InstructTime: Advancing Time Series Classification with Multimodal Language Modeling

Mingyue Cheng, Yiheng Chen, Qi Liu*, Zhiding Liu, Yucong Luo, Enhong Chen

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

Motivation

- **Limitations in Label Representation**: Traditional time series classification approaches typically use one-hot encoding to represent target categories, which fails to capture the comparability and similarity between different labels.
- **Difficulty in Cross-Domain Model Transfer**: Existing time series classifiers struggle with transfer ability and generalization across different domains, which significantly limits their practicality and universality.
- **Inconsistency in Modal Inputs**: There is a fundamental difference between continuous time series data and the discrete textual data processed by language models, leading to inconsistencies in modal inputs that can adversely affect model performance.

Contribution

- We introduce **InstructTime**,which redefines time series classification by treating it as a learning-to-generate task, utilizing a multimodal language understanding approach to produce labels for each input instance, and achieved the goal of cross-domain data simultaneous training.
- We employed **vector quantization** techniques to map time series to a discrete vocabulary for each domain, while constructing an embedding codebook rich in dense temporal information. This can makes it possible to transfer model parameters among different domains.
- We conducted extensive experiments on several commonly used physiological signal datasets, and the results show that InstructTime bring **great improvements** to the time series classification task, indicating its potential impact for this area. At the same time, we also demonstrated through experiments that this model has strong capabilities in few-shot and zero-shot tasks as well.

Methodology

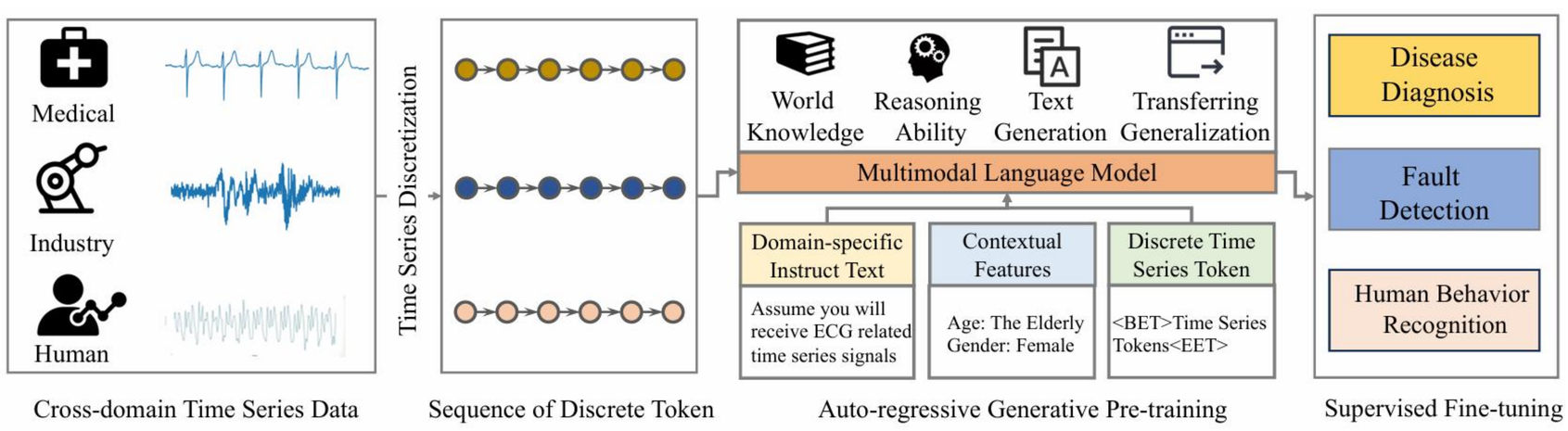

Figure 1: The illustration of the network architecture of the InstructTime.

 # Question Input:
You will receive electrocardiogram(ECG) related signals.
Electrocardiogram signals: <BET><TS-Tokens><EET>.
Age: The Elderly.
Gender: Female.
Describe the potential health issue(s) and associated symptom(s) the patient may be experiencing based on the provided information.

 # Answer Output:
The person's electrocardiogram pattern is indicative of **abnormal ECG**.

Figure2: An input-output example of the InstructTime

- **Time Series Discretization:** Train the ***Vector Quantization*** network based on the TCN architecture for unsupervised reconstruction tasks, and use the codebook in VQ, which contains dense and effective sequence information, as the vocabulary for time series.
- **Representation Alignment Pre-training**: Pretrain the PLM using ***Auto-Regressive*** approach to **fully integrate** time series tokens with text tokens, and learn cross-domain general representations.
- **Supervised Fine-tuning:** Employ ***domain-specific*** supervised generative fine-tuning to tailor the pre-trained base model to the target domain effectively.

