
GRUNDLAGEN DES MASCHINELLEN LERNENS

Was wird gemacht? Wie wird es gemacht?

Ist das wirklich Lernen?



**Außenstelle für Verarbeitungsmaschinen und
Verpackungstechnik**

Dresden, 25. September 2017

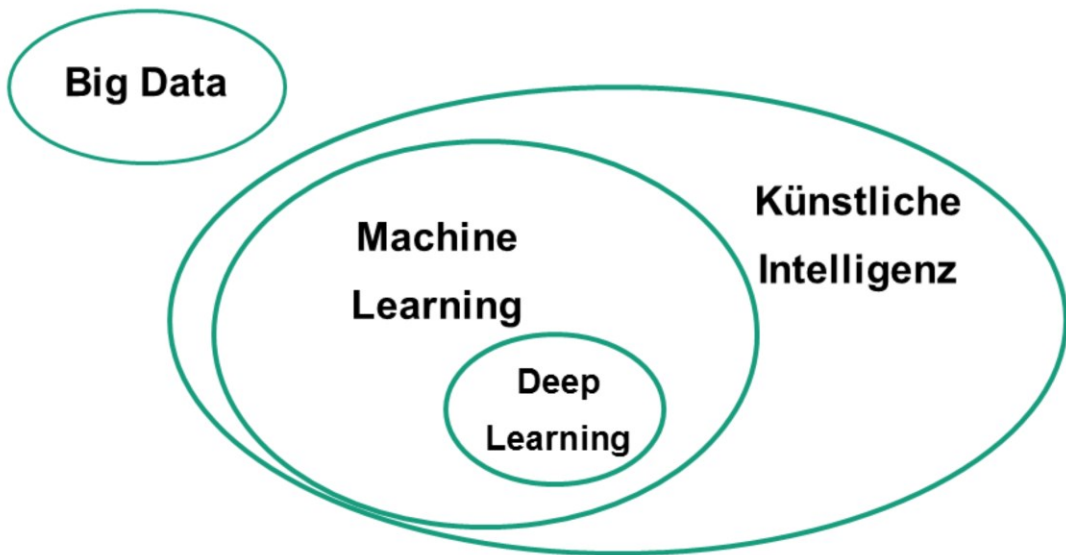
Dipl.-Ing. Tilman Klaeger

© Fraunhofer

- Tilman Klaeger
 - Dipl.-Ing. Mechatronik
 - Forschungsgebiet Assistenzsysteme, maschinelles Lernen am Fraunhofer IVV seit Juli 2016

Big Data, Machine Learning, Künstliche Intelligenz?

Was bedeutet das?



© Fraunhofer

 **Fraunhofer**
IVV Dresden

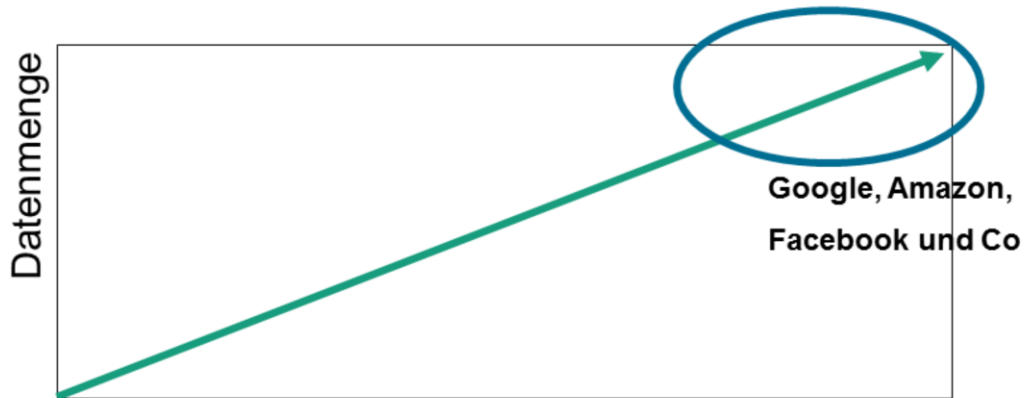
- **Big Data:**
 - Mehr Daten als in den Arbeitsspeicher passen
 - Spezielle Methoden erforderlich, z.B. MapReduce
 - Beispiel: Google Suchindex => Auf einem Rechner ist diese Datenmenge nicht handhabbar.

