Long-tailed Time Series Classification via Feature Space Rebalancing

Pengkun Wang [1], *Xu Wang* [1], *Binwu Wang* [1], *Yudong Zhang* [1], *Lei Bai* [2,*], *and Yang Wang* [1,*]

[1] University of Science and Technology of China (USTC), China
[2] Shanghai AI Laboratory, China

Reporter: Pengkun Wang

DASFAA 2023

Outline

- Background
- Our Method
- Experiment
- Conclusion

Background

What is *Long-tailed Distribution?*

- **Long-tailed Class Distribution**
  - 80/20 Principle (Pareto Principle)
  - In real-world applications, training samples typically exhibit a long-tailed class distribution, where a small portion of classes have massive sample points but the others are associated with only a few.

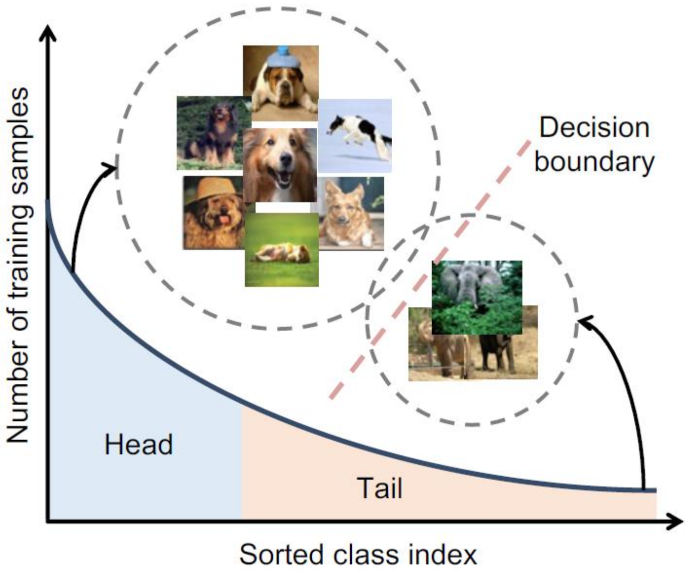


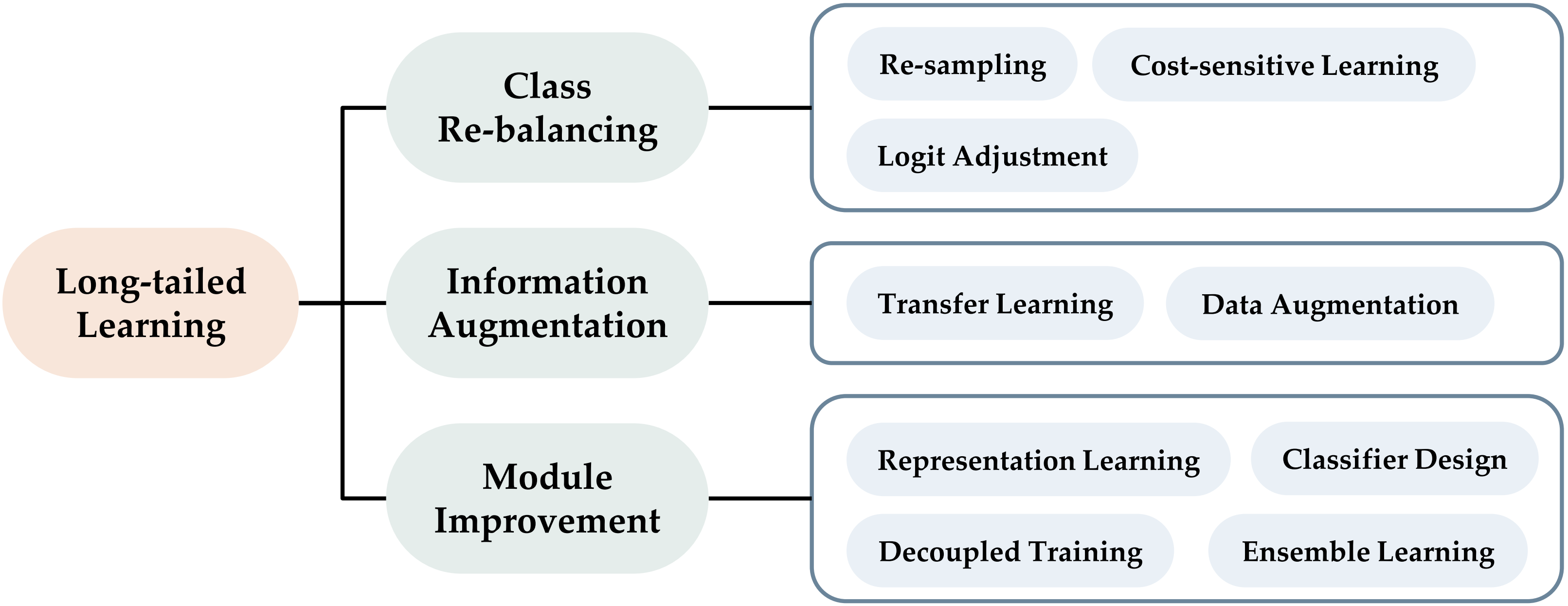
Fig. 1 The label distribution of a long-tailed dataset [1].

