

텍스트마이닝 기반 아동 감정 분석 시스템 및 아동용 감정 사전 구축 방안

박영준* · 김선용** · 김요한***

A Child Emotion Analysis System using Text Mining and Method for Constructing a Children's Emotion Dictionary

Young-Jun Park* · Sun-Young Kim** · Yo-Han Kim***

요약

급격하게 변화되는 사회 속에서 현대인들은 다양한 스트레스를 경험하고 있으며, 아동 또한 정신 건강 진료량이 눈에 띄게 증가하고 있다. 소아정신건강장애 등 아동의 정신 건강 문제를 예방하기 위해서는 감정 상태를 빠르게 파악해야 하지만, 유아기 아동들은 몇 가지 단어만을 사용하여 자신의 감정을 표현하는 경우가 많기에 어려움이 있다. 본 논문에서는 아동 심리 상태를 우울, 불안, 외로움, 두려움 등 4가지의 감정으로 세분화하고 아동 심리 전문가의 점수를 기반으로 한 아동용 감정 사전 구축 방안을 제안한다. 또한, STT 및 텍스트 마이닝 기반의 아동 감정 분석 시스템을 제안하고 실제 음성 데이터로 성능을 평가하였다. 평가 결과는 제안한 아동용 감정 분석 시스템이 아동 감정 상태를 정확하게 파악할 수 있음을 보여준다.

ABSTRACT

In a society undergoing rapid change, modern individuals are facing various stresses, and there's a noticeable increase in mental health treatments for children as well. For the psychological well-being of children, it's crucial to swiftly discern their emotional states. However, this proves challenging as young children often articulate their emotions using limited vocabulary. This paper aims to categorize children's psychological states into four emotions: depression, anxiety, loneliness, and aggression. We propose a method for constructing an emotion dictionary tailored for children based on assessments from child psychology experts.

키워드

Child Sentiment Dictionary, Sentiment Analysis, Text-Mining, Speech to Text
아동 감정 사전, 감정 분석, 텍스트 마이닝, STT

1. 서론

최근 코로나19 팬데믹과 급격하게 변화되는 사회 속에서 현대인들은 다양한 스트레스를 경험하고 있다. 스트레스는 모든 연령대에서 발생할 수 있으며, 아동

또한 중년기 성인 못지않은 다양한 스트레스를 경험하고 있다[1]. 2019~2021년 동안 9세 이하 아동의 우울증 진료는 32.4%, 스트레스 진료는 58%로 대폭 증가하는 등 소아청소년의 정신 건강 진료량이 눈에 띄게 증가하고 있다[2]. 이러한 스트레스는 아동에게 정

* 동서대학교 소프트웨어학과 (ppp7545@naver.com)

** 동서대학교 소프트웨어학과 (sykim@dongseo.ac.kr)

*** 교신저자 : 동서대학교 소프트웨어학과

• 접수일 : 2024. 04. 29

• 수정완료일 : 2024. 05. 21

• 게재확정일 : 2024. 06. 12

• Received : Apr. 29, 2024, Revised : May. 21, 2024, Accepted : Jun. 12, 2024

• Corresponding Author : Yo-Han Kim

Dept. of Software, Dongseo University

Email : yhkim@dongseo.ac.kr

서적 부적응, 신체적 건강상태, 다양한 형태의 행동 문제를 발생시킬 수 있어 아동의 발달에 치명적이다.

국민의 생활 수준이 높아짐에 따라 유아의 심리적 건강 발달과 가족 상호작용에 대한 관심이 높아지고 있다[3]. 그러나, 나이가 어릴수록 자신이 느끼는 감정을 문장으로 정확하게 전달하기 어려우며, 심리 상태나 갈등을 그림으로 그리거나 행동으로 표출하기도 한다[4]. 말을 하기 시작하는 유아기 아동들도 단순히 몇 가지 단어를 연결해 자신의 감정을 표현하는 경우가 많다. 부모가 아동의 심리 상태를 파악하기 힘든 경우에는 아동 심리 전문가의 도움을 받아야 하지만, 항상 도움을 받기에는 어려움이 있어 전문가의 도움 없이도 아동의 감정 상태를 파악할 방안이 요구된다.

