Tutorial
Learning Metrics For Temporal Data

Ahlame Douzal (Ahlame.Douzal@imag.fr)

AMA-LIG, Université Joseph Fourier

(EPAT'2014)

Outline

1. Motivation
2. Temporal alignments
3. Values and behavior based metrics
4. Complex temporal data
   - Temporal kernels
   - Learning temporal matching

Motivation

Temporal Data

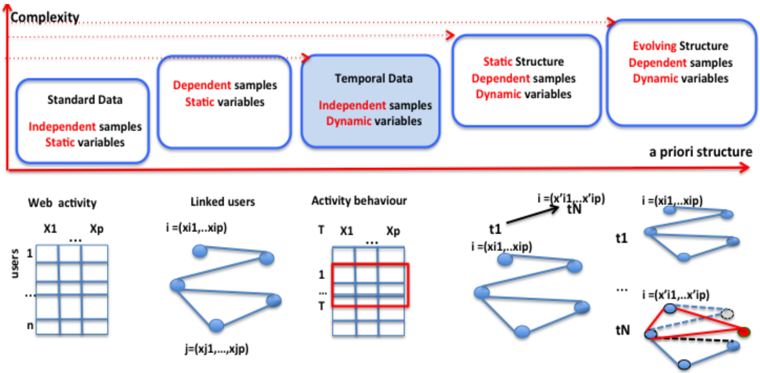
Definition

- A kind of sequence data:
  - an ordered set of elements
  - order criterion: time

Temporal data are ubiquitous

- User Behaviour Analysis
- Evolving social Networks
- Load curve Prediction
- Learning from sensor networks

Temporal data structures



- Classify, cluster Web usage classes

- Measure the dependency between the structure and the content
- Identify group of users of similar /dissimialr usages (hot spot / cold spot components)

- Activity prediction
- Clustering of usage dynamics

- Quantify the dependency between the dynamics and the srtucture
- Draw a map of the main behaviours

- Identify stable structures
- Link, node prediction

Temporal alignments

Temporal alignments

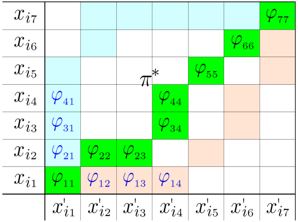
Let $\mathbf{x}_i = (x_{i1}, \ldots, x_{iT})$, $\mathbf{x}_{i'} = (x_{i'1}, \ldots, x_{i'T})$ be two time series of length T.

Definition

An alignment $\pi \in \mathbb{A}$ of length $|\pi| = m$ between two time series $\mathbf{x}_i$ and $\mathbf{x}_{i'}$ is defined as a sequence of m couples of aligned elements:

$$\pi = ((\pi_1(1), \pi_2(1)), (\pi_1(2), \pi_2(2)), \ldots, (\pi_1(m), \pi_2(m)))$$

- π defines a warping function that realizes a mapping from time axis of $\mathbf{x}_i$ onto time axis of $\mathbf{x}_{i'}$



Temporal alignments: conditions

1. No *a priori* knowledge about which sub-period contain important information
2. Continuity and monotonic conditions: π_1 and π_2 define applications from $\{1, ..., m\}$ to $\{1, .., T\}$ that satisfy $\forall j \in \{1, .., m-1\}$:

$$\pi_1(j+1) \leq \pi_1(j) + 1 \text{ and } \pi_2(j+1) \leq \pi_2(j) + 1,$$
$$(\pi_1(j+1) - \pi_1(j)) + (\pi_2(j+1) - \pi_2(j)) \geq 1.$$

3. Boundary conditions:

$$1 = \pi_1(1) \leq \pi_1(2) \leq ... \leq \pi_1(m) = T$$
$$1 = \pi_2(1) \leq \pi_2(2) \leq ... \leq \pi_2(m) = T$$

4. Adjustment window condition:

$$|\pi_1(j) - \pi_2(j)| \leq r, \quad r = 0, .., T \text{ the window length}$$

5. Slope constraint condition:

   - the slop intensity controlled by $p = \frac{r}{c} = 0, 1, 2, ...,$
   it imposes to a point that moves forward in the direction of one dimension consecutive c times, to step at least r times in the diagonal direction.

   - $p = 0$, there is no restrictions on the slope, $p = \infty$ the warping function π is restricted to diagonal.

Values and behavior based metrics

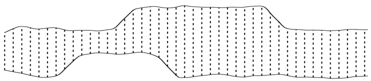
Metrics for temporal data

Euclidean alignment

The Euclidean alignment π *between* $\mathbf{x}_i$ *and* $\mathbf{x}_{i'}$ *alignes elements observed at the same time:*

$$\pi = ((\pi_1(1), \pi_2(1)), (\pi_1(2), \pi_2(2)), \ldots, (\pi_1(T), \pi_2(m)))$$

$$\forall\ k = 1, \ldots m, \quad \pi_1(k) = \pi_2(k) = k, \quad |\pi| = T$$



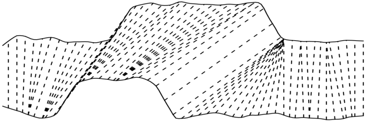
Euclidean Distance for Time Series

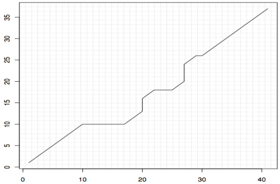
The Euclidean Distance (DE) distance between the time series $\mathbf{x}_i$ and $\mathbf{x}_{i'}$ is given by:

$$DE(\mathbf{x}_i, \mathbf{x}_{i'}) \stackrel{def}{=\!=} \frac{1}{|\pi|} \sum_{k=1}^{|\pi|} \varphi(x_{i\ \pi_1(k)}, x_{i'\ \pi_2(k)}) = \frac{1}{T} \sum_{t=1}^{T} \varphi(x_{it}, x_{i't})$$

φ taken as the euclidean norm.

