International Core Journal of Engineering Volume 6 Issue 9, 2020
ISSN: 2414-1895 DOI: 10.6919/ICJE.202009_6(9).0024

A Summary of Research on Deep Learning in Time Series Learning Methods

Rong Cao

School of Shanghai Maritime University, Shanghai 201306, China.

Abstract

In recent years, deep learning has developed rapidly, and end-to-end applications in the field of time series analysis are becoming more and more mature. Therefore, this article discusses the time series learning method based on deep learning, and summarizes the latest deep learning methods for task representation, prediction, classification, and anomaly detection of time series from the aspects of application, architecture, and ideas. The in-depth study of learning solutions and future development trends provide a reference.

Keywords

Deep learning, Time series prediction, Time series classification.

1. Introduction

In order to understand the development law of a certain thing, people usually observe a certain index of the thing at a certain sampling frequency. This sequence of recording data points in chronological order is called a time series. A common example is the daily fluctuation of the stock market. , Power consumption records, etc. It is of great academic value and practical significance to study how to characterize these data and effectively mine potential and valuable information from massive time series data. In recent years, with the rapid development of artificial neural networks, the end-to-end method has been used more and more in the field of time series learning, and the overall method has become more mature.

Deep learning stems from the research of artificial neural networks, and its development is largely the development of various neural network models that are more complex and reasonable above artificial neural networks, so time series learning based on deep learning is actually Application of a neural network model on time series problems.Deep learning methods have solved a lot of practical problems in academia and industry, such as robot processing [1], object recognition [2], speech and handwritten font recognition [3], and simultaneous translation [4].

As a complex data object with great mining value, time series widely exists in various fields. Streaming media file data, financial and financial data, meteorological observation data, census data, system log data, etc. are all forms of time series. . Time series data in different fields contains different analytical values: for example, doctors can understand the abnormality of patients' sleep patterns by analyzing patient's sleep data, and economists can predict stock market trends by analyzing stock market data. Different time series problems are solved by different methods, which can be summarized into three methods: classification, anomaly detection and prediction.

The structure of this article is: the first section introduces the theoretical basis of time series; the second section introduces the deep learning model commonly used for time series learning; the third section introduces the learning method of deep learning-based time series, including the time series Feature representation, prediction methods, classification methods, anomaly detection methods;

ISSN: 2414-1895 DOI: 10.6919/ICJE.202009_6(9).0024

finally, in Chapter 5, the application direction of deep learning in the field of time series is summarized and predicted.

2. Theoretical basis of time series

2.1 Definition of time series

Time series data is a collection of recorded data values generated according to the progress of the time axis. The interval between time series data values is a time scale, which can be seconds, minutes, hours, days, months, and years. It can be seen from the definition of time series that time series has two basic elements of time and value. In order to describe the time series more accurately, this article gives the definition of univariate time series and multivariate time series from the mathematical point of view.This section uses X_{uni} and X_{mul} to denote univariate time series and multivariate time series, respectively, to denote time steps. The definitions of X_{uni} and X_{mul} are shown as formula (1) and formula (2) respectively:

$$X_{uni} = \{\mathrm{p}_1, \mathrm{p}_2, ..., \mathrm{p}_\mathrm{T}\} \tag{1}$$

$$X_{mul} = \{\vec{\mathrm{q}}_1, \vec{\mathrm{q}}_2, ..., \vec{\mathrm{q}}_\mathrm{T}\} \tag{2}$$

For the $i(i=1,2,...,T)^{\mathrm{th}}$ component p_i of X_{uni}, p_i represents the observed value of the observed variable at the instant.

For the $i(i=1,2,...,T)^{\mathrm{th}}$ component $\vec{q}_i$ of X_{mul}, $\vec{q}_i$ represents the vector set of observed value of the observed variable at the instant.

