Temporal-Frequency Co-training for Time Series Semi-supervised Learning

Zhen Liu[1], **Qianli Ma**[1,2*], **Peitian Ma**[1], **Linghao Wang**[1]

[1]School of Computer Science and Engineering, South China University of Technology, Guangzhou, China
[2]Key Laboratory of Big Data and Intelligent Robot (South China University of Technology), Ministry of Education
cszhenliu@mail.scut.edu.cn, qianlima@scut.edu.cn, ma_scuter@163.com, wlhsama@gmail.com

Abstract

Semi-supervised learning (SSL) has been actively studied due to its ability to alleviate the reliance of deep learning models on labeled data. Although existing SSL methods based on pseudo-labeling strategies have made great progress, they rarely consider time-series data's intrinsic properties (e.g., temporal dependence). Learning representations by mining the inherent properties of time series has recently gained much attention. Nonetheless, how to utilize feature representations to design SSL paradigms for time series has not been explored. To this end, we propose a **T**ime **S**eries SSL framework via **T**emporal-**F**requency **C**o-training (**TS-TFC**), leveraging the complementary information from two distinct views for unlabeled data learning. In particular, TS-TFC employs time-domain and frequency-domain views to train two deep neural networks simultaneously, and each view's pseudo-labels generated by label propagation in the representation space are adopted to guide the training of the other view's classifier. To enhance the discriminative of representations between categories, we propose a temporal-frequency supervised contrastive learning module, which integrates the learning difficulty of categories to improve the quality of pseudo-labels. Through co-training the pseudo-labels obtained from temporal-frequency representations, the complementary information in the two distinct views is exploited to enable the model to better learn the distribution of categories. Extensive experiments on 106 UCR datasets show that TS-TFC outperforms state-of-the-art methods, demonstrating the effectiveness and robustness of our proposed model.

Introduction

Time series data is widely available in real-world scenarios, such as human activity recognition (Yang et al. 2015), fault diagnosis (Liu et al. 2016), and clinical analysis (Song et al. 2018). In recent years, time series classification algorithms have received much attention due to the powerful feature extraction capability of deep neural networks (Ismail Fawaz et al. 2019). However, time series data have complex dynamic properties that lead to over-reliance on human expert knowledge for annotation. As massive labeled time series data are difficult to obtain, the performance of deep learning models is easily limited. Semi-supervised learning (SSL) can make use of abundant unlabeled data, which can effectively alleviate the reliance of deep learning models on labeled data. Therefore, it is of great significance to explore SSL paradigms for time series.

Consistent regularization (Sajjadi, Javanmardi, and Tasdizen 2016; Laine and Aila 2017) and pseudo-labeling (Lee et al. 2013; Iscen et al. 2019) are two main strategies for SSL. Consistent regularization generally adds an unsupervised loss term to the learning objective, enabling the model to learn from unlabeled data. In the field of computer vision, consistent regularization of samples with strong and weak augmentations is a widely proven effective strategy (Berthelot et al. 2019). Nevertheless, related studies (Wen et al. 2021; Iwana and Uchida 2021) show that the selection and combination of data augmentation techniques that facilitate time series classification remains a challenge. Recently, MTL (Jawed, Grabocka, and Schmidt-Thieme 2020) and SemiTime (Fan et al. 2021) construct time-series forecasting loss and temporal relation prediction loss as regularization terms utilizing the relationship between sampled subseries, respectively. However, the sampling bias of the subseries limits the performance of the above methods.

Pseudo-labeling enables SSL by assigning labels to unlabeled data. Generating pseudo-labels by utilizing the classifier prediction results is a common approach (Lee et al. 2013), whereas the quality of pseudo-labels overly depends on the learning ability of the classifier. Meanwhile, based on the assumption of label consistency of nearest neighbor samples, some studies (Kamnitsas et al. 2018; Iscen et al. 2019) employ label propagation to create the nearest neighbor graph for labeled and unlabeled image data. Then, the node information of the labeled data is utilized to propagate pseudo-labels to the unlabeled data (Wang and Zhang 2007). However, the quality of pseudo-labels obtained by label propagation is susceptible to the influence of the distance of the sample distribution between categories. In addition, how to obtain high-quality pseudo-labels via feature representations by exploiting the intrinsic properties of time series has not been fully explored.

In recent years, time series representation learning has attracted much attention. For example, TS-TCC (Eldele et al. 2021) and TS2Vec (Yue et al. 2022) design universal representation learning frameworks based on contrastive learning for capturing temporal dependence of time series.

*Qianli Ma is the corresponding author.

Copyright © 2023, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

CoST (Woo et al. 2022), FEDformer (Zhou et al. 2022), and BTSF (Yang and Hong 2022) show that the representation learning capability of deep learning models can be effectively improved by combining time-domain with frequency-domain information from the same time series data source. Meanwhile, the aforementioned studies indicate that representation learning can obtain more discriminative features than the original time series by mining the inherent properties of the time series, thus facilitating downstream tasks.

Therefore, this paper proposes a **T**ime-**S**eries SSL framework via **T**emporal-**F**requency **C**o-training (**TS-TFC**). Specifically, we treat the time domain and frequency domain of the same time series data source as two distinct views, and train two deep neural networks separately. At the same time, label propagation is employed to create the nearest neighbor graph in the representation space to dynamically obtain pseudo-labels for each view's unlabeled data. Then, we allow the pseudo-labels obtained from time-domain and frequency-domain views to guide the training of each other's classifier, using the complementary information from two distinct views to improve the robustness of representations. To enhance the discriminative of representations between categories, we propose a temporal-frequency supervised contrastive learning module to mitigate the negative effect of feature distribution distance on propagating pseudo-labels. In addition, we integrate the learning difficulty of categories and dynamically select high-quality propagating pseudo-labels for co-training, thereby enabling TS-TFC to better learn the distribution of categories.

