

일간베스트 댓글에서 나타나는 혐오 표현 분석

신채원 20213021

소프트웨어학부

mirageciel@kookmin.ac.kr

요 약

본 프로젝트는 대용량의 일간베스트 댓글 텍스트를 분석하고, 자주 사용되는 혐오 표현들을 추출해내어 해당 표현들이 가리키는 혐오의 방향에 대해 분석을 진행한다. 이때 혐오 표현의 추출은 다양한 혐오 표현을 학습한 후 각 혐오 표현에 대한 빈도수를 파악하는 것이 아닌, 수집된 전체 텍스트에 대한 형태소 분석 후 높은 빈도수로 나타난 단어들 중에서 혐오 표현들을 추출해내는 방식으로 진행된다. 이를 통해 일간베스트 댓글 내에서 혐오 표현의 사용 빈도와, 혐오 표현이 향하는 방향이라는 두 가지 관점으로 연구 및 분석을 진행한다.

1. 연구목적 및 필요성

일간베스트라는 사이트는 무엇이며, 어떤 이들이 모인 커뮤니티인가? 수많은 보수 커뮤니티들 중 하나였을 뿐인 일간베스트는 2012년 대선 이후 지금과 같은 화제성을 얻게 되었다. 사회적 상식으로는 용납될 수 없는 이들의 게시글 문화가 인터넷에 퍼지고, 국정원 혹은 보수 기득권 세력을 뒤에 업고 있다는 의혹과 함께 일베는 극우 커뮤니티의 핵심으로 떠올랐다[1]. 극우 집단에서 그들 행위의 담론을 떠받치고 있는 감정은 혐오다[2]. 마찬가지로 일간베스트 내에는 혐오가 만연하며, 가장 대표적으로 보이는 것이 지역 혐오와 여성 혐오, 소수자 혐오이다. 일간베스트가 지금과 같은 위치를 갖게 되며, 커뮤니티 내에서 나타나는 혐오에 대해서는 꾸준한 연구가 진행되어 왔다.

그러나 사이트 내에서만 사용되던 혐오 표현들이 SNS를 통해 퍼져 나가면서 현실에서도 사용되는 빈도가 늘어나고 있다. 사람들은 어느 순간 일베에 관한 솔한 담론들에 익숙해졌으며, 일상적 경험 속에서 그것의 현실적 ‘힘’을 믿기 시작했다[3]. 최근에는 일간베스트 내에서 전 대통령을 조롱하기 위해 말끝마다 ‘-노’를 붙이는 현상이 현실에서의 유형이 되었으며, 해당 표현의 정확한 뜻도 알지 못한 채 ‘유형’이라는 이유로 사용하는 사람 역시 많아졌다.

일간베스트 내의 혐오가 현실에까지 영향을 미치는 지금, 그들의 혐오는 더 이상 일간베스트라는 커뮤니티 내의 문제가 아니게 되었다. 이러한 문제를 해결하기 위해 일간베스트 내의 댓글에서 등장하는 혐오 표현을 분석하고자 한다. 우선 댓글에서 높은 빈도로 나타나는 단어들 중 혐오 표현들을 정리하고, 해당 혐오 표현들의

정확한 의미와 이들이 향하고 있는 혐오의 방향을 분석하고자 한다. 이러한 과정을 통해 혐오의 본질적인 해결책에 대해 고민하고 탐구해 본다.

2. 연구배경 및 관련 연구

온라인 혐오 표현이 규모가 늘어나고 주류화 됨 [3]에 따라, 온라인 상에서 보이는 혐오 표현에 대한 연구들은 꾸준히 진행되어 오고 있다. 이러한 연구들은 주로 특정 주제나 인종, 성별, 성적 지향, 종교, 국적 등의 차이에 대한 혐오 발언에 초점을 두고 연구를 진행하였다.

여러 곳에서 온라인 혐오 표현들이 보이고 있지만, 한국 사회에서 나타나고 있는 온라인 혐오 표현의 대표적 사례는 ‘일베 현상’으로 설명[4]된다. ‘일베 현상’이란 ‘일간베스트 저장소’라는 이름의 인터넷 커뮤니티가 사회적 관심을 끌면서 나타난 일련의 현상을 지칭한다[5]. 국내 논문들을 살펴보면 이러한 일간베스트라는 특정 사이트에서 보이는 혐오들을 분석한 연구들 역시 다수 존재함을 알 수 있다.

