

이미지 객체 및 메타정보 기반 GPT 활용 SNS 문장 작성 보조 시스템

이동희¹, 문미경¹, 최봉준^{1*}

¹동서대학교 소프트웨어학과

GPT-enabled SNS Sentence writing support system Based on Image Object and Meta Information

Dong-Hee Lee¹, Mikyeong Moon¹, Bong-Jun Choi^{1*}

¹Department of Software, Dongseo University

요약 본 연구에서는 SNS와 같이 이미지와 함께 글을 작성하는 활동을 보조하기 위해 YOLO와 GPT를 활용한 SNS 문장 작성 보조 시스템을 제안한다. YOLO 모델을 활용하여 글 작성 시 삽입되는 이미지에서 객체를 추출하고 메타정보인 GPS 정보, 생성 시간 정보도 추출하여 함께 GPT의 프롬프트 값으로 사용한다. YOLO 모델을 사용하기 위해 양식 이미지 데이터로 학습하여 사용했으며 해당 모델의 mAP score는 평균 약 0.25이다. GPT는 ‘맛집 리뷰’ 주제의 1,000개의 블로그 텍스트 데이터를 학습하였으며, 본 연구에서 학습된 모델을 사용하여 이미지에서 추출한 2가지 타입의 키워드로 문장을 생성하였다. 생성된 문장의 실용성을 평가하기 위해 설문을 진행하였으며 설문 결과의 명확한 분석을 위해 폐쇄형 설문을 진행하였다. 삽입한 이미지와 키워드 문장을 제공하여 질문에 대해 3가지 평가 항목을 두어 진행하였다. 설문 결과 이미지의 핵심 키워드 경우 유의미한 문장을 생성한다는 결과를 얻을 수 있었다. 본 연구를 통해서 이미지 기반 문장 생성 시 이미지 키워드와 GPT 학습 내용과의 관계에 따라 결과물의 정확성이 달라진다는 결과를 얻을 수 있었다.

• 주제어 : GPT, YOLO, 메타정보, SNS, 보조시스템

Abstract In this study, we propose an SNS sentence writing assistance system that utilizes YOLO and GPT to assist users in writing texts with images, such as SNS. We utilize the YOLO model to extract objects from images inserted during writing, and also extract meta-information such as GPS information and creation time information, and use them as prompt values for GPT. To use the YOLO model, we trained it on form image data, and the mAP score of the model is about 0.25 on average. GPT was trained on 1,000 blog text data with the topic of ‘restaurant reviews’, and the model trained in this study was used to generate sentences with two types of keywords extracted from the images. A survey was conducted to evaluate the practicality of the generated sentences, and a closed-ended survey was conducted to clearly analyze the survey results. There were three evaluation items for the questionnaire by providing the inserted image and keyword sentences. The results showed that the keywords in the images generated meaningful sentences. Through this study, we found that the accuracy of image-based sentence generation depends on the relationship between image keywords and GPT learning contents.

• Key Words : GPT, YOLO, Meta data, SNS, Supporting system

Received 22 August 2023, Revised 21 September 2023, Accepted 25 September 2023

* Corresponding Author Bong-Jun Choi, Department of Software, Dongseo University, 47, Jurye-ro, Sasang-gu, Busan, Korea.
E-mail: bjchoi@gdsu.dongseo.ac.kr

I. 서론

2010년대를 들어서며 스마트폰이 현대사회에 널리 보급되었다. 이후 디지털화가 빠르게 진행되어 아날로그 시대에서 디지털 시대로 현대사회는 바뀌었다. 현재 많은 사람은 이미지를 활용한 SNS를 통하여 본인의 경험과 지식, 정보를 공유한다. SNS 활동 중 블로그 활동은 한국에서 2002년 11월 blog.co.kr가 시작하였다. 블로그 작성의 경우 수익 창출이 가능하여 많은 사람이 부업으로도 활용하고 있으며 수익 창출 목적뿐 아니라 일상 기록을 목적으로 많은 사람이 활동하고 있다. 2022년 N 포털의 블로그는 200만 개가 신설되었고 10·20대 월평균 사용자는 17% 증가하였다. 하지만 게시물과 같이 글을 작성하는 활동에 대해서 다수의 사람이 긍정적으로 받아들이지만 29.6%는 글쓰기에 대해 ‘괴로움’, ‘짜증’, ‘두려움’으로 느꼈다. 부정적으로 느낀 사람의 경우 ‘본인의 글 작성 실력이 부족하고 어떤 내용으로 작성해야 하는지 모른다.’와 같은 이유가 83.7%를 차지하였다[1]. 또한 직장인의 경우 게시물 작성 시간도 역시 문제점 중 하나이다. 게시물 1개를 작성하는 데 걸린 시간은 2014년 평균 2시간 24분, 2018년 3시간 28분, 2022년 4시간 10분이 걸렸다[2]. 이러한 문제들이 블로그 활동을 시작하려는 사람 또는 활동 중인 사람에게 영향을 미치고 있다[2].