In summary, the contributions of this paper are as follows:

- We propose a temporal-frequency co-training model for time-series SSL, utilizing the complementary information from two distinct views for unlabeled data learning. Specifically, we employ each view's pseudo-labels generated by label propagation in the representation space to guide the training of the other view's classifier.
- We propose a temporal-frequency supervised contrastive learning module for acquiring discriminative representations, thus mitigating the negative effect of feature distribution distance on generated pseudo-labels through label propagation. Also, we dynamically select high-quality pseudo-labels for co-training by integrating the learning difficulty of the categories.
- Extensive experiments on 106 UCR datasets show that TS-TFC outperforms state-of-the-art methods, which demonstrates the effectiveness and robustness of our proposed model. In addition, visualization analysis indicates that feature representations can capture the discriminative region of time series, thereby boosting accuracy.

Related Work

Semi-Supervised Learning in Time Series

Traditional time series SSL methods (Wei and Keogh 2006; Chen et al. 2013; Xu and Funaya 2015) are mainly based on the Dynamic Time Warping (DTW) (Müller 2007) distance, utilizing labeled data in the original time series space to estimate the pseudo-labels of unlabeled data. Although DTW distance can make good use of temporal dependence, traditional methods still have disappointing performance compared to deep learning-based time-series classification algorithms (Ismail Fawaz et al. 2019). With the advantage of the deep learning models, shapelet (Wang et al. 2019) is introduced to time series SSL. Meanwhile, MTL (Jawed, Grabocka, and Schmidt-Thieme 2020) designs an SSL strategy by combining time-series forecasting and classification tasks. TapNet (Zhang et al. 2020) constructs an attentional prototype network that can classify multivariate time series (MTS) data with limited labeled data. SMATE (Zuo, Zeitouni, and Taher 2021) proposes a semi-supervised spatio-temporal representation learning method on MTS. SemiTime (Fan et al. 2021) constructs an unsupervised temporal relation prediction loss for unlabeled time series data learning. Although the above studies alleviate the reliance of deep learning models on time-series labeled data, how to design a time-series SSL model using feature representations is still in the exploration stage.

Representation Learning in Time Series

Due to the powerful feature extraction capability of deep neural networks, time series representation learning has received much attention in recent years. For example, T-loss (Franceschi, Dieuleveut, and Jaggi 2019) employs an unsupervised triplet loss based on the assumption of subseries consistency to learn universal representations. TS-TCC (Eldele et al. 2021) constructs an unsupervised contrastive learning (Chen et al. 2020) framework based on temporal and contextual consistency assumptions. In addition, CoST (Woo et al. 2022) comprises both time and frequency domains contrastive learning loss to learn disentangled seasonal-trend representations for time series forecasting. BTSF (Yang and Hong 2022) designs an unsupervised contrastive learning module via iterative bilinear temporal-spectral fusion from both time and frequency domains. The above studies indicate that contrastive learning is effective in mining the inherent properties of time series for representation learning. However, the sampling bias between subseries may introduce false positive pairs for contrastive learning (Yue et al. 2022). Moreover, unsupervised contrastive learning strategies rely on effective data augmentation strategies (Chen et al. 2020). Unlike the above work, we propose a supervised contrastive learning module based on temporal and frequency representations. Without the aid of data augmentation, the entire time series samples are utilized for contrastive learning based on category information to learn discriminative representations.

Label Propagation

Label propagation (LP) is a graph-based transduction learning method that can create a graph via features to establish relationships between labeled and unlabeled samples (Xiaojin and Zoubin 2002; Zhou et al. 2003). Based on the assumption that neighboring samples have consistent labels, LP employs the information of labeled nodes in the graph to make inductive inferences about the labels of unlabeled nodes (Rohrbach, Ebert, and Schiele 2013; Liu et al. 2019). With this advantage, LPDeepSSL (Iscen et al. 2019) utilizes

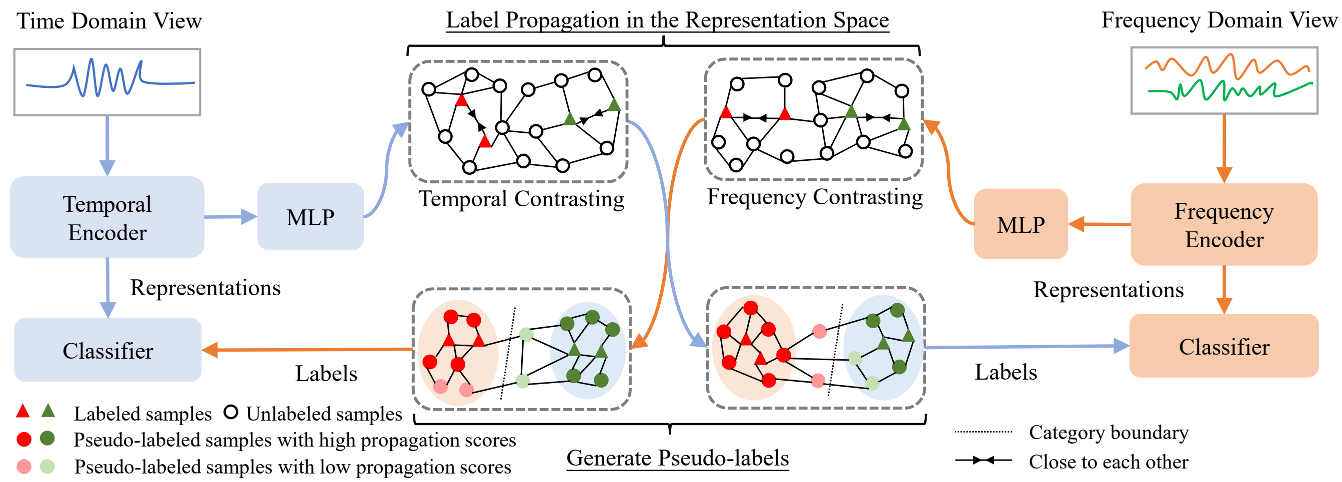


Figure 1: The general architecture of the TS-TFC framework. The oval shading indicates pseudo-labels are selected based on the category learning difficulty, while samples not in the oval shading are not used for classification training.

feature representations of image data as nodes and adopts LP to obtain pseudo-labels of unlabeled data. Although the above work achieved good performance in SSL, they do not consider the inherent properties of time series in representation learning. In addition, owing to the different learning difficulties of the categories in deep neural networks (Bengio et al. 2009; Zhang et al. 2021), samples of different categories may be adjacent to each other in the representation space, resulting in poor quality of the pseudo-labels generated by LP. Therefore, it remains a challenge to obtain robust time series representations to improve the quality of pseudo-labels on unlabeled data.

