

다중 애플리케이션 처리를 위한 경량 인공지능 하드웨어 기반 통합 프레임워크 연구

전석훈** · 이재학** · 한지수** · 김병수*

A Study of Unified Framework with Light Weight Artificial Intelligence Hardware
for Broad range of Applications

Seok-Hun Jeon** · Jae-Hack Lee** · Ji-Su Han** · Byung-Soo Kim*

요 약

경량 인공지능 하드웨어는 다양한 문제의 해결을 위해 멀티모달 센서 데이터를 입력받아 특징 선택, 추출, 차원축소, 정규화 과정을 수행한 후 인공지능 엔진으로 예측 결과를 도출한다. 다양한 애플리케이션에서 높은 성능을 달성하기 위해서는 이러한 경량 인공지능 하드웨어의 초 매개변수와 전체적인 전처리 시스템의 구성을 데이터에 맞춰 최적화할 필요가 있다. 본 논문에서는 경량 인공지능 하드웨어의 효율적인 제어 및 최적화를 위한 통합 프레임워크를 제안한다. 제안된 통합 프레임워크는 데이터 전처리 및 뉴로모픽 기반 경량 인공지능 엔진을 유연하게 재구성할 수 있으며, 최적의 모델을 생성할 수 있다. 기능검증을 위해 손글씨 이미지 데이터 세트와 관성 센서 데이터 기반의 낙상 검출 데이터 세트를 사용하였으며, 실험 결과 제안하는 통합 프레임워크가 각각의 데이터 세트에서 90% 이상의 정확도를 갖는 최적의 모델을 생성함을 확인하였다.

ABSTRACT

A lightweight artificial intelligence hardware has made great strides in many application areas. In general, a lightweight artificial intelligence system consist of lightweight artificial intelligence engine and preprocessor including feature selection, generation, extraction, and normalization. In order to achieve optimal performance in broad range of applications, lightweight artificial intelligence system needs to choose a good preprocessing function and set their respective hyper-parameters. This paper proposes a unified framework for a lightweight artificial intelligence system and utilization method for finding models with optimal performance to use on a given dataset. The proposed unified framework can easily generate a model combined with preprocessing functions and lightweight artificial intelligence engine. In performance evaluation using handwritten image dataset and fall detection dataset measured with inertial sensor, the proposed unified framework showed building optimal artificial intelligence models with over 90% test accuracy.

키워드

Lightweight Artificial Intelligence Engine, Software Framework, Preprocessing, Performance Optimization
경량 인공지능 엔진, 소프트웨어 프레임워크, 전처리, 성능 최적화

* 교신저자 : 전자부품연구원 SoC플랫폼연구센터 · Received : Sep. 06, 2019, Revised : Sep. 25, 2019, Accepted : Oct. 15, 2019

** 전자부품연구원(jhk507, jisuhan91, seokhun.jeon@ · Corresponding Author : Byung-Soo Kim

keti.re.kr)

SoC Platform Research Center, Korea Electronics Technology Institute

· 접수 일 : 2019. 09. 06

Email : bskim4k@keti.re.kr

· 수정완료일 : 2019. 09. 25

· 게재확정일 : 2019. 10. 15

I. 서론

현재 우리는 인간과 인간, 인간과 사물, 사물과 사물이 서로 네트워크로 연결된 초연결사회[1]에서 살고 있다. 초연결사회는 사물인터넷을 기반으로 구현되었으며 소비자 시장과 산업 시스템 전반에 다양한 서비스를 제공한다. 또한, 인공지능이 사물인터넷에 적용되어 다양하고 지능적인 서비스를 제공하는 사물지능[2]이 전략 기술의 추세로 주목받고 있다. 사물 지능을 담당하는 엣지 기기에서 뉴로모픽 기술이 적용된 경량 인공지능(LWAI : Lightweight Artificial Intelligence) 시스템은 처리속도, 전력소비, 저 복잡도 측면에서 상당한 이점을 갖는다[3]. 경량 인공지능 시스템은 낮은 코어성능, 메모리, 제한된 인터페이스 등 한정된 자원 활용의 극대화를 위해 다양한 전처리 기술이 필요하다[4-6]. 이와 더불어 경량 인공지능 하드웨어의 성능을 극대화하는 소프트웨어 프레임워크 기술 역시 필수적이다[7, 8].

