



지도기계학습을 이용한 트위터 뉴스의 프레임 특성 분석

코로나19 보도 프레임의 자동화 판별 방법을 중심으로

이주연 중앙대학교 미디어커뮤니케이션학과 석사과정

이신행 중앙대학교 미디어커뮤니케이션학과 조교수

Using Supervised Machine Learning to Uncover Framing Features in Twitter News

An Automated Frame Analysis of COVID-19 Coverage*

Ju Yeon Lee**

(Master's Student, Department of Media and Communication, Chung-Ang University)

Shin Haeng Lee(ORCID:<https://orcid.org/0000-0003-2817-3969>)***

(Assistant Professor, Department of Media and Communication, Chung-Ang University)

Using the supervised machine learning model, this study examined the coverage of COVID-19 by news frames—severity, susceptibility, benefits, and barriers—drawing upon the health belief model. In particular, linguistic features of each frame were automatically derived from the headline and lead of news and they were applied to explore framing features of Twitter messages posted by the major newspapers in South Korea. The data included news articles and tweets about COVID-19 from Chosun Ilbo, JoongAng Ilbo, Kyunghyang Shinmun, and Hankyoreh. To automatically identify news frames, we employed support vector machine(SVM) and naïve Bayes(NB) algorithms by evaluating the accuracy of classifying each frame in 2,000 randomly sampled articles. Furthermore, the optimal classification algorithm was applied to 2,000 randomly sampled tweets to evaluate the predication accuracy for each frame and reveal distinctive linguistic features of each frame on Twitter. Findings showed that perceived threat frames of severity and susceptibility were emphasized in the coverage to a greater extent than behavioral evaluation frames of benefits and barriers, highlighting the risk-aware aspects of Covid-19 prioritized over the costs and benefits of preventive behavior. But we also found that severity and barriers frames were not constructed by consistent and distinct features, given the reduced accuracy of models compared to susceptibility and benefits frames. Furthermore,

* This research was supported by the Chung-Ang University Research Scholarship Grants in 2020.
(이 논문은 2020년도 중앙대학교 연구장학기금 지원에 의한 것임)

** lgy9712@cau.ac.kr

*** shinlee@cau.ac.kr, corresponding author

news frames on Twitter were constructed in a more flexible and discriminatory manner insofar as the logic of social media engages more personalized and emotional use of language in the content. This study sets out a methodology whereby machine learning is employed to code news frames in large-scale news coverage of COVID-19 and identify the features of framing language in an automated, transparent, and reproducible way.

Keywords: Automated Frame Analysis, Supervised Machine Learning, Twitter News, COVID-19, Health Belief Model

1. 연구의 동기

뉴스는 개인이 국가적 재난 상황을 인지하고 대응하기 위한 정보 습득의 주요 수단이다. 더욱이, 소셜미디어를 통해 확산되는 뉴스는 그 전파방식의 신속성과 참여성에 따라 관련 정보의 유통은 물론 이용자의 태도 형성과 행동 촉진에 미치는 영향력이 커지고 있다. 이러한 이유로 코로나바 이러스감염증-19(이하 코로나19)의 대유행이라는 위기를 극복하기 위해 시민 개개인의 건강행동(예: 마스크 착용과 집합 금지 등의 사회적 거리 두기와 코로나19 감염 진단 검사) 참여에 언론보도가 어떠한 영향을 미치고 있는 것인지에 대한 논의가 커지고 있다. 무엇보다 다른 이슈에 비해 건강 관련 뉴스에 대한 정보 의존도가 높다는 점을 고려할 때(Talosis-Garcial & Davis, 2005), 이용자의 건강행동에 대한 인식과 판단에 영향을 미칠 수 있는 언론보도의 프레임에 대한 분석이 필요한 시점이다. 무엇보다 소셜미디어를 통한 건강 관련 정보의 확산이 빠르게 증가함에 따라 트위터를 통해서 유통되는 뉴스의 프레임과 그 구성 방식의 특성에 대한 진단 역시 요구된다. 이에 본 연구는 코로나19가 언론보도에서 구성되는 방식, 즉 프레임을 분석하고 이것이 트위터라는 소셜미디어 공간에서는 어떠한 차별성으로 발현되는지 분석했다.

사실, 언론보도를 통한 건강정보를 습득한 뉴스 이용자가 어떠한 인식 및 신념체계를 바탕으로 건강행동에 참여하게 되는지는 건강신념모델을 이론적 토대로 활발히 탐색되어져 왔다. 질병 정보를 제공하는 뉴스를 소비한 개인이 건강예방행동을 취함에 있어 지각된 위협(perceived threat)과 행동적 평가(behavioral evaluation)로 구분된 인식과 신념체계가 작동하고 있음을 밝혀낸 것이다(Entwistle, 1995; Rosenstock, 1974). 구체적으로, 지각된 위협은 개인이 느끼는 질병의 위협 정도에 대한 인지적 평가로써 감염으로 인한 결과의 심각성을 지각하는 수준을 나타내는 심각성(severity)과 자신의 위협 노출 정도와 감염으로 인한 건강 상실 가능성에 대한 인식 수준인 취약성(susceptibility)으로 구분되어 건강행동에 영향을 미치는 신념체계를 구성하고 있다(김지혜·조재희, 2019; Champion & Skinner, 2008). 반면, 행동적 평가는 예방적 건강행동이 가져올 결과에 대한 기대가치로써 질병예방행동으로 얻을 수 있는 이득(benefits)과 이를 위해 요구되는 고통, 불편의 정도, 비용 부담 등의 부정적 결과에 대한 인식을 의미하는 장애(barriers)로 구분되어 이러한 주관적 신념이 건강행동에 영향을 미치는 것으로 밝혀졌다(조성은·신호창·유선옥·노형신, 2012; Janz & Becker, 1984).

그러나 이러한 신념 변인에 영향을 미치는 뉴스를 프레임의 관점에서 분석한 연구는 상대적으로 빈약하다(하진홍·임혜준, 2020; Jung & Brann, 2014; Quick & Bates, 2010). 건강에 관련한 뉴스의 내용이 어떠한 주제적 특성으로 구성되어 있는지를 파악하는 방법으로 그

효과를 추론할 수 있음에도 말이다(Coleman & Thorson, 2002). 이러한 시각은 언론보도가 특정 이슈를 다룸에 있어 정보를 선택적으로 제시하거나 강조함으로써 문제에 대한 수용자의 인식과 태도에 영향을 미치게 된다는 프레이밍 이론에 기초한다(Entman, 1993). 이에, 안순태와 이하나(2016)가 뉴스 소비자의 건강행동에 언론보도의 직접적인 영향력이 간과된 경향을 지적하며 메시지가 어떠한 건강신념의 변인을 강조하는지를 분석하여 자살 예방을 위한 우울증 관련 정보가 관련 뉴스에서 다루지고 있는지, 그리고 어떠한 경향성과 문제점이 있는지를 고찰한 점은 주목할 만하다. 본 연구 역시 같은 문제의식에서 코로나19 뉴스가 사회적 거리두기와 진단 검사 및 백신 접종 같은 감염예방행동에 영향을 미치는 언론보도의 특성을 건강신념모델의 주요 변인으로 살펴보고자 뉴스기사의 프레이밍을 분석했다.

그러나 본 연구가 더욱 주목하고자 하는 점은 이러한 코로나19에 대한 언론보도 프레이밍 트위터라는 소셜미디어 환경에서는 어떻게 발현되고 있는가이다. 소셜미디어를 통한 뉴스 유통이 활발해짐에 따라 국내 주요 언론사들은 트위터 계정을 생성하고 자사의 뉴스를 게시하며 트윗에 메시지를 추가하고 있다. 그리고 이러한 보도 형태는 사용자들에게 뉴스의 헤드라인과 리드가 아닌 새로운 메시지를 통해 뉴스를 접하게 되는 가능성을 보여주며 그 내용적 차별성에 대한 분석을 요구한다. 이것은 개인 간의 정보와 의견이 자유롭게 교환되는 방식으로 뉴스가 유통되는 소셜미디어의 매체적 특성이 건강 관련 보도에 미치는 영향을 탐색할 필요성을 제시하는 것이다. 특히, 이 과정에서 개인들의 생각과 감정의 공유가 뒤섞여 일어나는 경향(Welbers & Opgenhaffen, 2019)이 뉴스보도의 프레이밍 특성에 반영될 수 있는 가능성에 주목했다. 따라서, 본 연구는 코로나19 관련 뉴스가 트윗으로 게시될 때 추가되는 메시지에서 어떠한 프레이밍이 강조되고 있는지를 분석하여 기사의 헤드라인과 리드에서 파악된 결과와 비교했다. 이를 통해 트윗 메시지에서 포착된 코로나19에 대한 프레이밍이 뉴스 헤드라인과 리드에서 도출된 건강신념모델의 프레이밍과 공명하는지, 아니면 소셜미디어의 특성으로 재구성되는지를 평가했다.

트위터가 전통적 뉴스매체와는 다른 콘텐츠 생성 원리를 보인다면 건강커뮤니케이션에서 작동하는 뉴스 프레이밍 역시 트윗 메시지에서는 어떻게 발현되는지에 대한 분석을 요구한다. 감염증 확산의 전개 상황과 예방 수칙은 물론 과급되는 문제에 대한 객관적이고 정확한 정보 제공이 중요한 코로나19 상황에서 트위터에 게시되는 언론보도의 프레이밍은 사용자의 위험 인식과 예방 행동에 미칠 수 있는 영향이 클 것으로 예상되기 때문이다. 그러나 건강정보의 유통과 이용자에 대한 효과에 있어 소셜미디어 역할을 진단해 온 연구는 주로 미디어 이용 정도에 초점이 맞춰져 왔다(김주미·최정화·박동진, 2018; 김지혜·조재희, 2019). 반면, 건강정보 습득과 질병에 대한 인식, 그리고 건강행동에 소셜미디어를 통해 제공되는 뉴스가 어떠한 영향을 미칠 것인지를

고찰하기 위해 그 내용적 특성을 파악한 연구는 부족한 실정이다. 특히, 코로나19에 대한 허위정보(disinformation)와 오정보(misinformation)가 소셜미디어를 통해 빠르게 확산하며 이용자의 정보 습득에 상당한 영향을 미치고 있다는 연구결과(Apuke & Omar, 2021)를 고려한다면 언론보도가 트위터를 통해 유통될 때의 프레임 차이를 발견하는 것은 건강행동에 대한 인식과 신념이 격차를 보이는 원인 진단에 기여할 수 있다.

이상의 문제 제기를 바탕으로 본 연구는 언론보도의 프레임 구성이 소셜미디어 환경에서 다르게 발현되고 있음을 주장하기 위해 코로나19 관련 뉴스에 대한 프레임을 분석했다. 우선, 코로나19 관련 언론보도의 프레임을 건강신념모델에 기반한 주요 네 가지 변인들인 심각성, 취약성, 이득, 장애로 구성했다. 그리고 뉴스기사에서 발견되는 각 프레임의 언어적 특성을 도출한 후, 이를 트윗 메시지의 프레임 판별에 적용했다. 이를 위해 지도기계학습을 활용한 자동화 방식의 프레임 판별을 다음과 같이 실시했다. 먼저, 텍스트의 언어적 특성에 따라 코로나19 관련 뉴스기사의 헤드라인과 리드에서 각 프레임을 판별하는 분류 알고리즘을 학습에 적용했다. 그리고 표본에 대한 예측 시험 성능을 인간 코더의 판별 결과와 비교하여 평가해 최적의 모델을 선택했다. 이렇게 훈련된 프레임 분류 알고리즘을 동일 뉴스에 대한 트윗 메시지에 적용하여 그 언어적 특성의 유사성을 측정했다. 이를 통해, 코로나19에 대한 언론 보도 프레임이 트위터와 같은 소셜 미디어의 사용자에게 제시될 때 발생하는 차별적 메시지 특성을 파악하고 그 함의를 고찰했다.

2. 문헌 검토

1) 코로나19의 확산과 미디어의 프레임 효과

2019년 말 중국 우한에서 첫 확진자가 보고된 코로나19는 호흡기를 통해 쉽게 전파되는 특징을 보이지만 백신과 치료제가 없어 전 세계적 유행이 빠르게 확산된 신종 감염증이다. 국내의 경우에도 2020년 1월 중국 우한에서 입국한 중국 국적의 35세 여성이 처음으로 코로나19 확진 판정을 받은 후 신천지 신도를 중심으로 대구·경북 지역에 수천 명에 달하는 확진자가 생겨남에 따라 전국적 유행 국면에 접어들게 되었다(최성호, 2020). 물론 코로나19 감염환자의 치사율은 평균 3%로 보고되고 있어 질병 자체의 위험이 높지 않다는 시각도 있으나,¹⁾ 기저질환자와 노인에게

1) 물론, 감염환자의 치사율을 최고 10.8%(이탈리아)에서 최저 0.8%(독일)로 국가별로 매우 상이한 것으로 보고되고 있다. (2020년 3월 29일 기준)

는 치명적일 뿐만 아니라 완치자의 상당수가 후유증을 겪고 있는 것으로 알려져 감염을 예방하는 행동에 시민들의 참여가 촉구되고 있다. 실제 코로나19의 확산을 막기 위해 정부는 마스크 착용 의무화와 5인 이상 사적모임금지 등의 사회적 거리두기와 감염 진단검사로 대표되는 방역 조치를 펼쳐오고 있다. 또한, 최근에는 집단 면역을 목표로 백신 접종에 적극 참여할 것을 홍보하고 있다. 그러나 다른 한편으로 강력한 방역 조치로 인해 시민들의 외출과 모임이 자제되면서 지역 경제는 물론 교육 및 종교와 여가생활에 걸친 사회 전반에 상당한 피해가 발생하고 있다는 지적이 있다.

언론은 이러한 코로나19의 성격과 확진자 규모 및 형태, 그리고 경제적, 사회적 피해 상황을 시시각각으로 보도해왔다. 그러나 이러한 중계식 보도 못지않게 감염증의 확산을 억제하고 예방하기 위한 건강행동, 즉 사회적 거리두기와 진단검사에 관한 정보전달이 중요하게 다뤄져야 하는 점 또한 사실이다. 즉, 질병관리청을 비롯한 책임있는 기관에서 제공하는 신뢰할 만한 정보가 언론보도를 통해 효과적으로 전달될 때 시민 개개인의 자발적 참여가 필수적인 사회적 거리두기와 진단검사 조치가 성공적으로 운영되어 코로나19의 확산 억제라는 목적을 달성할 수 있기 때문이다. 따라서 언론이 코로나19 관련 보도를 어떠한 시각과 초점에 맞춘 뉴스로 구성하고 제공하는 지가 중요한 문제로 떠오르고 있다. 이는 프레이밍 이론에 따라 질병 관련 뉴스 소비자의 신념은 물론 예방 행동이 언론보도에서 강조되고 있는 문제의 측면에 영향을 받는다는 관점에 기인한다. 다시 말해, 뉴스기사가 어떠한 정보를 선택하고 강조하는지에 따라 대중들의 질병에 대한 위험 인식과 예방 행동이 달라지기 때문이다(안순태·이하나, 2016; Coleman & Thorson, 2002; Yanovitzky & Blitz, 2000). 이러한 관점은 코로나19 관련 뉴스기사를 구성하는 메시지의 성격에 따라 사회적 거리두기와 진단검사라는 건강행동에 대한 자극과 촉진이 다르게 나타날 수 있다는 점에 주목하여 미디어 프레이밍을 분석하는 시도의 이론적 토대를 제공한다.

