

Pressure prediction and abnormal working conditions detection of water supply network based on LSTM

Zhe Xu, Zhihao Ying, Yuquan Li, Bishi He and Yun Chen

ABSTRACT

In this study, a deep learning model based on LSTM (Long Short-Term Memory) is used to predict the state of a water supply network due to its highly complex nonlinearity. The inputs of the model include state information on the pressures at measuring points, as well as control information on the water supply pressure and flow at each entry point. In order to enhance the performance of the model in feature extraction and identification and improve prediction accuracy, a parallel LSTM tandem DNN deep neural network model (PLDNN) is proposed. The experimental results indicate that the model has better learning performance and accuracy compared with traditional prediction methods (artificial neural networks, support vector machines, etc.) and general LSTM models.

Key words | anomaly detection, LSTM, pressure prediction, water supply network

Zhe Xu (corresponding author)
Zhihao Ying
Yuquan Li
Bishi He
Yun Chen
Hangzhou Dianzi University,
Hangzhou 310018,
China
E-mail: xuzhe@hdu.edu.cn

INTRODUCTION

A water supply network (WSN) is a large-scale multi-source/multi-node flow system to deliver water with sufficient pressure and quantity to users. It is characterized by a large variety and complexity, and in particular, its hydraulic dynamics are time-variant, spatially distributed and highly nonlinear (Rao & Alvarruiz 2007). To ensure the high efficiency, reliability and resilience of WSN, people have done a lot of research work on optimizing control and scheduling, leakage anomaly detection and location and pressure management and leakage reduction, etc. over the past two decades, recently in real-time control especially (Creaco *et al.* 2019) and on-line burst detection (Wu & Liu 2017).

Generally, nodal pressure and pipe flowrate are monitored in an urban water supply network. Since most pipe network monitoring points measure pressures, dispatchers usually observe the operation status of the water supply pipe network based on the measured pressure data, to control the pressure of the whole pipe network in a reasonable range, and judge any abnormality according to local pressure loss. Therefore, it is of great significance to predict the pressures at the measuring points accurately and rapidly for real-time scheduling and scientific management.

The traditional methods of predicting pressures at measuring points of the water supply network use statistical analysis to establish a correlation between the main variables based on the historical data of network operation (Lv *et al.* 2001; Luvizotto *et al.* 2012). Currently, machine learning models, such as artificial neural network (ANN) and support vector machine (SVM), have been applied and have achieved good results in predicting the state of water supply networks. Jamieson *et al.* (2007) and Rao & Salomons (2007) built a macro back-propagation (BP) neural network model by applying genetic algorithm (GA) on lots of simulation data from EPANET, which is a steady-state hydraulic simulator for water supply networks. Yu *et al.* (2005) and Ping *et al.* (2014) built support vector machine models on measuring point pressures of a water supply network. Zhang *et al.* (2011) built a BP neural network model on the inputs and outputs of a water supply network. Perea *et al.* (2019) combined dynamic ANN architecture, Bayesian framework and GA, and proposed a new method to predict short-term irrigation demand with limited data. Mehrparvar & Asghari (2018) used a modular method to build a model by combining support vector regression (SVR) and a modified

data assimilation (MDA) technique to partially correct the predicted values based on the observed data. Yang *et al.* (2014) proposed a method which combines embedded space technology and neural network modeling to predict pipe network pressure in time series. However, these models utilized shallow networks and exhibited limited ability to characterize complex phenomena (Bengio 2009).

Xu *et al.* (2015) proposed a NARX (Nonlinear Auto-Regressive with Exogenous Inputs) model for WSN real-time prediction and control. The model estimates the time-variable nodal demand equivalently by exploiting the real-time and historical operating data and establishes a functional relationship between the major variables in the network. The actual cases show that the model has good tracking and prediction performance. In essence, The NARX model is realized by a simple three-layer neural network, which increases the learning ability of time series data on the BP network.

In recent years, there have been some studies focusing on the application of deep learning in control systems (Punjani & Abbeel 2015). However, only Wu *et al.* (2015) attempted to apply the deep learning model on the modeling of an urban water supply system to simulate the water level of pools. The preliminary results showed that the Deep Belief Network (DBN) was better than the traditional ANN.

Currently, Long Short-Term Memory (LSTM) has been widely used in time series prediction. LSTM improves the recurrent neural network (RNN) and introduces long-term and short-term memory units. It can not only use current information, but also choose to use short-term and long-term historical information. It is more suitable for the modeling of such a complex nonlinear system in time series as a water supply network.

In this paper, LSTM is used to study pressure prediction in a water supply network. In addition, the structure of LSTM is modified for the characteristics of a water supply network. Specific contributions are as follows:

- A macroscopic mathematical model is established for the water supply network system with incomplete state observability. The deep learning model LSTM is applied to predict the state of an urban water supply network for the first time. In addition, the feasibility of using LSTM to predict the trends of pressure change is discussed based on a mathematical model.

- According to the characteristics of the water supply network, the structure of the LSTM network was improved and a parallel LSTM tandem DNN deep neural network model (PLDNN) is proposed. The model enhanced the performance to extract features of the control variables and state variables, and improved the prediction performance by combining the advantages of both LSTM and DNN. The model applicability in abnormal condition detection is also tested by pipe-burst experiments, which were generated by opening fire hydrants.

MATHEMATICAL MODEL OF WATER SUPPLY NETWORK AND LSTM NEURAL NETWORK

Mathematical model of water supply network

The water supply network system can be viewed as a nonlinear time-delay system with multiple inputs and outputs:

$$x(t+1) = f(x(t), u(t), d(t+1), \varepsilon(t+1)) \quad (1)$$

where x are the observable state variables, describing the pressure or flow at the monitoring point; u are control variables, describing the outlet pressure, outlet flow of pumping stations and the opening degree of telecontrolled valves; d is water demand or consumption at the nodes distributed in the network; ε denotes the white noise; t denotes the instantaneous values at sampling times; and f is a strongly nonlinear function of the hydraulic characteristics in the whole network.

