

STATATOR

Hoe machine learning tot nieuwe inzichten in leugenachtig gedrag kan leiden

De Bayesiaanse leercyclus

Investeren en wedden met Kelly

Ook in de toekomst moeten we niet blind varen op cijfers

Optimaal ontwerp voor het beschikbare vloeroppervlak door een slimme combinatie van optimalisatie en simulatie

Statistiek en causaliteit: voortschrijdende liaison of moeizame samenspraak

Gecombineerde radiotherapiebehandelingen – duur is niet altijd beter



STATOR

Jaargang 19, nummer 4, december 2018

STATOR is een uitgave van de Vereniging voor Statistiek en Operations Research (VVSOR). STATOR wil leden, bedrijven en overige geïnteresseerden op de hoogte houden van ontwikkelingen en nieuws over toepassingen van statistiek en operations research. Verschijnt 4 keer per jaar.

Redactie

Joaquim Gromicho (hoofdredacteur), Annelieke Baller, Ana Isabel Barros, Joep Burger, Nikky van Buuren, Kristiaan Glorie, Caroline Jagtenberg, Guus Luijben (eindredacteur), Richard Starmans, Gerrit Stermerink (eindredacteur) en Vanessa Torres van Grinsven. Vaste medewerkers: Johan van Leeuwen, John Poppelaars, Gerard Sierksma en Henk Tijms.

Kopij en reacties richten aan

Prof. dr. J.A.S. Gromicho (hoofdredacteur), Faculteit der Economische Wetenschappen en Bedrijfskunde, afdeling Econometrie, Vrije Universiteit, De Boelelaan 1105, 1081 HV Amsterdam, mobiel 06 55886747, j.a.dossantos.gromicho@vu.nl

Bestuur van de VVSOR

Voorzitter: prof. dr. Fred van Eeuwijk, db@vvsor.nl
Secretaris: Sophie Swinkels PhD, db@vvsor.nl
Penningmeester: dr. Ad Ridder, db@vvsor.nl
Algemeen: Nikky van Buuren MSc, webmaster@vvsor.nl
Voorzitters van de secties: prof. dr. ir. Mark van de Wiel (Biometrical Section); prof. dr. Albert Wagelmans (Section for Operations Research); dr. Eduard Belitser (Section Mathematical Statistics); prof. dr. Casper Albers (Social Sciences Section); dr. Michel van de Velden (Economics Section); Jonas Haslbeck MSc (Young Statisticians).

Leden- en abonnementenadministratie van de VVSOR

VVSOR, Postbus 1058, 3860 BB Nijkerk, telefoon 033 2473408, admin@vvsor.nl. Raadpleeg onze website www.vvsor.nl over hoe u lid kunt worden van de VVSOR of een abonnement kunt nemen op STATOR.

Voor advertenties

M. van Hootegem, hootegem@xs4all.nl
STATOR verschijnt in maart, juni, oktober en december.

Ontwerp en opmaak

Pharos, Nijmegen

Uitgever

© Vereniging voor Statistiek en Operations Research
ISSN 1567-3383

WERK AAN DE WINKEL

Het grootste deel van dit nummer zult u waarschijnlijk in de rust van de kerstvakantie lezen. Waarom dan de oproep Werk aan de Winkel! Simpel, omdat er nog een boel te doen valt. Wie de media een beetje volgt valt het op dat er voor statistici in de meest brede zin nog veel werk te doen is. Zelfs een basisbegrip als gemiddelde levert vaak misverstanden op. Regelmatig zijn er berichten waarvan je denkt: dat kan helemaal niet! Soms ligt het aan de journalist die een wat al te simpele weergave van een genuanceerd rapport heeft gemaakt. Maar helaas ligt de oorzaak dieper. Onbegrip en gebrek aan kennis over ons vakgebied zijn alomtegenwoordig. Zo verschijnt de laatste tijd op de website van *de Volkskrant* op gezette tijden een banner waarin wordt gesteld dat bij meer dan 35 % van de kinderen de prikkelverwerking afwijkt van het gemiddelde. Deze banner linkt door naar een *advertorial* van de Open Universiteit(!). Die advertorial maakt het nóg bonter: 'Uit enkele pilotstudies blijkt dat bij iets meer dan 35 procent van de kinderen de prikkelverwerking afwijkt van het gemiddelde. Daar zitten ook kinderen bij die prikkels beter dan gemiddeld verwerken, maar het grootste deel heeft er problemen mee.'! Ineens komt dan een herinnering aan een staatssecretaris van onderwijs naar boven die, jaren geleden, zonder blikken of blozen beweerde dat ze er naar streefde dat uiteindelijk alle scholen bovengemiddeld zouden presteren.

Gezien het bovenstaande is er geen gebrek aan werk voor onze leden en vooral niet voor de nieuwe sectie Statistics Communication die zich bezighoudt met allerlei aspecten van de presentatie van onderzoek. Dat is ook de intentie van STATOR. Al 19 jaar is het streven met toegankelijke artikelen de kennis over Statistiek en OR zo breed mogelijk te verspreiden. Het komende jaar zult u merken dat wij daarbij met andere media gaan samenwerken, dat alles met het doel de kennis over Statistiek en OR zo goed mogelijk te bevorderen. Werk aan de Winkel!

Dit nummer onderstreept onze ambitie, het bevat weer een grote verscheidenheid aan artikelen. U kunt lezen over de analyse van leugenachtig gedrag met toepassing van *machine learning* maar ook over het optimaal gebruik van een vloeroppervlak voor de opslag van staalrollen. Eén artikel willen we speciaal onder uw aandacht brengen: dat van Stefan ten Eikelder over de optimale combinatie van radiotherapeutische behandelingen bij kanker met als ondertitel 'Duur is niet altijd beter'. Stefan heeft voor dit belangrijke onderzoek de Jan Hemelrijk Award 2018 gekregen.

Aan de vooravond van onze 20e jaargang wensen we u veel leesplezier. En we hopen dat u in het nieuwe jaar opnieuw de noodzaak van Werk aan de Winkel inziet.

De redactie van STATOR wenst u een goed 2019!



4



8



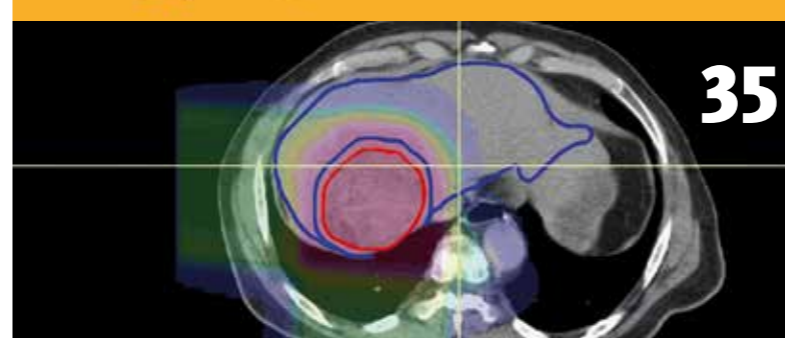
16



24



28



35

INHOUD

- 2 Redactioneel
- 4 Van top-down naar bottom-up: Hoe machine learning tot nieuwe inzichten in leugenachtig gedrag kan leiden | SOPHIE VAN DER ZEE & RONALD POPPE
- 8 De Bayesiaanse leercyclus | ERIC-JAN WAGENMAKERS & QUENTIN F. GRONAU
- 14 Investeren en wedden met Kelly – column | HENK TIJMS
- 16 Ook in de toekomst moeten we niet blind varen op cijfers | SANNE BLAUW
- 22 Gekunstelde intelligentie – column | JOHN POPPELAARS
- 24 Optimaal ontwerp voor het beschikbare vloeroppervlak door een slimme combinatie van optimalisatie en simulatie | DAAN MERKESTEIN & STEVO AKKERMAN
- 28 Statistiek en causaliteit: voortschrijdende liaison of moeizame samenspraak | RICHARD STARMANS
- 35 Gecombineerde radiotherapiebehandelingen – duur is niet altijd beter | STEFAN TEN EIKELDER
- 38 Oproep om kandidaten te nomineren voor de Jan Hemelrijk en Willem R. van Zwet Awards 2018
- 38 Young Statisticians are looking back at a great 2018
- 38 VVSOR Annual Meeting 2019
- 39 Statistiek van de sanseveria – column | GERARD SIERKSMA
- 40 LNMB & NGB Seminar 'Deep learning and its impact on Operations Research'

VAN TOP-DOWN NAAR BOTTOM-UP

Hoe machine learning tot nieuwe inzichten in leugenachtig gedrag kan leiden

SOPHIE VAN DER ZEE & RONALD POPPE

Mensen vinden betrouwbaarheid en eerlijkheid belangrijke persoonskenmerken. We streven vaak na 'een goed persoon te zijn', al is de invulling daarvan natuurlijk subjectief. We vinden het niet acceptabel als er tegen ons gelogen wordt. Je zou dus verwachten dat mensen weinig tegen elkaar liegen. Dat blijkt echter niet het geval. Er wordt volop gelogen, al is het moeilijk precies vast te stellen hoe vaak. In de wetenschappelijke literatuur wordt de frequentie waarmee mensen liegen meestal bepaald in dagboekstudies, waarin proefpersonen gevraagd worden om op te schrijven hoe vaak ze in de afgelopen 24 uur hebben gelogen. Mensen zelf rapporteren gemiddeld ongeveer twee leugens per dag te vertellen (Debey et al., 2015). Ongeveer de helft van de mensen beweert niet te hebben gelogen in de afgelopen 24 uur, en ongeveer de helft van de leugens wordt verteld door een kleine groep (9%) van *prolific liars*.

Om erachter te komen hoe goed deze onderzoeksresultaten de echte wereld weerspiegelen hebben we recentelijk meegewerkt aan een BBC Horizon documentaire over liegen en leugendetec­tie genaamd *The honesty experiment: A week of living truthfully* (uitgezonden door de

BBC op 29 augustus 2018). Drie proefpersonen hebben een dag leugendetec­tie-apparatuur gedragen in hun dagelijkse levens. Vervolgens probeerden ze een week lang te leven zonder liegen. *Spoiler alert*, dat is geen van allen gelukt. Liegen blijkt een stuk frequenter te gebeuren dan bleek uit de eerder genoemde dagboekstudies. De drie deelnemers hadden ieder een specifieke relatie met liegen. Mo, de marketing medewerkster, dacht dat ze bijna nooit loog. Ruth, de dominee die liever niet liegt, maar dat toch doet om iedereen te vriend te houden. En Ehiz, de vlogger die vaak en heel instrumenteel liegt en zelfs een geheel verzonnen online alter ego in stand houdt. Nadat we deze drie personen een week lang hebben gevolgd, bleek zelfs onze eerlijkste deelnemer Mo de hele dag door te liegen. Vaak loog ze om anderen niet lastig te hoeven vallen of om aardig gevonden te worden, maar ze loog dus wel regelmatig.

In deze documentaire hebben wij (vier leugendetec­tie-onderzoekers) twee spannende nieuwe dingen gedaan. Allereerst hebben we leugendetec­tie in de praktijk toegepast. Normaal gesproken wordt onderzoek naar leugendetec­tie in het lab gedaan. Door de omstandighe-

den zo kaal en consistent mogelijk te houden, wordt de kans op ruis in de data verkleind. Als de verdachte een beweging maakt, kunnen we aannemen dat deze veroorzaakt wordt door gesprekskenmerken zoals liegen en niet door een externe factor. Tijdens de documentaire hebben we leugendetec­tie toegepast onder geheel verschillende omstandigheden, zoals tijdens een werkmeeting, in een café en op een feestje. Tijdens deze interacties werd het gedrag van onze proefpersonen niet alleen bepaald door het gesprek, maar ook door gelijktijdige activiteiten zoals eten, drinken en wandelen. Dit bleek problematisch voor het analyseren van de fysiologische reacties. Niet alleen liegen, maar ook een sprintje trekken voor de bus leidt namelijk tot een fysiologische stressreactie. Dit is een breder probleem binnen leugendetec­tie. Bepaalde gedragingen of reacties die een teken kunnen zijn dat iemand liegt, kunnen ook veroorzaakt worden door andere factoren. Dat is bijvoorbeeld een van de redenen waarom de *polygraaf*, de leugendetec­tiemethode bekend uit Hollywood, goed is in het detecteren van leugens, maar minder in het detecteren van waarheden (i.e., een hoge *false positive rate*; Meijer & Verschuere, 2010). Hierdoor lopen juist onschul-

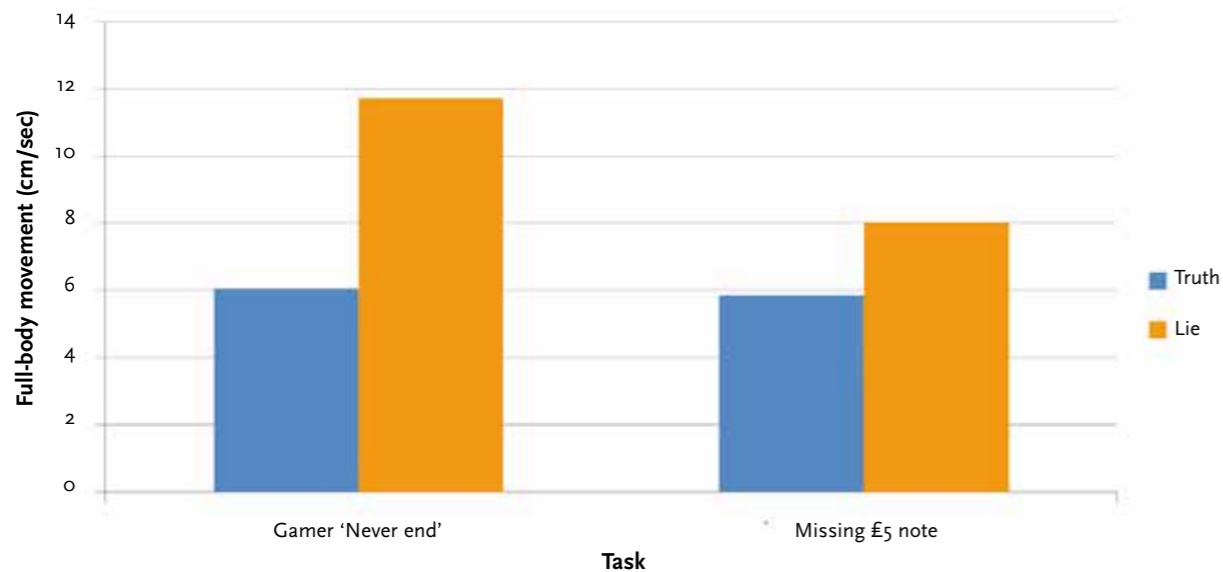
digen risico als ze instemmen met een polygraaf­test.

Het tweede unieke aan dit experiment is de combinatie van drie verschillende leugendetec­tiemethodes. Onderzoek naar leugendetec­tie richt zich doorgaans op een specifiek element van leugenachtig gedrag. Dit is echter niet erg bruikbaar in de praktijk. Voor TNO heeft Van Der Zee in 2016 onderzoek gedaan naar de kennis en behoeftes van publieke en private partijen in de Nederlandse praktijk en daaruit bleek dat er vooral behoefte is aan een multimodale aanpak waarin verschillende leugendetec­tiemethodes met elkaar worden gecombineerd tot één optimale *leugendetector* (Van Der Zee et al., 2016). Iedereen liegt namelijk op een andere manier en verschillende meetmethoden hebben allemaal hun beperkingen in accuraatheid en robuustheid. Met een combinatie van methoden krijgen we zowel een completer als nauwkeu­riger beeld van iemands gedrag.

Voor deze documentaire hebben we daarom tegelij­kertijd leugenachtig gedrag automatisch gemeten met *motion-capture pakken*, hebben we een linguïstische spraakanalyse toegepast en hebben we fysiologische reacties zoals hartslag en huidgeleiding gemeten met een



De drie proefpersonen Ehiz, Ruth en Mo met hun leugendetector. Foto: BBC/Thoroughly Modern Media



Figuur 1. Aantal lichaamsbewegingen bij het vertellen van waarheden en leugens

geavanceerde smartwatch. Door verschillende methoden te combineren is het moeilijker om succesvol de *leugen-detector* te misleiden. Dit bleek heel nuttig. Er was bijvoorbeeld een fragment waarin Mo tijdens een meeting met klanten werd gevraagd of hun project vertraging op zou lopen en welke barrières er waren. Twee van de drie leugendetectiemethoden gaven aan dat Mo's reactie niet waarheidsgetrouw was. De taalanalyse liet echter zien dat ze hoog scoorde op *cognitive processing*, waardoor we konden achterhalen dat ze niet loog, maar hard moest nadenken omdat het een lastige vraag was.

Tijdens de BBC-documentaire, maar ook in onze wetenschappelijke onderzoeken, meten we leugenachtig gedrag automatisch. De meeste onderzoeken naar leugenachtig gedrag zijn namelijk gedaan door proefpersonen een waarheid en/of leugen te laten vertellen voor een videocamera en vervolgens handmatig te annoteren welke gedragingen de persoon wel en niet maakt. Meta-analyses naar de honderden onderzoeken die op deze manier zijn gedaan hebben geleid tot inconsequente en soms zelfs tegenovergestelde resultaten en de effecten die zijn gevonden hebben kleine *effect sizes* (DePaulo et al., 2003). Daar zijn twee mogelijke verklaringen voor. Ofwel, mensen veranderen hun gedrag niet op consequente wijze wanneer zij liegen, ofwel, zij veranderen wel hun gedrag maar de manier waarop non-verbaal gedrag traditioneel werd gemeten is niet in staat om deze gedragsveranderingen aan het licht te brengen. De laatste optie is niet onaannemelijk. Gedrag is namelijk een stuk complexer dan het wel of niet maken van een bepaalde beweging. De grootte, snelheid, richting en duur van een beweging kunnen ook relevant zijn, maar deze informatie werd eerder nooit meegenomen.

Om te bepalen of liegen leidt tot gedragsveranderingen, hebben wij uitgezocht hoe je zo betrouwbaar en precies mogelijk menselijke beweging kunt meten en analyseren (Poppe et al., 2014). *Motion-capture*-apparatuur, waarbij de locatie van gewrichten vele malen per seconde worden vastgelegd, bleek een goede manier om gedetailleerd en objectief lichaamsbeweging te meten. Vervolgens hebben we deze methode toegepast op het identificeren van leugenachtig gedrag. Dit hebben we zowel top-down aangepakt op basis van theorie (Van Der Zee et al., 2015) als bottom-up met behulp van *machine learning* (Poppe et al., 2015).

Vanuit een theoretisch oogpunt (top-down-methode) hebben we de hoeveelheid lichaamsbeweging gemeten, omdat de effecten die liegen op iemand kan hebben een verschil in de hoeveelheid beweging kan veroorzaken. Samengevat kan liegen drie verschillende effecten hebben: het verhogen van de cognitieve belasting (liegen is vaak moeilijker dan de waarheid vertellen), emotionele reacties opwekken (zoals angst en schuldgevoel) en leugenaars proberen vaak zo eerlijk mogelijk over te komen waardoor ze hun gedrag proberen te controleren (DePaulo et al., 2003). Cognitieve belasting en het controleren van gedrag kunnen zorgen voor een vermindering van beweging, terwijl stress juist kan zorgen voor meer nerveuze bewegingen. In een experiment met 180 proefpersonen vertelde de geïnterviewde twee leugens of twee waarheden tegen een interviewer terwijl hun beider non-verbale gedrag automatisch werd gemeten door middel van *motion capture* pakken (Van Der Zee et al., 2015). Vervolgens hebben we voor iedere geïnterviewde berekend hoeveel ze bewogen tijdens het interview en de bewegingen vergeleken van waarheidssprekers en leugenaars. Op basis van een binaire logistische regressie konden we in totaal 82,2% van

de waarheden en leugens correct classificeren, met een accuraatheid van 89,9% voor waarheden en 75,6% voor leugens. Over het algemeen bewogen proefpersonen over hun gehele lichaam meer wanneer zij logen, dan wanneer zij de waarheid spraken (zie ook figuur 1).

Om te onderzoeken of er misschien nog onontdekte *cues to deceit* lagen, hebben we ook een *machine learning* methode toegepast op deze data (bottom-up-onderzoek). Door het gebruik van één enkele afhankelijke variabele zoals de hoeveelheid lichaamsbeweging die we hierboven gemeten hebben, kunnen details of patronen die indicatief zijn voor liegen, gemist worden. Om daar inzicht in te krijgen, hebben we een grote hoeveelheid *features* uitgerekend die de data op verschillende manieren beschrijven. Zo hebben we gekeken naar de symmetrie in de positie van ledematen, naar de hoeken en afstanden tussen paren van gewrichten en naar de gemiddelde hoeveelheid beweging van ledematen over verschillende tijdsintervallen. In totaal hadden we 335 verschillende features. Een van de belangrijkste gevaren van *machine learning* is de geringe hoeveelheid observaties (data) voor het aantal variabelen (features). Door een complex model te *fitten* op een beperkte hoeveelheid data, is het goed mogelijk om *chance findings* te doen. Zeker bij onderzoek naar leugendetectie is dit een risico aangezien in experimenteel onderzoek het interviewen van proefpersonen tijdsintensief is en daardoor de beschikbare data beperkt zijn. Data uit de praktijk zijn slechts zelden bruikbaar voor leugendetectie-doelstellingen omdat de *ground truth* – wat er echt is gebeurd – niet bekend is.

We hebben dit risico ingeperkt door alleen features te gebruiken die statistisch significant verschillend zijn tussen de data waarin gelogen werd en de data waarin de waarheid gesproken werd. Daarnaast hebben we de data gerooteerd door middel van *leave-one-out cross-validation*. Daarbij wordt een model telkens op een andere set gemaakt en vervolgens toegepast op andere *out-of-sample* data. Dit verkleint de effecten van een toevallige scheve sampling. Ten slotte hebben we een simpel classificatiemodel gebruikt: Gaussian Naive Bayes. Hierbij worden alleen gemiddelde en standaarddeviatie van elke feature gebruikt. Door te variëren met sets van features, kunnen alsnog inzichten verkregen worden. Zo vonden we dat de armen grotendeels bepalend waren voor de classificatie. Ook zagen we dat features die de symmetrie, beweging, pose en afstanden tussen gewrichten beschreven, complementaire informatie dragen. De hoog-dimensionale metingen van gedrag, typisch voor automatische observatie, kunnen dus leiden tot nieuwe inzichten.

