Symbol-Temporal Consistency Self-supervised Learning for Robust Time Series Classification

Kevin Garcia, Cassandra Garza, Brooklyn Berry, Yifeng Gao
Department of Computer Science, University of Texas Rio Grande Valley

***Abstract*—The surge in the significance of time series in digital health domains necessitates advanced methodologies for extracting meaningful patterns and representations. Self-supervised contrastive learning has emerged as a promising approach for learning directly from raw data. However, time series data in digital health is known to be highly noisy, inherently involves concept drifting, and poses a challenge for training a generalizable deep learning model. In this paper, we specifically focus on data distribution shift caused by different human behaviors and propose a self-supervised learning framework that is aware of the bag-of-symbol representation. The bag-of-symbol representation is known for its insensitivity to data warping, location shifts, and noise existed in time series data, making it potentially pivotal in guiding deep learning to acquire a representation resistant to such data shifting. We demonstrate that the proposed method can achieve significantly better performance where significant data shifting exists.**

I. Introduction

The growing importance of time series data in the digital health domains [10], [12], [16], especially data collected from motion sensors, has fueled the need for advanced methodologies that can extract meaningful patterns. While many representations have been explored [15], [22], Self-Supervised Contrastive Learning(SSL) has emerged as a promising approach [3], [13], [24] because it trains models without human annotation, and is versatile enough to be applied to tasks such as classification and anomaly detection. However, challenges lie in systematic data shifting that is prevalent in time series data mining [5], [20]. For example, in sensor readings from human activities, a male may exhibit different behavior compared to a female due to varying weights, paces, and fitness levels.

Inspired from the bag-of-pattern style representations [19], which is known to be insensitive to data warping, location shifts, and noise within time series data, we aim to help deep learning models acquire a representation resistant to such shifts. Our motivation is demonstrated in Figure 1. Two activity time series [18] representing the same activity in two different people, shown in blue and orange respectively. Visually, these examples seem dissimilar due to data shifting, warping, and extra noise. But converting the samples into symbolic representations (Fig. 1.bottom) reveals their similarity. Driven from this observation, we propose a symbol-temporal consistency based SSL framework to address these challenges. The framework aims to encourage the representation consistency between embeddings generated from the symbolic representations and original temporal representations, forcing the model learn a warping-invariant representation.

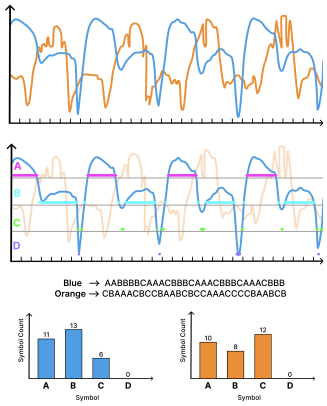

Fig. 1: Sample Transformation to Symbolic Representation

II. Related Work

Recently, there has been a growing trend in designing self-supervised learning frameworks [1], [9] for time series data [4], [6], [21], [23]. Many of these models are based on the unsupervised contrastive learning framework SimCLR [1]. Yue et al. [23] introduced a self-supervised learning framework named ts2vec, which employs time-stamp-wise and instance-wise contrastive co-learning to train the model. Additionally, Jiang et al. focused on enhancing SSL performance by utilizing both time and frequency domains [11]. Zhang et al. [25] introduced a SSL framework named Time-Frequency Consistency (TFC) to perform a time-frequency domain independ self-supervised learning. Furthermore, other works focus on designing augmentation techniques. The existing literature explores various augmentation strategies such as jittering, scaling, masking and time-shifts to augment time-based embeddings [2], [4], [6], [21].

However, there has not been any SSL research that incorporates symbolic information. In the realm of time series data mining, symbolic representation techniques have traditionally been employed for tasks such as motif discovery [7], classification [8], [19], and clustering [17] but they have yet to be utilized in SSL. This paper bridges the gap between symbolic and temporal domains, drawing inspiration from the symbolic representation counting paradigm to enrich contrastive learning [14] in time series data.

