



HAL
open science

Apprentissage de couplages pour la discrimination de séries temporelles

Cédric Frambourg, Ahlame Douzal-Chouakria, Éric Gaussier, Jacques Demongeot

► **To cite this version:**

Cédric Frambourg, Ahlame Douzal-Chouakria, Éric Gaussier, Jacques Demongeot. Apprentissage de couplages pour la discrimination de séries temporelles. CAp 2011 - Conférence Francophone d'Apprentissage, May 2011, Chambéry, France. pp.151-166. hal-00744205

HAL Id: hal-00744205

<https://hal.science/hal-00744205v1>

Submitted on 22 Oct 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Apprentissage de couplages pour la discrimination de séries temporelles

Cédric Frambourg^{1,2}, Ahlame Douzal-Chouakria¹, Eric Gaussier¹, Jacques Demongeot²

¹ LIG (Lab. d'Informatique de Grenoble)

BP 53 - 38041 Grenoble cedex 9, France

cedric.frambourg, ahlame.douzal, eric.gaussier@imag.fr

² TIMC-IMAG (Techniques de l'Ingénierie Médicale et de la Complexité - Informatique, Mathématiques et Applications de Grenoble)

Domaine de la Merci - 38710 La Tronche, France

jacques.demongeot@imag.fr

Résumé : En vue de la discrimination de classes de séries temporelles, ce papier propose l'apprentissage de couplages entre séries. Cette discrimination repose sur un nouveau critère de variance visant à maximiser la compacité des classes et leur séparabilité. Pour ce faire, les couplages appris tendent à lier les séries temporelles selon des caractéristiques, communes au sein des classes et différentielles entre les classes. L'idée principale de cet apprentissage consiste à renforcer ou à pénaliser les liens en fonction de leur contribution à la variabilité intra et inter classe. Pour cela, l'expression usuelle de la variance/covariance est étendue à des ensembles puis à des partitions de séries temporelles. Les expérimentations menées sur plusieurs jeux de données publics et simulés rendent compte de l'efficacité des couplages appris par rapport aux couplages standards pour la discrimination de séries temporelles. **Mots-clés** : Analyse de Données Temporelles, Couplages de séries temporelles, discrimination de séries temporelles, classification, variance/covariance

1. Introduction

La classification de séries temporelles repose souvent sur la définition de couplages entre les séries. Différents travaux dans la littérature présentent la détermination de couplages comme un problème de recherche d'alignements minimisant l'écart entre paires de séries (e.g., Das *et al.* (1997), Nanopoulos *et al.* (2001), Rodriguez & Alonso (2004), Gusfield (1997), Levenshtein

(1966), Navarro (2001), Sankoff & Kruskal (1983)). De telles approches décomposant l'apprentissage par paires de séries, indépendamment de la structure globale des classes, peuvent restreindre la discrimination. En effet, il est d'intérêt en discrimination, d'une part, de rapprocher les séries de même classe selon des caractéristiques communes, d'autre part, d'éloigner les séries de classes différentes sur la base de caractéristiques différentielles. Ainsi, nous proposons une nouvelle stratégie d'apprentissage de couplages discriminants fondée sur un critère de variance. Il s'agit de renforcer ou de pénaliser les liens en fonction de leur contribution à la variance intra et inter classe. La mesure de la variance/covariance est un critère classiquement utilisé dans de nombreuses approches en analyse de données multivariées. Parmi ces approches, on note quelques travaux pionniers étendant le critère de variance/covariance pour la généralisation, entre autres, de l'analyse en composante principale, de l'analyse factorielle et de l'analyse discriminante, à des structures de graphes (Wartenberg (1985), Banet & Lebart (1984), Banet (1988), Thioulouse *et al.* (1995), et Mom (1988)). Ce travail s'articule comme suit : dans la section 2., nous proposons tout d'abord une extension de l'expression standard de la variance/covariance à un ensemble, puis à une partition de séries temporelles. Ces définitions permettent d'introduire dans la section 3., deux algorithmes pour l'apprentissage des couplages temporels intra et inter classe. La section 4. évalue la pertinence des couplages appris, en vue de la discrimination des séries, sur trois jeux de données publics et un jeu simulé. Enfin, les résultats obtenus et les perspectives sont discutés dans la section 5.

2. Variance généralisée à des données temporelles

La notion de variance/covariance est bien connue dans le cadre de données numériques non structurées, tant sous sa forme matricielle qu'analytique. Nous allons généraliser ces définitions à un ensemble de séries temporelles, puis plus particulièrement à une partition de séries.