According to (1), if the control vector and state vector of the system are known at the current time, as well as the nodal demand and disturbance at the next moment being well estimated, the states at the next moment can be determined. Information on x and u can be collected by the SCADA (Supervisory Control and Data Acquisition) system. However, the nodal water demand d is affected by many highly stochastic factors, which makes it difficult to estimate directly. Considering the entire network, there is a complex correlation among the water demand at each node, the node pressures, the water supply pressure and flow. According to Equation (1), the estimated value $\bar{d}(t)$ at

time t can be determined as:

$$\bar{d}(t) = g(x(t), x(t-1), u(t-1)) \quad (2)$$

According to the nonlinear mapping function g , if the input and state variables of the system at the previous moment, and state variables at the current moment are known, the water consumption at the current time can be calculated. Since the nodal water consumption is a sample in a time series, it has its trend and periodicity. The water demand at the next sampling time can be obtained by the autoregressive model according to the historical water consumption time series:

$$\bar{d}(t) = h(\bar{d}(t-1), \bar{d}(t-2), \dots, \bar{d}(t-n_d)) \quad (3)$$

where $h()$ represents the regression function, and n_d represents the length of the time series of historical water consumption. Equations (2)–(3) can be simplified to Equation (4) as:

$$\bar{d}(t) = h(g(x(t), x(t-1), \dots, x(t-n_d), u(t-1), \dots, u(t-n_d))) \quad (4)$$

For an actual water supply system, the state information can be partially known. A feasible method is to select long historical state information and control information to compensate for the information. According to the length of historical state information n_a and historical control information n_b , the mathematical model of the water supply network system is obtained as:

$$x(t+1) = f(x(t), \dots, x(t-n_a), u(t-1), \dots, u(t-n_b)) \quad (5)$$

where f is a complex nonlinear function, which is extremely hard to establish as an exact expression. As is known, a neural network has the ability of infinite approximation to any nonlinear function, and has good robustness and stability. Therefore, researchers have used neural networks to approximate the nonlinear function (Jamieson et al. 2007 and so on). However, Equation (5) shows that the current WSN state evolves from the historical operating conditions. The evolution of the WSN state is the result of external effect (control input) on the state and endogenous

evolution. Simple feedforward neural networks (such as BP), even the NARX model, lack the memory and processing ability for time-series data, that is, they need to explicitly take the time-delay variable as the input of the model to process time-series data, which may lead to an over-fitting problem.

LSTM neural network

Deep learning technology originates from the development of the neural network, and the recurrent neural network is one of the most popular models in deep learning. It has a very strong nonlinear mapping ability in analysis and prediction based on time series problems.

Theoretically, the RNN can accept historical information with arbitrary length. However, the length of historical information will be expanded into the corresponding number of layers in the training process, which is equivalent to a multi-layer feedforward neural network. The large number of layers will make the gradient vanish and cause other problems during training (Sutskever 2013). As a result, RNN can only accept a very limited amount of historical information. In order to solve the problem, Hochreiter & Schmidhuber (1997) proposed a new recursive neural network called LSTM.

The pressure prediction of a water supply network is a prediction problem based on a nonlinear system in a time series. The LSTM model, which is an advanced RNN model, can not only utilize the historical state information of pipe network pressure, but also accept historical information of arbitrary length. Therefore, LSTM is selected as the model for prediction analysis in this paper.

LSTM PREDICTION MODEL AND ITS APPLICABILITY IMPROVEMENT

LSTM prediction model

According to Equation (5), the inputs of the model include the historical state information of the monitoring points, as well as the control information on water supply pressure and flow. Thus, the water supply network pressure prediction model based on LSTM is shown in Figure 1, where

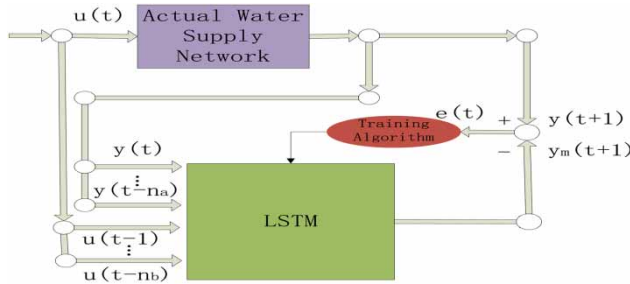


Figure 1 | LSTM prediction model of a water supply network.

$y(t - n_a), \dots, y(t)$ is the state information of each observation point from time $t - n_a$ to time t ; $u(t - 1), \dots, u(t - n_b)$ is the control information of inlet pressure and flow between time $t - 1$ and time $t - n_b$, which are the inputs to the deep learning model; and $y(t + 1)$ and $\bar{y}(t + 1)$ are the actual output of the pressure value at each monitoring point, and the predicted output given by the deep learning model at the time of $t + 1$, respectively. In the training process, mean square error between the predicted value \bar{y} and actual value y is generally used as the loss function, given as:

$$loss = \frac{1}{n} \sum_{i=1}^n (\bar{y}_i - y_i)^2 \tag{6}$$

where n is the number of neuron nodes in the output layer.

