

Machine Learning und Complex Event Processing: Effiziente Echtzeitauswertung am Beispiel Smart Factory

Jonas Wanner¹, Christopher Wissuchek¹, Christian Janiesch¹

¹ Julius-Maximilians-Universität Würzburg, Würzburg, Germany
{jonas.wanner, christian.janiesch}@uni-wuerzburg.de,
christopher.wissuchek@gmail.com

Abstract. Durch die Verbindung zwischen physischen Maschinenteilen und digitalen Services werden mit Cyber-physischen Systemen in Smart Factories viele datenbasierte Optimierungen möglich. Ein wichtiger Bestandteil dieser sogenannten Smart Factories kann die Technologie Complex Event Processing (CEP) sein. CEP erlaubt Echtzeitauswertungen komplexer Events, i. S. v. kombinierten Datenwerten aus unterschiedlichen Quellen. Damit können u. a. anomale Prozessabläufe identifiziert und lokalisiert werden. Eine aktuelle Beschränkung der Wirkungsfähigkeit ist die hauptsächlich deklarative und reaktive Implementierung von CEP. Eine Erweiterung um Ansätze aus dem Machine Learning (ML) ist daher vielversprechend. Es fehlt jedoch an einer aktuellen Übersicht zu Verbindungen von CEP und ML innerhalb der Forschung sowie deren Transferfähigkeit auf Smart Factories. Unser Beitrag liefert (1) eine Synthese der bislang erforschten CEP-ML-Kombinationen, wobei sich Supervised Learning als überwiegender Kombinationsansatz zeigt, und (2) eine Übertragung der Potenziale für die Verwendung in Smart Factories. Hier zeigten sich reaktive Maßnahmen als bisheriger Forschungsschwerpunkt.

Keywords: Machine Learning; Complex Event Processing; Real-time Event Processing; Smart Factory; Literature Review

1 Einleitung

Die nutzbare Datenmenge in der modernen Industrie steigt stetig an [1]. Inzwischen beschränkt sich die Nutzung der Analysepotenziale nicht nur auf historische Datenbestände, sondern verstärkt auf die Auswertung von Echtzeitdaten. Dies erlaubt eine situationsabhängige und dynamische Entscheidungsfindung [2]. Ein Realisierungsansatz dafür kann Complex Event Processing (CEP) sein, ein Technologie-Paradigma zur Verarbeitung kontinuierlicher Datenströme auf Basis von Ereignissen [3].

Gerade in Smart Factories lassen sich mit dieser Technologie Einsparungen durch die Echtzeitauswertung von Sensordaten erzielen [4]. Trotz der großen Potenziale von CEP existiert jedoch eine fundamentale Einschränkung: CEP verarbeitet und analysiert Ereignisse auf Basis manuell vorsezifizierter, deklarativer Regeln [5, 6]. Durch die regelbasierte Umsetzung [6] ist eine Proaktivität schwierig umsetzbar. Ebenso erfordert

der reaktive Charakter teure manuelle Identifikations- und Implementierungsprozesse [7].

Forscher aus der Domäne des CEP greifen infolgedessen zuletzt verstärkt auf Machine Learning (ML) zurück, um die (teure) deklarative Natur von CEP zu verändern und zu verbessern [8-12]. ML ist ein Ansatz, in dem Computersysteme automatisiert auf Basis von Algorithmen aus Erfahrung lernen [13].

Zusammenfassend sollen die aktuellen CEP-Restriktionen durch geeignete ML-Möglichkeiten aufgeweicht und wenn möglich ganz gegenstandslos werden. Der Fokus dieser Arbeit liegt dabei auf der Anwendbarkeit in Cyber-physischen Systemen moderner Smart Factories. Es lassen sich damit zwei Forschungsfragen ableiten:

- (1) *Welche ML-Verfahren werden in Kombination mit CEP eingesetzt und wie sind diese zu systematisieren?*
- (2) *Welche Potenziale leiten sich aus dem Einsatz von ML im CEP von Industriefertigungsanlagen in Smart Factories ab?*

Zur Beantwortung der Forschungsfragen ist dieser Beitrag wie folgt aufgebaut: In Kapitel 2 wird ein Überblick zu den theoretischen Grundlagen von Smart Factories, CEP und ML gegeben. Kapitel 3 erläutert die angewendete Forschungsmethodik. Kapitel 4 beantwortet Forschungsfrage (1) auf Basis einer literaturgestützten Taxonomie zur Lokalisierung potenzieller Anwendungsmöglichkeiten. Die erlangten Erkenntnisse werden anschließend innerhalb von Kapitel 5 zur Beantwortung von Forschungsfrage (2) auf Herausforderungen und Verbesserungspotenziale für Anlagen in Smart Factories übertragen. In Kapitel 6 erfolgt eine kritische Stellungnahme zu den erarbeiteten Erkenntnissen. Ebenso werden erkannte Forschungslücken aufgezeigt. Die Arbeit schließt mit einem Fazit.

2 Smart Factories, CEP und ML

Cyber-physische Systeme und Smart Factories. Die zentralen Produktionsfaktoren der Industrie 4.0 sind Informations- und Kommunikationstechnologien [14]. Sie werden dazu genutzt Cyber-physische Systeme (CPS) zu entwickeln. Dabei handelt es sich u. a. um Objekte, Gebäude oder Produktionsanlagen, die durch eingebettete Systeme kommunikationsfähig werden und ihre Umgebung mittels verbundener Sensorik erfassen können. Sie sind vernetzbar und offenbaren über echtzeitfähige Messungen von Ereignissen (sog. Events) neue Optimierungs- und Automatisierungspotenziale [14].

Complex Event Processing. Eine Möglichkeit der Nutzenmachung dieser Events ist CEP, ein Technologie-Paradigma zur Echtzeitverarbeitung kontinuierlicher Ereignis-Datenströme [3]. Die CEP-Referenz-Architektur besteht aus drei Komponenten. Der (1) *Event Producer* überträgt die gesammelten Events an den (2) *Event Processor*, die zentrale Verarbeitungseinheit. Dieser verarbeitet Events regelbasiert auf Basis manuell und deklarativ zu definierender Agenten zur Erkennung anomalen Verhaltens [7]. Die Verbesserung der Fähigkeiten dieser sogenannten *Event Processing Agents* durch ML ist der Fokus des vorliegenden Beitrags. Nach der Verarbeitung

werden die Ergebnisse an festgelegte (3) *Event Consumer* übergeben, welche die Ereignisbehandlung übernehmen [2, 15].

