

daan wordt aan alle controleregels. Het oorspronkelijke principe van Fellegi en Holt is een speciaal geval van deze formulering, voor een specifieke verzameling toegelaten aanpassingen.

Recent is gebleken dat ook dit uitgebreide foutlocalisatieprobleem is te schrijven als MILP-probleem (Daalmans en Scholtus, 2018). Het kan daarom op dezelfde manier worden opgelost als het oorspronkelijke probleem van Fellegi en Holt. Resultaten op data met gesimuleerde fouten suggereren dat hiermee een verbetering in de kwaliteit van automatisch gecorrigeerde data kan worden bereikt, mits men in staat is om relevante toegelaten aanpassingen te vinden.

Slot

De twee besproken uitbreidingen maken de automatische foutlocalisatie flexibeler, maar ze introduceren ook extra keuzes die per toepassing moeten worden gemaakt, zoals de precieze kostenterm voor geschonden zachte regels en de keuze van de toegelaten aanpassingen. Vervolgonderzoek zal zich richten op het uitwerken en testen van deze methoden voor concrete toepassingen bij het Centraal Bureau voor de Statistiek.

LITERATUUR

- Daalmans, J., & Scholtus, S. (2018). *A MIP Approach for a Generalised Data Editing Problem*. Discussion paper, Den Haag: Centraal Bureau voor de Statistiek.
- Fellegi, I.P., & Holt, D. (1976). A Systematic Approach to Automatic Edit and Imputation. *Journal of the American Statistical Association*, 71, 17–35.
- Jonge, E. de, & Loo, M. van der (2014). *Error Localization as a Mixed Integer Problem with the Editrules Package*. Discussion paper 2014-07. Den Haag: Centraal Bureau voor de Statistiek.
- Scholtus, S. (2013). Automatic Editing with Hard and Soft Edits. *Survey Methodology*, 39, 59–89.
- Scholtus, S. (2016). A Generalized Fellegi-Holt Paradigm for Automatic Error Localization. *Survey Methodology*, 42, 1–18.
- Scholtus, S. (2018). *Editing and Estimation of Measurement Errors in Administrative and Survey Data*. Proefschrift, Amsterdam: Vrije Universiteit. <http://dare.uvu.vu.nl/handle/1871/55568>.
- Waal, T. de, Pannekoek, J., & Scholtus, S. (2011). *Handbook of Statistical Data Editing and Imputation*. Hoboken, NJ: John Wiley & Sons.

SANDER SCHOLTUS is methodoloog bij het Centraal Bureau voor de Statistiek in Den Haag. In maart 2018 is hij cum laude gepromoveerd aan de Vrije Universiteit op een proefschrift dat onder andere gaat over het onderwerp van dit artikel. E-mail: s.scholtus@cbs.nl.



Het is inmiddels alweer een jaar geleden dat aan onze verliefdheid op 'het algoritme' een eind kwam.¹ We hielden het niet langer vol. Het algoritme zorgde voor filterbubbels, beïnvloedde verkiezingen, propageerde vooroordelen en elimineerde bovenal onze privacy. Cathy O'Neil heeft ons de ogen geopend, het 'algoritme' is manipulatief en heeft niet het beste met ons voor.² Nee, het kon niet langer zo, het moest stoppen. Nu we een jaar verder zijn, komt het besef dat de breuk wellicht wat impulsief was. Zijn we te generalistisch geweest door alle algoritmes in de ban te doen? Maar hoe kunnen we elkaar dan weer vertrouwen?

In mijn werk als *analytics consultant* kom ik, jammer genoeg, steeds meer organisaties tegen die het vertrouwen in het gebruik van data en algoritmes aan het verliezen zijn. In de afgelopen jaren hebben ze, veelal op aangeven van strategieconsultants en technologieleveranciers, geïnvesteerd in opslagcapaciteit en analysesoftware en zijn ze als een dolle data gaan verzamelen en analyseren. Dit heeft geleid tot een overspannen vraag naar statistici en econometristen terwijl al deze inspanningen tot niet meer dan een open-deurenschouw³ van inzichten hebben geleid. Er zijn ook maar weinig echt impactvolle toepassingen van data science en kunstmatige intelligentie in de praktijk te vinden. Andrew Ng geeft aan dat bijna alle economische waarde die gegenereerd wordt met machine-leren op het conto komt van *supervised learning*⁵ toegepast op online *display ad*.⁴ Een kleine verbetering van de *click through rates* kan in deze industrie veel geld opleveren.

