

소셜 미디어에서의 MBTI 성격 유형별 관심 주제 차이 분석

WU YUHANG*, LI WENQI**

Analyzing the Differences in Interest Topics among MBTI Personality Types in Social Media

WU YUHANG*, LI WENQI**

요 약

본 연구는 SNS에서 MBTI 성격 유형에 따른 관심 주제의 차이를 분석하여, 인간 성격의 이해와 응용에 도움이 되는 정보를 제공하고자 하였다. 데이터 수집 및 사전 처리 후, 8600개의 행 데이터를 사용하여 원형 그래프와 워드 클라우드를 이용한 시각화 분석을 진행하였다. 또한 LDA 토픽 모델링을 활용하여 각 MBTI 유형별 관심 주제를 더 깊이 파악하였다. 연구 결과, 다양한 성격 유형이 주제에 대한 관심 차이를 확인할 수 있었으며, 이를 통해 인간 성격의 이해와 응용에 도움이 될 수 있다. 이 연구는 소셜 미디어 마케팅, 팀 빌딩, 인사 자원 관리 및 심리학 등 다양한 분야에 유용한 정보를 제공할 수 있다.

Abstract

This study aimed to analyze the differences in interest topics according to MBTI personality types on SNS, in order to provide information that can help in understanding and applying human personality. After data collection and preprocessing, we conducted a visualization analysis using circular graphs and word clouds with 8,600 rows of data. Furthermore, we utilized LDA topic modeling to gain a deeper understanding of the interest topics for each MBTI type. As a result, we were able to confirm the differences in topic interests among various personality types, which can contribute to a better understanding and application of human personality. This research can provide valuable information for various fields such as social media marketing, team building, human resource management, and psychology.

Key words

Myers-Briggs Type Indicator (MBTI), Interest topics, text analysis, LDA topic modeling, visualization.

1. 서 론

인간의 성격은 개인의 행동과 선호도에 중요한 영향을 미치는 요인이며, 이에 대한 연구와 분석은 심리학, 사회학, 교육학 등 다양한 분야에서 활발하

게 이루어지고 있다. 본 연구에서는 소셜 네트워크에서 MBTI 다양한 유형의 사람들의 주제를 수집한 텍스트 데이터를 통해 MBTI 유형별로 사용자들이 관심 있는 주제를 분석하고, 이를 시각화하여 각 성격 유형별로 주제에 대한 관심 차이를 탐색한다. 이

* 부경대학교 신문방송학과, wuyuhang@naver.com

** 부경대학교 신문방송학과(교신저자), liwenqi@7745@gmail.com

를 위해, 본 연구는 데이터를 수집하여 사전 처리한 후 8600개의 행 데이터를 사용하여 분석을 진행한다.

II. 관련 연구

Myers-Briggs 유형 지시자(MBTI)는 4개의 특성이분법 쌍을 사용하여 16개의 인격 유형을 만드는 인격 유형론이다. 네 가지 쌍은 외향성-내향성(E-I), 감각-직관(S-N), 사고-감정(T-F), 판단-인식(J-P)이다. 결과적으로 각각의 네 쌍에서 이진 법으로 표현되며, 이를 결합하여 인격 유형을 만든다. Choong, En Jun, Kasturi Dewi Varathan, 2021)

MBTI 성격유형 검사는 이러한 용의 이론을 바탕으로 개인이 쉽게 응답 가능한 자기 보고 형식의 문항을 통해 인식하고 판단할 때의 선호 경향을 찾는 검사이다. 따라서 접근성이 쉬울 뿐만 아니라, 이러한 선호 경향들이 합쳐져 개인이 행동할 때 어떤 영향을 미치는지를 파악하여 실생활에서 응용할 수 있도록 제작되었다. (김인겸, 임영훈, 2002)

LDA 토픽모델링은 하나의 문서에 복수의 토픽이 내재된 상황을 전제로 하여 문서 내 여러 단어의 조합을 토대로 문서의 토픽들을 확률적으로 파악하는 통계적 분석 방법이다(Blei, D. M., 2012). 토픽모델링은 단순히 단어의 출현 빈도수를 기반으로 하는 텍스트 분석과는 다르며, 토픽이라는 변수를 통한 확률 기반의 분석을 함으로써 과적합 문제가 적고, 새로운 데이터가 투입되어도 분석 가능하다는 장점이 있다(박영욱, 정규엽, 2021).

