

Bedankt voor het downloaden van dit artikel. De artikelen uit de (online)tijdschriften van Uitgeverij Boom zijn auteursrechtelijk beschermd. U kunt er natuurlijk uit citeren (voorzien van een bronvermelding) maar voor reproductie in welke vorm dan ook moet toestemming aan de uitgever worden gevraagd.

Boom

Behoudens de in of krachtens de Auteurswet van 1912 gestelde uitzonderingen mag niets uit deze uitgave worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand, of openbaar gemaakt, in enige vorm of op enige wijze, hetzij elektronisch, mechanisch door fotokopieën, opnamen of enig andere manier, zonder voorafgaande schriftelijke toestemming van de uitgever.

Voor zover het maken van kopieën uit deze uitgave is toegestaan op grond van artikelen 16h t/m 16m Auteurswet 1912 jo. Besluit van 27 november 2002, Stb 575, dient men de daarvoor wettelijk verschuldigde vergoeding te voldoen aan de Stichting Reprorecht te Hoofddorp (postbus 3060, 2130 KB, www.reprorecht.nl) of contact op te nemen met de uitgever voor het treffen van een rechtstreekse regeling in de zin van art. 16l, vijfde lid, Auteurswet 1912.

Voor het overnemen van gedeelte(n) uit deze uitgave in bloemlezingen, readers en andere compilatiewerken (artikel 16, Auteurswet 1912) kan men zich wenden tot de Stichting PRO (Stichting Publicatie- en Reproductierechten, postbus 3060, 2130 KB Hoofddorp, www.cedar.nl/pro).

No part of this book may be reproduced in any way whatsoever without the written permission of the publisher.

info@boomamsterdam.nl
www.boomuitgeversamsterdam.nl

De *Methodenstreit* in tijden van Big Data

Reinoud Bosch*

Hoewel het epigram '*Plus ça change, plus c'est la même chose*' nooit volledig opgaat, heeft de nieuwste reïncarnatie van de *Methodenstreit* in de context van het Big Data-fenomeen er wel wat van weg. Kort door de bocht gesteld ging de *Methodenstreit* aan het einde van de twintigste eeuw over het verschil in belang van het unieke versus het gemiddelde, wat zich vertaalde in een tegenstelling tussen historisch georiënteerde casestudies versus het generaliserende effect van (statistische) benaderingen gericht op het blootleggen van gemiddelden en de verbanden tussen gemiddelden (Albert, [1963] 2012); Kelle, 2008). Hoewel regelmatig gesteld is dat zowel gemiddelden als unieke verschijnselen relevant zijn om de sociale wereld te begrijpen – niet in het minst door Weber (1904) – dreigt dat perspectief iedere keer weer onder te sneeuwen. En dat ligt geenszins alleen aan kwantitatieve onderzoekers, maar vaak ook aan kwalitatieve onderzoekers.

Om een relevant voorbeeld te noemen: op kwantitatieve benaderingen gebaseerde schaak- en go-computers zijn er inmiddels in geslaagd de beste menselijke spelers ter wereld te verslaan – een prestatie van formaat voor kwantitatief onderzoek gericht op artificiële intelligentie van besluitvorming. En dat is natuurlijk maar een enkel voorbeeld – zie voor andere voorbeelden het essay van Beuving. Toch bestaat er vaak een afkeer ten opzichte van dit soort kwantitatieve implementaties onder kwalitatieve onderzoekers. De andere kant van het verhaal is uiteraard de regelmatige verwaarlozing van het unieke bij kwantitatief onderzoek. Statistiek is gebaseerd op gemiddelden (van een verdeling, van een correlatie, ten opzichte van een gemiddelde, et cetera) en unieke verschijnselen worden in een kwantitatieve analyse vaak terzijde geschoven. Daarnaast worden bij kwantitatief onderzoek de betekenissen van gegevens vaak verwaarloosd. Kwantitatieve gegevens – indicatoren en uitkomsten – vereisen interpretatie om te kunnen begrijpen waar ze over gaan.