De complementaire karakters van theorie-gedreven, validerend onderzoek (top-down) en datagedreven, ex-

ploratief onderzoek (bottom-up) kunnen goed gecombineerd worden, zoals we hier aantonen. Daarbij kan een *machine learning* aanpak leiden tot nieuwe inzichten die vervolgens na rigide analyse kunnen leiden tot nieuwe theorie. In de toekomst zullen we deze combinatie van top-down- en bottom-up-onderzoek vaker zien. Zo kunnen we ontrafelen hoeveel van het gedrag dat we meten wordt veroorzaakt door liegen, afhankelijk is van het individu en hoe gedragspatronen over tijd veranderen.

LITERATUUR

- Debey, E., De Schryver, M., Logan, G. D., Suchotzki, K., & Verschuere, B. (2015). From junior to senior Pinocchio: A cross-sectional lifespan investigation of deception. *Acta Psychologica*, 160, 58-68.
- DePaulo, B. M., Lindsay, J. L., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, 129, 74-118.
- Meijer, E. H., & Verschuere, B. (2010). The polygraph and the detection of deception. *Journal of Forensic Psychology Practice*, 10(4), 325-338.
- Poppe, R. W., Van Der Zee, S., Taylor, P. J., & Heylen, D. K. J., (2014). AMAB: Automated Measurement and Analysis of Body Motion. *Behavior Research Methods*, 46, 625-633.
- Poppe, R. W., Van Der Zee, S., Taylor, P. J., & Veltkamp, R. (2015). Mining bodily cues to deception. *Conference Proceedings of the Rapid Screening Technologies, Deception Detection and Credibility Assessment Symposium*, 48th HICSS (5-8 January, 2015).
- Van Der Zee, S., Poppe, R. W., Taylor, P.J., & Anderson, R. (2015). To freeze or not to freeze: A motion-capture approach to detecting deceit. *Conference Proceedings of the Rapid Screening Technologies, Deception Detection and Credibility Assessment Symposium*, 48th HICSS (5-8 January, 2015).
- Van Der Zee, S., van der Kleij, R., van Rest, J., & Bouma, H. (2016). Toekomstverkenning leugendetectie: Relevante toepassingen en implementatievormen in het Nederlandse veiligheidsdomein. *Ongerubriceerd TNO Rapport R10924*.

SOPHIE VAN DER ZEE is gepromoveerd op het automatiseren van leugendetectie aan de Lancaster University (Security Lancaster) waar psychologie en informatica nauw samenwerken om veiligheidsvraagstukken te beantwoorden. Sindsdien heeft ze onderzoek gedaan naar liegen, oneerlijk gedrag en cybercriminaliteit in verschillende disciplines (psychologie, informatica en gedragseconomie). Ze is universitair docent gedragseconomie aan de Erasmus Universiteit. E-mail: vanderzee@ese.eur.nl

RONALD POPPE is gepromoveerd aan de Universiteit Twente op het automatisch annoteren van gedrag uit videobeelden. Voor zijn publicaties op dit gebied ontving hij tweemaal de citatie-award van het tijdschrift *Image and Vision Computing*. Hij is universitair docent aan de Universiteit Utrecht, waar hij onderzoek doet naar machine-learningmodellen voor de automatische analyse van interactief gedrag uit video en andere sensoren. E-mail: r.w.poppe@uu.nl



DE BAYESIAANSE LEERCYCLUS

Piranha's. Foto: Today I Found Out

De regel van Bayes lijkt op het eerste gezicht weinig inspirerend. De regel (niet eens een wet) volgt direct uit de definitie van conditionele kans, en geeft aan hoe de kans op A gegeven B (bv. de kans dat een willekeurige Ajax-fan ook een Nederlander is) zich verhoudt tot de kans op B gegeven A (de kans dat een willekeurige Nederlander ook een Ajax-fan is). Hoe bedrieglijk, hoe oppervlakkig is deze eerste indruk! In werkelijkheid laat de regel van Bayes zien hoe organismen middels het toetsen van voorspellingen hun kennis optimaal kunnen bijwerken. Dit universele leerproces kan behalve door dieren en software programma's ook gebruikt worden om wetenschappelijke vragen te beantwoorden.

ERIC-JAN WAGENMAKERS & QUENTIN F. GRONAU

'Zelfs een ezel stoot zich in het gemeen niet tweemaal aan dezelfde steen.' Ezels leren namelijk van hun ervaring, en dit aanpassingsvermogen deelt de ezel met alle ons bekende diersoorten – katten, salamanders, spinnen, ja zelfs eencellige slijkwammen zijn in staat om te leren. Dat kan ook moeilijk anders, want de evolutie boetseert de natuur op een meedogenloze manier: orga-

nismen die zich niet kunnen aanpassen aan hun omgeving wacht de vergetelheid.

Maar hoe kunnen organismen leren van hun omgeving? In algemene zin kan er alleen geleerd worden als er meerdere concurrerende hypothesen bestaan. Als er maar een enkele hypothese bestaat dan spreekt men ook wel van een religieuze overtuiging, een rotsvaste mening

die door geen enkel feit aan het wankelen kan worden gebracht. Om te leren moeten we dus beginnen met meerdere concurrerende hypothesen, ieder met een zekere mate van plausibiliteit of waarschijnlijkheid. Een jonge piranha in de Amazone voelt van veraf een beweging in het water; de ene hypothese is dat het gaat om een mogelijke prooi, de andere dat het gaat om een mede-piranha. Om meer te weten te komen zwemt onze piranha dichterbij. Zo verzamelt de piranha nieuwe gegevens, en deze gegevens zouden aanleiding moeten geven tot leren, dat wil zeggen tot een aanpassing van de relatieve waarschijnlijkheid van de hypothesen. Het is intuïtief dat hypothesen toenemen en afnemen in waarschijnlijkheid al naar gelang hun voorspellende verdienste: de 'prooi-hypothese' voorspelt een heftig spartelen, en de 'mede-piranha hypothese' voorspelt een meer gelijkmatige beweging. Wanneer de nieuwe gegevens nu wijzen op een heftig spartelen dan neemt de plausibiliteit van de prooi-hypothese toe en die van de mede-piranhahypothese af.

Op basis van dergelijke algemene overwegingen komen we tot de volgende kwalitatieve wetmatigheid:

$$\text{bijgewerkte kennis over de wereld} = \text{bestaande kennis over de wereld} \times \text{relatieve voorspellende verdienste}$$

Dit zegt dat het leerproces – het bijwerken van kennis op basis van geobserveerde data – gestuurd wordt door de voorspellende verdienste van de concurrerende hypothesen. Deze redenering van het gezonde verstand wordt precies gemaakt door de regel van Bayes:

$$\text{Posterior overtuiging voor } \theta = \text{Prior overtuiging voor } \theta \times \frac{p(\text{data} | \theta)}{p(\text{data})}$$

Relatieve voorspellende verdienste voor θ

De term 'prior' staat voor de bestaande overtuiging voordat de data zijn geobserveerd, 'posterior' staat voor de bijgewerkte overtuiging, dus nadat de data zijn geobserveerd, en de Griekse letter θ (theta) representeert een

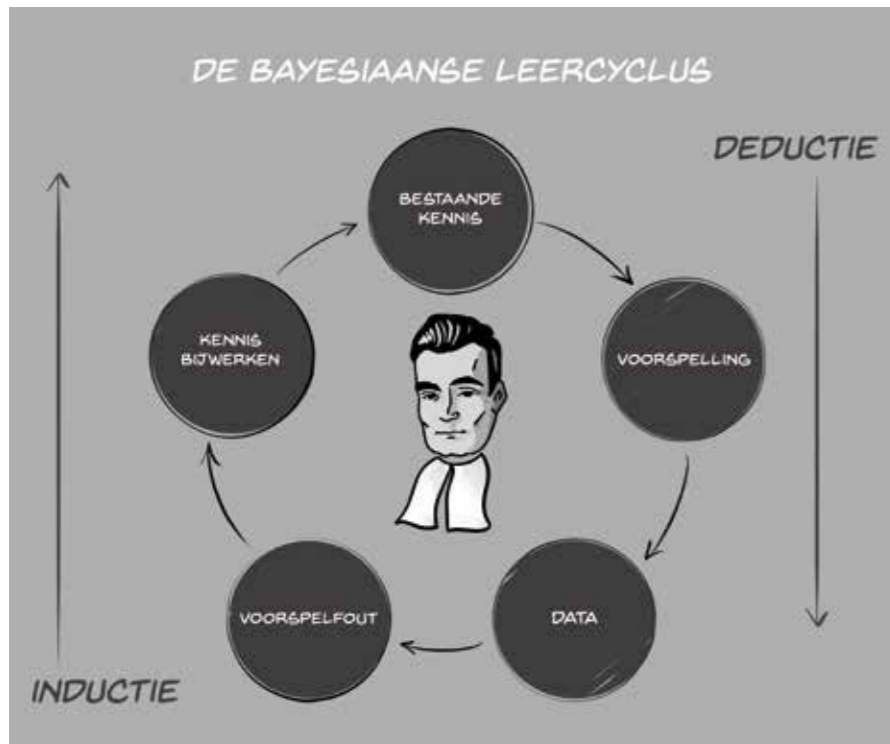
van meerdere hypothesen of proposities waarover we iets willen leren. De overtuiging voor een bepaalde hypothese θ neemt toe naarmate zij de data *beter* voorspelt dan gemiddeld, en neemt af naarmate zij de data *slechter* voorspelt dan gemiddeld (zie ook Rouder & Morey, 2017).

Het is tenslotte nog van groot belang dat het leerproces nooit hoeft te stoppen; de posterior-overtuiging voor θ , de bijgewerkte kennis dus, dient als prior-kennis voor de voorspelling van de volgende reeks gegevens. Dit is niet alleen theoretisch elegant, maar voor een organisme zoals onze piranha, die een leven lang geconfronteerd wordt met een continue stroom gegevens, is dit ook praktisch relevant: nadat de kennis is bijgewerkt hebben de oude data hun werk gedaan en kunnen ze veilig worden vergeten – het enige wat de piranha hoeft te doen is de nieuwe data gebruiken om zijn al bestaande kennis bij te schaven. Figuur 1 geeft een schematisch overzicht van de Bayesiaanse leercyclus waarin voorspellen (een deductief proces) en kennis bijwerken (een inductief proces) elkaar telkens afwisselen.

Op middelbare scholen en universiteiten wordt de regel van Bayes vaak anders onderwezen, namelijk als truc om van $p(A|B)$ te gaan naar $p(B|A)$. Stel we weten dat er in 2010 ongeveer 16,6 miljoen Nederlanders waren en 737 miljoen Europeanen, dus $p(\text{NL}) = 16,6/737 \approx 0,023$. Stel we weten ook dat er in Europa 7,1 miljoen Ajax-fans zijn, dus $p(\text{Ajax-fan}) = 7,1/737 \approx 0,010$. Ten slotte weten we dat van die 7,1 miljoen fans er 4,1 miljoen Nederlanders zijn, dus $p(\text{NL} | \text{Ajax-fan}) = 4,1/7,1 \approx 0,577$. Figuur 2 beeldt deze situatie uit als een Venn-diagram. Dan kunnen we via de bovenstaande regel van Bayes uitrekenen wat de kans is dat een willekeurige Nederlander ook fan is van Ajax: $p(\text{Ajax-fan} | \text{NL}) = p(\text{Ajax-fan}) \times p(\text{NL} | \text{Ajax-fan}) / p(\text{NL}) = 0,010 \times 0,577 / 0,023 \approx 0,251$. De regel van Bayes is natuurlijk onmisbaar voor de kansrekening, maar het is jammer om de daarmee verbonden Bayesiaanse leerfilosofie vervolgens over te slaan.

Een voorbeeld van Bayesiaans leren: Het getal π

Om de eigenschappen van Bayesiaans leren te verduidelijken behandelen we nu een concreet probleem rondom

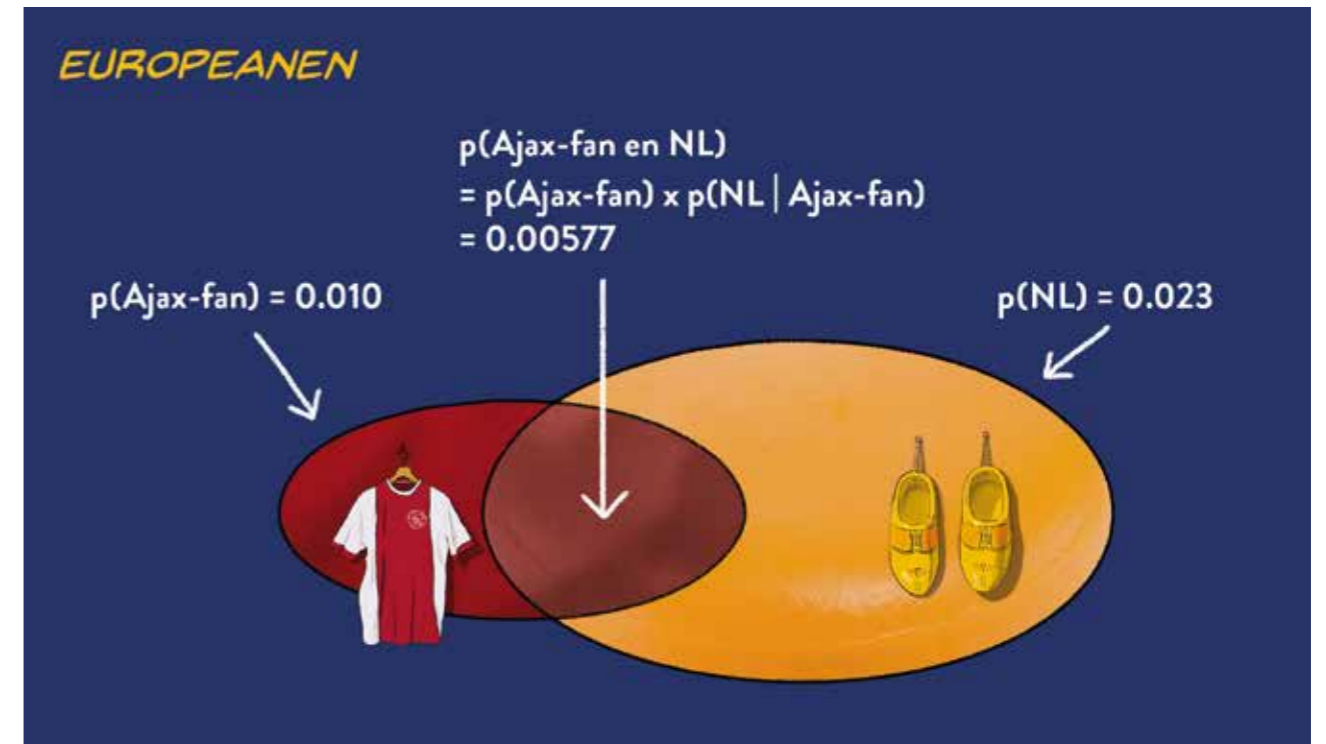


Figuur 1. Een conceptuele representatie van de Bayesiaanse leercyclus: vanuit bestaande kennis wordt een deductieve voorspelling gedaan; confrontatie met geobserveerde data leidt tot een voorspelfout, en deze geeft aanleiding tot het inductief bijwerken van de bestaande kennis; hierna kan de leercyclus worden hervat, met dien verstande dat de bijgewerkte kennis de plaats inneemt van de eerdere bestaande kennis (zie ook Jevons, 1874). Illustratie: Viktor Beekman [instagram.com/viktordepiktor](https://www.instagram.com/viktordepiktor)

het irrationale getal π , de verhouding tussen de omtrek en de diameter van een cirkel. Het getal π is 3,1415926535... en zo voort tot in de oneindigheid. Een prangende vraag die wiskundigen al langere tijd uit hun slaap houdt is of π 'normaal' is, dat wil zeggen, dat iedere specifieke reeks decimalen van een bepaalde lengte even vaak voorkomt. Er is een sterk vermoeden dat π inderdaad normaal is, maar niemand heeft het nog wiskundig kunnen bewijzen. Wij kunnen dat ook niet, maar wel kunnen we kwantificeren in hoeverre de data ondersteuning bieden voor de concurrerende hypothesen. Voor de eenvoud zullen we hier enkel onderzoeken of de even decimalen van π net zo vaak voorkomen als de oneven decimalen, en nemen we alleen de eerste honderd decimalen in beschouwing (voor een uitgebreider benadering zie Gronau & Wagenmakers, 2018).

Voordat we het leerproces kunnen starten dienen we eerst onze concurrerende hypothesen te formuleren, en wel zo specifiek dat ze ieder concrete voorspellingen kunnen maken. Onze eerste hypothese, H_0 , is dat π normaal is, en dit houdt in dat de even decimalen net zo vaak voorkomen als de oneven decimalen. Deze hypothese voorspelt dat de kans op een even decimaal telkens precies 0,5 is. Op dit moment schrikt de klassiek geschoolde statisticus misschien wakker: 'Hoe kun je nu spreken van de kans op een even decimaal? Een gegeven decimaal is

immers even of oneven, en dat kun je zo opzoeken of uitrekenen!' Dat is natuurlijk waar. De decimalen van π zijn deterministisch bepaald, en het lijkt misschien vreemd om hier een statistische methode toe te passen. Maar deze schijn bedriegt. De decimalen van π mogen dan deterministisch bepaald zijn, de uitkomst van de worp van een eerlijke munt is dat evenzeer! Met perfecte kennis van de onderliggende fysische processen (met name de verticale kracht, de draaiimpuls, en wellicht nog de luchtweerstand) kan de uitkomst feilloos worden voorspeld. Jevons (1874, pp. 197-198) verwoordde het deterministisch perspectief als volgt: 'Kans bestaat niet (...) De bliksem twijfelt niet over waar hij zal inslaan; in de zwaarste storm is geen willekeur; geen korreltje zand ligt op het strand, of oneindige kennis zou kunnen verklaren waarom het zich daar bevindt; en de vlucht van ieder vallend blad wordt bepaald door de mechanische principes die regeren over de bewegingen van de hemellichamen. Kans bestaat derhalve niet in de natuur, en valt niet te verenigen met kennis; zij is enkel een uitdrukking, zoals Laplace al opmerkte, van onze onwetendheid over de betreffende oorzaken, en ons daaruit voortvloeiend onvermogen om het resultaat te voorspellen, of om het feilloos voort te brengen. In de natuur is het geschieden van een gebeurtenis voorbestemd vanaf de eerste constructie van het universum. *Waarschijnlijkheid behoort volledig tot de*



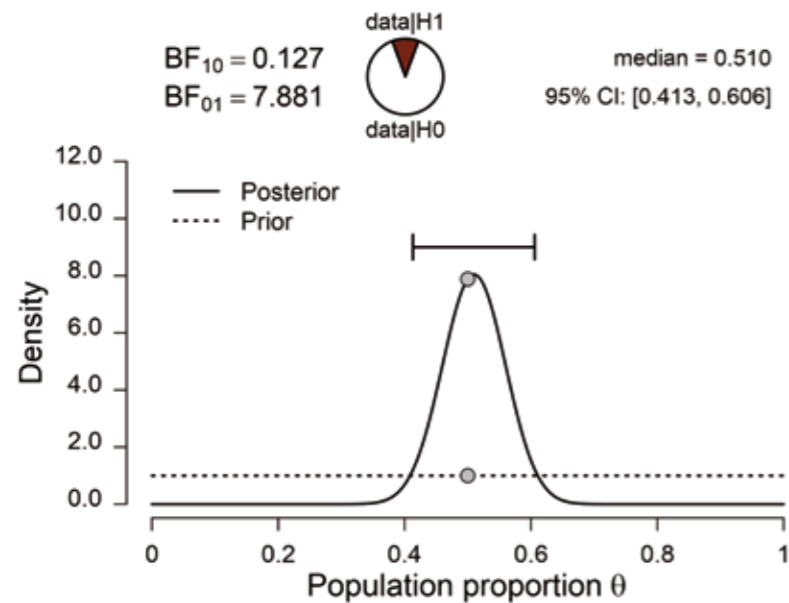
Figuur 2. In dit Venn-diagram kan het overlappende gebied (Nederlandse Ajax-fans) worden uitgedrukt hetzij als proportie van Ajax-fans (gegeven dat iemand Ajax-fan is, hoe groot is de kans dat hij of zij Nederlander is?), hetzij als proportie van Nederlanders (gegeven dat iemand Nederlander is, hoe groot is de kans dat hij of zij Ajax-fan is?); dit ligt ten grondslag aan de regel van Bayes. Illustratie: Viktor Beekman [instagram.com/viktordepiktor](https://www.instagram.com/viktordepiktor)

gedachtenwereld.' (onze vertaling; de laatste zin in het origineel luidt: *Probability belongs wholly to the mind*). Met andere woorden, als we aannemen dat π normaal is, en u wordt gevraagd wat de waarschijnlijkheid is dat de 99ste decimaal van π even is, dan kunt u vanuit uw onwetendheid – even aangenomen dat u het antwoord niet uit uw hoofd kent – niet anders doen dan deze waarschijnlijkheid inschatten op 0,5.