Each point in the time series can be marked as TP, FN, FP, TN according to the real situation and the predicted situation, then the evaluation criteria accuracy rate precision and recall calculation method are shown in formula (3):

$$precision = \frac{TP}{TP+FP}, recall = \frac{TP}{TP+FN} \tag{3}$$

F1 Score is another index to measure the accuracy of the binary classification model. It takes into account both precision and recall. It is the average of the two. The calculation method is shown in formula (4):

$$F1 = 2 \times \frac{precision \times recall}{precision + recall} \tag{4}$$

2.2 Characteristics of time series

(1) Each data point in the time series changes with time, and the value of each data point is the result of the combined action of different factors that affect the change at the same time;

(2) From a local point of view, the data points on each time node are random, but there is a certain correlation between the data points, this correlation reflects the inherent operation of the observed object or system;

(3) On the whole, the changes of data points in the time series are often followed by rules, such as showing trend, periodicity and comprehensiveness. Trendiness means that the sequence changes show a slow and long-term upward or downward trend. The reason for the periodicity is the periodic influence of external factors on the system, such as the season; comprehensive means that the sequence change is the result of a combination of factors;

ISSN: 2414-1895 DOI: 10.6919/ICJE.202009_6(9).0024

(4) The past and current data points in the time series may indicate the development trend of future activities, and there will be no sudden jump changes, and it will move forward in a relatively small step.

3. Introduction of deep learning models

Through the introduction of the characteristics of time series in Section 2.2, we can see that there must be some regularity behind the change of time series data point values, and this regularity may be the result of a combination of factors. A neural network is a mathematical model that organizes related elements through a "layer" structure. Its geometric structure and functional properties are inspired by the human brain. Through neural network model training to find the regularity of time series numerical changes, so as to achieve the purpose of prediction or classification.

3.1 BP neural network

A neural network is a mathematical model that organizes related elements through a "layer" structure. Its geometric structure and functional properties are inspired by the human brain. Backpropagation neural network is a concept proposed by scientists led by Rumelhart and McClelland in 1986. It is a multi-layer feedforward neural network trained according to the error backpropagation algorithm[5]. BP neural network does not need to determine the mapping relationship between input and output in advance, but learns some rules through training. Eventually, when a certain input value is given, the neural network can output a result close to the actual expected value. The BP neural network is composed of an input layer, a hidden layer, and an output layer. The input layer and the output layer generally have only one layer. The hidden layer may be one layer or multiple layers. A classic three-layer BP neural network structure is shown in Figure 1, from which the gradient of the error function is calculated and the weights are set iteratively to minimize the error.

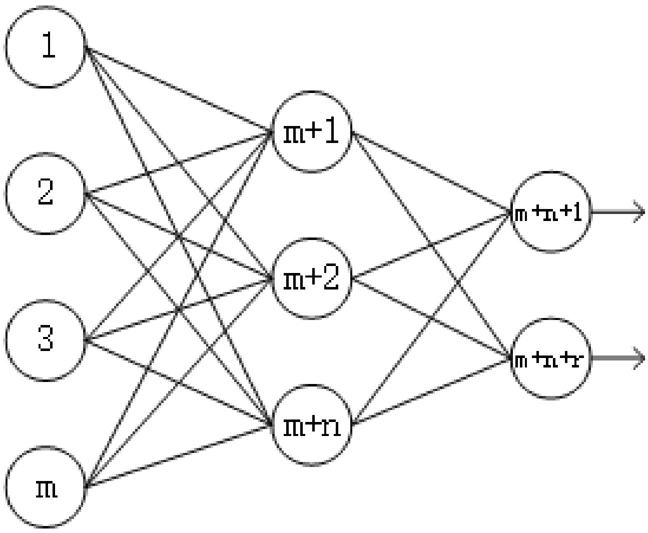

Figure 1. BP neural network structure

3.2 Convolutional Neural Network

Convolutional neural network(CNN) [6] is a neural network designed specifically for image recognition problems, and is also implemented by a back propagation algorithm. Convolutional neural networks usually use multiple channels and multiple convolution kernels to extract historical information of the data. For time series data, different features correspond to different channels of convolution kernels. Its sequence modeling ability has always been considered weak, but in 2016, Google proposed the WaveNet model in voice data processing, introducing hollow causal convolution, jump connection and residual network, adding more hidden layers to increase the experience , So that it has excellent results in sequence modeling [7].

ISSN: 2414-1895 DOI: 10.6919/ICJE.202009_6(9).0024

3.3 Recurrent neural network

Recurrent neural network (RNN) is a network structure based on chain structure to achieve parameter sharing. Traditional neural networks can only process inputs one by one, and there is no relationship between the previous input and the next input. However, many information in the real world have a strong temporal correlation, such as when people understand a sentence It is often necessary to combine the context of that word with a word in. RNN solves the problem that traditional neural networks cannot model time-dependent inputs.In recent years, scholars have used the RNN model to realize the application of time series prediction in weather, power load, website traffic, traffic flow, sales, etc. [8-11].

