

Bachelorthesis-Aufgabe

Machine Learning "End-to-End" - Verarbeitungsprozesse im Spannungsfeld zwischen Cloud und intelligenten Devices

ID	IDSA2-1-17
Studierende	Michael Herren
Betreuer	Dr. Andreas Danuser
Experten	Dr. Federico Flueckiger
Aufgabe	Einleitung / Umfeld und Ausgangslage

Mit Hilfe von Neuronalen Netzen ("neural networks"), welche in geeigneter Weise angelernt werden können ("machine learning") lassen sich künstlich intelligente Systeme realisieren, mit deren Hilfe z.B. Bilder analysiert oder Verhaltensmuster erkannt werden können. Vor ca. 30 Jahren hatte diese Technologie eine erste Blütephase; diese wurde unterstützt durch geeignete Technologien, wie z.B. die Transputers-Mikroprozessoren und die Programmiersprache Occam. Dennoch war die Leistungsfähigkeit dieser Systeme limitiert und daher ebenso der Erfolg der Neuronalen Netze damals.

Nun, 30 Jahre später stehen wir wieder in einer Blütephase dieser Technik, angetrieben vom Bedürfnis, Bilder und Situationen analysieren, Muster erkennen und Systeme autonom machen zu können. Diese Blütephase wird - wiederum - ermöglicht durch die Verfügbarkeit von Halbleiter-Bausteinen von höchster Leistungsfähigkeit: ursprünglich für die Bearbeitung von Grafik ausgelegte Prozessoren (z.B. Nvidia) ermöglichen die effiziente Ausführung von Algorithmen auf Basis von neuronalen Netzwerken. Noch immer sind aber Energie- und Platzbedarf dieser Hardware-Geräte verhältnismässig gross, so dass heute die entsprechende Rechenleistung hauptsächlich auf Server-Computer in der Cloud bereitgestellt wird.

Der Wunsch aber besteht, Algorithmen zur Erkennung von Bildern und Mustern in Echtzeit auch auf kleineren Geräten ("Devices"), wie z.B. SmartPhones, zu ermöglichen, da diese Geräte heute veritable "Sensor-Zentralen" sind und Bilder, Töne, Bewegungen uvm. erfassen können. Der im September 2017 angekündigte Kirin 970 Chip von Huawei, für den Einsatz in kleinen Devices ausgelegt, unterstreicht diese Entwicklung. Der Kirin 970 Chip umfasst neben einer herkömmlichen CPU und GPU auch eine sog. NPU - eine Neural Processing Unit. Immer mehr ist es also möglich, leistungsfähige neuronale Netze teilweise oder ganz auf mobilen Endgeräten ausführen zu können.

So steht folgende **Frage** im Raum: "Inwieweit können und sollen heute kleine Geräte oder SmartPhones die Aufgaben rund um Neuronale Netze / Machine Learning teilweise oder ganz abarbeiten? Was kann auf einem solchen Gerät geleistet werden und was soll / muss in der Cloud stattfinden?"

Aufgabe

In der vorliegenden BSc-These soll diese Frage systematisch und schrittweise analysiert werden. Mit einer geeigneten typischen Anwendung soll die Aufteilung der Aufgaben auf intelligente Geräte und Cloud illustriert werden. (Hinweis: Der im Titel verwendete Ausdruck "Machine Learning" umfasst alle damit verbundenen Aspekte, also sowohl die Verarbeitung von Informationen mittel neuronalen Netzen wie auch das geeignete maschinelle Trainieren dieser Netze.)

Der Diplomand soll grosse Freiheit zur Bearbeitung dieser Aufgabe haben; im Sinne einer möglichen Leitlinie könnte man die Aufgabe in folgende Teile gliedern:

1. Kurze Einführung in das Thema "Neural Networks and Machine Learning"
2. Analyse der durch typ. "Cloud-Plattformen" angebotenen entsprechenden Funktionalitäten (z.B. von Google / Android)
3. Analyse / Recherche betr. Rechenaufwand für die Teilaufgaben "Rechnen" und "Lernen"
4. Analyse / Recherche betr. Leistungsfähigkeit verschiedener HW-Plattformen
5. Analyse / Exploration bez. Frage "Was kann auf dem mobilen Endgerät geleistet werden?"
6. Freie Auswahl und Umsetzung eines geeigneten / illustrativen Beispiels unter Einbezug mobiler Endgeräte und Cloud-basierter Server-Infrastruktur. Dieses

Beispiel soll als Basis zur Beantwortung der oben genannten Teilaufgaben dienen.