

Automatisierung im Prozess der Wissensentdeckung in Simulationsdaten – Charakterisierung der Ergebnisdaten

Automation within the process of knowledge discovery in simulation data – Characterization of the result data

Jonas Genath, Sören Bergmann, Niclas Feldkamp, Steffen Straßburger,
TU Ilmenau, Ilmenau (Germany),
jonas.genath@tu-ilmenau.de, soeren.bergmann@tu-ilmenau.de,
niclas.feldkamp@tu-ilmenau.de, steffen.strassburger@tu-ilmenau.de

Abstract: The traditional application of simulation in production and logistics is usually aimed at changing certain parameters in order to answer clearly defined objectives or questions. In contrast to this approach, the method of knowledge discovery in simulation data (KDS) uses a simulation model as a data generator (data farming). Subsequently using data mining methods, hidden, previously unknown and potentially useful cause-effect relationships can be uncovered. So far, however, there is a lack of guidelines and automatization-tools for non-experts or novices in KDS, which leads to a more difficult use in industrial applications and prevents a broader utilization. This paper presents a concept for automating the first step of the KDS, which is the process of characterization of the result data, using meta learning and validates it on small case study.

1 Einleitung

Die traditionelle Anwendung der Simulation in Produktion und Logistik zielt meist auf die Veränderung bestimmter Parameter ab, um im Vorfeld klar definierte Zielstellungen bzw. Fragestellungen zu beantworten (März et al. 2011; Law 2013). Die Richtlinie 3633 des VDI nennt als typische zu bearbeitende Zielstellungen beispielsweise das Finden von Kapazitätsengpässen und die Bewertung von unterschiedlichen Maschinenbelegungen in einem Prozess. Bei diesem Vorgehen hängen die Ergebnisse stark von dem Expertenwissen der für die Durchführung der Simulationsstudie verantwortlichen Personen ab, da diese auf der Basis von eigener Erfahrung die Spanne und die zu variierenden Parameter festlegen. Die resultierenden Simulationsergebnisse werden anschließend analysiert und es können entsprechende Rückschlüsse über Maßnahmen für das reale System erarbeitet

werden. Obwohl diese Vorgehensweise zur Beantwortung von im Vorhinein festgelegten Fragestellungen als Best-Practice-Ansatz gilt, könnte, da der abgedeckte Modellraum nur sehr gering ist, über das grundsätzliche Modellverhalten nur wenig Wissen generiert werden (Feldkamp et al. 2015). Sind die Zielstellungen der Simulationsstudie weniger konkret, wie z. B. „Welche Parameter besitzen den größten Einfluss auf eine Kenngröße des Systems?“ oder besteht das Interesse an einem verbesserten Gesamtverständnis über das Systemverhalten sowie den Beziehungen zwischen Parametern, muss bereits im Experimentdesign für einen umfassenderen und besser abgedeckten Modellraum gesorgt werden (Law 2013; Kleijnen et al. 2005). Aus dieser Idee, die Simulation für mehr als nur die Beantwortung bestimmter Fragestellungen eines Systems zu nutzen, ergibt sich ein neues Anwendungsgebiet: Das von Brandstein und Horne für die Gefechtssimulation entwickelte Data Farming (Brandstein und Horne 1998).

Im Data Farming wird das Simulationsmodell als Datengenerator genutzt, indem im Experimentdesign ein umfangreicher Wertebereich für die Eingangsparameter gewählt und dadurch der möglichst vollständige Ergebnisraum des Modells abgedeckt wird (Sanchez 2014). Hiermit können zur Unterstützung der Entscheidungsfindung positive und negative Wirkzusammenhänge entdeckt werden, welche unter Betrachtung eines spärlich abgedeckten Modellraums verborgen geblieben sind (Horne und Meyer 2004). Unter anderem erkannten Feldkamp et al. (2017) die ungenutzten Analysemöglichkeiten des Data Farming für die Simulation in der Produktion und Logistik und entwickelten das Konzept der Wissensentdeckung in Simulationsdaten.

