

Themenliste LS Wirtschaftsinformatik SoSe 2022

1. Individuelles Thema.....	2
2. Maschinelles Lernen / Offener Block	3
3. Ein Vergleich von Random Forest und Gradient Boosting als Regressionsverfahren	4
4. Random Forest Hyperparameter Optimization Techniques.....	5
5. Ein Vergleich von Genetischer Programmierung und Entscheidungsbäumen als Regressionsverfahren	6
6. Analyse des Einflusses von Selektionsverfahren auf das Generalisierungsverhalten in Genetischer Programmierung.....	7
7. Autoencoder als probabilistische Modelle in Estimation of Distribution Genetic Programming.....	8
8. Genetic Programming für Arcade Games	9
9. Programmsynthese mit Genetic Programming	10
10. Analyse des Einflusses von Selektionsverfahren auf das Generalisierungsverhalten in Genetischer Programmierung.....	11
11. Anomalie-Erkennung durch nicht überwacht maschinelles Lernen	12
12. Prognose der Überlebensfähigkeit von Krebspatienten	13
13. Deep Reinforcement Learning	14
14. Natural Language Processing using Deep Learning	15
15. Deep Feed Forward Neural Networks.....	16
16. Interaktion in Social live streaming services (SLSS).....	17
17. Geschäftsmodelle in Streaming und E-Sport	18
18. Explainable AI.....	19
19. Data Analytics anhand eines E-Learning Datensatzes	20
20. Techno-Eustress	21
21. Datensatzanalyse einer Online-Mediathek mit Empfehlungssystem	22

1. Individuelles Thema

Sofern Sie einen eigenen Themenwunsch haben (bspw. eine Arbeit mit betrieblichem Kontext), können Sie bei der Themenwahl als Erstpräferenz die (1) angeben. Voraussetzung für die Vergabe ist, dass ein Mitarbeiter des Lehrstuhls die Annahme des Themenvorschlags befürwortet und sich zur Betreuung bereiterklärt. Es wird daher vorab eine eigenständige Kontaktaufnahme mit einem in Frage kommenden Mitarbeiter empfohlen. Sofern der Mitarbeiter die Bearbeitung des Themas befürwortet, wird er dies an den Koordinator des Seminars weitermelden. Falls zum Zeitpunkt der Themenvergabe keine Befürwortung durch einen Lehrstuhlmitarbeiter vorliegt, wird automatisch das Thema der Zweit- oder Drittpräferenz vergeben.

- Um eine angemessene fachliche Betreuung sicherstellen zu können, sollte das Wunschthema möglichst Bezugspunkte zur Wirtschaftsinformatik haben und/oder im Zusammenhang mit den Forschungsschwerpunkten des potentiellen Betreuers stehen.
- Für betriebliche Arbeiten gilt: Ergebnisse sollten über den unternehmerischen Zusammenhang hinaus generalisierbar sein. Bei „Projektarbeiten“ ist dies häufig ein kritischer Punkt.

WICHTIG: Melden Sie sich bitte vor der Wahl des Themas bei den Betreuern, um das individuelle Thema vorab festzulegen!!!

2. Maschinelles Lernen / Offener Block

(Bachelorseminar/-arbeit, Masterseminar, Masterarbeit)

Betreuung: David Wittenberg, Martin Briesch, Philipp Röchner, Louisa Schwarz

In diesem offenen Block können eigene Themen für Seminar-, Bachelor- und Masterarbeiten aus dem Bereich des maschinellen Lernens (Machine Learning) oder der künstlichen Intelligenz (Artificial Intelligence) vorgeschlagen werden.

Beispielsweise können Künstliche Neuronale Netze (KNNs) verwendet werden. Diese finden derzeit vielfältige Anwendung in den verschiedensten Gebieten:

- Bilderkennung (z.B. <http://benchmark.ini.rub.de/?section=gtsrb> oder <https://github.com/udacity/self-driving-car-sim>),
- Generieren von Texten (z.B. <http://karpathy.github.io/2015/05/21/rnn-effectiveness>),
- Komponieren von Musik,
- gesprochene Worte in Text umwandeln (und umgekehrt).

Dafür kommen, je nach Aufgabenstellung, unterschiedliche Architekturen zur Anwendung, darunter Convolutional Neural Networks (CNNs), Long short-term memory networks (LSTMs) oder die Transformer Architektur.

Für diese Arbeit soll ein praktisches, gemeinsam mit dem Betreuer zu definierendes Projekt umgesetzt werden. Sie können dabei bestehende Software-Projekte verwenden und ausprobieren sowie ergründen, wie diese funktionieren. Ebenso können Sie aber auch von Grund auf ein eigenes Projekt implementieren.

Erste Programmier-Erfahrung ist hilfreich. Das genaue Thema sollte frühzeitig abgesprochen werden. Die Arbeit kann auf Englisch oder Deutsch verfasst werden.

