

현대 전자상거래 플랫폼에서의 멀티 모달리티 정보검색 정확도 향상 연구

오윤정

성균관대 데이터사이언스학과
yj.oh@g.skku.edu

The 2020 KDD Cup Comopetition: Modern E-commerce platform multimodality recall algorithm research and development

Yoon jung OH

Data Science Convergence Department of SungKyunKwan University
yj.oh@g.skku.edu

KEY WORDS: E-commerce, Multi-modality, ImageBERT, Banary Classification, nDCG@5, Optimal Retrieval Search in E-commerce

ABSTRACT: 현대 이커머스 시장의 핵심화두인 멀티모달리티 연구를 KDD cup 대회의 데이터셋과 검증방식(nDCG@5)를 통해 최적의 검색값을 제공하는 모델을 구현하고 성능을 분석한다. 연구방법은 다음과 같다. 1) ImageBert를 활용한 input data 전처리를 통해 product label과 boxfeature 임베딩해서 데이터 청크의 대표값을 뽑고, 2) 질의어를 BERT 통해 임베딩한 데이터청크의 대표값을 뽑아 최종적으로 이진분류(Binary Classification)한 결과의 정확도를 측정한다. 이때 질의어와 이미지 임베딩값은 동일한 웨이트 값을 공유하여 평가지표인 nDCG@5의 최적 추천도로 0.5999의 결과 값을 얻을 수 있다. 이 모델의 개선점과 향후 성능향상 방안을 통해 이커머스 시장의 최적추천 검색 (Optimal Retrieval Search) 고도화에 기여하고자 한다.

Multimodal: "...a phenomenon or a system is observed using multiple instruments, measurement devices or acquisition techniques. In this case, each acquisition framework is denoted as a modality and is associated with one dataset. The whole setup, in which one has access to data obtained from multiple modalities, is known as multimodal." [Lahat et al., 2015]

1. Introduction

전 세계적으로 e-commerce 시장의 거래액은 3.46천조원에 이르며 전년대비 84.6%증가한 2.93천조원으로¹ 매년마다 가파른 성장세를 띄고 있다. 또한 국내시장역시 19년 기준 109조원의 거래액을 기록하며 '100조'억원²의 고지를 넘어섰다. 시장의 볼륨이 증가함에따라 소비자들의 행동분석과 유입률 강화는 e-commerce 시장의 중요 지표가되었다. 연구에 따르면 관련 품목에 비일치도가 높은 페이지나 검색결과에대한 이탈율(Bounce Rate)에 따라 고객의 구매까지 이어지는 구매 깔대기(Digital Marketing Purchasing Funnel)는 단일

페이지당 영향도가 높은것³으로 조사되었다. 시장이 고도화됨에 따라, 고객 맞춤형(Personalised Customer) 동적 웹페이지에 대한 요구가 늘어가고 있는 것이다. 이에 따라 이커머스 포탈내의 자연어 질의와 추천 상품의 연관도 및 상관관계 추적이 중요한 이슈로 떠오르고 있다.

1-1. 대회설명

중국의 최대 전자상거래 업체 알리바바(Alibaba.com)에서 고객의 질의어에 가장 적합한 이미지를 추천하는 파이프라인 모델 구현을 목표로 KDD cup의 track1의 주제로 multimodalities recall 대회를 오픈하였다. 데이터의 구성은

¹ Digitalcommerce360.com, Jessica Young(Nov,13,2019)

² 전자신문,윤희석기자(16,12,2019),
<https://m.etnews.com/20191216000231>

³ Rui Machado et el, Building Retrieval System for a Fashion e-commerce Company (2018)

약 500개의 질의어 데이터셋과 30Giga byte 2만5천장에 달하는 샘플 품목 이미지가 주어진다. 대회종료 2주전까지 주어진 test A training set으로 제출 가능하며, 대회종료(6/16) 2주전인 6월 5일 최초공개된 test B 서브셋으로 트레이닝된 모델을 검증하여 최종 스코어보드에 제출하면된다. Multimodalitie recall의 추천정확도에 검증방법은 nDCG@5(n = 5)의 검증방식을 사용하여 최종 고득점자에게 수상을 부여한다. 질의어(쿼리)에 가장 높은 정확도를 보인 5개의 이미지를 추천하는 알고리즘을 구현하여 Cross-modal Retrieval Accuracy를 향상하는 Feature Engineering에 초점을 맞추고 있다



