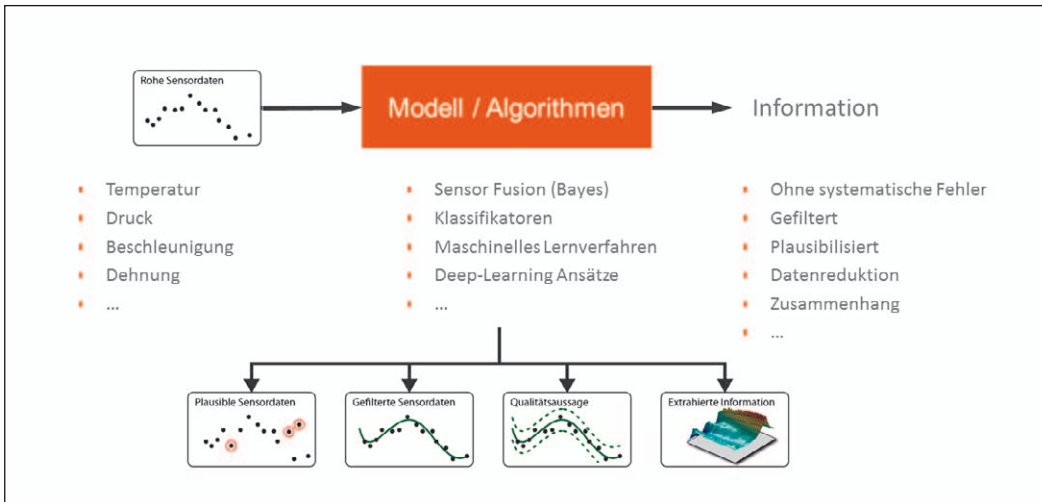


## Herausforderungen und Ansätze bei der Datenverarbeitung



**Bild 1: Konversion der Sensordaten auf die gewünschte Information**

Bereits heute und verstärkt in den nächsten Jahren werden die von einem technischen Prozess erhobenen Messdaten dazu dienen, Kosten beim Betreiben des Prozesses zu reduzieren. Die Erkennung von Einsparpotentialen bei der Optimierung des Prozesses oder die Verringerung von Stillstandzeiten führen dabei zur Kostenersparnis.

Das Kernthema hier ist die Verarbeitung von historischen und aktuellen Daten, um damit eine kontinuierliche Erfassung der Information über einen (technischen) Prozess zu ermöglichen. Dabei kommen eine Vielzahl grundlegender Techniken aus der Informationstheorie, dem maschinellen Lernen und Data Mining zur Anwendung. Aus der Analyse historischer Messwerte werden unbekannte Zusammenhänge offengelegt, oder Prognosen über das zukünftige Systemverhalten bestimmt, wie es beispielsweise in Bild 1 dargestellt ist. Diese gewonnenen Erkenntnisse ermöglichen eine kontinuierliche Optimierung von Systemen, Prozessen, und Strategien.

### Heterogene Daten erfordern spezielle Algorithmen

Allerdings weisen die erhobenen Messdaten einen heterogenen Charakter auf, was eine besondere Herausforderung für die Datenverarbeitung darstellt. Das Verhalten und die Charakteristik der Prozess-

daten unterscheidet sich sehr stark je nach physikalischer Herkunft der Messgröße, beispielsweise Druck, Temperatur oder Luftfeuchte. Weiterhin können Prozessdaten in der Anzahl der Dimension sehr stark variieren, beispielsweise müssen die Daten von Beschleunigungssensoren und Kamerabilder unterschiedlich verarbeitet und miteinander verknüpft werden. Der Einsatz von RFIDs, Barcodes, und die diskreten Prozessschritte während der Wertschöpfung machen weiterhin die Unterscheidung in kontinuierliche und zeit-, wertdiskrete Prozessdaten notwendig. Hinzu kommen qualitative Unterschiede bei den eingesetzten Sensoren bzw. den Prozessgrößen: Beispielsweise müssen Größen aus einer Sichtprüfung anders bewertet werden als von einem physikalischen Sensor; auch bei den Sensoren untereinander gibt es Unterschiede in Genauigkeit und Güte die bei der Verarbeitung und Verknüpfung berücksichtigt werden müssen. Zudem müssen unterschiedliche Abstraten, Latenzen und Datenausfällen bei der Datenverarbeitung berücksichtigt werden. Diese Heterogenität stellt eine große Herausforderung dar, mit denen die Algorithmen umgehen müssen.