본 연구는 이러한 문제를 해결하기 위해 SNS에 이미지와 함께 문장을 작성하는 특성을 활용하여 사용자가 게시물 작성 시 삽입하는 이미지를 통해 게시물 작성을 보조하고자 한다. 이미지에서 정보를 추출하고 이를 기반으로 GPT를 활용하여 이미지와 관련된 문장을 생성하여 게시물 작성을 보조하는 시스템을 제안한다. 제안 시스템은 YOLO 모델을 사용하여 사용자가 삽입한 이미지를 분석하여 객체와 이미지의 메타정보 키워드를 추출한다. 추출 키워드를 학습된 GPT 모델의 입력값으로 사용하여 이미지와 관련된 문장을 생성한다.

II. 관련연구

이미지의 객체 정보와 메타정보 기반 GPT 활용 SNS 작성 보조 시스템은 이미지 객체 추출을 위해 YOLO 또는 Vision API를 활용하여 구성할 수 있다.

2.1 YOLO 모델

YOLO(You Only Look Once)는 이미지 또는 영상 내에서 객체를 실시간으로 탐지하여 자동차 자율주행 분야, CCTV에 탑재하여 차량 탐지, 범죄 탐지 등 각종 사고를 예방할 수 있는 기술로 활용되고 있다[3]. 단일 단계 방식의 객체 탐지 알고리즘의 대표 모델인 YOLO는 이미지를 한 번만 보고 이미지에 존재하는 객체와 객체의 위치를 예측할 수 있다. YOLO 모델은 입력받은 이미지와 동일한 프레임의 크기의 그리드로 분할한다. 각 그리드 셀은 N개의 바운딩 박스(Bounding box)와 바운딩 박스에 객체가 어떤 물체의 클래스일 확률 즉, 바운딩 박스의 예측 정확도를 나타내는 confidence score(p_c)를 예측한다. confidence score는 다음 수식으로 정의한다.

$$P_C = P_r(Objct) * IoU_{pred}^{truth}$$
 에서 P_r 은 특정한 바운딩 박스 내에 객체가 존재하지 않는다면 $P_r = 0$ 이고 가장 이상적인 P_r 은 1의 값을 가진다. IoU (Intersection Over Union)은 예측된 바운딩 박스와 실제 물체의 바운딩 박스 사이의 겹치는 영역의 비율을 나타낸다. 일반적으로 0과 1 사이의 값을 가지며 두 바운딩 박스가 완전히 일치할 때 1의 값을 가진다. 그림 1에 YOLO 모델을 사용하여 객체를 추출하는 과정을 나타내었다[4].

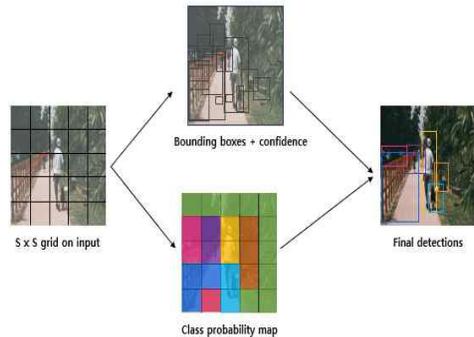


Fig. 1. YOLO Object Detection algorithm

2.2 Vision API

Vision API는 Google에서 제공하는 머신러닝 기반의 이미지 분석 API이다. Vision API에서 제공하는 기능은 객체 탐지 기술을 통해서 이미지의 광학 문자 인식(OCR), 라벨 인식, 얼굴 인식, 객체 위치 인식 등 여러

기능을 제공한다[5]. Vision API는 완성형 모델이기에 객체 탐지를 활용한 시스템을 개발할 경우 해당 시스템의 개발 가능 여부를 판단하기 위한 테스트 모델로 사용하기에 적합하다. 하지만 본 시스템에서 Vision API를 사용하여 객체의 라벨링 정보를 추출할 경우 완성된 모델 API를 활용만 할 수 있어 이미지를 통해 추출된 객체의 라벨링 정확도에 대한 문제가 발생한다. 추가적인 학습을 통해서 정확도를 향상 시킬 수 있는 모델이 아니기 때문에 정확도에 대한 신뢰가 떨어졌을 경우 객체 탐지 모델을 새로 구현하여 본 시스템에 적용해야 한다. 또한 Vision API를 사용하기 위해서는 이미지 요청 수에 따라 추가적인 비용이 발생한다. 그림 2는 Google Cloud에서 사용자의 이미지를 전달받아 가공 후 처리되는 과정을 표현한 그림이다.



