

KUIAI

데이터 기반 패션 인플루언서 추천 알고리즘

안암어스(Anam-Us)

산업경영공학부 2020170815 김경호

산업경영공학부 2020170823 윤화평

산업경영공학부 2020170812 임정섭

산업경영공학부 2020170810 최대원

목차

1. 개요	03
가. 프로젝트 배경	03
나. 프로젝트 필요성 및 기대효과	04
2. 프로젝트 내용	04
가. 추천시스템 활용 목표	04
나. 데이터 크롤링 방법 설명	04
다. 데이터 분석 내용 및 결과	05
라. 추천 알고리즘 내용	07
3. 프로젝트 결과	09
가. 추천시스템 흐름도	09
나. 추천시스템 결과 및 고찰	10
4. 구성원별 역할 및 개발일정	13
5. 참고문헌	13

1. 개요

가. 프로젝트 배경

인플루언서(Influencer)란, ‘영향력 있는 사람’이란 뜻으로, 인스타그램, 유튜브 등 SNS에서 수십만 명의 구독자(팔로워)를 보유한 ‘SNS 유명인’을 말한다. 최근 인스타그램이 쇼핑검색의 플랫폼으로 진화함에 따라 인스타그램을 활용한 커머스, 마케팅 플랫폼 서비스들이 급격히 증가하였다. 특히 우리나라의 젊은 세대(10대 ~ 30대)의 경우, SNS 이용 비율 중 인스타그램의 비율이 가장 높다는 조사 결과와 인플루언서 시장의 성장 추세를 보았을 때 앞으로 그 영향력은 더욱 커질 것이라 예상할 수 있다.

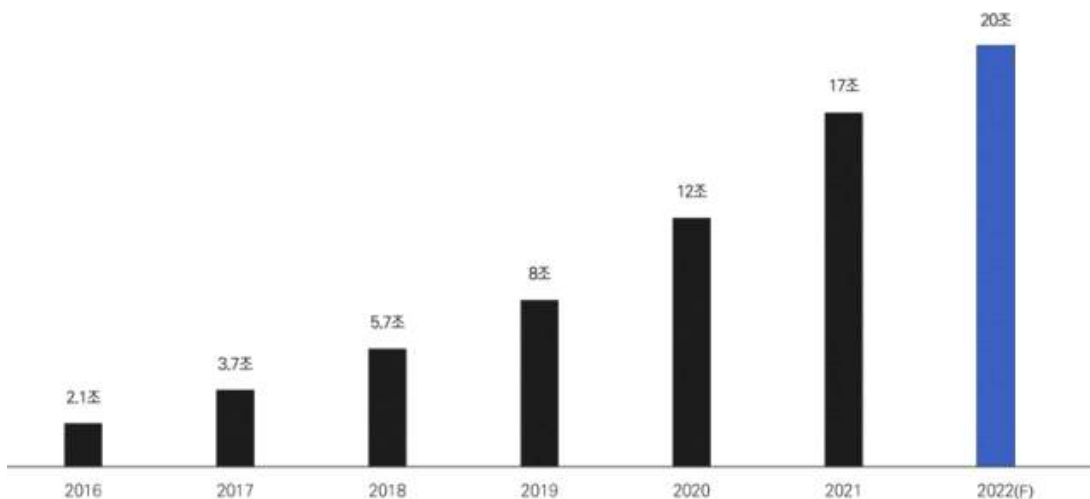
연령별 가장 많이 이용하는 소셜 미디어 Top3 (단위: 명)

	10대	20대	30대	40대	50대
1위	 (212만 6,377명)	 (501만 7,263명)	 (451만 2,998명)	 (467만 3,910명)	 (544만 7,025명)
2위	 (184만 2,504명)	 (339만 1,023명)	 (279만 655명)	 (309만 763명)	 (241만 7,511명)
3위	 (135만 8,667명)	 (168만 9,323명)	 (233만 4,458명)	 (258만 5,065명)	 (190만 2,108명)

Note: 월 평균 이용자 수는 2021년 1분기(1~3월) 내 월별로 발생한 모바일(Android App+iOS App) 이용자 수의 산술평균값

1)

글로벌 인플루언서 마케팅 시장 규모
인플루언서 마케팅허브, 2022



2)

