Caliphate Journal of Science & Technology (CaJoST)

ISSN: 2705-313X (PRINT); 2705-3121 (ONLINE) Research Article

Open Access Journal available at: https://cajost.com.ng/index.php/files and https://www.ajol.info/index.php/cajost/index

DOI: https://dx.doi.org/10.4314/cajost.v5i3.14

Article Info

Received: 10th June 2023

Revised: 4th November 2023

Accepted: 7th November 2023

Online Anomaly Detection with Uncertainty Estimation and Concept Drift Adaptation using Quantile Regression

Ahmad I. Tambuwal[1*] and Habiba M. Sani[2]

[1]Faculty of Physical and Computing Sciences, Usmanu Danfodiyo University Sokoto

[2]Faculty of Physical and Computing Sciences, Usmanu Danfodiyo University Sokoto

***Corresponding author's email:**

idris.ahmad@udusok.edu.ng

Deep learning algorithms played an important role in big data applications that open other opportunities to study their applicability in anomaly detection. These algorithms are used in prediction-based anomaly detection methods for detection of anomalies in time series data. Various Recurrent Neural Networks (RNN) structures particularly based on Long Short-Term Memory (LSTM) or Gated Recurrent Unit (GRU) have been reported to be used for time series anomaly detection. Despite the performance of these methods, they are affected using a fixed threshold, and the assumption of Gaussian distribution on the prediction error to identify anomalous values. In addition, these techniques do not consider uncertainty in their predictions that may lead to over-confident predictions especially when there is limited training data. This impression motivates our previous research work that proposed a new anomaly detection method called Deep Quantile Regression Anomaly Detection (DQR-AD) that used confidence interval to identify anomalies in time series. However, the speed at which the time series data arrives and the dynamic change of normal behavior in a non-stationary environment will affect the offline training of DQR-AD using historical data. To mitigate these problems, this paper proposed an online DQR-AD that will enable the adaptation of concept changes in the data. Experiments conducted indicate that online DQR-AD method has better performance than its counterpart methods with relatively 10% margin. This result demonstrates how concept drift adaptation strategies adopted in the proposed method improve the performance of anomaly detection in time series.

Cite this: *CaJoST, 2023, 3,* 343-352

Keywords: Time Series, Prediction Interval, Uncertainty Estimation, Quantile Regression, Concept Drift Adaptation.

1. Introduction

Deep learning algorithms played an important role in big data applications which open other opportunities to study their applicability in anomaly detection [1]. These algorithms are used in prediction-based anomaly detection methods for detection of anomalies in time series data. Various deep learning techniques particularly based on Long Short-Term Memory (LSTM) [2], [3], [4], Gated Recurrent Unit (GRU) [5], Convolutional Neural Network (CNN) [6] and Autoencoder [7], [8], [9], [10], [11], [12] have been reported to be used for time series anomaly detection. However, the dynamic nature of time series will affect the performance of deep learning-based anomaly detection methods due to changes in the distribution of data that result in concept drift. Concept drift means changes in the characteristics of data over time where the characteristic of the new data is different from the previous data [13], [14], [15]. For example, when the computer's software is updated or its configuration is changed, data such as CPU utilization and the speed of reading or writing data in the disk will change. In anomaly detection system, the definition of abnormal behavior often changes with the change in data characteristic. As such, anomaly detection methods should be able to adapt to the new data and redefine the meanings of abnormal behaviors to accurately detects anomalies in the new data. These challenges results in the need for online anomaly detection methods which are able to adapt to concept drift [13], [16], [17]. The key idea of these methods is to adapt to concept changes in the data by updating model parameters in an incremental manner as the new data arrives. Examples of models that fall under this category includes Hierarchical Temporal Memory (HTM) [13], RNN [5], Autoencoder [11], and Sparse Gaussian Process (SGP) [16].

However, most of these methods does not take into account uncertainty in their predictions which may lead to over-confident predictions especially when there is limited training data[18]. Quantifying uncertainty is particularly important in critical applications such as clinical diagnosis [18], where a realistic assessment of uncertainty is important in determining disease status and appropriate treatment. In addition, such approach is most welcome in anomaly detection applications requesting better-informed decisions and mitigate against false anomaly alerts.

Various approaches have been developed to address uncertainty in deep neural networks for anomaly detection. They range from Bayesian approach [19] to interpreting dropout as performing variation inference [20], [21]. These methods, used prediction interval (PI) in regression tasks that quantifies the level of uncertainty associated with the point forecasts, thereby offering an interval of confidence for a prediction of lower and upper bounds [21]. In another approach, conditional quantile regression is used which capture statistical uncertainty on probability-weighted outcomes [22]. The aim of conditional quantile regression is to estimate quantiles of interest. In an attempt to find out which method can detect anomalies in an unlabeled multivariate time series data in real-time settings, Van de wiel et al [23] compared both univariate and multivariate regression-based anomaly detection methods. Their results show that multivariate anomaly detection methods perform better out of which quantile regression-based methods was overall the best approach for time series anomaly detection. This result motivates our previous work which created a new method called DQR-AD that used confidence interval to identify anomalies in time series data [24]. Despite the performance of this method compared with other anomaly detection methods, the speed at which the time series data arrives and the dynamic change of normal behavior in a non-stationary environment will affect its anomaly detection performance.

