

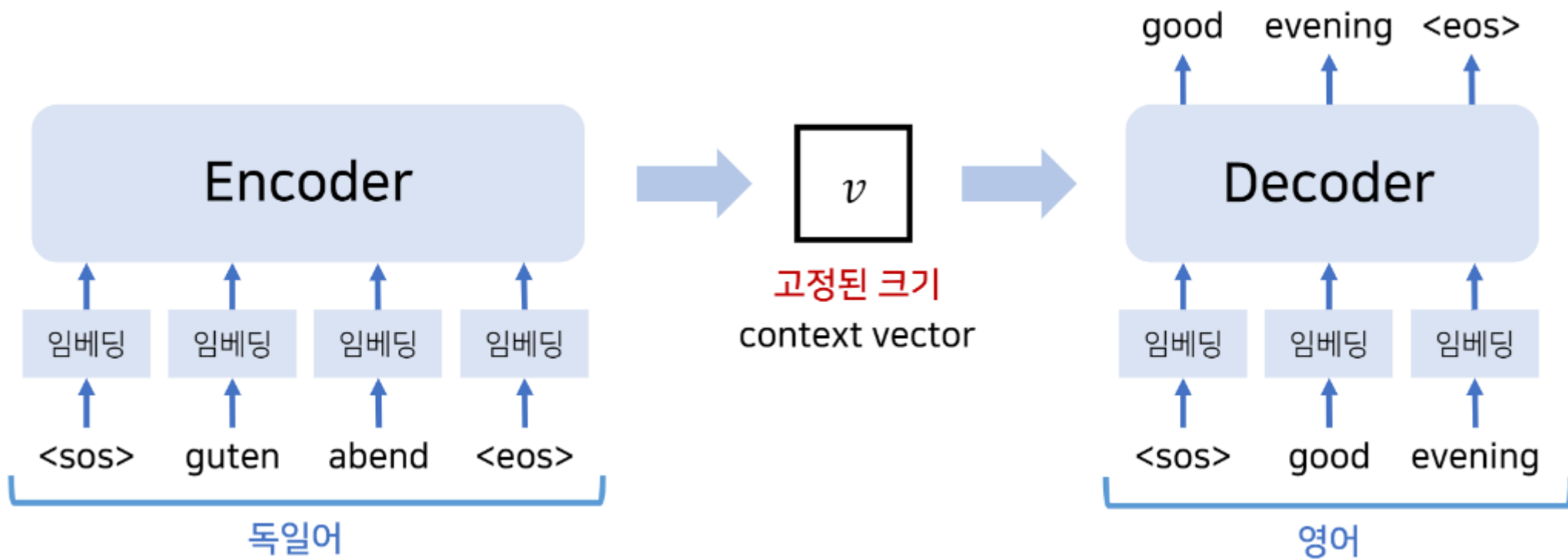
Seq2Seq : Sequence to Sequence Learning with Neural Networks

