Signal Classification Using Deep Learning

Conference Paper · July 2019

DOI: 10.1109/SENSORSNANO44414.2019.8940077

2 authors:

Hiromitsu Nishizaki
University of Yamanashi
93 PUBLICATIONS **394** CITATIONS
SEE PROFILE

K. Makino
University of Yamanashi
77 PUBLICATIONS **132** CITATIONS
SEE PROFILE

Signal Classification Using Deep Learning

Hiromitsu Nishizaki
Graduate School of Interdisciplinary Research
University of Yamanashi
4-3-11 Takeda, Kofh-shi, Japan
hnishi@yamanashi.ac.jp

Koji Makino
Graduate School of Interdisciplinary Research
University of Yamanashi
4-3-11 Takeda, Kofh-shi, Japan
kohjim@yamanashi.ac.jp

***Abstract*—Internet-of-Things (IoT) devices have rapidly become important in understanding conditions in an environment. The sensed data from an IoT (or sensor) device generally form a time sequential signal where the values vary with time. This study describes time sequential signal processing using a recurrent-based neural network and particularly focuses on two sorts of signal classification tasks: a sound classification and a tennis swing motion classification. We will introduce these classification tasks and their evaluation results using recurrent neural networks. The experimental results show that the recurrent neural networks could well classify the signals. Moreover, the bi-directional analysis is critical to achieving high-performance classification.**

***Index Terms*—deep learning, neural network, signal processing, signal classification**

I. Introduction

We have recently become capable of obtaining the sensing data from various Internet-of-Things (IoT) devices in real time and storing them to storage. The sensing (signal) data from these devices are then analyzed, and the analyses results are utilized for various objects. Deep learning framework has recently received much attention. Deep learning is a very effective technology for understanding various sorts of data. Object recognition [1], [2] on a photo/movie, speech recognition of an utterance [3], and machine translation [4] have been developed by using the deep learning framework. Many research papers on deep learning are published every year.

This study will focus on signal processing in a deep learning framework among various media. We particularly deal with two types of signal processing tasks: sound classification and tennis swing motion classification.

Sound data are generally made by catching an oscillation wave that travels through the air from a sound source, such as human mouth(s) using a microphone sensor device. The sensed analog signal by a microphone is translated into a digital signal in a suitable sampling frequency and a quantization bit rate. Fig. 1 shows an example of a sound waveform. Similar to sound, the sensing data from an IoT device, such as an electroencephalograph and a thermometer, are time sequential. Therefore, these sensed data can be represented as Fig. 1.

A signal waveform is characterized by the wave contour at a certain time. In other words, the waveform varies depending on the time. Therefore, when classifying a signal waveform, we should consider the continuous change of the waveform with time. To deal with sound classification with deep learning, we should adopt a neural network that can consider a time sequential change. The recurrent neural network (RNN) is very suitable in dealing with time sequential data, such as signal waveform.

Therefore, many studies deal with signal processing using RNN [5]–[9]. The audio processing research field focuses on the acoustic event detection tasks. For example, the Detection and Classification of Acoustic Scenes and Events (DCASE) challenges[1] have been held every year since 2014. Many techniques with RNN were proposed in the previous DCASE workshops [8], [9]. Meanwhile, in the music information processing research field, the music classification task [7], [10] is a well-known signal classification task. Many RNN-based approaches are proposed for these tasks [11], [12]. Apart from that, many tasks have been proposed for signal processing, and various sorts of datasets on signals have been released on Keggle[2] or other websites.

We will pick up the following two signal-classification tasks: sound classification and tennis swing motion classification. We will also propose the usage of RNN-based approaches to deal with these signal classification tasks herein.

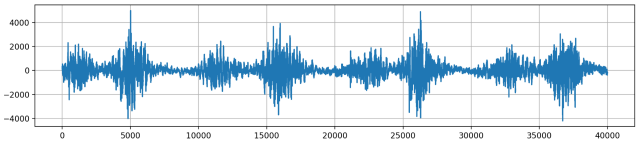

Fig. 1. Example of a sound waveform.

