SEMI-SUPERVISED TIME SERIES CLASSIFICATION BY TEMPORAL RELATION PREDICTION

Haoyi Fan[1], *Fengbin Zhang**[1], *Ruidong Wang*[1], *Xunhua Huang*[1], *Zuoyong Li**[2]

[1]School of Computer Science and Technology,
Harbin University of Science and Technology, Harbin 150080, China.
[2]Fujian Provincial Key Laboratory of Information Processing and Intelligent Control,
Minjiang University, Fuzhou 350121, China.
{*isfanhy, zhangfengbin*}*@hrbust.edu.cn*, {*iswangrd, ishuangxh*}*@gmail.com*, *fzulzytdq@126.com*.

ABSTRACT

Semi-supervised learning (SSL) has proven to be a powerful algorithm in different domains by leveraging unlabeled data to mitigate the reliance on the tremendous annotated data. However, few efforts consider the underlying temporal relation structure of unlabeled time series data in the semi-supervised learning paradigm. In this work, we propose a simple and effective method of **Semi**-supervised **Time** series classification architecture (termed as **SemiTime**) by gaining from the structure of unlabeled data in a self-supervised manner. Specifically, for the labeled time series, SemiTime conducts the supervised classification directly under the supervision of the annotated class label. For the unlabeled time series, the segments of past-future pair are sampled from time series, where two segments of pair from the same time series candidate are in positive temporal relation, while two segments from the different candidates are in negative temporal relation. Then, the temporal relation between those segments is predicted by SemiTime in a self-supervised manner. Finally, by jointly classifying labeled data and predicting the temporal relation of unlabeled data, the useful representation of unlabeled time series can be captured by SemiTime. Extensive experiments on multiple real-world datasets show that SemiTime consistently outperforms the state-of-the-arts, which demonstrates the effectiveness of the proposed method. Code and data are publicly available at `https://haoyfan.github.io`.

Index Terms— Time series classification, semi-supervised learning, self-supervised, temporal relation.

1. INTRODUCTION

Time series classification as a fundamental task in machine learning and signal processing, has gained significant attention over past decades with many applications such as intelligent fault diagnosis for electric machine [1], ECG analysis for human healthcare [2], and cyber-security for power systems [3].

Deep supervised learning models have achieved remarkable performance on different time series analysis tasks [4–6]. However, those supervised deep models rely on the availability of large amounts of labeled training data, where labeling time series data is often labor-intensive and time-consuming. Thus, there is a large research effort dedicated to learn with not only limited labeled data but abundant easily accessible unlabeled data from many realistic settings. Within this effort, interest in semi-supervised time series representation learning has recently increased, which shows promising results on time series classification [7–11].

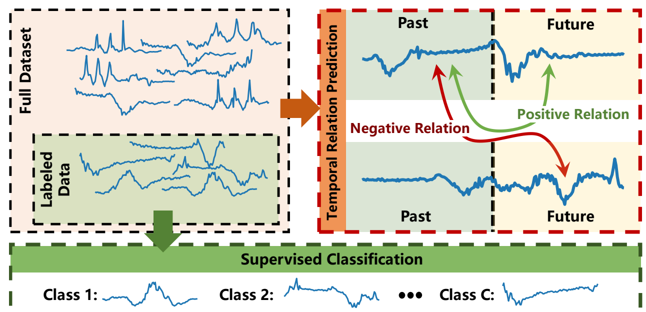

Fig. 1. Schematic illustration of the proposed semi-supervised techniques: Both labeled and unlabeled data are utilized in a semi-supervised paradigm, where labeled data is classified with supervision, and unlabeled data is trained by self-supervised temporal relation prediction of past-future segment pair.

