

Seq2Seq

| | |
|----------|--------------------------|
| ■ 날짜 | @24/07/2025 → 30/07/2025 |
| ■ Select | 과제 |
| ■ Status | In progress |
| ■ Tags | 2차시 |

1. LSTM 기반 Seq2Seq 모델에서 디코딩할 때 사용하는 Beam Search 동작 방식에 대해서 설명

Beam search는 해당 시점에서 가장 확률이 높은 K개를 골라 유지하고 매 시점마다 가능한 모든 단어로 확장한 뒤 또 상위 K개를 골라 진행하는 방식이다. 여기서 K가 1이면 항상 가장 확률이 높은 출력을 선택하는 greedy decoding 방법과 같다.

빔 서치 과정을 <eos> 토큰을 만날 때까지 반복해서 확률이 높은 K개의 출력을 선택하는 과정을 반복하고 <eos> 토큰을 만난 빔이 K개가 되면 총 K개의 후보 중에서 가장 높은 누적 확률을 가진 빔을 최종적으로 선택한다.

2. Seq2Seq with LSTM 모델은 Attention이 없던 시절 제안된 구조임. 기본 Seq2Seq 모델의 한계와 이후 Attention 메커니즘이 이 한계를 어떻게 보완했는지 설명

Seq2Seq는 문장이 매우 길어져도 정보를 고정된 길이의 C에 정보를 담아야하기 때문에 중요한 정보가 디코더에 잘 전달되지 않고 결국 장기 의존성 문제(long-term dependencies problem)이 발생한다.

Attention은 이러한 장기 의존성 문제를 해결한다. 출력할 때 입력 전체 hidden state를 모두 참조하기 때문에 중요한 정보를 놓치지 않도록하고 중요한 특정 부분에 집중할 수 있도록 한다.

Transformer 구조로 확장되면서 성능이 상승했고 이후엔 이러한 attention 구조가 널리 사용되게 되었다.