

Prof. Dr.-Ing.habil.

Dirk Joachim Lehmann

Lehrstuhl für Data Science in IoT

Fakultät für Informatik

Tel. +49 (0) 5331 939 32800

Fax +49 (0) 5331 939 32802

di.lehmann@ostfalia.de

Interaktives Visuelles Machine Learning

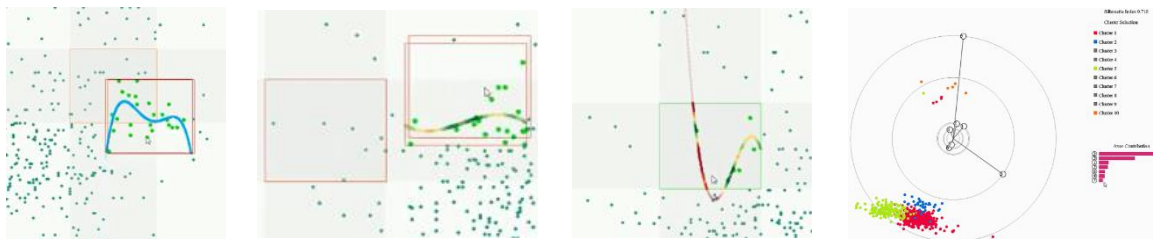
Hintergrund

Im Machine Learning kennen wir konkrete - regelmäßig wiederkehrende - Datenanalyseaufgaben. Hierunter fallen beispielsweise die Erstellung von [Regressionsmodellen](#), [Clustermodellen](#) (inkl. Ausreißeridentifikation), [Klassifikationsmodellen](#), [Assoziationsmodellen](#), oder Aufgaben der geeigneten [Merkmalsselektion](#) (Feature Engineering). Ein Ziel im Machine Learning ist es, solche Modelle – soweit möglich – automatisiert aus vorliegenden Datensätze zu generieren.

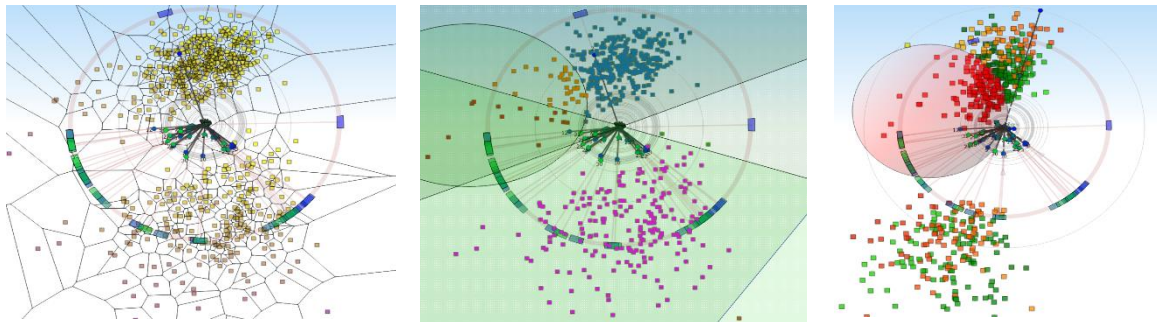
Es ist weiterhin Forschungsgegenstand, inwieweit das vollautomatische Erstellen solcher Modelle aus den Daten heraus insofern immer möglich ist, dass die entstehenden Modelle für die Domäne, welche den Daten zugrunde liegt, repräsentativ sind. Es mehren sich Hinweise, dass dies nicht notwendigerweise der Fall sein kann [1,2].

Eine Idee ist es daher ebenfalls den Nutzer/den Menschen in den Generierungsprozess für Modelle mit einzubinden und somit ein semi-automatisches Machine Learning zu unterstützen, bei welchem zusätzliches externes Domänenwissen mit in der Modellerstellungsprozess eingeht.

