

빅데이터 분석과 네트워크 기반 온라인 소비자 커뮤니티 참여 행동 분석

박 순 화¹ · 이 점 숙¹ · 박 은 경¹ · 김 은 경¹ · 서 영 건^{2*}

¹경상국립대학교 컴퓨터공학과 강사

^{2*}경상국립대학교 컴퓨터공학과 교수

Analysis of Participation Behavior in the Online Consumer Community Using Big Data Analysis and Netnography

Soon-Hwa Park¹ · Jeom-Suk Lee¹ · Eun-Kyoung Park¹ · Eun-Kyoung Kim¹ · Yeong Geon Seo^{2*}

¹Lecturer, Department of Computer Science, Gyeongsang Nat'l University, Jinju 52828, Korea

^{2*}Professor, Department of Computer Science, Gyeongsang Nat'l University, Jinju 52828, Korea

[요 약]

빅데이터 분석은 방대한 데이터에서 빠르게 통계적 패턴과 경향을 파악하기는 쉽지만, 데이터가 가진 심층적 의미와 인과관계를 확인하는 데는 어려움이 있다. 반면 네트워크 기반 분석과 같은 질적 분석은 시간과 노력이 많이 소요되지만, 소비자의 의견과 경험을 직접 파악하여 인과관계를 이해하고 심층적인 통찰력을 제공할 수 있다. 본 연구는 빅데이터 분석 결과의 정확한 해석과 심층적인 통찰력을 제공하기 위해 빅데이터 분석에 나타난 데이터의 패턴과 특이점을 활용하여 질적 분석을 수행하는 BIGNET 통합 분석 모델을 제안하였다. 연구의 타당성과 실용성을 검증하기 위해 국내 대표 천기저귀 온라인 소비자 커뮤니티를 대상으로 실증 연구를 수행하였으며, 연구 결과 제시된 통합 분석 모델이 온라인 소비자 커뮤니티에 나타난 현상의 원인을 파악할 수 있음을 확인하였다. 이는 빅데이터 분석 결과만으로 자칫 잘못 해석하거나 오인할 수 있는 온라인 소비자 행동 중 많은 부분을 효과적으로 설명할 수 있을 것이라 기대된다.

[Abstract]

Big data analysis easily identifies statistical patterns and trends. However, relying solely on big data analysis poses challenges in validating the profound meanings and causal relationships underlying the data. On the other hand, qualitative analysis such as netnography requires substantial time and effort but offers direct insights into consumer opinions and experiences, facilitating a deeper understanding of causality and providing in-depth insights. To provide accurate interpretation and in-depth insight into big data analysis results, this study proposes the BIGNET integrated analysis model that performs qualitative analysis, utilizing the patterns and singularities of the data considered in big data analysis. To verify the validity and practicality of the study, an empirical study was conducted targeting Korea's representative online cloth diaper consumer community, thereby confirming that the proposed integrated analysis model can identify the cause of the phenomena appearing in the online consumer community. This study is expected to effectively explain many aspects of online consumer behavior that can easily be misinterpreted or misunderstood based on big data analysis results alone.

색인어 : 빅데이터 분석, 네트워크, 혼합 방법 연구, 질적 분석, 천기저귀

Keyword : Big Data Analysis, Netnography, Mixed-Method Research, Qualitative Analysis, Cloth Diaper

<http://dx.doi.org/10.9728/dcs.2024.25.5.1325>



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 08 March 2024; **Revised** 25 March 2024

Accepted 09 May 2024

***Corresponding Author; Yeong Geon Seo**

Tel: +82-55-772-1392

E-mail: young@gnu.ac.kr

I. 서론

최근 생성형 AI의 등장과 성공은 기업들에게 AI와 빅데이터 분석에 대한 관심을 더욱 고조시키고 있다. 자연어 처리, 머신러닝, 딥러닝과 같은 AI 기술을 활용한 빅데이터 분석은 기업 경영의 패러다임을 변화시키며, 데이터를 중심으로 한 의사결정과 자동화된 상호작용을 통해 기업은 고객서비스, 마케팅, 경영전략 등을 개선하고 혁신과 경쟁력을 확보하고 있다[1],[2].

빅데이터 분석과 머신러닝은 많은 양의 데이터를 기반으로 객관적이고 빠르게 통계적인 패턴을 찾고 추세나 예측에서 아주 훌륭한 성능을 보이는 유용한 도구이지만, 분석 결과만으로 데이터가 가진 심층적 의미를 완전히 이해하고 활용하는 데는 한계가 있다. 왜냐하면, 통계적으로 유의미한 결과가 나왔지만, 그 이유를 명확히 설명하기는 어렵기 때문이다[3]. 빅데이터 분석 결과를 정확히 해석하여 빠르게 의사결정에 활용하기 위해서는 기계적인 접근뿐만 아니라 인간의 직관과 경험을 결합한 종합적인 접근이 필요하다. 많은 연구자는 사람의 직관과 경험을 기반으로 한 통찰이 정량적인 데이터 분석으로부터 얻어진 발견과 조화를 이룰 때 시너지 효과가 창출된다고 주장한다[4]-[6].

빅데이터 분석의 이런 한계를 보완하기 위해 혼합 방법 연구에서는 맥락에 기반한 질적 분석과의 통합이 활발히 연구되고 있다[7],[8]. 왜냐하면, 우리가 찾고자 하는 의미는 언제나 맥락 속에 존재하며, 사실을 분절적 데이터로 나누면 상황적 의미를 잃고 잘못된 해석을 할 가능성이 크기 때문이다. 특히, 소셜미디어에서 사용되는 텍스트 데이터의 경우, 사용자의 언어적 특성, 문화적 배경, 사회적 상황 등이 텍스트의 의미에 영향을 미친다. 단순히 텍스트 데이터의 빈도나 연관성만을 고려하여 분석한다면, 맥락을 이해하는데 필요한 정보를 놓치고 잘못된 해석을 할 수 있다[3],[9].

온라인에서 소비자 행동을 연구하는 질적 분석 연구자들은 방대한 데이터로 인해 분석할 콘텐츠의 서브셋을 추출하여 질적 분석을 수행하지만, 효과적이고 정확한 질적 분석 대상을 추출할 수 있는 프로세스에 대한 구조는 많지 않은 실정이다[10]. 특히 연구 질문이 명확하지 않았을 때 분석 대상 선정은 더욱 어려움을 겪는다.

본 연구에서는 빅데이터 분석 결과의 패턴과 특이점을 활용하여 심층적 분석이 필요한 연구 질문을 찾고, 선정된 연구 질문에 대한 효과적이고 정확한 질적 분석 대상을 추출하여 네트노그래피(Nethography)를 수행하는 BIGNET(Big Data Analysis & Netnography)통합 분석 모델을 제안한다.

제안된 분석 모델의 실용성과 연구의 타당성을 검증하기 위해 실제 국내 대표 천기저귀 온라인 커뮤니티를 대상으로 사례분석을 수행하여 분석 모델의 우수성과 실효성을 검증하고자 한다.

II. 관련 연구

2-1 빅데이터 분석의 한계

빅데이터 분석은 현대 사회에서 온라인 소비자를 연구하는 아주 중요한 도구로 자리 잡았지만, 예측 분석의 힘과 이를 통해 얻은 통찰력을 실제 비즈니스에 적용하는 데 몇 가지 한계도 가지고 있다.

첫째, 데이터의 품질 검증이다. 빅데이터는 종종 구조화되지 않은 형태로 존재하며, 빠진 정보, 오류, 또는 왜곡된 데이터를 포함할 수 있다. 이는 분석 결과의 정확성과 신뢰성을 저해시키며, 이러한 경우에는 인간의 판단과 경험에 의한 데이터의 정확성, 완전성, 일관성을 평가하고 필요한 전처리 및 정제 작업을 수행해야 한다.