III. 연구 방법

python 프로그래밍 언어와 nltk 라이브러리를 사용하여 텍스트 데이터를 전처리한다. 전처리 후 텍스트 내용을 네 가지 차원으로 나누어, 각각에서 단어 빈도를 계산한다. 그 다음, 각 차원에서 상위 20개의 키워드를 추출하여 후속 분석 및 시각화를 진행한다.

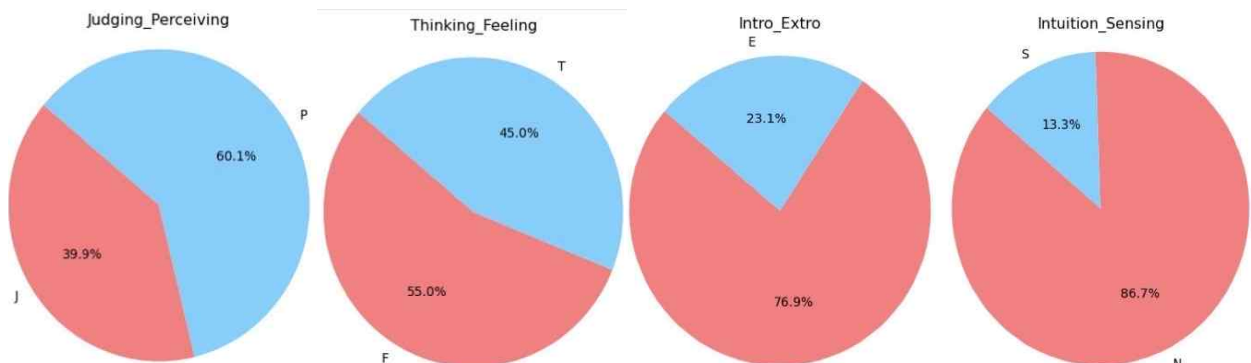
본 연구에서 사용한 데이터셋은 Myers-Briggs Type Indicator (MBTI) 테스트를 기반으로 분류된 사용자 유형을 포함한다. 이러한 사용자들의 게시물을 텍스트 분석하여 그들이 관심 있는 주제나 화제를 찾아내고, 분석 결과를 시각화한다.

IV. 분석 결과

4.1 단어 빈도 통계

네 가지 차원(내향/외향(Introversion/Extroversion), 직관/감각(Intuition/Sensing), 사고/감정(Thinking/Feeling), 판단/지각(Judging/Perceiving))의 단어 빈도 통계 결과를 분석하여 각 차원을 시각화로 표현하였다. 각 차원에서 Matplotlib 라이브러리의 pie 함수를 사용하여 원형 그래프를 그린다.(그림 1참조)

이러한 방식을 통해 각 차원의 키워드와 그 등장 횟수를 얻어, 각 차원의 특성과 의미를 더 깊이 이해하고 분석하는 데 도움이 된다.



<그림 1> 단어 빈도 통계

4.2 워드 클라우드

단어 빈도 통계가 완료된 후, 각 차원에서 상위 20개의 키워드를 추출했다. 이러한 키워드를 더 직관적으로 보여주기 위해 WordCloud 라이브러리와 matplotlib 라이브러리를 사용하여 각 차원의 워드클라우드를 생성했다.(그림 2,3 참조)워드클라우드를 통해 각 차원의 키워드 분포를 명확하게 관찰할 수 있으며, 이를 통해 각 차원의 특성과 의미를 더욱 분석하고 이해할 수 있다.

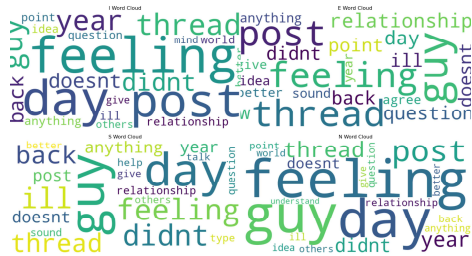
4.3 LDA 토픽 모델링 결과

각 MBTI 유형이 다양한 주제에 대해 얼마나 관심을 가지고 있는지 깊이 분석하기 위해, scikit-learn 라이브러리의 LatentDirichletAllocation을 사용하여

LDA 모델을 구축하였고, 주제의 수를 10으로 설정하였다. 그 다음, 이 모델을 문서-단어 행렬에 적용하여 각 문서의 주제 분포를 얻다. 각 사용자의 게시물에 대해 주제 분류를 진행하고, 분류 결과를 "topic"이라는 새로운 열에 기록한다.

