

Mess- und Datenplanung für Modelle des maschinellen Lernens an Bestandsanlagen

Christopher Schnur^{1,2}, Steffen Klein¹, Anne Blum², Tizian Schneider^{1,2}, Rainer Müller^{2,3} und Andreas Schütze^{1,2}

¹Lehrstuhl für Messtechnik, Universität des Saarlandes, Saarbrücken, Deutschland

²Zentrum für Mechatronik und Automatisierungstechnik gGmbH, Saarbrücken, Deutschland

³Lehrstuhl für Montagesysteme, Universität des Saarlandes, Saarbrücken, Deutschland

Kontakt: c.schnur@lmt.uni-saarland.de

Einleitung

Um maschinelles Lernen im industriellen Kontext zuverlässig anwenden zu können, spielen neben den entsprechenden Algorithmen auch die Daten und deren Qualität eine entscheidende Rolle. Während die Algorithmen und deren Entwicklung momentan im Vordergrund stehen, wird den Daten, deren Erfassung und Annotation insbesondere in der Industrie nicht genügend Aufmerksamkeit gewidmet. In diesem Beitrag liegt der Fokus auf der produzierenden Industrie und speziell auf der Montage.

Aktuell herrscht in der Industrie eine hohe Diskrepanz zwischen idealen Daten, welche einen hohen Informationsgehalt beinhalten und bestmöglich die Realität abbilden, und realen Daten, die an den Anlagen aufgezeichnet werden. Ausschlaggebend für diese Diskrepanz sind diverse Herausforderungen, die Wilhelm et al. in die vier Bereiche produktionspezifische (z. B. heterogene Produktionslandschaften), datenspezifische (z. B. fehlende Eindeutigkeit und Zuordenbarkeit der Daten), analytische (z. B. Ursache-Wirkungs-Analyse) und softwarespezifische Herausforderungen (z. B. fehlende Datenschnittstellen) einteilt [1]. Die Bewältigung dieser Herausforderungen resultiert oftmals in einem gesteigerten Aufwand bei der Aufbereitung der Daten von schätzungsweise 80%, während lediglich 20% des Arbeitsaufwandes für die eigentliche Modellierung des maschinellen Lernmodells benötigt werden (80/20-Regel) [2]. Oftmals besteht auch ein Interessenskonflikt zwischen den direkten Zielen der Industrie, wie bspw. einer hohen Produktionsstückzahl bei gleichzeitig geringen Ausschusszahlen, und den Anforderungen an hochqualitative Daten.

Methode

Im Rahmen der Projekte „Messtechnisch gestützte Montage“ (MessMo) und „iTecPro – Erforschung und Entwicklung von innovativen Prozessen und Technologien für die Produktion der Zukunft“ wurde daher eine Checkliste entwickelt, die den Anwender bei der Bearbeitung maschineller Lernprojekte (ML-Projekte) an Bestandsanlagen der Montage unterstützt [3]. Nachfolgend werden beide Projekte vorgestellt.

Projekt MessMo

Im Projekt MessMo wurden drei Serienprodukte eines großen deutschen Herstellers hinsichtlich des Toleranzmanagements, der statistischen und physikalischen Analyse,

dem maschinellen Lernen und dem Condition Monitoring vom Lehrstuhl für Messtechnik und dem Lehrstuhl für Montagesysteme der Universität des Saarlandes untersucht. Zwei dieser Produkte wurden vollständig und ein Produkt teilweise auf Brownfield-Anlagen (Bestandsanlagen) gefertigt. Im Projekt bestätigte sich einerseits die 80/20-Regel und andererseits waren weitergehende Datenanalysen aufgrund einer geringen Datenqualität nicht möglich. Die Hauptgründe für die geringe Datenqualität waren:

- Keine eindeutige Zuordnung der Daten zu den Produkten bzw. Prozessen möglich (z.B. durch fehlende oder doppelte IDs).
- Fehlende Synchronisation zwischen Stationen und Prozessen.
- Fehlende Schnittstellen zum Auslesen von Daten und proprietäre Datenformate.
- Prozesstrends wie z. B. Mittelwertverschiebungen ändern die ursprüngliche Verteilungsfunktion.
- Systematische Muster und unzureichende Datenbasis.
- Unvollständige Daten und Metadaten

Abb. 1 zeigt die schematische Nachbildung einer Messreihe eines realen Prozesses mit Mittelwertverschiebungen und systematisch auftretenden Messfehlern. Die dargestellte Messreihe enthielt keinen Ausschuss (fehlerhafte Produkte) und eine Ursachen-Rückführung war nur mit hohem Aufwand möglich.

