테크니컬 리포트

2020년 1차
인공지능
학습용
데이터 구축

미디어 영역

대용량 동영상 콘텐츠

개요: 대용량 동영상 콘텐츠 AI 학습데이터셋이란?

미디어 분야는 AI 기술이 가장 활발하게 적용되는 분야로 FANG(Faceboo, Amazon, Netflix, Google)과 네이버, 카카오로 대변되는 거대 미디어 플랫폼이 AI 기술을 주도하고 있다. 특히 AI 르네상스 시대를 연 딥러닝(deep learning) 기반 AI 기술은 행동데이터를 넘어서 자연어처리, 음성인식, 영상인식 등 의미데이터 처리에 널리 활용 중이다.

동영상 분야의 대규모 학습데이터셋으로는 YouTube 8M 이 610 만개 클립, 35 만시간 분량의 동영상을 원천데이터(raw data)로 활용해 3862 개 클래스에 대해 26 억 개의 오디오/비디오 피쳐를 제공하고 있다. 그럼에도 불구하고 동영상 분야는 사진을 학습시키는 경우보다 훨씬 더 많은 양의 학습데이터가 필요하다.

언론사 등 미디어 기업은 저작권(copyright)을 확보한 의미 데이터(semantic data)로서 콘텐츠를 대거 보유하고 있다. 뿐만 아니라 보도 등 미디어 동영상은 인위적으로 촬영된 재연 동영상이 아닌, 상당수가 실제 동영상이라는 강점을 갖고 있으며, 수십년 간 촬영된 다양한 객체와 상황을 담고 있어 과적합 이슈를 극복할 수 있다는 장점을 갖고 있다.

이번에 소개할 데이터는 특히 방송사가 보유한 미디어 동영상을 원천 데이터로 대거 활용해 동영상 내 다양한 객체, 행동, 상황의 인식을 위한 AI 학습데이터를 본격적으로 구축했다는 점에서 의의가 있다.

데이터셋의 구성

데이터셋은 크게 원천데이터와 학습데이터로 나눠진다. 학습데이터는 학습동영상 클립(mp4), 어노테이션 데이터셋(json), 카테고리 정의서(csv), 클립리스트(csv)으로 구성된다.

원천동영상(mp4)은 학습데이터와 함께 공개된다. 매경미디어그룹에 속한 MBN(매일방송), 매경 TV, 매일경제신문이 제공하는 1630 시간 분량의 실제 방송 동영상이다. 크게 보도, 예능, 교양, 유튜브 등 4 개 장르에서 객체 및 상황별로 세분화된 30 종 42,474 개 클립으로 구성되어 있다. 보도 동영상은 메인 뉴스인 MBN 종합뉴스를 활용했으며 스크립트가 포함된 자연어처리 데이터를 함께 제공한다. 원천동영상은 모든 인물의 안면을 자동 블러 처리했다.

표 1 원천동영상의 구성과 시간

번호	동영상 종류	프로그램명	수량(개)	분량(시간)	평균시간(초)	매체
1	보도_경제	MBN종합뉴스_경제	41,285	1,145.94	100	MBN
2	보도_국제	MBN종합뉴스_국제				MBN
3	보도_문화	MBN종합뉴스_문화				MBN
4	보도_부동산	MBN종합뉴스_부동산				MBN
5	보도_사회	MBN종합뉴스_사회				MBN
6	보도_생활건강	MBN종합뉴스_생활건강				MBN
7	보도_스포츠	MBN종합뉴스_스포츠				MBN
8	보도_연예	MBN종합뉴스_연예				MBN
9	보도_정치	MBN종합뉴스_정치				MBN
10	보도_증권	MBN종합뉴스_증권				MBN
11	보도_대담	경제노트	115	18.01	564	MBN
12	교양_건강	천기누설	53	58.41	3,967	MBN
13	교양_다큐	나는자연인이다	52	50.24	3,478	MBN
14	교양_주거	행복한家	10	8.13	2,926	MBN
15	교양_안전	안전대한민국	88	2.69	110	MBN
16	예능_강아지	우리집에해피가왔다	14	16.23	4,174	MBN
17	예능_건강	엄지의제왕	52	58.09	4,022	MBN
18	예능_동물	기막힌동물원	13	11.40	3,156	MBN
19	예능_뷰티	111뷰티	164	4.10	90	MBN
20	예능_상품	카트쇼	25	26.94	3,880	MBN
21	예능_요리	123요리	59	2.81	171	MBN
22	예능_음식	알토란	53	70.57	4,784	MBN
23	예능_음악	보이스퀸	13	33.74	9,345	MBN
24	예능_한복	역사드라마쇼왕과여자	10	11.62	4,184	MBN
25	유튜브_경제	와이드경제	233	43.67	675	매일경제TV
26	유튜브_DIY	이렇게만들죠	59	33.32	2,033	매일경제
27	유튜브_정치	정치라면	49	12.05	886	매일경제
28	유튜브_여행	꿀잼트래블	103	10.24	358	매일경제
29	유튜브_질병	코로나19	16	9.62	2,166	매일경제
30	유튜브_IT	실리콘밸리 리포트	8	2.22	998	매일경제

