

Research Article

A Harmonic Linear Dynamical System for Prominent ECG Feature Extraction

Ngoc Anh Nguyen Thi,¹ Hyung-Jeong Yang,¹ SunHee Kim,² and Luu Ngoc Do¹

¹ Department of Computer Science, Chonnam National University, Gwangju 500-757, Republic of Korea

² Department of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Correspondence should be addressed to Hyung-Jeong Yang; hjyang@chonnam.ac.kr

Received 27 September 2013; Revised 6 January 2014; Accepted 10 January 2014; Published 26 February 2014

Academic Editor: Imre Cikajlo

Copyright © 2014 Ngoc Anh Nguyen Thi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Unsupervised mining of electrocardiography (ECG) time series is a crucial task in biomedical applications. To have efficiency of the clustering results, the prominent features extracted from preprocessing analysis on multiple ECG time series need to be investigated. In this paper, a Harmonic Linear Dynamical System is applied to discover vital prominent features via mining the evolving hidden dynamics and correlations in ECG time series. The discovery of the comprehensible and interpretable features of the proposed feature extraction methodology effectively represents the accuracy and the reliability of clustering results. Particularly, the empirical evaluation results of the proposed method demonstrate the improved performance of clustering compared to the previous main stream feature extraction approaches for ECG time series clustering tasks. Furthermore, the experimental results on real-world datasets show scalability with linear computation time to the duration of the time series.

1. Introduction

Clustering multiple time series data have received considerable attention in recent years in various applications, such as industries of finance, business, science domains, and medicine [1–8]. Clustering in time series is the unsupervised mining of grouping similar time series into a same cluster. Specifically, clustering finds groups of homogenous patterns into a cluster so that these patterns within the cluster bear strong similarity to other ones and dissimilarity to patterns in other clusters [9, 10].

Since the quality of clustering results relies strongly on good features extracted from multiple time series, a very important processing step is to identify compact features extracted from multiple coevolving time series. This step can be used to not only convert a series of original values to more meaningful information, such as more understandable and interpretable information, but it also can lead time series to a lower dimensionality with the most relevant features.

This paper is motivated by mining these essential features of the medicine applications, namely, electrocardiography (ECG) time series. The problems are studied based on the

challenges across time series applications such as time shifts effects, nearby frequencies, and harmonics. By exploiting the temporal evolving trends and the correlation characteristics of coevolving time series, meaningful features which help to achieve the best clustering accuracy can be captured.

Feature extraction can efficiently describe time series since suitable representations reduce the feature spaces. This provides highly efficient features for knowledge discovery so that they help to improve performance as well as the speed of the mining algorithms.

Many time series representations have been proposed to extract features. The well-known dimension reduction approaches, namely, Principal Component Analysis (PCA) and Singular Value Decomposition (SVD), are powerful tools to discover linear correlations across multiple time series [11–15]. They effectively give the optimal low-rank approximation of the dataset by removing the variables with lower energy [14]. However, these methods ignore time-evolving patterns since their characteristics are not designed for tracking the ordering of the rows or the columns. Moreover, the projection of the multiple time series into low-dimensional principal components is not easy to interpret. They do not clearly

exhibit the characteristic of each pattern. As a result, applying clustering methodology with these reduction methods to multiple time series leads to poor performance.

One of the popular alternative approaches for feature extraction on multiple time series is Discrete Fourier Transform (DFT), where the original time series data is projected into the frequency domain [16–18]. The main advantage of DFT is to capture frequencies in a single time sequence. However, DFT lacks the dynamics; hence, clustering based on Fourier coefficients is not suitable. Another method based on Linear Predictive Coding Cepstrum (LPCC) which also provides distinguished features in time series is well known in signal recognition [8, 19, 20]. It is efficient for clustering performance by means of a few coefficients. However, the representations of achieved LPC cepstrum features are hard to interpret.

The Linear Dynamical System, known as Kalman filters, has been commonly used for time series analysis because of its simple implementation and extensibility [21–23]. Kalman filter can capture correlations among multiple time series and learn their evolving dynamics. However, the time shift effects, which are temporal gaps between two time series, cannot be handled. The representations of the extracted features are not also clear. The resulting model parameters thus lead to poor clustering performance.

Another popular feature extraction method, known as Dynamic Time Warping (DTW), can handle time shifts across the sequences. However, this approach ignores temporal dynamics. Therefore, applying DTW directly for clustering purpose cannot give good results [24].

Autoregression Moving Average (ARMA) is also a method for feature extraction. However, ARMA parameters do not provide a reliable method since different sets of parameters can be obtained even from time series with similar structures. As a result, the clustering performances will be affected dramatically [25, 26].

In this paper, we applied an approach of prominent feature extraction based on a Harmonic Linear Dynamical System (HLDS) [21] for unsupervised learning of time series in biomedical applications, specifically in ECG. Since the applications have characteristics of lag correlation and temporal dynamics, HLDS works well for the prominent feature extraction of the time series in the same period with different lags and light amplitudes. In addition, HLDS is expected to identify periodic patterns and group them into one cluster. These series of frequencies are known as harmonic series which contain several mixtures of frequencies, especially in sensor measurements, such as in human voice and in human motion domain [23, 27].