- **Machine Learning**
 - Automatisches Erkennen von Zusammenhängen in Datensätzen
 - Kerntechnologie der künstlichen Intelligenz

- **Künstliche Intelligenz**
 - Agiert autonom und verhält sich intelligent [Hecker 17].

- **Deep Learning**
 - Einsatz großer (tiefer) künstlicher Neuronaler Netze
 - Vorteil: Kein menschliches Wissen über Prozess im Modell erforderlich

Zur Menge der Daten



© Fraunhofer

 **Fraunhofer** 
IVV Dresden

- Aufregende Ergebnisse v.a. aus den USA
 - Datenmenge für Google, Facebook riesig, damit gigantisches Potenzial
 - Alpha GO: Computer gewinnt im Go-Spiel. Datensätze gewonnen indem Computer gegen Computer spielt

Was heißt „viele Daten“?

Handschrift



■ 60.000 Datensätze

■ Fehlerrate 0,21 %

■ *MNIST*

Bilder



■ >14 Millionen

■ Fehlerrate <5%

■ *ImageNet*

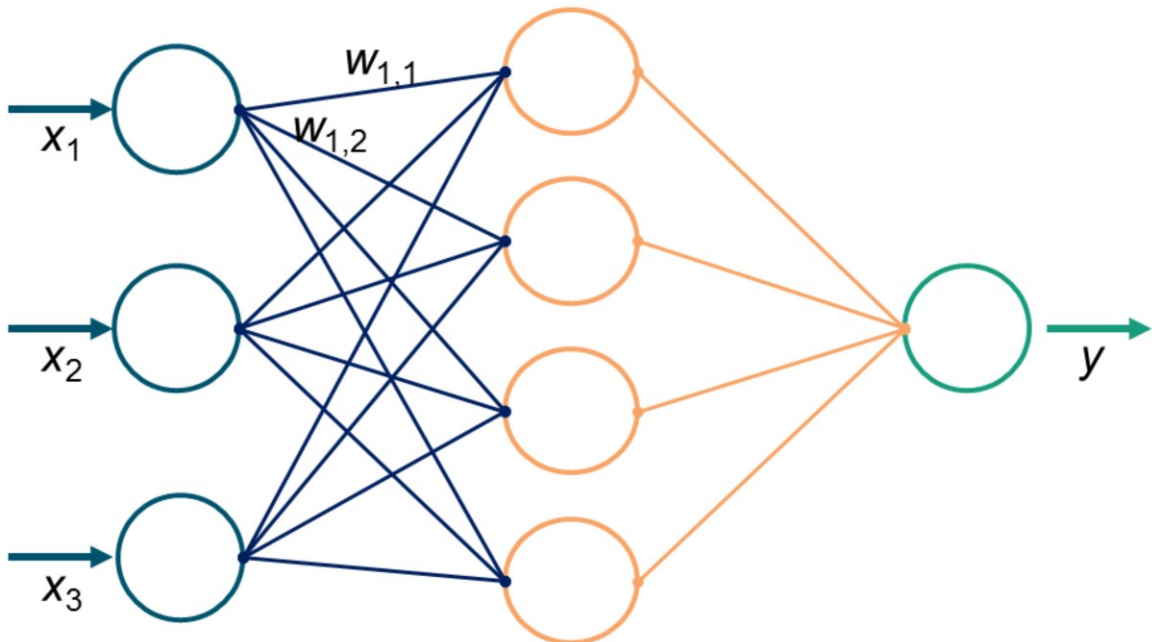
© Fraunhofer

 **Fraunhofer** 
IVV Dresden

Große Beispieldatensätze im Internet verfügbar

- regelmäßige Wettbewerbe unter Informatikern
- **MNIST**
 - Kleiner Datensatz für Handschrifterkennung, 60.000 Datensätze
 - Fehlerrate bei 0,21% [Vadim 16]
- **ImageNet**
 - >14 Millionen manuell zugeordnete Bilder
 - Regelmäßige Nutzung für Wettbewerbe zur Bilderkennung
 - Ergebnis: Fehlerrate unter 5 % [Ioffe and Szegedy, 2015], besser als der Mensch

