



Summer NLP

Lec 06. Sequence Modeling Part 2

파이토치로 배우는 자연어 처리 CHAPTER 7~8

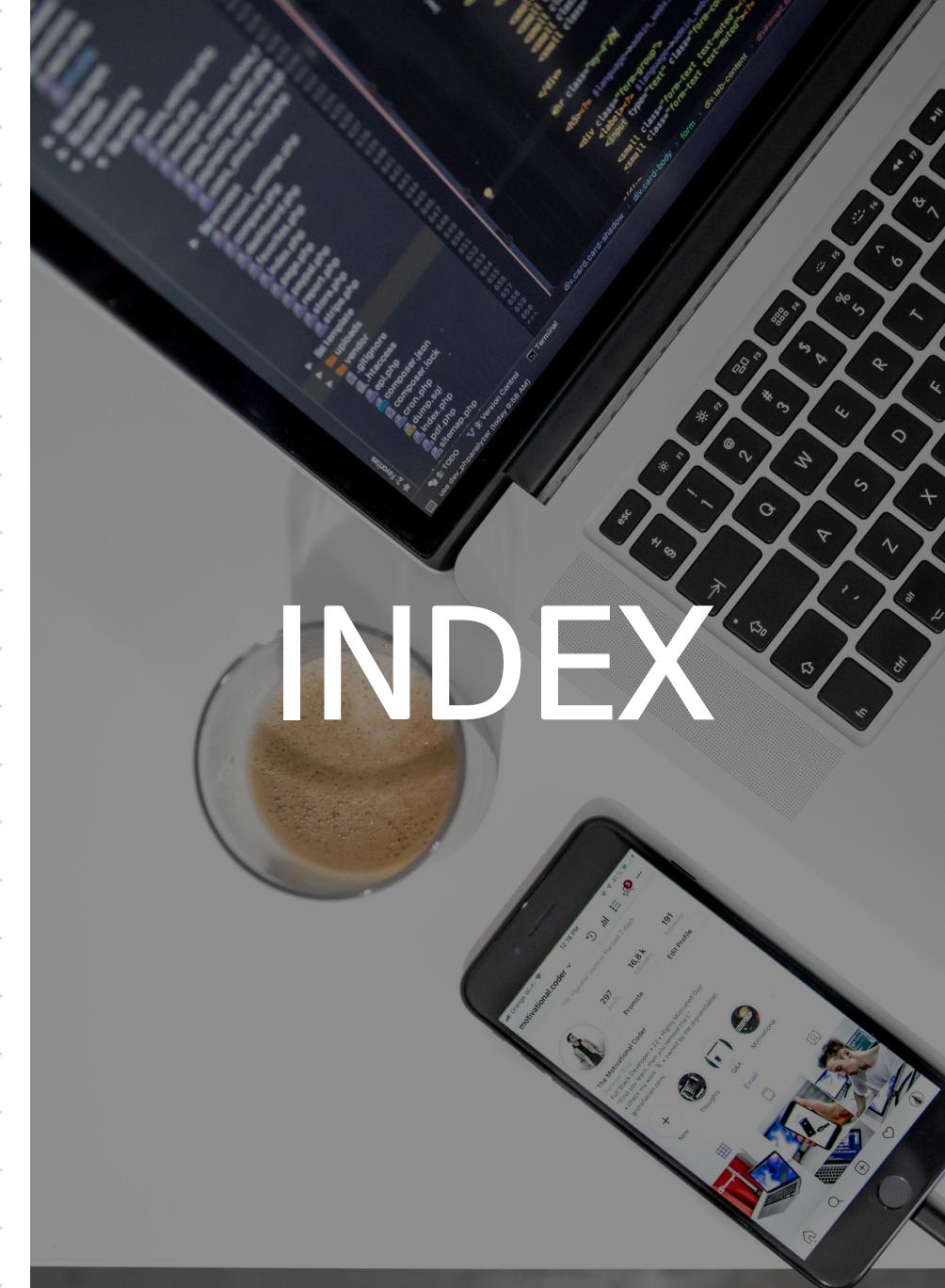
| Sequence to Sequence

| Bi-directional RNNs

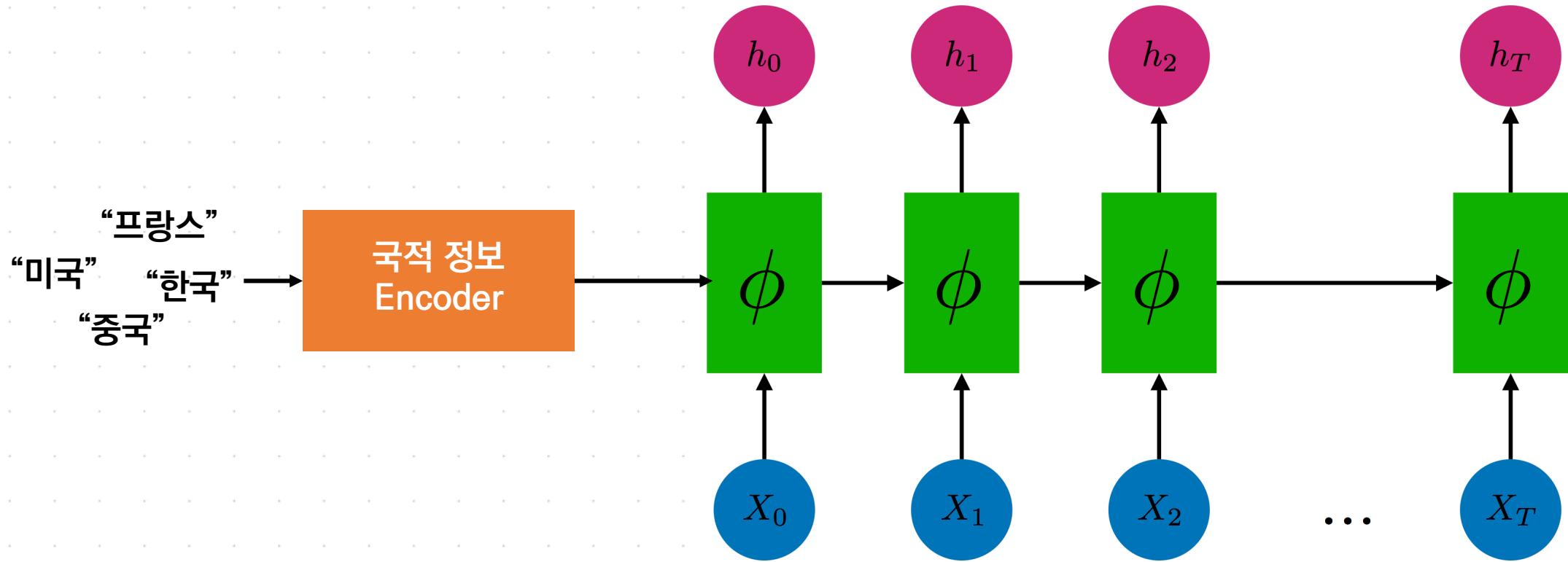