Dieser Beitrag widmet sich der Forschungsfrage, inwieweit die Charakterisierung der Ergebnisdaten in der Wissensentdeckung in Simulationsdaten unterstützt bzw. automatisiert werden kann. Der Schwerpunkt der Analyse liegt hier auf der Filterung von relevanten Ergebnisparametern, welche zu finden für den unerfahrenen Anwender entweder hohen Aufwand bedeuten oder ihm sogar möglicherweise verborgen bleiben. In diesem Kontext werden unter relevanten Ergebnisparametern, die verstanden, welche statistische Auffälligkeiten besitzen und dadurch im Verlauf anschließender Analysen zur einem Informationsgewinn über das Modell führen können.

Das Ziel ist, dass auch Anwender, welche mit dem Prozess der Wissensentdeckung nicht vertraut sind, oder erfahrene Anwender bei dem Erkennen von relevanten Ergebnisparametern einer Data-Farming-Studie unterstützt werden und diese somit für weitere Analyseschritte nutzen können.

Der Aufbau des Beitrags gliedert sich entsprechend wie folgt: Zuerst werden die theoretischen Grundlagen des Prozesses der Wissensentdeckung in Simulationsdaten und der Methode des Meta Learning vorgestellt. Im Anschluss folgt das konzeptionelle Kapitel, welches das vorgeschlagene Konzept zur Automatisierung der Charakterisierung der Ergebnisdaten vorstellt und sich mit der Entwicklung auseinandersetzt. Abschließend werden die Ergebnisse der Konzeptentwicklung anhand eines Realweltszenarios evaluiert. Der Beitrag wird mit einem Fazit und einem Ausblick abgeschlossen.

2 Grundlagen der behandelten Methodiken

In den folgenden zwei Unterkapiteln werden die Grundlagen der in diesem Beitrag verwendeten Methoden vorgestellt, um dem Leser ein tieferes Verständnis über das vorgeschlagene Konzept zu ermöglichen.

2.1 Wissensentdeckung in Simulationsdaten

Die Wissensentdeckung in Simulationsdaten hat ihren Ursprung im Prozess der Knowledge Discovery in Databases zur Auswertung von großen Datenmengen nach Fayyad et al. (1996). Feldkamp et al. (2017) adaptierten diesen Prozess an das Data Farming, da sich die generierten Daten auf Grund ihres Umfangs nicht mehr manuell auswerten lassen. Die in der Wissensentdeckung in Simulationsdaten zur Datenanalyse verwendeten Methoden stammen aus dem Spektrum des Data Mining und wurden bisher im Kontext von Simulationsdaten nur wenig genutzt. Die anschließende Ergebnis- und Datenrepräsentation erfolgt im Sinne des eng mit dem Data Mining verzahnten Konzepts der Visual Analytics. Unter Visual Analytics wird die interaktive Visualisierung von großen und komplexen Datensätzen zur Unterstützung der Mustererkennung sowie Entscheidungsfindung verstanden (Keim et al. 2008). Die Wissensentdeckung in Simulationsdaten lässt sich in drei Prozessschritte aufteilen. Bei diesen handelt es sich um die Charakterisierung der Ergebnisdaten, die Mustererkennung und Klassenbildung sowie die Untersuchung der Beziehung zwischen Eingangs- und Ergebnisdaten.

Die Hauptaufgabe in der Charakterisierung der Ergebnisdaten besteht in der Identifizierung von interessanten Ergebnisparametern. Diese bilden entweder die Basis oder die Eingangsparameter für die darauffolgenden Data-Mining-Verfahren, welche für die Suche nach Wissen über das Modellverhalten genutzt werden. Das Vorgehen erfolgt aktuell manuell und Anwender müssen die statistischen Lagemaße sowie dessen Visualisierungen individuell auswerten, um die relevanten Ergebnisparameter zu identifizieren. Bei der im Data Farming üblichen großen Menge an Ergebnisparametern kann diese Aufgabe zu einem hohen zeitlichen Aufwand führen. Dieser Beitrag konzentriert sich daher auf die Automatisierung dieses Vorgehens. Eine Betrachtung der anschließenden Prozessschritte in der Wissensentdeckung in Simulationsdaten ist aktuelles Forschungsgebiet.

