

Causal Impact 분석 기법을 접목한 COVID-19 팬데믹 전·후

메타버스 애플리케이션 리뷰의 토픽 변화 분석 (Analysis of Topic Changes in Metaverse Application Reviews Before and After the COVID-19 Pandemic Using Causal Impact Analysis Techniques)

이소원*, 노미진**, 한무명초***, Kim, YangSok****

(Lee, Sowon, , Mijin Noh, MuMoungCho Han, YangSok Kim)

요약

가상환경 기술의 발전과 COVID-19 팬데믹으로 언택트 문화가 부상함에 따라 메타버스(Metaverse)가 주목받고 있다. 본 연구에서는 최근 메타버스 서비스로 주목받는 “제페토” 애플리케이션에 대한 사용자들의 리뷰를 분석하여, COVID-19 팬데믹 이후 메타버스에 대한 요구사항의 변화를 확인하고자 하였다. 이를 위해 2018년 9월부터 2023년 3월까지 구글플레이스토어에 작성된 “제페토” 애플리케이션 리뷰 109,662건을 수집하였으며, LDA 토픽모델링 기법을 활용하여 토픽을 추출하고, COVID-19 팬데믹이 선언된 “2020년 3월 11일”을 기준으로 전·후로 토픽이 어떻게 변화했는지 Causal Impact 기법을 사용하여 분석하였다. 분석 결과 애플리케이션 기능적 문제(토픽1), 보안 문제(토픽2), 애플리케이션 내 가상화폐(Zem)에 대한 불만 사항(토픽3), 애플리케이션 성능(토픽4), 개인정보 관련 문제(토픽5) 등 5가지 토픽이 추출되었으며, 이들 중 보안 문제(토픽2)가 COVID-19 팬데믹에 가장 큰 영향을 받았음이 확인하였다.

■ 중심어 : Causal Impact ; LDA ; 토픽모델링 ; COVID-19 ; 메타버스 애플리케이션

Abstract

Metaverse is attracting attention as the development of virtual environment technology and the emergence of untact culture due to the COVID-19 pandemic. In this study, by analyzing users' reviews on the “Zepeto” application, which has recently attracted attention as a metaverse service, we tried to confirm changes in the requirements for the metaverse after the COVID-19 pandemic. To this end, 109,662 reviews of “Zepeto” applications written on the Google Play Store from September 2018 to March 2023 were collected, topics were extracted using LDA topic modeling technique, and topics were analyzed using the Causal Impact technique to examine how topics changed before and after based on “March 11, 2020” when the COVID-19 pandemic was declared. As a result of the analysis, five topics were extracted: application functional problems (topic1), security problems (topic 2), complaints about cryptocurrency (Zem) in the application (topic 3), application performance (topic 4), and personal information-related problems (topic 5). Among them, it was confirmed that security problems (topic 2) were most affected by the COVID-19 pandemic.

■ keywords : Causal Impact ; LDA ; Topic Modeling ; COVID-19 ; Metaverse Application

I. 서론

2019년 12월 중국 우한에서 시작된 COVID-19는 높은 전염력을 지닌 새로운 감염병으로 중국뿐만 아니라 전 세계로 빠른 속도로 확산하였다. 발생 3개월 만인 2020년 3월 11일 세계보건기구(WHO)는 COVID-19를 전염병 세계적 대유행인 팬데믹으로

선언하였다. 이에 따라 방역 정책으로 시행된 사회적 거리두기는 감염자와 비감염자가 섞이지 않도록 하여 감염병의 유행을 억제하는 핵심적인 방법으로 채택되었다[1].

사회적 거리두기의 실시와 지속으로 외부 활동이 제한되고 집에 머무르는 시간이 증가함에 따라 활동 공간은 실외에서 실내로 옮겨지면서, 이에 대한 대

* 학생회원, 계명대학교 경영정보학과

** 정회원, 계명대학교 디지털경영학부 경영빅데이터전공

*** 정회원, 동국대학교 와이즈캠퍼스 교양융합교육원

**** 정회원, 계명대학교 디지털경영학부 경영정보학전공

접수일자 : 2023년 12월 19일

수정일자 : 2024년 01월 23일

게재확정일 : 2024년 01월 30일

교신저자 : Kim, Yang Sok e-mail : yangsok.kim@kmu.ac.kr

응으로 비대면 문화가 활성화되었다. 이러한 상황에서 비대면 서비스의 수요가 급증하고, 디지털 전환이 가속화되면서, 가상환경 기술을 기반으로 한 메타버스(Metaverse)에 대한 관심과 주목이 증가하였다[2]. 메타버스는 'Universe(현실 세계)'와 'Meta(초월적인)'의 합성어로, 현실 세계를 넘어선 가상환경과 현실을 결합한 다중 사용자 환경으로 디지털 객체와 사람 간에 다중감각적 상호 작용(Multi-modal interactions)을 가능케 하는 특징을 가지고 있다[3,4]. 메타버스 관련 서비스·개발 엔진·하드웨어·네트워크 등의 생태계는 기하급수적으로 성장하고 있어 베인컴퍼니는 2030년 메타버스 시장이 최대 9,000억 규모로 성장할 것이라는 낙관적인 전망을 내놓고 있다[5,6].

