Learning Multiple Temporal Matching for Time Series Classification

Cedric Frambourg, Ahlame Douzal-Chouakria, Eric Gaussier

Learning Multiple Temporal Matching for Time Series Classification

Cedric.Frambourg, Ahlame Douzal-Chouakria, and Eric Gaussier

Université Joseph Fourier, Grenoble 1 / CNRS / LIG.
{Cedric.Frambourg,Ahlame.Douzal,Eric.Gaussier}@imag.fr

Abstract. In real applications, time series are generally of complex structure, exhibiting different global behaviors within classes. To discriminate such challenging time series, we propose a multiple temporal matching approach that reveals the commonly shared features within classes, and the most differential ones across classes. For this, we rely on a new framework based on the variance/covariance criterion to strengthen or weaken matched observations according to the induced variability within and between classes. The experiments performed on real and synthetic datasets demonstrate the ability of the multiple temporal matching approach to capture fine-grained distinctions between time series.

1 Introduction

The problem of exploring, classifying or clustering multivariate time series arises in a natural way in a lot of domains, inducing a notable increase activity in this area of research these last years. The Dynamic Time Warping (DTW) [1] is frequently and successfully used in many domains to classify time series that share similar global behaviors within classes subject to some delays. However it fails on complex time series, namely, that present different global shapes within classes, or similar ones between classes. In fact, the applied DTW alignment yields a local view, as it is performed in light of a single pair of time series, ignoring all other time series dynamics within and between clusters; furthermore, the alignment process used is achieved regardless of the analysis process (as clustering or classification), weakening its efficiency on complex data. Several variants of DTW have been proposed to improve performance in classification or clustering. They mostly aim to more finely estimate the DTW parameters, namely, warping constraints, the time weighting, or the underlying divergence function between mapped values. Without being exhaustive, the first part of these works mainly rely on the Sakoe-Chiba, Itakura or Rabiner [2] approches to constrain globally or locally the DTW warping space [3]. The second propositions concentrate on the estimation of time weighting functions [4], whereas the last works pay particular attention to the definition of adaptive divergence functions involving both values and behaviors components of time series [5]. Although these approaches yield more accurate temporal alignments, time series of the same class are assumed to share a single global behavior. In real application time series are generally of

more complex structure. In particular, time series may exhibit different global behaviors within classes, or similar ones between classes. Consequently, for classification purpose, it appears important that the temporal alignment relies on the commonly shared features within the classes and the most differential ones between classes. Such challenging problem is addressed in a recent work [6] [7] to learn pattern graphs from sequential data. For time series, such linkages are hardly reachable by conventional alignments strategies that are mainly limited to monotone warping functions preserving temporal order constraints [1].
To do so, we propose a new approach for multiple temporal alignment that highlights class-specific characteristics and differences. The main idea rely on a discriminant criterion based on variance/covariance to strengthen or weaken links according to their contributions to the variances within and between classes. The variance/covariance measure is used in many approaches, including exploratory analysis, discriminant analysis, clustering and classification [8]. However, to the best of our knowledge, it has never been investigated to define temporal alignment for time series classification. To this end, we propose a new formalization of the classical variance/covariance for a set of time series, as well as for a partition of time series (Section 2). In Section 3, we present a method for training the intra and inter class time series matching, driven by within-class variance minimization and between-class variance maximization. Subsequently, the learned discriminative matching is used to define a locally weighted time series metric that restricts the time series comparison to discriminative features (Section 4). In Section 5, the experiments carried out on both simulated and real datasets reveal the proposed approach able to capture fine-grained distinctions between time series, all the more so that time series of a same class exhibit dissimilar behaviors.

2 Variance/covariance for time series

We first recall the definition of the conventional variance/covariance matrix, prior to introducing its formalization for time series data. Let X be the $(n \times p)$ data matrix containing n observations of p numerical variables. The conventional $(p \times p)$ variance/covariance matrix expression is:

$$V = X^t (I - UP)^t P (I - UP) X \tag{1}$$

where, I is the diagonal identity matrix, U the matrix of ones, and P a diagonal weight matrix of general term $p_i = \frac{1}{n}$ for equally weighted observations. In the following, we provide a generalization of the variance/covariance expression Eq.(1) to multivariate time series observations.