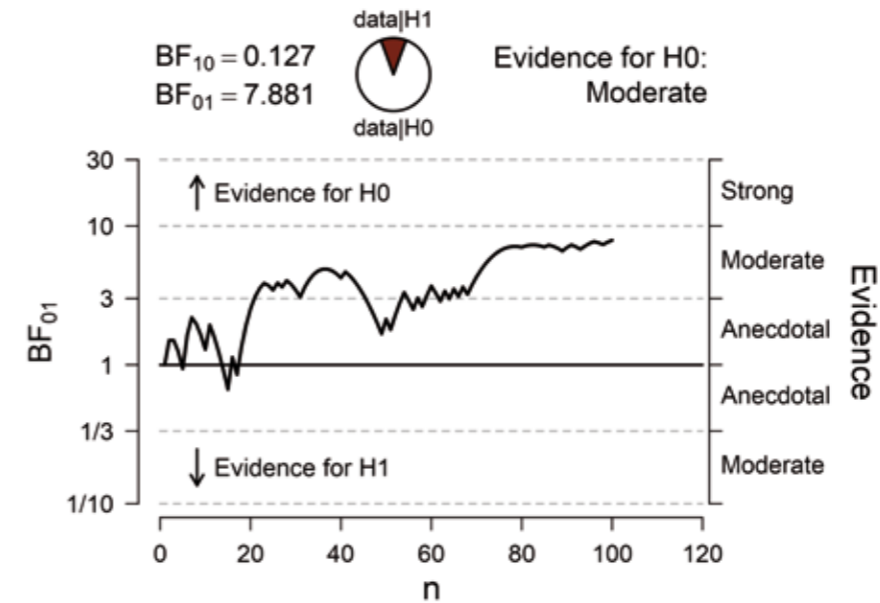
De ' π is normaal' hypothese, H_0 , is star en kent de onbekende proportie θ van even decimalen een enkele waarde toe. De tweede hypothese, H_1 , is flexibeler en staat toe dat de proportie θ alle mogelijke waardes aanneemt tussen 0 en 1. Om H_1 -voorspellingen te kunnen laten maken moet het duidelijk zijn wat de a priori waarschijnlijkheid is van ieder van die waardes voor θ . Er zijn nu plotseling twee soorten prior waarschijnlijkheid, en het is van belang deze scherp te onderscheiden. De eerste 'prior' is die voor de discrete hypothesen H_0 en H_1 . Welke hypothese is waarschijnlijker? Komen even decimalen van π net zo vaak voor als oneven decimalen, of is de verhouding toch scheef? We analyseren hier de eerste 100 decimalen van π , dus we reizen mentaal terug in de tijd tot de periode van Archimedes. Het lijkt dan niet gemakkelijk om aan H_0 en H_1 zinvolle, niet arbitraire prior kansen toe te kennen. We worden daarom iets minder ambitieus en stellen ons als doel om de

relatieve voorspellende verdienste van H_0 versus H_1 te kwantificeren, dat wil zeggen de evidentie die de data aandragen – de mate waarin de kennis bijgewerkt moet worden. De relatieve voorspellende verdienste van H_0 versus H_1 wordt ook wel de 'Bayes factor' genoemd, en wij korten hem af als BF_{01} . De tweede prior is die voor de continue set van deelhypothesen θ die vallen onder de overkoepelende hypothese H_1 . Deze prior is een verdeling over de waardes van θ die aangeeft 'onder de aanname dat even decimalen van π niet net zo vaak voorkomen als oneven decimalen, wat kunnen we dan verwachten over de mate waarin dit het geval is?' Een populaire keuze is de uniforme prior, die deze vraag beantwoordt met 'geen idee; iedere waarde voor de proportie θ is even waarschijnlijk'.

We vergelijken nu dus twee concurrerende hypothesen die ieder concrete voorspellingen maken. De starre H_0 stelt dat $\theta = 0,5$, en de flexibele H_1 stelt dat iedere waarde van θ tussen 0 en 1 even waarschijnlijk is. De Bayesiaanse leercyclus passen we nu toe op de eerste honderd decimalen van π ; hiervan zijn er 51 even. Een gemakkelijke manier om deze Bayesiaanse analyse uit te voeren is met het gratis statistisch softwareprogramma JASP (jasp-stats.org) dat in onze groep aan de Universiteit van Amsterdam wordt ontwikkeld. Figuur 3 geeft een overzicht van het resultaat (zie ook



Figuur 3. Een Bayesiaanse analyse van de vraag of het getal π net zoveel even als oneven decimalen kent, gebaseerd op de eerste 100 decimalen (waarvan 51 even). De figuur is gegenereerd door JASP, jasp-stats.org



Figuur 4. Naarmate er meer decimalen van π worden toegevoegd aan de analyse stijgt the evidentie voor de nulhypothese dat even en oneven getallen net zo vaak voorkomen. De figuur is gegenereerd door JASP, jasp-stats.org

<https://osf.io/xbuqc/>). De twee belangrijkste conclusies zijn als volgt:

1. De data bieden steun voor H_0 ; $BF_{01} = 7,881$, hetgeen betekent dat H_0 de data bijna 8 keer zo goed voorspelde als H_1 . De 'pizza plot' poogt de sterkte van de evidentie intuïtief te maken. Het witte 'mozzarella' gedeelte van de pizza representeert de posterior-model kans voor H_0 (onder een 50-50 prior-model kans) en het rode 'pepperoni' gedeelte representeert de posterior-model kans voor H_1 . Stel je prikt met je wijsvinger blindelings in de pizza, en hij landt in het niet-dominante beleg (in dit geval de pepperoni). Hoe verrast ben je? De intensiteit van je ingebeelde verrassing is een maat voor de sterkte van de evidentie.
2. Indien we aannemen dat H_1 waar is (een aanname die door de data dus enigszins wordt weersproken) dan kunnen we de continue prior- en posterior-verdeling voor θ interpreteren. Waardes van θ lager dan 0,4 en hoger dan 0,6 zijn minder waarschijnlijk geworden door de data, terwijl waardes tussen 0,4 en 0,6 juist waarschijnlijker zijn geworden. De waarde van $\theta = 0,5$ is 7,881 keer zo waarschijnlijk geworden, en dit bevestigt dat H_0 (die immers gedefinieerd is als $\theta = 0,5$) de data bijna 8 keer zo goed voorspelde als H_1 . Verder valt 95% van de posterior verdeling tussen $\theta = 0,413$ en $\theta = 0,606$, wat betekent dat we 95% zeker zijn dat de ware waarde zich in dit interval bevindt.

De evidentie voor H_0 versus H_1 kan ook worden bijgehouden terwijl de decimalen een voor een aan de analyse worden toegevoegd. De leerproces wordt dus 100 keer doorlopen, en na iedere decimaal wordt opgeslagen hoe goed H_0 en H_1 de data tot dat punt voorspelden. Het re-

sultaat van deze sequentiële analyse staat in figuur 4. Het is duidelijk dat de evidentie voor H_0 over het algemeen toeneemt naarmate er meer decimalen worden meegenomen in de analyse. De categorieën van evidentie op de tweede y-as zijn afkomstig van de Bayesiaanse pionier Harold Jeffreys; ze zijn bedoeld als grove heuristiek, als hulpmiddel bij interpretatie, maar niet als harde drempel voor alles of niets beslissingen.

De voordelen van Bayesiaans leren voor de wetenschappelijke praktijk

De Bayesiaanse analyse van π hierboven is bijzonder relevant voor de praktische onderzoeker. Ten eerste laat het voorbeeld zien dat het mogelijk is om evidentie te verzamelen in het voordeel van de nulhypothese. Als de nulhypothese de geobserveerde data beter voorspelt dan de alternatieve hypothese, dan is dit evidentie voor de afwezigheid van een effect; dit contrasteert met het scenario waar beide hypotheses de geobserveerde data ongeveer even goed voorspellen, zodat er sprake is van afwezigheid van evidentie. In de wetenschap is het vaak cruciaal om evidentie te kunnen verzamelen voor de afwezigheid van een effect: 'veroorzaakt vaccinatie autisme of doet het dat niet?'; 'replceert het effect van *power posing* of replceert het *niet*?'; 'ondersteunen de data de hypothese dat lijstlengte *niet* van invloed is op de geheugenprestatie?'

Ten tweede laat het voorbeeld zien dat evidentie bijgehouden kan worden terwijl de data geleidelijk binnenvallen. Als we niet tevreden zijn met de evidentie na 100 decimalen kunnen we eenvoudigweg meer decimalen

toevoegen tot we wel overtuigd zijn. Conceptueel is dit hetzelfde als de piranha die, nog onzeker over de bron van de beweging die hij heeft waargenomen, dichterbij gaat zwemmen om meer informatie te verzamelen en zo zijn onzekerheid te verkleinen. Ter illustratie hebben de eerste 100 miljoen decimalen van π geanalyseerd (hier van zijn er 50.006.474 even); de relatieve voorspellende verdienste van H_0 was aanzienlijk groter dan die van H_1 – de Bayes factor BF_{01} is ongeveer 3451. Voor de pragmatische onderzoeker betekent dit een toename in efficiëntie: wanneer een experiment in een vroeg stadium al doorslaggevende evidentie levert kan de datacollectie eerder worden stopgezet dan oorspronkelijk gepland.

De kern van het Bayesiaanse gedachtegoed is eenvoudig: verklaringen worden waarschijnlijker als ze de data relatief goed voorspellen, en minder waarschijnlijk als ze de data relatief slecht voorspellen. De uitdaging voor Bayesiaanse onderzoekers is om concurrerende modellen zo te specificeren dat ze zinnige voorspellingen maken, voorspellingen die de moeite van het toetsen waard zijn. De aard van de voorspellingen wordt gedeeltelijk bepaald door de prior verdeling, en dit vormt zowel de achilleshiel als de unieke kracht van het Bayesiaanse paradigma. Hoe zorgvuldiger de prior verdeling wordt gekozen, hoe zinniger de voorspellingen, en hoe betekenisvoller de statistische conclusies.

In ons voorbeeld van π hebben we een uniforme verdeling aangenomen voor de proportie θ van even decimalen. Andere keuzes zijn mogelijk en leiden onherroepelijk tot andere resultaten. Is dit een probleem? Wij denken van niet. Stel er is een aantal verschillende prior verdelingen, die allemaal verdedigbaar zijn; als de resultaten kwa-

litatief overeen komen dan hebben we ontdekt dat onze conclusies robuust zijn tegen redelijke veranderingen van prior aannames; als daarentegen de resultaten kwalitatief sterk verschillen dan moeten we erkennen dat de data onvoldoende doorslaggevend zijn, zodat plausibele, verdedigbare veranderingen in prior kennis leiden tot grote verschillen in de uiteindelijke conclusie. Dergelijke onzekerheid dient te worden benoemd en gerespecteerd, en niet onder het tapijt geveegd.

We hopen dat er in de toekomst meer aandacht wordt besteed aan de Bayesiaanse leerproces, zowel in onderzoek als in onderwijs. In de woorden van James Clerk Maxwell: 'de ware Logica van deze wereld is de Kansrekening.'

LITERATUUR

- Gronau, Q. F., & Wagenmakers, E.-J. (2018). Bayesian evidence accumulation in experimental mathematics: A case study of four irrational numbers. *Experimental Mathematics*, 27, 277-286.
- JASP Team (2018). JASP (Version 0.9) [Computer software].
- Jevons, W. S. (1874). *The principles of science: A treatise on logic and scientific method*. London: MacMillan.
- Rouder, J. N., & Morey, R. D. (2017). Teaching Bayes' theorem: Strength of evidence as predictive accuracy. *The American Statistician*, DOI: 10.1080/00031305.2017.1341334.
- ERIC-JAN WAGENMAKERS is hoogleraar Neurocognitieve Modelering aan de Faculteit der Maatschappij- en Gedragwetenschappen (programmagroep Psychologische Methodenleer) van de Universiteit van Amsterdam.
E-mail: E.M.Wagenmakers@uva.nl
- QUENTIN F. GRONAU is PhD-student aan de Faculteit der Maatschappij- en Gedragwetenschappen (programmagroep Psychologische Methodenleer) van de UvA.
E-mail: quentin.f.gronau@gmail.com



INVESTEREN EN WEDDEN MET KELLY

In zijn boek *A Mathematician Plays the Stock Market*, Basic Books, 2003, voert John Allen Paulos het volgende denkbeeldige scenario ten tonele. Een investeerder krijgt de gelegenheid om 52 weken lang elke maandagochtend geld te beleggen in een aandeel van een dotcom-bedrijf en het aandeel dan de vrijdagmiddag daarop te verkopen, waarna aan het begin van de week daarop weer een nieuw aandeel in een andere dotcom-bedrijf gekocht kan worden. Voor elke week geldt dat de prijs van het aangeschafte aandeel in die week 80% stijgt met kans 1/2 en 60% daalt met kans 1/2 onafhankelijk van het verloop van de markt in voorgaande weken. Dit betekent dat gemiddeld genomen de stijging van de waarde van het aangeschafte aandeel gelijk is aan $0,8 \times 0,5 - 0,5 \times 0,6 = 0,1$, ofwel een gemiddeld rendement van 10% per week. De investeerder, die een beginkapitaal van 10 duizend dollar heeft voor de investeringen in de komende 52 weken, aarzelt geen moment en besluit aan het begin van elke week zijn totale kapitaal in aandelen van een dotcom-bedrijf te steken. Na 52 weken blijkt dat de investeerder minder dan 2 dollar over heeft van zijn oorspronkelijke kapitaal van 10 duizend dollar. De investeerder staat perplex. Het beleg-

gingsresultaat is echter niet zo verrassend wanneer je beseft dat werken met gemiddeldes heel gevaarlijk kan zijn in situaties met onzekere factoren. In een meer dan gemiddeld 25 cm diep is, kun je nog steeds verdrinken.

In situaties met onzekere factoren moet je niet werken met gemiddeldes maar met kansen! Door te werken met kansen is het simpel in te zien dat met grote kans het beginkapitaal van 10 duizend dollar bijna helemaal verdampt als je elke keer je totale kapitaal investeert. Het meest waarschijnlijke pad dat zich in 52 weken zal ontwikkelen is een pad waarin het aandeel de helft van de tijd in waarde stijgt en de helft van de tijd in waarde daalt. In zo'n pad is het kapitaal na 52 weken gelijk aan $1,8^{26} \times 0,4^{26} \times \$ 10.000 = \$ 1,95$. Simuleren we 100 duizend keer de investeringen over de 52 weken, dan vinden we dat het eindkapitaal niet meer dan \$ 1,95 zal zijn met een kans van ongeveer 50%, terwijl de kans om met meer dan het beginkapitaal 10 duizend dollar te eindigen een schamele 5,8% is.

Misleid door de zonnig ogende gemiddeldes, belegde de investeerder elke keer zijn totale kapitaal. Blijkbaar was de hij niet op de hoogte van de Kelly-strategie. Deze

strategie investeert niet je gehele kapitaal, maar investeert elke keer eenzelfde vaste fractie van je beschikbare kapitaal. In bovenstaande situatie zou de Kelly-strategie voorgeschreven hebben om elke keer een fractie 5/24 van je beschikbare kapitaal te investeren. Bij deze strategie was de kans praktisch gesproken nul geweest om met minder dan \$ 1,95 te eindigen na 52 weken. De investeerder had dan na 52 weken met een kans van ongeveer 70% meer dan 10 duizend dollar gehad en met een kans van 44% meer dan 20 duizend dollar. De getalwaarde $f^* = 5/24$ volgt door in de beroemde Kelly-formule

$$f^* = \frac{pr_1 + (1-p)r_2 - 1}{(r_1 - 1)(1 - r_2)}$$

de waarden $p = 0,5$, $r_1 = 1,8$ en $r_2 = 0,4$ te nemen. Deze formule heeft betrekking op de volgende algemene situatie. Stel je kunt herhaald weddenschappen afsluiten op een spel dat 'gunstig' voor je is, waarbij gunstig betekent dat de verwachtingswaarde van je winst per spel positief is. Voor elke dollar die je bij een uitvoering van het spel inzet, krijg je r_1 dollar terug met kans p en r_2 dollar met kans $1 - p$, waarbij $0 < p < 1$, $r_1 > 1$ en $0 \leq r_2 < 1$. De aanname is dat $pr_1 + (1 - p)r_2 > 1$, dat wil zeggen, het spel is gunstig voor je op de lange duur. Je begint met een gegeven kapitaal en aangenomen wordt dat je elke keer elk bedrag kan inzetten tot het maximum van je huidige kapitaal. Wil je de groeivoet van je kapitaal op de lange duur maximaliseren, dan vertelt de Kelly-formule je dat je elke keer dezelfde vaste fractie f^* van je beschikbare kapitaal moet inzetten. Voor het speciale geval dat $r_2 = 0$ wordt de Kelly-fractie gegeven door $f^* = (pr_1 - 1) / (r_1 - 1)$ en kan dan geïnterpreteerd worden als het quotiënt van de verwachte netto winst per ingezette dollar en de zogeheten odds van de uitbetaling.

De inzet onder de Kelly-strategie kan in bepaalde gevallen erg groot zijn waardoor het verloop van de hoogte van je kapitaal heel grote schommelingen kan vertonen, wat bij de meeste mensen niet bevorderlijk is voor de nachtrust. Bijvoorbeeld als $p = 0,95$, $r_1 = 2$ en $r_2 = 0$, dan wordt elke keer 90% van je kapitaal ingezet onder de Kelly-strategie en bij een verlies duikt je kapitaal dus dan met 90%. Vandaar dat in de praktijk veelal een fractionele Kelly-strategie wordt gebruikt. Onder zo'n strategie zet je elke keer een fractie $f = cf^*$ van je kapitaal in, waar-

bij c een getal tussen 0 en 1 is. Empirisch is gevonden dat onder de fractionele Kelly-strategie het lange-termijn rendement bij benadering met een factor $1 - c(2 - c)$ afneemt in vergelijking tot de Kelly-strategie. Een fractionele Kelly strategie $f = cf^*$ is minder risicovol en dit wordt tot uitdrukking gebracht door de benaderingsformule $(1 - b^{2/c-1}) / (1 - (b/a)^{2/c-1})$ voor de kans dat je kapitaal eerder boven de waarde aW_0 stijgt dan onder de waarde bW_0 daalt voor waarden van a en b met $0 < b < 1 < a$, waarbij W_0 je beginkapitaal is. Bijvoorbeeld, door elke keer de halve Kelly-fractie in te zetten in plaats van de volle Kelly fractie, geef je een kwart van je maximale lange-termijn rendement op maar vergroot je de kans van 0,67 tot 0,89 om tot een verdubbeling van je beginkapitaal te geraken zonder dat het eerst een keer gehalveerd is. Bij een fractionele Kelly-strategie met de in de praktijk veel gebruikte waarde $c = 0,3$ is de reductie van je maximale lange-termijn rendement ongeveer 49% en is de kans ongeveer 98% om tot een verdubbeling van je beginkapitaal te geraken zonder dat het eerst een keer gehalveerd is.

De Kelly-strategie werd voor het eerst gebruikt door de wiskundige Edward Thorp in casino's waar hij zijn winnende blackjacksysteem uitprobeerde. Later paste Thorp en met hem beroemde beleggers waaronder Warren Buffett, de Kelly-strategie succesvol toe op de aandelenmarkt. Tegenwoordig wordt een gegeneraliseerde versie van de Kelly-strategie veelvuldig gebruikt bij weddenschappen op sportevenementen zoals paardenrennen en voetbalwedstrijden waarin op hetzelfde evenement meerdere weddenschappen tegelijk kunnen worden afgesloten. De huidige toepassingen van het nu zogeheten Kelly systeem had John Kelly Jr. niet in gedachten toen hij in 1956 zijn beroemde formule ontwikkelde op het onderzoekslaboratorium Bell Labs. Zoals maar weer eens blijkt, fundamenteel onderzoek leidt vaak tot onvoorziene toepassingen.

HENK TIJMS is emeritus hoogleraar Operations Research aan de Vrije Universiteit en auteur van diverse leerboeken over operations research en kansrekening. Deze column is een verkorte Nederlandse versie van hoofdstuk 16 uit zijn net verschenen boek *Surprises in Probability; Seventeen Short Stories*, Chapman & Hall, 2018. E-mail: h.c.tijms@xs4all.nl



OOK IN DE TOEKOMST MOETEN WE NIET BLIND VAREN OP CIJFERS

Bij het standaardiseren, verzamelen én analyseren van cijfers kan er van alles misgaan. Er kunnen vooroordelen verstopt zitten in meetmethoden, steekproeven zijn niet representatief en er kunnen verbanden worden gelegd die er in werkelijkheid niet zijn. Maar spelen die problemen ook in tijden van big data? Moeten we ons nog zorgen maken over cijfers nu de hoeveelheid data zo groot is en de algoritmes zo precies zijn? Spoiler alert: ja. Waarom, dat vertel ik in dit artikel.

SANNE BLAUW

Maak kennis met de 65-jarige Keniaanse Jenipher. Ze verdiende jarenlang haar geld met het verkopen van eten in het zakendistrict van Nairobi. De zaken liepen prima, maar ze kon niet investeren. Het probleem? De bedragen die ze via microfinanciering kon krijgen waren te klein, de rentes bij woekeraars te hoog. En een gewone bank zag een lening aan haar niet zitten, want ze had geen onder-

pand. Bovendien miste ze iets: een kredietscore.

In 1956 begonnen Bill Fair en Earl Isaac hun bedrijf Fair, Isaac and Company (FICO). FICO werd opgericht vanuit een simpele gedachte: met data kun je beter inschatten of mensen hun lening zullen terugbetalen. Tot dan toe werd zo'n beslissing genomen op basis van wat mensen over je zeiden, hoe je overkwam in een gesprek

en hoe de onderbuik van de bankier aanvoelde. Dat pakte niet voor iedereen goed uit. In oude Amerikaanse kredietrapporten kun je lezen dat 'voorzichtigheid [geboden is] in grote transacties met joden'.

Fair en Isaac bedachten een formule die niet keek naar je achtergrond, maar naar je financiën. Op basis van die gegevens berekenden ze een score die aangaf hoe groot de kans was dat je een lening zou terugbetalen. De FICO-score bleek een uitkomst: miljoenen mensen kregen toegang tot leningen en kredietverstrekkers verdienden meer geld, omdat de score veel beter dan zichzelf voorspelde wie wanbetalers zouden zijn.

