The Journal of The Institute of Internet, Broadcasting and Communication (IIBC)
Vol. 22, No. 3, pp.69-74, Jun. 30, 2022. pISSN 2289-0238, eISSN 2289-0246

https://doi.org/10.7236/JIIBC.2022.22.3.69
JIIBC 2022-3-11

음성 데이터 전처리 기법에 따른 뉴로모픽 아키텍처 기반 음성 인식 모델의 성능 분석

Performance Analysis of Speech Recognition Model based on Neuromorphic Architecture of Speech Data Preprocessing Technique

조진성*, 김봉재**

Jinsung Cho*, Bongjae Kim**

요 약 뉴로모픽 아키텍처에서 동작하는 SNN (Spiking Neural Network) 은 인간의 신경망을 모방하여 만들어졌다. 뉴로모픽 아키텍처 기반의 뉴로모픽 컴퓨팅은 GPU를 이용한 딥러닝 기법보다 상대적으로 낮은 전력을 요구한다. 이와 같은 이유로 뉴로모픽 아키텍처를 이용하여 다양한 인공지능 모델을 지원하고자 하는 연구가 활발히 일어나고 있다. 본 논문에서는 음성 데이터 전처리 기법에 따른 뉴로모픽 아키텍처 기반의 음성 인식 모델의 성능 분석을 진행하였다. 실험 결과 푸리에 변환 기반 음성 데이터 전처리시 최대 84% 정도의 인식 정확도 성능을 보임을 확인하였다. 따라서 뉴로모픽 아키텍처 기반의 음성 인식 서비스가 효과적으로 활용될 수 있음을 확인하였다.

Abstract SNN (Spiking Neural Network) operating in neuromorphic architecture was created by mimicking human neural networks. Neuromorphic computing based on neuromorphic architecture requires relatively lower power than typical deep learning techniques based on GPUs. For this reason, research to support various artificial intelligence models using neuromorphic architecture is actively taking place. This paper conducted a performance analysis of the speech recognition model based on neuromorphic architecture according to the speech data preprocessing technique. As a result of the experiment, it showed up to 84% of speech recognition accuracy performance when preprocessing speech data using the Fourier transform. Therefore, it was confirmed that the speech recognition service based on the neuromorphic architecture can be effectively utilized.

Key Words : Artificial Intelligence, Deep Learning, Neuromorphic Computing, Speech Recognition

*준회원, 충북대학교 전기·전자·정보·컴퓨터공학부
**정회원, 충북대학교 컴퓨터공학과
접수일자 2022년 5월 20일, 수정완료 2022년 6월 2일
게재확정일자 2022년 6월 10일

Received: 20 May, 2022 / Revised: 2 June, 2022 /
Accepted: 10 June, 2022
**Corresponding Author: bjkim@chungbuk.ac.kr
Department of Computer Engineering, Chungbuk National University, Korea

Ⅰ. 서 론

뉴로모픽 컴퓨팅은 인간의 뉴런의 동작을 모방한 회로를 만들어 인간의 인지 기능을 모사하려는 컴퓨터공학분야이다. 뉴로모픽 컴퓨팅을 위한 SNN(Spiking Neural Network)은 뉴로모픽 아키텍처에서 동작 가능한 3세대 인공신경망 모델이다[1]. 뉴로모픽 아키텍처 기반 SNN 모델은 2세대 인공신경망인 DNN(Deep Neural Network) 기반의 기존 딥러닝 기법을[2][3] 대체할 차세대 인공지능 모델로 주목받고 있다[4][5].

인간의 신경계는 뉴런이라 불리는 신경 세포가 전기 신호가 임계값을 넘었을 때 신호를 다른 뉴런으로 전달하는 방식으로 동작한다. 뉴로모픽 아키텍처는 이런 인간의 신경계 뉴런을 모방하여 만들어졌다. 뉴로모픽 아키텍처는 생물학적 뉴런을 모방하였기 때문에 동작에 낮은 전력이 소모된다는 장점을 갖는다. 이와 다르게 2세대 인공신경망인 DNN(Deep Neural Network) 기반의 기존 딥러닝 기법들은 일반적으로 전력 소모가 큰 고성능 컴퓨팅 장치인 다수의 GPU나 CPU를 필요하다는 단점이 존재한다[6][7]. 이러한 이유로 뉴로모픽 컴퓨팅을 지원하는 뉴로모픽 아키텍처와 SNN을 활용해 다양한 문제를 해결하고자 하는 연구가 활발히 일어나고 있다[8][9][10][11].