ML-Verfahren. Die bisherigen CEP-Einschränkungen sollen durch Möglichkeiten des ML verbessert werden. Dies erlaubt die automatisierte CEP-Regelbildung für bekannte und ggf. unbekannte Anomalien. Zum Beispiel könnte eine zu definierende, Schwellenwert-basierte Filterung durch eine ML-trainierte Klassifikation ersetzt werden. ML bietet dazu unterschiedliche Verfahren zur Schulung von Algorithmen. *Supervised Learning* umschreibt einen Lernansatz, bei welchem Inputs bewusst vordefinierten Outputs zugeordnet werden. Durch die Übergabe der Input-/ Output-Paare zu Trainingszwecken kann die Maschine künftig neue Fälle automatisiert zuordnen [16-18]. Diese können weiter in Klassifizierungs- und Regressionsansätze unterkategorisiert werden. Beim *Unsupervised Learning* wird auf die Übergabe der Input-/ Output-Paare verzichtet. Der verwendete Algorithmus erhält ausschließlich Inputs. Als Ziel gilt das Erkennen bisher unentdeckter Muster [17, 19, 20]. Hier wird i. d. R. mit Clustering und Dimensionsreduktion gearbeitet. Das *Reinforcement Learning* ist ein weiterer Lernansatz zum Verstehen und zur Automatisierung von zielgerichtetem Lernen und Entscheidungsfindungsprozessen [21]. Durch Versuch und Irrtum in Kombination mit Belohnungen wird ein Lernagent geschult, sodass er je nach angenommenen Zuständen und Aktionen Belohnungen maximiert [18, 22, 23].

3 Methodisches Vorgehen

Zur Beantwortung der Forschungsfragen wurde das Rahmenwerk zur strukturierten Literaturrecherche nach vom Brocke et al. [24] verwendet.

Rechercheumfang. Der festzulegende Rechercheumfang basiert auf der Taxonomie nach Cooper [25]. Als Fokus gelten *Forschungsergebnisse* und *Anwendungen*. Das Ziel ist die *Integration* der Forschungsbereiche ML und CEP. Es wird eine *neutrale Darstellung* mit *repräsentativer Abdeckung* gewählt. Die Ergebnisse selbst werden *konzeptionell* zur Beantwortung der Forschungsfragen organisiert. Als Zielgruppe haben wir *spezialisierte Forscher* aus den Bereichen von ML und oder CEP sowie *allgemeine Forscher* aus der Wirtschaftsinformatik bestimmt.

Konzeptualisierung. Das Suchergebnis muss ein Verfahren aus der Domäne des ML im CEP einsetzen und auf Anwendungen in Smart Factories übertragbar sein.

Literatursuche. Bei der *Datenbankauswahl* wurden sieben Datenbanken aus den Bereichen der Informatik und Betriebswirtschaftslehre gewählt (vgl. *Abbildung 1*). Für die *Keyword-Suche* wurden gängige Begriffe aus dem CEP genutzt (Kurzform: *Event & Processing | -driven*) und mit dem engl. Oberbegriff „Machine Learning“ verbunden.¹

¹ Eine Übersicht der definierten Suchterme je wissenschaftlicher Datenbank sowie die vollständigen Referenzen zu Tabelle 1 können im digitalen Anhang abgerufen werden unter: <https://bit.ly/2qSdhP8>.

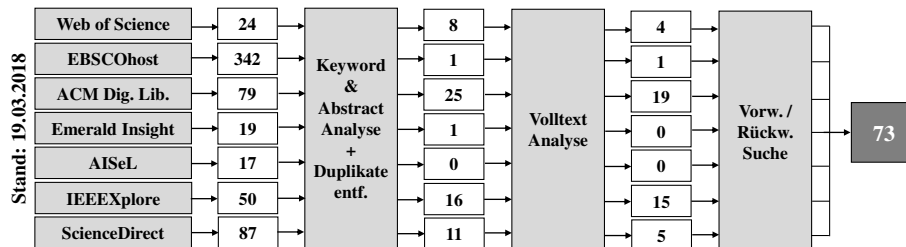


Abbildung 1: Prozess der Literatursuche

Es wurden im ersten Schritt 618 potentiell relevante Werke identifiziert. Nach der Analyse aller Werke bzgl. Titel, Abstract, Keywords und Duplikaten konnten 62 relevante Publikationen ermittelt werden. Die Volltextanalyse reduzierte die Anzahl auf 44. Ein Gros an Suchergebnisse fand sich in informatiknahen Datenbanken (*IEEEExplore* und *ACM Digital Library*). Letztere publiziert u. a. Beiträge der für das Thema hoch relevanten Konferenz Distributed Event-Based Systems (DEBS). Die *Rückwärtssuche* wurde auf Basis der extrahierten Quellen aus den Werken durchgeführt, die *Vorwärtssuche* über Google-Scholar-Zitationsdaten. Dies führte zu weiteren 29 Ergebnissen, sodass insgesamt 73 Beiträge gefunden werden konnten.

4 Auswertung der Literaturanalyse

Ergebnis der Synthese. Die Synthese (vgl. *Abbildung 2*) zeigt, dass das Lernparadigma *Supervised Learning* im CEP dominiert ($n=46$). Der Großteil der Arbeiten behandelt dabei den algorithmischen Ansatz der *Klassifizierung* ($n=23$).

Das Lernparadigma *Unsupervised Learning* findet hingegen weniger Anwendung ($n=6$). *Clustering* zeigt sich hier als dominanter Ansatz ($n=5$), wobei ein Beitrag einen *dimensionsreduktionsbasierten Ansatz* verfolgt.

Beliebter ist die Kombination (Englisch: *Combined Learning*) aus *Unsupervised* und *Supervised Learning* ($n=13$). Die meisten Autoren ($n=12$) nutzen hierbei eine Verbindung aus *Clustering* und *Klassifikation*. Ein Ansatz behandelt *Clustering* erweitert um ein *probabilistisches Modell* ($n=1$).

