

효과적인 가짜 뉴스 탐지를 위한 텍스트 분석과 네트워크 임베딩 방법의 비교 연구

박성수¹, 이건창^{2*}

¹성균관대학교 경영대학 박사

²성균관대학교 글로벌경영학과/삼성융합의과학원 융합의과학과 교수

A Comparative Study of Text analysis and Network embedding Methods for Effective Fake News Detection

Sung Soo Park¹, Kun Chang Lee^{2*}

¹PhD, SKK Business School, Sungkyunkwan University

²Professor, Global Business Administration/Dept of Health Sciences & Technology, SAIHST(Samsung Advanced Institute for Health Science & Technology), Sungkyunkwan University

요약 가짜 뉴스는 소셜 미디어와 같이 사용자가 상호작용하는 미디어 플랫폼에서 정보가 빠른 속도로 확산되는 이점을 가지는 오류 정보(misinformation)의 한 형태이다. 최근 가짜 뉴스의 증가로 인해 사회적으로 많은 문제가 발생하고 있다. 본 논문에서는 이러한 가짜 뉴스를 탐지하는 방법을 제안한다. 이전의 가짜 뉴스 탐지는 텍스트 분석을 사용한 연구가 주로 수행되었다. 본 연구는 소셜 미디어의 뉴스가 확산되는 네트워크에 초점을 두고, 네트워크 임베딩 방법인 DeepWalk 로 자질을 생성하고 로지스틱 회귀분석을 사용하여 가짜 뉴스를 분류한다. 인터넷에 공개된 뉴스 211개와 120만개의 뉴스 확산 네트워크 데이터를 사용한 가짜 뉴스 탐지에 대한 실험을 수행하였다. 연구 결과 텍스트 분석에 비하여 네트워크 임베딩을 사용한 가짜 뉴스 탐지의 정확도가 최소 1.7%에서 최대 10.6% 더 높게 나타났다. 또한, 텍스트 분석과 네트워크 임베딩을 결합한 가짜 뉴스 탐지는 네트워크 임베딩에 비해 정확도의 상승이 나타나지 않았다. 본 연구의 결과는 기업이나 조직은 온라인 상에서 확산되는 가짜 뉴스 탐지에 효과적으로 활용될 수 있다.

주제어 : 가짜 뉴스 탐지, 텍스트 분석, 네트워크 임베딩, DeepWalk, 뉴스 확산 네트워크.

Abstract Fake news is a form of misinformation that has the advantage of rapid spreading of information on media platforms that users interact with, such as social media. There has been a lot of social problems due to the recent increase in fake news. In this paper, we propose a method to detect such false news. Previous research on fake news detection mainly focused on text analysis. This research focuses on a network where social media news spreads, generates qualities with DeepWalk, a network embedding method, and classifies fake news using logistic regression analysis. We conducted an experiment on fake news detection using 211 news on the Internet and 1.2 million news diffusion network data. The results show that the accuracy of false network detection using network embedding is 10.6% higher than that of text analysis. In addition, fake news detection, which combines text analysis and network embedding, does not show an increase in accuracy over network embedding. The results of this study can be effectively applied to the detection of fake news that organizations spread online.

Key Words : Fake news detection, Text analysis, Network embedding, DeepWalk, News diffusion network.

*Corresponding Author : Kun Chang Lee(kunchanglee@gmail.com)

Received February 21, 2019

Revised March 24, 2019

Accepted May 20, 2019

Published May 28, 2019

1. 서론

한국과학기술기획평가원(KISTEP)은 2017년 3월 20일, 인공지능 사실 확인(Fact Checking) 기술을 10대 미래유망기술로 꼽았다. 소셜 미디어 플랫폼의 발달로 인해 전통적인 뉴스 미디어보다 소셜 미디어에서 뉴스를 찾아 소비하는 경향이 증가하고 있다. 예를 들어 2012년에는 미국 성인의 49%만이 소셜 미디어에서 뉴스를 소비했지만, 2016년에는 62%가 소셜 미디어에서 뉴스를 제공받는 것으로 나타났다. 그러나 소셜 미디어를 통해 온라인으로 뉴스를 제공하는 것은 저렴하고 보급하기가 쉽기 때문에 많은 양의 가짜 뉴스, 즉 의도적으

로 잘못된 정보가 담긴 뉴스 기사는 금융 시장의 교란, 정치적 비방 등과 같은 다양한 목적으로 온라인에서 제작된다. 예를 들어 2016년 미국 대선이 끝날 때까지 1백만 건이 넘는 트위터의 가짜 뉴스 "Pizzagate"와 관련이 있는 것으로 추산되었다. 국내에서도 미래창조과학부가 주관하는 2017년도 인공지능 R&D 챌린지의 과제가 '가짜 뉴스 찾기'인 정도로 뜨거운 이슈가 되고 있다.

