



Patroonherkenning helpt strategisch, tactisch én operationeel

Politiekorps voorspelt incidenten via datamining

Robbert Hoeffnagel

Waar houdt het buikgevoel van de individuele politieagent op en kunnen we misschien beter een beroep gaan doen op gestructureerde methoden voor 'knowledge discovery'? Die vraag is zeer actueel geworden nu een aantal politieregio's in ons land in toenemende mate een beroep doet op technieken als datamining. "Op basis van een brede reeks van gegevens kunnen wij uitstekend voorspellen waar we bij voetbalrellen een ME-peloton het beste kunnen opstellen."

"Iedereen weet natuurlijk dat als de politie een hangplek van drugsverslaafden schoonveegt, die mensen weliswaar van deze plek verdwijnen, maar dat zij even later elders in de stad weer opduiken. De vraag is alleen wel: waar dan precies?"

Als geboren en getogen Amsterdammer kan Mick Hoff over dit soort onderwerpen mooie anekdotes vertellen. Hij heeft 28 jaar bij de politie in Amsterdam gewerkt. Na een uitstapje van enkele jaren naar het bedrijfsleven is hij inmiddels wederom werkzaam bij de politie, maar dan in het zuiden van het land. Hoff is momenteel projectleider datamining in de politieregio Midden en West Brabant.

Zijn functie laat zich voor een belangrijk deel omschrijven als die van 'technology evangelist'. "Ik probeer het gebruik van datamining te bevorderen en help waar nodig met het oplossen van problemen. De laatste vier jaar dat ik bij het korps Amsterdam-Amstelland werkzaam was, hield ik mij bezig met voorspellen. Daar had men zo rond 2004 goed in de gaten dat op basis van allerlei historische gegevens goed voorspellingen vielen te maken over verkeerssituaties. Een politiekorps verzamelt heel wat gegevens.

De vraag is dan vervolgens natuurlijk: wat doen we daar verder mee? Je wilt er immers meer mee doen dan 'alleen maar' managementrapportages opstellen. Eigenlijk wil je het functioneren van een korps verbeteren door lessen te trekken uit het verleden. En dan liefst niet alleen op strategisch niveau, maar ook op tactisch niveau en zelfs operationeel."

De bronsystemen die een politiekorps tot zijn beschikking heeft bestaan voor tachtig procent uit databases, terwijl nog eens circa twintig procent bestaat uit tekstuele systemen. Logisch dus dat de gedachten in Amsterdam destijds al snel uitgingen naar dataminingstechnieken. "Hamvraag is dan alleen wel", vertelt Hoff, "kan de politie dat ook zelf? Of hebben we daar iedere keer