Unconstrained temporal alignments





Unconstrained Dynamic Time Warping ([SK83], [KL83])

The Dynamic Time Warping (*DTW*) dissimilarity measure between the time series $\mathbf{x}_i$ and $\mathbf{x}_{i'}$ is given by :

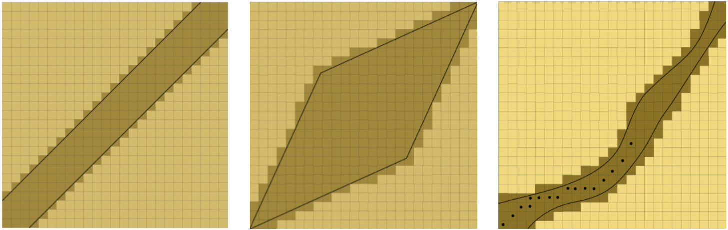
$$DTW(\mathbf{x}_i, \mathbf{x}_{i'}) \stackrel{def}{=\!=} \min_{\pi \in \mathbb{A}} C(\pi)$$

$$C(\pi) \stackrel{def}{=\!=} \frac{1}{|\pi|} \sum_{k=1}^{|\pi|} \varphi(x_{i\ \pi_1(k)}, x_{i'\ \pi_2(k)}) = \frac{1}{|\pi|} \sum_{(t,t') \in \pi} \varphi(x_{it}, x_{i't'})$$

φ taken as the euclidean norm.

Temporal alignments under global/local constraints

Sakoe-Chiba-band [SC78] Itakura [Ita75] Rabiner [Rab89]



Global window Slope constraints Local constraints

Metrics for temporal data: Sakoe-Chiba constraint

Sakoe-Chiba Dynamic Time Warping [SC78]

The Sakoe-Chiba band Dynamic Time Warping (DTW_{SC}) dissimilarity measure between the time series $\mathbf{x}_i$ and $\mathbf{x}_{i'}$ is given by:

$$DTW_{SC}(\mathbf{x}_i, \mathbf{x}_{i'}) \stackrel{def}{=\!=} \min_{\pi \in \mathbb{A}} C(\pi)$$

$$C(\pi) \stackrel{def}{=\!=} \frac{1}{|\pi|} \sum_{k=1}^{|\pi|} w_{\pi_1(k),\pi_2(k)}\, \varphi(x_{i\ \pi_1(k)}, x_{i'\ \pi_2(k)}) = \frac{1}{|\pi|} \sum_{(t,t') \in \pi} w_{t,t'}\, \varphi(x_{it}, x_{i't'})$$

$$w_{t,t'} = 1,\ if |t - t'| < c, \quad \infty\ if |t - t'| \geq c$$

- φ taken as the euclidean norm,
- $w_{t,t'}$ weights that constrain $\mathbb{A}$ to a subset of alignments
- c being the Sakoa-Chiba band width

Temporal alignment

Characteristics

- Dynamic programming alignments deal with delays or time differences
- Pairwise alignments
- Comparison involves the whole observations (no *a priori* knowledge about informative sub-periods)
- Values-based metrics
- Usage in classification/clustering: assumption of similar dynamics within classes

Lack of !!

- Behavior-based metrics
- Comparison involves sub-period importances
- Multiple temporal alignments
- Address time series of complex dynamics

Behavior-based metrics

Behavior-based metrics

Definition (Similar / Opposite behavior)

- Two time series are said similar if, for each period $[t_i, t_{i+1}]$, they increase or decrease simultaneously with the same growth rate
- Two time series are said opposite if, for each period $[t_i, t_{i+1}]$, when one time series increases, the other decreases and (vice-versa) with the same growth rate (in absolute value)
- Two time series are said of different behaviors if not similar nor opposite (linearly and stochastically independent)

Some contributions

- Derivative-based for Slope comparison [KP01], [MLKCW03], [XW10]
- Correlation coefficient-based
  - Kendall coefficient, qualitative distance [cTCK02], [SB08]
  - Spearman coefficient elements rank comparison [AT10], [CVMW07], [RBK08]
  - Autocorrelation-based temporal kernel [GHS11]
  - Temporal Correlation [DCN07], [DCDG09], [DCA12]