Our prime objective is to exploit two common characteristics of multiple coevolving time series: correlations and temporal dynamics for meaningful feature extraction. Correlation reflects the relationships among multiple time sequences, while dynamic property discovers the temporal moving trends of multiple time series by automatically identifying a few hidden variables. For example, a particular medical signal of physiological records in ECG application characterizes a specific symptom of a patient such as a malignant ventricular arrhythmia person. Therefore, each

time series differs from the others in dynamics since time series encodes temporal dynamics along the time ticks. By capturing correlations, we can achieve good interpretable features in the presence of time shift effects and small shift in frequency. By exploiting the evolving temporal components, we can find the clusters of time series by grouping them with similar temporal patterns.

In order to evaluate the effectiveness of the applied method considering clustering accuracy, reliability, and complexity aspects, this paper demonstrates the clustering results of ECG time series using k -means algorithm. The method we applied discovers useful and understandable features from time series for clustering purposes. Moreover, its computational time scales up to the duration of the time sequences. We demonstrate comparative study to assess the performance of the proposed method against previous feature extraction algorithms such as PCA [11, 12], DFT [17, 18], original Kalman filter [22, 23], and LPCC [19, 20].

The rest of the paper is organized as follows. Background of the underlying Linear Dynamical System theory and the proposed model setup are given in the upcoming Section 2. In Section 3, the experimental time series dataset of ECG used in this paper are illustrated and the empirical evaluations have been conducted to assess effectiveness of the proposed feature extraction method in clustering multiple time series. Finally, we present our conclusions in Section 4.

2. Materials and Methods

2.1. Fundamentals of Linear Dynamical System. Multidimensional time series dataset Y is an ordered sequence of data points measured at equal time intervals, denoted by $Y = \{y_1, y_2, y_3, \dots, y_T\}$, where each vector $y_1, y_2, y_3, \dots, y_T$ is formed of M observations recorded at time ticks $t = 1, t = 2, \dots, t = T$, respectively. Time series collections can be formed as the matrix below:

$$Y = \begin{bmatrix} y_1^1 & y_1^2 & \dots & y_1^i & \dots & y_1^M \\ y_2^1 & y_2^2 & \dots & y_2^i & \dots & y_2^M \\ \vdots & \vdots & & \vdots & & \vdots \\ y_t^1 & y_t^2 & \dots & y_t^i & \dots & y_t^M \\ \vdots & \vdots & & \vdots & & \vdots \\ y_T^1 & y_T^2 & \dots & y_T^i & \dots & y_T^M \end{bmatrix}. \quad (1)$$

In particular, each row vector includes all of the observations for one certain time tick which is an ordered sequence of T vectors each having dimensionality of M . Each column denotes the observations for one particular measurement, which is an unordered collection of M one-dimensional time series vector [1, 2].

Linear Dynamical System (LDS), known as Kalman filter, has been used to model multidimensional time series data. By taking the definition of time series above as a matrix for a dynamical system, this means that multidimensional time series data can be presented by a matrix $Y_{M \times T}$ of the variables M and observed time ticks T . LDS builds a statistical model to represent the state of the hidden variables

which are evolving to a linear transformation leading to the observed numerical time sequences. LDS captures the correlations among multiple signals by means of choosing a proper number of hidden variables so that the model can learn the dynamics of time series data [13, 14, 28]. LDS for a multidimensional time sequence is modeled by the following equations:

$$z_1 = \mu_0 + \omega_0 \quad (2)$$

$$z_{n+1} = A \cdot z_n + \omega_n \quad (3)$$

$$y_n = C \cdot z_n + \varepsilon_n, \quad (4)$$

where $\theta = \{\mu_0, Q_0, A, Q, C, R\}$ is the set of parameters. Vectors y_n and z_n denote observed data sequences and hidden variables at time n , respectively. μ_0 is an initial state for hidden variables of the whole system. The transition dynamic matrix A relates to the transition of the state from the current time tick to the next time tick with noise $\{\omega_n\}$. The matrix C is the observation projection with noise $\{\varepsilon_n\}$ at each time t . All noise ω_0 , ω_i , and ε_i ($i = 1, \dots, T$) are zero-mean and normally distributed random variables with covariance matrices Q_0 , Q , and R , respectively [29]. The Expectation Maximization (EM) algorithm is utilized to learn the component models and estimate hidden variables [30]. In the model, each row of output matrix C corresponds to one sequence and, therefore, can be used as features in clustering. However, the clustering quality is not good since these features cannot provide high interpretability excluding information of time shift. Time shift effect or phase shift indicates that two time series have the same frequency but different phase lags.

2.2. Proposed Feature Extraction Methodology. In this section, the proposed method for ECG dataset is set up to illustrate how to exploit the interpretability of prominent features extracted from multiple time series in order to improve the clustering quality. Since each row of output transition matrix C of the LDS model does not present distinct characteristic of the corresponding series, we explore further expecting to discover deeper hidden patterns for each sequence of LDS by learning the straight forward transition matrix A and output projection matrix C .