학습동영상은 559 시간 12,1536 개 클립, 283,751,365 개(프레임 기준)의 바운딩 박스(bounding box)에 객체, 행동, 상황 정보가 레이블링되어 제공된다. 학습동영상의 길이는 평균 16.6 초다. 행동, 상황,

보도 동영상의 기자 135 명에 대해서는 초상권 동의를 받아 블러 처리 없이 레이블링되어 있다. 나머지 동영상에서는 인물이 모두 수작업으로 블러 처리되어 있다. 어노테이션

데이터셋은 바운딩박스의 위치 정보 및 레이블링 정보 등을 담고 있다. 바운딩박스의 최소 크기는 40*40pix 이다.

표 2 학습데이터셋 구성

설계분류	설계 및 규칙
Format&codec	- H.264 - MPEG-4
Resolution	- SD : 640 X 480(360) - HD : 1280 X 720
FPS&Length	- 30fps - 10Sec
Labeling	- Bounding box Spec. min. 40pix - .json

데이터셋의 설계 기준과 분포

데이터셋 설계는 원천동영상 수집 단계에서부터 다양성과 균형을 동시에 고려할 수 있도록 데이터셋을 설계했다. 객체, 행동, 상황의 카테고리 정의는 널리 알려진 카테고리 체계를 최대한 참조하되, 미디어에 최적화된 형태로 수정했다. 우선 원천동영상 종수는 보도의 경우 정치, 경제, 사회, 문화, 국제 등 뉴스 카테고리 10종, 보도 이외는 주요한 객체나 상황 등과 매핑되는 프로그램별로 구성했다.

객체 카테고리는 ImageNet의 카테고리를 기반으로 19개 대분류, 184개 중분류, 1763개 소분류로 나누었다. 행동 카테고리는 DeepMind의 Kinetic400를 참조하여 18개 대분류, 58개 중분류, 358개 소분류로 정의했다. 상황 정보는 스크립트가 있는 보도 동영상 대상으로 내용을 중심으로 레이블링했다. 분류 체계가 매우 다양하기 때문에 소분류 수준에서는 어느 정도 쏠림 현상이 있지만, 원천동영상 수집 단계에서부터 객체와 상황이 구별되는 30종으로 구성했다.

동영상 종별 분포는 <표 3>과 같다.

표 3 동영상 종별 분포

분류 (총 4분류)	소항목 (총 30종)	클립 수	시간 (초)
보도 (총 11개 소항목)	경제	8,322	157,752
	국제	4,112	71,530
	문화	913	17,151
	부동산	8	215
	사회	31,368	684,498
	스포츠	3,626	53,168
	정치	26,014	535,461
	증권	2	59
	기타 (생활건강, 연예, 대담)	2,320	57,438
교양 (총 4개 소항목)	건강	1,292	6,881
	다큐	2,497	13,985
	주거	895	6,805
	안전	303	4,210
예능 (총 9개 소항목)	강아지	2,241	39,640
	건강	12,979	118,957
	동물	1,161	7,016
	뷰티	423	2,849
	상품	4,662	57,224
	요리	269	7,140
	음식	6,717	101,097
	음악	7,808	40,537
	한복	3,193	19,882
유튜브 (총 6개 소항목)	경제	102	4,001
	DIY	17	146
	정치	196	3,610
	여행	30	177
	질병	48	620
	IT	18	277

객체와 행동의 카테고리별 분포는 <표 4>와 같다.