[1] Zhang, Yifan, et al. "Deep long-tailed learning: A survey." arXiv preprint arXiv:2110.04596 (2021).

Background

Long-tailed Learning

- **Challenges**
  - The trained model is easily biased towards head classes
  - The trained model performs poorly on tail classes
- **Classic Methods**



Background

Long-tailed Time Series Classification (Long-tailed TSC)

- **General TSC Methods**
  - focus on learning decision boundaries from artificially balanced datasets
- **Imbalanced TSC Methods**
  - only explored simple tasks with two categories or seriously ignore the long-tailed nature

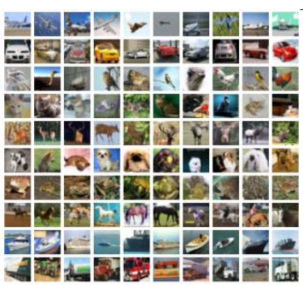
I. What are the challenges of Long-tailed TSC relative to the other domain (especially vision)?

II. Are general long-tailed recognition methods (GLR) applicable to the time series domain?

III. How to realize efficient Long-tailed TSC ?

Our Method

- **Long-tailed TSC *vs.* GLR (Question 1 & 2)**
  - Similarities (ignoring the data dimension)
    - Long-tailed TSC and GLR can be regarded as homogeneous problems
    - Existing methods have certain generalizations to Long-tailed TSC
  - Differences (considering the data dimension)
    - Time series have unique temporal properties
    - Existing methods can not model temporal and correlation information between variables
- **Data Limitations in Long-tailed TSC**



CIFAR-100-LT
ImageNet-LT
iNaturalist 2018
… …

Dataset in GLR

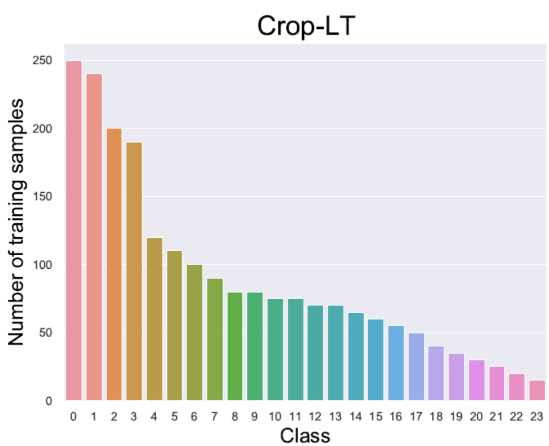
Dataset in Long-tailed TSC

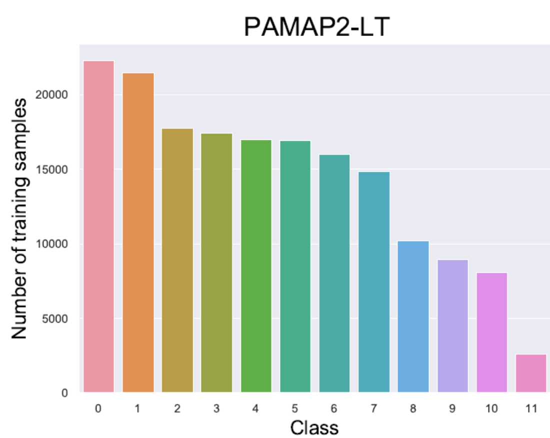
Our Method (Question 3)

