

Process mining is een relatief nieuwe methode om bedrijfsprocessen accuraat in kaart te brengen. Modellen worden automatisch gegenereerd op basis van daadwerkelijke gebeurtenissen, zoals die zijn geregistreerd in informatiesystemen. Het vakgebied is genoemd naar procesmodellering en data mining, en is interessant voor wetenschap én commercie. De TU Eindhoven loopt voorop. Door Karina Meerman



Wanneer organisaties hun processen in kaart brengen, maken zij modellen op basis van de vermeende – vaak ook gewenste – werkelijkheid. Medewerkers worden geïnterviewd, bestaande modellen aangepast. Een dergelijke aanpak brengt niet de uitzonderingen in beeld en resulteert in modellen die niet de realiteit vertegenwoordigen. 'Bedrijven zien meestal wat ze willen zien, niet wat ze moeten zien,' zegt prof. dr. ir. Wil van der Aalst, hoogleraar aan de Technische Universiteit Eindhoven en grondlegger van process mining. Een PowerPoint-werkelijkheid, noemt hij het ook wel.

Andere werkelijkheid

Een voorbeeld uit de praktijk: een grote overheidsinstantie wilde haar bedrijfsprocessen verbeteren. Een probleem was het op tijd betalen van facturen. Volgens het bestaande procesmodel moesten facturen voor betaling altijd voor akkoord worden voorgelegd aan de aannemer. De medewerkers deden dat per e-mail. Uit onderzoek bleek dat hier de bottleneck zat. Aannemers zijn in werkelijkheid altijd op pad en zitten zelden



Wil van der Aalst: 'De tijd is rijp voor process mining.'



Anne Rozinat: 'Geweldig dat het mogelijk is om processen automatisch naar boven te halen.'

achter hun computer. Dat was de reden dat het verkrijgen van akkoord, en dus de betalingen, weken vertraging opliepen. Nu bellen de overheidsmedewerkers met de aannemers en worden facturen wél op tijd betaald.

Universitair docent dr. ir. Boudewijn van Dongen geeft nog een voorbeeld: 'In het procesmodel van de Nederlandse Vereniging van Gemeentes ligt vast dat een ID-document pas mag worden geprint wanneer het is betaald. Als je het proces gaat herontwerpen op basis van dat model, heb je een probleem. Want in werkelijkheid rekent de klant meestal pas af nádat hij zijn paspoort of rijbewijs heeft gekregen.'

Business Intelligence (BI) is ook niet de oplossing. Deze commerciële software werkt niet met event logs, maar met van tevoren gedefinieerde 'key performance indicators' (KPI's) en andere meetpunten. Van der Aalst zou BI zelfs onintelligent willen noemen. 'Vaak gaat het om vrij triviale analyses van gestructureerde data.' Wetenschappelijk programmeur dr. ir. Eric Verbeek vult aan: 'KPI's doen niets met het onderliggende proces. Ze geven doorlooptijden aan. Als je van tevoren weet waar de beslispunten zitten kan je hoogstens meten dat 40 procent linksom gaat en 60 procent rechtsom. Maar met process mining komen die beslispunten vanzelf wel naar boven.' Dat betekent een betere – want realistische – weergave van je werkelijke bedrijfsprocessen.

Algoritmen

Algoritmen die veel bij process mining worden gebruikt zijn alpha miner, fuzzy miner, multiphase, heuristics, transition system miners en genetische algoritmen. Die laatste werken met een evolutionair proces, geïnspireerd door concepten uit de natuur zoals natuurlijke selectie, reproductie, mutatie en recombinatie. Kortom, kindjes maken, die kindjes maken, die weer kindjes maken enzovoort. Uit de bovenste 20 procent van events genereert het algoritme een nieuw model en gebruikt daarvan dan weer de bovenste 20 procent voor nieuwe modellen. Verbeek: 'Het doel is om een procesmodel te vinden dat het log het beste beschrijft, dat de hoogste graad van fitness heeft.'

Exponentiële datagroei

Process mining brengt dus de reële processen in beeld, doordat het gebruik maakt van echte gebeurtenissen die zijn geregistreerd in de eventlogs van informatiesystemen. Alles wat een mens doet laat tenslotte sporen achter. En bestelling bij een webwinkel, het opladen en gebruiken van de OV-chipknip, het boeken van een reis, maar ook de behandeling van een patiënt in een ziekenhuis. Elke handeling, elke 'transactie', wordt vastgelegd, ook de uitzonderingen en ook de overdrachtsmomenten waar geen mensen aan te pas komen, zoals de interactie tussen websites. Aan data geen gebrek dus. 'De hoeveelheid data groeit exponentieel,' zegt Van der Aalst. 'Vroeger ging je naar een reisbureau en kreeg je een papieren ticket. Het aantal overdrachtsmomenten was beperkt. Wanneer je nu een ticket online boekt, gaat die site contact aan met meerdere luchtvaartmaatschappijen en met een betaalsysteem. Ook die gebeurtenissen worden vastgelegd.'

