Soft Contrastive Learning for Time Series

Seunghan Lee[1], *Tayeoung Park*[1,2], *Kibok Lee*[1,2]

[1]*Department of Statistics and Data Science, Yonsei University,* [2]*Department of Applied Statistics, Yonsei University*

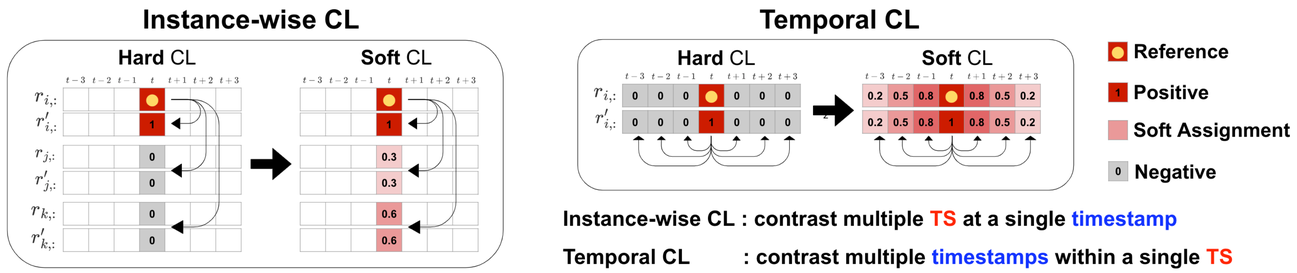
1. Contrastive Learning in Time Series

Contrastive Learning (CL): maximize similarities between positive pairs while minimizing similarities between negative pairs

- **Instance-wise CL**: contrasts the representations of TS instances
- **Temporal CL:** contrasts the representations of timestamps within a single TS
- **Soft CL**: assigns soft labels (between 0 and 1) to pairs

	T-Loss (NeurIPS 2019)	Self-Time (arxiv 2020)	TNC (ICLR 2021)	TS-SD (IJCNN 2021)	TS-TCC (IJCAI 2021)	TS2Vec (AAAI 2022)	Mixing-Up (PR Letters 2022)	CoST (ICLR 2022)	TimeCLR (KBS 2022)	TF-C (NeurIPS 2022)	CA-TCC (TPAMI 2023)	SoftCLT (Ours)
Instanse-wise CL	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓
Temporal CL		✓	✓		✓	✓					✓	✓
Hierarchical CL						✓						✓
Soft CL												✓

2. Soft Contrastive Learning for Time Series (SoftCLT)



Instance-wise CL : contrast multiple TS at a single timestamp

Temporal CL : contrast multiple timestamps within a single TS

- Two views of the same sample are denoted as r and $\tilde{r}$, respectively.
- Assign soft labels using sigmoid function: $\sigma(a) = 1/(1+\exp(-a))$

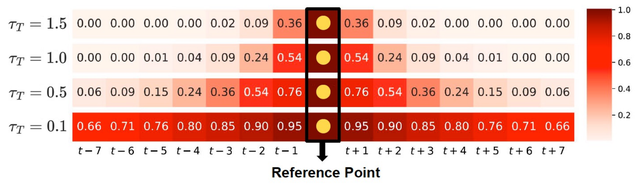
(1) Soft Instance-wise CL

- **Soft labels** : $w_I(i,i') = 2\alpha \cdot \sigma\left(-\tau_I \cdot D(x_i, x_{i'})\right)$
- **Distance:** min-max normalized distance metric $D(\cdot,\cdot)$ between two TS in data space (i.e. Euclidean distance, Cosine similarity, DTW)
- **Soft instance-wise contrastive loss**:
  - For conciseness, let $r_{i,t} = r_{i+2N,t}$ and $\tilde{r}_{i,t} = r_{i+N,t}$
  - Softmax probability of the relative similarity out of all similarities; for a pair of TS indices (i,i'): $p_I((i,i'),t) = \dfrac{\exp(r_{i,t} \circ r_{i',t})}{\sum_{j=1,j\neq i}^{2N} \exp(r_{i,t} \circ r_{j,t})}$

$$\ell_I^{(i,t)} = -\log p_I((i,i+N),t) - \sum_{j=1,j\neq\{i,i+N\}}^{2N} w_I(i,j \bmod N) \cdot \log p_I((i,j),t)$$

