

Week1: Transformer Report

Joonho Kim

March 17, 2025

1 Abstract

Transformer 모델을 학습하면서, WandB를 통해 하이퍼파라미터 조합이 모델 성능에 미치는 영향을 분석한 결과를 정리해보았다. `learning_rate`, `batch_size`, `d_model`을 조정하며 최적의 조합을 찾는 것이 목표였다.

2 Dashboard Analysis

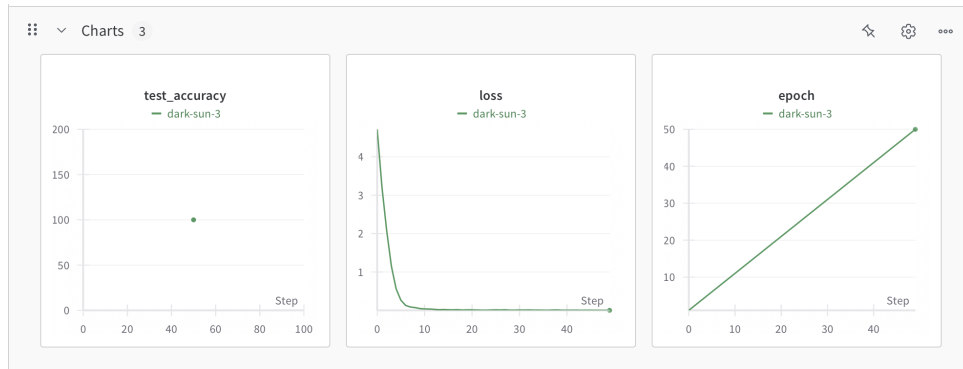


Figure 1: WandB 대시보드: 손실값(Loss) 및 테스트 정확도 변화

2.1 Test Accuracy

- 최종 정확도가 100%로 기록됨.
- 이는 모델이 학습 데이터의 패턴을 효과적으로 학습했음을 의미한다.
- 그러나 오버피팅 가능성을 고려하여 추가적인 검증이 필요할 것 같다.

2.2 Loss

- 초기 손실값이 4.5에서 빠르게 감소하며, 10 Epoch 이후 0 근처로 수렴함.
- 이는 설정된 학습률(0.0001)이 적절했음을 시사함.
- 지나치게 높은 학습률은 손실값이 불안정하게 변동할 위험이 있으므로, 현재 설정이 안정적인 학습을 지원하는 것으로 판단됨.

2.3 Epoch

- 50 Epoch 동안 지속적인 손실 감소가 이루어짐.
- 하지만 10 Epoch 이후 손실값이 거의 일정해지는 것으로 보아, 조기 종료(Early Stopping)를 고려할 수 있음.

3 Hyperparameter Settings and Effects

- Table1 참조

하이퍼파라미터	설정값	선택 이유 및 효과
learning_rate	0.0001	손실값이 안정적으로 감소하며, 너무 빠르게 수렴하지 않음.
batch_size	32	16보다 일반화 성능이 좋고, 64보다 학습 속도가 빠름.
d_model	32	메모리 사용량과 성능의 균형을 맞추기 위해 선택.
num_epochs	50	손실값이 10 Epoch 이후 거의 변화가 없어, Early Stopping 고려 가능.

Table 1: Hyperparameter 설정 및 효과

4 Conclusion

- 현재 설정된 하이퍼파라미터는 손실값 감소 및 정확도 면에서 효과적이었다.
- 다만, 학습이 너무 빠르게 수렴하는 경향이 있으므로, Early Stopping을 적용하여 불필요한 Epoch 낭비를 줄일 필요가 있을 것 같다. `d_model=64` 또는 `learning_rate=0.0005`로 설정하여 일반화 성능 개선을 시도해 볼 수 있을 것 같다.
- 이번 실험을 통해 안정적인 학습률과 손실 감소 패턴을 확인했으니, 다음 번엔 모델의 일반화 성능 향상을 중점적으로 고려해 보자.