

소셜미디어를 통한 우울 경향 이용자 담론 주제 분석*

An Analysis of the Discourse Topics of Users who Exhibit Symptoms of Depression on Social Media

서하림 (Harim Seo)**

송민 (Min Song)***

초 록

우울증은 전 세계적으로 많은 사람들이 겪고 있으며, 최근 다양한 분야에서 꾸준히 우울증에 대한 연구가 수행되고 있다. 특히 사람들이 본인의 스트레스나 감정 상태에 대해 소셜미디어에 공유한 글을 통해 그들의 심리나 정신건강에 대해 파악해보고자 하는 맥락에서 소셜미디어를 활용한 연구 역시 유의미하게 받아들여지고 있다. 이에 본 연구에서는 우울 경향의 이용자와 그렇지 않은 이용자들의 2016년부터 2019년 2월까지의 트위터 데이터를 수집하여 어떤 주제적, 어휘 사용의 특성을 보이는지 보고자 하였으며, 우울 경향의 시기별로도 어떤 차이를 보이는지 살펴보기 위해 우울 경향 관측 날짜를 기준으로 하여 이전(before) 시기와 이후(after) 시기를 구분하여 실험을 수행하였다. 토픽모델링, 동시출현 단어분석, 감성분석 방법을 통해 우울 경향과 비(非)우울 경향 이용자의 텍스트의 주제적 차이를 살펴보고, 감성 반응에 따라 사용한 어휘에 대해서도 살펴봄으로써 어떠한 특성이 있는지 확인해 보았다. 데이터 수집 단계에서 '우울' 표현을 포함한 텍스트 데이터 수집방법을 통해 비교적 긴 기간, 많은 양의 데이터를 수집할 수 있었고, 또한 우울 경향의 여부와 시기적 구분에 따른 관심 주제에 대한 차이도 확인할 수 있었다는 점에서 유의미하다고 볼 수 있다.

ABSTRACT

Depression is a serious psychological disease that is expected to afflict an increasing number of people. And studies on depression have been conducted in the context of social media because social media is a platform through which users often frankly express their emotions and often reveal their mental states. In this study, large amounts of Korean text were collected and analyzed to determine whether such data could be used to detect depression in users. This study analyzed data collected from Twitter users who had and did not have depressive tendencies between January 2016 and February 2019. The data for each user was separately analyzed before and after the appearance of depressive tendencies to see how their expression changed. In this study the data were analyzed through co-occurrence word analysis, topic modeling, and sentiment analysis. This study's automated data collection method enabled analyses of data collected over a relatively long period of time. Also it compared the textual characteristics of users with depressive tendencies to those without depressive tendencies.

키워드: 소셜미디어, 텍스트 마이닝, 트위터, 우울증, 토픽모델링, 동시출현단어 분석, 감성분석
social media, text mining, twitter, depression, topic model, co-occurrence, sentiment analysis

* 본 연구는 석사학위논문을 수정 및 요약한 것임.

** 연세대학교 문헌정보학과 석사(seohari@yonsei.ac.kr) (제1저자)

*** 연세대학교 문헌정보학과 교수(min.song@yonsei.ac.kr) (교신저자)

■ 논문접수일자: 2019년 11월 17일 ■ 최초심사일자: 2019년 12월 10일 ■ 게재확정일자: 2019년 12월 19일
■ 정보관리학회지, 36(4), 207-226, 2019. <http://dx.doi.org/10.3743/KOSIM.2019.36.4.207>

1. 서론

1.1 연구배경 및 목적

우울증은 정신질환 중에서 가장 일반적인 형태 중 하나이며, 미국에서만 일 년에 약 1,600만 명 정도가 우울증을 겪고 있는 것으로 추정된다. 그러나 우울증과 관련하여 사람들이 지적하는 가장 큰 문제점으로는 우울증 환자들이 가지고 있는 정신적 문제점의 발견과 이에 대한 적절한 치료가 제대로 이루어지지 않는 경우가 많다는 것이다. 이는 우울증과 관련한 증상들이 쉽게 방치될 수 있기 때문에, 대부분 사람들이 그들의 정신 건강문제에 대해 제대로 인식조차 못하는 경우가 많기 때문이다(Wang, 2017). 이러한 이유로 점차 사람들의 일상에서 그들의 우울증 증상을 포착해낼 수 있는 특성을 파악하고자 하는 연구가 많이 수행되고 있는데, 특히 사람들의 생각과 의견이 솔직하게 공유되는 소셜미디어 상에서 우울증을 앓고 있는 사람들을 판별해내기 위한 연구가 다수 이루어지고 있다. 그러나 우울증과 관련하여 지금까지 진행된 연구들은 대체적으로 연구 대상이 되는 우울증 사람들의 정보를 충분히 확보하기 쉽지 않다는 문제와 영어 데이터 위주로 진행되었다는 한계점이 있다. 따라서 본 연구에서는 한글 데이터를 대상으로 토픽모델링, 동시출현단어 분석, 감성분석의 기법을 통해 우울 경향을 보이는 사람들의 트위터에서 공유한 텍스트 데이터를 기반으로 그들의 성향과 관심 주제를 살펴보고, 이를 통해 온라인 공간에서 우울 증세에 영향을 미칠 수 있는 주제에 대해 파악해보고자 하였다. 또한 향후 우울증 증상 판단

에 있어서 소셜미디어 데이터를 활용하여 유의미한 결과를 도출해 낼 수 있다는 새로운 가능성을 보고자 하였다.

본 연구에서 설정한 연구 질문은 다음과 같다.

- 연구질문 1. 트위터 상에서 우울 경향을 보이는 이용자와 그렇지 않은 이용자들 간 관심있는 주제에 대해 어떠한 차이가 있는가?
- 연구질문 2. 트위터 상에서 보이는 우울 경향 이용자들의 특성이 특정 시기에 따라 어떠한 차이점을 보이는가?
- 연구질문 3. 트위터에서 관찰된 우울 경향의 이용자들은 어떠한 감성표현의 특성을 보이는가?

1.2 연구의 범위 및 배경

본 연구에서 우울 경향 이용자들의 트위터 텍스트 분석을 위해 사용한 여러 방법 중 토픽 모델링(Topic Modeling)은 구조화되지 않은 거대한 문헌 집단에서 주제를 찾아내기 위한 알고리즘 중 하나로, 대량의 데이터를 분석할 수 있고 패턴을 찾아낼 수 있다는 특징을 가지고 있다. 또한 동시출현 단어들을 통해 우울 경향을 보이는 이용자와 그렇지 않은 이용자 간 소셜미디어에서의 담론 내용에 어떤 차이를 보이는지 살펴보고자 하였다. 그리고 수집한 우울 경향 이용자들의 트위터 데이터에 대해 시기별 차이에 대해서도 보기 위하여 먼저 스스로 본인의 우울증에 대해 명확히 언급한 표현이 포함되어 있는 트위터 계정을 대상으로 한 다음, 우울증 진단 표현이 포함된 트위터를 기

준으로 해당 월을 decision 구간으로 설정하고, 이전에 해당하는 기간을 before 구간, 이후 기간에 대해서는 after 구간으로 설정하였다. 예를 들어 2018년 7월에 우울 경향 진단 단서와 관련된 메시지를 언급한 트위터 이용자가 있다면, 해당 이용자 계정을 타겟으로 하여 이 계정의 트위터 중 2018년 7월 한 달 동안 게시한 트위터 데이터가 decision 구간에 해당하며 2016년 1월부터 2018년 6월까지의 데이터가 before, 2018년 8월 이후부터의 데이터가 after 구간에 속하게 된다. 보다 쉬운 이해를 위하여 우울 경향 데이터에 대한 특정 기간 구분에 대한 내용은 다음 <그림 1>과 같이 나타내었다.

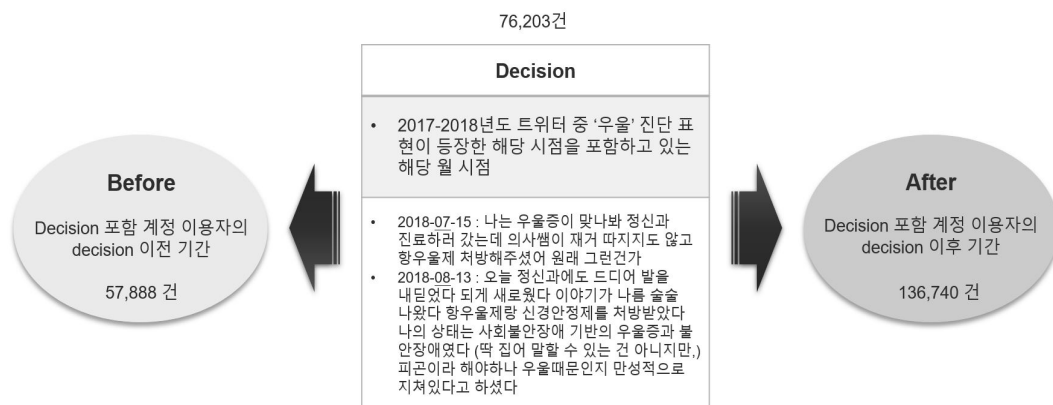
이와 같이 기간별로 구분하는 과정을 통해 우울 경향 사용자들의 시간의 흐름에 따른 특성 및 양상의 변화를 살펴보고자 하였으며 수집한 데이터들은 전처리 과정을 거친 뒤, 토픽 모델링 등을 통해 우울 경향 이용자가 소셜미디어 상에서 보이는 특성에 대해 감성 값이나 어휘표현을 살펴보았다.

