HAND GESTURE RECOGNITION USING BIOACOUSTIC RESPONSE

Shizuru Iida
Tokyo University of Science, Faculty of Science and Technology, Chiba, Japan
email: 7517603@ed.tus.ac.jp

Takumi Asakura
Tokyo University of Science, Faculty of Science and Technology, Chiba, Japan

In recent years, wearable devices such as the smartwatch and activity tracker are becoming more popular. They are unique from the viewpoints that they are worn on the body, therefore they are applicable as always-available interface for input and output. Especially, such a device located on the wrist is advantageous for capturing biosignals such as the biometric and bioacoustic signals. Some studies on estimation of hand gestures by using biosignals has been conducted, however, few studies have focused on the problem that the number of measuring times for the estimation increases in proportion to the number of gestures. In this study, we propose the method to reduce the number of measuring times for estimating the hand gestures. We recognize the hand gestures by actively applying the biometric vibrations. First, we developed a wearable wristband sensor provided with a transducer and a contact microphone. For estimating the hand gestures, the transducer inputs the unnoticeable vibration to the wrist and the contact microphone receives the transmitted signal. We estimated energy spectral density(ESD) of the received signals and measured the subset hand dimensions(SHD), the 4 hand dimensions independent of each other. we made the individual multi-layer perceptron (MLP) model by fine-tuning the pre-trained MLP model, applying it to estimate hand gestures. We input the dataset of ESD and SHD as the feature values to individual MLP model. We confirmed the recognition accuracy and the mean value of the accuracy reached more than 80% (7 participants, 7 kinds of gestures). As future work, we planned to consider estimating the SHD from recorded bioacoustics signals.

Keywords: wearable device, machine learning, bioacoustics, gesture recognition

1. Introduction

1.1 Wearable devices

In recent years, wearable devices are becoming more popular. Wearable devices are smart electronic devices used by being worn on a body such as an arm or a head. Wearable devices can collect various data on daily activities. By analyzing the collected data, A variety of services are being studied for various fields and subjects. Nowadays, various products and services have appeared in the fields of business use, health care, sports, medical, etc.

1.2 Bioacoustic signal

Typical examples are smartwatch and activity trackers worn on a wrist. According to the research by IDC (Table 1, [1]), smartwatches and wristbands accounts for less than 95% of the total shipments in 2017; Therefore, most of the wearable device is to wear on the wrist. The human wrist is an ideal site for users to collect, display, interact with data and executing tasks using applications. Because of this, wearable devices on the wrist are applicable as always-available interface for input and output. There are various biosignals collected...
from the wrist such as biometric signals (blood pressure, heart rate, body temperature) and electrical biosignals (electroencephalogram(EEG), electrocardiogram(ECG), electromyogram(EMG) and Electrooculography(EOG)). In terms of biosignals, mechanical vibrations are mechanical biosignals, especially called “bio-acoustic” signals.

1.3 Hand gesture recognition

With the spread of virtual reality technology and wearable devices, there has been a growing interest in the studies about hand gesture recognition. In these studies, supervised learning such as support vector machine (SVM) and multi-layer perceptron (MLP) are one of popular methods and need the training dataset collected by wearable devises on the wrist. It is reported that Hand gesture recognition has been studied by widely sensing approaches [2]:

- Accelerometer: using multiple orientation sensors to recognize arm gestures [3]
- Glove: using embedded strain gauges to measure finger orientation [4]
- Vision: using wearable camera to finger gestures [5]
- Electrical biosignals: using EMGs [6]
- Bioacoustic (by accelerometer): sensing bioacoustic signals by overclocked smartwatch accelerometers [7]
- Active bioacoustic (by contact microphone and vibration transducer): sensing bioacoustic signals and reflections by the contact microphone for capturing vibrations from the transducer [8]
- Ultrasound: using the ultrasound imaging array [9]

In all of these studies, the diversity of the anthropometry of humans forces users multiple times to measure for make the model to recognize hand gesture and still remains an open research problem. However, as far as we know, these have been few reports about this problem.

1.4 The purpose of this study

The purpose of this study is to reduce the number of measurements for make the model to recognize hand gestures. For this purpose, we adopted fine-tuned MLP as the recognition method and feature values input the MLP model are energy spectral density(ESD) and subset hand dimensions(SHD).

