

텍스트 마이닝을 이용한 소셜 미디어 상의 MBTI 성격 유형별 관심 주제 차이 분석

우위항*, 남인용**, 이문기***

Analyzing Differences in Interest Topics According to MBTI Personality Types on Social Media using Text Mining

Yu-Hang Wu*, In-Yong Nam**, and Wen-Qi Li***

요약

디지털 시대의 지속적인 발전과 함께, 네트워크 사용자의 개별화된 요구와 행동 패턴을 이해하는 것이 점점 더 중요해지고 있다. 이 연구는 서로 다른 MBTI 성격 유형의 사용자가 네트워크에서 남긴 콘텐츠 댓글을 분석하여, 그들이 관심을 갖고 있는 주제와 분야에 차이가 나타나는가를 살펴보았다. Python을 사용하여 네트워크에서 수집한 다양한 MBTI 유형의 사용자 댓글을 텍스트 마이닝 및 분석하여 관심 주제를 추출하였고, 그 결과를 통해 서로 다른 MBTI 유형 간의 관심 분야 차이를 확인하였다. 이용할 수 있는 정보가 폭발적으로 증가함에 따라 다양한 성격 유형별로 사용자의 요구와 행동을 더욱 심층적으로 이해하는 것은 네트워크 서비스와 제품을 최적화하는 데 도움이 된다. 이 연구는 학문적으로는 MBTI 유형별로 사용자의 요구를 이해하는 새로운 관점과 도구를 제공하고, 실무적으로는 사용자 중심의 콘텐츠를 배포하고 서비스를 혁신하는 실증적 근거가 될 것이다.

Abstract

As the digital age continues to evolve, understanding network users' individualized needs and behavior patterns is becoming increasingly important. This study explores the comments left by different MBTI personality types on the network to reveal differences in their interests. Using Python, we analyzed various MBTI types' comments to extract interest topics and identified differences between them. In an era of abundant information, understanding diverse personality types' needs and behaviors will assist in optimizing network services and products. This research not only academically provides new perspectives and tools to understand individual user needs but also practically serves as a valuable reference for promoting user-centered content distribution and service innovation.

Keywords

myers-briggs type indicator, MBTI, interest topics, text analysis, LDA topic modeling, visualization

* 부경대학교 미디어커뮤니케이션학과 석·박사과정
- ORCID: <https://orcid.org/0009-0009-0235-8683>
** 부경대학교 미디어커뮤니케이션학과 교수(교신저자)
- ORCID: <https://orcid.org/0000-0002-7754-0158>
*** 부경대학교 미디어커뮤니케이션학과 박사과정
- ORCID: <https://orcid.org/0009-0009-4949-1705>

· Received: Aug. 02, 2023, Revised: Aug. 20, 2023, Accepted: Aug. 23, 2023
· Corresponding Author: In-Yong Nam
Dept. of Media Communication (48513) 45, Yongso-ro, Nam-Gu, Busan,
Korea
Tel.: +82-51-629-5383, Email: tarzan@pknu.ac.kr

1. 서 론

인터넷의 빠른 발전과 디지털 시대의 도래에 따라, 네트워크 사용자의 선호와 행동을 연구함으로써 사용자의 요구와 행동을 더 잘 이해하고, 네트워크 서비스와 제품을 최적화할 수 있다. 정보가 폭발적으로 증가하는 시대적 배경에서, 네트워크 환경에서 사용자가 어떻게 정보를 선택하고 소비하는지를 이해하는 것은 다수의 학술단체와 기업들이 주목하는 주제가 되었다. 개인화 서비스와 제품 제공이 점점 중요해지는 오늘날에, 사용자의 선호와 행동을 심층적으로 파악하면 과학기술과 인간 생활의 더 나은 융합을 촉진하고 상업적 효과를 높이는 데 도움이 된다.

사람들의 성격 유형은 그들의 기호와 행동을 결정하는 중요한 요소 중 하나로, 심리학과 사회학 분야에서 널리 연구되어왔다. 그중에서도 MBTI (Myers-Briggs Type Indicator)는 인기 있는 성격 테스트 도구로, 사람들의 성격을 16가지 서로 다른 유형으로 분류한다. 각 유형은 사람들이 각각의 상황에서 보여줄 수 있는 행동 경향을 반영한다.

그러나, 네트워크 환경에서 사람들의 성격이 그들의 의사소통 방식과 행동에 어떻게 영향을 미치는지에 대한 논의는 아직 충분하게 이루어지지 않았다. 특히, 온라인 커뮤니티와 소셜 네트워크 같은 복잡한 상호작용 환경에서, 사람들의 성격 유형은 그들의 정보 선택, 의사소통 방식, 심지어는 다른 사람들과의 상호작용 스타일 등과 밀접하게 관련되어 있을 수 있다. 따라서, MBTI의 16가지 성격 유형별로 사용자들이 네트워크 커뮤니티에서 다양한 주제에 대해 어떻게 관심을 갖는가를 분석하고 이해함으로써, 사람들의 가상 세계에서의 행동 패턴을 밝혀낼 수 있을 뿐만 아니라, 네트워크 콘텐츠의 개인화 추천, 커뮤니티 관리, 사용자 경험 최적화 등에 대한 참고 자료를 제공할 수 있다.

이 연구는 바로 이런 관점에서 MBTI의 16가지 성격 유형이 사용자들이 네트워크에서 각각 어떤 주제에 주목하는가, 어떤 행동 경향을 보이는가를 살펴보았다. 이 연구의 목적은 텍스트 마이닝 기술을 사용하여 대량의 텍스트 데이터를 분석하여 사

람들의 성격 유형과 온라인 커뮤니티 및 소셜 네트워크에서의 행동 사이의 연관성을 밝혀내는 것이었다.

