

De verbeelding van de ruimte

Jan Kraak j.kraak@rug.nl

Deel 2

Kijken in de N-dimensionale ruimte

Net als de 'gewone' driedimensionale ruimte, is ook de N-dimensionale ruimte ($N > 3$) te 'verbeelden' in een vlak. Als illustraties dienen een nieuwe visualisatiemethode voor EEG-data en een vierdimensionale kubus.

Wie deel 1 schrijft, moet ook deel 2 schrijven. In deel 1 behandelde ik verschillende methoden om de 'gewone' driedimensionale ruimte in een plat vlak af te beelden, zodat we ons die ruimte later weer kunnen 'verbeelden'¹. Het tweede deel gaat over de verbeelding van de N-dimensionale ruimte met $N > 3$.

Al lang zijn mensen gefascineerd door het idee dat er meer dan drie dimensies kunnen zijn. In de fysica wordt de tijd als de vierde dimensie beschouwd. Edwin A. Abbott schreef in 1884 het boek "Flatland – A romance of many dimensions"², een soort cultboek avant la lettre, over een tweedimensionale wereld met de mogelijkheid om in de derde dimensie te reizen.

Ik beschouw hier de N-dimensionale ruimte als een gedachteconstructie. Net zoals je de tweedimensionale ruimte uit kunt uitbreiden tot de driedimensionale ruimte, kun je die op zijn beurt uitbreiden tot een vierdimensionale ruimte, enzovoort. Ook in de N-di-

dimensionale ruimte is een afstand te berekenen tussen punten.

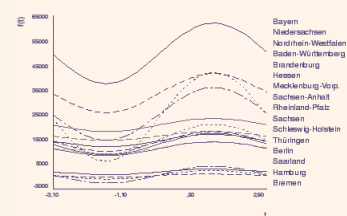
Veel verschijnselen zijn te beschrijven door een aantal (N-) variabelen die een zogenaamd multivariaat gegeven vormen. De N-variabelen kan men opvatten als de coördinaten van een punt in de N-dimensionale ruimte. In een hierna te geven voorbeeld worden de Duitse deelstaten beschreven door 4 variabelen: oppervlakte, bevolkingsdichtheid, werkloosheidspercentage en opkomstpercentage bij de Bondsdagverkiezing. Elke deelstaat correspondeert dus met een punt in de vierdimensionale ruimte.

Voor $N \leq 3$ kunnen we ons die ruimte voorstellen, maar voor $N > 3$ moeten we een afbeelding op het platte vlak zien te vinden. Het is nu de kunst om een zodanige afbeelding te vinden, dat we dingen over de hoogdimensionale ruimte te weten komen.

In de loop van de tijd zijn er al heel wat methoden bedacht die soms wat wonderlijk aandoen. Sommige werken niet, maar weer andere blijken erg nuttig te zijn.

Looking at multivariate data

Omstreeks 1980 bestond er in de multivariate statistiek grote belangstelling voor visualisatiemethoden, er werden conferenties aan gewijd. In maart 1980 bezocht ik met Anne Boomsma (Sociologie) de conferentie 'Looking at multivariate data'



Figuur 1. Andrews' plot van vier gegevens van de 16 Duitse deelstaten, let op de clustering.

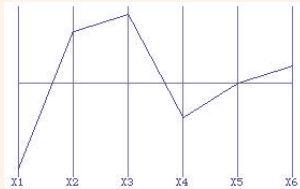
in Sheffield. De eerste spreker was David Andrews. Met behulp van een wiskundige formule³ beeldde hij multivariate gegevens af op curven in een grafiek. Deze afbeelding heeft een interessante eigenschap: punten die in de N-dimensionale ruimte dicht bij elkaar liggen, corresponderen met curven die in de grafiek dicht langs elkaar lopen. Dit biedt de mogelijkheid om visueel clustering te ontdekken (figuur 1)⁴. Andrews' plots zijn met Excel te maken.

Hoofdspreker was John W. Tukey, auteur van "Exploratory Data Analysis"⁵.

