

University of Groningen

Dynamic clustering

Ernst, Anja Franziska

DOI:
[10.33612/diss.196176258](https://doi.org/10.33612/diss.196176258)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2022

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Ernst, A. F. (2022). *Dynamic clustering: classifying people through ecological momentary assessment*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen.
<https://doi.org/10.33612/diss.196176258>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Samenvatting

Vandaag de dag stellen smartphones en tablets onderzoekers in staat om door middel van zogenaamde ecological momentary assesment tijdsintensieve data te verzamelen. Bij een ecological momentary assesment worden mensen gevraagd om meerdere keren per dag hun ervaringen, zoals bijvoorbeeld emoties, gedurende een aantal weken, vast te leggen. Dit levert intensieve longitudinale data op. De beschikbaarheid van dergelijke data heeft binnen de sociale en medische wetenschappen en de onderwijswetenschappen een verschuiving teweeggebracht in de richting van het bestuderen van de processen die zich in een individu afspelen. Binnen de sociale wetenschappen heeft de transitie van longitudinale data met slechts enkele tijdstippen per individu naar intensieve longitudinale data een verandering teweeggebracht in hoe longitudinale modellen worden gebruikt. De focus van de analyse is hierbij verlegd van het beschrijven van groei en afname naar het beschrijven van de dynamiek van variabelen die door de tijd heen fluctueren. Dit stelt onderzoekers in staat om nieuwe hypothesen te ontwikkelen en te toetsen om zo fenomenen zoals depressie, angst en blijdschap beter te begrijpen.

Sociale wetenschappers zijn in het bijzonder geïnteresseerd in de dynamiek van de interactie tussen de emoties (bijv. een blij gemoed), de klinische symptomen (bijv. zwaarmoedigheid) en de trans-diagnostische factoren (bijv. rumineren) van een individu door de tijd heen. Deze dynamiek reflecteert, bijvoorbeeld, de neiging tot het voortduren van emoties, of de verhoging van het niveau van intensiteit van emoties. De laatste jaren heeft de sociale wetenschap veel vooruitgang geboekt door het bestuderen van deze dynamische processen, in het bijzonder betreffende de ontwikkeling van emoties en psychopathologie door de tijd heen.

Vaak willen onderzoekers op een data-gedreven manier subgroepen bestaande uit individuen die structureel andere kenmerken laten zien in kaart brengen. Deze subgroepen zijn a priori niet bekend en worden daarom in kaart gebracht door middel van exploratieve clustering. Bij clustering draait het erom om een heterogene groep onder te verdelen in homogene subgroepen (Lubke & Muthén, 2005). Het doel van clustering is dus om de verschillende subgroepen, of clusters, binnen de dataset te identificeren, en om de verschillende individuen onder te brengen binnen deze clusters. Op deze manier houdt clustering rekening met interindividuele verschillen waarvan de oorzaken onduidelijk of niet geobserveerd zijn.

Het doel van dit proefschrift is om nieuwe dynamische clusteringsprocedures te ontwikkelen. De procedures moeten rekening houden met kwalitatieve interindividuele verschillen in intra-individuele dynamiek. We doen dit door het ontwikkelen van clusteringsprocedures die clusters van individuen met kwalitatief verschillende dynamische processen in hun longitudinale data zichtbaar kunnen maken. Zodoende kunnen subgroepen worden geïdentificeerd, waarvan de individuen die tot dezelfde subgroep behoren vergelijkbare dynamiek laten zien. Dynamische clustering stelt ons daarom in staat om de informatie van verschillende individuen te combineren, waarbij men rekening houdt met de kwalitatieve heterogeniteit tussen individuen in de onderliggende dynamiek.

