Anomaly Detection for Drinking Water Quality via Deep BiLSTM Ensemble

Xingguo Chen, Fan Feng, Jikai Wu, and Wenyu Liu
Jiangsu Key Laboratory of Big Data Security & Intelligent Processing, Nanjing University of Posts and Telecommunications.
State Key Laboratory for Novel Software Technology, Nanjing University.
Nanjing, Jiangsu Province, China.
chenxg@njupt.edu.cn,ffeng1017@gmail.com,wujikai983@gmail.com,and543359819lwy@gmail.com.

ABSTRACT

In this paper, a deep BiLSTM ensemble method was proposed to detect anomaly of drinking water quality. First, a convolutional neural network (CNN) is utilized as a feature extractor in order to process the raw data of water quality. Second, bidirectional Long Short Term Memory (BiLSTM) is employed to handle the time series prediction problem. Then, a linear combination of t-time-step predictions weighted by a discount factor was utilized to ensemble the final output of event. Finally, cost-sensitive learning combined with Adam optimization was applied to learn the model according to the imbalance property of the event label.

KEYWORDS

Convolutional Neural Networks, Bidirectional LSTM, and Discounted Ensemble.

ACM Reference Format:

Xingguo Chen, Fan Feng, Jikai Wu, and Wenyu Liu. 2018. Anomaly Detection for Drinking Water Quality via Deep BiLSTM Ensemble. In *Proceedings of the Genetic and Evolutionary Computation Conference 2018 (GECCO '18 Companion),* Jennifer B. Sartor, Theo D'Hondt, and Wolfgang De Meuter (Eds.). ACM, New York, NY, USA, Article 4, 2 pages. https://doi.org/10.1145/3205651.3208203

1 INTRODUCTION

Drinking water quality is of great importance to us human beings. However, in many places the quality of drinking water cannot be ensured [5]. The quality of drinking water can be affected by many factors, e.g., chemical pollutions, physical properties, etc. Then, it is meaningful for a machine learning algorithm to automatically detect anomaly event based on the monitor data of the drinking water [1].

Recently, water quality monitoring methods based on machine learning have been extensively studied. Support Vector Machine (SVM), Logistic Regression (LR), Linear Discriminant Analysis (LDA) have been applied to monitor water quality by using several water features as inputs [4]. From the contest results of GECCO '17, these methods have shown a great potential in anomaly detection. However, the performances of the above methods are still not ideal in practice.

1.1 Motivation

According to our observations, all the data features including event have time series characteristics. Thus, machine learning algorithms that handles time series problems are promising to achieve a better performance.

1.2 Our Approach

The data was minutely collected with properties of water quality and these properties vary with time. Such circumstance fits the typical application of a number of time series methods. The performance of predictions depends on how the model we are going to use extracts features hidden in the time series data. Therefore, we use convolutional neural network as a feature extraction method and the bidirectional long short term memory method to interpret the time series data. We choose the Adam algorithm to optimize the parameters of the BiLSTM. Based on the t-time-step predictions of deep BiLSTM, an ensemble of these predictions is summed linearly with a discount factor. From the data, we also found that the label EVENT is not distributed evenly. To deal with the problem of imbalanced labels, cost-sensitive learning is applied to enhance the overall performance.

2 DEEP BILSTM ENSEMBLE

We now present our Deep BiLSTM Ensemble. Although the BiLSTM layer has proven powerful in handling temporal correlation problems, it is not able to cope with spatial data. To address this problem, we propose an extension of BiLSTM where convolutional neural networks (CNN) structures are followed by the BiLSTM structure. We call this structure as "Deep BiLSTM". The whole structure of our proposed Deep BiLSTM is shown in Figure 1.

2.1 Convolutional Neural Networks

Our CNN structure consists of 2 convolutional and 2 max-pooling layers followed by 2 fully connected layers with both 128 neurons. At the input level, our training data are fed to the network as input. At the first convolutional layer, 20 kernels with stride of 1 pixel applied in the input data. And after local response normalization and max pooling operations, the feature maps are obtained. Then we add a flatten layer to make multidimensional input into one-dimensional input, which is commonly used in transition from convolution layer to fully connection layer.

GECCO '18 Companion, July 15–19, 2018, Kyoto, Japan

ACM ISBN 978-1-4503-5764-7/18/07…$15.00
https://doi.org/10.1145/3205651.3208203

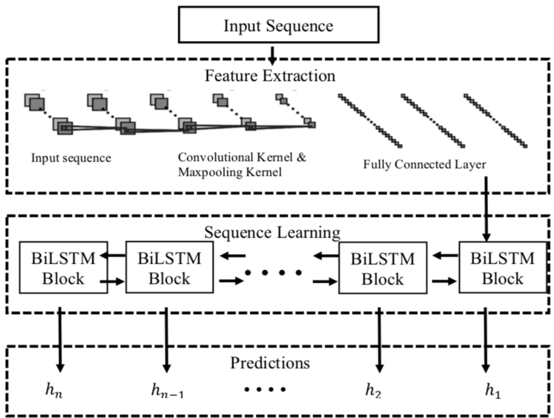

Figure 1: The Deep-BiLSTM structure, combined with a CNN module to extract information and a BiLSTM module to learn from sequences.

