

노동정책연구  
2021. 제21권 제2호 pp.197~229  
한국노동연구원  
<http://doi.org/10.22914/jlp.2021.21.2.007>

연구논문

## 코로나19 확산에 따른 언론의 일자리 주제 분석

양지윤\*  
손미나\*\*  
노승민\*\*\*

본 연구는 코로나19 팬데믹으로 인한 고용 위기에 대응하는 정부정책과 코로나19의 국내 확산 양상을 중심으로 언론 기사를 분석했다. 이를 위해 일자리 주제와 주제별 감정이 어떻게 변화하는지 텍스트 마이닝 기법을 활용했다. 일자리 주제는 ‘맞춤형 지원’, ‘기업투자’, ‘정치’, ‘정부재정지원’, ‘고용동향’, ‘노사관계’, ‘취약계층복지’, ‘세계경제’ 등 8개로 구분했다. 코로나19가 발병한 2019년 1월부터 일자리 기사의 주제를 살펴보면 ‘맞춤형 지원’, ‘기업투자’, ‘고용동향’은 감소하고 ‘정부재정지원’, ‘정치’, ‘세계경제’, ‘취약계층복지’는 증가했다. 감염병 유행 직후 긍정 감정을 나타내는 ‘기업투자’와 ‘맞춤형 지원’이 감소하며 기사 감정이 부정적으로 변화하였으나, 3월부터 ‘정부재정지원’이 늘어나면서 긍정 방향으로 변했다. 코로나19 이후 주제가 정치, 정부재정지원 및 취약계층복지, 고용동향과 세계경제, 기업투자와 노사관계로 순차적으로 급증세를 보였다. 선행연구와 마찬가지로 일자리 기사 주제는 정부정책을 중심으로 설정되며 내외부 요인에 민감하게 반응하였다. 특히, 코로나19 팬데믹 상황에서 주요 일자리 주제가 정부정책과 감염병 확산 양상에 따라 탄력적으로 변화하였음을 보여주었다.

핵심용어 : 일자리, 토픽분석, 감성분석, 코로나19

논문접수일: 2021년 3월 5일, 심사의뢰일: 2021년 3월 24일, 심사완료일: 2021년 5월 18일

\* (제1 저자) 한국고용정보원 연구위원(sun1004@keis.or.kr)

\*\* (공동저자) 한국고용정보원 연구원(sonm35@keis.or.kr)

\*\*\* (교신저자) 한국고용정보원 책임연구원(reaching@keis.or.kr)

## I. 서론

뉴스는 국민의 관심을 반영한다는 측면에서 매우 유용한 정보다. 뉴스는 독자에게 신속하게 전달하는 것이 중요하다. 그러므로 뉴스의 정확성과 신뢰성은 의료나 금융 및 행정정보에 비해 낮지만, 국민여론에 영향을 미칠 뿐만 아니라 정부정책에 반영되므로 중요하다. 본 연구에서는 뉴스 중 일자리 기사에 초점을 맞추었다. 일자리는 노동자의 소득과 직결된 문제일 뿐만 아니라 국가 경제와 밀접한 주제이다. 그러나 뉴스는 비정형 자료이므로 분석이 까다롭다. 일반적인 정형 데이터 방법으로는 분석이 어려워 비정형 데이터 분석 방법 중에서 토픽 분석과 감성 분석을 활용해 일자리 기사의 주요 주제를 파악하고 기사의 감정 상태를 확인했다.

최근 일자리와 관련한 큰 변화는 코로나19의 발생이다. 2020년 전 세계는 코로나19 팬데믹이라는 유례없는 보건 위기를 맞이했다. 2019년 12월 중국에서 처음 보고된 이후에 2020년 1월에 세계로 확산해 1월 30일 세계보건기구가 국제 공중보건 비상사태를 선언한 바 있다(Sohrabi, Alsafi, O'Neill, Khan, Kerwan, Al-Jabir, Iosifidis, and Agha, 2020). 코로나19 팬데믹은 곧바로 일자리 위기 상황을 만들었다. 바이러스 확산 방지를 위해 실시한 봉쇄와 완화 조치(containment and mitigation policy)로 경제활동이 위축되고 불확실성이 증가하면서 경제와 노동시장에 나쁜 영향을 미쳐 일자리 위기가 시작됐다(OECD, 2020).

본 연구는 이러한 고용 위기 상황에서 각종 언론에 게시되는 일자리 기사가 어떤 주제로 구성되어 있고 주제별 감정의 변화는 어떠한지 살펴보고 코로나19 확산에 따른 일자리 관련 주제 변화에 대해 논의하고자 한다. 이를 통해 정부나 학계에서 일자리와 관련된 주요 의제를 파악하고 코로나19 확산에 따라 어떤 주제가 관심을 받는지 파악하고자 한다.

## II. 이론적 배경

### 1. 일자리 기사 텍스트 마이닝 연구 동향

최근 일자리 언론 기사를 텍스트 마이닝 하여 의제를 도출하거나 트렌드를 살피는 연구가 증가하고 있다. 특히 고용정책의 목표 대상인 노인, 장애인, 청년 등을 중심으로 연구가 진행되고 있다. 2020년에는 노인 일자리 언론 기사를 텍스트 마이닝 기법을 활용하여 분석한 여러 연구가 발표되었다. 이소정(2020)은 2004년부터 2019년 사이의 노인 일자리 기사를 토픽 분석하여 ‘일반적 정보 전달’, ‘지자체 사업홍보’, ‘노후생활’, ‘고용효과’, ‘시장연계’라는 다섯 가지 잠재 토픽의 확인을 통해 문재인 정부 출범 이후 노인일자리사업의 고용효과 의제가 증가했음을 보여주었다. 김영아·박윤희·차수민·이혜진(2020)은 2010년부터 약 10년의 노인일자리 기사를 수집하여 토픽 트렌드를 분석하였다. 그 결과 노인 일자리 단어의 검색 횟수는 꾸준히 증가했으며 ‘사회활동지원사업’과 ‘보건복지부’를 자주 출현한 키워드로 파악했다. 그러나, 2016년부터 2020년 사이의 노인 관련 기사에서 ‘노인 일자리’ 키워드는 4% 미만의 트래픽을 차지하며 대중적 관심이 미미하였고 주요 트래픽이 ‘안전지킴이’, ‘공익활동’ 등에 치우치면서 소득증대와 동떨어진 노인 일자리의 인식이 컸다(김양우, 2019). 감성 분석 결과 노인 일자리 기사는 긍정 감정이 우세했지만, 저임금 용어 관련 문장에서는 부정 감정이 압도적이었다. 저임금 관련 부정 문장은 일자리 확대가 저임금 일자리로 이어져 노인 빈곤으로 연결된다는 내용을 포함했다. 자료 수집 기간의 차이는 있지만 이들 연구는 노인 일자리 기사가 정부정책을 중심으로 보도되고 있으며 건강한 노후생활을 영위하도록 일자리로 사회참여를 독려하는 노인복지 목적의 사업에 초점을 맞춘다는 것을 보여준다. 다만, 최근 들어 노인일자리사업의 고용효과에 관한 관심이 증가하여 사회참여보다 소득증대와 일자리 질적 향상에 목표를 둔 사업이 주목받고 있다.

장애인 고용과 관련한 고용동향을 분석한 연구도 있었다. 2018년 9월부터 12

월까지 언론 기사를 토픽 분석한 결과 ‘편의점 직업교육’, ‘장애인 고용 우수기업’, ‘장애인 표준 사업장’, ‘진로교육’, ‘일자리 확대’ 등 정부와 공공기관에서 운영하는 사업이나 프로그램, 사업 주체와 관련된 7가지 잠재 주제가 나타났다(전혜영, 2019). 감성 단어 연결을 분석한 결과 장애인 일자리를 만들기 위한 업무 체결과 일자리 창출은 긍정적이었지만, 장애인 의무고용에 따른 부담금 납부에는 부정적인 감정을 드러냈다. 다른 연구는 2003년부터 2018년까지 장애인 일자리 기사를 수집하여 노동의제를 분석했는데 언론이 정부정책을 최우선으로 하여 의제를 설정하며 장애인 일자리의 양적 증대에 관한 더 높은 관심을 보인다고 주장했다(정지원 · 이재민 · 최소연, 2018).

고용정책의 폭넓은 의제를 확인한 연구는 2010년부터 2018년까지의 고용 관련 언론 기사를 수집하여 토픽 모델링과 네트워크 분석을 시행한 것이다(김경동 · 이시영 · 고길곤, 2020). 수집된 언론 기사 텍스트에서 ‘대내환경’, ‘고용정책’, ‘청년고용’, ‘대외환경’, ‘고용갈등’, ‘고용동향’ 등 6개 의제가 도출되었으며 시간의 흐름과 공론장(예: 중앙지, 지방지)에 따라 의제의 중요도와 프레임이 달라졌다.

국내 일자리 언론 기사를 이용해 텍스트 마이닝 기법을 적용한 연구 문헌의 시사점은 다음과 같다. 일자리 기사의 주요 의제는 정부 일자리 정책과 관련이 깊으며 기사 수, 토픽 비중, 검색어 트래픽 등의 추이 변화가 정부정책 기조나 사회 관심의 변화와 맞닿아 있다. 또한, 언론 기사의 긍·부정의 감정보다 요소가 평균 추이보다는 특정 주제나 키워드에서 더 명료하게 해석되는 특징을 보인다.

## 2. 코로나19와 일자리 의제

세계 여러 국가는 코로나19로 인한 일자리 위기를 해결하기 위해 업무처리 방식 전환, 기업 지원, 고용유지지원 등의 정책을 폈다(OECD, 2020). 텔레워킹 등 비대면 업무처리 방식을 확산하기 위한 재정 지원을 확대했고 사업장 폐쇄 같은 봉쇄 조치로 생산활동이 중단되어 유동성 제약을 겪게 된 기업에 유동성을 지원했다. 또한, 근로시간 단축을 장려하고 지원금 소득대체율을 인상하는

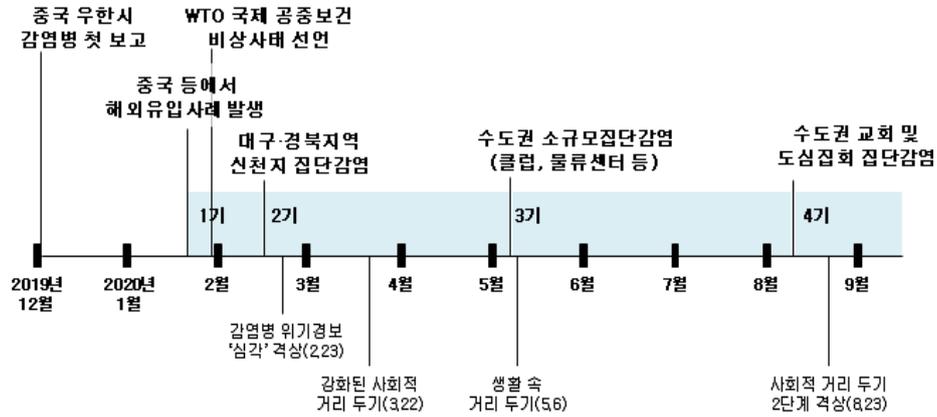
등 고용유지 지원책을 시행했으며 코로나19의 영향을 받는 노동자 중에서 고용유지제도 등의 적용이 어려운 자영업자, 비전형근로자 등에 대한 소득 보조와 실업자의 구직급여 기간 연장 등을 조치했다.

