

SPOORGLORIEN



NS-PROJECT AURORA VOORSPELT VOOR ELKE TREIN HET AANTAL REIZIGERS

JAN HOOGENRAAD, AART DE VOS & JAAP DE VRIES

De NS organiseerde in 2009 een wedstrijd tussen drie wiskundige teams. De opdracht was om in twee maanden zo goed mogelijk in te schatten hoeveel reizigers er per traject en tijdstip in de treinen hebben gezeten. Het ging om treinen die in 2008 in de Kop van Noord-Holland hebben gereden. De wedstrijd is gewonnen door econometristen van de Vrije Universiteit Amsterdam (VU) maar vormde ook de start van het project AURORA. In *STATOR* werd verslag gedaan van de wedstrijd.¹

De wedstrijd heeft een ruim vervolg gekregen: AURORA verschaft al jaren voorspellingen voor iedere trein en ieder traject. Twee miljoen voorspellingen per twee maanden op basis van honderden miljoenen chipkaartreizen. Sinds 2011 zijn deze prognoses de basis voor de planning van de capaciteit van treinen. Dat laatste gaat nu met een geheel taling lineair programmeringsprobleem, waardoor de planning meermalen per jaar kan worden bijgesteld. De noodzakelijke handmatige bijstel-

ling van prognoses en plannen is veel kleiner geworden. De treinen zijn gemiddeld korter geworden, vooral midden op de dag, en het aantal problemen (te volle treinen) dat een gevolg is van foute prognoses is afgenomen (maar nog niet nul). Kortom: AURORA is een succes.

Het project begon in 2009 als een voorspelwedstrijd. Aan de hand van 'Meten in de Trein' (MidT) gegevens. In iedere wagon worden vier maal per jaar de kaartsoorten geteld ten behoeve van de verdeling van opbrengsten tussen vervoerders. Sinds 2007 was de dienstregeling nauwelijks veranderd, dus waren er 200.000 datapunten beschikbaar voor een model.

De opzet was eenvoudig: 10% van de data werd geheimgehouden en wie bij het voorspellen de laagste Root Mean Square Error (RMSE) had won. Als voorbeeld was er een referentiemodel, gebaseerd op een trainings-sample van 10% en allerlei ideeën om zo goed mogelijk hierop te scoren. Op de geheime set had dat dan een wat

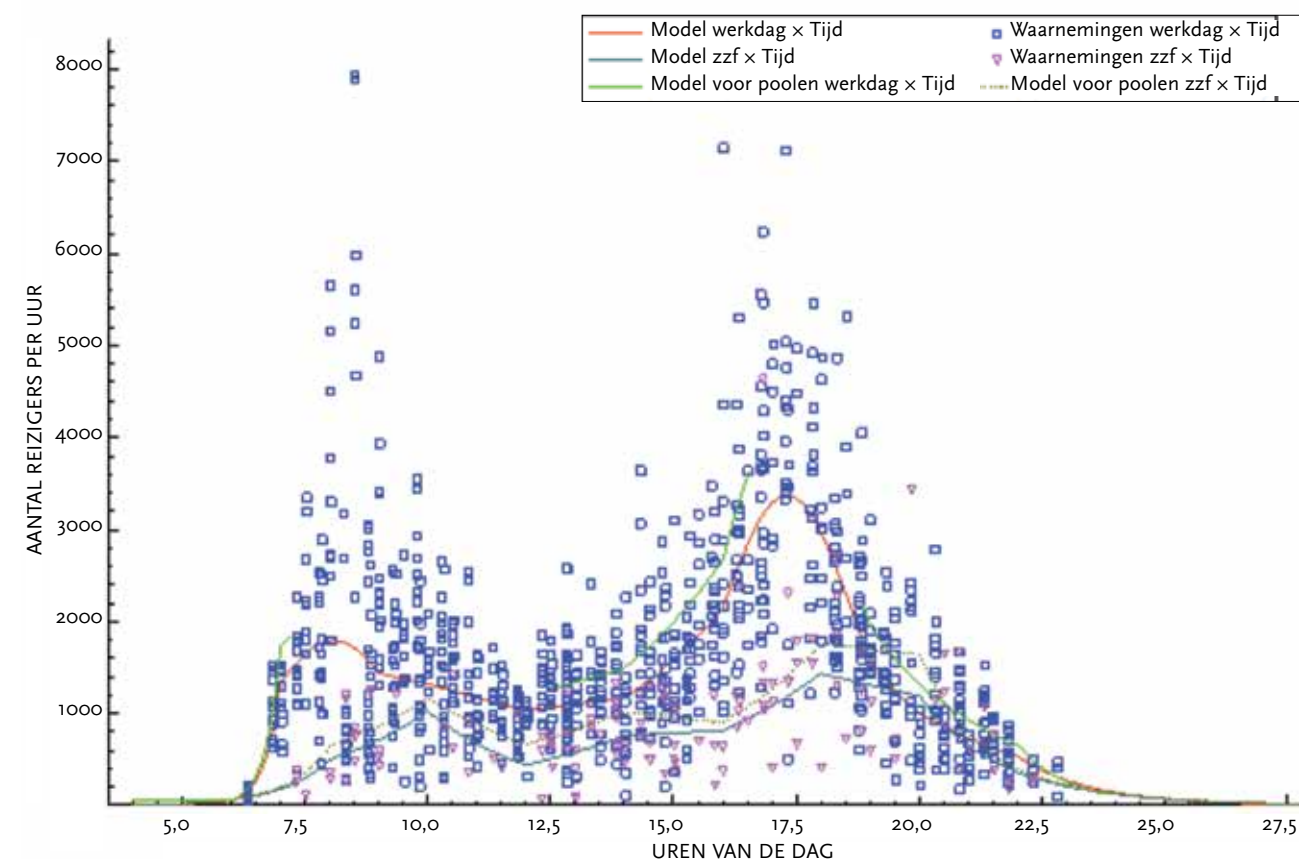
lagere score door overfitten, maar het werkte best aardig. Twee teams – CQM in Eindhoven (Consultants in Quantitative Methods) en de VU – kwamen bij de score van het referentiemodel in de buurt met multiplicatieve modellen voor tijd (splines)- en trajecteffecten. Het derde team, LIME van de Technische Universiteit Eindhoven, poogde reizigersgedrag te modelleren, maar dat bleek met de beperkte tijd en de schaarse data niet goed haalbaar.²