2.2 Meta Learning

Damit eine automatisierte Identifikation der statistisch auffälligen Ergebnisparameter erreicht werden kann, wird die Methodik des Meta Learning vorgeschlagen. Für das Meta Learning gibt es in der Wissenschaft noch keine einheitliche Definition. Dieser Beitrag orientiert sich daher an der Interpretation nach Brazdil et al. (2009), die Meta Learning als Methode definiert, welche Domänenwissen und Metafeatures benutzt, um die Effizienz von Prozessen mittels der Adaption von Machine-Learning-Modellen oder Data-Mining-Verfahren zu steigern. Bei Metafeatures handelt es sich um Charakteristika eines Datensatzes, Algorithmus oder Machine-Learning-Modells (Brazdil et al. 2009). Das Domänenwissen wird für das Meta Learning mittels spezifischer definierter Kennzahlen repräsentiert. Beide Aspekte lassen sich unter dem Begriff Metadaten zusammenfassen. Grundsätzlich kann Meta Learning als das Anwenden von Machine Learning zur Verbesserung von Machine-Learning-Modellen beschrieben werden.

Im Bereich der prognostizierenden Modelle oder Hyperparameteroptimierung wurde das Konzept des Meta Learning bereits mit guter Erfolgsquote umgesetzt (Ren et al. 2018). Diese Anwendungen behandeln jedoch nur akademische Klassifikations- und Regressionsprobleme, bei denen es eindeutige Ergebnisse gibt. Bei dem vorgestellten Prozessschritt handelt es sich jedoch um einen komplexen Ablauf mit einem starken explorativen Charakter, welcher von dem Simulationsmodell und dem Experimentdesign abhängig ist. Daher sind die Ergebnisse nicht immer eindeutig „richtig oder „falsch“.

Das typische Vorgehen bei der Entwicklung eines Meta-Learning-Modells lässt sich in drei Abschnitte einteilen. Dabei handelt es sich als Erstes um das Extrahieren der Metadaten. Dieser Schritt besitzt eine hohe Relevanz, da nur mit einer geeigneten Repräsentation und Auswahl an Metadaten das Meta-Learning-System in der Lage ist, sinnvolle und hilfreiche Unterstützung zu liefern. Vanschoren (2011) und Brazdil et al. (2009) empfehlen für die Extraktion die Entwicklung eines Metadaten-Generators. Der nächste Schritt ist die Auswahl eines geeigneten Algorithmus für das Meta Learning. Grundsätzlich sind die meisten traditionellen Machine-Learning-Algorithmen auch für die Verwendung als Meta Learner geeignet. Für die verfügbaren Trainingsdaten des Meta-Learning-Algorithmus empfiehlt Lemke et al. (2015) eine erweiterbaren Metadatenbank. Die Erweiterbarkeit ist dahingehend wichtig, dass ein kontinuierlicher Verbesserungsprozess der Prädikationen des trainierten Modells realisiert werden kann.

3 Konzeptentwicklung

In dem folgenden Abschnitt werden das Konzept und dessen Entwicklungsschritte zur Lösung der genannten Probleme erläutert. Das Konzept wurde nach den drei von Lemke et al. (2015) empfohlenen Abschnitten erarbeitet und ist in Abbildung 1 dargestellt. Dabei wird das Konzept zusätzlich in zwei Phasen aufgeteilt. In der Offline-Phase werden aus den Ergebnissen bereits durchgeführter Data-Farming-Studien mittels Metadaten-Generators die festgelegten Metadaten extrahiert. Diese dienen als Trainingsdatensatz für das Meta-Learning-Modell, welches anschließend der Online-Phase zur Verfügung gestellt wird. In der Online-Phase werden neue Data-Farming-Studien ausgewertet. Zuerst extrahiert der Metadaten-Generator die definierten Metadaten aus den Simulationsdaten und stellt diese dem trainierten Modell zur Verfügung. Nur die simulationsspezifischen Metadaten müssen von dem Anwender mittels einer einfachen Maske abgefragt werden. Mit Hilfe der extrahierten Metadaten inferiert das trainierte Modell eine Rangliste über die prognostizierte Relevanz der Ergebnisparameter. Diese liefert dem Anwender Informationen über mögliche statistische Auffälligkeiten und eine Basis für die folgenden Data-Mining-Verfahren.

Die drei Abschnitte der Konzeptentwicklung werden in den folgenden Ausführungen detailliert erläutert.

