**인공지능 과제 리포트**

과제 제목 : 리뷰 감성분석 기반

홍대 맛집 추천 서비스

학번 : B811027

이름 : 김성겸

1. 과제 개요

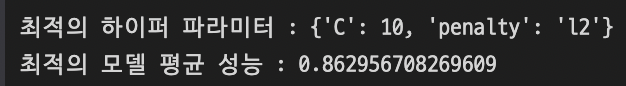
* 주제 : 식당 리뷰 댓글의 감성 분석을 통한 맛집 추천 서비스
* 웹 크롤링 및 여러가지 자연어 처리 모델을 사용하여 성능 비교 후 모델 활용

1. 구현 환경

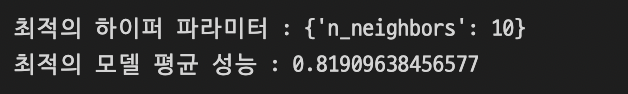
* vscode : 웹 크롤링 및 데이터 정제
* colab : 모델 구현 및 모델 학습

1. 알고리즘에 대한 설명

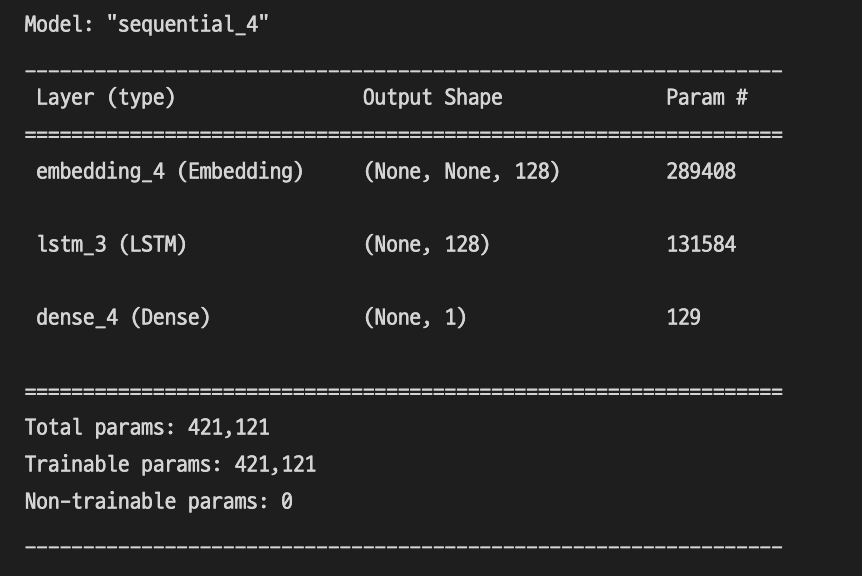
* Logistic Regression



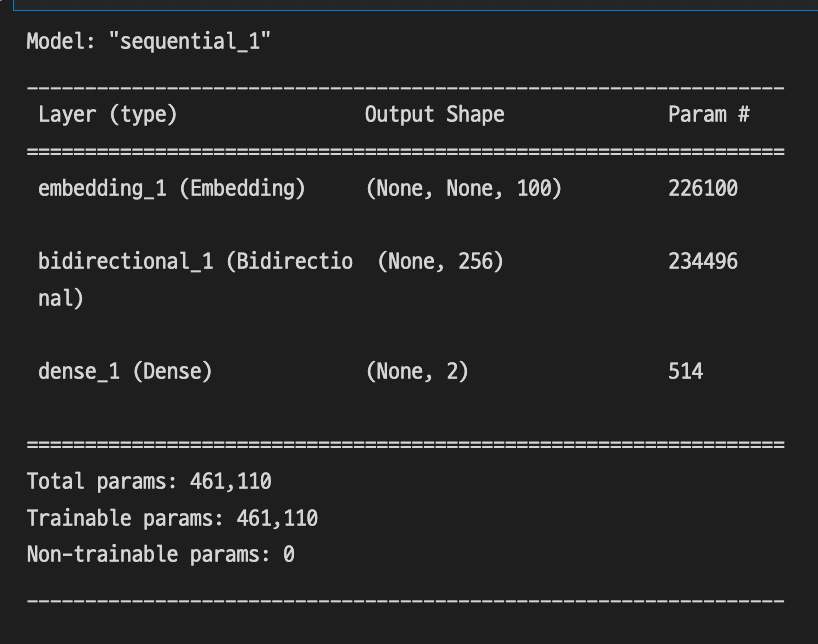
* 그리드 서치 결과 c는 10, panalty는 'l2'일때 최대성능
* KNN



