

그래프 임베딩을 활용한 코로나19 가짜뉴스 탐지 연구 - 사회적 참여 네트워크의 이용 여부에 따른 탐지 성능 비교*

정이태

국민대학교 비즈니스IT전문대학원
(jeongiatae@kookmin.ac.kr)

안현철

국민대학교 비즈니스IT전문대학원
(hcahn@kookmin.ac.kr)

인터넷 및 모바일 기술의 발달과 소셜미디어의 확산으로 인해 다량의 정보들이 온라인 상에서 생성, 유통되고 있다. 이 중에는 대중에게 도움이 되는 유익한 정보들도 있지만, 역기능을 하는 이른바 가짜뉴스들도 함께 유통되고 있다. 지난 2020년 코로나19의 전세계적인 확산 이후, 온라인 상에는 이와 관련한 수많은 가짜뉴스들이 유통되었다. 다른 가짜뉴스들과 달리 코로나19와 관련된 가짜뉴스는 사람들의 건강, 나아가 생명까지 위협할 수 있다는 점에서 그 심각성이 매우 크다고 할 수 있다. 때문에 코로나19와 관련한 가짜뉴스를 자동으로 탐지하고, 이를 예방하는 지능형 기술은 사회적 건강도를 제고하는데 매우 의미 있는 연구주제라 할 수 있다. 이러한 배경에서 본 연구에서는 코로나19 관련 가짜뉴스 탐지를 효과적으로 수행하기 위해 그래프 임베딩 방법 중 하나인 Graph2vec을 활용한 방법을 제안한다. 가짜뉴스 탐지에 대한 주류 방법은 뉴스 콘텐츠 기반 즉, 텍스트에 대한 특징 분석으로 진행되었으나 본 연구에서는 사회적 참여 네트워크 내에서의 정보 전달 관계를 추가로 활용함으로써 보다 효과적으로 코로나19와 관련된 가짜뉴스를 탐지할 수 있었으며 성능 측면에서 정확도 향상을 확인할 수 있었다.

주제어 : 가짜뉴스 탐지, 그래프 임베딩, 사회적 참여 네트워크, Graph2vec, 코로나19

논문접수일 : 2021년 11월 26일 논문수정일 : 2022년 2월 26일 게재확정일 : 2022년 3월 8일

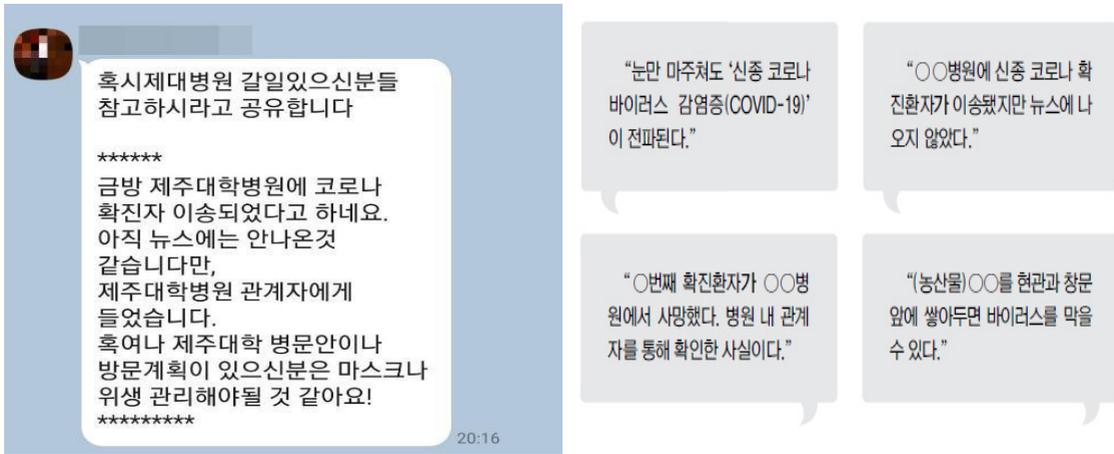
원고유형 : 학술대회용 Fast Track 교신저자 : 안현철

1. 서론

디지털 기술이 발달함에 따라 다양한 정보들이 TV, 라디오와 같은 전통적인 매체뿐 아니라 페이스북, 카카오톡, 인스타그램 등 이른바 소셜 미디어와 같은 온라인 채널을 통하여 전파, 확산되고 있다. 근래에는 코로나19의 영향으로 비대

면으로 활동하는 사람들이 증가하여 온라인 채널 특성인 편리함, 신속함 등으로 인해 소셜미디어에서 확산되는 정보가 오프라인 대비 더 빠른 확산세를 보이고 있다. 이렇게 소셜미디어 상에서 확산되고 있는 정보 중에는 정보 수용자에게 이익이 되는 좋은 정보가 있는 반면 해가 되는 부정적인 정보도 존재한다. 이 중 대표적인 부정

* 본 논문은 2021년 한국지능정보시스템학회 추계학술대회에서 Fast Track 심사대상으로 선정된 동일 저자의 발표논문을 확장, 발전시킨 것임. 본 연구는 정보통신산업진흥원의 2021인공지능 고성능 컴퓨팅 자원사업의 연구결과로 수행되었음. 본 논문은 교육부 및 한국연구재단의 4단계 두뇌한국21 사업(4단계 BK21 사업)으로 지원된 연구임.



<그림 1> 소셜미디어 상의 코로나19 가짜뉴스 사례(출처: 김종광, 2020; 황승민, 2020)

적인 정보라 할 수 있는 가짜뉴스는 개인을 넘어 사회 전반적인 부분에 부정적인 영향을 유발하고 있다. 특히 지난 2020년 코로나19의 발발 이후, 소셜미디어 상에서 전파되고 있는 다양한 종류의 코로나19와 관련된 가짜뉴스들은 심각한 사회문제가 되고 있다.

Shejkh et al.(2021)에 따르면 코로나19는 감염자 1명이 전염시킬 수 있는 전파력 지수분석에서 평균 2~4명사이로 나타났다. 이는 유사한 감염성 바이러스인 메르스의 전파력 0.8명~1.2명과 비교해 매우 높은 전파력을 가진 것이라 할 수 있으며, 최근 사람의 세포로 들어갈 때 중요한 역할을 하는 스파이크 단백질에 따라 알파, 델타, 델타플러스, 그리고 최근의 오미크론에 이르기까지 다양한 바이러스의 변이가 지속해서 탄생하고 있어 인류 전체에 큰 위협이 되고 있다. 이처럼 코로나19는 우리의 건강, 나아가 생명과도 직결되는 전염병이기에 많은 대중이 코로나 추이에 관해 관심을 가지는 중이다. 하지만 주목을 받는 만큼 발생하는 다양한 정보들 또한 존재하

며 그 중에는 다음 <그림 1>과 같이 가짜뉴스들도 다수 유통되고 있다.

이러한 코로나19 관련 가짜뉴스들은 사회에 다음과 같은 부정적 현상들을 초래할 수 있다.

첫째, 근거 없는 낭설들을 전파함으로써 혼돈을 야기한다. 방역 및 예방이 중요한 시기에 가짜뉴스들은 코로나19 백신에 대한 부작용을 과대 증폭시킴으로써 사람들로 하여금 백신에 대한 판단을 흐리게끔 만들어 예방접종에 대한 부정적인 영향을 미치고 있다(고정민, 2021). 또한 의료인을 사칭하여 전문성을 악용하는 행위들로 인해 대응과 위기 극복이 절실한 요즘 공동체 협력에 차질을 빚고 있다.

둘째, 가짜뉴스 대다수는 기존 뉴스 대비 다소 자극적인 점을 관찰할 수 있다. 이에 대중들이 언론이나 미디어에 대한 수용자극의 임계치가 높아짐으로써 수용자의 심리적 기체에 부정적인 영향을 줄 수 있다. 이는 거시적으로 오직 사실에만 집중해야 할 언론 보도에 대해 사람들의 반응을 바꿀 만큼 심각한 영향을 초래한다.

셋째, 특정 소수 집단에만 이익을 주는 사회 불공정한 이익을 발생하여 경제적으로 부정적인 영향을 야기한다. 예를 들면 특정 약품 혹은 식품이 코로나에 효능이 있다는 근거 없는 가짜뉴스를 전파함으로써 해당 제품에 대한 수요를 폭증시켜 불공정 거래를 유발하는 일이 종종 발생하고 있다(조승한, 2021). 이것은 국가적으로 심각한 경제적 혼란을 초래할 수 있으며, 잘못된 정보로 인해 범국가적, 전세계적으로 많은 사람들의 사망까지 유발할 수 있는 위험이 있어 그 폐해가 심각할 수 있다.