메타버스 서비스는 주로 모바일 기반의 애플리케이션 형태로 제공되며, 대표적인 메타버스 서비스에는 제페토, 로블록스 등이 있다. 메타버스 서비스에 대한 사용자의 요구사항 파악은 추후 메타버스 관련 서비스와 시스템의 개발·유지·보수에 필수적인 정보로 여겨진다. 애플리케이션의 사용자 리뷰는 애플리케이션 사용자의 요구사항을 파악하여 서비스의 품질을 개선하고 보완하는 데 필요한 유용한 정보를 제공하지만[7,8,9], 현재까지 메타버스 사용자 요구사항에 대한 리뷰 분석 연구는 부족한 상황이다.

애플리케이션 리뷰 분석은 애플리케이션 운영에 필수적인 사항이지만, 매일 새롭게 작성되는 방대한 양의 리뷰를 일일이 정제하고 분석·해석하는 일은 많은 시간과 인력이 필요한 어려운 작업이다. 최근 토픽모델링 기법을 활용하여 리뷰를 자동으로 분석하여 토픽을 추출하는 연구가 활발히 진행되고 있다. 이전 연구들은 LDA(Latent Dirichlet Allocation)토픽모델링 기법을 사용하여 애플리케이션 만족도에 영향을 미치는 요인 및 이슈 분석을 중심으로 수행되었다[8,10,11]. 그러나 이러한 기존 연구들은 주로 토픽 키워드의 단순 비교분석에 그치고 있어, 특정 현상 발생이 토픽에 미치는 영향을 체계적으로 분석하는 데에 한계가 있다. 본 연구에서는 Causal Impact 분석 기법[12]을 LDA 토픽모델링

기법과 접목하여 특정 사건의 발생이 토픽에 미치는 영향을 체계적으로 분석하는 방법을 제안하고자 한다. 본 연구 결과는 추출한 토픽 및 토픽 변화를 통해 COVID-19 팬데믹 전·후 메타버스 서비스에 대한 사용자의 요구사항을 종합적으로 이해하고 추후 메타버스 애플리케이션 품질 향상에 기여할 것으로 기대한다.

II. 이론적 배경

1. 애플리케이션 리뷰 분석

애플리케이션 사용자 리뷰는 애플리케이션에 대한 사용자의 직접적인 피드백으로, 사용자의 관점에서 애플리케이션 운영에 중요한 정보를 제공한다[8,9]. 애플리케이션 리뷰 분석은 적기에 애플리케이션 관련 이슈를 탐지하고 대응할 수 있으며, 버그 수정·새로운 기능 추가·업데이트 시기 결정 등 애플리케이션의 유지·보수·개선에 유용한 정보를 준다[7,8,9]. Mudambi & Schuff[13]는 Amazon의 고객 리뷰에서 도움이 되는 리뷰의 특성을 파악한 연구를 진행하였고, 특히 리뷰의 길이·긍정적 표현·다양한 어휘 사용이 도움(Helpfulness)이 되는 경향으로 나타났다. Hong[14]는 온라인 리뷰 도움 정도의 결정 요인에 대한 메타분석으로 리뷰의 톤·길이·리뷰어의 신뢰성이 중요하다는 결론을 도출하였으며, 온라인 소비자들이 어떤 리뷰가 도움이 되는지를 인식하고 평가하는지에 대한 통찰력을 제공하였다. 정지훈[7]은 배달 애플리케이션인 배달의민족 리뷰 데이터에 LDA 토픽모델링 기법과 시계열 이상치 탐지 모형인 ARIMA(Autoregressive Integrated Moving Average) 모델을 결합하여 감성분석을 수행하였고, 리뷰 건수의 이상치가 발생한 요일에 대한 이슈 요인 분석으로 이슈 대응 프로세스를 수립하였다. 김선주[11]은 숙박 공유 애플리케이션인 Airbnb 리뷰 작성 매체별로 구분하여 국내 COVID-19 팬데믹 영향 전·후 시기별로 시기를 구별하여 LDA 토픽모델링 기법으로 비교 분석하였다.

2. LDA 토픽모델링

토픽모델링은 통계적 추론을 사용하여 문장 속에서 잠재된 주제를 찾고 데이터 및 텍스트 문서 간의 관계를 파악하여 의미 있는 토픽을 추출하여 다양한 관점을 제공하는 기술이다. 본 연구는 Blei[15]가 제안한 LDA 토픽모델링 기법을 사용한다. 그림 1은 LDA 토픽모델링 기법의 간략한 구조로 해당 모델의 컬렉션 내에 K개의 토픽이 있다. 각 토픽은 어휘에 대한 디리클레(η)에서 뽑힌 다항 분포를 갖고, 각 문서에 대하여 문서 내에서 토픽에 대한 분포 θ_d 를 디리클레(α)에서 선택한 후 문서 내 각 단어에 대한 토픽 인덱스 $Z_{d,n}$ 을 θ_d 에서 뽑아 선택된 토픽에서 관찰된 단어 $W_{d,n}$ 을 추출한다[17].