Soit $X(x_{ij})$, une matrice de données ($n \times p$), caractérisant n observations par p variables numériques X_1, \dots, X_p . La matrice de variance/covariance usuelle $V_{p \times p}$ s'écrit :

$$V = X^t(I_n - 1_n 1_n^t D)^t D (I - 1_n 1_n^t D) X \quad (1)$$

avec I_n la matrice identité, $1_n 1_n^t$ la matrice unité d'ordre n et $D(p_i)$ la matrice diagonale des poids vérifiant $\sum_{i=1}^n p_i = 1$. La variance V_j de la variable X_j

est donnée par la formule suivante :

$$V_j = \sum_{i=1}^n p_i (x_{ij} - \bar{x}_j)^2 = \sum_{i,i'} \frac{1}{2} p_i p_{i'} (x_{ij} - x_{i'j})^2 \quad (2)$$

où $\bar{x}_j = \sum_{i=1}^n p_i x_{ij}$ est la moyenne de X_j . Dans le cas d'observations indépendantes, il est aisé de voir dans l'expression de droite de V_j (Eq. 2) que l'estimation de la variance fait intervenir toutes les différences de valeurs des observations (i, i'). Ainsi, la matrice $1_n 1_n^t$ peut être vue comme une matrice d'adjacence indiquant les couples d'observations impliqués dans l'estimation de la variance.

2.1. Variance d'un ensemble de séries temporelles

Soit X la matrice de données ($nT \times p$) décrivant n séries temporelles S_1, \dots, S_n , de longueur T (nombre d'instant), par p variables X_1, \dots, X_p .

$$X = \begin{matrix} & X_1 & \dots & X_p \\ S_1 & \left(\begin{array}{ccc} x_{11}^1 & \dots & x_{1p}^1 \\ \dots & \dots & \dots \\ x_{T1}^1 & \dots & x_{Tp}^1 \end{array} \right) \\ \vdots & & & \\ S_n & \left(\begin{array}{ccc} x_{11}^n & \dots & x_{1p}^n \\ \dots & \dots & \dots \\ x_{T1}^n & \dots & x_{Tp}^n \end{array} \right) \end{matrix} \quad (3)$$

Soit $M(M_{ll'})$ la matrice ($nT \times nT$) constituée de n^2 blocs matriciels $M_{ll'}(m_{ii'}^{ll'})$. Chaque bloc $M_{ll'}$ explicite le couplage entre S_l et $S_{l'}$, où $m_{ii'}^{ll'} \in [0, 1]$ exprime l'intensité du lien entre l'observation i de S_l et l'observation i' de $S_{l'}$ avec $\sum_{i'} m_{ii'}^{ll'} = 1$. Ainsi, $M_{ll'}$ pondère les écarts entre les observations de S_l et $S_{l'}$ intervenant dans l'estimation de la variance/covariance (Eq. 2).

$$\mathbf{M}_{ll'} = \begin{matrix} & & S_{l'} \\ S_l & \begin{matrix} 1 & \dots & T \\ \dots & \dots & \dots \\ T & \dots & \dots \end{matrix} \left[\begin{array}{ccc} m_{11}^{ll'} & \dots & m_{1T}^{ll'} \\ \dots & m_{ii'}^{ll'} & \dots \\ m_{T1}^{ll'} & \dots & m_{TT}^{ll'} \end{array} \right] \end{matrix} \quad (4)$$

On note en particulier trois couplages classiques : le couplage "complet" définit par $M_{ll'}^0 = \frac{1}{T} U$, (U matrice unité) où toutes les observations sont liées

et équipondérées, le couplage "Identité" défini par $M_{ll'}^I = I$ (I matrice identité) où seules sont liées les observations effectuées aux mêmes instants et le couplage "DTW" dans lequel $M_{ll'}^{DTW}$ est défini par l'alignement obtenu par la dissimilarité usuelle *Dynamic Time Warping* (DTW) entre S_l et $S_{l'}$ (Sankoff & Kruskal (1983)).

Soit M la matrice des couplages définis entre les n séries temporelles. On définit V_M la matrice de variance-covariance induite par l'ensemble des séries et généralisant l'expression usuelle définie à l'Eq. 1.

$$V_M = X^t(I_n - P)^t D(I - P)X \quad (5)$$

avec, I_n la matrice identité, $D(p_i)$ la matrice diagonale des poids, avec $p_i = \frac{1}{nT}$ sous l'hypothèse d'équipondération des observations. On note $P(p_{i,i'})^{ll'}$ la matrice M normalisée en ligne, avec $p_{i,i'}^{ll'} = \frac{m_{ii'}^{ll'}}{\sum_{l'} \sum_{i'} m_{ii'}^{ll'}}$ et V_{M_j} la variance de la variable X_j .

$$V_{M_j} = \sum_{l=1}^n \sum_{i=1}^T p_i (x_{ij}^{S_l} - \overline{x_{ij}^{S_l}})^2 \quad (6)$$

$$\overline{x_{ij}^{S_l}} = \sum_{l'=1}^n \sum_{i'=1}^T p_{i,i'}^{ll'} x_{i'j}^{S_{l'}} \quad (7)$$

où $\overline{x_{ij}^{S_l}}$ représente la moyenne des valeurs prises par X_j dans le voisinage de l'observation i de la série temporelle S_l .

