Bosch Project - Team Bosch Bisch

Claudio Milanesi, Luca De Felice, Matteo Riva

1 Root Cause Analysis of Part Failures Using Causal Graph Discovery

1.1 Objective

This project aims to identify the root causes of part failures by modeling causal relationships among measurements collected at multiple production stations. The significance of this analysis lies in its potential to improve production quality, reduce downtime, and enhance operational efficiency. By understanding the causal factors of failures, we can provide actionable insights for production adjustments and preemptively address the conditions that lead to part failure.

1.2 Data

The dataset comprises 2,500 instances of successful parts and 2,500 instances of failed parts, with measurements taken by sensors at five sequential stations. Causal dependencies in this multi-station setup are assumed to flow unidirectionally, where measurements at an earlier station can influence those at subsequent stations but not vice versa.

1.3 Approach

Our approach to the problem can be described in three main blocks:

1. Causal Graph Discovery: the goal was to find the Directed Acyclic Graph (DAG) considering the sequential constraints of measurements in the dataset;
2. Utilize the estimated DAG to detect the root causes of the observed change in Station5_mp_85;
3. Implement a web application to allow potential Bosch employees to utilize our technology in a smart and intuitive way.

1.3.1 Causal Graph Discovery

The causal graph discovery step involved setting some constraints on the DAG matrix to be predicted, such that some edges are excluded a priori from the possible causality relationships. For example, a sensor in Station 5 could not influence (have causality) sensors in the previous stations.

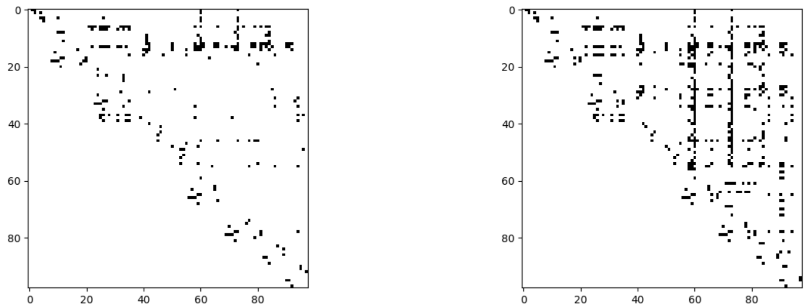

Figure 1: [Left] Predicted DAG matrix. [Right] Ground truth.

We tested many different approaches, such as: PC, DAGGNN, Lingam, GES, FCI, and DirectLingam. We mostly used python libraries, e.g. Causal Discovery Toolbox, gCastle, causal-learn. The result of our estimated matrix compared with the ground truth provided by the Bosch team is shown in Figure 1. Our prediction does not match the ground truth perfectly, but it serves an initial estimate. The DAG is found 'DAG_MATRIX.txt'.

1.3.2 Utilize the estimated DAG to detect the root causes of the observed change in Station5_mp_85

To detect the parent root causes for changes in the values of Station5_mp_85 sensor, we made use of our DAG matrix as possible causality graph. We thus used the DoWhy python package to create a Structural Causal Model, that we used to fit the data from the sensor readings. Then we sample some of the entries from our data and compute anomaly attribution scores. This approach allows us to obtain scores of the variables that affected the sensor reading at each sensor. Specifically, at sensor Station5_mp_85, we obtain that the problems are originated in station 2. In the folder 'Claudio/Pickles', you can find a list of scores for each sensor. These scores are store as a dictionary in a pickle file, wherekeys represent ancestors, and the values are arrays whose means represent the weight between the key and the sensor of the file.

1.3.3 Grpah Visualization

To visualize the graph, we implemented a web application that shows the causality weigths for each sensor. Screenshots are presented in Figure 2. Once a sensor from the dropdown list is clicked, all the related causes are colored based on how much they influenced the outcome of the selected sensor.

