A Deep Learning Approach to Human Activity Recognition Based on Single Accelerometer

Ashna Jain(201501008)
ashna.j.btechi15@ahduni.edu.in

Dhruti Chandarana (201501015)
dhruti.c.btechi15@ahduni.edu.in

Janvi Patel(201501072)
janvi.p.btechi15@ahduni.edu.in

Nishi Patel(201501076)
nishi.p.btechi15@ahduni.edu.in

Charmi Chokshi (201501021)
charmi.c.btechi15@ahduni.edu.in

Hena Ghonia (201501032)
hena.g.btechi15@ahduni.edu.in

Abstract—Proposed method in this paper is Human Activity Recognition using Convolutional Neural Network (CNN), which is based on an acceleration. Constructed CNN model will adapt the characteristic of the tri-axial acceleration signals. Acceleration signals are generated using accelerometers sensor worn by individual subjects. Constructed dataset consists of 31688 samples from eight typical activities. The experiment results show that the CNN works well, which can reach an average accuracy of 87.9%.

Keywords-tri-axial acceleration signal; human activity recognition; deep architecture; convolution kernel

I. INTRODUCTION

Human activity recognition is a significant research topic and it is full of challenges.With improved technology , we now have more accurate sensors and faster processors with lower power consumption.Recently HAR Dataset has attracted more and more attention both from industry as well as from academia.

A HAR system can be divided into several modules , sensing , segmentation , feature extraction , classification and post pre-processing.Generally, according to sensing method, we can classify HAR system into two types: vision-based and acceleration-based methods. Vision-based method usually uses one or more cameras to collect data, while acceleration-based method asks the users to wear several accelerometers for data collecting. The advantage of vision-based system is that it works without placing any sensors with users, but its recognition performance highly depends on light condition, visual angle and other outer factors. On the contrary, acceleration-based system requires users to wear a device, but almost eliminates all those outer interferences.

In the recent years deep architectures due to their dramatically encouraging performance, have been involved in solving series of problems[4]. Some of the applications based on deep architecture have already been brought into service. With the aim of developing a HAR system with high accuracy, good robustness and quick response, we successfully built up a deep architecture for acceleration-based HAR system. The model has been evaluated on a large dataset (with 31688 samples from 8 typical activities). The result is promising, which reaches a higher accuracy than former methods. And more notably, the deep model directly operates on raw data. In other words, the model doesn't need an extra procedure of feature extraction, which matches our goal of quick response. In this paper, we briefly describe our used CNN approach.

II. CNN APPROACH TO ACCELERATION-BASED HAR

Recently, deep neural network architectures have made significant improvements in many fields of pattern recognition. Especially the CNN, one of the most powerful deep architectures is widely used in computer vision and image recognition.

A. Convolution Kernel

A conventional CNN approach is hardly utilized in tri-axial acceleration data without any changes in the basic model, because the width of the data greatly restricted the construction of a CNN architecture. There are two alternatives, resizing the data or resizing the convolution kernel. Resizing the data might be easier to implement, but it's not an appropriate approach. On the one hand, it may lose the relevance between adjacent acceleration values, which go against the design idea of CNN. On the other hand, it is inconvenient for generalization, the resized length and width are difficult to be determined when facing different acceleration data with varied length.

B. Locate the best epochs

Once the structure of the CNN is determined, we have to train the network, tuning the best parameters for activity recognition. Generally, the error rate of training set will gradually reduce as the training proceeds. At the beginning, the error rate on the test set will decrease, after a specific epoch, the error rate will stop decreasing, sometimes even increase. This phenomenon is called over fitting, it's caused by over training the network.

C. Preprocessing

The recorded raw 3D acceleration signal stream is cropped into the same size with an overlap of 50%. Every sample is a matrix with the size of 256x3. Since the sampling rate is around 100 Hz, one sample records the three dimension acceleration data with a length of about 2.56 seconds. After the simple preprocessing, we have 31688 labeled samples, 27395 are used for training, the rest 4293 are for testing. It's reasonable to consider the sample in the training set and the test set are independent because the signals of the training set and test set are collected by different volunteers.

D. Architecture of CNN

The detail of the CNN model is shown in following figure. It contains 3 convolution layers and 3 pooling layers. The model works directly on the raw acceleration signal without any other processing.

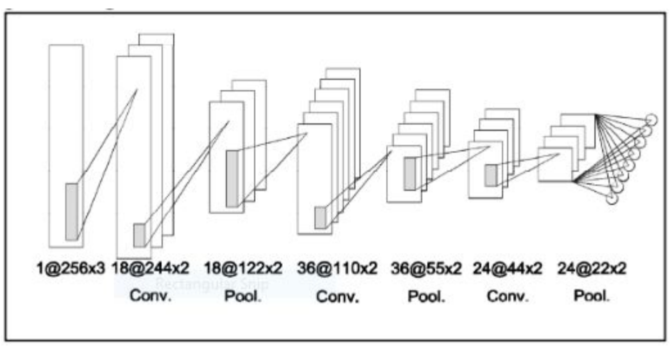

Fig1. CNN Architecture

E. Details about Datasets

Dataset used in this paper is constructed with the help of single accelerometer. It is different from UCI HAR dataset as it UCI HAR uses multiple sensors to capture the data. UCI HAR dataset uses many signal processing approaches to find out feature vector. DCT (Discrete Fourier Transform), FFT (Fast Fourier Transform), PCA (Principal Component Analysis), AR (Autoregressive Model) and Haar filters are used to find out feature vector of constructed dataset. Dataset consists of 31688 samples and 8 different activities. Using this dataset they are able to achieve accuracy of 87.9%.