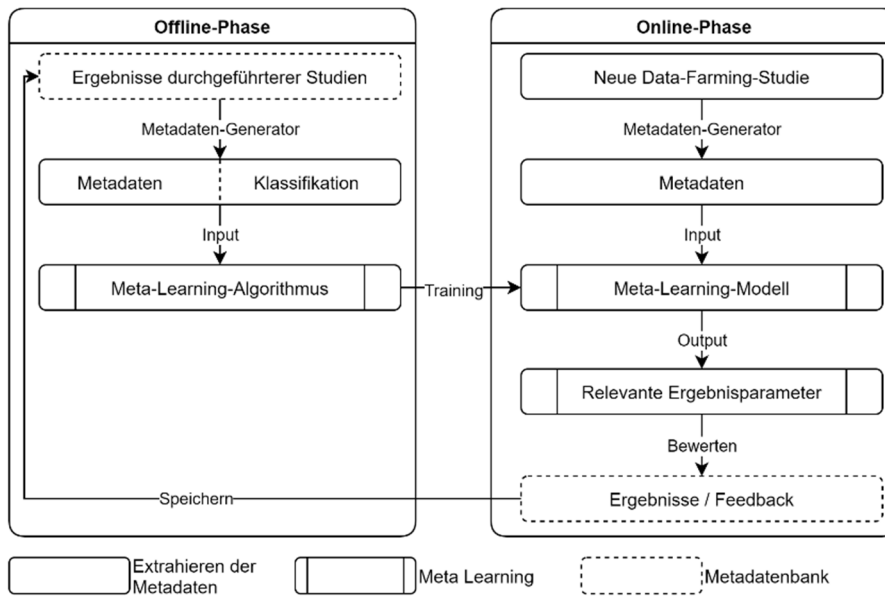


Abbildung 1: Konzeptionelles Vorgehen zur automatisierten Charakterisierung der Ergebnisdaten.

Extrahieren der Metadaten

Die Schwierigkeit beim Meta Learning besteht in der Definition der beschreibenden Domäneninformationen und wie diese dem für das Meta Learning verwendeten Machine-Learning-Modell zugänglich gemacht werden. Hierbei gilt es vor allem zu beachten, dass ein Machine-Learning-Modell keine Informationen darüber hat, welchen Kontext der Ergebnisparameter beschreibt. Beispielsweise wird eine hohe Auslastung einer Maschine meist als positiv erachtet aber die Rüstzeiten in einem System sollten eher niedrig ausfallen. Ähnliches gilt für die Unterscheidung zwischen einer zeitlichen und mengenbezogenen Kennzahl. Einzig anhand des numerischen Wertes ist ein Machine-Learning-Modell zu dieser Interpretation nicht in der Lage.

Rivoli et al. (2018) definieren über 100 standardisierte Metafeatures und teilen diese in sechs Gruppen ein. Im Kontext des Ziels und der verfügbaren Daten eignen sich hierfür nur die Metafeatures der statistischen und allgemeinen Gruppierung. Dazu zählen unter anderem verschiedene Quartile, Min- bzw. Max-Werte und die statistische Schiefe. Die domänenspezifischen Informationen müssen je nach Anwendungsfall experimentell bestimmt werden. Hierfür wurde eine Reihe von Experimenten mit simulationsspezifischen Informationen durchgeführt. Die besten Resultate wurden mit Metadaten über den oben beschriebenen Kontext der Kennzahlen erreicht. Zu diesem Zweck wurde eine Kategorisierung der Kennzahlen nach der Richtlinie 3633 des VDI vorgenommen. Die Einteilung ist über den Typ (Zeit, Menge, Quotient, Prozent) sowie die Gruppierung (Produktivität, Nutzungsgrad, Systemlastentitäten) des Parameters definiert. In den Eingangsdaten für das Modell werden diese Eigenschaften über binäre Einträge repräsentiert. Die Metadaten werden dem Meta-Learning-Algorithmus in Tabellenform übergeben. Tabelle 1 liefert einen beispielhaften Ausschnitt der bereitgestellten Daten.

Tabelle 1: Ausschnitt ausgewählter Metadaten.

ID	Ergebnisdaten	Min	Max	Mean	Menge	Zeit	Systemlastentität	Nutzungsgrad
1	Durchlaufzeit	12,3	22	17,35	0	1	1	0
2	Auslastung	0,78	0,99	0,80	0	0	0	1
3	Rüstvorgänge	5	12	6	1	0	0	1

Meta Learning

Der verwendete Meta-Learning-Algorithmus muss verschiedene Anforderungen erfüllen, um für den vorliegenden Kontext geeignet zu sein. Zu diesen zählen zum einen, dass eine Rangordnung der Ergebnisse feststellbar sein muss. Das heißt die Zuordnung muss quantifizierbar sein. Weiterhin müssen bereits wenige Trainingsdaten klassifizierbar und Ergebnisse präzifizierbar sein können.

