

Unstrukturierte Daten aus der Serienproduktion mit Process Mining zur Fehleranalyse und Prozessoptimierung nutzen

Utilize unstructured data from serial production with process mining for error analysis and optimization

Wolfgang Kratsch, Institutsteil Wirtschaftsinformatik des Fraunhofer FIT, 95444 Bayreuth, Deutschland, wolfgang.kratsch@fit.fraunhofer.de

Gregor Stengel, RAPA Automotive GmbH & Co. KG, 95100 Selb, Deutschland, gstengel@rapa.com

Andreas Egger, Institutsteil Wirtschaftsinformatik des Fraunhofer FIT, 95444 Bayreuth, Deutschland, andreas.egger@fit.fraunhofer.de

Tobias Fehrer, Institutsteil Wirtschaftsinformatik des Fraunhofer FIT, 95444 Bayreuth, Deutschland, tobias.fehrer@fit.fraunhofer.de

Kurzfassung

Business Process Mining (PM) analysiert Unternehmensdaten zur Prozessverbesserung und wird immer bedeutender. Bisher beruht die Analyse meist auf Daten aus Kern-IT-Systemen (z.B. ERP-Systemen), weil diese leicht verfügbar sind und viele andere Datenquellen unzugänglich sind. Mit dem X2Log Framework soll eine Datengrundlage geschaffen werden, um bisher nicht nutzbare Datenpunkte bei sichergestellter Datenqualität für die Analyse von Prozessen zugänglich zu machen. Es umfasst die Auswahl der Datenquellen, die Extraktion von Rohdaten, deren Vorverarbeitung und Umwandlung in Prozessereignisse sowie das Zusammenfügen von Ereignissen zu Aktivitäten. Damit können etablierte PM-Werkzeuge genutzt werden. Der Automobilzulieferer RAPA sieht hierzu Potential in seiner End-of-Line-Prüfung. Zwar werden aktuell relevante Prozessdaten aufgezeichnet, aber herkömmliche Ansätze können Informationen ineffizient korrelieren, unter anderem aufgrund uneinheitlicher Abstraktionslevel. Eine Prozessanalyse ist nur bedarfsgetrieben, rückwärtsgewandt und mit hohem manuellem Aufwand möglich. Hier soll das X2Log Framework den Produktionsprozess transparent machen.

Abstract

Business Process Mining (PM) analyses company data for process improvement and is becoming increasingly important. So far, the analysis is mainly based on data from core IT systems (e.g., ERP systems), as these are easily available and many other data sources are inaccessible. The X2Log Framework aims to provide a data foundation to make previously unusable data points accessible for process analysis while ensuring data quality. It includes the selection of data sources, the extraction of raw data, its preprocessing and conversion into process events, and aggregating events into activities. This allows established PM tools to be utilized. The automotive supplier RAPA here sees potential in its End-of-Line check. Currently, relevant process data are recorded, but conventional approaches can correlate information inefficiently, partly due to inconsistent levels of abstraction. A process analysis is only demand-driven, retrospective, and requires high manual effort. Here, the X2Log Framework should make the production process transparent.

1 Einleitung

Business Process Mining (PM) nutzt digitale Ereignislogs, um Geschäftsprozesse zu identifizieren und analysierbar zumachen [1]. Die technischen Ansätze im PM ermöglichen es auch Unternehmen im Produktionsumfeld, ein neues Ende-zu-Ende-Verständnis ihrer Prozesse zu bilden und dienen als quantitative Grundlage für Maßnahmen zur Effizienzsteigerung und Stabilisierung [2]. Entscheidend für die Analyse ist, die zugrundeliegenden Prozessdaten verfügbar zu machen [3, 4]. Gerade kleine und mittlere Unternehmen stehen vor der Herausforderung, dass Prozessdaten zwar in Größenordnungen von Big Data anfallen, jedoch keine Integration von Prozessdaten aus unterschiedlichen und teils fragmentierten Quellen in Systeme zur späteren Auswertung vorgesehen ist.

