KUBIG 2022-1

① 첫번째, NLP 분반

분반장: 쿠빅 13기 통계학과 19 기다연

분반원: 쿠빅 12기 원윤정, 13기 배은지, 14기 제갈예빈, 이연정, 15기 김진수

KUBIG NLP





스터디 OT

- 스터디 진행방식
- 코드 구현과제 예시
- 주차별 커리큘럼 복습





아이스 브레이킹

- 간단한 자기소개
- 사용 프레임워크/코딩 툴 조사
- NLP 공부/프로젝트 관련 경험





1주차

- 1. 딥러닝 복습
- 2. 자연어처리의 큰 흐름
- 3. 자연어처리의 응용분야

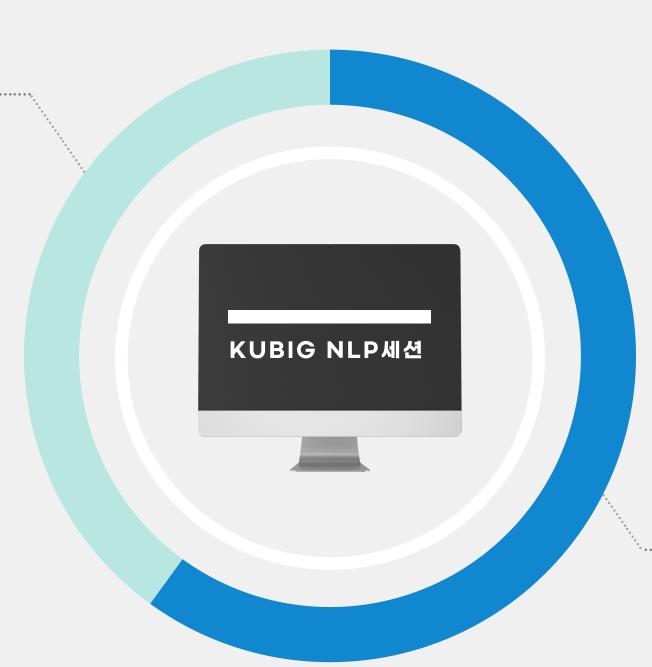
PART 01

- 1. 스터디 진행방식
- 2. 코드 구현과제 예시
- 3. 주차별 커리큘럼 복습

전반 30분 ~ 1시간

코드구현 과제 공유

각자 이전 주에 나간 코드구현 과제 공유 (코드구현 과제는 배운 내용의 복습 과제)



후반 1시간 ~ 1시간 30분

스터디식 강의

주차별 커리큘럼에 따라 준비해온 내용 발표 (배우는 내용은 이전 주에 미리 예습 진행)

* 모든 스터디원들은 캠을 필수적으로 켜주셔야 합니다!

2주차 코드 구현과제의 일부

RNN/LSTM

- 1. 스팸 메일 분류하기: 이진 분류(Binary Classification)
 - https://wikidocs.net/22894
 - 스팸 메일 분류는 정상 메일과 스팸 메일 중 스팸 메일을 찾아내는 이진 분류 모델입니다.
- 2. 리뷰 감성 분류하기: 감성 분류(Sentiment Analysis)
 - https://wikidocs.net/24586, https://wikidocs.net/44249, https://wikidocs.net/94600
 - 감성 분류는 데이터 셋의 텍스트의 긍부정 여부를 판단하는 모델입니다. 위 세가지 사이트의 감성 분류는 LSTM이나 GRU를 통해 이루어져 있지만, model building 시의 LSTM과 GRU를 simpleRNN으로 대체해도 모델 학습이 가능합니다. 더 좋은 모델을 위해 LSTM이나 GRU를 사용하셔도 됩니다.
- 3. [Kaggle] 텍스트 생성하기: 언어 모델(Language Models)
 - https://www.kaggle.com/roblexnana/generating-text-for-nlp-using-simplernn-with/
 - 관련 데이터 셋 다운로드: https://www.kaggle.com/roblexnana/alice-wonderland-dataset
 - 텍스트 생성은 소설 '이상한 나라의 엘리스' 텍스트를 기반으로, Test Data가 되는 문자들이 입력되었을 때 그 다음에 나올 내용(문자)들을 생성해주는 언어 모델을 형성하는 코드입니다. 소설 텍스트가 짧아서 위 코드에서는 문자(character) 단위로 진행하였지만, 더 긴 텍스트 등을 사용하거나 해서 단어(word) 단위로 진행하면 더 재밌는 과제가 될 수 있을 것 같습니다.
- 4. [Kaggle] 주가 예측하기: 시계열 데이터 분석 및 예측
 - https://www.kaggle.com/ozkanozturk/stock-price-prediction-by-simple-rnn-and-lstm
 - 주가 예측은 Tesla 사의 주가 데이터를 활용해 이후 시가를 예측하는 코드입니다. RNN과 LSTM 중에 RNN 부분만 진행하셔도 무방하고, 더 공부하시기 위해서 LSTM도 해보시는 것을 추천합니다. 이외에 번역 등 순환신경망을 이용한 다른 활용 방법을 찾아보시는 것 역시 환영합니다.