Table 1: World wearable device market forecast by product category: Unit shipments (in millions)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Watch</td>
<td>67.4</td>
<td>55.4%</td>
<td>154.4</td>
<td>67.3%</td>
<td>+23.0%</td>
</tr>
<tr>
<td>Wristband</td>
<td>47.6</td>
<td>39.1%</td>
<td>51.3</td>
<td>22.4%</td>
<td>+1.9%</td>
</tr>
<tr>
<td>Shoes/Cloth</td>
<td>2.8</td>
<td>2.3%</td>
<td>11.6</td>
<td>5.1%</td>
<td>+42.8%</td>
</tr>
<tr>
<td>Ear-hung</td>
<td>1.8</td>
<td>1.5%</td>
<td>10.5</td>
<td>4.6%</td>
<td>+54.4%</td>
</tr>
<tr>
<td>Modular</td>
<td>1.6</td>
<td>1.3%</td>
<td>1.5</td>
<td>0.6%</td>
<td>-2.8%</td>
</tr>
<tr>
<td>Others</td>
<td>0.4</td>
<td>0.3%</td>
<td>0.2</td>
<td>0.1%</td>
<td>-10.5%</td>
</tr>
<tr>
<td>Total</td>
<td>121.7</td>
<td>100%</td>
<td>229.5</td>
<td>100%</td>
<td>+17.2%</td>
</tr>
</tbody>
</table>

2. Method

2.1 Principle

2.1.1 Fine-tuning

The multi-layer perceptron (MLP) is a method of supervised learning. The universal approximation theorem for neural networks states that every continuous function that maps intervals of real numbers to some output interval of real numbers can be approximated arbitrarily closely by a multi-layer perceptron with just one hidden layer [11]; therefore, the MLP is used for classification of various complex phenomena. A one-hidden-layer MLP is shown in Fig. 1. The output of nth unit is shown in Eq. (1).
\[ y_j(x,w) = f\left(\sum_{i=0}^{N} w_{ji} x_i \right). \] (1)

The input of \( i \)th unit includes \( x_1, \ldots, x_N \) and the bias unit \( x_0=1 \). \( w_{ji} \) is weight between the input of \( i \)th unit and the output of \( j \)th unit. \( f \) is called activation function and used non-linear functions such as sigmoid, softmax and rectified linear unit (ReLU). The MLP is a hierarchical connection of this unit so that the output of the lower layer becomes the input of the upper layer. As a result, the final output is shown in Eq. (2).

\[ y_k(x,w) = f^2\left(\sum_{j=0}^{N} w_{kj} f^1\left(\sum_{i=0}^{N} w_{ji} x_i \right)\right). \] (2)

Fine-tuning means taking weights of a pre-trained neural network and use it as initialization for a new model being trained on data from the same kind. It is used to speed up the training and overcome small dataset size.

In this study, we used the two-hidden layer MLP. The numbers of units of [input, hidden1, hidden2, output] are [174, 500, 250, 7] units. Activation function is the ReLU shown in Eq. (3). The optimization algorithm is Adam. The rate of dropout is 0.5 at the output of hidden layers.

\[ f(x) = \log(1 + \exp x). \] (3)

![Figure 1: the one-hidden-layer perceptron](image)

2.1.2 Energy spectral density (ESD)

The energy spectral density (ESD) shows how the energy of signals are distributed with respect to frequency. The ESD is often used for evaluating impulse like finite energy. If a discrete signal has the value represented by \( f_n = f(n, dt) \) and continues infinitely, the ESD of the signal are shown in Eq. (4). \( F(\omega) \), \( dt \), and \( f_s \) is discrete-time fourier transform (DTFT), sampling interval, and sampling frequency of \( f_n \).

\[ \text{ESD} = \left| \frac{dt}{2\pi} \sum_{i=-\infty}^{\infty} f_se^{-i\omega_i} \right|^2 = \frac{1}{2\pi f_s^2} F_d^2(\omega) F_d^2(\omega). \] (4)

2.1.3 Subset hand dimensions (SHD)

The subset hand dimensions (SHD, Fig. 2) are 4 hand dimensions independent of each other defined in the study by T. Nohara et. al. [10]. They used it for the estimation of 55 fullset hand dimensions to deform a polygon hand model by multiple regression analysis. The estimation error from the value by manual measurement are 0.23 mm and very small.

In this study, we regarded the SHD as feature values about the shape of a hand, one of the diversity of the anthropometry of humans.
2.2 System

2.2.1 Wristband sensor

For collecting bioacoustics signals, we implemented the wristband sensor shown in Fig. 3a. The wristband sensor has one transducer (COM-10917, SparkFun) and one contact microphone (CM-01B, TE Connectivity). In measurement shown in Fig. 1b., the transducer transmits up-time stretched pulse (up-TSP) signal \( n = 2^{11} \), filtered 1-2 kHz: the band of this frequency is unnoticeable) mechanical vibration into a wrist, then the complex of direct, bone-conducted, reflexed vibrations, etc. were captured by contact microphone and recorded to PC.

2.2.2 System flowchart

The flowchart of this study is shown in Fig. 4.

In step(i), we created the training dataset for training the MLP model. The training dataset consists the ESD, SHD and the correct levels of hand gestures from the following measurement. We recorded bioacoustic signals by the wristband sensor. We convolved recorded data with inverted time stretched pulse (ITSP, the signal inverted the signal transmitted to the transducer) and obtained impulse response (IR). We estimated the normalized ESD from IRs. The SHD are manual measured by a digital caliper (19975, Shinwa Rules Co., Ltd.).