As such, this paper proposed an online DQR-AD that incorporates uncertainty estimation and concept drift adaptation to improve the performance of anomaly detection in time series data. In summary, this paper makes the following contributions:

1. To carryout uncertainty estimation, we used quantile regression model to estimate quantile values. Literature [18] shows that quantile regression can capture aleatoric uncertainty.
2. To obtain a probabilistic threshold, the proposed model estimates lower and upper quantiles which are used to threshold the input sequence for anomalies. Specifically, the anomaly detection is done by directly estimating a 95% confidence interval.
3. To provide concept drift adaptation, we compute anomaly likelihood using Q-function to define the abnormal degree of the current data point based on the previous data points. The likelihood is used to update the model parameters to adapt to the changes observed aver a significant period.

The rest of the paper is organized as follows: Section 2 provides review of related work. In section 3, we provide a detailed description of materials and methods. Section 4 describes the experiments and comparison of proposed method with other state-of-the-art methods. Finally, conclusion and future works are discussed in section 5.

2. Literature Review

Researchers have proposed some new methods for online anomaly detection [25]. These methods are generally categorized into two based on the approach they used. First is to build a new online model using a single incremental learning algorithm that will be train and test in real time [26], [27] or a stationary model that is initially train with historical dataset which is later updated to capture the features of new incoming data instance [5]. While the second involves an ensemble learning theory [28], [25] that trains multiple individual models for different part of the data streams. Research in the literature have proved the performance of online ensemble learning in handling concept drift in data streams [29]. The performance is due to their utmost predictive accuracy and stability-elasticity property. This property makes it easy for them to incorporate new data into the model, by training and adding new members to the ensemble, and naturally, forget irrelevant knowledge by removing the old members from the ensemble [30]. Despite this performance, they are slow and have higher computational cost. As such, the focus of this review will be on deep learning methods that used single incremental learning algorithm. The review is structured into paragraphs each highlighting our review on the methodology used for identifying anomalies in each technique.

To handle sequential and temporal nature of time series, a Hierarchical Temporal Memory (HTM) is used for an online anomaly detection in data streams [13]. This method models the temporal nature of the data stream at a given time and makes predictions for the next time step. At each step, the actual instance is compared with the predicted instance to compute anomaly score that is threshold to determine whether the point is anomalous or not. The likelihood of a point to be anomaly is determined by assumption of Gaussian distribution on prediction error. In a similar context, a study in [31] proposed a decision support mechanism for outlier detection in the concept drifting environment. This is achieved by implementing resistance learning concept with envelop module using single layer feed-forward neural network. A sequence-based moving window is also used to demonstrate incremental learning process where incoming data streams are added into learning process while older ones are discarded. This will allow the model to adapt with the dynamic changes in the time series and differentiate them with outliers.

In a different approach, a deep learning-based anomaly detection method is proposed where RBF classifier is used to identify engine fault using real data analysis [32]. This is done by training an offline model whose parameters are updated in real time to adapt with model uncertainties and dynamic changes caused by environment or mechanical wear of engine part. Fault detection is done using a Gaussian basic activation function that model the prediction error as normal distribution to identify faulty data when the prediction error exceeds a specified threshold. A more recent approach that used Recurrent Neural Network to detect anomalies in real time was proposed in [5]. This approach used instances of data that arrives continuously and trained the model incrementally thereby adapting to the changes that may occur in the data distribution. The authors used model prediction error to determine when to update the model based on the changes that occur in the data and whether those changes are anomalies or not. The predicted result is compared with the actual observed sequence and prediction error will be used for anomaly score computation thereby updating next step RNN model using BPTT. The likelihood of the predicted window to be anomaly is determined by assumption of Gaussian distribution on prediction error. The challenge remains as to find out whether the incoming data distribution matches the normal distribution. However, data do not necessarily follow a clear distribution and defining or assuming a distribution in modelling step is often difficult or inappropriate.

An alternative approach is the use of prediction interval that is computed by taking into accounts the uncertainty in both the data and data driven model. Hill and Minsker proposed an anomaly detection method that combine model prediction and its corresponding prediction interval to detect anomalies in data streams [33]. This method provides a principal framework for selecting a threshold where a data point is classified as anomalous or not based on whether it fall outside a given PI. The type of prediction interval used is t-interval that relies on Student's t-distribution where the prediction levels guide the selection of the interval width. This demonstrates the benefit of using PI over an arbitrary threshold value. The method also used AD strategy for processing future data points after flagging an anomaly. AD was used because its long-term performance is unaffected by previous miss-classifications.