| Attention Mechanism

| 실습 : Machine Translation

| 고전모델, 최신모델, 더 배울 것들



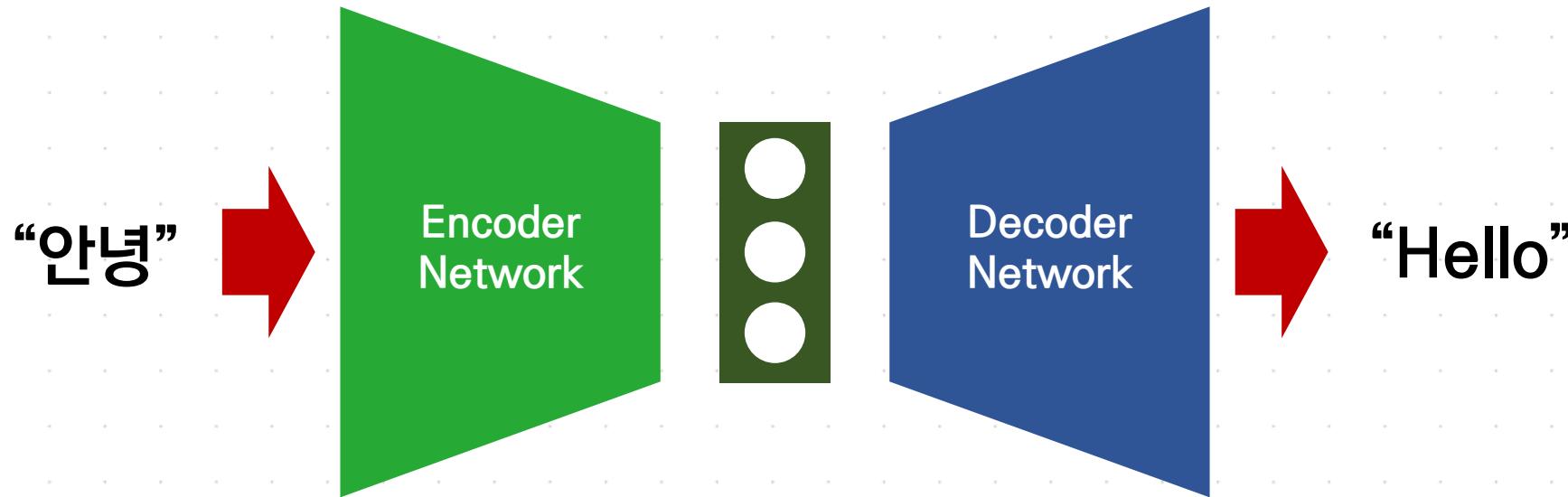
Sequence to Sequence



조건부 생성 모델

국적 정보를 임베딩한 값을 RNN Hidden State의 초기값으로 사용하여 Inference