Semi-supervised time series classification aims to combine a small amount of labeled data with a large amount of unlabeled data during training to boost the final classification performance. In [7, 8], Euclidean Distance (ED) and Dynamic Time Warping Distance (DTW-D) based one-nearest-neighbor classifiers are used respectively, to make predictions based on the similarities between labeled and unlabeled time series. Different from those time domain based similarity measure using ED and DTW-D, in [9], the Maximum Diagonal Line of the Cross-Recurrence Quantification Analysis (MDL-CRQA) is applied on the time series phase space for semi-supervised time series classification. Moreover, in [10], both labeled and unlabeled time series data are engaged by employing the least squares regression, the power of the pseudo-labels, shapelet regularization, and spectral analysis of time series. More recently, multi-task learning is introduced in [11] by classifying labeled samples and forecasting future series values of unlabeled samples jointly. Despite the existing methods achieve encouraging results, they ignored the underlying temporal relation structure of time series data, which makes the utilization of unlabeled data under-explored.

In this paper, we argue that the underlying temporal relation of time series data is a significant supervision signal, which can be utilized in the semi-supervised learning to supervise the learning of

*Corresponding authors: Fengbin Zhang, Zuoyong Li. This work is partially supported by National Natural Science Foundation of China (61972187, 61772254, 61172168), Natural Science Foundation of Fujian Province (2020J02024 and 2019J01756).

unlabeled time series data. Consequently, we propose a general semi-supervised time series classification framework, by exploring the semantic feature from unlabeled data in a self-supervised manner. As shown in Figure 1, for the labeled time series, SemiTime conducts supervised classification directly under the supervision of the annotated class label. For the unlabeled time series candidates, past segment and future segment are sampled from each time series to construct past-future segment pairs, where two segments of pair from the same candidate are in positive temporal relation while two segments from the different candidates are in negative temporal relation. Then, the temporal relation between those segments is predicted by SemiTime in a self-supervised manner. Finally, by jointly classifying labeled data and predicting the temporal relation of unlabeled data, more discriminative feature and useful representation of unlabeled time series can be captured by SemiTime. We conduct extensive experiments on multiple real-world datasets from diverse data sources, experimental results show that SemiTime consistently outperforms the state-of-the-arts, which demonstrates the effectiveness of the proposed method.

In sum, the main contributions of this paper are as follows:

- We propose a general semi-supervised time series classification framework, by exploring the semantic feature from unlabeled data in a self-supervised manner.
- We design a simple but effective temporal relational segments sampling strategy, and based on the sampled relational segments, the useful semantic feature can be extracted from the unlabeled time series data.
- We comprehensively evaluate the effectiveness of SemiTime on multiple real-world datasets, and the results demonstrate that our proposed method outperforms the state-of-the-arts.

2. RELATED WORK

Semi-supervised Time Series Classification. In the last decades, semi-supervised learning has attracted lots of attention in different domains seeking to learn from both unlabeled and labeled data [12, 13]. Recently, different semi-supervised time series classification methods have been proposed by learning the underlying structure of the unlabeled time series. Distance based one-nearest-neighbor classifier [7, 8] are used to make predictions, those methods use euclidean distance or dynamic time warping distance to measure the similarities between labeled and unlabeled time series. Maximum Diagonal Line of the Cross-Recurrence Quantification Analysis (MDL-CRQA) is applied to time series phase space in [9] for semi-supervised time series classification. In [10], both labeled and unlabeled time series data are engaged by employing the least squares regression, the power of the pseudo-labels, shapelet regularization, and spectral analysis of time series. In [11], a multi-task learning framework is introduced by classifying labeled samples and forecasting the unlabeled samples jointly for semi-supervised learning.

Self-supervised Learning. Self-supervised learning aims to extract the underlying useful representation of unlabeled data by designing effective pretext tasks. Recently, self-supervised techniques have a broad range of applications in different domains such as computer vision [14–18], and audio/speech processing [19–22]. For visual data, various pretext tasks are designed including solving jigsaw puzzles [14], rotation prediction [15] and visual contrastive learning [16] for image, and frame order validation [17] and pace prediction [18] for video. For audio/speech data, different self-supervised techniques include contrastive learning on audio/speech data [19, 20], and multi-task learning from raw audio by predicting a number of handcrafted features such as MFCCs, prosody, and waveform [21, 22].