2. 이론적 배경

2.1 소셜미디어에서의 우울증 연구

소셜 미디어를 통해 정신건강에 대해 파악하고자 한 연구들은 소셜미디어에서의 언어 사용, 사회적 표현, 그리고 정신건강의 척도가 될 수 있는 사람들과의 상호작용의 모습을 통해 연구의 가능성을 보여준다. 또한, 소셜미디어는 온라인상에서 말 그대로 사회의 속성인 사회적 패턴과 정신건강 및 질환의 부분까지 볼 수 있다는 특징이 있다. 그러므로 트위터와 기타 소셜미디어는 사람들의 행동에 대해 정량화할 수 있는 고유한 관점을 제공할 수 있다는 점에서 정신건강과 관련된 연구자들에게 유용한 도구로 활용되고 있다(Coppersmith, Dredze, & Harman, 2014).

그 중, 우울증에 관한 연구는 전 세계 인구를 대상으로 흔하게 볼 수 있는 정신질환 중 하나이며, 또한 앞으로 점차 우울증을 앓는 사람들의 수가 증가할 것이고 우울증이 초래하는 다



<그림 1> 우울 경향 이용자 대상 시기별 데이터 구분 설정

른 증상들로 상황이 더 심각해진다는 예측이 제시되면서 다양한 분야에서 꾸준히 연구되고 있다. De Choudhury, Gamon, Counts, & Horvitz (2013)의 연구에서는 대중들의 참여를 통해 솔루션을 얻는 방법인 크라우드 소싱(Crowdsourcing) 방법(Snow, O'Connor, Jurafsky, & Ng, 2008)을 통해 우울증 진단을 받은 이용자들을 모집하여, 1년 여 간 그들의 트위터 글을 분석하여 행동적 특징을 분석했다. 또한 우울증을 지닌 이용자와 그렇지 않은 이용자 그룹의 차이점을 통해 우울증 예측 프레임워크를 위한 자질 벡터(feature vector)도 개발하였는데, 이러한 실험을 통해 소셜미디어 활동이 개인의 우울증 증상과 진단을 예측하고 판단하는데 유용한 근거가 될 수 있다는 의미를 지니고 있다. 이와 같이 소셜미디어에서 드러나는 우울증 이용자들의 특성을 밝혀내고 이를 발전시켜 향후 우울증 진단 시 판단의 근거로 사용할 수 있는지 그 가능성을 보여주는 연구들은 많이 있다. 국내에서 소셜미디어를 통해 우울증에 대해 살펴보고자 진행한 사례로는 송호윤, 박한철, 양원석, 박종철(2017)의 연구가 있다. 그들은 대부분 기존연구에서 텍스트 상에서 포착되는 우울증 관련 증상과 수반되는 감정에 관련된 어휘 "자살, 무기력, 우울감" 등을 중심으로 사용자의 우울증 여부를 예측한 시도들에 해 분석하였다. 그 내용으로는 비정형 데이터인 소셜미디어 텍스트의 특성과 보통 신조어이거나 사전 미등록 어휘로 나타날 가능성이 많은 우울증 및 감정 관련 어휘를 대상으로는 우울증 관련 어휘를 모두 포착하기는 어렵다는 한계를 지적하고 있다. 이에 어휘 기반 예측 모델이 지닌 한계점을 해결하기 위해, 문헌적 고찰과 관찰을 토대

로 우울증을 겪고 있는 사용자에게서 나타나는 언어학적 패턴을 제시하였다. 그리고 이를 적용한 어휘 모델을 구축함으로써 소셜미디어 사용자의 우울증 예측에 있어서 언어학적 패턴 사용의 효과를 검증하였다.

2.2 동시출현단어 분석과 토픽모델링

동시출현단어 분석은 한 문헌에서 두 키워드가 동시에 출현하였을 때 주제적으로 연관이 있다고 보고, 동시출현단어 빈도를 이용하여 단어 간의 관계를 분석하는 기법이다. 즉, 같은 문맥 안에서 자주 함께 등장한 단어는 서로 밀접한 관계가 있다고 볼 수 있는 것이다. 이러한 동시출현단어 분석 기법은 다양한 분야에서 널리 사용되고 있다. 특히 소셜미디어 데이터를 활용하여 동시출현단어와 관련된 실험을 수행한 연구를 살펴보자면, Jiang과 Yang(2013)은 점차 증가하고 있는 온라인 상에서의 건강 관련 소셜 웹사이트를 대상으로 하여 이용자들의 건강 관련 어휘에 대하여 분석하였다. 웹 공간에서의 건강 관련 소셜미디어 사이트는 이용자들이 건강과 관련된 정보를 찾아보고, 직접 정보를 생산하며 더 많은 건강 정보를 공유할 수 있는 공간을 제공한다. 그러나 이때 이용자들은 의학 또는 약물과 관련하여 전문적인 지식이 없기 때문에 동일한 질환이나 증상에 대해서도 여러 표현을 사용한다는 문제점을 보인다. 이에 해당 연구에서는 동시출현단어 분석을 통하여 이용자가 직접 제공하고 사용한 건강 관련 표현을 확인함으로써 고객이 사용하는 건강 관련 어휘의 범위를 확장하고자 하였다. 그리고 국내 연구에서는 트위터를 대상으로 하여,

특정 상품명을 포함하는 트위터 데이터를 수집하고, 이를 활용하여 하나의 트위터 게시글에서 동시에 출현한 단어의 쌍을 이용하여 동시출현단어 분석을 실시한 내용이 있다. 진설아, 허고은, 정유경, 송민(2016)은 이러한 동시출현단어 분석 결과를 통해, 트위터 상에서 토픽 키워드와 토픽이 변화하는 시점과 패턴을 분석하고자 했다.

DMR(Dirichlet-multinomial Regression) 모델은, 문헌과 주제 분포를 기반으로 저자, 발행 날짜, 저널명 등 문헌의 메타데이터 특성을 제3의 파라미터로 설정하여 결과를 도출하는 문헌 집합 내에서 주제를 발견하는 토픽 모델이다(Mimno & McCallum, 2008). 이러한 토픽모델링을 적용한 연구는 다양한 분야에서 이루어지고 있다.

특히 소셜미디어와 접목하여 토픽모델링을 활용한 연구 사례로는 소셜미디어 중 한 가지 형태인 트위터와 New York Times라는 전통적인 뉴스 매체를 이용하여 토픽모델링 분석을 통해 주제 분야와 유형에 대해 분석한 연구가 있다(Zhao, Jiang, Weng, He, Lim, Yan, & Li, 2011). 그들의 연구에서 이러한 토픽 분석을 통해 유용한 정보매개체로서의 트위터의 역할에 대해 보여주었다. 또한, 국내에서 트위터를 대상으로 동시출현 단어 검색, 트렌드 변화에 대한 토픽모델링, 그리고 트위터 멘션(mention) 기반 이용자 네트워크 분석등의 기능을 통해 2012년 한국 대선을 대상 사례로 선정하여 트위터에서 생성되는 사회적 이슈 동향을 추적할 수 있게 하는 트렌드 마이닝 시스템을 개발한 연구가 있다(배정환, 손지은, 송민, 2013). 특히 이 연구에서는 사회적 이슈의 출현과 변화를

반영하는 트위터 이용자 네트워크 추출과 그 특성을 밝히기 위해 이용자들이 작성한 트윗(멘션)과 그 트윗의 방향성에 대한 분석을 위해 DMR을 사용하였다.

2.3 감성 분석

텍스트 데이터에 포함되어있는 작성자의 감성이나 태도, 의견을 파악하는 감성 분석(Sentiment Analysis)은 텍스트마이닝, 기계학습, 자연언어 처리 등에서 많은 양의 비정형 데이터를 분석하는 기법이다. 감성 분석의 대상이 되는 분야로는 상품에 대한 만족도 및 태도 조사부터 사회적, 정치적 대상에 대한 여론 조사까지 다양하게 활용되고 있다(Wang, Can, Kazemzadeh, Bar, & Narayanan, 2012; 서상현, 김준태, 2016). 최근 감성 분석에 대한 연구는 기계학습(machine learning)을 기반으로 진행되고 있으며, 크게 지도학습 방법(supervised learning)과 비지도 학습(unsupervised learning) 접근방법으로 구분되고, 첫 번째 방법인 지도학습 방법에는 나이브 베이즈 분류기, 최대 엔트로피, 서포트벡터 분류기, 의사결정 나무, 인공신경망 등이 있다(Pang, Lee, & Vaithyanathan, 2002).

세부적인 주제의 감성 분석 연구 예시로는 영화 리뷰(이상훈, 최정, 김종우, 2016)나 상품 후기(장재영, 2009), 경제 뉴스(이현서, 송민, 2018)와 같이 여러 분야에서 다양한 주제를 대상으로 연구되어 왔다. 그 중 소셜미디어에서 우울증과 관련된 감성 분석 연구로는 트위터로부터 수집한 우울증과 관련된 텍스트 데이터에서 감정을 전달하는 단어들의 빈도에 따라 우

울증에 대한 예측이 가능한지 보고자 한 연구가 있다(De Choudhury, 2013). 그리고 감성분석을 통해 우울한 이용자들은 그렇지 않은 이용자에 비해 부정적인 감정 어휘들을 더 많이 사용하고 화를 잘 표현하는 경향을 보인다는 것을 밝히고, 게다가 우울한 이용자들은 그들의 우울증 치료 이력등과 같은 개인 정보를 소셜미디어에 기록해두기도 했다는 것을 보여준 연구도 있다(Park, Cha, & Cha, 2012).