Behavior-based metrics: Slope comparison

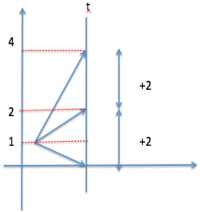
Derivative Dynamic Time Warping [KP01]

$$DDTW(\mathbf{x}_i, \mathbf{x}_{i'}) \stackrel{def}{=\!=} \min_{\pi \in \mathbb{A}} C(\pi)$$

$$C(\pi) \stackrel{def}{=\!=} \frac{1}{|\pi|} \sum_{k=1}^{|\pi|} \varphi(\Delta_{i\ \pi_1(k)}, \Delta_{i'\ \pi_2(k)}) = \frac{1}{|\pi|} \sum_{(t,t') \in \pi} \varphi(\Delta_{it}, \Delta_{i't'})$$

$$\Delta_{it} = \frac{(x_{i\,t} - x_{i\,t-1}) + (x_{i\,t+1} - x_{i\,t-1})/2}{2}$$

- *ignore the sign of the slope*
  (e.g. $\Delta_{it} = +1$, $\Delta_{jt} = +3$, $\Delta_{kt} = -1$, *and* $\varphi(\Delta_{i\,t}, \Delta_{j\,t}) = \varphi(\Delta_{i\,t}, \Delta_{k\,t}) = +2$*)*



Behavior-based metrics: Pearson correlation coefficient

Pearson correlation coefficient

$$\mathbf{x} = (x_1, \ldots, x_n),\ \mathbf{y} = (y_1, \ldots, y_n)$$

$$Cor(x,y) = \frac{\sum_{i,i'}(x_i - x_{i'})(y_i - y_{i'})}{\sqrt{\sum_{i,i'}(x_i - x_{i'})^2}\sqrt{\sum_{i,i'}(y_i - y_{i'})^2}}$$

+ / -

+ Similar, opposite, different $\Rightarrow$ $Cor = 1$, -1 and 0
- Higher $Cor \not\Rightarrow$ similar dynamics
- Involve all the couples i, i' (ignore the temporal dependency)
- Overestimate the similarity (tendency effects, drifts,...)

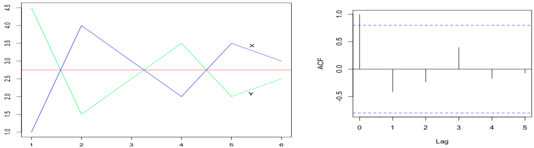
Behavior-based metrics: autocorrelation

Difference between Auto-Correlation Operators [GHS11]

$$\mathbf{x} = (x_1, \dots, x_n), \quad \mathbf{y} = (y_1, \dots, y_n), \quad \tilde{x} = (\rho_1(x), \dots, \rho_K(x)), \quad \tilde{y} = (\rho_1(y), \dots, \rho_K(y))$$

$$\rho_\tau(x) = \frac{\sum_{i=1}^{T-\tau}(x_i - \bar{x})(x_{i+\tau} - \bar{x})}{\sum_{i=1}^{T}(x_i - \bar{x})^2}, \quad d_{DACO}(x, y) = \|\tilde{x} - \tilde{y}\|^2$$

+ Divergence measure between correlogrammes (usefull for model selection)
- Close autocorrelation ρ_τ (lower d_{DACO}) $\not\Rightarrow$ similar behaviors !



$d_{DACO}(x, y) = 0$ for x, y of opposite behaviors as

$$\tilde{x} = \tilde{y} = (1.000, -0.415, -0.234, 0.394, -0.170, -0.074)$$

Behavior-based metrics: temporal correlation

Temporal correlation coefficient $Cort(x, y)$ of order r [DCN07], [DCDG09], [DCA12]

$$Cort(x, y) = \frac{\sum_{i,i'} m_{ii'}(x_i - x_{i'})(y_i - y_{i'})}{\sqrt{\sum_{i,i'} m_{ii'}(x_i - x_{i'})^2}\sqrt{\sum_{i,i'} m_{ii'}(y_i - y_{i'})^2}}$$

$m_{ii'} = 1$ si $|i' - i| \leq r$, 0 otherwise (temporal dependency within r)

+ / -

- + Similar, opposite, different $\Leftrightarrow$ $Cort = 1$, -1 and 0
- + Non sensitive to tendency and drifts (lower r advised)
- \- Sensitive to noise (higher r advised)

Illustration (1)

15 synthetic time series
3 classes: $F_1 = \{1, ..5\}$, $F_2 = \{6, ..10\}$ and $F_3 = \{11, ..15\}$

$$
\begin{aligned}
F_1 &= \{f_1(t)/f_1(t) = g(t) + 2t + 3 + \epsilon\} \\
F_2 &= \{f_2(t)/f_2(t) = \mu - g(t) + 2t + 3 + \epsilon\} \\
F_3 &= \{f_3(t)/f_3(t) = 4g(t) - 3 + \epsilon\}
\end{aligned}
$$

- g(t): a random discrete function,
- $\mu = E(g(t))$
- $\epsilon \rightsquigarrow N(0, 1)$,
- $2t + 3$: a linear trend effect.

