유튜브 댓글 성별 성향 분석

DeepSharkLab 학부연구생 이지언

연구 배경 및 목적

기존인식	연구 질문	연구 목적
• 유튜브 채널마다 시청자 성별 성향이 뚜렷하게 다름 • Ex) 뷰티/브이로그 채널은 여성 시청자가 많고, UFC/게임 채널은 남성 시청자가 많음	• 댓글 텍스트만으로 채널 시청자의 성별 성향을 파악할 수 있는가?	• 댓글 말투 패턴을 데이터로 직접 학습해서 채널별 성향을 시각화

데이터 소개

유튜브 채널 = 10개

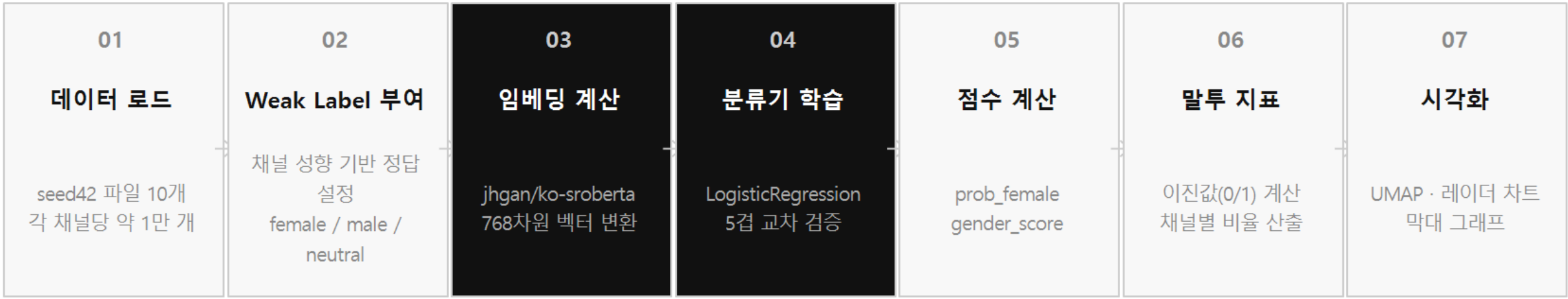
채널당 댓글 수 = 각 10,000개

최종 분석 댓글 (2글자 이하 제거) = 98,571개

랜덤 샘플링 고정값 = seed42

성향	채널명	콘텐츠 유형	분류 근거
여성향	힙으뜸	러닝 / 피트니스	여성 운동 채널, 국내 1위 여성 크리에이터
여성향	해쭈 [HAEJOO]	커플 브이로그	코미디·일상·요리 콘텐츠, 호주 거주 한국인
여성향	회사원A	뷰티 / 메이크업	메이크업·헤어 + 개그 요소 뷰티 채널
여성향	예랑가랑	라이프스타일 브이로그	부산 여대생 둘이 운영하는 일상 채널
남성향	피지컬갤러리	피트니스 / 도전	전직 UDT 출신 김계란, 가짜사나이 제작
남성향	차도르	UFC / 격투	종합격투기 전문, tvN UFC 해설위원
남성향	모트라인 MOTline	자동차	하체·구조 분석 + 서킷 리뷰 자동차 채널
남성향	랄로	게임 (LCK/롤)	롤 방송 BJ, 경상도 사투리 + 걸걸한 입담
남성향	[꿀템TV] 클템 유튜브	LCK / 롤 분석	LCK 해설위원 CloudTemplar 채널
중립	애타【ETA】	오타쿠 / 서브컬처	버튜버 합동방송, 게임·그림·수다

전체 분석 파이프라인



데이터를 로드해서 채널 성향 라벨을 붙이고, 한국어 임베딩 모델로 댓글을 768차원 벡터로 변환. 그 벡터를 입력으로 분류기를 학습시켜서 댓글마다 여성향 확률을 계산하고, 말투 지표도 별도로 계산한 뒤, UMAP으로 시각화.

WEAK LABEL이란?