II. Snack Sound Classification

A. Dataset

The first signal processing task was the sound classification (snack name identification) task. We focused on a small dataset, in which a set of sounds recorded by shaking a bag of snack was included. This dataset included six types of snack sound with five sorts of microphone devices. Fig. 2 shows the six sorts of popular Japanese snacks used in the dataset. All of them are sold in Japan.

Table I lists the microphone devices used for recording the snack sounds. The duration of each recorded snack

[1]http://dcase.community
[2]https://www.kaggle.com/datasets

978-1-5386-5619-8/19/$31.00 ©2019 IEEE

Fig. 2. Six sorts of snacks in the dataset for the sound classification task. Top left: Tongari Corn; top middle: Kappa Ebisen; top right: Babystar; bottom left: Sapporo Potato Vegetable; bottom middle: Sapporo Potato BBQ; and bottom right: Calbee Potato Chips.

sound by a microphone device was 300 s. Each sound was segmented into 150 WAVE files, with each file having a 2 s duration. The sampling frequency and the quantization bit rates were 44.1 kHz and 16 bits/sample, respectively. A total of 900 WAVE files for each microphone device were used for the neural network training. For the evaluation, we also prepared 60 s of snack sound for each snack and microphone device. Each sound was segmented into 30 WAVE files for testing.

This snack sound classification task was tough for humans. Fig. 3 shows examples of the power spectrograms of the two types of snack sounds. As shown in Fig. 3, we can hardly find the characteristic differences between them.

TABLE I. FIVE SORTS OF MICROPHONE DEVICES

Mic. ID	Mic. type	Model number
01	Microphone	JVC-KENWOOD MZ-V8
02	USB-mic	BUFFALO BSHSM05KB
03	Laptop built-in mic	FMVS90PWD1
04	Voice recorder	Olympus DS-850
05	Smartphone	VAIO VPA0511S

B. Neural network model

We used an RNN for the snack sound classification. Fig. 4 depicts the architecture of the RNN-based neural network.

The neural network was based on a single layer of long short-term memory (LSTM) [14]. An LSTM is well-known to be effective in classifying the time sequential data because it can keep a long-span history compared to a normal RNN.

Before inputting the sound waveform into the NN, the sound waveform was made to undergo a short-time Fourier transformation with a 1024-point window width to extract a 512-dimensional power spectrogram vector. Fig. 5 shows the pre-processing procedure of a sound signal using short-time Fourier transformation.

An input data (i.e., set of 512-dim at time i) were input to the LSTM layer. The LSTM layer output the 512-dimension

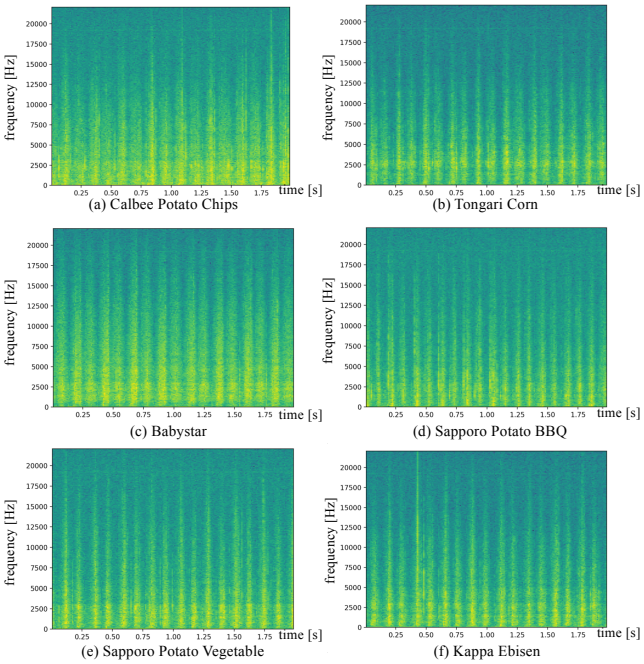

Fig. 3. Power spectrograms of the two sorts of snack sounds.