Similarly, Chebyshev's inequality that proves most of the distribution values are clustered around the distribution mean can also be used to define an interval for identification of anomalies. The study in [34] used Chebyshev inequality condition to define a prediction interval that provides the same anomaly detection result without any assumption of the distribution of the data. The proposed method combines the application of TEDA for fault detection in industrial process. This density-based anomaly detection approach analyses the density of each data sample by computing its distance from all other samples read so far. Similarly, Ferdowsi et al [35] avoids the use of fixed threshold for identification of anomalies. The authors proposed an online outlier detection system that defined an interval using Chebyshev's inequality to declare outliers. Because of dynamic changes in an online system which lead to change in the distribution of the data, the authors used a fixed time window and assumed the measurement in each time window to have a fixed distribution. With this a data point at given time can be identified as outlier when it exceeds the interval defined using Chebyshev's inequality. In another approach, Reunanen et al, also utilized Chebyshev's inequality to identify anomalies in data streams [11]. This method combines Autoencoder and Logistic Regression for outlier detection and prediction in sensor data streams. The Autoencoder reconstruct the input data and produce hidden representation of the input that can be used to create the required labels for Logistic Regression to classify anomalous points. A data point is classified as anomalous when its reconstruction cost exceeds the expected reconstruction cost with three standard deviation that represents an upper bound of the inequality. Although no assumption of the distribution of the data is required but the method assumes the descriptive statistics of the unknown normal values to be initially defined.

As such, deep learning algorithms combined with quantile regression are required to predict target value accompanied with its quantile values. To demonstrate the performance of regression-based anomaly detection methods, van de Wiel et al [23], carried out experimental comparison real-time quantile regression-based anomaly detection methods for both univariate and multivariate time series data. All the anomaly detection methods used in the paper are quantile regression-based where the prediction of a target variable is accompanied with the quantile values. Their results shows that multivariate anomaly detection methods perform better out of which quantile regression-based methods was overall the best approach for time series anomaly detection. This impression motivates our research work in [24] where we proposed a new method called DQR-AD that used quantile interval compared with a fixed threshold to identify anomalies in time series data. Despite the performance of this method compared with other anomaly detection methods, the speed at which the time series data arrives and the dynamic change of normal behavior in a non-stationary environment will affect its anomaly detection performance.

As such, there is need to update DQR-AD to an online method that can handle concept changes that can occur in the time series. Specifically, this paper will propose an online DQR-AD method that incorporates uncertainty estimation and concept drift adaptation to improve the performance of anomaly detection in time series data.

3. Materials and Methods

Let's consider a multivariate time series $x = \{x_1, x_2, \dots, x_t\}$, where t is the length of the time series and each point $x_i \in R^m$ (for $i = 1 \dots t$) in the time series is an *m*-dimensional vector corresponding to the *m* features. To formulate the anomaly detection problem in an online setting, we used a sliding window to segment the time series with the aim to predict the next time step given a window of previous time steps. As such, a window of *w* previous time series points $h_w = \{x_{t-w+1}, \dots, x_{t-1}, x_t\}$ is used to forecast the next sequential time series point x_{t+1}. An observe point is then classified as anomaly if it deviates from its forecasted value using h_w as input. In summary, the proposed approach involves the following steps beginning at time t: (1) Use DQR model that takes h_w as input to forecast upper and lower quantiles that corresponds to the expected value at time $t+1$. (2) The upper and lower quantiles are used to calculate the quantile interval (QI) within which the actual point should lie. (3) When the current data point at time $t+1$ arrives, it is compared with the QI and when it is outside the QI range, it is classified as anomalous otherwise classified as normal data point. (4) The anomaly likelihood is determined using Q-function. (5) when the likelihood is less than or equal to a threshold, the predicted mean μ_{t+1} is added to h_w using ADAM strategy, otherwise, the current data point x_{t+1} is added to the h_w using AD strategy to update DQR model. This will allow the model to learn new characteristics of the data and redefine the meanings of abnormal behaviors. (6) Repeat steps 1 to 5. These steps are illustrated using a flowchart in fig.1. which are fully describe in the following subsections.

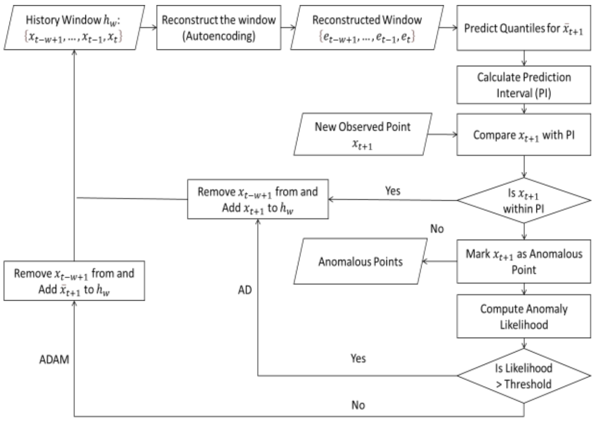

Figure 1 Schematic representation of proposed online DQR-AD method

3.1 Autoencoding

Prior to time series prediction, we first employ autoencoder that can extract meaningful representations from the time series. The Autoencoder model consists of an encoder and a decoder with two layers of LSTM each. The basic idea involves non-linear mapping of the input sequence to a fixed dimensional vector representation through an encoder, which is then followed, by another non-linear mapping of the vector representation back to the time series sequence using decoder. Specifically, the encoder read the history window h_w of t time steps and constructs a fixed-dimensional state from which the decoder constructs the following sequence of representations $\{e_{t-w+1}, \dots, e_{t-1}, e_t\}$. After the autoencoder perform the feature extraction, the DQR model then works on the reconstructed sequence received from the decoder to forecast the next time series point.

