

Accurate Time Series Classification Using Partial Dynamic Time Warping

Haemwaan Sivaraks, Phongsakorn Sathianwiriya khun, Thapanan Janyalikit, Chotirat Ann Ratanamahatana

Abstract— Dynamic Time Warping (DTW) has been widely used in time series domain as a distance function for similarity search. Several works have utilized DTW to improve the classification accuracy as it can deal with local time shiftings in time series data by non-linear warping. However, some types of time series data do have several segments that one segment should not be compared to others even though DTW can naturally warp across those segments. In this paper, we propose PartialDTW distance measure that utilizes domain knowledge about special characteristics of different sections of the data to limit the warping path. The experiment shows that our PartialDTW has much better performance when compare with other well known algorithms.

Keywords- Dynamic time warping, DTW, PartialDTW, time series classification

I. Introduction

Dynamic Time Warping (DTW) is a powerful similarity measure for time series data that is widely used in many applications such as electrocardiogram (ECG) classification, financial prediction and biological classification [1-6]. DTW allows non-linear alignment so the optimal warping path between two time series can be found. This warping path can represent the corresponding regions and the similarity (or dissimilarity) between two time series [4]. The drawbacks of DTW include its $O(n^2)$ time complexity and possible invalid warping path from cross-segment alignment. Although DTW can exploit global constraints to limit warping path and somewhat decrease computation time, it is still problematic in some domains. DTW with global constraint only concerns about how much warping each individual data point is allowed to warp away. However, in some domains where each time series sequence is composed of two or more segments, only the data within the same segment is allowed to align to each other. For example, a typical DTW in ECG classification [3][5] cannot obtain the ideal warping path because a P-wave segment in one ECG could be aligned to a QRS-complex segment in another ECG, as shown in Fig.1. This is undeniably meaningless in clinical diagnosis since different portions of the ECG signal represent different activities of the heart.

Therefore, our main objective is to develop a PartialDTW distance measure that prevents unreasonable warpings by allowing only data points of the two time series sequences that belong to the same segments to be warped

together. Our PartialDTW uses prior knowledge about the data to complete a segmentation phase before the DTW distance calculation, which is performed partially according to the segments determined in the previous step.

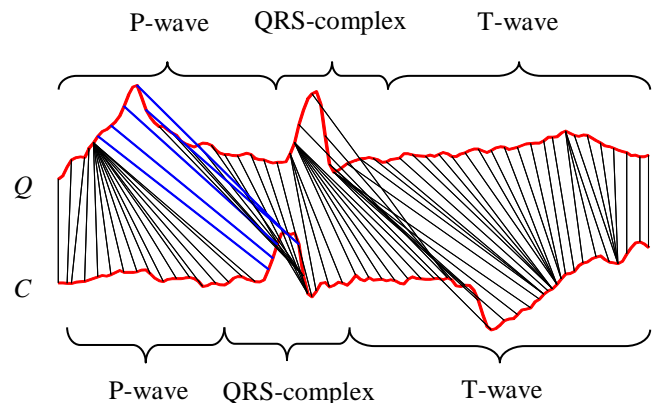


Figure 1. An example of an unreasonable DTW alignment between two ECG subsequences Q and C . P-wave of a sequence Q is aligned to QRS-complex portion of a sequence C .

The rest of the paper is organized as follows. The next section explains the DTW methodology and some related works. Section 3 provides details of our proposed PartialDTW. Section 4 provides the experiment setup details, evaluation, results, and discussion. The last section provides conclusion.