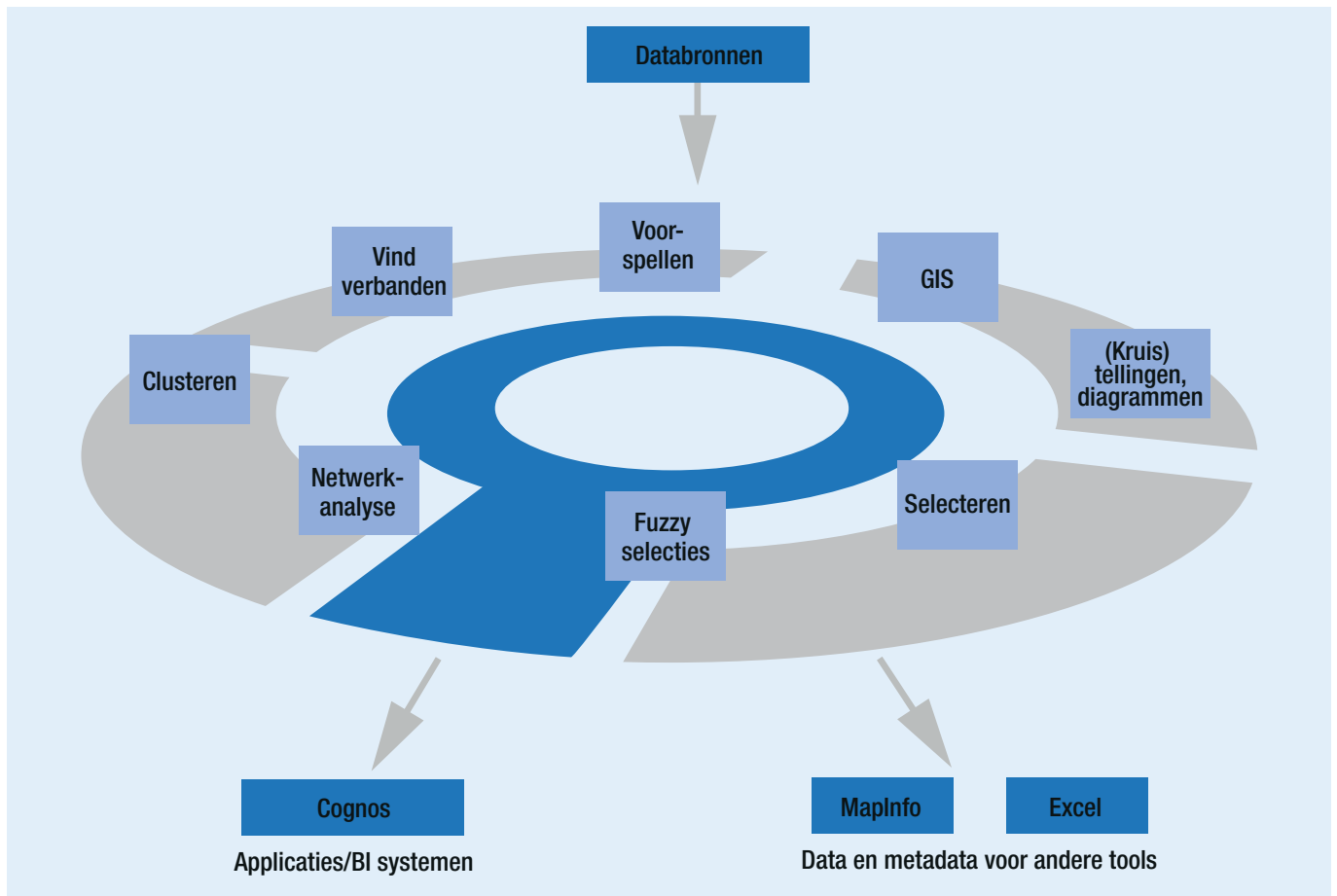
de ICT-afdeling of een dataminingspecialist bij nodig? Wie datamining op allerlei brongegevens tactisch en operationeel wil inzetten, zal voor een manier van werken moeten kiezen waarbij de politie die analyses zelf kan doen."

Ondersnemen

"We keken destijds in Amsterdam ook naar mogelijkheden om voorspellingen te doen over bijvoorbeeld veelplegers. Voor het toenmalige programmabureau ABRIO hebben we toen een inventarisatie gemaakt van tools waarmee we een datawarehouse konden inrichten, waarop we dan vervolgens een datamining tool konden loslaten. Die inventarisatie leverde uiteindelijk drie softwareproducten op die functioneel gezien behoorlijk aan elkaar gewaagd waren: SPSS Clementine, SAS Enterprise en DataDetective van Sentient. Wat echter opviel was dat twee van de drie tools toch meer gericht waren op gebruik door specialisten, waardoor deze voor gebruik door een regiokorps toch wat minder geschikt waren.

Alle functionaliteit die we zochten, was bij SAS en SPSS weliswaar aanwezig, maar de toegankelijkheid van de pakketten voor de relatieve leek was toch wat minder. Wat we vooral niet wilden was dat de gebruiker ondergesneeuwd zou raken in de aangeboden technologie. Vandaar dat de keuze uiteindelijk viel op DataDetective. Overigens was ook het gemak waarmee maatwerk aan dit pakket kan worden toegevoegd een belangrijk argument om voor dit product te kiezen."