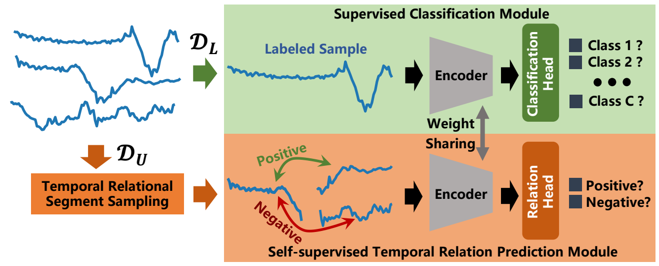

Fig. 2. Architecture of SemiTime.

3. METHOD

In this section, we present our proposed SemiTime in detail. As shown in Figure 2, SemiTime consists of three modules including temporal relational segment sampling module, supervised classification module, and self-supervised temporal relation prediction module. The input of SemiTime is a training set of labeled input-target pairs $(\boldsymbol{t}_i, y_i) \in \mathcal{D}_L$ and unlabeled inputs $\boldsymbol{t}_i \in \mathcal{D}_U$, where $\mathcal{D}_U = \{\boldsymbol{t}_i | \boldsymbol{t}_i = (t_{(i,1)}, ...t_{(i,T)})\}_{i=1}^{N}$ is a set of T-length time series, and $\mathcal{D}_L$ is subset of $\mathcal{D}_U$. In the supervised classification module, the labeled data $\mathcal{D}_L$ are fed to a backbone encoder f_θ following a classification head h_μ for supervised classification task. For unlabeled data $\mathcal{D}_U$, their temporal segments are sampled from temporal relational segment sampling module, and then feed to self-supervised temporal relation prediction module, which consists of a shared backbone encoder f_θ for feature extraction, and a relation head h_φ for segment relation prediction task.

3.1. Training on labeled data

Given the labeled input-target pairs $(\boldsymbol{t}_i, y_i) \in \mathcal{D}_L$, a backbone encoder f_θ takes the time series as input to extract the feature embedding $\boldsymbol{z}_i = f_\theta(\boldsymbol{t}_i)$, and then the following classification head h_μ makes the final classification output $p_i = h_\mu(\boldsymbol{z}_i)$. The supervised training loss is defined as a cross-entropy loss:

$$\mathcal{L}_{cls} = -\frac{1}{|\mathcal{D}_L|} \sum_{i=1}^{|\mathcal{D}_L|} y_i \cdot \log(p_i) \tag{1}$$

3.2. Training on unlabeled data

To explore unlabeled data during feature learning, we use self-generated temporal segment relation as supervisory signal and conduct temporal relation prediction task on unlabeled time series. Formally, given unlabeled inputs $\boldsymbol{t}_i \in \mathcal{D}_U$, each time series $\boldsymbol{t}_i$ is split into two parts where the front B-length part of $\boldsymbol{t}_i$ denotes past segment $\boldsymbol{s}_{i,\alpha}$ and the rear $(T - B)$-length part denotes future segment $\boldsymbol{s}^+_{i,\alpha}$, where $B = \lfloor \alpha * T \rfloor$ and α is a past-future segment split ratio. To sample temporal relation among those segments, given an anchor segment $\boldsymbol{s}_{i,\alpha}$, we select its future counterpart $\boldsymbol{s}^+_{i,\alpha}$ from the same time series $\boldsymbol{t}_i$ as its positive sample, and select another future segment $\boldsymbol{s}^-_{j,\alpha}$ from a different time series $\boldsymbol{t}_j$ as its negative sample. Based on the sampled temporal segment relation, the shared backbone encoder f_θ takes the sampled anchor segment $\boldsymbol{s}_{i,\alpha}$, positive segment $\boldsymbol{s}^+_{i,\alpha}$ and negative segment $\boldsymbol{s}^-_{i,\alpha}$ as inputs, to extract the feature embedding $\boldsymbol{z}_{i,\alpha} = f_\theta(\boldsymbol{s}_{i,\alpha})$, $\boldsymbol{z}^+_{i,\alpha} = f_\theta(\boldsymbol{s}^+_{i,\alpha})$, $\boldsymbol{z}^-_{j,\alpha} = f_\theta(\boldsymbol{s}^-_{i,\alpha})$, and then the relation head h_φ conducts temporal relation prediction between segments, where $p_{2i-1} = h_\varphi([\boldsymbol{z}_{i,\alpha}, \boldsymbol{z}^+_{i,\alpha}])$ for positive relation prediction, and $p_{2i} = h_\varphi([\boldsymbol{z}_{i,\alpha}, \boldsymbol{z}^-_{i,\alpha}])$ for negative relation prediction. Here, $[\bullet, \bullet]$ is concatenation operation. The self-supervised relation prediction training loss is defined as a binary cross-entropy loss:

$$\mathcal{L}_{rel} = -\frac{1}{2|\mathcal{D}_U|} \sum_{i=1}^{2|\mathcal{D}_U|} \tilde{y}_i \cdot \log(p_i) + (1 - \tilde{y}_i) \cdot (1 - \log(p_i)) \quad (2)$$

where $\tilde{y}_i = 1$ denotes positive relation and $\tilde{y}_i = 0$ negative relation.

The mini-batch training algorithm of SemiTime is provided in Algorithm 1.

Algorithm 1 SemiTime Mini-batch Training.

Require:
Labeled input pair $(\boldsymbol{t}_i, y_i) \in \mathcal{D}_L$, and unlabeled input $\boldsymbol{t}_i \in \mathcal{D}_U$. Encoder backbone f_θ; Classification head h_μ; Relation head h_φ; Learning rate η.

for each epoch **do**
  for each labeled minibatch B_L **do**
    $\boldsymbol{z}_i = f_\theta(\boldsymbol{t}_{i \in B_L})$ ▷ Embedding of labeled inputs.
    $p_i = h_\mu(\boldsymbol{z}_{i \in B_L})$ ▷ Label classification.
    $\mathcal{L}_{cls} = -\frac{1}{|B_L|} \sum_{i=1}^{|B_L|} y_i \cdot \log(p_i)$ ▷ Cross-entropy loss.
    $\theta = \theta - \eta \nabla_\theta \mathcal{L}_{cls}$, $\mu = \mu - \eta \nabla_\mu \mathcal{L}_{cls}$ ▷ Update models.
  end for
  for each unlabeled minibatch B_U **do**
    $\boldsymbol{z}_{i,\alpha} = f_\theta(\boldsymbol{s}_{i \in B_U, \alpha})$ ▷ Embedding of anchor segments.
    $\boldsymbol{z}^+_{i,\alpha} = f_\theta(\boldsymbol{s}^+_{i \in B_U, \alpha})$ ▷ Embedding of positive segments.
    $\boldsymbol{z}^-_{j,\alpha} = f_\theta(\boldsymbol{s}^-_{j \in B_U, \alpha})$ ▷ Embedding of negative segments.
    $p_{2i-1} = h_\varphi([\boldsymbol{z}_{i,\alpha}, \boldsymbol{z}^+_{i,\alpha}])$ ▷ Positive relation prediction.
    $p_{2i} = h_\varphi([\boldsymbol{z}_{i,\alpha}, \boldsymbol{z}^-_{i,\alpha}])$ ▷ Negative relation prediction.
    $\mathcal{L}_{rel} = -\frac{1}{2|B_U|} \sum_{i=1}^{2|B_U|} \tilde{y}_i \cdot \log(p_i) + (1 - \tilde{y}_i) \cdot (1 - \log(p_i))$ ▷ Binary Cross-entropy loss.
    $\theta = \theta - \eta \nabla_\theta \mathcal{L}_{rel}$, $\varphi = \varphi - \eta \nabla_\varphi \mathcal{L}_{rel}$ ▷ Update models.
  end for
end for
return Encoder backbone f_θ and classification head h_μ

4. EXPERIMENTS

In this section, we will describe the experimental setups and then analyze the experimental results.

4.1. Experimental Setup

Datasets. To evaluate the effectiveness of the proposed method, in the experiment, we use different categories of time series including three public datasets CricketX, UWaveGestureLibraryAll, and InsectWingbeatSound from the *UCR Time Series Archive*[1], along with two real-world bearing datasets XJTU[2] and MFPT[3] [23], and a EEG dataset EpilepticSeizure[4] [24]. All six datasets consist of various numbers of instances, signal lengths, and number of classes. In the experiment, we set train-validation-test split as 60%-20%-20%. The statistics of six datasets are shown in Table 1.

