

Zelflerende software detecteert opvallende transacties

Onderzoek naar de mogelijkheden om concepten uit de kunstmatige intelligentie in te zetten voor analyse van financiële gegevens

Quintra Rijnders, Patrick Özer, Vivian Blankers en Thom Eijken

SAMENVATTING Het in kaart brengen van opvallende transacties in het kader van de accountantscontrole wordt vaak gedaan vanuit een traditionele benadering van een kwalitatieve risicoanalyse en een kwantitatieve (data-)analyse gebaseerd op ervaringsregels. Hierbij worden risico-indicatoren die niet betrokken zijn in voornoemde kwalitatieve risico- en data-analyse buiten beschouwing gelaten. In dit artikel wordt aangetoond dat 'unsupervised' zelflerende software op effectieve en efficiënte wijze uitzonderingen op bestaande datastructuren in een grootboek in kaart brengt. De software vormt een aanvulling op het pallet van oplossingen dat kan worden ingezet in het kader van het voortdurend monitoren van de bedrijfsprocessen.

RELEVANTIE VOOR DE PRAKTIJK De software die in dit artikel onderzocht is, detecteert zelflerend uitzonderingen op bestaande datastructuren, waardoor een grootboek integraal wordt geanalyseerd op mogelijke opvallende transacties. De software kan onder andere ingezet worden in het kader van *continuous monitoring*.

1 Inleiding

De populariteit en inzetbaarheid van data-analyse bij audits en fraudeonderzoeken zijn in de afgelopen jaren sterk toegenomen. Schouten (2007) geeft aan dat er veel goede toepassingsvoorbeelden voorhanden zijn, zowel binnen de accountantscontrolepraktijk als daarbuiten. Ook in het fraudeonderzoek blijkt data-analyse in de praktijk een uitstekend middel bij het snel doorgronden van grote hoeveelheden financiële informatie om de spreekwoordelijke speld in de hooiberg te vinden. Dit heeft geleid tot de ontwikkeling van veel verschillende data-analyse-software voor toepassing binnen de accountantscontrole en daarbuiten. Deze tools werken op basis van regels die specificeren waar de dataset op doorzocht dient te worden. De set van regels is tot stand gekomen vanuit de ervaringen in audits en fraudeonderzoeken over de jaren heen.

Naast de inzet van data-analyse bij repressieve opdrachten, zoals audits, fraude- en feitenonderzoeken, wordt data-analyse steeds meer op detectieve basis ingezet, als middel voor risico-identificatie. Met de hedendaagse technologie kan de binnen bedrijven aanwezige data integraal in beschouwing genomen worden. Vervolgens kan gezocht worden naar indicatoren die duiden op het omzeilen van de interne beheersingsmaatregelen. Daaruit kunnen risico's worden geïdentificeerd. Voorbeelden van dergelijke risico-indicatoren zijn boekingen door ongeautoriseerde medewerkers of het boeken van kosten in een reeds gesloten periode.

De repressieve en detectieve eigenschappen van data-analyse worden gecombineerd in *continuous auditing* en *monitoring*, waarbij bedrijfsinformatie (bijna) real time wordt gescreend op risico-indicatoren en afwijkingen ten opzichte van gedefinieerde interne beheersingsmaatregelen. Dit betekent dat bijvoorbeeld bij het aannemen van een boeking van een inkoopfactuur direct de 'three way match' wordt uitgevoerd, waarbij uitzonderingen boven een bepaalde drempel middels een e-mail worden gemeld aan de interne accountantsdienst. Uitzonderingen die door de 'three way match'-controle zijn gedetecteerd, kunnen nader worden onderzocht (repressief) en kunnen input geven voor het aanscherpen van risico's en beheersingsmaatregelen binnen het inkoopproces (detectief).

Het begrip *continuous monitoring* is al geruime tijd onderwerp van discussie (Dassen, 2003; Mashaie, 2008; Strikwerda, 2008; zie ook dossier over XBRL op Accountant.nl). De afgelopen jaren is in de praktijk een beweging zichtbaar richting het daadwerkelijk nagenoeg real time inzetten van *continuous monitoring*-software en -technieken.

De voornoemde trends op het gebied van data-analyse en *continuous monitoring* laten één vraag onbeantwoord:

wat zijn nu de onontdekte signalen van risico's die gelopen worden? De Haas (2003) vroeg zich al af wat de waarde van zekerheid is in een digitale omgeving. Zowel repressieve en detectieve data-analyse software als continuus monitoring-systemen zijn geprogrammeerd met ervaringsregels en leren niet – of in zeer beperkte mate – zelfstandig op basis van de data waarmee ze worden gevoed. Dit impliceert dat risico-indicatoren die niet binnen de voornoemde ervaringsregels vallen, niet worden gedetecteerd door deze software en systemen.

Bij creditcardmaatschappijen en in de verzekeringsbranche is het inzetten van zelflerende systemen reeds enkele jaren gebruikelijk. Creditcardmaatschappijen gebruiken zelflerende systemen voor het detecteren van creditcardfraude. Deze systemen bepalen een gebruikerspatroon voor iedere kaarthouder aan de hand van de transacties die deze verricht. De creditcardmaatschappij neemt direct contact op met de kaarthouder indien hij een transactie verricht die afwijkt van het gebruikerspatroon.

Verzekeringsmaatschappijen zetten zelflerende systemen in om declaraties te screenen en zodoende externe fraude-signalen te detecteren. In een interview noemt Verkruijsse (Koopmans, 2005) ook de inzet van neurale netwerken¹ bij een financiële dienstverlener, om online aanvragen te toetsen aan een norm. Voornoemde neurale netwerken herkennen patronen in de opgeslagen gegevens en vergelijken de nieuwe aanvragen met de bestaande patronen, om zodoende afwijkende aanvragen door een medewerker te laten controleren.

