

Al voor het echie

Afscheidsrede, 10 november 2017

Dit is de tekst die uitgesproken wordt. Voor de noten en verantwoording wordt verwezen naar de Engelse versie in boekvorm.

Arnold W.M Smeulders

1. De grote vraag

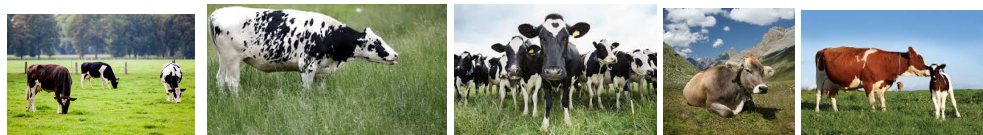
Dit is een plaatje uit de oratie van 22 jaar geleden.



uit de oratie, 23 jaar geleden

De vraag is: wat is het? Dit is het mysterie dat me mijn hele leven achtervolgd heeft. Hoe kunnen woorden de complexe wereld om ons heen beschrijven?

Om een oneindig complexe wereld vast te leggen, hebben we categorieën nodig. Zonder categorieën kunnen we de wereld niet begrijpen. Categorieën zijn een poging de wereld om ons heen te beheersen. In het Nederlands is dat duidelijk: *de wereld onder woorden brengen*. Door verwoording voegen we structuur toe aan de wereld zoals het ons uitkomt.



koeien

In computer visie hebben we onlangs het categorisatie probleem opgelost.

Neem voldoende voorbeeldfoto's elk met het label *koe*. Op elke afbeelding staan één of meer koeien (en misschien nog andere dingen). Met een kleine hoeveelheid plaatjes kan het nog lijken dat een koe wel bruin moet zijn. Een computer kan alleen de variaties leren in een categorie die het heeft gezien. Daarom is meer plaatjes beter. En, meer variatie is ook beter. Om te weten wat het is, is het niet nodig om te weten waar in het plaatje de koe staat.



geen koeien

Op een plaatje staan ook andere dingen zoals bomen, gras en lucht. De computer weet het verschil niet. Aanvankelijk kan aan de hand van de voorbeelden een koe worden gezien als groen en gestreept als gras. Daarom zijn er ook nog tegenvoorbeelden nodig met het label *geen koe*. Deze verzameling dient om *gras* uit de definitie van een koe te halen. Het kan niet tegelijkertijd onderdeel zijn van *koe* en van *geen koe*. En zo wordt de overgang van zwart-witte huid naar gras is een belangrijk deel van de visuele definitie van *koe*.

Van voorbeeld plaatjes en hun labels is de computer zo in staat om te leren wat essentieel is voor een koe. Op deze manier, zoals aangegeven door het positieve bewijs in geel, wordt een zeilboot gevonden door het water en door een verticale streep tegen lucht, veel meer dan de boot zelf (die blauw is gekleurd om negatief bewijs aan te geven). Er zijn immers veel andere objecten die van dezelfde materialen en vormen gemaakt zijn als een boot.

Visueel bewijs voor een zeilboot.

2. Zinnen om de wereld vast te leggen

In taal zijn de zelfstandige naamwoorden voor statische dingen, spullen, staten en scènes. Werkwoorden zijn voor toestanden, voor acties, voor transacties en voor oorzaken. *koe* is een zelfstandig naamwoord, en *grazen* is een werkwoord dat een toestand aanduidt. Zelfstandige naamwoorden en werkwoorden worden in de taal geheel verschillend behandeld, maar vanuit computer visie zijn ze vrijwel hetzelfde. Dit komt omdat ze beide de wereld categoriseren. Daarom wordt grotendeels één en hetzelfde algoritme gebruikt om zelfstandige naamwoorden en werkwoorden te beschrijven.

De taal kent 3.000 werkwoorden en meer dan 100.000 zelfstandige naamwoorden waarvan er 15.000 gangbaar zijn maar u kan op 1000 zelfstandige naamwoorden en 30 werkwoorden wel overleven.



Lakenvelder



Long Horn



Blaarkop

Het is duidelijk dat gewone woorden niet voldoende zijn om de wereld te vangen. De categorisatie van de wereld moet hiërarchisch zijn. Voor communicatie is een kern van categorieën voldoende. Subcategorieën van *koe* zijn een *Lakenvelder*, een *Friese blaarkop*, of *long horn*. Gedetailleerde subcategorieën zijn alleen nodig in kleine gemeenschappen, zoals beroepen, gezinnen en experts.



Koe bij het hek

Een verdere verfijning van de wereld kan niet worden bereikt door steeds verdere subcategorisatie. De lijst van categorieën zou dan net zo complex worden als de wereld zelf. Er

zijn meer expressie mogelijkheden nodig. En daarmee verschijnen voorzetsels ten tonele die zelfstandige naamwoorden kunnen verbinden: *de koe naast het hek*.



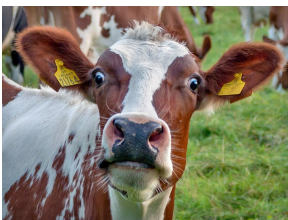
Welke kleur heeft de auto achter de aangegeven auto, blauw of rood?

