Trident: Change Point Detection for Multivariate Time Series via Dual-Level Attention Learning

Ziyi Duan, Haizhou Du(✉), and Yang Zheng

Shanghai University of Electric Power, Shanghai 200090, China
duhaizhou@shiep.edu.cn

Abstract. Change point detection is an important subset of anomaly detection problems. Due to the ever-increasing volume of time-series data, detecting change points has important significance, which can find anomalies early and reduce losses, yet very challenging as it is affected by periodicity, multi-input series, and long time series. The performance of traditional methods typically scales poorly.

In this paper, we propose Trident, a novel prediction-based change point detection approach via dual-level attention learning. As the name implies, our model consists of three key modules which are the prediction, detection, and selection module. The three modules are integrated in a principled way of detecting change points more accurately and efficiently. Simulations and experiments highlight the effectiveness and efficacy of the Trident for change point detection in time series. Our approach outperforms the state-of-the-art methods on two real-world datasets.

Keywords: Time series · Change point detection · Attention mechanism

1 Introduction

With the explosive development of big data analysis, anomaly detection in time-series is also increasingly important. Change point detection is an important subset of anomaly detection problems. Due to the ever-increasing volume of time-series data that must be efficiently analyzed, it is becoming a mainstream study in a wide variety of applications, including finance, energy, meteorology, medicine, aerospace, etc.

Change points are the moments when the state or property of the time series changes abruptly [2]. Increasing the detecting accuracy is beneficial to operational efficiency in many aspects of society [20,21], such as power load detection, online sales analysis, or weather forecasting. We could mine the potential mutations and take corresponding preventative measures early to reduce financial and time losses.

However, detecting change points in modern applications is particularly challenging, affected by the following complex factors: periodicity, multi-series, and

N. T. Nguyen et al. (Eds.): ACIIDS 2021, LNAI 12672, pp. 799–810, 2021.
https://doi.org/10.1007/978-3-030-73280-6_63

long series. Traditional methods cannot adaptively select relevant series and achieve feature extraction for multiple input series. And may lead to error accumulation and inefficient computation on the long-term series. Moreover, the real-world time-series data may contain a large number of change points and outliers. There are fundamental differences between them [22]. Among them, change points are the moments of the original series' property or state change abruptly. And outliers refer to the sudden single peak or decrease in the series [6]. Distinguishing them has always been one of the difficulties in change point detection.

The current methods of change point detection are mainly split up into probability and statistics-based, classification-based, and prediction-based. The traditional methods are ineffective in modeling complex non-linear time series data [16,23]. The prediction-based method is one of the most commonly used methods [19]. Recently, the deep learning-based approaches have demonstrated strong performance in time series modeling [4,10,17]. However, the existing methods only focus on improving the ability to learn nonlinear features, while they ignore the problem of feature and information loss.

To address these aforementioned issues, inspired by the multimodal features fusion [5,8] and the hierarchical attention networks [14,15], which are the latest progress of attention mechanisms [3,9], we proposed a prediction-based change point detection approach via dual-level attention learning, which we call Trident.

In this paper, Trident consists of three key modules: prediction, detection, selection. Accordingly, the key contributions can be summarized as follows.

- We propose Trident, a change point detection approach for time series employing dual-level attention learning. It could detect change points accurately and timely in long periodic series with multiple relevant input series.
- In the input attention stage, we integrate the novel multi-series fusion mechanism. To the best of our knowledge, this is the first time proposing the idea in the change point detection task, ensuring we can adaptively extract features from multivariate time series.
- In the temporal attention stage, in order to prevent error accumulation in long-series detection, we use the Bi-LSTM decoder to better capture the long-term temporal dependencies of the time series and improve accuracy.
- In the change point selection module, we propose a novel and simple algorithm. By setting the number of consecutive abnormal points in the series, we can identify change points and outliers, so that reduces the computation complexity and improves interpretability.

To demonstrate the effectiveness of Trident, we conducted experiments on two public datasets in different domains. Extensive experimental results show that our approach outperforms current state-of-the-art models on the two real-world datasets.

The remainder of the paper is organized as follows. We introduce the overview and the details of Trident in Sect. 2. Then we present the evaluation results and analyze the performance in Sect. 3. Lastly, we conclude our paper and sketch directions for the possible future work in Sect. 4.

