WILEY | Hindawi

Research Article

Complexity to Forecast Flood: Problem Definition and Spatiotemporal Attention LSTM Solution

Yirui Wu [ID],[1] Yukai Ding,[1] Yuelong Zhu,[1] Jun Feng [ID],[1] and Sifeng Wang [ID][2]

[1]*College of Computer and Information, Hohai University, Nanjing, China*
[2]*School of Information Science and Engineering, Qufu Normal University, Rizhao, China*

Correspondence should be addressed to Sifeng Wang; sfwang@qfnu.edu.cn

With significant development of sensors and Internet of things, researchers nowadays can easily know what happens in physical space by acquiring time-varying values of various factors. Essentially, growing data category and size greatly contribute to solve problems happened in physical space. In this paper, we aim to solve a complex problem that affects both cities and villages, i.e., flood. To reduce impacts induced by floods, hydrological factors acquired from physical space and data-driven models in cyber space have been adopted to accurately forecast floods. Considering the significance of modeling attention capability among hydrology factors, we believe extraction of discriminative hydrology factors not only reflect natural rules in physical space, but also optimally model iterations of factors to forecast run-off values in cyber space. Therefore, we propose a novel data-driven model named as STA-LSTM by integrating Long Short-Term Memory (LSTM) structure and spatiotemporal attention module, which is capable of forecasting floods for small- and medium-sized rivers. The proposed spatiotemporal attention module firstly explores spatial relationship between input hydrological factors from different locations and run-off outputs, which assigns time-varying weights to various factors. Afterwards, the proposed attention module allocates temporal-dependent weights to hidden output of each LSTM cell, which describes significance of state output for final forecasting results. Taking Lech and Changhua river basins as cases of physical space, several groups of comparative experiments show that STA-LSTM is capable to optimize complexity of mathematically modeling floods in cyber space.

1. Introduction

As more sensors are applied to acquire variant data from physical space, researchers try to build a corresponding cyber space to describe inherent mathematical relationship between sensor acquired factors and results, which provides users a great deal of convenience to find novel solutions for problems in the real world. However, mathematical and technical complexity and challenge rise in both procedures, i.e., transforming problem-related data from physical space to cyber space, and utilizing models to solve problems in cyber space. Inspired by data-driven and artificial intelligent idea to solve problem in physical space, we intend to smartly solve the problem of flood forecasting by means of complexity modeling and optimization.

Flood often occurs with sudden and devastating nature, causing huge life and economic losses to human society.

Therefore, it is of significance to forecast flood disasters in advance. To minimize impacts brought by floods, researchers have proposed quantity of methods to accurate forecasting in the past decade [1, 2]. Based on core ideas to forecast, we divide their proposed methods into two categories: physical model [3, 4] and data-driven model [5, 6]. Physical model explains hydrological procedures with conceptual math equations, such as rain, evaporation, and flow concentration. Afterwards, a highly nonlinear function system is constructed to model complex flood process from hydrological clues to result of large run-off values. We can find carefully designed physical models in works of Fan et al. [7] and Pontes et al. [8, 9], where their models well fit in special areas to handle complexity of flood forecasting.

Data-driven model directly models mathematical interactions between different hydrological factors and run-off

values based on historical observations. In other words, data-driven models learn mapping between flooding cues and flow rates without considering detailed physical processes, which is the main difference between physical models and data-driven models [10]. Due to rapid development of machine learning technology, many novel data-driven forecasting methods have been proposed and practiced, including Bayesian network [1], SVM model [11], neural network [6], and their variations and integrations.

It is noted that both physical and data-driven models are sensitive to their internal parameters, which requires both quantity of convinced data and great deal of manual efforts from researchers to adjust. In other words, the main difficulty to apply models on small or medium river basins lies in the fact of insufficient data to support accurate forecasting. Moreover, small or medium river basins are generally short of special research in most developing countries, which leads to difficulties in designing appropriate physical models for forecasting. Based on all these discussions, we aim to optimize complexity to forecast flood in small or medium river basins by methods of data-driven models.

Recently, deep learning structures gained lots of attentions by their significant classification and regression results on visual tasks of text detection, object categorization, language translation, and so on. Following the big progress of deep learning technology on accuracy, we concentrate on LSTM network for the goal of more accurate flood forecasting results. Essentially, LSTM is developed on the basis of Recurrent Neural Network (RNN), which could handle long-time sequential data with special designs of gates. Based on its high potential property to forecast time-varying variables, we apply LSTM to model inherent and complex relationship between hydrological factors and run-off values.

Attention mechanism is widely used in prediction tasks, due to its novelty to borrow idea from human visual attention, namely, that humans purposely view parts of environment or picture with the context information or advanced semantic knowledge. Inspired by core ideas of attention mechanism, we study flood formation of small and medium river basins by firstly gathering related hydrological data of different locations and timings. Afterwards, a novel LSTM structure embedding a spatiotemporal attention module, named as STA-LSTM, is constructed to dynamically select hydrological features for accurate forecasting.

STA-LSTM takes advantages of original LSTM structure, which is capable to handle long-time time-varying data. Meanwhile, the embedded spatiotemporal attention module dynamically assigns spatial-wise weights for input hydrological factors acquired from different locations at first. After that, it allocates temporal-wise weights to hidden output of each LSTM cell, which is comprehended as context information in Liu et al. [12]. Such spatial- and temporal-wise weights result in capability to dynamically characterize significance of hydrological factors obtained at any timings and locations. According to case study experiments in Europe Lech and China Changhua river basins, STA-LSTM could realize accurate flood forecasting by constructing contexture-based weighting schemes.

We conclude two contributions of this paper as follows:

(i) Facing complexity of modeling cyber-physical interaction, a novel LSTM model embedding spatial-temporal attention module is proposed, which is capable of accurately predicting run-off values in cyber space based on hydrological data acquired from physical space.

(ii) We design a novel temporal attention module, which is built on contextual information to compute weights for each LSTM cell output. Incorporating with both spatial and temporal context information, the proposed attention module helps describe how hydrology factors interact to form flood in physical space and appropriately builds such process in cyber space by constructing weighting schemes in STA-LSTM.

2. Related Work

In this section, we introduce related methods with two categories, i.e., data-driven model for flood forecasting and introduction to attention model.

2.1. Data-Driven Model for Flood Forecasting. Early, Juliang et al. [13] propose an accelerated genetic algorithm (AGA). Their method utilizes a Back Propagation Neural Network (BPNN) to optimize initial parameters, which brings advantages of better and faster convergence performance. Inspired by development of support vector machine (SVM), Yu et al. [11] compare performance on flood forecasting between artificial neural network (ANN) and SVM. After performing a number of comparison experiments, they draw a conclusion that SVM is slightly better than ANN in forecasting floods. Later, Minghua et al. [14] emphatically compare experimental results achieved by Xinanjiang model (a famous physical model) and ANN model. Afterwards, they conclude that ANN could reflect time-varying characteristics of hydrological process, which is an advantage by comparing with abstract representation of hydrology process in Xinanjiang model.