본 논문에서는 뉴로모픽 아키텍처 기반 SNN 모델의 구성 형태와 음성 데이터의 전처리 기법에 따른 음성 인식 성능을 분석하였다. 전처리 기법에는 다운 샘플링(Down sampling), RMSE(Root Mean Square Energy), 푸리에 변환(Fourier transform)을 사용하였다[12][13][14]. SNN 모델의 입력 데이터 크기에 따른 정확도 측정에는 196, 400, 784 의 벡터 크기를 사용하였다. 실험에는 직접 녹음한 단어 음성 데이터를 사용하였다. 데이터셋은 총 다섯 개의 단어 클래스로 구성되어 있으며 각 단어 클래스는 150개의 데이터로 구성되어 있다. SNN 모델의 정확도 측정에는 K-Fold[15] 검증 방식을 사용하였다. 실험 결과 다운 샘플링과 RMSE 기법에서 대체로 낮은 음성 인식 성능을 보였다. 반면 푸리에 변환을 이용하여 음성 데이터를 전처리한 경우 최대 84% 정도의 인식 정확도를 보였다. 따라서 음성 인식 기반의 서비스가 뉴로모픽 아키텍처와 SNN 모델을 기반으로 구현 가능함을 확인하였다.

이후 논문의 구성은 다음과 같다. 2장에서는 실험에 사용된 데이터셋과 전처리 기법을 서술한다. 3장에서 실험 환경 및 결과를 서술하고 4장 결론을 끝으로 논문을 마친다.

Ⅱ. 데이터셋 및 음성 데이터 전처리

1. 데이터셋 구성

뉴로모픽 아키텍처 기반 SNN 모델의 음성 인식 성능을 분석하기 위하여 총 다섯 가지 단어를 한국어로 녹음하여 사용하였다. 표 1은 학습 및 테스트에 사용된 단어 클래스의 종류와 사용된 데이터의 수를 보여준다. 단어는 “위”, “아래”, “가운데”, “왼쪽”. “오른쪽”으로 구성되었으며 총 다섯 개의 클래스이다. 각 단어는 150개의 데이터로 구성되어 있으며, 전체 데이터의 개수는 총 750개이다. 각 단어 클래스의 음성 데이터는 1채널의 WAV 포맷 형태를 가진다. 또한 각 원본 음성 데이터의 샘플링 레이트(Sampling rate)는 44,100Hz 이다.

표 1. 음성 데이터 클래스의 종류 및 그 개수
Table 1. Types of data classes and the numbers of their

단어	데이터 개수
위	150
아래	150
가운데	150
왼쪽	150
오른쪽	150
합계	**750**

2. 음성 데이터 전처리 기법

구축된 음성 데이터셋의 각 음성 데이터의 샘플링 레이트는 44,100Hz 이다. 각 음성 데이터를 뉴로모픽 아키텍처 기반 SNN 모델의 입력 데이터로 사용하는 것은 적합하지 않다. 따라서 각 음성 데이터를 SNN 모델의 입력 크기로 변화하는 과정이 필요하다. 이와 같은 음성 데이터 전처리 기법으로 다운 샘플링 기법, RMSE 기법, 푸리에 변환 기법을 적용한다.

첫 번째로 사용한 기법인 다운 샘플링 기법은 녹음된 파일의 샘플링 주기를 낮추는 방법으로 원본을 최대한 유지함과 동시에 데이터의 축소가 가능하다. 예를 들어 1ms 단위로 샘플링한 데이터를 10ms 단위로 샘플링할 경우 1/10 수준으로 데이터의 크기를 축소하는 것이 가능하다. 다운 샘플링 기법은 음성 이외에 이미지, 비디오에서도 사용되는 기술로 넓은 범용성을 가지고 있다. 두 번째 기법은 RMSE 기법으로 샘플링된 값에서 연속된 특

정 구간의 음성 데이터 값의 평균 제곱근을 이용하여 음성 데이터의 크기를 줄이는 전처리 기법이다. RMSE를 적용하는 음성 데이터 구간의 크기를 조절하여 다양한 크기로 음성 데이터의 축소가 가능하다. 마지막으로 푸리에 변환 기법은 시간 도메인을 갖는 데이터를 주파수 도메인으로 해당 음성 데이터 신호를 변환한다. 다시 말해서 특정 음성 데이터의 주요 주파수 성분을 얻어낼 수 있다. 일반적으로 음성 데이터 분석 및 데이터 변환에 많이 사용되는 기법이다.