We studeren nu op naamwoordelijke bepalingen. In de gewone taal zijn er maar zo'n 100 voorzetsels, maar elk van hen heeft een sterk karakter. Ze gedragen zich visueel heel anders dan zelfstandige naamwoorden en werkwoorden. In de foto, welke kleur heeft de auto achter de rode auto? Het antwoord hangt ervan af of *achter* betrekking heeft op het standpunt van de waarnemer of op de oriëntatie van de auto zelf. Voor deze foto zouden de meeste mensen stemmen voor rood, en sommigen voor blauw. Maar de vraag hier is: hoe kan de computer dit weten als het nooit doelloze gesprekken heeft gevoerd?

COW

**The cow and the tree stood together,
one happy to be, the other one waiting.** *Zin over een koe*

De expressiviteit van zinsneden groeit met een factor 10.000 voor elk zelfstandig naamwoord dat wordt toegevoegd, en met een factor 3000 voor elk werkwoord. Het getal doet er niet meer toe, omdat er nu alleen nog woorden worden geselecteerd als ze in de zin passen. Betekenis en metafoor verbinden de primaire en secundaire betekenissen van woorden. Ik weet niet van een methode in computer visie die metaforen aankan, maar interessant is het zeker.



Voorbij de taal

En ten slotte, om de wereld te beschrijven, verlaten we het rijk van de taalkundige beschrijvingen (categorieën, subcategorieën, bepalingen met voorzetsels, zinnen) en komen we bij de ruimte tussen de woorden. Om een zeer complexe wereld te beschrijven, moet taal wel dubbelzinnig zijn.

3. De waarnemer toevoegen

Er zijn maar een stuk of 1000 bijvoeglijke naamwoorden.

Er zijn favoriete combinaties met zelfstandige naamwoorden zoals *gele taxi* of *mooie zonsondergang* die zich gedragen alsof de combinatie een zelfstandig begrip is. Maar in het algemeen gedragen bijvoeglijke naamwoorden zich geheel anders dan zelfstandige naamwoorden, werkwoorden of voorzetsels.



Lang is afhankelijk van grootte maar ook van de context

Bijvoeglijke naamwoorden zijn geladen met persoonlijke beleving. Wat *lang* is in een *lange man*, hangt af van de grootte van de man, maar ook van de intersubjectieve beleving van lange mensen, de context en het toevallige standpunt.



oud schip



oud huis



oude auto



oude man

Zo wordt de perceptie van een *oude man* gedeeltelijk bepaald door de leeftijd van de waarnemer. Het bijvoeglijke deel is subjectief, en het zelfstandige deel is voornamelijk objectief. Een *schip* is visueel heel anders *oud* dan een *oud huis*. *Oud schip* verwijst meestal naar leeftijd, zelden naar een dysfunctionele staat. *Oude auto* kan beide betekenissen hebben.

Een *oude man* verwijst meestal naar leeftijd met een ondertoon van een dysfunctionele staat. Een *oud kind*, *oude kaas* en *oude universiteit* zijn visueel weer heel anders *oud* dan oude boot. Zijn er duizend verschillende visuele transformaties van het zelfstandige naamwoord, één voor elk bijvoeglijk naamwoord? Totdat we het mechanisme van overdracht begrijpen, moeten er 1000 keer 10.000 categorieën worden geleerd, een voor elk bijvoeglijk naamwoord bij elk zelfstandig naamwoord.

Het is duidelijk dat mensen veel meer kunnen zeggen over de visuele wereld dan computers. Een computer kan elke zin uitspreken door grote hoeveelheden zinnen te bestuderen. Een computer kan visueel categoriseren wat kan worden waargenomen uit voorbeelden. Een computer kan misschien dingen omschrijven die lijken op wat ze eerder heeft gezien. Maar, een computer kan niet beschrijven wat het nog niet eerder gezien heeft. Een computer heeft geen richting om onbekende dingen te leren. Een computer ontbeert de wil de wereld te beheersen door dingen te benoemen.



4. Leren zien

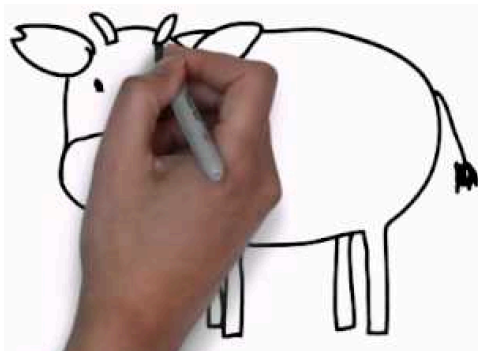
We keren terug naar de vraag: wat is dit? Laten we het eens zijn: het leven is eenvoudig, dit is (de mannelijke versie van) een koe.

Maar wacht! Zulke koeien bestaan helemaal niet.



Echte koeien lijken er op of en helemaal niet op

Getekende koeien zijn heel anders dan een echte koe. Hoe kunnen dan beide soorten hetzelfde perceptuele systeem doorlopen en terecht komen in dezelfde categorie: *koe*?



Is tekeningen lezen aangeboren of wordt het geleerd? Is het een culturele conventie zoals schrijven, of het is een natuurlijke weergave van de werkelijkheid? Feit is dat een tekening van de rand de meest informatieve lijn van een koe is die men kan tekenen.