Ondanks dat deze toepassing van machine-leren kennelijk veel oplevert, moet ik bekennen dat ik nog steeds niet erg onder de indruk ben van de relevantie van de advertenties die ik online te zien krijg. Ik blijf me verbazen dat er in de media zoveel aandacht geschonken wordt aan de beloftes van data science en kunstmatige intelligentie, terwijl tastbare resultaten schaars zijn. De voorbeelden

zijn beperkt en hebben niet de potentie ook in de praktijk veel impact te hebben. Zeg nu zelf, zou je een algoritme⁶ dat GO kan spelen en winnen de opdracht geven je auto te besturen of je agenda te beheren? Vragen naar bewijs voor de gouden bergen die worden beloofd⁷ worden bij voorkeur genegeerd, het lijkt wel een religie.

Dat de inzet van statistiek en operations research juist veel impact heeft, komt minder vaak over het voetlicht. Onze overheid bespaart miljarden euro's doordat we met statistiek en operations research kunnen vaststellen welke dijkhoogte⁸ optimaal is en de hoge benutting van ons spoorwiel kan alleen bereikt worden omdat de NS operations research gebruikt om treinen te *schedulen*. Begin juli was ik als jurylid van de EURO Excellence in Practice Award⁹ (EEPA) op de EURO-conferentie in Valencia. Ik vond het fantastisch om te zien hoeveel praktische toepassingen van operations research gepresenteerd werden en te horen welke impact die hebben gehad. Toepassingen als het opstellen van een eerlijk spelschema voor de kwalificatie van het WK-voetbal, het bepalen van de beste manier om UNESCO-erfgoed te conserveren, het optimaal aansluiten van offshore windmolens op het energienet, het dynamisch plannen van melkcollectieritten, het verbeteren van het kindbeschermsbeleid en het bepalen van een marktmechanisme voor de verkoop van visserijrechten. Stuk voor stuk praktische toepassingen met veel impact, zowel maatschappelijk als economisch. Ik heb er alleen niets over gelezen in de media.

Ondanks dat ik al lang in ons vakgebied werkzaam ben blijf ik het lastig vinden aan iemand van buiten ons vakgebied uit te leggen wat operations research nu precies is. Voor hen is er geen verschil tussen operations research, data science, analytics, machine learning of kunstmatige intelligentie. In elk van deze disciplines gebruik je vanuit hun optiek data en wiskunde om inzichten te genereren die besluitvorming ondersteunen. Als gevolg projecteren

ze hun teleurstellende ervaringen op allemaal en dat is niet terecht. Om hun vertrouwen te herwinnen is het belangrijk ze te laten zien wat de impact van operations research werkelijk is. Dat kan in mijn optiek het beste aan de hand van voorbeelden zoals de inzendingen voor de EEPA, de Franz Edelman Award¹⁰ of natuurlijk je eigen ervaringen. Het is belangrijk dat duidelijk wordt wat operations research te bieden heeft. Zeker nu de wereld steeds complexer wordt en veranderingen zich sneller aandienen is de behoefte aan inzicht en op feiten gebaseerde, gestructureerde besluitvorming groter dan ooit. Wij moeten de wereld laten weten dat operations research daarbij een instrumentele rol kan spelen. Het verleden heeft dat al vele malen bewezen en – je kunt me vertrouwen – in dit geval zijn resultaten uit het verleden een garantie voor de toekomst.

NOTEN

- <https://www.wired.com/story/2017-was-the-year-we-fell-out-of-love-with-algorithms/>
- <https://www.wired.com/2016/10/big-data-algorithms-manipulating-us/>
- Een van mijn klanten vatte de resultaten van het data science initiatief binnen zijn organisatie zo samen
- <https://www.mckinsey.com/featured-insights/artificial-intelligence/how-artificial-intelligence-and-data-add-value-to-businesses>
- https://en.wikipedia.org/wiki/Supervised_learning
- <https://www.wired.com/2016/03/two-moves-alphago-lee-sedol-redefined-future/>
- <https://www.gartner.com/newsroom/id/3872933>
- <https://www.cpb.nl/persbericht/3213331/franz-edelman-award-voor-project-optimale-dijkhoogtes-nederland>
- <https://www.euro-online.org/web/pages/209/excellence-in-practice-award-eepea>
- <https://www.informs.org/Recognizing-Excellence/INFORMS-Prizes/Franz-Edelman-Award>

JOHN POPPELAARS is Practice leader of the Advanced Analytics and Business Intelligence unit of BearingPoint. E-mail: john.poppelaars@bearingpoint.com