

Fake News Detection Using CNN-based Sentiment Change Patterns

Tae Won Lee[†] · Ji Su Park^{††} · Jin Gon Shon^{†††}

ABSTRACT

Recently, fake news disguises the form of news content and appears whenever important events occur, causing social confusion. Accordingly, artificial intelligence technology is used as a research to detect fake news. Fake news detection approaches such as automatically recognizing and blocking fake news through natural language processing or detecting social media influencer accounts that spread false information by combining with network causal inference could be implemented through deep learning. However, fake news detection is classified as a difficult problem to solve among many natural language processing fields. Due to the variety of forms and expressions of fake news, the difficulty of feature extraction is high, and there are various limitations, such as that one feature may have different meanings depending on the category to which the news belongs. In this paper, emotional change patterns are presented as an additional identification criterion for detecting fake news. We propose a model with improved performance by applying a convolutional neural network to a fake news data set to perform analysis based on content characteristics and additionally analyze emotional change patterns. Sentimental polarity is calculated for the sentences constituting the news and the result value dependent on the sentence order can be obtained by applying long-term and short-term memory. This is defined as a pattern of emotional change and combined with the content characteristics of news to be used as an independent variable in the proposed model for fake news detection. We train the proposed model and comparison model by deep learning and conduct an experiment using a fake news data set to confirm that emotion change patterns can improve fake news detection performance.

Keywords : Convolutional Neural Networks, Sentiment Change Patterns, Fake News

CNN 기반 감성 변화 패턴을 이용한 가짜뉴스 탐지

이 태 원[†] · 박 지 수^{††} · 손 진 곤^{†††}

요 약

최근 가짜뉴스는 뉴스 콘텐츠 형식을 가장하고 중요한 사건이 발생할 때마다 등장하여 사회적 혼란을 초래한다. 이에 가짜뉴스를 탐지하기 위한 연구로 인공지능 기술이 사용된다. 자연어 처리를 통해 가짜뉴스를 자동으로 인지 및 차단하거나, 네트워크 인과 추론과 결합함으로써 허위 정보를 확산시키는 소셜미디어 인플루언스 계정을 감지하는 등의 가짜뉴스 탐지 접근법이 딥러닝을 통해 구현될 수 있었다. 그러나 가짜뉴스 탐지는 여러 자연어 처리 분야 중에서도 해결이 어려운 문제로 분류된다. 가짜뉴스가 가지는 형식 및 표현의 다양성으로 특성 추출의 난도가 높고, 뉴스가 속한 범주에 따라 하나의 특성이 서로 다른 의미를 가질 수도 있는 등 다양한 한계점이 존재한다. 본 논문에서는 가짜뉴스를 탐지하기 위한 추가적인 식별 기준으로 감성 변화 패턴을 제시한다. 합성곱 신경망을 가짜뉴스 데이터 세트에 적용하여 콘텐츠 특성에 기반한 분석을 수행하고, 감성 변화 패턴을 추가로 분석함으로써 성능이 개선된 모델을 제안한다. 뉴스를 구성하는 문장에 대하여 감성 극성을 산출하고 장단기 메모리를 적용함으로써 문장 순서에 의존적인 결괏값을 얻을 수 있다. 이를 감성 변화의 패턴으로 정의하고 뉴스의 콘텐츠 특성과 결합하여 가짜뉴스 탐지를 위한 제안 모델의 독립변수로 활용한다. 제안 모델과 비교 모델을 딥러닝으로 학습시키고 가짜뉴스 데이터 세트를 이용한 실험을 진행하여 감성 변화 패턴이 가짜뉴스 탐지 성능을 개선할 수 있음을 확인한다.

키워드 : 합성곱 신경망, 감성 변화 패턴, 가짜뉴스

1. 서 론

익명성을 악용한 허위 정보의 유포는 인터넷의 대표적인

역기능 중 하나이다. 특히 가짜뉴스는 인터넷에 산재한 수많은 허위 정보 중에서도 뉴스 형식을 갖추고 있는 날조된 정보를 의미한다. 가짜뉴스는 정치, 경제, 사회, 문화, 스포츠 등 다양한 분야에서 영향력을 확장하고 있으며 지역감정, 지리적 위치에 따른 정당 선호도의 양극화, 성별, 종교, 인종, 세대, 빈부격차 등으로 인한 차별과 혐오는 가짜뉴스가 대중들의 관심을 끌기 쉬워지게 된 원인이 되고 있다[1].

가짜뉴스 문제를 해결하기 위한 노력은 다양한 분야에서 시도되고 있다. 대한민국은 언론중재법, 정보통신망법, 공직

※ 이 논문은 한국방송통신대학교 학술연구비 지원을 받아 작성된 것임.

† 준 회 원 : 한겨레신문(주) 독자서비스국 독자기획부 차장

†† 중 심 회 원 : 전주대학교 컴퓨터공학과 교수

††† 중 심 회 원 : 한국방송통신대학교 컴퓨터과학과 교수

Manuscript Received : January 31, 2023

Accepted : February 3, 2023

* Corresponding Author : Jin Gon Shon(jgshon@knou.ac.kr)

선거법 등 여러 현행 법률에 따라 가짜뉴스를 규제하고 있으며[2], 가짜뉴스 유포 웹사이트의 경영에 불이익을 주기 위한 광고 중단이 글로벌 IT 기업에서 시행되기도 하였다[3]. 최근에는 언론의 고의 및 중과실로 인한 허위 보도에 대하여 피해자에게 발생한 손해액을 배상 청구할 수 있도록 하는 언론중재법 개정안이 논의되기도 하는 등 가짜뉴스 문제를 해결하려는 노력은 계속해서 강화되고 있다. 그러나 가짜뉴스가 확산하는 것을 효과적으로 방지하기 위해서는 가짜뉴스를 정확하게 탐지하는 것이 무엇보다 중요하다.

가짜뉴스 탐지에는 인공지능 연구의 한 분야인 딥러닝 기반의 자연어 처리를 이용할 수 있다. 딥러닝 기반의 가짜뉴스 탐지에 관한 다수의 선행 연구는 뉴스의 콘텐츠 정보를 포함하고 있는 주제목, 부제목, 리드(lead), 본문 등을 주요 분석 대상으로 선정하여 가짜뉴스에서 공통으로 나타나는 특성을 추출하고자 하였다. 그러나 이러한 데이터들은 그 내용의 전체 또는 일부가 쉽게 수정될 수 있고, 이에 따라 다량의 가짜뉴스 데이터를 학습한 딥러닝 모델이라 하더라도 우수한 정확도를 장기간 유지하기 어려울 수 있다.

이러한 문제를 해결하기 위해 본 논문은 합성곱 신경망(Convolutional Neural Network; CNN)을 이용하여 가짜뉴스의 콘텐츠 특성을 분석하고, 장단기 메모리(Long Short-Term Memory; LSTM)를 통해 가짜뉴스의 감성 변화 패턴을 추출한다. 감성 변화 패턴은 문장의 등장 순서에 의존적인 감성 극성의 집합으로서 가짜뉴스 유포자가 가짜뉴스를 작성할 때 보이는 표현 및 작법과도 연관이 있다. 감성 변화 패턴을 추출하기 위하여 뉴스를 문장별로 토큰화(tokenization)하고, 각 토큰에 감성분석을 적용하여 극성을 할당한 후 LSTM을 사용한다. 마지막으로 콘텐츠 특성과 감성 변화 패턴을 결합하여 가짜뉴스 식별에 활용한다.