Reinforcement Learning für CEP wurde bisher nicht beachtet und scheint damit potentieller möglicher Gegenstand weiterer Forschung. Weitere alternative Ansätze, welche sich nicht in das gewählte Klassifikationsschema einordnen ließen, wurden unter *Weitere Beiträge* ($n=8$) aufgelistet. Beispiele finden sich mit Algorithmen aus der Data Mining Domäne [26] oder Eigenentwicklungen wie bei Margara et al. [27] mit deren Windows Learner.

Zeitliche Entwicklung. Die Forschungsdomäne kann als jung bezeichnet werden mit dem frühesten Beitrag aus dem Jahr 2007 [28]. Über die Folgejahre zeigt sich ein steigendes Forschungsinteresse mit Höhepunkten in den vergangenen Jahren 2016 ($n=18$) und 2017 ($n=17$). Das Jahr 2018 ist nicht vollständig erfasst. Diese Tendenz verdeutlicht die Aktualität und Themenrelevanz sowie die verschiedenartigen Anwendungsmöglichkeiten des ML im CEP.

Auf Ebene der Lernparadigmen ist festzustellen, dass anfangs vor allem *Supervised Learning* betrachtet wurde. Hierbei lag der Fokus zunächst auf Klassifizierungsansätzen. Ab dem Jahr 2013 findet eine Diversifikation statt. Weitere Ansätze, wie bspw. regressionsbasierte Verfahren, finden vermehrt Betrachtung. Dennoch bleiben Klassifikationsverfahren durchgehend dominant. *Unsupervised Learning* ist in der CEP-Forschung hingegen (bisher) unterrepräsentiert. Die erste Publikation aus dem Jahr 2010 verfolgt einen dimensionsreduktionsbasierten Ansatz. Ab 2013 werden vereinzelt Clustering-basierte Beiträge veröffentlicht, wobei das Maximum im Jahr 2017 (n=2) liegt. Ein stärkeres Interesse zeigt sich bei der *Kombination* beider Paradigmen. Nach nur einer Publikation im Jahr 2014, finden sich 2016 fünf und 2017 sieben Beiträge.

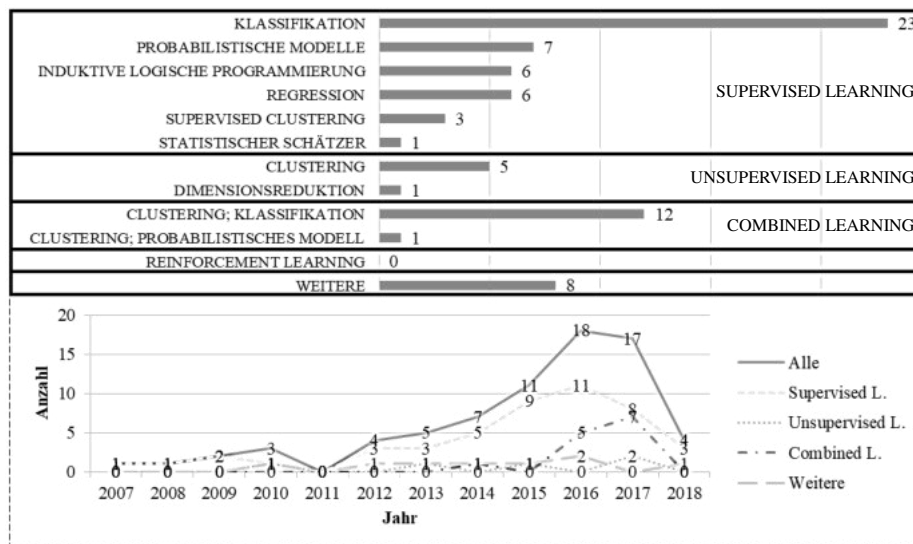


Abbildung 2: Zuordnung der Beiträge zu den ML-Lernparadigmen

Fazit. Im CEP findet überwiegend *Supervised Learning* Anwendung. Dies lässt sich aufgrund des reaktiven Charakters in Kombination mit historischen Eventdaten erklären. *Unsupervised Learning* wird dementsprechend meist in *Kombination* eingesetzt. Damit lassen sich Vorteile aus beiden Paradigmen nutzen, wobei hierbei weiterer Forschungsbedarf besteht. Eine fundierte Eignungsprüfung des Ansatzes *Reinforcement Learning* für CEP fehlt hingegen. Es zeigt sich aus der historischen Analyse, dass die Forschung im Bereich CEP und ML an Wirkungskraft gewinnt und nicht nur punktuell einzelne Verfahren betrachtet. Es ist vielmehr ein Trend zu erkennen, dass verschiedene Verfahren kombiniert werden, um einen erkennbaren Mehrwert in der Echtzeitverarbeitung von Events zu schaffen.

5 Anwendungspotenziale in Smart Factories

5.1 Zielbereiche von Produktionsanlagen

Aus betriebswirtschaftlicher Perspektive bestehen in Produktionsanlagen vier Zielbereiche [29]: Kosten, Qualität, Zeit und Flexibilität. Unter *Kosten* wird mehrheitlich die bestmögliche Allokation der eigenen Ressourcen verstanden. *Qualität* umschreibt die Annahme und Schätzung einer Übereinkunft des Produkts mit den eigenen Zielvorgaben basierend auf dem Produktionsprozess. *Zeit* definiert sich als Quelle eines Wettbewerbsvorteils i. S. v. Ansätzen wie Just-in-time und als fundamentale Messinstanz für die Produktionsperformance. *Flexibilität* beinhaltet die Fähigkeit eines Industrieherstellers, sein Produktionssystem effizient zu verändern – sowohl hinsichtlich der Produktvielfalt als auch der Herstellmenge [30].

Dieselben Zielbereiche bestehen auch in Smart Factories [4]. Durch den Einsatz von CPS und der verbundenen Infrastruktur sollen potenzielle Verbesserungen in allen vier Bereichen realisierbar sein [4]. Um Potenziale durch die Kombination aus CEP und ML aufzuzeigen, wird im nachfolgenden Kapitel von derartigen Industrieanlagen ausgegangen.