- **문제 상황**

  개별 댓글에 어떤 성별이 썼다는 정답이 없음

  유튜버가 시청자 성별 정보를 공개하지 않음

  성별을 안다고 해도 댓글 하나하나에 라벨을 붙이는 건 현실적으로 불가능

- **해결 (Weak Label 사용)**

  채널 단위 시청자 성향을 댓글의 정답으로 활용

  회사원 A 채널 댓글 전체 -> female 라벨

  차도르 채널 댓글 전체 -> male 라벨

  이를 Weak Supervision(약한 지도학습)이라 함

-장점

라벨링 비용 없이 대규모 학습 데이터 확보 가능

-단점

노이즈 포함 (같은 채널에도 반대 성별 존재)

임베딩 & 분류기 학습

임베딩 (Embedding)

모델	jhgan/ko-sroberta-multitask
특징	한국어 특화 Sentence-BERT 수백만 개 한국어 문장으로 사전 학습
변환	댓글 1개 → 숫자 768개 벡터 의미가 비슷한 댓글 = 가까운 벡터
출력	(98,571 × 768) 행렬

분류기 학습

알고리즘	LogisticRegression
입력 X	98,571 × 768 임베딩 벡터
정답 y	female / male / neutral

분류기 결과 - PROB_FEMALE

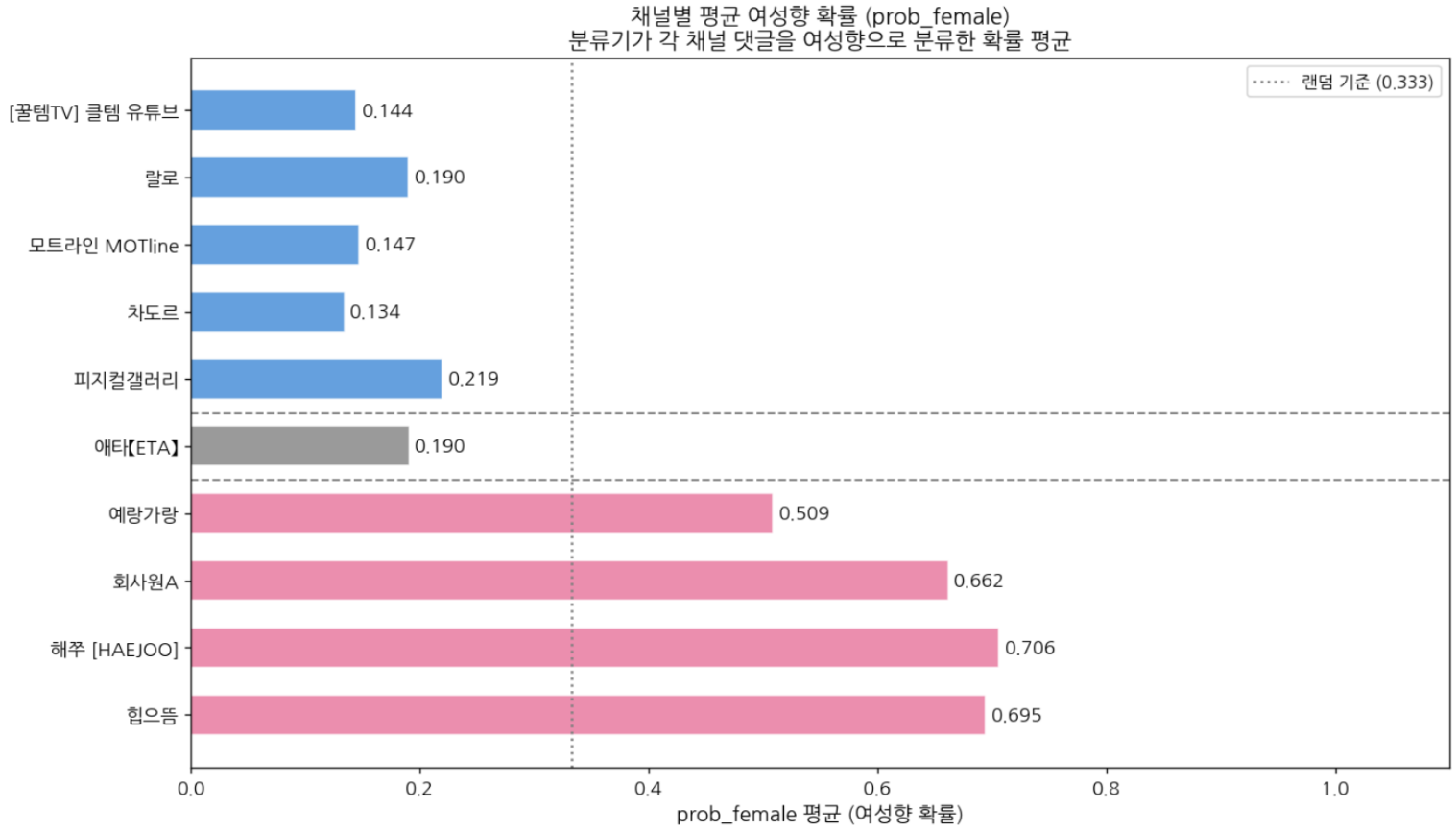
- prob_female이란?

