



Universiteit
Leiden
The Netherlands

Unraveling temporal processes using probabilistic graphical models

de Paula Bueno, M.L.

Citation

De Paula Bueno, M. L. (2020, February 11). *Unraveling temporal processes using probabilistic graphical models*. *SIKS Dissertation Series*. Retrieved from <https://hdl.handle.net/1887/85168>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/85168>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/85168> holds various files of this Leiden University dissertation.

Author: De Paula Bueno, M.L.

Title: Unraveling temporal processes using probabilistic graphical models

Issue Date: 2020-02-11

SAMENVATTING

Temporele processen beschrijven gebeurtenissen in het dagelijks leven, zoals lopen, slapen en eten, etc., maar ook meer gecompliceerde situaties zoals medische behandelingen, de jaargetijden, gebeurtenissen in een workflow, etc. We zijn gewoonlijk geïnteresseerd om te begrijpen hoe bepaalde aspecten van objecten zich ontwikkelen, zoals de ziektesymptomen van een patiënt. Aan de ene kant hebben we *expressieve* modellen nodig om complex gedrag vast te leggen. Aan de andere kant moeten dergelijke modellen *compacte* beschrijvingen van processen opleveren om inzicht te verwerven. Het vinden van de juiste balans tussen deze twee kenmerken van modellen is niet triviaal, want door het verhogen van de expressiviteit komt men vaak tot complexere modellen, die wellicht minder interpreteerbaar zijn. Vaak is het ook nog nodig dat geschikte beschrijvingen van processen expliciet rekening houden met onzekerheid.

In dit proefschrift willen we de expressiviteit van het model vergroten, geïnspireerd door de complexiteit van reële problemen, met behoud van de interpreteerbaarheid van het model. Daartoe beschrijven we drie nieuwe verschillende gezichtspunten op processen op basis van probabilistisch grafische modellen.

We geven eerst een nieuw procesperspectief op basis van *latente toestanden*, die kunnen worden gezien als abstracte representaties van de waarneembare data. Latente toestanden kunnen helpen bij de interpretatie, omdat ze fungeren als een instrument voor *dimensionaliteitvermindering*.

In hoofdstuk 3 introduceren we asymmetrische hidden Markov modellen voor het vastleggen van de lokale structuur tussen de waarneembare variabelen. Dit wordt gedaan door elke latente toestand te associëren met een Bayesiaans netwerk. Asymmetrische hidden Markov-modellen leiden vaak tot een betere kwaliteit van modellen en meer inzicht in het domein, terwijl de noodzaak van het kiezen van een a priori modelarchitectuur wordt verminderd. Een empirische evaluatie werd uitgevoerd met behulp van gesimuleerde en echte datasets.

In hoofdstuk 4 stellen we een semi-automatisch raamwerk voor om ziekteprocessen te begrijpen op basis van de dynamiek van latente toestanden binnen de hidden Markov modellen. We hebben het raamwerk toegepast op gegevens die verkregen zijn bij de behandeling van patiënten met psychotische depressie, waarbij latente toestanden als patiëntgroepen fungeren, die symptomen voorspellen als onderdeel van de prognose van de patiënt.

In hoofdstuk 5 leren we hidden Markov modellen uit gegevens van gebeurtenissen in de gezondheidszorg. Er wordt gebruik gemaakt van een casus op basis van gebeurtenissen die te maken hebben met aderverkalking. De grootte van dergelijke datasets, in vergelijking met het kleine aantal mogelijke gebeurtenissen in de datasets, maakt dat dezelfde gebeurtenis wordt geassocieerd met meerdere latente toestanden, een begrip dat we clustering van latente toestanden noe-

men. We laten zien dat gebeurtenissen in een cluster geassocieerd worden met patiënten met een verschillende ernst van de ziekte.

Het tweede gezichtspunt op processen is gebaseerd op de identificatie van procesveranderingpunten of regimeverandering. De uitdaging ligt in het uitbreiden van modellen die tijdsinvariant zijn (zoals dynamische Bayesiaanse netwerken) voor het vastleggen van regimeverandering op een eenvoudige manier, die geschikt kan zijn wanneer de beschikbare dataset klein is.

In hoofdstuk 6 stellen we gepartitioneerde dynamische Bayesiaanse netwerken voor om modellen te kunnen bouwen waarvoor de tijdshomogeniteitsaanname niet geschikt is. Gepartitioneerde dynamische Bayesiaanse netwerken zijn een verzameling van dynamische Bayesiaanse netwerken waarbij afkappunten heuristisch worden geïdentificeerd. Deze modellen werden geëvalueerd in een brede verzameling experimenten.

In het laatste procesperspectief wordt getracht deelverzamelingen van tijdsgegevens te ontdekken die samenhangen met modellen die substantieel afwijken van het model dat uit de hele dataset wordt verkregen. Dit kan worden gezien als het identificeren van belangrijke subprocessen.

In hoofdstuk 7 introduceren we dynamische Bayesiaanse netwerken voor het representeren van uitzonderlijke temporele modellen uit deelverzamelingen van de data. Dit biedt een algemene representatie voor subprocessen binnen de context van de ontdekking van subgroepen en *exceptional model mining*. We evalueren de voorgestelde aanpak door middel van gesimuleerde data en een casus rond subsidieaanvragen in de agrarische sector.