Variance induced by a set of time series For a set of time series, let X be the $(nT \times p)$ matrix providing the description of n multivariate time series $S_1, ..., S_n$ by p numerical variables at T time stamps. The matching between n time series can be described by a matrix M of positive terms composed of n^2 block matrices $M^{ll'}$ $(l = 1, ..., n; l' = 1, ..., n)$. A block $M^{ll'}$ is a $(T \times T)$ matrix that specifies the matching between S_l and $S_{l'}$, of general term $m_{ii'}^{ll'} \in [0, 1]$

giving the weight of the link between the observation i of S_l and i' of $S_{l'}$. Then, the $(p \times p)$ variance/covariance matrix V_M induced by a set of time series $S_1, ..., S_n$ connected to one another according to the matching matrix M can be defined on the basis of Eq.(1), as:

$$V_M = X^t (I - M)^t P (I - M) X \tag{2}$$

where P is a $(nT \times nT)$ diagonal matrix of weights, with $p_i = \frac{1}{n\,T}$ for equally weighted observations. Note that for a complete linkage matching, M is equal to UP and V_M leads to a conventional variance covariance V Eq.(1). For clarity and to simplify notation, we focus for the theoretical developments on univariate time series. The extension to the multivariate case is direct and will be used in the experiments. Thus, let x_i^l be the value of the variable X taken by S_l $(l = 1, ..., n)$ at the ith time stamp $(i = 1, ..., T)$.

Definition 1. *The variance V_M of the variable X is given by:*

$$V_M = \sum_{l=1}^{n} \sum_{i=1}^{T} p_i (x_i^l - \sum_{l'=1}^{n} \sum_{i'=1}^{T} m_{i\,i'}^{l\,l'} x_{i'}^{l'})^2 \tag{3}$$

Note that each value x_i^l is centered relative to the term $\sum_{l'=1}^{n} \sum_{i'=1}^{T} m_{i\,i'}^{l\,l'} x_{i'}^{l'}$ estimating the average of X in the neighborhood of the time i of S_l. The neighborhood of i is the set of instants i' of $S_{l'}$ $(l' = 1...n)$ connected to i with $m_{ii'}^{ll'} \neq 0$. We now proceed to define the variance within and between classes when the set of time series is partitioned into classes.

Variance induced by a partition of time series Let us now consider a set of time series $S_1, ..., S_n$ partitioned into K classes, with $y_i \in \{1, ..., K\}$ the class label of S_i and n_k the number of time series belonging to class C_k. The definition of the *within variance* (i.e. the variance within classes) and the *between variance* (i.e. the variance between classes) induced by K classes is obtained by using the expression given in Eq.(2) based on a matching M specified below.

Definition 2. *The* **within variance** *with an intra-class matching matrix M is given by:*

$$WV_M = \frac{1}{nT} \sum_{k=1}^{K} \sum_{l=1}^{n_k} \sum_{i=1}^{T} (x_i^l - \sum_{l'=1}^{n_k} \sum_{i'=1}^{T} m_{ii'}^{ll'} x_{i'}^{l'})^2$$

with

$$M^{ll'} = \begin{cases} \mathbf{I} & \textit{if } l = l' \\ \neq \mathbf{0} & \textit{if } y_l = y_{l'} \textit{ and } l \neq l' \\ \mathbf{0} & \textit{if } y_l \neq y_{l'} \end{cases} \tag{4}$$

where $\mathbf{I}$ *and* $\mathbf{0}$ *are the* $(T \times T)$ *identity and zero matrices, respectively.*

The general setting for the blocks $M^{ll'}$ of the intra-class matching M is based on three considerations: (a) the Euclidean alignment $(M^{ll} = \mathbf{I})$ linking each time series to itself ensures a variance of zero when comparing a time series with itself, (b) time series within the same class should be connected, while (c) time series of different classes are not connected, as they do not contribute to the within variance. Similarly, we have:

Definition 3. *The* **between variance** *with an inter-class matching matrix M is given by:*

$$BV_M = \frac{1}{nT}\sum_{k=1}^{K}\sum_{l=1}^{n_k}\sum_{i=1}^{T}(x_i^l - (m_{ii}^{ll}x_i^l + \sum_{k'\neq k}\sum_{l'=1}^{n_{k'}}\sum_{i'=1}^{T} m_{ii'}^{ll'}x_{i'}^{l'}))^2$$

with

$$M^{ll'} = \begin{cases} \mathbf{I} & \text{if } l = l' \\ \mathbf{0} & \text{if } y_l = y_{l'} \text{ and } l \neq l' \\ \neq \mathbf{0} & \text{if } y_l \neq y_{l'} \end{cases} \quad (5)$$

where $\mathbf{I}$ *and* $\mathbf{0}$ *are the* $(T \times T)$ *identity and zero matrices, respectively.*

The setting of the inter-class matching M is symmetric with respect to the preceding one, matching between time series of the same class being forbidden, while matching between time series of different classes is taken into account.
As one can note, the matching matrix M plays a crucial role in the definition of the within and between variances. The main issue for time series classification is therefore to learn a discriminative matching that highlights shared features within classes and distinctive ones between classes. To do so, we look for the matching matrix M, under the general settings given in Eqs. (4) and (5), that minimizes the within variance and maximizes the between variance. We present an efficient way to do this in the following section.

3 Learning discriminative matchings

We present here an efficient method to learn the matching matrix M, so as to connect time series based on their discriminative features. The proposed approach consists of two successive phases. In the first phase, the intra-class matching is learned to minimize the within variance. The learned intra-class matching reveals time series connections based on class-specific characteristics. In the second phase, the learned intra-class matching is refined to maximize the between variance.

Learning the intra-class matching We are interested in inferring commonly shared structure within classes, that is in identifying the set of time stamps i' connected to each time stamp i regardless of their weights. Thus, the problem of learning the intra-class matching matrix M to minimize the within variance. We introduce here an efficient approach that iteratively evaluates the contribution of each linked observation (i, i') to the within variance; the weights $m_{ii'}^{ll'}$ are then penalized for all links (i, i') that significantly increase the within variance. For a given class C_k, this process, called *TrainIntraMatch*, is described in Algorithm 1 and involves the following steps.

Algorithm 1 *TrainIntraMatch*(X, α, k)

M = complete intra-class matching — Step 1
for all (l, l') *with* $y_l = y_{l'} = k$ and $l \neq l'$ **do**
  for all $(i, i') \in [1, T] \times [1, T]$ **do**
    $C^{ll'}_{ii'}$ evaluation with Eq. (7) — Step 2
  end for
end for
repeat
  $LinkRemoved = false$
  for all $(i, l) \in [1, T] \times [1, n]$ **do**
    $Link = \arg\max_{i', l'}(C^{ll'}_{ii'})$ satisfying Eq. (9) — Step 3
    if $Link \neq \emptyset$ **then**
      Remove $Link$ ($m^{l,l'}_{i,i'} = 0$) and
      Update weights with Eq. (8)
      Update contributions
      $LinkRemoved = true$
    end if
  end for
until $\neg LinkRemoved$ — Step 4
$return(M_{Intra} = M)$

1. **Initialization (Step 1)** A complete linkage is used to initialize the intra-class matching matrix M, to ensure that all possible matchings are considered and that no *a priori* constraints on the type of matching one should look for are introduced.

$$M^{ll'} = \begin{cases} \mathbf{I} & \text{if } l = l' \\ \frac{1}{T}\mathbf{U} & \text{if } y_l = y_{l'} \text{ and } l \neq l' \\ \mathbf{0} & \text{if } y_l \neq y_{l'} \end{cases} \tag{6}$$

2. **Computing link contributions (Step 2)** We define the contribution $C^{l_1 l_2}_{i_1 i_2}$ of the link (i_1, i_2) between S_{l_1} and S_{l_2} ($y_{l_1} = y_{l_2}$) as the induced variation on the within variance after the link (i_1, i_2) has been removed:

$$C^{l_1 l_2}_{i_1 i_2} = WV_M - WV_{M \backslash (i_1, i_2, l_1, l_2)} \tag{7}$$

where $M \backslash (i_1, i_2, l_1, l_2)$ denotes the matrix obtained from M by setting $m^{l_1 l_2}_{i_1 i_2}$ to 0 and re-normalizing its i_1^{th} row:

$$m^{l_1 l'}_{i_1 i'} \leftarrow \frac{m^{l_1 l'}_{i_1 i'}}{1 - m^{l_1 l_2}_{i_1 i_2}} \tag{8}$$

The evaluated contributions reveal two types of links: the links of positive contribution $C^{ll'}_{ii'} > 0$ that decrease the within variance if removed, and the links of negative contribution $C^{ll'}_{ii'} < 0$ that increase the within variance if removed.

