CrossTimeNet: Cross-Domain Pre-training with Language Models for Transferable Time Series Representations

Presenter： Xiaoyu Tao

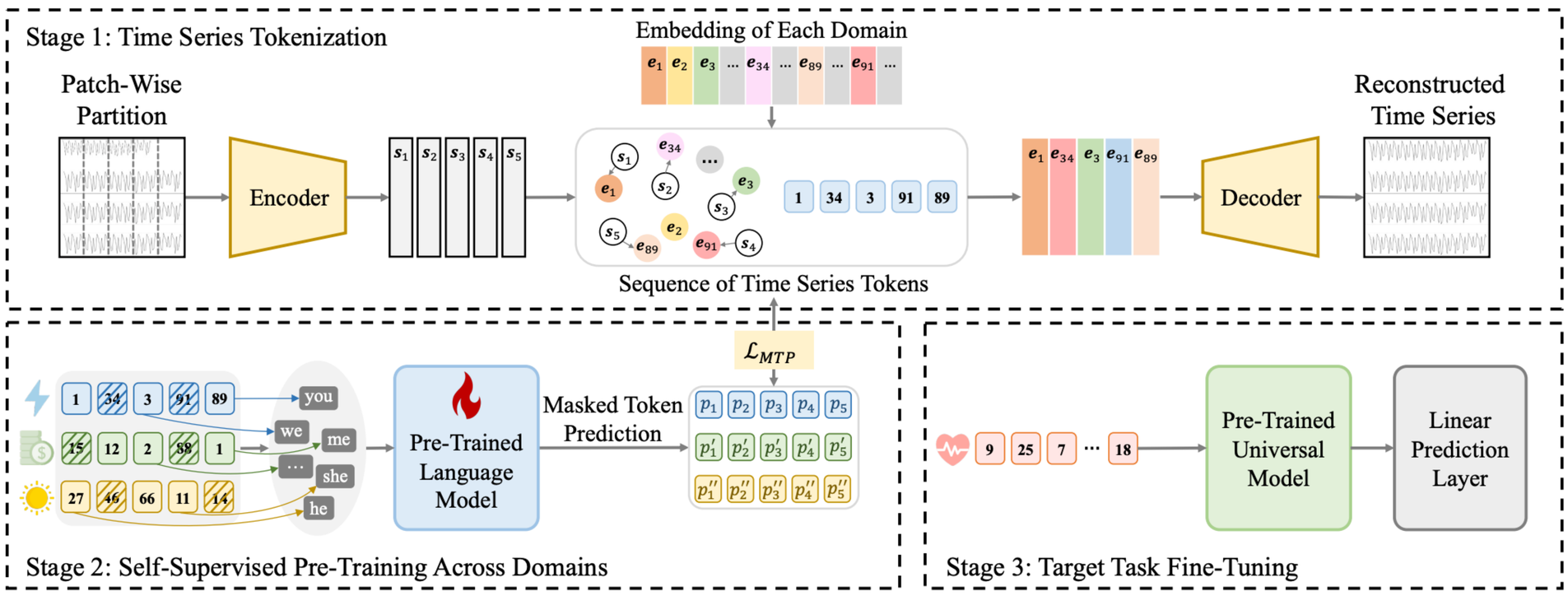
Mingyue Cheng, Xiaoyu Tao, Qi Liu, Hao Zhang, Yiheng Chen, Defu Lian

State Key Laboratory of Cognitive Intelligence, University of Science and Technology of China, Hefei, China

Contribution

1. We introduce **CrossTimeNet**, a cross-domain self-supervised pre-training framework for time series representation learning. Our methodology emphasizes the discretization of time series data, making cross-domain self-supervised pre-training possible.
2. We utilize a **pre-trained language model** as the backbone network, replacing the conventional randomly initialized model. This strategic choice allows us to leverage the extensive knowledge embedded within the language model.
3. We implement a **bidirectional masking optimization strategy** throughout the self-supervised pre-training phase and conduct comprehensive experiments across various classification and prediction tasks. The results indicate that CrossTimeNet significantly enhances the efficacy of time series representation learning, underscoring its potential impact within this field.

