20190741 김지수

소비자 (구매) 행동 예측을 위한 Multimodal 모델

: 문화적 특성에 따른 소비자 행동의 차이를 고려한 구매 예측 모델

전통적인 소비자 (구매) 행동 예측에 관한 연구

1) 소비자행동론

: 인간의 경제적 소비행동(상품 및 서비스의 구매, 사용 및 폐기)을 해석하기 위하여 실험적으로 얻은 행동 원칙의 사용을 연구하는 학문으로 경제 심리학과 마케팅 과학의 교차점에 있음

관련 과목 수강: 마케팅최신주제및사례연구 <23-Spring> 하영원 교수님

- Judgment Under Uncertainty: Heuristics and Biases. Tversky & Kahneman, D. (1974) Science
- Choices, Values, and Frames, Kahneman, D. & Tversky, A. (ed.) 2000 외 논문 여러편 (16편 이상)
- 키워드: Biases in Judgment, Choice Heuristics, Framing Effect, Mental Accounting, Transaction Decoupling, Confirmation Bias, Envy, Regret & Self-Control, Diversification Bias, Time Perception 등
- 주로 소비자 구매 및 의사결정 행동에 대하여 Case 실험을 통해 행동양상을 설명하는 연구가 해당 분야에서 진행됨.

오프라인에서 온라인 소비로 이동

<국내 온라인 커머스 동향>

- 가전(53.7%), 가구(49.7%), 서적/문구(49.3%), 신발/가방(40%), 화장품(37.3%) 등의 분야에서 거래액 비중으로 소매 판매액 대비 온라인 판매가 상당부분을 차지 (2024년, 통계청)
- 유통업체를 기준으로는 전체 매출 대비 온라인 비중이 55.6%(25년 2월, 산업통상자원부)로 오프라인보다 높음

(출처: 연합뉴스, https://www.yna.co.kr/view/AKR20250123032400003?)

- 온라인 구매 데이터를 사용한 직접적인 분석이 활발하게 수행되고 있음

소비자 문화적 특성에 따른 영향의 가능성

- 지속가능한 패션 소비에 대한 국가별 차이 분석 Jung, H.J.; Oh, K.W.; Kim, H.M
- 국가 문화가 소비자 의사결정 스타일에 미치는 영향 Chan Yie Leng, Delane Botelho (2020)
- Amazon 거래 데이터와 같은 데이터셋을 활용한 모델에 대비하여 자체 데이터(소비자의 국적 등을 포함)를 활용하여 튜닝한 모델을 사용하였을 때의 성능 차이가 얼마나 유의미하게 영향을 줄 수 있는지 궁금증을 가짐.

-> 산학 협력을 통한 연구 혹은 더 나아가면 디지털 트윈 및 시뮬레이션과도 관련

온라인 환경에서의 소비자 구매 행동에 관한 연구 분석

최근의 인공지능 기반 연구

- Meta AI의 Commerce Multimodal Shopping model (이미지와 텍스트 기반, 2022)
- 네이버 쇼핑 e-CLIP 모델(e-CLIP: Large-Scale Vision-Language Representation Learning in E-commerce) Multimodal pre-trained 모델을 네이버 쇼핑 데이터베이스에서 2억 7천만개의 이미지-텍스트 쌍을 활용

☞ 상품 클러스터링 (유사 상품 묶는 과제), 상품 매칭, 상품 속성 추출, 카테고리 분류, 성인 상품 인식 등의 과제에 활용하고 있음
소비자 구매 촉진 측면보다는 상품 관리의 측면에서의 활용

- A Multimodal In-Context Tuning Approach for E-commerce Product Description Generation(2024), Yunxin Li 외
- Jeong Euiju, Xinzhe Li, Angela (Eunyoung) Kwon, Seonu Park, Qinglong Li, and Jaekyeong Kim. 2024. "A Multimodal Recommender System Using Deep Learning Techniques Combining Review Texts and Images"
- 이커머스 고객 행동 이벤트 시퀀스 패턴 분석 기반 구매 전환 예측 모델링 방안에 관한 연구 (이유림, 유헌창) 2023.
- 인공지능 분류모델을 활용한 농축수산물 전문 쇼핑 모바일 앱의 구매고객 예측: 고객 행동 데이터를 기반으로 (신우석 외, 2024)

: 온라인 커머스의 상품 이미지, 상품 설명(LLM 학습에 사용), 상품 이름 번역, 상품 추천, 구매 예측, 소비자 만족도 조사, 상품 분류 등 폭 넓은 관련 분야에서 다방면으로 연구가 활발하게 이루어지고 있다. 추천시스템은 주로 상품 이미지와 리뷰 텍스트를 기반으로 하고, 예측의 경우 로그 데이터(행동 데이터)를 기반으로 하고 있다.