Die vorgeschlagenen vier Elemente des Anything-to-Log (X2Log) Frameworks umfassen die Auswahl passender

Datenquellen, die Extraktion von Rohdaten, die Vorverarbeitung und Strukturierung der Informationen, die Transformation von Rohdaten in Prozessereignisse sowie die Zusammenführung von Ereignissen zu Aktivitäten. Hierdurch wird die Nutzung etablierter PM-Werkzeuge ermöglicht.

Bei RAPA, einem Zulieferer für die Automobilindustrie, ist die End-of-Line-Prüfung ein Prozessschritt mit Potential für die Nutzung von Prozessdaten im X2Log Framework. Trotz der Aufzeichnung relevanter Prozessdaten aus Wartung und Logistik, sowie hochaufgelösten Sensordaten der einzelnen Maschinen können Informationen mit herkömmlichen Ansätzen nicht effizient korreliert werden – unter anderem weil kein einheitliches Abstraktionslevel in der Datenhaltung besteht. So ist derzeit lediglich eine bedarfsgetriebene, rückwärtsgewandte und mit erheblichem manuellem Aufwand verbundene Prozessanalyse möglich. Das X2Log Framework nutzt die bisher ungenutzten, heterogenen und unstrukturierten Datenquellen, um ein umfassenderes Bild des

Ist-Fertigungsprozesses zu erstellen und Analysen zu ermöglichen. Herausforderungen sind dabei kaskadierte Fertigungsprozesse und die Verschiedenartigkeit der Daten, die einen Einfluss auf die Ergebnisse der End-of-Line-Prüfung haben oder haben könnten. Auf der einen Seite werden teilweise Messdaten für jedes Bauteil einzeln erfasst und abgespeichert und auf der anderen Seite handelt es sich teilweise um Chargeninformationen, die unmittelbar zugeordnet werden können. Zudem sind teils Unschärfen in den Daten zu berücksichtigen, wenn z.B. Reparaturen oder Umstellungen nicht mit der gleichen Präzision zugeordnet werden können. Ähnliches gilt für Umgebungsbedingungen. Großes Potential ergibt sich durch eine – zumindest teilweise – automatisierte Analyse von Hypothesen. Im ersten Schritt hilft dies bei der Fehlersuche und in weiteren Schritten für prädiaktive Eingriffe.

Das X2Log Framework wird derzeit bei RAPA prototypisch angewandt und evaluiert. Durch die modulare Architektur soll das Framework aber auf vielfältige Domänen anwendbar sein, etwa auch in einer Softwarelösung für den Kundenservice eines weiteren Verbundpartners.

2 Process Mining & Daten

2.1 Process Mining und Datenqualität

PM zielt darauf ab, Geschäftsprozesse zu identifizieren, zu analysieren, auf Konformität zu überprüfen und zu verbessern, indem Wissen aus digitalen Ereignislogs gewonnen wird [1]. Ein Ereignislog ist eine fortlaufende Liste von Geschäftsereignissen, die die folgenden Mindestattribute enthalten:

- *Activity Name*: Bezeichnung der im Event aufgetragenen Aktivität. Z.B., „Auftrag anlegen“, „Bauteil wiegen“.
- *Case ID*: Zugehörigkeit einer Aktivität zu einer Prozessinstanz. Z.B. Verkaufsauftrag, Seriennummer eines Bauteils.
- *Timestamp*: Zeitstempel, der in der Regel den Abschluss einer Aktivität angibt.

Bereits anhand dieser Standardattribute können Geschäftsprozesse auf ihr Verhalten untersucht werden und etwa häufig vorkommende Ablaufvarianten, Engpässe (sog. „bottlenecks“) oder unerwünschte Aktivitäten identifiziert werden. Zusätzliche Attribute auf der Ebene von Events oder Cases liefern weitere Möglichkeiten bei der Datenauswertung. **Tabelle 1** visualisiert ein beispielhaftes Ereignislog, wie es aus der Produktion bei RAPA generiert werden könnte.