Experiments

Compared Models	EEG		ECG		HAR		FD		RWC	
	Accuracy	F1 Score	Accuracy	F1 Score	Accuracy	F1 Score	Accuracy	F1 Score	Accuracy	F1 Score
Transformer	0.7940	0.5178	0.1821	0.3691	0.9148	0.9149	0.9451	0.9564	0.7158	0.7152
Patch Transformer	0.8076	0.5460	0.2465	0.3883	0.8704	0.8683	0.9390	0.9471	0.7552	0.7545
FormerTime	0.8356	0.5828	0.3712	0.5233	0.9199	0.9198	0.9732	0.9853	**0.7803**	**0.7796**
MCDCNN	0.8102	0.5395	0.0929	0.1735	0.8873	0.8862	0.9396	0.9545	0.7762	0.7759
TCN	0.7525	0.3927	0.1014	0.1654	0.9002	0.8997	0.7962	0.7248	0.7113	0.7109
MiniROCKET	0.8318	0.5638	0.2689	0.3900	0.9173	0.9153	0.9412	0.9569	0.7569	0.7556
TimeMAE	0.8248	0.5865	0.2546	0.3834	0.9294	0.9284	0.9878	0.9904	0.7690	0.7664
TS-TCC	0.7291	0.4347	0.1778	0.3780	0.8832	0.8815	0.9296	0.9363	0.6979	0.6931
GPT-As-Classifier	0.7689	0.4929	0.2253	0.3557	0.8973	0.8963	0.9489	0.9598	0.7554	0.7553
InstructTime-Universal	0.8067	0.5007	0.3402	0.4820	0.8990	0.8944	0.9619	0.9656	0.7307	0.7299
InstructTime-Adapt	**0.8452**	**0.6240**	**0.4121**	**0.5547**	**0.9298**	**0.9307**	**0.9901**	**0.9917**	0.7599	0.7578

Table 1: Classification results of all compared methods

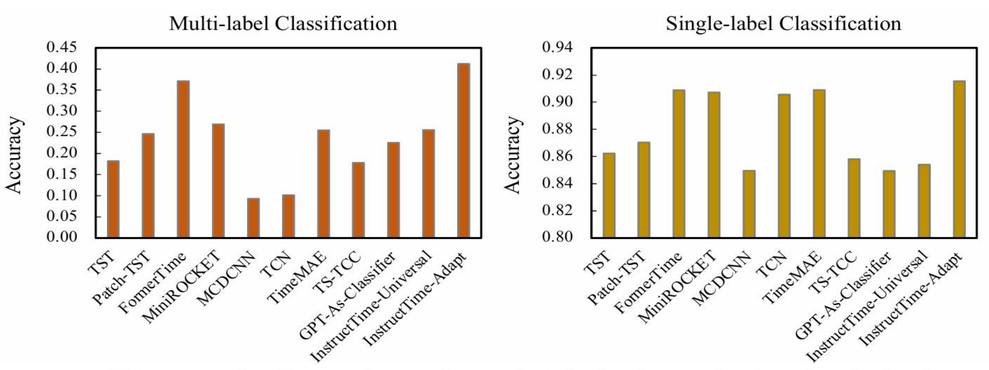

Figure 3: Results of multi-label and single-label ECG classification in terms of all compared methods

Model Variants	EEG		ECG		HAR		FD		RWC	
	Accuracy	F1 Score	Accuracy	F1 Score	Accuracy	F1 Score	Accuracy	F1 Score	Accuracy	F1 Score
w/o Pre-training	0.7854	0.4854	0.2554	0.3751	0.8341	0.8296	0.9092	0.9203	0.7270	0.7268
w/ Pre-training	**0.8452**	**0.6240**	**0.4121**	**0.5547**	**0.9298**	**0.9307**	**0.9901**	**0.9917**	**0.7599**	**0.7578**

Table 2: Performance comparison of InstructTime-Adapt in terms of w/o or w/ auto-regressive pre-training stage

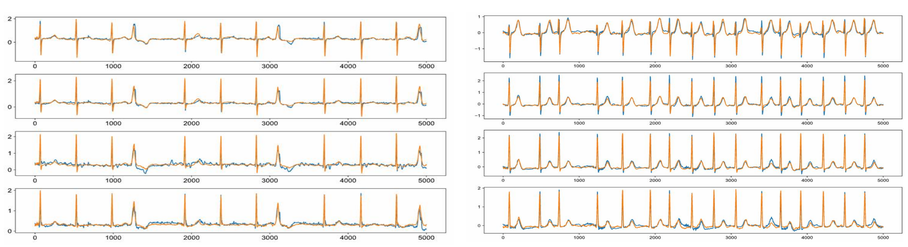

Figure 4: Visualizing the reconstruction of ECG signals in the vector quantized networks.

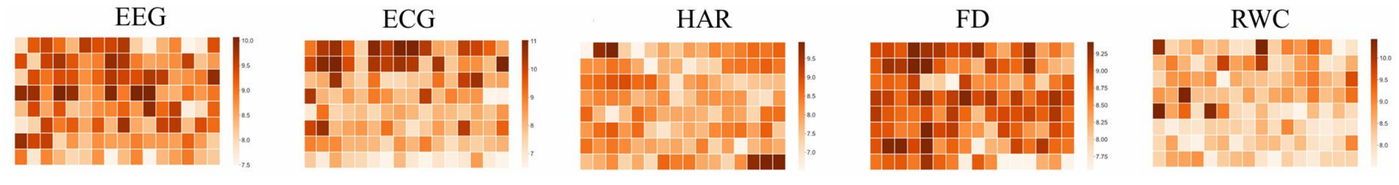

Figure 5:Statics of used token frequency in terms of five datasets, with color intensity indicating the frequency of token appearance.

Code

The code is avaliable at https://github.com/Mingyue-Cheng/InstructTime.
feel free to concact us (mycheng@ustc.edu.cn) if you have any suggestions!

paper: code: WeChat: