SMATE: Semi-supervised Spatio-Temporal Representation Learning on MultivariAte Time Series

Jingwei Zuo, Karine Zeitouni, Yehia Taher

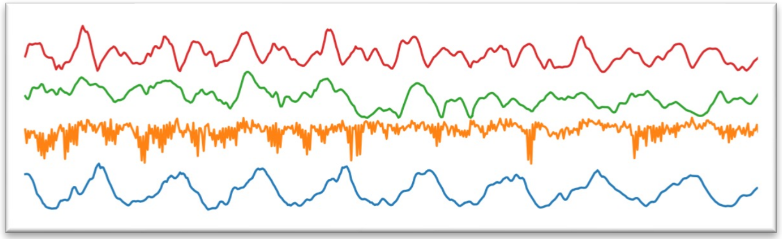
DAVID Lab, UVSQ,
University of Paris-Saclay

IEEE International Conference on Data Mining (ICDM) 2021

Multivariate Time Series (MTS)

Definition

- A sequence where each element is a M-d vector, M is the variable number:
  $\mathbf{x} = (t_1, t_2, \ldots, t_i, \ldots, t_T)$, where $t_i \in \mathbb{R}^M$.
- When M=1, we call it *Univariate Time Series (UTS)*



A MTS sample representing "Running" activity (with 4 sensors) of the SHL[1] dataset.

1. H. Gjoreski, M. Ciliberto, L. Wang, F. J. O. Morales, S. Mekki, S. Valentin, D. Roggen. "The University of Sussex-Huawei Locomotion and Transportation Dataset for Multimodal Analytics with Mobile Devices." IEEE Access 6 (2018): 42592-4260

Problem Statement

- Complex structure in Multivariate Time Series (MTS)
- Label shortage when learning from MTS

Objectives

1. Explore the features and learn an appropriate representation of Multivariate Time Series (MTS)
   - The **temporal dependencies** on time axis: temporal dynamics
   - The **dynamic interaction**s between the variables: spatial dynamics
2. Remedy to label shortage when learning from MTS
   - Huge cost on labelling MTS in real life
   - Weak supervision is envisaged

Multivariate Time Series Features

1. Combine features from each variable (i.e., 1-D series)
   - Global features: **1NN-DTW$_I$** [Yekta et al., DMKD'15]
   - Shapelet features: **Shapelet Ensemble** [Cetin et al., SDM'15], **M-Shapelet Discovery** [Grabocka et al., KAIS'16]
   - Motif features: **WEASEL+MUSE** [Schafer et al., AALTD'18], **Global Discriminative Patterns** [Nayak et al., SDM'18]
   - Deep Representation features: **Multi-Channels CNN** [Zheng et al., WAIM'14]
2. Extract features directly from all variables
   - Global features: **1NN-DTW$_D$** (**1NN-DTW$_A$**) [Yekta et al., DMKD'15]
   - Motif features: Symbolic Representation for MTS (**SMTS**) [Baydogan et al., DMKD'15]
   - Deep Representation features: Modified DNN approaches for Univariate TSC, e.g., **LSTM-FCNs** [Karim et al., ArXiv'19], **InceptionTime** [Fawaz et al., DMKD'20], **ROCKET** [Dempster et al., DMKD'20], etc.
3. Consider the interactions between the variables
   - Variable correlation: **MLSTM-FCNs** [Karim et al., Neural Networks'19]
   - Attention Mechanism: **CA-SFCN** [Hao et al. IJCAI'20]
   - 2D-CNN with 1D-CNN: **MTEX-CNN** [Assaf et al., ICDM'19], **XCM** [Fauvel et al., ArXiv'20]
   - Graph Pooling: **MTPool** [Xu et al., ArXiv'20]
   - Etc.

All these approaches are fully supervised!

Semi-supervised Learning (SSL) on TS

1. Self-training or Positive Unlabeled Learning-based models
   - **SSTSC** [Wei and Keogh, KDD'06]
   - **LCLC** [Nguyen et al., IJCAI'11]
   - **DTW-D** [Chen et al., KDD'13], etc.
   - **SSSL** [Wang et al., Pattern Recognit'19]
2. Clustering-based model
   - **SUCCESS** [Marussy et al., ICAISC'13]
3. Self-Supervised Learning-based model
   - **MTL** [Javed et al., PAKDD'20]

⚠ All these SSL approaches are designed for Univariate Time Series!

Semi-supervised Learning on MTS

- **USRL** [Franceschi et al., NeurIPS'19]
- **TapNets** [Zhang et al., AAAI'20]

Proposal: SMATE

- **S**emi-supervised Spatio-Temporal Representation Learning on **M**ultiv**A**riate **T**ime **SE**ries

- **Representation Learning** on $\mathbf{x} \in \mathbb{R}^{T \times M}$
  - Learn a low-dimensional representation $\mathbf{h} \in \mathbb{R}^{L \times D}$, where $L < T,\ D < M$
  - **h** embeds the spatial* and temporal features of **x**
- **Semi-supervised regularization** in the embedding space $\mathcal{H}$
  - Combine both labelled and unlabelled samples
  - Learn class-separable representations for downstream tasks, e.g., MTS classification

* We use the term "spatial" in this work to represent the variable axis, instead of indicating the physical spatial locations

