Supplementary Material: Learning Representations for Time Series Clustering

Qianli Ma
South China University of Technology
Guangzhou, China
qianlima@scut.edu.cn

Jiawei Zheng*
South China University of Technology
Guangzhou, China
csjwzheng@foxmail.com

Sen Li *
South China University of Technology
Guangzhou, China
awslee@foxmail.com

Garrison W. Cottrell
University of California, San Diego
CA, USA
gary@ucsd.edu

A Dataset Introduction

In Section 4, we report the experimental results of DTCR on 36 UCR datasets [1]. Here, we show the statistics of these 36 datasets.

Table 1: Statistics of the benchmark time series datasets

No.	Dataset	#Train/Test	Length	#classes	No.	Dataset	#Train/Test	Length	#classes
1	Arrow	36/175	252	3	19	Mid.phal.outl.correct	291/600	81	2
2	Beef	30/30	471	5	20	Mid.phal.TW	154/399	81	6
3	BeetleFly	20/20	513	2	21	MoteStrain	20/1252	85	2
4	BirdChicken	20/20	513	2	22	OSULeaf	200/242	428	6
5	Car	60/60	578	4	23	Plane	105/105	145	7
6	ChlorineConcentration	467/3840	167	3	24	Prox.phal.outl.ageGroup	400/205	81	3
7	Coffee	28/28	287	2	25	Prox.phal.TW	205/400	81	6
8	DiatomsizeReduction	16/306	346	4	26	SonyAIBORobotSurface	20/601	71	2
9	Dist.phal.outl.agegroup	139/400	81	3	27	SonyAIBORobotSurfaceII	27/953	66	2
10	Dist.phal.outl.correct	276/600	81	2	28	SwedishLeaf	500/625	129	15
11	ECG200	100/100	97	2	29	Symbols	25/995	399	6
12	ECGFiveDays	23/861	137	2	30	ToeSegmentation1	40/228	278	2
13	GunPoint	50/150	151	2	31	ToeSegmentation2	36/130	344	2
14	Ham	109/105	432	2	32	TwoPatterns	1000/4000	129	4
15	Herring	64/64	513	2	33	TwoLeadECG	23/1139	83	2
16	Lighting2	60/61	638	2	34	Wafer	1000/6164	153	2
17	Meat	60/60	449	3	35	Wine	57/54	235	2
18	Mid.phal.outl.agegroup	154/400	81	3	36	WordsSynonyms	267/638	271	25

B Details of Baseline Methods

We compare DTCR with 11 recently representative time series clustering methods and 2 state-of-the-art non time series deep clustering methods (DEC [2], IDEC [3]). The details are as follows:

- K-means: Use K-means on the entire time series.
- UDFS [4]: Unsupervised discriminative feature selection that simultaneously explores the manifold structure, local discriminative information, and feature correlations.
- NDFS [5]: Non-negative discriminative feature selection that adopts $l_{2,1}$ regularised regression and non-negative spectral analysis as a joint framework for selecting features.

*Two authors have equal contribution.

33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

- RUFS [6]: Robust unsupervised feature selection that uses robust orthogonal non-negative matrix factorization to perform feature learning jointly.
- RSFS [7]: Robust spectral learning for unsupervised feature selection, which joins spectral regression with sparse graph embedding.
- KSC [8]: Uses K-means for clustering by adopting a pairwise scaling distance measure and computing the spectral norm of a matrix for centroid computation.
- k-DBA [9]: Adopts K-means and dynamic time warping distance to obtain centroids, via a DBA method.
- k-shape [10]: Adopts a scalable iterative refinement procedure to explore the shapes of time series that have a normalized cross-correlation measure.
- u-shapelet [11]: A time series clustering method that deliberately ignores the rest of the data and only uses local patterns to cluster the time series.
- DTC [12]: Takes the KL divergence between predicted and target distribution as guidance to learn non-linear features in a deep framework.
- USSL [13]: Integrates the strengths of shapelet learning, shapelet regularization, spectral analysis, and pseudo-labels to help to cluster unlabeled time series better.
- DEC [2]: Learns a mapping from the data space to a lower-dimensional feature space in which it iteratively optimizes a clustering objective.
- IDEC [3]: Manipulates feature space to scatter data by optimizing a KL divergence-based clustering loss and maintains the local structure carefully.

C Comparison with State-of-the-art Methods

Table 2 reports the metric Normalized mutual information (NMI) of each algorithm on the 36 UCR time series datasets. The best performance for each dataset is highlighted in bold. And our method also achieves the lowest average rank of 2.2500 while USSL achieve 2.3472 on the NMI metric.