이상의 문제들로 인해 보건당국에서 직접 웹 혹은 앱을 통해 국민들에게 정확한 정보를 전송하고자 하는 노력하고 있는 등 가짜뉴스 전파에 대한 상황들을 타개하고자 여러 방면에서의 조치가 시도되고 있다. 그리고 다른 한 편에서는 가짜뉴스를 전문가 기반, 인공지능 기반 등을 통하여 판별하고자 하는 시도들도 진행되고 있다. 그 중, 전파되는 뉴스의 텍스트, 즉 콘텐츠에 대한 정보들인 게시자 특성, 내용, 생성일 등을 토대로 가짜뉴스를 탐지하고자 하는 방법론인 텍스트 기반 탐지 방법론이 주류를 이루며 학계에서 활발하게 연구가 진행되고 있다. 하지만 이러한 텍스트 위주의 접근법은 언어적 특성에만 과도하게 의존한다는 제한이 있다(Shim et al., 2021). 특히 뉴스 자체의 특성만을 고려하여 탐지하기에 그 뉴스가 어떤 방식으로 크게는 우리 사회, 좁게는 우리의 가치관 및 인식에 접근할 수 있었는가에 대한 사항은 전혀 고려하지 못한다는 한계점이 있다.

더불어 기존 단방향 정보교류에서 양방향성으로의 방향성 확장을 토대로 사회적 참여를 제안한 플랫폼인 소셜 네트워크에서는 정보의 전파력이 훨씬 빠르다. 때문에 가짜뉴스 역시 소셜

네트워크에서 더욱 빨리 전파된다. 권오성(2017)에 따르면 가짜뉴스는 소셜미디어에 특화되어 더욱 타 매개체에 비해 빠르게 전파됨을 확인할 수 있었다. 이는 SNS 이용률이 세계 2위인 우리나라에서 가짜뉴스 탐지가 시급하게 다루어질 중대한 문제임을 시사한다(이영호, 2021).

이에 본 연구에서는 소셜미디어 내에서의 가짜뉴스는 본질적으로 상호 간의 교류를 통해 전달되므로 정보 전파자, 즉 네트워크 내에서 발생하는 참여 관계에 대한 접근을 추가로 고려하는 것이 유리할 것이라는 전제 하에 이러한 참여 관계를 그래프화 하여 추가로 활용한 그래프 임베딩(graph embedding) 기반 가짜뉴스 탐지 방법을 제안한다. 그래프 임베딩 방법론이란 그래프 데이터를 컴퓨터가 효율적으로 이해하기 위해 벡터화(vectorize)한 방법론으로서, 본 논문에서는 Graph2vec이라는 방법론을 활용한다. 본 연구에서는 코로나19에 특화된 가짜뉴스 데이터인 COAID를 가공하여 그래프화 한 후 진짜 및 가짜뉴스에 대한 참여를 토대로 형성된 네트워크를 벡터화하여 이를 가짜뉴스 탐지에 추가로 활용하고자 한다. 본 연구에서 제안하는 Graph2vec 기반 가짜뉴스 탐지 방법론은 모든 유형의 가짜뉴스 탐지에 적용될 수 있는 방법이지만, 본 연구의 실증분석 대상으로 삼은 코로나19 가짜뉴스의 경우 코로나19에 대한 공포감을 갖고 있는 일반 대중들을 통해 SNS상에서 빠르게 확산되는 특징이 있기 때문에 사회적 참여 네트워크의 이용 여부를 추가로 고려하는 본 연구의 제안모형이 특별히 더 의미가 있는 접근법이 될 것으로 판단하였다.

이후 본 논문은 다음과 같이 구성된다. 2장 관련된 연구에서는 가짜뉴스의 정의 및 유래 그리고 탐지 방법론에 대한 기존 연구들에 대해 살펴볼

것이며, 3장에서는 본 연구에서 제안하는 모델에 대해 소개한다. 이어 4장 실증분석에서는 앞서 제안한 방법을 실제 코로나19와 관련한 진짜 및 가짜뉴스 데이터에 대해 적용했을 때, 탐지 성능이 어떻게 개선되었는지를 살펴본다. 끝으로 5장 결론에서는 본 논문에서 제안한 방법이 갖는 학술적, 실무적 의의 및 향후 연구 방향에 대해 논의한다.

2. 관련연구

2.1. 가짜뉴스

우리나라에서 가짜뉴스에 대한 관심은 2016년 미국 대선 과정에서 문제가 된 'Fake News'의 'Fake'를 '가짜'라고 번역하면서 높아지기 시작했다. 가짜뉴스가 신조어로서 등장한 지 얼마되지 않은 만큼 가짜뉴스에 대한 정의는 다양하며 정확한 정의는 정착되지 않았다. 명확하지 않은 가짜뉴스의 개념을 학계나 언론계에서 연구자들이 각자 다양한 관점에서 정의를 내리고 있으나, 가장 널리 통용되는 정의에 따르면, 가짜뉴스는 “정치·경제적 이익을 목적으로 사실이 아닌 내용과 정보를 언론 보도의 형식으로 만든 후 의도적으로 유포한 거짓 정보”로 정의된다(엄정윤, 정세훈, 2019; 황용석, 권오성, 2017). 최근에 등장한 가짜뉴스의 정의에서는 소셜미디어 상에서 전파되는 특성을 강조하기도 한다. 예를 들어, 이완수(2017)는 가짜뉴스에 대해 정의를 내리기를 “소셜 미디어를 통해 유통되는 확인되지 않은 불확실한 뉴스 또는 허위정보”의 개념이라 주장했다.

이처럼 소셜미디어는 최근 가짜뉴스의 주요

근원지로 간주되고 있다. 가짜뉴스의 목적은 재정적 또는 정치적으로 이득을 얻기 위해 작성되고 발간되며, 종종 주목을 끌기 위해 선정주의, 과장됨 또는 간과한 거짓 표제를 사용한다. 그러다 보니 정보 공유가 손쉬운 소셜미디어 상에서 유포될 때, 일반 대중들에 의해 더 쉽게 전파될 수 있다. 관련하여 Vosoughi et al.(2018)에서는 약 300만명이 총 450만회 이상 트윗한 12만 6천여건의 데이터를 토대로 정보 확산 양상에 대해 실증분석 하였는데, 그 결과 가짜뉴스는 진짜뉴스보다 리트윗되는 비율이 70%가량 높았으며 전파 속도는 진짜 뉴스보다 최대 20배가량 빨랐다는 결과를 확인할 수 있었다. 또한 소셜 미디어로 인해 범국가적으로 정보전달이 가능하게 되었다. 차미영(2020a; 2020b)에서는 각국의 소셜 네트워크 서비스내에서의 게시물을 분석한 결과 국가적으로 정보들이 재생산되어 범국가적으로 발생함을 확인하였고, 특히 소셜미디어는 정보의 사실성과 정확성에 대한 자체 검증 기제가 부족하기 때문에 가짜뉴스의 전파와 확산에 구조적으로 취약할 수밖에 없음을 지적했다.

이런 가운데 2020년 코로나19가 전세계적으로 유행을 시작한 이후에는 다양한 코로나19 관련 가짜뉴스들이 전세계 곳곳에서 유포되고 있다. 다음 <표 1>은 디지털 시대로 인해 범국가적으로 발생한 대표적인 코로나19 가짜뉴스 사례들을 소개하고 있다.

<표 1>에서 확인할 수 있듯이 지역사회, 세계, 중국, 아시아 각 나라에서 유포된 코로나19 관련 가짜뉴스는 그 종류가 다양하다. 우리나라에서는 최근 코로나19 백신 예방접종과 더불어 가짜뉴스가 유행하고 있는데, 국가에서 가짜뉴스 유통으로 인해 신속하게 접종이 이루어져야 하는 근래 혼란을 야기하는 상황이 발생한다.

<표 1> 지역별로 전파된 코로나19 관련 가짜뉴스의 대표사례(출처: 차미영, 2020a; 2020b)

세계로 퍼진 가짜뉴스	알코올 음료 섭취로 바이러스를 죽일 수 있다.
	헤어드라이어로 열을 가하면 바이러스가 죽는다.
	마늘, 참기름을 섭취하거나 코에 바르면 예방할 수 있다.
	소금물 가글로 예방할 수 있다.
	담배의 열기로 바이러스를 죽일 수 있다.
중국에만 퍼진 가짜뉴스	불꽃놀이로 바이러스를 소멸시킬 수 있다.
	울금(중국 약제)이 치료에 효과적이다.
	항생제나 항고혈압제를 복용해 치료할 수 있다.
아시아에만 퍼진 가짜뉴스	열약한 의료진 치우
	품질이 낮은 마스크를 여러 겹 써도 효과 있다.
지역사회에 특화된 가짜뉴스	미국 뉴욕 기차역이 문을 닫을 것이다.
	베이징을 봉쇄한다.
	한국시외버스터미널을 폐쇄한다.