Seit der wissenschaftlichen Begründung von PM um die Jahrtausendwende wurden zahlreiche Ansätze entwickelt und publiziert und kommerzielle Anbieter wie z.B. Celonis haben diese mit großem Erfolg in die Praxis übertragen und weiterentwickelt [5]. Dadurch wurde die Analyse von Prozessdaten für fachliche Anwender wesentlich erleichtert. Heute profitieren Unternehmen durch den Einsatz von PM in ihren Geschäftsprozessen von Kosteneinsparungen, resistenteren Lieferketten, zufriedeneren Kunden und höherer Prozesskonformität [6].

Ein zentrales Erfolgskriterium für PM ist die Verfügbarkeit und Qualität von Prozessdaten. Im Fokus von PM stehen bisher häufig Prozesse, die in Workflow Management Systeme, prozessgestützte Informationssysteme (engl. Process-Aware Information Systems, PAISs) und prozessorientierte Informationssysteme wie etwa Enterprise Resource Planning (ERP)- oder Customer Relationship Management (CRM)-Anwendungen eingebettet sind. In den wohlstrukturierten, relationalen Datenbanken dieser Systeme hinterlassen Prozessaktivitäten Datenspuren, die über vorgefertigte Extraktions-, Transformations- und Lademethoden (ETL) für die PM Analyse als Ereignislogs aufbereitet werden können.

Trotz der zunehmenden Verfügbarkeit von ETL-Hilfen für Standardprozesse in kommerziellen PM-Lösungen nimmt die Vorbereitung der Daten (engl. Preprocessing) nicht selten 80% des Projektaufwands in Anspruch. Dies liegt insbesondere darin begründet, dass die Qualität der erzeugten Ereignislogs eine entscheidende Rolle für die PM-Analyse spielt.

Um die Nutzbarkeit vorliegender Ereignislogs für das PM zu beurteilen, haben van der Aalst et al. [1] fünf Reifegrade anhand der Qualitätskriterien Belastbarkeit, Vollständigkeit, Semantik und Sicherheit definiert. Je höher der Reifegrad und somit die Qualität der Ereignislogs ist, desto höher ist auch das Potential sinnvoller Analyseergebnisse mittels PM zu erzielen. So können Ereignisse in Ereignislogs mit dem niedrigsten Reifegrad, dem Reifegrad I, von der Realität abweichen oder fehlen. Da die zugrundeliegenden Daten oftmals händisch zusammengetragen werden, sind Analysen wenig belastbar und somit ungeeignet für PM-Anwendungen. Beim Reifegrad II werden Ereignislogs ohne systematischen Ansatz als Nebenprodukt eines Anwendungssystems automatisch generiert. Deshalb besteht auch hier die Möglichkeit, dass Ereignisse fehlerhaft sind oder fehlen. Auch wenn die protokollierten Ereignisse in Reifegrad III nicht vollständig sind, kann die Belastbarkeit im Gegensatz zum zweiten Reifegrad als hoch eingeschätzt werden. Für Reifegrad IV gilt, dass Ereignisse verlässlich, automatisch und systematisch mit klarem Prozessbezug aufgezeichnet werden. Dem Reifegrad V entsprechen Ereignislogs aus prozessnahen Informationssystemen von höchster Qualität mit klar definierter Semantik und interpretierbaren Ereignissen. Hier werden neben den für den vierten Reifegrad definierten Aspekten auch Datenschutz und -sicherheit berücksichtigt. Aufgrund der gesteigerten Belastbarkeit können Ereignislogs ab dem dritten Reifegrad sinnvoll für PM-Anwendungen genutzt werden.

2.2 Datenquellen im Process Mining

Bisher werden überwiegend Ereignislogs aus PAISs oder prozessorientierten Informationssystemen ab dem dritten Reifegrad für die PM-Analyse verwendet. Viele wichtige Prozessaktivitäten in Unternehmen spielen sich allerdings außerhalb von prozessgestützten Informationssystemen ab und entziehen sich damit systematischer Protokollierung. Beispiele sind die Kommunikation mit Lieferanten (via Telefon oder Text), die Beratung durch den Kundenservice