III. Methodology

A. Problem Description

Given a pre-training dataset $\mathcal{D}_{pret} = \{x_i \mid i = 1, \ldots, N\}$ which consists of N unlabeled time series samples $x_i \in$

$R^{K\times L}$. Self-supervised learning aims to leverage $\mathcal{D}_{pret}$ to train a model $f : x \rightarrow z$ such that by fine-tuning model parameters on a dataset $D_{tune} = \{x'_i \mid i = 1, \ldots, N'\}$ where $x'_i \in R^{K\times L}$, the fine-tuned model can adapt to various downstream tasks. Specifically, in the task of time series classification, we assume the labels of the data in $\mathcal{D}_{tune}$ are unknown and denoted as $y_i \in \{1, \ldots, C\}$.

B. Overall Framework

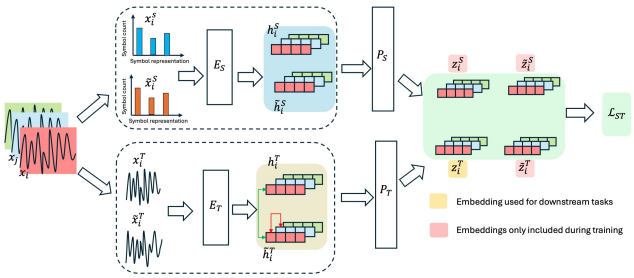

Fig. 2: Overall Model Training Framework

The overall framework of our proposed method is shown in Figure 2 and can be described by five steps: pre-processing, augmentation, encoding, self-contrastive learning, and symbol-temporal consistency learning. Given an input time series data, the framework first performs pre-processing to produce two copies of data representations: the bag-of-symbols representation, and its raw temporal representation. These representations are then augmented using symbol modification and random noise respectively. Next, each symbolic representation is+ passed through a symbolic encoder and each temporal representation is passed through a time encoder to obtain the embedding series representations. These embeddings are then used to perform self-contrastive learning on the symbolic and temporal embeddings independently to enhance robustness. Finally, embeddings from different data representations are aligned via symbol-temporal consistency loss. Through the above steps, we're able to guide the model to learn high-level features robust to both symbolic and temporal perturbations.

C. Bag-of-Symbol Representation x^S:

In addition to its raw temporal representation x^T, x^T is transformed into a symbolic representation x^S to better learn warping and shift invariance. For each element in x^T, we employ a discretization function that transforms real values into symbols. In this paper, the discretization function is designed to produce n symbols, with $n-2$ of those symbols evenly spaced within the range of $\pm 3\sigma$ (standard deviations) of the data with equal-width gap (i.e. $c_i = -3\sigma + \frac{6\sigma}{N-2} i$). By placing the remaining two symbols beyond this range, one above and one below, we account for all possible ranges, allowing for more accurate depictions on patterns in our data.

Specifically, given a time series data $x_{i,t,k}$ where t represents the sample index, k represents k_{th} dimension, and cut lines c_j where j represents the symbol index, the discretization can be represented as follows:

$$s_{i,t,k} = i \quad c_i \geq x_{i,t,k} > c_{i+1} \tag{1}$$

We aggregate symbol counts within the discretized time series, ignoring order to simplify the representation and improve robustness to distribution shifts. The count is computed as:

$$s_{i,j} = \sum_{t=1}^{L} \mathbb{I}\{\text{symbol_series}_{itk} = j\} \tag{2}$$

where s_{ij} denotes the count of symbol i in sample j, and $\mathbb{I}$ is the indicator function. By focusing on symbol frequency rather than temporal order, this approach reduces sensitivity to noise and suits tasks where timing is less critical. Thus, symbolic representations enhance the model's capacity to capture key shifts and warping invariant behavior in time series data.

