

Literatuurstudie

Wetenschappelijke vorming

Wannes Croes & Lowie Debois

Faculteit Wetenschappen
KU Leuven
Informatica

May 4, 2024



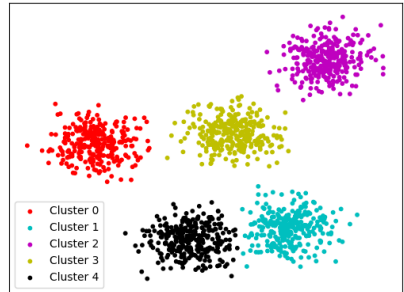
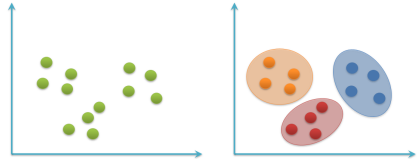
Bronnen

- ▶ **Masterthesis:** T. Vanhoof, Adaptieve tensor factorisaties om versneld tijdreeksen te clusteren, 2023
- ▶ **Dataset:** T. Decroos, K. Schutte, T. Beéck, B. Vanwanseele, and J. Davis. AMIE: Automatic Monitoring of Indoor Exercises: European Conference, ECML PKDD 2018, Dublin, Ireland, September 10-14, 2018, Proceedings, Part III, pages 424–439. 01 2019.

Tensor time series clustering

Clusteren

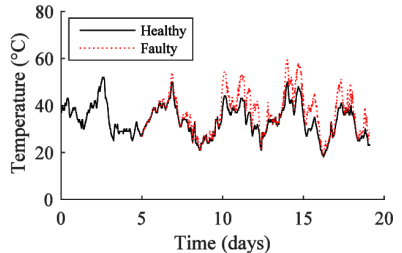
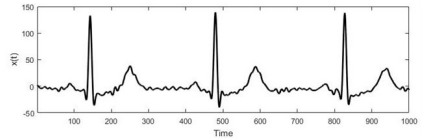
- ▶ groeperen van data
- ▶ automatische feedback
- ▶ anomaliedetectie



Type data

Time Series = informatie van 1 parameter doorheen de tijd

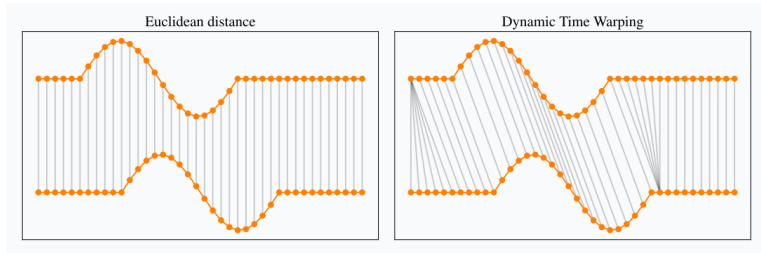
- ▶ temperatuur
- ▶ aandelen (de beurs)
- ▶ elektrocardiogram (hartslagmeter)
- ▶ positie (2D, 3D...)



Vergelijken van time series

Euclidische afstand

- ▶ one-to-one
- ▶ snel ($O(n)$)
- ▶ niet gewenste resultaat



Dynamic Time Warping (DTW)

- ▶ many-to-many (optimaal)
- ▶ kostelijk ($O(n^2)$)
- ▶ gewenste resultaat

Afstandstensenoren

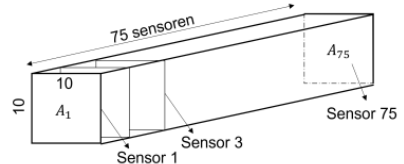
Matrix (Tensor orde 2)

- ▶ afstanden als elementen
- ▶ paarsgewijs vergelijken
- ▶ symmetrische matrix
- ▶ computationeel inefficiënt

$$\begin{bmatrix} d(T_1, T_1) = 0 & d(T_1, T_2) & \dots & d(T_1, T_{n-1}) & d(T_1, T_n) \\ d(T_2, T_1) & 0 & \dots & d(T_2, T_{n-1}) & d(T_2, T_n) \\ \dots & \dots & \dots & \dots & \dots \\ d(T_{n-1}, T_1) & d(T_{n-1}, T_2) & \dots & 0 & d(T_{n-1}, T_n) \\ d(T_n, T_1) & d(T_n, T_2) & \dots & d(T_n, T_{n-1}) & 0 \end{bmatrix}$$

Tensor orde 3

- ▶ afstanden als elementen
- ▶ data groeperen per dimensie
- ▶ symmetrische slices
- ▶ minder elementen → efficiënt
- ▶ bevat meer informatie



Clusteren met afstandstensenoren

Informatie in tensor

- ▶ gelijkaardige data \rightarrow afstand klein
- ▶ afstand klein \rightarrow klein element in tensor
- ▶ veel gelijkaardige elementen \rightarrow lage rang (zoals bij matrices)

Conclusie: gelijkaardige data \rightarrow lage rang tensor.

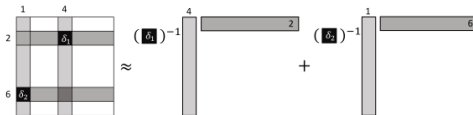
Takeaway: details in afstandstensor overbodig.

