**Anomaly Detection in Health Data Based on Deep Learning**

Ning Han1, Sheng Gao2, Jin Li3, Xinming Zhang4, Jun Guo5

1,2,3,4,5 Beijing University of Posts and Telecommunications, Beijing 100876, China

hanning.bupt@foxmail.com,{ gaosheng, lijin0718, zhangxinming, guojun}@bupt.edu.cn

**Abstract:** Anomaly detection in health data means finding out abnormal human status automatically. In this paper, we propose a scheme which can be used for building monitoring system to promote quality of independent living and reduce the consequences of falls and diseases for elderly. We choose non-contact sensors to collect health data and build system using deep learning algorithms. This scheme includes two approaches, approach based on raw data which aims at abnormal activities and approach based on spectrogram which aims at abnormal status. Convolutional neural network is used to classify activities that predicted by support vector machine later, and recurrent neural network is used to predict signals directly. Through experiments, we evaluate performance of the scheme which proves it can solve the task successfully. Based on this result, it can be expected that our scheme will be utilized in reality.

**Keywords:** Deep Learning; Anomaly Detection; Health Monitoring; Convolutional Neural Network; Recurrent Neural Network

**1 Introduction**

Nowadays, aging population becomes a serious problem of our society. Elderly living alone increases rapidly. The risk of falls and diseases that occur when they stay alone need to be paid attention to. Mean-time, as health monitoring sensors and information systems in hospitals continue developing in today’s health-care environment, we can get a huge amount of health information [1].

There are already some researches aiming to fall prediction as anomaly detection based on vision data [2]. However, taking privacy and data size into consideration, sometimes video monitoring system is difficult to be expanded. Instead of camera, using sensors to monitor health status of the elderly in nurse house can also acquire health status data consists of wave signals about heartbeat, breath and body movement to build a monitoring system [3]. Using microwave radar to detect physiological movements can be traced back to the early 1970s [4]. This technique enabled non-contact detection of vital signs of humans or animals from a distance away, without any sensor attached to the body. Compared to either infrared or visible light, microwave has greater penetration capability through the building materials, which brings unique property to many applications. Monitoring system using microwave sensors is easier to be accepted and can collect health data more conveniently.