Door het identificeren van dergelijke dynamische clusters zijn onderzoekers bijvoorbeeld in staat om resultaten naar de populatie te generaliseren ondanks dat individuen kwalitatief verschillende dynamische processen laten zien (Voelkle e.a., 2014). Het samenvoegen van de gegevens van verschillend individuen, terwijl men rekening houdt met de essentiële verschillen tussen deze individuen, heeft een aantal voordelen. Een aantal van de voordelen is dat overfitting wordt voorkomen (Bulteel e.a., 2018), het onderscheidend vermogen van het model toeneemt (Mansueto e.a., 2020), en dat individuen die structureel van elkaar verschillen, toch worden beschreven met hetzelfde statistische model (zie bijv. Hamaker e.a., 2016).

Om de dynamische interactie tussen variabelen in intensieve longitudinale data te beschrijven maken onze dynamische clusteringsprocedures gebruik van vector-autoregressieve

(VAR) modellen (Lütkepohl, 2005). De coëfficiënten van een VAR model van een individu kunnen de dynamiek van emoties beschrijven. Onze dynamische clusteringsprocedures zullen het VAR model uitbreiden, zodat het kan omgaan met clusters van individuen.

De tot dusver bekende statische methoden die zijn voorgesteld voor dynamische clustering, nemen aan dat de onderliggende genererende processen voor individuen van een bepaald cluster gelijk zijn. Om deze zeer restrictieve aanname te vermijden stellen wij in hoofdstuk 2 een clusteringsprocedure voor die gebaseerd is op een mixture model (McLachlan & Peel, 2004). Onze dynamische clusteringsprocedure bestaat uit twee stappen. In stap 1 wordt de dynamiek van elk individu gevangen in een apart VAR model, vervolgens worden in stap 2 de VAR coëfficiënten van elk individu geclusterd door middel van een mixture model. Door gebruik te maken van een mixture model voor het clusteren van de VAR coëfficiënten stelt deze dynamische clustering ons in staat om rekening te houden met de continue heterogeniteit in de VAR coëfficiënten binnen de clusters. In een simulatiestudie vergelijken wij de prestaties onze dynamische clusteringsprocedure met de prestaties van een dynamische clusteringsprocedure die geen rekening houdt met de heterogeniteit in de VAR coëfficiënten binnen de clusters (zoals voorgesteld door Bulteel e.a., 2016). We laten zien dat wanneer er variatie is in de VAR coëfficiënten binnen clusters, onze dynamische clusteringsprocedure betere prestaties levert dan de clusteringsprocedure die geen rekening houdt met de heterogeniteit binnen de clusters. Beide procedures worden geïllustreerd door middel van een ecological momentary assessment studie die de emotionele dynamiek van 366 individuen onderzoekt. We beschrijven de resulterende clusters in termen van hun emotionele dynamiek en hun scores op maatstaven van depressie en angst.

Om dynamische clusters op een accurate wijze te beschrijven kan het nuttig zijn om de twee stappen — d.w.z. het beschrijven van de individuele dynamiek en het clusteren — niet onafhankelijk uit te voeren, maar deze in plaats daarvan te combineren in een enkele schattingsstap. In hoofdstuk 3 presenteren wij een dynamische clusteringsprocedure, waarbij schatting op een dergelijke adaptieve manier plaatsvindt. Omdat de beschrijving van de intra-individuele dynamiek zich aanpast tijdens de schatting van de clusters, wordt voortijdige compressie van de longitudinale data voorkomen. Deze adaptieve schattingsprocedure maakt een clusterspecifieke beschrijving van de intra-individuele dynamiek mogelijk, waarbij de verschillende clusters kunnen worden beschreven door kwalitatief verschillende longitudinale modellen, bijvoorbeeld VAR modellen met verschillende aantallen lags. Wij stellen een adaptief algoritme voor om een nieuwe dynamische clusteringsprocedure te schatten. Dit algoritme is beschikbaar gesteld aan empirische onderzoekers door middel van een R functie. De prestaties van de voorgestelde schattingstechniek zijn beoordeeld doormiddel van een simulatiestudie. De dynamische clusteringsprocedure is geïllustreerd door deze toe te passen op ecological momentary assessment data waarbij de emoties van 410 individuen een maand zijn gemeten. We beschrijven de gemiddelde emotionele dynamiek van de resulterende clusters en de gemiddelde scores van elk cluster op externe maatstaven van depressie en angst.