최근 AI와 데이터 분석 기술의 발전으로 다양한 분야에서 텍스트 마이닝 기법이 활용되고 있다[5-6]. 텍스트 마이닝 기법 중 하나인 감정 분석은 다양한 분야에서 활발하게 연구가 진행되고 있다. KNU 감성 사전은 단어의 감정을 긍정, 부정, 중립으로 세분화한 감정 단어 사전을 구축하였다[7]. 이러한 감성 사전의 경우 특정 도메인이나 연령대에 따라 감정의 정도가 달라진다는 특징이 있어 아동에게 적용하기에는 무리가 있다. 예를 들면, ‘열심히’라는 단어는 성인에게는 긍정일 수 있지만, 아동에게는 폭력적이고 두려운 감정을 느끼게 할 수도 있다. 따라서 아동의 감정을 정확하게 분석하기 위해서 아동에 적합한 데이터를 바탕으로 감정 사전을 구축할 필요가 있다.

기존의 감정 분석 연구들은 대부분 감정 단어를 긍정 또는 부정의 감정으로 양극화하는 극성 분류가 대부분이었으나, 최근에는 빅데이터 분석 및 인공지능 기술의 발전으로 다양한 감정으로 분류하기 위한 인공지능 기반 감정 분류 기술이 제안되었다[8-9]. 그러나, 문장의 단어 형태소 하나씩에 대한 세분화된 분석이 어렵고, 문장 전체를 읽어야만 감정 점수를 부여할 수 있다는 점에서 데이터수집에 어려움이 있다. 또한, 이러한 문제는 단어 단위로 대화를 이어가는 아동을 대상으로 하는 경우 감정 분류 정확도가 떨어질 수 있으며, 단어 단위로 실시간 감정 분석을 실현하기 위한 알고리즘의 확장성이 낮아질 수 있다.

본 논문에서는 소아정신건강장애를 겪는 아동의 대화로부터 정확한 심리 검사를 수행하기 위해 아동 심리 전문가 자문을 바탕으로 아동의 감정 상태를 우울

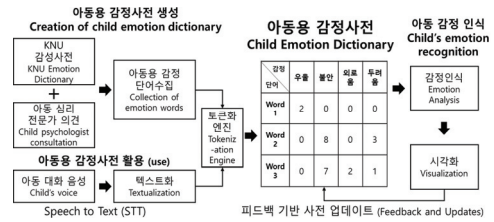


그림 1. 아동 감정 분석 시스템 모델
Fig. 1 Child Emotion Analysis System Model

(과거 및 현재에 대한 상실감), 불안(미래의 상실감에 대한 예측), 외로움(주변과 분리되어 고립된 감정), 두려움(위협이나 위험을 느끼는 감정) 등 4가지의 감정으로 세분화하였다. 아동의 감정 상태를 분석하기 위해 4가지 감정에 대한 전문가 점수 기반 아동용 감정 사전 구축 방안을 제안하고, STT(Speech-to-Text) 및 텍스트 마이닝 기반의 아동 감정 분석 기법을 통해 아동의 감정 상태를 파악한다. 아동의 감정 분석 결과는 본 연구에서 구축한 제안한 아동용 감정 사전 및 텍스트 마이닝 기반의 감정 분석이 아동의 감정을 쉽게 파악하고 빠른 대처가 가능하게 함을 보여준다.

II. 아동 감정 사전 구축 및 아동 감정 분석

2.1 아동용 감정 분석 시스템

그림 1은 본 논문에서 제안하는 아동 감정 분석 시스템을 나타낸다. 아동 심리 전문가의 의견을 반영해 KNU 감성 사전을 기반으로 아동용 감정 사전을 구축하고, STT를 통해 텍스트화된 아이들의 대화 내용으로부터 감정 사전을 활용해 감정 분석을 수행하였다. 최종적으로 분석된 아동의 감정 상태는 확률 그래프 형태로 시각화된다.