Fig 1 Multimodal retrieval image sample

이커머스 플랫폼에서 고객의 첫번째 행동은 원하는 품목에 대한 검색일 것이다. 'leopard print woman shoes'를 검색한 고객에게 가장 적합한 이미지를 제공하는 추천 알고리즘이 필요하다. 만일 빨간 구두 이미지가 상단에 뜬다면 고객은 첫 세션에서 바로 이탈하거나, 2-3번의 세션이동을 통한 추가 탐색이후 플랫폼을 이탈하게 될 것이다. 단일세션 이탈은 고객의 구매와 직결되는 행동 패턴으로서 현대의 이커머스 플랫폼에서 가장 민감하게 받아들일 수밖에 없는 Key-Factor이다.

2. THE DATASET

2-1. Consist of dataset

Multimodal_train_sampleset.zip (82.53MB)

Multimodal_labels.txt (8056MB)

Multimodal_train.zip(26.19GB) – unzipped from that file, train.tsv can be accessed for training records. Before released the official whole set of training data, sampled set containing 10k pair of data. Each row in .tsv files represents a single sample pair. It has 9 columns separated by tabs

2-2. Dataset files

Multimodal_validpics.zip(2.41GB)

This file has the ground-truth product images can access the original image dataset.

Multimodal_valid.zip(137.2MB)

Multimodal_testA.zip(269.38MB) – testB set will be uploaded before 2weeks when competition end comes up. (finalized performance test will be testB set)

2-3. Multimodal_submit_example_testA.zip(25.73KB)

2-4. Multimodal_validpics.zip(2.41GB)

This file has the ground-truth product images can access the original image dataset.

Multimodal_valid.zip(137.2MB)

Multimodal_testA.zip(269.38MB) – testB set will be uploaded before 2weeks when competition end comes up. (finalized performance test will be testB set)

Multimodal_submit_example_testA.zip(25.73KB)

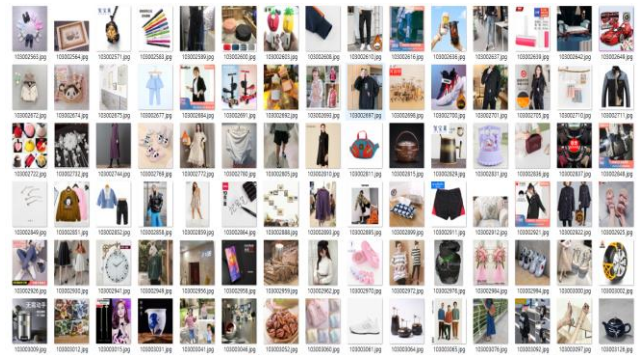


Fig 2 Top 30 ranked product image sample

3. RELATED WORK

해외의 멀티모달리티의 연구는 최근 ImageBERT 위주로 논의가 활발해지고 있다. Di Qi et al.(2020)의 ImageBERT 분류기를 활용한 cross modal pre-training의 연구,

Tadas Baltrušaitis et al.(2017)의 Taxonomy classification등이 관련분야의 파생연구를 촉발하고 있는 상황이다. 또한 도메인 knowledge도 유통, 금융, e-commerce, 제조등의 다양한 분야에서 주로 B2C소비 분석을 위한 용도로 폭넓게 연구되고 있다. 반면 국내의 경우는 대표적인 연구로 임정선, 한미경, 윤현진(2018), 멀티모달 딥러닝 모델을 이용한 실감효과 구간 검출과 최승진 외(2017), 멀티모달 데이터 분석을 위한 기계학습 기법 연구를 가장 최근의 연구된 내용으로 찾아볼 수 있었다. 상대적으로 머신러닝의 테크니컬한 멀티모달 분석에

초점이 맞추어져 Domain application과 관련분야 응용에 대한 아쉬움이 크다. 국내의 멀티모달리티와 크로스 모달리티에 대한 도메인 확장과 다변화가 촉발되길 바란다.

