

de jonge wetenschapper **Hugo Proença**



**‘Computers zijn nog
steeds best dom’**

Supercomputers en zelfs kwantumcomputers: machine learning neemt steeds ongrijpbaarder vormen aan. Zelfs datawetenschappers weten niet precies hoe black boxes tot hun resultaten komen. Om te begrijpen hoe deze modellen beslissingen nemen, duikt **Hugo Proença** (29) in 'interpreteerbare modellen'.

Voorspellen wanneer vluchten vertraging oplopen en daar rekenmodellen voor ontwikkelen: dat is kort gezegd het onderzoeks-onderwerp van Proença. Niet omdat hij zelf zoveel vertragingen meemaakt, maar omdat hij van jong af aan al de wetenschap in wilde. 'Mijn twintig jaar oudere neef was mijn grote voorbeeld', vertelt hij. 'Toen ik op de middelbare school zat, deed hij een PhD in *mechanical engineering*. Als we bijvoorbeeld Formule 1 keken, kon hij precies vertellen wat er zo goed was aan de motor en de wagen. Hij sprak met zoveel passie over zijn vak, dat ik besloot dezelfde opleiding te doen. Alleen koos ik voor robotica als master.' Helaas zal zijn neef Proença's promotie niet meer meemaken. 'Hij is jong overleden, heel verdrietig. Gelukkig weet mijn tante dat hij me geïnspireerd heeft en kan ze daar een beetje troost uit putten.' Hoewel wetenschap hem op het lijf geschreven leek, vindt Proença juist de combinatie met het bedrijfsleven interessant. Zijn huidige onderzoek is onderdeel van een publiek-private samenwerking. 'Het nadeel is dat een bedrijf minder geïnteresseerd is

in hoe je methodes ontwikkelt', vertelt hij, 'zolang ze de techniek maar direct kunnen toepassen. Aan de andere kant is die bedrijfsmatige insteek prettig, want mijn onderzoek leidt tot een eindproduct dat praktisch inzetbaar is. Wetenschap voor de wetenschap is heerlijk, maar ik wil graag een bijdrage leveren aan de samenleving. Door een bedrijf te betrekken bij je onderzoek, blijf je dichtbij de realiteit.'

Digitaal kerkhof

Voordat Proença iets met praktisch nut ontwikkeld heeft, zijn er talloze ideeën vergeefs uitgewerkt. 'Mijn werkproces begint met een probleem waar ik allerlei oplossingen voor bedenkt. Het beste idee werk ik uit en ga ik programmeren. Dan loop ik meestal tegen fouten aan, of blijkt het in de testfase niet te werken. Dan moet ik terug naar het oorspronkelijke idee, om te schaven. Al die documenten met abstract denkwerk, aanzetten en vastgelopen projecten, bewaar ik in een speciale map op mijn computer: mijn digitale kerkhof.'

Al die documenten in het kerkhof vinden nooit hun weg naar de buitenwereld. Toch zijn ze noodzakelijk in

het wetenschappelijke proces. 'Het is erg jammer dat onderzoek niet wordt gepubliceerd als resultaat uitblijft', betoogt Proença. 'We zouden zoveel meer van elkaar kunnen leren als wetenschappers transparanter zijn over mislukkingen. Niet alleen kun je voorkomen dat anderen dubbel werk doen, ook kan een ander iets zien waardoor het wel slaagt. Langzaam aan komt er meer openheid over dat lange proces vol pogingen, maar het blijft een moeilijk onderwerp.'

Zelflerende modellen

Praten over fouten is één, maar wat als je niet eens weet waar het fout gaat? Dat is bij zelflerende modellen vaak onduidelijk. 'Het kan zelfs problematisch worden als je niet weet hoe stabiel de output is, bijvoorbeeld bij systemen voor de gezondheidszorg of luchtvaart', zegt Proença. 'Met collega's praat ik vaak over dit soort ethische kwesties. Het is erg spannend om in een veld te werken dat de wereld aan het veranderen is.' Bang dat computers die wereld overnemen, is hij niet. 'Computers zijn eigenlijk nog steeds best dom. Een mens hoeft maar één keer een kat te zien om die daarna te kunnen herkennen. Een computer weet het na honderd foto's nog niet.' Mensen blijven altijd nodig, gelooft hij. 'Je gebruikt de machine, leert van de data, begrijpt het, maakt een nieuwe machine, en die machine leert steeds meer. Het belangrijkste van interpreteerbare modellen is het geloof dat mens en machine samen beter werken dan ieder afzonderlijk.' **□**



Wat is... **Machine learning?**

Machine learning is een vorm van kunstmatige intelligentie: algoritmes

kunnen leren van data die ze krijgen als input. In plaats van een persoon die de modellen aanpast om betere voorspellingen te doen, verbeteren deze algoritmes zelf de parameters voor hun model-

len. Als het aantal parameters dat de machines gebruiken, zo groot wordt dat voor wetenschappers niet meer duidelijk is hoe ze beslissingen nemen, worden ze 'black boxes' genoemd. Programmeurs

kunnen algoritmes ook dwingen om eenvoudiger te opereren: interpreteerbare modellen. Zo kunnen mensen begrijpen hoe computermodellen tot bepaalde resultaten komen en hiervan leren.