Computer visie als onderzoeksgebied begon ongeveer 50 jaar geleden. Toen ik 40 jaar geleden begon vulden we een klaslokaal. De natuurkunde, microbiologie, pathologie, radiologie, de kunsten, geschiedenis, televisie, geografie en astronomie ontdekten dat ze voornamelijk uit beeldinformatie bestonden.

In het begin geloofden we dat het honderd jaar zou duren om visuele categorisatie op te lossen. In de intreerede uit 1995 beloofde ik dat ik ontslag zou nemen als het probleem geheel opgelost was. In 2004 zagen we voor het eerst een teken dat het antwoord beter was

dan random. Slechts tien jaar later kunnen we alle dingen in de wereld met menselijke precisie visueel uit elkaar houden: koeien, koelkasten, bruiloften, steden. Tot onze stomme verbazing, van collega's over de hele wereld en mijzelf, hebben we visuele categorisatie in het afgelopen decennium opgelost.

Daarbij hebben de beschikbaarheid van digitale data, digitale netwerken, en labels een belangrijke rol gespeeld, maar ook de open innovatie competities, de uitvinding van nieuwe algoritmen, en software- en hardware platforms.

Veel van deze factoren komen voort uit een gezamenlijke activiteit. Zelfs de uitvinding van algoritmen is vaak gezamenlijk. Het belang van het individu in de wetenschap wordt in deze dagen gemakkelijk overschat. In dit deel van de wetenschap is samenwerken, zowel lokaal als internationaal, even belangrijk als individuele vonken.

5. Wat is er nog meer te zien?

We keren nog eenmaal terug naar het majestueuze beeld waarmee we begonnen zijn.



Nu we het type kunnen herkennen, zijn wij vrij andere vragen te stellen: Is het zonnig? Hoe ziet ie eruit aan de andere kant? Wat is de kleur? Welke snelheid? Op welke grond?

Ja het was zonnig. De lijn aan de onderkant is dikker dan de bovenste lijn. In feite zou de lijn bovenop bijna onzichtbaar kunnen zijn. Die kan worden weggelaten. Dikke lijnen zijn een objectieve realiteit: de grens aan de onderkant is dikker vanwege de schaduw.

We weten niet hoe de koe er van de andere kant uitziet, maar een veilige gok is:



We hebben deze kant nog nooit gezien, maar volgens onze kennis opgedaan met veel dieren moet dit wel juist zijn. De kennis om deze vraag te beantwoorden komt niet uit deze foto maar uit externe bron.

Hetzelfde geldt voor de snelheid. De snelheid moet op het moment van de foto ergens liggen tussen 20 en 30 kilometer per uur. En de kleur is waarschijnlijk bruin, zeker niet paars, zoals sommige chocoladefabrikanten je willen doen geloven.

Dus er informatie uit de waarneming, al is dit beeld verschillend van elk beeld in werkelijkheid. En er is externe kennis in woorden of in iconen. En er is visuele verwachting. De scène is 22.000 jaar oud maar de waarneming, de kennis en de verwachting combineren tot een vrij goed idee van waar we hier mee te maken hebben.

Nu we categorisatie onder de knie hebben, zijn we bezig met uit te vinden hoe dingen in een video bewegen of hoe dingen er aan de andere kant uitzien. Maar in het algemeen is er nog weinig gedaan aan verwachting. Het menselijke observatiesysteem is in veel situaties te traag om te kunnen handelen. Visuele vooroordelen zijn nodig want zonder vooroordelen zou u veel meer informatie en veel meer tijd nodig hebben om te kunnen handelen.

De echte kunst is natuurlijk zonder vooroordeel door het leven te gaan. Het is interessant om te zien hoe Facebook met behulp van kunstmatige intelligentie juist het tegendeel dient door berichten te selecteren op basis van persoonlijke voorkeuren. Vooroordelen zijn onontbeerlijk om te (over)leven, maar tegelijkertijd maken voordelen het leven lastig.

6. Wat wordt er geleerd?

Computer visie kan nu woorden toekennen aan foto's. Het kan een koe herkennen. De betekenis van *koe* blijft een mysterie voor een computer, omdat het geen enkele andere kennis van een koe heeft. Koe is voor een computer gewoon een code voor een zelfstandig naamwoord.

Bij het leren werd de computer niet geïnstrueerd door regels om koeien te herkennen. De eerste 25 jaar volgden we het verkeerde pad. Het formuleren van regels *een koe is een middelgroot dier met vier poten met een gevlekte of bruine huid* werkt stomweg niet. Beschrijvende regels zijn alleen maar achteraf van toepassing, nadat u de koe al heeft herkend. Als de observatie samengevat moeten worden in regels zou de computer zeggen: *een koe is een gat in het gras met altijd opvallende details zoals de overgang van zwart-wit vlekken naar gras of bruin naar gras, en soms koeie-poten en/of koeie-ogen*. En u eigenlijk ook.

Als er geen voorbeelden zijn met koe-labels, kan de computer niet achterhalen wat er toe doet en wat er niet toe doet om een koe te herkennen. Er is dan een andere leerprocedure nodig. Als je een hele tekst aan een foto plakt in plaats van een label, wat is dan het resultaat? Zou dat leiden tot hetzelfde resultaat? Ik weet het niet, maar verwacht het antwoord binnenkort. Mijn gok is dat het beste resultaat bereikt wordt wanneer de kernbegrippen van de taal expliciet worden gelabeld, zeg duizend woorden (met daarbij inbegrepen koeien). Dat is de kern van de taal waarmee je dan alle andere dingen kunt uitleggen. Na het leren van de kern onder toezicht, kunnen de overige 15 tot 100 duizend categorieën voortvloeien uit onbenoemde ervaring en algemene teksten.