3. 연구 방법

3.1 데이터 수집

Coppersmith, Dredze와 Harman(2014)의 연구에서는 많은 양의 트위터 데이터로부터, 스스로 정신질환을 밝힌(self-expression of mental illness diagnoses) 데이터들을 찾아내는 방법을 시도하였다. 이러한 방법은 단순히 기존 설문 기반의 데이터 수집 방법을 대체하는 것이 아닌, 상호보완의 역할을 할 수 있는 데이터 수집 방법으로서, 통계 분석 등을 위해 충분한 양의 데이터를 수집하는 데 큰 역할을 할 수 있다는 점에서 의미가 있다. 따라서 본 연구에서도 위와 같은 수집방법을 사용하여 “나 우울증이래”, “우울증 진단”, “우울증 처방” 등과 같이 명확한 진단 근거가 있는 표현을 질의어로 하여 총 2609명의 1,679,287건의 트위터 데이터를 수집하였다. 또한, 우울 경향을 보이지 않는 이용자에 해당하는 트위터를 수집하기 위하여 특정 표현을 포함하고 있는 우울 경향 이용자 데이터 수집 방법과는 달리, 보다 일반적인 질

의어를 사용하여 수집을 진행하였다. 예를 들어 텍스트 성격을 한정지을 수 있는 주제 단어가 아닌 문장에서 반드시 등장하는 “은/ 는/ 이/ 가”와 같은 조사를 수집 키워드로 사용하여 일상적인 텍스트 데이터들을 수집할 수 있도록 하였다. 조건을 만족하는 광범위한 이용자의 트위터 데이터를 우울 경향 이용자 데이터와 유사한 조건으로 맞추기 위하여, 같은 시간 조건(2016년부터 2019년 2월)과 수집하고자 한 트위터 이용자 수는 3,000명으로 조건을 주고, 총 1,447,302건의 텍스트 데이터를 수집하였다.

3.2 연구 절차

3.2.1 연구 방법

수집한 트위터 데이터에 대해서 우울 경향 이용자와 그렇지 않은 이용자로 구분하고 또한 우울 경향 이용자 그룹에서 시기에 따른 기준으로 구분하여 정제한 데이터에 대해서 먼저 전처리를 수행하였다. 그 후 본 연구에서는 수집한 우울 경향 및 그렇지 않은 이용자들의 트위터 데이터를 대상으로 실험을 위해, DMR(Dirichlet Multinomial Regression)토픽모델링을 수행하였다. 하나의 문헌의 주제 분포와 주제의 단어 분포가 다리클레 분포를 따른다고 가정하고, 실제 관측되는 문헌의 단어로부터 문헌별 주제 분포와 주제별 단어분포를 추론해내는 LDA(Latent Dirichlet Allocation)과는 달리 DMR에서는 문헌에 할당된 명목변수를 활용하여 주제별 단어분포를 파악할 수 있는 특징이 있기 때문에 본 연구에서는 시간에 따른 주제 분포 변화를 분석하기 위하여 DMR모

형을 사용하였다(Mimno & McCallum, 2008).

또한, 동시출현단어 분석도 실시하여 우울 경향 이용자와 그렇지 않은 이용자 간의 트위터에서 사용된 단어에 대한 전반적인 동향을 비교하고자 하였고, 우울 경향 이용자 중에서 기간별로 그룹을 나누어, 우울 경향 관측 시기에 따라 어떠한 양상을 보이는지 분석해보았다. 전체적인 연구 방법 및 과정에 대해서는 다음 <그림 2>로 나타내었다.

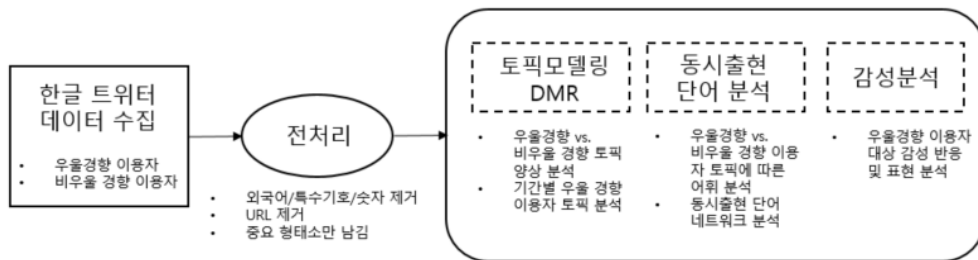
3.2.2 전처리 과정

본 연구에서 분석의 대상이 되는 텍스트들은 모두 전처리 과정을 거쳤다. 먼저 트위터의 특성상, 외국어나 특수기호 그리고 URL 등의 텍스트 형태가 많기 때문에 이에 해당하는 내용부터 제거하는 단계를 거쳤다. 그 후 형태소 분석은 코모란(Komoran)을 사용하였는데, 코모란은 Shineware의 기술로 연구되고 개발된 한국어 형태소 분석기로써 java library(jar) 형태로 제공된다. 발전된 형태인 Komoran 3.0은 기존 버전보다 속도와 정확도가 개선되었으며,

타 형태소 분석기와는 달리 여러 어절을 하나의 품사로 분석을 가능하게 함으로써 적용 분야에 따라 더 정확하게 분석할 수 있다는 장점이 있다(SHINEWARE, 2013). 이러한 이유로 코모란을 사용하여 형태소 분석을 수행하였으며 조사나 어미만을 제외하고 그 외의 품사들에 해당하는 단어들만 남기는 과정을 거쳤다. 또한 이후에 형태소의 문맥과 의미 등을 고려하면서 연구에 불필요하다고 생각되는 불용어들도 제거하여 최종 실험대상 데이터를 구축하였다.

3.2.3 동시출현단어 분석

우울 경향 이용자와 그렇지 않은 이용자들이 트위터 상에서 어떠한 주제의 이야기를 나누는지 전체적인 주제를 파악해보고, 세부 양상에 대해서도 알아보하고자 각 이용자 그룹을 대상으로 동시출현단어 분석을 실시하였다. 동시출현 단어는 두 단어가 같은 맥락에서 함께 사용된 것이므로 이들끼리 매우 밀접한 관계로 의미를 지니고 있다고 할 수 있다.



분석한 내용을 기반으로 우울 경향 이용자의 관심 주제 분석
및 향후 연구에서의 활용 가능성 확인

<그림 2> 연구 방법 및 과정

본 연구에서는 먼저 전처리 과정을 거친 각 이용자 그룹별 텍스트를 대상으로 하여 동시출현 빈도를 기반으로 단어 쌍의 출현 빈도를 계산하고 그 결과로 동시출현 단어를 구하였다. 이 때, 동시출현한 횟수가 많을수록, 함께 등장한 단어들의 관계성이 높다는 가정하에 일정한 기준(threshold)을 주고 그 기준을 충족하는 단어 쌍만 남기고 남은 단어 쌍들은 필터링하였다. 또한 동시출현 단어 분석의 특징을 활용하여 추출된 단어 쌍들은 Gephi(Bastian, Heymann, & Jacomy, 2009) 0.9.2 판을 활용하여 시각화를 하였다.

3.2.4 토픽모델링

우울 경향 이용자와 그렇지 않은 이용자의 텍스트를 바탕으로, 각각 해당하는 이용자들이 어떤 주제에 대해 소셜미디어 공간에서 이야기를 나누며 또한, 이들의 트위터 데이터에서 확인할 수 있는 텍스트 특성의 차이점을 살펴보기 위하여 토픽모델링을 수행하였다. 보다 세밀한 분석을 위하여 토픽모델링 실험을 위해 '비우울 경향 이용자'와 '우울 경향 이용자'에서도 시기별로 구분한 'before 우울', 'decision 우울', 'after 우울'의 총 네 그룹을 대상으로 하여 DMR 실험을 수행하였다. LDA 토픽모델링과는 달리 DMR 토픽모델링의 경우 주제에 대한 사전분포인 알파(α) 즉, 토픽 결과로 문헌에 들어가는 사전 확률(prior probability) 값은 관찰된 문헌의 특성(feature)으로 간주하므로, 문헌의 고유한 메타데이터 특성 값에 바뀔 수 있다(송민, 2017). 이러한 DMR 토픽모델링의 특성을 고려하여, 주제의 갯수와 기타 하이퍼파라미터들의 조정을 통해 가장 적절하게 결과

가 나오는 값인 $\sigma=1$, $\beta=0.01$, $\varepsilon=0.005$ 일 때의 주제 20개 결과를 기준으로 하였다.

3.2.5 감성 분석

본 연구에서 우울 경향을 보이는 이용자와 그렇지 않은 이용자의 트위터 데이터를 대상으로 감성 분석을 실시해 보고, 이후 실험 결과로 나오는 긍정과 부정에 해당하는 표현에서 어떠한 맥락에서 감성을 보여주는 표현을 하였는지 세부적으로 파악해 보고자 하였다. 특히 본 연구에서 감성 분석 실험을 위해 감성 사전(sentiment lexicon)을 사용하였는데, 기존에 사용되던 한글 감성어 사전을 바로 사용하기에는 분석 대상 데이터의 성격이 고려되지 않아 감성 분석의 결과가 제대로 나오지 않았다.

즉, 분석 대상 텍스트로부터 감성을 명확하게 추출하기 위해서는 언어의 사용 과정에서 발생하는 다양한 문맥적 의미와 새롭게 등장한 표현 등이 고려되어야 할 필요가 있다(유은지, 김유신, 김남규, 정승렬, 2013). 따라서 우울 경향 이용자의 상황을 고려하여 더욱 정교한 감성 사전을 구축하는 작업을 진행하였다. 이를 통해, 보다 정확도가 향상된 감성평가 결과를 얻고자 하였다.