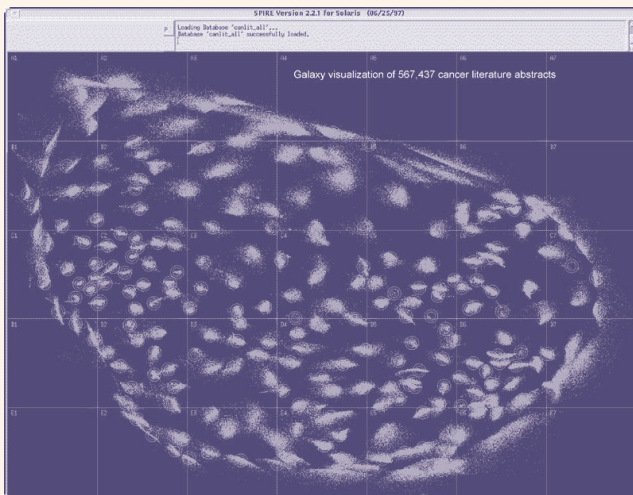
Niet lang na deze conferentie verminderde de belangstelling voor multivariate visualisatie. Na verloop van tijd kregen analisten van grote databases weer belangstelling, ze spraken van *data mining* en *knowledge discovery*. Ook de algemene term informatievevisualisatie deed opgang. Over dat onderwerp schreef ik eerder^{6 7 8}.

Visual Analytics

Kortgeleden werd me een foldertje toegestuurd over de workshop 'Visual Analytics' op 4 juni 2005 in Darmstadt bij het Fraunhofer Instituut. Met visual analytics wordt het samengaan bedoeld van data mining en informativisualisatie, om greep te krijgen op enorme hoeveelheden data. Dus: 'oude wijn in nieuwe zakken'. In het foldertje wordt het National Visualization and Analysis Center (NVAC)⁹ genoemd, dat is gehuisvest in het Pacific NorthWest National Laboratory. Daar wordt software ontwikkeld voor informativisualisatie, zoals de 'sterrenwolkvisualisatie' van artikelen over kanker (figuur 2). Verwante artikelen klonteren samen. Het onlangs door president George W. Bush opgerichte ministerie voor binnenlandse veiligheid (Department of Homeland Security) is de sponsor van het NVAC. In het kader van de binnenlandse veiligheid kan men immers sterrenwolkvisualisatie e.d. toepassen op de e-mails van personen, die van terrorisme zijn verdacht.



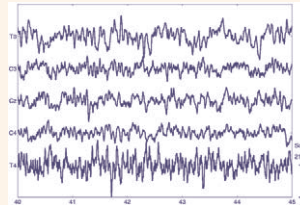
Figuur 3. De afbeelding van een zesdimensionaal punt met parallelle coördinaten.



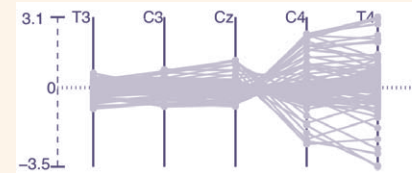
Figuur 2. Sterrenwolkvisualisatie van de 567.437 artikelen over kanker.

Parallele coördinaten

Ik ga nu verder met de multidimensionale visualisatie van punten volgens de methode van de parallelle coördinaten, die is uitgevonden door Alfred Inselberg¹⁰ in 1981. De methode is in wezen



Figuur 4. De potentialen op vijf elektroden uitgezet tegen de tijd.



Figuur 5. Parallelle coördinaten representatie van de potentialen op vijf elektroden voor 100 tijdstappen.

doodsimpel. In plaats van langs onderling loodrechte coördinaten, waarvan er maar hooguit drie zijn in 3D, zet Inselberg multivariate data af langs onderling parallelle coördinaten (PC), die op gelijke afstand van elkaar liggen. Het aantal assen is onbeperkt, de onderlinge volgorde doet er niet toe. Met de PC-methode zijn punten, lijnen en vlakken af te beelden¹¹. Ik beperk me tot punten. In figuur 3 staat de representatie van het zesdimensionale punt (-5, 3, 4, -2, 0, 1).

Er zijn verschillende programma's beschikbaar om data met parallelle coördinaten te visualiseren, zoals Xmdv-Tool¹². Omdat onze geometrische intuïtie hier niet meer opgaat, moeten we nieuwe regels leren om deze methode goed te kunnen gebruiken.

