

텍스트와 이미지 데이터를 활용한 관세 품목 분류 모델 분석

김시현¹, 이은지^{1,2}, 김선동², 차미영^{2,1}

¹한국과학기술원 전산학부 ²기초과학연구원 데이터 사이언스 그룹
{sihk, mk35471}@kaist.ac.kr, {sundong, mcha}@ibs.re.kr

Text and Image Data Analysis for Customs Item Classification

Sihyeon Kim¹, Eunji Lee^{1,2}, Sundong Kim², Meeyoung Cha^{2,1}

¹School of Computing, Korea Advance Institute of Science and Technology

²Data Science Group, Institute for Basic Science

요 약

글로벌 무역 거래가 증가하고 기술 개발로 거래되는 상품도 다양화·복잡화되는 상황에서 수많은 물품들을 분류할 수 있는 기준이 바로 국제통일상품명부호체계(이하 'HS')이다. HS란 물품을 분류하는 국제적인 기준으로 모든 교역물품은 국제 협약에 따라 5,387개(6자리 소호 기준)의 품목번호 중 하나로 '분류'된다. 본 연구에서는 상품에 대한 텍스트와 실제 모습을 담은 이미지를 활용한 품목 분류기를 소개한다. 결과를 바탕으로 텍스트와 이미지 데이터가 품목 분류에 어떤 영향을 미치는지 분석하고 향후 연구를 위한 방향성을 제시한다.

1. 서론

WCO(세계관세기구) 통계에 따르면 '20년 기준 전 세계의 수출입신고 건수는 5억여 건에 달한다. 특히 코로나19로 비대면 해외 직구가 활성화됨에 따라 같은 해 우리나라의 전자상거래물품 수출입 건수(목록통관 기준)는 6천3백만여 건에 이른다. 모든 수출입 물품은 품목 분류 결과(HS코드)에 따라 물품의 관세율이 정해지게 되며, 이는 물품 가격에 직접적인 영향을 미쳐 물품의 글로벌 경쟁력을 결정하기도 한다. 관세당국은 수출입 기업을 대상으로 정식 수출입신고 전에 품목 분류 심사를 관세청에 의뢰할 수 있도록 하는 품목 분류 사전심사 제도(관세법 제86조)를 운영하고 있으나, 교역물품의 종류가 다양해지고 복합적인 기능을 지닌 물품이 늘어남에 따라 의뢰 물품의 품목 분류 소요 기간은 '18년 이래 매년 증가('18년 20.4일 → '20년 25.9일)하고 있다. 이에 따라, 수출입 신고의 정확도를 높이고, 업무 효율성을 높이기 위한 품목 분류 알고리즘의 활용이 대두되고 있으며, 상품의 텍스트 [1] 및 이미지 정보 [2]를 활용한 품목 분류기의 개발이 진행되고 있다. 본 논문에서는 품목 분류 국내 사례 데이터를 기반으로, 상품의 텍스트와 이미지 정보를 함께 활용하였을 때, 품목 분류 알고리즘이 어떠한 성능을 보이는지 분석하고, 향후 연구를 위한 방향성을 제시한다.

2. 품목 분류 국내 사례 데이터

HS코드는 첫 두 자리는 류, 네 자리는 호, 여섯 자리는 소호라고 불리며 여섯 자리까지는 전 세계가 공통된 분류 체계를 따른다. 본 연구에서는 관세법령정보포털에서 제공하는 품목 분류 사례 데이터 중 84, 85, 90류에

해당하는 14,800 건의 데이터를 활용하였다. 84류는 기계류, 85류는 전자기기, 90류는 광학·정밀 의료기기를 포함하며, 항목 간의 경계가 모호하여 고도화된 품목 분류 알고리즘을 필요로 한다. 각 사례 데이터에는 물품 설명과 물품 이미지, 그리고 결정된 HS코드 등이 기록되어 있으며, 물품 설명은 “구리합금의 선을 직조하여 만든 클로스에 링 형태의 프레임에 결합한 여과망”과 같은 전문적인 단어들을 포함한다. 84, 85, 90류를 통틀어 호는 163개, 소호는 925개로 이루어져 있으며, 그림 1에서 볼 수 있듯이 각 호에 대한 데이터는 불균일한 분포를 따른다. 빈도 상위 10개의 호가 전체 데이터의 약 35%를 차지하며, 반대로 하위 10개의 호는 1~2개의 사례만 존재한다.