그런점에서 중국 최대의 전자상거래 플랫폼인 Tianchi(알리바바 계열사)의 Modern e-commerce platform: Multimodalities recall은 보다 정교한 모델 빌드를 통해서 소비자들의 만족도를 높일 수 있는 검색결과를 제공하는것이 요지이다. 기존 국내외의 연구 사례와 ImageBERT분석 위주로 연구를 참고하여 본 연구를 진행하였다.

4. Proposed Method / Access

4-1. Model Overview

Cross modality search는 일반적으로 두 가지 방법을 사용한다.

- ① 각기 다른 modality data를 동일한 하이퍼 피쳐 공간에 매핑(mapping)→ 다른 modality data 유사성(해석가능 거리: uclidian distance) 계산
- ② 다른 modality data를 각각 다른 피쳐 공간에 매핑한 다음 latent layer통해서 각 피쳐를 학습해서 설명할 수 없는 거리 함수를 train 하는 방법

방법1)의 장점은 음수샘플(네가티브 샘플) 구성할 필요가 없고 해석가능거리에 대한 계산량(computational complex)이 작다는 점이 있다. 온라인 모드(on-line mode)에 사용 적합한 방향이다.

방법 2)는 분석 모델은 functional 상호 작용효과가 있지만 네가티브 샘플을 별도 구성해야하는 번거로움이 존재한다. 그러나, 이번 이미지검색 최적화 목표에는 ‘인터랙티브 커넥션’의 초점을 맞추었으므로 네가티브 샘플을 구성하는 작업을 하더라도 두번째 기능적 방법을 사용하는 것이 옳은 방향일 것이다. 본 연구의 접근은 방법2)로 진행하게 된다.

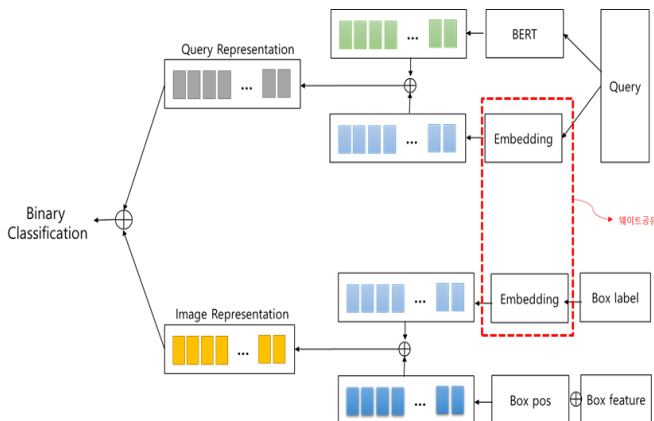


Fig 3 Model Over View

인풋 데이터는 자연 질의어(query)와 이미지 feature의 임베딩 웨이트를 공유하여 동일한 값의 최적값을 추적한다. 이 연구의 최종목표는 Binary Classification과 밀접하며 각각 Image chunk(약 3만개)의 대표 샘플과 Query chunk (약 3만개)대표 샘플의 매핑으로 Optimal Decision Boundary를 결정한다.

4-2. EDA & Preprocessing

위 세트는 랭킹 제품 30개의 후보 풀을 제공한다.

이 데이터 집합의 각 컬럼은 후보제품 - 쿼리의 training / validation 세트를 한 스키마로 제공한다. 피쳐설정이나 테스트셋 쿼리는 개별적 응용이 가능하지만 product ID는 valid set이나 test set에서 모두 동일하게 써야하는 것이 주 포인트이다.

‘Valid_answer.json’ 파일의 콘텐츠는 실제 제품 기반으로 소비자가 질의하는 자연어 쿼리를 반영한다. (정렬순서와 상관없음)

사전 데이터 작업으로는, Train 데이터의 chunksize - 100,000으로 묶어준 뒤 공간 효율성을 위해 일부 필드에는 base64 전환 인코딩을 한다.