Lastly, to ensure a smooth training process, the scale of the count can be challenging to feed into deep learning models. Thus, x_i^S is obtained by performing z-score normalization on top of s_i:

$$x_i^S = \frac{s_i - \mathbb{E}\{s_i\}}{\sigma_i^2} \tag{3}$$

where σ_i denotes the standard deviation of s_i.

D. Time-wise and Symbolic Representation Augmentation

Following the classical SSL learning framework [1], [25], x_i^s and x_i^T are augmented to $\tilde{x}_i^S$ and $\tilde{x}_i^T$ respectively. $\tilde{x}_i^s$ is produced through symbolic modifications (insertion or deletion) to x_i^s, and $\tilde{x}_i^T$ is produced by injecting noise into x_i^T.

E. Data Encoding

Next, we describe the proposed model which is trained based on the consistency loss structure introduced by Zhang et al. [25] to encourage the model to project time series with similar temporal or symbolic histogram characteristics into similar embedding representations. Specifically, x_i^T, $\tilde{x}_i^T$ pass through the time encoder to the latent embedding space h_i^T and $\tilde{h}_i^T$, where $h_i^T = E_T(x_i^T)$ and $\tilde{h}_i^T = E_T(\tilde{x}_i^T)$. Similarly, the symbolic component x_i^S and its augmentation sample $\tilde{x}_i^S$ are fed into the symbolic encoder E_S, mapping the samples to symbolic embeddings $h_i^S = E_S(x_i^S)$ and $\tilde{h}_i^S = E_S(\tilde{x}_i^S)$. Next, we project all the embeddings into a unified latent space via the following functions: $z_i^S = P_S(h_i^S)$, $\tilde{z}_i^S = P_S(\tilde{h}_i^S)$, $z_i^T = P_T(h_i^T)$, and $\tilde{z}_i^T = P_T(\tilde{h}_i^T)$.

The creation of this joint time-symbolic space allows our model to unite time series and symbolic representations, allowing for improved generalization across diverse datasets. In the testing phrase, only z_i^T will be treated as the model output to perform fine-tuning tasks.

F. Self-contrastive Learning

Inspired by prior work [25], we propose a consistency contrastive loss with three components: a time-based loss using time-domain embeddings, a symbolic-based loss using bag-of-symbol embeddings, and a consistency loss aligning both. Positive pairs (original and augmented samples) are encouraged to be close in the embedding space, while negative pairs (original and other samples) are pushed apart. This encourages the model to capture underlying patterns from both time and symbolic views. Specifically, the time-based loss is described as:

$$L_{T,i} = -\log \frac{\exp(\text{sim}(h_i^T, \hat{h}_i^T)/\tau)}{\sum_{j \in \mathcal{D}_{pret}} \mathbb{I}_{i \neq j} \exp(\text{sim}(h_i^T, G_T(x_j))/\tau)} \quad (4)$$

where $sim(a, b) = a^T b / \|a\| \|b\|$ is the cosine similarity, $\mathbb{I}_{i \neq j}$ is an indicator function that equals 0 when $i = j$ and 1 otherwise, and τ is a temporal parameter used to adjust the scale.

The symbol-based loss aims to ensure that the symbolic encoder E_S generates embeddings that are invariant to symbolic perturbations by using positive and negative pairs to guide the model in learning symbolic representations.

$$L_{S,i} = -\log \frac{\exp(\text{sim}(h_i^S, \tilde{h}_i^S)/\tau)}{\sum_{j \in \mathcal{D}_{pret}} \mathbb{I}_{i \neq j} \exp(\text{sim}(h_i^S, E_S(x_j))/\tau)} \quad (5)$$

