

인공지능 기술을 활용한 부동산 허위매물 필터링 시스템*

문 정 경*

요 약

최근 온라인을 중심으로 광고를 수행하고 오프라인에서 실제로 물건을 매매하는 O2O 기반의 부동산중개 웹 사이트 혹은 앱이 폭발적으로 증가하고 있는 추세이다. 이로 인해서 기존의 오프라인 기반의 부동산중개 환경에서 온라인 기반으로 환경이 변화됨으로써 소비자들이 시간, 비용, 편리성 측면에서 매우 높은 호감을 얻고 있다. 하지만, 온라인 기반의 부동산중개 서비스들의 편리함 이면에 잘못된 정보 또는 악의적인 허위정보로 인해서 사용자가 시간, 금전적으로 피해를 보게 되는 경우도 자주 발생하고 있다. 그러므로 본 연구에서 O2O 기반의 부동산중개 서비스에서 발생 가능한 소비자의 피해를 줄이고자 인공지능 기술을 활용해 등록된 매물정보에 대한 진위 여부를 판별할 수 있는 허위 매물정보 필터링 시스템을 설계하였다. 제안한 연구방법을 통해서 온라인 부동산 서비스에 등록되는 매물정보에 대한 진위 여부를 판별할 수 있을 뿐만 아니라, 소비자의 시간적, 금전적 피해를 줄일 수 있음을 보였다.

A Design of Estimate-information Filtering System using Artificial Intelligent Technology

Jeong-Kyung Moon*

ABSTRACT

An O2O-based real estate brokerage web sites or apps are increasing explosively. As a result, the environment has been changed from the existing offline-based real estate brokerage environment to the online-based environment, and consumers are getting very good feelings in terms of time, cost, and convenience. However, behind the convenience of online-based real estate brokerage services, users often suffer time and money damage due to false information or malicious false information. Therefore, in this study, in order to reduce the damage to consumers that may occur in the O2O-based real estate brokerage service, we designed a false property information filtering system that can determine the authenticity of registered property information using artificial intelligence technology. Through the proposed research method, it was shown that not only the authenticity of the property information registered in the online real estate service can be determined, but also the temporal and financial damage of consumers can be reduced.

Key words : Estimate Information, Filtering System, AI, Web-crawling, Deep learning

접수일(2021년 03월 02일), 수정일(1차: 2021년 03월 13일),
게재확정일(2021년 03월 30일)

* 호서대학교 혁신융합학부(주저자)

★ 본 논문은 2020년도 호서대학교의 재원으로 학술연구비 지원을 받아 수행된 연구임(과제번호: 20200409).

1. 서 론

초고속 인터넷 환경과 스마트폰 공급의 급격한 증가로 우리 삶에 많은 변화들이 발생하고 있다. 그 중에서도 소비자가 원하는 부동산 매물에 관한 정보를 빠르고 정확하게 검색할 수 있게 되었다. 특히 기존에 오프라인에서 부동산 중개업자를 만나서 원하는 부동산 매물정보를 확인하고 직접 선택한 부동산 매물에 찾아가서 물건을 확인해야만 했던 시간적·비용적인 불편함이 크게 해소되었다.

온라인 부동산 서비스 앱이나 웹 게시판을 통해서 고객이 필요로 하는 정보를 제공하고, 오프라인으로 실제 부동산 매매로 이어질 수 있도록 하는 O2O(Online to Offline) 서비스는 사용자 편리성을 매우 높였다. 높은 부동산 중개 앱 혹은 플랫폼 서비스가 국내·외에서 다양하게 출시되어 사용자들의 편의성을 높이는 데 기여하고 있다. 국내에서는 2012년부터 부동산 매물정보를 스마트 앱과 웹 플랫폼 서비스가 제공되면서 국내·외에서 가장 발전된 서비스를 제공하고 있다[1].

하지만 위에서 언급한 바와 같이 국내에서 다방, 직방, 방콜과 같이 대표적인 3대 스마트 앱과 네이버 부동산, 부동산114와 같은 웹 플랫폼 서비스들이 잘못된 정보(허위 매물정보)를 제공함으로써 인해서 발생하는 부작용이 증가하고 있다[2].

