

Nationaal verkeerskundecongres 2017

Data Science in de verkeerskunde, tijd voor een vernieuwde blik op verkeerskundige data

George Stern

(Vialis)

Samenvatting

Big Data is al jaren een hot item in de verkeerskundige wereld. Echter, de ontwikkelingen om relevante informatie uit Big Data te halen lagen tot nu toe stil of zijn nog niet gestart. Waar speculanten op de beurs intussen al rijk worden met parkeerdata als indicator voor omzetcijfers en EBITA (indicator van financiële gezondheid) van winkelcentra en daarmee van huuropbrengsten van winkelpanden, lieten verkeerskundigen 'informatie door datawetenschappelijke analyse', liggen. Maar die tijd is voorbij.

Resultaten van datawetenschap op verkeerskundige logging worden in deze paper toegelicht. De use case gaat over de automatisering van verkeerskundige configuratie van een kruispunt binnen de Kwaliteitscentrale. Ook worden er andere toepassingsgebieden geschetst waarvoor implementatie van kunstmatige intelligentie wordt ontwikkeld.

Trefwoorden

Verkeerslichtenregeling, kruisconfiguratie, topologie, Big Data

ORGANISATIE 2017



Introductie

Big Data is al jaren een hot item in de verkeerskundige wereld. In vele vakgebieden heeft Big Data - en met name de informatie die uit Big Data wordt gehaald – tot revoluties geleid. Echter liggen de ontwikkelingen om informatie uit Big Data te halen in de verkeerskundige wereld stil of zijn misschien nog niet gestart.

Ook bij de systemen van Vialis zien we ‘verwaarlozing’ van data. Van de verkeersregelininstallaties die aangesloten zijn op centrale systemen, is een behoorlijk aantal aangesloten op data-verwerkingssystemen zoals Kwaliteitscentrale of Vivaldi, VIMS en IVERA telgegevens, hierbij zijn de gegevens van het NDW en het nieuwe dataplatform uit cluster 2 van TalkingTraffic nog niet meegerekend. Er ligt echter meer dan 6 TB data waar bijna niets (meer) mee wordt gedaan, dat is ongeveer 60.000.000 uur, ofwel 7000 jaar aan verkeersdata.

De tijd is rijp voor een vernieuwde blik op deze verkeerskundige data. De vraag is, hoe deze data op wetenschappelijke methode mogelijk is. Resultaten van datawetenschap op verkeerskundige logging worden in deze paper toegelicht. De use case gaat over de automatisering van verkeerskundige configuratie van een kruispunt. Ook worden er andere toepassingsgebieden geschetst waarvoor Vialis implementatie van kunstmatige intelligentie ontwikkelt. Op deze manier brengen we de verkeerskundige wereld een stapje vooruit en voeren we onze diensten (nog) beter en efficiënter uit en zijn daardoor competitiever. Ook onderzoeken we zo nieuwe mogelijkheden voor klanten en de eindgebruikers.

In deze paper wordt: datawetenschap als discipline toegelicht, daarna wordt ingegaan op een use case die is gebruikt om de datawetenschappelijke methode toe te passen binnen het verkeerskundige domein. Vervolgens worden de resultaten weergegeven en daarop volgt een vooruitblik op wat er met deze resultaten kan worden gedaan, alsmede wat de datawetenschappelijke methode de verkeerskunde nog meer te bieden heeft.

Datawetenschap

Datawetenschap, vaak ook onvertaald: data science, is een interdisciplinair onderzoeksveld met betrekking tot wetenschappelijke methoden, processen en systemen om kennis en inzichten uit data die zich in verschillende gestructureerde of ongestructureerde vormen bevinden te bestuderen en te integreren.

Datawetenschap is een concept om statistieken, data-analyse en aanverwante methoden te verenigen. Het maakt gebruik van technieken en theorieën ontleend uit vele velden binnen het brede gebied van wiskunde, statistiek, informatiekunde en computerwetenschappen. In het bijzonder de subdomeinen van machinaal leren, classificatie, cluster-analyse, datamining, databases en visualisatie zijn belangrijke hulpvakken¹.

Verkeerskundige configuratie: een datawetenschappelijke use case

Voor het gebruik van datawetenschap in het verkeerskundig domein kunnen de volgende onderdelen worden bereikt:

1. Efficiënter werken door werkzaamheden te laten ondersteunen;
2. Werkzaamheden inhoudelijk verbeteren;
3. Overige ongekende mogelijkheden onderzoeken.

