

댓글 정렬 방식이 댓글러의 분포와 댓글 내용의 다양성에 미치는 영향

2022. 5. 13.
연세대학교 최고은, 이상엽

연구 배경

2

□ 연구 배경

- ▣ 주요 포털 (네이버, 다음 등)에서 제공되는 신문기사에 달리는 댓글 정렬방식의 차이
- ▣ 각 기사의 댓글 정렬방식을 신문사가 직접 결정
- ▣ 특히, 정치 섹션의 기사의 댓글 정렬 방식 차이 존재
 - 즉, 일부 언론사는 최신순만, 다른 언론사들은 최신순 + 순공감순

신문사	경향신문	국민일보	동아일보	문화일보	서울신문	세계일보	조선일보	중앙일보	한겨레	한국일보
정렬방식	최신순	최신순	최신순	순공감순 최신순 공감비율순	순공감순 최신순 공감비율순 답글순 과거순	순공감순 최신순 공감비율순	최신순	순공감순 최신순 공감비율순 답글순 과거순	순공감순 최신순	순공감순 최신순

〈기준 = 네이버〉

연구 배경

3

□ 연구 배경

- ▣ 다른 사람들이 공감한 정도에 따라 독자들이 댓글을 인식하는 방식이
다름
 - 베스트 댓글을 주요한 여론으로 간주 (이은주, 2011)
 - 침묵의 나선이론 (Noelle-Neumann, 1974) 에 따르면
베스트 댓글과 다른 의견을 갖는 사람들의 경우는 베스트 댓글을 다수의
의견으로 간주할 수 있기 때문에 의견을 잘 표현하지 않을 가능성 존재
 - 이는 정렬 방식에 따라서 댓글을 통해 나타나는 의견이나 내용의 다양성이
달라질 수 있다는 것을 의미

연구 목적

4

□ 연구 목적

- ▣ 정치 기사의 댓글 정렬 방식에 따라서
댓글을 다는 사람들 (이하, 댓글러)의 분포와 댓글이 달리는 정도,
그리고 댓글 내용의 다양성이 어떻게 달라지는지를 분석
 - 댓글러 분포
 - 성별
 - 연령별
 - 댓글이 달리는 정도
 - 댓글의 수
 - 대댓글의 수
 - 댓글 당 대댓글 수
 - 댓글 내용의 다양성

주요 선행 연구

5

- Social media 또는 SNS 상에서도
침묵의 나선이론이 적용된다는 선행연구들 다수 존재
 - ▣ ‘지각된 의견지지’ 정도는 SNS 에서의 사회적담론 관련 의견 표현에 정의 영향을 미친다. 즉, ‘지각된 의견지지’ 정도가 클수록 의견 표현을 더 적극적으로 한다 (김은진, 2019; 신동인·곽기영, 2019).
 - ▣ Facebook 상에서 인지된 반대 의견의 정도가 클수록 정치적 표현 자제 (Kim, 2016)
 - ▣ SNS 상에서 자신의 의견과 동일한 정치적 콘텐츠에 노출될수록 본인의 의견을 적극적으로 표현하는 반면, 반대되는 정체적 콘텐츠에 노출되는 경우 의견 표현 자제 (Gearhart & Zhang, 2015)
- 하지만, 신문기사 댓글 관련 연구는 부재

연구 방법

6

□ 데이터

▣ 신문기사 수집

- 기간: 2021.9.1 – 2021.12.31
- 대상: 주요 일간지 (9개의 신문사)
- 데이터 수집 포털: 네이버

- 각 신문사에 대해 일별로 사람들이 많이 읽은 5개의 기사를 수집
 - 전체 기사수: 5,040
 - 전체 댓글수: 5,235,244

- 이 중 정치 기사만 선별
 - 정치기사의 수: 1,303

연구 방법

7

□ 데이터 (cont' d)

▣ 각 기사의 댓글 및 댓글러 정보 수집

405개의 댓글

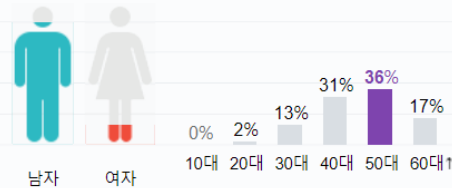
접기 | 내 댓글

현재 댓글
380

작성자 삭제
23

규정 미준수
2

누가 댓글을 썼을까요?