둘째, 빅데이터 분석은 상관관계와 인과관계를 혼동할 수 있다. 두 변수 간의 관련성이 확인된다고 하여, 이것이 두 변수 사이의 직접적인 인과관계를 보장하지 않는다. 따라서 실질적인 인과관계를 확인하기 위해서는 추가적인 연구와 분석이 필요하다.

셋째, 인간의 직관적 판단과 경험이 필요한 경우이다. 빅데이터 분석의 원리와 기술을 활용하는 생성형 AI 또한 아직까지 인간의 판단과 경험을 완전히 대체하지는 못한다. 특히 빅데이터 분석에서 도출된 결과만으로 자칫 잘못 해석하거나 오인의 소지가 있기 때문이다. 따라서 분석 결과의 정확한 해석을 위해서는 데이터가 가진 심층적 의미를 이해하는 것이 중요하다. 통계적으로 유의미한 결과가 나왔더라도 그 결과가 실제로 비즈니스에 어떤 영향을 미칠지 이해하기 위해서는 도메인 지식이 필요하며, 단순히 데이터의 통계적 패턴을 분석하는 것을 넘어 인간의 직관적 판단과 경험이 중요한 요소로 작용하기 때문이다[10].

이렇듯, 빅데이터 분석은 인간의 주관적인 판단을 최소화하고 빠르고 객관적 분석이 가능하다는 큰 장점에도 불구하고 단일 방법론만으로는 정확한 결과 해석에 제한이 있을 수 있다. 이를 보여주는 대표적인 사례 중 하나가 2015년에 중단된 GFT(Google Flu Trends)서비스이다. 초기에는 독감 확산을 예측하는 유용한 도구로 인식되었지만, 모델이 검색어 데이터에만 의존했기 때문에 검색어 사용 패턴이 변하거나 특정 이벤트에 영향을 받을 때 정확성이 떨어지고, 데이터 분석에서 독감 자체에 대한 어떠한 전문성도 필요하지 않았기 때문에[11] 왜, 그 단어가 독감 발생과 관련이 있는지도 설명할 수 없었다. 해당 검색어와 독감 발생이라는 두 대상이 우연히 비슷하게 움직였을 가능성도 내포하고 있었다[12],[13]. 또한, 모델이 특정 기간의 데이터에 과도하게 적응되면서 예측의 불일치가 발생하는 과적합 문제도 있었다. 즉, 모델이 특정 상황에 지나치게 민감해지면서 일반성이 떨어지게 되었기 때문이다. 이를 계기로 혼합 연구에 대한 필요성이 더욱 부각 되었고, 전염병 감시 및 예측 연구에서는 빅

데이터 분석과 AI는 물론, 더 많은 차원에서의 이해를 위해 다양한 연구 방법론을 혼합하여 예측하는 경향이 나타나고 있다.

2-2 온라인 소비자 커뮤니티와 네트워크 그래피

소셜미디어와 온라인 소비자 커뮤니티는 수많은 소비자가 게시글과 댓글을 통해 다양한 의견과 정보를 실시간으로 공유하는 플랫폼이다. 이를 분석하면 소비자 행동의 패턴과 의미, 그리고 그들의 심층적인 욕구를 파악할 수 있다[14]. 이는 기업이나 조직이 제품 또는 서비스를 빠르게 개선하고 발전시키는 데 큰 도움이 될 수 있다. 이런 온라인 소비자 커뮤니티의 소비자 행동을 분석하는 최적의 연구 방법이 네트워크 그래피라고 볼 수 있다[15].

네트워크 그래피는 Kozinets에 의해 개발된 연구 방법으로, Internet(Net)과 Ethnography의 합성어이며 기존 에스노그래피에 디지털 시대의 특징을 반영하여 전통적인 에스노그래피를 확장한 질적 연구 방법이다. 에스노그래피가 문화나 사회 집단 내에서 참여자로서 직접 참여하고 경험을 바탕으로 조사하는 방법이었다면, 이를 온라인 커뮤니티나 소셜미디어, 포럼, 블로그와 같은 디지털 플랫폼에 적용한 것이 네트워크 그래피이다. 네트워크 그래피(Netnographer)는 온라인 커뮤니티에 참여 관찰을 통해 상호작용하며 그들의 행동, 대화, 문화적 특징들을 이해한다. 최근에는 데이터 수집 범위가 확장되어, 특정 주제에 대한 폐쇄된 온라인 커뮤니티뿐만 아니라 다양한 소셜미디어에서 다양한 형태의 자료를 수집, 분석하여 특정 주제나 문화적 현상에 대한 인사이트를 얻고 있다. 이런 흐름을 반영하여 네트워크 그래피의 최근 정의는 “다양한 소셜 미디어에서 나타난 문화적 경험을 질적으로 이해하려는 연구 방법”으로 정의되고 있다[16].

전통적 에스노그래피가 연구자의 간섭이나 편향 오류가 발생할 수 있어, 분석 결과의 신뢰성에 영향을 미치지만, 네트워크 그래피는 연구자의 개입 없이 소비자가 자발적으로 제공한 데이터를 분석하기 때문에 고객의 욕구와 의미를 왜곡 없이 파악할 수 있다[14],[17]. 또한, 시공간적 제약을 받지 않으며, 분석 시간과 비용 측면에서도 더욱 효율적인 연구 방법이다. 그러나 이런 장점에도 불구하고 온라인상의 방대한 데이터, 즉 빅데이터를 대상으로 사람이 일일이 질적 분석을 수행하는 것은 현실적으로 많은 어려움이 있다.

III. 연구 모델 설계

본 연구는 선행된 통합 분석 모델[9]의 후속 연구로서 확장과 검증 측면을 강화하여 새로운 BIGNET 모델을 그림 1과 같이 설계하고 실증 연구를 수행하였다.

3-1 데이터 수집

연구 대상 소셜미디어를 선정하고 데이터를 수집하는 과정은 철저한 사전조사를 통해 이루어지며, 이 과정은 네트워크 그래피 수행 절차의 시작이기도 하다. 본 연구에서는 사전 조사를 통해 육아 관련 온라인 소비자 커뮤니티인 네이버 카페 ‘천기저귀 홀릭맘(천홀)’을 연구 대상으로 선정하였다. 대상 선정 기준은 데이터 정제, 내용 타당성, 분석 대상의 적절성과 같은 네트워크 그래피 연구 적합성[17]을 판단하여 선정하였다.

네트워크 그래피 접근법에 따라 네트워크 그래피는 연구대상 카페를 일정 기간 지속적 온라인 참여 관찰을 진행한 후 커뮤니티에서 사용되는 언어, 형식, 그리고 그들만이 의식할 수 있는 의미와 문화를 이해하고, 게시판과 게시글의 특성을 파악하여