After analyzing various flood forecasting models, Cheng et al. [15] utilize quantum particle swarm optimization method to solve complexity of defining parameters for ANN, which is later examined by experiments to forecast daily run-off values of reservoirs. Lima et al. [16] conduct flood frequency analysis with a hierarchical Bayesian framework, which estimates Generalized Extreme Value (GEV) distribution parameters in a local sense for explicitly modeling and uncertainties reduce. Recently, Wang et al. [10] proposed a Bayesian-based method, which establishes a posterior distribution for daily flow rate forecasts and uncertainty quantifications.

With the idea of coupling the strength of physical model and data-driven model, O'Connel et al. [17] use paleo hydrologic information constraints to effectively reduce the uncertainty during flood frequency analysis. Following such idea, Biondi et al. [18] firstly simulate the hydrologic response by a rainfall-run-off model named as Infiltration and Saturation Excess (RISE) and then utilize the extracted

hydrological information for later deterministic Bayesian Forecasting. Recently, Wu et al. [5] successfully transformed hydrological process described by Xinanjiang model into entities and connections of Bayesian network, which offers a solution to integrate expert knowledge in a data-driven model. In order to offer a task-specified computing service, data-driven models nowadays have been developed accompany with Internet of things [19, 20], cloud-edge computing [21–23], big data [24, 25], and other technologies [26–28].

2.2. Introduction to Attention Model. Core idea behind attention model is to select informative and significant information for task goal, which coincides with principle of human selective visual mechanism. Existing attention models for deep learning can be divided into two groups: hard and soft attention. Hard attention can be comprehended as spatial selection for salient regions, which leads the input areas to be processed as different parts with values of 0 (ignore areas) or 1 (concentrate areas). Meanwhile, soft attention assigns flexible weight values between 0 and 1 to parts of input data.

Mnih et al. [29] introduce general idea of hard attention by optimally selecting salient regions from input images based on pre-defined selection rules. Their proposed method performs recognition tasks on selected salient regions with a novel RNN structure. Following idea of hard attention, He et al. [30] propose a convolutional neural network, named as Text-CNN, to involve attention scheme for scene text detection. Specifically, their scheme not only extracts salient regions as informative parts of input images, but also particularly selects informative features from feature pools for more accurate detection.

Soft attention is flexible and efficient to be an additional and functional part for deep learning networks. For example, Song et al. [31] propose spatial attention module to accurately and robustly recognize human actions. Their proposed method firstly constructs a spatial-wise weight scheme to pay attention on informative joints in each RGB-D skeleton frame and then assigns spatial attention weights to guide the construction of feature map of the corresponding frame. To utilize global attention information for higher accuracy and robustness, a globally context-aware attention LSTM [12] is built, which successfully constructs and optimizes global attention information for each RGB-D human action sequence from dataset [32].

Most recently, Yeung et al. [33] utilize soft attention model to assign frame-wise weights for frames captured by a sliding window, which helps fuse multiframe information for recognition tasks. Chen et al. [34] build an attention model on the basis of a novel network architecture combining advantages of CNN and RNN, which successfully extracts informative and modality-specific feature for human activity recognition. They claim that their extracted feature is able to represent high-level visual information, even training with an imbalanced and limited size dataset. Anderson et al. [35] construct a bottom-up and top-down attention model on top of Faster R-CNN, which is capable of

assigning weights in object or salient image region level. After conducting experiments on several public datasets, they have achieved state-of-the-art performance on image caption task. Inspired by above attention models, we designed the proposed spatial-temporal attention model to allocate attention weights for temporal and spatial dimensions.

3. The Proposed Method

We firstly introduce the experimental small river basins, i.e., Lech and Changhua river basin. Then, we introduce overall network structure of the proposed STA-LSTM model. Finally, a novel spatial-temporal attention module is proposed to show how context information is extracted.

3.1. Introduction to Experimental River Basins. We take two river basins, i.e., Lech and Changhua, as experimental areas, due to their small and complex nature for flood formation. Due to significant development of remote sensing and sensor technologies, we build our prediction model on data collected from remote sensing imageries and sensors. Specifically, data about Lech river basin are achieved from European Centre for medium-range Weather Forecasts (ECWMF), which is free to download worldwide weather and hydrological information. Meanwhile, we get hydrological data about Changhua river basin from cooperation China government.

Figure 1(a) refers to the map of Lech river basin, where we suppose latitude and longitude range for Lech river basin is from (10.68E, 47.65N) to (10.94E, 48.73N) with 0.01×0.01 radius precision. Originating from northwest slope of Lysitar mountain, Lech river finally flows into Danube river at 40 km north of Augsburg. Total length, basin area, and estuary average annual flow of Lech river are 263 km, 4126 km^2, and 120 m^3/s, respectively. Weather in Lech river basin areas is warm and humid throughout the year.

Figure 1(b) shows the map of Changhua river basin, where we can find that Changhua river originates from Jixi County and finally flows into Xinanjiang river. Total length, basin area, and estuary average annual flow of Changhua river are 96 km, 905 km^2, and 146.651 m^3/s, respectively. Daily run-off value of Changhua river could vary from 0.58 m^3/s to 2100 m^3/s.

Our goal for both river basins is to realize forecasting of surface run-off at their converge locations (represented as red circles at Figure 1) through the proposed STA-LSTM model. Specifically, the proposed model adopts multiple hydrological factors as inputs, including precipitation, evaporation, soil tension water, temperature, and wind.

3.2. Network Architecture Design. We firstly offer a brief introduction to mathematical theory of LSTM cell. Then, we explain how attention scheme improves accuracy of flood forecasting. Afterwards, we design a novel LSTM network architecture involving context information to complete task of flood forecasting. At last, we offer pseudocode of STA-LSTM for readers' convenience.

(a)   (b)
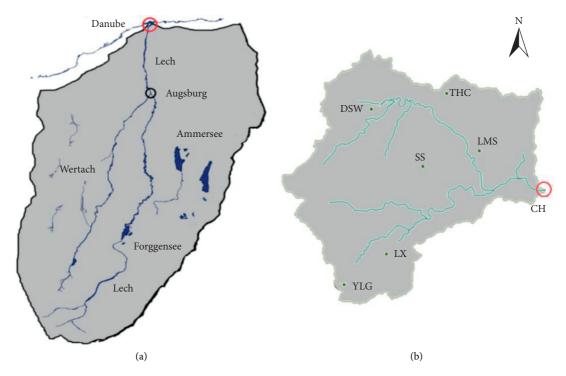
FIGURE 1: Illustration of Lech (a) and Changhua (b) river basins, where we need to predict flow rate values of converge locations labeled by red circles.

3.2.1. Mathematical Theory of LSTM Cell.

Due to difficulty in maintaining long-distance dependency information, LSTM modifies RNN by designing gate structure to keep long-term state. Typical structure of a LSTM cell is represented in Figure 2, where we can observe input gate i, output gate o, input modulation gate g, forget gate f, and memory cell c. Each LSTM cell is responsible to update its hidden output representation h at each state t with the following function:

$$h_t = \sigma\left(W\binom{x_t}{h_{t-1}}\right), \tag{1}$$

where x represents input signal and function $\sigma()$ refers to operation of Sigmoid.