In Nederland hebben we sinds 1965 de Stichting Bureau Krediet Registratie (BKR). Als je een nieuwe lening wilt afsluiten, dan is de bank verplicht je gegevens daar op te vragen. De stichting berekent op aanvraag van kredietverstrekkers ook een persoonlijke score. Ook in veel andere landen worden kredietscores gebruikt. Toch hebben nog altijd miljoenen mensen er geen. Maar sinds een paar jaar bestaat er een mogelijkheid om ook de Jeniphers een kredietscore te geven, vertelt Shivani Siroya in 2016 in een TED-talk.¹ Siroya is de CEO van Tala, een start-up die big data gebruikt om leningen toe te kennen. Jenipher mocht dan tot een paar jaar geleden geen kredietscore hebben, ze had wel een mobiele telefoon die allerlei data over haar bijhield. Op een dag overtuigde Jeniphers zoon haar om de Tala-app te installeren. Ze vroeg een lening aan en, op basis van haar data, kreeg ze binnen de kortste keren een lening. Twee jaar later is haar leven compleet veranderd: ze runt drie kraampjes en heeft plannen voor een restaurant.

Een van de gevaarlijkste ideeën van dit moment

Jeniphers verhaal is hartverwarmend. En al is het een promotieverhaal van Tala, het vertelt alles over de ontwikkeling waar we nu midden in zitten: de big data-revolutie. Big data worden vaak omschreven met de vier v's: *volume*, *velocity*, *variety* en *veracity*. De verwachtingen zijn hoog. Beleidsmakers, bedrijfslieden en publieke intellec-

tuelen stellen dat we met big data het klimaatprobleem kunnen oplossen, de gezondheidszorg kunnen transformeren en honger de wereld uit kunnen helpen. Met big data zouden we zelfs de democratie kunnen redden. We hebben niets aan verkiezingen als veel mensen toch niet stemmen, stelde universiteitsbestuurder Louise Fresco in 2016 in NRC. 'Wat als we democratische verkiezingen vervangen door een systeem van kunstmatige intelligentie?'

Dit gedachte-experiment mag bizar lijken: big data-algoritmes krijgen al steeds meer macht. Verzekeraars gebruiken algoritmes om te berekenen welke premie je moet betalen, de Belastingdienst om in te schatten of je fraude zal plegen en Amerikaanse rechters om te beoordelen of een gevangene eerder vrijgelaten wordt. De aanname dat we cijfers prima kunnen laten beslissen over ons leven is gevaarlijk. Achter deze opvatting gaat namelijk een ernstig misverstand schuil: dat de data altijd overeenkomen met de werkelijkheid.

Waarover we praten als we over algoritmes praten

Maar eerst een kijkje onder de motorkap. Net zoals vroeger gemiddelden en grafieken werden bedacht om de grote bergen informatie te doorgronden, zo bedenken slimmeriken nu methoden om de triljoenen bytes aan informatie te temmen. Die technieken – algoritmes – beslissen welke zoekresultaten je krijgt op Google, welke posts je ziet op Facebook en wie er een lening krijgt van bedrijven als Tala. Eigenlijk is een algoritme niet meer dan een aantal stappen die je zet om een bepaald doel te bereiken. Op een computerscherm ziet het er droogjes uit: regel na regel schrijft een software-ontwikkelaar in computertaal welke stappen er moeten worden gezet onder welke omstandigheden. Zo'n regel kan bijvoorbeeld een als-dan-commando zijn: 'als iemand haar lening heeft terugbetaald, dan gaat haar kredietscore tien punten omhoog.'

Hoe werkt een algoritme? De Amerikaanse wiskun-

dige en auteur Cathy O'Neil legt het uit in haar boek *Weapons of Math Destruction* aan de hand van een praktisch voorbeeld: koken voor haar gezin. Ze is tevreden als haar gezin (a) voldoende eet, (b) het eten lekker vindt en (c) genoeg voedingsstoffen binnenkrijgt. Door elke avond te evalueren hoe het staat, weet ze hoe het beter kan. Dat haar kinderen spinazie laten staan maar broccoli verslinden, bijvoorbeeld, helpt haar om ze gezonder te laten eten. Bij het bereiken van haar doelstellingen heeft ze wel een aantal beperkingen. Haar man mag geen zout en een van haar zoons lust geen hamburgers. Ook heeft ze niet eindeloos veel budget, tijd en zin om te koken. O'Neil heeft het proces na jaren oefening goed in de vingers. Stel nu dat een computer haar taak komt overnemen. Daarvoor moet ze een manier bedenken om haar doelstellingen te *standaardiseren*. Om te weten of haar gezin genoeg, lekker en gezond eet, kan ze bijvoorbeeld kijken naar (a) aantal calorieën, (b) tevredenheidsscores en (c) percentage van de aanbevolen dagelijkse hoeveelheid van elke voedingsstof.

Als ze eenmaal heeft bedacht wat en hoe ze wil gaan standaardiseren, dan kan ze de data gaan *verzamelen*. Ze zou kunnen beginnen met een lijst van alle mogelijke recepten, inclusief bereidingstijd, prijs en voedingswaarde. Steeds kan ze noteren hoe het eten scoort op hoeveelheid en gezondheid, en ze kan haar gezinsleden vragen om elk gerecht een cijfer te geven. Met die data zou O'Neil een programma kunnen schrijven dat precies uitspelt wat haar gezin elke dag van het jaar moet eten. Ze kan het programma ook zelflerend laten worden. Zolang alles in cijfers gegoten is, kan de computer *analyseren* wat de samenhang is tussen de gerechten en de doelstellingen. Uiteindelijk kan het algoritme misschien zelfs nieuwe gerechten gaan suggereren op basis van de doelstellingen die O'Neil heeft gegeven. Op die manier gebruikt haar computer *machine learning*, een vorm van kunstmatige intelligentie, om een taak te leren die niet stap voor stap voorgeprogrammeerd is. Kortom, O'Neil maakte haar kooktaak meetbaar, verzamelde de cijfers en liet software de gegevens analyseren. Maar met de algoritmes die O'Neill schreef kan veel misgaan: abstracte concepten worden in een cijfer gevat; de herkomst van big data kan schimmig zijn; en correlatie is nog altijd niet gelijk aan causaliteit. Het verhaal van kredietcores laat zien waarom.

1. Abstracte concepten worden in een cijfer gevat

Er zijn in de financiële sector meer bedrijven als Tala die big data gebruiken om iemands kredietwaardigheid in te schatten. Neem Zest Finance, dat sinds 2009 meer dan driehonderd miljoen individuen een score gaf. Het bedrijf,

opgericht door de voormalig Googlebestuurder Douglas Merrill, stelt dat het traditionele kredietcoresysteem beperkt is door 'little data'. Conventionele kredietcores, zoals ooit bedacht door Fair en Isaac, maken gebruik van 'minder dan vijftig datapunten'. Zest gebruikt daarentegen meer dan drieduizend variabelen om iemand te beoordelen, aldus het bedrijf. Ook in Nederland gebruiken tal van bedrijven big data om de betaalmoraal van klanten te meten. Zo geeft de Nederlandse datahandelaar Focum iedereen een cijfer tussen één en elf. Heb je een rekening nog niet betaald? Tien punten eraf, of het nou om twintig of twintigduizend euro gaat.

Nu kun je je afvragen: wat is hier mis mee? Kredietcores bieden immers ook mogelijkheden, zoals het verhaal van Jenipher laat zien. Maar big data-algoritmes proberen vaak abstracte concepten te meten, en daar kan het fout gaan. In het Amerikaanse rechtssysteem wordt berekend wat de kans is dat een veroordeelde opnieuw de fout ingaat. Die berekeningen hebben grote gevolgen: ze spelen mee bij de beslissing of iemand vervroegd wordt vrijgelaten. Maar als iets abstract is en lastig te meten, dan is het wel wat er in de toekomst gaat gebeuren. Als we vergeten dat zulke voorspellingen slechts een *benadering* zijn van iemands gedrag, dan veroordelen we mensen op basis van gebrekkige cijfers.

Bij kredietcores is nog iets aan de hand. Vaak worden ze gebruikt om ook een ander concept uit te drukken dan toekomstig gedrag, iets dat minstens even abstract is: betrouwbaarheid. Uit een Amerikaans onderzoek onder humanresources-medewerkers uit 2012 bleek dat zo'n 47 procent van de werkgevers de kredietgeschiedenis van sollicitanten controleert. Een ander onderzoek onder Amerikaanse huishoudens met creditcardschulden constateerde dat één op de zeven respondenten met een slechte kredietgeschiedenis te horen kreeg dat hij of zij een baan niet kreeg vanwege hun verleden. Terwijl er geen enkel bewijs is voor een verband tussen je leengedrag en je prestaties op de werkvloer.

Onderzoeker Jeremy Bernerth en collega's deden een van de weinige studies die beschikbaar zijn. Ze legden in 2012 individuele FICO-scores naast persoonlijkheidstesten. Mensen met een hogere kredietcore scoorden beter op zorgvuldigheid, maar waren juist minder servicegericht. Op andere kenmerken was er geen enkel verschil. Belangrijker: er was geen enkel verband tussen kredietcores en frauduleuze praktijken. Niet voor niets is het in elf Amerikaanse staten inmiddels verboden om als werkgever naar kredietgegevens te vragen. In Nederland mogen alleen kredietverstrekkers die zijn aangesloten bij BKR je betalingsgeschiedenis inzien.

2. De herkomst van big data kan schimmig zijn

Big data kunnen helpen om fundamentele problemen op te lossen bij dataverzameling. Zo is steekproefgrootte geen probleem meer. Zeker in een land als Nederland zit intussen bijna iedereen op internet. Bovendien houden allerlei apparaten – thermostaten, auto's, fitbits – bij wat we doen. Ook de bedrijven achter kredietcores weten dat in tijden van big data de persoonlijke gegevens voor het oprapen liggen. Zij hoeven deze gegevens meer niet via de officiële wegen op te vragen, maar kunnen het internet afstruinen voor data. Soms zijn de gegevens die ze verzamelen openbaar, zoals inschrijfgegevens bij de Kamer van Koophandel. Soms heb jij ooit - vaak zonder dat je het doorhad - toestemming gegeven tot het delen van informatie.

Regelmatig komen de data uit obscuurdere hoeken. In oktober 2017 publiceerden *De Groene Amsterdammer* en *Investico* een onderzoek van journalisten Karlijn Kuijpers, Thomas Muntz en Tim Staal naar datahandelaren in Nederland. Ze ontdekten dat sommige bedrijven rechtstreeks gegevens ontvangen van incassobureaus. Zo kwamen schuldenaren zonder hun medeweten in een database terecht en kon hun financiële geschiedenis – ook lang nadat ze hun schulden hadden afbetaald – hen blijven achtervolgen. Vaak is niet te achterhalen of de gebruikte data wel kloppen, omdat het onduidelijk is welke gegevens zijn gebruikt. Om de proef op de som te nemen, schakelden de journalisten tien mensen in die hun eigen gegevens bij drie databureaus opvroegen. Ze ontvingen bijna niets. Maar toen de journalisten zich vervolgens voordeden als een klant en de data over diezelfde mensen kochten, kregen ze ineens uitgebreide datarapporten.

Het staat buiten kijf dat er regelmatig fouten voorkomen in data. De Amerikaanse Federal Trade Commission constateerde in 2012 dat in hun steekproef maar liefst een kwart van de mensen een fout had in een kredietrapport van een van de drie grote bureaus. Voor één op de twintig was de afwijking zo ernstig, dat deze mensen waarschijnlijk onterecht een hogere rente moesten betalen voor leningen. En tussen 2009 en 2010 leken er in het Verenigd Koninkrijk 17.000 zwangere mannen te wonen. Jawel, zwangere *mannen*. De code waarmee hun medische behandeling was geregistreerd was verwisseld met die van een verloskundige procedure. Uit *iOverheid*, een rapport uit 2011 van de Wetenschappelijke Raad voor het Regeringsbeleid, bleek dat zulke datafouten ook in Nederland gemaakt worden. Verkeerde adresgegevens in de gemeentelijke basisadministratie, foutieve inkomsten bij de Belastingdienst en het UWV, een onterechte registratie als crimineel in een politiedatabase - overal duiken mis-

sers op. Een oud adagium in de statistiek luidt: *Garbage in, garbage out*. Je kunt nog zo'n glad *machine learning*-algoritme bouwen, je hebt er niets aan als de gebruikte gegevens niet deugen.

3. Correlatie is nog altijd niet gelijk aan causaliteit

Een traditionele kredietcore, zoals de FICO-score, is gebaseerd op data over jou. Of je ooit geld hebt geleend, hoeveel je leende en of je het op tijd terugbetaalde. Die factoren, is de gedachte, kunnen voorspellen of je in de toekomst je lening terug zal betalen. Er is genoeg reden om deze redenering onrechtvaardig te vinden. Schulden worden regelmatig veroorzaakt door hoge medische kosten of ontslag. Sommige mensen kunnen zulke tegenslagen opvangen met hun spaargeld, maar niet iedereen heeft daar genoeg vermogen voor. Zo is een kredietcore niet alleen een maatstaf van betrouwbaarheid, maar ook van mazzel.

De berekening van big data-kredietcores gaat nog een stap verder. Neem weer Jenipher en haar eetkraam. Hoe bepaalde Tala dat de Keniaanse een lening mocht krijgen? Daarvoor moest Jenipher het bedrijf via een app toegang geven tot haar telefoon. Haar telefoongegevens lieten zien dat ze geregeld belde met familie in Oeganda. Bovendien communiceerde ze met wel 89 verschillende mensen. Factoren die volgens Tala's algoritme de kans verhogen dat Jenipher haar lening gaat terugbetalen. Dat ze regelmatig contact heeft met dierbaren, bijvoorbeeld, verhoogt die kans volgens de analyse met 4 procent. Ook een vast dagelijks patroon en het hebben van meer dan 58 contacten lijken gunstige signalen.

Big data kredietcores werken dus anders dan traditionele scores. De algoritmes kijken niet alleen naar wat jij hebt gedaan, maar naar wat *mensen zoals jij* hebben gedaan. Ze zoeken naar verbanden – correlaties – in de data en voorspellen daarmee wat jij zal gaan doen. Daarbij zijn alle cijfers welkom, zolang ze maar goed voorspellen. Zelfs de woorden in iemands aanvraag kunnen al veelzeggend zijn. Douglas Merrill van Zest Finance stelde in 2013 dat een aanmelding met enkel hoofdletters – of juist alleen kleine letters – een indicatie kan zijn van slecht betaalgedrag. Ook je koopgedrag of je sociale media-accounts kunnen worden gebruikt om je betaalmoraal te voorspellen.

Ooit lieten bankiers hun beslissing om iemand al dan niet een lening te verstrekken beïnvloeden door vooroordelen over ras, sekse en klasse. De FICO-scores moesten daar een eind aan maken. Maar met big data-kredietcores lijken we weer precies hetzelfde te doen als die ouderwetse bankier: iemand beoordelen aan de hand van de groep waartoe hij of zij behoort. Alleen worden die groepen nu

gedefinieerd als de Hoofdletterschrijvers, de Koopjesjagers, de Vriendlozen. Kijk je onder het oppervlak van de cijfers dan zie je dat er weinig nieuws aan is. Het schrijven in hoofdletters is waarschijnlijk gecorreleerd met je onderwijsniveau. Het hebben van LinkedIn-contacten met het hebben van een baan. En waar je winkelt zegt veel over je inkomen. Zo maken algoritmes vaak precies hetzelfde onderscheid als die ouderwetse bankier: laag- of hoogopgeleid, met of zonder baan, arm of rijk. Statistici noemen het correlaties, anderen noemen het vooroordelen.

Hoe zit het met correlatie en causaliteit nu we big data hebben? Volgens Chris Anderson, de voormalig hoofdredacteur van technologie tijdschrift *Wired*, hoeven we ons daar geen zorgen meer over te maken. De verklaring voor bepaalde verbanden is onbelangrijk, schreef hij in 2008 in zijn invloedrijke artikel 'The End of Theory'. 'Googles basisfilosofie is dat we niet weten waarom deze pagina beter is dan die: als de statistieken [...] zeggen dat het zo is, dan is dat goed genoeg.' Dat correlatie niet gelijk is aan causaliteit doet er volgens Anderson niet meer toe. 'Petabytes staan het ons toe te zeggen: 'Correlatie is genoeg.'

Een zeer naïeve uitspraak. Ook in het big datatijdperk is correlatie niet genoeg. Neem Google Flu Trends, het algoritme dat in 2008 met veel tamtam werd geïntroduceerd. Aan de hand van zoekopdrachten beloofde Google te kunnen voorspellen waar, wanneer en hoeveel griepgevallen er zouden komen. Twee of drie jaar lang voorspelde het model vrij nauwkeurig wanneer en waar de griep zou toeslaan. Maar in de jaren die volgden zat het algoritme er steeds naast, met in 2013 het dieptepunt toen het algoritme meer dan twee keer te veel griepgevallen voorspelde. De bouwers van het algoritme hadden uit vijftig miljoen zoektermen de vijfenveertig gekozen die het sterkst correleerden met de bewegingen in de griepgolf. Vervolgens hielden ze de zoekacties op die termen in de gaten. Dat klinkt logisch, maar: als je maar lang genoeg zoekt, vind je altijd wel een verband. Sterker nog, juist bij big data heb je last van dit probleem. Want hoe meer datapunten je hebt, hoe meer verbanden je zult vinden die significant zijn. Zo vonden de onderzoekers een sterk verband tussen de zoekterm 'high school basketball' en de verspreiding van griep.

Een ander probleem met het algoritme was dat bouwers belangrijke ontwikkelingen negeerden. Veranderingen in het ontwerp van Googles eigen zoekmachine bijvoorbeeld. Zo liet de website vanaf 2012 mogelijke diagnoses zien als iemand bijvoorbeeld 'hoesten' of 'koorts' opzocht. Een van die diagnoses? Griep. Hierdoor gingen mensen waarschijnlijk vaker op zoek naar informatie over de ziekte en overschatte het Google Flu-algoritme

de griepgolf. Ook kredietbureaus, zoals we eerder zagen, doen aan voorspellen. In die voorspelling liggen net zo goed toevallige correlaties op de loer en kunnen belangrijke ontwikkelingen evenzeer roet in het eten strooien. Als eenmaal bekend wordt dat je bepaalde woorden in een aanmelding moet gebruiken, bijvoorbeeld, dan kunnen mensen daarop inspelen en zeggen de correlaties weinig meer.

Maar stel nu dat we ons in de toekomst over deze twee valkuilen geen zorgen meer hoeven te maken. Dat we manieren vinden om toevallige correlaties te herkennen en we veranderingen *real time* in de gaten houden. Kunnen we dan wel blind op cijfers vertrouwen?

Cijfers die de werkelijkheid niet vangen, maar vervangen

'Ik ga echt niet jarenlang naar school als jij me toch niet wilt aannemen.'

'Ik ga je niet aannemen als jij niet voldoende onderwijs hebt.'

In 2003 werd dit gesprek gehouden in de Amerikaanse staat Virginia. Het had een heftige discussie kunnen zijn tussen een werkgever en een sollicitant. Misschien werd de werkzoekende wel afgewezen op basis van huidskleur. Of had de werkgever een blik op het cv geworpen en geconcludeerd: onvoldoende onderwijs. De sollicitant was alleen niet zwart, hij was paars. En de twee waren geen echte werkzoekende en werkgever, maar studenten. Ze deden mee aan een experiment van Harvard hoogleraar Roland Fryer en collega's. Hun studie zou laten zien hoe snel een gelijke wereld kan ontsporen als je blind op cijfers stuurt.

In het experiment kregen studenten willekeurig een rol toegedeeld als 'werkgever', 'groene werkzoekende' of 'paarse werkzoekende'. Elke ronde moest een werkzoekende kiezen of zij zou investeren in haar eigen onderwijs of niet. Aan de ene kant had zo'n investering een nadeel: de studenten ontvingen voor hun deelname een vergoeding en het 'onderwijs' zou hen geld kosten. Aan de andere kant hadden ze ook meer kans op een hoge score op de 'test' (die bestond uit een soort gewogen dobbelsteen die vaker in hun voordeel zou beslissen als ze hadden geïnvesteerd in onderwijs) en daarmee maakten ze meer kans op extra geld. Werkgevers wilden namelijk het liefst werkzoekenden met een goede score, want een goed opgeleide werknemer leverde meer geld op. Maar omdat een werkgever alleen de testscore kon zien, was het nooit

100 procent duidelijk of de sollicitant daadwerkelijk onderwijs had genoten. Zo lijkt het experiment op de werkelijkheid: een werkgever weet nooit zeker of een sollicitant geschikt is, maar kan het wel inschatten aan de hand van imperfecte graadmeters zoals schoolcijfers.

In de eerste ronde investeerden paarse werkzoekenden net wat minder geld in onderwijs. Dit had niets te maken met hun paarse kenmerk, want de kleur was willekeurig bepaald. In de volgende ronde konden werkgevers de statistieken inzien. Paarse werkzoekenden, dachten ze, konden ze beter niet hebben. Toen de paarse deelnemers op hun beurt weer zagen dat hun groene collega's vaker werden aangenomen, besloten ze minder te investeren. Want die investering leek niet hun kans op een baan te vergroten.

Het gekke was: iedereen gedroeg zich rationeel. Maar binnen twintig rondes ontstond een vicieuze cirkel, die uitmondde in een extreem ongelijke wereld. 'Ik was verbaasd. De kids waren echt boos', vertelde onderzoeker Fryer aan Tim Harford, die over het experiment schreef in zijn boek *The Logic of Life*. 'De aanvankelijke ongelijkheden kwamen door toeval, maar mensen bleven zich eraan vastklampen en lieten niet meer los.'

De wereld is natuurlijk veel complexer. Maar het experiment illustreert: cijfers zijn zowel gevolg als oorzaak van hoe de wereld eruitziet. En hoe meer cijfers onze wereld gaan beheersen, zoals nu gebeurt met big data, des te meer ze onze wereld zullen veranderen.

Neem *predictive policing*, algoritmes die door politie worden ingezet om uit te vinden wie er mogelijk crimineel is. Amerikaanse cijfers laten een duidelijk verband zien tussen zwarte arme jongemannen en criminaliteit. Op basis van die algoritmes zul je je als politie richten op de wijken en individuen die voldoen aan dit signalement. Het gevolg? Etnisch profileren, waarbij ook veel onschuldige mensen worden aangehouden. En als je bepaalde mensen vaker aanhoudt, dan belanden ze automatisch vaker in de statistieken. De rijke witte criminelen zie je immers over het hoofd, want die vallen buiten je werkgebied. Niet vreemd dan, dat je ook in de volgende statistieken een - misschien nog wel sterker - verband ziet tussen huidskleur en criminaliteit.