본 연구가 코로나19 언론보도 분석에 적용하는 프레이밍 효과 이론은 뉴스 소비자의 인식과 태도를 결정하는 해석의 틀이 미디어를 통해 선택적으로 정보가 제공되는 방식에 기초한다는 전제로 미디어의 효과를 뉴스가 해당 이슈를 어떻게 다루는지에 초점을 맞춘다(Gamson & Modigliani, 1989; Pan & Kosicki, 1993; Scheufele, 1999). 미디어에서 보도되는 뉴스의 정보가 구성되는 틀, 즉 프레이밍이 뉴스 소비자가 해당 이슈 혹은 사건을 해석하는 방식을 결정한다는 입장이다(이준웅, 2001; Edy & Meirick, 2007; Mcleod & Detenber, 1999). 물론, 프레이밍은 메시지를 통한 정보전달 방식은 물론 뉴스보도를 구성짓는 과정과 이용자의 정보 처리와 인지의 구조, 혹은 이러한 모든 과정의 상호작용을 의미할 수 있다(김종화·유홍식, 2012a; 이준웅, 2000). 하지만 본 연구가 주목하는 해석적 방식으로서의 프레이밍(Goffman, 1974)은 뉴스를 구성하는 정보들이 선택적으로 강조되는 패턴으로써 이용자가 언론보도를 통해 현실을 바라보고 그 의미를 이해하는 틀로 작용

한다는 점에 기반한다(Entman, 1993; Price, Tewksbury, & Powers, 1997). 이러한 차원에서 뉴스 프레임을 분석해 온 연구자들은 대중의 인식과 태도 및 행동의도는 언론보도가 강조하는 특정 정보와 관점이 선택적으로 제시되는 텍스트적 속성에 영향을 받는다는 가정하에 해당 이슈의 어떠한 측면과 맥락이 부각되고 있는지를 파악하여 그 효과를 추론했다(김은정·유홍식·한규준, 2019; 김효정, 2017; 이준웅·김성희, 2018; Scheufele, 1999).

프레이밍 효과 이론은 건강 혹은 위험에 관련된 뉴스보도가 이용자의 예방 행동에 미치는 효과를 파악한 연구에 활발히 적용돼 왔다(김종화·유홍식, 2012a; 2012b; 김효정, 2017; 정재선·이동훈, 2012; 조혜림·정민수, 2019). 이러한 프레이밍연구는 카네만과 티버스키(Kahneman & Tversky, 1979)의 '예사이론(Prospect Theory)'이 제시하는 불확실한 위험 상황에서 관련 메시지의 획득·손실 프레이밍에 따라 이용자의 행동 선택이 달라지는 효과에 근거한다(김수진·차희원, 2009). 가령, 건강행동으로 인한 확실한 이득이 강조된 메세지(획득 프레이밍)에 노출된 의사결정자는 위험을 예방 혹은 회피하려는 경향을 띄는 한편, 행동의 결과가 이득으로 확실히 이어지지 않는 경우에는 이를 취하지 않았을 경우에 예상되는 피해가 강조된 메세지(손실 프레이밍)에 더 반응하여 행동의도가 고취되는 점에 착안해 프레이밍 분석을 시도하여 미디어 효과를 검증한 것이다(김종화·유홍식, 2012b; Tversky & Kahneman, 1981). 위 연구결과는 코로나19와 같은 건강위험을 다루는 뉴스의 프레이밍이 감염 예방행동에 대한 이용자들의 인식과 평가 및 참여의도에 영향을 미칠 수 있음을 보여준다. 하지만 획득·손실 프레이밍의 이분법적 메시지 구성 방식을 넘어 질병 예방을 위한 행동 결정에 영향을 미치는 세부적인 인지 요인들을 프레이밍 효과 이론에 기초해 분석한 연구는 부족하다(안순태·이하나, 2016; Quick & Bates, 2010). 따라서 본 연구는 건강신념모델에 기초해 코로나19 예방 행동을 설명하는 네 가지 인지적 요인(심각성, 취약성, 이득, 장애)으로 관련 뉴스의 프레이밍을 파악하고자 한다.

2) 건강신념모델과 뉴스 프레이밍

코로나19가 빠르고 광범위하게 확산됨에 따라 질병에 대한 이해와 감염 예방이 중요해지자 대중에게 관련 정보를 전달하고 사회적 거리두기와 진단검사로 대변되는 건강행동 촉진에 언론의 역할이 주목받고 있다. 질병 관련 언론보도를 구성하는 뉴스 프레이밍에 따라 대중의 위험 인식은 물론 예방과 치료를 위한 행동 유발이 직접적으로 영향받기 때문이다(김수진·차희원, 2009; 김지혜·조재희, 2019; 정재선·이동훈, 2012). 따라서 코로나19 뉴스가 어떤 프레이밍으로 구성되어 있는지를 건강신념모델에 기초해 분석해 예방 행동 의도에 미치는 영향을 고찰해 볼 필요가 있다. 이에 본 연구는 언론이 대중의 건강행동에 미치는 영향력을 고려하여 건강신념모델을 이론적

분석 틀로 활용해 코로나19 관련 뉴스의 특성과 효과를 고찰하고자 한다.

베커(Becker, 1974)와 로젠스탁(Rosenstock, 1974)에 의해 헬스커뮤니케이션 분야의 연구에 본격적으로 활용되기 시작한 건강신념모델은 어떠한 동기적 요인으로 질병을 예방하거나 진단하는 행위를 선택하는지를 설명하는 이론이다. 이 이론은 개인의 노력에 따라 기대되는 보상이나 이익에 따라 행위 발생의 동기가 형성된다는 기대이론에 기초하여 건강 관련 행동을 유발하는 심리적 기제를 설명한다. 그리고 건강행동을 결정하는 데 있어 질병에 노출되어 위험에 빠질 수 있는 가능성을 평가하는 지각된 위협(perceived threat)과 이를 예방하는 행동의 결과에 대한 기대가치인 행동적 평가(behavioral evaluation)로 구분된 개인의 태도와 신념이 작용하고 있음을 제시한다(Andreeva, Unger & Pentz, 2007; Tanner-Smith & Brown, 2010). 이중 지각된 위협은 심각성(severity)과 취약성(susceptibility)으로 구분되어 보다 세부적인 요인 탐색이 이뤄져 왔다. 심각성은 질병에 노출되었을 때 개인의 신체에 미치는 영향 또는 사회적 측면에서 감염 및 그 결과의 심각성에 대한 주관적 신념을 의미하는 것으로 인지된 심각성이 높을수록 질병에 대한 적극적 예방 및 치료 행동이 이뤄진다(김지혜·조재희, 2019; 장해·박주식·이경식, 2020). 취약성은 이러한 질병 위험에 자신이 노출될 가능성에 대한 신념으로 높은 심각성에 더해 취약성에 대한 지각 수준이 높을 경우 매우 적극적으로 건강행동을 실천한 의도를 가진다(Janz & Becker, 1984). 이러한 관점에서 안순태와 이하나(2016)는 뉴스 메시지가 심각성과 취약성을 강조할 경우 자살을 예방하는 행동으로의 우울증 치료 의도가 촉진될 것으로 봤다. 본 연구 역시 같은 시각에서 언론이 코로나19에 대한 심각성(예: 치명률과 부작용 혹은 사회적 경제적 피해)과 취약성(예: 확진 현황 중계 및 일상적 감염경로)을 강조하는 뉴스 프레임으로 이용자가 지각하는 위협 수준을 높여 건강행동을 자극하고 유도하는 맥락을 살피고자 한다.

행동적 평가는 예방 행동의 결과에 대한 주관적 신념으로 지각된 이득과 장애로 구분된다(Janz & Becker, 1984). 우선, 지각된 이득(perceived benefits)은 질병의 위험을 회피하거나 예방함으로써 얻을 수 있는 이득에 대한 기대 수준을 의미한다(Rosenstock, 1974). 즉, 질병을 예방하는 행위의 결과가 얼마나 효과적이고 긍정적인 것인가에 대한 개인의 인식으로써 그 정도가 높을수록 건강에 대한 예방적 행동이 손쉽게 일어날 수 있다는 것이다(한규훈, 2011). 지각된 장애(perceived barriers)는 건강행동 과정에서 발생할 것으로 예측되는 부정적 측면에 대한 개인의 인식을 의미한다(Rosenstock, 1974). 건강행동에 수반되는 물질적 비용은 물론 불쾌감이나 통증 및 부작용에 대한 염려와 같은 정신적 장애를 모두 포함하는 주관적 신념으로써, 높은 장애 지각은 질병을 예방하기 위한 행위 결정을 저해하는 것으로 밝혀졌다(김명혜, 1997; Hyman, Baker, Ephraim, Moadel & Philip, 1994). 뿐만 아니라 지각된 이

특과 장애는 치료나 진단 목적이 아닌 예방을 위한 건강행동인 경우에 더 주요한 요인으로 작용하고 있음이 드러났다(Carpenter, 2010). 이 결과는 건강행동의 종류에 따라 각 주관적 신념이 미치는 영향이 다르다는 점이 고려되어야 함을 의미하는 것으로(Janz & Becker, 1984), 코로나19를 예방하기 위한 사회적 거리두기와 진단검사를 촉진하는 과정에는 이 행위가 수반하는 이득과 장애에 대한 신념이 중요한 영향을 미칠 수 있음을 제시한다.

뉴스보도는 건강과 관련한 정보를 제공해 대중들의 관련 지식 향상 뿐만 아니라 질병을 예방하기 위한 행위를 유도하기도 한다(Coleman & Thorson, 2002; Entwistle, 1995). 가령, 스트라이커(Stryker, 2003)의 연구에 따르면 마리화나 사용의 부정적 결과를 강조한 뉴스보도의 증가는 청소년들의 마리화나 절제를 촉진하는 반면, 긍정적 결과를 다루는 뉴스보도의 증가는 청소년들의 마리화나 절제를 저해하는데 유의미한 효과를 일으켰다. 이는 뉴스보도가 건강 행위에 직접적인 영향을 미치고 있음을 보여준 연구로 미디어의 메시지가 어떠한 관점으로 구성되어 있는지, 즉 뉴스 프레임에 따라 대중의 인식과 태도는 물론 질병으로부터 건강을 보호하기 위한 행동의 변화가 일어날 수 있음을 제시한다(Clarke, McLellan, & Hoffman-Goetz, 2006). 이는 감염병의 확산으로 공중보건의 위기가 발생했을 때 대중이 이를 어떻게 인식하고 이에 대해 어떠한 행동을 결정할 것인지를 예측하기 위해서는 미디어가 문제를 다루는 프레임을 분석하는 것이 선행되어야 함을 의미한다(Shih, Wijaya, & Brossard, 2008). 특히, 건강신념모델에서 제시하는 네가지 신념요인인 심각성, 취약성, 이득, 장애는 질병 예방을 위한 행위의 가치 판단에 영향을 미쳐 실질적 행동 변화의 의도를 유발하는 효과가 있기 때문에(김지혜·조재희, 2019), 각 요인이 미디어에서 어떻게 강조되고 있는지를 프레임 분석으로 살펴봄으로써 건강행동에 대한 효과를 추론할 수 있다. 실제, 안순태와 이하나(2016)는 '자살'과 '우울증'의 키워드로 검색된 관련 뉴스기사를 건강신념모델에 기초해 내용분석하여 자살의 심각성은 강조되고 있는 반면 우울증에 대한 일반적인 취약성은 간과되고 있고, 우울증 치료의 이득과 장애에 대한 정보 제공 또한 미흡한 점을 발견하여 자살예방 행동을 유도하는 미디어의 효과가 미진함을 추론했다. 한편, 하진홍과 임혜준(2020)은 전염병 이슈를 다루는 뉴스보도를 내용분석하여 언론이 문제의 심각성을 강조하는 프레임을 가장 많이 사용하고 있지만 대중의 위기의식은 이에 미치지 못함을 발견했다. 이는 질병의 심각성이나 취약성과 같은 부정적 프레임이 이용자의 인식 변화에 갖는 효과가 제한적이므로 건강행동으로 인한 이득을 강조하거나 이를 저해하는 장애를 지적하는 긍정적 프레임의 중요성을 의미한다.

3) 언론사의 트위터를 통한 뉴스 제공

디지털 네트워크의 확산에 따라 빠르게 발전 중인 소셜미디어는 재난 커뮤니케이션의 주요 채널로

써 주목받고 있다(최재웅, 2012). 특히, 트위터는 140자로 이뤄진 간결한 메시지를 연속적으로 게시할 수 있는 마이크로블로그 서비스이자 팔로잉(following)하는 사용자의 트윗이 제공되는 소셜미디어로서 신속한 커뮤니케이션이 이뤄질 뿐만 아니라 해시태그를 통한 검색 기능으로 재난 상황에서 효과적이고 조직적인 정보의 교류와 공유가 가능한 구조로 평가받는다(이영재, 2009). 가령, 배병걸, 이보람, 최선화(2015)는 트위터를 통한 지진 관측이 기상청 발표보다 약 4분, 전통 미디어 매체보다는 약 10분 정도 빠른 것으로 관찰됐다고 밝혔다. 더 나아가, 트위터는 메시지 생산자와 이용자의 경계가 모호한 “참여문화”를 바탕으로 오프라인 환경에서는 관찰되지 않던 새로운 사회적 연결망에 따른 변화하는 정보 교류 양상을 촉진하는 것으로 여겨진다(Jenkins, 2006). 실제, 박효찬과 박한우(2017)는 2016년 경주지진이 트위터를 통해 어떻게 소통되고 있는지 분석하여 지진 발생 직후 네트워크의 영향력 차이(팔로워의 수)를 불문하고 모두가 정보 교류에 참여하는 공동체의 구조를 보였음을 발견했다. 이러한 시각에서 박준형, 류범모, 오효정(2017)은 가슴기 살균제 사건이라는 사회적 재난 상황에서 트위터가 신속한 대응을 위한 여론 수렴의 소통창구로 역할 할 수 있음을 제시했다. 더욱이 소셜미디어에서는 이용자의 공유를 통해 새로운 뉴스가 소비되고 확산되는 현상이 빈번하게 일어나고 있으며, 이에 전통적 매체에 기반한 언론사 역시 뉴스 보도의 주요 창구로써 트위터와 같은 소셜미디어를 활발하게 이용하고 있다. 물론, 트위터를 통해 구성된 사회적 관계망은 가짜뉴스로 대변되는 허위 정보(disinformation)와 오정보(misinformation)의 무분별한 생산과 빠른 전파가 일어나는 여건을 제공한다는 비판적 시각 또한 존재하는 것이 사실이다(유재원·금희조, 2018). 그럼에도 트위터는 신속하고 참여적인 정보 교류에 용이한 소셜미디어로서 재난 상황에 그 활용 가치가 큰 커뮤니케이션 수단이다.