코드 구현과제는 3가지 중 **하나**를 수행해주시면 됩니다.

- 1. 코드를 보고 똑같이 따라서 구현해본다.
- 2. 주어진 코드나 주제를 응용해본다.
- 3. 주어진 데이터셋을 활용해 원하는 모델을 구현해본다.

각 주차별 노션에 올린 코드 구현과제의 예시 주제/데이터셋 외에도 자율적으로 원하시는 코드가 있거나, 흥미로운 주제가 있으시면 구현해와 주시면 됩니다!

구현하시면서 모르는 부분이나 궁금한 점이 생기면,

- 노션 Q&A 페이지
- NLP 분반 카카오톡 단톡방

적극 활용해주세요!

다 하시면, 노션의 '스터디 과제 정리' 페이지에 링크 (공유권한!)를 남겨주시면 됩니다!

* 저도 이렇게 매주 배운 내용을 이론에서 그치는 것이 아니라 실제 코드로 어설프게나마 구현해보니 실력이 크게 성장한 것 같아요!

노션 워크스페이스 링크

설연휴



KUBIG CONTEST?

각 팀이 정한 NLP 주제를 이용해 팀끼리 회의 진행, 프로젝트 완성하는 경험

* 저번 NLP 분반 기수들의 프로젝트는 노션의 KUBIG CONTEST 페이지 참조!

KUBIG _ NLP



- 1. 간단한 자기소개
- 2. 사용 프레임워크/코딩 툴 조사
- 3. NLP 공부/프로젝트 관련 경험



저희 NLP 2022-1학기 방학분반은 총 5분이 신청해주셨습니다! 오늘부터 2달간 함께하실 분들이기 때문에 친해지길 바래요! (저도..) 순서대로 각자

- 이름 / 학과 / 나이와 같은 기본적인 정보
- 좋아하는 것!싫어하는 것!
- 취미 (쿠빅/데이터 사이언스/유사 답안 제외..)
- NLP 공부/프로젝트 관련 경험
- NLP 분반에 대한 포부..?

에 대해서 맘껏 소개해주세요 -!

단순히 코드 구현과제와 KUBIG CONTEST를 위한 수요조사입니다 ! 코드 구현과제의 경우, 자신이 편한 툴을 이용해주시면 됩니다.

KUBIG CONTEST를 위한 프로젝트의 경우, 팀원들과 함께 코워킹이 가능한 툴을 이용해주시면 됩니다.



- 1. PyTorch
- 2. Tensorflow



- 1. jupyter notebook
- 2. Google colab
- 3. Pycharm
- 4. Visual Studio Code
- 5. 기타..?

KUBIG _ NLP

PART O3 T T T



- 1. 딥러닝 복습
- 2. 자연어처리의 큰 흐름
- 3. 자연어처리의 응용분야

끎 ♬ ② ☆ …

PPT보다 노션에서의 정리가 더 자신있어 모든 내용은 노션 페이지를 통해 공유해드리도록 하겠습니다. 노션 페이지는 꼭! 복제하셔서 사용해주세요!

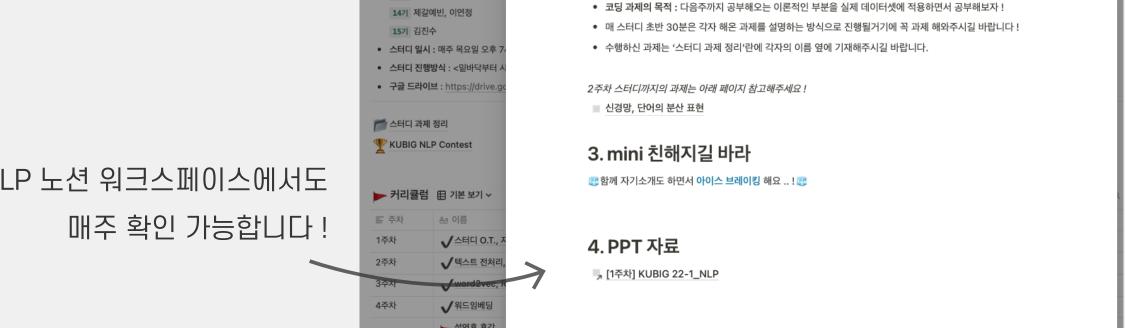
복제하신 다음에, 더 공부하면서 내용을 수정/삭제/삽입하시길 바랍니다!



https://zoeylaboratory.notion.site/1-KUBIG-22-1_NLP-96d033247c9b451d9dc91e6035beccec

12기 원윤정

13기 배은지



Attention, Transforme

2022년 2월 10일 💹 사전학습모델 (BERT, GPT)

☑ 페이지로 열기

저희 KUBIG 2022-Winter NLP 노션 워크스페이스에서도 매주 확인 가능합니다!

KUBIG 2022-1

감사합니다!

저희 2달간 앞으로 같이 열심히 해봐요!



쿠빅 13기 | 기다연

TEL. 010-8535-0353

E-mail dayeonki@gmail.com