II. Related Work and Background

A. Related works

Dynamic time warping (DTW) is a similarity measure for time series data that is widely used in many applications. However, some data do contain two or more segments that should not be warped crossing segments to each other. Besides, a typical DTW has no mechanism to detect nor to separate segments and sometime mistakenly warp across them. For example, a single beat of normal ECG data consists of a P-wave, a QRS-complex, and a T-wave. Using a classic DTW to align two ECG sequences, there are many chances that a portion of P-wave will be warped to a portion of QRS-complex or a portion of QRS-complex will be warped to a portion of T-wave, which is considered illogical in clinical practice

There are some works that use DTW on ECG data [1][3][5] or other data that consist of many segments such as gun point [7]. However, DTW in numerous cases does warp across segments, which affects the classification accuracies. To improve the accuracy of classification, this work proposes a method called PartialDTW that performs a sequence segmentation and then utilizes the segment information to perform DTW alignment within the same segment of the two sequences.

Haemwaan Sivaraks, Phongsakorn Sathianwiriya khun,
Thapanan Janyalikit, Chotirat Ann Ratanamahatana
Department of Computer Engineering, Chulalongkorn University,
Phayathai Rd., Pathumwan, Bangkok, Thailand, 10330

B. Dynamic Time Warping (DTW)

Dynamic Time Warping (DTW) is one of the widely used similarity measures for time series data. DTW uses dynamic programming techniques to find optimal warping path. DTW can compare time series data with different lengths and handle non-linear alignments or local time shiftings [8]. Consequently, DTW is more suitable for time series domain than a basic method for similarity measures such as Euclidian distance.

Given two time series sequences, a sequence Q of length n and a sequence C of length m , as follows:

$$Q = q_1, q_2, \dots, q_i, \dots, q_n \quad (1)$$

$$C = c_1, c_2, \dots, c_j, \dots, c_m \quad (2)$$

A matrix n -by- m is created to hold cumulative distances between any pair of data point in the sequences Q and C . The cumulative distance $Y(i, j)$ is calculated from a sum of the distance in a current cell and the minimum of the cumulative distance of the three adjacent elements as follows:

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

where

$$d(q_i, c_j) = d(q_i - c_j)^2 \quad (4)$$

An optimal warping path is a path with the lowest warping cost from cumulative matrix as follows:

$$W = w_1, w_2, \dots, w_k, \dots, w_K \quad (5)$$

where

$$w_1 = (1,1) \text{ and } w_K = (m,n) \quad (6)$$

Finally, the DTW distance values can be calculated as follows:

$$DTW(Q, C) = \min \left\{ \sqrt{\sum_{k=1}^K w_k} \right\} \quad (7)$$

where w_k is the element $(i,j)_k$ of the matrix and also belongs to k^{th} element of a warping path W .

For example, given a subsequence $A = [3,1,2,3,2,4,5,2]$ and a subsequence $B = [2,1,3,2,1,4,2,3,5]$ their corresponding cumulative distance matrix is shown in Fig. 2 (a). An optimal warping path is shown in grey. The DTW distance is calculated as in (7). The illustration of subsequences A and B and their corresponding warping path is shown in Fig. 2 (b).

III. Proposed Work

Our PartialDTW framework consists of three major steps: segmentation, DTW distance calculation, and classification. An overview of our proposed work is shown in Fig. 3.

