Temporal data clustering via weighted clustering ensemble with different representations

Yun Yang and Ke Chen, *Senior Member*, *IEEE*

Appendix 1

In this appendix, we present our empirical studies on investigating the capability and the limitation of our weighted clustering ensemble (WCE) algorithm based on the algorithm analysis described in Sect. 3.3.

As pointed out in Sect. 3.3, (13) critically determines the performance of our WCE via the quantities $|\mu_m - w_m|$. As a result, we need both μ_m and w_m for a given data set $X = \{\mathbf{x}_n\}_{n=1}^N$. While w_m is achieved by applying one of clustering validation criteria or their combination to input partitions, μ_m is generally unavailable unless we know both the ground-truth partition and all possible partitions of the given data set.

In reality, there are only a subset of partitions, $\mathbf{P} = \{P_m\}_{m=1}^M$, returned by initial clustering analysis, we approximate μ_m by using only these partitions via a partition similarity measure, Normalized Mutual Information (NMI), although other similarity measures can be used. Thus, we estimate μ_m corresponding to P_m by

$$\hat{\mu}_m = \frac{\mathrm{NMI}(P_m, C)}{\sum_{m=1}^M \mathrm{NMI}(P_m, C)}, \tag{A.1}$$

where C is the ground-truth partition of X and

$$\mathrm{NMI}(P_m, C) = \frac{\sum_{i=1}^{K_m}\sum_{j=1}^{K_C} N_{ij}^{mc} \log(\frac{N N_{ij}^{mc}}{N_i^a N_j^b})}{\sum_{i=1}^{K_m} N_i^m \log(\frac{N_i^m}{N}) + \sum_{j=1}^{K_c} N_j^c \log(\frac{N_j^c}{N})}. \tag{A.2}$$

Here, K_m and K_C are the number of clusters in P_m and C, respectively. N_{ij}^{mc} is the number of entities shared by two clusters $C_i^m \in P_m$ and $C_j^c \in C$, where there are N_i^m and N_j^c entities in C_i^m and C_j^c.

As described in Sect. 3.1.B, we employ three clustering validation criteria to generate weights but do not combine three weighted similarity matrices directly. Instead we combine only three optimal partitions yielded by cutting dendrogram trees constructed with three weighted similarity matrix at the longest lifetime. Strictly speaking, the similarity matrix of the final consensus partition is a binary version of the multiple-criteria based weighted similarity matrix by applying a threshold. Nevertheless, we firmly believe that this binary version inherits most characteristics of the original multiple-criteria based weighted similarity matrix. Hence, we approximate its weight by

$$\overline{w}_m = \frac{1}{3}\sum_{\pi} w_m^{\pi}, \tag{A.3}$$

where π = {DVI, MHT, NMI} defined in Sect. 3.1.

In general, data distribution and underlying cluster shapes may be arbitrarily complex. Therefore, it is impossible to examine all kinds of data sets exhaustedly. In our empirical studies, we apply two of the most important cluster properties, *compactness* and *separability*, as a guideline to produce data sets. As the Gaussian mixture model (GMM) can approximate any kind of distribution, we employ the GMM of four Gaussian components to produce data sets of four clusters (K^*=4). By altering parameters, mean, co-variance and mixture proportion, in the GMM, we can produce data sets of different data distributions and cluster shapes. For visualization, we produce three 2-D data sets of clearly distinct properties.

In our experiments, we use K-mean in initial clustering analysis (ICA) with the procedure described in Sect. 4.2; K is randomly chosen from a range $1 \le K \le 8$ and 20 partitions are produced on different initial conditions for a given data set. We name such an initial clustering analysis ICA1. To simulate the limitation of initial clustering analysis, we also use another range of K: $1 \le K \le 8$ and $K \ne K^*$ to produce 20 partitions for a given data set. We denote such an ICA to be ICA2. The purpose of our experiment is two-fold: investigating the capacity and the limitation of our WCE and verifying the benefit of using multiple validation indexes (MVI).

