

핀테크 2021 딥러닝 조별 프로젝트 설명

프로젝트 설명

프로젝트

- 전체 프로젝트의 진행과 발표, 평가는 기계학습 프로젝트와 동일합니다.
- 딥러닝 수업 시간에 배운 여러 알고리즘이 실제 데이터에 어떻게 적용될 수 있는지 알아보고, 각 모델 별 차이를 알아보는 시간을 가지는 데 목표가 있습니다.
- 수업시간에 배웠던 Ensemble 방법론처럼 다양한 분야의 학생들이 모였으니, 팀원들간의 충분한 토의를 통해 더 나은 결과물을 보여주세요!
- 개인 GPU가 있으시다면 GPU를 사용하셔도 무방하며, 그렇지 않다면 Colab에서 런타임을 GPU로 바꾸어서 사용하시길 권장드립니다. (다만 Colab의 경우 시간 제한이 있어 중간에 사용한다고 있다는 확인을 체크해주셔야 하거나 너무 큰 모델은 사용 시간을 초과할 수도 있습니다.)

프로젝트 기간

- 03.05(금): 프로젝트 설명
- 03.29(월): 프로젝트 발표 (발표 15~20분 + Q&A 10분), 1차 코드 제출
- 04.04(일): 프로젝트 보고서 제출 (보고서 파일은 발표에 사용된 ppt 파일에 피드백을 수정해서 코드와 함께 제출해주세요) - 이 일정은 핀테크 과정의 다른 과제의 일정에 따라 단축/연장될 수 있음을 미리 알려드립니다.

프로젝트 평가

- 발표: 발표자료(5), 발표자의 태도(3), 질의 답변 및 발표시간 준수(2) 에 따라 교수님 , 조교, 상호 평가(조별 발표자)가 평가
- 보고서: 독창성, 유의성, 발표 피드백 반영 에 따라 교수님 , 조교가 평가

프로젝트 구성

아래는 예시일 뿐이며, 다른 순서와 내용으로 프로젝트를 구상하셔도 무방합니다.

Data Selection

- kaggle, 기타 public에 공개된 데이터 사용 (크롤링 가능)
- 데이터 설명과 택한 이유 등 설명
- visualization 등등
- 딥러닝은 기계학습에 비해 training time이 훨씬 오래 걸립니다. 따라서, 너무 큰 데이터를 고르면 학습에 너무 많은 시간이 걸릴 수 있기 때문에 적당한 데이터를 고르길 바랍니다.

Preprocessing

- NaN 제거
- 이상치 제거
- 원하는 데이터 추출
- 데이터의 개수가 맞지 않을때 불균형 해소

Modeling

- CNN(이미지), RNN(시계열), GAN, Autoencoder 등등 사용 무방합니다.
- 여러 모델을 사용해보거나, 만일 이미지 데이터를 골라 CNN을 고른다면 LeNet, ResNet 등을 여러 모델을 바꿔가며 실험을 진행하여 결과를 확인해도 무방합니다.

Result & Discussion(*기계학습 부분에서는 이 부분이 미흡한 경우가 많았음*)

- 얻은 결과와 이에 대한 해석 공유

- 어떤 모델이 잘 작동하는지, 왜 잘 작동하는지에 대한 해석
- 딥러닝의 결과가 데이터에 대해서 어떤 부분을 설명해주는 지에 대한 해석

참고 자료

작년 프로젝트

2020 핀테크 딥러닝 프로젝트 중 4가지 프로젝트를 github로 공유하니 참고하시기 바랍니다.

- https://github.com/SLCFLAB/FinTech2021/tree/main/Projects_딥러닝
- 작년 프로젝트는 NLP에서 BERT를 중심으로 프로젝트를 진행하였으나, NLP 수업이 따로 분리되어 다양한 모델들 중 데이터에 맞는 원하는 모델을 사용하셔도 됩니다.

프로젝트를 하다 궁금한 점이나 해결되지 않는 부분이 있으면 조교 박진성(jinseong@snu.ac.kr)로 보내주세요!