Poster: Comparative Study of Transformer Models on a Large Multivariate Time Series HAR Dataset

Hyungtaek Kwon, Zongxing Xie, Mengjing Liu and Fan Ye
Electrical and Computer Engineering
Stony Brook University, NY, USA
{Hyungtaek.Kwon, Zongxing.Xie, Mengjing.Liu, Fan.Ye}@stonybrook.edu

***Abstract*—In Activities of Daily Living (ADL) research, which has gained prominence due to the burgeoning aging population, the challenge of acquiring sufficient ground truth data for model training is a significant bottleneck. This obstacle necessitates a pivot towards unsupervised representation learning methodologies, which do not require many labeled datasets. The existing research focused on the tradeoff between the fully supervised model and the unsupervised pre-trained model and found that the unsupervised version outperformed in most cases. However, their investigation did not use large enough Human Activity Recognition (HAR) datasets, both datasets resulting in 3 dimensions. This poster extends the investigation by employing a large multivariate time series HAR dataset and experimenting with the models with different combinations of critical training parameters such as batch size and learning rate to observe the performance tradeoff. Our findings reveal that the pre-trained model is comparable to the fully supervised classification with a larger multivariate time series HAR dataset. This discovery underscores the potential of unsupervised representation learning in ADL extractions and highlights the importance of model configuration in optimizing performance.**

***Index Terms*—Multivariate Time Series, Transformer, Classification, Pre-training, Activities of Daily Living**

I. Introduction

ADLs have become increasingly prevalent in the medical sector, as it significantly aids healthcare for elders. An issue with this kind of data is label scarcity. Since creating labels for datasets for training machine learning models is costly and time-consuming, recent studies focus on reducing the dependency on such labels. In the existing paper, HAR datasets chosen for the experiments only consisted of 3 dimensions, each containing 1000 samples and 4471 samples, so it is hard to tell whether the models would work on larger dimensional datasets. For this poster, we conducted experiments that leverage unsupervised representation learning with a Transformer model [1] to extract ADL from larger multivariate HAR time series data compared to supervised learning. Orange4home [2] was chosen as the main dataset for the following experiments, which is a dataset that captures routines of daily living in a smart home environment for 4 constant weeks using more than 200 heterogeneous sensors with around 380000 samples and 260 dimensions. As we experimented with existing Transformer models with a dataset larger in volume, we had the chance to observe the performance tradeoff between different HAR datasets and between two types of models. In addition, we are trying different combinations of parameter values, comparing them with the default combination to observe which combinations resulted in the best performance.

II. Related Works

There has been similar research that used a Transformer for extracting ADLs. One research utilized wearable sensor data to compare the performance of Long-Short-Term Memory (LSTM) and the Transformer model [3]. The Transformer model contains two steps. Atomic activities are initially recognized and probability scores in the first step. The Transformer is used in the second step to classify ADLs. Although the paper showed better performances of the Transformer model over the LSTM version, all the cases' accuracy does not exceed 80 percent. Another research focused on detecting behavior change in older adults [4] by using Bi-Directional Encoder Representations from Transformers (BERT)-based and observing behavior anomalies from sequences of ADLs. The proposed model has two steps: first behavior representation where category numbers result as an output, and those numbers are fed to BERT for detecting the behaviors. This paper showed a great extraction from sequences of ADLs but the actual accuracy of the cases does not exceed 90 percent.

III. Method

To preprocess the Orange4home [2] dataset for model training, we first converted the dataset structure from a list table to a cross-table. We filled out the missing values by replacing them with the latest ones appearing on the according feature. Then we employed standardization and one-hot encoding techniques. Standardization was applied to numerical features to ensure they have a mean of 0 and a standard deviation of 1, and one-hot encoding was used for transforming categorical features into a machine-readable format. We also used 90 percent of the full dataset to reduce noisiness and computation time and allocated the first 70% of the dataset for training purposes and the remaining 30% for testing. The dataset is framed into 50 samples to reduce the computation time. Regarding the model, we mostly followed the instructions from the existing Transformer paper by Zerveas et al [1]. Our approach leverages a transformer encoder architecture for multivariate time series data. The pre-training is achieved through an autoregressive task of denoising the input: parts of the input data are masked, and the model is tasked with predicting these masked values. This pre-trained model is then followed by

TABLE I
TABLE OF RESULT METRICS USING DIFFERENT COMBINATIONS OF BATCH SIZES AND LEARNING RATE

Batch Size/Learning Rate	**Accuracy**				**Loss**				**Classification Time (Seconds)**	
	Supervised	***SD***	***Pre-trained***	***SD***	***Supervised***	***SD***	***Pre-trained***	***SD***	***Supervised***	***Pre-trained***
64/1e-3	***0.939***	0.006	0.937	0.01	0.795	0.095	***0.649***	0.026	***2.82***	3.81
128/1e-4	***0.943***	0.005	0.942	0.01	0.471	0.05	***0.431***	0.08	***2.47***	3.67
128/1e-3 (Default)	***0.93***	0.01	0.922	0.02	0.692	0.09	***0.594***	0.12	***2.47***	3.65
128/1e-2	0.923	0.004	***0.928***	0.01	0.689	0.025	***0.589***	0.066	***2.24***	3.24
256/1e-3	***0.934***	0.008	0.931	0.0075	***0.496***	0.03	0.5	0.07	***2.16***	3.15