이러한 연구들은 주로 특정 주제, 혹은 이슈와 혐오 표현 사이의 연관성을 분석하는 방향으로 이루어져 왔다. 일간베스트 사이트 내에서 나타나는 혐오를 인지하고, 해결 방안을 고안해내는 것에 대한 논의 역시 활발하게 진행되고 있는 중이다.

이러한 연구들을 바탕으로 본 프로젝트에서는 일간베스트 댓글에서 나타나는 혐오 표현들의 종류와 그 빈도수를 분석하고, 이를 바탕으로 일간베스트 내의 혐오의 발생 환경과 혐오의 방향을 분석하며, 이에 대한 해결책을 모색해보기로 한다.

3. 방법론

해당 프로젝트는 대용량의 일간베스트 댓글 데이터를 수집하고, 이들을 분석하여 높은 빈도수를 보이고 있는 표현들을 정리하고 해당 표현들의 혐오의 방향을 확인하기로 한다.

3.1 데이터 수집

해당 프로젝트에서는 국민대학교 강승식 교수님의 네이버 카페인 nlp study 카페에 올라온 일간베스트 댓글 데이터셋을 사용하기로 한다. 해당 데이터 세트는 일간베스트 사이트에 올라온 각 게시글의 좋아요 수, 싫어요 수, 날짜, 댓글에 대해 수집이 되어 있는 상태이다. 데이터는 총 1,196,345개로, 1,388,264개의 댓글을 수집한 후 중복이 제거되었다.

본 프로젝트에서는 해당 자료의 좋아요 수, 싫어요 수, 날짜, 댓글이라는 4개의 항목 중 ‘댓글’에 해당되는 데이터만을 추출해서 분석을 진행하였다.

댓글 데이터를 확인한 결과 완전한 단어(ex. 인정) 대신 초성 글자의 형태로 작성(ex. ㅇㅈ)하거나, 감정 표현을 반복되는 초성 글자의 나열(ex. ㅋㅋ)이라는 형태로 작성하고 있음을 알 수 있었다. 또한 다양한 줄임말(ex. 싸이코패스 대신 싸패 사용)과 신조어(ex. 흙수저)를 사용하는 모습을 보이고 있었다. 또한 온라인 커뮤니티라는 특성 상 문장의 띄어쓰기 및 맞춤법이라는 측면에서 많은 오류가 존재함을 볼 수 있었다.

이러한 특징들은 형태소 분석을 진행할 때 부정확한 결과로 이어질 수 있으므로, 텍스트 전처리 단계에서 맞춤법 검사 및 교정을 함께 수행해주어야 할 필요가 있다.

3.2 데이터 전처리 및 형태소 분석

형태소 분석이 정확하게 이루어지지 않는 문제를 방지하기 위해, 데이터 전처리 단계에서는 불필요한 특수문자 제거와 맞춤법 검사 및 교정을 함께 수행해 주었다. 이러한 작업에는 파이썬 내장 라이브러리인 re와 네이버 맞춤법 검사기 api를 사용하여 구현한 한글 맞춤법 검사 라이브러리인 hanspell이 사용되었다.

데이터 전처리를 위해 우선 대문자 A부터 Z, 소문자 a부터 z, 숫자 0부터 9, 한글 완성형 인코딩의 한글 구현 범위인 ‘가’부터 ‘힣’, 한글 초성 ㄱ부터 ㅎ까지를 제외한 모든 문자들을 re 라이브러리를 사용하여 제거하였다. 이는 맞춤법 검사기를 실행할 때 맞춤법 검사기가 문장을 이해하지 못하고 맞춤법 검사를 수행하지 못할 확률을 조금이라도 낮추기 위해서 진행된 작업이다.

파이썬에서 re.sub()을 이용하여 특수문자를 모두 제거하면, 불필요한 문자는 모두 제거가 완료된 상태이지만, 맞춤법과 띄어쓰기가 정확하지 않은 탓에 여전히 형태소 분석에서 어려움을 겪을 것임을 예측할 수 있다. 이를 해결하기 위해 파이썬의 hanspell 라이브러리를 사용하여 맞춤법 검사를 수행해 주었다.

hanspell이 교정할 수 없는 문장에 대해서는 예외 처리를 통해 별도의 파일에 저장하였다. 해당 파일 내의 문장은 수동으로 맞춤법을 교정한 후 형태소 분석을 진행할 파일에 추가하는 방식으로 진행하였다.

맞춤법 교정 작업을 통해서 올바르지 않은 방향으로 교정된 단어에 대해서는 추후 word2vec을 이용한 워드 임베딩 작업에서 min_count 인자값 설정을 통해 학습에서 제외시킬 계획이다.