First of all, the hidden dynamics are learned via Linear Dynamical System (LDS), capturing the series of hidden variables which are evolving according to the linear transformation A ; they are linearly transformed to the observed numerical sequences [13, 28]. This implies that the series of hidden variables, z_n , are evolving over time ticks with linear transition matrix A . Also, the observed data sequences, y_n , are generated from these series of hidden variables with a linear output projection matrix C [21].

Secondly, after achieving the hidden variables from the LDS system, the canonical form of the hidden variables is identified. However, these hidden variables are hard to interpret since they are mixed in the observation sequences. Therefore, we need to make them compact and uniquely identify. Equation (3) represents the hidden variables depending on the eigenvalues of the transition matrix A [21]. Those eigenvalues can capture the amplitude and frequencies of the

underlying signals of hidden variables which are referred to as harmonics. As a result, normalizing the transition matrix can directly reveal the amplitude, frequency, and the mixtures of the given set of sequences. Harmonic Linear Dynamical System (HLDS) uses eigendecomposition on transition matrix A in order to find harmonics as well as the mixing weight of harmonics. The eigendecomposition corresponds to the diagonal matrix (Λ) of eigendynamics and eigenvector (V) of A as follows:

$$A = V\Lambda V^*. \quad (5)$$

In LDS, the output projection matrix C illustrates how the hidden variables are translated into observation sequences with linear combinations. Therefore, we need to compensate C matrix to achieve the harmonic mixing matrix C_h in order to obtain the same observation sequences from eigendynamics (Λ) as the transition matrix:

$$C_h = C \cdot V. \quad (6)$$

The canonical hidden variables will be

$$\begin{aligned} \mu_0^{\text{new}} &= V^* \cdot \mu_0, \\ z_n^{\text{new}} &= V^* \cdot z_n. \end{aligned} \quad (7)$$

V contains conjugate pairs of columns corresponding to the conjugate pairs of eigenvalues in Λ . Therefore, the harmonic mixing matrix C_h must contain conjugate pair of columns corresponding to the conjugate pairs of the eigenvalues in Λ :

$$\begin{aligned} z_n^{\text{new}} &= \Lambda^{n-1} \cdot \mu_0^{\text{new}} + \text{noise}, \\ y_{\text{new}} &= C_h \cdot \Lambda^{n-1} \cdot \mu_0^{\text{new}} + \text{noise}. \end{aligned} \quad (8)$$

From the equation, we can obtain all hidden variables z_n , canonical hidden variables z_n^{new} , and observation y_n . They are mixtures of a set of scaling such as growing, shrinking, or stationary sinusoid signals of data-dependent frequencies which are referred to as harmonics [21]. Their characteristics of frequencies and amplitudes are completely defined by the eigenvalues of the transition matrix A .

Thirdly, the harmonic mixing matrix C_h is calculated in the second step. From here, the contribution of each harmonic to the resulting observation sequences is found. This means that each row of C_h represents the characteristic of each sequence in the domain of the harmonics and thus, it can be used to cluster sequences. However, the harmonic mixing matrix will fail to group similar sequences with phase shifts or time shifts because it tells not only the strength of each eigen dynamic, but also encodes the required phases for different sequences. To eliminate the phase/lag effect by taking the magnitude of the harmonic mixing matrix C_h , we will obtain the same column for the conjugate column of C_h . By dropping these duplicated columns, we will obtain the harmonic magnitude matrix C_m , which tells how strong each harmonic base participates in the observation time sequences and solves lag challenge as well.

Lastly, in order to obtain the interpreted features for each sequence, we apply the dimension reduction approach with