LDA 토픽모델링 기법은 트위터 이슈 트래킹[18], 금융 소비자의 불만 분석[19], 헬스케어 분야 텍스트 분석[20] 등에 활용되었다. 따라서 본 연구에서는 LDA 토픽모델링 기법으로 제페토 애플리케이션 사용자에게 대한 토픽을 추출하고자 한다.

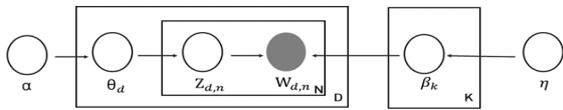


그림 1. LDA 토픽모델링 기법 구조(LDA Topic Modeling Technique Structure)[16]

3. Causal Impact 분석 기법

Causal Impact 분석 기법은 Brodersen[12]가 제안한 기법으로, 시계열 데이터에서 의도된 개입의 인과관계 효과추정을 Bayesian Structural Time - Series (BSTS) 모델을 기반으로 한다. A/B 테스트와 같은 일반적인 인과관계 추정 방법이 존재하지만, 마케팅 효과·웹 서비스 운영 등 분야에서 인과관계를 측정하거나 추정이 어려울 수가 있다. Causal Impact 분석 기법은 이 문제를 해결하기 위해 BSTS 모델 기반의 시계열 추정 방법을 활용한다. BSTS 모델은 상태-공간 모델과 베이지안 추론을 결합하여 시계열 데이터의 동적인 특성을 고려하며, 대조 사실이 없을 때도 적절한 예측 및 인과 추론을

가능하게 한다. 모델은 개입 발생을 반영하여 새로운 상태-공간 모델을 학습하고, 선발 기간과 후발 기간 간의 차이로 해당 개입의 영향을 추정한다. 결과는 주로 인과적인 영향의 크기와 방향을 나타내어 특정 이벤트가 시계열 데이터에 미치는 효과를 정량적으로 이해할 수 있다[12].

Martin[21]은 Google Play에서의 애플리케이션 출시(업데이트)와 사용자 행동에 미치는 영향을 효과적으로 분석하기 위하여 Causal Impact 분석 기법을 활용하였다. 이 연구는 Google Play에서의 애플리케이션 업데이트가 사용자들의 행동과 경험에 미치는 효과를 정량적으로 이해하고자 하며, 새로운 기능 도입이나 버그 수정과 같은 출시 영향 평가에 기여하였다. Feroze[22]는 COVID-19 팬데믹의 패턴을 예측하고 인도·브라질·미국·러시아·영국 등 전세계의 상위 다섯 국가의 봉쇄 조치에 대한 인과적 영향을 비교 분석하기 위하여 Causal Impact 분석 기법과 ARIMA 모델을 사용한 연구를 진행하였다. 연구 결과 Causal Impact 분석 기법이 ARIMA 모델보다 더 높은 정확도를 보였다. Navas[23]은 Causal Impact 분석 기법과 ARIMA 모델을 활용하여 COVID-19 팬데믹의 추세를 예측하고 백신의 인과적 영향을 정량적으로 분석하였다. 연구 결과는 몇 가지 예외를 제외하고 제안된 모델의 예측 정확도가 일반적으로 사용하는 ARIMA 모델보다 우수하다는 것을 시사하였다. 따라서 본 연구에서는 COVID-19 팬데믹의 영향을 정량적으로 분석하기 위해 Causal Impact 분석 기법을 활용한다.

III. 연구 방법

본 연구는 다음 그림 2의 연구 절차대로 연구를 수행한다.

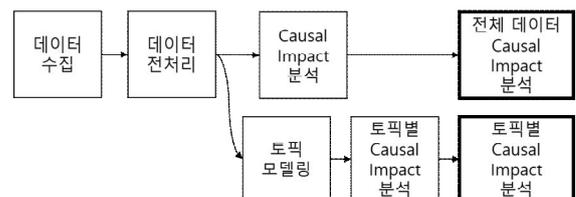


그림 2 연구 절차(Research Procedure)

1. 데이터 수집 및 전처리

데이터 수집 단계에서는 대표적인 애플리케이션 플랫폼인 구글플레이스토어를 선정하여 제페토 애플리케이션의 영어 리뷰를 수집한다. 이를 위하여 구글플레이스토어 애플리케이션의 사용자 리뷰 크롤러인 Google Play Scraper를 사용하여 2018년 9월부터 2023년 3월까지 작성된 총 109,662건의 리뷰를 수집한다. 수집 데이터의 항목은 사용자 ID·리뷰 작성 날짜·리뷰 내용·평점 등이다. 데이터를 수집한 후 데이터 전처리 단계에서는 중복적인 의견을 포함할 가능성이 큰 평점 3점을 제거한다. 문장 수가 적은 리뷰는 문맥이 부족하여 의미 있는 주제를 도출하기 어려울 수가 있어 단어 수가 5 이하인 문장을 제외한다. 정규 표현식을 사용하여 문장 부호 및 이모티콘 등을 삭제하고, NLTK·Sparse 패키지 등을 활용하여 토큰화를 진행하고 불용어를 제거한다. NLTK 패키지에 정의된 불용어 사전 활용뿐만 아니라 제페토·앱 등과 같이 빈도는 높지만, 분석에 큰 영향을 미치지 않는 단어 또한 불용어에 포함하여 제거하며, 표제어 추출로 단어의 표준화를 통하여 분석의 일관성을 유지한다.