2. 관련 연구

2.1 딥러닝 기반의 가짜뉴스 탐지

가짜뉴스 탐지를 위한 여러 선행 연구로부터 확인할 수 있듯이 뉴스의 콘텐츠를 구성하고 있는 주제목, 부제목, 리드, 본문 등은 지금 보고 있는 뉴스가 가짜뉴스인지 아닌지 판단할 수 있는 가장 기본적인 통찰력을 제공한다. 그러나 가짜뉴스는 다양한 주제, 범주, 표현, 길이, 작법으로 생산되고 있어 뉴스 콘텐츠만을 기반으로 학습하는 가짜뉴스 탐지 모델은 정확도를 높이는 데에 한계가 있다. 이러한 한계를 극복하기 위한 대안으로 선행 연구에서 제안된 기법들은 뉴스의 콘텐츠 정보를 새로운 관점에서 응용하는 방법과 외부 지식을 활용하는 방법으로 나눌 수 있다.

토픽 모델링을 활용한 가짜뉴스 탐지 기법은 뉴스로부터 10개의 이슈 그룹을 추출함으로써 콘텐츠에 대한 새로운 시각을 제공하였다. 추출된 이슈 그룹과 언론사 메타데이터를 서포트 벡터 머신에 적용하고 학습 및 검증함으로써 공정성

을 확보할 수 있게 하였다[4]. 또한 Fasttext 워드 임베딩 기법과 Shallow-and-wide CNN을 이용하여 뉴스를 구성하는 제목, 본문 간의 일관성을 분석함으로써 클릭 유도를 목적으로 하는 미끼 기사를 탐지하는 기법이 제안되기도 하였다[5]. 이러한 제안 방법들은 가짜뉴스 유포자가 의도적으로 가짜뉴스의 제목, 본문, 메타데이터 등을 수정할 경우, 가짜뉴스 탐지 모델의 분류 정확도가 쉽게 낮아질 수 있다는 한계점이 존재한다.

외부 지식을 활용하는 방법으로는 뉴스의 콘텐츠 정보와 페이스북, 트위터 등의 SNS 데이터를 통합하여 활용하는 가짜뉴스 탐지 모델이 기존 모델과 비교했을 때 성능이 28% 향상됨을 보였고[6], 맥락적 단어 표현, 문장 내 엔티티 정보, 엔티티 정보에 대한 지식 그래프를 활용하고 어텐션을 수행하여 가짜뉴스를 최종 분류하는 기법은 데이터 세트 편향으로 인한 영향력을 완화하여 신뢰성을 높였다[7]. 가짜뉴스 탐지를 위해 외부 지식을 활용하는 접근 방법은 성능 향상의 효과가 입증되었으나 충분한 외부 지식 데이터 확보가 필수적이며, 신뢰할만한 지식 데이터 구축에 많은 시간과 비용을 투자해야 한다는 단점이 있다.

2.2 감성분석

허위 정보는 공격 대상이 되는 독자의 신뢰를 훼손하기 위한 목적으로 과도한 불안감을 부추긴다. 또한 불확실성 관리 이론에 의하면 개인에 의해 지각된 불확실성은 부정적 감정을 형성하고, 추가정보 탐색 의지를 강하게 하는 것으로 확인되었다[8, 9]. 특히 재난 시 소셜미디어에 의한 소문의 유포 및 확산에 관한 연구[10]에 의하면 부정적 감성은 가짜뉴스의 확산과 밀접하게 관련된 것으로 나타났다.

본 연구에서 가짜뉴스 탐지를 위한 접근법 중 하나로 사용하게 될 감성분석은 입력데이터로부터 주관적인 의견을 주요 분석 대상으로 선정하고 긍정 또는 부정의 극성으로 분류하는 자연어 처리 기법이다. 감성을 분석하기 위하여 그동안 주로 사용된 방법은 감성 사전을 구축하고, 단어들의 출현 빈도를 고려하는 것이었다. 또한 가짜뉴스 탐지에 있어서 다양한 소스의 결합이 성능의 향상으로 이어질 수 있다. 발화자, 발화자의 신뢰도, 외부 전문가의 보고서 등 다양한 메타데이터를 결합하거나[11] 뉴스에 삽입된 사진 및 이미지 분석 결과를 결합함으로써[12] 탐지 정확도가 개선될 수 있음을 여러 선행 연구에서 확인하였다.

2.3 SNU 팩트체크

언론사들이 검증한 공적 관심사를 대중들에게 알리기 위한 목적으로 서울대학교 언론정보연구소에서 공개하고 있는 SNU팩트체크는 SNU팩트체크원칙을 준수하는 국내 31개 언론사와 제휴하여 검증을 수행하는 비정치적, 비영리적 공공 정보 서비스이다[13]. SNU팩트체크에 참여하고자 하는 언론사가 제휴를 신청하면 SNU팩트체크센터는 언론사의 팩트체

크 뉴스 제작 현황을 모니터링하고, SNU팩트체크원칙 준수 여부, 활동의 지속성 등을 검토하여 승인을 결정한다. SNU팩트체크 콘텐츠를 등록할 때의 절차는 팩트를 추가하고 검증하는 2단계로 크게 구분된다. 팩트체크를 수행하고자 하는 콘텐츠가 이전에 검증된 적 없는 내용일 때 언론사는 발언 주제, 테마, 유형, 토픽, 헤드라인, 보충 설명, 출처 등의 정보를 신규로 입력하며, 그 결과로 하나의 뱃지(badge)가 생성된다. 추가된 뱃지는 다수의 타 언론사를 통해 검증되며 본문, 평가점수, 검증 기사, 근거자료가 검증 과정에서 추가로 입력된다[14].

뱃지는 판정 결과에 따라 '전혀 사실 아님', '대체로 사실 아님', '절반의 사실', '대체로 사실', '사실', '판단 유보'의 6개 클래스로 분류된다. 또한 하나의 뱃지에 대하여 한 언론사가 '대체로 사실'로 판정한 상황에서 다른 언론사가 '전혀 사실 아님'으로 판정한 경우와 같이 두 개 이상 언론사에 의한 각 판정 결과가 3단계 이상의 차이를 보이면 '논쟁 중'으로 표시될 수 있다.

3. 가짜뉴스 탐지 모델

3.1 가짜뉴스 탐지 모델 설계

본 연구에서 제안하는 가짜뉴스 탐지 모델은 콘텐츠 특성 분석, 감성 변화 패턴 분석, 콘텐츠 특성과 감성 변화 패턴의 결합 등 크게 3가지 요소로 구성되어 있다. 콘텐츠 특성 분석은 가짜뉴스 탐지에 있어서 기본적인 통찰력을 제공하는 부분으로 CNN을 NLP에 최적화한 알고리즘인 1D CNN을 이용하여 콘텐츠 특성을 추출한다. 이 과정은 딥러닝 기반의 NLP를 이용한 선행 연구에서 주로 제안되었던 접근법으로서 다량의 훈련 데이터가 제공되는 조건에서 워드 임베딩을 결합함으로써 일정 수준의 성능을 보인다.