LSTM introduces a long-term memory structure c to maintain long-term information for each cell. Furthermore, it decides whether to forget information inside memory based on the following equations:

$$\begin{pmatrix} i \\ f \\ o \\ g \end{pmatrix} = \begin{pmatrix} sigm \\ sigm \\ sigm \\ sigm \end{pmatrix}\left(W\binom{x_t}{h_{t-1}}\right), \tag{2}$$

$$c_t = f \odot c_{t-1} + i \odot g, \tag{3}$$

$$h_t = o \odot \operatorname{Tan} h(c_t), \tag{4}$$

where $\odot$ refers to element-wise multiplication. From equation (3), we can notice that internal memory cell c_t would be updated, if forgetting gate f is activated. After
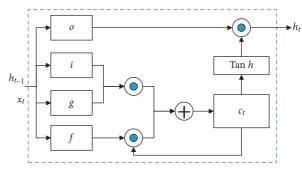

FIGURE 2: Structure of a typical LSTM unit.

activating f, c_t will be assigned with signal controlled by input gate i and input modulation gate g. Afterwards, LSTM cell will update hidden output h_t on the basis of output gate o and current memory cell c_t, which is described in equation (4). With above designs of memory cell and different gates, output of LSTM can be associated with previous input signals to memorize long-time sequential information [36].

3.2.2. Function of Attention Scheme in Flood Forecasting.

After years of research on applying data-driven models to forecast flood, we find adopting all hydrological data for forecasting could not help achieve satisfactory forecasting results, since some hydrological features are useless or even independent with run-off predictions. For example, soil tension water is an important factor for initial state of floods in humid areas. When value of soil tension water increases and exceeds maximum amount that soil can maintain during raining, it would have no impact on variation of run-off

values. Furthermore, soil tension water does not affect river flow values in dry locations with sandy soil, due to bad capability to maintain water of sandy soil. All these facts can be found in hydrological simulation studies or physical models [37].

Due to high spatial and temporal variation of hydrological processes, it is highly recommended to collect hydrological factors by a dense network of hydrometeorological stations. Built on the basis of sufficient stations to collect hydrological features, modeling informativeness degrees of hydrological features would contribute to accurate forecasting. In other words, selectively utilizing informative factors acquired at significant timings and locations is the key idea to adopt attention scheme in flood forecasting. With the ability of focusing on key features and ignoring irrelevant features, data-driven models can appropriately integrate different factors to fit in process of floods, instead of expert knowledge used in physical models. Besides, irrelevant factors could bring in noise to decrease forecasting accuracy.

Based on above discussions, we thus establish a dynamic feature selection mechanism, i.e., attention mechanism, to describe informativeness degrees of flood factors, so that different combinations or weights can be applied on input hydrological features based on context information, i.e., inherent characteristics of river basin for flood formation. It is noted that there exists a trend in deep learning domain that researchers should design all functions by one single network, which brings advantages of less computation and high optimization efficiency. Following such trend, we aim to design a novel LSTM network, i.e., STA-LSTM, to complete task of choosing variables by attention module, which works with the same function of Principal Component Analysis (PCA) indeed.

3.2.3. Structure of STA-LSTM.
Network structure of the proposed STA-LSTM is shown in Figure 3, where the proposed attention module allocates dynamic weights to both input and output of STA-LSTM cells for usage of selection on informative features. After building attention module, hidden outputs of all STA-LSTM cells would be concatenated to form F for prediction on increase or decrease in run-off values. It is noted that we prefer to predict based on all hidden outputs, since LSTM structure is restricted in preserving global contextual information by designing forgetting gate. Considering flood forecasting as a global regression problem, we thus prefer to perform forecasting on all hidden outputs, rather than hidden output for final state.

As shown in Figure 3, we firstly acquire hydrological raw data from a small or median river basin to construct input dataset $X = \{x^i \mid i = 1, 2, \ldots, n\}$, where i and n refer to index and total number of samples. There exists n flood records in input dataset X. Afterwards, raw sample x^i is normalized to construct corresponding hydrological feature set $\widehat{x^i}$:

$$\widehat{x^i} = f_{\text{norm}}(x^i) = \{I_t \mid t = 1, 2, \ldots, \tau\}, \text{ where } I_t \in R^{H \times W},$$
(5)

where function $f_{\text{norm}}()$ represents normalization function, τ represents total number of states over the whole network, and $H \times W$ refers to the size of input feature for each state and is formed by variant hydrological factors.

At state t which refers to the part labeled by blue rectangle in Figure 3, the corresponding tth LSTM cell would compute hidden output of current state h_t with

$$h_t = f_{\text{lstm},t}(\widehat{I}_t, \widehat{h}_{t-1}),$$
(6)

where function $f_{\text{lstm},t}()$ represents processing of the tth LSTM cell to maintain long-term information, $\widehat{I}_t$ represents weighted input feature computed by spatial attention module S, and $\widehat{h}_{t-1}$ refers to weighted former hidden computed by temporal attention module T. It is noted that number of LSTM cells is the same with total number of states τ.

Specifically, weighted input feature $\widehat{I}_t$ in equation (6) is processed by the proposed spatial attention module S with

$$\widehat{I}_t = \alpha_t \otimes I_t = \varphi_{t-1}^S(\widehat{h}_{t-1}) \otimes I_t,$$
(7)

where α_t refers to spatial attention weight for state t, $\otimes$ denotes element-wise operation, and function φ_{t-1}^S means operations inside $t - 1$th spatial attention module to compute α_t. It is noted that number of either spatial or temporal attention module is $\tau - 1$.

Meanwhile, hidden output $\widehat{h}_t$ is processed by the proposed temporal attention module T with

$$\widehat{h}_t = \beta_t \odot h_t = \varphi_{t-1}^T(I_{t-1}, I_t) \odot h_t,$$
(8)

where I_{t-1} and I_t are original input features for state $t - 1$ and t, respectively, β_t refers to spatial attention weight for state $t - 1$, $\odot$ operation represents element-wise multiplication, and function φ_{t-1}^T represents operations inside $t - 1$th temporal attention module to compute β_{t-1}.

3.2.4. Pseudocode of STA-LSTM.
After describing steps of building STA-LSTM network with mathematic functions, we provide detail pseudocode in Algorithm 1, where readers can easily understand the process of experiment and implementation details of STA-LSTM model.

3.3. Spatial-Temporal Attention Module.
Structure of the proposed spatial-temporal attention module is shown in Figure 4. Compared with traditional physical models which rely on expert knowledge and experience to manually assign factor weights, the proposed attention model can automatically select informative factors to forecast based on inherent characteristics of collected data, which is more flexible for different application scenarios.