3. **Link deletion (Step 3)** The deletion of a link with positive contribution ensures that the within variance will decrease. In addition, if all links within a row have a negligible contribution to the variance, one can dispense with

removing them in order to (a) avoid overtraining and (b) speed up the process. Thus, a link (i, i') between S_l and $S_{l'}$ is deleted if it satisfies:

$$C_{ii'}^{ll'} > \alpha.WV_{M_1} \text{ and } \sum_{i''=1,(i''\neq i')}^{T} m_{ii''}^{ll'} > 0 \tag{9}$$

where $\alpha \in [0, 1]$ and WV_{M_1} is the initial within variance.
Because the normalization in Eq.(8) performed after the deletion of (i_1, i_2) impacts only the weights of the i_1^{th} row, deleting a single link per row at each iteration of the process guarantees that the global within variance will decrease. Thus, at each iteration one can simply delete the link on each row of maximal contribution compliant with Eq.(9).

4. **Stopping the learning process (Step 4)** The algorithm iterates steps 2, 3 and 4 until there are no more links satisfying the conditions specified in Eq.(9).

From the learned intra-class matching obtained at step 4, noted M_{Intra}, one may induce for each time series S_l one intra-block $M_{Intra}^{l.}$ to indicate the characteristic linkage between S_l and time series of the same class. This intra-block is obtained by summing the block matrices learned for S_l, as follows:

$$M_{Intra}^{l.} = \sum_{l' \in 1,\ldots,n_k} M_{Intra}^{ll'} \tag{10}$$

Learning the inter-class matching The goal of this second phase is to refine the highlighted connections in M_{Intra} (i.e., that connects shared features within classes) to capture the links that are additionally differentiating classes. For this, we refer to a similar algorithm called *TrainInterMatch*, where the inter-class matching is initialized with M_{Intra}, then trained to maximize the between variance BV_M of Definition 3. As for the within variance minimization problem, we adopt the same approach, which consists in iteratively evaluating the contribution of each linked observations (i, i') to the between variance; the weights $m_{ii'}^{ll'}$ are then penalized for all links (i, i') significantly decreasing the between variance. We now turn to the application of the learned matching matrix to time series classification.

4 Time series classification based on the learned matching

Our aim here is to present a way of using learned discriminative matching to locally weight time series for k-nearest neighbor classification. The purpose of the proposed weighting is to restrict the time series comparison to the discriminant (characteristic and differential) features. Let M_* be the discriminative matching learned by the *TrainIntraMatch* and *TrainInterMatch* algorithms, where discriminant linkages are highly weighted. For each S_l of the training sample, we note $M_*^{l\,\cdot}$ the average of the learned matrices $M_*^{ll'}$ $(y_{l'} \neq y_l = k)$:

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

It defines the linkage schema of S_l to a given time series (of the same or of different class) according to S_l own discriminative features. To damp the effect of outliers, the geometric mean could be used for $M_*^{l\cdot}$ as well.
In k-nearest neighbor classification, one can compare a new time series S_{test} to a sample series S_l of C_k based on its learned discriminative matching $M_*^{l\cdot}$. This can be achieved by looking for the delay r that leads to the minimal distance between S_{test} and S_l:

$$D_l(S_l, S_{test}) = \min_{r\in\{0,..,T-1\}} \Big(\sum_{|i-i'|\leq r;\,(i,i')\in[1,T]^2} \frac{m_{ii'}^{l\cdot}}{\sum_{|i-i'|\leq r} m_{ii'}^{l\cdot}} (x_i^l - x_{i'}^{test})^2 \Big) \quad (11)$$

where r corresponds to the Sakoe-Chiba band width [2]. Note that for $r = 0$, D_l defines a locally weighted Euclidean distance involving the diagonal weights $m_{ii}^{l\cdot}$.