2.2 STT 기반 아동 음성 인식 모델

분석 아동들의 실제 대화 음성을 텍스트로 변환하기 위한 STT 기술은 사람의 음성을 컴퓨터로 해석하여 문자형 데이터로 전환하는 기술이다. 다만, 발화자의 인종, 나이, 스타일에 따라 인식률이 낮아질 수 있다. 아동 발화자의 STT 음성 인식을 위해 AI-Hub에서 제공하는 ‘한국어 아동 음성 데이터’를 활용하였으며, 높은 인식 성능을 보이는 Wav2Vec 기반 모델을

Google Speech Recognition thinks you said : 까를까를 내 볼
 Google Speech Recognition thinks you said : 어찌 간직간질 우리 아기 발
 Google Speech Recognition thinks you said : 디리 쓰려면 환참 기다려야 됩디다
 Google Speech Recognition thinks you said : 때찌 해 줄까 때찌
 Google Speech Recognition thinks you said : 생김만 해도 신나지 않니
 Google Speech Recognition thinks you said : 영소가 풍글풍글 들을 토목토목 놀디니
 Google Speech Recognition thinks you said : 두 손을 활짝 펴서
 Google Speech Recognition thinks you said : 열미가 손을 벌여서 하루에 겨우 한 끼를 먹어요
 Google Speech Recognition thinks you said : 숨이는 까민의 일함을 취 벗어던지고는
 Google Speech Recognition thinks you said : 친구들은 숨이의 일함을 보고 예뻐하고 했어요

그림 2. STT 기반 아동 음성 텍스트화
 Fig. 2 Child's Voice STT Conversion

		우 울				식별번호 :	
0							
1	거울	글	내려오다	도하지	명	별나라	약능
2	살	야구	의사	지하도	행복	수영장	행권
3	가슴	그날	나무	담장	말다	발목	올리다
4	선물	여기	알어서다	조동약쟁	책시	열매	통화
5	(주사름)뺏아	귀신	곧나다	달라	마음	밖	오전
6	과일	꼬리	다르다	똑같다	만들래	빚방울	운전
7	개	그러다	남자	도보	멋지다	버섯	재미
8	색칠	어떤	이십	착하다	화나다	시내	세수
9	빠다	아중마	원피스	주장	학교	세탁소	자르다
10	사과	안	위험하다	중국	완식	소리	잔
		불 안				식별번호 :	
1	생각나다	어깨	인터넷	쪽	말다	수	우유
2	생	것말	유모차	준다	호수	소파	운동복
3	사다리	그네	이름	중국	완색	산방	하루
4	산	기쁨	날짜	지우하다	모르다	빨간색	작년
5	시버스	아랫밤	늪다	지나가다	미끄러웠	비다	왜
6	삼겹	얼다	곧다	달력	행복	식탁	오도바이
7	사위	꼬리	나눠가져	집게	학교	사탕물	온온
8	고수	꼭	이름	당면	바나나	빚방울	움동장
9	생일	어머니	유지원	중국집	마르다	비누방울	자리
10	석유	깨끗하다	이분	중간	휴가	시계	말십
		외 로움				식별번호 :	
1	고양	귀찮다	음료수	도시락	목걸이	벗어	요일
2	과자	깜깜해	남	도토리	회	시원하다	자동판매기
3	선배	그만	내일	첫째	행동	블라우스	통화
4	개구리	곧씨	날말	뜨겁다	햇빛	식방	장미
5	결혼	어름	다람쥐	초록색	형	숨	옷장
6	곤	규칙	내기	닫고기	마음	식물	모장
7	거울	악어	유령	단종일	맛있다	시상	왜냐하면
8	같다	여섯	의자	준비물	말	불가사리	재미
9	색	그때	노안	등장	미안하다	밤	아까
10	거짓말	급하다	이렇다	질문	목소리	부끄럽다	우표
		두 려움				식별번호 :	
1	공놀이	그리고	나들이	찬물	학생증	시	자꾸
2	갯벌	그네	일요일	떨어뜨려	화장품	배방이	집시
3	개	에이스크림	다쳐	드디어	호	방	성함
4	곤	극장	이모	담	웃생기다	배추	피해
5	선물	꽃다발	내려오다	똑같이	한가위	시작	표지판
6	거울	기회	다리	틀어가지	행동	블라우스	통화
7	가지	어서	노래	대회	메시지	수	요리사
8	생	깨나다	이름	중심	환자	발표	출발
9	사다	연재	나비	등	명	병원	포도
10	서점	그러면	이웃	라면	한식	손가락	우체국

그림 3. 아동용 4가지 감정 단어 시트지
 Fig. 3 4-Emotion Word Sheet for Children

활용해 STT 음성 인식을 수행하였다. 그림 2는 STT 모델을 기반으로 아동 음성파일 데이터를 텍스트 데이터로 변환한 결과이다. 아동의 음성 데이터가 잘 인식되어 출력되는 것을 확인할 수 있다.

