

개요: 동영상 콘텐츠 하이라이트 편집 및 설명(요약) 데이터셋이란

한국인을 대상으로 하는 인공감성지능(Artificial Emotional Intelligence) 모델 개발을 위한 학습 데이터셋으로 KAIST에서 구축했으며, 총 6,000시간 이상의 동영상 데이터와 이에 대응하는 레이블링 데이터로 구성되어 있다.

미래 사회에서 개인과 사회 모두 영상 자료를 효과적으로 활용 및 탐색하기 위해서 동영상 하이라이트 편집 및 설명(요약) 기술에 대한 필요가 대두되었다. 이를 위해서는 인공지능 기술 개발에 필요한 양질의 학습용 영상 데이터를 구축할 필요가 있다. 학습시키는 input 데이터 세트의 품질은 결국 output의 품질로 이어지며, 양질의 데이터 세트를 구축한다면 이를 활용한 유용한 기술 개발이 활발할 것으로 예상되므로 이러한 근거를 바탕으로 본 데이터를 구축했다.

데이터셋의 구성

- KBS 뉴스 방송본과 취재원본을 더해 총 9,528개, 약 1,000시간 분량의 영상 수집
- 유튜브 영상 총 23,909개, 약 5,000시간 분량 영상 수집, KBS 뉴스와 유튜브를 합쳐 33,437건, 약 6,000시간
- 영상당 3명의 어노테이터가 작업하여 100,000장 이상의 레이블링 데이터 확보

원천데이터 기준				
출처	개수	합계 : 원천 영상 길이(초)	합계 : 원천 영상 길이(분)	합계 : 원천 영상 길이(시)
KBS 뉴스	9,528	3,603,302	60,055	1,001
방송본	6,490	1,314,033	21,901	365
취재원본	3,038	2,289,269	38,154	636
유튜브	23,909	17,999,223	299,987	5,000
총합계	33,437	21,602,525	360,042	6,001
가공데이터 기준				
출처	개수	합계 : 작업 영상 길이(초)	합계 : 작업 영상 길이(분)	합계 : 작업 영상 길이(시)
KBS 뉴스	30,266	11,236,981	187,283	3,121
방송본	21,149	4,368,656	72,811	1,214
취재원본	9,117	6,868,325	114,472	1,908
유튜브	71,735	54,004,120	900,069	15,001
총합계	102,001	65,241,102	1,087,352	18,123
카테고리 비율				
출처/카테고리	개수			
KBS 뉴스	9,528			
기타	9,528			
유튜브	23,909			
기타	4,200			
반려동물/동물	2,240			
스타일링/뷰티	2,509			
스포츠	3,658			
여행	2,720			
음식	2,506			
일상	3,976			
자동차	2,100			
총합계	33,437			

데이터셋의 설계 기준과 분포

본 데이터셋은 어노테이터가 영상의 하이라이트 구간을 태깅한 뒤, 해당 구간에 대해 장소 21종, 인물관계 12종, 행동 65종, 감정 7종을 라벨링했다. 장소 카테고리의 항목들은 관련 연구인 Places 365 데이터셋을 바탕으로 선정되었다. 인물관계 또한 관련 연구를 참조하여 중복되거나 주관적 판단이 개입되는 항목을 추려 선정되었다. 행동 카테고리는 해외 관련 데이터

셋(ActiveNet, HMDB, UCF101)과 국립국어원 언어정보나눔터를 참고했다. 감정 카테고리는 Susan David의 감정 분류를 준수하였다.

*장소 카테고리 21종

번호	대분류	소분류
1	교육시설	학교, 실험실, 강의실, 도서관 등
2	음식점	식당, 바, 푸드코트, 레스토랑 등
3	자연 경관	바닷가, 협곡, 강, 숲 등
4	인공조성환경	외부 발코니, 분수, 정자, 과수원, 마당 등
5	거주시설	부엌, 거실, 서재, 발코니, 침실 등
6	숙박시설	호텔, 캠핑장 등
7	종교/의료시설	병원, 보건소, 교회, 사찰 등
8	공공시설	관공서, 은행, 공원 등
9	교통/이동수단	택시, 기차, 자가용, 비행기 등
10	상업시설	재래시장, 대형마트, 쇼핑몰 등
11	문화재 및 유적지	문화재, 유적지
12	스포츠 관람 및 레저시설	야구장, 아이스링크, 볼링장, 체육관 등
13	오락시설	오락실, PC방, 놀이동산 등
14	공연시설	콘서트장, 무대, 야외극장 등
15	행사/사무공간	사무실, 컨퍼런스 룸, 결혼식장, 구내식당 등
16	건물 내부시설	복도, 계단, 엘리베이터 등
17	도심 환경	도심지 대로변, 빌딩 숲 등
18	도로 및 교통시설	육교, 다리, 횡단보도 등
19	예술/전시공간	미술관, 전시회 등
20	산업시설	공장, 댐, 급수시설 등
21	알 수 없음	위의 20가지에 해당하지 않는 장소