Table 1. Statistics of Datasets.

Dataset	Sample	Length	Class
CricketX	780	300	12
XJTU	1920	1024	15
InsectWingbeatSound	2200	256	11
MFPT	2574	1024	15
UWaveGestureLibraryAll	4478	945	8
EpilepticSeizure	11500	178	5

Baselines. We compare SemiTime against several state-of-the-art semi-supervised baselines: (1) *Supervised*: a fully supervised baseline using the same encoder backbone and linear classifier as SemiTime, which is trained only on labeled data. (2) *Pseudo-Label* [12]: is an SSL method that uses the pseudo label generated from unlabeled data to enlarge training set for supervised training. (3) Π-*Model* [13]: is an SSL model that uses self-ensembling to form a consensus prediction of the unknown labels under different regularization and input augmentation conditions. (4) *MTL* [11] is an SSL model that leverages features learned from the self-supervised time series forecasting task on unlabeled data.

Implementation. All experiments were performed using PyTorch (v1.4.0). A simple 4-layer 1D convolutional neural network with ReLU activation and batch normalization was used as the backbone encoder f_θ for SemiTime and all other baselines, and use a linear layer as classification head h_μ and a two-layer fully-connected networks with 256 hidden neurons as relation head h_φ respectively. Adam optimizer was used with a learning rate of 0.01. We train all models 1000 epochs with an early-stopping callback of 200 patience epochs to monitor the validation metric and stop the training when no improvement is observed. The batch size is set as 128. During training, we use data augmentations (magnitude warping and time warping [25]) for all models. Classification accuracy is used as the evaluation metric.

4.2. Ablation Study

Firstly, we evaluate the effectiveness of the proposed temporal relation prediction module by investigating the impact of different past-future segment split ratios α on time series classification. As shown in Figure 3, where blue bar indicates temporal relation prediction accuracy on training data (Rel. Pred. ACC), and brown line indicates classification accuracy on test data (Class. ACC). With the increase of α, Class. ACC keeps increasing until $\alpha = 0.4$, and we find that small split ratio values ($\alpha < 0.2$) or big split ratio values ($\alpha \geqslant 0.7$) will drop the classification performance. One possible reason behind this is that an imbalanced past-future segment split makes the relation prediction task too difficult for the model to learn useful representation on unlabeled data, and therefore results in worse Rel. Pred. ACC. Overall, by taking the temporal relation prediction as pretext task, SemiTime consistently outperforms the supervised baseline, which demonstrates the effectiveness of the proposed temporal relation prediction module.

[1]https://www.cs.ucr.edu/~eamonn/time_series_data_2018/
[2]https://biaowang.tech/xjtu-sy-bearing-datasets/
[3]https://www.mfpt.org/fault-data-sets/
[4]https://archive.ics.uci.edu/ml/datasets/Epileptic+Seizure+Recognition

Table 2. Test classification accuracy (%, averages of 10 runs) for supervised baseline and semi-supervised learning on different datasets.All methods use the same 4-layer convolutional backbone. Best results are marked in **red** and the second-best in **blue**.