먼저 우울 경향 이용자 데이터를 대상으로 하여 전처리를 수행한 결과를 바탕으로 긍정과 부정단어 목록을 만들었다. 이 때, 부정을 구성하는 단어들을 살펴보면 기존에 다른 연구에서도 활용되던 한글 부정어뿐 아니라, 사회적 배경과 맥락적 의미로 부정적인 의미를 지닌 새로운 표현들이 많이 등장했기 때문에 수동 접근방법(manual approach)으로 사전을 구축하였다(김재봉, 김형중, 2017). 다만 데이터를 확

인하여 담론의 맥락 속에서 가지는 의미를 기반으로 구축하였기에 객관성이나 타당성에 있어서 한계점이 있을 수 있으나, 사회적 배경과 시기적 문화가 반영된 단어도 고려되었다는 점에서 유의미하다고 본다. 해당 긍/부정 단어 감성사전은 이후 부록에 첨부하였다. 그리고 이를 기반으로 하여 감성 값을 도출하기 위한 방법으로, 한 문장에서 긍정의 단어가 등장하는 횟수만큼 +1, 부정의 단어가 등장할 때는 -1을 계산할 수 있도록 하여 최종 값을 해당 트위터 문장의 감성값으로 설정하였다. 우울 경향 이용자에 대해 우울 경향 기간에 따른 긍정/부정 감성 분석 실험 과정을 거친 후에는, 도출한 결과를 바탕으로 동시출현 단어 네트워크를 그려보았다. 앞서 실행한 동시출현단어 분석 네트워크를 만들 때와 마찬가지로 Giant Component를 구하고 모듈러리티 클래스(Modularity class)를 기준으로 나타내었다.

4. 분석 결과

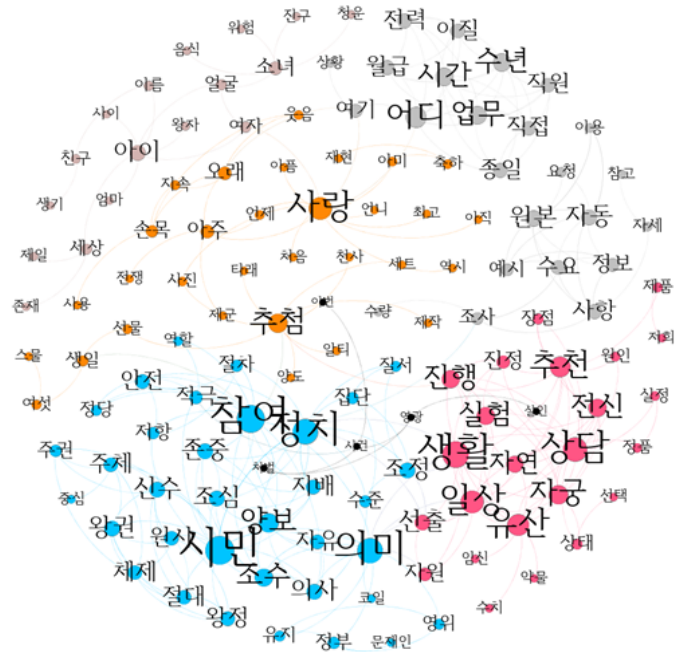
4.1 동시출현단어 분석

본 연구에서는 우선 우울 경향 이용자와 그렇지 않은 이용자가 트위터에서 어떤 단어들을 중심으로 이야기하는지 전반적인 동향을 살펴보기 위해 각 그룹별 트위터를 대상으로 동시출현 단어 분석을 수행하였다. 분석 대상은 각 데이터 집합에서 동시출현 빈도횟수를 기준으로 하여 상위 1,000개의 단어 쌍을 구하였다. 이를 바탕으로 시각화를 통해 나타내었는데, 그래프 상에서 우울 경향을 보이지 않는 이용자의 네트워

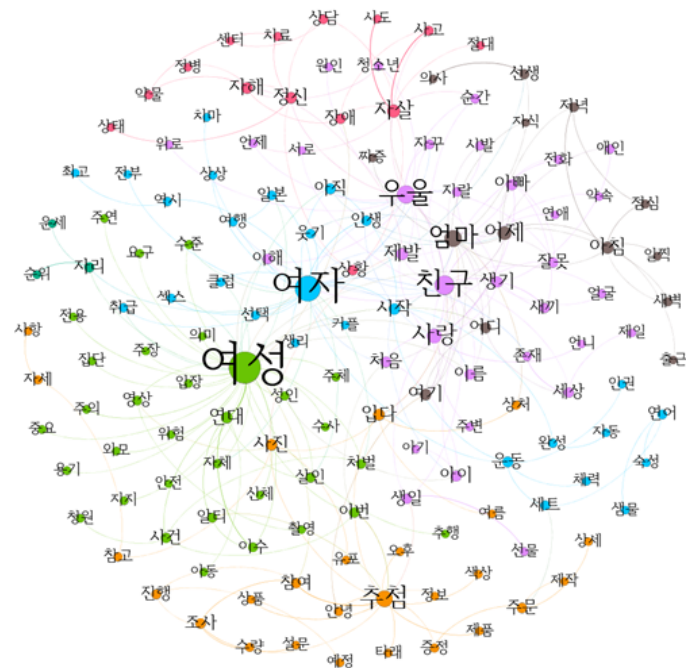
크의 경우 139개의 노드와 312개의 에지를 가지고 있으며, 우울경향 이용자의 네트워크에서는 151개의 노드와 249개의 에지를 포함한다. 우울 경향을 보이지 않는 이용자의 동시출현 네트워크를 나타내면 <그림 3>과 같다.

비우울 경향 이용자의 데이터에서는 적은 비중의 다양한 주제 단어들인 다수의 모듈러리티를 구성하고 있으므로, 좀 더 효과적인 시각화를 위해 일정 기준에 충족하는 클래스들만 그래프로 나타내었다. 그 결과, 그래프를 통해 해당 이용자의 네트워크에서 확인할 수 있는 전반적인 군집들의 내용들은 “정치”, “시민”, “참여”, “생활” 등과 같이 보다 일반적이고 매우 다양한 주제들이 고르게 분포되어 있는 모습을 볼 수 있다.

우울 경향 이용자를 대상으로 동시출현 네트워크를 나타낸 결과, <그림 4>와 같은 모습을 볼 수 있었다. 먼저, 가장 눈에 띄는 키워드 군집에는 “여성”, “여자” 등이 있는데, 동시출현한 다른 단어와의 관계와 전체적인 문맥을 고려했을 때 최근 사회이슈로 떠오른 여성 인권 문제나, 여성을 대상으로 한 성범죄 등의 요인으로 인해 언급된 경우가 많은 편인 것으로 보여진다. 또한 “우울”과 관련된 키워드와 동시출현한 단어들의 군집도 크게 형성되어 있는데, 연관성이 높은 다른 단어들을 살펴보면 “엄마”, “친구”, “사랑” 등 주변 인간관계와 관련된 단어들이 크게 형성되어 있었다. 이와 같이 우울 경향을 보이지 않는 이용자와 비교했을 때 우울 경향 이용자의 데이터에서 인간관계와 관련된 단어들이 더 비중 있게 등장한 이유로는, 우울한 사람들은 우울하지 않은 사람들에 비해, 대인관계가 비교적 특징적인 패턴을



〈그림 3〉 비우울 경향 이용자 동시출현 단어 네트워크



〈그림 4〉 우울 경향 이용자 동시출현 단어 네트워크

보이고, 종종 부부간이나 부모-자녀 간의 갈등 등의 모습으로 나타나기도 한다는 연구 결과(Hammen & Brennan, 2002)와 같이 주변 인간관계에서 우울증의 원인이 되거나 의지할 수 있는 존재가 있을 수도 있기 때문에 더 쉽게 영향을 받기 때문으로 보여진다. 그리고 “자살”, “자해”와 같은 단어들 역시 우울증과 관련된 증상과 고민 중 하나로 연관성이 높은 군집으로 보인다.

4.2 DMR 토픽모델링

우울 경향을 보이는 이용자와 우울 경향을 보이지 않는 이용자를 기간별로 구분한 그룹(before/ decision/ after)을 대상으로 DMR을 수행하여 분류된 토픽들은 자동으로 묶여진 그룹 내에서도 비슷한 성격의 단어들을 데이터를 직접 확인하여 대표할 수 있는 주제명으로 선정하여 작성하였고 그 결과는 <표 1>과 같다. 다만, 하나의 주제 단어로 대표 주제명을 선정하기 모호한 경우에는 등장한 단어들을 복수 제시하여 주제명을 지정하였다. 이 때, 주제 항목마다 normal(비우울)/ before/ decision/ after 등의 그룹에 따라 값에 어느 정도 차이를 가지고 있는 것을 볼 수 있는데, 먼저 우울 경향을 보이지 않는 이용자들은 <그림 5>와 같이 ‘기록(글, 사진, 그림)’, ‘느낌’, ‘표정/행동’, ‘방송/연기’ 등의 주제에 더 높은 반응을 보였다. 이와 반대로 우울 경향 이용자들은 <그림 6>에서와 같이 ‘젠더’, ‘성범죄’, ‘외모/패션’, ‘우울’이라는 주제에 비우울 경향 이용자와 비교하여 전반적으로 더 큰 반응을 보이며 많은 담화가 이루어진 것으로 보여진다.