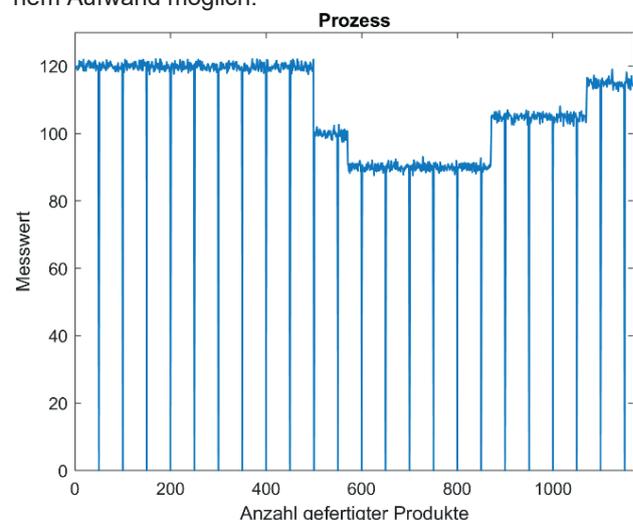


Abb. 1: Beispielhafte Ergebnisse eines realen, kritischen Prozesses.

Ein Ergebnis des Projekts MessMo war daher die Identifizierung der Notwendigkeit einer Mess- und Datenplanung im industriellen Kontext. Aus den gewonnenen Erkenntnissen der beiden Lehrstühle wurde zusammen mit dem Projektpartner eine erste Checkliste zur Steigerung der Datenqualität entwickelt.

Projekt iTecPro

Im Projekt iTecPro, in dem die Erforschung und Entwicklung von innovativen Prozessen und Technologien für die Produktion der Zukunft im Fokus stand, wurde ein Arbeitspaket explizit der Mess- und Datenplanung gewidmet. In diesem Kontext wurde die Checkliste aus dem Projekt MessMo grundlegend überarbeitet und hinsichtlich der FAIR-Prinzipien ergänzt. Die FAIR-Prinzipien stammen aus dem Forschungsdatenmanagement und stehen für die Attribute [4, 5]:

- Findable (auffindbar)
- Accessible (zugänglich)
- Interoperable (interoperabel)
- Reusable (wiederverwendbar)

Grundlegend basiert die Checkliste auf den sechs Phasen (Verständnisaufbau vom Business, Verständnisaufbau von den Daten, Datenaufbereitung, Modellbildung, Evaluation und Anwendung) des Cross-Industry Standard Process for Data Mining (Crisp-DM) [6].

Ergebnisse

Ähnlich dem Crisp-DM enthält die Checkliste Punkte, die parallel und iterativ durchgeführt werden müssen. Die erarbeitete Checkliste umfasst dabei die Themengebiete:

- Vorbereitung und Projektplanung
- Mess- und Datenplanung
- Datenaufnahme
- Datenprüfung und Datenbereinigung
- Datenauswertung und Modellbildung
- Projektabschluss

Nachfolgend werden die Inhalte der einzelnen Themengebiete kurz erläutert.

1. Vorbereitung und Projektplanung

Ein wichtiger Schritt zur Durchführung von ML-Projekten ist eine gründliche Vorbereitung und Projektplanung. In diesem Themengebiet werden die Ziele des Projektes definiert und dessen Rahmenbedingungen festgelegt.

2. Mess- und Datenplanung

Bei der Mess- und Datenplanung handelt es sich in der Regel um einen iterativen Prozess, bei dem zunächst Prozesswissen aufgebaut werden muss, welches dann zur Gestaltung des Aufbaus der Daten genutzt werden kann. Weiterhin müssen geltende Standards (wie z. B. Normen) recherchiert und beachtet werden.

