

'INSTELLINGEN MOETEN WEL IN STAAT ZIJN DE WAARDE ERUIT TE HALEN'

Big data in de zorg: niet alleen verzamelen

Big data zijn hot, ook in de gezondheidszorg. In de Verenigde Staten worden ze zelfs al verzameld door een supercomputer, Watson. Maar als we er echt van willen profiteren, moet het datadenken doorsijpelen in alle haarvaten van een instelling, zegt Nicky Hekster, big data-deskundige van IBM.

Watson onderzoekt 25 patiënten met een multiform glioblastoom, een fatale vorm van hersenkanker. Hij vergelijkt het genoom van iedere patiënt met het DNA van de tumor, en onderzoekt in de literatuur welke geneesmiddelen de meeste kans van slagen zullen hebben, en welke onderzoeken nog nodig zijn. Het bijzondere is: Watson is geen arts, maar een computer van IBM, die wordt ingezet in het New York Genome Center bij de behandeling van deze zeldzame hersentumor.

De analyse van al die genetische gegevens in combinatie met de medische literatuur is een bovenmenselijke taak, en voor deze supercomputer is het goed te doen. De computer treedt op als adviseur, wat betekent dat artsen geen 'absolute' antwoorden krijgen van de computer, maar adviezen met een kansberekening plus achterliggende argumentatie. Artsen geven vervolgens feedback aan de computer over wat zij vinden van de adviezen, en daarvan leert Watson op zijn beurt weer bij.

Het is een aansprekend voorbeeld van

big data, hét modewoord in de ICT van dit moment. Big data gaat over het geautomatiseerd analyseren, en correleren van grote hoeveelheden diverse data, om daar vervolgens kennis, wijsheid en voorspellingen uit te halen. De geneeskunde kan wel wat hulp gebruiken van big data-analyses, volgens IBM, want 80 procent van de artsen besteedt niet meer dan vijf uur per maand aan het lezen van de medische literatuur, en artsen zitten één op de vijf keer (deels) fout bij het stellen van de diagnose, aldus de analyse van IBM. Maar met een goed ICT-netwerk alleen ben je er nog niet. 'Je kunt wel heel veel data en een mooie dataopslag hebben, maar dan heb je nog geen kennis of wijsheid', waarschuwt Nicky Hekster, technical leader Healthcare & Lifesciences voor IBM Benelux. 'Zorginstellingen verzamelen van oudsher veel gegevens, maar ze zijn vaak onvoldoende in staat daar de waarde uit te halen.'

Vier uitdagingen

Wie waarde wil halen uit een databank met zorggegevens krijgt te maken met vier uitdagingen, oftewel de vier v's,

volgens Hekster. De eerste v staat voor *volume*, want er is een minimum aan volume nodig om betrouwbare uitspraken te kunnen doen. Dat zit over het algemeen wel goed in de zorg. Het Amerikaanse gezondheidszorgsysteem alleen al produceerde 150 exabytes in 2011, (1 exabyte = 10^{18} bytes). Hekster schat de productie door Nederlandse ziekenhuizen op 2 petabyte per jaar (1 petabyte = 10^{15} bytes). De hoeveelheid data verdubbelt elke vijf jaar. Medische beelden vragen overigens relatief veel bytes – ze nemen 30 procent van alle computeropslagcapaciteit ter wereld in beslag.

Tweede uitdaging is de *variëteit* in al die gegevens. Een medisch dossier bestaat vanuit ICT-oogpunt uit een bonte verzameling van bestanden, afkomstig van de radiologie, cardiologie, pathologie, het klinisch-chemisch lab, uit gesprekken met de patiënt, uit verpleegkundige dossiers, enzovoort. Zelfs binnen bijvoorbeeld CT-scans bestaat variatie, omdat de standaarden die ervoor zijn afgesproken vaak niet zijn nageleefd. Conclusie:

Met een goed
ICT-netwerk
alleen ben je er
nog niet



BIG DATA – VIER NIVEAUS

- 1 Met de 'simpelste' *databanken* kun je rapportages maken, of een antwoord vinden op een vraag als: hoeveel heupoperaties vonden plaats in 2013?
- 2 Een trapje complexer krijg je de zogenaamde *datawarehouses*, zoals de klinische databanken van de Dutch Institute for Clinical Auditing. Ze geven bijvoorbeeld spiegelinformatie aan deelnemers. Hierin kun je antwoord vinden op de vraag: hoeveel heroperaties heeft mijn ziekenhuis ten opzichte van ziekenhuizen in de omgeving? Op dit niveau bevindt de Nederlandse zorg zich nu, volgens IBM.
- 3 Nog een stap verder komt de *beslis-ondersteunende software*. Op basis van klinische uitkomsten en literatuur krijgt de arts adviezen over een behandelplan. Hiermee krijg je antwoord op de vraag: hoe doelmatig is onze richtlijn voor heupoperaties in de praktijk?
- 4 Het kan echter nog slimmer, met *voorspellende analyses*. De computer kan dan met de databank op de patiënt toegesneden adviezen geven op basis van grote hoeveelheden gegevens. Hoeveel risico loopt mevrouw B van 62 jaar op een heroperatie? Als er ook nog social media-gegevens bij worden gebruikt, en de tellingen van de stappentellers van soortgelijke lotgenoten, kan er ook worden voorspeld wat deze mevrouw weer zal kunnen over een jaar.