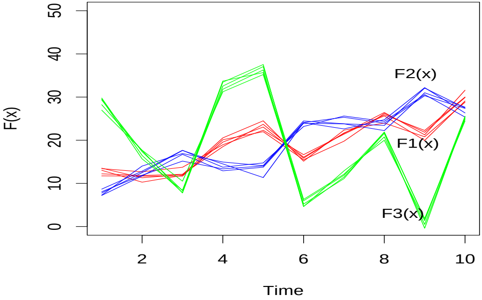
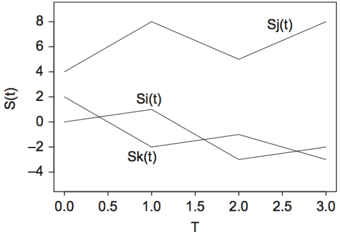


Illustration (2)



- Both the Euclidean distance and the dynamic time warping give S_i closer to S_k than to S_j,
- $d_E(S_i, S_k) = 4.24 < d_E(S_i, S_j) = 15.13 < d_E(S_j, S_k) = 16.15$
- $d_{dtw}(S_i, S_k) = 6 < d_{dtw}(S_i, S_j) = 29 < d_{dtw}(S_j, S_k) = 29$

Clustering time series

Hierarchical clustering

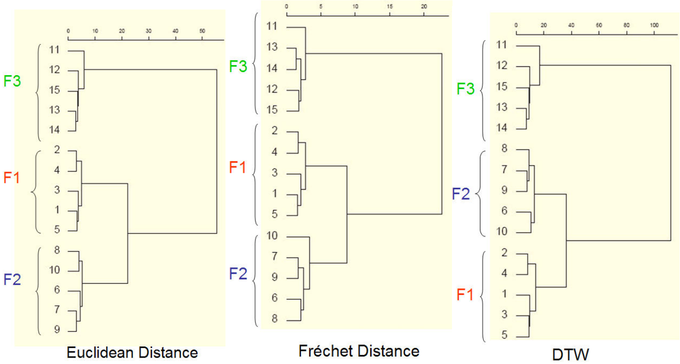
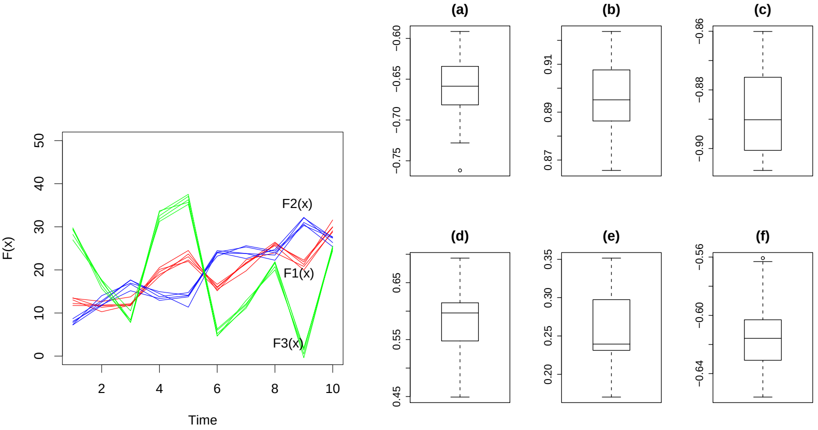


Illustration (3) : *cor* vs *cort*



(a) $\textsc{cort}(F_1, F_2)$, **(b)** $\textsc{cort}(F_1, F_3))$ **(c)** $\textsc{cort}(F_2, F_3)$,
(d)$\textsc{cor}(F_1, F_2)$, **(e)** $\textsc{cor}(F_1, F_3)$, **(f)** $\textsc{cor}(F_2, F_3)$

Dynamic programming alignments

Characteristics

- Dynamic programming alignments deal with delays or time differences
- Pairwise alignments
- Comparison involves the whole observations (no *a priori* knowledge about informative sub-periods)
- Values-based metrics
- Usage in classification/clustering: assumption of similar dynamics within classes

Lack of !!

- Behavior-based metrics
- Comparison involves sub-period importances
- Multiple temporal alignments
- Address time series of complex dynamics

Complex temporal data !

* Temporal kernels

- Learning temporal matching

Temporal kernels: under Euclidean alignment

Temporal Correlation Kernel [DCA12]

$$k_{cort}(\mathbf{x}_i, \mathbf{x}_{i'}) \stackrel{def}{=\!=} Cort(\mathbf{x}_i, \mathbf{x}_{i'})$$

Cort is a linear kernel (p.d.)

Autocorrelation Kernel [GHS11]

$$k_{DACO}(\mathbf{x}_i, \mathbf{x}_{i'}) \stackrel{def}{=\!=} e^{-\frac{1}{\sigma^2} d_{DACO}(\mathbf{x}_i, \mathbf{x}_{i'})}$$

k_{DACO} is a gaussian kernel (p.d.)