Het inzetten van het Bayesiaanse gedachtegoed² ten behoeve van het analyseren van financiële data, zelfs binnen de accountantscontrole, is niet nieuw. Veenstra (2005, p. 83) beschreef al hoe Bayesiaanse statistiek kan worden gebruikt bij het uitvoeren van steekproeven. Veenstra en Heertje (2006) beschrijven hoe Bayesiaanse statistiek gebruikt kan worden voor het bepalen van steekproefgrootte door harde voorinformatie te interpreteren als kansverschijnsel. Echter, in dit artikel beschrijven wij ons onderzoek naar de inzet van zelflerende software gebruikmakend van Bayesiaanse netwerken met als specifiek doel het detecteren van opvallende transacties. Wij testen deze zelflerende software op zes datasets van vijf organisaties die gebruikmaken van Oracle Financials. Het doel van dit artikel is om te onderzoeken of zelflerende software geschikt is om op effectieve en efficiënte wijze opvallende transacties te detecteren. De door ons onderzochte zelflerende software is gebaseerd op het concept van een Bayesiaans netwerk, dat inhoudelijk nader zal worden toegelicht bij het theoretisch raamwerk.

In dit artikel worden eerst de theoretische uitgangspunten behandeld waarop de software is gebouwd. Voorts worden

onze onderzoeksmethodiek, de resultaten, de conclusie en beschouwingen op ons onderzoek beschreven. Het artikel eindigt met een paragraaf over toekomstig onderzoek waardoor functionaliteiten van de onderzochte software verder uitgebreid zouden kunnen worden.

2 Theoretisch raamwerk

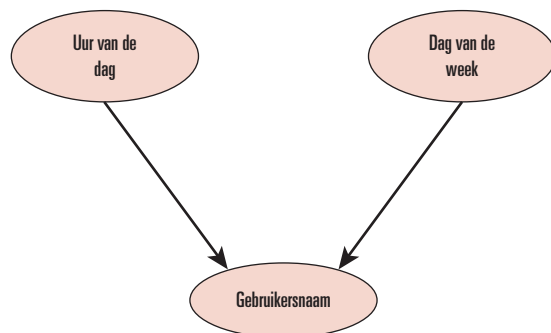
In deze paragraaf gaan wij dieper in op de onderliggende methodiek van de onderzochte zelflerende software. Deze methodiek komt voort uit een stroming binnen de kunstmatige intelligentie genaamd 'data mining'. Data mining is het (semi)automatische proces van het ontdekken van verbanden in een vaak substantiële verzameling van gegevens (Witten en Frank, 2005). Deze verbanden kunnen vervolgens worden gebruikt voor het behalen van economische voordelen. Data mining-technieken kunnen 'supervised' of 'unsupervised' zijn.

Een supervised methode kan gebruikt worden voor het maken van beslissingen op basis van historische gegevens. Verzekeringsmaatschappijen maken gebruik van supervised data mining-technieken om te beoordelen of een declaratie mogelijk frauduleus is, gebruikmakend van kenmerken van frauduleuze en legitieme declaraties uit het verleden als ervaringsgegevens. Hierbij geldt uiteraard de voorwaarde dat men weet welke declaraties uit het verleden frauduleus en welke legitiem waren.

In gevallen waarin men niet weet wat frauduleus en legitiem is, kan een unsupervised methode gebruikt worden. Een unsupervised methode kan uitzonderingen in een dataset opsporen en behoeft daarvoor geen historische gegevens. Grootboekdata vereisen over het algemeen een dergelijke unsupervised aanpak, aangezien men veelal weinig tot geen gegevens heeft over opvallendheden in het verleden. Daarnaast hoeven de opvallendheden die in het verleden hebben plaatsgevonden geen voorspellende karakteristieken te bevatten die relevant zijn voor de toekomst. Over het algemeen zullen opvallendheden, indien zij ontdekt zijn en zeker als er sprake is van mogelijke onregelmatigheden, geredresseerd zijn, waardoor herhalen wordt voorkomen. Bovendien kunnen zowel fouten in het financiële proces als mogelijke onregelmatigheden op zeer uiteenlopende wijzen in de financiële verantwoording tot uitdrukking komen. Met betrekking tot fraude wordt dit al geïllustreerd door de verschillende soorten fraude die benoemd worden in de toepassingsgerichte en overige verklarende teksten van NV COS 240. In deze teksten staan uitwerkingen en voorbeelden van frauduleuze financiële verslaggeving en het ontvreemden van waarden. In bijlage 1 van COS 240 worden, uitgesplitst naar de componenten van de fraudedriehoek, meer dan 60 voorbeelden van frauderisico's genoemd.

De basis van de unsupervised data mining-techniek die ten grondslag ligt aan de hier beschreven software is een 'Bayesiaans netwerk' (hierna: BN). Een BN is een datastructuur die gebruikt wordt voor het modeleren van kansverdelingen. Een voorbeeld van een versimpelde versie van een BN voor het maken van een boeking in een financieel programma is gegeven in figuur 1.

Figuur 1 Voorbeeld netwerk (vereenvoudigd)



Een aantal regels uit de bijbehorende eenvoudige voorbeelddataset is gegeven in tabel 1. Elke knoop in een BN correspondeert met een kolom in de dataset. Voor figuur 1 zijn dit de kolommen 'Uur van de dag', 'Dag van de week' en 'Gebruikersnaam'.³ De pijlen in een BN geven aan welke kolommen in de dataset van elkaar afhangen. Men zou verwachten dat de gebruikersnaam afhankelijk is van de dag van de week en het uur van de dag waarop de boeking plaatsvindt, aangezien de gebruikers in een financieel systeem vaak min of meer vaste werktijden hebben. Naast de knopen en de pijlen bevat een BN ook informatie over de manier waarop de pijlen werken, oftewel: op welke uren/dagen maakt men de boekingen? Deze informatie is verwerkt in de kansverdelingen in het BN. Deze kansverdelingen kunnen vervolgens gebruikt worden om te bepalen wat de kans is dat er bijvoorbeeld op zaterdagmiddag om drie uur een boeking wordt gemaakt door een bepaald persoon.

