



REVUE ALGERIENNE DES SCIENCES

Section -A-

Sciences de la Nature et de la Vie

Sciences techniques



Accelerating Quasi Periodic Dynamic Time Warping For Similarity Search in ECG Time Series

Imen Boulnemour*, Bachir Boucheham a*, Abdelmadjid Lahreche a*

^aUniversité 20 août 1955 de Skikda, Laboratoire d'Electronique de Skikda (LRES), Faculté des Science, Département d'Informatique, BP 26, Skikda, DZ21000, Algérie.

Informations	Abstract
KEYWORDS: Time Series Similarity search QP-DTW *Correspondence : Boulnemourimen@hotmail.fr Reçu, Révisé, Accepté.	Several methods have been proposed for similarity search in time series (TS). Our previously published method QP-DTW (Quasi Periodic Dynamic Time Warping) has been proposed as a rival method to DTW for quasi periodic TS alignment. However the method is expensive in terms of time complexity. Consequently it is not appropriate for similarity search in TS, where the methods with less time complexity are needed. In this paper we propose an accelerated version of the QP-DTW method and we use it to solve the problem of searching similar ECG (Electrocardiogram) time series in a data base. We have focused on ECG TS for its utility in various medical domains, including Diagnostic of cardiac disease and biometry. The proposed Accelerated QP-DTW (A-QPDTW) used Haar Wavelet Transform as a technique to reduce the dimensionality of the TS and the KNN (K nearest neighbor) approach to search for TS that are similar to a given query. Experimental results show that the proposed method is more efficient than QPDTW method with comparative accuracy. All results have been obtained using the ECG TS of the UCR (University of California at Riverside) time series Dataset, universally admitted as the first Benchmark in time series analyses.

1. Introduction

There are many phenomena that are represented in time series forms (TS) such as the values of stock exchange (companies shares), prices of goods and services, the ECG signal (Electrocardiogram), the traffic density in a computer network, multimedia data, etc..

A time series is a collection of data obtained sequentially in time. Formally, a time series is a continuation of couples $\langle (v_1, t_1), (v_2, t_2), \dots, (v_i, t_i), \dots \rangle$ where v_i is a value or a vector of values taken at time index t_i . This notation can be abbreviated in $\langle v_1, v_2, \dots, v_i, \dots \rangle$ when the reference to the time does not need to be clarified [1].

Proximity or similarity searching is the problem of, given a data set and a similarity measure, finding the elements from the set that are close to a given query. This is a natural extension of the classical problem of exact searching [2].

It has many applications. Some examples are: speech recognition [3], handwriting

recognition [4], content-based image retrieval [5], slice music similarity search [6], etc.

Literature shows that the most used types of similarity search queries are: (a) range query and (b) nearest neighbors query. Range query consists in retrieval of all time series in a data base within a certain distance r from the query object, while k nearest neighbors query returns the k time series most similar to the query time series [7].

In both cases, the performance of similarity search depends highly on the distance measure used to compare the time series [8]. It is known that DTW (Dynamic Time Warping) is the most suitable measure for many applications of time series, including similarity search [9]. Chen et al, for example adapt DTW at Individual Query for Electrocardiogram Analysis in Cardiology [10].

In another context, our previously published method QP-DTW (Quasi Periodic Dynamic Time Warping) has been proposed as a rival method to DTW for ECG alignment. However the method is expensive in terms of time

complexity. Consequently it is not appropriate for similarity search in ECG TS, where methods with less time complexity are needed, particularly in large databases. In this paper we propose an accelerated version of the QP-DTW method and we use it to resolve the problem of searching similar ECG time series within a given database. We have focused on ECG for its utility in various domains including Diagnostic of cardiac diseases and biometry.

Diagnostic applications for people with risk of heart dysfunctions are more and more used nowadays. For instance, people with cardiac high risk factors bear wearable sensors in their daily lives. These sensors continuously monitor health conditions like ECG and send the data to Automatic Diagnostic Systems, where they are processed in quasi real-time diagnosis [11]. Regarding the field of biometry, it is suggested that ECG trace is unique to each person and it has been introduced as one of the most powerful biometric tool for personal identification [12], [13].