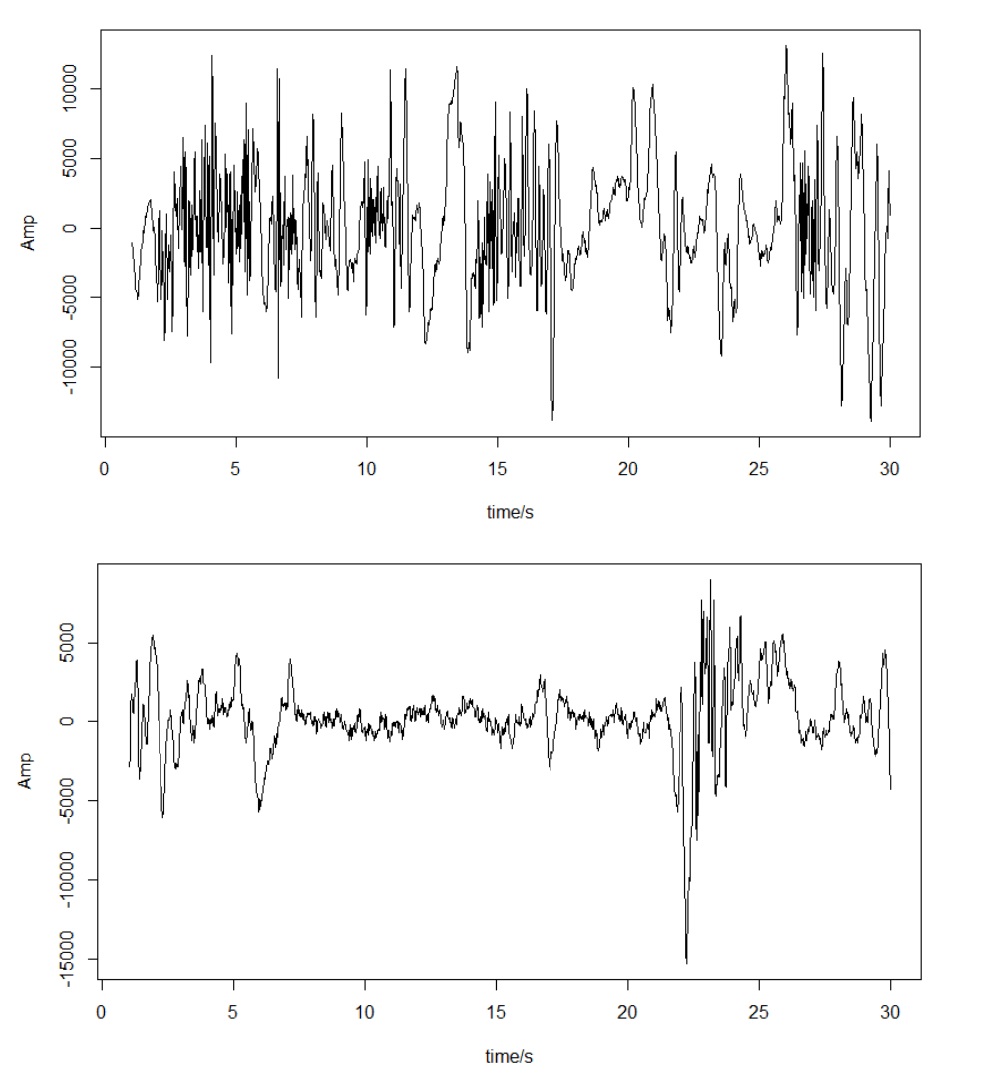
Another problem is that how to choose suitable algorithms to process the huge amount of sequence data. When using traditional machine learning algorithms to do anomaly detection, two-phase approach is necessary [5]. First, a preprocessing transforms the time-series data into a vector of features. Data segmentation, time window management, data normalization and dimensionality reduction are used to do such work. Then, feature vector is used as input of traditional classification algorithms to train model. However, in this procedure, the parameterization of feature extractors remains a manual process, and short-time signal analysis is intuitively problematic.

Deep learning means neural networks contain more than one hidden layer to learning tasks. Deep structure helps it to abstract more complex information from input data. Not like traditional machine learning algorithms, it doesn’t need to extract features from input data at first phase which means that we will keep information from input data as more as possible. Deep learning has been applied to fields including computer vision, speech recognition, natural language processing, audio recognition, social network filtering and machine translation where they produced results comparable to and in some cases superior to human experts [6]. We believe that deep learning algorithm can be helpful to detect anomaly in health data.

In this paper we propose a scheme using deep learning to solve anomaly detection task in health data. Section 2 introduce how we collect data by non-contact microwave sensors. The system and methods are explained in Section 3. We analyze our experimental results in Section4 and give conclusions in Section 5.

**2 Dataset**

We put two non-contact microwave sensors and a camera in volunteer’s room to collect data. One non-contact microwave sensor is on the bed and another one is on the table. Non-contact microwave sensor and camera that put in the room to collect human signals and videos conveniently in 2 days. The frequency of the non-contact microwave sensors we used is 50 Hz, which means that the sensors will emit microwave and record the amp of reflected wave 50 times every second. And the sensors sent data to storage device every 10 seconds. In other words, we receive data consists of 500 components every 10 seconds.



**Figure 1** Data samples for eating

Then we label all human signals collected by non-contact microwave sensors with different type of activities according to video. We divide activities to 12 classes. Activities set *A* = {housework, toilet, watch video, be away, eat, REM sleep, Non REM sleep, get up, lie, work, play game, move}. We label the activity every 10 seconds as this is the time interval for sensors to send raw data. Figure 1 shows a segment of wave signals from raw data we collected for eating. In our dataset, the amount of data for each activity is different because of the difference of time length for each activity. For example, time for sleeping must be longer than eating, so the set of data for sleeping is larger than the set of data for eating.

It need to be mentioned that abnormal data are not in our data set. Because generating anomaly artificially such as pretending to fall down is unreasonable, the human status will be totally different with falling down suddenly. Instead, we generate some random signals in experiments used as test data.