2. LDA 토픽모델링

LDA 토픽모델링을 위해 본 연구에서는 gensim을 활용한다. 토픽모델링에서 토픽의 개수를 정하는 것은 중요한 문제다. 이를 위해 일반적으로 Wallach[24]가 제안한 토픽의 혼잡도를 나타내는 Perplexity score와 Newman[25]가 제시한 토픽의 일관성을 나타내는 Coherence score로 최적의 토픽 개수를 결정한다. Perplexity score는 주어진 모델이 얼마나 좋은 예측을 하는지를 나타내는 지표로, 산출된 값이 낮은 값일수록 모델의 예측이 더 정확하다고 판단한다[24]. 반면 Coherence score는 Perplexity score의 한계인 결과 해석의 어려움을 극복하는 방안으로 제안되었으며, 토픽 간의 의미적 유사성 측정으로 토픽이 의미론적으로 일관되게 구

성되었는지를 평가하는 지표로[25] 산출된 Coherence score 값이 클수록 각 토픽이 유사한 단어로 구성되었다고 해석한다. 따라서 본 연구는 Coherence score로 최적의 토픽 개수를 결정하고, 이는 토픽 모델의 의미적 일관성을 강화하고 해석의 용이성을 높일 수 있음을 기대한다.

3. Causal Impact 분석

Causal Impact 분석은 개입이 발생한 특정 시점에서의 반응 변수(이하 Y)의 counter-factual 값을 예측하고, 예측한 예상 시계열 데이터와 관찰 시계열 데이터 간의 차이로 특정 개입에 대한 Y의 예상 효과를 측정하여 인과관계를 분석한다. Causal Impact 분석에서 X는 개입 또는 반응 변수에 영향을 주는 변수로 처리 변수를 의미하고, Y는 개입에 영향을 받는 변수로 반응 변수를 의미한다. 즉, X는 시계열 데이터에서 관찰되는 특정 사건이나 처리를 뜻하는 개입의 시점 혹은 정도를 나타냄으로 본 연구에서는 X를 CDC에서 발표한 COVID-19 팬데믹 선언일인 “2020년 3월 11일”로 설정한다. COVID-19 팬데믹이 제페토 애플리케이션 사용에 영향을 미쳤는가를 수행하기 위한 Y는 전체 리뷰의 연도별 주별 합계값으로, COVID-19 팬데믹은 제페토 애플리케이션 사용자에게 대한 토픽의 변화에 영향을 미쳤는지에 대한 Y는 추출한 토픽의 연도별 주별로 합산한 값으로 설정한다. 이를 통하여 COVID-19 팬데믹이 제페토 사용에 미친 영향과 사용자의 토픽에 대한 변화에 어떠한 영향을 미쳤는지 측정하고 분석한다.

IV. 연구 결과

1. LDA 토픽 모델링

본 연구는 토픽의 개수를 결정하는 지표 중 하나로 토픽의 유사성을 측정하는 지표인 Coherence score로 최적의 토픽 개수를 결정하였다. 토픽 개수

를 2부터 10까지로 설정하여 각각 토픽 개수에 대한 Coherence score를 측정하였고, 토픽 개수가 5개일 경우 0.515로 가장 높아 최적의 토픽 개수는 5개로 판단하여 토픽을 추출하였고, 토픽별 키워드는 다음 표 1과 같다

토픽 1은 애플리케이션 로딩과 강제 종료·랙 그리고 계정 접속 문제 등 기능상의 문제임을 알 수 있으며 코드 최적화와 버그 수정으로 안정성을 강화하고, 서버 업그레이드와 최적화로 랙을 줄이며, Facebook 로그인 문제와 계정 접속 버그에 대한 대응책 마련에 관한 내용을 시사하고 있음을 알 수 있다.

토픽2는 보안 관련 우려를 표현하고 있다. 사용자들은 아이들 감시를 경고하며, 해커의 존재와 헤드폰으로 들리는 숨소리, 개인정보 노출 가능성에 대한 우려를 표현하였다. 이에 따라 애플리케이션의 보안 강화를 위한 보안 헤더·주기적인 보안 테스트·사용자 인증 및 권한 관리를 강화하는 대책이 필요하며, 불필요한 데이터 접근을 방지하는 방향으로 보완이 필요함을 제시하고 있다.

