Karlsruhe Institute of Technology

Institut für Automation und angewandte Informatik

IAI

Bachelorarbeit

Uncertainty in Deep Learning

Beim Lösen von Problemen mittels maschinellem Lernen (ML) ist es oft von Vorteil zu wissen, wie sicher ein Modell in den getroffenen Vorhersagen ist. Viele Algorithmen können diese Sicherheiten bzw. Unsicherheiten mithilfe der Bayesschen Statistik darstellen. Zusätzlich verspricht ein bayessches Framework mehr Robustheit gegen sogenannte „Adversarial Attacks", welche versuchen den Algorithmus zu überlisten.

Deep Learning ist eine Methodik, die sich dank neuartiger Hardware und Verfügbarkeit großer Datensätze für schwierige ML-Probleme eignet. Ein offenes Problem ist jedoch, dass tiefe neuronale Netze schwer interpretierbare Black-Box-Modelle sind, die meist Millionen von Parametern aufweisen. Dies macht eine exakte bayessche Behandlung und damit eine bessere Interpretierbarkeit unmöglich. Mit Hilfe von Variational Inference (VI) wird eine vereinfachte Verteilung approximiert und an die gewünschte Verteilung angenähert. Dies macht eine Untersuchung der Unsicherheiten in den vereinfachten Verteilungen möglich und lässt Rückschlüsse auf die echte Verteilung zu. Auf diese Weise könnten Unsicherheiten in den Ausgaben von neuronalen Netzen quantifiziert und somit das Problem des Overfittings entschärft werden. Da die Netzwerkparameter in einem bayesschen Framework als Zufallsvariablen behandelt werden, müssen die Netzwerkstrukturen umfänglich geändert werden. Die Unsicherheit wird somit durch die Varianz in den Netzwerkparametern repräsentiert und soll bewertet werden.

Ziel dieser Arbeit ist es, ein Framework für bayessche neuronale Netze zu implementieren und diese auf Datensätzen (z.B. CIFAR10 und medizinische Daten) auszuwerten, mit herkömmlichen neuronalen Netzen zu vergleichen und den zusätzlichen Rechenaufwand zu bewerten.

Aufgaben:

- Implementierung eines Frameworks für bayessche neuronale Netze
- Bewertung der Reaktion der Netze unter „Adversarial Attacks"
- Bewertung der Reaktion der Netze auf neue, ungesehene Daten
- Vergleich von bayesschen Netzen mit herkömmlichen neuronalen Netzen.
- Bewertung von Rechenzeit und Implementationsaufwand für weitere Datensätze

Markus Reischl
Arbeitsgruppe ML4HOME
Tel.: 0721 / 608 25749
E-Mail: markus.reischl@kit.edu

Institut für Automation und angewandte Informatik (IAI)
Karlsruher Institut für Technologie, Campus Nord
Hermann-von-Helmholtz-Platz 1
76344 Eggenstein-Leopoldshafen