Aan de hand van een use case, over het efficiënter maken van verkeerskundige configuraties, worden de mogelijkheden van datawetenschap aangetoond. Verderop in deze paragraaf wordt de use case en de methode toegelicht.

Doelstelling use case

Voor de toepassing van datawetenschap in het verkeerskundige domein is een use case ontworpen. Het doel in deze use case is het efficiënter maken van verkeerskundige configuraties door kennis te

¹ <https://nl.wikipedia.org/wiki/Datawetenschap>

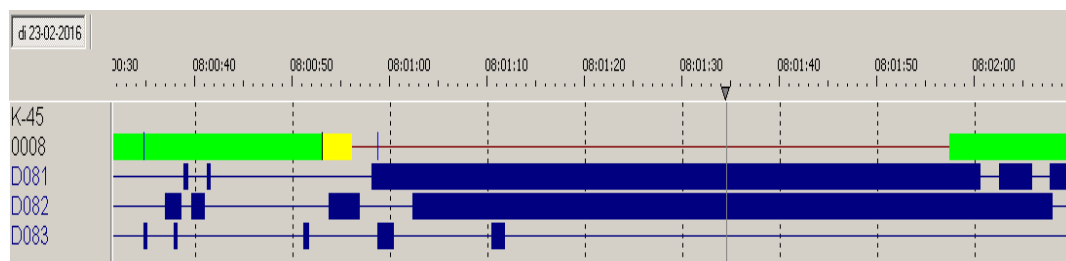
verkrijgen van eigenschappen van de ingangen van een VRI. Dit wordt gedaan aan de hand van mv-files (measured value files). Met de eigenschappen van ingangen wordt het volgende bedoeld:

- Type ingang, verdeeld in:
 - Massadetector
 - Druknop
 - Overige ingang
- Afstand tot stopstreep
- Lengte van detector
- Toepassingsgebied (één of meerdere rijstroken)

Wanneer datawetenschap wordt toegepast binnen het verkeerskundig vakgebied zal de data op een bepaalde methode moeten worden gepresenteerd, hierop wordt in de volgende paragraaf ingegaan.

Beschrijving van verkeerskundige informatie

Voor een verkeerskundige is het vrij eenvoudig om aan de hand van de werking van ingangssignalen de eigenschappen te herkennen. Zij het met beperkt detail. In onderstaand figuur een voorbeeld van drie ingangssignalen.

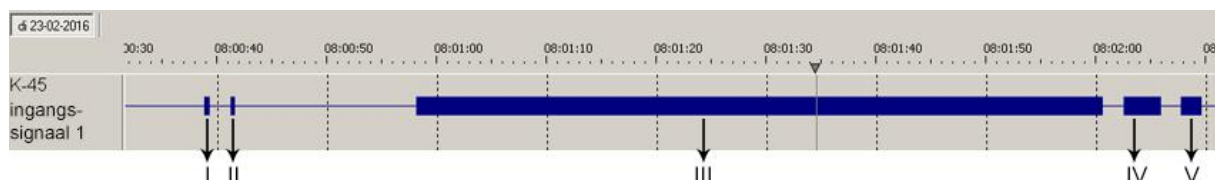


Figuur 1: ingangssignaal (1) – D081, ingangssignaal (2) – D082, ingangssignaal (3) – D083 in combinatie met de lantaarnaansturing.

Er is een duidelijk aanrijdpatroon te herkennen. Eerst wordt D083 aangereden, vervolgens D082 en daarna D081. Dit duidt al op de afstand van de detectoren ten opzichte van de stopstreep. De eerste twee activeringen van D082 zijn in het algemeen langer dan die van D081 en D083, hierdoor is D082 vermoedelijk een langere detector dan D081 en D082. Dat D081 langer geactiveerd wordt tijdens rood, duidt erop dat een voertuig stilstaat voor de stopstreep. Detector D082 wordt pas na een extra voertuig geactiveerd, maar is vervolgens ook lang geactiveerd. De lange lus ligt dan ook waarschijnlijk kort voor de stopstreep.