Einstiegsliteratur:

Bishop (2007). Pattern Recognition and Machine Learning. Springer.

Goodfellow, Bengio, Courville (2016). Deep Learning. MIT Press.

Murphy (2010) Machine Learning - A probabilistic Approach. MIT Press.

3. Ein Vergleich von Random Forest und Gradient Boosting als Regressionsverfahren

(Bachelorseminar/-arbeit, Masterseminar, Masterarbeit)

Betreuung: David Wittenberg

Random Forest und Gradient Boosting sind Ensemble-Methoden, die sich zu einem fundamentalen Werkzeug für die Lösung von Prognoseproblemen entwickelt haben. Beide Methoden basieren auf Entscheidungsbäumen. Im Fall von Regressionen spricht man auch von Regressionsbäumen. Während Random Forest als Weiterentwicklung des Bagging Algorithmus typischerweise viele, unabhängige Entscheidungsbäume miteinander kombiniert, zeichnet sich Gradient Boosting durch die iterative Entwicklung von Entscheidungsbäumen aus.

Ziel dieser Arbeit ist die theoretische Darstellung der Ensemble-Methoden Random Forest und Gradient Boosting und die anschließende praktische Implementierung und Gegenüberstellung beider Methoden (bspw.) anhand eines online verfügbaren Datensatzes über Immobilienpreise aus der Stadt Boston, USA (<https://archive.ics.uci.edu/ml/machine-learning-databases/housing/>). Welche Methode erzielt, gemessen am RMSE, leistungsstärkere Prognoseergebnisse? Für die praktische Implementierung empfiehlt sich das statistische Programm R.

Erste Programmier-Erfahrung ist hilfreich. Die Arbeit kann auf Englisch oder Deutsch verfasst werden.

Einstiegsliteratur:

James, G., Witten, D., Hastie, T., Tibshirani, R. (2013). An Introduction to Statistical Learning: with Applications in R. Berlin, Heidelberg: Springer Science & Business Media.

Hastie, T., Tibshirani, R. & Friedman, J. H. (2009). The elements of statistical learning: data mining, inference, and prediction. Second Edition. Berlin, Heidelberg: Springer Science & Business Media.

4. Random Forest Hyperparameter Optimization Techniques

(Bachelorseminar/ -arbeit, Masterseminar, Masterarbeit)

Betreuung: David Wittenberg

Entscheidungsbäume können zur Klassifikation oder Regression genutzt werden und werden u.a. von Banken eingesetzt, um vorherzusagen, ob ein Kunde einen Kredit nicht bedienen kann. Sogenannte Random Forests stellen eine Weiterentwicklung dieser Entscheidungsbäume dar, indem eine bestimmte Menge an Entscheidungsbäumen unabhängig voneinander generiert wird, um die Klassifikationsentscheidung oder Regressionsentscheidung zu treffen.

Es existieren verschiedene Hyperparameter, die das Verhalten des Algorithmus beeinflussen. Diese Hyperparameter können mittels verschiedener Verfahren, darunter Grid Search, Random Search oder Sequential Model Based Optimization (SMBO) optimiert werden, um die Prognosegüte zu erhöhen. Es ist zu erwarten, dass SMBO bei gleicher Anzahl an Hyperparameterevaluationen die besten Ergebnisse erzielt. Die Berechnung bei SMBO kostet allerdings Rechenzeit. Das Ziel der Arbeit ist die Gegenüberstellung der Methoden Random Search und SMBO am Beispiel eines Random Forest, wobei neben der Anzahl der Hyperparameterevaluationen die Rechenzeit berücksichtigt werden soll.

Grundlegende Programmierkenntnisse sind hilfreich. Die Arbeit kann auf Englisch oder Deutsch verfasst werden.

Einstiegsliteratur:

Breiman, Leo (2001). Random Forests. Machine Learning, Vol. 45. 5-32.

Hutter F, Hoos HH, Leyton-Brown K. (2011). Sequential model-based optimization for general algorithm configuration. 5th international conference on Learning and intelligent optimization. LION, Vol. 6683 of lecture notes in computer science. 507–523.

5. Ein Vergleich von Genetischer Programmierung und Entscheidungsbäumen als Regressionsverfahren

(Bachelorseminar/ -arbeit, Masterseminar, Masterarbeit)

Betreuung: David Wittenberg

Genetische Programmierung (GP) hat zum Ziel automatisch Computerprogramme oder mathematische Ausdrücke zur Lösung eines definierten Problems zu erzeugen. Die Idee ist es eine Menge aus zufällig generierten Lösungen über mehrere Generationen durch Selektion und Variationsoperatoren zu evolvieren und dadurch eine gute Lösung für das gegebene Problem zu identifizieren. Da mit GP Lösungen variabler Länge generiert werden können, ist die Symbolische Regression ein klassisches Anwendungsfeld.