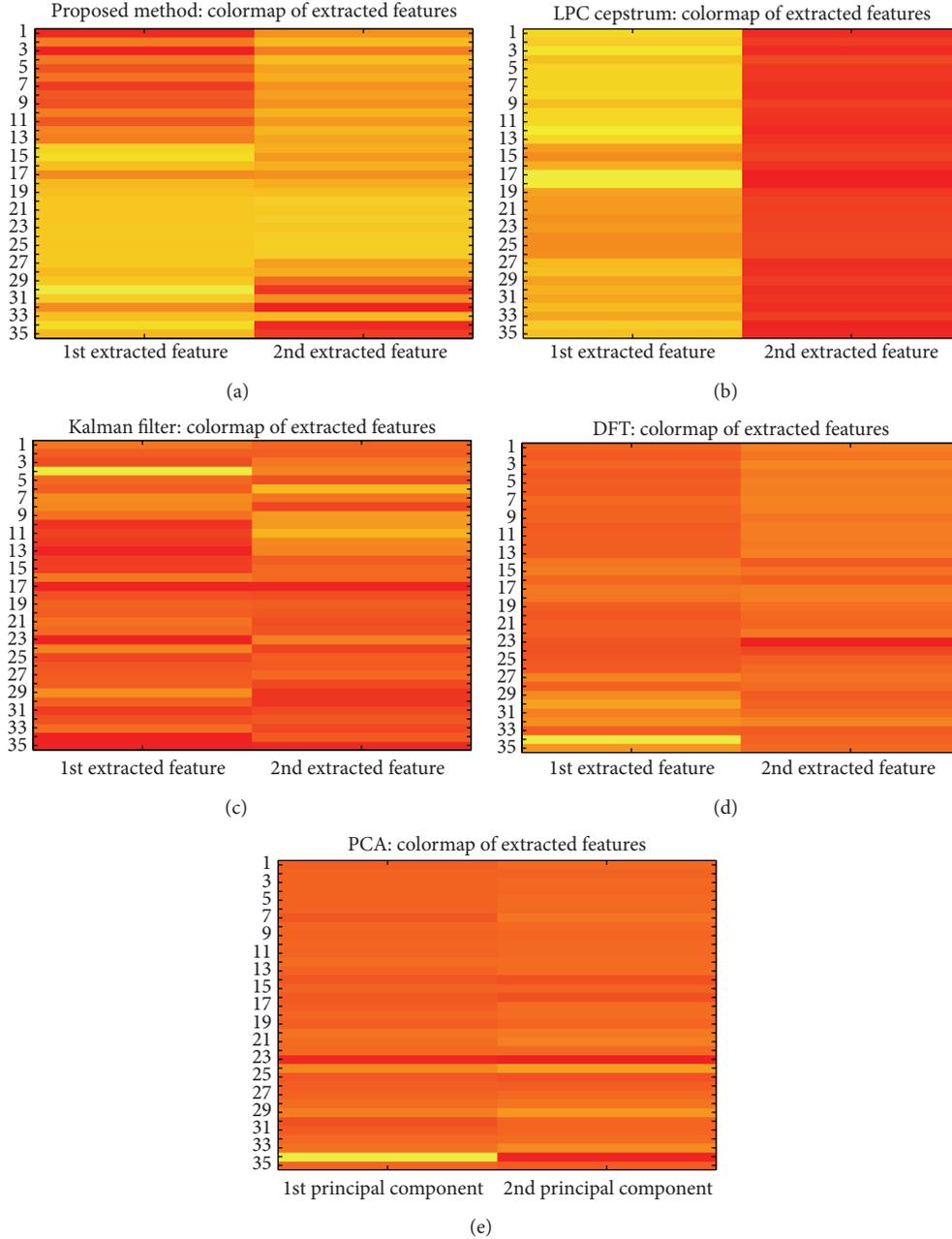


FIGURE 1: The first two extracted features of algorithms on EEG dataset.

SVD on the harmonic magnitude matrix, $\tilde{C}_m \approx U_k \cdot S_k \cdot V_k^T$, where \tilde{C}_m is the column centered from C_m , U_k and V_k are orthogonal matrices with k columns, and S_k is a diagonal matrix. The interpretability of the prominent features can be obtained as $U_k \cdot S_k$. For the proper number of hidden dimension h in our ECG application, we use Fukunaka's principle rule [15, 21]. We choose h as the one with the 98th percentile of the total sum of squared singular values with s_i 's being the singular values of Y in descending order. The formula is as follows:

$$h \leftarrow \arg \frac{\sum_{j=1}^h s_j^2}{\sum_{i=1}^m s_i^2}. \quad (9)$$

In summary, HLDS includes four steps: (1) learning hidden variables using LDS, (2) taking eigendecomposition on transition matrix A to find the canonical form of hidden variables, which helps to find harmonics and mixing weight of harmonics, (3) taking the magnitude of the harmonic mixing matrix to eliminate phase shift, and (4) using SVD to combine harmonics.

3. Experimental Results and Discussion

In order to show the validation of the clustering by HLDS, we carry out experiments on real ECG dataset taken from