As shown in Fig. A.1(a). Dataset 1 can be viewed as a representative of a class of data sets that have the tight compactness and the high separability, an easy task for clustering analysis. Therefore, such properties should be easily captured with any clustering validation criterion. As observed in Fig. A.1(b) and A.1(c), our WCE based on the MVI yields nearly perfect partitions even when the initial clustering analysis returns no correct partitions given that the fact that there are no partitions of four clusters retuned by ICA2. As expected, the use of a single criterion in the WCE is enough to produce satisfactory partitions as illustrated in Fig. A.1(d)-A.1(i) except Fig. A.1(h) but a single criterion is not robust against an inadequate initial clustering analysis as demonstrated in Fig. A.1(h) where the WCE based on the MHT criterion only fails to produce a correct partition. Fig. A.1(j) and A.1(k) show the dissimilarity between the ideal weight defined with the ground truth partition, μ_m defined in (A.1), and a weight based on one or more clustering validation criteria, w_m, collectively. It is observed from Fig. A.1(j) and A.1(k) that the dissimilarity between μ_m and $\overline{w}_m$ defined in (A.3) is much smaller than that between μ_m and

w_m^{π} overall, while the dissimilarity between μ_m and w_m^{π}, π = {DVI, MHT, NMI}) varies across 20 partitions. According to our algorithm analysis in Sect. 3.3, (13) suggests that the smaller collective dissimilarity between μ_m and w_m results in a lower cost. Therefore, experimental results here confirm the benefit of using MVI to measure the contribution of a partition for combination. In addition, Fig. A1.(k) shows that overall the dissimilarity between μ_m and w_m^{MHT} is significantly larger than others. This explains why the use of the MHT criterion only in the WCE fails to produce a correct partition by combining 20 partitions returned from ICA2.

Dataset 2 shown in Fig. A.2(a) is different from Dataset 1 in terms of the number of entities in different clusters, compactness and separability. Although this data set still has some identifiable properties, the intra-cluster variability gets higher and the inter-cluster variability becomes lower in contrast to Dataset 1, which leads to a difficult clustering analysis task. From Fig. A.2(b) and A.2(c), it is observed that our WCE yields a satisfactory partition and detects the correct number of clusters by combining partitions returned by ICA1 but fails to produce a partition of the intrinsic structure by combining partitions returned by ICA2. The ambiguity arises when the separability between different clusters is low. Incorrect initial clustering analysis inevitably misleads the clustering ensemble to produce a wrong partition due to ambiguity. This result suggests that initial clustering analysis plays a critical role particularly when there appears low separability between different clusters. In other words, a clustering ensemble itself cannot detect the intrinsic structure underlying a data set unless input partitions carry such information. As the ambiguity appears, the WCE based on a single criterion is no longer reliable; it is seen from Fig. A.2(d)-Fig. A.2(f) that the WCE based on the MHT criterion yields a partition of four clusters but the WCE based on the DVI and the NMI criteria produces two different partitions of three clusters. It implies that due to the ambiguity a single criterion does not always recognize partitions of the intrinsic structure even though an initial clustering analysis returns such partitions. Again, this evidence justifies our motivation on the joint use of multiple clustering validation criteria in our weighting scheme. Likewise, the WCE based on a single criterion fails to produce a partition of four clusters owing to the same reason seen for the partition shown in Fig. A.2(c), which is illustrated in Fig A.2(g)-A.2(i). Fig. A.2(j) and A.2(k) shows the dissimilarity between μ_m and w_m of 20 partitions returned from ICA1 and ICA 2, respectively. It is clearly shown from Fig. A.2(j) and A.2(k) that the larger dissimilarity appears as a clustering validation criterion mismatches the intrinsic structure underlying a given data set.