Ein Vorteil von GP gegenüber state-of-the-art Regressionsverfahren wie Random Forest oder Gradient Boosting ist, dass die finale Lösung interpretierbar ist, ähnlich wie die Lösungen von einfachen Entscheidungsbäumen. In dieser Arbeit soll daher der Fokus auf der Interpretierbarkeit von Lösungen am Beispiel eines Standard Regressionsdatensatzes (z.B. Airfoil self-noise, <https://archive.ics.uci.edu/ml/datasets/airfoil+self-noise>) gelegt werden. Dabei sollen Lösungen der GP mit einfachen Entscheidungsbäumen verglichen werden, wobei der Fokus sowohl auf der Prognosegüte, als auch auf die Interpretierbarkeit liegt.

Grundlegende Programmierkenntnisse sind hilfreich. Die Arbeit kann auf Englisch oder Deutsch verfasst werden.

Einstiegsliteratur:

Goodfellow, Bengio, Courville (2016). Deep Learning. MIT Press.

Kingma & Welling (2014). Auto-Encoding Variational Bayes. In 2nd International Conference on Learning Representations (ICLR) 2013.

I. Higgins, L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick, S. Mohamed, and A. Lerchner. beta-VAE: Learning basic visual concepts with a constrained variational framework. In Proc. ICLR, 2017.

6. Analyse des Einflusses von Selektionsverfahren auf das Generalisierungsverhalten in Genetischer Programmierung

(Bachelorseminar/-arbeit, Masterseminar, Masterarbeit)

Betreuung: David Wittenberg, Dominik Sobania

Genetische Programmierung (GP) hat zum Ziel automatisch Computerprogramme oder mathematische Ausdrücke zur Lösung eines definierten Problems zu erzeugen. Die Idee ist es eine Menge aus zufällig generierten Lösungen über mehrere Generationen durch Selektion und Variationsoperatoren zu evolvieren.

Das angewendete Selektionsverfahren hat dabei einen großen Einfluss auf die Suche. Bekannte Verfahren sind Tournament-Selection und Lexicase Selection, wobei letztere dafür bekannt ist stärker den Suchraum zu explorieren und oftmals "bessere" Lösungen zu finden. In der Literatur wird allerdings häufig nur die Leistung auf Trainingsdaten reported und das Generalisierungsverhalten der gefundenen Lösungen wird meist ignoriert.

Ziel dieser Arbeit ist es deshalb das Generalisierungsverhalten von Tournament Selection und Lexicase Selection für Standardprobleme der Symbolischen Regression zu analysieren. Als Beispiel kann bpsw. der Energy Heating Datensatz betrachtet werden (<https://archive.ics.uci.edu/ml/datasets/energy+efficiency>).

Erste Programmierkenntnisse sind hilfreich. Die Arbeit kann auf englisch oder deutsch verfasst werden.

Einstiegsliteratur:

McPhee, N. F., Poli, R., & Langdon, W. B. (2008). Field guide to genetic programming.

William La Cava, Lee Spector, and Kourosh Danai. 2016. Epsilon-Lexicase Selection for Regression. In Proceedings of the Genetic and Evolutionary Computation Conference. 741–748.

7. Autoencoder als probabilistische Modelle in Estimation of Distribution Genetic Programming

(Masterseminar, Masterarbeit)

Betreuung: David Wittenberg

Genetische Programmierung (GP) hat zum Ziel automatisch Computerprogramme oder mathematische Ausdrücke zur Lösung eines definierten Problems zu erzeugen. Die Idee ist es eine Menge aus zufällig generierten Lösungen über mehrere Generationen durch Selektion und Variationsoperatoren zu evolvieren. Klassisch werden die Variationsoperatoren Mutation und Rekombination eingesetzt.

Estimation of Distribution Genetic Programming (EDA-GP) ersetzt die Variationsoperatoren durch ein probabilistisches Modell. Die Idee ist es im ersten Schritt die Struktur von vielversprechenden Lösungen zu erlernen und im zweiten Schritt die erlernte Struktur auf neue Lösungen zu übertragen.

Unser Lehrstuhl hat im Jahr 2020 einen ersten Prototypen veröffentlicht (siehe Einstiegsliteratur), der weiterentwickelt werden kann. Mögliche Forschungsfelder sind:

- Vergleich der Transformer-Architektur zur bestehenden LSTM Architektur,
- Analyse von Variational Encodern als generative Modelle,
- Analyse des Einflusses von Denoising auf die Suche.

Programmierkenntnisse sind hierfür erforderlich. Die Arbeit kann auf englisch oder deutsch verfasst werden. Ein Framework wird zur Verfügung gestellt. Zur genauen Themenfindung bitte Kontakt mit dem Betreuer aufnehmen.

Einstiegsliteratur:

McPhee, N. F., Poli, R., & Langdon, W. B. (2008). Field guide to genetic programming.