The proposed AQP-DTW (Accelerated Quasi Periodic Dynamic Time Warping) uses Discrete Wavelet Transform (DWT) as a technique to reduce the dimensionality of the TS. The obtained TS are fed to QP-DTW and finally, the KNN (K nearest neighbor) approach uses the AQP-DTW procedure to search for the K TS that are most similar to a given query. Experimental results show that the proposed method is more efficient than the QPDTW method in the context of TS similarity search for a comparable accuracy.

All results have been obtained using the ECG TS of the UCR (University of California at Riverside) time series Dataset, universally admitted as the first Benchmark in time series analyses.

The rest of this paper is organized as follow. Section 2 presents related works, while Section 3 describes our proposed method. The 4th section is devoted to application of our method to similarity search and to experiment tests. Finally, Section 5 concludes this paper and gives perspectives of our research.

2. Materials and methods

The search for similarity is limited because of the large size of the time series BDDs. The best solution, according to [14], would be to use a dimensionality reduction technique on the data such as DFT (Discrete Fourier Transform) [1], DWT (Discrete Wavelet Transform [15], PAA (Piecewise Aggregate Approximation) [14], SAX (Symbolic Aggregate approximation) [16] or to

use index structures. Echihabi et al. [17] compared these methods in terms of efficiency under a single, unified experimental framework. Their work indicates that there is no single best method that outperforms all the rest. Recently Gogolou et al. [18] provides progressive whole-matching similarity search results on large time series collections. They show that there is a gap between the time the 1st Nearest Neighbour (1-NN) is found and the time when the search algorithm terminates. In other words, users often wait without any improvement in their answers. They further show that high-quality approximate answers are found very early with their method. In our case, we will use the wavelet transform to accelerate the similarity search with QP-DTW and KNN methods.

2.1. Discrete Wavelet Transform (DWT)

DWT or Discrete Wavelet Transform [15] is a mathematical tool for signals and time series modeling based on a hierarchical decomposition function. Meanwhile, it is used as an effective characterization and dimensional reduction technique for similarity searching in time series.

The basic idea of DWT is similar to that used by Agrawal for the DFT technique [1] in the sense that a time series can be represented only by a discrete wavelet transform, but by taking only the first N coefficients. This way, we can reduce the dimensionality and keep a large part of the information within the original time series. DWT has several advantages such as its ability to search at different scales and its temporal complexity which is $O(n)$, compared to that of DFT which is $O(n^2)$.

2.2. Quasi-Periodic Dynamic Time Warping (QP-DTW)

Quasi-Periodic Dynamic Time Warping is a similarity measure proposed by Boulnemour & Boucheham [19]. It combines the SEA [20] and the DTW [21], [22] methods. It was especially designed for alignment of quasi-periodic time series and specifically for heartbeats time series provided from ECG (Electrocardiogram) records. It is argued that Quasi-periodic time series are very complex TS for they are concatenations of quasi-similar forms called periods (Cycles). Such that, the alignment of these series is not possible beyond the SEA method and its derivatives (including QP-DTW) [20]. It was demonstrated in [20] and [19], that DTW completely fails when it comes to align quasi-periodic time series containing each a different number of quasi similar and phase shifted periods.

Fig.1(a) shows the original traces X and Y. Fig.1(b) illustrates the alignment done by QP-

DTW for X and its reconstructed time series (left of Fig. 1(b)) and for the time series Y and its reconstructed time series (right of Fig. 1(b)). Fig. 1 (c) presents the alignment done by the DTW method for the time series X and Y. Note that the two traces are from the same time series but that they are of different lengths. They are also phase shifted because they are not taken at the same time (offset of 35 seconds for Y). Fig. 2, illustrate the QP-DTW diagram.