80 procent van alle zorgbestanden is ongestructureerd en het kost een computer erg veel moeite om binnen deze ongestructureerde data patronen te herkennen. Alleen al de 'platte tekst' in medische dossiers vormt een uitdaging. De Watson-computer van het voorbeeld kent bijvoorbeeld de taxonomie van de Engelse taal goed, maar zou hopeloos in de war raken als hij teksten in het Nederlands zou moeten analyseren. Uitdrukkingen of ambiguïteit in onze taal zijn hem niet bekend. De computer moet daarbij bewust zijn van context. 'Zeker in een medische omgeving moet je dat echt trainen', aldus Hekster. 'En dat kunnen dit soort computers ook steeds beter, ze leren al doende.'

Het is de vraag of het Nederlands taalgebied niet te klein is voor een dure toepassing als Watson, met een totale investering van 1 miljard dollar. Kleine onderdelen van de Watson-computer zijn al wel vertaald naar het Nederlands. In het UMC Groningen werken de oncolo-

gen bij kanker van het hoofd-halsgebied bijvoorbeeld met de technologie. De computer zoekt naar correlaties tussen hoofd-halskanker en andere aandoeningen, in een cohort van zeshonderd geanonimiseerde patiënten. De databank kent een grote variatie, met onder meer ontslagbrieven, verwijsbrieven, medietiehistories, behandelingen, diagnosemethodieken en protocollen. De eerste resultaten worden binnenkort gepubliceerd.

Geen absolute waarheid

Bij big data wordt overigens vaak breder gekeken dan alleen naar medische dossiers en wetenschappelijke literatuur. Ook gegevens uit social media zoals Facebook of Twitter kunnen bijdragen aan nieuwe inzichten voor een patiënt. De Inspectie voor de Gezondheidszorg heeft laten uitzoeken hoe zij sociale media kan gebruiken voor toezicht. Daarbij genereren de medische apps op de telefoon voortdurend meetgegevens. Dit

Dankzij big data-analyses kunnen artsen anticiperen op de voortekenen van een infectie

soort data kunnen inzicht geven op populatieniveau in hoe een patiënt de behandeling heeft ervaren.

Derde uitdaging voor big data-wijsheid is de snelheid (*velocity*) waarmee de computer zijn analyses maakt. Hekster: 'In het Canadese Hospital for Sick Children worden big data gebruikt bij couveusekinderen op de intensive care. We weten bijvoorbeeld dat infecties bij premature baby's zich aankondigen met een verhoogde hartslag. Dankzij big data-analy-



HET GENOOM IN DE CLOUD

IBM is niet de enige partij die toepassingen maakt voor big data. ICT-bedrijf Google benadert universiteiten en ziekenhuizen, in een poging om gegevens te mogen opslaan van het menselijke genoom in een beveiligde 'cloud'. Google vraagt 25 dollar per jaar van de wetenschappers voor de opslag van een genoom. Onder de naam Google Genomics wil het bedrijf een enorme genendatabank opzetten, waar onderzoekers experimenten in mogen doen. De hoop is dat bijvoorbeeld iemand met longkanker in deze databank zijn genoom en dat van zijn tumor kan vergelijken met miljoenen andere genomen, en zo de gouden tip krijgt voor het meest effectieve geneesmiddel. Op dit moment zijn er al 3500 genomen opgeslagen door Google uit openbare bronnen, en daar wil het bedrijf genomen uit private bronnen aan toevoegen.

De *Cancer Genome Atlas* van het National Cancer Institute in Amerika omvat 2,6 petabytes. (1 petabyte is 1 miljoen gigabyte.) Het instituut heeft vorige maand 19 miljoen dollar betaald om dit bestand in de cloud te zetten; zij hebben dit gedaan bij Google en Amazon. Daarnaast zijn ook IBM en Microsoft bezig met genoomdatabanken, naast wat kleinere spelers.

Google zet de laatste tijd meer stappen op het gebied van de gezondheidszorg. Met het project *Baseline Study* wil het bedrijf de wetenschap helpen door het aanbieden van big data over gezonde mensen. Hierin zullen menselijke genomen worden opgeslagen, naast informatie uit meters die mensen bij zich dragen in hun mobiele telefoon.

ses van de klinische metingen bij deze kinderen kunnen artsen anticiperen op de voortekenen van een infectie. Dit stelt wel hoge eisen aan de snelheid van de datastroom en de snelheid van analyse, want anders ben je alsnog te laat.' Ten slotte is de waarheidsgetrouwheid, of onzekerheid van de data (*veracity*) een uitdaging voor computers. In medische gegevens bestaat er geen absolute waarheid; de ene bron is waarheidsgetrouwer dan de andere, en dat is voor een compu-

ter lastig te wegen. Dit is echt iets wat de computer moet 'leren' aan de hand van feedback van mensen.