Prinzip eines künstlichen Neuronales Netzes

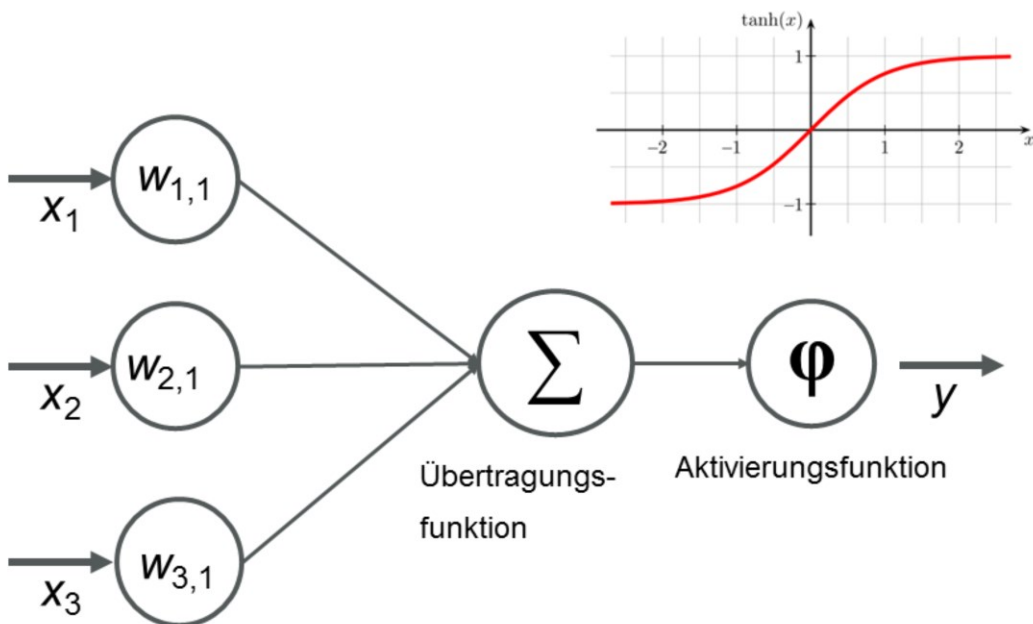


© Fraunhofer

Fraunhofer
IVV Dresden

- Feed Forward Netz mit einer versteckten Ebene (Hidden Layer) mit 4 Neuronen
 - Gewichtete Kanten zwischen den Neuronen
 - Jedes Neuron rechnet mit diesen Eingängen einen individuellen Zustand aus und gibt ihn an das nächste Neuron weiter
 - Durch geschickte Algorithmen anpassen der Wichtungen
 - Was im inneren passiert ist schwer zu verstehen, wird häufig als Blackbox betrachtet [Rothe 17].
- Für Deep Learning noch wesentlich größere Netze
 - Aktuell im Trend: Convolutional Netze mit zusätzlicher Verbindung vom Ausgang zum Eingang eines Neurons