DataDetective is een Nederlands pakket dat rechtstreeks kan werken op brondata die in SQL Server, Oracle of Microsoft Access zijn vastgelegd. Daarnaast kent het product nogal wat mogelijkheden om externe data te importeren. Rob van der Veer, directeur van ontwikkelaar Sentient, spreekt van 'duizenden soor-



Afbeelding 1: DataDetective functies.

ten bronnen', waaronder grote transactiedatabases, Excel-bestanden, tekst en SPSS. Aan de andere kant integreert het pakket met Cognos, Excel, Word en bijvoorbeeld Analyst's Notebook en – interessant uiteraard voor een politietoepassing – met MapInfo om kaarten te kunnen vervaardigen.

Hoewel de naam wellicht anders doet vermoeden, is het software-product breder in te zetten dan alleen voor misdaadanalyse. Van der Veer: "Een belangrijke toepassing bij de Nederlandse politie richt zich op het maken van voorspellingskaarten waarin typen criminaliteit worden uitgezet op basis van datum, tijd en bijvoorbeeld het weer. Hiermee kunnen clusters van incidenten worden gevonden waarbij sprake is van sterk vergelijkbare werkwijzen of bijvoorbeeld signalelementen.

Ook wordt de tool gebruikt om mogelijke verdachten te bepalen door een zaak of een serie van incidenten te vergelijken met situaties uit het verleden. Bovendien is het mogelijk om op zoek te gaan naar verklarende factoren voor een trend of bijvoorbeeld een bepaald type gedrag. We zien ook dat met DataDetective netwerken van criminelen in beeld kunnen worden gebracht en combinaties van tijd, plaats en omstandigheden kunnen worden achterhaald waarin een bepaald type criminaliteit speelt. Op die manier kunnen dus acties van de politie veel beter worden gestuurd."

Eis- en wenskolommen

DataDetective is opgebouwd uit een aantal onderdelen. De eerste is de zogeheten Organizer. In deze module worden alle dataselecties, selecties van variabelen, alle analyseresultaten en modellen beheerd. PowerQuery is de naam van het hulpmiddel waarin de gebruiker nieuwe selecties kan definiëren. Hierbij kan gebruik worden gemaakt van een zogeheten 'eiskolom' en 'wenskolom'. De uiteindelijke selectie bevat elementen die aan alle eisen voldoen en het meest overeenkomen of matchen met de wensen. Hiermee kan dus al gewerkt worden met situaties of personen die geheel of gedeeltelijk voldoen aan een bepaalde set van criteria. Bovendien kan de gebruiker zogeheten *Fuzzy Sets* aanmaken waarbij in de selectie elk element wordt gewogen op een bepaalde variabele.

In de analysemodule vindt het rekenwerk plaats. Hierbij kan de gebruiker kiezen uit een reeks van analysefuncties. Dat kan bijvoorbeeld een profielanalyse zijn. Hiermee is het mogelijk om in één klik de data te leren kennen door tegelijkertijd alle relevante kenmerken van een selectie weer te geven. Deze functie maakt het mogelijk om snel een beeld te krijgen van de typische kenmerken van een dataset. Deze profielanalyse werkt overigens op alle datatypes waar het pakket mee uit de voeten kan. Daarnaast wordt een Looking Glass geheten analysefunctie aangeboden. Van der Veer: "Looking Glass is een visuele cluster-

functie. Hiermee kan bijvoorbeeld worden bepaald wat voor klanten er bestaan of welke soorten incidenten er voorkomen. Elk element in de selectie wordt hierbij als een stip weergegeven. Als de clustering start bewegen de stippen die sterk vergelijkbare kenmerken kennen naar elkaar toe – het clusteren. Dit levert dus als het ware een interactieve visualisatie op van alle clusters en hun onderlinge samenhang. Door op een specifiek cluster in te zoomen, kan de gebruiker deze nader onderzoeken. Ook is het mogelijk om een cluster op basis van een specifieke variabele of doorsnede van de dataset te kleuren. Ook kan van een of meer clusters een profielanalyse worden gemaakt met behulp van het hiervoor genoemde hulpmiddel.”