Spatio-temporal Representation on MTS

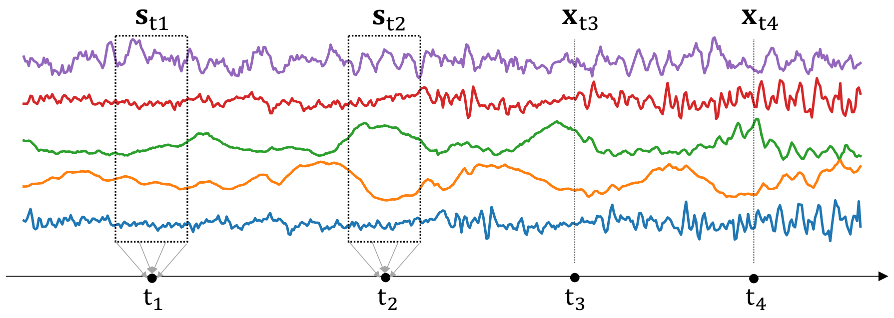
Intuitions

System status at time t

- Local value $\mathrm{x}_t \in \mathbb{R}^M$
- Neighbor values $s_t = [x_{t-m/2}, x_{t+m/2}]$

Spatio-temporal features

- Temporal Dynamic $p(\mathrm{x}_{t'} \mid \mathrm{x}_t)$
- Spatial Dynamic $p(\mathrm{s}_{t'} \mid \mathrm{s}_t)$
- Spatio-temporal dynamics $(\mathrm{x}_t, \mathrm{s}_t) \rightarrow (\mathrm{x}_{t'}, \mathrm{s}_{t'})$



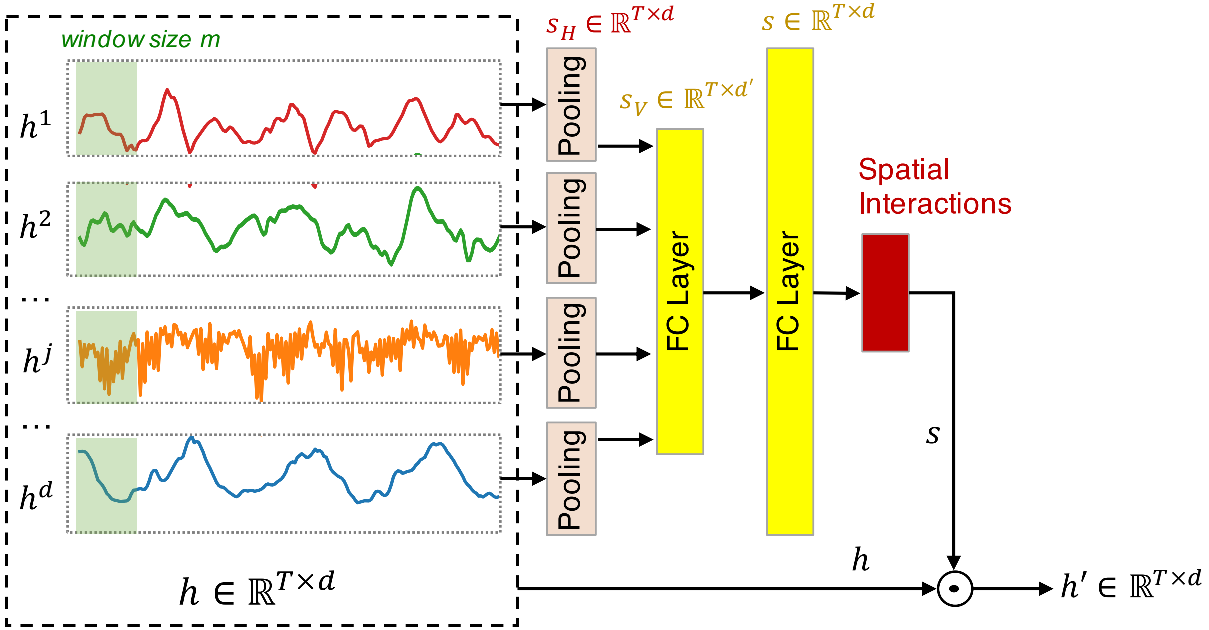
The spatial and temporal structure in a MTS sample representing "Walking" activity (with 5 sensors) of the SHL[1] dataset.

1. H. Gjoreski, M. Ciliberto, L. Wang, F. J. O. Morales, S. Mekki, S. Valentin, D. Roggen. "The University of Sussex-Huawei Locomotion and Transportation Dataset for Multimodal Analytics with Mobile Devices." IEEE Access 6 (2018): 42592-4260

Spatio-temporal Representation on MTS

Spatial Modelling Block (SMB)

- Capture the spatial interaction at **segment level**
- 1-d average Pooling: output the **temporal status**
- 2 Fully Connected (FC) layers: **interacting** the temporal status **in spatial direction**
- **Output:** the weighted MTS considering the spatial interaction