1) 닐슨 코리아클릭, 2021년 1분기 포털&SNS 보고서

인플루언서 마케팅이 가지는 가장 큰 특징 중 하나는 고객을 또 다른 마케터로 이용할 수 있다는 점이다. 이러한 특징은 기존 TV 광고와 구분되는 가장 큰 특징이다. 예를 들어, 인플루언서의 해시태그와 소비자가 똑같은 해시태그를 게시하는 경우 마케팅의 효과를 얻을 수 있기에 그 영향력이 더욱 크다고 볼 수 있다. 또 다른 특징으로는, 과거엔 단순히 팔로워 수와 비례하여 효율성을 생각했다면, 최근에는 팔로워 수뿐만이 아니라 브랜드와 제품 특성에 따른 팔로워 1만명 이상의 마이크로 인플루언서를 활용하는 사례도 늘어나고 있다.³⁾ 이러한 배경에 따라 인플루언서 마케팅은 그 규모와 영향이 확대되고 있으며 패션 상품도 인플루언서 마케팅의 주요 품목 중 하나로 인스타그램 인플루언서의 마케팅의 큰 영향을 받고 있다.⁴⁾

나. 프로젝트 필요성 및 기대효과

패션 제품의 측면에서 바라보았을 때, 제품을 판매하는 브랜드 입장에서는 제품의 분위기와 어울리는 인플루언서에게 광고를 맡기고 싶을 것이고, 소비자 또한 마찬가지로 본인과 가장 잘 매칭되는 인플루언서의 영향을 많이 받을 것이다. 이에 따라 상품 정보에 따른 인플루언서 추천, 소비자의 인스타그램 피드에 따른 성향을 통한 인플루언서 추천을 통해 각각의 니즈에 맞는 추천시스템을 구현할 수 있다. 이를 통해, 브랜드 입장에서는 마케팅 효율을 증진시키며, 소비자 입장에서는 효율적인 상품 구매가 이루어질 수 있도록 하는 데에 의의가 있다.

2. 프로젝트 내용

가. 추천시스템 활용 목표

- 1) 상품 정보에 따른 인스타그램 패션 인플루언서 추천
- 2) 소비자 정보에 따른 인스타그램 패션 인플루언서 추천

나. 데이터 크롤링 방법 설명

알고리즘은 기본적으로 주어진 데이터를 기반으로 하여 제작했다. raw data에 새로운 필드를 추가하진 않았다. 완성된 알고리즘에 Test 용도로 입력해볼 소비자 데이터를 추가 수집하였다. 수집한 데이터는 소비자의 인스타그램 피드 정보로, 주어진 소비자 데이터와 같은 형식(이미지 + 이미지명과 동일한 post_id, user_name, text, timestamp)를 가진다. 인스타 계정을 소유한 팀원 둘과 유명 연예인인 아이유의 인스타 계정에 대해 크롤링을 수행했다. timestamp는 완성된 알고리즘에서 활용되지 않으므로 값을 0으로 고정했다.

2) 인플루언서 마케팅허브, 2022

3) 인플루언서 마케팅시장 6년간 10배 성장(주간코스맥틱)

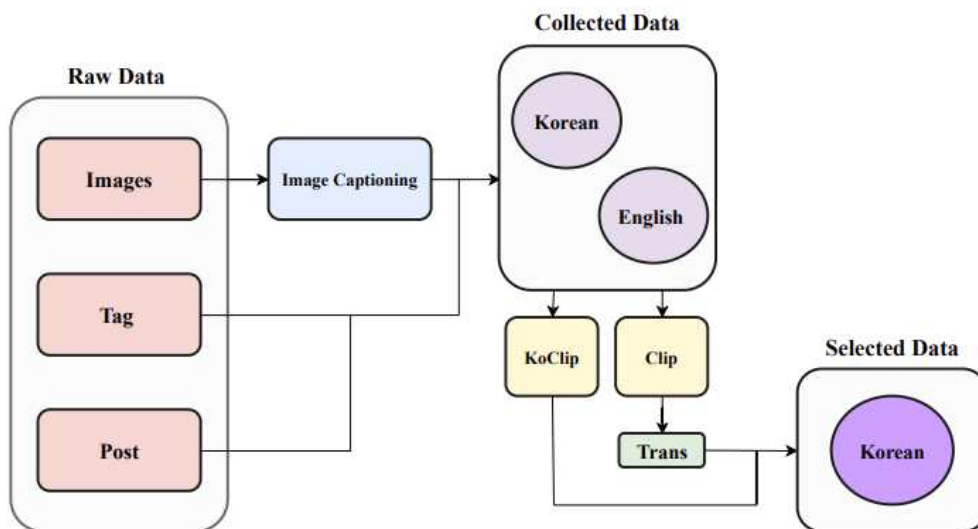
4) 위키백과 - 인플루언서 마케팅

다. 데이터 분석 내용 및 결과

프로젝트의 Raw Data로, 상품 정보(무신사 크롤링, 이미지 오픈 데이터셋), 소비자 정보, 인플루언서의 피드 데이터가 주어졌으며 추가적인 소비자의 경우 크롤링 방법에서 언급한 소비자의 인스타그램 피드정보를 크롤링하여 수집하였다. 상품 정보는 이미지, Object Key, Category, Look_name, Style_text, Tag, Product의 항목들로 이루어져 있으며, 하나의 상품에 대해 다양한 이미지가 존재하였다. 인플루언서 피드 데이터와 소비자 피드 데이터는, 피드 이미지, 인플루언서 ID(user_name), post_id, text, timestamp으로 구성되어있다. 이 중에서 우리가 사용한 데이터는 다음과 같다.