Fig. 2. Google Cloud Vision API

III. 이미지 기반 GPT 활용 작성 보조 시스템

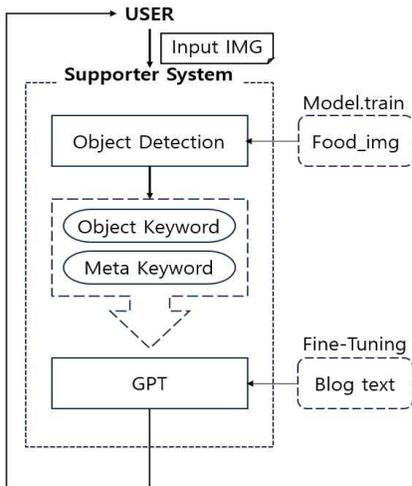


Fig. 3. System concept

본 논문에서 제안하는 이미지 기반 GPT 활용 SNS 작성 보조 시스템은 본 시스템에 사용자가 이미지를 삽입한다. 삽입된 이미지를 수집된 데이터로 학습된

YOLO 모델을 통해서 이미지 객체의 라벨링 정보를 추출하고 이미지의 메타 정보도 추출하여 2가지 타입의 정보들을 해당 키워드들을 활용 목적에 맞게 학습된 GPT 모델의 프롬프트 값으로 사용한다. 본 시스템의 결과물은 해당 키워드들을 GPT 모델에 삽입하여 생성된 문장이다. 그림 3은 본 논문에서 제안한 시스템에 대한 전체 개념도이다.

3.1 YOLO 학습 및 이미지 데이터 추출

본 시스템은 사용자가 ‘맛집 리뷰’ 주제의 블로그를 작성한다는 가정으로 YOLO 모델을 학습한다.

3.1.1 YOLO 학습 및 라벨링 정보 추출

YOLO 모델은 정확도와 속도 측면에서 v5 모델보다 우수하고 실시간 Object Detection 작업에 주로 사용되는 v8 모델을 사용하였으며 세부 모델은 v8 모델 중 성능과 속도 측면에서 준수한 s 모델을 사용하였다. 모델을 학습하기 위해 주제에 알맞은 데이터를 선별한다. Roboflow에서 CC0 라이선스(공개된 저작권 라이선스) 표시가 되어있는 일반 사용자의 라벨링 처리된 양식 이미지 데이터로 학습 데이터 셋을 구성하였다.

학습 조건으로 epoch(학습 수): 150, batch(배치 크기):32, imsz(이미지 크기): 416 값들로 학습하였다. 학습 결과 그림 4를 보면 평균 정밀도 mAP(Mean Average Precision) 값이 학습 10회~20회 사이 평균 약 0.19 값을 나타냈고 학습 중 가장 높은 mAP 값은 학습 49번째에 0.28212의 값으로 가장 높은 평균 정밀도 값을 나타내었다. 해당 모델을 사용하여 이미지에서 객체의 라벨링 키워드를 추출한다. 이미지를 삽입하여 YOLO 모델 결과로 [‘hamburger’ , ‘french fries’]를 추출할 수 있다.

3.1.2 이미지 메타정보 추출

본 시스템은 이미지 객체 라벨링 정보뿐 아니라 이미지의 메타 정보인 GPS 정보와 시간 정보도 추출한다. GPS 정보는 GPS 기능을 허용해 둔 스마트폰을 사용하여 사진 촬영하였을 때 메타 정보에 등록된다[6]. 이를 통해 생성된 GPS 정보는 경위도 각각 [도; 분; 초]의 형식으로 제공되는데 정보를 통해서 해당 건물에 대해서는 파악할 수 있으나 해당 건물에 여러 가게가 있을 경우 정확한 가게 정보를 파악하는 것에는 문

제가 있어 본 연구에서 키워드로 추출하여 사용하기에는 부적합한 정보이다[7]. 본 연구에서는 이미지의 생성 시간 정보는 ['07:00~10:00' : 아침], ['11:00~14:00' : 점심], ['17:00~20:00' : 저녁] 기준으로 해당 시간대에 맞는 키워드로 변환하여 추출한다. 그림5에서 ['2023-05-18 18:19:00']의 시간 정보에 맞는 ['저녁'] 키워드로 변환하여 추출하였다. 결과적으로 ['hamburger' , 'french fries' , '저녁'] 총 3가지 키워드를 추출하였다.