이 연구는 먼저 MBTI 성격 분류의 이론적 배경을 소개하고, 그다음으로 데이터 수집 및 분석 방법에 대해 자세히 설명한다. 그다음, 네트워크 환경에서 서로 다른 성격 유형의 사용자들의 행동 특성을 살펴보고, 이러한 특성이 그들의 정보 선택 및 사회적 상호 작용에 어떻게 영향을 미치는지에 대해 자세히 살펴볼 것이다. 마지막으로, 이러한 분석 결과가 개인화된 추천, 커뮤니티 관리, 사용자 경험 최적화 등의 분야에 어떻게 적용될 수 있는지에 대해 논의할 것이다.

전반적으로, 이 연구는 네트워크 사용자의 복잡한 행동을 이해하는 데 유용한 새로운 시각과 도구를 제공하고, 더 인간적인 네트워크 서비스와 제품을 실현하기 위한 실증적 근거를 제공하고자 하였다.

이 연구의 구성은 다음과 같다.

제1장 서론에서는 논문의 연구 배경과 논문의 구성에 대해 소개하였다.

제2장은 연구의 이론적 배경을 제시하기 위한 장으로, 총 세 부분으로 구성되어 있다. 첫 번째 부분에서는 MBTI의 네 가지 특성과 이에 해당하는 16가지 성격 유형에 대해 설명하였고, 두 번째 부분은 인터넷의 개인화서비스의 내용과 그 장점을 요약하였으며, 세 번째 부분에서는 LDA 주제 모델과 그 주요 용도를 소개하였다.

제3장은 공개된 데이터 세트를 활용해 MBTI 성격 유형에 따른 사용자의 게시물 주제를 분석하는 연구방법을 설명했다. 분석 도구로는 Python과 NLTK 라이브러리를 사용하였으며, 이를 통해 각 MBTI 유형에 따른 주요 주제와 관련 키워드를 파악할 수 있었다.

제4장은 데이터 분석 결과 부분이다. MBTI의 네 가지 주요 차원을 기반으로 한 단어 빈도 통계 분석을 수행하였다. 각 차원별 상위 20개의 키워드를 확인한 후, 워드 클라우드를 통해 시각화하였고, LDA 모델을 구축하여 각 사용자의 게시물에 대한 주제 분류를 진행하였다.

제5장 결론 및 논의에서는 이 연구의 분석 내용을 종합하였다. MBTI의 다양한 유형에 따라 사용자에 대한 개인화된 콘텐츠 제공의 중요성을 강조하며, 이를 통해 서비스의 만족도와 충성도를 향상시킬 수 있다는 점을 제안하였다.

II. 연구 배경

2.1 MBTI

MBTI는 마이어스(Myers)와 브릭스(Briggs)가 스위스의 정신분석학자인 카를 융(Carl Jung)의 심리 유형론을 토대로 고안한 자기 보고식 성격 유형 검사 도구이다[1]. MBTI 성격유형 검사는 융의 이론을 바탕으로 개인이 쉽게 응답 가능한 자기 보고 형식의 문항을 통해 자신이 인식하고 판단할 때의 선호 경향을 찾는 검사이다. 따라서 MBTI는 접근성이 쉬울 뿐만 아니라, 이러한 선호 경향들이 합쳐져 개인이 행동할 때 있어서 어떤 영향을 미치는가를 파악하여 실생활에서 응용할 수 있도록 제작되었다[2].

Myers-Briggs 유형 척도(MBTI)는 4개의 특성 이분법 쌍을 사용하여 16개의 인격 유형을 만드는 인격 유형론이다. 네 가지 쌍은 외향-내향(E-I), 감각-직관(S-N), 사고-감정(T-F), 판단-인식(J-P)이다. 결과적으로 각각의 네 쌍에서 이진법으로 표현되며, 이를 결합하여 인격 유형을 만든다[3].

노안영과 강영신이 2018년에 출판한 <인간 이해 및 성장을 위한>성격심리학>에 따르면, 융의 심리 유형론은 각 개인이 인식(감각과 직관) 및 판단(사고와 감정) 기능에서 자신만의 선호 방식을 가지고 있으며 이러한 선호 방식이 그들의 성격 유형을 형성하는 데 중요하다는 점을 강조한다. 어떤 사람에게 한 기능(예를 들어, 감각이나 사고)이 지배적이라면, 그 반대의 기능(예를 들어, 직관이나 감정)은 상응하여 약화된다[4].

2.2 인터넷의 개인화 서비스

인터넷의 개인화 푸시는 사용자 개인의 취향, 행동 및 요구를 기반으로 하는 정보 푸시 방식이다.

이 방식은 일반적으로 개인정보와 사용자의 행동 데이터(검색 이력, 검색 기록, 쇼핑 습관 등)를 분석하여 수행한다. 개인화 푸시의 장점은 사용자가 관심 있는 정보나 제품을 더 빨리 찾을 수 있도록 도와 사용자 경험과 만족도를 높일 수 있다는 것이다. 일부 고급 개인화 푸시는 기계 학습 및 인공지능 기술을 사용하여 푸시의 정확성과 효율성을 향상시킬 수도 있다.

경쟁우위를 차지하기 위한 마케팅 전략의 하나로, 기업에서는 고객의 욕구를 정확히 충족시킬 수 있는 맞춤형 혜택을 제공하는 다양한 개인화 서비스가 이루어지고 있다[5]. 개인화 서비스는 고객의 선호도와 행동에 대한 정보를 기반으로 개개인에게 맞춤형된 서비스와 정보를 제공하는 능력을 의미한다[6]. 개인화 서비스를 통해 고객은 자신에게 가장 적합한 정보를 편리한 방법으로 얻을 수 있게 되었으며, 고객에 관한 정보와 자료를 얻게 됨으로써 기업은 지속적인 이용과 관계 구축을 유도할 수 있게 되었다[7].

고객이 입력한 개인정보 혹은 기업에서 수집한 구매 관련 정보를 바탕으로, 상품과 광고, 메뉴 등의 콘텐츠를 제공하고, 고객의 니즈에 맞는 맞춤형된 웹페이지를 구현함으로써 고객의 선호에 더욱 적합한 혜택 제공이 가능해졌다[8].

2.3 LDA토픽모델링

토픽모델링은 텍스트마이닝(Text mining) 기법 중의 하나로 비구조화된 텍스트와 같이 비정형화된 자료들의 집합에서 의미 있는 토픽(주제)을 추출해주는 확률적 모델 알고리즘으로(박준형·오효정, 2017; Blei, Ng, & Jordan, 2003), 본문의 숨겨진 의미구조를 확인하기 위해 사용되는 기법이다[9]. 토픽모델링은 방대한 양의 문서집합에서 주요 토픽을 추출하고 각 토픽에 대응되는 문서를 식별하여 제공하며, 하나의 문서가 하나의 토픽으로만 할당되는 일반적인 군집화(Clustering) 기법과는 다르게, 하나의 문서가 여러 토픽에 동시에 대응될 수 있기 때문에 현실 세계의 모델링에 보다 적합한 기법으로 평가받고 있다[10].

토픽모델링 초기에는 확률적 잠재의미분석(pLSA, probabilistic Latent Semantic Analysis) 기법이 사용되었으나, 이 연구에서 채택한 LDA(Latent Dirichlet Allocation)분석은 잠재의미분석의 확률분포에 대한 한계점을 보완한 것으로 토픽모델링의 가장 대표적인 방법론으로 활용되고 있다[11]. LDA 토픽모델링은 하나의 문서에 복수의 토픽이 내재된 상황을 전제로 하여 문서 내 여러 단어의 조합을 토대로 문서의 토픽들을 확률적으로 파악하는 통계적 분석 방법이다[12]. 토픽모델링은 단순히 단어의 출현 빈도수를 기반으로 하는 텍스트 분석과는 다르며, 토픽이라는 변수를 통한 확률 기반의 분석을 수행함으로써 과적합 문제가 적고, 새로운 데이터가 투입되어도 분석 가능하다는 장점이 있다[13].

LDA는 문서, 단어 등 관찰된 변수를 통해 문맥이나 문서의 구조 등 보이지 않는 변수를 추론하는 방법[14]으로 토픽에 따라 문서 집합을 재구성할 수 있으며, 데이터, 이미지 및 소셜 네트워크에서 특정한 패턴을 찾는 데 사용될 뿐만 아니라[11], 인터넷 리뷰, 신문기사, 논문 등의 동향을 파악하는 등 잠재 주제를 파악하는데 유용하게 활용되고 있다[15].

III. 연구 방법

이 연구의 데이터는 Personality Cafe 포럼에서 수집한 공개 데이터 세트로, 이 포럼은 연구에 풍부한 자료를 제공한다. 데이터 세트는 총 8600행의 데이터인데, 다양한 사용자의 게시물과 관련 MBTI 유형 레이블을 포함하고 있다. 사용된 데이터 세트는 공공재성이므로 윤리적, 법적 측면에서도 허용된다. 이러한 데이터를 분석함으로써 포럼에서 사용자들의 다양한 성격 유형별로 그들이 논의하는 주제와 화제를 깊이 있게 이해할 수 있으며, 그들의 관심사와 행동 경향을 파악할 수 있다.

이 연구에서는 Python 프로그래밍 언어와 NLTK 라이브러리를 사용하여 데이터를 분석하였다. 먼저, 분석의 정확성과 일관성을 보장하기 위해 데이터 세트의 텍스트 내용에 대한 일련의 전처리 작업을 거쳤다. 구체적인 전처리 단계는 소문자로 변환, 구두점과 숫자 제거, 토큰화, 불용어 제거 및 어형 복

원을 포함하며, 이러한 단계는 주로 string 라이브러리와 NLTK 라이브러리의 관련 함수를 통해 수행되었다. 그 후 사용자 MBTI의 네 가지 주요 차원, 즉 내향/외향(Intro_Extro), 직관/감각(Intuition_Sensing), 사고/감정(Thinking_Feeling), 판단/지각(Judging_Perceiving) 별로 각 게시물을 분석하였으며, 각 차원 내 유형에 대한 단어 빈도 통계를 작성하고 가장 높은 빈도수를 가진 상위 20개의 핵심 단어를 추출하여 추가 분석 및 시각화를 수행하였다.

또한, CountVectorizer를 사용하여 문서-단어 행렬을 구축하고, scikit-learn의 Latent Dirichlet Allocation을 사용하여 10개 주제의 LDA 모델을 구축하였다. 각 사용자의 게시물에 주제 분류를 수행하고 이를 새로운 열인 'topic'에 기록한 다음, 사용자 유형과 주제별로 데이터를 그룹화하고 분석하여 각 사용자 유형이 관심을 가진 주제 또는 화제의 분포 상황을 통계화하였다. 이 과정에는 그룹화된 데이터를 2차원 표로 변환하고 각 사용자 유형의 주제 계수를 정규화하여 주제 분포 백분율을 얻는 작업도 포함되었다. 마지막으로, 각 MBTI 유형이 주목하는 주제 분포 상황을 보여주는 Data Frame을 출력하고, 각 주제 내의 키워드 분포를 확인함으로써 이러한 주제를 설명하였다.