데이터 항목	사용 데이터 항목
상품정보 데이터	[Image Data], [Category, Look_name, Style_text, Tag]를 상품정보에 대한 Tag Data 하나로 합쳐서 추출.
인플루언서 피드 데이터	인플루언서 ID Data, 게시물 Image Data, 게시물에 해당하는 Post Data(게시글), Tag Data(해시태그)만을 따로 추출.
소비자 피드 데이터	인플루언서 ID Data, 게시물 Image Data, 게시물에 해당하는 Post Data(게시글), Tag Data(해시태그)만을 따로 추출.

전체적인 데이터 전처리의 단계는 아래의 흐름도와 같이 진행하였다. 참고로 아래 전처리과정은 상품정보, 인플루언서 정보, 소비자 정보 종류에 관계없이 가지고 있는 정보 유형 중 어떤 세가지 정보 유형(Image, Tag, Post)에 속하는지만 판단하여, 필요한 파이프라인만 사용하면 된다. 가령 이미지 데이터가 존재하지 않는 경우, 아래 흐름도에서 Image 관련 파이프라인은 거치지 않아도 된다. (Tag, Post도 마찬가지)



Raw Data의 모든 Image(상품정보, 인스타그램 피드 이미지)에 대해 Image Captioning 모델을 적용하여 이미지에 해당하는 영어로 이루어진 문장 Text를 추출하였다. Image Captioning을 적용한 이유는 1) 인스타그램 게시물의 경우, 이미지와 상관관계가 거의 없는 Text Data(Tag, Post)만 존재하는 경우가 존재하고, 2) 인스타그램 게시물에 피드 이미지 외에 Text Data가 아예 존재하지 않는 경우가 있기 때문이다. Image Captioning은 이미지를 input으로 넣으면, 이미지에 대한 description을 output으로 추출할 수 있다. 이러한 description에는 이미지를 잘 표현할 수 있는 단어들이 존재하기에, 이를 활용하여 이미지에 대해서 부족한 Text Data를 보완할 수 있다. 사용한 Image Captioning 모델은 NIC(Neural Image Caption)모델⁵⁾로, output으로 한글이 아닌 영어로만 구성된 description이 나오게 된다. 참고로, Image Captioning 진행하기 전 이미지에서 패션과 관련된 부분만 Object Detection을 통해 먼저 뽑아내는 방식도 생각했지만, 이미지의 패션 외에도 배경과 같은 요소도 인플루언서 혹은 상품, 소비자의 분위기를 잘 나타낼 수 있는 요소라고 생각하여 원본 이미지를 그대로 활용하는 방향으로 결정했다. 즉, 우리는 단순히 패션만을 보고 인플루언서를 추천하는 방향은 옳지 않다고 생각했다.

Image Captioning을 통해 추출한 description, Tag, Post Data는 먼저 게시물 기준으로 모두 Join 했고, '#, 이모티콘'과 같은 특수문자를 공백으로 대체한 뒤, 공백을 기준으로 단어를 분리하는 과정을 거쳤다. 분리 과정을 거친 뒤에는 아래와 같이 한글만, 영어만, 한글+영어, 숫자만, 한글+숫자로 이루어진 경우로 데이터가 구분되었고, 히스토그램과 데이터 샘플을 관찰함으로써 최종적으로 한국, 영어로 구성된 데이터만 사용하기로 결정했다.

데이터 종류	사용 여부	이유
한글만	o	빈도수가 많고, 이미지를 표현할 수 있는 데이터가 충분히 많음.
영어만	o	빈도수가 많고, 이미지를 표현할 수 있는 데이터가 충분히 많음.
한글+영어	x	빈도수가 적고, 한글과 영어 중 선택하는 공수가 큼.
숫자만	x	빈도수가 적고, 숫자가 단독으로만 있는 경우에는 유의미한 정보를 담고 있지 않는다고 판단.
한글+숫자	x	빈도수가 적고, 숫자와 연관되지 않은 한글(=단위가 아닌 경우)라면 무의미하다.

