
KENNIS VERGARING IN COMMAND & CONTROL

SAMENVATTING VAN DE SCRIPTIE
KENNIS VERGARING IN COMMAND & CONTROL,
IMPLEMENTATIE VAN NON-SENSOR
INFORMATIE IN COMMAND & CONTROL

 Tlnt S.F.F. (Bas) van Odenhoven BSc.
Jr. Engineer F-16 Avionica

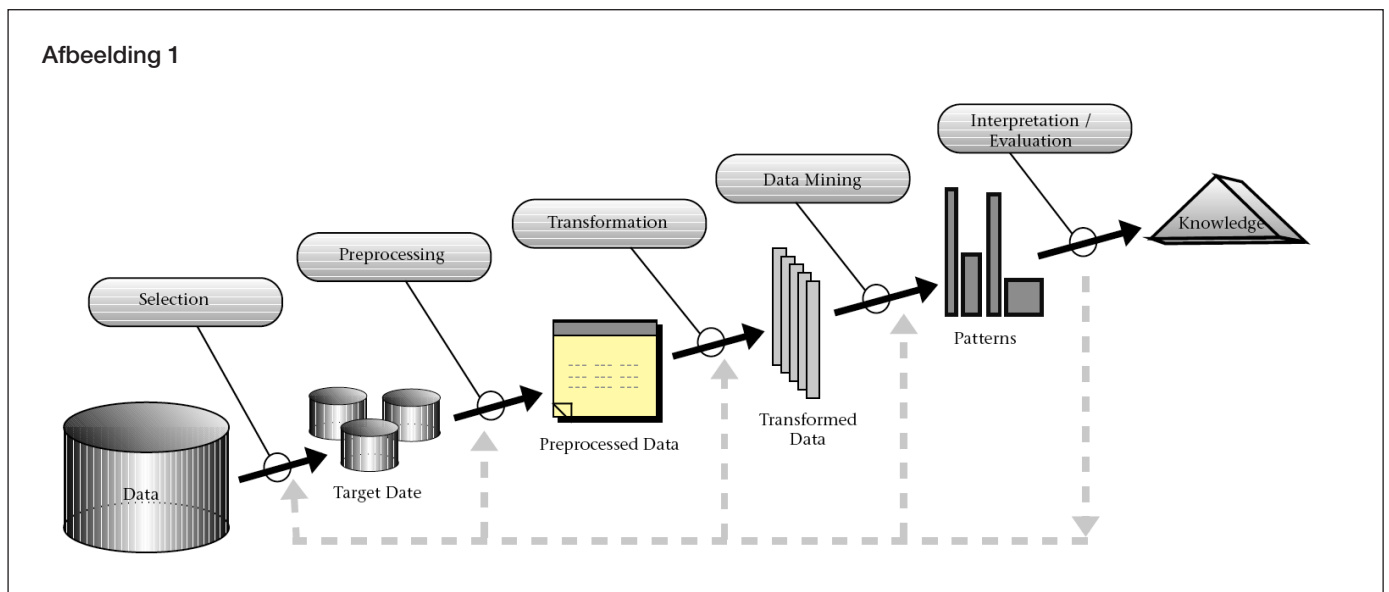
Hedendaagse Command & Control (C2) systemen dragen zorg voor het ontvangen, verwerken en weergeven van sensordata. Op basis van de projectie van deze data, wordt de data omgezet in informatie voor de operator van het C2-systeem. De uitdaging hiervoor ligt in het zoeken naar een generieke benadering voor het verzamelen, fuseren en verwerken van sensorinformatie met non-sensorinformatie. Hiertoe is voor het versterken van de inlichtingenpositie in het digitale domein van Defensie gezocht naar een generieke benadering om de Situational Awareness (SA) van de operator van een Command & Control systeem te verbeteren. →

De operator kan met deze informatie, middels het overwegen van opties, beslissingen nemen.