PLDNN prediction model

The state variables and control variables are two different types of feature information in a water supply network. If a single LSTM model is used to extract both types of features, the different effect of each type of feature on the model cannot be identified. As a result, a parallel LSTM model is used to extract and learn the characteristic information of each type of feature respectively. In addition, because DNN can map characteristic information to a higher-dimensional space, a parallel LSTM tandem DNN deep neural network model (PLDNN) is built. The model can process both pressure and water supply flow information. In this model, based on a certain historical event window, the control variables of water supply pressure and flow at the entry point, as well as the state variables of

the pressure data at measuring points, are used as input variables in two LSTM models, respectively. The outputs of these two LSTM models are then combined, and pressure at next moment is predicted by DNN as:

$$y_m(t + 1) = H(W_{DNN}[\bar{y}_1(t + 1)\bar{y}_2(t + 1)] + b_{DNN}) \tag{7}$$

where $y_m(t + 1)$ is the predicted output of the PLDNN model; $\bar{y}_1(t + 1), \bar{y}_2(t + 1)$ are the two LSTM model output values of the state variables and control variable input, respectively; $[]$ means to combine the two matrices with the same dimension on the time dimension; $H()$ is the activation function of the DNN model; and W_{DNN} and b_{DNN} are the weights and threshold value of the DNN model, respectively. LSTM and DNN are combined to utilize the advantages of each model and the overall architecture of the improved PLDNN model is shown in Figure 2.

In this paper, the dropout method (Srivastava et al. 2014) is used to prevent overfitting in PLDNN. Dropout is implemented at every layer of the model for better performance. The dropout discards a certain proportion of the nodes randomly in each iteration during the training process while restoring full connection during prediction. Without the dropout mechanism, there may be co-adaptation between certain hidden nodes, which will reduce the robustness of the feature-reuse of these nodes. Dropout will result in updating the network parameters randomly. The introduction of such randomness will increase the generalization ability of the model and prevent overfitting.

In this paper, the gradient descent method is used to optimize each parameter in the model. There are two methods to minimize the risk function. One is batch

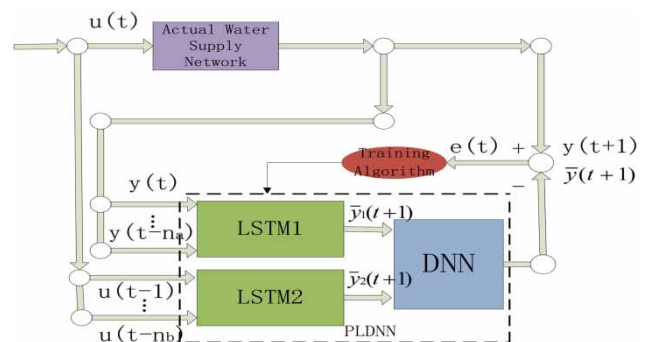


Figure 2 | PLDNN prediction model of a water supply network.

gradient descent, which uses all the data in the training set in each epoch and, thus, is computationally expensive. The other is the stochastic gradient descent method, which randomly samples from training data to calculate the loss function and update parameters. In order to overcome the shortcomings of these two methods, this paper uses an eclectic approach, the mini-batch gradient descent method, which divides the data into several batches and updates the parameters by calculating the loss function on each batch. Since a group of data in a batch is used to update the parameters, such a method reduces randomness and the number of calculations (Tsuruoka et al. 2009).

The activation function is used in the LSTM layer for input value calculation and the output layer after feature combination. Various types of functions can be used as activation function, such as tanh, softmax, linear, ReLU, etc. Equation (8) shows the non-saturated activation function ReLU (Nair & Hinton 2010):

$$y_i = \begin{cases} x_i, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (8)$$

The traditional saturated activation functions, such as sigmoid and tanh, will result in a vanishing gradient, while the non-saturated activation functions such as ReLU will not. Relative to the saturated activation function, a non-saturated activation function such as ReLU can accelerate the speed of convergence of the model. The deep learning model using ReLU can get similar or even better results without pre-training before supervised training.

INTEGRATION FRAMEWORK IMPLEMENTATION

Integrated framework of water supply network prediction model

The implementation framework of a monitoring point pressure prediction system in an urban water supply network is shown in Figure 3, where the arrow indicates the direction of information flow. The framework mainly collects pressure and flow data through the SCADA system, transforms the collected data into standard data by wavelet transform and normalization and saves it in the database. The

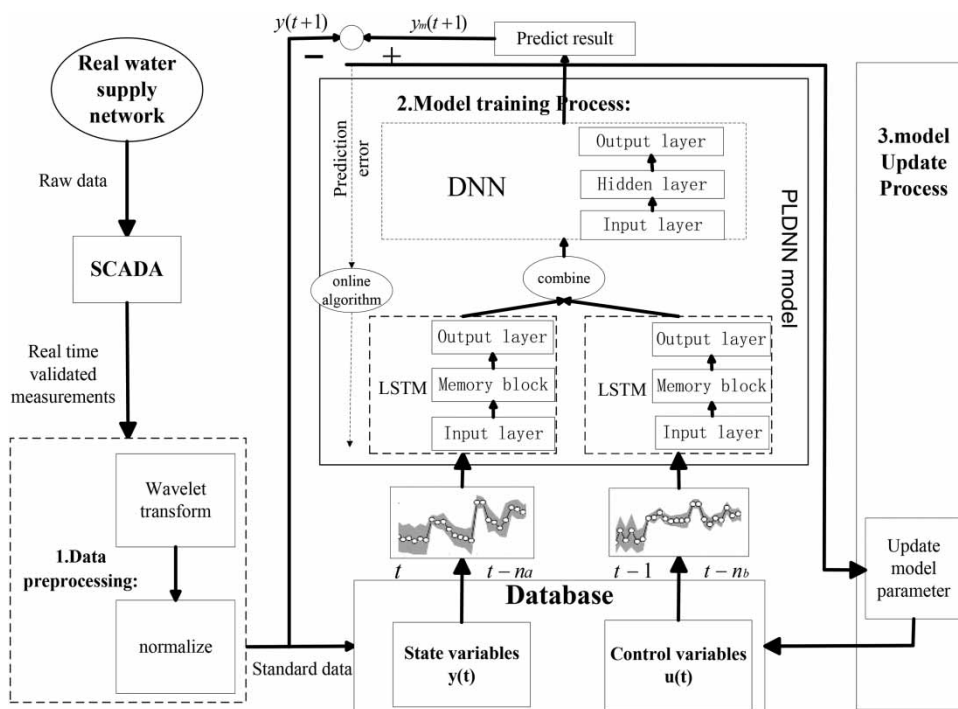


Figure 3 | Integration framework.

