Journal of Digital Convergence
Vol. 19. No. 12, pp. 353-358, 2021

ISSN 2713-6434 / eISSN 2713-6442
https://doi.org/10.14400/JDC.2021.19.12.353

Wav2vec을 이용한 오디오 음성 기반의 파킨슨병 진단

윤희진
장안대학교 IT학부 소프트웨어융합과 교수

Diagnosis of Parkinson's disease based on audio voice using wav2vec

Hee-Jin Yoon
Professor, Department of Software convergence, Jangan University

요 약 노년기에 접어들면서 알츠하이머 다음으로 흔한 퇴행성 뇌 질환은 파킨슨병이다. 파킨슨병의 증상은 손 떨림, 행동의 느려짐, 인지기능의 저하 등 일상생활의 삶의 질을 저하시키는 요인이 된다. 파킨슨병은 조기진단을 통하여 병의 진행 속도를 늦출 수 있는 질환이다. 파킨슨병의 조기진단을 위해 오디오 음성 파일 입력으로 wav2vec을 이용하여 특징을 추출하고 딥러닝(ANN)으로 파킨슨병의 유무를 진단하는 알고리즘을 구현하였다. 오디오 음성 파일을 이용하여 파킨슨병을 진단하는 실험 결과 정확도는 97.47%로 나타났다. 기존의 뉴럴네트워크를 이용하여 파킨슨병을 진단하는 결과보다 좋은 결과를 나타냈다. 오디오 음성 파일을 wav2vec 이용으로 간단하게 실험을 과정을 줄일 수 있었으며, 실험 결과 향상된 결과를 얻을 수 있었다.

주제어 : 파킨슨병, 오디오음성, wav2vec, 딥러닝, 분류

Abstract Parkinson's disease is the second most common degenerative brain disease after Alzheimer's in old age. Symptoms of Parkinson's disease are factors that reduce the quality of life in daily life, such as shaking hands, slowing behavior and cognitive function. Parkinson's disease that can slow the progression of the disease through early diagnosis. To diagnoze Parkinson's disease early, an algorithm was implemented to extract features using wav2vec and to diagnose the presence or absence of Parkinson's disease with deep learning(ANN). As a results of the experiment, the accuracy was 97.47%. It was better than the results of diagnosing Parkinson's disease using the existing neural network. The audio voice file could simply reduce the experiment process and obtain improved results.

Key Words : Parkinson's disease, human audio voice, wav2vec, deep learning, classification

1. 서론

파킨슨병은 알츠하이머병 다음으로 흔한 퇴행성 뇌 질환으로 노년기에 접어들수록 발병률이 증가한다. 파킨슨병으로 인해 노년기의 삶의 질을 낮추고, 치료와 요양으로 인한 사회적 경제적 손실이 크다[1-3]. 파킨슨병은 뉴런에 생긴 미토콘드리아 폐기물의 처리를 제어하는 신호에 이상이 생겨 폐기물이 과도히 쌓여 뇌 흑질의 도파민계 신경이 파괴됨으로 움직임에 장애가 나타나는 질환이다[4]. 도파민은 뇌의 기저핵에 작용하여 몸을 정교하게 움직일 수 있도록 하는 신경전달계 물질이다. 파킨슨병의 증상은 다양하게 나타나는데 손 떨림, 행동의 느려짐, 인

*This paper was supported by jangan University Research Grant in 2021.
*Corresponding Author : Hee-Jin Yoon(hjyoon@jangan.ac.kr)
Received November 5, 2021 Revised November 30, 2021
Accepted December 20, 2021 Published December 28, 2021

지기능 저하, 수면장애 등 자세가 불안정하며 근육이 경직되는 증상으로 나타난다. 이런 증상으로 인해 독립적인 생활에 어려움을 느끼는 파킨슨병은 삶의 질도 저하시킨다[5-6]. 파킨슨병은 조기 발견하여 조기 치료를 한 경우엔 생명에 밀접한 연관은 없으며 진행 속도도 늦출 수 있다. 파킨슨병의 진단 과정은 전문의가 병력을 듣고 임상적인 증상을 진찰하는 것이다[7]. 운동 장애 증상이 나타나고 난 후엔 이미 40~60% 정도의 도파민 신경의 손실이 나타난 후 알 수 있어 조기 치료를 할 수 있는 기회를 놓치게 된다[8]. 파킨슨병은 무엇보다도 조기진단이 중요하다. 파킨슨병 환자들은 마비말장애로 인해 호흡, 조음, 발성에 영향을 받아 거친 음성, 성대 떨림 등으로 음성장애를 나타낸다[9,10].

