특집 03

인공지능을 활용한 가짜뉴스 찾기 기술 및 정책적 해결방안에 대한 사례연구

(과기정통부의 인공지능 R&D챌린저를 중심으로)

강장묵 (남서울대학교 빅데이터산업보안학과)

목 차

1. 서 론

2015년 미국대통령 45대 선거 때부터 가짜뉴스에 대한 사회적 관심이 급증하였다. 우리나라에서는 2016년부터 가짜뉴스에 대한 미디어적인 접근과 사실 확인(fact check) 등을 언론인 및 전문가의 수작업에 기초하여 대응하고 있었다. 이 방법은 정책에 의한 가짜뉴스 해결방법으로 (그림 1)의 비기술적 접근에 해당한다. 각 언론사가 팩트체크라는 코너를 만들어 수작업에 기초하여 가짜뉴스를 판별하는 방법이다. 그러나 이 방안은 비용이 많이 들고 매번 사람이 관여하여야 하는 문제가 발생한다. 따라서 정부에서는 인공지능 기술로 사회 현안 문제인 가짜뉴스를 찾는 기술 개발의 필요성이 높다고 판단하고 이를 경진대회를 통해 해당 기술을 개발한 팀을 선정하여 해결하고자 하였다.

가짜뉴스를 탐지하는 방법은 (그림 1)과 같이 비기술적 방법과 기술적 방법이 있고 그 중 필자는 2017년과 2018년 해당 문제를 풀기 위해 인공지능 기반 탐지 기법을 적용하였고, 2019년 현재는 하이브리드 분석 기법(consider the source 분석)을 적용하여 가짜뉴스 탐지 알고리즘을 개발 중에 있다. 이를 이용한 시스템으로는 페이스북의 가짜뉴스 판별 시스템, 구글과 언론사가 연계한 가짜뉴스 판별 서비스, 카이트스-가짜뉴스 판별 알고리즘 연구, 서울대-팩트체크 사이트 운영, 미국 텍사스대와 미시시피대의 ‘클레임버스터’, 인디에나주립대의 지식그래포, 팩트마타 등 다양한 알고리즘과 서비스가 존재한다.

필자가 이 글을 작성하게 된 배경은 2017년 대한민국 1회 R&D 인공지능 경진대회(과학기술정보통신부 주최, 이하 과기정통부)에 개인자격으로 필자가 참여하면서, 경험한 사례를 소개하기 위해서이다. 대한민국 과기정통부에서는 어떤 자격 조건(초등학교 졸업생부터 대학교수까지 누

(그림 1) 가짜뉴스 탐지 기법 [1]

구든 참여 가능)이나 경쟁의 규모(개인과 연구소 인력 100여명이 경쟁한다 할지라도 상관없음)에 제한을 두지 않고, 개인, 기업, 연구소, 대학 등을 모아 국내 최초 인공지능 경진대회를 개최하였다. 경쟁의 순위에 따라 장관상, 기관장 상과 15억 원의 부상(1년 후 연차과제 단독선정 시 15억 전액 추가 지원) 등을 주었는데, 오직 평가 전에 공개된 정량평가 방식으로 수상하는 무한 베틀 방식이었다[2].

이 글은 인공지능을 활용한 가짜뉴스 경진대회에 대한 사례를 통해 필자가 경험한 인공지능 기술의 개발과 사회 현안 문제 해결을 위한 정책 방안들이 어떻게 조화를 이룰 수 있는지에 대한 시사점을 독자에게 드리고자 2년 동안 치러진 경진대회 과정을 복기하는 방식으로 작성하였다. 이미 미국에서는 정부 주도로 민간인 누구나 참여하는 경진대회를 열어, AI에 대한 관심을 높이고 더불어 큰 상금과 영광을 주는 챌린저 방식이 정착된 지 오래이다. 일예로 미국 방위 고등 연구 계획국(DARPA)은 도전적인 문제를 해결하기 위한 '그랜드 챌린지'를 개최하였고 2015년 카이스트에서 개발한 재난대응 로봇 '휴보'가 우승을 차지한 대회도 DARPA에서 주관한 로봇 챌린지였다. 이들 경진대회의 가장 큰 장점은 논문이나 특허 같은 실험실 수준의 정량적인 지표가 아닌 실제 문제 해결의 능력을 기준으로 한다는 점이다[3].

필자는 350만 건의 뉴스 데이터를 2017년과 2018년에 거쳐 웹 크롤링 (Web Crawling, 뉴스 기사를 파싱하여 분산 DB에 제목, 저자, 날짜 등으로 분류하여 저장)하고 이를 딥러닝시키면서 '가짜뉴스 찾기'의 정확도를 높였다. 실제 데이터를 처리하고 분석하는 세부 내용의 전부를 이 글에서 다룰 수는 없지만 그 과정에서 데이터 과학자가 겪을 몇 가지 경험과 내용을 토대로 통찰력을 공유하자 한다. 예를 들면, 우리나라 뉴스 기사의 특징은 종합지, 지방지, 중앙지, 잡지, 블로그, 카페글 등 매체에 따라 색인화 작업을 달리하면 결과도 차이를 보였고 동일한 해당 매체라 할지라도 경제, 과학, 정치, 사회면에 따라 가짜뉴스의 주요 패턴이 미디어적인 관점에서 차이가 있었다. 실제 그 차이는 반드시 앙상블 기반의 인공지능 알고리즘의 우수성이나 컴퓨팅 성능만으로 해결되는 것은 아니었다. 필자는 데이터를 직접 다루면서 데이터 이면에 해당 데이터의 성격 즉 특징을 파악하는 것이 그 무엇보다 중요함을 터득하였다.