PLDNN extracts the training data from a database in the training process and uses the state variables and control variables as input variables for the two LSTM models respectively based on certain historical events. Then the outputs of the two LSTM are merged and the final output is given by DNN. The prediction of PLDNN is compared with the actual output of the water supply network. The parameters of the model are adjusted until the error meets the convergence condition or no longer decreases.

Model parameter adjustment

Because the internal structure of the neural network is complex, there is no relevant theory to provide a selection method. Usually a most suitable result is given by experience

and a large number of experiments. This paper uses the parameter adjustment process shown in Figure 4.

Step 1: The ranges of values of parameters are determined according to experience and the result of preliminary parameter adjustment. $N_u, N_s \in \{1, 2, \dots, 12\}$, which indicates that the maximum time span of historical information is 60 minutes. $Layers \in \{1, \dots, 5\}$ is the number of hidden layers. Since long historical information will result in redundant input and it has little effect on the accuracy of prediction, increasing the number of hidden layers can improve the ability of feature extraction and learning. However, a large number of layers will increase the complexity of the model. $Neurons \in [0, 300]$ is the number of neurons, which determines the degree of nonlinearity of the network; $dropout\ rate \in \{0.1, \dots, 0.5\}$ represents the node discard

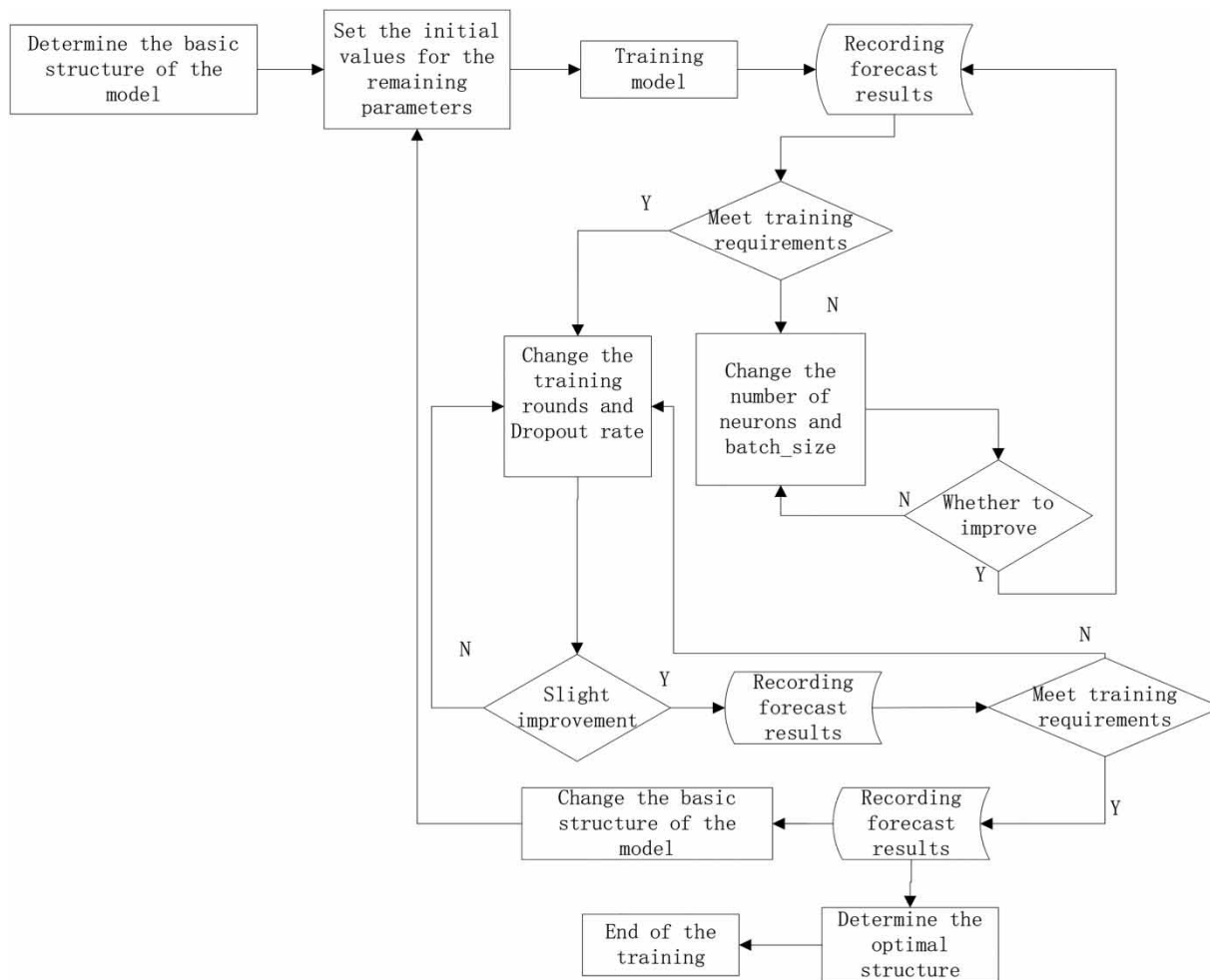


Figure 4 | Model parameter adjustment process.

ratio. If the ratio is too low, the effect will be limited; if the ratio is too high, the model will be underfitted; $epoch \in [20, 100]$ is the number of training. If the epoch is too small, the model will not fit well; if the epoch is too large, the time required for training will increase without any increase in the accuracy of prediction. $Mini\text{-}batch \in [1, 100]$ is the sample size of the small batch gradient descent method.

Step 2: Adjusting parameters by experiments and trial-and-error method, for each given $\{N_u, N_s, Layers\}$, the basic structure of the deep learning model can be obtained, and other parameters are adjusted according to the basic structure.

