Karlsruher Institut für Technologie
Institut für Angewandte Informatik und Formale Beschreibungsverfahren (AIFB)
Anna Nguyen
nguyen@kit.edu

15.04.2019

Thesis

„Explaining Approaches for Neural Networks“

Neuronale Netze sind aus dem Bereich des Maschinellen Lernens nicht mehr wegzudenken. Sie erleben in den letzten Jahren einen erheblichen Aufschwung. In fast allen Bereichen werden Neuronale Netze eingesetzt. Sie sind zwar mathematisch nachvollziehbar, aber das sog. „Blackbox-Problem“ besteht weiterhin. Während sie auch in Bereiche wie Recht, Finanzwesen und Medizin vordringen, wird es immer wichtiger auch zu „verstehen“, wie sie zu diesen Vorhersagen kommen. Dazu gibt es einige Ansätze in der Literatur.

Vor diesem Hintergrund ist es das Ziel dieser ausgeschriebenen Arbeit verschiedene Erklärungsansätze, wie z.B. Kernel SHAP[1], LIME[2] und L2X[3], aus der Literatur anzuwenden.

Die Idee und Schwerpunkt dieser Arbeit ist ein Wrapper (für Text), der die verschiedenen Erklärungsansätze auf ein gegebenes Neuronales Netz ausführt und dessen „Erklärung“ zurückgibt. Darauf aufbauend, sollen weitere Ansätze, evtl. aus den Bereichen Supervised Ensemble Learning oder Semantik, angewendet werden. Diese Arbeit zielt auf Masterstudierende mit sehr guten Programmierkenntnissen.

Mögliche Aufgabenbereiche umfassen (sind aber nicht begrenzt auf)**:**

- Erstellung eines Wrappers, der die verschiedenen Erklärungsansätze ausführt
- Maßzahl zum Vergleich der verschiedenen Ansätze definieren

Das sollten Sie mitbringen:

- Gute Englischkenntnis
- Strukturiertes sowie organisiertes Denken
- Interesse an semantischen Technologien und statistischen Verfahren
- Sehr gute Kenntnisse in Maschinellen Lernverfahren
- Sehr gute Python Kenntnisse

Sie haben Interesse?
Dann schicken Sie eine E-Mail mit Anschreiben, Lebenslauf, und aktuellem Notenauszug.

Kontaktperson:
Anna Nguyen
nguyen@kit.edu
Tel.: 0721/60845780

[1] http://arxiv.org/abs/1705.07874
[2] https://arxiv.org/pdf/1602.04938.pdf
[3] https://arxiv.org/pdf/1802.07814.pdf