Fig. A.3(a) shows Dataset 3 of the ground truth that has no identifiable properties and is full of ambiguity without the reference to the ground truth. In particular, the intra-cluster variability is far higher than the inter-cluster variability. Due to a lack of no identifiable properties, neither a clustering algorithm nor a clustering validation criterion works on such a data set. As anticipated, the WCE based on either a single criterion or the multiple criteria fails to yield a partition close to the ground truth as shown in Fig. A.3(b)-Fig.3(i). Also our weight dissimilarity index values illustrated in Fig. A.3(j) and A.3(k) clearly indicate the reason of failure.

In summary, the above experiment results suggest that our WCE based on multiple validation criteria performs well but heavily relies on the quality of input partitions returned by initial clustering analysis, in particular, when a given data set is of fewer identifiable cluster structural information, e.g., low separability, uneven size of clusters, high intra-cluster and low inter-class variability. In general, our empirical studies are consistent with our algorithm analysis presented in Sect.3.3. As demonstrated in plots (j) and (k) of Fig. A.1-A.3, the dissimilarity between the optimal "weights" μ_m and "weights" w_m generated via clustering validation criteria becomes a useful measure to understand the behaviors of our WCE algorithm for different data sets. On the other hand, experimental results further justify the benefit of using multiple clustering validation criteria in our weighting scheme when the ground truth is not available.

Appendix 2

In this appendix, we employ two common evaluation criteria for clustering analysis to assess the performance of all clustering algorithms used in Sect. 4.2 for time series data mining benchmarks.

One is the normalized mutual information (NMI) defined in (A.1) and (A.2). The other is the normalized Rand Index (ARI) defined by

$$\mathrm{ARI}(P_m, C) = \frac{\sum_{i,j}\binom{N_{ij}}{2} - \left[\sum_i \binom{N_i}{2}\sum_j \binom{N_j}{2}\right] / \binom{N}{2}}{\frac{1}{2}\left[\sum_i \binom{N_i}{2} + \sum_j \binom{N_j}{2}\right] - \left[\sum_i \binom{N_i}{2}\sum_j \binom{N_j}{2}\right] / \binom{N}{2}}. \quad \text{(A.4)}$$

Here, N is the number of data points in a given data set and N_{ij} is the number of data points of the label C_j assigned to clusters i in partition P_m. N_i is the number of data points is in cluster i of partition P_m and N_j is the number of data points in class j. In general, an ARI value lies between 0 and 1. The index value is equal to one only if a partition is completely identical to the intrinsic structure and close to 0 for a random partition.

Corresponding to Table 2 in Sect. 4.2, the performance of different clustering algorithms is tabulated in Tables A.1 and A.2 in terms of the NMI and the ARI measures. Likewise, Tables A.3 and A.4 list the performance of four clustering ensembles by the NMI and the ARI measures, which is the counterpart of Table 3 in Sect. 4.2. The notation here is the same as used in Tables 2 and 3 in Sect. 4.2.

A direct comparison between those counterparts, i.e., Table 2 versus Tables A.1 and A.2 as well as Table 3 versus Tables A.3 and A.3, shows that the performance measured by three different evaluation criteria is completely consistent each other for different clustering algorithms and four clustering ensemble algorithms on all 16 time series data mining benchmark tasks. Thus, all conclusions drawn in Sect. 4.2 are supported and augmented by additional evaluation results in Tables A.1-A.4.