3.3.1. Spatial Attention Module.
Acquired data from ECWMF sites is gathered with structure of grids ruled by latitude and longitude, which offers detailed information on spatial distribution of inputting hydrological factors. However, small radius precision, i.e., 0.01×0.01 radius, could greatly increase computation burden of the proposed model.
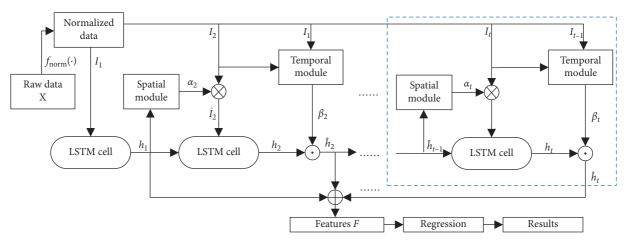
FIGURE 3: Network structure of the proposed STA-LSTM, which takes raw data as input and computes regression results to predict increase or decrease in run-off values. It is noted that we use blue rectangle to locate parts of STA-LSTM, which is described by equations and explanations in detail.

Input: Input dataset X with n samples, each sample refers to a hydrological feature set collected from total τ states, input initial super-parameters θ.
Output: Run-off value Prediction Y.
(1) Initialize STA-LSTM model with initial super-parameters θ and random network weights W_1.
(2)   Set sample index $i = 1$;
(3)   **for** $i \leq n$ **do**
(4)   Set current state $t = 2$, hidden output in the first state $h_1^i = f_{lstm,1}(X_1^i)$, weighted hidden output in the first state $\widehat{h}_1^i = h_1^i$;
(5)   **for** $t \leq \tau$ **do**
(6)   Input $\widehat{h}_{t-1}^i$ into $t-1$th spatial attention module to calculate spatial attention weight by $\alpha_t = \varphi_{t-1}^S(\widehat{h}_{t-1}^i)$;
(7)   Refine current input X_t^i with spatial attention weight α_t by $\widehat{X_t^i} = \alpha_t \otimes X_t^i$;
(8)   Input $\widehat{X_t^i}$ and $\widehat{h}_{t-1}^i$ into tth LSTM cell to compute current hidden output by $h_t^i = f_{lstm,t}(\widehat{X_t^i}, \widehat{h}_{t-1}^i)$;
(9)   Input X_{t-1}^i and X_t^i into $t-1$th temporal attention module to compute temporal attention weight by $\beta_t = \varphi_{t-1}^T(X_{t-1}^i, X_t^i)$;
(10)   Refine current hidden output h_t^i with spatial attention weight β_t by $\widehat{h_t^i} = \beta_t \odot h_t^i$.
(11) t ++;
(12)   **end for**
(13)   Increase or Decrease value in Run-off $Y^i = \text{Outlayer}(\widehat{h_1^i}, \widehat{h_2^i}, \ldots, \widehat{h_\tau^i})$
(14)   Calculate RMSE with Y^i and ground-truth value.
(15)   **if** RMSE decreases **then**
(16)   Update model weights with: $W_{i+1} = W_i - (\partial loss / W_i)$;
(17)   **else**
(18)   Continue;
(19)   **end if**
(20) i ++;
(21) **end for**
(22) End training process and Save model parameters.
(23) Input testing samples to output corresponding prediction value Y.
(24) Calculate RMSE, DC and MAPE for testing samples.

ALGORITHM 1: Pseudocode of STA-LSTM for implementation.

Therefore, we reconstruct organization form of the acquired data to keep balance on precision and effectiveness, where the reconstructed data structure is shown in Figure 5. We can notice that Figure 5(a) is abstracted from Figure 1 with flip operation and large spatial grids. After accumulating data from original and small grids into large grid, we finally achieve a novel and effective representation for input hydrological factors in Figure 5(b), where each factor can be viewed as a 3D dimension vector with feature, position, and time values inside.

Informativeness for input hydrological factors varies greatly in different locations. For example, regions near Lech river should be more important than regions far away, since rainfall near river can quickly be converged to increase run-off values. To utilize spatial property of input hydrological factors, a spatial attention module is constructed to assign weights for hydrological features acquired from different location grids. Essentially, spatial attention module explores interchannel relationship among features obtained from
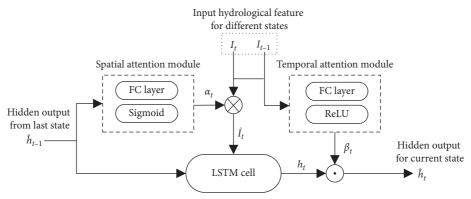
FIGURE 4: Architecture of the proposed spatial-temporal attention module.

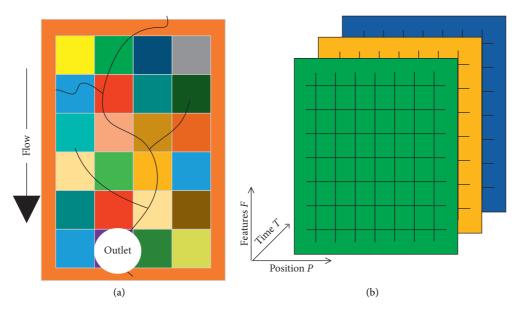(a)   (b)

FIGURE 5: Illustration of the reconstructed data structure for information acquired from ECWMF: (a) abstract Lech river basin with large spatial grids and (b) data representation for input hydrological factors with three independent dimensions to represent feature, position, and time, respectively.

different locations, which help STA-LSTM to pay more attention on salient grids for accurate forecasting.

As shown in Figure 4, input feature I_t is processed by a fully connected layer and a sigmoid function to output spatial-wise weight α_t:

$$\alpha_t = f_s\left(W_S \widetilde{h}_{t-1} + b_s\right), \tag{9}$$

where function $f_s()$ refers to sigmoid operation, W_S and b_s represent weighting matrix and bias parameters for fully connected network, respectively.

3.3.2. Temporal Attention Module. Considering that there generally exists a trend in sequential data, researchers design Holt-Winters double exponential smoothing filter to describe relationship between current and former observation values. Following such supposition and implementing it with a dynamically updating weight scheme, we utilize temporal attention module to assign weights for hidden output of STA-LSTM cells,

which acts as a relation modeling function between observations at different timings.

As shown in Figure 3, we utilize hidden output of current state and former state to construct temporal attention module, which explores the difference between two states to decide whether current input is informative. The detailed structure of the proposed temporal module is shown in the right part of Figure 4, where we compute temporal weight β_{t-1} for state $t-1$ as

$$\beta_{t-1} = f_R\left(W_{t-1,t-2} I_{t-2} + W_{t,t-1} I_{t-1} + b_{t-1}\right), \tag{10}$$

where function $f_R()$ represents ReLU function, $W_{t-1,t-2}$ and $W_{t-1,t-2}$ refer to parameter matrix required to be defined during training, and b_{t-1} is the bias vector. In fact, temporal-wise weight is key to control information passing through network from former hidden output to next cell. Therefore, temporal attention module is a beneficial complementary to spatial attention module.

4. Experiments

In experiment section, we firstly introduce dataset and measurements. Then, we design two groups of ablation experiments to discuss sensibility of hydrological features and effectiveness of the proposed attention module. Afterwards, we conduct comparative studies with several flood forecasting methods to compare effectiveness. Finally, we offer implementation details of STA-LSTM.

