

<https://doi.org/10.7236/JIIBC.2023.23.3.223>

JIIBC 2023-3-30

확장된 개인 데이터를 활용한 OTT 서비스 추천 시스템

Recommendation System of OTT Service using Extended Personal Data

유희정*, 김능회**

HeeJung Yu*, Neunghoe Kim**

요약 정보통신정책연구원에 따르면 OTT 서비스가 처음 등장한 2017년도로부터 4년간 33.4%라는 성장률을 보였다. 또, 한국수출입은행은 지난 2020년 국내 OTT 시장이 7801억원 규모라 발표하였다. 이런 OTT 시장의 발전은 OTT 서비스 플랫폼 간의 경쟁을 부추겨 OTT 서비스의 영상 추천 등과 같은 편의 기능에 대한 이용자 만족이 OTT 서비스 시장 경쟁에서 중요한 요소로 작용할 것으로 보인다. 현재 OTT 시장은 사용자의 맞춤형 추천을 위해 여러 데이터를 사용하고 있지만 앱 내부에서의 데이터만을 사용했다는 한계가 있다. 이에 개개인의 맞춤형 추천을 위한 앱 외부의 사용자 개인 데이터 활용을 제안하였으며, 설문 조사 결과 사용자의 만족도는 넷플릭스 추천 콘텐츠 대비 제안한 방법을 토대로 한 추천 콘텐츠가 23.72% 더 높은 것으로 나타났다.

Abstract According to the Korea Information Society Development Institute, OTT services grew at a rate of 33.4% in four years from 2017, when they were first launched. The Korea Export-Import Bank announced in 2020 that the domestic OTT market was worth 780.1 billion KRW. This growth of the OTT market is expected to stimulate competition among OTT service platforms, and user satisfaction with convenience features, such as video recommendations, seems to be acting as an important factor in the competition. Currently, the OTT market uses a variety of data for customized recommendations, but the limitation is that it only uses data collected within the app. Thereby we have proposed the use of personal data collected outside the app for personalized recommendations, and the survey results showed that user satisfaction was 23.72% higher for recommended content based on the proposed method than Netflix recommended content.

Key Words : ott(over-the-top) service, recommendation system, personal data

1. 서론

OTT(Over-The-Top)란 인터넷을 통해 방송 프로그램·영화·교육 등 각종 미디어 콘텐츠를 제공하는 서비스

를 말한다.^[1-2] 대표적으로 넷플릭스, 유튜브, 페이스북, 트위터, 틱톡 등이 있다. 지난 5월 30일 정보통신정책연구원(KISDI)의 'OTT 무료 및 유료(단·복수) 이용자 비교 분석' 보고서에 따르면 OTT가 처음 등장한 2017년도의

*학생회원 군산대학교 소프트웨어학부

**정회원 군산대학교 소프트웨어학부(교신저자)

접수일자 2023년 3월 29일, 수정완료 2023년 4월 29일

게재확정일자 2023년 6월 9일

Received: 29 March, 2023 / Revised: 29 April, 2023 /

Accepted: 9 June, 2023

**Corresponding Author: nunghoi@kunsan.ac.kr

Department of Software Science & Engineering, Kunsan National University, Korea

OTT 이용률은 36.1%였지만, 2021년도에는 69.5%로 4년간 33.4%의 성장률을 보였다.^[3] 국내에서 10명 중 7명이 OTT 서비스를 사용하고 있는 셈이다. 또, 한국수출입은행은 지난 2020년 글로벌 OTT 시장 규모는 1100억 달러(약 133조 8400억), 국내 OTT 시장이 7801억원 규모라 발표했으며,^[4] 디지털TV리서치는 2021년 1350억 달러 규모였던 전 세계 OTT 시장이 2027년에는 2240억 달러로 약 2배 이상으로 성장한다고 전망하고 있다.^[5] 이처럼 OTT 서비스는 시간이 지날수록 무섭게 성장하고 있다.