이 연구는 사용자들의 다양한 성격 유형별로 그들의 언어 패턴 및 행동 경향을 깊이 있게 이해할 수 있는 견고한 기반을 제공하기 위해 다양한 분석 방법을 사용하였다. 첫째, 공개 데이터 세트를 사용하여 다양한 성격 유형에 대한 세심한 언어 및 행동 분석을 수행할 수 있었다. 이 분석의 정확성과 일관성을 보장하기 위해 특히 NLTK 및 string 라이브러리를 통해 텍스트 데이터의 품질을 효과적으로 최적화하기 위해 엄격한 데이터 전처리를 수행했다. LDA 모델의 구축은 이 연구의 또 다른 핵심 연결고리이다. Latent Dirichlet Allocation을 이용하여 주제 분류를 실현했을 뿐만 아니라 사용자 유형별 관심사를 정확하게 파악하였다. 마지막으로 시각화 도구와 주제에 대한 설명을 사용하여 각 주제에 대한 키워드 분포를 직관적으로 보여줌으로써 연구 결과를 보다 생생하고 이해하기 쉽게 만들었다.

이 연구의 연구방법은 사용자들의 다양한 성격 유형별로 그들의 사용자의 언어와 관심사에 대한 이해를 심화시킬 뿐만 아니라 개인화된 서비스 제공의 중요성을 강조한다. 더 언급할 가치가 있는 이 방법은 유연성이 높으며 유사한 다른 연구 또는 다양한 연구 분야에 쉽게 적용할 수 있다.

IV. 분석 결과

4.1 단어 빈도 통계

Myers-Briggs 유형 지표(MBTI)의 네 가지 주요 차원을 기반으로 한 단어 빈도 통계 분석에서 각 차원별로 몇 가지 두드러진 차이점을 발견했다. 표 1과 같다.

내향/외향 차원에서, 내향형 유형의 표본은 총 265,624개로, 4,193,621번 나타났으며, 외향형 유형의 표본은 108,680개로, 1,258,843번 나타났다. 직관/감각 차원에서, 직관형 표본의 수는 291,131개로, 4,726,244번 나타났고, 감각형 표본은 72,349개로, 726,220번 나타났다. 생각/감정 차원에서, 감정형 표본은 200,158개로, 총 2,999,170번 나타났으며, 생각형 표본은 183,146개로, 2,453,294번 나타났다. 판단/지각 차원에서, 판단형 표본은 159,955개로, 2,175,760번 나타났고, 지각형 표본은 222,550개로, 3,276,704번 나타났다. Matplotlib 라이브러리를 사용하여 이러한 데이터를 시각화했다(그림 1 참조). 그림 1은 MBTI 네 차원의 단어 빈도 통계 결과를 자세히 보여준다.

표 1. 단어 빈도 통계 결과
Table 1. Statistical results of word frequency

type	number of samples	number of results
Intro (I)	265624	4193621
Extro (E)	108680	1258843
Intuition (N)	291131	4726244
Sensing (S)	72349	726220
Feeling (F)	200158	2999170
Thinking (T)	183146	2453294
Judging (J)	159955	2175760
Perceiving (P)	222550	3276704

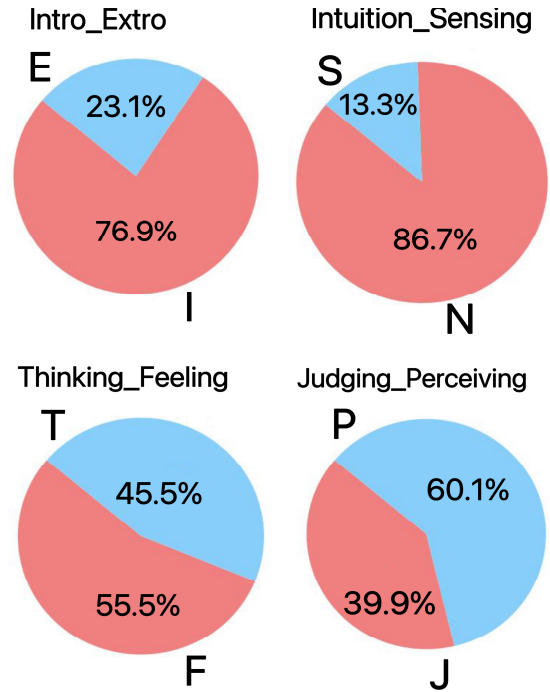


그림 1. 단어 빈도 통계
Fig. 1. Statistics of word frequency

이러한 분석 결과에서 다양한 MBTI 성격 유형이 언어 사용 면에서 뚜렷한 차이를 보인다는 점을 알 수 있으며, 서로 다른 성격 유형의 사람들이 자신의 표현과 의사소통에서 보여주는 이러한 차이의 구체적인 의미를 더 깊게 이해할 수 있다.

4.2 워드 클라우드

MBTI의 네 가지 주요 차원에 기반한 단어 빈도 통계 분석에서, 각 차원에서 상위 20개의 Top Keywords를 확인했으며 요약하면 다음과 같다.

첫째 차원은 내향/외향(Intro_Extro) 차원인데, 내향형(I)에서 'feeling', 'day', 'post', 'year' 등의 키워드는 내향형 개인의 감정, 시간, 자기 성찰에 대한 경향을 반영한다. 이 키워드들은 많은 내향형 사람들이 심사숙고와 개인적인 성찰을 하는 심리적 특징과 일치한다. 외향형(E)에서 'guy', 'thread', 'relationship' 등의 키워드는 외향형 사람들의 사회 활동과 인간 관계에 대한 주의를 강조한다. 이 키워드들은 외향형 사람들이 대인 교류에서 활발하다는 것을 보여준다(그림2).



그림 2. 내향/외향
Fig. 2. Intro_Extro

둘째 차원은 직관/감각(Intuition_Sensing) 차원인데, 직관형(N)에서 'feeling', 'guy', 'world', 'understand' 등의 키워드는 직관형 사람들이 이해, 감정, 인간관계의 동향을 탐색하고 추상적인 개념을 선호함을 보여준다. 감각형(S)에서 'guy', 'day', 'help', 'talk' 등의 키워드는 감각형 사람들이 현실, 구체적인 느낌, 타인을 돕는 것에 중점을 둔다는 것을 강조하며, 현실의 세부 사항에 중점을 둔 특성과 일치한다(그림3).



그림 3. 직관/감각
Fig. 3. Intuition_Sensing

셋째 차원은 사고/감정(Thinking_Feeling) 차원인데, 감정형(F)에서 'feeling', 'others', 'help', 'sound' 등의 키워드는 감정형 사람들이 대인 관계와 감정 표현에 중점을 둔다는 점을 보여주며, 동정과 공감의 가치관을 나타낸다. 사고형(T)에서 'thread', 'point', 'problem', 'thinking' 등의 키워드는 사고형 사람들의 논리, 분석, 비판적 사고 경향을 강조하며, 분석과 객관적 평가의 요구와 일치한다(그림4).



그림 4. 사고/감정
Fig. 4. Thinking_Feeling

넷째 차원은 판단/지각(Judging_Perceiving) 차원인데, 판단형(J)에서 'feeling', 'question', 'help', 'understand' 등의 키워드는 판단형 사람들이 계획, 조직, 이해에 중점을 둔다는 점을 반영하며, 질서와 구조의 필요성과 일치한다. 지각형(P)에서 'guy', 'idea', 'world', 'ever' 등의 키워드는 지각형 사람들의 유연성, 개방성, 탐험적인 특성을 나타내며, 개방된 사고와 자유로운 탐험의 경향과 일치한다(그림 5).

MBTI의 네 가지 차원을 나타내는 20개의 키워드는 단어 빈도 분석을 통해 선택되었으며, 서로 다른 MBTI 성격 유형이 언어 표현에서 독특한 경향을 갖고 있음을 보여주고, 각 차원은 사회적 상호작용과 개인행동에서의 고유한 특징을 강조한다.



그림 5. 판단/지각
Fig. 5. Judging_Perceiving

4.3 LDA 토픽 모델링 결과

각 MBTI 유형이 다양한 주제에 대해 얼마나 관심을 갖고 있는지 깊이 분석하기 위해, scikit-learn 라이브러리의 Latent Dirichlet Allocation을 사용하여 LDA 모델을 구축하였고, 주제의 수를 10으로 설정하였다. 그 다음, 이 모델을 문서-단어 행렬에 적용하여 각 문서의 주제 분포를 얻었다. 각 사용자의 게시물에 대해 주제 분류를 진행하고, 분류 결과를 "topic"이라는 새로운 열에 기록하였다. 표 2와 같다.

표 2. 토픽 0 ~ 9 관련 단어
Table 2. Words related to topic 0 ~ 9

Subject number	Related words
Topic 0	sent, using, iphone, tapatalki, tapatalk, emoji, ipad, nexus, new, tapatalkim
Topic 1	ive, work, love, well, day, year, even, going, thats, cant
Topic 2	love, song, day, year, friend, happy, school, going, night, last
Topic 3	friend, love, ive, type, lot, infj, someone, always, guy, also
Topic 4	life, never, love, even, always, friend, someone, ive, find, person
Topic 5	type, enneagram, tmlt, sx, tritype, socionics, forum, MBTI, sxso, spsx
Topic 6	welcome, god, perc, forum, believe, argument, need, could, thread, said
Topic 7	school, year, music, love, favorite, movie, also, ive, friend, song
Topic 8	jag, är, sodium, det, inception, och, danish, inte, så, song
Topic 9	type, function, intp, intj, also, youre, entp, ive, fe, could

- topic 0은 이동 장치, 응용 프로그램 및 통신과 관련이 있다.
- topic 1은 개인 경험, 감정 및 생활과 관련이 있다.
- topic 2는 우정, 사랑 및 일상 생활과 관련이 있다.
- topic 3과 4는 각각 우정, 사랑, 성격 유형, 그리고 인생, 감정, 대인 관계와 관련이 있다.
- topic 5는 성격 유형 이론 및 분류에 중점을 둔다.
- topic 6은 종교, 포럼 토론 및 견해와 관련이 있다.
- topic 7은 학교, 음악, 영화 및 문화 취향과 관련이 있다.
- topic 8은 일부 관찰 (예: 사용된 단어 또는 표현 방식)로 스웨덴어 및 문화와 관련이 있을 수 있다.
- topic 9는 성격 유형 이론 및 심리학과 관련이 있다.

