

# F1 score

transformer · 2024. 7. 12. 15:29

출처

<https://product.kyobobook.co.kr/detail/S000200330771>

f1 score 는 precision과 recall의 조화평균

F1 score 는 라벨이 불균형할때 정확도 보다 우선시 되는 metric

precision은 모델이 정답이라고 예측하는것중에 실제 맞춘거의 비율

recall은 데이터의 정답중에서 모델이 몇개나 찾아냈냐는 비율

label: 0 0 1 0 1 0 0 0 0 1 신 양성 = 1일때

pred : 0 0 0 0 0 0 0 0 0 1 acc: 80% -> 1이 모델이 맞다고 예측한것인데 실제 정답인 label  
에서 맞춤 -> label / pred -> 1/1

precision - 1이 모델이 맞다고 예측한것인데 실제 정답인 label에서 맞춤 -> label / pred ->  
1/1 모델이 예측한것중에 실제 맞춘것만 봄

recall - 정답(label)은 3개이고 모델이 1개 맞춰서 recall 은 1/3

f1 score =  $1/2 = 0.5$

precision과 recall 은 0과 1 사이의 값이고

f1 score는 precision과 recall의 조화평균이라서 0과 1사이값

분자 2는 precision 1개 recall 1개라서 2개의 평균이라 2

$$f1 = \frac{2}{\frac{1}{p} + \frac{1}{r}}$$

$$f1 = \frac{2}{\frac{1}{p} + \frac{1}{r}}$$

macro f1 label의 빈도는 모든 label의 평균되서 자주안나오는것도 성능에 반영됨

micro f1 자주나오는 label의 score가 많이 반영

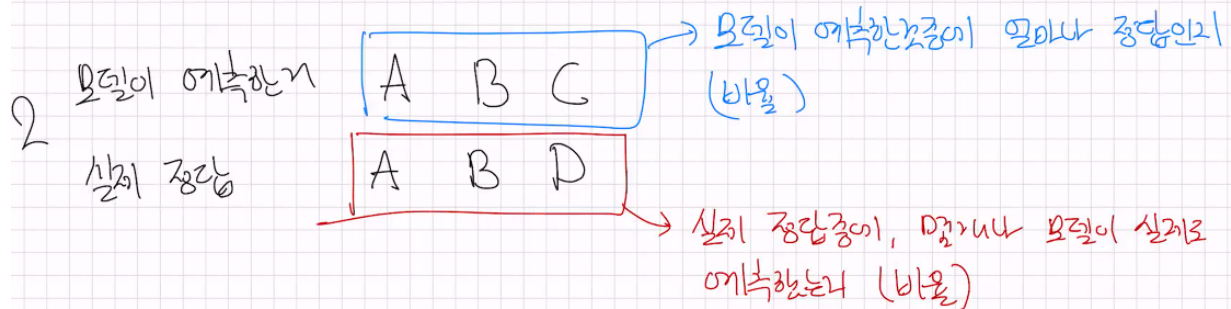
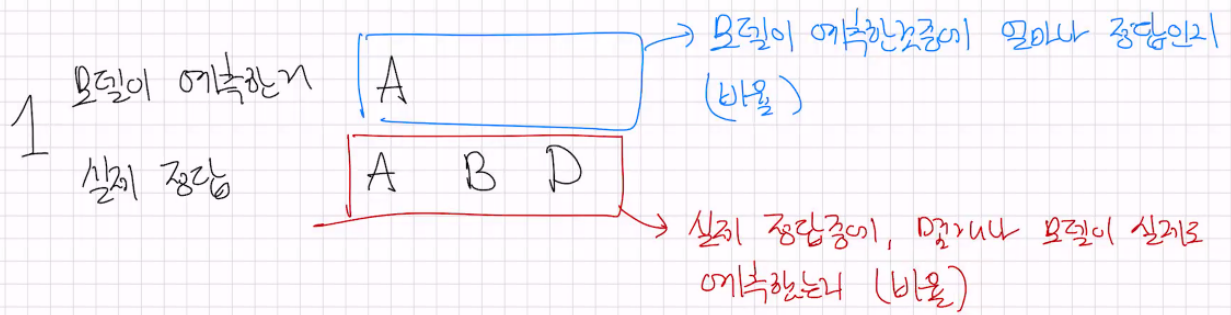
precision : 예측 4, 예측한것중 맞은게 3개 3/4