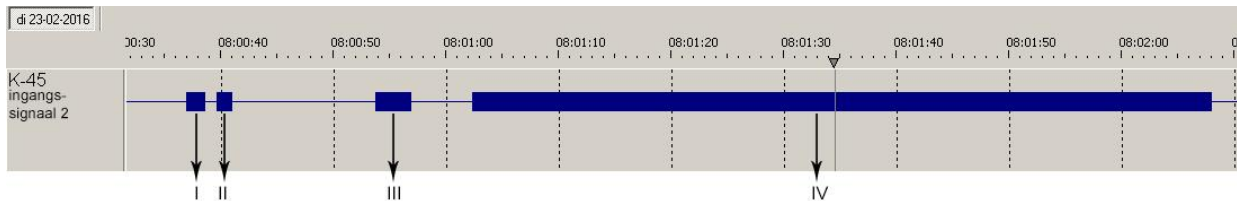
Beschrijving van datawetenschappelijke informatie

Datawetenschappelijk gezien kan het verband tussen de verschillende detectoren niet worden gelegd. In de drie figuren hieronder is een uitsnede zichtbaar van faselogging die de werking van drie verschillende ingangssignalen weergeeft, zonder de ondersteunde eigenschappen. Het gaat binnen deze methode over het representeren van de werking van het ingangssignaal in periodes van activatie en periodes van de-activatie.



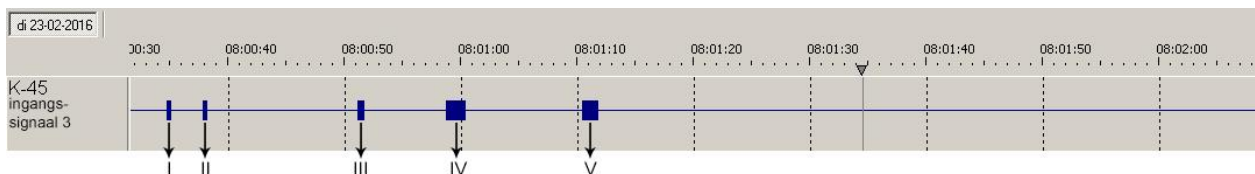
Figuur 2: werking van ingangssignaal (1)

In figuur 2 is te zien dat het ingangssignaal (1) twee keer zeer kort wordt geactiveerd (I en II), vervolgens één keer lang (III), en daarna steeds weer kort (IV en V). De tussenposen tussen (IV) en (V) is beperkt.



Figuur 3: werking van ingangssignaal (2)

In bovenstaand figuur 3 is te zien dat het ingangssignaal (2) twee keer kort wordt geactiveerd (I en II). Deze activeringen komen overeen met de eerste twee activering van ingangssignaal (1), maar duren iets langer. Vervolgens wordt het ingangssignaal één keer (III) iets langer geactiveerd en daarna lang (IV) geactiveerd. Er is geen vijfde activering te zien.



Figuur 4: werking van ingangssignaal(3)

In figuur 4 is te zien dat het ingangssignaal (3) alleen korte geactiveerd wordt. Sommige van deze activeringen zijn wel langer dan andere activeringen, maar niet veel langer.

De lange activeringen zijn te zien in figuur 2 (III) en figuur 3 (IV) en ontstaan wanneer voertuigen stilstaan. Dit gebeurt voornamelijk bij detectoren dicht bij de stopstreep. Het ontbreken van dergelijke lange activeringen (figuur 4) duidt op een grotere afstand van de stopstreep. De zeer korte activeringen die te zien zijn in figuur 2 (I en II) en in figuur 4 (I, II, III) duiden op een detectorlus met een beperkte lengte.

Hieruit is de datawetenschappelijke hypothese, die verkeerskundig wordt ondersteund:

- Ingangssignaal (1) is een koplus, bijvoorbeeld een lus van 1 meter lang op 1 meter van de stopstreep;
- Ingangssignaal (2) is een lange lus, bijvoorbeeld een lus van 20 meter lang op 20 meter van de stopstreep;
- Ingangssignaal (3) is een verweglus, bijvoorbeeld een lus van 1 meter lang op 60 meter.

Dit is echter niet de enige informatie die uit bovenstaande figuren is op te maken. De korte tussenposen tussen de activeringen (IV en V) in figuur 1 ondersteunen het afrijdpatroon bij een koplus. De activering (IV) van figuur 3 is langer dan de activering (II) van figuur 2. Toch is ingangssignaal (3) een korte lus en ingangssignaal (2) een lange lus. Dit komt doordat bij ingangssignaal (3) ook kortere activeringen optreden en bij ingangssignaal (2) niet.

Met enige training kan een verkeerskundige zonder ondersteuning van omgevingskennis dus de eigenschappen van een ingangssignaal herkennen. De essentie daarbij is een adequate representatie van de periode van activatie en de-activatie.