### Bereinigung von Störungseffekten

Eine umfassende, echtzeitbasierte und vor allem robuste Datenver-

arbeitung in der Produktion erfordert neben der Erhebung verschiedener Erfassungsdaten die Plausibilisierung und Bereinigung von Störungseffekten. Im Vergleich zu Messeinrichtungen in kontrollierter Umgebung unterliegen die mit Sensoren erfassten Messwerte an Produktionsanlagen starken mechanischen und thermischen Beanspruchungen. Dies führt dazu, dass die relevante Information in den Messsignalen durch Störsignale verschiedenster Art überlagert wird. Nur durch die frühzeitige Plausibilisierung von Sensordaten kann verhindert werden, dass offensichtlich fehlerhafte Messwerte für die Verarbeitung herangezogen werden. Die Wahrung der Datenintegrität ist eine weitere wichtige Herausforderung.

### Modellbasierte Ansätze

Für die Verarbeitung der Daten können modellbasierte Ansätze oder datengetriebene Ansätze verwendet werden. Eine Auflistung verschiedener Verfahren sind in Bild 2 dargestellt. Modellbasierte Verfahren, wie z.B. Bayes'sche Schätzverfahren, können zielführend zur Darstellung von Sensorinformationen bzw. Zustandsaussagen (u.a. durch Datenfusion), zur Plausibilitätssteigerung, Datenreduktion und Extraktion höherwertiger Information eingesetzt werden. Im Allgemeinen wird ein explizites mathematisches Modell angenommen, welches die physikalischen (bzw. chemischen) Eigenschaften und Zusammenhänge des zu verfolgenden Prozesses beschreibt: Beispielsweise zeitunabhängige Stoffgesetze, oder analytische Lebensdauermodelle. Je genauer die Modellierung des abzubildenden physikalischen Prozesses erfolgt, desto genauer ist in Folge das Ergebnis der Zustandsdaten. Hierzu ist eine Vielzahl an zunächst unbekanntem bzw. noch zu bestimmenden physikalischen und chemischen Parametern zu implementieren.

### Statistische Analyseverfahren

- insbesondere die maschinellen Lernverfahren - eignen sich insbe-

### Autor:

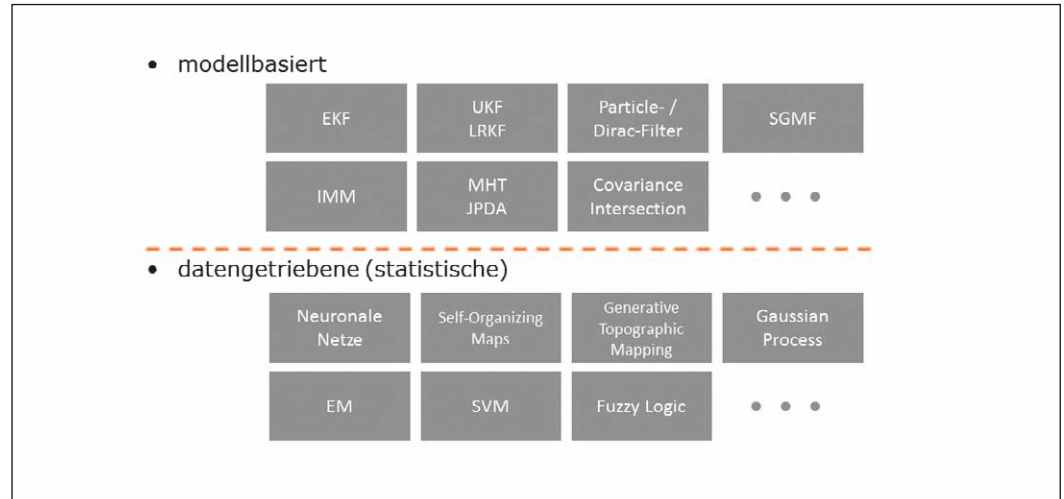


**Dr.-Ing. Frederik Beutler,**  
Geschäftsführer Knowtion  
UG – Spezialist für  
Sensorfusion und  
automatische Datenanalyse