Methodologisch is er dus niets nieuws onder de zon. Wat wel opvalt, is dat met de komst van Big Data kwantitatieve benaderingen aan kracht lijken te hebben gewonnen. Het boek van Stephens-Davidowitz (2017) is daar een goed voorbeeld van. Kritische kwalitatieve onderzoekers zullen niet veel tijd nodig hebben om de vinger te leggen op de (soms verbazingwekkende) tekortkomingen in het boek. Dat neemt niet weg dat gemiddelden niet voor niets gemiddelden zijn en dus inzichten kunnen bieden – over gemiddelden. En gemiddelden zijn belangrijk omdat ze zo gemiddeld zijn.

* Reinoud Bosch is coördinator eindredactie van KWALON. Hij begeleidt masterstudenten bij hun scripties bij de vakgroep Accountancy & Control van de University of Amsterdam Business School. E-mail: reinoudbosch@hotmail.com.

De algemeen economische benadering

De *Methodenstreit* speelde zich aanvankelijk af tussen algemene economen enerzijds en historische economen anderzijds. Stephens-Davidowitz is weliswaar geïnteresseerd in sociaalwetenschappelijke onderwerpen maar hij is, na een bachelor in de filosofie, gepromoveerd als algemeen econoom. Een bekende grap over algemene economen gaat als volgt:

‘There is a story that has been going around about a physicist, a chemist, and an economist who were stranded on a desert island with no implements and a can of food. The physicist and the chemist each devised an ingenious mechanism for getting the can open; the economist merely said, “Assume we have a can opener”!’ (Boulding, 1970)

Zoals gebruikelijk in de economische wetenschap veronderstelt Stephens-Davidowitz het één en ander, waarna hij kwantitatieve technieken loslaat op gegevens die door de *backends* van websites worden verzameld (waar zoekt iemand naar, wat typt iemand in, wat staat er in de verschillende gigantische databases, waar klikt iemand op, hoe verschuift iemand de muis, hoelang blijft iemand op een site, welke andere sites worden bezocht, enzovoort). De technieken van Stephens-Davidowitz beperken zich niet tot standaard correlaties, maar hij heeft het bijvoorbeeld ook over een zogenoemde *doppelganger*-techniek waarbij een voorspelling wordt gedaan over toekomstig ‘gedrag’ van een gebruiker op basis van gelijkennis met een ander type gebruiker (feitelijk een soort ideaaltypische analyse).

Dat Stephens-Davidowitz in zijn werk gebruikmaakt van aannames en dus in de woorden van Beuving ‘onkritisch (...) eigen vooronderstellingen op onderzoeksgegevens [projecteert]’ is dan ook niet verrassend. Dat doen economen nou éénmaal al zo ongeveer sinds het begin van de economische ‘wetenschap’ en ik ben zelden iemand tegengekomen die erin geslaagd is economen daarvan af te brengen. Een ander element typerend voor economen is een relatief simplistisch beeld van wetenschap:

‘Just about everybody agrees that physicists, biologists, and chemists are real scientists. They utilize rigorous experiments to find how the physical world works. In contrast, many people think that economists, sociologists, and psychologists are soft scientists who throw around meaningless jargon so they can get tenure. To the extent this was ever true, the Big Data revolution has changed that. (...) If a violent movie comes to a city, does crime go up or down? If more people are exposed to an ad, do more people use the product? If a baseball team wins when a boy is twenty, will he be more likely to root for them when he’s forty? These are all clear questions with clear yes-or-no answers. And in the mountains of honest data, we can find them.’ (Stephens-Davidowitz, 2017, p. 366-367)

Ook wat dit betreft ben ik zelden iemand tegengekomen die erin geslaagd is economen tot andere gedachten te brengen.