prob_female = 분류기가 이 댓글을 female로 예측할 확률 (0~1)

prob_male = 분류기가 이 댓글을 male로 예측할 확률 (0~1)

gender_score = prob_female – prob_male (-1~1) <- 양수면 여성향, 음수면 남성향

분류기 결과 - PROB_FEMALE



5가지 말투 지표 설계

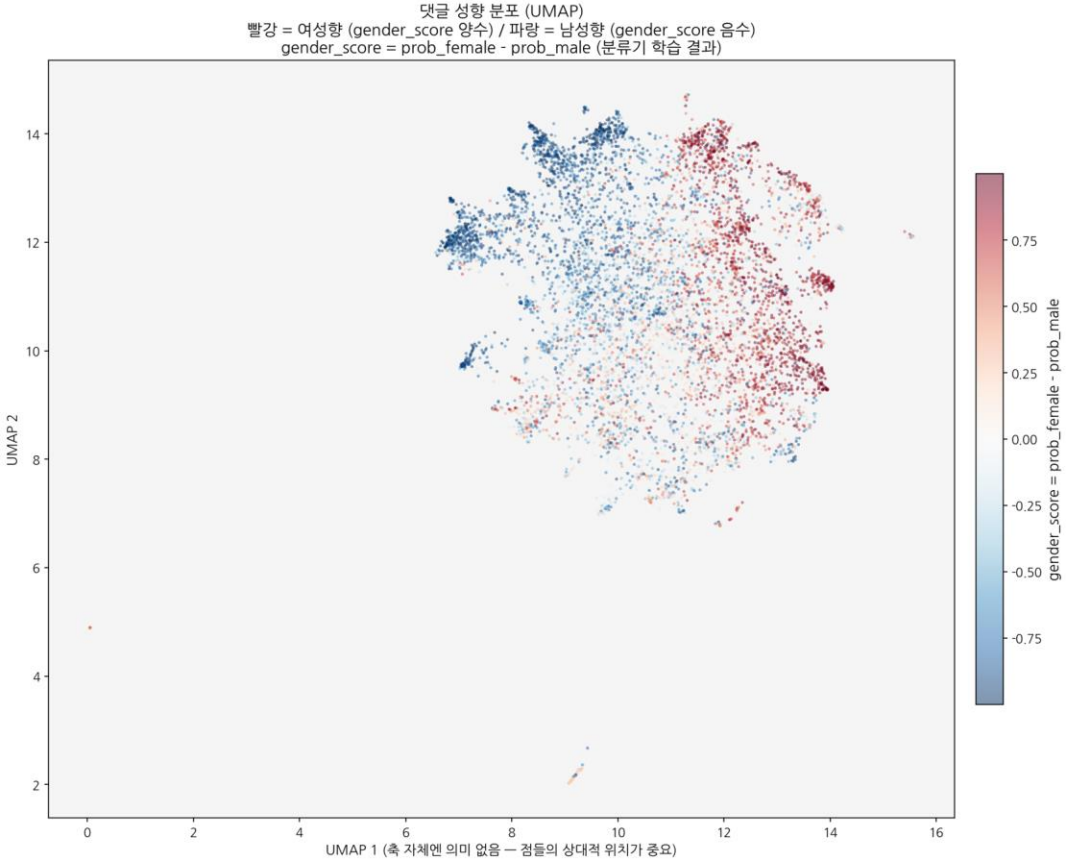
지표	구성 요소 (댓글에 포함되면 해당)	데이터 근거 (여-남 차이)	방향
감정표현	느낌표 (!) · 하트이모지 ♥💗· ㅠㅠ/ㅜㅜ · 감성이모지 😮‍💨	+14.6%	여성향 ↑
친밀감	님 호칭 · 언니 호칭 · 물결 (~)	+6.7%	여성향 ↑
공감성	"너무" · "진짜" · "저도"	+13.3%	여성향 ↑
정보성	30자 이상 댓글 · 숫자 포함	+18.7%	여성향 ↑
공격성	ㅅㅂ · ㅂㅅ · 존나 · 개소리	+0.7%	남성향 ↑