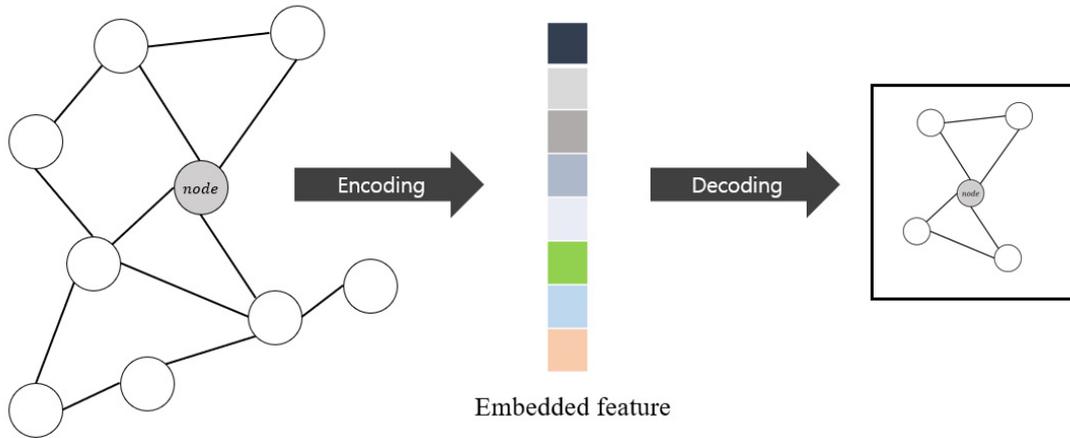
소셜 미디어는 본디 온라인상에서의 사람들 교류 혹은 개인의 관심 있는 정보 전달 및 습득을 위해 발생된 공간이다. 그러기에 소셜 미디어는 온라인에서 상호 간의 접점이 있거나 개인적 관심사에 기반해 미디어 활용이 이루어지기 때문에 오프라인 대비하여 개방적인 미디어지만 이용자들 간의 커뮤니티는 폐쇄적인 특징도 있는데, 이런 폐쇄적인 커뮤니티에서는 소셜미디어의 장점인 빠른 정보의 확산과 신속성이 앞서 언급한 정보공유에 대한 정확성 인지 와 같은 장치, 뉴스 가치 등들이 보장되지 않는다면 오히려 부작용을 일으킬 수 있다.

2.2. 그래프 임베딩

그래프(graph) 데이터는 노드(node)와 엣지(edge) 로 이루어진 데이터를 의미한다. 이때, 엣지란 특정 노드 간의 상호 작용의 집합을 의미한다. 그래프 데이터는 정해진 특정 조건에 맞게

저장되도록 설계된 데이터베이스인 정형 데이터와는 다르게 데이터 간 관계 중심의 표현 및 처리 가능한 구조를 지녔으며, NoSQL 계열의 데이터베이스로 구현된다. 그래프 데이터는 비정형 데이터로 불리며, 타 비정형 데이터(텍스트, 이미지)와는 다르게 데이터 자체를 노드와 엣지의 그래프 형태로 저장하여 데이터 간의 관계를 함축할 수 있다는 점에서 차별화된다. 사회 전반적인 영역에서 소셜미디어의 비중이 늘어남에 따라 소셜 데이터들도 여실히 증가세를 보이고 있으며 대다수 그래프 데이터 형태로 발생하고 있다. 증가하고 있는 그래프 데이터비중에 따라 이를 분석하기 위한 방법인 그래프 임베딩에 대한 관심이 학계 및 산업계에서 높아지고 있는 중이다.

그래프 임베딩이란 아래 <그림 2>에 묘사되어 있는 바와 같이 그래프화 되어 있는 고차원 데이터셋을 저차원 데이터셋으로 인코딩(encoding)하



〈그림 2〉 그래프 임베딩

여 컴퓨터가 이해할 수 있게 벡터화하는 기술을 의미한다. 그래프 임베딩을 인코더(encoder)-디코더(decoder) 관점에서 설명해보면 다음과 같다. 먼저 인코더 모델은 그래프 내의 각각의 노드를 임베딩 공간으로 맵핑을 해준다. 이후 디코더 모델은 인코더에서 발생한 저차원 노드 임베딩 값을 통해 기존 그래프에서 연결 되어있는 노드들 간의 정보를 활용하여 기존 그래프와 임베딩 된 값의 유사도를 비교하여 그래프를 임베딩 해 준다. 이 때 인코더에서 노드 및 엣지들에 대해 임베딩 공간으로 맵핑 과정에서 활용하는 정보의 활용치에 따라 얕은 임베딩(shallow embedding), 그래프 신경망(graph neural network)으로 나뉜다. 얕은 임베딩은 특정 노드의 인덱스에 대해서만 참조하여 인코딩을 해준다. 반면 그래프 신경망은 특정 노드의 인덱스뿐만 아니라 노드의 특성 그리고 노드와 연결된 노드들의 구조, 노드가 가지고 있는 특성 등과 같은 다양한 정보를 토대로 인코딩을 해 준다는 점에서 차이가 있다.

2.3. 가짜뉴스 탐지

정보의 홍수라 불리는 요즘 정보의 수요가 공급을 따라가질 못한다고 표현할 정도로 많은 정보들이 데이터화 되어 온라인상에서 공유되고 있다. 정보를 쉽게 접할 수 있는 기회가 많아 짐에 따라 유용한 정보를 기반으로 특정 문제해결에 이용하여 가치를 창출할 수 있다는 장점이 있긴 하나 정보의 공급이 늘어난 만큼 정확한 정보 인가에 대해 판단하는 능력이 중요해지고 있다. 허나 근거 불명확, 확증 편향 등과 같은 여러 방해 요소들로 인해 자체적으로 정보에 대한 판단이 어려워지고 있다.

이에 다양해진 뉴스의 콘텐츠로부터 정확한 뉴스인지 부정확한 뉴스인지에 대한 가짜뉴스 탐지 분야에 대해 활발한 연구가 진행 중이다. 윤영석 등(2017)은 가짜뉴스 탐지 연구에서 사용되는 접근법을 크게 비기술적 방법과 기술적 방법의 2가지 유형으로 분류하였다. 비기술적 접근으로는 전문가 판단, 집단 지성 판단의 방법론이 있으며, 기술적 접근으로는 언어 신호 분석,

네트워크 분석이 있다. 그 외에 비기술적 접근과 기술적 접근의 방법론을 둘 다 사용한 하이브리드 분석 기법도 존재한다. 이에 본 장에서는 본 연구에서의 제안 방법론과 관련 있는 인공지능 기반 탐지에 초점을 두고 Bondielli and Marcelloni(2019)의 구분에 따라 3가지 접근법, 즉 내용 기반(content-based), 맥락 기반(context-based), 혼합(hybrid)의 관점에서 기존 연구에서 활용된 가짜뉴스 탐지 기법들을 고찰하였다.

2.3.1. 내용 기반 가짜뉴스 탐지

내용 기반 탐지 기법은 콘텐츠 내의 2가지 특성인 언어적 특성, 시각적 특성을 가지고 추출하는 방법이다. 언어적 특성은 문자, 단어, 문장 등 문서 내의 특성들을 자연어 처리 방법론을 활용하여 다각적으로 추출한 특성이며, 시각적 특성은 뉴스 내에 등장하는 시각적 이미지 및 동영상에서의 특성을 의미한다. 현윤진과 김남규(2018)은 뉴스 자체적 특성, 즉 언어적 특성에 대한 과도한 의존을 한계점으로 보고, 외부 정보인 트위터 데이터를 결합한 가짜뉴스 탐지 기법을 제안하여 우수한 성능을 보였다. 허나 트위터 데이터 역시 토픽 모델링 방법론을 활용한 언어적 특성으로서 언어적 특성의 한계점인 콘텐츠에 의존성이 강하다는 점 존재한다. 윤태욱과 안현철(2018)은 텍스트 마이닝 방법의 하나인 토픽모델링과 TF-IDF를 이용해 짧은 문장을 정량화된 값인 벡터로 도출하여 기계학습에 적용하는 방법론을 제안하였다. 심재승 등(2019)은 언어적 특성을 기반으로 가짜뉴스를 탐지하는 방법을 제안하였는데, 구체적으로 문서 관점에서의 중요한 정보만을 요약하는 방법인 Doc2vec을 이용하여 생성된 요약된 데이터를 활용한 모델을 제안