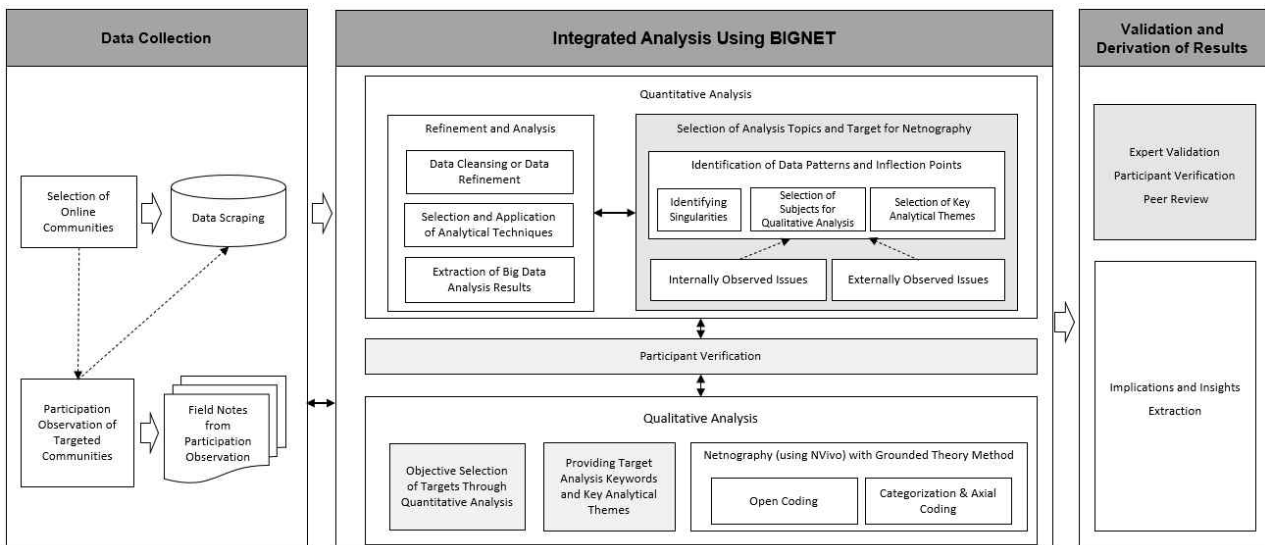


그림 1. BIGNET 통합 분석 모델

Fig. 1. BIGNET integrated analysis model

필드 노트를 작성한다. 온라인 소비자 커뮤니티의 소비자 참여 행동은 주로 본문과 댓글로 나눌 수 있는데, 본문과 댓글의 작성 행동은 관여 정도와 적극성에서 차이가 있다. 따라서 본 연구에서는 게시글과 댓글의 조회 수를 수집하였으며, 파이썬 웹크롤링을 통해 총 99,951건의 게시글을 수집, 분석하였다. 연구를 시작한 1차 수집 기간[2011. 11. 29.~2018. 7. 7]의 데이터 총 99,448건과 1차 분석 결과를 바탕으로 심층 분석의 결과 확인과 검증을 위해 다시 2차 수집 기간[2022.1.1~2023.7.31]의 503건의 데이터를 수집, 분석하였다. 2차 분석 대상의 수집 기간과 대상 데이터는 네트노그래피와 논의를 통해 이론적 포화 상태의 표본 크기로 정하였다.

3-2 통합 분석 모델 : BIGNET

1) 데이터 전처리

BIGNET 통합 분석 과정은 우선 수집된 데이터를 대상으로 정확성, 완전성, 일관성, 중복 여부 등을 확인하고, 이상치나 오류를 수정하여 데이터의 품질을 향상시키는 전처리 과정이 수행된다. 데이터 클리닝을 거쳐 데이터의 신뢰성을 높이고 분석에 적합한 상태로 데이터를 변화하는 과정이며, 본 논문에서는 빅데이터 결과의 신뢰성을 높이기 위해 전처리 과정에 연구자, 빅데이터 분석가, 네트노그래피의 삼각검증법을 사용하였다[5]. 전처리 과정에서는 맥락적으로 의미 있는 단어를 하나의 의미 단어로 묶는다든지, 무의미한 단어를 정제하는 과정에서 연구자의 주관적 판단이 불가피하게 개입된다. 특히 축약어와 다양한 신조어가 만들어지는 요즘의 인터넷 문화에서는 문화를 정확히 이해하는 전문가의 검증 과정은 필요하다. 종종 대량의 데이터를 정제하는 과정에서는 데이터에 대한 선지식이 높지 않은 경우, 의도치 않게 중요한 데이터들을 놓칠 우려가 있기 때문이다. 따라서, 수집된 데이터의 신뢰성과 타당성을 높이기 위한 삼각검증법은 온라인 소비자 행동 연구에서 유용한 시도가 될 것이다[18],[19].

2) 빅 데이터 분석과 네트노그래피 대상 지점 선정

BIGNET 모델은 빅데이터 분석 결과에 나타나는 데이터의 패턴을 통해 특이점이나 결과 해석이 필요한 키워드와 연구 주제를 빠르게 선정하여 심층적 분석의 병행이 가능하다. 전통적 질적 분석은 시간과 노력이 많이 드는 분석 방법이며 연구자의 주관적인 해석과 판단에 의존하는 연구 방법이기 때문에 연구자의 편향이 결과에 큰 영향을 미칠 수 있다. 그러나 본 연구에서는 빅데이터 분석 결과를 바탕으로 질적 연구 대상 데이터를 선정하고 심층 분석 주제를 설정하여 질적 연구를 진행하기 때문에 전통적 질적 분석 방법의 한계를 어느 정도 보완한 연구 방법이다[8]. 카페 내의 전체 게시글의 변동 추이를 확인하는 것은 거시적 관점의 흐름을 파악할 수는 있지만, 연구 대상이나 주제가 세분화되지 않으면 우리가 찾고자 하는 인사이트는 희석될 수 있다. 1차 자료 수집 기간 당시 천홀 카페는 회원들의 주요 테마와 관심사가 상이하게 다

른 점을 감안하여 총 45개의 세부 게시판이 운영되고 있었으며, 네트노그래피와 상호 논의 과정을 거쳐 소비자 행동에 대한 가장 많은 맥락을 내포하고 있으면서도 새로 유입되는 회원들의 상황을 자세히 파악할 수 있는 가입 인사 게시판을 대상으로 분석을 실시하여 게시글의 변동 추이를 확인하였다. 분석 결과에서 변동 추이 그래프에 이상 패턴을 보이는 지점과 이상 패턴에 영향을 미칠 수 있는 내외적 이슈를 종합 판단하여 분석이 필요한 정확한 지점과 범위를 선정하였다. 선행된 논문[9]의 통합 분석 모델은 빅데이터 분석 결과의 특이점을 이용하여 분석 대상을 선정하는 모델을 제안하였지만, 변동 추이 패턴에서 특이점 선정이 어려운 경우에는 분석 대상 선정에 어려움이 있었다. 이를 개선하여 본 연구에서는 변동추이 그래프에서 특이점 선정이 어려운 경우에도 다양한 분석 결과를 바탕으로 관심 주제와 핵심 키워드를 연구자가 지정할 수 있게 하여, 심층 분석 대상 데이터를 선정할 수 있도록 BIGNET 모델을 수정 보완하였다.

3) 분석 결과 검증과 심층 분석 연구 주제 선정

선정된 네트노그래피 대상 지점에 나타난 키워드 변동 추이를 통해 확인이 필요한 키워드와 심층 분석 연구 주제를 선정한다. 이것은 키워드 변동 추이뿐만 아니라 다양한 빅데이터 분석 결과의 이유나 해석이 어려운 주제가 심층 분석 주제로 선정될 수 있으며, 이 과정에서 빅데이터 분석 결과의 오류나 이상을 참여자 검증을 통해 확인할 수 있다.

4) 심층 분석 대상 선정과 네트노그래피

네트노그래피는 연구자가 문장을 읽고 일일이 코드를 부여하는 고전적인 연구 방식이지만 정확한 문맥과 빅데이터 분석에서 알 수 없었던 여러 이유를 직접 설명할 수 있게 해준다. 방대한 데이터인 경우에는 사람이 직접 분석 하는 네트노그래피는 시간적, 비용적 측면에서 현실적인 어려움을 가지고 있다. 따라서 본 연구에서는 최적의 네트노그래피 분석 대상을 빠르고 효율적으로 선정하기 위해 빅데이터 분석의 데이터 패턴과 특이점을 이용하여 네트노그래피 대상 지점을 선정하였으며, 네트노그래피의 숙련도에 따라 이론적 포화를 이룰 수 있는 심층 분석 대상의 범위나 대상을 선정한다. 연구의 방법과 목적에 따라 다양한 질적 연구 방법들이 시도될 수 있으며, 본 연구에서는 근거이론 접근법에 따라 개방 코딩 → 축 코딩 → 선택 코딩 과정을 지속적 수정 반복하며 현상에 따른 이론을 생성하였다.

