

## Juryrapport

### KNVI Scriptieprijs voor Informatica en Informatiekunde 2018

#### **M. (Meike) Nauta MSc (Universiteit Twente)**

##### *Temporal Causal Discovery and Structure Learning with Attention-Based Convolutional Neural Networks*

Sommige studenten krijgen een 10 voor hun scriptie. Wat dat betekent weet meestal niemand. Feitelijk is het een onhandigheid van de begeleider die woorden tekort schiet als hij of zij wil zeggen dat het toch wel èrg goed werk was. Beter dan die waarvoor ze ooit een 9,5 hadden toebedeeld. Meike Nauta heeft een 10 voor haar werk gekregen. Door alle drie haar begeleiders. Drie begeleiders die dus eigenlijk niet precies wisten wat ze hier mee moesten, behalve mogelijkerwijs dan nomineren voor de KNVI Scriptieprijs.

De jury had er minder problemen mee. Ten eerste waren ze erg blij met de nominatie. Tenslotte kan het heel veel werk schelen als er een scriptie is die toch wel snel de aandacht trekt. Nog mooier is het als deze de aandacht trekt tussen heel veel uitstekende nominaties. En bovendien de aandacht pakt van alle juryleden want dat betekent dat ze allemaal denken te begrijpen waar het over gaat. Dat kan alleen maar betekenen dat ongetwijfeld uiterst complexe materie uitstekend uitgelegd is. Goed uitleggen is meer dan een kunst: het kan alleen maar als je werkelijk begrijpt hoe de vork in de steel zit. Meike heeft overtuigend laten zien dat ze heel goed weet waar ze het over heeft.

De wetenschappelijke vraag is eenvoudig en al vaak gesteld: als je twee verschijnselen in gezamenlijkheid waarneemt, kun je dan iets zinnigs zeggen over hun oorzakelijk verband? Als molens harder draaien als het harder waait, betekent dat dat het harder waait doordat molens harder draaien?

Computers zijn tamelijk dom, maar zijn verschrikkelijk goed in het analyseren van onvoorstelbare hoeveelheden data. Zo kunnen we ze heel goed leren dat er een hoge correlatie bestaat tussen hard draaiende windmolens en hoge windsnelheden. Maar om ze te leren dat het draaien van die windmolens veroorzaakt wordt door windsnelheden en niet andersom is een hele andere zaak. Meike is hier diep ingedoken en heeft gekeken naar mogelijke oorzakelijke verbanden die voortvloeien uit tijdreeksen: Als iemand door een rood licht rijdt en even later een flitser zijn werk doet, zou dat kunnen komen omdat door rood licht rijden die flitser doet afgaan? Als dat heel vaak voorkomt, dan zou je bijna vermoeden van wel. En zo gemakkelijk als het is deze observatie te stellen, zo moeilijk is het om domme computers hetzelfde vermoeden te doen genereren.

Door Meike weten we nu hoe bepaalde, vermoede oorzakelijke verbanden blootgelegd kunnen worden. Het is een buitengewoon knap en toegankelijk geschreven scriptie die nota bene tot stand kwam nadat eerst werk aan een ander onderwerp was besteed. De jury was unaniem in haar oordeel: dit was de nummer 1.

**2<sup>e</sup> prijs G.G. (Govert) Brinkmann MSc (Universiteit Leiden)**  
*Interactive Visualization of Large Networks on a Tiled Display System*

Door de bomen het bos niet kunnen zien is een bekende uitdrukking die we gebruiken als we door te veel details het overzicht hebben verloren. Computers kunnen ons helpen bij het grafisch weergeven van zeer veel gedetailleerde gegevens. Het menselijk brein is namelijk zeer goed getraind om mogelijke structuren in afbeeldingen te herkennen. Het doel is inzicht te geven in de mogelijke structuur van de gegevens. Echter, inzicht en intuïtie zijn lastig te omschrijven begrippen, die niet voor iedereen hetzelfde betekenen, zodat we het niet altijd eens kunnen zijn hoe een computer zijn taak goed kan vervullen. We zullen veel moeten experimenteren om menselijk inzicht en intuïtie te toetsen. Eén ding is echter wel zeker: een zeer groot scherm (display) is nodig om een goede grafische weergave van massale hoeveelheden samenhangende data te geven.