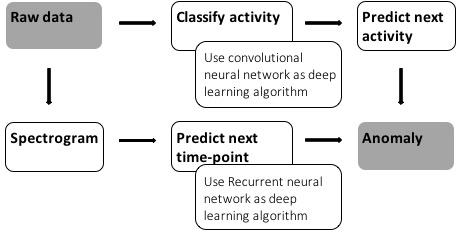
**3 Scheme and Methods**

**3.1 Scheme**

We try to propose a scheme that can detect two kinds of anomaly: abnormal activity and abnormal data. Abnormal activity means that the activity is normal but the changing of activities over time is abnormal such as excessive sleeping. Abnormal data means that the data is totally different with previous pattern, maybe some activities can not be classified correctly appear.

Corresponds to these two types of anomaly, we design the scheme with two approaches. One is based on raw data. In monitoring system just use non-contact microwave sensor, we can not figure out the activity happened directly. So at first we build a model to classify activities. Then try to predict what next activity is normally. By comparing prediction and classification from real-time raw data, we can find anomaly. Another is based on spectrogram. A spectrogram is a visual representation of the spectrum of frequencies of signal as they vary with time. Then we model the spectrogram sequence and predict next spectrogram segment. We can find anomaly when spectrogram calculated by real-time data is far away from prediction. Figure 2 shows the structure of whole scheme.

**3.2 Approach based on raw data**



**Figure 2** Scheme structure

**3.2.1 Convolutional neural network**

We choose convolution neural network (CNN) to do classification. Convolutional neural network (CNN) is a class of deep, feed-forward artificial neural network that have successfully been applied to analyzing visual imagery [7]. CNN uses a variation of multilayer perceptron designed to require minimal preprocessing. A CNN consists of an input and an output layer, as well as multiple hidden layers. The hidden layers are either convolutional, pooling or fully connected. Convolutional layer will compute the output of neurons that are connected to local regions in the input, each computing a dot product between their weights and a small region they are connected to in the input volume. Convolutional layer makes CNN can capture the relationship between neighbor regions and also reduce the amount of parameters.

Use convolutional neural network to classify activity. A training set for CNN model in our research is raw signals sequence before time-point *t* of length *T* from *N* sensors and activity label .

|  |  |
| --- | --- |
|  | (1) |

When the model receives new data sent by microwave sensors, it will classify the activity for this time-point and we can get the classification label: .

**3.2.2 Activity prediction**

Because activity sequence is very simple with a small set A, and the pattern of the activities changing in someone’s regular life is obvious, we choose traditional machine learning algorithm support vector machine to predict next activity. Use traditional machine learning algorithm(SVM) to predict activity.

SVM is a typical machine learning algorithm which have perfect performance in many supervised learning tasks [8]. Input of SVM model is an activity sequence before the current time-point t of length *T.*

|  |  |
| --- | --- |
|  | (2) |

By minimize

|  |  |
| --- | --- |
|  | (3) |

where *w* and *b* are parameters, means label and means component of feature vector, SVM can build a model to give prediction of next time-point activity .

If we find that the prediction of next time-point activity , this time point will be considered as anomaly.

**3.3 Approach based on spectrogram**

**3.3.1 Spectrogram**

In this approach, we transform signal waves to spectrograms at first. Figure 3 visualized the transformation from raw data to spectrogram. The spectrogram has two geometric dimensions: the horizontal axis represents time; the vertical axis is frequency. A third dimension indicating the amplitude of a particular frequency at a particular time is represented by color of each point in the image.

**3.3.2 Recurrent neural network**

Recurrent neural networks (RNN) are models that can capture the dynamics of sequences via cycles in the network of nodes. Unlike standard feed forward neural networks, recurrent networks retain a state that can represent information from an arbitrarily long context window [9]. Although recurrent neural networks have traditionally been difficult to train, and often contain millions of parameters, recent advances in network architectures, optimization techniques, and parallel computation have enabled successful large-scale learning with them.

