

시계열 데이터 분류와 NAS를 통한 손동작 인식

김기덕⁰, 김미숙^{*}, 이학만^{**}

⁰부산대학교 전기전자컴퓨터공학과,

^{*}부산대학교 멀티미디어협동과정,

^{**}부산대학교 전자계산학과

e-mail: kimsjpk@hanmail.net⁰, sook99@pusan.ac.kr^{*}, bruce@pusan.ac.kr^{**}

Hand Gesture recognition through NAS and time series classification

Gi-Duk Kim⁰, Mi-Sook Kim^{*}, Hackman Lee^{**}

⁰Dept. of Electricity and Electronic Computer Engineering, Pusan National University,

^{*}Dept. of Multimedia, Pusan National University,

^{**}Dept of Computer Science, Pusan National University

● 요약 ●

본 논문에서는 손동작 데이터에서 추출한 데이터를 다변수 시계열 데이터 분류를 자동으로 찾는 NAS 모델에 적용하여 손동작 인식 모델을 찾는 방법을 제안한다. NAS를 통해 모델을 구하는 과정은 프로그래머의 시간과 노력을 절감시켜준다. 손동작 인식을 위해 DHG-14/28 데이터셋과 SHREC'17 Track 데이터셋에 논문에서 제안한 방법을 적용하여 손동작 인식 정확도가 기존의 모델보다 높은 손동작 인식률을 얻음을 실험을 통하여 확인하였다. 실험에서 DHG-14/28 데이터셋의 손동작 인식 정확도는 96.38%, 96.63%, SHREC'17 Track 데이터셋의 정확도는 96.88%, 96.57%를 얻었다.

키워드: 손동작 인식(hand gesture recognition), NAS(neural architecture search), mcfly(mcfly)

I. Introduction

손동작 인식은 컴퓨터와 인간의 상호작용을 위한 수단으로 관심을 받고 있다. 예를 들면 마우스와 키보드를 대체할 수 있는 인식 수단, Kinect와 같은 센서에서의 데이터를 처리하여 게임이나 VR(Virtual Reality)를 가능하게 해주며 말을 할 수 없는 장애인이나 손동작함으로써 의사 표현을 인식하는 도구로써 사용될 수 있다.[1] 손동작 인식의 방법으로 센서를 부착한 장갑을 통한 손가락 움직임의 인식, 카메라에서 피부 색상을 추출하여 손 영역을 추출한 뒤 머신러닝이나 딥러닝 방법을 사용한 손동작 인식, Kinect와 depth camera를 사용하여 손가락의 스켈레톤 정보와 깊이 정보로부터 손동작 인식이 사용되고 있다. 색상을 적용한 방법은 조도의 영향을 받고 주위 배경에 영향을 받아 이를 제거하기 위한 별도의 알고리즘이 요구된다는 단점을 지니고 있다. 이를 극복하려는 방법으로 센서 카메라를 사용한 방법이 연구되고 있다. 본 논문에서는 센서 카메라를 사용하여 손가락의 스켈레톤 정보를 추출하고 NAS(neural architecture search)[2]를 적용한 딥러닝 모델을 사용하여 손동작 인식을 진행하였다. NAS를 사용하면 모델의 구조와 파라미터를 자동으로 찾아주어 프로그래머의 시간과 노력을 절감해 주는 장점이 있다. 실험 데이터로 사용자의

손동작을 Intel RealSense 카메라를 사용하여 손의 스켈레톤 정보를 저장한 DHG-14/28 데이터셋[3]과 SHREC'17 Track 데이터셋[4]에 시계열 데이터 분류 모델을 자동으로 찾아주는 mcfly[5] 프로그램을 (<https://github.com/NLeSC/mcfly>)을 적용하여 손동작 인식을 진행하였다.

II. Preliminaries

1. Related works

1.1 스켈레톤 기반 손동작 인식

센서 카메라에서 손가락 관절의 위치와 각도 정보를 기하학적 분석을 통하여 공간 정보 및 시계열 데이터로 저장한다. 이를 SVM이나 HMM과 같은 머신러닝[6]과 CNN과 LSTM과 같은 딥러닝[7]을 사용하여 손동작 인식이 이루어진다. 센서 카메라를 사용하기 때문에 조도의 영향을 받지 않으며 배경 정보를 제거할 수 있는 장점을

자니고 있다.

1.2 neural architecture search

신경망을 데이터에 학습시키기 위해서 데이터마다 적합한 신경망 모델의 구조와 하이퍼 파라미터를 일일이 찾는 노력을 줄이기 위한 방법으로 NAS가 사용되고 있다. 이때 조정되는 하이퍼 파라미터로 학습률, 가중치 초기화 방법, 적층할 레이어의 수 등이 있다. 하이퍼 파라미터 탐색 방법으로 베이지안 최적화 방법이 주로 연구되고 있다.

1.3 mcfly

시계열 데이터 분류를 위한 파이썬 기반의 딥러닝 라이브러리이다. 4가지 딥러닝 모델(CNN, DeepConvLSTM, ResNet, InceptionTime)에서 임의로 학습률과 레이어 계층 등을 특정 개수만큼 정한다. 정한 모델에서 데이터를 입력받아 사용자가 정한 학습 epoch 만큼 학습을 하고 가장 높은 validation accuracy를 갖는 모델을 자동으로 추출해 준다.