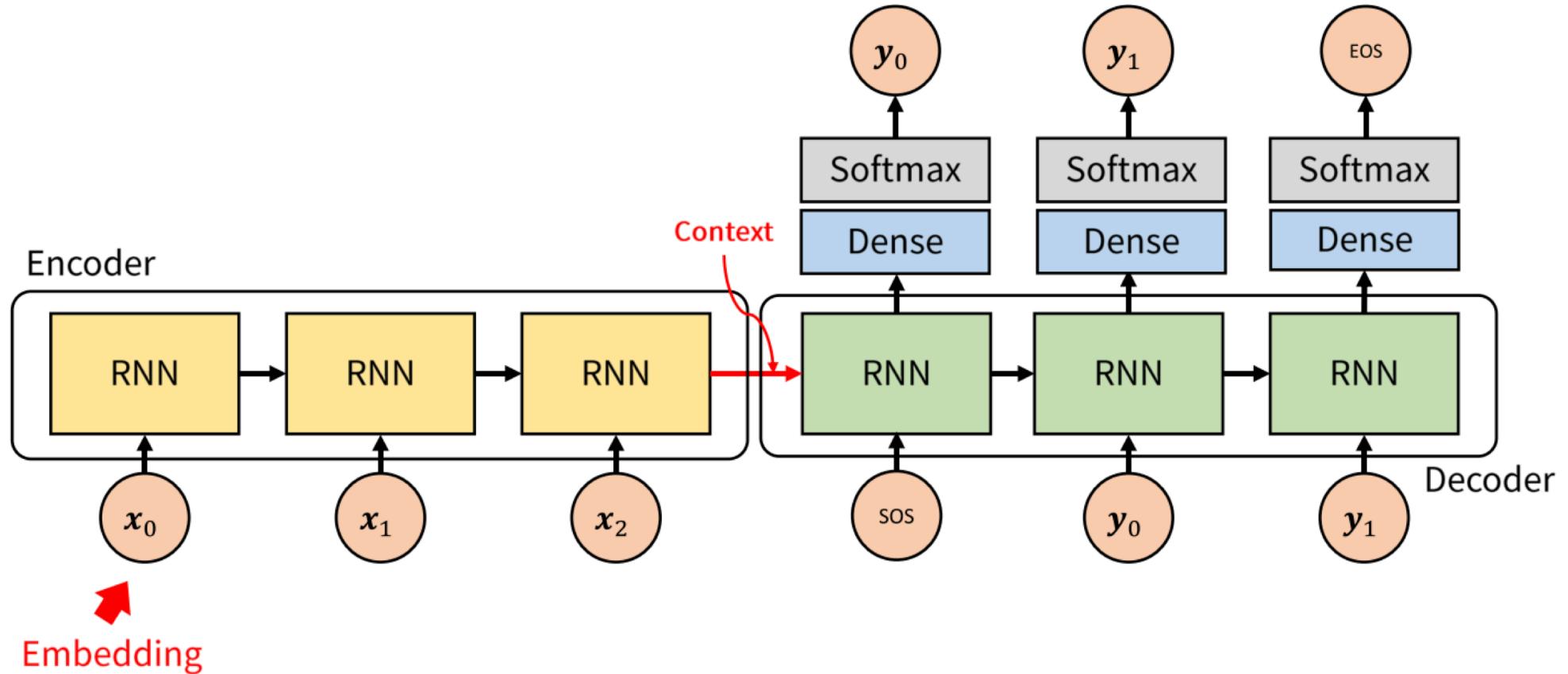
Sequence to Sequence



Encoder–Decoder Architecture

Encoder는 주어진 입력으로 고정 길이의 context vector를 생성,
Decoder는 context vector를 입력으로 원하는 출력을 생성

