

## <AI 보안 과제#2>

### <1. 과제 개요 및 목표>

#### <개요>

RNN모델을 활용한 텍스트 분류이며 IMDB Dataset을 이용하여 텍스트 기반 감정분류입니다. 제공된 실습 코드를 개선하거나 혹은 모델링을 설계하여 RNN 구조에 대한 이해 및 정확도 90%이상 올리겠습니다.

#### <목표>

단순한 RNN이 아닌 다양한 모델링을 통해 모델을 향상시켜 정확도를 90%이상 올리겠습니다. 저는 BERT+GRU 결합한 모델로 설계하여 정확도를 90%이상 나오도록 목표하였습니다.

### <2. 전처리 과정 설명>

IMDB 데이터셋을 활용한 감정 분석 모델 구축을 위해 텍스트 데이터를 전처리 과정을 하였습니다. HTML 태그와 알파벳 이외의 특수 문자를 제거하여 불필요한 정보를 삭제하였습니다. 모든 텍스트를 소문자로 변환해 일관성을 유지하고, 공백을 기준으로 단어를 분리한 후, 재조합했습니다. 사전 학습된 BERT 토큰라이저를 사용해 텍스트를 토큰화하고, 최대 길이를 200으로 설정해 긴 문장은 잘라내고 짧은 문장은 패딩했습니다. 레이블은 긍정(1)과 부정(0)으로 인코딩하여 정수 형태로 변환했습니다. 데이터셋을 8:2 비율로 학습용과 테스트용으로 분할했습니다. 이러한 전처리 과정은 모델의 입력을 정제하고 일관성을 높여 성능 향상에 도움되도록 하였습니다.

### <3. 기존 모델 설명(RNN)>

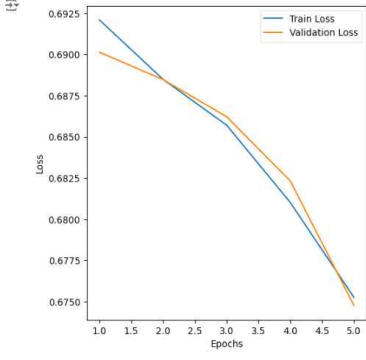
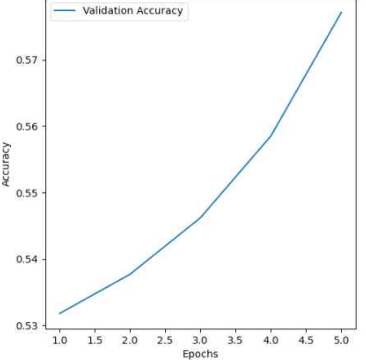
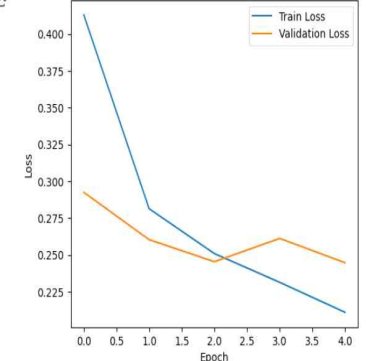
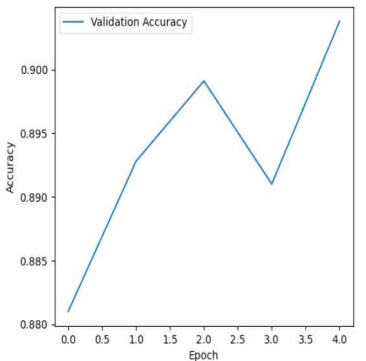
제공된 기존 모델은 기본적인 RNN모델이기 때문에, 순차 데이터(시계열 데이터, 자연어 등)의 패턴을 인식하도록 설계된 모델입니다. 이 모델은 Self-Connection을 통해 이전 입력에 대한 메모리를 유지할 수 있도록 구성되어 있으며, 각 입력을 순차적으로 처리하기 때문에 문맥과 순차 정보가 중요한 작업에 특히 유용합니다. 구조적으로 RNN은 노드 간의 연결이 순환적이며, 이는 이전 데이터의 결과를 학습 및 전파하는 과정에서 영향을 미치도록 합니다. 하지만 RNN은 학습 과정에서 Vanishing Gradient 문제에 직면할 수 있습니다. 이 문제는 기울기가 점점 작아져서 학습이 더 이상 진행되지 않거나, Chain Rule에 의해 가중치 업데이트 값이 0으로 수렴하게 되는 현상입니다. 또한, 이러한 한계로 인해 장기 의존 문제를 해결하는 데 어려움이 있습니다. 단순한 모델을 극복하기 위해 LSTM이나 GRU 모델이 개발되었고 정보의 흐름을 더 효과적으로 제어하기 때문에, Vanishing Gradient 문제를 완화하고 장기 의존성을 처리할 수 있습니다.