Decompositie: benaderen van een lage rang matrix

Adaptive Cross Approximation (ACA)

- ▶ apart nemen 'grote' rijen/kolommen
- ▶ som van vectorproducten
- ▶ rang = aantal termen
- ▶ grootte van decompositie « grootte van matrix

Takeaway: meer termen \rightarrow betere decompositie \rightarrow berekende matrix lijkt op volledige matrix



Decompositie: benaderen van een lage rang tensor

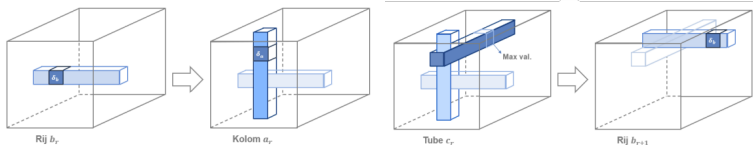
Adaptive Cross Approximation for Tensors (ACA-T)

- ▶ apart nemen 'grote' fibers
- ▶ som van outerproducten
- ▶ rang = aantal termen

Takeaway: meer termen \rightarrow betere decompositie

\rightarrow berekende tensor lijkt op afstandstensor

\rightarrow meer DTW operaties



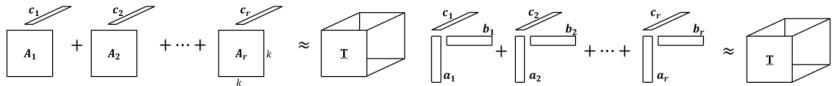
Adaptive Cross Approximation for Tensors

Matrix decompositie:

- 1 $\sum \text{rij} \otimes \text{kolom}$
- 2 that's it

Tensor decompositie:

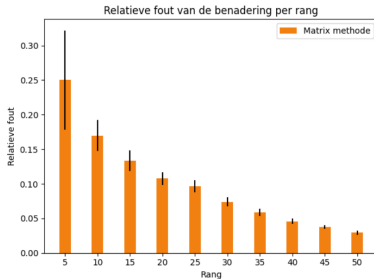
- 1 $\sum \text{matrix} \otimes \text{vector}$
- 2 $\sum \text{rij} \otimes \text{kolom} \otimes \text{tube}$
- 3 ...



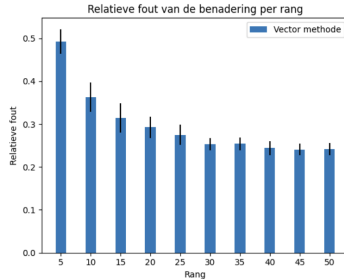
Reeds uitgevoerde experimenten

Resultaten thesis: verband relatieve fout met rang

- ▶ methode 1: rel. fout ≈ 0.03 (rang 50)
- ▶ methode 2: rel. fout ≈ 0.24 (rang 30)



(A) Matrix methode



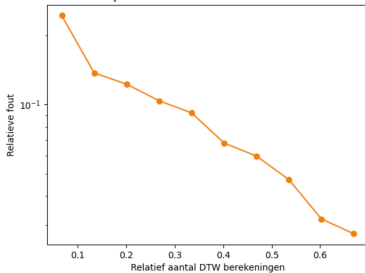
(B) Vectoren methode

Reeds uitgevoerde experimenten

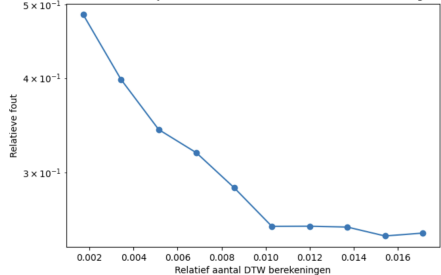
Resultaten thesis: verband relatieve fout met DTW berekeningen

- ▶ methode 1: tot 60% van aantal DTW berekeningen
- ▶ methode 2: relatief weinig berekeningen (1.6%)

Relatieve fout bij een toenemend relatief aantal DTW berekeningen



Relatieve fout bij een toenemend relatief aantal DTW berekeningen



Reeds uitgevoerde experimenten

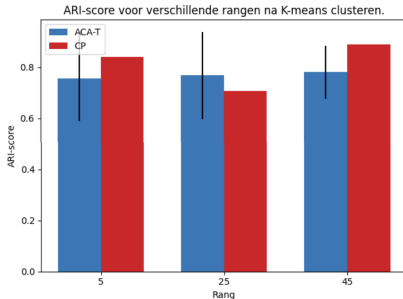
Hoe zit het met clusteren? (*de eigenlijke essentie*)

Evaluatie via: Adjusted Rand Index (ARI)

- ▶ maat voor correcte classificatie $[-1, 1]$

Grafiek

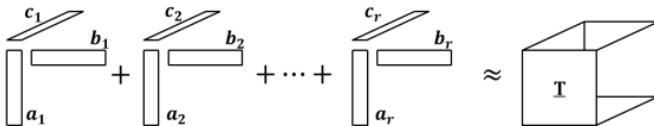
- ▶ 1 dimensie gekozen
- ▶ 3 clusters
- ▶ verschillende rangen
- ▶ ACA-T: ARI-score ≥ 0.75



Hypothese

Onderzoeksmogelijkheden: andere methodes uittesten

- ▶ AMIE dataset gericht samplen
- ▶ methode toepassen op similarity tensor



Bronnen

Afbeeldingen

- ▶ <https://medium.com/@mygreatlearning/clustering-algorithms-d7b3ae040a95>
- ▶ <https://medium.com/analytics-vidhya/k-means-clustering-optimizing-cost-function-mathematically-1ccae156299f>
- ▶ https://www.researchgate.net/publication/308864495_Comparison_of_different_modelling_approaches_of_drive_train_temperature_for_the_purposes_of_wind_turbine_failure_detection
- ▶ P. Čepulionis and K. Lukoševičiūtė, “Electrocardiogram time series forecasting and optimization using ant colony optimization algorithm,” *Mathematical Models in Engineering*, Vol. 2, No. 1, pp. 69–77, Jun. 2016.
- ▶ <https://rtavenar.github.io/blog/dtw.html>