

Time Series Clustering

1. input variables

variable selection

*field- and concept specific.
e.g.:*

- **affect, behavior, cognition, desire**
- biopsychosocial domains
- emotion circumplex

data preparation

data cleaning

data exclusion

data transformation

2. feature engineering

(A) feature extraction

psychological time series:

- diversity
 - **multivariate**
- (in)stability
 - mean sum squared differences
 - **mean absolute change**
- central tendency
 - mean
 - **median**
- variability
 - variance / standard deviation
 - **median absolute deviation**
- self-similarity
 - **autocorrelation (lag-1)**
 - periodicity/seasonality (e.g., day, week)
- linear trend
 - **OLS regression slope**
- nonlinearity
 - **GAM spline estimated degrees of freedom (line wiggleness)**

(B) feature selection

- filter
 - univariate (e.g., Laplacian Score)
 - multivariate (e.g., var-covar)
- wrapper
 - sequential (e.g., forward selection)
 - bio-inspired (e.g., ELSA)
 - iterative (e.g., Feature Saliency)
- hybrid (e.g., filter + wrapper)

(B) feature projection

- linear
 - **Principal Component Analysis (PCA)**
 - Factor Analysis (FA)
 - Linear Discriminant Analysis (LDA)
- nonlinear
 - t-distributed Stochastic Neighbor Embedding (t-SNE),
 - Multidimensional Scaling (MDS)
 - Isometric mapping (Isomap)

3. clustering

one of

centroid-based
(e.g., k-means)

distribution-based
(e.g., gaussian mixture model)

density-based
(e.g., dbscan)

hierarchical-based
(e.g., ward hierarchical)

hybrid
(e.g., hybridHclust)

4. evaluation

choice of

visualization / human judge

internal index
(e.g., BIC, SSE, ASW)

external index
(e.g., accuracy, adj. Rand index)