Table 2: Normalized Mutual Information (NMI) comparisons on 36 time series datasets (the values in parentheses represent standard deviations)

Dataset	K-means [14]	UDFS [4]	NDFS [5]	RUFS [6]	RSFS [7]	KSC [8]	KDBA [9]	k-shape [10]	u-shapelet [11]	DTC [12]	USSL [13]	DEC [2]	IDEC [3]	DTCR
Arrow	0.4816	0.5240	0.4997	0.5975	0.5104	0.5240	0.4816	0.5240	0.3522	0.5000	**0.6322**	0.3100	0.2949	0.5513(0.0022)
Beef	0.2925	0.2718	0.3647	0.3799	0.3597	0.3828	0.3340	0.3338	0.3413	0.2751	0.3338	0.2463	0.2463	**0.5473**(0.0013)
BeetleFly	0.0073	0.0371	0.1264	0.1919	0.2795	0.2215	0.2783	0.3456	0.5105	0.3456	0.5310	0.0308	0.0082	**0.7610**(0.0001)
BirdChicken	0.0371	0.0371	0.3988	0.1187	0.3002	0.3988	0.2167	0.3456	0.2783	0.0073	**0.6190**	0.0160	0.0082	0.5310(0.0035)
Car	0.2540	0.2319	0.2361	0.2511	0.2920	0.2719	0.2691	0.3771	0.3655	0.1892	0.4650	0.2766	0.2972	**0.5021**(0.0020)
chlorineConcentration	0.0129	0.0138	0.0075	**0.0254**	0.0159	0.0147	0.0164	0.0000	0.0135	0.0013	0.0133	0.0009	0.0008	0.0195(0.0010)
coffee	0.5246	0.6945	**1.0000**	0.2513	**1.0000**	**1.0000**	0.0778	**1.0000**	**1.0000**	0.5523	**1.0000**	0.0120	0.1431	0.6277(0.0015)
diatomsizeReduction	0.9300	0.9300	0.9300	0.8734	0.8761	**1.0000**	0.9300	**1.0000**	0.4849	0.6863	**1.0000**	0.8030	0.5140	0.9418(0.0027)
dist.phal.outl.agegroup	0.1880	0.3262	0.1943	0.2762	0.3548	0.3331	0.4261	0.2911	0.2577	0.3406	0.3846	0.4405	0.4400	**0.4553**(0.0007)
dist.phal.outl.correct	0.0278	0.0473	0.0567	0.1071	0.0782	0.0261	0.0199	0.0527	0.0063	0.0115	0.1026	0.0011	0.0150	**0.1180**(0.0028)
ECG200	0.1403	0.1854	0.1403	0.2668	0.2918	0.1403	0.1886	0.3682	0.1323	0.0918	**0.3776**	0.1885	0.2225	0.3691(0.0028)
ECGFiveDays	0.0002	0.0600	0.1296	0.0352	0.1760	0.0682	0.1983	0.0002	0.1498	0.0022	0.6502	0.0178	0.0223	**0.8056**(0.0034)
GunPoint	0.0126	0.0220	0.0334	0.2405	0.0152	0.0126	0.1288	0.3653	0.3653	0.0194	**0.4878**	0.0020	0.0031	0.4200(0.0013)
Ham	0.0093	0.0389	0.0595	0.0980	0.0256	0.0595	0.0265	0.0517	0.0619	0.1016	**0.3411**	0.1508	0.1285	0.0989(0.0036)
Herring	0.0013	0.0253	0.0225	0.0518	0.0236	0.0027	0.0000	0.0027	0.1324	0.0143	0.1718	0.0306	0.0207	**0.2248**(0.0016)
Lighting2	0.0038	0.0047	0.0851	0.1426	0.0326	0.1979	0.0850	0.2670	0.0144	0.1435	**0.3727**	0.0600	0.1248	0.2289(0.0014)
Meat	0.2510	0.2832	0.2416	0.1943	0.3016	0.2846	0.3661	0.2254	0.2716	0.2250	0.9085	0.5176	0.2250	**0.9653**(0.0009)
Mid.phal.outl.agegroup	0.0219	0.1105	0.0416	0.1595	0.0968	0.1061	0.1148	0.0722	0.1491	0.1390	0.2780	0.2686	0.2199	**0.4661**(0.0017)