Gebruik in Nederland

Een echt voorbeeld van big data in Nederland is het Center for Personalized Cancer Treatment. Daar vergelijken instellingen het genetisch materiaal van duizenden patiënten. Uit de analyse rolt een advies voor chemotherapie dat misschien niet geschikt is voor de grote bulk van de

‘Ik ben voor het bevrijden van alle data’

PATIËNTGEGEVENS TE KOOP

Big data is ook big money. In Engeland is ophef ontstaan nadat de National Health Service geanonimiseerde patiëntgegevens had verkocht aan commerciële partijen. Critici zeggen dat de data weliswaar anoniem zijn, maar dat door combinatie van verschillende databanken toch te achterhalen is om wie het gaat.

Het mag van de wet, volgens Sjaak Nouwt, adviseur gezondheidsrecht van de KNMG, zolang de gegevens inderdaad maar anoniem zijn. Hij is het echter wel eens met de critici dat anonieme gegevens, mede door de huidige technologische mogelijkheden, steeds vaker herleidbaar tot individuele personen kunnen worden gemaakt. ‘En een arts of instelling mag herleidbare patiëntgegevens alleen doorgeven aan iemand anders, met toestemming van de patiënt’, benadrukt Nouwt. Voor anonieme gegevens is dat niet nodig, maar het is wel belangrijk om de gegevens zo anoniem mogelijk te maken, zodat ze niet achteraf toch herleidbaar blijken. Nouwt: ‘Een geboortedatum plus woonplaats kan bijvoorbeeld al herleidbaar zijn tot een individu. Wanneer een partij om deze gegevens vraagt, kun je voorstellen om alleen de leeftijd of het geboortjaar door te sturen. Vaak kunnen de onderzoekers daar ook al mee uit de voeten en hebben ze de exacte geboortedatum niet nodig.’

patiënten, maar wel voor een specifiek individu. Verder gebruiken we over het algemeen in Nederland niet het echte big data-niveau 4 (zie *kader*) waar we hier over spreken, aldus Hekster. Instellingen die hiermee aan de slag willen, moeten beginnen bij de bron. Voor dit soort complexe data-analyses is namelijk niet alleen een slimme computer nodig, het begint met het volgens standaarden vastleggen van klinische gegevens, en het bewustzijn van wat er later nog met de data moet kunnen. ‘Het datadenken moet doorsijpelen in alle haarvaten van een instelling, om echt profijt te hebben van de big data’, aldus Hekster. ‘Er zullen ook deskundigen moeten komen die het verband tussen klinische praktijk en de data begrijpen, de zogenaamde *clinical data scientists*. Gelukkig worden er ook in

Nederland hoogleraren aangesteld die de kennis over big data op een hoger plan brengen.’

Leefstijldata

De leefstijl van een patiënt blijft nu vaak grotendeels buiten beeld van de arts, zegt ook Hekster van IBM. Patiënten zelf verzamelen steeds meer gezondheidsgegevens met *wearables*, oftewel meet-apparaatjes die de patiënt met zich meedraagt in zijn telefoon, smartwatch of los als stappenteller. Met de *wearables* komt er een nieuwe stroom data op gang over dagelijkse gewoonten van patiënten. Van de 30.000 medische apps in de App-store gaat 90 procent over preventie, levensstijl, fitness en eetadvies. De heilige graal voor big data-analisten is de koppeling van de metingen uit de goedgekeur-

de gezondheidsapps aan het elektronische patiëntendossier. In het Meander Medisch Centrum in Amersfoort is bijvoorbeeld de Activiteitenweger-app ontwikkeld, speciaal voor het ziekenhuis. Ergotherapeuten krijgen hiermee meer inzicht in de dagelijkse activiteiten van revaliderende patiënten, als de patiënt daar toestemming voor geeft. Op hoger niveau worden ook samenwerkingen gesloten. IBM heeft hiervoor afspraken gemaakt met Apple Healthkit, en ook epd-producent Epic Systems heeft dergelijke afspraken gemaakt.

Een koppeling tussen apps en epd is niet eenvoudig te realiseren. ‘Er is veel verzui-ling in de zorg, de data zijn verkokerd’, zegt Hekster. Het probleem zit er niet zozeer in dat bestanden slecht uitwisselbaar zijn, dat is in de praktijk vaak nog wel op te lossen. De crux zit hem bij de ontsluiting van de data, want die houden artsen of instellingen graag voor zichzelf. Hekster: ‘Ik ben voor het bevrijden van alle data, maar ik hoor vaak van artsen dat zij het liever zelf houden, bijvoorbeeld omdat er een aios aan het promoveren is op basis van het databestand. Zonde, want als we bijvoorbeeld alle oncologische patiëntendata centraal zouden verzamelen, zouden we behandelingen meer kunnen toespitsen op de individuele patiënt dan op de gemiddelde patiënt.’ ■

web

Eerdere MC-artikelen over dit onderwerp en verwijzingen naar websites met aanvullende informatie vindt u onder dit artikel op medisch-contact.nl/artikelen.