이러한 트위터의 속성은 코로나19 대확산의 위기를 극복하기 위한 사회적 거리두기와 진단 검사라는 정부의 방역지침이 신속하고 폭넓게 전파되기 위해 트위터상에서 관련 정보가 어떻게 제시되고 있는지를 파악할 필요성을 제기한다. 이는 위기커뮤니케이션의 주요 주체로써 언론이 코로나19 관련 보도를 전달함에 있어 트위터라는 커뮤니케이션 도구를 어떻게 사용하고 있는가에 대한 분석으로 문제에 접근할 수 있음을 의미한다. 트위터로 제공되는 뉴스는 쌍방향 커뮤니케이션을 통한 참여적 소통 방식으로 전통적 매체와 구별된 방식의 정보의 교류와 공유를 촉진하기 때문이다(Hermida, Fletcher, Korell, & Logan, 2012). 우선, 트위터에서는 사용자 누구나 정보의 생산자로서 참여할 수 있을 뿐만 아니라 리트윗 기능을 이용하여 각자의 사회적 네트워크로 전달하는 전파자로서의 역할도 가능하다. 리트윗은 작성한 글을 전달받는 사람이 자신의 팔로워에게 재전송하여 다른 사람도 글을 볼 수 있도록 하는 기능으로 정보 확산에 중요한 역할을 하고 있다(김지현, 2015). 또한, 해시태그 기능을 통해 특정 이슈에 대한 트윗으로 연결되

이 조직화된 공동체 구성을 가능하게 하여 트위터 사용자의 정보 습득과 태도 형성은 물론 행동 참여에도 영향을 미칠 수 있다(Bennett & Segerberg, 2013). 실제로 트위터 내에서 많은 팔로워를 가진 인플루언서(influencer)는 그 메시지가 전통적인 뉴스매체에서 보도될 만큼의 영향력을 행사하는데, 사회적 관계망을 통해 정보를 확산할 뿐만 아니라 의제를 설정하여 공론장을 조성하는 의견지도자(opinion leader)로서의 역할을 담당하고 있다(박효찬·박한우, 2017).

따라서 트위터를 통해 제공되는 언론사의 메시지는 그 구성 방식이 전통적 매체 환경의 그것과는 다른 원리로 이뤄진다. 전통적으로 뉴스기사의 헤드라인과 리드는 수동적 메시지 이용자인 독자 혹은 시청자의 주목을 받는 것과 함께 보도내용의 핵심을 간결히 전달하는 목적을 가진다(Andrew, 2007; Kuiken, Schuth, Spitters & Marx, 2017). 그러나 트위터와 같은 소셜미디어 환경에서는 사용자의 관심 끌기나 효과적인 정보 전달을 넘어 ‘댓글달기’, ‘리트윗’, ‘좋아요’ 등으로 표현되는 적극적 형태의 커뮤니케이션 참여 유도를 통한 메시지의 바이럴(viral) 확산이 언론의 프레이밍 효과에 선행되는 조건이 된다(Garcia-Perdomo, Salaverria, Kilgo & Harlow, 2018; Hermida et al., 2012). 이런 관점에서 어떠한 메시지의 특성이 소셜미디어에서의 바이럴 뉴스를 야기하는 지에 대한 분석이 산업적 차원에서는 물론 학계에서도 활발히 논의되고 있다. 예를 들어, 뉴스가 공유할 만한 속성(shareability)이 있는지가 소셜미디어 환경에서의 새로운 뉴스 가치로 부상했음을 주장하며(Harcup & O'Neill, 2017), 이는 그 내용이 웃음을 유발하거나 분노를 자아내는 등의 감정적 동요를 일으킬 때 사용자들의 공유 행동이 발생할 가능성이 크다는 점에 주목한다(Al-Rawi, 2019; Newman, 2011). 실제 언론보도의 언어적 구성을 비교 분석한 결과, 전통적 매체 환경에서의 헤드라인과 리드에 비해 트위터와 같은 소셜미디어를 통한 뉴스 게시 시 추가되는 상태 메시지에서 주관적 의미의 어휘와 긍정적 감정 표현이 많이 사용되고 있음을 발견한 연구는 시사하는 바가 크다(Welbers & Opgenhaffen, 2019). 뉴스의 전파 과정이 언론사와 수용자의 일방적 관계가 아닌 소셜미디어의 사용자들 간의 자유롭고 일상적인 연결관계를 통함에 따라 개인의 의견과 감정 표현의 주관적 언어가 사실 위주의 객관적 언어보다 선호되는 경향을 보여주기 때문이다.

위의 논의를 종합해보면, 코로나19 확산에 따른 보건위기에 대한 올바른 정보전달은 물론 감염 예방행동을 촉진하기 위한 건강커뮤니케이션의 주요 행위자인 언론의 역할을 진단하고 평가하기 위해서는 뉴스보도의 프레이밍이 트위터라는 소셜미디어 환경에서 어떻게 나타나고 있는지를 파악하는 것이 중요한 과제임을 알 수 있다. 트위터를 통한 언론의 프레이밍은 감염병에 관련한 뉴스를 접하는 사용자들의 인식과 태도는 물론 예방을 위한 행동의 변화를 유도할 수 있기 때문이다(장해·박주식·이경식, 2020). 즉, 코로나19의 대유행이라는 재난 상황을 간접적으로 ‘경험’

하는 트위터 사용자는 뉴스보도의 프레임에 따라 건강신념이 영향받는다라는 점을 고려하면 언론이 어떠한 메시지를 제공하는지에 대한 분석은 사회적 거리두기와 진단검사를 촉진하는 언론의 건강 커뮤니케이션 효과를 추론하기 위해 필요하다. 특히, 전통적인 미디어 환경에서 헤드라인과 리드를 통해 언론이 뉴스의 요점을 전달하고 독자의 관심을 유도해온 것에 비교하여 마찬가지로 함축적인 메시지가지만 공유될 만한 가치를 포함하기 위해 트윗은 보다 감정적인 언어로 구성될 것이라는 가정을 기반으로 이러한 소셜미디어의 특성이 건강신념 프레임에 어떠한 영향을 미치는지에 대한 실증적 분석이 요구된다. 이를 통해 갈수록 확장되고 있는 소셜미디어 환경에서 언론이 코로나19와 같은 감염병의 대유행 위기를 대중에게 효과적으로 전달하고 사회적 거리두기와 진단검사라는 예방행동에 대중의 참여를 유도하는 건강커뮤니케이션에 관련한 제언을 도출하고자 한다.

3. 연구문제

본 연구는 코로나19 관련 언론보도에서 이용자의 인식구조를 결정하는 해석적 틀로서의 프레임에 건강신념모델에 기초해 탐색하고 트위터라는 소셜미디어 환경에서 발견되는 프레임의 차별성을 분석했다. 트위터의 차별적 프레임을 추적하는 시도는 언론사의 기사 헤드라인과 리드에 비해 같은 뉴스의 소셜미디어 게시글이 주관적 감정 표현을 많이 포함하고 있다는 연구결과(Welbers & Opgenhaffen, 2019)에 기반한다. 그리고 그 원인으로 소셜미디어의 메시지가 개인의 감정이 드러난 자기표현일수록 사용자들의 소통 참여가 증가한다는 점도 감안했다(Berger & Milkman, 2010). 결국, 언론사의 트윗 게시글이 사용자 개인에게 소구함으로써 공감을 표시하거나 자신의 사회적 관계망에 공유(혹은 리트윗)하는 등의 관여도를 높이기 위해 링크로 제공되는 뉴스기사에 감정적 표현의 메시지를 추가할 수 있음을 제시한다. 이는 소셜미디어 환경에서 사용자의 공감과 댓글달기, 그리고 공유가 정보 확산에 중요한 역할을 맡게 됨에 따라(Klinger, 2013), 뉴스기사의 형식은 물론 내용 역시 매체적 특성에 맞춰 변화하고 있음을 의미한다(Stieglitz & Dang-Xuan, 2013). 결국, 트위터를 통해 게시되는 뉴스기사와 함께 작성되는 메시지가 헤드라인과 리드와는 구별된 내용적 초점으로 프레임을 구성하는지 살펴볼 필요성을 보여준다.

물론 본 연구는 코로나19와 같은 건강위험 관련 뉴스의 경우 정치적 이슈가 아닌 국민의 안전과 생계에 직결된 문제임에도 보수와 진보로 나뉜 언론사의 이념적 지향 차이가 언론보도 프레임에 영향을 미칠 수 있음을 고려했다(이민규·이예리, 2012; 이준웅·김성희, 2018). 또한, 이념적 지향과 같은 조직 차이는 물론 코로나19라는 위협의 전개 상황에 따라 뉴스 프레임이 시기

별 차이를 보이는 점도 참고했다(조민정·이신행, 2021; Shih et al. 2008). 결국, 코로나19 뉴스의 프레임에 영향을 미칠 수 있는 언론사별, 시기별 차이를 넘어 소셜미디어 환경에서 발현되는 프레임의 특성을 포착하는 것을 목표로 전통적 매체 환경에서 확인된 뉴스기사 프레임의 언어적 특성을 헤드라인과 리드에서 도출하고 이를 트윗 메시지와 비교했다. 결국, 소셜미디어의 정보 확산 원리를 고려해 건강 커뮤니케이션에서의 언론이 가진 역할이 변화할 가능성을 타진하고 그 함의를 논의하고자 했다. 이러한 목적을 바탕으로 본 연구는 다음의 연구문제를 제시하였다.

RQ 1. 코로나19 관련 언론보도의 경우 어떠한 언어적 특성과 알고리즘이 건강신념모델의 심각성, 취약성, 이득, 장애 요인으로 구성된 프레임을 기계학습모델로 판별하는데 더 정확한 결과를 도출하는가?

RQ 2. 코로나19 관련 언론 보도에서 나타난 심각성, 취약성, 이득, 장애 프레임은 각각 트위터 환경에서 발현된 뉴스 프레임으로 어떻게 표현되고 있고 그 차이는 무엇인가?

4. 연구방법

1) 분석대상

본 연구의 분석대상으로 국내 주요 종합일간지 조선일보, 중앙일보, 한겨레, 경향신문이 2020년 1월 20일부터 2021년 1월 19일까지 보도한 코로나19 관련 뉴스를 수집했다. 뉴스매체 선정은 전국종합일간지로서의 보도 영향력은 물론 보수와 진보의 이념적 정파성을 대표하는 언론사를 포함했다. 또한, 자료수집 기간은 코로나19의 국내 첫 확진자가 발생한 시점을 기준으로 감염 확산세가 등락을 거듭하며 위기가 지속돼 온 1년이다. 이 기간은 코로나19의 위기 상황에 따라 1차 대유행 - 안정세 - 2차 대유행 - 3차 대유행의 4시기로 구분된다. 1차 대유행 시기는 국내 확진자 발생이 시작된 2020년 1월 20일부터 대구·경북 지역의 신천지 집단을 중심으로 이어진 확산세가 잦아든 시점인 2020년 5월 5일까지이다. 이후의 안정세는 2020년 8월 23일까지로 이 시기는 정부의 방역조치가 생활 속 거리두기인 1단계에 머물렀다. 2차 대유행 시기는 2020년 8월 24일부터 수도권을 중심으로 시작된 감염 확산에 의해 사회적 거리두기 조치가 2단계로 상향된 기간으로 전국적 감염 확산의 조짐으로 정부의 고강도 방역조치가 시작된 2020년 11월 24일 이전까지이다. 이후 시기는 3차 대유행이 지속돼 온 2021년 1월 19일까지의 기간으로 선정했다. 그리고 코로나19의 감염 확산에 따른 건강 문제를 다루는 뉴스보도를 수집하기 위해 “코로나”와 “확진”

이라는 키워드를 선정함과 동시에, “우한 폐렴” 또한 코로나19의 확산 초기에는 공식 명칭과 함께 사용됐기 때문에 뉴스 검색 키워드에 포함하여 다음과 같은 방법으로 자료를 수집했다.

우선, 종합일간지의 뉴스기사 수집을 위해 한국언론진흥재단 데이터베이스인 <빅카인즈>를 이용해 검색 키워드가 헤드라인에 포함되어 있는 뉴스를 검색하여 조선일보 12,644건, 중앙일보 9,713건, 한겨레 2,635건, 경향신문 3,987건, 총 28,979건의 뉴스 기사를 CSV파일로 다운로드 받았다. 또한 본 연구의 분석도구로 활용한 R 환경에서 ‘rtweet’ 패키지를 이용해 각 언론사의 트위터 계정에서 키워드가 메시지에 포함된 트윗을 검색하여, 조선일보 621건, 중앙일보 1,252건, 한겨레 1,903건, 경향신문 1201건으로 구성된 총 4,977건의 트윗 데이터를 수집했다.²⁾

2) 종속변인

본 연구는 코로나19의 감염 예방을 위한 사회적 거리두기를 건강행동으로 전제하고, 언론보도를 취약성, 심각성, 혜택, 장애의 건강신념 프레임으로 구성된 분석 틀로 살펴보고자 했다. 특히, 소셜미디어를 통한 코로나19 언론보도의 특징을 파악하고자 언론사 뉴스기사에서 포착된 프레임이 같은 기간 트윗에서 공명하는지를 지도기계학습을 이용하여 분석했다. 지도기계학습을 이용해 프레임별로 뉴스 기사를 자동적으로 분류하는 방법은 시간과 공간의 제약에 구애받지 않고 디지털화된 대단위 자료를 효율적으로 분석할 수 있다는 장점뿐만 아니라 인간 코딩의 분류와 동일한 결과를 보이는 정확성 측면에서도 신뢰할 만한 성능을 보인다는 점에서 프레임 분석방법으로서의 활용 가치가 높아지고 있다. 실제로 연구자에 의해 미리 정해진 조각적 정의에 따라 뉴스기사에서 채택된 프레임을 판별하는 방법을 학습한 지도기계학습모델은 인간 코딩을 모방해 대단위 자료를 효율적이고 투명하게 처리하는데 그 적합성이 확인되어 왔다(Burscher, Odijk, Vliegenthart, Rijke & Vreese, 2014; Kananovich, 2018; Opperhuizen, Schouten, & Klijn, 2019). 프레임 분석을 위한 지도기계학습은 인간코더가 표집한 뉴스기사에 대해서 확인되는 프레임의 존재 여부를 기록한 훈련(training) 데이터 셋에서 개별 어휘들의 등장 혹은 조합으로 도출된 언어적 특성으로 새로운 데이터 셋의 프레임 존재 여부를 예측 후 판별하는 방식이 대표적이다(De Grove, Boghe & De Marez, 2020). 다시 말해, 인간코더가 표집된 뉴스기사에 존재하는 프레임을 확인하고 판별한 규칙을 컴퓨터가 기계학습하여 이를 재현하는 성능을 검증하고 대단위의 모집단 자료에 대한 자동화 프레임 분석에 적용하는 방법이다. 따라서

2) ‘rtweet’ 패키지를 이용해 트윗에 대한 상태 아이디(status id), 게시 날짜, 계정 이름(screen name), 트윗 메시지(text) 등의 정보가 포함된 데이터 프레임 형식의 트윗 데이터를 다운로드 받았다.

본 연구는 프레임 탐지를 위해 활용되는 규칙이 어휘적 특성에 기반한다는 전제하에 각 뉴스기사에서 도출가능한 어휘적 특성을 독립변인으로 설정해 종속변인인 건강신념모델의 4가지 프레임이 각각 등장하는지의 여부를 예측하는 지도기계학습을 실행했다.