Mid.phal.outl.correct	0.0024	0.0713	0.0150	0.0443	0.0321	0.0053	0.0760	0.0349	0.0253	0.0079	**0.2503**	0.1005	0.0083	0.1150(0.0005)
Mid.phal.TW	0.4134	0.4276	0.4149	0.5366	0.4219	0.4486	0.4497	0.5229	0.4065	0.1156	**0.9202**	0.4509	0.3444	0.5503(0.0006)
MoteStrain	0.0551	0.1187	0.1919	0.1264	0.2373	0.3002	0.0970	0.2215	0.0082	0.0094	**0.5310**	0.3867	0.3821	0.4094(0.0041)
OSULeaf	0.0208	0.0200	0.0352	0.0246	0.0463	0.0421	0.0327	0.0126	0.0203	0.2201	**0.3353**	0.2141	0.2412	0.2599(0.0010)
Plane	0.8598	0.8046	0.8414	0.8675	0.8736	0.9218	0.8784	0.9642	**1.0000**	0.8678	**1.0000**	0.8947	0.8947	0.9296(0.0033)
Prox.phal.outl.ageGroup	0.0635	0.0182	0.0830	0.0726	0.0938	0.0682	0.0377	0.0110	0.0332	0.4153	**0.6813**	0.2500	0.5396	0.5581(0.0029)
Prox.phal.TW	0.0082	0.0308	0.2215	0.1187	0.0809	0.1919	0.2167	0.1577	0.0107	0.6199	**1.0000**	0.5864	0.3289	0.6539(0.0017)
SonyAIBORobotSurface	0.6112	0.6122	0.6112	0.6278	0.6368	0.6129	0.5516	**0.7107**	0.5803	0.2559	0.5597	0.2773	0.4451	0.6634(0.0027)
SonyAIBORobotSurfaceII	0.5444	0.4802	0.5413	0.5107	0.5406	0.5619	0.5481	0.0110	0.5903	0.4257	**0.6858**	0.2214	0.2327	0.6121(0.0017)
SwedishLeaf	0.0168	0.0082	0.0934	0.0457	0.0269	0.0073	0.1277	0.1041	0.3456	0.6187	**0.9186**	0.5569	0.5573	0.6663(0.0019)
Symbols	0.7780	0.7277	0.7593	0.7174	0.8027	0.8264	**0.9388**	0.6366	0.8691	0.7995	0.8821	0.7421	0.7419	0.8989(0.0018)
ToeSegmentation1	0.0022	0.0089	0.2141	0.0880	0.0174	0.0202	0.2712	0.3073	0.3073	0.0188	**0.3351**	0.0010	0.0010	0.3115(0.0008)
ToeSegmentation2	0.0863	0.0727	0.1713	0.1713	0.1625	0.0863	0.2627	0.0863	0.1519	0.0096	**0.4308**	0.0065	0.0118	0.3249(0.0033)
TwoPatterns	0.4696	0.3393	0.4351	0.4678	0.4608	0.4705	0.4419	0.3949	0.2979	0.0119	**0.4911**	0.0195	0.0142	0.4713(0.0019)
TwoLeadECG	0.0000	0.0004	0.1353	0.1238	0.0829	0.0011	0.0103	0.0000	0.0529	0.0036	**0.5471**	0.0017	0.0010	0.4614(0.0009)
wafer	0.0010	0.0010	0.0546	**0.0746**	0.0194	0.0010	0.0000	0.0010	0.0010	0.0008	0.0492	0.0165	0.0188	0.0228(0.0001)
Wine	0.0031	0.0045	0.0259	0.0065	0.0096	0.0094	0.0211	0.0119	0.0171	0.0000	**0.7511**	0.0018	0.1708	0.2580(0.0038)
WordsSynonyms	0.5435	0.4745	0.5396	**0.5623**	0.5462	0.4874	0.4527	0.4154	0.3933	0.3498	0.4984	0.4134	0.4387	0.5448(0.0003)
AVG Rank	10.6806	9.6389	7.6389	6.9028	6.8472	7.2639	7.5278	7.9444	8.0833	9.8889	2.3472	8.8750	9.1111	**2.2500**
p-value	1.6803E-7	3.2916E-7	2.2576E-6	1.4131E-6	1.7880E-6	3.8576E-6	3.0288E-7	2.3773E-5	3.3140E-6	1.8292E-7	3.1370E-2	3.2916E-07	3.0.288E-7	-