(via Service-Hotlines, Techniker und Chatbots) oder die Ausführung von (manuellen) Arbeitsschritten in der Fertigung. Diese Prozessaktivitäten hinterlassen dennoch Spuren in Randsystemen wie z.B. E-Mail-Archiven, PDF-Dokumenten, Messreihen (z.B. Sensordaten) oder Mediendateien, die beispielsweise durch Videokameras oder Mikrofone erzeugt werden. Die in den Spuren enthaltenen Datenpunkte sind meist wenig strukturiert oder unstrukturiert (d.h. Daten ohne formalisierte Struktur) und nicht mit Kernsystemen korreliert. Da bisher keine systematischen Ansätze existieren, um derartige digitale Spuren in Ereignislogs mit Reifegrad III oder höher zu überführen [7], ist die integrierte Analyse von Prozessen abseits prozessgestützter Informationssysteme nicht möglich [8, 9].

Vergleichbar mit toten Winkeln beim Autofahren (engl. blind spots) navigieren Analysten also regelmäßig mit eingeschränkter Sicht durch den Prozess und sind folglich nicht in der Lage, handlungsrelevante Daten zu berücksichtigen. Mit bestehenden PM-Werkzeugen können diese sogenannten Blind Spots nur in vereinzelten Anwendungsfällen und mit immensum Aufwand bei der Datenerhebung und im Preprocessing berücksichtigt werden. Da das Volumen unstrukturierter Daten sehr stark steigt, ist für die Zukunft zu erwarten, dass sich der Anteil direkt nutzbarer Daten noch weiter verringern wird [9].

Wie bei allen datengetriebenen Analyseverfahren ist der Erfolg von PM nur so gut, wie es die Datenlage zulässt. Dies betrifft sowohl die Qualität der Daten als auch die möglichst umfassende Abdeckung der Realwelt. Um zukünftig Prozesse in voller Breite und Tiefe auf Basis einer integrierten Datengrundlage zu analysieren, müssen daher auch bisher nicht genutzte sowie unstrukturierte Datenquellen für PM erschlossen werden [7]. In einer durchgeführten Delphi-Studie zu relevanten Prozessmanagement-Fähigkeiten für das Digitalzeitalter wurde die Relevanz der Erschließung schwach strukturierter Daten im Prozessmanagement von Teilnehmenden aus der Wissenschaft und Praxis bestätigt [10, 11].

In der bestehenden PM-Literatur gibt es bereits vereinzelte Ansätze, die sich mit unstrukturierten oder nicht-klassischen Datenquellen beschäftigen. Beispielsweise stehen in einigen Publikationen unstrukturierte Textdaten im Fokus. Holz et al. [12] nutzen textuelle Sachbearbeiter-Berichte, um mit deren Hilfe Prozessabläufe zu entdecken und Kecht et al. [13] entwickelten einen Ansatz, der aus Konversationen in Foren Prozesslogs generiert. Andere Ansätze legen den Fokus auf Videodaten, wie beispielsweise [4], die aus Videos Aktivitäten extrahieren und zu Prozesslogs weiterverarbeiten. Auch Verfahren zur Nutzung von Sensordaten für PM lassen sich in der Literatur finden, wie beispielsweise in Hemmer et al. [14].

3 Das X2Log Framework

Das X2Log Framework bildet realweltliche Prozesse in einer für den Anwendungsfall spezifischen Abstraktion ab. Die Ausgangsbasis für das X2Log Framework bilden Geschäftsprozesse jeglicher Art (beispielsweise Vertriebs- oder Logistikprozesse). Um eine möglichst vollständige

und belastbare Datengrundlage für Prozessanalysen mittels PM zu schaffen, schließt das X2Log Framework direkt an bestehenden Datenquellen an. Das Framework besteht aus vier Komponenten, welche sukzessiv durchlaufen werden.