Prinzip eines künstlichen Neurons

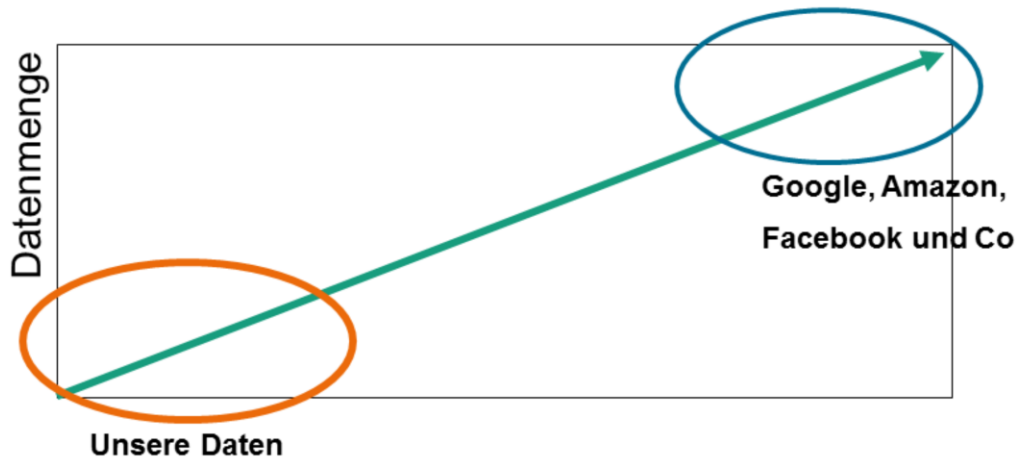


© Fraunhofer

Fraunhofer
IVV Dresden

- Eingänge eines Neurons werden aufsummiert und in Aktivierungsfunktion geschickt
- Aktivierungsfunktion häufig \tanh und andere Funktionen mit einem Ausschlag v.a. im Bereich $[-1;1]$

Zur Menge der Daten



© Fraunhofer

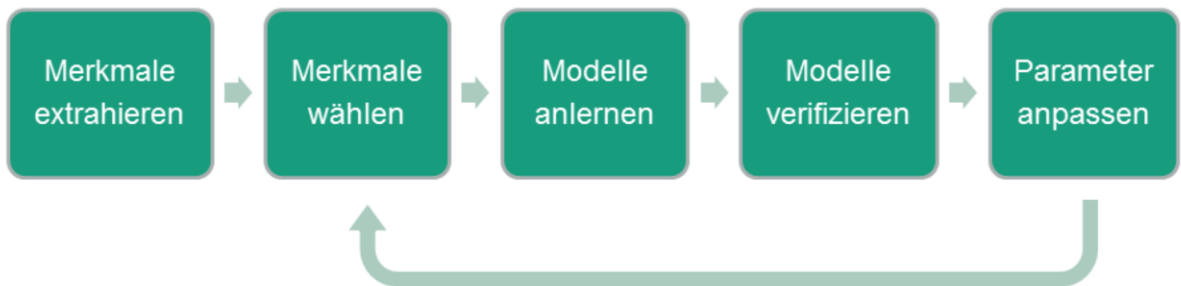
 **Fraunhofer** 
IVV Dresden

- Problem in Forschung: Nur relativ wenig Daten verfügbar
- Analog in vielen Maschinen: Viel Qualitätsprodukt, wenig Störungen => Sehr unausgewogene Datensätze, wenig Daten zum Lernen von Störungen

Vorgehen ohne Deep Learning

- Menschliches Wissen über Datensatz erforderlich:

Welche Daten könnten für eine Entscheidungsfindung nützlich sein?