위와 같은 일련의 과정을 통해 모아진 Data는 1) 한글만 혹은 2) 영어만 이루어진 데이터이고, 특수문자가 제거된 데이터이다. 먼저 정확한 형태소 분석을 위해, 게시물마다 공백으로 구분했던 단어를 다시 공백을 기준으로 Join 했다. 1) 한글로만 이루어진 데이터의 경우, 한글형태소 분석기 OKT를 이용하여 형태소 단위로 토큰화하였고, 한글 불용어 사전을 이용해 불용어에 해당하는 토큰을 제거했다. 2) 영어로만 데이터의 경우, 영어 전용 자연어처리 라이브러리인 nltk의 토큰라이저를 이용해 토큰화를 진행했고, stopwords 목록을 통해 불용어를 제거했다. 최종적으로 수집된 데이터는 게시물마다 토큰화된 한글/영어 Text 데이터이다.

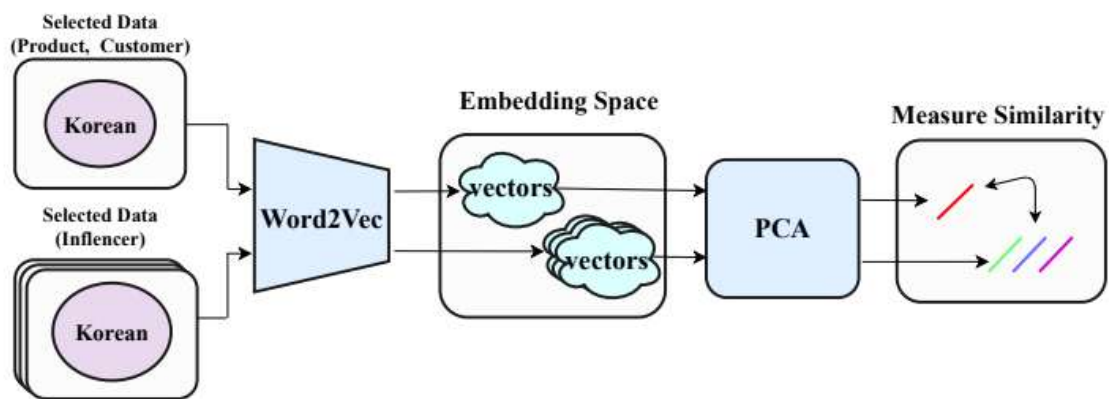
5) Vinyals, Oriol, et al. "Show and tell: A neural image caption generator." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.

위 과정을 거친 뒤에도, 게시물의 피드 이미지와 전혀 관계가 없는 Text(ex. img: 자전거-tag: 맛팔)는 존재하였다. 이를 제거하기 위해서 CLIP⁶⁾을 사용하였다. CLIP은 Image와 Text pair를 받아서, 각각을 encoder를 통과시킴으로써, 임베딩 공간으로 projection한 뒤, 공간 상에서 distance를 측정할 수 있는 도구이다. 여기서 distance는 곧 Input으로 들어온 Image와 Text 사이의 연관성으로 해석할 수 있다. 우리는 이러한 CLIP을 이용하여 게시물의 피드 이미지와 Text 사이의 연관성을 수치화한 다음, 이를 기준으로 관련있는 Text만을 뽑아서 사용하기로 했다. 구체적으로 사용 모델에 대해서 설명하자면, 영어 Text는 기존 CLIP을 사용했고, 한글 Text는 한글 Text를 Input으로 받을 수 있는 CLIP인 KoCLIP⁷⁾을 사용했다.

CLIP을 통한 연관성 반환 값을 기준으로, 한글/영어 Text에 대해서 각각 상위 6개의 단어를 각각 추출하여 비교집단으로 구성하였다. 두 집단을 설정한 이유는, 추천시스템 알고리즘 적용 단계에서 두 집단의 비교를 통해 더 좋은 성능을 보이는 파라미터를 선택하기 위함이다. 이후, 최종적인 추천 알고리즘 단계에서의 구현을 위해 Python의 Google Trans 패키지를 활용하여 영어는 한글로 변환하였고, 최종적으로는 CLIP을 통해 피드 이미지와 연관성이 높은 토큰화된 한글 Text 데이터를 얻을 수 있었다. 참고로 영어를 먼저 한글로 번역한 뒤에 koCLIP을 사용하지 않고, 한글, 영어 각각을 CLIP, koCLIP을 돌린 뒤에 추출된 영어 토큰만 번역한 이유는 영어를 먼저 한글 Text로 번역하는 과정에서 오역으로 인해, 피드 이미지를 나타내는 중요 토큰이 사라질 수도 있겠다고 판단했기 때문이다.