Hetzelfde risico loop je bij kredietscores: mensen met bepaalde kenmerken krijgen moeilijker een lening dan anderen, waardoor deze mensen sneller in armoede terecht komen, waardoor ze moeilijker een lening kunnen krijgen, waardoor ze nog sneller arm worden, enzovoorts enzoverder.

De cijfers die de werkelijkheid hadden moeten vangen hebben haar vervangen.

Wat wil je met cijfers bereiken?

Uiteindelijk draaien big data – net als kleine data – om dezelfde vraag: wat wil je met de cijfers bereiken? We zagen eerder dat de Federal Trade Commission concludeerde dat één op de twintig kredietrapporten ernstige fouten bevatte. Consumer Data Industry Association (CDIA), de beroepsvereniging van onder meer kredietbureaus, zag dit als een positief bericht: 95 procent van de consumenten had immers geen last van fouten. Maar is 5 procent nu veel of weinig? Het is maar net wat je beoogt met de scores. Kredietverstrekkers zijn over het algemeen commerciële partijen. Hun doelstelling: winst. Door die bril gezien is 95 procent inderdaad netjes. Of het rechtvaardig is, is voor hen minder belangrijk. De lener is immers niet de klant, maar het product. Zo blijft het opletten. Ook in Nederland worden we uitgebreid gescoord. Kredietbeoordelaars proberen te berekenen of we met geld kunnen omgaan, de Belastingdienst of we fraude plegen, de politie of we een misdaad begaan, verzekeraars of we gezond blijven. Telkens hebben die berekeningen gevolgen voor ons dagelijks leven: je wordt geweigerd voor een lening, je krijgt een aanmaning, je wordt staande gehouden, je moet een hogere premie betalen. En vaak zijn juist die mensen de dupe die toch al een kwetsbare positie in de samenleving hebben.

Big data kunnen de wereld mooier maken. Kijk maar naar Jenipher, die dankzij een lening een beter bestaan kon opbouwen. Maar diezelfde algoritmes die mensen als Jenipher helpen kunnen eeuwenoude ongelijkheden in stand houden en nieuwe creëren. Het zijn dan ook niet de algoritmes die 'goed' of 'slecht' zijn, maar de manier waarop wij ze gebruiken. Daarom is het van levensbelang om mee te praten over de vraag: Waar willen we de algoritmes voor gebruiken?

Algoritmes zullen nooit objectief zijn, hoe betrouwbaar de data ook mogen worden en hoe geavanceerd kunstmatige intelligentie ook zal zijn. Als we deze eigenschap van algoritmes vergeten, dan laten we morele beslissingen over aan mensen die toevallig een talent hebben voor computers. En die al programmerend beslissen over wat goed en fout is.

NOOT

1. <https://www.youtube.com/watch?v=mJ63-bQc9Xg&t=3s>

SANNE BLAUW heeft econometrie gestudeerd aan de Erasmus School of Economics en is gepromoveerd op geluk en inkomensongelijkheid. Ze schrijft voor *De Correspondent*. Dit artikel is gebaseerd op een hoofdstuk uit haar boek *Het bestverkochte boek ooit (met deze titel)* dat onlangs is verschenen bij *De Correspondent*. E-mail: sanne@decorrespondent.nl

GEKUNSTELDE INTELLIGENTIE

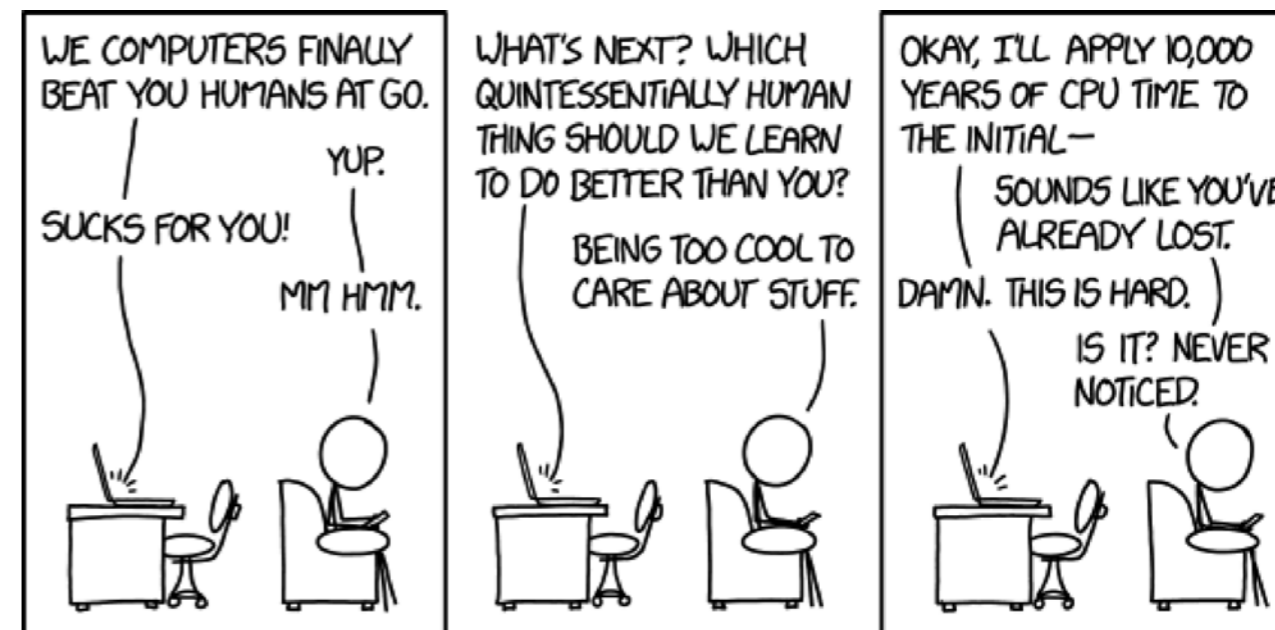
Al sinds mijn jeugd ben ik geïnteresseerd in wat je met wiskunde en computers kunt doen. Wellicht geïnspireerd door stripverhalen als *Storm* die ik als jongen las, waarin de hoofdfiguur van de strip zich moest verweren tegen zelfdenkende machines', een strijd die hij met creativiteit en intuïtie in zijn voordeel wist te beslechten. Die interesse leidde me naar boeken als *Gödel, Escher, Bach* (GEB) van Douglas Hofstadter² waarin de mogelijkheid van kunstmatige intelligentie wordt onderzocht. Het fascineerde me enorm dat een machine, als het algoritme in de machine maar complex genoeg zou zijn, een bewustzijn zou kunnen ontwikkelen en intelligent gedrag zou gaan vertonen. Later kwam daar *The Emperor's New Mind* van Roger Penrose³ bij die juist het tegenovergestelde van Hofstadter beweerde. Toen ik tijdens mijn studie de mogelijkheid had Kunstmatige Intelligentie nader te bestuderen, was de keuze dan ook snel gemaakt. Helaas was het vak minder spannend dan ik verwacht had, de mogelijkheden die Kunstmatige Intelligentie op dat moment te bieden had stonden veraf van wat ik me had voorgesteld. Veel verder dan het oplossen van Rubiks kubus met technieken als *gradient descent* en *backtracking* kwam het niet. De technieken, die in mijn ogen bijna gelijk stonden aan trial en error zoekmethoden, stonden ver af van het leren, redeneren en structureren dat ik verwachtte dat de computer zou laten zien. Nee, Kunstmatige Intelligentie leverde me niet de inzichten en besliskracht die ik met technieken uit de Operations Research en Statistiek kon bereiken.

Inmiddels zijn we 30 jaar verder, Kunstmatige Intelligentie lijkt zichzelf in die periode opnieuw te hebben uitgevonden en is van onderzoeksterrein met weinig potentieel uitgegroeid tot een miljarden business. Het is dé technologie geworden voor internetgiganten als Google, Facebook en Amazon en heeft op vele manieren invloed om ons dagelijks leven. Of het nu het nieuws in je tijdslijn op Facebook, de voorgestelde websites als antwoord op je Google zoekopdracht, het volgende nummer in je Spotify playlist of je volgende date via een dating site is, Kunstmatige Intelligentie lijkt overal de key enabler te

zijn. De verwachtingen over wat Kunstmatige Intelligentie nog meer te bieden heeft zijn dan ook hooggespannen. De vraag is niet langer of, maar wanneer ons wegennet alleen nog autonoom rijdende auto's bevat. De verwachting is dat zelflerende algoritmes steeds krachtiger zullen worden en ons steeds meer werk en beslissingen uit handen gaan nemen. Gaat de computer, zoals Hofstadter in GEB beweert, inderdaad binnenkort een bewustzijn ontwikkelen en intelligent gedrag vertonen?

Hofstadter zelf geeft in zijn artikel 'The Shallowness of Google Translate'⁴, daarop al het antwoord. Hoewel de snelheid waarmee Google Translate zinnen van de ene taal naar de andere taal omzet en de veelheid aan talen waarvoor het dat kan tot de verbeelding spreken, concludeert Hofstadter dat er nog een hele weg te gaan is. De complexe, geneste neurale netwerken waar Google Translate gebruik van maakt behoren tot het beste wat Kunstmatige Intelligentie op dit moment te bieden heeft, maar stellen Google Translate niet in staat te 'begrijpen' waar de tekst die vertaald moet worden over gaat. Gevolg is een ruwe, vaak woord voor woord, vertaling van de zin waardoor de dieperliggende betekenis verloren gaat. Geen probleem als je op zoek bent naar de vertaling van dat ene woord voor in je artikel of email, maar ondermaats als ook de diepere betekenis van de zin van belang is. Google Translate staat daarmee nog ver af van wat je een intelligent algoritme zou kunnen noemen. Het duurt nog wel even voordat we algoritmische intelligentie bereiken volgens Hofstadter.

Het recente succes van Kunstmatige Intelligentie is voor een belangrijk deel terug te voeren op de herontdekking van een algoritme dat ooit ontwikkeld is om de stuwkracht van de raketmotoren van de Apollo landingsmodule⁵ tijdens maanlandingen te optimaliseren, Backpropagation. Gecombineerd met een grote hoeveelheid data en krachtige computers stelt het algoritme ons in staat systemen te maken die ons nadoen of zelfs taken beter dan ons doen. Denk aan het visueel herkennen van objecten, spraakherkenning, automatisch vertalen, robo-



Figuur 1. Computers vs Humans. Illustratie: <https://kkcd.com/1875/> (CC)

tica en het spelen van spellen. Recent nog versloeg DeepMind's AlphaGO algoritme de wereldkampioen GO⁶, een strategisch bordspel dat qua complexiteit het schaakspel ver overstijgt. Het feit dat die systemen ons 'verslaan' doet velen van ons aannemen dat die systemen ook daadwerkelijk het spel begrijpen en dus een vorm van intelligentie hebben ontwikkeld. Niets is minder waar. Als we net als Dorothy uit de *Wizard of Oz* achter het gordijn dat door slimme marketing wordt opgeworpen kijken zien we systemen die op basis van patroonherkenning (*supervised learning*) of via een feedbackmechanisme (*reinforcement learning*) 'leren' een zo goed mogelijke voorspelling te geven van het antwoord. Dit maakt die systemen niet minder nuttig, maar van intelligent gedrag of begrip is echter geen sprake. Ten opzicht van de trial en error zoekmethoden van 30 jaar geleden is er dus niet veel veranderd, het zoeken gaat alleen veel sneller.

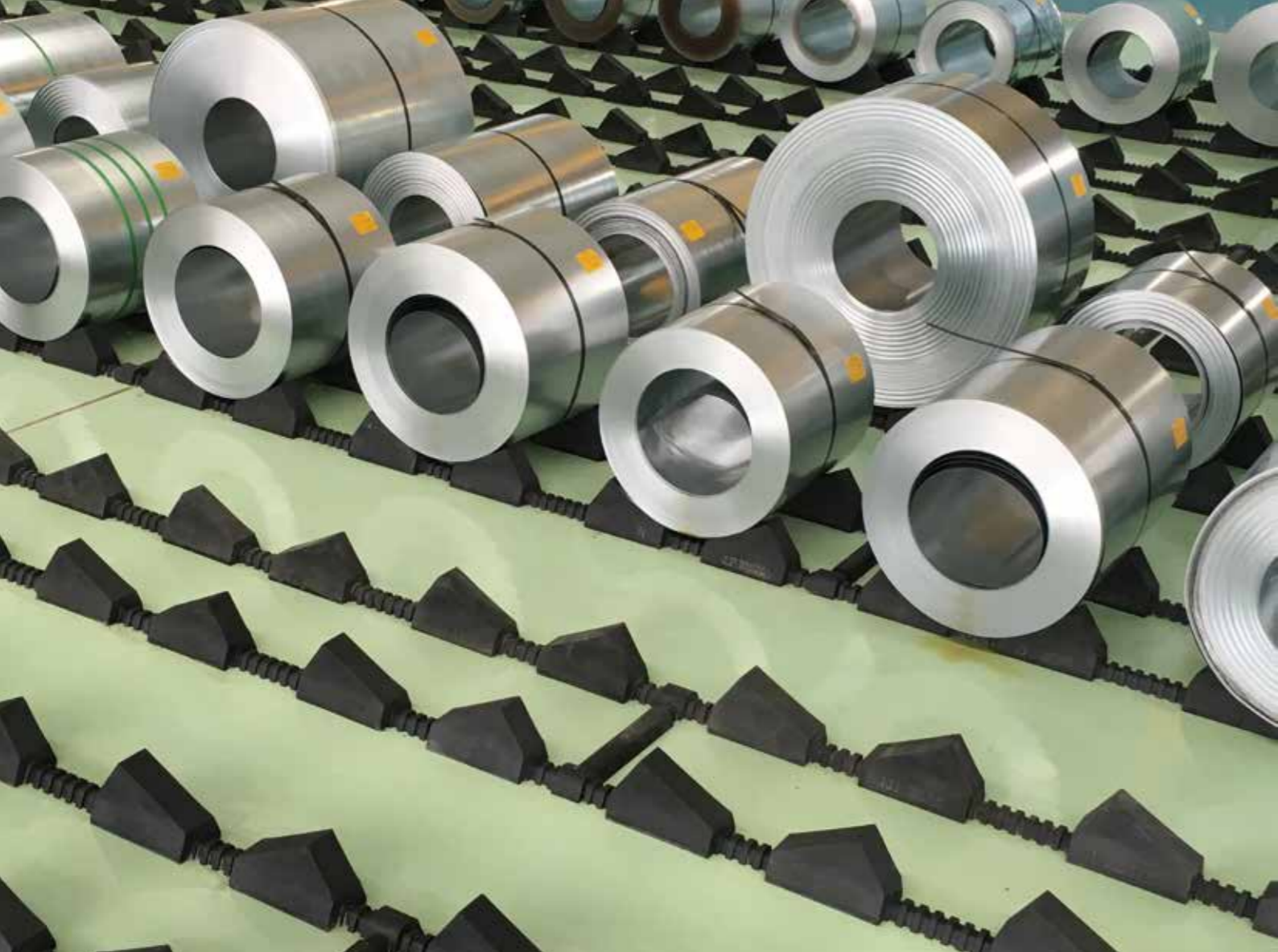
De hoge verwachtingen over wat Kunstmatige Intelligentie kan, leiden ertoe dat bedrijven het als de oplossing zien voor bijna ieder vraagstuk. Het 'zelflerend' vermogen suggereert dat het algoritme autonoom kan opereren en zich kan aanpassen aan veranderende omstandigheden. Iedereen kent inmiddels de filmpjes van algoritmes die videogames⁷ leren spelen en in no time onze high scores verpulveren. Wat mensen zich niet realiseren is dat vele uren aan computersimulatie nodig zijn waarin de (*reinforcement learning*) algoritmes 'leren' het spel te spelen. Yann Lecun, Chief AI scientist van Facebook, twijfelt dan ook aan het praktische nut van dit soort technieken. Hij geeft bijvoorbeeld aan dat zo'n algoritme een auto 50.000 keer van een klif⁸ afstuurt voordat het door heeft

dat het niet wenselijk is. Er bestaan in de praktijk maar weinig omgevingen en vraagstukken waarin er de ruimte is om keer op keer iets te proberen om het algoritme van de bereikte resultaten te laten leren. In Online Marketing of bij services als Google Translate kan ik me dat voostellen, maar moet er niet aan denken als dergelijke technieken gebruikt gaan worden om algoritmes te ontwikkelen die supply chains aan sturen, verkeerstromen managen of medische diagnoses stellen. Specialisten als Yann Lecun geven aan dat het nog wel enkele decennia duurt voordat deze technieken volwassen genoeg zijn om besluitvorming in de echte wereld te ondersteunen. Gelukkig is er een goed alternatief.

NOTEN

- https://nl.wikipedia.org/wiki/De_laatste_vechter
- https://en.wikipedia.org/wiki/G%C3%B6del,_Escher,_Bach
- https://en.wikipedia.org/wiki/The_Emperor%27s_New_Mind
- <https://www.theatlantic.com/technology/archive/2018/01/the-shallowness-of-google-translate/551570/>
- Klumpp A.R. (1971). *Apollo Guidance, Navigation, and Control: Apollo Lunar-Descent Guidance* (Charles Stark Draper Lab, TR R-695). Cambridge, MA: Massachusetts Institute of Technology.
- <https://www.theguardian.com/technology/2017/may/23/alphago-google-ai-beats-ke-jie-china-go>
- <https://youtu.be/TmPftjtdgg> DeepMind speelt Breakout
- <https://youtu.be/otEhw5t6rhc?t=1258>, Yann LeCun - Power & Limits of Deep Learning

JOHN POPPELAARS is Practice Leader Advanced Analytics bij BearingPoint.
E-mail: john.poppelaars@bearingpoint.com



Rollen staal en cradles. Foto: Lankhorst Mouldings

OPTIMAAL ONTWERP VOOR HET BESCHIKBARE VLOEROPPervlak DOOR EEN SLIMME COMBINATIE VAN OPTIMALISATIE EN SIMULATIE

DAAN MERKESTEIN & STEVO AKKERMAN

Grote investeringsvraagstukken in het bedrijfsleven gaan vaak gepaard met grote onzekerheden. Voor logistieke projecten hebben deze onzekerheden veelal betrekking tot de afhankelijkheden tussen verschillende onderdelen in de logistiek. Ook zijn aankomstpatronen van entiteiten in de praktijk niet mooi volgens een exponentiële verdeling, is er variabiliteit over de tijd in de processen en heeft men te maken met pieken en stringen van apparatuur.

Een model van het logistieke proces kan duidelijkheid scheppen in deze onzekerheid door het gehele proces, met alle afhankelijkheden en relevante onderdelen van de logistiek, te simuleren.

Een mooi voorbeeld van het toepassen van een simulatiemodel is de bouw van de nieuwe opslaghal door staalproducent Tata Steel in IJmuiden¹. In samenwerking met Systems Navigator² is Tata Steel begin 2018 begon-

nen met de bouw van een simulatiemodel om de logistiek van de nieuw te bouwen zogeheten LA-hal te analyseren en te optimaliseren.

Slim indelen

Tata Steel is continu bezig met het verbeteren van zijn (logistieke) processen. De bouw van een nieuwe opslaghal voor rollen hoogwaardig staal, de LA-hal, is onderdeel van dit verbeterproces. De LA-hal zal gebruikt worden voor de tijdelijke opslag van staalrollen. De afgewerkte rollen komen binnen in de hal via treinen en worden opgeslagen in de hal tot er een uitgaand transport beschikbaar is naar de klant. Om het aantal manuele handelingen te minimaliseren worden de rollen binnen de hal vervoerd door twee automatische kranen, die worden aangestuurd door een *warehouse* managementsysteem.

Voor de efficiëntie en de opslagcapaciteit van de hal is de indeling erg belangrijk. Deze halindeling, een zogeheten *cradle set-up*, is het vraagstuk waar de focus op ligt in dit artikel. Een *cradle* is de wieg waar een staalrol op staat in de opslaghal en zorgt ervoor dat de rol niet wegrolt. Een bepaalde *cradle* kan alleen gebruikt worden door rollen die aan bepaalde voorwaarden voldoen qua breedte, diameter en gewicht. Het is niet verstandig om elke *cradle* voor elke rol geschikt te maken, aangezien elk *cradle* dan geschikt moet zijn voor de grootste rollen, waardoor ruimte verloren gaat als er een kleine rol op gelegd wordt. Er mogen, onder bepaalde voorwaarden betreffende de breedte, het gewicht en de diameter met de te stapelen rol, ook rollen gestapeld worden op de rollen die op de *cradles* liggen. Hierbij is het een voordeel als er rollen met ongeveer dezelfde dimensies naast elkaar liggen, want dit zorgt voor een grotere kans op een stapelmogelijkheid voor een derde rol bovenop/tussen de twee rollen op de *cradles*.

Omdat het aanpassen van de halindeling inhoudt dat de hele hal leeggehaald moet worden en alle operaties stilgelegd moeten worden om veiligheidsredenen, is het belangrijk om de halindeling zo slim mogelijk op te zetten. Hierbij moet ook gekeken worden naar mogelijke veranderingen in de karakteristieken van de rollen die opgeslagen zullen worden en de variabiliteit van de inkomende stroom rollen. Er moet gestreefd worden naar een halindeling waarin zoveel mogelijke potentiële binnenkomende stromen met verschillende types rollen kunnen worden voorzien.

Waarom simuleren?

Er zijn vele factoren die de kwaliteit van een halindeling beïnvloeden. Denk bijvoorbeeld aan de volgorde van binnenkomst van de rollen of de beschikbaarheid van treinen en medewerkers in de hal. Maar ook de variatie in het soort rollen dat de hal binnenstroomt, de modaliteit waarmee deze rollen weer vertrekken en de prioritering van de kranen. Daarnaast wordt de optimale halindeling bereikt door de juiste keuzes te maken in de afmetingen van de verschillende *cradles*, het aantal van deze *cradles*, de plaatsing van *cradles* in de hal en in welke volgorde/batches de rollen de hal binnenkomen. Een probleem met dermate veel complexe variabelen en factoren optimaliseren is een schier onmogelijke taak.