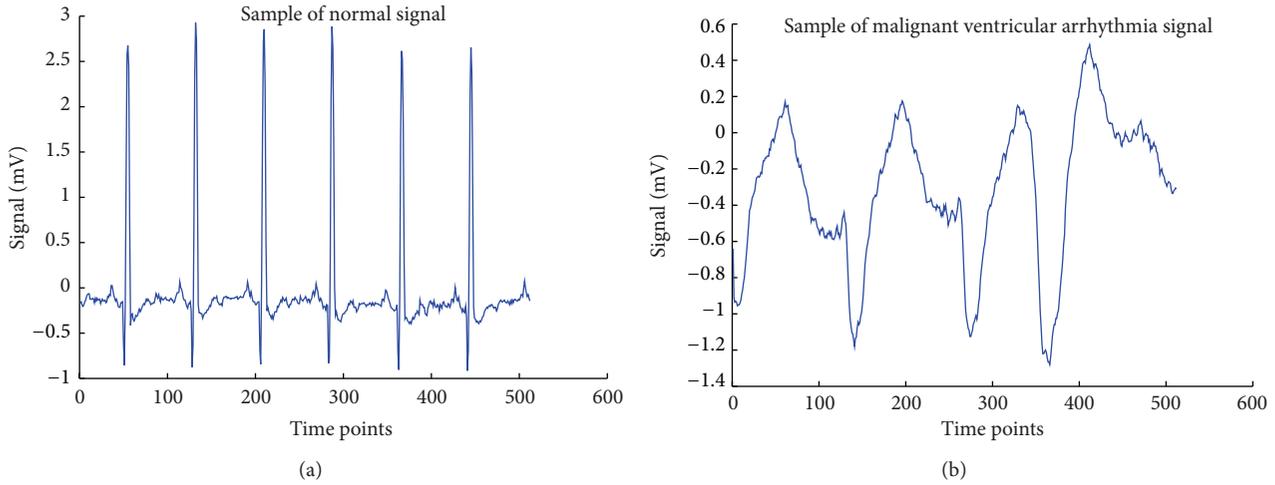


FIGURE 2: Original samples of normal and malignant ventricular arrhythmia time series, respectively.

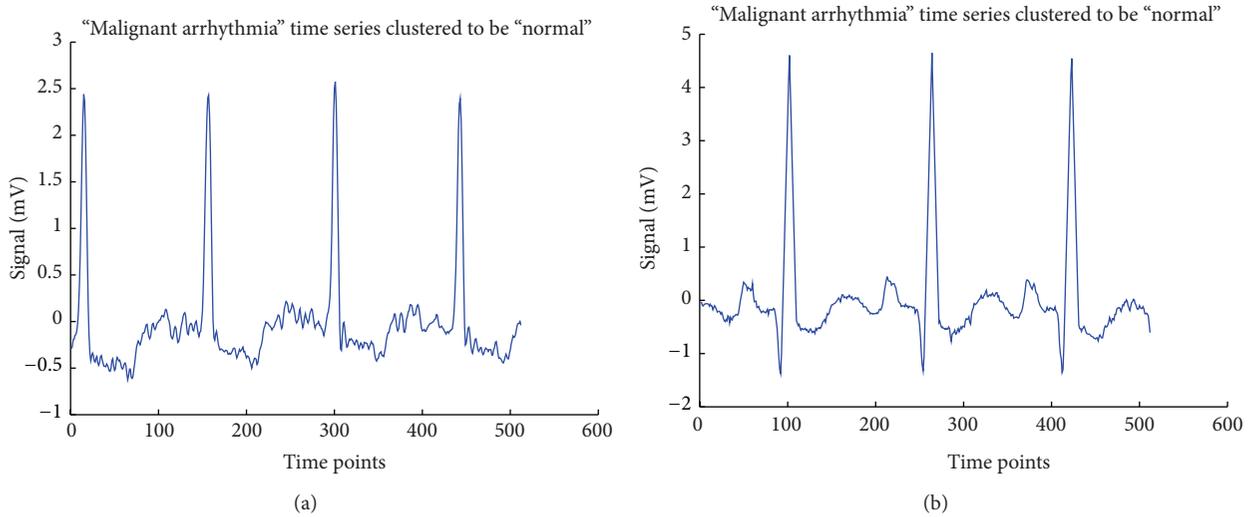


FIGURE 3: The 17th and 32nd malignant ventricular arrhythmia patterns clustered to be wrong cluster.

PhysioNet <http://www.physionet.org/physiobank/database/> [31, 32]. A Harmonic Linear Dynamical System for the prominent feature extraction of ECG application is investigated successfully since the applications have the characteristics of lag correlations and temporal dynamics. These extracted features can group time series patterns in the same period with different lags and light amplitudes. Moreover, the applied method is expected to identify periodic patterns and group them into one cluster. ECG dataset contains three different groups of ECG time series: 13 time series of ECG recordings of healthy people, 22 time series of people having malignant ventricular arrhythmia, and 30 time series of people having supraventricular arrhythmia, taken from the following specific links:

- (i) MIT-BIH Healthy/Normal Sinus Rhythm Database <http://www.physionet.org/physiobank/database/nsrdb/>;

- (ii) MIT-BIH Malignant Ventricular Arrhythmia Database <http://www.physionet.org/physiobank/database/vfdb/>;
- (iii) MIT-BIH Supraventricular Arrhythmia Database <http://www.physionet.org/physiobank/database/svdb/>.

There are two collections which are investigated. Collection 1 contains a group of healthy people and malignant ventricular arrhythmia while collection 2 is obtained by the group of healthy and supraventricular arrhythmia people. To evaluate the effectiveness of the feature extraction for the clustering, both the quality and scalability of normalized time series are considered against previous feature extraction approaches such as LPCC, PCA, DFT, and original Kalman filter. Normalization is carried out to compensate the differences in level and scale of dataset to a zero-mean and unit variance. In the experiment, we use the first two coefficients and cluster them by k -means algorithm with