이런 OTT 시장의 발전은 기존 미디어 시장의 발 빠른 변화를 요구하였으며, OTT 서비스 플랫폼 간의 경쟁을 부추겨 각 OTT 서비스의 차별화 전략과 서비스 개선이 시급한 시점이라 볼 수 있다.^[6-7] 특히, OTT 서비스의 영상 추천, 검색, 영상 정보와 같이 이용자의 편의를 돕는 편의 기능에 대한 이용자 만족은 OTT 서비스 시장 내 경쟁에 있어서 중요한 요소로 작용할 것이며, 궁극적으로 이용자의 지속적 이용 의도에 영향을 미칠 수 있는 서비스적 특성이 될 수 있을 것으로 보인다.^[8-9]

하지만 기존의 MZ세대를 대상으로 한 연구에서 사용자들은 OTT 서비스의 영상 콘텐츠 추천에 대한 만족도가 크지 않은 것으로 나타났다.^[10] 이에 자체적으로 OTT 서비스인 넷플릭스 추천 시스템에 관한 소수의 만족도 조사를 진행했다. 인원은 총 4명이며 각 인원에게 넷플릭스 홈에서 사용자를 위한 추천 콘텐츠에 해당하는 배너 '○○ 님의 취향 저격 베스트 콘텐츠'(이하 취향 저격 콘텐츠) 중 보고 싶은 콘텐츠를 선택하게 하였다. 이때 취향 저격 콘텐츠의 전체 콘텐츠 개수는 사용자마다 다르며 조사에 참여한 사용자 평균 39.5개로 나타났다. 이를 백분율로 환산한 만족도는 평균 40.93%이다. 위 조사는 현재 OTT 서비스의 개인을 위한 맞춤형 추천 기능을 개선할 필요가 있다는 것을 알린다.

한국언론진흥재단에 따르면 대표적인 OTT 서비스 중 하나인 유튜브의 추천 알고리즘이 사람들이 무엇을 보는지, 얼마나 많은 사람이 해당 영상에서 시간을 보내는지, 얼마나 빠르게 인기가 높아지는지, 얼마나 새로운 영상인지, 사람들이 플랫폼에서 얼마나 많은 시간을 보내는지, 좋아요, 싫어요, 공유 등의 피드백을 종합하여 작동한다고 한다.^[11] 이 데이터들은 앱 내부의 사용자 활동과 영상 정보 등을 수집하여 활용하였다는 특징이 있다. 이에 추천하는 과정에서 단순히 앱 내부의 사용자 활동 데이터나 영상 정보 등의 내부 정보가 아닌 앱 밖 사용자의 개인 데이터를 활용하여 추천할 필요가 있다.

II. 배경

『콘텐츠 선호 모형을 결합한 행렬 분해 기반 영화 추천시스템』에서는 추천 알고리즘의 성능이 낮은 것을 보완하기 위하여 협업 필터링에 기반한 행렬 분해 방법론에 콘텐츠 기반 필터링에 사용되는 아이템에 대한 메타데이터 정보(배우, 감독, 장르)를 결합한 하이브리드 추천 기법(CPMF)을 제안했다.

위 논문에서는 추천 개선을 위해 과거 시청 이력을 토대로 사용자의 배우, 감독의 선호도를 계산하여 가중치를 주고, 최종적으로 상위 n개의 영화를 추천할 때 사용자의 장르 선호도가 높은 장르를 대상으로 추천 리스트를 구성한다. 사용자의 선호 여부를 과거 시청 이력으로 판단하기 때문에 정말 사용자가 선호하는 데이터인지 추천하기에 미흡하다. 또, 추가적으로 사용되는 메타데이터 정보가 OTT 서비스 내에 내장된 정보는 OTT 서비스 외부의 방대한 사용자 개인 데이터와 비교했을 때 사용자의 개인적인 부분을 담고 있다고 보기 어렵다는 한계가 있다.^[12]

『메타 가중치 학습을 활용한 내용 기반의 맞춤형 영화 추천시스템 설계 및 구현』은 영화를 고를 때 일반적으로 중요시 보는 속성들을 활용하여 영화의 특징 벡터를 구성하고, 이를 사용자 평가와 결합하여 개인의 선호에 대한 특징 벡터를 구성하는 방법을 제안했다. 이때 영화를 고를 때 일반적으로 중요시 보는 속성들이란 줄거리, 장르, 감독, 배우, 인지도를 의미하며, 이 속성들에 대한 사용자 평가가 결합된 특징 벡터가 바로 개인화된 메타 가중치이다.

