EMERGING TECH CONFERENCE – Edge Intelligence

Volume 03, 2024, Pages 17–23

Proceedings of Emerging Tech Conference:
Edge Intelligence 2024

Leveraging Empirical Mode Decomposition for Time Series Anomaly Detection in IoT and Smart City Applications

Ioannis Stasinopoulos[1] and Ilias Theodorakopoulos[1]
[1] *Democritus University of Thrace, Greece.*
ioanstas@ee.duth.gr, *iltheodo@ee.duth.gr*

Abstract

This paper proposes a novel forecasting-based anomaly detection method that leverages Empirical Mode Decomposition (EMD) to break down complex univariate time series into Intrinsic Mode Functions (IMFs). Gated Recurrent Units are tasked with processing IMFs to predict future values. By reconstructing the time series from predicted IMFs, anomalies are detected when deviations between predicted and observed values of the time-series occur. The proposed method is validated on real-world datasets from smart city applications, demonstrating its efficiency in handling noisy data and multiscale seasonal trends while maintaining low computational overhead, making it suitable for deployment in resource-constrained IoT environments.

1 Introduction

Anomaly detection in univariate time series (TS) is a critical task across various domains, including finance, healthcare, and industrial monitoring. The primary approaches to anomaly detection involve forecasting, reconstruction, representation learning, and hybrid models [1]. Forecasting-based techniques predict future data points based on historical trends, flagging anomalies when actual observations deviate significantly from predictions. This method is particularly effective for real-time monitoring applications, such as detecting abnormal traffic in data centers, monitoring urban environments and financial markets or identifying equipment malfunctions in industrial settings. Reconstruction-based techniques, on the other hand, focus on capturing the underlying structure of TS data by reconstructing it from a low-dimensional representation. Anomalies are detected when the reconstruction fails to accurately represent the data. These techniques are especially useful in complex or noisy environments, such as medical diagnostics, where subtle deviations from the norm can indicate significant issues. Representation learning techniques, which involve learning a compact and informative representation of the TS, are adept at capturing intricate patterns in data. They are often employed in applications like fraud detection, where understanding the relationships between transactions is crucial. Hybrid models combine these approaches to leverage the strengths of each, making them effective in scenarios where both temporal and spatial data characteristics are important, such as in the management of multivariate TS in power grids.

In the context of Internet of Things (IoT) and smart city applications like traffic monitoring, the choice of anomaly detection technique must consider the limited computational resources available. Forecasting-based methods present a compelling solution for such applications, since in some configurations such methods can be computationally efficient and operate effectively on devices with constrained processing capabilities. By focusing on predicting the next data point in a TS, forecasting based models can both promptly identify anomalies with small computational footprint, as also benefit from training without the need of data painstakingly annotated in quest of often rare and indistinct anomalies. The scalability of forecasting-based models also makes them ideal in domains such as smart cities, for widespread use across a city's sensor network. Each sensor can be equipped with a locally trained model that is tailored to the specific traffic patterns of its location, enabling context-aware anomaly detection. This localized approach not only enhances the accuracy of anomaly detection but also ensures that the system remains efficient and responsive, even as the number of deployed sensors increases.2

In this work, we present a lightweight forecasting-based method for anomaly detection in univariate TS, leveraging signal decomposition via Empirical Mode Decomposition (EMD) [2]. The employment of EMD allows the separation of a TS into multiple simpler components, thus enabling improved forecasting with small recurrent neural models in the form of Gated Recurrent Units (GRU) architecture. The rest of the paper is organized as follows: In Section 2 we provide a brief outline of the literature in forecasting-based anomaly detection in TS and comment on the pros and cons of each neural architecture. In section 3 we describe the proposed method for anomaly detection. In section 4 we provide experimental results on two popular datasets from the field of traffic monitoring in urban environments and highlight the efficiency of decomposition-based approaches in forecasting with small computational footprint. Finally, conclusions are drawn in section 5.

2 Related work

Forecasting-based models are fundamental for anomaly detection in univariate TS, wherein a model learns historical patterns to predict future values or sequences. Anomalies are flagged when the deviation between the predicted and actual values exceeds a defined threshold. This approach is particularly effective for real-world anomaly detection, where normal behavior is plentiful, and anomalous instances are rare [3]. By forecasting one step at a time, these models can efficiently identify deviations in both short-term and long-term dependencies within the TS.