On note dans la suite $\sum_{j=1}^n V_{M_j} = tr(V_m)$

2.2. Variance induite par une partition de séries temporelles

On considère à présent que l'ensemble des séries temporelles S_1, \dots, S_n est partitionné en K groupes, avec $y_i \in \{1, \dots, K\}$ la classe d'appartenance de la série S_i . Pour une partition de séries temporelles, la variance intra évalue la dispersion induite par les séries au sein de la classe. De ce fait, l'estimation de la variance doit se limiter aux observations i de S_l et i' de $S_{l'}$, pour lesquelles $y_l = y_{l'}$ (i.e., $S_l, S_{l'}$ sont de la même classe).

Ainsi, une matrice M définissant les couplages temporels de séries d'une même classe s'écrit

$$M_{ll'} = \begin{cases} I_T & \text{si } l = l' \\ \neq 0 & \text{si } y_l = y_{l'} \\ 0 & \text{si } y_l \neq y_{l'} \end{cases} \quad (8)$$

avec 0 la matrice nulle ($T \times T$) et I_T la matrice identité ($T \times T$). La matrice de variance/covariance V_M introduite à l'Eq. 5 définit alors une extension de la variance/covariance intra classique à une classe de séries temporelles. De façon similaire, la variance inter mesure la séparabilité induite par les séries de classes différentes. Elle fait intervenir les observations i de S_l et i' de $S_{l'}$, pour $y_l \neq y_{l'}$ (i.e., $S_l, S_{l'}$ sont de classes différentes). Ainsi, une matrice M définissant les couplages temporels de séries de classes différentes s'écrit :

$$M_{ll'} = \begin{cases} I_T & \text{si } l = l' \\ 0 & \text{si } y_l = y_{l'} \\ \neq 0 & \text{si } y_l \neq y_{l'} \end{cases} \quad (9)$$

La matrice V_M définit une extension de la variance/covariance inter usuelle à des classes de séries temporelles.

On note dans la suite M_W et M_B les matrices de couplage intra et inter fondées respectivement sur les blocs matriciels définis en 8 et 9, et notés $Mw_{ll'}(mw_{ii'}^{ll'})$, $Mb_{ll'}(mb_{ii'}^{ll'})$.

3. Apprentissage de couplages discriminants

A chaque couplage intra et inter classes de séries temporelles correspond une estimation de la variance intra et de la variance inter. En vue de la discrimination de classes de séries temporelles, notre objectif consiste à apprendre des couplages discriminants M_W et M_B , à savoir minimisant la variance intra V_{M_W} et maximisant la variance inter V_{M_B} .

L'idée général pour l'apprentissage de tels couplages discriminants consiste à évaluer de manière itérative la contribution de chaque couple (i, i') à la variance intra ou inter estimée. Les poids $m_{ii'}^{ll'}$ sont alors pénalisés pour tous les couples (i, i') dégradant le critère de discrimination. Ce processus est réitéré jusqu'à stabilisation de la variance intra ou inter induite.

Nous introduisons dans ce qui suit deux algorithmes *LearnWitAlig* et *LearnBetAlig* proposés respectivement pour l'apprentissage de couplages intra et inter classes. La partie 3.3. discute de la convergence du processus d'apprentissage proposé.

3.1. Apprentissage des couplages entre séries au sein d'une classe

Pour la minimisation de la variance intra V_{M_W} , on fait appel à la procédure *LearnWitAlig* (Algorithm 1). Elle prend comme paramètres d'entrée

la matrice de données X , le vecteur décrivant les classes d'appartenance Y , le paramètre d'arrêt α , et la matrice d'initialisation des couplages intra MW^0 (i.e. couplage complet), fondée sur les blocs matriciels suivants :

$$Mw_{ll'}^0 = \begin{cases} I_T & \text{si } l = l' \\ \frac{1}{T}U & \text{si } y_l = y_{l'} \\ 0 & \text{si } y_l \neq y_{l'} \end{cases} \quad (10)$$

L'algorithme *LearnWithAlign* comporte deux phases. Dans la première phase (ligne 4 à 9), les contributions à la variance intra de chaque couple (i, i') sont évaluées. Pour cela, l'effet induit (augmentation, diminution de la variance) suite à la suppression d'un lien (i, i') est mesuré. Notons que la suppression d'un lien engendre par effet de normalisation la redistribution de son poids sur les voisins. On note $M_{W \setminus (i, i', l, l')}$ la matrice des couplages intra privée du lien (i, i') entre S_l et $S_{l'}$, c'est à dire avec $mw_{ii'}^{ll'} = 0$ et $y_l = y_{l'}$. Ainsi, on définit la contribution $WC_{ii'}^{ll'}$ de (i, i') à la variance intra :

$$WC_{ii'}^{ll'} = tr(V_{M_W}) - tr(V_{M_{W \setminus (i, i', l, l')}}) \quad (11)$$

On note \mathcal{E} l'ensemble des liens (i, i') dont la contribution $WC_{ii'}^{ll'}$ est positive (dont la suppression engendre une diminution de la variance). Ils constituent des liens à pénaliser car augmentant la variance intra.

