

Process mining is vertrekpunt voor verandering

DE SLIMME MIJNWERKER

We dorsten naar informatie, maar verdrinken in een zee van data. De tendens naar informatisering, outsourcing en selfservicing heeft er zelfs voor gezorgd dat we onze processen niet meer begrijpen. Een nieuwe vacature komt dan ook vrij: Mijnwerker (M/V). Hij graaft niet met een pikhouweel in donkere mijnschachten, maar met geavanceerde tools in bedrijfsdata. Wij deden veldwerk in diverse sectoren, en groeven naar de voor- en nadelen van het nieuwe goud: 'Process Mining'.

Door Jonas Buyle, Gert Linthout en Bram Vanschoenwinkel

Een organisatie is zo doeltreffend als haar processen, daarover heerst een algemene consensus. Procesgeoriënteerde organisaties slagen er namelijk in om weg te stappen van het silodenken en wendbaar in te spelen op steeds veranderende klantenbehoeften en marktomstandigheden. Process management heeft dan ook een ware revolutie doorgemaakt: waar de focus initieel lag op modellering en standaardisatie, ligt deze nu op koppeling met kwantitatieve principes (Lean, Six sigma), architectuur (SOA) en ondersteunende systemen (BPMS).

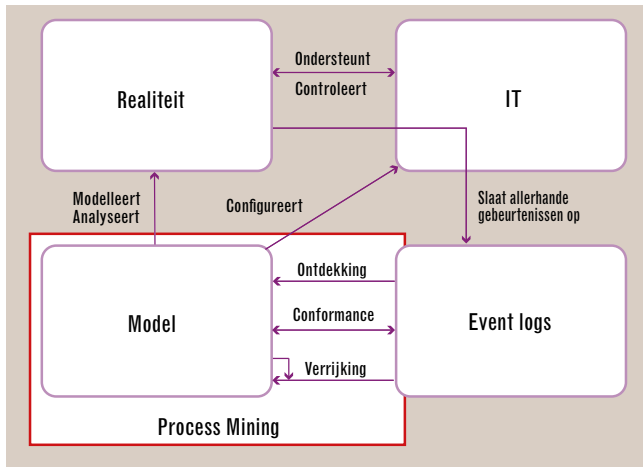
Het zijn manieren om het procesdenken daadwerkelijk te vertalen naar bedrijfsvoering en daarop continu te sturen. Het proces vormt de ruggengraat, maar blijkt steeds vaker ook de achilleshiel te zijn! We meten op zoveel punten dat we niet meer vinden wat we zoeken, zelfs onze eigen processen niet. Ook al werden ze minutieus uitgetekend in de beginfase, zijn ze na talrijke veranderingen niet meer te herken-

nen of blijkt de praktijk niet overeen te stemmen met het oorspronkelijke model.

De symptomen zijn herkenbaar: sociale secretariaten *outsourcen* een deel van het intake proces, maar spenderen de gewonnen tijd aan *rework*. Selfservice ratio's in contactcenters blijven onder de verwachtingen. In ziekenhuizen blijkt het moeilijk om de – nochtans gedetailleerde – patiëntgegevens te koppelen aan klinische verzorgingspaden. Dat een brede schare medewerkers of zelfs de klant de actoren zijn in het proces, bemoeilijkt de beheersing ervan. En wat doe je als je je proces niet beheerst? Graven!

Graven met stijl

Door de ver doorgedreven IT-ondersteuning van processen slaan de gebruikte systemen veel gegevens op over het reële proces in zogenaamde *event logs*. Dit wordt schematisch voorgesteld in afbeelding 1. Bij de webingave van een loondossier bijvoorbeeld wordt elke toetsaanslag, elk scherm en elke keuze opgeslagen. Dergelijke logs bevatten veel nuttige infor-



Afbeelding 1: Situering van process mining. Bron, zie [1].