데이터 프레임으로 불러온 확장자 ‘.tsv’ 데이터셋의 swifter 패키지를 통한 수동마스크를 생성한다. Train 데이터셋 컬럼의 query는 실제 검색 텍스트를 의미하며 해당 품목의 이미지 너비, 높이등의 공간정보를 보유하고 있다. 품목 이미지 박스의 위치와 feature에 대해 2048픽셀의 이미지 매핑을 통해 각 이미지박스 범주를 나타내는 레이블 컬럼또한 존재한다.

4-3. Feature Selection for Input Data

이 문제는 압축해제시 약 300G인 전체 3만개의 훈련데이터를 제공하며, 모든 데이터를 읽을수 없으므로 baseline method에 따른 지정된 청크 크기를 사용하게 된다.

Column name	contents
product_id	the index of the product
image_h	the height of the product image
image_w	the width of the product image
num_boxes	the number of detected object bounding boxes for the image
Boxes	a [num_boxes, 4] shaped 2-D array specifying the location of each object bounding box in the image. For each row of this array, the 4 values specify the top/left/bottom/right coordinates of the corresponding box. This array is encoded into a base64 string , you can recover it into a numpy-array using <code>np.frombuffer(base64.b64decode(THIS_ENCODED_BASE64STRING), dtype=np.float32).reshape(NUM_BOXES, 4)</code> in python
Features	a [num_boxes, 2048] shaped 2-D array specifying the 2048-dimension feature computed by the detector of each object bounding box in the image. This array is also encoded into a base64 string, you can recover it into a numpy-array using <code>np.frombuffer(base64.b64decode(THIS_ENCODED_BASE64STRING), dtype=np.float32).reshape(NUM_BOXES, 2048)</code> in python
class_labels	a [num_boxes] shaped 1-D array specifying the category-id of each object. There are 33 object classification categories in this dataset. The correspondence between each category-id and its category name is given in multimodal_labels.txt. The array is also encoded and can be recovered by <code>np.frombuffer(base64.b64decode(THIS_ENCODED_BASE64STRING), dtype=np.int64).reshape(NUM_BOXES)</code>
query	a natural language query which matches the corresponding product
query_id	the index of the query

Table 1 제공 컬럼의 Training/Valid dataset 설명

제공되는 데이터의 포맷은 Training set: 300만 쌍의 검색정보 샘플과 모델 트레이닝용 제품 이미지 포함 객체감지를 위한 위치 및 분류 레이블 등이 포함되어있는 데이터이다. 확장자는 'tsv' 파일이며 총 9개컬럼은 표와 같은 정보를 갖고있다. 이중 인풋데이터로 4개의 feature를 선정한다.

- Product_id: product index
- Image_h: height
- Image_w: image weight
- Num_boxes: 상자수의 개체 경계 상자이미지 감지

$$\text{vector } \mathbf{c}^{(i)} = \left(\frac{x_{tl}}{W}, \frac{y_{tl}}{H}, \frac{x_{br}}{W}, \frac{y_{br}}{H}, \frac{(x_{br}-x_{tl})(y_{br}-y_{tl})}{WH} \right)$$

Formula 1 Image Vector in ImageBERT

Num_boxes의 경계상자 이미지감지에 대한 벡터는 위의 vector $\mathbf{c}^{(i)}$ 의 수식을 참고한다.⁴ 전반적인 진행 과정은 다음과 같다.

- ① 텍스트 부분은 쿼리 표현을 형성하기 위해 bert 및 사용자 정의 임베딩을 통한 쿼리 스티칭작업을 시행한다.
- ② 이미지 부분에서는 먼저 여러 상자의 레이블을 단 어 표현으로 변환한다. 이러한 방식으로 레이블 시

퀀스가 구성되고 레이블 매개 변수의 표현 은 공유 매개 변수를 포함하여 형성된다.

- ③ 이미지 기능의 경우 여러 상자를 시퀀스로 취급하여 NLP와 비슷한 통합 방식으로 처리할 수 있다. 상자 기능 시퀀스와 상자 위치 시퀀스를 동일한 위 도에 매핑하고 추가하여 상자 시퀀스(image box pixel)을 추출할 수 있다. 이미지 박스는 제품의 랭 크와 연결되고 쿼리 라벨링으로 replace된다. 쿼리 표현은 product-image 매칭 및 image box pixel정보 와 상호작용한다.