본 논문은 다중 애플리케이션 처리를 위해 뉴로모픽 기반 경량 인공지능 엔진에 다양한 전처리 기능이 추가된 경량 인공지능 시스템을 위한 통합 프레임워크를 개발하였다. 통합 프레임워크는 그림 1과 같이 호스트 PC에 위치하며 경량 인공지능 시스템이 적용된 하드웨어 플랫폼과 UART 인터페이스 기반의 시리얼 통신으로 제어된다. 2장에서는 경량 인공지능 시스템의 각 기능을 상세히 설명하였고, 3장에서 제안된 통합 프레임워크에서 경량 인공지능 시스템을 위한 최적 모델 생성 방법에 대해 설명하였다. 4장은 서로 다른 특성을 갖는 손글씨 이미지 데이터와 관성 센서 기반 낙상 데이터를 사용하여 제안된 통합 프레임워크가 최적화된 인공지능 모델을 생성함을 검증하였다.

II. 경량 인공지능 시스템

본 논문에서 제안하는 경량 인공지능 시스템은 그림 2와 같이 멀티모달 센서 데이터 인터페이스가 포함된 센서 서브시스템(sensor subsystem), 다차원 데이터에서 의미 있는 데이터를 추출하고 데이터 차원을 줄이는 전처리 시스템(preprocessor), 뉴런의 학습 및 인지 과정을 수행하는 경량 인공지능 엔진(LWAI

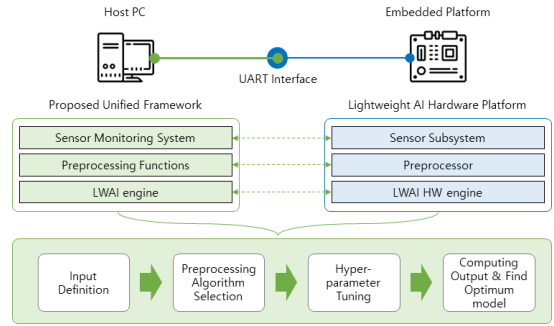


그림 1. 경량 인공지능 하드웨어 플랫폼과 연동된 통합 프레임워크 시스템 아키텍처

Fig. 1 Unified framework system architecture with lightweight artificial intelligence hardware platform

engine)으로 구성된다. 이 시스템은 그림 2 (a) Cortex M3 기반 임베디드 프로세서와 (b) Xilinx FPGA 기반 하드웨어 플랫폼으로 제작되었고, 통합 프레임워크에서 UART 인터페이스 기반의 시리얼 통신으로 제어된다.

2.1 센서 서브시스템

멀티모달 센서 데이터 처리를 위해 SPI, I2C, I2S, UART, GPIO 총 5종의 통신 인터페이스와 오차 및 잡음 보정을 위한 신호 보정 회로로 구성된다. 통신 인터페이스로 다양한 센서의 초깃값(initial value)이 설정되며, 센서에서 감지된 데이터가 출력되고, 출력된 데이터의 오차 및 잡음이 신호 보정 회로[9]를 통해 보정된다.

2.2 데이터 전처리 시스템

다양한 입력 데이터 처리를 위해 전처리 시스템은 특징 선택 (feature selection), 특징 생성 (feature generation), 특징 추출 (feature extraction), 정규화 (normalization) 4가지로 구성된다. 특징 선택 기능은 전체 데이터에서 불필요한 특징 데이터를 제거한다. 특징 생성은 시간에 따른 특징 정보로 구성된 2차원 데이터를 해당 라벨을 대표하는 특징 벡터로 변환하는 기능으로 전처리 시스템에 대표적으로 사용되는 최댓값, 최솟값, 평균 등의 기능이 적용되었다. 다음으로 특징 추출 혹은 차원축소를 사용하여 크기가 큰 특징 데이터를 압축 혹은 축소하여 새로운 특징 데이터로 변환한다. 마지막으로 데이터를 정규화 하여 특