국내의 대표적인 O2O부동산 서비스를 제공하는 업체들의 스마트 앱이나 웹 플랫폼에서 발생하고 있는 허위매물 정보의 유형으로는 (1) 공개된 매물의 주소, 사진 정보들이 실제와 일치하지 않는 경우, (2) 매물에 대한 실제 시세보다 확연히 저렴한 가격 정보를 게시하는 경우, (3) 이미 거래가 완료된 매물을 아직도 거래가 가능한 것처럼 게시하는 경우, (4) 실제 매물이 갖추고 있는 옵션 사항과 게시된 매물정보의 옵션이 다른 경우이다[3][4].

그리고 선행연구자들의 연구 결과를 자세히 검토해본 결과, 위와 같은 부동산 허위매물정보가 해결되지 않는 원인은 아래와 같이 3가지로 요약할 수 있다.

(1) 현행 부동산 매물정보의 게시 및 관리는 사단법인 부동산중개업협회에서 자율적으로 관리하도록 할 뿐 법적으로 부동산 매물정보에 대한 법적 게시 기

준 및 허위 매물정보에 대한 법적 제재 기준이 상이하 다[5].

(2) 국토교통부에서 공인중개사법에 따라서 부동산 매물정보에 대한 규제 및 감시를 수행하고 있기는 하나, 금지행위에 대한 법적 제재나 처벌 등에 대해서는 강제성을 가지고 있지 않다.

(3) 부동산 매물정보에 대한 사전 검열방식은 공정거래위원회에서 제안한 표시·광고의 공정화에 관한 법률에 의해서 표현의 자유, 재산권에 대한 침해로 헌법재판소의 결정에 의해서 금지되고 있다.

위의 사항들에 대해서 소비자원이 지난 2016년 관련 실태조사를 통해서 문제점을 파악하고, 이에 대한 해결 가능방안에 대해서 법적, 사회적 인식 변화의 필요성을 충분히 제시되었다. 하지만 기술적 해결 방안은 아직 미비한 실정이다. 이를 해결하기 위해서 먼저 법적인 근거 마련, 사회적 인식의 변화, 부동산중개 관련 업자들의 자발적인 허위매물 정보에 대한 개선을 위한 노력이 필요하다[5].

그러므로 본 연구에서 인공지능 기술인 딥러닝 기술을 활용해서 허위매물 정보들 중에서 여러 스마트 앱에 존재하는 동일한 부동산 매물정보들을 주기적으로 추출하여 삭제하고자 한다. 뿐만 아니라, 국토교통부와 같은 제3의 신뢰 기관이 부동산 중개업자와 부동산 매물에 대한 신뢰할 수 있는 인증체계를 마련할 수 있는 방안에 대해서 제시하고자 한다.

본 논문의 구성은 다음과 같다. 2장은 관련연구로 기존에 다방, 직방과 같은 스마트 앱의 개발현황과 웹 플랫폼 서비스의 특징과 장·단점에 대해서 검토한 내용을 요약한다. 3장은 제안하고자 하는 연구 방안을 2가지로 구별해서 제안한다. 첫 번째는 인공지능 기술을 기반으로 한 허위매물 정보 추출에 대한 것이고, 두 번째는 제3의 신뢰 기관을 중심으로 한 부동산 매물정보에 대한 인증체계에 대한 제안이다. 4장은 앞서 제안한 연구방법들에 대한 설계와 구현 내용을 기술한다. 마지막으로 앞서 제안한 연구방법들의 현실적 기여도에 대해서 살펴본다.