온라인 환경에서의 소비자 구매 행동에 관한 연구 분석

간학문적인 연구

온라인 커머스의 상품 이미지, 상품 설명(LLM 학습에 사용), 상품 이름 번역, 상품 추천, 구매 예측, 소비자 만족도 조사, 상품 분류 등 폭 넓은 관련 분야에서 다방면으로 연구가 활발하게 이루어지고 있다.
한편, 조사 대상이 되는 소비자의 군집특성이 (구매) 행동에는 영향을 미친다는 연구결과(가격 민감도, 리스크 회피성향, No.1 선호현상, 집단주의/개인주의 등) 는 많이 존재하나, 이러한 이유에서 발생할 수 있는 데이터셋의 차이에 따른 모델 성능은 직접적으로 연구되지 않았음.
- 5개국의 스마트폰 센서 데이터를 활용한 성격 특성을 예측하는 연구, Hofstede의 문화 차원 이론을 기반으로 다양한 국가의 군중 행동을 예측하는 모델(CC-ANN) 등 다른 분야에서는 관련 연구가 존재

구체적인 예시로, 아마존에서 개발한 데이터셋을 사용한 모델의 사용으로 선택된 적합한 추천 이미지, 추천 제품 등과 국내 소비자 데이터셋(산학 협력 필요)을 활용하였을 때의 결과가 실제로 얼마나 유의미한 차이가 있는지 연구
(아마존과 메타 데이터셋에서 origin_country와 같이 제품의 원산지는 포함하나 예측대상의 직접적인 국적은 개인정보 이슈로 위치 데이터와 같이 간접적인 데이터가 사용되고 있다. 추가 해당 데이터셋은 영국, 동일, 일본, 프랑스, 이탈리아, 스페인 6개국의 데이터만 포함)

Multimodal 소비자 구매 행동(의사결정) 예측 문제

1) 데이터

- 텍스트: 상품 리뷰, 검색어, 상품 설명
- 이미지: 제품 이미지, 유저 리뷰 이미지
- 행동 데이터: 체류 시간, 구매 여부, 클릭 로그
- 소비자 데이터: 국적, 나이, 성별 등
- 그 외: 가격, 브랜드, 할인율 등

2) 모델

- Cross-modal transformer (MMBT 등)
- Attention weight 분석/Contribution 분석

3) 분석 지표

- AUC, nDCG, Precision 등
- Cutural-fit score(?)

소비자 구매 행동(의사결정) 예측 문제

Research Question 1) 한국 온라인 쇼핑 환경에서 Multimodal 데이터를 활용한 모델이 기존 각각의 모델과 비교하여 구매 행동을 얼마나 정확하게 예측할 수 있는가?

Research Question 2) Multimodal 모델(Transformer, Clip, ViLT, Meta AI commerce mm등)의 소비자 구매 행동은 어떤 데이터 요인에 의해 더 설명되는가? (XAI 기법을 사후 적용/Modal 별로 별도의 모델을 만들어 기여도 비교 등)

Research Question 3) 소비자의 사회문화적 특성(Global vs Local model)이 모델 성능에 유의미한 차이를 보이는가?

Research Question 4) 어떤 Feature (할인율, 카테고리, 무료배송 여부, 썸네일 이미지, 리뷰 내용 등)가 문화에 따라서 더 민감하게 반응하는지, 실제로 차이가 있는가? 특성을 기존의 연구와 매핑이 가능한지?

Research Question 5) 국가 간 데이터 부족 문제가 있는데, 모델을 확장할 때 전체 모델에 대비하여 특정 국적을 추가하였을 때 성능 향상의 폭이 어느정도 발생하는지? (글로벌 모델은 지역 특화 학습 없이도 잘 작동할 수 있는가?)

Kaggle에서 Multimodal 예측 모델 사용 예시

Shopee - Price Match Guarantee
Determine if two products are the same by their images

대회 배경

- 온라인 쇼핑 시에 같은 제품이라도 가격이 다른 경우가 많음
- 소매업체는 경쟁사 제품과 정확히 일치하는 제품 매칭을 통해서 가격 경쟁력을 유지하려고 함
- 자동화하기 위해서 텍스트와 이미지 정보를 통합한 ML접근법이 필요

