

# Verbesserung mobiler Arbeitsprozesse mit Methoden von Big Data und Data Analytics

Heinrich Warkentin<sup>1</sup>, Thilo Steckel<sup>2</sup>, Alexander Maier<sup>3</sup> und Ansgar Bernardi<sup>4</sup>

**Abstract:** Im Gegensatz zu industriellen Fertigungsprozessen werden landwirtschaftliche Prozesse immer noch vorwiegend erfahrungsbasiert geplant und gesteuert. Wesentliche Ursachen hierfür sind die erschwerte Vorhersagbarkeit von Umgebungsbedingungen durch die hohe Umweltexposition und der Mangel an quantitativen Informationen zur Bestimmung präzise beschriebener Handlungsvorschläge. Die sich schnell entwickelnde Fähigkeit zur Erfassung, Übertragung, Analyse und anwendergerechten Aufbereitung großer, maschinenbezogenen Datenmengen in kurzen Zeiträumen liefert zumindest in Teilbereichen Verbesserungsmöglichkeiten. Dieser Beitrag beschreibt die Vorgehensweise zum Aufbau einer geeigneten Infrastruktur und der Analyse von großen Datenmengen am Beispiel des Mähdreschers mit dem Ziel der Produktivitätsverbesserung.

**Keywords:** Mähdrescher, Prozess, Big Data, Data Analytics, Anomalieanalyse

## 1 Einleitung

Beim Vergleich von Daten zur Einsatzproduktivität von Mähdreschern (t/h) fällt auf, dass beste Ergebnisse und durchschnittliche Ergebnisse um den Faktor 2 abweichen. Großmähdrescher zeigen unter optimalen Bedingungen mit über 80 Tonnen pro Stunde einen doppelt so hohen Durchsatz, wie vergleichbare Maschinen aus einer großen Stichprobe (CLAAS Telematics). Auch wenn diese Zahlen noch keine Aussage über den wirtschaftlichen Erfolg zulassen, ist erkennbar, dass Verbesserungspotenziale vorhanden sind. Typische Schwachstellen lassen sich zwar auf allgemeiner Ebene benennen. Eine nachhaltige Beseitigung ist aber nur möglich, wenn Prozesse in größerer Detailtiefe beschrieben und dadurch besser analysierbar gemacht werden. Im Rahmen des vom Bundesministerium für Forschung und Bildung geförderten Projektes AGATA werden eine Infrastruktur und mathematische Verfahren zur Identifikation und Analyse von Maschinen- und Prozessanomalien entwickelt, die im weiteren Fortgang eine solche Verbesserung ermöglichen.

---

<sup>1</sup> CLAAS E-Systems, Advanced Development, Bäckerkamp 21, 33330 Gütersloh, heinrich.warkentin@claas.com

<sup>2</sup> CLAAS E-Systems, Advanced Development, Bäckerkamp 21, 33330 Gütersloh, thilo.steckel@claas.com

<sup>3</sup> Fraunhofer IOSB-INA, Langenbruch 6, 32657 Lemgo, alexander.maier@iosb-ina.fraunhofer.de

<sup>4</sup> Deutsches Forschungszentrum für Künstliche Intelligenz, Trippstadter Str. 122, 67663 Kaiserslautern, ansgar.bernardi@dfki.de

## 2 Status quo bei der Verarbeitung von Maschinendaten

Aktuelle landwirtschaftliche Arbeitsmaschinen verfügen über eine CAN-Bus Architektur zur Erfassung von Betriebs- und Umgebungszuständen sowie Steuerung und Dokumentation von Prozessen. Ursprüngliche Motivation zur Entwicklung dieser Technologie war die Automatisierung von Selbstfahrern und Kombinationen von Zugmaschine und Anbaugerät. Anwendungen im Bereich Precision Farming und Telematik erweiterten diesen Ansatz, so dass Maschinendaten unter Anwendung drahtloser Kommunikationstechnologien auch auf verteilten, räumlich entfernten Systemen verarbeitet werden. Die Anzahl von in den Maschinen eingeführten Attributen hat im Laufe der Zeit stark zugenommen. Der Großmähdrescher CLAAS LEXION verfügt über einen Satz von über 4.300 Attributen. 70 dieser Attribute kommen - periodisch erfasst - in CLAAS-Telematics zur Anwendung. Die prinzipielle Möglichkeit zur ereignisgesteuerten Erfassung (teilweise < 25 ms) sämtlicher Attribute (leistungsfähige Logger), ihrer Übertragung (LTE) und stark parallelisierten Verarbeitung (Map-Reduce) führte zu der Überlegung, weitere Attribute zur Analyse, Bewertung und Verbesserung des Maschinenverhaltens zu nutzen. Aktuell genutzte Logger, Kommunikationsstandards unterhalb LTE und relationale Datenbanken sowie bislang genutzte Algorithmen stoßen bei der Verarbeitung großer Datenmengen mit heterogenen Strukturen an Grenzen. Folglich wurde im Rahmen des Projektes eine neue Systemarchitektur entwickelt und erprobt.