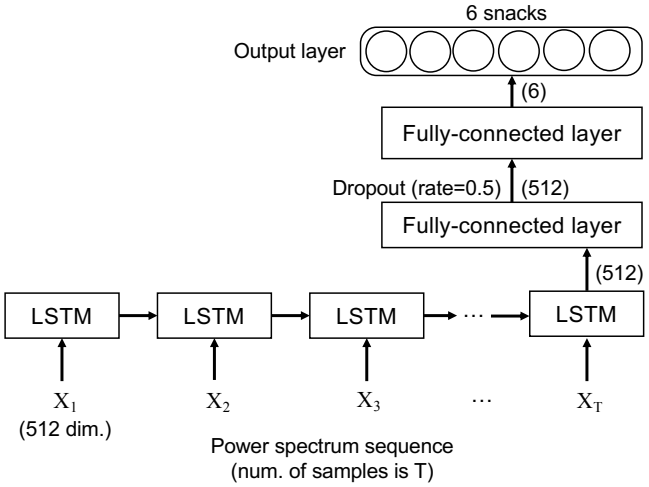

Fig. 4. Architecture of the RNN-based neural network for the snack sound classification task.

vector in a power spectrum vector at time T. They were also input to the first fully-connected (FC) layer. The first FC layer output the 512-dimensional hidden vector before proceeding to the second FC layer. The number of output nodes was six because the number of snacks was six. The dropout process [15] with 0.5 of dropout ratio was adapted at the outputs of both the LSTM and the first FC layer. Table II presents the training condition of the neural network.

The neural network model was implemented using Chainer[3] ver.5.3.0, which is a Python framework for deep learning.

[3]https://chainer.org/

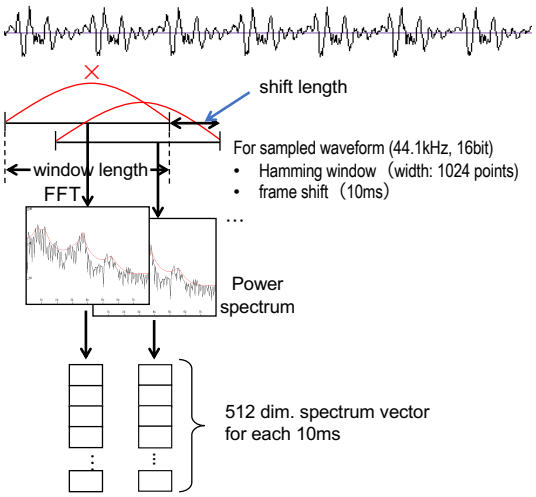

Fig. 5. Pre-process for the sound signal waveform.

TABLE II. TRAINING CONDITION OF THE NEURAL NETWORK

Mini-batchsize	16
Num. of epochs	50
Activation at the 1st FC	ReLU [16]
Batch normalization [13]	Yes (the LSTM and the 1st FC)
Loss func.	Softmax cross entropy
Optimaizer	MomentumSGD
Learning rate	0.001

C. Classification result

Table III shows the accuracy rates of the snack sound classification. Table III shows the confusion matrix between the microphone devices. For example, when the neural network model was trained from the sound signals recorded by microphone 01, the accuracy for the sound signals recorded by the same microphone 01 (matched condition) was 85.7%. However, the unmatched condition to the microphone environments drastically created damages to the classification accuracies. Under the matched condition of the microphone environments, the RNN-based neural network achieved the high performances of sound classification. In addition, the accuracy rates improved more when all the recorded signals from all the microphone devices were used for model training.

TABLE III. SNACK SOUND CLASSIFICATION ACCURACY RATES [%]

Mic. ID	01 (test)	02 (test)	03 (test)	04 (test)	05 (test)
01 (train)	85.7	21.2	50.6	21.7	28.5
02 (train)	35.4	86.3	18.4	19.3	29.0
03 (train)	24.4	16.2	87.8	30.6	35.6
04 (train)	35.4	17.3	50.2	87.4	25.7
05 (train)	37.6	16.9	29.3	39.2	79.3
All devices	91.1	88.8	94.0	86.7	84.3

III. TENNIS SWING MOTION CLASSIFICATION

A. Dataset

The second task was the tennis swing motion classification. This task challenged to estimate those who swing a tennis racket. The swing data from a player were obtained by a three-dimensional (3D) accelerometer attached to the player's hand. The data format was X, Y, and Z directions from the output of the accelerometer and saved as the time sequential data. The sampling frequency rate was 40 Hz. Therefore, the sampled data were a 3D vector. Fig. 7 shows an example of a swing signal.