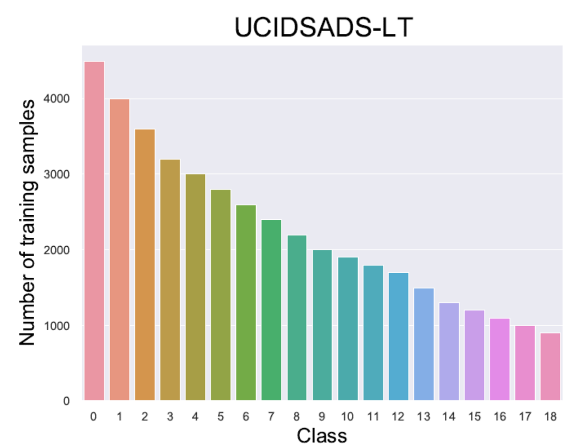
Derived Long-tailed TSC

- Three derived datasets
  - Crop-LT
  - PAMAP2-LT
  - UCIDSADS-LT
- More classes
  - 12~24
- Higher imbalance rate
  - 5~17

Dataset	Crop-LT	PAMAP2-LT	UCIDSADS-LT
Variable⋆	1	36	45
Class⋆	24	12	19
Length	46	20	20
Training set⋆	2,145	173,309	42,692
Validation set⋆	1,200	6,000	9,500
Test set⋆	16,800	12,000	19,000
Head classes⋆	2 (# >200)	2 (# >20000)	4 (# >3000)
Medium classes⋆	5 (200≥ # ≥100)	6 (20000≥ # ≥11000)	9 (3000≥ # ≥1700)
Tail classes⋆	17 (# <100)	4 (# <11000)	6 (# <1700)
Imbalance ratio	17	9	5



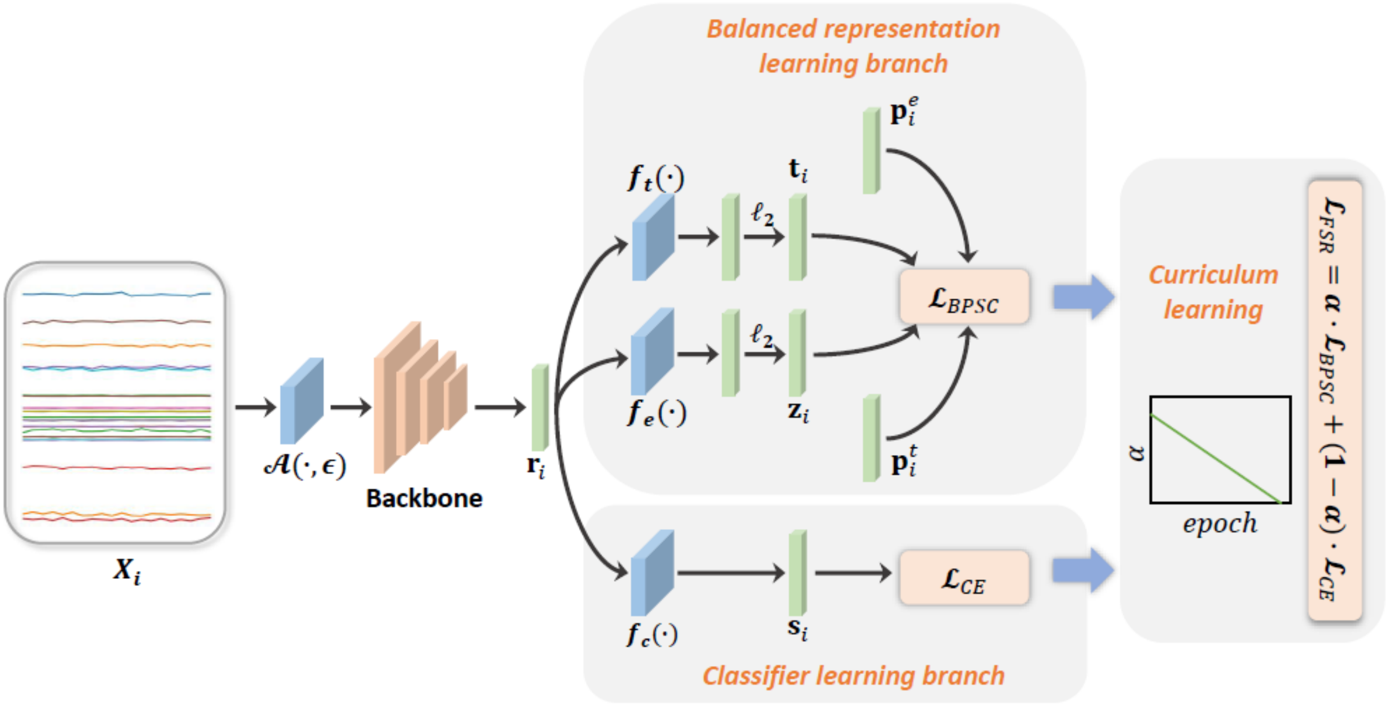




Feature Space Rebalancing (FSR)

FSR: A Hybrid Network

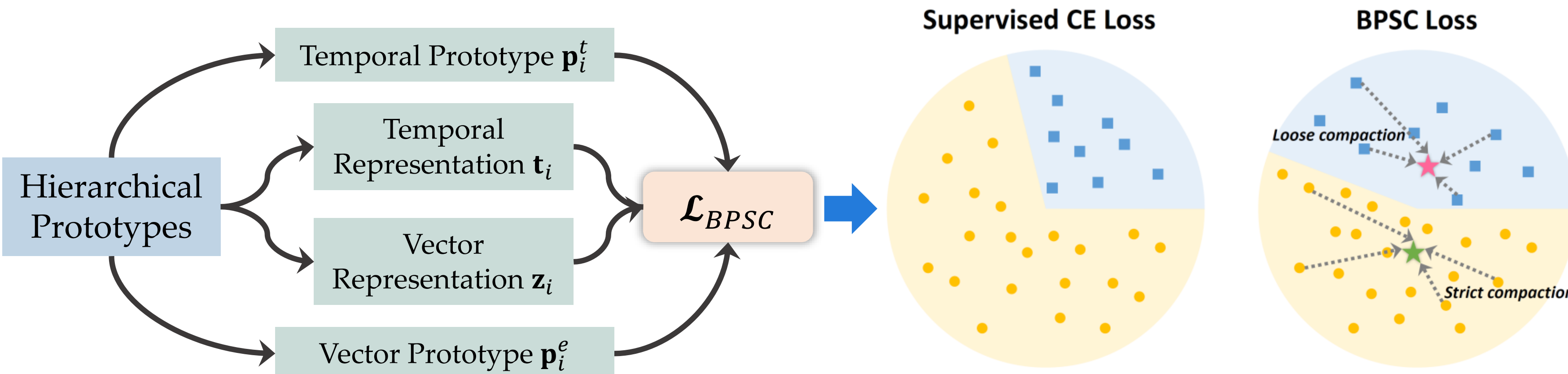
- Balanced representation learning branch
- Classifier learning branch
- Curriculum learning
- Adaptive temporal augmentation



Feature Space Rebalancing (FSR)

Balanced Contrastive Learning (BCL)

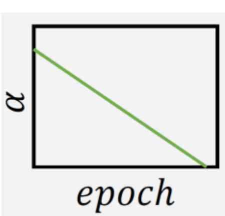
- Learns a balanced feature space that balances head and tail classes while achieving intra-class compactness and inter-class separability



Cross-Entropy Loss (CE) + Balanced Prototypical Supervised Contrastive Loss (BPSC)

$$\mathcal{L}_{BPSC}(\mathbf{z}_i) = -\,log\frac{\omega_i \cdot e^{(\mathbf{z}_i \cdot \mathbf{p}^e_{y_i}/\tau)}}{\sum_{j=1, j\neq y_i}^{C} e^{(\mathbf{z}_i \cdot \mathbf{p}^e_j/\tau)}} + \beta \cdot \left(-log\frac{\omega_i \cdot e^{(\mathbf{z}_i \cdot \mathbf{p}^t_{y_i}/\tau)}}{\sum_{j=1, j\neq y_i}^{C} e^{(\mathbf{z}_i \cdot \mathbf{p}^t_j/\tau)}}\right)$$

$$with \quad \omega_i = e^{-n_{max}/n_i \cdot \rho_b}$$



Curriculum Learning

$$\mathcal{L}_{FSR} = \alpha \cdot \mathcal{L}_{BPSC} + (1-\alpha) \cdot \mathcal{L}_{CE}$$

Feature Space Rebalancing (FSR)

Adaptive Temporal Augmentation (ATA)

