Few-shot Classification for Human Activity Recognition

Chaohao Yuan   Aoyang Liu   Hao Xu

TBSI, Tsinghua University, China

Abstract

- **Apply few-shot learning in sensor-based activity recognition: output a classifier given only a few labeled examples of the unseen activity.**
- **Implement LSTM to capture temporal dynamics, which improves the accuracy of activity recognition.**
- **Introduce prototypical network to HAR few-shot learning scenarios.**

Introduction

Human activity recognition (HAR) is essential to society. In this project, we develop a few-shot learning model which can learn a classifier for unseen activities.

- The advantages of few-shot learning in HAR:
  - Although large scale datasets perform well in training deep learning models, it is a limitation to collect data with intensive human labor.
  - Few-shot learning is closer to human learning, where new concepts can be learned from very few examples.

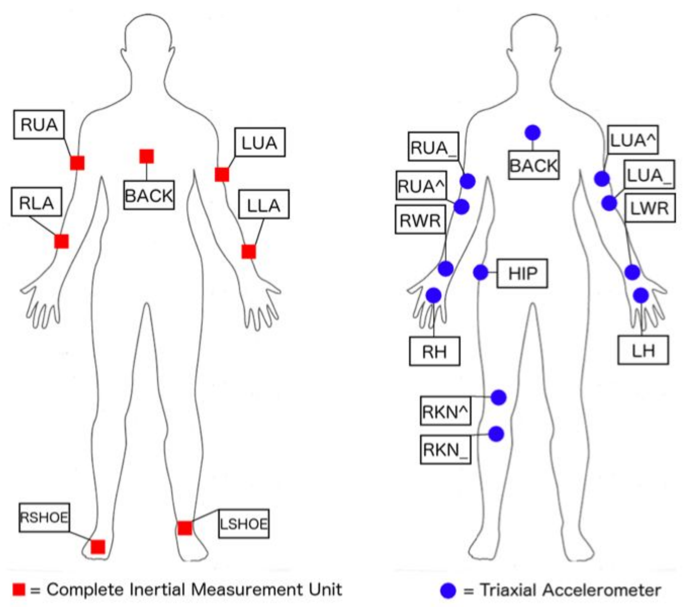

Figure 1. Collection of dataset

- Introduction to the dataset:
  - The dataset consists of digital sequence collected by different sensors attached to the body.
  - Known activities have full training data, while unknown activities have only a small number of information.

Method

To construct our HAR model, we combine LSTM with prototypical network to implement few-shot learning.

