

# Data Mining von multidimensionalen Qualitätsdaten aus einer computerintegrierten industriellen Fertigung zur visuellen Analyse von komplexen Wirkzusammenhängen

Frederick Birnbaum<sup>1</sup> Christian Moewes<sup>2</sup> Daniela Nicklas<sup>3</sup> Ute Schmid<sup>4</sup>

**Keywords:** Data Mining, industrielle Fertigung, visuelle Analyse

## 1 Motivation

Die industrielle Fertigung ist ein komplexer Prozess. Heutzutage sind viele Faktoren am Fertigungsprozess von Teilen beteiligt, darunter Maschinen, Menschen, Zulieferer sowie auch zahlreiche Umweltfaktoren. Der Produktionsprozess ist dabei durch Spezifikationen für jedes zu produzierende Teil vorgegeben und kann unter anderem mit Hilfe von Modellierungssprachen zur besseren Übersicht erfasst werden [QKZ11]. Diese Spezifikationen basieren jedoch auf Erfahrungen, Annahmen und Expertenwissen [H+06] was eine zentrale Anpassung der Maschinen deutlich erschwert. Ein komplettes Verständnis des Produktionsprozesses ist heutzutage ohne Hilfsmittel nicht mehr möglich [AK05]. Es ist nicht klar wie sich die Veränderung einer Maschineneinstellung auf andere Funktionen in der Produktionslinie auswirkt. Dazu kommt, dass bisher noch unbekannt Zusammenhänge, z.B. zwischen Mensch und Maschinen nicht richtig erfasst werden können. Im Zuge des Wandels zur Industrie 4.0, welcher eine völlig autonom gesteuerte Fabrik ermöglicht [Lu13], fallen riesige Datenmengen an. Diese können nun auch für Qualitätsverbesserungen im Produktionsprozess verwendet werden. Solche Qualitätsverbesserungen in Bezug auf Ausschussfrüherkennung sollen in dieser Arbeit besonders betrachtet werden.

## 2 Multidimensionale Qualitätsdaten

Für die Fertigung relevante Daten entstehen aus mehreren Informationssystemen. So erfassen an den Maschinen angebrachte Sensoren die jeweiligen Umweltfaktoren (wie Temperatur, Feuchtigkeit) in Echtzeit [PKP16]. Sogenannte Manufacturing Execution Systems (MES) bilden die digitale Fertigungsplanung auf der produktiven Ebene ab. Die dabei verwendeten Stammdaten, also nicht veränderbare Werte, sind währenddessen in Enterprise Resource Planning-Systemen (ERP) abgespeichert, um einen Bezug zum Prozess

---

<sup>1</sup> Otto-Friedrich-Universität Bamberg, 96045 Bamberg, frederick-leon.birnbaum@stud.uni-bamberg.de

<sup>2</sup> Robert Bosch GmbH, 96050 Bamberg, Christian.Moewes@de.bosch.com

<sup>3</sup> Otto-Friedrich-Universität Bamberg, 96045 Bamberg, daniela.nicklas@uni-bamberg.de

<sup>4</sup> Otto-Friedrich-Universität Bamberg, 96045 Bamberg, ute.schmid@uni-bamberg.de

herstellen zu können [WMK05]. Es besteht weiter die Möglichkeit, die Daten in einem Data Warehouse [IGG03] zu erfassen, zu integrieren und zu Analyseziwecken für das Management oder Produktionsmitarbeiter aufzubereiten.

### 3 Geplante Data Mining-Methoden

Auf ein Data Warehouse aufbauend entstehen nun neue Möglichkeiten des Data Minings. Das Sammeln und der Einsatz der Maschinendaten mit passenden Verfahren bringt Wirkungszusammenhänge der einzelnen Produktionsfaktoren hervor. Diese wirken sich entweder direkt oder indirekt auf die Qualität des Fertigungsprozesses aus. Vorgegebene Toleranzwerte aus ERP und MES-Systemen helfen dabei, passende Qualitätsregeln aufzustellen. Darunter fällt z.B. das heuristische Einstellen der Parameter einer Maschine in einer Linienfertigung. Eine lokale Veränderung der Einstellgrößen einer Maschine kann auf nachfolgenden Maschinen zum Ausschuss führen, auch wenn diese nicht angepasst werden. Dies liegt vornehmlich daran, dass der Fertigungsprozess heutzutage so komplex ist, dass linearen Abhängigkeiten zwischen Prozessparametern auf unterschiedlichen Maschinen faktisch nicht existieren. Jedoch kann der produzierte Ausschuss als Erfahrung dienen, um gerichtete Suchverfahren für bessere Einstellungen heranzuziehen.

In der Industrie hat sich hierzu der “Cross Industry Standard Process for Data Mining” (CRISP-DM) über die Jahre als Standard durchgesetzt. Dieser gliedert den Data Mining Prozess in die Phasen Business Understanding, Data Understanding, Data Preperation, Modelling, Evaluation und Deployment [Ch00].

Basierend auf dem CRISP-DM Ansatz befasst sich diese Arbeit mit der Anwendung passender Data Mining Verfahren auf diskrete und numerische Daten, um möglichst früh Ausschuss zu erkennen. Die Ergebnisse sind Wirkungszusammenhänge, die als Assoziationsregeln aufgestellt und in geeigneter Reporting Software visualisiert werden können. Da diese Regeln lediglich auf diskrete Daten ausgelegt sind, muss hierbei auch der Vergleich zu herkömmlichen Data Mining-Verfahren gezogen werden. Dies geschieht mit Hilfe der damit verbundenen Gütemaße.