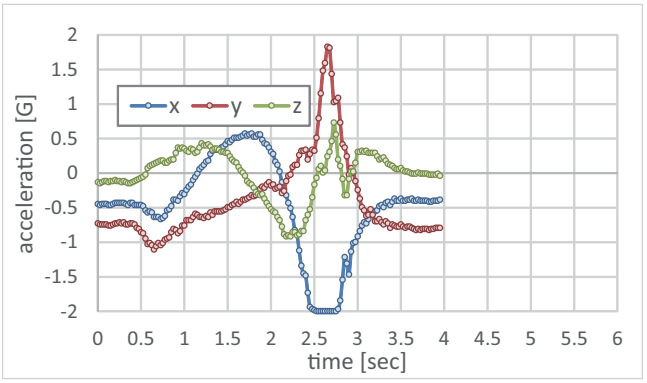

Fig. 6. An example of a tennis swing signal.

The number of subjects who swung a racket was 10. Each subject swung a racket for 50 times. The total number of swing signals was 500. The 500 swing signals were separated into 400 and 100 signals for training and testing, respectively.

B. Neural network model

We also used an RNN for the classification of the tennis swing motion. Fig. 7 shows the architecture of the RNN-based neural network, which was almost the same as the model for classifying a snack sound signal. The difference between the tennis swing model and the snack sound model was the utilization of a bi-directional LSTM. In addition, the input signal did not undergo any pre-processing like the Fourier transformation. The sampled raw data (sequence of 3D vectors) were directly inputted to the neural network.

The training conditions were almost the same as those in Table II. However, we did not perform any batch normalization processes. We also used an RNN for the classification of the tennis swing motion.

The neural network model was implemented using Chainer ver. 5.3.0.

C. Classification result

Table IV shows the classification results of the tennis swing signals. A∼J denotes the subject IDs. The table is represented as the confusion matrix, in which the subject ID on the horizontal axis depicts whose tennis swing signal was input to the neural network, while the IDs on the vertical axis denote the output from the neural network.

Table IV illustrates that most tennis swing signals can be correctly classified using the bi-directional RNN model. The accuracy rate for all the subjects' swing signals was 97% (97/100). We also applied the same neural network model (the uni-directional LSTM model) in Fig. 4 to the

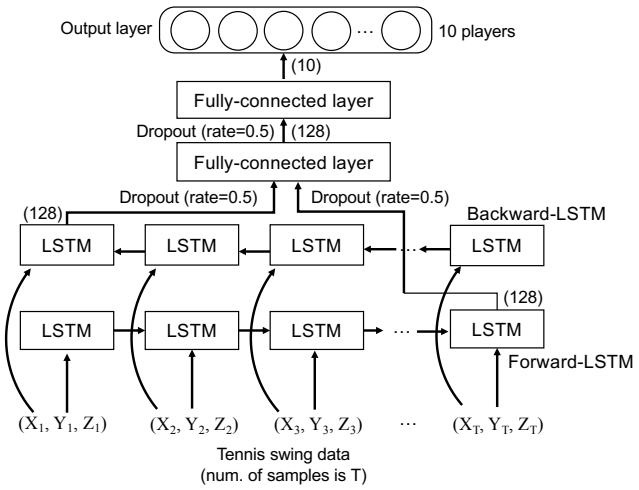

Fig. 7. Architecture of the RNN-based neural network for the tennis swing motion classification.

tennis swing classification task. The accuracy rate for all the subjects was 83% (83/100), which is substantially worse than that for the bi-directional LSTM model. The tennis swing classification was a competitively hard task compared to the snack sound classification; however, the bi-directional LSTM model can capture the characteristics of a tennis swing signal.