In dem vorliegenden Beitrag wurde sich für einen *k-nearest neighbors classifier* entschieden. Dieser erfüllt die Anforderungen und hat in den durchgeführten Experimenten zu guten Ergebnissen geführt. Es sind aber auch andere Machine-Learning-Algorithmen wie z. B. Support-Vector-Maschinen geeignet. Als Zielgröße der Klassifizierung wurden qualitative Klassen („Sehr relevant“, „Nähere Betrachtung sinnvoll“, „Nicht relevant“, ...) gebildet, welche die Relevanz des Ergebnisparameters wiedergeben. Dafür wurden die verwendeten Trainingsdaten manuell ausgewertet und in die entsprechenden Klassen eingeteilt. Die verwendeten Trainingsdaten stammen aus bereits durchgeführten Data-Farming-Studien.

Metadatenbank

In der Kombination mit den extrahierten Metadaten werden die Zielklassen in der Metadatenbank gespeichert und dem Algorithmus zur Verfügung gestellt. Im Falle einer neuen Data-Farming-Studie arbeitet der Anwender mit den präzifizierten Ergebnissen, welche er auf Basis der Rangliste auswählt. Dadurch erfolgt im weiteren Verlauf der Wissensentdeckung automatisch eine Evaluation der ausgewählten Ergebnisparameter. Bei Bedarf können zum Abschluss der Studie die präzifizierten Klassen und Metadaten der evaluierten Ergebnisparameter als Erweiterung der Trainingsdaten in der Metadatenbank gespeichert werden.

4 TestszENARIO

Im Folgenden wird das trainierte Meta-Learning-Modell an einem TestszENARIO evaluiert und erste Ergebnisse der Data-Farming-Studie werden diskutiert.

4.1 Simulationsmodell

Zur Evaluation des trainierten Modells wird ein reales Szenario aus dem Logistikbereich herangezogen. Hierbei handelt es sich um eine Umladestation aus dem Automobilzuliefererbereich. Konkret werden Lastkraftwagen (LKW) an zwei Ladedocks mit jeweils unterschiedlicher Beladung (*Produkt A/Produkt B*) durch einen Gabelstapler entladen und mit dem produktneutralen Leergut wieder beladen. Für diese Aufgabe stehen für das entladene Gut sowie das Leergut mehrere Puffer zur Verfügung. Abbildung 2 zeigt einen Screenshot des Simulationsmodells aus Siemens Plant Simulation.

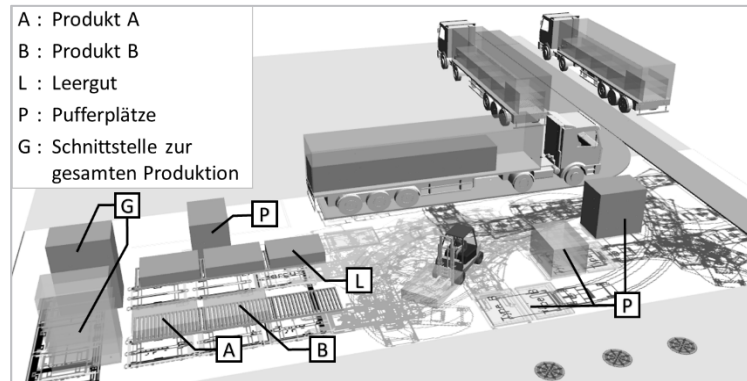


Abbildung 2: Screenshot des 3D-Simulationsmodell.