4-4. Image embedding and Box Featurizing

json파일에 product image의 좌상, 좌하, 우상, 우하의 4개 공간에 대한 feature정보가 제공되어 product image의 input dataset으로 투입된다.

4-5. Chunk Dataset to Extract Representative Data

베이스라인 모델의 쿼리를 BERT 입력 후 BERT의 첫 출력은 ['CLS']컬럼의 아웃풋 값으로 생성된다. BERT는 성능적 제한을 고려하여 Early training 된 데이터셋을 사용한다. 이를 통해 Box Label Feature, Box Feature, Query의 대표 데이터셋을 추출 한다.

4-6. Binary Classification

최종 binary classification 모델에 각 피처를 투입하여 trainX, trainY과 testX, testY값을 추출한다. (상세 패키지와 앙상블모델 은 코드내용을 참조한다)

4-7. Expected Results

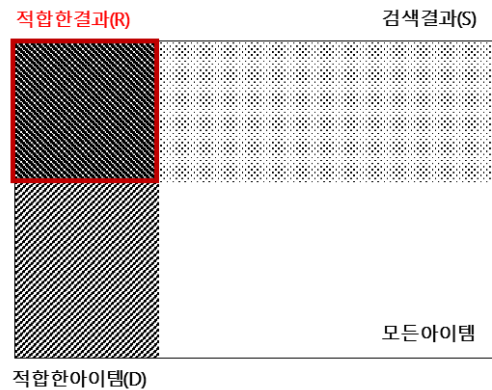


Fig 4 적합한 아이템, 검색결과 적합한 결과 사이의 관계⁵

⁴ Di Qi et al.(2020),ImageBERT: Cross-Modal Pre-Training with Large-Scale week-supervised Image-Test Data

⁵ 일리야캐서브, 알고리즘 마케팅(2017) 4.2.1. 적합성 지표,p.212

질의 쿼리에 대한 검색요청의 결과 기댓값은 다음의 세 가지로 정의할 수 있다. D는 시스템에서 인덱스처리된 특정 검색의도에 대한 적합한 아이템의 숫자, S는 검색 결과에 등장하는 아이템의 숫자, R은 검색 결과 중 적합한 아이템의 숫자이다. 이 세가지관계는 Fig4로 나타낼 수 있다. 즉 정밀도(precision)와 우리가 이 대회에서 궁극적으로 구하려고하는 재현율(recall)에 대한 최적값을 산출 해 낼 수 있게된다.

5. Experimental Results

5-1. Experimental Setup

5-1-1. Data

검증 세트에는 약 500 개의 쿼리와 총 14,720 개의 데이터가 있으며 valid_answer는 후보에게 각 검증 세트 쿼리에 대해 5 개의 제품 ID를 제공한다. 테스트 세트에는 약 1,000 개의 쿼리와 총 17428 개의 데이터가 있으므로 제출할 각 쿼리에 대해 후보 5 개의 제품 ID를 계산해야한다.

1	data['words_len'].describe()	1	data['num_boxes'].describe()
count	3.000000e+06	count	3.000000e+06
mean	3.864156e+00	mean	3.770936e+00
std	1.280866e+00	std	3.316120e+00
min	1.000000e+00	min	1.000000e+00
25%	3.000000e+00	25%	2.000000e+00
50%	4.000000e+00	50%	3.000000e+00
75%	5.000000e+00	75%	5.000000e+00
max	2.100000e+01	max	9.100000e+01
Name:	words_len, dtype: float64	Name:	num_boxes, dtype: float64

Fig 5 Dataset Describe

대부분의 쿼리는 약 5이고 가장 긴 쿼리는 21이다. 상자의 수는 대부분 약 5의 값에 분포되어 있다.

