

# textNAS의 다변수 시계열 데이터로의 적용 및 손동작 인식

김기덕\* · 김미숙 · 이학만

부산대학교

## TextNAS Application to Multivariate Time Series Data and Hand Gesture Recognition

Gi-duk Kim\* · Mi-sook Kim · Hack-man Lee

Pusan National University

E-mail : kimsjpk@naver.com / sook99@pusan.ac.kr / bruce@pusan.ac.kr

### 요 약

본 논문에서는 텍스트 분류에 사용된 textNAS를 다변수 시계열 데이터에 적용 가능하도록 수정하여 이를 통한 손동작 인식 방법을 제안한다. 이를 사용하면 다변수 시계열 데이터 분류를 통한 행동 인식, 감정 인식, 손동작 인식 등 다양한 분야에 적용 가능하다. 그리고 분류에 적합한 딥러닝 모델을 학습을 통해 자동으로 찾아줘 사용자의 부담을 덜어주며 높은 성능의 클래스 분류 정확도를 얻을 수 있다. 손동작 인식 데이터셋인 DHG-14/28과 Shrec'17 데이터셋에 제안한 방법을 적용하여 기존의 모델보다 높은 클래스 분류 정확도를 얻을 수 있었다. 분류 정확도는 DHG-14/28의 경우 98.72%, 98.16%, Shrec'17 14 class/28 class는 97.82%, 98.39%를 얻었다.

### ABSTRACT

In this paper, we propose a hand gesture recognition method by modifying the textNAS used for text classification so that it can be applied to multivariate time series data. It can be applied to various fields such as behavior recognition, emotion recognition, and hand gesture recognition through multivariate time series data classification. In addition, it automatically finds a deep learning model suitable for classification through training, thereby reducing the burden on users and obtaining high-performance class classification accuracy. By applying the proposed method to the DHG-14/28 and Shrec'17 datasets, which are hand gesture recognition datasets, it was possible to obtain higher class classification accuracy than the existing models. The classification accuracy was 98.72% and 98.16% for DHG-14/28, and 97.82% and 98.39% for Shrec'17 14 class/28 class.

### 키워드

textNAS, Multivariate Time Series Classification, Neural architecture Search, Hand Gesture Recognition

### 1. 서 론

손동작 인식은 인간과 컴퓨터의 상호작용(HCI, Human Computer Interaction)을 위해 활발히 연구되고 있다. 그 외에도 게임의 입력, 수화 등에 사용되어진다. 손동작 인식의 앞선 연구로 이미지 기반 입력[1]과 스켈레톤 기반 입력[2]을 통한 연구가 이루어져 왔다. 스켈레톤 기반 입력 방법은 조도나

손의 일부가 가려져 있을 때도 강건하게 손의 특징을 구하여 손동작 인식에 사용될 수 있는 장점을 지니고 있다. 키넥트나 인텔 RealSense 카메라를 이용하여 손의 스켈레톤을 정확하게 추출하여 손동작 인식[3]이 연구되고 있다. 본 논문에서도 인텔 RealSense 카메라에서 추출된 스켈레톤 정보를 사용하여 손동작 인식을 진행하였다. 스켈레톤을 사용한 손동작 인식은 동적 이미지 입력을 대상으로 특징을 그래프로 변형한 방법[4]이 연구되고 있다. 본 논문에서는 그래프를 적용하지 않고 데이터

\* corresponding author

세트에 저장된 특징을 입력으로 하여 손동작 인식을 하였다. 본 논문에서는 손동작 인식 데이터셋으로 DHG-14/28 데이터셋[5], SHREC'17 Track 데이터 세트[6]를 사용하였다. 그리고 프로그래머가 데이터 세트에 적합한 딥러닝 모델을 찾는 수고를 줄일 수 있도록 NAS(Neural Architecture Search)[7] 방법이 활발히 연구되고 있는데 본 논문에서는 텍스트의 클래스 분류에 사용된 textNAS[8]를 다변수 시계열 데이터 분류에 적용할 수 있도록 수정하고 이를 손동작 인식에 적용하는 방법을 제안한다.