Waar deze sensorinformatie veelal bestaat uit kinematische informatie zoals positie, snelheid, versnelling en koers van een doel en gezien het feit dat missies van de huidige krijgsmacht met toenemende mate complexer worden, wordt de behoefte van de operator aan meer dan alleen sensorinformatie ook groter. Deze additionele, non-sensor informatie kan onder meer verkregen worden door het toepassen van data mining, waarbij gezocht wordt naar verbanden in verscheidene on- en offline databases,

seerd. Vervolgens wordt er een data set gemaakt door middel van selectie. Door het selecteren van – voor het doel – relevante data wordt er een focus gelegd op subsets van databases. Op deze subsets wordt de kennisvergarings toegepast.

Aansluitend wordt de data gefilterd van ruis (onbruikbare data) en gemodelleerd in de preprocessing fase. Daarna wordt de pre-processed data getransformeerd tot bruikbare data, die gebruikt kan worden om het doel van de KDD te bereiken. Vaak wordt dit gedaan door transformatiemethodes om het aantal variabelen



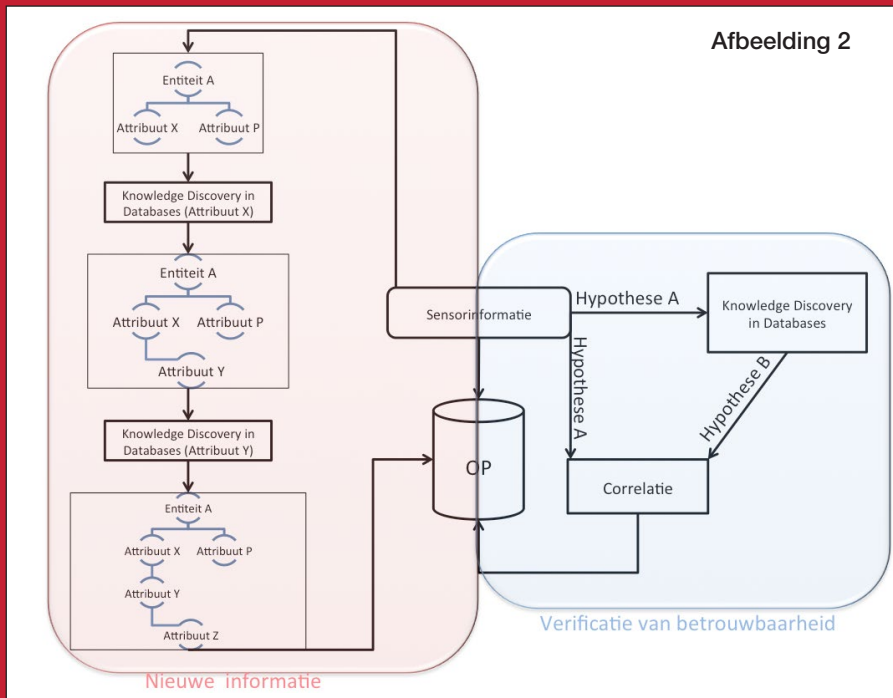
Knowledge Discovery in Databases model

expertise van operators, spraak en tekst. Om sensorinformatie aan te vullen met non-sensorinformatie – zonder de operator onnodig te belasten –, is een concept ontwikkeld met behulp van het Knowledge Discovery¹ model. De schematische weergave van het Knowledge Discovery model is te zien in afbeelding 1 en wordt getoetst aan het opereren in C2-concept. Het vergaren van kennis door middel van het zoeken in databases kan worden aangeduid met het begrip Knowledge Discovery in Databases (KDD). Dit is een niet-triviaal proces waarbij valide, nieuwe, potentieel bruikbare en uiteindelijk begrijpelijke patronen in data geïdentificeerd worden. Het KDD proces is niet-triviaal aangezien het zoeken naar bruikbare data geen vooraf gedefinieerde grootte heeft. De theorie beschrijft een proces waarbij automatisch grote hoeveelheden data doorzocht worden op patronen die kennis over de data kunnen geven. Door het doorlopen van een aantal iteratieve stappen, wordt er kennis vergaard.