### 4 Vorgehen

In einem ersten Schritt befasst sich diese Arbeit mit der Analyse von Diskretisierungsverfahren, um eine passende Data Preparation nach CRISP-DM durchzuführen. Direkt im Anschluss soll die Menge und die Art der zur Verfügung stehenden Daten analysiert werden, um eine geeignete Auswahl an Data Mining Verfahren zu ermitteln.

Im zweiten Schritt werden die ausgewählten Verfahren (Assoziationsregeln, Naive Bayes-Klassifikator, . . .) auf die Maschinendaten angewandt. Hierbei dienen die Toleranzwerte als Mittel, um den Ausschuss zu klassifizieren. Im Prozess sollen dabei mehrere Simulationen der Verfahren auf unterschiedlichen Datenmenge erfolgen.

Die Ergebnisse der Durchläufe werden im dritten Schritt mit der Reporting-Software Tableau<sup>5</sup> anschaulich visualisiert. Die Gütemaße der eingesetzten Verfahren werden ebenfalls erfasst und untereinander verglichen. Der Prozess und die Visualisierung der Assoziationsregeln erfolgt dabei basierend auf der Vorarbeit von [St13], wo bereits auf die Analyse und Visualisierung von Wirkungszusammenhängen mittels Assoziationsregeln eingegangen wird.

Im letzten Schritt soll eine Beurteilung der eingesetzten Verfahren erfolgen. Dazu spielen die bereits erwähnten Gütemaße eine Rolle. Außerdem wird auf die Performance und Analysierbarkeit der Visualisierungen in Tableau eingegangen und ein Fazit gezogen. Dabei ist zu erwarten, dass z.B. der Einsatz von Assoziationsregeln höhere Kosten verursachen wird als herkömmliche Verfahren, da der Einsatz von Diskretisierungsverfahren eine Vorbedingung ist. Auf der positiven Seite stehen dafür eine sehr deutliche Visualisierung der Wirkungszusammenhänge bei Ausschuss und ein leichtes Erkennen von Veränderungen.

## 5 Zusammenfassung und Ausblick

Insgesamt ergibt sich aus dem skizzierten Vorgehen ein hohes Potential, Fertigungsprozesse und deren Wirkungszusammenhänge durch Data Mining besser verstehen zu können.

Für einen produktiven Einsatz müssen jedoch noch einige zentrale Herausforderung gemeistert werden. Zum einen ist die korrekte Auswahl passender Simulationsdaten aus dem riesigen multivariaten Datensatz des Data Warehouses zu nennen. Zum anderen stellt sich die Frage nach einem optimalen Satz an Variablen- bzw. Features [GE03].

Bei der Übertragung auf den Echtzeit-Einsatz müssen zum einen Fragen des Concept Drift und der kontinuierlichen Anpassung der Analysemethoden an die aktuelle Produktion betrachtet werden; zum anderen müssen die gelernten Methoden geeignet umgesetzt werden, um auch mit niedriger Latenz und hohem Durchsatz in Echtzeit arbeiten zu können. Hierfür müssen inkrementelle Verfahren und der Einfluss unterschiedlicher Fenstergrößen für die Analysen erforscht werden.

Schließlich muss untersucht werden, welche Visualisierungen am besten geeignet sind, um auch Domänenexperten ohne besonderes Data Science-Vorwissen einen Zugang zu dem reichen Wissensschatz zu ermöglichen, der mit Data Mining-Methoden aus multidimensionalen Qualitätsdaten in der industriellen Fertigung gewonnen werden kann.

## Literatur

- [AK05] Agard, B.; Kusiak, A.: Data mining for selection of manufacturing processes. In: Data Mining and Knowledge Discovery Handbook. Springer, 2005, S. 1159–1166.

---

<sup>5</sup> <http://www.tableau.com>

- [Ch00] Chapman, P.; Clinton, J.; Kerber, R.; Khabaza, T.; Reinartz, T.; Shearer, C.; Wirth, R.: CRISP-DM 1.0 Step-by-step data mining guide./, 2000.
- [GE03] Guyon, I.; Elisseeff, A.: An introduction to variable and feature selection. *Journal of machine learning research* 3/Mar, S. 1157–1182, 2003.
- [H+06] Harding, J.; Shahbaz, M.; Kusiak, A. et al.: Data mining in manufacturing: a review. *Journal of Manufacturing Science and Engineering* 128/4, S. 969–976, 2006.
- [IGG03] Imhoff, C.; Gallemmo, N.; Geiger, J. G.: *Mastering data warehouse design: relational and dimensional techniques*. John Wiley & Sons, 2003.
- [Lu13] Lucke, D.: Smart Factory. In: *Digitale Produktion*. Springer, 2013, S. 251–269.
- [PKP16] Petersen, S. A.; van der Kooij, R.; Puhar, P.: Connecting Business Processes and Sensor Data in Proactive Manufacturing Enterprises. In: *Working Conference on Virtual Enterprises*. Springer, S. 101–109, 2016.
- [QKZ11] Qiao, L.; Kao, S.; Zhang, Y.: Manufacturing process modelling using process specification language. *The International Journal of Advanced Manufacturing Technology* 55/5-8, S. 549–563, 2011.
- [St13] Steinbrecher, M.: *Discovery and visualization of interesting patterns*, Diss., Magdeburg, Universität, Diss., 2013, 2013.
- [WMK05] Wiendahl, H.-H.; Mußbach-Winter, U.; Kipp, R.: *MES-Lösungen: Wann braucht man sie, wozu sind sie nützlich?* In: *automatisierungs-ATLAS 2005/06*. Technik-Dokumentations-Verlag, 2005.