Fuzzy matching

Looking Glass gebruikt een algoritme dat door Sentient in eigen beheer is ontwikkeld. Het gebruikt, zo vertelt Van der Veer, de complete multidimensionale ruimte om clusters te vinden. “Het probeert deze vervolgens zo goed mogelijk te projecteren op twee dimensies. Het onderliggende fuzzy match-systeem zorgt ervoor dat Looking Glass goed kan omgaan met foutieve data en ontbrekende waarden.

Fuzzy matching is in feite een soort associatief geheugen”, legt Van der Veer uit. “De menselijke intelligentie is voor een groot deel afhankelijk van het vermogen om te associëren. Met andere woorden: wat lijkt op elkaar en welke vergelijkbare situaties en aspecten zijn relevant? Dit is voor de politie uiteraard van cruciaal belang. Geen twee incidenten zijn volledig identiek, maar kennen vaak wel kenmerken die relevant zijn om – bijvoorbeeld – een mogelijke dader te kunnen bepalen.”

Dit associatieve vermogen wordt binnen DataDetective gebruikt om via fuzzy matching te kunnen zoeken, bij het maken van fuzzy selecties en fuzzy sets, om clusters te vinden, what-if analyses op te stellen en voorspellende modellen te maken. “Het is hierbij vooral interessant dat er weinig eisen aan het dataformaat worden gesteld. Sommige gegevens mogen ontbreken en complexe datatypes als teksten of verzamelingen kunnen in analyses worden meegenomen. Niet onbelangrijk is dat de voorbewerking van de data in tegenstelling tot andere voorspellingstechnieken relatief weinig tijd kost.”

Beslisbomen en diagrammen

Een andere analysefunctie wordt gevormd door beslisbomen. Hiermee kan de gebruiker stapsgewijs en met steeds meer detail onderzoeken welke kenmerken bepalend zijn voor het behoren tot een bepaalde groep. Bovendien kan hiermee worden voorspeld of een nieuw geval tot die groep zal behoren. Tot de eveneens beschikbare modevaluaties behoren onder andere Gains charts, root mean squared error, correlatie, ROC-curves en migratiematrixes. Erg relevant voor een toepassing bij politiekorpsen is uiteraard de mogelijkheid om de resultaten van een analyse te kunnen uitzetten op een kaart. Hiertoe is voorzien in een koppeling naar MapInfo als GIS (geografisch informatiesysteem) oplossing. GIS-analyses kunnen binnen DataDetective worden

uitgevoerd. Het is hierdoor mogelijk om zogeheten hotspotkaarten te maken, evenals puntkaarten, verschilkaarten en bijvoorbeeld vlakkenkaarten. Naast typische technieken voor datamining, beschikt het pakket ook over statistische functies als tellingrapporten, draaitabellen, kruistabellen, correlaties en descriptieven. Verder kunnen frequentiediagrammen, rapportdiagrammen, histogrammen, scatterdiagrammen, oppervlakte-diagrammen en seriediagrammen worden gebruikt. Hierbij kan gekozen worden voor alle gangbare diagramtypen.

