Jubel.be | Powered by KNOPSPublishing.com

NIEUWE TECHNOLOGIEËN EN HET RECHT: DE IMPACT VAN ARTIFICIËLE INTELLIGENTIE OP DE RECHTSPRAKTIJK

DEEL II: Machine learning

Rémy Bonnaffé [1]

Jubel.be, 7 november 2019.

[1] Advocaat bij Freshfields Bruckhaus Deringer.

Remy Bonaffé, Nieuwe technologieën en het recht: De impact van artificiële intelligentie op de rechtspraktijk, Deel 1 (Jubel.be, 7 november).

Machine learning en de mogelijkheid om onzichtbare regels te vinden

In onze vorige bijdrage bespraken we de zogenaamde *expert systems*. *Expert systems* vertalen menselijke expertise in een algoritme. Een algoritme kan gedefinieerd worden als een "reeks instructies die worden uitgevoerd om de input te veranderen naar een output"[2]. Het gebruik van een *expert system* veronderstelt evenwel dat de expert in kwestie de regels *kent*, gezien de menselijke expert vereist is om deze regels te programmeren in het *expert system* als zogenaamde instructies. Het is echter niet altijd mogelijk om alle regels te kennen. Daarenboven is het in sommige gevallen eveneens te omslachtig om alle relevante regels te bepalen of te programmeren in een systeem. Een goed, weliswaar niet-juridisch, voorbeeld in dit verband zijn spam-e-mails. Welke regels bepalen welke e-mail spam is of niet? Proberen om deze regels te doorgronden is bijzonder tijdrovend (indien dit überhaupt mogelijk is) en het *expert systeem* zou nog steeds niet erg accuraat zijn gezien een menselijke expert alle mogelijke situaties zou moeten voorzien. Daarenboven is het mogelijk dat het concept spam over de tijd heen kan veranderen en ook van individu tot individu kan variëren. Dit zorgt ervoor dat men een quasi oneindig aantal situaties zou moeten kunnen voorzien, hetgeen weinig mogelijk is.

Het voorbeeld van de spam-e-mails kunnen we ook transponeren naar een juridische context. Hoe zou een jurist bijvoorbeeld het identificeren van een *change of control*-bepaling kunnen programmeren in een *expert system*? Met andere woorden: welke 'harde' regels kan men gebruiken om deze bepalingen te identificeren? Een *change of control*-bepaling is een contractuele clausule die bepaalt dat de wijziging in de controle van een bepaalde vennootschap een bepaald gevolg teweegbrengt, veelal de mogelijkheid om de overeenkomst in kwestie vroegtijdig op te zeggen. Veel vennootschapsjuristen zullen in de context van fusies en overnames geconfronteerd worden met deze *change of control*-bepalingen en zullen in dit verband doorheen de overeenkomsten van de desbetreffende vennootschap moeten zoeken of er dergelijke bepalingen aanwezig zijn. Hoewel het zeker mogelijk is om enkele regels op te sommen die een computer kan gebruiken om een paragraaf te identificeren als zijnde een *change of control*, zullen deze regels veelal *over-inclusive* (identificeren als een *change of control* bepaling, terwijl het er in de realiteit geen is) of *under-inclusive* (geen identificatie van een paragraaf als een *change of control* bepaling, terwijl het in de realiteit wel een dergelijke bepaling is) zijn. Het gebruiken van *expert systems* voor toepassingen zoals het identificeren van *change of control* bepalingen is bijgevolg onderworpen aan een aantal belangrijke beperkingen. Vooreerst, moet een menselijke expert talrijke uren besteden om op basis van deze regels een *expert system* op poten te zetten dat een aanvaardbare accuraatheid heeft. Vervolgens zal in dit proces de juridische expert moeten samenwerken met een IT-expert,

[2] E. ALPAYDIN, *Machine Learning*, Cambridge, MIT Press, 2016, 16.

hetgeen niet altijd eenvoudig is. Bepaalde IT-bedrijven proberen echter wel toepassingen te ontwikkelen die het mogelijk maakt om als IT-leek *expert systems* te ontwikkel, zoals Neota Logic en Berkeley Bridge. Ten slotte is het trouwens mogelijk dat de verwoording van een *change of control*-bepaling mettertijd in de rechtspraktijk wijzigt, waardoor de volledige oefening vaak opnieuw gemaakt moet worden.