Label Ratio	**10%**	**20%**	**40%**	**100%**	**10%**	**20%**	**40%**	**100%**	**10%**	**20%**	**40%**	**100%**
Dataset	CricketX				XJTU				InsectWingbeatSound			
Supervised	33.62±0.95	38.79±2.08	52.64±2.53	62.98±2.01	69.71±1.96	83.32±1.59	94.03±1.56	97.92±0.61	50.96±1.58	55.95±0.76	61.41±0.96	**66.27±1.30**
Pseudo-Label [12]	38.87±2.26	44.44±2.91	53.39±2.18	-	74.88±2.78	85.19±1.82	93.97±2.79	-	43.16±3.20	48.35±1.81	55.32±2.04	-
Π-Model [13]	38.61±2.29	48.18±2.07	54.73±1.04	-	**75.96±0.52**	85.93±0.91	**95.03±1.34**	-	**51.47±0.36**	56.14±1.32	**62.20±0.53**	-
MTL [11]	**40.94±1.97**	**50.12±1.22**	**55.10±1.12**	**63.58±1.72**	73.22±1.86	**86.64±1.78**	94.02±1.65	**98.15±1.04**	50.45±1.01	**56.43±0.88**	60.90±0.87	64.14±1.08
Ours	**44.88±3.13**	**51.61±0.66**	**58.71±2.78**	**65.66±1.58**	**84.61±1.39**	**93.93±0.49**	**97.79±0.33**	**98.46±0.25**	**54.96±1.61**	**59.01±1.56**	**62.38±0.76**	**66.57±0.67**
Dataset	MFPT				UWaveGestureLibraryAll				EpilepticSeizure			
Supervised	50.88±0.32	57.14±0.54	69.76±0.48	81.63±0.15	75.81±0.84	81.53±0.54	85.81±0.66	89.5±0.68	68.40±0.43	70.77±0.70	73.49±0.60	77.77±1.13
Pseudo-Label [12]	**63.90±2.62**	65.39±1.70	69.60±2.27	-	75.72±1.85	81.66±0.74	**86.45±1.20**	-	68.57±0.50	72.92±0.48	74.60±0.65	-
Π-Model [13]	55.41±0.65	59.68±0.43	70.15±0.88	-	**77.26±0.31**	**82.87±0.64**	86.17±0.91	-	**69.60±0.34**	71.58±0.64	74.54±0.55	-
MTL [11]	56.11±1.25	**66.20±1.18**	**74.25±1.01**	**82.81±1.06**	76.35±0.56	81.77±0.94	86.01±0.68	**89.76±0.96**	68.71±0.94	**73.17±0.81**	**74.77±0.75**	**78.53±0.62**
Ours	**64.16±0.85**	**69.84±0.94**	**76.49±0.54**	**84.33±0.50**	**81.46±0.60**	**84.57±0.49**	**86.91±0.47**	**90.29±0.32**	**74.86±0.42**	**75.54±0.63**	**77.01±0.79**	**79.26±1.20**

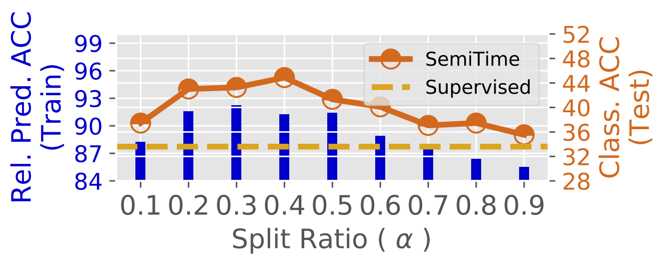

Fig. 3. Impact of different past-future segment split ratios α on CricketX dataset (10% labeled data).

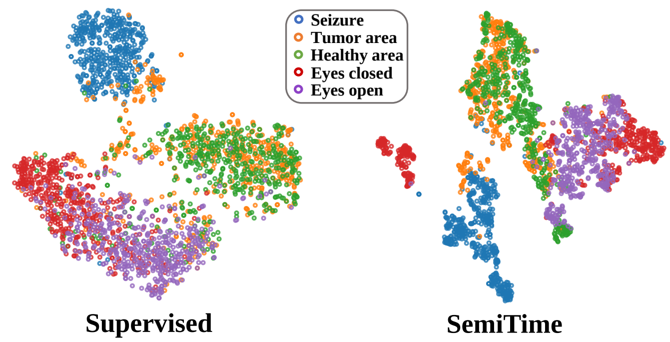

Fig. 4. t-SNE visualization of the learned embedding on EpilepticSeizure dataset. Different colors indicate different labels.

