Bio-Medical Materials and Engineering 26 (2015) S1095–S1105
DOI 10.3233/BME-151406
IOS Press

A novel similarity comparison approach for dynamic ECG series

Hong Yin[a,b,*], Xiaoqian Zhu[a], Shaodong Ma[c], Shuqiang Yang[a] and Liqian Chen[a]
[a]*College of Computer, National University of Defense Technology, Changsha 410073, China*
[b]*Xiangyang School for NCOs, Xiangyang 441118, China*
[c]*School of Engineering, The University of Hull, Cottingham Road, Hull, United Kingdom*

Abstract. The heart sound signal is a reflection of heart and vascular system motion. Long-term continuous electrocardiogram (ECG) contains important information which can be helpful to prevent heart failure. A single piece of a long-term ECG recording usually consists of more than one hundred thousand data points in length, making it difficult to derive hidden features that may be reflected through dynamic ECG monitoring, which is also very time-consuming to analyze. In this paper, a Dynamic Time Warping based on MapReduce (MRDTW) is proposed to make prognoses of possible lesions in patients. Through comparison of a real-time ECG of a patient with the reference sets of normal and problematic cardiac waveforms, the experimental results reveal that our approach not only retains high accuracy, but also greatly improves the efficiency of the similarity measure in dynamic ECG series.

Keywords: Similarity, DTW, ECG series, MapReduce, parallel

1. Introduction

Dynamic Electrocardiogram (ECG) monitoring, also known as Holter ECG [1] is used to detect hidden but vital changes in the heart sound such as transient ischemic episodes, cardiac arrhythmias, and for arrhythmic risk assessment of patients [2]. A typical ECG signal frame is shown in Figure 1. Holter ECG recording typically lasts for 24 hours, to investigate unusual causes of heart failure and associated symptoms (e.g., cardiorespiratory abnormality) through a cumulative recording of daily cardiac motions. A single piece of the Holter ECG recording is composed of more than 100,000 data points; analyzing dynamic ECG data is extremely time-consuming, increasing time spent on diagnosis, even the analysis is computer-aided. Many algorithms have been developed to detect, classify and automatically analyze electrical cardiac complexes, such as wavelet decomposition [3, 4] and radial basis function (RBF) modeling [5]. Automatic classification methods are applied to the Holter ECG signal analysis for labeling of cardiac waveform features (e.g., neural networks) [6], Hidden Markov Models [7] and Support Vector Machines [8]. In order to achieve high recognition accuracy, many methods focus on exact temporal and amplitude features from the QRST wave [9].

* Address for correspondence: Hong Yin, College of Computer, National University of Defense Technology, Changsha 410073, China. Tel.: +86 186 7036 1345; Fax: 86-731-84467256; E-mail: yinhonggfkd@aliyun.com.

For the quasiperiodic nature of ECG series, most algorithms split the long-term ECG series into thousands of segments in which an individual segment contains a complete period of a heartbeat with the QRS complex or the RR interval (an ECG signal from one R peak to the next). Accordingly, the segmented series will be individually compared and analyzed, somewhat reducing the computational complexity. However, the series segmentation is likely to result in faulty feature detections due to regular division of the waveforms which may contain irregular, periodic pulses. On the contrary, the global trend feature of a long-term ECG series that may be critical to cardiac malfunction prevention will become invisible if the ECG series are segmented, whereas comparing the global long-term ECG series without segmentation can help reduce the accumulated faulty analysis and find useful hidden medical diagnoses over time. For example, in the satellite fault diagnosis, a long-term series is analyzed using the classification algorithm, and the nominal data is classified into clusters representing different modes of the system. Whether an anomaly occurred is determined by measuring the similarity in the new series from sensors and the enrolled template [10]. This method can also be applied to ECG signal analysis. In this paper, a highly efficient similarity comparison method, Dynamic Time Warping based on MapReduce (MRDTW), is proposed for analysis of long-term ECG series of more than 100,000 points. The feasibility of making prognoses for the possible cardiopathy of patients will be investigated by comparing the real time ECG of a patient with the enrolled templates, including both nominal and diseased series.