3-3 결과 도출 및 검증

BIGNET 통합 분석 과정을 통해 도출된 결과는 참여자와 전문가 검증을 거친다. 연구 참여자들에게는 분석 결과를 제시하고 그들의 의견이나 해석을 확인하는 검토 과정을 통해 참여자 검증을 실시하며, 신뢰성 향상을 위해 추가적으로 공신력 있는 외부 전문가 피어 리뷰를 실시하여 연구의 외부 신

피도를 향상시킬 수 있도록 설계하였다.

이를 통해 BIGNET 통합 분석 모델이 다양한 연구 방법들의 통합과 세부적인 검증 과정을 거쳐 빅데이터가 가지고 있는 숨은 가치의 연결 고리들을 확인하고 온라인 소비자 참여 행동에 관한 시사점과 인사이트 도출이 가능함을 분석 결과를 통해 확인할 수 있었다.

IV. 분석 과정 및 결과

4-1 이상 패턴 확인과 심층 분석 대상 지점 선정

1차 수집된 자료를 바탕으로 전체 게시글의 변동 추이 그래프를 우선 확인하였다. 그림 2는 가입 인사 게시글의 연도별 변동 추이 그래프와 출생아 수 변동 추이 그래프이다. 가입 인사 게시글 수는 출생아 수에 영향을 받을 수 있으므로 이상 징후 확인을 위해 인구 통계 자료와 대조하여 변동 추이를 확인하였다. 시계열 분석에서 출생아 수는 급격한 감소 추세를 보이지만 게시글 수는 2017년 급증한 이상 패턴을 보인다. 그림 3은 가입 인사 게시글의 월별 변동 추이 그래프이며 급격한 변화를 보이는 3곳(동그라미로 표시)의 정확한 일별 변곡점을 확인하였다. 특이점은 일반적인 값을 벗어나 이상치를 보이는 지점으로 변곡점의 꼭짓점이 아니라 변곡점의 시작 지점이 특이점이다. 이렇게 이상 징후의 원인 확인이 필요한 후보 지점을 2017.1.31, 2017.8.23, 2017.10.17로 1차 선정하였다.

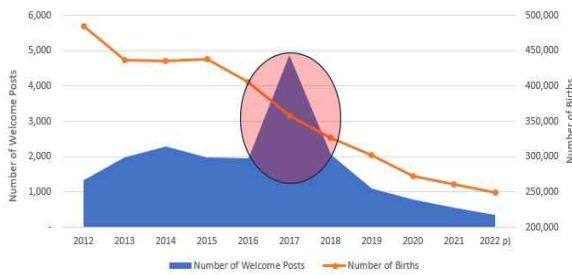


그림 2. 가입 인사 게시글 수와 출생아 수 변동 추이
Fig. 2. Relationship between the number of welcome posts and the trend in birth rates

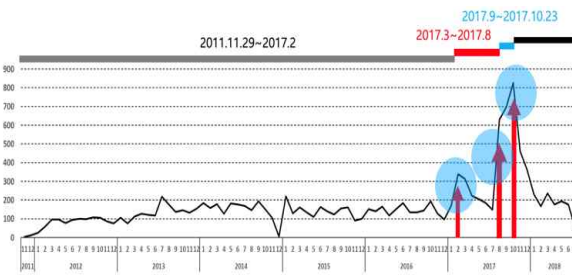


그림 3. 가입 인사 게시글의 월별 변동 추이 그래프
Fig. 3. Graph of monthly changes in greetings posts

표 1. 네트노그래퍼의 참여 관찰 일지에서 확인된 이슈 지점
Table 1. Issue points identified in the Netnographer's observation field notes

Date	Description
2017.1.30	Rapid increase in the number of subscribers (Pa**s diaper issue)
2017.8.23	The bulletin board of the use of cotton sanitary pad was activated, and public opinion on cotton sanitary pad was actively formed.
2017.10.16	Explosive increase in subscribers at cafes after TV broadcast on carcinogenic diapers
2017.10.23	Change of internal regulations: Must attend 10 times and write 2 greetings (500 characters or more) to confirm that you are a cloth diaper user

1차 선정된 지점의 정확성과 타당성 확인을 위해 빅데이터 분석 결과에 영향을 미칠 수 있는 내외적 이슈를 함께 확인하였다. 우선 내적 이슈는 카페에서 발생한 여러 이슈 중 표 1과 같이 4개의 큰 내적 이슈와 날짜를 확인하였다. 참여 관찰 일지는 카페의 내부동향과 이슈들을 자세히 분석한 일지이다. 다음으로 특이점 구간에 영향을 미칠 수 있는 사회적 여론과 같은 외적 이슈 확인을 위해, 본 논문에서는 네*버 데이터 랩의 키워드 검색을 시행하였다. 분석 대상 기간은 이상 구간으로 나타난 2017년 1년간의 데이터이며, 분석 키워드는 기저귀, 천기저귀, 생리대이다. 키워드 분석 결과 2017년 당시 천기저귀에 대한 가장 강한 사회적 여론을 형성한 외적 이슈는 생리대 이슈(2017.8.21.)였으며, 발암물질 기저귀 이슈(2017.10.16.), 팜*스 기저귀 이슈(2017.1.30.) 순으로 나타났다.

표 1에 나타난 네트노그래퍼의 참여 관찰 일지와 날짜를 대조해 보면 이런 강한 사회적 외적 이슈는 카페의 즉각적 여론을 형성하고, 카페의 구조적 변화를 일으킬 수 있음을 확인하였다. 실제로 생리대 이슈 후 카페 내에 면생리대 관련 게시글이 새롭게 추가되었다. 대부분은 이런 카페의 구조적 변화와 원인은 빅데이터 분석만으로는 감지하기 힘든 부분이다. 이렇게 선택된 1차 후보 지점과 내외적 이슈를 종합 판단하여 현미경적 심층 분석이 필요한 대상 지점 A, B, C를 각각 2017.1.30, 2017.8.21, 2017.10.23로 선정하였다.

4-2 대상 지점의 키워드 변동 추이 확인

패턴의 이상을 보인 특이점 A, B, C 지점의 전후 키워드 변동 추이를 확인하여 심층 분석 연구 질문을 선정하였다. 대조군으로 2016~2018년의 가입 인사 게시글 8,106건에 대한 워드클라우드 분석 결과(그림 4)와 A, B, C 전후 한 달 간의 워드클라우드 변동 추이를 살펴보았다. 비교 분석 결과 심층적 이해나 설명이 필요한 키워드로 땅콩, 공부, 생리대, 둘째, 첫째, 면생리대, 건조기, 아들이라는 키워드를 선정하였다. 키워드 분석 결과에서 둘째가 첫째보다, 아들이 딸보다 더 높은 키워드 빈도를 보이지만, 왜 이런 결과가 나타나는지에



*We don't translate into English for the accurate conveyance of the word's meaning

그림 4. 워드클라우드 분석 결과(2016.1.1 ~ 2018.8.1)

Fig. 4. Word cloud analysis results



*We don't translate into English for the accurate conveyance of the word's meaning

그림 5. B 지점의 전후 한 달간의 워드클라우드 분석 결과

Fig. 5. Word cloud analysis results for the one month period before and after 2017.8.21.

대한 정확한 이유를 빅데이터 분석만으로 는 정확히 설명하기는 어려우므로 심층 분석 연구 질문으로 선정하였다.

그림 5는 생리대 이슈가 있었던 특이점 B 전후 각 한 달간의 구간별 키워드 분석 결과이다. 이슈 전에는 거의 등장하지 않던 “생리대”라는 키워드가 이슈 이후 급격한 변화를 보임을 확인할 수 있다. 이처럼 특이점 선택이 정확해지면 키워드의 변동 추이를 정확하게 확인할 수 있다.