이 글의 목적은 인공지능에 연구경험이 일천

한 필자가 국내 최우수 연구팀인 카이스트, 서울대, 한국전자통신연구원(ETRI) 등이 참여한 경진대회에서 우승한 것은 사회과학의 한 방편인 뉴미디어에 대한 이해로 코퍼스(corpus;말뭉치)를 고도화한 것, 가짜뉴스 문제를 제작하는 과정에서 필요로 했던 인문학적 지식의 요구 사항 등을 소개하여 공학도들에게 인공지능을 활용하여 사회현안문제를 해결하고자 한다면 사회과학, 인문학과의 융복합이 절대적으로 필요하다는 것을 강조하기 위해서이다. 실제 이상의 방법은 필자가 인공지능 개발 과정에서 습득한 대단한 발견이 아니라, 미국 MIT의 미디어랩 또는 캐나다 토론토 대학교와 공동으로 연구를 수행중인 토론토 인공지능 연구소 등에서 인문학, 사회학 등과의 숙의를 거쳐 인공지능을 적용한 사례를 따라간 것뿐이다.

필자는 2년 동안 동일한 형태의 성능을 국가성능평가 공인 기관으로부터 평가받고 2018년 12월에는 1등을 하여 현재, 후속 연구개발(R&D) 비용을 지원받았다[4]. 이상의 사례에서 필자는 2017년 150만건 기사의 딥러닝, 2018년도에는 200만건 내외의 기사를 딥러닝시켜, 알고리즘의 정확도를 높였다. 그리고 그 결과로 작성 중이거나 출판한 10여건의 SCI 및 30여건의 SCOPUS 그리고 20여건의 특허 출원 등은 소문, 구술, 민담 등 고전에서 발견한 가짜 뉴스에 대한 인문학적 함의를 융/복합하는 세계적인 인공지능 연구 추세이기도 하다.

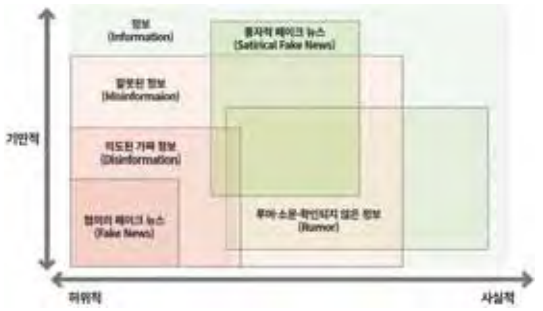
엄밀하게 필자는 언론의 자유 및 국민의 알권리 그리고 프라이버시 등을 통해 미디어적 가치를 바라본 사회과학의 변인들을 어떻게 코딩해 나갈지에 대한 성찰을 통해 가짜뉴스 찾기 알고리즘의 정확도를 높였다. 그 이면의 내용(경진대회 방식, 경진대회가 바라본 가짜뉴스의 정확도 측정 방법 등)을 살피면 다음 장과 같다.

2. 가짜뉴스의 이해

2.1 가짜뉴스의 정의와 실태

국내 가짜뉴스에 대한 개념적인 분석은 황용석(2017, 2018)의 연구에서 체계적으로 정리되어 제시되었다. 그 후 여러 미디어 및 언론학자들이 황용석의 글을 차용하거나 인용하여 가짜뉴스에 대한 추가적인 분석을 시도하였다. 그러나 여전히 가짜 뉴스의 개념 확립은 진행 중인 것처럼 보인다. 가짜뉴스는 풍자적 가짜 뉴스(satirical fake news), 루머(rumor), 허위정보(disinformation), 거짓정보(hoax), 오인정보(misinformation), 패러디(parodies) 등을 지칭하는 여러 개념과 용어가 혼재되어 사용되고 있다[5]. 따라서 가짜뉴스를 찾아주는 인공지능이 단순히 사실 확인 (Fact Check)라는 기능적인 접근만을 한다면, 뉴스 독자들은 사회현상을 복잡한 풍경에서 고찰하고 분석한 결과라고 판단하지 않을 것이다. 예를 들어 논란의 대상인 풍자적 가짜 뉴스(satirical fake news)는 '잘못된 정보'의 유형 중 한가지로서 그 실익과 사회 폐해 간에 상충되는 측면이 존재한다. 여기서 잘못된 정보란 사실과 전체 또는 부분적으로 다른 정보를 말하며, 가장 포괄적 용어이다. 풍자적 가짜 뉴스란 상대의 결점을 비유를 들어 비웃으면서 공격하고 폭로하는 것이다. 패러디(parodies)란 풍자와 유사한 맥락으로 이해할 수 있다[6].