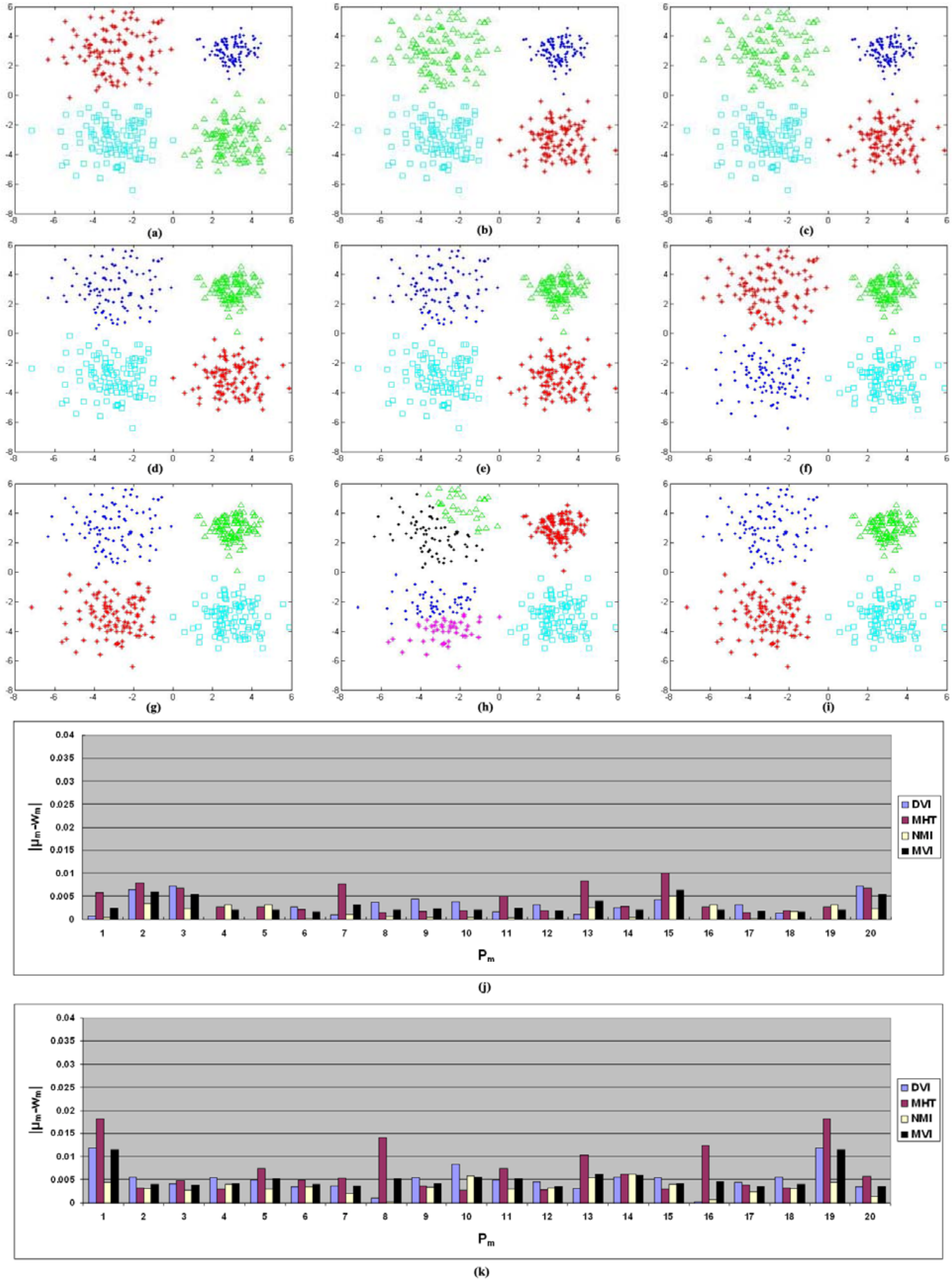

Fig. A.1. Results on Dataset 1. (a) Ground truth. (b) Partition by WCE based on multiple validation indexes (K-mean, $1 \le K \le 8$). (c) Partition by WCE based on multiple validation indexes (K-mean, $1 \le K \le 8$ and $K \neq 4$) (d)-(f) Partitions by WCE based on DVI, MHT and NMI (K-mean, $1 \le K \le 8$). (g)-(i) Partitions by WCE based on DVI, MHT and NMI (K-mean, $1 \le K \le 8$ and $K \neq 4$). (j) Dissimilarity between $\boldsymbol{w}_m$ and $\boldsymbol{\mu}_m$ of 20 partitions (K-mean, $1 \le K \le 8$). (k) Dissimilarity between $\boldsymbol{w}_m$ and $\boldsymbol{\mu}_m$ of 20 partitions (K-mean, $1 \le K \le 8$ and $K \neq 4$).