We use recurrent neural network to do anomaly by predicting spectrogram with *F* components. An input for RNN model is a spectrogram sequence before time point *t* of length *T* from *N* sensors:

|  |  |
| --- | --- |
|  | (4) |

And the network predict spectrogram for next time point . RNN calculates the sequence to the sequence by iteration

|  |  |
| --- | --- |
|  | (5) |

Where *f* is activating function, is parameter matrix. Maximize sequence generation conditional probability to estimate parameter.

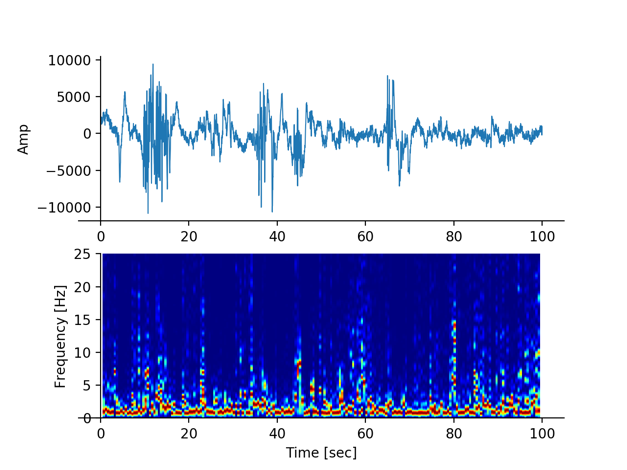
(6)

(7)

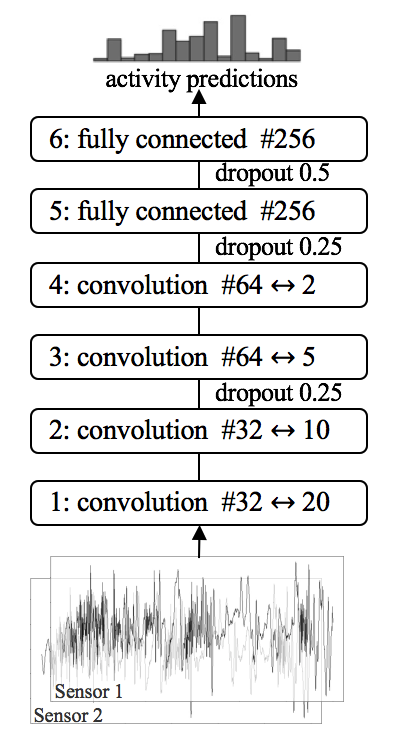
In our research, we choose Long Short Term Memory which is capable of learning long-term dependencies [10]. LSTM is explicitly designed to avoid the long-term dependency problem. Remembering information for long periods of time is practically their default behavior.

After prediction we calculate mean absolute error be- tween prediction with real spectrogram for next time point.

(8)



**Figure 3** Signal waves and spectrogram



**Figure 4** The convolution neural network architecture we used. The filter sizes (↔) and numbers of units (#) are indicated.

If , where is a pre-chosen threshold for mean absolute error, the time point *t* + 1 will be considered as anomaly.

**4 Experimental Results**

**4.1 Approach based on raw data**

**4.1.1 Classification**

In experiments we segment data by 10 time intervals, use raw data from two sensors as two layers of input data. And we set 4 convolution layers and 2 full connecting layers. We also add regularization and dropout to avoid over fitting. Figure 4 shows the best structure and parameters we tuned. To solve the problem of unbalanced data set, we over sample the data with rare activities.

We choose random forest as traditional machine learning algorithm to do baseline experiments. We extract heartbeat, breathe and body-movement data according to frequency from raw signal and segment them using time window. And for features, we choose average value, max coefficient and frequency after FFT (Fast Fourier Transformation).

