Context and Temporal Aware Attention Model for Flood Prediction

Zhaoyang Liu[1], Yirui Wu[1,2], Yukai Ding[2], Jun Feng[2], and Tong Lu[1](✉)

[1] National Key Lab for Novel Software Technology, Nanjing University, Nanjing, China
zyliumy@gmail.com, lutong@nju.edu.cn

[2] College of Computer and Information, Hohai University, Nanjing, China
{wuyirui,fengjun}@hhu.edu.cn, 291809390@qq.com

Abstract. To minimize damages brought by floods, researchers pay special attentions to solve the problem of flood prediction. Multiple factors, including rainfall, soil category, the structure of riverway and so on, affect the prediction of sequential flow rate values, but factors are not always informative for flood prediction. Extracting discriminative and informative features thus plays a key role in predicting flow rates. In this paper, we propose a context and temporal aware attention model for flood prediction based on a quantity of collected flood factors. We build our model on top of Long Short-Term Memory (LSTM) networks, which selectively focuses on informative factors and pays different levels of attentions to the outputs of different cells. The proposed CT-LSTM network assigns time-varying weights to input factors at all the cells of LSTM network, and allocates temporal-dependent weights to the outputs of each LSTM cell for boosting prediction performance. Experimental results on a benchmark flood dataset with several comparative methods demonstrate the effectiveness of the proposed CT-LSTM network for flood prediction.

Keywords: Attention model · Context and temporal aware
Flood prediction

1 Introduction

Flood, as one of the most common and largely distributed natural diasters, happens occasionally and brings large damages to life and property. In the past decades, researchers have proposed a quantity of models for accurate and robust flood prediction. We generally category them into two types, namely, physical models [7,9,10] and data-driven models [4,14,16]. Physical models generally describe the formation of flood by using functions to represent complex hydrology processes from clues to results. However, such models are extremely sensitive to parameters [17], which require large research efforts of experts to adjust. On the contrary, data-driven models directly explore relations between river flow

R. Hong et al. (Eds.): PCM 2018, LNCS 11164, pp. 545–555, 2018.
https://doi.org/10.1007/978-3-030-00776-8_50

and flood factors from historical observations, without considering physical processes. Since the complex mechanism of flood results in large computations of physical models, data-driven models are more efficient and costless for flood prediction.

Inspired by the significant performance [8,13] of deep LSTMs, we intend to utilize such an architecture to discover the inherent relations between flood factors and flow rates. Due to the development of internet of things, researchers can gather a large set of relevant flood factors for prediction. However, not all the collected factors are representative and informative for flood prediction. For example, the water retained in soil, named as soil tension water, has great effects on formation of floods in humid areas, while it is irrelevant with flood in places with sandy soil [9]. This is because soil in humid areas contain a great amount of water, meanwhile sandy soil is quite low in capacity to contain water. The informativeness degrees of each flood factor may vary at different time points during the same flood. Take soil tension water as an example, its value is highly relevant with flow rate values in humid areas at the beginning of a flood. Once its value exceeds the maximum water containing capacity of soil in the middle of the flood, the value of soil tension water no more changes and contributes little to the variations of flow rate values. Therefore, we propose a context-aware attention module, which automatically focuses on discriminative factors for flood prediction. The learned attention to factors are content-dependent and allowed to vary over time. This selectively focusing mechanism has been demonstrated to be very effective in various applications, such as speech recognition [3] and action recognition [11].

Furthermore, we often get predictions on flow rate values under a reasonable assumption that there exists a trend in historical flow data. We thus propose a temporal-aware attention module for simulations of the trends embedded in historical flow data. For a sequence of floods, the proposed temporal-aware attention module explicitly learns and allocates content-dependent weights to predicting flow rate values at each time point. In fact, the idea of the proposed temporal-aware attention module is similar with Holt-Winters double exponential smoothing [15], which assigns higher weights to the nearby observations for more convinced predictions. Flow rate predictions at different time points thus have different degrees of importance and robustness to variations. Moreover, some flow rate predictions can be unreliable induced by noises of input factors. Learning weight distribution for flow rate predictions under a trend assumption can help exclude such unreliable predictions.