4-3 빅데이터 분석 결과 검증과 연구 질문 선정

도출된 분석 결과에서 심층 분석이 필요한 키워드에 대한 참여자 검토 과정을 통해 각 키워드가 가지는 의미와 이유를 확인하였다. 해당 카페 내에서는 땅콩 모양의 천기저귀를 “땅콩”으로 대부분 지칭하고 있으며, 천기저귀가 다양해지면서 카페내 사용 용어나 천기저귀 사용법에 대해 “공부”를 많이 언급하는 것으로 나타났다. “건조기” 키워드는 세탁에 가장 어려움이 나타나는 천기저귀가 팬티형 천기저귀 모양인 일체형으로 바뀌면서 그 두께가 두꺼워져 세탁뿐만 아니라 건조에 대한 어려움이 있었다. 그런 이유로 현재는 건조기가 천기저귀 관련 육아 필수로 인식되는 현상이 카페 내에서 나타나고 있음을 확인하였다. 그러나 여전히 빅데이터 분석 결과만으로 둘째가 첫째보다 높은 비율로 언급되는 이유나 임신과 출산을 거치며 생리대 사용이 줄어드는 데 반해 엄마의 면생리대 키워드가 높은 이유, 그리고 아들이 딸이나 여아보다 더

많은 키워드 빈도를 보이는 이유에 대해서는 상호논의를 통해 심층적 분석이 필요하다고 판단하였다. 인구 통계적으로 아들의 비율이 딸의 비율보다 아주 약간 높긴 하지만 거의 비슷한 비율이기 때문이다. 오히려 네트노그래피의 검토 의견은 천기저귀 사용은 남아보다 여아 언급이 더 높다는 상반된 의견이 제기되었다.

빅데이터 분석 결과의 이상을 확인하고 자료 재분석을 수행하였으며, 이 과정에서 수집된 자료의 전처리 과정에서 한 글자 단어의 불용 처리 과정의 오류가 있음을 확인하였다. 전처리 과정에서는 맥락적으로 의미 있는 단어를 하나의 의미 단어로 묶고, 무의미한 단어를 정제하는 과정에서 연구자의 주관적 판단이 불가피하게 개입된다. 대량의 데이터를 정제하는 과정에서는 데이터에 대한 선지식이 높지 않은 경우나 의도치 않게 일괄 처리 과정에서 이런 오류가 발생할 가능성도 있다. 대다수의 무의미한 단어(예: ‘이’, ‘저’, ‘나’, ‘것’ 등)가 포함되는 한 글자 단어 중, 중요한 한 글자 단어인 “딸”이라는 의미 있는 데이터를 놓치면서 잘못된 빅데이터 분석 결과를 초래하였다. 이 사례를 통해 빅데이터 결과의 신뢰성을 높이기 위해서는 분석의 재료가 되는 데이터 수집과 정제 과정은 더 엄격하고 세심하며, 전문가의 검증이 필요하다는 것을 알 수 있다. 이와 같이 제안 통합 분석 모델은 삼각 검증을 통해 빅데이터 분석 결과의 이상을 감지하고 검토과정을 거쳐 결과의 정확성과 신뢰성을 강화할 수 있도록 설계하였다.

재분석된 가입 인사 게시글에 나타난 성별 언급 관련 키워드 빈도 분석 결과는 남아(7.3%)와 여아(7.2%)로 거의 비슷하게 나타나지만, 딸(60.2%)과 아들(25.3%)을 칭하는 키워드 빈도에서는 큰 차이를 보인다. 도출된 결과만 보면 여아 비율이 훨씬 더 높은 것으로 해석하거나 착각할 수 있다. 그리고 여전히 딸이라는 키워드가 아들보다 훨씬 더 높은 비율로 언급되는 성별 언급 불균형의 이유를 분석 결과만으로는 명확히 밝힐 수는 없었다. 따라서, 빅데이터 분석 결과를 바탕으로 심층 분석이 필요한 연구 질문을 다음과 같이 선정하였다.

- (1) 왜 딸이 아들보다 높은 언급 비율을 보이는가?
- (2) 첫째보다 둘째 언급이 높은 이유는 무엇인가?
- (3) 면생리대 사용 경험이 천기저귀 사용 결정과 어떤 관계를 가지는가?

4-4 네트노그래피 분석 대상 선정과 분석 결과

심층 분석을 위해 가입 인사 게시글의 특이점 A,B,C 를 중심으로 전후 각 1달간의 데이터 총 648건을 대상으로 선정하였다. 대상 데이터의 기간이나 양은 네트노그래피의 숙련도에 의존적이며 이론적 포화를 이룰 수 있는 범위나 양으로 선정하였다. 사람의 경험과 직관에 의존적인 네트노그래피는 양적 연구와 달리 주관적인 견해가 연구에 포함되는 한계를 가지고 있다. 코딩의 객관성을 유지하기 위하여 연구자와 다른 분석자 2인이 각기 코딩한 결과물을 상호 비교하여 코더간 일치율을 검증하였으며, 범주를 확정된 후에는 카페 운영진에게 결

표 2. 특이점 전후에 나타난 코드 비율 변동 표
Table 2. Code ratio fluctuation table before and after singularity

Coding	A1	A2
Community membership reason	9 29.03%	34 40.00%
Current situation	10 32.26%	28 32.94%
Decision to use cloth diapers	6 19.35%	14 16.47%
Reactions from surroundings	2 6.45%	4 4.71%
Purchased cloth diapers	2 6.45%	3 3.53%
Impressions of cloth diapers	1 3.23%	2 2.35%
Drawbacks of cloth diapers	1 3.23%	0 0.00%
Coding	B1	B2
Community membership reason	15 45.45%	55 37.16%
Current situation	7 21.21%	43 29.05%
Decision to use cloth diapers	3 9.09%	24 16.22%
Date of Joining	0 0.00%	15 10.14%
Impressions of disposable diapers	0 0.00%	6 4.05%
Reactions from surroundings	2 6.06%	3 2.03%
Impressions of cloth diapers	4 12.12%	2 1.35%
Considerations regarding cloth diapers	2 6.06%	0 0.00%
Coding	C1	C2
Current situation	203 22.99%	128 21.12%
Decision to use cloth diapers	194 21.97%	89 14.69%
In Use or planned use of cloth diapers	51 5.78%	81 13.37%
Impressions of cloth diapers	95 10.76%	76 12.54%
Impressions of the community	72 8.15%	59 9.74%
Community membership reason	111 12.57%	58 9.57%
Reactions from surroundings	48 5.44%	25 4.13%
Drawbacks of cloth diapers	26 2.94%	24 3.96%
Impressions of disposable diapers	40 4.53%	19 3.14%
Reasons for purchasing cloth diaper	7 0.79%	18 2.97%
Current usage status of cloth diapers	16 1.81%	18 2.97%
Concerns about using cloth diapers	8 0.91%	4 0.66%
Advantages of cloth diapers	0 0.00%	4 0.66%
Reasons for diaper failure	12 1.36%	3 0.50%

과물을 알리고 승인을 받아 상호 주관성을 확보하였다.

표 2는 분석의 상위 범주화 코드와 그 비율 변동 표이다. A 지점 전(A1), 후(B1)의 코드 변동 비율은 카페 가입 이유, B 지점에서는 자신의 현재 상태와 천기저귀 결심 이유, 마지막 카페 규정 강화 이슈가 있었던 C 지점 전후로는 사용 중이거나 사용 예정인 천기저귀, 천기저귀 구매 이유, 천기저귀 사용의 걱정 등 소비자의 천기저귀와 관련된 더 자세하고 많은 정보가 언급되는 것으로 나타났다. 이를 통해 생리대 사건, 발암 물질 기저귀 파동과 같은 강한 사회적 여론은 카페 내의 여론을 형성하고, 천기저귀 사용 결심 이유를 예전보다 더 자세하고 비중 있게 언급하는 것을 알 수 있다.