분석결과, 다음과 같은 흥미로운 경향을 관찰하고 통계 검정을 통해 그 중요성을 확인하였다: 그림 6 참조. 예를 들어, ENFJ 유형의 사람들은 주제 3(우정, 사랑 및 성격 유형, 58.42%) 및 주제 1(개인 경험 및 감정, 17.37%)에 중점을 둔다. ISTP 유형의 사람들은 주제 1(개인 경험, 감정 및 생활)에서 비율이 가장 높으며 53.41%에 달한다. INFP 유형의 사람들은 주제 4(인생, 감정 및 대인 관계, 37.99%) 및 주제 1(21.56%)에서 널리 분포한다. ESFJ 유형의 사람들은 주제 3(우정, 사랑 및 성격 유형)의 비율이 61.90%에 달한다.

이러한 분석결과는 각 MBTI 유형이 소셜 미디어에서 선호하는 특정 관심 주제를 나타내며 일부 심리학 이론과도 일치하는데, 예를 들어 일부 유형은 보다 대인화된 주제에 집중한다. 흥미롭게도, 주제 9는 많은 유형에서 높은 비율을 나타내어 유형 간의 보편성을 나타냈다.

이와 같은 분석결과에 근거해서 소셜 미디어에서의 사용자의 행동과 관심을 더 잘 이해함으로써, 개인화된 추천, 팀 구축 및 심리 분석 등의 분야에도 활용할 수 있다. 또한, 서로 다른 문화적 배경에서의 유사한 추세 및 가능한 실제 응용 프로그램을 더욱 심층적으로 연구할 수 있다.

소셜 미디어에서 다른 성격 유형의 사용자가 갖고 있는 관심 경향에 대한 이해를 높이는 것에서 더 나아가 주제 분석의 심층 연구를 수행할 수 있다.

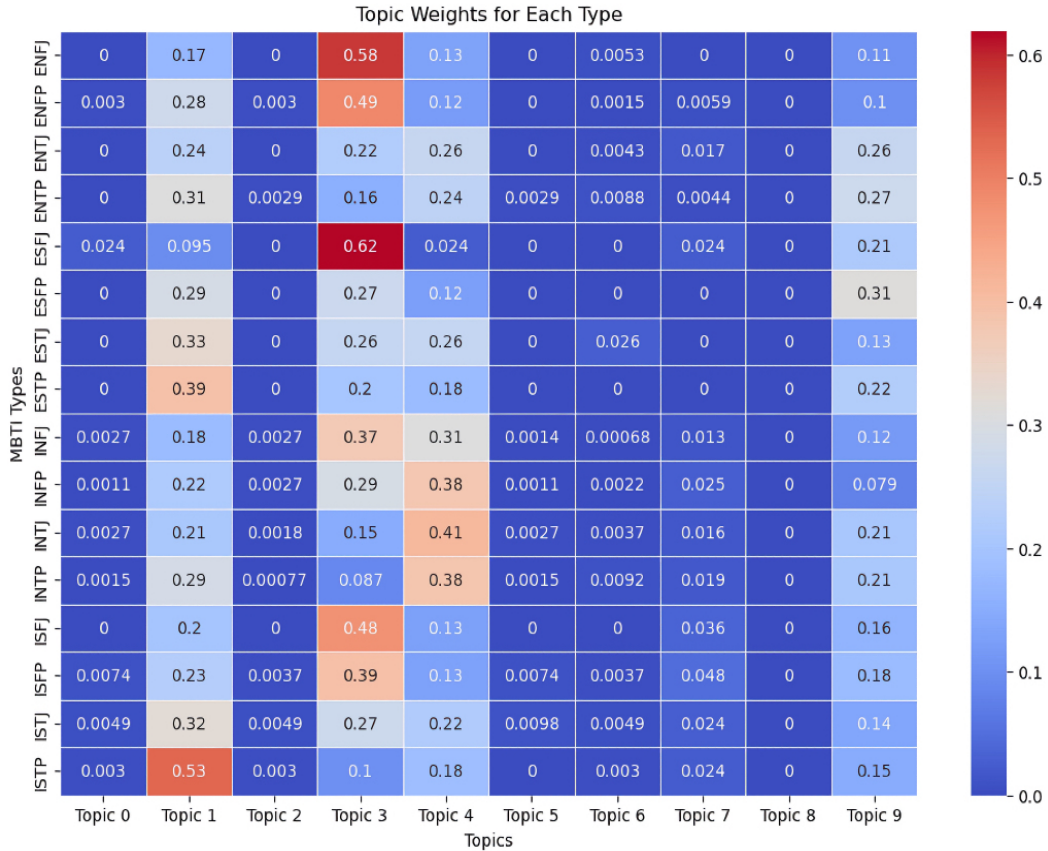


그림 6. 히트맵
Fig. 6. Heat map

이를 통해, 개인화된 서비스 및 정신 건강 개선과 같은 미래의 응용 프로그램에 대한 강력한 실증적 근거를 제공할 수 있다.

V. 결론 및 논의

이 연구의 주요 목적은 MBTI의 16가지 성격 유형에 따라 인터넷에서 보여주는 다양한 선호 경향의 차이를 분석하고, 선호 카테고리 및 활동성 등 관련 정보를 획득하여, 사용자의 MBTI 성격 유형을 기반으로 사용자에게 필요한 인터넷 콘텐츠를 정확하게 추천하는 것이다. 이를 통해 인터넷 산업에 종사하는 사업자와 사용자는 지속적인 성장의 실마리를 찾을 수 있으며, 인터넷 발전과 직접적, 간접적으로 관련이 있는 여러 산업에 참고 자료를 제공할 수 있다. 따라서, 이 연구에서는 소셜 네트워크에서 수집한 텍스트 데이터를 활용하여 MBTI 유형별 사용자들의 관심 주제를 분석하고 시각화하였다. 원형 그래프와 워드 클라우드를 통해 각 성격 유형별 주

제에 대한 선호 차이를 탐색하였다. 마지막으로, LDA를 활용하여 더 깊은 분석을 실시하고, 10개의 사용자 토론 주제를 도출하였으며, 이를 통해 사용자의 다양한 성격 유형별로 그들이 선호하는 주제의 차이를 이해하였다.