Als eerst moet er een begrip worden ontwikkeld van de omgeving waarover kennis moet worden gegenereerd. Hierbij worden relevante voorkennis en het doel van het KDD proces geanaly-

in de dataset te verminderen, of om irrelevante representaties voor data daar uit te halen. Aansluitend wordt een data mining algoritme toegepast op de data. Welk algoritme dit is, hangt af van het doel van de operator. Tijdens het data minen wordt er gezocht naar patronen in de data met een gedefinieerde vorm. Deze patronen zijn expressies in een taal die een subset van de data beschrijft. Na het data mining proces worden de patronen geïnterpreteerd door de gebruiker. De patronen gaan daarna opnieuw door het KDD proces voor meer iteraties. Een iteratie geeft het systeem de gelegenheid om meer kennis te genereren over voorafgaande patronen en informatie. De gebruiker van het KDD proces kan bepalen hoeveel iteraties er benodigd zijn bij voldoende informatie. De gebruiker kan zelf bepalen of er genoeg iteraties doorlopen zijn, of kan aangeven dat er een (on)zekerheidsgraad behaald moet worden door het proces. Ten slotte worden de patronen geëvalueerd door de gebruiker waarbij kennis ontstaat (zie afbeelding 1).

De vergaarde kennis kan in verschillende databases of systemen worden opgeslagen voor documentatie en verslaglegging, of



Hypothese analyse door KDD

voor andere doeleinden. Uit patronen die gevonden worden uit het KDD proces, wordt nieuwe data gegenereerd. Deze data bevat een bepaalde zekerheidsgraad. Een patroon kan nieuwe kennis genereren wanneer deze potentiële kennis voldoet aan een – door de gebruiker opgestelde – grens.

Doelen Knowledge Discovery

Ten eerste zorgt KDD voor verificatie van de nauwkeurigheid² en betrouwbaarheid³ van informatie en hypothesen. De hypothese A van het C2-systeem – gebaseerd op sensorinformatie – is bijvoorbeeld: 'Entiteit A heeft klasse X en identiteit Y'. Door Knowledge Discovery toe te passen, kan deze hypothese bevestigd of ontkracht worden: het KDD proces geeft de volgende hypothese B: 'Entiteit A heeft klasse X en identiteit Z'.

Wanneer de sensorinformatie gecorrigeerd wordt met kennis uit het KDD, kan er bepaald worden wat de nauwkeurigheid en betrouwbaarheid is van de sensorinformatie of van de informatie uit het KDD proces. Wanneer de vergelijking tussen deze twee hypothesen in overeenstemming is, geeft dit bevestiging aan de

operator van het C2-systeem. Hierdoor wordt het classificeren- en/of identificatieproces versneld. Indien de vergelijking tussen beide hypothesen A en B niet in overeenstemming zijn met elkaar, geeft dit een discrepantie tussen beide hypothesen. Als een discrepantie optreedt, moet de operator een uitgebreidere analyse van de entiteit maken. Hij kan er bijvoorbeeld voor kiezen om andere sensoren aan te sturen.

Ten tweede kan Knowledge Discovery, zoals de term al impliceert, zorgen voor ontdekking van nieuwe informatie over entiteiten of nieuwe hypothesen. Sensorinformatie geeft bijvoorbeeld een attribuut X van entiteit A. Met het datamining proces wordt gevonden dat X een verband heeft met attribuut Y. Dit attribuut wordt vervolgens opgeslagen in het Operational Picture (OP). Daarna wordt Y onderzocht, en blijkt dat attribuut Z een verband heeft met attribuut Y, terwijl X en Z geen direct verband hebben met elkaar. Hierdoor ontstaat er een nieuwe hypothese: 'Entiteit A bezit de attributen X, Y en Z. Deze nieuwe hypothese kan, indien de operator dit nodig vindt, opnieuw worden getest.