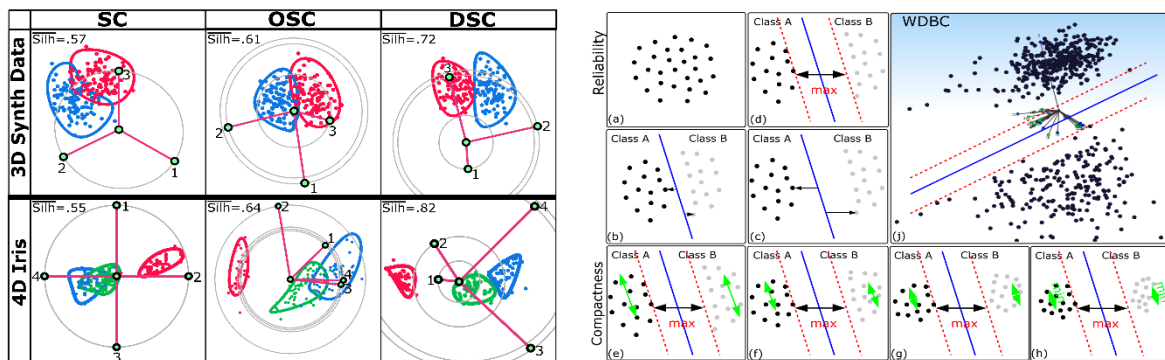
Letztendlich ist damit ein Prozess verbunden, bei welchem die Daten und Modelle dem Nutzer visuell repräsentiert werden müssen und es dem Nutzer möglich sein muss, relevante Eingaben - innerhalb dieses semi-automatischen Modellerzugesprozesses - interaktiv durchführen zu können, was über die Bezeichnung **Interaktives Visuelles Machine Learning** beschrieben ist.



Interaktive Regressionsmodelle und Merkmalsselektionen im Interaktiven Visuelles Machine Learning



Interaktive Erstellung von Clustermodellen im Interaktiven Visuelles Machine Learning



Interaktive Erstellung von Klassifikationsmodellen im Interaktiven Visuelles Machine Learning

Angebot

Im Rahmen ihres *Praxisprojektes*, *Masterseminars*, *Masterprojektes*, ihrer *Bachelorarbeit*, *Masterarbeit* oder ähnlichen Studienleistungen - wie einem *interdisziplinären Digitalisierungsprojekt* und Ähnliches - können Sie gerne einen Themenschwerpunkt im Rahmen des **Interaktiven Visuelles Machine Learning** bearbeiten. Melden Sie sich einfach bei mir unter: di.lehmann@ostfalia.de

Mit welchem Bereich Sie sich im Detail beschäftigen möchten, stimmen wir gemeinsam gerne ab, in Abhängigkeit Ihrer Interessenlage, dabei können Sie sich aussuchen welches Feld für Sie am besten passt.

Zur Wahl steht, dass interaktive Erstellen von

- [Regressionsmodellen](#)
- [Clustermodellen](#) (inkl. Ausreißeridentifikation)
- [Klassifikationsmodellen](#)
- [Assoziationsmodellen](#), oder Themen der geeigneten
- [Merkmalsselektion](#) (Feature Engineering)

Aufgaben

Je nach Themengebiet und Umfang Ihrer Studienleistung, welche Sie absolvieren möchten

- (a) erstellen sie – oder erweitern sie – ein Softwaresystem um weitere interaktive Features, darunter fallen: Erweiterung von User-Interfaces, Integration weiterer Modelle, Laden und Speichern von Daten und Modellen und Ähnliches.

Seite 3 von 4

- (b) Testen der Qualität der erstellten Modelle an realen Datensätzen konkreter Domänen, wie der Medizin, Klimadaten, Agrar etc.
- (c) Vergleichen sie Ihren Ansatz semi-automatisch Modelle zu generieren mit State-of-the-Art Methoden des Maschine Learning, um Modelle voll-automatisch zu generieren.
- (d) Sie recherchieren den aktuellen Stand der Forschung (Paper, Bücher, Online, Podcasts, etc.).
- (e) Sie erstellen eigenständig eine wiss. schriftliche Arbeit zum Thema.