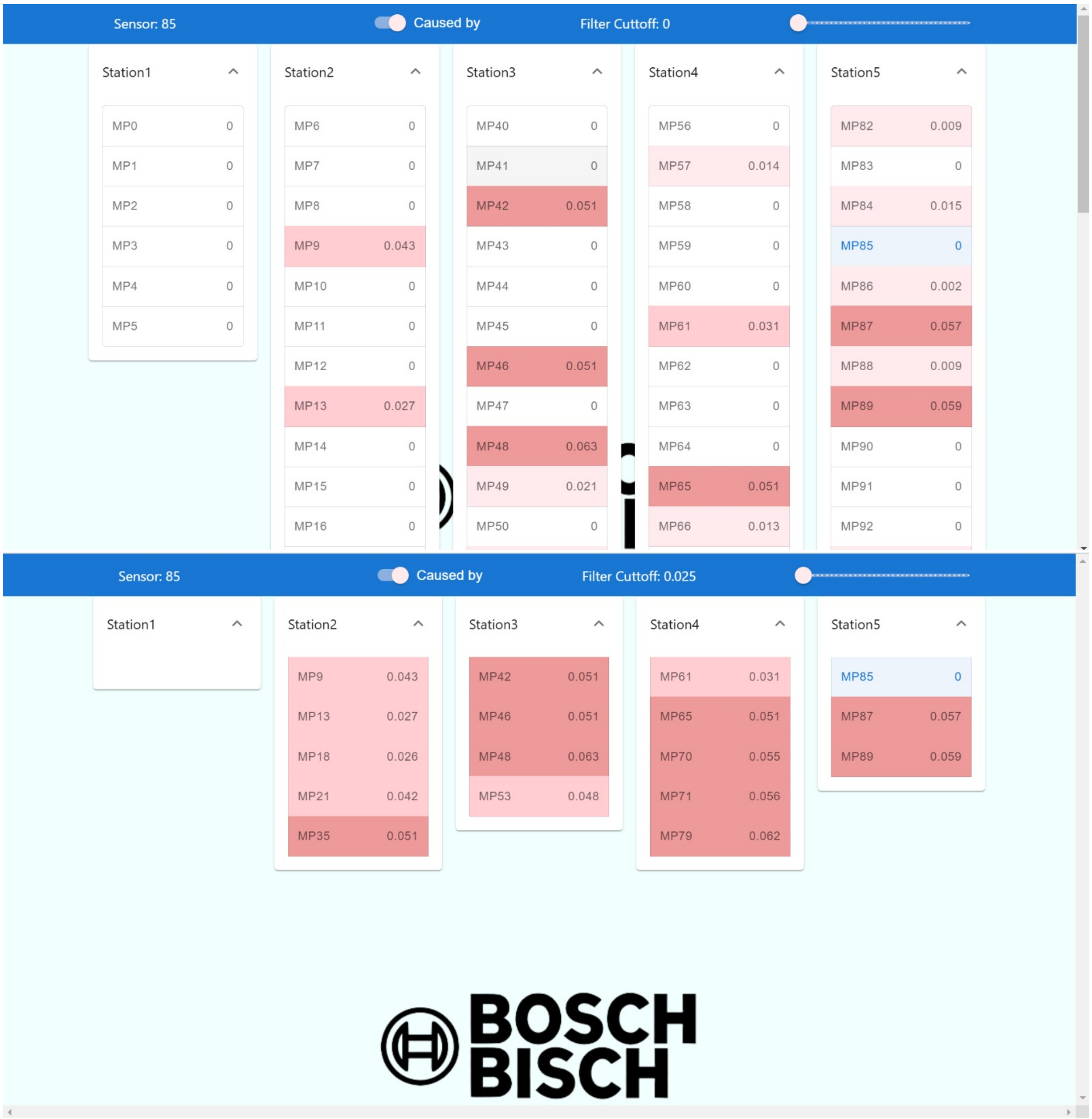

Figure 2: Web application.