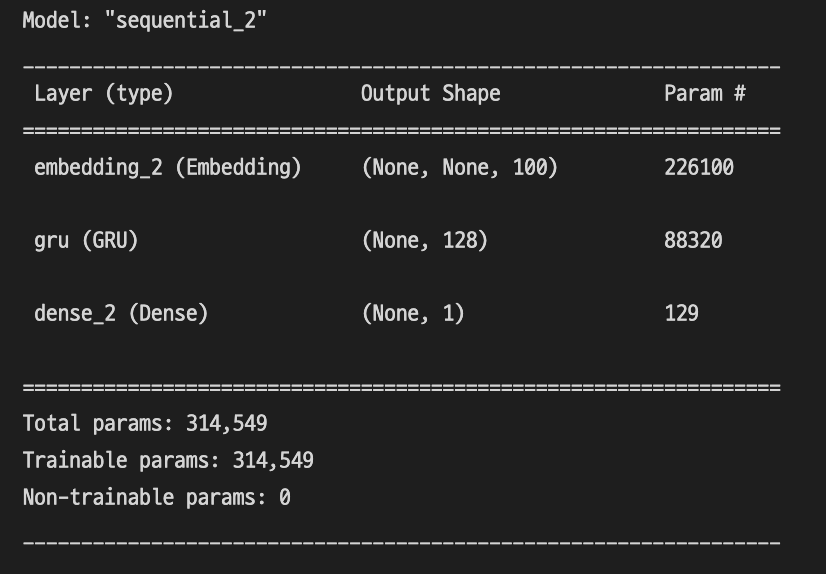
* 그리스 서치 결과 n=10일때 최대성능
* LSTM



* Drop out 0.2 적용
* Bidirectional LSTM

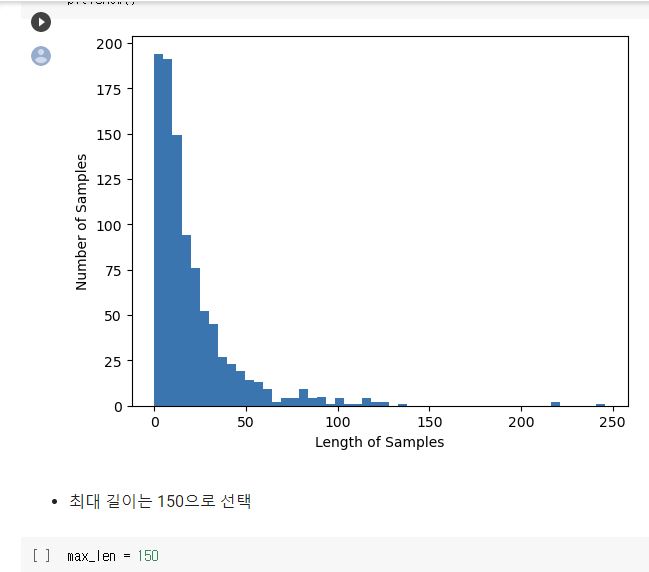


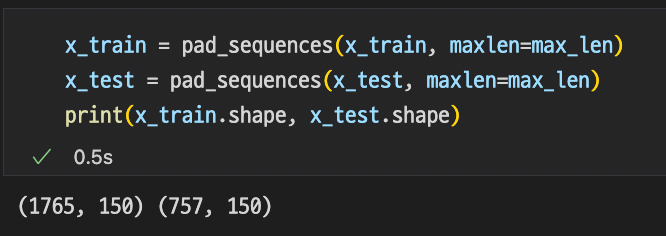
* GRU



* KoBERT

1. 데이터에 대한 설명
   1. Input Feature

* 홍대 인근 식당의 리뷰 댓글(data type : str)
* 차수 : 150
* 



* 패딩 마친 최종 데이터 결과
  1. Target Output

Binary cross entropy : 긍정 = 1 , 부정 = 0

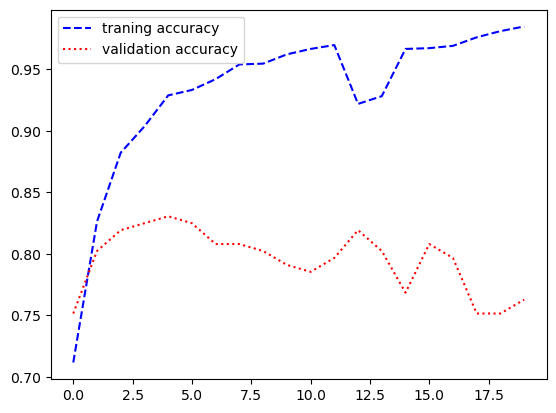
Sparse categorical cross entropy : 긍정 = 2, 중립 = 1, 부정 = 0

