Journée Régionale des Doctorants en Sciences de la Mer et Automatique

8 juillet 2024 au Centre Universitaire Le Musée à Boulogne-sur-Mer

Anomaly Detection for IoT multivariate time series data with statistical and machine learning techniques

Thi Thuy Van Nguyen[a,b], Kim Phuc Tran[a], Cédric Heuchenne[b] , Guillaume Tartare[a]
[a]University of Lille, ENSAIT, GEMTEX - Laboratoire de Génie et Matériaux Textiles, F-59000 Lille, France
[b]HEC Management School - University of Liège, Belgium

Introduction

- Good decisions are critical for business success in today's competitive global economy.
- IoT applications aid in decision-making and automating daily tasks for convenience
- Embedded AI integrates AI into electronic systems, such as home automation systems, smart wearables, and autonomous vehicles
- Anomaly detection enables quick identification of anomalies or unexpected patterns for effective decision-making.
- Control charts (in SPC): well-established and reliable method, easy construction, and user-friendly usage.
- Deep learning (DL) has capability to automatically learn and adapt to complex patterns and variations in data

→ uncover anomalies that may be overlooked by traditional statistical methods

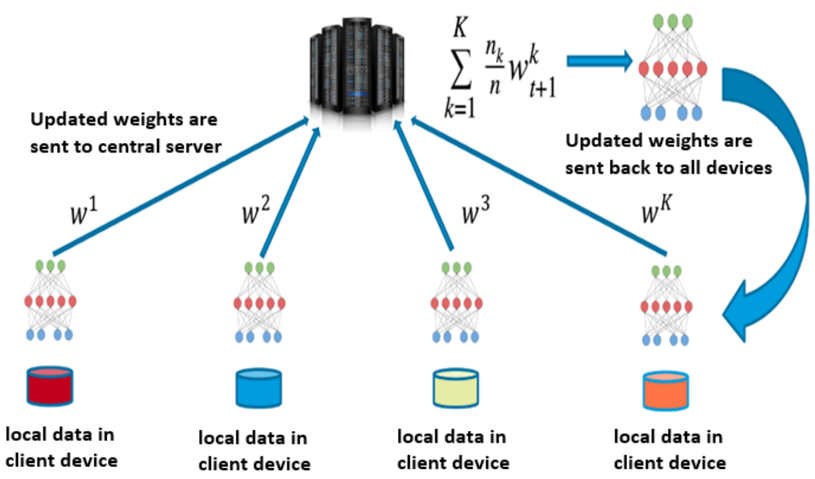
→ provide a higher level of accuracy compared to traditional statistical methods.

Problems

- **Control charts** rely on data distribution assumptions & may struggle with complex data → **Deep Learning** can handle these struggles, however:
- Large amounts of industrial data cannot be easily collected from a single silo.
- Collecting labeled data for training DL models can be expensive and time-consuming
- DL models are like black-box, hard to interpret
- DL models can be too large and resource-intensive → inefficiencies in data storage and processing for certain applications
- Training DL models normally requires a large amount of data and it can lead to privacy concerns.