첫째, 16가지 성격 유형은 외향(E) - 내향(I), 감각(S) - 직관(N), 사고(T) - 감정(F), 판단(J) - 인식(P)의 네 가지 차원을 기준으로 분류된다. 이 네 가지 차원은 사람들이 정보를 어떻게 수용하고 처리하는지, 어떻게 결정을 내리는지, 그리고 그들이 외부 세계와의 상호작용을 어떻게 보고 조직하는지에 관련되어 있다. 키워드 빈도 분석과 워드클라우드 계산의 결과는 MBTI 유형이 네 가지 차원에서 각각의 독특한 언어 경향을 갖고 있음을 보여준다. 이것은 그들이 선택하여 사용하는 어휘와 표현 방식에서 나타난다. 외향형 인간은 텍스트에서 더욱 적극적이고 활발한 어휘를 사용하는 경향이 있고, 내향형 인간은 더욱 성찰적이고 내성적인 언어 표현을 사용하는 경향이 있다.

그들의 언어 경향을 분석함으로써, 그들의 소셜 미디어 또는 온라인 커뮤니티에서의 행동과 상호작용 방식을 이해하는 데 도움이 될 수 있다.

둘째, 시각화된 워드 클라우드 분석 결과를 통해 MBTI에서 각종 키워드에 대한 단어 사용 빈도를 확인할 수 있다. 분석 결과는 인터넷 상에서 다양한 MBTI 유형의 사용자들이 의사소통하고 자신을 표현하는 독특한 방식과 경향을 반영하는데 그치지 않고, 그들의 표현 선호도를 확인할 수 있게 해준다. 이는 네트워크 커뮤니티 관리자가 커뮤니티를 더 잘 관리하고 사용자의 요구와 선호도에 맞는 경험을 제공하는 데 도움이 된다.

셋째, LDA를 활용하여 MBTI 유형별로 다른 분야에 대해 얼마나 관심이 있는지 분석함으로써, 소셜 네트워크와 온라인 커뮤니티에서 사용자의 MBTI 유형과 그들이 다른 분야에 얼마나 관심이 있는지를 이해할 수 있다. 이러한 분석결과를 바탕으로 커뮤니티 활동을 더 원활하게 구성하고 조정하여, 커뮤니티의 활동성과 사용자 만족도를 높이는 데 도움이 된다. MBTI 성격 유형과 사용자의 특정 영역에 대한 관심의 관계를 살펴보면, 우리가 다른 성격 유형의 사용자의 관심사와 선호도를 이해하는 데 도움이 되며, 이는 사용자 행동을 이해하고 그에 따라 제품과 서비스를 최적화하는 데 매우 중요하다.

마지막으로, 분석 결과를 종합하면, 서로 다른 MBTI 유형이 소셜 네트워크에서 주로 어떤 주제와 화제에 관심을 갖는지를 파악할 수 있다. 각 MBTI 유형이 선호하는 키워드들은 언어 표현, 감정 경향, 취향 등에서 나타나는 MBTI 유형별 차이를 반영하며, 개인의 성격과 사회적 행동 간의 관계에 대한 심도 있는 연구를 위한 실증적 근거를 제공한다.

이 연구는 MBTI 성격 유형에 따른 사용자의 네트워크 콘텐츠 댓글 행동을 새로운 시각에서 깊이 분석하였다. 인터넷 카페에서 수집된 데이터를 활용하여 사용자의 실제 네트워크 활동에서 나타나는 관심 주제를 중심으로 조사하였다. 기존의 연구에서는 MBTI 성격 유형의 일반적인 행동 패턴에 중점을 둔 반면, 이 연구는 성격 유형과 언어 사용 패턴 사이의 관계를 탐구하였다.

이런 통합적 연구방법을 통해, 우리는 다른 성격 유형의 사람들이 어떤 주제나 화제에 집중하는지 파악할 수 있었다. 이러한 연구결과는 네트워크 서비스 제공자나 마케터들에게 매우 유용하게 작용한다. 사용자에게 개인화된 콘텐츠를 제공하면, 그들의 서비스 이용 경험을 향상시키며, 정보 검색에 소요되는 시간과 노력을 줄일 수 있다. 이는 결국 사용자의 서비스 만족도와 충성도를 높이는 효과를 가져올 것이다.

더불어, 이와 같은 분석은 인터넷 기업들에게도 큰 도움이 된다. 기업들은 제공한 콘텐츠에 대한 사용자의 반응을 모니터링하고 데이터를 수집, 분석함으로써 사용자의 필요와 행동 패턴을 더 정확하게 파악할 수 있다. 이를 바탕으로 서비스 최적화 및 마케팅 전략을 개발할 수 있다.

그러나, 이 연구는 연구방법에서 몇 가지 제한점이 존재한다. 분석에 사용된 자료가 인터넷 카페에서만 수집되었기 때문에 이 연구의 분석결과를 다양한 미디어 환경에 일반화하기에는 한계가 있을 수 있다. 또한, MBTI 유형별 데이터의 분포 불균형 문제도 제한점으로 고려되어야 한다. 이러한 제한점들은 향후 연구에서 보완이 필요하며, 그 과정에서 이 연구의 결과를 더욱 풍부하게 할 것으로 기대된다.