코로나19로 인한 고용 충격에 맞서 대한민국 정부가 시행한 고용안정 특별대책의 방향은 크게 ① 재직 중인 근로자의 고용을 유지하기 위해 개인과 기업을 대상으로 한 정책, ② 새 일자리 창출, ③ 실업자와 취업지원이다.<sup>1)</sup> 기업 대상으로는 기존의 고용유지지원금 대상을 확대했고 코로나19로 직접 피해가 심각한 여행·관광·항공·교통 관련 업종을 특별고용지원 업종으로 신규 지정해 고용유지 지원을 받을 수 있도록 했다. 개인 대상으로는 이전까지 고용보험을 적용받지 못하던 특수고용·프리랜서·영세 자영업자에게 긴급고용안정지원금을 지급했다. 일자리 창출을 위해 비대면·디지털 중심으로 공공일 자리를 확대하기도 했다. 일자리 기사의 의제 변화는 앞선 텍스트 마이닝 연구 결과처럼 정부정책 발표나 운영 절차와도 관련이 있을 것이며 감염병 양상과도 밀접한 관계를 지닐 가능성이 크다. 국내 감염병 확산 방지의 기본 원칙인 사회적 거리두기가 감염병 유행에 따라 곧바로 단계가 조정되므로 비대면 업무 전환이 어려운 업종의 생산활동에 직접 영향을 주기 때문이다. OECD는 감염병 유행에 따라 우리나라의 고용상황을 전망했다.<sup>2)</sup> 국내 감염병 양상은 [그림 1]처럼 2020년 1월 국내 첫 확진자를 기점으로 9월까지 4단계를 거쳤다.<sup>3)</sup> 1기는 소규모 해외유입과 개별 감염이 발생한 2020년 1월 20일부터 2월 17일까지이다. 2기는 종교시설에서 대규모 집단 감염이 발생한 2월 18일부터 5월 5일까지이다. 3기는 소규모 집단감염이 발생한 5월 6일부터 8월 10일까지이다. 4기는 도심 집회와 관련된 집단감염이 발생한 8월 이후이다.

본 연구는 팬데믹으로 야기된 고용 위기에 관한 정부 대책과 코로나19의 확

1) 대한민국 정책브리핑(2020. 6. 1), 「코로나19 경제 대책」. <https://www.korea.kr/special/policyCurationView.do?newsId=148872965#L2-5> (검색일: 2021년 2월 25일)  
 2) OECD는 2020년 고용전망보고서에서 1차 유행으로 한국 내 고용률이 0.6% 하락하고 실업률은 2020년 하반기 4.8% 증가할 것으로 예측했다.  
 3) 중앙방역대책본부에서 2020년 12월 31일 발표한 「방역당국이 분석한 2020년 코로나19 현황과 새해 전망」 보도자료에서는 국내 감염병 발병 양상을 5개 시기로 구분하고 있으며 이 중 2020년 9월 이전의 네 시기를 소개함. 자세한 내용은 홈페이지 참조 <https://www.korea.kr/special/policyFocusView.do?newsId=148882004&pkgId=49500742> (검색일: 2021년 3월 2일)

(그림 1) 코로나19 유행 양상



자료: 중앙방역대책본부(2020.12.31.)의 자료 재구성.

산 양상을 중심으로 언론의 일자리 의제와 의제별 감정이 어떻게 변화하는지를 살피고자 한다.

### Ⅲ. 연구 방법

#### 1. 분석 대상

본 연구는 2019년 1월 1일부터 9월 17일까지의 뉴스를 한국언론재단의 빅카인즈 통합 데이터베이스에서 수집하였다. 빅카인즈(BIGKinds)는 한국언론진흥재단에서 제공하고 있는 뉴스 분석 서비스로 종합일간지와 지역일간지 및 경제지를 포함하여 54개 언론사의 뉴스 데이터베이스를 구축하고 이를 빅데이터 기술로 분석하여 결과를 제공한다. 연구진은 정치, 경제, 사회 분류에서 제목이나 내용에 ‘일자리’ 키워드를 포함하는 기사를 수집하였다. 수집한 기사 중에서 인사 동정 기사 3,061건을 제외하고 51개 언론사의 기사 13만 399건을 최종 분석 대상으로 설정하였다. 데이터의 수집, 가공, 분석은 모두 Python 3.7을 활용하여 비정형 데이터 분석을 진행했다.

## 2. 분석 방법

### 가. 자료 가공 및 정제

비정형 분석의 첫 단계는 데이터 정제이다. 일자리 관련 기사는 보도자료를 기반으로 작성하므로 내용이 유사한 중복 기사가 많은 편이나 기사 재생산 수준도 국민 관심을 반영하는 중요한 지표가 될 수 있으므로 중복 기사는 정제하지 않았다. 먼저, 코로나바이러스감염증-19의 다양한 동의어를 ‘코로나’로 치환하는 작업을 진행하였다. 치환된 단어는 ‘코로나 바이러스’, ‘신종 코로나바이러스’, ‘covid-19’, ‘covid19’, ‘코비드’였다. 이후 특수문자, 숫자, 영어, 한자 등 한글이 아닌 문자를 먼저 제거하기 위해 `soynlp(0.0.492)` 라이브러리의 `only_hangle` 함수를 로드하였다. 토큰화에는 은전한닢 프로젝트<sup>4)</sup>의 `Mecab-ko` 형태소 분석기를 사용하였다. ‘비대면’, ‘코로나’, ‘언택트’와 같은 신조어를 사전에 추가하였고 정보로서 의미 없는 단어와 한 글자 단어(예: 것, 일, 수) 및 자료 수집 키워드였던 ‘일자리’는 불용어로 처리하였다. 토픽 분석은 단일 토큰으로 구성된 유니그램(`uni-gram`)을 활용하였고 감성 분석은 범용 오픈소스인 군산대학교 감성사전을 활용하였다.<sup>5)</sup> 이 사전은 표준국어대사전을 구성하는 형용사, 부사, 동사, 명사의 뜻풀이와 그 외 여러 출처에서 긍·부정어를 추출하여 다양한 어구와 문형, 축약어를 포함하며 음식이나 여행지 및 영화 후기 분석 등 여러 영역에서 활용되었다. 사전은 총 1만 4,843개의 단어로 구성되어 있고 각 단어의 극성(긍정, 부정, 중립)과 강도에 따라 -2점에서 2점까지 1점 간격의 점수가 매겨져 있다. 또한, 최대 3개의 토큰을 결합한 트라이그램(`tri-gram`)을 226개나 포함하고 있어 추출 범위를 트라이그램까지 확대하였다. 감성 단어를 추출하기 위해 토큰화 함수를 별도로 구성하였는데 뜻풀이에 활용된 품사를 참고하여 형태소 분석기에서 판별 가능한 일반명사, 동사, 형용사, 보조용언, 일반부사, 접속 부사, 명사 파생 접미사, 동사 파생 접미사, 형용사 파생 접미사를 대상으로 단어를 추출하였으며, 복합어는 분리하여 어근만 추출하는 기능을 구현하였다. 마지막으로 앞서 준비된 전처리 함수, 토큰화 함수, 감성사전을 `gensim`

4) <http://eunjeon.blogspot.com>

5) <http://dilab.kunsan.ac.kr/knusl.html>

(3.8.1) 라이브러리의 CountVectorizer 파라미터인 preprocessor, tokenizer, vocabulary에 대응시키고 ngram\_range=(1, 3)으로 설정하여 fit\_transform을 수행하면 감성사전에 정의된 단어에 한하여 기사별 토큰화가 수행된다. 이후 토큰별 극성을 확인하여 기사별 감정점수를 계산하였다.

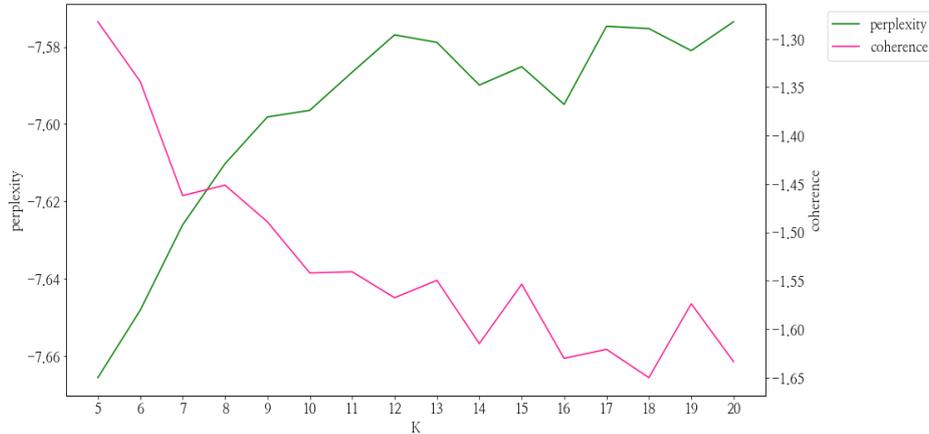
#### 나. 토픽 분석

언론 기사에서 주요 정책 이슈를 탐색하기 위하여 토픽(주제) 분석 방법을 활용했다. 토픽 모형(Topic modeling)은 문서에 관한 사전 정보 없이 주제를 찾아내므로 본 연구의 분석 대상인 대량의 텍스트 데이터 분석에 적합하다. 토픽 분석은 토픽 모형 알고리즘을 활용하여 주제를 찾아내는 통계 방법이다. 본 연구에서는 많은 양의 신문기사를 분석할 때 유용한 것으로 알려진 잠재적 디리클레 할당(Latent Dirichlet Allocation : LDA)을 활용하였다(Jacobi, Van Atteveldt, and Welbers, 2016). LDA는 문서 집합에 존재하는 특정한 단어 집합이 발생할 확률을 기반으로 잠재 토픽을 생성하는 생성적 확률모델(generative probabilistic model)로 각 문서는 디리클레 분포를 따르는 여러 주제의 혼합체(mixture)로 표현된다(Blei, Ng and Jordan, 2003). 또한, 단순성과 데이터 차원 축소의 유용성 및 해석이 가능하며 의미적으로 일관된 주제를 만들어내는 능력 때문에 널리 알려진 모형이다(Mimno, D. and McCallum, A., 2008).