Traditional data augmentation

Step 1. Define a parametric temporal augmentation

$$\hat{x}_i = \mathcal{A}(x_i, \epsilon)$$

Step 2. Assign same degrees of augmentation to each class

$$\epsilon_i' = \epsilon$$

✓ Enlarge the feature space

✗ Balance feature space

Not suitable for long-tailed learning

Adaptive data augmentation

Step 1. Define a parametric temporal augmentation

$$\hat{x}_i = \mathcal{A}(x_i, \epsilon)$$

Step 2. Assign different degrees of augmentation to each class according to the sample size

$$\epsilon_i' = \epsilon \cdot e^{-\frac{n_i}{n_{max}} \cdot \rho_a}$$

✓ Enlarge the feature space

✓ Balance feature space

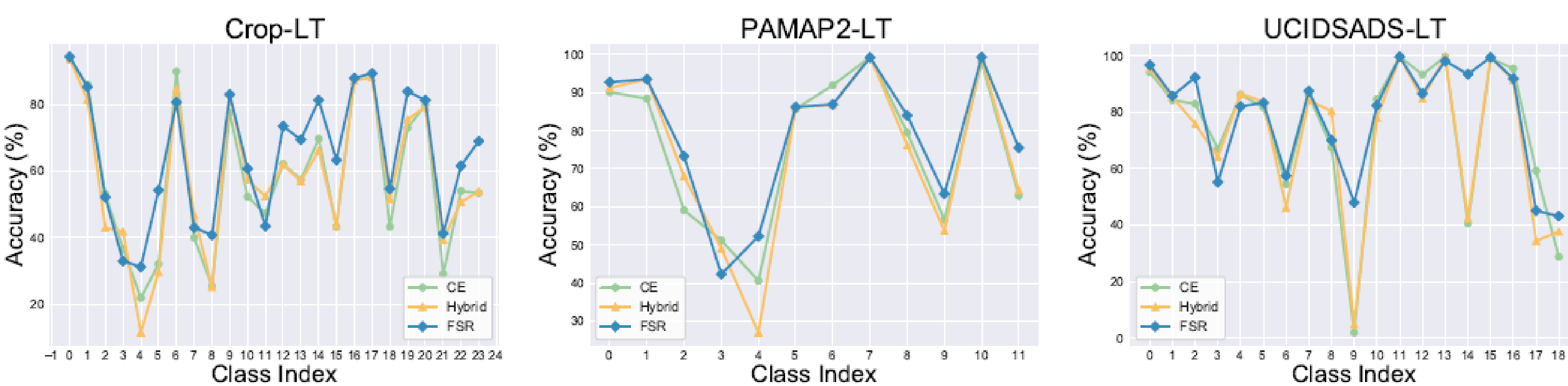
Suitable for long-tailed learning

Experiment

- Compare with state-of-the-art methods

Method	Crop-LT				PAMAP2-LT				UCIDSADS-LT			
	Head	Medium	Tail	All	Head	Medium	Tail	All	Head	Medium	Tail	All
CE [10]	89.65	46.77	57.93	58.25	89.25	71.30	74.25	75.28	82.15	72.87	70.51	74.08
Focal [20]	90.15	46.37	53.42	55.01	93.90	72.90	73.83	76.71	85.45	70.06	65.45	71.84
CB [6]	79.43	38.63	63.19	59.43	90.75	70.87	76.73	76.13	79.85	72.67	72.77	74.21
LDAM [3]	85.93	45.11	62.14	60.58	85.65	71.17	73.38	74.32	77.95	71.23	66.82	71.25
BS [22]	81.14	44.60	63.71	61.18	89.35	69.80	78.30	75.89	78.08	76.28	67.35	73.84
Seesaw [23]	85.29	47.69	61.53	60.62	92.35	72.12	75.03	76.46	79.70	72.14	80.53	76.38
Hybrid [24]	87.57	42.14	59.96	58.55	92.35	69.40	73.40	74.56	80.22	72.04	67.53	72.34
KCL [14]	88.21	47.98	62.84	61.85	92.15	70.82	75.27	75.86	80.33	71.57	71.95	73.53
TSC [19]	88.01	48.10	63.27	62.17	91.80	71.03	77.24	76.56	81.59	74.37	72.79	75.39
FSR	89.75	50.29	66.31	**64.93**	93.10	73.70	80.55	**79.05**	82.55	77.48	78.57	**78.89**