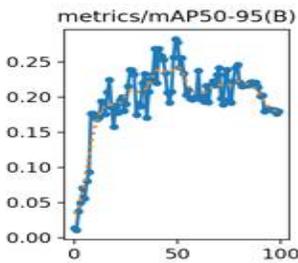


Fig. 4. mAP graph



Fig. 5. YOLO v8 result label

3.2 GPT 모델 학습

본 시스템은 블로그에 삽입한 이미지에 맞는 문장을 GPT를 활용하여 블로그 형식의 글과 동일하게 생성하는 것이다. 따라서 목적에 맞는 데이터를 수집하여 GPT를 목적에 맞게 Fine-Tuning 과정이 필요하다[8].

본 시스템에서 사용할 GPT 모델은 SKT-AI에서 개발한 KoGPT2 모델을 사용한다. GPT의 부족한 한국어 성능을 향상시키기 위해 40GB 이상의 한국어 텍스트로 학습된 한국어 모델이다[9]. KoGPT2 모델로 블로그에 사용되는 문장을 생성하기 위해 Fine-Tuning을 진행한다. Fine-Tuning을 위해 사용할 데이터는 ‘맛집 리뷰’ 주제의 1,000개 블로그 텍스트 데이터를 사용한다. GPT에 raw 데이터로 사전 학습 결과 학습 데이터의 문장 구조와 비슷한 구조로 문장이 생성되었다. 따라서 활용성 있는 문장을 생성하기 위해 2가지 학습

데이터 전처리 과정을 진행한다. 첫 번째 작업은 불필요한 줄바꿈 제거이다. raw 데이터에 불필요한 1줄 이상의 공백이 있어 문장과 문장 사이에 줄 바꿈을 의미하는 이스케이프 문자 ‘\n’ 이 포함되는 경우가 존재하였다. 두 번째 작업은 한 줄에 한 문장 구조로 변환한다. 한 줄에 여러 문장 또는 3줄에 걸쳐 한 문장인 경우 문장이 완성되지 않았음에도 줄바꿈, 무의미한 띄어쓰기 현상이 발생하는 경우도 존재했다. 두 번째 과정은 raw 데이터의 문장을 형태소 분석하여 종결 어미(EF), 마침표(SF), 붙임표(SO) 3가지를 기준으로 구분하였다. 이러한 전처리 과정을 거친 데이터로 100회 반복하여 GPT 모델을 학습한다.

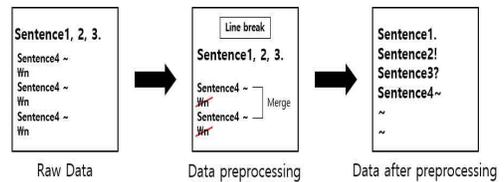


Fig. 6. Data preprocessing

3.3 문장 생성

Section 3.1에서 추출한 객체 키워드와 메타 정보 키워드를 Fine-Tuning 된 GPT 모델에 삽입하여 문장을 생성한다. GPT 모델에서 동일 키워드를 반복하여 문장을 생성하는 것을 방지하기 위해 1.0 ~ 2.0의 값을 가지는 repetition_penalty 값을 설정한다. repetition_penalty 값을 설정하기 위해 ‘하늘’ 키워드로 속성값 1.0, 1.5, 2.0 3가지 값으로 문장을 생성하여 비교하였다. 1.0의 값의 결과는 문장의 끝이 “~어요!”, “~어요~!”, “~네요~!” 로 반복되었고 1.5와 2.0의 값은 반복되는 단어 없이 동일한 문장을 생성하여 1.5의 속성값으로 진행하였다. 문장의 최대 생성 길이를 결정하는 속성인 max_length 값을 학습 데이터의 문장 평균 길이인 50으로 동일하게 설정하여 문장을 생성하였다.