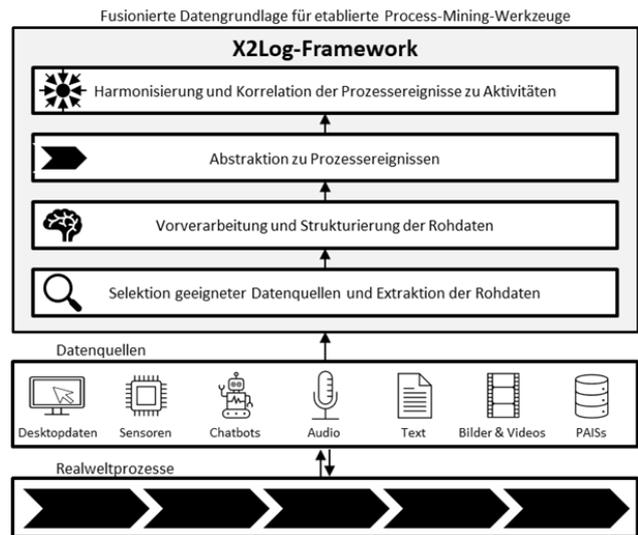


Bild 1 Schematische Darstellung des X2Log Frameworks

3.1 Selektion geeigneter Datenquellen und Extraktion der Rohdaten

Innerhalb der ersten Komponente wird eine Vorauswahl der relevanten Datenquellen getroffen, welche im Rahmen des X2Log Frameworks weiterverarbeitet werden sollen. Diese Selektion ist notwendig, da aufgrund der Vielzahl und Heterogenität der im Organisationskontext verfügbaren Datenquellen eine vollständige Erschließung nicht sinnvoll ist. Während der Integrationsaufwand mit jeder weiteren Datenquelle steigt, nimmt der Beitrag zur Vollständigkeit und Belastbarkeit für die gesamte Datengrundlage mit jeder zusätzlich berücksichtigten Datenquelle ab. Im nächsten Schritt werden aus den ausgewählten Datenquellen Rohdaten unter Berücksichtigung von Datenschutzaspekten extrahiert und für die weitere Verarbeitung gespeichert. Abhängig vom Datenformat, dem Datenvolumen und der Geschwindigkeit, mit der neue Daten generiert werden, müssen passende Ansätze für die Speicherung bestehender Rohdaten und die kontinuierliche Einspeisung neuer Rohdaten gewählt werden. Für große Datenmengen können Systeme wie Data-Lakes, welche sich insbesondere für unstrukturierte Daten eignen, in Betracht gezogen werden [15].

3.2 Vorverarbeitung und Strukturierung der Rohdaten

In der zweiten Komponente werden die ausgewählten Rohdaten in ein strukturiertes Format überführt, welches prozessbezogene Inhalte wie beispielsweise Zeitstempel oder den Namen des ausgeführten Prozessschrittes umfasst.

Tabelle 1 Schematisiertes Ereignislog aus der Produktion bei RAPA

Case ID	Activity Name	Timestamp	Resource	Measurement Result	Rel. Pressure	Product Type
1503	Pick part	2023-11-29@23:04:54	M1			A
			...			
1504	Pick part	2023-11-30@12:28:31	M1	PASS		B
1502	Test Magkraft	2023-11-30@12:28:32	M3	FAIL	-0.9	A
			...			
1503	EOL Test	2023-12-08@09:10:20	EOL Nest 1	FAIL		A

Vergleichbar mit den Funktionalitäten etablierter PM-Software müssen deshalb Konnektoren für die unterschiedlichen Datenquellen und -typen bereitgestellt werden. Die notwendigen Vorverarbeitungsschritte innerhalb der Konnektoren sind vom Datentyp der Ursprungsdatenquelle abhängig. Prozessdaten, die beispielsweise innerhalb von PAISs in relationalen Datenbanken gespeichert werden, liegen bereits in einem strukturierten Format vor und können mithilfe etablierter Algorithmen in das erforderliche Format überführt werden. Insbesondere zur Überführung unstrukturierter Datentypen (beispielsweise Sensordaten, Audio oder Video) in das für das PM erforderliche strukturierte Format gibt es nur vereinzelte allgemeingültige Ansätze. So zeigen [4] wie prozessrelevante Inhalte aus Videodaten gewonnen werden können und setzen dabei fortgeschrittene KI-Technologien aus dem Bereich des Deep Learning ein. Obwohl auch für andere Datentypen wie Sensordaten gängige Vorverarbeitungsschritte für die Datenanalyse existieren [16], fehlen allgemeingültige Konzepte für die PM-Domäne. Deshalb müssen auf Basis wissenschaftlicher Recherche geeignete Verfahren zur Datenextraktion identifiziert und für das PM erweitert werden. Hier bietet sich der Einsatz von Deep Learning oder Vision Transformer Architekturen an, welche oftmals auf Mustererkennung und Generalisierung beruhen. Als Ergebnis des Forschungsprojekts soll zudem ein aussagekräftiger Vergleich der Eignung verschiedener Deep Learning und Transformer Architekturen für verschiedene Datentypen unter Berücksichtigung von Skalierbarkeit und Ressourcenverbrauch erstellt werden.