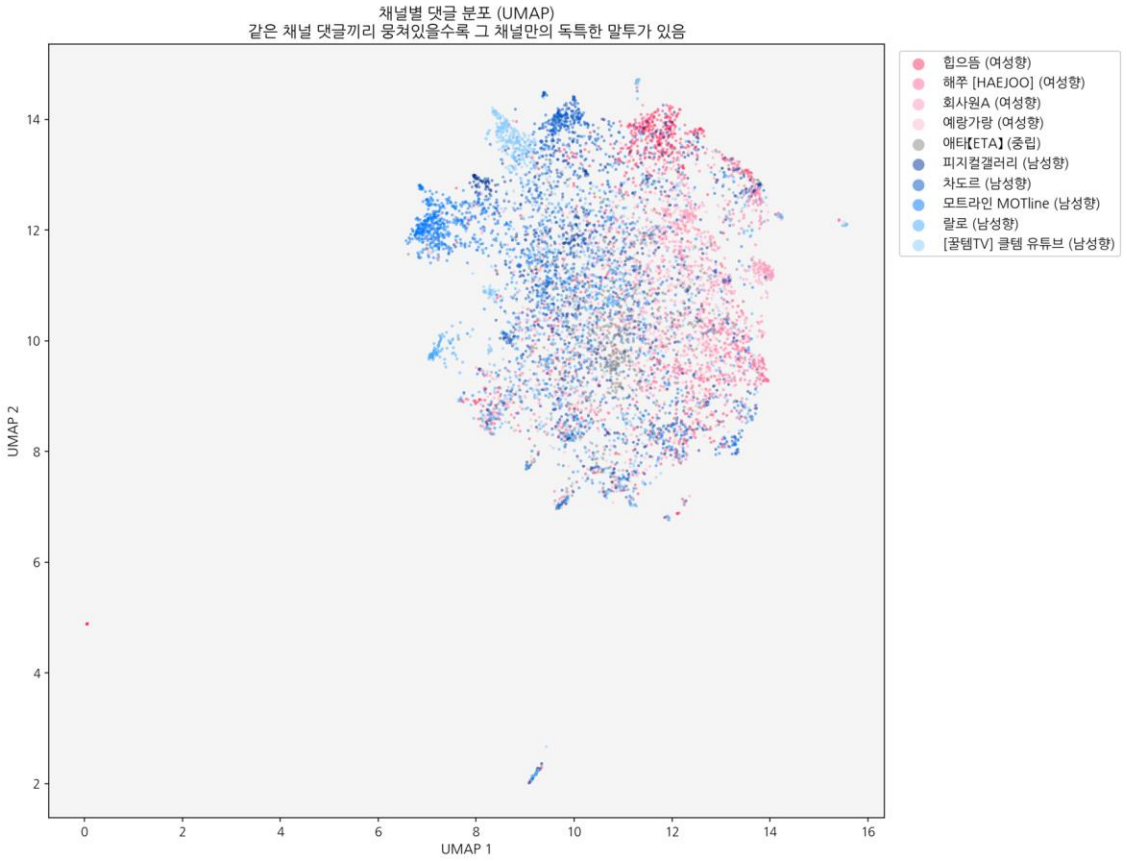
채널별 지표 비율 결과

채널별 지표 비율 (단위: %)

