Using Barycenters as Aggregate Representations of Repetition-Based Time-Series Exercise Data

Bahavathy Kathirgamanathan[1], James Davenport[2,3], Brian Caulfield[2,3], and Pádraig Cunningham[1]

[1] School of Computer Science, University College Dublin, Ireland
[2] School of Public Health, Physiotherapy and Sports Science, University College Dublin, Dublin, Ireland
[3] Insight Centre for Data Analytics, University College Dublin, Dublin, Ireland
bahavathy.kathirgamanathan@ucdconnect.ie

Abstract. This paper introduces the use of time-series barycenter averaging as a means of providing aggregate representations of repetition-based exercises. Time-series averaging is not straightforward as small misalignments can cause key features to be lost. Our evaluation focuses on the Forward Lunge exercise, an exercise that is used for strengthening, screening and rehabilitation. The forward lunge is a repetition-based movement so assessment entails comparing multiple repetitions across sessions. We show that time-series barycenters produced using Dynamic Time Warping are effective for this application. The barycenters preserve the key features in the component time-series and are effective as an aggregate representation for further analysis.

Keywords: Barycenter · Multiple-rep Exercise · Time-Series averaging

1 Introduction

The use of wearable sensors for the detection and evaluation of exercise has become widespread in recent years [6]. Evaluating exercise performance usually involves analysis of multiple repetitions of the same target movement profile [9]. The sensor data is typically high frequency and represents movement in three anatomical frames so each set of repetitions produces a significant amount of data.

In this paper we present a preliminary evaluation of the use of time-series barycenters (TSBs) to aggregate the multiple repetitions into a single 'average' signal. Averaging time-series is not straightforward as small shifts in the phase can cause key shared features to be *smeared out*. In this paper we consider time-series averaging incorporating Dynamic Time Warping (DTW) that stretches (warps) the time-series to find good alignments [11]. TSBs were first introduced to tackle clustering problems in time series research. Since then, TSBs have been used mostly for data mining tasks such as data reduction to produce faster and more accurate classification [10]. TSBs have also been used as an averaging strategy in applications such as signature template matching where a mean template is calculated using a time series barycenter[7].

Our analysis compares the DTW based averaging approaches of DTW Barycenter averaging (DBA) and Soft-DTW Barycenter with the more commonly used Euclidean averaging technique in preserving the original features of the time series from the lunge exercises.

Functional exercises in general can be used to determine injury risk and can aid the clinician in recommending injury prevention strategies [3]. The Forward Lunge is a functional exercise that is representative of the lower limb function during activity and hence is commonly used in strength and conditioning, injury risk screening, and rehabilitation [8]. In most clinical settings, interpretation of the forward lunge remains subjective, depending on visual interpretation. Limitations associated with this is the poor level of intra/inter rater reliability and inability to assess multiple aspects simultaneously. 3D motion capture offers an in depth objective alternative but is expensive, time consuming, and restricted to a laboratory environment [8]. Wearable sensors such as Inertial Measurement Units (IMUs) offer a cost-effective and objective method to capture the movement during such exercises. However, this sensor data is time series in nature and due to the vastness of the data being outputted, can be difficult to visualize in a meaningful way. In this paper, we propose that the barycenter can act as an aggregate measure of the time series whilst also preserving the key features required to analyse the quality of the lunge.

The following section describes in detail the data used for evaluation and the concept of the time series barycenter and some of the different approaches that can be used to compute the barycenter. This is followed by an evaluation of the suitability of the different approaches in the representation of multi-rep lunges and the suitability to use barycenters for further data mining tasks.

2 Background

2.1 Multiple rep exercises

In sports science, it is common practice to perform repetitions of a functional movement exercise to gain insight into multiple components of that individual's movement pattern or strategy. The data used for this analysis is from a current study focusing on assessing the characteristics of the forward lunge and changes due to alterations in motor function using IMU's (under review). The data was collected using IMU sensors (Shimmer, Ireland) with sensors placed on the shanks, thighs, and the lumbar spine.

The data was collected from 25 participants performing forward lunges before and after an exercise bout, causing central and peripheral fatigue resulting in short term alteration to motor function. Three sets of five repetitions were performed both pre- and post-fatigue. Participants were fatigued through a modified 60-second wingate protocol using a cycle ergometer (Lode, Netherlands) [4]. The lunges were performed bilaterally at time intervals 0, 10, and 20 minutes pre- and post-fatigue. For the analyses carried out in this paper, the signals from the sensor placed on the shanks were used.

