

연세대학교 대나무숲을 통해 본 대학생들의 감정 공유 양상 분석

팀: 오픈톡 (노경준, 원서연, 이지영)

1. 서론

정보통신기술의 발달로 시공간의 제약이 희미해져가는 오늘날, 커뮤니티의 개념과 양상 또한 변화를 거듭하고 있다. 과거의 전통적 오프라인 커뮤니티가 혈연/지연과 같은 객관적 특성에 의해 정의되고 면대면 상호작용을 근간으로 했다면, 현대의 커뮤니티는 비슷한 관심이나 가치를 중심으로 주관적으로 형성되며, 온라인 공간이라는 새로운 사회적 공간을 그 무대로 삼고 있다 (Knapp & Daly, 2002; Brown et al. 2007).

온라인 커뮤니티에 대한 정의는 학자마다 다양하지만, '다양한 상호작용으로서의 지속적인 관계성'을 중요시 한다는 점에서는 그 궤를 함께 한다. 이는, Rheingold (1993)가 '감정 공유'와, '인간적 유대'를 온라인 커뮤니티에서의 가장 중요한 특징이라 언급한 바와 같이, 해당 커뮤니티 내에서도 전통적인 오프라인 커뮤니티와 같은 '높은 수준의 커뮤니티 감정'이 공유되고 있음을 시사한다. 온라인 환경의 발달과 스마트폰의 대중화 등으로 SNS (Social Network Service)가 더 많은 사람들의 커뮤니티 참여를 이끌어 내면서 이러한 양상은 더욱 도드라지고 있다.

SNS란 개인들로 하여금 특정 시스템 속에 자신들의 프로필을 만들도록 해주고 그것들을 타인에게도 소개해주어 이를 기반으로 사람들이 타인과 연계하고, 소통할 수 있도록 만들어주는 웹 기반의 서비스로 정의된다 (Boyd & Ellison, 2008). 이를 바탕으로 SNS는 시공간의 제약을 뛰어넘어 보다 높은 접근성과 참여 기회를 제공함으로써 타인과의 감정 공유와 인간적 유대를 더욱 용이하게 하는 온라인 커뮤니티로서의 기능을 수행하고 있다. 한 발 더 나아가, 사람들이 오프라인에서의 면대면 커뮤니케이션 상황보다 더 능동적이고 솔직한 의견을 피력할 수 있게 만든 요인이 있는데, 온라인 커뮤니티의 '차별점'이라 할 수 있는 '익명성'이 바로 그것이다. '트위터'나 '페이스북' 등의 SNS는, 아이디(ID)나 닉네임 혹은 자신을 감추는 익명을 활용하여 의견을 표출할 수 있는 기능을 제공하고 있는데, 이는 사람들의 보다 적극적이고 능동적인 참여를 촉진하는 중요한 요소로 작용한다.

본 연구팀은 이러한 온라인 익명 커뮤니티의 익명성에 기반한, 보다 활발하고 정제되지 않은 감정 공유 및 인간적 유대에 주목하고자 한다. 이 중에서도, 동일 연령대인 대학생들의 온라인 익명 커뮤니티 상에서의 감정 표현과 공감 등의 교류 양상을 파악하기 위해 페이스북 페이지 '연세대학교 대나무숲'의 커뮤니티를 중심으로 연구를 구상하였다¹. 특히, 좋아요 수가 높은 게시글들을 분석 대상으로 삼았는데, 이를 통해 정보가 범람하는 SNS 상에서 특별한 기준을 따로 마련하기 어려울 수 있다는 우려를 줄임과 동시에, 객관적 수치로 나타나는 보편적 정서를 확인하고자 하였다. 이에 따라, 수집한 게시물들의 특성(극성, 길이 등)과 커뮤니티 참여자들의 반응 (좋아요 수, 댓글 수) 등을 양적 차원과 질적 차원에서 종합적으로 살펴보고자 한다.

