

Tensor Time Series Clustering

De Dataset

AMIE: Automatic Monitoring of Indoor Exercises

- Metingen van lichaamsgewichtoefeningen
 - Squats
 - Forward lunges
 - Side lunges
- **Tijdsreeksen** als datapunten: Posities bij de uitvoering gemeten doorheen de tijd



Puur illustratief
voorbeeld van een
tijdsreeks

Datapunten vergelijken

- **Dynamic Time Warping (DTW)**: berekent afstand tussen 2 tijdsreeksen
 - Gelijkaardige data krijgen een kleine afstand
 - $O(n^2)$ tijd
- **Afstandstensor**: elk element is een DTW afstand tussen 2 datapunten
→ Elementen berekenen kostelijk!

Benadering van een Tensor

Candecomp/Parafac (CP) decompositie

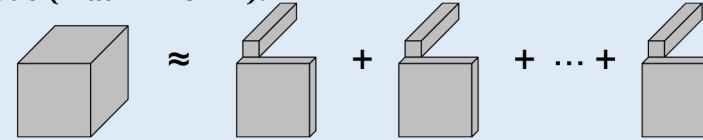
- Som van uitwendige producten van vectoren
- Alternating Least Squares algoritme
- De volledige tensor nodig => Kostelijk

Adaptive Cross Approximation for Tensors (ACA-T) decompositie

- Som van uitwendige producten van matrices/vectoren
- **Adaptief** = Gericht vectoren in Tensor kiezen
- 'Klein' deel van de tensor nodig

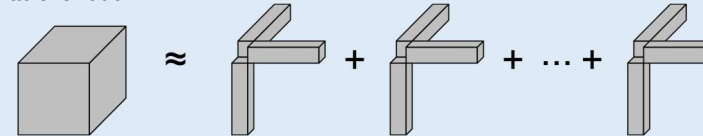
Matrix Methode (Matrix ACA-T):

- Kostelijk

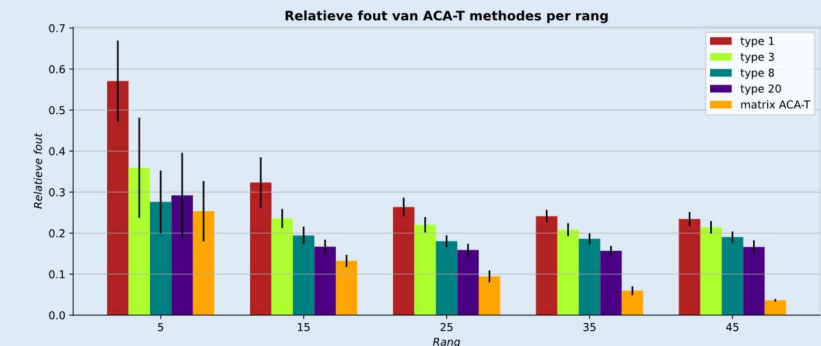


Vectoren Methode (Vector ACA-T type 1):

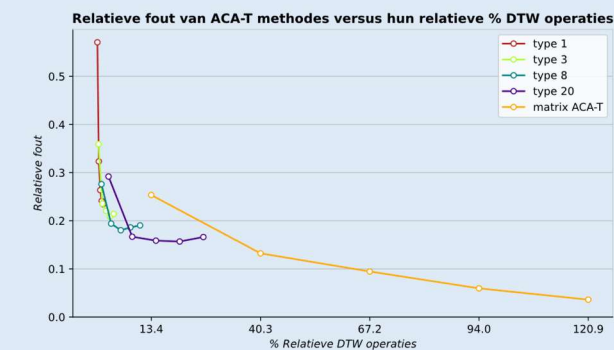
- Hogere relatieve fout



Resultaten: Tensor benadering



Bovenstaande figuur toont de relatieve fout van enkele types van de uitgebreide vectoren methode per rang. De bar is het gemiddelde en de zwarte streep de standaard afwijking ($n = 50$). De rang van een decompositie is gelijk aan het aantal termen. **Opmerkelijk:** Hogere types liggen dicht bij de matrix methode bij lage rang, en dicht bij de vectoren methode (type 1) bij hoge rang.



De y-as bevat opnieuw het gemiddelde van de relatieve fout zoals in bovenstaande figuur (zonder standaard afwijking)

Op de x-as staat het percentage DTW operaties nodig om deze relatieve fout te verkrijgen. 100% betekent dat we de volledige tensor hebben berekend.

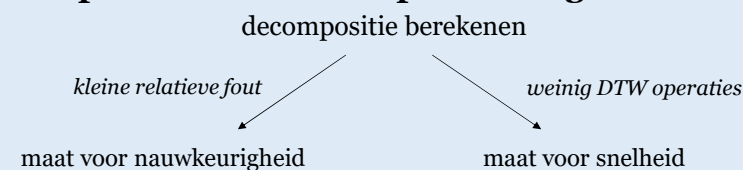
We willen een methode die in de linker onderhoek ligt.

