Digital Security

Radboud University

Prof. dr. Mireille Hildebrandt
Chair of Smart Environments, Data Protection and the Rule of Law

Faculty of Science
Institute for Computing and Information Sciences (iCIS)
Radboud University

Toernooiveld 212
6525EC Nijmegen
The Netherlands

☎ +31 24 3652077

www.ru.nl/fnwi
m.hildebrandt@cs.ru.nl

Vaste commissie voor Justitie en Veiligheid
Tweede Kamer der Staten-Generaal

Our reference	Your reference	Direct number	Date
		+ 31 24 365 2076	21 maart 2018

Subject
STANDPUNT
ARTIFICIELE INTELLIGENTIE EN RECHT

E-mail
m.hildebrandt@cs.ru.nl

1. Kunstmatige intelligentie is een wat ongelukkige term, waar van alles en nog wat onder wordt begrepen. Het roept hoge verwachtingen op en leidt gemakkelijk tot misverstanden omtrent mogelijkheden en beperkingen. Praktisch gezien zal het gaan om de inzet van **machinaal leren (ML)** in de rechtsvinding, bij de voorbereiding en/of het nemen van bestuursrechtelijke besluiten, en bij het ontwerpen van regelgeving. Vanwege de beperkte ruimte ga ik niet in op blockchain (distributed ledger technologies), maar ik verwacht dat beide gecombineerd gaan worden.
2. ML is een techniek om **patronen te herkennen** in relevante en liefst zo compleet mogelijke databestanden met als doel patronen in de werkelijkheid te ontdekken, die in casu gebruikt kunnen worden bij (1) het detecteren en evalueren van mogelijk relevant bewijsmateriaal, (2) het traceren van (a) relevante rechtsbronnen, (b) van argumenten en argumentatielijnen in relevante rechtspraak en doctrine, (c) conflicten tussen toepasselijke regelgeving, en (3) bij het voorspellen van rechterlijke uitspraken.
3. Het ontdekken van statistisch-wiskundige patronen in datasets (hoe compleet ook) staat allerminst garant voor het detecteren van relevante patronen in de werkelijkheid. Bij het ontwerpen van ML applicaties moet een hele serie **ontwerpbeslissingen** worden genomen die **allerhande trade-offs** hebben. Het is cruciaal dat domein experts (juristen) hecht samenwerken met ML experts en de betreffende trade-offs benoemen, zodat uitkomsten op de juiste wijze gerelativeerd kunnen worden. Men zegt dat een machine leert: van ervaring E (training data), met betrekking tot een klasse van taken T (bv correct voorspellen rechterlijke uitspraken), met performance maatstaf P (bv percentage juist voorspelde uitspraken), als die performance t.a.v. taken T, zoals gemeten door P, verbetert dankzij E.[1] Het mag duidelijk zijn dat het nogal uitmaakt wie, hoe, welke taken, training data en performance maatstaven vaststellen. Daarbij speelt ook

[1] Thomas Mitchell, *Machine Learning* (New York: McGraw-Hill Education, 1997).

een rol dat training data meestal handmatig moet worden gelabeld en ruis moet worden verwijderd ('cleansen'). Zowel het labelen als het 'cleansen' is een klus voor domein experts, die het vaak oneens zullen zijn en dan knopen door zullen hakken die doorwerken in de uitkomsten.

4. De inzet van ML voor de onder 2 genoemde doelen is **vooralsnog experimenteel van aard**. Het is zaak om tijd en middelen vrij te maken voor nuttige experimenten en die op de juiste wijze te documenteren, zodat helder is in hoeverre de uitkomsten interessant en valide zijn. Bijvoorbeeld: Hoe zijn de data verkregen, opgeschoond en gelabeld? Welke verschillende taakformuleringen zijn uitgeprobeerd? Welke type wiskundige functies zijn ontworpen om verbanden te testen? Welke verschillende performance maatstaven zijn getest (met welk resultaat)? **Exploratief onderzoek** levert hypotheses op, bevestiging is nog iets anders.[2]
5. Voorzover ML daadwerkelijk zou worden ingezet op een wijze die rechtsgevolg of andere impact heeft moet sprake zijn van **bevestigend onderzoek**. Dat vereist een vooraf geregistreerd onderzoeksontwerp waarin de bovenstaande keuzes zijn vastgelegd en de uitkomsten narekenbaar verantwoord. Nu het hier gaat om het recht, zou zo'n ontwerp altijd openbaar beschikbaar moeten zijn, zodat conclusies weersproken kunnen worden (en andere ontwikkelaars van vergelijkbare technologie ermee verder kunnen werken). Vergelijk de medisch onderzoek, waarbij negatieve resultaten in de la bleven liggen, met alle resultaten van dien. Dit heeft geleid tot een vergelijkbare eis van voorregistratie.
6. ML kan worden ingezet om **juridische kennis te verrijken**, door onverwachte inzichten te bieden op basis van het doorzoeken van grote hoeveelheden tekst. Dat kan tegenwicht bieden aan vastgeroeste ideeën, omissies en inconsistenties zichtbaar maken en nieuwe argumentatielijnen aandragen. ML is echter niet gebaseerd op inzicht in de materie of begrip voor wat op het spel staat, het is een puur statistisch-wiskundige exercitie met een sterke hang naar behavioristische uitgangspunten (het tekst-gestuurde gedrag van juristen wordt vanuit een extern perspectief doorgerekend, dat is alles). Daarbij kunnen zich allerlei bugs en onzichtbare afhankelijkheden voordoen die de uitkomst hachelijk maken (we zien dat nu bij medisch onderzoek),[3] en kunnen problematische vooringenomenheden worden versterkt (ongewenste bias in de training data keren terug in de uitkomsten).[4] Vaak is dat niet zomaar te achterhalen. Om die reden is niet verstandig ML in te zetten ter vervanging van juridische expertise, maar kan het dienen om dezelfde data (bv juridische teksten) op verschillende manieren te doorzoeken zodat de relevante dilemma's snel boven water komen. ML moet vooral niet worden gezien als een bezuinigingsoperatie, eens dat duidelijk is dat het nieuwe problemen schept zal de inzet al gauw meer kosten meebrengen. Selectieve inzet is dus aangewezen.
7. ML kan ook worden ingezet om **de toegang tot het recht** te vergroten. Om bovenstaande redenen is dat een hachelijke zaak, maar ik zie daar wel kansen.

Zie nog mijn 'Oordeelsvorming door mens en machine: heuristieken, algoritmes en legitimatie':
https://works.bepress.com/mireille_hildebrandt/38/

---

[2] Jake M. Hofman, Amit Sharma, and Duncan J. Watts, 'Prediction and Explanation in Social Systems', *Science* 355, no. 6324 (3 February 2017): 486–88, https://doi.org/10.1126/science.aal3856; Mireille Hildebrandt, 'Law As Computation in the Era of Artificial Legal Intelligence. Speaking Law to the Power of Statistics', SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, 7 June 2017), https://papers.ssrn.com/abstract=2983045.

[3] Rich Caruana et al., 'Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-Day Readmission', in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '15 (New York, NY, USA: ACM, 2015), 1721–1730, https://doi.org/10.1145/2783258.2788613.

[4] Julia Angwin et al., 'Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks.', ProPublica, 23 May 2016, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.