귀중하게	귀중	2
귀중하다	귀중	2
귀중한	귀중	2
귀중한 보	귀중 보배	2
귀중한 자	귀중 자원	2
귀중히	귀중히	2
귀중히 여	귀중히 여	2
귀중히 여	귀중히 여	2
귀차니즘	귀	-1
귀찮거나	귀찮	-1
귀찮게	귀찮	-1

그림 4. KNU 감성사전 예시
 Fig. 4 Example of KNU Emotion Dictionary

2.3 아동용 감정 사전 구축

본 논문에서는 KNU 감성 사전을 기반으로 4가지 감정 상태(우울, 불안, 외로움, 두려움)에 대한 아동용 감정 사전 구축 방안을 제안한다.

STT를 통해 변환된 아동 음성 데이터는 감정 상태를 나타내는 여러 단어로 이루어진다. 따라서, 그림 3과 같이 아동 심리 전문가를 통해 4가지 감정 상태와 연관된 단어가 세분화된 감정 시트지를 도출하였다. 사용된 단어는 아이들이 특정 감정에 대한 단어를 직접 체크하도록 하여 수집되었으며, 아동 심리 전문가가 아이들의 감정 상태에 따라 단어별로 감정 점수를 1~5점 사이로 부여하도록 하였다. 하나의 단어는 여러 가지 감정에 대한 점수를 동시에 가질 수 있도록 설계하였다. 예를 들면, '위험'이라는 단어는 '우울' 감정에 4점, '불안' 감정에 3점, '두려움' 감정에 5점이 부여된다.

아이들을 통해 직접 수집되어 4가지 감정으로 세분화된 단어는 실질적인 감정의 의미를 나타내는 실질 형태소만으로 구성되어 있어 다양한 형태의 감정 단어에 적용할 수 있도록 형태를 다양화하여 감정 사전을 구축해야 한다. 그림 4는 KNU 감성 사전에서 실질형태소인 '귀중'과 관련된 단어를 출력한 것이며, 각 단어에 대한 점수는 감정을 긍정/부정으로 양분화하여 5단계(-2, -1, 0, 1, 2)로 점수화한 것이다. 일반적으로 명사는 단어의 형태가 '친구' 또는 '노래'와 같은 가지지만, 감정을 기반으로 한 동사, 형용사, 부사 등은 그림 4에서와 같이 '귀중하다'는 뜻을 가진 단어가 '귀중한' 또는 '귀중히 여겨'와 같이 다른 단어로 표현되는 등 한 가지 뜻과 감정 상태를 드러낼 수 있는 단어임에도 다양한 단어로 표현될 수 있다[10].

word	word_root	polarity	depression	anxiety	loneliness	violence
목숨	목숨	0	5	0	0	0
하늘	하늘	0	5	0	0	0
끝나다	끝나	0	5	0	0	0
마음	마음	0	5	0	0	0
기본	기본	0	5	0	4	0
기억	기억	0	5	0	4	0
내려오다	내려오	0	4	0	0	0
메모	메모	0	4	0	0	0

그림 5. 아동용 감정 사전 및 단어별 감정 점수
Fig. 5 Emotion Dictionary for Children with Emotion Scores per Word

본 논문에서는 그림 3에서 도출된 감정 단어를 다 양화하여 아동 감정 분석의 정확도를 높이기 위하여 KNU 감정 사전의 형태소 분석을 통해 그림 5와 같이 실질형태소 매칭을 수행하였다. 감정 단어(word)를 Kkma 형태소 분석기를 통해 형태소 단위로 분석하여 실질 형태소(word_root)로 저장하고, 아동 감정 단어 시트지와 KNU 감정 사전에서 같은 의미를 갖는 단어들이 동일하게 간주되도록 설계하였다. 또한 기존 KNU 감정 사전에서 단순히 긍정/부정으로 양분화된 감정 분석을 우울, 불안, 외로움, 두려움의 4가지로 세분화하여 아동의 감정을 구체적으로 판단할 수 있도록 감정 사전을 구축하였다. 최종적으로 구축된 아동용 감정 사전은 아동의 대화로부터 형태소 분석을 통해 감정 단어를 토큰화하고, 전체 대화에서 감정 점수를 도출하여 아동이 4가지 감정 중에서 어떤 감정 상태인지를 결과로 보여준다.