Dans la seconde phase (lignes 10 à 19), la pénalisation des liens i, i' de \mathcal{E} se traduit par une diminution du poids $p_{ii'}^{ll'}$ d'un couple sélectionné aléatoirement dans \mathcal{E} proportionnellement à $WC_{ii'}^{ll'}$.

On note $M_W^s = P$ la matrice des poids mise à jour à l'itération s et $V_{M_W^s}$ la variance intra induite. Tant que cette dernière n'est pas stabilisée (ligne 23), on réitère les phases 1 et 2 du processus de pénalisation. Dans le cas contraire, la procédure renvoie le meilleur couplage intra classe appris.

3.2. Apprentissage des couplages entre séries de classes différentes

L'apprentissage des couplages pour des séries de classes différentes se fait de manière symétrique. La matrice d'initialisation M_B^0 est fondée sur des couplages complet entre les séries de classes différentes, selon les blocs suivants :

$$Mb_{ll'}^0 = \begin{cases} I_T & \text{si } l = l' \\ 0 & \text{si } y_l = y_{l'} \\ \frac{1}{T}U & \text{si } y_l \neq y_{l'} \end{cases} \quad (12)$$

Algorithm 1 *LearnWitAlig*(X, Y, M_W^0, α)

```

1:  $s = 0$ 
2:  $P(p_{ii}^{ll'}) = M_W^s$ 
3: repeat
4:   for all  $(S_l, S_{l'}) : y_l = y_{l'}$  and  $l \neq l'$  do
5:     for all  $(i, i') \in [1, T] \times [1, T]$  do
6:       {évaluation des contributions intra}
7:        $WC_{i,i'}^{ll'} = tr(V_{M_W^s}) - tr(V_{M_W^s \setminus (i, i', l, l')})$ 
8:     end for
9:   end for
10:   $\mathcal{E} = \{(i, i') / WC_{i,i'}^{ll'} > 0\}$ 
11:  Choisir aléatoirement  $(i, i') \in \mathcal{E}$ 
12:  {pénalisation du poids  $(i, i')$ }
13:   $p_{i,i'}^{ll'} = p_{i,i'}^{ll'} \cdot (1 - \frac{WC_{i,i'}^{ll'}}{\sum_{i', l'} |WC_{i,i'}^{ll'}|})$ 
14:  {normalisation de la ligne i}
15:  for all  $S_k : y_l = y_k$  and  $l \neq k$  do
16:    for all  $r \in [1, T]$  do
17:       $p_{i,r}^{lk} = \frac{p_{i,r}^{lk}}{\sum_{i'} p_{i,r}^{lk}}$ 
18:    end for
19:  end for
20:   $s = s + 1$ 
21:  {mise à jour de  $M_W$ }
22:   $M_W^s = P$ 
23: until  $\frac{tr(V_{M_W^{s-1}}) - tr(V_{M_W^s})}{tr(V_{M_W^{s-1}})} \leq \alpha$  {répéter l'apprentissage jusqu'à stabilisation de la variance intra}
24: return( $M_W^s$ )

```

On définit la contribution d'un lien (i, i') à la variance inter :

$$BC_{i,i'}^{ll'} = tr(V_{M_B^s}) - tr(V_{M_{B \setminus (i,i',l,l')}}^s) \quad (13)$$

\mathcal{E} est l'ensemble des liens (i, i') dont la contribution $BC_{i,i'}^{ll'}$ est négative, c'est-à-dire les liens qui tendent à diminuer la variance inter. À chaque itération, un lien (i, i') est choisi aléatoirement dans \mathcal{E} , et son poids est pénalisé proportionnellement à sa contribution $BC_{i,i'}^{ll'}$.

$$p_{ii'}^{ll'} = p_{ii'}^{ll'} \cdot \left(1 + \frac{BC_{i,i'}^{ll'}}{\sum_{ii',ll'} |BC_{i,i'}^{ll'}|}\right) \quad (14)$$

L'algorithme converge lorsque l'augmentation de la variance descend sous le seuil α .

$$\frac{tr(V_{M_W^{s-1}}) - tr(V_{M_W^s})}{tr(V_{M_W^{s-1}})} \leq \alpha \quad (15)$$

3.3. Convergence du processus d'apprentissage des couplages temporels.

Par symétrie des deux algorithmes *LearnWitAlig* et *LearnBetAlig*, notre discussion va se concentrer sur la convergence de *LearnWitAlig*.