3. Datenaufnahme

Bevor die eigentliche Datenaufnahme startet, wird eine Testmessung mit anschließender Beurteilung der Datenqualität durchgeführt. Auf diese Weise können frühzeitig Änderungen und Korrekturen angewandt werden. Anschließend folgt eine Langzeitdatenaufnahme mit regelmäßigen

Prüfintervallen, um die ordnungsgemäße Funktion der Sensoren und eine konstant hohe Qualität der Daten zu gewährleisten.

4. Datenprüfung und Datenbereinigung

Nach Abschluss der Langzeitdatenaufzeichnung muss der Datensatz auf Plausibilität und Vollständigkeit (vor allem der Metadaten) geprüft und ggf. ergänzt werden. Fehlerhafte Messungen gilt es zu bereinigen.

5. Datenauswertung und Modellbildung

Der geprüfte und bereinigte Datensatz kann nun analysiert werden. Zunächst wird ein Verständnis durch statistische Analysen und Visualisierungsmethoden über die Daten aufgebaut. Anschließend kann ein maschinelles Lernmodell gebildet und abschließend validiert werden. Aufgrund der hohen Anzahl verfügbarer Algorithmen wird die Modellbildung im Rahmen der Checkliste nur grundlegend behandelt.

6. Projektabschluss

Im Projektabschluss werden die erlangten Erkenntnisse in einem Abschlussbericht dokumentiert, Lessons Learned formuliert und alle notwendigen Unterlagen zentral und leicht auffindbar abgelegt.

Aufbau der Checkliste

Jedes der aufgezählten Themengebiete wird mittels eines kurzen Abschnittes eingeführt, bevor anschließend spezifischere Unterkapitel mit den jeweiligen Checkpunkten folgen. Weiterhin besitzt jeder der Checkpunkte eine Beschreibung. Abb. 2 zeigt die Struktur der Checkliste. Bei den Checkpunkten wird zwischen Pflicht-Checkpunkten, deren Erfüllung zwingend erforderlich ist, und Best-Practice-Checkpunkten, deren Erfüllung optional, aber empfehlenswert ist, unterschieden. Weiterhin enthält die Checkliste zusätzliche Tipps und Hinweise, die den Anwender im entsprechenden Themengebiet unterstützen. Die entsprechenden Symbole sind in Abb. 3 dargestellt.

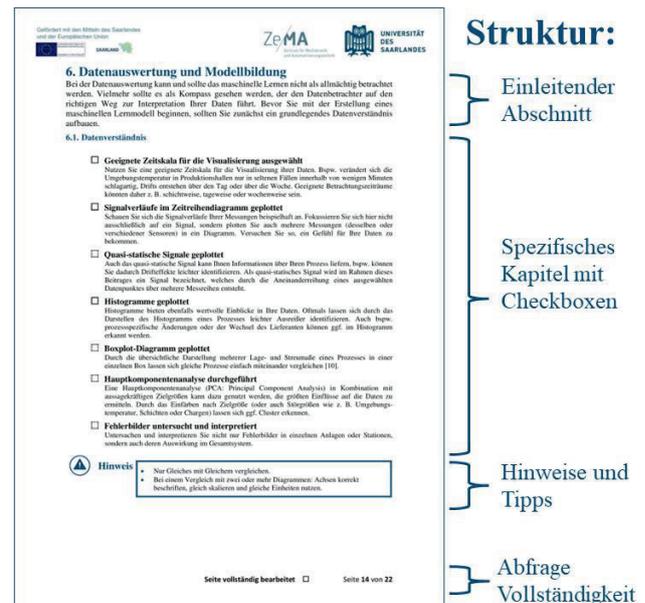


Abb. 2: Struktur der Checkliste.

Legende:



Abb. 3: Legende der Checkliste.

Die Checkliste wurde unter der Creative Commons Attribution 4.0 International Lizenz beim Online-Speicherdienst Zenodo veröffentlicht [7].

Diskussion

Obwohl die vorgestellte Checkliste zur Mess- und Datenplanung für Modelle des maschinellen Lernens an Bestandsanlagen anhand von Produktionsprozessen in der Montage entwickelt wurde, ist sie auch in anderen Domänen gültig und kann in diese übertragen werden. Um dies zu ermöglichen, geht die Checkliste einen Kompromiss hinsichtlich einer detaillierten Beschreibung der Vorgehensweisen zugunsten ihrer Allgemeingültigkeit ein. Durch weiterführende Literaturquellen bzw. Literaturempfehlungen (wie z.B. [8-11]) ist der Aufbau von tiefgehendem Verständnis möglich.

