이진분류

202055518 김병현

데이터 불균형과 평가 착시의 원인

- 데이터셋의 불균형: 모델을 학습하거나 평가할 때, 특정 클래스에 속하는 데이터가 다른 클래스에 비해 압도적으로 많을 때, 모델이 주로 빈도 높은 경우만 예측해도 성능이 높아 보일 수 있다. 예를 들어, 긍정 리뷰가 90%, 부정 리뷰가 10%인 데이터셋에서 모델이 긍정 리뷰만 예측해도 90%의 정확도를 얻을 수 있다.
- 평가 착시: 모델이 실제로는 다양한 클래스에 대한 예측을 잘하지 못하더라도, 전체적인 성능 평가 지표(예: 정확도)가 높은 경우, 실제 모델의 성능을 과대평가할 수 있다. 이렇게 모델이 성능이 좋아 보이는 착시가 발생한다.

착시가 없는 평가 방법

불균형한 데이터로 정확도 처럼 단순한 지표를 사용하면 문제를 감지하기가 어렵다. 착시 없는 평가를 위해서는 데이터의 불균형을 고려한 평가 지표를 사용해야 한다.

1. 정확도 **(Accuracy)**

```
accuracy = safe_div(tp+tn, tp+tn+fp+fn)
```

정확도는 모델이 올바르게 예측한 모든 샘플(tp + tn)의 비율을 나타낸다. 그러나 데이터셋이 불균형하면 높은 빈도 클래스에 대해 항상 맞추는 모델도 높은 정확도를 얻게 되어 착시가 발생할 수 있다.

2. 정밀도 **(Precision)**

```
precision = safe_div(tp, tp+fp)
```

정밀도는 모델이 긍정적으로 예측한 것 중에서 실제로 긍정인 샘플의 비율이다. 즉, 잘못된 긍정 예측을 줄이기 위해 중요한 지표이다. 불균형 데이터셋에서도 높은 정밀도는 모델이 과도하게 한쪽으로 치우치지 않았음을 보장할 수 있다. 그러나 정밀도만으로는 재현율을 함께 고려하지 않으면 특정 클래스에 대한 성능을 모두 반영하지 못할 수 있다.

3. 재현율 **(Recall)**

```
recall = safe_div(tp, tp+fn)
```

재현율은 실제로 긍정인 샘플 중에서 모델이 얼마나 많이 맞췄는지를 의미한다. 데이터 불균형에서 중요한 역할을 한다. 만약 재현율이 매우 낮다면, 모델이 소수 클래스를 거의 예측하지 못한다는 신호이다. 이로 인해 데이터 불균형에 대한 착시가 줄어들고, 모델이 얼마나 잘 예측하는지를 더욱 명확하게 파악할 수 있다.

4. F1 점수 (F1 Score)

```
f1 = 2 * safe_div(recall*precision, recall+precision)
```

F1 점수는 정밀도와 재현율의 조화 평균이다. 이는 정밀도와 재현율 간의 균형을 잡아주며, 데이터 불균형에서도 모델 성능을 보다 정확하게 평가할 수 있는 지표이다. F1 점수가 높을수록 모델이 한쪽으로 치우치지 않고 전반적으로 균형 잡힌 성능을 발휘하고 있음을 나타낸다.

데이터 조정 했을 때, 안 했을 때 결과의 차이

```
pulsar_exec()
Final Test: final result = 0.967,0.976,0.649,0.780

pulsar_exec(adjust_ratio=True)
Final Test: final result = 0.915,0.909,0.919,0.914
```

adjust_rate = False (미조정 상태)

결과: `0.967, 0.976, 0.649, 0.780`

- 정확도 **(Accuracy):** 0.967
- 정밀도 **(Precision):** 0.976
- 재현율 **(Recall):** 0.649
- **F1** 점수 **(F1 Score):** 0.780

이 경우 정확도와 정밀도는 매우 높지만, 재현율은 낮다. 이는 모델이 대부분의 예측에서 주로 다수 클래스에 집중하고, 소수 클래스는 거의 예측하지 못했음을 나타낸다. 데이터셋의 불균형으로 인해 모델이 특정 클래스만 잘 예측하는 경향이 있으며, 그 결과 정밀도는 높으나 재현율이 낮게 나타난다.

adjust_rate = True (조정 상태)

결과: `0.915, 0.909, 0.919, 0.914`

- 정확도 **(Accuracy):** 0.915
- 정밀도 **(Precision):** 0.909
- 재현율 **(Recall):** 0.919
- **F1** 점수 **(F1 Score):** 0.914

이 경우는 정밀도와 재현율이 더 균형 있게 나타난다. 비록 정확도는 조금 낮아졌지만, 정밀도와 재현율이 거의 비슷하게 나오면서 모델이 모든 클래스에 대해 고르게 예측하고 있다는 것을 의미한다. 이는 데이터 불균형이 조정되었음을 보여준다. **F1** 점수도 이전보다 개선되어, 모델이 클래스 간 균형 잡힌 성능을 발휘하고 있음을 알 수 있다.

epoch_count, mb_size, learning_rate 조정하며**,** 성능 측정

```
LEARNING_RATE = 0.001
pulsar_exec(epoch_count=20, mb_size=10, report=1, adjust_ratio=True)
Final Test: final result = 0.928,0.948,0.902,0.925
```

```
LEARNING_RATE = 0.0001 로 수정
pulsar_exec(epoch_count=30, mb_size=16, report=1, adjust_ratio=True)
Final Test: final result = 0.930,0.959,0.897,0.927

pulsar_exec(epoch_count=60, mb_size=16, report=1, adjust_ratio=True)
Final Test: final result = 0.925,0.944,0.901,0.922
```

여러 지표 점수가 오르며 성능이 조금씩 오르다가 점차 수렴하게 된다.