Um eine modulare softwaretechnische Realisierung zu ermöglichen, findet eine konzeptionelle Unterscheidung zwischen dem Basis-Framework, das PM-Funktionalitäten unabhängig vom Datentyp abbildet, und datentypspezifischen PM-Modulen, die jeweils sämtliche ETL-Funktionalitäten für einen spezifischen Datentyp kapseln, statt (siehe Bild 1).

4 Anwendungsfall: RAPA

Im Projekt X2Log soll parallel zur Erarbeitung des Frameworks dieses auch in der praktischen Anwendung getestet werden. Hierzu wird in drei Stufen vorgegangen. Zunächst muss die Datenqualität der Rohdaten überprüft und ggfs. verbessert werden. Daraufhin müssen die Daten sinnvoll korreliert und zugreifbar gemacht werden. Erst dann erfolgt eine automatisierte Bewertung.

Die Taktzeiten in der Produktion liegen im Bereich weniger Sekunden bis hin zu wenigen Minuten. Im Fall von Grenzwertverletzungen mussten früher manuell Messdaten einzelner Maschinen, wie z.B. Pressendaten, ausgewertet und zu einem Gesamtbild zusammengesetzt werden. Je nachdem, welche Einflussfaktoren berücksichtigt werden sollen, wird dies sehr zeitintensiv. An einem Beispiel soll dies verdeutlicht werden:

Produkt A wird auf den Maschinen M1 bis M5 hergestellt und am End of Line Prüfstand (EOL) an 10 Nestern geprüft. Hier kommt es nun zu Verletzungen der Toleranzgrenzen. Zur Fehleranalyse werden im einfachsten Fall nur die Messwerte des EOLs betrachtet und von dort auf mögliche Ursachen geschlossen. Basierend auf einem Simulationsmodell von Produkt A wird sich der weitere Analysefokus auf einige Messwerte der Maschinen richten. Im Beispiel sei dies eine Kraft F an Maschine M4, die an zwei Nestern gemessen wird. Damit ergeben sich bereits 20 mögliche Kombinationen, die zu unterschiedlichen Ergebnissen führen können.

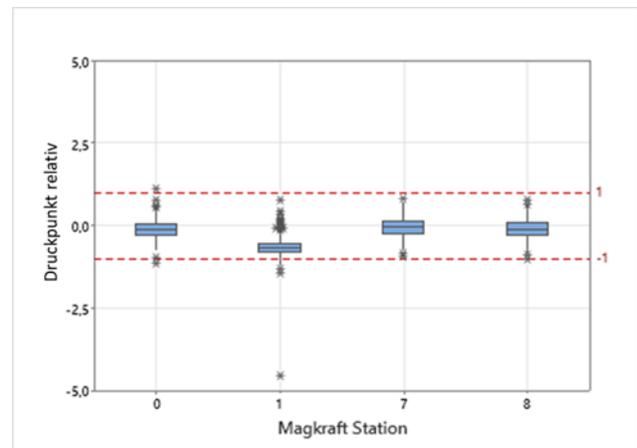


Bild 2 Bsp.: Einfluss der Nester zweier Anlagen auf den EOL

In der Praxis wird man versuchen, jeweils die Nester zueinander abzugleichen. Ziel im Projekt ist jedoch, durch eine permanente Bewertung schon frühzeitig auf Basis statistischer Auswertungen kleinste Änderungen zu detektieren. Umgesetzt ist dies inzwischen an allen neuen Anlagen für alle Messwerte, die an den eigenen Anlagen aufgezeichnet werden. Damit ist es bereits möglich, alle Produkte anhand der Seriennummern, die sich wiederum anhand der Produktionszeit auf einer Maschine bestimmen lassen, bis auf alle Unterbaugruppen hin zu verfolgen. Kommt es

beispielsweise an M1 zu einem Messwertsprung, kann dessen Auswirkung auf die nachgelagerten Maschinen verfolgt werden.