CASE STUDY

In this paper, the water supply network in SX City YC District is taken as an example for analysis. The network is shown in Figure 5. The area of the network is about 106.7 square kilometres, and the daily water supply

volume is about 150,000 cubic metres. YMB, YMN, TYMY and PSDD are the four main water inlet ports in the area. The 17 black triangles in the figure represent the positions of key pressure monitoring points.

The LSTM neural network for the water supply network in SX city takes the historical control information of water supply pressure and flow of four inlet ports, and the historical and current state information of 17 pressure measuring points as inputs; the pressures at the next moment of 17 monitoring points are taken as outputs. The data set is sampled from June 1, 2016, to June 28, 2016, and the sampling interval is five minutes. Data from June 1 to June 26 are used for training, while data from June 27 to June 28 are used for testing.

Data preprocessing

Because the data collected from the SCADA system have certain problems such as loss of data and noise, linear

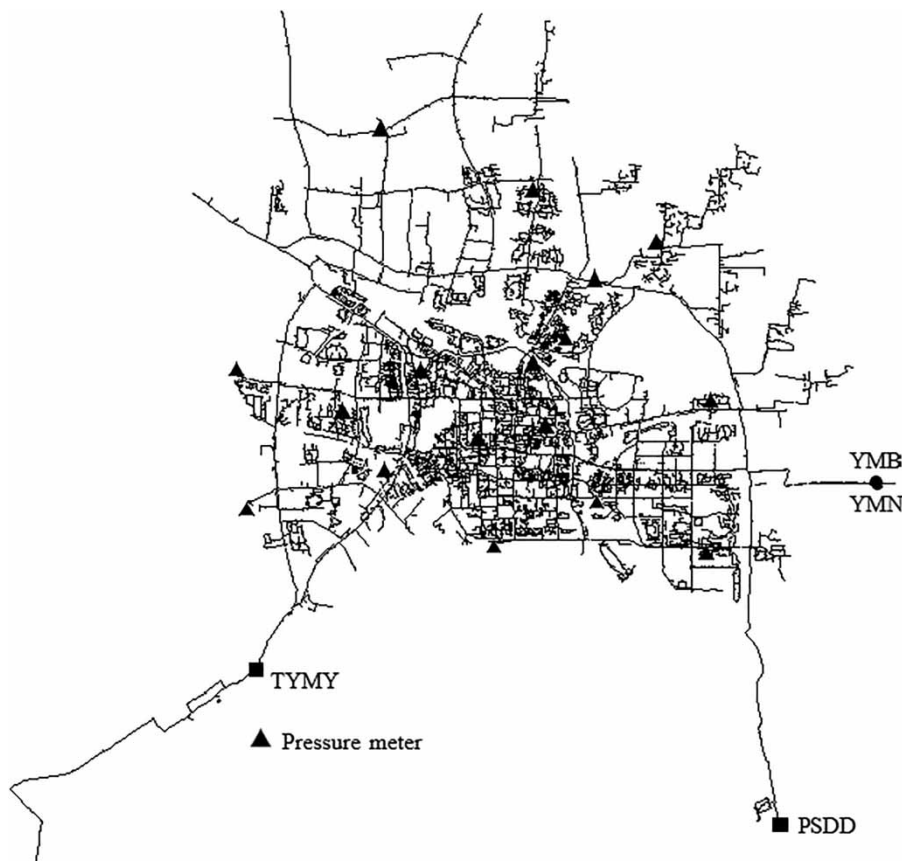


Figure 5 | SX City YC District water supply network.

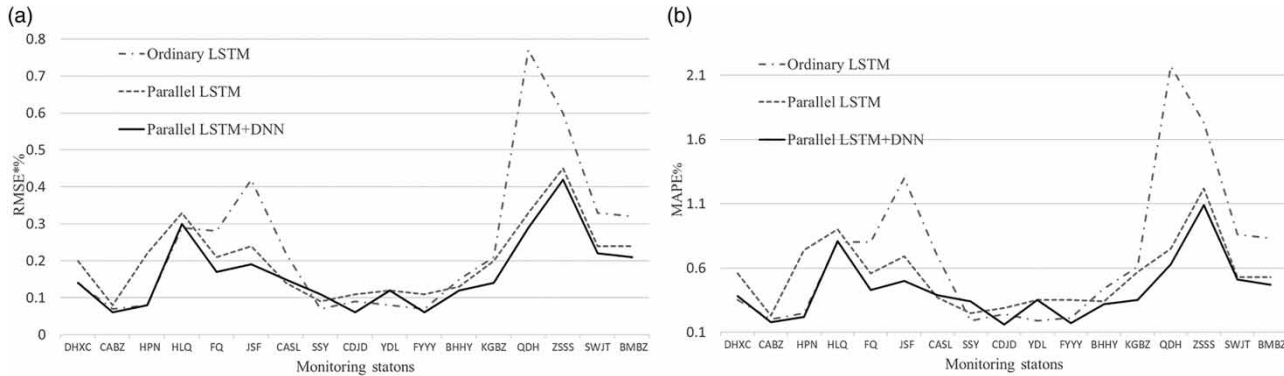


Figure 6 | MAPE and RMSE error performance index of each monitoring point: (a) MAPE, (b) RMSE*100.

interpolation is used to complete the missing data and the noise data are filtered out by wavelet transform.

The point pressure within a day exhibits periodic fluctuation and shows a clear period and trend, including early peaks, late peaks and off-peak period. However, for a specific moment, the magnitude of the change is not the same. The short-term pressure prediction of the water supply network is to predict the future pressure for a certain period of time (several minutes or hours).

As the pressure and flow data in a water supply network have different physical meanings and orders of magnitude, in order to alleviate the difficulty of network training, the data are pre-processed by normalization, which means the inputs and outputs of the network are limited to [0,1].