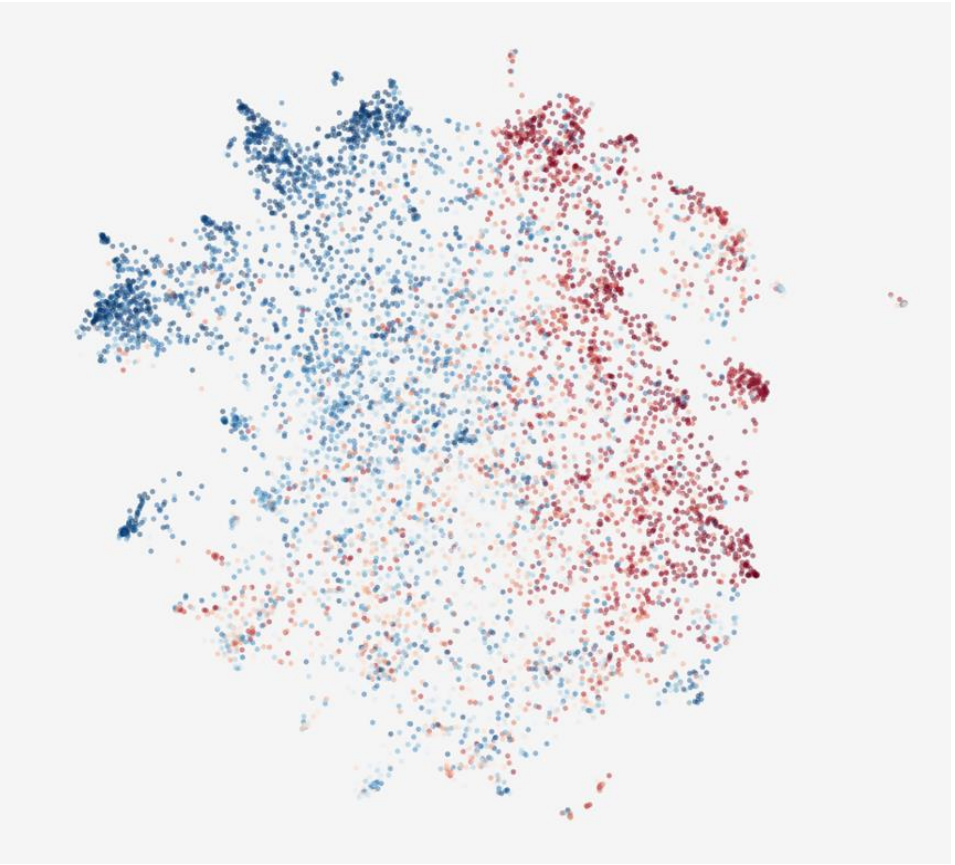
channel_label	channel_name	감정표현	친밀감	공감성	정보성	공격성
female	예랑가랑	26.5	13.9	25.8	62.0	0.5
	해쭈 [HAEJOO]	50.1	22.4	29.5	62.7	0.2
	회사원A	46.1	34.0	30.4	67.3	0.2
	힙으뜸	42.6	32.0	20.6	67.8	0.0
male	[꿀템TV] 클템 유튜브	11.8	16.6	13.8	53.6	0.7
	랄로	6.5	7.0	8.2	35.8	3.3
	모트라인 MOTline	11.2	16.2	10.6	59.1	1.0
	차도르	9.8	11.9	9.2	43.2	1.5
	피지컬갤러리	12.1	15.2	10.4	45.6	1.1
neutral	애타【ETA】	23.1	16.9	24.4	56.3	0.2