D The Process of Learning Representation

In section 4.3.2 (The Process of Learning Representation) of the main text, we visualize the learning process of DTCR on 2 specific datasets, demonstrating that DTCR is capable of learning the cluster-specific representations as the iterative process proceeds. Here we report the experimental results on all other datasets by recording the improvements on the metrics. As shown in Table 3, the performance is gradually improving during the learning process (indicating by an upward arrow) demonstrating that the learned representations are more and more suitable for clustering.

Table 3: The improvements of the performance during the training process

Dataset	#Epoch 0 (RI/NMI/ACC)	#Epoch 30 (RI/NMI/ACC)	#Epoch 50 (RI/NMI/ACC)
Arrow	0.4126 / 0.3317 / 0.4152	0.4952 / 0.4119 / 0.5137↑	0.5717 / 0.4687 / 0.5796↑
Beef	0.4828 / 0.3371 / 0.3406	0.5942 / 0.4025 / 0.4205↑	0.6778 / 0.4650 / 0.4815↑
BeetleFly	0.5403 / 0.4566 / 0.5105	0.6639 / 0.5640 / 0.6213↑	0.7640 / 0.6436 / 0.7115↑
BirdChicken	0.4876 / 0.3189 / 0.5100	0.5980 / 0.3911 / 0.6182↑	0.6870 / 0.4483 / 0.7011↑
Car	0.4503 / 0.3023 / 0.3798	0.5532 / 0.3676 / 0.4691↑	0.6235 / 0.4244 / 0.5337↑
chlorineConcentration	0.3216 / 0.0279 / 0.5368	0.3996 / 0.0333 / 0.6688↑	0.4632 / 0.0385 / 0.7572↑
coffee	0.5600 / 0.4892 / 0.5784	0.6605 / 0.6055 / 0.7185↑	0.7665 / 0.7098 / 0.8239↑
diatomsizeReduction	0.5837 / 0.5666 / 0.5475	0.7059 / 0.6864 / 0.6614↑	0.8235 / 0.7810 / 0.7517↑
dist.phal.outl.agegroup	0.4701 / 0.2730 / 0.4832	0.5801 / 0.3379 / 0.5910↑	0.6614 / 0.3870 / 0.6830↑
dist.phal.outl.correct	0.3658 / 0.0707 / 0.4249	0.4526 / 0.0851 / 0.5163↑	0.5118 / 0.0973 / 0.5788↑
ECG200	0.3987 / 0.2218 / 0.4825	0.4962 / 0.2679 / 0.5924↑	0.5557 / 0.3121 / 0.6888↑
ECGFiveDays	0.5783 / 0.4856 / 0.5123	0.7029 / 0.5874 / 0.6160↑	0.8035 / 0.6802 / 0.7196↑
GunPoint	0.3839 / 0.2526 / 0.4719	0.4712 / 0.3137 / 0.5755↑	0.5342 / 0.3562 / 0.6645↑
Ham	0.3222 / 0.0592 / 0.3829	0.3961 / 0.0730 / 0.4735↑	0.4582 / 0.0831 / 0.5407↑
Herring	0.3459 / 0.1354 / 0.4220	0.4191 / 0.1690 / 0.5220↑	0.4869 / 0.1965 / 0.5885↑
Lighting2	0.3555 / 0.1375 / 0.4330	0.4284 / 0.1640 / 0.5244↑	0.4900 / 0.1923 / 0.6150↑
Meat	0.5863 / 0.5797 / 0.5872	0.7254 / 0.7145 / 0.7261↑	0.8190 / 0.8185 / 0.8213↑
Mid.phal.outl.agegroup	0.4792 / 0.2799 / 0.4619	0.5763 / 0.3474 / 0.5719↑	0.6508 / 0.4034 / 0.6380↑
Mid.phal.outl.correct	0.3369 / 0.0689 / 0.4040	0.4110 / 0.0841 / 0.4886↑	0.4837 / 0.0954 / 0.5483↑
Mid.phal.TW	0.5183 / 0.3303 / 0.3568	0.6316 / 0.3965 / 0.4394↑	0.7195 / 0.4597 / 0.4996↑
MoteStrain	0.4611 / 0.2459 / 0.5036	0.5605 / 0.3009 / 0.6206↑	0.6492 / 0.3458 / 0.7159↑
OSULeaf	0.4645 / 0.1563 / 0.2658	0.5747 / 0.1969 / 0.3202↑	0.6628 / 0.2235 / 0.3716↑
Plane	0.5752 / 0.5585 / 0.3781	0.7034 / 0.7019 / 0.4766↑	0.8019 / 0.8039 / 0.5378↑
Prox.phal.outl.ageGroup	0.4895 / 0.3450 / 0.4684	0.5895 / 0.4256 / 0.5713↑	0.6641 / 0.4753 / 0.6562↑
Prox.phal.TW	0.5416 / 0.4186 / 0.3782	0.6594 / 0.5151 / 0.4517↑	0.7607 / 0.5885 / 0.5139↑
SonyAIBORobotSurface	0.5264 / 0.4597 / 0.5570	0.6391 / 0.5627 / 0.6731↑	0.7238 / 0.6562 / 0.7735↑
SonyAIBORobotSurfaceII	0.5015 / 0.3699 / 0.5435	0.6269 / 0.4574 / 0.6540↑	0.7025 / 0.5243 / 0.7726↑
SwedishLeaf	0.5570 / 0.4017 / 0.2289	0.6731 / 0.5034 / 0.2697↑	0.7699 / 0.5753 / 0.3119↑
Symbols	0.5520 / 0.5395 / 0.3901	0.6723 / 0.6656 / 0.4881↑	0.7808 / 0.7470 / 0.5503↑
ToeSegmentation1	0.3396 / 0.1869 / 0.3395	0.4145 / 0.2338 / 0.4151↑	0.4736 / 0.2633 / 0.4684↑
ToeSegmentation2	0.4988 / 0.2340 / 0.5408	0.6084 / 0.2841 / 0.6457↑	0.7065 / 0.3243 / 0.7477↑
TwoPatterns	0.4210 / 0.2836 / 0.2555	0.5099 / 0.3470 / 0.3155↑	0.5987 / 0.3992 / 0.3641↑
TwoLeadECG	0.4279 / 0.2770 / 0.5178	0.5059 / 0.3419 / 0.6290↑	0.5814 / 0.3883 / 0.7188↑
wafer	0.4420 / 0.0135 / 0.4070	0.5314 / 0.0163 / 0.5022↑	0.6085 / 0.0186 / 0.5618↑
Wine	0.3779 / 0.1732 / 0.4782	0.4589 / 0.2199 / 0.5773↑	0.5330 / 0.2458 / 0.6541↑
WordsSynonyms	0.5406 / 0.3271 / 0.0977	0.6582 / 0.3953 / 0.1167↑	0.7572 / 0.4593 / 0.1338↑