종속변인으로써 각 뉴스기사에 등장하는 프레임을 예측하기 위해 다음의 절차로 기계학습을 위한 표집된 뉴스기사의 헤드라인과 리드³⁾의 조합 2000건과 트윗 메시지 2000건에 대해 각 프레임 존재 여부를 기록한 훈련용 데이터 셋을 생성했다.⁴⁾ 이 데이터 셋은 언론사 간의 코로나19 보도량 차이가 프레임 분포에 영향을 미칠 가능성을 고려하여 각 언론사의 뉴스기사 500건씩을 빅인즈와 트위터 데이터 셋에서 단순 무작위 방식으로 표집하는 방법으로 구성했다. 그리고, 프레임 확인은 뉴스기사의 경우 헤드라인과 리드, 그리고 트위터의 경우 트윗 메시지에 기반해 진행했는데 기사의 보도내용을 이해하고 사안에 대한 평가를 내리기 위해 사용되는 인지적 틀짓기, 즉 프레이밍에 헤드라인이 중요한 역할을 수행한다는 연구결과(Andrew, 2007)를 고려해 분석단위를 결정했다. 또한, 제주지역의 신문들이 ‘국제자유도시’와 ‘평화의 섬’ 의제를 어떠한 프레임으로 다루는지를 분석하기 위해 관련기사의 헤드라인과 리드를 분석하여 프레임을 판별한 고영철과 최낙진(2006)의 연구도 참고했다. 또한, 트윗의 텍스트 메시지는 링크로 첨부된 언론사의 기사를 소비하는 사용자의 뉴스에 대한 인식에 영향을 미치는 프레이밍 효과를 가질 것으로 보고 프레임을 분석했다. 이러한 분석단위의 개념화를 바탕으로 2명의 인간코더가 건강신념모델에서 도출된 심각성, 취약성, 이득, 장애 프레임을 각각 뉴스기사에서 확인하고 그 존재 여부를 판단하기 위해 분석항목에 대한 조작적 정의를 체계적이고 반복적으로 숙지하는 과정을 거쳤다.⁵⁾ 그리고 코더 간 신뢰도 검증을 위해 훈련용 데이터 셋 중 150건에 대한 네가지 프레임 확인 결과의 상호간 일치정도를 측정하였고, Cohen’s Kappa 계수가 심각성은 0.76, 취약성은 0.96, 이득은 0.90, 장애는 0.96으로 그 평균값이 0.89로 나타나 신뢰할 만한 수준인 것으로 확인됐다.

3) 빅인즈는 각 기사의 헤드라인은 물론 본문의 앞머리 200자를 제공하고 있는데, 본 연구에서는 이를 기사의 리드로 조작화하여 분석하였다.

4) 기계학습의 훈련용 데이터 셋을 구성하기 위해 추출된 뉴스기사의 헤드라인과 리드의 조합 혹은 트윗의 내용이 보건 복지부나 질병관리청 등 정부 기관에서 발표한 보도 자료에 기반하여 작성되어 상당히 비슷한 내용으로 작성된 사례들은 첫 1건을 제외한 나머지를 제거하였고, 키워드를 포함하고 있지만 코로나19와 관련한 뉴스보도와 직접적 연관이 없는 사례들도 제외하였다.

5) 2인의 인간코더는 언론학 전공자로 뉴스 텍스트에 대한 내용분석 경험이 있는 석사 과정 학생들이다. 코더 훈련을 위해 본 연구자들은 코더들에게 코딩의 목적과 각 프레임에 대한 개념과 조작적 정의를 설명한 후, 실제 사례에 대한 예비 검사(pilot-test)를 통해 측정방법을 공유했다. 그리고 연역적으로 도출된 심각성, 취약성, 이득, 장애 프레임에 대한 조작적 정의 항목을 활용해 2000건의 뉴스기사와 2000건의 트윗 메시지에 대해 각 프레임의 출현 여부를 각각 측정하였고, 코더 간 측정값이 불일치하는 사례에 대해서는 연구자와의 토론으로 최종 측정값을 결정했다.

Table 1. Framing Typology of Health Belief Model in COVID-19 News Coverage

Frames	Definitions	Subcategories
Severity	This frame focuses on damages caused by the disease and infection	Fatality of the disease
		Serial infection
		Economic damage such as unemployment, close-down
		Social damage such as social blame, stigma
Susceptibility	This frame is characterized by the likelihood of being easily and widely exposed to infection threats	Aftereffect
		Risk of (mass) infection in everyday life
		Dangerous facilities such as church, gym, or sauna
Benefits	This frame stresses the rewards of performing a healthy behavior	Vulnerable groups by age, religion, or hobby
		Effectiveness of social distance
		Effectiveness of vaccination
Barriers	This frame refers to barriers to performing a healthy behavior	Incentives for social distance
		Social barriers such as family gathering or holiday going out
		Psychological barriers such as loneliness or fear
		Physical barriers such as shortage of masks or job responsibility

뉴스기사에서 제시되는 건강신념모델의 심각성, 취약성, 이득, 장애 프레임의 존재 여부를 각각 판별하기 위해 사용한 분석항목의 조작적 정의는 선행연구들(안순태·이하나, 2016; Jung & Brann, 2014; Quick & Bates, 2010)을 참고해 <Table 1>과 같이 마련되었다. 우선, 지각된 위협 차원에서의 심각성 프레임은 코로나19 확산의 위험이 국민의 삶 전반에 영향을 미치는 과정과 그 결과의 심각성을 부각하는 프레임으로 정의하여 일상적 감염의 위험과 이에 따른 파급효과로 구분하였다. 또 다른 위협 요인으로서의 취약성 프레임은 일반적 취약성과 집단적 취약성의 두 가지 경우를 모두 고려하여 코로나19의 확산에 따라 질병과 국민의 관련성을 부각하는 프레임으로 규정했다. 그리고 행동적 평가 차원에서의 이득 프레임은 건강행동을 취했을 때 행위자가 얻을 수 있는 이익을 강조한 프레임으로 장애는 건강행동을 하는 데 있어 행위자가 느끼는 어려움의 원인, 즉 건강행동을 가로막는 방해요소를 부각한 프레임으로 정의하였다.

3) 독립변인

종속변인으로서의 프레임을 예측하는데 사용된 독립변인은 4개 종합일간지와 코로나19 유행의 4시기를 각각 구별하는 더미변수(dummy variable)들과 뉴스기사의 헤드라인과 리드, 그리고 트위터 메시지에서 도출된 언어적 특성이다. 엔트만(Entman, 1993)에 따르면 뉴스 프레임은 기사의 텍스트에 등장하는 특정 키워드나 상투적인 표현으로 발현되기 때문에 헤드라인과 리드

에서 어떠한 어휘들이 사용되고 있는가로 프레임 예측을 하는 방법은 적합하다고 보였다. 특히, 본 연구는 프레임이 해당 이슈에 대한 언급에서 선택되는 단어와 비유(metaphor)로 발현된다는 전제로 대규모 텍스트에서 등장하는 단어들의 공기(co-occurrence) 패턴으로 그 의미 관계를 파악하여 분석했다(Sagi, Diermeier, & Kaufmann, 2013). 가령, “테러”라는 단어가 “범죄”가 아닌 “전쟁”이라는 프레임으로 사용되는 경우 “테러”와 함께 사용되는 혹은 가까이 등장하는 단어(공기어)들은 “범죄”보다는 “전쟁”과 공기하는 단어들과 보다 비슷하게 구성된다는 것이다(Sagi et al., 2013). 따라서 뉴스를 구성하는 단어들이 비슷한 맥락으로 사용될 때 즉, 공기어들이 비슷할 때 가까운 의미 관계를 가지고 있다는 전제로 같은 프레임을 구성하는 단어들의 사용을 추적하고자 다음의 과정으로 각 기사 및 트윗의 언어적 특성을 추출했다.

우선, 비정형 데이터인 뉴스 텍스트 내용의 의미와 직접적 관련 없는 문장 부호와 언론사명 혹은 코너 제목 등을 제거했다. 트윗 메시지의 경우 HTML 태그와 이모티콘은 물론 URL을 삭제했다. 또한, 언어적 의미가 구성되는 기본 단위로 단어를 설정하고 “Mecab” 형태소 분석기를 이용해 텍스트의 문자열을 형태소 단위로 토큰화(tokenization)하는 과정을 거쳤다.⁶⁾ 이후 단어(명사)들의 사용 패턴에서 프레임 판별을 위한 규칙이 발생한다는 전제로 단어주머니(bag of words)와 단어 임베딩(word embedding)의 방법으로 특성을 추출했다.⁷⁾ 단어주머니 방법으로 뉴스기사와 트위터 자료에서 등장하는 전체 단어들 중 헤드라인과 리드 조합 및 트윗 메시지에서 등장하는 단어들의 빈도수를 기록한 문서-용어 행렬(document-term matrix)을 생성했다. 그리고 전체 문서에서 드물게 등장하는 단어들은 프레임 판별에 유의미한 영향을 갖지 못할 것으로 판단하고 빈도수가 5회 이하인 단어들은 제거했다. 또한, 많은 문서에서 공통적으로 빈번히 등장하는 단어들은 관성적으로 사용되어 내용적 구분에 유용하지 않아 그 가중치를 낮추고 특정 문서에서만 많이 등장해 해당 문서의 특성을 주어 다른 기사들과의 차별성을 높이는 단어들의 가중치는 높이는 TF-IDF(Term Frequency-Inverse Document Frequency)를 적용하여 어휘 특성에 기초한 프레임 판별의 정확도를 높이하고자 했다. 더 나아가, 개별 어휘 단위를 넘어 복합명사나 구와 절과 같이 복수의 형태소가 결합된 바이그램(bigram)을 추출하여 보다 풍부한 언어적 특성에 기반한 프레임 판별의 성능을 비교했다.

6) “Mecab” 형태소 분석기의 성능은 다른 형태소 분석기와 비교했을 때 정확성 측면에서는 비슷하고 처리 속도에서는 큰 강점을 보여 한국어 텍스트 분석을 위한 토큰화 작업에 활발히 활용되어 왔다(Lee & Lim, 2019). 본 연구에서는 R 환경의 “RcppMeCab” 패키지를 이용했다.

7) 한국어 텍스트는 명사와 조사 또는 동사 어간과 어미가 결합되어 있어 외형적으로 의미 구분이 어렵기 때문에 많은 경우 명사만을 추출하여 의미분석을 하는 방식이 사용된다(이종혁·길우영, 2019).

단어 임베딩(word embeddings)은 비슷한 의미의 단어들은 같은 의미 맥락에서 함께 등장하는 단어들인 유사하다는, 즉 비슷한 문맥에서 등장한다는 전제로 단어들의 공기(co-occurrence) 패턴으로부터 포착된 의미적 관계를 다차원의 벡터 공간에 분산 표상(distributed representation)하는 방법이다(Mikolov, Sutskever, Chen, Corrado & Dean, 2013). 이러한 단어 임베딩으로 결과로써 단어벡터(word vector)를 도출하면 단어들간의 의미적 거리를 측정할 수 있다. 실제로, 단어들의 공기 패턴으로부터 표상화된 맥락 정보를 사용해 그 의미 관계를 측정하는 방법은 정치적 담화는 물론 뉴스보도에서 특정 이슈가 해석되는 경향을 추론하는데 유용한 방법으로 활용돼 왔다(Sagi et al., 2013; Sagi & Dehghani, 2014). 이에 본 연구는 단어 차원에서의 의미적 관계를 고려한 문서 분류를 위해 빅카인즈 데이터 셋의 키워드 항목⁹⁾에 표시된 단어 목록에 “GloVe” 알고리즘(Pennington, Socher, & Manning, 2014)을 적용하는 단어 임베딩 방법을 사용했다.¹⁰⁾ 우선, 수집된 전체 뉴스기사 본문에서 5회 이상 등장하는 43,876개 단어(명사)들의 목록을 추려 5개의 연속된 단어로 정의된 동일 맥락을 기초로 단어 간 공기패턴을 측정할 후, 100개의 차원으로 구성된 다차원 공간상에 각 단어의 의미 정보를 표현하는 벡터를 도출했다. 그리고 각 기사와 트윗에 등장하는 단어들에 대해 빈도수 대신 단어 임베딩으로 도출된 단어 벡터를 대입하고 각 문서의 단어 벡터들의 평균 벡터를 구하는 방식으로 각 문서에 대한 언어적 특성을 측정했다.

위와 같은 단어 임베딩 방법으로 측정한 기사와 트윗의 언어적 특성은 내용이 아닌 메시지 길이 및 단어 출현 확률의 차이로 발생하는 프레임 판별의 문제를 다음과 같이 극복했다. 우선, 각 문서에서 등장하는 모든 단어 벡터들의 평균을 구해 문서 특성을 도출함으로써 메시지 길이 차이가 아닌 단어들의 의미 정보로 문서 특성을 구성했다. 또한, 보도내용 전체에서 단어들의 맥락 정보를 도출해 유의어와 반의어와 같은 어휘적 의미 관계를 반영한 문서 특성을 구성하여 매체 간 단어 분포 차이를 상쇄했다. 예를 들어, “불안”이라는 단어는 같은 맥락에서 등장하는 단어들이 “걱정”이라는 단어와 같은 맥락에서 등장하는 단어들과 유사하기 때문에 “불안”이 등장한 헤드라인과 “걱정”이 등장한 트윗은 비슷한 특성값을 가지게 된다. 그리고 이러한 방식으로 기사와 트윗에서 등장하는 단어들의 특성값을 평균하여 문서의 언어적 특성을 도출함으로써 매체 간 단어 분포는 물론 메시지 길이의 차이로 인한 프레임 판별의 오류를 방지하고자 했다. 실제로, 기사와 트윗

9) 빅카인즈 서비스는 키워드 검색의 결과로써 해당 뉴스기사의 헤드라인과 앞머리 200자 뿐만 아니라 기사에 등장하는 모든 명사를 순서대로 추출하여 “키워드” 항목에 표시하고 있다.

10) R 환경의 “text2vec” 패키지를 이용하여 단어 임베딩을 실시했다.

에 대해 문서 단위로 측정된 평균 벡터 전체에 대해 측정된 코사인 유사도는 평균이 0.86($SD = 0.12$)이므로서 1이 똑같은 문서 간의 수치임을 고려할 때 기사와 트윗 간의 차이가 크지 않았다. 더욱이, 조선일보($M = 0.79, SD = 0.17$)를 제외하곤 동일 언론사의 기사와 트윗 유사도 평균이 전체 문서 간의 유사도 평균보다 더 높고 표준편차도 작았다(중앙일보 $M = 0.87, SD = 0.10$; 한겨레 $M = 0.88, SD = 0.08$; 경향신문 $M = 0.89, SD = 0.07$). 이는 헤드라인과 리드에서 도출한 기사의 언어적 특성과 트윗의 특성이 같은 언론사일 경우 그 차이가 줄어들고 있음을 의미하는 것으로 문서의 특성을 측정하고 비교하기 위한 단어 임베딩 방법의 적합성을 보여준다.

4) 모델링

본 연구는 건강신념모델에 기초해 코로나19 관련 뉴스보도를 구성하는 프레임이 해당 문서에서 등장하는지를 언어적 특성에 따라 판별하는 지도기계학습모델을 이용했다. 이를 위해, 각기 다른 조합으로 구성된 독립변인으로서의 문서 특성에 자동화 문서 분류 알고리즘을 적용한 프레임 판별이 인간 코더의 프레임 분석 결과와 얼마나 일치하는지를 측정하는 성능 비교를 수행하여 정확성(accuracy) 측면에서 가장 뛰어난 모델로 예측 작업을 진행했다. 사용된 알고리즘¹⁾은 문서 분류에 있어 우수한 성능으로 미디어 텍스트 분석에도 활발하게 적용되고 있는 나이브 베이즈(Naive Bayes, 이하 NB)와 서포트 벡터 머신(Support Vector Machines, 이하 SVM)이다(De Grove, Boghe, & De Marez, 2020). NB와 SVM은 특히 뉴스에서 관찰되는 성차별적 언어 특성(Leavy, 2019)이나 트윗 메시지에서 확인되는 혐오 표현(Vidgen & Yasseri, 2020)을 분류하는데 다른 알고리즘에 비해 상대적으로 뛰어난 성능을 보여 기사와 트윗에서 발견되는 프레임을 판별하는데 적합하다고 판단했다.