In summary, we aim to construct a context and temporal aware attention LSTM (CT-LSTM) network for accurate and robust flood prediction. The context-aware attention module learns weight schemes for input factors based on hidden output of each LSTM cell (representing contextual information [5] between two nearby cells) in a local sense. Meanwhile, the temporal-aware attention module learns weight structures for flow rate predictions of each LSTM cell in a global sense. We have made the following three main contributions in this work.

- To the best of our knowledge, this is the first context and temporal aware attention model designed based on the LSTM architectures for flood prediction.
- A temporal-aware attention module is designed to allocate content-dependent attention to different predictions under a reasonable trend assumption.
- The proposed method is powerful to discover the inherent patterns between input factors and flow rates, especially for regions whose flood formation mechanism is too complex to construct a convinced physical model.

2 Related Work

Considering the relevance to the proposed CT-LSTM network, we introduce the data-driven model for flood prediction and attention model in this section.

Data-driven Model. Early, Yu *et al.* [20] utilize the support vector machine to establish a real-time flood forecasting model, which applies a two-step grid search method to find the optimal parameters for SVM. Later, Cheng *et al.* [2] perform daily runoff forecasting by training artificial neural network with quantum-behaved particle swarm optimization, which achieves much better forecast accuracy than the basic ANN model. Recently, Wu *et al.* [16] construct a hierarchical Bayesian network for flood predictions of small rivers. They establish entities and connections of Bayesian network to represent variables and physical processes of the Xinanjiang model, *i.e.*, a famous physical model, which appropriately embeds hydrology expert knowledge for high rationality and robustness.

Due to high potentials of discovering distinctive features from data, researchers try to utilize deep learning architectures for flood prediction. For example, Zhuang *et al.* [21] design a novel Spatio-Temporal Convolutional Neural Network (ST-CNN) to fully utilize spatial and temporal information and automatically learn underlying patterns from data for extreme flood cluster prediction. Liu *et al.* [4] propose a deep learning approach by integrating stacked auto-encoders (SAE) and back propagation neural networks (BPNN) for the predictions of stream flow, which simultaneously takes advantages of the powerful feature representation capability of SAE and superior predicting capacity of BPNN. Most recently, Wu *et al.* [14] propose context-aware attention LSTM network to accurately predict sequential flow rate values based on a set of collected flood factors. The proposed method is built on it and involves the combination of context and temporal aware attention over all the steps of LSTM network for higher predicting accuracy.

Attention Model. When observing the real-world, human perception focuses selectively on parts of a scene to acquire information at specific places and times. The exploitation of an attention model has attracted increasing interests in various fields, such as machine translation, image recognition and action recognition. Their proposed attention models are generally constructed as a dimension of interpretability into internal representations by selectively focusing on specific

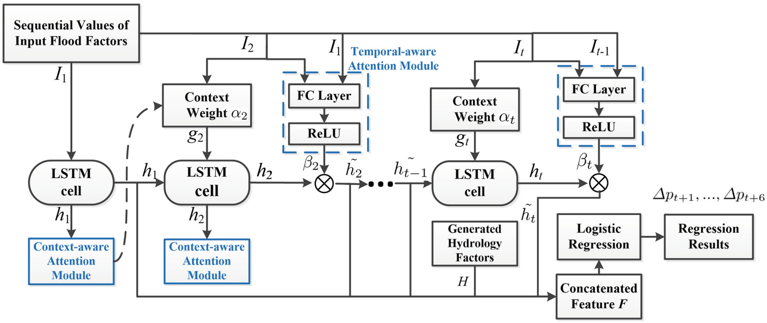

Fig. 1. Illustration of the proposed context and temporal aware attention LSTM network for flood prediction.

information. We categorize attention models into two classes, *i.e.* hard attention [19] and soft attention [1]. Hard attention mechanically chooses parts of the input data as focuses. For example, Mnit *et al.* [6] propose a hard attention model for image recognition, which adaptively selects a sequence of regions and processes the selected regions as inputs for RNN network.