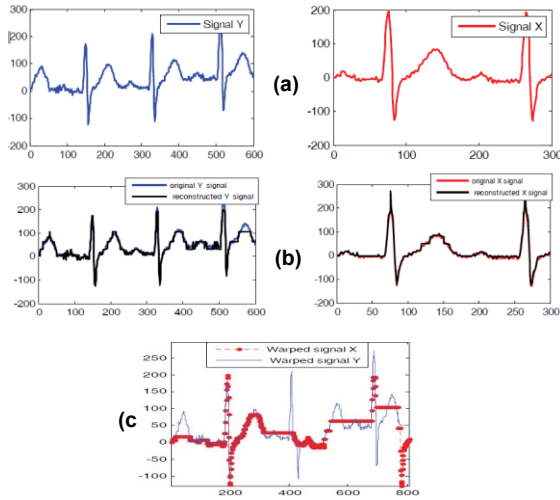


Fig.1. Comparison of the QP-DTW and the DTW methods for the case of quasi-periodic time series with phase shift (rec.215). (a) Original signals. (b) Original signals vs. reconstructed signals by QP-DTW (c) Original signal vs. warped signal by DTW [19].

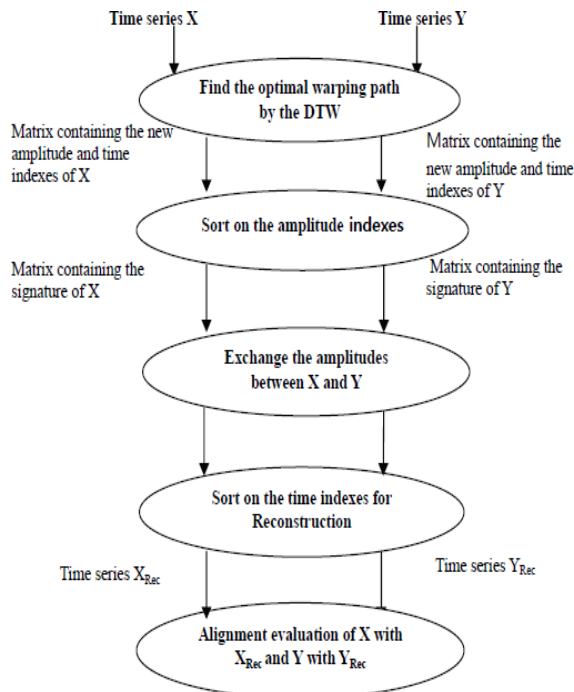


Fig.2. Illustrative diagram of the QP-DTW method.

3. Accelerated Quasi Periodic Dynamic Time Warping (A-QPDTW)

In this section, we describe our proposed method: Accelerated Quasi Periodic Dynamic Time Warping (A-QPDTW). A-QPDTW is mainly composed of two steps. It is described in Fig. 3.

A. Feature Extraction

In literature, there are several types of wavelet transforms. We have used the OneLevel Haar Wavelets Transforms due to its simplicity, efficiency (Linear order of complexity) and effectiveness in extracting the discriminate features and reducing the dimensionality of quasi-periodic time series (QP-TS).

The Haar wavelet transform is a series of matching and differentiation operations. In the matching operations, we obtain the approximation coefficients. In the differentiation operations, we get the detail coefficients. To have an idea of how the Haar wavelet transform works, let's consider the following 4-dimensional TS:

$X = [10, 6, 5, 7]$. We obtain the detail (cD) and the approximate coefficients (cA) as follows:

$$cA = \left\{ \frac{10+6}{2}, \frac{5+7}{2} \right\}, cD = \left\{ \frac{10-6}{2}, \frac{5-7}{2} \right\}$$

The one level Haar Wavelet transform of X using cA is then [2, -1] and using cD is [8, 6]. Fig. 3, shows the original TS and its reduced TS using approximation coefficients of Haar wavelets Transform. We see that, the length of the reduced TS is the half of the original TS.

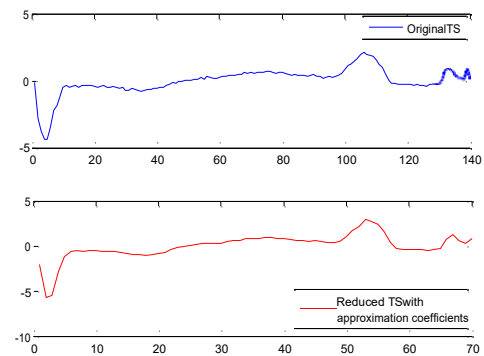


Fig.3. Original TS vs. Reduced TS with approximation coefficients.