**Table I** Accuracy comparison between two classification models

|  |  |  |  |
| --- | --- | --- | --- |
| Activities | CNN | RF | |
| Housework | 1.0000 | 0.4063 |
| Toilet | 0.9420 | 0.1477 |
| Watch video | 0.9928 | 0.6353 |
| Be away | 1.0000 | 0.5714 |
| Eat | 0.9589 | 0.6086 |
| REM sleep | 0.9783 | 0.5977 |
| Non-REM sleep | 0.9981 | 0.9660 |
| Get up | 1.0000 | 0.7470 |
| Lie in bed | 0.9835 | 0.8256 |
| Work | 0.9638 | 0.9567 |
| Move | 0.9960 | 0.9226 |

**Table II** Accuracy comparison between different data sources

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Activities | Position of sensor | | | |
| **Table + Bed** | **Table** | **Bed** | |
| Housework | 1.0000 | 1.0000 | 1.0000 |
| Toilet | 0.9420 | 0.9855 | 0.8841 |
| Watch video | 0.9928 | 0.9783 | 0.9638 |
| Be away | 1.0000 | 1.0000 | 1.0000 |
| Eat | 0.9589 | 0.9452 | 0.9178 |
| REM sleep | 0.9783 | 0.9946 | 0.9293 |
| Non-REM sleep | 0.9981 | 0.9889 | 0.9926 |
| Get up | 1.0000 | 1.0000 | 1.0000 |
| Lie in bed | 0.9835 | 0.9835 | 0.9571 |
| Work | 0.9638 | 0.9805 | 0.9721 |
| Move | 0.9960 | 0.9759 | 0.9880 |

Table I shows the accuracy of the classification. As it showed, the performance of CNN model is better than traditional approach in all activities.

To certify robustness of CNN model, we change the data source of network. Using same structure and parameters of network, we train model in dataset consists of signals collected by one sensor and test classification performance. Table II shows the comparison. It can be found that although use data collected by two sensors together get the best performance, just use data collected by one sensor also classify activities well. Especially by using data collected by sensor on table, the model can achieve almost same performance with using data collected by two sensors.

**4.1.2 Activity prediction**

In this task, we set sequence length T as 1000, which means SVM model use 1000 consequent activities to predict next activity. Table III shows the performance of activity predicting. We can find that the average accuracy of prediction is over 95%, which means that when an abnormal activity occurs the system can detect it with error rate less than 5%.

**4.2 Approach based on spectrogram**

Figure 5 shows the structure and parameters we used in spectrogram prediction task. RNN model uses spectrogram segments which length is 10 time points to predict spectrogram for next time point.

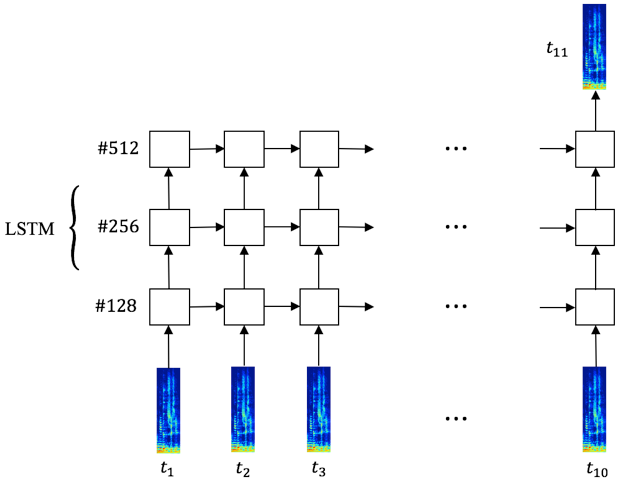
Since there is no anomaly in our data set, we need to generate test data set to test performance of RNN model

**Table III** Performance of activity predicting

|  |  |  |
| --- | --- | --- |
| Activities | Accuracy | |
| Housework | 0.9318 |
| Toilet | 0.9688 |
| Watch video | 1.0000 |
| Be away | 1.0000 |
| Eat | 1.0000 |
| REM sleep | 1.0000 |
| Non-REM sleep | 0.9981 |
| Get up | 0.8514 |
| Lie in bed | 0.9858 |
| Work | 0.9608 |
| Move | 1.0000 |

**Table IV** Comparison of MAE between different test data sets

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| MAE | Test data | | | |
| Normal class | Abnormal class | Random data | |
| Average | 0.0023 | 0.0096 | | 0.0104 |
| Deviation | 0.0029 | 0.0147 | | 0.0053 |