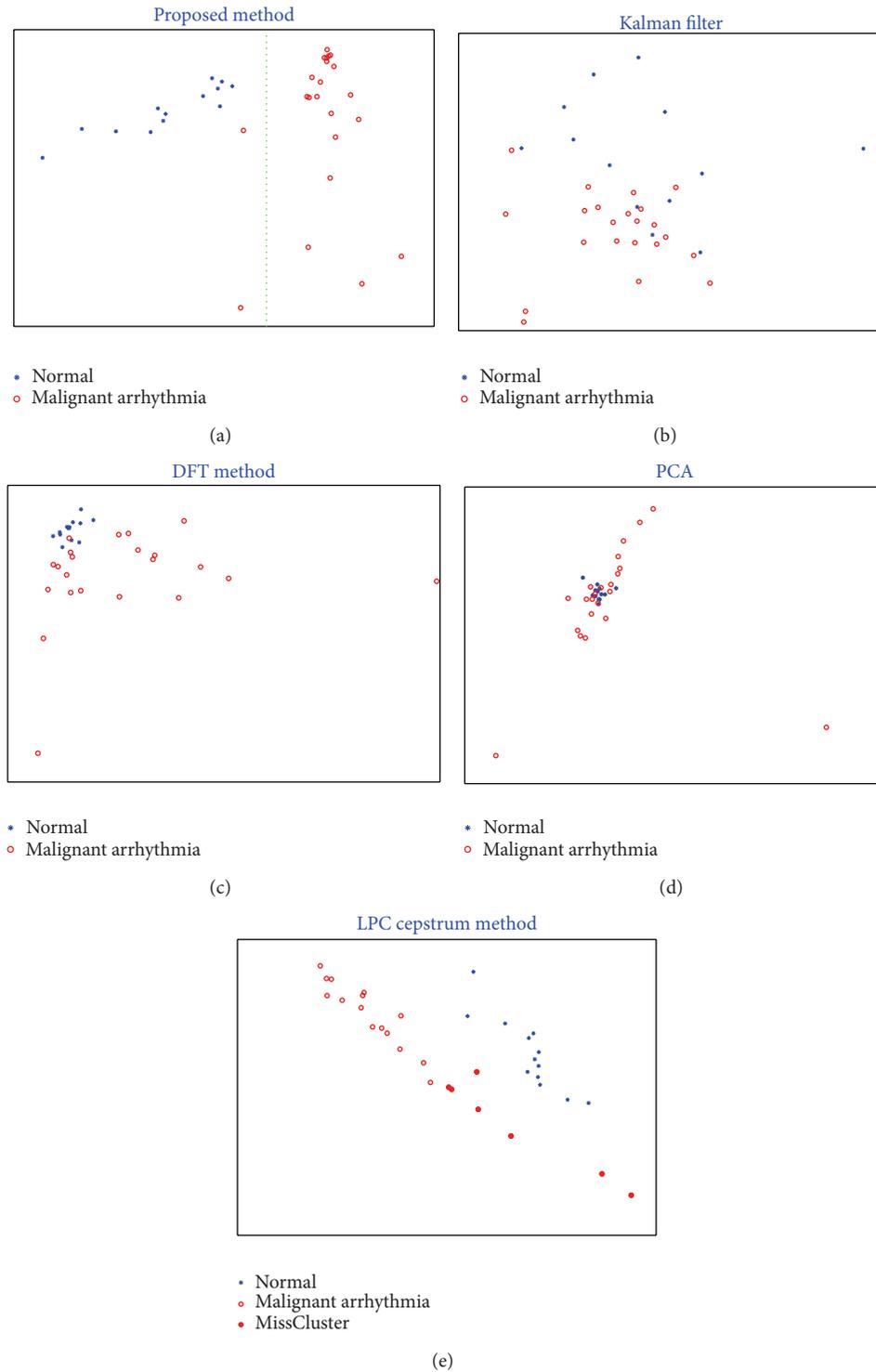


FIGURE 4: Scatter plot of clustering visualization.

Euclidean distance. For choosing the proper number of hidden variables, we set 98% of energy in each original data which is corresponding to the hidden dimension of 20 and 35 for collection 1 and collection 2, respectively.

To evaluate the quality of the clustering results, we use the confusion matrix since we know the ground truth labels

of each sequence. The clustering performance of different methods on collection 1 of real ECG time series is recorded as follows: the proposed method (94.29%), LPCC (85.71%), KF (42.9%), DFT (57.14%), and PCA (40%). Compared to the previous feature extraction methods, the average performance of applied HLDS on real ECG datasets demonstrates

significant performances, that is, 9.1%, 54.5%, 39.39%, and 57.57% clustering improvement against the LPCC, original Kalman filter, DFT, and PCA, respectively.

Since this method can discover deeper hidden patterns which can capture correlation and temporal dynamics successfully, it provides the group of distinct harmonics that helps to handle the presence of the time shift effect with small shifts in frequency. These harmonic groups represent good resulting features which lead to good clustering as well as visualization.

In more detail, Figure 1 shows the comparison of the first two extracted features of ECG dataset by the proposed method, LPCC, original Kalman filter, DFT, and PCA. In each heat map of Figure 1, the rows represent the number of sequences of dataset and the columns show the first two features extracted from each method. Only the proposed method clearly illustrates the characteristic of the two clusters, normal and malignant ventricular arrhythmia especially in the 1st extracted features in Figure 1(a), where the rows have similar feature values colored in orange and yellow, so that they are grouped in the same cluster. On the other hand, the four remaining methods, Figures 1(b), 1(c), 1(d), and 1(e), do not show meaningful interpreted features.

