MQTT based Time-Series Anomaly Detection System for Smart Grid

Shijing Zhu And Bugong Xu
School of Automation Science and Engineering, South China University of Technology, Guangzhou 510640, P. R. China
201820115384@mail.scut.edu.cn

Abstract. Developing smart grid system is a crucial thing and relies heavily on modern technology. Powered by cloud computing and artificial intelligence, electrical grid can make a generational leap towards a more efficient and smart future. Meanwhile, machine learning can get great result in fields like anomaly detection and load forecasting. However, power grid equipment is usually quite underpowered, can’t perform deep learning algorithm locally. In this paper, we proposed a new stream processing architecture based on MQTT protocol, therefore time-series anomaly detection algorithms running on the cloud. Allowing any anomaly event can be detected fast and accurately.

1. Introduction
Electrical grid system is one of the most essential system in modern society. Electrical grid system should take advantage of cloud computing and machine learning to make the system more modern and improve efficiency greatly.

Power grid equipment always consists of microcontrollers and single board computers, which don’t have enough computational power to process and persist a large amount of data. To overcome this difficulty, data from edge devices should be uploaded to cloud and processed as a stream. Machine learning has achieved great results in areas like anomaly detection and load forecasting which fits very well with smart grid. In order to develop a smart grid system capable of machine learning, we propose a new stream processing architecture which can detect time-series anomaly instantly and can add other machine learning service easily.

In this paper we will also examine different time-series anomaly detection system algorithms. Including methods like SR (Spectral Residual) [1] and Seq2Seq (Sequence to Sequence) [2]. These algorithms will be test at industrial standard large labelled dataset. However, in real life, labelling a large amount of data requires a lot of manpower and even the samples are labelled, the distribution of positive samples and negative samples may not fit machine learning algorithm training process. Therefore, method like artificially injecting anomaly sample is used in this paper.

2. System Overview

2.1 Message streaming
Traditional power grid equipment generates a lot of data, usually can only perform simple processing task, and display on local display panel. In order to control device, monitor real time data and take the advantages of machine learning, it should be integrated with the cloud more closely.
The architecture is very important, every component should be loosely coupled. So heterogeneous systems can work together, and new type of device can intergrade into system easily. Figure 1 demonstrates the architecture of the whole system.

![Figure 1. System Architecture](image)

First challenge is message buffering and queueing. Since data generation and processing speed never match perfectly, simple TCP connection can’t be used. And since anomaly detection task requires low latency, saving to database and query it later doesn’t fit as well. To build a stream processing system, message queue (MQ) is used as a stream processing central hub [3]. Different message queue middleware has different traits and features [4]. To build a system specially for smart grid, MQTT based broker is chosen for many reasons. Following are the main reasons:

- MQTT client is way simpler than other clients, since most of work is done by broker.
- MQTT is push-based, which means the latency is much lower than pull-based systems.
- Topic don’t need to be created in advance, and subscriber can subscribe topic by regular expressions.

MQTT (Message Queuing Telemetry Transfer) protocol is an ISO standard publish-subscribe network protocol. MQTT is a protocol and message broker that support MQTT like Mosquitto and HiveMQ is actually used in this system. Comparing to industry standard MQ middleware like Apache Kafka, MQTT is optimised for IoT (Internet of Things) devices, and fits perfectly for smart grid system. Message Queue like Apache Kafka [5] is very advanced as distributed system, but fits poorly in smart grid. In smart grid, there are a lot of devices and messages, every device should have its own topic, Otherwise, consumers have to receive messages from every single device and can’t only specify target device.

The main difficulty building message transfer system is most of power grid devices are single board computers. Some runs Linux, but others runs RTOS, which only give very basic support for task managing. In order to make every device capable of integrating with this system. MQTT client is written for various operating system. Because creating thread is very hard on RTOS, our MQTT publisher and subscriber shares the same thread. And the lock provided by OS is unfair lock. Ticket lock algorithm is used make lock fair, so the thread can be shared.

2.2 Online processing
In smart grid, low latency is very critical. To eliminate latency overhead, MQTT broker’s WebSocket functionality is enabled, browser can subscribe to MQTT topic directly. However, raw data should be processed. To reduce bandwidth, power grid equipment publishes binary data, but browser and other tasks prefer JSON format. Therefore, an event-driven architecture is chosen. Format converter service
and other machine learning service, subscribe to input topics and publish processed data to output topics. With this framework-free architecture, multiple programming language can be used. Traditional service like format converter is written in Java to maximize performance. Machine learning service is written in python to utilize GPU resource. Most of service works independently, if some service like anomaly detection service isn’t needed in certain system, other functionality remains unaffected.

Time-series anomaly detection requires a sliding window of time-series data [6]. And in different tasks, the original sampling rate and detection rate may not match. For example, sampling may be second-level, but minute-level detection is more proper. A service that can sample and aggregate raw sequence into multiple rate of target sequences is needed.