Leerzaam proces

Er volgde AURORA-10 waarbij het hele land in kaart werd gebracht. Het team van de VU modelleerde op het niveau van treinen, maar een trein die het halve land doorgaat heeft in de loop van zijn traject te maken met veel rolwisselingen. Het CQM-team knipte het land in stukjes, dat ging beter. Duidelijk werd dat de codering van data cruciaal was. Treinnummers veranderen soms, er komen nieuwe stations, en er zijn allerlei verschillen in administratieve details in de loop der jaren. Op basis van de geleverde prognoses werd in 2010 voor het eerst

een aparte materieelomloop voor de zomervakantie gemaakt.

Op 1 juli 2010 was het eindelijk zo ver dat AURORA-11 (voorspellen voor 2011) van start ging. Een project om op de basis van MidT-gegevens te komen tot prognoses. Het werd een bijzonder project. Een botsing tussen methodologieën. Aan de ene kant de methode van het referentiemodel, waarbij razendsnelle programma's (in PERL) spreadsheets met in bakjes opgesplitste uitslagen produceerden en creatief werden gemanipuleerd om een zo laag mogelijke score in de RMSE te behalen. Aan de andere het team op de VU dat gecompliceerde statistische trucs toepaste, op logaritmen, en met een computertaal Ox, een objectgeoriënteerde taal die alles in matrices doet maar veel aanpassingen vergde om met grote databestanden vol strings en uitzonderingen te werken. En daartussenin CQM die met het pakket STATA zeer snel nieuwe aspecten van de data tevoorschijn toverde. Het wekelijks overleg kwam voor sommige aanwezige NS'ers over als één grote ruzie maar voor de betrokkenen als een leerzaam proces. Samengevat met de volgende limerick



Figuur 1. Reizigersstroom op een traject als functie van de vertrektijd, apart voor werkdagen en voor zaterdagen/zondagen/feestdagen (zzf). Dit plaatje is uit 2011, met alleen de data van Meten in de Trein.