본 연구에서는 파킨슨병의 조기진단을 위해 음성 오디오 파일을 wav2vec을 이용하여 특징을 선택한 후 딥러닝을 이용하여 파킨슨병을 진단하는 알고리즘을 제안한다. wav2vec은 많은 양의 음성 오디오 파일을 비지도학습(unsupervisor)로 학습하며 CNN(Convolution Neural Network)으로 구성되어있다, 또한 학습을 통해 특징추출을 한다.[11,12] 실험 데이터로 파킨슨병을 가진 사람의 음성 오디오 파일과 정상인의 오디오음성 파일을 wav2vec으로 처리한 후 딥러닝(ANN, Artificial Neural Network)으로 파킨슨병의 유무를 진단하는 알고리즘을 제안하였다. 논문의 구성은 2장 관련 연구로 wav2vec에 대해 설명한다. 3장은 wav2vec을 이용한 파킨슨병 유무 진단에 대한 실험데이터와 알고리즘 구현, 4장 성능평가 5장 결론으로 구성되었다.

2. wav2vec

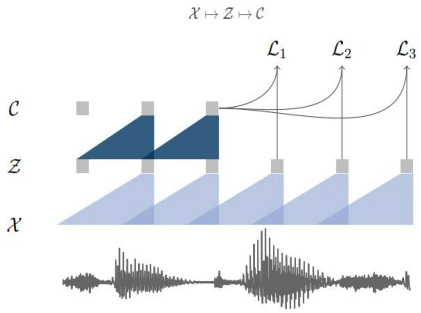

Fig. 1. structure of wav2vec [12]

데이터에 라벨 없이 자기 자신의 특성을 배우는 학습이 자기지도학습(Self-Supervised Learning)이다. wav2vec는 뉴럴네트워크 기반에서 음성 오디오를 입력으로 모델을 최적화하여 Pre-training 한 후 입력된 음성 오디오 파일에서 특징을 추출하고 샘플을 예측할 수 있다[13]. wav2vec은 모든 언어의 음성에서 특징을 추출한다. Fig. 1는 wav2vec을 나타낸 것이다.

wav2vec의 구조를 encoder network와 context network로 나눌 수 있다. Fig.2에서 X는 음성오디오 데이터가 입력 되어진 영역으로 encoder(CNN network)에 특정 vector로 변환한다. Z는 숨겨진(latent)영역으로 학습을 시킨다, C는 (context) 영역이며 오디오 음성 데이터의 공통성을 갖고 있는 mutual information을 최대한 할 수 있다.

- $f : \chi \rightarrow Z$ encoder network
- $g : Z \rightarrow C$ context network

파킨슨병의 유무 진단을 위해 파킨슨병을 가진 사람과 정상인의 오디오 음성데이터의 특징추출을 위해 wav2vec을 사용하였다.

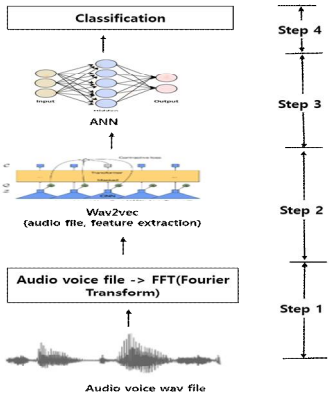

Fig. 2. Structure of Process

3. 실험 및 결과

3.1 실험데이터

음성데이터를 wav2vec을 이용하여 파킨슨병을 진단하는 알고리즘을 구현하는데 사용된 데이터는 이탈리안 50명의 오디오음성 데이터로 2그룹을 사용하였다.

첫 번째 그룹은 건강한 60대 이상의 건강한 사람으로 22명의 음성이다. 성비의 구성은 남성 10명 여성 12명의 데이터이다. 두 번째 그룹의 데이터는 파킨슨병을 갖고 있는 사람의 오디오 음성데이터로 남성 9명 여성 28명의 음성을 실험데이터로 사용하였다[13]. 오디오음성 녹음은 다음 Table. 1과 같은 조건으로 다양한 모드와 실행시간으로 녹음한 데이터로 786개의 데이터를 생성하였다.