가짜 뉴스는 뉴스가 널리 유통되기 시작한 것과 거의 같은 시간 동안 매우 오랜 시간 동안 존재 해 왔다. 그러나 "가짜 뉴스"라는 용어에 대한 합의된 정의는 없다. 일반적으로 가짜 뉴스는 뉴스 내용의 진위성이나 의도에 따라 거짓 정보를 사용하여 뉴스 소비자를 오도하려는 부정직한 의도로 만들어진 뉴스를 말한다[1]. 이 정의는 최근 연구에서 광범위하게 사용되었다[2-4]. 본 연구는 가짜뉴스를 다음과 같이 정의한다. 가짜 뉴스는 "소셜 미디어와 같이 사용자가 상호작용하는 미디어 플랫폼에서 정보가 빠른 속도로 확산되는 이점을 가지는 오류 정보(misinformation)의 한 형태"이다.

가짜 뉴스의 광범위한 확산은 개인과 사회에 심각한 부정적인 영향을 미친다. 첫째, 가짜 뉴스는 뉴스 생태계의 균형을 깨뜨리게 된다. 예를 들어, 가장 인기 있는 가짜 뉴스가 2016년 미국 대통령 선거에서 가장 인기 있는 정통 주류 뉴스보다 페이스북 북에서 훨씬 더 널리 퍼져 나갔다. 둘째, 가짜 뉴스는 의도적으로 소비자들이 편향되거나 거짓된 믿음을 받아들이도록 설득한다. 가짜 뉴스는 일반적으로 정치·경제적 이득을 얻을 수 있는 선전가들에 의해 조작된다. 셋째, 가짜 뉴스는 사람들이 실제 뉴스를 해석하고 이에 반응하는 방식을 바꾼다. 예를 들어, 일부 가짜 뉴스는 사람들의 불신을 유발하고 혼란스럽게 만들어 실제와 가짜를 판별하는 판단력을 흐리게 한다. 가짜 뉴스로 인해 유발된 사용자 간의 사회적 논쟁은 크

고 불완전하며 구조가 복잡하고 시끄러운 데이터를 생성하게 된다[5].

이러한 가짜 뉴스를 빠르게 탐지하고 확산을 방지하는 것이 중요한 이슈이지만 현재까지 명확한 방법을 제시하는 연구는 찾아보기 어렵다. 가짜 뉴스 탐지가 어려운 것은 가짜 뉴스의 주제, 스타일 및 미디어 플랫폼 측면에서 다양한 방법으로 진실을 왜곡하면서 동시에 진짜 정보를 가리기 때문이다. 가짜 뉴스가 비사실적 주장을 뒷받침하기 위해 다른 맥락의 진짜 증거를 인용하는 경우, 텍스트 분석만으로 가짜 뉴스 탐지는 충분한 성과를 내기 어렵다. 따라서, 본 연구는 가짜 뉴스의 탐지에 초점을 두고, 최근 활발한 연구가 이루어지는 네트워크 임베딩 방법을 적용한 가짜뉴스 탐지 방법에 대한 다음과 같은 연구 문제(Research Question)를 다룬다.

RQ1: 가짜 뉴스가 온라인에서 확산된 네트워크를 결합한 자질을 사용하여 보다 효과적인 가짜 뉴스 탐지 방법을 제시한다.

RQ2: 가짜 뉴스를 구성하는 텍스트를 텍스트 분석 가짜 뉴스 탐지 방법과 비교하여 네트워크 임베딩을 사용한 가짜 뉴스 탐지 방법의 효과적임을 밝힌다.

RQ3: 텍스트 분석과 네트워크 임베딩을 결합한 가짜 뉴스 탐지 방법의 효과를 검증한다.

2. 관련 연구

2.1 가짜뉴스 탐지(Fake News Detection)

가짜 뉴스로 인한 부정적인 영향을 줄이기 위해 소셜 미디어에서 가짜 뉴스를 자동으로 탐지하는 방법을 개발하는 것은 중요한 이슈이다. 가짜 뉴스를 탐지하기 위해 가장 많이 사용되는 방법은 텍스트 분석 접근법이다. 이 방법은 뉴스를 구성하는 문자, 단어, 문장 및 문단에서 자질(Feature)을 추출하여 가짜 뉴스의 여부를 분류하는 것이다. 주로 연구되는 자질은 (i) 어휘 자질 - 전체 단어, 단어 당 문자, 단어 출현 빈도 및 고유 단어와 같은 단어 수준 자질 - 과, (ii) 구문론적 자질 - 문장을 구성하는 단어를 원 핫 인코딩(One-hot encoding) 형태로 나타내거나 품사 (POS) 태깅과 같은 단어의 기능적 특성 및 단어의 출현 빈도를 자질로 표현 -11; 하는 방법이 있다. 텍스트 분석을 사용한 가짜뉴스 탐지에 대한 연구는 아직 체계적이고 명확한 연구 성과를 내지 못하고 있다.