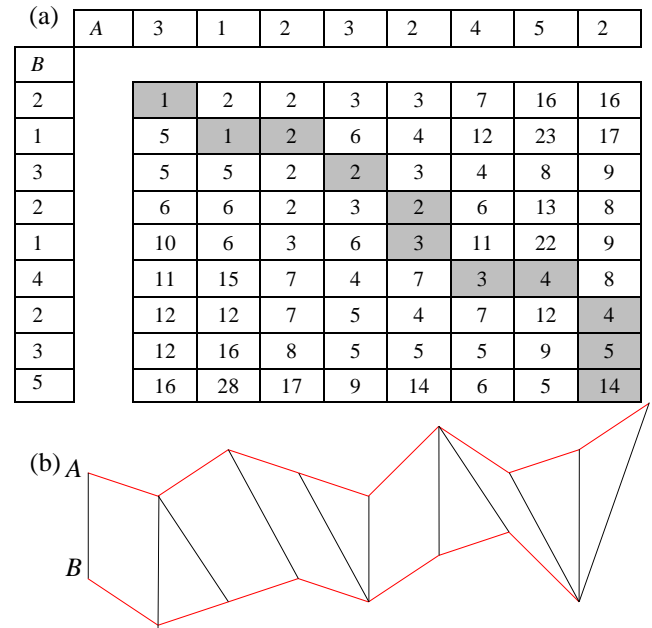


Figure 2. (a) The cumulative distance matrix between subsequences A and B. Grey cells indicate the warping path. (b) Corresponding alignment between subsequences A and B. Lines between two subsequences represent the warping path/alignment.

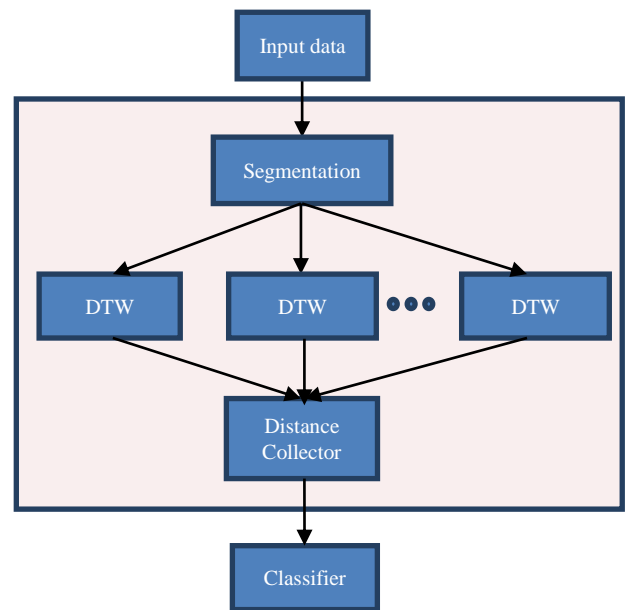


Figure 3. An overview of our proposed framework.

A. Segmentation

We segment the whole time series of length n into k segments according to its characteristics. Other time series sequences to be compared must also have the same number of k segments. However, each segment may have different lengths, and the number of data points within each segment may also vary across different sequences. This segmentation step is a crucial step in PartialDTW because we will need the domain knowledge to be able to segment the data properly.

B. Partial Dynamic Time Warping (PartialDTW)

Traditional DTW measures the distance between two time series sequences, by taking the whole time series into the cumulative distance calculation, as described in section II. However, in PartialDTW, we use domain knowledge to partition each time series sequence into k segments before performing k individual DTW calculations.

The idea behind PartialDTW is similar to the sliding window approach in [9][10], but instead of using the sliding window to find suitable subsequence for partial alignment, we use the domain knowledge to mark the partitions, which is evidently more accurate. PartialDTW then calculates DTW distance separately within each individual partition before combining the distances from all k partitions through the distance collector.

C. Classification

In this step, we employ the 1-Nearest Neighbor (1-NN) classifier, which is simple and widely used in time series classification tasks [11][12]. The 1-NN classification algorithm first determines the most similar candidate sequence to the query sequence and returns the class label as being the same class as that of the most similar candidate sequence. As shown in Fig. 4, if we want to obtain a class for the query sequence Q , we first calculate the distance between the query sequence Q and each candidate sequences 1, 2, 3, and 4. After that, the most similar candidate sequence to the query sequence which is indicated by the smallest distance will be obtained. Finally, the class A of sequence 2 is assigned to sequence Q .

iv. Experiments

In this section, we use 5 datasets to compare the accuracies using 1-NN with DTW distance and other well known classification algorithms including Naive Bayes (NB), Bayesian Network (BN), Multilayer Perceptron (MLP) and Support Vector Machine (SVM). The rival classification algorithms were run on the latest stable version of Weka (v. 3.6.12) [13]. We also report the speedup of PartialDTW over the traditional DTW by comparing their computation time.