In **Bild 2** ist beispielhaft der Einfluss einer Messung in zwei Anlagen in je zwei Nestern auf die Ergebnisse im EOL dargestellt. Durch reine Auswertung, ohne zusätzliche Messung wird deutlich, dass Nest „1“ betrachtet werden muss. Ziel der Analyse im Projekt ist, dass dieses Phänomen automatisch überwacht werden kann und dann zu einem Alarm führt. Sukzessive sollen weitere Datenquellen mit in die Analyse einbezogen werden:

- Chargendaten von Einzelteilen, ggfs. mit statistischen Verteilungen der Lieferanten und Daten aus der Wareneingangskontrolle
- Fehlerdatenbanken der Maschinen
- Datenbanken zu Wartungs- und Instandhaltungseinsätzen
- Schichtprotokolle
- Äußere Einflüsse, wie Temperatur oder Luftfeuchte
- usw.

Insgesamt ist die Datenanalyse kein Selbstzweck, sondern muss dabei helfen, die Prozesse zu optimieren und letztlich die Produktionskosten senken. Hierzu hilft die Kenngröße Gesamtanlageneffektivität (engl. Overall Equipment Effectiveness, OEE) als Produkt aus Verfügbarkeit, Leistung und Qualität. Eine Korrelationsmatrix basierend auf Messwerten der verschiedensten Anlagen kann bei der Analyse von Qualitätsproblemen helfen (siehe **Bild 3**). Die Analyse der Zeitstempel lässt eine Bewertung der kritischen Stationen in Bezug auf die Taktzeiten und damit auf die Leistung der Anlage zu. Durch Auswertung von Störungen und Fehlerdatenbanken lassen sich Rückschlüsse auf die kritischen Elemente für die Verfügbarkeit ziehen.

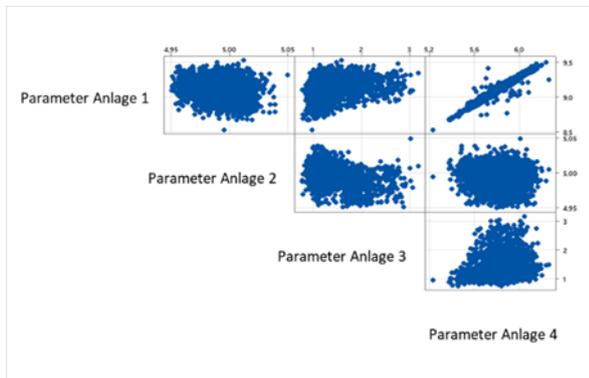


Bild 3 Korrelation von Messwerten unterschiedlicher Anlagen

5 Danksagung

Dieser Artikel wurde durch die BFS (Bayerische Forschungsförderung) im Rahmen der Sachbeihilfe „Anything-to-Log: Systematische und Software-gestützte Vervollständigung von Ereignislogs (AZ-1550-22)“ gefördert. Wir danken an dieser Stelle für die Unterstützung.

6 Zusammenfassung

Das X2Log Framework abstrahiert realweltliche Prozesse für die Analyse in verschiedenen Anwendungsfällen. Bei RAPA wird X2Log für die Analyse eines Produktionsprozesses instanziiert. Durch die Implementierung eines Forschungsdemonstrators möchte RAPA zukünftig Fehler in der Produktion erkennen und eine teilautomatisierte Ursachenanalyse durchführen. Die Implementierung des Demonstrators sowie seine Evaluation sind derzeit laufend. Eine wissenschaftliche Begleitung des Projektes soll die Wiederverwertbarkeit der Erkenntnisse sicherstellen.