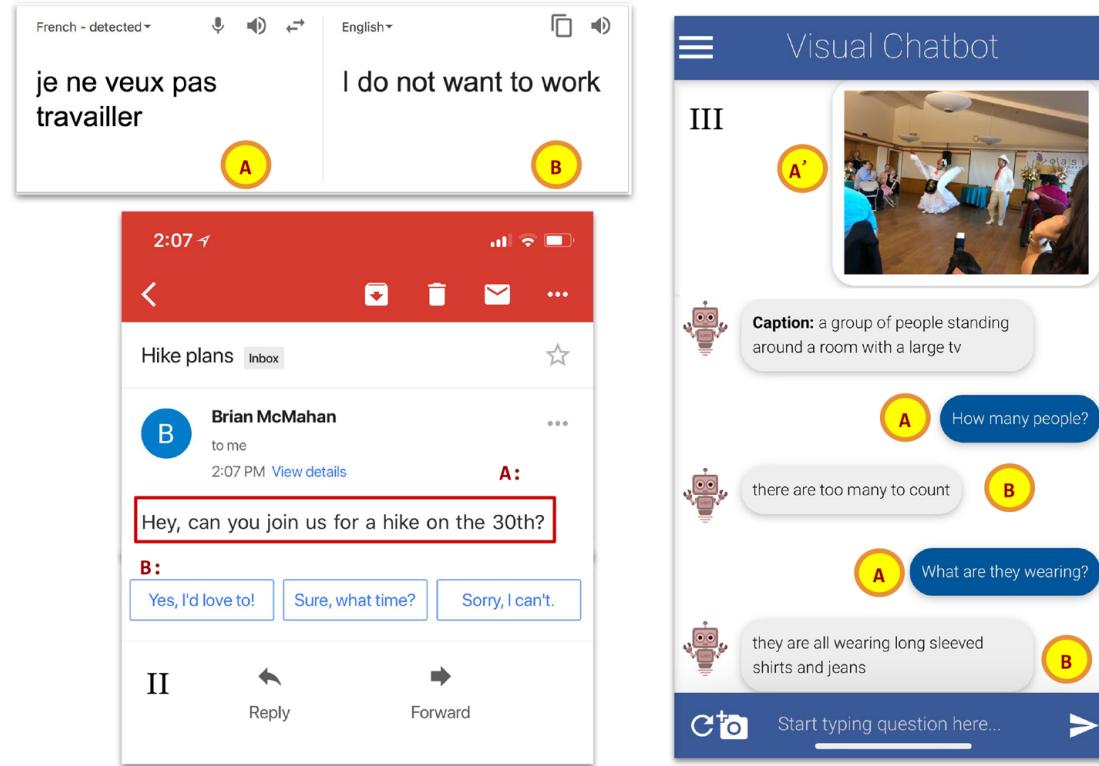
Sequence to Sequence



Sequence to Sequence Model

가변 길이의 Sequence를 입력으로 받고, 가변 길이의 Sequence를 출력으로 뱉는 모델,
Encoder-Decoder Architecture를 사용한다.