B. Similarity Measure

The second step consists in applying the QP-DTW similarity measure on the reduced data and to search for the similar TS to a given query within a given database, using KNN

approach. Fig.4, describes the AQP-DTW general diagram.

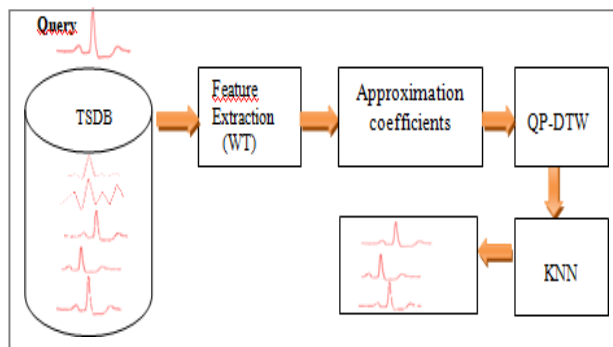


Fig.4. Illustrative diagram of the AQP-DTW method.

4. Experimentation

4.1. Experimental configuration

In this section, we evaluate the efficiency and the effectiveness of our algorithm on the problem of similarity search on ECG time series. The ECG time series dataset, for experiment are selected from the UCR Time Series Data Archive [Chen, 201523]. This data archive is admitted as a benchmark for testing and evaluating time series similarity measures. The UCR contains 7 ECG time series dataset and each one have a standard partitions of training and testing set. A description of the dataset used in our study is presented in Table 1.

Table 1. Description of the used ECG datasets

Dataset Name	Number of Classes	Size of Training Set	Size of Testing Set	Time Series Length
ECG200	02	100	100	96
ECGFiveDays	02	23	861	136
ECG5000	05	500	4500	140
CinC_ECG_torso	04	40	1380	1639
TwoLeadECG	02	23	1139	82
NonInvasiveFatalECGThorax1	42	1800	1965	1965
NonInvasiveFatalECGThorax2	42	1800	1965	750

All applications in this study have been performed on an experimental environment with the following characteristics: PC Intel(R) Core™ i7 CPU, 2.10 GHz, 4Go of main memory, OS Windows 7 (64 bits) and MATLAB 2014 environment.

4.2. Experimental results

First we start by evaluating the proposed method AQP-DTW, in terms of efficiency (run time) with regard to the QP-DTW, on the problem of similarity search. The execution time is obtained by calculating the average execution time of all the queries in the testing dataset. Fig.5, shows the execution time in seconds for both A-QPDTW and QP-DTW, on all ECG time series in UCR. We notice that, the method AQP-DTW is more efficient than the QP-DTW method. In fact, it is about three times faster than QP-DTW method in all datasets. Moreover, in the case of long time series, we show that our method is even faster. As an example, for the dataset CinC_ECG_torso, the time execution of the AQP-DTW is 1046, 20 seconds while it is 4228,75 seconds for the QP-DTW (four times faster). Another example, for the small dataset ECG200, the time execution of the AQP-DTW is 0, 37 seconds and it is 1,12 seconds for the QP-DTW.

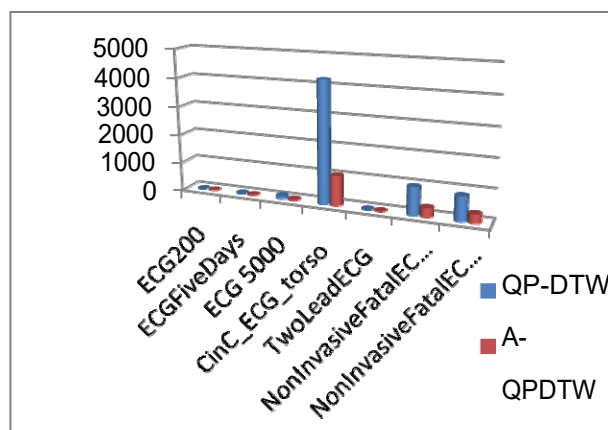


Fig.5. Execution times for AQP-DTW Vs QP-DTW.