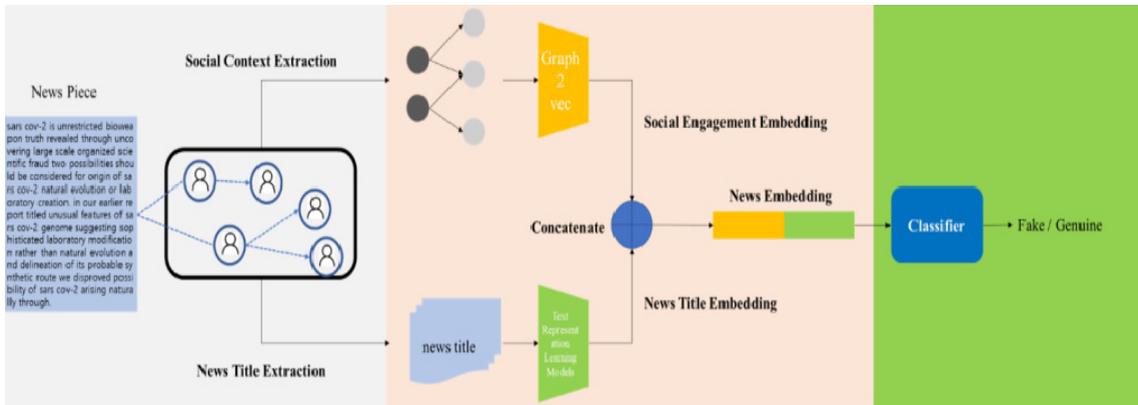
하였다. Wang et al.(2018)은 시각적 특성과 언어적 특성 모두 이용한 멀티모달(multi-modal) 특성을 활용하여 가짜뉴스를 탐지하였다.

2.3.2. 맥락 기반 가짜뉴스 탐지

맥락 기반 탐지 기법은 소셜 네트워크 내에서의 맥락을 토대로 가짜뉴스를 탐지하는 방법이다. 이때, 맥락은 사용자, 글(post), 그리고 네트워크를 의미하며 구체적인 의미는 다음과 같다. 사용자의 특성은 소셜 네트워크 내에서 활동하는 사용자 자체의 특성을 의미한다. 예를 들자면 생일, 팔로워 수 등을 뜻한다. 글 특성은 소셜 미디어 내에 올린 글에 사용자가 직접적으로 표현한 감정 혹은 의견을 의미한다. 네트워크 특성은 소셜 미디어 내에서의 흥미, 토픽, 그리고 관계로부터 형성된 네트워크 형태를 의미한다. 관련 연구로 Shu et al.(2019)은 소셜미디어 내의 정보 전달 관계를 ‘작성자’, ‘뉴스’, ‘독자’로 범주를 지정하고, 각 범주 간의 관계를 임베딩하는 방법론인 ‘TriFN’을 제안하였다.

2.3.3. 혼합 가짜뉴스 탐지

혼합 방법론은 내용 기반 방법론과 맥락 기반 방법론에서의 특성들을 조합하여 가짜뉴스를 탐지하는 방법이다. 예를 들어 Ren and Zhang(2020)은 가짜뉴스 데이터를 그래프화 하여 가짜뉴스를 탐지하였다. 이들은 노드 수준(node-level), 스키마 수준(schema-level) 마다 계층적 특성을 반영하고자 어텐션(attention) 메커니즘을 적용하였다. 이때 어텐션 메커니즘이란 모델로부터 출력된 값을 입력값과 대조하여 출력값과 입력값이 얼마나 연관이 있는지 유추하여 특정 부분에 집중함으로써 효율적으로 추론하고자 제안된 알고



〈그림 3〉 제안 모델의 구조

리즘이다(Bahdanau et al., 2014). Ren et al.(2020)은 가짜뉴스 문제점 중 하나로 가짜뉴스가 급격하게 사라지고 나타나는 시간적 특성에 대해 언급하며, 특정 시간이 지나면 특정 가짜뉴스에 대한 효능이 떨어진다는 점을 한계점으로 지적하였다. Han et al.(2020)은 시간적 특성에 따른 부족한 데이터셋에 대해 극복해보고자 데이터를 이중그래프화 한 후 적대적 신경망 기술을 적용하였으며 우수한 결과를 얻었다. 박건수 등(2021)은 가짜뉴스를 탐지하는 데 있어 언어적 특성에 의존하는 점을 문제점으로 제시하여 사용자 인증 여부, 사용자 타임라인, 팔로워 수, 친구 수 등과 사용자 고유 특성을 토대로 진행한 연구이다. Shim et al.(2021)은 텍스트 정보와 가짜뉴스가 유통되었던 플랫폼 링크를 파생변수로 가공하여 혼합한 특성을 활용하 가짜뉴스를 탐지하였다.

3. 제안 모델

본 연구에서 제안하는 모델은 아래 <그림 3>

에 제시된 바와 같이 크게 ‘그래프 생성 및 자질 추출’, ‘텍스트 및 그래프 임베딩’ 그리고 ‘모델 입력 및 예측’의 3가지 단계에 걸쳐 작동된다. 각 단계에 대한 상세 설명은 아래와 같다.

3.1. 그래프 생성 및 자질 추출

본 단계에서는 먼저 사회적 참여 그래프를 구축하고자 소셜 네트워크 서비스인 트위터에서 뉴스에 대해 전파 받은 사용자의 게시 그리고 해당 게시물에 댓글을 통해 참여한 사용자 각각을 노드로 간주하였다. 이 때 해당 게시물 댓글 여부를 통해 형성된 관계를 엣지로 간주하여 하나의 그래프로 구축하였으며, 본 연구에서는 이를 사회적 참여 네트워크라 명명하였다.

3.2. 임베딩

본 단계에서는 뉴스 제목에 대한 정보와 사회적 참여 네트워크를 임베딩 해 주게 된다.

3.2.1. 텍스트 임베딩

본 연구에서는 뉴스 및 기관에서 제시한 뉴스의 제목을 DistilRoberta(Liu et al., 2019)를 활용하여 임베딩 하였다. DistilRoberta(A distilled version of BERT)은 텍스트 임베딩 기법인 BERT에 경량화 기법 중 하나인 지식증류(knowledge distillation) 기법을 활용하여 파라미터(parameter)의 수를 압축 및 모델 최적화를 바탕으로 경량화한 모델이다(Hinton et al., 2015). 이 때 지식 증류 기법이란, 미리 잘 학습된 teacher network의 지식을 실제로 사용하고자 하는 작은 네트워크와 상호교류하며 지식전달을 하는 방법을 의미한다. 지식증류 기법을 활용하고자 한 기본 모델로는 Roberta(Liu et al., 2019)를 활용하였는데 Roberta는 BERT 대비 다이내믹 마스크(dynamic masking), 그래디언트(gradient) 증첩과 같은 기술들을 이용하여 성능 향상을 보였다.

3.2.2. 그래프 임베딩

본 연구에서는 사회적 참여 그래프를 임베딩 하기 위해 Graph2vec 방법을 활용하였다. Graph2vec은 그래프를 저차원 공간에 임베딩 하기 위해 고안된 방법론이다(Narayanan et al., 2017; 이도경 등, 2019). Graph2vec은 Doc2vec의 개념을 차용하여 그래프를 임베딩한다. Doc2vec은 문서를 벡터화하기 위한 알고리즘으로서 주어진 ‘문서’ 그리고 특정 단어를 예측하고자 전체 문서 혹은 단어를 학습하여 도출된 맥락을 토대로 특정 문서 혹은 단어들어 어떤 문서 혹은 단어일지에 대해 추론하는 방법론이다(이윤주 등, 2020). 이와 유사하게 Graph2vec은 각각의 그래프의 각각의 노드들로 연결되어 파생된 서브 그래프(subgraph)를 학습한다. 이때 서브 그래프

는 특정 노드와 연결된 노드들의 연결 집합체를 의미하며 그 과정은 다음과 같다.

우선, 서브 그래프를 형성을 해 준다. 이 때 서브 그래프라 함은 그래프 내부에서 어떤 그래프의 꼭지점과 변 가운데 일부로 이루어진 그래프를 의미한다. 본 논문에서의 서브 그래프 구축은 그래프 내부에 각각 노드들끼리 연결된 그래프를 하나의 집합으로 간주하여 진행된다. 두 번째, 서브그래프를 사용하여 Doc2vec알고리즘으로 학습시킨다. 세 번째, 학습된 Doc2vec 모델에서의 은닉층에 포함된 정보인 가중치를 활용하여 새로 들어온 그래프의 값을 추론한다.