In het eenvoudige voorbeeld dat geïllustreerd is in tabel 1 en figuur 1 kan een BN handmatig geconstrueerd worden. In de praktijk zijn datasets niet zo eenvoudig als het voorbeeld gegeven in tabel A, waardoor het handmatig bepalen van een BN een bijna onmogelijke taak is.

Tabel 1 Voorbeeld dataset (vereenvoudigd)

Dag van de week	Uur van de dag	Gebruikersnaam
Maandag	9.15	Jan
Donderdag	15.26	Piet
...	...	...

De meeste datasets die in de huidige tijd bij organisaties worden bijgehouden, zijn onderdeel van complexe ERP-systemen, zo ook boekhoudkundige data. Om die reden zijn er methoden ontwikkeld ten behoeve van het leren van een netwerk dat van toepassing is op de data, op basis van de data zelf (zie voor een uitgebreid overzicht: Heckerman, 1996). Dit leerproces bestaat uit twee stappen.

De eerste stap is het leren van de structuur van het BN. Hierbij worden de onderlinge afhankelijkheden van de verschillende kolommen in een dataset berekend. Toegepast op het voorbeeld in figuur 1 wordt door de software bepaald tussen welke knopen pijlen moeten worden geplaatst. Voor deze stap maakt de software gebruik van een algoritme met de naam 'Max-Min Hill-Climbing' (Tsamardinos, Brown en Aliferis, 2006). Deze methode beperkt eerst het aantal plekken waar een pijl mag komen te staan aan de hand van een statistische toets. Vervolgens zoekt het algoritme welke combinatie van pijlen de hoogst verklarende waarde bevat voor de dataset.

De tweede stap is het leren van de kansverdelingen binnen het BN. Hiervoor maakt de software gebruik van een Maximum Likelihood Estimator (hierna: MLE) (Aldrich, 1997). Dit is een methode waarmee een schatting wordt gemaakt van de kansverdelingen tussen de knopen van de eerder gevormde structuur. De MLE maakt de schatting van kansverdelingen op basis van hoe vaak bepaalde waarden binnen kolommen in de dataset voorkomen.

Nadat een BN is geleerd, gebruikmakend van de twee hiervoor beschreven stappen, kent de software een kans toe aan elke regel in de dataset. Deze kans geeft aan hoe gebruikelijk het is dat een regel in de dataset voorkomt. Omgekeerd kan deze kans geïnterpreteerd worden als een score voor opvallendheid. Hoe kleiner de kans die toegekend wordt aan een regel, hoe opvallender de combinatie van waarden in de kolommen van de regel is. Om deze reden kunnen regels met een significant kleinere kans worden aangewezen als zijnde opvallende regels (zie ook paragraaf 3.3).

Het leren van een BN is een generiek proces, waardoor de software in theorie gebruikt kan worden voor het signaleren van opvallendheden in iedere dataset. Echter, om tot bruikbare resultaten te kunnen komen, moeten de kolommen in de dataset wel zinvol zijn voor analyse. Het detecteren van opvallendheden in een personeelsbestand met daarin uitsluitend de kolommen personeelsnummers en NAW-gegevens zal bijvoorbeeld weinig resultaten van betekenis opleveren, omdat in deze data geen (procesmatige) patronen aanwezig zijn.

3 Onderzoeksmethode

3.1 Zelflerende software

De zelflerende software die is geïmplementeerd voor ons onderzoek werkt volgens een aantal stappen. Eerst wordt een door de gebruiker meegegeven dataset ingelezen. Vervolgens wordt een BN geleerd op basis van de ingelezen dataset en gebruikmakend van de in paragraaf 2 beschreven methodiek. Daarna worden regels uit de dataset geselecteerd die aan de hand van het BN opvallend zijn. Deze opvallende regels en de patronen in het BN worden tot slot aan de gebruiker teruggekoppeld. Wij hebben de werking van de software gecontroleerd volgens de theoretische benadering zoals deze is beschreven in paragraaf 2, gebruikmakend van vereenvoudigde datasets.

3.2 Onderzoekspopulatie

Om te onderzoeken of de zelflerende software effectief en efficiënt opvallende transacties detecteert in een dataset uit de praktijk, hebben wij vijf organisaties bereid gevonden om ons een digitale kopie van hun grootboek te verstrekken. Enerzijds hebben wij gekozen voor vijf organisaties uit verschillende typologieën, om te voorkomen dat de resultaten alleen op een specifieke typologie van toepassing zijn. De vijf organisaties zijn anderzijds geselecteerd op het gebruik van het ERP-pakket Oracle Financials. Hoewel de software niet afhankelijk is van een bepaald ERP-pakket of van bepaalde datastructuren richt het onderhavige onderzoek zich op Oracle Financials-datasets. Wij hebben gekozen voor één ERP-pakket om de verschillen tussen de onderliggende betekenis van de brongegevens zo klein mogelijk te houden.

Binnen de onderzoekspopulatie hebben wij ons gericht op alle financiële boekingen in het grootboek.⁴ De onderzochte software baseert zich niet op ervaringsregels en leert op basis van de ingelezen data. Naarmate er meer data beschikbaar is, kan de software meer patronen en meer opvallende transacties ontdekken. Voor ons onderzoek hebben we ervoor gekozen om zowel de handmatige boekingen als de automatische (batch) boekingen te analyseren. Met behulp van beide type boekingen leert de software tevens het verschil tussen de twee type boekingen.

3.3 Voorbereiding van data voor analyse

De vijf organisaties hebben elk grootboekdata voor het financiële boekjaar 2009 aangeleverd. Eén van de organisaties heeft daarnaast grootboekdata voor het eerste kwartaal van 2010 aangeleverd. Een andere organisatie leverde naast de grootboekdata voor het jaar 2009 ook grootboekdata voor het jaar 2008 aan. Bij deze laatstgenoemde organisatie is het daardoor mogelijk om de patronen in de datasets uit verschillende jaren te vergelijken. Daarom hebben we

ervoor gekozen om voor deze organisatie de jaren 2008 en 2009 apart te analyseren. Als gevolg daarvan omvat onze test zes datasets.