On the contrary, soft attention takes the entire input into account by weighting each part or step dynamically. The fusion of neighboring frames within a sliding window with learned attention weights is proposed by Yeung *et al.* [18] to enhance the performance of dense labeling of actions in RGB videos. Liu *et al.* [5] propose a global context-aware attention LSTM for RGB-D action recognition, which recurrently optimize the global contextual information and further utilizes it as an informative function to assist accurate action recognition. Song *et al.* [12] achieve the goal of action recognition from skeleton data by selectively focusing on discriminative joints of skeleton within each frame of the inputs and assigning different levels of attention to the outputs of different frames. By designing context and temporal aware attention model as a soft attention scheme, the proposed method is reasonable to solve the regression problem of flow rate prediction.

3 LSTM Network with Context and Temporal Aware Attention Model

Take a typical river, *i.e.*, Changhua, as an example, we show its general information in Fig. 2, where we can notice 7 rainfall stations, 1 evaporation station and 1 river gauging station. In our work, we aim to predict the flow rate values at the river gauging station CH for the next 6 h with the proposed CT-LSTM network. The input set of flood factors consists of real rainfalls observed at rainfall stations, predicted rainfalls, evaporation observed at evaporation station SS and former river runoff observed at CH. We also utilize several intermediate variables such as total surface runoff, total interflow runoff and total groundwater runoff computed by a famous physical model, namely, the Xinanjiang Model. In

(a)

Station Name	Catchment Area (km^2)
Changhua(CH)	20.8
Longmengsi(LMS)	120.4
Taohuacun(THC)	133.0
Shuangshi(SS)	96.8
Daoshiwu(DSW)	256.1
Lingxia(LX)	116.7
Yulingguan(YLG)	161.2
Total	905.0

(b)

Fig. 2. Illustration of the Changhua watershed, where (a) is the map for various kinds of stations and (b) represents catchment areas corresponding to the listed rainfall stations. Note that we need predict the flow rate values of river gauging station CH and station SS functions as an evaporation station.

Xinanjiang Model, the outflow of a watershed can be subdivided into three components, including surface runoff, interflow runoff and groundwater runoff. Using these three components for prediction will provide more information about the watershed, which will be informative about the flood formation that cannot be precisely measured by sensors. In total, we prefer 7 features for prediction.

We propose an LSTM network with context and temporal aware attention mechanisms for flood prediction as shown in Fig. 1. We only feed the features mentioned above and last hidden state to the LSTM cell of our proposed network. The designed local context-aware attention and global temporal-aware attention module help automatically select relevant and informative features from the views of factors and the trend embedded, respectively. After paying different levels of attention on inputs and outputs of LSTM cells, we concatenate sets of the hidden outputs of cells $\{h_1, \tilde{h}_2, ..., \tilde{h}_t\}$ and generated hydrology factors H as a novel feature F for prediction. The reason to predict with sets of hidden outputs lies in the restriction of LSTM in perceiving the global contextual information with forgetting mechanism. However, the forgotten contextual information is important for the global regression problem.

3.1 Context-Aware Attention Module

Inspired by [6] which considers the attention problem as the sequential decision process of how an agent interact with a visual environment, the "interaction level" for the proposed context-aware attention module is essentially described by weights assigned to each feature. The context-aware attention module thus recurrently defines the corresponding weight vector α_t for input factors I_t as

$$\alpha_t = Nor(sig(W_{c,t-1}h_{t-1} + b_{c,t-1})) \tag{1}$$

where $W_{c,t-1}$ is the learnable parameter matrix, $b_{c,t-1}$ is the bias vector, h_{t-1} is the hidden output for each cell representing context information, function $sig()$ and $Nor()$ represent sigmoid function and normalization function, respectively.