코로나 발병 시점을 전후로 데이터를 분리하지 않고 전체 데이터를 통합하여 토픽 분석을 수행하였다. 그 이유는 기간을 분리하여 토픽을 분석하면 이전의 토픽과 이후 토픽의 연결성 및 트렌드 변화를 살필 수 없고, 코로나19 전의 주요 토픽과 코로나19 후의 주요 토픽을 각각 서술하는 것만 가능하기 때문이다. LDA는 자료의 시점을 고려하지 않으므로 전 기간을 합쳐 진행하더라도 코로나19 이후 새롭게 등장하는 토픽이 하나의 토픽을 구성할 수 있다.

LDA를 활용하여 토픽 모형을 생성할 때 토픽의 수(k)와 외생 모수인  $\alpha$ 와  $\beta$  설정이 필요하다. 토픽의 수를 설정할 때는 모델적합도에 해당하는 혼란도(perplexity)와 토픽별 상위 단어 간의 유사성을 의미하는 토픽 일관성(topic coherence)을 모두 고려하여야 한다(Blei, Ng and Jordan, 2003; Newman, Lau, Grieser, and Baldwin, 2010). 즉, 모델적합도(혼란도)와 함께 사람이 해석 가능한

(그림 2) 혼란도와 토픽 일관성(K=5~20)



토픽 수의 설정이 필요하므로 토픽 일관성 지표도 함께 고려해야 한다(Chang, Gerrish, Wang, Boyd-Graber, and Blei, 2009). 이와 함께 LDAvis로 토픽 분석 결과를 시각화했을 때, 토픽 간 중첩이 최대한 적게 적절히 범주화되는지를 토픽 수 결정에 활용한다(김경동·이시영·고길근, 2020).

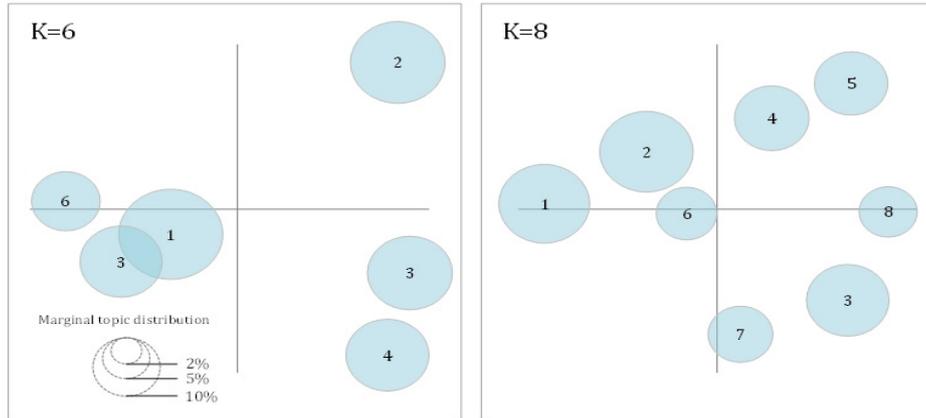
본 연구는 토픽 수 결정을 위해 K값을 5에서 20까지 하나씩 늘려가면서 1차로 혼란도와 토픽 일관성 지표를 확인했고, 2차로 시각화를 통한 범주화 정도를 파악하였다. 토픽 수 증가에 따른 혼란도와 토픽 일관성의 변화는 [그림 2]와 같다. 이를 평가하는 기준은 혼란도는 낮을수록 토픽 일관성은 높을수록 좋은 모델이라고 할 수 있다.

두 지표를 고려하여 K값을 5~8개로 좁히고, [그림 3]처럼 결과를 시각화하여 토픽 간 중첩 없이 분리되는지를 확인하였다. 본 연구에서는 시각화 데이터 생성을 위해 LDAvis의 Python wrapper인 pyLDAvis (2.1.2) 라이브러리 prepare 함수를 이용하였다. 토픽의 좌표는 prepare 함수 내부에서 주성분 분석(Principal Component Analysis)을 이용하여 2차원으로 압축되어 표현된다.

K값이 5일 때와 8일 때 범주화가 잘 되었으나, 기사의 다양한 주제가 토픽 수 5보다 8에서 잘 드러난다고 판단하여 8개를 최종 선정하였다.

이때, 외생변수는 Gensim의 LdaMulticore에서의 symmetric 옵션(1.0/토픽 수)을 적용해  $\alpha=0.125$ 와  $\beta=0.125$ 로 지정하였다. 토픽별 키워드와 해당 토픽

(그림 3) 토픽 간 거리 시각화



비중이 높은 대표기사의 내용을 참조하여 연구자가 토픽 이름을 부여하였다.

#### 다. 토픽 트렌드 분석

토픽 트렌드 분석은 토픽 비중이 시간에 따라 어떻게 변화하는지를 보여주는 방법이다(김영아 · 박윤희 · 차수민 · 이혜진, 2020; 방한솔 · 문호석, 2019; 이소정, 2020). 본 연구에서는 <표 1>처럼 토픽 분석의 세 가지 방법을 소개하고 각각의 장단점을 서로 보완하고자 하였다.

먼저, 가장 많이 활용하는 토픽 트렌드 분석은 각 문서의 토픽별 비중을 시점 별로 모두 더하는 비중 합계 방식이다(이소정, 2020). 이 방식은 문서 내에서 토픽 비중이 높아지는 것과 특정 토픽을 다룬 문서 수가 증가하는 것을 반영할 수 있다. 그러나 토픽이라는 추상 개념을 수치화한 것이므로 직관적이지 못하다. 또한, 짧은 기사의 비중과 긴 기사의 비중을 같은 값으로 처리하므로 기사 분량을 반영하지 못한다는 단점이 있다.

두 번째 방법은 토픽지수를 활용하는 것이다. 토픽지수는 토픽별 키워드 빈도의 합을 기사 수로 나눈 값이다. 고희이 · 오승원 · 백장선(2020)은 수집한 시계열 뉴스 자료에서 특정 시점을 기준으로 언론사가 추가되어 기사의 수가 증가한 것을 시계열 분석에 고려하려고 토픽지수를 계산하여 KOSPI 200 지수를 예측한 바 있다. 이 방법을 변형하여 본 연구에서는 토픽별 상위 20개 키워드의

월별 문서 집합 내 등장 횟수를 월 전체 기사 수로 나누어 토픽지수를 계산하였다. 즉, 토픽지수는 해당 월에 토픽별 주요 단어의 기사당 평균 출현 횟수를 의미한다. 기사 수의 효과를 배제하면서 키워드별 빈도수의 증감을 관찰할 수 있다는 장점이 있다. 그러나 원래 출현 빈도가 높은 단어가 토픽지수에 크게 영향을 준다. 토픽별 핵심 단어가 중복을 허용하므로 같은 단어를 공유하는 토픽끼리 비슷한 추세를 보일 수 있다.

마지막 방법은 토픽별 기사 수(문서 빈도)이다. 토픽별로 비중 0.5 이상을 차지하는 기사 수를 월별 추이로 확인할 수 있다. 비중 0.5 이상을 차지하는 토픽 없이 여러 토픽의 작은 비중으로 구성된 기사는 다중토픽 기사로 분류하였다. 이렇게 비중 조합에서 값이 큰 하나를 대표로 삼는 코딩 과정을 거치면 자료를 직관적으로 이해할 수 있게 된다. 그러나 토픽 모형이 갖는 강점인 모호함(fuzziness) 허용을 이점으로 살리지 못한다. 작은 비중을 차지하는 토픽이나 여러 토픽이 균등한 비중을 차지하는 문서는 해석에서 제외되기 때문이다. 특정

〈표 1〉 토픽 분석 방법

종류	정의	특징
토픽비중 (합계)	모든 문서의 토픽 비중을 토픽별로 합한 값	(+) 문서 내 비중과 기사 수의 증감을 모두 고려할 수 있어, 토픽의 실질 증감을 가장 잘 나타냄 (-) 단문 기사와 장문 기사 비중을 동일하게 계산하므로 기사 분량 효과를 반영하지 못함. 비중 계산 값이므로 직관적이지 않음.
토픽지수	토픽별 상위 키워드 20개의 문서 집합 내 개수 합을 월별 총 기사 수로 나눈 값	(+) 기사 수 증감에 따른 효과를 배제하면서 기사 분량 효과는 반영할 수 있음. (-) 주제 간 단어 중복을 허용하므로 특정 주제에 독점적인 정보라고 볼 수 없음. 빈도가 높은 단어에 예민함.
토픽별 기사 수	토픽별 비중이 0.5 이상인 기사 수 * 비중 0.5 이상을 차지하는 토픽이 없는 경우 다중 토픽 기사로 분류	(+) 주요 주제별 기사 수를 파악할 수 있어 직관적임. (-) 여러 토픽으로 구성된 기사의 추이를 해석할 수 없음. 기사 분량 효과를 반영하지 못함.

문서의 토픽 비중이 소수의 토픽에 집중되는 정도는 연구자가 설정하는  $\alpha$ 값에 따라 달라지므로(백영민, 2020) 연구자의 추정에 따라 토픽별 문서 수가 달라지는 것도 단점이 된다.

본 연구는 토픽 비중 합계를 우선하여 분석 결과를 해석하되 필요한 경우에 토픽별 기사 수와 토픽지수를 활용하여 해석을 보완하였다.

#### 라. 감성 분석

뉴스 기사에서 나타난 정서를 분석하기 위하여 감성 분석(sentiment analysis)을 활용하고자 한다. 감성 분석은 텍스트에 나타나는 태도, 의견, 성향 같은 주관적 요소나 정서적 특징을 정량화하여 분석하는 방법이다(Pang and Lee, 2008). 주로 소셜 미디어나 상품 후기, 영화평, 뉴스 등의 텍스트 데이터를 분석하여 장소, 제품, 브랜드, 인물, 정책 등에 대한 평판과 여론을 파악하는 데 쓰인다(이정훈, 2018).

감성 분석은 어휘 기반 접근 방법과 지도 기계학습 방법으로 나뉜다(김영민·정석재·이석준, 2014). 어휘 기반 접근은 단어 사전을 활용하여 문서의 주관적 특성을 파악하는 방법이다. 단어 목록에 특정 감정(예: 분노, 슬픔 등)이나 정서의 극성(긍정-부정)을 값으로 매겨둔 단어 사전을 이용하여 문서 내에 사용된 단어의 감정점수를 매겨 문서의 정서를 파악한다. 단어 사전은 직접 구축하거나 기존 사전을 활용할 수 있으며 정교하고 방대한 사전을 활용할수록 좋은 결과를 얻을 수 있다. 미리 정의된 단어 사전을 활용하면 비교적 쉽고 빠르게 감정을 파악할 수 있다. 사전을 이용하는 방법은 범용 어휘를 기반으로 다양한 주제를 다루는 언론 보도 분석에 적절하며 사전의 어휘 확장이 쉬워 높은 유연성을 갖는다(최창식·임영호, 2021). 반면, 특정 어휘나 표현이 자주 사용되는 분야에서는 판별력이 떨어질 수 있다는 우려도 있다(Zhang and Lju, 2017).