3.4 LSTM (Long Short-Term Memory Network)

Recurrent neural networks are very powerful dynamic systems, but training them has proven to be problematic because the gradient of backpropagation either grows or shrinks at each time step, so at many time steps they usually explode or disappear.

Long Short Term Memory Network (LSTM) was first designed to solve the problem of gradient disappearance or explosion of RNN, which was later widely used in various applications and solved many problems. Compared with general RNN, LSTM avoids the long-term dependence problem through deliberate design. LSTM naturally stores long-term information without paying a very high price. LSTM is different from a single neural network layer. There are four neural network layers that interact in a very special way. LSTM has the ability to remove or add information to the state of the cell through a carefully designed structure called a "gate". The design of three kinds of thresholds (forgetting threshold, input threshold and output threshold) solves the difficulty of long sequence learning, gradient disappearance and gradient explosion in RNN. The network structure of LSTM is shown in Figure 2.

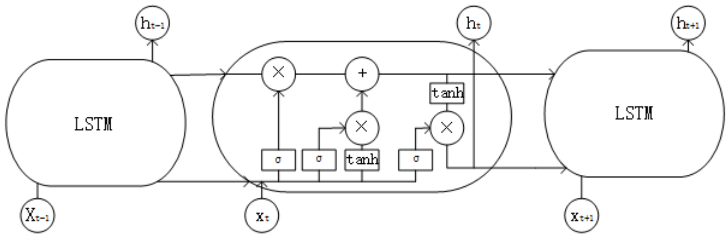

Figure 2. LSTM network structure

4. Time series analysis method based on deep learning

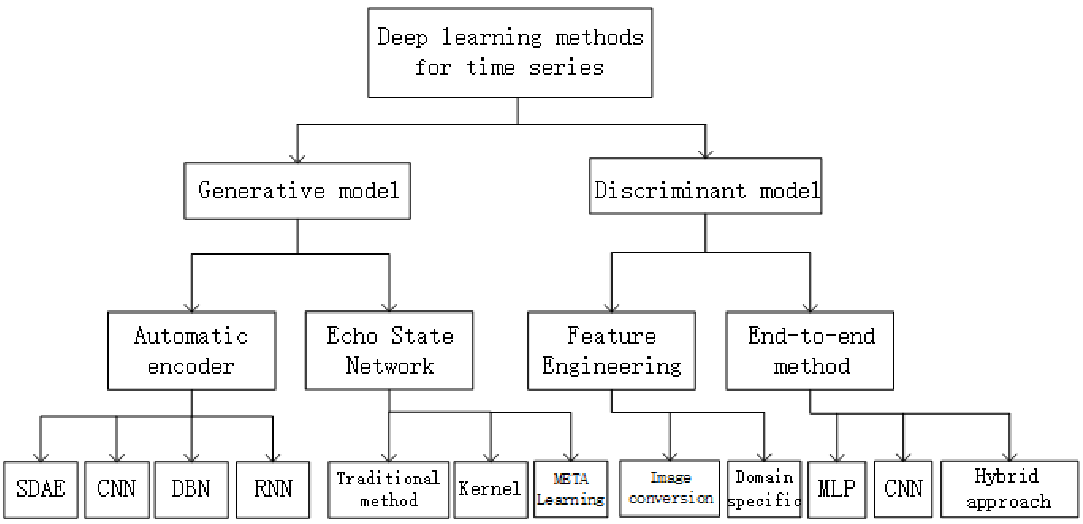

Figure 3. Classification of deep learning in time series analysis methods

ISSN: 2414-1895 DOI: 10.6919/ICJE.202009_6(9).0024

This section summarizes recent deep learning methods that analyze multiple aspects of multiple tasks in time series. Figure 3 shows the classification of deep learning methods under the time series analysis method.

4.1 Feature representation of time series based on deep learning

The time series feature representation refers to how to effectively represent the time series data before the actual data mining, and to describe the original time series data in another data form, requiring the new data to represent the key information of the original time series in the new data space , The results of feature representation can generally project high-dimensional and complex time series data to a lower-dimensional feature space. The low-dimensional feature representation with original time series feature information can complete time series data mining faster and more efficiently jobs.