Method

Problem Definition

Given a set of time series $\mathcal{X} = \{x_1, x_2, \ldots, x_N\}_{n=1}^{N}$, each time series $x_n \in R^{T \times K}$, where T denotes the sequence length and K is the feature dimension. For the time series semi-supervised classification problem, we assume that there are L labeled time series, denoted as $D^L = \{(x_1, y_1), (x_2, y_2), \ldots, (x_L, y_L)\}_{l=1}^{L}$, where y_L denotes the label of the sample x_L. Also, there are U unlabeled time series, denoted as $D^U = \{x_1, x_2, \ldots, x_U\}_{u=1}^{U}$, where $L + U = N$. The goal of this paper is to generate pseudo-labels of samples in D^U to assist the model to better learn the distribution of different categories of samples in $\mathcal{X}$. For samples containing labels and pseudo-labels, we employ cross-entropy for training, which is defined as:

$$\mathcal{L}_{cls}(\mathcal{X}, \mathcal{Y}) = -\frac{1}{N}\sum_{n=1}^{N}\sum_{c=1}^{C}\mathbf{I}\left(y_n = c\right)\log\left(P\left(\hat{y}_n = c \mid x_n\right)\right), \quad (1)$$

where $\mathcal{Y}$ denotes the set of lables, C indicates the number of category, $\mathbf{I}(b) = 1$ if b is true otherwise 0.

Model Architecture

The overall architecture of TS-TFC is shown in Figure 1. Time domain view represents the original time series $\mathcal{X}$. Frequency domain view consists of the amplitude and phase of the original time series $\mathcal{X}$ after converting it to the frequency domain using the Fast Fourier Transform (Nussbaumer 1981). Temporal encoder and frequency encoder have the same network architecture and are utilized to learn temporal and frequency representations of the time series $\mathcal{X}$, respectively.

The temporal (frequency) encoder is composed of a three-layer Fully Convolutional Network (FCN) (Wang, Yan, and Oates 2017), which can achieve excellent performance in the field of time series classification (Ismail Fawaz et al. 2019). In addition, we employ an MLP layer to perform a nonlinear transformation of the feature representations acquired by the encoder for contrastive learning, which has been proven to be effective in improving the learning ability of the encoder (Chen et al. 2020; Khosla et al. 2020). Further, the nearest neighbor graph is created utilizing the representations output from the MLP layer after performing temporal-frequency contrasting, and the pseudo-label of unlabeled data is generated by label propagation. Meanwhile, we utilize curriculum pseudo-labeling to select high-quality propagating pseudo-labels for co-training to mitigate the negative impact of incorrect pseudo-labels. Also, the MLP consists of two-layer linear neural networks combined with a nonlinear ReLU function and a dropout operation, while the classifier consists of a single-layer linear neural network.

Temporal-Frequency Contrasting

Label propagation is based on the assumption of label consistency of neighboring samples to generate pseudo-labels. However, samples between different categories may be adjacent in the representation space, leading to the generation of low-quality propagating pseudo-labels. To address this issue, we propose a temporal-frequency contrastive learning module, which makes the samples of different categories in the representation space more dispersed and the samples of the same category closer. Specifically, we utilize samples from the same category as positive samples and samples from different categories as negative samples.

Temporal Contrasting We employ r_l to denote the temporal representation obtained by the time-domain data x_l via an encoder. Also, the MLP layer's output is defined as $\hat{r}_l = MLP(r_l)$, where x_l is the labeled sample in D^L. Hence, a temporal supervised contrastive loss can be formulated as:

$$\mathcal{L}_{tem}^{\sup}\left(\mathcal{X}_{tem}^L, \mathcal{Y}^L\right) = \sum_{l=1}^{L} \frac{-1}{|P(l)|} \sum_{p \in P(l)} \log \frac{\exp(\hat{r}_l \cdot \hat{r}_p/\tau)}{\sum_{a \in A(l)} \exp(\hat{r}_l \cdot \hat{r}_a/\tau)}, \quad (2)$$

where $\mathcal{X}_{tem}^L$ is the representation of time-domain labeled data through MLP nonlinear transformation, and $\mathcal{Y}^L$ denotes the set of labels for the labeled data. $A(l) = \{1, 2, \ldots, L\}_{l=1}^{L}$ is the set of labeled sample subscripts in D^L, $P(l) = \{p \in A(l) : y_p = y_l\}$ represents the set of positive sample subscripts belonging to the same category, and $\tau \in R^+$ is a temperature coefficient.

Frequency Contrasting We utilize the Fast Fourier Transform to convert the original time series data $\mathcal{X}$ to the frequency domain view and represent it as S. However, S is a set of complex numbers, which can not be directly input into the encoder for training. Therefore, we adopt S to calculate the amplitude $|S|$ and the phase $\emptyset(S)$. Compared with only utilizing amplitude or phase as frequency domain data for classification, employing data $\mathcal{S} = [|S|, \emptyset(S)]$ composed of amplitude and phase can effectively improve classification accuracy. For detailed reasons, we provide ablation studies in the experiments. Further, we adopt q_l to denote the representation obtained by the frequency domain data $\mathcal{S}_l$ via the encoder, and set $\hat{q}_l = MLP(q_l)$. Thus, the frequency supervised contrastive loss is defined as:

$$\mathcal{L}_{feq}^{sup}\left(\mathcal{S}_{feq}^L, \mathcal{Y}^L\right) = \sum_{l=1}^{L} \frac{-1}{|P(l)|} \sum_{p \in P(l)} \log \frac{\exp(\hat{q}_l \cdot \hat{q}_p/\tau)}{\sum_{a \in A(l)} \exp(\hat{q}_l \cdot \hat{q}_a/\tau)}, \quad (3)$$

where $\mathcal{S}_{feq}^L$ denotes the representation of the frequency-domain labeled data by MLP nonlinear transformation.