E Robustness Analysis

In section 4.3.3 (Robustness Analysis) of the main text, we explore the robustness of our model on one specific dataset (*SonyAIBORobotSurface*). Here we report the experimental results on all datasets to verify the point that our model is capable of correcting mistakes with the help of the term of $\mathcal{L}_{reconstruction}$ ($\mathcal{L}_{res}$) by recording the changes of the metrics.

As shown in Table 4, the experimental results are consistent with the Robustness Analysis of the main text. Three metrics in Table 4 from left to right are RI, NMI and ACC, respectively. The results in Table 4 can be divided into Group A and B, consisting of three columns a, b, c and a, e, f, respectively. **More intuitively, Groups A and B correspond to the results of the first and second rows in Figure 4 of the main text, respectively.** We also use arrows to indicate the changes in the performance. Within the Group A, from column a to b, when we train DTCR with the loss term of shuffled K-means and the classification, the wrong clustering information does mislead the learning, decreasing the clustering performance. However, once putting the $\mathcal{L}_{res}$ term back (column c), the clustering performance is improved, indicating the learning of the model was corrected. Similarly, we do that again to check what happens without the $\mathcal{L}_{cls}$ loss term. Within the Group B, from column a to e, even without the $\mathcal{L}_{cls}$ term, but with the help of the $\mathcal{L}_{res}$ term, the performance is still improved and less confused. Comparing column b and e, it is clear that the $\mathcal{L}_{res}$ term enables our model to correct mistakes. Comparing column c and f shows that the earlier and longer the $\mathcal{L}_{res}$ term is used, the stronger the ability to prevent being misled by K-means and thus obtaining better performance.