- Visualization of accuracy on each class

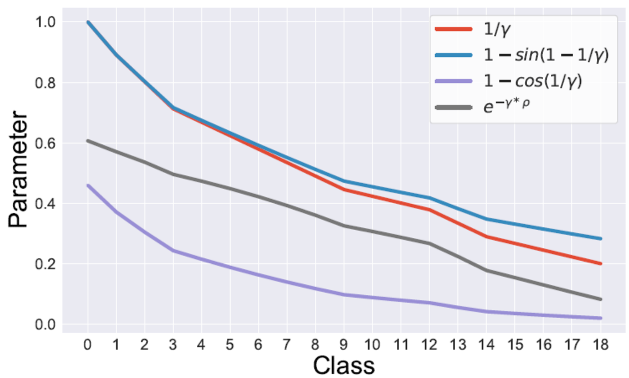


Experiment

- Mitigative compaction factor in BCL

$$\omega_i = e^{-n_{max}/n_i \cdot \rho_b}$$

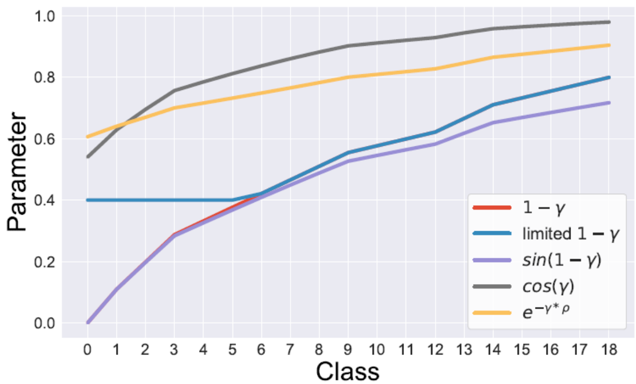
$$\gamma = n_{max}/n_i$$



δ	Head	Medium	Tail	All
Hybrid	80.22	72.04	67.53	72.34
$1/\gamma$	80.17	74.39	72.82	75.11
$1-sin(1-1/\gamma)$	80.30	74.87	73.62	75.62
$1-cos(1/\gamma)$	80.89	76.78	74.05	76.78
$e^{-\gamma\cdot\rho}$	81.75	77.50	76.43	78.06

- Mitigative augmentation factor in ATA

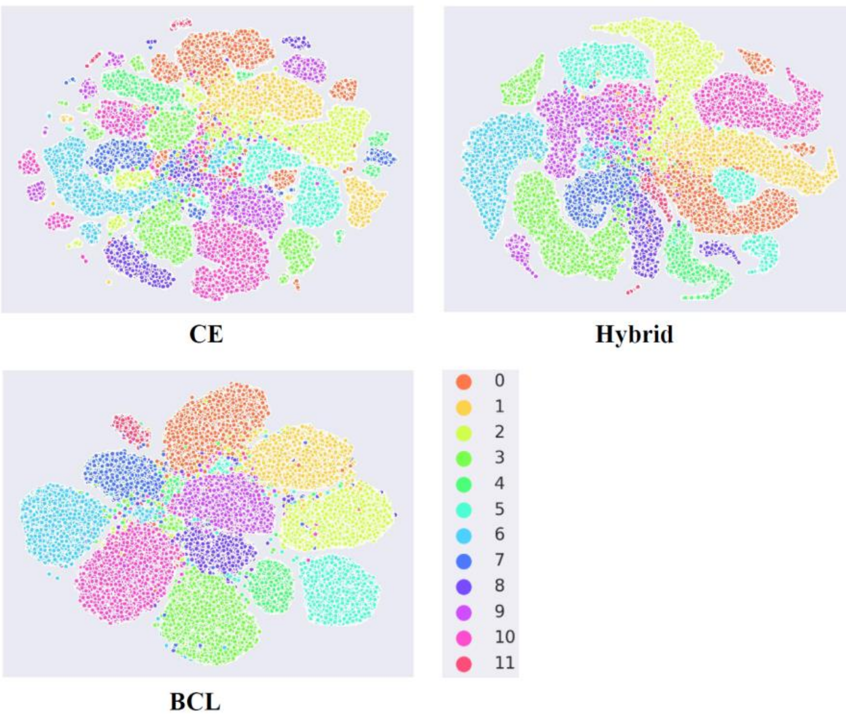
$$\epsilon_i' = \epsilon \cdot \delta = \epsilon \cdot e^{-\gamma \cdot \rho_a}$$