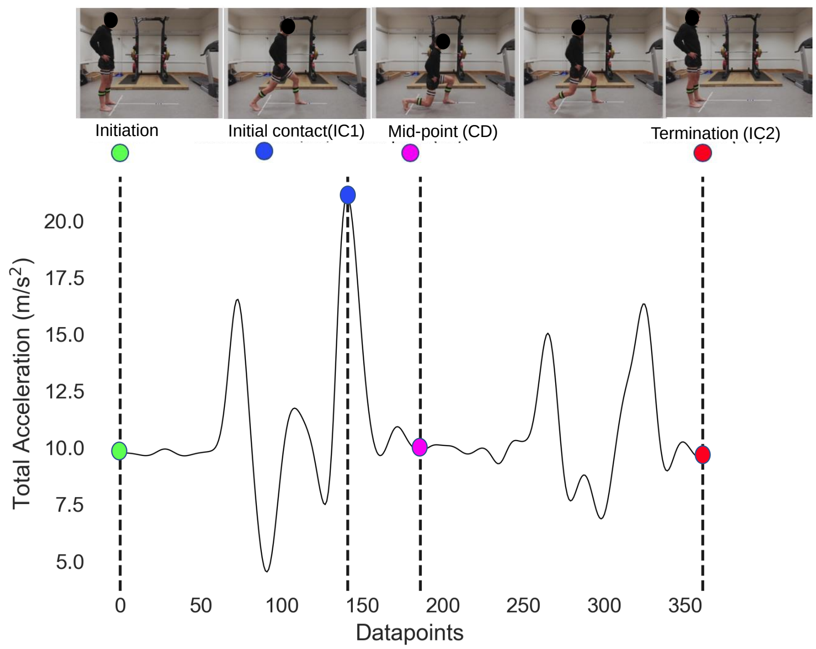


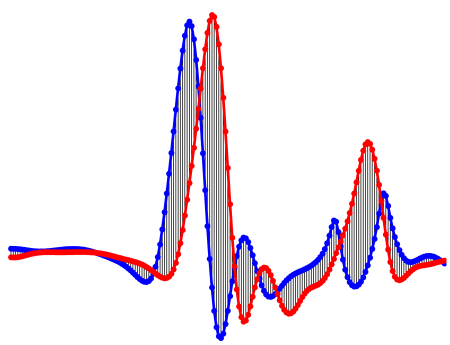
Fig. 1: Time series of forward lunge with corresponding image demonstrating key features used in identification of key lunge characteristics

The collected data is a set of time series capturing the duration of the lunge. Figure 1 shows a single time series with different stages of the lunge. Each of the features is associated with a functional event during the lunge and hence preserving these features is important in understanding performance. In particular, the initial contact (IC1) is the point where peak acceleration occurs and is a particular point of interest whilst analysing performance.

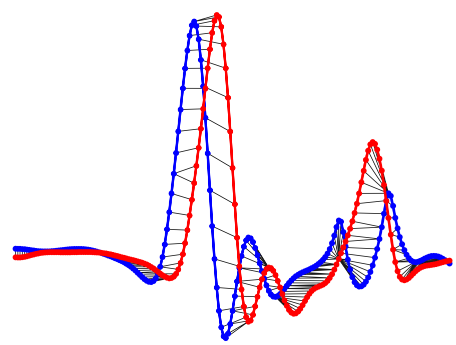
2.2 Time-series Barycenter

The time series barycenter is an average measure of a collection of time series. Time series averaging strategies are generally classified as either local or global. Local averaging strategies use pairwise averaging through which a collection of series are iteratively averaged into one series. However, local averaging strategies are dependent on the order in which the series are averaged and hence changing the order could lead to different results. To tackle the problems associated with local averaging strategies, recent advances have looked at global averaging methods which aim to compute the average of the set of time series simultaneously. Two such global averaging strategies are the DTW Barycenter Averaging (DBA) proposed by Petitjean *et al.* [11] and Soft-DTW Barycenter proposed by Cuturi *et al.* [2].