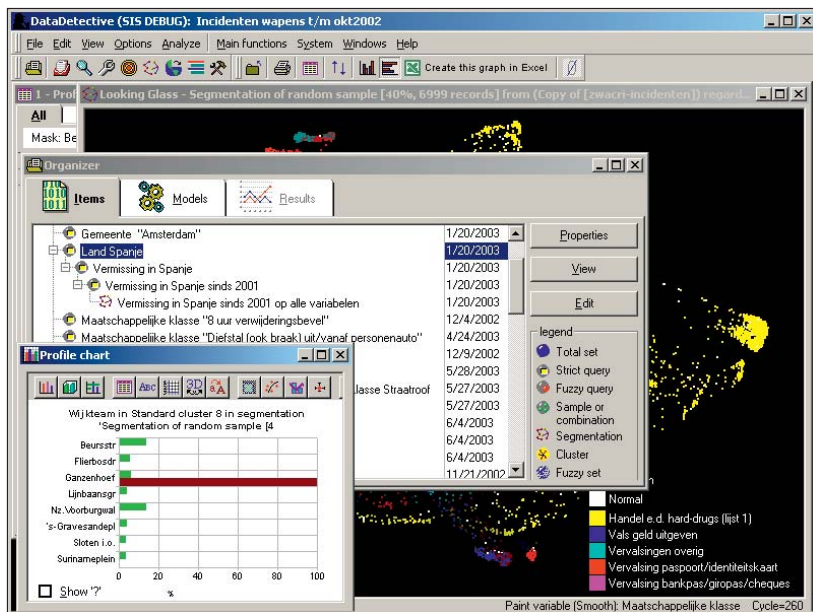
Modellen bouwen

Naast de analysemodule beschikt DataDetective ook over een zogeheten Model Studio. Hiermee kunnen modellen worden gemaakt en gewijzigd. Wizards helpen de gebruiker om stap voor stap een model op te bouwen of aan te passen. In de Application Designer kan vervolgens de koppeling met data worden gelegd. Hierbij worden typen data vastgesteld en normalisatiekaders voor fuzzy matching ingesteld. Ook kunnen hier labels worden aangebracht en kan de schermlyout worden bepaald. Hiermee kunnen dus – zeg maar – verticale dataminingapplicaties worden vervaardigd zonder dat hiervoor geprogrammeerd hoeft te worden. De laatste module is de zogenaamde Sentient Assistant. “Wij noemen dit wel ons dataminingreceptenboek”, vertelt Van der Veer. “In feite is het een startpagina voor de gebruiker die met datamining aan de slag wil gaan. Van hieruit wordt de gebruiker stap voor stap door het proces van datamining geleid. Hoe dit proces er precies uitziet, is uiteraard afhankelijk van de specifieke applicatie. Organisaties stellen die ‘recepten’ dus zelf op, compleet met verwijzingen naar tools en documentatie en dergelijke.”

Oud en Nieuw

Hoe gebruikt de regio Midden en West Brabant dit hulpmiddel nu precies? Hoff spreekt van datamining als een *hypothesegenerator*. “Die hypothesen zijn bedoeld voor één doel: de pakkans vergroten. De politie werkt in de dagelijkse praktijk met speerpunten. Een mooi voorbeeld is Oud en Nieuw. In veel gemeenten is dit een periode die voor vrij veel onrust en tal van incidenten zorgt. Een speerpunt kan dus zijn: tot een rustig verlopende jaarwisseling komen.”

De 48 uur rondom dat ene middernachtelijk moment vereist een voorbereiding die vaak al maanden van tevoren begint. Hoff: “In eerste instantie wordt zo’n project vooral strategisch opgepakt. Met DataDetective kunnen we vooraf al uitgebreide analyses maken: waar vonden in het verleden incidenten plaats? Om wat voor soort incidenten ging het: brandstichting, vechtpartijen, overlast door vuurwerk, noem maar op. Welke personen waren daar vooral bij betrokken? Dit kunnen we allemaal in kaart brengen en analyseren. Daar komen voorspellingen uit, waarbij rekening wordt gehouden met nieuwe gegevens. Zit bij wijze van spreken iemand die vorig jaar voor problemen zorgde nu in de gevangenis, dan heeft dat uiteraard gevolgen voor dit jaar. De rapportage zal daar rekening mee houden. Naarmate december dichterbij komt, schakelen we van een strategisch niveau over op



Afbeelding 2: DataDetective GUI.

een meer tactisch gerichte aanpak die vervolgens uitmondt in een operationeel document: welke mensen en middelen verwachten we op welke locaties en tijdstippen te moeten inzetten?"