4.2.1 젠더이슈 및 성범죄 관련 우울 경향 이용자 분석

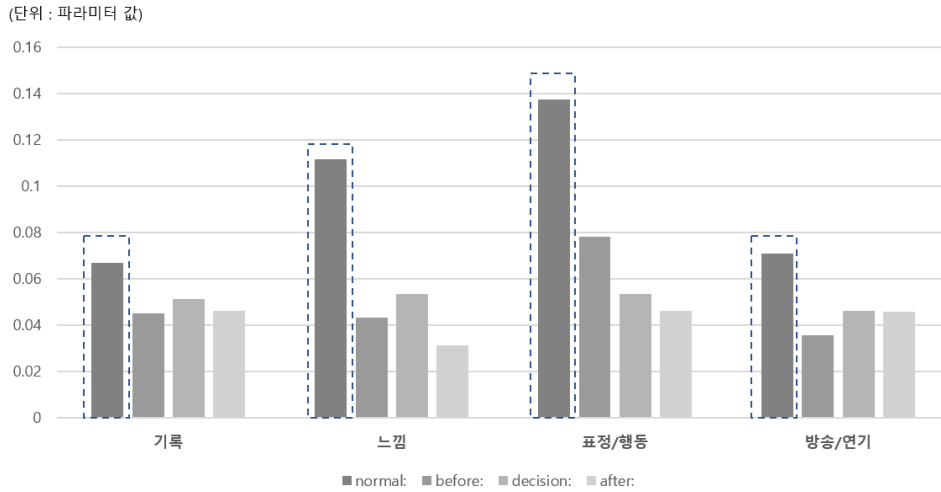
우울 경향 이용자를 대상으로 동시출현 단어 분석 및 토픽모델링을 수행한 결과, 우울 경향을 보이지 않는 이용자와 비교하여 눈에 띄는 특징으로는 우울 경향을 보이는 사람들이 ‘여성’과 관련된 젠더이슈 및 ‘성범죄’와 관련된 주제에 더 많은 관심과 언급을 보인다는 점이다. 그러나 데이터 수집 대상 기간이 시기적으로 당시 사회 전체적으로 성범죄 이슈 및 페미니즘 논란으로 인해 관련 담론이 많이 발생하였다. 이에 우울 경향과 비우울 경향 이용자들의 데이터 모두에서 이와 관련된 내용이 토픽이 많이 발생한 것에 영향이 있었던 것으로 추측된다. 그러나 실험대상의 우울 경향과 비우울 경향 데이터를 비교했을 때 우울 경향 이용자 데이터에서 비교적 더 많은 양과 주제의 담화가 이루어지는 것이 포착되므로 이와 관련된 내용을 더욱 심층적으로 분석해보았다.

그동안의 연구에 따르면, 우울증은 성별과 밀접한 관련성을 보이는 특성이 있다. 이와 같은 여성과 우울증이 연관되어 등장하는 우울증 담론이 2007년 이후로 폭증하였으며 ‘여성’이 이슈가 되고 문제화하는 사회적 분위기가 오히려 또다시 우울한 여성들을 만들어내는 요인이 되기도 한다는 연구 결과도 있다(이유림, 2016).

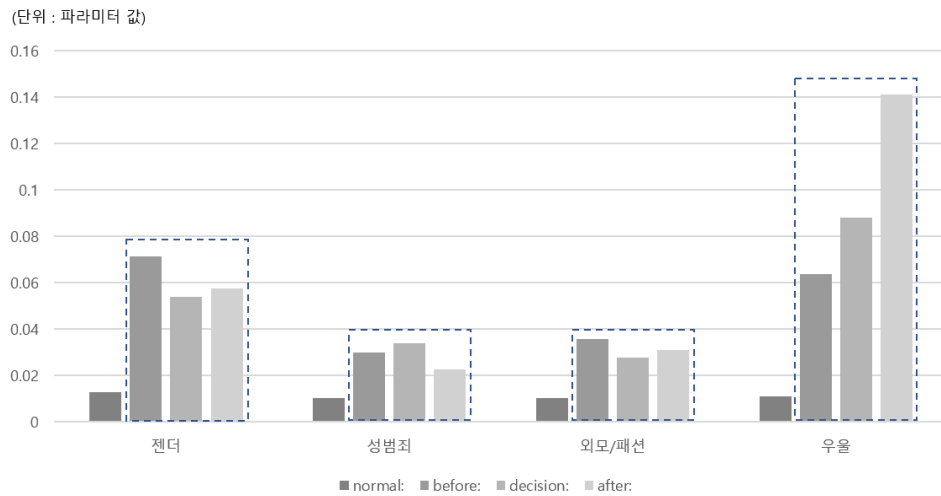
이와 같이, 성별과 관련하여 우울증은 강한 연관성을 보인다는 것을 선행연구들을 통해 알 수 있었다. 본 연구에서는 우울 경향 이용자의 트위터 데이터를 월별로 나누어 토픽모델링을 수행한 후 몇 가지 특정 토픽 - ‘여성/젠더’, ‘혐오(남성)’, ‘성범죄’, ‘우울증’ 등에 대해 어떤

〈표 1〉 우울 & 비우울 경향 이용자 트위터 DMR 수행 결과

주제	주제명	주제 단어
1	반려동물/아이	고양이, 꽃, 눈, 귀엽다, 동물, 아이, 예쁘다, 이름, 사랑, 찾다, 하늘, 길, 강아지, 사진, 별, 봄, 빛, 키우다, 물, 날, 바다, 좋아하다, 세상, 작다, 아름답다
2	기록(글/그림/사진)	트다, 그리다, 그림, 마음, 쓰다, 캐다, 올리다, 드리다, 글, 사진, 당, 친하다, 남기다, 메다, 탐라, 감사, 맛, 흔적, 찍다, 찾아가다
3	학교	학교, 공부, 대학, 반, 선생, 학생, 시험, 친구, 다니다, 영어, 수업, 교수, 학년, 배우다, 듣다, 학원, 쓰다, 졸업, 수학, 싫다, 읽다, 고등학교, 수능, 싸다, 책
4	일상	하루, 내일, 일어나다, 잠, 먹다, 밤, 아침, 날, 꿈, 나가다, 깨다, 어제, 쉬다, 보내다, 새벽, 친구, 놀다, 꾸다, 일찍, 늦다, 만나다, 출다, 술, 기분, 피곤
5	젠더	여자, 남자, 여성, 성, 남성, 쓰다, 혐오, 한국, 당하다, 결혼, 아이, 차별, 폭력, 싫다, 페미니즘, 엄마, 맛, 글, 딸, 본인, 듣다, 섹스, 낳다, 남편, 이야기
6	성범죄	사건, 피해자, 여성, 성, 경찰, 폭행, 성폭행, 가해자, 당하다, 공론, 살인, 처벌, 범죄, 신고, 성범죄, 성폭력, 촬영, 연대, 피해, 불법, 이수, 남성, 기사, 여자, 부탁
7	외모/패션	입다, 옷, 스다, 사다, 트다, 추천, 신상, 경험, 예쁘다, 피터, 블다, 공유, 토니, 주문, 바지, 이벤트, 사이즈, 프리, 판매, 드리다, 디자인, 가격
8	국가	한국, 일본, 서울, 여행, 선수, 미국, 부산, 중국, 세계, 나라, 한국인, 독일, 올림픽, 팀, 러시아, 지역, 제주, 경기, 프랑스, 대전, 도시, 이기다, 축구
9	종교	성, 가능, 세계, 필요, 가장, 능력, 모든, 한국, 높다, 불, 나라, 신, 존재, 중요, 다르다, 시대, 이해, 역사, 종교, 의미, 이야기, 사용, 개인
10	감정 - 느낌	귀엽다, 사랑, 스다, 캐다, 후, 미치다, 예쁘다, 쓰다, 울다, 아악, 야, 웃기다, 있다, 음, 이쁘다
11	감정 - 생각	사랑, 당신, 마음, 죽다, 버리다, 행복, 좋아하다, 느끼다, 힘들다, 모든, 세상, 듣다, 감정, 친구, 만나다, 삶, 기억, 이야기, 순간, 싫다, 믿다, 바라다
12	표정/행동	손, 눈, 머리, 얼굴, 버리다, 잡다, 입, 웃다, 앓다, 울다, 옆, 웃, 입다, 치다, 뒤, 쓰다, 나가다, 맛, 아프다, 묻다, 들어가다, 표정
13	우울	죽다, 먹다, 우울, 힘들다, 우울증, 아프다, 약, 엄마, 싫다, 병원, 자해, 자살, 울다, 버리다, 무섭다, 아빠, 기분, 정신, 제발, 심하다, 괜찮다, 불안, 치료, 상태, 스트레스
14	방탄소년단	드리다, 이벤트, 생일, 방탄, 보내다, 부탁, 감사, 축하, 소년단, 참여, 폼, 입금, 투표, 아미, 호석, 기념, 선물, 진행, 태형, 팬, 판매
15	정치	문, 대통령, 국민, 당, 정부, 재인, 한국, 뉴스, 김, 나라, 의원, 정치, 권, 민주당, 일본, 김경수, 정권, 북한, 대표, 박근혜, 국가, 법, 국회, 대한민국
16	팬/친구	듣다, 노래, 부르다, 팬, 좋아하다, 친구, 치다, 쓰다, 미치다, 돌, 웃기다, 언니, 새끼, 시발, 영상, 아이, 덕, 글, 오빠, 올리다, 목소리
17	돈	쓰다, 돈, 사다, 책, 읽다, 글, 게임, 폰, 사진, 낳다, 올리다, 벌다, 카드, 버리다, 가능, 찾다, 사용, 회사, 정리, 찍다, 추천
18	운세	색, 자리, 머리, 좋아하다, 별, 컬러, 운세, 핑크, 쓰다, 닉네임, 행운, 발, 럭키, 열쇠, 출처, 염색, 어울리다, 바꾸다, 글자, 톤, 눈, 별자리
19	방송/연기	영화, 독자, 신, 배우, 최, 캐다, 드라마, 좋아하다, 왕, 박, 사랑, 장면, 작가, 연기, 쓰다, 작품
20	음식	먹다, 맛, 밥, 마시다, 맛있, 사다, 술, 낳다, 음식, 좋아하다, 물, 커피, 고기, 치킨, 빵, 요리, 저녁, 우유, 라면, 맥주, 딸기, 굽다, 케이크, 초, 피자



