論文 / 著書情報
Article / Book Information

論題(和文)	
Title(English)	Alzheimer's Disease Prediction Using Audio Gated Convolutional Neural Network
著者(和文)	Warnita Tifani, 井上 中順, 篠田 浩一
Authors(English)	Tifani Warnita, Nakamasa Inoue, Koichi Shinoda
出典(和文)	2018年 秋季研究発表会 講演論文集, , , pp. 1223-1224
Citation(English)	ASJ 2018 Autumn Meeting, , , pp. 1223-1224
発行日 / Pub. date	2018, 8

Alzheimer's Disease Prediction Using Audio Gated Convolutional Neural Network *

○ Tifani Warnita, △ Nakamasa Inoue, Koichi Shinoda (Tokyo Institute of Technology)

1 Introduction

Alzheimer's Disease (AD) is the common cause of dementia caused by neurodegenerative disease (1). Numerous automatic approaches to predict AD incorporate linguistic information (2; 3) since language deficits become perceivable symptoms in the patients. However, applying the approaches to other languages might be difficult, especially to the low resource languages. In this paper, we present our language-independent approach of detecting AD by utilizing utterance-level prediction based on the paralinguistic cues emerged. For this task, we employed convolutional layers with gating mechanism.

2 Proposed Method

2.1 Features

Some studies addressed the problem of detecting AD using the linguistic information given by patients such as taking into account the syntactic complexity, vocabulary richness, information content, and repetitiveness (2). However, AD detection task can also be considered close to paralinguistic aspects, i.e. aspects related to non-word things in speech communication of humans. Therefore, we used INTERSPEECH 2010 Paralinguistic Challenge Features (IS10) (4) feature set. The feature set consists of 76 low-level descriptors (LLDs) such as pitch, energy, jitter, and shimmer. After being applied by statistical functionals, we get the total of 1582 features for one utterance.

2.2 Gated Convolutional Neural Network

In this paper, Gated Convolutional Neural Network (GCNN) is employed. A gate can manage information flow to the succeeding layers. The gate has the advantage of avoiding the vanishing gradient problem (5).

We gave the input $X \in \mathbb{R}^{T \times F}$, where F and T represent the dimension of LLDs of each utterance segment and the number of time frames respectively. The sliding window length is N, spanning over the feature dimension. We used a sigmoid function as the activation function g, which is multiplied by a linear gate. The output at time i is then defined as,

$$y_i = \left(\sum_{f=1}^{F} \sum_{n=1}^{N} v_{f,n} x_{f,i-n} + e \right) \cdot g \left(\sum_{f=1}^{F} \sum_{n=1}^{N} w_{f,n} x_{f,i-n} + b \right), \quad (1)$$

where $v_{c,d}$ and $w_{c,d}$ represent the (c,d) elements of the matrices V and W respectively. The $V, W \in \mathbb{R}^{K \times F}$ and $e, b \in \mathbb{R}$ are the kernel weight matrices and biases.

The visualization of our GCNN can be seen in Figure 1. In the figure, one gated convolution layer lies

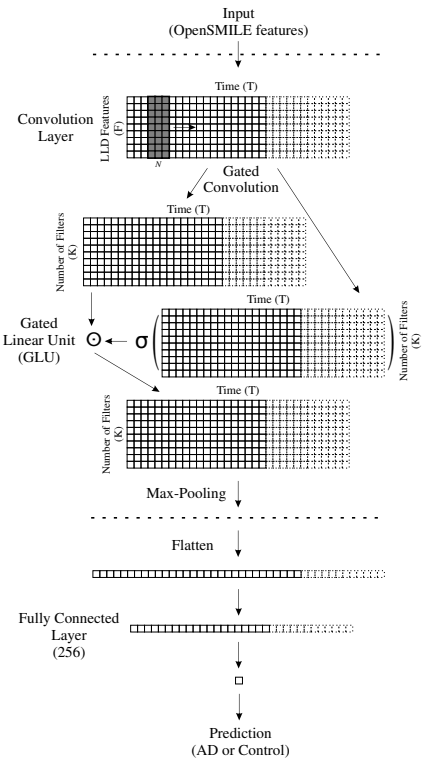

Fig. 1 GCNN with depth = 1. The kernel window in the convolutional layer is colored in gray.