Het is in dat verband dat *machine learning* bijzonder nuttig wordt. *Machine learning* maakt het immers mogelijk om de kennis van een menselijke expert te vervangen door data. *Machine learning* verwijst naar *"the ability of computer systems to improve their performance by exposure to data without the need to follow explicitly programmed instructions*"[3]. *Machine learning* systemen zijn in staat om tendensen te ontdekken en maken op basis van deze tendensen in de data-algoritmen. Het gevolg is dat geen menselijke expert nodig is om de relevante regels in het systeem te programmeren. Dit is nuttig omdat, zoals zojuist aangetoond, de expert deze regels niet kent of te veel tijd moet besteden om deze regels te ontdekken en in het systeem te programmeren.

Gesuperviseerd leren of hoe *machine learning* 'leert'

In het voorgaande hoofdstuk werd uiteengezet dat *machine learning* systemen in staat zijn om tendensen te vinden in data. De volgende vraag is bijgevolg: hoe worden deze patronen gevonden? Een computer moet eerst en vooral op een bepaalde manier feedback krijgen om te kunnen leren. Feedback kan aan de computer gegeven worden op drie onderscheiden manieren: *supervised learning, reinforcement learning* en *unsupervised learning*[4]. Wij focussen ons in dit hoofdstuk op *supervised learning* of gesuperviseerd leren.

Door het proces van gesuperviseerd leren toe te passen, wordt de computer twee reeksen data voorgeschoteld: een reeks trainingsdata en een reeks nieuwe data. De trainingsdata onderscheidt zich van de nieuwe data doordat de trainingsdata vooraf geïdentificeerde data bezit. Dit wil zeggen dat de reeks data die wordt gebruikt reeds geïdentificeerd is op basis van de gewenste uitkomst[5]. Een computer krijgt, bijvoorbeeld, 1.000 contractuele bepalingen voorgeschoteld dewelke een voor een (door een expert of een daardoor specifiek opgeleide derde partij) correct geïdentificeerd zijn als zijnde al dan niet *change of control*-bepalingen. De trainingsdata geven met andere woorden de computer de 'correcte antwoorden'.

De computer zal vervolgens elke instantie van de reeks data doorzoeken om een tendens te vinden die het mogelijk maakt om de juiste uitkomst te bepalen voor nieuwe data. De computer zal statistische modellen gebruiken en zal de parameters van deze modellen

[3] D. SCHATSKY & C. MURASKIN, "Demystifying Artificial Intelligence", *Deloitte University Press* 2014, 6.
[4] E. ALPAYDIN, *Machine Learning*, Cambridge, MIT Press, 2016, 694-695.
[5] H. SURDEN, "Machine Learning and Law", *Washington Law Review* 2014, 87, 93.

aanpassen om de accuraatheid van het voorspellingsmodel te verbeteren. Dit proces is repetitief en incrementeel in karakter. Deze parameters kunnen bepaalde attributen zijn (in het voorbeeld van de spam-e-mails, het land van oorsprong van de e-mail), de aanwezigheid van bepaalde woorden (in het voorbeeld van de *change of control* bepaling, de aanwezigheid van de woorden 'controle', 'wijziging' of 'controlewijziging') of de nabijheid van dergelijke woorden. Eens de computer het meest accurate model en de bijhorende parameters heeft 'geleerd' (hetgeen de computer kan weten gezien het trainingsdata heeft), is de computer klaar om het model toe te passen op een nieuwe reeks data. In het algemeen is het zo dat hoe meer aan de computer trainingsdata ter beschikking wordt gesteld, hoe robuuster het leermodel zal zijn.

Om te bepalen hoe performant een specifiek model is (gemeten op basis van diens mogelijkheid om accuraat te voorspellen), is het noodzakelijk om op empirische wijze het model te evalueren. Om het model te evalueren wordt gebruik gemaakt van een *k-fold cross validation*. Dit wil zeggen dat een klein deel van de trainingsdata opzij wordt gehouden alvorens het computermodel zichzelf begint te trainen (bijvoorbeeld 20 % van de totale reeks trainingsdata). Eens het model klaar is, zal het model toegepast worden op deze subgroep van de reeks trainingsdata om de prestatie van het model te evalueren. Dit is, zoals eerder aangeduid, mogelijk omdat de computer gebruik maakt van trainingsdata waarbij de juiste uitkomst van de data reeds door een persoon werd aangeduid.