$s_{H_i} \in \mathbb{R}^d$

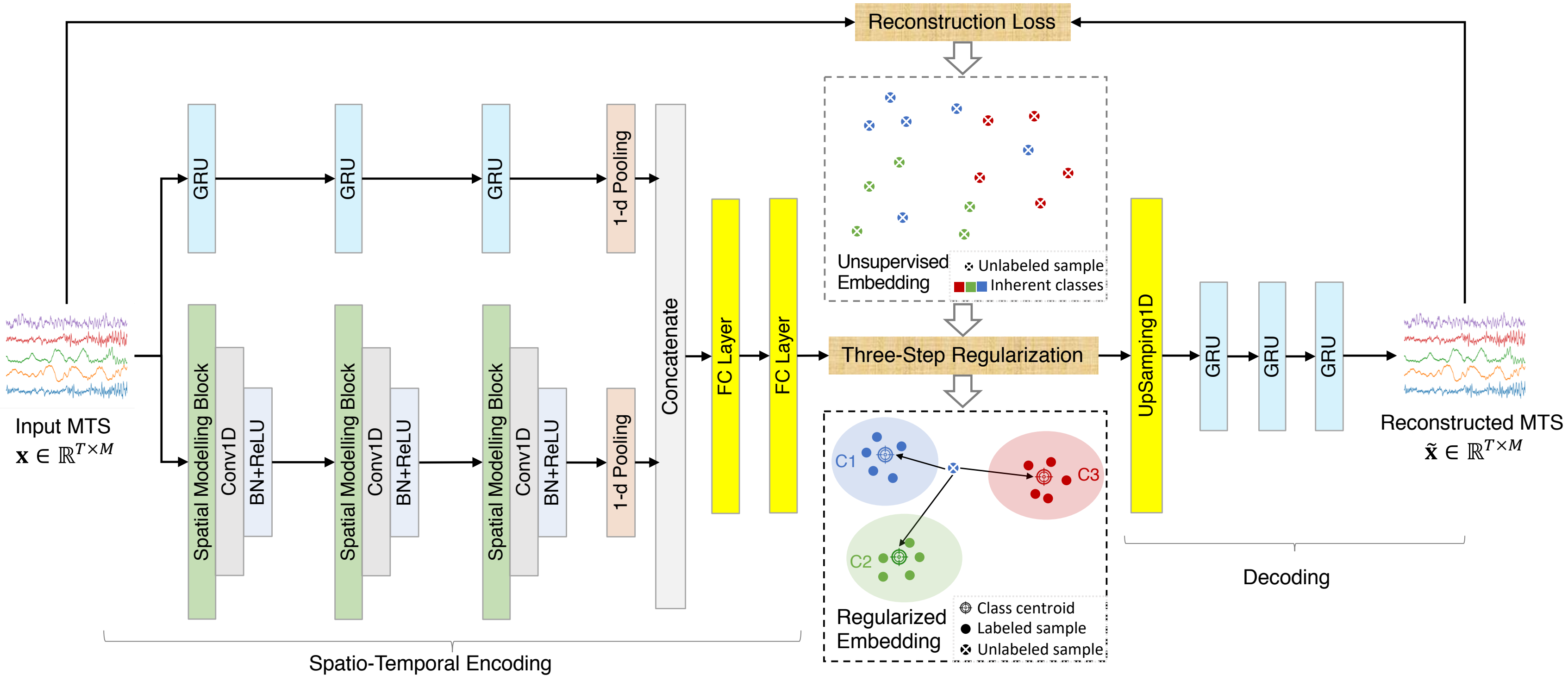
- Horizontal temporal status

$s_i \in \mathbb{R}^d$

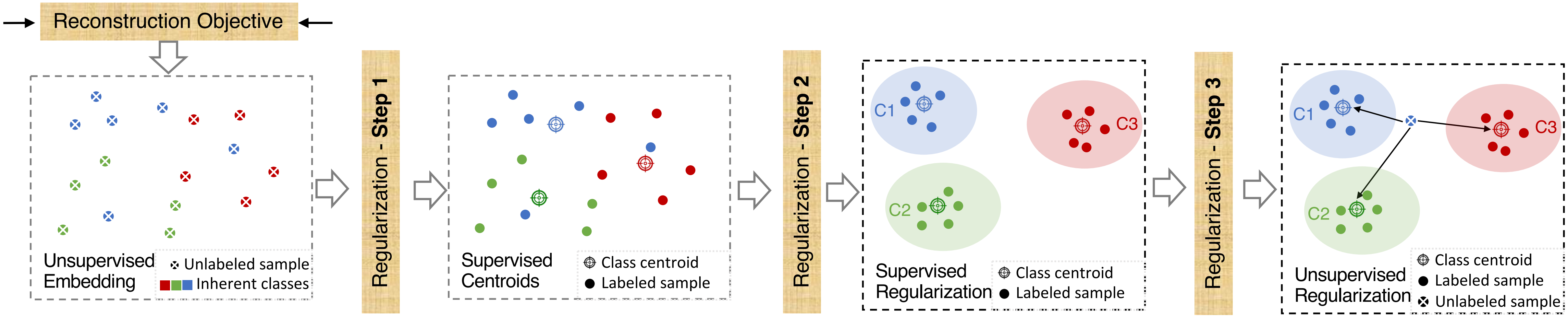
- Spatial interaction weights

SMATE: Model Structure

- Based on an **asymmetric auto-encoder** structure
- Two channels for **Spatio-temporal encoding**
  - GRU: $p(\mathrm{x}_{t'} \mid \mathrm{x}_t)$
  - SMB + Conv1D: $p(\mathrm{s}_{t'} \mid \mathrm{s}_t)$
- **Three-Step Regularization**



SMATE: Three-Step Regularization



Step 1: Supervised Centroids Initialization

The embedding collection of X^k :