자동화 프레임 판별 모델은 인간코더가 프레임 판별 작업을 마친 2,000건의 뉴스기사 표본 중 무작위로 선택된 75%의 훈련 데이터 셋에 대한 학습으로 결정했다. 이 때, 신뢰성있는 모델 선택을 위해 훈련 데이터 셋을 동일한 크기의 10개 하위 데이터 셋으로 구분한 후 반복된 학습 결과의 평균 정확도를 교차 검증하여 최적의 판별 모델을 도출하는 10겹 교차검증(10-fold cross-validation)을 거쳤다.²⁾ 그리고 훈련 과정을 통해 최종적으로 선택된 최적의 판별 모델

11) 기계학습모델에는 R 환경의 "tidymodels" 패키지를 활용했다. SVM 모델은 "kernlab" 패키지를 통해 linear 모델에 비해 더 정확한 분류를 기대할 수 있는 RBF 커널(kernel)기반 기계학습 방법을 이용했다.

12) 10겹 교차검증에 사용된 10개의 하위 데이터 셋은 각각 동일한 언론사 비율로 구성되었다. 그러나 프레임 분포는 불균형 상태로서 교차 검증시 각 항목의 비율이 균형을 이루도록 조정하는 계층적 10겹 교차검증(stratified 10-fold cross-validation)을 실시했다.

을 뉴스기사 표본의 25%에 해당하는 시험(test) 데이터 셋에 적용하여 그 성능을 평가했다. 이 때, 문서분류 알고리즘은 분류 방식을 학습 과정에서 각 항목의 비율이 동일한 것으로 전제하기 때문에 훈련과 검증, 그리고 시험 과정이 적용되는 데이터 셋에서 각 프레임의 존재가 확인되는 비율이 절반씩 차지하게 표본을 조정하는 균형화 작업을 진행했다. 이러한 과정을 통해 선정된 최적 모델을 전체 뉴스기사 데이터 셋에 적용하여 각 프레임의 사용 여부를 판별하는 작업에 활용했다. 마지막으로, 신문기사의 프레임 판별에 적용한 최적 모델을 2,000건의 트윗 메시지의 프레임 판별에 적용하여 그 적합성을 평가하고 트위터를 통한 뉴스보도에서 차별성을 보이는 각 프레임의 언어적 특성을 살폈다. 특히, 기계학습모델로 프레임을 판별한 결과를 인간 코딩 결과와 비교하여 그 차이를 살핌으로써 트위터에서 나타나는 코로나19 뉴스보도의 차별성을 객관적이고 신뢰할 수 있는 방법으로 도출했다.

5. 분석 결과

1) 표본에 대한 인간 코더의 분석 결과

코로나19와 관련한 언론의 뉴스기사와 트윗에서 각각 2,000건씩 표집한 데이터에 대해 인간코더가 프레임 판별 작업을 실시하였고 그 분포를 파악했다(〈Figure 2〉 참조). 우선, 뉴스기사와 트윗에서 모두 심각성 프레임이 가장 많이 확인되고 있었고 취약성, 이득, 장애 프레임의 순서대로 등장하고 있음을 발견했다. 2,000건의 뉴스기사 표본 중 심각성 프레임이 934건(47%), 취약성 프레임은 786건(39%), 이득 프레임은 364건(18%), 장애 프레임은 582건(29%)에서 확인됐다. 마찬가지로 2,000건의 트윗 표본 중 심각성 프레임은 825건(41%), 취약성 프레임은 792건(40%), 이득 프레임은 288건(14%), 장애 프레임은 619건(31%)에서 확인됐다. 또한, 지각된 위협 차원에서의 심각성 및 취약성 프레임의 출현 빈도는 뉴스기사 표본의 경우 약 85%의 문서에서 트윗 표본의 경우 약 81%의 문서에서 확인되어 약 47%의 뉴스기사와 45%의 트윗에서 확인된 행동적 평가의 이득 및 장애 프레임에 비해 월등히 많은 비율로 제시되고 있음을 발견했다. 더 나아가, 이득과 장애 프레임은 특히 뉴스기사와 트윗에서 출현하는 빈도가 그렇지 않은 경우에 비해 현저히 낮은 불균형 데이터임을 확인했다. 따라서 인간코더의 프레임 판별 결과를 지도기계학습을 위한 훈련 데이터 셋으로 활용하기 위해 뉴스기사와 트윗에서 확인된 프레임의 출현 빈도를 각각 반영하여 동일한 수의 프레임의 사용이 확인되지 않은 뉴스기사와 트윗을 무작위 추출 방식으로 표집하는 균형화 작업을 실시했다.

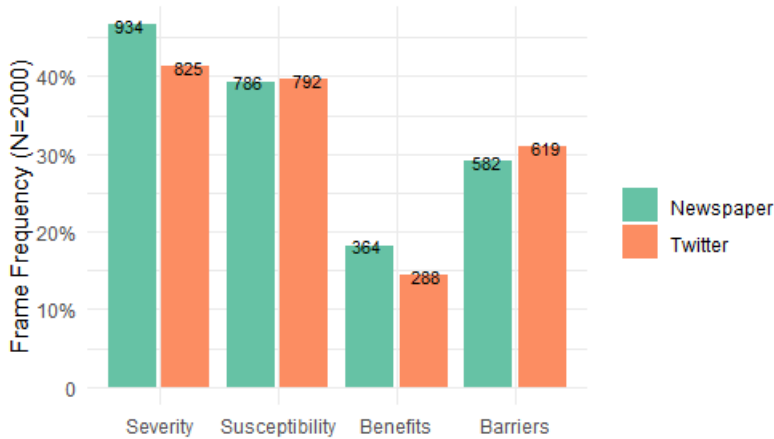


Figure 2. The frequency of frames in covid-19 new coverage, identified by human coders

2) 뉴스기사의 프레임 판별

인간코더가 분석한 뉴스기사 표본의 75%를 임의로 추출하여 훈련 데이터 셋으로 활용해 프레임을 판별하는 최적의 자동화 분류 모델을 선택하기 위해 10겹 교차검증을 실시하여 그 결과를 비교했다(〈Table 2〉 참조). 다시 말해, 훈련 데이터 셋의 뉴스기사에서 추출한 특성으로 프레임을 판별하는 성능 비교를 위해 10겹 교차검증을 실시하여 각 모델이 보인 판별 정확도(accuracy)의 평균값을 비교하여 최적 모델을 도출했다. 판별 정확도는 기계학습모델이 각 프레임이 등장하는 것으로 예측한 기사와 그렇지 않은 것으로 예측한 기사가 인간 코더에 의해 프레임 사용이 확인된 기사와 그렇지 않은 기사에 일치하는 비율이다. 이러한 모델의 성능 정확도를 비교한 결과, 뉴스기사 프레임 판별에 있어 SVM이 NB보다 더 뛰어난 정확도를 보이고 있었다. 그리고 헤드라인과 리드 조합의 텍스트에서 추출한 단어들의 빈도수 특성으로 프레임을 판별한 기준 모델(baseline model)에 비해 시기와 언론사 정보를 추가한 모델(Model 1)이 모든 프레임 판별에 있어 더 정확한 결과를 도출함을 발견했다. 그러나 뉴스기사의 어휘적 특성을 보다 풍부하게 추출하기 위해 바이그램을 추출해 구성한 어휘 특성 모델(Model 2)과 시기 및 언론사 정보를 추가한 모델(Model 3)은 취약성과 이득 프레임에서만 그 성능이 향상되었다. 그러나, 어휘 빈도수가 아닌 단어임베딩의 방법으로 뉴스기사의 특성을 추출한 모델(Model 4)은 이득 프레임 판별에서 그리고 여기에 시기와 언론사 정보를 추가한 모델(Model 5)은 심각성과 취약성, 그리고 장애 프레임 판별에 있어 가장 뛰어난 정확도를 보여주었다.

Table 2. Selecting the Training Model of Frame Classifier over 10 Folds

Feature set	Algorithm	Severity	Susceptibility	Benefits	Barriers
Baseline Model: Text (Headline & Lead)	NB	55.5%	86.3%	71.8%	64.2%
	SVM	58.6%	87.9%	76.3%	66.2%
Model 1: Text + Phase + Organization	NB	56.1% (0.6%)	86.5% (0.2%)	72.0% (0.2%)	64.8% (0.6%)
	SVM	62.3% (3.7%)	88.2% (0.3%)	77.1% (0.8%)	66.8% (0.6%)
Model 2: Text (Bigram)	NB	54.6% (-0.9%)	86.7% (0.4%)	71.8% (0.0%)	64.0% (-0.2%)
	SVM	57.1% (-1.5%)	88.2% (0.3%)	79.7% (3.4%)	65.8% (-0.4%)
Model 3: Text (Bigram) + Phase + Organization	NB	56.3% (0.8%)	86.7% (0.4%)	72.1% (0.3%)	63.9% (-0.3%)
	SVM	61.6% (3.0%)	88.4% (0.5%)	78.4% (2.1%)	66.8% (0.6%)
Model 4: Word embeddings	NB	62.3% (6.8%)	87.3% (1.0%)	77.5% (5.7%)	67.8% (3.6%)
	SVM	65.0% (6.4%)	89.5% (1.6%)	82.8% (6.5%)	70.7% (4.5%)
Model 5: Word embeddings + Phase + Organization	NB	63.1% (7.6%)	87.3% (1.0%)	77.2% (5.4%)	67.6% (3.4%)
	SVM	67.8% (9.2%)	89.6% (1.7%)	82.1% (5.8%)	71.0% (4.8%)

Note. Average values of classification accuracy of 10-fold cross-validation. Net change in accuracy from baseline in parenthesis.

위에서 제시한 대로 훈련 데이터 셋을 이용해 각 모델의 성능을 검증했고, 각 프레임 판별에 가장 정확한 결과를 도출한 모델을 선택하여 시험 데이터 셋에 적용해 그 성능을 평가했다(〈Table 3〉 참조). 뉴스기사의 프레임 판별에 적용된 최적의 기계학습모델이 인간 코더에 비교해 신뢰할 만한 결과를 도출하는지 그 성능을 평가하기 위해 정확도는 물론 정밀도(precision)와 재현율(recall), 그리고 F1 스코어를 검토했다. 이러한 평가 지표 중 정밀도는 기계학습모델이 프레임이 제시되는 것으로 판별한 기사 중에 인간 코더에 의해 프레임이 확인된 건의 비율을 의미하고, 재현율은 인간 코더에 의해 프레임이 확인된 기사 중에 기계학습모델이 프레임의 출현으로 판별한 건의 비율을 뜻한다. 그리고 F1 스코어는 정밀도와 재현율을 결합한 지표이다. 이러한 성능 평가 지표를 바탕으로 건강신념모델에 기초한 프레임이 뉴스기사에서 나타나고 있는지에 대한 여부를 자동적으로 판별하는 기계학습모델의 성능을 검토했다. 그 결과, 취약성 프레임이 가장 높은 정확도로 판별되고 있었고 이득과 장애, 그리고 심각성 프레임의 순서로 뉴스기사에 대한 정확한 프레임 판별이 이뤄지고 있음을 발견했다. 무엇보다, 심각성 프레임을 제외하고 모든 프레임 판별의 정밀도가 70%를 상회하고 있어 자동화 모델이 신뢰할 만한 성능을 보여주고 있음을 확인했다(Van Rijsbergen, 1979). 물론, 심각성 프레임 판별 정확성의 저조는 포괄적인 조작적 정의로 인해 인간 코더 간의 판별 불일치가 높았던 점에 기인하는 것으로 판단된다. 또한, 이득과 장애 프레임 판별의 경우 재현율이 정밀도에 비해 상대적으로 떨어지고 있음을 발견했는데 이는 시험 데이터 셋에서 포착된 뉴스기사의 특성이 훈련 데이터 셋으로 학습된 모델에 충분히 반영되지 못하고 있음을 의미한다. 그리고

그 원인으로는 이득과 장애 프레임의 경우 감염 확산의 위기 상황이 지속되면서 건강행동의 효용에 대한 평가가 변화했고, 또한 백신(치료제) 도입과 사용에 따라 새로운 정보를 토대로 뉴스기사의 프레임이 재구성되었지만 훈련 과정에서 과소 평가됐을 가능성을 추론할 수 있다.

Table 3. Performance of Frame Classifiers on Test Data from Headline and Lead

	Severity	Susceptibility	Benefits	Barriers
Accuracy	68.7%	88.8%	80.2%	72.8%
Precision	68.8%	87.3%	88.6%	77.5%
Recall	68.2%	90.8%	68.9%	66.7%
F1 Score	68.5%	89.0%	77.5%	71.7%

3) 트윗의 프레임 판별

뉴스기사의 프레임을 판별하는데 사용한 기계학습모델을 인간 코더에 의해 프레임이 판별된 트윗 데이터 셋에 적용하여 그 성능을 평가했다. 그 결과, 모든 프레임에 있어 판별 정확도가 떨어졌음을 발견했는데 이는 뉴스기사의 헤드라인과 리드의 조합으로 이루어진 텍스트에서 추출된 프레임의 언어적 특성은 트윗 메시지에서 나타나는 프레임의 특성과 분명한 차이를 보여주고 있음을 의미한다(Table 4 참조). 특히, 행동적 평가 영역인 이득과 장애 프레임에 비해 지각된 위협으로서의 심각성과 취약성 프레임 판별의 정확도 저하가 두드러졌다. 이는 트위터에서 언론이 코로나19로 인한 위협을 보도하는 방식에 차이가 있음을 드러내는 결과로써 심각성 및 취약성 프레임의 언어가 이득과 장애 프레임보다 더 소셜미디어 환경에서의 뉴스 생성 원리가 가진 차별성이 반영된 결과로 추론된다.

Table 4. Performance of Frame Classifiers on Test Data from Tweet Message

	Severity	Susceptibility	Benefits	Barriers
Accuracy	59.7% (-9.0%)	79.6% (-9.2%)	71.9% (-8.3%)	68.2% (-4.6%)
Precision	58.6% (-10.2%)	87.9% (0.6%)	68.1% (-20.5%)	60.0% (-17.5%)
Recall	59.9% (-8.3%)	75.4% (-15.4%)	73.7% (4.8%)	71.8% (5.1%)
F1 score	59.2% (-9.3%)	81.2% (-7.8%)	70.8% (-6.7%)	65.3% (-6.4%)

실제, 심각성 프레임은 정밀도와 재현율에 있어 그 성능이 각각 10.2%와 8.3%씩 저하되어 뉴스기사의 헤드라인과 리드에서 표현된 방식은 트윗 메시지에서 확인된 심각성 프레임의 그것과 상당한 차이를 보이고 있음을 알 수 있다. 정밀도의 하락은 기계학습모델이 특정 프레임들

트윗 메시지에서 확인한 것으로 판별했지만, 인간 코더의 판별은 달랐던 경우가 많았다는 의미로써 뉴스기사의 프레임에 사용된 언어적 특성이 트위터에서는 다르게 나타나고 있음을 보여준다. 이러한 정밀도의 저하는 프레임 판별에 있어 거짓 양성(false positives)의 사례가 증가했을 때 나타나는 결과이므로 이에 해당하는 대표적 사례들을 검토하여 트위터를 통한 코로나19 관련 뉴스보도의 특징을 살펴봤다.

1. 예산안이 통과되면 전국에서 처음으로 재난기본소득이 지급된다. 코로나19 여파로 정상적인 경제활동이 막힌 비정규직 노동자 등 5만여 명이 대상이다.