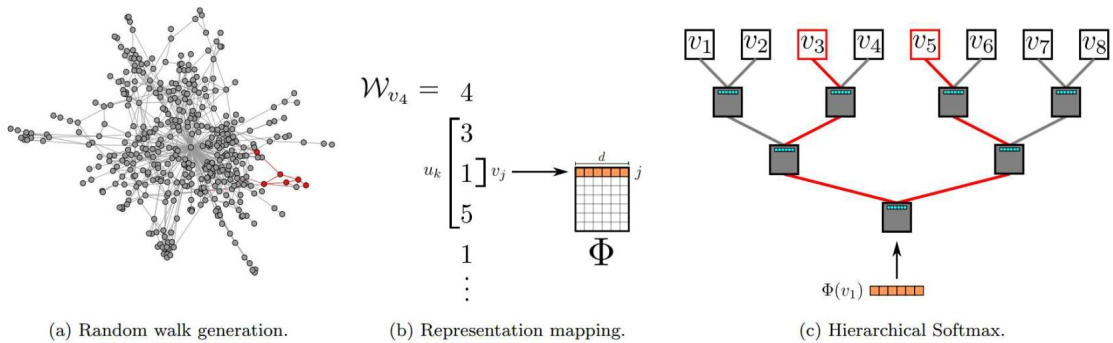


Fig. 1. DeepWalk Overview

자연 언어 처리(NLP) 및 가짜 뉴스 탐지의 최신 발전이 사기성 뉴스를 탐지하는 데 도움이 될 수 있다는 연구들은 진행 중이다[6]. 가짜 뉴스는 가짜 뉴스 제목에 명사와 동사가 많을수록 정지 단어(Stop words)와 명사가 적은 패턴을 나타내는 경향이 있다[7]. 연구자들은 텍스트의 복잡성, 감정 단어 및 인과 단어, 작자의 문체 스타일, 텍스트 구문 등을 자질로 하는 SVM기반의 가짜 뉴스 예측 모형을 제시하여 가짜 뉴스 탐지 정확도는 71%를 달성하였다. 또한, 텍스트 분석 방법 중 하나인 N-gram과 머신러닝을 결합한 가짜 뉴스 탐지에 대한 연구가 수행되었고[8], 소셜 네트워크 상의 텍스트를 분석해 사기 그룹을 탐지하는 연구 등이 수행되었다[9].

가짜 뉴스를 탐지하는 또 다른 연구 흐름인 네트워크 분석 기반의 가짜 뉴스 탐지는 소셜 미디어에 게시물을 게시한 사용자간에 특정 네트워크를 구성하여 추출된 자질을 사용하는 것이다. 사용자는 관심사, 주제 및 관계 측면에서 소셜 미디어에 서로 다른 네트워크를 형성하게 된다. 네트워크를 구성하는 방식에 따라 다양한 연구가 수행되었다. 자세 네트워크(stance network)는 뉴스와 관련된 모든 트윗을 나타내는 노드와 자세의 유사성의 가중치를 나타내는 간선으로 구성한다[10, 11]. 다른 유형의 네트워크는 사용자 참여를 기반으로 사용자가 동일한 뉴스 기사와 관련된 게시물을 작성하는지 여부를 계산하여 작성된 동시 발생 네트워크(co-occurrence network) 이다[12]. 또한 우정 네트워크(friendship network)는 관련 트윗을 게시하는 사용자 간의 following-follower 구조를 나타낸다[13]. 이 우정 네트워크의 확장된 형태로 뉴스 확산 궤도를 추적하는 확산 네트워크(diffusion network)가 있다. 여기서 노드는 사용자를 나타내고 간선은 정보 확산 경로를 나타낸다. 즉, (1) u_j 가 u_i 를 따르고, (2) u_j 는 u_i 가 게시한 뉴

스를 게시하는 경우에 u_i 와 u_j 사이의 확산 경로가 존재한다. 이러한 네트워크는 네트워크의 위상(Topology)을 사용하는 사회 연결망 분석(Social Network Analysis) 방법의 하나인 정도(degree)와 클러스터링 계수(clustering coefficient)를 사용해 가짜 뉴스를 분류하는 연구가 이루어졌다(Kwon et al. 2013). 다른 접근법은 SVD 또는 네트워크 전파 알고리즘(network propagation algorithms)을 사용하여 네트워크의 잠재 자질(latent feature)을 사용하는 가짜 뉴스 탐지에 대한 연구들이 수행 되었다[12,10].