Temporal kernels: under dynamic warping alignment

Dynamic Time Warping-based Kernels ([BHB02]

$$k_{DTW}(\mathbf{x}_i, \mathbf{x}_{i'}) \stackrel{def}{=\!=} e^{-\frac{1}{t} DTW(\mathbf{x}_i, \mathbf{x}_{i'})}$$

- non p.d. kernel, t a normalization parameter

Sakoe-Chiba Dynamic Time Warping Kernel

$$k_{SC}(\mathbf{x}_i, \mathbf{x}_{i'}) \stackrel{def}{=\!=} e^{-\frac{1}{t} DTW_{SC}(\mathbf{x}_i, \mathbf{x}_{i'})}$$

- non p.d. kernel, t a normalization parameter

Temporal kernels: under dynamic warping alignment

Dynamic Temporal Alignment Kernel [SNNS01]

$$DTAK(\mathbf{x}_i, \mathbf{x}_{i'}) \stackrel{def}{=\!=} \max_{\pi \in \mathbb{A}} C(\pi)$$

$$C(\pi) \stackrel{def}{=\!=} \frac{1}{|\pi|} \sum_{j=1}^{|\pi|} \varphi(x_{i\ \pi_1(j)}, x_{i'\ \pi_2(j)}) = \frac{1}{|\pi|} \sum_{(t,t') \in \pi} \varphi(x_{it}, x_{i't'})$$

$$\varphi(x_{it}, x_{i't'}) = k_\sigma(x_{it}, x_{i't'}) = e^{-\frac{1}{\sigma^2} \|x_{it} - x_{i't'}\|^2}$$

- non p.d. kernel, but positive semidefinite matrices (sufficient in an experimental context)

Temporal kernels: under dynamic warping alignment

Global Alignment Kernel [CVBM07]

$$K_{softmax}(\mathbf{x}_i, \mathbf{x}_{i'}) \stackrel{def}{=\!=} \sum_{\pi \in \mathbb{A}} \prod_{j=1}^{|\pi|} k(x_{i\ \pi_1(j)}, x_{i'\ \pi_2(j)})$$

$$k(x, y) \stackrel{def}{=\!=} \frac{\frac{1}{2} e^{-\frac{1}{\sigma^2}\|x-y\|^2}}{1 - \frac{1}{2} e^{-\frac{1}{\sigma^2}\|x-y\|^2}}$$

+ non p.d. but, the property $\frac{k}{1+k}$ yields positive semidefinite matrices

- Diagonally dominant Gram matrix (cause of non p.d. property, may be rescaled)

Global Alignment Kernel [Cut11]

$$K_{GA}(\mathbf{x}_i, \mathbf{x}_{i'}) \stackrel{def}{=\!=} \sum_{\pi \in \mathbb{A}} \prod_{j=1}^{|\pi|} k(x_{i\ \pi_1(j)}, x_{i'\ \pi_2(j)})$$

$$k(x, y) \stackrel{def}{=\!=} e^{-\phi_\sigma(x,y)}, \quad \phi_\sigma(x, y) \stackrel{def}{=\!=} \frac{1}{2\sigma^2}\|x - y\|^2 + log(2 - e^{-\frac{1}{2\sigma^2}\|x-y\|^2})$$

Temporal kernels: under dynamic warping alignment

Triangle Global Alignment Kernel [Cut11]

$$K_{TGA}(\mathbf{x}_i, \mathbf{x}_{i'}) \stackrel{def}{=\!=} \sum_{\pi \in \mathbb{A}} \prod_{j=1}^{|\pi|} k(x_{i\ \pi_1(j)}, x_{i'\ \pi_2(j)})$$

$$k(x_{i\ \pi_1(j)}, x_{i'\ \pi_2(j)}) \stackrel{def}{=\!=} \frac{w_{\pi_1(j),\pi_2(j)}\ k_\sigma(x_{i\ \pi_1(j)}, x_{i'\ \pi_2(j)})}{2 - w_{\pi_1(j),\pi_2(j)}\ k_\sigma(x_{i\ \pi_1(j)}, x_{i'\ \pi_2(j)})}$$

w a radial basis kernel on N (a triangular kernel for integers):

$$w(j, j') = \left(1 - \frac{|j - j'|}{c}\right)_+$$

c being the Sakoe-Chiba band width.

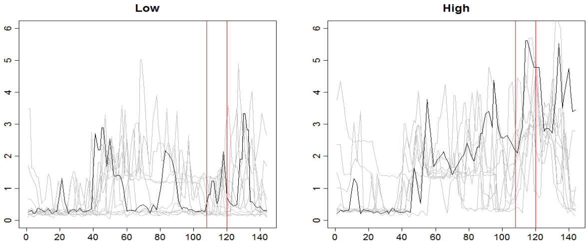
Complex temporal data !

- Temporal kernels

* Learning temporal matching

Time Series: complex data !

Real Data: UCI ML Household Electrical load consumption



Data characteristics:

- Each time series gives a daily load consumption
- In Low (vs. High) class the average consumption between 6-8 pm is lower (resp. higher) than the annual average consumption
- Consumption profiles are different within class

Objective and challenges

Objective

- The early classification (before 6 pm) of a load consumption to predict consumer demand on 6-8pm

Standard approaches

- Based on a standard time series metric (DTW)
- Assign a time series to the class of similar consumption profiles

Challenge

- Load consumption exhibit different global behaviors within classes or nearly similar ones between classes

Learning temporal matching for time series classification

Objective

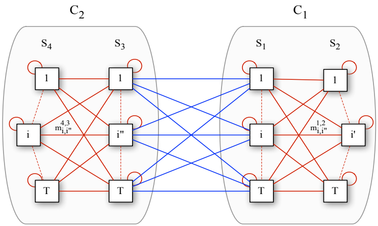
- Complex time series classification: different dynamics within classes, slight differences between classes

For this,

- Enlarge time series alignments to a less constrained temporal matching
- The learning process involves the whole dynamics within and between classes
- Match time series on their shared features within classes and distinctive ones between classes
- Derive a metric based on the highlighted discriminative features to be used for the time series classification.