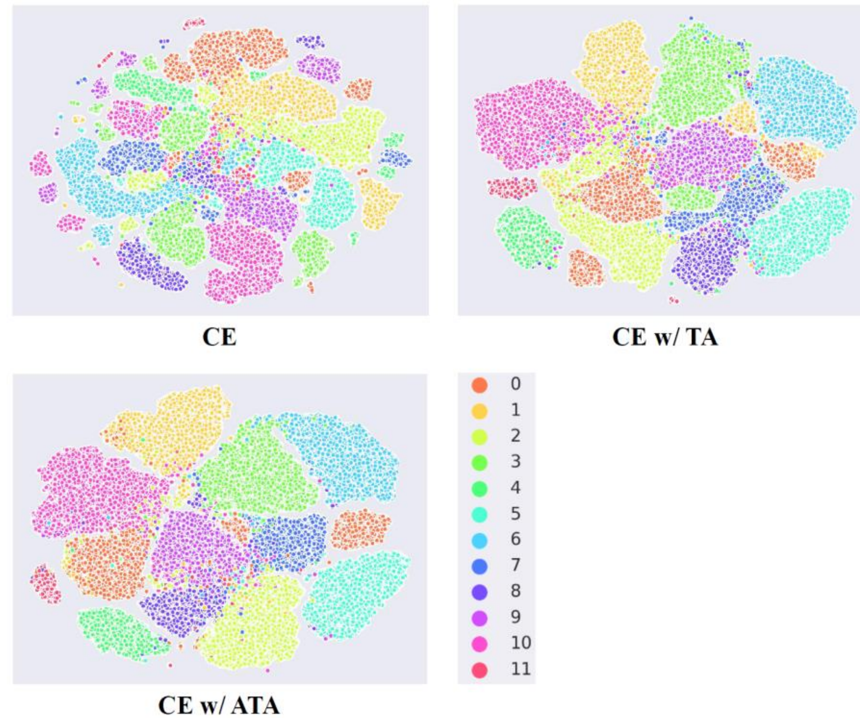
δ	Head	Medium	Tail	All
CE	82.15	72.87	70.51	74.08
$1-\gamma$	79.50	73.43	75.33	75.30
$limited\ 1-\gamma$	81.14	73.60	76.21	76.01
$sin(1-\gamma)$	80.16	73.57	75.92	75.70
$cos(\gamma)$	81.50	73.81	76.47	76.27
$e^{-\gamma\cdot\rho}$	82.55	74.00	77.30	76.84

Experiment

- **T-SNE Visualization**



CE, Hybrid, and BCL



CE, CE with TA, and CE with ATA

Experiment

- Ablation study for BCL

Method	Head	Medium	Tail	All
BCL	92.40	71.45	78.23	77.20
w/o BPSC & HP	92.35	69.40	73.40	74.56
w/o BPSC	92.48	70.75	76.22	76.20
w/o HP	92.27	71.24	77.87	76.96

- Ablation study for ATA

Method	Crop-LT				UCIDSADS-LT			
	Head	Medium	Tail	All	Head	Medium	Tail	All
CE	89.65	46.77	57.93	58.25	82.15	72.87	70.51	74.08
w/ TA	90.28	46.06	58.57	58.61	77.07	73.40	75.77	74.92
w/ ATA	89.72	42.08	60.97	59.43	82.55	74.00	77.30	76.84
BCL	87.86	49.29	63.96	62.90	81.75	77.50	76.43	78.06
w/ TA	86.07	45.14	64.29	62.11	81.33	75.57	74.95	76.59
w/ ATA	89.75	50.29	66.31	64.93	82.55	77.48	78.57	78.89

Conclusion

- We discuss the long-tailed time series classification learning and construct three long-tailed datasets. To the best of our knowledge, this is the first long-tailed time series classification work, which fills a gap in the field.
- To address the above Long-tailed TSC, we propose a novel Feature Space Rebalancing (FSR) strategy. First, we design a Balanced Contrastive Learning (BCL) to avoid imbalanced feature spaces by introducing compaction factors and hierarchical prototypes in the supervised contrastive loss. Second, we rethink traditional data augmentation and propose an Adaptive Temporal Augmentation (ATA) to balance the augmented feature space.
- We conduct extensive experiments on the three proposed datasets and demonstrate that the proposed FSR is more suitable for long-tailed time series classification than existing methods.

Thanks for your listening!

For more details, please refer to our paper!

Reporter: Pengkun Wang

E-mail: pengkun@mail.ustc.edu.cn