Determination of network parameters

According to the method of adjusting parameters proposed in the section above on ‘Model parameter adjustment’, the values of parameters are set as the following: $N_u = N_y = 6$; $Layers = 1$ and $Neurons = 96$ for LSTM; $Layers = 1$ and $Neurons = 96$ for DNN; $Mini-batch\ size = 32$, $epoch = 50$ for the training process, $dropout\ rate = 0.2$.

Controlled experiment and analysis

In order to analyze and compare the prediction performance of the proposed model, traditional LSTM and parallel LSTM are used for prediction as well. MAPE (mean absolute percentage error) and RMSE (root mean

square error) are used as the indexes to evaluate performance. The results are shown in Figure 6.

Figure 6 shows that the ability of prediction of each LSTM model is different at different monitoring points, but the accuracies of prediction results for each monitoring point meet the requirements. The prediction results of monitoring points close to the water source, such as CABZ, FYYY and SSY, are better than those close to the end of the network, such as HLQ and ZSSS. Since there are fewer nodes for the monitoring points upstream, the change of the flow rate is small and the pressure is mainly affected by the water supply pressure at the water plant, so, the pressures at these points are predictable. At the end of the pipe network, due to the large number of nodes downstream and different characteristics of variation in water consumption at each node, the hydraulic conditions are complicated and the pressure change is difficult to predict. Then, the errors of predictions at these points are larger. Based on the PLDNN, the deep learning model can extract and learn the data of state variables and control variables in the water supply network system separately and keep the characteristic information of the original data as much as possible. As a result, it can simulate complex non-linear phenomena and has higher prediction accuracy than traditional LSTM.

Table 1 | Comparison between PLDNN and traditional prediction methods

Prediction method	BP	SVM	VAR	NARX	PLDNN
RMSE*100	0.56	0.43	0.29	0.28	0.17

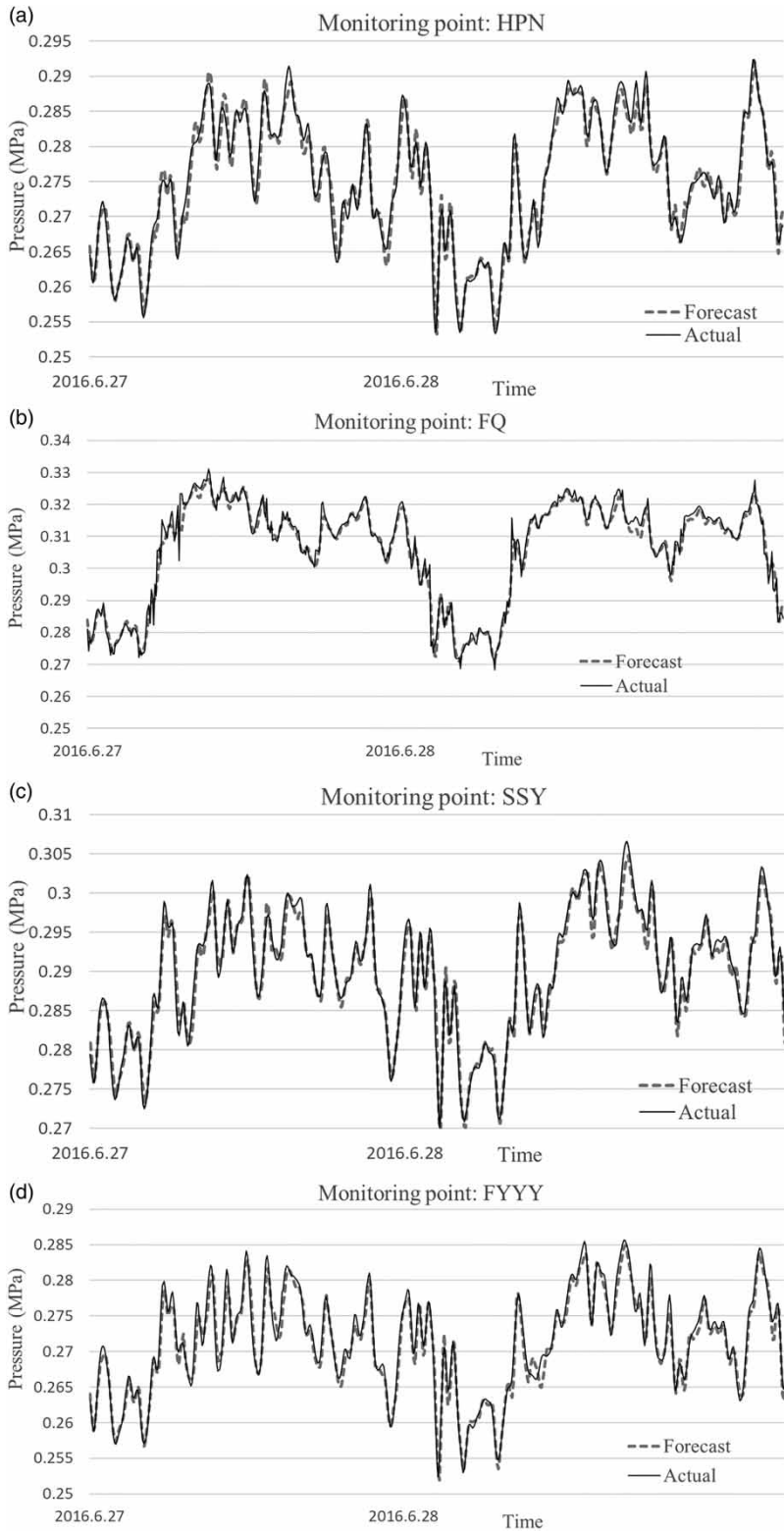


Figure 7 | Pressure prediction effect of four monitoring points.

In order to compare with the traditional prediction model, BP, SVM, VAR, NARX and other models were implemented respectively, and the results are compared and shown in Table 1.

The prediction results of the PLDNN model are significantly better than the results of the traditional BP and SVM models. Since VAR and NARX are suitable for a class of problems such as time series prediction, their prediction performance is often better than those of other shallow learning models, but still inferior to that of the deep learning model. Figure 7 shows the predictive effect of the PLDNN model for some monitoring points.