Note that the proposed context-aware attention module determines the importance of input flood factors based on the hidden variables from an LSTM layer. In our work, the context-aware attention subnetwork actually composes of a fully connected layer and a normalization unit as suggested by Eq. 1.

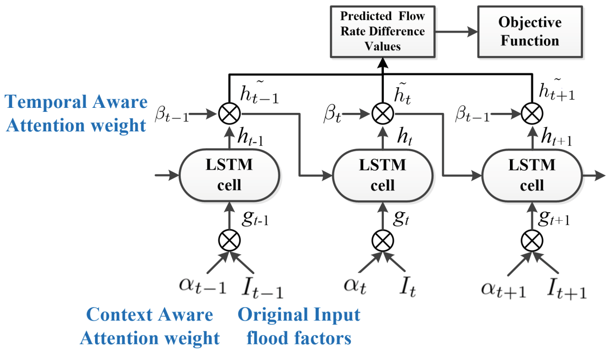

Fig. 3. Illustration of how context-aware attention weight α, temporal-aware attention weight β and objective function influence the CT-LSTM network.

The resulting weight vector α_t leads to the attention on informative factors, where Fig. 3 explains how the context-aware attention module works by a local way. We can find that the sequential input features $\{I_1, ..., I_t\}$ are separatively fed to all cells of CT-LSTM as the original time-varying description of flood factors. With the feature-wise weight vector α_t, the input of the informative flood factors g_t for each cell can be represented as:

$$g_t = I_t \bigotimes \alpha_t \tag{2}$$

where $\bigotimes$ represents the element-wise multiplication.

3.2 Temporal-Aware Attention Module

Holt-Winters double exponential smoothing filter considers there exists a trend behind a time-varying variable and utilizes a updating weight scheme to describe how prediction interacts with former observations. It has been successfully applied on smoothing of skeleton action data [15]. Follow the idea of Holt-Winters double exponential smoothing filter, we propose to use temporal-aware attention module to simulate the trend by globally assigning different levels of weights β_t to output of all the cells h_t as shown in Fig. 3. In fact, the hidden variable h_t contains information of past time points, benefiting from the merit of LSTM which is capable of exploring temporal long range dynamics. The weight vector β_t computed by temporal-aware attention module thus can adjust the input for the next cell $\tilde{h_t}$ based on information from a temporal long range:

$$\tilde{h_t} = h_t \bigotimes \beta_t. \tag{3}$$

As shown in Fig. 1, the temporal-aware attention module is composed of a fully connected layer and a ReLU nonlinear unit. The temporal weight vector β_t thus can be computed as

$$\beta_t = sig(W_{m,t-1}I_{t-1} + W_{m,t}I_t + b_{m,t}) \tag{4}$$

which depends on the former and current input flood factors I_{t-1} and I_t, respectively. We use the non-linear function of sigmoid due to its good convergence performance. The temporal weight vector control the amount of information of former predictions to be used for making the final prediction.

3.3 Design of Objective Function

How the context-aware attention module acts on the input flood factors and how the temporal-aware attention module acts on the hidden output of LSTM cells are given in Fig. 3. Constrained by the objective function, the main LSTM network, the context and temporal aware attention subnetwork can be jointly trained to implicitly learn the model. We thus formulate the final objective function of the context and temporal aware attention network with a regularized cross-entropy loss for a sequence of flood factors as

$$L_t = -\sum_{i=1}^{C}\sum_{t=1}^{6} loss(y_{i,t}, \tilde{p_i} + \Delta p_{i,t}) + \lambda \|W_N\|_2 \tag{5}$$