3.3. 모델 입력 및 예측

앞선 과정에서 도출된 텍스트 특성과 네트워크 특성을 분류기에 적용하는 과정이다. 이 때 텍스트 특성과 네트워크 특성은 각각 128 차원이며, 활용한 분류기는 다음과 같다.

3.3.1. LinearSVC

분류를 목적으로 선을 정의하여 지도학습에서 활용되는 모델인 SVM(support vector machine)에 속한 모델로서, SVM 파라미터 중 하나인 커널(kernel)을 선형으로 지정해 준 알고리즘이다(Joachims, 1998).

3.3.2. Passive Aggressive Classifier

Crammer et al.(2006) 에 의해 제안된 방법론으로서 다양한 예측 문제를 해결하기 위해 새로운 데이터를 빠르게 업데이트 해주는 방법인 온라인 러닝(online learning) 알고리즘에 기반하여 데이터 학습 시 발생하는 누적 손실에 대해 최적

의 마진을 제시하는 알고리즘이다.

3.3.3. Perceptron

Rosenblatt(1958)에 의해 제안된 방법론으로서 실제 뇌를 구성하는 신경 세포 뉴런의 작동 방식을 모방하는 알고리즘이다. 이는 다수의 입력으로부터 하나의 결과를 내보내는 형식으로 예측을 수행한다.

3.3.4. AdaBoost Classifier

Rätsch et al.(2001)에 의해 제안된 방법론으로서 데이터를 분류기에 적합시키며 발생하는 잘못 분류된 값에 가중치를 두어 후속 분류기가 학습을 거듭할수록 어려운 사례에 집중할 수 있도록 설계된 알고리즘이다.

3.3.5. K-Neighbor Classifier

데이터를 n 차원의 공간에 점으로 개념화하여 유사한 특성을 가진 데이터들끼리는 거리가 가깝다는 가정을 토대로 지속적으로 분류기 학습을 하며, 비슷한 데이터들끼리의 응집으로 추론을 하는 알고리즘이다(Goldberger et al., 2004).

3.3.6. Random Forest Classifier

랜덤 포레스트는 다양한 분류기들의 결과를 집계해 추론하는 앙상블 학습 방법의 일종이다. 훈련 과정에서 다수의 의사결정 나무를 만들고, 그 나무들의 분류 또는 평균 예측치를 집계하여 최종적으로 결과를 추정하는 알고리즘이다(Pal, 2005).

4. 실증분석

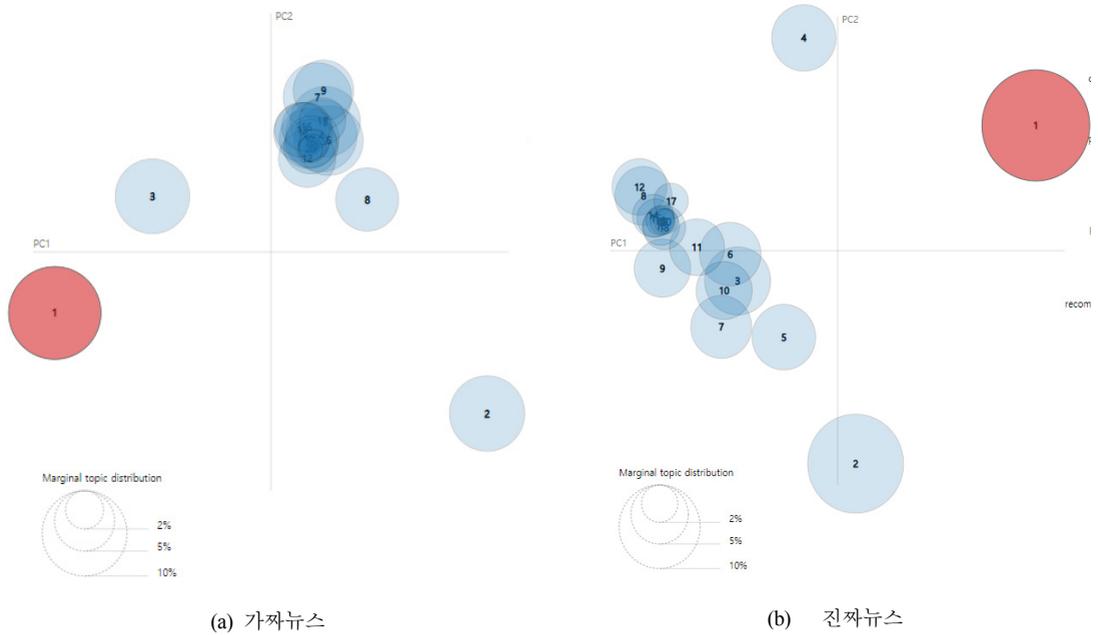
4.1. 연구 데이터

본 연구의 목적인 텍스트의 특성과 뉴스 공유를 통해 형성된 사회적 참여 네트워크 특성의 결합이 가짜뉴스 탐지에 도움이 되는가를 확인하고자 코로나 가짜뉴스 탐지만을 위해 수집된 COAID-19(Cui et al., 2020) 데이터셋을 사용하였다. 본 데이터셋은 보건전문기관(who, who.twitter, MNT)에서의 특정 뉴스에 대해 진짜 / 가짜에 대해 주장한 데이터도 포함되어 있어 신빙성을 더 한다는 장점이 있다. 또한 소셜 미디어에서 공유된 데이터도 포함되어 있어 기존의 뉴스에만 집중된 데이터셋과 다르게 소셜 네트워크 관점에서의 트위터 내의 사용자 간의 사회적 참여 관계도 특징으로 활용할 수 있다는 장점을 갖고 있다.

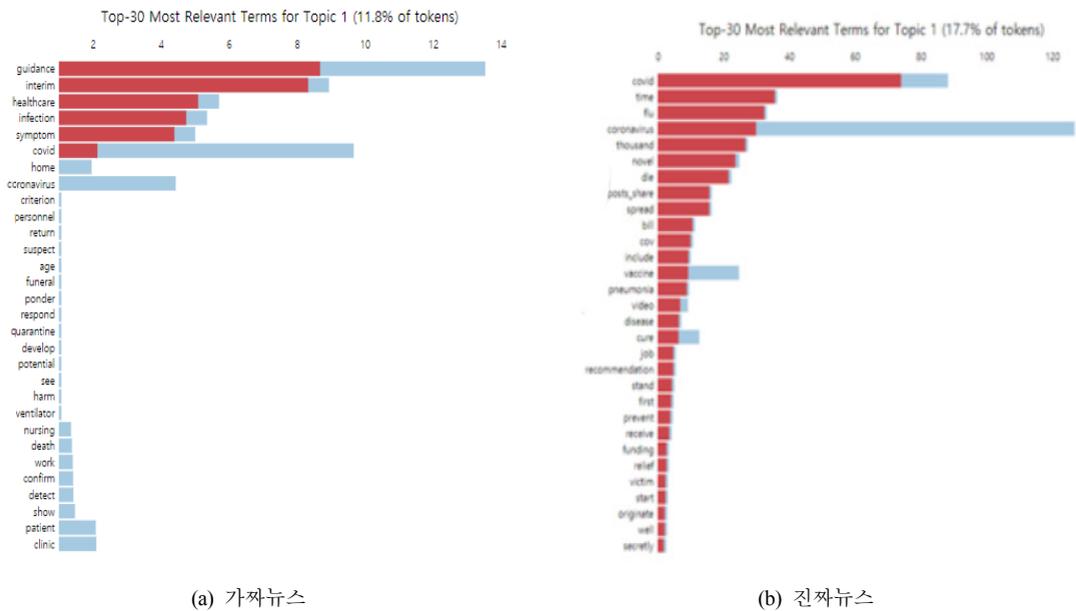
4.1.1. 텍스트

가짜뉴스 그리고 진짜뉴스의 특성 중 하나인 텍스트에 대한 데이터를 관찰해보고자 토픽 모델링의 알고리즘 중 하나인 잠재 디리클레 할당(Latent Dirichlet Allocation, LDA) 방법론을 활용하였다. 이 때 잠재 디리클레 할당 방법론은 문서들이 여러 토픽 즉 단어들의 반복 및 혼합으로 이루어져 있으며 그 결과 토픽들은 확률 분포에 기반하여 단어들을 생성한다고 가정한다. 앞선 특성을 토대로 데이터가 주어지면, 잠재 디리클레 할당은 문서가 생성되던 과정을 역추적하는 식으로 토픽을 찾아낸다.