Vorkenntnisse

Es ist hilfreich – aber keine Voraussetzung – wenn Sie Vorkenntnisse mitbringen in

- Softwareentwicklung
- Datenanalyse + Machine Learning
- Lineare Algebra

Vorarbeiten

Im Rahmen meiner Forschungstätigkeiten und der Arbeit des Lehrstuhls *Data Science in IoT* sind zum Thema bereits eine Anzahl an Vorarbeiten entstanden, auf welche Sie für Ihre Arbeit zurückgreifen können, um Ihre Arbeit – je nach Wunsch - auf diesen aufzubauen:

[Yunhai Wang, Jingting Li, Feiping Nie, Holger Theisel, Minglun Gong and Dirk J. Lehmann: Linear Discriminative Star Coordinates for Exploring Class and Cluster Separation of High-Dimensional Data, Computer Graphics Forum \(EuroVis\), \[5\], 2017](#)

Hier finden Sie die Arbeit:

http://46.38.235.241/webpage/dirkfiles/publications/Lehmann_2017_cgfc.pdf

Hier finden Sie ein Video des Systems auf YouTube:

<https://www.youtube.com/watch?v=Zgs5IHdo4rs>

[M. Chegini, L. Shao, R. Gregor, D. J. Lehmann, K. Andrews and T. Schreck: Interactive Visual Exploration of Local Patterns in Large Scatterplot Spaces, Computer Graphics Forum \(EuroVis\), \[4\], 2018](#)

Hier finden Sie die Arbeit:

http://46.38.235.241/webpage/dirkfiles/publications/Lehmann_2018_EV.pdf

Hier finden Sie ein Video des Systems auf YouTube:

<https://youtu.be/gi4mlkniAzI>

[Lin Shao, Aishwarya Mahajan, Tobias Schreck and Dirk J. Lehmann: Interactive Regression Lens for Exploring Scatter Plots, Computer Graphics Forum \(EuroVis\), \[3\], 2017](#)

Hier finden Sie die Arbeit:

http://46.38.235.241/webpage/dirkfiles/publications/Lehmann_2017_cgfb.pdf

Hier finden Sie ein Video des Systems auf YouTube:

<https://www.youtube.com/watch?v=5Wm75sNGG1I>

[Hannes Seibt: Visual Analytics Framework for Training and Evaluation of Online Random Forests for Tissue Classification in Large Histopathological Slides, Masterarbeit](#)

Hier finden Sie die Arbeit:

http://46.38.235.241/webpage/dirkfiles/betreutearbeiten/Seibt_MA_2017.pdf

Hier finden Sie ein Video des Systems auf YouTube:

<https://www.youtube.com/watch?v=5e70vdhsh-M>

Literatur & Referenzen

- [1] Jon Kleinberg, An Impossibility Theorem for Clustering, 2002, 446--453, MIT Press
- [2] B. Karer, H. Hagen, D. J. Lehmann, Insight Beyond Numbers: The Impact of Qualitative Factors on Visual Data Analysis, IEEE Transactions on Visualization and Computer Graphics (Proc. IEEE VAST), 2020
- [3] Lin Shao, Aishwarya Mahajan, Tobias Schreck and Dirk J. Lehmann: Interactive Regression Lens for Exploring Scatter Plots, Computer Graphics Forum (EuroVis), 2017
- [4] M. Chegini, L. Shao, R. Gregor, D. J. Lehmann, K. Andrews and T. Schreck: Interactive Visual Exploration of Local Patterns in Large Scatterplot Spaces, Computer Graphics Forum (EuroVis), 2018
- [5] Yunhai Wang, Jingting Li, Feiping Nie, Holger Theisel, Minglun Gong and Dirk J. Lehmann: Linear Discriminative Star Coordinates for Exploring Class and Cluster Separation of High-Dimensional Data, Computer Graphics Forum (EuroVis), 2017

Bei Interesse melden Sie sich bitte unter:

Prof. Dr.-Ing.habil.
Dirk Joachim Lehmann
Lehrstuhl für Data Science in IoT
Fakultät für Informatik
di.lehmann@ostfalia.de