De zes datasets zijn geïmporteerd in een testomgeving. Om er zeker van te zijn dat de datasets juist en volledig zijn geïmporteerd en geen journaalregels dubbel aanwezig zijn of ontbreken, is een aansluiting gemaakt met de saldobalans voor het desbetreffende boekjaar.

De software werkt met bepaalde kolommen van de data uit het grootboek. Het gaat hierbij om de kolommen die interessant zijn voor analyse. Het nummer van de journaalpost is bijvoorbeeld heel nuttig om een boeking te identificeren, maar niet interessant voor het identificeren van patronen en afwijkingen daarop. Voor ons onderzoek zijn de interessante kolommen onder andere het bedrag, de grootboekrekening, het type transactie, de gebruikersnaam en de verschillende datumkolommen. Deze kolommen zijn onder andere de input voor de analyse door de software.

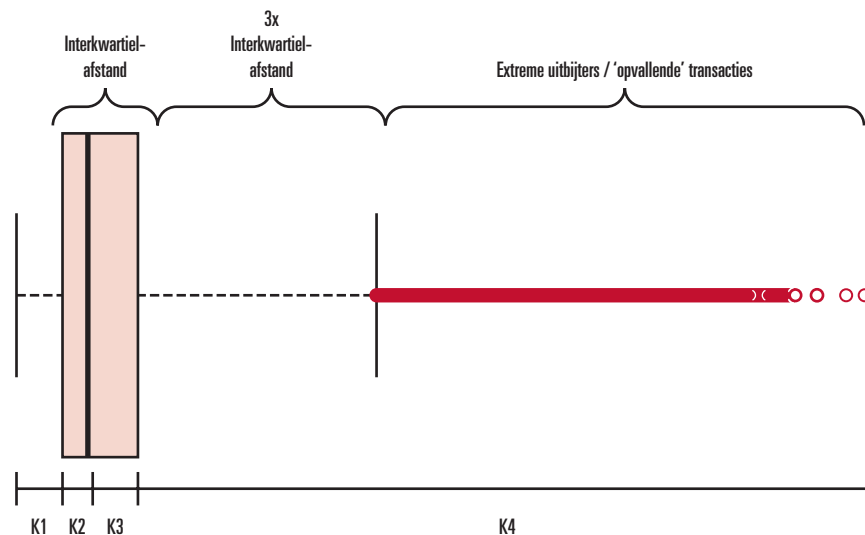
Zoals beschreven in het theoretisch raamwerk wordt door de software met behulp van het geïdentificeerde BN een score voor opvallendheid toegekend aan elke regel van het ingelezen grootboek. Voorts sorteert de software de data op basis van opvallendheid. De software groepeerde de data in vier even grote delen op aantallen transacties. Elk van deze groepen wordt een kwartiel genoemd. De afstand van het einde van het eerste kwartiel tot het begin van het vierde kwartiel wordt de interkwartielafstand (Upton en Cook, 1996, p. 55) genoemd. Deze afstand wordt gemeten in opvallendheid.

De interkwartielafstand kan gevisualiseerd worden door middel van een boxplot⁵, zie figuur 2. In ons onderzoek liggen opvallende regels meer dan drie keer de interkwartielafstand van de box (die de interkwartielafstand weergeeft) af. Drie keer de interkwartielafstand is een algemeen uitgangspunt in de statistiek om extreme uitbijters aan te duiden. Deze extreme uitbijters definiëren wij ten behoeve van dit onderzoek als ‘opvallend’.

De efficiency van de onderzochte software om opvallende transacties te identificeren is te meten aan de hand van de computertijd die de software nodig heeft om een BN te vinden en een lijst met opvallende transacties te genereren. Voorafgaand aan de test verwachtten wij een computertijd die afhankelijk van de omvang van de dataset uiteen zou lopen tussen enkele uren en enkele dagen.

3.4 Evalueren van de analyse

De software levert per ingelezen dataset het gevonden BN op en een lijst met opvallende transacties. Het gevonden

Figuur 2 Boxplot

netwerk en de opvallende transacties zijn per dataset teruggekoppeld aan de respectievelijke organisaties. De effectiviteit van de software komt vervolgens tot uitdrukking in de mate waarin de in de test betrokken organisaties de geïdentificeerde opvallende transacties aanmerken als zijnde opvallende transacties en nader onderzoek verrichten naar de opvallende transacties.

4 Resultaten

Uit ons onderzoek blijkt dat het aantal opvallende transacties⁶ in de data bij de organisaties uiteenloopt van 0,1% tot 2,5%. Over het algemeen vertegenwoordigen de opvallende transacties 0,1% tot 28,5% van de euro's in de populatie. Deze hoge bovengrens van euro's kan worden verklaard door het feit dat boekingen met grote bedragen niet vaak voorkomen en zodoende niet binnen de vaste patronen in het netwerk vallen. Opvallend is dat daarbij bij één organisatie 52,8% van de journaalposten één of

meerdere als opvallend aangeduide regels bevat. Dit houdt in dat de opvallende regels zijn opgenomen in relatief kleine journaalposten en zodoende een klein deel van de boekingen een groot percentage van de journaalposten representeert.

De computertijd die de software nodig had om tot resultaten te komen, was conform de onder 'Onderzoeksmethode' gedefinieerde verwachting.