2.3 Anomaly detection
Different systems have very different patterns, some sequences are seasonal, some sequences are stable. In order to build a robust anomaly detection system, many algorithms are examined in this paper. Including algorithm like Spectral Residuals, Seq2Seq. Even though both of them focus on time series detection, but the underlying mechanics are very different.

Since many sequences in grid system is system is seasonal. Outlier points in time-domain can be very hard to detect, but in frequency domain, they can be very significant in certain frequency. To obtain features in frequency domain, FFT (Fast Fourier Transform) is used to transform time series into frequency domain. However, using raw spectral as feature can only predicts which sequence contains anomaly. Spectral Residual method takes a step further. First the spectral residual of the log amplitude of the transformed signal is computed, then apply the Inverse Fast Fourier Transform to map the sequence back from the frequency to the time domain.

Sequence to sequence method takes a different approach. The main idea came from machine translation. It has an encoder-decoder architecture, first encode a time sequence into latent vector, then decode it back into time sequence. It trains in sequences without outlier. If the input contains outlier, it can’t encode and decode well, resulting in high reconstruction error.

These methods are drastically different and have its own advantages and disadvantages. For example, Seq2Seq method is LSTM based, can’t handle long sequences very well, and takes hours to train. Spectral Residual method can be really fast, but its accuracy can be relatively low. In this system, both of the algorithms can be deployed independently.

3. Applications
Putting these efforts together, a scalable cloud platform for smart grid is built. Every device’s data can be monitored in real time and many machine learning services including anomaly detection service and load forecasting service intergrade very well with this system.

![Figure 2. Grafana front end for data monitoring](image)

Data go through MQTT broker. Browser and many services can get ultra-low latency data from it. A MQTT to InfluxDB forwarder is used to persist data. As shown in Figure 2, we can use Grafana to monitor history data and real-time data. When anomaly happens, anomaly detection service can call
alert service, sending email or text message to administrator. Even though forecasting isn’t main focus of this system, ARIMA is used to do one step ahead forecast as a service in this system.

4. Methodology
Real-time time-series anomaly detection is our main focus. The problem is defined as below.

Given an input sequence of values, \( x = x_1, x_2, x_3, ..., x_n \), using algorithm to produce a sequence of \( y = y_1, y_2, y_3, ..., y_n \), where \( y_i \) denotes whether \( x_i \) is an anomaly point.

Before running any algorithm, since the input sequence can be any length, sliding window must be applied.

Even though there are many algorithms can detect anomaly, but many of them focus on image data or tabular data, rather than time series. Spectral Residual method and Sequence to Sequence framework are used to solve this problem. Spectral Residual utilizes feature in frequency domain. Sequence to Sequence framework encodes input time series into latent vector then reconstructs it back, if the original network contains outlier points, the reconstruction error at that point will be high.

4.1 Spectral Residual
Spectral residual method was original used in saliency detection, but saliency detection and anomaly detection are very similar. The formula that computes saliency map of input sequence \( x \) is shown as below:

\[
A(f) = Amplitude(F(x))
\]

\[
P(f) = Phase(F(x))
\]

\[
R(f) = L(f) - h_q(f) \cdot \log(A(f))
\]

\[
S(x) = ||F^{-1}(\exp(R(f)) + iP(f)))||
\]

First, apply Fast Fourier Transform to the input sequence. Then, compute spectral residual in frequency domain. spectral residual is computed by log amplitude subtracts averaged log amplitude. Then use Inverse Fast Fourier Transform to transform it back to time domain. The output result is saliency map. After retaining saliency map, based on threshold \( \tau \) and the following formula, the output sequence can be computed.

\[
O(x_i) = \begin{cases} 
1, & \text{if } \frac{S(x_i) - \overline{S(x)}}{\overline{S(x)}} > \tau \\
0, & \text{otherwise} 
\end{cases}
\]

The window size and threshold are the only hyperparameter need to be tuned in this algorithm. The window size should be several periods long if the time series is periodical. But the threshold is rather hard to decide. One practical method is getting the outlier distribution in the training dataset, then pick the outlier amount of highest score data, the threshold should near this. Sometimes, the outlier won’t get high score, another empirical way dealing with this situation is simply choose the threshold near 3, which suits most condition. If an empirical threshold is chosen, the method will become a fully unsupervised and can serve multiple purpose. Since there are many types of power equipment in smart grid, and most of the time, we don’t have enough resource labeling them manually. Using Spectral Residual with a constant threshold can be a great default anomaly detection method.

Another problem this method facing is near the start and the end of the sequence, the score will be higher than usual, which may lead to many unnecessary false positive. Therefore, before retaining saliency map, many points are padded to the end of the sequence. And at the end of the saliency map, the value should be compensated.

4.2 Sequence to Sequence
Sequence to Sequence is a neural machine translation framework at the beginning. It uses Bidirectional LSTM [7] as encoder and decoder. First encode into latent vector, then decode into another sequence. In this specific use case, the input output should be the same, which shares a lot of similarity with Variational Autoencoder [8], but Seq2Seq can handle time series as input very well.
During training process, data without outlier will be windowed as input, minimizing reconstruction loss. During test process, the outlier points will result in high MSE (Mean Squared Error).

