Anomaly Detection in Multivariate Time Series from Industrial Data

Brijesh Gupta Chagi Ramesh
gupta.chagiramesh@tu-dortmund.de

***Abstract*—With the emergence of AI, IoT, Industry 4.0 and Smart Manufacturing, the number and types of sensors used in factories continue to grow every day. Thus, the analysis of the Multivariate Time Series Data generated by them becomes important. However, extracting important features from such data is still a daunting task because it simultaneously takes into account the correlation between the pair of sensors and temporal information of each time series. This master thesis aims to detect Anomalies at an early stage of a production line, in an unsupervised way, to save cost and time wasted by a possible production failure.**

I. Motivation

The demand for automation paired with rapid advancements in the field of sensor technology, machine learning and artificial intelligence has revolutionized the Manufacturing Industry. Factories all around the world are quickly adopting IoT, Industry 4.0 and Smart Manufacturing to improve their profits and safety standards for their employees. This transforms the industries into highly complex systems generating massive amounts of multivariate time series data.

An important task in managing these systems is to detect anomalies in the system during operation. Doing so enables predictive maintenance of the system which reduces the frequency of maintenance and avoid potential failures. This is especially important because the probability of failures grow with an increase in scale and complexity of manufacturing systems. It has been reported that a minute of downtime of a completely automated manufacturing plant may cost up to 20,000 USD, thus making accurate anomaly detection crucial to avoid serious financial losses.

For building systems that automatically perform anomaly detection, a major hurdle is a high imbalance in the dataset. This occurs due to the fact that anomaly records are very few, their patterns are highly irregular and their detection in a timely fashion is crucial. Historically, rule-based policies have been applied for their detection. These rules are created based on experience, domain knowledge or ad hoc data analysis. This makes it susceptible to unseen anomalies and unable to generalise to other environments or domains.

With the rise of Big Data and Artificial Intelligence, machine learning techniques have become more common. A few of them are based on time series analysis models like Autoregressive Integrated Moving Average. In other cases, classification algorithms like KNN or Dynamic Time Warping are used. But, classifiers are built to maximise accuracy and thus perform poorly on imbalanced datasets.

To address these challenges unsupervised Artificial Neural Network models based on LSTM Auto-encoders are built. The ability of LSTM cell to capture temporal dependencies makes them well suited for this task. The LSTM Auto-encoder is trained to learn the normal functioning of the system without the need for labelled data. Once trained, online data is fed in real-time as a set of sliding windows to perform anomaly detection. Alerts are triggered when the online data deviates from training data based on reconstruction error of the Auto-Encoder thus detecting the Anomaly.

Generative Adversarial Networks(GANs) is another well researched deep learning framework that can perform similar tasks. They have been very successful in the field of Image and Video processing. Thus, several researchers have tried to use them in the field Multivariate Time Series to achieve similar success. This master thesis aims to evaluate and hopefully improve the performance of such LSTM Auto-encoder and GANs on actual industrial data.

II. Recent Work

Anomaly Detection on Multivariate Time Series data is a challenging and actively researched task due to its high level of importance for the industry. Various approaches have been researched over the years. These can be mainly classified into the following categories.

A. Rule Based Approach

This system consists of a set of rules that normally take the form of if-then rules. These rules can be formed based on past experiences or by domain experts [1]. [2] and [3] are also examples where rule-based systems were used to deal with anomalies.

B. Classification Approach

Classification based approaches aim to detect patterns in time series data. Methods based on distance paired with k-Nearest Neighbours [4] has been proven to provide good results in some cases. Dynamic Time Warping (DTW) [5] has proven to be the best distance measure to use along with k-NN. Ensemble methods such as Isolation forests [6] have also been used in some cases. Other methods such as the One-class SVM [7] learns the density distribution of the data and classifies new data as normal or anomalous.