텔레비전을 중심으로 발달한 풍자적 가짜 뉴스는 정치의 무거운 측면을 들어내고 사실과 유머를 적절히 버무려 시의성 높은 정치적 사안을 논의하는 정치담론 양식이다. 정치 비판이라는 순기능을 갖고 있다[7]. 이런 풍자적 가짜뉴스는 SNS를 통해 쉽게 세상에 전파되고 그 효과가 크다. 루머 역시 잘못된 정보라 할 수 있다. 루머는



출처: 황용석(2017.3.20.)

(그림 2) 가짜뉴스와 유사 개념의 관계를 보여주는 도식

진위가 확인되지 않은 진술이 사람들 사이에 떠도는 것인데, 불확실성, 위험 상황, 혹은 잠재적 위협 상황에 놓인 사람들이 그 위험을 통제하고 이해하기 위해 공유하는 것으로 정의한다[8].

잘못된 정보지만 풍자적 가짜 뉴스나 패러디, 루머를 가짜 뉴스로부터 제외시키는 경향이 존재한다. 오인정보(misinformation) 또한 잘못된 정보를 생산한 해당 언론에 책임을 물을 수 있다는 이유로 가짜 뉴스에서 제외된다. 이러한 유형의 정보를 가짜 뉴스로부터 배제하지 않고 통제할 경우, 표현의 자유 침해나 언론의 자유를 침범하는 일종의 검열이 될 수 있기 때문이다.

이상의 문제 인식은 공적 영역에서 표현의 자유와 국민의 알권리 등에 대한 문제와 갈등하면서 (그림 1)과 같이 비기술적인 정책 방법으로 가짜뉴스를 찾게 된다. 따라서 사회과학현안 문제인 가짜뉴스를 해결하기 위해서는 공학도 역시 표현의 자유와 국민의 알권리 등 사회과학이론에 대한 이해를 기초로 변수 설정과 가중치 그리고 이를 바탕으로 학습데이터를 구축할 때 분류 등을 달리하게 되고 그런 노력이 실제 가짜뉴스 탐지 정확도가 높아질 수 있다. 기술이 기능적 해결에만 집중하면 복합적 사회현상인 가짜뉴스를 해결하는데 한계를 보인다.

그리고 사회현상을 탐구하는 정책집단에 의해 연구비를 지원받는 공학의 경우, 당장 해당 연구의 성과가 미흡해 보인다는 이유나 여타 연구비가 사회현안, 산업 등에 미치는 다양한 파급 효과를 인지하지 못하는 경우가 비일비재하기 때문이다. 가짜뉴스는 정치적으로 민감한 사인뿐만 아니라, 전 사회적인 병폐이고 전 국민이 이것으로 인해 한 두번 고통을 겪은 중대한 사인이다. 특히 우리나라에서는 가짜뉴스를 빌미로 인신공격성 정보 소통이 사회에 큰 피해를 주고 있으나 사적 영역인 개인 간 비밀통신(SNS)라는 이유로 법의 사각지대에 놓여 있다.

우리나라와 같이 도시 인구 밀집도가 높고 경쟁이 치열한 경우에는, 불필요한 사적 참견과 사적 일탈(성범죄가 아닌 간통 등)에 대해서도 뒷담화 형식으로 개인을 비난하는 카카오톡, 텔레그램방 등이 은밀하게 성행하는 중이다. 이곳에서 이루어지는 불법적인 따돌림, 왕따, 루머 등에 의한 피해는 사례와 그 정신/경제적 비용을 측정할 수 없는 실정이다.

따라서 가짜뉴스는 다음과 같은 공적 영역과 사적영역에서의 폐해로 나뉘고 실제 연구 대상은 공적 영역인 뉴스였지만, 전문가 자문 과정에서 우리나라의 문화특성 상 '수많은 사인들이 사인에 의해 단톡방 등에서 매도당하고 억울한 소리를 듣는 경우'가 비일비재한 경우를 정성적으로 분석할 수 있었다. 실제 그로인해 오해를 받고 경제적 사회적 지위를 잃는 경우까지 발생하고 있어 이에 대한 연구가 시급한 실정이다.

이처럼 가짜뉴스는 사회학자 사이에서도 그 범위, 일탈의 수준, 해악의 정도와 실제 제도적인 보완 등에 논란이 존재하나, 공학자가 가짜뉴스를 찾는 알고리즘을 설계하거나 개발하는 과정에서 이에 대한 이해가 실제 데이터 분석의 결과가 주는 의미를 찾지 못할 때 합당한 가설을 세

우는데 기여하는 측면이 크다. 필자는 이 측면이 공학자가 사회과학적 이론을 일정 수준 이해하여야 하는 이유라 여기고 융합적 이해가 실제 코딩 과정이나 개발과정에 트러블 슈팅 (Trouble shooting)을 가능하게 한다고 여긴다. 그렇다면 가짜뉴스의 정의에 포함되는 다양한 잘못된 정보 중 피해 대상을 중심으로 나누면 다음과 같다.

2.2 공적 영역 vs 사적 영역

흔히 가짜뉴스는 정치적인 이슈에서 '갑론을박'하는 경우가 많다. 그러나 필자의 정성 평가 결과 중 영남대학교 언론정보학과 박한우 교수의 주장에 따르면, 가짜뉴스는 정치적 측면뿐만 아니라 경제적 측면에서도 피해 규모와 사례가 크다. 흔히 가짜뉴스라고 하면 사회의 공적 영역에 대한 혼란 및 피해에 대한 경우를 연상한다. 특히, 뉴스라는 공식적 채널과 보도의 형태를 띠고 있어 가짜뉴스의 폐해는 명예훼손 등과 같은 형법으로 다투는 공적 영역으로 다루게 된다.