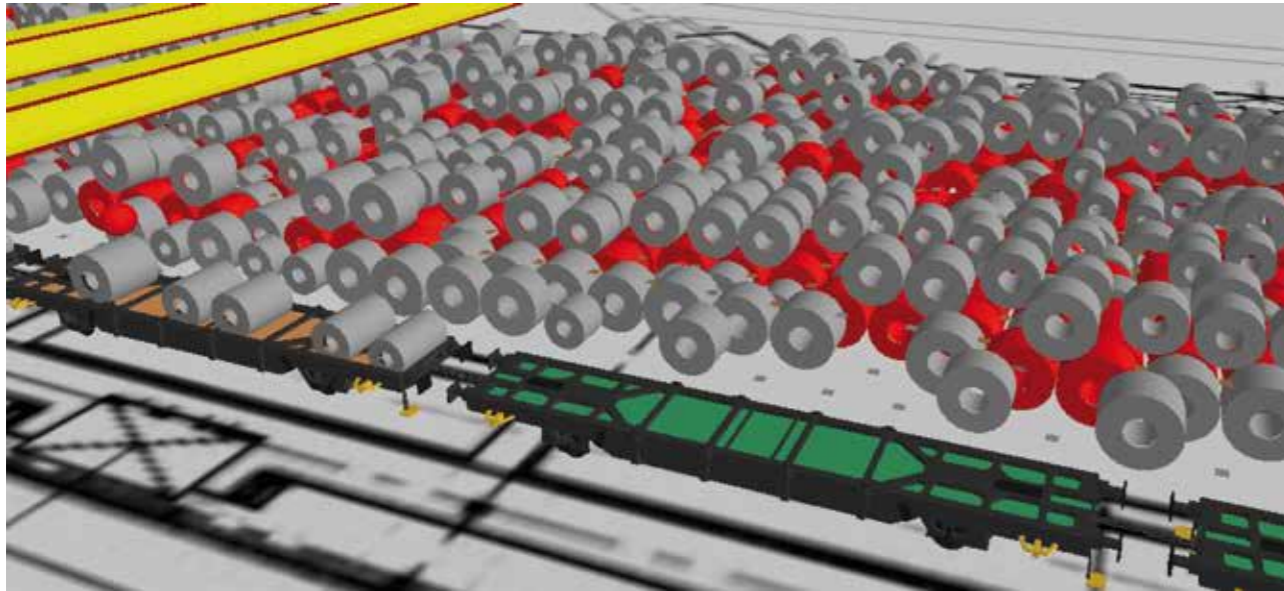
Daarom is er gekozen voor een aanpak met behulp van discrete event-simulatie. De oplossingsruimte van mogelijke halindelingen is doorzocht door telkens een set halindelingen met specifieke karakteristieken op te stellen en deze te vergelijken middels simulatie. In een simulatiemodel zijn de bovenstaande onzekere factoren goed mee te nemen. Uit een analyse van deze simulaties kan een conclusie worden getrokken waarmee de oplossingsruimte verkleind wordt.

Het simulatiemodel

Het simulatiemodel voor de LA-hal is gemaakt in het softwarepakket Simio. Simio werkt op basis van discrete event simulatie en is object-georiënteerd, waardoor de intelligente objecten met elkaar kunnen communiceren en op die manier hun informatie uitwisselen. Als bijvoorbeeld een rol opgepakt moet worden, communiceert hij met de kranen zijn oppak-verzoek. De kranen zullen daarna zelf onderling bepalen welke kraan de rol gaat oppakken. Er is dus geen overkoepelend proces nodig dat al deze aansturing doet. Simio werkt met een 3D-omgeving waar echte afstanden, tijden en snelheden worden meegenomen en er zo realistische object bewegingen en transporttijden worden gewaarborgd. Een voorbeeld van een 3D-animatie uit Simio wordt weergegeven in figuur 1.

Zonder input geen output

Wat voor elk simulatie model geldt is dat de output niets waard is als de input niet correct en volledig is (*Garbage*



Figuur 1. 3D-animatie uit Simio

in, garbage out). Voor veel investeringsprojecten zijn er historische data aanwezig van een proces dat enigszins lijkt op het proces dat gesimuleerd wordt. Zo kunnen bijvoorbeeld historische scheepsaankomsten bij een olieterminal een prima representatie zijn voor de toekomstige scheepsaankomsten als deze terminal investeert in een extra kade of tank.

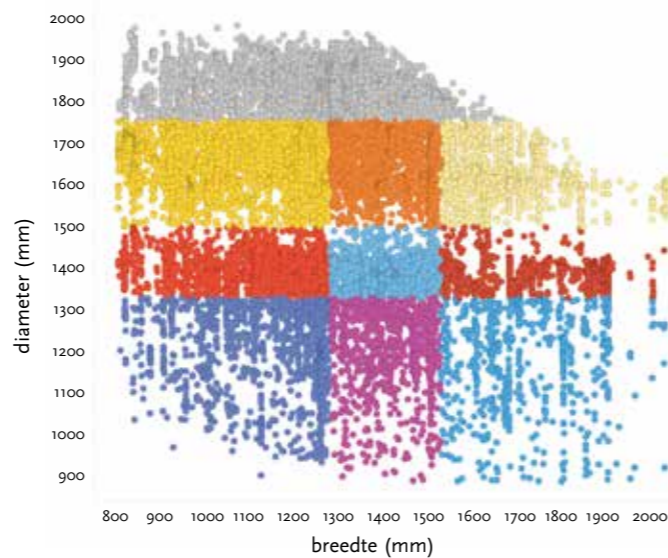
Voor de LA-hal is er gekeken naar een dataset van alle rollen die er in 2017 ergens op het terrein van Tata Steel in opslag hebben gelegen. Met deze dataset als basis is er door het projectteam gekeken naar de rollen die potentieel naar de LA-hal *gegaan zouden zijn* als de hal er in 2017 al geweest was. Op deze manier is er een representatieve set van 90.000 rollen samengesteld. Voor de overige parameters in het model is er onder andere gekeken naar technische specificaties van de hal, treinbeladingen, truckbeladingen en bijbehorende restricties en de vaarschema's richting de klanten van Tata Steel.

Het bepalen van de cradle-types

De cradles worden gemonteerd op rails met inkepingen waarin ze worden vastgezet. Hierdoor zijn er een bepaald aantal mogelijke cradle-types mogelijk, afhankelijk van het aantal inkepingen dat er tussen de cradle in wordt gelaten. De afstand tussen de twee rails bepaalt de mogelijke breedte van de cradle, hierin is meer vrijheid.

Uit de data-analyse kwam de scatterplot in figuur 2 van de breedte/diameter van de rollen uit de dataset: De

eerste stap was om vast te stellen welke opdeling in de afstelling van de breedte en diameter van de cradles het beste presteert. Een bepaalde afstelling van een groep cradles leidt tot een breedte- en diametergroep. Uiteindelijk is er gekozen om 6 verschillende combinaties van breedte/diameter afstellingen voor de cradles te analyseren. Deze combinaties onderscheiden zich van elkaar door het aantal groepen waarin de breedtes en de diameters verdeeld zijn.



Figuur 2. Scatterplot van de breedte/diameter van de rollen, met per kleur één cradle-type

Voor elke van deze 6 combinaties werden de optimale groepsafmetingen bepaald met behulp van het verwachte voorraadniveau van deze groepen. Door het minimaliseren van de lost space (de hoeveelheid die een rol kleiner is dan de grootste rol die op het cradle zou passen), werden de optimale afmetingen van de groepen, gegeven het aantal groepen, gevonden.

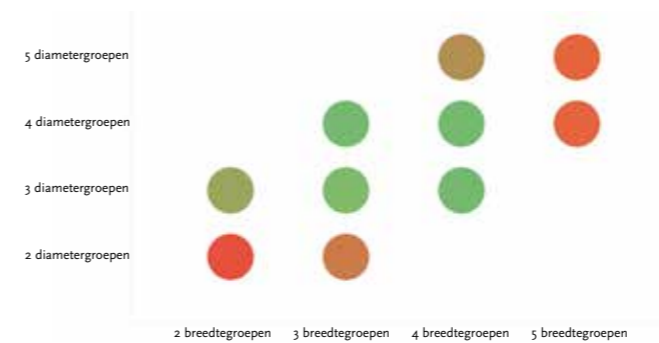
Om het verwachte voorraadniveau per breedte/diameter-afstelling te bepalen wordt er bij de analyse rekening gehouden met het aantal aankomsten en de verblijfsduur van de rollen in de opslag. Volgens de wet van Little geeft het product van deze twee waarden het verwachte gemiddelde benodigde voorraadniveau per cradle type. De uitkomsten van de simulatie analyseren we op verschillende gekozen prestatie-indicatoren, zoals het aantal rollen waar geen plek voor was, de gemiddelde halbelegging en de efficiëntie van de kranen. De resultaten worden weergegeven in figuur 3. Uit de analyse bleek dat een combinatie van 4 breedtegroepen x 3 diametergroepen de beste halindeling gaf. Dit resulteerde in 12 verschillende cradletypes.

Door deze strategie, het analyseren van de geoptimaliseerde halindelingen met specifieke cradletype en doorzet karakteristieken met behulp van het simulatiemodel, nog een aantal keer toe te passen, is de oplossingsruimte telkens te verkleinen. Na zo'n 10 iteraties van dit proces is er een optimaliseerde oplossing gevonden.

Voor het LA-hal-project is het advies opgeleverd in de vorm van een geoptimaliseerde cradle setup lay-out, welke nu een-op-een geplaatst kan worden door de leverancier.

Toegevoegde waarde

Om de toegevoegde waarde van het simulatiemodel te toetsen is de initiële cradle setup, gemaakt op basis van common sense en ervaring, ook gesimuleerd en geanaly-



Figuur 3. Prestatie van verschillende combinaties van breedte- en diametergroepen

seerd. Uit de resultaten kwam naar voren dat er in de geoptimaliseerde lay-out 50 cradles meer in de hal passen ten opzichte van de initiële setup. Door de geoptimaliseerde lay-out, met goede stapelmogelijkheden voor de rollen, kunnen nu ongeveer 7,5% meer rollen opgeslagen worden in de hal. Dit alles verhoogt de piek-opslagcapaciteit met 13,5%. Maar nog belangrijker; de uiteindelijke lay-out is ook beter in staat om de variabiliteit in aankomsten op te vangen en houdt rekening met een veranderend rolaanbod in de toekomst.

Conclusie

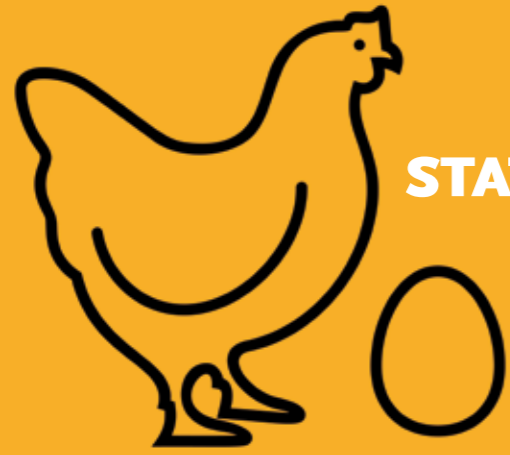
Het LA-hal-project heeft bewezen dat het gebruik van een simulatiemodel een goed middel is om bij investeringsprojecten van complexe logistieke systemen antwoorden en adviezen te verkrijgen, nog vóór dat de bouw van het project is gestart. Omdat het model generiek is opgezet, kan het in de toekomst worden gebruikt om andere opslaghallen binnen Tata Steel (opnieuw) in te delen. Momenteel wordt het model verder ontwikkeld zodat het gebruikt kan worden voor de ondersteuning van operationele planningsvraagstukken, zoals het toekennen van opstelplaatsen voor inkomende treinen, zetten van prioriteiten voor taken, de inzet van operationele medewerkers, alsmede de analyse van ruimtegebruik. Vergelijkbare studies zijn gedaan voor de opslag van groente en fruit, voor contactlenzen en voor een distributiecentrum voor postpakketten.

DAAN MERKESTEIN, afgestudeerd aan de VU (Econometrics & Operations Research) is consultant bij Systems Navigator in Delft. E-mail: daan.merkestein@systemsnavigator.com

STEVO AKKERMAN, afgestudeerd aan de VU (Econometrics & Operations Research) is manager logistieke projecten bij Tata Steel in IJmuiden. E-mail: stevo.akkerman@tatasteel.eu.com

SYSTEMS NAVIGATOR is een software consultancy bedrijf uit Delft en is wereldleider op het gebied van geavanceerde beslissingsondersteunende technologie. Systems Navigator helpt zijn klanten bij het nemen van betere beslissingen omtrent investeringen, onder andere door middel van simulatie modellen die de impact van veranderingen demonstren.

TATA STEEL produceert, fabriceert en distribueert hoge kwaliteit staal voor producten in verscheidene industrieën, zoals automobiel, verpakkingen en bouw. Bij Tata Steel in IJmuiden werken ruim 9.000 werknemers aan een jaarlijkse productie van bijna 7 miljoen ton staal.



STATISTIEK EN CAUSALITEIT

voortschrijdende liaison of moeizame samenspraak

RICHARD STARMANS

Doorloopt men de ideeëngeschiedenis vanaf het prille begin der pre-socratische natuurfilosofen tot aan het tijdperk van data science en big data, dan stuit men onvermijdelijk op het hardnekkige en complexe probleem van de causaliteit. Voor sommigen betreft dit een volstrekt obsoleet begrip, een ten onrechte gekoesterd relict uit langvervlogen tijden. Zo vertrouwen eigentijdse sceptici in het kielzog van Google's onderzoeksdirecteur Peter Norvig liever op 'the unreasonable effectiveness of data' (Havelly, 2011) of stellen vast dat causaliteit definitief van zijn voetstuk als primaire bron van betekenis is gestoten (Mayer Schönberger, 2013). Sommigen proclameren zelfs *The End of Theory* en belijden dat 'correlation supersedes causation' (Anderson, 2009) of zij vestigen hun hoop op *The Master Algorithm* (Domingos, 2016). Anderen daarentegen beschouwen causaliteit nu juist als 'the most immediate and vital element of the world' (Mumford, 2011), 'the cement of the universe' (Mackie, 1980) of typeren causale relaties als 'the fundamental building blocks both of reality and of human understanding of that reality' (Pearl, 2000). Zij bekennen zich daarmee ontegenzeggelijk tot

een eigentijdse vorm van metafysisch realisme. De AI-wetenschapper en Turing-award-laureaat Judea Pearl (1936), gaat nog een stap verder. In een recente publicatie stelt hij onomwonden dat zich momenteel een 'Causal Revolution' voltrekt, die de wereld zal veroveren middels een 'new science of cause and effect' gebaseerd op een 'causal mathematical language' (Pearl, 2018). Zoals wel vaker blijven meer gematigde en genuanceerde stellingnames enigszins onderbelicht wanneer krachtige metaforen en retorisch geweld in overmaat aanwezig zijn en pathos en overdrijving niet worden geschuwd. In dit essay belichten we enkele aspecten van causaliteit, dat vanuit historisch-wijsgerig perspectief een weerbarstig begrip blijkt en welhaast permanent in een staat van crisis lijkt te hebben verkeerd en nog verkeert. We gaan eerst kort in op de reikwijdte van het begrip. Vervolgens komt summier de genealogie van de wetenschappelijke en filosofische kritiek op c.q. crisis van de causaliteit aan de orde. Tot slot schetsen we de opkomst van de statistiek en haar rol bij pogingen de crises te bezweren door een soort voortschrijdende liaison met causaliteit aan te gaan. Deze

heeft evenwel niet kunnen voorkomen dat zich vandaag de dag een scheiding der geesten heeft voltrokken, waarbij een moeizame samenspraak overblijft als surrogaat voor eenheid en eensgezindheid.

Cultuur, taal en evolutie

Oorzaak en gevolg, het principe van oorzakelijkheid en causale relaties manifesteren zich in een veelvoudigheid aan verschijningsvormen: causale processen en mechanismen, uitspraken en argumenten, theorieën en modellen, redeneringen en inferenties. Dit geschiedt tegen de achtergrond van concepties als logische noodzakelijkheid, fysische noodzakelijkheid of determinisme en 'duale', evenzeer weerbarstige noties als contingentie, toeval, waarschijnlijkheid, onzekerheid, onbepaaldheid, vrije wil en verantwoordelijkheid. Al deze concepten constitueren het begrip causaliteit, zijn ervan afgeleid of worden er steevast historisch mee in verband gebracht. Causaliteit speelt uiteraard niet alleen een rol in voorwetenschappelijke theorieën, metafysica, kennisleer en wetenschap, maar kent een nog imposantere cultuurhistorische reikwijdte. Wie geïnteresseerd is in causale allusies in de narratieve traditie (variërend van mythen en volksverhalen tot Dostojevski's *Schuld en Boete* en Nabokov's *Lolita*), in sacrale religieuze teksten, maar evenzeer in kwantummechanica, genetica en psychoanalyse kan te rade gaan bij Stephen Kerns onvolprezen studie *A Cultural History of Causality* (2004).

In weerwil van eeuwenoude preoccupaties van erudiete denkers met causaliteit, lijkt het allereerst een intuïtief concept, een common sense notie of 'natuurlijke' categorie (in de aristotelische of kantiaanse zin van het woord), die we als intelligente organismen voortdurend gebruiken om ons te oriënteren in de werkelijkheid, aan te passen aan de toevalligheden en grilligheid van het bestaan en zo adaptief te zijn in de struggle for life. Maar ook om onze ervaringen te verklaren en onszelf, onze gesitueerdheid en de contingenties van het menselijk bestaan te begrijpen. Bijgevolg is oorzakelijkheid voortdurend manifest in het alledaagse taalgebruik. Expliciet gebeurt dit bij voorbeeld bij uiteenlopende voegwoorden (omdat, doordat, opdat, zodat, hoewel, toch, maar, ondanks, desalniettemin, et cetera) en talloze 'causale' werkwoorden (veroorzaken, induceren, leiden tot, bewerkstelligen, genereren, beïnvloeden, produceren, et cetera). Impliciet

causaal zijn feitelijk alle transitieve werkwoorden. De natuurlijke interpretatie ervan vooronderstelt een subject of handelende agens c.q. abstracte entiteit die een verandering teweegbrengt bij een object c.q. in de werkelijkheid, bijvoorbeeld 'Jan eet de sandwich'. Ook de bijbehorende vraag is causaal: 'Wie eet de sandwich?' Antwoord: de entiteit/actor die een *interventie* pleegt en een *mechanisme* in werking zet waardoor de sandwich betrokken raakt in een *proces* van gegeten worden, daarbij een *verandering* ondergaat en uiteindelijk geraakt in de *eindtoestand* van 'opgegeten zijn' en 'geen sandwich meer zijn', een zichtbaar *verschil* met de *begintoestand* van 'nog niet opgegeten zijn' en 'nog wel een sandwich zijn'. Transitieve werkwoorden leiden tot causale vraagzinnen. Bijgevolg zijn ook vele wetenschappelijke en alledaagse vraagstellingen intrinsiek causaal.

Gelet op de alomtegenwoordigheid van causaliteit in het menselijk handelen en de zeer diverse verankering ervan in de taal lijkt een linguïstisch zoeklicht op causaliteit wenselijk om een compleet beeld ervan te verwerven. Deze linguïstische invalhoek is manifest in *ordinary language philosophy* (*meaning is use*, Wittgensteins taalspelen) en *speechact theory*, maar ook in de zogenaamde informele logica, argumentatietheorie, *critical thinking movement* en *discourse analysis*, die vooral de laatste veertig jaar opgang hebben gemaakt. Kortom, een down-to-earth-benadering, die bovendien een weldadig antidotum vormt tegen de soms zwaarwegende metafysische of kennistheoretische benaderingen van causaliteit.

Daarnaast heeft causaliteit zowel in de wijsbegeerte (metafysica en kennisleer) als in de wetenschappen een merkwaardig ontwikkelingstraject doorlopen; de persistentie van het gebruik van de term door de eeuwen heen staat in fors contrast met de talloze conceptuele verschuivingen die het onderging. Het kameleontische en bovenal *context-gevoelige* karakter van het begrip blijkt als we handboeken metafysica en epistemologie tot medio 19e eeuw overzien. Steevast worden daarin uiteenlopende denkers als Lucretius (en andere atomisten), Chrysippos (en andere stoïci), Plato, Aristoteles, Thomas van Aquino, Bacon, Descartes, Galilei, Spinoza, Hobbes, Newton, Leibniz, Locke, Hume, Kant, Stuart Mill en Charles Sanders Peirce als erflaters of boegbeelden van het denken over causaliteit beschouwd. Die opmerkelijke eensgezindheid lijkt weliswaar enige canonvorming mogelijk te maken, maar relevanter is dat vele van bovengenoemde denkers het begrip causaliteit stipulatief definieerden en

– zoals het goede filosofen betaamt – een intellectuele vadermoord pleegden door radicaal met de traditie te breken of in ieder geval nauwelijks voort te bouwen op illustere voorgangers. In feite werd causaliteit keer op keer opnieuw uitgevonden, juist omdat elke conceptualisering nauw verbonden was met de metafysische en epistemische standpunten van de betreffende denker (Starmans, 2018a). Wanneer een bepaalde wijsgerige positie onder druk kwam te staan, geraakte het bijbehorende causaliteitsbegrip als vanzelfsprekend ook in het beklaagdenbankje. Omdat theorieën en standpunten in de filosofie zelden worden gefalsificeerd en vervangen, zijn veel onderling tegenstrijdige visies naast elkaar blijven bestaan. Deze merkwaardige ‘evolutie’ blijft echter zichtbaar in vele hedendaagse benaderingen en perspectieven: regulariteitstheorieën, *counterfactuals*, interventionist-benaderingen, dispositionele, actor- en procesgeoriënteerde benaderingen, mechanistische visies, *difference-making*, *potential outcomes*, instrumentele variabelen, Bayesianisme, et cetera (Russo, 2016). Zoals uiteengezet in (Starmans, 2018b) zijn vrijwel al deze hedendaagse invalshoeken terug te voeren tot voornoemde illustere denkers.

