

유튜브 데이터 마이닝을 통한 영상 소비자들의 관심주제 분석: 유아동 콘텐츠의 사례

박영준, 송용백, 이승윤, 이준송, 윤장혁 교수*

건국대학교 산업공학과

1. 서론

Motivation

- 온라인 오픈 플랫폼이자 구글에 이은 세계 제 2위의 정보 검색 서비스 유튜브
- 2018년 기준 유아, 아동 관련 유튜브 콘텐츠 시장은 4조에 달하는 규모로 성장
- 영상 소비자의 실질적 반응을 고려한 소비자 니즈 분석연구 미비
- 프로슈머의 성격을 가지는 유튜브 제작자, 소비자들의 관심과 니즈의 파악은 관심 트렌드를 대변

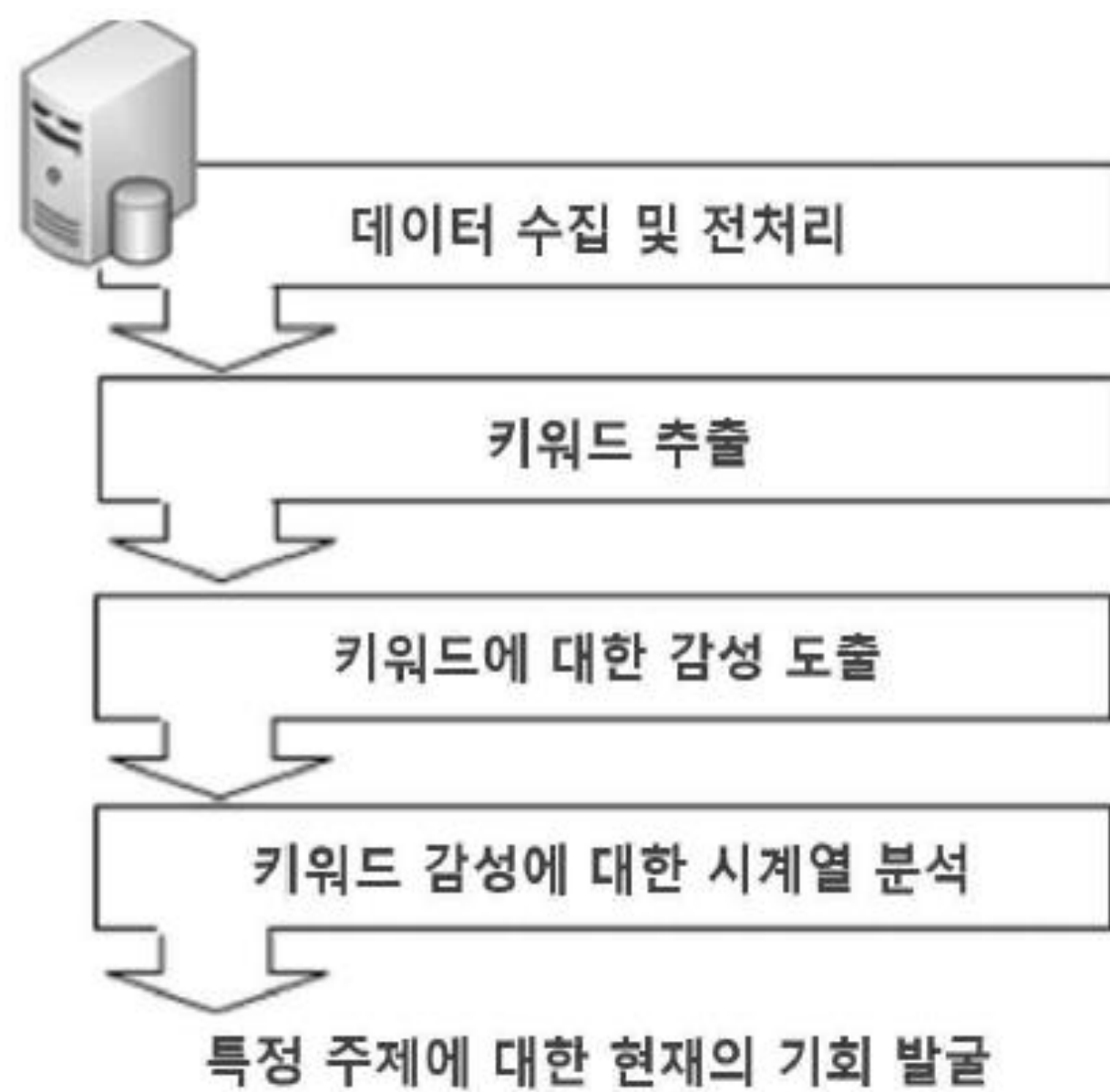
Goal

- 영상의 제목과 Description에서 추출한 여러 키워드 중 소비자의 관심을 나타내는 핵심단어를 “관심 주제 ” 로 정의
- 영상의 주제와 소비자 반응을 분석하여 유튜브 소비자들의 관심주제 파악
- 관심주제의 파악을 통해 미래에 영향력 있는 콘텐츠를 제작하고자 하는 유튜버 및 다중 채널 네트워크(MCN) 비즈니스에 인텔리전스 제공