JUHYEONG LEE, KNUAIR

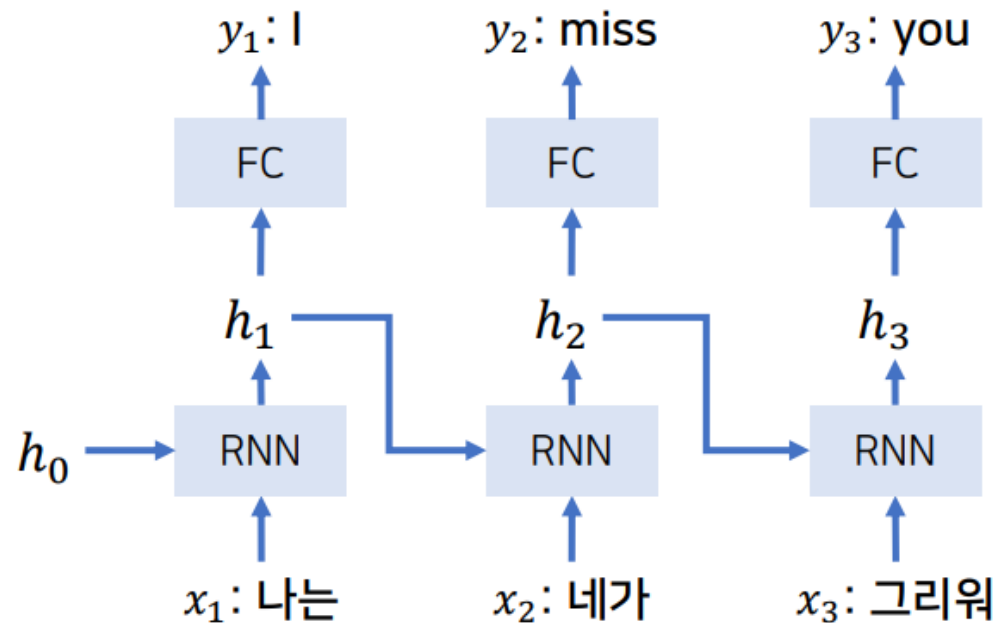
Motivation



기계 번역의 구원자



RNN 기반 기계 번역

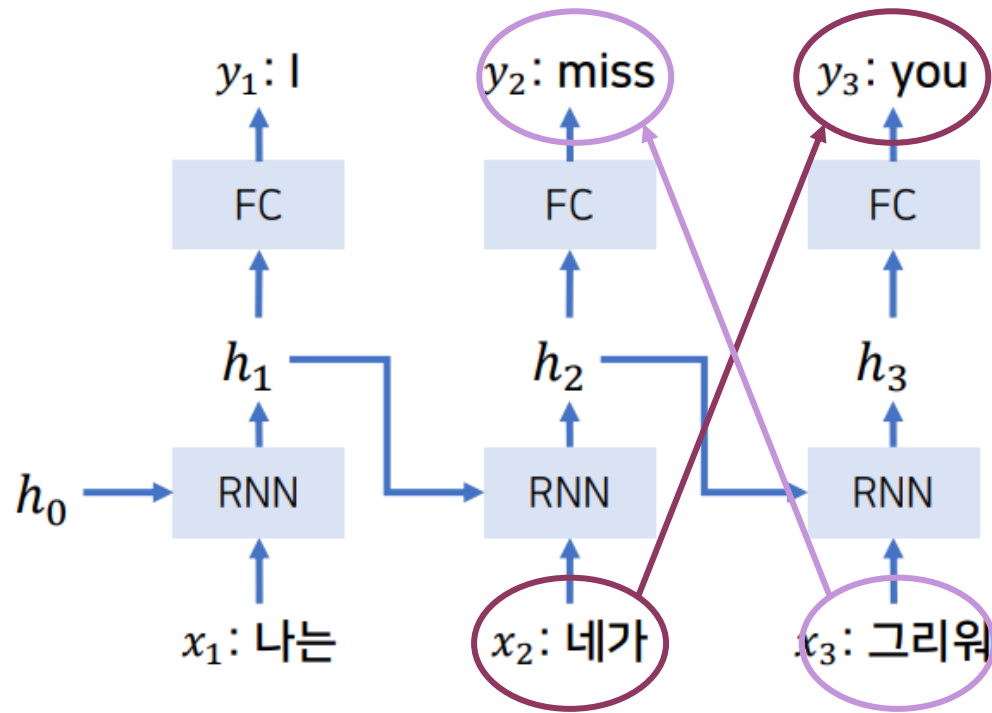


- 입/출력 크기가 같다고 가정
 - Input : (x_1, x_2, \dots, x_n)
 - Output : (y_1, y_2, \dots, y_n)
 - $h_t = \sigma(W^{hx}x_t + W^{hh}h_{t-1})$
 - $y_t = W^{yh}h_t$

RNN 기반 기계 번역 : 문제점

- 가정 : 입/출력 크기가 같다.
 - 출발 언어와 도착 언어의 단어 개수는 같을 수 없다.
 - 애초에 함수의 관계도 성립하지 않음
 - 예) 배 → ① boat ② stomach ③ pear ...
 - 예) book → ① 책 ② 예약하다 ...

