ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

Multidimensional CDTW-based features for Parkinson's Disease classification

Ferhat Attal, Nicolas Khoury and Yacine Amirat

LISSI Laboratory, Paris-Est Creteil University, France

Abstract. This paper presents an improvement of the Unidimensional Continuous Dynamic Time Warping (UCDTW) method for diagnosing Parkinson's Disease (PD) based on multidimensional time series data. These data include recordings of vertical Ground Reaction Forces (vGRFs) collected from eight force sensors per shoe sole during the walk. Leveraging gait cycle patterns, the proposed approach distinguishes between healthy and PD subjects by assessing gait cycle repetition through Multidimensional CDTW. Several classification methods, including supervised (K-NN, DT, RF, SVM) and unsupervised (GMM, K-means), are used to classify the healthy and PD subjects, using MCDTW distances extracted from the gait cycles. The obtained results show a significant improvement in terms of classification performances when using MCDTW-based features compared to unidimensional ones.

1 Introduction

Parkinson's disease (PD) is a chronic neurodegenerative disorder affecting the human central nervous system. The Parkinson's disease diagnosis is a difficult, subjective task, mainly in the early stages, and there is no available biomarker or specific test for such a diagnosis. The PD is responsible of destroying the dopaminergic neurons which produce neurotransmitters known as dopamine. These neurotransmitters play an important role in the transmission of brain signals in order to control human balance and movement [1, 2, 3]. Thus, most PD patients suffer from movement disorders affecting their walking ability [1, 2]. Tremors, stiff muscles, and changes in walking gait pattern are among the symptoms that can be observed in patients with PD [2]. A gait cycle consists of two phases: stance phase and swing phase, representing 60 % and 40 %, respectively, of the gait pattern. For healthy persons, the gait represents a cyclical and repetitive activity in which one stride (gait cycle) follows the other one continuously. However, the PD subjects often present significant variations in their gait patterns from one cycle to another [5]. In this work, to assess the repetition of gait cycles, the similarity of time-series data corresponding to stance phases is characterized. MCDTW, an enhanced version of UCDTW technique proposed in [7], is proposed to estimate the similarity between time-series during the stance phases. The MCDTW distances, extracted from gait cycles, serve as classifier inputs to discriminate between healthy subjects and those with PD. To evaluate the proposed approach, three online sub-datasets including gait data collected from 72 healthy subjects and 93 PD subjects were used. These datasets were provided by Yogev et al. [8] (29 PD subjects and

ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

18 healthy ones), Hausdorff et al. [9] (29 PD and 25 healthy subjects) and Frenkel-Toledo et al. [10] (35 PD subjects and 29 healthy ones). The gait data consist of vertical Ground Reaction Forces (vGRFs) recordings collected from eight force sensors embedded in each shoe sole worn by the subjects, resulting in sixteen signals per recording. In this paper, the time-series corresponding to the sixteen force signals provided by the vGRFs sensors as well as the mean of the outputs from the 8 sensors of each foot are considered to characterize the gait pattern of each subject. This paper is structured as follows: Section 2 outlines the proposed approach. Section 3 presents and discusses the performance of this approach. Finally, the paper concludes with a summary and outlines some future perspectives in the last section.

2 Proposed approach

In our previous work [7], unidimensional CDTW based features were used to assess the gait cycle similarity using unidimensional time-series data. Only the mean of the eight force sensors placed under each foot are considered. For a more accurate measure of similarity between stance phases, it's crucial to consider all the data characterizing the gait cycle. For this, multidimensional Continuous Dynamic Time Warping (MCDTW) based features are proposed. The objective is to enhance discrimination between healthy subjects and those with PD. In this paper, for each foot, we consider the eight force signals provided by the vGRF sensors, as well as their mean value, resulting in a total of nine signals per foot.

First, the standard multidimensional DTW will be reviewed before presenting MCDTW. Formally, let $U = (u_1, u_2, ..., u_l, ..., u_d)$ and $V = (v_1, v_2, ..., v_l, ..., v_d)$ be two multidimensional time-series, where $u_l = (u_l(1), u_l(2), ..., u_l(i), ..., u_l(m))$ and $v_l = (v_l(1), v_l(2), ..., v_l(j), ..., v_l(n))$ represent the l^{th} unidimensional time-series. d represents the number of signals per foot, $i = 1, 2, ..., m$, $j = 1, 2, ..., n$ where m and n represent respectively the lengths of the time series U and V that could be represented under the following matrices.