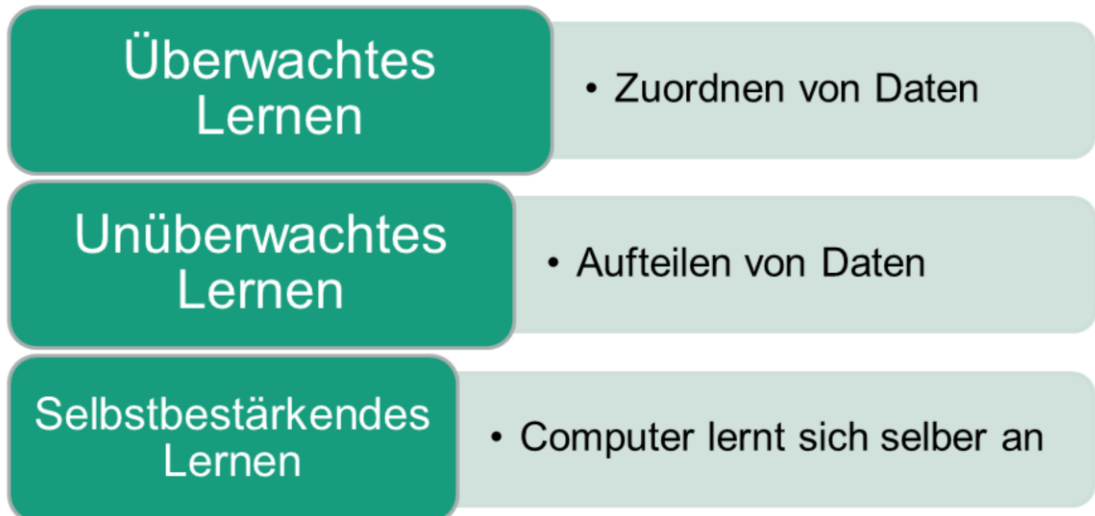


© Fraunhofer

 **Fraunhofer** 
IVV Dresden

- Am Anfang steht der Mensch, Extraktion der Merkmale aufwendige manuelle Tätigkeit [Domingos 12]
 - Welche Daten sind verfügbar?
 - Was kann ich daraus evtl. ableiten?
 - Welche Daten sind sinnvoll?
- Mit diesen Daten können Modelle angelernet und anschließend verifiziert werden
- Typisch: Iterativer Prozess mit Anpassung der Parameter und häufig auch veränderter Merkmalsauswahl

Überwachtes Lernen oder unüberwacht?



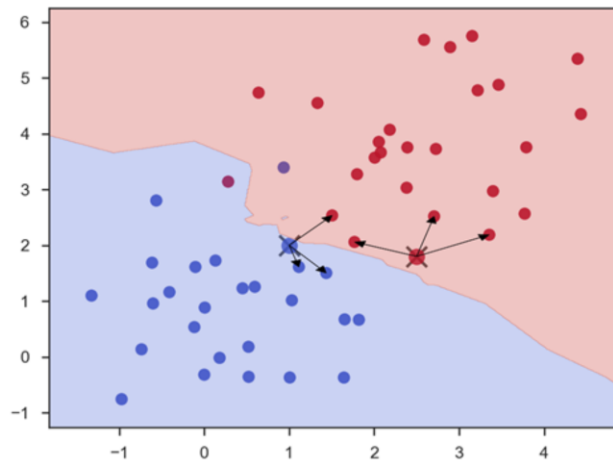
© Fraunhofer

 **Fraunhofer** 
IVV Dresden

- Überwachtes Lernen
 - Ein Trainingsdatensatz mit bekannten Ergebnisse existiert
- Unüberwachtes Lernen
 - Die Daten werden automatisch in Klassen aufgeteilt.
 - Auch im Bereich Anomalie-Erkennung
 - Überwacht: Bsp Amazon: Wer gerne bestimmte Musik kauft wird auch ein Interesse an schwarzer Kleidung haben. Name der Gruppe egal
- Selbstbestärkendes Lernen
 - Der Computer schafft sich selber Trainingsdaten und lernt an diesen
 - z.B. Alpha Go: Wenn das Spiel gewonnen wurde als positiv klassifizieren

***k* Nearest Neighbour**

- Zugehörigkeit über Distanz von Punkten zu den *k* nächsten Nachbarn, z.B. nach Euklid $d = \sqrt{x^2 + y^2}$