세 가지 전처리 기법을 사용하여 전처리된 데이터는 데이터가 갖는 값의 범위가 서로 상이하다. 이러한 문제를 해결하기 위해서 0에서 1 사이 값으로 정규화한다.

Ⅲ. 실험 환경 및 결과

1. 실험 환경

본 논문의 실험에서는 음성 데이터 전처리 기법에 따른 뉴로모픽 아키텍처 기반 SNN 모델의 음성 인식 정확도를 분석하기 위하여 표 2와 같은 다양한 상황을 고려하였다. 이후 모델의 정확도 성능의 측정에서는 K-Fold 검증 기법을 적용하였다. K값을 5로 지정하여 전체 데이터를 8:2의 비율로 학습 데이터와 검증 데이터로 나누어 실험을 진행하고 평균 정확도를 산정하였다.

표 2. 인식 정확도 분석을 위한 실험 파라미터
Table 2. Experimental parameters for recognition accuracy analysis

항목	내용
SNN 모델 입력 데이터 크기	196, 400, 784
SNN 모델의 인식 가능 클래스의 수	위, 아래
	위, 아래, 가운데
	위, 아래, 가운데, 왼쪽, 오른쪽
SNN 모델의 내부 뉴런의 수	81, 40, 20

뉴로모픽 아키텍처 기반 인공신경망 SNN 모델의 구동을 위하여 FPGA기반 뉴로모픽 아키텍처 보드인 DE1-SoC 보드를 사용하였다. DE1-SoC은 FPGA 칩을 프로그래밍 하는 것으로 뉴로모픽 아키텍처와 동일한 동작이 가능하다. DE1-SoC에서 동작이 가능한 SNN 모델의 규모는 입력 데이터의 크기와 내부 뉴런 수의 곱이 16,000보다 작아야 한다. 또한 내부 뉴런 수와 출력 데이터의 크기의 곱 역시 16,000보다 작아야 한다. 표 3과 표 4는 각각 실험에 사용된 뉴로모픽 아키텍처인 DE1-SoC와 뉴로모픽 아키텍처 추상화 및 제어용 Host PC의 세부 사양 정보이다.

표 3. DE1-SoC 사양 정보
Table 3. DE1-SoC specs

항목	내용
OS	Linux Angstrom 2014.12
CPU	Dual-core ARM Cortex-A9
메모리	DDR3 1GB
네트워크	10/100/1000 Megabit Ethernet
FPGA	Cyclone V 5CSEA5

표 4. Host PC 사양 정보
Table 4. Host PC specs

항목	내용
OS	Ubuntu 20.04 LTS
CPU	AMD 라이젠 스레드리퍼 3960X
메모리	DDR4 64GB
네트워크	10/100/1000 Megabit Ethernet

DE1-SoC의 특성상 DE1-SoC에서 동작하는 SNN 모델의 형태는 입력 데이터 크기, 내부 뉴런의 수, 출력 데이터의 크기에 따라 달라진다. 앞서 설명했듯이 입력 데이터 크기가 커지면 사용 가능한 뉴런의 수가 줄어든다. 본 실험에서는 인식해야 하는 클래스의 수가 최대 5개이므로 입력 데이터 크기에 따라 SNN 모델의 뉴런 수가 결정된다. 입력 데이터의 크기에 따른 SNN 모델의 뉴런의 최대 사용 가능 수는 수식(1)과 같으며 이때 $Neuron_{max}$는 16,000이다. 예를 들어 입력으로 196 크기의 벡터를 사용할 경우, DE1-SoC에서 동작 가능한 SNN 모델의 뉴런의 개수는 ⌊ 16,000/196 ⌋ 으로 81개가 된다.

$$Num_{neuron} = \lfloor Neuron_{max} / InputData_{size} \rfloor \quad (1)$$

그림 1은 호스트(Host) PC와 DE1-SoC를 이용한 음성 인식 과정을 보여준다. 음성 인식을 위한 SNN 모델은 DE1-SoC에서 탑재되어 동작한다. Host PC와 이더넷 통신으로 전처리된 음성 인식 데이터와 해당 데이터에 대한 인식 결과를 송수신하는 형태로 동작한다. 실험

에서 Host PC와 DE1-SoC의 통신 및 뉴로모픽 아키텍처 기반 SNN 모델 동작을 위하여 Python 프로그래밍 언어 기반의 SNN 신경망 구축 및 테스트 프레임워크인 Nengo FPGA[16]를 사용하였다.