Kangil Kim, Yin Shan, Xuan Hoai Nguyen, and R. I. McKay. 2014. Probabilistic model building in genetic programming: a critical review. Genetic Programming and Evolvable Machines.15, 2 (2014), 115–167.

David Wittenberg, Franz Rothlauf, and Dirk Schweim. 2020. DAE-GP: denoising autoencoder LSTM networks as probabilistic models in estimation of distribution genetic programming. In Proceedings of the 2020 Genetic and Evolutionary Computation Conference. 1037–1045.

8. Genetic Programming für Arcade Games

(Bachelorseminar/-arbeit, Masterseminar, Masterarbeit)

Betreuung: Dominik Sobania

Mit Genetic Programming ist es möglich, automatisiert kleine Computerprogramme oder mathematische Ausdrücke zur Lösung eines definierten Problems zu generieren. So kann man auch Programme generieren, die z.B. alte Atari-Games spielen. Ziel dieser Arbeit ist es - basierend auf dem aktuellen Stand der Forschung - einen Genetic Programming-Ansatz zu implementieren, der für ein konkretes Spiel funktioniert.

Programmierkenntnisse sind hierfür erforderlich. Die Wahl der Programmiersprache bleibt den Studierenden überlassen.

Einstiegsliteratur:

McPhee, N. F., Poli, R., & Langdon, W. B. (2008). Field guide to genetic programming.

Langdon, W. B., & Poll, R. (2005, September). Evolutionary solo pong players. In Evolutionary Computation, 2005. The 2005 IEEE Congress on (Vol. 3, pp. 2621-2628). IEEE.

9. Programmsynthese mit Genetic Programming

(Bachelorseminar/ -arbeit, Masterseminar)

Betreuung: Dominik Sobania

Mit Genetic Programming können kleinere Computerprogramme für eine vordefinierte Aufgabe generiert werden. Meist wird hierzu eine sehr problemspezifische Syntax verwendet, möglich ist jedoch auch die Generierung von Programmen in höheren Programmiersprachen wie z.B. Python, C/C++ oder Java.

Praktische Ausrichtung:

Ziel dieser Arbeit ist Implementierung und Vergleich von verschiedenen Programmsyntheseverfahren basierend auf Genetic Programming. Für den Vergleich kann die Benchmark-Suite von Helmuth und Spector (siehe Einstiegsliteratur) verwendet werden.

Programmierkenntnisse sind für diese Arbeit erforderlich. Die Wahl der Programmiersprache bleibt den Studierenden überlassen.

Einstiegsliteratur:

McPhee, N. F., Poli, R., & Langdon, W. B. (2008). Field guide to genetic programming.

Helmuth, T., & Spector, L. (2015, July). General program synthesis benchmark suite. In Proceedings of the 2015 Annual Conference on Genetic and Evolutionary Computation (pp. 1039-1046). ACM.

10. Analyse des Einflusses von Selektionsverfahren auf das Generalisierungsverhalten in Genetischer Programmierung

(Bachelorseminar/-arbeit, Masterseminar, Masterarbeit)

Betreuung: David Wittenberg, Dominik Sobania

Genetische Programmierung (GP) hat zum Ziel automatisch Computerprogramme oder mathematische Ausdrücke zur Lösung eines definierten Problems zu erzeugen. Die Idee ist es eine Menge aus zufällig generierten Lösungen über mehrere Generationen durch Selektion und Variationsoperatoren zu evolvieren.

Das angewendete Selektionsverfahren hat dabei einen großen Einfluss auf die Suche. Bekannte Verfahren sind Tournament-Selection und Lexicase Selection, wobei letztere dafür bekannt ist stärker den Suchraum zu explorieren und oftmals “bessere” Lösungen zu finden. In der Literatur wird allerdings häufig nur die Leistung auf Trainingsdaten reported und das Generalisierungsverhalten der gefundenen Lösungen wird meist ignoriert.

Ziel dieser Arbeit ist es deshalb das Generalisierungsverhalten von Tournament Selection und Lexicase Selection für Standardprobleme der Symbolischen Regression zu analysieren.

Erste Programmierkenntnisse sind hilfreich. Die Arbeit kann auf Englisch oder Deutsch verfasst werden.

Einstiegsliteratur:

McPhee, N. F., Poli, R., & Langdon, W. B. (2008). Field guide to genetic programming.

William La Cava, Lee Spector, and Kourosh Danai. 2016. Epsilon-Lexicase Selection for Regression. In Proceedings of the Genetic and Evolutionary Computation Conference. 741–748.

