



ANOMALIEDETEKTION IN DER GERÄTEERPROBUNG MITTELS ANALYSE VON ZEITREIHEN

NICK SCHREIBER (MASTERSTUDIUM INFORMATIK)

Betreuer: Prof. Dr. Marcel Tilly, Prof. Dr. Markus Breunig

Diese Masterarbeit untersucht, inwieweit Anomalien in Getränkeprozessen von Kaffeevollautomaten anhand aufgezeichneter Sensordaten zuverlässig erkannt werden können. Ziel ist die Entwicklung eines datengetriebenen Verfahrens zur Anomaliedetektion, das zur frühzeitigen Identifikation von unsachgemäßem Verhalten eingesetzt werden kann, um eine vorausschauende Wartung zu ermöglichen. Als Datengrundlage dienen reale Sensordaten aus der Geräteerprobung. Die Arbeit behandelt die notwendigen Schritte der Datenvorverarbeitung und stellt einen Ansatz vor, mit dem zunächst ungelabelte Zeitreihendaten gruppiert und anschließend effizient manuell gelabelt werden können. Der Schwerpunkt der Arbeit liegt auf der Entwicklung und dem Vergleich verschiedener Autoencoder-Modelle, die die Struktur fehlerfreier Getränkeprozesse erlernen und Abweichungen davon als Anomalien erkennen. Die entwickelten Verfahren erzielen vielversprechende Ergebnisse und zeigen, dass in den Zeitreihen auftretende Anomalien zuverlässig erkannt werden können.

Motivation

Kaffeevollautomaten sind komplexe Systeme, in denen mechanische und elektronische Komponenten während der Zubereitung eines Getränks präzise zusammenarbeiten müssen. Bereits geringfügige Abweichungen, beispielsweise durch Verschleiß, Verkalkung oder mechanische Defekte, können den Getränkeprozess beeinflussen und die Getränkequalität oder die Funktionsfähigkeit des Geräts beeinträchtigen. Während jedes Getränkeprozesses zeichnen zahlreiche Sensoren kontinuierlich verschiedene Messwerte, beispielsweise Temperatur oder Pumpenleistung, auf. Die daraus entstehenden Zeitreihen bilden den tatsächlichen Ablauf detailliert ab und enthalten wertvolle Informationen über den Zustand des Geräts.

Trotz einer umfangreichen Datenbasis an Aufzeichnungen existiert bislang kein Verfahren, das automatisch beurteilen kann, ob ein Getränkeprozess erwartungsgemäß verlaufen ist oder nicht. Eine zuverlässige Anomaliedetektion würde es ermöglichen, fehlerhafte Geräte frühzeitig zu identifizieren und potenzielle Defekte bereits vor einem Ausfall zu erkennen. Dadurch könnte die Geräteerprobung gezielt unterstützt und gleichzeitig die Grundlage für Verfahren der vorausschauenden Wartung geschaffen werden.

Datenbasis und Datenvorverarbeitung

Die Grundlage der Arbeit bilden Sensordaten aus der Geräteerprobung von Kaffeevollautomaten. Die aufgezeichneten Zeitreihen weisen je nach Gerät, Getränktyp und Geräteeinstellungen erhebliche Unterschiede auf. Dadurch entsteht eine hohe Datenvielfalt mit zahlreichen Einflussfaktoren, die eine allgemeingültige Anomaliedetektion deutlich erschweren. Um die Problemstellung unter definierten Rahmenbedingungen untersuchen zu können, wird der entwickelte Ansatz auf ausgewählte Sensordaten sowie die Getränke Milchschaum und Kaffee eingegrenzt. Der betrachtete Datensatz umfasst dabei etwa 100.000 Getränkeprozesse.

Bevor maschinelles Lernen eingesetzt werden kann, müssen die Sensordaten aufbereitet werden. Da die Sensoren ihre Messwerte nicht in festen Zeitabständen erfassen, werden die Zeitreihen auf eine einheitliche Abtastrate interpoliert. Anschließend werden die Messwerte skaliert, sodass die Daten aller Sensoren trotz unterschiedlicher Wertebereiche vergleichbar sind und als gemeinsame Eingabe für die nachfolgenden Verfahren genutzt werden können.