2 Trident Design

The overall workflow of our approach is presented in Fig. 1. As we introduced in Sect. 1, Trident consists of three major modules.

We first predict the target series. Then, based on the deviation between the actual value and the predicted value, we determine the threshold and detect abnormal points. Lastly, we identify change points and outliers. The theories and details of the three core modules will be introduced below.

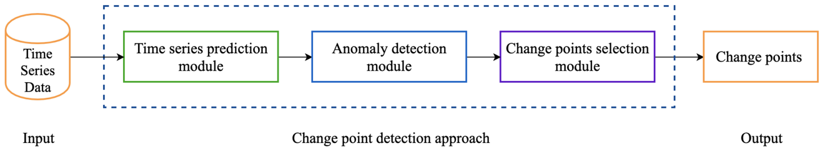


Fig. 1. The overall workflow of Trident

2.1 Time Series Prediction Module

In this module, we propose a novel dual-level attention-based approach for time series prediction. In the encoder, we employ a novel input attention mechanism and propose a multi-series fusion mechanism, which can adaptively learn the relevant input series and achieve feature extraction. In the decoder, a temporal attention mechanism is used to automatically capture the long-term temporal dependencies of the time series. The description of the proposed model is shown in Fig. 2.

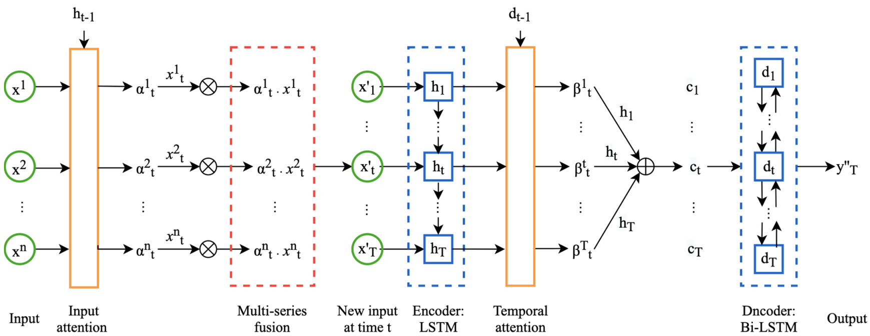


Fig. 2. The architecture of time series prediction module

Encoder with Input Attention. In our model, the encoder part is an LSTM, aiming to better capture the long-term dependencies of time series.

Given the input sequence $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T)$ with $\mathbf{x}_t \in \mathbb{R}^n$, where n is the number of input series, T is the length of window size or time steps. The encoder can be applied to learn a mapping from $\mathbf{x}_t$ to $\mathbf{h}_t$ (at time step t) with: $\mathbf{h}_t = f_1(\mathbf{h}_{t-1}; \mathbf{x}_t)$, where $h_t \in \mathbb{R}^p$ is the hidden state of the encoder at time t, p is the size of the hidden state, x_t is the input of each time step, and f_1 is a non-linear activation function. Since we use an LSTM unit as f_1, so we can summarize the update as follows: $\mathbf{h}_t = LSTM(\mathbf{h}_{t-1}; \mathbf{x}_t)$.

We propose an input attention mechanism that can adaptively extract features from the relevant input series through multi-series fusion, which has practical meaning. Given the k-th input series $\mathbf{x}^k = (x_1^k, x_2^k, \ldots, x_T^k)^T \in \mathbb{R}^T$, we construct the input attention mechanism by referring to the previously hidden state $\mathbf{h}_{t-1}$ and the cell state $\mathbf{s}_{t-1}$ in the encoder LSTM unit. So the input attention weight is computed as follow:

$$e_t^k = \boldsymbol{\omega}_e^T tanh(\mathbf{W}_e[\mathbf{h}_{t-1}; \mathbf{s}_{t-1}] + \mathbf{U}_e \mathbf{x}^k + \mathbf{b}_e) \tag{1}$$

and

$$\alpha_t^k = \frac{\exp(e_t^k)}{\sum_{i=1}^{n} \exp(e_t^i)} \tag{2}$$

where $\mathbf{W}_e \in \mathbb{R}^{T \times 2p}$ and $\mathbf{U}_e \in \mathbb{R}^{T \times T}$ are matrices, $\boldsymbol{\omega}_e \in \mathbb{R}^T$ and $\mathbf{b}_e$ are vectors. They are parameters to learn.