Because time series has the characteristics of large data volume and high dimension, time series data mining requires a lot of computing resources, and it faces severe tests in terms of storage space and computing time. On the other hand, the time series obtained through practical applications in real life usually carry a lot of noisy data and outliers, which will seriously affect the performance of data mining. Therefore, before the actual implementation of data mining, you usually need to characterize the acquired time series data, thereby reducing the dimension of the time series data and reducing the noise and outliers in the data.

Chen et al. [12] proposed a spatiotemporal deep belief network (ST-DBN) for unsupervised learning of unchanging spatiotemporal features from video. Our method is based on the previous deep learning method and uses the convolution-limited Boltzmann machine (CRBM) as the basic processing unit. It alternates the aggregation of spatial and temporal information so that higher layers can capture longer-range statistical dependencies in space and time. Geoffrey E. Hinton [13] et al. extended RBM into a model TRBM capable of processing time series data by adding autoregressive connections between the visible and hidden layers in the previous time series, and obtained Bine Markov through experiments. The model's more powerful feature representation capabilities. Takayuki Katsuki [14] and others originally proposed a method of fuzzy time stamping event time series, that is, time-discounting convolution. The method uses time convolution mechanisms with specific parameters to process them, effectively expressing event dependencies in a time-shift-invariant manner, while overcoming the effects of past events. Z decouples static input features from cyclic sequence features, The unsupervised time series feature representation learning for sequence prediction problems is studied, that is, giving beautiful input samples of previous history for high-dimensional input sequences. The article introduces three models based on generative stochastic networks (GSN) for unsupervised sequence learning and prediction.

4.2 Time series prediction method based on deep learning

The traditional time series forecasting method is based on statistics, and has gradually evolved from modeling univariate, homoscedastic and linear time series to modeling multivariable, heteroscedastic and non-linear time series.

The traditional time series forecasting method has a certain forecasting ability. However, it does not consider the causal relationship between external influencing factors and the predicted object, but instead combines the results of all the influencing factors into time, and only predicts from the time dimension. Therefore, when some influencing factors change greatly, the prediction accuracy of traditional time series prediction methods will be greatly reduced.

In recent years, researchers have often combined new theories and methods with existing neural network models to achieve better prediction results. Reference [15] uses an attention mechanism-based LSTM for stock price prediction. The introduced attention mechanism can assign corresponding weights to inputs in different time periods. Through the distribution of weights, the model pays more attention to the input that is more related to the predicted target, which greatly improves the prediction effect of the model。

Literature [16] uses Deep Belief Network (DBN) with Restricted Boltzmann Machines (RBM) to deal with time series prediction problems; literature [17] compares DBN and integrated denoising The performance difference of self-encoders (Stacked Denoising Auto-encoders) in time series prediction; literature [18] uses deep networks to predict indoor temperature; literature [19] uses deep networks to predict traffic flow. The deep learning models in these literatures are just simple and direct applications, and do not combine the nature of time series data. The literature [20] makes a good summary of the problems of using deep learning to complete various time series predictions. In addition to using data to train new deep learning network structures to complete prediction tasks, you can also modify existing deep learning models to achieve prediction goals. Borovykh et al. modified the existing deep learning model WaveNet to realize the use of the relationship between a wide range of historical data and sequences.

4.3 Time series classification method based on deep learning

Deep learning methods have been widely used in natural language processing, speech recognition and other fields. At the same time, researchers have noticed that the two types of data, natural language and speech, are related in time, which is also the main feature of time series data. Therefore, in recent years, deep learning methods have been continuously used to solve time series classification problems.Cui Z et al. [21] first introduced deep learning methods in time series classification tasks. The article innovatively considers the characteristics of time series that usually have different time scales, and proposes an end-to-end neural network model-multi-scale volume Product Neural Network (MCNN) combines feature extraction and classification in one framework. Wang Z et al. [22] used deep learning to provide a simple but powerful baseline model for time series classification tasks.The model is implemented purely end-to-end and does not require tedious preprocessing operations on the original data. It provides a good starting point for future deep learning algorithms in time series. At the end of the article, the generalization of the model, network structure and classification semantics are also discussed. Wang Z et al. [23] introduced an attention mechanism on the basis of deep learning encoder to automatically learn the characteristics of time series for classification.Mariusz Zebik et al. proposed a convolutional neural network for time series classification. Unlike other feature-based classification methods, CNN can discover and extract suitable internal structures, and automatically generate deep features of the input time series by using convolution and pooling operations. And experiments show that this method is superior to other methods in terms of classification accuracy and noise tolerance.TWIESN (Time Warping Invariant Echo State Network) is a variant that can be used for time series classification on the basis of ESN. ESN was originally proposed for time series prediction. After modification, the new network structure can directly use the original input time series and predict Probability distribution of class variables. The latest progress in time series classification tasks systematically discusses time series classification based on deep transfer learning for the first time.