- $H^k \in \mathbb{R}^{N_k \times L \times D}$

The centroid of class k

- $c_k = mean(H^k),\ c^k \in \mathbb{R}^{L \times D}$

Step 2: Supervised Centroids Adjustment

Intuition:

- The *near-by* samples have *larger contribution weights* to the class centroids

$$W_{k,i} = 1 - \frac{dist(h_\theta(\mathbf{x}_i), c_k)}{\sum_{i=1}^{K} dist(h_\theta(\mathbf{x}_i), c_j)}$$

$$c_k = \sum_{i=1}^{N_k} W_{k,i} \cdot h_i^k, \quad h_i^k \in H^k$$

Step 3: Unsupervised Centroids Adjustment

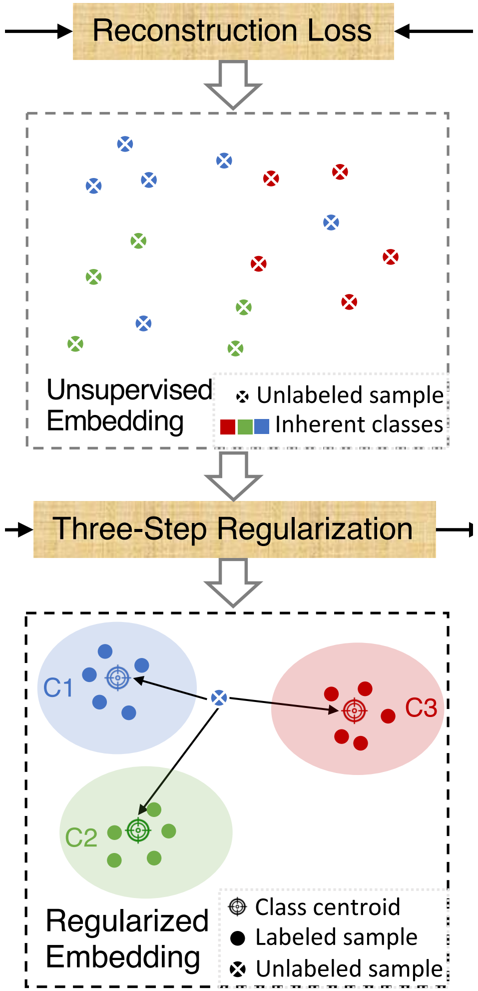
- The propagated label from the distance-based class probability

$$\hat{p}_\theta(y = k|\hat{\mathbf{x}}_i) = 1 - \frac{dist(h_\theta(\hat{\mathbf{x}}_i), c_k)}{\sum_{j=1}^{K} dist(h_\theta(\hat{\mathbf{x}}_i), c_j)}$$

- The class centroid c_k is further adjusted:

$$c_k = \frac{N_k}{N_k + \hat{N}_k} \sum_{i=1}^{N_k} W_{k,i} \cdot h_i^k + \frac{\hat{N}_k}{N_k + \hat{N}_k} \sum_{i=1}^{\hat{N}_k} \hat{p}_{k,i} \cdot \hat{h}_i^k$$

SMATE: Joint Model Optimization



Regularization loss:

- With labelled samples
- With class centroids regularized by both labelled and unlabelled samples

$$L_{Reg}(\theta) = -\sum_{k} log W_{\theta}(y = k|\mathbf{x})$$

Reconstruction loss:

$$L_R = \mathbb{E}_{\mathbf{x}_{1:T}}\left[\sum_t \|\mathbf{x}_t - \tilde{\mathbf{x}}_t\|_2\right]$$

Objective function:

$$\min_{\theta}(L_R + \lambda L_{Reg})$$

Experiments

Objectives

1. Evaluating SMATE for fully supervised representation learning

30 datasets from UEA Archive[1]

13 baselines:

- Distance-based 1-NN classifier on non-normalized (*non-norm*) or normalized (*norm*) MTS
  - **1NN-ED** (*non-norm & norm*)
  - **1NN-DTW$_I$** (*non-norm & norm*); **1NN-DTW$_D$** (*non-norm & norm*)
  - **1NN-DTW$_A$** (*norm*) [Yekta et al., DMKD'15]
- Bag-of-patterns classifier
  - **WEASEL+MUSE** [Schäfer et al., AALTD'18]
- Deep Learning-based classifier:
  - **SMATE$_{NR}$**: SMATE without supervised Regularization
  - **MLSTM-FCNs** [Karim et al., Neural Networks'19], **USRL**[Franceschi et al., NeurIPS'19], **TapNet** [Zhang et al., AAAI'20], **CA-SFCN** [Hao et al., IJCAI'20]