The attention weight measures the importance of the k-th input feature (driving series) at time t. A softmax function is applied to α_t^k to ensure all the attention weights sum to 1.

Multi-series Fusion Mechanism. Based on the attention weights, we propose a multi-series fusion mechanism that aims to fuse different input series, that is, different periods of history, to better extract features.

Specifically, first, we divide the time series into multiple sub-series. Each sub-series represents a complete period, and multiple relevant sub-series are used as input to our model. Then at each future time step, with the attention weights calculated by Eq. (2), we combine them by learning the relative importance of each series. Therefore, we adaptively extract the most relevant input features, achieve multi-series fusion, and get new input at time t.

$$\mathbf{x}_t^{'} = (\alpha_t^1 x_t^1, \alpha_t^2 x_t^2, \ldots, \alpha_t^n x_t^n)^T \tag{3}$$

Then the hidden state at time t can be updated as:

$$\mathbf{h}_t = f_1(\mathbf{h}_{t-1}; \mathbf{x}_t^{'}) \tag{4}$$

where f_1 is an LSTM unit with $\mathbf{x}_t$ replaced by the newly computed $\mathbf{x}_t^{'}$.

With the proposed input attention mechanism and multi-series fusion mechanism, we can selectively focus on certain driving series rather than treating all the input driving series equally.

Decoder with Temporal Attention. Following the encoder with input attention, a temporal attention mechanism is used in the decoder to adaptively select relevant encoder hidden states across all time steps. Specifically, the attention weight of each encoder hidden state at time t is calculated based upon the previous decoder hidden state $\mathbf{d}_{t-1} \in \mathbb{R}^p$ and the cell state of the LSTM unit $\mathbf{s}^{'}_{t-1} \in \mathbb{R}^p$ with:

$$v_t^i = \boldsymbol{\omega}_v^T tanh(\mathbf{W}_v[\mathbf{d}_{t-1}; \mathbf{s}^{'}_{t-1}] + \mathbf{U}_v \mathbf{h}_i + \mathbf{b}_v) \tag{5}$$

and

$$\beta_t^i = \frac{\exp(v_t^i)}{\sum_{j=1}^{T} \exp(e_t^j)} \tag{6}$$

where $\mathbf{W}_v \in \mathbb{R}^{p \times 2q}$ and $\mathbf{U}_v \in \mathbb{R}^{p \times p}$ are matrices, $\boldsymbol{\omega}_v \in \mathbb{R}^p$ and $\mathbf{b}_v$ are vectors. They are parameters to learn.

Since each encoder hidden state $\mathbf{h}_i$ is mapped to a temporal component of the input, the attention mechanism computes the context vector $\mathbf{c}_t$ as a weighted sum of all the encoder hidden states $\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_T$:

$$\mathbf{c}_t = \sum_{i=1}^{T} \beta_t^i \mathbf{h}_i \tag{7}$$

After getting the weighted summed context vectors, we can combine them with the target series:

$$\tilde{y}_t = \tilde{\boldsymbol{\omega}}^T [y_t; \mathbf{c}_t] + \tilde{\mathbf{b}} \tag{8}$$

The newly computed $\tilde{y}_t$ can be used for the update of the decoder hidden state at time t.

In our approach, a bi-directional LSTM (Bi-LSTM) is used as the decoder backbone. Because the traditional unidirectional LSTM will ignore the dynamic future information, which could have a strong influence on the time series forecast in practice. Therefore, the Bi-LSTM decoder aims to prevent error accumulation and improve accuracy for long-horizon forecasting.

It is composed of two LSTMs that allow both backward (past) and forward (future) dynamic inputs to be observed at each future time step. Then the hidden states of Bi-LSTM are fed into a fully-connected layer to produce final predictions. Formally, we define the formulations as follows:

$$\mathbf{d}_t = BiLSTM(\tilde{y}_t; \mathbf{d}_{t-1}, \mathbf{d}_{t+1}) \tag{9}$$

After updating the hidden state of the decoder, our model produces the final prediction result, denoted as $\hat{y}_t$.

Moreover, in the training procedure, we use the minibatch stochastic gradient descent (SGD) together with the adaptive moment estimation (Adam) optimizer to optimize parameters. We implemented our approach in the TensorFlow framework.

2.2 Anomaly Detection Module

After the final prediction result $\hat{y}_t$ is produced by the time series prediction module, the predicted values of the target series are delivered to the anomaly detection module. We calculate the deviation between the actual and the predicted value as $l_t = y_t - \hat{y}_t$, where y_t is the actual value, and $\hat{y}_t$ is the predicted value.

The absolute value of l_t is used as the anomaly score, denoted as e_t. The larger the anomaly score, the more significant the anomaly at the given time step. Therefore, we need to define a threshold based on the target series for anomaly classification.

In our approach, we adopt the Gaussian distribution-based method to determine the threshold. Extensive results in [13] show that anomaly scores fit the Gaussian distribution very well in a range of datasets. This is due to the fact that the most general distribution for fitting values derived from Gaussian or non-Gaussian variables is the Gaussian distribution according to the central limit theorem. Motivated by this, we define the following parameters based on Gaussian distribution:

$$(e_1, e_2, \ldots, e_n) \sim \boldsymbol{N}(\mu; \sigma^2) \tag{10}$$

$$\mu = \frac{1}{n}\sum_{t=1}^{n} e_t; \sigma = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(e_t - \mu)^2} \tag{11}$$

Based on the Pauta criterion, we determine the range of the threshold according to the proportion of normal points in the dataset. Then we determine e as the threshold when the performance is optimal.

2.3 Change Points Selection Module

After the anomaly detection module classifies the data in the target series, this module will select the change points from the anomaly points. We propose a novel and simple approach to distinguish change points and outliers.

We have discussed the discrepancy between change points and outliers and the current research status in Sect. 1. In our approach, we use a simple strategy. We use the series containing the abnormal mark as input to this module. Once an abnormal point occurs, we declare it as an outlier. If a certain number of outliers are continuously found, the first point of this outlier series is declared as a change point. Thus, this method could reduce the computation complexity and improve interpretability. We define the parameter N (we set $N = 3$, which can achieve the best performance), which represents the minimum number of consecutive outliers.

3 Experiment and Evaluation

Based on the above approach, we designed the following experiment scheme. In this section, we conducted experiments on the three modules of the proposed method. Extensive results on two large real-world datasets show the effectiveness and superiority of Trident.

3.1 Experimental Environments

All the experiments were executed on a single computer with the following specifications: 3.1 GHz Intel Core i5 CPU, 16 GB 2133 MHz LPDDR3.

The software details are shown below: Model implementation: TensorFlow 1.13.2, Keras 2.1.0; Operating System: macOS Catalina 10.15.3.

3.2 Datasets Introduction

To demonstrate the effectiveness of Trident, we conducted experiments on two public datasets in different domains: The load Forecasting dataset and the Air-Quality dataset.

- **Load Forecasting Dataset:**[1]
  The dataset is hourly load data collected from utilities in 20 zones of the United States. The time period is from 2004 to 2008. It contains 33,000 instances with 29 attributes. This dataset was used in the Global Energy Forecasting Competition held in 2012 (GEFCom 2012) [7].
  We chose 4 areas of similar magnitude to conduct experiments. We took the power load in 2007 as the target series and the power load in 2004–2006 as the three relevant input series.
- **Air-Quality Dataset:**[2]
  The second dataset is the Air-Quality dataset, which can be found in the UCI Machine Learning Repository. It includes hourly data of 6 main air pollutants and 6 relevant meteorological variables at 12 air-quality monitoring sites in Beijing. This dataset covers the period from March 1st, 2013 to February 28th, 2017, and contains 420,768 instances with 18 attributes.
  We chose the ambient temperature as the prediction object. We used the 2016 data as the target series and the 2013–2015 data as the three relevant input series.
  The data in the dataset contains negative numbers and zeros, where zero will cause the metric MAPE to be unable to calculate. Therefore, we converted the representation of temperature in the dataset from degree Celsius to thermodynamic temperature: $T(K) = t(°C) + 273.15$, where T is thermodynamic temperature, and t is degree Celsius.

3.3 Results 1: Prediction Performance

In this section, we demonstrated the predicting performance of the time series prediction module of Trident, and proved that our approach can solve the three challenges we proposed well.

[1] https://www.kaggle.com/c/global-energy-forecasting-competition-2012-load-forecasting/data.

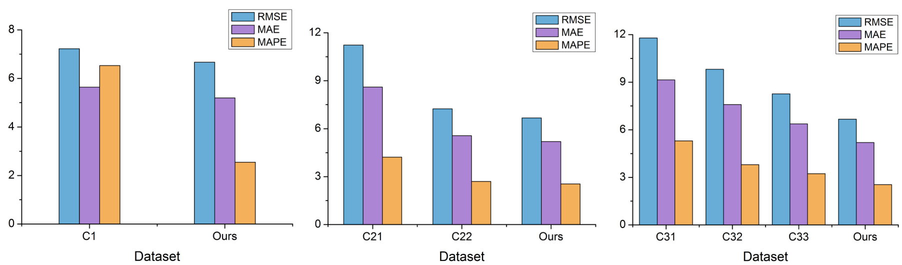
[2] http://archive.ics.uci.edu/ml/datasets/Beijing+Multi-Site+Air-Quality+Data.

Competing Methods and Evaluation Metrics. To demonstrate the effectiveness of Trident in predicting performance, we compared it against four baseline methods: Bi-LSTM [18], attention mechanism (Attn) [3], BiLSTM-Attn (the setting of Trident that does not employ the input attention mechanism), and DA-RNN (a dual-stage attention-based recurrent neural network) [15].

Moreover, to measure the effectiveness of the time series prediction module, we considered the three evaluation metrics: root mean squared error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE).

Results and Analysis. Firstly, we used the Load Forecasting dataset to verify the effectiveness of our approach for the proposed challenges.

For each experiment about the challenge, we set an experimental group and control groups. The experimental group uses 4 zones of data for 4 years (contains 3 input series and 1 target series). For the periodicity experiment, the control group C1 has no periodicity, predicting the data of 1 zone through the data of 3 zones. Similarly, for the long series experiment, the 3 control groups are denoted as C21–C23. Each sub-dataset use one-zone, two-zone, and three-zone data, respectively (with different length of the series in the sub-dataset). And for multi-series experiments, the control groups are denoted as C31 and C32, using two-year (1 input series) and three-year (2 input series) data, respectively. Other parameter settings are the same. The experimental results are shown in Fig. 3.



(a) Periodicity (b) Multi-series (c) Long series

Fig. 3. The comparative experiment results for the three proposed challenges.

Figure 3 shows that, with the increase of series length and the number of input series, or the improvement of data's periodicity, the three metrics have improved significantly. Trident shows good performance in solving each proposed challenge. Therefore, our approach has obvious advantages for the problems of long, periodic, and multiple input series.

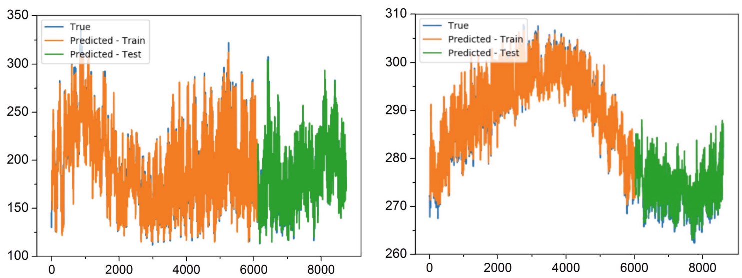
Next, to measure the predicting performance of Trident, we conducted extensive experiments on Trident and all baseline methods on two real-world datasets. Table 1 summarizes the results.

Table 1. The predicting performance of Trident and all baseline methods.

Models	Load forecasting dataset			Air-quality dataset		
	RMSE	MAE	MAPE (%)	RMSE	MAE	MAPE (%)
Bi-LSTM	23.894	18.853	9.100	8.663	7.904	2.877
Attn	16.678	13.290	6.333	8.394	7.672	2.791
BiLSTM-Attn	14.432	10.806	5.336	8.286	7.561	2.746
DA-RNN	11.007	8.712	4.188	4.092	3.411	1.241
Ours	**6.665**	**5.196**	**2.547**	**3.239**	**2.615**	**0.950**

Table 1 shows that our approach outperforms the other baseline models markedly on the two datasets. The RMSE, MAE, and MAPE values improved by 40% and 22% on average on the two datasets, respectively. These experimental results demonstrated the superiority of Trident compared with the state-of-the-art methods.