III. The Proposed Scheme

Intel RealSense 카메라에서 추출된 스켈레톤 정보로부터 일정한 특정 크기와 프레임 길이를 갖는 데이터로 나누어 저장한다. 이때 데이터 증대를 위하여 전체 길이로부터 무작위로 시작지점을 지정하고 이후 특정 프레임 길이 만큼 저장을 하여 데이터 가공을 하였다. DHG-14/28의 경우 20프레임을 SHREC'17 Track 데이터셋의 경우는 7프레임을 각각 저장하였다. DHG-14/28 데이터셋에서는 20개의 무작위 인덱스를 지정하여 데이터를 저장하였고 데이터 모양은 (56,000, 20, 66)이다. 그리고 SHREC'17 데이터셋에서는 40개의 무작위 인덱스를 지정하여 데이터를 저장하였고 데이터 모양은 (112,000, 7, 66)이다. 저장된 데이터의 모양이 (N, L, C) 3차원 데이터이며 이는 다변수 시계열 데이터의 모양과 같다. 그래서 mcfly 라이브러리를 사용하여 다변수 시계열 데이터에 적합한 딥러닝 모델을 찾아 학습을 진행하였다. 임의의 하이퍼 파라미터와 레이어 계층을 가진 30개의 모델을 정하고 20epoch 만큼 데이터에서 임의로 추출한 5,000개의 데이터에 학습하여 가장 높은 validation accuracy를 갖는 모델을 최적 모델로 정하였다. 최적 모델로 정해진 하이퍼 파라미터를 사용하여 전체 데이터의 학습을 200 epoch 진행하였다. 데이터는 train 데이터와 test 데이터를 4:1로 나누어 학습 데이터로 사용하였다. 학습 후 각 데이터셋에 대한 손동작 분류 정확도는 표 1과 표2와 같다.

Table 1. DHG-14/28 Dataset Accuracy

method	accuracy	
	14 class	28 class
SoCJ+HoHD+HoWR[8]	86.86	84.22
DG-STA[9]	91.9	88
MLF-LSTM[10]	96.07	94.4
제안한 방법	96.38	96.63

Table 2. SHREC'17 Track Dataset Accuracy

method	accuracy	
	14 class	28 class
STA-Res-TCN[11]	93.6	90.7
ST-GCN[12]	92.7	87.7
DG-STA[9]	94.4	90.7
제안한 방법	96.88	96.57

IV. Conclusions

본 논문에서는 손동작 인식 데이터셋인 DHG-14/28 데이터셋과 SHREC'17 Track 데이터셋에서 무작위 인덱스로 추출된 데이터를 다변수 시계열 분류 모델을 자동으로 찾아주는 mcfly를 사용하여 데이터에 적합한 모델을 자동으로 찾아주는 방법을 제안하였다. 탐색을 통하여 찾아낸 모델로 학습한 결과 기존의 손동작 인식에 사용된 모델보다 높은 손동작 인식 정확도를 얻을 수 있었다.

REFERENCES

- [1] Oudah, Munir, Ali Al-Naji, and Javaan Chahl. "Hand gesture recognition based on computer vision: a review of techniques." journal of Imaging 6.8 (2020): 73.
- [2] Zoph, Barret, and Quoc V. Le. "Neural architecture search with reinforcement learning." arXiv preprint arXiv:1611.01578 (2016).
- [3] Dynamic Hand Gesture Recognition using Skeleton-based Features , Quentin De Smedt, Hazem Wannous and Jean-Philippe Vandeborre, 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)
- [4] De Smedt, Quentin, et al. "Shrec'17 track: 3d hand gesture recognition using a depth and skeletal dataset." 2017.
- [5] Van Kuppevelt, D., et al. "Mcfly: Automated deep learning on time series." SoftwareX 12 (2020): 100548.
- [6] Chen, Yanmei, et al. "A real-time dynamic hand gesture recognition system using kinect sensor." 2015 IEEE International Conference on Robotics and Biomimetics (ROBIO). IEEE, 2015.
- [7] 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 (2018): 80-94.
- [8] De Smedt, Quentin, Hazem Wannous, and Jean-Philippe Vandeborre. "Heterogeneous hand gesture recognition using 3D dynamic skeletal data." Computer Vision and

Image Understanding 181 (2019): 60-72.

- [9] Chen, Yuxiao, et al. "Construct dynamic graphs for hand gesture recognition via spatial-temporal attention." arXiv preprint arXiv:1907.08871 (2019).
- [10] Do, Nhu-Tai, et al. "Robust Hand Shape Features for Dynamic Hand Gesture Recognition Using Multi-Level Feature LSTM." Applied Sciences 10.18 (2020): 6293.
- [11] Hou, Jingxuan, et al. "Spatial-temporal attention res-TCN for skeleton-based dynamic hand gesture recognition." Proceedings of the European Conference on Computer Vision (ECCV). 2018.
- [12] Yan, Sijie, Yuanjun Xiong, and Dahua Lin. "Spatial temporal graph convolutional networks for skeleton-based action recognition." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 32. No. 1. 2018.