Doel van het Onderzoek

Tijdsreeksen clusteren

- 1) Tensor decompositie berekenen
 - *CP decompositie: benadert tensor zeer goed*
 - *Adaptieve methodes: steunen op het ACA algoritme*
- 2) Rijen/kolommen/tubes uit decompositie als feature vectoren kiezen
 - *Dimensie bepaald wat we clusteren*
 - *Bij AMIE-dataset: clusteren op sensoren of personen*
- 3) K-means clustering algoritme gebruiken met deze feature vectoren

DTW operaties van decompositie laag houden



Adaptive Cross Approximation (ACA) uitbreiden

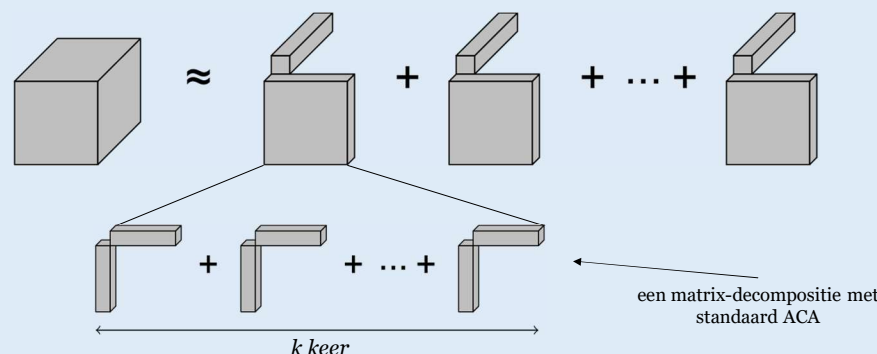
- Decompositie algoritme voor matrices
- **Adaptief:** Focust op grootste en dus belangrijkste elementen
- **Cross:** Iteratief een rij en kolom aan de decompositie toevoegen
- Kleinere elementen niet nodig
→ *Ideaal voor een decompositie van de afstandstensor*
→ *ACA methodes voor tensoren ontwikkelen*

Ons Onderzoek

Hypothese: De uitgebreide vectoren methode zal voor een kleinere relatieve fout zorgen maar het aantal DTW operaties laag houden.

De uitgebreide vectoren methode

- Starten vanuit de matrix methode
- De matrix in elke term opnieuw benaderen met ACA
→ notatie: *type k* = er zijn k termen in elke matrix decompositie
- elke term bevat:
 - **Tube:** een vector mode-3
 - **Matrix-decompositie:** som van termen met 2 vectoren mode-1 en 2

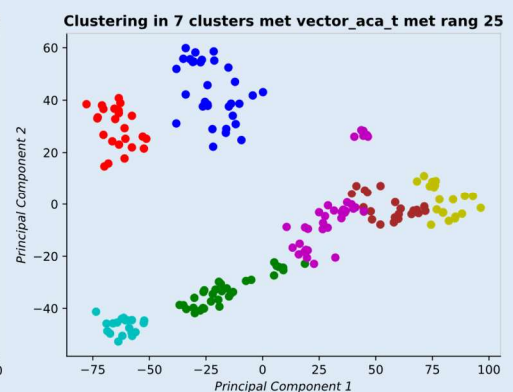
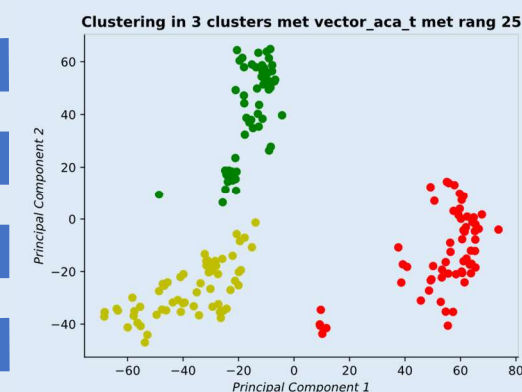


Resultaten: Clusteren

Person	Exercise	Cluster
Person8	squat	1
Person8	squat	1
Person8	squat	1
Person8	squat	1
Person8	squat	1
Person8	squat	1
Person8	squat	1
Person8	lunge	0
Person8	lunge	0
Person8	lunge	0
Person8	lunge	0
Person8	lunge	0
...
Person5	sidelunge	2

Person	Exercise	Cluster
Person8	squat	2
Person8	squat	2
Person8	squat	4
Person8	squat	4
Person8	squat	4
Person8	squat	4
Person8	squat	4
Person8	lunge	0
Person8	lunge	0
Person8	lunge	0
Person8	lunge	0
Person8	lunge	0
...
Person5	sidelunge	1

Resultaten clusteren ($k = 3$ en $k=7$ resp.) met uitgebreide vectoren methode rang 25 type 3.



Visualisatie van de resulterende clusters door middel van *Principal Component Analysis* om de dimensie te reduceren, waarbij we de rijen als feature-vectoren nemen.