그러나 필자가 연구한 결과 실제 측정하기 곤란한 영역(언론 기사로 소개된 경우가 아닌)인 사적 영역에서 가짜뉴스의 폐해는 심각한 수준에 이른다.

초연결 사회로 급격히 이행되면서 소셜 네트워크 서비스와 수많은 단톡(카카오톡 등)방이 성행하게 되었다. 수많은 사람들이 자신을 숨기거나 몇몇 친분 있는 끼리끼리의 모임방을 이용하여, 귓속말로 특정 사인의 프라이버시(비리의혹, 내연관계, 말투 등 인격적 모욕을 일방적으로 전파하는 방법)와 명예훼손을 침해하는 사례는 이루 헤아릴 수 없이 많다.

농경사회의 전통과 유교주의적 교리가 신자본주의와 기형적으로 결탁한 2019년도의 대한민국에서는 사인에 의한 사인의 혐오 발언, 사생활을 통한 모독, 특정인을 왕따 시켜 자신의 사회적 성취를 이루려는 삐뚤어진 경쟁심리, 교만과 방조 등 다양한 동인으로 정신병적인 뒷담화(타인을 흉보거나 욕보이는 행위)가 공공연하게 이루어진다.

단톡방에서 이런 흉을 보고 그 방을 나가거나 그 방에서 그런 대화를 금지시키는 건전한 비판을 하기 보단 묵시적으로 읽음으로 동조하거나 침묵의 나선형을 통해 몇몇에 의한 악의적 입소문(viral)이 형성된다.

그러나 카카오톡의 단톡방은 개인정보보호(방송이 아닌 사인 간 통신의 영역으로 간주하는 법)법에 의해 그 대화 내용을 볼 수 있거나 건전한 감시 자체에 대한 논의가 표현의 자유와 심각한 충돌을 야기하여 아노미 공간으로 남겨져 있다.

필자는 철저하게 공적 영역인 공개된 뉴스 기사 중 참인 기사를 선정하여 이를 다시 아르바이트생을 고용하여 가짜기사(객체 변환, 수동과 능동 변환, 주어와 목적어 변환 등)를 생성하고 이를 학습시켜 알고리즘의 유효성, 실제 정확도 등을 높였다. 그러나 사적 영역에서 이루어지는 가짜뉴스는 접근자체가 불가능하고 실제로 그 피해가 커서 이에 대한 연구의 필요성을 연구를 수행할수록 커지는 것도 사실이다. 이에 2019년 후속 연속 과제로 선정된 가짜뉴스 찾기 연구팀은 역할 기반의 챗봇을 통해 메신저에서 뉴스에 대한 기본적인 진위를 자동 판별하고 특별한 공유 방식이나 링크가 가짜 뉴스 등의 진원지라는 사실 공지 그리고 가짜뉴스나 사인비방의 주요 문장이나 단어 그리고 표현을 딥러닝 방식으로 선별하여 밑줄 또는 중간줄 표현식으로 제공하는 것 등으로 '메신저에서의 가짜뉴스 탐지 및 방지'에 관한 해당 문제를 해결하는 특허를 출원하고 관련 논문을 현재 해외 저널지에 투고하여 심사를 받는 중에 있다.

(그림 3) 출원된 챗봇 기반의 가짜뉴스 찾기 서비스 UX

다음은 실제 가짜뉴스 경진대회의 평가 알고리즘, 방식, 임무 등에 대해 설명한다.

3. 가짜뉴스 찾기의 방법 및 성능 측정

3.1 가짜뉴스 찾기 알고리즘의 방법 및 성능[9]

서론에서 가짜뉴스를 찾는 기술적 방법과 비기술적 방법을 소개하였다. 그 방법 중 인공지능만으로 해결할 수 없는 분야를 'consider the source'를 통해 해결하는 방법 연구가 최근 미국 등에서 활발하게 소개되고 있는 연구이다.

그러나 필자가 경진대회에서 다룬 방법론은 임무 1과 2를 이미 공개된 평가 방법론(출제될 가짜뉴스 예제를 경쟁자 공통으로 충분하게 분석한 후)에 최적화된 알고리즘을 개발하는 방법을 사용하였고, 코드의 부정을 막기 위해 1등한 팀의 전체 소스 코드는 해당 경쟁팀에게 현장에서 공개하여 소스 코드에 부정이 없음이 판단될 때, 해당 팀이나 개인에게 연구비를 지원하는 방식이다. 따라서 본 글에서 필자가 개발한 여타 기술의 상세한 내용까지 기술하는데는 다소 불편함을 느껴, 통찰력을 줄 수 있는 방법론적 팁을 공유한다.

2017년의 경진대회는 해당 T(rue)와 F(aulse) 명제만을 고려하였다. 뉴스 자동 판별 솔루션의 성능은 "ROC 곡선의 하단면적"을 계산하여 평가한다. ROC 곡선의 하단면적이란 "Area under the ROC curve" (AUROC)을 뜻한다. AUROC는 1에 가까울수록 우수하다.