콘텐츠 특성을 추출하기 위해 분석되었던 뉴스는 감성 변화 패턴의 추출을 위하여 다시 한번 활용된다. 뉴스를 구성하고 있는 장문의 텍스트를 문장 단위로 분할하고, 각 문장에 감성분석이 적용된 결괏값의 집합을 LSTM에 입력함으로써 문장 순서에 의존적인 뉴스의 감성 변화 패턴을 추출할 수 있다. 앞서 추출한 콘텐츠 특성과 감성 변화 패턴을 결합하여 가짜뉴스 탐지 성능을 향상하는 것이 본 연구의

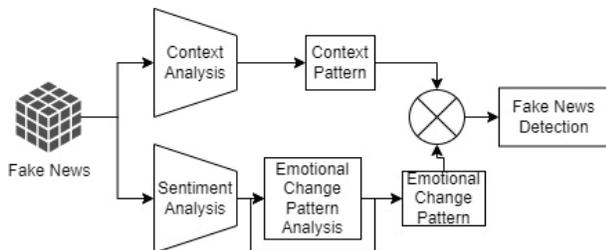


Fig. 1. Fake News Detection Model

Table 1. Examples of Fake News Data

News	Content of the Article	Label
Training Data A	미국, 영국 질병청에 보고된 백신 부작용 중증 사망자가 코로나 사망자 수를 훌쩍 넘기고 있다. 현 상황에서 백신은 답이 아닌 것으로...	Fake
Verification Data B	경찰이 무력으로 백신 강제 접종할 수 있다. 문을 부수는 영상과 함께 한국에서도 경찰력을 동원한 백신 강제 접종이 이뤄질 것이라고...	Fake

핵심이라고 할 수 있으며, 제안하는 가짜뉴스 탐지 모델은 Fig. 1과 같다.

Table 1은 제안하는 가짜뉴스 탐지 모델의 동작 과정을 설명하기 위한 실제 가짜뉴스의 표본이다. A, B 모두 가짜뉴스로 레이블링 되어있으며, 코로나19 백신에 대해 불안감을 조성하고 있다. A를 훈련 데이터, B를 검증 데이터로 하는 테스트에서 가짜뉴스 탐지 모델이 B를 가짜뉴스로 분류하면 정확하게 탐지한 것으로 판정한다.

3.2 콘텐츠 특성 분석

본 논문에서 정의하는 가짜뉴스 탐지는 가짜뉴스 여부가 레이블링 된 훈련 데이터를 학습한 모델에 가짜 여부를 식별하고자 하는 뉴스 데이터를 입력하고, 그 결괏값을 바탕으로 사실 여부를 판단하는 이진 분류 문제이다. 제안하는 감성 변화 패턴의 적용이 가짜뉴스를 정확하게 분류하는 데 있어서 유의미한 성능 향상을 가져올 수 있는지 확인하려면 우선 콘텐츠 특성을 정밀하게 분석해야 한다.

자연어 상태의 훈련 데이터 및 식별 대상 데이터를 입력하기 위하여 컴퓨터에서 처리하기에 적합한 형태로 워드 임베딩을 진행해야 한다. 워드 임베딩은 주변 단어를 바탕으로 중심 단어를 예측하는 CBOW(Continuous Bag of Words) 방식의 Word2Vec를 이용하여 분산 표현하는 것을 의미한다. 본 논문에서는 실험의 편의를 위해 사전 훈련된 Word2Vec을 이용하여 워드 임베딩을 얻는다. 워드 임베딩으로 변환된 뉴스 콘텐츠 데이터는 $x \in \mathbb{R}^{n \times d}$ 로 정의된다. n 은 단일 뉴스 데이터에 포함된 단어의 개수, d 는 워드 임베딩의 차원이다.

가짜뉴스의 특성을 추출하기 위해 딥러닝 알고리즘의 하나인 합성곱 신경망을 사용하며, 텍스트를 대상으로 특성을 추출할 수 있는 1D CNN을 사용한다. 합성곱 연산을 위한 커널은 $k^m \in \mathbb{R}^{l \times d}$ 로 정의한다. l 은 커널의 길이를 의미하며 m 개의 멀티 커널을 적용한다.

$$Y(i, j) = (X \times W)(i, j) = \sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} X(i+k_1, j+k_2) W(k_1, k_2) \quad (1)$$

Equation (1)에 따라 x 와 k 의 합성곱을 이용함으로써 커

Table 2. Content Information of Fake News

News	Content Information
Training Data A	Vaccine Side Effects Deaths > Corona Deaths
Verification Data B	Compulsory vaccination is possible

널의 수용영역에 해당하는 단어의 로컬 특성을 반영하는 f 를 얻을 수 있으며, 수식 표현은 Equation (2)와 같다. 결과적으로 단어의 로컬 특성은 $f^m \in \mathbb{R}^{n-l+1}$ 형식의 행렬이 된다.

$$f = \sum_{i=1}^l \sum_{j=1}^d x(i,j)k(i,j) \quad (2)$$

Equation (3)의 t_i 은 커널의 수용영역에 해당하여 합성곱 연산의 대상이 되는 텍스트 시퀀스이다. '미국 영국 백신 부작용 사망 보고가 코로나 사망자보다 많다'라는 문장에서 $i = 3, l = 3$ 일 때 합성곱 연산 대상 t_i 는 '백신, '부작용' '사망'이 된다. t_i 에 합성곱 연산을 수행한 Equation (4)의 결과에 따라 커널 수 m 개에 해당하는 텍스트 로컬 특성 f_i^m 을 얻을 수 있다.

$$t_i = \{x_i, x_{i+1}, \dots, x_{i+l-1}\} \quad (3)$$

$$f_i^m = t_i * k + b \quad (4)$$

신경망의 연산을 효율화하고 과적합(overfitting)을 억제하기 위하여 f_i^m 에 풀링(pooling)을 적용한 후, 마지막으로 이진 분류를 위한 시그모이드(sigmoid) 활성화 함수를 통해 가짜뉴스 여부를 판단한다. 예를 들어 Table 1의 뉴스 A, B에 1D CNN을 적용한 결과값은 Table 2와 같은 콘텐츠 정보를 포함하고 있을 수 있다. 그러나 콘텐츠 특성에 기반한 가짜뉴스 탐지 모델이 A가 가짜뉴스임을 학습했어도 B가 가짜뉴스임을 예측할 수 있는 근거는 여전히 충분하지 않으며, 탐지 정확도 향상에 한계가 존재하게 된다.

3.3 감성 변화 패턴 분석

뉴스를 구성하는 장문의 텍스트를 대상으로 감성분석을 수행하여 긍정 또는 부정의 극성을 판단할 수 있다. 그러나 본 연구에서는 뉴스로부터 감성 변화 패턴을 추출하기 위하여 가장 먼저 가짜뉴스 입력데이터 $x \in \mathbb{R}^{n \times d}$ 을 문장 단위로 토큰화한다. 각 뉴스는 문장 토큰의 집합으로 정의되며, 각 문장 토큰에 사전 학습된 감성분석 모델을 적용함으로써 감성분석 결과에 따라 긍정과 부정으로 분류된 극성 집합 $s \in \mathbb{R}^{o \times e}$ 를 구할 수 있다.