Proposal's key [FDCG13], [FDCG+14]

Given a set of linked time series (alignment, temporal matching,...)

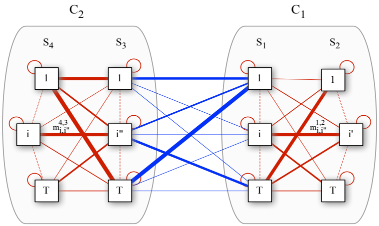


Idea

- Each link induces a variability corresponding to the divergence between the connected values
- To reveal shared features within a class, we minimize the within variance by removing links between non shared features
- To reveal differential features between classes, we maximize the between variance by removing links between shared features

How?

- A new formalization of the classical variance/covariance for a set of time series, as well as for a partition of time series
- Strengthen or weaken links according to their contribution to the variances within and between classes



Variance/Covariance formalization for time series data

- $S_1, \ldots, S_n$ multivariate time series, of length T describing p variables
- X: description of $S_1, \ldots, S_n$ by p variables
- Assume time series linked through DTW alignment, temporal matching, ...

- We define $M_{(n,n)}(M_{ll'})$ as an adjacency block matrix
- A block $M_{ll'}$ specifies the linkage between S_l and $S_{l'}$
- A term of $M_{ll'}$ $m_{ii'}^{ll'} = 1$ if the instants i and i' of S_l and S_l' are aligned, 0 otherwise.

Variance induced by a set of time series

- **Variance/covariance induced by a set of time series**

$$V_M(X) \quad = \quad X^t(I - M')^t P(I - M')X$$

M': row normalized matrix of M

$(I - M')$: Laplacian matrix of the graph defined by the connected observations

Each observation is centered relative to the average of its neighborhood.

Remark: V_M leads to the total Variance/Covariance

- For a complete linkage defined by a unit matrix $M = 1$
- If each time series shrinks to one point

Variance induced by a partition of time series

Variance/covariance within et between classes of time series

$$
\begin{aligned}
V_{M_W}(X) &= X^t(I - M'_W)^t P(I - M'_W)X \\
V_{M_B}(X) &= X^t(I - M'_B)^t P(I - M'_B)X
\end{aligned}
$$

intra-class matching M_W:

$m_{ii'}^{ll'} = 1$ if the linked time series belong to the same class, 0 otherwise.

inter-class matching M_B:

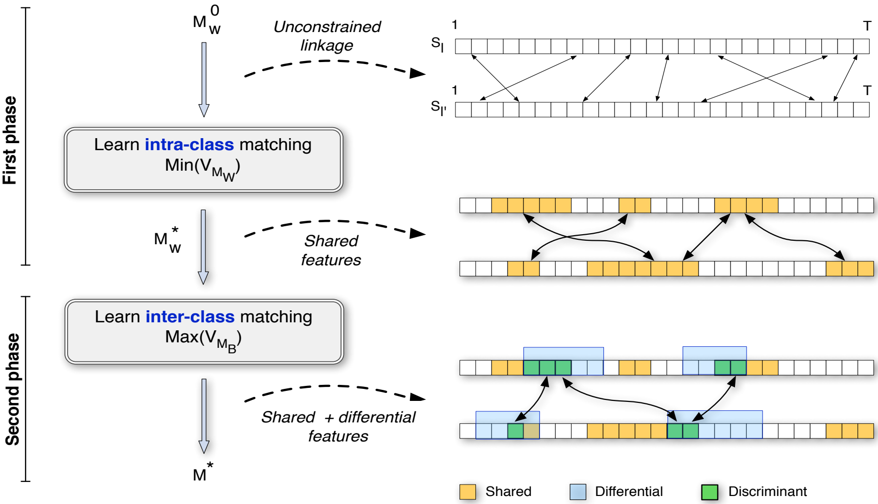
$m_{ii'}^{ll'} = 1$ if the linked time series belong to different classes, 0 otherwise.

Remark: V_{M_W}, V_{M_B} lead to the within, between Variance/Covariance

- For a complete linkage defined by a unit matrix $M_W = 1$, $M_B = 1$
- If each time series shrinks to one point

Learning a discriminative temporal matching

Two consecutive phases algorithm



Learning the intra-class temporal matching

$S_l = (x_1^l, \ldots, x_T^l)$, $S_{l'} = (x_1^{l'}, \ldots, x_T^{l'})$ belonging to C_k ($|C_k| = n_k$)

$M \backslash (i, i', l, l')$: M after the removal of the link (i, i') between S_l and $S_{l'}$ ($m_{ii'}^{ll'} = 0$)

Outlines of the algorithm

1. Initialise M_W as a complete linkage

$$\forall\ i, i' \in \{1, \ldots T\} \text{ and } S_l, S_{l'} \text{ of the same class } m_{ii'}^{ll'} = 1$$

