

자생 매미 음성 분류를 위한 딥러닝 접근 : 주파수 변화 분석과 모델 최적화

장동석, 홍윤식*
인천대학교 컴퓨터공학과

Deep Learning Approach for Autonomous Cicada Sound Classification: Frequency Variation Analysis and Model Optimization

Dong-Seok Jang, Youn-Sik Hong*

Dept. of Computer Science & Engineering Incheon National University

요약 본 논문은 한국에서 서식하는 매미과 속인 말매미 등 12종의 국내 자생종 매미의 음성 데이터를 활용하여 딥러닝 기법을 통해 매미 종을 분류하는 새로운 접근 방식을 제시하였다. 표준화된 데이터 전처리 및 시간에 따른 주파수 변화를 시각적으로 나타내는 그래픽 표현방식인 스펙트로그램(Spectrogram)의 적용을 통해 주파수 변화와 시간적 변화를 동시에 시각화하여 데이터의 특성을 파악하고 활용하며 딥러닝 모델인 ResNet34, ResNet50, AlexNet 모델을 적용하였다. 드롭아웃 기법을 적용하여 과적합(Overfitting)을 방지하며, 다양한 학습률(Learning Rates)을 적용하여 모델의 학습 및 검증 과정을 최적화하였다. 이러한 접근을 통해 98% 이상의 높은 정확도로 매미 종을 식별을 검증하였다. 본 연구는 인공지능 기술인 CNN(Convolutional Neural Network)를 활용하여 생물 다양성 보존과 종 식별의 정확성을 높이기 위해 수행하였으며, 음성 데이터 기반의 딥러닝 시스템이 생태학적 연구와 환경 모니터링에 크게 기여할 수 있음을 시사한다. 나아가 본 연구는 생태계 보존 및 관리에 중요한 도구로 활용될 수 있음을 보여주며, 인공지능 기술과 생물 분류학을 결합하여 향후 생물 다양성 연구와 환경 보호를 위한 새로운 방법을 제시할 수 있다.

Abstract This paper presents a novel approach to classifying cicada species by using deep learning techniques that utilize acoustic data of 12 cicada species found in Korea, including Meimuna opalifera. Standardized data preprocessing and the application of spectrograms, which visually represent frequency changes over time, were used to simultaneously visualize both frequency and temporal changes, allowing for species identification from data characteristics. Deep learning models such as ResNet34, ResNet50, and AlexNet were applied. Dropout techniques were employed to prevent overfitting, and various learning rates were applied to optimize the training and validation processes of the models. The approach successfully identified cicada species with an accuracy of over 98%. This study enhances the accuracy of species identification and conservation of biodiversity by using the artificial intelligence technology of a convolutional neural network. It suggests that deep learning systems based on acoustic data can significantly contribute to ecological research and environmental monitoring. Furthermore, this study has the potential for use as an essential tool in ecosystem conservation and management, combining AI and taxonomy to propose new methods for future biodiversity research and environmental protection.

Keywords : Artificial Intelligence, Deep-learning, CNN, Classification, Spectrogram

*Corresponding Author : Youn-Sik Hong(Incheon National Univ.)

email: yshong@inu.ac.kr

Received April 25, 2024

Accepted July 5, 2024

Revised May 23, 2024

Published July 31, 2024

1. 서론

1.1 연구의 필요성 및 목적

1.1.1 자생 매미 종의 중요성과 보존

매미는 다양한 생태계에서 중요한 역할을 하는 곤충이다. 이들 종의 분포는 생태계의 건강을 반영하며, 종 다양성과 개체군의 상태를 정확하게 파악하는 것은 물론이고, 생태계의 변화를 모니터링하고 예측하는 데 필수적이라고 할 수 있다[1]. 특히 생태계에서 중요한 먹이사슬의 일부인 매미는 환경 변화에 따라 개체 수의 증감이 생태계에 미치는 영향을 시사할 수 있다. 이러한 변화를 조기에 감지하고 이해하기 위해서는 특정 종에 대한 고정밀 동정(Classification) 기술이 필요하며 현재 매미를 포함한 많은 생물 종의 동정은 전문가의 시각적·경험적 사고에 의존하고 있고, 이와같은 동정 방법은 시간이 많이 걸리고 주관적인 오류를 일으킬 수 있다. 이를 방지하기 위하여 인공지능 기술중에 각광받는 딥러닝 기술을 활용 및 적용하면 종 동정과정을 자동화 할 수 있고, 기존 동정방법 보다 빠르고 정확하게 객관적인 결론을 도출할 수 있다.

