한국컴퓨터정보학회 동계학술대회 논문집 제28권 제1호 (2020. 1)

가짜뉴스 판별 기법 및 해결책 고찰

이혜진[O], 김진영[*], 백주련(교신저자)[*]
[O]평택대학교 데이터정보학과,
[*]평택대학교 데이터정보학과
e-mail: chocobling10@naver.com[O], wlsdud1517@naver.com[*], jrpaik@ptu.ac.kr[*]

Survey of Fake News Detection Techniques and Solutions

HyeJin Lee[O], Jinyoung Kim[*], Juryon Paik(corresponding author)[*]
[O]Dept. of Digital Information & Statistics, Pyeongtaek University,
[*]Dept. of Digital Information & Statistics, Pyeongtaek University

● 요 약 ●

인터넷 상에서의 허위정보 생산과 유통은 주로 가짜 뉴스를 통하여 이루어진다. 과거에는 신문이나 공중파 TV등 뉴스 기사의 생산과 유통이 매우 제한적이었지만 지금은 인터넷의 발달로 누구나 쉽게 뉴스를 생산하고 유통할 수 있다. 뉴스 생산의 용이성은 정보 공유의 즉각성과 수월성이라는 장점을 제공하지만 반대로 불확실한 뉴스 남발로 인한 정보의 신뢰성 하락과 선량한 피해자를 양산하는 단점 또한 존재한다. 이는 가짜 뉴스가 사회적 문제로 대두되고 있는 이유이다. 에이전트나 스파이더 등의 소프트웨어를 통해 인터넷으로 급속도로 전파되는 가짜 뉴스를 전통 방식인 소수의 전문가가 수동으로 잡아내는 것은 불가능하다. 이에 기술발달로 잡아내기 힘들어진 가짜뉴스에 대해, 역으로 발달된 기술을 활용하여 잡아내려는 시도가 늘어나고 있다. 본 논문에서는 가짜뉴스를 판별하는 다양한 기법들을 탐색하고 해결방안을 제시하고자 한다.

키워드: 가짜뉴스(fake news), 빅 데이터(big data), 인공지능(AI)

I. Introduction

최근 온라인에서 조작적 허위정보의 생산과 유통이 중요한 사회적 문제로 떠오르고 있다[1]. 소수의 사람들만으로도 여론을 쉽게 독점하여 거짓 정보를 퍼트릴 수 있다. 인터넷 오픈소스 백과사전 같은 일반인들의 자발적인 참여로 형성되는 플랫폼을 살펴보면 집단지성 형성과정의 흐름을 확인할 수 있다. 집단지성이란 수많은 개인들이 협력과 각자의 정보 토론 등을 통해 지식을 채워나가는 과정과 결과를 말한다. 소수 개인 자능의 융합보다 더 큰 영향력을 발휘하는 창발현상(창발(創發)또는 떠오름 현상은 하위 계층(구성 요소)에는 없는 특성이나 행동이 상위 계층(전체 구조)에서 자발적으로 돌연히 출현하는 현상이다.) 은 집단지성에서 시작될 가능성이 많다고 보고된다. 위키백과나 오픈소스 프로젝트가 대표적인 예들이다. 하지만 최근 이러한 오픈소스 플랫폼들이 가짜뉴스 공격에 취약성이 드러나고 있다. 가짜뉴스는 이미 전문가가 직접 하나 하나 검증할 수 없는 수준을 뛰어넘어 생산되고 있으며, 진짜 정보를 가려내는 속도는 가짜뉴스의 생성 속도를 따라가지 못한다. 또한 파급된 거짓 정보를 정정하여 다시 전파하는 것은 희박하며 설혹 정정된 뉴스가 오픈된다하더라도 이미 관련자에게 돌이킬 수 없는 정신적·물질적 피해를 입힌다는 것이다. 따라서 생성된 정보들에 대한 신속한 판별이 필요하다.

본 논문은 가짜뉴스의 심각한 폐해를 최소화하기 위해 연구된 가짜뉴스 자동탐지 시스템에 대한 다양한 기법들을 소개하고 기법 간의 융합을 통한 새로운 방식을 제안한다. 분석된 기법들은 빅 데이터를 활용하여 사용자들 간의 연결, 지리적 위치 및 개인 성향을 고려해 진짜와 가짜 정보의 온라인 상에 남기는 데이터 흔적의 특징 및 차이점들을 분석하여 판별을 가능하게 한다.