7 Literatur

- [1] W. M. P. van der Aalst *et al.*, "Process Mining Manifesto," in *Business process management workshops*, Clermont-Ferrand, 2012, pp. 169–194.
- [2] A. Corallo, M. Lazoi, and F. Striani, "Process mining and industrial applications: A systematic literature review," *Knowl Process Manag*, vol. 27, no. 3, pp. 225–233, 2020, doi: 10.1002/kpm.1630.
- [3] T. Grisold, J. Mendling, M. Otto, and J. vom Brocke, "Adoption, use and management of process mining in practice," *BPMJ*, vol. 27, no. 2, pp. 369–387, 2021, doi: 10.1108/BPMJ-03-2020-0112.
- [4] W. Kratsch, F. König, and M. Röglinger, "Shedding light on blind spots – Developing a reference architecture to leverage video data for process mining," *Decision Support Systems*, vol. 158, p. 113794, 2022, doi: 10.1016/j.dss.2022.113794.
- [5] W. van der Aalst, "Academic View: Development of the Process Mining Discipline," in *Process Mining in Action*, L. Reinkemeyer, Ed., Cham: Springer International Publishing, 2020, pp. 181–196.
- [6] L. Reinkemeyer, Ed., *Process Mining in Action*. Cham: Springer International Publishing, 2020.
- [7] D. Beverungen *et al.*, "Seven Paradoxes of Business Process Management in a Hyper-Connected World," *Bus Inf Syst Eng*, vol. 63, no. 2, pp. 145–156, 2021, doi: 10.1007/s12599-020-00646-z.
- [8] D. Davis, *AI Unleashes the Power of Unstructured Data*. [Online]. Available: <https://www.cio.com/article/3406806/ai-unleashes-the-power-of-unstructured-data.html> (accessed: Oct. 12 2021).
- [9] B. Marr, "What Is Unstructured Data And Why Is It So Important To Businesses? An Easy Explanation For Anyone," *Forbes*, 16 Oct., 2019. <https://www.forbes.com/sites/bernardmarr/2019/10/16/what-is-unstructured-data-and-why-is-it-so-important-to-businesses-an-easy-explanation-for-anyone/> (accessed: Jan. 24 2024).
- [10] G. D. Kerpedzhiev, U. M. König, M. Röglinger, and M. Rosemann, "An Exploration into Future Business Process Management Capabilities in View of Digitalization," *Bus Inf Syst Eng*, vol. 63, no. 2, pp. 83–96, 2021, doi: 10.1007/s12599-020-00637-0.
- [11] N. Martin *et al.*, "Opportunities and Challenges for Process Mining in Organizations: Results of a Delphi

Study," *Bus Inf Syst Eng*, 2021, doi: 10.1007/s12599-021-00720-0.

- [12] F. Holz, B. Lantow, and M. Fellmann, "Towards a Content-Based Process Mining Approach in Personal Services," in *Lecture Notes in Business Information Processing*, vol. 421, *Enterprise, Business-Process and Information Systems Modeling*, A. Augusto, A. Gill, S. Nurcan, I. Reinhartz-Berger, R. Schmidt, and J. Zdravkovic, Eds., Cham: Springer International Publishing, 2021, pp. 62–77.
- [13] C. Kecht, A. Egger, W. Kratsch, and M. Röglinger, "Event Log Construction from Customer Service Conversations Using Natural Language Inference," in *3d International Conference on Process Mining*, 2021.
- [14] A. Hemmer, M. Abderrahim, R. Badonnel, J. Francois, and I. Chrisment, "Comparative Assessment of Process Mining for Supporting IoT Predictive Security," *IEEE Trans. Netw. Serv. Manage.*, vol. 18, no. 1, pp. 1092–1103, 2021, doi: 10.1109/TNSM.2020.3038172.
- [15] N. Miloslavskaya and A. Tolstoy, "Big Data, Fast Data and Data Lake Concepts," *Procedia Computer Science*, vol. 88, pp. 300–305, 2016, doi: 10.1016/j.procs.2016.07.439.
- [16] H. Zhu, S. Samtani, H. Chen, and J. F. Nunamaker, "Human Identification for Activities of Daily Living: A Deep Transfer Learning Approach," *Journal of Management Information Systems*, vol. 37, no. 2, pp. 457–483, 2020, doi: 10.1080/07421222.2020.1759961.