토픽3은 메타버스 상의 구매와 관련된 문제를 언급하고 있다. 실제 사용자의 리뷰를 보면 Zem을 이용한 물건 구매 시 Zem이 계속 사라지는 문제·Zem 구매 후 미수령 및 연결 확인 메시지 등 Zem 관련 다양한 문제가 발생하고 있음을 알 수 있다. 사용자들은 이러한 문제로 불편함을 표출하며, 해당 문제들에 대한 신속한 해결을 요구하였다. 특히, 게임 내에서 Zem을 획득하기가 어렵다는 점과 새로운 아이템 가격에 대해 언급되었다. 애플리케이션 제작자는 Zem 관련 문제를 해결하여 사용자들의 만족도 향상에 주력해야 함을 시사하고 있다.

토픽4에서는 애플리케이션 성능과 업데이트 관련 문제에 대한 사용자의 불만이 나타났다. 특히, 사용자들은 다른 애플리케이션 다운로드 유도·로딩 속도 지연·화면 표시 문제와 업데이트로 인해 더 나빠진 경험을 언급하였다. 따라서 서버와 네트워크 인프라 최적화를 통한 성능 개선과 디스플레이 관련 버그 수정·사용자 피드백 반영 등을 통한 업데이트로 성

능 및 화면 개선에 대한 필요성이 있다고 판단된다.

토픽5에서는 개인정보와 관련 우려를 표현하고 있다. 사용자는 랙으로 폰이 꺼져 다시 켜야 했고, 제페토 애플리케이션을 사용할 때 개인정보 감시 및 스톱이 의심에 대한 경험에 대한 언급이 있다. 개인정보 보안에 대한 우려가 있음을 시사하며, 개인정보 보안 강화·애플리케이션의 감시 기능 강화·악성 행위 대응·이용자 교육 등으로 개인정보 보안을 강화할 필요가 있음을 보여준다.

표 1. 토픽별 키워드와 가중치 (Keywords And Weights For Each Topic)

토픽1	토픽2	토픽3	토픽4	토픽5
let (0.027)	hear (0.060)	game (0.073)	game (0.087)	download (0.045)
account (0.024)	download (0.059)	zem (0.037)	play (0.038)	game (0.043)
log (0.015)	game (0.037)	give (0.024)	hate (0.018)	play (0.025)
main (0.011)	breathing (0.022)	money (0.015)	time (0.016)	really (0.021)
play (0.011)	track (0.020)	buy (0.014)	fix (0.015)	phone (0.018)
email (0.009)	hacker (0.019)	play (0.013)	download (0.015)	install (0.017)
fix (0.009)	delete (0.017)	fix (0.013)	load (0.015)	fix (0.014)
give (0.009)	people (0.016)	item (0.013)	update (0.014)	give (0.013)
work (0.009)	watch (0.016)	star (0.012)	open (0.013)	work (0.013)
put (0.008)	breathe (0.014)	like (0.009)	let (0.013)	delete (0.012)

2. Causal Impact 분석

COVID-19 팬데믹 선언일인 “2020년 3월 11일” 기점을 연구의 핵심 시점으로 선택하여 Causal Impact 분석을 진행하였다. 시계열 예측을 위한 모델링은 팬데믹 선언 전 1년 2개월의 데이터를 사용하였고, 예측은 팬데믹 선언 후 1년 8개월로 설정하여 분석을 수행하였다. 분석 결과는 표 2와 그림 3에 정리하였다.

COVID-19 팬데믹에 따른 제페토 애플리케이션

사용의 변화를 분석한 결과, 실제 데이터의 평균이 159.5이고 누적값이 21,373인 반면, 모델은 평균 83.2를 예측하였다. 개입의 결과로 나타난 변화의 크기를 나타내는 절대적 효과는 76.3이고, 상대적 효과는 96.4%임으로 COVID-19 팬데믹 이후의 제페토 사용이 약 96.4% 증가했다고 해석할 수 있다. 이에 대한 P-Value가 0.001로 통계적으로 유의미한 결과로, COVID-19 팬데믹이 제페토 사용 증가에 영향을 미쳤을 가능성이 매우 크다고 볼 수 있다.

COVID-19 팬데믹이 각 토픽에 어떤 영향을 미쳤는지를 분석한 결과, 5개 토픽에 대한 P-Value가 모두 0.001로 통계적으로 유의미하며, COVID-19 팬데믹에 따른 결과가 우연히 발생할 가능성이 낮다고 판단된다. 따라서 COVID-19 팬데믹의 영향을 받아 토픽의 변화가 있을 가능성이 크며, 토픽별 Causal Impact 분석 결과는 다음과 같다.