De VU levert schatters met dakjes
En Jan stopt alles in bakjes
En wat zit daar in?
De RMS BIN
Misschien begrijpen we 't strakjes

De MidT-tellingen leverden per traject plaatjes als in figuur 1 (een log-lineaire spline met spitsparabolen). Er doken allerlei problemen op. Soms was de melding 'treinstel te vol' of 'slechts een deel geteld', passagiers zijn vaak ongelijk verdeeld over de trein (het 'voor, midden, achter effect'), er was niet bijgehouden welk deel eerste-klasseplaatsen (overdag vaak leeg) betrof, enz. En daarbovenop waren er allerlei effecten van verstoorde treinenloop, popfestivals, studenten die allemaal in dezelfde trein zitten, en dergelijke. En weinig tellingen in weekends en 's morgens vroeg en 's avonds laat. Daar kan je op allerlei manieren mee omgaan. Waarbij niet alleen de statistische fit maar ook het doel van de exercitie (capaciteit van de treinen plannen) telt. Op de 10% van de data die niet gebruikt is voor de modellen wordt de betrouwbaarheid van de voorspellingen geëvalueerd. Daaruit wordt berekend welke capaciteit nodig is om bijvoorbeeld met 90% kans voldoende plaatsen te bieden.

Resultaat van het werken met meer modellen was de 'mix', een gewogen gemiddelde voorspelling op basis van de beste performance. Ook werd een systeem opgezet waarin je kunt zien wanneer de modellen sterk verschilden bij een voorspelling. Met die inzichten konden de modellen verbeterd worden, en een bereedere keuze gemaakt welke prognoses mee te nemen in de mix voor de planning. Bij het uiteindelijk mondeling overleg met mensen van NS was dit systeem zeer belangrijk.

Een doorbraak binnen de NS organisatie volgde toen 1200 volledige tellingen werden uitgevoerd en bleek dat de voorspellingen een factor twee nauwkeuriger waren dan de gebruikelijke op basis van conducteurstellingen. Ook kwam de werkelijke nauwkeurigheid overeen met de binnen AURORA (met de 10% toetsdata) bepaalde.

Eindeloos waren de discussies over transformaties (de logaritme of misschien toch de worteltransformatie?) en de relevante beoordelingscriteria. Waarbij vooral opviel dat de rangorde van de scores weinig gevoelig is voor het criterium dat men kiest. Uiteindelijk werd gekozen voor de RMSratio, een transformatie van de bekende R^2 . Zonder logaritmen, die leefden slechts voort in de modellen en in de volgende limerick.

Ln(OV)

Er was een geleerde uit Amsterdam
Die steeds de logaritme nam
En op de vraag
Is dat niet traag?
Zei hij: veel sneller dan de **Tram**

Eens bij een reis naar Nieuwegein
Sprak hij: kijk dat is fijn
Een passagiersaantal van nul
Geeft logaritmisch flauwekul
En hij stapte in een lege **Trein**

Hij legde uit: zo zit het dus
De log maakt van een maal een plus
Zo dat men zonder hinder
Werkt met percentages meer of minder
En het klopt altijd als een **Bus**

Alles bij elkaar was AURORA-11 een groot succes. Het bleek mogelijk om ondanks de schaarste aan data reizigersaantallen goed te voorspellen en daarmee beter de nodige capaciteit van treinen te plannen. Binnen NS kwam hiermee ook een proces op gang om sneller en nauwkeuriger te plannen. In dit proces vervulde de projectleider AURORA ook de rol van data scientist: het begeleiden in het gebruik van de data.

Puzzels

AURORA-12 had de ambitie om de OV-chipkaartgegevens te verkennen. Maar de OV-chipkaart kwam maar langzaam op gang; gemiddeld 12% van de reizigers reisde in 2011 er mee. Hiermee omvatte het project pas echt Big Data: we analyseren nu 13 miljoen reizen per maand. Uit de gegevens van stations en tijdstippen waar in- en uitgecheckt wordt is zeer goed af te leiden hoeveel OV-chipkaartreizigers zich in iedere trein bevinden. Hiertoe is een algoritme ROCKT ontwikkeld. De hoeveelheid data is enorm, maar na twee dagen rekenen levert het programma de gewenste output. De implementatie is in handen van NS, het AURORA-team hielp bij de ontwikkeling en onderzocht de voorspellende kwaliteit. Ook werden weeggegevens toegevoegd: een aantal treinen is uitgerust met weegapparatuur, waardoor we nog meer verschillende gegevens krijgen.