Abnormal condition experiment

Pipe bursts occur frequently in the urban water supply network. Because the PLDNN prediction model predicts trends in principle and the pressure change has its own trends and periodicity under normal conditions, it is hard to track the trend pressure under abnormal conditions, where the trend of the pipe network is affected by the change of hydraulic operation conditions. However, the model can in turn be used to test whether there are abnormalities in the water supply network. If the pipeline network is under normal conditions, the predicted value of PLDNN is basically consistent with the measured value of SCADA. If there is an anomaly in the pipe network and such an anomaly has an impact on the pressure at the monitoring point, there will be a big error between the prediction value and the measured value. If the error exceeds the normal range, then it can be determined that there is an anomaly near the monitoring point at this moment.

Pipe-burst experiments are often generated by opening fire hydrants. SX Water Company conducted five pipe-burst experiments in YC District on April 3, 2015. At that time, the data acquisition was strengthened, and the sampling interval was one minute. We used the data from March 30 to April 2, 2015, to train the PLDNN model, and predict the pressure on April 3. With the help of the PLDNN prediction value, five abnormal pipe-burst events can be successfully detected by the TBA method (Romano 2012). For other traditional prediction methods, there are false alarms and missed alarms, as shown in Table 2.

Figure 8 shows the pressure signal of the DHXC monitoring point, which is the nearest to the burst point, in the second pipe-burst experiment. Before 9:35 AM, in normal working conditions, the predicted value is close to the actual value. In the abnormal period of 9:35 AM–10:07 AM, there is a big deviation between the predicted value and the actual value (the deviation $>3\delta$). After 10:07 AM, with the end of the second pipe-burst experiment, the predicted value is soon very close to the actual value. Therefore, according to the PLDNN, the dynamic behavior under normal conditions can be well tracked, and abnormal hydraulic conditions (pipe burst or fire-hydrant use) can be judged.

CONCLUSION

In this paper, the deep learning model LSTM is used for pressure prediction of a water supply network, and the LSTM model is adaptively improved according to the characteristics of the water supply network. The structure of parallel LSTM tandem DNN is proposed to take into

Table 2 | Detection effect of five prediction methods in pipe-burst experiments

Pipe-burst experiment	BP	SVM	VAR	NARX	PLDNN
Burst1 (3 April 2015 09:16)	4 mins	–	9 mins	6 mins	5 mins
Burst2 (3 April 2015 09:35)	6 mins	5 mins	10 mins	5 mins	5 mins
Burst3 (3 April 2015 10:12)	–	–	5 mins	–	5 mins
Burst4 (3 April 2015 10:33)	–	–	–	–	6 mins
Burst5 (3 April 2015 10:54)	12 mins	–	–	–	4 mins
False alarm on that day	24 times	5 times	12 times	11 times	2 times

Gray-shaded events were detected. The number in gray is the detection process time (detection time minus burst time).

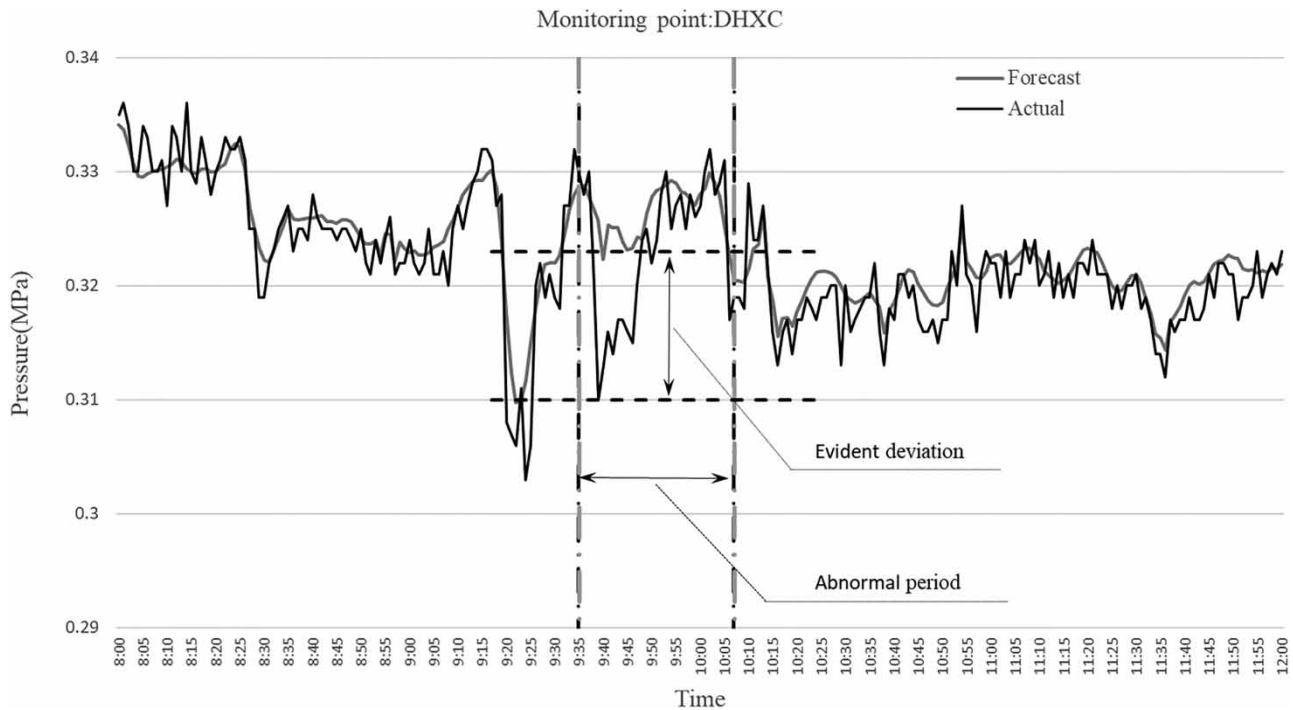


Figure 8 | Pressure prediction effect of DHXC monitoring point.

account the two different types of input information on water supply pressure/flow and monitoring point pressure. Dropout and other deep learning methods are used to improve the prediction performance of the model.