3.4 생성 문장 평가

Section 3.3에서 각 키워드에 대한 문장을 생성하였다. 본 시스템은 이미지를 기반으로 블로그와 같은 SNS 작성을 보조해주는 시스템인 만큼 실질적으로 문장 활용 가능성이 있어야 한다. 실용 가능성을 평가하

기 위해서 본 논문에서는 보기를 보여주는 폐쇄형 설문을 통한 평가를 진행한다[10]. 평가는 52명의 20대를 대상으로 진행하였으며 평가 방식은 다음과 같다. 이미지와 [단어 : 생성 문장]의 쌍으로 보여주고 각 문장에 대해 평가 항목 3가지 중 1가지를 선택하는 방식으로 설문한다. 설문의 질문은 “블로그 작성 시 해당 이미지에 대한 글을 작성할 때 GPT가 생성한 문장을 사용할 수 있습니까?”이다. 평가 항목은 [‘① 사용할 수 있다.’, ‘② 수정이 필요하다.’, ‘③ 사용할 수 없다.’]이며 Google Forms를 활용하여 설문을 진행하였다.

3.5 시스템 수행 결과

본 논문에서 제안하는 시스템을 통해서 Section 3.1의 추출 값인 키워드, Section 3.3의 추출 값인 GPT 생성 문장, Section 3.4에서 실시한 문장 평가 설문 조사 결과 총 3가지 결과물을 도출하였다.

Table 1. Image extraction Keywords

Keywords		
hamburger	french fries	점심

Table 2. Generated sentence

Keyword	Generated Sentence
햄버거	“햄버거집으로 유명한 로스터리 본점!!(한번쯤은 다들 먹어봤을 듯!)”
감자튀김	“감자튀김, 감자, 새우구이, 소시지빵, 텐더, 샐러드, 병아리콩? 그런거 다 있어요 ㅎㅎ”
저녁	“저녁을 카페에서 간단한 식사로 즐기곤 한다.\n\n빵도 맛있지만,카페푸어에서도 먹을 수 있는 빵!이렇게”

Table 3. Survey results for generated sentences

항목 \ 키워드	햄버거	감자튀김	저녁
	사용 가능	41	16
수정 필요	6	25	18
사용 불가	5	11	24

위의 표 1, 2, 3은 각각 Section 3.1, 3.3, 3.4의 결과물이다. 표 1의 추출 키워드는 이미지에서 객체 탐지를 통해 정확한 라벨 정보를 추출하였다. 본 시스템에

서 영어 키워드를 Papago API를 통해 한국어로 번역 후 KoGPT2 모델의 키워드로 사용되었다. 표 2를 통해 번역된 키워드별로 생성된 문장을 확인할 수 있다. 생성된 문장이 실용성이 있는지 확인하기 위해 표 3을 보면 전체 키워드에서 ‘햄버거’의 ‘사용 가능’ 선택이 41명으로 가장 많다. 반면 ‘감자튀김’과 ‘저녁’의 경우는 역시 ‘사용 가능’보다 ‘수정 필요’, ‘사용 불가’ 선택이 가장 많다. 설문자가 설문 조사 시 평가 항목의 선택 이유를 자유롭게 작성할 수 있었는데 ‘감자튀김’의 경우 약 30% 설문자가 ‘햄버거 생성 문장보다는 이미지와의 관련성은 조금 떨어지지만 그렇다고 사용할 수 없는 정도의 문장은 아니다’라는 의견을 가지고 있었다. ‘저녁’의 경우는 약 40% 정도의 설문자가 ‘햄버거 이미지인데 카페 리뷰에 대한 글이다. 주제와 완전히 다른 문장이다.’라는 의견을 가지고 있었다.