COVID-19 팬데믹 이후의 제페토 사용에 대한 토픽1의 연구 결과, 실제 데이터의 평균이 198.5이고 누적값이 29,378인 반면, 모델은 평균 119.1로 예측하였다. 절대적 효과는 79.4로 나타났고, 상대적 효과는 67.5%로 COVID-19 팬데믹 이후의 토픽1의 발생이 이전 대비 약 67.5% 증가하였다고 볼 수 있으며, 이에 대한 P-Value가 0.001로 통계적으로 유의미한 것으로 나타났다. 이에 따라 토픽1은 COVID-19 팬데믹의 영향으로 추이가 변화하였다고 볼 수 있다.

토픽2를 분석한 결과, COVID-19 팬데믹 이후 토픽2의 발생이 이전 대비 약 89.9% 증가하였다. 실제 데이터의 평균은 198.5이고 누적값이 29,378인 반면, 모델은 105.6으로 평균 예측하였다. 이에 따른 절대적 효과는 92.9이며, 상대적 효과는 89.9%이다. 토픽2의 P-Value가 0.001로 통계적으로 유의미하며, COVID-19 팬데믹의 영향을 받았음을 의미한다.

COVID-19 팬데믹과 토픽3의 관계를 분석한 결과, 실제 데이터의 평균이 198.5이고 누적값이 29,378인 반면, 모델은 평균 116.1로 예측하였다. 개입 이후의 절대적 효과는 82.4로 나타나며, 상대적

효과는 72.7%이고, 토픽3의 P-Value가 0.001로 통계적으로 유의미하며, COVID-19 팬데믹의 영향을 받아 토픽3의 발생이 이전 대비 72.7% 증가하였음을 알 수 있다. 이로써 COVID-19 팬데믹은 토픽3에 영향을 주었음을 사하고 있음을 알 수 있다.

표 2. Causal Impact 분석 결과 (Result of Causal Impact Analysis)

		Actual	Prediction (s.d.)	Absolute effect (s.d.)	Relative effect (s.d.)	P-Value
전체	Avg.	159.5	83.2 (12.07)	76.3 (12.07)	96.4% (33.1%)	0.001
	Cum.	21,373	11,152.1 (1,616.76)	10,220.9 (1,616.76)	96.4% (33%)	
토픽 1	Avg.	198.5	119.1 (8.54)	79.4 (8.54)	67.5% (12.2%)	0.001
	Cum.	29,378	17,634.1 (1,263.5)	11,743.9 (1,263.5)	67.5% (12%)	
토픽 2	Avg.	198.5	105.6 (11.06)	92.9 (11.06)	89.9% (17.9%)	0.001
	Cum.	29,378	15,624.7 (1,636.47)	13,753.3 (1,636.48)	89.9% (18%)	
토픽 3	Avg.	198.5	116.1 (11.4)	82.4 (11.4)	72.7% (17.8%)	0.001
	Cum.	29,378	17,184 (1,687.67)	12,194 (1,687.67)	72.7% (18%)	
토픽 4	Avg.	198.5	117.4 (9.7)	81.1 (9.7)	70.3% (14.4%)	0.001
	Cum.	29,378	17,373.4 (1,435.58)	12,004.6 (1,435.58)	70.3% (14%)	
토픽 5	Avg.	198.5	101.8 (13.14)	90.5 (13.14)	86.3% (21.7%)	0.001
	Cum.	29,378	15,986 (1,944.31)	13,392 (1,944.31)	86.3% (22%)	

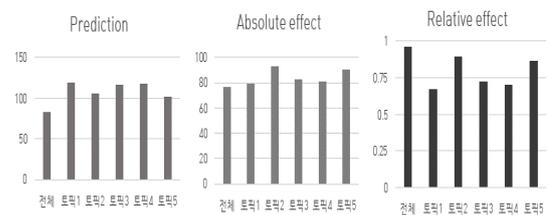


그림 3. Causal Impact 분석 결과 (Result of Causal Impact Analysis)

토픽4에 관한 연구 결과, COVID-19 팬데믹 이후 토픽4의 발생이 이전 대비 약 70.3% 증가하였다. 실제 데이터의 평균이 198.5이고 누적값이 29,378인 반면, 모델은 평균 117.4로 예측하였고, 이에 따른

절대적 효과는 81.8이며, 상대적 효과는 70.3%이다. 토픽4의 P-Value가 0.001로 통계적으로 유의미하며, COVID-19 팬데믹의 영향을 받았음을 알 수 있다.

마지막으로, COVID-19 팬데믹 이후 토픽5의 발생이 이전 대비 약 86.3% 증가하였다. 실제 데이터의 평균은 198.5이고 누적값이 29,378인 반면, 모델은 평균 101.8로 예측하였으며, 이에 따른 절대적 효과는 90.5로 나타나며, 상대적 효과는 86.3%이다. 토픽5의 P-Value가 0.001로 통계적으로 유의미하며, COVID-19 팬데믹의 영향을 받아 토픽5가 변화하였다고 볼 수 있다.