TABLE IV. CLASSIFICATION MATRIX AMONG SUBJECTS

Subject ID	A	B	C	D	E	F	G	H	I	J
A	9	0	0	0	0	0	0	1	0	0
B	0	9	0	0	0	0	1	0	0	0
C	0	0	10	0	0	0	0	0	0	0
D	0	0	0	10	0	0	0	0	0	0
E	0	0	0	0	10	0	0	0	0	0
F	0	0	0	0	0	10	0	0	0	0
G	0	0	0	0	0	0	9	1	0	0
H	0	0	0	0	0	0	0	10	0	0
I	0	0	0	0	0	0	0	0	10	0
J	0	0	0	0	0	0	0	0	0	10

IV. CONCLUSIONS

This study introduced the signal classification tasks using the deep learning framework. We showed that the recurrent-based neural network was very effective in understanding and classifying the signal. In particular, the bi-directional LSTM can realize a robust classification.

Although we dealt with the snack sound classification and the tennis swing classification, the RNN-based model can be widely applied to the time sequential data from various IoT devices. In the future work, we will develop an environment-understanding system for in-home [17] or forest environments [18] with IoT devices and the deep learning framework.

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant-in-Aid for Scientific Research (B) Grant Number 17H01977.

REFERENCES

[1] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, and Li Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009, pp. 248–255.

[2] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779—788.

[3] G. Hinton et al., "Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups," IEEE Signal Process. Mag., vol. 29, no. 6, pp. 82–97, 2012.

[4] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," arXiv:1810.04805, 2018.

[5] J. Nam, K. Choi, J. Lee, S.-Y. Chou, and Y.-H. Yang, "Deep Learning for Audio-Based Music Classification and Tagging: Teaching Computers to Distinguish Rock from Bach," IEEE Signal Process. Mag., vol. 36, no. 1, pp. 41–51, 2019.

[6] J. Lee, J. Park, K. L. Kim, and J. Nam, "Sample-level deep convolutional neural networks for music auto-tagging using raw waveforms," in the Sound and Music Computing Conference, pp.220–226, 2017.

[7] G. Tzanetakis, and P. Cook, "Musical genre classification of audio signals," IEEE Trans. Speech Audio Processing, vol. 10, no. 5, pp. 293—302, 2002.

[8] I. Himawan, M. Towsey, and P. Roe, "3D convolutional recurrent neural networks for bird sound detection," in Workshop on Detection and Classification of Acoustic Scenes and Events (DCASE2018), 2018.

[9] Y. Guo1 M. Xu, J. Wu, Y. Wang, and K. Hoashi, "Multi-scale convolutional recurrent neural network with ensemble method for weakly labeled sound event detection," in Workshop on Detection and Classification of Acoustic Scenes and Events (DCASE2018), 2018.

[10] M. Defferrard, K. Benzi, P. Vandergheynst, and X. Bresson, "FMA: A dataset for music analysis," in 18th International Society for Music Information Retrieval Conference, 2017, pp. 316–323.

[11] K. Choi, G. Fazekas, K. Cho, and M. Sandler, "Convolutional recurrent neural networks for music classification," in the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2017, pp. 2392–2396.

[12] J. Dai, S. Liang, W. Xue, C. Ni, and W. Liu, "Long Short-term Memory Recurrent Neural Network based Segment Features for Music Genre Classification," in the 10th International Symposium on Chinese Spoken Language Processing (ISCSLP), 2016, pp. 1–5.

[13] S. Ioffe and C. Szegedy, "Batch normalization: Accelarating deep network training by reducing internal covariate shift," arXiv:1502.03167v3, 2016.

[14] S. Hochreiter and J. Schmidhuber, "Long Short-Time Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[15] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," J. Mach. Learn. Res., vol. 15, no. 1, pp. 1929—1958, 2014.

[16] X. Glorot, A. Bordes and Y. Bengio, "Deep Sparse Rectifier Neural Networks," in 14th International Conference on Artificial Intelligence and Statistics (AISTATS), 2011, vol. 15, pp. 315–323.

[17] A. S. Abdull Sukor, A. Zakaria, N. A. Rahim, L. M. Kamarudin, R. Setchi and H. Nishizaki, "A hybrid approach of knowledge-driven and data-driven reasoning for activity recognition in smart homes," J. Intell. Fuzzy Syst., vol. 36, no. 5, pp. 4177—4188, May 2019.

[18] R. Gunasagaran,L. M. Kamarudin,E. Kanagaraj, A. Zakaria and A. Y. M. Shakaff, "Internet of Things: Solar power under forest canopy," in the 2016 IEEE Student Conference on Research and Development (SCOReD), 2016.