댓글을 작성하려면 로그인 해주세요

인상 깊었던 댓글, 이젠 팔로우 해보세요.

순공감순 최신순 공감비율순 답글순 과거순



- 전체 댓글수
- 삭제된 댓글수
- 규정 미준수 댓글수
- 성별 분포
- 연령별 분포
- 정렬 방식

연구 방법

8

□ 데이터 (cont' d)

▣ 각 댓글에 대해서 다음과 같은 정보 수집

**bart****** 
2022.04.27. 20:56:01

저 때 검사들이 고졸 대통령 업신여기는게 너무 인상에 깊게 남아서 검사 집단에 대한 부정적 이미지 발단 됐고, 대통령 수사 당시 창문에 모여서 비웃는 듯한 사진 찍히게 검찰 악마화에 결정적 계기 됐습니다.

답글 17  603  66

**jjun****** 
2022.04.27. 21:23:05

국민이 뽑은 대통령 공개적으로 무시한 대표적인 검사. 저 검사들 보고 경악했죠

 37  4

moon****
2022.04.27. 21:33:06

① 작성자에 의해 삭제된 답글입니다.

**mrha****** 
2022.04.27. 21:38:46

그게 문재인운동권들이 검사에게 씌운 프레임이죠 고졸이라고 업신여기고 조롱한다는건 자신들의 생각이고 증거가 있는 근거있는 말도 아닌데 그 말로 선동해 검사를 악마화프레임을 씌운것이 팩트임

 30  21

- 댓글
 - 댓글러 아이디
 - 댓글 작성 시간
 - 댓글 내용
 - 대댓글의 수
 - 공감 정보
- 대댓글
 - 댓글러 아이디
 - 각 대댓글의 내용
 - 공감 정보

연구 방법

9

□ 데이터 (cont' d)

- 분석에서 신문사의 정치 성향을 고려하기 위해 다음과 같이 정렬 방식이 서로 다른 보수 2개, 진보 2개 신문사를 최종 선택 (총 기사수 = 634)



東亞日報

경향신문

중앙일보

한겨레
hani.co.kr

□ 데이터 (cont' d)

- 분석에서 신문사의 정치 성향을 고려하기 위해 다음과 같이 정렬 방식이 서로 다른 보수 2개, 진보 2개 신문사를 최종 선택 (총 기사수 = 634)

신문사	정치 성향	댓글 정렬 방식
동아일보	보수	최신순
중앙일보	보수	순공감순 + 최신순
경향신문	진보	최신순
한겨레	진보	순공감순 + 최신순

□ 분석 방법

- ▣ 종속변수의 유형에 따라 서로 다른 회귀모형 사용
- ▣ 종속변수가 fractional 변수인 경우 (즉, 취하는 값이 0 -1 인 경우), fractional logit 모형 사용
 - 여성 댓글러 비중, 연령별 댓글러 비중
- ▣ 그 외 종속변수의 경우,
선형회귀모형 with robust standard error estimation
- ▣ 주요 독립변수: 순공감순 배열 여부 (순공감순 배열시 1, 아니면 0)
- ▣ 주요 통제변수: 언론사 ID (언론사의 특성 (예, 정치 성향)을 통제하기 위해 사용)

□ 분석 방법

▣ 군집화

- 내용의 다양성을 파악하기 위해 군집화 방법 사용
- 군집화를 위해 각 댓글을 벡터로 표현
 - Doc2vec 방법 사용 ⇒ 학습데이터의 양이 많을 수록 더 정확한 결과를 얻을 수 있기 때문에 5,040 기사의 댓글을 모두 사용 (전체 댓글수 = 5,235,244)
 - Doc2vec은 특정 단어를 기준으로 이웃한 단어들이 유사할수록 문서(댓글) 벡터의 원소값이 유사
 - 즉, 벡터가 유사할수록 내용이나 사용된 단어가 유사
- 군집화 방법
 - K-Means, Agglomerative hierarchical clustering, DBSCAN