## 3 Nutzung von Big Data Technologien zur Maschinenanalyse

### 3.1 Erfassung von Maschinendaten

In 2 landwirtschaftlichen Betrieben wurden 7 Mähdrescher der Baureihe LEXION mit Datenloggern ausgestattet, die entsprechend einer vorgegebenen Konfiguration (Attribut, Erfassungsregel) definierte CAN-Nachrichten erfassen, vorverarbeiten, speichern und übertragen, sofern eine Netzverfügbarkeit gegeben ist. Durch den Austausch von Datenbankdateien zur Attributbeschreibung können die Logger für unterschiedliche Maschinen angepasst und genutzt werden. Die Datenerfassung erfolgt ereignisorientiert und ermöglicht eine genauere Zuordnung als in bislang praktizierten Verfahren mit fest definierten Intervallen. Die Hardware basiert auf einer ARM-Architektur, so dass Implementierungen auch auf dem kompatiblen Raspberry Pi vorgenommen werden können. Die Datenerfassung erfolgte in den Erntekampagnen 2014 und 2015. Im ersten Jahr wurden die Daten zunächst noch lokal abgespeichert, im Folgejahr via LTE direkt auf die Zielumgebung gestreamt.

Weiterhin wurden 5 der 7 Maschinen mit Videokameras (GoPro) ausgestattet, die das Vorfeld der Maschine erfassen. Die Kameras (1 Bild pro Minute) wurden für die spätere Verarbeitung mit der jeweiligen Systemzeit der Maschinen synchronisiert.

### 3.2 Übertragung von Maschinendaten

Für die Datenübertragung wurde ein LTE-Modul integriert. Maschinen- und Serverseitig erfolgte die Anbindung wahlweise mit KAFKA und MQTT-Schnittstellen. Die Netzverfügbarkeit war in den meisten Fällen gegeben, so dass die Daten unmittelbar zur Weiterverarbeitung verfügbar waren. Datentarife mit einem monatlichen Volumen von 5 Giga-byte reichten aus.

### 3.3 Serverseitige Verarbeitung der Maschinendaten

Zur serverseitigen Verarbeitung der Maschinendaten wurde ein Hadoop-System aufgebaut. Hadoop ist ein Framework, das die Verteilung rechenintensiver Prozesse auf mehrere Server ermöglicht. Hierzu wird mit Hilfe des *map-reduce*-Algorithmus ein großer Berechnungsprozess in kleine Prozesse zerlegt, zur Abarbeitung auf separate Server verteilt und danach wieder zusammen geführt. Im Projekt kamen 1 Management-Node und 4 Data-Nodes (5 Server, 24 Kerne je Knoten, 64 GB RAM je Knoten) zu Einsatz. Die Anzahl der Data-Nodes kann je nach Anforderung beliebig erhöht werden.

### 3.4 Analyse und Visualisierung von Maschinen- und Prozessdaten

Für die Analyse der erfassten Daten sind bekannte Anwendungen wie Excel nur sehr eingeschränkt geeignet. Die Beschränkung auf 1 Mio. Zeilen, fehlende Unterstützung für die Parallelisierung und der Mangel an fortgeschrittenen Algorithmen erfordert leistungsfähigere Werkzeuge. Die Online-Analyse erfolgt mit Apache-Storm und –Spark, die Offline-Analyse R und Python. Sie integrieren sich nahtlos in der Serverumgebung. Eine Erweiterung des Hadoop-Frameworks über Module ist möglich.