1. 소스코드에 대한 설명

* 주석 참조

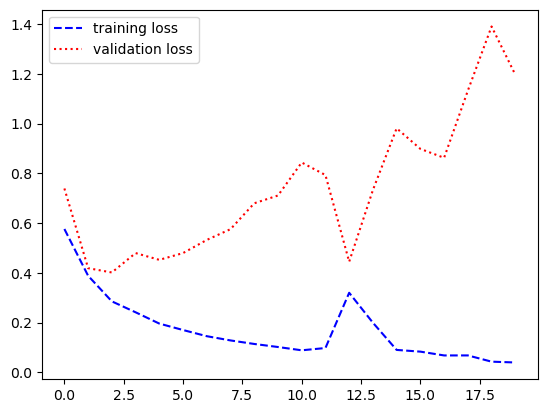
1. 학습 과정에 대한 설명

* 시도해본 여러 모델 중 LSTM의 학습 과정을 보면

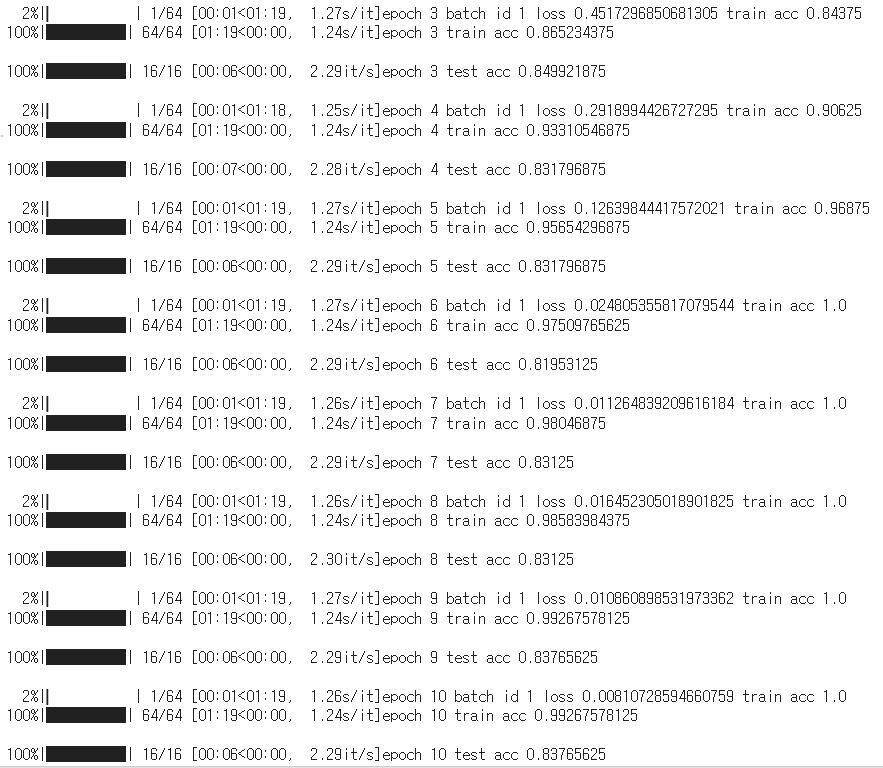


* 검증셋에 대한 loss는 12번째 에포크 정도에서 오버피팅이 발생하는 현상을 볼 수 있다.

1. 결과 및 분석



* Costfunction값 추이 : training loss값은 계속해서 줄어들고 있으나 validation loss 값은 에폭 5~10 이후로 급격히 증가하는 것으로 보아 최적의 학습 에폭은 5에서 10 사이로 보임



* 결과 : 여러 모델을 비교해본 결과 KoBERT 모델의 성능이 가장 좋은 것으로 확인이 되었다. 하지만 KoBERT모델 내에서 데이터 수를 2배 이상 늘렸음에도 정확도는 향상되지 않았다.
* 분석 : 먼저 다른 모델에 비해 KoBERT의 성능이 뛰어났던 이유를 알아본 결과 먼저 기존의 단방향으로 처리하는 자연어 처리 모델과 달리 BERT모델은 양방향으로 보기 때문에 다른 모델에 비해 정확도가 높아진 것 같고, 특히 한국어에 특화된 KoBERT를 사용한 것도 성능 차이에 영향을 끼친 것 같다

1. 실행 매뉴얼

* 홍대\_맛집\_리뷰\_감정분석\_kobert.ipynb
* Colab 환경에서 전체 실행 권장(오류시 프리미엄GPU및 고용량 RAM 사용)
* 홍대\_맛집\_리뷰\_감정분석\_knn\_logistic.ipynb

1. 코랩환경에서 실행, 맨윗줄 주석해제

2. JAVA\_HOME 변수에 사용자 컴퓨터의 jdk경로 설정

* 홍대\_맛집\_리뷰\_감정분석\_lstm.ipynb
* Knn-logistic과 동일
* 데이터 ver1,ver2,ver5는 1,0 라벨링(binary,or sparse사용)
* 데이터 ver3는 2,1,0라벨링 (sparse)
* 카카오맵스크래퍼는 맨위 필요패키지 모두 다운로드 후 실행
* 인터넷환경 좋은 곳에서 사용하길 추천
* (모델 할당후 get\_data 연속으로 실행시 오류가능성 있으니 천천히 하나씩 실행 추천)