Table. 1 structure of dataset[13]

sortation	description
a	2 readings of a phonemically alanced text spaced by a parse(30 sec)
b	execution of the syllable 'a'(5 sec), pause(20 sec), execution of the syllable 'ta'(5 sec)
c	2 phonation of the vocal 'a'
d	2 phonation of the vocal 'e'
e	2 phonation of the vocal 'i'
f	2 phonation of the vocal 'o'
g	2 phonation of the vocal 'u'
h	reading of some phonemically balanced words, pause(1 min), and reading of some phonemically balanced phrases

단어 구성은 이탈리아어로 *pipa, buco. topo, dado, casa, gatto, filo, waso, muro, neve, rete, zero, calendariom gioso, momotono, sdraio, sbrigo, frate, spesa, stufa, classe, flotta, gnomo prestigioso* 단어로 이루어졌다. 다음 Table 2는 60세 이상의 건강한 사람의 데이터 파일 구성을 나타낸 것이다.

Table 3은 28명의 파킨슨병 환자들의 음성 녹음 파일의 형태를 나타낸 것이다. Table 2와 Table 3에서 보는 것과 같이 데이터는 세 부분으로 나누어 실험데이터를 녹음하였다. time 1과 time 2는 텍스트를 읽는데 time 3은 단어를 읽는 것 그리고 각각 time에 대한 CPS(Characters per second)도 녹음 파일로 구성되어있다.

3.2 제안 알고리즘

파킨슨병을 가진 사람과 정상인의 오디오 음성데이터로 wav2vec을 이용하여 특징추출을 하고 딥러닝을 이용하여 파킨슨병을 진단하는 알고리즘의 전체적인 흐름은 다음 Fig. 2와 같다. 처리단계는 크게 4단계로 나누었다.

- step 1. 실험을 위해 오디오 음성데이터 파일의 구성은 이탈리안 50명을 정상인 그룹과 파킨슨병을 가진 사람

Table 2. Experimental phase of Elderly Healthy[13]

person		text1 Reading		text2 Reading		Text3 Reading	
sex	age	time1	CPS1	time2	CPS2	time3	CPS3
F	69	57.12	9.07	49.99	10.36	47.11	5.96
F	62	100.95	4.72	77.26	5.94	66.4	3.92
F	65	70.87	7.31	57.71	8.68	43.16	6.51
M	68	59.55	8.70	55.08	9.33	43.47	6.46
F	68	55.97	9.25	53.01	9.77	51.98	5.41
M	70	–	–	–	–	64.5	3.77
M	60	60	8.33	54.3	9.36	38.23	7.22
F	60	59.49	7.88	54.97	8.53	43.96	6.39
F	61	66.92	7.74	53.89	9.61	43.77	6.42
M	68	58.92	8.38	56.31	8.77	33.8	8.31
F	63	70.3	6.69	71.26	6.60	44.8	6.27
M	68	58.57	8.74	51.26	9.85	42.02	6.69
F	69	64.45	7.87	54.34	9.53	33.26	8.45
M	76	56.09	9.24	53.15	9.75	49.15	5.72
M	77	67.97	7.59	59.75	8.27	67.33	4.17
F	63	75.88	6.68	62.73	7.94	54.65	5.03
M	69	65.88	7.86	63.09	8.00	38.59	7.28
F	61	59.09	8.55	53.51	7.61	56.81	4.95
F	70	148.83	3.00	104.51	4.37	64.67	4.04
M	62	92.7	4.99	70.11	7.03	49.26	5.58
M	75	68.74	7.26	65.43	7.90	41.19	6.82
F	72	67.37	7.53	67.35	7.53	67.64	4.15

Table 3. Experimental phase of people with Parkinson's se Reading [13]

person		text1 Reading		text2 Reading		Text3 Reading	
sex	age	time1	CPS1	time2	CPS2	time3	CPS3
F	63	//	//	//	//	60.64	4.63
M	50	71.73	7.22	53.82	9.62	39.75	7.07
F	61	53.40	9.70	51.4	10.08	49.25	5.71
M	68	84.05	6.16	63.32	8.18	56.66	4.96
F	40	60.92	7.76	52.4	9.89	54.58	5.15
M	65	52.40	9.89	50.23	10.31	40.2	6.99
M	73	79.35	6.53	71.22	7.27	69.62	4.04
M	56	86.81	5.97	66.64	7.77	58.7	4.79
M	77	64.75	8.00	60.9	8.51	50.41	5.57
M	71	59.84	8.66	56.76	9.13	53.52	5.25
F	71	66.22	7.82	48.38	10.71	55.56	5.06
M	71	49.95	10.37	45.35	11.42	36.86	7.62
M	73	70.98	7.30	65.07	7.96	56.87	4.94
M	75	62.20	8.33	56.58	9.16	52.98	5.30
M	68	85.37	6.07	63.35	8.18	48.67	5.77
M	71	62.33	8.31	52.56	9.86	66.38	4.23
F	65	242.50	2.14	180.09	2.88	167.83	1.67
F	80	169.29	3.06	//	//	101.19	2.78
M	73	66.90	7.74	63.5	8.16	53.04	5.30
M	70	65.46	7.91	60.4	8.58	48.25	5.82
F	67	79.30	6.53	73.8	7.02	71.67	3.92
F	54	55.00	9.42	49.8	10.40	54.7	5.14
F	78	163.60	3.17	//	//	108.3	2.59
M	72	117.60	4.40	98.8	5.24	87.51	3.21
M	65	164.10	3.16	//	//	151.3	1.86
M	65	233.00	1.76	//	//	217.3	0.96
M	70	112.00	4.63	106.6	4.86	68.31	4.11
M	70	68.30	7.58	61.47	8.43	64.5	4.36