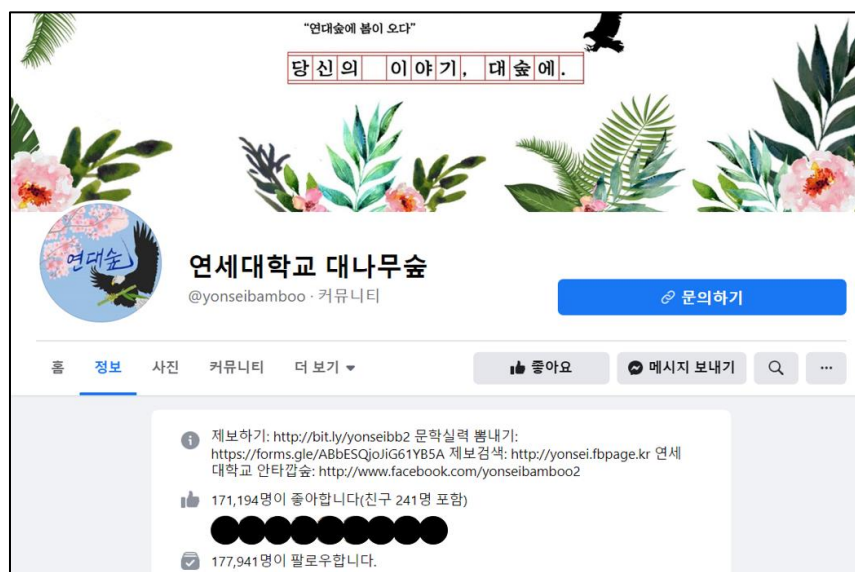
¹ 연세대학교 대나무 숲으로 분석을 한정한 이유는, 필자들이 몸 담고 있는 커뮤니티와 구성원이기 때문에, 보다 심도 있게 이해할 수 있다는 장점을 갖기 때문이다.

위의 연구 목적을 달성하기 위해 몇 가지 선행연구를 앞서 살펴보았다. 이론적으로는 익명성(Anonymity)과 친밀도(Intimacy)에 따른 자기노출(Self-disclosure) 정도의 차이를 확인한 연구(Ma, Hancock, & Naaman, 2016)를 참고했다. 방법론적으로는 동시출현단어분석(Co-word Analysis)과 시계열 분석, LDA (Latent Dirichlet Allocation) 토픽 모델링을 통해 트위터 데이터를 분석해 토픽의 변화 시점 및 패턴을 파악한 연구(진설아, 허고은, 정유경, 송민, 2013)를 참고하여 연구를 설계, 진행하였다. 해당 연구에 대한 본 논문은 현재 1장 서론을 시작으로 다음과 같이 구성된다. 2장에서는 본 연구에서 사용한 데이터와 변수 등 사전 정보에 대한 내용을 소개하고, 3장에서는 워드 클라우드, 키워드 네트워크 데이터, 그리고 LDA에 대한 분석과 함께 본 연구팀의 해석을 제시한다. 끝으로, 4장에서는 본 연구의 결론, 한계 및 발전 방향에 대해 논하며 마무리한다.

2. 데이터 구성 및 기초 분석

본 연구의 목적은 크게, 대학생들이 그들만의 익명커뮤니티 속에서 1) 어떠한 주제에 대해 주로 이야기 하고 공감을 받는지, 2) 해당 게시물들이 어떤 감정들을 담고 있으며 이에 따라 어떤 양상을 보이는지를 파악해보는 것이다. 해당 내용들을 비교 분석하기 위하여, 본 연구팀은 페이스북 페이지 중 하나인 '연세대학교 대나무숲'에 게시된 텍스트들을 수집하였다. 보통 각 학교마다 존재하는 대나무 숲은, 공동 계정의 형태로 운영되며 투고자의 익명성이 보장되는 방식을 채택하고 있다. 이와 같은 익명의 형태로 SNS에 올라오는 게시물들은, 일반적인 SNS 계정에 올라오는 게시물과 다른 행태를 보인다. 이는 해당 커뮤니티가 제공하는 익명성이라는 특성이 사람들로 하여금 '참 자아'를 보다 쉽게 드러낼 수 있는 환경을 제공해주기 때문이다(Bargh et al. 2002).