표 4 객체별 행동별 AI 학습데이터 분포

	클립 수		클립 수		클립 수
객체 합계	**116,794,966**	**행동 합계**	**23,058,893**	**상황 합계**	**143,897,506**
건축	1,420,927	개인용모 관리	20,613	내용_경제	12,857
공간	585,587	개인운동	47,607	내용_국제	9,054
군사	160,461	꾸미기	6,878	내용_기타	142,618,170
문화	252,443	농경, 조경	3,518	내용_노동 및 산업 진흥 행정	829
바이러스	222	동물과 상호작용	79,395	내용_문화	336,719
사람	39,497,700	레포츠	5,198	내용_범죄	18,563
사물	25,382,777	만들기	11,132	내용_사고	7,024
식품	1,204,805	무용	32,833	내용_사법 및 공공 질서 행정	1,542
운동선수	165,951	스포츠	109,790	내용_사회	191,181
육상생물	1,348,536	식생활	347,634	내용_사회 및 산업정책 행정	793
의류/악세서리	23,905,612	신체 기본동작	6,544,195	내용_사회이슈	37,943
이동수단	3,319,054	연주하기	342,628	내용_스포츠	5,575
인물	3,840,157	오락	3,304	내용_외무 및 국방행정	3,530
장소	763,017	육아, 보육	8,706	내용_입법 및 일반 정부 행정	457
주류	23,524	의사소통	15,260,136	내용_재해	3,429
지형	441,647	이동하기 (운전)	47,318	내용_정보통신	10,459
직업	14,419,048	정리하기	8,426	내용_정치	149,919
해양생물	36,834	콘텐츠	179,582	내용_지역	635
화장품	25,870			시간	74,779
	794			장소	313,824
				콘텐츠 소비	100,224

데이터 구조

어노테이션 데이터셋의 항목과 설명은 아래 표와 같다.

표 5 어노테이션 데이터셋의 구조

NO	항목		길이	타입	필수여부
	한글명	영문명			
1	비디오ID	video_id		String	Y
2	메타정보	metas		Object	Y
2-1	유형코드	category_0	45	String	Y
2-2	대분류코드	category_1	45	String	Y
2-3	중분류코드	category_2	45	String	Y
2-4	소분류코드	category_3	45	String	Y
2-5	비디오시작프레임	start_frame		Number	Y
2-6	비디오종료프레임	end_frame		Number	Y
2-7	박스정보	bbox_list		Array	Y
2-7-1	박스 x 좌표값	x		Number	Y
2-7-2	박스 y좌표값	y		Number	Y
2-7-3	박스 너비	width		Number	Y
2-7-4	박스 높이	height		Number	Y

데이터 예시

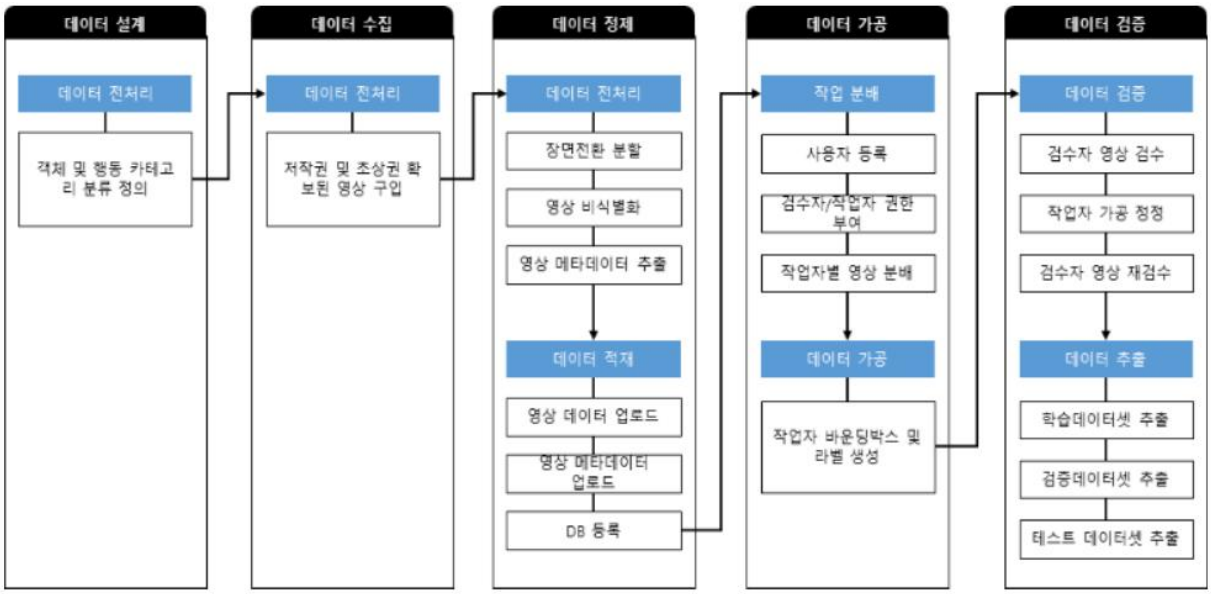
레이블링된 학습동영상의 데이터의 예시는 아래와 같다.