Table 3. Emotion Change Patterns of Fake News

News	Emotional Change Patterns
Training Data A	negative → negative → negative → negative → negative → negative → negative
Verification Data B	negative → negative → negative → negative → negative → negative → negative

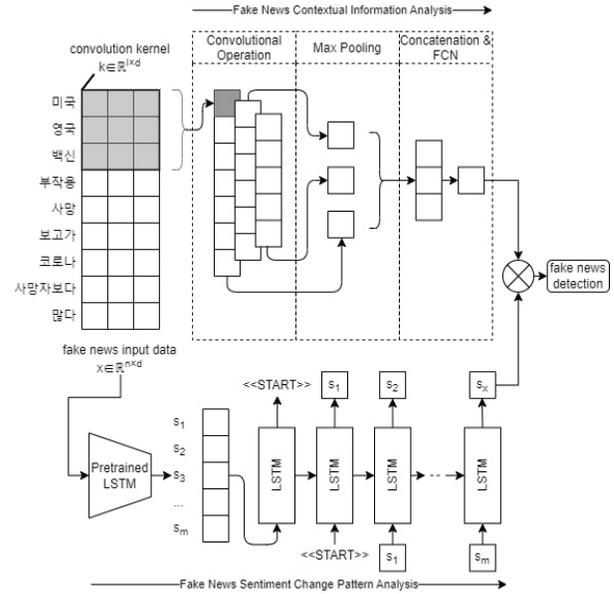


Fig. 2. Sentiment Change Pattern for Improving Fake News Detection Performance

극성 집합은 문장이 등장하는 순서에 의존적이며, 극성 집합을 구성하는 원소는 긍정 또는 부정의 감성을 나타내고 있으므로 '뉴스가 긍정적인 감성의 문장으로 시작해서 부정적인 감성의 문장이 여러 번 등장했다가 마지막은 긍정적인 감성의 문장으로 마무리되었다'라는 식의 감성 변화 패턴을 표현하는 것이 가능하다. Table 3은 가짜뉴스 A, B 모두가 부정적 감성을 일관적으로 나타내고 있음을 보여주는 감성 변화 패턴의 예이다.

가짜뉴스 탐지 모델을 학습시키기 위하여 감성 변화 패턴을 순서, 시간 등의 데이터를 처리하는 모델의 학습에 강점을 보이는 LSTM에 입력한다. 감성 변화 패턴에 LSTM을 적용한 결과값과 뉴스의 콘텐츠 특성을 결합하여 단일 벡터로 변환한다. 가짜뉴스 탐지 성능을 개선하기 위하여 뉴스의 콘텐츠 특성에 감성 변화 패턴을 추가로 적용하는 전체 과정은 Fig. 2와 같다.

3.4 콘텐츠 특성과 감성 변화 패턴의 결합

순전파(forward propagation) 및 역전파(back propagation)를 수행하는 과정에서 가중치와 편향을 업데이트하기 위하여 뉴스의 콘텐츠 특성과 감성 변화 패턴의 분석 결과값

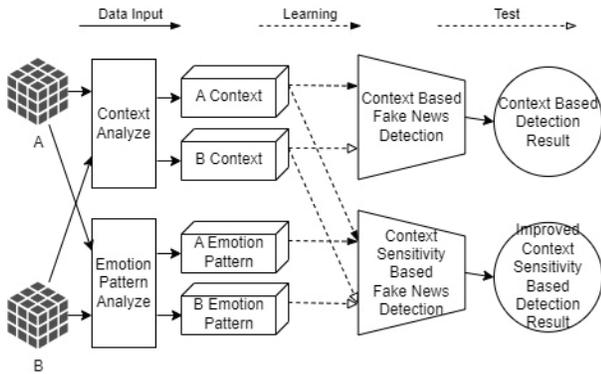


Fig. 3. Combination of Content Information and Emotional Change Patterns

을 단일 벡터(single vector)로 결합하고, 이를 가짜뉴스의 특성으로 정의하여 분석 대상으로 주어진 뉴스가 가짜뉴스일 확률을 계산한다.

Table 1의 뉴스 A, B는 코로나19 바이러스 및 백신에 관한 가짜뉴스이다. 그러나 두 가짜뉴스 모두 코로나19 바이러스와 관련되어 있다는 점을 제외하면 특성의 유사성을 찾기 어렵다. 예를 들어 NLP를 통해 가짜뉴스 A, B로부터 Table 2의 콘텐츠 정보를 성공적으로 추출했다고 가정했을 때, 다량의 가짜뉴스 훈련 데이터를 충분히 학습한 신뢰성 있는 모델이 A 또한 가짜뉴스임을 학습했다고 하더라도 B를 가짜뉴스로 분류하기에는 여전히 근거가 미흡할 수 있다.

반면 콘텐츠 특성 외에 감성 변화 패턴을 추가로 분석함으로써 A, B 모두 코로나19 백신에 부정적인 뉴스임을 알 수 있다. 또한 Table 3에서 확인할 수 있듯이 감성 변화 패턴이 일관성 있게 부정적인 것을 고려하였을 때 부정적 감성을 의도적으로 형성하고 있음을 A, B의 공통적인 특성으로 추출할 수 있다.

이처럼 감성 변화 패턴을 콘텐츠 특성과 결합하는 제안 모델은 가짜뉴스 간의 추가적 유사성을 얻을 수 있게 한다. 이로 인해 훈련 데이터 A가 가짜뉴스임을 학습한 가짜뉴스 탐지 모델이 데이터 B를 검증하는 과정에서 가짜뉴스로 분류할 수 있는 근거가 좀 더 명확해질 수 있고, 가짜뉴스 탐지 정확도가 향상되는 결과로 이어질 수 있다. 가짜뉴스 탐지 성능을 개선하기 위하여 콘텐츠 특성과 감성 변화 패턴을 결합하는 과정은 Fig. 3과 같다.

4. 실험 및 성능 평가

4.1 실험 데이터 세트

제안 모델의 검증을 위하여 가짜뉴스 여부가 확인된 다량의 뉴스 데이터가 필요하다. 그러나 가짜뉴스는 공개된 데이터의 수가 많지 않고, 주류 언론사의 경우 가짜뉴스로 인한 문제가 제기되는 시점에서 기사를 삭제하고 빠른 정정보도로 대응하고 있어 수집에 어려움이 따른다. 대안으로 본 논문

Table 4. Fake News Data Structure

Talker	Naver blog users
Title	경찰이 무력으로 백신 강제 접종할 수 있다.
Content	지난 9월 10일 한 블로그에 방호복을 입은 사람들이 문을 부수는 영상과 함께 한국에서도 경찰력을 동원한 백신 강제 접종이 이뤄질 것이라고 주장하는 글이 올라왔다. 해당 게시물은 응급의료에 관한 법률 개정안, 경찰관 직 무집행법 개정안이 통과되면 경찰이 가정을 찾아와 백신을 강제로 접종할 것이라고 주장했다. 작성자의 말대로 경찰을 동원한 백신 강제 접종이 가능해질까.
Lable	not true at all

에서는 실험을 위한 훈련 데이터로서 다수의 전문가가 사실 확인을 제안하고, 객관적인 자료를 바탕으로 검증할 수 있는 SNU팩트체크를 선정하였다.

실험은 2017년 3월 29일부터 2021년 10월 15일까지 SNU팩트체크에서 수집한 3,002개의 데이터 세트를 대상으로 하였다. 데이터는 발화자, 제목, 내용, 레이블로 구성되어 있으며 레이블은 '사실, 대체로 사실, 절반의 사실, 대체로 사실 아님, 전혀 사실 아님, 판단 유보'의 6가지 클래스로 분류된다. Table 4은 백신 강제 접종에 대한 가짜뉴스의 데이터 구조이다.