II. Preliminaries

1. 가짜뉴스 형식 및 습성

사람들은 다양한 SNS를 통하여 무수히 많은 가짜뉴스에 노출된다. 과거에는 SNS로 뉴스를 이용하는 사람들은 소수였지만, 지금은 뉴스의 중요 매개체로 SNS가 자리 잡았으며 대표적인 예로 트위터, 페이스북, 유튜브 등을 들 수 있다 [2]. 비슷한 성향을 가진 대중들에게 아주 매력적인 뉴스 제공자로 신뢰받으며 급속도로 빠르게 성장했으며 현재도 이용자 수는 꾸준히 증가하고 있다. 가짜뉴스는 SNS의 이러한 습성을 이용해 대중들에게 친근하게 다가가 정치적, 혹은 여론몰이를 달성하는 수단으로써 가짜뉴스 유통 플랫폼으로 SNS를 사용한다.

SNS 플랫폼을 통한 가짜뉴스는 대표적인 두 가지의 속성을 갖는다. 첫째는 해당 뉴스가 단일 여론 형성의 도구로 이용된다는 것이며 둘째는 가짜뉴스 전파 집단의 우월함을 과시하는데 이용된다는 것이다. 작위적인 여론 형성과 조작을 막고 확실하고 검증된 정보 생산을 최대화하기 위해서는 어느 것이 가짜뉴스인지를 정확하게 판별할 수 있는 기술이 필요하다.

2. 가짜뉴스 판별을 위한 현 기술들

무분별한 가짜 뉴스들이 생산되는 가운데 이를 판별하기 위한 많은 기술들이 연구되어 있다. 본 논문에서는 다양한 기법들 중에서 논문 [3]과 [4]에 설명된 언어적 특징 기반 접근법, 문서형태분석 기술, 출처신뢰도검증 기술, 콘텐츠교차검증 기술, 딥러닝 모델, 그리고 Jin 등에 의해 제안된 방식을 소개한다.

언어적 특징 기반 접근법[3]은 가짜 뉴스의 주요 언어적 특징을 추출하여 사용하는 방법론이다. 언어에는 다양한 형태들과 기법들이 있으며, 대표적으로 구두점, 심리 언어적 특징, 가독성, 구문 등을 활용하는 방법 중심으로 활용한다. 구두점은 가짜 뉴스 알고리즘이 사기성이 짙은 텍스트와 진실성이 있는 텍스트를 구별하는데 도움을 주며, 탐지를 통해 구현되는 여러 가지 유형의 구두점을 수집한다. 또한 심리 언어적 특징은 언어의 품사 분류 등을 결절하는 데 도움을 주며, 심리적 구조 같은 여러 특징을 갖는 집단들의 유사성을 바탕으로 하여 데이터를 그룹으로 나누는 것이 가능하다. 가독성은 콘텐츠 특징 추출이 포함되며, 문법의 경우는 하나로 이어지는 문법의 특징을 추출하고, 추출된 특징은 상위 연결 포인트들과 결합된 어휘 생성 규칙에 의존한다.

문서형태분석 기술은 문서형태가 사전에 합의된 형태가 아닐 경우 가짜뉴스로 판별해내는 기술이다. 국제뉴스기사 표준 규격 또는 공인된 표준문서 틀을 기준으로 형태가 올바르지 않는 기사를 가짜뉴스라고 판단한다.

출처신뢰도검증 기술은 뉴스기사가 게시된 웹사이트나 출처라고 정의할 수 있는 웹사이트에 관한 데이터베이스 바탕으로 별도로 신뢰성을 확인하여 가짜뉴스를 판별하는 방식이다. 이 방식은 가짜뉴스를 생성하는 곳은 지속적으로 가짜뉴스를 생성하는 습성이 있으며, 진짜뉴스를 생성하는 곳은 지속적으로 진짜뉴스를 게시하는 습성이 있다는 관찰결과를 바탕으로 작동한다. 따라서 잘 알려져 있거나 믿을 수 있는 언론사와 엘로저널리즘의 요약본들을 구분하여 출처 신뢰도를 판단한다고도 볼 수 있다. 내용 안에 포함되어 있는 사진 및 문서, 상기되어 있는 출처에 대한 데이터베이스 기반 신뢰성 검증을 실행하는 경우도 포함 될 수 있다.

콘텐츠교차검증 기술은 뉴스 기사의 키워드를 추출 후 검증할 만한 가치가 있는 문장을 추출하고, 추출된 문장들을 중심으로 검색엔진을 통한 크롤링으로 유사한 뉴스들을 추출하여 해당 뉴스들의 유사성을 검증하여 분석 비교하는 방식이다. 추출된 유사도가 높은 글들은 다른 방식들과 결합하여 재확인하는 방법으로 한 번 더 검증을 수행하여 최소 두 가지 판단 방법으로 가짜뉴스여부를 판별한다.