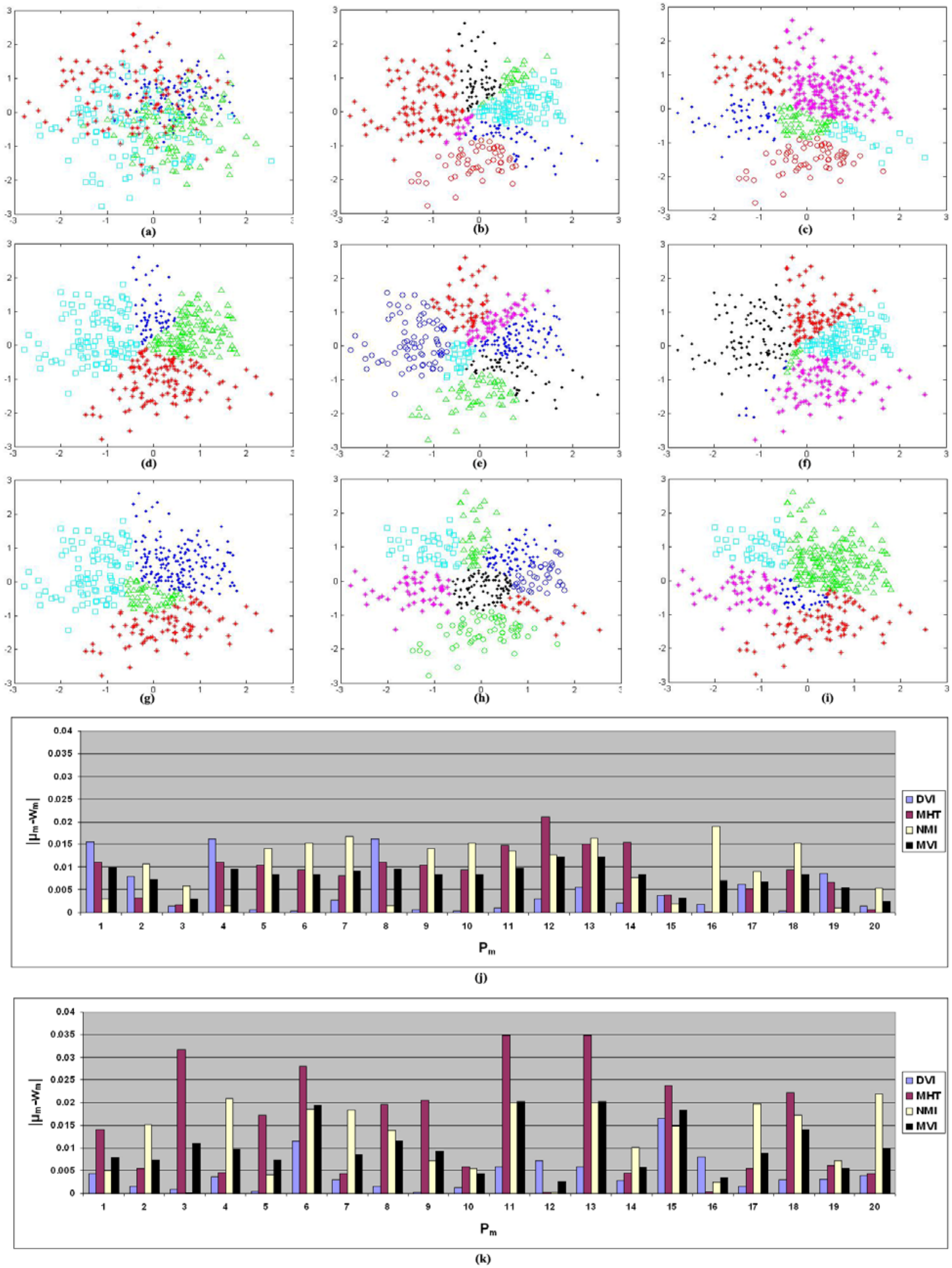

Fig. A.3. Results on Dataset 3. (a) Ground truth. (b) Partition by WCE based on multiple validation indexes (K-mean, $1 \le K \le 8$). (c) Partition by WCE based on multiple validation indexes (K-mean, $1 \le K \le 8$ and $K \neq 4$) (d)-(f) Partitions by WCE based on DVI, MHT and NMI (K-mean, $1 \le K \le 8$). (g)-(i) Partitions by WCE based on DVI, MHT and NMI (K-mean, $1 \le K \le 8$ and $K \neq 4$). (j) Dissimilarity between $\boldsymbol{w}_m$ and $\boldsymbol{\mu}_m$ of 20 partitions (K-mean, $1 \le K \le 8$). (k) Dissimilarity between $\boldsymbol{w}_m$ and $\boldsymbol{\mu}_m$ of 20 partitions (K-mean, $1 \le K \le 8$ and $K \neq 4$).

TABLE A.1
ARI ACHIEVED BY DIFFERENT CLUSTERING ALGORITHMS

Data Set	Time Series			Single Representation								Different Representations		
				K-mean				DBSCAN				WCE		
	K-mean	HC	K-HMM	PCF	DFT	PLS	PDWT	PCF	DFT	PLS	PDWT	K-mean	HC	DBSCAN
Syn Control	0.663	0.502	0.675	0.488	0.583	0.625	0.670	0.318	0.689*	0.692*	0.463	**0.862±0.018***	0.711*	0.749*
Gun-Point	0.477	0.376	0.405	0.410	0.397	0.421	0.439	0.432*	0.461*	0.398	0.448*	0.492±0.019*	**0.565***	0.481*
CBF	0.671	0.411	0.653	0.580	0.402	0.532	0.621	0.589*	0.405	0.423	0.673*	0.683±0.024*	**0.707***	0.676*
Face (all)	0.514	0.330	0.519	0.329	0.311	0.330	0.390	0.273	0.464	0.290	0.337	**0.751±0.019***	0.724*	0.415