where C is the total number of training samples, $y_{i,t} = \{y_{i,t+1}, ..., y_{i,t+6}\}$ denotes the groundtruth flow rate values for the next 6 h corresponding to the ith training sample, $\tilde{p_i} = \sum_{j=0}^{4} y_{i,t-j}$ implies the mean of observed flow rate values for former 5 h and current time, $\Delta p_{i,t} = \{\Delta p_{i,t+1}, ..., \Delta p_{i,t+6}\}$ refer to the predicted difference flow rate values computed by the CT-LSTM, and function $loss()$ is defined as the smooth L1 loss function. The regularization item with L2 norm is to reduce overfitting of the networks. W_N denotes the connection matrix (merged to one matrix here) in the networks, including $W_{c,t}$ in Eq. 1 and $W_{m,t}$ in Eq. 4. Note that we use the back-propagation through time (BPTT) algorithm to minimize the loss function and adopt smooth L1 loss function. This is because the smooth L1 loss function makes the loss value convergent in a faster and more stable way comparing with adopting Root Mean Square Error.

4 Experimental Results

4.1 Dataset and Settings

Changhua Dataset. We collect hourly data of 40 floods happened from 1998 to 2010 in Changhua river as our original dataset. We use samples of flow rates every 11 h as to increase dataset size. After augmentation, the number of flood samples is increased to 8555. We utilize 8-fold cross validation and Root Mean Square Error (RMSE) to evaluate predictions:

$$RMSE = \sqrt{\frac{1}{n}\sum_{k=1}^{n}\sum_{t=1}^{6}(y_{k,t} - \tilde{p_k} - p_{k,t})^2} \quad (6)$$

where n refers to the number of testing samples. Note that smaller values of RMSE imply better performance the predicting achieves.

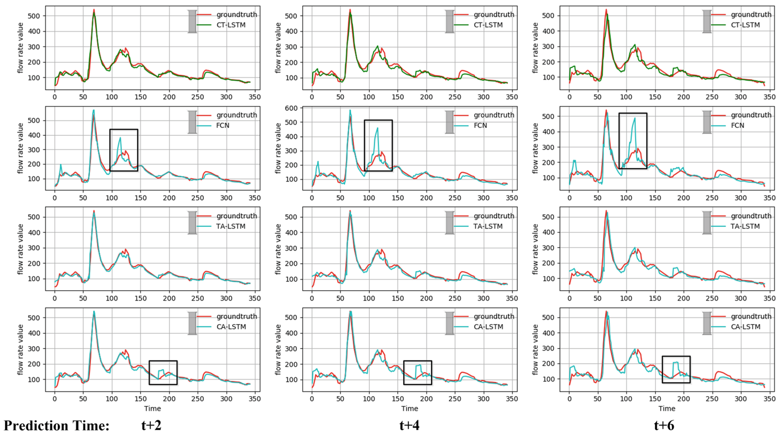

Fig. 4. Comparison with the ground-truth flow rate values and predicted flow rate values during a flood, where each row represents prediction results of the proposed CT-LSTM, FCN, TA-LSTM and CA-LSTM, respectively. Note that the rectangles indicate several obvious wrong predictions.

Implementation Details. For constructing the CT-LSTM network, we select t as 11, the dimension of hidden output as 128 and λ as 0.00005, respectively. We train the CA-LSTM network by setting learning rate, weight decay, epoch iterations and batch size as 0.00225, 10^{-6}, 500 and 100, respectively. The proposed CT-LSTM network runs on a workstation (2.4 GHz 6-core Xeon CPU, 60 G RAM and Nvidia GeForce GTX 1080Ti) for all the experiments.

4.2 Performance Analysis

We implement Fully-connected Network, CT-LSTM network without attention module (LSTM), CT-LSTM with only context-aware attention module (CA-LSTM), and CT-LSTM with only temporal-aware attention module (TA-LSTM) for comparisons. The main structures and training parameters of LSTM, CA-LSTM and TA-LSTM are exactly the same as CT-LSTM, meanwhile FCN is designed with 3 fully-connected layers. We compare the flow rate values prediction results of CT-LSTM, FCN, TA-LSTM and CA-LSTM in Fig. 4. We can see the CT-LSTM and TA-LSTM achieve nearly the same flow rates as the observed results. For CA-LSTM, we find it get obvious wrong predictions labeled by rectangles. We also view wrong predictions are enlarged when predicting with FCN.