4.3. Time Series Classification

In this section, we evaluate the proposed method by comparing with other semi-supervised state-of-the-arts on time series classification task. Following previous studies [11, 13], we randomly select partial samples (10%, 20%, 40%, and 100%) from the training set as labeled data and use the whole training set as unlabeled data for model training. As shown in Table 2, our proposed SemiTime consistently outperforms all the best baselines across all datasets. For example, given 10% labeled data, SemiTime improves the accuracy over MTL by 9.62% on CricketX, over Π-Model by 11.38% on XJTU and 6.78% on InsectWingbeatSound, respectively. Given 20% labeled data, SemiTime improves the accuracy over MTL by 5.49% on MFPT and over Π-Model by 7.55% on EpilepticSeizure. The results demonstrate that by only generating pseudo-label or using self-ensembling to predict unlabeled data cannot effectively capture the underlying temporal structure of time series, which is important for semi-supervised representation learning of time series data. Moreover, we also evaluate the performance of SemiTime by using 100% training data as both labeled and unlabeled data for supervised and self-supervised training, experimental results show that SemiTime consistently outperforms Supervised baseline and another self-supervised learning based MTL, which demonstrates that forecasting pretext task of MTL cannot effectively capture the useful structure of unlabeled time series, while our designed temporal segment relation prediction is able to capture the underlying intra-temporal relation of unlabeled time series.

4.4. Visualization

To qualitatively evaluate the learned representations, we use the trained backbone to extract the feature embedding and visualize them in 2D space using t-SNE [26] to verify the semantic consistency of the learned representations. Figure 4 shows the visualization results of embedding from the supervised baselines and the proposed SemiTime on EpilepticSeizure dataset. It is obvious that by capturing temporal relation structure and gaining the useful representation from unlabeled data, SemiTime learns more semantic representations and results in better clustering ability for time series data, where more semantic consistency is preserved in the learned representations by our proposed method. Interestingly, we also find that the EEG records from the tumor brain area and the healthy brain area are hard to be discriminated by both models, although SemiTime does better, and this finding provides more insight about data checking and model refinement.

5. CONCLUSION

We propose a general semi-supervised time series classification framework, by exploring the semantic feature from unlabeled data in a self-supervised manner. We design a simple but effective temporal relation sampling strategy, and based on the sampled temporal relation, the useful semantic feature can be extracted from the unlabeled time series data. Our main finding is that the supervisory signal of self-generated temporal segment relation facilitates the better representation of unlabeled time series, and this finding motivates further thinking of how to design better self-supervised pretext task to assist semi-supervised time series classification. Our experiments on multiple real-world datasets show that our proposed method consistently outperforms the state-of-the-arts of semi-supervised models. In the future, we aim to design more effective temporal relation sampling strategy and conduct semi-supervised representation learning on multivariate time series by considering inter-variable relationships.

6. REFERENCES

[1] Myeongsu Kang, Jaeyoung Kim, Linda M Wills, and Jong-Myon Kim, "Time-varying and multiresolution envelope analysis and discriminative feature analysis for bearing fault diagnosis," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 12, pp. 7749–7761, 2015.

[2] Antônio H Ribeiro, Manoel Horta Ribeiro, Gabriela MM Paixão, Derick M Oliveira, Paulo R Gomes, Jéssica A Canazart, Milton PS Ferreira, Carl R Andersson, Peter W Macfarlane, Meira Wagner Jr, et al., "Automatic diagnosis of the 12-lead ecg using a deep neural network," *Nature communications*, vol. 11, no. 1, pp. 1–9, 2020.

[3] Gian Antonio Susto, Angelo Cenedese, and Matteo Terzi, "Time-series classification methods: Review and applications to power systems data," in *Big data application in power systems*, pp. 179–220. Elsevier, 2018.

[4] Alex Graves, Abdel-rahman Mohamed, and Geoffrey Hinton, "Speech recognition with deep recurrent neural networks," in *IEEE international conference on acoustics, speech and signal processing, ICASSP*, 2013, pp. 6645–6649.

[5] Satya Narayan Shukla and Benjamin M Marlin, "Interpolation-prediction networks for irregularly sampled time series," in *7th International Conference on Learning Representations, ICLR*, 2019.

[6] Vincent Fortuin, Matthias Hüser, Francesco Locatello, Heiko Strathmann, and Gunnar Rätsch, "Som-vae: Interpretable discrete representation learning on time series," in *7th International Conference on Learning Representations, ICLR*, 2019.

[7] Li Wei and Eamonn Keogh, "Semi-supervised time series classification," in *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2006, pp. 748–753.