La convergence de l'algorithme est liée à la décroissance de la variance au cours du processus d'apprentissage. Or, l'estimation de la variance V_{M_W} dépend du poids $p_{ii'}$ du lien (i, i') . La variance peut alors s'exprimer comme une fonction V , avec $V(p_{ii'}) = V_{M_W}$. Soit f définie par :

$$\begin{aligned} f : [0, 1] &\rightarrow \mathbb{R} \\ \alpha &\mapsto V_{M_W} - V(\alpha p_{ii'}) \end{aligned}$$

On peut montrer que la fonction f est polynomiale, avec $f(1) = 0$ et $f(0) = WC_{ii'} > 0$. 1 est une racine de f , et en pratique, les autres racines du polynôme sont grandes (de l'ordre de $\frac{1}{p_{ii'}}$). Il n'y a donc pas de zéros du polynôme entre 0 et 1. En particulier, la fonction f est monotone et de signe constant sur $[0, 1[$. Soit α_0 le facteur de pénalisation dans l'algorithme.

$$\alpha_0 = \left(1 - \frac{WC_{i,i'}^{ll'}}{\sum_{ii',ll'} |WC_{i,i'}^{ll'}|}\right) \quad (16)$$

$f(\alpha_0)$ est compris entre 0 et $WC_{ii'}$. En particulier, $f(\alpha_0)$ est positif. Donc, $V_{M_W} > V(\alpha_0 p_{ii'})$. La pénalisation entraîne une diminution de la variance, et donc, la variance chute au cours du processus itératif, ce qui assure la convergence de l'algorithme *LearnWitAli* et par symétrie, celle de *LearnBetAli*.

4. Applications et étude comparative.

4.1. Description des jeux de données.

Les algorithmes proposés pour l'apprentissage des couplages discriminants (*LearnWitAli* et *LearnBetAli*) sont appliqués à des données fréquemment utilisées en classification de séries temporelles, CBF (Saito (1994)), CC et TRAJ (Asuncion & Newman (2007)).

Ces données usuelles partagent des caractéristiques communes : chaque classe identifie un profil distinct, les classes sont facilement séparables selon le profil global des séries, et au sein d'une même classe, les séries varient dans des domaines de valeurs relativement proches. Il est évident que les séries temporelles rencontrées dans des applications réelles peuvent présenter des caractéristiques beaucoup plus complexes. Ainsi, pour étendre le processus de validation à des données moins basiques, on introduit un jeu de données supplémentaire BME, caractérisé par des séries ayant des profils globaux distincts au sein d'une même classe.

BME comprend trois classes de séries (Figure 1). Dans la classe Begin, les séries partagent une signature commune caractérisée par l'apparition d'une cloche apparaissant en début de trajectoire, et peuvent diverger sur la trajectoire restante selon que la cloche principale soit orientée vers le haut ou vers le bas. La classe Middle est constituée de séries partageant un comportement global similaire caractérisé par une grande cloche centrale. Les séries de la classe End partagent un événement commun caractérisé par une cloche située en fin de trajectoire, et dont le comportement global peut différer selon que la cloche principale soit orientée vers le haut ou vers le bas.

Le Tableau 1 précise les principales caractéristiques des quatre jeux de données ci-dessus, indiquant pour chacun : l'origine des données (source=1 : simulés selon le papier d'origine, source=2 : téléchargés du site Machine Learning Repository, source=3, simulés selon les préconisations du présent papier), la taille du jeu (Taille), le nombre de classes (Nb. cla), le nombre de séries par classe (Nb. ST/cla), la longueur des séries (long ST), ainsi que la nature multivariée (Multi.) ou réelle (Reel) des séries.

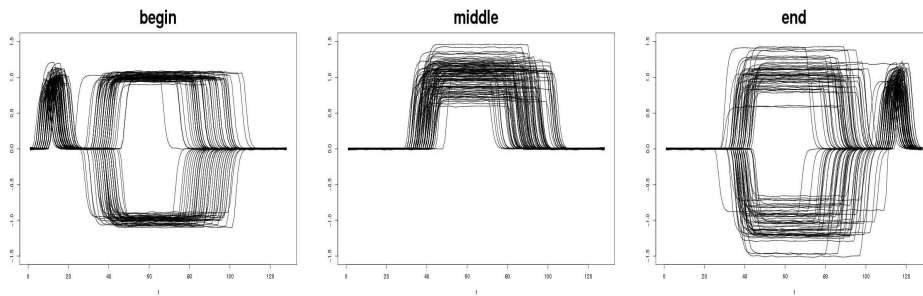


FIGURE 1: Les classes du jeu BME

Nom	Source	Taille	Nb.cla	Nb. ST/cla	Long ST	Multi.	Reel
CBF	1	300	3	100	128	Oui	Non
CC	2	600	6	100	60	Non	Non
TRAJ	2	1000	20	50	20	Oui	Oui
BME	3	300	3	100	128	Non	Non

TABLE 1: Description des jeux de données

4.2. Apprentissage de couplages discriminants.

La Figure 2 nous permet d’illustrer un exemple de couplage appris entre deux séries de la classe *Cylinder* du jeu CBF. D’une part, la figure de gauche visualise les liens entre une observation de S_l et toutes les observations de $S_{l'}$. Les traits les plus épais traduisent des liens de poids forts, ici essentiellement entre l’observation i de S_l et les observations i' du plateau de $S_{l'}$. Les traits hachurés correspondent à des liens de très faibles poids ($mw_{ii'}^{ll'} \approx 0$). D’autre part, la figure de droite visualise l’intensité des connexions entre les observations de S_l et $S_{l'}$. Les cellules les plus claires identifient des régions correspondant à des liens de poids forts. Par exemple, le carré central clair illustre une zone de couplage fort entre les plateaux des deux séries de la classe *Cylinder*.