본 연구는 가짜뉴스와 가짜가 아닌 뉴스로 이진 분류하는 것이 목적임을 고려하여 '대체로 사실 아님, 전혀 사실 아님'으로 레이블링 된 데이터를 가짜뉴스로, 그 외는 가짜가 아닌 뉴스로 정의하였다. 최종 분석 대상 데이터는 보충 설명이 비어있는 데이터를 제외한 2,658개의 데이터이다. 이 중 가짜뉴스는 1,668개, 가짜가 아닌 뉴스는 990개로 구성되어 있어, 전체 데이터에서 가짜뉴스와 가짜가 아닌 뉴스가 차지하는 비율은 각각 63%와 37%이다. SNU팩트체크 데이터의 상당수가 가짜뉴스임을 의미하며, 이처럼 분포가 불균형한 데이터일수록 훈련 데이터와 테스트 데이터를 분리할 때 각 레이블의 분포를 균일하게 유지하지 않으면 좋은 성능을 기대할 수 없으므로 훈련 데이터와 테스트 데이터를 나눌 때 가짜뉴스와 가짜가 아닌 뉴스의 수를 각각 1,334개/792개, 334개/198개로 구성하여 같은 비율이 유지되게 하였다. 소셜미디어, SNS, 소셜미디어 이용자 등 서로 다른 단어가 같은 의미를 가진 경우 단일화하였다.

4.2 실험

1) 데이터 세트 전처리

전처리 과정으로 정규 표현식을 이용하여 분석 대상이 아닌 불필요한 문장 부호, 특수 문자를 제거하고 형태소 분석기를 이용하여 토큰화한다. 형태소 분석기는 Python 패키지 중 한국어 정보처리를 위한 KoNLPy에서 제공하는 Okt(Open Korean Text) 형태소 분석기를 사용한다[15]. 조사, 접속사

Table 5. Preprocessing of Fake News

Contents Information	MorphologicTokenization	['경찰', '무력', '백신', '강제', '접종', '블로그', '방호복', '입다', '사람', '영상', '한국', '경찰', '동원', '백신', '강제', '접종', '이뤄지다', '해당', '게시', '응급의료', '법률', '개정안', '경찰관', '직무', '집행', '개정안', '통과', '되다', '경찰', '가정', '찾아오다', '백신', '강제', '접종', '주장', '작성자', '경찰', '동원', '백신', '강제', '접종', '가능']
	Morphological Encoding	[232, 2791, 20, 524, 66, 1189, 2920, 741, 72, 226, 6, 232, 1144, 20, 524, 66, 875, 80, 121, 14440, 630, 282, 4929, 1364, 911, 282, 357, 2, 232, 2179, 4438, 20, 524, 66, 4, 1522, 232, 1144, 20, 524, 66, 70]
	Morphological Padding	[0, 0, ..., 0, 232, 2791, 20, 524, 66, 1189, 2920, 741, 72, 226, 6, 232, 1144, 20, 524, 66, 875, 80, 121, 14440, 630, 282, 4929, 1364, 911, 282, 357, 2, 232, 2179, 4438, 20, 524, 66, 4, 1522, 232, 1144, 20, 524, 66, 70]
Emotion Change Pattern	Sentence Tokenization	['경찰이 무력으로 백신 강제 접종할 수 있다.', '지난 9월 10일 한 블로그에 방호복을 입은 사람들이 문을 부수는 영상과 함께 한국에서도 경찰력을 동원한 백신 강제 접종이 이뤄질 것이라고 주장하는 글이 올라왔다.', '해당 게시물은 응급의료에 관한 법률 개정안, 경찰관 직무집행법 개정안이 통과되면 경찰이 가정을 찾아와 백신을 강제로 접종할 것이라고 주장했다.', '작성자의 말대로 경찰을 동원한 백신 강제 접종이 가능해질까?']
	Polarity	[2, 2, 2, 2]
	Polar Padding	[0, 0, 0, 0, 0, 0, 2, 2, 2]
	Talker Encoding	[71]
	Talker Padding패딩	[0, 71]

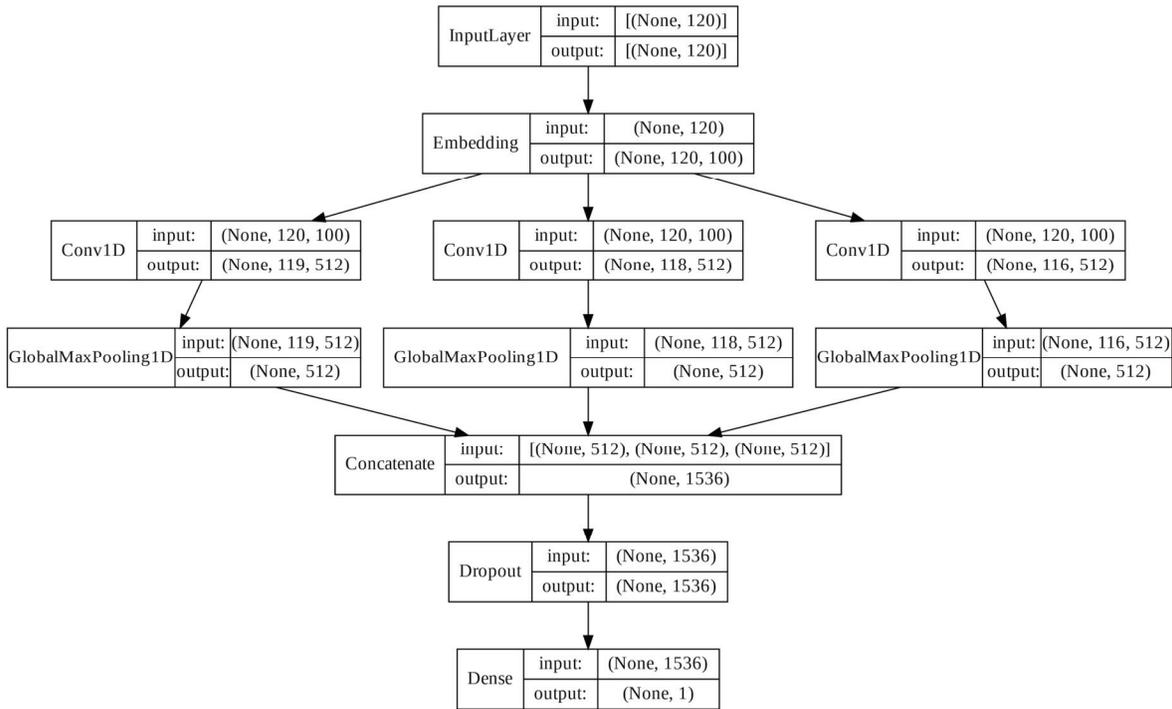


Fig. 4. Fake News Detection Model based on Content Characteristics

등의 불필요한 단어는 불용어 사전에 등록하여 형태소 분석 과정에서 제거될 수 있도록 하고 고유명사, 전문용어, 외래어, 신조어 등 사전에 존재하지 않아 형태소 분석 결과를 부자연스럽게 하는 단어가 존재하면 사용자 정의 사전에 등록 후 형태소 분석을 다시 수행한다. 형태소 분석 결과는 최종적으로 정수 인코딩을 통해 컴퓨터에서 처리하기에 적합한 정수

시퀀스 형태로 변환된다. 문장 토큰화는 Korean Sentence Splitter(KSS)[16]를 이용하였다. 정수 시퀀스의 길이는 행렬 곱을 이용한 병렬 연산이 가능하도록 모든 정수 시퀀스의 길이를 똑같이 맞춰준다. Table 5는 Table 4의 가짜뉴스를 전처리한 결과이다.