<https://t.co/XJJep5HQvE> 경향신문, 2020/03/10

2. 코로나19 상황이 확인한 2차 확산기에 접어들면서, 세계 경제 전망도 다시 어두워지고 있다. 코로나19 추세에 따라 경제 전망이 낙관과 비관 사이를 오가는 것은 지난 9개월 동안 반복된 현상이다. V, W, U, K형 회복 전망 중 세계 경제 회복 전망 정답은?

<https://t.co/CBhwLd4yDH> 한겨레, 2020/10/02

위의 두 트윗 메시지는 기계학습모델이 가장 높은 확률로 심각성 프레임으로 예측했으나 인간 코더의 판별 결과는 그렇지 않았던 대표적 거짓 양성의 사례이다. 이들에게서 공통적으로 발견되는 특징은 메시지의 내용은 코로나19로 인한 경기침체와 극복 방안에 대한 논의로써 심각성 프레임의 낮은 정확성이 발생하게 된 원인을 보여준다. 즉, 질병 감염이 개인의 신체에 미치는 영향 혹은 이로 인한 사회적 비난과 경제적 손실은 물론 이로 인한 심리적 고통 등이 심각성 프레임의 주된 특성인데 위 메시지는 개인적 측면을 넘어 사회적 차원에서 감염증 대유행의 파급효과에 주목하는 보도로써 이에 대한 명확한 구분이 인간 코딩은 물론 기계학습 과정에서 이루어지지 않았음을 의미한다.

또한, 재현율의 저하는 인간 코더가 트윗 메시지에서 확인한 프레임을 기계학습모델이 예측하지 못한 경우, 즉 거짓 음성(false negatives) 사례들이 증가했음을 나타낸다. 따라서 기계 예측과 인간 판별 간의 큰 격차를 보이는 아래의 대표적 거짓 음성 사례들에 기초해 같은 심각성 프레임의 뉴스보도가 트위터 공간에서 어떻게 다른 특징을 보이는지 고찰했다.

1. 코로나19 확진자들의 후유증을 조사한 결과 3개월 뒤엔 탈모와 운동 때 숨찬 증상이, 6개월 뒤엔 피로감이 상대적으로 높게 나타난 것으로 조사됐다. 일부 확진자들에게선 폐 기능 저하나 폐 섬유화도 관찰됐고, 정신적 우울감과 외상 후 스트레스 장애 증상도 보였다.

<https://t.co/AxVxFanekF> 한겨레, 2021-01-14

2. “코로나를 빌미로 이웃 간 감시가 일상화되는 ‘스몰 브러더 시대’가 됐다.”

<https://t.co/gNKj4xyVIL> 조선일보, 2020-12-30

위 거짓 음성 사례를 통해 심각성 프레임의 뉴스기사 헤드라인 및 리드에서 추출한 특성을 학습한 자동화 판별 모델로는 같은 프레임의 트윗 메시지를 구별해 낼 수 없는 원인을 살펴보았다. 즉, 트위터 공간에서만 발견되는 심각성 프레임의 뉴스보도가 보이는 특성을 다음과 같이 고찰했다. 첫째, 트윗 메시지는 뉴스의 헤드라인과 리드에서 발견되는 전체 기사에 대한 개괄 및 요약보다는 자극적인 표현과 극단적 예시를 제시함으로써 사용자의 주목을 이끌고자 했다. 가령, 코로나19가 후유증을 발생시킬 수 있다는 뉴스의 요점이 아닌 후유증의 구체적 예시를 열거하는 방식이다. 둘째, 링크로 연결된 언론사의 웹 페이지로 트위터 사용자의 방문을 유도하는 클릭미끼(clickbait)로 트윗 메시지가 활용된다는 점이다. 예를 들어, “코로나를 빌미로 이웃 감시가 일상화되는” 이유와 사례에 대한 언급이 없는 메시지는 사용자의 궁금증을 유발하고 내용 확인을 위한 링크 클릭을 유도할 수 있다는 점이다.

한편, 취약성 프레임의 경우 정밀도의 차이는 미미한 반면 재현율에 있어 프레임 판별의 성능이 크게 저하된 점을 고려할 때 트위터 공간에서만 인간 코더에 의해 확인된 뉴스 프레임의 언어적 특성이 존재하고 있음을 알 수 있다. 아래의 두 트윗 메시지는 인간 코더가 취약성 프레임을 확인했음에도 기계학습모델이 가장 높은 확률로 같은 프레임의 출현을 부정한 거짓 음성 사례로써 뉴스기사로는 예측할 수 없는 트위터 보도 프레임만의 특성을 보여준다.

1. 수많은 미국인이 자신은 이미 코로나19에 걸렸을 것으로 확인하는 현상이 나타나고 있다고 합니다. <https://t.co/TSLf577SNc> 중앙일보, 2020-05-07
2. “전국 어디서나, 어느 공간에서나, 누구나 코로나19에 노출될 수 있는 위험한 상황입니다.” <https://t.co/znNLRaAjTa> 한겨레, 2020-08-21

위 사례에서 공통적으로 발견할 수 있는 특징은 메시지의 작성 방식이 짧고 건조하게 표현되는 일반 스트레이트 기사와는 달리 방송 리포트 형식의 존칭 구어체로 구성됐다는 점이다. 특히, 첫 번째 사례는 링크된 기사에는 존칭 없는 문어체로 작성되어 있는 문장을 존칭을 붙인 구어체로 변환한 점이 주목할 만한 특징으로 이는 트윗 작성자가 뉴스를 전달하는 방식을 방송 리포트와 같이 글이 아닌 말로 전달하는 형태로 친근하고 편안한 정보전달에 주안점을 두고 있음을 알 수 있다. 또한, 인용문으로 트윗 메시지를 대신한 두 번째 사례는 정은경 중앙방역대책 본부

장의 언급만 트윗 내용으로 전달해 책임당사자의 주관적 판단이 드러나는 의견으로 심각성 프레임 을 나타내고 있다. 이는 뉴스 정보의 전달방식에 있어 사실 위주의 객관적 보도가 아닌 가치판 단이 개입된 감정적 언어가 강조되는 형태로써 언론사의 뉴스보도가 소셜미디어에서는 개인화된 감정표현이 바이럴 확산에 유리한 특성에 맞춰져 있음을 제시한다.

추가적으로 주목해야 할 부분은 이득과 장애 프레임의 경우 뉴스기사 판별 모델의 성능이 정밀도에서만 크게 저하되고 재현율에 있어서 오히려 상승했다는 점이다. 즉, 기계학습모델이 트 윗 메시지에서 이득 및 장애 프레임이 사용된 것으로 판별한 사례들 중 상당수가 인간 코더의 판 단과는 달랐던 반면, 인간 코더에 의해 확인된 프레임 출현 사례들 대부분을 기계학습모델이 동일 하게 예측했다는 결과다. 이는 거짓 양성에 해당하는 사례들이 증가했다는 의미로써, 뉴스기사 헤 드라인과 리드에서 확인된 이득 및 장애 프레임의 특성이 동일 프레임의 트윗 메시지에서 도출한 특성을 정확히 반영하지 못하고 있음을 보여준다. 따라서 기계학습모델이 높은 확률로 이득 및 장 애 프레임의 트윗 메시지로 판별했지만, 인간 코더의 결과는 반대였던 사례들을 살펴보았다.

1. 정세균 국무총리가 29일 오후 4시 30분 정부서울청사에서 코로나19 관련 긴급 기자 간담회 를 연다고 총리실이 밝혔다. 수도권에서 사회적 거리두기 2단계가 일주일째 유지되는데도 코로나 19 확산세가 꺾이지 않고 있어, 추가적인 방역 강화 방침이 발표될지 관심이 모인다.
<https://t.co/wrt1FNCwy0> *한겨레 2020-11-29*
2. 코로나19 대유행 속에 세계의 은행들이 1245억달러(약 135조원)의 수수료 수익을 벌어들 였다. 수수료 수익으로는 역대 최대치다.
<https://t.co/b9nkcotNrc> *경향신문 2020-12-30*

위에서 제시된 트윗 메시지는 이득 프레임에 대한 대표적 거짓 양성 사례이다. 첫 번째 사 례의 경우 사회적 거리두기를 통해 감염 확산을 방지하는 이득의 차원에서 코로나19 보도를 구 성한 것이 아닌 오히려 그러한 방역 방침이 기대되는 효과를 내고 있지 못한 상황을 지적하고 있 다. 이는 기계학습모델이 사회적 거리두기라는 건강행동과 그 영향에 대한 언급을 바탕으로 이득 프레임의 출현을 예측한 결과로 추론된다. 두 번째 사례 역시 코로나19로 인한 세계적 위기에서 은행은 최대치의 수수료 수익을 벌어들였다는 보도로 개인의 건강행동이 가져올 혜택과는 무관 한 측면을 다루고 있다. 이득의 주체가 개인이 아닌 은행 기관이기 때문이다. 결국, 기사의 헤드 라인과 리드에서 파악된 이득 프레임의 어휘적 특성이 트위터를 통한 코로나19 보도에서는 내용 적 의미의 변화를 만들고 있음을 알 수 있다. 다시 말해, 트위터의 뉴스 속성은 언론보도의 프레

임을 반영하기도 하지만 시안에 대한 초점을 바꾸거나 다른 관점을 제시하는 등의 내용으로 보다 사용자들의 이목과 참여를 유도한다는 것이다.

1. 연구팀은 “어린이의 권리와 기본적 필요를 충족시키기 위해 공중보건 개입은 코로나19 봉쇄의 초점을 등교 중지에서 학교의 사회적 가치를 달성하고 교육을 지속하는 데까지 넓혀야 한다”고 밝혔다. <https://t.co/XiaWKtM3si> *경향신문 2021-01-20*
2. 식당과 상가는 손님이 없어 한산하다. 모임과 행사도 잇따라 취소된다... ‘신종 코로나’가 아니라 불안과 공포 때문에 모두 망하겠다는 얘기가 과장이 아니다
<https://t.co/VPffd9SLTx> *한겨레 2020-02-11*

마지막으로, 기계학습모델이 판별한 장애 프레임의 트윗 메시지 중 대표적 거짓 양성으로 파악된 위의 사례들에 기초해 그 특징을 고찰했다. 장애 프레임은 건강행동에 수반되는 비용과 방해요소에 대한 강조로써 거리두기 등의 방역 방침은 개인의 권리는 물론 등교 교육과 같은 사회적 필요를 제한하는 어려움을 발생시키고 있기 때문에 이러한 언급은 판별에 중요한 특성으로 작용했을 가능성이 크다. 하지만, 트윗 메시지의 내용은 건강행동에 대한 어려움보다는 건강행동의 목표가 중요한 사회적 가치를 훼손하는 방향으로 설정되어야 함을 역설한다. 즉, 이득 프레임의 거짓 양성 사례와 마찬가지로 뉴스기사의 헤드라인과 리드에서 장애 프레임 강조하는 주요한 언어적 특성이 트위터에서는 해당 프레임의 차원으로만 다뤄지는 것이 아닌 오히려 건강행동으로 인한 사회적, 경제적 부담에 대해 초점을 맞추고 있음을 보여준다. 결국, 국내 언론의 코로나19 관련 보도에서 강조되는 건강 행동 프레임이 트위터에서 그대로 재현되기보다는 시안에 대한 뚜렷한 입장과 주장을 제시하는 방식으로 프레임의 재구성을 구현하고 있음을 제시한다. 이러한 결과는 소셜미디어 사용자의 공감이나 공유가 뉴스 유통에 필수적인 조건에서 언론사 역시 중립적이고 객관적이기 보다는 편향된 언어의 명확한 프레임으로 같은 시각의 이용자들에게 소구하여 적극적 참여를 유도하고 있음을 시사한다.

6. 결론과 논의

본 연구의 결과와 함의를 논의하기에 앞서 코로나19 관련 뉴스의 언어적 특성으로 프레임을 판별하기 위한 지도기계학습모델의 방법적 의미를 설명하고자 한다. 우선, 대단위 뉴스 본문에서

발견되는 공기어 패턴에 기초해 설정된 단어들의 의미 관계에 따라 언어적 특성을 도출하는 단어 임베딩 방법을 자동화 뉴스 프레임 분석에 적용해 측정의 신뢰도 및 타당성을 높였다. 그리고 분류 알고리즘으로 프레임 출현 가능성을 예측하는 확률 기반 모델을 활용하여 계량적 분석기법으로서 기사와 트윗에서 프레임이 발현되는 차이를 객관적으로 측정하여 인간 코더들의 주관적 해석에 따른 내용분석에 의존하는 분석방법의 한계를 극복하고자 했다. 물론, 이러한 자동화 프레임 판별 모델의 정확성은 프레임 간 격차가 존재했다. 가령, 심각성과 장애 프레임의 경우 취약성과 이득 프레임에 비해 모델의 성능이 낮은 결과를 보였는데 이 결과는 컴퓨터를 이용해 뉴스에서 프레임을 자동적으로 판별하는 방법의 타당성 부족을 의미할 수 있다. 그러나 심각성 프레임 판별의 부정확성은 코더 간 판별의 불일치에 기인하는 것으로 훈련 데이터 셋이 일관적인 기계학습의 조건을 제공하지 못한 점이 고려될 필요성이 있다. 더 나아가, 본 연구는 이러한 결과를 지도기계학습모델에 기반한 프레임 분석의 한계로 평가하기보다 다른 프레임과의 차별성을 객관적으로 진단할 수 있는 방법으로 사용할 수 있음을 드러냈다. 즉, 심각성과 장애 프레임은 취약성과 이득 프레임에 비해 프레임에 대한 추상적 개념을 식별하기 위한 실체화 된 언어적 특성이 다양할 뿐더러 프레임이 표현되는 방식 또한 복잡하게 사용되고 있다는 특성이 있음을 제시한 것이다.

또한, 본 연구는 소셜미디어 뉴스가 코로나19에 대한 건강 관련 프레임에 미치는 영향을 파악하기 위해 기사의 헤드라인과 리드에서 확인된 언어적 특성으로 프레임을 판별한 모델로 트윗에서의 프레임 출현을 예측하였고 그 정확성을 검증했다. 이를 통해 기사에서 확인된 프레임의 특성이 트윗 메시지에서 얼마나 유사하게 발현되는지를 평가했고 그 예측 성능의 변화가 어떠한 언어적 차이 혹은 의미적 변화에 기인하는지를 추적했다. 결국, 전반적인 모델 성능의 하락을 발견해 트위터에서 뉴스 프레임을 표현하는 어휘의 종류와 구성 방식에 많은 차이가 있음을 포착했다. 또한 본 연구는 프레임 차이의 원인을 추론하고자 모델의 정밀도와 재현율을 통해 인간 코딩과 알고리즘의 예측이 불일치한 사례들을 살펴보았다. 가령, 정밀도의 저하는 인간 코딩과 달리 알고리즘 예측으로 프레임이 출현한 것으로 판별된 사례인 거짓 양성이 증가한 결과로써 기사의 헤드라인과 리드에서 도출한 프레임의 어휘적 특성이 트위터에서는 다른 의미를 표현하기 위해 사용된 경우가 늘어났음을 의미한다. 이는 헤드라인과 리드에서 도출된 어휘들이 비교적 고정된 방식으로 프레임을 구성하는 반면 트위터에서는 그러한 어휘들이 보다 유연하고 자유로운 방식으로 다양한 의미를 전달하고 있음을 시사한다. 반면, 재현율의 저하는 인간 코딩이 트윗 메시지에서 확인한 프레임의 출현을 알고리즘 판별 모델은 예측하지 못한 거짓 음성 사례가 증가한 결과로써 트위터 뉴스의 차별성을 보여주는 방식으로 프레임이 사용되고 있음을 나타낸다. 이러한

관점에서 본 연구는 기계학습모델을 인간 코딩의 대용으로만 사용한 것이 아닌 프레임에 대한 부정확한 예측에 영향을 미친 메시지 특성을 탐색할 수 있는 방법으로 활용해 결과의 함의를 도출함으로써 프레임 분석 방법의 지평 확장에 기여하고자 했다(De Grove et al., 2020).