최근까지 가짜 뉴스를 탐지하는 다양한 방법이 연구되었지만 더욱 정교하고 다양하게 변화하는 환경에 적응 가능한 가짜 뉴스 탐지 방법에 대한 연구가 필요하다. 본 연구는 네트워크를 벡터 공간에 표현하는 네트워크 임베딩을 사용한 가짜 뉴스 탐지 방법으로 효율적인 가짜 뉴스 탐지 방법을 실증한다.

2.2 네트워크 임베딩(Network Embedding)

일반적으로 네트워크를 $G = \langle V, E \rangle$ 그래프로 표현하는데, 여기서 V 는 네트워크의 노드를 나타내는 정점 세트(Vertex)이고, E 는 노드 간의 관계를 나타내는 간선(Edge)이다. 전통적인 네트워크 표현은 오늘날 대규모 네트워크 처리 및 분석에서 한계에 도달하였다. 전통적인 표현에서 간선 세트를 사용하여 명시적인 관계를 표현하는 것이 가장 큰 한계점이다.

이 문제를 해결하기 위해 네트워크 노드에 대한 저차원 벡터 표현과 같은 새로운 네트워크 임베딩 방법에 대한 다양한 연구들이 진행되었다. 전통적인 네트워크 분석은 네트워크의 위상을 분석에 바로 사용하지만, 네트워크 임베딩은 벡터 공간에 노드들 사이의 거리와 노드의 토폴로지 및 구조적 특성과 같은 노드들 간의 관계를 표현

하고 이를 응용하는 것이다.

네트워크 구조를 유지하는 것은 네트워크 임베딩을 위한 기본적인 요구 사항으로 노드의 로컬 구조 특성을 설명하는 고유 구조는 네트워크 임베딩에 중요한 요소이다. 인접 행렬(Adjacency Matrix)은 노드의 1 차 근방 구조를 부호화하지만 대용량 네트워크에서는 희소성(sparseness)의 특성으로 인해 일반적으로 희소하고 이산적인 고차원 벡터로 표현된다. 이러한 표현 방법은 이어지는 네트워크의 응용에도 상당한 문제가 발생한다. 자연어 처리(Neural Language Processing)의 단어 표현(Word Representation)은 이와 비슷한 문제점을 나타낸다. Mikolov et al.이 제안한 Word2Vec은 희소성이 높고 이산형의 고차원 벡터를 고밀도의 연속형인 저차원 벡터로 변환하여 단어 표현의 효율성을 크게 향상시켰다[14]. 다시 말해, Word2Vec은 동시 발생 비율로 정의된 인접 단어를 학습해서 단어 벡터로 재구성하는 것이다. 특히, Skip-gram은 특정 단어의 앞, 뒤에 나타나는 단어를 학습하는데 효과적인 성능을 보인다.

최근의 네트워크 임베딩은 이러한 언어 모델을 네트워크 데이터에 응용하였다. 네트워크 임베딩의 핵심적인 문제는 네트워크에서 노드간의 '이웃'을 정의하는 방법이다. Word2Vec에서 단어들이 연속적으로 출현하는 것과 유사하게, 랜덤 워크 모델로 네트워크의 무작위 경로(random path)를 생성하는 방법을 사용했다. 이 방법은 노드를 단어로 간주하고 무작위 경로로 생성된 노드의 순서는 문장으로 보는 것이다. 노드 인접 노드는 Word2Vec의 동시 출현 비율(co-occurrence rate)로 식별 할 수 있다. 이러한 방식의 대표적인 네트워크 임베딩 방법을 처음 제안한 DeepWalk[15]와 간선의 가중치 계산을 보완한 Node2Vec[16]이 대표적이다.

DeepWalk는 Fig. 1.과 같이 노드의 인접 구조를 보존 할 수 있는 네트워크의 노드 표현을 학습하기 위해 제안되었다. DeepWalk는 짧은 무작위 경로에 나타나는 노드의 분포가 자연 언어로 된 단어의 분포와 유사하다는 것에 착안하여 널리 사용되는 단어 표현 학습 모델인 Skip-Gram 모델[14]을 채택하여 네트워크 표현을 학습한다. DeepWalk는 일련의 걷기 순서(walk sequence)를 생성함에 있어 걷기 순서가 무한대로 증가하는 것을 방지하기 위해 네트워크에서 제한된 랜덤 워크(truncated random walk) 방법을 적용한다. Skip-Gram에 따라 각 걷기 순서 $s = \{v_1, v_2, \dots, v_s\}$ 에 대해 DeepWalk는 노드 v_i 의 이웃 노드의 확률을 최대화하는 것을 목표로 학습한다.