2. Related works

Due to the non-stationary feature of an ECG series, the Euclidean distance is not suitable for comparing the similarity of two ECG series. In fact, heartbeat variation can produce nonlinear time fluctuation of ECG series; because the Euclidean distance is very sensitive to small distortions in the time axis, it may fail to produce an intuitively correct measure of similarity between two series. In order to eliminate this fluctuation, time normalization is frequently employed in ECG classification. Dynamic time warping (DTW), which has been used successfully in speech recognition [11], is a pattern-matching algorithm with a nonlinear time-normalization effect. The DTW was utilized to obtain a template from the left ventricular volume ECG signal by Caiani et al. [12], and was also used to track the diastasis onset and offset on a cycle-by-cycle basis. Hu et al. proposed a new method of processing a noisy ECG signal contaminated by electromyogram, which depends on the derivation of a subspace filter. In this method, the alignment of ECG cycles of different length can also be accomplished by DTW [13]. By combining pattern matching with the DTW, Syeda-Mahmood et al. developed a method of capturing the perceptual pattern similarity of ECG waveforms. Cardiovascular diseases based on pattern matching can be analyzed from the results associated with ECG pattern similarity [14].

A major disadvantage of using DTW to analyze ECG time series is that due to the quadratic O(N^2) time and space complexities, it is difficult for DTW to measure similarity in large series with more than one thousand data points. For example, Chadwick et al. found that due to time constraints, DTW did not perform well for large series [15].

Likewise, Bemdt and Clifford realized that the performance of DTW is seriously constrained when anaylizing large databases [16], and it has been determined insufficient for scaling DTW methods to process massive databases [17]. To reduce the computational complexity, many improved algorithms have been proposed for DTW searching in large databases, such as the Piecewise Dynamic Time Warping (PDTW) [18], Iterative Deepening Dynamic Time Warping (IDDTW) [19] and Segmented Dynamic Time Warping (SDTW) [20], which functions as approximate search technology. The main

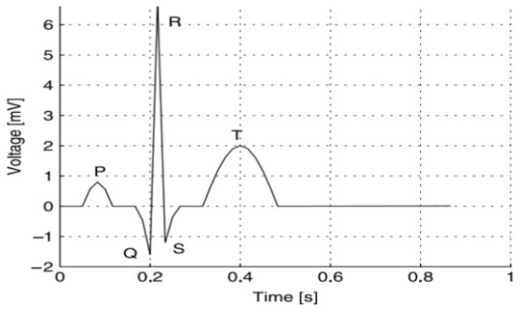


Fig. 1. A complete period of a heartbeat.

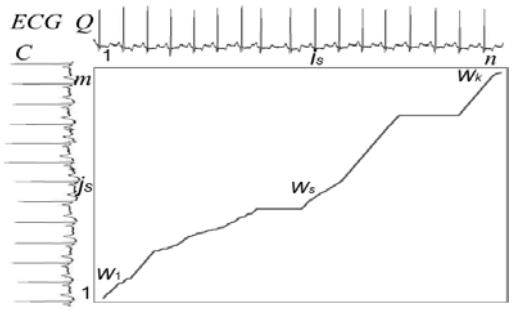


Fig. 2. An example of a warping path for ECG series.

limitation of the approximate search is that it is not guaranteed to find the optimal solution [21]. MapReduce proposed by Google is a parallel programming model for processing massive data sets with a distributed algorithm in a cluster [22]. Many applications benefit from MapReduce, such as Hadoop [23] which is a distributed system infrastructure, and MongoDB [24] which is a database based on a distributed file storage. Accordingly, incorporating MapReduce with DTW will present a novel way to improve the efficiency and accuracy of similarity comparison.

3. The similarity measure of ECG series

In this section, the preliminary concepts of experimental data and test conditions will be defined for a record of a Holter register, which is typically composed of more than 1,000,000 data points.

Definition 1. *ECG Series*: An ECG series $T=t_1,t_2\ldots t_m$ is a sequence of m samples collected at regular intervals over a period.

Definition 2. *ECG Subsequence*: Given an ECG series T of length m, an ECG subsequence C of T is a sampling of length n ($n \leq m$) of contiguous positions from T, that is, $C=t_p,\ldots\ldots t_{p+n-1}$, $1 \leq p \leq m-n+1$.

Definition 3. *Similarity of ECG Series*: Consider two ECG series S_1 and S_2; their similarity function is $Sim(S_1, S_2)$, given a threshold ε, if: $Sim(S_1, S_2) \leq \varepsilon$ established, and S_1 and S_2 are similar to the case of the ε boundary.

After de-noising and normalization for ECG, if two ECG series are similar, it is believed that the two ECG have the same syndrome at one time.

3.1. Dynamic time warping (DTW)

Dynamic Time Warping (DTW) was first introduced by Bemdt and Clifford to analyze the time series. For thoroughness, the classic DTW algorithm is reviewed here.