1.1.2 음성 데이터의 생물학적 보존가치

음성 데이터는 동물의 행동, 생태, 생리 상태를 반영하는 중요한 생물 정보를 포함한다. 특히 포유류, 조류, 곤충 등 다양한 분류군에서 나오는 소리 데이터는 종의 식별, 개체군의 건강 상태, 생태적 상호작용을 파악하는 데 중요한 지표로 활용되는데, 이러한 원시 생물 데이터의 특징값 추출을 통해 전처리하고 데이터 세트로 구축한 후 인공지능 모델에 적용함으로써, 보다 광범위한 결과를 도출할 수 있으며, 인공지능과의 융합 연구는 생물학적 연구와 환경 보존 전략 수립에 기여한다. 본 연구에서는 데이터의 분석을 통해 생물 다양성의 체계적 보존 및 관리에 필수적인 정보를 제공한다. 기존 수행되었던 전통적인 생물학적 조사 방법보다 시간과 비용을 훨씬 절감할 뿐만 아니라, 접근하기 어려운 지역이나 위험한 환경에서도 연구를 수행할 수 있으며, 이러한 인공지능 기술의 발전은 생물학과 융합하여 연구 방법론을 크게 변화시켜 보다 정확하고 효율적인 생태계 관리와 보존을 가능하게 한다[2].

1.2 기존 연구와의 차별점

1.2.1 기존 연구에서의 접근 방식과 한계

인공지능 기술은 크게 머신러닝과 딥러닝으로 구분되며, 머신러닝은 결정트리(decision tree), 서포트 벡터 머신, 랜덤 포레스트(Random Forest) 등 전통적인 알고리즘을 통해 선형 및 비선형 관계의 데이터를 학습한다. 그러나 이러한 전통적 접근법은 복잡한 데이터 구조에서 고차원 패턴과 상호작용을 모델링하는 데 한계가 있는 반면, 딥러닝 기반의 컨볼루션 신경망[3,4]은 입력 데이터의 다양한 특징을 자동추출하고 학습하며, 여러 층을 통한 점진적 추상화를 통해 패턴을 보다 효과적으로 파악한다. 이는 머신러닝보다 높은 정확도와 유연성을 제공한다는 것을 의미하며, 본 논문에서 사용한 매미 음성 데이터셋을 딥러닝 모델 적용 전 머신러닝 모델을 적용하여 분석한 결과 Fig. 1과 같이 높은 정확도를 보였으며, 모델별 하이퍼파라미터에 대한 설정은 다음과 같다. Random Forest에 경우 200개의 트리, 최대 깊이 10, 내부 노드 분할에 필요한 최소 샘플 수를 5, 리프 노드에 필요한 샘플 수를 4, 랜덤 시드는 42로 설정하였고, Logistic Regression에 경우, 0.1의 규제강도를 설정, liblinear의 알고리즘을 사용, 결과 수렴을 위한 최대 반복 횟수는 1,000번으로 설정하였다. 마지막 SVM에 경우, 규제 매개변수는 1.0, 커널 계수는 Scale, 랜덤 시드는 42로 설정하여 아래와 같은 실험 결과를 확인 하였다.

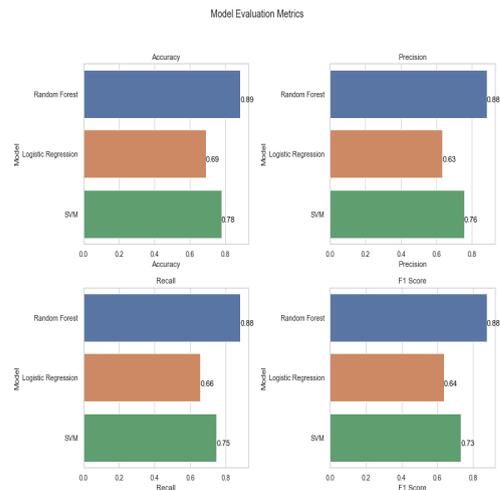


Fig. 1. M/L analysis results using voice data
(a) Recall, Accuracy, Precision, F1-Score