2.4 아동용 감정 사전 기반 아동 감정 분석

본 논문에서 구축한 감정 사전을 통해 아동 감정 분석을 진행하기 위해서는 입력된 아동의 대화를 STT를 통해 텍스트 형태로 추출하고 형태소 분석을 통해 단어 토큰화를 수행해야 한다. 그림 6은 아동 대화의 일부를 Kkma 기반 형태소 분석을 통해 토큰화한 결과를 보여준다. 숫자는 문장 내에서 등장한 형태소 단어에 대한 빈도수를 나타낸다. 예를 들어, 전체 문장에서 '하늘' 단어는 4번, '끝나' 단어는 1번 사용된 것을 확인할 수 있다.

형태소 분석이 끝난 아동 대화 데이터는 그림 7과 같이 본 논문에서 구축한 아동 감정 사전과의 매칭을 통해 감정 점수가 부여된다. 마지막으로 그림 8에서와 같이 4가지 감정 상태에 대한 점수 결과를 막대 그래

```

(하늘: 4, '별나': 1, '마음': 14, '기본': 10, '기억': 1, '흔재': 6, '위험': 2, '꿈': 5, '살': 9,
침울: 5, '그날': 1, '있': 1, '토': 4, '개': 10, '그리': 11, '이상': 11, '질문': 2, '공부':
4, '물음': 1, '불편': 1, '못': 12, '살': 5, '어울': 4, '재미없': 1, '지루': 3, '질': 2, '그프': 5,
'원하': 6, '미안': 2, '쉬': 1, '검은색': 6, '떨어지': 7, '울': 5, '슬프': 5, '갈': 1, '빠
지': 5, '내려가': 1, '죄송': 1, '무감': 1, '눈물': 4, '발': 5, '갈': 5, '부담': 2, '바위': 2, '화나':
1, '그려나': 2, '술': 1, '그릴지만': 2, '올리': 3, '아나': 17, '더럽': 2, '모르': 3, '쓰
레기': 3, '생각': 1, '생각': 20, '조심': 4, '동물': 4, '이름': 1, '짙깔': 1, '있게': 1, '오르': 2,
무섭': 3, '피해': 1, '기반': 3, '자주': 1, '죽': 3, '떨어뜨리': 1, '싸움': 1, '갈아지':
1, '인형': 5, '결혼': 2, '사랑': 2, '반감': 2, '사랑': 11, '우린': 68, '안속': 2, '돌아오': 3, '중아': 1,
    
```

그림 6. Kkma 기반 형태소 분석 결과 예시
Fig. 6 Example of Kkma-based Morphological Analysis

word	word_root	polarity	depression	anxiety	loneliness	violence
0	하늘	하늘	0	5	0	0
1	하늘	하늘	0	5	0	0
2	하늘	하늘	0	5	0	0
3	하늘	하늘	0	5	0	0
4	끝나다	끝나	0	5	0	0
...
673	어름	어름	0	0	0	2
674	어름	어름	0	0	0	2
675	어름	어름	0	0	0	2

그림 7. 문장 내 단어에 대한 감정 점수 부여
Fig. 7 Assigning Emotion Scores to Words

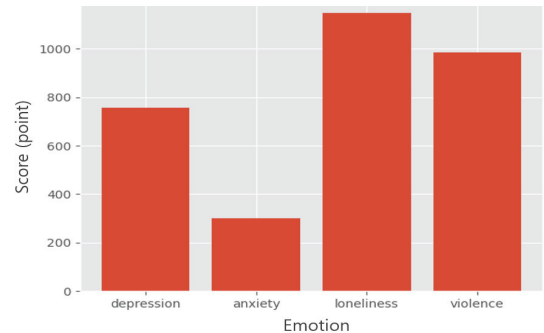


그림 8. 감정 점수 시각화
Fig. 8 Emotion Score Visualization

프로 시각화하여, 제안하는 아동용 감정 분석 시스템을 활용하는 사용자들이 아동의 감정 상태를 한눈에 파악할 수 있도록 구현하였다. 그림 8의 score는 문장 전체에 등장하는 단어들의 감정 점수를 더한 결과를 보여준다. 결과는 아동이 '기분', '그리', '아빠', '울', '생각' 등의 단어를 많이 사용하였고, 이에 따라 외로움 감정에 가장 높은 점수를 얻은 것을 보여준다.