G. Symbol-Temporal Consistency Learning

Lastly, the time-symbolic loss enforces consistency between the time-based z_i^T, $\tilde{z}_i^T$ and symbolic-based embeddings z_i^S, $\tilde{z}_i^S$. We first compute all the across domain contrastive loss:

$$L_{a,b} = -\log \frac{\exp(\text{sim}(a, b)/\tau)}{\sum_{j \in \mathcal{D}_{pret}} \mathbb{I}_{i \neq j} \exp(\text{sim}(a, b))/\tau} \quad (6)$$

and the consistency loss is defined as:

$$L_{TS,i} = \sum_{a \in \{z_i^T, \tilde{z}_i^T\}} \sum_{b \in \{z_i^S, \tilde{z}_i^S\}} (L_{t,s} - L_{a,b} + \delta) \quad (7)$$

where δ denotes the hyper-parameter. Finally, the overall loss function is:

$$L = \sum_{x_i \in \mathcal{D}_{pret}} L_{T,i} + L_{S,i} + \lambda L_{TS,i} \quad (8)$$

where λ allows us to control the importance of $L_{TS,i}$.

Jointly optimizing these losses enables the model to learn embeddings that are robust to noise, adaptable to perturbations, and rich in temporal and symbolic patterns, boosting downstream performance and generalizability.

IV. Experimentation

We evaluate our implementation through comparison experiments conducted on a T4 Nvidia GPU. Using the Adam optimizer (learning rate: $5e-4$) and a ReduceLROnPlateau scheduler, we base our experiments on the PAMAP2 dataset. Our discretization function produces 64 symbols using 62 cutlines. For MLP and Linear Regression testing, we use only the z_t embeddings, as shown in Figure 2, to obtain accuracy.

Dataset characteristics: We use the PAMAP2 dataset [18], a Human Activity Recognition dataset, to evaluate the performance of our model. The PAMAP2 dataset comprises recordings from 9 subjects, capturing various activities. In our experiments, we focus on a classification task specifically designed for classifying standing, walking, and running. To ensure meaningful evaluations, we only included subjects with at least 90 seconds of recorded data in all three activities, resulting in the selection of 5 our of 9 candidates (subjects 1, 2, 5, 6, 8). The experiment is conducted using pairwise classification tasks, where the model is trained on data from candidate i and tested on candidate j. Following the evaluation protocol in [23], after training the model using the SSL framework, we use the embeddings to train a logistic regression model for performance comparison. We compared our proposed work with 3 baselines: **TFC** [25]: Time-Frequency Consistency based SSL, and two models that training from scrap, Multi-layer perceptron (**MLP**) and Logistic Regression (**LR**).

Source	Target	Proposed	TFC	MLP	LR
2	1	**0.909**	0.825	0.662	0.352
5	1	0.740	**0.908**	0.433	0.389
6	1	0.932	0.878	**0.953**	0.636
8	1	**0.910**	0.863	0.819	0.562
1	2	**0.896**	0.864	0.605	0.127
5	2	0.888	**0.918**	0.544	0.137
6	2	**0.915**	0.878	0.137	0.137
8	2	**0.954**	0.934	0.572	0.122
1	5	0.853	**0.922**	0.611	0.399
2	5	**0.920**	0.888	0.654	0.347
6	5	**0.955**	0.883	0.639	0.295
8	5	0.908	**0.935**	0.593	0.336
1	6	**0.922**	0.874	0.576	0.297
2	6	**0.886**	0.855	0.450	0.328
5	6	**0.959**	0.907	0.825	0.639
8	6	**0.938**	0.929	0.827	0.583
1	8	**0.905**	0.825	0.662	0.352
2	8	0.819	**0.908**	0.433	0.389
5	8	0.899	0.878	**0.953**	0.636
6	8	0.929	0.885	**0.930**	0.562
Average:		**0.901**	0.888	0.644	0.381

TABLE I: Baseline comparisons. Source and Target correspond to Subject ID, where the model is trained on the Source data and tested on the Target Data.

