Exploring Contrastive Learning in Human Activity Recognition for Healthcare

Chi Ian Tang
Ignacio Perez-Pozuelo, Dimitris Spathis, Cecilia Mascolo

UNIVERSITY OF CAMBRIDGE

NEURAL INFORMATION PROCESSING SYSTEMS

Exploring the effectiveness of the SimCLR framework on sensor-based HAR data

Human Activity Recognition (HAR) constitutes one of the most important tasks for wearable and mobile sensing given its implications in human well-being and health monitoring. Motivated by the limitations of labeled datasets in HAR, particularly when employed in healthcare-related applications, this work explores the adoption and adaptation of SimCLR, a contrastive learning technique for visual representations, to HAR.

Highlights

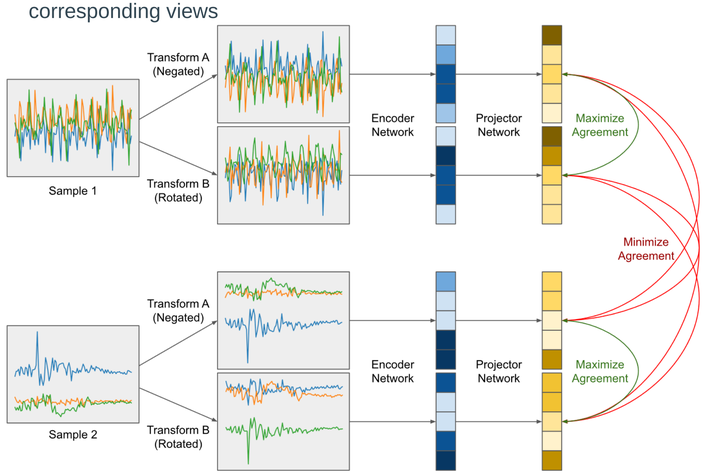
- Adapted the SimCLR framework, which trains the representations of corresponding views to be more similar, and those of non-corresponding views to be more different, to sensor-based HAR
- Extensive evaluation of different combinations of sensor time-series augmentations for contrastive learning
- Strong performance achieved (0.942 in F1), which is higher than pure fully-supervised and self-supervised training approaches
- Indicative of the potential of contrastive learning for healthcare data

Background

The SimCLR contrastive learning framework

SimCLR consists of four main components

- A probabilistic transformation function, which transforms data into different views
- A neural network base encoder
- A projection head
- A contrastive loss function, which maximizes the agreement between corresponding views



An illustration of the SimCLR contrastive learning framework

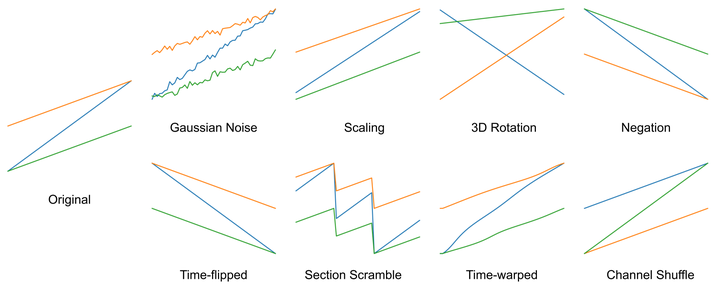
SimCLR for HAR

Adaptations made to enable contrastive learning for HAR

Probabilistic Transformation Function.

8 augmentation functions designed for time-series sensor data, which mimic common noises in sensor data, are chosen to replace the image augmentation operators:

1. Adding random Gaussian noise
2. Scaling by a random factor
3. Applying a random 3D rotation
4. Inverting the signals
5. Reversing the direction of time
6. Randomly scrambling sections of the signal
7. Stretching and warping the time-series
8. Shuffling the different channels



An illustration of the 8 different augmentation functions for time-series sensor data

These functions are used to compose the 64 different augmentation functions by applying different pairs of functions in different orders.

Model.

In this work, we have adopted a relatively lightweight neural network architecture, TPN, which consists of three 1D-convolution layers, as the base encoder. A three-layer fully connected MLP was used as the projection head.

Contrastive Loss

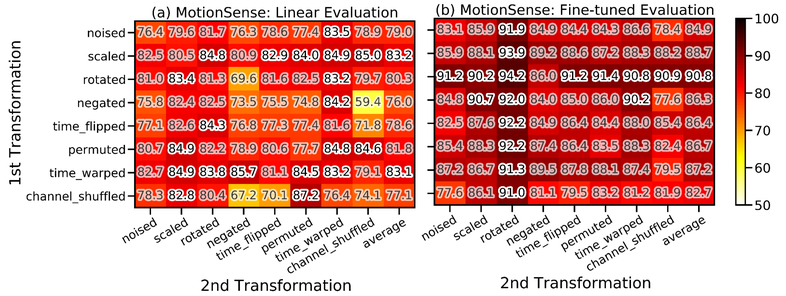
The NT-Xent (normalized temperature-scaled cross entropy loss) was adopted in this work, which trains the model to maximize agreement between positive pairs (corresponding views).

$$L = -\log \frac{\exp(\text{sim}(\boldsymbol{x}, \boldsymbol{x}_+)/\tau)}{\exp(\text{sim}(\boldsymbol{x}, \boldsymbol{x}_+)/\tau) + \sum_{\boldsymbol{x}_- \in \boldsymbol{X}_-} \exp(\text{sim}(\boldsymbol{x}, \boldsymbol{x}_-)/\tau)}$$

The NT-Xent loss function

Evaluation Results

Different Combinations of augmentation functions.



Average weighted F1 scores (in percent) of models trained by different combinations of transformation functions for SimCLR on the MotionSense dataset across 5 independent runs. The diagonal entries correspond to using only a single transformation, and the last column is the average performance of the corresponding rows

Comparison with baseline models

Model	Pure Supervised	Self-Supervised	SimCLR for HAR
Weighted F1	0.922	0.923	**0.942**

Findings

- **The choice of transformation function has a significant impact on performance.** In linear evaluation, the scaled and the time-warped transformations generally performed well when used alongside other transformations. The highest performing models were trained by combining channel shuffling and permutation, with an average F1 score of 0.872. Some combinations perform significantly worse, where the performance difference can be as high as 0.278.

- **Modest performance gain over previously proposed methods indicate its potential of being generalized to other types of data.** Compared to models trained using pure supervised learning and self-supervised learning pipelines, in which the models are pre-trained to identify transformation rather than optimizing the contrastive objective, our adaptation of SimCLR for HAR resulted in a performance gain of up to 0.020 compared to fully supervised models, which is indicative of the potential of transferring SimCLR to mobile sensing settings and other health data, especially due to the modality-agnostic nature of the method.

References

Chen, Ting, et al. "A simple framework for contrastive learning of visual representations." arXiv:2002.05709 (2020).

Um, Terry T., et al. "Data augmentation of wearable sensor data for parkinson's disease monitoring using convolutional neural networks." In: Proceedings of the 19th ACM International Conference on Multimodal Interaction. 2017.

Saeed, Aaqib, Tanir Ozcelebi, and Johan Lukkien. "Multi-task self-supervised learning for human activity detection." In: Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 3.2 (2019): 1-30.