

wandb 사용법

wandb란?



- train 과정에서 온라인 tracking과 visualization 해주는 툴임.
- hyper-parameter tuning, 모델별 loss, score비교, 학습 종료 지점 등 파악 가능

wandb 사용법

우리 코드에서 wandb는 4파트로 나뉜다. 1) import, 2) init, 3) log, 4) finish

결국 logging을 위해 나머지 3파트가 존재한다.

1. import

- wandb를 사용하고자 하는 파이썬 파일 상단에 `wandb` 를 import 해야한다(ex. train.py)
- `wandb` 와 `dotenv` 를 import하기 위해 아래와 같이 코드 내에서 `pip install` 하게 하였다

```
import sys
import subprocess as sp
sp.check_call(["sudo", sys.executable, "-m", "pip", "install", "python-dotenv"])
sp.check_call(["sudo", sys.executable, "-m", "pip", "install", "wandb"])
import wandb
from dotenv import load_dotenv
```

2. init

- wandb를 run 하기 위해 epoch 돌기 직전에 wandb를 시작하는 단계

```
load_dotenv(dotenv_path="wandb.env")
WANDB_AUTH_KEY = os.getenv("WANDB_AUTH_KEY")
wandb.login(key=WANDB_AUTH_KEY)
wandb_name = save_path.split('/')[-1]
wandb.init(
    entity="rond-7th",
    project="fashion-how",
    group=f"sub-task1",
    name=wandb_name
)
```

- `wandb.init` 함수 내부에서 설정은 다음과 같다
 - `entity`: 우리 그룹 명으로, `rond-7th` 로 정했다 (고정)
 - `project`: 는 우리 그룹이 진행하는 프로젝트로, `fashion-how` 로 정했다 (고정)
 - `group`: 프로젝트 내의 실험 그룹 분류로, 태스크마다 지정하면 된다. 우선 `sub-task1`
 - `name`: 이번 run의 이름. yaml 이름과 통일 시키면 된다.
- 위 코드가 최초 실행시 터미널에서 author 코드를 물어 본다.

- entity 내에서 어느 계정으로 올렸는지 정보를 주고, 권한을 부여하기 위한 코드
- <https://wandb.ai/authorize> 에서 코드 확인 가능
- 연구소에서 wandb 접속 안되니 휴대폰 등을 이용해 코랩 어딘가에 저장해두기

3. log

- 원하는 시점에 `wandb.log` 를 호출하면 parameter인 dict가 로깅이 된다.
- key는 wandb에서 보게 될 변수이름이고, value는 로깅하고 싶은 값 (float, image, ..) 이다.

```
train_info = {
    "train/loss": epoch_losses[0],
    "train/loss_daily" : epoch_losses[1],
    "train/loss_gender": epoch_losses[2],
    "train/loss_embel" : epoch_losses[3]
}
wandb.log(train_info)
```

- 실시간으로 휴대폰으로 그래프를 확인할 수 있다.

4. finish

- 학습이 종료되면 wandb 꺼야함. 자동으로 꺼지지만 그래도 해주자.

```
wandb.finish()
```