C. Forecasting or Prediction based Approach

Forecasting models predict future values based on the given sequence of inputs. These models can also be applied for Anomaly Detection. This is achieved by training the models on anomaly free data. Once well trained, anomalies can be detected by comparing the predicted and actual values of real-time data. Deep Neural Networks with Long Short-Term Memory cells [8] [9] have proven to be really effective for this application.

Also, Generative Adversarial Networks(GANs) based anomaly detection methods have come up as the state of the art in the field of Image and Video processing. Inspired by this, GANs based models have also been implemented to perform Anomaly Detection in the field of Multivariate Time Series Data [10] [11].

III. METHODOLOGY

For the majority of Anomaly Detection problems, the datasets are highly imbalanced i.e. anomalous data is limited. Adding to this anomalous patterns could be irregular making them harder to recognize. Due to these limitations classification or supervised models are not suitable. Rather, unsupervised models are more suited for this task.

This thesis aims to evaluate two prominent unsupervised approaches, an LSTM based Auto-encoder approach and a Generative Adversarial Network approach, on the real-world industrial data.

A. LSTM based Auto-encoder

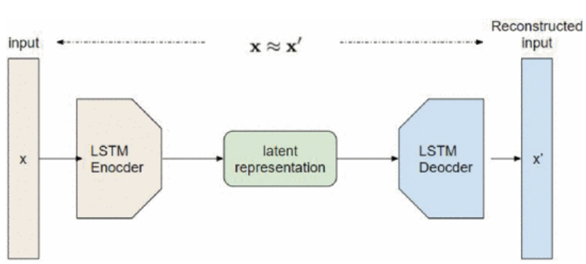

Fig. 1. Architecture of the LSTM Auto-encoder.

The model architecture of the Auto-encoder is as shown in Fig.1. The Auto-encoder model is made up of an encoder and a decoder network with both based on recurrent neural networks. The approach can be broken down into training and evaluation phases. During the training phase, the Auto-encoder is trained purely on historical data free from anomalies. The Encoder part of the model encodes the input data and maps it to a latent space. This latent space representation encompasses the information and features that represent the input. The Decoder uses these latent space representations to reconstruct the input data as best as possible. The output of the Auto-encoder is the reconstruction error of the given inputs.

Thus, the loss function of the Auto-encoder is calculated based on the reconstruction error. The optimizer attempts to train both the encoder and the decoder simultaneously to minimize the reconstruction error. In doing so, the Auto-encoder learns the normal behaviour of the system.

During the evaluation phase, the actual data with the anomalies present is given to the model. The reconstruction errors for a normal pattern should be small, whereas an unseen anomalous pattern is likely to have a higher reconstruction error. Hence, by using a threshold ϵ, anomalous patterns can be identified. The value of ϵ is assigned based on the reconstruction errors during the training phase.

B. Generative Adversarial Network

Instead of feeding the Multivariate Time Series data directly as input, they are first converted into an image like structure called Signature Matrices. Considering a time segment from time $t - w$ to t, an nXn matrix is constructed by taking pairwise inner-product of two time series within this time segment. The equation to perform this calculation is given by 1 This is done to ensure that the inter-correlation between different pairs of signals are captured.

$$m_{ij}^{t} = \frac{\Sigma_{\delta=0}^{w} x_i^{t-\delta} x_j^{t-\delta}}{w} \tag{1}$$

Along with capturing the value scale correlations and shape similarities, the signature matrix is also robust to input noise.

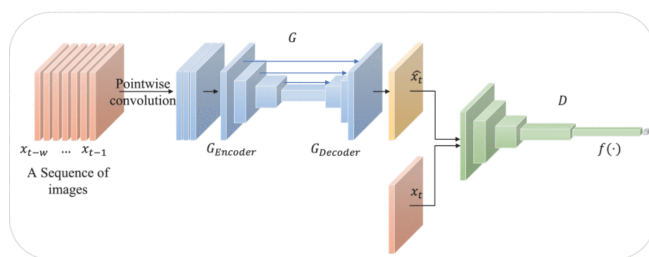

Fig. 2. Architecture of the LSTM Auto-encoder.

The basic architecture of the GAN itself can be seen in Fig.2. It consists of a Generator part and a Discriminator part. The goal here is to train the Generator to learn the mapping of signature matrices to enable it to predict the next signature matrix in that time window. This is achieved by training the GAN on purely normal data that is free from anomalies. The generator, consisting of an encoder and a decoder, learns the latent vector space of the normal data distribution. Simultaneously the Discriminator tries to distinguish the input images(real) from reconstructed images(fake) as accurately as possible. Thus, the overall loss used to train the GAN is a combination of the adversarial loss and the reconstruction loss.

Since the GAN has only been exposed to normal data during training, the reconstruction loss will increase when it is exposed to data containing anomalies during evaluation. Additionally, by computing the $L2$ distance of feature vectors of the input image and the generated image, we can measure the severity of the anomaly.

IV. Dataset

The data source for this thesis is obtained from a real-life production plant that deals with manufacturing Webs. The anomalies in this case are the Web Breaks that occur from time to time. These breaks not only halt production but also sometimes damage the machinery. The goal here is to predict these breaks before they occur so as to minimize downtime and reduce costs.

The plant contains about two hundred sensors measuring temperature, pressure, velocity and other parameters all across the production line. The data generated by each of them produce one time series. Combining them together produces a Multivariate Time Series of size two hundred. This sensor data is sampled at a frequency of 50 ms i.e. 20 data-points per second. Sampling the sensor data at such high frequencies generates an enormous amount of data. This makes data handling and cleaning a challenging task in itself. Finding efficient solutions to these challenges will also be a part of this thesis.

References

[1] Han Liu, Alexander Gegov and Mihaela Cocea, "Rule-based systems: a granular computing perspective", Granular Computing, vol. 1, no. 4, pp. 259-274, Dec 2016.

[2] Jean Roy, Rule-based expert system for maritime anomaly detection, 2010.

[3] Weng-Keen Wong, Andrew Moore, Gregory Cooper and Michael Wagner, "Rule-based anomaly pattern detection for detecting disease outbreaks", Eighteenth National Conference on Artificial Intelligence, pp. 217-223, 2002.

[4] Hautamaki, Ville, Ismo Karkkainen, and Pasi Franti. "Outlier detection using k-nearest neighbour graph." Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.. Vol. 3. IEEE, 2004.

[5] S. Seto, W. Zhang and Y. Zhou, "Multivariate time series classification using dynamic time warping template selection for human activity recognition", 2015 IEEE Symposium Series on Computational Intelligence, pp. 1399-1406, Dec 2015.

[6] Liu, Fei Tony, Kai Ming Ting, and Zhi-Hua Zhou. "Isolation forest." 2008 Eighth IEEE International Conference on Data Mining. IEEE, 2008.

[7] Manevitz, Larry M., and Malik Yousef. "One-class SVMs for document classification." Journal of machine Learning research 2, no. Dec (2001): 139-154.

[8] Malhotra, P., Vig, L., Shroff, G., Agarwal, P. (2015, April). Long short term memory networks for anomaly detection in time series. In Proceedings (Vol. 89, pp. 89-94). Presses universitaires de Louvain.

[9] Zhang, Chuxu, et al. "A deep neural network for unsupervised anomaly detection and diagnosis in multivariate time series data." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 33. 2019.

[10] Khoshnevisan, Farzaneh, and Zhewen Fan. "RSM-GAN: A Convolutional Recurrent GAN for Anomaly Detection in Contaminated Seasonal Multivariate Time Series." arXiv preprint arXiv:1911.07104 (2019).

[11] Choi, Yeji, et al. "GAN-Based Anomaly Detection and Localization of Multivariate Time Series Data for Power Plant." 2020 IEEE International Conference on Big Data and Smart Computing (BigComp). IEEE, 2020.