matie over het daadwerkelijk procesverloop, maar zijn vaak moeilijk te interpreteren en enorm omvangrijk. Process mining is een techniek die toelaat om die miljoenen lijnen aan beschikbare gegevens om te zetten naar interpreteerbare informatie. Aan de hand van een model worden de event logs gebruikt om de realiteit te reconstrueren. Het is een specifieke toepassing van technieken uit de meer algemene discipline van datamining. De verbanden en patronen die je hiermee blootlegt slaan op het werkelijke procesverloop. Het is logischerwijs alleen relevant voor systeemondersteunde bedrijfsprocessen. Het toepassingsgebied gaat van de bank- en verzekeringssector naar telecom en utilities tot manufacturing en assembly. Al naar gelang de context, doe je dit om één van de volgende doelstellingen te bereiken:

- Traditioneel werd deze techniek ontwikkeld om de stappen in het achterliggende proces te ontdekken en uit te tekenen. In een manufacturingomgeving bijvoorbeeld is het reële procesverloop vaak complex en moeilijk te achterhalen. Een onderdeel ondergaat daar een aantal vooraf bepaalde bewerkingsstappen. Omdat voor elke stap verschillende machines in aanmerking komen, groeit het aantal mogelijkheden exponentieel met het aantal stappen;
- Een conformance-analyse legt de verschillen bloot tussen het bedoelde procesmodel en de realiteit en traceert waar het fout loopt. Een telefonisch keuzemenu is bijvoorbeeld ontworpen om klanten naar de agent met specifieke vaardigheden te routeren. Wanneer de klant verward raakt door een onduidelijke verwoording van de verschillende mogelijkheden, volgt hij in de praktijk een andere weg in het keuzemenu en komt hij bij de verkeerde agent;
- Het gekende proces wordt verrijkt met data uit de event logs om bottlenecks te identificeren. Wanneer klanten bijvoorbeeld zelf loongegevens invullen in hun dossiers, blijkt een aantal schermen specifiek te leiden tot incorrecte combinaties. De initiële doelstelling, met name werklastverlaging bij de medewerkers, heeft in dat geval meer *rework* in een latere fase tot gevolg.

In wat volgt gaan we dieper in op de toepassing van de techniek zelf, aangevuld met de bevindingen uit twee reële projecten: *Team A vs. Team B* en *De klant kiest altijd juist* (zie de kaders). Naast de specifieke voordelen vestigen we de aandacht op een aantal vaak voorkomende valkuilen.

Klik en klaar?

Bottlenecks, doorlooptijden en de (on)zin van verschillende stappen in het proces krijg je niet met een druk op de knop. Het vergt heel wat kennis en voorbereiding om er aan te beginnen. Process mining steunt op de aanwezige data en is dus sterk afhankelijk van de kwaliteit van die data. Van meet af aan moet het doel van de analyse duidelijk vastliggen en moet de analist de beschikbare data goed begrijpen. Om het in de termen van de mijnbouw te formuleren: we hebben een land van honderd hectare aangekocht om te ontginnen, maar in het wilde weg gaan graven naar het kostbare goud maakt de kans op succes erg klein. Het graven begint dus pas wanneer we het doel helder voor ogen hebben.

In het manufacturingproject *Team A vs. Team B* was het doel om het outputvolume te verhogen via best practices, door verschillen in performance tussen diverse ploegen te begrijpen. Wanneer een klant inbelt via het keuzemenu wilden we in het project *De klant kiest altijd juist* begrijpen waarom de klant andere keuzes maakt dan voorzien, en dit gebruiken om de werking te verbeteren.

Globaal gezien verloopt dit in drie stappen: in de eerste stap wordt het pad geëffend en moeten de beschikbare data klaar-gemaakt worden zodat ze ontgonnen kunnen worden. In de tweede stap kan het echte werk, het feitelijke ontginningsproces aanvangen. Finaal dienen we de output correct te interpreteren en te vertalen in verbeteracties, waarbij we vaak beroep doen op complementaire technieken.