Seq2Seq is LSTM based means that it is state-dependent. And LSTM can’t handle long sequence very well, increasing window length may slow the training speed and get worse result.

5. Experiments

5.1 Datasets
First, we use synthetic data to examine both methods’ performance, and the influence of hyperparameters.

Then two different real datasets are picked specially for each method. These datasets are industry standard and periodic which is quite representative for smart grid.

Yahoo S5 dataset is a labeled anomaly detection dataset by yahoo lab. It contains synthetic part and real part. Real part is actually used and is from yahoo’s real-life internal service. The interval is one hour. And it’s one of the most popular anomaly detection datasets. This dataset is used for evaluate Spectral Residual method.

Electrocardiograms (ECG’s) dataset is a dataset contains 5000 electrocardiograms. It’s formally known as BIDMC Congestive Heart Failure Database. The dataset’s length is 140 and very periodic which fits very well with Seq2Seq method.

5.2 Synthetic Benchmark
First generate sin wave with Gaussian noise, then inject outlier with specified amount, strength and length into it. In this case, we used 0.25Hz as sin wave frequency, 0.1 as Gaussian noise standard deviation, and 10% outliers. Figure 3 shows the outlier injection process. Different window size of different algorithms is examined. Below is the input we produced.

![Synthetic data with outlier injection](image)

**Figure 3.** Synthetic data with outlier injection

Table 1 shows the comparison results of different method and different window size on synthetic dataset:

| Method | Window size | F1-score | Precision | Recall |
|--------|-------------|----------|-----------|--------|
| SR     | 20          | 0.936    | 0.978     | 0.898  |
| SR     | 50          | 0.982    | 0.982     | 0.983  |
| SR     | 150         | 0.965    | 0.980     | 0.951  |
| Seq2Seq| 20          | 0.917    | 0.966     | 0.873  |
| Seq2Seq| 50          | 0.958    | 0.987     | 0.923  |
| Seq2Seq| 150         | 0.935    | 0.971     | 0.902  |
Both of the algorithm can detect anomaly very effectively, and longer window doesn’t always result in better result. And Seq2Seq takes much longer to train especially when window size is long.

5.3 Yahoo dataset

Yahoo dataset is a real-life dataset collect and labeled by yahoo. It contains 67 real sequence from Yahoo services. The anomaly is sparse and random, which challenges anomaly detection algorithms.

![Figure 4. Spectral Residual on Yahoo Dataset](image)

Figure 4 shows the saliency map and threshold of a certain sequence from yahoo dataset. In this particular time series, SR algorithm works really well, can achieve F1-score greater than 0.90. However, some outliers can be very hard to detect. Table 2 shows the overall result of SR method on Yahoo Dataset.

| Window size | F1-score | Precision | Recall |
|-------------|----------|-----------|--------|
| SR          | 50       | 0.582     | 0.502  | 0.694  |

5.4 ECG dataset

Seq2Seq method tend to works better on medium length and periodic time series, such as ECG dataset. Original ECG dataset contains 5 categories. One of them should be viewed normal and others should be anomaly. Table 3 shows the overall result of Seq2Seq method on ECG Dataset.

| Window size | F1-score | Precision | Recall |
|-------------|----------|-----------|--------|
| Seq2Seq     | 140      | 0.953     | 0.965  | 0.943  |

6. Conclusion

To build smart grid, we proposed a new architecture capable of processing a large amount of data. Only the system may not be enough, the algorithm is what make smart grid smart. In this paper, we focus on anomaly detection, which is very crucial to smart grid. The system will keep improving in the future.

Reference

[1] Hou, X. and Zhang, L., 2007, June. Saliency detection: A spectral residual approach. In 2007 IEEE Conference on computer vision and pattern recognition (pp. 1-8). Ieee.
[2] Sutskever, I., Vinyals, O. and Le, Q.V., 2014. Sequence to sequence learning with neural networks. In Advances in neural information processing systems (pp. 3104-3112).

[3] Stonebraker, M., Çetintemel, U. and Zdonik, S., 2005. The 8 requirements of real-time stream processing. ACM Sigmod Record, 34(4), pp.42-47.

[4] Naik, N., 2017, October. Choice of effective messaging protocols for IoT systems: MQTT, CoAP, AMQP and HTTP. In 2017 IEEE international systems engineering symposium (ISSE) (pp. 1-7). IEEE.

[5] Ren, H., Xu, B., Wang, Y., Yi, C., Huang, C., Kou, X., Xing, T., Yang, M., Tong, J. and Zhang, Q., 2019, July. Time-Series Anomaly Detection Service at Microsoft. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 3009-3017).

[6] Laptev, N., Amizadeh, S. and Flint, I., 2015, August. Generic and scalable framework for automated time-series anomaly detection. In Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1939-1947).

[7] Gers, F.A., Schmidhuber, J. and Cummins, F., 1999. Learning to forget: Continual prediction with LSTM.

[8] Kingma, D.P. and Welling, M., 2013. Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114.