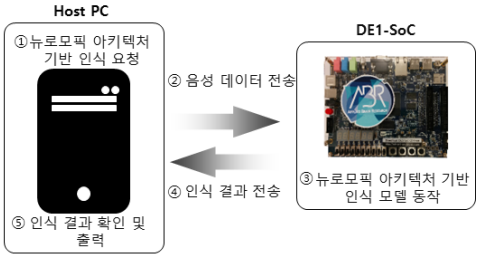

그림 1. DE1-SoC 기반 음성인식 모델의 동작 순서
Fig. 1. Operation sequence of DE1-SoC-based speech recognition model

2. 실험 결과

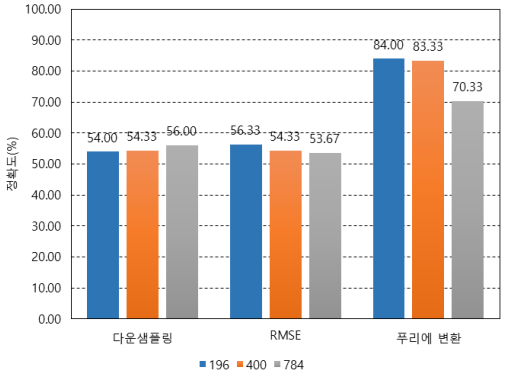

그림 2. 두 가지 클래스(위, 아래) 인식 정확도
Fig. 2. Recognition accuracy results of two classes (up and down)

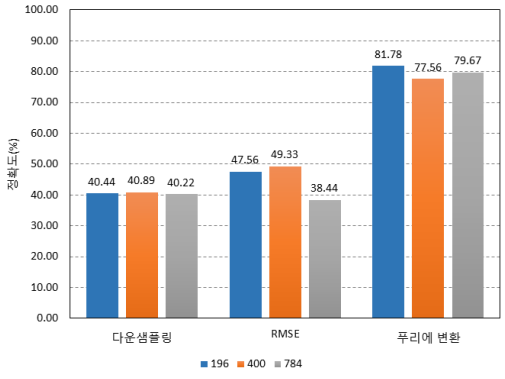

그림 3. 세 가지 클래스(위, 아래, 가운데) 인식 정확도
Fig. 3. Recognition accuracy results of three classes (up, down, and center)

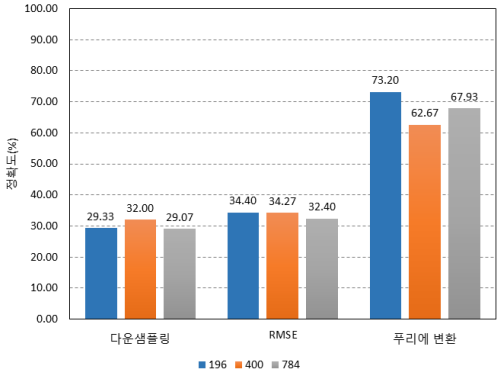

그림 4. 다섯 가지 클래스(위, 아래, 가운데, 왼쪽, 오른쪽) 인식 정확도
Fig. 4. Recognition accuracy results of five classes (up, down, center, left, and right)

그림 2, 그림 3, 그림 4는 인식하려는 단어의 클래스 수와 전처리된 입력 데이터의 크기, SNN 모델의 뉴런의 수에 따른 뉴로모픽 아키텍처 기반 SNN 모델의 음성 인식 정확도이다. 그림 2는 2개 클래스, 그림 3은 3개 클래스, 그림 4는 4개 클래스를 인식하는 SNN 모델의 음성 인식 정확도를 보여준다. 입력 데이터의 크기가 각각 196, 400, 784인 경우의 SNN 모델의 뉴런의 수는 81, 40, 20이다.