4.3. Évaluation des variances induites par les couplages M_W^* et M_B^*

Pour évaluer le pouvoir discriminant des couplages appris M_W^* et M_B^* , nous estimons pour chaque jeu de données les variances intra $V_{M_W^*}$, inter $V_{M_B^*}$ et les ratios intra/inter ρ^* induits. L’efficacité de ces couplages est ensuite

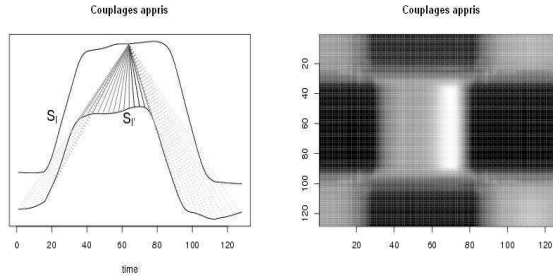


FIGURE 2: Les couplages appris entre deux séries de la classe *Cylinder* (jeu CBF)

étayée par comparaison aux couplages standards : euclidien (M_W^I) et DTW (M_W^{DTW})

Ces résultats sont résumés dans le Tableau 2, permettant de comparer les critères de compacité, de séparabilité et de discrimination sur l'ensemble des jeux de données, et pour les trois couplages principaux (appris, euclidien et DTW).

Jeux	Compacité			Séparabilité			Discrimination		
	$V_{M_W^*}$	$V_{M_W^I}$	$V_{M_W^{DTW}}$	$V_{M_B^*}$	$V_{M_B^I}$	$V_{M_B^{DTW}}$	ρ^*	ρ^I	ρ^{DTW}
CBF	0.119	1.771	0.163	18.441	4.844	1.004	0.006	0.366	0.162
CC	1.732	14.597	2.587	212.339	130.001	107.818	0.008	0.112	0.024
TRAJ	0.057	0.341	0.145	10.830	1.902	0.739	0.005	0.305	0.196
BME	22.161	65.955	22.734	199.476	109.089	35.548	0.111	0.605	0.640

TABLE 2: Comparaison des pouvoirs discriminants des couplages appris, Euclidien, et DTW

Les Figures 3 et 4 illustrent, par exemple, pour les jeux de données CC et TRAJ, la progression des variances intra et inter durant le processus d'apprentissage. Les figures de gauche montrent la décroissance significative de la variance intra $V_{M_W^s}$ et la comparent aux variances intra $V_{M_W^I}$, $V_{M_W^0}$ et $V_{M_W^{DTW}}$ fondées respectivement sur les couplages euclidien, complet et DTW.

De manière similaire, nous constatons dans les Figures 3 droite et 4 droite une croissance drastique de la variance inter $V_{M_B^s}$ comparée aux variances inter $V_{M_B^I}$, $V_{M_B^0}$ et $V_{M_B^{DTW}}$. L'étoile indique les variances intra et inter apprises $V_{M_W^*}$ et $V_{M_B^*}$, retenues lors de l'apprentissage pour un seuil d'arrêt $\alpha = 10^{-3}$.

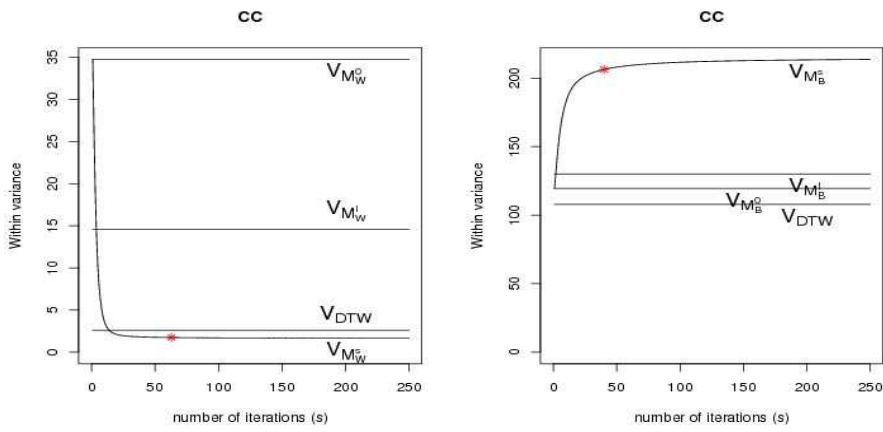


FIGURE 3: Évolution des variances intra et inter classe au cours du processus d'apprentissage ($\alpha = 10^{-3}$) pour le jeu CC.