### <4. 수정된 코드 설계 및 이유>

기존 모델과 달리, BERT + GRU를 결합한 모델을 설계했습니다. 선택한 이유는 이 CNN + LSTM, GRU + LSTM, LSTM, GRU등 여러 모델을 설계하였지만, BERT + GRU 모델이 가장 높은 성능이 나와 이 모델을 선택하였습니다. 모델은 BERT의 임베딩 출력을 입력으로 받아, 양방향 GRU를 통해 시퀀스의 시간적 의존성을 학습하도록 구현하였습니다. 이 모델의 구성은 사전 학습된 BERT 모델을 활용하여 각 리뷰를 임베딩 벡터로 변환합니다. BERT의 모든 파라미터는 학습되지 않도록 고정하여 훈련 시간을 단축하고, 모델의 안정성을 높였습니다. 이후 BERT의 임베딩 출력은 양방향 GRU에 입력됩니다. GRU는 시퀀스의 앞뒤 맥락을 모두 고려하여 더 풍부한 의미를 학습할 수 있도록 설계되었습니다. GRU의 최종 은닉 상태는 양방향의 두 방향을 결합한 후, 드롭아웃을 적용해 과적합을 방지합니다. 마지막으로, 결합된 은닉 상태는 fully connected layer로 전달되어 감정(긍정 또는 부정)을 예측합니다. 이 모델은 BERT의 강력한 텍스트 이해 능력과 RNN 계열인 GRU의 시간적 의존성 학습 능력을 결합하여, 감정 분석과 같은 자연어 처리 작업에서 우수한 성능을 기대할 수 있었습니다.

### <5. 학습 과정 및 성능 비교>

모델 학습은 Adam 옵티마이저 및 BCEWithLogitsLoss 손실 함수를 이용하여 진행했습니다. 학습률은 1e-3로

기존 RNN-Loss	기존 RNN-Accuracy	기존 RNN-Accuracy
		Validation Accuracy: 53.18% Validation Accuracy: 53.77% Validation Accuracy: 54.62% Validation Accuracy: 55.85% Validation Accuracy: 57.71%
		Validation Accuracy: 88.10% Validation Accuracy: 89.28% Validation Accuracy: 89.91% Validation Accuracy: 89.10% Validation Accuracy: 90.38%
개선한 GRU-Loss	개선한 GRU-Accuracy	개선한 GRU-Accuracy

설정하고, ReduceLROnPlateau 스케줄러를 적용해 검증 손실이 개선되지 않을 경우 학습률을 10배 감소시켰습니다. 배치 사이즈는 128, 학습 에폭 최대 5회로 설정했습니다. 매 에폭마다 학습 손실과 검증 손실, 검증 정확도를 측정하고 기록했습니다. 학습 과정에서 Early Stopping을 적용하여, 5번의 에폭 동안 성능이 개선되지 않을 경우, 학습을 조기 종료하도록 설정하였습니다. 모델의 예측은 시그모이드 함수를 통해 0~1 사이의 확률로 변환된 후, 0.5를 기준으로 긍정 또는 부정으로 분류했습니다. 학습률 스케줄러는 손실 개선이 멈춘 시점마다 학습률을 줄여 최적화를 미세하게 진행했습니다. 학습 초반에는 손실이 빠르게 감소했지만, 후반부에는 성능 개선이 미세해졌습니다. BERT 파라미터를 고정한 덕분에 학습이 안정적이었고, GRU와 fully connected layer만 학습되어 훈련 시간이 단축되었습니다. 1에폭부터 높은 성능을 보이다가 마지막 5에폭에서 정확도 90%를 넘어 성공적으로 목표한 수치에 도달 하였습니다.

## <6. 결과 요약 및 분석>

### <결과 분석>

BERT-GRU 모델은 IMDB 감성 분석에서 85~90%의 검증 정확도를 달성하며 안정적인 성능을 보였습니다. 양방향 GRU는 시퀀스의 앞뒤 맥락을 모두 학습해 긴 리뷰에서도 높은 예측 성능을 발휘했습니다. 학습 초반에는 손실이 빠르게 감소했지만, 후반부에는 개선 폭이 줄어들었습니다. BERT 파라미터를 고정한 덕분에 모델은 빠르고 안정적으로 학습되었으며, 드롭아웃과 학습률 스케줄링을 통해 과적합을 방지할 수 있었습니다. Early Stopping이 적절히 작동하여 성능이 정체된 시점에서 학습을 종료했습니다. 이 모델은 텍스트의 맥락과 순서를 잘 반영하여 긍정/부정 감성 예측에 적합한 성능을 보였습니다.

### <요약>

BERT-GRU 모델은 BERT의 임베딩 생성 능력과 GRU의 시퀀스 학습 능력을 결합해 IMDB 감성 분석에서 90%이상의 정확도를 달성했습니다. 학습률 스케줄링과 Early Stopping을 통해 학습 효율성을 높이고 과적합을 방지했습니다. 이 모델은 긴 텍스트에서도 맥락을 잘 파악해 긍정/부정 감성 예측에 적합한 성능이 나왔습니다.