Zu den variierbaren Eingangsfaktoren des Simulationsmodells gehören die Anteile der Produkte A (*ProduktionAnteilA*) und B (*ProduktionAnteilB*) an der gesamten Produktion. Die Darstellung der gesamten Produktion ist für den Anwendungsfall nicht weiter relevant und ist daher in dem Modell nicht abgebildet. Weitere Faktoren des Simulationsmodells sind das gesamte Produktionsvolumen, die Anlieferzyklen der LKWs an jeweiligen Docks (*ZykluszeitA*, *ZykluszeitB*) und die Größe des Puffers für Vollgut (*Pufferplätze_Vollgut*). Im Gegensatz zu der geringen Anzahl an Eingangsparametern stehen 24 Ergebnisparameter zur Analyse des Szenarios zur Verfügung. Die Ergebnisparameter erfassen unter anderem Daten über den Gabelstapler (*Auslastung*, *Aufträge*, *Blockierzeit*), die LKWs (*Min-/Max-/Durchschnitts-Beladung*, *Backlog*, *Zykluszeit überschritten*), die Puffer und weitere allgemeine Eigenschaften (*Verbrauch Produktion*, *Stillstand Produktion*, *DLZ*). Die Anzahl der in dem vorliegenden Testszenario durchgeführten Experimente der Data-Farming-Studie beläuft sich auf 5000. Um eine gute Raumbdeckung durch die Experimente zu erzielen, wurde ein Latin-Hyper-Cube-Sampling durchgeführt. Die Grenzwerte der variierbaren Faktoren sind in Tabelle 2 dargestellt.

Tabelle 2: Experimentdesign.

Eingangsfaktor	Min-Wert	Max-Wert	Einheit
Produktionsvolumen	25	33	Einheiten/Stunde
ProduktionAnteilA	4,75	50	Prozent
ProduktionAnteilB	0,25	50	Prozent
ZykluszeitA	45	120	Minuten
ZykluszeitB	45	120	Minuten
PufferplätzeVollgut	3	10	Einheiten

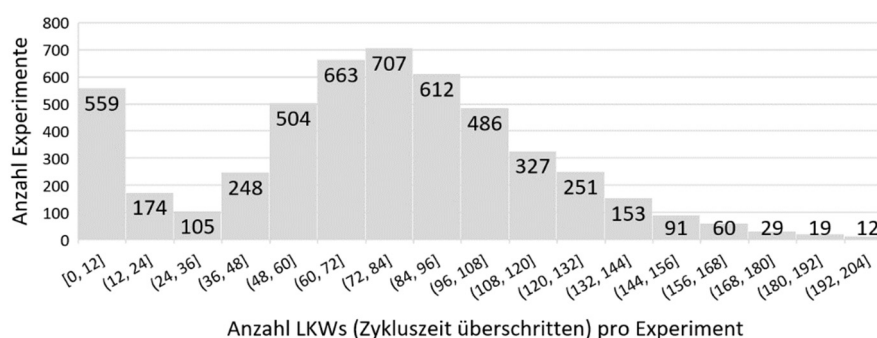
4.2 Ergebnisse

Auf die akkumulierten Simulationsdaten der Data-Farming-Studie wurde das trainierte Meta-Learning-Modell zur Filterung der Ergebnisdaten angewendet. Die Resultate der Prädikation der 24 Ergebnisparameter werden in Tabellenform ausgegeben. Einen ersten Überblick über die interessantesten drei Ergebnisparameter liefert Tabelle 3.

Tabelle 3: Ausschnitt der Prädiktionen des Meta-Learning-Modells.

ID	Parameter	Klasse	Distanzmaß	Rank
13	Zykluszeit überschritten	Sehr relevant	0,914	1
18	Verbrauch Produktion	Sehr relevant	0,815	2
3	Durchschnittliche Stapelhöhe	Sehr relevant	0,764	3

Die am höchsten bewerteten Ergebnisparameter sind *Zykluszeit überschritten*, *Verbrauch Produktion* und *Durchschnittliche Stapelhöhe*, welche als sehr relevant eingestuft werden. Das Distanzmaß stellt einen Kehrwert der auf das Intervall von Null bis Eins transformierten Entfernungen zu dem Zentrum der jeweiligen Klasse dar. D. h. je höher der Wert, desto zuverlässiger ist die Zugehörigkeit zu der eingeordneten Klasse. In Abbildung 3 ist die Verteilung des Parameters *Zykluszeit überschritten* dargestellt. Dieser Parameter beschreibt, wie viele LKWs ihre Zykluszeit in dem Experimentlauf überschritten haben. Es ist deutlich erkennbar, dass die Normalverteilung von einer Spitze im Intervall $[0, 12]$ durchbrochen wird. Das führt zu der Vermutung, dass es Wechselwirkungen zwischen den Faktoren gibt, die zu einer deutlich stabileren Einhaltung der Zykluszeit führen. Eine genauere Untersuchung im Laufe der Data-Farming-Studie in Bezug auf diesen Ergebnisparameter ist daher sinnvoll.