4.4 Time series anomaly detection method based on deep learning

In actual engineering, most time series data sets have data skew. For example, many companies use various sensors to continuously monitor various key performance indicators of the production process. In this case, the automatic detection of abnormal behavior in the collected data may have a great impact. Hawkin gave the definition of time series anomalies, which refers to data that deviates from most normal data in a large data set, which makes people suspect that these data are generated by different mechanisms rather than random of deviation.

Time series anomaly detection tasks have many similarities with classification and prediction tasks. For example, [24] converts anomaly detection tasks into classification tasks. Anomaly detection is easily transformed into a time series forecasting task, and in the case of a given model, it is an intuitive idea to treat the part where the difference between the predicted value and the real value is too large as an outlier. Using a Stacked LSTM-based network for time series anomaly detection, the idea is to determine the confidence of the anomaly by the difference between the predicted value and the true value. The network is first trained under normal data and used as a predictor for multiple time steps,

ISSN: 2414-1895 DOI: 10.6919/ICJE.202009_6(9).0024

and then the resulting prediction error is modeled as a multivariate Gaussian distribution for evaluating the possibility of abnormal behavior. In order to solve the aforementioned problem of insufficient instances of time series anomalies, Vincent Vercruyssen et al. discussed the use of transfer learning in time series anomaly detection tasks. The algorithm proposed in this paper attempts to transfer the labeled examples from the source domain to the target domain where no label is available. This method uses unusual and unexpected insights to decide whether to transfer tag instances from the source domain to the target domain. After the transfer is completed, we construct the nearest neighbor classifier in the target domain and use dynamic time warping as the similarity measure. Finally, experimental evaluations on many real-world data sets show that the overall method is promising and that it is superior to unsupervised anomaly detection in the target domain. Yahoo [25] proposes a general and extensible framework for automatic anomaly detection of large-scale time series data. Being able to detect system anomalies at an early stage plays a critical role in maintaining user data consistency and protecting enterprises from malicious attacks. In order to solve the problem of abnormal location in the multi-dimensional index system, Yongqian Sun et al. [26] subtly reduced the problem to a search problem, and then used the Monte Carlo search algorithm and hierarchical pruning to locate the abnormality.

5. Summary and outlook

Deep learning has gradually become the most advanced method in various tasks, and its application in time series has become more and more mature. No matter in the tasks of time series feature extraction, prediction, classification or anomaly detection, more and more novel deep model structures are proposed, and the specific exploration in various time series scenarios is also more perfect. But the application of deep learning in time series analysis still has a lot of work to study. First of all, there is no standard general model in terms of models. Secondly, the application of deep transfer learning in time series has yet to be studied for the improvement of model effects. Finally, the current time series analysis systems are mostly dedicated systems, which are not particularly common. Therefore, in the future, automatic machine learning systems dedicated to time series analysis will also have great research significance. This article introduces in detail the latest research progress of deep learning in the field of time series, which provides good help for further learning.

References

[1] King S Y, Hwang J N. Neural network architectures for robotic applications[J]. IEEE Transactions on Robotics & Automation, 1989, 5(5):641-657.

[2] Szegedy, C., Toshev, A., Erhan, D.: Deep neural networks for object detection. In: Advances in Neural Information Processing Systems. pp. 2553{2561 (2013).

[3] Graves, A., et al.: Supervised sequence labelling with recurrent neural networks, vol. 385. Springer (2012)

[4] Akmeliawati R, Ooi P L, Kuang Y C. Real-Time Malaysian Sign Language Translation using Colour Segmentation and Neural Network[C]// Instrumentation & Measurement Technology Conference. IEEE, 2007.

[5] Wen Xin, Zhang Xingwang, Zhu Yaping, et al. Intelligent fault diagnosis technology: MATLAB application [M]. Beijing University of Aeronautics and Astronautics Press, 2015.