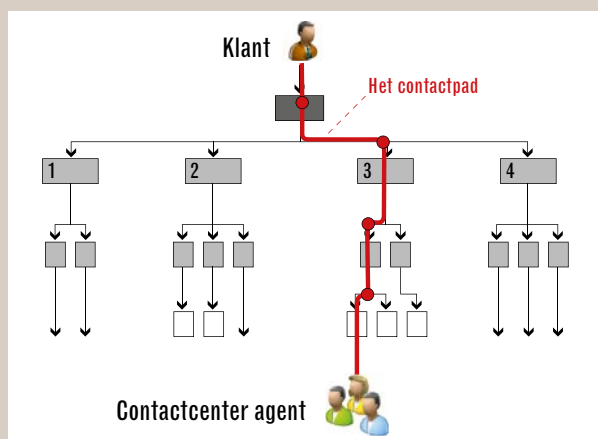
Stap 1: het pad effenen

Het klaarmaken van de data heeft twee doelstellingen. In de eerste plaats zal het de kwaliteit van de data verhogen door het verwijderen van ruis. In een manufacturingomgeving bijvoorbeeld wordt de productie-installatie elke maand onderworpen aan een reeks testen waarbij de goede werking van de verschillende machines wordt gecontroleerd. Het verloop van dergelijke tests vertoont sterk afwijkend gedrag van het normale verloop van het productieproces en daarom is het belangrijk dat we de gegevens die tijdens deze tests worden opgeslagen verwijderen uit de event logs. Daarnaast is het belangrijk om het formaat van de data af te stemmen op de tool die gebruikt zal worden bij de ontginning. XML of CSV zijn veelvoorkomende formaten die hiervoor aangewend worden.

Stap 2: het echte werk

Net zoals in de mijnbouw zijn de technieken die je zult gebruiken voor de ontginning ook verschillend afhankelijk van de ondergrond waarin je moet graven. Bij process mining vertaalt dit zich in een aantal keuzes. Om te beginnen moet de analist

De klant kiest altijd juist



In een contactcenter omgeving worden klanten voortdurend gevraagd om keuzes te maken: Wat is uw voorkeurtaal? Over welk onderwerp heeft u een vraag? De klant legt zo een contactpad af door het keuzemenu heen, gebaseerd op de sequentie van zijn keuzes (het 'proces'). Dit heeft een dubbel doel: zorgen dat de klant terecht komt bij een agent die de kennis heeft om de vraag te beantwoorden; de klant leiden naar de selfservice mogelijkheden in het keuzemenu (bijvoorbeeld een detail van uw factuur opvragen).

Wanneer de structuur en de bewoording van het keuzemenu niet duidelijk zijn, vindt de klant de juiste weg niet meer. Dit leidt tot frustratie bij de klant en oplopende kosten voor het bedrijf. Klanten vinden de kostbare selfservice opties niet, of ze moeten doorverbonden worden omdat ze bij de verkeerde agent terechtkomen. Deze problemen lopen in een groot contactcenter op tot miljoenen euro per jaar. We zetten process mining in om te ontdekken waar het misloopt in het keuzemenu. Vaak zijn er immers honderden contactpaden mogelijk. Via process mining leggen we de meeste frequente keuze-paden bloot, en ontdekken we ongewenste sequenties of keuzes.

kieszen welk specifiek process mining algoritme hij gaat gebruiken. Dit hangt nauw samen met het doel dat we willen bereiken, zoals eerder vermeld. Naast de vraag die beantwoord moet worden hangt de keuze ook af van de inhoud van de event logs. Wat niet kan teruggevonden worden in de data kan ook niet onderzocht worden.

Ongeacht de keuze van de techniek, gaat elk algoritme uit van hetzelfde principe: het is onmogelijk om alle gegevens uit de event logs uitputtend te onderzoeken. Daarom zal het algoritme de zoekruimte op een 'slimme' manier beperken en een