〈그림 5〉 비우울 경향 이용자 우세 토픽 비교



〈그림 6〉 우울 경향 이용자 우세 토픽 비교

한 양상을 보이는지 확인했다. 특정 주제를 구성하는 주제 단어들은 〈표 2〉에 나타내었다. 앞서 수행했던 토픽모델링 실험에 비해, 데이터 수가 충분하지 않아 불명확한 주제 구성 단어들도 등장하였으나, 여러 번 시도 후 가장 결과가 잘 나오는 토픽 20개를 추출하였다.

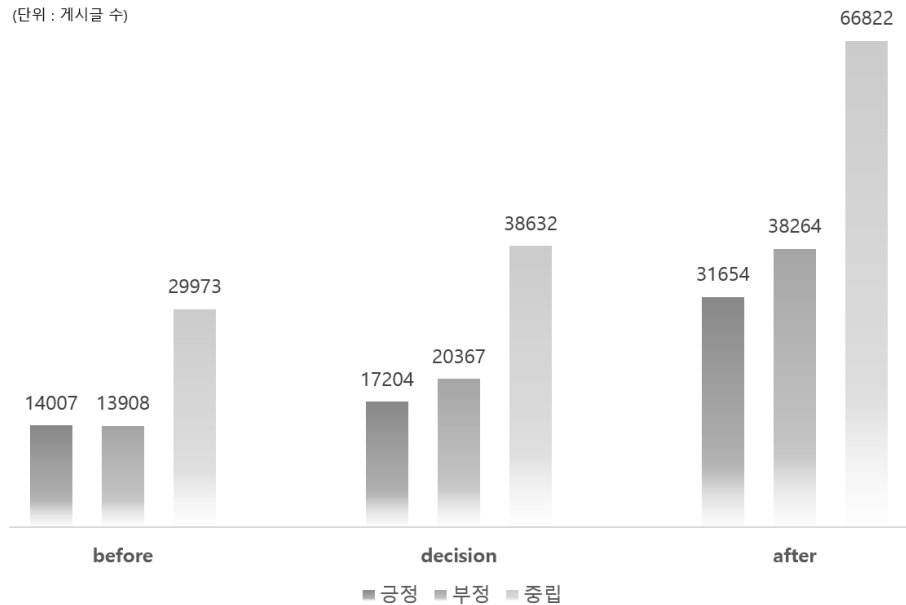
4.3 우울 경향 이용자 대상 감성 분석

우울 경향을 보이는 이용자들을 대상으로 기간에 따라 감성에도 어떠한 특성을 보이는지 확인 해보기 위하여 감성분석을 실시하였다. 기간별 감성별 추이를 살펴보기에 앞서 감성

〈표 2〉 젠더이슈 & 성범죄와 관련된 주제와 이를 구성하는 주제 단어 예시

주제명	4. 여성/젠더	14. 혐오(남성)	18. 성범죄	19. 우울증	기타 주제명
주제 단어	여성, 성, 여자, 가능, 한국, 남성, 모든, 쓰다, 다르다, 느끼다, 가장, 존재, 높다, 페미니즘, 능력, 의미 ...	여자, 남자, 싫다, 새끼, 시발, 치다, 씨발, 놈, 친구, 욕, 한국, 당하다, 남자들, 버리다, 좆 ...	여성, 사건, 피해자, 성, 폭행, 당하다, 가해자, 경찰, 미투, 여자, 이수역, 성폭력, 범죄, 남성, 처벌, 촬영, 불법, 강간, 남자 ...	죽다, 우울증, 힘들다, 자살, 아프다, 자해, 싫다, 병원, 울다, 버리다, 무섭다, 약, 불안, 기분, 정신, 치료, 엄마, 환자, 심하다, 상담 ...	1-음식, 2-학교, 3-글(취미), 5-일상, 6-방송/연예, 7-반려동물/아이, 8-노이즈(감정), 9-노이즈(이벤트), 10-기분/감정, 11-외모, 12-청소, 13-자연, 15-국가, 16-연기/방송, 17-트위터, 20-가족

(단위 : 게시물 수)



〈그림 7〉 기간별 우울 경향 이용자 감성분석

값을 추출하는 방법으로는 수집한 데이터를 기반으로 구축한 사전을 통해 긍정 단어가 등장할 때 +1, 부정 단어가 등장할 때 -1 값을 주고 연산하여 해당 트위터의 감성 값을 결정하게 되는 특성 상, 상쇄되는 경우(+1과 -1의 연산)이 많이 발생하여 중립의 갯수가 제일 많이 등장하였다(〈그림 7〉 참조).

본 연구에서는 감성 분석 실험 후, 긍정/부정의 결과 값을 단순히 수치로 비교하는 것이

아닌 동시출현 단어 분석을 통해 어떤 단어가 구성되어 있는지 확인해 보고자 하였다. 먼저, before 시기의 긍정에서 확인할 수 있는 특징으로는 크게 '사랑', '행복', '감성'의 단어가 중심을 이루는 것을 알 수 있었다. 반면에 부정 감성을 지닌 단어 구성들을 살펴보면 '여성', '성', '장애' 등의 단어가 중심을 이루어 구성되어 있는 것을 볼 수 있는데 이 단어들은 긍정의 감성에서도 확인할 수 있는 단어이다. 하지만

긍정 그룹에서는 '평등', '다양', '권리'와 같은 존중의 의미를 지닌 단어들 이 등장하였으나 부정 그룹에서는 '추행', '범죄', '낙태' 등과 같은 단어들과 연관되어 등장하였다는 점에서 차이가 있다고 할 수 있겠다. 다음으로 decision 기간에서 긍정에 해당하는 단어들은 '사랑', '마음'이라는 단어 이외에도 '우울증'과 관련된 단어들이 주 그룹을 이루고 있는데, 이는 decision 기간 그룹 자체가 '우울'을 중심으로 수집하고 정제한 데이터를 바탕으로 이루어졌기 때문으로 보인다. 그러나 연관된 단어들이 '감정', '의지', '행복' 등의 단어들이므로 미루어보아 우울의 상태를 극복하고 나아지기 위한 이야기들이 포착된 것으로 추측된다. 부정 그룹에서는 우울증이 '자해', '자살', '사건', '성' 등의 단어들과 연관되어 나타났다. 이는 이전의 before 그룹과 비교했을 때, 더욱 구체적인 우울증 증상 - '자해', '자살' 등의 단어들이 두드러졌고, '사건'과 관련한 단어들의 경우 단순히 성범죄, 성폭행과 같이 포괄적으로 묘사하는 것이 아닌 '연대', '이수'와 같이 구체적인 범죄 사건에 대해 언급하고 있었다. 이는 해당 시기와 관련하여 페미니즘 논쟁 및 이수역 사건에 대해 언급한 것으로 보여진다. 계속해서 after 기간의 긍

정 그룹에서는 '사랑'이라는 주제가 가장 크게 형성되어 있었다. 그리고 '행복', '친구'에 대한 내용도 관련 단어들과 그룹을 형성하고 있었지만, 이전 다른 기간들과 비교했을 때는 비교적 덜 다양한 주제들이 형성되어 있는 것을 볼 수 있다. 반대로 after 기간의 부정 그룹에서는 이전 기간의 그룹들과 유사한 주제들 - '우울증', '여자/남자', '성', '사건' 등이 등장하였지만 훨씬 다양한 단어들까지 포함하여 등장하였다. 구체적인 사건과 관련된 키워드 - '이수', '연대'를 포함하여 '경찰', '색출' 등과 같이 그 후속 사건 처리와 관련된 내용까지 활발하게 이야기가 나오고 있는 것으로 파악된다.

지금까지 살펴본 우울 경향 이용자들의 기간별 감성분석 결과는 다음 <표 3>에 요약하여 나타내었다.

