

자동화기반의 가짜 뉴스 탐지를 위한 연구 분석

좌희정¹, 오동석², 임희석^{3*}

¹고려대학교 컴퓨터학과 박사과정, ²고려대학교 Human-inspired 복합지능 연구센터 연구원, ³고려대학교 컴퓨터학과 교수

Research Analysis in Automatic Fake News Detection

Hee-Jung Jwa¹, Dong-Suk Oh², Heui-Seok Lim^{3*}

¹Ph. D. Candidate, Dept. of Computer Science and Engineering, Korea University

²Researcher, Human-inspired AI & Computing Research Center, Korea University

³Professor, Dept. of Computer Science and Engineering, Korea University

요 약 가짜 정보를 탐지하기 위한 연구는 2016년 미국 대통령 선거 이후 본격적으로 시작되었다. 정확한 출처를 알 수 없는 정보들이 뉴스 형식으로 생산되고, 이는 자극적이고 흥미로운 소재에 많은 관심을 보이는 대중의 특성에 따라 빠른 속도로 확산되고 있다. 또한, 소셜 네트워크 서비스 등 정보를 전달하기 쉬운 플랫폼의 대중화는 이러한 현상을 더욱 악화시킨다. Poynter는 IFCN(International Fact Checking Network)를 만들어 숙련된 전문가들이 사실 여부를 판단할 수 있는 가이드라인을 제시하고, 팩트 체크 기관을 위한 강령을 제공하고 있다. 하지만 이러한 접근 방법은 하나의 기사에 대한 진위 여부를 검증하기 위해 다수의 전문가 인력이 투입되어야 하므로 시간 및 금전적 비용이 크다. 따라서 지속적으로 증가하는 가짜 뉴스에 효율적으로 대응할 수 있는 자동화된 가짜 뉴스 탐지 기술에 대한 연구가 주목받고 있다. 본 논문에서는 최근 딥러닝 기술의 접목으로 인해 빠르게 발전하고 있는 가짜 뉴스 탐지 시스템과 연구들을 정리 및 분석한다. 또한, 많은 연구가 필요한 본 분야에 연구자들이 쉽게 접근할 수 있도록 다양한 형태로 주어지는 학습 말뭉치 및 챌린지들도 정리한다.

주제어 : 가짜 뉴스, 가짜정보, 가짜 뉴스 챌린지, 머신러닝, 딥러닝

Abstract Research in detecting fake information gained a lot of interest after the US presidential election in 2016. Information from unknown sources are produced in the shape of news, and its rapid spread is fueled by the interest of public drawn to stimulating and interesting issues. In addition, the wide use of mass communication platforms such as social network services makes this phenomenon worse. Poynter Institute created the International Fact Checking Network (IFCN) to provide guidelines for judging the facts of skilled professionals and releasing "Code of Ethics" for fact check agencies. However, this type of approach is costly because of the large number of experts required to test authenticity of each article. Therefore, research in automated fake news detection technology that can efficiently identify it is gaining more attention. In this paper, we investigate fake news detection systems and researches that are rapidly developing, mainly thanks to recent advances in deep learning technology. In addition, we also organize shared tasks and training corpus that are released in various forms, so that researchers can easily participate in this field, which deserves a lot of research effort.

Key Words : Fake news, Fake Information, Fake News Challenge, Matching Learning, Deep Learning

*This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program (IITP-2018-0-01405) supervised by the IITP(Institute for Information & communications Technology Planning & Evaluation)

*Corresponding Author : Heuiseok Lim(limhseok@korea.ac.kr)

Received May 20, 2019

Revised June 21, 2019

Accepted July 20, 2019

Published July 28, 2019

1. 서론

가짜 뉴스는 정치·경제 이익을 위해 의도적으로 언론 보도의 형식을 하고 유포된 거짓 정보이다. 가짜 뉴스의 범위가 정확하게 구분되어 있지 않아서 단어가 광범위하게 사용되고 있지만 가짜 뉴스를 판별하기 위해 뉴스를 가짜 뉴스와 진짜 뉴스가 아닌 것으로 구분할 수 있다는 점에서 뉴스를 다루는 언론 매체와 많은 관련이 있다고 볼 수 있다.