그림 2에서 확인할 수 있듯이 두 가지 클래스 분류에서 다운 샘플링 기법과 RMSE 기법을 사용하여 전처리를 진행한 경우 모두 60% 이하의 정확도를 보였다. 반면 푸리에 변환은 최고 84.00%의 정확도를 보였다. 그림 3의 상황에서도 마찬가지로 푸리에 변환의 경우 81.78%로 가장 좋은 성능을 보였다. 마지막으로 그림 4에서 확인할 수 있듯이, 다섯 개의 클래스를 인식하는 SNN 모델의 성능은 두 가지, 세 가지 클래스를 인식할 때 보다 대체로 낮은 성능을 보였다. 하지만 푸리에 변환 기법으로 데이터를 전처리한 경우 다섯 가지 클래스를 분류하는 상황에서도 73.20%의 정확도를 보였다. 실험 결과 모든 상황에서 푸리에 변환이 우수한 성능을 보였다. 따라서 뉴로모픽 아키텍처 기반 SNN 모델의 음성 데이터 전처리에 푸리에 변환이 다운 샘플링 기법이나 RMSE 기법보다 적합하다는 것을 확인할 수 있다.

실험에서 가장 우수한 성능을 보인 푸리에 변환의 경우, 입력 데이터 크기가 196의 크기를 가지는 벡터일 때 가장 좋은 성능을 보였다. 196의 입력 데이터 크기를 사용할 경우 FPGA 기반 뉴로모픽 아키텍처인 DE1-SoC의 특성상 SNN 모델의 뉴런을 최대 81개 사용할 수 있

다. 반면 400의 입력 데이터 크기와 784의 입력 데이터 크기는 각각 40개와 20개의 뉴런만 사용할 수 있다. 따라서 SNN 모델의 내부 뉴런의 수도 인식 성능에 영향을 주는 요소이고 이 때문에 SNN 모델의 정확도 성능이 감소하였다고 예측된다. 그림 3과 그림 4에서 푸리에 변환을 사용하여 784의 입력 크기로 변환하였을 때 400의 입력 크기를 사용하였을 때보다 정확도가 더욱 높은 것을 확인할 수 있었다. 이를 통해 일정한 상관관계가 있는 것이 아니라 최적의 성능을 보이는 입력 데이터의 크기와 SNN 모델의 뉴런수가 존재할 수 있음을 알 수 있다.

Ⅳ. 결 론

본 논문에서는 음성 데이터 전처리 기법에 따른 뉴로모픽 아키텍처 기반 SNN 모델의 음성 인식 정확도를 측정하고 분석하였다. 다운 샘플링과 RMSE 기법을 사용하여 전처리하는 경우, 푸리에 변환 기반 데이터 전처리 기법과 비교할 때 대체로 낮은 정확도를 보였다. 푸리에 변환을 사용하여 주파수 도메인으로 음성 데이터 전처리를 진행하였을 때 최대 84.00%의 정확도 성능을 보였다. 이를 통해 뉴로모픽 아키텍처 기반 SNN 모델을 이용하여 음성 인식과 같은 서비스의 구현 및 제공이 가능함을 확인하였다.

향후 연구에서는 뉴로모픽 아키텍처의 음절 단위 음성 인식 가능 여부를 실험할 것이다. 이외에도 뉴로모픽 아키텍처 기반 SNN 모델의 단순 인식, 분류 문제 이외에도 다양한 분야에 사용될 가능성을 확인하기 위한 기법을 설계하고 성능을 분석할 것이다.

References

[1] Oh, K. I., et al., "Trend of AI Neuromorphic Semiconductor Technology", Electronics and Telecommunications Trends, Vol. 35, No. 3, pp. 76-84, 2020.
DOI: https://doi.org/10.22648/ETRI.2020.J.350308

[2] Chang-Bok Kim, "Deep Learning Model for Prediction Rate Improvement of Weather Data using Parallel Merge Structure", The Journal of KIIT, Vol. 20, No. 4, pp. 131-140, 2022.
DOI: https://doi.org/10.14801/jkiit.2022.20.4.131

[3] Youngjoon Cho, Jongwon Kim, "A Study on The Classification of Target-objects with The Deep-learning Model in The Vision-images", Journal of the Korea Academia-Industrial cooperation Society(JKAIS), Vol. 22, No. 2, pp. 20-25, 2021.
DOI: http://dx.doi.org/10.5762/KAIS.2021.22.2.20