2. 관련연구

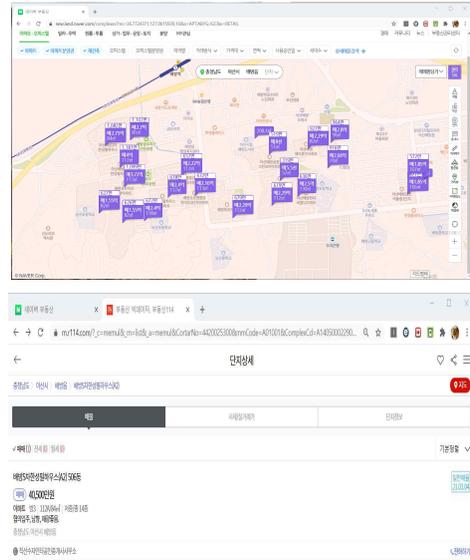
2.1 부동산 웹-플랫폼 서비스

국내에서는 네이버 부동산, 부동산114 등이 웹 서비스를 기반으로 한 부동산 웹-플랫폼 서비스이다. 이와 같은 부동산 웹 플랫폼 서비스는 부동산 중개업자들이 해당 웹 플랫폼에 매매하고자 하는 매물정보를 매물의 형태에 따라서 전세, 매매와 같이 2개 종류로 구분해서 등록하게 된다. 이때 부동산 중개업자가 부동산 소유주로부터 전달받은 올바른 매물정보를 등록하고, 거래가 완료시 해당 매물정보를 웹 플랫폼에서 삭제하여야만 한다. 이런 과정 중에 부동산 소유자가 여러 부동산 중개업소에 해당 매물을 거래 요청하는 경우 또는 부동산 중개업자가 이미 거래된 부동산 매물정보를 곧바로 삭제, 관리하지 않음으로 인해서 부동산 매물정보에 대한 오인에 관한 문제가 발생할 수 있다.

이런 부동산 웹 플랫폼 서비스는 기존의 오프라인 부동산중개 거래와 비교해서 온라인으로 원하는 매물정보를 확인하고 오프라인으로 부동산 중개업자와 만나서 실제 매물을 확인한 후 거래를 수행한다. 그리고 부동산 중개업자에게 수고한 부분을 보상하는 중개수수료를 제공함으로써 서비스가 종료된다. 하지만 이런 과정 중에서 많은 부동산 중개업자들이 정보의 공개성, 과열 경쟁 등의 이유로 인해서 부동산 중개물에 대한 잘못된 정보를 고의적 혹은 실수로 제공하게 되는 경우가 수시로 발생할 수 있다는 단점을 가지고 있다. 이와 같은 현상이 발생하는 이유는, 기존의 오프라인 중개업소들만이 경쟁하던 환경과 비교해서 온라인 서비스를 통해서 고객이 원하는 정보를 다른 중개업자보다 빠르고 편리하게 제공하는 것 이외에 경쟁력을 확보할 수 없기 때문인 것으로 이해된다. 뿐만 아니라 부동산 중개업자 입장에서 한 달에 웹 플랫폼 서비스 제공자에게 매달 납부하는 정보 공개 청구요금은 동일한데, 부동산 매물정보를 최대한 많이 제공하여 고객들을 호객하는 것이 경제활동에 조금이라도 도움이 되는 어쩔 수 없는 현실 때문인 것으로 파악된다. 그리고 중개업소마다 법적으로 허용하는 범위 내에서 중개수수료를 조금이라도 적게 받는다는 식으로 고객들을 호객함으로써 자체적인 내부 경쟁이 심화되어 경제활동에 많은 어려움이 있는 것으로 파악된다.

대표적인 2가지 부동산 웹-플랫폼 서비스는 유사

한 점도 많지만, 서로의 장점들을 자랑하고 있다. 각각의 서비스에 따라서 좀 더 많은 부동산 매물정보를 제공하고 있다는 점과 사용자 편의성을 최대한 고려하였음을 주장하고 있다.



(그림 1) 네이버부동산과 부동산114

2.2 Deep learning system

국내에서는 10년 전부터 부동산 중개 플랫폼과 앱들에 대한 연구가 활발하게 진행되고 있다. 특히 다방, 직방으로 대표되는 2가지 앱은 스마트폰을 기반으로 널리 알려진 사용되고 있는 추세이다. 하지만 이런 부동산 매매 플랫폼 및 앱들은 보안상 많은 허위정보를 배포하는 부작용이 발생하고 있다[6][7].

이를 해결하기 위해서 많은 국내외 연구자들이 인공지능 기술, 그중에서도 딥러닝 기술을 활용하여 허위매물 정보에 대해 찾아내고자 하는 노력이 같이 급증하고 있다. 하지만 부분적인 연구가 이뤄지고 있을 뿐 아직까지 광범위한 딥러닝 기술을 탑재한 부동산 허위매물 정보 필터링 시스템이 존재하지 못하고 있다. 이에 부동산 허위매물을 골라낼 수 있는 딥러닝 기술에 대해서 조사한 결과를 요약한다.

부동산 매물정보는 2가지 측면으로 구분해서 허위 정보 여부를 판별할 수 있다. 첫번째는 부동산 매물정

보에 대한 문자열 정보를 판단하는 기술과 두 번째로 부동산 매물정보에 대해서 위치정보나 구성정보를 사진으로 나타내는 것이다[8].