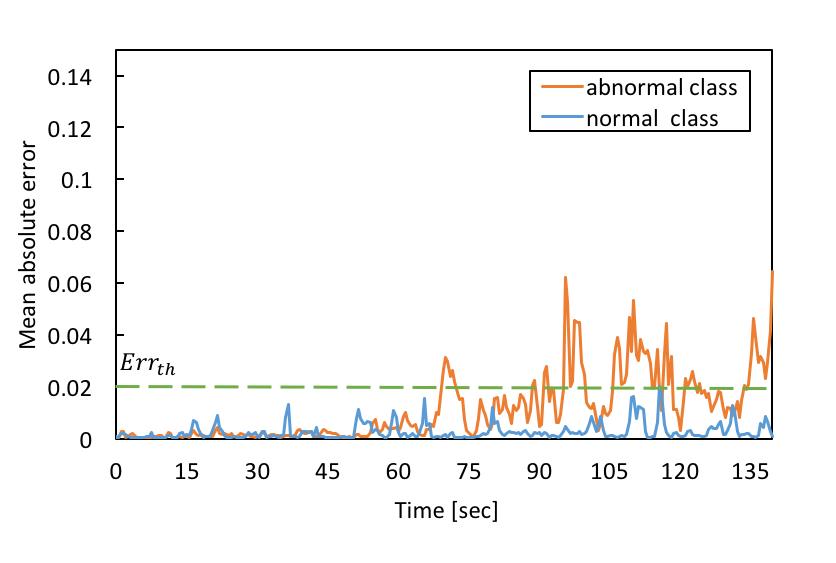


**Figure 5** The recurrent neural network architecture we used. The numbers of units (#) are indicated.

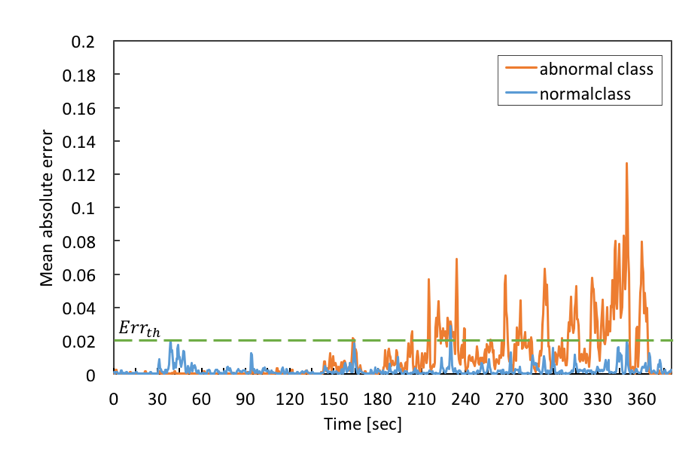
for anomaly detection. We propose two ways to imitate abnormal data. One is that we set housework as abnormal activity and train the model just with data belongs to other 11 activities. Another is to generate random data with normal distribution in raw signals. We also split test data set from normal part. Use the same network to predict next time point in these test data sets and calculate mean absolute errors. Table IV shows average and deviation of error in test data. We can find that average error of two kinds of imitation anomaly is larger than normal data.

With setting threshold of error as 0.02, our model can detect both these imitation anomalies. Figure 6 shows two examples of comparing mean absolute error between using data from normal class with abnormal class. Figure 7 shows generating of random signal waves and mean absolute error comparison. After 100 seconds we replace original signals with random data. And we can find that spectrogram become chaotic and error increase obviously with exceeding the threshold.

