

DATA CREATOR CAMP

2024 데이터 크리에이터 캠프

starter

대학부

출제문제





구성

문제
조건

Mission 1
Mission 2
Mission 3

제출 파일 안내



문제. 인공지능은 사람의 마음을 이해할 수 있을까?

■ 연도별 패션 스타일 이미지 및 추천 데이터

2022

#의류 이미지 #패션 #포즈 #웨이프리스의류

NEW 연도별 패션 선호도 파악 및 추천 데이터

분야: 영상이미지 유형: 이미지

구축년도: 2022 갱신년월: 2023-10 조회수: 4,960 다운로드: 379 용량: 333.22 GB

다운로드 ↓ 샘플 데이터 ?

■ 데이터셋 외 코드 제공되지 않음



조건.

- Image Classifier 및 Image Feature Vector의 아키텍처(architecture)는 ResNet-18을 사용해야 한다.
- Pre-trained Weights 사용할 수 없다.
- 제출파일(ipynb)에서 최종결과와 관련 없는 코드 및 cell은 삭제한다.
- 마크다운 및 주석으로 각 cell의 역할을 설명한다.
- 성능 평가지표는 Top-1 accuracy를 사용한다.
- Tensorflow, Keras, PyTorch 등 에서 제공하는 전이학습 네트워크를 사용할 수 없다.



Mission 1. 패션 스타일 이미지 분류

- 1-1. 주어진 이미지 데이터의 파일명은 아래와 같은 형식이다. “{W/T}_{이미지ID}_{시대별}_{스타일별}_{성별}.jpg”에 기반하여 “이미지ID” 수 기준으로 “성별 & 스타일” 통계치를 아래 표 형식으로 기입한다.

※ Training, Validation 데이터에 대해서 각각 통계표를 작성한다.

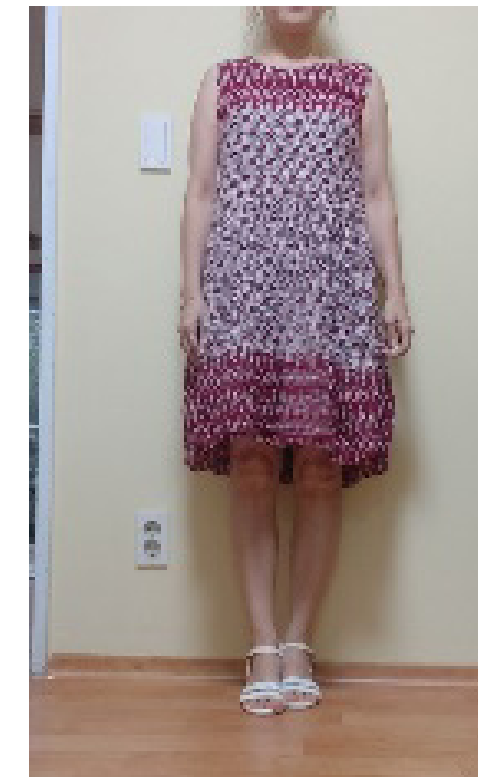
성별	스타일	이미지 수
여성	feminine	
	classic	
	minimal	
	popart	
	...	
남성	ivy	
	mods	
	hippie	
	bold	
	...	

```
W_96469_60_minimal_W.jpg
W_96472_19_lounge_W.jpg
W_96487_19_lounge_W.jpg
W_96507_19_lounge_W.jpg
W_96514_60_minimal_W.jpg
W_96538_60_minimal_W.jpg
W_96574_19_normcore_W.jpg
W_96596_60_popart_W.jpg
W_96599_60_popart_W.jpg
W_96600_60_minimal_W.jpg
W_96606_60_popart_W.jpg
W_96607_60_popart_W.jpg
W_96612_60_popart_W.jpg
W_96616_60_minimal_W.jpg
W_96617_60_minimal_W.jpg
W_96619_60_minimal_W.jpg
W_96625_60_minimal_W.jpg
W_96626_60_minimal_W.jpg
W_96632_60_minimal_W.jpg
W_96634_60_minimal_W.jpg
W_96637_60_minimal_W.jpg
W_96643_60_minimal_W.jpg
```




Mission 1. 패션 스타일 이미지 분류

- 1-2. ResNet-18를 활용하여 “성별 & 스타일” 단위로 클래스 분류를 수행하고 Validation 데이터에 대한 정확도를 제시한다.
 - ResNet-18의 parameters는 무작위로 초기화하여 사용한다.
(즉, pretrained weights는 사용할 수 없음)
 - 성능을 높이기 위해 object detection, image cropping 등의 다양한 데이터 전처리 기법을 활용해도 무방하다.
(데이터 전처리 단계에 한해서는 외부 라이브러리 활용 가능)



여성 에콜로지룩



여성 오리엔탈룩



여성 농코어룩



여성 클래식룩



Mission 2. 패션 스타일 선호 여부 예측

- 2-1. 주어진 라벨링 데이터의 파일명은 아래와 같은 형식이다. “{W/T}_{이미지ID}_{시대별}_{스타일별}_{성별}_{설문ID}.json” 이에 기반하여 “설문ID” 수 기준으로 “성별 & 스타일” 통계치를 아래 표 형식으로 기입한다.