1. www.timeseriesclassification.com

Experiments

Fully supervised representation learning

- SVM on learned representation

SMATE performs the best among all baselines, especially on EEG/MEG applications

Performance Comparison for MTS classification over UEA MTS archive

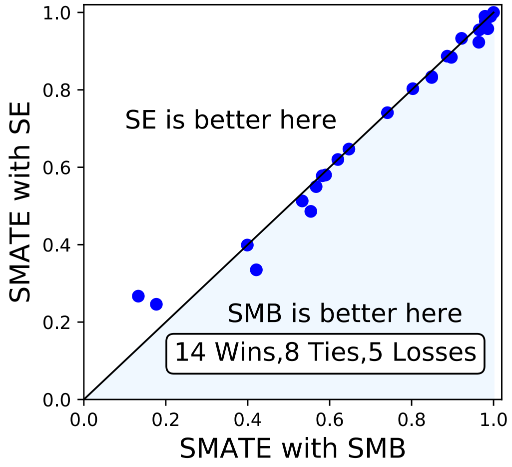
Dataset	SMATE	$SMATE_{NR}$	USRL	TapNet	MLSTM -FCN	CA-SFCN	WEASEL +MUSE	1NN-ED	1NN-DTW_I	1NN-DTW_D	1NN-ED (*norm*)	1NN-DTW_I (*norm*)	1NN-DTW_D (*norm*)	1NN-DTW_A (*norm*)
ArticularyWordR.	**0.993**	0.987	0.987	0.987	0.973	0.97	0.99	0.97	0.98	0.987	0.97	0.98	0.987	0.987
AtrialFibrillation	0.133	0.133	0.133	**0.333**	0.267	**0.333**	**0.333**	0.267	0.267	0.2	0.267	0.267	0.22	0.267
BasicMotions	**1**	**1**	**1**	**1**	0.95	**1**	**1**	0.675	**1**	0.975	0.676	**1**	0.975	**1**
CharacterTrajectories	0.984	**0.997**	0.994	**0.997**	0.985	0.988	0.99	0.964	0.969	0.99	0.964	0.969	0.989	0.989
Cricket	0.986	0.968	0.986	0.958	0.917	0.972	**1**	0.944	0.986	**1**	0.944	0.986	**1**	**1**
DuckDuckGeese	N/A	N/A	**0.675**	0.575	**0.675**	N/A	0.575	0.275	0.55	0.6	0.275	0.55	0.6	0.567
EigenWorms	N/A	N/A	0.878	0.489	0.504	N/A	**0.89**	0.55	0.603	0.618	0.549	N/A	0.619	N/A
Epilepsy	0.964	0.946	0.957	0.971	0.761	0.986	**1**	0.667	0.978	0.964	0.666	0.978	0.964	0.979
ERing	**0.981**	0.904	0.88	0.904	0.941	0.856	0.964	0.93	0.93	0.93	0.93	0.93	0.93	0.93
EthanolConcentration	0.399	0.373	0.236	0.323	0.373	0.323	**0.43**	0.293	0.304	0.323	0.293	N/A	0.323	0.316
FaceDetection	**0.647**	0.556	0.528	0.556	0.545	N/A	0.545	0.519	0.513	0.529	0.519	0.5	0.529	0.529
FingerMovements	**0.62**	0.55	0.54	0.53	0.58	0.59	0.49	0.55	0.52	0.53	0.55	0.52	0.53	0.509
HandMovementD.	**0.554**	0.365	0.27	0.378	0.365	0.324	0.365	0.279	0.306	0.231	0.278	0.306	0.231	0.224
Handwriting	0.421	0.335	0.533	0.357	0.286	0.322	0.605	0.371	0.509	**0.607**	0.2	0.316	0.286	0.601
Heartbeat	0.741	0.615	0.737	0.751	0.663	**0.756**	0.727	0.62	0.659	0.717	0.619	0.658	0.717	0.571
InsectWingbeat	N/A	N/A	0.16	**0.208**	0.167	N/A	N/A	0.128	N/A	0.115	0.128	N/A	N/A	N/A
JapaneseVowels	0.965	0.924	**0.989**	0.965	0.976	0.973	0.973	0.924	0.959	0.949	0.924	0.959	0.949	0.959
Libras	0.849	0.834	0.867	0.85	0.856	0.89	0.878	0.833	**0.894**	0.872	0.833	**0.894**	0.87	0.879
LSST	0.582	0.568	0.558	0.568	0.373	**0.674**	0.59	0.456	0.575	0.551	0.456	0.575	0.551	0.551
MotorImagery	**0.59**	**0.59**	0.54	**0.59**	0.51	N/A	0.51	0.39	N/A	0.5	0.51	N/A	0.5	0.5
NATOPS	0.922	0.87	0.944	0.939	0.889	**0.956**	0.87	0.86	0.85	0.883	0.85	0.85	0.883	0.883
PEMS-SF	**0.803**	0.744	0.688	0.751	0.699	N/A	N/A	0.705	0.734	0.711	0.705	0.734	0.711	0.73
PenDigits	0.98	0.98	**0.983**	0.98	0.978	0.975	0.948	0.973	0.939	0.977	0.973	0.939	0.977	0.977
Phoneme	0.177	0.19	**0.246**	0.175	0.11	0.19	0.19	0.104	0.151	0.151	0.104	0.151	0.151	0.151
RacketSports	0.849	0.816	0.862	0.868	0.803	0.875	**0.934**	0.868	0.842	0.803	0.868	0.842	0.803	0.858
SelfRegulationSCP1	**0.887**	0.874	0.771	0.739	0.874	0.734	0.71	0.771	0.765	0.775	0.771	0.765	0.775	0.786
SelfRegulationSCP2	**0.567**	0.539	**0.556**	0.55	0.472	N/A	0.46	0.483	0.533	0.539	0.483	0.533	0.539	0.539
SpokenArabicDigits	0.979	0.967	0.956	0.983	**0.99**	0.982	0.982	0.967	0.96	0.963	0.967	0.959	0.963	0.963
StandWalkJump	**0.533**	0.4	0.4	0.4	0.067	0.2	0.333	0.2	0.333	0.2	0.2	0.333	0.2	0.333
UWaveGestureLibrary	0.897	0.869	0.884	0.894	0.891	0.8	**0.916**	0.881	0.868	0.903	0.81	0.868	0.903	0.9
Avg. Rank	3.85	6.19	5.9	4.73	7.33	5.45	4.66	9.3	7.43	6.37	9.37	7.88	6.83	6.21
Wins (Ties)	**11**	**3**	6	5	2	5	8	0	2	2	0	2	1	2

