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

24.03.2020

Thesis

„Exploring Neural Networks with Activation Atlas“

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. Im Bereich der Bildverarbeitung spielen „Convolutional Neural Networks“ eine große Rolle. Trotz der vielen Erklärungsansätze, die die Features dieser Netze visualisieren[1], sind sie meist für den Menschen unverständlich.

Vor diesem Hintergrund ist es das Ziel dieser ausgeschriebenen Arbeit den Erklärungsansatz mit „Activation Atlas“[2] zu implementieren, semantisch zu labeln und Erklärungen menschenverständlich herauszuziehen. Da die Visualisierungen auf Neuronen-Ebene für Menschen eher unverständlich sind, wäre eine Möglichkeit diese auf Filter-Ebene darzustellen. Dazu können Implementierungen wie Deep Visualization Toolbox[3] oder DeConv[4] benutzt werden.

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

- Anwendung des Ansatzes „Activation Atlas“ auf Filter
- Visualisierungen semantisch annotieren

Das sollten Sie mitbringen:

- Englischkenntnis
- Strukturiertes sowie organisiertes Denken
- Kenntnisse in Maschinellen Lernverfahren
- 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] https://distill.pub/2018/building-blocks/
[2] https://distill.pub/2019/activation-atlas/
[3] https://github.com/yosinski/deep-visualization-toolbox
[4] https://arxiv.org/abs/1311.2901