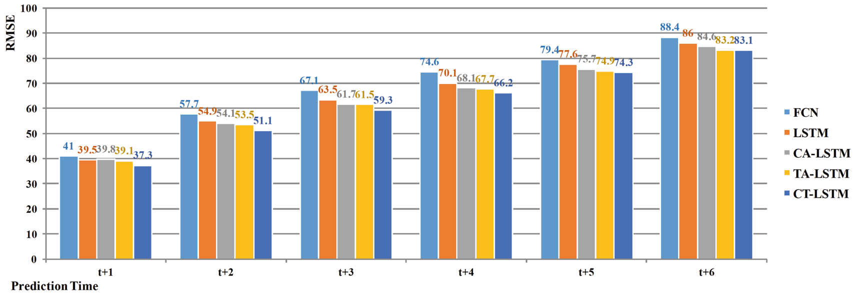

Fig. 5. Comparison of RMSE on Changhua Dataset computed by the proposed CT-LSTM, CA-LSTM, TA-LSTM, LSTM and FCN.

Figure 5 gives the detailed statistics of the proposed CT-LSTM network and several comparative methods on the Changhua dataset. As shown in Fig. 5, CT-LSTM network achieves the lowest RMSE values except for prediction at t+6. In fact, LSTM is designed to solve the problem of local dependencies with the forgetting structure, which implies LSTM network can not handle prediction with a rather long interval or delay. We thus observe that CT-LSTM performs nearly the same with TA-LSTM at t+6 due to the limitation of LSTM structure. FCN is not suitable for the time-varying prediction problem proved by much higher RMSE values comparing with other four LSTM-based methods. With the context-aware or the temporal-aware attention module, we find CA-LSTM and TA-LSTM achieve smaller RMSE than the conventional version of LSTM, which proves the effectiveness of the proposed context-aware and temporal-aware attention model. By jointly designing attentional module, the proposed CT-LSTM achieves the lowest RMSE, which proves the advantages of the structure of paying different levels of attentions on the input and output of LSTM cells for prediction.

5 Conclusions

In this paper, we extend the original LSTM network to achieve a context and temporal aware attention LSTM network for flood prediction, which is capable to selectively focus on informative flood factors and nearby predicted flow rate values. Experiment results on the Changhua dataset show the proposed method outperforms several comparative methods. Our future work includes the exploration on other hydrology purposes with the proposed method, such as mid-term flood predicting and flood frequency analysis.

Acknowledgment. This work was supported by the National Key R&D Program of China under Grant No. 2018YFC04000401, the Natural Science Foundation of China under Grant 61702160, 61370091, 61672273, the Fundamental Research Funds for the Central Universities under Grant 2016B14114, the Science Foundation of JiangSu under Grant BK20170892, the Science Foundation for Distinguished Young Scholars of Jiangsu under Grant BK20160021, the open Project of the National Key Lab for Novel Software Technology in NJU under Grant KFKT2017B05, and Scientific Foundation of State Grid Corporation of China (Research on Ice-wind Disaster Feature Recognition and Prediction by Few-shot Machine Learning in Transmission Lines).