Temporal-Frequency Co-Training

Co-training is first applied to SSL by (Blum and Mitchell 1998), and indicates that complementary information between different views of the same instance can have a positive impact on model training. Specifically, the authors utilize the prediction results of classifiers on the unlabeled data from different views to generate pseudo-labels. Motivated by the aforementioned work, we leverage label propagation to generate pseudo-labels in the temporal-frequency representation space to fully exploit the inherent properties of time-series data. At the same time, we dynamically select high-quality pseudo-labels considering the learning difficulty of different categories, thus employing complementary information from temporal-frequency representations for time series SSL. Next, we describe the specific process of label propagation and curriculum pseudo-labeling.

Label Propagation Suppose $V = \{r_1, r_2, \ldots, r_M\}_{m=1}^{M}$ denotes the set of representations containing M time-domain view data, and $W \in R^{M \times M}$ represents a symmetric adjacency matrix, where the element W_{ij} is the euclidean distance between samples r_i and r_j. Then, we utilize the normalized graph Laplacians (Chung and Graham 1997) on W, that is, $\mathcal{W} = Q^{-1/2} W Q^{-1/2}$, where $Q = diag(W 1_n)$ is a diagonal matrix. Also, we define a one-hot encoded label matrix $Y \in R^{M \times C}$ and set $Y_{mc} = 1$ if sample r_m belongs to category c, otherwise set $y_{mc} = 0$. We employ the top k values of each row in $\mathcal{W}$ to create the nearest neighbor graph, which is denoted as $\mathcal{W}_{topk}$. Based on Y, label propagation estimates the pseudo-label of each node in the nearest neighbor graph by iteratively solving the following equation:

$$F_{t+1} = \alpha \mathcal{W}_{topk} F_t + (1 - \alpha) Y, \quad (4)$$

where $F_t \in R^{M \times C}$ represents the predicted pseudo-labels of the t-th iteration, and $\alpha \in (0, 1)$ denotes the parameter that controls the propagation information. Naturally, the sequence $\{F_t\}$ has a closed-form solution (Zhou et al. 2003), which can be formulated as:

$$\mathcal{F} = \left(I - \alpha \mathcal{W}_{topk}\right)^{-1} Y, \quad (5)$$

where $\mathcal{F} \in R^{M \times C}$ is the final predicted soft pseudo-labels and I denotes the identity matrix. The time complexity of solving $\mathcal{F}$ by flipping the matrix $\left(I - \alpha \mathcal{W}_{topk}\right)^{-1}$ is $O\left(k \times M^2\right)$, which is inefficient for larger M. To address this issue, we perform label propagation within every mini-batch of the model training. Further, to improve the quality of pseudo-labels acquired within the same mini-batch, we utilize a queue to dynamically store data from up to three consecutive mini-batches for label propagation.

Curriculum Pseudo Labeling Due to the different learning difficulties of categories (Bengio et al. 2009), samples from the same category may remain more scattered in the feature representation space. In other words, it may be difficult for TS-TFC to learn the correct data distribution by directly co-training with all pseudo-labels of unlabeled data , which inevitably include incorrect labels. To this end, we employ a dynamic threshold $\mathcal{T}$ to select high-quality propagation pseudo-labels in $\mathcal{F}$ based on the learning difficulty of categories (Zhang et al. 2021), thus reducing the error rate of the selected pseudo-labels. The updated process of the threshold $\mathcal{T}_e(c)$ is defined as:

$$\mathcal{T}_e(c) = \mathcal{M}\left(\frac{\delta_e(c)}{\max\left(\delta_e\right)}\right) \gamma, \quad (6)$$

where e denotes the number of iterations of the model training, $c \in [0, C)$ is the category, γ denotes a fixed threshold for select pseudo-labels, $\delta_e(c)$ represents the number of selected samples larger than the threshold γ. Also, $\mathcal{M}(x) = x/(2 - x)$ is a nonlinear mapping function that improves the stability of the model training (Zhang et al. 2021). Finally, the set of pseudo-labeled samples selected for co-training only if $\max\left(\mathcal{F}_m\right) > \mathcal{T}_e(c)$, otherwise the pseudo-labeled sample r_m is not used for classification training (please refer to the oval shading in Figure 1).

Overall Loss Function Ultimately, the training losses $\mathcal{L}_{tem}$ and $\mathcal{L}_{feq}$ for the temporal encoder and frequency encoder are defined as follows:

$$\mathcal{L}_{tem}\left(\mathcal{X}, Z^{feq}\right) = \mathcal{L}_{cls}\left(\mathcal{X}, Z^{feq}\right) + \lambda \mathcal{L}_{tem}^{sup}\left(\mathcal{X}_{tem}^L, \mathcal{Y}^L\right), \quad (7)$$

Labeling Ratio	10%			20%			40%		
Method	Avg Acc	Avg Rank	P-value	Avg Acc	Avg Rank	P-value	Avg Acc	Avg Rank	P-value
Supervised	0.738	5.06	1.16E-04	0.798	4.86	4.68E-06	0.839	4.97	1.12E-08
Pseudo-Label	0.745	4.56	1.71E-04	0.800	4.36	6.28E-05	0.840	4.68	7.48E-09
TE	0.749	4.27	4.94E-04	0.803	4.50	1.76E-04	0.842	4.61	7.73E-08
LPDeepSSL	0.729	5.43	6.55E-09	0.769	6.37	1.54E-09	0.823	5.84	8.57E-10
TS-TCC	0.513	9.64	2.05E-27	0.569	9.51	1.58E-24	0.610	9.55	1.57E-26
MTL	0.650	7.81	5.78E-18	0.681	7.98	8.94E-18	0.716	8.19	2.55E-19
SemiTime	0.751	4.30	8.53E-03	0.807	3.84	3.48E-03	0.852	3.82	4.09E-04
TS-T	0.754	3.74	9.01E-03	0.808	3.49	1.12E-04	0.847	3.23	2.17E-05
TS-F	0.694	6.85	1.47E-17	0.756	6.67	1.04E-11	0.818	6.21	2.71E-13
TS-TFC	**0.769**	**2.44**	-	**0.822**	**2.19**	-	**0.867**	**2.00**	-

Table 1: Test classification results compared with baselines on 106 UCR time series datasets. P-value < 0.05 (Demšar 2006) represents that TS-TFC is significantly superior to the baseline.

$$\mathcal{L}_{feq}\left(\mathcal{S}, Z^{tem}\right) = \mathcal{L}_{cls}\left(\mathcal{S}, Z^{tem}\right) + \mu \mathcal{L}_{feq}^{sup}\left(\mathcal{S}_{feq}^{L}, \mathcal{Y}^{L}\right), \quad (8)$$

where Z^{feq} (Z^{tem}) denotes the labels acquired by frequency-domain (time-domain) view, and λ and μ are hyperparameters to adjust loss's weight. For details of TS-TFC training, please refer to Algorithm 1 in the Appendix.