Consider two ECG series Q and C, of length n and m respectively, where:

$$Q=q_1, q_2,\ldots, q_i,\ldots, q_n \qquad C=c_1, c_2,\ldots, c_j,\ldots, c_m \tag{1}$$

First, an n-by-m distance similarity matrix is constructed, where the $(i^{\text{th}}, j^{\text{th}})$ element of the matrix forms the distance vector $d(q_i,c_j)$ between the two points q_i and c_j(Manhattan Distance is used here, so $d(q_i,c_j)=|q_i - c_j|$). Each matrix element (i, j) corresponds to the alignment between points q_i and c_j., as illustrated in Figure 2.

Definition 4. *Warping path*: A warping path, W, is a contiguous set of matrix elements which defines a mapping between Q and C. The k^{th} element of W is defined as $w_k = (i,j)_k$, so that: $W=w_1, w_2,..., w_k,... w_K$, and K satisfies: $max(m,n) \le K \le m+n$-1

A warping path must match the constraints of boundary conditions, continuity, and monotonicity. There are a significant number of potential warping paths that can meet the above conditions; however, the path which has the minimum warping cost is measured as:

$$DTW(Q,C)=\min\left\{\left(\sqrt{\sum_{k=1}^{K} w_k}\right)/K\right\} \tag{2}$$

The parameter K in Eq. (2) is used to reduce the effect caused by warping paths with different lengths. The cumulative distance $\gamma(i,j)$ must be defined before searching the optimum warping path. $\gamma(i,j)$ is defined using the distance $d(i,j)$ which can be determined by the current cell and the minimum of the cumulative distances of the adjacent elements. In order to find the warping path more efficiently, dynamic programming is used to compare $\gamma(i,j)$ recursively.

$$\gamma(i,j)= d(q_i,c_j)+\min\{\gamma(i-1,j-1), \gamma(i-1,j), \gamma(i,j-1)\} \tag{3}$$

3.2. The similarity comparison for clinical decision making

To develop clinical prognoses of heart diseases, the test ECG series are usually compared with many or all biometric templates stored in the database; that is to say that a patient's condition can be predicted by comparing the similarity of ECG series. It is important to obtain the compared results very quickly in order to make a good clinical decision for the patient, especially those who are critically ill. Using the DTW to compare the similarity of biometric series, a threshold denoted as ε is set to distinguish whether the test series belongs to the same classification. The DTW distances between the test series and all enrolled templates are calculated. If the lowest distance is higher than the threshold ε, then the test series is not considered similar to all the templates. If the lowest distance is smaller than the threshold ε, then the test series and the enrolled template with the lowest DTW distance are considered to be a match, meaning that the enrolled template of longer length can be used to predict the subsequent trend of the test series with shorter length. This process is illustrated in Figure 3.

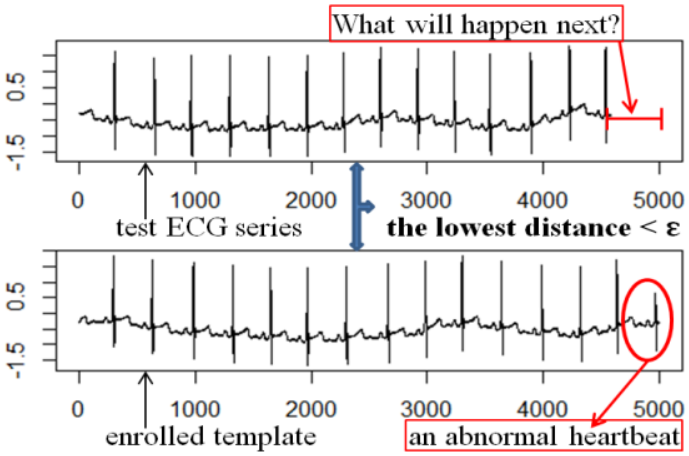
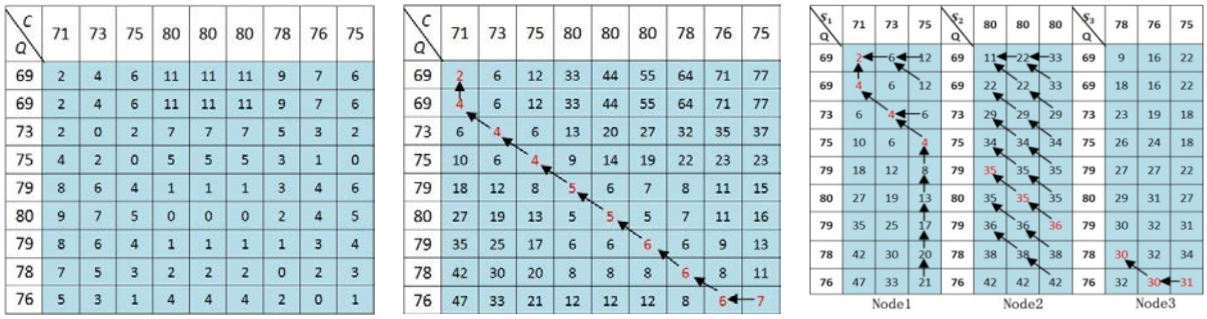