Table 4: Robustness Analysis of DTCR

Dataset	Initial state (a)	shuffle kmeans+$\mathcal{L}_{cls}$ (b)	putting $\mathcal{L}_{res}$ back (c)	shuffle kmeans+$\mathcal{L}_{res}$ (e)	putting $\mathcal{L}_{cls}$ back (f)
Arrow	0.5717 / 0.4687 / 0.5796	0.5342 / 0.4361 / 0.5523↓	0.6240 / 0.5140 / 0.6361↑	0.6057 / 0.4997 / 0.6168↑	0.6323 / 0.5203 / 0.6437↑
Beef	0.6778 / 0.4650 / 0.4815	0.6475 / 0.4365 / 0.4505↓	0.7530 / 0.5140 / 0.5277↑	0.7303 / 0.4928 / 0.5151↑	0.7596 / 0.5178 / 0.5323↑
BeetleFly	0.7640 / 0.6436 / 0.7115	0.7253 / 0.6124 / 0.6766↓	0.8347 / 0.6968 / 0.7806↑	0.8169 / 0.6732 / 0.7603↑	0.8477 / 0.7099 / 0.7893↑
BirdChicken	0.6870 / 0.4483 / 0.7011	0.6510 / 0.4273 / 0.6664↓	0.7485 / 0.4894 / 0.7699↑	0.7309 / 0.4737 / 0.7473↑	0.7592 / 0.4950 / 0.7808↑
Car	0.6235 / 0.4244 / 0.5337	0.5956 / 0.4027 / 0.5150↓	0.6918 / 0.4604 / 0.5821↑	0.6566 / 0.4491 / 0.5691↑	0.7008 / 0.4685 / 0.5919↑
chlorineConcentration	0.4632 / 0.0385 / 0.7572	0.4434 / 0.0363 / 0.7227↓	0.4978 / 0.0428 / 0.8252↑	0.4875 / 0.0413 / 0.8007↑	0.5011 / 0.0432 / 0.8353↑
coffee	0.7665 / 0.7098 / 0.8239	0.7237 / 0.6618 / 0.7817↓	0.8451 / 0.7571 / 0.8949↑	0.8209 / 0.7338 / 0.8677↑	0.8548 / 0.7693 / 0.9041↑
diatomsizeReduction	0.8235 / 0.7810 / 0.7517	0.7848 / 0.7434 / 0.7212↓	0.8897 / 0.8663 / 0.8432↑	0.8632 / 0.8415 / 0.8076↑	0.8987 / 0.8815 / 0.8521↑
dist.phal.outl.agegroup	0.6614 / 0.3870 / 0.6830	0.6299 / 0.3700 / 0.6366↓	0.7203 / 0.4137 / 0.7411↑	0.7027 / 0.4030 / 0.7203↑	0.7284 / 0.4198 / 0.7536↑
dist.phal.outl.correct	0.5118 / 0.0973 / 0.5788	0.4924 / 0.0931 / 0.5574↓	0.5658 / 0.1078 / 0.6551↑	0.5493 / 0.1044 / 0.6264↑	0.5727 / 0.1095 / 0.6664↑
ECG200	0.5557 / 0.3121 / 0.6888	0.5328 / 0.2958 / 0.6509↓	0.6120 / 0.3407 / 0.7464↑	0.5917 / 0.3291 / 0.7309↑	0.6203 / 0.3455 / 0.7556↑
ECGFiveDays	0.8035 / 0.6802 / 0.7196	0.7665 / 0.6489 / 0.6874↓	0.8800 / 0.7384 / 0.7890↑	0.8587 / 0.7201 / 0.7655↑	0.9012 / 0.7513 / 0.7987↑
GunPoint	0.5342 / 0.3562 / 0.6645	0.5128 / 0.3392 / 0.6220↓	0.5904 / 0.3889 / 0.7219↑	0.5762 / 0.3753 / 0.7006↑	0.5992 / 0.3935 / 0.7330↑
Ham	0.4582 / 0.0831 / 0.5407	0.4359 / 0.0792 / 0.5207↓	0.4935 / 0.0904 / 0.5776↑	0.4797 / 0.0881 / 0.5649↑	0.5013 / 0.0920 / 0.5837↑
Herring	0.4869 / 0.1965 / 0.5885	0.4637 / 0.1835 / 0.5673↓	0.5349 / 0.2076 / 0.6425↑	0.5172 / 0.2038 / 0.6189↑	0.5389 / 0.2104 / 0.6513↑
Lighting2	0.4900 / 0.1923 / 0.6150	0.4690 / 0.1833 / 0.5699↓	0.5438 / 0.2097 / 0.6672↑	0.5232 / 0.2052 / 0.6526↑	0.5530 / 0.2138 / 0.6780↑
Meat	0.8190 / 0.8185 / 0.8213	0.7833 / 0.7764 / 0.7887↓	0.8957 / 0.8871 / 0.8861↑	0.8736 / 0.8582 / 0.8641↑	0.9157 / 0.9000 / 0.9097↑