## II. 관련 연구

### 스켈레톤 기반 손동작 인식

전통적인 스켈레톤 손동작 인식 방법으로 손의 특징 서술자 구하기에 집중하였다. Smedt[5]는 손의 스켈레톤 특징으로 연결된 관절 서술자 방법을 제안하였다. 최근에는 딥러닝을 사용하여 손동작 인식이 이루어지고 있다. CNN(Convolution Neural Network)나 LSTM(Long Short Term Memory)을 사용해 이미지나 시계열 특징을 입력으로 하는 손동작 인식 방법[9]이 연구되고 있다.

## III. 제안하는 방법

데이터 세트 내 skeleton.txt 파일에 저장된 특징에서 데이터 증대를 위해 임의의 인덱스를 지정하는데 DHG-14/28 데이터 세트의 경우 20개의 임의의 인덱스를 지정하고 인덱스 이후 20개 프레임 특징을 저장하였다. SHREC'17 데이터 세트의 경우 40개의 임의의 인덱스를 지정하고 인덱스 이후 7개 프레임 특징을 저장하였다. DHG-14/28의 경우 데이터 모양은 (56000, 20, 66)이며 SHREC'17의 경우 (112000, 7, 66)이다. textNAS의 경우 단어를 임베딩하여 데이터를 저장하는데 다변수 시계열 데이터의 경우 임베딩이 필요 없어 임베딩 과정을 생략하였다. 그리고 textNAS에서 특정 길이를 지정하여 단어의 개수가 이보다 적은 경우 패딩을 하

여 길이를 맞춘다. 단어가 있을 때는 마스크를 1로, 단어가 없는 패딩된 영역에 대해 마스크를 0으로 지정하는데 다변수 시계열 데이터에는 패딩이 없으므로 모든 프레임에 대해 마스크를 1로 지정하였다. 예로 배치 사이즈가 128, 길이가 20인 DHG-14/28 데이터 세트에는 (128, 20) 크기의 1이 저장된 배열을 마스크로 지정하였다. 수정된 내용에 관한 파이썬 코드와 데이터 학습 과정은 블로그(<https://blog.naver.com/kimsjpk/222448001178>)에 올려두었다. 학습 과정은 먼저 train/validation/test 데이터를 4:1:1로 나누고 데이터 세트에 적합한 신경망 구조를 찾는다. 찾은 신경망 구조는 enas[10]의 macro 구조와 유사한 구조를 갖는다. enas의 경우 노드 내부에 (identity, 3 x 3 or 5 x 5 conv filter, 3 x 3 average pooling or max pooling)이 들어가는데 textNAS의 경우 (1-D convolutional operator with filter size (1, 3, 5, 7), recurrent operator (bi-directional GRU), self-attention operator, pooling operator (max/average))로 구성된다. 학습 후 찾은 모델의 예시는 그림 1과 같다.

## IV. 실험 결과

textNAS를 통해 데이터 세트의 train, validation 데이터를 사용해 적합한 신경망 구조를 찾고 train, test 데이터를 사용하여 신경망의 파라미터를 학습하게 된다. 학습 결과는 표 1, 2 와 같다.

표 1. DHG-14/28 클래스 분류 정확도

적용한 방법	14 class	28 class
DG-STA[11]	91.9%	88.0%
MLF-LSTM[12]	96.07%	94.4%
CGSM-LFM[13]	98.57%	88.29%
제안한 방법	98.72%	98.16%

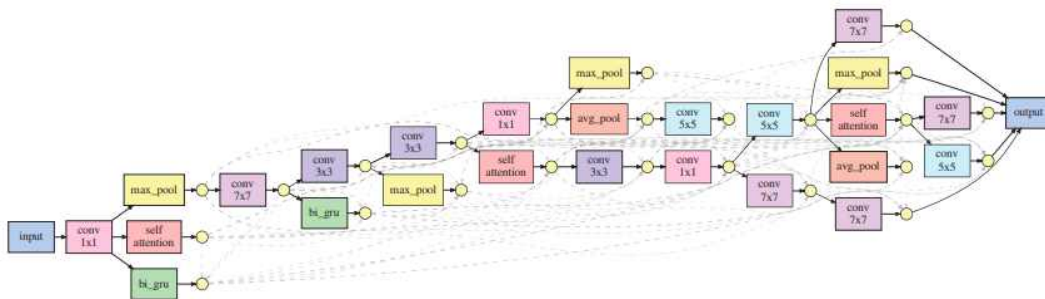


그림 1. textNAS에서 찾은 신경망 구조 예시

표 2. SHREC'17 Track 클래스 분류 정확도

적용한 방법	14 class	28 class
STA-Res-TCN[14]	93.6%	90.7%
ST-GCN[15]	92.7%	87.7%
DG-STA[11]	94.4%	90.7%
제안한 방법	97.82%	98.39%