IV. 결론

본 논문에서 게시물 작성에 대한 어려움과 작성 시간에 대한 문제를 해결하기 위해 이미지 정보 기반 GPT 활용 블로그 작성 보조 시스템을 구현하였다. 본 시스템을 통해 이미지를 분석하여 정보를 추출하는 과정에서 mAP 값에 비해서 이미지의 라벨링 정보는 매우 정확했다. 그리고 KoGPT2로 생성한 한국어 문장은 학습을 통해서 생각보다 자연스러운 문장을 생성하였다. 하지만 YOLO의 경우 YOLO 공식 문서에서 제공한 모델 성능표와는 다르게 mAP 수치가 조금 부족하였다. 0~100 사잇값을 기준으로 평균 44.9로 공개하였으나 이번 연구에서 가장 높은 값이 28.2의 값으로 아쉬운 성능을 보여주었다[11]. 그리고 생성 문장의 경우 ‘햄버거’를 제외한 ‘감자튀김’, ‘저녁’의 경우는 주제와는 관련이 있으나 이미지와 연관성이 조금 부족하여 무의미한 결과를 도출하게 되었다. YOLO의 경우는 조금 더 정확성 있는 v8의 l, x모델을 사용하여 조금 더 다양한 데이터 셋을 활용하여 생성한다면 더 좋은 성능을 가진 모델을 만들 수 있을 것이다. 그리고 GPT문장 생성 성능을 높이기 위해 GPT를 학습시켜 활용하기보다 대화형 모델인 ChatGPT 모델을 사용하여 이미지 키워드로 질문 문장을 만들어 ChatGPT 모델에 삽입한다면 더욱 연관성 있는 문장 결과를 얻을 수 있다고 생각한다.

ACKNOWLEDGMENTS

이 논문은 2023년도 동서대학교 “Dongseo Cluster Project” 지원에 의하여 이루어진 것임 (DSU-20230005)

REFERENCES

- [1] Keum Youn Han and Seon Kyung Choi, “A Research on the responses and demands of freshman about writing education as college’s required course”, Korean Journal of General Education, Vol. 10, No. 3, (33), pp. 549-576, 2016
- [2] A. Crestodina (2022, September 22), New Blogging Statistics: What Content Strategies Work in 2022? We asked 1016 Bloggers. [Online]. Available: <https://www.orbitmedia.com/blog/blogging-statistics>
- [3] Yong Hwa Jo, Hyuek Jae Lee, Young Hun Kim, “Implementation of a Classification System for Dog Behaviors using YOLI-based Object Detection and a Node.js Server”, KICSP, Vol. 21, No. 1, pp. 29-37. 2020
- [4] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, “You Only Look Once: Unified, Real-time Object Detection.”, IEEE Conference on, pp. 779-788, 2015
- [5] Sayak Paul (2018, December 10), Beginner’s Guide to Google’s Vision API in Python. [Online], Available: <https://www.datacamp.com/tutorial/beginner-guide-google-vision-api>
- [6] JEIDA, “Digital Still Camera Image File Format Standard(Exchangeable image file format for Digital Still Cameras: Exif) Version 2.1” JEIDA, 1998
- [7] Sung Jin Hong, “A Study on Manipulation Detection of Exif GPS Information in Photographic Files”, KDFS, 5, No.1, pp. 41-54, 2011
- [8] A. Radford, K. Narasimhan, T. Salimans, I. Stuskever “Improving Language Understanding by Generative Pre-Training”, NIPS, 2018
- [9] wicidocs, (2022, Jan 13) Creating a koGPT2 chatbot, [Online], Available: <https://wikidocs.net/157001>
- [10] Ji Bum Kim, So Ri Kim, Jeong Han Kang, “Survey Experiment on Close-Ended and Open-Ended Questions: 2016 Korean General Social Survey

- (KGSS)”, KASR, Vol. 18, No. 4, pp. 127-147, 2017
- [11] (2023, April 25), Check out YOLO V8 performance. [Online], Available: <https://developer-lionhong.tistory.com/62>, <https://docs.ultralytics.com/>

저자소개

이 동 희 (Dong-Hee Lee)



2018년 3월~현재 : 동서대학교
소프트웨어학과(학사)
관심분야 : 웹, WebRTC, GPT,

문 미 경 (Mikyeong Moon)



1990년 이화여자대학교
전자계산학과 졸업(이학사)
1992년 이화여자대학교 대학원
전자계산학과 졸업(이학석사)
2005년 부산대학교 대학원
컴퓨터공학과 졸업(공학박사)

2008년 동서대학교 컴퓨터공학부 교수
관심분야 : 소프트웨어공학, AI융합기술 응용

최 봉 준 (Bond-Jun Choi)



2013년 2월 : 동서대학교
컴퓨터정보공학부(공학사)
2015년 2월 : 연세대학교
컴퓨터과학과(공학석사)
2014년 12월 ~ 2019년 2월 :
LG 전자
2019년 2월 ~ 2021년 1월 :
Aalborg University Operation
Research Lab

2021년 3월 ~ : 동서대학교 소프트웨어융합대학 조교수
2022년 12월 ~ : ㈜구현솔루션 대표이사
관심분야 : 데이터 분석, 데이터 생성, GAN, 자율주행,
중등SW교육