Source	Target	z_t	z_t + z_s
2	1	**0.926**	0.853
5	1	**0.948**	0.920
6	1	0.923	**0.955**
8	1	**0.912**	0.908
1	2	0.872	**0.896**
5	2	0.833	**0.888**
6	2	0.42	**0.915**
8	2	0.882	**0.954**
1	5	0.887	**0.922**
2	5	**0.888**	0.886
6	5	**0.966**	0.959
8	5	0.925	**0.938**
1	6	0.808	**0.905**
2	6	0.794	**0.847**
5	6	**0.881**	0.874
8	6	0.839	**0.905**
1	8	0.905	**0.909**
2	8	**0.819**	0.740
5	8	0.899	**0.932**
6	8	**0.929**	0.905
Average		**0.901**	0.863

TABLE II: z_t+z_s combined embedddings vs z_t embeddings only

A. Result

Comparison Result: The results, compared with all baselines, are shown in Table I . The best-performing implementation is bolded in Table I , and the second best is underlined. From the table, our implementation consistently performs well against the TFC baseline method. It is notably superior when the target is subject 2, supporting our claim our implementation is useful for addressing the distribution offset often characterized within genders. The result demonstrates that the proposed

approach can outperform the existing self-supervised learning framework.

Ablation Test: Next, we conduct an ablation test to evaluate the effectiveness of only using z_s in the proposed framework. The results of using a combination of z_s and z_t, the strategy used in the TFC framework, along with our performance, are shown in Table II . Throughout experimentation, we do note that when omitting embeddings z_s, we observe an improved performance when compared to combined z_t and z_s embeddings. This leads us to believe that by omitting Symbolic z_s embeddings during testing, we can achieve better performance.

V. Conclusion

In this paper, we proposed a Self-Supervised Learning framework called Symbol-Time Consistency(STC), which integrates the bag-of-symbol representation and proves robustness against warping, location shifting, and noise within data. We demonstrate that the proposed work can enhance the generalization of deep learning models for time series, specifically in the task of activity recognition. Through extensive experiments, we demonstrate the effectiveness of our framework, particularly in handling distribution shifts arising from characteristics of unseen users.

VI. Acknowledgements

This work is supported by Google CASHI IRP and National Science Foundation (NSF) under Grant IIS-2348480. The computational resource is supported by Google Cloud Platform (GCP) credit.

References

[1] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR, 2020.

[2] Jiaxiang Dong, Haixu Wu, Haoran Zhang, Li Zhang, Jianmin Wang, and Mingsheng Long. Simmtm: A simple pre-training framework for masked time-series modeling. *Advances in Neural Information Processing Systems*, 36:29996–30025, 2023.

[3] Emadeldeen Eldele, Mohamed Ragab, Zhenghua Chen, Min Wu, Chee Keong Kwoh, Xiaoli Li, and Cuntai Guan. Time-series representation learning via temporal and contextual contrasting. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, pages 2352–2359, 2021. Temporal and Contextual Contrasting.

[4] Emadeldeen Eldele, Mohamed Ragab, Zhenghua Chen, Min Wu, Chee Keong Kwoh, Xiaoli Li, and Cuntai Guan. Time-series representation learning via temporal and contextual contrasting. *arXiv preprint arXiv:2106.14112*, 2021.

[5] Hassan Ismail Fawaz, Germain Forestier, Jonathan Weber, Lhassane Idoumghar, and PierreAlain Muller. Deep learning for time series classification: a review. data mining and knowledge discovery.

[6] Jean-Yves Franceschi, Aymeric Dieuleveut, and Martin Jaggi. Unsupervised scalable representation learning for multivariate time series. *Advances in neural information processing systems*, 32, 2019.

[7] Yifeng Gao and Jessica Lin. Efficient discovery of time series motifs with large length range in million scale time series. In *2017 IEEE International Conference on Data Mining (ICDM)*, pages 1213–1222. IEEE, 2017.

[8] Yifeng Gao and Jessica Lin. Hime: discovering variable-length motifs in large-scale time series. *Knowledge and Information Systems*, 61:513–542, 2019.

[9] Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Guo, Mohammad Gheshlaghi Azar, et al. Bootstrap your own latent-a new approach to self-supervised learning. *Advances in neural information processing systems*, 33:21271–21284, 2020.