그룹 두 그룹으로 나눈 786개의 오디오 음성데이터 파일을 사용하였다. 본 실험에서 처리 속도를 위해 786개의 데이터를 푸리에변환을 하였다.

- step 2. step 1에서 처리된 오디오음성 파일을 wav2vec을 이용하여 특징 추출한다.
- step 3. 딥러닝(ANN)을 이용하여 50명의 오디오 음성 파일을 학습데이터와 테스트 데이터로 나누어 학습하고 테스트한다. 오디오 파일 786개의 실험데이터 중 학습데이터는 628개의 오디오 파일로 사용하였고, 158의 오디오 파일을 테스트 데이터로 사용하였다.
- step 4. 정상인과 파킨슨병 환자를 분류한다.

푸리에변환은 오디오음성 파일을 다양한 주파수 성분으로 분해하여 표현한다. 푸리에변환은 음성, 이미지 분석 등 여러 분야에서 사용된다. wav2vec의 오디오 음성 데이터의 처리 속도를 높이기 위해 푸리에변환을 하였다.

Fig. 3는 정상인의 오디오음성 파일 중 1분 정도의 긴 파일에 대한 FFT(fast Fourier transform)를 나타낸 그림이다.

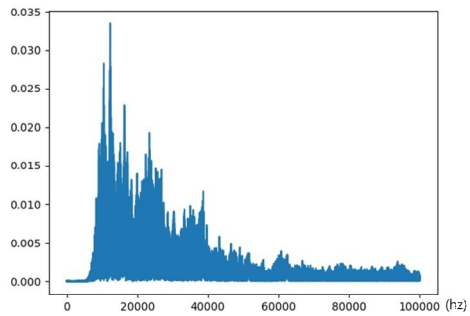

Fig. 3. FFT of Normal_long

Fig. 4는 파킨슨병을 갖고있는 오디오음성 파일 중 1분 정도의 긴 파일에 대한 FFT를 나타낸 그림이다.

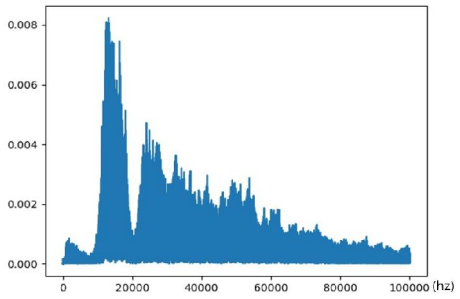

Fig. 4. FFT of diseased_long

4. 성능평가

wav2vec을 이용하여 오디오 음성데이터 기반에서 파킨슨병을 진단하는 실험에서 오디오 음성데이터 파일을 푸리에로 변환하여 wav2vec으로 특징을 추출하여 파킨슨병을 가진 사람과 정상인을 분류하였다. 700hz에서 97.4% 분류를 할 수 있었다. 기존의 순환신경망(RNN, recurrent neural network)를 이용한 방법[14]에 비해 오디오음성 파일로 간단한 절차로 향상된 결과를 얻을 수 있었다. 실제값과 예측값의 차이를 MSE(Mean Squared Error)를 이용하여 실험 결과를 확인해 보았다.

$$MSE = \frac{1}{N}\sum_{i=1}^{n}(yi - \hat{y})^2$$

Fig. 5는 700hz의 97.47%의 정확도를 나타낼 때의 MSE를 나타낸 손실함수 그래프이다.