III. 성능 평가

3.1 감정에 대한 점수 성능 평가 결과

그림 9와 그림 10은 본 논문에서 제안한 아동용 감정 분석 시스템을 활용하여 우울, 불안, 외로움, 두려움 감정에 대한 소아정신건강장애를 겪고 있는 아동

석에 활용 가능하도록 연구를 진행할 예정이다. 또한, STT 기반 인공지능 스피커를 활용하여 아동의 발화 즉시 감정 분석을 할 수 있도록 실시간 텍스트 분석 기술에 대한 연구를 진행하고자 한다.

감사의 글

"본 연구는 2024년 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학사업의 연구결과로 수행되었음"(2019-0-01817)

References

[1] J. Park and M. Chung, "Relationships Between Strategies for Coping with Stress and Behavior Problems in School - age Children," *Korean Journal of Child Studies*, vol. 22, no. 3, pp. 107-122, 2001.

[2] Salary Information Management Department, "Analysis of Depression and Anxiety Disorder Treatment Status in the Recent 5 Years (2017-2021)," *Health Insurance Review & Assessment Service* - press release, 2022.

[3] J. Song, J. Lee, and L. Kim, "The Effect of Mother's Self-reflection on Children's Resilience: The Mediating Effect of Co-parenting," *Korean Journal of Child Care and Education Policy (KJCCPEP)*, vol. 17, no. 3, pp. 3-21, 2023.

[4] H. Oh, "A Study on Children's Emotion and Psychology about Color By Age Group," *Master's Thesis of Cheju National University*, 2005.

[5] J. Choi and S. Kim, "Early Detection Assistance System for Rare Diseases based on Patient's Symptom Information," *Journal of the Korea Institute of Electronic Communication Sciences*, vol. 18, no. 2, pp 373-378, 2023.

[6] J. Lee, M. Moon, Y. Jeong, and S. Park, "Development of Dog Name Recommendation System for the Image Abstraction", *Journal of the Korea Institute of Electronic Communication Sciences*, vol. 18, no. 2, pp 313-320, 2023.

[7] S. Park, C. Na, M. Choi, D. Lee, and B. On, "KNU Korean Sentiment Lexicon - Bi-LSTM-based Method for Building a Korean Sentiment Lexicon," *Journal of Intelligence and Information Systems*, vol. 24, no. 4, pp. 219-240, 2018.

[8] M. Nam, E. Lee, and J. Shin, "A Method for User Sentiment Classification using Instagram Hashtag," *Journal of Korea Multimedia Society*, vol. 18, no. 11, 2015.

[9] Y. Kim and M. Song, "A Study on Analyzing Sentiments on Movie Reviews by Multi-Level Sentiment Classifier," *Journal of Information System*, vol. 22, no. 3, pp. 71-89, 2016.

[10] S. Park, C. Eom, B. On, and D. Jeong, "An AR-KNU Sentiment Lexicon-based University Reputation Assessment Using Online News Data," *Journal of Korean Institute of Information Technology*, vol. 17, no. 3, pp. 11-21, 2019.

저자 소개



박영준 (Young-Jun Park)

2017년~ 동서대학교 소프트웨어학과 학부생
2021년~ 동서대학교 AI공학 학과 학부생
※ 관심분야 : 인공지능, 데이터마케팅



김선용(Sun-Yong Kim)

2014년 한동대학교 전산전자공학부 졸업(공학사)
2019년 University of Maryland, Baltimore County 방문연구원
2020년 GIST 전기전자컴퓨터공학부 졸업(공학박사)
2020년 GIST AI연구소 연구원
2020년~현재 동서대학교 소프트웨어학과 조교수
※ 관심분야 : 스마트그리드, 강화학습



김요한(Yo-Han Kim)

2015년 충북대학교 정보통신공학부 졸업(공학사)
2021년 광주과학기술원(GIST) 전기전자컴퓨터공학부 석박사통합과정 졸업(공학박사)
2022년 한국과학기술정보연구원 선임연구원
2022년~현재 동서대학교 소프트웨어학과 조교수
※ 관심분야 : 차세대 통신(B5G, 6G), 강화학습