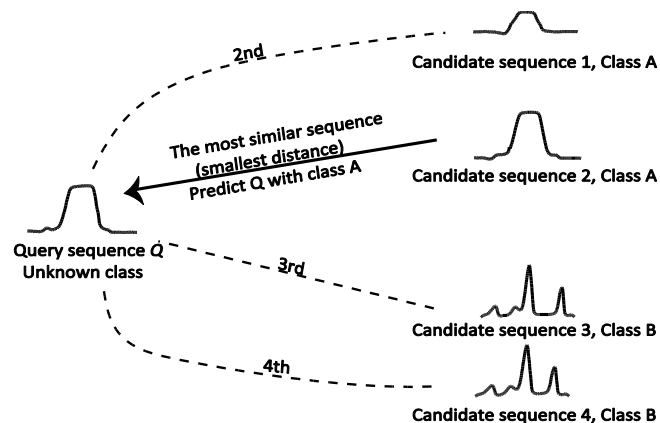


Figure 4. An example of using 1-NN classification to predict a class label of the sequence Q by labeling it with the class of its most similar sequence.

A. Experiment Setup

1) Datasets

We use 5 datasets, two from the UCR Time Series Classification/Clustering archive [14] and three from Physionet [15]. The first two are ECG and Gun-Point. The other three are extracted from the St. Petersburg Institute of Cardiological Technics: 12-lead Arrhythmia Database, a two-class database labeled as normal (class -1) or abnormal (class 1). In the pre-processing step, z-normalization is performed on all time series sequences so that every data sequence will have a zero mean and standard deviation of 1. The number of instances, the length of time series, and the number of classes are shown in Table I. For evaluation, we use the test sets provided by the archive in the first two datasets. However, Physionet's data do not have a train-test split. So, we use the leave-one-out cross-validation method to evaluate the performance of the data.

Table I. Details of 5 datasets used in our experiments.

Name	Number of classes	Size of training set	Size of test set	Time series Length
ECG	2	100	100	96
Gun-Point	2	50	150	150
INCART01	2	-	48	200
INCART02	2	-	168	202
INCART03	2	-	75	180

2) Data Segmentation

We segment the time series in each dataset using its domain knowledge. For example, in ECG data, following the expert's annotation, an ECG morphology consists of 5 partitions, segmented by a PQRST element; a P-wave appears before an R-Peak, followed by a minimum down peak called an S-peak, and ended with a T-wave as shown in Fig. 5. For the Gun-Point dataset, we segment the data into 3 parts indicating 1) the gun grasping (if any), 2) hand raising up to perform a pointing/shooting act followed by a hand lowering gesture, and 3) putting the gun back to the holster (if any) [7].

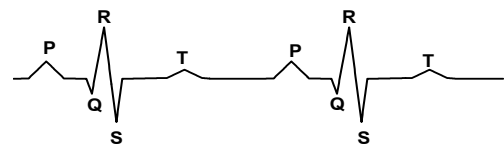


Figure 5. A morphology annotation in two beats of ECG data illustrating that a P-wave always appears first, followed by a QRS-complex, and ended with a T-wave in each beat.

B. Experiment Results and Discussion

We compare the performance of each algorithm by the classification accuracy. The results are shown in Table II, comparing our proposed PartialDTW among 5 other well-known classifiers, 1-NN with DTW, Naive Bayes, Bayesian Network, Multilayer Perceptron, and SVM. Our PartialDTW clearly outperforms all other algorithms by large margins. This is mainly because PartialDTW does exploit the domain knowledge in the segmentation step to make the data more compatible during the alignment.

Table II. Classification accuracy (%) for all five datasets using DTW, PartialDTW, Naïve Bayes, Bayesian Network, MLP and SVM. Bold figures denote the winning algorithm.