In step (ii), we created the individual MLP model and used it for recognition of hand gestures. In the first, we trained the non-trained MLP model with the training dataset and obtained the parameters of pre-trained MLP. Second, we fine-tuned the pre-trained MLP with the calibration data and created the individual MLP model. Third, we input the ESD and SHD to the individual MLP model and recognized hand gestures from the maximum output of the individual MLP model.
2.3 Experiment

We conducted an experiment for verifying the accuracy of the constructed system. The subjects are 7 males of twenties. The subjects wore the wristband sensor and we recorded 90 bioacoustic signals when the subjects taking each hand gesture as shown in Fig. 6. We created the dataset of estimated normalized ESD of the hand gestures (170 dimensions), measured SHD (4 dimensions), and the correct levels of the hand gestures. The dataset of all were divided as shown in Fig. 5: the dataset of the hand gesture (a) to (d) divided into 60 training data A and 20 test data B. the dataset of the hand gesture (e) divided into 60 training data A, 20 test data B, and 10 calibration data C, and the dataset of each hand gesture of (e) to (G) are 60 calibration data C and 20 test data B.

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
<th>(d)</th>
<th>(e)</th>
<th>(f)</th>
<th>(g)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training data</td>
<td>2100</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>(for training non-trained MLP)</td>
<td>(60<em>5kinds</em>7subjects)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calibration data</td>
<td>130</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>(for fine-tuning pre-trained MLP)</td>
<td>(10+60*2, each subject)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test data</td>
<td>140</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>(for testing individual MLP)</td>
<td>(20*7kinds, each subject)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 5: The dataset used in this study

3. Result

3.1 Measured data

Figure 6 shows the estimated normalized ESD corresponding to the hand gestures. The ESD of each hand gesture is different from each other especially between 1-1.6 kHz.
The measured value of SHD is shown in Table 2. Compared with the SHD of T. Nohara (338 subjects (102 males and 336 females) [10], there is a tendency that the mean of our SHD is a little larger and the SD of our SHD is slightly smaller. Regarding the difference about the mean is because the subject was only male, and the size of the hand was larger than the average of male and female combined.

Table 2: The subset hand dimensions

<table>
<thead>
<tr>
<th>Value [mm]</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T. Nohara [10]</td>
<td>Mean 177.5</td>
<td>79.6</td>
<td>17.3</td>
<td>16.5</td>
</tr>
<tr>
<td></td>
<td>SD 10.3</td>
<td>5.8</td>
<td>1.4</td>
<td>1.4</td>
</tr>
<tr>
<td>This study</td>
<td>Mean 185.0</td>
<td>82.9</td>
<td>18.0</td>
<td>16.2</td>
</tr>
<tr>
<td></td>
<td>SD 9.5</td>
<td>4.8</td>
<td>1.8</td>
<td>1.2</td>
</tr>
</tbody>
</table>

3.2 Accuracy of the system for hand gesture recognition

Across all 7 subjects and 7 gestures, our system achieved a mean accuracy of 86.4% (SD=5.4%). Figure 7 shows a confusion matrix for each gesture set. The hand gestures between (a) to (d), used training for the non-trained MLP model, scored higher accuracy among the whole. The hand gestures (f) and (g), used fine-tuning for the pre-trained MLP model, scored lower accuracy among the whole. The hand gestures (e), used training and fine-tuning the MLP models, is similar accuracy of the hand gestures (f) and (g) because of the similarity to the hand gesture (g).

In the result, if the pre-trained MLP model is available, the individual MLP model can recognize the pre-trained gestures with practical accuracy without the measurement. Nevertheless, if the ESD of the hand gestures added recognized classes is similar to ESD of pre-trained gestures, it is concerned that the accuracy of added hand gestures is lower than the accuracy of the pre-trained gestures.
4. Conclusion

We proposed the method for the purpose to reduce the number of measurements for make the model to recognize hand gestures. We adopted fine-tuned MLP as the recognition method and the feature values as energy spectral density (ESD) and subset hand dimensions (SHD). We conducted the experiment for verifying the accuracy of the constructed system. We created the pre-trained MLP model from the measurement of ESD and SHD and the individual MLP model by fine-tuning it using few calibration data per-subjects.

In the result, across all 7 subjects and 7 gestures, our system achieved a mean accuracy of 86.4% (SD=5.4%) with few calibration per-subjects. Nevertheless, if the ESD of the hand gestures added recognized classes is similar to ESD of pre-trained gestures, it is concerned that the accuracy of added hand gestures is lower than the accuracy of the pre-trained gestures.

As future work, we planned to consider estimating the SHD from recorded bioacoustics signals.

REFERENCES

1. IDC Worldwide Quarterly Wearable Device Tracker, September 2017
10. Nohara, R., Endo, Y., Murai, A., Takemura, H. and Tada, M., MoCap driven contact analysis with individual finite element hand model synthesized from a small number of measured dimensions, the Robotics and Mechatronics Conference 2017 (ROBOMECH 2017), Hukushima, Japan, 10-13 May, (2017).