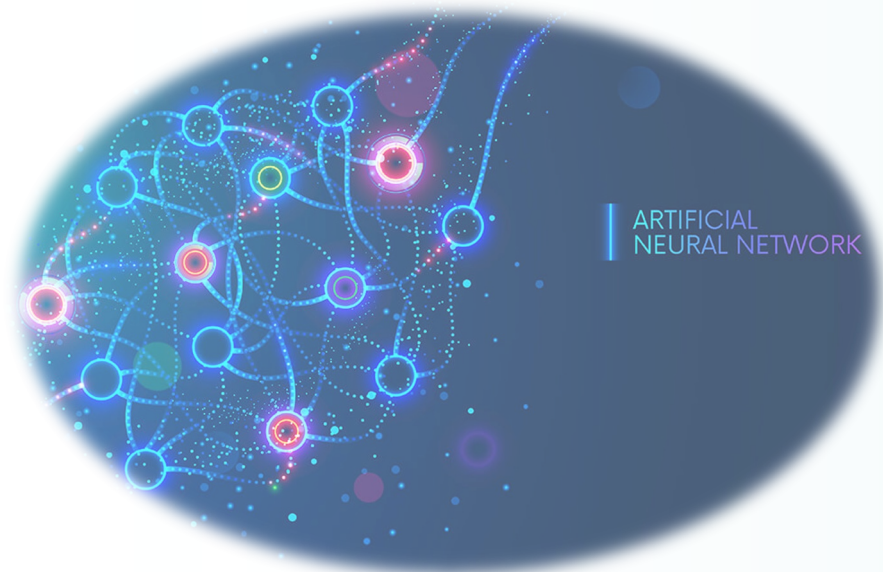


〈프로젝트 #2〉

시퀀스 데이터 모델 및 CNN 전이학습 구현

패턴인식과 딥러닝
2022학년도 2학기



INDEX

1. 시퀀스 데이터 모델 구현
2. 전이학습을 이용한 CNN 분류기 구현
3. 프로젝트 제출 방법
4. 프로젝트 관련 질의 응답

1. 시퀀스 데이터 모델 구현

1) 과 제 내 용

- 강의자료 11 ~ 13 에 있는 IMDb 데이터셋 모델 실습 구현

2) 목 표

- 텐서플로우와 케라스를 이용하여 시퀀스 데이터 전처리 및 시퀀스 데이터 모델 숙지
- IMDb 데이터셋 전처리
 - 강의자료에 있는 IMDB 데이터셋을 도메인 지식 활용하여 전처리한 데이터(인코딩 된 입력 데이터의 하위 100개 시퀀스 사용)를 이용
 - Bidirectional RNN, Bidirectional LSTM, 트랜스포머 모델 실습

3) 과제 1의 제출 방법

- 각 모델별 실습 결과 **노트북 파일** 제출
 - 한 노트북에 여러 모델을 학습할 경우 메모리 문제가 발생할 수 있어 모델별로 노트북 생성하여 실습

1. 시퀀스 데이터 모델 구현

4) 과제 1 제출 방법 (계속..)

- 노트북 파일 제출
 - 각 모델에 대한 3개의 노트북 파일을 만들어 제출

노트북 파일명	실습 내용
1_bidirectional_rnn.ipynb	Bidirectional RNN 의 구현 및 학습/평가 결과
2_bidirectional_lstm.ipynb	Bidirectional LSTM 의 구현 및 학습/평가 결과
3_transformer_model.ipynb	트랜스포머 모델의 구현 및 학습/평가 결과

2. 전이학습을 이용한 CNN 분류기 구현

- CIFAR100 데이터셋
 - 이미지의 크기, 데이터의 수는 CIFAR10 데이터셋과 동일
 - 각 이미지의 크기는 32 x 32 의 컬러이미지
 - 훈련 데이터셋: 50,000장
 - 테스트 데이터셋: 10,000장
 - <https://www.cs.toronto.edu/~kriz/cifar.html>
 - 레이블링은 100개의 Class로 되어있으며 100개의 클래스 중 유사한 것들을 묶어 20개의 Super class 로 Labeling