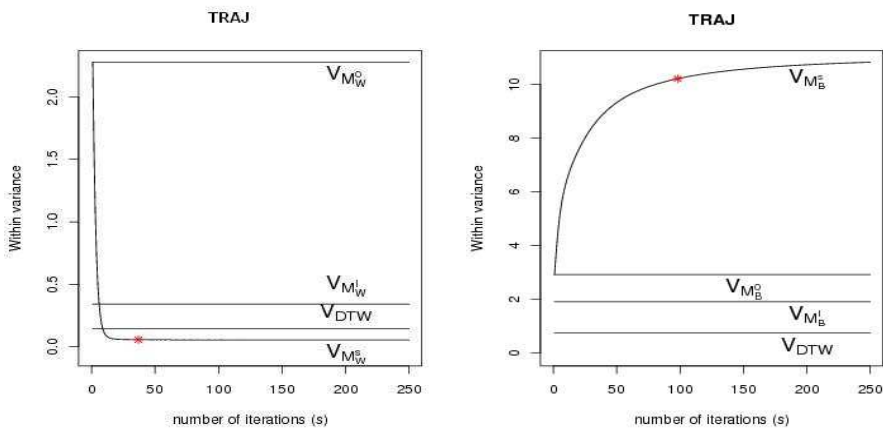


FIGURE 4: Évolution des variances intra et inter classe au cours du processus d'apprentissage ($\alpha = 10^{-3}$) pour le jeu TRAJ.

4.4. Classification fondée sur les couplages appris

Notre objectif est de tester l'efficacité d'une classification (ici 1-Kppv) fondée sur les couplages appris en comparaison avec celles fondées sur les couplages standards. Pour ces derniers, nous considérons la distance euclidienne et la DTW comme métrique. Pour les couplages M^* , nous introduisons un indice de proximité D^* entre une nouvelle série S^* et une série S^l de l'échantillon d'apprentissage.

$$D^*(S^*, S^l) = \|(S^* - \mathbb{S}^l)\|_2 \quad (17)$$

$$\text{avec } \mathbb{S}_l = Mw_l \cdot X \quad (18)$$

où Mw_l correspond à la sous-matrice de Mw constituée de l'ensemble des blocs $Mw_{ll'}$, l' parcourant l'ensemble des séries de la classe. Ainsi, \mathbb{S}_l définit le profil moyen au voisinage de S^l . En particulier, la valeur de l'observation i du profil moyen \mathbb{S}_l se note $\mathbb{S}_{li} = \overline{x_{ij}^{S^l}}$ (i.e. la valeur moyenne au voisinage de i).

Le Tableau 3 résume les taux d'erreur obtenus avec l'algorithme 1-kppv, pour la métrique fondée sur D^* et pour les métriques euclidienne (DE) et Dynamic Time Warping (DTW).

	Métrique	CBF	CC	TRAJ	BME
Taux d'erreur	D^*	9.2%	3.8%	1.3%	8.4%
	DE	10.5%	4.7%	1.2%	17.7%
	DTW	33.2%	9.6%	1.9%	13.4%

TABLE 3: Taux d'erreurs d'une classification 1-kppv

5. Discussion et perspectives.

Dans un premier temps, nous discutons et analysons les comportements des processus d'apprentissage présentés dans les Figures 3 et 4. Les deux figures de gauche indiquent une décroissance qui traduit la chute de la variance intra $V_{M_W^s}$ au cours du processus itératif d'apprentissage. La valeur optimale retenue pour la variance intra est significativement inférieure à $V_{M_W^I}$ et

à $V_{M_W^{DTW}}$. De même, il ressort des Figures 3 droite et 4 droite que la croissance de la variance inter $V_{M_B^s}$ est régulière au cours du processus d'apprentissage. La valeur optimale de la variance inter apprise $V_{M_B^*}$ est significativement plus forte que $V_{M_B^I}$ et $V_{M_B^{DTW}}$. Pour un seuil d'arrêt $\alpha = 10^{-3}$, le processus d'apprentissage converge en moins de 50 itérations pour la variance intra, entre 50 et 100 itérations pour la variance inter.

Le comportement décroissant de la variance intra (qui s'oppose au comportement croissant des variances inter) à travers l'ensemble des jeux de données révèle la pertinence de la pénalisation adoptée en vue de la maximisation de la compacité et de la séparabilité des classes.

La progression drastique de la variance en début d'apprentissage et son ralentissement jusqu'à stabilisation met en lumière deux caractéristiques : la convergence du processus d'apprentissage, et sa capacité à moduler l'intensité des modifications, fortes en début d'apprentissage, quand le critère est loin de l'optimalité, et plus faibles au fil des itérations.

On remarque sur la Figure 4 que les valeurs obtenues pour les couplages euclidien, DTW et appris, sont assez proches. En effet, au travers de la Figure 5, nous constatons que les matrices de couplages apprises pour le jeu TRAJ révèlent un alignement diagonal correspondant à un couplage euclidien.

