An Investigation on Feature Extraction and Feature Fusion Methods for Wearable Sensor-Based Human Activity Recognition

Nguyen Thi Hoai Thu, Dong Seog Han
School of Electronics and Electrical Engineering, Kyungpook National University

thunguyen@knu.ac.kr, dshan@knu.ac.kr

Abstract

Feature extraction is usually considered as one of the most essential parts in wearable sensor-based human activity recognition (HAR) and classification tasks in general. In this paper, we carry out an investigation into feature extraction methods of both conventional machine learning and deep learning as well as feature fusion of these two approaches for wearable sensor-based human activity recognition. The hand-crafted features and automatically learning features are combined in order to provide the most useful information for the classification task. The experimental results on a benchmark dataset indicate that using hand-crafted features with deep learning models can give a better performance compare to other feature extraction methods.

Ⅰ. Introduction

Recently, human activity recognition (HAR) has gained great attention as its contribution to the domain of healthcare and human-computer interaction. With the ubiquity of smart wearable devices which contain powerful sensors, human activities, and abnormal behaviors can be automatically detected using the sensor data.

Research on human activity recognition can be generally grouped into two main approaches: conventional machine learning (ML) approach and deep learning (DL) approach. Conventional ML methods have been widely applied to HAR for the last two decades, in which the system contains two main parts: feature extraction and activity classification. Essential features are extracted from the sensor data by using several feature extraction methods in both time domain (e.g., mean, standard deviation) and frequency domain (e.g., Fourier Transform, Wavelet Transform) before being fed into some conventional classification models such as *k*-nearest neighbors (kNN) and support vector machine (SVM) [1]. Although this approach has succeeded in gaining significant achievements, it still has some limitations as the feature extraction often requires domain knowledge.

In the last few years, in the rapid growth of deep learning algorithms and powerful computational resources, several studies have delved into applying DL to human activity recognition [2, 3]. With extraordinary architectures such as convolutional neural networks (CNN) and long short-term memory (LSTM), deep learning has opened a new approach for human activity recognition where the features can be automatically extracted without expert knowledge. In addition, these deep features also help improve the performance of HAR, especially in complex activity recognition. However, there is an opening question which is whether DL automatic feature extraction methods always outperform the conventional methods.

Ⅱ. Method

In order to answer the question, we carry out several experiments on different feature extraction methods: The Wavelet transform, CNN, and LSTM on the public HAPT dataset [4]. The dataset contains data collected from accelerator and gyroscope embedded in a smartphone mounted at the waist of the users. Thirty participants carried out 12 activities: 6 basic activities and 6 postural transitions.