3.2 Deep Quantile Regression (DQR)

We consider a regression model that can forecast next time step based on previous time steps to be suitable method for learning temporal characteristics of time series. However, the focus here is to develop a regression model that can forecast time series points considering uncertainty. This is achieve via quantile regression that estimate conditional quantiles as oppose to classical regression which estimate the conditional mean of the respond variable [18]. The goal of quantile regression is to estimate conditional quantiles of interest which is applied in cases where parametric likelihood cannot be specified [43]. From the literature reviewed, it has been shown that quantile regression is able to captured aleatoric uncertainty [18]. As such, we quantify uncertainty in our model estimates using conditional quantile regression that estimate multiple quantiles of interest. To perform quantile regression on our LSTM model, we replace the Gaussian likelihood term in the LSTM loss function with the quantile loss that penalizes errors based on the quantile values generated. In this context, we specifically estimate lower quantile (LQ) and upper quantile (UQ) which corresponds to the median and one standard deviation from the mean respectively. These LQ and UQ are then used to calculate the mean μ and variance δ that can be used to obtain the corrected p-values for anomaly detection. We also reduce the chance of quantile overlapping via bootstrapping which allow the regression model to be iterated n times, thereby storing the predicted values in an array that is finally used to compute the desired quantiles.

3.3 Online Anomaly Detection

To obtain a probabilistic threshold, we train the DQR model to estimate LQ and UQ. We adapted the approach in [18] where we assume the output of DQR to be Gaussian. As such, we can use the two estimated quantiles to characterize the Gaussian distribution. Specifically, DQR estimate 0.15-th and 0.5-th quantiles that corresponds to the median and one standard deviation from the mean respectively. These LQ and UQ are then used to calculate the mean (μ) and variance (δ) which can be used to obtain the corrected p-values. The p-values are used for anomaly detection task where a probabilistic threshold is set at α= 0.05 significance level. This means, we perform the anomaly detection task by estimating a 95% confidence interval where an approximate αlevel prediction interval is computed by $[\mu - z_{\alpha/2}\delta,\ \mu + z_{\alpha/2}\delta]$, where $z_{\alpha/2}$ is the upper $\alpha/2$ quantile of standard normal. The prediction interval will serve as the threshold for identifying whether the data point is anomalous or not. The new data point is classified as normal when it is within the prediction interval; otherwise, it is classified as anomaly. In order to reduce the number of false positives, the threshold is selected based on the corrected p-values calculated to control the False Discovery Rate (FDR) [18]. Specifically, we chose the threshold corresponding to an FDR corrected p-value of 0.05.

3.4 Concept Drift Adaptation

To adapt to concept drift, the model needs to be updated in incrementally as new data arrives. To achieve this, this paper employs anomaly likelihood that is computed using Q-function to define the abnormal degree of the current data point based on the previous data points. The likelihood of the data point at time $t + 1$ is define in equation 1:

$$L_{t+1} = Q\left(\frac{x_{t+1} - \mu_{t+1}}{\delta_{t+1}}\right) \quad (1)$$

Where x_{t+1} is the actual observe point at $t + 1$. The Q-function measure the abnormal degree of the data point relative to the previous data points. The smaller the value of the Q-function, the higher the abnormal degree of the current data point and vice versa. As such, when the current data point at time $t + 1$ is classified as anomaly, L_{t+1} is computed which determine when the model will be updated to adapt to the new changes. This is achieved by comparing L_{t+1} with user defined threshold ϵ. When $L_{t+1} < \epsilon$, the predicted mean μ_{t+1} is added to h_w using ADAM strategy, otherwise, the current data point x_{t+1} is added to the h_w using AD strategy. When x_{t+1} is added to the h_w, the earliest data point x_{t-w+1} is removed from h_w. Specifically, when concept drift occurs, our proposed method will mark the current data point as anomalous. However, when the abnormal behavior continues for a longer period, the abnormal degree of the current data point will be low compared to the previous data points using L_{t+1}. As such, the current data point is added to the h_w to retrain the DQR model. This will allow the DQR model to learn the new characteristics of the data and hence adapt to the concept changes thereby redefining the abnormal behavior.