이 방식은 검출 분야에서 널리 사용되고 있는 평가 방식으로, True Positive Rate (TPR)와 False Positive Rate (FPR)를 이용하여 제안된 알고리즘의 성능을 평가하는 방식이다. 가짜뉴스는 참과 거짓이라는 True와 False를 검출하는 것이다. 이의 정확도인 ROC 커브를 통해 정량 평가 받는다.

TPR은 실제 True를 True로 정확하게 예측한 비율을 의미한다. FPR은 실제 False를 True로 잘못 예측한 비율을 의미한다. TPR과 FPR의 수식은 다음과 같다.

$$TPR = \frac{TP}{TP + FN}, \quad FPR = \frac{FP}{FP + TN}$$

2017년에는 해당 T(rue)와 F(aulse) 명제만을 고려하여 알고리즘을 설계하였으나 2018년과 2019년에는 문맥 기반으로 가짜 뉴스 문제의 정확도를 향상시키고 궁극적으로는 의미추론을 가능하게 하는 알고리즘을 설계 및 개발하고 있다. 자연스럽게, 'FPR(가짜뉴스를 참 뉴스로 예측한 비율)' 뿐만아니라, 진짜뉴스를 가짜뉴스라고 판단하는 경우 등에 대한 문제 처리의 중요성을 판단해야 하는 필요성을 발견하였다.

아래 <표 1>과 같이 2017년에는 가짜를 가짜라 하는 경우, 가짜를 진짜라 하는 경우, 진짜를 진짜라 하는 경우, 진짜를 가짜라 하는 경우에

〈표 1〉 가짜/진짜뉴스를 가짜/진짜라고 검출하는 경우

		Label	
		True	False
Prediction	True	TP (True Positive)	FP (False Positive)
	False	FN (False Negative)	TN (True Negative)

대해서 TP, FP, FN, TN의 의미로 정리하였고 각 경우에 차이를 두지 않았다.

현재 연구에서는 가짜라고 판명된 기사에 대한 심층 분석 또는 진위 여부 판단 첫 단계에서의 중요도 처리 등을 달리하는 적절한 방법론을 적용하고 그 결과 값을 비교 분석 하는 하이브리드 방식으로 가짜뉴스를 탐지하고 있다.

3.2 ROC 곡선의 하단 면적과 임무별 가중치[11]

정량 목표 임무 1은 명제 기반 사실 불일치 검출 정확도이다. 자세하게는 외부 소스에서 입력된 사실을 지지 또는 부정하는 명제를 검출하는 것이다.

두 번째 임무 2는 문맥 기반 사실 불일치 검출 정확도이다. 자세하게는 외부 소스에서 입력된 사실을 지지 또는 부정하는 문맥(명제 집합)을 검출하는 것이다.

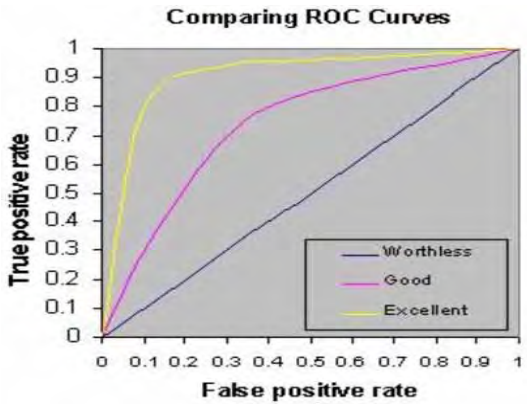
구체적으로는 TP, FP, FN, TN은 예측 경계값(cut-off value, threshold)에 따라 다른 수치를 가지게 된다. 그리고 다른 수치에 따라 TPR과 FRP의 수치도 영향을 받는다.

ROC 곡선은 예측 경계값으로 FPR을 0에서 1까지 변화시킬 때 TPR이 어떻게 변하는지 나타낸 그래프이다.

ROC 곡선이 좌상향으로 향할수록 성능이 좋다고 판단하며 이를 수치로 비교하기 위해 “ROC 곡선의 하단면적”을 사용한다[12].

과학기술정보통신부에서 선정한 심사위원회는 각 참가팀이 제출한 2개 임무에 대해 ROC 곡선의 하단면적을 계산하여 평가한 후, 임무별 가중치를 두어 최종 점수로 환산한다.

이때 평가값은 배치파일로 구성된 자동화 방법으로 값을 넣으면 결과가 자동으로 도출되는 방식으로 부정을 막았다. 임무별 가중치는 위 <표 2>와 같다.



출처: http://gim.unmc.edu/dxtests/roc3.htm

(그림 4) ROC 커브 [11]

〈표 2〉 임무 1/2별 가중치

	임무 1	임무 2
가중치	0.50	0.50

4. 가짜뉴스 해결 방안

4.1 Consider the Source

3장에서 필자는 가짜뉴스를 검출하기 위해 실제 뉴스로부터 가짜뉴스를 학습 데이터로 만드는 작업을 소개하였다. 그리고 이를 바탕으로 실제 진짜 뉴스를 가짜뉴스로 조작한 문제를 풀어 그 정확도에서 입상한 이력을 설명하였다.