Python 3.8.18 and NVIDIA GeForce RTX 3090 graphics card utilized for all experiments

supervised fine-tuning to enhance performance across various tasks. The number of epochs is always higher for the fully supervised classification than for the fine-tuning since pre-training is performed before fine-tuning. All the models were executed 3 times for robustness and standard deviations are written as SD and recorded in every column followed by the accuracy and loss. We made changes to two parameters: batch size and learning rate. We used batch sizes bigger and smaller than the default value of 128 and the same for the learning rate with the default value of 1e-3. The combinations of batch size and learning rate were chosen based on both empirical evidence and theoretical considerations. Previous studies in the domain of time series analysis and other areas of deep learning suggest that certain combinations can lead to faster computation and better performance.

IV. RESULTS

Table I shows the performance results for two types of models. Each table row shows the different combinations of batch size and learning rate. A bold number indicates the better of the two between supervised and pre-trained models. The metrics are divided into accuracy and loss ranging from 0 to 1 and average computation time for every epoch.

Based on the results in the table, there are a few things to investigate and notice. First of all, in most cases, the accuracy showed a better number for the fully supervised model than for the pre-trained model. However, the differences were almost negligible. The accuracy difference did not exceed 1 percent. However, the loss metric differed greater between the fully-supervised and pre-trained models. Losses were generally lower for pre-trained models. Although the poster's observations are not the same as the existing paper, the pre-trained model comes close to the supervised model. In addition, better performances were found by trying different combinations of batch size and learning rate. The best combinations for both the supervised and the unsupervised model were the batch size of 128 and the learning rate of 1e-4. The worst combinations were considered from the same batch size but with bigger learning rates. Although a bigger learning rate reduced computation time, this led to the conclusion that smaller learning rates are more appropriate for both models. In regards to batch size, although increasing the batch size did not bring big advantages in accuracy compared to the smaller batch size, loss, and computation time had much better records, leading to the conclusion that using a bigger batch size is more appropriate to achieve better performance with less computation material used. To conclude, using a larger multivariate time series HAR dataset led to similar observations from the existing research and better performances can be achieved by different combinations of batch size and learning rate, but reducing fluctuations and experimenting on different parameters stand as future works.

V. CONCLUSION

In this poster, we executed experiments using a large multivariate time series HAR dataset with various combinations of batch size and learning rate on the Transformer models implemented from the existing paper. In general, the poster showed the unsupervised model's performance was comparable to the supervised model. This tendency occurred in all the other combinations, proving that existing work's observations hold with larger datasets. This shows a potential for future research on extracting ADLs to reduce dependency on data labels with less computation required. Future research can focus on trying out different parameters not mentioned heavily in the poster and reducing fluctuations.

ACKNOWLEDGMENT

This work was supported in part by the National Science Foundation (NSF) under Grant Nos. 2119299 and 1951880. The authors are affiliated with the Mobile Computing and Applications (MoCA) Lab at Stony Brook University.

REFERENCES

[1] Transformers: Zerveas, G., Jayaraman, S., Patel, D., Bhamidipaty, A., & Eickhoff, C. (2021). A transformer-based framework for multivariate time series Representation learning. Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery Data Mining. https://doi.org/10.1145/3447548.3467401.

[2] Julien Cumin, Grégoire Lefebvre, Fano Ramparany, and James L. Crowley. "A Dataset of Routine Daily Activities in an Instrumented Home". In 11th International Conference on Ubiquitous Computing and Ambient Intelligence (UCAmI), 2017.

[3] Gabriela Augustinov, Muhammad Adeel Nisar, Frédéric Li, Amir Tabatabaei, Marcin Grzegorzek, Keywan Sohrabi, and Sebastian Fudickar. 2023. Transformer-Based Recognition of Activities of Daily Living from Wearable Sensor Data. In Proceedings of the 7th International Workshop on Sensor-based Activity Recognition and Artificial Intelligence (iWOAR '22). Association for Computing Machinery, New York, NY, USA, Article 9, 1–8. https://doi.org/10.1145/3558884.3558895.

[4] F. Akbari and K. Sartipi, "A Transformer-based Model for Older Adult Behavior Change Detection," 2022 IEEE 10th International Conference on Healthcare Informatics (ICHI), Rochester, MN, USA, 2022, pp. 27-35, https://doi.org/10.1109/ICHI54592.2022.00016.