$$U = \begin{bmatrix} u_1(1) & u_1(2) & ... & u_1(i) & ... & u_1(m) \\ u_2(1) & u_2(2) & ... & u_2(i) & ... & u_2(m) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u_l(1) & u_l(2) & ... & u_l(i) & ... & u_l(m) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u_d(1) & u_d(2) & ... & u_d(i) & ... & u_d(m) \end{bmatrix} \quad V = \begin{bmatrix} v_1(1) & v_1(2) & ... & v_1(j) & ... & v_1(n) \\ v_2(1) & v_2(2) & ... & v_2(j) & ... & v_2(n) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ v_l(1) & v_l(2) & ... & v_l(j) & ... & v_l(n) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ v_d(1) & v_d(2) & ... & v_d(j) & ... & v_d(n) \end{bmatrix}$$

The local distance $d_{\text{local}}(U(i), V(j))$ between the elements of the time-series in the multidimensional domain is computed using the squared Euclidean distance, expressed as:

$$d_{local}(U(i), V(j)) = \sum_{l=1}^{d} (u_l(i) - v_l(j))^2 \tag{1}$$

ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

Where $U(i)$ and $V(j)$ represent the i^{th} and the j^{th} columns of U and V respectively. Then the optimal alignment path between the two time series U and V corresponds to the minimum distance warping path that is calculated based on the $d_{local}(U(i), V(j))$.

In the context of multidimensional data, the cumulative distance can be formulated as follows:

$$\mathcal{D}(i,j) = d_{local}(U(i), V(j)) + min \begin{cases} \mathcal{D}(i-1, j-1) \\ \mathcal{D}(i-1, j) \\ \mathcal{D}(i, j-1) \end{cases} \tag{2}$$

Where $\mathcal{D}(1,1) = d_{local}(U(1), V(1))$ represents the initial condition.

The optimal warping path $\mathcal{W} = (w_u(k), w_v(k))$ can be calculated using the recursion equation 2. It corresponds to the minimum sum of distance from $(1,1)$ to (m,n), which indicates how the two multidimensional time series U and V stretch or shrink along the time axis. In the following, two new multivariate time-series $\overline{U}$ and $\overline{V}$ are calculated as follows:

$$\begin{cases} \overline{U}(k) = U(w_u(k)) \\ \overline{V}(k) = V(w_v(k)) \end{cases} , k = 1, 2, ..., p \tag{3}$$

Where $w_u(k)$ and $w_v(k)$ are the indexes in time-series U and V respectively, and p is the length of the warping path. $(w_u(k);w_v(k))$ indicates that the $w_u(k)$ th element in time series U corresponds to the $w_v(k)$ th element in time series V (mapping between $w_u(k)$ th and $w_v(k)$ th).

Using the warping path $\mathcal{W}$, the original time-series U and V will be mapped to $\overline{U}$ and $\overline{V}$. Thus, the multidimensional DTW distance measure can be expressed as follows :

$$\begin{aligned} DTW(U,V) = \sum_{k=1}^{p} d_{local}(U(w_u(k)), V(w_v(k))) &= \sum_{k=1}^{p} d_{local}(\overline{U}(k), \overline{V}(k)) \\ &= \sum_{k=1}^{p} (\overline{U}(k) - \overline{V}(k))^2 \end{aligned} \tag{4}$$

The MCDTW can be formulated by considering the process of mapping used in multidimensional DTW in the continuous domain. As in the case of UCDTW, intermediate points are added in the time-series using linear interpolation. The MCDTW differs from the standard MDTW by the fact that a sample point in one of the time-series could match a point in-between two samples in the other time-series. Thus, the warping path $\mathcal{W}$ can take non-integer values. For example, the sample point $v_l(j)$ can be mapped to a new point which is obtained using an orthogonal projection of $v_l(j)$ onto the segment $[u_l(i); u_l(i+1)]$.

Using the orthogonal projection, denoted as $v_l(f)$, of sample point $u_l(i)$ of the time series U onto the segment $[v_l(j-1); v_l(j)]$, and conversely the orthogonal projection , denoted as $u_l(e)$, of sample point $v_l(j)$ of the time-series V onto the

ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

segment $[u_l(i-1); u_l(i)]$, the coordinates of the intermediate matching points $v_l(f)$, $u_l(e)$ can be expressed as follows:

$$u \begin{cases} e = (i-1) + r_u.\dfrac{\triangle_i}{\triangle_1} \\ u_l(e) = u_l(i-1) + r_u.\dfrac{\triangle_u}{\triangle_1} \end{cases} \quad with \quad \begin{array}{l} \triangle_i = (i) - (i-1) = 1 \\ \triangle_u = u_l(i) - u_l(i-1) \\ \triangle_1 = \sqrt{\triangle_i^2 + \triangle_u^2} \\ r_u = \dfrac{(j-(i-1))}{\triangle_1} + \dfrac{(v_l(j) - u_l(i-1)).\triangle_u}{\triangle_1} \end{array}$$

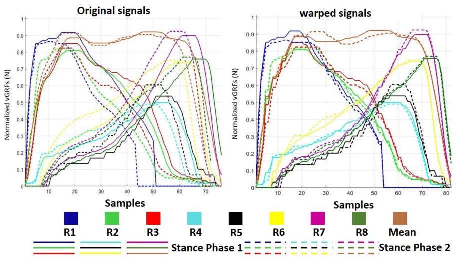
$$v \begin{cases} f = (j-1) + r_v.\dfrac{\triangle_j}{\triangle_2} \\ v_l(f) = v_l(j-1) + r_v.\dfrac{\triangle_v}{\triangle_2} \end{cases} \quad with \quad \begin{array}{l} \triangle_j = (j) - (j-1) = 1 \\ \triangle_v = v_l(j) - v_l(j-1) \\ \triangle_2 = \sqrt{\triangle_j^2 + \triangle_v^2} \\ r_v = \dfrac{(i-(j-1))}{\triangle_2} + \dfrac{(u_l(i) - v_l(j-1)).\triangle_v}{\triangle_2} \end{array}$$

3 Results and discussion

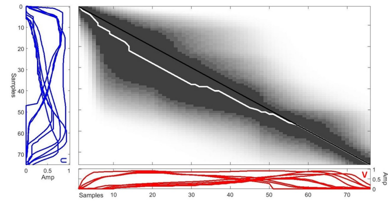
Figures 1(a) and 1(c) present the matching process between multidimensional time-series during the stance phases of the right foot for a healthy an a PD subject, respectively. In the case of healthy subject, it can be observed that there is a matching between the consecutive stance phases as shown in figure 1(a). However, in the case of PD subject, the matching between the consecutive stance phases is less important as shown in figure 1(c). These observations are better illustrated in figures 1(b) and 1(d), which present the optimal warping paths for both a healthy and a PD subject, respectively. The warping path obtained for the PD subject illustrates a dissimilarity between consecutive stance phases compared to that observed in the case of healthy subject. The black lines represent the 'optimal' warping paths in the case of identical stance phases. For a healthy subject, the 'optimal' warping path closely aligns with optimal one (figure 1(b)). Conversely, for a PD subject, the 'optimal' warping path is highly distorted and deviates significantly from the optimal one (figure 1(d)).

In this paper, only the stance phases are considered as the vGRFs values in swing phases (foot up) are equal to zero. In order to estimate the similarity of gait cycles, the MCDTW technique is applied on successive stance phases extracted from gait cycles of each foot. Thus, Two MCDTW distance vectors characterizing this similarity are estimated. Mean and std of the MCDTW distance vectors are considered leading to four features namely Mean and std of the MCDTW distance of the left and right foot. These feature are then used as classifier inputs. In order to maximize the classifier performances in terms of PD classification, parameters tuning step was carried out. In this study, finding parameter settings is conducted similarly as in [7], using a grid search method. The performances of the classifiers are evaluated using a 10-fold cross-validation technique. Table 1 shows the classifiers performances in terms of accuracy and its standard deviation obtained using MCDTW and UCDTW features. A comparison of the classifiers performances in the case of MCDTW, shows that k-NN and SVM allow achieving high accuracy rates (higher than 98%) with a small std (less than 1.02%). In the case of Yogev et al and Frenkel-

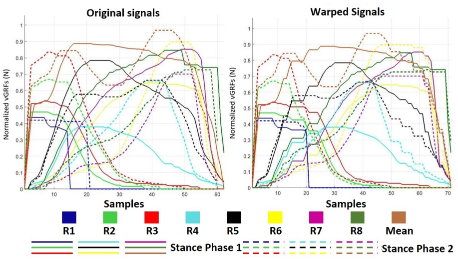
ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.



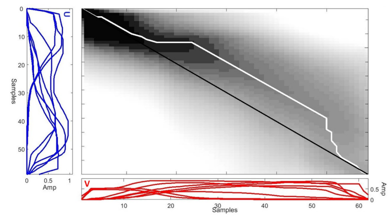
(a) Healthy matching Stance Phases



(b) Healthy Optimal path



(c) PD matching Stance Phases



(d) PD Optimal path

Fig. 1: Results obtained with consecutive stance phases of right foot using MCDTW. $R_1, \ldots, R_8$ represent the vGRF signals.

Toledo sub-datasts the SVM outperforms the other classifier methods followed by KNN, RF and DT. For Hausdorff sub dataset, KNN outperforms the other classifier methods followed by SVM, RF and DT. By comparing supervised and unsupervised methods, it can be noted that supervised ones outperform largely the unsupervised methods. It is worth noting that in the case of unsupervised methods, the GMM methods outperforms the K-means one in the case of the tree sub-datasets.