3. 연구 방법

3.1 연구 데이터

본 연구의 목적인 텍스트와 네트워크 임베딩 방법을 결합한 가짜 뉴스 탐지를 위해 소셜 네트워크 상에 전파되는 사실이 입증되지 않는 가짜 뉴스 및 무머를 수집한다. 소셜 네트워크인 트위터에서 가짜 뉴스와 관련된 키워드를 언급하는 트윗과 트위터안의 리트윗 관계를 추출한다. 가짜 뉴스와 비교하기 위해 관련된 주제의 뉴스를 전송하는 트윗과 트위터안의 데이터도 같이 수집한다. 이후 언론사들이 협업해 주요 뉴스의 사실 관계를 확인해 게시하는 팩트 체크 서비스에서 가짜 뉴스를 식별하고 뉴스에 대한 참-거짓을 라벨링 한다. 이러한 방법으로 수집된 가짜 뉴스 네트워크(Fake News Net) 데이터 셋을 사용하여 연구의 실험으로 사용하였다[17].

가짜 뉴스 네트워크 데이터는 뉴스 게시자, 뉴스 콘텐츠 및 소셜 네트워크 정보가 포함 된 두 가지의 가짜 뉴스 데이터 셋이 포함되어 있다. 두 가지 데이터 셋은 가짜와 진짜 뉴스를 구별하는 라벨링을 한 기준으로 생성되었다. BuzzFeed 데이터 셋은 저널리스트 전문가가 뉴스의 사실 여부를 확인하는 곳이고, PolitiFact는 잘 알려진 사실 확인(Fact Check) 웹 사이트이다. 각 데이터 셋에 대한 통계량은 다음 Table 1과 같다.

Table 1. Statistics for fake news detection data set

Platform	BuzzFeed	PolitiFact
Candidate news	182	240
True news	91	120
Fake news	91	120
Users	15,257	23,865
Social Links	634,750	574,744
Publisher	9	91

3.2 연구 절차

본 연구의 목적인 네트워크 임베딩 기반의 가짜 뉴스 탐지 절차는 다음과 같으며 Fig. 2.으로 나타냈다.

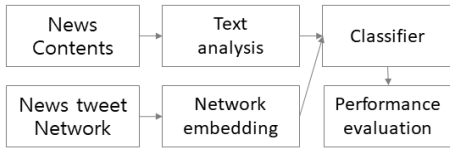


Fig. 2. Fake news detection procedures that combine text and network embedding methods

- 1단계: 뉴스가 언급된 트윗의 트윗터리안과 연관되어 있는 트윗터리안들의 네트워크 데이터를 추출한다. 즉, 뉴스와 트윗터리안을 모두 네트워크의 노드로 두고 뉴스와 뉴스를 언급한 트윗터리안과의 간선과 트윗터리안 간의 Follower, Followee관계를 간선으로 하는 뉴스 확산 네트워크(News diffusion network)를 생성한다. 네트워크 데이터는 네트워크 분석이 가능한 인접 리스트(adjacent list) 형태로 변환해 한다. 또한, 텍스트 분석을 통한 가짜 뉴스 탐지 실험을 위하여 트윗에 언급된 뉴스에 대한 뉴스 본문을 추출한다.

- 2단계: 네트워크 임베딩으로 네트워크 자질을 추출한다. 네트워크 임베딩 방법은 Perozzi가 제안한 DeepWalk를 사용한다. 각 노드 당 랜덤 워크 횟수는 10회, 워크의 길이는 80회, 생성할 벡터 차원의 수는 128차원으로 설정한다. 분석도구는 인터넷에 공개된 네트워크 임베딩 프로그램을 사용하여 연구 목적에 맞게 수정하여 사용한다. 또한, 제안한 방법과 비교를 위해 뉴스 본문의 텍스트에 대한 자질을 생성한다. 텍스트 리뷰는 파이썬을 사용하여 자주 출현하는 다빈도 단어 상위 500개를 선별한다. 이를 각 뉴스 당 출현 빈도를 원-핫 인코딩(One-hot Encoding)해서 텍스트 자질을 생성한다.