AURORA-13 volgde. Er waren nog vele puzzels op te lossen, en antwoorden te vinden op vragen die vroeg of laat gesteld worden. Hoogtepunt van dit project was de ontwikkeling van een algoritme om het effect van

AURORA: MODELLEN EN DEELNEMERS

In de AURORA-projecten werkten enkele afdelingen van NS samen. Voor het maken van modellen werden externe organisaties ingehuurd. De volgende personen werkten daaraan:

Model-C, gemaakt door CQM, Consultants in Quantitative Methods:

- Monique van den Broek (wiskundige, senior consultant bij CQM)
- Marcel van Vuuren (wiskundige, senior consultant bij CQM)
- Marnix Zoutenbier (principal consultant bij CQM)

Model-R (Referentie), voorbereiding, evaluatie en mixen:

- Jan Hoogenraad (natuurkundige, ervaring als programma manager, test- en kwaliteitsmanager en change manager)

Model-V (in eerste instantie ontwikkeld op de VU)

- Aart de Vos (UHD econometrie, nu gepensioneerd).
- Kai Ming Lee – tot en met AURORA-10 (econometrist).
- Jaap de Vries – vanaf AURORA-11 (econometrist met veel kennis van openbaar vervoer)
- Zes studenten econometrie – in AURORA-11

NS-afdelingen:

- Commercie (Edwin Eshuis, Bert de Vries, Ad Ligtenberg, Henri Kroesen)
- Finance Center (Victor van den Broek, Kees Jong, Karen Slijkhuis)
- Logistiek (Hans van Uden, Herman Poelen, Ruud Drabbe)
- Proceskwaliteit en Innovatie (Ellis Leijte, Dirk Jekkers, Ramon Lentink)

Vanaf de start van AURORA-13 (juli 2012) werken Jan Hoogenraad, Aart de Vos en Jaap de Vries samen in een nieuwe bv 'Spoorglaren'. Dit is het team dat sinds het voorjaar van 2013 is overgebleven voor AURORA.

grote veranderingen in de dienstregeling te voorspellen, op basis van OV-chipkaartgegevens. Eind 2012 is de Hanzelijn in gebruik genomen, en is op veel plaatsen in het land de dienstregeling ingrijpend gewijzigd. Om de gewijzigde reispatronen te kunnen voorspellen was een uitbreiding van de AURORA-methodiek nodig.

Die kon worden ontwikkeld en getest op een wijziging een jaar eerder: in 2012 werden treinen van Amsterdam naar Rotterdam via Haarlem geleid in plaats van Schiphol. Hoe voorspel je wat mensen gaan doen? Je neemt de chipkaartreizen van 2011, neemt aan dat mensen rond dezelfde tijd hun reis willen maken, voegt een preferentiefunctie in met onder meer voorkeur voor weinig overstappen, en schat dat om zo goed mogelijk 2012 te voorspellen. Dit gebruikten we om de gevolgen van de nieuwe Hanzelijn te voorspellen. Ook lukte het te laten zien hoeveel effect verstoringen hebben op aantallen passagiers in treinen.

Conclusie

Hoeveel mensen zitten er in de trein? Een eenvoudige vraag. Maar het is een fantastische uitdaging om met allerlei bronnen van informatie er iets over te zeggen. Een uitdaging die alles bevat wat een modellenbouwer zich kan wensen. Tijd en plaats en alles wat er mis kan gaan. En geen theorie, maar harde en zeer voorstelbare werkelijkheid. Een fout maken kan betekenen dat mensen niet in de trein kunnen. Goede planning spaart elektriciteit en geld (verbetering van de bezettingsgraad leverde een aanzienlijk bijdrage aan de energiebesparing van NS van de afgelopen twee jaar). En bestaande data slim gebruiken is vele malen goedkoper dan het verzamelen van nieuwe.

NOTEN

1. Hoogenraad, J., Lentink, R. & De Vries, B. (2010). Hoeveel reizigers zitten er in de trein? Wedstrijd NS groot succes! *STATOR*, 11(1), 8-10.
2. LIME werd onderdeel van Sioux, en deed in het vervolg niet meer mee.

JAN HOOGENRAAD is directeur van Spoorglaren, een bureau voor dienstverlening aan de spoorsector.
E-mail: <jan.hoogenraad@spoorglaren.nl>

AART DE VOS was universitair hoofddocent aan de VU en heeft een adviesbureau Ecstasystems.
E-mail: <ecstasystems@gmail.com>

JAAP DE VRIES werkt als consultant bij Spoorglaren en bereidt een proefschrift voor aan de VU.
E-mail: <jaap.devries@spoorglaren.nl>