앞서 언급한 바와 같이 학습 세트의 쿼리 길이와 상자 수를 계산한다. 4.Proposed Method에서 고민한 2가지 방법 중 각 이미지셋을 각기 다른 공간벡터에 맵핑한 뒤 latent layer를 각각 학습시켜 개별 feature간의 거리를 구하기 위해서는 부정데이터셋에 대한 추가생성이 필요하다.

- 네거티브 샘플 구성: 각 배치에서 동일한 양의 네거티브 샘플을 무작위로 샘플링

5-1-2. Models

Uclidian distance를 구하는 idx 변수를 사용해 seq_gather(x)함수와, 시퀀스와 벡터간의 거리를 구하는 seq_and_vec(x)함수, maxpooling을 구하는 seq_maxpool(x)함수 총 3개의 함수로 구현한다.

매개변수 x는 시퀀스(seq)와 거리(idx) 로 구성되어졌다.

5-2. Performance Comparison

5-2-2. Processing Time

- 1 model training set learning 19hr / 1 model valid set 4.5hr (Nvidia K80 GPU기준)

5-3. Evaluation and Result

5-3-1. MAP (Mean average Precision)

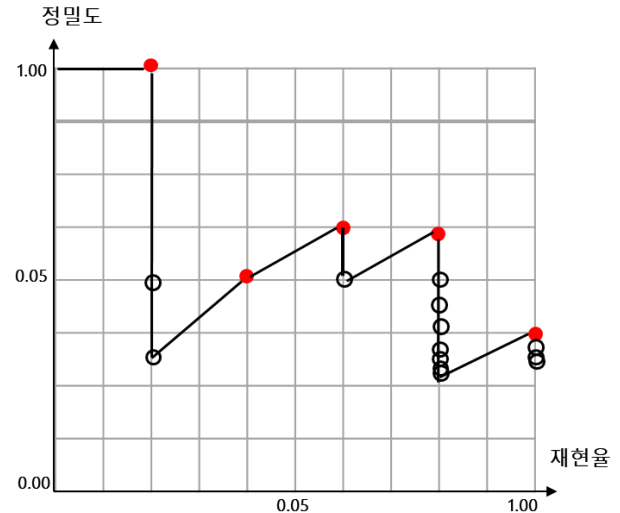


Fig 6 KDD cup데이터셋의 정밀도-재현율 곡선

일반적인 최적 추천검색 결과값의 재현율 검증방법은 평균정밀도(MAP)를 이용한다. 연구 데이터셋의 정밀도- 재현율곡선은 Fig6의 결과값과 같다. 검색결과에서 빨간색 원은 적합한 아이템에 해당하고, 나머지 원은 적합하지 않은 아이템에 해당한다. 검색서비스의 전체성능을 한눈에 가늠할 수 있는 지표이다. 그러나 본 대회에서는 할인누적이득 지표에대한 Evaluation을 제안하므로 다음으로 넘어가도록 한다.

5-3-2. nDCG@5

만약 추천시스템을 평가할 때 한 사용자의 추천목록과 해당 결과만 사용하여 평가할 수 없음은 당연하다. 그렇다고 전체 테스트셋에서 한 사용자만의 해당 추천목록 결과를 평가하는 것은 불가능하다.

즉, 다른 사용자의 추천목록 평가점수를 반영하기 위한 정규화 작업이 필요하다. 그래서 NDGC(Normalized Discounted Cumulative Gain)을 사용한다.

$$NDCG_u@k = \frac{DCG_u@k}{IDCG_u}$$

Formula 2 NDCGu@k

여기서 IDCG = Ideal DCG 즉, 사용자의 추천시스템에 반환된 최상의 권장결과 목록을 반환을 의미한다.