© Fraunhofer

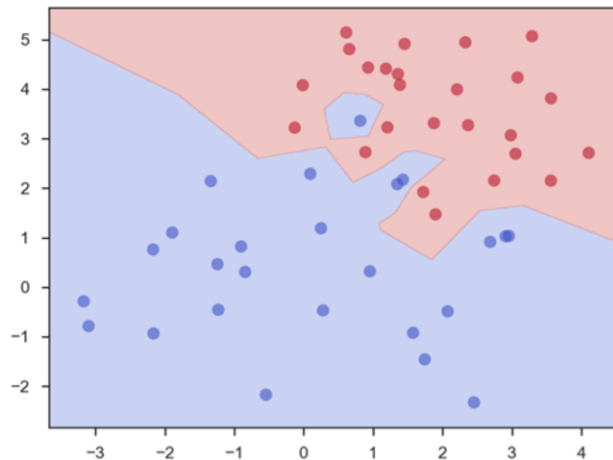
Fraunhofer 
IVV Dresden

- Zu jedem Punkt werden die *k* (hier *k*=3) nächsten Nachbarn gesucht und damit die Entscheidungsgrenze festgelegt
- Darstellung im Zweidimensionalen Raum gut möglich. Die Mathematik ändert sich in höher-dimensionalen Räumen nicht, nur die Vorstellbarkeit
- Einsatz trotz Einfachheit nach wie vor, häufig mit anderen Distanz-Metriken (Manhattan, Minkowski, Hamming, ...)
 - Auch eigene Distanz-Funktionen denkbar, damit gute Anpassung an eigene Problemstellung möglich

Verringern der Fehlklassifizierung

- Der Fehler wird kleiner, aber passt das Modell auch noch für neue Daten?

1 Nachbarn - 0 fehlklassifiziert



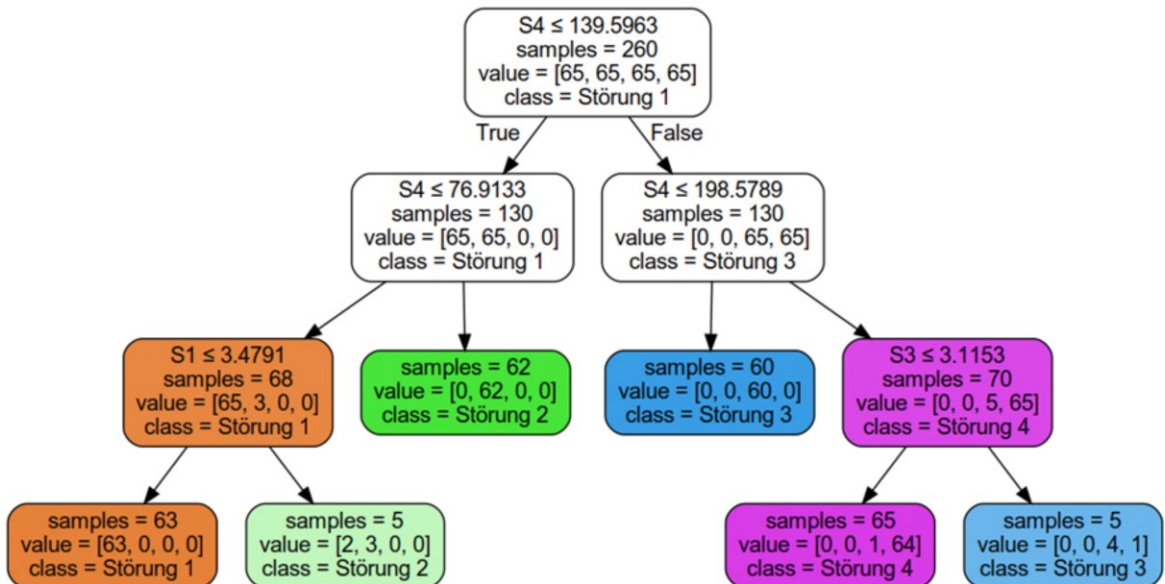
© Fraunhofer

 **Fraunhofer** 
IVV Dresden

- Durch Anpassen von k kann die Generalisierung festgelegt werden.
 - Mehr Nachbarn: Allgemeineres Modell
 - Weniger Nachbarn: Modell lernt immer spezifischer den Datensatz => Überanpassung

