

UNSUPERVISED TIME-SERIES CLUSTERING OF DISTORTED AND ASYNCHRONOUS TEMPORAL PATTERNS

Simon Mure¹, Thomas Grenier¹, Charles R.G. Guttmann², Hugues Benoit-Cattin¹

¹Université de Lyon, CREATIS
CNRS UMR5220 ; Inserm U1044
INSA-Lyon ; Université Lyon 1, France

²Center for Neurological Imaging
Department of Radiology, Brigham and Women's Hospital
Harvard Medical School, Boston, MA 02115, USA

ABSTRACT

Most time-series clustering methods, such as *k-means* or *k-medoids*, are initialized by prior knowledge about the number of classes or by a learning step. We propose an unsupervised clustering technique based on *spatiotemporal mean-shift* and optimal time series warping using *dynamic time warping* (DTW). Our main contribution consists in combining a spatiotemporal filtering technique, which gathers similar and synchronized temporal patterns in image sequences, with a clustering algorithm that applies a trajectory constraint on the DTW associations, thereby discriminating between similar time-series that are temporally shifted or warped. We assess the method's robustness on synthetic data, and demonstrate its versatility on brain magnetic resonance and multispectral satellite image sequences.

Index Terms— Unsupervised clustering, Dynamic time warping, Spatiotemporal mean-shift, Time-series alignment

1. INTRODUCTION

As for static data clustering, time-series clustering has to face two major issues: the choice of a clustering algorithm and the choice of an appropriate distance measure to compare the samples. These choices are mostly driven by the underlying applications, which are numerous when it comes to time evolving features (e.g., finance, economics, GPS tracking, health care).

Data clustering approaches can be divided into five families: partitioning, hierarchical, density based, grid based and model based methods [1]. According to [2], three of these five families have specifically been adapted to time-series clustering: partitioning, hierarchical and model based methods. These methods' major drawback is that prior knowledge about the number of clusters [3, 4] or about the data evolution [5] is required.

Thanks to CNRS grant PEPIS INS2I for funding. This work was performed within the framework of the LABEX PRIMES (ANR-11-LABX-0063) of Université de Lyon, within the program "Investissements d'Avenir" (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR).

In addition, the use and the efficiency of these algorithms are dependent on the choice of a time-series similarity measure. A major challenge is to define a similarity metric that is robust to time-shifting and stretching of temporal patterns. To address these issues the dynamic time warping (DTW) metric has been proposed [6, 7] and its relevance to various applications demonstrated [8, 9].

Hence, we propose a new method allowing unsupervised spatiotemporal clustering of time-series, i.e. without any prior information about the number of classes or the data evolution. As shown in Fig.1, we propose to combine the method named spatiotemporal mean-shift (STM-S, [10]) to smooth data evolution and cluster samples evolving synchronously, with a second clustering step based on the DTW metric which will allow to cluster similar time-evolution patterns that are temporally distorted or shifted.

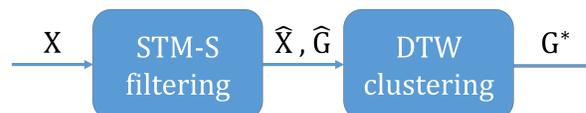


Fig. 1. Proposed clustering pipeline. \mathbf{X} will be the samples and \mathbf{G} the classes.

The STM-S filtering is briefly presented in section 2. Our main contribution to clustering based on DTW optimal pairings is detailed in section 3. The efficiency of our approach is demonstrated on synthetic and real data in section 4.

2. SPATIOTEMPORAL MEAN-SHIFT FILTERING

Prior to defining temporal clusters, the multidimensional extension of the STM-S framework [10] is used to filter the spatiotemporal data.