(리턴 결과가 관련성에 따라 정렬되었다 가정하면 가장 관련성이 높은 결과가 이전순서의 IDCG이다.) 따라서 DCG값은 (0, IDCG)사이값 이므로 NDCG(0,1)사이의 값이고, 사용자의 NDCG@k (k=5)는 아래의 식으로 정의된다.

$$NDCG@k = \frac{\sum_{u \in \mathcal{U}^{te}} NDCG_u@k}{|\mathcal{U}^{te}|}$$

Formula 3 NDCG@k

본 대회에서는 누적할인이득을 통한 최종 추천 상위 5개 키워드 기반의 정규화된 할인 누적이득 값을 scoring한다.

nDCG(Discount Cumulative Gain)은 CG(Cumulative Gain) → DCG(Discount Cumulative Gain)을 통해 발전한 등식이다. 누적이득(Cumulative Gain)은 일반적인 추천시스템에서는 누적이득에 해당하는 결과값의 추천결과에 대한 관련성 점수를 총 추천의 누적값으로 계산한다. 여기서, rel-i는 위치 i에서의 추천 결과의 관련성을 나타내고, k는 검사할 추천리스트의 크기를 나타낸다.

그러나 CG의 단점은 추천결과가 전체 추천효과에 대해 다른 위치에 미치는 영향을 고려하지 않는다는 것이다. 상관관계가 낮은 결과가 최상위 위치에 있으면 사용자에게 심각한 영향을 미치므로 DCG의 공식을 통한 검증방법을 고려해본다. DCG(Discounted Cumulative Gain)으로 기준 위치영향의 요소가 나타난다. 추천결과에 추천효과에 대한 할인처리(Discount) 참조하여 순위를 조정한다.

$$DCG_k = \sum_{i=1}^k \frac{2^{rel_i} - 1}{\log_2(i + 1)}$$

Formula 4 DCG@k

위 CG와 DCG공식을 통해 두 가지 결론에 도달하게 된다.

- 추천결과와 상관관계가 클수록 DCG가 커진다.
- 상관관계가 양호하고 추천리스트가 앞에있으면, 추천효과가 좋을수록 DCG가 커진다.

그러나 DCG도 여전히 한계가있는 다른 권장 사항의 목록간에 평가수행이 어렵다는 단점이 있다. 만약 추천시스템을 평가할 경우, 한 사용자의 추천목록과 해당결과만 사용하여 평가할 순 없을 것이다. 전체 테스트셋에서 사용자의 해당 추천 목록 결과를 평가하는 것은 불가능하다.

즉, 다른 사용자의 추천목록 평가점수를 반영하기위한 정규화 작업이 필요하게되고, 최종적으로 NDGC(Normalized Discounted Cumulative Gain)을 사용하게 된다.

5-4. Limitations

① Natural Language Processing에대한 BERT raw data를 full로 돌릴 수 있는 GPU 성능한계 개선 (KDD컵이 대기업 연구소를 대상으로 하는 컴페티션을 자주 개최하는 만큼 개인들의 GPU성능에 대한 기대치가 다소 상향 조정되어있는 편이다.)

② 방대한 대용량 데이터셋 (최소 200G)에 대한 인프라적 서포트 개선 필요

→ 훈련 세트 압축 패키지 100g, 압축 해제 및 디코딩 200g +적어도 256g 메모리가있는 서버

③ BERT Text의 크기가 대용량이상 수준이어서 청크로 쪼갬된 대표 데이터만 사용함. (쿼리 데이터는 전체 훈련 데이터의 1/4 만 사용됨)

6. Discussion

6-1. Learning Multimodal Knowledge

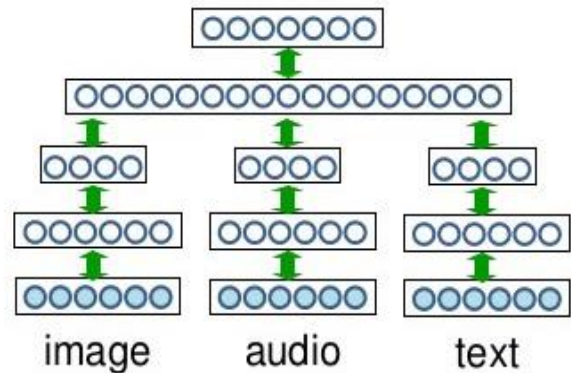


Fig 7 Multimodality Concept Model

{Query : Image}의 멀티모달리티셋에 대한 연구와 관심이 지속적으로 증가하고있다. 이것은 현대 이커머스 플랫폼의 성장과 더불어 필연적인 수요이며, 이에 따라 text BERT 알고리즘에 대한 Decision Boundary 명확화가 필요하다.