Methodology



Framework of our solution

➢Time Series Tokenization

➢Encoding patches

$$\mathbf{z}_l = \text{EncoderTCN}(\mathbf{s}_l)$$

➢Quantizing latent representations

$$\mathbf{z}_l^q = \text{argmin}\ \mathbf{e}_k \in \mathbf{e}|\mathbf{z}_l - \mathbf{e}_k|^2$$

➢Reconstructing the time series

$$\hat{\mathbf{x}} = \text{Decoder}_{\text{TCN}}(\mathbf{z}^q)$$

➢Pre-trained Language Model as Encoders

➢Masked Token Prediction

$$\mathcal{L}_{MTP} = -\sum_{l \in \mathcal{M}} \log p(p_l|\mathbf{x}_{\setminus \mathcal{M}}; \Theta)$$

➢Downstream Task Adaptation

Experiments

◆ Main results in both classification and forecasting tasks.

Table 1: Experimental results of time series classification task evaluated by Accuracy and F1 Score.

Models / Metric	TNC Accuracy	TNC F1 Score	TS-TCC Accuracy	TS-TCC F1 Score	TS2Vec Accuracy	TS2Vec F1 Score	SimMTM Accuracy	SimMTM F1 Score	TST Accuracy	TST F1 Score	TST-Zero Accuracy	TST-Zero F1 Score	TST-Plus Accuracy	TST-Plus F1 Score	CrossTimeNet Accuracy	CrossTimeNet F1 Score
HAR	0.8961	0.8951	0.8832	0.8815	0.8968	0.8957	0.9200	0.9220	0.9203	0.9203	0.9121	0.9120	0.8550	0.8520	**0.9335**	**0.9347**
EEG	0.7603	0.4457	0.7291	0.4347	0.7565	0.4449	0.8165	0.6123	0.8086	0.5516	0.7938	0.5211	0.7929	0.5426	**0.8541**	**0.6402**
ECG	0.2081	0.3310	0.1178	0.3780	0.1302	0.2064	0.2565	0.3562	0.2206	0.3317	0.1810	0.3861	0.2134	0.3246	**0.4378**	**0.6278**

Table 2: Experimental results of time series forecasting task evaluated by MSE and MAE.

Models / Metric	TNC MSE	TNC MAE	TS-TCC MSE	TS-TCC MAE	TS2Vec MSE	TS2Vec MAE	SimMTM MSE	SimMTM MAE	TST MSE	TST MAE	TST-Zero MSE	TST-Zero MAE	TST-Plus MSE	TST-Plus MAE	CrossTimeNet MSE	CrossTimeNet MAE
ETTh1	0.6401	0.5561	0.5962	0.5375	0.6775	0.5690	0.5991	0.5352	0.5718	0.5244	0.6463	0.5571	0.5778	0.5351	**0.5009**	**0.4944**
ETTh2	0.4087	0.4378	0.4122	0.4403	0.4099	0.4395	0.4537	0.4555	0.4171	0.4446	0.4208	0.4496	0.4169	0.4451	**0.3835**	**0.4233**
ETTm1	0.6618	0.6056	0.6323	0.5947	0.5897	0.5765	0.4935	0.4691	0.6030	0.5774	0.6125	0.5869	0.5507	0.5003	**0.4028**	**0.4171**
ETTm2	0.3123	0.3550	0.3225	0.3668	0.3175	0.3611	0.3827	0.3929	0.3113	0.3561	0.3031	0.3514	0.3146	0.3630	**0.2929**	**0.3474**
Weather	0.3523	0.4065	0.3023	0.3895	0.3502	0.4172	0.3134	0.3223	0.3067	0.3888	0.3184	0.4035	0.3028	0.3346	**0.2794**	**0.3089**
Exchange	0.5970	0.5606	0.6079	0.5644	0.6540	0.5904	**0.5684**	**0.5345**	0.6000	0.5605	0.6249	0.5734	0.6540	0.5908	0.5927	0.5499

- Crosstimenet achieves SOTA performance in both classification and forecasting tasks.

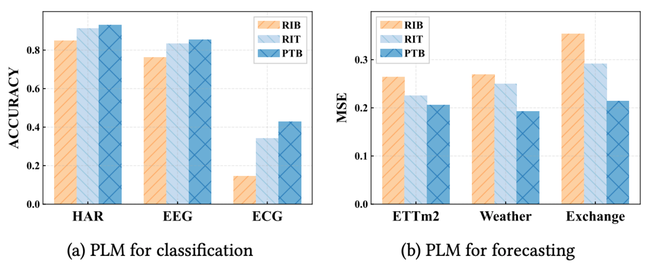
◆ Ablation Study

- **Pre-training Across Domains**

Compared Models	HAR Acc	HAR F1	EEG Acc	EEG F1	ECG Acc	ECG F1
w/o Cross Domain	0.9305	0.9305	0.8541	0.6327	0.4287	0.6161
w/ Cross Domain	**0.9335**	**0.9347**	**0.8541**	**0.6402**	**0.4378**	**0.6278**