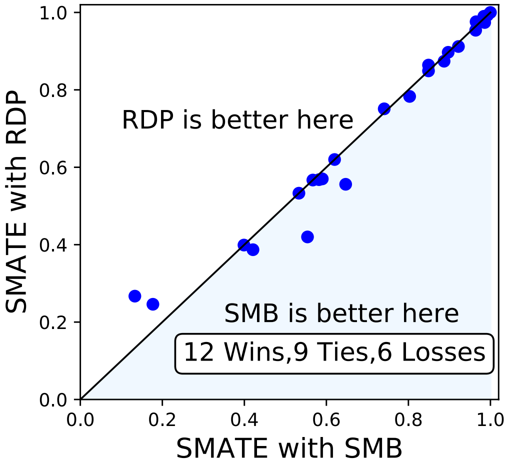
Experiments

Objectives

2. Effectiveness of Spatial Modelling Block (**SMB**)

- **27 datasets from UEA archive** in which SMATE has successfully executed
- SMATE **with SMB** *Vesrsus* SMATE **without SMB**:
  - [**17 Wins** | 8 Ties | 2 Losses]
- **SMB** *Versus* **others**:
  - Squeez-and-Excitation (**SE**) in MLSTM-FCNs [Karim et al., Neural Networks'19]
  - Random Dimension Permutation (**RDP**) in TapNet [Zhang et al., AAAI'20]





Experiments

Objectives

3. Effectiveness over competitors of Semi-supervised representation learning

Four datasets on different domains from UEA Archive[1]

- *ArticularyWordR*. (Motion)
- *Epilepsy* (Human Activity)
- *Heartbeat* (Audio Spectra)
- *SelfRegulationSCP1* (EEG/MEG)

Baselines:

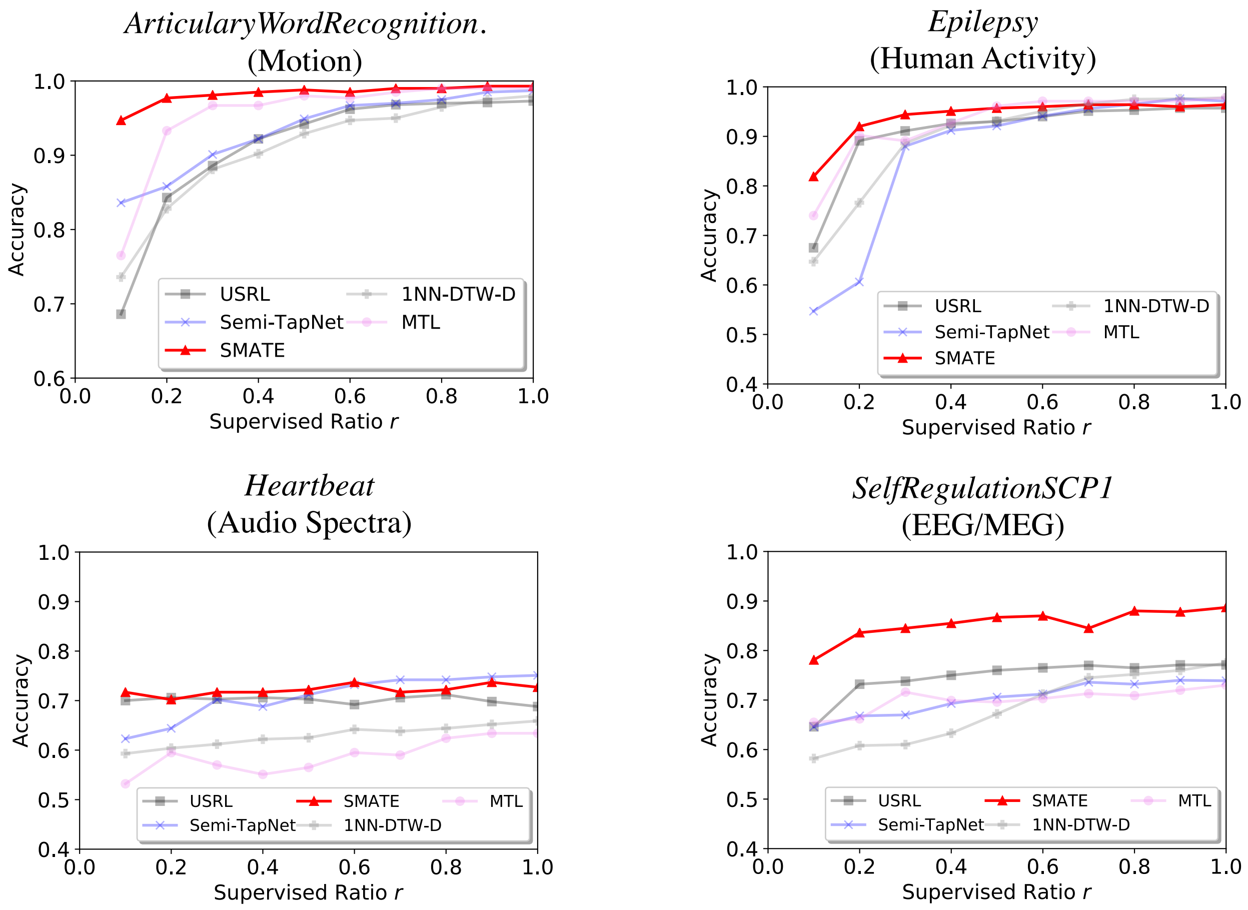
- **1NN-DTW-D** [Chen et al., KDD'13]
  - Initially designed for UTS
  - We adjust the distance measure with DTW_D which is designed for MTS
- **USRL** [Franceschi et al., NeurIPS'19]: SVM on unsupervised representation
- **Semi-TapNets** [Zhang et al., AAAI'20]: Learning unlabeled samples via Attentional Prototype Network
- **MTL** [Javed et al., PAKDD'20]: Multi-task learning with self-supervised features from forecasting task