5 Experiments

Synthetic datasets The first objective of these experiments is to show through challenging synthetic datasets that the proposed approach successes to recover the *a priori* known discriminative features. For this, two synthetic datasets BME and UMD are considered, where a given class may be composed of time series of different global behaviors and including amplitude and delay variations. BME is composed of three classes *Begin*, *Middle*, and *End* of time series of length 128. Figure 1 illustrates the time series variability within each class, it shows the profile of one time series (in black) compared to the remaining time series (in grey) of the class. In the *Begin* (respectively the *End*) class, time series are characterized by a small bell arising at the initial (respectively final) period. The overall behavior may be different within a same class depending on whether the large bell is up or down positioned. Furthermore, time series of the *Begin* and the *End* classes composed of an up-positioned large bell are quite similar to the *Middle* class time series. The second dataset UMD, composed of three classes *Up*, *Middle*, and *Down* (time series length of 150), introduces an additional complexity with the *Up* and *Down* classes characterized by a small bell that may occur at different time stamps, as illustrated in Figure 1.

Electric power consumption classification The proposed approach is motivated by a classification problem of a real electrical power consumption of customers, to adequately meet consumer demands. To classify such challenging data, we refer to the proposed approach to: a) localize the periods that characterize the daily power consumption of each class, b) highlight periods that differentiate the power consumption of different classes, c) and classify new power consumption based on the learned discriminative features.
The application relies on two public datasets[1] CONSLEVEL and CONSSEASON providing the electric power consumption recorded in a personal home over almost

[1] These data are available at http://bilab.enst.fr/wakka.php?wiki=HomeLoadCurve, and analyzed in [9]

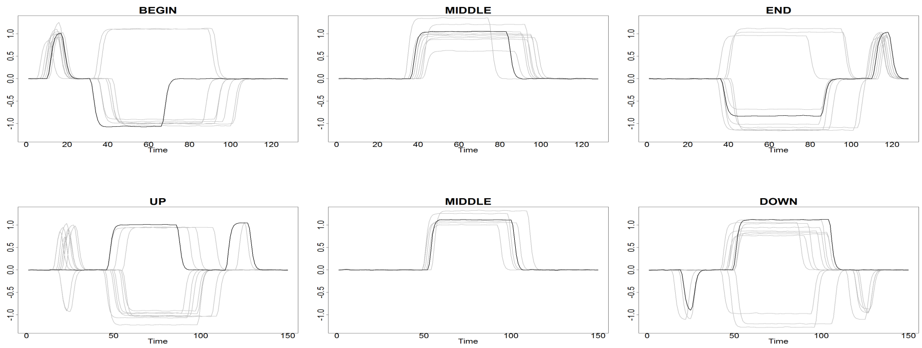

Fig. 1. BME (top three classes) and UMD (bottom three classes) datasets

one year (349 days). Each time series consists of 144 measurements that give the power consumption of one day with a 10 minute sampling rate. CONSLEVEL is composed of 349 time series distributed in two classes (*Low* and *High*) depending on whether the average electric power during the peak demand period [6:00pm-8:00pm] is lower or greater than the annual average consumption of that period. Figure 2 shows the electric consumption profiles within the CONSLEVEL classes; the red frames delineate the time interval [108,120], corresponding to the peak period [6:00pm-8:00pm]. On the other hand, CONSSEASON is composed of 349 time series distributed in two season classes (*Warm* and *Cold*) depending on whether the power consumption is recorded during the warm (from April to September) or cold (from October to March) seasons (Figure 2). Note that the electric power consumption profiles differ markedly within classes in both datasets.

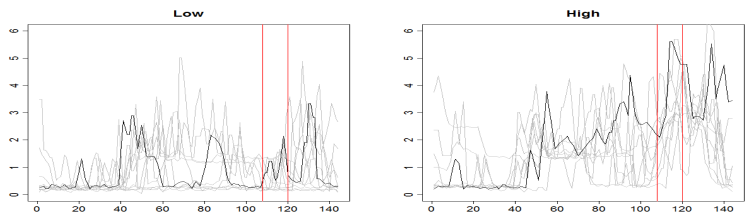

Fig. 2. The electrical power consumption of *Low* and *High* CONSLEVEL classes

Character trajectories classification The objective of this latter dataset is to verify whether the proposed approach can recover standard time series structures within classes, namely, when the classes are mainly composed of time series of similar global behaviors. For this, we have considered a standard dataset on character trajectories TRAJ [10], where time series share a quite similar global behavior within classes (20 classes of 50 time series each).