Entscheidungsbaum

- Ausgangslage: 4 Sensoren, 4 Störungen

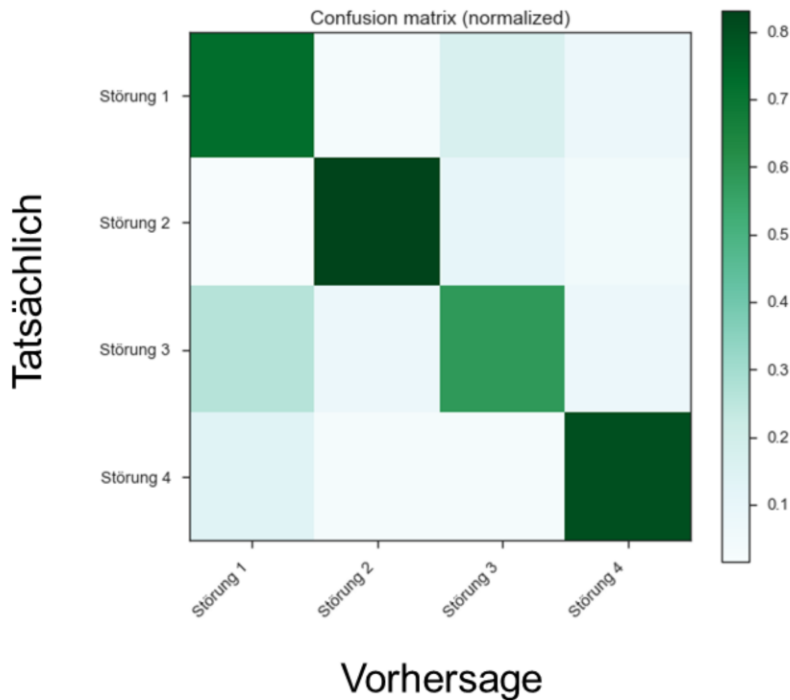


© Fraunhofer

Fraunhofer
IVV Dresden

- Algorithmus findet Entscheidungskriterien nach überwachtem Lernen
- Wenn Wert Sensor S4 < 76.6 dann Störung 1
- Einsatz eines einzelnen Baums selten, aber viele Modelle nutzen eine Vielzahl kleiner Bäume (Random Forest, Gradient Boosted Decision Trees)
- Visualisierung bei mehr als 2 Dimensionen schwierig
 - Berechnen von Kennwerten wie Genauigkeit: Wieviel Prozent wird mit dem Modell passend zugeordnet. Hier 95%
- Überanpassung: Bei so einem kleinen Baum unwahrscheinlich
- Möglichkeit zur Erweiterung: Viele kleine Bäume ergeben einen Wald, z.B. Random Forest [Breiman 2001]