1. www.timeseriesclassification.com

Experiments

Semi-supervised representation learning

- SVM on learned representation



SMATE performs generally the best among all semi-supervised models, especially under weak supervision

Experiments

Objectives

4. Computational Efficiency

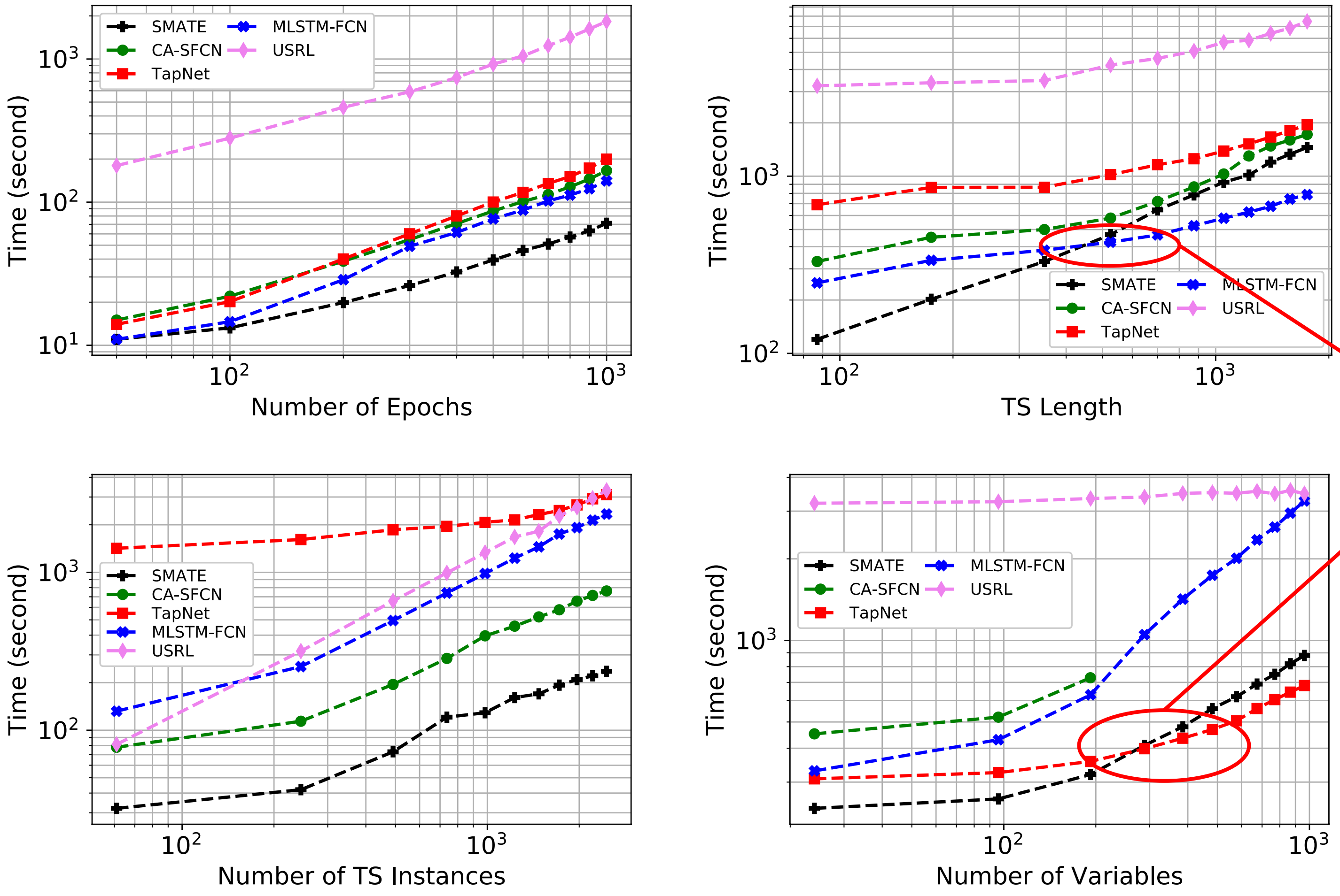
Factors	Dataset	(N, M, T)
Number of training epochs	*ArticularyWordRecognition*	(275, 9, 144)
TS length (**T**)	*EthanolConcentration*	(261, 3, **1751**)
Number of TS instance (**N**)	*LSST*	(**2459**, 6, 36)
Number of variables (**M**)	*PEMS-SF*	(267, **963**, 144)

Only compare with the Deep Learning models:

- **MLSTM-FCNs** [Karim et al., Neural Networks'19]
- **USRL**[Franceschi et al., NeurIPS'19]
- **TapNet** [Zhang et al., AAAI'20]
- **CA-SFCN** [Hao et al. IJCAI'20]

Experiments

Computational Efficiency



Globally, **SMATE** performs more efficiently than the competitors.