Dit proces is schematisch weergegeven in afbeelding 2.

Hypothese analyse door KDD

Bovendien zorgt Knowledge Discovery er voor dat de overige taken in het Information Management proces (estimate risk, detect anomaly, recognise intent) ondersteund worden middels de verificatie van betrouwbaarheid en vergaring van nieuwe informatie. Voor het toepassen van KD heeft het proces databases nodig waar informatie uit gehaald kan worden.

Ontwerp database

Voor het zoeken naar data in datasets zijn databases benodigd. Er zijn een tal van geschikte databases die voor KD gebruikt kunnen worden. Militaire platformen kunnen op twee manieren databases benaderen: offline databases, welke niet in verbinding staan met het internet, en online databases die middels het internet geraadpleegd kunnen worden.

Een online database kan in de vorm zijn van een semantisch web. Het semantisch web is gebaseerd op ontologieën, welke de structuur en een model van data weergeven met bijvoorbeeld de standaard Resource Description Framework (RDF). Een RDF is een standaard model voor het uitwisselen van meta-data op het web.

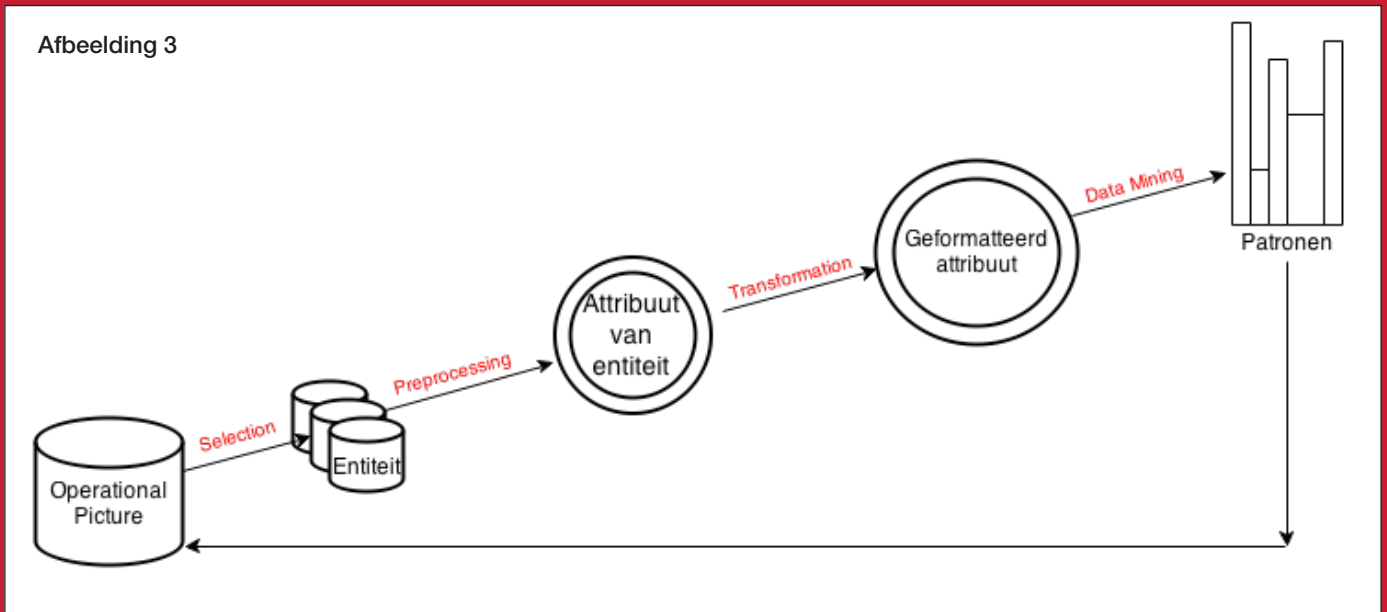
Met deze taal wordt er een structuur opgebouwd met 'platte' data met behulp van tags. Een attribuut wordt beschreven door de waarde binnen twee tags te zetten op de volgende manier:

<attribuut>waarde</attribuut>

Het attribuut wordt begonnen met een opening-tag (<>) en afgesloten met een closing-tag (< / >).