4. Experiments

This section describes experiments conducted to evaluate the performance of the proposed online DQR-AD method. We evaluate the performance of the method via comparison with other state-of-the-art anomaly detection methods. Specifically, we compare our proposed method with VAE-LSTM [12], NumentaTM [13], online RNN-AD [5], data-driven-AD [35], AE-AD [11], SGP-Q [16], and DQR-AD [146]. The section starts with dataset description, followed by experimental design, and ends with experimental results and discussions.

4.1 Datasets

The proposed method is evaluated on two commonly used anomaly benchmark datasets in the literature. These datasets include Yahoo Webscope dataset and Numenta Anomaly Benchmark (NAB) dataset. The choice of these datasets is because of the availability of the anomaly labels that we can used to validate our model.

4.2 Experimental Design

4.2.1 Models Training

To train all the models for time-series anomaly detection, records of each time-series are organized in 80% training set and 20% test set. From the training set, 20% of it will be used for validation. Fig. 2. illustrate the training and validation loss behavior on the nyc_ taxi dataset. A sliding window technique is used to segment both training, testing and validation sets into 24 hours samples of history h_w and prediction p_w windows. The size of the sliding window is set up by looking at the data and notice the presence of hourly pattern in all the time-series. Each input to the model is scaled between 0 and 1 for numerical stability during model training. The h_w is pass as input to the autoencoder which produced reconstructed sequence r_w. The DQR model will then takes the reconstructed sequence as input and predicts the quantile values. An autoencoder is constructed with two layers of LSTM cells that comprises 128 and 32 hidden states for both encoder and decoder. Deep quantile regression model is constructed using LSTM with two fully connected layers of 64 and 16 hidden units with Adam activation function. We also train each models using mini-batch gradient descent where a batch size of 128 is used and 100 number of epochs. For DQR-AD and online DQR-AD, we make use of bootstrapping by reactivating the dropout with value 0.5 and iterate for 100 times, thereby

storing the predicted values in an array which is finally used to compute the desired quantiles.

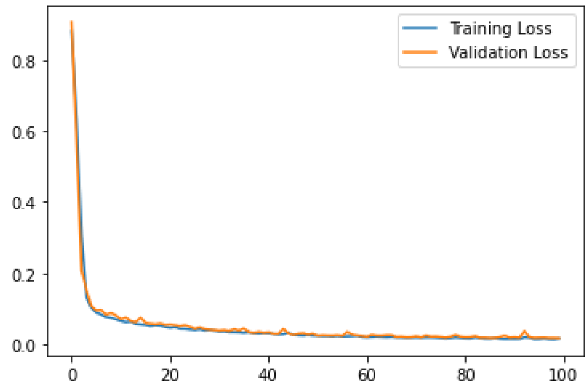

Figure 2 Training and Validation Loss

4.2.1 Anomaly Detection

The trained DQR model estimates 0.15-th and 0.5-th quantiles which corresponds to the median (μ) and one standard deviation (δ) from the mean respectively. The mean and variance can then be used to obtain the corrected p-values. The p-values are used for anomaly detection task where a probabilistic threshold is set at an α= 0.05 significance level. This means, we perform the anomaly detection task by estimating a 95% confidence interval where an approximate αlevel prediction interval is computed by $[\mu - z_{\alpha/2}\delta,\ \mu + z_{\alpha/2}\delta]$, where $z_{\alpha/2}$ is the upper $\alpha/2$ quantile of standard normal. Fig. 3. shows the actual data plot against the prediction interval for nyc_taxi test set. The prediction interval will serve as the threshold for identifying whether the data point is anomalous or not. The new data point is classified as normal when it is within the prediction interval; otherwise, it is classified as anomaly.

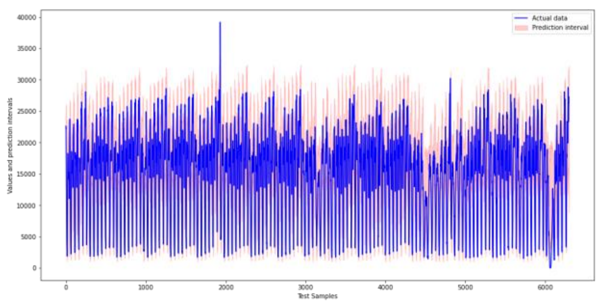

Figure 3 Predicted test set sample from nyc taxi data. Actual values are shown in blue with prediction interval shown in orange.