Euclidean distance is generally considered the simplest distance between two series and is simply calculated by summing the point to point distances along the time series. However, Euclidean distance has many drawbacks, in particular when there is an offset in the time series or when we have variable length time series. Hence other similarity measures to deal with time series data have been proposed to improve the overall robustness [1]. One such similarity measure which will be focused on in this paper is Dynamic Time Warping (DTW) [13]. DTW allows for a mapping of the time series in a non-linear way and works to find the optimal alignment between both series. Euclidean mapping is generally a one-to-one mapping between two curves whereas DTW can be considered as a one-to-many mapping. This is illustrated in Figure 2 where it can be seen that DTW aligns the series while Euclidean simply maps based on the time point.



(a)



(b)

Fig. 2: a) Euclidean mapping and b) DTW mapping on part of a lunge exercise. These images were generated by modifying existing code for use with our data[4].

DTW Barycenter averaging (DBA) and Soft-DTW barycenter are two averaging strategies that use the DTW metric and these are described in detail below:

DTW Barycenter Averaging (DBA) computes the optimal average sequence within a group of series in DTW space whereby the sum of squared DTW between the average sequence and the group of series considered is minimised [11]. So computing the DTW barycenter of a set of time series D is the optimisation problem outlined in Equation 1.

$$\min_{\mu} \sum_{x \in D} DTW(\mu, x)^2 \tag{1}$$

Where x corresponds to a series belonging to the set D and μ is a candidate

[4] $https://github.com/e{-}alizadeh/medium/blob/master/notebooks/intro\_to\_dtw.ipynb$

barycenter. In this optimisation, the DTW distance between each time series and a temporary average sequence (candidate barycenter) is calculated and the association between the coordinates of the average sequence and the time series are sought. This is then used to update the temporary average sequence until the optimal average sequence is found [14].

Soft-DTW barycenter computes the average sequence within a group of series whereby the weighted sum of the Soft-DTW distance between the average sequence and the group of series is minimised. The weights can be different for each sequence in the set but normalised to one [2]. The Soft-DTW approach is an extension of the DBA method where the min operator is replaced by the soft-min. This gives the advantage of being differentiable with respect to all of its inputs. Furthermore, where DTW alone only considers the optimal alignment, Soft-DTW considers all possible alignments. Soft-min can be computed as shown by Equation 1 [16].

$$\text{softmin}_\gamma(a_1, ..., a_n) = -\gamma \log \sum_i e^{-a_i/\gamma} \tag{2}$$

γ acts as the smoothing parameter and hence as $\gamma \to 0$, the result gets closer to that of DTW.

3 Preliminary Evaluation

To evaluate the suitability of using barycenters to aggregate multi-rep exercise data, the ability of the barycenters to preserve key features and be representative of the entire set of the lunge data was analysed. Signals streamed from the wide range tri-axial accelerometer ($\pm 16g$) at a sampling rate of 102.5 Hz were used in this study. The total acceleration was derived as the magnitude of the acceleration independent of the direction as shown in Equation 3.

$$a_{total} = \sqrt{a_x^2 + a_y^2 + a_z^2} \tag{3}$$

The signals were passed through a low-pass Butterworth filter with order of 6 and cut-off frequency of 10 Hz to remove the high frequency noise and ensure that only the signals corresponding to the movement due to the lunge was considered. The series were all segmented prior to the analysis through an automated algorithm that searches for the key features in the lunge. The analysis used the entire time series from initiation to termination and hence all the lunges are aligned based on the initiation point. Hence any misalignments between sets of lunges will be minor and mostly due to varying tempos at which each lunge is performed. This is of importance as DTW is known to be sensitive to misalignments at the start or end of the time series. Although not within the scope of this paper, in cases where it is not possible to align the start point of the time series, psi-DTW which allows relaxation of endpoint can be considered [15]. The