Als de computer niet werd geïnstrueerd om een koe te herkennen door het aantal poten te tellen, om bruine of zwart-witte huid te zoeken te zoeken, heeft hij deze feiten dan per ongeluk geleerd? Ja dat kan, maar alleen als we de computer vragen om ook nog te sorteren tijdens het leren. Neem schoenen. Wanneer we de computer dwingen om zoveel mogelijk schoenen te groeperen tijdens het leren, komt er een sub-categorie structuur aan het licht:



De computer heeft de schoenen zelf onderverdeeld in dezelfde ontwerpelementen. Tijdens het leren hebben we de computer niet verteld welke schoenen bij elkaar horen, en we hebben ook niet bij elk plaatje verteld welk type schoen het is. De computer heeft dat allemaal zelf geleerd door honderden schone voorbeelden van verschillende kanten te zien, en dan de resultaten sorteren in dezelfde subcategorieën. In het resultaat per rij zien we inderdaad hoge hakken, sneakers en pantoffels verschijnen. Schoenen is een sterk gestructureerde categorie. Dus dit resultaat uit 2015 kan niet zomaar worden gegeneraliseerd naar bijvoorbeeld natuurlijke gewassen. We studeren nu op vragen als: koeien en tafels hebben alle twee poten, wat hebben ze visueel gemeen? Het is een moeilijke vraag.

7. Wat kan geleerd worden?

Er zijn een aantal stadia van leren, elk met een verschillende stijl van instructie.

<i>method van leren</i>	<i>Instructie</i>	<i>resultaat</i>
1. imitatie	waar te kijken	reproduceer waar en waarom ergens te kijken.
2. bemoediging	aanmoediging	reproduceer de aanmoediging op vergelijkbare plaatjes.
3. supervisie	categorie labels	reproduceer labels voor onbekende plaatjes.
4.		
5. abstractie	regelboek	reproduceer de regels die een onbekend plaatje afhandelen.
6.		

Voor de mens is leren door imitatie het belangrijkste in de vroege stadia van het leven. Leren door imitatie leidt tot standaardkennis, zoals rechts rijden en hoe zich te bewegen in sociale kringen: doe zoals anderen doen. Maar ook later worden spiervaardigheden doorgaans geleerd door imitatie: *een komkommer snijden doe je zo*.

Zodra we de instructeur hebben geïdentificeerd, wordt meer geleerd door aanmoediging: *goed gedaan / doe dat niet*. Ook leren met aanmoediging is het makkelijkst zichtbaar op jonge leeftijd maar later in het leven ook nog belangrijk.

Kernbegrippen leren we door toezicht: *koe, tafel, huis*. Zoals we boven zagen, leren we koeien niet herkennen uit regels over poten, bruine, en een zwart-witte huid maar als geheel. Die regels voor een volledige beschrijving zouden oneindig complex zijn en de onderdelen zouden evengoed visueel ongedefinieerd zijn.



poten, huid, hoorns



geen benen, huid in orde, geen hoorns

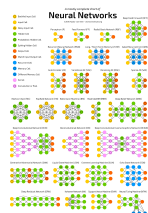
Een koe is uiteindelijk een koe. Zodra koe is geleerd kan specialisatie en abstractie volgen door regels: een Lakenvelder is een zwarte koe met een wit vel.

Voor computers is leren door supervisie de standaard. Wij hebben nodig

1. Heel veel voorbeelden en
2. Labels voor elk gegeven.

Voorbeelden kunnen van alles zijn: beelden, geluiden, de tijd tussen klikken op een toetsenbord, lichaamsparameters, de chemische samenstelling van de lucht, of wat dan ook. *Label* impliceert dat elk data-item een naam, een diagnose, een therapie, een parameter, een algemene beschrijving heeft, of welk doel dan ook. Foutvrije labels zijn goed, meer labels zijn beter ook als ze niet foutvrij zijn. Dit is diep leren: redeneren van de originele data aan het begin naar de labels aan het einde. Je stopt er een plaatje in, en er komt een diagnose uit, geleerd uit vele voorbeelden.

De voorbeelden en labels zijn sinds 2000 explosief gegroeid, omdat sensoren nu digitaal en mobiel zijn. En ze groeien nog steeds voor die bedrijven die midden in het netwerk zitten.



Strategieën geven aan hoe men kan redenen van begin tot einde.

Data alleen is niet genoeg. Voor het leerproces hebben we nodig:

3. Een repertoire van strategieën om te leren,
4. Een aanpak om een strategie om te zetten in dit bijzondere geval, en
5. Een maat om te bepalen of we vordering maken met het leren.

De maat geeft richting aan het leren: wat is een verbetering, wat is geen verbetering? Elk probleem heeft een eigen strategie, een eigen leeraanpak en een eigen maat. Er zijn vaste keuzes maar de (menselijke) kunst van het machine leren en computer visie vult nog steeds conferenties.