Method Dataset	1-NN DTW	1-NN Partial DTW	Naive Bayes	Bayesian Network	MLP	SVM
Gunpoint	90.60	96.00	78.67	85.33	93.33	76.67
ECG	77.00	84.00	77.00	75.00	82.00	81.00
INCART01	70.83	95.83	70.83	75.00	75.00	75.00
INCART02	82.73	98.81	82.74	78.57	95.83	80.95
INCART03	78.67	98.67	60.00	85.33	90.67	85.33

Specifically, PartialDTW separates the P-wave from the QRS-complex and the T-wave, such that when it looks for its nearest neighbor, it maps P-wave to only P-wave, QRS-complex to only QRS-complex, and T-wave to only T-wave, as illustrated in Fig. 6; no warping is allowed crossing different segments. On the contrary, in classic DTW, the first peak of an ECG sequence Q is an onset of a P-wave but is contaminated by noise and hence is warped to align with a part in the QRS-complex instead of a P-wave of another ECG sequence C .

We also compare the running time between DTW and PartialDTW, as shown in Fig. 7. INCART02 has the largest running time due to its longer sequence and larger test set. The speedup of PartialDTW is calculated by dividing the

overall running time of DTW by the running time of PartialDTW, and is shown in Table III.

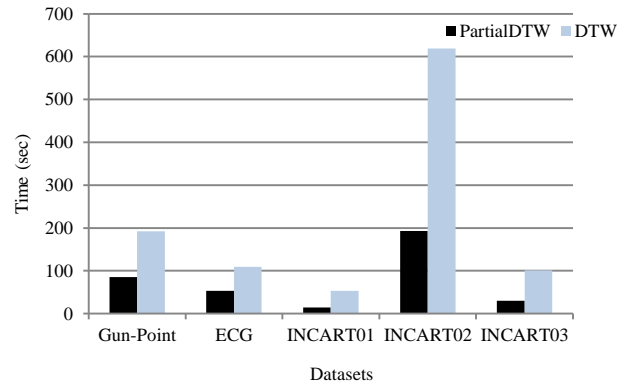


Figure 6. Time comparison between our proposed PartialDTW algorithm and DTW

Table III. Speedup of PartialDTW over a traditional DTW.

Dataset	Speedup of the computation time
Gun-Point	2.259
ECG	2.057
INCART01	3.786
INCART02	3.207
INCART03	3.367