**Abbildung 3:** Häufigkeitsverteilung der LKWs, welche die Zykluszeit überschreiten.

Auch bei anderen als interessant eingeordneten Ergebnisparametern lassen sich Auffälligkeiten finden. Z. B. fällt bei dem Verbrauch der Produktion im Bereich des Medians ein deutlicher Einbruch in der Verteilung auf. Auch hier kann eine tiefere Analyse zu Wissensgewinn führen, da der Verbrauch der gesamten Produktion einen essenziellen Informationsgehalt über die Performanz der einzelnen Experimentläufe besitzt.

Um die prädizierten Ergebnisse zu validieren, wurde eine mehrdimensionale Klassifizierung des Modellverhaltens mittels eines Gaussian-Mixture-Modells und den genannten Ergebnisparametern durchgeführt. Bei fünf Clustern wurde auf Basis des Silhouettenkoeffizienten durch das Modell eine gute Strukturierung der Daten erreicht. Werden die klassifizierten Daten auf das Intervall $[0, 12]$ des auffälligen Ergebnisparameters *Zykluszeit überschritten* gefiltert, wird deutlich, dass in diesem Abschnitt fast sämtliche Experimente dem Cluster 5 zugeordnet wurden. Gleichzeitig führen diese Experimente auch zu den höchsten Ergebniswerten des Verbrauchs der gesamten Produktion. Daraus schlussfolgernd können die

Experimente, die diesem Cluster angehören, als sehr performant eingestuft werden. Anschließend stellt sich die Frage: Welche Faktoren führen zu diesen performanten Experimenten? Abbildung 4 zeigt die Eingangsfaktoren *Produktionsvolumen*, *ProduktionAnteilA* und *ProduktionAnteilB* nach Cluster 5 gefiltert. Es wird deutlich, dass diese Faktoren eine große Bedeutung in Bezug auf die Clusterzuordnung haben. Insbesondere das Produktionsvolumen scheint einen großen Einfluss auszuüben, da in diesem Cluster ausschließlich Ausprägungen größer oder gleich 30 vorkommen. Entgegengesetzt sind die Faktoren *ProduktionAnteilA* und *ProduktionAnteilB* im niedrigen Bereich wesentlich stärker ausgeprägt. Daraus kann folgende erste Schlussfolgerung über die Systemperformanz gezogen werden: Ein hoher Verbrauch in der gesamten Produktion und ein zuverlässiges Einhalten der Zykluszeit durch die LKWs resultiert aus einem hohen Produktionsvolumen bei relativ niedrigen Produktanteilen von A und B. Dieser Zusammenhang eröffnet die Vermutung, dass ein Engpass in dem abgebildeten System vorliegt, da invers hohe Produktanteile zu einem niedrigen Gesamtverbrauch führen könnten.

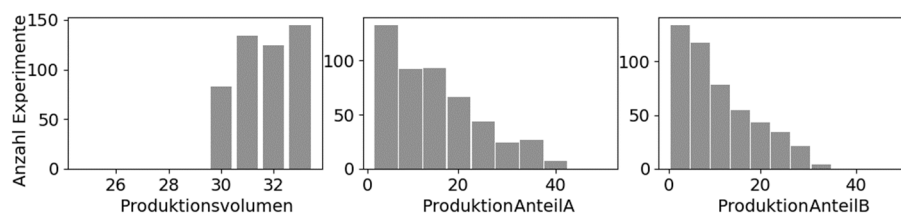


Abbildung 4: Eingangsfaktoren gefiltert nach Cluster 5.

5 Fazit und Ausblick

Dieser Beitrag hat ein Konzept zur automatisierten Identifizierung von statistisch auffälligen bzw. relevanten Ergebnisparametern im Kontext der Wissensentdeckung in Simulationsdaten vorgeschlagen und das entwickelte Modell an einem Anwendungsfall evaluiert. Dadurch konnte die prinzipielle Funktionsfähigkeit des entwickelten Konzepts sowie des trainierten Modells validiert werden. Kritisch zu hinterfragen ist der mögliche Bias, welcher durch die manuelle Klassifizierung der Trainingsdaten entstehen könnte sowie deren geringe verfügbare Menge. Durch eine breite Anwendung und einer online verfügbaren Metadatenbank, könnte dieser Kritikpunkt auf Grund einer kontinuierlichen Erweiterung der Trainingsdaten negiert werden. Ein weiterer notwendiger Schritt im Sinne der Grundidee des Beitrags ist die Unterstützung der anschließenden Prozessschritte der Wissensentdeckung in Simulationsdaten.