Archetypen

De verkeerskundige is nu in staat om de ingangssignalen te herkennen omdat hij / zij kennis heeft van de verschillende eigenschappen van de ingangssignalen en zeker van de meest voorkomen verschijningsvormen. Deze verschijningsvormen worden archetypen genoemd. Enkele voorbeelden van archetypen zijn:

- Drukknoppen
- Luslengte 1 meter, stopstreepafstand 1 meter (één rijstrook), voor motorvoertuigen (mvt)
- Luslengte 10 meter, stopstreepafstand 10 meter (één rijstrook), mvt
- Luslengte 15 meter, stopstreepafstand 20 meter (één rijstrook), mvt
- Luslengte 20 meter, stopstreepafstand 20 meter (één rijstrook), mvt
- Luslengte 1 meter, stopstreepafstand 50 meter (één rijstrook), mvt

- Luslengte 1 meter, stopstreepafstand 60 meter (twee rijstrook), mvt
- Luslengte 1 meter, stopstreepafstand 1 meter (één rijstrook), fietsers
- Etc.

Voor deze use case zijn totaal 36 archetypen gedefinieerd.

Algoritmekeuze en datasets

Datawetenschap houdt zich op verschillende subdomeinen bezig. Aangezien er veel archetypen bestaan, is gekozen voor een multi-classificatiemodel op basis van een neuraal netwerk² dat dient te worden getraind. Het trainen van het netwerk gebeurt door de datawetenschapper. Dit doet hij / zij door statische waarden van de werking van de ingangssignalen samen met de eigenschappen van dezelfde ingangssignalen aan te bieden.

De statische waarden van de werking van de ingangssignalen zijn dus de adequate representatie van de periode van activatie en de-activatie van het ingangssignaal én ook zijn de eigenschappen samengevat door de archetypen.

Het trainen van een neuraal netwerk kan alleen met historische data; data waarvan de verkeerskundige het archetype heeft kunnen bepalen.

Om het succes van de training van het neurale netwerk te controleren is er een crossvalidatie-dataset nodig. Hiermee worden verschillen tussen instellingen van het neutrale netwerk gecontroleerd en wordt datawetenschappelijke willekeur (bias) voorkomen.

Als laatste wordt een test dataset aangeboden. Ook van deze dataset heeft de verkeerskundige het archetype bepaald, maar deze set wordt bij de training buiten beschouwing gelaten. Het percentage van correct herkende ingangssignalen door deze test dataset is de score van de het neurale netwerk. Het gebruik van verschillende datasets voorkomt willekeur in het onderzoek en maakt het resultaat reproduceerbaar.

Resultaten

Binnen het onderzoek is een neuraal netwerk gemaakt voor het herkennen van 36 verschillende archetypen van vier verschillende kruispunten. Er is een dataset gemaakt van 1200 regels met de gerepresenteerde werking van ingangssignalen en de bijbehorende archetypen. Deze dataset is verdeeld in:

- train set – 60% van de dataset;
- cross-validatie set – 20% van de dataset;
- test set – 20% van de dataset.

Het percentage van correct herkende ingangssignalen is de score van het neurale netwerk. Dit percentage is tot nog toe 90%. Dit betekent dat van 90% van de ingangssignalen het precieze archetype kan worden bepaald.

Vooruitblik

Met het huidige getrainde neuraal netwerk kan 90% van de test set adequaat worden voorspeld. Meer data en verdere training zullen de voorspelling mogelijk verbeteren. Met de voorspelde ingangssignalen kunnen kruispuntconfiguraties (topologieën) worden ingevuld.

Deze werkmethode is:

- efficiënt, aangezien de werkzaamheden (deels) geautomatiseerd worden gedaan;
- effectief, doordat er direct wordt gecontroleerd of de detectoren conform ontwerp zijn gerealiseerd.
- Het inzetten van datawetenschap heeft zodoende een toegevoegde waarde binnen het verkeerskundig domein.

Naast de verkeerskundige configuraties die gemaakt kunnen worden met behulp van de voorspelde archetypen, kan met deze kennis meer worden gedaan. Wanneer dagelijks het archetype van een ingangssignaal wordt bepaald, kunnen ook afwijkingen worden geconstateerd. Deze afwijkingen kunnen te maken hebben met:

² Voor meer informatie over neurale netwerken zie: https://nl.wikipedia.org/wiki/Neuraal_netwerk

- (aankomende) Detectiestoringen, zoals ondergedrag, bovengedrag, jutteren;
- Foutief oplevering na onderhoud.

Door het constateren van deze afwijkingen kan vroegtijdig worden ingegrepen.

Op deze manier heeft de datawetenschappelijke benadering het verkeerskundig domein nog veel te bieden.