라. 추천 알고리즘 내용

전처리 후에 진행되는 추천 알고리즘의 전체적인 과정은 다음과 같다.



6) Radford, Alec, et al. "Learning transferable visual models from natural language supervision." *International Conference on Machine Learning*. PMLR, 2021.

7) Park, Sungjoon, et al. "Klue: Korean language understanding evaluation." *arXiv preprint arXiv:2105.09680* (2021).

다.에서 추출된 최종 데이터(위 그림에서 Selected Data로, 상품, 소비자, 인플루언서 정보)는 각 게시물의 피드이미지와 높은 연관성을 보이는 한글 토큰들로만 구성된다. 하지만 토큰들 같은 경우, 특히 인플루언서의 경우, 30개의 게시물마다 존재하기에 추천 알고리즘을 만들기에 여전히 많은 수가 존재한다. 이를 해결하기 위해서, 많은 수의 토큰들을 representation 할 수 있는 하나의 feature를 추출할 필요가 있다고 판단했다. 이를 위해, 먼저 토큰들을 적절한 feature space로 projection 시킬 수 있는 Word2Vec 모델을 사용하였다. Word2Vec은 단어(=토큰)들 사이의 연관성을 고려하여 단어를 적절한 임베딩 공간으로 projection하는 대표적인 자연어 처리 방법이다. Word2Vec은 여러 가지 방법론이 존재하지만, 우리는 추출된 데이터가 한글로만 구성된 데이터라는 점에서 충분히 큰 크기의 한글 Coupus에서 Pretrain 된 Word2Vec 모델을 load한 뒤, 가지고 있는 데이터(상품, 소비자, 인플루언서)를 활용해 이를 FineTuning하는 방식을 택했다. 위 과정을 통해 도출된 Word2Vec을 사용하여 데이터의 한글 토큰들을 모두 임베딩 공간을 projection했다. 단어들이 vector로서 임베딩 공간으로 projection된 뒤에는, 이러한 vector들을 representation하는 하나의 feature를 찾아야 했다. 이는 하나의 차원축소 문제라고 판단했고, 대표적인 차원축소방법인 PCA를 선택했다.

PCA를 활용하면, 주어진 데이터의 분산을 최대로 보존할 수 있는 Component를 계산할 수 있다. 우리는 이 Component를 많은 수의 토큰 vector를 대표할 수 있는 하나의 feature로 고려했다. PCA를 통해 상품데이터, 소비자데이터, 100명의 인플루언서를 각각 대표하는 하나의 feature를 계산할 수 있었고, 도출된 대표 feature를 이용해 최종적으로 추천 시스템의 종류에 따라 다음과 같이 3명의 인플루언서를 추천할 수 있었다.

1) 상품정보를 input으로 받는 인플루언서 추천 시스템

-> 상품 정보를 대표하는 feature를 100명의 인플루언서의 feature와 cosine similarity를 계산한 뒤, similarity가 높은 순서대로 3명의 인플루언서 추천