실제 판단은 부울 값으로 처리되는데, 0과 1사이의 소수점 6째 자리까지의 실수로 진짜뉴스와 가짜뉴스의 예측 경계값을 표시한다. 가짜도 진짜도 아닌, 판단을 할 수 없는 경우에는 0.5라는 점수가 부여된다.

이 기술을 실제 뉴스 현장에 적용한다면 어떤 문제가 발생할 것인가. 가짜뉴스일 확률이 높은 뉴스를 식별, 삭제하는 강제적 방법이 가능할 것이다.

출처: http://laverne.libguides.com/fakenews

(그림 5) 프로세스기반 가짜뉴스 판별에 필요한 체크 포인트[13]

그러나 2장에서 가짜뉴스에 대한 정의, 실태, 공적 뉴스와 사적 뉴스 등에 대한 분석에서 가장 큰 문제는 표현의 자유 등과의 충돌 문제이다. 이를 해결하기 위해서는 기술에 기반을 둔 검출 정확도와 함께 프로세스 상에서 가짜를 구별해 내는 절차적 방안이 보완될 필요가 있다. 이를 위해 Consider the Source를 고려한 프로세스가 필요하다. 그 Source로는 출처, 저자의 평판, 채널, 날짜, 공유 수 등 다양한 변인으로 구성된다.

(그림 5)는 어떤 단계 즉 순서로 가짜뉴스를 판단하는 것이 합리적인지를 도식화한 것이다. 이를 해결하기 위해서 소통 과정에서 걸러지는 선순환 구조에 대한 모형을 많이 제시할 필요가 있다. (그림 5)는 그 중 하나로 현재 수행되는 연구에서 각 프로세스의 모듈로 어떤 체크포인트를 둘지를 결정할 때 참조한 모델이다. 이 외에도 다수의 선행 모델을 분석한 후 변인에 데이터를 대입하여 최적의 단계를 선택하고 이를 기반으로 가장 최적의 모델을 결정하고 있다.

4.2 가짜뉴스 해결방안

가짜뉴스를 해결하는 것은 우선 자연어 처리 기술이 완벽해질 때 가능하다. 그 완벽이란 인공지능이 한글을 스스로 학습하여 언어를 습득하고 해독해 낼 수 있는 상황을 말한다. 이를 위해서는 완벽한 문맥기반 의미 추론 및 분석 그리고 자연어를 스스로 학습하여 인간과 같이 사고할 수 있는 지능의 출현에 있다. 현재까지는 언어의 통계학적 방법과 미리 축적한 라이브러리 기반의 딥러닝을 수행하고 있지만, 향후에는 언어 습득과 이해를 통한 지식의 아카이브 축적이 인공지능에 의해 처리될 것이다. 이를 빌게이츠는 '컴퓨터에 읽기 가르치는 인공지능'이라고 말하였다[14]. 컴퓨터에 읽기를 가르치는 인공지능을

(그림 6) 외부지식 수집 서버

개발하기 위해서는 (그림 6)과 같은 외부지식 수집 서버로부터 양질의 데이터를 수집하는 것이 요구된다.

현재까지는 언어추론이나 자연어 중심으로 가짜뉴스 찾기 서비스를 개발하는 것뿐만 아니라, 미디어 배포나 뉴스 소비의 프로세스 상에서 발견되는 가짜 유형에 대한 보조적 접근으로 해결하는 것이 보다 정확도를 높일 것으로 여겨진다. 단기적 해결방안은 기술 개발에 박차를 가하는 것이고 가짜뉴스 찾기로 습득한 원천기술을 다양한 다른 플랫폼 등에 적용하여 산업을 발전시키는 것이다.

(그림 7)은 2019년 12월 릴리즈될 페이크 킬러 인공지능 서비스 (Fake Killer AI Service)의

(그림 7) 페이크 킬러 인공지능 서비스 (Fake Killer AI Service)

도식도이다. 이 서비스에는 한 문장 내에서 주어나 목적어의 도치, 한 문장 내에서 능동/수동의 변경, 한 문장 내에서 육하원칙 중 한 요소 변경, 사실에 근거하지 않는 루머 판별 등과 같은 기본적인 가짜뉴스 찾기 요소 기술뿐만 아니라, 저자 평판과 출처 평판 그리고 이미지 자료의 위치와 해상도 등을 종합하여 학습한 AI가 가짜뉴스를 판별한다.

장기적으로는 미디어에 대한 이해 즉 미디어 리터러시에 대한 개념을 코딩할 수 있는 문제 해결방안이 강구되어야 한다.

미디어리터러시의 개념은 1992년 '미디어리터러시 미 지도자회의'에서 "다양한 형태의 커뮤니이션에 접근하고, 분석하고, 평가하고, 발신하는 능력"이라고 정의한 바 있다. 데이비드 버킹엄은 미디어리터러시는 "미디어를 사용하고, 해석하기 해 요구되는 지식, 기술, 능력을 말한다."고 하며, 분석, 평가, 비 성찰을 포함하는 개념으로 리터러시라고 정의한다[15]. 리터러시(Literacy)란 흔히 문해력이라고 정의하고 읽고 쓸 수 있는 능력을 말하는데 장래에 우리나라에서 문해력을 갖는 인공지능을 개발하여 모든 컴퓨터에 문해력을 가르칠 수 있다면 가짜뉴스 찾기 뿐만 아니라, 원천기술 확보와 여타 산업 측면에서 지대한 기여를 할 것으로 사료된다.