2. Calculate the contribution $C_{ii'}^{ll'}$ to the variance V_{M_W} of each link i, i' between S_l et $S_{l'}$

$$C_{ii'}^{ll'} = V_{M_W} - V_{M_W \backslash (i, i', l, l')}$$

3. Delete links (i, i') ($m_{ii'}^{ll'} = 0$) of positive contributions $C_{ii'}^{ll'} > 0$
4. Iterate steps 2 and 3 until V_{M_W} stabilization

Non degenerate and convergence conditions

$\forall k \in \{1, ..., K\}, \ \forall (l, l') \in C_k, \ \forall (i, i') \in [1, T]^2$

Variance definition

1- $m_{ii}^{ll} > 0$

2- M_W row-normalized : $\sum_{l'=1}^{n_k} \sum_{i'=1}^{T} m_{ii'}^{ll'} = 1$

Non-degenerate variance

3- Each obs. of S_l should be linked to at least one obs. of $S_{l'}$: $\sum_{i'=1}^{T} m_{ii'}^{ll'} > 0$

Convergence of the variance minimization process

4- The delete of (i, i') impacts the i et i' neighborhoods (rows i and i'): at each iteration, delete the link of maximal positive contribution per row

Derive discriminative metric

M_*: the learned discriminative matching

- Let $M_*^{l\cdot}$ be the average matching to S_l:

$$M_*^{l\cdot} = \frac{1}{(n - n_k)\,T} \sum_{l'} M_*^{ll'}$$

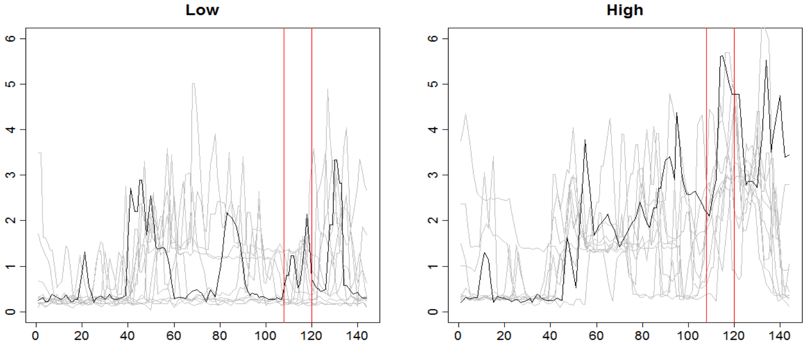
  with $y_{l'} \neq y_l = k$

- The discriminative dissimilarity between S_{New} and S_l

$$D_l(S_l, S_{New}) = \min_{r \in \{0,..,T-1\}} \Big(\sum_{|i-i'| \leq r;\, (i,i') \in [1,T]^2} \frac{m_{ii'}^{l\cdot}}{\sum_{|i-i'| \leq r} m_{ii'}^{l\cdot}} (x_i^l - x_{i'}^{New})^2 \Big)$$

  where r corresponds to the Sakoe-Chiba band width.

Classification of the household electric power consumption

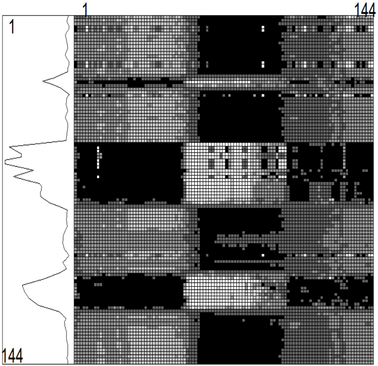


Objective: Early classification of consumption profiles for consumer demand prediction on 6-8pm

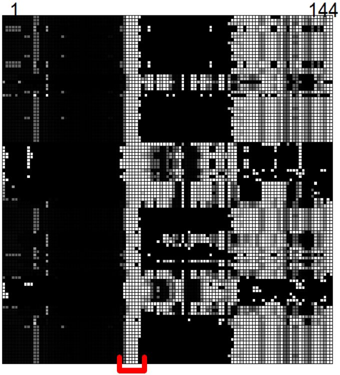
Classification of the household electric power consumption

Learned discriminant matching (CONSLEVEL)

M_W^* (Low)



M_B^* (Low vs. High)

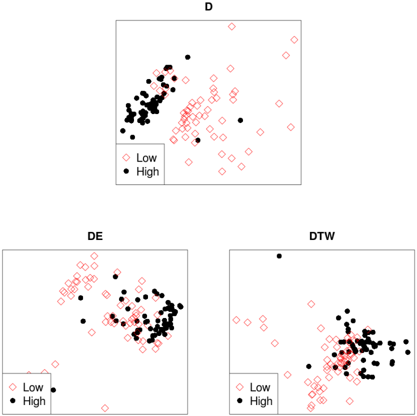


Preliminary results

- Classes compactness/separability by MDS

Table 2: k-Nearest Neighbor classification error rates

	k	D	DE	DTW
BME	1	**0.032**	0.165	0.130
	3	0.034	0.208	0.132
	5	0.062	0.234	0.136
	7	0.079	0.297	0.191
UMD	1	**0.055**	0.173	0.121
	3	0.111	0.333	0.177
	5	0.173	0.343	0.225
	7	0.222	0.378	0.274
CONSLEVEL	1	**0.056**	0.306	**0.289**
	3	**0.044**	0.267	**0.261**
	5	**0.028**	0.233	**0.239**
	7	**0.017**	0.233	**0.233**
CONSSEASON	1	**0.094**	0.239	0.283
	3	0.128	0.228	0.311
	5	0.205	0.200	0.300
	7	0.111	0.222	0.306
TRAJ	1	0.014	**0.012**	0.019
	3	0.018	0.017	0.022
	5	0.022	0.021	0.028
	7	0.019	0.021	0.026