Conventionele EEG-visualisatie

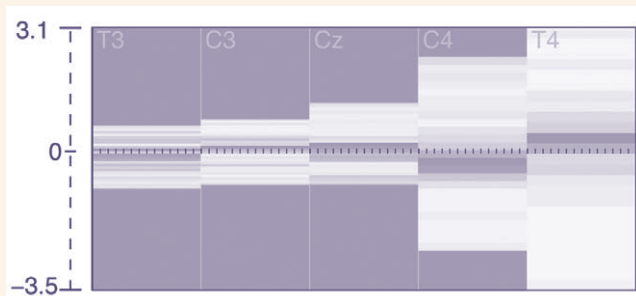
Michael ten Caat, als AIO werkzaam bij Informatica en bij het Neuro-imaging Centrum van de RUG, paste de PC-methode toe op EEG (electro-encefalografische) gegevens. Tijdens een EEG-experiment wordt de elektrische activiteit van de hersenen gemeten met behulp van (maximaal 256) elektroden, die zijn aangebracht in een aantal vaste punten op de schedel. Op alle elektroden wordt *simultaan* de elektrische potentiaal gemeten met een sample frequentie van maximaal 2000 Hz. Een klinisch experiment

duurt tussen de 15 en 30 minuten, terwijl sommige slaapexperimenten uren lang kunnen duren. Gedurende experimenten wordt de patiënt aan bepaalde stimuli bloot gesteld, zoals lichtflitsen of elektrische prikkels, om bepaalde hersenreacties op te wekken.

Het is gebruikelijk om de potentialen op de elektroden in een grafiek boven elkaar tegen de tijd uit te zetten. In figuur 4 staan de gegevens van vijf elektrodes. Het moge duidelijk zijn dat het veel ervaring vereist om hier iets uit af te lezen, vooral bij veel curven. Reden genoeg dus om eens naar een andere visualisatiemethode te zoeken, die hopelijk beter is.

Tiled parallel coordinates (TPC)

Michael ten Caat bedacht samen met Natasha Maurits (UMCG) en Jos Roerdink (Informatica) de zogeheten *tiled parallel coordinates* (TPC) methode. Deze wordt toegepast op de hersenresponses op de elektrische stimulatie van de centrale zenuw in de pols, wat aanleiding geeft tot activiteit in de tegenovergestelde hersenhelft. In plaats van de conventionele manier, waarbij de potentialen op de N-elektrodes tegen de tijd worden uitgezet, worden nu *tijdens elke tijdstap de potentialen op de N-elektrodes opgevat als de coördinaten van een punt in de N-dimensionale ruimte*. Dat kan worden gevisualiseerd met N-parallelle coördinaten. Elke punt geeft aanleiding tot een lijn. Bij



Figuur 6. Samenvatting in tiles van de data uit figuur 4.

veel tijdstappen in dezelfde figuur krijgt men een chaos van lijnen op zoals figuur 5 toont.

Beter is het om de ruimte om één parallele as als een *tile* (tegel) te beschouwen behorende bij een bepaalde elektrode, waarop een *samenvatting* is getekend van de bij de tijdstappen behorende lijnen. De extrema zijn aangegeven door het zwarte deel, waarin geen lijnen lopen. Tevens wordt de dichtheid van de lijnen doormiddel van grijswaarden weergegeven. In figuur 6 staan samenvattingen van de gegevens van vijf elektroden voor 100 tijdstippen.

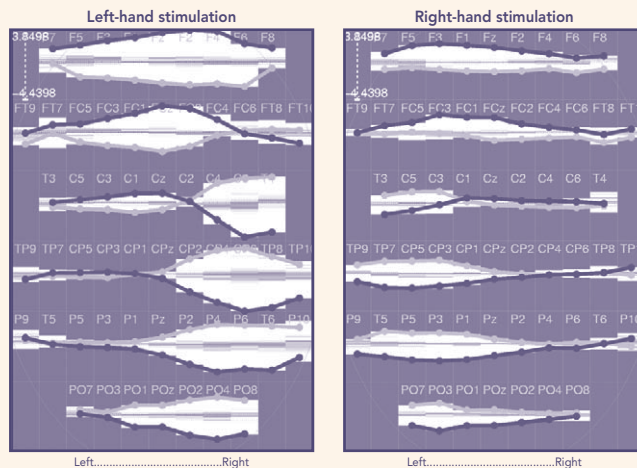
De tiles worden nu zodanig gerangschikt, dat hun posities ongeveer overeenkomen met de werkelijke plaats van de elektroden op de schedel. Figuur 7 toont 58 elektroden. Links zijn de responses op de elektrische prikkeling van de linker pols getekend en rechts responses voor prikkeling van de rechter pols. De verwachting is dat neurologische 'defecten' zichtbaar zullen worden als een asymmetrie in de weergave. Voor meer uitleg verwijs ik naar de paper die Michael ten Caat heeft gepresenteerd op het Eurographics/IEEE visualisatiesymposium in Leeds, begin juni 2005¹³.