해결 과제

- 두 제품이 같은 제품인지 여부를 판별해야 한다.
- 같은 제품이어도 이미지, 제목, 설명이 다르게 작성될 수 있음

평가

- Mean F1 Score 각 샘플에 대해 예측한 그룹과 정답을 비교하여 계산

Competition Host
Shopee

Prizes & Awards
$30,000
Awards Points & Medals

Participation
16,388 Entrants
3,032 Participants
2,426 Teams
51,077 Submissions

Tags
Image
Text
Retail and Shopping
Custom Metric

Kaggle에서 Multimodal 예측 모델 사용 예시

From Embeddings to Matches

1. 모델 예측 성능 정리

- **Baseline(기본)**: 0.70 (이미지만 사용), 0.64 (텍스트만 사용)
- 이미지 임베딩 + 텍스트 임베딩 **concate 후 정규화**: 0.724
- **min2 전략 적용**: 0.743
- **각각을 정규화 후 concat** (순서 변경): 0.753
- **전체 데이터로 학습**: 0.757
- **이미지/텍스트 매칭 결과 합친후 임계값 튜닝**: 0.776
- **INB 기법 + 다양한 텍스트 모델 추가**: 0.784
- **INB 1단계에서 이미지/텍스트/조합 임베딩 모두 활용 + 임계값 공동 튜닝**: 0.793

Kaggle 사례

2. 최종 모델 설명

- 이미지 인코더와 텍스트 인코더를 각각 사용해서 이미지와 텍스트 데이터를 임베딩해 이를 활용하여 매칭을 수행함

이미지 인코더: timm 라이브러리의 eca_nfnetl1 모델 2개 사용

이미지 feature는 GAP(Global Average Pooling) 후

BatchNorm1D, Normalization 단계를 거쳐서 최종 1792 크기로 생성

텍스트 인코더는 Hugging Face 플랫폼에 있는 아래의 모델 사용

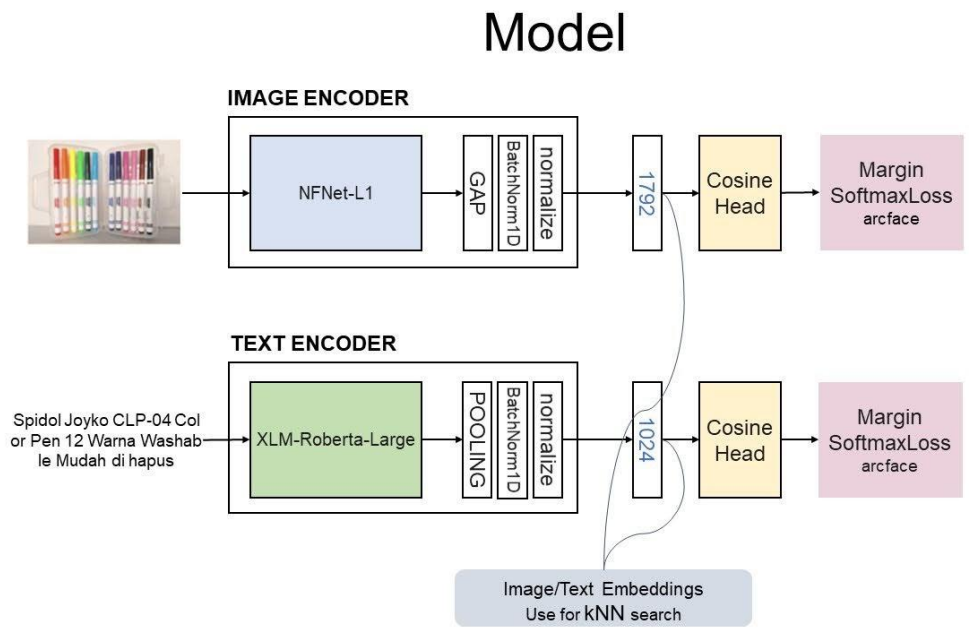
- xlm-roberta-large, xlm-Roberta-base
- cahya/bert-base-Indonesian-1.5G (인도네시아 언어)
- Indobenchmark/indobert-large-p1
- Bert-base-multilingual-uncased
- 텍스트 featur는 pooling, batchnorm1d, Normalization을 통해 1024 크기로 생성

Cosine similarity 계산 후 결과에 Arc Margine 적용 후

Softmax Loss를 사용하여 학습을 진행함

최종적으로 얻은 image/text embeddings가 kNN search에 활용되어

이미지-텍스트 간 유사도를 기반으로 매칭이 수행됨.