어노테이션 데이터 예시는 아래와 같다.

```
[{
  "video_id":"1591556",
  "metas":[{
    "category_0":"1",
    "category_1":"011",
    "category_2":"002",
    "category_3":"002",
    "start_frame":581,
    "end_frame":582,
    "bbox_list":[
      [17.5,116.71875,158.75,163.12499999999994],
      [17.515060240963855,116.73273450946643,158.78334767641996,163.14759036144574]]
    },
    {
      "category_0":"1",
      "category_1":"006",
      "category_2":"001",
      "category_3":"001",
      "start_frame":587,
      "end_frame":588,
      "bbox_list":[
        [54.375000000000014,135.46875,185.63139931740616,202.03125],
        [54.362113402061865,135.4236469072165,185.6248900460927,202.0763530927835]],
      }
  ]
}]
```


데이터 구축 과정

데이터 구축 과정은 '데이터 설계 → 데이터 수집 → 데이터 정제 → 데이터 가공 → 데이터 검증' 등 총 5단계로 구성된다. 설계 단계에서는 객체/행동/상황 카테고리를 정의한다. 수집 단계에서는 저작권과 초상권 이슈를 해결한 원천 동영상을 구매한다. 정제 단계에서는 영상을 분할하고 불용 동영상을 제거하며 안면 블러 처리 등 비식별화 작업을 수행한다. 가공 단계에서는 작업자에게 정제된 동영상을 제공하고 바운딩 박스 생성 및 레이블링을 수행한다. 빠른 레이블링을 위해 오토 레이블링 및 전이학습을 통한 모델 개선 과정을 병행했다. 검증 단계에서는 전수 수작업으로 타 작업자의 작업량을 교차로 검수한다.



검수와 품질 확보

검수 절차는 아래와 같다. 이 사업에서는 5 개의 사회적 기업이 참여했다. 최근 인종 차별 등 AI 편향성 이슈가 제기되고 있다. 노인, 발달장애인, 경력 단절 여성, 청년 등 취약 계층 참여는 이러한 문제를 개선하는데 기여했다. 이밖에 품질 제고를 위해 다양한 시도가 있었다. 우선 레이블러의 역량에 따라 작업 물량을 차등 배분했다. 각 사의 레이블러는 저작도구 개발사와 끊임없이 소통하면서 저작도구의 사용성을 개선했다. AI 전문 기관이 레이블링 과정에서 모호한 부분에 명확한 가이드라인을 제공하고 해당 내용을 반영해 매뉴얼을 수시로 업데이트했다. 각 사의 관리자가 레이블러들의 작업 결과를 검수했으며, 그 결과는 저작도구에서 실시간 모니터링이 됐다. 1 차 가공 완료 후에는 다른 팀이 작업한 결과물을 전수 수작업 검수하고 필요시 추가 레이블링 또는 수정 등을 진행했다.

데이터 구축 담당자

수행기관(주관): ㈜케이디엑스한국데이터거래소 (전화: 02-2000-5936), 이메일: webmaster@kdx.kr

수행기관(개발): ㈜씨이랩 (전화: 02-2039-3145), 이메일: sales@xiilab.com