References

- [1] Myers, I. B, "The Myers-Briggs Type Indicator: Manual (1962)", Consulting Psychologists Press, Jan. 1962. <https://doi.org/10.1037/14404-000>.
- [2] I. G. Kim and Y. H. Lim, "A Study on the Dynamic Web Design Based on MBTI : Focusing on Users in Their 20s", The Korean Journal of Animation, Vol. 18, No. 4, pp. 70-89, Dec. 2022. <https://doi.org/10.51467/ASKO.2022.12.18.4.70>.
- [3] E. J. Choong and K. D. Varathan, "Predicting judging-perceiving of Myers-Briggs Type Indicator (MBTI) in online social forum", Peer J, Vol. 9, pp. 1-27, Jun. 2021. <https://doi.org/10.7717/peerj.11382>.
- [4] A. Noh and Y. S. Kang, "Personality Psychology",

- Hakjisa, pp. 182-183, Jun. 2018.
- [5] Y. S. R. Kim and S. J. Lee, "A Study on the Consumers' Perceptions toward Personalized Service", *The Korean Journal of Advertising*, Vol. 27, No. 7, pp. 31-57, Oct. 2016.
- [6] C. F. Surprenant and M. R. Solomon, "Predictability and Personalization in the Service Encounter", *Sage Journals*, Vol. 51, No. 2, pp. 86-96, Apr. 1987. <https://doi.org/10.1177/002224298705100207>.
- [7] C. H. Shin, Y. O. Yang, and S. L. Han, "A Study on the Effect of Perceived Personalization and Switching Barriers on Loyalty in Internet Shopping Malls", *Journal of Marketing Management Research*, Vol. 13, No. 1, pp. 23-43, Jan. 2008.
- [8] G. B. Yu, H. M. Lee, and S. K. Kim, "The impact of mobile personalized service perception on mobile shopping satisfaction and loyalty: Focusing on the moderation effect of perceived control", *The Academy of Customer Satisfaction Management*, Vol. 18, No. 1, pp. 1-21, Apr. 2016.
- [9] J.-H. Kim and Y. J. Kim, "Analysis of Research Trends in Sales Management Using Topic Modeling", *Korean Journal of Marketing*, Vol. 36, No. 1, pp. 99-120, Feb. 2021. <https://doi.org/10.15830/kjm.2021.36.1.99>.
- [10] Y. Hyun and N. Kim, "Detecting Spam Data for Securing the Reliability of Text Analysis", *The Journal of Korean Institute of Communications and Information Sciences*, Vol. 42, No. 2, pp. 493-504, Jan. 2017. <https://doi.org/10.7840/kics.2017.42.2.493>.
- [11] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation", *Journal of Machine Learning Research*, Vol. 3, pp. 993-1022, Mar. 2003. <https://dl.acm.org/doi/10.5555/944919.944937>.
- [12] D. M. Blei, "Probabilistic Topic Models", *Communications of the ACM*, Vol. 55, No. 4, pp. 77-84, Apr. 2012. <https://doi.org/10.1145/2133806.2133826>.
- [13] Y.-U. Park and K.-Y. Chung, "A Study on the Analysis of Customer's Sentiment Using DMR (Dirichlet Multinomial Regression) Topic Modeling Based on Online Review Big Data: Focusing on the Foreign Customer's Reviews of Domestic 5 Star Deluxe Hotels", *Korean Journal of Hospitality & Tourism*, Vol. 30, No. 2, pp. 1-20, Feb. 2021. <https://doi.org/10.24992/KJHT.2021.2.30.02.1>.
- [14] J. H. Park and H.-J. Oh, "Comparison of Topic Modeling Methods for Analyzing Research Trends of Archives Management in Korea: Focused on LDA and HDP", *Journal of Korean Library and Information Science Society*, Vol. 48, No. 4, pp. 235-258, Dec. 2017. <https://doi.org/10.16981/kliss.48.4.201712.235>.
- [15] H. I. Jo, J. W. Kim, and B. G. Lee, "A Study on Research Trends of Blockchain Using LDA Topic Modeling: Focusing on United States, China, and South Korea", *Journal of Digital Contents Society*, Vol. 20, No. 7, pp. 1453-1460, Jul. 2019. <https://doi.org/10.9728/dcs.2019.20.7.1453>.

저자소개

우 위 향 (Yu-Hang Wu)



2018년 3월 ~ 2020년 6월 :
부경대학교
미디어커뮤니케이션학과(학사)
2021년 3월 ~ 현재 : 부경대학교
미디어커뮤니케이션학과
석·박사과정
관심분야 : 텍스트 마이닝,
빅데이터, 뉴미디어, SNS 커뮤니케이션

남 인 용 (In-Yong Nam)



1990년 2월 : 서울대학교
언론정보학과(문학사)
1992년 2월 : 서울대학교
언론정보학과(문학석사)
1998년 2월 : 서울대학교
언론정보학과(문학박사)
1999년 3월 ~ 2003년 8월 :
신라대학교 광고홍보학과 교수
2003년 9월 ~ 현재 : 부경대학교
미디어커뮤니케이션학과 교수
관심분야 : 광고효과, SNS 커뮤니케이션, 설득전략,
미디어 효과, 도시 이미지, 정치커뮤니케이션

이 문 기 (Wen-Qi Li)



2018년 8월 :동명대학교
국제통상학과(무역학사)
2020년 8월 :동명대학교
언론영상광고학과(언론학석사)
2021년 9월 ~ 현재 : 부경대학교
미디어커뮤니케이션학과
박사과정
관심분야 : 감성 분석, 대중매체, 미디어 효과, 영화분석,
미디어커뮤니케이션