Kaggle 사례

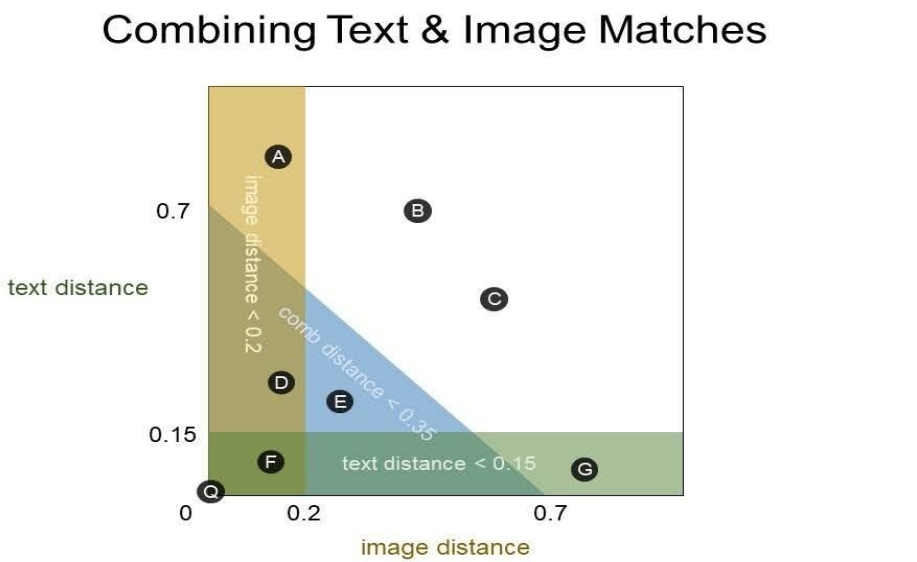
3. ArcFace Tuning (생략)

- 이전 과정에서 Arc Margin 적용할 때 임베딩에 충분히 마진을 주는 것이 중요하다는 내용 (논문을 기반으로)

4. 이미지 & 텍스트 매치 혼합 – Concatenation & Union

- 이미지와 텍스트 두 입력을 어떻게 잘 결합하는지가 성능에 매우 중요
- Image matches + Text matches + Comb maches -> Union 이 가장 좋다는 결론을 얻음

- 임베딩을 정규화한 이후 Concatenation하고 Comb Similarity를 계산 (이미지 유사도와 텍스트 유사도를 평균낸 것과 동일함)
- 이미지 기반으로 강하게 추천된 항목, 텍스트 기반으로 강하게 추천된 항목, 둘 다 중간 정도로 추천된 항목을 모두 받아들이는 전략
- 이미지와 텍스트 인코더를 Joint training도 시도해보았으나 성능은 개별 학습 후 합친 경우보다는 낮았음



for each query item Q, accept items in colored region → matches for Q: [Q, A, D, E, F, G]

Kaggle 사례

5. Iterative Neighborhood Blending (INB)

핵심 아이디어: 임베딩 간의 유사도를 기반으로 이웃을 정의하고, 해당 이웃 정보를 활용해서 임베딩을 반복적으로 개선하는 것 (더 정밀한 군집화와 매칭 향상)

INB 구성요소)

1. K-Nearest Neighbor Search (k=51)

- faiss(https://github.com/facebookresearch/faiss) 사용
- cosine similarity로 유사한 top-50 이웃 검색 (문제 제출 최대 50개)

2. Thresholding (기준치 정하기)

- cosine similarity → cosine distance(= 1 - similarity)로 변환
- 일정 threshold보다 가까운 이웃만 유지
- min2: 최소 2개의 매칭을 확보하도록 두 번째 이웃까지는 threshold를 완화