Fig. 3. Example of test ECG series and enrolled template with high similarity.



(a) (b) (c)

Fig. 4. (a) The distance matrix of *C* and *Q*; (b) The warping path of *C* and *Q*; (c) The result of parallel DTW.

4. Optimization of DTW in parallel

4.1. Segmentation of ECG series

Theoretically, the brute-force search method can be adopted to find the paths which meet the constraints mentioned above. It could, however, lead to complete exhaustion that is impractical in ECG series; therefore, one of the ECG series *C* is segmented. The segmented sub-sequences are given as: $S_1, S_2, \ldots S_l$. Each sub-sequence is located on different nodes from a distributed cluster, allowing the use of the parallel mechanism to improve the DTW algorithm. The ECG series Q is not segmented, and is distributed on every node.

In this paper, an adaptive segment algorithm to recognize the periodic feature is proposed to split the ECG series [18], an important aspect for analyzing time series with a setperiod, such as a long-term ECG series. The sub-sequences obtained by the segment algorithm retain most of the original ECG series modes, and each sub-sequence is relatively independent. Comparing the similarity between sub-sequences S_i and Q with the DTW method, the final results will be integrated to reflect the similarity between *C* and *Q*.

4.2. A numerical example of parallel DTW

Consider an example to illustrate the principal operations of parallel comparison. The series: *C*={71,73,75,80,80,80,78,76,75}, *Q*={69,69,73,75,79,80,79,78,76}. First, the distance matrix *D* is computed (using Manhattan Distance), results shown in Figure 4(a). Using Formula (3) to compute the cumulated distance $\gamma(i,j)$ of matrix *D*, the obtained warping path *W* is shown in Figure 4(b): $\{(1,1),(1,2),(2,3),(3,4),(4,5),(5,6),(6,7),(7,8),\ (8,9),(9,9)\}$. Before the parallelization of DTW, the series *C* is grouped as three subsequences: $S_1 = \{71,73,75\}$, $S_2 = \{80,80,80\}$, $S_3 = \{78,76,75\}$.

$S_1$and S_2 respectively represent the mode of rising and invariant, and S_3 represents the mode of descending. Each sub-sequence represents a mode of original sequence. Supposing that there are three nodes--Node 1, Node 2, Node 3--in a cluster, the sub-sequences, S_1, S_2, S_3, are located on the three nodes respectively, and series Q is located on every node. The process of parallelization is, in fact, comparing the similarity of S_1 and Q, S_2 and Q, S_3 and Q simultaneously. The result of parallel DTW is shown in Figure 4(c).

On Node 1, because the first point of warping path at (1,1), is fixed due to the boundary conditions, the path marked with red numbers is the shortest, W_1={(1,1), (1,2), (2,3), (3,4)}. On Node 3, because S_3 is the last sub-sequence, according to the boundary conditions of DTW, the last point of the warping path is (*n,m*); in this this example, W_3 = {(7,8),(8,9),(9,9)}. On Node 2, the choice of the path is complicated, as the first and the last point are both uncertain. All possible paths of Node 2 are given as shown in Figure 4(c): {(4,7),(5,8),(6,9)}, {(4,6),(5,7),(6,8)}, {(4,5),(5,6),(6,7)}, {(4,4),(5,5),(6,6)}, {(4,3),(5,4),(6,5)}, {(4,2), (5,3),(6,4)}, {(4,1),(5,2),(6,3)}, {(4,1),(5,1),(6,2)}, {(4,1),(5,1),(6,1)}.