김정화·윤서연(2018)은 뉴스와 신문 기사에서 도로교통서비스 관련 정책 이슈 및 여론의 변화를 파악함에 있어서 이용자 만족도 조사를 보완하는 수단으로 감성 분석을 수행했다. 이때 군산대 감성사전을 활용하였으며 단어의 극성을 파악하여 문장 수준에서 긍정은 1, 부정은 -1로 분류하고 전체 문장의 분류값을 합산하여 기사의 긍·부정을 분류하였다. 최창식·임영호(2021)는 전국

지와 지역지 정치 기사에서 표출되는 감성을 살펴보기 위해 군산대 한국어 감성사전을 이용하였는데 기사 속 단어의 감정점수를 합산하고 기사의 총 단어 수로 나누어 기사별 감성지수를 산출하여 신문 간 차이를 비교하였다.

기계학습 방법은 문서나 문장에 대한 감정 평가 자료를 수집하여 기계학습 알고리즘을 적용하여 분류모델을 구축하는 것이다. 이 경우 학습 데이터를 사전에 준비해야 하며 학습되지 않은 분야의 자료 분석에는 정확도가 떨어질 수 있다(Zhang and Lju, 2017). 분석 대상 텍스트에 사전 평가 자료가 없다면 연구자가 직접 자료의 감성을 분류해야 하는데 데이터 규모가 클수록 비용이 많이 들며 잘못된 분류로 정확도가 낮아질 수 있다. 물론 표본추출을 통한 소규모 데이터를 인간(코더)이 긍·부정으로 분류한 후에 이를 학습시킨다고 하더라도 뉴스 기사는 다양한 표현과 주제를 포괄하므로 코더의 역량이 결과에 영향을 미칠 수 있다는 문제가 있다. 그래서 본 연구에서는 기계학습 방법이 적절하지 않다고 판단하였다. 또한, 감정 트렌드 분석은 특정 시점의 점수가 중요한 것이 아니라 전월(년) 대비 증가나 감소의 추세가 중요하다. 감정점수가 전월 대비 감소하였다는 의미는 뉴스 기사에 대한 부정 기류가 나타난 것이고 증가하였다는 의미는 긍정 기류를 보이는 것이다.

토픽별 뉴스 기사의 감성을 분석하고 앞선 토픽 분석 결과와 감정 분석 결과를 통합하여 토픽별 감정의 시계열 추이를 확인하였다. 독자는 신문을 기사 단위로 읽게 되므로 기사를 최소 분석 단위로 설정하여 토픽별 감정 변화를 분석하였다. 한 기사에서 비중이 0.5 이상인 토픽이 있는 문서만을 추려낸 후에 토픽과 시점별 감정점수의 합계를 계산하였다. 그리고 토픽별 비중 합계와 감정점수 합계를 시계열로 표현하였다.

## IV. 분석 결과

### 1. 토픽

토픽을 해석하기 위해서는 토픽을 구성하는 상위 키워드와 토픽을 대표하는

기사들을 확인하면 된다. 토픽별 상위 20개 키워드는 <표 2>와 같이 나타났다. 토픽 제목은 키워드와 토픽별 대표 기사를 참고하여 연구자가 붙인 것이다.

<표 2> 토픽별 키워드

(단위: 건, %, 점)

토픽	상위 20개 키워드	문서 빈도	문서 비율	감정 점수합
토픽 1 (맞춤형 지원)	청년, 지원, 사업, 지역, 취업, 교육, 창업, 센터, 기업, 참여, 대학, 운영, 제공, 창출, 여성, 도시, 프로그램, 대상, 진행, 채용	33,383	25.6	230,519
토픽 2 (기업투자)	기업, 산업, 투자, 기술, 사업, 혁신, 경제, 지원, 지역, 창출, 개발, 성장, 생산, 에너지, 시장, 벤처, 추진, 분야, 계획, 구축	24,202	18.6	207,880
토픽 3 (정치)	대통령, 경제, 국민, 의원, 정부, 정책, 민주당, 국회, 후보, 정치, 대표, 청와대, 위원장, 미래, 총선, 국가, 문재인, 선거, 통합, 강조	16,106	12.4	13,542
토픽 4 (정부재정지원)	지원, 코로나, 정부, 경제, 고용, 예산, 추경, 대책, 지급, 재정, 사업, 확대, 위기, 규모, 지원금, 자금, 안정, 회의, 대상, 보험	14,531	11.1	45,065
토픽 5 (고용동향)	고용, 경제, 증가, 감소, 소득, 정부, 정책, 임금, 지난해, 취업자, 인구, 기업, 실업, 성장, 올해, 조사, 이후, 대비, 노동, 분기	14,029	10.8	3,345
토픽 6 (노사관계)	채용, 광주, 고용, 노동, 기관, 기업, 노조, 공공, 지역, 노사, 경영, 임금, 정규직, 노동자, 직원, 회장, 위원회, 한국, 비정규직, 전환	10,883	8.3	3,373
토픽 7 (취약계층복지)	장애, 노인, 서비스, 코로나, 여성, 복지, 건강, 활동, 서울, 가족, 의료, 대표, 병원, 시작, 운영, 필요, 시설, 직원, 학교, 어르신	10,055	7.7	-44,085
토픽 8 (세계경제)	미국, 중국, 코로나, 세계, 경제, 한국, 트럼프, 달러, 일본, 시장, 무역, 대통령, 기업, 글로벌, 국가, 해외, 위기, 국제, 현지, 투자	7,210	5.5	-16,695
전 체		130,399	100.0	

여덟 개 토픽 중 맞춤형 지원 기사가 전체의 25% 정도로 가장 많았다. 다음은 기업투자, 정치, 정부재정지원, 고용동향, 노사관계, 취약계층복지, 세계경제 순으로 나타났다. 문서 비율이 10%에 달하지 못한 토픽은 노사관계, 취약계층복지, 세계경제였다.

첫째 토픽은 맞춤형 지원이었다. ‘청년’, ‘지원’, ‘취업’, ‘교육’, ‘창업’, ‘대학’ 등이 주요 단어로 나타났다. 가장 중요도가 높은 단어는 ‘청년’이었지만, ‘대학’, ‘여성’, ‘도시’, ‘창업’ 등 다양한 취업 지원 대상과 취업 지원 형태를 포함한다. 관련 기사는 청년 일자리뿐만 아니라 여성, 노인, 장애인 일자리 특화, 대학 일자리센터 프로그램, 직업교육, 취업 및 창업 지원 등을 다룬다. 문서 비율이 25.6%로 토픽 중 가장 높은 비중을 차지한다.

둘째 토픽은 기업투자였다. ‘기술’, ‘혁신’, ‘창출’, ‘개발’, ‘성장’, ‘에너지’, ‘추진’ 등 기업과 산업의 투자 및 성장과 관련된 단어로 구성되었다. 이 토픽을 주로 다룬 문서는 전체 문서의 18.6%였다. 기업투자로 분류된 기사는 규제 완화, 투자 확대, 기업 전략 변화, 지자체의 산업육성, 정부-기업 간 협약 등 일자리 창출에 수반되는 정부 및 지자체의 정책과 기업 전략을 주제로 삼고 있다.

셋째 토픽은 정치였다. ‘대통령’, ‘의원’, ‘민주당’, ‘국회’, ‘후보’, ‘정치’, ‘총선’ 등 정치 관련 키워드가 등장했고, 관련 기사에서 다양한 정치 이슈들을 다루고 있었다. 주요 키워드는 ‘대통령’, ‘의원’, ‘후보’, ‘대표’, ‘위원장’, ‘문재인’과 같이 인물을 지칭하는 단어가 많았다. 분석 대상 키워드인 일자리는 행정부의 정책 수립이나 예산편성, 정치권에서의 현안이나 2020년 4월에 치러진 21대 총선 공약으로 언급되었다.

넷째 토픽은 정부재정지원이었다. 키워드에는 ‘지원’, ‘정부’, ‘예산’, ‘추경’, ‘지급’, ‘지원금’ 등이 나타났고 대표기사에는 고용과 관련한 정부의 추경 편성이나 지원금 관련 내용이 실렸다. 토픽 비중이 높은 대표기사는 대부분 국내 신종 코로나바이러스 감염증 발생 이후 게시된 것으로 기업과 소상공인 금융지원, 사회보험료 완화, 긴급재난지원금과 같은 중앙정부 추진 사업과 지방자치단체에서 마련한 지원책을 포함한다. 코로나19 발생 이전 기사도 일자리안정자금, 고용장려금 등 재정지원 일자리 사업에 관한 것이었다.

다섯째 토픽은 고용동향이었다. ‘증가’, ‘감소’, ‘지난해’, ‘취업자’, ‘인구’, ‘조사’, ‘대비’, ‘분기’ 등 고용 통계를 설명하는 키워드를 포함했다. 대표기사는 취업자 수나 소득 지표를 설명하는 정보 전달성 기사였다.

여섯째 토픽은 노사관계였다. 키워드에는 ‘채용’, ‘광주’, ‘고용’, ‘노동’, ‘기업’, ‘노조’, ‘노사’ 등이 나타났다. ‘채용’은 기업의 채용 소식 기사와 비정규직의 정규직 전환 관련 기사에서 언급되었다. ‘광주’는 광주형 일자리 기사에서 나타났다.<sup>6)</sup> 또한, 해당 토픽으로 분류된 기사에는 인천국제공항공사, 수원시 등 ‘비정규직’의 ‘정규직’ ‘전환’ 내용이 포함되었다.

일곱째 토픽은 취약계층복지였다. 키워드는 ‘장애’, ‘노인’, ‘서비스’, ‘코로나’, ‘여성’, ‘복지’ 등으로 고용 취약계층에 대한 ‘복지’와 ‘서비스’로 요약될 수 있다. ‘의료’, ‘병원’ 등 의료서비스와 관련된 단어도 함께 나타났다. 주요 기사를 살펴보면 무인 주문시스템 도입으로 인한 취약계층 일자리 위협, 조선산업의 위축으로 인한 거제도 산업 쇠퇴, 청계천 제조업 거리 재개발, 노인복지와 일자리, 성평등과 여성 일자리 등 다양한 이슈를 다루고 있으며, 대체로 고용 위기에 취약한 계층, 산업, 지역을 포괄했다.

여덟째 토픽은 세계경제였다. ‘미국’, ‘중국’, ‘일본’ 등 국가명과 ‘세계’, ‘글로벌’, ‘해외’, ‘국제’ 등 세계를 지칭하는 단어가 키워드로 나타났다. 세계경제 기사는 미-중 무역전쟁, 전염병 팬데믹, 세계경제 등 초국가적 이슈가 글로벌 일자리에 미치는 영향에 관한 것이다. 반면, 세계 이슈와 국내 일자리 문제를 직접 연관시킨 기사는 적었다. 글로벌 이슈가 국내 산업 및 기업 활동에 영향을 미치고 그에 따라 국내 일자리 문제에 영향을 줄 수 있지만, 토픽 분석 결과로 그 관계를 확인하기는 어려웠다. 여덟째 토픽은 가장 작은 비중을 차지하며 주제 측면에서 국내 일자리와는 다소 독립적인 영역이라고 볼 수 있다.