The proposed method gives wrong clustering results for the 17th and 32nd signals. Even though the 17th and 32nd signals are malignant ventricular arrhythmia signals actually, the shapes of these signals are very similar to the normal case. Therefore, when applying method, they discover similar features as the normal cases; consequently, they are clustered to the normal cluster.

To verify them, the 17th and 32nd time sequences of Figure 1(a) are plotted in Figure 2 where the patterns of 17th and 32nd from Figure 1 are in very similar shape compared to the normal signal shown in Figure 3(a) where Figure 3(b) is a sample signal of malignant ventricular arrhythmia case.

Figure 4 shows the scatter plots of the first feature and the second feature from different methods for ECG clustering. On the other hand, the first two of extracted features are plotted to evaluate how much these features contribute to ECG recognition. It is observed that the HLDS shows clear separation of dots compared to the other approaches.

The proposed method again performs the best clustering accuracy for collection 2 which consists of healthy and supraventricular arrhythmia persons. The error rates for all of experimental methods are shown as follows: proposed method (0.139), LPCC (0.2326), KF (0.4419), PCA (0.3256), and DFT (0.3953).

The computational complexity of the HLDS is shown in Figure 5, executed on collection 1. It can be observed that wall clock times lie on almost a straight line over the duration of sequences since it would improve the scalability in ECG dataset. The computation time of HLDS is satisfied when the time duration T of the time sequences is increased.

4. Conclusions

The proposed method, HLDS, considers the problem of handling the challenges of time series, namely, time shift,

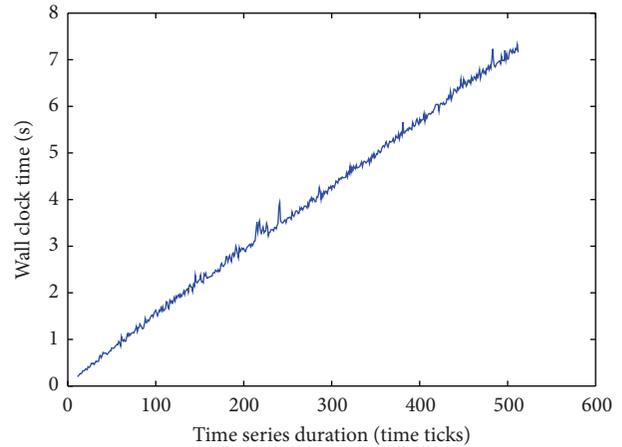


FIGURE 5: Linear execution time with respect to the length of sequences.

nearby frequencies, and harmonics. The applied method demonstrates the efficiency for solving these challenges in real applications of ECG time series domain. Interpretability of prominent features was discovered for the clustering as well as visualization. In most cases, HLDS gives the best result compared to the other feature extraction techniques such as Kalman filter, LPC cepstrum, DFT, and PCA. Moreover, the performance results show almost a linear speedup as we increase the input of the dataset.

For further study, we will investigate the harmonic linear dynamical system over much longer time series with missing values in various applications.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This research was supported by the MKE (The Ministry of Knowledge Economy), Korea, under the ITRC (Information Technology Research Center) support program supervised by the NIPA (National IT Industry Promotion Agency) (NIPA-2013-H0301-13-3005). This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2013-052849).

References

- [1] R. T. Olszewski, *Generalized feature extraction for structural pattern recognition in time series data [Ph.D. thesis]*, School of Computer Science, Carnegie Mellon University, 2001.
- [2] S. Papadimitriou and M. Vlachos, *Hand on Time Series Analysis with Matlab*, ICDM, Hong Kong, 2006.
- [3] H. S. Guan and Q. S. Jiang, "Cluster financial time series for portfolio," in *Proceedings of the International Conference on*