Compared Models		ETTh2 MSE	ETTh2 MAE	ETTm1 MSE	ETTm1 MAE	Exchange MSE	Exchange MAE
w/o Cross Doamin	96	0.3706	0.4101	0.3790	0.4105	0.2142	0.3387
	192	0.3982	0.4302	0.4046	0.4269	0.3329	0.4264
	336	0.4065	0.4386	0.4523	0.4495	0.5140	0.5336
	720	0.4400	0.4612	0.5139	0.4799	1.2366	0.8358
w/ Cross Domain	96	**0.3589**	**0.4046**	**0.3430**	**0.3818**	0.2444	0.3674
	192	**0.3755**	**0.4167**	**0.3714**	**0.4046**	**0.3276**	**0.4171**
	336	**0.3731**	**0.4183**	**0.4200**	**0.4180**	0.6046	0.5830
	720	**0.4265**	**0.4537**	**0.4766**	**0.4638**	**1.1940**	**0.8321**

- **Masked-style PLM**

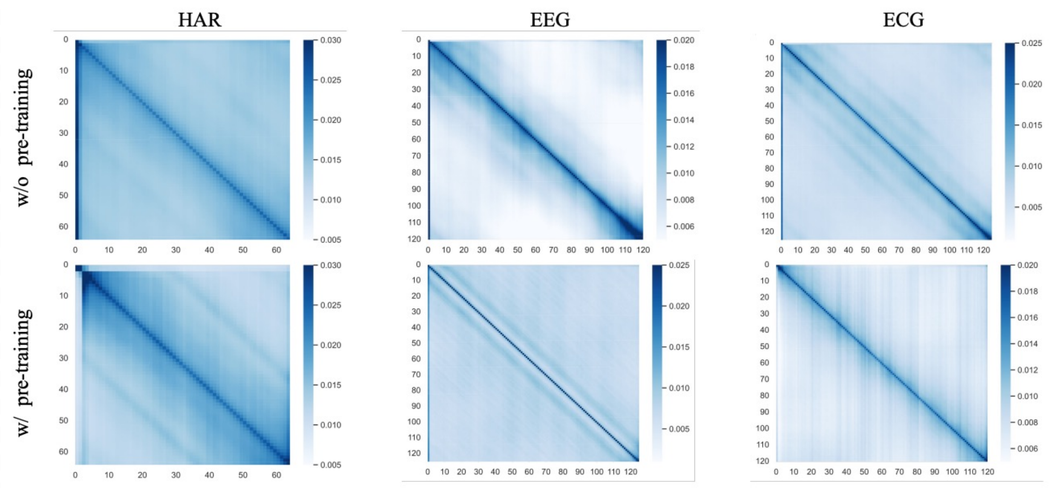


(a) PLM for classification (b) PLM for forecasting

RIB: indicates the use of a pre-trained BERT model as the backbone.
RIT: signifies the use of a randomly initialized BERT model as the backbone.
PTB: denotes the employment of a randomly initialized Transformer model as the backbone.

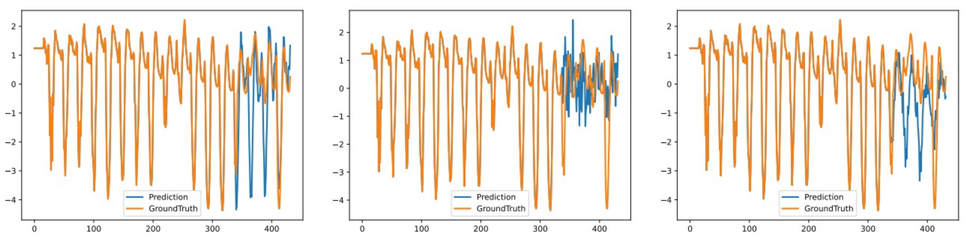
◆ Case Study

- **Attention on Positions**



w/o pre-training: integration of cross-domain information leads to slight improvements in performance.

- **Forecasting Case**



(a) CrossTimeNet (b) TST (c) Fine-Zero

Code https://github.com/Mingyue-Cheng/CrossTimeNet

Contact us mycheng@ustc.edu.cn txytiny@mail.ustc.edu.cn 15656578587