V. 결 론

본 연구는 COVID-19 팬데믹이 메타버스 애플리케이션에 대한 사용자 요구사항에 미치는 영향을 LDA 토픽모델링과 Causal Impact 분석을 통하여 종합적으로 이해하고자 하였다. 이를 위해 메타버스의 대표 애플리케이션인 제페토 리뷰를 활용하였다. 분석 결과, LDA 토픽모델링으로 애플리케이션 기능적 문제·보안 문제·애플리케이션 내 가상화폐(Zem)에 대한 불만 사항·애플리케이션 성능·개인정보 관련으로 부정적인 토픽을 추출하였다. 또한, COVID-19 팬데믹 이후의 제페토 사용이 약 96.4% 증가한 것으로 확인되었으며, COVID-19 팬데믹은 메타버스 애플리케이션 사용에 유의미한 영향을 미쳤으므로 나타났다. COVID-19 팬데믹에 따른 토픽의 변화를 분석한 결과, 5개 토픽 모두 COVID-19 팬데믹의 유의한 영향을 받아 증가하였음을 알 수 있으며, 이 중 보완과 관련된 토픽2가 COVID-19 팬데믹의 영향을 특히 많이 받은 것으로 나타났다. 이에 대한 대응책 마련이 필요함을 시사하고 있다. 본 연구는 제페토 애플리케이션의 사용자 리뷰에 LDA 토픽모델링을 수행하여 토픽을 추출하였다. 메타버스 애플리케이션 사용자의 관점에서 발생한 이슈를 구체적으로

이해함으로써 서비스 개발자와 운영팀에게 서비스 개선 및 대응책 제시하였다. 사용자 피드백을 기반으로 서비스 품질 향상과 사용자 요구사항 충족에 기여하였다. 또한, 본 연구는 Causal Impact 기법을 활용하여 COVID-19 팬데믹과 메타버스 애플리케이션에 대한 영향을 구체적으로 분석한 점에서 의의가 있다.

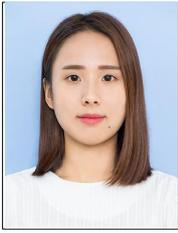
본 연구는 다음과 같은 한계점을 가진다. 첫째, 연구 대상을 "제페토" 애플리케이션에만 국한하여 다양한 메타버스 애플리케이션을 고려하지 않았다. 따라서 결과의 일반화에 한계가 있으므로 향후 연구에서는 다양한 애플리케이션을 고려하여 일반적인 패턴을 파악할 필요가 있다. 둘째, 본 연구는 과거 데이터 기반으로 진행되었으며, "제페토" 리뷰는 현시점이 아닌 과거 시점에서 작성되었다. 따라서 현재 상황 및 새로운 토픽 반응에 한계가 있으며, 향후 다양한 시점을 비교하여 토픽의 변화를 조사하는 연구가 필요하다. 마지막으로 본 연구에서는 Causal Impact 분석 시 COVID-19 팬데믹 선언 시점만 고려하였다. 이로 인해 세부적인 개입에 따른 사용자 토픽의 변화를 놓칠 수 있으며, 향후 연구에서는 더 세분화된 시점을 고려하여 결과를 정확히 이해할 필요가 있다.

REFERENCES

- [1] 최성호, "코로나 19 유행의 방역," *대한내과학회*, 제 95권, 제3호, 134-140쪽, 2020년 6월
- [2] 김태중, 안부영, 이원철, 강혜진, "뉴스 빅데이터 활용 메타버스 트렌드 분석," *Journal of Digital Contents Society*, 제23권, 제2호, 203-216쪽, 2022년 2월
- [3] 서성은, "메타버스 개발동향과 발전전망 연구," *한국 HCI 학회 학술대회*, 1450-1457쪽, 2008년 2월
- [4] Mystakidis, S, "Metaverse. Encyclopedia," vol. 2, no. 1, pp. 486-497, Feb. 2022.
- [5] 이한진, 구현희, "메타버스 플랫폼 사용성 평가체계 구축에 관한 델파이연구 - 로블록스, 제페토, 게더타운 사례를 중심으로," *한국콘텐츠학회논문지*, 제 22권, 제9호, 179-193쪽, 2022년
- [6] 베인앤드컴퍼니, "메타버스 시장은 2030년 최대 9,000억 달러 규모로 성장 가능하나 시간 걸릴 것,"(2