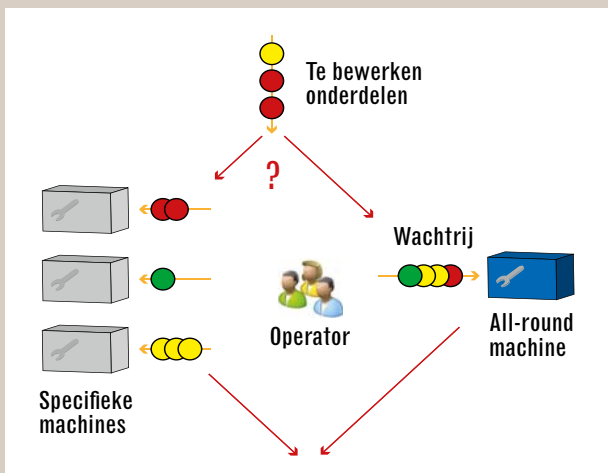
formeel (wiskundig) model opbouwen dat in staat is om een veralgemening te maken van de onderliggende patronen in de data. Dit betekent een afweging tussen de complexiteit van het model en de precisie. Een te complex model zal een zeer precieze weergave van de data bevatten maar is niet in staat om een veralgemening te maken van de onderliggende patronen in de data en leert ons dus weinig. Dit probleem staat bekend als 'overfitting'. Een te eenvoudig model daarentegen laat te veel informatie verloren gaan waardoor we een te gesimplificeerd beeld krijgen van de werkelijkheid, ook wel 'underfitting' genoemd. Beschouw in dat opzicht het project rond de keuze van de klant. Een model dat alle mogelijke contactpaden bevat is even complex als het oorspronkelijke probleem. Dit leert ons weinig over de meest frequent gevolgde paden die tot de verkeerde agent leiden. Een model dat daarentegen slechts één enkel (bijvoorbeeld het meest frequente) contactpad weergeeft is te eenvoudig, omdat er waarschijnlijk nog enkele andere frequente paden zijn die minstens even interessant zijn om te onderzoeken. De analist is echter wel geïnteresseerd in het model dat er in slaagt om de tien meest frequente paden, die consequent naar de verkeerde agent leiden, bloot te leggen. Maar hoe bepaal je de juiste balans?

Om het zoekproces en de opbouw van het model te sturen, wordt elk process mining algoritme gekenmerkt door een set van parameters die eigen zijn aan het algoritme in kwestie. De juiste waarde voor deze parameters bepalen is niet eenvoudig omdat de verschillende parameters elkaar kunnen beïnvloeden en kleine verschillen in numerieke waarde een grote invloed kunnen hebben op het resultaat. Dit probleem staat bekend als parameterselectie, en vergt voor elke techniek de nodige ervaring en kennis. Bovendien neemt het aantal mogelijke combinaties exponentieel toe in relatie tot het aantal parameters van het algoritme. Het is de taak van de analist om meerdere combinaties uit te proberen en zo tot de ideale configuratie te komen. Vaak zal het onmogelijk zijn om alle parametercombinaties te onderzoeken, maar er bestaan verschillende methoden en technieken die de analist hierbij kunnen ondersteunen.

Stap 3: output interpreteren

Als eenmaal het goud opgegraven is, gaat de smid aan de slag om er een mooie ring van te maken. De uitkomst van het process mining algoritme kan namelijk verschillende vormen aannemen en vergt verdere interpretatie en verwerking om tot een bruikbaar resultaat te komen. Voor een juiste interpretatie hebben we minstens twee elementen nodig: de technische expertise en de kennis van de omgeving waarin het proces plaatsvindt. Vaak blijkt zelfs dit niet voldoende: het is niet eenvoudig om uit de 'ontgonnen' procesmodellen duidelijke conclusies te trekken, of de onderliggende oorzaak af te leiden. Waarom kiest de operator machine x voor de bewerking, terwijl machine y beter lijkt? Waarom kiest de klant in het keuzemenu de optie 'factuur' terwijl hij een vraag heeft over zijn internetaansluiting?

Team A vs. Team B



In manufacturing- en assemblageomgevingen zijn routingproblemen gangbaar. Onderdelen worden in een sequentie van stappen omgezet van ruw materiaal naar eindproduct. Het proces is hier een opeenvolging van machinebewerkingen, waarbij eenzelfde bewerking vaak kan worden uitgevoerd op verschillende machines.

In het voorbeeld lijkt de problematiek eenvoudig: als een onderdeel arriveert (drie types), op welke machine moet dit bewerkt worden? Elk onderdeel heeft twee mogelijkheden: de allround machine (die alle onderdelen kan bewerken), of één van de drie onderdeelspecifieke machines.