5. 결론

5.1 결론 및 제언

본 연구는 트위터 데이터를 대상으로 우울 경향을 나타내는 표현을 통해 이용자들의 성향

<표 3> 우울 경향 기간별 감성 분석에 따른 동시출현단어 분석 결과

	Before 기간	Decision 기간	After 기간
긍정 군집핵심 단어	'사랑', '행복', '감사', '성', '여성', '이벤트', '추침'	'사랑', '마음', '우울증', '추침'	'사랑', '행복', '친구', '게이', '이벤트'
부정 군집핵심 단어	'여자', '남자', '여성', '성', '장애', '우울증'	'우울증', '우울', '자해', '자살', '성', '여성', '연대', '이수'	'우울증', '성', '사건', '여자', '남자', '여성', '우울', '강간', '판결'
특징	긍정/부정에서 형성된 군집 핵심 단어를 중심으로 다양한 범위에 해당하는 내용 언급	기간 설정을 위해 쿼리로 포함시킨 '우울증'과 관련된 내용 뿐 아니라, 성범죄 관련 구체적 사건('연대', '이수역')에 대한 언급	'우울증', '여성', '성범죄'와 관련하여 추상적인 내용이 아닌, 구체적인 사건과 후속처리('경찰', '판결')와 관련된 전반적인 내용에 대해 언급

및 관심 주제를 파악해보고자 하였다. 단순히 우울 경향 이용자와 우울 경향을 보이지 않는 이용자와의 차이점을 보이는 데만 그치는 것이 아니라, 우울 경향 이용자 중에서도 구간을 나누어 우울 경향의 기간별 특성에 대해서도 살펴보았다.

이러한 분석을 위해 트위터에서 우울 경향의 이용자들의 데이터와 우울 경향을 보이지 않는 이용자들의 데이터를 수집하고 동시출현 단어와 토픽모델링, 감성 분석을 통해 그들의 성향이나 관심 주제에 대한 차이점을 보았다. 연구 결과로 먼저 우울 경향 이용자와 비우울 경향 이용자가 트위터 상에서 공유하고 관심을 가지는 이야기의 주제적 차이를 확인할 수 있었다. 토픽모델링과 동시출현 단어분석을 통한 주제 단어들을 살펴보면, 보다 일상적이고 다양한 성격의 주제 어휘들을 많이 사용하는 비우울 경향 이용자에 비해 우울 경향 이용자들은 '여성' 및 '성범죄' 등 여성 인권, 사회 이슈, 인간관계와 관련된 단어, 그리고 우울증 증상인 '자살', '자해' 등 성격이 뚜렷한 주제 단어들이 많이 보이는 것을 포착하였다. 또한 우울 경향 이용자들의 시기에 따른 세부적 특성을 살펴보았을 때, 시기의 변화에 따라서도 트위터에서 나누는 답변의 내용에 미세한 차이가 있다는 것을 알 수 있었다. 우선 시기에 상관없이 공통적으로는 우울증과 관련하여 사람들이 고민하고 힘들어하는 상황에 대한 내용이나 우울증 증상, 그리고 우울증과 관련된 주변 환경을 의미하는 단어들이 등장하였다. 세부적으로 before 그룹에서는 비교적 '약', '병원', '의사'와 같이 우울증 질환에 대해 병원 진단과 치료 과정에 대한 주제가 보였고, decision 그룹에서는 우울증을 겪으며 도

움이 될 수 있는 '반려동물'이나 '서점'과 같이 의지하고 위로받을 수 있는 존재에 대한 언급을 확인할 수 있었다. 마지막으로 after 그룹에서는 '담배'나 '술'과 같이 정신적, 감정적으로 힘들 때 의존하게 되는 대상들이 등장하였으며 '죽음'과 같이 극단적인 어휘도 발견할 수 있었다. 이 밖에도 감성분석의 결과 우울 경향의 이용자들이 보이는 감성 표현의 특성을 살펴보았을 때, 긍정/부정의 값을 형성하는 단어들이 유사한 성격의 단어들 중심으로 형성되더라도, 기간이 before에서 after 시기로 갈수록 더 넓은 주제 범위의 이야기를 나누는 경향이 있었다.

본 연구는 우울증이라는 정신질환에 대해서 소셜미디어 중 트위터를 통해 기존에 주로 영어로만 수집되고 수행되었던 연구와는 달리, 한글로 된 많은 양의 트위터 데이터를 수집하고 분석을 시도하였고 우울 경향을 보이는 이용자와 그렇지 않은 이용자 간 트위터에서 보이는 성향 및 관심 주제의 차이뿐만 아니라 우울 경향의 시기별로도 특성을 보고자 한 점에서 유의미하다고 할 수 있다. 그러나 본 연구에서 설정한 데이터 수집 및 구분 기준에 대해 수동적으로 처리하는 과정에서 오류나 편향된 내용이 발생하였을 수도 있기 때문에, 대상 비교에 있어서 객관성이나 분석 내용에 한계가 있다는 문제점이 있다. 따라서 후속 연구에서는 짧은 길이의 트위터 텍스트 데이터 분석에 더욱 적합한 텍스트마이닝 기법을 적용하고, 우울증 이용자에 대한 기준이나 분석 내용의 근거에 있어 여러 번의 반복 및 통계적 검증을 통한 검토를 통해 타당성을 확보하고 연구가 이루어진다면 더욱 유의미한 결과 발견의 향상을 기대해 볼 수 있을 것이다.

참 고 문 헌

- 김재봉, 김형중 (2017). 주가지수 방향성 예측을 위한 도메인 맞춤형 감성사전 구축방안. 한국디지털콘텐츠학회 논문지, 18(3), 585-592.
- 배정환, 손지은, 송민 (2013). 텍스트 마이닝을 이용한 2012년 한국대선 관련 트위터 분석. 지능정보연구, 19(3), 141-156. <https://doi.org/10.13088/jiis.2013.19.3.141>
- 서상현, 김준태 (2016). 딥러닝 기반 감성분석 연구동향. 한국멀티미디어학회지, 3, 8-22.
- 송민 (2017). 텍스트마이닝. 서울: 청람.
- 송호윤, 박한철, 양원석, 박중철 (2017). 언어학적 패턴을 이용한 소셜 미디어 사용자의 우울증 증세 예측. 한국정보과학회 학술발표논문집, 625-627. 부산: 한국정보과학회.
- 유은지, 김유신, 김남규, 정승렬 (2013). 주가지수 방향성 예측을 위한 주제지향 감성사전 구축 방안. 지능정보연구, 19(1), 95-110.
- 이상훈, 최정, 김중우 (2016). 영역별 맞춤형 감성사전 구축을 통한 영화리뷰 감성분석. 지능정보연구, 22(2), 97-113. <https://doi.org/10.13088/jiis.2016.22.2.097>
- 이유림 (2016). 정서의 약료화와 우울증 경험의 구성. 페미니즘 연구, 16(1), 81-117.
- 이현서, 송민 (2018). 트럼프 취임 관련 국내 언론에서 나타난 감성과 거시 경제 지수 간 영향 관계. 언론과학연구, 18(3), 129-169. <https://doi.org/10.14696/jcs.2018.09.18.3.129>
- 장재영 (2009). 온라인 쇼핑몰의 상품평 자동분류를 위한 감성분석 알고리즘. 한국전자거래학회지, 14(4), 19-33.
- 진설아, 허고은, 정유경, 송민 (2016). 토픽 모델링 기반 정보학 분야 학술지의 학제성 측정 연구. 정보관리학회지, 33(1), 7-32. <https://doi.org/10.3743/kosim.2016.33.1.007>
- Bastian, Heymann, & Jacomy (2009). Gephi: An open source software for exploring and manipulating networks. Icwsm, 361-362.
- Coppersmith, G., Dredze, M., & Harman, C. (2014). Quantifying mental health signals in twitter. Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, 51-60. <https://doi.org/10.3115/v1/w14-3207>
- De Choudhury, M. C. (2013). Social media as a measurement tool of depression in populations. Proceedings of the 5th Annual ACM Web Science Conference, 47-56. ACM. <https://doi.org/10.1145/2464464.2464480>
- De Choudhury, M., Gamon, M., Counts, S., & Horvitz, E. (2013). Predicting depression via social media. ICWSM, 13, 1-10.
- Hammen, C., & Brennan, P. A. (2002). Interpersonal dysfunction in depressed women: Impairments

- independent of depressive symptoms. *Journal of Affective Disorders*, 72(2), 145-156.
[https://doi.org/10.1016/s0165-0327\(01\)00455-4](https://doi.org/10.1016/s0165-0327(01)00455-4)
- Jiang, L., & Yang, C. C. (2013). Using co-occurrence analysis to expand consumer health vocabularies from social media data. *IEEE International Conference on Healthcare Informatics*, 75-81. <https://doi.org/10.1109/ichi.2013.16>
- Mimno, D., & McCallum, A. (2008). Topic models conditioned on arbitrary features with dirichlet-multinomial regression. *The 24th Conference on Uncertainty in Artificial Intelligence*, 411-418. Helsinki: UAI.
- NAMI. (2014). Tell me about depression. Retrieved from NAMI - National Alliance on Mental Illness: <https://www.nami.org/Learn-More/Mental-Health-Conditions/Depression>
- Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up?: sentiment classification using machine learning techniques. *Proceedings of the ACL-02 conference on Empirical methods in natural language processing*, 10, 79-86. Association for Computational Linguistics.
- Park, M., Cha, C., & Cha, M. (2012). Depressive moods of users portrayed in twitter. *Proceedings of the ACM SIGKDD Workshop on Healthcare Informatics*, 1-8. New York: ACM.
- SHINEWARE. (2013, 3, 20). Products-komorán. Retrieved from SHINEWARE:
<https://www.shineware.co.kr/products/komorán/>
- Snow, R., O'Connor, B., Jurafsky, D., & Ng, Y. A. (2008). Cheap and fast---but is it good?: evaluating non-expert annotations for natural language tasks. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 254-263. Association for Computational Linguistics. <https://doi.org/10.3115/1613715.1613751>
- Wang, H., Can, D., Kazemzadeh, A., Bar, F., & Narayanan, S. (2012). A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. *Proceedings of the ACL 2012 System Demonstrations*, 115-120. Association for Computational Linguistics.
- Wang, Z. (2017). *Machine learning methods for finding textual features of depression from publications*. Georgia: Georgia State University.
- Zhao, W. X., Jiang, J., Weng, J., He, J., Lim, E. P., Yan, H., & Li, X. (2011). Comparing twitter and traditional media using topic models. *European conference on information retrieval*, 338-349. Berlin: Springer.
- Zhou, Z., Wang, W., & Wang, L. (2012). Community detection based on an improved modularity. *Springer*, 638-645. Berlin: Heidelberg. https://doi.org/10.1007/978-3-642-33506-8_78