Experiments

We conduct experiments utilizing the UCR time series archive (Dau et al. 2019), which is widely employed for time series classification studies (Ismail Fawaz et al. 2019). The number of test samples in many datasets in the UCR archive is much higher than the training samples. As suggested by (Dau et al. 2019; Wang et al. 2019), we merge the original training and test sets, and then divide the train-validation-test set using a five-fold cross-validation method in the ratio of 60%-20%-20% for evaluation. Following the previous work (Wang et al. 2019; Fan et al. 2021), we randomly select 10%, 20% and 40% of the samples in the training set as labeled data. To obtain stable evaluation results, we limit the average number of samples included in each category to at least 30. Therefore, we utilize 106 datasets from the original 128 UCR datasets for experimental analysis, and Appendix A.1 demonstrate the details of datasets. Adam is used as the optimizer, and the learning rate is 0.001. The maximum batch size is 1024, and the maximum epoch is 1000. The temperature coefficients τ in Eq. 2 and Eq. 3 are set to 50, the hyperparameters α is set to 0.99 and 5. And top k in Eq. 4 for temporal and frequency encoder are set to 40 and 30, respectively. The fixed threshold γ is set to 0.95. The hyperparameters λ and μ are set to 0.05. Further, we employ labeled data for the warm-up training in the first 300 epochs, mitigating the learning bias of the model for unlabeled data. All experiments are repeated five times with five random seeds, and are conducted on Pytoch 1.10 platform with 2 NVIDIA GeForce RTX 3090 GPUs. Our implementation of TS-TFC is available at https://github.com/qianlima-lab/TS-TFC.

Comparison with State-of-the-art Methods

We compare TS-TFC with seven approaches (Supervised, Pseudo-Label (Lee et al. 2013), Temporal Ensembling (TE) (Laine and Aila 2017), LPDeepSSL (Iscen et al. 2019), TS-TCC (Eldele et al. 2021), MTL (Jawed, Grabocka, and Schmidt-Thieme 2020), and SemiTime (Fan et al. 2021)). Among them, Supervised indicates that only labeled samples are employed for SSL. In addition, we construct TS-T (w/o co-training in TS-TFC, only time domain view) and TS-F (w/o co-training in TS-TFC, only frequency domain view) models as comparison methods. For details about baselines, please refer to Appendix A.2. TS-TFC has two classifiers in the time domain and frequency domain views, and we choose the classifier with the best classification performance on the validation set to obtain the test results. The test classification results on 106 UCR time series datasets are shown in Table 1, and please refer to Appendix B.1 for detailed results. The data marked in bold represent the best results. *Avg Acc* denotes the average accuracy on 106 UCR time series datasets, and the value in parentheses is the standard deviation. *Avg Rank* is the average ranking.

As shown in Table 1, TS-TFC achieves the best performance. Meanwhile, TS-T obtains suboptimal classification performance in *Avg Rank* metric, demonstrating the superiority of using feature representations for SSL and the effectiveness of the temporal-frequency co-training mechanism. In addition, we find that the *Avg Rank* of TS-TFC, TS-T, and TS-F increased gradually with increasing labeling ratio, which may be due to the increase in labeling ratio improving the performance of the temporal-frequency contrasting module. Further, we find that LPDeepSSL, TS-TCC, and MTL perform worse than Supervised, which may be attributable to the large bias introduced by their designed strategies on the learning of unlabeled time series data. Among them, TS-TCC delivers strong and weak augmentation on the original time series data. However, the above data augmentation strategy may introduce learning bias to UCR time series datasets (Wen et al. 2021; Iwana and Uchida 2021). Also, we find that TS-TFC can obtain the best accuracy/training time tradeoff on 106 UCR time series datasets (please refer to Appendix B.1).

Ablation Studies

To verify the validity of each component in TS-TFC, we conduct ablation studies using 106 UCR time series datasets with 10% labeling ratio, and the ablation statistical results

Method	Avg Acc	Avg Rank	P-value
	Time Domain View		
TS-TFC	**0.769**	**2.17**	-
TS-T	0.754 (-1.5%)	2.79	9.01E-03
w/o warmup	0.696 (-7.3%)	4.32	2.66E-10
w/o queue	0.744 (-2.5%)	3.56	3.77E-02
w/o contrasting	0.745 (-2.4%)	3.85	4.10E-02
w/o curriculum	0.747 (-2.3%)	3.78	8.57E-04
	Frequency Domain View		
TS-TFC	**0.769**	**1.35**	-
TS-F	0.694 (-7.5%)	2.88	1.47E-17
w/o contrasting	0.685 (-8.4%)	3.98	2.57E-17
w/o curriculum	0.687 (-8.2%)	3.41	1.33E-18
only amplitude	0.682 (-8.7%)	3.69	9.78E-14
only phase	0.581 (-18.8%)	5.52	2.51E-27

Table 2: Ablation results on 106 UCR datasets with 10% labeling ratio.

are shown in Table 2 (containing two different sub-tables). Please refer to Tables 5 and 6 in Appendix B.2 for the detailed results. Among them, 1) w/o warmup removes the warm-up training strategy; 2) w/o queue indicates that only the current mini-batch samples are employed to perform label propagation; 3) w/o contrasting removes the temporal or frequency contrastive learning loss; 4) w/o curriculum removes the curriculum pseudo-labeling strategy and utilizes all pseudo-labels generated by label propagation.

As shown in Table 2, the warmup and queue mechanisms can effectively improve the classification performance of the model. Additionally, we find that the warmup mechanism can also improve the classification performance of Pseudo-Label, Temporal Ensembling, LPDeepSSL, MTL, and SemiTime. Therefore, we utilize a consistent warmup strategy for the above baselines in Table 1. The ablation of the contrasting module shows that expanding the differences in the distribution of samples between categories in the representation space can effectively improve the classification performance of TS-TFC. The ablation of co-training, contrasting, and curriculum results indicate the robustness of TS-TFC to resist the false propagated pseudo-labels. In addition, we find that using a combination of amplitude and phase as frequency domain data improves the performance of TS-F compared with merely using amplitude or phase.