위 논문에서는 사용자가 영화를 고를 때 일반적으로 보는 속성을 줄거리, 장르, 감독, 배우, 인지도로 정의하였으며 이에 대한 사용자의 선호도를 조사하여 개인화된 메타 가중치로 사용하였다. 이 메타 가중치는 사용자가 위 5가지 특징에 대해 점수화한 것으로, OTT 서비스를 벗어난 사용자의 일상 속의 데이터를 비교했을 때 해당 데이터가 사용자 개인화를 위한 데이터라고 보기에 미흡하다.^[13] 사용자 조사를 통해 강화된 사용자 파악을 수행하였으나 내부 데이터만을 활용하였다는 한계가 있다.

III. 확장된 개인 데이터를 활용한 OTT 서비스 추천 시스템

실험을 위해 선별한 OTT 서비스와 사용자의 개인 데

이터는 각각 넷플릭스와 카카오톡 대화 파일이다. 넷플릭스는 22년도 와이즈앱에서 발표한 사용자가 가장 많은 OTT 앱에서 압도적인 1위를 차지하여 가장 우수한 OTT 서비스라 가정하고 선정하였다.^[14] 카카오톡 대화 파일은 22년도 와이즈앱에서 발표한 한국인이 가장 많이 사용하는 앱 1위와 한국인이 가장 자주 사용하는 앱 1위를 차지하였으며, 이는 채팅 서비스 특성상 개인 데이터가 풍부할 것이라 가정하고 선정하였다.^[15]

이용자 개개인을 위한 맞춤형 데이터 추출 시스템의 전체 흐름은 그림 1과 같다. 본 시스템은 카카오톡 대화 내용 텍스트 파일과 분석할 대상의 카카오톡 이름을 입력받아 해당 사용자의 취향을 담은 키워드를 반환한다. 해당 시스템의 큰 흐름은 1) 카카오톡 대화 파일 분석, 2) 넷플릭스 콘텐츠 데이터 크롤링, 3) 사용자 만족도 조사와 같다.

1. 카카오톡 대화 파일 분석

카카오톡 대화 파일과 카카오톡에 저장된 분석할 대상의 이름을 입력받으면 해당 사용자가 가장 많이 사용한 명사를 추출한다. 먼저 데이터 전처리 과정에서 카카오톡 채팅방에서 공통으로 나타나는 의미 없는 문자와 단어들을 제거한다. 여기에는 ‘ㄴ’나 ‘ㄷ’, ‘이모티콘’, ‘사진’, ‘동영상’ 등이 포함된다. 이후 데이터 정제 과정에서 날짜와 시간 정보를 삭제하고, 특정 사용자의 이름이 들

어난 문자열만 파이썬 리스트에 저장한다. 이렇게 정제된 문자열은 파이썬의 한국어 정보처리를 위한 패키지인 Konlpy(Korean Natural Language Processing in Python)의 Okt(Open Korean Text) 모듈을 사용하여 형태소를 분석하고 명사만 추출한다.^[16] 이때, 추출된 명사 중 명사가 아닌 불용어(예를 들어, 다른 사용자들의 카카오톡 이름이나 ‘지금’, ‘혹시’와 같은 명사가 아닌 단어)들을 미리 제작한 불용어 사전을 통해 제거해준다. 이렇게 정제된 단어는 디셔너리의 key 값에 저장되고 value 값에는 해당 단어의 빈도수를 카운팅하여 저장한 뒤, 빈도수가 높은 순으로 정렬한다. 이 디셔너리에 저장된 값은 사용자가 평소 자주 사용하는 단어로서 사용자의 개인적인 정보를 반영한 데이터라고 할 수 있다. 위 단어와 빈도수는wordDB에 저장된다.

2. 넷플릭스 콘텐츠 데이터 크롤링

넷플릭스와의 비교를 위해 취향 저격 콘텐츠 개수와 동일하게 데이터를 크롤링한다. 크롤링은 사용자가 자주 쓰는 단어 순으로 정렬된 wordDB의 단어를 순서대로 넷플릭스에 검색하는 방법으로 진행한다. 이때 다양한 단어를 검색하기 위해 한 단어당 크롤링 개수는 5개로 제한한다. 예를 들어 취향 저격 콘텐츠 개수가 40개인 사람은 최빈다 단어 상위 8개 * 단어당 크롤링할 콘텐츠 개수는 5개이며, 37개인 경우 최빈다 단어 상위 8개 * 단