4.3 Experimental Results and Discussion

To evaluate the performance of online DQR-AD based on uncertainty estimation and concept drift adaptation, we conducted two levels of the same experiment. On the first level, we compare online DQR-AD with VAE-LSTM [12] and DQR-AD [24] using precision, recall, and F-score (2) which are the most commonly used metrics to evaluate the performance of anomaly detection methods in terms of false-positive rate [24].

$$F\_score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (2)$$

Table 1 shows the metrics scores of each of the three methods on the seven time series data used. It can be observed from this table that online DQR-AD has good performance in terms of f-score where out of seven datasets, online DQR-AD is minimally 18% better than both DQR-AD and VAE-LSTM on five datasets as indicated in bold. This is because online DQR-AD can identify large number of true anomalies with a smaller number of false positives. Although, DQR-AD and VAE-LSTM have better precision and recall respective in most of the datasets, online DQR-AD obtained relatively good score in both precision and recall. In order to evaluate the performance of online DQR-AD in terms concept drift adaptation, its compared with other online anomaly detection methods that includes NumentaTM [13], online RNN-AD [5], data-driven-AD [35], AE-AD [11], and SGP-Q [16]. Table 2 shows the AUC score of the five methods with best performance result indicated in bold. In this table, online DQR-AD method has better performance than its counterpart methods with relatively 10% margin on six out of the seven datasets. This performance is due to feature extraction and concept drift adaptation strategies adopted in the proposed method which improve its anomaly detection performance in time series.

Table 1: Comparative evaluation of Online DQR-AD with two other anomaly detection methods (DQR-AD and VAE-LSTM) on 7 time series. F-Score, Precision, and Recall are reported in this Table.

Datasets	DQR-AD			VAE-LSTM			Online DQR-AD		
	F-Score	Prec	Recall	F-Score	Prec	Recall	F-Score	Prec	Recall
Nyc_taxi	0.49	1	0.33	1	1	1	0.87	0.98	0.78
Amb Temp	0.11	1	0.06	0.89	0.81	1	0.99	0.98	1
CPU	0.67	1	0.50	0.81	0.69	1	0.98	0.97	0.99
EC2	0.82	1	0.70	0.99	0.99	1	0.99	0.99	1
Mach Temp	0.50	0.50	0.50	0.72	0.56	1	0.99	0.99	1
Key hold	0.09	1	0.05	0.03	0.02	0.10	0.85	1.00	0.74
Key updown	0.17	0.50	0.1	0	0.11	0	0.96	0.99	0.93

Table 2: Comparative evaluation of Online DQR-AD with five other online anomaly detection methods (NumentaTM, Online RNN-AD, Data-driven-AD, AE-AD, and SGP-Q) using AUC.

Datasets	NumentaTM	Online RNN-AD	Data-driven-AD	AE-AD	SGP-Q	Online DQR-AD
Nyc_taxi	0.74	0.89	0.48	0.61	0.42	0.97
Amb Temp	0.83	0.83	0.56	0.66	0.88	0.98
CPU	0.54	0.71	0.80	0.73	0.55	0.72
EC2	0.70	0.89	0.72	0.71	0.90	0.99
Mach Temp	0.33	0.55	0.50	0.60	0.54	0.71
Key hold	0.43	0.33	0.22	0.50	0.48	0.64
Key updown	0.40	0.24	0.23	0.42	0.38	0.59

5. Conclusion and Future Works

This paper proposed a new online anomaly detection method that incorporate feature extraction, uncertainty estimation, and concept drift adaptation to improve the performance of anomaly detection in time series data. This is achieved by combining autoencoder with quantile regression model to estimates lower and upper quantiles which are used to threshold the input sequence for anomalies. In addition, an anomaly likelihood is computed using Q-function that is used to update the model parameters for concept drift adaptation. To evaluate the performance of online DQR-AD based on uncertainty estimation, it is compared with DQR-AD and VAE-LSTM anomaly detection methods using seven time series from the NAB datasets. Results in Table 1 shows that online DQR-AD has good performance in terms of f-score where online DQR-AD is 18% better than both DQR-AD and VAE-LSTM on five datasets. To demonstrate the performance of online DQR-AD on concept drift adaptation where it is compared with five other online anomaly detection method. Results in Table 2 shows that online DQR-AD method has better performance than its counterpart methods with relatively 10% margin on six out of the seven datasets. This result demonstrates how feature extraction and concept drift adaptation strategies adopted in the proposed online DQR-AD improve the performance of anomaly detection in time series.

Our future work will involves extending the proposed method to detect a continuous faulty data from faulty sensor as an anomaly. In addition, computational complexity of the algorithms will be considered in the future as one of the metrics to be use for performance evaluation.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

The authors acknowledge the effort of the reviewers who help in reshaping the content of the paper to a better one.

References

[1] G. Pang, C. Shen, L. Cao, and A. Van Den Hengel, "Deep Learning for Anomaly Detection: A Review," ACM Computing Surveys, vol. 54, no. 2. 2021. **DOI**: https://doi.org/10.1145/3439950

[2] P. A. P Malhotra, L Vig, G Shroff and M. Rinard, "Long Short Term Memory Networks for Anomaly Detection in Time Series," in Proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 22-24 April 2015., 2015. **DOI**: http://www.i6doc.com/en/

[3] M. S. Aldosari, "Unsupervised Anomaly Detection in Sequences Using Long Short Term Memory Recurrent Neural Networks," PhD Diss. Georg. Mason Univ., p. 98, 2016. **DOI**: https://hdl.handle.net/1920/10250