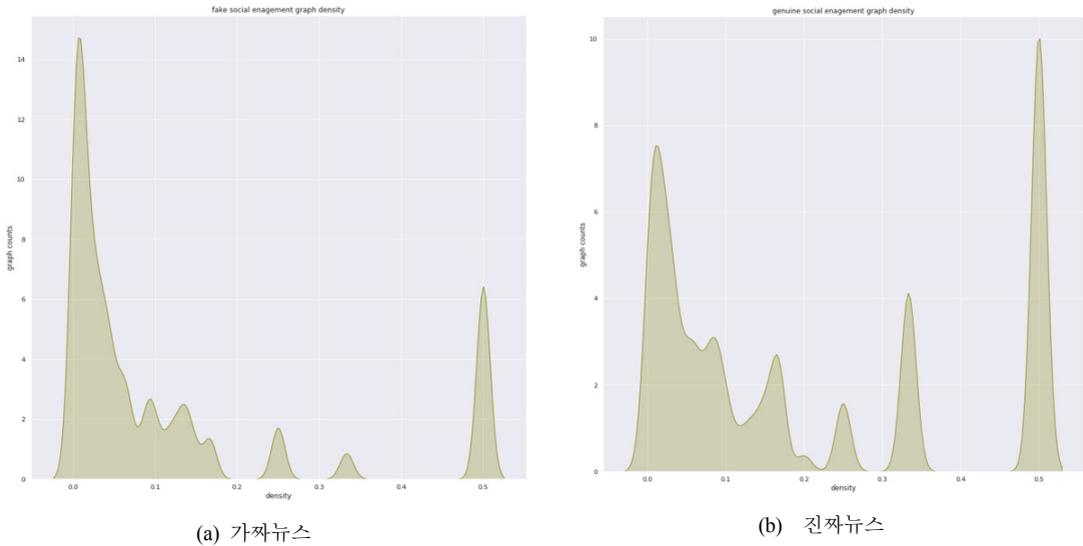
<그림 4>에서 좌·우측의 원들은 각각 가짜뉴스, 진짜뉴스 각각의 20개 토픽을 나타낸다. 이 때, 원과의 거리는 각 토픽이 서로 얼마나 다른



〈그림 4〉 벡터공간 상에서의 텍스트 분포



〈그림 5〉 특정 토픽 내에서의 단어 분포도



<그림 6> 사회적 참여 네트워크 밀도 분포

지를 보여준다. 만약 두 개의 원이 겹친다면 이는 특정 공간상에서 두 원이 유사하다는 것을 의미하며 이는 곧 두 개의 토픽은 유사한 토픽이라는 것을 의미한다. 위의 그림에서 좌측의 가짜뉴스는 특정 원들의 분포도가 고루 되어있긴 하나 상대적으로 오른쪽 진짜 뉴스와 대비해 다수의 원이 모여 있어 밀집도가 높음을 보인다. 이는 곧 가짜뉴스에서 발생하는 토픽 즉, 단어들의 범위가 좁다는 것을 의미하며 특정 단어들이 자주 활용된다는 것으로 보인다.

한편 가짜뉴스 그리고 진짜 뉴스 각각에서 발생한 토픽에서의 단어 발생 분포는 다음의 <그림 5>에 제시되어 있다. 가짜뉴스에서 직접적으로 covid를 언급한 분포는 상대적으로 낮은 편에 속하였으며 guidance, symptom, infection 와 같이 대중이 상대적으로 관심이 있을 만한 추세 전염 그리고 증상에 대해 다룬 것을 관찰할 수 있었다. 대조적으로 진짜 뉴스는 covid, time, flu, coronavirus와 같이 직접적으로 covid에 언급하는

것을 볼 수 있었다.

4.1.2. 그래프

네트워크의 밀도(density)는 노드 간의 전반적 연결 정도의 수준을 의미하는 지표이다. 허나 연결 정도 즉, 연결의 개수만을 측정하면 네트워크의 크기가 누락되기에 네트워크에 존재하는 엣지의 개수를 모든 노드끼리 전부 연결되어 있다고 가정한 상태에서 엣지 수로 나누어 이를 보완하였다. <그림 6>에서 우측 진짜 뉴스의 밀도에 비해 좌측 가짜뉴스의 밀도가 낮은 수치에 속해 있는 것의 분포가 많음을 관찰할 수 있다. 이는 곧 가짜뉴스의 연결 정도 즉, 정보 전달에 대해 연결함으로써 발생하는 커뮤니티가 열은 것이라 해석할 수 있다.

제안 방법론의 성능을 확인하기 위해 그래프 실험에 활용한 데이터 통계는 아래 <표 2>와 같다.

〈표 2〉 그래프 데이터 통계

	Genuine news	Fake news
Candidate news	339	113
Tweet user	3,034	1,696
Reply user	13,637	7,959
Social Links	14,531	8,079

〈표 3〉 실험 결과

Model	Accuracy			F1 Score		
	Text	Graph	Unified	Text	Graph	Unified
Linear SVC	0.87	0.74	0.91	0.87	0.74	0.91
Passive Aggressive Classifier	0.88	0.71	0.91	0.88	0.72	0.91
AdaBoost Classifier	0.86	0.76	0.90	0.85	0.76	0.90
K-NN Classifier	0.81	0.73	0.91	0.82	0.68	0.91
Random Forest Classifier	0.88	0.81	0.90	0.87	0.77	0.89
Perceptron	0.89	0.76	0.90	0.88	0.75	0.89

4.2. 실험평가 지표

본 연구의 목적은 이진 분류(binary classification) 과제이며 이에 따라 최종 성능 확인을 위한 성능 지표로 정확도(accuracy)와 F1(F-measure)을 선정하였다. 적용하고자 하는 데이터셋인 COAID-19의 구성이 불균형하기에 정확도 뿐 아니라 F1을 적용하여 엄밀하게 모델 간의 성능을 확인하여야 했다. 이때, F1은 정밀도(precision), 재현율(recall)의 두 가지 지표에 대한 조화평균을 토대로 도출된 평가 척도이다. 예측 모형 평가 척도에 대한 계산식은 아래와 같다.

$$Accuracy = \frac{TruePositives + TrueNegatives}{TruePositives + TrueNegatives + FalsePositives + FalseNegatives} \quad (1)$$

$$F1 = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \quad (2)$$

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives} \quad (3)$$

$$Recall = \frac{TruePositives}{TruePositives + FalseNegatives} \quad (4)$$

상기 산식에서 True Positives는 분류기가 참을 참으로 분류한 건수이고, True Negatives는 분류기가 거짓을 거짓으로 분류한 건수이다. False Positives는 분류기가 참을 거짓으로 분류한 경우이며 False Negatives는 분류기가 거짓을 참으로 분류한 경우이다.

4.3. 모델 성능비교

다음의 <표 3>에 제시된 바와 같이, 본 연구에서 제안한 모델(Unified)은 어떤 분류기를 사용하

든 항상 가장 우수한 정확도와 F1값을 제공함을 확인할 수 있었다. 특히 전반적인 성능으로 봤을 때 Linear SVC, Passive Aggressive Classifier, K-NN Classifier를 사용했을 때 F1값이 0.91로 가장 우수하게 산출되는 것을 확인할 수 있으며, 그래프를 단독으로 사용했을 때는 정확도와 F1이 0.7~0.8 사이의 상대적으로 낮은 값을 보이는 반면, 그래프를 추가로 사용했을 때 정확도와 F1이 눈에 띄게 개선되는 것으로 나타나, 그래프 정보는 그 자체로서 독립된 정보원으로서 더 큰 의미가 있음을 알 수 있다.

5. 결론

본 연구는 인터넷 기술이 발달함에 따라 증가하는 정보 그리고 그 정보에 대해 교류하는 플랫폼인 소셜 네트워크에서의 가짜뉴스 탐지에 대해 효율적인 방안을 제시하는 것을 목적으로 하였다. 가짜뉴스 본연의 특성인 텍스트 특성에 의존한 기존 가짜뉴스 탐지 기법을 확장하여 소셜 네트워크에서 발생하는 관계들을 네트워크화 하여 추출한 그래프 특성을 추가로 접목하여 뉴스 내에서 발생한 사용자 간의 참여인 그래프 특성이 가짜뉴스를 탐지하는데 있어서 효과적으로 활용됨을 확인하였다. 다시 말해 기존 주류 가짜뉴스 탐지 연구들에서는 가짜뉴스의 고유한 내부 특성인 텍스트 즉, 유사 문장 및 단어만을 토대로 뉴스의 진위에 대해 파악하는 것에 있어 해당 정보에 대한 전과 관계에 대해 추적이 불가능하며 문맥에 의존한다는 한계점이 존재하였는데, 본 연구에서는 네트워크 관점을 보완하여 특정 뉴스에 관심있는 사람들 간의 정보 교류에 대

한 특성도 추가함으로써 보다 정확한 가짜뉴스 탐지가 가능해 짐을 실험을 통해 확인하였다.