Exceptions:

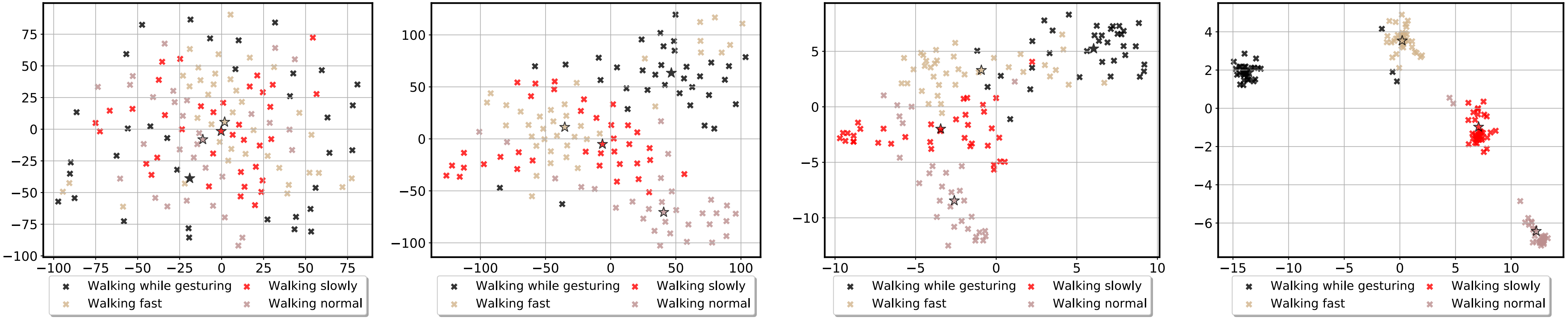
- on Extra Long TS
- on TS with Extra Huge Variable Numbers

Experiments

Objectives

5. Visualization & Interpretation of the Representation Space

- Interpretable for the effect of the weak supervision
- Interpretable for the classification results



(a) Without **any regularization**

(b) Regularization step 1: supervised initialization of the class centroids

(c) Regularization step 2: supervised adjustment of the class centroids

(d) Regularization step 3: unsupervised adjustment of the class centroids

Figure: The t-SNE visualization of the representation space for the *Epilepsy* dataset. We set the supervised ratio to 0.1. The colors of the embeddings represent their inherent labels, which are not fully adopted for training. The class centroids are marked by ⋆.

Conclusion

- SMATE allows learning a validated Spatio-temporal representation on MTS
    - Spatial Modelling Block (SMB) captures the **evolving interactions** between 1-D segments i.e., *Spatial Dynamics*, which represent better the *spatial interactions* than the State-of-the-art
- SMATE allows an efficient representation learning and classification for MTS
- SMATE learns an interpretable representation for:
    - The effect of the weak supervision
    - The classification results
- SMATE allows weak supervision on the embedding space
    - Combine both labelled and unlabelled samples
    - A low amount of labelled samples give promising classification result

Github page

Thank you for your attention

References

1. L. Wei and E. Keogh, “Semi-supervised time series classification,” in Proc. ACM SIGKDD’06.
2. M. N. Nguyen, X. L. Li, and S. K. Ng, “Positive unlabeled learning for time series classification,” in IJCAI, 2011.
3. Y. Chen, B. Hu, E. Keogh, and G. E. A. P. A. Batista, “DTW-D: Time Series Semi-Supervised Learning from a Single Example,” in KDD’13.
4. P. Schäefer and U. Leser, “Multivariate Time Series Classification with WEASEL+MUSE, AALTD’18.
5. F. Karim, S. Majumdar, H. Darabi, and S. Harford, “Multivariate lstm-fcns for time series classification,” Neural Networks, vol. 116, 2019.
6. J.-Y. Franceschi, A. Dieuleveut, and M. Jaggi, “Unsupervised Scalable Representation Learning for Multivariate Time Series,” NeurIPS 2019.
7. Y. Hao and H. Cao, “A New Attention Mechanism to Classify Multi- variate Time Series,” in IJCAI, 2020.
8. X. Zhang, Y. Gao, J. Lin, and C.-T. Lu, “TapNet: Multivariate Time Series Classification with Attentional Prototypical Network,” in AAAI’20.
9. S. Jawed, J. Grabocka, and L. Schmidt-Thieme, “Self-Supervised Learning for Semi-Supervised Time Series Classification,” in PAKDD’20.
10. M. Shokoohi-Yekta, B. Hu, H. Jin, J. Wang, and E. Keogh, “Generalizing DTW to the multi-dimensional case requires an adaptive approach HHS Public Access,” Data Min Knowl Discov, vol. 31, no. 1, pp. 1–31, 2017.