2. 전이학습을 이용한 CNN 분류기 구현

- CIFAR100 데이터셋
 - 20개의 Superclass 와 100개의 Class

Superclass

aquatic mammals
fish
flowers
food containers
fruit and vegetables
household electrical devices
household furniture
insects
large carnivores
large man-made outdoor things
large natural outdoor scenes
large omnivores and herbivores
medium-sized mammals
non-insect invertebrates
people
reptiles
small mammals
trees
vehicles 1
vehicles 2

Classes

beaver, dolphin, otter, seal, whale
aquarium fish, flatfish, ray, shark, trout
orchids, poppies, roses, sunflowers, tulips
bottles, bowls, cans, cups, plates
apples, mushrooms, oranges, pears, sweet peppers
clock, computer keyboard, lamp, telephone, television
bed, chair, couch, table, wardrobe
bee, beetle, butterfly, caterpillar, cockroach
bear, leopard, lion, tiger, wolf
bridge, castle, house, road, skyscraper
cloud, forest, mountain, plain, sea
camel, cattle, chimpanzee, elephant, kangaroo
fox, porcupine, possum, raccoon, skunk
crab, lobster, snail, spider, worm
baby, boy, girl, man, woman
crocodile, dinosaur, lizard, snake, turtle
hamster, mouse, rabbit, shrew, squirrel
maple, oak, palm, pine, willow
bicycle, bus, motorcycle, pickup truck, train
lawn-mower, rocket, streetcar, tank, tractor

2. 전이학습을 이용한 CNN 분류기 구현

▪ CIFAR100 데이터셋

- tensorflow_datasets API 를 이용한 CIFAR100 데이터셋
 - tfds 의 builder 메서드 호출 시 매개변수로 'cifar100' 을 사용
 - Training Set 50,000 장 중 일부는 Validation Set 으로 하여 학습
 - 성능 평가는 Test Set 10,000 장에 대한 Accuracy, Precision, Recall, Average Precision 평가

1) 과 제 내 용 및 목표

- CIFAR100 데이터셋의 Superclass(20개의 클래스) 의 분류 문제를 해결하기 위한 CNN 설계 및 실험 분석을 통하여 실전 모델 설계 능력 배양
- CNN 은 전이학습(Transfer learning, Fine-tuning) 방법을 이용하여 설계 및 구현

2. 전이학습을 이용한 CNN 분류기 구현

2) 보고서 작성 방법

- 전이학습을 이용한 CIFAR100 분류기 설계부터 실험결과 및 분석 과정을 코드, 그림, 표, 그래프 등을 이용하여 상세히 서술(온라인 제출로 분량 제한 없음)
- 표지를 포함한 문서 형식으로 작성
- 최초 모델 설계부터 성능 향상을 위한 다양한 시도를 해보고 해당 내용을 서술
ex) 네트워크 레이어 구성, 규제 적용, 학습 파라미터 튜닝 등
- 보고서 작성 시 업로드한 양식을 표지로 하여 작성(워드, 한글 파일 모두 제공)

2. 전이학습을 이용한 CNN 분류기 구현

3) 과제 2 제출 방법

- **보고서**

- 보고서는 양식 표지가 포함된 한글 또는 워드로 작성 후 PDF 파일로 변환하여 제출 권장
 - PDF 파일 제출이 제한적인 경우 한글 또는 워드 원본 파일을 제출해도 무관
- 보고서 작성 방법
 - 보고서 작성 시에는 주피터 노트북의 코드와 결과 화면이 보이는 캡처 화면을 문서 내에 삽입
 - 과제 진행에 있어 문제 해결 과정에서 다양한 실험과 결과 분석들을 자유롭게 작성
 - 보고서 성실성에 따라 채점 결과 차이 있음