4.1. Dataset and Measurement. We utilize two datasets to prove the effectiveness of STA-LSTM, i.e., Lech and Changhua river basins. It is noted that they have differences in features, since they are collected from ECWMF and cooperation government departments, respectively. Specifically, we utilize the tool provided by ECWMF to collect 7360 hydrological instances of Lech river basin varying from May 1, 2002, to January 1, 2018. These instances have shown significant increase of run-off values, which provide raw data to detect patterns of variation for run-off factor. Meanwhile, we collect 8555 samples varying from January 1, 1998, to December 31, 2010, which represent 40 floods and are manually recorded hydrological data from rainfall, evaporation, and gaging station in Changhua watershed. It is noted that Lech dataset contains sufficient information on adopted hydrological features, i.e., precipitation, evaporation, soil tension water, temperature, and wind. Meanwhile, Changhua dataset is short of information on temperature and wind. Shortage of these two minor hydrology factors does not have a great impact on accuracy of flooding results. Moreover, we achieve data of soil tension water according to calculation of Xinanjiang model (Short for XAJ model), which is a famous physical model to forecast flood in semihumid regions.

Table 1 offers descriptive statistics for flow and rainfall data collected from Lech and Changhua river basin, where we can observe that data representation for both datasets is different. This is due to their distinctive data collection operations, where Lech dataset is constructed from remote sensing imageries and data in Changhua Dataset is collected manually. From Table 1, we can notice obvious difference in data distribution between Lech and Changhua rivers, since characteristics of different river dataset varies greatly from one to another. This phenomenon brings large difficulty to accurately forecast river run-off values, since it requires models to describe relation function between input hydrological features and run-off values without overfit performance.

Due to the nature of ECWMF, i.e., they collect hydrological data every 3 hours, we refer each state in STA-LSTM as 3 hours for modeling. As represented in Figure 5, we utilize such 3D feature by defining time value (state value) as 6, which results in an input feature to describe hydrological factors in 18 hours. After training, we perform regression task with STA-LSTM on run-off values for next 1, 2, and 3 states based on input of hydrological features of former 6 states.

We use standard quality measures, i.e., Root Mean Square Error (RMSE), Mean Absolute Percent Error (MAPE), and Deterministic Coefficient (DC) to measure the quality of flood forecasting. These three measurements are formulated as

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - q_i)^2},$$

$$\text{MAPE} = \frac{100}{n} \sum_{i=1}^{n} \left| \frac{y_i - q_i}{y_i} \right|, \quad (11)$$

$$\text{DC} = 1 - \frac{\sum_{i=1}^{n} (y_i - q_i)^2}{\sum_{i=1}^{n} (q_i - \overline{q})^2},$$

where y_i and q_i refer to forecasting run-off values and corresponding ground-truth run-off values, $\overline{q}$ refers to average of ground-truth run-off values, and n is number of testing instances. It is noted that higher DC value implies more convinced flood forecasting results. Meanwhile, RMSE and MAPE are used to quantify similarity between forecasting results and groud-truth values, where smaller values in RMSE and MAPE indicate high accuracy on flood forecasting quality.

4.2. Performance Analysis. We design three comparative experiments to analyze performance of STA-LSTM. Specifically, the first experiment is designed to estimate sensibility of input hydrological features, the second one is used to compare the effectiveness of STA-LSTM with or without attention module, and the last experiment aims to compare performance of STA-LSTM with comparative methods. For all experiments in this paper, input time period is settled form T-5 to T, which makes input data contain hydrological features of 6 states before T.

4.2.1. Feature Sensibility Experiment. We show related experimental data on feature sensibility in Table 2. In each case of experiment, we eliminate one input hydrological feature and keep other inputs remain same, which could show sensibility of specific feature in obtaining accurate forecasting results. Due to shortage of two hydrology features in Changhua dataset, we prefer Lech dataset to complete feature sensibility experiment. To better compare results, we offer two more statistics data in tables, i.e., mean and bias value represented with subscripts of A and σ, where the latter one calculates the difference value between result under current running condition and achieved by STA-LSTM.

From Table 2, we can notice smallest value in RMSE is achieved by Wo-Temperature, and the best performance in MAPE and DC are achieved by Wo-Wind. In other words, eliminating factors of either temperature or wind have a small impact on forecasting results. Based on this observation, we can conclude these two hydrological factors are less related with formatting of floods. Meanwhile, the worst performance in three measurements is all achieved by Wo-

TABLE 1: Descriptive statistics of daily flow and relevant data for Changhua and Lech dataset, where evaluation SD refers to standard deviation and p and R represent flow rate values and rainfall values, respectively. It is noted that subscripts c and l refer to gauging stations in Chuanghua and Lech River, respectively; subscript g refers to mean evaluation on basin areas of Lech river; subscripts DSW, THC, and other abbreviations represent names of rainfall stations for Changhua dataset. Note that unit for p is defined as m^3/s and units for R is settled as mm/h.

Evaluation	Changhua river basin								Lech river basin	
	p_c	R_{DSW}	R_{THC}	R_{LMS}	R_{SS}	R_{LX}	R_{YLG}	R_{CH}	p_l	R_g
Mean	146.651	0.596	0.618	0.704	0.786	0.712	0.822	0.631	433.395	0.931
SD	202.501	2.411	2.303	2.636	2.553	2.560	2.666	3.405	107.557	2.659
Median	80.320	0	0	0	0	0	0	0	48.780	0
Skewness	3.987	11.026	7.946	10.700	7.819	8.599	7.484	27.617	2.875	12.916
Kurtosis	24.388	198.996	89.910	190.660	104.110	120.549	100.293	1124.804	0.574	213.598

TABLE 2: Comparison experiments on Lech dataset to show sensibility of different input hydrological features, where Wo- refers to operation of eliminating one specifical hydrological feature, ST-Water is short for hydrological feature of soil tension water, subscripts A and σ represent mean value and bias value respectively, R, M and D represents measurements of RMSE, MAPE, and DC, respectively. Note that bold texts indicate best performance among comparative methods. We define unit for M as %.

Methods	R_{T+3}	R_{T+6}	R_{T+9}	R_A	R_σ	M_{T+3}	M_{T+6}	M_{T+9}	M_A	M_σ	D_{T+3}	D_{T+6}	D_{T+9}	D_A	D_σ
Wo-precipitation	117.0	121.3	145.3	127.9	61.88	29.61	31.96	36.31	32.63	16.12	0.35	0.33	0.28	0.32	0.49
Wo-evaporation	65.82	74.08	85.16	75.02	9.00	16.49	18.53	22.39	19.14	2.63	0.80	0.76	0.64	0.73	0.08
Wo-ST-water	80.44	74.74	77.92	77.70	11.68	19.91	18.92	19.64	19.49	2.98	0.71	0.80	0.73	0.75	0.06
Wo-temperature	63.38	68.51	75.50	69.13	**3.11**	15.94	17.18	19.04	17.39	0.88	0.82	0.79	0.74	0.78	0.03
Wo-wind	63.81	68.12	75.91	69.28	3.26	15.98	17.06	18.32	17.12	**0.61**	0.82	0.78	0.76	0.79	**0.02**
STA-LSTM	**60.94**	**65.56**	**71.56**	**66.02**	—	**15.24**	**16.39**	**17.89**	**16.51**	—	**0.84**	**0.81**	**0.77**	**0.81**	—

Precipitation, which proves that rainfall is the most significant factor to accurately forecast flood.