토큰화된 형태소는 모델의 임베딩 층(embedding layer)

에서 Word2Vec를 이용한 100차원의 연속 표현(continuous representation) 벡터로 변환한다. 실험에 사용된 Word2Vec은 한국어 위키백과 데이터베이스[17]를 기반으로 사전 훈련하였다.

2) 콘텐츠 정보에 기반한 분석

가짜뉴스는 인코딩과 패딩을 거쳐 컴퓨터를 이용한 병렬 연산에 적합한 형태인 2D 정수 텐서 형식이 되며, 가짜뉴스 하나는 120차원의 정수 집합이 된다.

2D 정수 텐서 형식으로 변환된 가짜뉴스 데이터를 입력받은 임베딩 층은 데이터에 포함된 단어를 밀집 표현(dense representation)으로 변환한다. 정수 인코딩된 단어는 사전 훈련된 Word2Vec을 통해 100차원의 밀집 벡터(dense vector)로 변환되며, 결과적으로 임베딩 층은 가짜뉴스 데이터의 수, 가짜뉴스 데이터의 차원 120, 워드 임베딩 차원 100의 크기를 갖는 3D 텐서를 출력한다.

합성곱 신경망에서는 입력데이터 및 커널의 길이와 스트라이드, 패딩이 주어졌을 때 출력 결과의 크기를 알 수 있으므로[18], 1D CNN 층에서 입력받은 3D 텐서에 대하여 2, 3, 5 크기의 다중 커널에 합성곱 연산을 수행했을 때 커널의 크기 별로 119, 118, 116 길이의 결과가 출력되는 것을 알 수 있다. 각 출력 결과는 커널 개수에 해당하는 512개씩 생성되며 GlobalMaxPooling1D를 통해 가장 큰 벡터를 선정하여 반환한다.

다중 커널을 이용한 1D CNN의 출력값들을 1,536 길이의 단일 벡터로 결합하고 dropout을 50%로 설정하여 학습 과정에서 무작위로 절반의 뉴런을 사용하지 않도록 한다. dropout은 특정 뉴런 간의 연결에 과도하게 의존함으로써 과적합이 일어나는 문제를 완화한다[19]. 마지막으로 이진 분류를 위하여 sigmoid가 활성화 함수로 설정된 전결합층(Fully-Connected Layer, FC)을 추가한다. <Fig. 4>은 콘텐츠 특성 기반 가짜뉴스 탐지 모델의 구조이다.

3) 감성 변화 패턴에 기반한 분석

뉴스의 콘텐츠 특성을 기반으로 하는 모델에 감성 변화 패턴을 추가로 결합하기 위한 첫 번째 단계로 문장 단위의 극성 집합을 입력받는다. 극성 집합은 10의 길이를 갖는 1차원 배열이면서 가짜뉴스 데이터의 개수만큼 존재하므로 입력데이터는 2D 정수 텐서의 형식이 된다. 또한 LSTM은

입력데이터로 3차원 형식의 데이터가 필요하므로 임베딩 층을 배치하여 극성 집합을 가짜뉴스 데이터의 개수, 극성 집합의 길이, 극성의 차원으로 구성된 3차원 텐서로 변환한다. 이 중 극성 집합의 길이는 LSTM에서 시점(time step)을 의미하는데 각 시점의 은닉 상태(hidden state) 및 셀 상태(cell state)를 재귀적으로 전달한다. 재귀적 전달의 마지막 시점에서 출력되는 결과값은 극성 집합의 순서에 의존적인 특성으로서 본 논문에서는 이를 감성 변화 패턴으로 정의하고 있다.

감성 변화 패턴과 긴밀하게 관련된 요소인 발화자 정보 또한 추가로 입력받는다. 단일 뉴스에 대하여 그 발화자의 수가 많지 않음을 고려하여 발화자 정보는 2의 길이를 가지며, 임베딩 층을 통해 가짜뉴스 데이터의 개수, 발화자의 길이 2, 워드 임베딩 차원 100의 크기를 갖는 3D 텐서로 출력한다. 또한 1D CNN 층에서는 크기가 2인 커널을 32개 사용하고, GlobalMaxPooling1D 및 30%의 dropout을 거쳐 발화자의 특성을 추출한다. 이어서 concatenate 층에서는 앞 단계에서 수행했던 콘텐츠 및 발화자 특성의 분석, 감성 변화 패턴의 추출 등의 결과를 결합한다. 결과적으로 콘텐츠 특성만을 분석했을 때보다 층이 많아진 것과 관련하여 FC 및 배치 정규화(Batch Normalization) 층을 추가로 배치한다. 배치 정규화는 기울기 소실 및 폭주를 완화하고, dropout을 생략할 수 있게 하여 모델의 학습 속도를 빠르게 한다[20]. 마지막으로 입력데이터가 가짜뉴스인지 아닌지를 판단하는 이진 분류를 위하여 sigmoid가 활성화 함수로 설정된 FC를 추가한다. Fig. 5는 감성 변화 패턴 분석이 추가된 가짜뉴스 탐지 모델의 구조를 나타낸다.

본 논문에서 해결하고자 하는 가짜뉴스 탐지 문제는 이진 분류 문제에 해당하므로 손실 함수는 이진 분류에 적합한 Binary Cross Entropy를 사용한다. 또한 최적값을 찾아가는 방향과 학습률 두 가지를 모두 고려하면서 가중치를 업데이트해나갈 수 있도록 아담(adam) 옵티마이저를 적용하였다. 배치 크기는 64로 설정하였다.

4.3 성능 평가

CNN을 이용한 콘텐츠 기반 분석 모델과 LSTM을 이용하여 감성 변화 패턴을 추가로 분석한 모델을 각각 구현하고 실험을 진행하였다. 제안 모델 및 비교 대상 모델의 성능을 평가하기 위한 기준으로 기계학습 모델의 성능을 측정하는 대표적인 지표인 정밀도(Precision), 재현율(Recall), 정확도(Accuracy)를 참고하였으며, Precision과 Recall을 균형적으로 반영하는 F1-score를 주요 검증지표로 사용하였다. F1-score를 주요 검증지표로 선정한 것은 F1-score가 불균형 데이터 구조에서 모델의 성능을 정확하게 평가할 수 있고, 훈련 및 테스트 데이터에서 가짜뉴스가 가짜가 아닌 뉴스보다 상대적으로 많은 상황을 고려하였다.