- **노트북 파일(.ipynb)**

- 과제 2를 수행한 주피터 노트북 파일
- 노트북 파일명은 'project.ipynb' 로 제출

3. 프로젝트 #1 제출 방법

- 제출 방법 및 기한

- 과제 1의 노트북파일과 과제 2의 보고서 및 노트북 파일을 모두 합쳐 하나의 압축파일로 제출
 - 파일명은 'pr2_학번_성명.zip' 으로 제출
ex) 'pr2_4100000000_김송실.zip'
- LMS(스마트 캠퍼스) 에 온라인 제출
 - [강의콘텐츠], 15주차 항목 중 '과제' 콘텐츠를 통하여 개별 온라인 제출
- 제출 기한: 12월 14일 수요일 23시 59분
 - 12월 8일부터 제출가능

4. 프로젝트 평가 방법

■ 배점 기준

- 본인 자체 평가 20점, 교수 평가 80점 총 100점
 - 배점은 보고서 표지 양식 참고
- 본인 자체 평가는 프로젝트 표지에 표기

프로젝트 #2

제 목: 시퀀스 데이터 모델 및 CNN 전이학습 구현

1. 프로젝트 자체 평가(본인 평가)

① 완성도 및 성실성(20점) : 상(20)() / 중(18)() / 하(16)()

2. 프로젝트 외부 평가(교수 평가)

① 과제 #1의 완성도(15점) : 상(15)() / 중(12)() / 하(12)()

② 과제 #2의 완성도(35점) : 상(35)() / 중(30)() / 하(25)()

③ 보고서 작성 성실성(30점) : 상(30)() / 중(26)() / 하(22)()

④ 제출일 미준수로 인한 감점(-20점) : 제출일 미준수(-20)()

소계: () 점

※ 총점: () 점

과 목 명: 패턴인식과 딥러닝

학 과: ○ ○ ○ ○ ○ ○

학 번: ○ ○ ○ ○ ○ ○

성 명: 김 승 실

제 출 일: 2022년 12월 8일(수)

4. 프로젝트 평가 방법

■ 자체 평가 20점

- 수강생 본인이 수행한 프로젝트에 대한 '완성도와 성실성'에 대한 자체 평가 점수를 상(20), 중(18), 하(16)로 나누어 평가하여 보고서 표지에 표기

■ 교수 평가 80점

- 과제 #1의 완성도: 15점
 - 상(15), 중(12), 하(10)
- 과제 #2의 완성도: 35점
 - 상(35), 중(30), 하(25)
 - 상: 모델 구현과 학습 및 평가 성공
 - 중: 학습 및 평가는 하였지만 성능이 어느 정도 수준을 넘지 못함
 - 하: 학습 및 평가 진행이 불가능한 경우
- 보고서 작성 성실성: 30점
 - 상(30), 중(26), 하(22)
 - 상: 학습 결과에 대한 분석 및 다양한 실험 시도를 서술
 - 중: 코드와 결과 화면만 복사

프로젝트 #2

제 목: 시퀀스 데이터 모델 및 CNN 전이학습 구현

1. 프로젝트 자체 평가(본인 평가)

① 완성도 및 성실성(20점): 상(20)() / 중(18)() / 하(16)()

2. 프로젝트 외부 평가(교수 평가)

① 과제 #1의 완성도(15점) : 상(15)() / 중(12)() / 하(12)()
② 과제 #2의 완성도(35점) : 상(35)() / 중(30)() / 하(25)()
③ 보고서 작성 성실성(30점) : 상(30)() / 중(26)() / 하(22)()
④ 제출일 미준수로 인한 감점(-20점): 제출일 미준수(-20)()
소계: () 점

※ 총점: () 점

과 목 명: 패턴인식과 딥러닝

학 과: ○ ○ ○ ○ ○ ○

학 번: ○ ○ ○ ○ ○ ○

성 명: 김 승 실

제 출 일: 2022년 12월 8일(수)

5. 프로젝트 관련 질의 응답

- 15주차 강의 기말고사 주로 대체 및 질의 응답 진행
 - 15주차 수업(12월 8일 19:30)은 기말고사 주로 대체하여 강의 진행은 없으며 프로젝트 관련 실시간 질의 응답을 수행
 - 질문이 있는 수강생은 실시간 화상 회의에 참여하여 질의 응답을 진행
 - 화상 회의에 참여하지 않더라도 전원 출석 처리
 - 이 외의 시간의 질문은 e-mail 을 통하여 가능하지만 답변이 늦어질 수도 있음