토픽에서 드러나는 감정을 보면 <표 2>에서 알 수 있듯이 맞춤형 지원과 기업투자 기사의 감정이 가장 긍정적이었지만, 취약계층복지와 세계경제는 부정

6) 대한민국 정책브리핑(2019. 1. 31), 「광주형 일자리」. 광주형 일자리란 지역 일자리 창출을 위해 광주광역시와 제안한 것으로 기업은 상대적으로 낮은 임금으로 근로자를 고용하여 일자리를 늘리고, 근로자의 낮은 임금을 정부와 지자체가 주거·문화·복지·보육시설 등 후생 복지 비용으로 지원하는 상생형 일자리 모델이다. <https://www.korea.kr/special/policy/CurationView.do?newsId=148858104> (검색일: 2021년 3월 4일)

적이었다. 이는 일자리 사업 자체의 특성과 밀접한 관계가 있다. 즉, 취약계층 복지 주제의 기사는 취약계층을 대상으로 하므로 부정 단어가 연관어로 자주 사용되기 때문이다.

토픽별 대표기사는 토픽 비중이 높은 기사를 말하는 것으로 <표 3>에서 기사 제목과 비중을 보여준다.

<표 3> 토픽별 대표기사

번호	일자	기사 헤드라인	비중
1	2019-12-10	마포구고용복지지원센터, 경단여성 재취업 위한 ‘어린이집 보육교사 양성사업’ 성황리 마무리	0.9971
	2019-01-15	광주 북구, 소통·공감 ‘청년 친화 도시’ 조성	0.9966
	2019-05-28	SBA, ‘모든 기업은 학교다’ 캠페인 가동	0.9966
	2019-08-18	서울시, 9월 말까지 총 21회 ‘여성일자리 박람회’... 530여 개 일자리 창출	0.9966
	2020-05-22	동작구, 팻시터 등 맞춤형 일자리 전문가 양성	0.9965
2	2020-05-28	탈원전 배후도시 새만금, 재생에너지·수소·전기차에역량 집중	0.9988
	2019-08-09	‘기회의 땅’ 송도... 바이오·헬스산업 메카 도약	0.9986
	2019-07-10	세종 자율주행 특화도시... 차근차근 시동건다	0.9986
	2019-09-20	신안 해상풍력발전단지 조성... ‘블루에너지’ 핵심 전략사업으로	0.9985
	2019-06-11	인천 송도와 남동산단 묶는 ‘바이오헬스밸리’ 속도 낸다	0.9985
3	2019-04-24	패스트트랙 정국... 국회는 ‘전쟁터’	0.9995
	2019-09-16	‘포스트 추석’ 정국 격돌... ‘조국 대전’	0.9995
	2019-06-19	인사청문회 계기 국회 정상화 되나	0.9995
	2019-05-13	나경원 ‘달창’ 발언 사과... 진짜 몰랐나?	0.9994
	2020-03-24	의원님들이 ‘당현·당규’만 지켜도 ‘타락한 진영의식’ 사라진다	0.9993
4	2020-04-22	홍남기 “3차 추경 불가피... 대부분 적자국채 발행해 충당”	0.9982
	2020-03-19	문대통령 “서민경제 위한 50조원 특단 금융조치”	0.9980
	2020-02-28	정부, 코로나19 지원에 ‘20조원+α’ 투입... 상반기 신용카드 소득공제 2배로 확대	0.9978
	2020-09-10	4차 추경 포함 12.4조 푼다... 소상공인 200만 원 지원 등에 투입	0.9976
	2020-05-05	서울시, 50인 미만 사업체에도 ‘고용유지지원금’ 50만 원 지원	0.9956

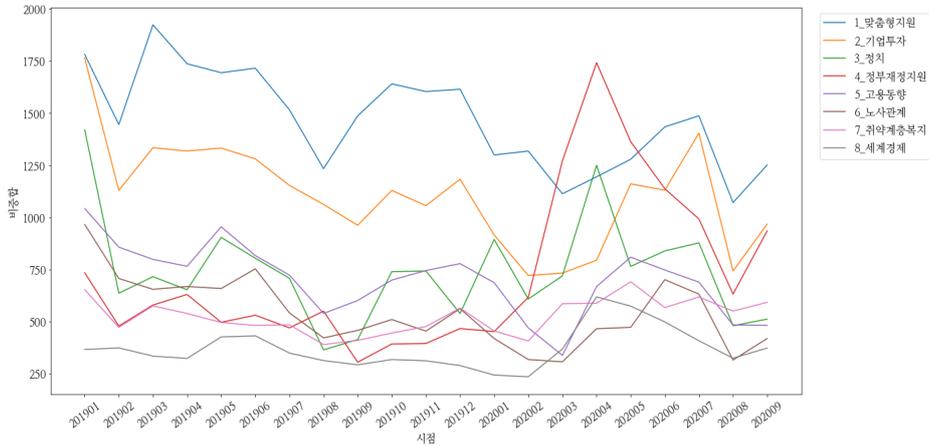
〈표 3〉의 계속

번호	일자	기사 헤드라인	비중
5	2020-02-21	송강호보다 807만 원 더 버는 이선균 ... '기생충'은 현실이다	0.9981
	2019-08-28	심화된 소득 양극화의 근본 문제는 급속한 고령화	0.9981
	2019-08-22	상·하위 20% 격차, 분기기준 최대 역대 ... "최하위층에 자영업자 증가 영향"	0.9976
	2019-12-11	11월 취업자 33만 1,000명 증가 ... 정부 '올해 20만명 증가' 목표 '청신호'	0.9976
	2020-09-09	코로나·장마에 갈수록 줄어드는 취업자 ... "9월은 더 걱정"	0.9975
6	2019-03-10	현대자동차·현대중공업 노사 '다른 행보'	0.9974
	2019-01-30	광주형 일자리 '2라운드', 이번엔 현대차 투자 이끌어낼까	0.9972
	2020-08-07	인천공항 자회사 직원들 실직 위기 ... 인천공항공사 "구제 어렵다"	0.9967
	2020-07-01	인국공 이어 ... 철도研 비정규직, 정규직 전환 논란	0.9960
	2019-05-09	버스 대란 폭풍전야 ... 광주·대구도 파업 압도적 찬성	0.9954
7	2019-07-01	조현병 환자도 사람입니다	0.9982
	2020-04-23	'엄마를 이렇게 살게 해선 안될 거 같아' 그녀의 이혼을 돕기로 했다, 우린 동거에 돌입했다	0.9980
	2019-01-09	사라진 왕국의 성에 그와 나	0.9980
	2020-09-05	금촌고시원 오 원장은 도망자의 짐을 버리지 않는다	0.9973
	2019-04-17	"정말 꿀자" 이 악물고 나온 날 ... '출소뽕' 한방에 다시 지옥행	0.9964
8	2019-07-17	트럼프 '中 관세폭탄' 협박에 급제동 ... S&P 0.3%↓	0.9981
	2020-04-06	'저유가 장기화' 전망, "미국도 감산 나서라"	0.9977
	2019-06-14	"대중 무역전쟁 제발 그만" ... 美 600여 개 기업·단체 트럼프에 서한	0.9974
	2019-05-31	'무질서의 세계' 대비해야 한다	0.9972
	2019-11-12	트럼프 "대선에 도움 안 돼" ... 유럽車 관세 부과 결정 또 미뤘다	0.9970

## 2. 토픽 및 감성 트렌드

분석 대상 기간의 일자리 기사 변화 특성을 파악하기 위해 토픽 비중의 합계 추세와 감성 추세를 활용했다. [그림 4]처럼 토픽 비중의 합계 추세에서 가장

(그림 4) 토픽별 비중 합계 월별 추이

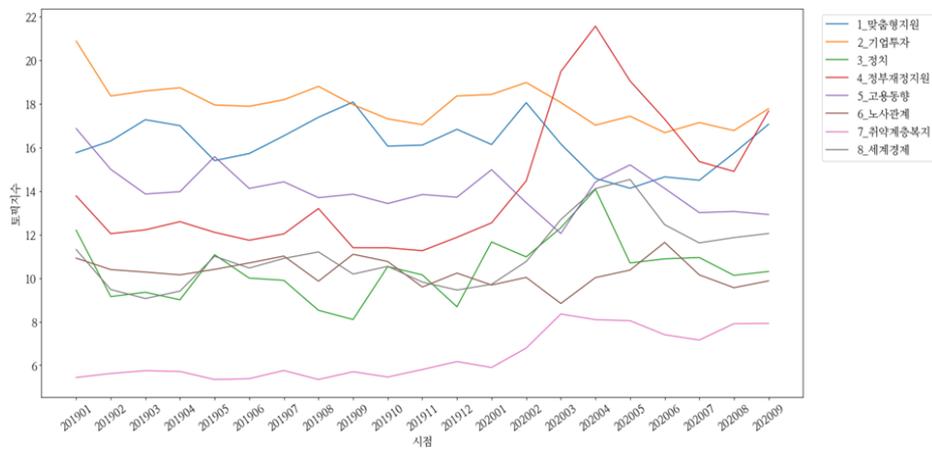


큰 변화는 2019년 12월부터 2020년 8월 사이에 나타났으며, 이는 코로나19의 발병과 유행 기간에 해당한다. 먼저, 원래 비중이 가장 큰 맞춤형 지원은 2019년 12월 이후 감소하여 2020년 3월에 저점을 보였다. 기업투자와 고용동향도 2019년 12월부터 2020년 3월까지 감소하였다. 반면, 정부재정지원과 정치 및 세계경제가 2020년 1월부터 증가하여 4월에 정점에 올랐고 7월까지 감소세를 보였으나 대체로 2019년보다 높은 비중을 유지했다. 비슷한 시기에 취약계층복지 토픽도 증가했다. 특히 정부재정지원은 2019년 비중이 높지 않았으나 2020년 3월 가장 비중이 큰 토픽으로 나타났다.