[6] Lecun Y L, Bottou L, Bengio Y, et al. Gradient-Based Learning Applied to Document Recognition[J]. Proceedings of the IEEE, 1998, 86(11):2278-2324.

[7] Van Den Oord A, Dieleman S, Zen H, et al. WaveNet: A Generative Model for Raw Audio[J]. SSW, 2016, 125.

[8] Qing X, Niu Y. Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM[J]. Energy, 2018, 148: 461-468.

[9] Bouktif S, Fiaz A, Ouni A, et al. Single and Multi-Sequence Deep Learning Models for Short and Medium Term Electric Load Forecasting [J]. Energies, 2019, 12(1): 149.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202009_6(9).0024

[10] Petluri N, Al-Masri E. Web Traffic Prediction of Wikipedia Pages[C]// Web Traffic Prediction of Wikipedia Pages. 2018 IEEE International Conference on Big Data (Big Data). IEEE, 2018: 5427-5429.

[11] Bandara K, Shi P, Bergmeir C, et al. Sales Demand Forecast in E-commerce using a Long Short-Term Memory Neural Network Methodology[J]. arXiv preprint arXiv: 190104028, 2019.

[12] Bo Chen, Jo-Anne Ting, B. Marlin, N. de Freitas, Deep learning of invariant spatio-temporal features from video, in: NIPS 2010 Deep Learning and Unsupervised Feature Learning Workshop, 2010.

[13] I. Sutskever, G. Hinton, Learning multilevel distributed representations for high-dimensional sequences, Technical Report, University of Toronto, 2006.

[14] Katsuki, Takayuki & Osogami, Takayuki & Koseki, Akira & Ono, Masaki & Kudo, Michiharu & Makino, Masaki & Suzuki, Atsushi. (2018). Time-Discounting Convolution for Event Sequences with Ambiguous Timestamps. 10.1109/ ICDM.2018.00139.

[15] Cheng L C, Huang Y H, Wu M E. Applied attention-based LSTM neural networks in stock prediction [C]// International Conference on Big Data. IEEE, 2018:4716-4718.

[16] KUREMOTOT, KIMURAS, KOBAYASHIK, etal. Time Series Forecasting Using a Deep Belief Network with Restricted Boltzman Machine [J]. Neurocomputing, 2014, 137(15):47-56.

[17] TURNERJT. Time Series Analysis Using Deep Feed Forward Neural Networks [D]. Baltimore: University of Maryland, 2014.

[18] ROMEU P, ZAMORA-MARTINEZ F, BOTELLA-ROCAMO-RAP, etal. Time-Series Forecasting og Indoor Temperature Using Pre-trained Deep Neural Networks [c]//International Conference on Artificial Neural Networks. Berlin: Springer, 2013:451-458.

[19] LV Y, DUAN Y, KANG W, etal. Traffic Flow Prediction with Big Data: A Deep Learning Approach [J]. IEEE Transactions on Intelligent Transportation Systems, 2015, 16(2):865-873.

[20] LANGKVIST M, KARLSSONL, LOUTFIA. A Review of Un-supervised Feature Learning and Deep Learing for Time-Series Modeling [J]. Pattern Recognition Letters, 2014,42(1): 11-24.

[21] Cui Z, Chen W, Chen Y. Multi-Scale Convolutional Neural Networks for Time Series Classifification [J]. 2016.

[22] Wang Z, Yan W, Oates T. Time Series Classifification from Scratch with Deep Neural Networks: A Strong Baseline [J]. 2016.

[23] Mariusz Ze bik, Korytkowski M, Angryk R, et al. Convolutional Neural Networks for Time Series Classifification [J]. Journal of Systems Engineering & Electronics, 2017, 28(1):162- 169.

[24] Malhotra, P., Vig, L., Shroff, G., Agarwal, P.: Long short term memory networks for anomaly detection in time series. In: European Symposium on Artifificial Neural Networks. vol. 23.

[25] Laptev N, Amizadeh S, Flint I. Generic and Scalable Framework for Automated Time-series Anomaly Detection[J]. 2015.

[26] Ma, Minghua & Zhang, Shenglin & Pei, Dan & Huang, Xin & Dai, Hongwei. (2018). Robust and Rapid Adaption for Concept Drift in Software System Anomaly Detection. 13-24. 10.1109/ ISSRE.2018.00013.