For experiment of Wo-Evaporation, we can notice bias values in three measurements increase with larger forecasting time. On the contrary, we can find bias values corresponding to Wo-ST-Water decrease with larger forecasting time. Based on this observation, we could know that importance of evaporation gradually becomes larger for long-time forecasting, while Wo-ST-Water mostly contributes to short-time forecasting. Essentially, ST-Water defines the initial state before formation of flood, which makes it significant to short-time forecasting. Meanwhile, evaporation affects the formation of flood throughout the whole process of flood. Last but not least, we should notice that both factors of evaporation and ST-Water are not major features for accurate prediction, when comparing with precipitation.

4.2.2. Ablation Experiment. Details of comparative experiment on effectiveness of attention modules are shown in Table 3. Specifically, we perform three cases of experiments with spatial attention module only, with temporal attention module only, and with both modules, respectively.

From Table 3, we can observe that measurement-related performance obtained by STA-LSTM is larger when comparing with spatial or temporal attention module only, which proves effectiveness of attention module to involve context information for hydrological feature enhancement. We can also notice that most cases of second best performance and best performance in bias value are achieved by method with temporal module, which proves that temporal attention information contributes more to forecasting than

spatial attention information. Essentially, flood is a complex procedure of run-off generation, separation, and routing, which leads timing to be an important factor for flood forecasting. Therefore, temporal context information extracted from sequential data contributes more to accurate forecast flood.

4.2.3. Experiment with Comparative Methods. Tables 4 and 5 offer the detailed statistics by testing STA-LSTM and comparative methods on Lech and Chuanghua dataset, respectively. Specifically, we implement SVM, LSTM, and 10 layers FCN (fully connected network) to work as comparative methods. For fair comparisons, structure and parameters of LSTM are settled to be the same with STA-LSTM. Moreover, we implement XAJ model as a comparative study in Table 5 to offer data for comparing data-driven models with physical models. Reason to abandon usage of XAJ model on Lech dataset lies in the fact that XAJ model is specially designed for Changhua river basin or other semihumid regions, which is not fit for Lech river basin under our supposition.

As proved by R_A, M_A, and D_A in Table 4, STA-LSTM achieves the best performance in Lech dataset. Meanwhile, Table 5 shows that STA-LSTM achieves the best performance in DC and second best performance in RMSE and MAPE after conducting experiments on Changhua dataset. By comparing between XAJ model and STA-LSTM in Table 5, we can notice that specifically designed physical model, i.e., XAJ model, is capable of obtaining a higher accuracy on flood forecasting, especially in long-time period forecasting. Such phenomenon can be explained by the truth that flood is a complex process for data-driven modeling under a limited

TABLE 3: Comparison experiments on Changhua dataset to test effectiveness of attention module. Note that bold and underline texts indicate best and secondary best performance among comparative methods, respectively. We define unit for M as %.

Methods	R_{T+3}	R_{T+6}	R_{T+9}	R_A	R_σ	M_{T+3}	M_{T+6}	M_{T+9}	M_A	M_σ	D_{T+3}	D_{T+6}	D_{T+9}	D_A	D_σ
With spatial	54.51	79.77	95.26	76.51	1.42	13.59	19.93	23.81	19.11	0.34	0.84	0.69	0.60	0.71	0.02
With temporal	54.33	78.46	94.61	75.80	**0.71**	13.55	19.64	23.67	18.95	**0.18**	0.84	**0.71**	0.61	0.72	**0.01**
STA-LSTM	**52.37**	**78.36**	**94.55**	**75.09**	—	**13.09**	**19.59**	**23.64**	**18.77**	—	0.85	0.71	0.62	0.73	—

TABLE 4: Comparisons on Lech dataset with several comparative methods. Note that bold texts indicate best performance among comparative methods.

Models	R_{T+3}	R_{T+6}	R_{T+9}	R_A	M_{T+3} (%)	M_{T+6} (%)	M_{T+9} (%)	M_A	D_{T+3}	D_{T+6}	D_{T+9}	D_A
FCN	**51.27**	82.54	123.4	85.74	**12.82**	20.64	30.85	21.44	**0.88**	0.70	0.32	0.63
SVM	63.99	79.98	92.49	78.82	16.00	20.00	23.12	19.71	0.82	0.72	0.62	0.72
LSTM	68.92	75.12	80.85	74.96	17.23	18.78	20.21	18.74	0.79	0.75	0.71	0.75
STA-LSTM	60.94	**65.56**	**71.56**	**66.02**	15.24	**16.39**	**17.89**	**16.51**	0.84	**0.81**	**0.77**	**0.81**

TABLE 5: Comparisons on Changhua dataset with several comparative methods. Note that bold texts indicate best performance among comparative methods.

Models	R_{T+3}	R_{T+6}	R_{T+9}	R_A	M_{T+3} (%)	M_{T+6} (%)	M_{T+9} (%)	M_A (%)	D_{T+3}	D_{T+6}	D_{T+9}	D_A
FCN	56.87	88.36	129.7	91.63	14.22	22.09	32.41	22.91	0.84	0.64	0.31	0.60
SVM	60.89	86.34	110.6	85.93	15.22	21.59	27.64	21.48	0.82	0.67	0.53	0.67
LSTM	66.11	84.49	109.2	86.60	16.53	21.12	27.30	21.65	0.79	0.68	0.54	0.67
XAJ model	77.30	**68.31**	**69.25**	**71.62**	19.26	**17.09**	**17.45**	**17.93**	0.65	0.70	**0.71**	0.69
STA-LSTM	**52.37**	78.36	94.55	75.09	**13.09**	19.59	23.64	18.77	**0.85**	**0.71**	0.62	**0.73**

size of data. Therefore, utilizing insufficient data to fit flood process without building inherent and knowledge-embedded relations could not work well for long-time period forecasting. Furthermore, errors and noise for data-driven are easy to accumulate without appropriate means of error optimization during forecasting process.

By comparing STA-LSTM with other data-driven models in Tables 4 and 5, we can notice that FCN performs better than STA-LSTM for forecasting at $T + 3$ in Lech dataset. However, it fails to obtain consistent performance for long-time period forecasting, i.e., $T + 6$ and $T + 9$. In fact, ten layer structure of FCN with limited size of parameters makes it suitable to fit in cases of relatively simple short-time forecasting. However, complexity increases in a large degree with longer forecasting period, which results in worse performance with insufficient parameters of FCN to model and optimize complexity. SVM is widely used to handle cases of learning with limited size of data. However, traditional SVM is not appropriate to conduct regression inference based on complex and sequential data, which leads to worse performance achieved by SVM than STA-LSTM. Original LSTM performs worse than STA-LSTM in all cases, which proves that attention module is of significance in improving accuracy by focusing on informative hydrological factors and timings, especially in forecasting with small dataset.