-> 모델이 예측한것중 얼마나 정밀한가(몇개나 맞았나)

recall: 정답이 7개 모델이 맞춘게 3개 3/7

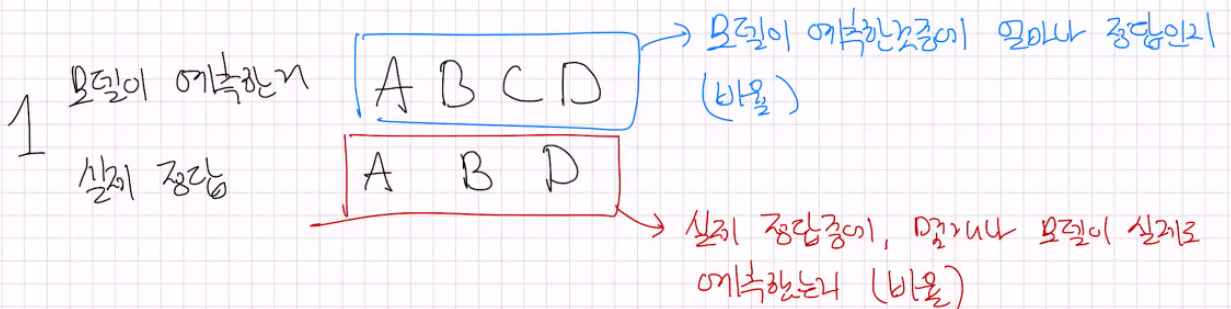
-> 실제정답중 모델이 몇개나 예측했는지(정답에서 모델이 맞춘것)

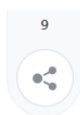
precision만 쓰면 1개만 예측해서 다맞으면 100% 맞음 -> errorrecall 도 모든 경우를 다예측하면 --> 이것도 100% 만점 --> error그래서 이 두개를 잘 조화한것이 F1 score



조화평균

$$f1 = \frac{2}{\frac{1}{p} + \frac{1}{r}}$$



		실제 정답	
		True	False
분류 결과	True	True Positive	False Positive
	False	False Negative	True Negative

출처 <https://velog.io/@jadon/F1-score%EB%9E%80>

f1

label: 0 0 1 0 1 0 0 0 0 1 신 양성 = 1일때

pred : 0 0 0 0 0 0 0 0 0 1 acc: 80% -> 1이 모델이 맞다고 예측한것인데 실제 정답인 label에서 맞춤 -> label / pred -> 1/1