국제 언론자유 감시단체인 국경 없는 기자회견(RSF:Reporters Sans Frontieres)[1]가 2018년 발표한 언론 자유 지수(Press Freedom Index)를 보면 한국은 조사 대상국 180개 국가 중 43위로 5개 그룹 중 두 번째 그룹인 Fairly good(yellow)에 속했다. 언론자유 지수는 87개 질문 문항과 자체 수집하고 있는 각국의 언론사와 언론인에 대한 학대 및 폭력 관련 데이터를 종합해 도출한 것이다.

또 디지털 뉴스 리포트[2]는 2018 리포트에서 설문조사 결과 한국사람은 소셜 미디어를 포함해 온라인으로 뉴스를 얻는 비율이 84%이고, 이 중 70%가 스마트폰으로 뉴스를 접한다고 밝혔다. 가장 인기 있는 메시징 앱인 카카오톡은 사용자들 사이에 뉴스와 정치 정보가 공유되고 확산되는 채널 역할을 한다.(39%)

반면 뉴스에 대한 신뢰도는 Fig. 1과 같이 25%로 37개국 중 가장 낮다. 네이버와 다음과 같은 검색 엔진을 통한 뉴스 또한 23% 신뢰하고, 개인적으로 뉴스를 습득하기 위해서 주로 이용하는 매체의 뉴스에 대한 신뢰도는 29%, 소셜 미디어에 대한 신뢰(19%)는 훨씬 낮았다.

DIFFERENT TYPES OF TRUST

News overall	25%	37 th /37
News in search	23%	
News I use	29%	
News in social	19%	

Fig. 1. Reuters Institute Digital News Report 2018

조사 결과를 정리해 보면 한국의 언론은 자유가 어느 정도 보장되지만 언론에 대한 신뢰도는 낮은 나라로 분

류되었다. 한국사람은 새로운 모바일 매체에 빠르게 적응하고 있고, 대부분의 뉴스를 온라인을 통해 얻는다. 하지만 언론 매체의 뉴스라고 해서 신뢰하지 않고 있으며 가짜 뉴스가 확산되고 있다는 것을 충분히 인지하고 있다.

또한 인터넷에서 실제로 무엇이 가짜인지 구분하기 위해 미디어 회사(79%), 플랫폼(77%), 정부(73%)는 더 많은 일을 해야 한다고 응답한 결과를 보면 가짜 뉴스를 판별하는 시스템 구축의 필요성도 느낀다고 볼 수 있다.

지금까지 가짜 뉴스 탐지 시스템의 필요성에 대해 알아보았다. 이 논문에서는 가짜 뉴스 탐지 시스템에 대한 연구가 어떻게, 어떤 방향으로 이뤄지고 있는지 연구동향을 정리하고, 실제로 가짜 뉴스를 판별하기 위해 서비스되고 있는 사례들을 소개한다. 마지막으로 결론에서는 한계점과 더 나은 가짜 뉴스 탐지를 위한 제안을 한다.

2. 가짜 뉴스의 종류

Fig. 2와 같이 가짜 정보는 동영상, 사운드, 이미지, 텍스트의 형태로 나타날 수 있고, 가짜 텍스트는 가짜 뉴스, 소셜네트워크서비스, 연설문, 문서 등의 형태로 데이터가 사용되고 있다.

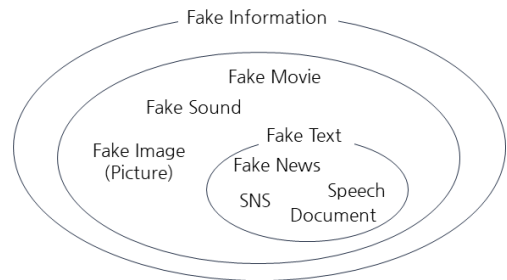


Fig. 2. Fake Information

가짜정보의 유형에는 허위정보, 거짓정보, 오인정보, 루머, 패러디, 풍자적 정보가 있으며, 이 중에서 가짜 뉴스는 허위정보, 거짓정보 형태로 많이 나타난다. 소셜네트워크서비스의 확산으로 뉴스형식을 갖추지 않은채 오인정보나 루머의 형태로 퍼지기도 한다[3].

2.1 허위정보

허위정보는 의도적으로 만들어지고, 오해를 부르는 정보이다. 정보의 내용은 사실이지만 표현을 달리하거나 과

장해서 전달함으로써 구독자가 다른 해석을 할 수 있게 하는 경우도 포함된다.