RNN 기반 기계 번역 : 문제점

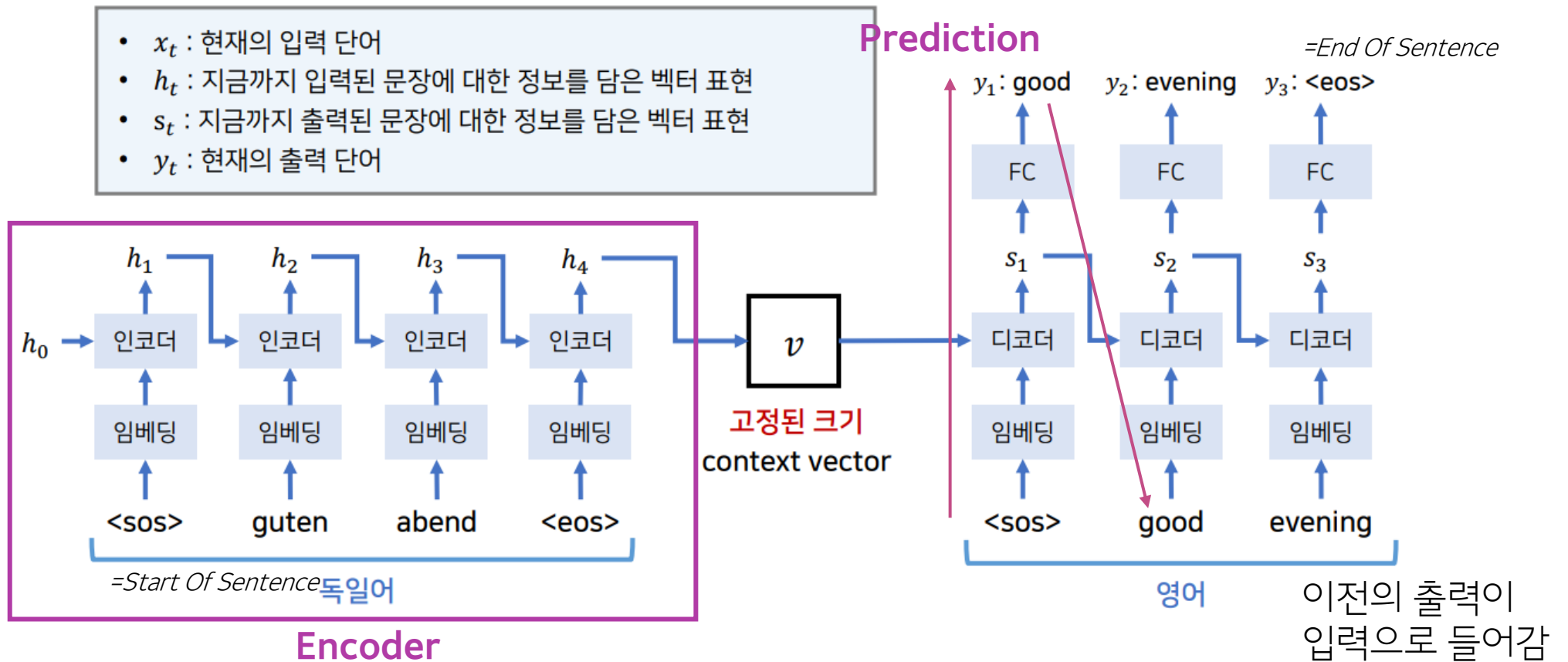


- 언어마다 어순의 차이가 있다.
 - 한국어 : $S + O + V$
 - 영어 : $S + V + O$
- 네가 \rightarrow miss,
그리워 \rightarrow you ???