<그림 1> 연세대학교 대나무숲 페이지



페이스북의 대나무 숲은 2010년대 초반부터 현재까지 활발하게 운영되어 온 대학 커뮤니티로 평가 받는데, 특히 연세대학교 대나무숲의 사용자 수가 서울대와 고려대에 이어 세 번째로 많다는 점에서 어느 정도 분석 결과의 일반화 가능성이 보장될 것이라 기대하였다<그림 1>. 앞 단에 설명하였 듯, 사람들의 공감을 자아내는 보편적 감정을 담은 글을 중심으로 살펴보기 위해 좋아요 수를

기준으로 게시물을 나열한 후, 가장 많은 좋아요를 받은 게시물 300건²의 내용과 관련 정보들을 수집하였다. 해당 데이터는 파이썬의 Beautifulsoup를 활용하여 <한양대 컴퓨터전공 허상민>이 운영하는 연세대학교 대나무숲 인기 제보 페이지³에서 수집하였으며, 수집된 게시물의 게시 시점은 2014년 10월 5일부터 2019년 8월 23일까지이다. 수집 데이터에 2019년 하반기와 2020년 상반기의 게시물이 포함되지 않은 것은, 2018년 하반기부터 대두된 페이스북의 침체와 이에 따른 사용자들의 타 SNS로의 대거 이탈 등에 의한 영향 때문인 것으로 예상된다. 대나무숲에서 수집한 정보들은 아래 <그림 2>의 이미지에 제시된 것과 같이 게시물의 제보 내용, 제보 날짜, 좋아요 수 및 댓글 수, 그리고 가장 많은 추천을 받은 베스트 댓글의 세 가지이다.

<그림 2> 대나무숲 인기제보 페이지



크롤링한 300개의 게시물들의 '제보 내용'은 일차적으로 문법, 보편적 어휘에 부합하도록 본 연구자들이 1차 정제하였으며, 이후 텍스트들의 감정 분석을 위해 기준 데이터 값을 선정하였다. 정제 과정에서는 이모티콘과 특수문자, 마크업 기호, '연대숲 #번째 외침'과 같은 무의미하거나 불필요한 단어를 삭제하고, 본문이 없고 사진으로만 이루어진 게시글을 제외시켰다. 극성(Valence) 판단을 보다 정밀하고 객관적으로 진행하고자, 무작위로 선별한 30명의 연세대학교 학생들에게 직관적으로 긍정/부정/중립의 양상이 상대적으로 뚜렷한 100개의 게시물들에 대한 샘플 라벨링을 시행 했으며, 이렇게 확보된 기준을 바탕으로 텍스트를 활용, 앞서 정제한 전체 300개의 데이터를 '긍정', '부정', '중립'의 세 가지 클러스터로 분류하였다. 해당 텍스트들의 기초통계량은 아래 <표 1>에서 확인할 수 있다.

² 300위 안에 들더라도 이미지 파일은 언어 사용 양상을 보기 위한 탐색의 목적과 맞지 않아 다음 순위의 데이터를 수집하였다.

³ <http://yonsei.fbpage.kr/#/recommend>

감정분석 결과, 긍정적인 글이 127개, 부정적인 글이 56개, 중립적인 글이 117개로 분류 되었다. 300건의 게시물 중에 부정적인 글이 가장 적은 비율을 차지했는데, 부정적인 글은 다른 긍정과 중립의 클러스터에 비하여 사람들의 일반적인 공감을 이끌어내기 어려운 것으로 해석할 수 있다. 이는 글쓴이가 익명으로 제보를 하더라도 구독자의 개인 신원은 여전히 식별될 수 있는 페이스북 상에 게시된다는 특성을 고려해본다면 놀라운 결과는 아니다. 좀 더 부연하자면, '나'가 소셜네트워크 상에서 공감과 지지를 보내는 행위가 타인에게 목격될 수 있기에 부정적인 내용의 콘텐츠일수록 의사 표현을 내비치는 것이 상대적으로 더 부담스러웠을 거라 추측할 수 있다.

<표 1> 대나무 숲 인기 게시물 300건 기초 통계량

| 감정 | 개 수 | 글자 수 | 체언 수 | 좋아요 수 | 댓글 수 | 베스트 댓글 ⁴ 글자수 |
|----|-----|---------|--------|---------|--------|-------------------------|
| 긍정 | 127 | 1448.18 | 198.31 | 2214.74 | 331.72 | 100.53 |
| 부정 | 56 | 1467.89 | 210.63 | 1844.30 | 311.25 | 166.57 |
| 중립 | 117 | 256.70 | 39.88 | 1814.32 | 327.26 | 70.89 |
| 총계 | 300 | 987.18 | 138.82 | 1989.43 | 326.16 | 101.30 |