In a sample set $\mathbf{X} = \{\mathbf{x}_i\}_{i=1\dots n}$, each sample is defined by its position $\mathbf{x}_{s,i}$ and the time-course $\mathbf{x}_{t,i}$ of each of its features C_f :

3. DTW BASED CLUSTERING

$$\begin{aligned} \mathbf{x}_i &= [\mathbf{x}_{s,i} \ \mathbf{x}_{t,i}] & \mathbf{x}_{s,i} &\in \mathbb{R}^S: \text{spatial domain} \\ \mathbf{x}_{t,i} &= [\mathbf{x}_{t,i}^{C_1} \ \dots \ \mathbf{x}_{t,i}^{C_f}] & \text{with } f &= 1, \dots, F: \text{feature index} \\ & & \mathbf{x}_{t,i} &\in \mathbb{R}^{T \times F}: \text{time domain} \end{aligned} \quad (1)$$

This allows to define the STM-S filtering equation of the samples at each iteration as:

$$\mathbf{x}_i^{[k+1]} = \frac{\sum_{j=1}^n Sp_{i,j}(\mathbf{x}_{s,i}^{[k]}, \mathbf{x}_{s,j}^{[k]}) \cdot Ra_{i,j}(\mathbf{x}_{t,i}^{[k]}, \mathbf{x}_{t,j}^{[k]}) \cdot \mathbf{x}_j^{[k]}}{\sum_{j=1}^n Sp_{i,j}(\mathbf{x}_{s,i}^{[k]}, \mathbf{x}_{s,j}^{[k]}) \cdot Ra_{i,j}(\mathbf{x}_{t,i}^{[k]}, \mathbf{x}_{t,j}^{[k]})} \quad (2)$$

with $Sp_{i,j}(\cdot)$ and $Ra_{i,j}(\cdot)$ being the weighted distances between two samples \mathbf{u} and \mathbf{v} in the spatial and temporal domains, respectively:

$$Sp_{i,j}(\mathbf{u}_s, \mathbf{v}_s) = g\left(\left(\mathbf{u}_s - \mathbf{v}_s\right)' \mathbf{H}_s^{-1} (\mathbf{u}_s - \mathbf{v}_s)\right) \quad (3)$$

$$Ra_{i,j}(\mathbf{u}_t, \mathbf{v}_t) = g\left(\left\|\mathbf{H}_r^{-\frac{1}{2}} (\mathbf{u}_t - \mathbf{v}_t)\right\|_\infty^2\right) \quad (4)$$

where \mathbf{H}_s is the spatial bandwidth matrix of size $S \times S$, which is diagonal, squared and positive definite. As for \mathbf{H}_r , it is the range bandwidth matrix used to weight the time-course of each feature. \mathbf{H}_r is of size $TF \times TF$ and is diagonal. It is important to note that we use the diagonal form of \mathbf{H}_r by blocks of size $T \times T$, enabling for independent features to be handled separately.

Using the distance defined in (4) allows to find the biggest difference between two samples in the temporal feature space, with values scaled by \mathbf{H}_r . The samples that will be kept for the mean computation (2) are the ones close enough to the sample of interest \mathbf{u} with regards to the evolution of their features. Thus, the samples participating in the spatiotemporal mean-shift computation are selected with respect to the closeness of their spatial position and of their temporal trajectories to the sample of interest.

For simplicity of understanding, we use the same profile function g to weight both the spatial distance d_s^2 and the range distance d_r^2 :

$$g(d_s^2(\cdot)) = g(d_r^2(\cdot)) = \begin{cases} 1 & \text{if } d_s^2(\cdot), d_r^2(\cdot) \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

Equations (3, 4) ensure that $Sp_{i,j}^{[k]}$ or $Ra_{i,j}^{[k]}$ become zero if the spatial distance between two samples or the distance between two trajectories at a given time-point exceed 1, respectively. Consequently, samples spatially or temporally far from the reference will not be taken into account in the STM-S computation.

This section describes our main contribution: we define a new way to benefit from DTW associations. First, all the samples that have converged after STM-S computation are clustered. Thus, only one representative sample per class (the mode) will be kept for the DTW clustering, thereby significantly reducing the number of samples and increasing computational speed.

