

코로나 백신에 정부 정책 발표에 대한 여론 형성에 관한 Causal Impact 분석

구현지*, 박채연*, 최지우*, 노성진*

Causal Impact Analysis on Public Opinion of Government Policy Announcement on COVID-19 Vaccine

Gu Hyeonji* and Park Chaeyeon* and Choi Jiwoo* and Noh Seongjin*

요 약

본 연구에서는 정부의 코로나 백신 정책이 사람들의 여론에 영향을 얼마나 미쳤는지 알기 위해, 빅인즈의 뉴스와 유튜브의 댓글 데이터를 수집하여 분석하였다. 먼저 여론의 동향을 알아보기 위해서 Word Cloud 분석을 수행하였고, Causal Impact 분석을 통해 정부 정책이 여론에 얼마나 영향을 분석하였다. 분석 결과 정부의 정책이 뉴스를 통해 통계적으로 유의미한 영향을 미치는 것을 확인하였다.

Abstract

In this study, to find out how much the government's COVID-19 vaccine policy affected people's public opinion, BigKinds's news and YouTube's comment data were collected and analyzed. First, Word Cloud analysis was performed to find out the trend of public opinion, and how much influence the government policy has on public opinion was analyzed through causal impact analysis. As a result of the analysis, it was confirmed that the government's policy had a significant through the news statistically.

Key words

Causal Impact, Word Cloud, COVID-19, Vaccine

1. 서 론

2019년 12월, 중국 우한에서 처음으로 코로나 바이러스의 변이체인 COVID-19가 처음 발견되었고 이후 새로운 유형의 변이 바이러스가 계속 나타나

고 있다. COVID-19의 빠른 종식을 위하여 국내에서는 2020년 7월 29일부터 코로나 백신에 대한 언급이 시작되었다. 2021년 2월 26일부터 백신 접종이 이루어지고 있다. 이에 따라 정부는 코로나 백신 관련 정부 지침을 계속해서 발표하고 있다. 따라서 본

* 이름: 구현지 소속: 계명대학교 email: ghj_0728@naver.com

* 이름: 박채연 소속: 계명대학교 email: pcy0721@naver.com

* 이름: 최지우 소속: 계명대학교 email: dnwchl741@naver.com

* 이름: 노성진 소속: 계명대학교 email: seongjin2221@daum.net

연구에서는 Causal Impact를 사용하여 정부가 발표한 정책이 사람들 인식에 얼마나 영향을 미쳤는지 알아보고자 한다.

II. 여론 분석을 위한 데이터 수집

코로나 정책들과 백신의 영향을 알아보기 위해 백신과 관련된 빅카인즈 뉴스와 유튜브 댓글 데이터를 수집하였다. 정부에서 코로나 백신 관련하여 발표한 것 중 첫 백신을 접종하기 시작한 2021년 02월 26일을 연구의 핵심 시점으로 선택하여 분석을 시작한다. 시계열 모델의 학습을 위해 시작일 기준 3개월 전 데이터를 수집하여 사용하였고, 검증은 시작일 후 2주간의 데이터를 사용하였다. 데이터는 일자별로 수집되었고, 빅카인즈 뉴스의 개수를 독립 변수로 유튜브 댓글 개수를 종속변수로 사용하였다.

III . 데이터 분석 방법

본 연구에서는 인과 분석을 위해 Causal Impact (코잘임팩트)를 사용하였다[1,5]. 코잘임팩트는 베이시안 시계열 모델을 활용해 타임시리즈에 대해 설계된 개입의 인과 관계를 추정하는 방법으로 R 프로그래밍 언어로 구현되었다[2].

첫 백신 접종과 관련하여 뉴스의 흐름을 파악하기 위해 연구 시점으로부터 오차범위 7일의 빅카인즈 뉴스 데이터 키워드를 사용하여 워드클라우드 (Word Cloud) 하겠다. 워드클라우드는 키워드를 직관적으로 파악할 수 있도록 빈도가 높은 핵심 단어를 시각적으로 돋보이게 해주는 시각화 기법이다[3, 4]. Python의 counter함수로 키워드의 빈도를 추출하여 워드클라우드 했다. <그림1> 빈도가 높은 상위 7개는 ‘접종’, ‘백신’, ‘코로나’, ‘앵커’, ‘국민’, ‘상황’, ‘시작’ 이다.

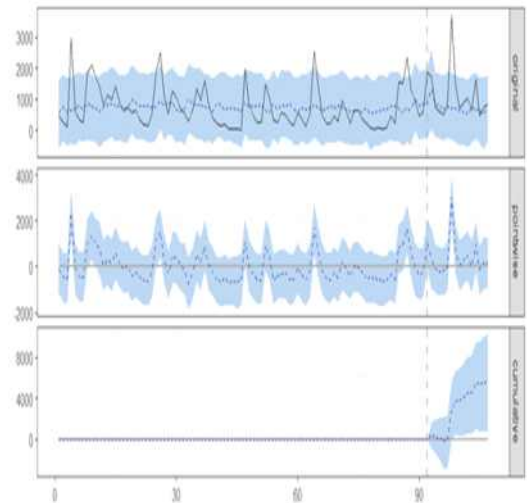
첫 백신 접종에 대한 사람들의 여론을 알아보기 위해 유튜브 댓글 데이터도 워드클라우드 하였다. 그 결과 ‘백신’, ‘먼저’, ‘접종’, ‘국민’, ‘대통령’, ‘사람’, ‘부작용’이 높은 빈도를 가진 것을 볼 수 있었다. <그림2> ‘먼저’와 ‘대통령’의 키워드를 통해 국민들은 대통령이 먼저 백신을 접종해야 한다는 여

론과 ‘접종’과 ‘부작용’의 키워드로 부정적인 여론이 형성된 것으로 보인다.