Hyperparameter Analysis

We mainly analyze the contrastive loss weights λ and μ, the temperature coefficient τ, and the hyperparameter top k used for label propagation. Since the training time to adjust the parameters using 106 UCR time series datasets is very long, we select 18 UCR time series datasets for hyperparameter analysis based on the number of samples, sequence length, number of categories, and application domains of the datasets. The influence of hyperparameters analysis is shown in Figure 2. When performing a hyperparameter analysis, we fix other hyperparameter values. For the selected dataset information, fixed hyperparameter values setting, and detailed classification results, please refer to Appendix B.3.

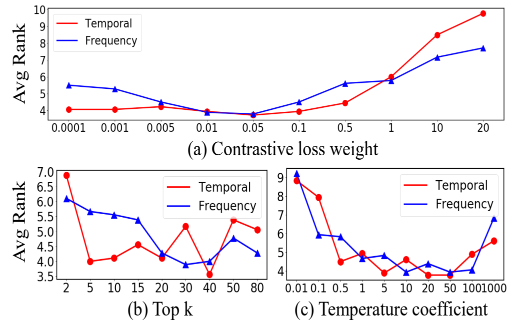


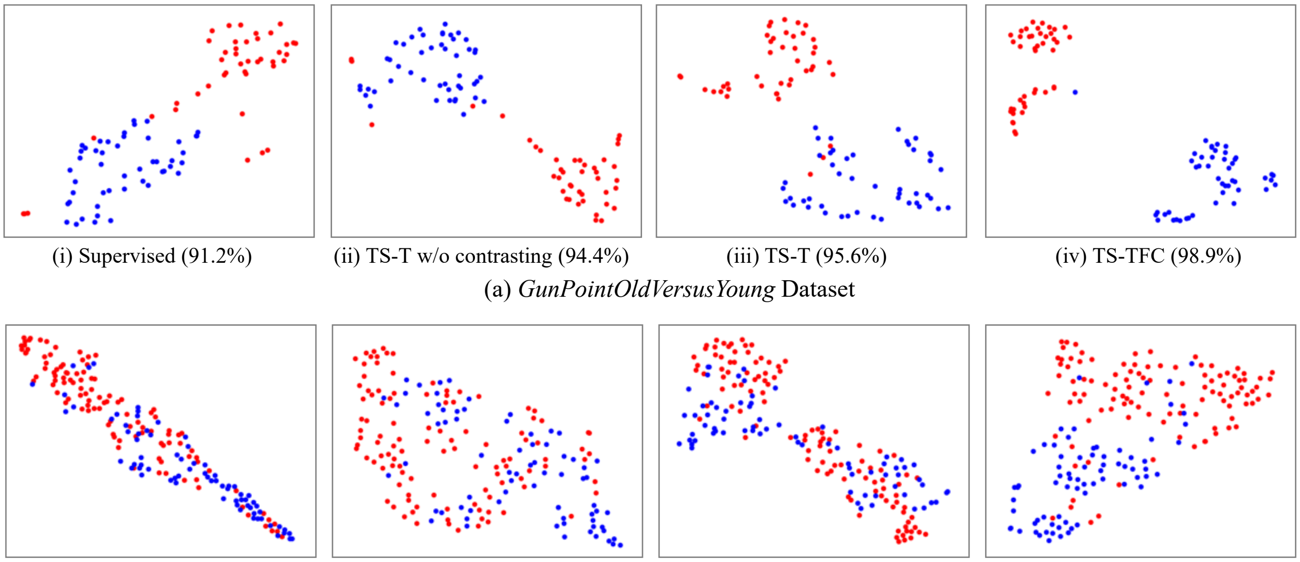
Figure 2: The influence of hyperparameters analysis.

As shown in Figure 2, we find that *Avg Rank* is the lowest when the contrastive learning loss weights λ, μ are taken as 0.05. About hyperparameter top k, the temporal and frequency encoders reach their lowest *Avg Rank* when top k is set to 40 and 30, respectively. In terms of temperature coefficients τ, the lowest *Avg Rank* values are reached at values of 50 for both temporal and frequency encoders.

Visualization Analysis

TS-TFC utilizes label propagation in the representation space to create nearest-neighbor graphs for labeled and unlabeled data. Yet, the distribution of representations has a significant impact on the quality of the propagated pseudo-labels on unlabeled data. To explore the distribution of representations, we employ t-SNE (Van der Maaten and Hinton 2008) for dimensionality reduction visualization. We select the *GunPointOldVersusYong* and *SemgHandGenderCh2* datasets for visualization, which perform well in semi-supervised classification. Specifically, we train the encoder using the first fold training set in the five-fold cross-validation. Then, we employ this encoder to obtain representations on the corresponding test set for visualization, as shown in Figure 3. Concretely, Figure 3 (a) (ii,iii) denotes the results of the representation visualization obtained from the time domain view, while Figure 3 (b) (ii,iii) indicates the results of the representation visualization obtained from the frequency domain view, where the values in parentheses indicate the test classification accuracy. As shown in Figure 3, the temporal-frequency contrasting module improves the discriminative of the representations between categories. Meanwhile, the representations obtained by TS-TFC are more discriminative than TS-T and TS-F. This further indicates that co-training can facilitate TS-TFC to capture complementary information from two distinct views, making the learned representations more discriminative.