On the other hand, we compare the precision of our method AQP-DTW with regard to the QP-DTW method within the problem of similarity search. In that respect, we mention that the precision criteria measures the fraction of relevant objects returned in a given query in relation to the total of returned objects [Garcia, 201424]. Fig.6, shows an example of a given query and the 20 returned similar TS using our proposed method. the parameter $k = 20$ has been chosen empirically.

We clearly see that all the returned TS have the same shape as the query shape.

Table 2 shows experimental results of the precision comparison for both A-QPDTW and QP-DTW on all ECG time series in UCR. Results show that the precisions of the two methods are comparable to each other. For the dataset, "CinC_ECG_torso", the precision of QP-DTW is 0.275 and the one of AQP-DTW is 0.273. For the dataset "NonInvasiveFatalECG_Thorax1", the precision of QP-DTW is 0.698 and the one of AQP-DTW is 0.693.

Indeed, AQP-DTW wins in two datasets and loses in five datasets with comparison to QP-DTW.

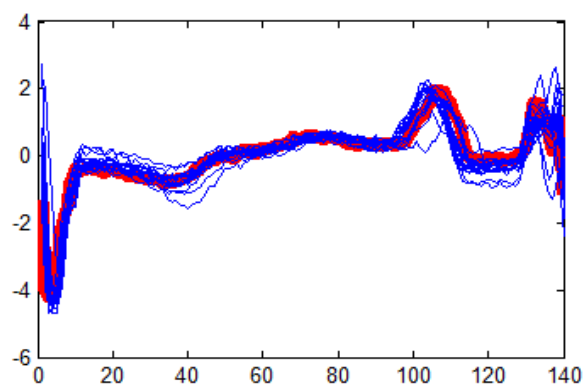


Fig.6. Precision of AQP-DTW method. The red TS is the query object and the blue TS are the 20 returned objects which are similar to the query ($k=20$).

Table 2. Precision of QPDTW Vs AQP-DTW

Data sets	QP-DTW	A-QPDTW
ECG200	0.693	0.632
ECGFiveDays	0.576	0.552
ECG 5000	0.720	0.728
CinC_ECG_tors	0.275	0.273
TwoLeadECG	0.532	0.578
NonInvasiveFatal ECG_Thorax1	0.698	0.693
NonInvasiveFatal ECG_Thorax2	0.730	0.712

5. Conclusion

In this paper, we have focused on the problem of similarity search in ECG database. In this regards, we have developed an efficient similarity measure. The obtained AQP-DTW is based on Haar Wavelet Transform and QPDTW similarity measure. Experimental results show that the proposed method is about three times faster than QP-DTW with practically no loss of precision. As future work, we aim to apply the proposed method in biometric field.

Référence :

[1] Agrawal, R, Faloutsos, C, Swami, A. (1993) Efficient Similarity Search in Sequence Databases, In *Proceedings of the International Conference on Foundations of Data Organization and Algorithms*, 1993.
 [2] Paredes, R, Reyes, N. (2009) Solving similarity joins and range queries in metric spaces with the