11. Anomalie-Erkennung durch nicht überwachtes maschinelles Lernen

(Bachelorseminar/ -arbeit, Masterseminar, Masterarbeit)

Betreuung: Philipp Röchner

Anomalien sind Beobachtungen, die im Vergleich zu den restlichen Daten anders aussehen, selten vorkommen und meist einen anderen Ursprung haben. In der Anwendung können Anomalien Hinweise auf Qualitätsprobleme in der industriellen Produktion, Betrugsversuche in Banken oder Angriffe auf Netzwerke liefern. Daher sind Verfahren zum Auffinden von Anomalien von großem Interesse. Da es viele unterschiedliche Ursachen für Anomalien gibt, unterscheiden diese sich meist deutlich untereinander und lassen sich nicht einheitlich definieren. Außerdem können in laufenden Prozessen immer neue, noch unbekannte, Arten von Anomalien auftreten. Daher sind zur Anomalie Erkennung Verfahren von Interesse, die ohne Vorgabe von Beispiel-Anomalien oder einer expliziten Definition, aus Daten lernen Anomalien zu erkennen. Solche Verfahren gehören zum nichtüberwachten maschinellen Lernen.

Ziel der Seminararbeit ist es unterschiedliche Verfahren zur Anomalie Erkennung durch nichtüberwachtes maschinelles Lernen zu vergleichen. Dies beinhaltet die Auswahl geeigneter Verfahren aus der Literatur, die theoretisch Darstellung und Anwendung dieser auf einen frei verfügbaren, strukturierten Datensatz. Beispiele für solche Verfahren sind isolation Forests, variational Autoencoder und Loda Modelle. Quellen zu den Verfahren und eine Sammlung von möglichen Datensätzen sind weiter unten angegeben.

Für die praktische Implementierung empfiehlt sich R Statistik oder Python. Erste Erfahrung in der Programmierung ist wünschenswert. Die Arbeit kann auf Englisch oder Deutsch verfasst werden.

Einstiegsliteratur:

Pang, Guansong, et al. "Deep learning for anomaly detection: A review." arXiv preprint arXiv:2007.02500 (2020).

Liu, Fei Tony, Kai Ming Ting, and Zhi-Hua Zhou. "Isolation-based anomaly detection." ACM Transactions on Knowledge Discovery from Data (TKDD) 6.1 (2012): 1-39.

Boehmke, Brad, and Brandon M. Greenwell. Hands-on machine learning with R. CRC Press, 2019.

Pevný, Tomáš. "Loda: Lightweight on-line detector of anomalies." Machine Learning 102.2 (2016): 275-304.

Boehmke, Brad, and Brandon M. Greenwell. Hands-on machine learning with R. CRC Press, 2019.

12. Prognose der Überlebensfähigkeit von Krebspatienten

(Bachelorseminar/-arbeit, Masterseminar/-arbeit)

Betreuung: Philipp Röchner, Louisa Schwarz

Die Anwendung von Methoden des maschinellen Lernens im medizinischen Bereich birgt ein hohes Zukunftspotential. Beispielsweise können mithilfe von Algorithmen Tumore auf CT-Bildern selbstständig erkannt werden oder durch die Analyse von Krankheitsverläufen der Ausgang einer Erkrankung vorhergesagt werden. Diese Methoden können als Hilfestellungen für Ärzte dienen und diese entlasten. Die Grundlage für diese Analysen bilden große Datensätze. Beispielsweise werden Daten zum Krankheitsverlauf und zur Therapie von Krebspatienten in Krebsregistern gesammelt und verarbeitet.

Im Rahmen des „Surveillance, Epidemiology, and End Results (SEER) Program“ der USA werden Daten aus verschiedenen Krebsregistern der USA zusammengeführt. Diese beinhalten medizinische Angaben zum Tumor und zur Therapie. Die dort gesammelten Daten sind öffentlich zugänglich und dienten bereits in der Vergangenheit als Grundlage zahlreicher Forschungsarbeiten.

Im Rahmen der Arbeit soll zunächst eine Aufbereitung der SEER-Daten erfolgen. Hierfür sollen Brust- oder Lungenkrebsdaten verwendet werden. Die zu untersuchenden Variablen werden von der Betreuerin definiert.

Im Anschluss daran erfolgt die Anwendung eines Algorithmus mit dem Ziel der Voraussage der Überlebensfähigkeit. Hierfür wird der Datensatz anhand eines Zeitraums von fünf Jahren in zwei Klassen unterteilt: Ist ein Patient innerhalb von fünf Jahren nach dem Zeitpunkt der Diagnose gestorben, so hat dieser die Krankheit nicht überlebt. Das Ziel ist es, den Ausgang der Krankheit mit einer hohen Genauigkeit richtig vorherzusagen.

Als Algorithmen können zwei (oder im Rahmen einer Masterarbeit auch weitere) der folgenden verglichen werden:

- Entscheidungsbäume
- Random Forest
- logistische Regression
- Support Vector Machine
- Künstliche neuronale Netze

Erste Programmiererfahrungen ist wünschenswert. Als Programmiersprache wird Python 3 oder R Statistik empfohlen. Die Arbeit kann auf Englisch oder Deutsch verfasst werden

Einstiegsliteratur:

Delen, D., Walker, G., & Kadam, A. (2005). Predicting breast cancer survivability: a comparison of three data mining methods. *Artificial intelligence in medicine*, 34(2), 113-127.