2.2 거짓정보

거짓정보는 고의로 조작한 정보를 사실로 가장한다. 허위정보와 개념을 혼동하기 쉬우나 거짓정보는 판단유류, 전설과는 다른 개념이다.

2.3 루머

루머는 근거없이 퍼지는 소문으로, 불확실한 내용이 대부분이다. 자연어처리 연구의 관점에서 볼 때 루머는 근거가 없기 때문에 가짜인지 아닌지를 판별하기에 어려운 점도 있다. 사람들에게는 정보의 사실성이 중요하지 않을 수 있다.

2.4 그 외(오인정보, 패러디, 풍자적 가짜 뉴스)

오인정보는 정보를 얻는 사람과 전파하는 사람이 정보의 사실성을 인식하지 못한 채 전파되는 가짜 정보이다. 패러디/풍자적 뉴스는 사람들이 정보의 대상이 허구임을 인지한 상태에서 주고 받는다.

3. 국내외 가짜 뉴스 탐지 사례

Poynter[4]는 언론인을 배출하기 위한 미디어 교육 기관으로써 뉴스의 사실을 확인하고 지역 뉴스 회사의 역량을 강화시키기 위해 노력하는 등 언론이 올바른 역할을 할 수 있도록 방향을 제시해 주고 있다. Poynter에 의해 설립된 International Fact Checking Network(IFCN)은 매년 전 세계의 팩트체크 기관을 위한 포럼을 개최하고, 각국의 팩트체크에 대한 리포트를 제공한다. 또한 팩트체크 기관을 위한 강령을 제공하고 있어 전세계 많은 팩트체크 기관은 IFCN이 제시하는 기준을 따르거나 이와 비슷한 기준을 제시하며 팩트 체크를 하고 있다.

3.1 영어 뉴스 탐지 사이트

3.1.1 Politifact

Poynter에서 운영하고 있는 팩트 체크 사이트이다. 뉴스기사를 True, Mostly True, Half True, Mostly False, False, Pants On Fire(새빨간 거짓말) 6단계로 나눠 라벨링한다. 라벨링 기준은 1) 기사의 내용이 사실

인가? 2) 다른곳에서 기사를 접할 방법이 있는가? 해당 기사의 해석이 개방적인가? 3) 기사내용에 대한 증거가 있는가? 그 증거가 사실인가? 4) 과거에 비슷한 기사내용을 어떻게 다루었는가? 해당 기사의 내용에 대해 판단할 수 있는 규정이 있는가? 이다. 에디터 3명과 리포터가 위 기준에 따라 투표 방식으로 다수결로 라벨을 결정한다. 의견이 갈리면 리포터가 의견 제시 후 다시 논의하게 된다.

3.1.2 Snopes

에디터들이 인터넷에 있는 많은 종류의 루머에 대해서 조사하여 True, Mixture, False, Mostly True, Outdated, Misattributed, Mostly False, Miscalcaptioned, Scam, unproven, Legend, Correct Attribution 12개 중 하나로 결과를 낸다.

3.1.3 Claimbuster

2016년 미국 대선과 관련한 기사를 팩트 체크하는 데 모를 소개한다. 원하는 정보를 입력하면 관련 내용과 가장 비슷한 기사를 찾아 기존에 판단했던 결과를 보여준다. 이 결과 역시 단계별로 나눠 라벨링 기준에 따라 보여준다[5].

3.1.4 Fiskkit

코멘트를 원하는 온라인 기사의 링크를 입력하면 기사 안에 있는 문장에 대한 태그가 달린다. True/False(True, False, Matter of Opinion), Fallacy(Overly General, Cherry Picking), Descriptive(Unsupported, Overly Simplistic, Biased Wording), Complimentary(Insightful, Well Reserched, Funny)가 링 형태로 나타난다.

3.2 한국어 뉴스 탐지 사이트

3.2.1 Factcheck.snu

서울대 언어정보연구소에서 운영하는 한국어기반 팩트 체크 시스템으로 Politifact의 6가지 라벨을 사용한다. 검증방법은 언론사가 내용을 업로드하면 다른 언론사와 함께 팩트 체크를 한다.