긍정으로 분류된 게시물들의 세부 항목을 먼저 살펴보면, 해당 군집은 가장 많은 좋아요와 댓글 수를 기록했는데, 댓글 수의 경우에는 집단간의 유의미한 차이가 발견되지는 않았다($p>0.05$). 하지만, 긍정 클러스터의 경우에는 비긍정(중립과 부정) 클러스터에 비하여 보다 많은 '좋아요'를 받은 것으로 나타났다($p<0.05$). 이를 통해, 긍정적인 글들이 다수 집단을 형성함과 동시에, 상위에 선정된 다른 극성의 클러스터들에 비하여 평균적으로 더 많은 공감을 이끌어 낸 것을 알 수 있었다. 부정적인 글의 경우, 다른 집단과 댓글 길이에 있어 유의미한 차이를 보였는데($p<0.05$), 이는 부정적인 글을 읽은 이들이 진심 어린 조언과 위로, 또는 관련된 자신의 경험을 길게 서술하게 되는 경우가 많기 때문인 것으로 보인다. 실제로 댓글의 내용을 살펴보면, 긍정 혹은 중립적인 게시물에 달린 댓글이 평균적으로 더 무겁지 않은 내용으로 구성되어 있었는데, 이는 사람들이 타인의 '부정적' 감정을 다룸에 있어 비교적 신중한 태도로 접근한다는 추론을 가능하게 한다.

중립적인 글의 경우, 다른 집단에 비해 짧은 길이 및 적은 체언 수를 가지고 있는 것으로 나타났는데, 이는 중립적인 글이 일종의 정보성 글로서 기능하기 때문인 것으로 보인다. 중립을 제외한 극단적인 극성을 가진 긍정과 부정의 클러스터의 경우, 개인적이고 감정적인 글이 주를 이루는 것을 고려한다면 이는 충분히 타당한 해석일 수 있다. 그렇지만, 중립적인 클러스터로 분류된 게시물들은 그 길이가 상대적으로 짧기에, 극성을 판단하기 어려워 '중립'으로 분류되었을 가능성 또한 고려되어야 한다. 흥미롭게도 중립 클러스터로 분류된 게시물들에 달린 베스트 댓글의 경우에도 타 클러스터의 댓글에 비해서 상대적으로 짧은 내용을 담고 있는 것으로 확인되었다($p<0.05$).

⁴ 베스트 댓글의 경우에는 각 게시물에서 가장 많은 추천을 받은 3개 댓글을 선택하였다.

3. 텍스트 데이터 분석: 워드 클라우드 / 키워드 네트워크 / LDA

<그림 3> 워드 클라우드



3.1 워드 클라우드 - 기초적인 정량 분석에서 나아가 정성적으로도 분석을 시도하기 위하여 본 연구팀은 일차적으로 워드 클라우드를 구현하였다. 워드 클라우드는 본문에 빈번하게 사용된 주요 단어를 직관적으로 보여주는 시각화 도구인데, 이를 활용하기 위해 각각의 문장들을 MeCab 분석기를 사용해 체언(일반 명사, 고유명사, 의존명사, 단위명사, 수사, 대명사)과 동사, 형용사, 관형사로 추려내는 전처리 과정을 진행했다. 이 때, 각 키워드 의미의 직관적 파악을 위해 2음절 이상의 최빈출 단어 30개로 한정하였으며 (양민석 & 김한길, 2018), 결과의 구체적인 양상은 위의 <그림 3>을 통해 확인할 수 있다. 이미지의 색은 차례대로 긍정(파랑), 중립(검정), 부정(빨강)의 클러스터를 나타낸다.