- 3단계: 생성된 자질 데이터 셋을 예측 모형에 입력하여 가짜 뉴스와 진짜 뉴스를 분류한다. 목표 변수는 각 뉴스에 대한 사실 여부이고, 입력된 자질 셋은 네트워크 자질은 가짜 뉴스를 언급한 트윗터리안과 연결된 소셜 네트워크를 네트워크 임베딩으로 벡터 표현한 자질이고, 텍스트 자질은 뉴스 본문에 대한 키워드 출현 빈도를 벡터 표현한 자질이다. 본 연구에서 사용된 예측 방법은 범주형 로지스틱 리그레션으로 모형의 적합성을 확보하기 위해 5 Fold 교차 검증법을 사용한다.

- 4단계: 각 자질 생성 방법에 따른 가짜 뉴스 탐지 성능을 비교한다. 각 자질 데이터 셋 별 예측 모형의 평가는 정밀도(Precision), 재현율(Recall)과 이 두 가지 지표를 조화평균하는 F1(F-measure), 정확도(Accuracy)를 사용한다. 예측 모형 평가 척도에 대한 계산식은 아래와 같다.

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$F1 = \frac{2 \times (Precision + Recall)}{(Precision \times Recall)} \quad (3)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

TP(True Positive)는 분류기가 참을 참으로 분류한 건수이고, FP(False Positive)는 분류기가 참을 거짓으로 분류한 건수이다. TN(True Negative)는 분류 결과가 거짓을 거짓으로 분류한 건수이고, FN (False Negative)은 거짓을 참으로 분류한 건수 이다.

4. 실험 결과

가짜 뉴스 탐지 분석 실험 결과는 Table 2와 같다. 본 연구의 벤치마크인 텍스트 분석을 통한 가짜 뉴스 탐지 결과는 PolitiFact 데이터 셋의 탐지 정확도는 정밀도, 재현율, F1, 정확도가 0.838이고, 또 다른 가짜 뉴스 데이터 셋인 BuzzFeed 데이터 셋은 0.750으로 나타났다. 본 연구에서 제시한 방법인 네트워크 임베딩을 사용한 가짜 뉴스 탐지 정확도는 PolitiFact 데이터 셋에서 정밀도, 재현율, F1, 정확도가 0.943이고, BuzzFeed 데이터 셋의 경우 0.767이다.

Table 2. Fake news detection experiment results using text analysis and graph embedding

Methods	Data Set	Precision	Recall	F1	ACC
Text Analysis	PolitiFact	0.842	0.838	0.837	0.838
	BuzzFeed	0.751	0.750	0.750	0.750
Network Embedding	PolitiFact	0.944	0.943	0.943	0.943
	BuzzFeed	0.770	0.767	0.766	0.767

각 분석 방법에 따른 정확도 차이는 PolitiFact 데이터 셋은 텍스트 분석 대비 네트워크 임베딩의 정확도가 0.106 더 높게 나타났고, BuzzFeed 데이터 셋은 0.017

상승한 것으로 나타났다.

추가적인 연구로 텍스트 분석과 네트워크 임베딩을 결합한 자질을 생성해서 가짜 뉴스 탐지 예측 모형을 생성하고 성과를 측정하였다. 텍스트 자질과 네트워크 임베딩 자질은 뉴스 아이디를 기준으로 두 자질 데이터 셋을 통합하였다. 텍스트 자질과 네트워크 임베딩 자질을 사용한 가짜 뉴스 탐지 분석 결과는 Table 3과 같다. PolitiFact 데이터 셋은 정밀도, 재현율, F1, 정확도가 0.876이고, BuzzFeed 데이터 셋의 경우 0.750을 보였다.

Table 3. Fake News Detection Analysis Results Combining Text Analysis and Graph Embedding

Data set	Precision	Recall	F1	Accuracy
PolitiFact	0.896	0.896	0.896	0.896
BuzzFeed	0.751	0.750	0.750	0.750

텍스트 자질과 네트워크 임베딩 자질을 사용한 가짜 뉴스 탐지 분석 결과를 개별 자질 데이터 셋의 결과와 비교하면 PolitiFact 데이터 셋은 텍스트 자질에 비해 정확도가 0.058 상승하였으나 네트워크 임베딩 자질에 비해 0.048 낮은 정확도를 보였다. BuzzFeed 데이터 셋은 텍스트 자질에 비해 정확도의 변화는 없고, 네트워크 임베딩 자질에 비해 정확도가 0.017 낮아졌다.