본 연구의 학술적, 실무적 기여는 다음과 같다. 우선 학술적 측면에서 본 연구는 그간 가짜뉴스 탐지 연구에 있어 크게 주목받지 못했던 배경정보(context) 중 하나인 소셜 네트워크에서의 참여 네트워크 정보가 매우 유용하게 활용될 수 있음을 제시하고, 네트워크 임베딩을 추가로 활용했을 때 텍스트 분석만 활용한 접근법보다 가짜뉴스 탐지에 더 효과적일 수 있다는 점을 실증 분석을 통해 확인했다는 점에서 의의가 있다. 본 연구와 유사하게 가짜뉴스 탐지를 목적으로 네트워크 임베딩 방법과 텍스트 분석을 동시에 적용해 보고 그 성과를 비교했던 박성수와 이건창(2019)의 연구에서는 네트워크 임베딩을 결합했을 때 텍스트 분석만 단독으로 사용하는 것에 비해 오히려 성능이 떨어지는 것으로 확인되었지만, 본 연구에서는 그 반대의 결과가 실증분석에서 확인되었다. 물론 모델의 성능은 적용된 데이터셋에 따라 달라질 수 있겠지만, 그럼에도 불구하고 경우에 따라 네트워크 임베딩 접근법이 가짜뉴스 탐지에 유의미한 접근이 될 수 있음을 확인한 점은 본 연구의 주요한 학술적 의의가 될 수 있을 것으로 기대된다.

두 번째, 실증분석을 최근 사회적으로 이슈가 되고 있는 ‘코로나19’ 관련 가짜뉴스를 대상으로 수행했다는 점에서 의의가 있다. 물론 연구에서 제안하고 있는 모델은 모든 유형의 가짜뉴스 판별에 적용될 수 있는 범용성 있는 모델이지만, 본 연구에서는 최근 우리 사회에 가장 큰 피해를 야기하고 있는 ‘코로나19’ 관련 가짜뉴스 탐지를 대상으로 실증분석을 수행하여 그 효과성을 확인하였다. 코로나19 가짜뉴스의 경우 코로나19에 대한 공포감을 갖고 있는 일반 대중들을 통해

SNS상에서 빠르게 확산되는 특징이 있기 때문에 본 연구에서 제안하는 사회적 참여 네트워크의 이용 여부를 추가로 고려하는 접근법이 특별히 더 유효한 성능을 발휘했을 것으로 예상된다.

실무적인 측면에서는 단방향성 정보 교류가 주류였던 이전과 다르게 쌍방향성 정보 교류의 장소인 소셜 네트워크에서 발생한 데이터를 그래프 임베딩 방법을 통해 활용할 수 있는 방법을 제안하였다는 점에서 실무적 가치가 높은 연구라고 판단된다. 지금까지 가짜뉴스를 탐지하고자 한 주류 연구들은 뉴스의 콘텐츠(텍스트)만을 주로 대상으로 연구하였고, 그 뉴스를 둘러싼 배경정보(context)에 대해서는 다소 등한시해 온 측면이 있었다. 하지만 본 연구는 가짜뉴스의 여러 배경정보 중 하나인 소셜 네트워크에서의 정보 전달에 초점을 두어 가짜뉴스를 판별하였을 때 보다 효과적으로 가짜뉴스에 대한 효과적인 선제적 대응이 가능함을 시사하고 있다. 따라서 자동화된 지능형 가짜뉴스 탐지를 고려하고 있는 관계 당국이나 소셜 미디어 서비스 제공자들은 뉴스의 콘텐츠뿐 아니라, 이러한 배경정보 역시 그들의 탐지 모델에서 활용될 수 있도록 적극 수용함으로써 성능 개선을 도모할 필요가 있다.

이상 여러 의의가 있지만, 본 연구가 가진 한계점도 있다. 우선 첫째, 게시자에 대한 정보가 존재하였으나 적용하지 않았다. 게시자의 인구통계학적 근거 그리고 주변 상황들로 이루어진 데이터를 추가로 사용한다면, 더 높은 정확도를 달성할 수 있을 것으로 예상된다. 둘째, 진짜뉴스에 대비하여 가짜뉴스의 빈도수가 부족하여 불균형적인 데이터 한계가 존재했다. 이를 극복하기 위해 향후 연구에서는 데이터 증강 방법을 활용하는 시도를 진행해 볼 예정이다. 셋째, 그래프 임베딩 알고리즘이 본 연구에서 적용한

Graph2vec 외에 다양하게 존재하는데, 본 연구에서는 Graph2vec, 즉 얇은 임베딩을 활용하여 그래프 구조적 특성만을 반영하였다. 향후 네트워크 내의 사용자들 간의 고유 특징들을 결합할 수 있는 방법론인 그래프 신경망을 적용하였다면 더욱 효과적인 가짜뉴스 탐지가 가능할 것으로 예상된다.

참고문헌(References)

[국내 문헌]

- 고정민. (2021). "코로나보다 백신이 더 위험하다?"...끝나지 않는 '가짜뉴스'전쟁, 청년의 사. Retrieved February 21, 2022, from <https://www.docdocdoc.co.kr/news/articleView.html?idxno=2013957>
- 권오성. (2021). 좋아요의 함정...가짜뉴스 원하는 SNS, 한겨레. Retrieved February 21, 2022, from https://www.hani.co.kr/arti/science/science_general/785227.html
- 김선호, 김위근. (2019). 유튜브의 대약진 <Digital News Report 2019>, 한국언론진흥재단. Retrieved February 21, 2022, from https://www.kpf.or.kr/front/board/boardContentsView.do?board_id=246&contents_id=000344023CF07421F30ED65B47114EE4
- 김종광, (2020). 경찰, 신종 코로나 바이러스 가짜뉴스 엄정 대응, 제주일보. Retrieved February 21, 2022, from <https://www.jejunews.com/news/articleView.html?idxno=2156143>
- 박성수, 이진창. (2019). 효과적인 가짜 뉴스 탐지를 위한 텍스트 분석과 네트워크 임베딩 방법의 비교 연구. *디지털융복합연구*, 17(5), 137-143.
- 심재승, 원하람, 안현철. (2019). A Study on the

- Effect of the Document Summarization Technique on the Fake News Detection Model, *지능정보연구*, 25(3), 201-220.
- 염정윤, 정세훈. (2019). 가짜뉴스 노출과 전파에 영향을 미치는 요인. *한국언론학보*, 63(1), 7-45.
- 윤영석, 엄태원, 안재영, 이현우, 허재두. (2017). 페이스북 뉴스 탐지 기술 동향과 시사점. ICT 신기술 주간기술동향, 정보통신기술진흥센터, 13.
- 윤태욱, 안현철. (2018). 텍스트 마이닝과 기계 학습을 이용한 국내 가짜뉴스 예측. *Journal of Information Technology Applications & Management*, 25(1), 19-32.
- 이도경, 김민태, 김우주. (2019). 의존 구문 분석을 이용한 질의 기반 정답 추출. *지능정보연구*, 25(3), 161-177.
- 이영호. (2021). *한국 SNS 이용률 세계 2위... 10~30대 인스타그램, 40~50대 밴드*, 한국경제TV. Retrieved February 21, 2022, from <https://www.wowtv.co.kr/NewsCenter/News/Read?articleId=A202106160022>
- 이완수. (2018). 가짜뉴스 (fake news)란 무엇인가? *미디어와 인격권*, 4(2), 173-214.
- 이윤주, 원하람, 심재승, 안현철. (2020) A Hybrid Collaborative Filtering-based Product Recommender System using Search Keywords, *지능정보연구*, 26(1), 151-166.
- 조승한. (2020). 코로나19 가짜뉴스로 1~3월 사이 800명 숨졌다, 동아사이언스. Retrieved February 21, 2022, from <http://dongascience.donga.com/news.php?idx=39006>
- 차미영. (2020a). *코로나바이러스와 인포테크*, 기초과학연구원(IBS). Retrieved February 21, 2022, from https://www.ibs.re.kr/cop/bbs/BBSMSTR_000000000971/selectBoardArticle.do?nttId=18234&pageIndex=2
- 차미영. (2020b). 코로나19 가짜뉴스에 맞선 데이터 과학, 기초과학연구원(IBS). Retrieved February 21, 2022, from https://www.ibs.re.kr/cop/bbs/BBSMSTR_000000000971/selectBoardArticle.do?nttId=18985&pageIndex=1&searchCnd=&searchWrd=
- 현윤진, 김남규. (2018). 뉴스와 소셜 데이터를 활용한 텍스트 기반 가짜 뉴스 탐지 방법론. *한국전자거래학회지*, 23(4), 19-39.
- 황송민. (2020). ‘신종 코로나’ 가짜뉴스 확산... “믿지 마세요”, 농민신문. Retrieved February 21, 2022, from https://www.nongmin.com/news/NEWS/FLD/CNT/319737/view?site_preference=normal
- 황용석, 권오성. (2017). 가짜뉴스의 개념화와 규제수단에 관한 연구: 인터넷서비스사업자의 자율규제를 중심으로. *언론과법*, 16(1), 53-101.