서론에서는 데이터 셋으로 선택된 자생 매미 종의 중요성과 보존, 음성 데이터의 생물학적 보존가치, 기존 연구에서의 접근 방식과 한계에 대하여 설명한다. 결과적으로 음성 데이터 셋을 활용하여 Random Forest,

Logistic Regression, SVM 모델을 적용하고, 각각의 하이퍼파라미터를 최적화하여 머신러닝 모델의 정확도를 확인하였고, 이를 통해 선행 연구된 머신러닝 연구의 결과와, 본 연구에서 진행한 딥러닝 모델과의 적용을 통하여 실험 결과를 검증해보고자 한다[2].

2. 본론

2.1 데이터 수집 및 전처리

2.1.1 매미 음성 데이터의 수집

본 논문에서는 현장에서 매미의 소리를 녹음하여 오디오 데이터를 수집하였다. 이 과정에서 국내 서식하는 12종의 매미[1] 음성 데이터를 확보하였으며, 각 종별로 평균 재생시간이 150분 이상 되도록 구축하였고, 세부 데이터 현황은 Table 1에 표현한다.

Table 1. Status of collected cicada vocalization data

종	파일 수	총 재생시간
소요산매미	174개	2:41:32
털매미	304개	4:53:38
유지매미	175개	2:52:09
말매미	284개	4:21:53
애매미	421개	6:33:49
쓰름매미	176개	2:46:48
참매미	277개	4:23:51
참깡매미	162개	2:32:20
늦털매미	217개	3:29:06
풀매미	163개	2:37:48
호좁매미	172개	2:44:10
세모배매미	155개	2:31:10

2.1.2 데이터의 전처리 및 표준화 과정

종별 매미 울음소리 데이터는 원시 상태에서 바로 딥러닝 모델에 적용될 수 없기 때문에 전처리 과정을 통해 소음 제거, ffmpeg 라이브러리를 사용한 샘플링 레이트 및 비트 레이트 변환 등을 수행하여 데이터의 일관성을 유지하였다. 본 논문에서 사용된 음성 데이터는 소음 제거(Noise Reduction)와 샘플링 레이트를 44,100Hz로 고정하는 등의 전처리 과정을 거쳐 일관성을 보장하였다.

2.2 데이터 분석 방법

2.2.1 스펙트로그램 및 데이터 셋 구축

스펙트로그램은 시간에 따른 주파수의 변화를 그래픽으로 시각화하는 도구이며 주파수, 즉 소리의 높낮이 변화를 이미지로 표현함으로써 소리의 주파수 내용과 시간적 특성을 효과적으로 파악할 수 있다. 본 논문에서는 Python을 기반으로 한 오디오 및 음악 신호 처리 라이브러리인 Librosa를 사용하여 MFCC(Mel-frequency cepstral coefficients)를 추출하고, 이를 통해 12종[1] 매미의 울음소리로부터 2,680개의 데이터 셋을 종별로 구분하여 구축하였다(Fig. 2 참조).

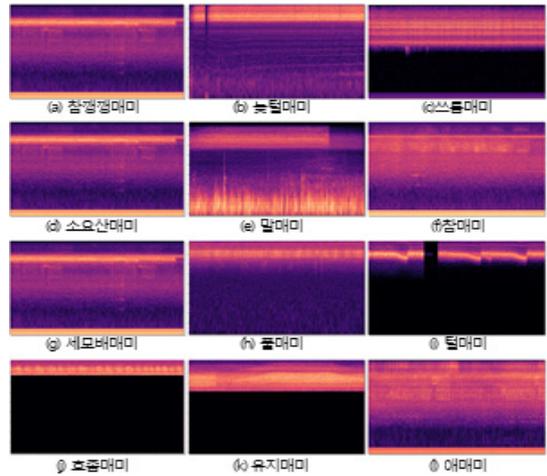


Fig. 2. Extract features from specific cicada vocalization data
(a) ~ (l) cicada Scientific name

2.3 딥러닝 모델의 구축 및 학습

2.3.1 딥러닝 모델

매미 울음소리 데이터셋은 매미의 종류를 식별하고 분류하는 딥러닝[3-6] 모델을 개발하기 위한 중요한 자원이며, ResNet34, ResNet50[5,6], AlexNet과 같은 주요 딥러닝 아키텍처를 사용하여 매미 울음소리 데이터셋을 융합하고 모델을 구축하는 방법을 설명한다[2,3].