4.3. Pruning of the warping path

As illustrated in the above section, there are several possible paths in the middle nodes. The selection of the optimal path will be generated by pruning the possible warping paths. To achieve the goal, the following conditions are defined:

Definition 5. *Flexibility*: the flexibility *F* of a warping path is defined as:

$$F = (\sum_{k=1}^{K} |i_k - j_k|)/K \tag{4}$$

In the process of an optimum warping path search, a threshold ε must be determined. The condition $F \geq \varepsilon$ suggests that the model of the two time series cannot be a dynamic warping match. In this paper, the ε is set to *K*.

Definition 6. *Continuity*: The elements in a warping path are continuous, and the path in the time dimension does not jump, satisfying: $i_s - i_{s-1} \leq 1$ and $j_s - j_{s-1} \leq 1$.

According to definition 5, the paths {(4,7), (5,8), (6,9)},{(4,1), (5,2), (6,3)}, {(4,1), (5,1), (6,2)} and {(4,1), (5,1), (6,1)} demonstrate that the model of each sub-sequence is significantly different, and are therefore discarded. For the remaining five paths, in descending order of their distance value, are listed as: {(4,6), (5,7), (6,8)},{(4,5), (5,6), (6,7)},{(4,4), (5,5), (6,6)}, {(4,3), (5,4), (6,5)} and {(4,2), (5,3), (6,4)}. The distances of the five paths are all similar. Nevertheless, according to continuity, the first point of the path following W_1 can be (4,4) or (4,5), and the last point of the path W_3 must be (6,7) or (6,8). Therefore it is demonstrated that only the path {(4,5), (5,6), (6,7)} satisfies the continuity, and W_2={(4,5), (5,6), (6,7)} is the optimal path. Combining the paths W_1,W_2 and W_3, the final warping path is obtained as *W* = {(1,1), (1,2), (2,3), (3,4), (4,5), (5,6), (6,7), (7,8), (8,9), (9,9)}. This result is the same as that derived by a standard DTW method.

4.4. Algorithm implementation of MRDTW

The algorithm implementation of parallel DTW is based on the MapReduce technique (MRDTW) which is an excellent framework for massive data processing. A more detailed description of the MapReduce method can be found in [22]. The realization of the Map function used in the proposed parallel DTW can be accomplished by the following procedures: comparing the similarities of the ECG sub-sequence with another ECG series in parallel on each distributed node, and obtaining the possible warping paths respectively. The path of the first and the last sub-sequences can be determined immediately; the other is pruned by setting flexibility and slope constraints.

The implementation of the Reduce function is the merging ofthe end-to-end warping paths according to the principle of continuity, which satisfies $i_s - i_{s-1} \leq 1$ and $j_s - j_{s-1} \leq 1$. After calculating the total length of the warping path, the path with the minimum length is the solution.

5. Experimental results

5.1. Efficiency of MRDTW to analyze ECG series

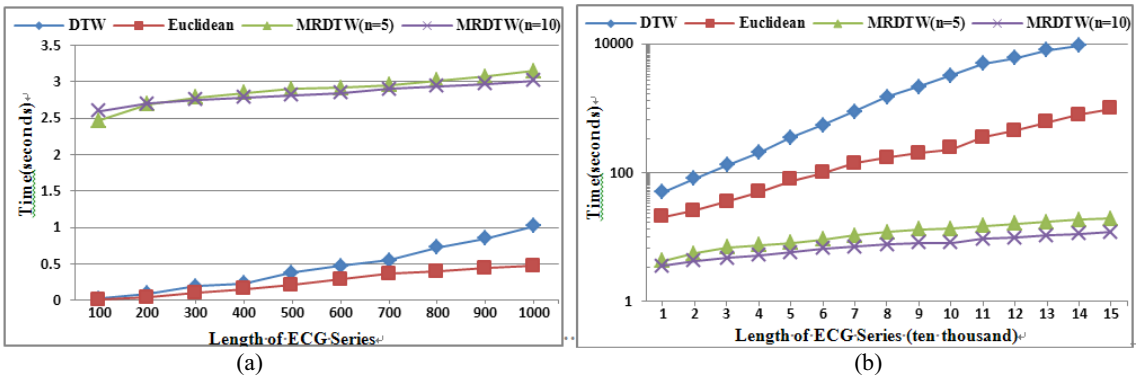
The test data chosen to evaluate the efficiency of MRDTW is the ECG data in the Physionet MIT-BIH Long-Term database which contains six two-channel ECG signals sampled at 128 Hz per channel, and one three-channel ECG sampled at 128 Hz per channel. The duration of the seven recordings varies from 14 to 24 hours each, involving more than 100,000 data points in this experiment.