Class activation map (CAM) (Zhou et al. 2016) is a way to visualize CNNs and analyze which regions of the input data the CNN-based model focuses on. Hence, we utilize CAM to visualize the region of interest of the original sequence by temporal and frequency encoders in TS-TFC. We choose *Gunpoint*, *ToeSegmentation1*, and *MidddlePhalanxOutlineCorrect* datasets for the visualization of temporal, amplitude, and phase sequences, respectively. Specifically, we employ the sample with the smallest variance



(i) Supervised (91.2%) (ii) TS-T w/o contrasting (94.4%) (iii) TS-T (95.6%) (iv) TS-TFC (98.9%)

(a) *GunPointOldVersusYoung* Dataset

(i) Supervised (68.9%) (ii) TS-F w/o contrasting (70.0%) (iii) TS-F (74.5%) (iv) TS-TFC (80.6%)

(b) *SemgHandGenderCh2* Dataset

Figure 3: The t-SNE visualization of the learned representations.

within each category for visualization, as shown in Figure 4. The left side of Figure 4 represents the original two types of sequences from three datasets, and the right means the CAM visualization results. For the temporal sequence (*Gunpoint* dataset), we follow (Lines et al. 2012) to choose the discriminative region. For the amplitude and phase sequences (*ToeSegmentation1* and *MidddlePhalanxOutlineCorrect* datasets), we obtain the discriminative region based on people's observations. Compared with the original temporal, amplitude, and phase sequences, the encoder of TS-TFC can give higher attention to the discriminative subsequence. The above results demonstrate that the encoder can exploit the discriminative region between categories of temporal and frequency information. Compared with SSL's usage of the original time series directly, TS-TFC can obtain high-quality pseudo-labels for SSL using discriminative representations.

Conclusion

This paper proposes a time-series SSL framework via temporal-frequency co-training, leveraging label propagation to generate pseudo-labels from time-domain and frequency-domain views to guide the training of each view's classifiers. Furthermore, we propose a temporal-frequency contrastive learning module for enhancing the discriminability of representations between categories, while combining the learning difficulty of categories to select high-quality pseudo-labels for co-training. Extensive experiments on 106 UCR datasets indicate that TS-TFC achieves state-of-the-art performance. In addition, the ablation study and visualization analysis demonstrate the effectiveness and robustness of each module in the TS-TFC. In the future, we will explore consistent regularization strategies for time series SSL.

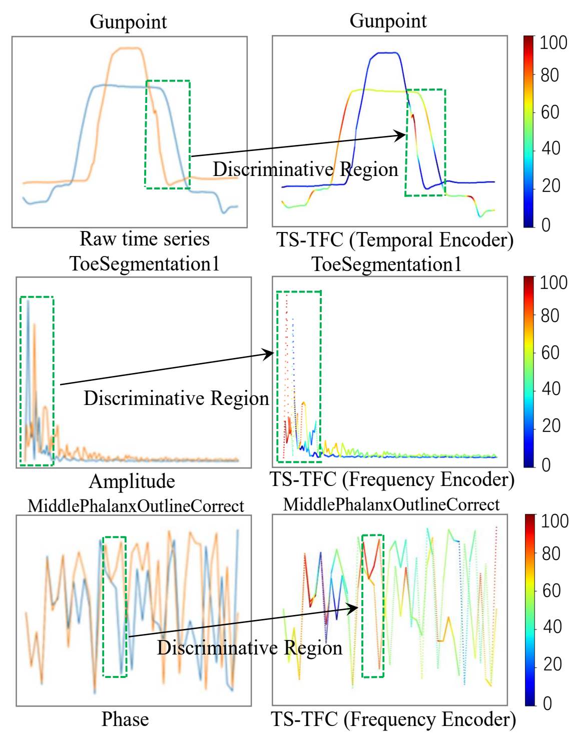


Figure 4: The CAM visualization of the contribution of each time-domain and frequency-domain series region using the FCN encoder. Red represents high contribution and blue indicates almost no contribution in the representation learning.

Acknowledgments

We thank the anonymous reviewers for their helpful feedback. The work described in this paper was partially funded by the National Natural Science Foundation of China (Grant Nos. 62272173, 61872148), the Natural Science Foundation of Guangdong Province (Grant Nos. 2022A1515010179, 2019A1515010768). The authors would like to thank Zhenjing Zheng for the technical discussions.

References

Bengio, Y.; Louradour, J.; Collobert, R.; and Weston, J. 2009. Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning*, 41–48.

Berthelot, D.; Carlini, N.; Goodfellow, I.; Papernot, N.; Oliver, A.; and Raffel, C. A. 2019. Mixmatch: A holistic approach to semi-supervised learning. In *Advances in neural information processing systems*, volume 32.

Blum, A.; and Mitchell, T. 1998. Combining labeled and unlabeled data with co-training. In *Proceedings of the eleventh annual conference on Computational learning theory*, 92–100.

Chen, T.; Kornblith, S.; Norouzi, M.; and Hinton, G. 2020. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, 1597–1607. PMLR.

Chen, Y.; Hu, B.; Keogh, E.; and Batista, G. E. 2013. Dtw-d: time series semi-supervised learning from a single example. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, 383–391.

Chung, F. R.; and Graham, F. C. 1997. *Spectral graph theory*, volume 92. American Mathematical Soc.

Dau, H. A.; Bagnall, A.; Kamgar, K.; Yeh, C.-C. M.; Zhu, Y.; Gharghabi, S.; Ratanamahatana, C. A.; and Keogh, E. 2019. The UCR time series archive. *IEEE/CAA Journal of Automatica Sinica*, 6(6): 1293–1305.

Demšar, J. 2006. Statistical comparisons of classifiers over multiple data sets. *The Journal of Machine learning research*, 7: 1–30.

Eldele, E.; Ragab, M.; Chen, Z.; Wu, M.; Kwoh, C. K.; Li, X.; and Guan, C. 2021. Time-series representation learning via temporal and contextual contrasting. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*.

Fan, H.; Zhang, F.; Wang, R.; Huang, X.; and Li, Z. 2021. Semi-Supervised Time Series Classification by Temporal Relation Prediction. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 3545–3549. IEEE.

Franceschi, J.-Y.; Dieuleveut, A.; and Jaggi, M. 2019. Unsupervised scalable representation learning for multivariate time series. In *Advances in neural information processing systems*, volume 32.

Iscen, A.; Tolias, G.; Avrithis, Y.; and Chum, O. 2019. Label propagation for deep semi-supervised learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5070–5079.

Ismail Fawaz, H.; Forestier, G.; Weber, J.; Idoumghar, L.; and Muller, P.-A. 2019. Deep learning for time series classification: a review. *Data mining and knowledge discovery*, 33(4): 917–963.

Iwana, B. K.; and Uchida, S. 2021. An empirical survey of data augmentation for time series classification with neural networks. *Plos one*, 16(7): e0254841.