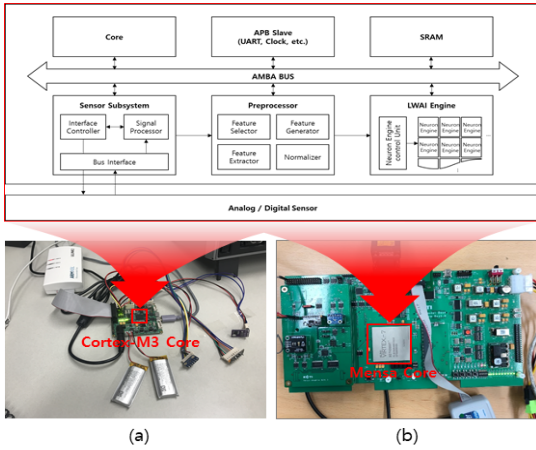


그림 2. 경량 인공지능 시스템 블록 다이어그램 및 (a) Cortex-M3 core를 사용한 임베디드 플랫폼, (b) Mena core를 사용한 Xilinx FPGA 하드웨어 플랫폼
 Fig. 2 LWAI system block diagram for hardware platform using (a) Coretexas-M3 core and (b) Mena core

정 데이터 분포를 경량 인공지능 엔진의 입력 범위로 조정하고, 개별 특징의 가중 편차로 인한 오차를 최소화한다. 이 4가지 전처리 기능들은 입력 데이터의 특성에 맞춰 사용자가 유연하게 재구성할 수 있다.

2.3. 경량 인공지능 엔진

경량 인공지능 엔진은 거리 기반의 군집 알고리즘을 사용하여 라벨을 구분하는 RCE-NN(Restricted Coulomb Energy Neural Network)[10] 기반으로 설계되

었다. 학습 상태에서 가변적으로 네트워크의 수정이 가능하므로 다양한 센서 애플리케이션에 유연하게 적용할 수 있다. 입력 데이터와 군집 중심점과의 거리를 기반으로 학습 및 인지가 수행되므로 군집 공간 반지름이 중요한 초 매개변수가 된다. 따라서 군집 공간 반지름의 최솟값 (min- influence field : IF)과 최댓값 (max IF)은 경량 인공지능 엔진의 인지 성능을 결정한다.

III. 통합 프레임워크

제안된 통합 프레임워크는 Windows 및 Linux 기반 응용 프로그램으로 구현하였고, 경량 인공지능 시스템의 모델 생성과 성능 평가를 그림 3과 같이 수행한다. 경량 인공지능 시스템은 통합 프레임워크에서 UART 인터페이스 기반의 시리얼 통신으로 제어된다. 센서 데이터 모니터링 (sensor data monitoring), 데이터 전처리 시스템 (preprocessor), 경량 인공지능 엔진 (LWAI engine)의 단계별 시뮬레이션이 가능하므로, 사용자가 유연하게 데이터 전처리 기능을 조합하고, 경량 인공지능 엔진의 초 매개변수 값을 변경할 수 있다. 시뮬레이션 결과, 인공지능 모델 구성 및 초 매개변수 값을 파일 관리 시스템 (file management)을 통해 파일로 저장된다. 또한, 통합 프레임워크는 경량 인공지능 시스템의 다양한 기능들을 소프트웨어 라이브러리 형태로 구현하여 하드웨어 플랫폼 없이도 다양한 애플리케이션의 최적화 모델 생성과 성능 검증이 가능하다.