토픽지수 월별 추이를 보면 [그림 5]처럼 일부 구간에서 토픽 비중 합계와 다른 패턴이 나타난다. 토픽지수는 연중 변화폭이 완만한 편이다. 이는 시의성이 적은 편인 핵심 단어(각 토픽의 주요 단어)만 활용하여 토픽지수를 계산했기 때문이다. 그러나 정부재정지원은 토픽 비중 합계 방식(그림 4)과 마찬가지로 [그림 5]처럼 2020년 이후 크게 상승하여 4월에 정점을 보였다. 이는 정부재정지원 기사에 관한 관심이 폭증했음을 시사한다. 비슷하게 취약계층복지 토픽 비중 방식은 변화가 [그림 4]처럼 완만했지만, 토픽지수는 [그림 5]처럼 2020년 1월부터 3월까지 급격하게 증가하여 9월까지 증가한 수준을 유지했다. 이는 장애, 노인, 서비스, 코로나, 여성, 복지 등 취약계층과 관련된 단어가 일자리 기사 전반에서 늘었다는 것을 의미한다. 반면 기업투자의 토픽 비중은 2020년 초 급

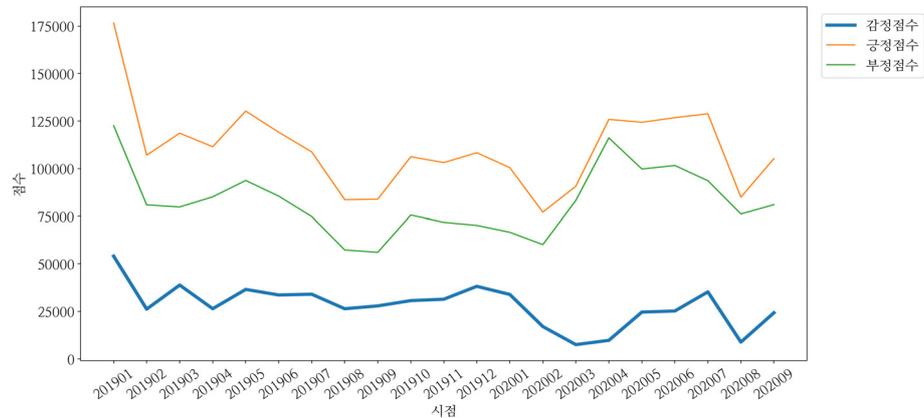
격히 감소했으나, 토픽지수는 대체로 일정하여 일자리 기사가 주제를 가리지 않고 기업투자 관련 단어를 꾸준히 활용하는 것으로 나타났다. 마지막으로, 맞춤형 지원 토픽에서 토픽 비중과 토픽지수가 다른 패턴을 보인다. 토픽 비중 합에서는 2020년 3월에 저점을 보인 뒤 7월까지 분명하게 증가했지만, 토픽지수는 같은 기간에 계속 낮은 수준을 유지했다. 이는 유사한 제목과 내용의 기사가 많았음을 의미한다.

[그림 5] 토픽지수 월별 추이



[그림 6] 감정점수 월별 추이

(단위: 점)

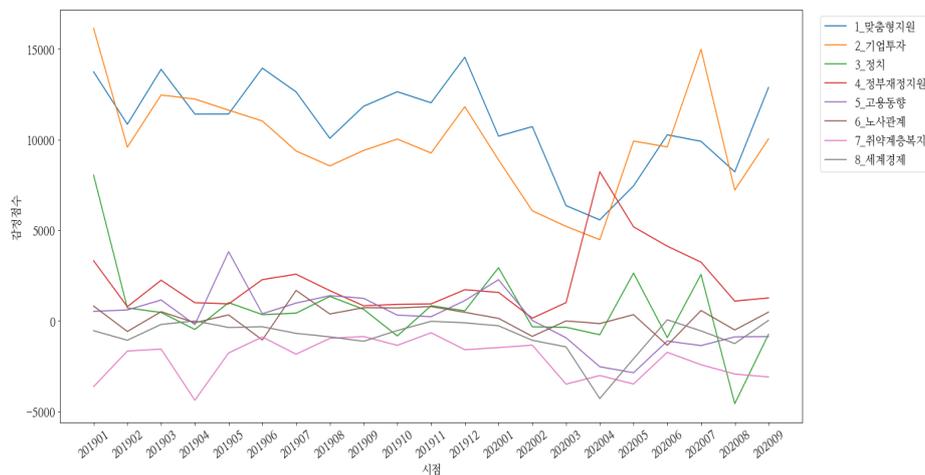


감성 트렌드는 긍정점수에서 부정점수를 뺀 감정점수의 월별 변화로 확인한다.7) [그림 6]에서 알 수 있듯이 분석 대상 기간 내내 긍정점수가 부정점수보다 높았으나, 두 점수의 격차는 시점에 따라 달라졌다. 토픽과 마찬가지로 감정도 2019년 12월부터 2020년 8월 사이에 변화가 컸다. 감정점수는 2019년 12월부터 계속 감소하여 2020년 3월에 가장 낮았다. 이후 점수가 올라서 7월에는 2019년과 비슷한 수준이었고 8월에 급격히 낮아졌다.

[그림 7]에서 토픽별 감정점수를 보면 평균적으로 가장 높은 감정점수를 보이는 주제는 기업투자와 맞춤형 지원이었다. 두 주제의 감정점수는 2019년 12월을 기점으로 2020년 4월까지 감소한 후에 다시 회복했다. 반면, 정부재정지원은 2020년 2월부터 4월까지 증가했으나, 4월부터 8월까지 계속 감소했다. 즉, 2019년 12월부터 모든 토픽의 감정점수가 하락했으나, 정부재정지원의 감정점수가 2020년 2월부터 큰 폭으로 증가하면서 종합 감정점수(그림 6)가 3월부터 반등했다. 이는 코로나19로 인한 정부정책 발표 및 시행과 관련이 깊은 것으로 보인다.

(그림 7) 토픽별 감정점수(합계) 월별 추이

(단위: 점)



7) 긍정과 부정은 같은 추세로 변화하는데 이는 기사 수의 증감과 대체로 일치한다. 감정점수는 텍스트 분량에 비례하기 때문이다. 따라서 긍정점수에서 부정점수를 빼면 기사량에 의한 효과를 상쇄할 수 있다.

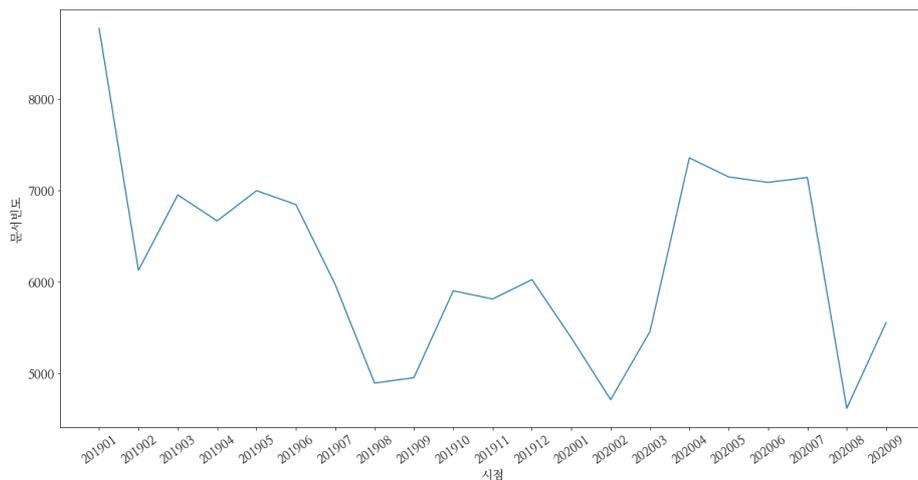
### 3. 코로나19 발병과 일자리 토픽 특성

일자리 관련 기사에 나타난 토픽과 감정점수의 변화는 코로나19의 발병과 전염 양상에 따른 일자리 이슈 변화를 드러낸다. [그림 8]에서 알 수 있듯이 국내에서 코로나19 전염병 문제가 부각된 2020년 1월부터 2기가 시작되는 2월까지 일자리 기사 수가 줄었고 2기가 진행된 3~4월에 일자리 기사 수가 증가하면서 토픽 간 비중 변화가 커졌다. 3기가 시작되는 8월(그림 1)에는 전체 일자리 기사 수가 감소하면서 변곡점을 보인다.

<표 4>와 같이 2020년 들어 전월 대비 비중이 가장 많이 증가한 토픽은 1월에 정치(66.1%), 2월에 정부재정지원(36.4%), 3월에 정부재정지원(106.3%)과 취약계층복지(44.1%), 4월에 고용동향(97.7%), 정치(73.7%), 세계경제(68.0%), 5월에 기업투자(46.1%), 6월에 노사관계(48.5%)였다. 7월은 변화가 크지 않았고, 8월에는 모든 토픽 비중이 감소했다. 이는 팬데믹 이후 일자리 관련 관심이 코로나19 초기 대응을 위한 정치 논의에서 시작해 재정지원과 취약계층복지에 관한 논의로 이어졌고 1/4분기 이후 고용동향과 세계경제 이슈가 주목받은 후에 기업투자와 노사관계로 이어지는 순서를 보였다. 또한, 2020년 4월에 치러진 21대 총선도 영향을 준 것으로 보인다.

[그림 8] 월별 일자리 기사 수

(단위: 건)



<표 4> 토픽별 비중 합계

토픽	연월	2020								
	2019	1	2	3	4	5	6	7	8	9
맞춤형지원	1615.2	1299.7	1318.3	1113.5	1194.7	1279	1434.2	1487.6	1071.1	1251.5
기업투자	1183.2	915.5	720.9	732.7	<b>794.5</b>	<b>1161</b>	1130.6	1404.4	742.6	967.7
정치	<b>538.5</b>	<b>894.3</b>	608.3	719.6	<b>1250.1</b>	765	839	877.8	480.1	511.8
정부재정 지원	466.6	<b>452.1</b>	<b>616.4</b>	<b>1271.4</b>	1742.3	1363.4	1137.8	992.3	632.1	935.1
고용동향	778.2	686.8	469.7	<b>337.9</b>	<b>667.9</b>	809.1	747.5	689.3	483.8	481.3
노사관계	562.1	418.8	317.8	307	465.9	<b>472.1</b>	<b>701.3</b>	631.8	315.5	418.8
취약계층 복지	564.9	454.8	406.7	585.9	588.2	690.9	567.2	618	550.4	592.4
세계경제	289	243.3	235.4	<b>368.3</b>	<b>618.6</b>	574	498.5	407.9	324.3	373.1

<표 5> 토픽별 '코로나' 등장 빈도

(단위: 건, %, 회)

토픽	문서 빈도	'코로나' 포함 문서 빈도	'코로나' 포함 문서 비율	문서 내 '코로나' 단어 빈도	문서당 '코로나' 출현 빈도
맞춤형지원	28,404	3,425	12.1	7,807	0.27
기업투자	18,519	2,723	14.7	7,550	0.41
정치	12,221	2,346	19.2	7,677	0.63
정부재정지원	10,434	5,955	57.1	25,454	2.44
고용동향	10,756	2,510	23.3	12,136	1.13
노사관계	7,214	935	13.0	2,604	0.36
취약계층복지	6,703	1,763	26.3	8,205	1.22
세계경제	5,154	2,218	43.0	12,245	2.38

코로나19와 관련이 높은 토픽은 <표 5>처럼 정부재정지원, 세계경제, 고용동향, 취약계층복지였다. 그러나 세계경제는 국내 일자리 문제와 독립적이었다. 고용동향도 통계 중심의 정보전달 기사가 많았다. 따라서 정부재정지원과 취약계층복지 기사를 중심으로 살펴보았다.