- Wavelet Analysis and Pattern Recognition (ICWAPR '07)*, pp. 851–856, November 2007.
- [4] R. Krishnapuram, A. Joshi, O. Nasraoui, and L. Yi, “Low-complexity fuzzy relational clustering algorithms for web mining,” *IEEE Transactions on Fuzzy Systems*, vol. 9, no. 4, pp. 595–607, 2001.
 - [5] X. Zhang, J. Liu, Y. Du, and T. Lv, “A novel clustering method on time series data,” *Expert Systems with Applications*, vol. 38, no. 9, pp. 11891–11900, 2011.
 - [6] C. S. Möller-Levet, F. Klawonn, K. H. Cho, and O. Wolkenhauer, “Fuzzy clustering of short time-series and unevenly distributed sampling points,” in *Advances in Intelligent Data Analysis V*, vol. 2810 of *Lecture Notes in Computer Science*, pp. 330–340, 2003.
 - [7] C. Goutte, L. K. Hansen, M. G. Liptrot, and E. Rostrup, “Feature-space clustering for fMRI meta-analysis,” *Human Brain Mapping*, vol. 13, no. 3, pp. 165–183, 2001.
 - [8] K. Kalpakis, D. Gada, and V. Puttagunta, “Distance measures for effective clustering of ARIMA time-series,” in *Proceedings of the 1st IEEE International Conference on Data Mining (ICDM '01)*, pp. 273–280, San Jose, Calif, USA, November–December 2001.
 - [9] Y. Xiong and D. Y. Yeung, “Time series clustering with ARMA mixtures,” *Pattern Recognition*, vol. 37, no. 8, pp. 1675–1689, 2004.
 - [10] T. Warren Liao, “Clustering of time series data—a survey,” *Pattern Recognition*, vol. 38, no. 11, pp. 1857–1874, 2005.
 - [11] I. T. Jolliffe, *Principal Component Analysis*, Springer, 2002.
 - [12] C. Ding and X. He, “K-means clustering via principal component analysis,” in *Proceedings of the 21st International Conference on Machine Learning (ICML '04)*, pp. 29–32, July 2004.
 - [13] L. Li, J. McCann, N. Pollard, and C. Faloutsos, “DynaMMo: mining and summarization of coevolving sequences with missing values,” in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '09)*, pp. 507–515, New York, NY, USA, July 2009.
 - [14] L. Li and B. A. Prakash, “Time series clustering: complex is simpler!” in *Proceedings of the 28th International Conference on Machine Learning (ICML '11)*, pp. 185–192, July 2011.
 - [15] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, San Diego, Calif, USA, 1990.
 - [16] Y. Wu, D. Agrawal, and A. E. Abbadi, “A comparison of DTF and DWT based similarity search in time series databases,” in *Proceeding of the 9th ACM International Conference on Information and Knowledge Management (CIKM '00)*, pp. 488–495, 2000.
 - [17] M. Jahangiri, D. Sacharidis, and C. Shahabi, “Shift-split: I/O efficient maintenance of wavelet-transformed multidimensional data,” in *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD '05)*, pp. 275–286, June 2005.
 - [18] F. Morchen, “Time series feature extraction for data mining using dwt and dft,” Tech. Rep. 33, Philipps-University Marburg, 2003.
 - [19] L. Rabiner and B. H. Juang, *Fundamentals of Speech Recognition*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1993.
 - [20] S. N. Navnath and S. H. Raghunath, “DWT and LPC based feature extraction methods for isolated word recognition,” *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2012, article 7, 2012.
 - [21] L. Li, B. Aditya Prakash, and C. Faloutsos, “Parsimonious linear fingerprint for time series,” *Proceedings of the VLDB Endowment*, vol. 3, pp. 385–396, 2010.
 - [22] L. Ralaivola and F. D’Alché-Buc, “Time series filtering, smoothing and learning using the kernel Kalman filter,” in *Proceedings of the International Joint Conference on Neural Networks (IJCNN '05)*, pp. 1449–1454, Montreal, Canada, August 2005.
 - [23] A. Jain, E. Y. Chang, and Y. F. Wang, “Adaptive stream resource management using Kalman Filters,” in *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD '04)*, pp. 11–22, June 2004.
 - [24] D. Gunopulos and G. Das, “Time series similarity measures and time series indexing,” in *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD '01)*, Santa Barbara, Calif, USA, May 2001.
 - [25] K. Deng, A. W. Moore, and M. C. Nechyba, “Learning to recognize time series: combining ARMA models with memory-based learning,” in *Proceedings of the 1997 IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA '97)*, pp. 246–251, July 1997.
 - [26] X. Wang, K. Smith, and R. Hyndman, “Characteristic-based clustering for time series data,” *Data Mining and Knowledge Discovery*, vol. 13, no. 3, pp. 335–364, 2006.
 - [27] E. J. Keogh, T. Palpanas, V. B. Zordan, D. Gunopulos, and M. Cardle, “Indexing large human-motion databases,” in *Proceedings of the 13th International Conference on Very Large Data Bases (VLDB '04)*, pp. 780–791, 2004.
 - [28] N. T. N. Anh, S. H. Kim, H. J. Yang, and S. H. Kim, “Hidden dynamic learning for long-interval consecutive missing values reconstruction in EEG time series,” in *Proceedings of the IEEE International Conference on Granular Computing (GrC '11)*, pp. 653–658, Kaohsiung, Taiwan, November 2011.
 - [29] R. E. Kalman, “A new approach to linear filtering and prediction problems,” *Transactions of the ASME Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.
 - [30] R. H. Shumway and D. S. Stoffer, “An approach to time series smoothing and forecasting using the EM algorithm,” *Journal of Time Series Analysis*, vol. 3, pp. 253–264, 1982.
 - [31] K. Kalpakis, D. Gada, and V. Puttagunta, “Distance measures for effective clustering of ARIMA time-series,” in *Proceedings of the 1st IEEE International Conference on Data Mining (ICDM '01)*, pp. 273–280, San Jose, Calif, USA, December 2001.
 - [32] H. Zhang, T. B. Ho, Y. Zhang, and M. S. Lin, “Unsupervised feature extraction for time series clustering using orthogonal wavelet transform,” *Informatica*, vol. 30, no. 3, pp. 305–319, 2006.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