3.1. DTW metric

DTW was originally introduced in [6, 7], and aims at finding optimal alignments between all time-points within a query numerical sequence and the closest in a reference numerical sequence. It captures the shape similarities between two sequences even when they are shifted or temporally stretched.

Let us first recall that the optimal step pattern $P_{\mathbf{u},\mathbf{v}}^* = \{\mathbf{p}_l^*\}_{l=1,\dots,L}$ is the ordered set of temporal index pairs that provides the best mapping between two time-series \mathbf{u} and \mathbf{v} . P^* is minimizing the DTW distance D :

$$D(\mathbf{u}, \mathbf{v}, P^*) = \frac{\sum_{l=1}^L d(\mathbf{u}, \mathbf{v}, \mathbf{p}_l^*) \cdot w_l}{\sum_{l=1}^L w_l} \quad (6)$$

with w_l the weights associated to each pair distance, which in turn is dependent on the restrictions chosen to determine P^* . From P^* , we can deduce the warped version of \mathbf{u} and \mathbf{v} noted $P_{\mathbf{u},\mathbf{v}}^*(\mathbf{u})$ and $P_{\mathbf{u},\mathbf{v}}^*(\mathbf{v})$, respectively. These two vectors shortened as $P^*(\mathbf{u})$ and $P^*(\mathbf{v})$ are of size L . The distance d is defined as in [11], as the euclidean distance between two paired multidimensional time-series:

$$d(\mathbf{u}, \mathbf{v}, \mathbf{p}_l^*) = \|P^*(\mathbf{u})_l - P^*(\mathbf{v})_l\| \quad (7)$$

3.2. On a different way to use DTW

As seen in section 3.1, DTW is an integrative metric as it computes the sum of all the weighted distances between pairs of points. Thus, DTW only provides a global metric on whether two time-series are similar or not. However this is not, in our view, a sufficiently accurate metric to use in an unsupervised time-series clustering procedure.

Before undertaking any clustering step, our aim is to determine locally, i.e. for each optimal point coupling, if the time-series are similar. To do so, the same trajectory constraint as in section 2 is used on the samples trajectory. The difference is that the infinity norm is not applied on the distance vector calculated between the values of corresponding time-points, but rather on a distance vector calculated between the values of DTW optimal couplings P^* . The distance d_∞ is then defined as:

$$d_\infty(\mathbf{u}, \mathbf{v}, P^*) = \|\mathbf{H}_r^{-\frac{1}{2}} (P^*(\mathbf{u}) - P^*(\mathbf{v}))\|_\infty \quad (8)$$

This equation quantifies the biggest distance separating two time-series couplings. The clustering is then based on the kernel function:

$$g(d_{\infty}^2(\cdot)) = \begin{cases} 1 & \text{if } d_{\infty}^2(\cdot) \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

As in (5), this equation ensures that if the range distance between two time-series couplings ever exceeds 1, they will not be considered as similar. On the one hand the time-series similarity constraint has been increased and, on the other hand, a decision threshold has been introduced by means of combining DTW couplings and the infinity norm. Therefore, a clustering algorithm of similar temporal patterns is proposed in algorithm 1.

Algorithm 1 Clustering based on symmetric DTW

Require: $\hat{\mathbf{X}} = \{\hat{\mathbf{x}}_i\}$: modes of STM-S results
1: $\hat{\mathbf{G}} = \{\hat{g}_i\}$: Classes initialized as singletons
2: **repeat**
3: $permut \leftarrow true$
4: **for all** $\hat{\mathbf{x}}_i \in \hat{\mathbf{X}}$ **do**
5: $(\hat{\mathbf{x}}_{t,j}^*, P^*) = arg \min_{\hat{\mathbf{x}}_j \in \hat{\mathbf{X}} \setminus \hat{\mathbf{x}}_i} D(\hat{\mathbf{x}}_{t,i}, \hat{\mathbf{x}}_{t,j}, P^*)$
6: **if** $g(d_{\infty}(\hat{\mathbf{x}}_{t,i}^*, \hat{\mathbf{x}}_{t,j}^*, P^*)) = 1$ **then**
7: $g_i \leftarrow g_j$
8: $permut \leftarrow true$
9: **end if**
10: **end for**
11: **until** $permut$ is true
12: **return** \mathbf{G}^*