2.3.2 ResNet34, ResNet50, AlexNet

ResNet34는 깊은 신경망 아키텍처로, 34개의 층을 갖춘 구조에서 잔여 블록(residual block)을 통해 그래디언트 소실 문제를 해결한다. 각 층은 컨볼루션, 배치 정규화, ReLU 활성화 함수, 맥스 풀링, 드롭아웃으로 구

성됩니다. ResNet50은 50개의 층을 갖는 더 깊고 강력한 신경망으로, ResNet34와 유사하나 더 많은 잔여 블록을 포함한다. AlexNet은 8개의 계층으로 구성된 모델로, 컨볼루션, 맥스 풀링, ReLU, 드롭아웃을 특징으로 하며, 복잡한 이미지 특징 학습에 효과적이다. ResNet50은 ResNet 시리즈의 한 종류로, 50개의 레이어로 구성된 깊고 강력한 신경망 아키텍처이다. Resnet34와 비슷한 구조를 가지고 있지만, 더 깊고 복잡한 네트워크로 구성되어 있고, Resnet34의 잔여 블록과는 다르게 잔여 연결(residual connections)을 사용하여 그래디언트 소실 문제를 해결하고, 매우 깊은 네트워크에서도 효율적인 학습을 가능하게 할 수 있다. AlexNet은 2012년 ImageNet Large Scale Visual Recognition Challenge(ILSVRC)에서 우수한 모델이며, Resnet34,50과는 다르게 8개의 계층으로 구성되어 있고, 컨볼루션(Convolution), 맥스 풀링(Max Pooling), ReLU 활성화 함수(ReLU), 드롭아웃(Dropout) 등의 레이어로 구성되어 있다. 이 아키텍처 역시 복잡한 이미지 특징을 학습하는데 효과적이며, 매미 울음소리 데이터셋과 융합하여 매미의 종류를 식별하고 분류를 위해 3개의 모델을 각각 선택하여 가장 높은 정확도를 갖는 결과를 도출한다.

2.3.3 모델 초기화 및 드롭아웃 기법의 적용

모든 딥러닝 모델은 효과적인 학습을 위해 초기화 과정이 필수적이며, 본 논문에서는 사전 학습된 모델을 초기화하고, 모델의 마지막 fully connected layer를 새로운 구성으로 교체하여 드롭아웃 기법을 적용함으로써 과적합을 방지하였다. 2.3.2.에서 제시한 모델의 적용을 위하여 사전 학습된 모델을 초기화하고 마지막 fully connected layer를 새롭게 설정하는 초기화 과정이 필요하며 이를 모델 초기화라고 본 논문에서는 정의한다. 초기화 과정이 진행되면 선택된 모델은 사전 학습(pretrained)된 모델로 초기화 되고, 모델의 마지막 fully connected layer를 새로운 fully connected layer로 대체하여 구성한다. 새롭게 생성된 fully connected layer에는 과적합(Overfitting)을 방지하기 위한 dropout layer가 추가되어 있으며, 본 논문에서 수행하는 실험에는 0.5로 비율을 설정하여 과적합을 방지하고 모델을 학습 및 테스트 하였다.

2.3.4 다양한 학습률의 적용과 모델 최적화

본 연구에서는 다양한 학습률을 적용하여, 대규모 데

이터셋으로 사전 훈련된 모델을 특정 데이터셋에 적용하는 전이 학습(Transfer learning) 방법을 사용하였다. 훈련횟수(epoch=25,50)은 모델별로 적용하여 총 6번을 수행하였다. learning rate는 [1e-06, 1e-05, 1e-04, 1e-03]으로 음의 기울기(순간변화율)가 가장 큰 점으로 실험을 수행하였고, batch size는 16으로 설정하였다 [2].

2.4 모델 검증 및 결과 분석

2.4.1 모델검증을 위한 환경설정

이전 장에서 제시한 학습률 및 모델에 따라 매미 울음소리를 기반으로 종 분류하는 연구를 진행하였으며, Table 2와 같이 model, 훈련횟수, learning rate 3가지 환경변수로 설정하고 최종 실험을 수행하였다. 실험은 Anaconda Jupyter Notebook을 사용한 로컬 환경에서 진행되었으며, CPU는 INTEL i9-9900K(4.9GHz), 메모리는 32GB, 그리고 NVIDIA RTX 2080ti 11GB GPU를 사용하였다. 또한, 메모리 누수를 방지하기 위해 가비지 콜렉션을 적용하여 모델 검증시 발생하는 오류를 사전에 방지하였다.