해당 데이터를 단어 단위로 빈도 분석을 진행하기 위해 각 문장들에 대해 형태소 분석을 진행하였다. 형태소 분석 작업에 대해 코드 내에 두 가지 옵션을 두었다. (1) konlp.kma의 Kkma를 사용한 형태소 분석, (2) KLT2000의 index2018.exe를 사용한 형태소 분석이 바로 그 두 가지이다.

분석 대상이 되는 텍스트 파일은 utf-8 인코딩으로 작성되어 있기에, (2)번 방법으로 형태소 분석을 진행하기 위해서는 utf-8에서 cp949로의 인코딩 변환이 우선 실행되어야 한다. 분석 결과 파일 역시 cp949로 저장되기 때문에, 이후의 워드 임베딩을 위해서는 다시 utf-8로 인코딩 변환을 해 주어야 한다. (1)번 방법의 경우 결과 파일이 utf-8로 저장되기 때문에 별다른 인코딩 변환 없이 그대로 진행하면 된다.

3.3 빈도 분석

전체 댓글 중에서 혐오 표현이 차지하는 비율을 알기 위해, 단어들의 등장 횟수를 계산하는 작업을 진행하였다. 단어들의 빈도수 계산 작업에는 pandas 라이브러리를 사용하였다.

높은 빈도로 출현하는 단어들 중 상위 100개를 출력한 후, 다시 해당 단어들 중에서 혐오를 내재하고 있는 단어들을 별도 추출하여 저장하였다. 그 결과 상위 100개의 단어들 중에서 총 16개의 혐오 표현이 추출되었다. 아래의 표는 각 혐오 표현에 대한 등장 횟수와 전체 빈도 분석에서의 순위를 정리한 것이다.

순위	단어	출현 횟수	전체 순위
1	새끼	108446회	1위
2	병신	43295회	11위
3	년	29765회	24위
4	ㄹㅇㅎ	29679회	25위

5	시발	29157회	26위
6	개	19979회	47위
7	흥어	19900회	48위
8	늬	19258회	52위
9	여자	18015회	60위
10	전라도	16064회	67위
11	게이	14692회	77위
12	지랄	14117회	79위
13	중국	12302회	91위
14	대통령	11947회	94위
15	좃	12302회	96위
16	문재인	11666회	98위

<표 1> 혐오 표현 출현 빈도수 계산 결과
표에서도 나타나듯이, 높은 빈도로 등장하고 있는 상위 100개의 단어들 중 혐오 표현이 약 15%를 차지하고 있음을 볼 수 있다. 이를 통해 일간베스트 사이트 내에서 혐오 표현은 결코 적지 않게 사용되고 있음을 알 수 있다.

3.4 워드 임베딩 및 결과 시각화

각 혐오 표현들에 대해, 이들이 다른 단어들과 어떻게 연결되어 있는지를 알아보기 위해 유사어를 탐색하고, 이에 대한 결과를 시각화 하고자 했다

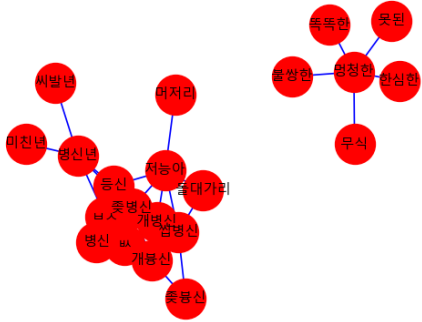
유사어 탐색을 위한 워드 임베딩은 word2vec을 통해 이루어졌으며, word2vec.Word2Vec() 함수의 min_count 인자를 2로 설정함으로써 잘못 교정된 단어는 학습에서 제외되도록 하였다. 워드 임베딩 결과는 *.model의 형태로 저장되며, 해당 파일과 word2vec의 most_similar() 함수를 이용하여 유사어 탐색을 진행하였다. 다음은 혐오 표현 중 일부에 대해 유사어 탐색을 진행하고, 그 결과로 얻은 유사어와 유사도를 정리한 표이다.