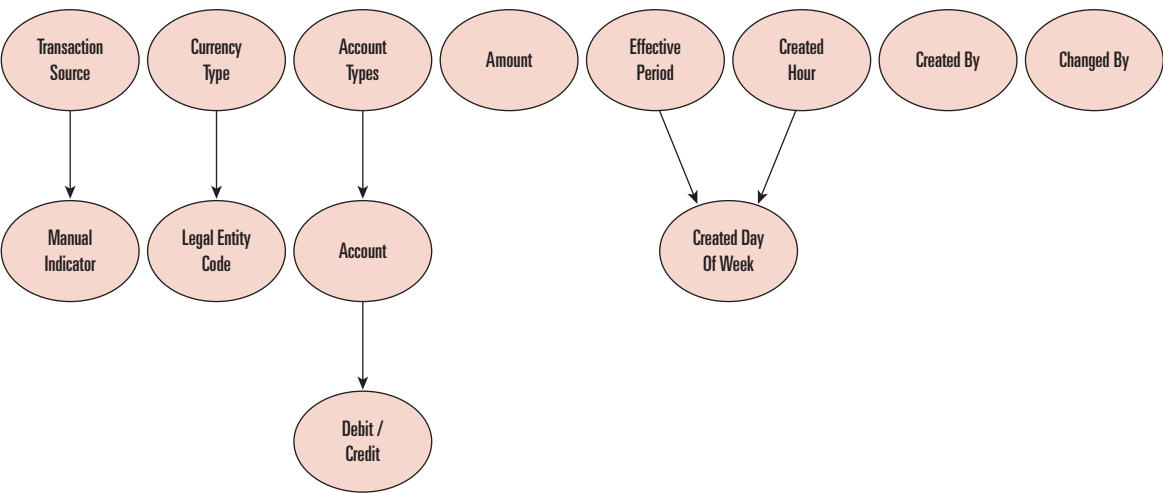
De software brengt de verschillende relaties die in de data bestaan in kaart via een BN, zoals uitgelegd in paragraaf 2. De software vindt daarbij patronen waarvan de organisaties en wij verwachtten dat zij aanwezig zouden zijn in het grootboek, zoals een verband tussen account en accounttypes. Deze verwachte patronen geven een indicatie dat de door de software gebouwde netwerken via een valide methode worden geconstrueerd. Daarnaast geeft het netwerk ook onverwachte relaties weer, zoals een verband tussen debet/credit en de dag waarop een boeking heeft plaatsgevonden of een verband tussen de periode en de dag van de boeking. De software vond twee verschillende BN's voor de organisatie die voor ons onderzoek grootboekdata voor twee verschillende jaren opleverde. Dit kan veroorzaakt zijn door procesmatige veranderingen.

De relaties die door de software gevonden worden, komen uit de reeds aanwezige data naar voren. Omdat deze relaties niet eerder inzichtelijk waren voor de organisatie, biedt deze manier van het representeren van de data de organisaties nieuwe inzichten in hun financiële organisatie. Een voorbeeld is dat een van de organisaties inzicht kreeg in het werknemersgedrag op

Tabel 2 Resultaten

Dag van de week	% opvallende transacties	opvallende transacties in % van journaalposten	% opvallende transacties in EUR volume
Pilot 1 (2009)	0,1	3,2	0,3
Pilot 2 (2009)	2,4	24,0	1,9
Pilot 3 (2009)	0,1	0,8	0,1
Pilot 4 (2009-2010)	2,5	52,8	28,5
Pilot 5a (2008)	0,1	1,9	0,5
Pilot 5b (2009)	1,0	21,6	18,1

Figuur 3 Een voorbeeld van een netwerk dat door de software kan worden gevonden



woensdag- en vrijdagmiddag, waar het volume van boekingen onevenredig daalde met het aantal parttimers op die dagen. Een voorbeeld van een BN is opgenomen in figuur 3.

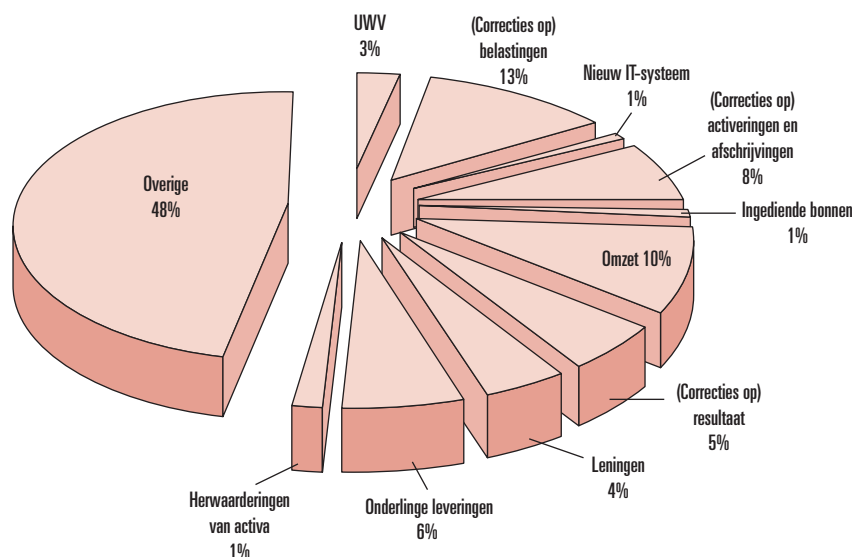
Zoals beschreven in de onderzoeksmethode brengt de software naast de verschillende relaties in de data, ook de opvallendheid van elke regel in kaart. Aan elke organisatie hebben wij een top tien van meest opvallende regels in hun grootboek teruggekoppeld. Daarnaast hebben wij ook een top tien van grootste bedragen op opvallende grootboekregels teruggekoppeld. De minder gebruikelijke handmatige journaalposten wijkten het meest af. Daarnaast komen ook fouten in systeemboekingen naar voren, zoals onvolledig of niet-tijdig verwerkte batchboekingen. In veel gevallen herkenden de betrokken organisaties de boekingen in de top tien. De betrokken organisaties hebben nader onderzoek ingesteld naar boekingen in de top tien, indien deze niet bekend waren maar volgens onze analyse en die van de organisatie, wel opvallend werden gevonden.