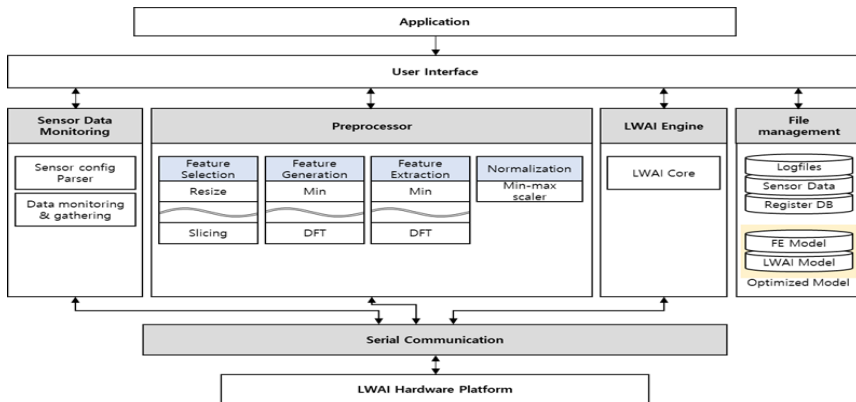


그림 3. 통합 프레임워크 구조
 Fig. 3 Unified framework structure

3.1 최적 모델 생성 워크플로우

그림 4는 통합 프레임워크를 사용한 애플리케이션

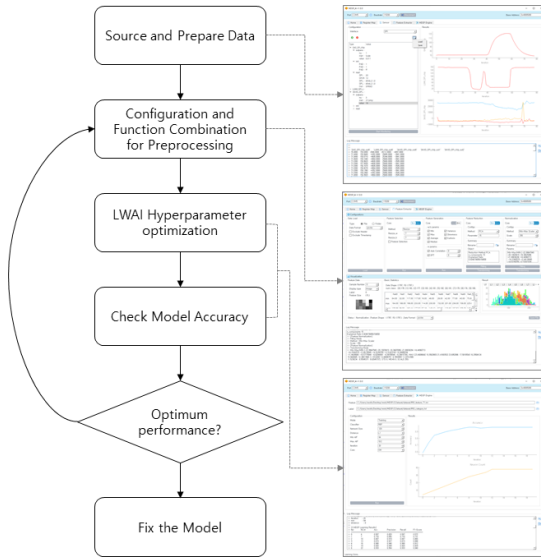


그림 4. 최적화된 경량 인공지능 시스템 모델 생성을 위한 플로우

Fig. 4 Work flow for optimized lightweight AI system model

최적화 모델 생성 과정을 나타낸다. 첫째, 데이터 획득 및 준비과정 (source and prepare data)에서 센서 서브시스템 혹은 외부에서 획득한 데이터는 통합 프레임워크의 형식에 맞게 변환된다. 둘째, 데이터 전처리 설정 (configuration and function combination for preprocessing)에서는 데이터 특성에 맞게 전처리 기능을 조합하고 초 매개변수를 조정한다. 셋째, 경량인공지능 엔진의 초 매개변수를 최적화(LWAI hyperparameter optimization) 하여 모델의 정확도를 평가 (check model accuracy)한다. 최적의 성능을 만족할 때까지 전처리 기능을 다양하게 조합하고, 경량 인공지능 엔진의 성능을 평가하여 최종적으로 애플리케이션에 최적화된 경량 인공지능 시스템 모델을 찾는다.

3.2. 경량 인공지능 엔진 초 매개변수 최적화

거리 기반 군집 알고리즘을 사용하는 경량 인공지능 엔진에서 군집 공간 반지름의 최소, 최댓값은 경량 인공지능 엔진 성능과 밀접하게 관련된 초 매개변수이다. 경량 인공지능 엔진의 저 전력, 저 복잡도 측면에서 최적화된 성능을 갖는 초 매개변수를 찾기 위해

훈련 데이터의 에러율 (E_{train}) 과 뉴런 유닛의 활성화 비율($N_{committed}/N_{samples}$) 을 사용한다. 최소한의 뉴런 유닛의 사용하면서 훈련 데이터의 에러가 최소화 되는 지점이 최적의 초 매개변수 파라미터가 되며 실제 테스트 데이터에서 최적의 과적합 특성과 높은 정확도를 보인다. 넓은 범위의 최소, 최대 반지름 값에서 최적화 값의 추론을 위해 그림 5와 같이 테스트 모드(test mode)를 coarse와 fine 두 단계로 구분하여