The standard DTW algorithm and Euclidean distance are selected as the comparative methods. The MRDTW algorithm will be tested with the node numbers 5 and 10. The lengths of the ECG series vary from 100 to 160,000. The MRDTW algorithm was significantly faster than the standard DTW algorithm with the larger ECG series; MRDTW is approximately 230-390 times faster than standard DTW when the ECG series have lengths of 160,000 points (and the node numbers are set as 5 and 10 respectively). A sample of the MRDTW algorithm performance is reflected in Table 1.

From Table 1, it can be concluded that Euclidean distance is more efficient than DTW in simple calculation. With the time consumed for the allocation of the map tasks in the cluster, MRDTW is slower than DTW and Euclidean distance when the length of the ECG series is less than 1,000. However, MRDTW is suitable for analyzing ECG series with rapid increase in the length of ECG series and a gentle growth in time consumption. Additionally, with an increasing number of nodes, the efficiency of the MRDTW algorithm can be further improved. The compared results are illustrated in Figure 5. Note that Figure 5(a) is scaled linearly, but the *y*-axis of Figure 5(b) has log-log scaling for convenience of illustration.

Table 1
A comparison of different methods of efficiency

Algorithm	Length of ECG Series			
	100	1,000	10,000	100,000
DTW	0.02s	1.02s	48.82s	3212.43s
Euclidean	0.01s	0.47s	20.73s	232.14
MRDTW (*node*=5)	2.47s	3.15s	4.29s	13.45s
MRDTW (*node*=10)	2.60s	3.02s	3.54s	8.21s



(a) (b)

Fig. 5. The compared results of efficiency (a) on small ECG series and (b) on large ECG series.

5.2. Accuracy of MRDTW in classification of ECG series

The accuracy of the MRDTW algorithm can be measured by the classification of ECG data. Three classes of data from the MIT-BIH ECG database are selected as experimental data: Arrhythmia ECG, Atrial Fibrillation ECG and Normal Sinus Rhythm ECG. The numbers of ECG series belong to the three classes as shown in Table 2. The length of each series is approximately 7,600 data points.

The PDTW [18], an excellent similarity measure method, is selected to test in comparison to MRDTW. From the results shown in Figure 6, the classified accuracy of MRDTW is higher than PDTW with the node set as 5. In each class, PDTW misclassifies more series than MRDTW. Although the PDTW makes the calculation more easily, it only uses the approximation piecewise to represent the original series; much information about the ECG series is ignored, which will reduce the classified accuracy of PDTW. For ECG series, the parallel mechanism of MRDTW is very useful to measure the similarity, and is highly accurate in classification.

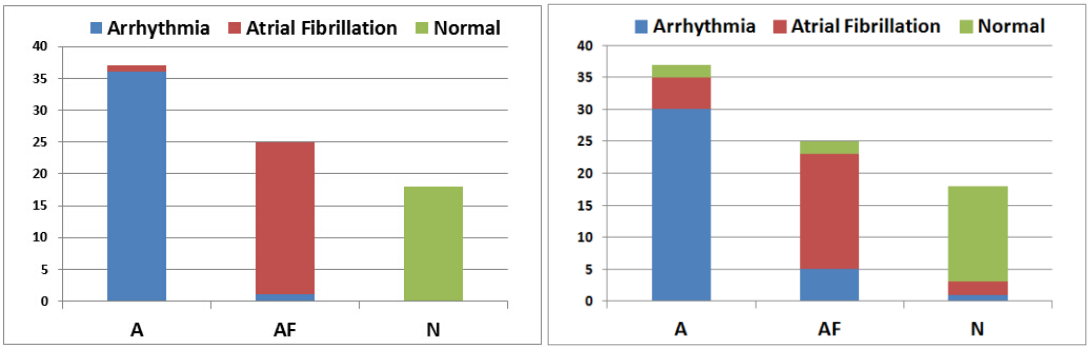
5.3. The accuracy of prediction for clinical decision making

Section 3.2 introduced the idea of using similarity comparison to aid clinical decision-making. By comparing the similarity between test ECG and template ECG, the former can be simulated by the latter with longer length. In this experiment, the prediction accuracy for clinical decision-making is tested. The MIT-BIH ECG series mentioned above are used as the test data; prediction results are as shown in Figure 7. Images (a), (b), and (c) illustrate the real ECG series, and images (d), (e), and (f) represent the same series with prediction results marked in red lines, according to the similarity comparison using MRDTW. Results demonstrate that the prediction results are similar to the real ECG series, increasing the the number of templates and positively influencing clinical decision-making.