Jawed, S.; Grabocka, J.; and Schmidt-Thieme, L. 2020. Self-supervised learning for semi-supervised time series classification. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 499–511. Springer.

Kamnitsas, K.; Castro, D.; Le Folgoc, L.; Walker, I.; Tanno, R.; Rueckert, D.; Glocker, B.; Criminisi, A.; and Nori, A. 2018. Semi-supervised learning via compact latent space clustering. In *International conference on machine learning*, 2459–2468. PMLR.

Khosla, P.; Teterwak, P.; Wang, C.; Sarna, A.; Tian, Y.; Isola, P.; Maschinot, A.; Liu, C.; and Krishnan, D. 2020. Supervised contrastive learning. In *Advances in Neural Information Processing Systems*, volume 33, 18661–18673.

Laine, S.; and Aila, T. 2017. Temporal ensembling for semi-supervised learning. In *International Conference on Learning Representations*.

Lee, D.-H.; et al. 2013. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *Workshop on challenges in representation learning, ICML*, 896–902.

Lines, J.; Davis, L. M.; Hills, J.; and Bagnall, A. 2012. A shapelet transform for time series classification. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, 289–297.

Liu, R.; Meng, G.; Yang, B.; Sun, C.; and Chen, X. 2016. Dislocated time series convolutional neural architecture: An intelligent fault diagnosis approach for electric machine. *IEEE Transactions on Industrial Informatics*, 13(3): 1310–1320.

Liu, Y.; Lee, J.; Park, M.; Kim, S.; Yang, E.; Hwang, S.; and Yang, Y. 2019. Learning to propagate labels: Transductive propagation network for few-shot learning. In *International Conference on Learning Representations*.

Müller, M. 2007. Dynamic time warping. *Information retrieval for music and motion*, 69–84.

Nussbaumer, H. J. 1981. The fast Fourier transform. In *Fast Fourier Transform and Convolution Algorithms*, 80–111. Springer.

Rohrbach, M.; Ebert, S.; and Schiele, B. 2013. Transfer learning in a transductive setting. In *Advances in neural information processing systems*, volume 26.

Sajjadi, M.; Javanmardi, M.; and Tasdizen, T. 2016. Regularization with stochastic transformations and perturbations for deep semi-supervised learning. *Advances in neural information processing systems*, 29.

Song, H.; Rajan, D.; Thiagarajan, J.; and Spanias, A. 2018. Attend and diagnose: Clinical time series analysis using attention models. In *Proceedings of the AAAI conference on artificial intelligence*.

Van der Maaten, L.; and Hinton, G. 2008. Visualizing data using t-SNE. *Journal of machine learning research*, 9(11).

Wang, F.; and Zhang, C. 2007. Label propagation through linear neighborhoods. *IEEE Transactions on Knowledge and Data Engineering*, 20(1): 55–67.

Wang, H.; Zhang, Q.; Wu, J.; Pan, S.; and Chen, Y. 2019. Time series feature learning with labeled and unlabeled data. *Pattern Recognition*, 89: 55–66.

Wang, Z.; Yan, W.; and Oates, T. 2017. Time series classification from scratch with deep neural networks: A strong baseline. In *2017 International joint conference on neural networks (IJCNN)*, 1578–1585. IEEE.

Wei, L.; and Keogh, E. 2006. Semi-supervised time series classification. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, 748–753.

Wen, Q.; Sun, L.; Yang, F.; Song, X.; Gao, J.; Wang, X.; and Xu, H. 2021. Time series data augmentation for deep learning: A survey. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*.

Woo, G.; Liu, C.; Sahoo, D.; Kumar, A.; and Hoi, S. 2022. CoST: Contrastive Learning of Disentangled Seasonal-Trend Representations for Time Series Forecasting. In *International Conference on Learning Representations*.

Xiaojin, Z.; and Zoubin, G. 2002. Learning from labeled and unlabeled data with label propagation. *Tech. Rep., Technical Report CMU-CALD-02–107, Carnegie Mellon University*.

Xu, Z.; and Funaya, K. 2015. Time series analysis with graph-based semi-supervised learning. In *2015 IEEE international conference on data science and advanced analytics (DSAA)*, 1–6. IEEE.

Yang, J.; Nguyen, M. N.; San, P. P.; Li, X. L.; and Krishnaswamy, S. 2015. Deep convolutional neural networks on multichannel time series for human activity recognition. In *Twenty-fourth international joint conference on artificial intelligence*.

Yang, L.; and Hong, S. 2022. Unsupervised Time-Series Representation Learning with Iterative Bilinear Temporal-Spectral Fusion. In *International Conference on Machine Learning*, 25038–25054. PMLR.

Yue, Z.; Wang, Y.; Duan, J.; Yang, T.; Huang, C.; Tong, Y.; and Xu, B. 2022. Ts2vec: Towards universal representation of time series. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 8980–8987.

Zhang, B.; Wang, Y.; Hou, W.; Wu, H.; Wang, J.; Okumura, M.; and Shinozaki, T. 2021. Flexmatch: Boosting semi-supervised learning with curriculum pseudo labeling. In *Advances in Neural Information Processing Systems*, 18408–18419.

Zhang, X.; Gao, Y.; Lin, J.; and Lu, C.-T. 2020. Tapnet: Multivariate time series classification with attentional prototypical network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 6845–6852.

Zhou, B.; Khosla, A.; Lapedriza, A.; Oliva, A.; and Torralba, A. 2016. Learning deep features for discriminative localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2921–2929.

Zhou, D.; Bousquet, O.; Lal, T.; Weston, J.; and Schölkopf, B. 2003. Learning with local and global consistency. *Advances in neural information processing systems*, 16.

Zhou, T.; Ma, Z.; Wen, Q.; Wang, X.; Sun, L.; and Jin, R. 2022. FEDformer: Frequency enhanced decomposed transformer for long-term series forecasting. In *International conference on machine learning*.

Zuo, J.; Zeitouni, K.; and Taher, Y. 2021. SMATE: Semi-Supervised Spatio-Temporal Representation Learning on Multivariate Time Series. In *2021 IEEE International Conference on Data Mining (ICDM)*, 1565–1570. IEEE.