표 1. 품목 분류 국내 사례 데이터 통계

	호 개수	소호 개수	데이터 개수
84류	85	516	5,403
85류	46	265	6,834
90류	32	144	2,563

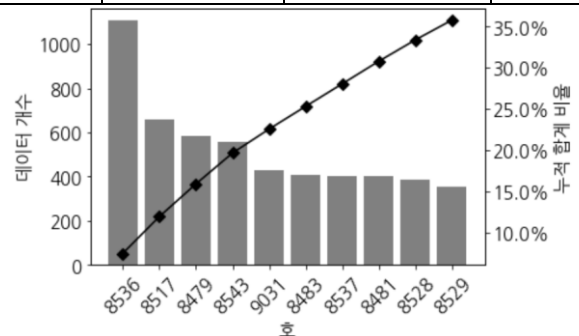


그림 1. 호에 따른 이미지 데이터 개수 분포

3. 모델 설명

품목 분류기는 물품 설명과 이미지를 입력받으며, 텍스트 처리에는 한국어 언어 모델 KoELECTRA를, 이미지 처리에는 Resnet50을 이용하였다. 물품 설명에 해당하는 텍스트와 이미지로 각각 파인튜닝된 모델을 함께 사용하여 최종 모델의 학습을 진행하였다.

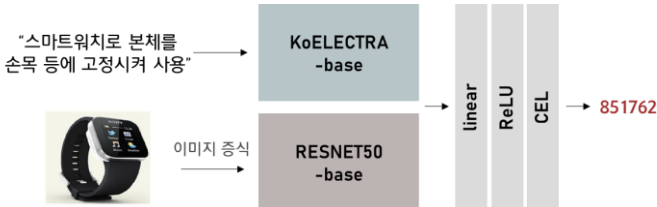


그림 2. 이미지와 텍스트를 입력 받는 모델의 개요

3.1. 텍스트 처리 모델

KoELECTRA를 파인튜닝하여 모델이 물품 설명을 입력받아 알맞은 HS코드로 분류할 수 있도록 학습했다. 각 품목에 대한 확률 p_{pred} 를 받고, 정답 HS코드와 크로스 엔트로피 손실을 계산하여 학습한다.

$$loss = CrossEntropyLoss(p_{pred}, y_{answer})$$

최종적인 HS코드로는 가장 높은 확률을 가지는 품목을 선택한다.

$$c_{pred} = argmax(p_{pred})$$

3.2. 이미지 처리 모델

이미지 증식을 위해서 랜덤하게 수평 뒤집기와 회전, 밝기 및 대조, 채도, 색조의 변화를 사용했으며, 마지막에 패딩을 더해 비율을 유지했다. Resnet50을 이용해 벡터를 생성하고 텍스트 처리 모델과 동일한 방법으로 학습하여 파인튜닝했다.

3.3. 분류 값 예측

물품 설명과 사진을 모두 이용해 분류를 진행한다. 텍스트 처리 모델에서 계산된 임베딩 e_{text} 과 이미지 처리 모델에서 계산된 e_{image} 를 이용해 최종 HS코드를 예측하기 위해 두 가지 방법을 이용했다. 첫번째 방법은 두 벡터를 결합하는 방법이다. $linear$ 함수와 활성화 함수를 통해 class 크기만큼의 y_{pred} 벡터를 생성한다.

$$y_{pred} = ReLU(linear(concat(e_{text}, e_{image})))$$

두번째로는 e_{image} 에 가중치 w 를 부여하고 가중합으로 더하여 벡터를 만든다.

$$y_{pred} = ReLU(linear(e_{text} + w * e_{image}))$$

마찬가지로 크로스 엔트로피 손실 함수를 이용하여 모델을 학습시켰다. 최종 HS코드를 결정할 때는 가장 큰 값을 가지는 class를 선택한다. 학습에 대한 정확도와 손실값이 수렴할 때까지 학습시킨 후, 검증 데이터의 정확도가 높은 모델을 선정하여 테스트 데이터로 모델의 정확도를 측정했다.

3.6. CLIP 모델

이 연구에서는 텍스트와 이미지 형태의 데이터를 함께 사용해 분류 문제를 해결하고자 했다. 여기서 활용된 데이터셋과 일반적인 데이터셋의 특성을 비교하기 위해 멀티모달 모델인 CLIP[3]을 활용해 분류 정확도를 확인해 보았다. CLIP은 해당하는 이미지와 문장 간의 임베딩 유사도를 최소화하는 대조 학습을 이용하여 학습되었으며, 방대한 양의 이미지와 텍스트를 통해 사전 학습되어 추가적인 파인튜닝 과정 없이도 새로운 문제를 해결할 수 있는 제로-샷 학습이 가능하다.

4. 실험 및 고찰

모델의 정확도

테스트 데이터에 대한 호와 소호의 정확도는 표 2와 같다. 각각 텍스트 처리 모델, 이미지 처리 모델, 그리고 이 두 모델을 이용한 결합 모델, 가중합 모델에 대한 결과이다.

표 2. 사전 학습 모델을 사용해 학습한 모델의 정확도

	호 정확도	소호 정확도
텍스트 처리 모델	0.704	0.614
이미지 처리 모델	0.303	0.257
결합 모델	0.692	0.425
가중합 모델	0.738	0.558

소호 분류의 경우 물품 설명만을 사용한 모델의 성능이 가장 높았지만, 호 분류 문제에서는 물품 설명과 이미지를 모두 사용한 가중합 모델이 가장 높은 정확도를 보여 텍스트와 이미지 데이터를 함께 사용했을 때의 효용성을 확인할 수 있었다.