5. 논의 및 결론

본 연구는 효율적인 소셜 네트워크와 같은 온라인에서 급속도로 증가하는 가짜 뉴스에 대한 탐지 방법을 제시하는 목적한다. 이를 위해 최근 활발한 연구가 수행되는 네트워크 임베딩을 가짜 뉴스 탐지에 적용하였다. 연구 결과 네트워크의 노드간의 연결 관계와 구조를 벡터 공간에 표현하는 네트워크 임베딩 기반의 가짜 뉴스 탐지가 더 효과적인 것으로 나타났다. 두 가지 가짜 뉴스 데이터 셋에 대한 가짜 뉴스 탐지 분석 실험으로, 네트워크 임베딩을 사용한 가짜 뉴스 탐지 방법이 텍스트 분석 기반의 가짜 뉴스 탐지에 비해 최소 1.7%에서 최대 10.6%의 정확도 상승이 나타났다. 이것은 소셜 네트워크 상에서 가짜 뉴스가 일정한 네트워크 경로에 따라 확산되고 있는 것을 의미한다. 유사한 목적이나 정보에 대한 신뢰가 동일한 사람들끼리 소셜 네트워크 상에서 관계를 맺고 뉴스에 대한 정보가 유포되는 것이다. 또한, 다른 면에서 가짜 뉴스가 진짜 뉴스와 유사한 단어나 문장 구조를

가지고 작성되는 경우 텍스트 분석으로 가짜 뉴스의 탐지가 제한적인 수준을 넘기 어려운 것을 의미한다.

본 연구의 의의는 다음과 같다. 첫째, 최근 급부상하고 있는 가짜 뉴스의 탐지에 대한 연구에 네트워크 분석 기법인 네트워크 임베딩을 결합한 최초의 연구이다. 네트워크 임베딩 연구는 컴퓨터 과학 분야에서 활발하게 많은 연구가 이루어지고 있으나 연구용 데이터를 벗어나 실제 데이터의 활용과 적용에 대한 연구는 극히 적은 실정이다. 본 연구는 네트워크 임베딩을 실제 데이터에 적용하고 효율적인 가짜 뉴스 탐지가 가능함을 실증 결과로 제시하였다. 둘째, 텍스트 분석 위주의 연구가 이루어지는 가짜 뉴스 탐지 분야에 대해 네트워크 분석을 통한 새로운 방향을 제시하였다. 이전의 가짜 뉴스, 루머, 오정보와 관련된 네트워크 분석 연구들은 가짜 뉴스 탐지나 분류에 대한 연구 보다는 뉴스의 확산이나 가짜 뉴스 커뮤니티 식별에 대한 연구가 활발하게 수행되었다. 본 연구는 이전의 연구에 비해 가짜 뉴스 탐지의 관점에서 네트워크 분석을 적용하였다. 셋째, 가짜 뉴스 탐지에 텍스트 분석과 네트워크 임베딩을 결합하는 것은 바람직하지 않음을 실증했다. 사용되는 데이터와 절차가 다른 두 가지 가짜 뉴스 탐지 방법 - 텍스트 분석, 네트워크 임베딩 - 을 결합하는 것은 각 방법 간에 특징들이 융합되지 않고, 오히려 가짜 뉴스 탐지 모형의 성능의 유지나 하락이 됨을 밝혔다.

본 연구는 실무자들에게 다음과 같은 시사점을 제공한다. 첫째, 소셜 네트워크 상에서 확산되는 가짜 뉴스는 몇몇 목적을 공유하는 사용자들 중심으로 확산되는 경향이 있다. 본 연구의 결과를 사용하여 가짜 뉴스의 최초 유포자나 가짜 뉴스를 공유하는 그룹의 사용자를 밝힐 수 있다. 둘째, 기업들은 자사의 상품과 서비스에 대한 평판을 좋게 유지하기 위해 많은 비용을 지불하고 있다. 하지만 악의적이고 공격적인 가짜 뉴스로 인하여 기업의 이미지가 훼손되고 자산 가치에 손상을 입는 경우가 발생한다. 기업들은 본 연구에서 제시한 가짜 뉴스 탐지 모형을 사용하여 가짜 뉴스가 확산되기 시작하는 초기에 신속하고 효과적인 대응이 가능하다. 셋째, 네트워크 임베딩을 사용하여 대규모의 네트워크에 대한 가짜 뉴스 탐지가 가능하다. 네트워크 임베딩은 고차원의 네트워크 데이터를 저차원의 벡터 공간에 표현하는 방법으로 임베딩된 네트워크는 상대적으로 낮은 자원과 시간으로 분석이 가능하다. 따라서 이전의 네트워크 분석 방법에 비해 효과적인 소셜 네트워크 상의 가짜 뉴스 탐지가 가능하다. 본 연구의 한계는 다음과 같다. 첫째, 제한적인 데이터에만 가짜 뉴스 탐지 방법을 적용하였다. 추후 연구로는 보다 더

많은 데이터에 대한 가짜 뉴스 탐지 방법에 대한 연구가 수행되어야 할 것이다. 둘째, 본 연구는 다양한 네트워크 임베딩 방법 중 대표적으로 많이 사용되는 DeepWalk를 적용하였다. 나날이 변화하고 진화하는 가짜 뉴스를 파악하기 위한 다양한 네트워크 임베딩 방법의 응용에 대한 연구가 필요하다.