[4] Z. C. Lipton, D. C. Kale, C. Elkan, and R. Wetzel, "Learning to diagnose with LSTM recurrent neural networks," in 4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings, 2016. **DOI:** https://doi.org/10.48550/arXiv.1511.03677

[5] S. Saurav, "Online Anomaly Detection with Concept Drift Adaptation using Recurrent Neural Networks," Proc. ACM India Jt. Int. Conf. Data Sci. Manag. Data - CoDS-COMAD '18, pp. 78–87, 2018. **DOI:** https://doi.org/10.1145/3152494.3152501

[6] M. Munir, S. A. Siddiqui, A. Dengel, and S. Ahmed, "DeepAnT: A Deep Learning Approach for Unsupervised Anomaly Detection in Time Series," IEEE Access, vol. 7, pp. 1991–2005, 2019. **DOI:** https://doi.org/10.1109/ACCESS.2018.2886457

[7] T. Amarbayasgalan, B. Jargalsaikhan, and K. Ryu, "Unsupervised Novelty Detection Using Deep Autoencoders with Density Based Clustering," Appl. Sci., vol. 8, no. 9, p. 1468, Aug. 2018. **DOI:** https://doi.org/10.3390/app8091468

[8] P. Malhotra, A. Ramakrishnan, G. Anand, L. Vig, P. Agarwal, and G. Shroff, "LSTM-based Encoder-Decoder for Multi-sensor Anomaly Detection," arXiv.org, Jul. 2016. **DOI:** https://doi.org/10.48550/arXiv.1607.00148

[9] M. Schreyer, T. Sattarov, D. Borth, A. Dengel, and B. Reimer, "Detection of Anomalies in Large Scale Accounting Data using Deep Autoencoder Networks," arXiv:1709.05254, Sep. 2017. **DOI:** https://doi.org/10.48550/arXiv.1709.05254

[10] W. Lu et al., "Unsupervised Sequential Outlier Detection with Deep Architectures," IEEE Trans. Image Process., vol. 26, no. 9, pp. 4321–4330, Sep. 2017. **DOI:** https://doi.org/10.1109/TIP.2017.2713048

[11] N. Reunanen, T. Räty, J. J. Jokinen, T. Hoyt, and D. Culler, "Unsupervised online detection and prediction of outliers in streams of sensor data," Int. J. Data Sci. Anal., vol. 9, no. 3, pp. 285–314, 2020. **DOI:** https://doi.org/10.1007/s41060-019-00191-3

[12] S. Lin, R. Clark, R. Birke, S. Schonborn, N. Trigoni, and S. Roberts, "Anomaly Detection for Time Series Using VAE-LSTM Hybrid Model," in ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 2020, vol. 2020-May, pp. 4322–4326. **DOI:** https://doi.org/10.1109/ICASSP40776.2020.9053558

[13] S. Ahmad, A. Lavin, S. Purdy, and Z. Agha, "Unsupervised real-time anomaly detection for streaming data," Neurocomputing, vol. 20, no. 0, pp. 53–0, 2016. **DOI:** https://doi.org/10.1016/j.neucom.2017.04.070

[14] J. Gama, I. Žliobaitė, A. Bifet, M. Pechenizkiy, and A. Bouchachia, "A survey on concept drift adaptation," Comput. Surv., vol. 46, no. 4, pp. 1–37, 2014. **DOI:** https://doi.org/10.1145/2523813

[15] A. Tsymbal, "The problem of concept drift: definitions and related work," Comput. Sci. Dep. Trinity Coll. Dublin, 2004.

[16] M. Gu, J. Fei, and S. Sun, “Online anomaly detection with sparse Gaussian processes,” Neurocomputing, vol. 403, pp. 383–399, 2020. **DOI:** https://doi.org/10.1016/j.neucom.2020.04.077

[17] S. Saurav et al., “Online anomaly detection with concept drift adaptation using recurrent neural networks,” in Proceedings of the ACM India Joint International Conference on Data Science and Management of Data - CoDS-COMAD ’18, 2018, pp. 78–87. **DOI:** https://doi.org/10.1145/3152494.3152501

[18] H. Akrami, A. Joshi, S. Aydore, and R. Leahy, “Quantile Regression for Uncertainty Estimation in VAEs with Applications to Brain Lesion Detection,” in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2021, vol. 12729 LNCS, pp. 689–700. **DOI:** https://doi.org/10.1007/978-3-030-78191-0_53

[19] J. Pang, D. Liu, Y. Peng, and X. Peng, “Anomaly detection based on uncertainty fusion for univariate monitoring series,” Meas. J. Int. Meas. Confed., vol. 95, pp. 280–292, Jan. 2017. **DOI:** https://doi.org/10.1016/j.measurement.2016.10.031

[20] L. Zhu and N. Laptev, “Deep and Confident Prediction for Time Series at Uber,” in IEEE International Conference on Data Mining Workshops, ICDMW, 2017, vol. 2017-Novem, pp. 103–110. **DOI:** https://doi.org/10.1109/ICDMW.2017.19