Om een koe te herkennen moesten we tussen 2004 en 2014 allerlei kleine stap voor stap uitvinden: Gauss-filters, selective search, SIFT-functies, soft assignment, codebooks, en zo verder. De recente convolutie netwerken omvatten van al deze stappen, omdat ze benaderd kunnen worden door convoluties. Het duurde vier decennia hard werk om te komen waar we zijn, maar nu kan een computer tenminste alle dingen om hem heen herkennen.

Dit alles zou in de academie gebleven zijn, als niet

6. De hardware platforms en

7. De software suites,

in het laatste decennium ook spectaculair waren verbeterd. Dat maakt het ook praktisch toepasbaar, in ieders leven.

Dit brengt de vraag naar voren: wat kan geleerd worden? Het primitieve antwoord is: als er voldoende hardware en handige software suites zijn, voldoende observatiedata, goed gelabeld, een goed repertoire van leeralgoritmen, en een goede aanpak met een goede maat voor het probleem, dan kan in principe alles geleerd worden.

8. Wat kan niet geleerd worden?

Het laatste antwoord kan diep teleurstellend zijn. Is er geen principe belemmering om alles te leren? Zal de AI ons overnemen? Is er een plek voor ons waar we mensen uniek zijn? Meer dan machines? Zullen we onze banen verliezen?

Let op: ik heb niet gezegd dat alles kan worden geleerd. Alleen als het probleem goed is ingekaderd onder alle zeven omstandigheden dan is er een kans.

Voor vertalingen van en naar Frans op het middelbare schoolniveau zijn er veel zinnen met dezelfde betekenis in beide talen beschikbaar. Er zijn labels: *dit is een goede/foute vertaling*. Er is een repertoire van leeralgoritmes voor vertalen met en zonder expliciete syntaxis. Door de ambiguïteit van de taal zijn maten wat een goede vertaling is nog niet eenvoudig maar op middelbare schoolniveau moet dat te doen zijn. Het zal nog wel even duren maar middelbare school Frans moet te leren zijn.

Hetzelfde geldt voor de middelbare schoolwiskunde: er zijn voldoende wiskunde problemen en labels: *dit is het goede antwoord*. Er is een repertoire van wiskunde problemen, en er zijn maten wat een goede en wat een slechte oplossing is. Een strategie om het leren te verbeteren zal reuze niet meevallen, maar ik denk dat ook wiskunde examens van de middelbare school te leren vallen.

Dit betekent overigens niet dat we wiskunde en Frans moeten staken maar dat is een ander onderwerp.

Een interessantere vraag is: wat kan niet worden geleerd? In een zoektocht naar objectieve resultaten, streeft de middelbare school naar doelstellingen die in belangrijke mate kunnen worden geformaliseerd. En voor de middelbare school zijn data en labels in overvloed aanwezig anders zou het niet op de middelbare school worden onderwezen. Bijna per definitie kunnen alle cursussen op de middelbare school met AI worden geleerd. Hetzelfde geldt voor human resource management dossiers en de selectie van kandidaten wanneer u denkt dat een volledig digitaal dossier de hele waarheid bevat.

Maar we zijn er nog niet.



Wanneer Facebook vraagt om een bericht te liken, bent u 1 seconde hun medewerker die wordt uitgenodigd om de labels te maken die zij missen. Ze hebben al de hardware, de software, het repertoire van leeralgoritmes, en de succesmaten. Daarom kunnen ze uit 250 likes je voorkeur beter te voorspellen dan je beste vriend zou kunnen. Niets nieuws, een goede winkelier zou in de 19e eeuw haar klanten goed genoeg kennen om hun gedrag te voorspellen. Maar de schaal, invloed en concentratie van Facebook is helemaal nieuw en onvoldoende doordacht.

Dus de vraag is: wat is er moeilijk te leren?

1. *Als er geen expliciet doel is, zijn er geen labels en is er geen leren onder supervisie.* Wat je na schooltijd aan kennis verwerft, is veel moeilijker te leren voor een computer.

2. *Als er geen gegevens zijn, is er geen leren onder supervisie.* Er zijn geen gegevens wanneer het evenement heel zeldzaam is. En hoe groter het detail in de vraag, hoe groter de kans dat die vraag nog niet eerder is gesteld. Er is nieuw onderzoek zonder voorbeelden, maar de vragen zijn dan nog erg dichtbij dingen die we eerder hebben gezien.

3. *Het huidige leren onder supervisie leert weinig van soortgelijks maar verschillende problemen.* De grootste lege verwachtingen ontstaan als we ervan uitgaan dat de computer slim is omdat het een wereld go-speler kan verslaan. Ja dit is slim, zonder twijfel, maar we gaan er onterecht vanuit dat de computer nu ook slim is in veel andere dingen. Dat is niet zo. Er is een heel beperkte kennisoverdracht naar andere problemen. Om een koe te herkennen is één ding, maar in de huidige AI helpt dat nauwelijks bij het herkennen van varkens of dieren in het algemeen of Lakenvelders in het bijzonder.

4. *Er is weinig generalisatie.* Frans van de middelbare school, Facebooks likes en gemiddelde wiskundesommen kunnen worden geleerd. Maar allemaal apart, ieder voor zich. Er wordt geen kennis gedeeld tussen die oplossingen. Wat moeilijk te leren is, zijn niet-expliciete doelen (het meeste wat u doet), ongedocumenteerde doelen (wat u vaak doet), verwante doelen (wat u niet doet) en inzicht (wat u nooit gedaan hebt).