*인물관계 카테고리 12종

번호	대분류	소분류
1	사회 속 관계	서로 아는 사이
2	사회 속 관계	유명인과 일반인
3	사회 속 관계	모르는 사이
4	혈연관계	부모 자식 관계
5	혈연관계	형제 자매
6	혈연관계	부모 자식/형제 자매 외 혈연관계
7	친구 - 적 관계	친구 관계
8	친구 - 적 관계	적대 관계
9	친구 - 적 관계	연인 및 배우자
10	직장 내 관계	직장 내 상하관계
11	직장 내 관계	직장 동료
12	기타	알 수 없음 (1인칭 포함)

*행동 카테고리 65종

행동					
번호	분류	번호	분류	번호	분류
1	먹다	23	보다	45	붙이다
2	마시다	24	듣다	46	모으다
3	놀다	25	(냄새를) 맡다	47	끝다
4	쉬다	26	만지다	48	씻다
5	타다	27	만들다	49	옷을 입다
6	요리하다	28	옮기다	50	꾸미다
7	산책시키다	29	찾다	51	제모하다
8	(공을) 차다	30	갖다	52	만나다
9	그리다	31	주다	53	헤어지다
10	연주하다	32	사다(구매하다)	54	말하다
11	(사진을) 찍다	33	팔다	55	일하다
12	오다	34	잡다	56	회의하다
13	가다	35	놓다	57	협업하다
14	서다	36	흔들다	58	경쟁하다
15	앉다	37	돌리다	59	싸우다
16	눕다	38	고치다	60	배우다
17	오르다	39	던지다	61	읽다
18	내려오다	40	날다	62	쓰다
19	자다	41	죽다	63	악수를 하다
20	걷다	42	웃다	64	(의견을)나누다
21	뛰다	43	울다	65	알 수 없음(기타)
22	다치다	44	자르다		

*감정 카테고리 7종

번호	대분류	소분류
1	분노	툼툼대는, 좌절한, 짜증내는,

		방어적인, 악의적인, 안달하는, 구역질나는, 노여워하는, 성가신
2	슬픔	실망한, 비통한, 후회되는, 우울한, 마비된, 염세적인, 눈물나는, 낭패한, 환멸을 느끼는
3	불안	두려운, 스트레스 받는, 취약한, 헛갈리는, 당혹스러운, 회의적인, 걱정스러운, 조심스러운, 신경쓰이는
4	상처	질투하는, 배신당한, 격리된, 충격 받은, 궁핍한, 희생된, 억울한, 괴로워하는, 버려진
5	당황	격리된, 시선 의식하는, 외로운, 열등한, 죄책감의, 부끄러운, 혐오스러운, 한심한, 헛갈리는
6	기쁨	감사하는, 믿는, 편안한, 만족한, 흥분한, 느긋한, 안도하는, 신이 난, 자신하는
7	중립	무표정, 감정이 0인 상태

데이터 구조

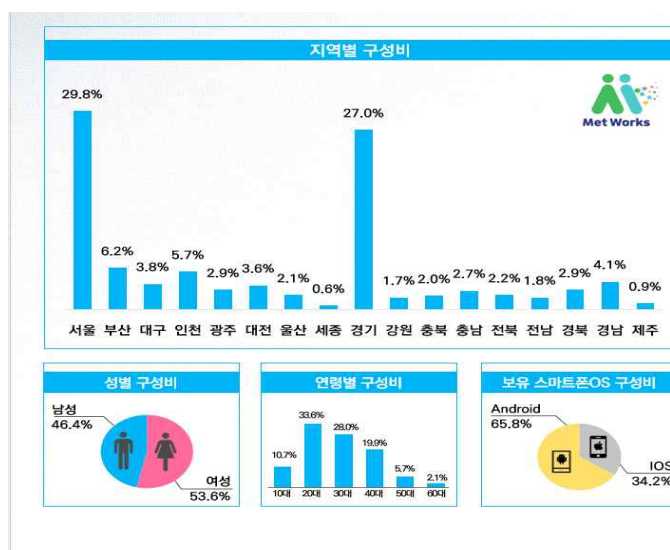
본 데이터셋은 KBS와 유튜브에서 수집한 동영상 데이터(MP4, AVI)와, 해당 동영상에 대한 메타데이터 및 레이블링 데이터(JSON) 파일을 제공한다.