2. 선행연구

- 대표적인 소셜 미디어로써 유튜브에 대한 많은 연구가 진행되었으나 조회수를 종속 변수로 상정하고 독립변수를 분석하는 연구가 대다수를 이룬다.
- 영상의 조회 수에 영향을 끼치는 요인들 중 댓글 수, 즐겨찾기 등록 수, Like 수가 가장 높은 상관관계를 가진다는 분석이 있었다.Chatzopoulou et al.(2010)
- Susarla(2012)는 댓글 수, 평점, 좋아요 수와 같은 일반적인 요인 뿐만 아니라 친구 네트워크, 구독자 네트워크 등 네트워크 요인을 분석하여 영상 게시 초기에는 구독자 네트워크가 관심의 주도적 영향력을 가지고 후반에는 친구 네트워크 방식이 영향을 끼침을 확인하였다.
- 소비자의 반응을 통해 인기 콘텐츠가 가진 주제를 파악하는 연구는 미비하다.

3. 연구방법



데이터 수집 및 전처리

- 유튜브 아동관련 top 100 채널 속 약 8만 여개 영상과 830만개의 댓글 추출 키워드 추출

키워드 추출

- 그 중 3 글자 이상의 유의미한 단어로 필터링 후, 핵심성을 파악하기 위해 Applied TF 계산

키워드 감성 도출

- 댓글에 대한 감성분석 (Sentiment Analysis)
- 본 연구에서는 감성분석을 위해 딥러닝 모델을 활용
- ‘네이버 영화평 리뷰 데이터’ 를 Dataset으로 선정(학습 15만 개, 테스트 5만 개)
- 리뷰를 감정적 단어와 문맥을 고려하여 특성 벡터(feature vector)로 변환 후 학습 [정확도(accuracy) 83.86%, 정밀도(precision) 84.03%, 재현율(recall) 83.88%]

키워드 감성에 대한 시계열 분석

- 관심주제 별로 총 댓글 수, 긍정 댓글 수, 부정 댓글 수에 대한 시계열 그래프를 작성

4. 연구결과

Weighted Term Frequency Matrix

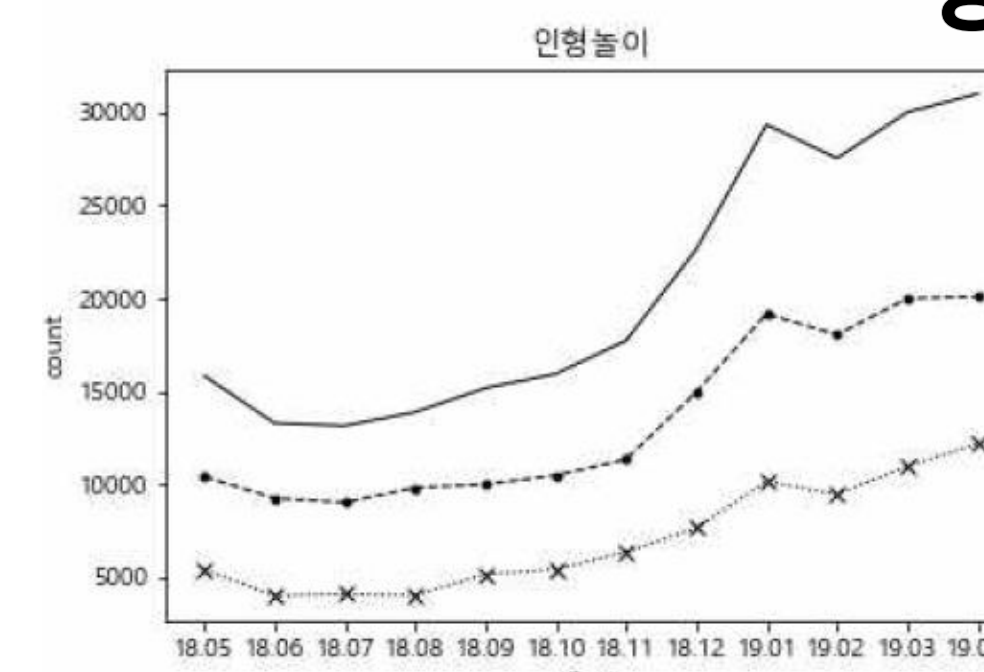
- 3글자 이상의 단어 36,059개에 대하여 영상의 조회수를 가중치로 설정하여 그 가중치를 곱해준 키워드들의 Term Frequency를 계산하였고, 372개의 관심주제 도출