그림 6은 가입자의 현재 상황에 대한 네트워크 분석 결과이다. 가입자들은 임신 중이거나 아이가 돌 전 아기 엄마의 비율이 높으며, 남아보다는 여아 언급이 높고, 첫째보다는 둘째 언급 비율이 높게 나타나고 있다. 빅데이터 분석과 네트워크 분석에서 아이의 성별 키워드 빈도 분석 결과는 여아 비율이 높은 것으로 나타났다. 그러나 심층 분석 과정에서 차원을 높여 아기(아이, 아가, 애기)를 지칭하는 키워드 빈도 분석 결과와 함께 다시 살펴보면 그 해석은 달라질 수 있다. 왜냐하면, 가입자들의 현재 상황 분석에서 임신 중이라고 응

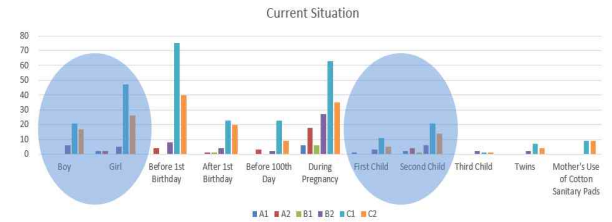


그림 6. 가입자의 현재 상황
Fig. 6. The current state of subscribers

표 3. 성별 언급 코드 비율
Table 3. Analysis code ratio for mentioning gender

Code	Keyword	All(2011~2018)	2016	2017
Girl	daughter(딸)	1842	205	707
	girl(여아)	223	21	102
Girl ratio		15.1%	15.0%	15.1%
Boy	son(아들)	775	86	301
	boy(남아)	219	21	104
Boy ratio		7.3%	7.1%	7.6%
Baby	baby(아기)	4124	479	1713
	baby(아이)	2844	318	1151
	baby(아가)	2838	297	997
	baby(애기)	768	82	275
Unknown gender ratio		77.6%	77.9%	77.3%

*We don't translate into English for the accurate conveyance of the word's meaning

답한 비율이 높고, 심층 분석 결과 임신 중인 여성은 아이를 지칭할 때 딸, 아들보다는 태명이나 아기(아이, 아가, 애기)라는 중성적인 단어를 더 많이 사용하는 것으로 나타났다.

표 3은 네트워크 분석 코드 범주화 과정에 나타난 성별 언급 비율이다. 이때 성별을 알 수 없는 중성적인 키워드인 아기로 현재 상황을 나타낸 비율이 77.6%나 된다. 그러나 수치적인 차이는 줄었지만, 여전히 여아(딸 포함) 언급 비율(15%)이 남아(아들 포함) 언급 비율(7%)보다 2배 정도 더 높게 나타나며 그 이유의 증거는 인용구의 예시에서 찾을 수 있다.

“마침 또 딸아이라 어릴 때부터 최대한 환경호르몬으로부터 멀리해 주고 싶어서 천기저귀 도전하고 싶다는 마음이 더해진 것 같습니다”
 “환경도 환경이고, 제가 피부가 약해서 면생리대를 쓰다 보니 좋은 점이 한둘이 아니어서, 내 딸에게도 천기저귀를 써야겠다 마음먹었어요.”
 “배 속에 아이가 딸이라네요^^ 아무래도 제가 여성 질환경으로 힘들었고, 생각해 보면 모계 쪽이 모두 임신이 쉽지가 않았기에 천기저귀를 사용하면 아기한테 더 좋지 않을까 생각이 듭니다.”

인용구의 예시에서 확인되듯이 가입 인사 게시글에서 여아를 더 많이 언급하는 이유는 경험적으로 엄마는 여아의 상황을 더 잘 이해한다. 그리고 생리대 이슈 등으로 환경호르몬에 대한 우려가 더 커졌기 때문임을 알 수 있다. 또한, 외부적인 요인으로는 최근 우리 사회에 나타나고 있는 여아 선호 현상을 들 수 있다. 한국 리서치의 연구 결과에 의하면 딸 선호 현

상이 특정 세대나 성별의 현상이 아닌 보편적인 정서라는 연구 통계를 확인할 수 있다[20]. 분석 결과와 네트워크, 인구 통계, 외부 연구 결과를 종합해, 첫 번째로 “천기저귀 사용 결심 이유로 왜 여아 언급 비율이 남아 언급 비율보다 높은가?”에 대한 결과를 도출하였다.

가입 인사 게시글의 50% 정도가 임신 중이며 남아나 여아와 같이 정확한 성별을 언급하지 않은 비율이 무려 전체의 약 77%인 것을 종합해보면 빅데이터 분석 결과만으로 여아이기 때문에 더 많이 천기저귀를 선호한다고 보기에는 일반화에 무리가 있을 수 있다. 다만 여아일 경우 가입 인사 게시글에서 더 적극적으로 딸임을 알리는 현상이 나타나고 있으며, 요인중의 하나는 여아 선호 현상이 강해진 사회적 영향일 수 있다.

연구 명제 1:

“가입 인사 게시글의 여아 언급 비율이 더 높게 나타나는 현상은 여아라서 더 높은 가입률을 가진다고 단순히 해석하기보다는, 사회적으로 나타나는 여아 선호 현상으로 여아일 경우 게시글에 더 적극적으로 성별을 언급하는 경향을 보일 수 있다”

두 번째로 빅데이터 분석 결과에서 자녀의 호칭이 첫째보다 둘째가 더 많이 언급되는 이유를 분석하였다. 출생 순서와 관련된 호칭은 다자녀를 둔 엄마에게서 주로 언급되며, 외동이를 키우는 엄마의 경우 순서 호칭은 거의 사용하지 않았다. 따라서 첫째라는 호칭은 둘째가 있는 경우에만 언급되기 때문에 천기저귀 사용자인 어린 둘째가 더 많이 언급되는 것을 확인하였다. 이렇듯 콘텐츠의 맥락과 상황을 자세히 들여다보지 않고 키워드 분석만으로 판단한다면 오해의 소지가 있을 수 있다.

다음은 세 번째 심층 분석 결과이다. 1차 자료 수집 기간에 나타난 분석 결과에서는 면생리대 사용 언급 코드 비율은 그림 7과 같다. 전체 658건 중 3.4%에 해당하는 22건의 면생리대 사용 경험에 대한 언급이 있었다. 그러나, 이론을 형성하기에는 분석을 위한 자료의 양이 적고, 특히 1차 자료 분석 기간인 2017년은 생리대 이슈가 강한 사회적 여론을 형성하고 있었던 시기이기 때문에, 최근의 데이터 503건의 가입 인사 게시글을 추가 수집하여 그 흐름을 살폈다. 2차 수집 데이터의 빅데이터 분석에서 1차 분석과 비교하면 환경, 면생리대, 남편이라는 키워드가 크게 상승한 흐름을 보였다. 1차 분석에서는 대상 선정에 변동 추이의 특이점을 찾고 대상을 선정하는 방법을 사용하였다. 이 방법은 대상 온라인 소비자 커뮤니

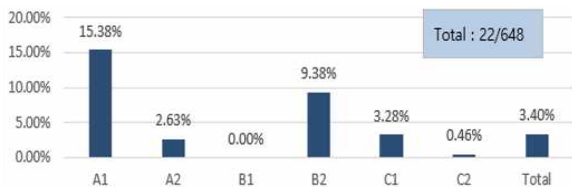


그림 7. 1차 수집기간에 나타난 엄마의 면생리대 사용 언급 비율
Fig. 7. The ratio of mother's cotton sanitary pad usage mentioned during the 1st collection period

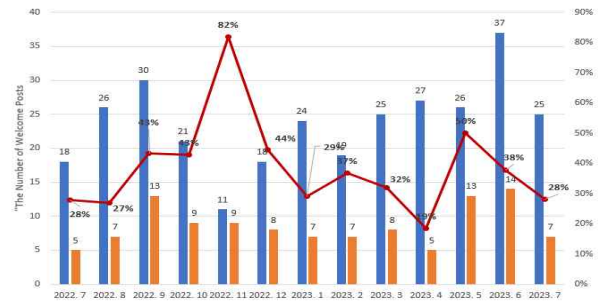


그림 8. 2차 수집 기간에 나타난 엄마의 면생리대 사용 언급 비율
Fig. 8. The ratio of mother's cotton sanitary pad usage mentioned during the 2nd collection period

터를 처음 분석하는 경우에는 무엇을 살펴보아야 할지 정확한 연구 주제나 질문이 없는 경우 유용한 방법이다. 그러나, 2차 수집된 데이터는 면생리대 사용 경험이라는 연구 질문이 명확하고 중심 키워드가 존재하기 때문에 수집된 데이터에서 키워드를 내포하는 대상 게시글만을 빠르게 선정할 수 있다.

