies are mainly caused by changes in the annual variation and phase of the data. In the two years after the first two abnormal changes, large earthquakes higher than level 6 occurred, which must be a medium-term earthquake anomaly. This result is consistent with that of the underground fluid anomalies generally characterized by the medium and short-term anomalies. However, the anomaly observed during July–October 2007 is more likely to be because of the coseismic effect of the 6.4 magnitude earthquake in Pu'er, Yunnan on 6/3/2007. Therefore, the proposed model is applicable for the anomaly detection of long-term periodically changing data, and is capable of accurately capturing the subtle variations in cyclical changes.

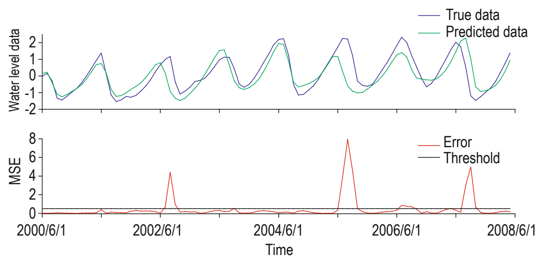

Fig. 11 Water level data anomaly detection result.

Conclusions

This study proposed a novel prediction model based on the LSTM network for the earthquake precursor data, which can be employed for automatic anomaly detection. To test the reliability and generalizability of this method, three different data sets from three disciplines of earthquake precursors were used as the testing datasets.

The findings of our experiments are as follows: 1) The LSTM network model effectively alleviates the gradient disappearance problem, and the anomaly recognition method is applicable for data sets of different lengths. 2) The detection method does not require specific background knowledge and preprocessing of data, and can be easily applied to different subject areas. and 3) The detection method can not only determine the type of anomalies, but also quantitatively describe them.

We observed from experiments that the amount of training data is an important factor for the accuracy of prediction results. The model needs a large amount of normal data to learn the characteristics of normal behaviors. With the accumulation of precursor observation data, further increasing the amount of data in the training set and using Gaussian distribution to model the prediction error of the training dataset will be investigated to improve the accuracy and reliability of anomaly detection. On the other hand, because the observation data differ for each station, establishing a prediction model to observe the data of each station is more accurate. We intend to employ Gated Recurrent Units (GRU) instead of LSTM to simplify the architecture of the neural network in the future. The GRU integrates the input and forget gates into an update gate, which is less computationally complex than the LSTM networks, and can produce similar results as the LSTM networks.

Acknowledgments

The authors appreciate the helpful discussion and technical support provided by the DDM Research Group of the University of Miami.

References

Bengio, Y., Simard, P., and Frasconi, P., 1994, Learning long-term dependencies with gradient descent is difficult: IEEE Transactions on Neural Networks, **5**(2), 157−166.

Chung, J., Gulcehre, C., Cho, K., and Bengio, Y., 2014, Empirical evaluation ofgated recurrent neural networks on sequence modeling.

Dey, R., and Salemt, F.M., 2017, Gate-variants of gated recurrent unit (GRU) neural networks: IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS), 1597−1600.

Frinken, V., and Uchida, S., 2015, Deep BLSTM neural networks for unconstrained continuous handwritten text recognition: 13th International Conference on Document Analysis and Recognition (ICDAR), IEEE, 911−915.

Gers, F., Schmidhuber, J., and Cummins, F., 2000, Learning to forget: continual prediction with LSTM: Neural Computation, **12**(10), 2451−2471.

Graves, A., Mohamed, A.-r., and Hinton, G., 2013, Speech recognition with deep recurrent neural networks: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 6645−6649.

Greff, K., Srivastava, R.K., Koutník, J., Steunebrink, B.R., and Schmidhuber, J., 2017, LSTM: A search space odyssey: IEEE Transactions on Neural Networks Learning Systems, **28**(10), 2222−2232.

Grossmann, A., and Morlet, J., 1984, Decomposition of hardy functions into square integrable wavelets of constant shape: SIAM Journal on Mathematical Analysis, **15**(4), 723−736.

Gupta, M., Gao, J., Aggarwal, C.C., and Han, J., 2014, Outlier detection for temporal data: A survey: IEEE Transactions on Knowledge Data Engineering, **26**(9), 2250−2267.

Han, P., Hattori, K., Xu, G., Ashida, R., Chen, C.-H., Febriani, F., and Yamaguchi, H., 2015, Further investigations of geomagnetic diurnal variations associated with the 2011 off the Pacific coast of Tohoku earthquake (Mw 9.0): Journal of Asian Earth Sciences, **114**, 321−326.

Hermans, M., and Schrauwen, B., 2013, Training and analysing deep recurrent neural networks: Advances in neural information processing systems, 190−198.

Hochreiter, S., Bengio, Y., Frasconi, P., and Schmidhuber, J., 2001, Gradient flow in recurrent nets: the difficulty of learning long-term dependencies: A field guide to dynamical recurrent neural network, Wiley-IEEE Press, USA, 1−15.

Hochreiter, S., and Schmidhuber, J., 1997, Long short-term memory: Neural Computation, 9(8), 1735-1780.

Huang, F., Li, M., Ma, Y., Han, Y., Tian, L., Yan, W. and Li, X., 2017, Studies on earthquake precursors in China: A review for recent 50 years: Geodesy and Geodynamics, **8**(1), 1−12.

Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.-C., Tung, C.C., and Liu, H.H., 1998, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis: Proceedings of the Royal Society of London. Series A: Mathematical, Physical and

Engineering Sciences, **454**(1971), 903−995.

Huang, Q., 2011, Retrospective investigation of geophysical data possibly associated with the Ms8. 0 Wenchuan earthquake in Sichuan, China: Journal of Asian Earth Sciences, **41**(4-5), 421−427.

LeCun, Y., Bengio, Y., and Hinton, G., 2015, Deep learning: Nature, **521**(7553), 436−444.

Lu, K., Hou, M., Jiang, Z., Wang, Q., Sun, G., and Liu, J., 2018, Understanding earthquake from the granular physics point of view — Causes of earthquake, earthquake precursors and predictions: International Journal of Modern Physics B, **32**(07), 1−39.

Masci, F., and Thomas, J.N., 2015, Ultra low frequency (ULF) electromagnetic anomalies associated with large earthquakes in Java Island, Indonesia by using wavelet transform and detrended fluctuation analysis: Natural Hazards and Earth System Sciences, **15**(12), 2697−2701.

Ouyang, Z., Zhang, H., Fu, Z., Gou, B., and Jiang, W., 2009, Abnormal phenomena recorded by several earthquake precursor observation instruments before the Ms 8.0 Wenchuan, Sichuan earthquake: Acta Geologica Sinica-English Edition, **83**(4), 834−844.

Pouyanfar, S., Sadiq, S., Yan, Y., Tian, H., Tao, Y., Reyes, M.P., Shyu, M.-L., Chen, S.-C., and Iyengar, S., 2018, A survey on deep learning: algorithms, techniques, and applications: ACM Computing Surveys, **51**(5), 92.

Sak, H., Senior, A., and Beaufays, F., 2014, Long short-term memory recurrent neural network architectures for large scale acoustic modeling: 15th Annual Conference of the International Speech Communication Association, 338−342.

Shah, M., and Jin, S., 2018, Pre-seismic ionospheric anomalies of the 2013 Mw = 7.7 Pakistan earthquake from GPS and COSMIC observations: Geodesy and Geodynamics, **9**(5), 378−387.

Shirokov, V., Firstov, P., Makarov, E., and Stepanov, I., 2015, An Approach to the short-and long-term forecasting of strong earthquakes: A case study of the Mw = 9.0 Tohoku earthquake, Japan, March 11, 2011: Seismic Instruments, **51**(3), 229−241.

Soltau, H., Liao, H., and Sak, H., 2016, Neural speech recognizer: Acoustic-to-word LSTM model for large vocabulary speech recognition: arXiv preprint arXiv:1610.09975,

Tsai, Y.B., Liu, J.Y., Shin, T.C., Yen, H.Y., and Chen, C.H., 2018, Multidisciplinary earthquake precursor studies in Taiwan: A review and future prospects: Pre-earthquake processes: A Multidisciplinary Approach to Earthquake Prediction Studies, John Wiley & Sons, Inc., USA, 41−65.

Wang, J., Chen, X., Kang, B., and Li, X., 2017, Groundwater micro-dynamic extraction and its response to seismic activity based on EMD: Ninth International Conference on Digital Image Processing (ICDIP 2017), SPIE, 1−5.

Wang, K., Chen, Q.-F., Sun, S., and Wang, A., 2006, Predicting the 1975 Haicheng earthquake: Bulletin of the Seismological Society of America, **96**(3), 757−795.

Werbos, P.J., 1990, Backpropagation through time: what it does and how to do it: Proceedings of the IEEE, **78**(10), 1550-1560.

Williams, R.J., and Zipser, D., 1989, A learning algorithm for continually running fully recurrent neural networks: Neural Computation, **1**(2), 270−280.

Wu, Z., Ma, T., Jiang, H., and Jiang, C., 2013, Multi-scale seismic hazard and risk in the China mainland with implication for the preparedness, mitigation, and management of earthquake disasters: An overview: International Journal of Disaster Risk Reduction, **4**, 21−33.

Yang, D.-H., Yuan, J., Wang, X.-Y., Shen, X.-H., Teng, H.-T., Li, W.-J., Tan, Q., and Wei, Q., 2017, Identification of multi-anomalies of precursory deformation data: Chinese Journal of Geophys-Chinese Edition (in Chinese), **60**(12), 4623−4632.

Yao, Y., Chen, P., Wu, H., Zhang, S., and Peng, W., 2012, Analysis of ionospheric anomalies before the 2011 Mw 9.0 Japan earthquake: Chinese Science Bulletin, **57**(5), 500−510.

Ye, Q., Singh, R.P., He, A., Ji, S., and Liu, C., 2015, Characteristic behavior of water radon associated with Wenchuan and Lushan earthquakes along Longmenshan fault: Radiation Measurements, **76**, 44−53.

Zhang, W., Huang, W., Li, L., Liu, W., and Li, F., 2016, High resolution strain sensor for earthquake precursor observation and earthquake monitoring: Sixth European Workshop on Optical Fibre Sensors (EWOFS'2016), SPIE, 1−5.

Yin Cai is a senior engineer in Shandong Earthquake Agency and graduated from Dalian University of Technology in 2007, with a master's degree in software engineering. He is a Ph.D. candidate in Institute of Geophysics, China Earthquake Administration. His researches primarily focus on the application of deep learning methods and cloud computing technology in geoscience.

Email: caiyin555@icloud.com ;