

Bosch SI Whitepaper: „Everything all right with your devices?“

Ein Kommentar von SyroCon-Consultant Sebastian Kößler

Mit der stetig wachsenden Zahl an vernetzten Geräten und der kontinuierlich voranschreitenden Digitalisierung wächst auch der Anreiz zur Realisierung datengetriebener Anwendungen. Mit dem vorliegenden [Whitepaper](#) gibt Bosch SI Einblick in seine diesbezügliche Projekterfahrung und definiert die Kernelemente einer erfolgreichen Datenanalyse im IoT-Umfeld. Als Anwendungsbeispiel zieht das Software- und Systemhaus der Bosch Group einen Ansatz zur Echtzeitanalyse kontinuierlicher Sensordatenströme mittels Verfahren zur Anomalie-Erkennung heran.

Die Verwendung von Anomalie-Erkennung ermöglicht, frühzeitig auf fehlerhaftes Verhalten aufmerksam zu machen. Dadurch können manuelle Analysen gezielt getriggert werden, die Maschinen-Ausfällen vorbeugen. Das ist insbesondere für Lösungen zur vorrausschauenden Wartung im Industrie-Umfeld wichtig. Technisch wird die Anomalie-Erkennung über statistische Auswertungen und die Anwendung von maschinellen Lernalgorithmen realisiert.

DAS ANWENDUNGSBEISPIEL

Anhand des erwähnten Anwendungsbeispiels fasst Bosch im Folgenden die drei wichtigsten Schritte für ein erfolgreiches Vorgehen zusammen:

- 1) Datenvorverarbeitung
- 2) Datenanalyse
- 3) Datenvisualisierung

Die Vorverarbeitung wird im Whitepaper als Schlüsselement zur Realisierung einer datengetriebenen Anwendung identifiziert. Nach eigener Erfahrung mache dieser Schritt den Großteil des zeitlichen Projektaufwands aus. Während der Vorverarbeitung werden die Daten bereinigt und normiert. Anschließend werden anwendungsspezifische Merkmale extrahiert, wobei dies laut Bosch nur in enger Zusammenarbeit mit einem Domain-Experten gelingen kann.

In dem betrachteten Beispiel werden aus den kontinuierlichen Sensordatenströmen mithilfe von standardisierten Regeln Events extrahiert. Bosch verweist hierbei auf Regeln, die in der statistischen Prozesslenkung zur Optimierung von Produktions- und Serviceprozessen verwendet werden. Nach der erfolgreichen Extrahierung der Events werden feste Zeitfenster analysiert und mit den auftretenden Eventtypen sowie deren

Häufigkeiten beschrieben. Damit die gewonnenen Informationen vergleichbar sind, werden die Häufigkeiten in Bezug auf die Gesamthäufigkeiten normiert.

ALGORITHMEN ZUR ANOMALIE-ERKENNUNG

Unabhängig von seinem Anwendungsbeispiel stellt Bosch im Anschluss verschiedene Algorithmen zur Anomalie-Erkennung vor. Darunter Multidimensional Scaling (MDS), Cluster-Analysen und Support Vector Machines (SVMs). Diese Algorithmen können aus historischen Daten Charakteristiken lernen und diese als Entscheidungsmodell abbilden. Bei der Interpretation der Ergebnisse muss laut Bosch allerdings Folgendes beachtet werden:

Ein Datenpunkt setzt sich aus vielen verschiedenen Merkmalen zusammen und lässt sich daher nicht ohne weiteres visualisieren. Deshalb werden häufig Verfahren verwendet, welche die verschiedenen Merkmale auf zwei Dimensionen reduzieren. In dieser Form können die Daten dann beispielsweise in einem Streudiagramm grafisch dargestellt werden. Des Weiteren ist es möglich, die Modelle zur Anomalie-Erkennung als Entscheidungsgrenzen innerhalb dieser Grafiken abzubilden. Generell fällt es uns im zweidimensionalen Raum leichter, Datenpunkte in Gruppen einzuordnen. Im Umkehrschluss können somit auch Ausreißer problemlos erkannt werden. Dafür müssen jene Datenpunkte identifiziert werden, die nicht eindeutig einer bestimmten Gruppe zuzuordnen sind.

FAZIT UND AUSBLICK

Das Whitepaper beschreibt zwar die wichtigsten Aufgaben zur Realisierung datengetriebener Anwendungen, versäumt aber, den Bogen zur automatisierten Analyse zu spannen. Stattdessen wird darauf verwiesen, die Ergebnisse der Algorithmen visuell unter Berücksichtigung von Domain-Wissen zu interpretieren. Hierfür müssen jedoch die gefundenen Anomalien mit Metadaten angereichert werden, um dem Auswerter den richtigen Kontext zur Verfügung zu stellen.

Abschließend gibt Bosch den Ausblick, dass die Modelle zur Anomalie-Erkennung durch die gewonnenen Erkenntnisse der visuellen Analyse verbessert werden können. Im besten Fall könnte aus meiner Sicht durch die manuelle Klassifizierung von relevanten Anomalien sogar eine ausschließlich algorithmenbasierte Anomalie-Erkennung realisiert werden. Zu diesem Zweck könnten mit ausreichend gesammelten, annotierten Daten Klassifikationsverfahren (supervised learning) verwendet werden, um die Interpretation der gefundenen Anomalien zu automatisieren.