[10] David Howell, Louis Osternig, and Li-Shan Chou. Monitoring recovery of gait balance control following concussion using an accelerometer. *Journal of biomechanics*, 48(12):3364–3368, 2015.

[11] H. Jiang, D. Liu, L. Shan, and X. Chen. Contrastive learning for time series representation. In *Proceedings of the IEEE Transactions on Knowledge and Data Engineering*, 2020. contrastive learning framework for time series data. Leverages time-based augmentations and explores the temporal dynamics of sequences to learn effective representations.

[12] Dovin Kiernan, David A Hawkins, Martin AC Manoukian, Madeline McKallip, Laura Oelsner, Charles F Caskey, and Crystal L Coolbaugh. Accelerometer-based prediction of running injury in national collegiate athletic association track athletes. *Journal of biomechanics*, 73:201–209, 2018.

[13] Dani Kiyasseh, Tingting Zhu, and David Clifton. Clocs: Contrastive learning of cardiac signals across space, time, and patients. In *Proceedings of the International Conference on Machine Learning*, 2021. contrastive learning, self-supervised pre-training.

[14] Dani Kiyasseh, Tingting Zhu, and David A. Clifton. CLOCS: contrastive learning of cardiac signals. *CoRR*, abs/2005.13249, 2020.

[15] Yucheng Li and Derong Shen. A new symbolic representation method for time series. In *School of Computer Science and Engineering, Northeastern University, Shenyang 110169, China*, 2022. symbolic representation. fractional fourier approximation.

[16] Xinwen Liu, Huan Wang, Zongjin Li, and Lang Qin. Deep learning in ecg diagnosis: A review. *Knowledge-Based Systems*, 227:107187, 2021.

[17] Vasileios Megalooikonomou, Qiang Wang, Guo Li, and Christos Faloutsos. A multiresolution symbolic representation of time series. In *21st International Conference on Data Engineering (ICDE'05)*, pages 668–679. IEEE, 2005.

[18] Attila Reiss and Didier Stricker. Introducing a new benchmarked dataset for activity monitoring. In *2012 16th international symposium on wearable computers*, pages 108–109. IEEE, 2012.

[19] Pavel Senin and Sergey Malinchik. Sax-vsm: Interpretable time series classification using sax and vector space model. In *2013 IEEE 13th international conference on data mining*, pages 1175–1180. IEEE, 2013.

[20] Fan-Keng Sun, Chris Lang, and Duane Boning. Adjusting for autocorrelated errors in neural networks for time series. In *Advances in Neural Information Processing Systems 34*, 2021.

[21] Sana Tonekaboni, Danny Eytan, and Anna Goldenberg. Unsupervised representation learning for time series with temporal neighborhood coding. *arXiv preprint arXiv:2106.00750*, 2021.

[22] Hong Yin, Shu-qiang Yang, Shao-dong Zhu, Shao-Dong Ma, and Lu-min Zhang. Symbolic representation based on trend features for knowledge discovery in long time series. In *Zhejiang University and Springer-Verlag Berlin Heidelberg*, 2015.

[23] Zhihan Yue, Yujing Wang, Juanyong Duan, Tianmeng Yang, Congrui Huang, Yunhai Tong, and Bixiong Xu. Ts2vec: Towards universal representation of time series. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 8980–8987, 2022.

[24] Xiang Zhang, Ziyuan Zhao, Theodoros Tsiligkaridis, and Marinka Zitnik. Self-supervised contrastive pre-training for time series via time-frequency consistency. In *Proceedings of the 36th Conference on Neural Information Processing Systems (NeurIPS 2022)*, 2022. Affiliations: Harvard University, MIT Lincoln Laboratory.

[25] Xiang Zhang, Ziyuan Zhao, Theodoros Tsiligkaridis, and Marinka Zitnik. Self-supervised contrastive pre-training for time series via time-frequency consistency. *Advances in neural information processing systems*, 35:3988–4003, 2022.