13. Deep Reinforcement Learning

(Masterseminar, Masterarbeit)

Betreuung: Martin Briesch

Ein Teilbereich des Maschinellen Lernens ist das sogenannte Reinforcement Learning (dt. Bestärkendes Lernen). Dabei agiert ein Agent mit seiner Umgebung und versucht eine nutzenmaximierende Strategie zu erlernen. In der aktuellen Forschung werden dazu Modelle aus dem Deep Learning Bereich verwendet, um Kontrollprobleme zu lösen. So konnten vor allem im Bereich von Spielen neue state-of-the-art Modelle und Agenten entwickelt werden. Prominente Beispiele sind Algorithmen wie AlphaGo/AlphaZero, welche amtierende Weltmeister in den Spielen Go und Schach besiegen können. Auch für Computerspiele wie Arcade Games, Starcraft und Dota gibt es bereits Agenten, die auf oder über menschlichem Niveau spielen. Dieselben Algorithmen können auch in der Robotik verwendet werden, um bspw. eine Roboterhand zu steuern.

Im Zuge der Arbeit soll die Funktionsweise eines Deep Reinforcement Learning Algorithmus anhand der relevanten Literatur erarbeitet werden. Außerdem soll der Algorithmus mit Standardbibliotheken wie bspw. Tensorflow und OpenAI Gym implementiert und für eine Testumgebung trainiert werden. Die Ergebnisse sollen anschließend untersucht werden. Testumgebung und Untersuchung richten sich dabei nach dem Umfang der Arbeit. Als Programmiersprache wird Python3 empfohlen.

Erste Programmiererfahrungen sollten vorhanden sein. Kenntnisse zu Künstlichen Neuronalen Netzwerken sind hilfreich, können sich aber im Laufe der Arbeit angeeignet werden.

Einstiegsliteratur:

Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction. MIT press.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... & Petersen, S. (2015). Human-level control through deep reinforcement learning. nature, 518(7540), 529-533.

14. Natural Language Processing using Deep Learning

(Masterseminar/Masterarbeit)

Betreuung: Martin Briesch

Für die Verarbeitung von Sequenzen wie bspw. Text werden im Bereich des Deep Learnings meistens sogenannte Recurrent Neural Networks (RNN) verwendet. Dabei zeichnen sich RNNs grundsätzlich durch Rückkopplungen aus, die es ermöglichen zeitliche Dynamik zu berücksichtigen und damit Sequenzen als solche zu verarbeiten. Eine prominente Variante dieser RNNs sind Long Short-Term Memory Netzwerke (LSTM). Eine andere Methode sind Transformer Architekturen. Diese verwenden Self-Attention Mechanismen und sind in vielen Anwendungsfällen state-of-the-art. In der Praxis finden solche Modelle vor allem bei Sprache und Text Anwendung, sowohl zur Klassifizierung und Übersetzung als auch für die Generierung von neuem Text.

Im Zuge der Arbeit soll die Funktionsweise einer solchen Architektur anhand der relevanten Literatur erarbeitet werden. Das untersuchte Modell soll außerdem mit Standardbibliotheken wie bspw. Tensorflow implementiert, trainiert und untersucht werden. Das Testproblem ist dabei entsprechend dem Umfang der Arbeit mit dem Betreuer zu definieren. Als Programmiersprache wird Python3 empfohlen.

Erste Programmiererfahrungen sollten vorhanden sein. Kenntnisse zu Künstlichen Neuronalen Netzwerken sind hilfreich, können sich aber im Laufe der Arbeit angeeignet werden.

Einstiegsliteratur:

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems* (pp. 5998-6008).

15. Deep Feed Forward Neural Networks

(Bachelorseminar/-arbeit, Masterseminar, Masterarbeit)

Betreuung: Martin Briesch

Mehrschichtige Künstliche Neuronale Netzwerke, genannt Deep Learning, sind eine Methode aus dem Bereich des Maschinellen Lernens. Solche Netzwerke können Strukturen in komplexen Daten erlernen und so beispielsweise Klassifikations- und Regressionsprobleme lösen, indem sie einem Input einen Output zuordnen. Abhängig vom Fehler dieser Zuordnung werden die Parameter des Modells iterativ angepasst, bis ein lokales Optimum erreicht ist. Besonders für Bildverarbeitung und sequenzielle Daten konnten in den letzten Jahren mithilfe von Deep Learning neue state-of-the-art Modelle entwickelt werden.