이에 본 연구를 수행하면서 사진 연구를 검토해 본 결과, 부동산 114나 네이버 부동산 등과 같은 웹 플랫폼에서 크롤링을 통해서 수집된 문자열 정보를 기준으로 허위매물 여부를 판별할 수 있는 것으로 파악되었다. 두 번째로 네이버 부동산의 경우 부동산 매물에 대한 구성정보를 사진으로 보여주는 과정에서 허위정보를 구별해 낼 수 있을 것으로 파악되었다[9].

딤러닝 처리 알고리즘 중에서 KLDP를 활용한 부동산 정보에 대한 문자열들을 활용해서 동일 정보에 대한 반복이나 일부분 위.변조 사례에 대한 연구를 찾아볼 수 있었다. 특히 참고문헌 8에서는 부동산 매물 정보들에 대해서 독립된 분석 기호를 만들고, 이를 기준으로 동일하거나 유사한 문자열 정보로 구성된 허위 매물정보를 등록 기준 날짜를 기준으로 진위 여부까지는 완벽하게 파악하지 못하고 있으나, 허위정보일 가능성이 있음을 판별할 수 있는 방법을 제안하였다 [8][10][11].

두 번째로 참고문헌 12번에서는 GAN 알고리즘을 활용해서 부동산 매물에 대한 사진 정보들을 기준으로 유사한 사진이나 등록정보를 찾아낼 수 있는 방법을 참고하였다[12].

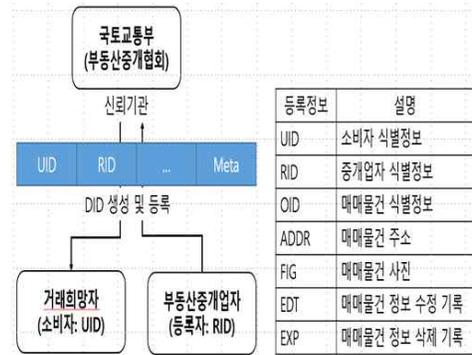
하지만 위와 같이 부동산 매물정보에 대한 부분적인 허위여부를 판별하거나 허위정보일 가능성을 파악하는 수준이지, 실질적인 서비스 가능성 여부는 매우 낮은 것으로 파악된다. 그러므로 본 연구에서는 국토교통부와 같은 신뢰 기관을 활용해서 전국의 부동산 매물들을 사전에 유일한 DID 등록정보를 등록, 수정, 관리할 수 있도록 하는 시스템을 설계 및 제안하고자 한다.

3. 제안시스템

3.1 제안시스템

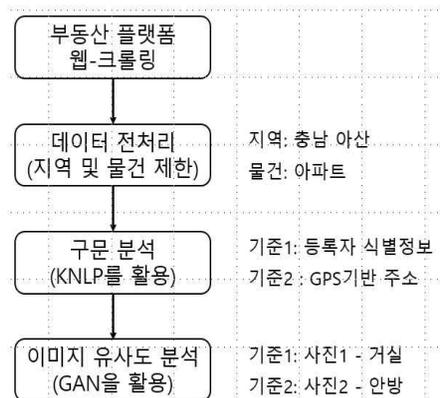
제안시스템은 (그림 2)와 같이 구성할 수 있다. 제안시스템은 독립된 DID를 갖도록 설계하였다. 이를 통해서 부동산 매물마다 고유의 등록정보를 사전에 가지도록 구성하였고, 고유 DID와 GPS 기반의 위치

정보, 부동산 구성정보, 사진 정보를 구성 필드로 갖는다.



(그림 2) 제안시스템 구성 및 등록정보

(그림 3)은 GAN을 사용한 부동산 구성의 사진 정보들을 기준으로 허위매물에 대한 검증방법을 보여준다. 지역과 물건을 충남 아산, 아파트로 제한하고, 문자열 기준으로 구문분석을 수행한다. 그리고 GAN 알고리즘을 통해서 거실 사진1과 안방 사진2로 제한해서 유사도를 비교하고자 한다.