One of the primary architectures used for forecasting-based anomaly detection in univariate TS is the Long Short-Term Memory (LSTM) [4] network, a variant of Recurrent Neural Networks (RNNs) designed to capture long-term dependencies in sequential data. For instance, LSTM-AD [5] is specifically designed to handle univariate TS (UTS) without labeled data, leveraging stacked LSTM layers to forecast multiple future time steps, allowing model to capture complex temporal patterns, making it more effective particularly when the distinction between normal and anomalous data depends on long-term behavior. Bontemps et al. [6] introduced a simpler LSTM architecture for collective anomaly detection in UTS, focusing on learning normal TS behavior and detecting collective anomalies by tracking prediction errors that exceed a threshold.

Besides recurrent neural architectures, feedforward topologies such as Convolutional Neural Networks (CNNs) and Transformers have also been proposed for forecasting-based anomaly detection. DeepAnt [7], a CNN-based model, applies a sliding window approach for forecasting in UTS, predicting future values based on historical windows and detecting anomalies by comparing predictions to actual observations. DeepAnt is particularly notable for its robustness to data contamination and its ability to detect small deviations in TS patterns in noisy or sparse datasets. TCN-ms [8] (Multiscale TCN) extends the standard Temporal Convolutional Network (TCNs) [9], by incorporating multiscale dilated convolutions to capture temporal dependencies at various resolutions, making it effective for detecting anomalies in quasi-periodic data, such as medical TS or industrial sensor data.

SAnD (Self-Attention Network for Anomaly Detection) [10] utilizes transformers' attention mechanisms to detect anomalies in UTS data. Unlike RNN-based models, SAnD processes entire sequences in parallel, capturing both short-term and long-term dependencies simultaneously. Furthermore, the self-attention mechanisms enhances its ability to detect anomalies in clinical and industrial applications where fine-grained analysis is essential.

While feedforward architectures offer robustness to noise and superior multi-scale mechanisms, they often require significant computational resources. An alternative approach to improving forecasting with simpler recurrent models is decomposing the TS into simpler components, where long term trends are removed or disregarded, allowing multiple simpler recurrent models to forecast each component separately. In that spirit, the recently proposed AD-LTI [11] combines Gated Recurrent Unit (GRU) [12] networks with the Prophet model [13] to enhance the detection of seasonal patterns in UTS. GRU, is employed to model temporal dependencies, while the Prophet model captures seasonal trends. A key innovation of AD-LTI is the Local Trend Inconsistency (LTI) metric, which measures the deviation of recent predictions from the expected trend, enabling more reliable anomaly detection, even when historical data includes noise or anomalies.

3 Proposed Method

A popular method for the decomposition of UTS into simple components is Empirical Mode Decomposition (EMD) [2]. EMD effectively and efficiently [14] decomposes a complex, non-linear TS signal into a finite number of simpler components, known as Intrinsic Mode Functions (IMFs). Each IMF captures specific oscillatory modes within the signal, isolating patterns such as noise, short-term fluctuations, and long-term trends. This decomposition simplifies the prediction task by enabling a GRU model to focus on predicting the evolution of less complex, individual IMFs rather than the entire raw signal.

EMD begins by decomposing the input TS $\boldsymbol{x} = \{x_0, x_1, \dots, x_t\}$ into a set of m IMFs**,** $\boldsymbol{I^1} = [I_0^1, I_1^1, \dots, I_t^1]$ ……. $\boldsymbol{I^m} = [I_0^m, I_1^{\mathrm{m}}, \dots, I_t^m]$. Each IMF represents a distinct component of the original signal, ranging from high-frequency oscillations to slower trends. This separation of timescales allows the forecasting model to handle different temporal characteristics of the data in a structured way. For instance, higher frequency IMFs, which capture short-term fluctuations, can be prioritized for prediction tasks, while lower-frequency IMFs representing long-term trends can be left unmodified, as they provide less information for detecting short-term anomalies.