		Supervised				Unsupervised	
	Dataset.	**k-NN**	**CART**	**RF**	**SVM**	**k-Means**	**GMM**
MCDTW	Yogev et al.	99.76 ± 0.5 %	94.24 ± 2.66 %	98.71 ± 1.18 %	**99.88 ± 0.38 %**	75.81 ± 7.88 %	77.79 ± 12.92 %
	Hausdorff et al.	**99.92 ± 0.26 %**	93.39 ± 2.08 %	96.03 ± 1.73 %	99.75 ± 0.4 %	69.42 ± 3.41 %	78.93 ± 1.31 %
	Frenkel-Toledo et al.	98.59 ± 1.02 %	92.81± 3.44 %	94.84 ± 2.06 %	**99.84 ± 0.49%**	73.75 ± 3.57 %	76.88 ± 7.72 %
UCDTW [7]	Yogev et al.	92.88 ± 2.03 %	82.96 ± 3.99 %	89.35 ± 2.62 %	93.57 ± 2.61 %	67.79 ± 1.10 %	65.58 ± 7.23 %
	Hausdorff et al.	97.52 ± 1.04 %	88.82 ± 3.09 %	90.02 ± 1.98 %	95.03 ± 1.85 %	65.29 ± 2.13 %	73.97± 4.90 %
	Frenkel-Toledo et al.	86.02 ± 3.09 %	80.02 ± 6.44 %	82.59 ± 4.92 %	87.32 ± 2.99 %	60.47 ± 1.48 %	67.19 ± 8.65 %

Table 1: Accuracy and its std obtained using proposed MCDTW and UCDTW based features.

By comparing the results obtained using MCDTW and UCDTW, it can be

ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

observed that the use of MCDTW gait features allow achieving best performances. The accuracy rate improvement obtained in the case of supervised classifiers varies from 2% to 12%. It can be also noticed a significant reduction of std measure. Same observation can be made also in the case of unsupervised methods with an accuracy rate improvement ranging from 4% to 13%.

4 Conclusion

This paper presents an enhancement of the unidimensional CDTW method for Parkinson's disease (PD) classification using multidimensional time series data. This approach exploits the repetitive patterns of human walking to discriminate between healthy and PD subjects. Healthy subjects present a repetitive gait cycles, whereas those with PD show dissimilarities from one cycle to another. To assess these dissimilarities, a multidimensional CDTW, an extension of the unidimensional CDTW, is proposed. The obtained results showed a significant classification accuracy improvements when using multidimensional CDTW-based features compared to unidimensional ones. Ongoing research focuses on exploring alternative models to the linear interpolation used in the MCDTW formulation to achieve better assessment of time-series similarity and data modelling.

References

[1] Davie C.A. review of Parkinson's disease. British medical bulletin, 2008, Volume 86, No. 1, pp. 109-127.

[2] Parkinson's disease Foundation, Available online:(`http://parkinson.org/understanding-parkinsons`).

[3] Ahlrichs C.; Lawo M. Parkinson's disease motor symptoms in machine learning: a review. Health Informatics-An International Journal (HIIJ), 2013, Volume 2, No. 4.

[4] Das, R.A. Comparison of multiple classification methods for diagnosis of Parkinson disease. *Expert Syst. Appl.* 2010, *37*, 1568–1572.

[5] Y. Nancy Jane, H. Khanna Nehemiah and K. Arputharaj. A Q-backpropagated time delay neural network for diagnosing severity of gait disturbances in Parkinson's disease. *Journal of biomedical informatics*, 60(C): 169-176, 2016.

[6] N. Khoury , F. Attal, Y. Amirat, L. Oukhellou, S. Mohammed. Data-Driven Based Approach to Aid Parkinson's Disease Diagnosis. *Sensors*, 2019, Volume 19, No. 2, pp. 242.

[7] N. Khoury , F. Attal, Y. Amirat, A. Chibani , S. Mohammed. CDTW-based classification for Parkinson's Disease diagnosis. In Proceedings of the 26th *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning* (ESANN), Bruges, Belgium, 25–27 April 2018; pp. 621–626.

[8] G. Yogev, N. Giladi, C. Peretz, S. Springer, E. S. Simon, and J.M. Hausdorff. Dual tasking, gait rhythmicity, and Parkinson's disease: which aspects of gait are attention demanding?, *European journal of neuroscience*, Volume 22(5): 1248-1256, 2005.

[9] J.M. Hausdorff, J. Lowenthal, T. Herman, L. Gruendlinger, C. Peretz, N. Giladi. Rhythmic auditory stimulation modulates gait variability in Parkinson's disease, *European Journal of Neuroscience*, 26(8): 2369-2375, 2007.

[10] S. Frenkel-Toledo, N. Giladi, C. Peretz, T. Herman, L. Gruendlinger, J.M. Hausdorff. Treadmill walking as an external pacemaker to improve gait rhythm and stability in Parkinson's disease, *Movement Disorders*, 20(9): 1109-1114, 2005.