**5 Conclusions**

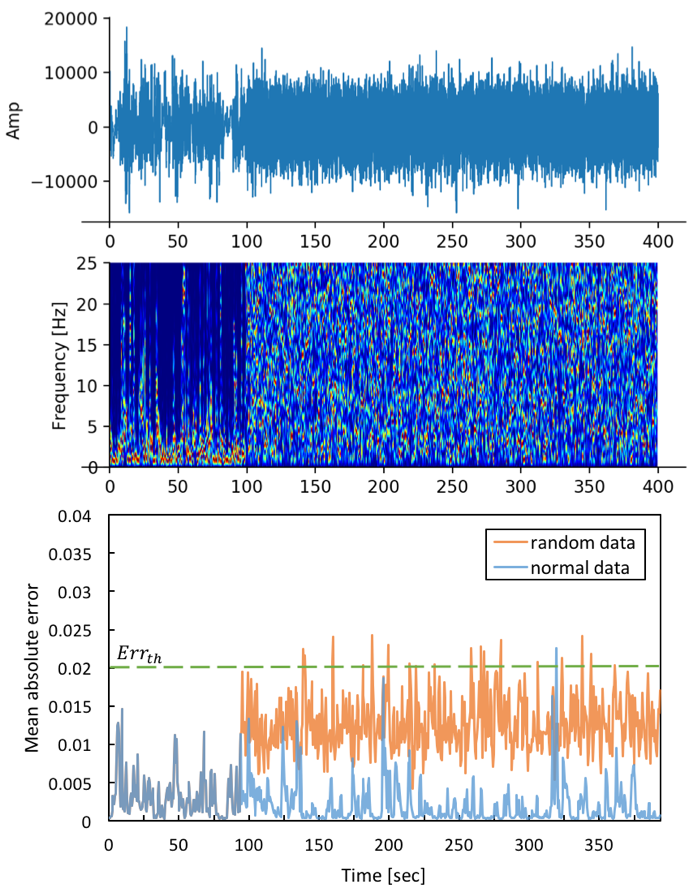
****

(a)



(b)

**Figure 6** MAE comparison between abnormal and normal activity, (a) and (b) are two examples



**Figure 7** Generate and test random signals as anomaly

In this paper, we have investigated whether deep learning algorithm can solve anomaly detection in health data task. Two approaches were proposed for this task, one based on raw data used CNN and SVM, another based on spectrogram used RNN. We proved that the performance of both approaches is good enough to detect anomaly from our data set, which showed that the deep learning networks are able model raw signal waves and spectrogram data.

In future work, we will expand our data set. Now there are just data of normal activities for two days of one person in a room. It is not enough to certify the robustness of our scheme. Before being utilized in reality, we need to collect more real data of elderly and evaluate the scheme using true abnormal data. Finally, we would like to establish a real monitoring system based on this research for elderly.

**Acknowledgements**

This work is supported by National Natural Science Foundation of China (61702047, 61300080), Beijing Natural Science Foundation (4174098) and the Fundamental Research Funds for the Central Universities (2017RC02).

The idea of applying deep learning algorithm to health data is obtained when I spent 1-year student exchange program at the University of Electro-Communications Tokyo Japan, where I was supervised by Prof. Numao to promote this research. His lab is involved in the development of watching systems for aged people. The microwave data was prepared by Ms. Tei, who was also master student at Numao Lab.

**References**

1. Barnes, Jody, “Intrusion Detection Systems in Hospitals: What, Why, and where. publisher not identified”, 2006.
2. R. Alazrai, Y. Mowafi and E. Hamad, “A fall pre- diction methodology for elderly based on a depth camera,” 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, 2015, pp. 4990-4993.
3. Numao, Masayuki, and Shuya Masuda, “Non- Restrictive Continuous Health Monitoring by Integration of RFID and Microwave Sensor,” AAAI Spring Symposium Series, 2016.
4. Lin, Jenshan, and Changzhi Li. “Wireless non-contact detection of heartbeat and respiration using low-power microwave radar sensor,” Microwave Conference, 2007.
5. Fuertes, Sylvain, et al, “Improving Spacecraft Health Monitoring with Automatic Anomaly Detection Techniques,” 14th International Conference on Space Operations, 2016.
6. LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton, “Deep learning,” Nature 521.7553 (2015): 436- 444.
7. Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton, “Imagenet classification with deep convolutional neural networks,” Advances in neural information processing systems, 2012.
8. Hsu, Chih-Wei, and Chih-Jen Lin, “A comparison of methods for multiclass support vector ma- chines,” IEEE transactions on Neural Networks 13.2 (2002): 415-425.
9. Lipton, Zachary C., John Berkowitz, and Charles Elkan, “A critical review of recurrent neural networks for sequence learning,” arXiv preprint arXiv:1506.00019 (2015).
10. Hochreiter, Sepp, and Jrgen Schmidhuber, “Long short-term memory,” Neural computation 9.8 (1997): 1735-1780.