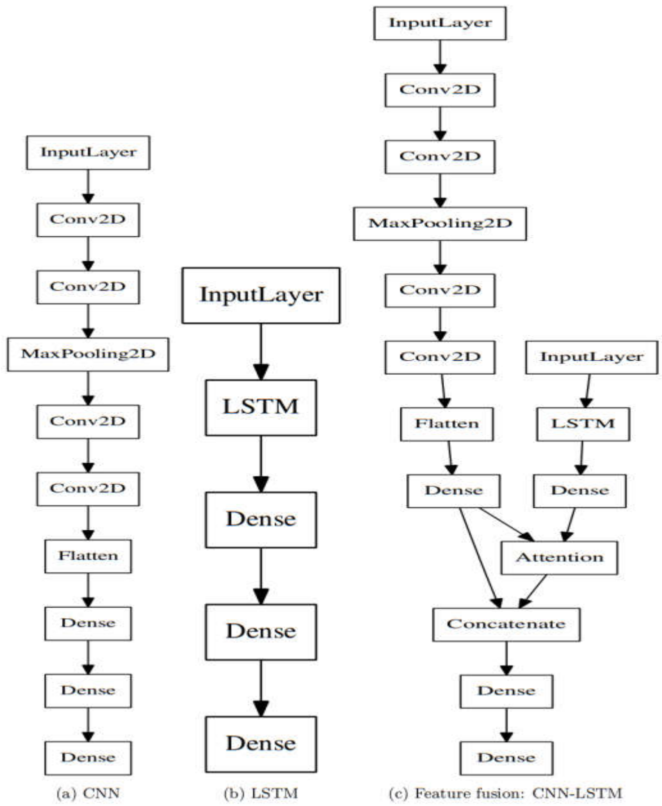

Figure 1. Three considered deep learning based HAR architectures

First, the data is split into fix-sized windows with an overlap of 50%. The Haar mother wavelet is used to extract the discrete wavelet transform (DWT) coefficients from the sensor data. Six HAR models have been made from the combination of 3 main architectures (CNN, LSTM, CNN-LSTM) and 2 types of input data (raw data, DWT coefficients). The detailed structures of the 3 main architectures are shown in Fig. 1. The CNN and LSTM models are implemented with standard sequential connections in which the CNN model contains several convolutional layers and maxpooling layer, followed by fully connected and softmax layers. In the feature fusion CNN-LSTM model, instead of connecting CNN and LSTM sequentially, the two sub-models are parallelly operated. An attention mechanism proposed by Luong *et al.* [5] is exploited to combine two outputs from the two sub-models.

The dataset is randomly split into 80% for training and 20% for validation. A $L1$-regularizer is used in all six models in order to avoid overfitting. In order to make a comprehensive investigation, two conventional ML methods: SVM and KNN ($k = 7$) are implemented. Each model is run for 10 experiments and the average accuracy is used as a performance metric. The results from the models are shown in Fig. 2. It can be clearly seen that, in most of the cases, the hand-crafted features give higher accuracy than the raw data except for the CNN model where the raw data achieve only 0.5% higher than the discrete wavelet transform features. In the feature fusion CNN-LSTM model, although both types of input data are used, it is not the one that gets the highest accuracy. The LSTM model which uses DWT data as input gets the highest accuracy and 3% higher than the raw input data.

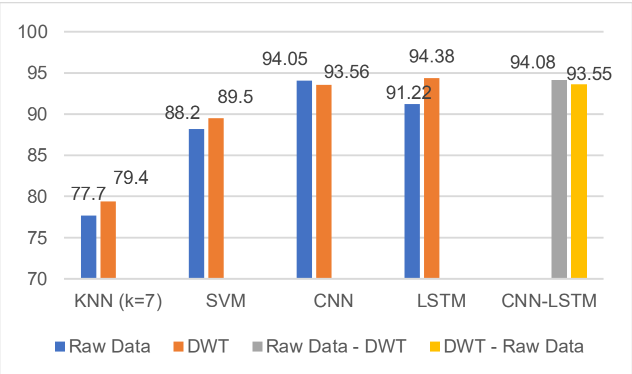

Figure 2. Performance comparison of different models with different input types

Ⅲ. Conclusion

In this paper, several feature extraction methods in both conventional and deep learning approaches have been implemented and applied to HAR. The experimental results indicate that although deep learning approach can automatically extract features from the raw data, in some cased, by exploiting the strength of domain knowledge in hand-crafted features, we can improve the performance of the system.

ACKNOWLEDGMENT

This work was supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (2016-0-00564, Development of Intelligent Interaction Technology Based on Context Awareness and Human Intention Understanding).

References

[1] O. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors," IEEE Communications Surveys & Tutorials, vol. 15, no. 3, pp. 1192-1209, 2012.

[2] O. Steven Eyobu and D. S. Han, "Feature representation and data augmentation for human activity classification based on wearable IMU sensor data using a deep LSTM neural network," Sensors, vol. 18, no. 9, p. 2892, 2018.

[3] N. T. H. Thu and D. S. Han, "Utilization of postural transitions in sensor-based human activity recognition," in 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), 2020, pp. 177-181.

[4] J.-L. Reyes-Ortiz, L. Oneto, A. Sam`a, X. Parra, and D. Anguita, "Transition-aware human activity recognition using smartphones," Neurocomputing, vol. 171, pp. 754-767, 2016.

[5] Luong, M. T., Pham, H., & Manning, C. D. Effective approaches to attention-based neural machine translation. arXiv preprint arXiv:1508.04025, 2015.