이어서, 사용자 유형과 주제별로 데이터를 그룹화하여 각 사용자 유형이 관심 있는 주제 또는 화제의 분포를 통계로 조사한다. 이러한 데이터를 더 직관적으로 보여주기 위해, 히트맵(그림 4 참조)을 그려 주제 모델에서 각 MBTI 유형의 주제 가중치 분포를 나타낸다.

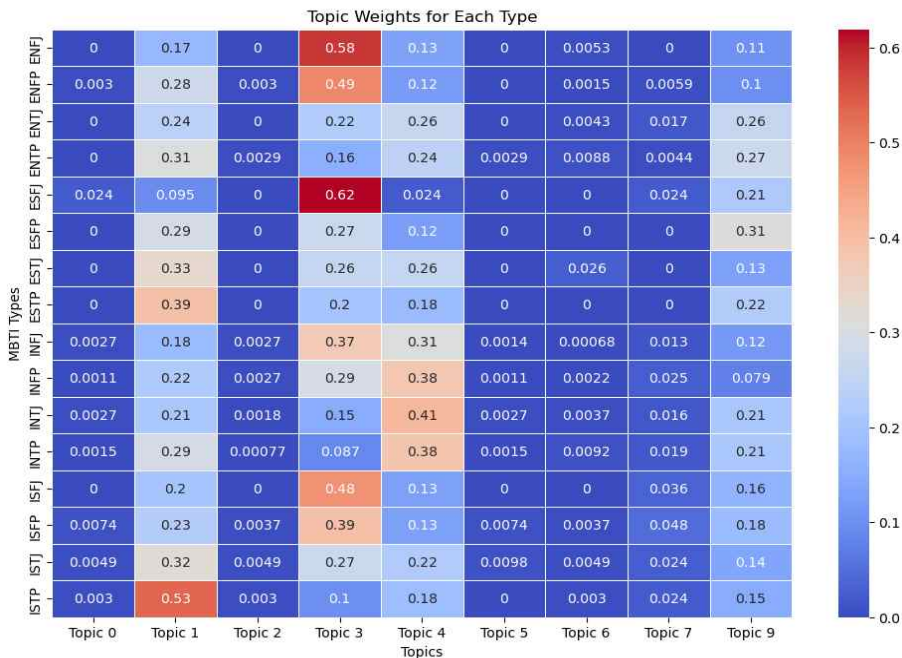
이 방법을 통해 다양한 MBTI 유형이 각 주제에 대한 관심도를 명확하게 관찰할 수 있으며, 그들의 관심사와 행동에 대한 추가 연구를 위한 강력한 근거를 제공한다.



<그림 2> 워드 클라우드



<그림 3> 워드 클라우드



<그림 4> 히트맵

V. 결론 및 논의

본 연구는 소셜 네트워크에서 수집한 텍스트 데이터를 활용하여 MBTI 유형별 사용자들의 관심 주제를 분석하고 시각화하였다. 원형 그래프, 워드 클라우드를 통해 각 인격 유형별 주제에 대한 관심 차이를 탐색하였다. 마지막으로 LDA를 사용하여 더 깊이 분석하고 9개의 사용자 토론 주제를 도출하였으며, 이를 비교하여 다양한 인격 유형이 주제에 대한 관심 차이를 이해하였다. 다양한 분야에 응용할 수 있는 가능성을 제시한다. 그러나 이 연구는 몇 가지 한계를 가지고 있습니다. 우선 사용하는 데이터셋은 인터넷 카페에서 수집되기 때문에 일반화에 한계가 있을 수 있다. 둘째, MBTI 유형의 데이터 분포 불균형도 한계다.

참 고 문 헌

- [1] Choong, En Jun, and Kasturi Dewi Varathan. "Predicting judging-perceiving of Myers-Briggs Type Indicator (MBTI) in online social forum", PeerJ 9, 2021.
- [2] 김인겸, 임영훈, "MBTI 성격유형에 따른 동적 웹 디자인 선호도 연구: 20 대 사용자를 중심으로", 애니메이션연구 18.4, pp. 70-89, 2022.
- [3] Blei, D. "Probabilistic Topic Models", Communications of the ACM, 55, pp. 77-84, 2012.
- [4] 박영욱, 정규엽, "DMR(Dirichlet Multinomial Regression) 토픽모델링을 이용한 온라인 리뷰 빅데이터 기반 고객감성 분석에 관한 연구: 국내 5성급 호텔의 외국인 이용객 리뷰를 중심으로", 호텔경영학연구 30.2, pp. 1-20, 2021.