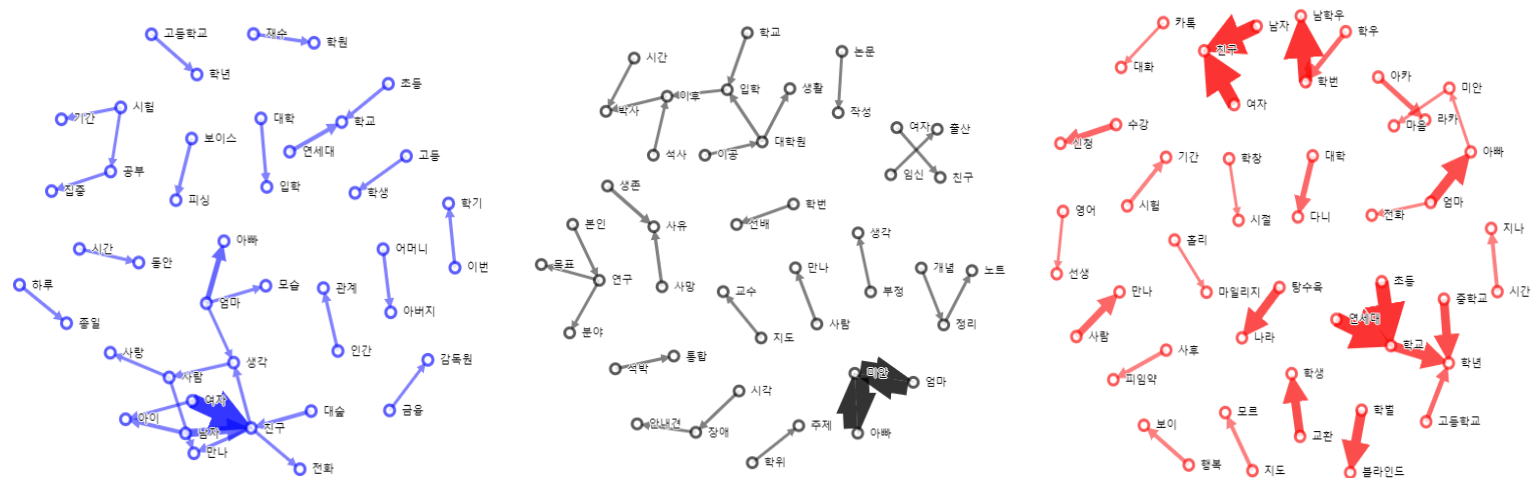
세 그룹에서 공통적으로 눈에 띄는 단어는 “엄마”와 “사람”이며, 그룹 간 다수의 공통된 어휘 사용이 목격되는 패턴을 고려하였을 때, 대나무 숲에서 좋아요를 많이 받는 게시물들은 일정한 주제군을 형성하고 있는 것으로 보인다. <표 2>의 내용에서 알 수 있듯, 세 집단 모두에서 목격되는 단어는 총 8개이며, 두 개의 집단에서 공통적으로 목격된 단어는 14개다. 주로 대학에 대한 내용 중 교수/이성 관계나 공부/일상에 관한 내용이 주를 이루는 것으로 보이는데, 아래의 어휘 군들의 사용에서 유추할 수 있는 주제에 대해서는 텍스트 네트워크와 LDA의 분석 결과를 참고하여 보다 정교하게 살펴보고자 한다.

<표 2> 워드 클라우드: 어휘 공통 및 개별 사용 양상

| | | | |
|-----------------------------------|---------|----------------|-----------------------------------|
| 긍정 / 중립 / 부정 | 긍정 / 중립 | 중립 / 부정 | 긍정 / 부정 |
| 공부, 마음, 사람, 생각, 시간, 아빠, 엄마, 학교 | 사랑, 정도 | 교수, 때문, 미안, 학생 | 남자, 대학, 아이, 애기, 여자, 전화, 친구, 행복 |

| 긍정 | 중립 | 부정 |
|--|---|---|
| 그러, 기억, 만나, 모르, 모습, 시작, 아버지, 얼굴, 연락, 예쁘, 이야기, 처음 | 경우, 논문, 대학원, 박사, 본인, 사유, 생활, 선배, 수능, 안내건, 연구, 연구실, 연대, 이상, 입학, 휴학 | 동안, 부모, 선생, 시험, 자신, 전공, 학년, 학번, 학점, 할아버지 |

<그림 4> 키워드 네트워크



3.2 키워드 네트워크 - 본 연구팀은 텍스트의 워드 클라우드(Word Cloud)를 활용하여 긍정, 부정, 중립의 글들의 내용을 구성하는 주요 어휘들이 무엇인지 구체적으로 살펴보았다. 또한, 해당 어휘들이 동일 문장 내 어떤 어휘와 함께 어울려 사용되는지 동시출현 단어(Co-Occurrence)들을 분석해 글의 전개 양상을 보다 직관적으로 파악할 수 있도록 키워드 네트워크(Keyword Network) 모델로도 구현해보았다.