Daarmee lijkt één historische bron van de kritiek en crisis geïdentificeerd, maar voor een begrip van de huidige situatie kan wellicht beter worden gezocht in de meer recente geschiedenis sedert de tweede helft van de 19e eeuw, toen vele wetenschapsgebieden verzelfstandigden en de probabilistische revolutie zich begon te voltrekken (Krüger, 1987; Gigerenzer, 1989). In de hedendaagse wetenschap en filosofie lijkt een pluralistische visie op causaliteit onvermijdelijk, zelfs als we ons in het kader van dit essay strikt beperken tot de in de methodologie dominante *probabilistische theorieën* over causaliteit. Deze kennen diverse verschijningsvormen die gemeen hebben dat een gepostuleerd of vermeend causaal verband c.q. de oorzaak-gevolg relatie wordt geanalyseerd en formeel gerepresenteerd met behulp van inzichten uit kansrekening en statistiek als formalismen om te redeneren met onzekere kennis: kansuitspraken, voorwaardelijke kansen, conditionele afhankelijkheden, likelihood-ratio's maar ook bijvoorbeeld schattingstheorie uit de inferentiële, parametrische statistiek. Probabilistische theorieën over causaliteit kennen inmiddels ook een respectabele wijsgerige traditie, die teruggaat tot de logisch-positivist Hans Reichenbach (1891–1953), de statisticus I.J. Good (1916–2009) en de filosoof Patrick Suppes (1922–2014) en hebben evenzeer hun weg gevonden in de moderne kennisleer (Williamson, 2009).

De geest, scepsis en eliminatie

De grote rol van causaliteit in het menselijk handelen en de wijsbegeerte laat onverlet dat de notie in de loop van de ideeëngeschiedenis problematisch bleek en regelmatig in een crisis geraakte. In de 18e eeuw was het vooral David Hume (1811–1776), die de zaak op scherp zette en vanuit zijn concept-empirisme de (metafysische) status van causaliteit betwistte. Causaliteit bestaat volgens Hume niet in de werkelijkheid. Daar is slechts constante opeenvolging van verschijnselen, die in onze gedachten worden geassocieerd door gewenning, vervolgens geprojecteerd in de werkelijkheid, waardoor onze ‘intuïtie’ van noodzakelijkheid wordt verklaard. Na het scepticisme van Hume was het in de 18e eeuw vooral Kant (1724–1804) die de notie in ere trachtte te herstellen en een verzoening tussen metafysica en natuurwetenschap nastreefde door causaliteit als *synthetisch a priori* kennis te beschouwen en te benoemen als een fundamentele categorie van het menselijk kenvermogen, die een voorwaarde vormt voor waarnemingen en natuurwetenschappelijke kennis. Met de anti-kantiaanse wetenschapsopvatting van de logisch positivisten in de 20e eeuw bleek ook deze conceptie van causaliteit problematisch, niet in de laatste plaats door de successen van de kwantummechanica, waarin causaliteit niet langer als een kantiaanse bouwsteen van de werkelijkheid werd beschouwd. In de 19e eeuw waren het vervolgens onder meer de fysisch/filosoof Ernst Mach (1838–1916) en de statisticus/filosoof Karl Pearson (1857–1936) die zich manifesteerden als prominente anti-causalisten. De belangrijkste criticaster diende zich evenwel aan in de 20e eeuw. Het was Bertrand Russell (1872–1970) die in 1913 nota bene in *The Proceedings of The Aristotelian Society* zijn artikel ‘On the notion of Cause’ publiceerde en stelde dat ‘*The law of causality, I believe, like much that passes muster among philosophers, is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm.*’

Belangrijk is dat deze houding ten opzichte van causaliteit niet op zichzelf staat. In de loop der tijden zijn vele concepten en begrippen en daarmee verbonden filosofische thema's in diskrediet geraakt of hebben in het gunstigste geval een specifieke abstracte of wiskundige duiding gekregen. Dat geldt onder meer voor noties als ruimte, tijd, beweging, massa, maar vooral voor begrippen als betekenis, intentionaliteit, zingeving, geest, vrije wil, bewustzijn en persoonlijke identiteit. Volgens sommige wetenschappers en naturalistische filosofen zullen

alle wijsgerige problemen uiteindelijk door de wetenschap worden ontrafeld en ontraadseld. Is het probleem goed gedefinieerd, dan zal het worden geanalyseerd en vervolgens worden opgelost. Is het niet goed gedefinieerd, dan wordt het als pseudoprobleem of als betekenisloos terzijde geschoven. Wijsgerige reflectie kan tijdens dat proces hooguit leiden tot een soort voorwetenschappelijke theorieën, die enige verklarende kracht of praktisch nut kunnen hebben, maar die uiteindelijk vervangen zullen worden door ware wetenschappelijke kennis. De concepten en begrippen die bij deze wijsgerige reflectie een rol spelen zullen dan doorgaans ook het veld moeten ruimen. Deze tendens om wetenschap en haar taal te zuiveren van metafysische begrippen, common sense noties, natuurlijke categorieën en alledaagse ervaringen kent een berucht hoogtepunt in de opvattingen van de filosoof en neurowetenschapper Paul Churchland (1942), die radicaal wil afrekenen met een traditie die soms pejoratief wordt aangeduid als ‘folk psychology’. Mensen trachten het gedrag van zichzelf en anderen te begrijpen, te verklaren en te voorspellen in termen van *causaal* relevante factoren, zoals motieven, intenties, overtuigingen en verplichtingen. Churchland pleit voor een radicaal ‘eliminatief materialisme’ inzake deze propositionele attitudes, en stelt dat ‘folk psychology’ met inbegrip van de notie van het bewustzijn, de menselijke geest en haar interne processen volkomen verkeerd benadert. Evenzeer betreurt hij de preoccupaties van filosofen met taal en de veronderstelde cruciale betekenis ervan voor het denken. Ontwikkelingen in de neurowetenschappen zullen volgens Churchland leiden tot eliminatie van deze ‘dwalingen’, die hij even relevant acht voor de wetenschap als de 18e-eeuwse flogistontheorie van Stahl voor de moderne scheikunde, of middeleeuwse opvattingen over hekserij voor de hedendaagse psychologie.

De extreme visie van Churchland past in een langdurige ontwikkeling in de ideeëngeschiedenis die bij de Presocratici een aanvang nam en een hoogtepunt bereikte in de hedendaagse naturalistische/fysicalistische epistemologie. Een gevolg hiervan is dat het huidige wereldbeeld veel van haar aanschouwelijkheid heeft verloren. Enerzijds is er de alledaagse, vertrouwde leefwereld der verschijnselen, met haar ervaringen (percepties, impressies), voorstellingen en ideeën en met haar (gepostuleerde) concrete objecten. Anderzijds is er het wetenschappelijke wereldbeeld met zijn abstracte, vaak wiskundige modellen, representaties van de ‘echte’ werkelijkheid, die verondersteld wordt achter deze ervaringen schuil te

gaan en die verondersteld wordt deze te veroorzaken c.q. te verklaren. Niet alleen lijken de ervaringen, de verschijnselen zoals die zich aan ons voordoen, geen betrouwbare basis om wetenschappelijke theorieën op te baseren, ook de intuïtieve concepten en natuurlijke categorieën lijken weinig gemeen te hebben met de onderliggende mechanismen, abstracte principes en wetten, die de ‘echte’ wereld regeren, zoals beschreven door middel van de taal en de nomenclatuur van de wetenschap. Het moge duidelijk zijn dat deze visie niet te rijmen valt met de linguïstische invalhoek die hierboven werd besproken. Ook de evolutie van causaliteit moet in dit spanningsveld worden begrepen. Na de mentalistische interpretatie van Hume en Kant, de mokerslagen van Mach, Pearson en Russell, en de churchlandiaanse drang tot eliminatie van obscure begrippen leek het lot van het begrip causaliteit in de wetenschap bezegeld.

Variatie, onzekerheid en statistiek

De omwenteling in het causaliteitsdenken kwam in de 19e eeuw op een nogal paradoxale manier tot stand door de convergentie van enkele ongelijksoortige ontwikkelingen. Allereerst vond volgens de filosoof Ian Hacking ‘een erosie van het determinisme’ plaats. Verworven kennis over de grilligheid van de (levende) natuur en de veelvuldigheid van haar verschijningsvormen, gepaard aan een historisering van het wereldbeeld vormden een obstakel voor een deterministische visie op de werkelijkheid. Variatie en verandering kwamen centraal te staan en mede door toedoen van de kruisbestuiving tussen biologie en statistiek ontstond een probabilistisch wereldbeeld. Daarmee leek de genadeslag te zijn toegebracht aan een conceptie van causale wetmatigheid die doorgaans met een deterministisch wereldbeeld werd geassocieerd. Feitelijk voltrok zich een emancipatieproces in het denken over onzekerheid in een tweetal stappen: eerst zouden Galton (1822–1911) en Pearson vanuit biologisch perspectief onzekerheid een belangrijke plaats geven in het wetenschappelijke wereldbeeld, daarna zou de fysisch Niels Bohr (1855–1962) met zijn Kopenhagen-interpretatie van de kwantummechanica laten zien dat onzekerheid een bouwsteen is van de natuur, irreducibel en niet te herleiden tot een gebrek aan kennis (Starmans, 2018a). De tweede ontwikkeling in het causaliteitsdenken ontstond vanuit methodologische hoek, waar behoefte ontstond aan een nieuwe, constructieve notie van causaliteit, die zich had los gemaakt van

zowel de oude metafysica als het laplaciaanse determinisme. Pogingen causaliteit in een meer pragmatische en experimentele context te verankeren zijn terug te vinden in het werk van onder meer C.S. Peirce (1839–1914) en vooral John Stuart Mill (1806–1873). Laatstgenoemde beschouwde causaliteit vanuit het geheel aan omstandigheden die moesten worden gekend, gecontroleerd of gemanipuleerd om causale relaties of een beoogd causaal effect vast te stellen. Zijn *System of Logic* (1843) bevat de befaamde ‘vijf methoden van Mill’ en was feitelijk een methodologisch handboek *avant la lettre*, waarin de auteur probeerde de kloof tussen de abstracte epistemologie en de feitelijke wetenschapsbeoefening te overbruggen, vooral inzake het denken over causaliteit. Dat bleek de jaren daarna hard nodig, juist omdat vele ‘variatie-en-onzekerheid-rijke’ wetenschappen als biologie, genetica, psychologie, sociologie en economie opkwamen, verzelfstandigden, nog weinig status genoten en daardoor een begrijpelijke grondslagencrisis doormaakten. Het proces van ontworsteling aan de speculatieve wijsgerige traditie vroeg om een geoperationaliseerd causaliteitsbegrip. Object, doel en methode moesten worden bepaald en gezocht werd naar een taal die recht kon doen aan variatie en covariatie, onzekerheid en de inherent causale onderzoeksvragen die daarmee verbonden waren; een formele methode die een *causale interpretatie* toeliet die recht doet aan de specifieke causale vraagstellingen binnen dat wetenschapsgebied. De liaison tussen causaliteit en de statistiek die Quetelet, Galton, Pearson, Yule en anderen in gang zouden zetten ontstond juist doordat veel statistische technieken in wisselwerking met de alledaagse praktijk en causale ‘grandeur en misère’ van die nieuwe disciplines zijn ontwikkeld. Door die wisselwerking maakten deze wetenschappen deel uit van de probabilistische revolutie (Krüger, 1987).

Steevast werden statistische technieken ontwikkeld, beoordeeld, herzien of verworpen afhankelijk van hun causale *fitness* of geschiktheid. De vernieuwing werd voortgestuwd doordat statistici steeds bedrevener bleken in het zoeken naar en creëren en analyseren van nieuwe bronnen van (co)varia(n)tie, het belang van interventie begrepen en door het ontwikkelen van complexere methoden en technieken die hieraan recht deden. Dit alles niet via *toy-examples*, maar doordat causale praktijkproblemen zich aandienen die met bestaande methoden niet adequaat werden benaderd. Dat begon nota bene met de anti-causalist par excellence Karl Pearson. Onmiddellijk na Pearsons klassieke paper startte de beroem-

de polemie met Udney Yule, vervolgens kwamen de partiële en semi-partiële correlatie, gekritiseerd op causale gronden door (Burks, 1926). Een volgende stap vormden regressie-analyse en padcoëfficiënten van Sewall Wright, die tot causale debatten leidden (Starmans, 2018b), de daaruit voortgekomen structurele vergelijkingsmodellen idem dito. Tegelijkertijd ontstond de variantieanalyse, gebaseerd op Fishers ideeën over interventie, manipulaties en experimenteel design, waarmee tevens het pad werd geëffend voor latere studies naar onderzoeksdesigns door Blalock, Cook, Cambell, de Groot et cetera. Maar ook factor-analyse en principale componenten analyse golden als technieken waarmee niet gemeten, maar causaal werkzaam geachte grootheden werden gemodelleerd, bijvoorbeeld de ‘sociale feiten’ die Emile Durkheim als niet reduceerbare causale entiteiten postuleerde. Het meest historisch saillante voorbeeld betreft het feit dat bij uitstek de voornoemde nieuwe wetenschappen data kenden die als hiërarchisch worden opgevat; zij zijn genest, ingebed, of gelaagd, dat wil zeggen kennen verschillende niveaus. De wisselwerking, interdependenties en interacties tussen de verschillende *levels* zijn essentieel in causale onderzoeksvragen en theorievorming. Het streven de verschillende ‘niveaus’ simultaan in een analyse te betrekken, leidde tot vele vormen van multilevel analyse. Of het nu gaat om *multilevel models in strikte zin, hierarchical (linear) models, nested data, mixed models, klassieke split-plot designs, random coëfficiënt of random-effect models, repeated measures*, et cetera. Het zoeken naar en creëren van nieuwe bronnen van varia(n)tie leidt tot modellen met meer random effects, minder fixed effects en een meer gesofisticeerde analyse van residuele error, dit alles binnen en tussen de verschillende niveaus. Tot slot zij opgemerkt dat vele van deze en andere technieken kunnen worden gekoppeld aan klassieke paradoxen en anomalieën (confounding, ‘spurious correlation’, Simpson-paradox, Robinsons ecologische drogreden, et cetera). Tot op de dag van vandaag domineren al deze methoden experimentele en observationele studies en vormen zij de basis voor causale uitspraken in vele empirische disciplines. Zonder de liaison met de statistiek was het begrip causaliteit in het licht van de gesignaleerde crises al lang verdwenen uit de wetenschappen.

Causaliteit, AI en data science

Dankzij fundamenteel onderzoek binnen de informatica

en de kunstmatige intelligentie vond aan het eind van de 20e eeuw een computationele wending plaats in het causaliteitsdenken. Allereerst zouden eind jaren 80 probabilistische of Bayesiaanse netwerken de regelgebaseerde kennissystemen verdringen en door velen al een *causale interpretatie* krijgen toebedeeld (Williamson, 2005). In aansluiting daarop publiceerde Pearl in 2000 zijn lijvige studie *Causality: Models, Reasoning and Inference*, dat in 2009 een grondige revisie beleefde. De benadering van Pearl gaat terug tot het werk van Sewall Wright en SEM, kent grafische representaties en een interventionistische benadering van causaliteit, waarbij de ongelijkheid *intervention ≠ conditioning* vooropstaat en de taal van de kansrekening wordt uitgebreid met de ‘do-operator’ als opmaat naar de bekende ‘do-calculus’. Pearl beschouwt zijn werk als een radicale breuk met de statistische traditie, waartoe hij feitelijk ook de mede door hem ontwikkelde Bayesiaanse netwerken rekent. Pearl kent ‘no greater impediment to scientific progress than the prevailing practice of focusing all of our mathematical resources on probabilistic and statistical inferences’, juist omdat in deze traditie volgens Pearl causale noties en nomenclatuur uit de formele, wetenschappelijke taal zijn verbannen. Hij spreekt dan ook de hoop uit dat na de 20e eeuw van ‘Probability’ de 21e eeuw een nieuw era van ‘Causality’ zal vormen. Blijkbaar gaat de bekering of omwenteling Pearl niet snel genoeg. In 2018 publiceert hij het op een groot publiek gerichte en nu reeds commercieel succesvolle *The Book of Why; the new science of cause and effect*. Hij creëert een zogenaamde ‘ladder van de causaliteit’, die bestaat uit een drietal treden, *associatie, interventie* en *counterfactuals* en betoogt op basis van een vluchtige blik op de geschiedenis dat de statistiek doorgaans niet verder is gekomen dan de eerste trede. Opmerkelijk is de polemische, verongelijkte en soms onverzoenlijke toon, waarbij historische nuanceringszins naar de achtergrond is verdrongen. In dit korte essay moeten een analyse van deze benadering en bovenal de *receptie* ervan in de theoretische statistiek buiten beschouwing blijven. We beperken ons hier tot de constatering dat Pearls eigen benadering tegen de achtergrond van het hier geschetste rijke causaliteitsbegrip evenzeer probabilistisch is als de andere varianten. Veel statistici hebben zich bovendien de laatste decennia gericht op *causal inference* in observationele data, beamen ook het nut van counterfactuals, die dankzij het gelijknamige boek van Lewis uit 1973 een enorme herwaardering beleven. Verschillende statistici omarmen een ‘hypothetische inter-

ventie’ in observationele studies, maar lijken vooralsnog de potential-outcome traditie van Neyman-Rubin als gedachtenexperiment te prefereren.

Hoe dan ook, statistiek lijkt daarmee tussen twee vuren te zijn geraakt, omdat in data science een anti-epistemische en anti-statistische wending zich lijkt te voltrekken, waarbij klassieke inferentie, modelbegrip en causaliteit voor hedendaagse data science apologeten uit den boze zijn (Starmans, 2018a). Voor de een is de statistiek ‘te causaal’, voor de ander ‘niet causaal genoeg’. In weerwil van de visie binnen data science moet ter verdediging worden opgemerkt dat de statistiek hier een grote traditie van zelfreflectie kent. Toen vader en zoon Pearson, Fisher, en Neyman al ruziënd de fundamenten van de inferentiële parametrische statistiek legden, dienden de anti-theses zich al spoedig aan: variërend van de eerste Bayesianen, de niet-parametrische statistiek van Wilcoxon, Whitney en anderen, John Tukeys’ *Future of Data Analysis* (1963) en *Exploratory data analysis* (1978) en meer recentelijk Leo Breimans moderne klassieker: *Statistical Modelling: the Two Cultures* uit 2001. De lijst kan moeiteloos worden uitgebreid.

Epiloog

Tot slot enkele kanttekeningen. Uiteraard kan met enige goede wil worden gesteld dat de band tussen statistiek en causaliteit reeds in de 17e en 18e eeuw manifest was, zoals bij Pascals beroemde weddenschap, Bernoulli’s wet van de grote getallen, maar vooral bij John Arbuthnots causale duiding van de gesignaleerde vaste proporties van jongens en meisjes in geboortecijfers en uiteraard bij Thomas Bayes’ exploratie van de omkering van de kansen (Starmans, 2018b). De specifieke combinatie van factoren in de 19e eeuw was echter nodig om deze liaison echt tot stand te brengen. Verder blijkt het causaliteitsdenken vandaag de dag springlevend, dit in weerwil van Russells opvattingen, de vele crises van het begrip en de huidige kritiek van sommige data science apologeten. Ook in de wetenschapsfilosofie vormen de interacties tussen causaliteit, wetenschappelijke verklaringen en wetmatigheden/ natuurwetten nog steeds een belangrijk aandachtsgebied. Daarnaast moet worden opgemerkt dat het pluralisme rond causaliteit vooral gestalte krijgt tegen de achtergrond van de driedelige gelaagde structuur van de queeste naar kennis; filosofie (metafysica, kennisleer),

methodologie en tot slot het gebied van afzonderlijke disciplines en toepassingsgebieden. De huidige moeizame samenspraak tussen statistiek en causaliteit uit zich vooral doordat in dit opzicht nauwelijks kruisbestuiving bestaat tussen kennisleer en methodologie en ook binnen de statistiek een scheiding der geesten manifest is (Starmans, 2018b). Pearls opvatting dat de statistiek te weinig causaliteit ‘bevat’ moet worden beschouwd tegen de achtergrond van zijn futurologische doel: het redden van de sterke AI. Meer dan in schattingstheorie of inferentiële statistiek lijkt hij geïnteresseerd in het uitrusten van robots met een menselijke notie van causaliteit. Anders dan bijvoorbeeld Churchland acht Pearl taal cruciaal voor het denken en een voorwaarde voor moral intelligent agents. Het maakt duidelijk dat het onderhavige debat deels past in de aloude AI-controverse *to represent or not*, dat in de klassieke publicatie van Rodney Brooks scherp is verwoord (Brooks, 1980). Probeer je causaliteit te formaliseren in de objecttaal (b.v. door een operator) of juist in de metataal door onderzoeksdesign, mechanisme van data genereren, door geavanceerdere statistische technieken of door te onderkennen dat het gaat om specifieke contexten die benaderd kunnen worden vanuit gespecificeerde language games.

Hoe dan ook in het licht van de *represent or not* tegenstelling kun je een causalist zijn zonder het expliciet te representeren, je kunt zelfs een causalist zijn zonder het begrip te willen gebruiken, zoals bij Pearson het geval was. Vanuit filosofische optiek is de gehele problematiek een prachtige conceptuele analyse van het rijke begrip van de causaliteit. In dit licht moet tot slot de reikwijdte van de probabilistische benadering in welke vorm dan ook niet worden overschat. Dat blijkt als men vertrekt vanuit een vijftal kernvragen die de causale relatie/vraag kunnen typeren: token versus type, individu/concreet versus groep/abstract, kwantitatief versus kwalitatief, deterministisch versus probabilistisch, fysisch versus mentaal. Wie aangemoedigd door Stephen Kerns cultuurhistorische exercitie vanuit de geschetste linguïstische optiek een causale relatie analyseert, merkt ongetwijfeld dat deze vaak *token*, *kwalitatief* en *deterministisch* is, zowel in het alledaagse taalgebruik als bij vele professionele toepassingen in het juridische, medische of technische domein, die rond het juiste gebruik van causaliteit elk een eigen taalspel spelen met regels en conventies die zich ten ene male onttrekken aan een standaard probabilistische reductie. In het huidige discours rond causaliteit wordt dat nog wel eens onderbelicht.