Im Zuge der Arbeit soll die Funktionsweise einer einfachen Feed Forward Architektur anhand der relevanten Literatur erarbeitet werden. Das untersuchte Modell soll außerdem mit Standardbibliotheken wie bspw. Tensorflow implementiert, trainiert und untersucht werden. Das Testproblem ist dabei mit dem Betreuer entsprechend dem Umfang der Arbeit zu definieren. Als Programmiersprache wird Python3 empfohlen.

Erste Programmiererfahrungen sind hilfreich, können sich aber im Laufe der Arbeit angeeignet werden.

Einstiegsliteratur:

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.
LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, 521(7553), 436-444.
Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, 61, 85-117.

16. Interaktion in Social live streaming services (SLSS)

(Bachelorseminar/-arbeit, Masterseminar, Masterarbeit)

Betreuung: Marc de Zoeten

Die drei großen Streamingportale Twitch (Amazon), YouTube Live (Google/Alphabet) sowie Facebook Video-Game-Streaming (Facebook) dominieren den durch enormes Wachstum gekennzeichneten Live-Streaming-Markt. Sie bieten gegenüber klassischen (kostenlosen) Video-on-Demand (VoD) Diensten sowie klassischen Medien starke Nutzer ↔ Streamer Live-Interaktionsoptionen. Welche theoretischen Grundlagen existieren zur (Live-)Interaktion im Streamingbereich?

Bachelorarbeit:

Innerhalb der Arbeit sollen zunächst die Konzepte der Co-Experience sowie der parasozialen Interaktion auf Basis der Literatur beschrieben werden. Aus den theoretischen Ausarbeitungen sollen in der Folge Hypothesen zur Wirkung und zum gezielten Einsatz (gezielte Erzeugung) von Co-Experience und parasozialer Interaktion bei SLSS entwickelt werden. Eine empirische Prüfung der erarbeiteten Thesen muss nicht stattfinden.

Masterseminar/Masterarbeit:

Innerhalb der Arbeit sollen zunächst die Konzepte der Co-Experience sowie der parasozialen Interaktion auf Basis der Literatur beschrieben werden. Aus den theoretischen Ausarbeitungen sollen Hypothesen zur Wirkung und zum gezielten Einsatz (gezielte Erzeugung) von Co-Experience und parasozialer Interaktion bei SLSS entwickelt werden. Diese Hypothesen werden in der Folge in mehreren Fallstudien geprüft.

17. Geschäftsmodelle in Streaming und E-Sport

(Bachelorseminar/-arbeit, Masterseminar)

Betreuung: Marc de Zoeten

E-Sport erfreut sich weltweit einer hohen und steigenden Popularität. E-Sport Turniere können sich bezogen auf die Größen Zuschauerzahlen und Preisgeld mit klassischen Turnieren längst messen oder diese sogar übertreffen. Selbst populäre Turniere wie der Superbowl oder Wimbledon fallen hinter Top-E-Sport-Turniere zurück. Auch Live-Streaming erfreut sich auf Seite der Rezipienten großer Beliebtheit (Juni 2020: 1,6 Milliarden Stunden wurden allein auf twitch.tv gestreamt).

Die Grundlage für die Analyse soll das Business Model Canvas von Osterwalder sein. Dieses kann im Rahmen der Arbeit durch andere theoretische Konzepte ergänzt werden (mit entsprechender theoretischer Begründung). Als Untersuchungsgegenstand sollen in der Analyse eigenständig Akteure aus dem Bereich E-Sport oder aus dem Bereich Streaming (oder Akteure, welche in beiden Bereichen aktiv sind) ausgewählt werden. Diese können auf verschiedene Aspekte, z.B. Gemeinsamkeiten und Unterschiede in den Geschäftsmodellen geprüft werden.

18. Explainable AI

(Bachelorseminar/-arbeit, Masterseminar, Masterarbeit)

Betreuung: Marc de Zoeten

Künstliche Intelligenz und insbesondere Machine Learning ermöglichen die Verarbeitung großer Datenmengen und unterstützen Entscheidungen in den unterschiedlichsten Bereichen. Dabei sind viele der eingesetzten Verfahren „Black-Box“-Verfahren, es ist für Menschen bzw. Entscheidungsträger entweder nicht oder kaum nachvollziehbar, wie das System zu einer Entscheidung oder Empfehlung kommt. In den vergangenen Jahren wurden Systeme entwickelt, die Nutzern dabei helfen sollen, die Outputs von entsprechenden Black-Box Systemen nachzuvollziehen. Für Recommender-Systeme liegt hierfür bereits Forschung vor, für andere Arten von Entscheidungsunterstützungssystemen besteht weiterhin hoher Forschungsbedarf.

Innerhalb der Arbeit sollen zunächst theoretische Grundlagen zu Explainable AI aufgearbeitet werden und ein Experiment mit Umfrage zu Explainable AI durchgeführt werden.

