

자연어처리 중간발표

14조

한윤
김건희
임특

목차

- 1. 프로젝트내용
 - 1. 한국어 감정분석
 - 2. 영어 감정분석

- 2. 계획 및 진행상황

프로젝트 내용

- 한국어 감정분석 :

binary classification 문제. NSMC 영화 리뷰 데이터셋을 바탕으로 긍정/부정을 예측하기

- 영어 감정분석 :


Multi-class classification 문제. Friends 데이터셋을 바탕으로 감정을 분류하기

한국어 감정분석 계획

1. 실험 환경 : Google Colab

2. 데이터 로드 : `open file(.txt) -> read data`

3. 데이터 전처리



품사태깅하여 .json파일에 저장



Count Vectorization, KoNLPy 내부 클래스 사용

3. 모델 정의와 학습

- 중간 레이어 사이에서는 **ReLU**를 활성화 함수로 사용하고, 출력층 레이어에서는 **sigmoid** 를 사용하여
0과 1 사이의 확률로 변환 시킴.
- 출력층에서는 **sigmoid** 를 거친 출력값을 출력.
- 손실함수와 **optimizer**의 선택 :

손실함수 : Binary cross entropy

Optimizer : Adam optimizer

4. 결과 확인

영어 감정분석 계획

Multi -class classification 과 binary classification 은 비슷하지만 2가지 차이점이 존재함.

- Binary classification 의 output은 sigmoid 함수를 이용하기 때문에 0 과 1사이의 확률값으로 존재함.
- Multi-class classification 경우에는 one hot encoding 이용해서 출력 레이어의 출력값을 벡터로 변환.

1. 데이터 로드 : .json 파일 읽기

2. 데이터 전처리

tokenizer, CountVectorization 으로 단어를 벡터로 변환, **nltk POS tagging**

3. 모델정의

keras로 LSTM를 만들기

4. 결과 확인

프로젝트 진행상황

1. NSMC 데이터 셋 binary classification

데이터 로드 완료, 데이터 전처리 작업 진행 중

2. Friends 데이터셋 multi-class classification

데이터 로드 및 데이터 전처리 작업 진행 중

구현 마일스톤

- 6월 2일 ~ 6월 6일 데이터 전처리 완료
- 6월 7일 ~ 6월 19일 모델 구현 및 학습 완료, 결과 분석
- 6월 20일 ~ 6월 24일 모델 튜닝 및 Final 보고서 작성

팀원 역할 배분

- 한 윤 : 팀장. NSMC 를 이용한 Binary classification 모델 구현, 발표
- 김 건희 : NSMC 를 이용한 Binary classification 모델 구현
- 임 특 : Friends 데이터셋을 이용한 Multi-class classification 모델 구현

감사합니다