Mid.phal.outl.agegroup	0.6508 / 0.4034 / 0.6380	0.6269 / 0.3858 / 0.6166↓	0.7228 / 0.4332 / 0.7044↑	0.7030 / 0.4231 / 0.6878↑	0.7340 / 0.4413 / 0.7197↑
Mid.phal.outl.correct	0.4837 / 0.0954 / 0.5483	0.4562 / 0.0910 / 0.5244↓	0.5231 / 0.1052 / 0.6039↑	0.5088 / 0.1020 / 0.5849↑	0.5296 / 0.1065 / 0.6150↑
Mid.phal.TW	0.7195 / 0.4597 / 0.4996	0.6964 / 0.4354 / 0.4782↓	0.7939 / 0.5078 / 0.5405↑	0.7655 / 0.4899 / 0.5296↑	0.8038 / 0.5159 / 0.5509↑
MoteStrain	0.6492 / 0.3458 / 0.7159	0.6178 / 0.3307 / 0.6803↓	0.7021 / 0.3767 / 0.7872↑	0.6815 / 0.3678 / 0.7656↑	0.7129 / 0.3813 / 0.7965↑
OSULeaf	0.6628 / 0.2235 / 0.3716	0.6305 / 0.2145 / 0.3519↓	0.7206 / 0.2409 / 0.4107↑	0.7027 / 0.2345 / 0.3985↑	0.7289 / 0.2435 / 0.4163↑
Plane	0.8019 / 0.8039 / 0.5378	0.7683 / 0.7615 / 0.5159↓	0.8824 / 0.8705 / 0.5772↑	0.8571 / 0.8570 / 0.5624↑	0.8956 / 0.8806 / 0.5854↑
Prox.phal.outl.ageGroup	0.6641 / 0.4753 / 0.6562	0.6422 / 0.4527 / 0.6212↓	0.7342 / 0.5340 / 0.7255↑	0.7091 / 0.5140 / 0.6968↑	0.7437 / 0.5424 / 0.7308↑
Prox.phal.TW	0.7607 / 0.5885 / 0.5139	0.7220 / 0.5624 / 0.4898↓	0.8238 / 0.6327 / 0.5767↑	0.7974 / 0.6197 / 0.5529↑	0.8344 / 0.6423 / 0.5846↑
SonyAIBORobotSurface	0.7238 / 0.6562 / 0.7735	0.6870 / 0.6209 / 0.7410↓	0.8045 / 0.7101 / 0.8437↑	0.7851 / 0.6938 / 0.8190↑	0.8168 / 0.7189 / 0.8594↑
SonyAIBORobotSurfaceII	0.7025 / 0.5243 / 0.7726	0.6828 / 0.4963 / 0.7289↓	0.7707 / 0.5603 / 0.8335↑	0.7456 / 0.5479 / 0.8105↑	0.7875 / 0.5728 / 0.8459↑
SwedishLeaf	0.7699 / 0.5753 / 0.3119	0.7300 / 0.5537 / 0.2938↓	0.8518 / 0.6160 / 0.3507↑	0.8339 / 0.6022 / 0.3373↑	0.8618 / 0.6291 / 0.3547↑
Symbols	0.7808 / 0.7470 / 0.5503	0.7336 / 0.7188 / 0.5266↓	0.8414 / 0.8272 / 0.5944↑	0.8240 / 0.7976 / 0.5783↑	0.8537 / 0.8402 / 0.6026↑
ToeSegmentation1	0.4736 / 0.2633 / 0.4684	0.4505 / 0.2520 / 0.4535↓	0.5190 / 0.2897 / 0.5171↑	0.5081 / 0.2793 / 0.4987↑	0.5256 / 0.2947 / 0.5245↑
ToeSegmentation2	0.7065 / 0.3243 / 0.7477	0.6813 / 0.3085 / 0.7082↓	0.7636 / 0.3536 / 0.8186↑	0.7463 / 0.3450 / 0.7932↑	0.7739 / 0.3591 / 0.8383↑
TwoPatterns	0.5987 / 0.3992 / 0.3641	0.5686 / 0.3803 / 0.3426↓	0.6455 / 0.4320 / 0.3940↑	0.6264 / 0.4198 / 0.3865↑	0.6529 / 0.4393 / 0.3985↑
TwoLeadECG	0.5814 / 0.3883 / 0.7188	0.5507 / 0.3681 / 0.6858↓	0.6426 / 0.4187 / 0.7905↑	0.6243 / 0.4104 / 0.7613↑	0.6542 / 0.4279 / 0.8032↑
wafer	0.6085 / 0.0186 / 0.5618	0.5696 / 0.0177 / 0.5343↓	0.6642 / 0.0203 / 0.6120↑	0.6466 / 0.0197 / 0.6008↑	0.6757 / 0.0207 / 0.6239↑
Wine	0.5330 / 0.2458 / 0.6541	0.5100 / 0.2351 / 0.6257↓	0.5821 / 0.2692 / 0.7267↑	0.5661 / 0.2604 / 0.7060↑	0.5915 / 0.2717 / 0.7395↑
WordsSynonyms	0.7572 / 0.4593 / 0.1338	0.7275 / 0.4327 / 0.1255↓	0.8215 / 0.4973 / 0.1498↑	0.8006 / 0.4909 / 0.1440↑	0.8392 / 0.5080 / 0.1521↑