OSU Leaf	0.544	0.551	0.584	0.492	0.421	0.490	0.514	0.280	0.353	0.376	0.559*	0.598±0.035*	0.604*	**0.614***
Swedish Leaf	0.491	0.417	0.454	0.411	0.399	0.424	0.450	0.156	0.197	0.387	0.392	**0.553±0.026***	0.504*	0.449*
50Words	0.412	0.385	0.396	0.367	0.482	0.405	0.363	0.315	0.291	0.309	0.194	0.372±0.021	**0.451***	0.345
Trace	0.535	0.480	0.563	0.488	0.490	0.492	0.519	0.405	0.496	0.526	0.550	0.582±0.019*	0.595*	**0.617***
Two Patterns	0.308	0.291	0.310	0.229	0.238	2.568	0.298	0.224	0.235	0.181	0.169	0.319±0.023*	**0.358***	0.264
Wafer	0.620	0.441	0.623	0.482	0.519	0.470	0.605	0.674*	0.452*	0.472	0.240	0.657±0.026*	0.625*	**0.686***
Face (four)	0.493	0.432	0.497	0.389	0.400	0.422	0.498	0.161	0.164	0.206	0.386	**0.692±0.031***	0.504*	0.400
Lightning-2	0.531	0.540	0.512	0.430	0.442	0.459	0.493	0.556*	0.496	0.389	0.375	**0.572±0.017***	0.561*	0.559*
Lightning-7	0.637	0.489	0.651	0.500	0.510	0.632	0.644	0.546	0.398	0.621*	0.425	**0.738±0.041***	0.668*	0.655*
ECG	0.640	0.510	0.641	0.506	0.526	0.531	0.580	0.434*	0.440*	0.429*	0.524*	0.620±0.019*	0.645*	**0.669***
Adiac	0.434	0.240	**0.440**	0.242	0.237	0.265	0.260	0.262	0.176	0.275	0.191	0.381±0.029	0.298	0.323*
Yoga	0.461	0.351	0.426	0.482	0.469	0.382	0.445	0.368	0.512*	0.294	0.331	0.495±0.026*	0.503*	**0.527**