3.2.2 조선일보

조선일보의 팩트 체크 시스템으로 2017 대선 후보들의 기사와 발언, 루머에 대한 검증을 하는데 사용되었다.

사실, 일부만 사실, 사실 아님, 말 바꿈, 판정보류로 라벨링되어 기사를 나눴다. 이 시스템은 2017년 선거운동이 진행되는 당시 일시적인 기간 동안만 운영하여 현재 추가 기사는 업데이트 되고 있지 않다.

4. 가짜 뉴스 연구동향

가짜 뉴스 탐지는 가짜 뉴스 데이터의 범위와 형식, 기술적 접근법 등에 따라 다양한 기준으로 분류하여 연구될 수 있다. Fig. 3은 최근 3년간 논문 중에서 가짜 뉴스 혹은 클릭베이트(Clickbait)의 키워드로 검색되는 논문 53편을 분석한 결과이다. 이 중에는 서베이 논문 3편과 데이터 구축 논문 2편이 포함되어 있다.

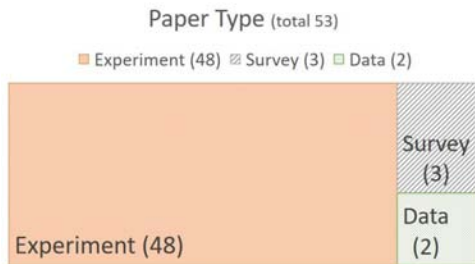


Fig. 3. Fake news Paper Type

4.1 데이터

가짜 뉴스 탐지를 위한 최근 연구논문에서 사용하는 데이터셋은 Fig. 4와 같이 뉴스 기사를 많이 사용한다. 서베이 논문 3편을 제외한 50편 중에서 뉴스 기사 데이터를 사용한 논문은 24편이고, SNS(Social Network Service) 데이터를 사용한 논문은 19편이다. 뉴스 기사와 SNS 데이터를 함께 사용한 논문은 5편이다. 뉴스 기사와 SNS 데이터를 함께 사용한 논문은 5편이다.

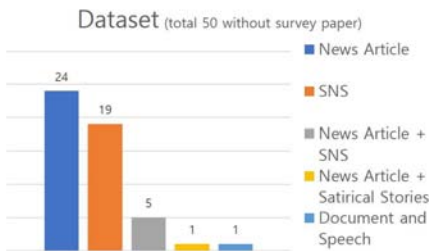


Fig. 4. Fake news Paper Dataset

뉴스 기사 데이터로 Fake News Articles, BuzzFeed, PolitiFact 등을 수집하였고, 데이터간의 균형을 맞추기 위해 3360개의 가짜 데이터와 3360개의 진짜 데이터를 사용한다[6].

뉴스 기사 외에도 소셜 네트워크 서비스의 트위터와 Reddit 데이터셋을 활용한다. 최근의 가장 인기있는 주제 9개를 선정하였고, 트레인 데이터 5217개, 개발 데이터 1485개, 테스트 데이터 1825개로 총 8529개이다[7].

또 2016년 미국 대통령 선거와 관련된 문서에서 임베딩 트레이닝 데이터셋 15,059개, 다른 모델과 비교하기 위한 평가 데이터셋 2,602 문장, 힐러리 클린턴과 도널드 트럼프의 연설문에서 Weakly Labelled 데이터셋 37,732 문장을 수집하였다[5].

Fig. 5와 같이 대부분의 데이터셋이 영어이지만 50편의 논문(without survey paper 3편) 중 중국어 3편, 한국어 3편, 불가리아어 1편이 데이터셋을 활용하였다.

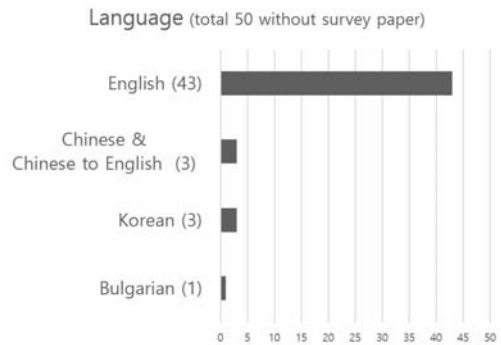


Fig. 5. Fakenews Paper Language

한국어로는 2016년 1월부터 2017년 10월까지 네이버와 다음에서 크롤링한 400만개의 뉴스 기사가 있다[8].