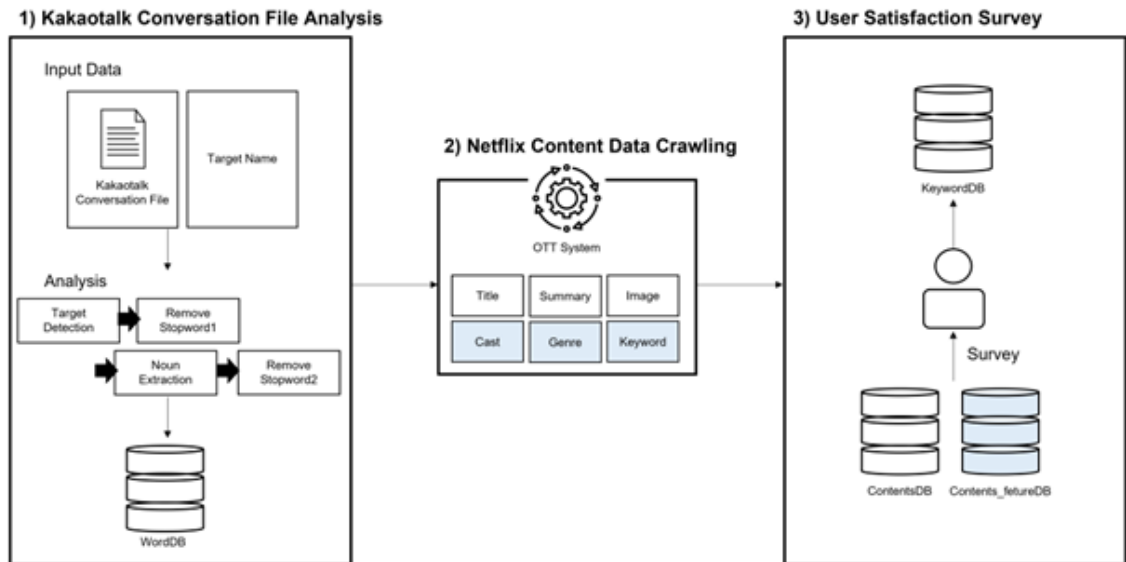


그림 1. 카카오톡 대화 파일을 활용한 넷플릭스 콘텐츠 추천 시스템
 Fig. 1. Netflix content recommendation system using Kakaotalk conversation file

어당 크롤링할 콘텐츠 개수는 5개이고, 맨 마지막으로 크롤링 된 콘텐츠 3개를 제거하는 식이다.

크롤링할 데이터는 콘텐츠의 제목과 포스터 이미지, 줄거리, 출연진, 장르, 시리즈 특징(해당 콘텐츠의 키워드 혹은 태그)이다. 만약 시리즈 특징이 없는 콘텐츠의 경우 검색한 키워드를 저장한다. 이 데이터 중 사용자 설문문에 사용될 콘텐츠 제목과 포스터 이미지, 줄거리는 contentDB에 저장하고, 출연진과 장르, 시리즈 특징은 그 특성상 여러 값이 나오므로 각각 테이블을 만들어 저장한다. 그리고 이 3개의 테이블의 기본키를 content_featureDB에 저장하며, 이 content_featureDB의 기본키는 contentDB에 저장한다. 여기서 기본키란 어떤 한 데이터베이스를 식별할 수 있는 값을 말한다.

크롤링은 파이썬 크롤링 모듈인 셀레니움(Selenium)을 사용한다. 셀레니움은 크롬, 인터넷 익스플로어, 파이어폭스와 같은 브라우저를 컨트롤하는 파이썬 모듈로 동적인 웹 페이지를 원격 조작하여 필요한 정보를 크롤링해오는 모듈이다.^[17]

3. 사용자 만족도 조사

크롤링이 끝나면 contentDB를 사용자에게 제공하여 선호도 여부를 실시한다. 이때 사용자가 호감으로 체크한 콘텐츠의 contentDB 기본키는 keywordDB에 저장한다. 이 keywordDB는 사용자의 개인 정보를 분석하여 추출한 키워드와 사용자가 호감으로 체크한 콘텐츠의 전반적인 정보를 가지고 있으며, OTT System에 제공되면 사용자 맞춤형 추천을 더 잘해줄 것으로 기대되는 데이터이다.