Ziel der Datenanalyse ist es, durch kombinierte Betrachtung vielfältiger Parameter Hinweise auf mögliche Ursachen für die gute (oder schlechte) Prozessleistung zu finden und so bisher unbekanntes Verbesserungsmöglichkeiten zu identifizieren. In diesem Sinne „interessante“ Auffälligkeiten können alleine auf Basis der vorhandenen Daten durch Anwendung von „überwachen“ oder „unüberwachten“ Verfahren des Maschinellen Lernens gefunden werden. Unüberwachte (rohdaten-basierte) Verfahren untersuchen Daten auf statistische Zusammenhänge. Vertreter dieser Kategorie sind z.B. Zeitreihenanalysen oder Clustering-Verfahren. Überwachte Verfahren werden zunächst mit bereits bewerteten Daten trainiert (also z.B. mit Datensammlungen, die eine von Domänenexperten als „hervorragend“ bewertete Performanz zeigen). In beiden Fällen werden anschließend eventuelle, vom erwarteten bzw. erlernten Verhalten abweichende Datenmuster erkannt („Anomalie-Detektion“), die Hinweise auf mögliche besondere Betriebszustände und damit Eingriffspotentiale geben. Für die Bewertung dieser Hinweise müssen dann Expertenmeinungen oder zusätzliche Informationen hinzugezogen werden. Ergänzend können modellbasierte Verfahren zum Einsatz kommen: Formalisiertes Modellwissen über den Drusch- oder Logistikprozess erlaubt die Berechnung von Vorhersagen und das sofortige Erkennen von Abweichungen.

### 3.5 Analyseergebnisse am Beispiel des Mähdruschprozesses

Am Beispiel der aufgezeichneten Mähdrusch-Datensätze konnte das Potential der automatisierten Datenanalyse exemplarisch erprobt werden: Zunächst werden aus der Vielzahl der verfügbaren Parameter einige geeignete ausgewählt. Unüberwachte Verfahren berechnen nun laufend Korrelation zwischen den Parametern und die zeitlichen Trends innerhalb der Parameter. Dabei fallen z.B. plötzliche Einbrüche im Massendurchsatz oder im Druschertrag (immer unter Berücksichtigung der Fahrgeschwindigkeit) als Anomalie in der Zeitreihe auf (siehe Abb. 1: Zeitreihendarstellung für Sensordaten aus Mähdrusch-Prozess). Ergänzend werten wir die aufgezeichneten Kamerabilder aus: Ein Bildanalysealgorithmus wurde auf das Erkennen von „Lagergetreide“ trainiert. Fällt nun die erkannte Durchsatz-Anomalie mit der im Bild erkannten Situation zusammen, gewinnen wir kleinräumig detailliertes Wissen über die mit dieser Situation verbundenen Performanceeinbußen. Gleichzeitig können die identifizierten Datenmuster ab sofort als bildunabhängige Identifikation der Problematik „Lagergetreide“ genutzt werden.

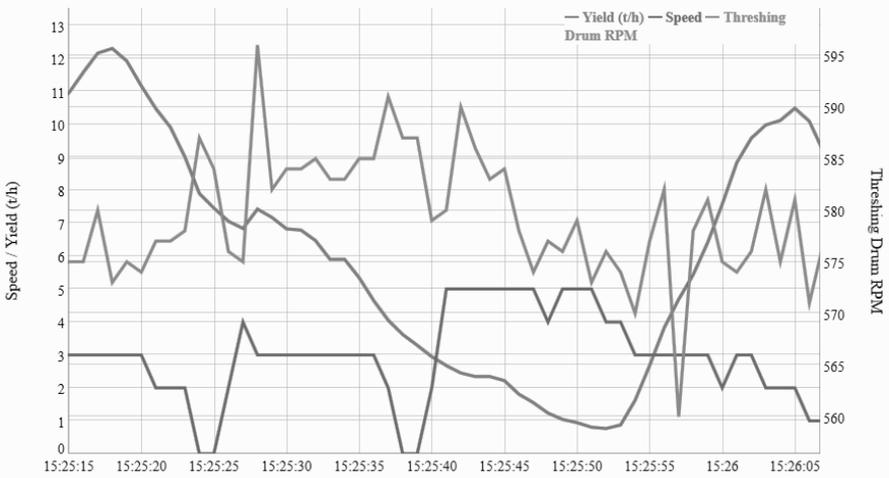


Abb. 1: Zeitreihendarstellung für Sensordaten aus Mähdrusch-Prozess

### 3.6 Ausblick

Beim Aufbau der Big Data-Plattform zu Identifikation von Prozessanomalien hat sich gezeigt, dass eine detaillierte Auseinandersetzung mit neuen Technologien, Methoden und Kompetenzen erforderlich ist. Nach ersten erfolgreichen Experimenten zeigt sich der Bedarf zur Entwicklung eines umfassenden Methoden-Baukastens, der Anforderungen aus allen Phasen des Maschinen-Lebenszyklus berücksichtigt.