[국의 문헌]

- Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.
- Bondielli, A., & Marcelloni, F. (2019). A survey on fake news and rumour detection techniques. *Information Sciences*, 497, 38-55.
- Bottou, L. (2012). Stochastic gradient descent tricks. In *Neural networks: Tricks of the trade* (pp. 421-436). Springer, Berlin, Heidelberg.
- Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785-794).

- Crammer, K., Dekel, O., Keshet, J., Shalev-Shwartz, S., & Singer, Y. (2006). Online passive aggressive algorithms. *Journal of Machine Learning Research*, 7, 551-585.
- Cui, L., & Lee, D. (2020). Coaid: Covid-19 healthcare misinformation dataset. arXiv preprint arXiv:2006.00885.
- Dai, E., Sun, Y., & Wang, S. (2020). Ginger cannot cure cancer: Battling fake health news with a comprehensive data repository. In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 14, pp. 853-862).
- Goldberger, J., Hinton, G. E., Roweis, S., & Salakhutdinov, R. R. (2004). Neighbourhood components analysis. *Advances in Neural Information Processing Systems*, 17.
- Grover, A., & Leskovec, J. (2016). Node2vec: Scalable feature learning for networks. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 855-864).
- Hinton, G. E. (1990). Connectionist learning procedures. In *Machine learning* (pp. 555-610). Morgan Kaufmann. <https://doi.org/10.1016/B978-0-08-051055-2.50029-8>
- Hamilton, W., Ying, Z., & Leskovec, J. (2017). Inductive representation learning on large graphs. *Advances in Neural Information Processing Systems*, 30.
- Hinton, G., Vinyals, O., & Dean, J. (2015). Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531, 2(7).
- Han, Y., Karunasekera, S., & Leckie, C. (2020). Graph neural networks with continual learning for fake news detection from social media. arXiv preprint arXiv:2007.03316.
- Joachims, T. (1998). *Making large-scale SVM learning practical* (No. 1998, 28). Technical report.
- Kipf, T. N., & Welling, M. (2016). Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907.
- Le, Q., & Mikolov, T. (2014). 6Distributed representations of sentences and documents. In *International conference on machine learning* (pp. 1188-1196). PMLR.
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692.
- Mahid, Z. I., Manickam, S., & Karuppayah, S. (2018, October). Fake news on social media: brief review on detection techniques. In *2018 Fourth International Conference on Advances in Computing, Communication & Automation (ICACCA)* (pp. 1-5). IEEE.
- Narayanan, A., Chandramohan, M., Venkatesan, R., Chen, L., Liu, Y., & Jaiswal, S. (2017). Graph2vec: Learning distributed representations of graphs. arXiv preprint arXiv:1707.05005.
- Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, 26(1), 217-222.
- Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature*, 592(7855), 590-595.
- Perozzi, B., Al-Rfou, R., & Skiena, S. (2014, August). Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp.

- 701-710).
- Pierri, F., & Ceri, S. (2019). False news on social media: a data-driven survey. *ACM Sigmod Record*, 48(2), 18-27.
- Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, 1(1), 81-106.
- Rätsch, G., Onoda, T., & Müller, K. R. (2001). Soft margins for AdaBoost. *Machine learning*, 42(3), 287-320.
- Ren, Y., & Zhang, J. (2020). HGAT: hierarchical graph attention network for fake news detection. *arXiv preprint arXiv:2002.04397*.
- Ren, Y., Wang, B., Zhang, J., & Chang, Y. (2020). Adversarial active learning based heterogeneous graph neural network for fake news detection. In *2020 IEEE International Conference on Data Mining (ICDM)* (pp. 452-461). IEEE.
- Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), 386.
- Shejkh, K. (2021). *How Bad will the Coronavirus Outbreak Get? Here Are 6 Key Factors*. Retrieved February 21, 2022, from <https://www.nytimes.com/interactive/2020/world/asia/china-coronavirus-contain.html?action=click&module=RelatedLinks&pgtype=Article> (Accessed 2021.10.13)
- Shim, J. S., Lee, Y., & Ahn, H. (2021). A link2vec-based fake news detection model using web search results. *Expert Systems with Applications*, 184, 115491.
- Shu, K., Mahudeswaran, D., Wang, S., Lee, D., & Liu, H. (2018). Fakenewsnet: A data repository with news content, social context and spatiotemporal information for studying fake news on social media. *arXiv preprint arXiv:1809.01286*.
- Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter*, 19(1), 22-36.
- Shu, K., Wang, S., & Liu, H. (2019). Beyond news contents: The role of social context for fake news detection. In *Proceedings of the twelfth ACM international conference on web search and data mining* (pp. 312-320).
- Tang, J., Qu, M., Wang, M., Zhang, M., Yan, J., & Mei, Q. (2015). Line: Large-scale information network embedding. In *Proceedings of the 24th International Conference on World Wide Web* (pp. 1067-1077).
- Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., & Bengio, Y. (2017). Graph attention networks. *arXiv preprint arXiv:1710.10903*.
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146-1151.
- Wang, W. Y. (2017). "liar, liar pants on fire": A new benchmark dataset for fake news detection. *arXiv preprint arXiv:1705.00648*.
- Wang, Y., Ma, F., Jin, Z., Yuan, Y., Xun, G., Jha, K., ... & Gao, J. (2018). EANN: Event adversarial neural networks for multi-modal fake news detection. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 849-857).

Abstract

A study on the detection of fake news - The Comparison of detection performance according to the use of social engagement networks

Iitae Jeong* · Hyunchul Ahn**

With the development of Internet and mobile technology and the spread of social media, a large amount of information is being generated and distributed online. Some of them are useful information for the public, but others are misleading information. The misleading information, so-called 'fake news', has been causing great harm to our society in recent years. Since the global spread of COVID-19 in 2020, much of fake news has been distributed online. Unlike other fake news, fake news related to COVID-19 can threaten people's health and even their lives. Therefore, intelligent technology that automatically detects and prevents fake news related to COVID-19 is a meaningful research topic to improve social health. Fake news related to COVID-19 has spread rapidly through social media, however, there have been few studies in Korea that proposed intelligent fake news detection using the information about how the fake news spreads through social media. Under this background, we propose a novel model that uses Graph2vec, one of the graph embedding methods, to effectively detect fake news related to COVID-19. The mainstream approaches of fake news detection have focused on news content, i.e., characteristics of the text, but the proposed model in this study can exploit information transmission relationships in social engagement networks when detecting fake news related to COVID-19. Experiments using a real-world data set have shown that our proposed model outperforms traditional models from the perspectives of prediction accuracy.

Key Words : Fake news detection, Graph embedding, Social engagement network, Graph2vec, COVID-19

Received : November 26, 2021 Revised : February 26, 2022 Accepted : March 8, 2022

Corresponding Author : Hyunchul Ahn

* Graduate School of Business IT, Kookmin University

** Corresponding author: Hyunchul Ahn

Graduate School of Business IT, Kookmin University

77 Jeongneung-ro, Seongbuk-gu, Seoul 02707, Korea

Tel: +82-2-910-4577, Fax: +82-2-910-4017, E-mail: hcahn@kookmin.ac.kr

저 자 소개



정이태

국민대학교 비즈니스IT전문대학원에서 공학석사 학위를 취득하였다. 주요 관심분야는 그래프 임베딩, 소셜네트워크분석, 추천시스템 등이다.



안현철

현재 국민대학교 비즈니스IT전문대학원 교수로 재직 중이다. KAIST에서 산업경영학사를 취득하고, KAIST 테크노경영대학원에서 경영정보시스템을 전공하여 공학석사와 박사학위를 취득하였다. 주요 관심 분야는 금융 및 고객관계관리 분야의 인공지능 응용, 지능형 의사결정지원시스템, 정보시스템 수용과 관련한 행동 모형 등이다.