Met het gebruik van attributen kunnen eigenschappen aan data gekoppeld worden, zoals bij de website Wikipedia gedaan wordt. Wikipedia is uitgegroeid tot de grootste open, online encyclopedie. Hierdoor zal er rekening gehouden ➔

Afbeelding 3



Generieke benadering Knowledge Discovery in C2-systemen

moeten worden met een grotere onzekerheidsgraad. Voor informatievergaring in C2-systemen is een generieke benadering te geven. In het model is de grootste database representatief voor het OP. De 'Target Data' uit het KDD model moet geselecteerd worden uit het OP. Dit is de entiteit die moet worden onderzocht door het KDD proces. Uit deze analyse moet een attribuut van de entiteit geselecteerd worden waarop Knowledge Discovery toegepast kan worden. Dit attribuut moet een formaat krijgen om door een query uitgevoerd te worden.

Met het geformateerde attribuut moet de agent datamining toepassen om patronen te vinden die bruikbaar zijn voor het C2-systeem. Met de meest betrouwbare en nauwkeurige bronnen waarover het C2-systeem beschikt, zoekt de agent naar patronen in bronnen aan de hand van de gevraagde informatie door de operator. De uitkomst van deze patronen wordt in het OP opgeslagen. De operator van het C2-systeem ziet vervolgens dat er additionele informatie over de entiteit in het OP staat. Indien deze informatie nuttig blijkt te zijn voor de operator, en hij vervolgacties kan nemen door deze additionele informatie, wordt de informatie gezien als kennis. Deze generieke benadering is weergegeven in afbeelding 3. Een operator kan met de verkregen kennis bepalen welke acties er ondernomen moeten worden.

Agent voor datamining

Omdat er geen afspraken zijn gemaakt over het invullen van attributen uit bronnen, wordt het bevragen van een database door bijvoorbeeld query's erg lastig. De operator weet immers niet wat voor type informatie de query terug geeft na bevraging. Hierbij zou een goal-based agent een uitkomst bieden. Dit type agent kan bepalen wat het resultaat is van een actie, alvorens deze

actie daadwerkelijk uit te voeren. Daarnaast kan de agent ook als een goal-based learning agent worden ontworpen, zodat in de toekomst niet dezelfde fouten begaan worden. Wanneer de agent bijvoorbeeld het attribuut 'kenteken' opzoekt, en de operator geeft aan dat 'kentekenplaat' hetzelfde attribuut is, kan de agent bepalen dat voortaan de resultaten van zowel 'kenteken' als 'kentekenplaat' weergegeven moeten worden.

Een operator van een C2-systeem weet niet altijd welke informatie een database bevat. Dit kan opgelost worden kan door middel van bepaalde labels aan attributen en databases te koppelen. Wanneer een operator bijvoorbeeld op zoek is naar kentekengegevens van een auto, voert hij de labels 'auto' en 'kenteken' in.

Hierna weet het C2-systeem welke databases er geraadpleegd moeten worden voor de informatie die de operator wil. In plaats van alle databases af te zoeken of deze wel 'kenteken' als attribuut erin hebben, hoeft het C2-systeem slechts de databases door te zoeken die de labels 'auto' en 'kenteken' bevatten. Hierdoor hoeft bijvoorbeeld de kentekendatabase voor motorfietsen of vrachtwagens niet doorzocht te worden.