list of twin clusters, *Journal of Discrete Algorithms*, Elsevier, vol. 7, n°1, pp. 18–35, 2009.
 [3] Obin, N, Roebel, A. (2016) Similarity Search of Acted Voices for Automatic Voice Casting, In *IEEE/ACM Transactions on Audio, Speech, and Language Processing* vol. 29, n°9, 2016.
 [4] Schimke, S, Vielhauer, C. (2007) Similarity searching for on-line handwritten documents, *Journal on Multimodal User Interfaces*, vol. 1, n°2, 2007.
 [5] Keogh, E, Wei, L, Xi, X, Lee, S, Vlachos, M. (2006) LB_Keogh Supports Exact Indexing of Shapes under Rotation Invariance with Arbitrary Representations and Distance Measures”, In *Proceedings of the 32nd international conference on Very large databases VLDB*, 2006.
 [6] Tavenard, R. (2005) Indexation de séquences de descripteurs pour exploitation audio et vidéo, Rapport de stage, Ecole Centrale de Lyon – Université de Lyon I, IRISA, 2005.
 [7] Cassisi, C, Montalto, P, Aliotta, M, Cannata, A, Pulvirenti, A. (2012) Similarity Measures and Dimensionality Reduction Techniques for Time Series Data Mining”, Chapter 3 published in *IntechOpen*, 2012.
 [8] Silva, D. F, Giusti, R, Keogh, E, Batista, G. (2018) Speeding up similarity search under dynamic time warping by pruning unpromising alignments, In *Data Mining and Knowledge Discovery*, vol. 32, n°2, 2018.
 [9] Movchan, A, Zymbler, M. (2015) Time Series Subsequence Similarity Search Under Dynamic Time Warping Distance on the Intel Many-core Accelerators, Chapter from book *Similarity Search and Applications*, In *proceedings of the 8th International Conference, SISAP*, Glasgow, UK, 2015.
 [10] Shen, D, Chi, M. (2020) An Initial Study on Adapting DTW at Individual Query for Electrocardiogram Analysis, In V. Lemaire et al. (Eds.): *AALTD 2019, LNAI 11986*, Springer, pp. 213–228, 2020.
 [11] Banaee, H, Ahmed, M, Uand Loutfi, A. (2013) Data Mining for Wearable Sensors in Health Monitoring Systems: A Review of Recent Trends and Challenges, In *Sensors*, vol. 13, n°12, pp. 17472–17500, 2013.
 [12] Kaur, G, Singh, D, Kaur, S. (2015) Electrocardiogram (ECG) as a Biometric Characteristic: A Review”, In *International Journal of Emerging Research in Management & Technology*, vol. 4, n°5, pp. 202–206, 2015.
 [13] Passos, d. S, Felipe, T, Duru, B, Oliveira, E. L, Pares, S, Lima, C. (2017) Symbolic representation of time series applied to biometric recognition based on ECG signals, In *proceedings of International Joint Conference on Neural Networks (IJCNN)*, 2017.
 [14] Keogh, E, Chakrabarti, K, Pazzani, M, and Mehrotra, S, “Dimensionality reduction for fast similarity search in large time series databases”, *Knowledge and Information Systems*, vol. 3, n°3, pp. 263–286, 2001.
 [15] Chan, K. P, Fu, A. W. C. (1999) Efficient time series matching by wavelets”, In *proceedings of the 5th International Conference on Data Engineering*, 1999.

- [16] Lin. J, Keogh. E, Li. W, and Lonardi. S. (2007) "Experiencing SAX: a novel symbolic representation of time series". *Data Mining and Knowledge Discovery*, vol.15, n°2, pp. 107–144, 2007.
- [17] Echihiabi, K, Zoumpatianos, K, Palpanas, T and Benbrahim, H. (2018) "The lernaean hydra of data series similarity search: An experimental evaluation of the state of the art". *The VLDB Journal*, vol.12, n°2, 2018.
- [18] Gogolou, A, Tsandilas, T, Palpanas, T, Bezerianos, A. (2019) "Progressive Similarity Search on Time Series Data", In *proceedings of the 2nd International Workshop on Big Data Visual Exploration and Analytics*, Mar 2019.
- [19] Boulnemour, I, Boucheham, B. (2018) QP-DTW: upgrading dynamic time warping to handle quasi-periodic time series alignment, *Journal of Information Processing Systems*, vol. 14, n°4, pp. 851-876, 2018.
- [20] Boucheham, B. (2008). Matching of quasi-periodic time series patterns by exchange of block-sorting signature, *Pattern Recognition Letters*, Elsevier, vol. 29, pp. 501-514, 2008.
- [21] Sakoe, H, Chiba, S. (1978) Dynamic programming algorithm optimization for spoken word recognition", *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol.26, n°1, pp.43-49, 1978.
- [22] Berndt, D. J, Clifford, J. (1994) Using dynamic time warping to find patterns in time series, In *Proceedings of AAAI Workshop on Knowledge Discovery in Databases (KDD'94)*, pp. 359-370, Seattle, Washington, 1994.
- [23] Chen, Y, Keogh, E, Hu, B, Begum, N, Bagnall, A, Mueen, A, Batista, G. (2015) *The UCR time series classification archive*, 2015.
- [24] Garcia de Andrade, C, Ribeiro, M, X. (2014). Similarity Search in Multidimensional Time Series using the Coulomb's Law, *Journal of Information and Data Management*, Vol. 5, No. 1, pp. 74–83, 2014.