5.2 Synthese potenzieller Lösungsansätze durch CEP und ML

Synthese. Auf Basis der Rechercheergebnisse lassen sich sowohl reaktive als auch proaktive Einsatzpotenziale für Smart Factories erkennen. Wie in *Tabelle 1* ersichtlich überwiegen Forschungsbemühungen zu reaktiven Möglichkeiten. Ebenso sind die in *Abbildung 2* unter *Weitere* aufgeführten Beiträge ausschließlich auf reaktive Maßnahmen gerichtet. Aus Platzgründen wurde jedoch auf die Aufnahme in *Tabelle 1* verzichtet, zumal es sich weitestgehend um Spezialfälle handelt oder kein spezifischer Algorithmus genannt wird. Ein Hinweis zu weiteren Details findet sich in Fußnote 1.

Reaktive Maßnahmen. Reaktive Maßnahmen von CEP in Kombination mit ML sollen effizientere Abläufe für Fehler- bzw. Schadensfälle ermöglichen. Dies erlaubt die Aufbereitung wichtiger Informationen für die Fehlerlokalisierung und -behebung. Bei Beginn des CEP-Monitorings in Industrieanlagen sind allerdings i. d. R. nicht alle Störfälle bekannt. Eine Fortschreibung der Mustererkennung ist daher essentiell. Dies erfordert eine manuelle Identifizierung neuer Anomalien. Die Aufgabe impliziert hohes technisches Wissen, bei vertiefender Domänenkenntnis [5, 6]. Eine Unterstützung kann ML bieten. Damit werden im Idealfall bisher unbekannte Eventmuster vollautomatisch aufdeckbar und für die Zukunft aufbereitet [28, 59].

Einen direkten Anwendungsbezug in Smart Factories bietet der Ansatz von Metz et al. [35]. Sie nutzen Produktionsdaten in Kombination mit Prozessdaten, um mit Hilfe einer ML-Komponente neue Regeln über Entscheidungsbäume zu klassifizieren und diese für die CEP-Einheit zu nutzen. Ebenso findet sich der Einsatz von Kamerasystemen und bildverarbeitenden Sensoren in Produktionsprozessen [42]. Diese bieten wertvolle Informationen für Smart Factories, um eine präzisere Mustererkennung und -adaption zu realisieren.

Tabelle 1: Potenziale beim Einsatz von CEP und ML in Smart Factories (für die vollständige Einordnung aller analysierten Quellen vgl. Fußnote 1)

Lernparadigma	ML-Ansatz	Algorithmus	Σ	Beispiel	Σ	Beispiel
Supervised Learning	Klassifikation	Support Vector Machine	4	[31]	1	[32]
		Shapelet-basierte Algorithmen	2	[33]	3	[34]
		Entscheidungsbaum	2	[35]	2	[36]
		Künstlich neuronale Netze	4	[12]	1	[37]
		Sliding Window, Aggressive and Pessimist Classifier	1	[38]	0	
		Markov Model	1	[7]	1	[39]
		Diskriminanzanalyse	3	[12]	0	
		K-nearest Neighbors	1	[31]	0	
		Naive Bayes	1	[31]	1	[36]
		FURLA	2	[40]	0	
		Bayessche Netzwerke	0		6	[36]
		Markov Model	0	[41]	0	
		Probabilistisches Modell	0		1	[42]
		Conditionale Density Estimation	6	[43]	0	
		Induktive logische Programm.	0		3	[44]
Support Vector Regression	0		1	[44]		
Unsupervised Learning	Clustering	Synthetic Regression	0		2	[45]
		Adaptive Moving Window Regression	0		1	[46]
		k-Means Adaption	0		0	
		Eigenentwicklung	1	[47]	0	
		Discrete Kalman Filter	1	[6]	0	
		Density-based Clustering	2	[49]	0	
		Complete-link and Single-link Clustering	1	[50]	0	
		Fixed-width Clustering	1	[51]	0	
		k-Means	1	[5]	0	
		Principal Component Analysis	1	[32]	0	
Combined Learning	Clustering; Klassifikation	k-Means; Markov Model	0		8	[53]
		k-Means; Gaussian Mixture Model	1	[54]	0	
		Hierarchical Clustering; Entscheidungsbaum; k-Means; Naive Bayes, SVM	0		1	[55]
		Unspezifizierter Clustering-Algorithmus; Naive Bayes	1	[56]	0	
		Density-based Clustering; Bag-of-words; Markov Model	0		1	[57]
		k-Means; Complementary Naive Bayes; ALSWR	0		1	[58]
		Clustering; Probabilistisches Modell; Topic Mining				
		Clustering; Klassifikation, Kollaboratives Filtern				

Darüber hinaus werden Sensornetze in Kombination mit Klassifizierungsalgorithmen genutzt [38]. Die Erkennung anomalen Verhaltens erfolgt dabei über Analysen von Events aus Druck-, Vibrations-, Temperatur- und Näherungssensoren [60]. Auch Ansätze aus weiteren Forschungsbereichen zeigen Potenzial für die Problemlösung auf. Einige der Autoren [41, 43] setzen auf Videoüberwachungs-Datensätze, um Aktivitäts- bzw. Eventmuster zu erlernen. Hierbei wird bis auf eine Ausnahme einer Markov-Model-Verwendung in [41] auf Induktive Logische Programmierung gesetzt. Die Anwendung erfolgt allerdings außerhalb von Smart Factories, ist aber potenziell übertragbar.

Weitere Autoren [40, 56] nutzen Beschleunigungssensor-basierte Ereignisströme, um automatisiert Bewegungsmuster von Personen zu erkennen. Für das ML wird bei [40] eine Klassifizierung mittels des Fuzzy Unordered Rule Induction Algorithm (FURIA) verwendet. [56] nutzen hingegen eine Kombination aus Clustering und Klassifikation auf Basis von Naive Bayes und eines unspezifizierten Clustering-Ansatzes. In produktionsspezifischen Szenarien bietet dies jeweils u. a. die Möglichkeit zur Überwachung und Steuerung von industriellen Roboterarmen [61]. Eine manuelle Implementierung der Bewegungsprofile als Events in einem CEP-System wäre dagegen mit großem zeitlichem Aufwand und hoher technischer Expertise (Kosten) verbunden.

Ein weiteres Anwendungsfeld ist die Netzwerksicherheit. Durch den Einsatz von CPS werden neue Cyber-Angriffsflächen eröffnet, welche eine Betriebsstörung oder Produktionsausfälle verursachen können [62]. Zur Prävention resultierender Zeitverluste und Kosten setzen [6, 26] auf CEP-basierte Intrusion Detection Systeme, welche um ML-Komponenten erweitert werden. [6] nutzen hierfür Discrete Kalman Filters, wohingegen [26] eine Kombination aus Markov Logic und Data Mining Ansätzen einsetzen. In beiden Fällen wird eine dynamische und adaptive Anpassung der Angriffsmuster möglich ohne die Notwendigkeit von manuellen Eingriffen und wiederkehrender, aufwändiger Nacharbeitung.