Validation protocol The proposed approach is applied for the classification of the above datasets. First, the discriminative features are localized, then used to define a locally weighted time series metric D as given in Eq.(11). The relevance of the learned discriminative features and of the induced metric is then studied through a k-nearest neighbor classification for several neighborhood sizes $k = 1, 3, 5, 7$. For BME and UMD datasets a training and test sets of 360 and 1440 time series, respectively, are considered. For the real datasets the performances are evaluated based on 10-fold cross-validation protocol. Finally, the results obtained are compared to two baselines: the Euclidean DE and dynamic time warping DTW distances (Table 1).

Results and discussion The algorithms *TrainIntraMatch* and *TrainInterMatch* are applied to the above datasets with $\alpha = 0.5\%$. As an example, let us first illustrate, for the BME dataset, the progression of the within and between variances during the learning processes (Figure 3). The clearly monotonically decreasing (respectively increasing) behavior of the within (respectively between) class variance, which ends at a plateau, assesses: a) the pertinence of the conducted links penalization to minimize the within variance and maximize the between variance, b) the convergence of the proposed algorithms.

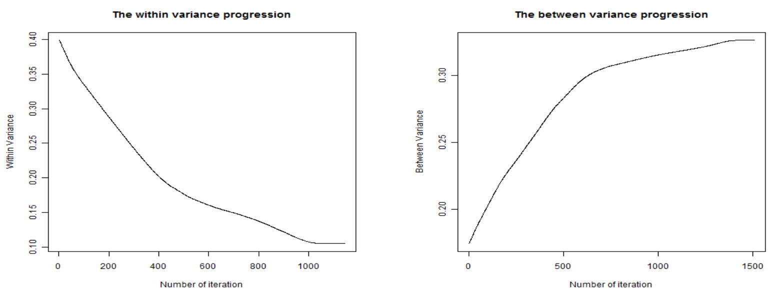

Fig. 3. The within and between variance progression for BME dataset.

For CONSLEVEL, Figure 4 shows the learned intra-class (left) and inter-class (right) blocks for a given time series of the *Low* class. The intra-class block reveals a checkerboard structure, indicating that the electric power consumption within the *Low* class alternates, in a daily period, between a low and a moderately high consumption. The corresponding inter-class block shows the discriminative matching between the considered *Low* class time series and time series of the *High* class (on column). This block displays many discriminative regions; for example, it shows that the power consumption within the *High* class within the period underlined in red (prior to 6:00pm-8:00pm) is especially important in predicting the consumption during the peak period. For each above described dataset, a

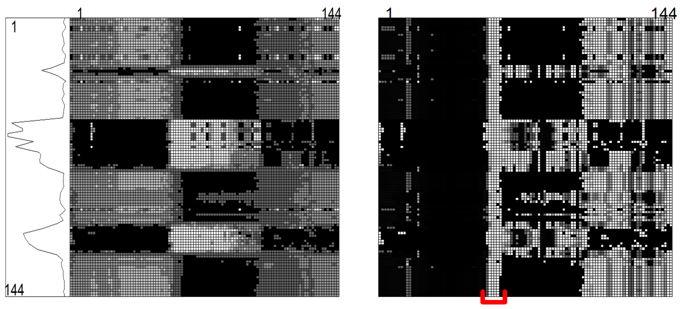

Fig. 4. The intra (left) and inter (right) class matching learned for *Low* class

locally weighted time series metric D is defined based the learned discriminative matching, as given in Eq.(11), then used for the time series classification. The relevance of the proposed approach and of the induced metric are studied according to the validation process described above. The results obtained are compared to two baselines: the Euclidean DE and dynamic time warping DTW distances.