Hoe anders was dat nog toen Van der Aalst zo'n tien jaar geleden begon met process mining. 'Ik bedacht dat we datamining en procesmodelleren wel konden koppelen, omdat ik vond dat modellen zo slecht de werkelijkheid weergaven. Alleen waren de gegevens nog nauwelijks aanwezig; ziekenhuizen werkten bijvoorbeeld nog met papieren briefjes.' Nu is de tijd rijp voor process mining. 'De digitale wereld zit heel

dicht op de echte wereld. De introductie van diagnose-behandelcombinaties in ziekenhuizen – waardoor vergoedingen alleen worden verstrekt op basis van geregistreerde gegevens – heeft een explosie aan data teweeggebracht. En als je bijvoorbeeld een boek bestelt bij Bol.com dan maakt het niet uit of het boek echt op voorraad is – dat je het op een plank ziet staan – want als het informatiesysteem zegt dat het er is, dan is het er ook.' Inmiddels heeft Van der Aalst een team van twintig postdocs en promovendi om zich heen verzameld.

Drie technieken

Process mining kent drie verschillende categorieën van technieken: Discovery, Conformance en Extension. Discovery is de meest magische van de drie: de technieken leiden automatisch modellen af van eventlogs. Een beetje te magisch, soms.

Van Dongen: 'Consultants in process mining merken dat klanten niet geloven dat modellen automatisch gegenereerd kunnen worden.' Niet de analist bepaalt van tevoren wat de werkelijkheid is, maar een algoritme, achteraf.

Conformance-technieken vergelijken een bestaand model met een eventlog, voor analyse van verschillen of afwijkingen in patronen. Zeer waardevolle technieken voor bijvoorbeeld financiële organisaties, die te maken hebben met steeds striktere compliance-eisen (Sox, Tabaksblatt). Van der Aalst: 'Bedrijven zijn gevoelig voor grotere efficiëntie en de maatschappij wil meer zekerheid over wat daadwerkelijk gebeurt.'

De derde categorie is extension. Met deze technieken kan een bestaand model worden verrijkt met nieuwe data. Stel dat discovery leidt tot de verschillende stappen in het proces, dan kunnen extensionstechnieken toevoegen wie de gebeurtenissen uitvoert. Van der Aalst: 'Dan wordt voorspellen mogelijk op basis van gedrag in het verleden.'

Opensource framework

Om al die data te kunnen modelleren ontwikkelde de TU/e een opensource framework, ProM. Dat ondersteunt twee formats van eventlogs, MXML (Mining XML) en XES (eXtensible Event Stream). 'XES is generieker en heeft daarom de voorkeur,' aldus Verbeek. Omdat de meeste informatiesystemen hun informatie opslaan in leverancierspecifieke databases, audit trails en transactielogs, ontwikkelden Van der Aalst en zijn team ook ProMimport. Hiermee kunnen de logs geladen worden van systemen als FLOWer, Staffware, Apache, CPN Tools, CVS, Eastman, PeopleSoft, Subversion, WebSphere en Adept. ProM wordt onder internationaal samenwerkende onderzoeksgroepen verspreid via processmining.org. Als wetenschapper gefinancierd door overheidsgeld, vindt Van der Aalst het niet meer dan logisch dat hij met open source werkt. 'Het gaat er niet om dat je als bedrijf patent op een stukje software hebt; het gaat om de ideeën.'



Boudewijn van Dongen: 'Klanten geloven vaak niet dat modellen automatisch kunnen worden gegenereerd.'



Eric Verbeek: 'Business Intelligence doet niets met het onderliggende proces, process mining wel.'

ProM ondersteunt tientallen algoritmes die resulteren in een model dat, afhankelijk van de hoeveelheid events en complexiteit van de processen, meer of minder visueel begrijpelijk is. Voor academische doeleinden kunnen doorrekeningen weken in beslag nemen. Voor het bedrijfsleven is dat ondenkbaar. Het vakgebied kent daarom inmiddels twee afsplitsingen die naast het ProM framework commerciële software ontwikkelen: de bedrijfjes Futura and Fluxicon.