F Hyper-parameter Analysis

Here, we perform an empirical evaluation of the effect of the hyper-parameter T. We perform this evaluation with randomly chosen 10 datasets. We do the evaluation by varying one parameter at a time while maintaining the other parameters fixed (hidden size: $[100, 50, 50]$, λ: 1, dilation: $[1, 4, 16]$). As shown in Table 5, our method is robust to the hyper-parameter T.

Table 5: Hyper-parameter Analysis of T

T / Dataset	10		20		30		40	
	RI	NMI	RI	NMI	RI	NMI	RI	NMI
diatomsizeReduction	0.9547	0.8801	0.9515	0.8874	0.9580	0.8910	0.9565	0.8901
Mid.phal.outl.agegroup	0.7966	0.4710	0.7967	0.4670	0.7969	0.4725	0.7981	0.4710
Mid.phal.TW	0.8635	0.4625	0.8632	0.4738	0.8630	0.4738	0.8636	0.4740
OSULeaf	0.7536	0.2859	0.7534	0.2532	0.7573	0.2398	0.7571	0.2640
Plane	0.9505	0.9296	0.9571	0.9295	0.9505	0.9296	0.9608	0.9296
Prox.phal.outl.ageGroup	0.7954	0.5377	0.7981	0.5596	0.7902	0.5491	0.8007	0.5628
Prox.phal.TW	0.8862	0.6299	0.8891	0.6342	0.8972	0.6517	0.8939	0.6571
SonyAIBORobotSurfaceII	0.8021	0.4813	0.8004	0.4767	0.8054	0.4864	0.8086	0.4932
SwedishLeaf	0.9088	0.5879	0.9074	0.5720	0.9111	0.5965	0.9094	0.5955
WordsSynonyms	0.8941	0.4287	0.8973	0.4402	0.8972	0.4300	0.8918	0.4180

References

[1] Yanping Chen, Eamonn Keogh, Bing Hu, Nurjahan Begum, Anthony Bagnall, Abdullah Mueen, and Gustavo Batista. The ucr time series classification archive, July 2015. `www.cs.ucr.edu/~eamonn/time_series_data/`.

[2] Junyuan Xie, Ross Girshick, and Ali Farhadi. Unsupervised deep embedding for clustering analysis. In *International conference on machine learning*, pages 478–487, 2016.

[3] Xifeng Guo, Long Gao, Xinwang Liu, and Jianping Yin. Improved deep embedded clustering with local structure preservation. In *IJCAI*, pages 1753–1759, 2017.

[4] Yi Yang, Heng Tao Shen, Zhigang Ma, Zi Huang, and Xiaofang Zhou. L2, 1-norm regularized discriminative feature selection for unsupervised. In *Twenty-Second International Joint Conference on Artificial Intelligence*, 2011.

[5] Zechao Li, Yi Yang, Jing Liu, Xiaofang Zhou, and Hanqing Lu. Unsupervised feature selection using nonnegative spectral analysis. In *Twenty-Sixth AAAI Conference on Artificial Intelligence*, 2012.

[6] Mingjie Qian and Chengxiang Zhai. Robust unsupervised feature selection. In *Twenty-Third International Joint Conference on Artificial Intelligence*, 2013.

[7] Lei Shi, Liang Du, and Yi-Dong Shen. Robust spectral learning for unsupervised feature selection. In *2014 IEEE International Conference on Data Mining*, pages 977–982. IEEE, 2014.

[8] Jaewon Yang and Jure Leskovec. Patterns of temporal variation in online media. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 177–186. ACM, 2011.

[9] François Petitjean, Alain Ketterlin, and Pierre Gançarski. A global averaging method for dynamic time warping, with applications to clustering. *Pattern Recognition*, 44(3):678–693, 2011.

[10] John Paparrizos and Luis Gravano. k-shape: Efficient and accurate clustering of time series. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, pages 1855–1870. ACM, 2015.

[11] Jesin Zakaria, Abdullah Mueen, and Eamonn Keogh. Clustering time series using unsupervised-shapelets. In *2012 IEEE 12th International Conference on Data Mining*, pages 785–794. IEEE, 2012.

[12] Naveen Sai Madiraju, Seid M Sadat, Dimitry Fisher, and Homa Karimabadi. Deep temporal clustering: Fully unsupervised learning of time-domain features. *arXiv preprint arXiv:1802.01059*, 2018.

[13] Qin Zhang, Jia Wu, Peng Zhang, Guodong Long, and Chengqi Zhang. Salient subsequence learning for time series clustering. *IEEE transactions on pattern analysis and machine intelligence*, 2018.

[14] John A Hartigan and Manchek A Wong. Algorithm as 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1):100–108, 1979.