2차 수집된 가입 인사 게시글에 나타난 면생리대 사용 경험 언급 비율은 그림 8과 같다. 면생리대 사용 경험 언급 비율은 수집기간 평균 1차 3.4%에서 2차 35.8%로 크게 상승하였다. 이것은 2차 수집된 총 가입 인사 게시글 517건 중 174건의 게시글에 자신이 면생리대 사용자이거나 면생리대 사용 경험이 있다는 것을 언급한 수치이며, 2017년에 사회적으로 큰 여론을 형성했던 생리대 이슈는 현재 천기저귀 카페 가입자 3명 중 1명이 면생리대 사용 경험이 있거나 사용자임을 언급하는 것으로 나타났다. 174건의 분석 결과, 엄마의 면생리대 사용 결정 요인은 생리통이나 여성 질환과 같은 건강 관련 이유를 언급한 비율 88%, 일회용품에 대한 걱정과 환경문제에 대한 소신 56%, 경제적 이점 28%, 냄새 감소 16%, 부드러운 촉감 12%, 기타 16%(복수 응답 포함)로 건강과 환경문제로 면생리대를 사용하거나 했던 것으로 나타났다.

면생리대 사용 시 느낀 긍정적 효과로는 건강 개선과 피부의 편안함, 환경 친화성, 냄새 감소, 부드러운 촉감, 경제적 이점, 소리 없음, 답답함과 불편함 개선, 심리적 안정감 등이었으며, 면생리대의 단점은, 세탁, 휴대성, 교체주기, 두께, 흡수력 등이 있었으며, 이것은 아기의 천기저귀에 대한 단점을 서술한 것과 거의 일치한다. 실제로 천기저귀 사용자인 아기는 사용상의 장단점을 표현하기 힘들다. 따라서 천기저귀의 느낌은 엄마의 면생리대 사용 경험이 아이의 천기저귀 사용 경험을 대신하는 것으로 보인다.

Linda의 연구에 의하면 여성의 최대 20% 정도가 일상생활에 지장을 주는 정도의 생리통에 시달린다고 한다[21]. 면생리대는 생리통에 대한 대안적 방법의 하나로 인식되고 있으며, 난임 여성들이 늘면서 환경호르몬은 여성들에게 더욱 민감한 주제가 되었다. 기업들은 일회용 생리대의 고급화 전략과 동시에 다회용 면생리대 제품의 다양화와 구입의 편리성을 제공하고 있다. 분석 결과, 면생리대 사용에서 좋은 경험을 가졌던 엄마는 아이의 천기저귀 사용 결정에 더욱 적극적

이고 긍정적인 것으로 심층 분석 결과에서 드러났다.

또한, 2차 자료 수집 기간에는 1차 기간과 달리 “남편”이라는 키워드 빈도가 높게 나타났다. 이는 주로 천기저귀 사용 결정의 맥락에서 언급되고 있으며, 남편 언급 비율에서 육아 동참과 천기저귀에 대한 적극적 지지와 같은 긍정적 태도를 보인 비율은 69.2%, 걱정하거나 부정적 태도는 19.2%, 모르겠음 11.5%로 나타났다. 그리고 환경이나 경제적 이점으로 남편이 먼저 천기저귀를 제안하는 경우도 11%나 되었다. 도출된 분석 결과에 대한 신뢰성과 일관성을 검증하였으며, 2차 분석 결과를 종합하여 다음과 같은 명제를 제시한다.

연구 명제 2:

“엄마의 면생리대 사용 경험이 긍정적일수록 배우자가 적극적 양육 태도를 보일수록 아이의 천기저귀 사용 결정에 더 적극적이고 긍정적인 참여 행동을 보인다.”

4-5 분석 방법론의 결과 비교

표 4는 분석 방법론의 결과 비교표이다. 네트워크 기반 단일 분석 방법은 인과관계 도출과 맥락적 의미 분석에 강점이 있지만, 사람의 직관과 경험을 기반으로 한 통찰이기 때문에 시간과 비용이 많이 소요되며, 빅데이터와 같이 방대한 데이터에 적용하는 것은 현실적으로 많은 어려움이 있다. 또한, 네트워크 기반 분석 역량이 결과에 큰 영향을 미칠 수 있다.

제안된 BIGNET 모델은 두 방법론의 상호 보완적인 장점을 활용하여 높은 수준의 분석 결과를 도출할 수 있음을 확인

표 4. 분석 방법론의 결과 비교

Table 4. Comparison of results in analysis methodologies

Attribute	Big Data Analysis	Netno-graphy	Proposed BIGNET
Causal relationship inference	△	◎	◎
Contextual understanding	△	◎	◎
Decision-making speed	○	○	◎
Speed of results generation	◎	△	○
Scientific analysis and visualization	◎	△	○
The impact of analytical skills on results	△	◎	◎
Data processing speed and quantity	◎	△	○
Objectivity of data	◎	△	○
Flexibility in data types	○	◎	◎
Interpretability of results	○	◎	◎
Real-world applicability	○	○	◎
User friendliness	◎	△	○
Detecting Structural changes in the community	△	◎	◎

◎:best ○:good △:normal

하였다. 제안된 모델은 단일 방법론에 비해 인과관계 도출, 맥락적 이해, 종합적이고 빠른 의사결정 지원, 데이터 타입에 대한 유연성, 결과 해석 가능성, 커뮤니티의 내부 구조적 변화 감지 면에서 단일 분석 방법론과의 비교 우위에 있음을 확인하였다.

V. 결론 및 향후 연구

본 논문은 온라인 소비자 커뮤니티의 참여 행동을 심층적으로 이해하기 위해 빅데이터와 네트워크 기반 통합 분석 모델인 BIGNET 모델을 제안하고, 실제 온라인 커뮤니티를 대상으로 실증 연구를 수행하여 모델의 타당성과 우수성을 검증하였다.

BIGNET 통합 분석 모델은 단순히 빅데이터를 활용한 분석 결과를 제공하는 것을 넘어, 빅데이터 분석 결과의 현상을 이해하고 정확히 해석하는 데 그 목적이 있다. 이에 국내 대표 천기저귀 온라인 소비자 커뮤니티의 가입 인사 게시글의 통합 분석 과정을 통해, 빅데이터 분석 결과만으로 이해하기 어려운 현상들을 네트워크 기반 분석 과정을 통해 그 이유를 설명할 수 있음을 보였다. 통합 분석 과정을 통해 빅데이터 분석 결과에서 3개의 연구 질문을 정하고, 그 질문을 분석한 결과는 다음과 같이 도출할 수 있었다. 첫째, 빅데이터 분석 결과에 나타난 여아 언급 비율이 남아 언급 비율보다 높은 이유를 분석하였다. 둘째, 빅데이터 분석 결과에서 자녀의 호칭이 첫째보다 둘째가 더 많이 언급되는 이유를 분석하였다. 셋째, 천기저귀 사용에 있어 엄마의 면생리대 사용 경험과 배우자 언급이 높은 이유에 대한 분석 결과이다.

단일 분석 방법론보다 제안된 BIGNET 모델은 인과관계 도출과 맥락적 분석뿐만 아니라 종합적이고 다층적 분석이 가능하여 빠르고 정확한 의사결정을 지원할 수 있다. 그리고 빅데이터 분석에서는 잘 드러나지 않는 커뮤니티의 내부 구조적 변화와 그 원인까지 감지할 수 있음을 분석 과정을 통해 확인할 수 있었다.