De dynamische clusteringsprocedure die is beschreven in hoofdstuk 3, houdt geen rekening met mogelijke heterogeniteit in VAR coëfficiënten binnen de clusters. Om deze beperking op te heffen introduceren wij in hoofdstuk 4 een dynamische clusteringsprocedure die de procedure uit Chapter 3 uitbreidt door ook variatie in de VAR coëfficiënten binnen de clusters te schatten. Zodoende is deze dynamische clusteringsprocedure in staat om zowel de kwalitatieve interindividuele heterogeniteit tussen de clusters, als de kwantitatieve (dat wil zeggen, continue) interindividuele heterogeniteit binnen de clusters, in kaart te brengen. Een aantal belangrijke zaken aangaande de modelspecificatie en de schatting van het model komen ter sprake. De voorgestelde dynamische clusteringsprocedure is ter illustratie toegepast op ecological momentary assessment data, verkregen uit een zeer heterogene steekproef,

bestaande uit jongvolwassenen en ouderen. We beschrijven de gemiddelde emotionele dynamiek in de geïdentificeerde clusters en tonen aan dat deze clusters grotendeels overeenkomen met de reeds bekende emotionele dynamieken die worden beschreven voor jongvolwassenen en ouderen. De gevonden afwijkingen van dit algemeen bekende patroon zijn interessant vanuit inhoudelijk oogpunt.

Het is voor empirische onderzoekers vaak lastig om verschillende statistische modellen te vergelijken. Dit is wel nodig om een goede keus te kunnen maken uit de veelheid aan beschikbare modellen. Er bestaat namelijk een grote verscheidenheid aan statistische modellen, afkomstig uit verschillende disciplines en gemaakt voor zeer ongelijksoortige datastructuren. In hoofdstuk 5 specificeren wij een modelraamwerk, waarbinnen wij de overeenkomsten benadrukken tussen de verschillende longitudinale modellen die kunnen worden gebruikt om op enige wijze rekening te houden met interindividuele verschillen. Dit overkoepelende raamwerk omspannt zowel de dynamische clusteringsprocedures die zijn besproken in hoofdstuk 3 en hoofdstuk 4, als een aantal van de meest prominente longitudinale modellen die worden gebruikt in de sociale en medische wetenschappen en de onderwijswetenschappen. Zodoende blijkt dat de dynamische clusteringprocedures uit hoofdstuk 3 en hoofdstuk 4 via het gespecificeerde modelraamwerk verbonden zijn met andere longitudinale modellen. Wij bespreken de verschillende aspecten van ons modelraamwerk, door een verscheidenheid aan longitudinale modellen als voorbeelden te gebruiken. Bijvoorbeeld, wij brengen verschillende aspecten in kaart van de longitudinale data, waarin ons modelraamwerk in staat is interindividuele verschillen te onderscheiden, zoals de dynamiek, groei, afname, en cyclische trends. De overeenkomsten en verschillen tussen de verscheidene modellen worden besproken en wij tonen aan dat er een zeer diverse set modellen kan worden uitgedrukt binnen ons modelraamwerk. We doen tenslotte enkele aanbevelingen aan empirische onderzoekers betreffende modelselectie en modelspecificatie.

In hoofdstuk 6 vergelijken wij de verschillende dynamische clusteringsprocedures die in deze dissertatie ter sprake zijn gekomen. We evalueren de verschillende schattingsprocedures voor elk van de dynamische clusteringsprocedures die wij hebben voorgesteld. Ook bespreken wij hoe het toepassen van dynamische clustering kan leiden tot verbeteringen binnen het psychologisch onderzoek. Tenslotte bespreken wij de mogelijkheden voor verder onderzoek.