4.2 접근법

위 내용과 같이 가짜 뉴스 탐지 데이터의 경우, 뉴스 기사 형태로만 연구가 이뤄지는 것이 아니라 다양한 형태의 데이터가 사용되고 있다. 또한 데이터 분석 뿐만 아니라 가짜 뉴스 생산자의 유형과 데이터가 확산되는 양이나 시간을 함께 적용하여 가짜 뉴스를 탐지한다. 2016년 미국 대통령 선거가 이뤄지면서 가짜 뉴스에 대한 연구가 많은 관심을 받게 되었다. 머신러닝을 활용한 방법이 많이 사용되었으나 최근 3년간은 Neural Networks, Directional Long Short-Term Memory network,

Bi-Directional Recurrent Neural Network, Gated Recurrent Units[9], Transformers, Ensemble[10]을 사용해서 대부분의 연구가 이뤄지고 있다.

Fig. 3의 실험논문 48편 중에서 어떤 접근법을 사용하였는지 분석하였다. Fig. 6에서 결과를 보여준다. 딥러닝 기반 25편, 머신러닝 기반 17편 그리고 룰베이스와 딥러닝&머신러닝을 함께 적용한 논문 각각 2편씩이었다.

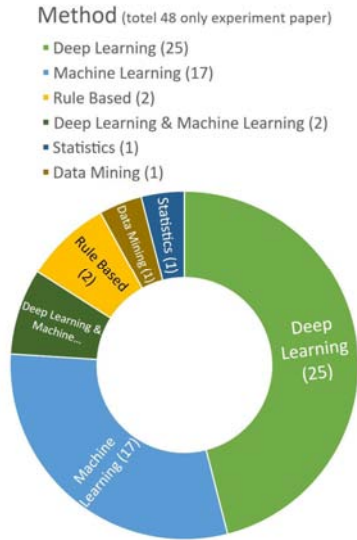


Fig. 6. Fake news Paper Method

가짜 뉴스를 탐지하는 방법은 크게 가짜 뉴스 탐지, 클릭베이트 탐지, 루머 탐지 3가지로 분류된다. 가짜 뉴스 탐지의 경우 뉴스 기사를 데이터로 사용하여 뉴스 헤드라인과 본문의 관계를 보는 것이 대부분이다[11]. 클릭베이트 탐지는 가짜 뉴스 탐지와 비슷한 접근법을 사용하

지만 낚시성의 문구로 구독자를 속이기 위한 목적이 있다[12]. 마지막으로 루머 탐지는 소셜 네트워크 서비스의 데이터를 많이 사용하게 된다. 경우에 따라 가짜인지 아닌지를 결정하기 위한 내용이 없을 수도 있다[7]. 챗터4에서 언급한 논문들을 Table 1에 정리하였다.

5. 국내외 챌린지

가짜 뉴스를 탐지하기 위한 챌린지도 다양한 방법으로 이뤄지고 있다. 챌린지의 태스크, 주어지는 데이터의 종류와 형식에 따라서 제안되는 모델과 접근법은 달라진다.

5.1 2017 Fake News Challenge

2017년 열린 Fake News Challenge는 토픽, 주장, 이슈와 관련된 두 텍스트의 상대적인 관점을 평가한다. 데이터는 헤드라인 2,587개, 본문 2,587개가 주어졌으며, 본문을 Agree, Discuss, Disagree, Unrelated로 분류하는 과제이다[13,14]. 1등 SOLAT 팀의 상대 점수(Relative Score)는 82.02, 2등은 81.97, 3등은 81.72 점이다.

5.2 2017 Clickbait Challenge

소셜미디어 포스팅 글이 클릭베이트인지 아닌지를 분류하는 과제이다. 데이터는 트위터의 트윗을 사용했으며 트레이닝 게시물 2495개, 레이블이 지정되지 않은 게시물 80012개, 검증 게시물 19538개가 주어졌다[15]. 1등 팀 albacore의 Mean Squared Error (MSE)값은 0.032이고, F1-Score (F1)값은 0.670으로 측정되었다.