9. Wat is er eigenlijk wel geleerd?

Leren met supervisie is eigenlijk leren te reproduceren. In leren met supervisie wordt een expliciet doel dat verborgen zit in de gegevens nagebootst. De computer leert eigenlijk zelf niks; het reproduceert menselijke kennis. Vandaar dat wat er geleerd wordt veelal gangbare kennis is, de gebruikelijke gevallen, de routine waarvan er ook veel gegevens zijn.

Maar voor dingen die 1 keer gebeuren, zoals in dit Youtube filmpje, helpt reproductie niet.





Een Youtube-filmpje van Mathieu Bourgarel.

Stelt u zich voor dat deze situatie werd gegeven aan een robot om op te lossen. Een domme kleine olifant is in een pijp gevallen. Een oudere olifant wil helpen. Wanneer is ze geleerd sociaal te zijn? De redder doet twee pogingen. Waar heeft ze geleerd niet op te geven? Ze verzint een ingenieuze oplossing. Waar heeft ze deze strategie vandaan? Laten we het over twee dingen eens zijn. Dit is serieus intelligent. Als een robot het zo zou doen, zou u het zeer intelligent noemen. En dit is serieus eenmalig. Dit geval wordt nooit geleerd op de olifanten school.

Unieke momenten (improvisatie, creativiteit, redding) en de combinatie van vele dimensies (politiek, mode, vertrouwen) kunnen niet door nabootsing worden geleerd. Kennis vervat in regels kan hier een expliciete of impliciete hulp zijn, maar die moeten dan wel worden verteld naar sensorische waarneming en instructie. Kennis-regel-systemen kunnen met enig succes redeneren, en data-systemen kunnen reproduceren, maar ze leren nog heel weinig van hun wederzijdse wijsheid. Wij mensen kunnen makkelijk schakelen tussen beide systemen, maar de computer nauwelijks.

We zagen al eerder dat de 3^e stap in de AI, leren door supervisie, elk probleem op zichzelf staat: koeien leren herkennen zegt niks over varkens, Lakenvelders of viervoeters in het algemeen. Maar wat nu in de AI het meest ontbreekt, is de 4^e stap, condensatie: hoe u expliciete kennis uit sensorische ervaring haalt. En de 6e stap is ook nauwelijks ontwikkeld: hoe kan je nieuwe kennis produceren in plaats van bestaande kennis reproduceren?

<i>Leren door</i>	<i>Instructie</i>	<i>Resultaat</i>
1. Imitatie		Reproduceer waar en waarom ergens te kijken.
2. Bemoediging		Reproduceer de aanmoediging op vergelijkbare plaatjes.
3. Supervisie		Reproduceer labels voor onbekende plaatjes.
4. Condensatie	 hebben vier poten.	Reproduceer gedeeltelijke en inter-subjectieve invarianten.
5. Abstractie	Een Lakenvelder is een koe met een vierkant wit veld op hun rug.	Reproduceer de regels die een onbekend plaatje afhandelen.
6. Uitvinding	- geen -	Produceer (persoonlijk) nieuwe kennis bij voorbeeld door metaforen.

10. AI voor het echie?

De vraag is of een autonome auto een allround bestuurder kan zijn. Ja, geassisteerd rijden zal komen. Tussen de lijnen blijven en nergens tegenaan botsen gaat makkelijk voor een computer dan voor een mens. Autonoom rijden zal er ook komen zolang het situaties betreft die de auto kan reproduceren uit wat het heeft geleerd. Op de snelweg in slaap

vallen, onwaarschijnlijk. Maar voor het samenvoegen van twee rijstroken in Amsterdam zijn andere vaardigheden nodig.

Er is een fundamenteel verschil is tussen wat de mens heeft geleerd en wat de machine leert. Mensen beginnen nederig in stap 1 en dan werken we onze weg naar stap 5 van abstractie en 6 inventie. Computers starten meestal op losse problemen op stap 3 of 5 (lenen daarbij menselijke kennis) en hebben geen idee van stap 1, 2, 4 en 6 en hoe dat te bereiken.

Als computers hetzelfde pad afgaan, van 1 naar 2 dan 3, zijn ze dan meer menselijk? Ik weet niet of dat haalbaar is en dan nog weet ik het antwoord niet. Maar wat ik wel weet is dat zolang ervaring met leven belangrijk is voor een goed functioneren, computers geen schijn van kans hebben om het over te nemen.

11. De stand van zaken

Mijn toga kan hier met pensioen. Aan het einde van deze tocht wil ik nog graag een paar woorden zeggen als privépersoon.

Internet heeft veel deuren op de wereld geopend en nog veel meer ramen. Dit heeft verwarring gebracht. Het voelt alsof wat elders gebeurt hier een onderdeel is.