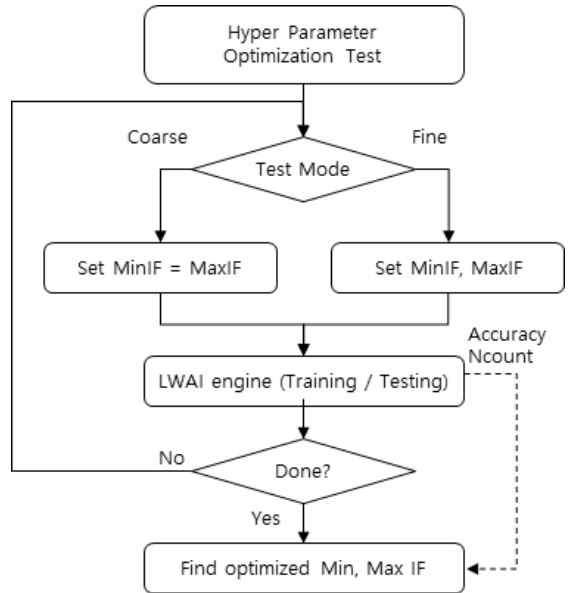


그림 5. 경량 인공지능 엔진의 초 매개변수 최적화 처리 과정

Fig. 5 Hyper-parameter optimization flow-chart for lightweight AI engine

수행한다. Coarse 테스트는 최소, 최대 반지름을 동일하게 설정하고, 균등 간격으로 성능을 측정한다. 최소, 최대 반지름이 동일할 때 학습 반복횟수를 1회만 하여도 데이터 판별을 위한 군집이 결정되므로 매우 빠른 테스트가 가능하다. Fine 테스트에서는 coarse 테스트에서 찾은 최적의 반지름 값을 기준으로 최소 최대 반지름을 다르게 설정하여 균등 간격으로 세밀하게 테스트 범위를 좁혀 테스트를 진행한다. Coarse와 fine 테스트에서 경량 인공지능 엔진의 성능 측정이 완료되면 수식 1과 같이 최적의 초 매개변수 값을 도출한다. 여기서 x 와 y 는 각각 min IF, max IF를 의미한다.

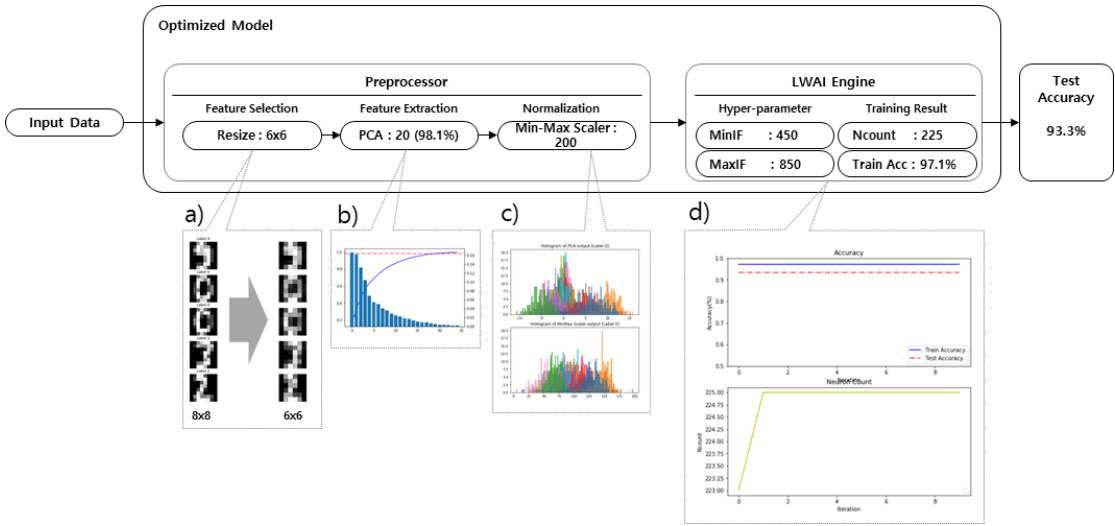


그림 6. 손글씨 데이터셋 최적 모델 생성 결과

Fig. 6 Result of optimized model for handwritten digit detection application

$$\operatorname{argmin}_{x, y \in AIF} \left(E_{train}(x, y) - \frac{N_{committed}(x, y)}{N_{sample}} \right) \quad (1)$$

IV. 실험결과

통합 프레임워크의 모델 생성 및 성능 검증을 위해 머신러닝 알고리즘 성능평가에 주로 사용되는 UCI 데이터 세트 중 손글씨 숫자검출을 위한 이미지 데이터[11]와 관성 센서 데이터를 이용한 낙상 검출 데이터[12]를 사용하였다. 3장에서 다룬 최적 모델 생성 과정과 경량 인공지능 엔진의 초 매개변수 최적화 방법을 사용하여 각각의 데이터 세트에 관한 최적화 모델을 결정하고 정확도를 확인하였다.

4.1 손글씨 이미지 데이터

그림 6은 경량 인공지능 하드웨어 시스템에 최적화된 손글씨 이미지 데이터 분류 모델을 나타낸다. 입력된 데이터(input data)는 a)부터 d)까지의 절차를 거쳐 처리된다. a) 한정된 하드웨어 입력 크기를 고려하여 이미지 사이즈 크기를 8×8 에서 6×6 로 줄이는 'resize' 기능을 사용하였다. b) 테스트 데이터 정확도를 고려하여 주성분 분석에서 변환된 데이터의 누적 분산 비율 (explained variance ratio)이 약 98.1%가

되는 20개의 특징 데이터로 축소하였다. c) 축소된 데이터는 0~200의 범위로 min-max scaler에 의해 정규화 되어 경량 인공지능 엔진의 입력으로 사용된다. d) 3.2장과 동일한 방법으로 경량 인공지능 엔진의 초 매개변수인 min, max IF의 최적 값을 각각 450, 850으로 도출하였다. 학습 데이터를 통해 225개의 뉴런이 사용 되었으며, 100%의 학습 정확도를 보이며, 테스트 데이터를 사용한 손글씨 이미지 데이터 분류 모델의 정확도는 93% 이상임을 확인하였다.