Gemiddeld was 90% van de meest opvallende regels een handmatige boeking, en 10% systeemboekingen. De grootste opvallende regels waren allemaal handmatige boekingen. De bedragen behorende bij de opvallende regels liggen tussen -/- 73 miljoen euro (als gevolg van zogenoemde ‘same side correcties’) en 137 miljoen euro (zie ook tabel 3). De gemiddelde opvallende regel bedraagt 225.000 euro.

Bij traditionele data-analyse is een veelvoorkomend nadeel dat er lange lijsten met uitzonderingen worden gegenereerd. Indien data-analyse in de context van de accountantscontrole wordt uitgevoerd, is aan de accountant de taak om deze uitzonderingen op te volgen. Dit kan een zeer arbeidsintensief proces zijn. Om voor de accountant de output van de software overzichtelijk te maken, zijn de opvallende regels geselecteerd die groter zijn dan de materialiteit die voor de betreffende organisatie redelijk te achten is.⁷ Zoals in tabel 3 te zien is, levert dit voor alle betrokken organisaties minder dan 500 regels aan opvallende boekingen op.

Tabel 3 Kengetallen van de populatie

	Kleinste bedrag opvallende regels (afgerond)	Grootste bedrag opvallende regels (afgerond)	Gemiddelde bedrag van opvallende regels	Aantal regels groter dan materialiteit
	In euro's	In euro's	In euro's	
Pilot 1	0,00	11.900.000,00	227.008,90	188
Pilot 2	0,00	18.950.000,00	10.604,29	95
Pilot 3	0,00	10.000.000,00	199.638,57	5
Pilot 4	-720.000,00	111.800.000,00	122.152,28	131
Pilot 5a	-20.000,00	55.290.000,00	700.055,02	27
Pilot 5b	-72.890.000,00	136.460.000,00	539.498,90	410

Figuur 4 Categorisering van de 600 meest opvallende regels

Voor elke betrokken organisatie is de top honderd van opvallendste regels gecategoriseerd in de categorieën zoals te vinden in figuur 4, aan de hand van de omschrijving. Uit deze figuur blijkt dat ongeveer de helft van de regels niet gegenereerd is uit de standaardboekingsgangen binnen processen, maar dat het wel gebruikelijk is om deze regels te boeken (zoals de boeking van het resultaat, correcties op de omzet of herwaarderings). De tweede helft van de regels (categorie 'overig') bestaat uit regels van (met name handmatige) journaalposten waarvan de oorsprong in eerste aanleg niet duidelijk is. Deze categorie komt zonder verdere voorkennis in aanmerking voor nader onderzoek.

De regels boven de materialiteitsgrens, gecombineerd met de regels in de categorie 'overig' en klantspecifieke kennis, leveren een overzichtelijke lijst met nader te onderzoeken transacties op.

Aan de hand van bekende boekingsgangen die binnen de traditionele data-analyse als mogelijk risicovol worden aangemerkt, is gecategoriseerd hoeveel opvallende regels deze boekingsgangen volgen. Deze opvallende boekingsgangen zijn te vinden in tabel 4. Uit tabel 4 blijkt dat het voorkomen van specifieke opvallende boekingsgangen varieert per betrokken organisatie. Voor de accountant zou deze informatie over boekingsgangen richtinggevend kunnen zijn voor de insteek van zijn controle.

Om meer inzicht te geven in de aard van de opvallende regels, is tabel 5 bijgevoegd. In tabel 5 is het voorkomen van bepaalde steekwoorden in opvallende regels per pilot weergegeven.

Uit de eerste testevaluaties met de betrokken organisaties bleek dat er een aantal verbeterpunten was met betrekking tot het terugkoppelen van de resultaten die de software

Tabel 4 Opvallende boekingsgangen

	Aantal journaalposten dat activa debiteert en kosten crediteert	Aantal journaalposten dat schulden debiteert en kosten crediteert	Aantal journaalposten dat schulden debiteert en omzet crediteert	Aantal journaalposten dat omzet debiteert en activa crediteert
Pilot 1	14	7	19	18
Pilot 2	5	163	3	6
Pilot 3	3	1	0	1
Pilot 4	10	2	408	1642
Pilot 5a	1	2	1	0
Pilot 5b	204	81	78	66

Tabel 5 Aantal keer dat keywords voorkomen in afgeletterde resultaten

	Pilot 1	Pilot 2	Pilot 3	Pilot 4	Pilot 5a	Pilot 5b
'fraud'	0	1	3	0	3	2
'corr'	237	13	12	837	93	1088
'restate'	0	0	0	48	0	0
'overboeking'	24	2	1	42	16	47
'reverse'	5	0	0	1086	3	13
'reclass'	0	0	0	344	50	686
'divers'	2	11	0	28	0	7
'various'	0	0	0	12	0	0
'wijziging'	3	1	0	0	0	1
'change'	0	0	0	5472	0	4

genereert. Twee van de betrokken organisaties gaven het advies om de als opvallend aangemerkte regels af te letteren alvorens deze terug te koppelen. Als de software de data analyseert, worden alle boekingen meegenomen in de resultaten. Echter, boekingen die een tegenboeking hebben, hebben per saldo geen effect op het resultaat in het grootboek. Daarom is besloten om deze boekingen tegen elkaar weg te laten vallen en alleen die regels in onze resultaten op te nemen die een impact hebben op de stromen en standen in het grootboek. Het proces van afletteren vindt achteraf plaats, dus na het berekenen van het netwerk en de kansverdelingen in de data. De in dit artikel gepresenteerde resultaten zijn na aflettering gegenereerd.

Daarnaast zijn boekingen met een bedrag van minder dan een euro vaak niet interessant voor een organisatie. Toch bleek dat de software regelmatig opvallende regels vindt met kleine bedragen. Immers, ook kleine bedragen kunnen opvallend zijn in combinatie met de waarden in andere kolommen. Aan de hand van het gesprek voorafgaand aan de analyse kan samen met de organisatie worden besloten om regels onder een bepaald bedrag te filteren uit de lijst met resultaten.