TABLE A.2
NMI ACHIEVED BY DIFFERENT CLUSTERING ALGORITHMS

Data Set	Time Series			Single Representation								Different Representations		
				K-mean				DBSCAN				WCE		
	K-mean	HC	K-HMM	PCF	DFT	PLS	PDWT	PCF	DFT	PLS	PDWT	K-mean	HC	DBSCAN
Syn Control	0.489	0.412	0.497	0.398	0.451	0.460	0.479	0.253	0.507*	0.513*	0.330	**0.726 ± 0.021***	0.535*	0.621*
Gun-Point	0.450	0.381	0.409	0.413	0.401	0.445	0.430	0.428*	0.465*	0.404	0.452*	0.483 ± 0.024*	**0.521***	0.476*
CBF	0.559	0.424	0.551	0.539	0.411	0.498	0.550	0.547*	0.215	0.441	0.568*	0.619 ± 0.028*	**0.655***	0.606*
Face (all)	0.297	0.208	0.301	0.224	0.196	0.210	0.292	0.060	0.287	0.065	0.210	**0.451 ± 0.031***	0.442*	0.260
OSU Leaf	0.238	0.244	0.279	0.209	0.200	0.220	0.225	0.080	0.117	0.159	0.251*	0.297 ± 0.018*	0.303*	**0.313***
Swedish Leaf	0.393	0.361	0.374	0.366	0.338	0.368	0.368	0.163	0.253	0.308	0.326	**0.521 ± 0.030***	0.439*	0.370*
50Words	0.524	0.503	0.512	0.400	0.464	0.369	0.483	0.372	0.354	0.359	0.314	0.491 ± 0.026	**0.569***	0.466
Trace	0.391	0.334	0.410	0.309	0.353	0.346	0.385	0.267	0.379	0.388	0.396	0.419 ± 0.023*	0.432*	**0.471***
Two Patterns	0.217	0.210	0.222	0.199	0.210	0.208	0.212	0.183	0.206	0.159	0.156	0.228 ± 0.019*	**0.236***	0.210
Wafer	0.291	0.227	0.293	0.260	0.281	0.251	0.289	0.403*	0.240*	0.274	0.157	0.381 ± 0.027*	0.300*	**0.421***
Face (four)	0.521	0.502	0.536	0.321	0.386	0.477	0.527	0.146	0.118	0.241	0.318	**0.648 ± 0.031***	0.553*	0.358
Lightning-2	0.529	0.533	0.492	0.438	0.447	0.461	0.472	0.542*	0.475	0.378	0.301	**0.686 ± 0.024***	0.656*	0.652*
Lightning-7	0.451	0.382	0.502	0.385	0.396	0.449	0.482	0.403	0.251	0.430*	0.373	**0.711 ± 0.029***	0.529*	0.516*
ECG	0.462	0.371	0.466	0.362	0.394	0.402	0.417	0.205*	0.207*	0.101*	0.394*	0.458 ± 0.019*	0.467*	**0.474***
Adiac	0.591	0.443	**0.598**	0.448	0.438	0.460	0.455	0.460	0.329	0.500	0.388	0.568 ± 0.023	0.505	0.513*
Yoga	0.487	0.407	0.468	0.533	0.505	0.437	0.479	0.425	0.602*	0.383	0.397	0.572 ± 0.015*	0.580*	**0.614**