They have also suggested HCII - SCUT dataset[3] which is also constructed using tri-axial accelerometer for the implementation of same application. It consists of 1278 samples of 44 different subjects and 10 different activities. It has also used DCT, FFT and AR model to construct the feature vector.

III. RESULTS

```
Epoch:98,batch:3216,loss:0.00000000,accuracy:1.00000000
Epoch:98,batch:4816,loss:0.00000000,accuracy:1.00000000
Epoch:98,batch:6416,loss:0.00000000,accuracy:1.00000000
Epoch:99,batch:016,loss:0.00002368,accuracy:1.00000000
Epoch:99,batch:1616,loss:0.00000000,accuracy:1.00000000
Epoch:99,batch:3216,loss:0.00000000,accuracy:1.00000000
Epoch:99,batch:4816,loss:0.00000000,accuracy:1.00000000
Epoch:99,batch:6416,loss:0.00000000,accuracy:1.00000000
Epoch:100,batch:016,loss:2.61394072,accuracy:0.93750000
Epoch:100,batch:1616,loss:0.00000000,accuracy:1.00000000
Epoch:100,batch:3216,loss:0.00000000,accuracy:1.00000000
Epoch:100,batch:4816,loss:0.00000000,accuracy:1.00000000
Epoch:100,batch:6416,loss:0.00000000,accuracy:1.00000000
Optimization finished!
Accuracy of testing:0.87987787
```

Fig 2. Result of implementation of CNN architecture

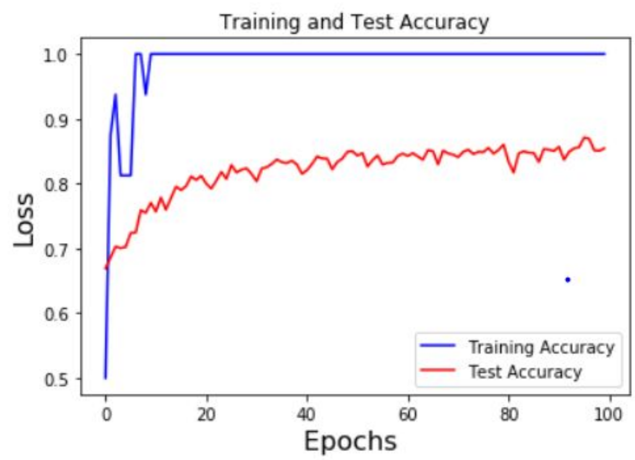

Fig 3. Training and test accuracy

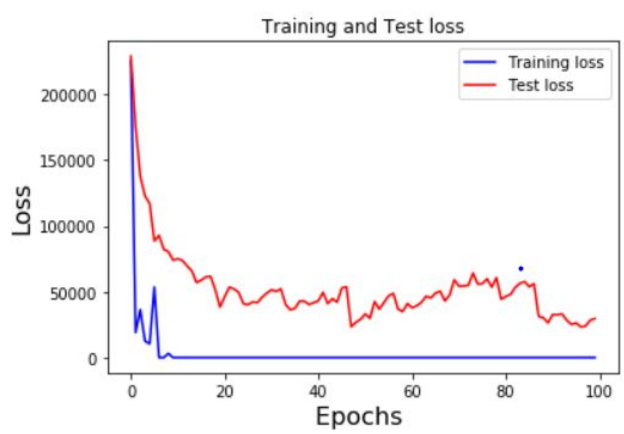

Fig 4. Training and test loss.

From above Fig 3. it can be seen that training accuracy becomes almost stagnant after 40-60 epochs and rarely changes as epochs increases. In the beginning, the validation accuracy was linearly increasing with loss, but then it did not increase much.

Changes done while implementing the approach used by author: We implemented this model on UCI HAR dataset and will do it on our dataset as well. We tried with all 9 channels instead of 3 channels. We decreased the input size to 128 instead of 256 and reduced to 2 convolutional layer and added one more fully connected layer to improve accuracy and decrease training and testing loss.

IV. CONCLUSION

System proposed an acceleration-based HAR algorithm using CNN, a popular used deep architecture in image recognition. According to the characteristic of acceleration data, we modified the conventional CNN structure.. The experiments are executed on a large dataset of eight kinds of typical activities with 31688 samples from 100 subjects. The results show that the CNN works well, reaches an accuracy of 87.9%. The proposed model is accurate and robust without any feature extraction, which is suitable for building a real-time HAR system on mobile platform.

REFERENCES

[1] David G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints”: Computer Science Department, University of British Columbia, Vancouver, B.C., Canada, lowe@cs.ubc.ca; January 5, 2004

[2] Coşkun, Musab, et al. "Face recognition based on convolutional neural network." *Modern Electrical and Energy Systems (MEES), 2017 International Conference on*. IEEE, 2017.

[3] Yang Xue, and Lianwen Jin, "A naturalistic 3D acceleration-based activity dataset & benchmark evaluations," Systems, Man and Cybernetics, pp. 4081-4085, 2010.

[4] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, "Imagenet classification with deep convolutional neural networks," Advances in neural information processing systems, pp. 1097-1105, 2012.