For further comparison, we showed the prediction results of Trident on the two datasets in Fig. 4. The blue, orange, and green line represents actual value, training part, and test part respectively. We can observe that our approach performs well in prediction performance.



(a) Load forecasting dataset (b) Air-quality dataset

Fig. 4. The final prediction results of Trident over the two datasets.

This is because Trident integrated the dual-level attention mechanism and multi-series fusion mechanism, which can adaptively select relevant input series and extract features. Meanwhile, it employs a Bi-LSTM decoder to solve the problem of error accumulation in long-term series, thereby greatly improves the predicting accuracy.

Overall, extensive experiments show the effectiveness and superiority of Trident. It can comprehensively solve the three proposed challenges and shows good predictive performance.

3.4 Results 2: Change Point Detection Performance

In this section, we proved the effectiveness of the anomaly detection module and the change points selection module. We demonstrated the superiority of Trident by comparing it with the other three state-of-the-art approaches.

Competing Methods and Evaluation Metrics. We compared Trident against three baseline methods: CNN-LSTM (classification-based method) [12], Bayesian online change point detection (BOCPD) (probability method) [1], and KLIEP (an online density-ratio estimation algorithm) [11].

Moreover, we evaluated the efficiency of the change point detection module based on the following metrics: Precision, Recall, and F1 Score.

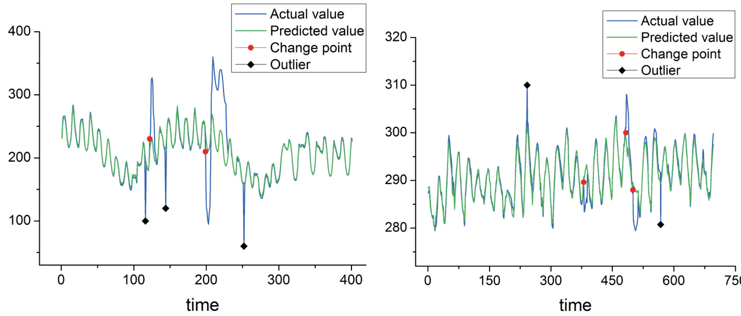
Results and Analysis. To measure the change point detecting performance of Trident, we conducted extensive experiments on Trident and other baseline methods on the two real-world datasets. Table 2 and Fig. 5 summarize the results.

Table 2. The change point detecting performance of Trident and all baseline methods.

Models	Load forecasting dataset			Air-quality dataset		
	Precision	Recall	F1	Precision	Recall	F1
KLIEP	0.963	0.887	0.924	0.975	0.816	0.888
BOCPD	0.967	0.934	0.950	0.976	0.854	0.911
CNN-LSTM	0.964	0.955	0.959	**0.977**	0.903	0.938
Ours	**0.989**	**0.985**	**0.987**	0.973	**0.977**	**0.975**

Table 2 shows that on the two real-world datasets, our approach outperforms the other baseline methods in performance. And as depicted in Fig. 5, we showed the change points detection results of Trident over the two datasets. We selected some periods of the target series that contain anomalies as examples. We can easily observe that performance of Trident in detecting change points is well. Trident can detect change points accurately and timely on the two datasets, and can identify change points and outliers.

In conclusion, Trident is a prediction-based change point detection approach. It employs the dual-level attention learning for time series prediction. And we used the Gaussian distribution-based method for the anomaly detection tasks, and combine the change point selection module. For the periodic long-term series and multiple relevant input series, our approach can detect change points accurately, and identify change points and outliers. Extensive experimental results prove the superiority of Trident compared with the state-of-the-art methods.



(a) Load forecasting dataset (b) Air-quality dataset

Fig. 5. The change point detection results of Trident on real-world datasets.

4 Conclusion

In this paper, we propose Trident, a novel change point detection approach in time series via dual-level attention learning. It consists of three key modules: time series prediction module, anomaly detection module, and change point selection module. In the time series prediction module, we use a dual-level attention learning model and integrate the multi-series fusion mechanism. It can adaptively extract features of input series. In the anomaly detection and change point selection module, we determine the threshold employing the Gaussian distribution-based method and identify change points and outliers. We verified the effectiveness of Trident on two public real-world datasets. Extensive experimental results show that our approach outperforms the state-of-the-art methods. In future work, we will further extend our approach to handle multivariate time series from different data sources. Due to the heterogeneity of data sources together with limited information about their interactions, exploring how to learn the complex dynamic correlations deserves our in-depth study.