5. 결 론

마셜 매클루언은 "미디어는 메시지다.", "미디어는 언어를 기반으로 하며, 인간의 물질 정신 확장이며, 새로운 환경으로의 확장이다."라고 정의한다[16]. 뉴스는 미디어라는 매체를 통해 가치를 전달한다. 이 가치를 통해 경제, 정치, 사회 평판 등 다양한 이해관계가 충돌한다. 이 때문에 가짜뉴스는 그 피해사례가 늘어나고 있고 사회적 난제로 선정되었다.

2019년 6월 19, 구글 검색엔진에서 '가짜뉴스'를 검색하였다. 0.43초만에 약 2260만건의 관련 콘텐츠가 검색된다. 검색어를 한 단계 심화하여, 검색 키워드를 '가짜뉴스 사례'로 분석하면, 약 178만건(0.38초)이 노출된다. 이를 정리하면 가짜뉴스 178만여건의 사례와, 해당 사례로 인한 '가짜뉴스'에 대한 전국민적 관심은 2260만건이 넘는다는 것이다.

이처럼 국가적 비용도 크고 국민적 피로감도 높은 가짜뉴스에 대한 해결방안은 인간의 수작업에 의한 팩트체크(언론인, 언론학자 등을 통해 해당 기사의 진위를 직접 아날로그로 밝히는 과정)와 인공지능에 의한 자동화방법으로 나뉜다.

그 중 필자는 인공지능 기술로 가짜뉴스를 탐지하는 자동화 방법을 수년 간 연구하였고 그 시작은 국가 챌린지 도전이었다.

오늘날 국가 과제 또는 경진 대회는 대학교수와 기업대표 등 전문가 그룹에 의해 정성적인 평가를 통하는 경우가 대부분이었고, 통상 국내 전문가는 그 규모가 작고 인적 네트워크가 첨밀하여 특정 학교나 학회 또는 그들만의 리그 안에 소속되지 않으면, 정보습득부터 비대칭적이고 심지어는 선정과정까지 불투명성이 높다. 필자는 무역학과를 학부와 석사로 공부하고 대기업과 대학교에서 프로그래밍을 가르치며 실무를 익힌 바, 전문가 중심의 인공지능 사업 선정에서는 필자가 제출한 제안서가 연구능력에 대한 의구심과 전공불일치 등의 사유로 번번이 떨어졌었다.

과기정통부 담당자에게 전화를 걸어 전문가에 의한 정성평가(제안서 평가 및 발표 등)가 아닌 ROC커브 기반으로 정량 값으로만 객관적인 평가를 하는지를 수차례 거듭 확인 한 후, 아이와즈 팀(개인 컨소시움, 대학교수라는 간판보다 데

이터를 다루고 처리하는 실력을 들고 경쟁하고 싶어, 신분을 숨기고 개인자격으로 출전)으로 6개월 동안 가짜뉴스 찾기 문제 유형 분석, 150기사의 기계학습, 딥러닝, 사전 지원, 1차 및 2차 경진대회 등을 거쳐 75개 팀중 최종 2위로 장관상을 거머쥐었다. 2018년도는 국가 AI R&D경진대회 수상자만을 모아, 국가공인시험 평가소에서 ROC 커브 면적과 정확도 등에 대한 정량 평가를 수행하였다. 그 결과, 2018년 12월 28일 최종 1등을 하였다.

현재 가짜뉴스를 해결하기 위해서는 문해력을 갖춘 자연어 처리 기술 즉 인공지능 기술의 도약이 요구된다. 연구자는 인공지능에 대한 지식이 일천함에도 건국대학교 언론대학원에서 뉴미디어 이론, 고려대학교에서 미디어 특론, UCC나비와 유비쿼터스 태풍(커퓨니케이션, 2009년) 등 미디어에 대한 강의와 집필로 '가짜뉴스'가 제작되고 유포되는 전반의 과정을 어느 공학자보다 잘 이해하고 있었다.

그러나 실제 가짜뉴스를 찾기 위해 진짜 뉴스를 가져다가 가짜뉴스를 제작하는 과정에서부터 여러 가지 어려움이 있었다.

필자는 소설가 지망생을 뽑아 가짜뉴스를 제작하게 하고 이를 워드2벡과 파이썬 등의 라이브러리를 이용하여 수백번 이상의 반복 실험(학습데이터를 딥러닝한 인공지능을 실제 뉴스 150만건과 250만건에 적용하여 가짜뉴스와 진짜뉴스를 검출)을 통해 그 정확도를 높여갔다.

이 과정에서 ETRI가 개발한 범용사전 등을 재수정하여 정확도를 높였고 동시에 앙상블 모델을 분석하여 최적의 모델링을 토대로 인공지능 알고리즘을 설계하고 이를 개발하였다.

실제 이 과정은 작은 부분에서의 가짜뉴스 문제(제목과 내용의 불일치, 전체 내용에서 거리가 먼 문장을 검출하는 방법) 몇 가지에 대해서는 그 기술력과 데이터 처리 능력을 검증받았으나, 이는 더 많은 새로운 신규 과제와 문제의 첫걸음에 불과하다.