4. EXPERIMENTS AND RESULTS

First, we evaluated the proposed pipeline on a synthetic dataset in order to assess its efficiency and accuracy with known feature evolution patterns. Then, we used both longitudinal MRI sequences and multispectral satellite acquisitions to show the pipeline’s efficiency with mono and multivariate feature evolutions. The symmetric form of DTW was used in the experiments, as it was shown by [7] to perform better at finding optimal alignments.

4.1. Synthetic data

Synthetic spatiotemporal data has been created in order to evaluate the clustering scheme. It is composed of two $176 \times 234 \times 36$ time-points image sequences based on the Shepp-Logan phantom [12], where a monovariate feature evolution is assigned to each region. These time-series evolutions were then corrupted by a Gaussian blur and an additive normal noise as in [10]. The corresponding regions in the sequences are supposed to contain the same evolution models excepted that they are temporally shifted and/or stretched, and slightly

dissimilar in amplitude. Thus, we expect the algorithm to find, without any prior knowledge, the right correspondences between the sequences.

Fig.2 shows the results after STM-S clustering of the synthetic data. In both cases the image regions correspond exactly with the ground truth. Nevertheless, they are not associated after the STM-S clustering because their time-series evolutions (Fig.3) are dissimilar with respect to (4).

In contrast, it can be seen on Fig.3 and 4 that DTW clustering performs well at clustering similar temporal patterns when they are not temporally synchronized: corresponding regions are adequately associated with respect to the ground truth.

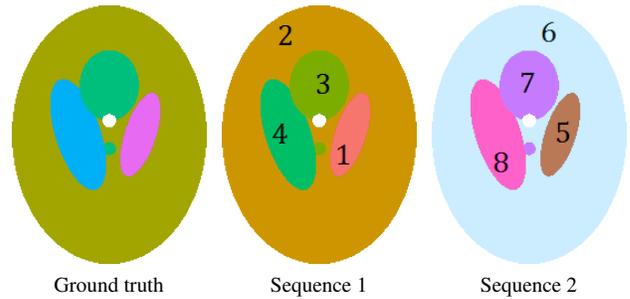


Fig. 2. STM-S Clustering of the synthetic image sequences. Scales were set to $\mathbf{H}_s = diag(\infty)$ and to $\mathbf{H}_r = diag(0.4)$.

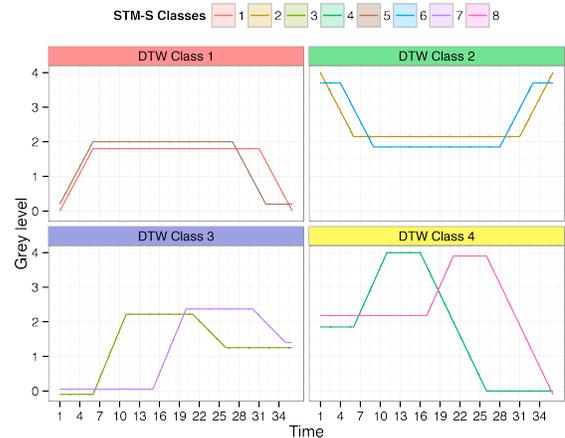


Fig. 3. Classes obtained after STM-S clustering of the synthetic data. The red, green, blue and yellow boxes represent the classes obtained after DTW clustering of the evolution patterns. The curve colors correspond to the colors in Fig.2 and the box colors correspond to the colors in Fig.4.