1. Overview

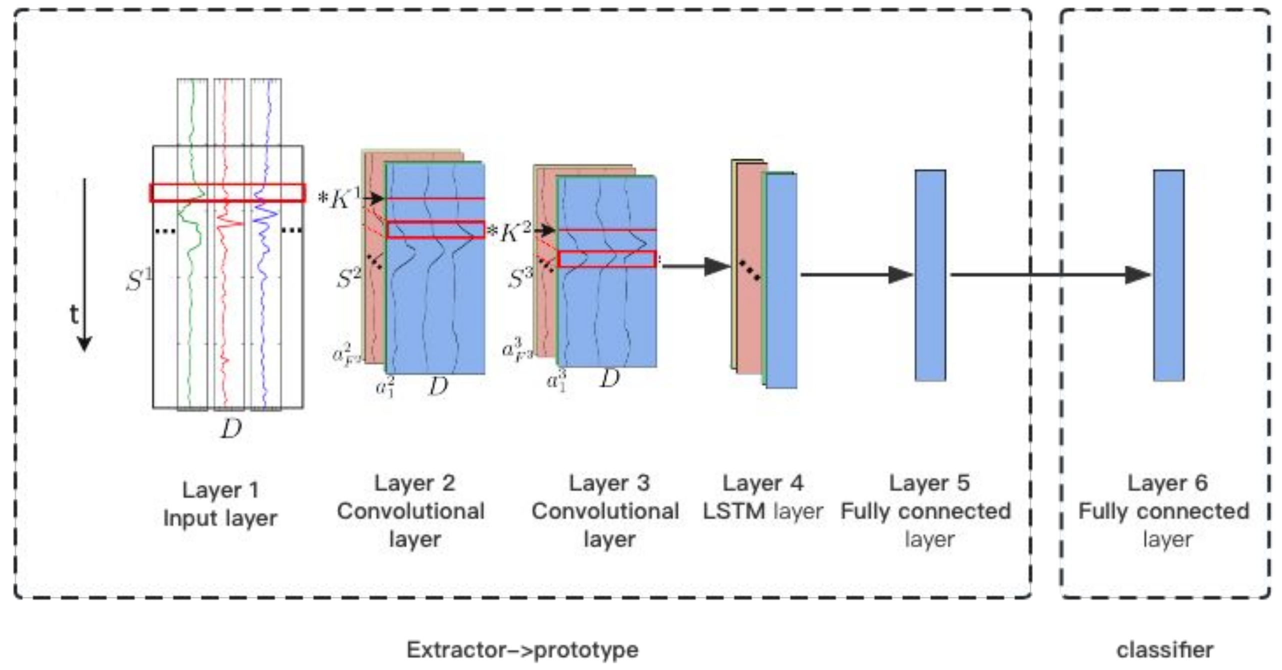

Figure 2. Architecture of our model

- Our model consists of 2 convolutional layers, 1 LSTM layer, and 2 fully connected layers.
- Take all but the last one layers as the feature extractor to obtain prototype.

2. LSTM

The connection weights and biases change once per episode of training; the activation patterns in the network change once per time-step.

3. Prototypical Network

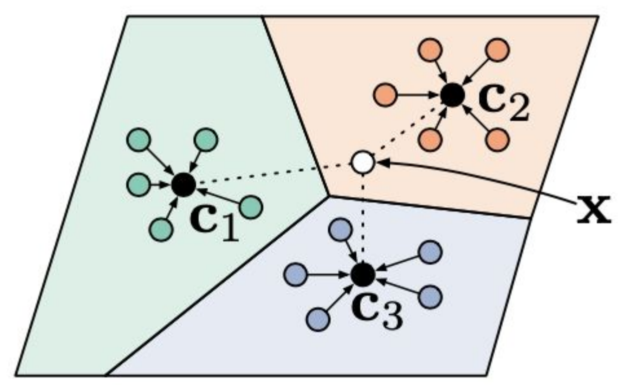

Figure 3. Prototypical network in the few-shot scenarios

- **Idea:** there exists an embedding in which points cluster around a single prototype representation for each class.
- **Method:** learn a non-linear mapping of the input into an embedding space using a neural network and take a class's prototype to be the mean of its support set in the embedding space.

Results

Activities	Methods		
	LSTM	FC + prototype	LSTM + prototype
Ascending Stairs	94.3%	49.1%	81.0%
Descending Stairs	50.0%	19.7%	63.6%
Jogging	85.2%	43.9%	94.5%
Jumping	90.6%	64.1%	91.7%
Lying	100.0%	100%	97.6%
Pushing	75.0%	31.8%	90.7%
Pushup (unseen)	25.0%	63.9%	91.0%
Sitting (unseen)	62.5%	87.1%	80.0%
Walking (unseen)	7.7%	21.6%	66.0%
total	**66.4%**	**52.36%**	**84.09%**
F1-score	**0.7089**	**0.5229**	**0.8389**

Table 1. Accuracy of different models on each activities

- LSTM performs better than FC in general.
- Prototypical networks bring great improvement in unseen activities: pushup, sitting, and walking.

Conclusion

- **LSTM** brings much **better** performance due to capturing temporal dynamics.
- **Prototypical** networks do well in **HAR few-shot** learning.
- Combining LSTM and prototypical networks can achieve satisfying performance.

References

[1] Ordóñez, F.J.; Roggen, D. Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition. Sensors 2016, 16, 115.

[2] Jake Snell, Kevin Swersky, Richard Zemel. Prototypical Networks for Few-shot Learning. Machine Learning (cs.LG), 2017.