가짜뉴스는 정책입안자 측면에서 정치적 이슈를 다루는 세상에 회자되는 수준의 논쟁으로 그쳐서는 안 된다. 자연어 중 영어는 그 형태소 분석기나 여타 기반 기술이 글로벌 언어인 관계로 발전 속도가 놀랍게 빠르다. 반면 우리나라 언어 즉 국어 기반의 형태소 분석이나 여타 문맥이해 더 나아가 특정 분야(법조문, 판결문, 피의자 조사서 등)에서 사용하는 국어 자체가 갖는 그 복잡성과 비문으로 구성된 1장에 1문장이 들어가는 문서(판결문 등) 등도 존재한다.

해당 분야에 가짜 뉴스 좁게는 사실이 아닌 주장, 주제를 벗어난 문장, 제목과 불일치하는 내용, 숫자의 불일치, 비논리적 표현의 반복 등을 인공지능이 검출해낸다면 사회적 비용과 실질적 효용이 클 것이다.

더 나아가 자연어는 고도의 인문학적 주제이다. 자연어 중 기능어에 한정한 연구도 실효성을 거둘 수 있을 것이다. 가짜뉴스를 검출하는 연구가 비단 가짜 뉴스만을 찾는 기능적 기술 발전에 그치지 않고 언어 속에 함의된 참과 거짓 그 경계를 어떻게 규정하고 분류해내는 것이 한국인의 얼과 대한민국의 민주주의적 가치에 합리적인지에 대한 거대 연구라는 측면에서 접근해야 할 것이다.

그간 연구의 작은 성과들은 실제로 기술적 탁월함(실제 파이썬은 라이브러리가 풍성하고 습득하기 수월한 언어로 구성되었고 텐서플로우는 프로세스를 정의하기에 수월했다)못지 않은 사회과학적 이해 즉 융복합 사고로 코딩했을 때 정확도를 높였다는 선례로 삼아, 공학 분야에서도 인문학과 사회학에 대한 이해가 높아졌으면 한다.

Acknowledgement

※ This work was supported by Institute for Information & communi-cations Technology Promotion(IITP) grant funded by the Korea gov-ernment(MSIP) (No.2018-0-00705, Algorithm Design and Software Modeling for Judge Fake News based on Artificial Intel-ligence).

※ This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2018S1A5A2A03038738 / Algorithm Design & Software Architecture Modeling to Judge Fake News based on Artificial Intelligence).

참 고 문 헌

[1] 윤영석, 엄태원, 안재영, 이현우, 허재두, “페이크뉴스 탐지 기술 동향과 시사점”, ICT 신기술 주간기술동향, 정보통신기술진흥센터, 2017.10, p. 13 [그림1].
[2] https://www.msit.go.kr/web/msipContents/contentsView.do?cateId=mssw315&artId=1348877
[3] https://spri.kr/posts/view/21943?code=column
[4] http://m.news.zum.com/articles/38820518
[5] 심홍진, “가짜뉴스와 민주주의”, KDF Report, 한국민주주의연구소, pp.4-7.
[6] 황용석(2017.3.20.). “페이크뉴스 현상과 인터넷 서비스 사업자 자율규제 현안”. KISO 포럼 정책세미나 발표문. p.4.
[7] Reilly, Ian(2012.). Satirical Fake News and/as American Political Discourse. Journal of American Culture , 35(3), 258-275.
[8] DiFonzo, N. and Bordia, P.(2007.). Rumor, gossip, and urban legend. Diogenes , 54, 19-35.
[9] 미래창조과학부 공고 제2017-0298호, “2017년도 인공지능 R&D 챌린지 대회 공고 안내”, 2017년 6월 29일 미래창조과학부 장관
[10] www.airndchallenge.com
[11] http://gim.unmc.edu/dxtests/roc3.htm
[12] https://en.wikipedia.org/wiki/Receiver_operating_characteristic
[13] http://laverne.libguides.com/fakenews
[14] 이운태, ‘컴퓨터에 익기 가르치는 AI회사’, 동아일보, 2019.06.25., URL: http://www.donga.com/news/article/all/20190625/96177545/1
[15] Douglas Kellner · Jeff Share, Critical Media Literacy is not an option(Learning Inquiry), Volume 1(2007)
[16] 마셜 매클루언, “미디어의 이해 –인간의 확장”, 커뮤니이션북스, 2011, p. 12.

저 자 약 력

강 장 묵

이메일 : kangjm@nsu.ac.kr

- 1996년 국민대학교 무역학과 (경제학사)
- 1999년 고려대학교 무역학과 (경영학석사)
- 2005년 고려대학교 정보보호대학원 (공학박사)
- 1996년~1997년 ㈜쌍용정보통신 컨설팅팀 / 컨설턴트
- 2013년~2017년 고려대학교 컴퓨터학과 연구교수
- 2001년~현재 남서울대학교 빅데이터산업보안학과 / 빅데이터산업보안센터 조교수 및 센터장
- 2017년~현재 과기정통부 국가 R&D AI, 사회현안문제 가짜뉴스 찾기 경진대회 2017년 장관상 수상 및 2018년 최종 1위(2019년 연속과제 단독 선정 및 총 과제비 25억 내외 연구책임자로 수행 중)
- 관심분야 : 자연어처리, 빅데이터, 인공지능, 블록체인, 프라이버시