Figure 1 illustrates the overview of our proposed approach for UTS anomaly detection. In this scheme, each IMF is used as input to a separate GRU model, with the task of predicting the next value in its respective series. The GRU models are trained in an unsupervised manner, using a reconstructive objective to minimize the difference between the predicted and actual IMF values. Once the GRU models have forecasted the next value for each IMF, the predicted IMFs are summed to reconstruct the overall forecasted value of the original TS. This allows the model to combine the predictive power of the GRUs for the relevant IMFs while potentially incorporating unmodified components, such as longterm trends, that are not critical for short-term anomaly detection.

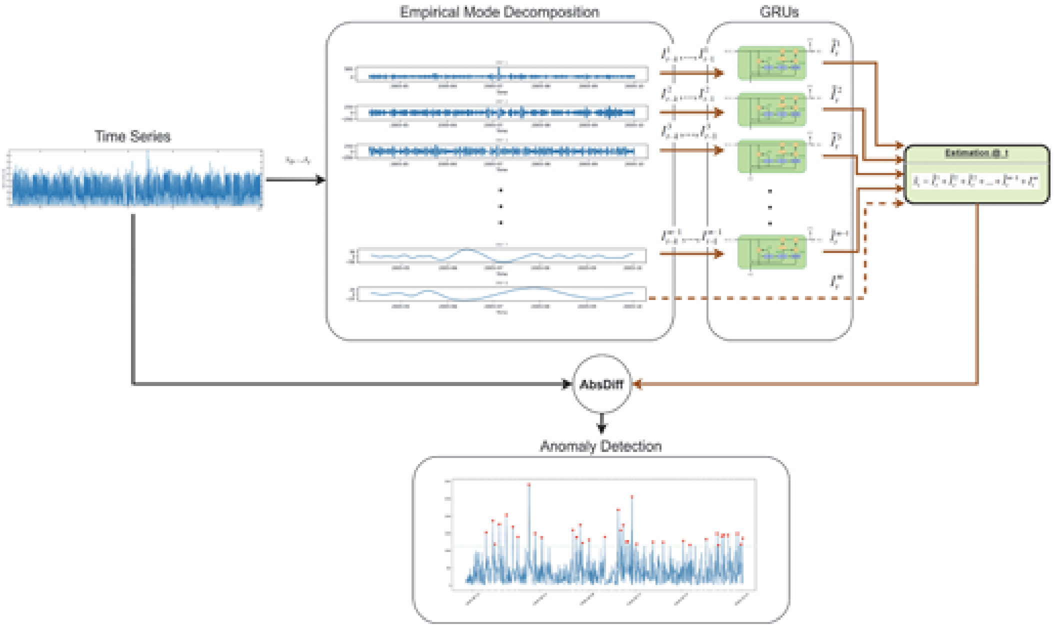


Figure 1: Overview of the proposed forecasting-base anomaly detection scheme

This framework also allows for flexibility in the choice of EMD variants. Ensemble EMD (EEMD) [15] can be used to improve robustness by addressing mode-mixing issues, often observed in standard EMD. Alternatively, in scenarios where current detection of anomalies is critical, online EMD [16] can be employed to ensure that the decomposition process can handle streaming data efficiently. Furthermore, depending on the prediction horizon required, the GRU models can be configured to process either the entire IMF sequence or a fixed-size time window, which captures the necessary temporal context for forecasting.

4 Experimental Evaluation

For the experimental evaluation, we used two real-world datasets, the Dodgers Loop Sensor[1] and Calit2[2] [17], to demonstrate the functionality of the proposed forecasting-based anomaly detection method in smart city applications. The Dodgers dataset consists of vehicle count data collected every 5 minutes over a span of six months from an inductive loop sensor on the Glendale on-ramp to the 101 North freeway in Los Angeles. With approximately 50,000 data points, this dataset focuses on comparing regular traffic patterns to those during baseball games at Dodger Stadium. However, it presents several challenges such as missing values due to sensor malfunctions and the presence of noise and irregularities caused by external factors like weather, holidays, and other special events that introduce significant variations in traffic patterns that are not explicitly marked in the data.

The Calit2 dataset captures building inflows and outflows from the main entrance of the Calit2 Institute at the University of California, Irvine, over a three-month period, recorded every 30 minutes, amounting to roughly 4,000 data points. This dataset monitors human activity entering and exiting the building, reflecting both regular daily patterns and sudden changes during organized events.

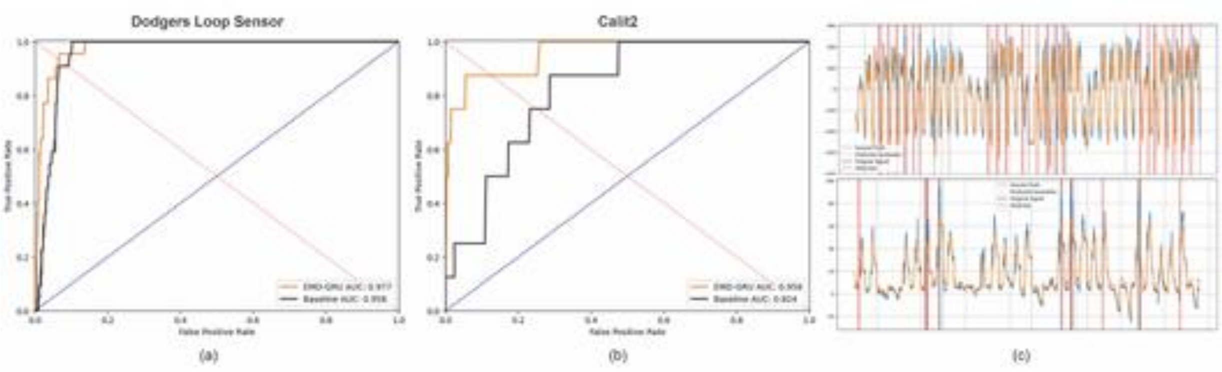
---

[1] https://archive.ics.uci.edu/dataset/157/dodgers+loop+sensor

[2] https://archive.ics.uci.edu/dataset/156/calit2+building+people+counts

Similarly to [11], we converted The Dodgers TS to 1-hour sampling intervals, by summing up consecutive 5-min samples. For the Calit2 dataset we combined the two streams (inflows - outflows) to a UTS by computing the cumulative sum on people in and out of the building each day. For the Dodgers dataset we employed standard EMD and for Calit2 the EEMD variant. In both cases, 2-layer GRUs were employed with the size of hidden state being 55 for the Dodgers dataset and 32 for the Calit2. GRUs were trained using a 48-hour time window of the IMFs. To test e challenging setting, the Initial 60% of Dodgers dataset and 50% of Calit2 dataset were used to train the respective GRUs with a reconstructive objective, without using the annotation to discard the anomalous events. For both datasets, 40% of the rest of the TS was used for validation and 60% was used for testing utilizing the provided annotations.

Figure 2 (a-b) illustrates the ROC curves derived from Dodgers and Calit2 datasets from the proposed model. For comparison, we evaluated a baseline model using a GRU corresponding to each dataset on the raw TS data prior to decomposition. The goal for both datasets is to detect the presence of the events, so for the Dodgers dataset we consider as true positive an anomaly detected within a 2-hour window prior or after the annotated time of the game, and for Calit2 1 hour prior and after an event. Figure 2 (c) illustrates the original TS, the forecasted TS for 1-sample ahead, the predicted anomalies based on the best threshold as defined in the validation set and the annotated anomalies. With obtained AUC of 0.977 for Dodgers and 0.958 for Calit2, it can be easily inferred that the EMD approach has significantly better performance from the baseline approach for both datasets (0.958 and 0.824 resp.), validating out assumption that by employing TS decomposition a lightweight recurrent model can offer remarkable performance in detecting anomalies in complex real-world scenarios.



(a) (b) (c)

Figure 2: Experimental results. ROC curve for Dodgers (a) and Calit2 (b) datasets for the proposed (orange) and baseline (black) models. (c) Forecasted TS and predicted anomalies detections vs ground truth for Dodgers (top) and Calit2 (bottom) datasets. Blue: Actual TS, Orange: Estimated TS, Red: GT, Dashed: Predicted Anomalies.

