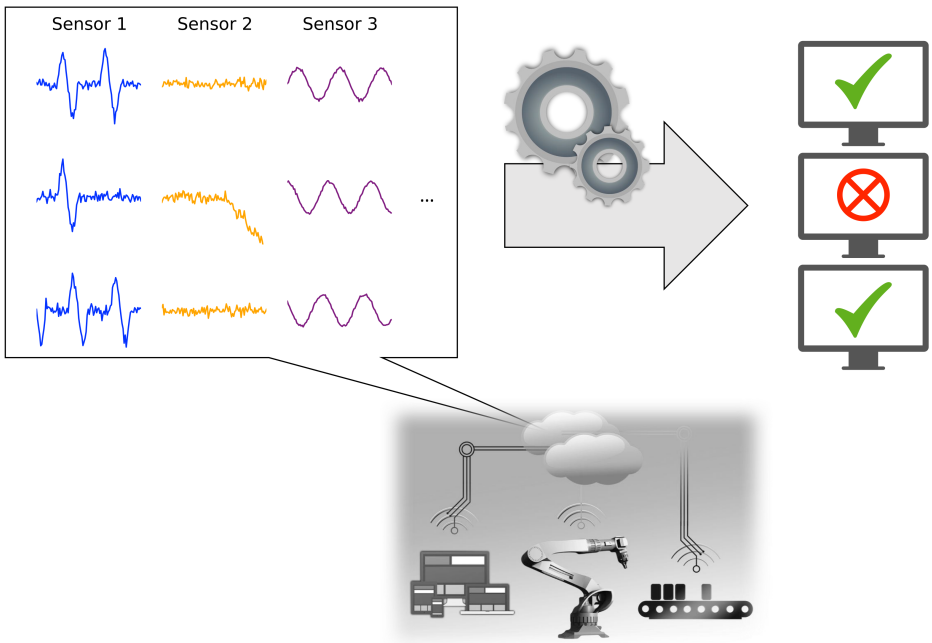
Knowledge Discovery and Data Mining
Prof. Dr. Emmanuel Müller
Master project Summer 2018

Unsupervised Anomaly Detection: Representation Learning for Predictive Maintenance over Time



Project description

Anomaly detection is the task of identifying patterns and points in the data that are *highly deviating*, *unexpected*, or *unusual* in comparison to the overall data distribution or in the context of a specific application. In this project we study the detection of **anomalies in multivariate time series data using deep learning techniques**. The task of finding anomalies in (multivariate) time series is posed in a great variety of applications such as industrial fault and damage detection for predictive maintenance, detecting abnormal patterns and extreme values in stock markets, monitoring network traffic for intrusion detection and cybersecurity, or detecting events in IoT sensor systems.

The challenge in many of those applications is that the types of anomalies and the patterns they show in the data are either unknown or very costly to obtain. In contrast, time series data recordings of "normal" states are usually cheap and plenty. For this reason, supervised methods cannot be used for detecting anomalies in most applications. Instead, an unsupervised approach has to be undertaken with the aim of describing the "normal". Deviations from such a description of the normal are then likely to be anomalies given that the characterization is accurate. Finding such an accurate characterization is not an easy task and still an open challenge for multivariate time series. In this project, we will tackle this challenge with representation learning approaches. Additional challenges that are posed in time series anomaly detection are (1) the different types of anomalies (e.g. shifts, bursts, trends, etc.) and (2) the properties of data streams (e.g. non-stationarity), multi-scale data and high-frequency measurements.

Imagine a company invested in a very complex and expensive machine for which it would like to set up a system for monitoring its health. To implement such a system, multiple sensors measuring

different quantities (e.g. temperature, pressure, etc.) are installed. Simulating all failure scenarios would be very costly and as the machine is very complex, many cases of failure would also be difficult to anticipate. On the other hand, monitoring the machine in its normal working state is cheap. The goal then is to characterize the normal operating state or states of the machine so that potential failures can be inferred by finding anomalous patterns which deviate from the description of the normal. In order to find a good characterization, data from the multiple sensors has to be considered as some anomalous patterns could only be detected from a combination of variables. For example the leakage in a pipe may only be detected if there is both, an abrupt decrease in pressure as well as temperature, which is why the multivariate view is crucial.

In this project, we will study different approaches to unsupervised anomaly detection on multivariate time series using the representation learning capabilities of deep recurrent neural networks. The key for learning a good characterization of normal states is to find the relevant features and factors of variation describing those states. In order to extract and learn those features, we will explore different recurrent architectures (e.g. recurrent neural networks (RNNs) or Long Short-Term Memory networks (LSTMs)) that have shown to be very potent predictors on sequences and time series, but have not yet been explored much for the task of anomaly detection. To evaluate the approaches, a comparison with traditional state-of-the-art anomaly detection methods will be conducted on synthetic as well as real-world data from predictive maintenance applications.

Project goals

1. Read and understand state-of-the-art deep recurrent networks literature.
2. Research existing approaches for anomaly detection in time series using deep learning.
3. Develop novel methods for anomaly detection in time series using deep learning.
4. Implement and evaluate the proposed methods using synthetic and real-world data.
5. Write a submission-ready scientific publication for a top data mining conference/workshop.
6. If successful, present the work at an international conference.

Required skills

- Interest in anomaly detection on time series and deep learning.
- Interest in doing research on the topic and in developing novel methods for the task.
- Good programming skills (preferably experience with deep learning libraries)
- Basic knowledge on anomaly detection concepts and time series are beneficial but not required.

Team and dates

The team will consist of 3–5 students. The project will start mid April, 2018. A preliminary meeting (February 6th, 2018 at 1 p.m. – 2 p.m., Office E1-01.2) will be held to answer potential questions.

Contacts

Please contact us to apply for the project or ask for more information:

- Prof. Dr. Emmanuel Müller (`emmanuel.mueller@hpi.de`) - Office: E1-01.2
- Lukas Ruff (`lukas.ruff@hpi.de`) - Office: E1-03.1
- Fabian Geier (`fabian.geier@hpi.de`) - Office: E1-02.3

Bibliography

[1] P. Malhotra, L. Vig, G. Shroff, and P. Agarwal. Long short term memory networks for anomaly detection in time series. In *ESANN, 23rd European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, 2015.

[2] D. T. Shipmon, J. M. Gurevitch, P. M. Piselli, and S. T. Edwards. Time series anomaly detection: Detection of anomalous drops with limited features and sparse examples in noisy highly periodic data. *arXiv preprint arXiv:1708.03665*, 2017.

[3] M. Soelch, J. Bayer, M. Ludersdorfer, and P. van der Smagt. Variational Inference for On-line Anomaly Detection in High-Dimensional Time Series. *ArXiv e-prints*, Feb. 2016.