De historisch-sociologische/antropologische benadering

Voor een voorbeeld van de historisch-sociologische/antropologische benadering volstaat een verwijzing naar het essay van Beuving en het boek van Beuving en De Vries (2015). De vraag die Beuving opwerpt, is of er naast de methodiek van Stephens-Davidowitz nog ruimte is of zou moeten zijn voor een interpretatieve benadering. Deze vraag is naar mijn mening niet zo heel interessant, want het antwoord is uiteraard ja. Niet alleen vereist een deugdelijke analyse van Big Data de nodige interpretatie (hoe kan online gedrag het best worden geïnterpreteerd, wat betekenen de verschillende typen data, moeten zoekopdrachten worden gezien als uitingen van eerlijkheid of is dat idee eerder een marketing truc van Stephens-Davidowitz, hoe kunnen de typen data het best worden geanalyseerd in de context van specifieke onderzoeksvragen, welke verschillende perspectieven kunnen er bestaan, enzovoort, enzovoort), daarnaast hebben mensen ook nog altijd een offline bestaan. En wat te denken van het effect van de dubbele hermeneutiek, de manier waarop onderzoeksuitkomsten het onderwerp van onderzoek beïnvloeden (Giddens, 1984, p. 284). Geldt een correlatie dan nog wel en hoe moeten we dit interpreteren? Vragen te over voor een interpretatieve benadering.

Zoals Beuving zelf al stelt, is de vraag eerder hoe die interpretatieve benadering zou kunnen worden gecombineerd, dan wel de plaats zou kunnen innemen van het type benadering dat Stephens-Davidowitz voorstaat. En daar is wel een weg te gaan.

Een constructieve benadering: denken over algoritmen

In de voetstappen van Weber (1904) kan een constructieve benadering bestaan uit een synthese van kwalitatieve en kwantitatieve benaderingen waarbij kwantitatief onderzoek onder andere theoretische sensitiviteit aanlevert voor sociaal-historisch antropologisch situationele interpretatie. Dat houdt in dat kwantitatief onderzoek, met alle beperkingen die daarvoor gelden, op waarde wordt geschat en wordt ingezet bij en gecombineerd met interpretatie (Bosch, 2012). In tijden van Big Data vereist dit denken over algoritmen. Algoritmen zijn logische stappen die computers doorlopen tijdens het uitvoeren van programma's. Bij kwalitatieve data-analyseprogramma's kun je hierbij denken aan automatisch coderen; bij kwantitatieve data-analyse gaat het naast statistische pakketten onder meer om de inzet van zogenoemde *neural networks* en *machine learning*. Het gebruik van algoritmen is onontkoombaar zodra datasets te groot worden voor handmatige analyse – en dat is precies waar het bij *Big Data Science* over gaat.

Wat betekent dit voor kwalitatief onderzoek in tijden van Big Data? In de eerste plaats is het essentieel dat kwalitatieve onderzoekers de specifieke stappen die zij volgen bij hun onderzoek, in heldere logische stappen uiteenzetten. Dat hoort niet te veel gevraagd te zijn; per slot van rekening ligt daarin de rechtvaardiging van de uitgevoerde analyse. Daarnaast moet worden verhelderd welke concepten bij de analyse worden ingezet en wat die concepten precies betekenen. Wil een computer-algoritme automatisch een analyse kunnen uitvoeren dan moeten de logische stap-

pen en de betekenissen van de gebruikte concepten volledig helder zijn. Dat kan met behulp van zogenoemde ontologieën – conceptuele schema’s die bij algoritmen worden ingezet als betekenisstructuren (voor een voorbeeld van een dergelijke ontologie gebruikt door Bing, Google, Yahoo! en Yandex zie <http://schema.org/docs/full.html>). Met behulp van heldere logica en (gebruikers-specifieke, historische, situationele en iteratief aanpasbare) ontologieën kunnen automatische coderings-algoritmes worden verbeterd, waarbij tevens ruimte wordt geboden aan (nuttige, goed uitgevoerde) kwantitatieve analyses (zie Bosch, 2016).