[21] A. Legrand, H. Trannois, and A. Cournier, “Use of uncertainty with autoencoder neural networks for anomaly detection,” Proc. - IEEE 2nd Int. Conf. Artif. Intell. Knowl. Eng. AIKE 2019, pp. 32–35, 2019. **DOI:** https://doi.org/10.1109/AIKE.2019.00014

[22] K. Yu and R. A. Moyeed, “Bayesian quantile regression,” Stat. Probab. Lett., vol. 54, pp. 437–447, 2001. **DOI:** https://doi.org/10.1016/S0167-7152(01)00124-9

[23] L. van de Wiel, D. M. van Es, and A. J. Feelders, “Real-time outlier detection in time series data of water sensors,” in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2020, vol. 12588 LNAI, pp. 155–170. **DOI:** https://doi.org/10.1007/978-3-030-65742-0_11

[24] A. I. Tambuwal and D. Neagu, “Deep Quantile Regression for Unsupervised Anomaly Detection in Time-Series,” SN Comput. Sci. 2021 26, vol. 2, no. 6, pp. 1–16, Sep. 2021. **DOI:** https://doi.org/10.1007/s42979-021-00866-4

[25] Z. Ding, M. Fei, D. Du, and F. Yang, “Streaming data anomaly detection method based on hyper-grid structure and online ensemble learning,” Soft Computing, pp. 1–13, 2016. **DOI:** https://doi.org/10.1007/s00500-016-2258-z

[26] M. Moshtaghi, C. Leckie, S. Karunasekera, J. C. Bezdek, S. Rajasegarar, and M. Palaniswami, “Incremental elliptical boundary estimation for anomaly detection in Wireless Sensor Networks,” in Proceedings - IEEE International Conference on Data Mining, ICDM, 2011, pp. 467–476. **DOI:** https://doi.org/10.1109/ICDM.2011.80

[27] D. Pokrajac, A. Lazarevic, and L. J. Latecki, “Incremental Local Outlier Detection for Data Streams,” in 2007 IEEE Symposium on Computational Intelligence and Data Mining, 2007, pp. 504–515. **DOI:** https://doi.org/10.1109/CIDM.2007.368917

[28] S. Ando, T. Thanomphongphan, Y. Seki, and E. Suzuki, “Ensemble anomaly detection from multi-resolution trajectory features,” Data Min. Knowl. Discov., vol. 29, no. 1, pp. 39–83, Jan. 2013. **DOI:** https://doi.org/10.1007/s10618-013-0334-x

[29] S. Ramírez-Gallego, B. Krawczyk, S. García, M. Woźniak, and F. Herrera, “A survey on Data Preprocessing for Data Stream Mining: Current status and future directions,” Neurocomputing, 2017. **DOI:** https://doi.org/10.1016/j.neucom.2017.01.078

[30] G. Ditzler, M. Roveri, C. Alippi, and R. Polikar, “Learning in Nonstationary Environments: A Survey,” IEEE Computational Intelligence Magazine, vol. 10, no. 4. pp. 12–25, 2015. **DOI:** https://doi.org/10.1109/MCI.2015.2471196

[31] S. Y. Huang, J. W. Lin, and R. H. Tsaih, “Outlier detection in the concept drifting environment,” in Proceedings of the International Joint Conference on Neural Networks, 2016, vol. 2016-Octob, pp. 31–37. **DOI:** https://doi.org/10.1109/IJCNN.2016.7727177

[32] M. S. Sangha, D. L. Yu, and J. B. Gomm, “Sensor fault diagnosis for automotive engines with real data evaluation,” Multicr. Int. J. Eng. Sci. Technol., vol. 3, no. 8, pp. 13–25, Aug. 2011. **DOI:** https://doi.org/10.4314/ijest.v3i8.2

[33] D. J. Hill and B. S. Minsker, “Anomaly

detection in streaming environmental sensor data: A data-driven modeling approach," Environ. Model. Softw., vol. 25, no. 9, pp. 1014–1022, 2010. **DOI:** https://doi.org/10.1109/ISBI45749.2020.9098543

[34] C. G. Bezerra, B. S. J. Costa, L. A. Guedes, and P. P. Angelov, "An evolving approach to unsupervised and Real-Time fault detection in industrial processes," Expert Syst. Appl., vol. 63, pp. 134–144, 2016. **DOI:** https://doi.org/10.1016/j.eswa.2016.06.035

[35] H. Ferdowsi, S. Jagannathan, and M. Zawodniok, "An online outlier identification and removal scheme for improving fault detection performance," IEEE Trans. Neural Networks Learn. Syst., vol. 25, no. 5, pp. 908–919, 2014. **DOI:** https://doi.org/10.1109/TNNLS.2013.2283456

[36] K. Yu and R. A. Moyeed, "Bayesian quantile regression," Stat. Probab. Lett., vol. 54, no. 4, pp. 437–447, Oct. 2001. **DOI:** https://doi.org/10.1016/S0167-7152(01)00124-9