2.4.2 데이터 세트 분할

실험에서 사용된 데이터 전처리된 데이터 2,595장(Class당 216개)의 이미지를 기반으로 국내 자생종 매미 12종에 대하여 테스트 데이터, 검증 데이터, 트레이닝 데이터로 분할하였다. 데이터의 분할 계산은 다음과 같이 표현한다.

Table 2. Calculate the split for Test/Valid/Train data

- **Test Data** : 각 클래스당 10개의 이미지
→ 12개 클래스 : 120개
- **Valid Data** : 각 클래스의 데이터 수의 20%
→ 클래스당 43개 * 12 = 516개
- **Training Data** : 나머지 데이터
→ 클래스당 163개 * 12 = 1,956개

위와 같이 전처리가 완료된 Table 2와 같이 데이터를 분할하여 데이터 셋을 구축하고 딥러닝 모델을 적용한다.

딥러닝 모델의 적용을 위한 모델과 훈련횟수, Learning rate는 Table 3과 같다[2].

Table 3. Set parameters and analyze

Var.#	Name	Range
#1	Model	Resnet34
		Resnet50
		alexnet
#2	epoch	25
		50
#3	Learning rate	1e-06
		1e-05
		1e-04
		1e-03

2.4.3 모델의 정확도 평가

그림 Fig. 3를 통해서 그래프를 통해 다양한 모델과 에폭 수에 따른 학습률 1e-06에서의 정확도 변화를 확인할 수 있다. AlexNet, ResNet34, 그리고 ResNet50 모델에 대한 25 에폭과 50 에폭에서의 정확도가 표시되어 있으며, 초기 에폭에서는 모든 모델의 정확도가 비슷한 증가 추세를 보이다가, 에폭 수가 증가함에 따라 차이가 드러나고 있다. ResNet50 에폭 50은 이 중 가장 높은 정확도에 도달하는 것으로 나타나, 더 깊은 네트워크가 장기적으로 더 좋은 성능을 보임을 시사하는 반면, AlexNet은 25 에폭과 50 에폭 사이에 큰 성능 향상을 보이지 않아, 복잡도가 낮은 모델이 특정 지점 이후에는 추가적인 학습으로 큰 이득을 얻지 못할 수 있음을 확인할 수 있다.

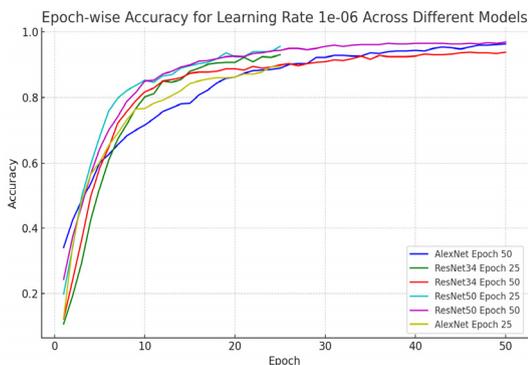


Fig. 3. learning rate 1e-06일 때 Accuracy

ResNet34는 중간 수준의 정확도 증가를 보여주며, 모델의 깊이와 학습의 효율성 간의 균형을 찾는 중요성을 강조한다. 이러한 실험 결과를 바탕으로 본 연구는 다양한 아키텍처가 주어진 데이터셋에 대해 어떻게 작용하는지를 평가하고, 특히 긴 학습 기간 동안 모델의 성능이

어떻게 변하는지를 분석한다. 결과는 모델 선택과 학습 전략을 결정할 때 중요한 정보를 제공하며, 이를 통해 최적의 정확도를 달성할 수 있는 학습 구조를 식별하는 데 도움을 줄 수 있을 것으로 사료된다.