4.2. Real data

We have shown in [10] that the STM-S procedure used with grey level evolution patterns of multiple sclerosis lesions allows to highlight the concentric evolution patterns observed

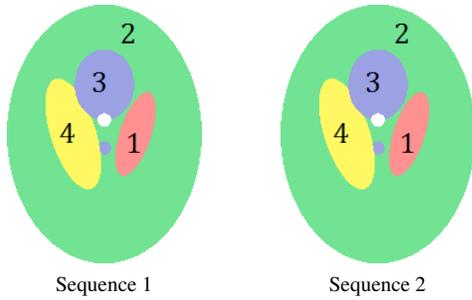


Fig. 4. DTW clustering of the synthetic image sequences after STM-S filtering of the pixel evolution patterns. The range scale is set to the same value as in the STM-S filtering.

by [13, 14] and reminiscent of the intensity evolution model proposed in [14]. However, we have not managed yet to automatically identify distinct clusters of lesion evolution patterns because of temporal shifts, due to the asynchronous nature of the disease, in which lesions are initiated at different times. Thus, DTW clustering is now used to find relationships, if any exist, between the evolution patterns in different lesions within or between patients.

The data used are four 3D regions of interest (ROI), each containing one new lesion evolving over time on T2-weighted MRI images of three multiple sclerosis patients. The utilized images are part of an MRI natural history study of multiple sclerosis patients [13].

The results in Fig.5 show that the DTW clustering of the four ROIs after STM-S filtering was able to group the feature temporal patterns of each region accordingly to expectations, i.e. associating pixels to the lesions centers (likely containing more destructive, neurodegenerative processes) in one cluster, and those in the periphery (likely consisting of more inflammatory, reversible swelling (edema)) of lesions to another cluster. Interestingly, one of the lesion centers does not co-cluster with the other three (lesion Center 1 in Fig.5). This appears to be explained by the fact that this lesion showed near-total signal recovery, with a higher initial peak amplitude and pixel grey levels retrieving their original, baseline values by the end of the follow-up, in contrast to the other three lesions that only recover partially. Our main interpretation regarding the outlier classes in Fig.5 is that they detect artifacts related to imprecise MRI co-registration, which is mostly due to the limited image resolution than to the registration process itself.

Finally, we also tested our pipeline on a 15 time-points satellite image sequence acquired over the year 2007 in the south-east of France. Each image is orthorectified and composed of four spectral bands, but we chose to use only three of them (near infrared, red and green) because of data correlations. The results in Fig.6 show the versatility and the efficiency of the proposed pipeline for the clustering of satellite images.

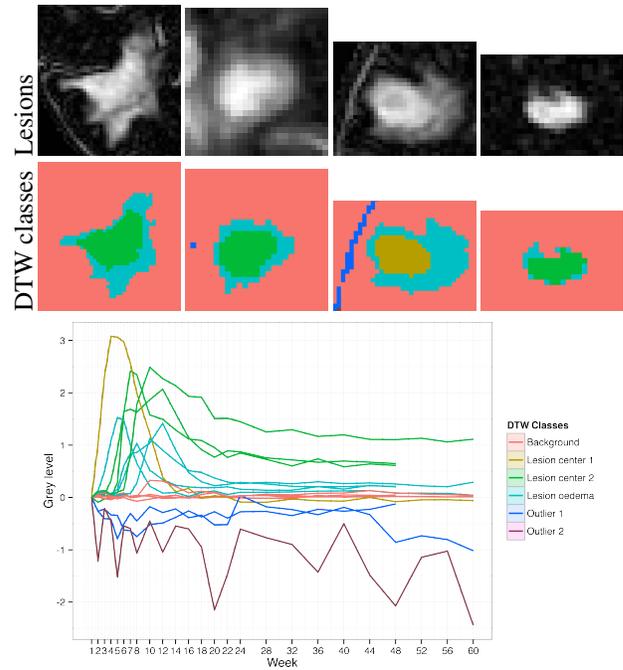


Fig. 5. 3D sections of the ROI's before and after DTW clustering. Each color in the 2D images of labels correspond to the colors of the evolution curves.