IV. 비교 및 평가

설문 대상자는 실험을 위해 평소 본인의 카카오톡 대화 파일과 넷플릭스 아이디, 패스워드를 제출해야 한다. 이는 개인 정보를 노출하는 것이므로 대상자를 모집하는 것이 어려웠다. 그 결과 전체 설문 대상자는 20대 대학생 4명으로, 넷플릭스를 1개월 미만 사용한 사람 1명, 3개월~6개월 사용한 사람 1명, 12개월 이상 사용한 사람 2명으로 구성되어 있다. 또, 설문지는 넷플릭스와 제안된 시스템이 각각 추천한 콘텐츠 중 사용자가 만족하는 콘텐츠의 비율을 알아보기 위해 아래와 같은 내용으로 구성하였다.

표 1. 설문지 목록
Table 1. Questionnaire list

Number	Content
1	Please select your favorite content among the content provided by Netflix.
2	Please select your favorite content among the content provided by the proposed method.

제안 데이터를 이용한 넷플릭스 콘텐츠 만족도 설문조사 진행 결과 사용자의 만족도는 평균 64.65%로, 넷플릭스 자체 추천 시스템과 비교했을 때 23.72% 더 높은 만족도를 보였다.

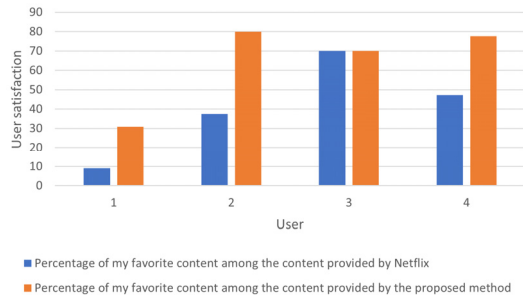


그림 2. 넷플릭스 추천과 제안 방식 추천에 대한 사용자 만족도 조사 결과

Fig. 2. Results of user satisfaction with Netflix recommendations and with the proposed method

V. 결론 및 향후 과제

기존의 OTT 추천 시스템은 미디어 서비스 플랫폼 안이라는 제한된 환경에서 수집한 데이터를 토대로 콘텐츠를 추천했다. 이는 앱 내부에서의 데이터를 활용한 것으로 앱 외부의 사용자 데이터와 비교했을 때 사용자 맞춤형 추천이라고 보기에는 미흡한 면이 있다. 또, 신규 가입자의 경우 서비스 이용에 앞서 좋아하는 장르 혹은 키워드를 고르게 하는 식으로 사용자의 취향을 파악하기 때문에 맞춤형 콘텐츠 추천이 어렵다. 이 결과, OTT 서비스 중 넷플릭스에 대한 사용자 추천 콘텐츠 만족도 조사 결과는 40.93%라는 결과가 도출되었다. 반면, 사용자의 개인데이터 중 카카오톡 대화파일을 이용하여 추천한 콘텐츠는 사용자의 '일상'이라는 폭넓은 범위에서 수집한 데이터로서 사용자의 개인적인 취향을 반영했을 뿐만 아니

라 사용자가 무의식적으로 사용하는 말 안에서 사용자도 몰랐던 관심사를 반영할 수 있다. 또한, 카카오톡 대화 파일을 토대로 추천된 콘텐츠의 만족도 조사 결과는 64.65%로 넷플릭스가 추천한 콘텐츠와 비교했을 때 사용자의 만족도가 높은 것을 알 수 있다. 또, 이렇게 추천한 콘텐츠 중 사용자가 관심있다 표현한 콘텐츠의 출연진, 장르, 태그 정보는 또 다른 사용자의 취향 데이터 정보가 되어 사용자에게 맞는 콘텐츠를 추천해줄 수 있는 기반이 된다. 이처럼 본 논문은 사용자 개인 데이터를 활용한 맞춤형 추천 콘텐츠와 해당 추천 콘텐츠를 통한 사용자의 선호도 조사를 통해 추가적인 사용자 맞춤형 데이터를 얻을 수 있다는 의의가 있다.