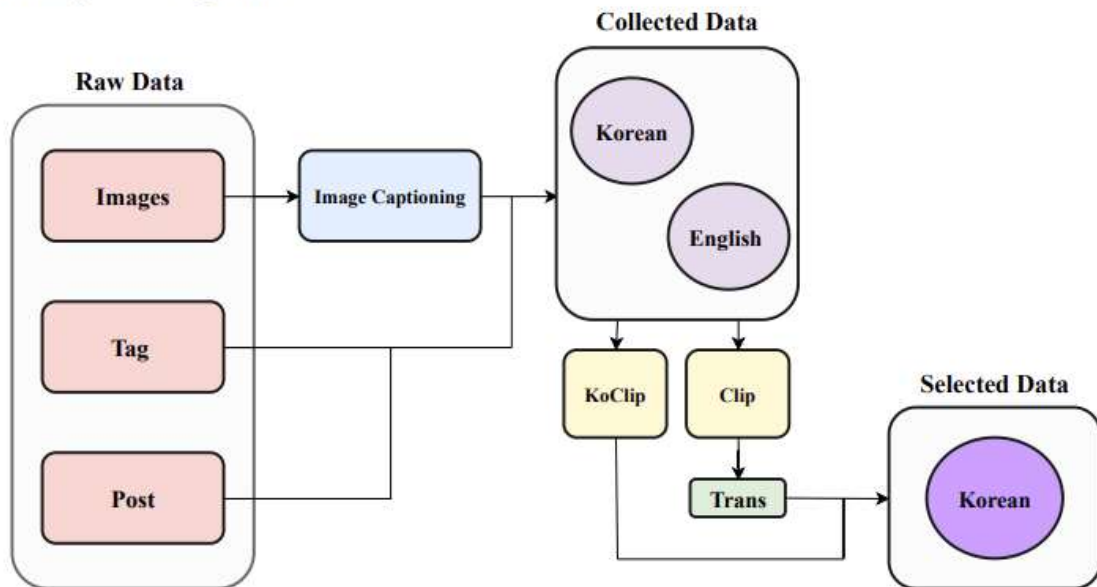
2) 소비자정보를 input으로 받는 인플루언서 추천 시스템

-> 소비자의 인스타그램 정보를 대표하는 feature를 100명의 인플루언서의 feature와 cosine similarity를 계산한 뒤, similarity가 높은 순서대로 3명의 인플루언서 추천

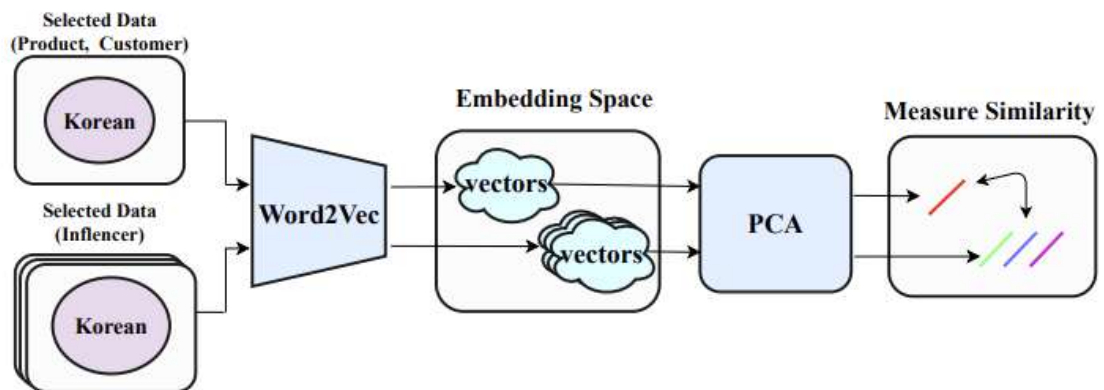
3. 프로젝트 결과

가. 추천시스템 흐름도

1. Preprocessing Phase



2. Recommending Phase



나. 추천시스템 결과 및 고찰

1) 추천시스템 결과

< 상품 - 인플루언서 간 유사도 결과 일부 > (각각의 row : 특정 product와 각 인플루언서 간의 유사도)

	product	@98.c_project	@ys971003_	@malko_bee	@hotneul	@jung_staas	@oneh6_	@dosirak_hansol	@j0ng_v
0	29960	0.934354	0.922696	0.917094	0.945321	0.961494	0.892225	0.926986	0.868224
1	29961	0.954035	0.923614	0.917064	0.921437	0.954260	0.926885	0.928888	0.903957
2	29202	0.874121	0.895390	0.884349	0.840655	0.880517	0.850780	0.896602	0.876683
3	29975	0.915689	0.891278	0.878399	0.912109	0.933238	0.869494	0.896074	0.840183
4	29977	0.941729	0.946088	0.923245	0.910751	0.949873	0.908564	0.941815	0.919372
5	29978	0.947014	0.961845	0.938965	0.893514	0.953039	0.919733	0.955614	0.935137
6	29980	0.975010	0.961818	0.970712	0.933595	0.972720	0.970750	0.972656	0.964776
7	29981	0.962463	0.925455	0.925174	0.927679	0.953441	0.939994	0.926446	0.904816
8	29469	0.950920	0.943327	0.945171	0.959543	0.975690	0.920609	0.952922	0.904782
9	28964	0.970763	0.949306	0.938004	0.930964	0.970794	0.942939	0.948055	0.922761
10	29349	0.936877	0.922609	0.926094	0.943902	0.966183	0.904584	0.940615	0.889729
11	29222	0.895690	0.879790	0.850900	0.868202	0.877112	0.870557	0.850767	0.871201

- 상품의 경우 하나의 상품에 대해 detail_<product>_?_????.jpg의 형식으로 여러 장의 사진이 존재해 product로 각각의 상품을 구분함

< 소비자 - 인플루언서 간 유사도 결과 일부>

	customer_name	@98.c_project	@ys971003_	@malko_bee	@hotneul	@jung_staas	@oneh6_	@dosirak_hansol	@j0ng_v
0	@kxungho00	0.9781163457	0.9826476826	0.9675105267	0.9268311325	0.9814590881	0.9563510578	0.9774128289	0.95259
1	@agree_seop	-0.6151088777	-0.6099531288	-0.5924101704	-0.4785684632	-0.5460134198	-0.6542346514	-0.581567022	-0.656036
2	@dlwlrma	0.9308180811	0.9207614896	0.8938477379	0.9005692303	0.9469854134	0.8806722904	0.9177863736	0.864003
3	@j_g_ok	-0.8246188025	-0.872770212	-0.8766349758	-0.7286292524	-0.8174754024	-0.8531032648	-0.8794519783	-0.8959