Toegankelijk maken

Fluxicon ontving in 2010 een Valorisation Grant van STW. Dr. Anne Rozinat is software engineer en runt het bedrijf samen met dr. Christian W. Günther. 'We formuleerden een goede visie op process mining en hoe we het toegankelijk willen maken,' geeft zij als reden voor het ontvangen van de financiële prikkel. Het extra geld geeft hen ruimte om dingen uit te zoeken en in feite meer tijd te kopen. 'Process mining is een belangrijk nieuw vakgebied dat enorm mooi en nuttig kan zijn, wanneer meer tools ontwikkeld zijn. Wij proberen process mining toegankelijk te maken voor mensen buiten de universiteit. Daar zijn grenzen aan, want we zitten op een academische plek, maar ik denk niet dat de kloof erg groot is tussen het bedrijfsleven en de academische wereld.' Zelf raakte ze gefascineerd door process mining tijdens haar studie aan de universiteit in Potsdam. 'Dat het mogelijk is om automatisch processen naar boven te halen: dat vond ik geweldig.'

Voor een academische tool is ProM heel succesvol, zegt Rozinat. 'Het wordt honderden keren per maand gedownload. Voor het bedrijfsleven moet er echter nog wel wat veranderen. Bedrijven willen statistieken, zij willen weten hoe lang een activiteit duurt, wie eraan werkt. Dat zijn wetenschappelijk totaal oninteressante gegevens, maar nodig om van processmining-software een commercieel succes te maken.'

Een andere hobbel voor bedrijven is inmiddels uit de weg geruimd. De vertaling van data naar een ProM-vriendelijk format kostte soms één of twee dagen, bleek uit Rozinats ervaring in advieswerk. Zo wierp de eerste stap van de analyse meteen een wel erg hoge drempel op. Met hun nieuw ontwikkelde conversie-gereedschap Nitro, dat data geautomatiseerd omzet van CSV of Excel naar MXML of XES, wordt dat nu gereduceerd tot enkele minuten. Uiteindelijk wil Fluxicon toe naar software die iedereen zonder achtergrond in process mining kan gebruiken.

Navigeren door processen

De favoriete metafoor van Van der Aalst voor de toegevoegde waarde van de processmining-techniek is die van het autonavigatiesysteem op basis van GPS. Traditionele procesmodellen zijn als landkaarten: ze zijn een representatie van de werkelijkheid maar niet de werkelijkheid zelf. Een dun lijntje is een weggetje, een dikkere lijn een snelweg; de verhouding klopt niet echt maar toch begrijpen mensen de informatie. Een model gebaseerd op eventlogs is als een navigatiesysteem. Het is niet alleen een goede landkaart, maar kent ook alternatieve routes, de gemiddelde reistijd en de actuele drukte op de weg. 'En als je TomTom zegt dat je naar rechts moet, kun je nog steeds linksaf. Die flexibiliteit, die lerende functie, is belangrijk.'

Process mining: the standard tool of the future

Process Mining combines the technologies of data mining and process modelling using event logs as input. It is a relatively new method of mapping business processes and one that is much more accurate than anything created by business intelligence. The technique uses actual events as documented in information systems to visualise processes, whereas more traditional techniques use supposed processes or existing models. The Eindhoven University of Technology is a global leader in process mining. Prof.dr.ir. Wil van der Aalst laid the foundation for this technique ten years ago, in response to the inaccuracies of the 'PowerPoint reality' that many organisations live in. Spin-off company Fluxicon is working on commercial software and received an STW Valorisation Grant in 2010 to support its efforts. Van der Aalst and team expect process mining to become a standard tool for organisations that wish to streamline even the most complex processes.

Standaard gereedschap

Process mining heeft de toekomst. De analisten van Gartner Research zien het als een belangrijke innovatie, en gaven in 2009 de titel 'Cool Vendor' aan IT-leverancier Pallas Athena. Deze partner van de TU/e heeft al enkele jaren een commercieel pakket voor process mining op de markt en timmert ook internationaal flink aan de weg. Mede door de samenwerking met de wetenschap kan het zijn product steeds verder verbeteren.

'Enerzijds hebben bedrijven een enorme behoefte om processen efficiënter te laten verlopen en zeker te stellen dat wordt gewerkt volgens bepaalde regels', verklaart Van der Aalst. 'Anderzijds gaat de dataexplosie gewoon verder en komen steeds meer gedetailleerde eventlogs beschikbaar.' Daarom is het volgens de hoogleraar evident dat over tien jaar veel organisaties gebruik zullen maken van process mining. 'Het wordt een gangbaar gereedschap voor auditors en consultants. Zeker op het moment dat deze functionaliteit standaard ingebouwd is in allerlei informatiesystemen. In tegenstelling tot eerdere technieken voor procesanalyse gaat het nu om échte data; dat maakt een wereld van verschil voor organisaties.' I/O