

1. LSTM 기반 Seq2Seq 모델에서 디코딩할 때 사용하는 Beam Search의 동작 방식

Beam Search는 디코딩 시점마다 가장 확률이 높은 단어 하나만 선택하는 Greedy 방식의 한계를 보완하기 위해 사용된다. 매 타임스텝마다 상위 k개의 후보 시퀀스를 유지하며, 각 후보에 대해 다음 단어의 확률을 예측하고 누적 확률 기준으로 다시 상위 k개를 선택한다. 이렇게 하면 보다 다양한 문장 구조를 고려할 수 있어 더 정확하고 자연스러운 출력이 가능해진다. 하지만 beam size가 너무 크면 계산 비용이 증가할 수 있다.

2. Seq2Seq with LSTM 모델의 한계와 Attention 메커니즘의 보완

기본 Seq2Seq 모델은 입력 시퀀스를 고정된 벡터로 압축하여 디코더가 이를 기반으로 출력을 생성한다. 이 방식은 문장이 길어질수록 정보 손실이 커지고 디코더가 중요한 입력 정보를 놓치기 쉽다. Attention은 디코딩을 할 때 마다 인코더의 모든 hidden state를 참고해 중요한 부분에 집중할 수 있게 하고 정보 손실을 줄인다. 이를 통해 긴 문장에서도 정확한 출력이 가능해졌, Seq2Seq 모델의 성능을 개선시켰다.