관심주제별 댓글의 감성분석

순위	키워드	값	순위	키워드	값
1	장난감	134148.4	13	미니특공대	72654
2	뽀로로	520435	14	포코스토이	67879
3	타닝메카드	268452	15	코코몽	61098
4	애니메이션	209425	16	인기동요	59428
5	헬로카봇	198868	17	슈퍼링스	58520
6	장난감놀이	189577	18	다이노코어	58382
7	포타버스 타로	132119	19	호이 스미스	57338
8	로보카폴리	111557	20	장난감 애니메이션	48047
9	파워레인저	103156	21	영어동요	44096
10	캐비키즈	99917	22	토이팩토리	43890
11	상상놀이터	81678	23	다이노포스	41938
12	장난감진구	81419	24	역체괴물	40483

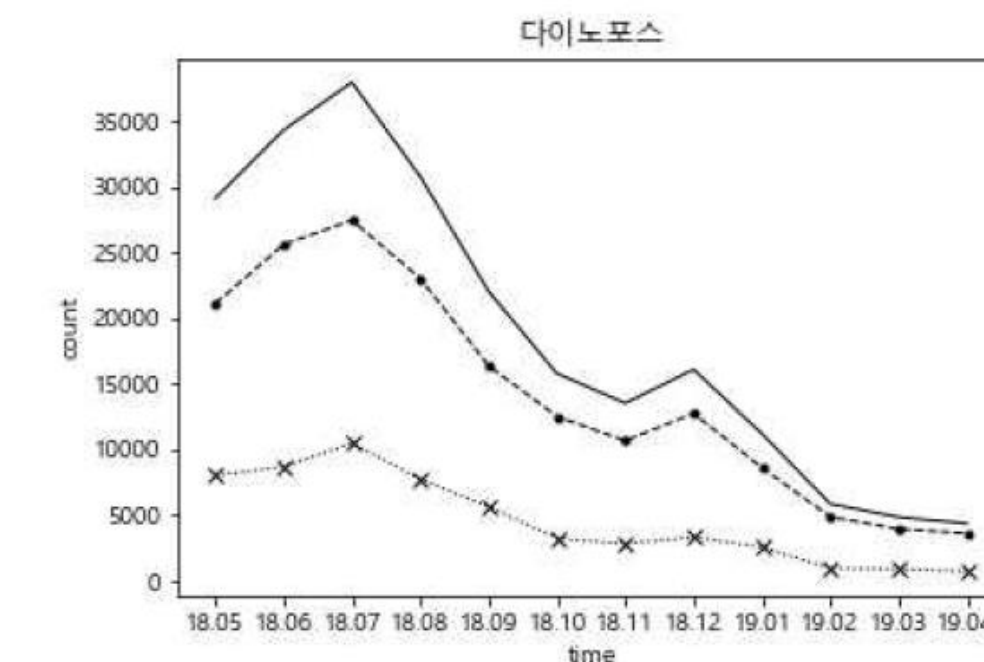
다이노포스			인형놀이		
time	긍정	부정	time	긍정	부정
18.05	21087	8025	18.05	10427	5421
18.06	25591	8714	18.06	9275	4029
18.07	27439	10491	18.07	9041	4123
18.08	23004	7756	18.08	9839	4054
18.09	16374	5649	18.09	10010	5175
18.10	12518	3265	18.1	10495	5450
18.11	10683	2852	18.11	11347	6381
18.12	12749	3346	18.12	14972	7667
19.01	8617	2581	19.01	19172	10169
19.02	4885	983	19.02	18055	9480
19.03	3967	905	19.03	19993	10980
19.04	3639	742	19.04	20103	12190