사례 연구-데이터 비교

텍스트와 이미지 형태의 데이터가 분류 모델의 성능에 미치는 영향에 차이가 있음을 확인한 만큼, 몇 가지 사례를 통해 이러한 결과의 원인을 분석해 보았다. 그림 3은 같은 HS 코드로 분류된 최신 사례 3건에 대한 물품 설명과 이미지를 보여준다. 텍스트의 경우 ‘부분품’, ‘ATM’, ‘인쇄물’, ‘부속품’과 같이 분류에 핵심적인 정보가 되는 단어를 공통점으로 포함하고 있는 반면, 이미지는 사람이 보더라도 유사성을 찾기 힘들 정도로 다양한 형태를 보인다. 이러한 이미지 데이터의 특성을 고려하면 하나의 6자리 HS 코드 당 평균 사례 건수 10건은

품목 학습에 충분하지 않은 양임을 알 수 있다.

품목 1

소형 ATM, 티켓 자동발권기 등을 위한 (중략) 지폐가 모이는 공간 역할을 하는 부속품

품목 2

등사기와 전원케이블로 연결되어 명령을 받아 (중략) 인쇄물의 부수를 분류해주는 전용 부속품

품목 3

ATM(현금자동화기기)의 LCD 및 Function Key 위에 부착되어 (중략) 보호하기 위한 플라스틱 커버 제품



그림 3. HS 코드 8473.40 - 인쇄기, ATM 등의 부분품에 해당하는 품목들의 물품 설명 및 이미지

사례 연구-CLIP 모델 이용

학습에 사용할 수 있는 데이터의 양이 충분히 많지 않더라도 ImageNet과 같은 일반적인 데이터셋과 유사한 특성을 가진다면 사전 학습된 모델을 이용해 적은 양의 데이터로도 높은 성능을 기대할 수 있다. 관세 데이터의 보편성을 확인하기 위해 CLIP 모델을 활용해 관세법령정보포털에서 제공하는 HS 해설서로부터 선정된 후보 텍스트 중 주어진 물품 이미지를 설명하는 텍스트를 예측하는 문제를 시험해 보았다. 사전 학습된 모델과 관세 데이터를 이용해 파인튜닝된 모델을 활용해 실험했다.

그림 4에서 주어진 이미지는 문장 6번에 해당하는 품목이지만 CLIP 모델은 5번 문장에 가장 높은 점수를 주었고 정답 문장에는 오히려 가장 낮은 점수를 부여했다. 품목 이미지와 물품 설명이 짝 지어지도록 한차례 파인-튜닝한 결과, CLIP 모델은 여섯 문장에 대해 거의 같은 점수를 부여함으로써 이미지와 텍스트의 관련 정보를 전혀 학습하지 못하는 모습을 보여주었다. 이를 통해 관세 물품 사례 데이터는 일반적인 분류 문제에 활용되는 데이터와는 매우 다른 특성을 가짐을 알 수 있었다.

5. 결론

서로 다른 네 가지 모델을 이용해 분류 작업을 진행했고, 관세청 데이터셋의 이미지는 모습이 다양하고 일반적인 이미지와는 다른 특성을 가지고 있음을 확인했다. 호 분류와 같은 상위 분류에서는 해당 이미지 데이터를 이용해 유의미한 결과를 얻을 수 있었다. 향후 학습에 사용할 수 있는 많은 양의 데이터와 물체 감지 알고리즘을 활용하여 이미지의 특성을 효과적으로 검출하고 이미지의 활용성을 높여 이미지의 특성을 효과적으로 검출하고 이미지의 활용성을 높이는 것을 제안한다.



정답 HS 코드 :

8462.62

(냉간 금속가공용 기계형 프레스)

1. 단조용 · 형 단조용 열간 성형기(프레스를 포함한다)와 열간 해머
2. 평판제품용 굽힘기 · 접음기 · 교정기 · 펼침기[프레스 브레이크를 포함한다]
- ...
5. 관 · 파이프 · 중공이 있는 형강 · 봉 가공용의 기계(프레스를 제외한다)
6. 냉간 금속가공용 프레스

파인-튜닝 모델

1. 0.165
2. 0.165
3. 0.175
4. 0.175
5. 0.165
6. 0.155

사전 학습 모델

1. 0.280
2. 0.113
3. 0.085
4. 0.051
5. 0.468
6. 0.003

후보 문장

CLIP 점수

그림 4. 물품 이미지가 주어졌을 때 각 후보 문장에 CLIP 모델이 부여한 예측 점수

사사문구

이 연구는 기초과학연구원[IBS-R029-C2, IBS-R029-Y4]과 관세청의 지원을 받았음.

참고 문헌

[1] Luppés, Jeffrey, Arjen P. de Vries, and Faegheh Hasibi. "Classifying short text for the harmonized system with convolutional neural networks." Radboud University (2019).

[2] 이동주, 최근호, and 김건우. "HS 코드 분류를 위한 CNN 기반의 추천 모델 개발." 경영과 정보연구 39.3 (2020): 1-16.

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