

Ensemble Learning of Models for Human Activity Recognition

¹Asahi Miyazaki, ²Kosuke Ukita, ³Tsuyoshi Okita

Department of Artificial Intelligence, Kyushu Institute of Technology, Iizuka, Japan

¹miyazaki.asahi676@mail.kyutech.jp, ²ukita.kosuke299@mail.kyutech.jp, ³tsuyoshi@ai.kyutech.ac.jp

Keywords: Deep learning, Ensemble learning, Activity recognition, Self-supervised learning

Introduction: Each SENVT [1] models have good performance. To improve their performance even more, we built 3 models with different patch sizes and combined them by ensemble learning. In this paper, we used pretrained models as base models for ensemble learning, which is not the typical ensemble learning form.

Approach: We used stacking as the ensemble learning method. The Fig. 1 shows the overview of this method. In stacking, we use some base models which has been already trained with the dataset. The weights of those base models are fixed during ensemble learning. We feed the same dataset to the base models to obtain their predicted results. Those predicted results were concatenated and used as the training data for the last ensemble module. This ensemble module accepts that concatenated results and predicts the label from those results. We used a vision-transformer based model named SENVT-u4[1]. We pretrained SENVT-u4 with the Capture24 dataset by self-supervised learning with patch sizes 2, 4, and 5. A patch size means the smallest length of time which the transformer processes. Fig.2 shows the architecture of SENVT-u4 and the relation between the time window and the patch. Those pretrained models were used as the base models.

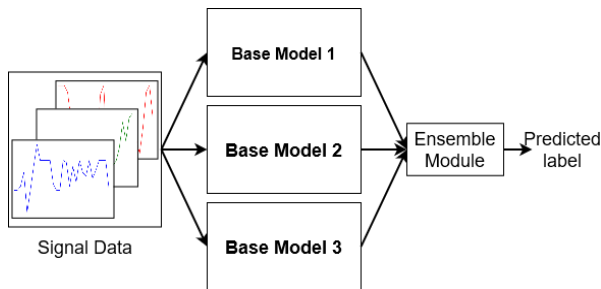


Fig. 1: The overview of the stacking ensemble learning method.

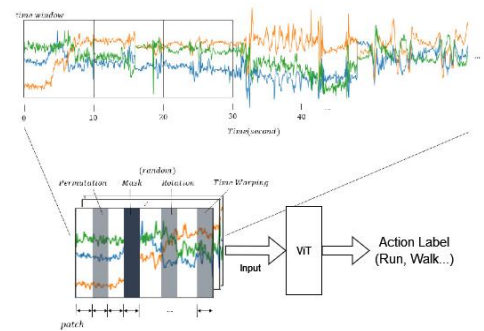


Fig. 2: SENVT-u4 architecture.

Results and Discussion: Table. 1 is the results of the single models for the ADL dataset. The most performant single model is the model with patch size 4. Table. 2 is the results of the meta model built with the base models with patch sizes 2, 4, and 5, respectively. The result of the ADL achieved the better performance than the single model with patch size 4.

Table. 1: The results of the single SENVT-u4 base models.

Patch size	Accuracy	F1-score
2	94.49±1.32	92.81±2.06
4	97.17±0.39	96.33±0.68
5	95.91±0.92	95.16±1.19

Table. 2: The results of the meta model built with the SENVT-u4 base models.

Dataset	Accuracy	F1-score
ADL	97.32±0.80	96.60±1.05
Oppo	82.95±0.82	83.93±1.15
Pamap	89.83±0.81	89.47±0.84
Realworld	93.82±0.18	94.46±0.13
Wisdm	93.22±0.24	93.19±0.24