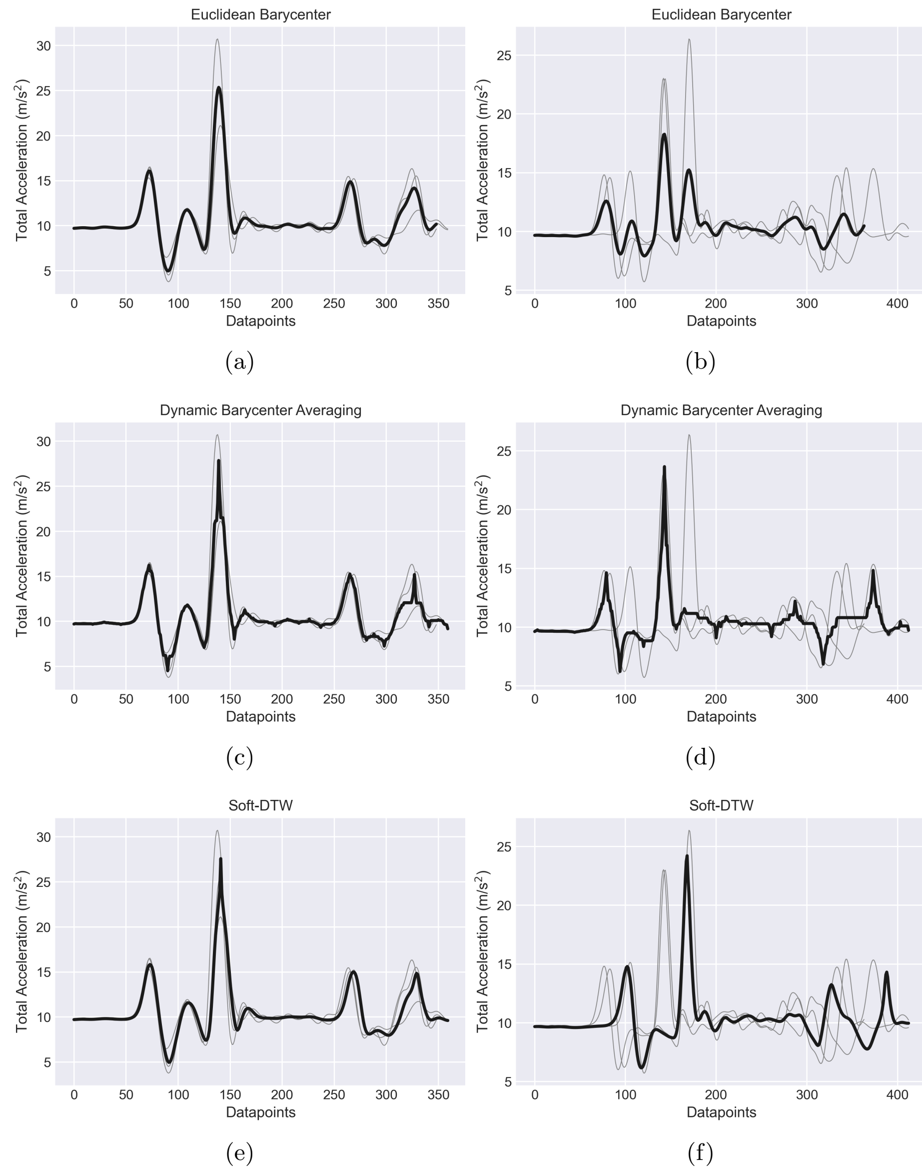


Fig. 3: Three repetitions of the lunge and the corresponding barycenter. Euclidean barycenter on a (a) well aligned set and (b) not well aligned set; DBA on (c) well aligned set and (d) not well aligned set; Soft-DTW barycenter on (e) well aligned set and (f) not well aligned set.

barycenters were all computed using the tslearn python package which provides machine learning tools for working with time series [16]. Soft-DTW barycenter, Euclidean averaging, and DBA were applied on the lunge data of 25 participants. The Soft-DTW was computed using a γ value of 10 and max iteration of 100. DBA was also computed with a max iteration of 100 and the default was used for all other parameters. These settings were kept constant throughout all the analysis presented in this paper.

3.1 Do Barycenters Preserve Key Features?

In order for barycenters to be useful in aggregating exercise data, the key features of the exercise must be preserved. When dealing with data that is time series in nature, lags in the time can make it difficult to compare repetitions of an exercise and hence difficult to get an aggregate measure. In Figure 3(a,c,e) we can see that when the time series are all nicely aligned, all the methods do a reasonable job of capturing the underlying trend. However once misalignments in the time series occur, the Euclidean barycenter smears out the key features and the result is not representative of the underlying data.

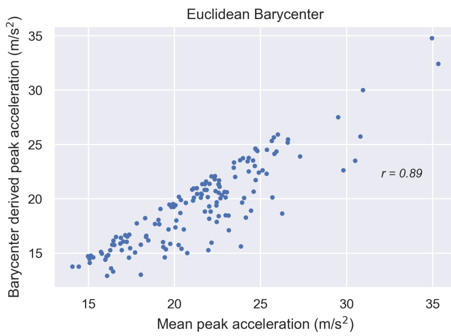
One point of interest for clinicians when analysing the total acceleration during the lunge is the peak acceleration of impact (IC1 in Figure 1). From Figure 3(b), it can be seen that Euclidean barycenter is not able to capture this peak whilst the other two techniques are able to preserve this peak. Additionally, the lunges considered were variable in length due to the differing tempos at which each lunge is performed. Euclidean barycenter does not work well with the variable length lunges as it only computes up till the end of the shortest lunge meaning that any part of the lunges after this point is not considered. Both DTW methods are able to take the variable lengths into consideration in their computation.

Although, DBA is able to capture the underlying pattern of the time series without the result being smeared out, spikes are seen in the series at certain points which are unrepresentative of the underlying trend. These spikes occur due to the algorithms used getting stuck in local minima. When using the Soft-DTW, even by using very small γ values which indicates the computation to be similar to that from DTW, the use of the softmin leads to the local minima being smoothed out and hence, although similar to the results from DBA, the barycenter will be much smoother and acts as a better representation of the underlying time series. A smoother time series representation is in particular ideal for visualisation purposes and will be more meaningful for the physio or end user.

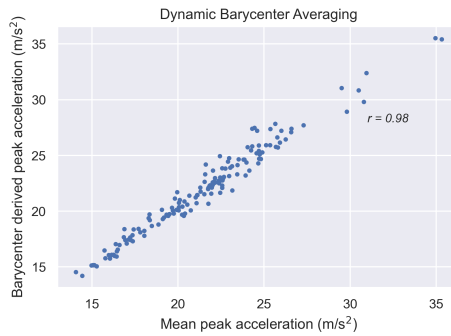
A comparison of the different methods indicates that the Soft-DTW provides the best visualisation of the underlying trends in the time series data of the lunge.

3.2 Using Barycenters to Represent Lunge Sets

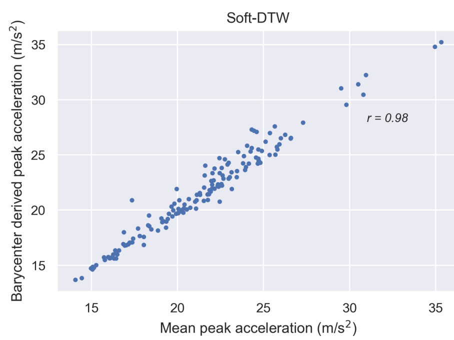
Having an aggregate measure of a set of lunges, helps to visualise the data and facilitates comparisons across sets. The last three lunges of each set of five repetitions were used in this analysis. This is a common practice while working with multiple repetitions of data to avoid any learning effects there may be in the first two lunges.



(a)



(b)



(c)

Fig. 4: Correlation plot between the mean peak acceleration from each three lunge set vs the peak acceleration from the corresponding barycenter for (a) Euclidean barycenter (b) DBA and (c) Soft-DTW barycenter

To test the effectiveness of the barycenter measure at an individual level, the correlation between the peak acceleration obtained from each set of three lunges and its corresponding barycenter were analysed. The mean peak for the three lunges was computed simply by averaging the three peak accelerations. Figure 4 (b) and (c) show that using a DTW measure preserves the peak acceleration across the sets of lunges with both having a Pearson's correlation coefficient of 0.98. The Euclidean barycenter (Figure 4 (c)) is also able to preserve some of the peaks very well whilst completely underestimating the peak for other sets of lunges and this can be attributed to how well the lunges are aligned to start with. As Soft-DTW shows the ability to preserve key features of the lunge and is representative of individual lunges sets, it is used as the choice of barycenter for all subsequent analysis.

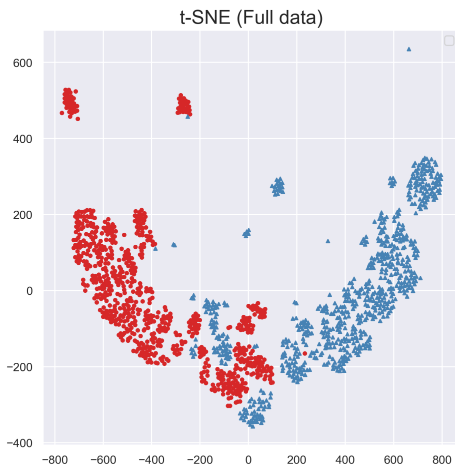
3.3 Using Barycenters for further analysis

Data collected from multiple-rep exercises are often used for data mining tasks such as classification or clustering. Multiple-rep data can be noisy where one 'bad' repetition of the exercise can affect the quality of the data. Aggregating the data can reduce this noise and hence barycenters may help give better results for data mining tasks. This hypothesis is tested by comparing how well the classes can be separated using the full data versus using Soft-DTW barycenter averaging.

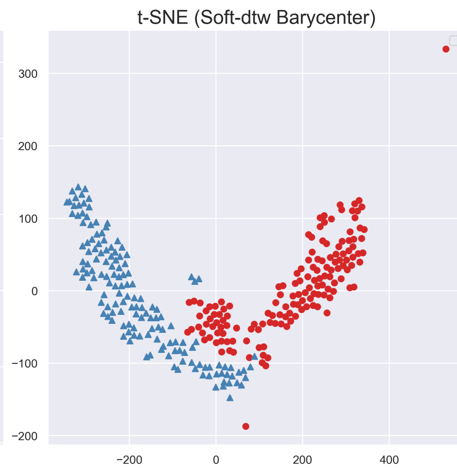
To visualise this capability to reduce noise, a t-distributed Stochastic Neighbour Embedding (t-SNE) [5] was used on the lunge data to reduce the 3D time series data into a 2D map. t-SNE is non-linear technique for dimensionality reduction and is generally used as a visualisation technique to represent high-dimensional datasets. For this analysis, the task of identifying left leg led vs right leg led lunges was considered. Figure 5 shows a t-SNE plot of the lunges with the two colours representing the different classes (Left leg vs Right leg). Each point in the plot represents a time series of the respective class. The Soft-DTW and DBA plots show the same data as in the full-data t-SNE plot but represented as aggregated barycenters. PCA initialization was used as it is known to be more globally stable than random initialization. It is clear that the DBA and Soft-DTW barycenters (Figure 5b,c) provide a clearer separation between the two classes than is possible when working with the full data (Figure 5a). This better separation will facilitate clustering. To test the goodness of the clusters of the two classes, the silhouette score was calculated [12]. The silhouette score measures the goodness of fit of clusters where the closer the score is to 1, the better separation there is between clusters and hence the classes should be more easily distinguishable. The full data from the t-SNE plot had a Silhouette score of 0.470 while the Soft-DTW barycenter had a Silhouette score of 0.492 and the DBA had a Silhouette score of 0.543. This suggests better separation when using the dtw techniques with DBA giving the best separation for this set of data.

4 Conclusions and Future Work

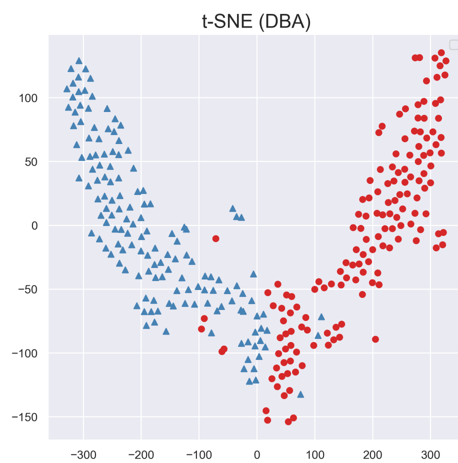
Preliminary evaluation shows the Soft-DTW barycenter to be an effective way of aggregating the time series from repetitions of an exercise into an 'averaged' series. The key features of multiple reps of the lunge were preserved in the Soft-DTW barycenters whereas the other approaches attempted did not perform as well in providing an accurate visualisation of the average. Particularly, the Soft-DTW approach performed well where there were phase shifts in the data. Furthermore, barycenter averaging shows potential for use in data mining tasks by allowing better separation between the classes.



(a)



(b)



(c)

Fig. 5: t-SNE plotted for (a) full data and (b) Soft-DTW Barycenter where the two colours represent two classes

The use of barycenters for the aggregation of time series can be useful in healthcare and sports studies where repeated trials are made as it helps to easily visualise the overall trend and better steer clinical decision making. This work can be extended for data reduction purposes, which could prove useful in clustering or classification tasks. Hence, as future work, an analysis of how classification/clustering performance is impacted by barycenter aggregation would be beneficial.