It is evident to observe that all data-driven methods perform better for short-time forecasting, i.e., $T + 3$, since core task of short-time forecasting for data-driven model is to fit data with suitable parameters and prevent over-fitting. Due to accumulations of uncertainty and errors sourced from models and input factors like weather forecast, there exist a decrease in performance with long-time forecasting. To deal with complexity of long-time forecasting, the proposed STA-LSTM is built on the basis of LSTM structure to resolve long-time dependencies, which designs cell memory to represent and memorize long-time dependencies. Moreover, STA-LSTM inherently models context information to better describe long-term memory and decreases impacts brought by noisy input. As a result, STA-LSTM is capable of forecasting flood with higher accuracy and longer time period than other data-driven models, which is proved by best performance at $T + 6$ and $T + 9$.

We adopt one instance in test datasets to compare forecasting performance on run-off values achieved by STA-LSTM and other comparative methods in Figure 6. According to plot for $T + 3$, we can find that FCN performs best, performance of STA-LSTM is close to LSTM, and SVM achieves the worst forecasting result. Essentially, the forecasting curve obtained by SVM is too smooth to coincide with great time-varying characteristics of ground-truth run-off values, due to its inherent modeling supposition that output values should be smooth to a certain extent. Due to smooth output property, we could find that forecasting results of SVM are low in RMSE, but fail to coincide with time-varying run-off forecasting plot. By comparing forecasting curves between short-time and long-time forecasting, we can notice obvious distortions for $T + 6$ and $T + 9$ due to large increase in complexity and difficulty of forecasting task. Among methods for long-time forecasting, STA-LSTM performs best, LSTM and SVM achieves slightly worse results, and FCN performs worst. Above all, overall forecasting performance of STA-LSTM is much more accurate and consistent
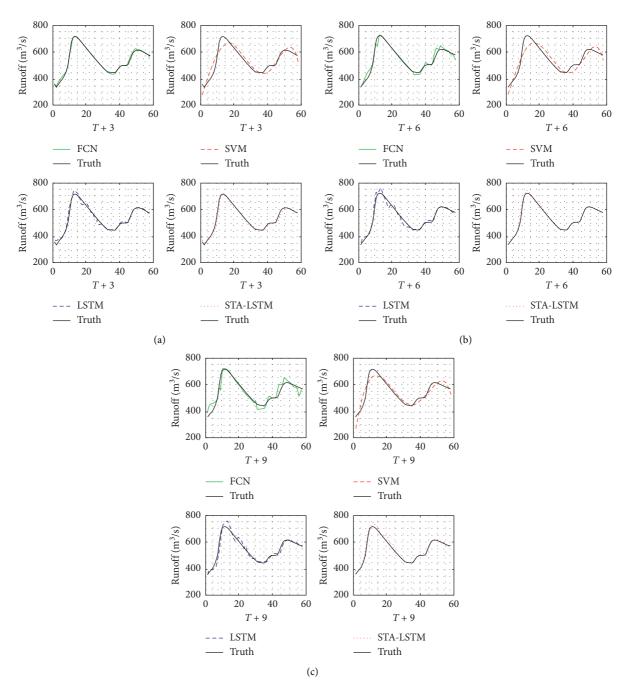
FIGURE 6: Comparisons in forecasting run-off values among STA-LSTM and other comparative methods, where three groups of plots refer to forecasting results performing for (a) $T + 3$, (b) $T + 6$, and (c) $T + 9$, respectively.

than comparative methods in this case study. Meanwhile, plots achieved by SVM and LSTM are close in visual effects, where we could find a certain and obvious degree of distortion. Last but not least, FCN obtain forecasting plot with serious distortion and large deviation in several key timings.

4.3. Implementation. Experiments were performed on a server (2.4 GHz 6-core Xeon CPU, 60 GB RAM and one Nvidia GeForce GTX 1080 Ti card). For Lech and Changhua datasets, we utilize 4-folder cross validation to test STA-LSTM and other comparative methods. During network

design, dimension of hidden state, and total state number of STA-LSTM network are defined as 128 and 32, respectively. During training, learning rate, weight decrease, and iteration times of STA-LSTM network are defined as 0.0025, 10^{-6}, and 500, respectively. We update the learning rate every 100 times, and the corresponding decrease rate is 0.01.

5. Conclusion

Facing difficulties in transforming problem-related data from physical space to cyber space and utilizing models to

solve problems in cyber space, we firstly define the problem of flood from views of both physical and cyber space, and then propose STA-LSTM by embedding spatial-temporal attention information for flood forecasting. STA-LSTM could selectively utilize informative hydrological features acquired from significant locations and timings. Experiments on Lech and Changhua river basins prove the effectiveness of STA-LSTM by comparing with several comparative studies. Our future work includes construction of a light-weight flood forecasting model by eliminating useless hydrological features, which not only boosts running speed of flood forecasting system, but also largely decrease complexity in collecting data and modeling feature relationship.

Data Availability

The image and acquired sensor data used to support the findings of this study were supplied by Yukai Ding under license and so cannot be made freely available. Requests for access to these data should be made to Yirui Wu (wuyirui@hhu.edu.cn).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] S. Han and P. Coulibaly, "Bayesian flood forecasting methods: a review," *Journal of Hydrology*, vol. 551, pp. 340–351, 2017.

[2] A. Kauffeldt, F. Wetterhall, F. Pappenberger, P. Salamon, and J. Thielen, "Technical review of large-scale hydrological models for implementation in operational flood forecasting schemes on continental level," *Environmental Modelling & Software*, vol. 75, pp. 68–76, 2016.

[3] W. Collischonn, R. Haas, I. Andreolli, and C. E. M. Tucci, "Forecasting river Uruguay flow using rainfall forecasts from a regional weather-prediction model," *Journal of Hydrology*, vol. 305, no. 1-4, pp. 87–98, 2005.

[4] V. A. Siqueira, R. C. D. Paiva, A. S. Fleischmann et al., "Toward continental hydrologic-hydrodynamic modeling in South America," *Hydrology and Earth System Sciences*, vol. 22, no. 9, pp. 4815–4842, 2018.

[5] Y. Wu, W. Xu, J. Feng, S. Palaiahnakote, and T. Lu, "Local and global Bayesian network based model for flood prediction," in *Proceedings of International Conference on Pattern Recognition*, Beijing, China, pp. 225–230, 2018.

[6] G. Corani and G. Guariso, "Coupling fuzzy modeling and neural networks for river flood prediction," *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*, vol. 35, no. 3, pp. 382–390, 2005.

[7] F. M. Fan, D. Schwanenberg, W. Collischonn, and A. Weerts, "Verification of inflow into hydropower reservoirs using ensemble forecasts of the TIGGE database for large scale basins in Brazil," *Journal of Hydrology: Regional Studies*, vol. 4, pp. 196–227, 2015.

[8] P. R. M. Pontes, F. M. Fan, A. S. Fleischmann et al., "MGB-IPH model for hydrological and hydraulic simulation of large floodplain river systems coupled with open source GIS," *Environmental Modelling & Software*, vol. 94, pp. 1–20, 2017.