데이터셋 예시

영상 구간 편집 및 라벨링 처리 작업물 예시	json 파일 예시
	

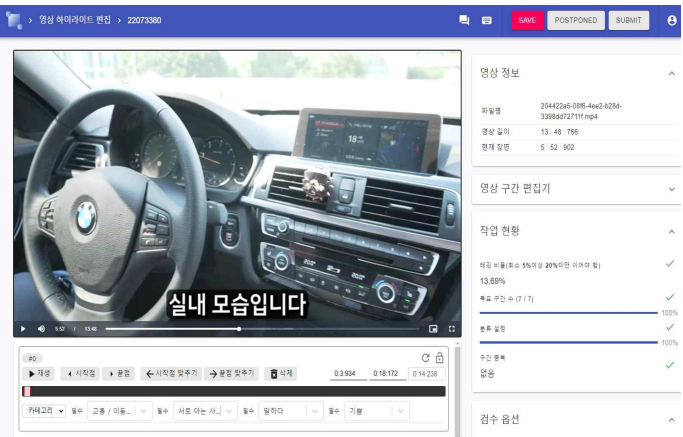
	<pre> { "metadata": { "file_name": "유튜브_자동차_0331.mp4", "type": "mp4", "length": 828.7666, "quality": "1280x720", "date": "2019-09-12", "license": "youtube", "class_code": 1, "annotator_id": 4988 }, "timelines": [{ "id": 0, "start": 3, "end": 18, "variation": "category", "attributes": { "place": 9, "action": 54, "emotion": 6, "relationship": 1 } }, { "id": 1, "start": 19, "end": 34, "variation": "category", "attributes": { "place": 9, "action": 54, "emotion": 6, "relationship": 1 } }] } </pre>
--	---

데이터 구축 과정



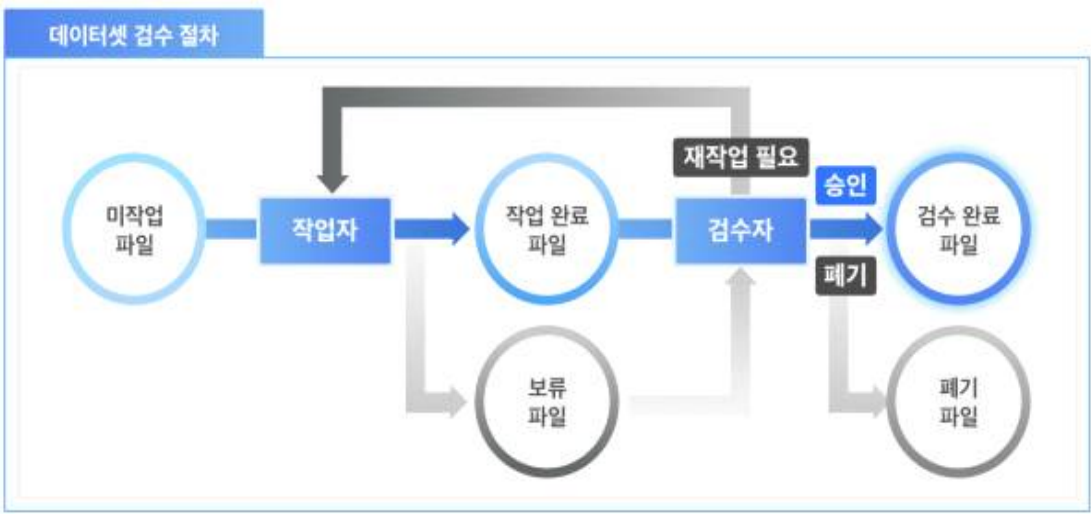
(클라우드 소셜 플랫폼 설명)

동영상 데이터는 클라우드 소싱 플랫폼을 활용하여 획득하였다. 영상길이 총 6,000시간 이상 분량의 동영상 데이터에 대해 역시 클라우드 소싱을 통해 3명의 어노테이터가 하이라이트 구간을 태깅하여 총 3배 분량의 레이블링 데이터를 획득하였다.



(하이라이트 구간 지정 도구)

검수와 품질 확보



(데이터 관점의 검수절차)

검수는 다음과 같은 절차를 통해 수행되었다.

- 1. 검수 가이드 작성
- 2. 검수 인력 배정
- 3. 검수 작업물 지정
- 4. 검수 항목 확인
- 5. 1차 검수
- 6. 2차 검수

검수의 기준은 다음과 같다.

구분		측정지표	검사기준	자동화	검사율
구문정확성	데이터 구조	어노테이션 포맷 정확도	99.9%	자동	전수검사
	데이터 형식	속성 값의 유효 범위 내 존재여부, 구성 데이터의 형식 준수 여부	99.9%	자동	
의미정확성	정확도	구간 하이라이트 지정의 적절성	90% 이상	수동 (육안)	샘플링 검사
	정밀도	생성된 어노테이션 중 정답 어노테이션 비율	95% 이상	수동 (육안)	
	재현율	모든 가능한 어노테이션 중 정답 어노테이션 비율	95% 이상	수동 (육안)	

데이터 구축 담당자

수행기관(주관) : 카이스트

(전화: 010-4028-0212, 이메일: nerdinary@kaist.ac.kr)