precision - 1이 모델이 맞다고 예측한것인데 실제 정답인 label에서 맞춤 -> label / pred -> 1/1

recall - 정답(label)은 3개이고 모델이 1개 맞춰서 recall 은 1/3

f1 score =  $1/2 = 0.5$

precision과 recall 은 0과 1 사이의 값이고

f1 score는 precision과 recall의 조화평균이라서 0과 1사이값

$$f1 = \frac{2}{\frac{1}{p} + \frac{1}{r}}$$

precision은 모델이 정답(true)이라고 예측하는것중에 실제 맞춘거의 비율

recall은 데이터의 정답중에서 모델이 몇개나 찾아냈냐는 비율

precision:  $TP / (TP + FP) = 1$

TP: 양성 1 했는데 맞은거 하나 1

FP: 양성 1 했는데 틀린거 없음 0

recall:  $TP / (TP + FN) = 1/3$

f1:  $2 / (1 + 3) = 1/2 = 50\%$

predi: 0 0 1 0 0 / 0 0 1 0 1 의사 false negative

의사가 음성(0)이라는데 틀렸고

true negative

음성

의사가 0이라는데 맞음 9

negative냐 positive 음성이냐 양성이나를 먼저보고 true (맞았냐) false(틀렸냐)판단

acc:  $9/10 \Rightarrow 90\%$

모델이 true -> positive

negative

의사가 검사해서 양성이라고 진단한게 오진확률가능성

양성, 음성 positive, negative

신이 있다면 true인지 false인지

신이 봤을때 의사진단이 맞으면 true, 틀리면 false

모델이 1이라고 했는데 ->맞았을때 true positive

모델이 false라 했는데 -> false가 맞았을때 false negative

모델이 false라했는데 -> 틀렸을때 false negative

True Positive: 2 p 양성 1 신이 1 2

False Positive: 1 positive: model이 1

negative: model이 0

p 양성 1 했는데 틀림 1

False Negative: 1 음성 틀림 0 1

True Negative: 6 음성 이라했는데 0 -> 맞은거 6

precision:  $TP / (TP+FP)$  의사가 양성이라고 예측한것중 몇퍼센트 실제양성이걸 맞췄냐

recall:  $TP / (TP+FN)$  신이알려줘서 실제 양성인것중에 의사가 몇퍼센트 찾아냈냐

precision:  $2 / (2+1)$

recall:  $2 / (2+ 1)$

f1 = (precision과 recall의 조화평균) -> 조화평균: 평균값중 하나

(일반적인게 산술평균)

$= 2 / ((1/precision) + (1/recall))$

$= 2 / (3/2 + 3/2)$

$= 2 / 3$

♡ 공감



0000

구독하기

**transformer** 카테고리의 다른 글

[Chapter\\_04\\_요약\\_multilingual-ner.ipynb](#) (0)

2024.07.26

[Chapter\\_03\\_요약\\_transformer 파해치기](#) <https://nlpinkorean.github.io/illustrated-transformer...> (0)

2024.07.12

[Chapter\\_02\\_요약 classification](#) (0)

2024.07.06

[Chapter04\\_요약\\_Ner\\_글내에서 어떤단어가 제품,사람, 장소인지 알려줄회사나 조직인지 분류 하는...](#) (0)

2024.07.05

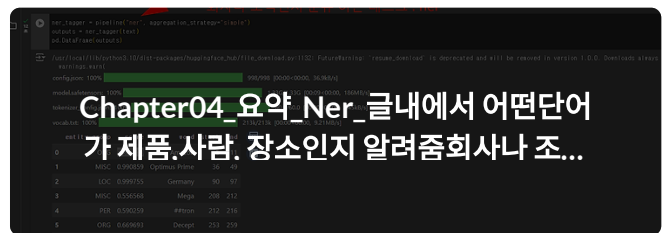
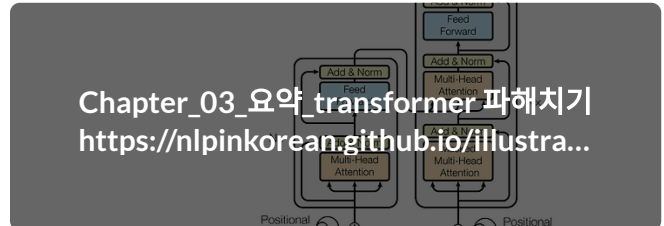
[introduction\\_transformer\\_요약](#) (0)

2024.06.28

## 관련글

[관련글 더보기](#)

Chapter_04_요약_multilingual_ner.ipynb							
RoBERTa							
with BOOKS + WIKI	16GB	8K	100K	93.6/87.3	89.0	95.3	
+ additional data (§3.2)	160GB	8K	100K	94.0/87.7	89.3	95.6	
+ pretrain longer	160GB	8K	300K	94.4/88.7	90.0	96.1	
	160GB	8K	300K	94.4/88.7	90.0	96.1	
BERT <sub>LARGE</sub>							
with BOOKS + WIKI	16GB	8K	100K	90.9/81.8	86.6	93.7	
XLNet <sub>LARGE</sub>							
with BOOKS + WIKI	13GB	256	1M	94.0/87.8	88.4	94.4	
+ additional data	126GB	2K	500K	94.5/88.8	89.8	95.6	



## 자연어(NLP)

네이쳐2024 님의 블로그입니다.

구독하기 +



댓글 0



이름

비밀번호

내용을 입력하세요.