< 상품 기준 추천 인플루언서 결과 일부 >

	product	1st	2nd	3rd
0	29960	@by_he.nique	@ap.s.fi1st	@_v.yuum.look_
1	29961	@lamode.seoul	@c_eunnnnnnn	@y_mood_h
2	29202	@yj_mark	@kimyannnnngh	@bloggerbok
3	29975	@ap.s.fi1st	@lamode.seoul	@by_he.nique
4	29977	@loolinmx	@yj_mark	@skuukzky
5	29978	@bloggerbok	@chaileeson	@skuukzky
6	29980	@so_j2	@sympa_young	@one_r_k
7	29981	@lamode.seoul	@y_mood_h	@337janggoon
8	29469	@_v.yuum.look_	@jung_staas	@0nefence
9	28964	@lamode.seoul	@c_eunnnnnnn	@y_mood_h
10	29349	@_v.yuum.look_	@so_love_so_	@ap.s.fi1st
11	29222	@so_h_appy	@jung_staas	@hotneul
12	29991	@jung_staas	@so_h_appy	@so_j2
13	29992	@so_h_appy	@jung_staas	@hotneul
14	29996	@jindalorian	@zxcvr0626	@jin_pyo_is
15	28973	@so_h_appy	@jung_staas	@hotneul
16	30000	@jin.wonder	@lamode.seoul	@_v.yuum.look_
17	30001	@rozley_y	@337janggoon	@so_j2
18	30006	@lamode.seoul	@0nefence	@c_eunnnnnnn

- 상품의 경우 하나의 상품에 대해 detail_<product>_?_????.jpg의 형식으로 여러 장의 사진이 존재해 product로 각각의 상품을 구분함

< 소비자 기준 추천 인플루언서 >

	customer_name	1st	2nd	3rd
0	@kxungho00	@0nefence	@y._dulcet	@so_love_so_
1	@agree_seop	@jin_pyo_is	@jindalorian	@dismas_
2	@dlwlrma	@jin.wonder	@_v.yuum.look_	@ap.s.fi1st
3	@j_g_ok	@jin_pyo_is	@jindalorian	@zxcvr0626

< 상품 기준 추천 인플루언서 결과 예시>

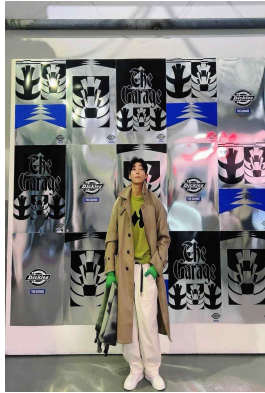
product(29202)

1st(@yj_mark)

2nd(@kimyannnnngh)

3rd(@bloggerbok)

MUSINSA



product(29992)

1st(@so_happy)

2nd(@jung_staas)

3rd(@hotneul)

MUSINSA



< 소비자 기준 추천 인플루언서 결과 예시>

customer(@agree_seop)

1st(@jin_pyo_is)

2nd(@jindalorian)

3rd(@dismas_)

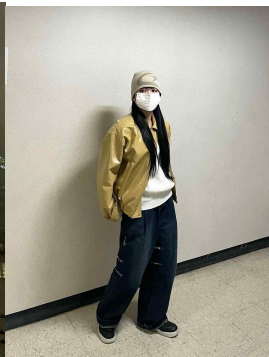


customer(@kxungho00)

1st(@0nefence)

2nd(@y._.dulcet)

3rd(@so_love_so_)



2) 추천시스템 고찰

이번 프로젝트에서 제시한 추천 시스템은 Image Captioning, CLIP을 활용함으로써 이미지, 텍스트와 같은 비정형 데이터를 효과적으로 반영할 수 있다는 점에서 가장 큰 의의가 있다. 또한 Word2Vec, PCA를 활용하여 상품, 소비자, 인플루언서 각각을 대표하는 하나의 feature를 계산한 뒤 코사인 유사도를 기준으로 추천하는 원리는 인스타그램 외에도 트위터, 유튜브 등 Image, text 데이터가 존재하는 어떠한 플랫폼에서도 General 하게 잘 작동될 수 있음을 알 수 있다.