6-2. Model Performance

우리는 위의 연구를 통해서 사용된 DNN알고리즘을 textCNN algorithm으로 성능 개선해볼 필요의 가능성을 열어두었다.

이미지 박스의 Feature Pixel에서 Bs, Len, Dim의 마지막 1 차원 변형으로 시퀀스를 생성하면 2048 x num_boxes의 데이터를 지정된 길이로 통합하는 것을 의미함을 활용해서 textCNN모델로서 발전가능성을 가늠해 볼 수 있다.

사실, 여기서는 쿼리의 분석은 BERT의 형태를 모방하기위한 것이지만 Decision Boundary에 대한 좀 더 명확한 기준선이 필요한 것을 발견했다.

또한 4-1. Model Overview에서 시퀀스 표현을 구성할 때 BiLSTM, Conv1D 등과 같은 다양한 옵션을 상기 베이스 라인 더 접합해볼 여지도 열어두고 있다. 마찬가지로 4-1. Model Overview에서 두 가지 데이터 청크의 결합 (쿼리 스티칭, 이미지 합산)이 parallel프로세싱이 아니라 시작부터 같이 상호 작용하는 모델을 생각할 수도 있다.

7. Conclusion

최종 training set을 통해 valid set으로 nDCG@5의 결과값은 다음과 같다.

- {Query : image} baseline recall : 0.5538
- Full build mode recall : 0.599

KDD Competition의 현시점(6/11) 스코어보드 1위: recall accuracy 0.77 (▲0.17)로 여전히 성능개선의 여지는 남아있다. 이렇듯 E-Commerce Platform의 Retrieval optimize Search 영역은 고객의 구매 행동 예측과 밀접한 연관되어 있어 향후 여러 알고리즘의 앙상블을 통한 개선의 여지가 무궁무진하다는 결론을 내릴 수 있다.

후 기

성능적인 한계로 fully data coverage access에 대한 아쉬움이 있지만, 이렇게 큰 사이즈의 데이터를 직접 분석해보고 모델을 설계해보는 경험은 많은 학습동기를 주었습니다. 데이터 사이언스응용과목을 통해서 머신러닝/딥러닝의 도메인응용 방법에 대한 현실적인 경험을 할 수 있는 기회였습니다.

참 고 문 헌

- [1] Arnabdi Viera, Inesting (2018). Optimization of Digital Marketing Campaigns. Figure6. Correlation with QGraph
- [2] Di Qi et al.(2020), ImageBERT: Cross-Modal Pre-Training with Large-Scale week-supervised Image-Test Data
- [3] Hieu Pham(2018). Implications of Multimodal Deep Learning for Textual and Visual Data
- [4] KDD cup 2020 Challenges for Mordern E-Commerce Platform: Multimodalities Recall
(<https://tianchi.aliyun.com/competition/entrance/231786/information>)
- [5] Louis-Philippe Morency et al.(2018)Tutorial on Multimodal Machine Learning, CMU MultiComp Lab Tutorials
- [6] Nicolas Audebert et al.(2019). Multi deep networks for text and image-based document classification.
- [7] MARCUS THORSTRÖM(2018). Applying machine learning to key performance indicators
(<https://odr.chalmers.se/bitstream/20.500.12380/250254/1/250254.pdf>)
- [8] Oriol Vinyals et al.(2015),Show and Tell: A Neural Image Caption Generator
- [9] Tadas Baltrusaitis et al.(2017), Multimodal Machine Learning: A survey and Taxonomy
- [10] Valentin radu et al.(2017) Multimodal Deep Learning for Activity and Context Recognition
- [11] Xiaojing Wang, Tianqi Liu, Jingang Miao(2019). A deep probabilistic model for customer lifetime value prediction.
(<https://arxiv.org/pdf/1912.07753.pdf>)
- [12] 임정선, 한미경, 윤현진(2018), 멀티모달 딥러닝 모델을 이용한 실감효과 구간 검출
- [13] 최승진 외(2017), 멀티모달 데이터 분석을 위한 기계학습 기법 연구