Het is de politie zelf die aangeeft hoe een rapportage er precies uit dient te zien. "Deze rapportages zullen altijd nog verder verrijkt moeten worden", vertelt Hoff. "Het zijn basisdocumenten. Het is dus niet zo dat de datamining tool leidend is. De analyse geeft de richting aan, het is de politieanalist zelf die vervolgens de feitelijke inhoud van een document voorbereidt dat voor tactische of operationele doeleinden wordt gebruikt."

Hoe meer, hoe beter

DataDetective wordt nu in het district Amsterdam-Amstelland en zes regio's in Brabant, Limburg en Zeeland gebruikt. Werkt het? Ja, zegt Hoff direct. "Maar het kan zeker nog beter, maar dat heeft vooral te maken met de bronsystemen die we nu kunnen gebruiken. Een belangrijk voordeel van datamining is dat we nuttige informatie waarmee we de dagelijkse operatie kunnen aansturen nu veel sneller beschikbaar hebben. Ik zou zeggen dat we van een periode van drie weken nu op maximaal twee dagen zitten. Maar voor datamining geldt natuurlijk heel duidelijk: hoe meer data, hoe beter. Ik zou bijvoorbeeld heel graag meer gegevens van het CBS willen toevoegen. Of allerlei geografische informatie. En zo kan ik nog tientallen bronsystemen bedenken die ik er graag bij zou willen hebben. Sommige aanvullingen zullen er ongetwijfeld komen, maar andere laten vanwege beperkte budgetten misschien nog een tijd op zich wachten. Andere systemen kunnen we niet toevoegen omdat de wet dat nu eenmaal niet toestaat."

Wat is voor de politiemanager en -vrouw op straat nu het effect van het feit dat de genoemde regio's met dataminingstechnieken aan de slag zijn gegaan? Hoff: "Dat is allereerst dat informatie sneller beschikbaar is. We hebben deze manier van werken inmiddels

dundanig onder controle dat in acht van de tien gevallen de rapportage die gegevens oplevert die de politie op dat moment ook daadwerkelijk nodig heeft. De beschikbare informatie is nu veel vaker werkelijk relevant voor de situatie waarin de politieagent zich bevindt. Deze snelheidswinst is zoals gezegd aanzienlijk. Uit eigen onderzoek weten we dat gegevens nu tot twintig maal sneller op de werkvloer beschikbaar zijn."

Een ander voordeel is voor Hoff: "De pakkans gaat inderdaad omhoog. Dat komt omdat de politie efficiënter kan werken. We kunnen een betere afbakening maken tussen informatie die wel relevant is en informatie die er voor het onderhavige incident niet toe doet. Dat brengt een flinke tijdswinst met zich mee en dat wil bij politiewerk zeggen dat de kans dat een dader gepakt wordt flink toeneemt."

Audio en video

Bovendien maakt datamining het mogelijk dat de politiemanager en -vrouw zich beter kan voorbereiden op situaties die zich kunnen voordoen. "Als we bij een bepaalde situatie bijvoorbeeld kunnen voorspellen volgens welke route mensen de plek van een incident zullen verlaten, weten we ook waar we – zeg – een ME-peloton moeten opstellen om de situatie niet verder uit de hand te laten lopen. Ook kunnen we dankzij die voorspelling zorgen voor een betere begeleiding van mensen die vertrekken."

Tot nu toe beperkt het gebruik van DataDetective zich nog tot informatie die vastligt in databases en in tekstuele documenten. Beelden van bewakingscamera's en de bijbehorende audio-opnames vallen nog niet binnen de scope van dit project.

J)Van der Veer van Sentient ziet een dergelijke combinatie zeker zitten. Sterker nog, de technologie om dergelijke data eveneens in een datawarehouse onder te brengen bestaat allang. "Er zijn helaas alleen nog geen klanten die er om vragen."

Robbert Hoeffnagel is freelance journalist.