Het effect van Internet is positief en negatief. Met één blik op de wereldgeschiedenis is er weinig reden om te geloven dat internet alleen maar goed zal doen. Binnen 30 jaar is het dagelijkse perspectief voor de meesten op aarde radicaal veranderd. Minuscule details van het leven die alleen te begrijpen zijn in context zijn nu toegankelijk voor iedereen elders. Dat is een shocktherapie van ongekeerde omvang. Enkele menselijke driften zoals jaloezie, xenofobie en privacy staan onder constante stress. Dus doordenkend is uiteindelijk een volledig vrij internet niet duurzaam- Regelgeving is niet perse het antwoord, maar toch moeten culturele gewoontes en omgangsvormen die in de gewone wereld al eeuwen bestaan op internet nog worden opgelost.

Gegevens zijn nu in een periode van 15 jaar digitaal geworden. En wat doen we? Wij geven ze weg. Dat is niet verstandig. Digitale data kunnen het best als handelswaar worden gewaardeerd. Het duurde 150 jaar om de handel in materiële goederen in wetten te regelen. Het heeft veel moeite gekost om de wetgeving uit te breiden naar niet-materiële goederen. Maar voor digitale data is er bijna geen kader. Google las tot voor kort mee met uw e-mails *om u beter te bedienen*. U stemde ermee in. Zou u ook instemmen als de postbode uw brieven las? Of het hotel waar u verblijft? Wat is de waarde van een klik? Wie bezit de klik? Onder welk land valt de klik? Is de juiste zorg voor uw klik inbegrepen? De kracht van gegevens zal onvermijdelijk leiden tot digitale data-scheiding: grenzen aan wat door één partij kan worden verzameld. En grenzen aan wat ermee gedaan kan worden, en wat moet worden overgedragen aan anderen.

AI kan alles oplossen. Oh nee, dat is niet waar. AI doet niets met de kennis die we op de kleuterschool hebben geleerd, waar we belangrijke kennis over het leven hebben opgedaan door eerst een jongen aan zijn haar te trekken en dan zien wat er gebeurt. Een computer heeft dat nog nooit gedaan.

12. De universiteit

Nog even terug naar de universiteit.

Het is merkwaardig dat informatica, 50 jaar na haar conceptie, nog gezien wordt als een jonge wetenschap. Hoeveel generaties moet je erover doen voordat ze je zien staan? OCW,

KNAW, NWO, en de adviesraden worden nog gedomineerd door leuke vakgebieden met leuke resultaten, maar het is weeffout in het systeem om geldstromen van nature zo te verdelen alsof het voor altijd in sterren staat geschreven.

De vergelijking van de kwaliteit tussen wetenschapsgebieden is notoir lastig. Absolute aantallen citaties zijn misleidend en beoordelingen door zelf-geselecteerde commissies volgen al gauw de stijl van waarderen binnen het vak- of deelgebied. Een beter criterium is citaties binnen het eigen vakgebied of grootschalige beoordeling door vele vakgenoten, maar dat beide methoden bieden geen houvast tussen vakgebieden. Sommige vakken zijn duurder dan andere. En langdurige structurele geldstromen leiden tot betere waarderingen, en dat is maar goed ook. Maar ook dat biedt dus geen houvast in de vergelijking tussen vakgebieden, sterker nog het trekt het alleen maar nog schever. Geen van deze indicatoren zegt of een vakgebied ook nu innoveert of op de lange termijn maatschappelijk noodzakelijk is. In de huidige mechanismen van beleidsvorming in de wetenschap is er te weinig ruimte voor innovatie van het wetenschapssysteem zelf. Ook in Nederland staan meer-objectieve internationale beoordelingen van vakgebieden los van de ontvangen geldstromen.

In deze rede geen jammerklachten over het Maagdenhuis want het Maagdenhuis dat zijn we zelf; geen jammerklacht over de publicatiedruk, want die zit vooral klem in de kaken van de eigen ambitie.

In deze oratie geen jammerklachten op het Maagdenhuis, want het Maagdenhuis dat zijn we zelf. Geen woord gewijd aan de verambtelijking, want die hebben we zelf toegelaten. Feit is dat het bestuur van een universiteit zelden soepel verloopt. Mijns inziens ligt dat aan het bestuursstructuur, maar evenzeer aan de wil van Den Haag te veel te willen sturen, aan de beter-wetende aard van het bedrijf en aan de onmogelijkheid de wetenschappen te ordenen in een objectieve rangorde. En daarmee zijn we weer terug waar ik begon: zij die meester zijn over het beeld en de taal zijn de meesters van de jacht, ook in de wetenschap, en zo zal het wel blijven.

citaat uit de vorige oratie

13. Dankwoord



In de eerste plaats wil ik de AIO's bedanken voor de vele plezierige momenten die we hebben doorgebracht. Geen betere tijd doorgebracht dan ongelijk hebben in een wetenschappelijke discussie. Wetenschap is niet altijd makkelijk, maar wie zou de promotie gratis willen krijgen? Sommigen worden beter bediend in dit spel door te knagen aan teveel zelfvertrouwen, en anderen zijn beter geholpen met richting wanneer het moeilijk is om te kiezen. Wetenschap is een vreselijk beroep. Op het

moment van tevredenheid met het resultaat, wordt alles weer in de waagschaal gelegd om nieuwe vragen te stellen.

Ik dank mijn promotor, Ab Bakker, van wie ik heb geleerd dat in strategische discussies grondslagen het moeten winnen van emotie. En ik dank mijn andere promotor, Bas Ploem, van wie ik heb geleerd dat emotie het moet winnen van grondslagen. Daarna was het leven eenvoudig.