Op basis van de zogenaamde statistische sensitiviteit en specificiteit kan men aan de hand van de *k-fold cross validation* de performantie van het model evalueren. Sensitiviteit en specificiteit slaan in dit verband op de zogenaamde *true positives*, *true negatives*, *false positives* en *false negatives*. *True negatives* (TN) is het totale aantal negatieve gevallen die correct als zijnde negatief werden voorspeld (een bepaling is geen *change of control* en werd ook zo correct door de computer geïdentificeerd). *True positives* (TP) is het totale aantal positieve gevallen die correct als zijnde positief werden voorspeld (een bepaling is een *change of control* en werd ook zo correct door de computer geïdentificeerd). *False negatives* (FN) is het totale aantal positieve gevallen die foutief als negatief werden voorspeld (een bepaling is een *change of control* en werd foutief niet als zodanig geïdentificeerd). *False positives* (FP) is het totale aantal negatieve gevallen die foutief als positief werden voorspeld (een bepaling is geen *change of control* en werd foutief als zijnde een *change of control* geïdentificeerd).

Een van de belangrijkste parameters om een model te evalueren op haar voorspellend vermogen is de accuraatheid van het model in kwestie. De accuraatheid is de ratio tussen correcte voorspellingen en het totale aantal voorspellingen (TN + TP)/(TN + TP + FN + FP). Het is belangrijk om in dit kader op te merken dat het niet realistisch is om een accuraatheid van 100% te verwachten van een *machine learning*-model. Daarenboven is het niet ongewoon dat

mensen een vooringenomenheid hebben in het nadeel van een computer en in het voordeel van menselijke experten, denkende dat een menselijke expert altijd accurater zal zijn in het voorspellen van een bepaalde uitkomst. Het is echter een realiteit dat, indien men een menselijk expert en een *machine learning*-model simultaan evalueert, dat het model accurater is. Bijvoorbeeld, Katz et al. (2014) creëerden een *machine learning*-model dat accurater uitspraken van het hooggerechtshof van de Verenigde Staten kon voorspellen in vergelijking met menselijke juridische experten. Het model voorspelde 69,7 % van de zaken correct over een periode van 60 jaar, in vergelijking met 59 % voor de juridische experten[6]. Hoewel het verschil tussen beiden niet drastisch is, toont het wel aan dat er geen reden is om een vooringenomenheid te hebben tegen een *machine learning*-model. De enige manier om de correcte analyse te maken over wie het best kan voorspellen, is door de proef op de som te nemen.

Beslisbomen

Een voorbeeld van een model dat een computer kan gebruiken om tendensen te vinden in data is door het gebruik van beslisbomen (*decision trees*). Dit model is een van de meest eenvoudige en succesvolle modellen voor *machine learning*-toepassingen[7]. Beslisbomen zijn ietwat gelijkaardig aan een *expert system* in de zin dat ook zij gebruikmaken van 'als-dan'-verklaringen. Het verschil is uiteraard dat, zoals hierboven reeds werd uiteengezet, de computer de parameters die deze 'als-dan'-verklaringen bepalen incrementeel kan aanpassen om op die manier tot het beste resultaat te komen. Het is dus niet de menselijke expert die deze verklaringen 'manueel' – op basis van diens kennis – bepaalt. Beslisbomen kunnen gevisualiseerd worden als meerdere *nodes* die elk vertakken naar nieuwe *nodes* (vandaar de benaming beslis*bomen*, waarbij elke *node* een blad of vertakking is van de boom).

Elke *node* in de boom is een 'als-dan'-verklaring gelinkt aan een specifieke verklaring of vraag[8]. Bijvoorbeeld, indien een e-mail meer dan 500 geadresseerden heeft, dan is het spam. Indien de e-mail in kwestie minder dan 500 geadresseerden heeft, dan is het geen spam. Het model zal vervolgens, na deze vertakking, een nieuwe *node* aanmaken (lees: een nieuwe 'test'). Door het geheel van deze *nodes* te combineren zal de computer in staat zijn om een (soms complex) geheel van regels te bepalen die de beoogde uitkomst (in dit geval: is de e-mail spam?) kunnen voorspellen met een zekere accuraatheid. Bijvoorbeeld, als een e-mail de woorden "gratis geld" (*node* 1), afkomstig zijn van land X (*node* 2) en meer dan 100 geadresseerden heeft (*node* 3), dan is er 90 % kans dat het spam is. Hoe deze beslisbomen toegepast worden in de context

[6] D. KATZ, M. BOMMARITO II EN J. BLACKMAN, "A General Approach for Predicting the Behavior of the Supreme Court of the United States", 2014, https://ssrn.com/abstract=2463244.