First, the training sequences data X was input into the CNN structure to extract the features of water quality data. Then, the features in sequences were entered into BiLSTM structure to perform the sequence learning and get the prediction vector $H = (h_1, h_2, \ldots, h_n)$. By forming and using this structure, a network model was built not only for the water quality prediction problem but also for more general spatio temporal sequence prediction problems.

2.2 Bidirectional Long Short Term Memory

As a variation of recurrent neural network (RNN), Bidirectional Long Short Term Memory (BiLSTM) was proposed to learn to store information over extended time intervals [2]. BiLSTM creates a new concept of a memory cell which essentially acts as an accumulator of the state information. If the input gets i_t is activated, information of new inputs will be accumulated. One of the advantages of BiLSTM is that its convergence performance is better compared to the basic RNN [6].

2.3 Linear Combination of t-time-step Prediction

The final output is defined as:

$$event = \begin{cases} 1, & \text{if p} \geq 0.5; \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

where p is the output of event probability,

$$p = \frac{\sum_{i=0}^{t-1} \gamma^i * h^{<i>}}{\sum_{i=0}^{t-1} \gamma^i}, \tag{2}$$

where $\lambda \in (0, 1)$ is a discount factor for $h^{<i>}$.

2.4 Cost-Sensitive Learning

According to our observations, the event label is imbalanced, where positive labels are much less than negative labels. Common methods to deal with this problem include: (i) generate positive samples in order that the numbers of positive labels and negative labels are equal; (ii) cost-sensitive learning with different costs for different misjudgements. Cost-sensitive learning is choosed here due to its conveniences and better performances [7]. The core of cost-sensitive learning is a cost matrix, i.e., $\begin{bmatrix} 0 & c(+,-) \\ c(-,+) & 0 \end{bmatrix}$, where $c(-,+) > c(+,-)$.

2.5 Stochastic Optimization

We choose Adam as our optimization method for LSTM network. Adam is an algorithm for first-order gradient-based optimization of stochastic objective functions, based on adaptive estimation of lower-order moments. Adam performs well when dealing with big-scale data [3]. The parameter updating process is shown as follows: $\theta_{t+1} = \theta_t - \frac{\alpha}{\sqrt{\hat{v_t}}+\varepsilon}\hat{m_t}$, where $\hat{m_t}$ is the mean gradient, $\hat{v_t}$ is the gradient variance, α is a learning rate, and ϵ is a small positive value.

3 CONCLUSION

The drinking water data contains not only the information besides features but also the information besides time series. The related works consider only the former information. The proposed "Deep BiLSTM Ensemble" is able to address both the features information and the time series knowledge. Thus, it is suitable for this problem.

ACKNOWLEDGEMENTS

The work was supported by National Natural Science Foundation of China (61403208), Natural Science Foundation of Jiangsu Province (BK20161516), Scientific and Technological Support Project (Society) of Jiangsu Province (BE2016776), and Science Foundation of Nanjing University of Posts and Telecommunications (NY214014).

REFERENCES

[1] George E. Dissmeyer. 2000. *Drinking water from forests and grasslands: a synthesis of the scientific literature.* Department of Agriculture, Forest Service, Southern Research Station.

[2] Alex Graves and Jürgen Schmidhuber. 2005. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks* 18, 5-6 (2005), 602–610.

[3] Diederik P Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *the 3rd International Conference for Learning Representations.*

[4] Fitore Muharemi, Doina Logofătu, Christina Andersson, and Florin Leon. 2018. Approaches to Building a Detection Model for Water Quality: A Case Study. In *Modern Approaches for Intelligent Information and Database Systems.* Springer, 173–183.

[5] World Health Organization. 2004. *Guidelines for drinking-water quality: recommendations.* World Health Organization.

[6] Ying Zeng, Honghui Yang, Yansong Feng, Zheng Wang, and Dongyan Zhao. 2016. A convolution BiLSTM neural network model for Chinese event extraction. In *Natural Language Understanding and Intelligent Applications.* Springer, 275–287.

[7] Zhi-Hua Zhou and Xu-Ying Liu. 2006. Training cost-sensitive neural networks with methods addressing the class imbalance problem. *IEEE Transactions on Knowledge and Data Engineering* 18, 1 (2006), 63–77.