

# Sequence to Sequence Learning with Neural Networks

<https://arxiv.org/pdf/1409.3215>

## 0. Introduction

- 자연어 처리(NLP) 및 시퀀스 변환 문제(번역, 요약, 음성 인식 등)에서 고정 길이 입력-출력 구조의 한계 존재
- 전통 RNN 기반 모델은 긴 시퀀스에서 정보 소실 문제가 있음
- 본 논문은 \*\*인코더-디코더 구조(Encoder-Decoder Architecture)\*\*를 제안
- 입력 시퀀스를 가변 길이 벡터로 압축하고 출력 시퀀스를 생성하는 방법을 소개

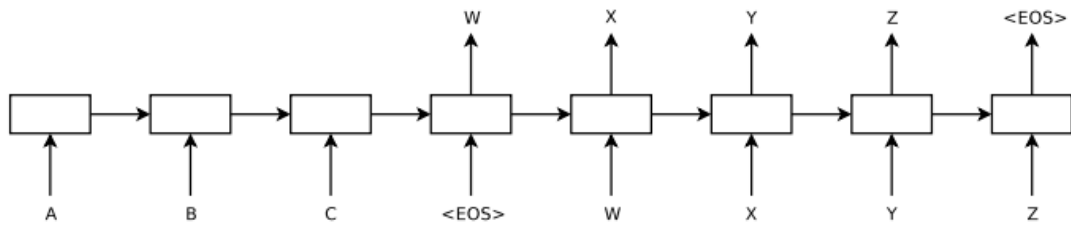
## 1. Overview

- Sequence-to-Sequence (Seq2Seq) 모델 :
  - 입력 시퀀스를 인코더 RNN이 고정 길이 벡터로 인코딩 → 디코더 RNN이 이를 기반으로 출력 시퀀스를 생성
- 주요 특징 :
  - 임의 길이의 입력 및 출력 처리 가능
  - RNN 구조(특히 LSTM)를 활용하여 긴 시퀀스 처리

## 2. Challenges

- 긴 시퀀스에서 중요한 정보 유지의 어려움
- 고정 길이 벡터로 모든 정보를 압축하는 과정에서 정보 손실 발생
- RNN 학습의 어려움: 기울기 소실(vanishing gradient), 학습 속도 문제
- 병렬 처리 한계

### 3. Method



- 인코더(Encoder) : 입력 시퀀스를 하나씩 읽어들이고 내부 상태(hidden state)를 업데이트
- 디코더(Decoder) : 인코더의 최종 상태를 초기 상태로 사용하여 출력 시퀀스를 생성
- LSTM(Long Short-Term Memory) 구조 사용 : 장기 의존성(long-term dependency) 문제 해결
- 학습 : 전체 시퀀스의 정답을 사용하여 최대 우도 추정(maximum likelihood estimation) 방식으로 진행
- Teacher Forcing 기법 적용 : 실제 정답을 디코더 입력으로 사용하여 학습 안정화