LITERATUUR

- Anderson, C. (2008). The end of theory: The data deluge makes the scientific method obsolete. *Wired Magazine*, 6. (www.wired.com/2008/06/pb-theory/)
- Bostrom, N. (2016). *Superintelligence; Paths, dangers and strategies*. Oxford: Oxford University Press.
- Brooks, R. (1991). Intelligence without representation. *Artificial Intelligence*, 47, 1-3.
- Burks, B. (1926). On the inadequacy of the partial and multiple correlation technique. *Journal of Educational Psychology* 17(8), 532-540
- Domingos, P. (2016). *The master algorithm; how the quest for the ultimate learning machine will remake our world*. New York.
- Halevy, A., Norvig, P., & Pereira, F. (2009). The unreasonable effectiveness of data. *IEEE Intelligent Systems*, 24(2), 8-12.
- Krüger, L., Daston, L., & Heidelberger, M. (Eds.). (1987). *The probabilistic revolution. Volume 1, Ideas in history*. Cambridge, MA: MIT Press.
- Krüger, L., Daston, L., Heidelberger, M., Gigerenzer, G., & Morgan, M. (Eds.). (1987). *The Probabilistic Revolution. 2 Volumes*. Cambridge, MA: MIT Press.
- MacKie, J. L. (1980). *The Cement of the Universe; a study of causation*. Oxford: Clarendon Press.
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work and think*. Boston: Houghton Mifflin Harcourt.
- Mumford, S. (2014). *Causality, a very short introduction*. Oxford.
- Pearl, J. (2009). *Causality: Models, reasoning and inference*. Cambridge: Cambridge University Press.
- Pearl, J., & MacKenzie, D. (2018). *The book of Why; The new science of cause and effect*. New York.
- Russo, F., & Illari, P. (2016). *Causality: Philosophical theory meets scientific practice*. Oxford.
- Starmans, R. J. C. M. (2016) the advent of data science; Some considerations on the unreasonable effectiveness of data. In Peter Buhlmann, e.a. (eds.), *Handbook of big data; Handbooks of modern statistical methods*. New York: Chapman & Hall/CRC, 2016.
- Starmans, R. J. C. M. (2018a). The predicament of truth: On statistics, causality, physics and the philosophy of science. In M. J. Van der Laan & S. Rose (Eds.), *Targeted learning in data science: Causal inference for complex longitudinal studies* (Springer Series in Statistics). Springer International Publishing.
- Starmans, R. J. C. M. (2018b). Prometheus unbound or Paradise regained: On causality in the contemporary AI-data science debate. *Journal of the French Statistical Society* (Special issue on Causality).
- Williamson, J. (2009). Probabilistic theories of causality. In Beebe, H., Hitchcock, C., & Menzies, P. (Eds.). *The Oxford handbook of causation* (pp. 185-212). Oxford: Oxford University Press.

RICHARD STARMANS is verbonden aan de Faculteit Bèta-wetenschappen (Department of Information and Computing Sciences) van de Universiteit Utrecht. Hij doet onderzoek op het snijvlak van filosofie, statistiek en informatica. E-mail: starmans@cs.uu.nl

Gecombineerde radiotherapiebehandelingen

– duur is niet altijd beter

STEFAN TEN EIKELDER

Radiotherapie is een verzamelnaam van meerdere bestralingsmethodes (modaliteiten) voor oncologische patiënten. Naast conventionele radiotherapie met hoogenergetische fotonstralen (X-ray) zijn er ook methodes die gebruik maken van geladen deeltjes, zoals geavanceerde protonentherapie. Protonentherapie is niet in alle gevallen beter dan conventionele foton radiotherapie, maar heeft wel een veel hoger prijskaartje (Bortfeld & Loeffler, 2017). Wiskundige optimalisatietechnieken uit de Operations Research worden gebruikt om een optimaal gecombineerde proton-foton behandeling op te stellen (Ten Eikelder, 2017). Dit biedt tevens inzicht in wanneer protonentherapie voordeel oplevert, en wanneer conventionele foton radiotherapie de beste optie is.

Verschil conventionele radiotherapie en protonentherapie

Een van de kenmerken van protonentherapie is de nauwkeurigheid. Door met een deeltjesversneller protonen te versnellen, deze te bundelen en de juiste energie mee te geven, kan zeer nauwkeurig bepaald worden waar in het lichaam de protonen hun energie afgeven, zie figuur 1. Het effect van de straling op weefsel is de geabsorbeerde dosis, gemeten in Gray (Gy), waarbij 1 Gy gelijk is aan 1 Joule energie per kilogram weefsel. Doordat de energieafgifte van protonenbundels nauwkeuriger te controleren is dan voor fotonenbundels, wordt er minder dosis toegediend aan omliggende gezonde weefsels, ook wel *organs-at-risk* (OARs) genoemd.

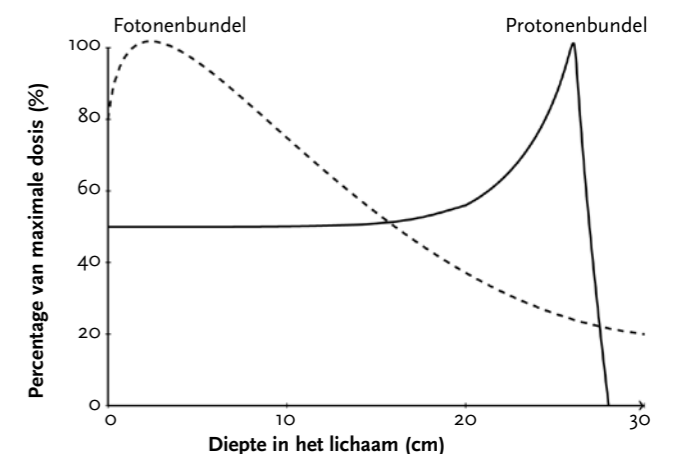
Als gevolg is er bij protonentherapie minder kans op complicaties in OARs dan bij conventionele radiotherapie. Echter, juist door de mogelijkheid om een hoge dosis zeer nauwkeurig toe te dienen spelen onzekerheden een grote rol. Denk hierbij onder andere aan onzekerheden in exacte orgaan- en tumorlocaties. In de huidige klinische praktijk worden deze meegenomen in de planning door grotere marges rondom de tumor aan te nemen voor protonentherapie dan voor conventionele foton radiotherapie. Hierdoor is de huidige implementatie van protonentherapie niet altijd beter dan conventionele foton ra-

diotherapie, en zal het per patiënt verschillen of protonen een verbetering opleveren of niet.

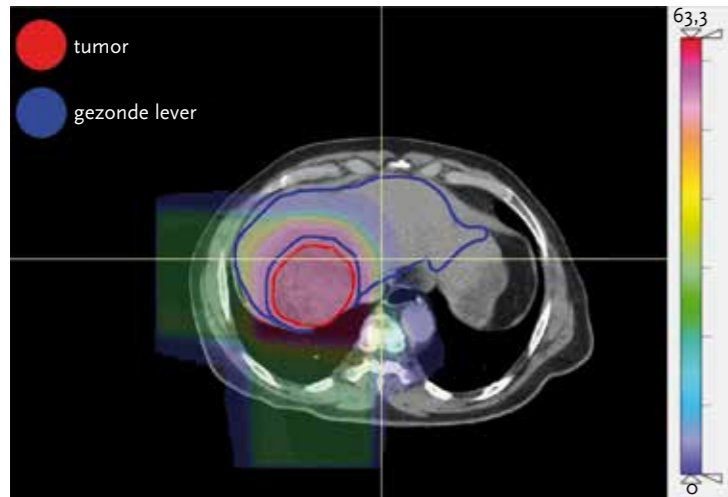
Gezien het prijskaartje van protonentherapie is het belangrijk te weten wanneer protonentherapie beter is dan foton radiotherapie, wanneer niet, en wanneer wellicht een gecombineerde proton-foton behandeling uitkomst biedt. Het belang hiervan wordt onderstreept door de ontwikkeling van protonentherapie in Nederland (ZON-PTC, 2018). Recentelijk is de eerste patiënt in het UMC Groningen behandeld, en ook de protoncentra in Delft en Maastricht verwachten binnenkort de eerste patiënten te behandelen. De beslissing om individuele patiënten al dan niet te behandelen met protonentherapie wordt in Nederland genomen op basis van modelmatig bewijs voor de verbetering in behandelingskwaliteit die protonentherapie kan bieden (Widder et al., 2016).

Treatment planning

Een belangrijk concept in radiotherapie *treatment planning* is fractionering, dat wil zeggen het opdelen van de bestraling in meerdere kleine fracties. Dit concept is gebaseerd op de observatie dat gezonde cellen de capaciteit hebben een hogere dosis te tolereren als deze geleverd



Figuur 1. Verschil in dosisafgifte tussen een fotonenbundel en een protonenbundel. De energie van de protonen wordt zo gekozen dat de piek van de protonenbundel zich bij de tumor bevindt



Figuur 2. Een proton dosisverdeling geprojecteerd op een CT scan slice van een lever patiënt, waar tevens de tumor (rood) en gezonde lever (blauw) zijn weergegeven. De kleur van de dosisverdeling geeft de dosis (in Gy) op een bepaalde locatie, volgens de legenda rechts

wordt in meerdere fracties verspreid over meerdere dagen. Het optimale aantal behandeldagen verschilt van patiënt tot patiënt, en is onder andere afhankelijk van de patiënt-specifieke anatomie en weefsel radiosensitiviteit. Bij een behandeling verspreid over meerdere dagen is het niet noodzakelijk dat elke dag hetzelfde stralingstype gebruikt wordt. Het is dus mogelijk om een behandeling voor te schrijven waarbij een deel van de bestraling met protonen gebeurt, en een deel met fotonen.

In een treatment plan wordt beschreven hoe een individuele patiënt behandeld wordt. Naast het fractioneringsschema worden hier onder meer de geometrische keuzes beschreven. Bestraling gebeurt doorgaans niet vanuit een enkele richting, maar stralenbundels (met fotonen of protonen) worden doorgaans vanuit verschillende hoeken op de patiënt gericht, om zo de schade aan OARs te beperken. Het maken van optimale keuzes met betrekking tot dit ruimtelijke aspect wordt tevens gedaan door middel van operations-researchtechnieken (zie voor een introductie Hoffmann & Den Hertog, 2018). De resulterende dosisverdeling geeft voor elke voxel (3-dimensionaal subvolume) van de tumor en OARs de dosis die het toegediend krijgt. Figuur 2 geeft een voorbeeld van een proton dosisverdeling geprojecteerd op een CT scan slice.

Een optimale combinatie

Voor een gegeven dosisverdeling staat de verhouding in dosis tussen voxels vast, en staat het schalen van de gehele dosisverdeling gelijk aan het verhogen of verlagen van de intensiteit van alle stralenbundels. Een gecombi-

neerde proton-foton behandeling wordt verkregen door in elke fractie een geschaalde variant van de proton of foton dosisverdeling toe te dienen.

Voor de optimale combinatie dient het biologische effect van fractionering meegenomen te worden. Het Biological Effective Dose (BED) model is het meest klinisch geaccepteerde model om dit effect te kwantificeren. Het model geeft het biologische effect voor een gegeven totale dosis en het aantal fracties waarin deze dosis geleverd wordt. Het BED-model is een veelgebruikt kwadratisch model dat ook wordt gebruikt om voor behandelingen met een enkele modaliteit (foton of proton) het optimale fractioneringsschema te bepalen (Saberian et al., 2016).

Aantoonbare optimaliteit van gecombineerde proton-foton behandelingen

Om een eerste beeld te krijgen van het nut van gecombineerde proton-fotonbehandelingen wordt gekeken naar simpele situaties waarbij maar een of twee OARs een rol spelen (Ten Eikelder, 2017). In deze situaties is het doel het maximaliseren van de tumor BED, onder gegeven BED toleranties op de OARs. De resulterende optimalisatie problemen kunnen analytisch opgelost worden, waaruit blijkt dat twee factoren een rol spelen. Ten eerste bestaat de mogelijkheid dat de proton en foton dosisverdeling verschillende OARs beter sparen. In dit geval kan een gecombineerde behandeling met deels proton fracties en deels foton fracties deze tegenstrijdige belangen het beste tegen elkaar afwegen.

Ten tweede is het zelfs zo dat, als de proton dosisverdeling beter is dan de foton dosisverdeling voor de meest relevante organen in termen van fysieke dosis, het in termen van het biologische effect (d.w.z., BED) toch beter kan zijn om een deel van de behandeling met fotonen te doen. Dit is mogelijk omdat, door het dure prijskaartje en/of de gelimiteerde capaciteit van een ziekenhuis, vaak het maximale aantal proton fracties dat toegewezen mag worden aan een patiënt beperkt is. Echter, fotonfracties zijn ruimer beschikbaar. In gecombineerde proton-fotonbehandelingen kunnen we dus in totaal meer fracties leveren, waardoor het effect van fractionering beter uit te buiten is.

Realistisch model

Om te bepalen of deze effecten daadwerkelijk optreden in meer realistische gevallen, wordt tevens een algeme-

ner model ontwikkeld. Dit model kan rekening gehouden met meerdere gezonde weefsels die risico lopen, en de stralingstolerantie van OARs kan op verschillende manieren worden geformuleerd. Seriële weefsels zoals de ruggengraat verliezen het grootste deel van hun functie als er ook maar een schakel tussenuit valt, hiervoor wordt de tolerantie gedefinieerd als een maximum punt dosis voorwaarde. Een parallel orgaan zoals de lever kan blijven functioneren als er een deel wegvalt, hiervoor is vooral de gemiddelde dosis belangrijk. Een andere veelgebruikte voorwaarde heeft betrekking op het dosis-volume histogram (DVH) van een behandelingsplan en specificeert het maximale weefsel volume dat een dosis boven een bepaalde bovengrens mag krijgen.

De complicerende factor in het resulterende optimalisatieprobleem is voornamelijk de DVH randvoorwaarden, aangezien het van tevoren niet duidelijk is welk deel van het volume het beste de hoge dosis kan krijgen. Het optimalisatieprobleem heeft hierdoor een binaire variabele per voxel, en voor volumes van vele duizenden voxels kunnen deze problemen niet (binnen acceptabele tijd) opgelost worden. Hiervoor is een algoritme ontwikkeld dat bestaat uit een eenvoudig oplosbaar deelprobleem en hoofdprobleem waarvoor een *pattern search* heuristiek gebruikt wordt (Audet & Dennis Jr., 2002).

Drie patiëntgroepen

Het ontwikkelde algoritme is getest op data van 17 leverpatiënten behandeld met protonen bij het Massachusetts General Hospital (Boston, VS), waar dit onderzoek ook is uitgevoerd. Uit de resultaten blijkt dat we het patiënten cohort grofweg kunnen opdelen in drie groepen (Ten Eikelder et al., 2018). Voor de eerste groep van 5 patiënten blijkt dat er een verbetering in tumor BED mogelijk is indien de patiënt niet alleen met protonen behandeld wordt, maar een deel van de behandeling met fotonen gedaan wordt. Het blijkt inderdaad dat voor deze patiënten meerdere organen een rol spelen, en voor sommige zijn protonen gunstiger, voor andere fotonen. Dit betekent onder andere dat de in de praktijk gebruikte behandelingen niet voor elke patiënt optimaal zijn (zie tabel 1).

GROEP	AANTAL PATIËNTEN
Verbetering	5
Alternatief	3
Geen verbetering	9
TOTAAL	17

Tabel 1. Resultaten voor cohort van 17 leverpatiënten

Voor de tweede groep (3 patiënten) levert een gecombineerde behandeling geen beter klinisch resultaat, maar kunnen ten opzichte van de protonbehandeling wel een deel van de proton fracties vervangen worden door foton fracties, met een zeer minieme vermindering in tumor BED als gevolg. Met andere woorden: een gecombineerde proton-foton behandeling is grofweg klinisch gelijkwaardig, maar een stuk goedkoper. Voor de derde groep (9 patiënten) blijkt protonentherapie inderdaad betere resultaten op te leveren dan foton radiotherapie of gecombineerde proton-foton behandelingen. Voor deze patiënten is het vervangen van (een deel van) de protonfracties door fotonfracties onverstandig.

Conclusie

De beslissing om patiënten al dan niet te behandelen met protonentherapie hoeft geen beslissing tussen uitersten te zijn: bij een behandeling die bestaat uit meerdere dagen (fracties) behoren ook gecombineerde proton-foton behandelingen tot de mogelijkheden. Door toepassing van operations-researchtechnieken is de potentie van dergelijke behandelingen onderzocht. Zowel uit analytische resultaten als uit simulaties op basis van klinische data blijkt dat gecombineerde behandelingen voordelen kunnen hebben ten opzichte van behandelingen met enkel fotonen of proton kunnen opleveren. Deze voordelen zijn te halen uit de kosten, de behandelingskwaliteit, en soms zelfs uit beide.

STEFAN TEN EIKELDER is promovendus aan Tilburg University, waar hij onderzoek doet naar robuuste optimalisatie in radiotherapie. In 2017 rondde hij zijn master in Operations Research en Management Science af en in 2018 zijn research master in Operations Research. Hij ontving de VVSOR Hemelrijk Award 2017 voor zijn master scriptie *Optimal fractionation for combined photon-proton treatments in radiation therapy*. E-mail: S.C.M.tenEikelder@uvt.nl

LITERATUUR

- Audet, C., & Dennis Jr., J. E. (2002). Analysis of generalized pattern searches. *SIAM Journal on Optimization*, 13(3), 889-903.
- Bortfeld, T., & Loeffler, J. (2017). Three ways to make proton therapy affordable. *Nature*, 549, 451-453.
- Homann, A., & den Hertog, D. (2008). Optimalisatie van bestralingsbehandeling tegen kanker: Operations Research toegepast voor radiotherapie. *STAtOR*, 9(4).
- Saberian, F., Ghate, A., & Kim, M. (2016). Optimal fractionation in radiotherapy with multiple normal tissues. *Mathematical Medicine and Biology*, 33(2), 211-252.
- Ten Eikelder, S. (2017). *Optimal fractionation for combined photon-proton treatments* (Master Thesis). Tilburg University.

vertellen dat die uit de mededeling dat de kans dat het morgen regent 50% is en dat de kans dat het overmorgen regent ook 50% is, concludeerde dat het 100% zeker is dat het de komende twee dagen gaat regenen.

Wij statistici hebben kennelijk de schone taak de mensheid op te voeden en ze te leren om onze statistieken te snappen. Welk een dedain. Daar zou zelfs de sanseveria de slappe lach van kunnen krijgen. Nou ja, volgens de Max Planck-geleerde kunnen we effectiever de journalisten opvoeden: die lui snappen vervolgens hoe je statistieken moet presenteren aan de rest van de mensheid. Zou de rest het dan wel snappen? De CBS-jongens geloven daar voorlopig niks van: betrouwbaarheidsintervallen, bijvoorbeeld, zijn voorlopig niet uit te leggen. Het CBS houdt het bij puntschattingen, net als mijn beppe bij de sanseveria.



P.S. Het toeval wil dat de sanseveria onlangs ter sprake kwam tijdens een etentje met het 'vorige' bestuur van mijn schaatsclub. Voor een goed begrip moet ik eerst nog even vertellen dat de aanmeldingsbrief die ik naar de schaakbond dacht te hebben gestuurd – dat was in mijn eerste jaar als Gronings wiskundestudentje – daar niet belandde. Bij het zoeken in de telefoongids (een boekwerk met namen, adressen en telefoonnummers) vond ik na enig speurwerk het adres van de KNSB. Na een week kreeg ik al bericht met een overzicht van alle schaatstrainingen van de, inderdaad ook, KNSB. Tja, zo heb ik inmiddels drie 11-stedentochten gereden en nooit Beverwijk bezocht. Zo nu en dan schaak ik met mijn kleinzoon van dertien. Ik heb hem geleerd dat de koploze zwarte koningin op zwart moet beginnen en de witte op wit. Ook heb ik hem het verhaal van oma beppe en de sanseveria verteld. Zijn schalkse blik deed me vermoeden dat hij me niet geloofde. Op een laag bakstenen muurtje tegenover onze tafel in het sfeervolle restaurantje stond een sanseveria. Nadere inspectie leerde me dat hij nep was. Ik heb me aangemeld bij het nieuwe statistiekclubje, de Statistics Communication Division van de VVSOR.

GERARD SIERKSMA is emeritus hoogleraar Kwantitatieve Logistiek en Sportstatistiek aan de Rijksuniversiteit Groningen. E-mail: g.sierksma@rug.nl

L

NMB & NGB SEMINAR

DEEP LEARNING AND ITS IMPACT ON OPERATIONS RESEARCH

Conference Center De Werelt in Lunteren,
January 16, 2019

Join the seminar and learn from experts all about the impact of Deep Learning and AI on Operations Research! With Max Welling and other speakers from industry and academia.

Deep Learning is booming in academia as well as business. It often works on the same problems as Operations Research. The program covers different aspects of linking Deep Learning with Operations Research; from a vision on the future to the current state of research, and from a tutorial to real life applications.

The list of speaker includes:

Eric Postma (Tilburg University, Jheronimus Academy of Data Science): The true AI revolution

Max Welling (University of Amsterdam): Learning to solve OR Problems

Johan van Rooij (CQM): Safe and efficient inspection of railway tracks using deep learning models (Winner Hendrik Lorentz Prize for innovative applications of Data Science)

Marco Lübbecke (RWTH Aachen University): (How) does Machine Learning impact Operations Research?

Eline Werkman (ORTEC), Advanced modelling techniques applied on a Revenue Management solution for a holiday park provider

... and more ...

More information on the invited speakers and the programme will follow in the next weeks. www.lnmb.nl/conferences/2019/announcementnbgblnmbseminar/