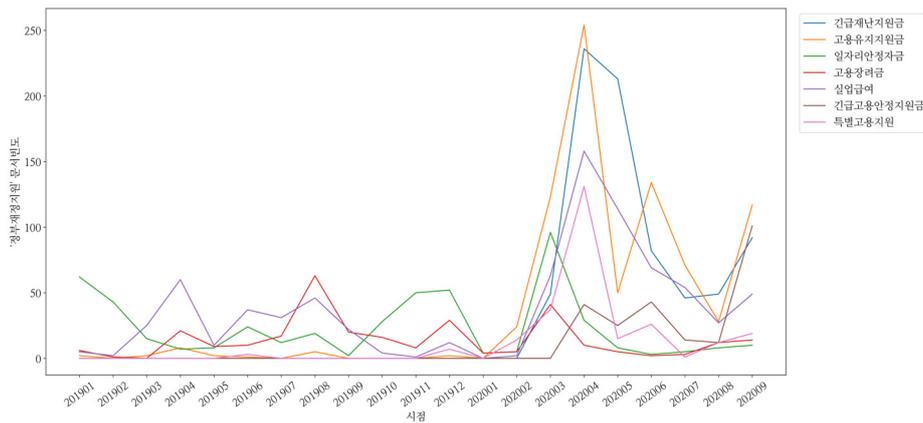
먼저, 정부재정지원 기사 중 <표 6>과 같은 주요 고용지원정책 명칭이 포함된 문서 빈도 변화를 확인한 결과는 [그림 9]와 같았다. 일자리안정자금 기사는 2020년 3월에 가장 많았고 고용유지지원금, 긴급재난지원금, 실업급여, 특별고용지원, 긴급고용안정지원금 기사는 4월에 가장 많았다. 기사 수가 가장 많았던 2020년 4월에는 고용유지지원금과 긴급재난지원금 제도에 관한 언급이 많았다.

<표 6> 코로나19 발병 이후 출현 빈도가 증가한 주요 고용지원책

명칭	특징
긴급재난지원금	<ul style="list-style-type: none"> <li>- 재난구호 및 재난복구를 위하여 지급함</li> <li>- 코로나19 위기 대책의 일환으로 재산, 소득에 관계없이 전 국민 가구를 기준으로 지급함</li> </ul>
고용유지지원금	<ul style="list-style-type: none"> <li>- 경기 변동 및 산업구조 변화로 고용조정이 필요한 경우 사업주의 경영 부담을 완화하고 근로자의 실직을 예방할 목적으로 지급함</li> <li>- 고용조정이 필요한 사업주가 근로자를 감원하지 않고 근로시간 조정, 휴업, 휴직 등 고용유지조치를 실시하여 고용을 유지하는 경우 지급</li> <li>- 코로나19로 지원금액, 업종 및 기간을 확대함</li> </ul>
일자리안정자금	<ul style="list-style-type: none"> <li>- 최저임금 인상에 따른 소상공인 및 중소기업의 경영 부담을 완화하고 노동자의 고용안정을 위하여 지급함</li> <li>- 최저임금을 준수하고 고용보험에 가입된 30인 미만 고용사업주에게 지급</li> </ul>
고용장려금	<ul style="list-style-type: none"> <li>- 장애인근로자의 직업생활 안정을 도모하고 고용촉진을 유도하기 위하여 지급함</li> <li>- 의무고용률을 초과하여 장애인을 고용하는 사업주에게 지급</li> </ul>
실업급여	<ul style="list-style-type: none"> <li>- 실업으로 인한 생계불안을 극복하고 생활안정을 도와 재취업 기회를 지원할 목적으로 지급함</li> <li>- 고용보험 가입 근로자가 실직하여 재취업 활동을 하는 기간에 지급</li> </ul>
긴급고용안정지원금	<ul style="list-style-type: none"> <li>- 고용보험의 보호를 받지 못한 취약계층의 생계안정을 도모할 목적으로 지급함</li> <li>- 코로나19로 소득이 감소한 특수형태근로종사자, 프리랜서, 영세자영업자, 무급휴직근로자에게 지급</li> </ul>
특별고용지원	<ul style="list-style-type: none"> <li>- 여행업, 관광숙박업 등 코로나19 피해 심각 업종에 고용유지지원금과 직업훈련 등 정부 지원 강화</li> </ul>

(그림 9) 정부재정지원 주제의 기사 빈도

(단위: 건)



특히, 긴급재난지원금은 정부재정지원 주제 기사에서 총 2,250번 나타났다. 이는 둘째로 높은 빈도인 505번을 기록한 정치 주제 기사보다 4배 이상 높은 빈도를 보였다. 국내에서 긴급 지원금에 관한 논의는 2020년 3월부터 시작되어 3월 30일 제3차 비상경제회의에서 처음으로 소득 하위 70%의 국민에 대한 지원금 지급을 결정했다.<sup>8)</sup> 4월에 지원금 대상에 관한 논의를 거쳐 4월 24일 전 국민 대상 긴급재난지원금 지급을 발표했다.<sup>9)</sup> 5월 4일부터 기초수급자 등 취약 계층에 지원금을 우선 지급하고 5월 11일부터 일반 국민도 지원금 신청을 시작했다.<sup>10)</sup> 이처럼 정부재정지원 토픽은 정치권의 긴급재난지원금 논의 시점과 맞물려 증가했으며 지급 신청이 시작된 5월부터 감소했다. 이는 권충훈·이형하(2020)의 연구에서 긴급재난지원금에 관한 주요 신문 기사가 5월부터 감소했다는 결과와도 일치한다.

8) 청와대 발표자료(2020. 3. 30), 「문 대통령 ‘소득 하위 70%, 4인 가구 기준 100만 원 긴급재난지원금’」. <https://www.korea.kr/news/policyNewsView.do?newsId=148870988> (검색일: 2021년 2월 27일)

9) 기획재정부 발표자료(2020. 4. 23), 「정부, 전 국민에 긴급재난지원금... 신청 않거나 기부 하면 세액공제」. <https://www.korea.kr/news/policyNewsView.do?newsId=148871819> (검색일: 2021년 2월 27일)

10) 기획재정부·행정안전부 발표자료(2020. 4. 30), 「긴급재난지원금 5월 11일부터 온라인 신청... 13일부터 지급」. <https://www.korea.kr/news/policyNewsView.do?newsId=148872032> (검색일: 2021년 2월 27일)

취약계층복지 기사는 <표 7>처럼 2019년 12월부터 2020년 2월까지 감소했다가 다시 5월까지 증가했다. 취약계층복지 토픽 비중이 0.5 이상을 넘는 기사는 월평균 354.6건이었고, 2020년 4월 350건에서 5월 497건으로 147건(42%)이 증가하였다. 6월은 327건으로 5월보다 170건이 줄었다. 코로나19 발생 후 확산세가 커진 2020년 3월부터 취약계층복지 기사의 절반 이상을 코로나 관련 기사가 차지했다. 취약계층복지 기사 중 코로나19와 관련한 내용에는 전염병으로 인한 노인 일자리 사업 중단, 경로원 및 어린이집 휴업, 근로시간 단축으로 인한 임시 일자리 소득 감소, 중년여성 일자리 감소, 불법 체류자 방역 문제, 소규모 여행업 타격 등이 있었다. 코로나19와 무관한 기사는 장애인 일자리 부족, 고령 근로자 처우, 플랫폼 노동자 보호 등이 있었다. 주로 고용 취약계층이거나 사각지대에 있는 사람들이 겪는 어려움에 관한 것이었다.

<표 7> 코로나19 발발 이후 취약계층복지 기사 변화

(단위: 건, %)

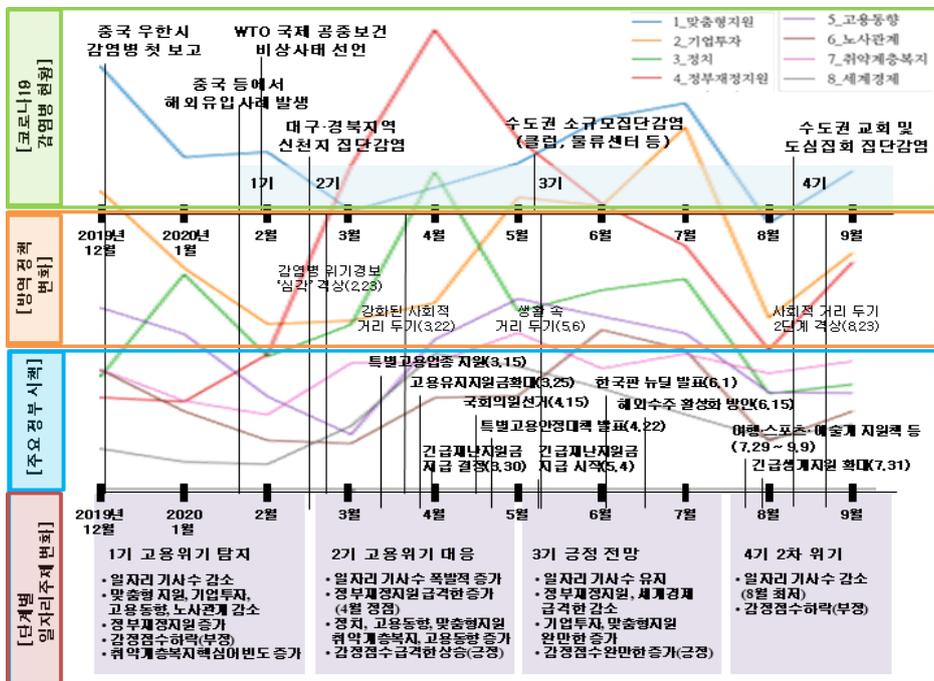
기간	문서 빈도	긍정 문서 빈도	‘코로나’ 포함 문서 빈도	‘코로나’ 포함 문서 비율
2019. 12	302	115	0	0
2020. 1	274	117	9	3
2020. 2	274	108	128	47
2020. 3	386	118	234	61
2020. 4	350	110	212	61
2020. 5	497	141	338	68
2020. 6	327	132	199	61
2020. 7	394	149	212	54
2020. 8	352	99	208	59
2020. 9	390	123	221	57

## V. 결론

언론 기사를 통해 파악한 일자리 이슈는 맞춤형지원, 기업투자, 정치, 정부재정지원, 고용동향, 노사관계, 취약계층복지, 세계경제와 관련된 주제인 것으로 나타났다. 이는 2010년부터 2018년까지의 고용 관련 언론 기사를 수집하여 토픽 모델링 분석을 통해 6개의 고용정책 의제를 확인한 연구(김경동·이시영·고길근, 2020)와 비교해 보면 정치와 정부재정지원 주제가 새롭게 등장했다. 이는 코로나19 같은 쉽게 통제할 수 없는 위험에 대해 정치권의 적극적 대응이 필요했고 이 결과로 정부재정투자가 이뤄진 것을 알 수 있다.

일자리에 대한 코로나19의 영향을 살펴보면 [그림 10]과 같다. 코로나19가

[그림 10] 일자리와 코로나19의 시점별 비교



자료: 중앙방역대책본부(2020. 12. 31)의 자료 재구성.