5 Conclusions

In this study, we introduced a novel forecasting-based anomaly detection method that combines EMD with GRUs for anomaly detection in univariate time series data. By decomposing the time series into simpler IMFs, we enhanced the GRU’s ability to predict short-term variations, improving anomaly detection performance in small computational budget. The experimental results on real-world smart city datasets demonstrate that the proposed method effectively detects anomalies, handling both noise and multiscale seasonal trends. In the future we intend to investigate alternative mechanisms of prediction such a Temporal Neural Networks and extend the proposed scheme into multivariant time series.

References

[1] Z. Zamanzadeh Darban, G. I. Webb, S. Pan, C. Aggarwal, and M. Salehi, “Deep Learning for Time Series Anomaly Detection: A Survey,” ACM Comput. Surv., Aug. 2024, doi: 10.1145/3691338.

[2] G. Rilling, P. Flandrin, and P. Gonçalves, “On empirical mode decomposition and its algorithms,” presented at the Proceedings of IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing NSIP-03, 2003.

[3] J. Ma and S. Perkins, “Online novelty detection on temporal sequences,” in KDD ’03. New York, NY, USA: Association for Computing Machinery, Aug. 2003, pp. 613–618.

[4] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” Neural Comput., vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[5] P. Malhotra, L. Vig, G. M. Shroff, and P. Agarwal, “Long Short Term Memory Networks for Anomaly Detection in Time Series,” presented at the European Symposium on Artificial Neural Networks, 2015.

[6] L. Bontemps, V. L. Cao, J. McDermott, and N.-A. Le-Khac, “Collective Anomaly Detection Based on Long Short-Term Memory Recurrent Neural Networks,” in Future Data and Security Engineering, T. K. Dang, R. Wagner, J. Küng, N. Thoai, M. Takizawa, and E. Neuhold, Eds., Cham: Springer International Publishing, 2016, pp. 141–152.

[7] 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,

[8] Y. He and J. Zhao, “Temporal Convolutional Networks for Anomaly Detection in Time Series,” J. Phys.: Conf. Ser., vol. 1213, no. 4, p. 042050, Jun. 2019.

[9] M. Thill, W. Konen, and T. Bäck, “Time Series Encodings with Temporal Convolutional Networks,” in Bioinspired Optimization Methods and Their Applications, B. Filipič, E. Minisci, and M. Vasile, Eds., Cham: Springer International Publishing, 2020, pp. 161–173.

[10] H. Song, D. Rajan, J. J. Thiagarajan, and A. Spanias, “Attend and diagnose: clinical time series analysis using attention models,” in AAAI’18/IAAI’18/EAAI’18. New Orleans, Louisiana, USA: AAAI Press, Feb. 2018, pp. 4091–4098.

[11] W. Wu et al., “Developing an Unsupervised Real-Time Anomaly Detection Scheme for Time Series With Multi-Seasonality,” IEEE Transactions on Knowledge and Data Engineering, vol. 34, no. 9, pp. 4147–4160, Sep. 2022.

[12] K. Cho, B. van Merriënboer, D. Bahdanau, and Y. Bengio, “On the Properties of Neural Machine Translation: Encoder–Decoder Approaches,” in Proceedings of SSST-8, Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 103–111.

[13] S. J. Taylor and B. Letham, “Forecasting at scale,” PeerJ Inc., e3190v2, Sep. 2017.

[14] Y.-H. Wang, C.-H. Yeh, H.-W. V. Young, K. Hu, and M.-T. Lo, “On the computational complexity of the empirical mode decomposition algorithm,” Physica A: Statistical Mechanics and its Applications, vol. 400, pp. 159–167, Apr. 2014.

[15] M. E. Torres, M. A. Colominas, G. Schlotthauer, and P. Flandrin, “A complete ensemble empirical

mode decomposition with adaptive noise," in 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 2011, pp. 4144–4147.

[16] R. Fontugne, P. Borgnat, and P. Flandrin, "Online Empirical Mode Decomposition," in ICASSP 2017, Mar. 2017, pp. 4306–4310.

[17] A. Ihler, J. Hutchins, and P. Smyth, "Adaptive event detection with time-varying poisson processes," in KDD '06. New York, NY, USA: Association for Computing Machinery, Aug. 2006, pp. 207–216.