[8] Yanping Chen, Bing Hu, Eamonn Keogh, and Gustavo EAPA Batista, "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*, 2013, pp. 383–391.

[9] Lucas de Carvalho Pagliosa and Rodrigo Fernandes de Mello, "Semi-supervised time series classification on positive and unlabeled problems using cross-recurrence quantification analysis," *Pattern Recognition*, vol. 80, pp. 53–63, 2018.

[10] Haishuai Wang, Qin Zhang, Jia Wu, Shirui Pan, and Yixin Chen, "Time series feature learning with labeled and unlabeled data," *Pattern Recognition*, vol. 89, pp. 55–66, 2019.

[11] Shayan Jawed, Josif Grabocka, and Lars Schmidt-Thieme, "Self-supervised learning for semi-supervised time series classification," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining, PAKDD*, 2020, pp. 499–511.

[12] Dong-Hyun Lee, "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in *Workshop on challenges in representation learning, ICML*, 2013, vol. 3.

[13] Samuli Laine and Timo Aila, "Temporal ensembling for semi-supervised learning," in *5th International Conference on Learning Representations, ICLR*, 2017.

[14] Mehdi Noroozi and Paolo Favaro, "Unsupervised learning of visual representations by solving jigsaw puzzles," in *European Conference on Computer Vision, ECCV*, 2016, pp. 69–84.

[15] Spyros Gidaris, Praveer Singh, and Nikos Komodakis, "Unsupervised representation learning by predicting image rotations," in *6th International Conference on Learning Representations, ICLR*, 2018.

[16] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton, "A simple framework for contrastive learning of visual representations," in *Proceedings of the 37th international conference on machine learning (ICML)*, 2020.

[17] Donglai Wei, Joseph J Lim, Andrew Zisserman, and William T Freeman, "Learning and using the arrow of time," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, 2018, pp. 8052–8060.

[18] Jiangliu Wang, Jianbo Jiao, and Yun-Hui Liu, "Self-supervised video representation learning by pace prediction," in *European conference on computer vision, ECCV*, 2020.

[19] Aaron van den Oord, Yazhe Li, and Oriol Vinyals, "Representation learning with contrastive predictive coding," *arXiv preprint arXiv:1807.03748*, 2018.

[20] Steffen Schneider, Alexei Baevski, Ronan Collobert, and Michael Auli, "wav2vec: Unsupervised pre-training for speech recognition," in *Proc. of the Conf. of the Int. Speech Communication Association (INTERSPEECH)*, 2019, pp. 3465–3469.

[21] Santiago Pascual, Mirco Ravanelli, Joan Serrà, Antonio Bonafonte, and Yoshua Bengio, "Learning problem-agnostic speech representations from multiple self-supervised tasks," in *Proc. of the Conf. of the Int. Speech Communication Association (INTERSPEECH)*, 2019, pp. 161–165.

[22] Mirco Ravanelli, Jianyuan Zhong, Santiago Pascual, Pawel Swietojanski, Joao Monteiro, Jan Trmal, and Yoshua Bengio, "Multi-task self-supervised learning for robust speech recognition," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 6989–6993.

[23] Zhibin Zhao, Tianfu Li, Jingyao Wu, Chuang Sun, Shibin Wang, Ruqiang Yan, and Xuefeng Chen, "Deep learning algorithms for rotating machinery intelligent diagnosis: An open source benchmark study," *arXiv preprint arXiv:2003.03315*, 2020.

[24] Ralph G Andrzejak, Klaus Lehnertz, Florian Mormann, Christoph Rieke, Peter David, and Christian E Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," *Physical Review E*, vol. 64, no. 6, pp. 061907, 2001.

[25] Terry T Um, Franz MJ Pfister, Daniel Pichler, Satoshi Endo, Muriel Lang, Sandra Hirche, Urban Fietzek, and Dana Kulić, "Data augmentation of wearable sensor data for parkinson's disease monitoring using convolutional neural networks," in *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, 2017, pp. 216–220.

[26] Laurens van der Maaten and Geoffrey Hinton, "Visualizing data using t-sne," *Journal of machine learning research*, vol. 9, no. Nov, pp. 2579–2605, 2008.