(2) Soft Temporal CL

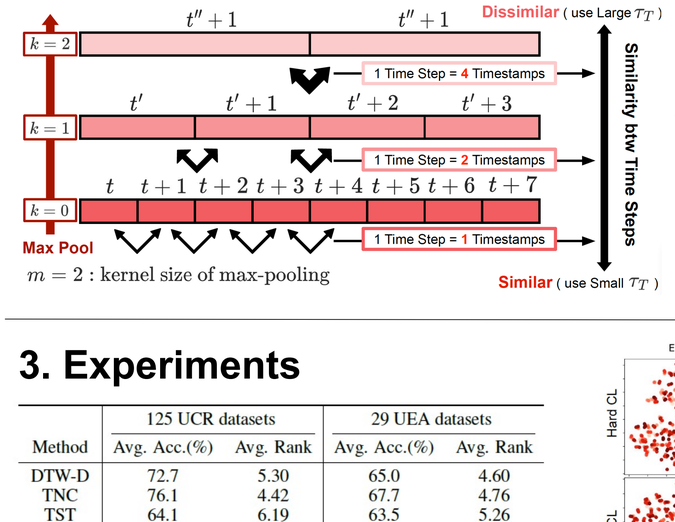
- **Soft labels** : $w_T(t,t') = 2 \cdot \sigma\left(-\tau_T \cdot |t-t'|\right)$
- **Distance:** difference between two time step pairs (t,t')
- **Soft temporal contrastive loss**:
  - For conciseness, let $r_{i,t} = r_{i,t+2T}$ and $\tilde{r}_{i,t} = r_{i,t+T}$
  - Softmax probability of the relative similarity out of all similarities; for a pair of time indices (t,t'): $p_T(i,(t,t')) = \dfrac{\exp(r_{i,t} \circ r_{i,t'})}{\sum_{s=1,s\neq t}^{2T} \exp(r_{i,t} \circ r_{i,s})}$



$$\ell_T^{(i,t)} = -\log p_T(i,(t,t+T)) - \sum_{s=1,s\neq\{t,t+T\}}^{2T} w_T(t,s \bmod T) \cdot \log p_T(i,(t,s))$$

(3) Hierarchical Soft Temporal CL

- Adopt the hierarchical contrastive loss proposed in TS2Vec (Yue et al., 2022)
- Losses are computed on intermediate representations after max-pooling along the temporal axis and then aggregated



(4) Total Loss

Weighted avg. of soft instance-wise and temporal CL loss

$$\mathcal{L} = \frac{1}{4NT} \sum_{i=1}^{2N} \sum_{t=1}^{2T} (\lambda \cdot \ell_I^{(i,t)} + (1-\lambda) \cdot \ell_T^{(i,t)})$$

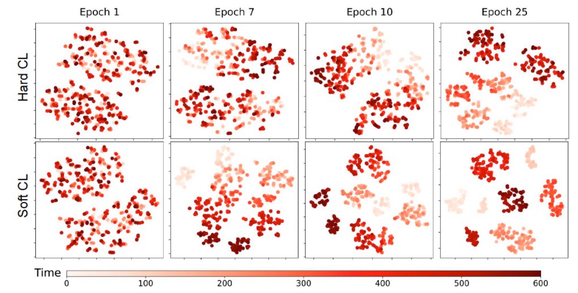
3. Experiments

Method	125 UCR datasets Avg. Acc.(%)	125 UCR datasets Avg. Rank	29 UEA datasets Avg. Acc.(%)	29 UEA datasets Avg. Rank
DTW-D	72.7	5.30	65.0	4.60
TNC	76.1	4.42	67.7	4.76
TST	64.1	6.19	63.5	5.26
TS-TCC	75.7	4.29	68.2	4.38
T-Loss	80.6	3.50	67.5	3.86
TS2Vec	83.0	2.80	71.2	3.28
+ Ours	**85.0(+ 2.0)**	**1.49**	**75.1(+ 3.9)**	**1.86**

(1) Classification Task

- 125 UCR datasets: 125 univariate TS
- 29 UEA datasets: 29 multivariate TS

(2) t-SNE Visualization of Temporal Representations



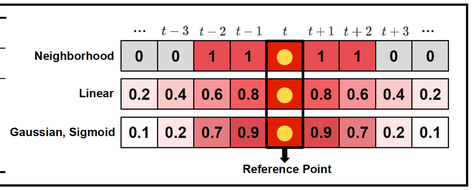
- Representations learned with hard & soft CL over training epochs
- Hard CL finds coarse-grained neighborhood relationships, while soft CL finds more fine-grained relationships

(3) Ablation Studies

Soft assignment: Instance-wise	Soft assignment: Temporal	UCR datasets Avg. Acc.(%)	UEA datasets Avg. Acc.(%)
		82.3	70.5
✓		83.9 (+1.6)	73.0 (+2.5)
	✓	83.7 (+1.4)	73.8 (+3.3)
✓	✓	**85.0** (+2.7)	**74.2** (+3.7)

(3-a)

Temporal CL: Method	Avg. Acc.(%)
Neighbor	76.1
Linear	77.2
Gaussian	83.5
Sigmoid	**83.7**



(3-b)

Inst. CL: Metric	Temporal CL: Hard	Temporal CL: Soft
COS	83.7	84.7
EUC	83.9	84.8
DTW	83.9	**85.0**
TAM	83.9	**85.0**

(3-c)

- (3-a) Applying soft assignments to instance-wise or temporal CL provides a performance gain, and applying them to both dimensions results in the best performance
- (3-b) Different assignments functions for temporal CL; Sigmoid results in the best performance
- (3-c) Difference choices of the distance metric; COS (cosine similarity), EUC (Euclidean distance), DTW (dynamic time warping), and TAM (time alignment measurement).

References

Yue, Zhihan, et al (2020). "Ts2vec: Towards universal representation of time series." In AAAI 2022