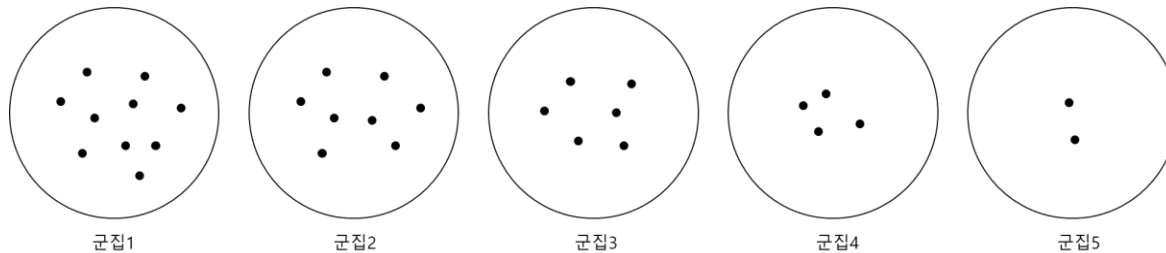
□ 분석 방법

▣ 군집화를 이용한 내용의 다양성 측정

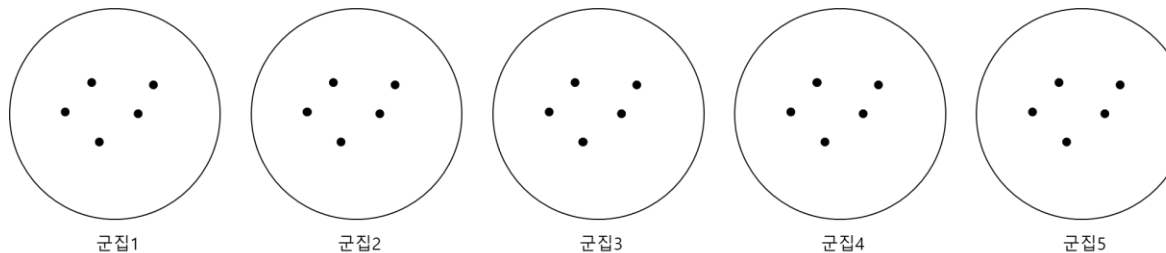
- K-Means 와 Agglomerative HC 방법은 찾고자 하는 군집의 수를 사용자가 지정
 - 본 연구에서는 10개로 지정
 - 같은 군집에는 상대적으로 내용이 유사한 댓글들이 포함
 - Agglomerative HC에서 linkage 방법은 Ward 사용 (Lelu & Cadot, 2019)
- 본 연구에서는 각 군집에 속하는 댓글수의 표준편차를 이용하여 내용의 치우침 정도를 파악 (다음 슬라이드 참고)

□ 분석 방법

- 본 연구에서는 각 군집에 속하는 댓글수의 표준편차를 이용하여 내용의 치우침 정도를 파악
 - 표준편차가 큰 경우 (아래는 예시) ⇒ 댓글 내용이 한쪽으로 치우치는 경향



- 표준편차가 작은 경우 ⇒ 댓글 내용이 균등하게 분포



□ 분석 방법

▣ DBSCAN

- 군집의 수 : 관측치의 유사도를 이용하여 관측치들을 연결하며 찾음
- DBSCAN을 이용한 내용의 다양성 측정
 - 군집의 수와 노이즈(noise)* 댓글수 사용
 - 군집의 수가 적다 = 유사한 의견/내용을 가진 댓글 그룹이 적다
 - 군집의 수가 많다 = 유사한 의견/내용을 가진 댓글 그룹이 많다
 - 노이즈 댓글의 수 증가
= 다른 의견과 유사도가 떨어지는 댓글이 더 많이 존재한다

노이즈 (Noise): 어떠한 군집에도 속하지 못한 관측치

→ 즉, 본 연구에서는 다른 댓글들과의 유사도가 떨어지는 댓글을 의미

결과

16

□ 정렬방식에 따른 댓글러 비중

Variable	댓글러 비중						
	여성댓글러	10대	20대	30대	40대	50대	60대 이상
순공감순정렬	0.269**	12.124**	0.867**	0.581**	0.038*	-0.146**	-0.184**
중앙일보	-0.076	-12.305**	-0.716**	-0.228**	0.238**	0.030	-0.253**
한겨레	-0.344*	-12.271**	-1.076**	-0.652**	0.166**	0.270**	-0.136**
경향신문	0.035	0.042	0.035	0.138**	0.259**	0.055*	-0.468**