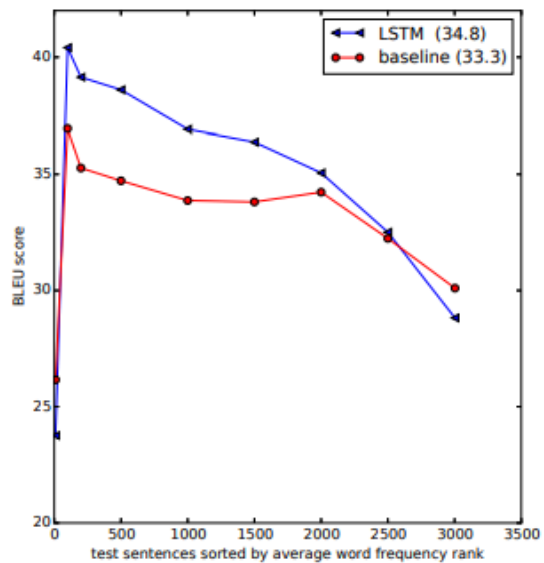
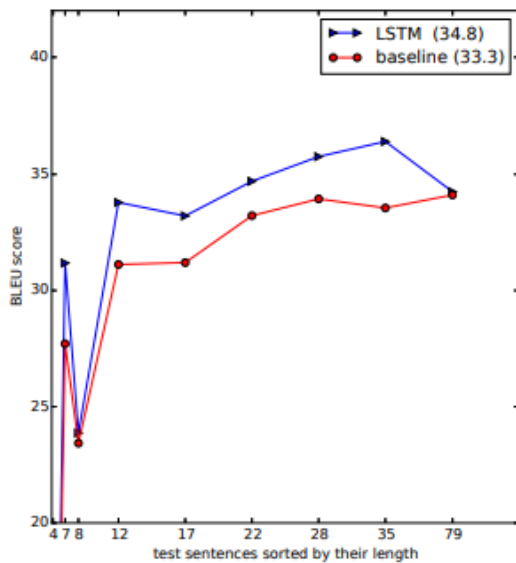
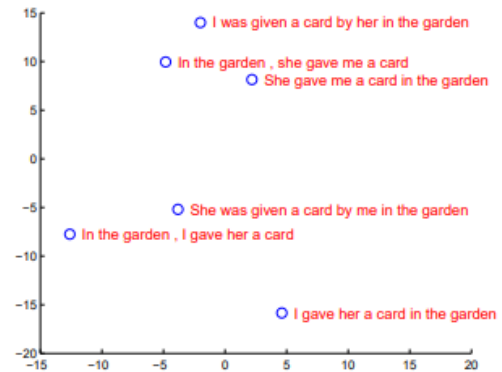
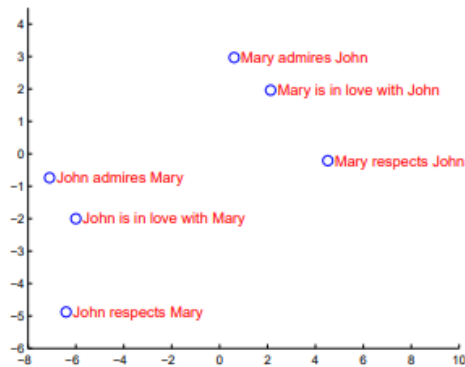
### 4. Experiments

- 데이터셋 : WMT'14 영어-프랑스어 번역 데이터셋 (12M 문장 쌍)
- 평가 : BLEU 점수(Bilingual Evaluation Understudy)
- 비교 baseline : 기존 구문 기반 기계 번역 시스템
- 다양한 네트워크 깊이 및 LSTM 층 수 실험

### 5. Results

Method	test BLEU score (ntst14)
Bahdanau et al. [2]	28.45
Baseline System [29]	33.30
Single forward LSTM, beam size 12	26.17
Single reversed LSTM, beam size 12	30.59
Ensemble of 5 reversed LSTMs, beam size 1	33.00
Ensemble of 2 reversed LSTMs, beam size 12	33.27
Ensemble of 5 reversed LSTMs, beam size 2	34.50
Ensemble of 5 reversed LSTMs, beam size 12	<b>34.81</b>

Method	test BLEU score (ntst14)
Baseline System [29]	33.30
Cho et al. [5]	34.54
Best WMT'14 result [9]	<b>37.0</b>
Rescoring the baseline 1000-best with a single forward LSTM	35.61
Rescoring the baseline 1000-best with a single reversed LSTM	35.85
Rescoring the baseline 1000-best with an ensemble of 5 reversed LSTMs	<b>36.5</b>
Oracle Rescoring of the Baseline 1000-best lists	~45



- Seq2Seq 모델은 기존 구문 기반 기계 번역 시스템 대비 경쟁력 있는 성능 달성
- 특히 깊은(Layered) LSTM 구조가 성능 향상에 효과적임
- 병렬 학습과 모델 크기 확장에 따라 번역 품질이 개선됨
- 긴 시퀀스 번역에서도 의미 유지 성능 확인

## 6. Insight

- Seq2Seq 구조는 NLP뿐 아니라 시퀀스 변환 문제 전반에 적용 가능
- 인코더-디코더 아키텍처는 이후 Attention Mechanism, Transformer 모델 발전의 핵심 기반이 됨
- 장기 의존성을 유지하는 RNN 구조(LSTM, GRU)의 중요성을 보여줌
- 기계 번역, 대화 시스템, 자동 요약 등 다양한 응용에서 혁신적 방법론 제시