Erstellung eines gelabelten Datensatzes

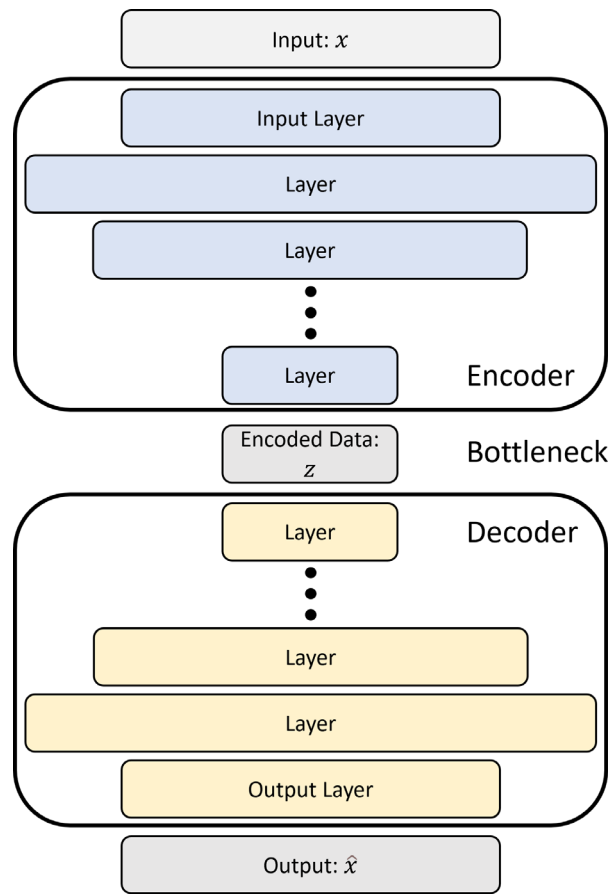
Eine zentrale Herausforderung besteht darin, dass zunächst nicht bekannt ist, welche der aufgezeichneten Zeitreihen fehlerfrei sind und welche eine Anomalie zeigen. Für die Entwicklung und Bewertung der Anomaliedetektion wird jedoch ein gelabelter Datensatz benötigt. Um nicht alle Getränkeprozesse einzeln manuell labeln zu müssen, werden die Zeitreihen zunächst mithilfe eines Clustering-Verfahrens gruppiert. Als Ähnlichkeitsmaß wird die Dynamic Time Warping (DTW)-Distanz verwendet. Der DTW-Algorithmus bestimmt eine optimale Zuordnung zwischen den Datenpunkten zweier Zeitreihen und berücksichtigt dabei zeitliche Verzerrungen und Verschiebungen. Dadurch können strukturell ähnliche Verläufe auch dann als ähnlich erkannt werden, wenn sie zeitlich versetzt auftreten oder gestaucht bzw. gestreckt sind. Da reale Abläufe in der Regel nicht exakt gleich lange dauern, ist die DTW-Distanz für den Vergleich solcher Zeitreihen besser geeignet als ein einfaches punktwises Distanzmaß. Anschließend werden die entstandenen Cluster manuell als fehlerfrei oder als Anomalie klassifiziert. Dadurch entsteht mit vergleichsweise geringem manuellem Aufwand ein vollständiger gelabelter Datensatz.

Anomaliedetektion mittels Autoencoder

Ein Autoencoder ist ein neuronales Netz, das Eingabedaten zunächst in eine kompakte Repräsentation überführt und daraus anschließend versucht, die ursprünglichen Daten wieder zu rekonstruieren. Die Abbildung veranschaulicht den

ROSENHEIMER INFORMATIKPREIS
INF-MASTER

grundlegenden Aufbau eines Autoencoders. Die Eingabe (Input) ist die Zeitreihe eines Getränkeprozesses. Der Encoder besteht aus mehreren Schichten (Layern), die den Input transformieren und schrittweise auf eine niedrigdimensionale Repräsentation, den sogenannten Bottleneck, komprimieren. Anschließend versucht der Decoder aus dieser komprimierten Repräsentation die ursprüngliche Zeitreihe wiederherzustellen.



Schematische Darstellung eines Autoencoders

Für die Anomaliedetektion wird ein Autoencoder ausschließlich mit fehlerfreien Getränkeprozessen trainiert. Dadurch lernt das Modell die charakteristischen Merkmale des Nor-

malverhaltens und kann entsprechende Zeitreihen mit einem geringen Rekonstruktionsfehler nachbilden. Treten hingegen Anomalien auf, weichen diese vom gelernten Normalverhalten ab und können in der Regel nicht korrekt rekonstruiert werden. Dies führt zu einem erhöhten Rekonstruktionsfehler, der als Maß zur Erkennung von Anomalien verwendet werden kann. Wichtig ist die Wahl des Schwellenwerts, ab dem ein Rekonstruktionsfehler als Anomalie eingestuft wird.

Ergebnisse

Im Rahmen der Arbeit wurden drei Autoencoder-Architekturen auf Basis von Long Short-Term Memory (LSTM)-, Transformer- und Convolution-Layern entwickelt und hinsichtlich ihrer Rekonstruktions- und Anomaliedetektionsleistung verglichen. Alle drei Architekturen erzielten gute Ergebnisse, wobei der Convolution-basierte Autoencoder insgesamt die beste Rekonstruktionsleistung zeigt und bei gleicher Bottleneck-Größe die höchste Übereinstimmung zwischen Input und Output erreicht. Der Transformer-basierte Autoencoder rekonstruiert die grundlegende Struktur der Zeitreihen, schwankt jedoch hochfrequent um die tatsächlichen Werte. Der LSTM-basierte Autoencoder rekonstruiert die Grundstruktur ebenfalls, verliert dabei jedoch Details.

Auch bei der Anomaliedetektion erzielt die Convolution-basierte Architektur die besten Ergebnisse. Auf Milchschaum-Testdaten von Kaffeevollautomaten, die nicht für das Training verwendet wurden, erreicht das Modell eine Genauigkeit von 97 % sowie einen F1-Score von 0,91. Dieses sehr gute Ergebnis ist jedoch teilweise darauf zurückzuführen, dass es bei Milchschaum einen häufig auftretenden Anomaliefall gibt, der deutlich vom Normalverhalten abweicht und daher vergleichsweise leicht zu identifizieren ist.

Für den Getränkeprozess Kaffee wird eine Genauigkeit von 92 % und ein F1-Score von 0,78 erreicht. Bei Anomalien, die sich nur geringfügig von fehlerfreien Zeitreihen unterscheiden, sowie bei uneindeutigen Grenzfällen, zeigt der entwickelte Ansatz Schwächen. Der Grund dafür ist, dass der Rekonstruktionsfehler häufig zu gering ausfällt, um eine zuverlässige Klassifikation als Anomalie zu ermöglichen. Insgesamt sind die erzielten Ergebnisse dennoch sehr positiv und bestätigen, dass Anomalien in Getränkeprozessen zuverlässig auf Basis von Sensordaten erkannt werden können.