Bitte kontaktieren Sie bevor Sie das Thema belegen mdezoet@uni-mainz.de

Einstiegsliteratur:

Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). Association for Computing Machinery, New York, NY, USA, 1135–1144. DOI:<https://doi.org/10.1145/2939672.2939778>

19. Data Analytics anhand eines E-Learning Datensatzes

(Masterseminar, Masterarbeit)

Betreuung: Julia Zielonka

E-Learning wird zunehmend zur Leistungsevaluation eingesetzt. Solche sog. E-Assessments können zum einen dazu dienen, den Leistungsstand nach Ablauf einer Lerneinheit zu bewerten und mit einem vorgegebenen Standard zu vergleichen (ergebnisorientierte Leistungsbewertungen). Zum anderen können E-Assessments aber auch für kontinuierliches Leistungsfeedback zur Verbesserung des Lernprozesses eingesetzt werden. Hierdurch ergeben sich neue Potenziale in Bezug auf die Leistungsbewertung. Wenn die Ergebnisse der Leistungserfassung als Feedback den Lernenden gegeben werden, erhalten diese während des Lernprozesses Informationen über ihre individuelle Leistungsveränderung (prozessorientierte Feedbackmechanismen).

Da Feedbacksysteme basierend auf ergebnisorientierten Leistungsbewertungen eine geringe Dynamik aufgrund der beschränkten Beeinflussbarkeit durch die Lernenden und zusätzlich eine verstärkende Wirkung bereits bestehender Leistungsunterschiede aufweisen, stellt sich die Frage, ob prozessorientierte Feedbackmechanismen nicht grundsätzlich besser in der Lage sind, motivations- und leistungssteigernde Effekte hervorzurufen.

Ziel der Arbeit ist die Darstellung der Effekte von prozessorientierten Feedbackmechanismen in E-Learning Anwendungen hinsichtlich der Motivationssteigerung, insbesondere im Vergleich zu ergebnisorientierten Leistungsbewertungen. Hierzu soll ein bestehender Datensatz analysiert werden.

Programmierkenntnisse und –erfahrungen (zur Aufbereitung und Analyse der Daten) sowie gute Statistikkenntnisse sind für diese Arbeit erforderlich.

Die Arbeit soll bevorzugt auf Englisch verfasst werden (Präsentationen können auch auf Deutsch gehalten werden). Grundlagenliteratur ist selbst zu identifizieren. Bevorzugt ist dieses Thema als Masterarbeit zu vergeben.

20. Techno-Eustress

(Bachelorseminar/-arbeit, Masterseminar)

Betreuung: Julia Zielonka

Sowohl für die Arbeit genutzte als auch im privaten Umfeld eingesetzte Technologien können für den Nutzer Stress erzeugen, sogenannter Technostress. Dieser wird durch starke psychische Belastungen aufgrund der Nutzung von technischen Geräten und Systemen hervorgerufen. Vorrangig ist der Begriff „Stress“ negativ konnotiert. Wissenschaftler konnten jedoch Stress auch eine positive Wirkung nachweisen, welcher in Verbindung mit Technologien „Techno-Eustress“ genannt wird.

In dieser Arbeit soll anhand einer Literaturrecherche der Stand der Forschung im Bereich Techno-Eustress erschlossen werden. Ziel ist die Zusammenfassung, Analyse und Systematisierung vorhandener Literatur.

Alternativ kann das Thema empirisch erforscht werden.

Die Arbeit soll bevorzugt auf Englisch verfasst werden (Präsentationen können auch auf Deutsch gehalten werden).

Einstiegsliteratur:

Tarafdar, M., Cooper, C. L., and Stich, J. (2019). “The technostress trifecta - techno eustress, techno distress and design: Theoretical directions and an agenda for research”, *Information Systems Journal* 29(1), 6–42.

Ayyagari R., Grover, V., and Purvis, R. (2011). Technostress: technological antecedents and implications. *MIS Quarterly* 35 (4), 831–858.

21.Datensatzanalyse einer Online-Mediathek mit Empfehlungssystem

(Masterseminar, Masterarbeit)

Betreuung: Marc de Zoeten

Empfehlungssysteme werden genutzt um Konsumenten Entscheidungen zu erleichtern. Dabei hat die Empfehlung unter Umständen eine intendierte lenkende Wirkung. Streaming Services wie Netflix erzielen sogar einen Großteil Ihrer Streams durch Empfehlungssysteme.

Innerhalb der Arbeit sollen zunächst theoretische Grundlagen zu Online-Empfehlungssystemen aufgearbeitet werden. In der Folge soll ein Datensatz zunächst bereinigt und in der Folge analysiert werden. Einfache deskriptive Auswertungen werden durch weitergehende Analysen ergänzt. Sowohl SPSS als auch STATA können als Analysesoftware verwendet werden. Die Verwendung von R oder Python ist nach vorheriger Rücksprache möglich.

Bitte kontaktieren Sie bevor Sie das Thema belegen mdezoet@uni-mainz.de