Acknowledgements

This work has emanated from research conducted with the financial support of Science Foundation Ireland under the Grant numbers 18/CRT/6183 and 12/RC/2289_P2. For the purpose of Open Access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

References

1. Cassisi, C., Montalto, P., Aliotta, M., Cannata, A., Pulvirenti, A.: Similarity measures and dimensionality reduction techniques for time series data mining (2012). https://doi.org/10.5772/49941, https://doi.org/10.5772/49941
2. Cuturi, M., Blondel, M.: Soft-DTW: A differentiable loss function for time-series. 34th International Conference on Machine Learning, ICML 2017 **2**, 1483–1505 (2017)
3. Garrison, M., Westrick, R., Johnson, M.R., Benenson, J.: Association between the functional movement screen and injury development in college athletes. International journal of sports physical therapy **10**(1), 21 (2015)
4. Johnston, W., O'Reilly, M., Coughlan, G., Caulfield, B.: Inertial Sensor Technology Can Capture Changes in Dynamic Balance Control during the Y Balance Test. Digital Biomarkers **1**(2), 106–117 (2018). https://doi.org/10.1159/000485470
5. Van der Maaten, L., Hinton, G.: Visualizing data using t-sne. Journal of machine learning research **9**(11) (2008)
6. Muro-De-La-Herran, A., Garcia-Zapirain, B., Mendez-Zorrilla, A.: Gait analysis methods: An overview of wearable and non-wearable systems, highlighting clinical applications. Sensors **14**(2), 3362–3394 (2014)
7. Okawa, M.: Template matching using time-series averaging and dtw with dependent warping for online signature verification. IEEE Access **7**, 81010–81019 (2019). https://doi.org/10.1109/ACCESS.2019.2923093
8. O'Reilly, M.A., Whelan, D.F., Ward, T.E., Delahunt, E., Caulfield, B.: Classification of lunge biomechanics with multiple and individual inertial measurement units. Sports Biomechanics **16**(3), 342–360 (2017). https://doi.org/10.1080/14763141.2017.1314544, http://doi.org/10.1080/14763141.2017.1314544
9. O'Reilly, M., Caulfield, B., Ward, T., Johnston, W., Doherty, C.: Wearable inertial sensor systems for lower limb exercise detection and evaluation: a systematic review. Sports Medicine **48**(5), 1221–1246 (2018)
10. Petitjean, F., Forestier, G., Webb, G.I., Nicholson, A.E., Chen, Y., Keogh, E.: Dynamic time warping averaging of time series allows faster and more accurate classification. In: 2014 IEEE international conference on data mining. pp. 470–479. IEEE (2014)
11. Petitjean, F., Ketterlin, A., Gançarski, P.: A global averaging method for dynamic time warping, with applications to clustering. Pattern Recognition **44**(3), 678–693 (2011). https://doi.org/10.1016/j.patcog.2010.09.013
12. Rousseeuw, P.J.: Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics **20**, 53–65 (1987). https://doi.org/https://doi.org/10.1016/0377-0427(87)90125-7, https://www.sciencedirect.com/science/article/pii/0377042787901257

13. Sakoe, H., Chiba, S.: Dynamic Programming Algorithm Optimization for Spoken Word Recognition. IEEE Transactions on Acoustics, Speech, and Signal Processing **26**(1), 43–49 (1978). https://doi.org/10.1109/TASSP.1978.1163055
14. Shi, K., Qin, H., Sima, C., Li, S., Shen, L., Ma, Q.: Dynamic Barycenter Averaging Kernel in RBF Networks for Time Series Classification. IEEE Access **7**, 47564–47576 (2019). https://doi.org/10.1109/ACCESS.2019.2910017
15. Silva, D.F., Batista, G.E.A.P.A., Keogh, E.: On the Effect of Endpoints on Dynamic Time Warping. SIGKDD MiLeTS'16 p. 10 (2016)
16. Tavenard, R., Faouzi, J., Vandewiele, G., Divo, F., Androz, G., Holtz, C., Payne, M., Yurchak, R., Rußwurm, M., Kolar, K., Woods, E.: tslearn: A machine learning toolkit dedicated to time-series data (2017), https://github.com/rtavenar/tslearn