sondere zur Plausibilitätssteigerung sowie der Extraktion höherwertiger Information, d.h. dem automatisierten Erkennen fachtechnischer Sachverhalte. Die besondere Eigenschaft der maschinellen Lernverfahren liegt im Gegensatz zu den modellbasierten Analyseverfahren in ihrer selbstständigen Lernfähigkeit. So können diese Verfahren basierend auf einem Trainingsdatensatz selbstständig bestimmte Aufgaben erlernen, ohne explizit dafür programmiert zu werden - es ist kein physikalisches Modell erforderlich. Die Leistungsfähigkeit, und die Möglichkeit der Erfüllung der gestellten Aufgaben, hängen von der Struktur und Anordnung der Verfahren ab. Die Herausforderung bei der Entwicklung derartiger Algorithmen besteht darin, die Struktur für bestimmte Aufgabenstellungen zu finden und zu optimieren. Ist diese einmal für eine bestimmte Problemklasse bzw. Sensortyp definiert, können die Verfahren diese Struktur automatisch durchführen und dessen Parametrierung und somit Verhalten selbstständig erlernen.

## Identifikations- und Erkennungsphase

In der Praxis findet die Datenverarbeitung typischerweise in zwei Phasen statt: Der Identifikationsphase und der Erkennungsphase. In der Identifikationsphase wird basierend auf den Messdaten ein Modell generiert. Dabei kann dies ein parametrisiertes Modell oder ein Datenmodell sein. Bei einem parametrisierten Modell wird ein mathe-



**Bild 2: Auflistung verschiedener modellbasierter und datengetriebener Verfahren**

mathematisches Modell angenommen und die Parameter des Modells werden in der Phase gelernt. Bei einem Datenmodell hängt das Modell rein von den Daten ab. Der größte Unterschied bei diesen Ansätzen ist, dass ein parametrisches Modell einfacher interpretiert und analysiert werden kann als ein Datenmodell. Als Resultat der Identifikationsphase erhält man ein Modell, welches das Normalverhalten des Systems beschreibt. Dieses Normalverhalten wird im Verfahren so repräsentiert, dass Aspekte der zeitlichen und örtlichen Abhängigkeit kodiert werden. Sobald diese abgeschlossen ist, ist das Verfahren in der Erkennungsphase in der Lage, Abweichungen vom gelernten Modell in Echtzeit zu erkennen und diese dem Benutzer als Anomalieindikatoren zur Verfügung zu stellen.

## Fazit

In diesem Artikel wurden Herausforderungen und Ansätze der Datenverarbeitung in Industrie-4.0-Umfeld dargestellt. Neben vielen Chancen besteht die größte Herausforderung in dem unterschiedlichen Verständnis und Sprache der informationstechnisch geprägten und der produktionstechnisch geprägten Unternehmen. In den meisten Fällen ist es so, dass die Unternehmen aus dem Produktionsumfeld zwar die eigenen Produktionsanlagen bis ins letzte Detail kennen und auch ein sehr gute Vorstellung davon haben, welche zusätzliche Prozessinformation das Potential für weitere Effizienzsteigerung haben, aber nicht wissen wie sie aus der Vielzahl an Prozessdaten die gewünschten Information extra-

hieren können. Hier besteht großes Potential in der engen Kooperation mit spezialisierten, informationstechnisch geprägten Anbietern, die verschiedenste Algorithmen entwickeln und softwaretechnisch umsetzen können.

Auf der Sensor+Test 2017 präsentiert die Knowtion UG weitere Informationen in folgenden Fachbeiträgen:

- Halle 1, 30.05.2017, 10:30-11:00, Datenverarbeitung in Industrie-4.0-Anwendungen
- Halle 5, 31.05.2017, 14:30-15:00, Erhöhung der Sensorgenauigkeit durch Software

**Sensor + Test  
Halle 1, Stand 253**

■ Knowtion UG  
[www.knowtion.de](http://www.knowtion.de)