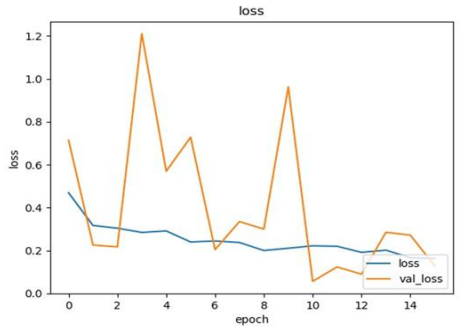

Fig. 5. Loss function graph of 97.47% accuracy

5. 결론

본 논문에서 wav2vec을 이용하여 음성의 특징을 추출하고 딥러닝을 이용하여 파킨슨병의 유무를 진단하는 알고리즘을 제안하였다. 정확도는 97.47% 분류의 정확도는 나타냈다. wav2vec은 특정된 언어에 대해 음성 특징추출을 하는 것이 아니라 모든 음성에 대해 특징추출이 가능하다. 향후 wav2vec을 이용하여 여러 딥러닝의 알고리즘을 설계하고, 파킨슨병 뿐 아니라, 다양한 분야에 오디오 음성데이터 파일을 적용하여 연구할 것이다.

REFERENCES

[1] kim, Dong Won, Bae, Eun sook,(2015), Factors Affecting Caregiver Burden in caregivers of Partients with Parkin's Disease, Korean Journal of Adult Nursing, Vol.27 No.3, 283-293
DOI:10.7475/kjan.2015.27.3.283

[2] Doyeon Lee, Yoseob Heo, Keunhwan Kim. (2020). Analysis of Technology Trends and Technology Covergence for Parkinson's Disease Therapeutics : Based on Global Patent Information . Journal of the Korea Convergence Society, 11(3), 135-143.

[3] Hyo-Lyun Roh, Se-Hyun Jang. (2021). Meta-analysis of the Effects of Untact Convergence Exercise Programs on Balance, Gait, and Falls Efficacy of Parkinson's Disease Patients . Journal of the Korea Convergence Society, 12(5), 39-50.

[4] Shin, Hee-Baek, Shim, Hee-Jeong et (2018), Characters of voice quality on clear versus casual speech in individuals with Parkinson's disease, pISSN 2586-5854 Vol.10 No.2 00.77-84
DOI:10.13064/KSSS.2018.10.2.077

[5] Martinez-Martin P.(1998) An introduction to the concept of quality of life in Parkinson's disease. J Neural;245 Suppl 1:2-6
DOI:10.1007/p.100007733

[6] Sung Reul Kim, R.N., Sun Ju Chung, M.D., sung Young Hee, M.D(2005), Factors Related to Quality of Life in Patients with Parkinson's Disease, J Korean Neurol assoc vol.23..6.2005. p.770-775
DOI:jkna.org/upload/pdf/20050606.pdf

[7] Seuk Kyung Hong, M.D, Kyung Won Park, M.D, Jae Kwan Cha, M.D, Quality of Life in Patients with Parkinson's Disease, 20(3);227-233, ISSN 1225-7044
DOI:10.1016/j-parkreldis.2004.12.005

[8] Pino, C. Ozsanocak, E. Tripoliti, S. Thobois, P.L.Dowsey, P.Auzou (2004), Treatments for dysarthria in Parkinson's disease, Lancet Neurology,3, pp. 574-56.

DOI:10.1016/S1474-4422(04)00854-3

[9] J.R.Duffy, Motor Speech Disorders:substrates, differential diagnosis and management, St Louis:Mosby, 2005

[10] Byung-Chul Cho, Sooyoung Cheon, Kab-Nyun Kim, Hyun-Seung Yuk. (2018). A policy study for the voice recognition technology based on elderly health care. Journal of Digital Convergence, 16(2), 9-17.

[11] Aliaksei Kolesau and Dmitrij Sesok,(2020), Unsupervised Pre-Training for Voice Activation, Sciences 10(23):8643
DOI:10.3390/app10238643

[12] Steffen Schneider, Alexei Baevski, Ronan Collobert, Michael Auli(2019), wav2vec: UNSUPERVISED PRE-TRAINING FOR SPEECH RECOGNITION, computer Science v4
DOI:arXiv-1904.05862[cs.CL]

[13] Giovanni Dimauro(2019), Assessment of Speech Intelligibility in Parkinson's Disease Using a Speech-To-Text System, IEEE
DOI:10.21227/aw6b-tg17

[14] Seung-Su, Gee Yeun Kim, Bon Mi Koo,(2019), Parkinson's ddisease diagnosis using speech signal and deep residual gated recurrent network network, Acoustical society of korea, 308-313
DOI:10.7776/ASK.2019.38.3.308

윤 희 진(Hee-Jin Yoon) [정회원]

· 2001년 2월 : 동국대학교 컴퓨터공학 (공학석사)
· 2015년 8월 : 가천대학교 전자계산과 (공학박사)
· 2013년 3월 ~ 현재 : 장안대학교
· 관심분야 : 인공지능, 빅데이터
· E-Mail : hjyoon@jangan.ac.kr