본 연구는 이처럼 지도기계학습모델을 이용해 자동화 프레임 분석이 인간 코딩을 그대로 재현하는가에 대한 타당성 평가에 그치는 것이 아니라 인간 코딩의 결과는 물론 다른 속성의 미디어 간에 프레임을 구성하는 언어적 특성이 차이가 나는 원인을 체계적이고 객관적으로 진단하는 방법을 제시했다. 이는 대단위 기사자료에서의 프레임을 자동화 방식으로 분석하는 방법론의 확장 차원에서 지도기계학습모델의 적용 가능성을 검증했을 뿐만 아니라 인간 코딩의 주관적 해석과 평가가 아닌 신뢰성과 재현 가능성을 수반하여 뉴스 프레임의 특징을 객관적으로 파악했다는 의의를 가진다.

이상의 방법적 의의를 바탕으로 본 연구는 코로나19 관련 언론보도의 프레임을 구성하는 언어적 특성과 이것이 트위터에서 차별적으로 발현되는 양상을 바탕으로 다음의 결론을 도출했다. 첫째, 본 연구는 코로나19 관련 뉴스기사의 표본을 대상으로 인간 코딩이 건강신념모델에 기초한 네 가지 프레임을 분석했다. 그리고 기사의 헤드라인과 리드에서 도출된 언론보도의 프레임 비중은 트윗 메시지에서도 눈에 띄는 차이가 없음을 확인했다. 코로나19의 심각성을 필두로 취약성을 강조하는 지각된 위협 프레임이 중점적으로 강조된 반면 이득과 장애의 행동적 평가 프레임은 상대적으로 부각되지 못한 것이다. 이는 뉴스 이용자가 갖는 코로나19에 대한 인식과 신념이 건강행동의 비용과 편익 차원에서 이뤄지기보다 질병과 감염에 대한 위협의식에서 형성될 수 있음을 의미한다. 그리고 뉴스기사의 헤드라인 및 리드에서 발견된 이러한 프레임 양상은 트윗 메시지에서도 재현되는바 소셜미디어가 뉴스 프레임의 선택에 있어 미치는 영향은 제한적임을 제시한다.

둘째, 본 연구는 코로나19 관련 기사의 헤드라인과 리드에서 프레임의 출현 여부를 예측하는 지도기계학습 모델을 활용해 심각성 프레임은 68.7%, 취약성 프레임은 88.8%, 이득 프레임은 80.2%, 그리고 장애 프레임은 72.8%의 정확성으로 판별해 심각성과 장애 프레임의 판별에 있어 기계학습모델의 정확성은 취약성과 이득 프레임에 비해 떨어짐을 발견했다. 이는 프레임을 구성하는 언어적 특성이 확연한 차별성을 보이는가의 문제로서, 코로나19의 심각성과 건강행동의 어려움을 강조하는 내용은 일관된 뚜렷한 특징을 보이지 않고 있음을 의미한다. 그리고 이러한 프레임 특성은 그 추상적 개념이 다양한 어휘와 표현 방식의 사례에 기반하고 있음을 시사한다. 또한, 위기의 양상에 따라 혹은 언론사별 차이에 의해 프레임의 언어적 패턴이 다양하게 나타나고 있음을 보여준다. 반면 감염의 취약성이나 건강행동에 따른 이익을 다룬 뉴스는 그 주제

적 특성을 표현하는 분명한 어휘군과 통일된 표현 방식이 사용되고 있다는 뜻으로 이러한 프레임의 언어적 특성이 언론사와 위기 양상에 따라 변화하는 폭은 상대적으로 작다는 것이다. 이 결과는 취약성과 이득 프레임은 일관된 방식으로 뉴스 이용자에게 제시되는 반면 심각성과 장애 프레임은 코로나19 확산의 양상에 따라 그 주제적 특성이 변화해갈 것임을 암시한다.

마지막으로, 본 연구는 뉴스 전파에 있어 그 영향력이 커지고 있는 소셜미디어의 역할을 진단하고자 트위터에서 발견되는 프레임의 특성을 탐색했고 다음의 함의를 도출했다. 우선, 트위터는 언론사의 뉴스 프레임의 선택에 미치는 영향은 제한적이지만 이를 구성하는 언어적 특성은 다르게 구성되어 있었다. 즉, 언론보도에서 강조하는 프레임은 트윗 메시지에서 그 현저성이 확인되었으나 뉴스의 헤드라인과 리드로 혼란된 프레임 판별 알고리즘이 트윗 메시지에서 나타나는 프레임을 예측하는 데 있어 그 성능이 크게 저하됨으로써 내용적 구성에 차이가 있음을 보인 것이다. 이러한 결과는 언론이 트윗 메시지를 통해 뉴스보도를 전달하는 방식의 차이를 드러낸 것으로 소셜미디어의 신속성은 물론 이용자의 참여가 메시지 전파의 중요 기제로 작용하여 보다 개인적인 차원에서의 감정적인 언어 사용이 프레임에 자주 드러나고 있음을 제시한다. 가령, 헤드라인과 리드에서 확인된 프레임에 비해 트윗 메시지는 자극적인 표현과 극단적 사례를 부각하여 이용자의 관심을 이끈 후 덧붙인 링크를 클릭해 기사의 구체적 내용을 확인하도록 유도하는 경향이 그것이다. 또한, 개인의 주관적 의견이 강조되거나 사안에 대한 명확한 입장으로 구성된 프레임은 같은 시각을 가진 이용자들의 참여로 촉진되는 소셜미디어 뉴스의 확산 원리에 맞춘 트윗 메시지의 특성으로 발견되고 있었다. 이 결과는 갈수록 확장되어 가는 소셜미디어 환경에서의 언론이 프레임을 구성하는 방식이 변화하고 있음을 시사한다. 특히, 트위터를 통해 전파되는 뉴스보도는 이미지나 영상이 헤드라인 및 리드와 함께 링크로 제공되는 멀티미디어 콘텐츠이지만 작성된 트윗 메시지의 내용이 기사와는 구별된 방식의 프레임으로 이용자의 인식과 태도에 영향을 미칠 가능성에 대해 검증해야 할 필요성을 제기한다. 이는 본 연구가 방법론의 지평 확장으로 이론적 논의 발전에 기여한 점으로 지도기계학습모델을 프레임 분석에 적용하는 후속 연구에 참조를 제공할 것으로 기대한다.

References

- Al-Rawi, A. (2019). Viral news on social media. *Digital Journalism*, 7(1), 63-79.
- An, S., & Lee, H. (2016). Media guidelines for suicide prevention: Content analysis of news stories on depression according to health belief model. *Health and Social Welfare Review*, 36(1), 529-564.
- Andreeva, V. A., Unger, J. B., & Pentz, M. A. (2007). Breast cancer among immigrants: A systematic review and new research directions. *Journal of Immigrant and Minority Health*, 9(4), 307-322.
- Andrew, B. C. (2007). Media-generated shortcuts: Do newspaper headlines present another roadblock for low-information rationality? *Harvard International Journal of Press/Politics*, 12(2), 24-43.
- Apuke, O. D., & Omar, B. (2021). Fake news and COVID-19: Modelling the predictors of fake news sharing among social media users. *Telematics and Informatics*, 56, 101475.
- Bae, B.-G., Lee, B.-R., & Choi, S.-H. (2015, June). The analysis of association between social bigdata and earthquake. Paper presented at the annual meeting of the Korean Institute of Information Scientists and Engineers, Jeju: Jeju National University.
- Becker, M. H. (1974). The health belief model and sick role behavior. *Health Education Monographs*, 2(4), 409-419.
- Bennett, W. L., & Segerberg, A. (2013). *The logic of connective action: Digital media and the personalization of contentious politics*. New York, NY: Cambridge University Press.
- Berger, J., & Milkman, K. (2010). Social transmission, emotion, and the virality of online content. *Wharton Research Paper*, 106, 1-52.
- Burscher, B., Odijk, D., Vliegthart, R., De Rijke, M., & De Vreese, C. H. (2014). Teaching the computer to code frames in news: Comparing two supervised machine learning approaches to frame analysis. *Communication Methods and Measures*, 8(3), 190-206.
- Carpenter, C. J. (2010). A meta-analysis of the effectiveness of health belief model variables in predicting behavior. *Health Communication*, 25(8), 661-669.
- Champion, V. L., & Skinner, C. S. (2008). The health belief model. In K. Glanz, B. K. Rimer, & K. Viswanath (Eds.), *Health Behavior and Health Education: Theory, Research, and Practice (4th ed.)* (p. 45-65). San Francisco, CA: Jossey-Bass.
- Cho, H. L., & Jung, M. (2019). An exploratory content analysis of media coverage and framing on naturalistic medicine. *Health and Social Welfare Review*, 39(2), 332-357.

- Cho, M., & Lee, S. H. (2021). Analyzing news frames in the coverage of COVID-19: Data-driven approach to frame analysis. *Journal of Speech, Media and Communication Research*, 20(1), 65-107.
- Choi, J. W. (2012). A study on improving the delivery system of emergency disaster broadcast using new media. *Broadcasting and Media Magazine*, 17(3), 24-39.
- Choi, S.-H. (2020). Preventive measures during outbreak of coronavirus disease 2019. *Korean Journal of Medicine*, 95(3), 134-140.
- Clarke, J. N., McLellan, L., & Hoffman-Goetz, L. (2006). The portrayal of HIV/AIDS in two popular African American magazines. *Journal of Health Communication*, 11(5), 495-507.
- Coleman, R., & Thorson, E. (2002). The effects of news stories that put crime and violence into context: Testing the public health model of reporting. *Journal of Health Communication*, 7(5), 401-425.
- De Grove, F., Boghe, K., & De Marez, L. (2020). (What) Can journalism studies learn from supervised machine learning? *Journalism Studies*, 21(7), 912-927.
- Edy, J. A., & Meirick, P. C. (2007). Wanted, dead or alive: Media frames, frame adoption, and support for the war in Afghanistan. *Journal of Communication*, 57(1), 119-141.
- Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, 43(4), 51-58.
- Entwistle, V. (1995). Reporting research in medical journals and newspapers. *BMJ*, 310(6984), 920-923.
- Gamson, W. A., & Modigliani, A. (1989). Media discourse and public opinion on nuclear power: A constructionist approach. *American Journal of Sociology*, 95(1), 1-37.
- García-Perdomo, V., Salaverría, R., Kilgo, D. K., & Harlow, S. (2018). To share or not to share: The influence of news values and topics on popular social media content in the United States, Brazil, and Argentina. *Journalism Studies*, 19(8), 1180-1201.
- Goffman, E. (1974). *Frame analysis: An essay on the organization of experience*. New York, NY: Harper & Row.
- Ha, J.-H., & Rim, H. (2020). Does audience optimism transcend the severity of news frame during health pandemic outbreaks? *Journal of Communication Science*, 20(1), 191-229.
- Han, K. H. (2011). The impact of health belief components on the effectiveness of women's cancer prevention campaigns: A cross-national examination towards Korean and Japanese women. *The Korean Journal of Advertising and Public Relations*, 13(2), 377-413.
- Harcup, T., & O'Neill, D. (2017). What is news? News values revisited (again). *Journalism Studies*, 18(12),

1470-1488.

- Hermida, A., Fletcher, F., Korell, D., & Logan, D. (2012). Share, like, recommend: Decoding the social media news consumer. *Journalism Studies*, 13(5-6), 815-824.
- Hyman, R. B., Baker, S., Ephraim, R., Moadel, A., & Philip, J. (1994). Health belief model variables as predictors of screening mammography utilization. *Journal of Behavioral Medicine*, 17(4), 391-406.
- Janz, N. K., & Becker, M. H. (1984). The health belief model: A decade later. *Health Education Quarterly*, 11(1), 1-47.
- Jenkins, H. (2006). *Fans, bloggers, and gamers: Exploring participatory culture*. New York, NY: New York University Press.
- Jo, S. E., Shin, H. C., Yoo, S. W., & Roh, H. S. (2012). The study of factors affecting tuberculosis preventive behavior intentions: An extension of HBM with mediating effects of self-efficacy and fear. *Journal of Public Relations Research*, 16(1), 148-177.
- Jung, J., & Lee, D. (2012). A study on frame effect in elaboration likelihood perspective: Focusing on the cancer related news. *Korean Journal of Journalism & Communication Studies*, 56(6), 278-309.
- Jung, T., & Brann, M. (2014). Analyzing the extended parallel process model and health belief model constructs in texting while driving: News coverage in leading US news media outlets. *International Journal of Health Promotion and Education*, 52(4), 210-221.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263-291.
- Kananovich, V. (2018). Framing the taxation-democratization link: An automated content analysis of cross-national newspaper data. *The International Journal of Press/Politics*, 23(2), 247-267.
- Kim, E., Yu, H., & Han, K. (2019). Analysis of frame types in the income-led growth news reports. *Korean Journal of Communication & Information*, 96, 7-36.
- Kim, H. J. (2017). Psychological reactance against news articles on nuclear energy: Effects of prospect frames and issue frames. *Korean Journal of Journalism & Communication Studies*, 61(5), 130-164.
- Kim, J., & Cho, J. (2019). Investigation of effects of individuals social viewing of fine dust information obtained through social media on behavioral intentions of disease prevention: Application of health beliefs model. *Korean Journal of Broadcasting and Telecommunication Studies*, 33(4), 37-65.
- Kim, J., & Yu, H. (2012). The effects of gain- or loss-framed health news and exemplars on the perception of reported issues and prevention intention. *Korean Journal of Journalism & Communication Studies*,

56(1), 5-30.

- Kim, J., & Yu, H. (2012). The effects of gain- or loss-framed internet health news and replies on the perception of reported issues and prevention intention. *Korean Journal of Broadcasting and Telecommunication Studies*, 26(3), 176-217.
- Kim, J., Choi, J., & Park, D.-J. (2018). The influence of health information overload caused by media use on health information distrust. *The Korean Journal of Advertising and Public Relations*, 20(2), 37-63.
- Kim, J.-H. (2015). A study on public information service using Twitter: Focused on Twitters of major metropolitans. *Journal of Korean Library and Information Science Society*, 46(1), 115-133.
- Kim, M. H. (1997). Health belief model approach to health beliefs, attitude, and health behaviors concerning HIV/AIDS. *Journal of Korean Society for Health Education and Promotion*, 14(2), 125-147.
- Kim, S.-J., & Cha, H. (2009). The effect of public segmentation and message framing on the health risk communication: Applying anger activism model. *Korean Journal of Journalism & Communication Studies*, 53(2), 231-253.
- Klinger, U. (2013). Mastering the art of social media: Swiss parties, the 2011 national election and digital challenges. *Information, Communication & Society*, 16(5), 717-736.
- Koh, Y. C., & Choi, N. J. (2016). Investigation on the diversity of news reported by Jeju local newspapers: Regarding the news on 'Free International City' and 'Island of Peace'. *Journal of Communication Science*, 6(2), 5-42.
- Kuiken, J., Schuth, A., Spitters, M., & Marx, M. (2017). Effective headlines of newspaper articles in a digital environment. *Digital Journalism*, 5(10), 1300-1314.
- Leavy, S. (2019). Uncovering gender bias in newspaper coverage of Irish politicians using machine learning. *Digital Scholarship in the Humanities*, 34(1), 48-63.
- Lee, J. H., & Kil, W. (2019). News agenda classification and media diversity analysis using topic modeling: Based on news on the Presidential New Year Press Conference. *Korean Journal of Broadcasting and Telecommunication Studies*, 33(1), 161-196
- Lee, M.-K., & Lee, Y.-R. (2012). A news frame study of domestic Korean newspapers' coverage on animal infectious disease: Focused on the case of foot-and-mouth disease reporting among national & regional newspapers. *Journal of Communication Science*, 12(2), 378-414.
- Lee, S. H., & Lim, T. Y. (2019). Connective action and affective language: Computational text analysis of Facebook comments on social movements in South Korea. *International Journal of Communication*,

13, 2960–2983.