[7] E. ALPAYDIN, *Machine Learning*, Cambridge, MIT Press, 2016, 77 en S. RUSSELL EN P. NORVIG, *Artificial Intelligence: A Modern Approach*, Upper Saddle River, N.J., Prentice Hall, 2010, 697.

[8] K. ASHLEY, *Artificial Intelligence and Legal Analytics*, Cambridge, Cambridge University Press, 2017, 694-695.

van een *change of control*-bepaling, wordt later besproken in het hoofdstuk over *natural language processing*.

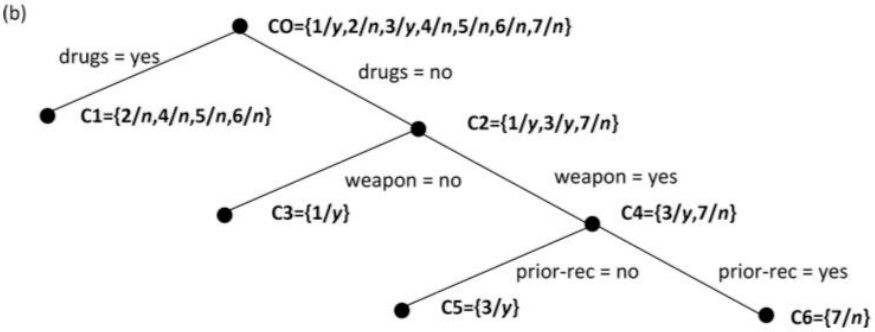
Hoe kleiner (lees: een laag aantal *nodes*) de beslisbomen, hoe gemakkelijker het is om de resultaten te interpreteren. Beslisbomen met meer dan 100 *nodes* kunnen daarbij bijzonder complex zijn om door de menselijke expert geïnterpreteerd te worden. Beslisbomen zullen in ieder geval op zo'n wijze gestructureerd worden dat de eerste *node* de meest belangrijke 'als-dan'-verklaring is[9]. Dit wil zeggen dat de eerste *node* het meest in staat is om de data te onderscheiden. Hierbij moet echter wel opgemerkt worden dat modellen gebaseerd op beslisbomen niet beogen om *nodes* te maken die 100 % accuraat zijn. Met andere woorden, de 'als-dan'-verklaringen zullen niet noodzakelijk passen voor alle data in de datareeks. Het model zal enkel trachten om *nodes* te maken die zo accuraat mogelijk zijn[10].

Voor een goed voorbeeld van beslisbomen kan verwezen worden naar Figuur 2, dewelke illustreert hoe beslisbomen opgesteld kunnen worden met betrekking tot het vrijlaten van beklaagden met borgtocht op basis van historische (beperkte) data. Merk op dat de eerste *node* in Figuur 2, met name of het misdrijf betrekking heeft op verdovende middelen, de meest belangrijke *node* is gezien het vier van de zeven instanties in de data kan voorspellen. De daaropvolgende *nodes* creëren meer precisie, en staan toe om ook de zaken te voorspellen waarbij geen betrokkenheid van verdovende middelen aanwezig is.

(a)

Case	Injury	Drugs	Weapon	Prior-record	Result
1	none	no	no	yes	yes
2	bad	yes	yes	serious	no
3	none	no	yes	no	yes
4	bad	yes	no	yes	no
5	slight	yes	yes	yes	no
6	none	yes	yes	serious	no
7	none	no	yes	yes	no

(b)



Figuur 2 Beslissingen met betrekking tot borgtocht waarop een beslisboommodel werd op toegepast[11].

[9] S. RUSSELL EN P. NORVIG, *Artificial Intelligence: A Modern Approach*, Upper Saddle River, N.J., Prentice Hall, 2010, 700.

[10] K. ASHLEY, *Artificial Intelligence and Legal Analytics*, Cambridge, Cambridge University Press, 2017, 114.

[11] K. ASHLEY, *Artificial Intelligence and Legal Analytics*, Cambridge, Cambridge University Press, 2017, 110.

