Offline anomaly detection with causal discovery

Sebastiano D'Arconso - Tommaso Del Prete

January 2024

1 Introduction

Offline anomaly detection with causal discovery aims to identify unusual or abnormal patterns in data using techniques from causal discovery. In the case of offline detection the analysis is applied to historical data, rather than in real-time. **Anomaly detection** involves identifying patterns in data that devies significantly from the expected behaviour. These anomalies could be indicative of errors, outliers or interesting events in the dataset. **Causal discovery** aims to uncover cause-and-effect relationships within a set of variables. It explores the directionality of relationships between variables, helping understand which variables influence others and in what manner.

2 Datasets

2.1 Swat

2.2 Boat

2.3 Pepper

2.4 Fourier analysis

3 Experiments

To perform offline anomaly detection with causal discovery we used the **PCMCI** algorithm, with both **PARCORR** and **KNN**. PCMCI (partial correlation and mutual information-based causal inference) is a framework that combines partial correlation and mutual information to infer causal relationships between variables. The idea is to use partial correlation to measure strength of linear relationships between variables while considering the influence of other variables. Mutual information is also employed to capture non-linear dependencies between variables. The PCMCI algorithm involves constructing a graph, often referred as causality graph, where nodes represent variables, and edges represent inferred causal relationships. The strength and direction of the edges are determined based on the partial correlation and mutual information values.

- **PARCORR**: Parcorr stands for partial correlation and it's a function used for computing partial correlation coefficients between time series variables. The partial correlation coefficient between two variables measures the strength and direction of their linear relationship while controlling for the influence of other variables in the dataset. It quantifies the association between two variables after removing the linear effect of all other variables.
- **KNN**: KNN is a machine learning algorithm, in the context of causal discovery it can be employed to enhance the identification of causal relationships between variables. The idea is to leverage the concept of proximity in feature space, it wants to obtain complex non-linear dependencies. KNN, used with PCMCI, considers the local structure of the data. KNN identifies the nearest neighbors of each data point in the feature space, and the relationships between variables can be assessed by taking into account the values of their neighbors. This allows the algorithm to capture non-linear dependencies that might be missed by traditional statistical measures.

3.1 Hyperparameters

To perform these experiments there were some hyperparameters to set.

- **tau max**: represents the maximum time lag or delay that the algorithm consider when examining the potential causal relationship between variables, it determines the maximum number of time steps by which one variable might influence another. In other words the algorithm will assess the relationship between variables at lags from 1 to tau max.
- **tau min**: represents the minimum time lag considered in the analysis. It sets lower limit for the range of time lags examined. Relationships with time lags smaller than tau min are not considered during time analysis.
- **pc alpha**: is the significance level or threshold used for conditional independence tests. This parameter determines the threshold for considering variables as conditionally independent.
- **alpha**: is the significance level for the causal graph.

4 Results

We did different analysis on all datasets, changing parameters such as **pc alpha**, **alpha**, and the number of neighbours in the case of **knn**, to see how the analysis would change.

4.1 Swat

For the swat dataset we had to analyze two different datasets, **normal** and **attack**.

- Swat, normal, linear, parcorr, pcalpha=0.01: link to image
- Swat, normal, linear, parcorr, pcalpha=0.05: link to image
- Swat, attack, linear, parcorr, pcalpha=0.01: link to image
- Swat, attack, linear, parcorr, pcalpha=0.05: link to image
- Swat, normal, not linear, knn=10, pcalpha=0.05: link to image
- Swat, normal, not linear, knn=10, pcalpha=0.03: link to image
- Swat, attack, not linear, knn=10, pcalpha=0.05: link to image
- Swat, attack, not linear, knn=10, pcalpha=0.03: link to image

4.2 Boat

- Boat, normal, linear, parcorr, pcalpha=0.01: link to image
- Boat, normal, linear, parcorr, pcalpha=0.03: link to image
- Boat, normal, linear, parcorr, pcalpha=0.05: link to image
- Boat, gpsdown, linear, parcorr, pcalpha=0.01: link to image
- Boat, gpsdown, linear, parcorr, pcalpha=0.03: link to image
- Boat, gpsdown, linear, parcorr, pcalpha=0.05: link to image
- Boat, stucked, linear, parcorr, pcalpha=0.01: link to image
- Boat, stucked, linear, parcorr, pcalpha=0.03: link to image
- Boat, stucked, linear, parcorr, pcalpha=0.05: link to image
- Boat, normal, not linear, knn=10, pcalpha=0.03: link to image
- Boat, normal, not linear, knn=3, pcalpha=0.05: link to image
- Boat, normal, not linear, knn=5, pcalpha=0.05: link to image
- Boat, normal, not linear, knn=10, pcalpha=0.05: link to image
- Boat, gpsdown, not linear, knn=5, pcalpha=0.03: link to image
- Boat, gpsdown, not linear, knn=10, pcalpha=0.03: link to image
- Boat, gpsodwn, not linear, knn=3, pcalpha=0.05: link to image
- Boat, gpsdown, not linear, knn=5, pcalpha=0.05: link to image
- Boat, gpsdown, not linear, knn=10, pcalpha=0.05: link to image
- Boat, stucked, not linear, knn=5, pcalpha=0.03: link to image

- Boat, stucked, not linear, knn=10, pcalpha=0.03: link to image
- Boat, stucked, not linear, knn=3, pcalpha=0.05: link to image
- Boat, stucked, not linear, knn=5, pcalpha=0.05: link to image
- Boat, stucked, not linear, knn=10, pcalpha=0.05: link to image

4.3 Pepper

- Pepper, normal, linear, parcorr, pcalpha=0.01: link to image
- Pepper, normal, linear, parcorr, pcalpha=0.03: link to image
- Pepper, normal, linear, parcorr, pcalpha=0.05: link to image
- Pepper, normal, linear, parcorr, pcalpha=0.05, alpha=0.000005: link to image
- Pepper, wheels, linear, parcorr, pcalpha=0.01: link to image
- Pepper, wheels, linear, parcorr, pcalpha=0.03: link to image
- Pepper, wheels, linear, parcorr, pcalpha=0.05: link to image
- Pepper, wheels, linear, parcorr, pcalpha=0.05, alpha=0.00005: link to image
- Pepper, joint, linear, parcorr, pcalpha=0.01: link to image
- Pepper, joint, linear, parcorr, pcalpha=0.03: link to image
- Pepper, joint, linear, parcorr, pcalpha=0.05: link to image