Literatur

1. WILHELM, Y.; SCHREIER, U.; REIMANN, P.; MITSCHANG, B.; ZIEKOW, H.: *Data Science Approaches to Quality Control in Manufacturing: A Review of Problems, Challenges and Architecture*. In: Dustdar S, ed. *Service-Oriented Computing*. Vol 1310. Communications in Computer and Information Science. Springer International Publishing; 2020:45-65. doi:10.1007/978-3-030-64846-6_4
2. MENDE, L.: *Merkmalentstehungs- und -wechselwirkungsanalyse (MEWA) für das prozessorientierte Toleranzmanagement in der Montage*. Dissertation. Universität des Saarlandes; 2020.
3. SCHNUR, C.; KLEIN, S.; BLUM, A.; SCHÜTZE, A.; SCHNEIDER, T.: *Steigerung der Datenqualität in der Montage*, 2022, Werkstattstechnik : wt, [submitted]
4. THE FAIRSHARING COMMUNITY; SANSONE, S.A.; MCQUILTON, P.: *FAIRsharing as a community approach to standards, repositories and policies*. *Nat Biotechnol*. 2019;37(4):358-367. doi:10.1038/s41587-019-0080-8

5. ALLEN, R.; HARTLAND, D.: *Fair In Practice - Jisc Report On The Findable Accessible Interoperable And Reuseable Data Principles*. Zenodo; 2018. doi:10.5281/ZENODO.1245568
6. CHAPMAN, P.; CLINTON, J.; KERBER, R.; et al.: *CRISP-DM 1.0: Step-by-step data mining guide*, 2000, Zugriff am 27 Oktober 2022, URL: <https://www.kde.cs.uni-kassel.de/wp-content/uploads/lehre/ws2012-13/kdd/files/CRISPWP-0800.pdf>
7. SCHNUR, C.; KLEIN, S.; BLUM, A.: *Checkliste – Mess- und Datenplanung für das maschinelle Lernen in der Montage*, 2022, doi:10.5281/ZENODO.6943476 URL: <https://zenodo.org/record/6943476#.Y10pYnZ-BxD8>
8. MATZKA, S.: *Crashkurs KI im Unternehmen: alles, was Sie über Data Science wissen müssen*, 1. Auflage, Haufe Group, 2021
9. METTERNICH, J.; BIEGEL, T.; CASSOLI, B.B.; et al.: *Leitfaden zur Expertise des Forschungsbeirats der Plattform Industrie 4.0 - Künstliche Intelligenz zur Umsetzung von Industrie 4.0 im Mittelstand*, Forschungsbeirat der Plattform Industrie 4.0 / acatech – Deutsche Akademie der Technikwissenschaften, 2021, Zugriff am 27 Oktober 2022, URL: https://www.vdma.org/documents/34570/4887324/Leitfaden_KI-fuer-I40-in-KMU.pdf/95df11bc-1c64-8a99-7e49-56a099e25ebc?t=1631272708548
10. RÖHLER, M.; HAGHI, S.: *Leitfaden Künstliche Intelligenz – Potenziale und Umsetzungen im Mittelstand*. VDMA Bayern, 2020, Zugriff am 27 Oktober 2022, URL: <https://www.vdma.org/documents/34570/1052572/Leitfaden+K%C3%BCnstliche+Intelligenz-Potenziale+und+Umsetzungen+im+Mittelstand.pdf/ce38a591-68cb-9775-101e-d7cad064b149?t=1615364023575>
11. BLUM, A.; WILHELM, Y.; KLEIN, S.; SCHNUR, C.; REIMANN, P.; MÜLLER, R.; SCHÜTZE, A.: *Ganzheitlicher Ablaufplan für wissensgetriebene Projekte des maschinellen Lernens in der Produktion*, 2022, *tm - Technisches Messen* 89, 363–383. <https://doi.org/doi:10.1515/teme-2022-0027>

Danksagung

Diese Arbeit wurde gefördert durch den Europäischen Fonds für regionale Entwicklung (EFRE) im Rahmen der Forschungsprojekte „iTecPro – Erforschung und Entwicklung von innovativen Prozessen und Technologien für die Produktion der Zukunft“ und „Messtechnisch gestützte Montage“.