References

1. Adams, R.P., MacKay, D.J.: Bayesian online changepoint detection. Stat. Optim. Inf. Comput. **1050**, 19 (2007)
2. Aminikhanghahi, S., Cook, D.J.: A survey of methods for time series change point detection. Knowl. Inf. Syst. **51**(2), 339–367 (2017)
3. Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. In: 3rd International Conference on Learning Representations, ICLR 2015 (2015)
4. Canizo, M., Triguero, I., Conde, A., Onieva, E.: Multi-head CNN-RNN for multi-time series anomaly detection: an industrial case study. Neurocomputing **363**, 246–260 (2019)
5. Fan, C., et al.: Multi-horizon time series forecasting with temporal attention learning. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 2527–2535 (2019)

6. Fearnhead, P., Rigaill, G.: Changepoint detection in the presence of outliers. J. Am. Stat. Assoc. **114**(525), 169–183 (2019)
7. Hong, T., Pinson, P., Fan, S., Zareipour, H., Troccoli, A., Hyndman, R.J.: Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond. Int. J. Forecast. **32**, 896–913 (2016)
8. Hori, C., et al.: Attention-based multimodal fusion for video description. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 4193–4202 (2017)
9. Hübner, R., Steinhauser, M., Lehle, C.: A dual-stage two-phase model of selective attention. Psychol. Rev. **117**(3), 759 (2010)
10. Hundman, K., Constantinou, V., Laporte, C., Colwell, I., Soderstrom, T.: Detecting spacecraft anomalies using LSTMs and nonparametric dynamic thresholding. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 387–395 (2018)
11. Kawahara, Y., Sugiyama, M.: Sequential change-point detection based on direct density-ratio estimation. Stat. Anal. Data Min. **5**(2), 114–127 (2012)
12. Kim, T.Y., Cho, S.B.: Web traffic anomaly detection using C-LSTM neural networks. Exp. Syst. Appl. **106**, 66–76 (2018)
13. Kriegel, H.P., Kroger, P., Schubert, E., Zimek, A.: Interpreting and unifying outlier scores. In: Proceedings of the 2011 SIAM International Conference on Data Mining, pp. 13–24. SIAM (2011)
14. Liang, Y., Ke, S., Zhang, J., Yi, X., Zheng, Y.: GeoMAN: multi-level attention networks for geo-sensory time series prediction. In: IJCAI, pp. 3428–3434 (2018)
15. Qin, Y., Song, D., Cheng, H., Cheng, W., Jiang, G., Cottrell, G.W.: A dual-stage attention-based recurrent neural network for time series prediction. In: Proceedings of the 26th International Joint Conference on Artificial Intelligence, pp. 2627–2633 (2017)
16. Sadouk, L.: CNN approaches for time series classification. In: Time Series Analysis-Data, Methods, and Applications. IntechOpen (2018)
17. Salinas, D., Flunkert, V., Gasthaus, J., Januschowski, T.: DeepAR: probabilistic forecasting with autoregressive recurrent networks. Int. J. Forecast. **36**, 1181–1191 (2019)
18. Schuster, M., Paliwal, K.K.: Bidirectional recurrent neural networks. IEEE Trans. Sig. Process. **45**(11), 2673–2681 (1997)
19. Sezer, O.B., Gudelek, M.U., Ozbayoglu, A.M.: Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. Appl. Soft Comput. **90**, (2020)
20. Su, Y., Zhao, Y., Niu, C., Liu, R., Sun, W., Pei, D.: Robust anomaly detection for multivariate time series through stochastic recurrent neural network. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 2828–2837 (2019)
21. Xu, H., et al.: Unsupervised anomaly detection via variational auto-encoder for seasonal KPIs in web applications. In: Proceedings of the 2018 World Wide Web Conference, pp. 187–196 (2018)
22. Zhang, A.Y., Lu, M., Kong, D., Yang, J.: Bayesian time series forecasting with change point and anomaly detection (2018)
23. Zhang, A., Paisley, J.: Deep Bayesian nonparametric tracking. In: International Conference on Machine Learning, pp. 5833–5841 (2018)