Table 6은 콘텐츠 기반 분석 모델과 감성 변화 패턴을 추가로 분석한 모델의 가짜뉴스 분류 성능을 비교한 결과이다. 같은 데이터와 하이퍼 파라미터 조건에서 모델별로 학습을 진행했을 때 감성 변화 패턴을 적용한 제안 모델이 F1-score에서 0.74를 보여 비교 모델의 0.68보다 8.8% 이상 좋은 성능을 보이는 것으로 나타났다. Table 7는 F1-score를 비롯하여 Accuracy, Precision, Recall을 기준으로 비교 대상 모델과 제안 모델의 성능을 종합적으로 평가한 결과이다. 실험 결과 제안 모델이 Precision에서는 0.66으로 비교 모델의 0.72보다 다소 낮은 결과를 보였고, Recall에서는 0.84로 비교 모델의 0.65보다 29.2% 이상 높은 성능을 보였다.

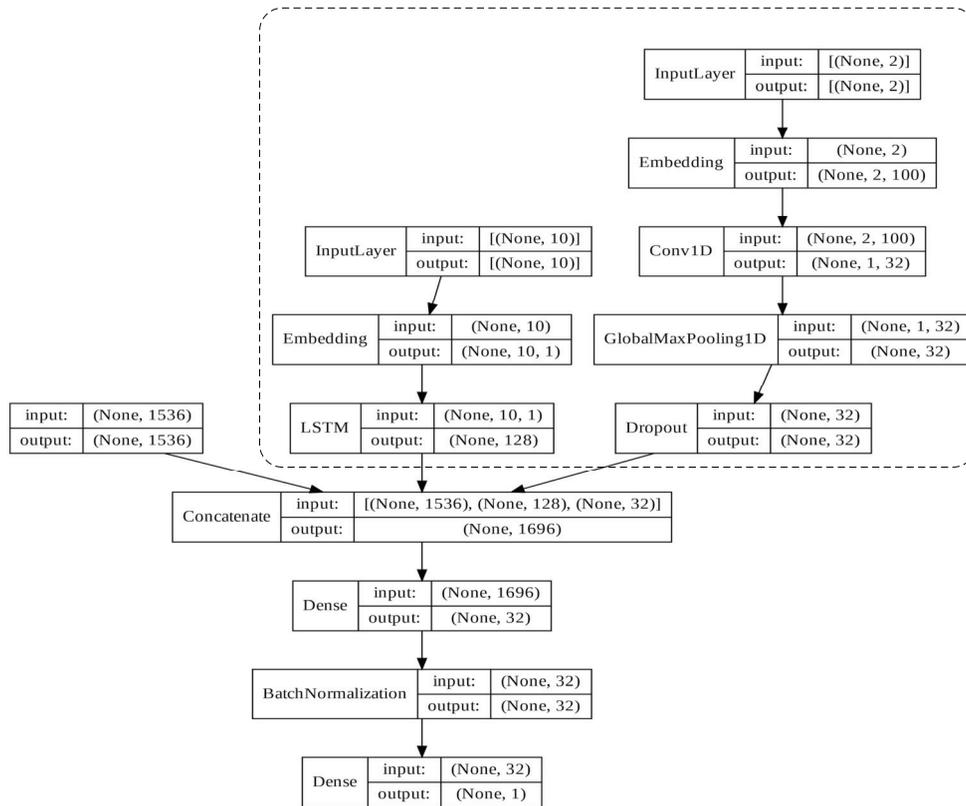


Fig. 5. Structure of Fake News Detection Model

Table 6. Comparison Model F1-score

Model	Embedding	Hyper Parameters				F1-score
		Batch	Filters	Kernels	Dropout	
Content	Word2Vec	64	512	[2, 3, 5]	0.5	0.68
Content + Sentiment	Word2Vec	64	512	[2, 3, 5]	0.5	0.74

Table 7. Comparison Model Performance Evaluation

Model	Accuracy	Precision	Recall	F1-score
Content	0.62	0.72	0.65	0.68
Content + Sentiment	0.63	0.66	0.84	0.74

본 논문의 실험에서 Precision은 테스트 단계에서 가짜뉴스로 레이블링 된 데이터를 가짜뉴스로 정확하게 분류한 비율, Recall은 가짜뉴스인 테스트 데이터 전체에서 가짜뉴스로 정확하게 분류된 비율을 의미한다. 제안 모델이 비교 대상 모델보다 Precision이 하락했지만, Recall에서는 큰 폭의 성능 개선을 보여 결과적으로 Precision과 Recall의 조화평균인 F1-score가 향상되었다.

Recall이 증가한 것은 입력데이터가 가짜뉴스로 판정되는 임계값이 낮아진 것을 의미하며, 가짜뉴스로 레이블링 된 데이터가 가짜가 아닌 뉴스로 잘못 분류되는 경우가 줄어들었다는 것을 의미한다. 가짜뉴스를 진짜라고 생각하는 뉴스 수

용자가 많아지면서 현실의 문제가 발생하는 것임을 고려했을 때, 가짜뉴스가 사실 확인된 뉴스로 잘못 분류되는 경우를 줄이는 것에 집중함으로써 현실 문제의 해결에 도움이 될 수 있다. 따라서 본 논문에서 제안하는 모델의 적용 및 실험의 결과로 확인할 수 있는 Recall의 증가는 가짜뉴스 근절이라는 목적에 부합하는 것으로 볼 수 있다.

또한, 사실 확인된 뉴스를 가짜뉴스가 아니라고 분류하는 것도 분류를 정확하게 수행한 것으로 볼 수 있다. 이 경우까지 고려한 Accuracy는 두 모델이 비슷한 결과를 보였다.

5. 결 론

본 논문에서는 뉴스를 구성하는 텍스트로부터 정량화된 콘텐츠 특성을 추출하고 활용하는 전통적 가짜뉴스 탐지 접근법과 비교하여 탐지 성능 면에서 개선된 감성 변화 패턴 기반의 가짜뉴스 탐지 모델을 제시하였다.

뉴스를 문장 단위로 토큰화하고, 각 문장 토큰을 대상으로

감성분석을 수행하여 얻은 감성 집합에 순환신경망을 적용한 결괏값을 감성 변화 패턴으로 정의하였으며 콘텐츠 특성과 감성 변화 패턴을 단일 벡터로 결합하여 가짜뉴스 탐지에 활용하였다. 콘텐츠 특성을 분석하기 위해서는 CNN을 자연어 처리에 최적화한 1D CNN을 이용하였고, 감성 변화 패턴을 분석하는 방법으로는 순서 의존적 특성 추출에 적합한 순환 신경망 LSTM을 활용하였다.

SNU팩트체크 데이터 세트를 이용한 비교 실험에서 감성 변화 패턴을 추가로 활용한 제안 모델이 콘텐츠 기반의 전통적 가짜뉴스 탐지 모델과 비교했을 때 성능 면에서 유의미한 개선이 있는 것으로 나타났다. 제안 모델은 감성 변화 패턴을 적용하기 전과 비교하여 재현율이 29.2%, F1 score가 8.8% 향상된 것으로 확인되었으며, 재현율과 F1 score의 향상은 감성 변화 패턴의 적용이 가짜뉴스 탐지 모델의 임계값을 낮추는 방향으로 영향을 주고 있음을 의미한다. 사실이 확인된 뉴스를 가짜뉴스라고 잘못 예측했을 때보다 가짜뉴스를 진짜라고 판단하는 경우의 위험이 상대적으로 큰 현실을 고려했을 때 감성 변화 패턴을 적용한 제안 모델은 가짜뉴스 근절이라는 목적에 한층 더 부합한다고 볼 수 있다.