위 과정에서도, 워드 클라우드와 마찬가지로 MeCab 분석기로 2음절 이상인 것들만 추출하여, 상위 30개까지의 키워드 쌍에 대한 네트워크로 시각화하였다. 위 과정에서 연구팀은 혐의를 거쳐, '엄마-엄마'와 같은 동일 어휘 쌍은 제외한 후 결과를 도출하였다. 일련의 정제과정 및 분석을 통해, 긍정 글의 경우 크게 '시간', '사람', '학교' 세 가지의 주제로 나뉜다고 판단하였다. 이 때, 각각의 주제를 구성하는 요소들을 살펴보면, 시간에 있어서는 [하루->종일], [시간->동안], 사람에 있어서는 [여자/남자->친구], [인간->관계], 학교에 있어서는 [연세대->학교], [시험->기간/공부]등과 같은 연결관계를 확인할 수 있었다. 전반적인 연결 양상을 살펴보면, 단편적이고 개별적 연결관계보다는 특정 키워드를 중심으로 다양한 연결 관계가 나타나는데 특히 '사람'을 축으로 복합적으로 얹혀 이어진 관계가 두드러진다. 이러한 점에서 미루어 보았을 때 긍정의 글은 과거에서 현재에 이르는 사적 영역의 경험적 요소들에 기반한다는 특징을 나타내는 듯 하다.

중립 글의 경우 [대학원->생활], [지도->교수], [연구->목표/분야] 등 대학원 관련 주제가 두드러졌고, 그 외 [임신->출산], [생존/사망->사유]와 같은 일반적 현상 및 객관적 대상에 대한 주제가 나타남을 확인할 수 있었다. 연세대학교 대나무 숲의 주 이용자가 연세대학교 학부생임을 전제했을 때, 중립의 글은 글쓴이의 현재에 대한 객관적 관심과 함께 미래에 대한 주관적 관심을 주로 담고 있는 것으로 보인다. 이는 워드 클라우드에서도 목격되는 양상인데, 중립 클러스터에서만 목격된 어휘인 '논문', '대학원', '연구', '입학', '휴학' 등을 통해서도 짐작 해볼 수 있다.

부정 글의 경우 긍정이나 중립적인 글에 비하여, 일상에서 접할 수 있는 다양한 요소들이 주제를 구성하고 있는 것으로 보인다. 구체적으로 살펴보면, 과거에서부터 현재와 미래를, 또한 사적 영역에서 공적 영역을 모두 아우르는 복합적인 주제가 관찰되는 만큼 타 클러스터에 비하여 단편적인 연결 관계가 나타난다. 다만, 키워드 네트워크 상에서 부정글은 두꺼운 화살표가 다른 클러스터들에 비해 상대적으로 많이 관측되는데, 이는 해당 감정으로 분류된 게시물의 개수가 타 클러스터 하의 게시물 개수에 비하여 적기 때문인 것으로 보인다. 지배적으로 사용되는 어휘 관계가 수치 상 상대적으로 적기 때문에 해석에서도 주의가 요구되는 부분이다.

키워드 네트워크의 연결 양상만으로 위와 같이 판단하는 것은 분석의 강건성(Robustness)을 보장하지 못할 수 있기 때문에, 본 연구팀은 LDA 분석을 추가로 진행하여 위의 판단을 뒷받침할 수 있을지 다각도로 살펴보기로 하였다.

3.3 LDA - 마지막으로, 본 연구자들은 대나무숲에서 많은 공감을 받는 주제군을 찾아보기 위하여 기존 선행 연구에서 활용된 LDA(Latent Dirichlet Allocation) 알고리즘을 이용하였다. LDA는 토픽 모델링(Topic Modeling) 기법 중 하나로, 게시자들이 글을 게시할 때 사용한 단어들이 특정 토픽에

수렴할 확률을 계산해 집합을 구성한다. 해당 방법론은 문헌 내의 잠재된 주제를 찾아내는 데 효과적인 것으로 알려져 있으며, 앞서 언급한 참고 논문을 포함한 2010년대 이후의 SNS 텍스트마이닝 관련 선행연구들에서도 활용되어 왔다. 아래의 표는, 대나무숲에서 긍정/중립/부정으로 판별된 5개의 토픽들 가운데 분석이 용이한 상위 내용 3개를 제시하며, 앞 단의 데이터와 마찬가지로 2음절 이상의 단어만 포함하였다⁵.