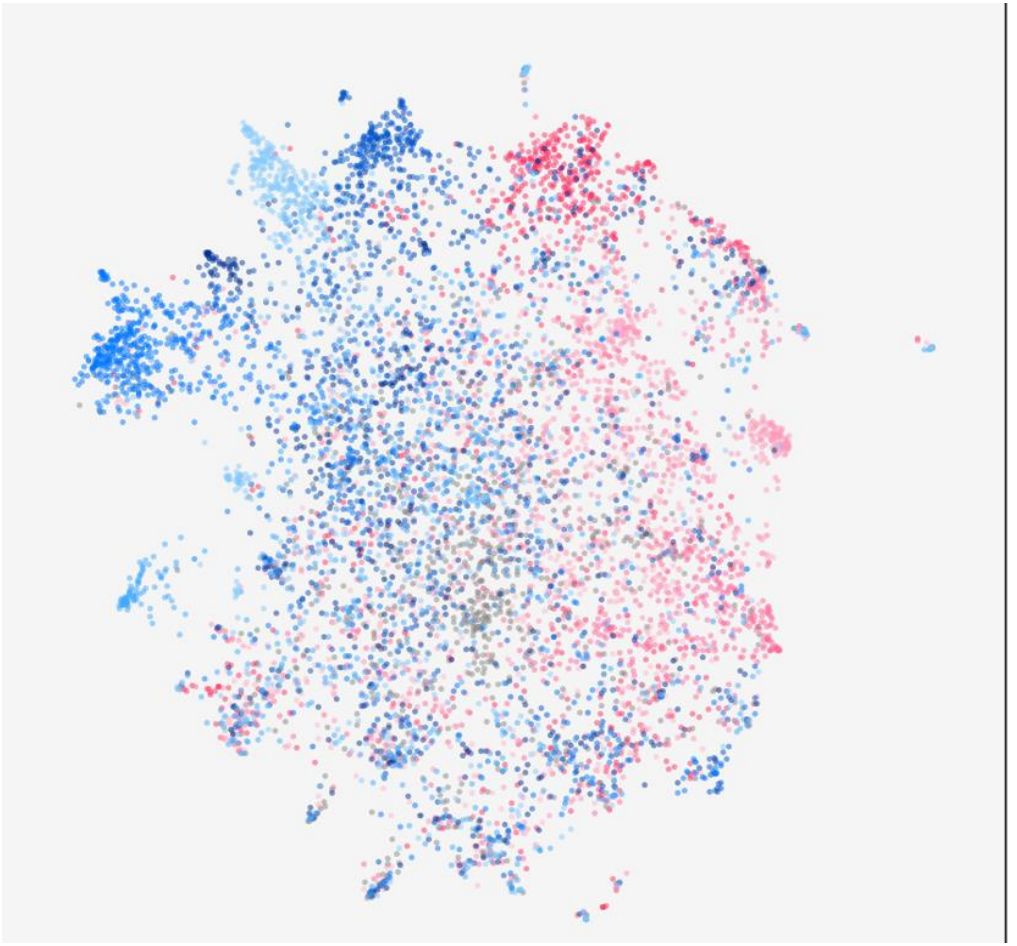
시각화 - UMAP


댓글 성향 분포 (UMAP)
빨강 = 여성향 (gender_score 양수) / 파랑 = 남성향 (gender_score 음수)
gender_score = prob_female - prob_male (분류기 학습 결과)
UMAP 2
UMAP 1 (축 자체엔 의미 없음 — 점들의 상대적 위치가 중요)
gender_score = prob_female - prob_male