Eén van de organisaties gaf aan dat de resultaten weinig zeiden over de organisatie als geheel, omdat er binnen de organisatie veel verschillende typen boekingen plaatsvinden. Om meer inzicht te krijgen, hebben wij voorgesteld om de data te verdelen in kleinere groepen, ingedeeld volgens het rekeningschema (kostensoorten, batensoorten). Op die manier waren voor de desbetreffende organisatie uitzonderingen beter te identificeren. Ook een mogelijke datagroepering, zoals in het voorgaande voorbeeld, kan voorafgaand aan de analyse met de organisatie worden besproken. Ook voor toepassing van deze software binnen de accountantscontrole kan datagroepering uitkomst bieden (zie paragraaf 6).

Nadat deze verbeterpunten zijn doorgevoerd in ons proces zijn de resultaten bij de volgende testevaluaties anders gepresenteerd. Dit leverde meer tevreden reacties op van de betrokken organisaties dan bij de eerdere evaluaties.

5 Beschouwing en conclusie

De inzet van zelflerende software om opvallende transacties uit financiële data te detecteren, werd niet eerder onderzocht. Kennis zit verscholen in de patronen in de data zelf. Op grond van de voornoemde patronen kunnen opvallende transacties worden geïdentificeerd.

De door ons onderzochte software kan de traditionele methoden van data-analyse niet vervangen. Met traditionele data-analysetechnieken wordt data op vooraf vastgestelde risico-indicatoren geanalyseerd. Het blijft zinnig om op bekende risico-indicatoren data te analyseren om inzicht te krijgen in de bedrijfsrisico's die worden gelopen. Het is namelijk mogelijk dat met betrekking tot bepaalde bedrijfsrisico's de data juist een gevestigd patroon volgt. Dit betekent dat de voornoemde data bij analyse met zelflerende software niet in de lijst met opvallende regels opgenomen zullen zijn. Echter, voor een compleet beeld is de door ons onderzochte software een belangrijke aanvulling op de traditionele methoden, omdat de gehele populatie vanuit de data zelf wordt geanalyseerd en daarmee de 'blinde vlekken' in de traditionele methoden worden ondervangen. Traditionele methoden analyseren namelijk niet op alle bestaande risico's, maar alleen op de vooraf aangegeven risico's.

De kosten voor het gebruik van de software zijn min of meer gelijk aan de kosten voor het gebruik van traditionele data-analyse, omdat gebruikgemaakt wordt van dezelfde data. Indien deze software en traditionele data-

analyse tegelijkertijd gebruikt worden, kan een groot efficiencyvoordeel behaald worden, omdat de ingelezen data voor twee doeleinden kan worden gebruikt. Het draaien van deze software of data-analyseprogramma's zelf kost met name computertijd (zie ook paragraaf 3.3 en 4). Door (traditionele) data-analyse te combineren met deze software, kan in potentie minder tijd besteed worden aan werkzaamheden in de interim-controle, terwijl tegelijkertijd meer zekerheid wordt behaald. Daarvoor zal de software echter wel verder ontwikkeld moeten worden (zie hiervoor paragraaf 6).

De opbrengst van de combinatie van traditionele data-analysemethoden en unsupervised zelflerende data mining-technieken is een totaalcontrole op de financiële kolom. Het inzetten van unsupervised data mining-technieken vormt hiermee een aanvulling op het pallet van oplossingen dat ingezet kan worden in het kader van het voortdurend controleren en monitoren van de bedrijfsprocessen. De concrete toepassing van deze software op audits dient nader onderzocht te worden. Dit is beschreven in paragraaf 6.

Het doel van dit onderzoek was om te testen of zelflerende software geschikt is om op effectieve en efficiënte wijze opvallende transacties te detecteren. Ons onderzoek toont dit aan. Door de software op de grootboeken van vijf organisaties te testen, is gebleken dat de software in staat is om boekingen te selecteren naar aanleiding waarvan de betrokken organisaties nader onderzoek hebben verricht (effectiviteit). Zoals hiervoor beschreven staan de kosten voor de inzet van deze software in verhouding tot de kosten van traditionele data-analyse en bieden de resultaten van de software toegevoegde waarde ten opzichte van de resultaten van voornoemde data-analyse (efficiency). Ook de geïdentificeerde patronen zijn voor de betrokken organisaties een interessante bron van informatie gebleken met betrekking tot hun bedrijfsprocessen. Een kanttekening hierbij is dat indien processen binnen een organisatie niet op eenduidige wijze verlopen, de software daardoor mogelijk geen of minder geschikte patronen zal kunnen identificeren.

6 Toekomstig onderzoek

Op dit moment is de software zo geprogrammeerd dat individuele regels vergeleken worden met de patronen die binnen de dataset aanwezig zijn. Het zou echter interessant zijn om te onderzoeken of het mogelijk is de software aanvullend zo te programmeren dat volgtijdelijke relaties tussen regels gevonden kunnen worden. Voor dit doel hebben wij tijdinformatie nodig in datasets over welke transacties elkaar opvolgen in de tijd. Daarmee

kunnen ook volgtijdelijke patronen worden onderkend en afwijkingen op deze patronen worden gedetecteerd. Een voorbeeld van een volgtijdelijk patroon is de standaardboekingsgang die binnen een kernproces van een organisatie gevolgd wordt. Afwijkingen hierop zouden met zelflerende software in kaart gebracht kunnen worden.

Ook interessant voor toekomstig onderzoek is het verkennen van het concrete verdere toepassingsbereik van zelflerende software binnen de jaarrekeningcontrole. Is deze software alleen geschikt om opvallende transacties te detecteren op het gehele grootboek? (Hoe) kan een redelijke mate van zekerheid⁸ verkregen worden met behulp van zelflerende technieken? Een combinatie van traditionele data-analysetechnieken en deze software is mogelijk op dit terrein. Hierbij valt te denken aan het inzetten van de software op stratificaties van het grootboek, zoals binnen kernprocessen. Door binnen kernprocessen naar opvallende boekingen te zoeken, kunnen uitzonderingen veel fijnmaziger gedetecteerd worden. De toepassing van de software binnen stratificaties van het grootboek in het kader van de jaarrekeningcontrole dient nader onderzocht te worden.

