

Van *top-down* naar *bottom-up*: Hoe machine learning tot nieuwe inzichten in leugenachtig gedrag kan leiden.

Dr. Sophie van der Zee (Erasmus Universiteit Rotterdam) & Dr. Ronald Poppe (Universiteit Utrecht)

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 op 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 het 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 omstandigheden 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 hele 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 leugendetectie methode 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 onschuldigen risico als ze instemmen met een polygraaftest.

Het tweede unieke aan dit experiment is de combinatie van drie verschillende leugendetectie-methodes. 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 geavanceerde smartwatch. Door verschillende methoden te combineren is het moeilijker om succesvol de *leugendetector* 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 vertelden 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.

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 een 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 erg tijdsintensief is en daardoor de beschikbare data beperkt is. Data uit de praktijk is slechts zelden bruikbaar voor leugendetectie doeleinden, 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 geroteerd 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. Tenslotte hebben we een simpel classificatie-model gebruikt: Gaussian Naive Bayes. Hierbij worden alleen de gemiddelde en standard-deviatie 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-dimensionele metingen van gedrag, typisch voor automatische observatie, kunnen dus leiden tot nieuwe inzichten.

De complementaire karakters van theorie-gedreven, validerend onderzoek (*top-down*) en data-gedreven, exploratief 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 de individu en hoe gedragspatronen over tijd veranderen.

Referenties

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*.

Contactgegevens

Dr. Sophie van der Zee
vanderzee@ese.eur.nl
Postbox 1738
3000 DR Rotterdam
Nederland

Biografie Sophie van der Zee

Van der Zee is gepromoveerd op het automatiseren van leugendetectie aan Lancaster University in het security centrum waar psychologie en informatica nauw samenwerkten om veiligheidsvraagstukken te beantwoorden. Sindsdien heeft ze onderzoek gedaan naar liegen, oneerlijk gedrag en cybercriminaliteit in verschillende disciplines (psychologie, informatica en gedragseconomie). Ze is momenteel werkzaam als universitair docent gedragseconomie aan de Erasmus Universiteit.

Biografie Ronald Poppe

Poppe is gepromoveerd aan de Universiteit Twente op het automatisch annoteren van gedrag uit videobeelden. Voor zijn overzichtspapers op dit gebied ontving hij tweemaal de citatie-award van het tijdschrift "Image and Vision Computing". Hij is momenteel universitair docent aan de Universiteit Utrecht, waar hij onderzoek doet naar machine learning modellen voor de automatische analyse van, met name, interactief gedrag uit video en andere sensoren.