또한, 제안된 모델은 방대한 데이터에서 질적 분석을 위한 데이터나 구간을 빠르고 정확하게 선정할 수 있다. 이것은 선행되는 빅데이터 분석 결과를 활용하여 심층 분석이 필요한 키워드나 연구 질문을 찾고, 빠르고 효율적으로 대상을 압축시켜 질적 분석을 수행할 수 있기 때문이다. 이런 통합 접근 방식은 빅데이터 분석의 한계를 보완하면서도 다소 주관적일 수 있는 네트워크 기반의 단점을 상호 보완할 수 있다. 본 논문에서는 실증적 사례분석을 실행함으로써 다양한 대상으로의 확장 가능성과 학문적 가치를 가진다. 이는 다양한 문제 해결을 위한 도구로써 실용 주체들에게 실질적 기준과 시사점을 제공해 줄 수 있을 것이라 기대된다. 향후 연구과제로는 다양한 소셜미디어에 관한 사례분석 결과들을 비교 분석하고, 네트워크 기반 범위나 대상을 자동으로 정확히 선정하는 방법에 관한 연구의 필요성이 있다.

감사의 글

이 연구는 2018년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2018S1A5B5A07071391)

참고문헌

- [1] F. H. Nah, R. Zheng, J. Cai, K. Siau, and L. Chen, "Generative AI and ChatGPT: Applications, Challenges, and AI-Human Collaboration," *Journal of Information Technology Case and Application Research*, Vol. 25, No. 3, pp. 277-304, 2023. <https://doi.org/10.1080/15228053.2023.2233814>
- [2] H. S. Sætra, "Generative AI: Here to Stay, But for Good?," *Technology in Society*, Vol. 75, 102372, November 2023. <https://doi.org/10.1016/j.techsoc.2023.102372>
- [3] C. Madsbjerg, *Sensemaking: The Power of the Humanities in the Age of the Algorithm*, New York, NY: Hachette, 2017.
- [4] E. Uprichard and L. Dawney, "Data Diffraction: Challenging Data Integration in Mixed Methods Research," *Journal of Mixed Methods Research*, Vol. 13, No. 1, pp. 19-32, 2019. <https://doi.org/10.1177/1558689816674650>
- [5] N. G. Fielding, "Triangulation and Mixed Methods Designs: Data Integration with New Research Technologies," *Journal of Mixed Methods Research*, Vol. 6, No. 2, pp. 124-136, 2012. <https://doi.org/10.1177/1558689812437101>
- [6] R. Y. Nooraie, J. E. M. Sale, A. Marin, and L. E. Ross, "Social Network Analysis: An Example of Fusion Between Quantitative and Qualitative Methods," *Journal of Mixed Methods Research*, Vol. 14, No. 1, pp. 110-124, 2020. <https://doi.org/10.1177/1558689818804060>
- [7] J. W. Creswell, *A Concise Introduction to Mixed Methods Research*, Sage Publications, 2021.
- [8] P. Ho, K. Chen, A. Shao, L. Bao, A. Ai, A. Tarfa, ... and M. Brauer, "A Mixed Methods Study of Public Perception of Social Distancing: Integrating Qualitative and Computational Analyses for Text Data," *Journal of Mixed Methods Research*, Vol. 15, No. 3, pp. 374-397, 2021.
- [9] S. Park, S. Park, and S. Oh, "How to Identify Customer Needs Based on Big Data and Netnography Analysis," *Information Systems Review*, Vol. 21, No. 4, pp. 175-195, November 2019. <http://dx.doi.org/10.14329/isr.2019.21.4.175>
- [10] P. Russom, Big Data Analytics, TDWI, TDWI Best Practices Report, Fourth Quarter 2011, 2011.
- [11] J. Ginsberg, and M. H. Mohebbi, R. S. Patel, L. Brammer, M. S. Smolinski, and L. Brilliant, "Detecting Influenza Epidemics Using Search Engine Query Data," *Nature*, Vol. 457, pp. 1012-1014, 2009. <https://doi.org/10.1038/nature07634>
- [12] S. Pollett, W. J. Boscardin, E. Azziz-Baumgartner, Y. O. Tinoco, G. Soto, C. Romero, ... and G. W. Rutherford, "Evaluating Google Flu Trends in Latin America: Important Lessons for the Next Phase of Digital Disease Detection," *Clinical Infectious Diseases*, Vol. 64, No. 1, pp. 34-41, January 2017. <https://doi.org/10.1093/cid/ciw657>
- [13] J. Y. Kim, "Big Data & Its Subjects - Digital Citizens Are Constructed," *Economy & Society*, No. 105, pp. 94-126, 2015. <https://doi.org/10.18207/criso.2015..105.94>
- [14] R. V. Kozinets, "The Field Behind the Screen: Using Netnography for Marketing Research in Online Communities," *Journal of Marketing Research*, Vol. 39, No. 1, pp. 61-72, 2002. <https://doi.org/10.1509/jmkr.39.1.61.18935>
- [15] R. Kozinets, Netnography: Radical Participative Understanding for a Networked Communications Society, in *The Sage Handbook of Qualitative Research in Psychology*, 2nd ed. Thousand Oaks, CA: Sage Publication Inc., 2017.
- [16] R. V. Kozinets, *Netnography: The Essential Guide to Qualitative Social Media Research*, 3rd ed. Thousand Oaks, CA: Sage Publication, 2019.
- [17] A. Choi and J. Y. Rha, "Consumers' Participation in Online Community: A Netnography Approach," *Journal of Consumer Studies*, Vol. 21, No. 1, pp. 1-31, 2010.
- [18] A. Hussein, "The Use of Triangulation in Social Sciences Research: Can Qualitative and Quantitative Methods Be Combined?," *Journal of Comparative Social Work*, Vol. 4, No. 1, pp. 106-117, 2009. <https://doi.org/10.31265/jcsw.v4i1.48>
- [19] N. K. Denzin and Y. S. Lincoln, *The Landscape of Qualitative Research*, 3rd ed. Los Angeles, CA: Sage, 2008.
- [20] Hankook Research. [Child/Childcare Awareness Survey] Preferred Child's Gender, Child Raising Awareness [Internet]. Available: <https://hrcopinion.co.kr/archives/23655>.
- [21] L. French, "Dysmenorrhea," *American Family Physician*, Vol. 71, No. 2, pp. 285-291, 2005.



박순화(Soon-Hwa Park)

1996년 : 경상대학교 컴퓨터과학과(이학사)
2015년 : 경상대학교 컴퓨터과학과(공학박사)

2015년~현 재: 경상국립대학교 강사
※ 관심분야 : 빅데이터, AI, Netnography



이점숙(Jeom-Suk Lee)

1994년 : 경상대학교 컴퓨터과학과(이학사)
2011년 : 경상대학교 컴퓨터과학과(공학박사)

2009년~현 재: 경상국립대학교 강사
※ 관심분야 : 멀티미디어, 컴퓨터교육, 빅데이터, AI



박은경(Eun-Kyoung Park)

2000년 : 경상대학교 컴퓨터과학과(이학사)
2010년 : 경상대학교 컴퓨터과학과(공학박사)

2010년~현 재: 경상국립대학교 강사
※ 관심분야 : 빅데이터, AI, Netnography



김은경(Eun-Kyoung Kim)

2000년 : 경상대학교 컴퓨터과학과(이학사)
2015년 : 경상대학교 컴퓨터과학과(공학박사)

2015년~현 재: 경상국립대학교 강사
※ 관심분야 : 빅데이터, AI, Netnography



서영건(Yeong Geon Seo)

1987년 : 경상대학교 전산과(이학사)
1997년 : 숭실대학교 전산과(공학박사)

1989년~1992년: 삼보컴퓨터
1997년~현 재: 경상국립대학교 컴퓨터공학과 교수
2022년~현 재: 경상국립대학교 정보전산처장
※ 관심분야 : 컴퓨팅 사고, AI, 영상 인식, 컴퓨터네트워크