Durch die aufgezeigten Maßnahmen werden primär die manuelle Überwachungstätigkeit und die Implementierung von deklarativen Regeln im CEP-System entfallen. Dies resultiert in einer Verbesserung der Reaktivität in Produktionsstätten und spart *Zeit* und *Kosten* ein.

Proaktive Maßnahmen. Zukünftige Events werden im CEP standardmäßig nicht berücksichtigt [42]. Eine proaktive und frühe Reaktion auf Events erhöht jedoch den Geschäftswert des Systems [32]. Mit der Erweiterung um ML lässt sich dies auch im CEP-Umfeld erreichen. Damit werden Vorhersagen und proaktives Handeln auf Basis historischer Eventdaten durch Schwellenbereiche möglich.

[42] setzen auf ein Conditional Density Estimation trainiertes ML-CEP-System, welches Eventdaten von Kamerasystemen und Sensoren erhält. Der Produktionsdurchlauf wird in Echtzeit auf anormale Veränderungen geprüft und mit Eintrittswahrscheinlichkeiten abgeglichen. Dieselbe Aufgabe findet sich als zentrales Thema der ACM DEBS 2017 Grand Challenge [53]. Zur Datenvorverarbeitung galt es Events auf Basis diskreter Zustände (z. B. Temperatur oder Druck der Produktionsmaschine) durch einen Clustering-Algorithmus zu gruppieren. Daraufhin musste ein Markov-Modell trainiert werden, um Eintrittswahrscheinlichkeiten von Anomalien in Echtzeit

zu berechnen. In allen Lösungsansätzen wurde mit Zustandsübergängen zwischen den Clustern gearbeitet. Das System benachrichtigt den Mitarbeiter sobald die Wahrscheinlichkeit über einem definierten Schwellenwert liegt.

[34] zeigen mit Hilfe eines Shapelet-basierten Ansatzes, dass sich die Vorhersagefähigkeit mit dem automatisieren Lernen von Regeln kombinieren lässt. Ein alternativer Ansatz [36] zielt auf die einheitliche Verwendung eines proaktiven Architekturstandards im industriellen Umfeld ab. Mit Hilfe eines Bayesschen Netzes und einer verbundenen Klassifizierung anhand von Entscheidungsbäumen werden in diesem Beispiel Probleme Zeit-terminierter Wartungsvorgänge unabhängig vom tatsächlichen Maschinenzustand angegangen. Durch eine CEP-Kombination mit einer ML-gestützten Echtzeitanalyse wird der Austausch bei faktischer Notwendigkeit eines Ersatzteils in Abhängigkeit der verbleibenden Lebensdauer angezeigt. Dies erlaubt eine entsprechende Optimierung des Wartungsintervalls.

Weitere Einsatzszenarien finden sich einerseits für intra-logistische Materialbewegungen wie der gezielten Überwachung von Förderbändern [63] sowie andererseits für extra-logistische Materialbewegungen wie z. B. im Straßenverkehr [8, 64, 65]. Dabei werden Eventdaten kontinuierlich über Sensoren, Kamerasystemen oder Induktionsschleifen erfasst, übertragen und zur Vorhersage u. a. von Verkehrsstaus herangezogen. [8] setzen dazu auf eine Adaptive Moving Window Regression, um ein kontinuierliches Lernen des CEP-Systems zu sichern. [64]. [65] nutzen hingegen Bayessche Netze, um probabilistische Complex Events vorherzusagen.

Die Proaktivität der aufgezeigten Ansätze führt vorrangig zur Optimierung der *Auslastung* der Fertigungs-, Wartungs- und Warenflussprozesse. Ebenso resultieren daraus *Kosteneinsparungen* durch proaktiv reduzierte Wartungsprozesse.

6 Diskussion und Fazit

Diskussion. Die unternommene Untersuchung zeigt, dass ein Einsatz von ML großes Potenzial für CEP besitzt. [66] bestätigen dies und bezeichnen den Einsatz von ML als bedeutenden Fortschritt in der Echtzeitauswertung von Events. Trotz der Bedeutung für die Forschungsdomäne fehlte bislang eine umfassende Betrachtung der unterschiedlichen Kombinationsansätze.

Auf Basis der durchgeführten Literaturrecherche konnte festgestellt werden, dass bei der Erweiterung von CEP um ML primär Supervised-Learning-Ansätze eingesetzt werden. Über die letzten Jahre finden sich aber vermehrt Kombinationen aus Supervised- und Unsupervised-Learning-Ansätzen. Dies birgt großes Potenzial für die Umsetzung proaktiven Verhaltens. Reinforcement Learning und weitere Alternativen sind hingegen kaum untersucht worden. Zukünftige Forschung könnte hier ansetzen, um bspw. die Wirksamkeit von belohnungsorientierten Ansätzen für das CEP zu untersuchen. Reinforcement Learning wird bereits in Multi-Agenten-Systemen verwendet. Da Event Processing Agents konzeptionell als autonome Agenten eines umfassenderen Event Processing Networks betrachtet werden, liegt es nahe, die Wirksamkeit dieses Ansatzes auch in diesem Umfeld zu untersuchen.

In einer zweiten Synthese wurden mögliche Einsatzpotenziale von ML in CEP für Smart Factories betrachtet. Für die betriebswirtschaftlichen Zielbereiche von Kosten, Zeit, Flexibilität und Qualität ergaben sich sowohl (1) reaktive als auch (2) proaktive Verbesserungsansätze. Das Potenzial wird überwiegend in der Reduzierung der Zeit (d. h. in Bezug auf bessere Maschinenauslastung) und der Kosten (d. h. in Bezug auf Automatisierungsansätze) gesehen und adressiert. Bisherige Forschung zielte primär auf reaktive Maßnahmen ab, wobei gerade in den letzten beiden Jahren verstärkt auch proaktive Maßnahmen erforscht wurde.