4.2 관성센서 기반의 낙상검출 데이터

그림 7은 낙상 검출을 위한 최적화된 경량 인공지능 시스템 모델을 나타낸다. 4.1장과 동일한 방식으로 a)부터 d)까지 수행되며, 데이터는 이미지 데이터가 아닌 센서의 원시 데이터를 사용함으로써 제시된 시스템의 활용성이 높음을 입증하였다.

입력 데이터는 9축의 관성 센서 6개가 25Hz로 수 초간 사람의 동작에 따라 측정된 값으로 구성된다. a) 동작 이벤트가 발생한 위치의 주변 데이터를 추출하기 위해 입력 데이터에서 허리 위치의 가속도 x, y, z 값을 수식 2를 참고하여 계산하고, 최대크기를 갖는 위치에서 앞, 뒤 50개의 표본을 추출하였다. 다음으로 b) 시간에 따른 54가지 특징 정보 (6개 센서 x 9개 특징)를 최솟값, 최댓값 등 총 7가지 특징생성 기능을

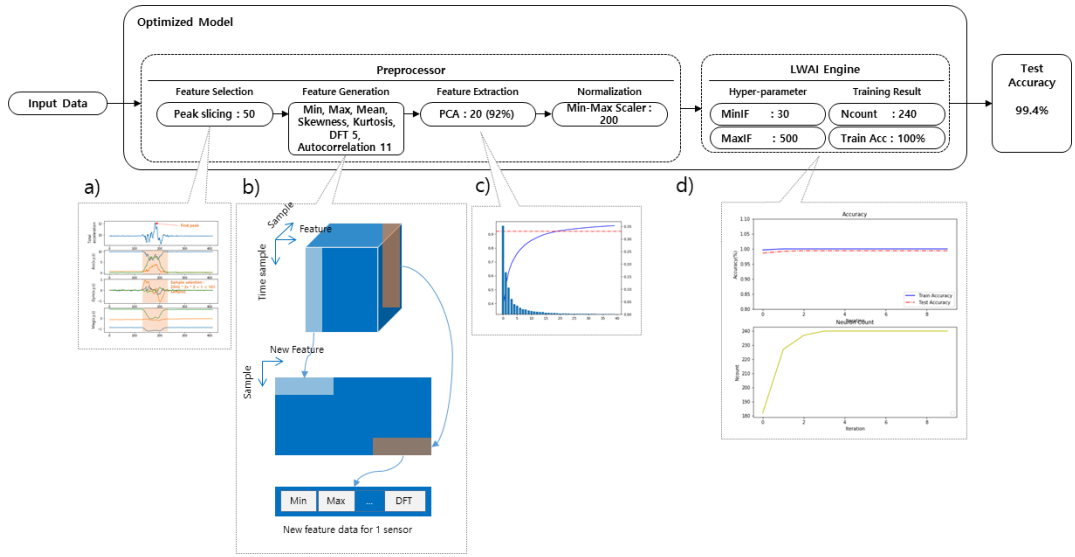


그림 7. 관성센서 기반 낙상 검출 데이터 최적 모델 생성 결과
 Fig. 7 Result of optimized model for fall detection application

사용하여 총 1404개의 새로운 특징으로 재구성한다. c) 재구성된 특징 데이터는 앞서 4.1과 동일하게 주성분 분석을 사용하여 92% 이상의 누적 분산 비율을 갖는 20개의 특징으로 데이터를 축소하고, 0~200의 범위로 정규화 한다. d) min, max IF의 최적화 값은 3.2장과 동일한 방법으로 각각 30, 500으로 도출하였고, 240개의 뉴런이 사용된 경량 인공지능 모델은 99.4% 이상의 정확도로 낙상 판별이 가능함을 보였다.

$$A_T = \sqrt{A_x^2 + A_y^2 + A_z^2} \quad (2)$$

V. 결 론

본 논문에서는 경량 인공지능 시스템을 위한 통합 프레임워크를 제안하였다. 제안된 통합 프레임워크는 멀티모달 센서 데이터의 사용이 가능하고, 다양한 전처리 기능 및 뉴로모픽 기반 경량 인공지능 엔진의 재구성이 가능하다. 따라서 다양한 애플리케이션에 범용적으로 적용되어 최적의 경량 인공지능 모델을 생성 한다. 일상생활 혹은 다양한 산업분야에서 사용되

는 옛지 기기에 경량 인공지능 시스템이 적용된다면 제안된 프레임워크를 통해 손쉽게 다양한 애플리케이션에 최적화된 경량 인공지능 모델의 생성이 가능할 것으로 예상된다.