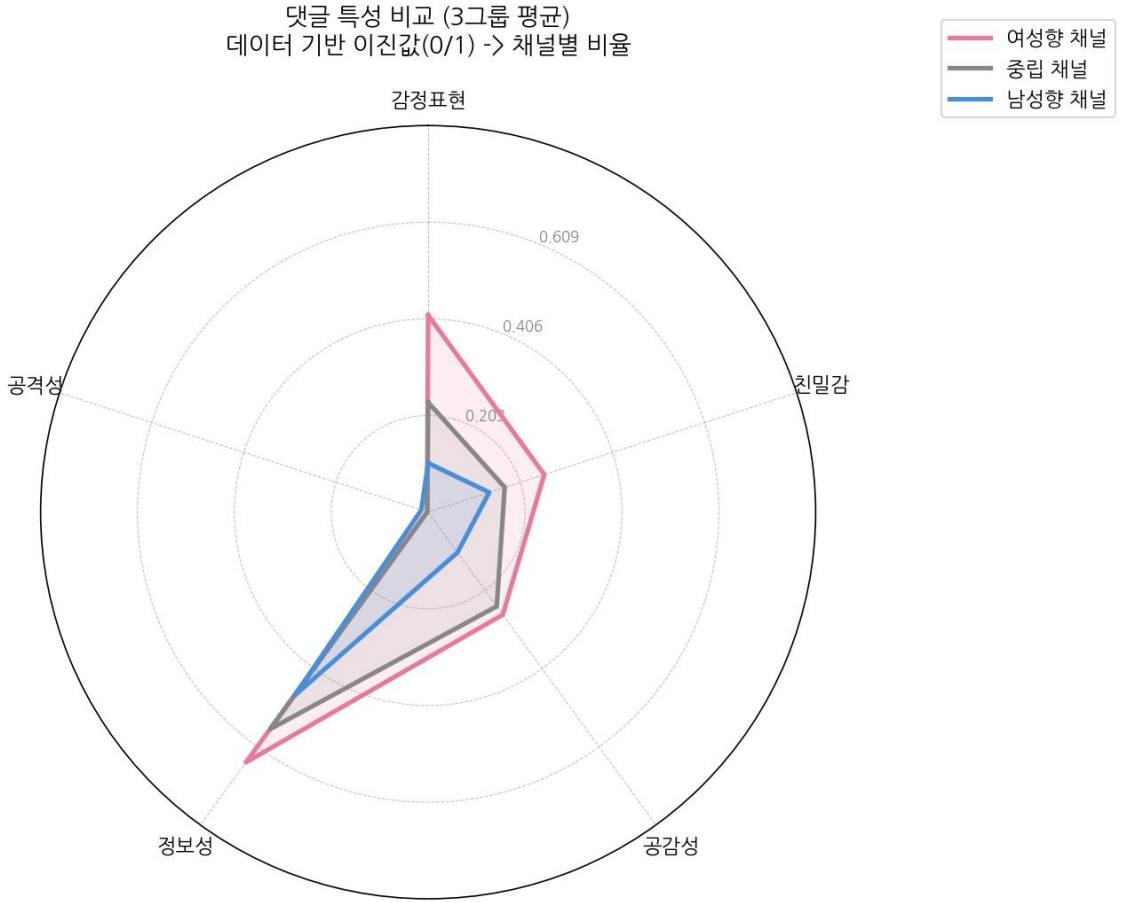

채널별 댓글 분포 (UMAP)
같은 채널 댓글끼리 뭉쳐있을수록 그 채널만의 독특한 말투가 있음
UMAP 2
UMAP 1
힙으뜸 (여성향)
해주 [HAEJOO] (여성향)
회사원A (여성향)
예랑가랑 (여성향)
애타【ETA】(중립)
피지컬갤러리 (남성향)
차도르 (남성향)
모트라인 MOTline (남성향)
랄로 (남성향)
[꿀템TV] 클템 유튜브 (남성향)