References

- [1] WIKIPEDIA, Feb 2023, "OTT Service", https://ko.wikipedia.org/wiki/OTT_%EC%84%9C%EB%B9%84%EC%8A%A4
- [2] Suk-Hee Han, "Analysis of Structure and Content of Entertainment OTT: Focus on WWE Network," The Journal of The Institute of Internet, Broadcasting and Communication, Vol. 19, No. 3, pp. 41-51, Jun. 2019. DOI: <https://doi.org/10.7236/JIIBC.2019.19.3.41>
- [3] KISDI, May 2022, "A Comparative Analysis of Free & Paid Users of OTT services", https://mediasvr.egentouch.com/egentouch.media/apiFile.do?action=view&SCHOOL_ID=1007002&URL_KEY=1e4f9199-b101-4863-a7ca-243fe4e48a1c
- [4] The Export-Import Bank of Korea, May 2021, "OTT Industry and K-Content Export", <https://keri.koreaexim.go.kr/HPHF0E052M01/63856?curPage=2>
- [5] DigitalTVResearch, May 2022, "Global OTT TV and Video Prediction", <https://digitaltvresearch.com/product/global-ott-tv-and-video-forecasts/>
- [6] Ji-Hun Yoo and Joo-Yeun Park, "A Study on the Factors Influencing Continuous Usage Intension based on OTT Service User", Journal of Broadcasting and Telecommunications Research, Vol., No. 102, pp. 46-79, Jan. 2018.
- [7] Cho-Do Eun, Si-Jung Kim and Yoon-Silc Kwak, "A Study of Personalized Contents Recommendation Method Based on User Preference Learning", Journal of Korean Institute of Information Technology, Vol. 9, No. 9, pp. 229-235, Mar 2011.
- [8] Debarun Chakraborty, Mujahid Siddiqui, Aaliyah Siddiqui, Justin Paul, Ganesh Dash and Francesca Dal Mas, "Watching is valuable: Consumer views - Content consumption on OTT platforms", Journal of Retailing and Consumer Services, Vol. 70, No. 103148, Jan. 2023. DOI: <https://doi.org/10.1016/j.jretconser.2022.103148>
- [9] Samala Nagaraj, Soumya Singh and Venkat Reddy Yasa, "Factors affecting consumers' willingness to subscribe to over-the-top (OTT) video streaming services in India", Technology in Society, Vol. 65, No. 101534, May. 2021. DOI: <https://doi.org/10.1016/j.techsoc.2021.101534>
- [10] Seck-Jun Jang, "A Study on the Influence of Service and Platform Attribute, Satisfaction and Continuance Usage Intention in OTT: Focus on MZ Generation", Journal of Speech, Media and Communication Research, Vol. 21, No. 3, pp. 71-114, Aug. 2022.
- [11] Korea Press Foundation, Nov 2019, "YouTube Recommendation Algorithm and Journalism", https://www.kpf.or.kr/front/research/selfDetail.do?seq=575347&link_g_homepage=F
- [12] Seo-In Baek and Dai-Ki Min, "Contents Preference Model Combined with Matrix Factorization for Movie Recommendation", Journal of the Korean Institute of Industrial Engineers, Vol. 47, No. 3, pp. 280-288, Dec. 2021. DOI: <https://doi.org/10.7232/JKIIIE.2021.47.3.280>
- [13] Hyeon-Woo An, Hea-Woon You and Dae-Yeol Kim, "Design and Implementation of Contents-based Customized movie recommendation system using meta weight learning", The Korean Institute Of Broadcast And Media Engineers, Vol. , No. , pp. 465-468, Jul. 2020.
- [14] WISEAPP, May 2022, "The OTT apps with the largest number of users are Netflix, Tving, Coupang Play, and Wave", <https://www.wiseapp.co.kr/insight/detail/190/>
- [15] WISEAPP, May 2022, "This is the app that Koreans use the most, for the longest period, and most frequently", <https://www.wiseapp.co.kr/insight/detail/188/>
- [16] Konlpy, URL: <https://konlpy.org/>
- [17] Selenium, URL: <https://www.selenium.dev/>

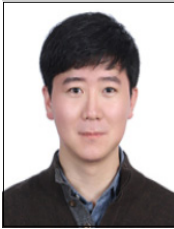
저 자 소 개

유 희 정(학생회원)



- 2019년 3월 ~ 현재 : 군산대학교 소프트웨어학부 학사과정
- 주관심분야 : 자연어처리, 소프트웨어공학

김 능 회(정회원)



- 2008년 8월 : 고려대학교 컴퓨터학과 (이학석사)
- 2017년 8월 : 고려대학교 컴퓨터 · 전파통신공학과 (공학박사)
- 2021년 3월 ~ 현재 : 군산대학교 소프트웨어학부 조교수
- 주관심분야 : 요구공학, 소프트웨어공학