The accuracy of MRDTW prediction is further tested using different numbers of nodes. An evaluation index is assigned to measure the accuracy of prediction: the Mean Distance Difference (MDD), the real ECG series represented by C, and the prediction series represented by Q. The formula of MDD is as follows:

Table 2
Heartbeat classes with the number of series

Class	Arrhythmia (A)	Atrial Fibrillation (AF)	Normal Sinus Rhythm (N)
Number	37	25	18



(a) (b)

Fig. 6. (a) The classification of MRDTW with node=5; (b) The classification of PDTW.

$$\mathrm{MDD} = \frac{\sum |c_i - q_j|}{n} \quad (n \text{ is the length of real ECG series}) \tag{5}$$

With a threshold θ, if the MDD<θ, the prediction result of the ECG series is considered to be acceptable. In order to evaluate the accuracy of the predictions, 250 ECG templates are selected, with the experimental results shown in Figure 8.

From the experimental results, it can be concluded that the accuracy of MRDTW for ECG prediction is altered by the number of nodes (n). With an increase in n, the accuracy declines; the middle results become larger with an increase in n, making the choice of optimal path more difficult. This will also occur with increased length of the ECG series. Still, the experimental results are inspiring: when n=4, the percentage of acceptable prediction is above 90%.

6. Conclusions

In this paper, a parallel scheme of similarity for ECG is proposed. Unlike the previous methods which focus on the segmentation of ECG, the MRDTW searches the entire ECG series to find similarity that identifies potential disease. Experimental results show that MRDTW is considerably suitable for analysis of ECG series. With increasing length of ECG series, the execution times of most of the algorithms rise exponentially; MRDTW presents a linear growth and retains a high accuracy compared to other methods.

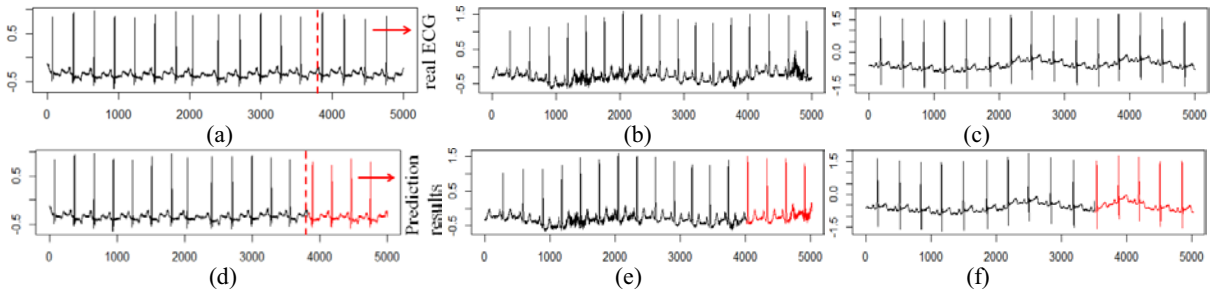


Fig. 7. The prediction results by MRDTW. Images (a), (b), (c) represent the real ECG series, while points (d), (e), (f) are the same series with predictions marked in red.

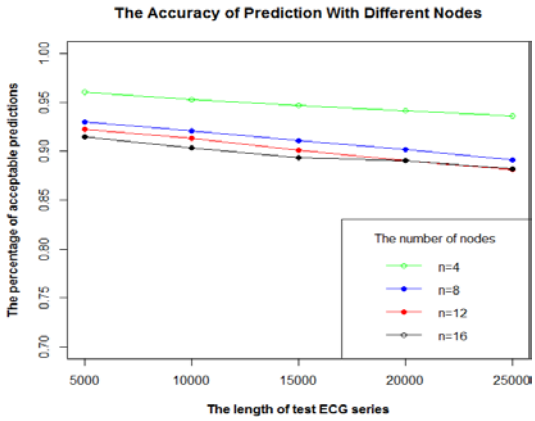


Fig. 8. The accuracy of prediction of MRDTW with different number of nodes.

Acknowledgments

This work was supported in part by National High-tech R&D Program of China under Grants No. 2012AA012600, 2012AA01A401 and 2012AA01A402.