* 音響信号を 入力とするゲート 付き畳み込みニューラルネットワークを用いたアルツハイマー病予測, Tifani Warnita, 井上 中順, 篠田 浩一 (東京工業大学)

Table 1 Result of the utterance-level and subject-level prediction represented in the average accuracy (%) from 10-fold cross validation.

Model	Utterance (%)	Subject (%)
SMO	67.5	66.0
GCNN-6	65.1	72.2
GCNN-8	66.3	**73.6**
GCNN-10	65.2	69.8

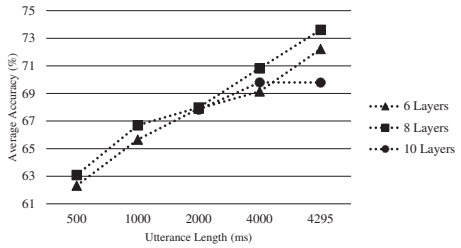

Fig. 2 GCNN with different utterance length.

between the dotted horizontal line which is followed by the max-pooling layer. Deeper networks consist of more gated convolution layers. The final verdict for each subject is based on the majority voting from the utterance-level prediction.

3 Experiments

3.1 Settings

For evaluating our proposed method, we used DementiaBank - Pitt Corpus (6), a multimedia database of conversations with people having dementia, specifically when the patients undergo a picture description task. From Pitt Corpus, we only selected data from the control subjects (people not having dementia) and also data from AD patients. This is resulting in 488 audio conversations (255 AD, 233 control) from 267 subjects (169 AD, 98 control).

3.2 Results

The comparison of the standard CNN and the gating mechanism is shown in Table 1. From the table, we got the best result of 73.6% using the GCNN with the depth of 8. This is better than the conventional SMO by 7.6 points.

We also investigated the importance of utterance length information by trying a set of different segment length L (500ms, 1000ms, 2000ms, 4000ms, and 4295ms). We segmented each subject data into segments with a predetermined length L. For data with length less than L, zero padding is added.

Figure 2 shows that shorter segment length yields worse results. However, we can still get close results using the utterance length of 4000ms with 69.1%, 70.8%, and 69.8%, compared to the cases when we use the oracle utterance length, 72.2%, 73.6%, and 69.8% for 6, 8, and 10 layers respectively. For 10 layers and the length of 4000ms, we got the similar result (69.8%) as the oracle utterance length. This result suggests that we can use the approach even if we do not have the transcription.

4 Conclusion

We proposed the combination of GCNN and paralinguistic features as a non-linguistic approach for detecting AD. By using only audio data, we got the best result of 73.6%. Future work will include the evaluation of the proposed method with other languages to further show its flexibility as language-independent approach.

Acknowledgements This work was supported by JSPS KAKENHI 16H02845 and by JST CREST Grant Number JPMJCR1687, Japan.

References

[1] A. Burns and S. Iliffe, "Alzheimer's disease," *BMJ*, vol. 338, 2009.
[2] K. C. Fraser, J. A. Meltzer, and F. Rudzicz, "Linguistic features identify Alzheimer's disease in narrative speech," *Journal of Alzheimer's Disease*, vol. 49, no. 2, pp. 407–422, 2000.
[3] S. Wankerl, E. Nöth, and S. Evert, "An N-gram based approach to the automatic diagnosis of Alzheimer's disease from spoken language," in *Proc. INTERSPEECH 2017*, 2017.
[4] B. Schuller, S. Steidl, A. Batliner, F. Burkhardt, L. Devillers, C. Müller, and S. Narayanan, "The INTERSPEECH 2010 paralinguistic challenge," *Proc. INTERSPEECH 2010*, pp. 2794–2797, 2010.
[5] Y. N. Dauphin, A. Fan, M. Auli, and D. Grangier, "Language modeling with gated convolutional networks," pp. 933–941, 2017.
[6] J. T. Becker, F. Boiler, O. L. Lopez, J. Saxton, and K. L. McGonigle, "The natural history of Alzheimer's disease: description of study cohort and accuracy of diagnosis," *Archives of Neurology*, vol. 51, no. 6, pp. 585–594, 1994.