

Applying Unsupervised Algorithms for Changepoint Detection in Manufacturing Data

Diplomand



Stefan Keller

Einleitung: Im Zeitalter von Industrie 4.0 und dem Internet der Dinge (IoT) fallen in industriellen Prozessen enorme Datenmengen an. Viele Unternehmen sammeln und speichern diese Daten, stehen jedoch oft vor der Herausforderung, deren volles Potenzial auszuschöpfen. Dabei könnten diese Daten durch fortschrittliche Analysen in wertvolle Erkenntnisse umgewandelt werden. Eine häufige Schwierigkeit besteht darin, dass die Datenmengen so gross sind, dass es kaum möglich ist, einen umfassenden, gelabelten Datensatz zum trainieren eines Klassifikators zu erstellen. Daher ist der Bedarf an unsupervised Methoden offensichtlich. Diese Methoden sind entscheidend, um Daten analysieren und Muster erkennen zu können, ohne auf vordefinierte Labels angewiesen zu sein.

Die Daten bestehen aus Sensordaten, die durch einen Prozess erzeugt werden, bei dem für jedes hergestellte Teil ein Datenpunkt aufgezeichnet wird, z. B. die maximal erforderliche Kraft bei einem Fügeprozess. Die wiederholte Durchführung dieses Prozesses erzeugt eine Zeitreihe, die das zeitliche Verhalten der Messgrösse darstellt. Diese Zeitreihe enthält naturgemäss ein Rauschen, das durch Faktoren wie Toleranzen und Messungenauigkeiten verursacht wird. Die zentrale Frage dabei ist, wann sich die Zeitreihe so verändert, dass Handlungsbedarf besteht. Das bedeutet, die Messgrösse zeigt Schwankungen, aber ab welchem Punkt wird eine solche Schwankung als signifikante Änderung interpretiert, die auf eine mögliche Abweichung oder Störung im Prozess hinweist? Es gilt zu bestimmen, ab wann eine Verschiebung in der Messgrösse nicht mehr als zufällige Variation, sondern als bedeutsamer Changepoint betrachtet werden sollte, der Massnahmen erfordert.

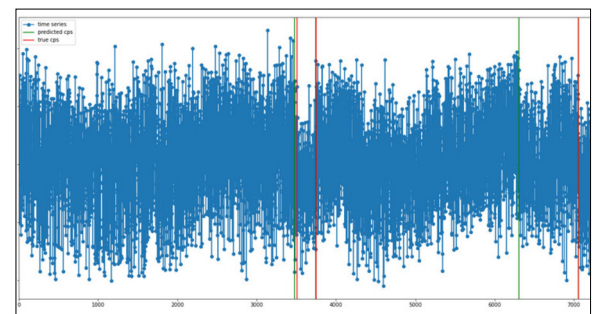
Ziel der Arbeit: Ziel dieser Arbeit ist es, den aktuellen Stand von unsupervised Algorithmen zur Erkennung von Changepoints in Zeitreihen zu analysieren, indem verschiedene Algorithmen an einem realen Datensatz getestet werden. Über mehrere Wochen hinweg wurden Sensordaten gesammelt und annotiert, um die Übereinstimmung der Vorhersagen der Algorithmen mit den Einschätzungen eines Experten zu überprüfen. Ein Evaluierungsschema wird entwickelt, um die Leistung der Algorithmen zu bewerten. Darüber hinaus wird eine Reihe von heuristischen Regeln vorgeschlagen, um die identifizierten Changepoints hinsichtlich ihrer Kritikalität und Bedeutung zu klassifizieren.

Vorgehen: Die Untersuchung der Zeitreihen erfolgt auf einer abstrakten Ebene, unabhängig vom spezifischen zugrunde liegenden Prozess oder Sensortyp, und ist auf einen allgemeinen Anwendungsfall ausgerichtet. Dieser Ansatz gewährleistet eine hohe Vielseitigkeit und Anwendbarkeit in verschiedenen Szenarien und

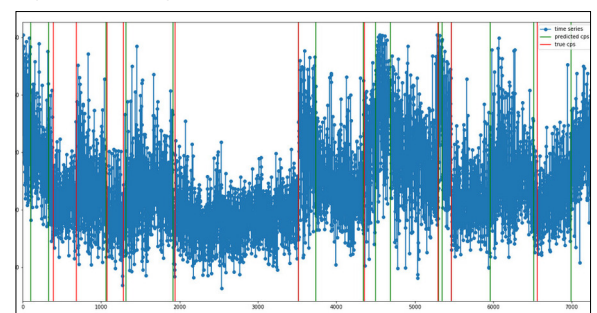
Branchen. Dadurch können verallgemeinerte Methoden und Rahmenwerke entwickelt werden, die sich flexibel an unterschiedliche Kontexte anpassen lassen. Dies verbessert die Skalierbarkeit und Praktikabilität von Methoden zur Erkennung von Changepoints und macht sie für eine breite Palette von Anwendungen nutzbar.

In den Abbildungen sind drei beispielhafte Zeitreihen mit gelabelten Changepoints (rot) und vorhergesagten Changepoints (grün) dargestellt.

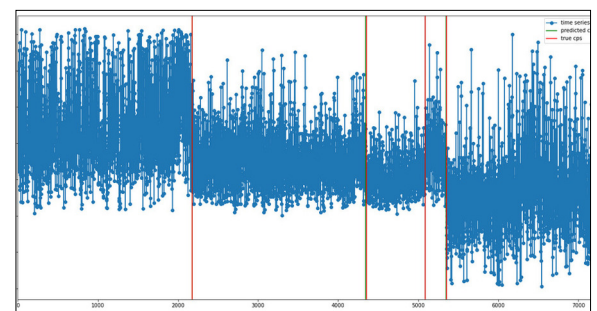
Eine Schwierigkeit liegt im Labeln der Daten, da kein festes Kriterium existiert, ab wann eine Änderung signifikant ist.
Eigene Darstellung



Die Vorhersagen stimmen im Allgemeinen gut mit den Experteneinschätzungen überein.
Eigene Darstellung



Bedeutende Changepoints werden fast immer korrekt erkannt.
Eigene Darstellung



Referent

Prof. Dr. Carlo Bach

Korreferentin

Dr. Kathrin Plankensteiner,
Fachhochschule Vorarlberg

Themengebiet

Data Science,
Computer Science