Ik dank mijn medereizigers, Frans Groen, Marcel Worrying, Theo Gevers, Cees Snoek en Dennis Koelma. Voor de zondvloed van data hebben we een ark gebouwd. Met Maarten de Rijke en Max Welling stoomt nu een kleine vloot op naar een zonnige toekomst.

Ik dank Harro Stokman en Jan Willem Klerkx voor de durf om bedrijven te beginnen op wetenschappelijke fundamenteën. Het gaat de universiteit revolutionair veranderen uiteindelijk ook in Nederland.

Ik dank Peter Apers voor onze genoeglijke omzwervingen door de wonderde wereld van het grote gelijk. En mijn dank gaat naar de makkers Inald Lagendijk, Geleyn Meijer en Johan Vos die besturen makkelijk maken. Jan Bergstra dank ik voor de eerlijkheid die uiteindelijk overwint.

Ik zal ook mensen hebben geïrriteerd. Ik hoop dat het redelijk was of, als het onredelijk was, dat u het kon hebben. Anders zou ik me schamen.

Ik dank Virginie Mes en de anderen zoals Mieke van de Berg en Suna van Bekkum, de tochtgenoten. Geen zee te hoog, geen woestijn te lang, altijd aangenaam. Elke dag weer bedroefd naar huis omdat de werkdag was afgelopen.

En tot slot dank ik mijn vrouw, mijn kinderen en mijn vrienden. Van hen heb ik het meest geleerd.

Ik heb gezegd.

Referenties

[1] *What is the spatial extent of an object?* J.R.R. Uijlings, A.W.M. Smeulders, R.J.H. Scha, CVPR, Miami 2009, see also [20].

[2] *Spatial-Aware Object Embeddings for Zero-Shot Localization and Classification of Actions*, P. Mettes and C.G.M. Snoek, ICCV, Venice, 2017.

[3] *Visual-Concept Search Solved?* C.G.M. Snoek, A.W.M. Smeulders, IEEE Computer, 43-6, 76-78, 2010.

[4] *Mediamill at TRECVID 2013: searching concepts, objects, instances and events in video*. C.G.M. Snoek, K.E.A. van de Sande, D. Fontijne, A. Habibiyan, M. Jain, S. Kordumova, Z. Li, M. Mazloom, S.L. Pintea, R. Tao, D.C. Koelma, A.W.M. Smeulders, NIST, Washington, 2013.

[5] *Tracking by Natural Language Specification*, Z. Li, R. Tao, E. Gavves, C.G.M. Snoek, A.W.M. Smeulders, CVPR, Honolulu, 2017.

[6] *Déjà vu motion prediction from static scenes*, S. Pintea, J. van Gemert, A.W.M. Smeulders, ECCV, Zurich, 2014.

[7] *Visual tracking, an experimental survey*, D.M. Chu, R. Cucchiara, S. Calderara, A. Dehghan, M. Shah, trans PAMI, 36-7, 1442-1468, 2014.

[8] *Siamese instance search for tracking SINT*, CVPR, Las Vegas, 2016.

[9] *Locality in Generic Instance Search from One Example*, R. Tao, E. Gavves, C.G.M. Snoek, A.W.M. Smeulders, CVPR, Columbus, 2014.

[10] *Attributes and categories for generic instance search from one example*, R. Tao, A.W.M. Smeulders, S.F. Chang, CVPR, Boston, 2015.

[11] *Socializing the Semantic Gap: A Comparative Survey on Image Tag Assignment, Refinement and Retrieval* X. Li, T. Uricchio, L. Ballan, M. Bertini, C.G.M. Snoek, A. Del Bimbo, ACM Computing Surveys, 49-1-14, 1-39, 2016.

- [12] *Dynamic Steerable Blocks in Deep Residual Networks*, J.H Jacobsen, B. de Brabandere, A.W.M Smeulders. BMVC, London, 2017.
- [13] *Latent embeddings for zero-shot classification*, Y Xian, Z Akata, G Sharma, Q Nguyen, M Hein, B Schiele, CVPR, Las Vegas, 2016.
- [14] *Pooling Objects for Recognizing Scenes without Examples*, S Kordumova, T Mensink, C.G.M Snoek, ACM ICMR New York, USA, 2016.
- [15] *Modeling video evolution for action recognition*, B Fernando, E Gavves, J.M Oramas, A Ghodrati, T Tuytelaars, CVPR, Boston, 2015.
- [16] *Content-based image retrieval at the end of the early years*, A.W.M Smeulders, M Worring, S Santini, A Gupta, R Jain, IEEE trans PAMI, 22-12, 1349-1380.
- [17] *Color-based object recognition*, T Gevers, A.W.M Smeulders, Pattern Recognition 32-3, 453-464, 1999.
- [18] *The challenge problem for automated detection of 101 semantic concepts in multimedia*, C.G.M Snoek, M Worring, J.C van Gemert, J.M Geusebroek, A.W.M Smeulders, ACM MM, Santa Barbara, 2006.
- [19] *Horus reference manual*, D Koelma, 1995-2015.
- [20] *Selective search for object recognition*, J.R.R Uijlings, K.E.A van der Sande, T. Gevers, A.W.M Smeulders, IJCV 104-2, 154-171, 2013.