Literaturverzeichnis

- Brandstein, A.G.; Horne, G.E.: Data Farming: A Meta-technique for Research in the 21st Century. *Maneuver Warfare Science* (1998), S. 93–99.
- Brazdil, P.; Giraud-Carrier, C.; Soares, C.; Vilalta, R.: *Metalearning: Applications to Data Mining*. Berlin, Heidelberg: Springer 2009.

- Fayyad, U.; Piatetsky-Shapiro, G.; Smyth, P.: Knowledge Discovery and Data Mining: Towards a Unifying Framework. In: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, 1996, S. 82–88.
- Feldkamp, N.; Bergmann, S.; Schulze, T.; Straßburger, S.: Data Farming im Kontext von Produktion und Logistik. In: Wenzel, S.; Peter, T. (Hrsg.): Simulation in Produktion und Logistik. Kassel: kassel university press 2017, S. 169–178.
- Feldkamp, N.; Bergmann, S.; Strassburger, S.: Knowledge Discovery in Manufacturing Simulations. In: Taylor, S.J.E.; Mustafee, N.; Son, Y.-J. (Hrsg.): Proceedings of the ACM SIGSIM Conference on Principles of Advanced Discrete Simulation, London, United Kingdom, 10.06. - 12.06., 2015, S. 3–12.
- Horne, G.E.; Meyer, T.E.: Data Farming: Discovering Surprise. In: Ingalls, R.G. (Hrsg.): Proceedings of the Winter Simulation Conference, Washington, D.C, 05.12. - 08.12, 2004, S. 792–798.
- Keim, D.; Andrienko, G.; Fekete, J.-D.; Görg, C.; Kohlhammer, J.; Melançon, G.: Visual Analytics: Definition, Process, and Challenges. In: Hutchison, D.; Fekete, J.-D.; Kanade, T.; Kerren, A.; Kittler, J.; Kleinberg, J.M. et al. (Hrsg.): Information Visualization. Berlin, Heidelberg: Springer 2008, S. 154–175.
- Kleijnen, J.P.; Sanchez, S.M.; Lucas, T.W.; Cioppa, T.M.: State-of-the-Art Review: A User's Guide to the Brave New World of Designing Simulation Experiments. INFORMS Journal on Computing 17 (2005) 3, S. 263–289.
- Law, A.M.: Simulation modeling and analysis. Dubuque: McGraw-Hill Education 2013.
- Lemke, C.; Budka, M.; Gabrys, B.: Metalearning: a survey of trends and technologies. Artificial intelligence review 44 (2015) 1, S. 117–130.
- März, L.; Krug, W.; Rose, O.; Weigert, G.: Simulation und Optimierung in Produktion und Logistik: Praxisorientierter Leitfaden mit Fallbeispielen. Heidelberg: Springer 2011.
- Ren, M.; Triantafillou, E.; Ravi, S.; Snell, J.; Swersky, K.; Tenenbaum, J.B.; Larochelle, H.; Zemel, R.S.: Meta-Learning for Semi-Supervised Few-Shot Classification. International Conference on Learning Representations (2018).
- Rivolli, A.; Garcia, L.P.; Soares, C.; Vanschoren, J.; Carvalho, André C. P. L. F.: Characterizing classification datasets: a study of meta-features for meta-learning (2018).
- Sanchez, S.M.: Simulation experiments: Better data, not just big data. In: Tolk, A. (Hrsg.): Proceedings of the Winter Simulation Conference, Savannah, GA, USA, 7.12. - 10.12, 2014, S. 805–816.
- Vanschoren, J.: Meta-Learning Architectures: Collecting, Organizing and Exploiting Meta-Knowledge. In: Kacprzyk, J.; Jankowski, N.; Duch, W.; Grąbczewski, K. (Hrsg.): Meta-Learning in Computational Intelligence. Berlin, Heidelberg: Springer Berlin Heidelberg 2011, S. 117–155.
- Verein Deutscher Ingenieure VDI 3633: Simulation von Logistik-, Materialfluß- und Produktionssystemen. Berlin: Beuth Verlag GmbH, 2000.