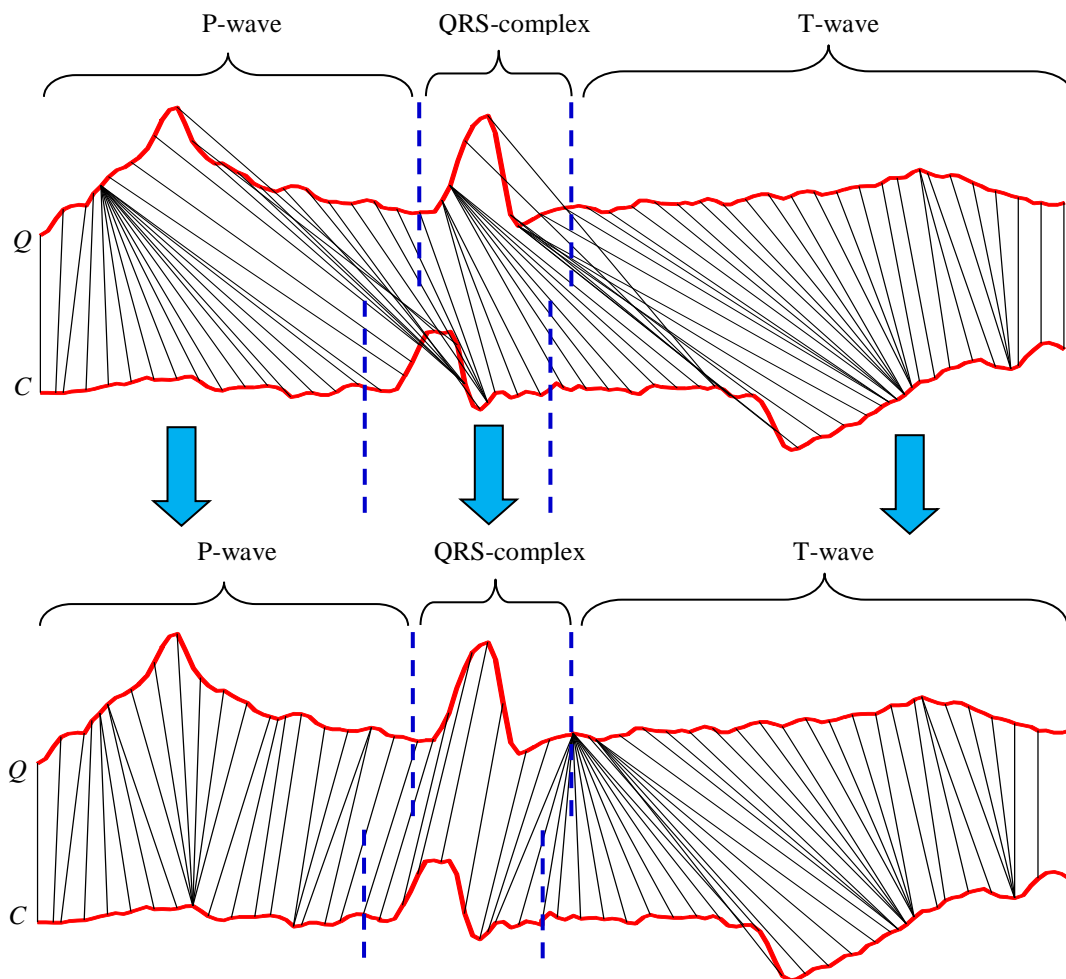


Figure 7. Comparison between classic DTW (top) and PartialDTW (bottom) for one beat of an ECG sequence. Our PartialDTW segments each beat of the ECG sequence into P-wave, QRS-Complex, and T-wave before warping, hence giving a more meaningful alignment and better accuracy.

For speedup, we can see that PartialDTW achieves large speedups in all datasets, especially in the last three datasets. This is because in the segmentation stage, the TP segment is discarded from the calculation according to the cardiologist's clinical diagnosis approach, decreasing the overall sequence's length by 25-35%, whereas the whole cardiac cycle length is used in traditional DTW.

V. Conclusion

In this paper, we propose the PartialDTW framework which is more robust than the traditional DTW. The essence of PartialDTW is in its segmentation phase where specific domain knowledge must be used to acquire accurate partitions for each time series sequence before computing its similarity. The warping alignment between time series sequence pair is then restricted to only within the same segment. The experiment results demonstrate that our PartialDTW can improve the performance, both in terms of accuracy and computation time, by outperforming many well known classifiers. Moreover, since PartialDTW calculates DTW within individual segment, we can exploit parallel computing to run all DTW calculations for all segments to further increase the speedup.

VI. Acknowledgements

This research is partially supported by Chulalongkorn University Dutsadi Phiphat Scholarship (to H. Sivaraks), CP Chulalongkorn Graduate Scholarship (to P. Sathianwiriyaakun) and the Thailand Research Fund and Chulalongkorn University given through the Royal Golden Jubilee Ph.D. Program (PHD/0057/2557 to T. Janyalikit).