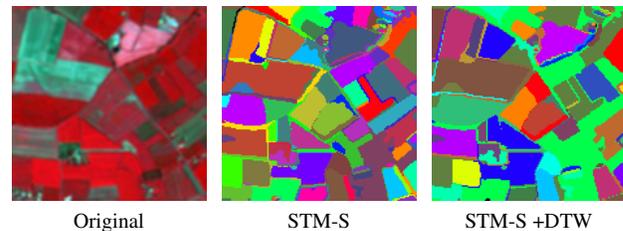


Fig. 6. Image extracted from the original satellite sequence represented in "false colors". Comparison of the clustering results obtained from STM-S and STM-S + DTW.

5. CONCLUSION

In this paper, we propose a new method combining spatiotemporal mean-shift and DTW clustering to handle the unsupervised classification of multivariate temporal patterns. We take advantage of DTW optimal mapping between the time-series and apply a specific trajectory comparison constraint on them, thanks to the infinity norm, to decide whether or not they are similar. Analysis of synthetic data suggests that the new approach improves classification accuracy compared to spatiotemporal mean-shift alone, without requiring prior knowledge of cluster number or time-evolution patterns. The versatility of our method has been demonstrated on both longitudinal magnetic resonance follow-ups acquired on several patients suffering from MS, and satellite image sequences.

6. REFERENCES

- [1] Jiawei Han, Micheline Kamber, and Jian Pei, *Data mining: concepts and techniques: concepts and techniques*, Elsevier, 2011.
- [2] T Warren Liao, "Clustering of time series data—a survey," *Pattern recognition*, vol. 38, no. 11, pp. 1857–1874, 2005.
- [3] François Petitjean, Alain Ketterlin, and Pierre Gançarski, "A global averaging method for dynamic time warping, with applications to clustering," *Pattern Recognition*, vol. 44, no. 3, pp. 678–693, 2011.
- [4] John Paparrizos and Luis Gravano, "k-shape: Efficient and accurate clustering of time series," in *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*. ACM, 2015, pp. 1855–1870.
- [5] Konstantinos Kalpakis, Dhiral Gada, and Vasundhara Puttagunta, "Distance measures for effective clustering of arima time-series," in *Data Mining, IEEE Proceedings of the International Conference on*, 2001, pp. 273–280.
- [6] Hiroaki Sakoe and Seibi Chiba, "A dynamic programming approach to continuous speech recognition," in *Proceedings of the seventh international congress on acoustics*, 1971, vol. 3, pp. 65–69.
- [7] Hiroaki Sakoe and Seibi Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 26, no. 1, pp. 43–49, 1978.
- [8] Toni M Rath and Raghavan Manmatha, "Word image matching using dynamic time warping," in *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*. IEEE, 2003, vol. 2, pp. II–521.
- [9] A Piyush Shanker and AN Rajagopalan, "Off-line signature verification using dtw," *Pattern recognition letters*, vol. 28, no. 12, pp. 1407–1414, 2007.
- [10] Simon Mure, Thomas Grenier, Dominik S. Meier, Charles R.G. Guttman, and Hugues Benoit-Cattin, "Unsupervised spatio-temporal filtering of image sequences. a mean-shift specification," *Pattern Recognition Letters*, vol. 68, Part 1, pp. 48 – 55, 2015.
- [11] Toni Giorgino, "Computing and visualizing dynamic time warping alignments in r: the dtw package," *Journal of statistical Software*, vol. 31, no. 7, pp. 1–24, 2009.
- [12] L Shepp and B. F. Logan, "The fourier reconstruction of a head section," *Nuclear Science, IEEE Transactions on*, vol. 21, pp. 21–43, 1974.
- [13] CR Guttman, Sungkee S Ahn, Liangge Hsu, Ron Kikinis, and Ferenc A Jolesz, "The evolution of multiple sclerosis lesions on serial mr.," *American journal of neuroradiology*, vol. 16, no. 7, pp. 1481–1491, 1995.
- [14] Dominik S Meier, HL Weiner, and Charles RG Guttman, "Mr imaging intensity modeling of damage and repair in multiple sclerosis: relationship of short-term lesion recovery to progression and disability," *American Journal of Neuroradiology*, vol. 28, no. 10, pp. 1956–1963, 2007.