최근에는 인공지능 기법의 고도화를 활용하여 한글 가짜뉴스 탐지를 위한 딥러닝 모델도 제시되었다. 충분한 데이터 셋을 이용해 정보가 많은 플랫폼을 설정하고 실시간으로 올라오는 정보들을 분석 후 검증한다. 이와 같은 과정을 위해서는 지속적인 관찰이 필요하다. 정확한 검증을 위해 통계적 기법도 적절히 활용하여야만 신뢰성 높은 딥러닝 기반 가짜뉴스 탐지 기술이 만들어 진다.

또 다른 기술로 인공지능 기법 기술인 이상 확산 패턴 탐지 기법을 효과적으로 적용한 Jin 등의 의해 제시된 기법이다. 트위터에서 서로 대치되는 관점을 발견한 후 신뢰성 전파 네트워크를 구축하여 반복적인 판단을 적용하여 가짜뉴스에 대한 가짜 정보 및 신뢰할 만한 정보에 대한 판별 결과를 생성하는 방법론이다. 주제들 간에 대한 관점들을 쌍으로 정하고 군집분석을 통해 두 개의 서로 반대되거나 대치되는 관점으로 분리한다. 또한, 트위터 멘션 간의 신뢰성 네트워크를 쌓아올리며 긍정적인 반응과 부정적인 반응에 따라 링크의 연결정도를 분석하였다. 하지만 이들의 접근법은 다른 SNS를 포함하지 않은 트위터 데이터들만 가짜뉴스의 데이터로 활용한다는 한계점을 갖는다.

Table 1은 가짜뉴스 판별을 위한 기술들에 장·단점을 정리한다.

Table 1. 가짜 뉴스 판별 기술

기술	장점	단점
언어적 특징 기반 접근법	언어에는 여러 가지 속뜻이 담겨 있으며 사람마다 쓰는 법도 각각 다르다. 따라서 그 유형의 분석 기법이 다양해 질 수 있으며, 유사성을 측정하기도 쉽다.	너무 어휘를 생성한 규칙에 의존한다는 게 단점이다.
문서형태분석 기술	정확하게 문서형태를 가지고 있으며, 명확성이 떨어지면 떨어질수록 분류하기 쉽다.	자유 형태를 가지고 쓰인 뉴스는 분류하기 힘들다.
출처신뢰도검증 기술	공인이 된 곳이나, 오보만을 올리는 신뢰성 격차를 많이 보인다면 효과가 있다.	특정 사이트나 사람을 일반화하는 것이기에 그만큼의 확실한 검정이 없다면 뉴스를 분류하기에는 적합하지 않다.
콘텐츠교차검증 기술	유사도가 높은 글들을 뽑아내 분석하는 방법 외에도 다른 방법을 겹쳐 확인하는 방법으로 2차 검증으로 인해 정확도가 높다.	다른 방법보다 검정 방법을 더 섞기 때문에 시간이 오래 걸린다.
딥러닝 기술	가짜뉴스에 대한 정보가 많으면 많을수록 실시간 성능 개선이 가능하다.	성과가 통계적 의미로 검증이 확실한지는 아직 미지수 이다.
Jin 등에 의해 제안된 기술	인공지능이 함께 융합되어있는 기법이기 때문에 패턴 탐지 기법이 적절히 결합되어 있어 완성도가 높다.	트위터는 익명성이기 때문에 정확도가 떨어지며, 트위터 외의 SNS는 구축되지 않았다.

III. Proposed Scheme

본 논문에서는 SACM (Sensing Automatic Classification Model) 기술과 전파패턴분석 기술을 응용해 새로운 판별 기술을 제안한다. 먼저 사람들이 많이 접할 플랫폼 분석이 수행되어야 한다. 때문에 다수의 사람들이 사용하는 SNS를 분석할 필요성이 있다. 그 기술로는 SACM이 적합하다[5]. SACM 기반 서비스 시스템은 수집된 수많은 SNS 상의 빅 데이터를 저장한 후 분석을 하거나 정보 처리에 사용하기 좋은 형태로 데이터를 제작하기 위해 데이터 전처리 모듈을 제공하는 것이다.