• 국문 참고문헌에 대한 영문 표기
(English translation of references written in Korean)

- Bae, Jung-hwan, Son, Ji-eun, & Song, Min (2013). Analysis of twitter for 2012 south Korea presidential election by text mining techniques. *Journal of Intelligence and Information Systems*, 19(3), 141-156. <https://doi.org/10.13088/jjis.2013.19.3.141>
- Chang, Jae-Young (2009). A sentiment analysis algorithm for automatic product reviews classification in on-line shopping mall. *The Journal of Society for e-Business Studies*, 14(4), 19-33.
- Jin, Seol-A, & Song, Min (2016). Topic modeling based interdisciplinarity measurement in the informatics related journals. *Journal of the Korean Society for Information Management*, 33(1), 7-32. <https://doi.org/10.3743/kosim.2016.33.1.007>
- Kim, Jae-bong, & Kim, Hyung-Joong (2017). A domain-specific sentiment lexicon construction method for stock index directionality. *Digital Contents Society*, 18(3), 585-592.
- Lee, Hyun-Seo, & Song, Min (2018). An analysis of the influences between sentiment values of korean online news and macroeconomic indicators using text mining. *Journal of Communication Science*, 18(3), 129-169. <https://doi.org/10.14696/jcs.2018.09.18.3.129>
- Lee, Sang Hoon, Jing Cui, & Kim, Jong-Woo (2016). Sentiment analysis on movie review through building modified sentiment dictionary by movie genre. *Journal of Intelligence and Information Systems*, 22(2), 97-113. <https://doi.org/10.13088/jjis.2016.22.2.097>
- Lee, Yu-Lim (2016). Pharmaceuticalization of emotion and structuring depression as experience: An analysis of the depression experiences of Korean women in their 20s. *Korea Womens Studies Institute*, 16(1), 81-117.
- Seo, Sang-hyeon, & Kim, Jun-Tae (2016). Deep learning based sentiment analysis research trends. *Korea Multimedia Society*, 3, 8-22.
- Song, Ho-Yun, Park, Han-chul, Yang, Won-seok, & Park, Jong-chul (2017). Predicting symptoms of depression for social media users via linguistic patterns. *Korea Information Science Society*, 625-627.
- Song, Min (2017). *Text mining*. Seoul: Chungram Books.
- Yu, Eun-Ji, Kim, Yoo-sin, Kim, Nam-gyu, & Jeong, Seung-ryul (2013). Predicting the direction of the stock index by using a domain-specific sentiment dictionary. *Journal of Intelligence and Information Systems*, 19(1), 95-110.

[부록] 우울 경향 이용자 대상 감성분석 사전 리스트

	긍정 단어 목록	부정 단어 목록
사전 구성 단어	<p>행복, 사랑, 노력, 감사, 예쁘다, 귀엽다, 좋아하다, 원하다, 쉽다, 놀다, 편하다, 도와주다, 도움, 용기, 웃다, 자유, 희망, 의지, 친하다, 착하다, 성공, 지키다, 고맙다, 맛있다, 응원, 안정, 나아지다, 인정, 위로, 웃기다, 이쁘다, 진심, 괜찮다, 다행, 재밌다, 즐겁다, 좋아, 낫다, 기쁘다, 이기다, 소중, 애정, 멋지다, 새롭다, 승리, 대박, 존중, 따뜻, 대단, 이루다, 강하다, 최고, 견디다, 꾸준, 배려, 공감, 아름답다, 선물, 즐기다, 보상, 힘내다, 즐겁다, 배려심, 순진, 잘생기다, 뿌듯, 상냥, 장점, 귀여움, 흐뭇, 짱, 깨끗, 편하다, 친절, 편안, 나이스, 도움, 호의, 근사, 완벽, 축하, 어울리다, 추천, 존맛탱, 강추, 달콤, 멋지다, 경이, 상냥, 당당, 기적, 예의, 깨닫다, 웃음소리, 존경, 감동, 환영, 흥미, 신선, 해맑다, 흡족, 편안, 만족, 재미, 즐거움, 부드럽다, 교훈, 다정, 감동, 맑다, 센스, 헤아리다, 섬세, 지지, 풍요, 안정, 깔끔, 천국, 행운, 특별, 가치, 아끼다, 기쁨, 빛나다, 승승장구, 극진, 미소, 반갑다, 웃음, 순수, 최선, 편하다, 감사, 용기, 들뜨다, 상쾌, 자유, 축하, 축복, 포근, 인연, 사귀다, 응원, 쿨트, 성취감, 해내다, 자랑, 친밀, 다독이다, 평온, 이해하다, 낫다, 인정, 실력, 합리, 신뢰, 양보, 천재, 꽃피우다, 성취, 귀, 인기, 대박, 매력, 예뻐죽겠, 청량, 스윗, 칭찬, 자신감, 용감, 순박, 영향력, 합격, 놀다, 줄기탱, 애교, 미모, 상큼, 신비, 훌륭한, 선호, 살아남다, 옳다, 해맑다, 기여움, 미인, 반갑다, 현모양처, 적절, 싱그럽다, 안심, 벽차다, 열심히, 당당, 정의, 단정, 반듯, 해피엔딩, 현명, 파스함</p>	<p>가해, 경악, 억압, 다투다, 상처, 한남춤, 강간, 중단, 털리다, 포주, 색출, 수사, 금하다, 낙태, 죄, 위험, 좇국, 그만두다, 실패, 사찰, 심각, 침해, 분노, 의심, 변질, 총공, 털다, 좇같다, 폐지, 미상, 좇같다, 박탈, 막무가내, 유린, 규탄, 취조, 밉다, 벌, 벌주다, 불편, 혈안, 망하다, 독하다, 죽다, 오류, 넘겨, 놀리다, 시위, 오점, 혐오, 얼룩지다, 중죄, 탄핵, 속상하다, 묻히다, 방관, 핑계, 각성, 탈교, 아가리, 편파, 폭력, 여혐, 불법, 전쟁, 국뽕, 하소연, 물어뜯다, 폐미, 싫다, 악마, 반성, 꼴, 성희롱, 지껄이다, 폐지, 개씹곤, 고소, 따위, 짜질함, 역겹다, 차별, 반말, 조롱, 당하다, 대자보, 더럽다, 덧씌우다, 소름, 갈등, 밀치다, 메갈년, 간음, 강제, 추행, 항소심, 유죄, 감옥, 음란물, 혐의, 조사, 편견, 단절, 성폭력, 모욕, 협박, 감금, 죽다, 거부, 위마, 악행, 딸년, 싸가지, 위기, 처하다, 공격, 비겁, 치졸, 숨기다, 뺨치, 공범, 힐난, 돼지, 찌들다, 무시무시, 몰카, 뺑다, 피살, 가해자, 세뇌, 피, 맘충, 징역, 죽이다, 사망, 징역, 범죄, 오염, 욕, 들끓다, 음란, 욕구, 거르다, 싸지르다, 놀, 무섭다, 패다, 빌미, 강제, 스킨미투, 엮기, 의심, 흐리다, 새끼, 실망, 물의, 죄송, 멍자국, 한녀, 왕따, 지우다, 징계, 사과문, 내쫓다, 울다, 괴해, 선동, 몸싸움, 트페미, 몸쓸, 폭행, 진압, 망하다, 뜯기다, 못되다, 충격, 추악, 질투, 망하다, 망신, 충돌, 무섭다, 비난, 과하다, 짜증, 멍청, 비판, 도촬, 욕, 걱정, 지긋지긋, 잘못, 욕하다, 드럽다, 잘못, 허구, 기만, 비속어, 어이없다, 추락, 사망, 죄, 격분, 시달리다, 압박, 멍청, 자살, 씹새끼, 노예, 살인, 위협, 유포, 추행, 문란, 폭행, 편파, 후려치다, 범행, 황당, 비하, 망신, 위협, 짜증, 타격, 모욕, 조롱, 씨발, 새끼, 시발, 끔찍, 흠치다, 뺑다, 나쁘다, 환멸, 싸우다, 극혐하다, 속이다, 모순, 망하다, 한남, 한심, 메갈, 개뻥, 파면, 서운, 서러움, 은폐, 고통, 깎깎, 아프다, 슬프다, 굶다, 지옥, 분노, 실패, 좇, 잔혹, 울적, 불안, 울다, 멍청이, 싫어하다, 심하다, 힘들다, 복잡, 슬픔, 답답, 역겹다, 일베들, 위마드, 끔찍, 불안, 스트레스, 혼란, 걱정, 못나다, 지치다, 별로, 불행, 상, 사악, 값이먹다, 외로움, 최악, 해악, 치욕, 무례, 수치, 쪽팔리다, 멸망, 지랄, 지겹다, 두렵다, 하찮다, 악몽, 이쁨, 때려치다, 앓다, 엉엉, 미친년, 구린내, 좇, 좇, 환멸, 창피, 불안증, 창녀, 탐욕, 안타까움, 외롭다, 고독, 착취, 자해, 죽음</p>