Het doel van Brinkmanns studie is het inzichtelijk grafisch weergeven van gestructureerde data. Een groot technisch probleem hierbij is dat de computer enorm veel data moet verwerken en omvormen tot een fraai, en zeer groot, beeld. Een beeld dat ook eenvoudig moet kunnen worden gemanipuleerd, zoals in- en uitzoomen en draaien. Vooral dat interactief manipuleren is een opgave. Draaien wil nog wel, maar inzoomen betekent dat nieuwe gegevens onthuld moeten worden. Dat wordt knap lastig als je niet van tevoren weet waar er ingezoomd moet gaan worden. Dat betekent dat er ter plekke uitgerekend moet worden wat er getoond moet worden, en waar dat getoond moet worden. Duidelijk is ook dat wat er al getoond werd grotendeels zal veranderen. Ergo: met een beetje pech betekent inzoomen dat je compleet beeld zal moeten veranderen en dat alles opnieuw uitgerekend moet worden. Wel een klus als je scherm een beetje groot is. Govert Brinkmann beschrijft in zijn afstudeerscriptie een computer die verbonden is met een groot beeldscherm, BigEye. Het beeldscherm is om economische en technische redenen opgebouwd uit twaalf kleinere schermen. Brinkmann gebruikt drie zeer snelle grafische processors (GPUs) om de data te converteren naar de benodigde 25 miljoen pixels van een beeld. De drie GPUs worden weer aangestuurd door een hoofdcomputer. De beelden worden zestig maal per seconde gegenereerd.

Govert Brinkmann heeft zeer veel geëxperimenteerd om te leren hoe een computer zijn taak, inzicht leveren in complexe datastructuren, goed kan vervullen. In zijn verslag heeft hij zeer duidelijk uiteengezet wat zijn doel was, en hoe hij dit doel heeft bereikt. De jury is zeer onder de indruk van zijn fraaie werk en het goede verslag.

**3<sup>e</sup> Prijs T.S. (Thijs) van Ede (Universiteit Twente)**  
*Detecting Adaptive Data Exfiltration in HTTP Traffic*

Aanvallen op computersystemen via het internet zijn een toenemend probleem. Deze zogenoemde 'malware' probeert computers binnen te dringen en kan grote schade veroorzaken of websites onbereikbaar maken. Het bijtijds detecteren van deze malware is dus een belangrijke zaak. Nu is dit inmiddels een kat-en-muis-spel geworden tussen aanvallers en verdedigers. De aanvaller (de malware dus) probeert zijn aanwezigheid steeds meer te verbergen door zich te gedragen als 'gewoon netwerkverkeer'.

Het afstudeerwerk van Thijs van Ede richt zich met name op het detecteren van dit niet gewenste verkeer, dat net doet of het gewoon verkeer is. Hij heeft een aantal nieuwe klassen van camouflagetechnieken voor vijandig netwerkverkeer ontworpen die door de huidige malware detectoren niet opgemerkt worden. Vervolgens heeft hij een eigen malware detector ontwikkeld, genaamd ABIDED, die in meer dan 93% van de gevallen wel in staat is dit vijandig verkeer te detecteren en een laag aantal vals-positieve gevallen genereert. ABIDED is hierbij extensief getest op bestaand netwerkverkeer.

De scriptie van Thijs is zeer helder geschreven en is naar de mening van de jury terecht met een 10 gewaardeerd. Het verbaast dus ook niet dat het werk van Thijs al in een topconferentie in het veld is gepubliceerd en dat een andere in de maak is. Bovendien is er grote interesse van het bedrijfsleven voor zijn malware detector ABIDED.

*Prof. dr. dr.h.c. ir. K.A. (Kees) Schouhamer Immink, emeritus hoogleraar informatie theorie Universiteit Essen-Duisburg, president Turing Machines BV Rotterdam*

*Prof. dr. ir. H.J. (Henk) Sips, hoogleraar programmatuuraspecten van parallelle en gedistribueerde systemen Technische Universiteit Delft*

*Prof. dr. ir. M.R. (Maarten) van Steen, hoogleraar informatica Universiteit Twente, wetenschappelijk directeur ICT Onderzoeksinstituut Universiteit Twente*

De jury vergaderde op 15 oktober 2018 onder leiding van Ir. M.C. (Maarten) van Veen, oud-voorzitter KHMW; tevens waren ter vergadering aanwezig Prof. dr. A.P. IJzerman, secretaris natuurwetenschappen en Drs. S. van Manen, secretaris.