Note. * $p < 0.05$, ** $p < 0.01$, $N = 634$

결과

17

□ 댓글과 리액션 관련

Variable	전체 댓글수	삭제된 댓글수	대댓글수	전체 리액션수	댓글 길이 (평균)	댓글 길이 (중앙값)
순공감순정렬	777.68**	233.04**	161.25**	1220.81**	-3.44**	-3.69**
중앙일보	401.55	65.568	163.12**	1418.45*	0.78	1.14
한겨레	-1118.69**	-347.26**	-180.61**	-3381.73**	3.17**	3.07**
경향신문	-199.64	-78.80*	75.823**	-1377.28**	-2.25**	-1.69**

Note. * $p < 0.05$, ** $p < 0.01$, $N = 634$

결과

18

□ 군집화 결과

Variable	KMeans std	Agglomerative std	DBSCAN 군집수	DBSCAN 노이즈수
순공감순정렬	0.002**	.0146**	-0.621**	-12.522**
중앙일보	-0.009**	-0.009**	-0.053	5.202
한겨레	-0.002**	-.0160**	0.800**	17.805**
경향신문	-0.001**	-.001	0.025	4.618

Note. * $p < 0.05$, ** $p < 0.01$, $N = 634$

결과

19

□ 댓글과 대댓글 관계 관련

Variable	댓글당 대댓글수	댓글당 대댓글수 표준편차	댓글과 대댓글간 거리 평균
순공감순 정렬	0.046**	0.402**	0.026**
중앙일보	-0.007	0.882**	-0.041**
한겨레	-0.024**	0.420**	-0.036**
경향신문	0.055**	0.125**	-0.011**

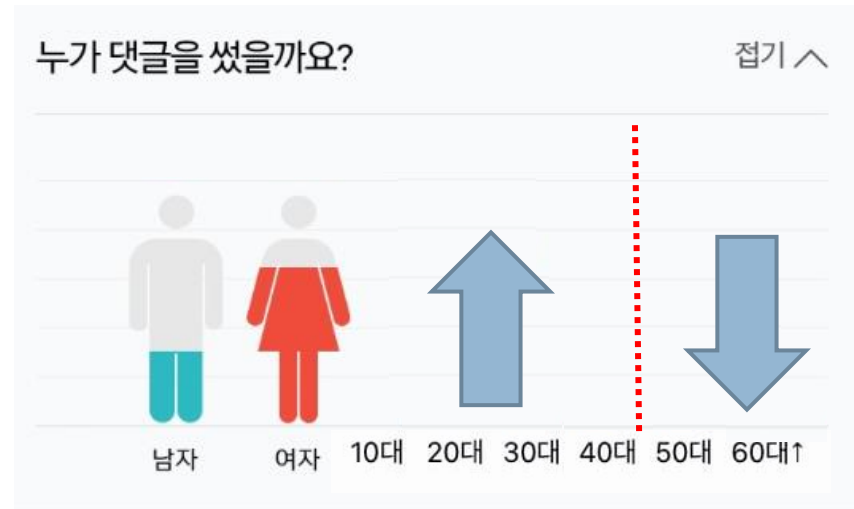
Note. * $p < 0.05$, ** $p < 0.01$, $N = 634$

□ 분석 결과 해석

▣ 순공감순 배열이 가능한 경우 (최신순 배열만 가능한 경우와 비교)

■ 댓글을 다는 사람들

- 여성 비중 ↑
- 10~40대 비중 ↑
- 5, 60대 비중 ↓



□ 분석 결과 해석

▣ 순공감순 배열이 가능한 경우 (최신순 배열만 가능한 경우와 비교)

■ 댓글

- 전체 댓글수, 삭제된 댓글수, 대댓글수, 전체리액션수 ↑
- 댓글당 대댓글수 증가, 댓글과 대댓글 사이의 내용 차이 ↑
- 댓글의 길이 ↓
- 내용의 다양성 ↓: 침묵의 나선이론 적용



- 뉴스 댓글이 중요함에도 불구하고 정렬 방식 연구 미흡
→ 해당 현상을 보다 잘 이해할 수 있는 연구 결과 제시
- 침묵의 나선이론
→ 댓글에 해당 이론이 적용된다는 것을 실증적으로 밝힘
- 댓글의 다양성 측정 위해 기계학습 기반의 방법 사용
→ 연구의 강건성 제고

Thank you!