In het onderhavige onderzoek hebben wij de software getest op grootboekdata. Op basis van onze bevindingen verwachten wij dat de software ook zal werken op andere procesmatig gegenereerde datasets dan een grootboek. De software zou derhalve getest moeten worden op andere procesmatig gegenereerde datasets dan een grootboek. ■

Q. Rijnders MSc. RA studeerde Bedrijfskunde aan de Erasmus Universiteit Rotterdam, gevolgd door een RA-traject bij ABS. Zij werkt als manager bij KPMG Forensic Technology met vraagstukken op het snijvlak van data-analyse en finance. P. Özer MSc. heeft Artificial Intelligence gestudeerd en is werkzaam bij KPMG Forensic Technology. Patricks werk richt zich voornamelijk op data-analyse en hij volgt daarnaast de studie EDP Auditing aan de ABS. V.L. Blankers heeft Forensic Artificial Intelligence gestudeerd aan de UvA. Voor haar indiensttreding bij KPMG heeft ze onder meer bij het Nederlands Forensisch Instituut onderzoek gedaan naar het automatiseren van handtekeningverificatie. T.A. Eijken heeft Business Mathematics & Informatics gestudeerd, is bij KPMG Forensic Technology afgestudeerd en daarna in dienst getreden. De tijdens zijn afstudeerproject ontwikkelde software heeft mede geleid tot dit artikel.

Noten

1 Een neurale netwerk, zoals gebruikt in de kunstmatige intelligentie, is een mathematisch model dat de werking van de neuronen, zoals wij deze kennen uit het menselijk brein, imiteert.

2 Het Bayesiaanse gedachtegoed maakt het mogelijk om onbekende voorwaardelijke kansen te berekenen door gebruik te maken van bekende voorwaardelijke kansen (Russell en Norvig 2002).

3 Het veld 'gebruikersnaam' verwijst naar personen die in het grootboek boekingen maken of anderszins muteren. Indien het onderzoek zich verplaatst van een algemeen onderzoek naar een onderzoek dat zich richt op het handelen of nalaten van (een) specifiek(e) perso(o)n(en), dient het onderzoek te worden uitgevoerd met inachtneming van de daarvoor geldende

praktijkhandleiding. Op 6 oktober 2010 is door het NIVRA praktijkhandreiking 1112 'Persoonsgerichte Onderzoeken voor Accountants-Administratieconsulenten/registeraccountants' uitgebracht.

4 Oracle kent ook niet-financiële boekingen, zoals Encumbrance en Budget-boeking.

5 Een boxplot is een grafiek waarin de data versimpeld is weergegeven met behulp van een mediaan, het eerste en het derde kwartiel.

6 'Opvallende' transacties zoals beschreven in paragraaf 3.3.

7 Omwille van geheimhouding geven wij hier niet de gehanteerde bedragen en berekening van de materialiteit weer.

8 Op grond van NV COS 200.17 blijkt dat de accountant een redelijke mate van zekerheid

dient te verkrijgen dat de financiële overzichten als geheel geen afwijkingen van materieel belang als gevolg van fraude of het maken van fouten bevatten. De laatste zin van COS 200.17 luidt: 'De redelijke mate van zekerheid heeft betrekking op het gehele controleproces.' Uit deze zin volgt dat ook de inzet van zelflerende software binnen de accountantscontrole zal moeten bijdragen aan het verkrijgen van een redelijke mate van zekerheid. In dit onderzoek hebben wij aan dit aspect van de inzet van de software binnen de accountantscontrole nog geen aandacht besteed. Derhalve wordt dit genoemd als onderwerp dat in acht genomen zou moeten worden bij toekomstig onderzoek.

Literatuur

■ Aldrich, J. (1997), R.A. Fisher and the making of maximum likelihood 1912-1922, *Statistical Science*, vol. 12, no. 3, pp 162-176.

■ Haas, M. de (2003), *Opinie: De accountant moet 'digitaal'*, De Accountant, maart, p. 58.

■ Heckerman, D. (1996), *A tutorial on learning with bayesian networks*, Technical report, Microsoft Research.

■ Koopmans, L. (2005), Real time science fiction?, *De Accountant*, februari, p. 22.

■ Mashaie, M. (2008), XBRL een katalysator voor continuous auditing?, *Maandblad voor Accountancy en Bedrijfseconomie*, vol. 82, no. 7/8, pp. 324-333.

■ Russell, S.J. en P. Norvig (2002), *Artificial intelligence: A modern approach*, 2nd Edition, Upper Saddle River, New Jersey: Prentice-Hall.

■ Schouten, R.P. (2007), De toegevoegde waarde van audit software, *Maandblad voor Accountancy en Bedrijfseconomie*, vol. 81, no. 11, pp. 521-530.

■ Strikwerda, J. (2008), De plaats en rol van de accountant in de 21e eeuw, *Maandblad voor Accountancy en Bedrijfseconomie*, vol. 82, no. 3, pp. 77-88.

■ Tsamardinos, I., L.E. Brown en G.F. Aliferis (2006), The max-min hill-climbing Bayesian network structure learning algorithm, *Machine*

Learning, vol. 65, no. 1, pp 31-78.

■ Upton, G. en I. Cook (1996), *Understanding statistics*, Oxford University Press.

■ Veenstra R.H. (2005), *De accountant beoordeeld*, Deventer: Kluwer.

■ Veenstra, R.H. en A. Heertje (2006), Doeltreffende administratieve controle door steekproeven, *Maandblad voor Accountancy en Bedrijfseconomie*, vol. 80, no. 12, pp. 611-619.

■ Witten, I.H. en E. Frank (2005), *Data-mining: Practical machine learning tools and techniques*, 2nd edition, San Francisco: Morgan Kaufmann.