References I

Z. Abraham and P. Tan, *An integrated framework for simultaneous classification and regression of time-series data.*, SIAM International Conference on Data Mining, 2010, pp. 653–664.

Claus Bahlmann, Bernard Haasdonk, and Hans Burkhardt, *Online handwriting recognition with support vector machines-a kernel approach*, Frontiers in Handwriting Recognition, 2002. Proceedings. Eighth International Workshop on, IEEE, 2002, pp. 49–54.

Ljup co Todorovski, Bojan Cestnik, and Mihael Kline, *Qualitative clustering of short time-series: A case study of firms reputation data*, IDDM-2002 (2002), 141.

Marco Cuturi, *Fast global alignment kernels*, Proceedings of the 28th International Conference on Machine Learning (ICML-11), 2011, pp. 929–936.

M. Cuturi, J.-P. Vert, Øystein Birkenes, and T. Matsui, *A kernel for time series based on global alignments*, the International Conference on Acoustics, Speech and Signal Processing, vol. 11, 2007, pp. 413–416.

F. Cabestaing, T.M. Vaughan, D.J. McFarland, and J.R. Wolpaw, *Classification of evoked potentials by pearsonís correlation in a brain-computer interface.*, Modelling C Automatic Control (theory and applications) **67** (2007), 156–166.

A. Douzal-Chouakria and C. Amblard, *Classification trees for time series*, Pattern Recognition **45** (2012), no. 3, 1076–1091.

A. Douzal-Chouakria, A. Diallo, and F. Giroud, *Adaptive clustering for time series: application for identifying cell cycle expressed genes*, Computational Statistics and Data Analysis **53** (2009), no. 4, 1414–1426.

A. Douzal-Chouakria and P.N. Nagabhushan, *Adaptive dissimilarity index for measuring time series proximity*, Advances in Data Analysis and Classification Journal. **1** (2007), no. 1, 5–21.

Cedric Frambourg, Ahlame Douzal-Chouakria, and Eric Gaussier, *Learning multiple temporal matching for time series classification*, Advances in Intelligent Data Analysis XII, Springer, 2013, pp. 198–209.

References II

Cédric Frambourg, Ahlame Douzal-Chouakria, Éric Gaussier, et al., *Learning temporal matchings for time series discrimination*.

A. Gaidon, Z. Harchaoui, and C. Schmid, *A time series kernel for action recognition*, British Machine Vision Conference, 2011.

F. Itakura, *Minimum prediction residual principle applied to speech recognition*, Acoustics, Speech and Signal Processing, IEEE Transactions on **23** (1975), no. 1, 67–72.

J.B. Kruskall and M. Liberman, *The symmetric time warping algorithm: From continuous to discrete. in time warps, string edits and macromolecules.*, Addison-Wesley., 1983.

E. Keogh and M.J. Pazzani, *Derivative dynamic time warping*, First International Conference on Data Mining, 2001.

C.S. Moller-Levet, F. Klawonn, K.H. Cho, and O. Wolkenhauer, *Fuzzy clustering of short time series and unevenly distributed sampling points*, 5th International Symposium on Intelligent Data Analysis (Berlin), aug 2003.

L.R. Rabiner, *A tutorial on hidden markov models and selected applications in speech recognition*, Proceedings of the IEEE **77** (1989), no. 2, 257–286.

J. Rydell, M. Borga, and H. Knutsson, *Robust correlation analysis with an application to functional mri*, IEEE International Conference on Acoustics, Speech and Signal Processing, no. 453-456, 2008.

Young Sook Son and Jangsun Baek, *A modified correlation coefficient based similarity measure for clustering time-course gene expression data*, Pattern Recognition Letters **29** (2008), no. 3, 232–242.

H. Sakoe and S. Chiba, *Dynamic programming algorithm optimization for spoken word recognition*, IEEE Transactions on Acoustics, Speech, and Signal Processing **26** (1978), no. 1, 43–49.

D. Sankoff and J.B. Kruskal, *Time warps, string edits, and macromolecules: the theory and practice of sequence comparison*, Addison-Wesley, 1983.

References III

Hiroshi Shimodaira, Ken-ichi Noma, Mitsuru Nakai, and Shigeki Sagayama, *Dynamic time-alignment kernel in support vector machine*, NIPS, vol. 14, 2001, pp. 921–928.

Y. Xie and B. Wiltgen, *Adaptive feature based dynamic time warping*, International Journal of Computer Science and Network Security **10** (2010), no. 1.