3. Neighborhood Blending (NB)

- 각 아이템의 이웃 임베딩을 **유사도 기반 가중치 합**하여 자신에게 더함
- 유사도가 edge weight인 그래프처럼 생각

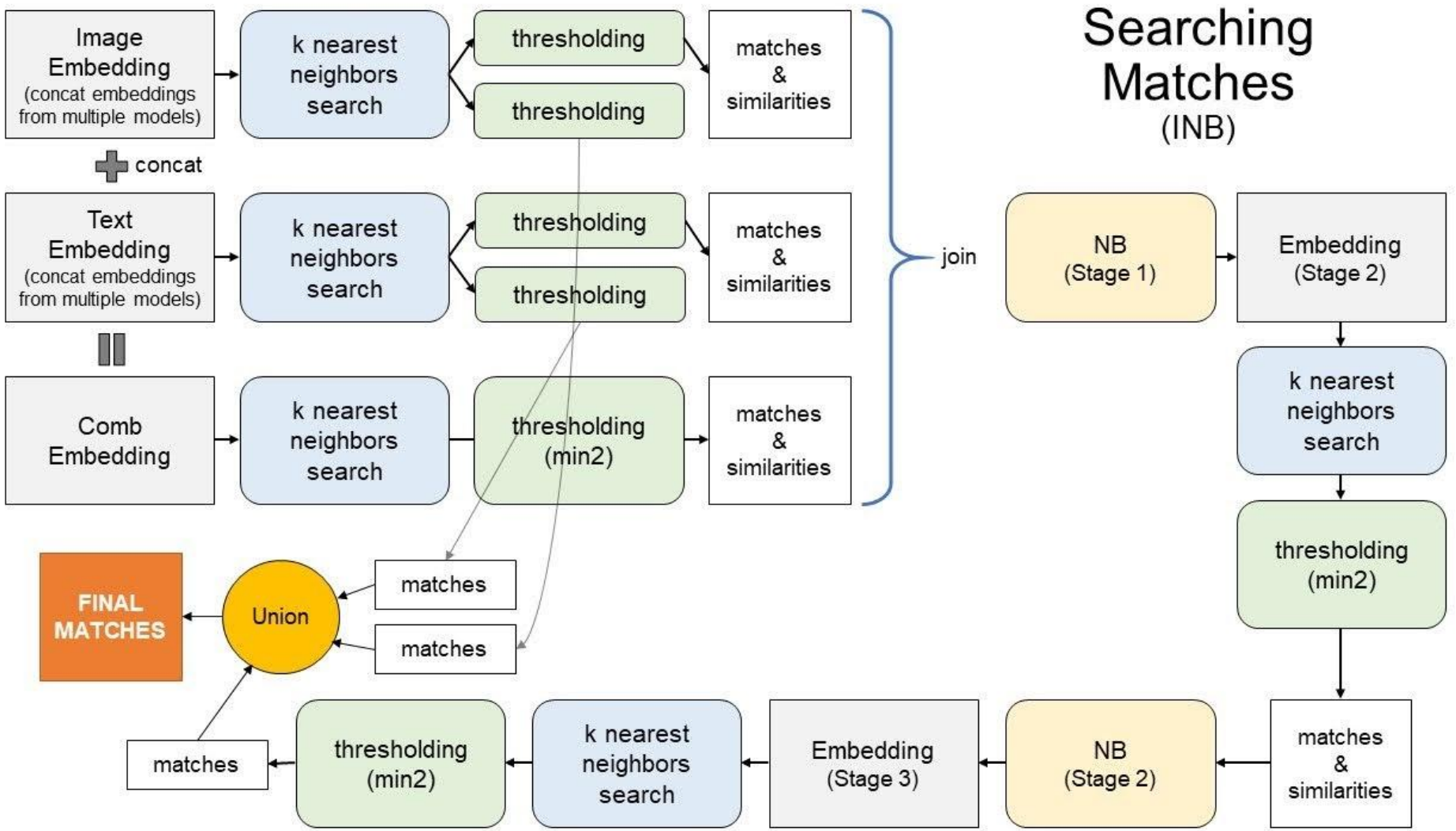
→ 결과적으로 **더 조밀한 임베딩 공간 형성**

4. Iterative NB

- 위 과정을 여러 번 반복 (stage1 → stage2 → stage3)
- 각 단계에서 refined embedding 생성 후 다시 kNN + NB
- 반복은 성능이 개선될 때까지만 진행함

Kaggle 사례

6. 전체적인 Iterative Neighborhood Blending (NIB)



7. Discussion

- **이렇게 Image, Text, Comb embedding 과 threshold를 jointly tuning한 모델의 성능이 0.793으로 가장 높았음.**
- product matching could support more accurate product categorization(분류) and uncover marketplace spam(이상한 제품) <- 이전에 Naver에서 Multimodal 활용한 것과 동일 Task
- **posting_id,matches**
- **test_123,test_123 <- 123번 id의 상품은 스스로만 동일**
- **Test_456,test_456 test_789 <- 456번 id의 상품은 789랑 동일 (최대 50개까지 매칭)**
- **0.780의 Private(비공개 테스트셋) F1 Score 달성 (평균적으로 78% 정확도 수준으로 매칭을 정확히 찾아냄)**
- Multimodal 모델에 대해서는 계속 연구가 이루어지고 있어서 추후 사용되는 분야가 넓혀지고 모델이 정교화된다면 더 좋은 성능을 기대해볼 수 있을 것
- 데이터는 의도하던 의도하지 않던 매우 방대하게 자동적으로 기록되고 있기 때문에, Multimodal model을 고려할 때 해결하고자 하는 문제에 어떤 데이터가 효과적일지 고민하는 것도 중요하다고 생각

감사합니다

2025. 6. 10. 화