

Nederlandse doorbraak in energiezuinige kunstmatige intelligentie

Dankzij een wiskundige doorbraak kunnen toepassingen van kunstmatige intelligentie, zoals spraakherkenning, gebarenherkenning en de classificatie van elektrocardiogrammen (ECG), een factor honderd tot duizend energiezuiniger worden. Dat betekent dat het mogelijk wordt om kunstmatige intelligentie in chips te stoppen, waardoor toepassingen bijvoorbeeld op een smartwatch of een smartphone kunnen draaien terwijl dat tot op heden in de cloud moest gebeuren.

Door kunstmatige intelligentie op een lokaal apparaat te draaien, worden de toepassingen bovendien robuuster en privacy-vriendelijker. Robuuster, omdat er geen netwerkverbinding met de cloud meer nodig is. En privacy-vriendelijker, omdat data lokaal kunnen blijven. De wiskundige doorbraak is bereikt door onderzoekers van het Centrum Wiskunde & Informatica (CWI) en het Eindhovense onderzoekscentrum IMEC/Holst Centre en is gepubliceerd in een paper van de International Conference on Neuromorphic Systems. De onderliggende wiskunde is open source beschikbaar.

Nieuw algoritme

Onder leiding van CWI-onderzoeker en UvA-hoogleraar cognitieve neurobiologie Sander Bohté ontwikkelden onderzoekers Bojian Yin en Federico Corradi een nieuw algoritme voor een zogeheten spiking neural network (gepulst neurale netwerk). Zulke netwerken bestaan al langer maar hadden tot nu toe het grote nadeel dat ze wiskundig moeilijk hanteerbaar waren waardoor ze in de praktijk niet konden worden toegepast. Het nieuwe algoritme bevat twee wiskundige doorbraken. De neuronen in het netwerk hoeven veel minder met elkaar te communiceren en daarnaast hoeft elk neuron ook nog eens minder te rekenen.

“In combinatie zorgen deze doorbraken ervoor dat kunstmatige intelligentie meer dan duizend keer energiezuiniger wordt, vergeleken met ouderwetse neurale netwerken en een factor honderd energiezuiniger dan de beste hedendaagse neurale netwerken”, zegt onderzoeksleider Bohté.

Zuinig als het menselijk brein

Bohté haalt zijn inspiratie en motivatie uit het feit dat het menselijke brein informatie verwerkt op een ongelofelijk energiezuinige manier (20 watt). Computers die neurale netwerken nabootsen hebben in de afgelopen jaren weliswaar prachtige toepassingen opgeleverd (beeldherkenning, spraakherkenning, automatisch vertalen, medische diagnoses), maar slurpen veel meer energie dan het menselijk brein.

De spiking neurale netwerken die Bohté en zijn collega's ontwikkelen verschillen van de klassieke neurale netwerken, die tot nu toe in AI-toepassingen zaten. Bohté: “In de klassieke neurale netwerken zijn de signalen continu en wiskundig makkelijk hanteerbaar. Spiking neurale netwerken werken met pulssjes en staan dicht bij de biologie van de hersenen. Dat betekent wel dat signalen discontinu zijn en wiskundig veel lastiger hanteerbaar.”

Nieuw soort chips nodig

Om spiking neurale netwerken efficiënt toepassingen te laten draaien zijn wel eerst aparte chips nodig. Bohté verwacht dat er binnen een jaar prototypen zijn: “Allerlei bedrijven werken daar hard aan, zoals onze projectpartner IMEC/Holst Centre.”

De spiking neurale netwerken van Bohté en zijn collega's kunnen momenteel zo'n duizend neuronen aan. Dat is beduidend minder dan klassieke neurale netwerken op dit moment aankunnen, maar wel genoeg voor velerlei toepassingen van spraakherkenning en ECG-analyse tot het herkennen van gebaren. De volgende uitdaging is dan ook om hun netwerken op te schalen naar 100.000 of een miljoen neuronen. Dat zal de toepassingsmogelijkheden nog verder uitbreiden.

Het onderzoeksproject van Bohté en zijn collega's vindt plaats binnen het NWO Perspectief-programma 'Efficient Deep Learning'.