VII. References

- [1] V. Tuzcu, and S. Nas, "Dynamic time warping as a novel tool in pattern recognition of ECG changes in heart rhythm disturbances," International Conference on Systems, Man and Cybernetics, vol.1 , 2005, pp.182-186.
- [2] J.Grabocka, E.Bedalli, and L. Schmidt-Thieme, "Efficient classification of long time-series", Advances in Intelligent Systems and Computing, vol. 207, pp.47-57, 2013.
- [3] G. P. Shorten, and M. J. Burke, "Use of dynamic time warping for accurate ECG signal timing characterization," Journal of Medical Engineering & Technology, vol. 38, pp. 188-201, 2014.
- [4] Y. Sakurai, M. Yoshikawa , and C. Faloutsos, "FTW: Fast similarity search under the time warping distance," Twenty-Fourth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, 2005, pp.326-337
- [5] G. Zhang, W. Kinsner, and B. Huang , "Electrocardiogram data mining based on frame classification by dynamic time warping matching," in Computer Methods in Biomechanics and Biomedical Engineering, vol. 12, pp.701-707, 2009.
- [6] H. Sivaraks and C. A. Ratanamahatana, "Robust and Accurate Anomaly Detection in ECG Artifacts Using Time Series Motif Discovery," Computational and Mathematical Methods in Medicine, vol. 2015, pp. 1-20, 2015.
- [7] C. A. Ratanamahatana, and E. Keogh. "Making Time-series Classification More Accurate Using Learned Constraints", in SIAM International Conference on Data Mining, 2004, pp. 11-22.
- [8] E. Keogh, and M. J. Pazzani, "Scaling up dynamic time warping for datamining applications," ACM SIGKDD international conference on Knowledge discovery and data mining, 2000, pp.285-289.
- [9] H. Li, "On-line and dynamic time warping for time series data mining," in International Journal of Machine Learning and Cybernetics, vol.6, pp.145-153, 2015.
- [10] R. Roopalakshmi, and G. Ram Mohana Reddy, "A novel spatio-temporal registration framework for video copy localization based on

multimodal features," in Signal Processing, vol.93, pp.2339-2351, 2013.

- [11] V. Kurbalija, M. Radovanović, Z. Geler, and M. Ivanović, "The influence of global constraints on similarity measures for time-series databases" in Knowledge-Based Systems, vol. 56, pp. 49-67, 2014.
- [12] X. Xi, E. Keogh, C. Shelton, L. Wei, and C. A. Ratanamahatana, "Fast time series classification using numerosity reduction," International conference on Machine learning, 2006, pp. 1030-1040.
- [13] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I. H. Witten "The WEKA Data Mining Software: An Update," SIGKDD Explorations, Volume 11, Issue 1, 2009.
- [14] E. Keogh, Q. Zhu, B. Hu, Y. Hao, X. Xi, L. Wei and C.A. Ratanamahatana. The UCR Time Series Classification/Clustering. 2011.
Homepage: www.cs.ucr.edu/~eamonn/time_series_data/
- [15] <http://www.physionet.org>.
- [16] Y. Chen, B. Hu, E. Keogh, and G. Batista, "DTW-D: Time series semi-supervised learning from a single example," ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2013, pp.383-39

About Authors :



Haemwaan Sivaraks received her B.Sc. from Thammasat University and received her M.Sc. and Ph.D. degrees from Dept. of Computer Engineering, Chulalongkorn University, where she was granted the Chulalongkorn University's Dutsadi Phiphat Scholarship. Her research interests include time series data mining, machine learning and artificial intelligence.



Phongsakorn Sathianwiriyaakun is a master's degree student in Computer Engineering at Chulalongkorn University. He received his B.Eng. in Computer Engineering from Chulalongkorn University. He was granted scholarships from CP Chulalongkorn Graduate Scholarship. His research interest is time series data mining.



Thapanan Janyalikit is a Ph.D. student in Computer Engineering at Chulalongkorn University. He received his B.Eng. in Computer Engineering from King Mongkut's University of Technology North Bangkok. He was granted a scholarship from the Thailand Research Fund through the Royal Golden Jubilee Ph.D. Program. His research interest is time series data mining.



Chotirat Ann Ratanamahatana is an assistant professor in Computer Engineering Department, Chulalongkorn University, Bangkok, Thailand. She received her undergraduate and graduate studies from Carnegie Mellon University and Harvard University, respectively, and received her Ph.D. from the University of California, Riverside. Her research interests include data mining, information retrieval, machine learning, and human-computer interaction.