The experiment results show that the LSTM model can overcome the shortcomings of the traditional artificial neural network in predicting complex situations and that of the support vector machine in predicting only a single measuring point. The PLDNN model can simulate highly nonlinear conditions and realize prediction at multiple measuring points. The PLDNN model extracts and learns the features of different types of information separately, which can improve the learning and generalization ability of the neural network and result in higher prediction accuracy than the traditional LSTM model. The dropout method can prevent overfitting in prediction models, which improves the generalization ability of the models. The performance is even better when implementing dropout at each layer. The PLDNN model can well track the normality of the operation state in a water supply network, and detect abnormal events in the network, but cannot track under abnormal conditions.

ACKNOWLEDGEMENTS

This work was funded by the National Natural Science Foundation of China (U1509205).

REFERENCES

- Bengio, Y. 2009 [Learning deep architectures for AI](#). *Foundations and Trends in Machine Learning* **2** (1), 1–127.
- Creaco, E., Campisano, A., Fontana, N., Marini, G., Page P, R. & Walski, T. 2019 [Real time control of water distribution networks: a state-of-the-art review](#). *Water Research* **161**, 517–530.
- Hochreiter, S. & Schmidhuber, J. 1997 [Long short-term memory](#). *Neural Computation* **9** (8), 1735–1780.
- Jamieson, D. G., Shamir, U., Martinez, F. & Franchini, M. 2007 [Conceptual design of a generic, real-time, near-optimal control system for water-distribution networks](#). *Journal of Hydroinformatics* **9** (1), 3–14.
- Luvizotto, E., Cavichia, M. C., Vatauvuk, P. & Andrade, J. G. P. 2012 [Nonmatrix gradient method for the simulation of water distribution networks](#). *Journal of Water Resources Planning & Management* **139** (4), 433–439.
- Lv, M., Zhang, S. & Li, H. 2001 [Dynamic combination prediction method of pressure measuring points in water distribution](#)

- network. *Chinese Theory and Practice of System Engineering* **21** (3), 139–144.
- Mehrpourvar, M. & Asghari, K. 2018 [Modular optimized data assimilation and support vector machine for hydrologic modeling](#). *Journal of Hydroinformatics* **20** (3), 728–738.
- Nair, V. & Hinton, G. E. 2010 Rectified linear units improve restricted Boltzmann machines. In: *Proceedings of the 27th International Conference on Machine Learning (ICML-10)* (J. Fürnkranz & T. Joachims, eds), Omnipress, Madison, WI, USA, pp. 807–814.
- Perea, R. G., Poyato, E. C., Montesinos, P. & Díaz, J. A. R. 2019 [Optimisation of water demand forecasting by artificial intelligence with short data sets](#). *Biosystems Engineering* **177**, 59–66.
- Ping, J., Wang, R., Sun, J. & Xiao, C. 2014 [Pressure prediction of a water distribution network based on SVM](#). In: *ICPTT 2014: Creating Infrastructure for a Sustainable World* (B. Ma, M. Najafi & H. Tang, eds), ASCE, Reston, VA, USA, pp. 155–168.
- Punjani, A. & Abbeel, P. 2015 [Deep learning helicopter dynamics models](#). In: *2015 IEEE International Conference on Robotics and Automation*, IEEE, pp. 3223–3230.
- Rao, Z. & Alvarruiz, F. 2007 [Use of an artificial neural network to capture the domain knowledge of a conventional hydraulic simulation model](#). *Journal of Hydroinformatics* **9** (1), 15–24.
- Rao, Z. & Salomons, E. 2007 [Development of a real-time, near-optimal control process for water-distribution networks](#). *Journal of Hydroinformatics* **9** (1), 25–37.
- Romano, M. 2012 *Near Real-Time Detection and Approximate Location of Pipe Bursts and Other Events in Water Distribution Systems*. Doctoral thesis, University of Exeter, Exeter, UK.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. 2014 Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research* **15** (1), 1929–1958.
- Sutskever, I. 2013 *Training Recurrent Neural Networks*. Doctoral thesis, University of Toronto, Toronto, Canada.
- Tsuruoka, Y., Tsujii, J. & Ananiadou, S. 2009 Stochastic gradient descent training for L1-regularized log-linear models with cumulative penalty. In: *Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP*, ACL, Stroudsburg, PA, USA, pp. 477–485.
- Wu, Y. & Liu, S. 2017 [A review of data-driven approaches for burst detection in water distribution systems](#). *Urban Water Journal* **14** (9), 972–983.
- Wu, Z. Y., El-Maghraby, M. & Pathak, S. 2015 [Applications of deep learning for smart water networks](#). *Procedia Engineering* **119**, 479–485.
- Xu, Z., Yang, J., Cai, H., Kong, Y. & He, B. 2015 [Water distribution network modeling based on NARX](#). *IFAC-PapersOnLine* **48** (11), 72–77.
- Yang, J., Xu, Z. & Kong, Y. 2014 [Chaos identification and prediction of pressure time series in water supply network](#). In: *Proceedings of the 33rd Chinese Control Conference*, IEEE, pp. 6533–6538.
- Yu, T., Zhang, T. & Lu, M. 2005 State estimation model of water distribution network based on SVM. *Journal of Harbin Institute of Technology* **37** (9), 1205–1208.
- Zhang, Z., Jiang, W. & Wang, Z. 2011 To base on MATLAB neural network's prediction of the urban supply water network's water pressure. *Journal of Heilongjiang Bayi Agricultural University* **2** (6), 13–15.

First received 1 August 2019; accepted in revised form 24 January 2020. Available online 7 February 2020