3. 결론

본 연구는 한국 자생 매미 종의 음성 데이터를 분류하기 위해 다양한 딥러닝 아키텍처의 적용 가능성을 탐색하였다. ResNet34, ResNet50, 그리고 AlexNet 모델의 성능을 비교 분석한 결과, 깊은 신경망 구조를 가진 ResNet50이 더 긴 훈련 기간(50 에폭) 동안에도 높은 정확도를 유지하며 가장 우수한 성능을 나타내었고, 이는 생태학적 데이터의 복잡성과 미묘한 변화를 감지하기 위해서는 더 깊고 강력한 모델이 필요함을 시사한다. AlexNet은 상대적으로 적은 수의 층을 가지고 있음에도 초기 단계에서 빠르게 학습을 진행하는 모습을 보였으나, 이후 정확도가 크게 향상되지 않는 한계를 확인할 수 있었다. 이는 자생 매미 음성 인식과 같은 고도의 특화된 작업에는 모델의 깊이와 복잡성이 중요할 수 있음을 나타낸다. ResNet34는 중간 정도의 모델 복잡성을 가지고 있으며, 성능은 AlexNet보다 나은 반면 ResNet50보다는 떨어지는 결과를 도출하였다. 이는 네트워크의 깊이가 성능을 향상시킬 수 있으나, 특정 지점 이후에는 이득의 크기가 감소할 수 있음을 확인한다. 종합적으로, 이 연구는 자생 매미의 음성 데이터에 기반한 식별과 분류를 위한 딥러닝 모델의 선정에 있어서, 균형 잡힌 아키텍처의 중요성을 강조하고, 더 나아가, 이는 생물 다양성의 보존과 생태계 모니터링에 있어서, 높은 정확도와 효율적인 자동화 프로세스를 달성하기 위한 최적의 모델 구조를 선별하는 데 중요한 기준을 제공한다.

References

- [1] S.J.Kim, J.H.W. Korean Cicada Ecological Guide. p.128, Nature and Ecology, 2017, pp.20-114.
- [2] Kang, M-J., Kim, Y-S., Shin, H-Y., Park, J., "Development of the Deep Learning System for Bird Classification Using Birdsong", *Journal of Knowledge Information Technology and Systems (JKITS)*, Vol. 15, No. 2, pp. 195-203, April 2020. DOI: <https://doi.org/10.34163/jkits.2020.15.2.005>
- [3] Islam, S. B., Valles, D., Hibbitts, T. J., Ryberg, W. A.,

Walkup, D. K., and Forstner, M. R. J., "Animal Species Recognition with Deep Convolutional Neural Networks from Ecological Camera Trap Images," *Animals*, Vol. 13, No. 9, pp. 1526, May 2023.
DOI: <https://doi.org/10.3390/ani13091526>.

- [4] B. M. Kim, "Image Classification Technology Based on Deep Learning", *The Journal of The Korean Institute of Communication Sciences*, 35(12), 8-14. 2018.
- [5] M.R.J.F., D.V., S.B.I., et al., "Animal Species Recognition with Deep Convolutional Neural Networks from Ecological Camera Trap Images", *Animals*, Vol.12, No.9, pp. 1153, September 2022.
DOI: <https://doi.org/10.3390/ani12091153>.
- [6] H. J. Shin, S. I. Lee, H. W. Jeoung, J. W. Park, "Indoor Plants Image Classification Using Deep Learning and Web Application for Providing Information of Plants", *Journal of Knowledge Information Technology and Systems (JKITS)*, vol.15, no.2, pp.167-175, 2020.
DOI: <https://doi.org/10.34163/JKITS.2020.15.2.002>
-

장 동 석(Dong-Seok Jang)

[정회원]



- 2022년 2월 : 인천대학교 컴퓨터 공학 (공학석사)
- 2022년 9월 ~ 현재 : 인천대학교 컴퓨터공학과 박사과정

<관심분야>

인공지능, 가상화, 빅데이터

홍 윤 식(Youn-Sik Hong)

[정회원]



- 1983년 2월 : 한양대학교 전자공학과 (공학석사)
- 1989년 2월 : 한국과학기술원 (KAIST) 전기 및 전자공학과 (공학 박사)
- 1989년 3월 ~ 1991년 7월 : LG 전자(주)우면연구소 선임연구원
- 1998년 3월 ~ 1999년 12월 : LG정통신(주)단말연구소 자문교수
- 1991년 8월 ~ 현재 : 인천대학교 컴퓨터공학부 교수

<관심분야>

모바일 컴퓨팅, 사물인터넷, 헬스케어