References

1. Baradel, F., Wolf, C., Mille, J.: Pose-conditioned spatio-temporal attention for human action recognition. CoRR abs/1703.10106 (2017)
2. Cheng, C., Niu, W., Feng, Z., Shen, J., Chau, K.: Daily reservoir runoff forecasting method using artificial neural network based on quantum-behaved particle swarm optimization. Water **7**(8), 4232–4246 (2015)
3. Chorowski, J., Bahdanau, D., Serdyuk, D., Cho, K., Bengio, Y.: Attention-based models for speech recognition. In: Proceedings of NIPS, pp. 577–585 (2015)
4. Liu, F., Xu, F., Yang, S.: A flood forecasting model based on deep learning algorithm via integrating stacked autoencoders with BP neural network. In: Proceedings of IEEE International Conference on Multimedia Big Data, pp. 58–61 (2017)
5. Liu, J., Wang, G., Hu, P., Duan, L., Kot, A.C.: Global context-aware attention LSTM networks for 3d action recognition. In: Proceedings of IEEE CVPR, pp. 3671–3680 (2017)
6. Mnih, V., Heess, N., Graves, A., Kavukcuoglu, K.: Recurrent models of visual attention. In: Proceedings of NIPS, pp. 2204–2212 (2014)
7. Paquet, E., Garavaglia, F., Garçon, R., Gailhard, J.: The schadex method: a semi-continuous rainfall-runoff simulation for extreme flood estimation. J. Hydrol. **495**, 23–37 (2013)
8. Redmon, J., Divvala, S.K., Girshick, R.B., Farhadi, A.: You only look once: unified, real-time object detection. In: Proceedings of IEEE CVPR, pp. 779–788 (2016)
9. Ren-Jun, Z.: The xinanjiang model applied in china. J. Hydrol. **135**(1–4), 371–381 (1992)
10. Rogger, M., Viglione, A., Derx, J., Blöschl, G.: Quantifying effects of catchments storage thresholds on step changes in the flood frequency curve. Water Resour. Res. **49**(10), 6946–6958 (2013)
11. Sharma, S., Kiros, R., Salakhutdinov, R.: Action recognition using visual attention. CoRR abs/1511.04119 (2015)
12. Song, S., Lan, C., Xing, J., Zeng, W., Liu, J.: An end-to-end spatio-temporal attention model for human action recognition from skeleton data. In: Proceedings of AAAI, pp. 4263–4270 (2017)
13. Wei, L., Wu, Y., Wang, W., Lu, T.: A novel 3D human action recognition framework for video content analysis. In: Schoeffmann, K., et al. (eds.) MMM 2018. LNCS, vol. 10704, pp. 42–53. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-73603-7_4
14. Wu, Y., et al.: Context-aware attention LSTM network for flood prediction. In: Proceedings of ICPR (2018)
15. Wu, Y., Lu, T., Yuan, Z., Wang, H.: Freescup: a novel platform for assisting sculpture pose design. IEEE Trans. Multimed. **19**(1), 183–195 (2017)

16. Wu, Y., Xu, W., Feng, J., Shivakumara, P., Lu, T.: Local and global bayesian network based model for flood prediction. In: Proceedings of ICPR (2018)
17. Yao, C., Zhang, K., Yu, Z., Li, Z., Li, Q.: Improving the flood prediction capability of the xinanjiang model in ungauged nested catchments by coupling it with the geomorphologic instantaneous unit hydrograph. J. Hydrol. **517**, 1035–1048 (2014)
18. Yeung, S., Russakovsky, O., Jin, N., Andriluka, M., Mori, G., Fei-Fei, L.: Every moment counts: Dense detailed labeling of actions in complex videos. IJCV **126**(2–4), 375–389 (2018)
19. Yeung, S., Russakovsky, O., Mori, G., Fei-Fei, L.: End-to-end learning of action detection from frame glimpses in videos. In: Proceedings of IEEE CVPR, pp. 2678–2687 (2016)
20. Yu, P., Chen, S., Chang, I.: Support vector regression for real-time flood stage forecasting. J. Hydrol. **328**(3), 704–716 (2006)
21. Zhuang, W.Y., Ding, W.: Long-lead prediction of extreme precipitation cluster via a spatiotemporal convolutional neural network. In: Proceedings of the 6th International Workshop on Climate Informatics: CI (2016)