(그림 3) 검증절차

3.2 부동산 매물의 진위여부 파악절차

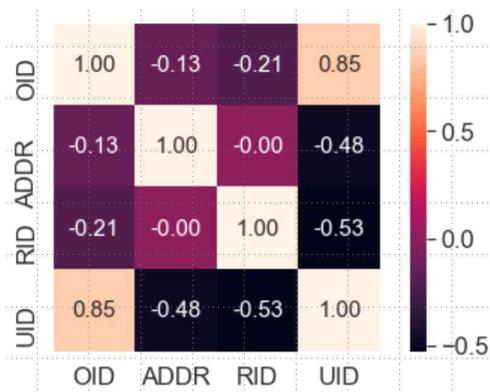
(그림 3)에 나타난 바와 같이 네이버 부동산 웹 플랫폼에 게시된 매물정보들을 수집하여 전처리 과정을

거친 후, 사전에 본 연구에서 제안한 식별기호들을 부여하고, 해당 식별기호를 바탕으로 1차 진위여부를 파악한다.

2차로 식별된 매물정보들 중에서 사진 정보를 중심으로 GAN알고리즘으로 학습시킨 필터링 시스템을 사용하여 1차 진위여부 파악에서 원본으로 등록된 사진과 1대 1로 유사성 여부를 파악하도록 시스템을 설계하였다. 이때 오역전과 알고리즘을 사용해서 잡음을 개선하고자 노력하였다.

4. 실험결과 및 고찰

본 연구에서 제안한 부동산 웹 플랫폼에 게시된 매물정보에 대해서 구글 크롤링을 통해서 정보를 수집하고, 전처리 과정에서 중복된 제목 필드를 갖는 정보나 등록자와 게시자가 다른 매물정보에 대해서는 필터링을 수행하였다. 해당 실험을 위해서 파이썬 아나콘다 주피터 노트북 환경에서 구문 분석(유사도 검사)한 결과를 seaborn 라이브러리를 사용해 피어슨 상관계수로 유사도를 분석하였다. 이를 다음과 같이 heatmap 그래프로 시각화 하였다.



(그림 4) 문자열 유사도 검색 결과

전처리 과정을 마친 매물정보들에 대해서 CSV 확장자를 갖는 파일로 문자열 정보들을 저장한 후, KNLP에 등록된 매물정보에 대한 한국어 인식 알고리즘을 사용해서 유사성을 가진 허위 매물정보들을

검색한 결과, 유사한 매물정보로 등록되었으나 게시자가 다르거나 주소지 정보가 다른 허위 매물정보를 15% 이상 검색할 수 있었다. 특히 주소지가 같은 매물에 대해서 등록된 소유자 정보가 다르거나 한 경우가 가장 높은 허위정보를 차지하고 있음을 알 수 있었다.

뿐만 아니라 2차로 이미지 판별 프로세스를 통해서 문자열 필터링에서 추출된 허위 매물정보들에 대해서 이미지 필터링 실험을 수행해 본 결과 약 80% 이상의 유사성을 가짐을 알 수 있었다.

본 연구에서 수행한 실험은 단순히 매물정보에 대한 문자열 유사성 판별과 이미지 단순 비교 알고리즘을 사용해서 본 실험을 진행하였다. 향후 단계별 문자열 유사성 판별 절차를 세분화하고, 이미지 분석을 위한 전문가들의 접근 방법을 적용함으로써 유사성 판별 능력을 개선할 수 있을 것으로 파악되었다.

5. 결론

인터넷과 사물인터넷 환경의 발전으로 인해서 부동산 매매거래 시장에도 급격한 변화가 발생하고 있다. 특히 부동산 매물정보를 사전에 인터넷 플랫폼이나 앱을 활용해서 확인하고, 자신이 원하는 부동산거래 후보 매물들을 정한다. 그리고 실제로 부동산 중개업자와 오프라인으로 혹은 실시간 온라인 회의 앱을 활용하거나 해서 부동산 정보를 체험하게 한 후 실질적인 거래를 수행하도록 하는 O2O의 활용 정도가 급격하게 증가하고 있는 추세이다.

하지만 이런 부동산 플랫폼 서비스에 공개되는 많은 부동산 매물정보들이 허위인 경우가 많다. 이로 인해서 사용자들이 자주 시간과 경제적인 피해를 자주 입는다. 이에 본 연구에서는 인공지능 기술들 중에서 딥러닝 기술인 GAN 알고리즘을 활용해서 부동산 플랫폼에 공개된 정보들을 크롤링한 후 이것들 중에서 허위매물들을 검색해 내는 필터링 시스템을 설계 및 제안하였다.