※ 이때 주어진 이미지 데이터에 존재하는 “이미지ID”를 식별하여 유효한 라벨링 데이터 대상으로만 통계치를 구해야 한다. (이미지ID 기준으로 라벨링 데이터에는 있지만, 이미지 데이터에는 없는 경우가 있음)

※ Training, Validation 데이터에 대해서 각각 통계표를 작성한다.

```
W_96626_60_minimal_W_008455.json
W_96626_60_minimal_W_234988.json
W_96632_60_minimal_W_008463.json
W_96632_60_minimal_W_234973.json
W_96634_60_minimal_W_008471.json
W_96634_60_minimal_W_234981.json
W_96637_60_minimal_W_008472.json
W_96637_60_minimal_W_234989.json
W_96643_60_minimal_W_018557.json
W_96643_60_minimal_W_234974.json
W_96645_60_minimal_W_018565.json
W_96645_60_minimal_W_234982.json
W_96646_60_minimal_W_018573.json
W_96646_60_minimal_W_234990.json
```

성별	스타일	이미지 수
여성	feminine	
	classic	
	minimal	
	popart	
	...	
남성	ivy	
	mods	
	hippie	
	bold	
	...	



Mission 2. 패션 스타일 선호 여부 예측

- 2-2. 2-1에서 구한 유효한 라벨링 데이터만 따로 분리하여 아래와 같이 100명 응답자의 “스타일 선호 정보표”를 구한다. 파일은 json 포맷으로 되어 있으며 json 필드 중, “응답자ID”는 “user>R_id”로 알 수 있고, “스타일 선호 여부”는 “item>survey>Q5”로 알 수 있다.

※ 스타일 선호도 값은 “1: 비선호”, “2: 선호”이다.

응답자 ID	Training		Validation	
	스타일 선호	스타일 비선호	스타일 선호	스타일 비선호
64747	W_07894_00_cityglam_W.jpg	W_44386_80_powersuit_W.jpg	W_05628_00_cityglam_W.jpg	W_34024_10_sportivecasual_W.jpg
	W_37160_70_punk_W.jpg	W_34573_10_sportivecasual_W.jpg	W_37491_70_military_W.jpg	W_11610_90_grunge_W.jpg
	W_39725_19_normcore_W.jpg	W_40876_70_punk_W.jpg	W_38588_19_genderless_W.jpg	W_47169_70_hippie_W.jpg

...				

```
"E_id": 18566,
"imgName": "W_96651_60_minimal_W.jpg",
"item": {
  "imgName": "W_96651_60_minimal_W.jpg",
  "era": 1960,
  "style": "minimal",
  "gender": "W",
  "survey": {
    "Q1": 2,
    "Q2": 1,
    "Q3": 1,
    "Q411": 1,
    "Q412": 2,
    "Q413": 2,
    "Q414": 2,
    "Q4201": 0,
    "Q4202": 0,
    "Q4203": 0,
    "Q4204": 4,
    "Q4205": 0,
    "Q4206": 0,
    "Q4207": 0,
    "Q4208": 0,
    "Q4209": 0,
    "Q4210": 0,
    "Q4211": 0,
    "Q4212": 0,
    "Q4213": 0,
    "Q4214": 14,
    "Q4215": 0,
    "Q4216": 0,
    "Q5": 2
  }
},
"user": {
  "R_id": 14762,
  "r_gender": 2,
  "age": 1,
  "mar": 1,
  "job": 5,
  "income": 1,
  "r_style1": 1,
  "r_style2": 1,
  "r_style3": 2,
  "r_style4": 2,
  "r_style5": 1
}
```




Mission 3. 패션 스타일 선호 여부 예측

- 3-1. 추천 시스템의 기본인 협업 필터링 (Collaborative Filtering)은 크게 user-based filtering, item-based filtering 방식으로 나뉘어져 있다. 각각에 대해서 이해하고, 2-2에서 구해 본 응답자의 “스타일 선호 정보표”를 토대로 Validation 데이터 내 응답자의 “스타일 선호 여부 예측” 문제를 2가지 기법으로 어떻게 적용해 볼 수 있고, 서로 비교하여 어떤 장단점을 갖는지 설명한다.

※ 설명을 용이하게 하기 위해 응답자의 스타일 선호도 예시를 들어서 설명해도 무방하다.



Mission 3. 패션 스타일 선호 여부 예측

- 3-2. 3-1에서 살펴 본 기법 중, item-based filtering을 직접 구현해본다. “이미지 간 유사도” (image2image)만을 활용하여 Validation 데이터 내 응답자의 “스타일 선호 여부 예측” 문제를 수행하고 성능을 측정한다.

※ Hint. 1-2에서 학습한 ResNet-18의 중간 layer 값을 활용하여 각 이미지의 feature vector를 구하고, 벡터 연산을 통해 이미지 간 유사도를 구해볼 수 있다.

※ 예측 문제에서 활용한 파라미터 및 임계 값 등의 수치를 정확하게 제시한다.



제출 파일 안내.

■ ipynb 파일

- 총 3개 (각 Mission 1,2,3에 해당)
- 본 모델 코드 포함
- 마크다운(Markdown) 및 주석으로 정리 필수

■ 결과 정리 PPT

- 전반적 미션 수행 내용 포함 (데이터 분석 내용, 모델 설명, 결과 분석 등)

감사합니다.

2024 데이터 크리에이터 캠프