Conclusies en aanbevelingen

De resultaten van het onderzoek tonen aan dat directe informatiebronnen automatisch aangevuld kunnen worden met indirecte informatiebronnen, zonder hiermee de operator onnodig te belasten. Hiermee wordt een betere Situational Awareness gecreëerd en wordt de inlichtingenpositie in het digitale domein versterkt. In dit onderzoek is slechts een klein deel van het Knowledge Discovery proces uitgelicht, waarmee aangetoond wordt dat directe informatiebronnen aangevuld kunnen worden met indirecte informatiebronnen. Uit het onderzoek is gebleken dat opensource databases zoals Wikipedia sterk vervuild kunnen zijn en daar-

mee niet erg betrouwbaar zijn. Hierdoor zal een database eerst geschikt gemaakt moeten worden voor gebruik.

Door implementatie van een parser in het C2-systeem kan de binnenkomende informatie geanalyseerd worden op juistheid. Door een open source database te complementeren met valide informatie uit betrouwbaardere bronnen wordt een hybride database gecreëerd. De implementatie van een softwarematige intelligent agent kan van grote waarde zijn. Daarnaast is tijdens simulaties de aanname gedaan dat er geen onzekerheid in de sensorinformatie zit. Een aanbeveling is om te kijken wat de uitwerking is van onzekerheid in sensorinformatie. Bij dit proces moeten andere processen zoals estimate risk, recognise intent en detect anomaly betrokken worden om een goede analyse te maken.

Een andere aanbeveling is om te kijken naar andere type databases en de structuur/ontologie van deze databases. Het is daarnaast een aanbeveling om een query builder in het C2-systeem te programmeren. Bovendien zorgt een geavanceerdere parser er voor dat de patronen die het C2-systeem inkomen, beter geanalyseerd kunnen worden op juistheid. Voor het concretiseren van het Knowledge Discovery model moet aan een operationele gebruiker van een C2-systeem gevraagd worden welke informatie in welk scenario het meest geschikt is.

Het is interessant om de data die open source bronnen leveren aan het C2-systeem, te kwalificeren aan informatie-kwaliteitsmodellen. In deze beoordeling wordt de kwaliteit van een indirecte informatiebron beschreven, waarmee deze vergeleken kan worden met andere bronnen. Op deze wijze kan de operator er voor kiezen welke criteria hij stelt aan de database.

Tot slot wordt er aanbevolen om te kijken hoe het Knowledge Discovery model



Bas van Odenhoven (links op de foto), 03-03-1991, tweede-luitenant bij de Koninklijke Luchtmacht, april 2014 afgestudeerd aan de Nederlandse Defensie Academie (NLDA)/Koninklijke Militaire Academie (KMA). Genoten van een driejarige Bachelor Militaire Systemen & Technologie, met afstudeerrichting C4I/C2 systemen te Den Helder. In opdracht van de NLDA en uit interesse van onderwerpen zoals Big Data, Command & Control- en wapensystemen, automatiseringsprocessen en communicatiesystemen heeft hij de afstudeerscriptie 'Knowledge Discovery in Command & Control' geschreven.

past in overige onderzoeken. Met name de onderzoeken met raakvlak aan de overige taken in het 'Information Management' proces zoals estimate risk, recognise intent en detect anomaly verdienen de voorkeur. Daarnaast is het interessant om te beoordelen hoe dit onderzoek zich vertaalt in een taak-georiënteerde pro-

grammeeromgeving in plaats van een object-georiënteerde programmeeromgeving. Mede de efficiëntie en potentie van deze programmeertalen kunnen bruikbaar zijn om het Knowledge Discovery proces op een juiste manier te implementeren in het C2-proces en daarmee te optimaliseren. 🔄

¹ U. M. Fayyad, G. Platetsky-Shapiro, P. Smyth, et al., "Knowledge discovery and data mining: Towards a unifying framework.," in Knowledge Discovery In Databases, vol. 96, 1996. 5, 6

² Nauwkeurigheid wordt gezien als de graad van overeenstemming van een gemeten of berekende hoeveelheid met zijn daadwerkelijke (ware) waarde. Hoe groter de nauwkeurigheid hoe kleiner de totale fout.

³ De betrouwbaarheid van informatie wordt gezien als een maat van zekerheid dat een gemeten waarde representatief is voor de werkelijke waarde.