Daarnaast is het zinvol de mystificatie van kwantitatieve Big Data-analyse te doorbreken. Het vlaggenschip van de kwantitatieve *Big Data Science* is *machine learning*, of, in de huidige nog spannendere terminologie, *deep learning* – gebruikt bij de ook al zo intrigerend klinkende *neural networks*. *Machine learning* wordt vaak zo gebracht alsof een computer zou leren van de gegevens die worden in- of teruggevoerd, maar dit is onzin. Een *machine learning*-algoritme bestaat uit een aantal logische stappen die worden uitgevoerd, met inputs en outputs. Het is niets anders dan een voorgeprogrammeerde regelkring. Voeg daar grote reken- en opslagcapaciteiten aan toe en je krijgt een programma dat als het doorlopen wordt dusdanige outputs creëert dat een wereldkampioen schaken of go kan worden verslagen. Dat is niet niets maar het is geen computer die zelf leert; het is een voorgeprogrammeerde regelkring. Het zou daarom beter zijn te spreken van *machine calibration* in plaats van *machine learning*, maar dat verkoopt natuurlijk een stuk minder. Toch is het doorbreken van dit soort mystificaties essentieel om een goed beeld te krijgen van wat computers feitelijk doen en hoe kwalitatief onderzoek met behulp van computers het best kan worden uitgevoerd. Er is behoefte aan goed ontworpen logische regelkringen die ruimte bieden voor de logica en ontologieën van kwalitatief onderzoek.

Het nadenken over kwalitatieve algoritmes en ontologieën in combinatie met het doorbreken van mystificaties leidt idealiter tot betere systemen voor data-analyse dan die op dit moment bestaan. Wil dit gebeuren dan moeten kwalitatieve onderzoekers bereid zijn zich daadwerkelijk te verdiepen in IT, en niet alleen in kant-en-klare kwalitatieve data-analysepakketten. Er moeten betere gebruikers-georiënteerde algoritmen en ontologieën worden ontwikkeld. Daarnaast is een openheid ten aanzien van de kracht van kwantitatief onderzoek van belang. Big Data-analyse wordt voor een groot deel, in ieder geval dat deel dat transparant is, beheerst door kwantitatieve benaderingen. Wil kwalitatief onderzoek een sterkere positie verwerven in Big Data-analyse, dan zullen er constructieve partners binnen *Data Science* moeten worden gezocht. Met andere woorden, IT’ers en kwantitatieve onderzoekers die openstaan voor aanvullingen op hun benaderingen. En die zijn er, want als systemen beter werken, kan er ontzettend veel geld worden verdiend. Toch is dit geen gemakkelijke opdracht, want waar kwantitatieve onderzoekers zich kunnen beperken tot een cursusje *R* en een cursusje *machine learning*, zullen kwalitatieve onderzoekers zich dieper in datastructuren, data-analyseprogramma’s en programmeertalen (zoals Java, Javascript, Python, C++, SQL en MongoDB) moeten verdiepen. Er moet nog te veel worden ontwikkeld.

Is er ruimte voor etnografie naast de economische benadering van Stephens-Davidowitz? Uiteraard; de vraag is vooral hoe ambitieus kwalitatieve onderzoekers willen zijn. Willen kwalitatieve onderzoekers een verschil gaan maken in de analyse van Big Data, dan moet er meer gebeuren dan kritiek leveren op kwantitatief onderzoek.

Literatuur

- Albert, H. ([1963] 2012). Model Platonism: Neoclassical economic thought in critical light (translated by D. Arnold & F. Maier-Rigaud). *Journal of Institutional Economics*, 8(3), 295-323.
- Beuving, J., & Vries, G. de. (2015). *Doing qualitative research: The craft of naturalistic inquiry*. Amsterdam: Amsterdam University Press.
- Bosch, R. (2012). *Wetenschapsfilosofie voor kwalitatief onderzoek*. Den Haag: Boom Lemma uitgevers.
- Bosch, R. (red.). (2016). Special issue: Qualitative research in the digital humanities. *KWALON* 61, 21(1).
- Boulding, K. (1970). *Economics as a science*. London: McGraw-Hill.
- Giddens, A. (1984). *The constitution of society: Outline of the theory of structuration*. Los Angeles: University of California Press.
- Kelle, U. (2008). *Die Integration qualitativer und quantitativer Methoden in der empirischen Sozialforschung: Theoretische Grundlagen und methodologische Konzepte* (2. Auflage). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Stephens-Davidowitz, S. (2017). *Everybody lies: Big data, new data, and what the Internet reveals about who we really are*. New York: Dey Street Books.
- Weber, M. (1904). Die 'Objektivität' sozialwissenschaftlicher und sozialpolitischer Erkenntnis. *Archiv für Sozialwissenschaft und Sozialpolitik*, 19, 22-87.