병신 (혐오 표현 2위, 전체 11위)	
유사어	유사도
좃병신	0.806771457195282
등신	0.779389500617981
개병신	0.6732909679412842
ㅂㅅ	0.6521783471107483
저능아	0.6465569734573364
병신년	0.5834298133850098
씹병신	0.5703725218772888
멍청한	0.5504761934280396

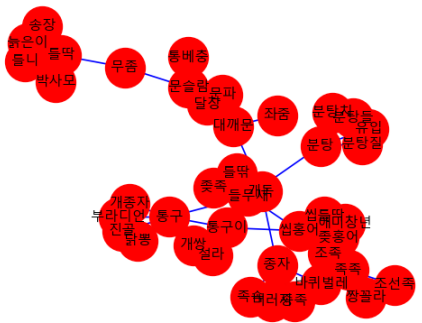
흥어 (혐오 표현 7위, 전체 48위)	
유사어	유사도
분탕	0.617926299571991
통구이	0.6037446856498718

들무새	0.6024805307388306
씹흥어	0.5826643705368042
대깨문	0.5765116810798645
틀딱	0.5649189352989197
종자	0.5580267310142517
문슬람	0.5476859211921692

<표 3> ‘흥어’ 유사어 탐색 결과
유사어 탐색에 대한 시각화는 networkx와 matplotlib을 이용하여 그래프의 형태로 나타나도록 하였다. 처음에는 주어진 혐오 표현에 대해 유사어를 탐색하도록 하고, 다시 각 유사어에 대해 새로운 유사어를 찾아서 이들을 그래프의 형태로 그리도록 하였다.
아래의 두 그림은 ‘병신’과 ‘흥어’라는 두 단어의 유사어 탐색에 대해 시각화를 진행한 결과이다.



[그림 1] ‘병신’ 유사어 탐색과 시각화



[그림 2] ‘흥어’ 유사어 탐색과 시각화
각 혐오 표현의 유사어로 또 다른 혐오 표현이 등장하며, 해당 유사어에 대한 유사어로도 혐오 표현 혹은 부정적인 의미를 내포하고 있는 단어가 검색됨을 볼 수 있다.

4. 결론

일간베스트 댓글에서 보이는 혐오 표현은 습관적이며, 다소 맥락이 부족하기까지 하다. 이는 높은 빈도수를 보이는 상위 100개의 단어들 중에서 혐오 표현이 15%나 차지하고 있었다는 분석 결과와 이어지고 있다.

또한, 일간베스트 내에서 사용되는 표현들은 단순한 욕설에서 그치는 것이 아니었다. 단순히 사용자의 분노 혹은 부정적인 감정을 표현하기 위해 사용되는 욕설일지라도, 그에 대한 유사어로는 지역 혐오 혹은 약자 혐오의 의미를 가지고 있는 단어들이 검색되고 있었다.

즉 이들의 욕설은 지역과 약자 혐오로 이어지며, 더 나아가 이들을 특정하여 비하하는 새로운 단어를 창조하기까지 이른다. 이는 사이트 내에서 사용되는 모든 혐오 표현은 상대를 향한 분노를 넘어서서, 혐오 표현이 가리키는 본질적 대상까지 혐오하고 있다는 것으로 해석할 수 있다.

습관적으로 사용하는 혐오 표현은 설령 목적이 혐오가 아니었다고 할지라도, 결국은 또 다른 혐오를 낳는다. 혐오 표현의 사용률이 특히나 높게 나타나는 일간베스트 사이트에 대해 경각심을 갖고, 이러한 커뮤니티 내에서 혐오를 줄일 수 있는 방법에 대해 추가적으로 논의해야 할 필요가 있다.

5. 참고문헌

- [1] 윤보라, 2013, 「일베와 여성 혐오: “일베는 어디에나 있고 어디에도 없다”」, 『진보평론』 가을 제57호, 뉴 래디컬 리뷰.
- [2] 석승혜, 장안식, 2017, 「극우주의의 프레임과 감정 정치: 언어네트워크방법론을 통한 일베커뮤니티 분석」, 『한국사회』 18-1, 고려대학교 한국사회연구소.
- [3] 이길호, 2014, 「‘일베’를 어떻게 인지할 것인가」, 『시민과세계』 25, 참여연대 참여사회연구소.
- [4] 김민정, 2014, 「일베식 “욕”의 법적 규제에 대하여: 온라인상에서의 혐오 표현에 대한 개념적 고찰」, 『언론과법』 13-2, 한국언론법학회.
- [5] 홍성수, 2013, 「표현의 자유와 한계: 혐오표현 규제의 정당성과 방법」, 대법원·법무부 국제인권법연구회 공동학술대회 발표문.
- [6] 강승식, 2017, 「한국어 원시 말뭉치 - 대용량 텍스트 파일」, nlp study - 정보검색, 자연어처리, 빅데이터, 기계학습, <https://cafe.naver.com/nlpk>