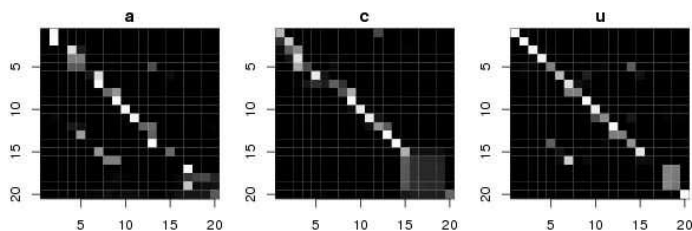


FIGURE 5: Les couplages appris M_W^* pour des classes du jeu TRAJ (“a”, “c”, “u”).

Le Tableau 2 fait état, sur tous les jeux, d'une meilleure discrimination (compacité et isolation) des classes pour M^* . Les alignements de la DTW donnent des classes plus compactes que le couplage euclidien, avec une isolation moindre et un critère de discrimination équivalent. En revanche, les couplages appris optimisent drastiquement le critère de discrimination pour tous les jeux. Le Tableau 3 résume les taux de classification obtenus sur l'ensemble des jeux. Ces résultats montrent l'apport de la métrique fondée sur les cou-

plages appris en comparaison des métriques standards. DE et DTW engendrent des taux d'erreur élevés, en particulier pour des jeux complexes tels que BME. La Dynamic Time Warping donne des taux d'erreur particulièrement élevés pour le jeu CBF. En effet, pour ce jeu, la structure des classes sous-jacente est fortement liée aux instants d'observation. Enfin, pour le jeu TRAJ, les taux de classement pour DE et D* sont équivalents puisque les couplages appris tendent vers des alignements euclidiens, tel que l'illustre la Figure 5.

Dans le cadre de la discrimination de séries temporelles, ce travail est une première avancée, puisqu'il permet la prise en compte de la structure globale des séries. Cependant, les apprentissages des couplages intra et inter demeurent des processus séparés. Les perspectives de ce travail visent à étendre la proposition actuelle pour un apprentissage des structures intra et inter simultanément.

6. Conclusion.

En vue de la discrimination de classes de séries temporelles, ce papier propose l'apprentissage de couplages entre séries. Cette discrimination repose sur un nouveau critère de variance. Pour cela, nous avons étendu l'expression usuelle de la variance/covariance à des ensembles puis à des partitions de séries temporelles. Au travers des expérimentations menées, nous avons montré l'apport des couplages appris, permettant une meilleure discrimination (compacité et isolation des classes) et un meilleur classement des séries temporelles comparés aux couplages standards.

Références

- ASUNCION A. & NEWMAN D. J. (2007). Uci machine learning repository, <http://www.ics.uci.edu/mllearn/mlrepository.html>.
- BANET T. A. (1988). Local and partial correspondence analysis : application to the analysis of electoral data. *Computational statistics quarterly*, 2, p. 89–103.
- BANET T. A. & LEBART L. (1984). Local and partial principal component analysis and correspondence analysis. *Computational Statistics*, p. 113–118.
- DAS G., GUNOPULOS D. & MANNILA H. (1997). Finding similar time series. In *Proc. of the Principles of Knowledge Discovery and Data Mining*, p. 454–456.

- GUSFIELD D. (1997). Algorithms on strings, trees, and sequences. *Computer science and computational biology*.
- LEVENSHTEIN V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Doklady Akademii Nauk SSSR*, 163(4), p. 707–710.
- MOM A. (1988). *Méthodologie statistique de la classification de réseaux de transport*. PhD thesis, U.S.T.L., Montpellier.
- NANOPOULOS A., ALCOCK R. & MANOLOPOULOS Y. (2001). Feature-based classification of time-series data. *International Journal of Computer Research*, p. 49–61.
- NAVARRO G. (2001). A guided tour to approximate string matching. *ACM Computing Surveys*, 33 (1), p. 31–88.
- RODRIGUEZ J. & ALONSO C. (2004). Interval and dynamic time warping-based decision trees. In *Proc of the ACM Symposium on applied computing*, p. 548–552.
- SAITO N. (1994). *Local feature extraction and its application using a library of bases*. PhD thesis, Department of Mathematics, Yale University.
- SANKOFF D. & KRUSKAL J. (1983). *Time warps, string edits, and macromolecules : the theory and practice of sequence comparison*. Addison-Wesley, Publishing Company, Reading, Massachusetts.
- THIOULOUSE J., CHESSEL D. & CHAMPELY S. (1995). Multivariate analysis of spatial patterns : a unified approach to local and global structures. *Environmental and Ecological Statistics*, 2, p. 1–14.
- WARTENBERG D. (1985). Multivariate spatial correlation : a method for exploratory geographical analysis. *Geographical Analysis*, 17 (4), p. 263–283.