

ABSTRACT

Data is collected permanently during operation, to prevent machine failures by detecting anomalies as quickly as possible. In order to pursue the best predictive maintenance, it is of interest not only knowing the device, but also knowing the environment of a machine. This leads to the question if it is possible to predict from the collected data whether two monitored machines are in a kind of neighborhood or not.

In this paper a *Bayesian Network* was built to identify if two machines are in spatial proximity to each other, are functionally integrated into a common process or even if two sensors are collecting data from the same machine. For this purpose, we define different metrics between two machines, which can be calculated from the collected data. Each kind of neighborhood and each metric corresponds to a single node in the Bayesian Network. Each edge of the network corresponds to a causality between kinds of neighborhood and metrics.

A particular challenge of the work was that there was no data available which indicates whether two machines are in a kind of neighborhood or not. These target variables were therefore generated on the basis of an analysis of the metadata using string metrics. Since the target values cannot be assigned with certainty, the nodes, which represent a kind of neighborhood, were optimized by an *Expectation-Maximization-Algorithm*.

In conclusion, the outcome of the Bayesian Network was compared to the outcome of the analysis of the metadata. This comparison allows the conclusion, that the network is able to find neighborhoods, but has difficulties distinguishing different kinds of neighborhood. Furthermore, the network cannot be trained completely independently of the analysis of the metadata.

Keywords: Bayesian-Network, Expectation-Maximization-Algorithm, machine, neighborhood, metric, Predictive Maintenance, Process-data

ZUSAMMENFASSUNG

Um Ausfällen bei Maschinen vorzubeugen, werden während des laufenden Betriebs Daten erhoben, damit Anomalien möglichst schnell erkannt werden können. Um möglichst gute *Predictive-Maintenance*-Modelle für Maschinen zu erstellen, ist es von Interesse, nicht nur die Geräte selbst, sondern auch deren Umgebung zu kennen. Es stellt sich die Frage, ob man aus den erhobenen Daten bestimmen kann, ob sich zwei überwachte Maschinen in einer Nachbarschaft zueinander befinden.

In dieser Arbeit wurde ein *Bayes-Netz* erstellt, welches bestimmen soll, ob zwei Maschinen in räumlicher Nähe zueinander stehen, funktional in einen gemeinsamen Prozess eingebunden sind oder ob sogar zwei Sensoren Daten von derselben Maschine erheben. Dafür wurden verschiedene Metriken zwischen den Maschinen definiert, welche sich aus den erhobenen Daten berechnen lassen. Jede Art der Nachbarschaft und jede Metrik entspricht einem einzelnen Knoten im Bayes-Netz. Die Kanten des Netzes entsprechen Kausalitäten zwischen den Nachbarschaften und den Metriken.

Eine besondere Schwierigkeit der Arbeit lag darin, dass für die untersuchten Daten nicht bekannt ist, ob die Maschinen sich tatsächlich in einer Nachbarschaft zueinander befinden. Diese Zielvariablen wurden daher aufgrund von einer Analyse der Metadaten über String-Ähnlichkeiten bestimmt. Da die Zielvariablen so nicht mit Sicherheit bestimmt werden können, wurden die Knoten, welche die Nachbarschaften bestimmen sollen, durch einen *Expectation-Maximization-Algorithmus* angepasst.

Die Ergebnisse des Bayes-Netzes wurden anschließend mit den Ergebnissen der Metadatenanalyse verglichen. Es stellte sich heraus, dass das Netz Nachbarschaften sehr gut erkennen kann, jedoch Schwierigkeiten damit hat, verschiedene Arten von Nachbarschaften zu unterscheiden. Außerdem kann das Netz nicht unabhängig von der Metadatenanalyse trainiert werden.

Schlüsselwörter: Bayes-Netz, Expectation-Maximization-Algorithmus, Maschinen, Nachbarschaft, Metrik, Predictive Maintenance, Prozessdaten