<표 3> 긍정/중립/부정 클러스터 별 LDA 분석 결과

긍정

| | |
|---------|--|
| 시간 (일상) | 공부, 마음, 시간, 사람, 친구, 남자, 아이, 다른, 보고, 정말, 지금, 생각 |
| 사람 (가족) | 엄마, 아빠, 우리, 아버지, 친구, 사람, 사랑, 모습, 마음, 생각 |
| 학교 (관계) | 남자, 친구, 선배, 후배, 너희, 사람, 생각, 다시, 사랑, 마음, 있는 |

중립

| | |
|----|--|
| 학교 | 휴학, 선배, 시간, 새내기, 수능, 학번, 정리, 경우, 기간, 그림, 생각, 정도 |
| 일상 | 훈련, 입학, 사유, 학교, 생각, 연구, 때문, 미안해, 엄마, 아빠, 안내견, 준비 |
| 진학 | 자신, 연구, 대학원, 교수, 박사, 시간, 하자, 생각, 경우, 사람, 많은, 있는 |

부정

| | |
|---------|--|
| 일상 (관계) | 연락, 남자, 전화, 사람, 친구, 생각, 있는, 그런, 얘기, 할아버지 |
| 사람 (가족) | 사람, 생각, 엄마, 아빠, 마음, 우리, 시간, 하나, 있는, 지금 |
| 학교 (미래) | 학번, 학우, 공부, 대학, 학벌, 우리, 블라인드, 학교, 있는, 입니다, |

본 연구팀은 수 차례에 걸쳐 음절 등 다양한 기준을 가지고 LDA 분석을 진행하였다. 이 때 주제 그룹 수를 크게 설정할 시 각각에 대해 뚜렷이 구별되는 토픽을 추출하기 어렵다는 점을 확인할 수 있었는데, 이를 통해 대나무숲에서 공감을 많이 받는 글의 내용들은 몇 가지 특정 주제로 한정되는 경향성을 띠는 것을 유추해볼 수 있다.

구체적으로 LDA 토픽모델링의 결과물을 살펴보자면 위의 표에 제시된 바와 같이, 긍정 게시물들의 경우 이전의 텍스트 네트워크의 분석 결과와 유사하게, 크게 시간(일상), 사람(가족 및 친구), 학교(관계)에 관한 내용이 주를 이루고 있음을 알 수 있다. 위의 표에 제시된 결과물을 통하여 중립과

⁵ LDA 기법은 복수의 선행연구에서 분석에 대한 결과 값과 그 해석에 신중성을 기해야 하는 방법론으로 분류 된다.

부정에서도 클러스터에서도 뚜렷한 차이가 관측되지 않는 양상을 확인할 수 있는데, 이를 통하여 대학 커뮤니티에서 일정 수준 이상의 공감을 보이는 내용들은 일상과 학교 등의 카테고리에서 한정되며, 일부 경우를 제외하고 비슷한 의미의 어휘들이 활용되어 내용이 전개된다는 것을 재확인해 볼 수 있었다. 이는 대숲대나무숲 에서 나타나는 보편의 일반적인 주제 양상을 <연애>, <관계>, <학업>, <일상생활>, <내적고민>이나 (보아즈, 2018), <연애>, <학교 생활>, <반했어요 & 짝사랑>, <고민 & 성찰>, <일상 공유>로 분류했던 기존의 분석과도 결을 같이 한다 (양민석 & 김한길, 2018).

4. 결론

본 연구에서는 온라인 익명 커뮤니티의 '감정 공유'와 '익명성'이라는 특징을 담고 있는 '연세대학교 대나무숲'에서 많은 공감을 이끌어 낸 게시물들을 수집하였다. 수집한 데이터는 텍스트 마이닝 기법을 활용하여 주요 토픽을 분석하였으며, 일련의 과정을 통하여 온라인 익명 커뮤니티에서 대학생들의 인간적 유대의 장으로서 활용되는 양상을 확인하였다 구체적으로, 해당 연구에서 커뮤니티 구성원들이 작성한 글들 중 가장 많은 공감을 이끌어낸 게시물 300건을 수집하였으며, 극성을 나누어 본 결과 긍정/부정/중립 클러스터에 제보 내용, 좋아요/댓글 수/베스트 댓글의 양상에 차이가 존재함을 확인할 수 있었다. 나아가 MeCab 분석기 통해 데이터 전처리를 진행한 후 워드 클라우드, 텍스트 네트워크, LDA를 구현해 정성적 차원에서도 다각도로 분석을 시도해보았다. 일련의 분석을 통하여 긍정/부정/중립의 각 군집을 이루는 텍스트들의 주요 토픽에 대한 보다 정교한 유추가 가능했으며, 이를 통해 연대 대나무 숲에서 공감을 많이 받는 글이 특정 주제 군의 내용에 대해 설명하며 일부 다른 양상의 이야기가 전개됨을 확인할 수 있었다.