이 시스템으로 사람들이 쉽게 접하는 SNS 상에 제공되는 다양한 뉴스들을 접한 대중들의 반응을 보고 가짜뉴스 관련 키워드들이 많이 포함되어 있는 뉴스 데이터를 관리 시스템이 병합하고 저장한다. 이후 해당 뉴스들에 가짜뉴스판독 기술을 시행한다. 이를 위해 전파패턴분석 기술[6]을 함께 응용한다. 이 기술은 뉴스기사가 널리 퍼지게 되는 패턴을 기반으로 가짜뉴스를 판단하는 기법이다. 특정 기사가 SNS나 여러 공유 등을 통해 퍼지게 되는 패턴을 추적하고 분석할 때, 보통 진짜뉴스의 경우는 처음 이슈가 된 짧은 시기에 전파되는 습성을 보이며 각 전파 연결 포인트 간 네트워크 연결 관계가 보이고 이슈메이커의 연결 포인트를 기점으로 확장되는 패턴을 보인다. 이와 반대로 가짜뉴스는 장기적인 시간에 걸쳐 다발적으로 전파되는 경향을 보이며 전파 연결 포인트 간 네트워크 연결 관계가 진짜 뉴스 보다는 적은 패턴을 보여준다. Fig. 1은 진짜뉴스와 가짜뉴스의 네트워크 패턴의 차이점을 도식화한 것이다.

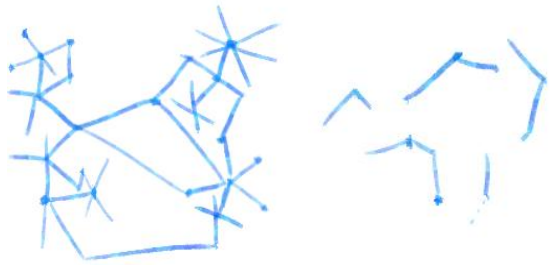

Fig. 1. 진짜뉴스(좌)와 가짜뉴스(우)의 네트워크 연결 패턴

각 패턴에서 네트워크 연결 관계가 보다 적은 뉴스들을 SACM 기술로 저장하고 분석하는 기법을 시행한다면 단시간에 가짜뉴스를 판별할 수 있는 응용기술을 가질 수 있다. 전파패턴분석은 단순하게 웹에 등록된 뉴스기사는 키워드를 뽑아 크롤링한 내용의 유사성 검토를 수행 후 날짜순 정렬로 원본을 찾아낸다. 그 후에 전파경로를 추적해야하기에 정확성이 떨어지는 편이다. 그리고 수많은 정보들을 검색하여 저장 후 패턴을 추측할 수 있기에 시간과 인력 등의 자원 소모가 많다는 단점을 가지고 있다. 그렇기 때문에 SNS를 주체로 분석하는 SACM이 선 수행에 적합하며 그 응용 기술을 하후 전파패턴 분석을 응용하여 수행한다면, 올바르지 못한 의도로 몇몇 소수의 집단 및 개개인이 근거 없이 거짓 정보를 퍼뜨리는 경우 가짜뉴스를 판별하기가 가장 적합하다는 장점을 살릴 수 있다고 생각한다.

IV. Conclusions

가짜뉴스의 파급력은 나날이 확대하고 있으며 선의의 피해자가 지속적으로 상승하고 있다. 인터넷 및 모바일 기술의 고도화는 다양하고 많은 정보들을 빠른 시간에 생산하여 급속도로 대중에게 공유된다. 그만큼 참과 거짓을 제대로 구분하기 어려워질 뿐더러 지식의 혼란을 야기하기도 한다. 가짜 정보들을 판별하기 여러 방안들이 최근 제시 및 적용되고 있지만 해당 뉴스들은 여전히 퍼지고 있다. 좀 더 신속하고 정확하게, 가짜 정보의 생성 즉시 참과 거짓을 판별할 수 있는 시스템이 지속적으로 연구되어야 한다.

ACKNOWLEDGEMENT

이 논문은 2019년도 정부 (과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (NRF-2017R1A2B1007015).

REFERENCES

[1] Article on The Science Times, December 20th, 2018

[2] H. U. Lee, “Responses of Users' Comments Depending on the Format and Issue of the Fake News”, Master Thesis, Mass Communication, Kyungpook National University, Aug. 2019.

[3] Yoonjin Hyun, “Text Analytics-based Fake News Detection Methodology Using News and Social Data”, Ph.D Thesis, Graduate School of Buisness IT, Kookmin University, Feb. 2019

[4] Tae-Ik Yoon and Hyunchul Ahn, “Fake News Detection for Korean News Using Text Mining and Machine Learning Techniques”, Journal of Information Technology Applications & Management Vol. 25, No. 1, pp. 19 - 32, Mar. 2018.

[5] Sungwhan Bae, “A Study on SNS Data-based Sensing Automatic Classification Model for Consumer Opinions on Social Big Data”, Ph.D Thesis, Aug. 2019.

[6] ChangKeun Ma, “A Proposal for Auto-Factchecking Model Categorization by Technique and Optimization by News Subject”, Master Thesis, Feb. 2018.