Konfusionsmatrix für den Entscheidungsbaum



© Fraunhofer

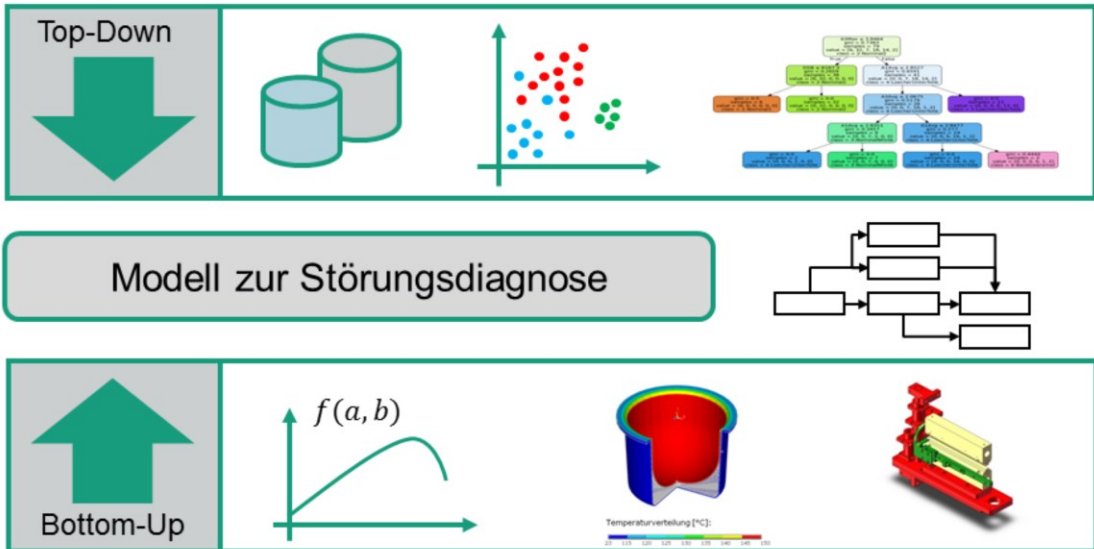
Fraunhofer
IVV Dresden

- Auf der linken Achse tatsächlich eingetragene Störung aus dem Datensatz
- Auf der unteren Achse Vorhersage durch das Modell
- Im Idealfall: Nur Diagonale Besetzt

- Eigentlich fast zu gut
 - Nach Prüfung wird festgestellt: S4 zeichnet den Durchmesser der Folienrolle auf.
 - Bei sequentieller Fehlerprovokation ist ein Fehler leicht über die abnehmende Dicke der Folienrolle erkennbar. Aber nur im vorliegenden Fall => Keine Generalisierung mit diesem Modell
 - Wäre bei großen Datenmengen nicht passiert. Hier sind die Daten nicht unabhängig => Scheinkorrelation

- Vorgehen: S4 ignorieren, der sollte auch später nie einen Einfluss haben

Notwendigkeit hybrider Modelle



© Fraunhofer

14

Fraunhofer IVV Dresden

- Benötigt wird ein Modell zur Störungsdiagnose
- Die ausschließliche Nutzung stochastischer Modelle (TOP-Down, bspw. mittels maschinellem Lernen) erfordert große Datenmengen → Die notwendige Wiederholung einzelner Störungen muss jedoch möglichst gering sein
- Ziel ist daher die Kombination mit numerischen und empirischen Modellen zum Prozess
 - Dadurch soll eine präzise Modellbildung auch mit wenigen Daten möglich sein → das System lernt deutlich schneller

Ihr Ansprechpartner

Fraunhofer IVV,
Außenstelle für Verarbeitungsmaschinen und Verpackungstechnik
Heidelberger Str. 20, 01189 Dresden

Tilman Klaeger
Tel.: 0351 / 436 14-33
tilman.klaeger@ivv-dresden.fraunhofer.de



© Fraunhofer



Quellen:

Breiman, L. (2001): Random Forests. Machine Learning. Vol. 45, Nr. 1, S. 5–32.

Domingos, P. (2012): A few useful things to know about machine learning. Communications of the ACM. Vol. 55, Nr. 10, S. 78–87.

Hecker, D. u. a. (2017): Zukunftsmarkt Künstliche Intelligenz – Potenziale und Anwendungen. Fraunhofer-Allianz Big Data. St. Augustin.

Ioffe, S., Szegedy, C. (2015): Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. Proceedings of the 32nd International Conference on Machine Learning (ICML-15). S. 448–456.

Klaeger, T., Schult, A., Majschak, J.-P. (2017): Lernfähige Bedienerassistenz für Verarbeitungsmaschinen. Industrie 4.0 Management. Vol. 33, S. 25–28.

Rothe, R. (2017): Applying Deep Learning to Real-world Problems, verfügbar unter <http://www.kdnuggets.com/2017/06/applying-deep-learning-real-world-problems.html>, abgerufen 24.07.2017.

Vadim, R. (2016): Training Data Expansion and Boosting of Convolutional Neural Networks for Reducing the MNIST Dataset Error Rate. Naukovi Visti NTUU KPI. Nr. 6, S. 29–34.