본 연구에서 제안한 연구방법을 적용함으로써 기존에 네이버 부동산 웹 플랫폼에 공개된 매물정보들을 6개월 동안 수집하고 전처리를 통해서 중복되거나 부정확한 정보들은 삭제한 후, 전처리를 마친 정보들 중

에서 부동산 매물정보에 사진 정보가 포함된 매물정보들 중에서 약 20% 이상의 부동산매물에 허위 정보가 존재함을 알아낼 수 있었다.

하지만 본 연구에서 아직 연구하지 못한 부분들이 많다. 특히 허위정보 여부를 파악하기 위해서 영상 판별의 방법들이 다양성 측면에서의 연구 시도가 더욱 필요할 것이다. 뿐만 아니라 사진 정보와 문자열 정보를 함께 학습시킴으로써 진위 여부의 판별력을 높이는 연구도 추가로 필요하다는 것을 알게 되었다.

뿐만 아니라 국내 네이버 부동산 플랫폼 서비스뿐만 아니라 국외 부동산 플랫폼 서비스에 대한 매물정보들을 좀 더 긴 시간 동안 수집하고 전처리 과정에서 발생할 수 있는 오류를 줄이는 방법에 대한 필요성도 인식하게 되었다. 향후 연구를 통해서 이런 문제들을 좀 더 개선하고자 한다.

참고문헌

- [1] 한국소비자원 편집부, “모바일 부동산 중개 서비스 이용 관련 실태조사”, 조사보고서, pp.1-64, 2016.05.
- [2] 황창근, “인터넷상 부동산 허위매물 광고 규제 의미와 한계”, KISO저널, 제40권, pp.18-22, 2020.10.
- [3] 서인숙, “O2O서비스 기반의 부동산중개 앱에 관한 실증연구”, 융복합지식학회논문지 제7권 제1호, 2019.1, pp.87-93
- [4] 최유진, 김승인, “모바일 부동산 서비스 검색 필터와 정보품질에 관한 연구 - 국내, 미국, 영국 사례를 중심으로”, 한국융합학회논문지, 제8권 제9호, pp.249-254, 2017.
- [5] 이상식, “모바일 부동산중개에서의 소비자 보호방안”, 국토연구원, 국토, pp.16-22, 2018.04.
- [6] 송인방, 양영식, “부동산거래에서 블록체인 스마트계약의 활용 가능성에 관한 연구”, 법학연구, 제18권 제4호, pp.1-26, 2018.12.
- [7] 한현구, 이은지, 박현주, “블록체인 기반의 부동산 거래 플랫폼”, 한국정보과학회 학술발표논문집, pp.1675-1677, 2019.12.
- [8] 장현준, 송규원, 김영훈, 김선겸, “온라인 부동산 플랫폼 실시간 매물 분석 시스템 개발”, 한국정보과학회 학술발표논문집, pp.1162-1164, 2020.7.

- [9] 문정경, 김진묵, 박성우, “BI기법을 이용한 부동산 허위매물 필터링 시스템”, 한국정보통신학회 종합학술대회 논문집, 제24권 제1호, pp.493-495, 2020.7.
- [10] 오서영, 이창훈, “부동산 시장의 신뢰성 향상을 위한 블록체인 응용 기술”, 한국전자거래학회지, 제22권 제1호, pp.51-64, 2017.
- [11] 권순재, 김성현, 탁운식, 정현희, “K-means clustering 알고리즘과 헤모닉 모형을 활용한 서울시 연립 다세대 군집분류 방법에 관한 연구”, 지능정보연구, 제23권 3호, pp.95-118, 2017.09.
- [12] 김대희, 최승완,곽수영, “딥 러닝 기반의 가짜 얼굴 검출”, 한국산업정보학회논문지, 23권 5호, pp.9-17, 2018.10.

[저자 소개]



문 정 경(Jeong-Kyung Moon)
 2013년 2월 : 공주대학교 컴퓨터공학과(공학박사)
 2012년 3월 ~ 2018년 2월 :
 선문대학교 IT교육학부 계약교수
 2018년 3월 ~ 2020년 2월 :
 가천대학교 소프트웨어중심대학
 초빙교수
 2020년 3월~현재 : 호서대학교
 혁신융합학부 조교수
 관심분야 : 에지 컴퓨팅, 정보보안,
 빅데이터 분석, 인공지능, 스마트 자동차
 E-mail : jkmoon@hoseo.edu