[4] Moon, S. E., et al., "Next-generation neuromorphic hardware technology", Electronics and Telecommunications Trends, Vol. 33, No. 6, pp. 58-68, 2018.
DOI: https://doi.org/10.22648/ETRI.2018.J.330607

[5] Ghosh-Dastidar, et al., "Spiking Neural Networks", International Journal of Neural systems, Vol. 19, No. 04, pp. 295-308, 2009.
DOI: https://doi.org/10.1142/S0129065709002002

[6] Silver, David, et al., "Mastering the game of Go with deep neural networks and tree search", Nature, Vol. 529, No. 7587, pp. 484-489, 2016.
DOI: https://doi.org/10.1038/nature16961

[7] Hong-Jin Park, "Trend Analysis of Korea Papers in the Fields of 'Artificial Intelligence' 'Machine Learning' and 'Deep Learning' ", Journal of Korea Institute of Information, Electronics, and Communication Technology, Vol. 13, No. 4, pp. 283-292, 2020.
DOI: http://dx.doi.org/10.17661/jkiiect.2020.13.4.283

[8] Park, Sangmin, Junyoung Heo, "Conversion Tools of Spiking Deep Neural Network based on ONNX", The Journal of The Institute of Internet, Broadcasting and Communication Vol. 20, No. 2, pp. 165-170, 2020.
DOI: https://doi.org/10.7236/JIIBC.2020.20.2.165

[9] Park, Kicheol, and Bongjae Kim, "Dynamic neuromorphic architecture selection scheme for intelligent Internet of Things services", Concurrency and Computation: Practice and Experience e6357, 2021.
DOI: https://doi.org/10.1002/cpe.6357

[10] Bing, Zhenshan, et al., "End to end learning of a multi-layered SNN based on R-STDP for a target tracking snake-like robot", 2019 International Conference on Robotics and Automation (ICRA). IEEE, 2019.
DOI: https://doi.org/10.1109/ICRA.2019.8793774

[11] Yang, Shuangming, et al., "BiCoSS: toward large-scale cognition brain with multigranular neuromorphic architecture", IEEE Transactions on Neural Networks and Learning Systems, 2021.
DOI: https://doi.org/10.1109/TNNLS.2020.3045492

[12] Shen, Minmin, Ping Xue, Ci Wang, "Down-sampling based video coding using super-resolution technique", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 21, No. 6, pp. 755-765, 2011.
DOI: https://doi.org/10.1109/TCSVT.2011.2130390

[13] Seryasat, O. R., F. Honarvar, Abolfazl Rahmani, "Multi-fault diagnosis of ball bearing using FFT, wavelet energy entropy mean and root mean square (RMS)", IEEE international conference on systems, man and cybernetics, 2010.
DOI: https://doi.org/10.1109/ICSMC.2010.5642389

[14] C. Rader, N. Brenner, "A new principle for fast Fourier transformation", IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 24, No. 3, pp. 264-266, 1976.
DOI: https://doi.org/10.1109/TASSP.1976.1162805

[15] Rodriguez, Juan D., Aritz Perez, Jose A. Lozano, "Sensitivity analysis of k-fold cross validation in prediction error estimation", IEEE transactions on pattern analysis and machine intelligence, Vol. 32, No. 3, pp. 569-575, 2009.
DOI: https://doi.org/10.1109/TPAMI.2009.187

[16] Morcos, Benjamin, "Nengofpga: an fpga backend for the nengo neural simulator", MS thesis, University of Waterloo, 2019.

저 자 소 개

조 진 성(준회원)

- 2021.3~: 충북대학교 전기·전자·정보·컴퓨터공학부 석사 재학
- 2020.2: 선문대학교 컴퓨터공학부 학사

김 봉 재(정회원)

- 2021.03~: 충북대학교 컴퓨터공학과 부교수
- 2016.03~2021.02: 선문대학교 컴퓨터공학부 조교수
- 2015.01~2016.02: 한국전자기술연구원 임베디드·SW 연구센터 선임연구원

※ "본 연구는 과학기술정보통신부 및 정보통신기획평가원의 지역지능화혁신인재양 (Grand ICT연구센터) 사업의 연구결과로 수행되었음" (IITP-2022-2020-0-01462)