- 023). <https://www.bain.com/ko>(accessed Dec., 03, 2023).
- [7] 정지훈, 정혜인, 이준기, “텍스트마이닝 기법과 ARI MA 모형을 활용한 배달의 민족 앱 리뷰 분석,” *디지털콘텐츠학회논문지*, 제22권, 제2호, 291-299 쪽, 2021년 2월
- [8] Gao, C., Zeng, J., Lyu, M. R., and King, I, “Online app review analysis for identifying emerging issues,” *In Proceedings of the 40th International Conference on Software Engineering*, pp. 48-58, May. 2018.
- [9] Dąbrowski, J., Letier, E., Perini, A., and Susi, A., “Finding and analyzing app reviews related to specific features: A research preview,” *In Proceedings of the REFSQ 2019*, pp. 183-189, Mar. 2019.
- [10] Islam, M. N., Islam, I., Munim, K. M., and Islam, A. N., “A review on the mobile applications developed for COVID-19: an exploratory analysis,” *IEEE Access*, vol. 8, pp. 145601-145610, Aug. 2020.
- [11] 김선주, 박대영, 김병수, “토픽모델링을 이용한 Airbnb 고객의 리뷰 분석: 코로나 시대 전과 후의 토픽 차이를 중심으로,” *인터넷전자상거래연구*, 제21권, 제4호, 115-130쪽, 2021년 8월
- [12] Brodersen, K. H., Gallusser, F., Koehler, J., Remy, N., and Scott, S. L., “Inferring causal impact using Bayesian structural time-series models,” *Ann. Appl. Stat.*, vol. 9, no. 1: pp. 247-274, Mar. 2015.
- [13] Mudambi, S. M., and Schuff, D., “Research note: What makes a helpful online review? A study of customer reviews on Amazon.com,” *MIS Quarterly*, Vol. 34, No. 1, pp. 185-200, Mar. 2010.
- [14] Hong, H., Xu, D., Xu, D., Wang, G. A., and Fan, W., “A meta-analysis on the determinants of online review helpfulness,” *WHICEB 2017*, pp. 319-327, May. 2017.
- [15] Blei, D. M., Ng, A. Y., and Jordan, M. I., “Latent dirichlet allocation,” *Journal of machine Learning research*, vol. 3, pp. 993-1022, Jan. 2003.
- [16] Blei, D. M., “Probabilistic topic models,” *Communications of the ACM*, vol. 55, no. 4, pp. 77-84, 2012.
- [17] Han, X., “Evolution of research topics in LIS between 1996 and 2019: An analysis based on latent Dirichlet allocation topic model,” *Scientometrics*, vol. 125, no. 3, pp. 2561-2595, Oct. 2020.
- [18] 배정환, 한남기, 송민, “토픽 모델링을 이용한 트위터 이슈 트래킹 시스템 Twitter Issue Tracking System by Topic Modeling Techniques,” *지능정보연구*, 제20권, 제2호, 109-122쪽, 2014년
- [19] Bastani, K., Namavari, H., and Shaffer, J., “Latent Dirichlet allocation (LDA) for topic modeling of the CFPB consumer complaints,” *Expert Systems with Applications*, vol. 127, pp. 256-271, 2019.
- [20] 김은정, 장석권, 이상용, “토픽모델링과 시계열 회귀분석을 활용한 헬스케어 분야의 뉴스 빅데이터 분석,” *경영정보학연구*, 제25권, 제3호, 163-177쪽, 2023년
- [21] Martin, W., Sarro, F., and Harman, M., “Causal impact analysis for app releases in google play,” *In Proceedings of the 2016 24th ACM SIGSOFT*, pp. 435-446, Nov. 2016.
- [22] Feroze, N., “Forecasting the patterns of COVID-19 and causal impacts of lockdown in top five affected countries using Bayesian Structural Time Series Models,” *Chaos, solitons & fractals*, vol. 140, pp. 110196, Aug. 2020.
- [23] Navas Thorakkattile, M., Farhin, S., and Khan, A. A., “Forecasting the trends of covid-19 and causal impact of vaccines using bayesian structural time series and arima,” *Annals of Data Science*, vol. 9, no. 5, pp. 1025-1047, Jun. 2022.
- [24] Wallach, H. M., Murray, I., Salakhutdinov, R., and Mimno, D., “Evaluation methods for topic models,” *In Proceedings of the ICML 2009*, pp. 1105-1112, Jun. 2009.
- [25] Newman, D., Lau, J. H., Grieser, K., and Baldwin, T., “Automatic evaluation of topic coherence,” *In Human language technologies: In Proceedings of the NAACL 2010*, pp. 100-108, Jun. 2010.

저자 소개



이소원(학생회원)

2022년 계명대학교 경영정보학과 학사 졸업.

2022년 계명대학교 경영정보학과 석사 재학.

<주관심분야 : 빅데이터분석, 인과분석, 텍스트마이닝>



노미진(정회원)

1999년 대구가톨릭대학교 경영학사 졸업.

2001년 경북대학교 경영학과 경영정보 전공 석사 졸업.

2006년 경북대학교 경영학과 경영정보 전공 박사 졸업.

<주관심분야 : 빅데이터분석, 데이터 시각화>



한무명초(정회원)

2006년 방송통신대학교 컴퓨터과학과 학사 졸업.

2009년 계명대학교 전산교육학과 석사 졸업.

2016년 계명대학교 경영정보학과 박사 졸업.

<주관심분야 : 데이터마이닝, 컴퓨팅사고, 인과추론, 정보기술>



김양석(정회원)

1995년 서울시립대학교 경제학과 학사 졸업.

2004년 University of Tasmania 컴퓨터 공학 석사 졸업.

2009년 University of Tasmania 컴퓨터 공학 박사 졸업.

<주관심분야 : Machine Learning and Data Analytics, Recommender Systems, Knowledge Engineering>