Table 1. Summary of works

Author	Dataset	Method
Casper Hansen et al., 2019 [5]	Document and Speech about 2016 U.S election Embedding Training Dataset:15,059 document	Neural Network
Hamid Karimi & Jiliang Tang. 2019 [6]	3,360 fake and 3,360 real news article	Bi-directional Long Short-Term Memory network
Martin Fajcik et al., 2019 [7]	Twitter and Reddit Dataset 8,529	Transformers
Seunghyun Yoon et al., 2019 [8]	4million Korean news articles	Neural Network
Amin Omidvar., et al., 2018 [9]	38,517 Twitter Tweets	Bi-directional Gated Recurrent Units
Lam Pham. 2019 [10]	320,767 news title pairs in Chinese	Ensemble
Jooyeon Kim et al., 2019 [11]	79,416 users and 175,389 tweets and retweets	Bayesian nonparametric
Ankesh Anand et al., 2016 [12]	15,000 news headlines	Bi-Directional Recurrent neural network Long Short Term Memory Gated Recurrent Units

5.3 2019 WSDM Challenge

가짜 뉴스 기사 제목 A와 가짜인지 아닌지 알 수 없는 뉴스 기사 제목 B를 비교하여 분류한다. B가 A와 같으면 AGREED로 분류되어 가짜 뉴스가 되고, 같지 않으면 DISAGREED로 가짜 뉴스가 아니며, 관계가 없으면 UNRELATED로 라벨링된다. 트레이닝 데이터셋 320,767개와 테스트 데이터셋 80,126개가 사용되었다 [16]. 1등 팀의 private score는 88.28%, public score는 88.19%로 나타났다.

5.4 2017 인공지능 R&D

임무1과 임무2로 나눠 진행되었다. 임무1은 제목과 본문 내용의 적합성을 판별하고 임무2는 본문 중 맥락에 관계 없는 내용을 검출한다. 트레이닝 데이터는 제공되지 않았으며 실제 한국어 뉴스 기사를 사용하였다.

6. 결론

기존에 가짜 뉴스 탐지와 관련해 많은 연구가 이뤄지고 있어 기술은 많이 소개되었으나 실제로 사람들이 가짜 뉴스인지 아닌지를 실시간으로 확인해 볼 수 있는 자동화 틀은 상용화되지 않았다. 가짜인지를 확인하려면 반드시 사람이 확인하는 절차를 거쳐야 하고, 시간이 흐르는 사이에 가짜 뉴스는 온라인과 미디어 플랫폼을 통해 확산될 수 있다. 또 사람도 가짜 뉴스에 대한 정확한 구분이 불가하며 한눈에 파악하기가 쉽지 않다. 지금도 많은 양의 뉴스가 끊임없이 작성되고 있고 실시간으로 생성되는 가짜 뉴스는 더욱 판별하기가 어렵다.

가짜 뉴스 판별을 위한 영어 데이터는 기존에 많은 연구자들이 수집을 하였지만 한국어로 된 데이터는 부족하다. 한국어 소셜네트워크서비스 데이터가 마련된다면 가짜 뉴스 확산의 외부적인 요인을 관찰 할 수 있을 것이다. 그 결과 한국문화가 반영되어 있음을 확인 할 수 있을 것이고, 가짜 뉴스 확산의 흐름을 파악할 수 있을 것이다.

본 논문에서는 가짜 뉴스 탐지에 관련한 연구동향을 조사하였다. 먼저 가짜 뉴스를 탐지하는 기준과 사례를 살펴봤고, 가짜 뉴스 탐지를 위해 어떤 연구들이 이뤄지고 있는지, 실제 탐지 기능이 어떻게 사용되고 있는지 확인했다. 향후 연구 계획으로는 2017 Fake News Challenge(chapter 5.1)에서 오픈한 데이터셋 FNC-1으로 구문정보를 활용하여 문장을 정확하게 이해하고 표

현하는 모델을 제안할 예정이다. 이를 위해서는 단어 임베딩이 중요하며, 가짜 뉴스에 적합한 언어적 구조를 활용하여 가짜와 진짜를 판별하기 위한 적합한 문장 표현 방법을 찾아야 한다.