Seq2Seq가 제시한 해결책

- 그럴 것 같으면 문장 통째로 처리하자
 - Encoder : 고정된 크기의 문맥 벡터(context vector) 추출
 - Encoder의 마지막 hidden state를 문맥 벡터로 활용
 - Decoder : 문맥 벡터로부터 번역 결과 추론
- Encoder와 Decoder는 서로 다른 가중치를 가짐

Seq2Seq Architecture



Training Techniques

- 디코더 구조 : 이전의 출력 결과가 다음 입력으로 들어감
 - 학습 단계에서 예측을 잘못해버리면 옆길로 샅 수도 있다.
 - 예) Expected result :
I think it'll be too *crowded on the first day of the sale.*
 - Prediction result :
I think it'll be too *busy to go there.*

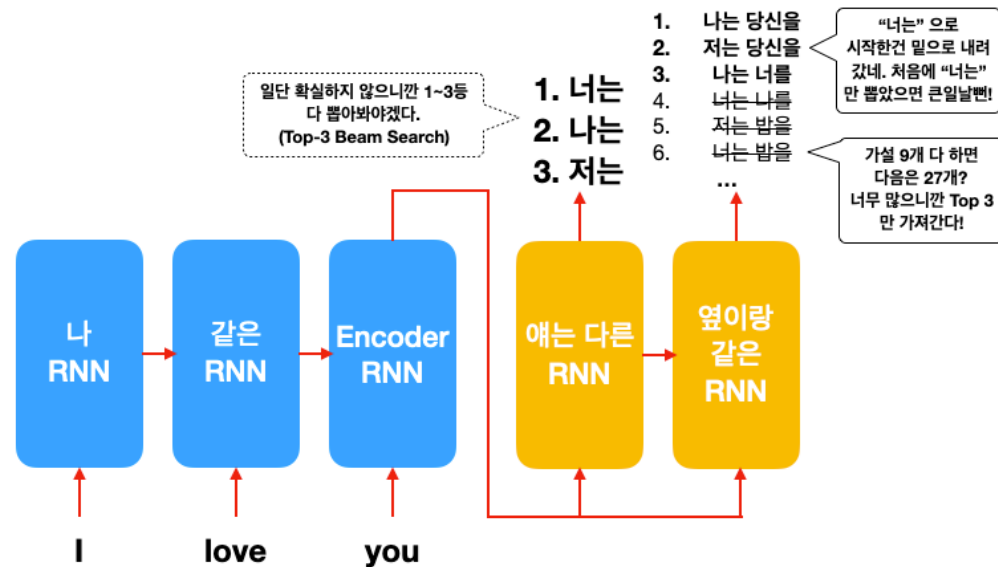


Training Techniques

- Teacher forcing
 - 예측이 틀릴 것을 대비하여, cheat sheet를 보면서 학습한다.
 - 확률적으로 이전 출력을 입력으로 쓸 지, 정답을 입력으로 쓸지 결정
 - 예) Expected result :
I think it'll be too *crowded on the first day of the sale.*
 - Prediction result :
I think it'll be too ~~*busy*~~ *to go there.*
→ *crowded on the first day of the sale.*
Teacher 개입

Training Techniques

- Top-K Beam Search
 - 예측을 할 때 확률이 **가장 높은 단어만 고르지 않고**,
상위 k개의 단어를 함께 출력
 - 다음 단어를 예측할 때 **k개의 단어를 고려하여 예측**



Seq2Seq의 한계점

- RNN 구조 자체의 한계 → 문장이 길어질수록...
 - Training Time : 학습시간이 너무 길어짐
 - Long-Term Dependency Problem : 학습 능력이 떨어짐
- 고정된 길이의 context vector
 - 5단어 길이의 문장이나, 100단어 길이의 문장이나 같은 크기로 인코딩됨
 - Bottle-neck : 문장을 압축하는 과정에서 정보 손실 발생

Q n A

Reference

- Original paper : <https://arxiv.org/pdf/1409.3215.pdf>
- <https://youtu.be/4DzKM0vgG1Y>
- <https://jiho-ml.com/weekly-nlp-22/>