본 연구의 의의는 다음과 같다. 첫째, 가짜뉴스를 탐지하기 위한 추가적인 통찰력을 제공하는 방법으로 감성 변화 패턴이라는 특성을 제시하였다. 감성 변화 패턴은 뉴스 분석에 감성분석을 활용하는 기존 접근법과 차별화되는 부분으로서 뉴스를 긍정 또는 부정의 극성으로 단순 분류하지 않고, 감성이 문장 단위에 따라 어떻게 변화하는지를 분석한다. 둘째, 가짜뉴스가 갖는 형식과 일탈적 내용이 뉴스 수용자에게 부정적 감성을 비롯하여 여러 가지 감성을 형성하는 것으로 확인되는 미디어 및 커뮤니케이션 분야의 가짜뉴스 선형 연구를 정보과학의 관점에서 실험을 통해 입증하였다. 특히 감성 변화 패턴은 가짜뉴스 탐지 성능과 연관이 있으며, 탐지 성능을 향상하는 데 활용될 수 있는 것으로 나타났다.

본 연구는 가짜뉴스 연구자 및 언론 전문가들에게 다음과 같은 시사점을 제공한다. 딥러닝에서 서로 다른 특성을 결합하는 것이 언제나 모델의 성능 향상으로 이어지는 것은 아니며 특성 간 상성에 따라 성능이 유지되거나 반대로 하락하기도 한다. 반면에 감성 변화 패턴은 콘텐츠 특성과 결합함으로써 성능을 향상하는 것으로 확인되었다. 이는 제안 모델을 실제 적용하는 데 있어서 콘텐츠 특성에 기반한 모델을 이미 운영하고 있는 경우에도 발생 가능한 성능 하락의 위험을 감소할 필요가 없음을 의미한다.

References

- [1] David M. J. Lazer et al., "The science of fake news," *Science*, Vol.359, No.6380, pp.1094-1096, 2018.
- [2] Y.-s. Hwang and O.-s. Kwon, "A study on the conceptualization and regulation of fake news: Focusing on self-regulation of internet service providers," *Media and Law*, Vol.16, No.1, pp.53-101, 2017.
- [3] N. Wingfield, M. Isaac, and K. Benner, "Google and Facebook take aim at fake news sites," *nytimes.com*. <https://www.nytimes.com/2016/11/15/technology/google-will-ban-websites-that-host-fake-news-from-using-its-ad-service.html> (accessed Nov. 30, 2021).
- [4] T.-W. Yun and H.-C. Ahn, "Prediction of domestic fake news using text mining and machine learning," *Journal of Information Technology Applications & Management*, Vol.25, No.1, pp.19-32, 2018.
- [5] D.-H. Lee, Y.-R. Kim, H.-J. Kim, S.-M. Park, and Y.-J. Yang, "Fake news detection using deep learning," *Journal of Information Processing Systems*, Vol.15, No.5, pp.1119-1130, 2019.
- [6] Y. Hyun and N. Kim, "Text-based fake news detection methodology using news and social data," *Journal of the Korean Society of Electronic Transactions*, Vol.23, No.4, pp.19-39, 2018.
- [7] J.-w. Cho, E.-b. Kim, H.-m. Kim, and M.-a. Son, "Development of attention network with added entity for fake news detection," *Proceedings of the Fall Conference of the Korean Industrial Engineering Society*, pp.2751-2755, 2019.
- [8] Dale E. Brashers, "Communication and uncertainty management," *Journal of Communication*, Vol.51, No.3, pp.477-497, 2001.
- [9] C. Choi, "The effect of fact-checking efforts on fake news efforts: Based on uncertainty management theory," *Communication Theory*, Vol.17, No.1, pp.5-53, 2021.
- [10] F. Liu, A. Burton-Jones, and D. Xu, "Rumors on Social Media in disasters: Extending Transmission to Retransmission," *Pacific Asia Conference on Information Systems*, pp.49, 2014.
- [11] H. Karimi, P. C. Roy, S. Saba-Sadiya, and J. Tang, "Multi-source multi-class fake news detection," *Proceedings of the 27th International Conference on Computational Linguistics*, pp.1546-1557, 2018.
- [12] Y. Yang et al., "TI-CNN: Convolutional neural networks for fake news detection," *arXiv:1806.00749*, 2018.
- [13] Seoul National University Press Information Research Institute SNU Fact Check Center, "SNU Fact Check," [Internet], <https://factcheck.snu.ac.kr> (downloaded on November 30, 2021).
- [14] S.-w. Oh and G.-h. Hwang, "An exploratory study on the requirements for the formal composition of facts: Suggestions through SNU FactCheck metadata analysis," *Press Information Research*, Vol.55, No.4, pp.54-98, 2018.
- [15] Open source Korean language processor. Yoo Ho-hyun. Downloaded on 30 November 2021. [Internet], <https://github.com/open-korean-text/open-korean-text>

[16] Korean Sentence Splitter. Sang-Kil Park. Accessed: Nov. 30, 2021. [Internet], <https://github.com/likejazz/korean-sentence-splitter>

[17] Wikipedia, The Free Encyclopedia and s.v. "Wikipedia: Database download," 2021. Distributed by Wikimedia. [Internet], <https://dumps.wikimedia.org/kowiki/latest/kowiki-latest-pages-articles.xml.bz2>

[18] Wikipedia, The Free Encyclopedia and s.v. "Convolutional neural network," [Internet], https://en.wikipedia.org/wiki/Convolutional_neural_network (accessed Nov. 30, 2021).

[19] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever and Ruslan Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, Vol.15, No.1, pp.1929-1958, 2014.

[20] Sergey Ioffe and Christian Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *Proceedings of the 32nd International Conference on Machine Learning*, Vol.37, pp.448-456, 2015.



이 태 원

<https://orcid.org/0000-0002-1675-7329>
 e-mail : hep@knou.ac.kr
 2013년 한국방송통신대학교 컴퓨터학과 (학사)
 2022년 한국방송통신대학교 정보과학과 (석사)

2010년 ~ 현 재 한겨레신문(주) 독자서비스국 독자기획부 차장
 관심분야: 뉴스, 빅데이터, 데이터 분석, 자연어 처리, 인공지능, 딥러닝



박 지 수

<https://orcid.org/0000-0001-9003-1131>
 e-mail : jisupark@jj.ac.kr
 2013년 고려대학교 컴퓨터교육과(박사)
 2015년 ~ 2018년 충남대학교 초빙교수
 2020년 ~ 현 재 전주대학교 컴퓨터공학과 교수

2020년 ~ 2021년 한국정보처리학회 이사
 2021년 ~ 현 재 한국정보처리학회 상임부회장
 2021년 ~ 현 재 KTSDE 편집위원장
 관심분야: 분산 시스템, 모바일 클라우드 컴퓨팅, e-Learning



손 진 곤

<https://orcid.org/0000-0002-0540-4640>
 e-mail : jgshon@knou.ac.kr
 1991년 고려대학교 전산학전공(박사)
 1991년 ~ 현 재 한국방송통신대학교 컴퓨터학과 교수
 1997년 ~ 현 재 State University of New York (Stony Brook) Visiting Professor

2000년 ~ 현 재 ISO/IEC JTC1/SC36 Korea Delegate
 2009년 ~ 현 재 이러닝학회 부회장
 2010년 한국정보처리학회 부회장
 관심분야: 컴퓨터통신망, 분산시스템, e-Learning, 정보기술 표준화