[9] P. R. M. Pontes, R. B. L. Cavalcante, P. K. Sahoo et al., "The role of protected and deforested areas in the hydrological processes of Itacaiúnas river Basin, eastern Amazonia," *Journal of Environmental Management*, vol. 235, pp. 489–499, 2019.

[10] H. Wang, C. Wang, Y. Wang, X. Gao, and C. Yu, "Bayesian forecasting and uncertainty quantifying of stream flows using Metropolis-Hastings Markov chain Monte Carlo algorithm," *Journal of Hydrology*, vol. 549, pp. 476–483, 2017.

[11] P.-S. Yu, S.-T. Chen, and I.-F. Chang, "Support vector regression for real-time flood stage forecasting," *Journal of Hydrology*, vol. 328, no. 3-4, pp. 704–716, 2006.

[12] J. Liu, G. Wang, P. Hu, L. Duan, and A. C. Kot, "Global context-aware attention LSTM networks for 3D action recognition," in *Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3671–3680, Honolulu, HI, USA, July 2017.

[13] J. Juliang, W. Yiming, and Y. Xiaohua, "Genetic algorithm based neural network and application in modeling vulnerability of flood disaster bearing body," *Journal of Natural Disasters*, no. 2, pp. 53–60, 1998.

[14] W. Minghua, D. Peng, and L. Zhijia, "Comparison between applications of artificial neural network model and Xinanjiang model," *Hydrology*, vol. 28, no. 6, pp. 33–35, 2008.

[15] C. T. Cheng, W. J. Niu, Z. K. Feng, J. J. Shen, and K. W. Chau, "Daily reservoir runoff forecasting method using artificial neural network based on quantum-behaved particle swarm optimization," *Water*, vol. 7, no. 8, pp. 4232–4246, 2015.

[16] C. H. R. Lima, U. Lall, T. Troy, and N. Devineni, "A hierarchical Bayesian GEV model for improving local and regional flood quantile estimates," *Journal of Hydrology*, vol. 541, pp. 816–823, 2016.

[17] D. R. H. O'Connell, "Nonparametric Bayesian flood frequency estimation," *Journal of Hydrology*, vol. 313, no. 1–2, pp. 79–96, 2005.

[18] D. Biondi, P. Versace, and B. Sirangelo, "Uncertainty assessment through a precipitation dependent hydrologic uncertainty processor: an application to a small catchment in southern Italy," *Journal of Hydrology*, vol. 386, no. 1-4, pp. 38–54, 2010.

[19] X. Xu, Y. Li, T. Huang et al., "An energy-aware computation offloading method for smart edge computing in wireless metropolitan area networks," *Journal of Network and Computer Applications*, vol. 133, pp. 75–85, 2019.

[20] L. Qi, Y. Chen, Y. Yuan, S. Fu, X. Zhang, and X. Xu, "A QoS-aware virtual machine scheduling method for energy conservation in cloud-based cyber-physical systems," *World Wide Web*, vol. 23, no. 2, pp. 1275–1297, 2020.

[21] X. Xu, X. Zhang, H. Gao, Y. Xue, L. Qi, and W. Dou, "Become: blockchain-enabled computation offloading for IOT in mobile edge computing," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 6, pp. 4187–4195, 2020.

[22] X. Xu, C. He, Z. Xu, L. Qi, S. Wan, and M. Z. A. Bhuiyan, "Joint optimization of offloading utility and privacy for edge computing enabled IoT," *IEEE Internet of Things Journal*, p. 1, 2019.

[23] L. Qi, R. Wang, C. Hu, S. Li, Q. He, and X. Xu, "Time-aware distributed service recommendation with privacy-preservation," *Information Sciences*, vol. 480, pp. 354–364, 2019.

[24] X. Xu, S. Fu, L. Qi et al., "An IoT-oriented data placement method with privacy preservation in cloud environment," *Journal of Network and Computer Applications*, vol. 124, pp. 148–157, 2018.

[25] X. Xu, Q. Liu, Y. Luo et al., "A computation offloading method over big data for IoT-enabled cloud-edge computing," *Future Generation Computer Systems*, vol. 95, pp. 522–533, 2019.

[26] X. Xu, Y. Chen, X. Zhang, Q. Liu, X. Liu, and L. Qi, "A blockchain-based computation offloading method for edge computing in 5G networks," *Software: Practice and Experience*, 2019.

[27] X. Xu, Y. Xue, L. Qi et al., "An edge computing-enabled computation offloading method with privacy preservation for internet of connected vehicles," *Future Generation Computer Systems*, vol. 96, pp. 89–100, 2019.

[28] L. Qi, X. Zhang, W. Dou, C. Hu, C. Yang, and J. Chen, "A two-stage locality-sensitive hashing based approach for privacy-preserving mobile service recommendation in cross-platform edge environment," *Future Generation Computer Systems*, vol. 88, pp. 636–643, 2018.

[29] V. Mnih, N. Heess, A. Graves, and K. Kavukcuoglu, "Recurrent models of visual attention," in *Proceedings of Neural Information Processing Systems*, pp. 2204–2212, Montreal, Canada, December 2014.

[30] T. He, W. Huang, Y. Qiao, and J. Yao, "Text-attentional convolutional neural network for scene text detection," *IEEE Transactions on Image Processing*, vol. 25, no. 6, pp. 2529–2541, 2016.

[31] S. Song, C. Lan, J. Xing, W. Zeng, and J. Liu, "An end-to-end spatio-temporal attention model for human action recognition from skeleton data," in *Proceedings of AAAI Conference on Artificial Intelligence*, pp. 4263–4270, San Francisco, CA, USA, February 2017.

[32] J. Liu, A. Shahroudy, M. L. Perez, G. Wang, L. Y. Duan, and A. K. Chichung, "Ntu RGB+D 120: a large-scale benchmark for 3D human activity understanding," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, p. 1, 2019.

[33] S. Yeung, O. Russakovsky, N. Jin, M. Andriluka, G. Mori, and F. Li, "Every moment counts: dense detailed labeling of actions in complex videos," *International Journal on Computer Vision*, vol. 126, no. 2–4, pp. 375–389, 2018.

[34] K. Chen, L. Yao, D. Zhang, X. Wang, X. Chang, and F. Nie, "A semisupervised recurrent convolutional attention model for human activity recognition," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–10, 2019.

[35] P. Anderson, X. He, C. Buehler et al., "Bottom-up and top-down attention for image captioning and visual question answering," in *Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6077–6086, Salt Lake City, UT, USA, June 2018.

[36] L. Ren, X. Cheng, X. Wang, J. Cui, and L. Zhang, "Multi-scale dense gate recurrent unit networks for bearing remaining useful life prediction," *Future Generation Computer Systems*, vol. 94, pp. 601–609, 2019.

[37] L. Jintao, L. Xiaopeng, C. Xi, and M. Can, "Soil moisture content distribution in intermittent rainfall and its effect on infiltration," *Journal of Soil and Water Conservation*, vol. 23, no. 5, pp. 96–100, 2009.