이러한 알고리즘의 특징을 바탕으로 플랫폼 서비스(가명: 모도리)를 제안한다. 모도리는 빈틈없이 아주 여무진 사람이라는 뜻이다. 모도리 서비스는 한 사람에 대해 빈틈없이 모든 흔적을 모으고, 다른 사람을 추천해주는 플랫폼이다. 모도리는 플랫폼의 플랫폼으로도 정의할 수 있다. 개인이 가입한 다양한 플랫폼을 연동하여 그 사람에 관련된 플랫폼 서비스를 모아둔다. 이때 연동 플랫폼은 SNS 뿐만 아니라 후기를 남기는 배달의 민족 등 사진과 글을 남기는 모든 서비스가 해당된다. 모도리 서비스의 성공은 연동 플랫폼을 구하는 일에 달려있다. 모도리는 상대 사용자를 추천할 때 그 사용자의 연동 플랫폼 계정 목록을 함께 보여주어 연동 플랫폼의 유입을 늘리는 효과가 있으므로 이러한 효과를 연동 플랫폼 모집에 적극 활용해야 한다.

한편 우리의 연구는 시간적인 제약으로 인해 몇 가지 측면에서 완벽하게 검증되지는 않았기에, 다음과 같은 세가지 연구방향을 제시하고자 한다. 첫 번째는 다양한 비교군의 대한 실험 가능성이다. 제시한 추천 시스템의 과정에서 여러 가지 비교군에 대한 실험이 필요하다. 먼저 ‘반영할 태그의 개수’ 측면에서 3개와 6개만 비교했지만 이 이외의 다른 개수의 Tag를 사용했을 때, 추천 시스템이 더 좋은 성능을 보일 가능성을 배제할 수 없다. 다음으로 ‘Feature Space의 선택’이다. 이번 추천 시스템에서는 Word2Vec을 사용하여 단어들을 임베딩 공간으로 보냈지만, BERT와 같은 Encoder를 통과시킨 Feature Space 등 여러 가지 Feature Space에 대한 선택지가 가능하다. 마지막으로 ‘유사도 측정 방식’의 선택이다. 코사인 유사도를 사용했지만 두 벡터사이의 유사도를 측정하는 방식은 Euclidean Distance, Manhattan Distance, Minkowski Distance 등 많은 선택지가 존재한다. 따라서 이러한 비교군 중 어떤 측정 방식이 우리의 추천시스템과 잘 matching되는 지는 확인해봐야 하는 문제이다. 두 번째는 비디오 영상 활용 문제이다. 주어진 샘플 데이터를 보면 인스타그램 게시물 중 비디오가 포함된 게시물들이 간혹 존재한다. 우리는 이러한 비디오 정보를 활용하기 위해서 프레임별로 쪼개서 이미지로 반영하는 등 여러 방식으로 생각했지만, 결과적으로 반영하지 못했다. 이러한 비디오 인플루언서, 상품, 소비자를 나타낼 수 있는 중요한 정보로 활용가능성이 있기 때문에 이에 대한 연구가 필요하다. 세 번째는 추천 시스템의 객관적인 평가지표의 부재이다. Input으로 넣어준 상품 혹은 소비자정보와 추천된 3명의 인플루언서가 얼마만큼 관련성이 있는지는 결국 사용자의 주관에 맡길 수 밖에 없다. 가령 위 결과에서 제시한 이미지들이 우리 팀 입장에서 Input 정보를 나름대로 표현하는 추천 인플루언서의 이미지라고 판단하고 가져왔지만, 사용자 혹은 제3자 입장에서 봤을때는 불충분하다고 느끼는 경우가 있을 수 있다. 이렇게 저마다 평가하는 것이 다르기에, 누구나 납득할 수 있는 객관화된 평가지표를 개발할 필요가 있다.

5. 구성원별 역할 및 개발일정



6. 참고문헌

닐슨 코리아클릭, 2021년 1분기 포털&SNS 보고서

인플루언서 마케팅허브, 2022

인플루언서 마케팅시장 6년간 10배 성장(주간코스매틱)

<https://www.geniepark.co.kr/news/articleView.html?idxno=49650>

위키백과 - 인플루언서 마케팅

https://ko.wikipedia.org/wiki/%EC%9D%B8%ED%94%8C%EB%A3%A8%EC%96%B8%EC%84%9C_%EB%A7%88%EC%BC%80%ED%8C%85

Vinyals, Oriol, et al. "Show and tell: A neural image caption generator." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.

Radford, Alec, et al. "Learning transferable visual models from natural language supervision." *International Conference on Machine Learning*. PMLR, 2021.

Park, Sungjoon, et al. "Klue: Korean language understanding evaluation." *arXiv preprint arXiv:2105.09680* (2021).