References

[1] N.J. Holter, New method for heart studies continuous electrocardiography of active subjects over long periods is now practical, Science **134** (1961), 1214-1220.
[2] M. Paoletti and C. Marchesi, Discovering dangerous patterns in long-term ambulatory ECG recordings using a fast QRS detection algorithm and explorative data analysis, Computer Methods and Programs in Biomedicine **82** (2006), 20-30.
[3] C. Li, C. Zheng and C. Tai, Detection of ECG characteristic points using wavelet transforms, IEEE Transactions on Biomedical Engineering **42** (1995), 21-28.
[4] J.P. Martínez, R. Almeida, S. Olmos, et al., A wavelet-based ECG delineator: Evaluation on standard databases, IEEE Transactions on Biomedical Engineering **51** (2004), 570-581.
[5] S. Suppappola, Y. Sun and S.A. Chiaramida, Gaussian pulse decomposition: An intuitive model of electrocardiogram waveforms, Annals of Biomedical Engineering **25** (1997), 252-260.
[6] B.P. Simon and C. Eswaran, An ECG classifier designed using modified decision based neural networks, Computers and Biomedical Research **30** (1997), 257-272.
[7] L. Senhadji, L. Thoraval and G. Carrault, Continuous wavelet transform: ECG recognition based on phase and modulus representations and hidden Markov model, Wavelets in Medicine and Biology, A. Aldroubi and M. Unser, eds., CRC Press, NewYork, 1996, 439-463.
[8] S. Jankowski, J. Tijink, G. Vumbaca, B. Marco and G. Karpinski, Morphological analysis of ECG holter recordings by support vector machines, Proceedings of the Third International Symposium on Medical Data Analysis, Springer-Verlag, 2002.
[9] T.W. Shen, W.J. Tompkins and Y.H. Hu, One-lead ECG for identity verification, Engineering in Medicine and Biology, 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society EMBS/BMES Conference **1** (2002), 62-63.
[10] D.L. Iverson, Inductive system health monitoring, International Conference on Artificial Intelligence, Las Vegas, NV, USA, 2004, pp. 605-611.
[11] H. Sakoe and S. Chiba, A dynamic programming approach to continuous speech recognition, Proceedings of the Seventh International Congress on Acoustics **3** (1971), 65-69.
[12] E.G. Caiani, A. Porta, G. Baselli, et al., Warped-average template technique to track on a cycle-by-cycle basis the cardiac filling phases on left ventricular volume, Computers in Cardiology **1998** (1998), 73-76.
[13] X. Hu and V. Nenov, A single-lead ECG enhancement algorithm using a regularized data-driven filter, IEEE Transactions on Biomedical Engineering **53** (2006), 347-351.
[14] T. Syeda-Mahmood, D. Beymer and F. Wang, Shape-based matching of ECG recordings, Engineering in Medicine and Biology Society (EMBS) 29th Annual International Conference of the IEEE, 2007, 2012-2018.
[15] N.A. Chadwick, D.A. McMeekin and T. Tan, Classifying eye and head movement artifacts in EEG Signals, 2011 Proceedings of the 5th IEEE International Conference on Digital Ecosystems and Technologies Conference (DEST), 2011, 285-291.
[16] D.J. Berndt and J. Clifford, Using dynamic time warping to find patterns in time series, KDD Workshop, Seattle, 1994, pp. 359-370.
[17] N. Adams, D. Marquez and G. Wakefield, Iterative deepening for melody alignment and retrieval, Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR), London, 2005.
[18] E. Keogh, K. Chakrabarti, M. Pazzani and S. Mehrotra, Locally adaptive dimensionality reduction for indexing large time series databases, Proceedings of ACM SIGMOD Conference on Management of Data **30** (2001), 151-162.
[19] S. Chu, E.J. Keogh, D. Hart and M.J. Pazzani, Iterative deepening dynamic time warping for time series, Proceedings of SIAM International Conference on Data Mining, 2002.
[20] E.J. Keogh and M.J. Pazzani, Scaling up dynamic time warping to massive datasets, In: Principles of Data Mining and Knowledge Discovery, Springer, Berlin Heidelberg, 1999, pp. 1-11.
[21] S. Salvador and P. Chan, Toward accurate dynamic time warping in linear time and space, Intelligent Data Analysis **11** (2007), 561-580.

[22] J. Dean and S. Ghemawat, MapReduce: Simplified data processing on large clusters, Communications of the ACM **51** (2008),107-113.
[23] T. White, Hadoop: The definitive guide, O'Reilly Media, Inc., 2010.
[24] C. Chodorow, Introduction to MongoDB, Free and Open Source Software Developers' European Meeting (FOSDEM), Brussels, Belgium, 2010.