Die Literaturanalyse zeigt, dass für reaktive Maßnahmen in Smart Factories vornehmlich auf Supervised Learning zurückgegriffen wird. Es handelt sich dabei insbesondere auf SVMs und künstlich neuronale Netze aus dem Bereich der Klassifizierung. Ansätze aus dem Unsupervised Learning sind bislang unterrepräsentiert, wobei jedoch erste Testversuche erfolgt sind. Combined Learning wiederum scheint für reaktive Maßnahmen hinsichtlich der fehlenden Verwendung vergleichsweise unbrauchbar. Dies steht im Gegensatz zu proaktiven Maßnahmen. Hier wird zunehmend mit Ansätzen aus dem Combined Learning gearbeitet. Primär erscheint eine Verbindung aus k-Means und dem Markov Model erfolgsversprechend. Auch Anwendungen aus dem Supervised Learning finden sich. Dabei werden überwiegend Bayessche Netze aus dem Bereich der probabilistischen Modelle eingesetzt. Ansätze aus dem Unsupervised Learning bestehen hingegen keine, was deren fehlende Eignung nahelegt.

Fazit und Ausblick. Aktuell werden einige Anwendungen für die kontinuierliche Echtzeitauswertung von Events auf CEP-Basis erprobt. Gerade im Umfeld von Smart Factories zeigen sich erste Umsetzungen primär in reaktiven Bereichen, wie der informationsgestützten Reparatur im Fehlerfall. Auch in Bezug auf Big Data werden Kombinationen aus CEP und ML als erfolgsversprechend angesehen [67]. Der Trend geht hierbei zu probabilistischen bzw. ungewissen Eventmodellen. Im Gegensatz zu deterministischen Eventmodellen sind dabei nicht alle Attribute bekannt oder mit Ungenauigkeiten versehen [67]. Erste Lösungsansätze fokussieren Bayessche und Markov-Netze allerdings mit bisher mäßigem Erfolg [68].

Zukünftige Forschung könnte hier ansetzen und über neue Ansätze wie bspw. das Online Machine Learning [69] weitere Verbesserungen erreichen. Beim Online ML steht die automatische Generierung der vormalig manuell spezifizierten deklarativen CEP-Abfragen im Mittelpunkt, um sowohl eine verbesserte Eventerkennung und -verarbeitung zu erreichen als auch um Kosten zu sparen. Auch die Kombination von CEP mit Ansätzen aus dem Reinforcement Learning ist wie oben beschrieben vielversprechend. Gerade für Smart Factories kann im Hinblick auf die Optimierung der Maschinenlauffähigkeit und für das damit verbundenen Wartungsmanagement eine Verbesserung erfolgen. Dies lässt sich insbesondere durch präskriptive Analyse prognostizieren, wobei hier nicht die Maschine als Einzelkomponente betrachtet werden sollte, sondern die jeweils beste betriebswirtschaftliche Entscheidung im Gesamtunternehmenskontext [70].

Förderhinweis

Dieses Forschungs- und Entwicklungsprojekt wird mit Mitteln des Bayerischen Staatsministeriums für Wirtschaft, Landesentwicklung und Energie (StMWi) innerhalb des Förderprogramms "Informations- und Kommunikationstechnik" (Kennzeichen IUK536/003) gefördert und vom Projektträger VDI+VDE Innovation + Technik GmbH betreut.

Literaturverzeichnis

1. Yin, S., Kaynak, O.: Big Data for Modern Industry: Challenges and Trends. Proceedings of the IEEE 103, 143-146 (2015)
2. Bruns, R., Dunkel, J.: Complex Event Processing: Komplexe Analyse von massiven Datenströmen mit CEP. Springer Vieweg, Berlin (2015)
3. Luckham, D.: Event Processing for Business: Organizing the Real Time Enterprise. John Wiley & Sons, Inc., Hoboken, NJ (2011)
4. Kagermann, H., Wahlster, W., Helbig, J.: Securing the future of German manufacturing industry: Recommendations for implementing the strategic initiative INDUSTRIE 4.0. Final report of the Industrie 4.0 Working Group. acatech (2013)
5. Akbar, A., Carrez, F., Moessner, K., Sancho, J., Rico, J.: Context-aware stream processing for distributed IoT applications. IEEE 2nd World Forum on Internet of Things, pp. 663–668. IEEE Press, New York (2015)
6. Turchin, Y., Gal, A., Wasserkrug, S.: Tuning complex event processing rules using the prediction-correction paradigm. Proceedings of the Third ACM International Conference on DEBS. ACM, New York (2009)
7. Mutschler, C., Philippsen, M.: Learning event detection rules with noise hidden Markov models. Proceedings of the 2012 NASA/ESA Conference on Adaptive Hardware and Systems, pp. 159–166. IEEE Press, New York (2012)
8. Akbar, A., Carrez, F., Moessner, K., Zoha, A.: Predicting complex events for pro-active IoT applications. IEEE 2nd World Forum on Internet of Things, pp. 327–332. IEEE Press, New York (2015)
9. Artikis, A., Paliouras, G., Portet, F., Skarlatidis, A.: Logic-based representation, reasoning and machine learning for event recognition. Proceedings of the Fourth ACM International Conference on DEBS, pp. 282–293. ACM, New York (2010)
10. Mehdiyev, N., Krumeich, J., Werth, D., Loos, P.: Sensor event mining with hybrid ensemble learning and evolutionary feature subset selection model. 2015 IEEE International Conference on Big Data, pp. 2159–2168. IEEE Press, New York (2015)
11. Mousheimish, R., Taher, Y., Zeitouni, K.: autoCEP: Automatic Learning of Predictive Rules for Complex Event Processing. In: Sheng, Q., Stroulia, E., Tata, S., Bhiri, S. (eds.) Service-Oriented Computing. ICSOC 2016. Lecture Notes in Computer Science, vol 9936, pp. 586–593. Springer, Cham, CHE (2016)
12. Widder, A., Ammon, R., Hagemann, G., Schönfeld, D.: An Approach for Automatic Fraud Detection in the Insurance Domain. 2009 AAAI Spring Symposium, pp. 98–100 (2009)
13. Samuel, A.L.: Some studies in machine learning using the game of checkers. IBM Journal of Research and Development 3, 210–229 (1959)
14. Bauernhansl, T.: Die Vierte Industrielle Revolution – Der Weg in ein wertschaffendes Produktionsparadigma. In: Vogel-Heuser, B., Bauernhansl, T., Hompel, M.t. (eds.) Industrie