References

- [1] Gartner Research, "The Hyper-connected Enterprise: Anticipating the Next Wave of Business," *Technical report*, Feb. 2008.
- [2] Gartner Research, "Gartner Identifies the Top 10 Strategic Technology Trends for 2019," *Technical report*, Oct. 2018.
- [3] B. Kim, J. Lee, T. Hwang, and D. Kim "Design of Lightweight Artificial Intelligence System for Multimodal Signal Processing," *J. of the Korea Institute of Electronic Communication Sciences*, vol. 13, no. 5, 2018, pp. 1037-1042.
- [4] A. Ozdemir and B. Barshan "Detecting Falls with Wearable Sensors Using Machine Learning Techniques," *J. of Sensors*, vol. 14, no. 6, Apr. 2014, pp. 10691-10708.
- [5] T. Baltrušaitis, C. Ahuja, and L. Morency,

“Multimodal Machine Learning: A Survey and Taxonomy,” *arXiv preprint arXiv:1705.09406*, 2017, pp. 1-20.

[6] J. Benediktsson and J. Sveinsson, “Feature extraction for neural network classifiers,” *Neurocomputation in Remote Sensing Data Analysis*, Berlin, Heidelberg, Germany, 1997, pp. 97-104.

[7] J. Cho, Y. Jung, S. Lee, and Y. Jung, “VLSI Implementation of Restricted Coulomb Energy Neural Network with Improved Learning Scheme,” *MDPI Electronics*, vol. 8, no. 563, 2019, pp. 68-73.

[8] H. Kim “A Study on The Real-Time Data Collection/Analysis/Processing Intelligent IoT,” *J. of the Korea Institute of Electronic Communication Sciences*, vol. 14, no. 2, 2019, pp. 317-322.

[9] J. Lee, B. Kim, H. Park, D. Kim, and J. Kwon “Genetic Algorithm Calibration Method and PnP Platform for Multimodal Sensor,” *J. of the Korea Institute of Electronic Communication Sciences*, vol. 14, no. 1, 2019, pp. 69-80.

[10] M. Buhmann, *Radial Basis Functions: Theory and Implementations*. Cambridge, United Kingdom, 2009.

[11] UCI Machine Learning Repository, “Optical Recognition of Handwritten Digits Data Set,” July 1998.

[12] UCI Machine Learning Repository, “Simulated Falls and Daily Living Activities Data Set,” June 2018.

저자 소개



전석훈(Seok-Hun Jeon)

2012년 숭실대학교 대학원 정보통신공학과 졸업(공학석사)
2017년 Seagate Korea Design Center Senior Engineer

2017년 ~현재 전자부품연구원 SoC플랫폼연구센터 연구원

※ 관심분야 : 인공지능 시스템, 최적화 알고리즘



이재학(Jae-Hack Lee)

2011년 아주대학교 대학원 전자공학과 졸업(공학석사)

2017년 아주대학교 대학원 전자공학과 졸업(공학박사)

2017년 ~현재 전자부품연구원 SoC플랫폼연구센터 선임연구원

※ 관심분야 : SoC 설계, 유무선 통신시스템

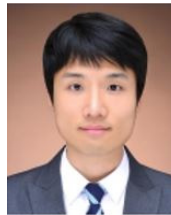


한지수(Ji-Su Han)

2016년 상명대학교 컴퓨터시스템공학과 졸업(공학사)

2017년 ~현재 전자부품연구원 SoC플랫폼연구센터 연구원

※ 관심분야 : 데이터 마이닝, 인공지능 시스템



김병수(Byung-Soo Kim)

2008년 인하대학교 정보통신공학과 졸업(공학석사)

2013년 인하대학교 정보통신공학과 졸업(공학박사)

2013년 ~현재 전자부품연구원 SoC플랫폼연구센터 선임연구원

※ 관심분야 : 인공지능 시스템 설계, SoC 설계