발생한 시점인 2019년 12월 이후부터 2020년 9월까지의 코로나19 감염병 현황에 따른 방역 정책과 정부 대책을 토픽별 트렌드와 비교해 일자리 주제 변화를 4단계(1기 고용위기 탐지, 2기 고용위기 대응, 3기 긍정 전망, 4기 2차 위기)로 구분했다.

1기 고용위기 탐지 단계에서는 일자리 문제보다 중국 우한시와 신천지 등에서 감염병 발생에 관한 관심으로 코로나19 이전부터 시행했던 청년이나 취약계층에 대한 맞춤형 정부지원정책과 기업의 투자계획 및 노사관계에 관한 관심이 상대적으로 줄어들었다. 반면에 정부가 사회경제의 불안 요소인 감염병 위기에 적극적으로 대응해야 한다는 필요성을 강조하는 정부재정지원 주제는 증가했다. 일자리 문제는 곧바로 취약계층의 복지에 영향을 주므로 취약계층복지와 같은 부정 용어의 출현 빈도가 증가해 전체적으로 감정점수가 하락하여 부정 기류가 보였다. 2기 고용위기 대응 단계에서는 코로나19로 인한 고용위기에 적극적으로 대응하던 기간으로 1기의 정부 대응 필요성을 강조하는 기사에서 실제 정부재정투입 관련 기사가 급격히 증가했다. 이는 외부 위험에 대응하는 정부 노동정책이 실제 현장에서의 집행으로 이어져 정치, 고용동향, 맞춤형 지원, 취약계층 복지, 고용동향 기사도 함께 증가했다. 이로 인해 감정점수가 1기에 비해 급격히 상승해서 전체적으로 긍정 기류가 보였다. 3기 긍정 전망 단계에서는 일자리 기사 수가 2기와 비슷했다. 이는 2기에 정부재정 투입이 시행된 결과로 정부재정지원 기사 수는 급격히 감소하는 반면에 실제 정부정책에 대한 단기 결과인 기업투자와 맞춤형 지원 기사 수는 완만한 증가세를 보였기 때문이다. 감정점수도 2기에 이어 완만히 증가해서 긍정 기류를 보였다. 4기의 2차 위기 단계에서는 일자리 기사 수가 8월 감소 후 증가로 이어져 언론의 관심이 늘어난 것을 확인했다. 특히 8월에 기사 수가 가장 적었다. 그러나 9월에 다시 일자리 기사 수가 증가했고 감정점수는 하락해서 부정 기류를 보였다.

이처럼 일자리 문제는 코로나19와 같은 외부 변수에 강한 영향을 받는 것으로 나타났다. 즉, 코로나19로 인한 집단감염이 발생할 때마다 거리두기와 같은 방역정책 시행으로 고용환경에 즉각적인 영향을 주었고, 정부는 위기마다 정부재정을 적극적으로 투입해서 고용안정에 힘을 쏟았다. 이런 정부의 위기 대응 노력에서 언론의 긍정 기류를 감지할 수 있었다. 이는 정부정책의 방향에

따라 언론의 관심이 긍정이나 부정으로 변화하는 것을 간접적으로 확인할 수 있으며 정부의 정책 방향은 재정 투입 방향에 크게 영향을 받는 것으로 나타났다.

코로나19는 현재 진행 중이다. 이 연구는 코로나19 발발과 확산의 초기 단계를 분석하였으므로 정부의 초기 정책 대응이 일자리 의제의 중심이 되었지만, 이러한 경향이 전염병을 극복해 나가는 과정에서도 비슷한 추세를 보일지는 향후 연구가 필요하다.

본 연구의 한계로는 다음과 같은 것들이 있다. 첫째, 감성 분석이 사전에 기반해서 이루어졌다는 점이다. 사전 기반 분석은 이 중 부정에 관한 내용을 명확히 파악할 수 없으므로 언론의 긍·부정 기류에 대한 해석에 난해한 측면이 있다. 향후 문장 기반으로 단어의 순서까지 고려하는 딥러닝 기반의 언어모델 분석 방법의 적용이 필요해 보인다. 둘째, 일자리에 영향을 줄 수 있는 코로나19 이외의 외부 사건을 고려하지 못했다. 대표적으로 2020년 4월 15일에 시행한 국회의원 총선 등이 어느 정도 일자리에 영향을 주었으리라 판단하지만, 그 영향을 고려하지 못했다.

## 참고문헌

- 고광이·오승원·백장선(2020). 「토픽 모델을 이용한 경기변동 토픽지수 개발 및 KOSPI200 지수에 대한 토픽지수 회귀모형」. 『한국데이터정보과학회지』 31 (4) : 579~594.
- 권충훈·이형하 (2020). 「주유 신문사 뉴스 기사를 통해 살펴본 2020년도 대한민국의 긴급재난지원금」. 『한국컴퓨터정보학회 하계학술대회 논문집』 28 (2) : 169~170.
- 김경동·이시영·고길곤(2020). 「텍스트 마이닝을 활용한 고용정책 프레임 연구: 언론기사 및 국회회의록의 비교·분석을 중심으로」. 『한국사회와 행정연구』 30 (4) : 135~163.
- 김양우(2019). 「인구절벽사회에서 노인과 일자리 감성분석」. 『한국콘텐츠학회

논문지』 20 (11) : 110~118.

김영민 · 정석재 · 이석준(2014). 「소셜 미디어 감성분석을 통한 주가 등락 예측에 관한 연구」. 『Entrue Journal of Information Technology』 13 (3) : 59~70.

김영아 · 박윤희 · 차수민 · 이해진(2020). 「뉴스기사 빅데이터를 활용한 노인일자리 트렌드 분석 : 2010~2019년」. 『대한고령친화산업학회지』 12 (1) : 13~20.

김정화 · 윤서연(2018). 「오피니언 마이닝을 이용한 도로교통서비스 정책이슈 및 만족도 변화 분석」. 『한국도로학회논문집』 21 (5) : 65~74.

방한솔 · 문호석(2019). 「텍스트마이닝을 이용한 텍스트의 주요 토픽을 시계열적으로 표현하는 방법론 연구」. 『한국데이터정보과학회지』 30 (6) : 1259~1276

백영민(2020). 「R을 이용한 텍스트 마이닝」. 한울아카데미.

이소정(2020). 「잠재디리클레할당 분석을 이용한 ‘노인일자리’ 관련 신문기사 토픽분석」. 『Journal of Digital Convergence』 18 (10) : 537~546.

이정훈(2018). 「감성분석 연구 동향」. 『춘계학술발표대회 논문집』 25 (1) : 359~361.

전혜영(2019). 「빅데이터를 통해 살펴본 장애인고용 동향분석 : SNS와 온라인 뉴스를 중심으로」. 『장애와 고용』 29 (2) : 55~82.

정지원 · 이재민 · 최소연(2018). 「텍스트마이닝 기법을 통한 언론에서의 장애인 노동의제 분석」. 『장애의 재해석』 pp.48~100.

최창식 · 임영호(2021). 「대통령 관련 보도의 감성 분석과 정파성의 지형 : 신문별 감성지수의 차이, 일관성, 대통령 지지도와의 관계를 중심으로」. 『한국언론학보』 65 (1) : 35~70.

Blei, D. M., A. Y. Ng, and M. I. Jordan(2003). “Latent Dirichlet Allocation”. *Journal of Machine Learning Research* 3 : 993~1022.

Chang, J., S. Gerrish, C. Wang, J. L. Boyd-Graber, and D. M. Blei(2009). “Reading Tea Leaves : How Humans Interpret Topic Models”. In *Advances in Neural Information Processing Systems*. pp.288~296.

- Jacobi, C., Van W. Atteveldt, and K. Welbers(2016). “Quantitative Analysis of Large Amounts of Journalistic Texts Using Topic Modelling”. *Digital Journalism* 4 : 89~106.
- Mimno, D. and A. McCallum(2008). “Topic Models Conditioned on Arbitrary Features with Dirichlet-multinomial Regression”. *The 24th Conference on Uncertainty in Artificial Intelligence*. pp.411~418.
- Newman, D., J. H. Lau, K. Grieser, and T. Baldwin(2010). “Automatic Evaluation of Topic Coherence”. In *Human Language Technologies : The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pp.100~108.
- OECD(2020). “OECD Employment Outlook 2020 : Worker Security and the COVID-19 Crisis”. OECD Publishing.
- Pang, B. and L. Lee(2008). “Opinion Mining and Sentiment Analysis”. *Foundations and Trends in Information Retrieval* 2 : 1~135.
- Sohrabi, C., Z. Alsafi, N. O’Neill, M. Khan, A. Kerwan, A. Al-Jabir, C. Iosifidis, and R. Agha(2020). “World Health Organization Declares Global Emergency : A Review of the 2019 Novel Coronavirus (COVID-19)”. *International Journal of Surgery* 76 : 71~76.
- Zhang, L. and B. Lju(2017). “Sentiment Analysis and Opinion Mining”. In C. Sammut & G. I. Webb (eds.), *Encyclopedia of Machine Learning and Data Mining*. Boston, MA : Springer.

군산대학교 한국어 감성사전(<http://dilab.kunsan.ac.kr/knusl.html>)

대한민국 정책브리핑(<http://www.korea.kr>)

## Employment News Agenda and Trends Analysis in the Face of the COVID-19 Pandemic using Text Mining

Yang, Jiyoun · Son, Mina · Noh, Seungmin

This study identified employment news agenda and changes of issue priorities and sentiment in the face of the Covid-19 pandemic using text-mining techniques. As a result of LDA-based topic analysis on employment news, eight topics emerged : Customized Employment Service, Corporate Investment, Politics, Government-funded Program, Employment Situations, Labor-management Relations, Support for the Disadvantaged, Global Economy. Priorities of the eight topics had significantly fluctuated since the Covid-19 occurred in Korea in January 2020. While Customized Employment Service, Corporate Investment, and Employment Situations had decreased, Government-funded Program, Politics, Global Economy, and Support for the Disadvantaged had increased in the aggregate topic weight. Positivity in text sentiment had declined as positively-valenced topics of Customized Employment Service and Corporate Investment shrunk in the initial stage of the crisis. However, positive sentiment had grown after Government-funded Program had been prioritized in the news agenda. Various topics had risen in chronological order after the Covid-19 occurrence, from Politics to Government-funded Program and Support for the Disadvantaged to Employment Situations and Global Economy, and lastly to Corporate Investment and Labor-management Relations. Consistent with the previous research on news topics and trends, our analysis revealed that employment news set agenda based on government policy and sensitively reacted to internal and external factors. Notably, the employment

agenda in the news showed flexible priorities in response to changing states of the Covid-19 spread and the progress of public policy discussions. Furthermore, our data collection with a keyword of ‘jobs(iljali)’ allowed a more extensive review of employment agenda than previous literature with a specific-policy focus.

Keywords : jobs, topic model, sentiment analysis, Covid-19