상향적 추세



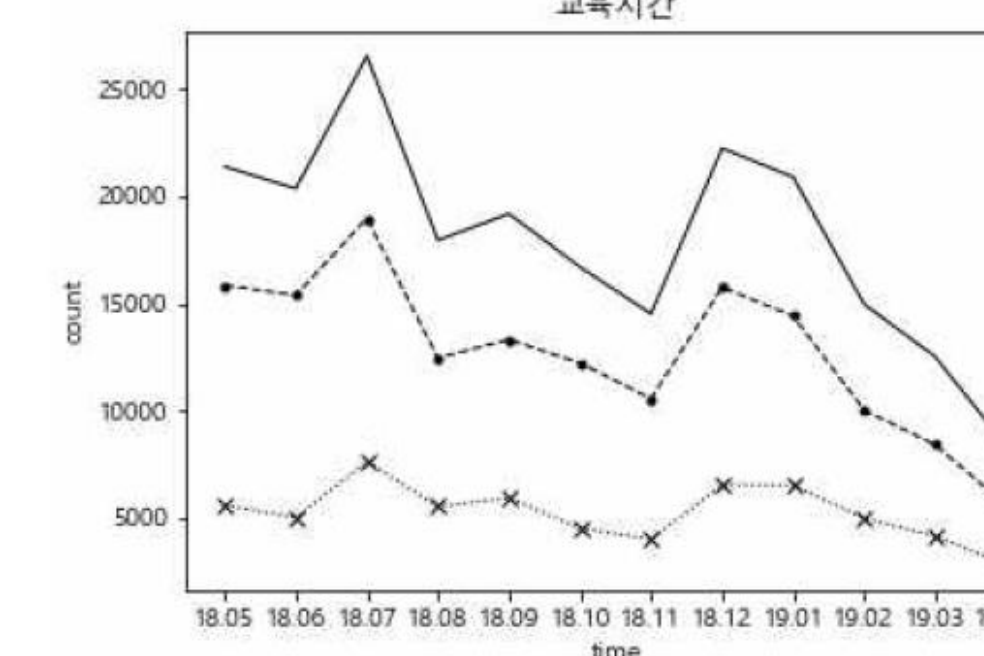
Max number of Comment (N)	Representative Keywords
N > 10000	인형놀이, 상황극, 상상이야기, 밀착중계, 사우스
N > 5000	병원놀이, 바비인형, 구급차
N > 1000	과학도감, 날카도, 영어동요, 캐비키즈, 추원영상
N > 100	관경 동영상, 라이센스, 래빗미디어, 상상터라기, 세계공룡싸움

하향적 추세



Max number of Comment (N)	Representative Keywords
N > 10000	다이노포스, 뽀로로, 시크릿유우, 터닝메카드, 파워레인저
N > 5000	변신로봇, 베이블레이드, 캐리엔송, 토이박스, 디즈니애니메이션
N > 1000	인기동화, 인형드라마, 스타워즈, 홈터머스, 마인크래프트
N > 100	로봇드레인, 코뿔안진이야기, 토이스토리, 활동동요, 미니피규어

계절적 추세



Max number of Comment (N)	Representative Keywords
N > 10000	겨울왕국, 교육시간, 포타버스, 인스타그램, 디즈니
N > 5000	대문밖장난감, 영어공부, 색칠공부, 신제품, 포켓몬스터
N > 1000	생화습관, 장려가족, 영이그림사전, 동화이야기, 놀이공원
N > 100	과학오기심, 바다탐험대, 적인놀이, 시사상식, 미니언즈

- 키워드 분석을 통해 도출된 가장 영향력 있는 단어는 ‘장난감’으로써 유아, 아동분야 유튜브에서 다양한 장난감을 활용한 콘텐츠가 활발히 제작됨을 확인
- 또한 ‘칠교놀이’, ‘색칠놀이’, ‘공룡탐험’, ‘과학도감’, ‘변신로봇’ 등의 관심주제는 과거 컬러북과 그림책으로 소비 되었던 각종 유아 대상 놀이들이 유튜브 영상으로 제작되어 아이들에게 제공됨을 확인

5. 토의 및 결론

- 본 연구는 유튜브 영상에 대한 소비자의 관심도를 시계열적으로 표현하고 추세에 대한 분류를 통해 시청자들의 관심주제를 파악한다.
- 상향적 추세를 보이는 관심주제를 포함한 콘텐츠는 가까운 미래에도 유아, 아동 영상의 주요 소비자인 아동과 그 부모에게 관심의 대상이 될 것임을 알 수 있다.
- 하향적 추세의 관심주제들은 지난 1년간 급격한 관심의 하향세에 있기 때문에 미래 유튜브 콘텐츠로의 제작은 경쟁력이 없음을 예측할 수 있다.
- 유튜브 크리에이터를 지원하고 콘텐츠 제작에 도움을 주는 MCN 비즈니스가 성장하고 있는 시기에 유망한 콘텐츠의 주제를 파악했다는 것에 의의를 가진다.