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

- 4.0 in Produktion, Automatisierung und Logistik, pp. 5-35. Springer Vieweg, Wiesbaden (2014)
15. Etzion, O., Niblett, P.: *Event Processing in Action*. Manning, Stamford, CT (2011)
 16. Marsland, S.: *Machine Learning: An Algorithmic Perspective*. Chapman & Hall/CRC (2015)
 17. Shalev-Shwartz, S., Ben-David, S.: *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press, New York, NY (2014)
 18. Wang, S., Chaovalitwongse, W., Babuska, R.: Machine Learning Algorithms in Bipedal Robot Control. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 42, 728–743 (2012)
 19. Ghahramani, Z.: Unsupervised Learning. In: Hutchison, D., Kanade, T., Kittler, J. (eds.) *Advanced Lectures on Machine Learning*, pp. 72–112. Springer, Berlin/Heidelberg (2004)
 20. Kubat, M.: *An Introduction to Machine Learning*. Springer International Publishing, Cham, CHE (2017)
 21. Sutton, R.S., Barto, A.G.: *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA (2017)
 22. Alpaydin, E.: *Introduction to Machine Learning*. MIT Press, Cambridge, MA (2004)
 23. Tokic, M.: Reinforcement Learning: Psychologische und neurobiologische Aspekte. *Künstliche Intelligenz* 27, 213–219 (2013)
 24. vom Brocke, J., Simons, A., Niehaves, B., Reimer, K., Plattfaut, R., Cleven, A.: Reconstructing the giant: on importance of rigour in documenting the literature search process. *ECIS 2019*, vol. 9, pp. 2206-2217, Verona (2009)
 25. Cooper, H.M.: Organizing knowledge syntheses. A taxonomy of literature reviews. *Knowledge in Society* 1, 104–126 (1988)
 26. Bhandari, S.: An adaptive event stream processing environment. *Proceedings of the on SIGMOD/PODS 2012 PhD Symposium*, pp. 57–62. ACM, New York (2012)
 27. Margara, A., Cugola, G., Tamburrelli, G.: Towards automated rule learning for complex event processing. (2013)
 28. Widder, A., Ammon, R., Schaeffler, P., Wolff, C.: Identification of suspicious, unknown event patterns in an event cloud. *Proceedings of the 2007 Inaugural International Conference on DEBS*, pp. 164–170. ACM, New York (2007)
 29. Shepherd, C., Günter, H.: Measuring supply chain performance: current research and future directions. *Behavioral Operations in Planning and Scheduling*, pp. 105-121. Springer, Berlin, Heidelberg (2010)
 30. Neely, A., Gregory, M., Platts, K.: Performance measurement system design: A literature review and research agenda. *International Journal of Operations & Production Management* 15, 80-116 (1995)
 31. Gad, R., Kappes, M., Medina-Bulo, I.: Analysis of the Feasibility to Combine CEP and EDA with Machine Learning using the Example of Network Analysis and Surveillance. *X Jornadas de Ciencia e Ingeniería de Servicios*, (2014)
 32. Schwegmann, B., Matzner, M., Janiesch, C.: preCEP: Facilitating Predictive Event-Driven Process Analytics. In: vom Brocke, J., Hekkala, R., Ram, S., Rossi, M. (eds.) *Design Science at the Intersection of Physical and Virtual Design. DESRIST 2013. Lecture Notes in Computer Science*, vol 7939, pp. 448–455. Springer, Berlin/Heidelberg (2013)
 33. Patri, O.P., Panangadan, A., Sorathia, V., Prasanna, V.: Sensors to Events: Semantic Modeling and Recognition of Events from Data Streams. *International Journal of Semantic Computing* 10, 461–471 (2016)

34. Mousheimish, R., Taher, Y., Zeitouni, K.: Automatic Learning of Predictive CEP Rules: Bridging the Gap between Data Mining and Complex Event Processing. Proceedings of the 11th ACM International Conference on DEBS, pp. 158–169. ACM, New York (2017)
35. Metz, D., Karadgi, S., Müller, U., Grauer, M.: Self-Learning Monitoring and Control of Manufacturing Processes Based on Rule Induction and Event Processing. The Fourth International Conference on Information, Process, and Knowledge Management, pp. 88–92. IARIA (2012)
36. Engel, Y., Etzion, O.: Towards proactive event-driven computing. Proceedings of the 11th ACM International Conference on DEBS, pp. 125–136. ACM, New York (2017)
37. Borkowski, M., Fdhila, W., Nardelli, M., Rinderle-Ma, S., Schulte, S.: Event-based failure prediction in distributed business processes. Information Systems (2017)
38. Shahad, R.A., Bein, L.G., Saad, M. H., Hussain, A.: Complex event detection in an intelligent surveillance system using CAISER platform. International Conference on Advances in Electrical, Electronic and Systems Engineering, pp. 129–133. IEEE Press, New York (2016)
39. Baldoni, R., Montanari, L., Rizzut, M.: On-line failure prediction in safety-critical systems. Future Generation Computer Systems 45, 123–132 (2015)
40. Mehdiyev, N., Krumeich, J., Werth, D., Loos, P.: Determination of Event Patterns for Complex Event Processing Using Fuzzy Unordered Rule Induction Algorithm with Multi-objective Evolutionary Feature Subset Selection. 49th Hawaii International Conference on System Sciences, pp. 1719–1728. IEEE Press, New York (2016)
41. Skarlatidis, A., Paliouras, G., Artikis, A., Vouros, G.A.: Probabilistic Event Calculus for Event Recognition. ACM Transactions on Computational Logic 16, 11–47 (2015)
42. Christ, M., Krumeich, J., Kempa-Liehr, A.W.: Integrating Predictive Analytics into Complex Event Processing by Using Conditional Density Estimations. 20th International Enterprise Distributed Object Computing Workshop, pp. 287–294. IEEE Press, New York (2016)
43. Katzouris, N., Artikis, A., Paliouras, G.: Incremental learning of event definitions with Inductive Logic Programming. Machine Learning 100, 555–585 (2015)
44. Lang, J., Capik, Z.: Prediction based on hybrid method in complex event processing. 12th International Symposium on Applied Machine Intelligence and Informatics, pp. 315–320. IEEE Press, New York (2014)
45. Akbar, A., Khan, A., Carrez, F., Moessner, K.: Predictive Analytics for Complex IoT Data Streams. IEEE Internet of Things Journal 4, 1571–1582 (2017)
46. Roledene, S., Ariyathilaka, L., Liyanage, N., Lakmal, P., Bamunusinghe, J.: GeniBux - event based intelligent Forex trading strategy enhancer. IEEE International Conference on Information and Automation for Sustainability, pp. 1–6. IEEE Press, New York (2016)
47. Lee, O.-J., Jung, J.E.: Sequence Clustering-based Automated Rule Generation for Adaptive Complex Event Processing. Future Generation Computer Systems 66, 100–109 (2017)
48. Cuzzocrea, A., Folino, F., Guarascio, M., Pontieri, L.: Predictive Monitoring of Temporally-Aggregated Performance Indicators of Business Processes against Low-Level Streaming Events. Information Systems In Press, Accepted Manuscript, (2018)
49. Junior, M.R., Endler, M., Silva e Silva, F.J.D.: An on-line algorithm for cluster detection of mobile nodes through complex event processing. Information Systems 64, 303–320 (2017)
50. Sugiura, K., Ishikawa, Y., Sasaki, Y.: Grouping Methods for Pattern Matching over Probabilistic Data Streams. IEICE Transactions on Information and Systems E100D, 718–729 (2017)
51. Gander, M., Felderer, M., Katt, B., Tolbaru, A., Breu, R., A., M.: Anomaly Detection in the Cloud: Detecting Security Incidents via Machine Learning. In: Moschitti, A., Plank, B.