- Lee, Y. J. (2009). The use of Twitter service as social media for information sharing in emergency disaster response. *Journal of Korean Society of Societal Security*, 2(3), 16-18.
- McLeod, D. M., & Detenber, B. H. (1999). Framing effects of television news coverage of social protest. *Journal of Communication*, 49(3), 3-23.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger, (Eds.). *Advances in Neural Information Processing Systems 26* (pp. 3111-3119). Lake Tahoe, NV: Curran Associates, Inc.
- Newman, N. (2011). *Mainstream media and the distribution of news in the age of social media*. Oxford, UK: Reuters Institute for the Study of Journalism.
- Opperhuizen, A. E., Schouten, K., & Klijn, E. H. (2019). Framing a conflict! How media report on earthquake risks caused by gas drilling: A longitudinal analysis using machine learning techniques of media reporting on gas drilling from 1990 to 2015. *Journalism Studies*, 20(5), 714-734.
- Pan, Z., & Kosicki, G. M. (1993). Framing analysis: An approach to news discourse. *Political Communication*, 10(1), 55-75.
- Park, H. C., & Park, H. W. (2017). Disaster communication networks on Twitter: Gyeongju earthquake. *Journal of the Korean Data Analysis Society*, 19(1), 291-302.
- Park, J., Ryu, P.-M., & Oh, H.-J. (2017). A collecting model of public opinion on social disaster in Twitter: A case study in `humidifier disinfectant`. *KIPS Transactions on Software and Data Engineering*, 6(4), 177-184.
- Pennington, J., Socher, R., & Manning, C. D. (2014, October). Glove: Global vectors for word representation. In *Proceedings of The 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1532-1543). Doha, Qatar: Association for Computational Linguistics.
- Price, V., Tewksbury, D., & Powers, E. (1997). Switching trains of thought: The impact of news frames on readers' cognitive responses. *Communication Research*, 24(5), 481-506.
- Quick, B. L., & Bates, B. R. (2010). The use of gain-or loss-frame messages and efficacy appeals to dissuade excessive alcohol consumption among college students: A test of psychological reactance theory. *Journal of Health Communication*, 15(6), 603-628.
- Rhee, J. (2000). Frames, interpretations and communication effects: An overview and assessment of frame

- studies, *Media & Society*, 29, 85-153.
- Rhee, J. (2001). Impacts of news frames in the coverage of conflicting issues on individual interpretation and opinion. *Korean Journal of Journalism & Communication Studies*, 46(1), 441-482.
- Rhee, J., & Kim, S.-H. (2018). News frames in the coverage of fine-dust disaster: Application of Structural Topic Modeling. *Korean Journal of Journalism & Communication Studies*, 62(4), 125-158.
- Rosenstock, I. M. (1974). The health belief model and preventive health behavior. *Health Education Monographs*, 2(4), 354-386.
- Sagi, E., Diermeier, D., & Kaufmann, S. (2013). Identifying issue frames in text. *PLoS ONE*, 8(7), e69185.
- Sagi, E., & Dehghani, M. (2014). Measuring moral rhetoric in text. *Social Science Computer Review*, 32(2), 132-144.
- Scheufele, D. A. (1999). Framing as a theory of media effects. *Journal of Communication*, 49(1), 103-122.
- Shih, T. J., Wijaya, R., & Brossard, D. (2008). Media coverage of public health epidemics: Linking framing and issue attention cycle toward an integrated theory of print news coverage of epidemics. *Mass Communication & Society*, 11(2), 141-160.
- Stieglitz, S., & Dang-Xuan, L. (2013). Emotions and information diffusion in social media—Sentiment of microblogs and sharing behavior. *Journal of Management Information Systems*, 29(4), 217-248.
- Stryker, J. E. (2003). Articles media and marijuana: A longitudinal analysis of news media effects on adolescents' marijuana use and related outcomes, 1977-1999. *Journal of Health Communication*, 8(4), 305-328.
- Talosis-Garcia, M., & Davis, S. W. (2005). Information-seeking behavior of minority breast cancer patients: An exploratory study. *Journal of Health Communication*, 10(1), 53-64.
- Tanner-Smith, E. E., & Brown, T. N. (2010). Evaluating the health belief model: A critical review of studies predicting mammographic and pap screening. *Social Theory & Health*, 8(1), 95-125.
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211(4481), 453-458.
- Van Rijsbergen, C. (1979, September). Information retrieval: Theory and practice. In *Proceedings of the Joint IBM/University of Newcastle upon Tyne Seminar on Data Base Systems* (pp. 1-14). Newcastle, UK: University of Newcastle.
- Vidgen, B., & Yasseri, T. (2020). Detecting weak and strong Islamophobic hate speech on social media. *Journal of Information Technology & Politics*, 17(1), 66-78.

- Welbers, K., & Opgenhaffen, M. (2019). Presenting news on social media: Media logic in the communication style of newspapers on Facebook. *Digital Journalism*, 7(1), 45-62.
- Yanovitzky, I., & Blitz, C. L. (2000). Effect of media coverage and physician advice on utilization of breast cancer screening by women 40 years and older. *Journal of Health Communication*, 5(2), 117-134.
- Yu, J., & Keum, H. (2018). Twitter as a risk information source: The effect of update recency and social endorsement on credibility judgments. *Korean Journal of Broadcasting and Telecommunication Studies*, 32(1), 33-65.
- Zhang, H., Park, J., & Lee, K. (2020). A study of factors affecting preventive behavior intentions of overseas infectious disease based on health belief model: Focusing on the moderating effects of SNS eWOM. *The Korean Journal of Advertising and Public Relations*, 22(2), 265-302.

최초 투고일	2021년	04월 07일
게재 확정일	2021년	06월 04일
논문 수정일	2021년	06월 07일

부록

- 고영철·최낙진 (2006). 제주지역 신문 뉴스의 다양성에 관한 탐색적 연구: '국제자유도시'와 '평화의 섬' 관련 기사를 중심으로. <언론과학연구>, 6권 2호, 5-42.
- 김명혜 (1997). 건강신념모델을 적용한 AIDS 예방의 건강신념, 태도 및 건강행동. <보건교육건강증진학회지>, 14권 2호, 125-147.
- 김수진·차희원 (2009). 공중유행과 메시지 프레이밍이 건강위험 커뮤니케이션 태도에 미치는 영향: 분노행동주의모델(Anger Activism Model)의 적용. <한국언론학보>, 53권 2호, 231-253.
- 김은정·유홍식·한규준 (2019). 국내 신문의 문재인정부 소득주도성장에 대한 뉴스보도 프레임 유형 분석. <한국언론정보학보>, 96권, 7-36.
- 김종화·유홍식 (2012a). 건강보도에서 획득·손실 프레임과 예시가 이슈의 지각과 예방행위 의도에 미치는 영향. <한국언론학보>, 56권 1호, 5-30.
- 김종화·유홍식 (2012b). 인터넷 건강보도에서 획득·손실 프레임과 댓글이 이슈 지각과 예방행위 의도에 미치는 영향. <한국방송학보>, 26권 3호, 176-217.
- 김주미·최정화·박동진 (2018). 미디어 이용에 따른 건강정보 과부하가 건강정보 불신에 미치는 영향. <한국광고홍보학보>, 20권 2호, 37-63.
- 김지현 (2015). 트위터를 활용한 공공 정보서비스 연구: 주요 광역도시 트위터들을 중심으로. <한국도서관·정보학회지>, 46권 1호, 115-133.
- 김지혜·조재희 (2019). 미세먼지에 대한 소셜 미디어 건강정보 사회적 시정미가 질병예방행위의도에 미치는 영향: 건강신념모델의 매개모형 적용을 중심으로. <한국방송학보>, 33권 4호, 37-65.
- 김효정 (2017). 원자력 기사 프레이밍이 수용자의 심리적 저항에 미치는 영향: 심리적 저항 이론 (Psychological Reactance Theory)을 중심으로. <한국언론학보>, 61권 5호, 130-164.
- 박준형·류범모·오효정 (2017). 사회적 재난에 대한 트위터 여론 수렴 모델: '가습기 살균제' 사건을 중심으로. <정보처리학회논문지. 소프트웨어 및 데이터 공학>, 6권 4호, 177-184.
- 박효찬·박한우 (2017). 트위터 데이터를 활용한 재난 커뮤니케이션 네트워크 분석: 2016 년 경주 지진. <Journal of The Korean Data Analysis Society>, 19권 1호, 291-302.
- 배병걸·이보람·최선화 (2015, 6월). <소셜 빅데이터와 지진의 연관성 분석>. 한국정보과학회 한국 컴퓨터종합학술대회. 제주: 제주대학교.
- 안순태·이하나 (2016). 자살예방을 위한 미디어 보도 방향: 건강신념모델을 통한 우울증 보도 내용 분석. <보건사회연구>, 36권 1호, 529-564.

- 유재원·금희조 (2018). 트위터의 업데이트 신속성과 사회적 지지가 정보원 신뢰도와 위험인식에 미치는 영향: 정보원 신뢰도의 매개효과를 중심으로. <한국방송학보>, 32권 1호, 33-65.
- 이민규·이예리 (2012). 국내 신문의 가축 전염병 위험 보도에 대한 프레임 연구: 중앙지와 지역지의 구계역 보도를 중심으로. <언론과학연구>, 12권 2호, 378-414.
- 이종혁·길우영 (2019). 토픽모델링을 이용한 뉴스 의제 분류와 미디어 다양성 분석. <한국방송학보>, 33권 1호, 161-196.
- 이영재 (2009). 긴급재난대응에서 정보공유를 위한 소셜 미디어(Social Media) 트위터(twitter) 서비스 활용. <한국재난관리표준학회지>, 2권 3호, 16-18.
- 이준웅 (2000). 프레임, 해석 그리고 커뮤니케이션 효과. <언론과 사회>, 29권, 85-153.
- 이준웅 (2001). 갈등적 이슈에 대한 뉴스 프레임 구성방식이 의견형성에 미치는 영향: 내러티브 해석 모형의 경험적 검증을 중심으로. <한국언론학보>, 46권 1호, 441-482.
- 이준웅·김성희 (2018). 미세면지 재해 보도의 프레임 분석: 구조적 주제모형(Structural Topic Modeling)의 적용. <한국언론학보>, 62권 4호, 125-158.
- 장해·박주식·이경식 (2020). 건강신념모델을 적용한 해외 감염병 예방 행동의도의 영향요인에 관한 연구: 외적 행위단서로서의 SNS 구전의 조절효과를 중심으로. <한국광고홍보학보>, 22권 2호, 265-302.
- 정재선·이동훈 (2012). 정교화 가능성 관점의 프레임 효과연구: 암 관련 보도기사를 중심으로. <한국언론학보>, 56권 6호, 278-309.
- 조민정, 이신행 (2021). 코로나19 관련 언론 보도 프레임 분석: 자료기반 자동화 프레임 추출 방법을 중심으로. <한국소통학보>, 20권 1호, 79-121.
- 조성은·신호창·유선욱·노형신 (2012). 결핵예방 행동의도에 영향을 미치는 요인에 관한 연구-자기효능감과 공포의 매개역할을 중심으로 한 건강신념모델의 확장. <홍보학 연구>, 16권 1호, 148-177.
- 조혜림·정민수 (2019). 자연주의 의료에 대한 언론보도와 미디어 프레임의 탐색적 연구. <보건사회연구>, 39권 2호, 332-357.
- 최성호 (2020). 코로나 19 유행의 방역. <대한내과학회지>, 95권 3호, 134-140.
- 최재웅 (2012). 뉴미디어를 활용한 재난방송 전달체계 개선 연구. <방송과 미디어>, 17권 3호, 24-39.
- 한규훈 (2011). 여성암 조기검진 촉진 메시지의 설득효과에 미치는 건강신념요인의 영향: 한국 여성과 일본 여성 간의 비교 고찰을 토대로. <한국광고홍보학보>, 13권 2호, 377-413.

하진홍·임혜준 (2020). 대규모 전염병에 관한 뉴스보도 프레임과 공중 위험인식 간의 관계 분석 연구. <언론과학연구>, 20권 1호, 191-229.

지도기계학습을 이용한 트위터 뉴스의 프레임 특성 분석 코로나19 보도 프레임의 자동화 판별 방법을 중심으로

이주연

(중앙대학교 미디어커뮤니케이션학과 석사과정)

이신행

(중앙대학교 미디어커뮤니케이션학과 조교수)

본 연구는 지도기계학습으로 코로나19 관련 언론보도를 건강신념모델에 기초한 심각성, 취약성, 이득, 장애의 프레임으로 분석했다. 특히, 기사의 헤드라인과 리드에서 확인된 각 뉴스 프레임의 언어적 특성으로 언론사의 트위터 게시물에서 드러나는 프레임을 판별하여 소셜미디어가 언론보도에 미치는 영향을 추론했다. 분석대상은 국내 주요 종합일간지인 조선일보, 중앙일보, 경향신문, 한겨레가 코로나19와 관련해 2020년 1월 20일부터 2021년 1월 19일까지 보도한 기사와 트윗이다. 지도기계학습을 이용한 자동화 방식의 프레임 판별을 위해 임의로 추출된 기사 표본 2000건에 대한 모델 정확성을 검증했고 이를 2000건의 트윗 표본에 적용하여 예측 정확성을 평가하고 언어적 특성의 차이를 살폈다. 그 결과, 심각성과 취약성의 지각된 위험 프레임이 이득과 장애의 행동적 평가 프레임에 비해 언론보도에서 부각되고 있어 코로나19에 대한 위험 인식 측면이 감염예방행동의 비용과 편익보다 강조되고 있음을 발견했다. 그러나 심각성과 장애 프레임은 취약성과 이득 프레임에 비해 기계학습모델의 예측 정확성이 저하된 점이 두드러져 프레임의 언어적 특성이 상대적으로 불규칙적이고 다양함을 포착했다. 더 나아가, 트위터에서는 이용자의 참여에 기반한 소셜미디어의 뉴스 확산 원리에 따라 프레임이 보다 유연하고 차별적인 방식으로 구성되어 개인적 차원에서의 감정적인 표현이 프레임에 자주 드러나고 있음을 확인했다. 본 연구는 기계학습을 활용해 대단위의 뉴스기사로부터 프레임을 분석하여 인간 코더의 주관적 해석이 아닌 투명하고 재현 가능한 자동화 방식으로 프레임 언어의 특징을 도출했다는 의미를 가진다.

핵심어 : 자동화 프레임 분석, 지도기계학습, 트위터 뉴스, 코로나19, 건강신념모델