	k	D	DE	DTW
	1	**0.032**	0.165	0.130
BME	3	0.034	0.208	0.132
	5	0.062	0.234	0.136
	7	0.079	0.297	0.191
	1	**0.055**	0.173	0.121
UMD	3	0.111	0.333	0.177
	5	0.173	0.343	0.225
	7	0.222	0.378	0.274

Table 1. k-Nearest Neighbor classification error rates on synthetic data

	k	D	DE	DTW
	1	0.056	0.306	0.289
CONSLEVEL	3	0.044	0.267	0.261
	5	0.028	0.233	0.239
	7	**0.017**	0.233	0.233
	1	**0.094**	0.239	0.283
CONSSEASON	3	0.128	0.228	0.311
	5	0.205	0.200	0.300
	7	0.111	0.222	0.306
	1	0.014	**0.012**	0.019
TRAJ	3	0.018	0.017	0.022
	5	0.022	0.021	0.028
	7	0.019	0.021	0.026

Table 2. k-Nearest Neighbor classification error rates on real data

The misclassification error rates obtained in Table 1 show the efficiency of the proposed locally weighted metric D in discriminating between complex time series classes, compared to standard metrics for time series. In particular, one can note that for all datasets but TRAJ, the best results (in bold) are obtained with D. For TRAJ, the three metrics lead to comparable results suggesting that the Euclidean alignment is an appropriate matching for this dataset. In Figure 5, we can see that the learned discriminative matching, for example, for "c", "o" and "i" characters is close to the Euclidean one, which shows the ability of the proposed approach to recover standard time series alignments. In addition, one can see that for nearly all datasets the best performances are obtained for $k = 1$. For CONSLEVEL, a slight improvement is reached for $k = 7$, indicating a great clusters overlap for this dataset.

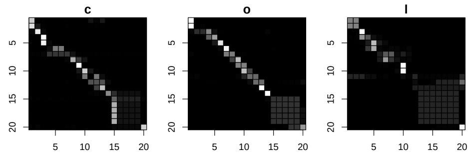

Fig. 5. The learned discriminative matching for the characters "c", "o", and "i" of TRAJ dataset.

Conclusion and future works Our future work will mainly focus on calculus complexity reduction to ensure the proposed method be useable for large scale data. The main idea consists to sparse the initial intra-class matching matrix M. Performances of the scalable variant of the approach will then be compared to alternative methods on large scale data. Furthermore, we aim to study new ways to define weighted metrics based on the discriminative masks $M_*^{l\cdot}$, for instance, by generalizing conventional DTW to achieve alignments limited to the discriminative regions of $M_*^{l\cdot}$.

References

1. J. Kruskall, M. Liberman, The symmetric time warping algorithm: From continuous to discrete. In Time Warps, String Edits and Macromolecules., Addison-Wesley., 1983.
2. H. Sakoe, S. Chiba, Dynamic programming algorithm optimization for spoken word recognition, IEEE Transactions on Acoustics, Speech, and Signal Processing 26 (1) (1978) 43–49.
3. D. Yu, X. Yu, Q. Hu, J. Liu, A. Wu, Dynamic time warping constraint learning for large margin nearest neighbor classification, Information Sciences 181 (2011) 2787–2796.
4. Y. Jeong, M. Jeong, O. Omitaomu, Weighted dynamic time warping for time series classification, Pattern Recognition 44 (2011) 2231–2240.
5. A. Douzal-Chouakria, C. Amblard, Classification trees for time series, Pattern Recognition 45 (3) (2012) 1076–1091.
6. S. Peter, F. Höppner, M. Berthold, Pattern graphs : A knowledge-based tool for multivariate temporal pattern retrieval, in: IEEE Conf. Intelligent Systems, 2012.
7. S. Peter, F. Höppner, M. R. Berthold, Learning pattern graphs for multivariate temporal pattern retrieval, in: Intelligent Data Analysis, Springer Verlag, 2012.
8. R. Fisher, The use of multiple measures in taxonomic problems, Annals of Eugenics 7 (1936) 179–188.
9. G. Hebrail, B. Hugueney, Y. Lechevallier, F. Rossi, Exploratory analysis of functional data via clustering and optimal segmentation, Neurocomputing 73 (2010) 1125–1141.
10. A. Asuncion, D. Newman. UCI, machine learning repository [online] (2007).