Dit bleek niet eenvoudig voor de operatoren, omdat de lengte van de wachtrijen, de sequentie van de onderdelen en de omsteltijden van de allround machine zeer bepalend zijn voor het outputvolume en de doorlooptijd. Er werden dan ook sterke volumeschommelingen tussen de verschillende ploegen vastgesteld van 40 tot 50 procent, en dit voor de bottleneck van de fabriek. We zetten process mining in om na te gaan welke keuzes leiden tot een verhoogd of verlaagd outputvolume en wat de optimale onderdeelsequentie is.

Bij een telecom provider konden we in *De klant kiest altijd juist* via process mining de frequente beslissingspaden blootleggen, naast de meest voorkomende ongewenste keuzes en keuzesequenties. In een tweede fase deden we een aantal meeluistersessies in het contactcenter, om de inhoudelijke problematiek beter te begrijpen. Daarna legden we een aantal scenario's voor aan een steekproef van klanten, aan wie we vroegen om deze uit te testen. Via real-time observatie stelden we vast *waarom* de klant bepaalde keuzes maakt. Dit bracht

naar voren dat de bewoording en de volgorde van de keuzemogelijkheden bepalend zijn voor een doeltreffende werking. Met deze verhelderende conclusies uit de studie slaagden we erin om het keuzemenu gericht aan te passen, zodat het aantal transfers tussen agenten en de selfservice ratio significant (meer dan 10 procent) verbeterden. Dit zorgt voor een jaarlijkse *return* van 1,5 tot 2,5 miljoen euro.

In het geval van het manufacturing project konden we via process mining vaststellen welke keuzes en onderdeelsequenties leidden tot een verlaagd outputvolume. De *good practices* konden we pas afleiden nadat we een aantal focusgroepen organiseerden met de verschillende operatorploegen en we zelf gedurende een bepaalde periode deelnamen aan de activiteit. Dit leidde tot een verhoging van de totale performance met 30 procent.

In bepaalde gevallen is het zelfs zinvol om zich letterlijk in de plaats van de klant te stellen, om op die manier inzicht te krijgen in de reële proceswerking. Deze *mystery* onderzoeken laten toe om vast te stellen hoe het proces in de praktijk verloopt en te begrijpen waarom dit zo is. Dit levert uiterst zinvolle informatie op die toelaat om de blootgelegde procesmodellen beter te begrijpen, niet vanuit het standpunt van de klant maar als klant zelf.

Conclusie

Door de wildgroei aan ERP-, CRM-, WMS- en andere ondersteunende systemen hebben we toegang tot een zee van informatie. Het menselijk brein schiet tekort om de hoeveelheid te bevatten. Process mining is ontstaan uit noodzaak om iets met deze informatie aan te vangen. De laatste jaren is er zeer veel vooruitgang geboekt in het onderzoek naar deze techniek en is de gebruiksvriendelijkheid van de toepassing ervan sterk verbeterd. Het blijft echter wel belangrijk te beseffen dat de magische doos die met één simpele druk op de knop alle processen tevoorschijn tovert nog niet bestaat.

Bovendien is het resultaat van deze oefening slechts nuttig als het gezien wordt in het licht van process management. Process mining mag geen standalone-analyse zijn, maar het vertrekpunt voor verandering van de ontdekte processen of pijnpunten. Omdat kwalitatieve technieken dieper ingaan op het 'waarom' achter de vastgestelde symptomen, is de combinatie van beide de sleutel tot succes.

Noot

1. R.S. Mans, M.H. Schonenberg, M. Song, W.M.P. van der Aalst & P.J.M. Bakker. Application of Process Mining in Healthcare – A Case Study in a Dutch Hospital, A. Fred, J. Filipe, and H. Gamboa (Eds.): BIOSTEC 2008, CCIS 25, pp. 425–438, 2008.

Jonas Buyle, Gert Linthout en Bram Vanschoenwinkel

Jonas Buyle (jonas.buyle@mobius.eu), Gert Linthout (gert.linthout@mobius.eu) en Bram Vanschoenwinkel (bram.vanschoenwinkel@mobius.eu) zijn consultant bij MÖBIUS Business Redesign.