REFERENCES

- [1] Reporters Without Borders. (2019). *Reporters Without Borders*. 2019 WORLD PRESS FREEDOM INDEX. www.rsf.org
- [2] Reuters Institute for the Study of Journalism. (2018). *Reuters Institute Digital News Report 2018*. Oxford : Reuters Institute for the Study of Journalism. www.digitalnewsreport.org
- [3] Korean Society for Journalism & Communication Studies. Korea Press Foundation. (2017). *Concepts and countermeasures of Fake News*, www.kpf.or.kr
- [4] The Poynter Institute. (1975). *Poynter*. About. www.poynter.org
- [5] C. Hansen, C. Hansen, S. Alstrup, J. G. Simonsen & C. Lioma. (2019). Neural Check-Worthiness Ranking with Weak Supervision: Finding Sentences for Fact-Checking. *International World Wide Web Conference*.
- [6] H. Karimi & J. Tang. (2019). Learning Hierarchical Discourse-level Structure for Fake News Detection. *Annual Conference of the North American Chapter of the Association for Computational Linguistics*.
- [7] M. Fajcik, L. Burget & P. Smrz. (2019). Determining the Rumour Stance with Pre-Trained Deep Bidirectional Transformers. *International Workshop on Semantic Evaluation*.
- [8] S. Yoon, et al. (2019). Detecting Incongruity Between News Headline and Body Text via a Deep Hierarchical Encoder. *AAAI Association for the Advancement of Artificial Intelligence*.
- [9] A. Omidvar, H. Jiang & A. An. (2018). Using Neural Network for Identifying Clickbaits in Online News Media. *2017 Clickbait Challenge*.
- [10] L. Pham. (2019). Transferring, Transforming, Ensembling: The Novel Formula of Identifying Fake News. *2019 WSDM International Conference on Web Search and Data Mining*.
- [11] J. Kim, D. Kim & A. Oh. (2019). Homogeneity-Based Transmissive Process to Model True and False News in Social Networks. *ACM International Conference on Web Search and Data Mining*.
- [12] M. M. U. Rony, N. Hassan & M. Yousuf. (2016). We used neural networks to detect clickbaits: You won't believe what happened next!. *ECIR European Conference on Information Retrieval*.
- [13] A. Hanselowski, et al. (2018). A retrospective analysis of the fake news challenge stance detection task. *2017 Fakenews Challenge*.

- [14] B. Riedel, I. Augenstein, G. P. Spithourakis & S. Riedel. (2017). A simple but tough-to-beat baseline for the Fake News Challenge stance detection task. *2017 Fakenews Challenge*.
- [15] Y. Zhou. (2017). Clickbait detection in tweets using self-attentive network. *2017 Clickbait Challenge*.
- [16] S. Liu, S. Liu & L. Ren. (2019). Trust or Suspect? An Empirical Ensemble Framework for Fake News Classification. *2019 WSDM International Conference on Web Search and Data Mining*.

좌 회 정(Hee-Jung Jwa) [장학원]



- 2008년 2월 : 경희대학교 신문방송학과 (학사), 국어국문학과 (학사)
- 2017년 8월 : 고려대학교 컴퓨터학과 (박사수료)
- 2017년 9월 ~ 현재 : 고려대학교 컴퓨터학과 NLP&AI Lab. 연구원
- 관심분야 : 자연어처리, 인공지능, 딥러닝

· E-Mail : jwaheejung@korea.ac.kr

오 동 석(Dong-Suk Oh) [장학원]



- 2014년 2월 : 충북대학교 정보통신공학부 (공학사)
- 2016년 2월 : 서강대학교 컴퓨터공학과 (공학석사)
- 2018년 8월 : 다이퀘스트 근무 전임연구원
- 2018년 3월 : NHN엔터네인먼트 근무 전임연구원

· 2019년 3월 ~ 현재 : Human-inspired 복합지능 연구센터 연구원

· 관심분야 : 인공지능, 지식표현, 자연어처리

· E-Mail : inow3555@gmail.com

임 희 석(Heui-Seok Lim) [장학원]



- 1992년 2월 : 고려대학교 컴퓨터학과 (이학사)
- 1994년 2월 : 고려대학교 컴퓨터학과 (이학석사)
- 1997년 2월 : 고려대학교 컴퓨터학과 (이학박사)
- 2008년 ~ 현재 : 고려대학교 정보대학 컴퓨터학과 교수

· 관심분야 : 자연어처리, 인공지능, 딥러닝

· E-Mail : limhseok@korea.ac.kr