<그림 1> 빅카인즈 뉴스 워드클라우드 <그림 2> 유튜브 댓글 워드클라우드

코잘임팩트 분석에서 백신 첫 접종의 유튜브 뉴스 수와 빅카인즈 뉴스 수에 변화가 있다는 가정을 세웠으며 모든 분석에서 학습은 3개월, 예측은 2주로 하였다. 코잘임팩트의 플롯을 보면 세 개의 패널이 포함되어 있다<그림3>. 첫 번째 패널은 사후 처리 기간에 대한 데이터와 반실 예측, 두 번째 패널은 관찰된 데이터와 반실 예측의 차이 그리고 세 번째 패널에서는 두 번째 패널의 포인트별 기여를 추가하여 개입의 누적 효과를 줄여준 플롯이 보인다.



<그림 3> 코잘임팩트 플롯

첫 백신 접종 시점 후 예측 기간 3주 동안 반응 변수의 평균값은 약 1.11K였다. 반대로, 개입이 없었다면, 우리는 평균 0.73K의 응답을 기대했다. 이

반 사실적 예측의 95% 구간은 [0.44K, 1.04K]이다. 관측된 반응에서 이 예측을 빼면 개입이 반응 변수에 미치는 인과적 효과의 추정치가 산출된다. 이 효과는 0.38K이고 95% 구간은 [0.07K, 0.67K]이다. 예측 기간의 개별 데이터 점을 합하면 반응 변수의 전체 값은 16.63K였다. 반대로 개입이 이루어지지 않았다면 우리는 총액 10.89K를 예상했다. 이 예측의 95% 구간은 [6.57K, 15.57K]이다. 위의 결과는 절대 숫자로 제시되어 있다. 상대적인 측면에서 반응 변수는 +53%의 증가를 보였다. 이 백분율의 95% 간격은 [+10%, +92%]이다.

이는 개입 기간 동안 관측된 긍정적 효과가 통계적으로 유의하며 무작위 변동에 의한 것 같지는 않다는 것을 의미한다. 다만 이번 인상에도 실질적인 의미가 있는냐는 문제는 절대적 효과(0.38K)를 근본적 개입이라는 당초 목표와 비교해야만 답할 수 있다는 점에 주목해야 한다.

우연히 이 효과를 얻을 확률은 매우 작다(베이지안 단측 꼬리 면적 확률 $p = 0.012$). 이것은 인과적 효과가 통계적으로 유의하다고 간주할 수 있다는 것을 의미한다.

< 표 1 > 코잘임팩트 분석 결과

	빅카인즈 뉴스	
	Average	Cumulative
Actual	1109	16633
Prediction(s.d.)	726 (162)	10895 (2434)
95% CI	[425, 1054]	[6371, 15808]
Absolute. Effect(s.d.)	383 (162)	5738 (2434)
95% CI	[55, 684]	[825, 10262]
Relative effect(s.d.)	53% (22%)	53% (22%)
95% CI	[7.6%, 94%]	[7.6%, 94%]
Posterior tail-area probability p	0.01119	
Posterior prob. of a causal effect	0.01119	

IV . 결 론

본 논문에서는 백신 첫 접종의 유튜브 뉴스 수와 빅카인즈 뉴스 수에 변화가 있다는 가정을 코잘임팩트로 분석한 결과 여론에 통계적으로 유의미한

영향을 미친 것을 확인하였다. 본 연구에서는 영향이 긍정인지 부정인지에 대해서는 확인하지 않았다. 이후의 연구에서는 감성분석을 통해 긍정/부정까지 고려한 영향을 도출하는 것을 연구할 것이다.

참 고 문 헌

- [1] Martin, William, Federica Sarro, and Mark Harman. "Causal impact analysis for app releases in google play.", Proceedings of the 2016 24th ACM SIGSOFT International Symposium on Foundations of Software Engineering, p435-446, November 2016.
- [2] <https://cran.r-project.org/web/packages/CausalImpact/vignettes/CausalImpact.html>
- [3] Haroon, DANISH, "Python Machine Learning Case Studies: Five Case Studies for the Data Scientist", Apress, p232-235, September 2017.
- [4] 원다예, "이것이 인포그래픽이다", 한빛미디어, p258-259, August 2016.
- [5] Brian H. Ross, "The Psychology of Learning and Motivation", Elsevier, vol. 57, p 24-26, July 2012.