시각화 - UMAP

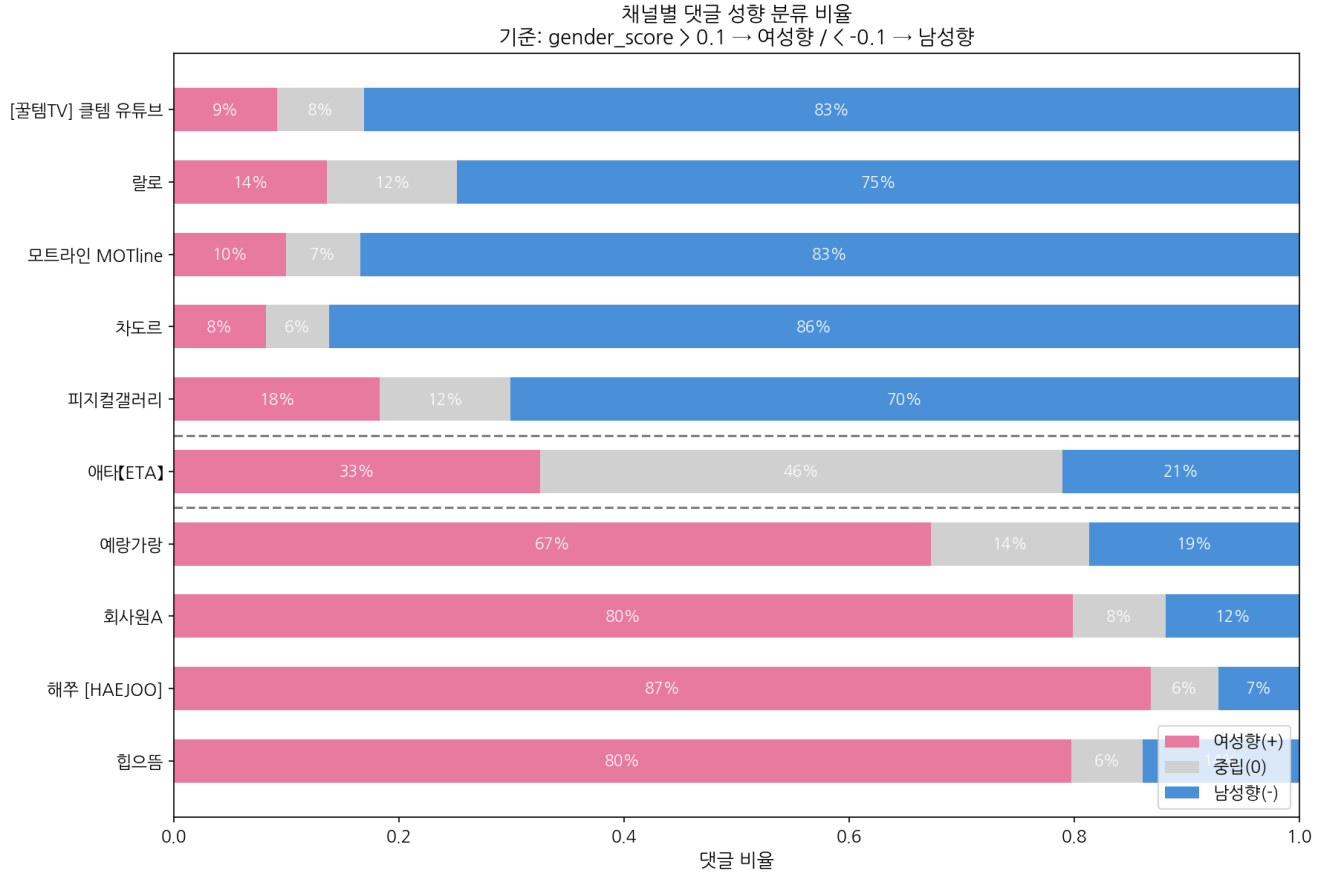



힙으뜸 (여성향)
해쭈 [HAEJOO] (여성향)
회사원A (여성향)
예랑가랑 (여성향)
애타【ETA】(중립)
피지컬갤러리 (남성향)
차도르 (남성향)
모트라인 MOTline (남성향)
랄로 (남성향)
[꿀템TV] 클템 유튜브 (남성향)

시각화 - 레이더 차트


댓글 특성 비교 (3그룹 평균)
데이터 기반 이진값(0/1) -> 채널별 비율
여성향 채널
중립 채널
남성향 채널
감정표현
친밀감
공감성
정보성
공격성
0.609
0.406
0.203

시각화 - 성향 분류 막대 그래프



한계점

1. Weak Label의 본질적 한계

채널 단위 정답을 쓰기 때문에, 같은 채널에도 반대 성별 시청자가 달린 댓글이 포함됨

-> 노이즈가 학습 데이터에 그대로 들어가는 문제

2. 애타[ETA]의 오분류

"대빵이", "오타쿠" 등 채널 고유 팬덤 호칭이 여성향 4개 채널에는 거의 등장하지 않아서 분류기가 이 표현을 "여성향이 아닌 신호"로 학습

-> 실제로는 여성 시청자가 많은 채널인데 남성향으로 분류됨

3. 공격성 지표의 한계

유튜브 댓글 정책으로 직접적 욕설이 드물어 남성향 신호로서 변별력이 약함 -> 실제 성별 차이를 반영하지 못하는 지표일 수 있음

향후 연구 방향

1. 팬덤 어휘 전처리 후 재학습

채널 고유 팬덤 호칭을 전처리 단계에서 제거한 뒤 분류기 재학습

-> 애타 같은 케이스의 오분류 감소 기대

2. LLM 기반 댓글 직접 분류

LLM에게 댓글 하나씩 어떤 말투인가를 판단하게 함

-> Weak Label 없이도 정밀한 분류 가능 / 단, 비용과 시간이 큰 단점

Q & A
감사합니다