TABLE A.3
ARI PERFORMANCE OF DIFFERENT ENSEMBLE ALGORITHMS

Data Set	CE	HBGF	SDP-CE	WCE
Syn Control	0.679±0.017	0.729±0.021	0.815+0.018	**0.862±0.018***
Gun-Point	0.481±0.014	0.488±0.023	0.472±0.007	**0.492±0.019***
CBF	0.590±0.026	0.698±0.017	**0.704±0.018**	0.683±0.024*
Face (all)	0.510±0.025	0.650±0.021	0.743±0.018	**0.751±0.019***
OSU Leaf	0.528±0.018	**0.617±0.025**	0.609±0.018	0.598±0.035*
Swedish Leaf	0.493±0.014	0.496±0.025	**0.581±0.021**	0.553±0.026*
50Words	**0.394±0.020**	0.379±0.019	0.380±0.019	0.372±0.021
Trace	0.558±0.017	0.511±0.017	0.576±0.018	**0.582±0.019***
Two Patterns	0.315±0.025	0.311±0.020	0.318±0.021	**0.319±0.023***
Wafer	0.618±0.029	**0.661±0.029**	0.648±0.025	0.657±0.026*
Face (four)	0.471±0.031	0.502±0.021	0.500±0.028	**0.692±0.031***
Lightning-2	0.528±0.024	0.520±0.014	0.550±0.012	**0.572±0.017***
Lightning-7	0.655±0.024	0.662±0.035	0.691±0.022	**0.738±0.041***
ECG	0.593±0.022	0.616±0.016	**0.624±0.014**	0.620±0.019*
Adiac	0.380±0.028	0.536±0.026	**0.549±0.017**	0.381±0.029
Yoga	0.439±0.025	0.483±0.024	**0.541±0.020**	0.495±0.026*

TABLE A.4
NMI PERFORMANCE OF DIFFERENT ENSEMBLE ALGORITHMS

Data Set	CE	HBGF	SDP-CE	WCE
Syn Control	0.501±0.023	0.606±0.018	0.691+0.016	**0.726 ± 0.021***
Gun-Point	0.465±0.016	0.476±0.019	0.459±0.010	**0.483 ± 0.024***
CBF	0.549±0.023	0.642±0.015	**0.651±0.024**	0.619 ± 0.028*
Face (all)	0.290±0.028	0.390±0.026	0.444±0.011	**0.451 ± 0.031***
OSU Leaf	0.311±0.013	**0.433±0.022**	0.412±0.015	0.297 ± 0.018*
Swedish Leaf	0.401±0.012	0.429±0.028	**0.542±0.017**	0.521 ± 0.030*
50Words	**0.510±0.018**	0.501±0.024	0.497±0.021	0.491 ± 0.026
Trace	0.404±0.021	0.385±0.016	0.414±0.015	**0.419 ± 0.023***
Two Patterns	0.225±0.020	0.224±0.018	0.226±0.017	**0.228 ± 0.019***
Wafer	0.588±0.021	**0.690±0.024**	0.608±0.019	0.381 ± 0.027*
Face (four)	0.513±0.028	0.550±0.028	0.548±0.022	**0.648 ± 0.031***
Lightning-2	0.510±0.024	0.502±0.016	0.619±0.015	**0.686 ± 0.024***
Lightning-7	0.510±0.031	0.520±0.029	0.593±0.028	**0.711 ± 0.029***
ECG	0.533±0.020	0.543±0.016	**0.569±0.015**	0.458 ± 0.019*
Adiac	0.458±0.021	0.562±0.019	**0.588±0.014**	0.568 ± 0.023
Yoga	0.475±0.021	0.561±0.019	**0.621±0.018**	0.572 ± 0.015*