Het *machine learning*-model leert niet echt maar gebruikt plaatsvervangende indicatoren

Wat betekent 'leren'? Wij definieerden artificiële intelligentie als de mogelijkheid voor een computersysteem om taken uit te voeren die normaal menselijke intelligentie vereisen. Voor wat betreft *machine learning* is de taak die de computer tracht te evenaren het leerproces. Met andere woorden, we proberen om de menselijke cognitie na te bootsen[12]. Het is belangrijk om in dit verband echter op te merken dat de huidige vaardigheden van *machine learning* (en artificiële intelligentie in zijn geheel) enkel de menselijke cognitie *evenaren*. Dat is omdat artificiële intelligentie zoals *machine learning* gebruik maakt van heuristieken en *proxies*, met inbegrip van statistische correlaties afgeleid van de tendensen in data, om een resultaat te bekomen dat gelijkaardige resultaten oplevert als wanneer een persoon een gelijkaardig probleem zou oplossen. Waar vandaag de dag een *machine learning*-model niet toe in staat is, is om abstracte concepten te begrijpen met betrekking tot de taken die ze toegewezen krijgt[13]. Neem als voorbeeld de vertaalprogramma's zoals Google Translate, dewelke gebaseerd zijn op *machine learning*. Deze systemen begrijpen niet de woorden die ze tegelijkertijd wel kunnen vertalen. Ze gebruiken data en statistiek om te bepalen hoe elk woord vertaald moet worden. Een ander goed voorbeeld zijn de aanbevelingsalgoritmen van Netflix en Amazon. De algoritmen van Amazon weten eigenlijk niet echt dat de boeken *Nineteen Eighty-Four* en *Brave New World* gelijkaardig zijn, maar op basis van eerder aankoopbedrag weet het algoritme wel dat de boeken door dezelfde personen gekocht worden en daardoor (als *proxy*) gebruik maakt van deze informatie om het een of het ander boek aan te raden aan personen die een van de twee boeken reeds hebben gekocht.

Het resultaat is dat het model niet het leerproces kan 'veralgemenen' op dezelfde manier waarop een persoon van vlees en bloed dat doet[14]. Het veralgemeend leren zoals hier beschreven kan worden aangeduid als *algemene* artificiële intelligentie. Artificiële intelligentie is vandaag de dag echter nog niet voldoende ontwikkeld om het te kunnen kwalificeren als zijnde algemeen. De sterktes van *machine learning* beperken zich momenteel tot beperkte taken, wat meteen ook de reden is waarom we de huidige versie van artificiële intelligentie vandaag de dag bestempelen als zijnde *beperkte* (*narrow*) artificiële intelligentie[15].

Ten slotte is het belangrijk op te merken dat de combinatie van deze vaststelling, met name dat *machine learning* niet de onderliggende concepten of tendensen begrijpt, en het feit dat *machine learning* gebruik maakt van historische data, mogelijk gevaarlijk is. De combinatie van deze twee eigenschappen zorgt er immers voor dat een bepaalde *bias* of vooringenomenheid

[12] E. ALPAYDIN, *Machine Learning*, Cambridge, MIT Press, 2016, 18.
[13] H. SURDEN, "Machine Learning and Law", *Washington Law Review* 2014 87, 95.
[14] R. BROOKS, "The Seven Deadly Sins of AI Predicitons", *MIT Technology Review* 2018, www.technologyreview.com/s/609048/the-seven-deadly-sins-of-ai-predictions/.
[15] J. MARGOLIES, R. RONANKI en D. STEIER, "Tech Trends 2018: The Symphonic Enterprise", *Deloitte Insights* 2018.

in stand gehouden kan worden. Door het gebruikmaken van historische data zal het *machine learning*-model er immers minder toe in staat zijn om bepaalde nieuwe evoluties op te sporen (die minder in hoeveelheid aanwezig zijn in de data), deze vervolgens niet in rekening nemen en bijgevolg verouderde regels toe passen op nieuwe data. Een menselijke expert zal in vergelijking met een *machine learning* meer in staat zijn om nieuwe tendensen op te sporen en deze bijgevolg toe te passen op nieuwe data. Daarmee verwant is ook het feit dat het gebruik van heuristieken en *proxies* soms problematisch kunnen zijn in een juridische context. Deze problematiek wordt reeds in de Verenigde Staten fors ter discussie gesteld[16], onder meer omdat een significant groter gedeelte van de gevangenispopulatie Afro-Amerikaans is. Indien men bijgevolg een *machine learning*-model wil creëren die de kans op criminele feiten kan voorspellen, zal dergelijke data in alle waarschijnlijkheid door het model als relevant beschouwd worden. De belangrijke vraag stelt zich evenwel of dit wenselijk is, en of dit niet een bepaalde problematiek in stand houdt.

In onze volgende bijdrage bespreken we hoe we *machine learning* kunnen toepassen op niet-gestructureerde data.

[16] ASSOCIATED PRESS, "Artificial intelligence is coming for both judges and defendants", *New York Post* 31 januari 2018 (29 augustus 2018), https://nypost.com/2018/01/31/artificial-intelligence-is-coming-for-both-judges-and-defendants/.