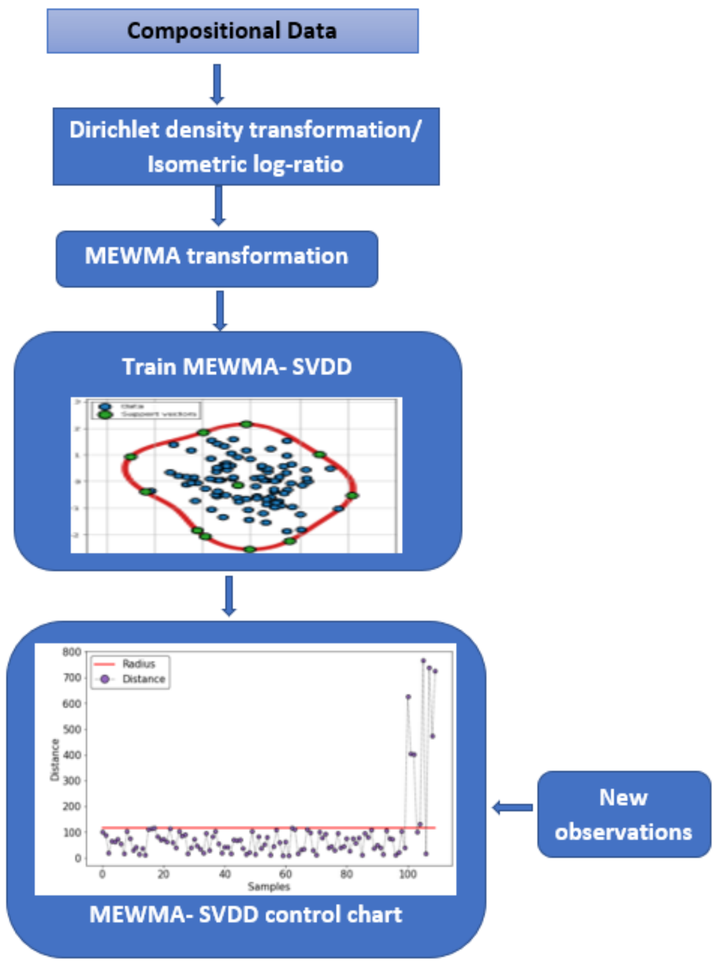
Objective

Develop a new Explainable Anomaly Detection (EAD) framework combining embedded AI with IoT technology in a federated setting for IoT Multivariate time series data using statistic and DL techniques



Results

SVDD control charts based on MEWMA technique for monitoring CoDa



- Dirichlet density transformation: eliminating inherent constraints, reducing data dimensionality in CoDa
- SVDD-EWMA control charts perform well without data distribution requirement
- Two SVDD-EWMA control charts outperform the classical MEWMA-CoDa for monitoring CoDa in terms of ARL.

Transformer-based VAE Anomaly Detection Approach for ECG Monitoring Healthcare System

Transformer-based VAE-MEWMA-SVDD Framework

- Combine transformer-based variational autoencoder (VAE) and MEWMA-SVDD control chart.
- Leverages the architecture introduced by Raza et al. for feature extraction using reconstruction error vectors
- Integrates MEWMA-SVDD for effective anomaly detection and false alarm management.

Experiment and Results

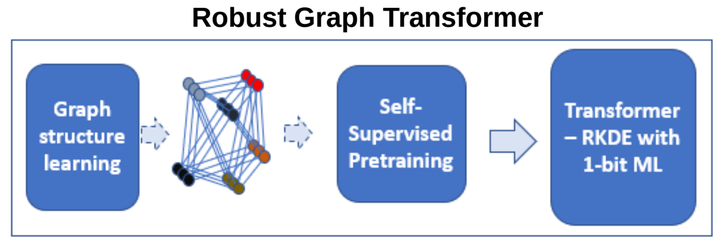
- **Dataset:** Hybrid dataset combining PhysioNet's BIDMC Congestive Heart Failure Database and MIT-BIH Normal Sinus Rhythm Database; total of 5,000 heartbeats (2,919 normal, 2,081 anomalies)
- **Training:** focused on normal data to detect anomalies via reconstruction loss
- **Results:**
  - ✓ Achieved test accuracy of 98.77%.
  - ✓ Reduced false alarms: 1 misclassification compared to 2 in Raza et al.'s result
  - ✓ No prior knowledge of data distribution is required.

Developing RGT-SVDD framework

Propose developing a **robust graph transformer** (RGT) combined with a MEWMA-SVDD control chart for AD in a federated setting:

- SSL: Utilize unlabeled data - learn representations, capture underlying patterns in data without explicit labels
- RGT: extract important information from the data
- 1-bit ML: Reduces model size & computational requirements
- MEWMA-SVDD: detect abnormal objects
- XAI (SHAP/XAI by design): enhances transparency, and trustworthiness
- Federated learning: handle data privacy

Robust Graph Transformer



Proposed framework

Conclusion

- Proposed SVDD-EWMA cc overcome data distribution assumptions, outperform traditional MEWMA-CoDa
- Future direction: an RGT in a federated setting together with XAI to detect anomalies in complex data

Reference

[1] T.T. Van Nguyen, C. Heuchenne, and K. P. Tran. "**Machine learning for compositional data analysis in Support of the Decision Making Process**", *Machine Learning and Probabilistic Graphical Models for Decision Support Systems*. CRC Press, 2021. 184-215

[2] T.T. Van Nguyen, C. Heuchenne, K.D. Tran and K.P. Tran. "**A Novel Transformer-Based Anomaly Detection Approach for ECG Monitoring Healthcare System**", *International Conference on Safety and Security in IoT.* Springer Nature Switzerland, 2023. 111-129