## V. 결 론

본 논문에서는 텍스트 분류에 사용된 textNAS를 다변수 시계열 데이터 분류에 사용할 수 있도록 수정하고 이를 손동작 인식에 사용한 방법을 제안한다. 제안한 방법을 통해 다변수 시계열 분류의 다양한 분야[16]에 적용할 수 있다. 본 논문에 제안한 방법을 적용하여 다변수 시계열 데이터 분류 정확도를 더욱 높일 수 있을 것으로 기대한다. 사용한 딥러닝 라이브러리는 nni 1.7, pytorch 1.9이며 ubuntu 20.04, Cuda 11.2 버전에서 실험을 진행하였다. 사용된 컴퓨터 사양은 i5-9300, GTX 1660Ti이다.

## References

- [1] FREEMAN, William T.; ROTH, Michal. Orientation histograms for hand gesture recognition. In: International workshop on automatic face and gesture recognition. IEEE Computer Society, Washington, DC, p. 296-301, 1995.
- [2] CHEN, Xinghao, et al. Motion feature augmented recurrent neural network for skeleton-based dynamic hand gesture recognition. In: 2017 IEEE International Conference on Image Processing (ICIP). IEEE, p. 2881-2885, 2017.
- [3] OBERWEGER, Markus; LEPETIT, Vincent. Deepprior++: Improving fast and accurate 3d hand pose estimation. In: Proceedings of the IEEE international conference on computer vision Workshops. p. 585-594. 2017.
- [4] SI, Chenyang, et al. Skeleton-based action recognition with spatial reasoning and temporal stack learning. In: Proceedings of the European Conference on Computer Vision (ECCV). p. 103-118. 2018.
- [5] DE SMEDT, Quentin; WANNOUS, Hazem; VANDEBORRE, Jean-Philippe. Skeleton-based dynamic hand gesture recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. p. 1-9. 2016.
- [6] DE SMEDT, Quentin, et al. Shrec'17 track: 3d hand gesture recognition using a depth and skeletal dataset. In: 3DOR-10th Eurographics Workshop on 3D Object Retrieval. p. 1-6. 2017.
- [7] ZOPH, Barret; LE, Quoc V. Neural architecture search with reinforcement learning. arXiv preprint arXiv:1611.01578, 2016.
- [8] WANG, Yujing, et al. Textnas: A neural architecture search space tailored for text representation. In: Proceedings of the AAAI Conference on Artificial Intelligence. p. 9242-9249. 2020.
- [9] NUNEZ, Juan C., et al. Convolutional neural networks and long short-term memory for skeleton-based human activity and hand gesture recognition. Pattern Recognition, 76: 80-94. 2018.
- [10] PHAM, Hieu, et al. Efficient neural architecture search via parameters sharing. In: International Conference on Machine Learning. PMLR, p. 4095-4104. 2018.
- [11] CHEN, Yuxiao, et al. Construct dynamic graphs for hand gesture recognition via spatial-temporal attention. arXiv preprint arXiv:1907.08871, 2019.
- [12] DO, Nhu-Tai, et al. Robust hand shape features for dynamic hand gesture recognition using multi-level feature LSTM. Applied Sciences, 10.18: 6293. 2020.
- [13] JIASHAN, L. I.; ZHONGHUA, L. I. Dynamic gesture recognition algorithm Combining Global Gesture Motion and Local Finger Motion for interactive teaching. IEEE Access, 2021.
- [14] HOU, Jingxuan, et al. Spatial-temporal attention res-TCN for skeleton-based dynamic hand gesture recognition. In: Proceedings of the European Conference on Computer Vision (ECCV) Workshops. p. 0-0. 2018.
- [15] YAN, Sijie; XIONG, Yuanjun; LIN, Dahua. Spatial temporal graph convolutional networks for skeleton-based action recognition. In: Thirty-second AAAI conference on artificial intelligence. 2018.
- [16] KIM, Gi-Duk. Action recognition, hand gesture recognition, and emotion recognition using text classification method. In: Proceedings of the Korean Society of Computer Information Conference. Korean Society of Computer Information, p. 213-216. 2021.