- (eds.) Trustworthy Eternal Systems via Evolving Software, Data and Knowledge. *EternalS 2012. Communications in Computer and Information Science*, vol 379, pp. 103–116. Springer, Berlin/Heidelberg (2013)
52. Teixeira, P., Clemente, R.G., Kaiser, R.A., Viera Jr., D.A.: HOLMES: an event-driven solution to monitor data centers through continuous queries and machine learning. *Proceedings of the Fourth ACM International Conference on DEBS*, pp. 216–221. ACM, New York (2010)
 53. Rivetti, N., Busnel, Y., Gal, A.: FlinkMan: Anomaly Detection in Manufacturing Equipment with Apache Flink: Grand Challenge. *Proceedings of the 11th ACM International Conference on DEBS*, pp. 274–279. ACM, New York (2017)
 54. Chandrathilake, H.M.C., Hewawitharana, H.T.S., Jayawardana, R.S., Viduranga, A.D.D., Dilum Bandara, H.M.N., Marru, S., Perera, S.: Reducing computational time of closed-loop weather monitoring: A Complex Event Processing and Machine Learning based approach. *Moratuwa Engineering Research Conference*, pp. 78–83. IEEE Press, New York (2016)
 55. Sejdovic, S., Hegenbarth, Y., Ristow, G.H., Schmidt, R.: Industry Paper: Proactive Disruption Management System: How not to be Surprised by Upcoming Situations. *Proceedings of the 10th ACM International Conference on DEBS*, pp. 281–288. ACM, New York (2016)
 56. Preuveneers, D., Berbers, Y., Joosen, W.: SAMURAI: A batch and streaming context architecture for large-scale intelligent applications and environments. *Journal of Ambient Intelligence and Smart Environments* 8, 63–78 (2016)
 57. Terroso-Sáenz, F., Cuenca-Jara, J., Gonzalez-Vidal, A.: Human Mobility Prediction Based on Social Media with Complex Event Processing. *International Journal of Distributed Sensor Networks* 12, 1–25 (2016)
 58. Zhang, W., Hirzel, M., Grove, D.: AQUA: adaptive quality analytics. *Proceedings of the 10th ACM International Conference on DEBS*, pp. 169–180. ACM, New York (2016)
 59. Peterson, E., To, M.A., Maag, S.: A novel online CEP learning engine for MANET IDS. *9th IEEE Latin-American Conference on Communications*. IEEE Press, New York (2017)
 60. Gungor, V.C., Hancke, G.P.: *Industrial Wireless Sensor Networks: Challenges, Design Principles, and Technical Approaches*. *IEEE Transactions on Industrial Electronics* 56, 4258–4265 (2009)
 61. Neto, P.P., N. J.; Moreira, A. P: Accelerometer-based control of an industrial robotic arm. *The 18th IEEE International Symposium on Robot and Human Interactive Communication*, pp. 1192–1197 IEEE Press, New York (2009)
 62. Mitchell, R., Chen, L.-R.: A survey of intrusion detection techniques for cyber-physical systems. *ACM Computing Surveys* 46, Article No. 55 (2014)
 63. Beamon, B.M.: System reliability and congestion in a material handling system. *Computers & Industrial Engineering* 36, 673–684 (1999)
 64. Akbar, A., Kousiouris, G., Pervalz, H., Sancho, J., Ta-Shima, P., Carrez, F., Moessner, K.: Real-Time Probabilistic Data Fusion for Large-Scale IoT Applications. *IEEE Access* 6, 10015–10027 (2018)
 65. Wang, Y., Gao, H., Chen, G.: Predictive complex event processing based on evolving Bayesian networks. *Pattern Recognition Letters* In Press (2017)
 66. Dayarathna, M., Perera, S.: Recent Advancements in Event Processing. *ACM Computing Surveys* 51, Article No. 33 (2018)
 67. Flouris, I., Giatrakos, N., Deligiannikis, A., Garfalakis, M., Kamp, M., Mock, M.: Issues in complex event processing: Status and prospects in the Big Data era. *Journal of Systems and Software* 127, 217–336 (2017)

68. Alevizos, E., Skarlatidis, A., Artikis, A., Paliouras, G.: Probabilistic Complex Event Recognition: A Survey. *ACM Computing Surveys* 50, 71-101 (2017)
69. Peterson, E., To, M.A., Maag, S.: An online learning based approach for CEP rule generation. 8th IEEE Latin-American Conference on Communications, pp. 1–6. IEEE Press, New York (2016)
70. Wang, G., Gunasekaran, A., Ngai, E.W., Papadopoulos, T.: Big data analytics in logistics and supply chain management: Certain investigations for research and applications. *International Journal of Production Economics* 176, 98-110 (2016)