Het zal duidelijk zijn dat met de TPC-methode veel meer elektroden en veel meer tijdstappen zijn te visualiseren dan met de conventionele methode. De TPC-methode

is al aan klinische EEG-experts van het UMCG voorgelegd. Binnenkort wordt de methode geëvalueerd, waarbij hij wordt vergeleken met een bestaande methode. De TPC-methode is ook toe te passen op andere tijdsafhankelijke multivariate gegevens, bijvoorbeeld van seismisch onderzoek.

Hyperkubussen

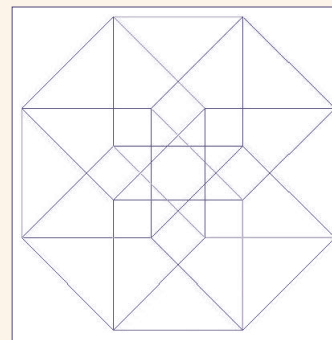
Na de multidimensionale punten, beschouw ik tot besluit de hyperkubussen van Jonathan Bowen als een voorbeeld van de visualisatie van hoogdimensionale figuren¹⁴. Net als bij parallele coördinaten worden de coördinaatassen in één vlak getekend, maar nu niet evenwijdig aan elkaar. We beginnen met een punt als een nuldimensionaal object. Vervolgens verplaatsen we dit punt vanuit de oorsprong langs de X-as over een afstand L. Dan krijgen we een lijn van lengte L, een eendimensionale kubus. Daarna verplaatsen we deze lijn over een afstand L langs de Y-as en verbinden de uiteinden van de lijn, en krijgen een vierkant ofwel een tweedimensionale kubus. In het vlak is de projectie van de Z-as getekend. We verplaatsen verder de hoekpunten van het vierkant over een afstand langs de Z-as en krijgen dan een 3D-kubus. Vervolgens kiezen we een nieuwe vierde W-as in het vlak, waarlangs we de punten van de 3D-kubus over een afstand L verplaatsen. Als we alle punten met elkaar verbinden krijgen we een 4D-kubus



Figuur 7. TPC-visualisatie van data op 58 elektroden voor responses op de elektrische prikkeling van beide polsen.

(figuur 8). Dit proces is te herhalen voor 5D-, 6D-, ... kubussen. Voor meer details verwijs ik naar het artikel van Bowen.

Bij Chemie zijn er plannen om met een soortgelijke methode vierdimensionale NMR-spectra, die worden gebruikt voor structuurbevestigingen, driedimensionaal te visualiseren in de Reality Cube van het universitair centrum voor High Performance Computing & Visualization (HPC&V)¹⁵.



Figuur 8. Vierdimensionale kubus.

Noten

- 1 J. Kraak, De verbeelding van de ruimte, deel 1, Pictogram april 2005, www.rug.nl/rc/organisatie/pictogram/2005-2/Verbeelding.pdf
- 2 www.geom.uiuc.edu/~banchoff/Flatland.
- 3 De variabelen worden opgevat als de coëfficiënten van een Fourierreeks die wordt getekend van -3.14 naar +3.14.
- 4 Beispiel zum Andrews' Plot, www.quantlet.com/mdstat/scripts/estat_eda/estat/bpreview/014_eda.html.
- 5 Zie 6.
- 6 J. Kraak, Voor de dolende webspeurder - Visualisatie van documenten, Pictogram 2001, www.rug.nl/rc/hpcv/people/kraak/kohonen.
- 7 J. Kraak, Informatievisualisatie met treemaps, Pictogram 2003, www.rug.nl/rc/organisatie/pictogram/2003-1/Visualisatie.pdf.
- 8 J. Kraak, Waarom werkt informatievisualisatie zo goed?, Pictogram 2005, www.rug.nl/rc/organisatie/pictogram/2005-1/Visualisatie.pdf.
- 9 NVAC, <http://nvac.pnl.gov>.
- 10 Home-pagina Alfred Inselberg, www.sdsc.edu/~nerona/Inselberg_A2.
- 11 Parallele coordinates, <http://catt.bus.okstate.edu/jones98/parallele.html>.
- 12 <http://davis.wpi.edu/~xmdv>.
- 13 M. ten Caat, N.M. Maurits, and J.B.T.M. Roerdink, Tiled Parallel Coordinates for the Visualization of Time-Varying Multichannel EEG Data, www.cs.rug.nl/~mtc.
- 14 J. Bowen, Hypercubes, 1981, www.jpbowen.com/publications/ndcubes.html.
- 15 HPC&V, www.rug.nl/rc/hpcv.