REFERENCES

[1] H. Allcott & M. Gentzkow. (2017). "Social media and fake news in the 2016 election". *Journal of Economic Perspectives*, 31(2), 211-236. DOI: 10.1257/jep.31.2.211

[2] N. J. Conroy, V. L. Rubin & Y. Chen. (2015). "Automatic deception detection: methods for finding fake news". *Proceedings of the Association for Information Science and Technology*, 52(1), 1-11.

[3] E. Mustafaraj & P. T. Metaxas. (2017). "The fake news spreading plague: Was it preventable?". *Proceedings of the 2017 ACM on Web Science Conference*, 235-239.

[4] M. Potthast, J. Kiesel, K. Reinartz, J. Bevendorff & B. Stein. (2018). "A stylometric inquiry into hyperpartisan and fake news". *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL'18)*, 231-240

[5] J. Tang, Y. Chang, & H. Liu. (2014). "Mining social media with social theories: A survey". *ACM SIGKDD Explorations Newsletter*. 15(2), 20-29.

[6] V. L. Rubin, Y. Chen & N. J. Conroy. (2015). "Deception detection for news: Three types of fakes". *Proceedings of the Association for Information Science and Technology*, 52(1), 1-11.

[7] B. D. Horne & S. Adali. (2017). "This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news". *Proceedings of AAAI*, 1-9.

[8] H. Ahmed, I. Traore & S. Saad. (2017). "Detection of online fake news using n-gram analysis and machine learning techniques". *International Conference on Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments*, 127-138.

[9] Z. Jia, L. Zhao & N. Zhou.. (2017). "The research on fraud group mining which based on social network analysis". *IEEE International Conference on Computer and Communications Analysis*, 2387-2391.

[10] Z. Jin, J. Cao, Y. Zhang & J. Luo. (2016). "News Verification by Exploiting Conflicting Social Viewpoints in Microblogs". *Proceedings of AAAI*, 2972-2978.

[11] E. Tacchini, G. Ballarin, M. L. Della Vedova, S. Moret, & L. de Alfaro. (2017). "Some like it hoax: Automated fake news detection in social networks," *CEUR Workshop Proceedings (1960)*, 1-12.

[12] N. Ruchansky, S. Seo & Y. Liu. "CSI: A hybrid deep model for fake news detection". *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (CIKM '17)*, 797-806.

[13] S. Kwon, M. Cha, K. Jung, W. Chen & Y. Wang. (2013). "Prominent features of rumor propagation in online social media". *IEEE 13th International Conference on Data Mining*, 1103-1108.

[14] T. Mikolov, K. Chen, G. Corrado & J. Dean. (2013). "Efficient Estimation of Word Representations in Vector Space". *Proceedings of Workshop at ICLR*, 1-12.

[15] B. Perozzi, R. Al-Rfou & S. Skiena. (2014). "DeepWalk: Online learning of social representations". *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '14)*, 701-710.

[16] A. Grover & J. Leskovec. (2016). "Node2Vec: Scalable feature learning for networks". *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*, 855-864.

[17] K. Shu, S. Wang & H. Liu. (2019). "Exploiting tri-relationship for fake news detection". *Proceedings of 12th ACM International Conference on Web Search and Data Mining (WSDM 2019)*, arXiv:1712.07709.

박 성 수 (Park, Sung Soo)

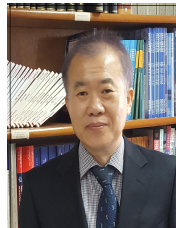
[정회원]



- 2010년 3월 : 성균관대학교 경영대학원 (경영학 석사)
- 2010년 2월 ~ 2019년 2월 : 성균관대학교 경영학과 박사과정 졸업
- 관심분야 : 데이터 마이닝, 감성분석, 인공지능
- E-Mail : volf74@empal.com

이 건 창 (Lee, Kun Chang)

[정회원]



- 1984년 2월 : 카이스트 경영과학과 (공학석사-의사결정지원)
- 1988년 8월 : 카이스트 경영과학과 (공학박사-인공지능)
- 성균관대학교 경영대학 및 삼성융합의과학원 (SAIHST) 융합의과학과 교수
- 관심분야 : 창의성과학, 인공지능, 헬스

인포매틱스, 감성분석 등
· E-Mail : kunchanglee@gmail.com