다만, 연구를 진행하며 아쉬웠던 점은, 한글 텍스트 분석의 어려움으로 인하여 기존의 선행 연구들에 준하는 정교한 연구 설계 및 진행이 이루어지지 못했다는 점이다. 한글 텍스트 분석 기법은 현재, 영어 텍스트 분석에 비하여 분석기의 종류와 기술적 완성도가 상대적으로 낮기에 신중한 접근이 요구된다. 물론, 극성 판단에 있어서도 판별의 완성도를 높이기 위하여 재학생들을 대상으로 사전 서버이를 거치는 등 분석의 완성도를 높이려 했으나 해석에 대한 신뢰성에는 여전히 한계가 존재한다.

이런 기술적인 한계 때문에 분석대상을 연세대학교 대나무숲으로 한정하였다. 다만 페이스북을 바탕으로 하는 대나무숲의 구조 특성 상 해당 대나무숲의 구성원이 아닌 사람들도 글을 남기고 공감을 표현할 수 있는 점을 고려했을 때 다른 대나무숲에 대한 탐색으로 폭 넓은 분석을 하지 않았던 것은 아쉬움으로 남는다. 특히, 텍스트 방법론 전반에 대한 이해가 높은 연구자라면 타 대나무숲과 비교하며 일반적 양상에 대해 보다 적극적으로 비교 분석을 진행해도 무리가 없을 수 있다. 따라서, 후속 연구팀들이 온라인 익명 커뮤니티에 대한 견고한 이해를 바탕으로 여러 대학의 대나무숲을 비교해 살펴본다면 보다 통찰력 있는 분석이 가능할 것으로 기대된다.

마지막으로, 페이스북의 시기별 인기 양상을 고려했을 때 데이터가 특정 기간에 집중 분포되었을 수 있음에도 불구하고, 페이스북 유저와 대나무숲 좋아요 수 등을 통제해 선별하지 못 했다는 점 또한

개선할 부분으로 남아있다. 게시물의 시간 데이터를 수집하여 대학생들의 연도별 양상 또는 계절별 양상을 분석하거나, 게시물 내의 가장 많은 공감을 받은 댓글들의 순차적인 양상까지도 종합적으로 살펴 본다면 대나무 숲 양상에 관하여 보다 심도 있는 이해가 가능할 것으로 기대된다.

참고문헌

보아즈. (2018). [시각화] 대학 생활 연구소 - 대학별 Facebook '대나무숲' 토픽 분석. URL:

<https://www.slideshare.net/BOAZbigdata/facebook-87924887>

양민석 & 김한길. (2018). 대나무숲에 드리운 익명성의 빛과 그림자. 단대신문. URL:

<http://pdf.dknews.dankook.ac.kr/1436/143603.pdf>

진설아, 허고은, 정유경, & 송민. (2013). 트위터 데이터를 이용한 네트워크 기반 토픽 변화 추적 연구. 정보관리학회지, 30(1), 285-302.

Bargh, J. A., McKenna, K. Y., & Fitzsimons, G. M. (2002). Can you see the real me? Activation and expression of the "true self" on the Internet. *Journal of social issues*, 58(1), 33-48.

Boyd, D. M., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of computer-mediated communication*, 13(1), 210-230.

Brown, J., Broderick, A. J., & Lee, N. (2007). Word of mouth communication within online communities: Conceptualizing the online social network. *Journal of interactive marketing*, 21(3), 2-20.

Knapp, M. L., & Daly, J. A. (Eds.). (2002). *Handbook of Interpersonal Communication*. Thousand Oaks, CA: Sage.

Ma, X., Hancock, J., & Naaman, M. (2016, May). Anonymity, intimacy and self-disclosure in social media. In *Proceedings of the 2016 CHI conference on human factors in computing systems* (pp. 3857-3869)