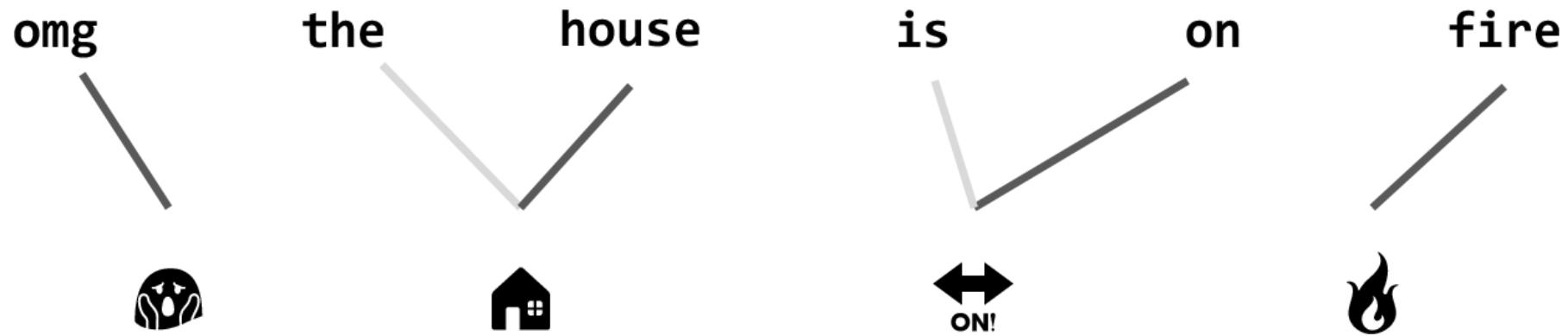
Sequence to Sequence



Examples of Sequence to Sequence Model

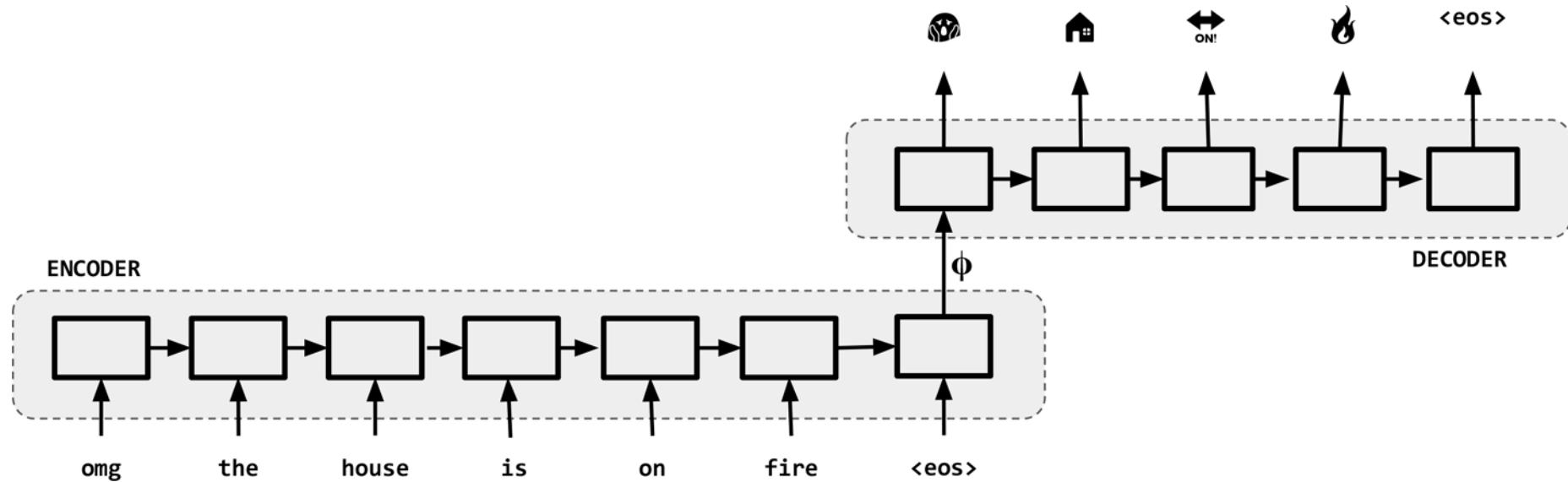
주어진 문장 A를 문장 B로 번역하거나, 대답을 만들거나,
대답들의 리스트를 추천받을 수 있다.

Sequence to Sequence



How Sequence to Sequence Model Works?
Seq2Seq의 Input과 Output은 가변적일 수 있으며,
입력과 출력 사이의 정렬(Alignment)를 학습한다.

Sequence to Sequence



How Sequence to Sequence Model Works?

문장의 시작은 <BOS> : Begin of Sentence Token으로, 문장의 끝은 <EOS> : End of Sentence Token으로 표현하여 문장 길이가 가변적일 수 있음.

Bi-directional RNNs

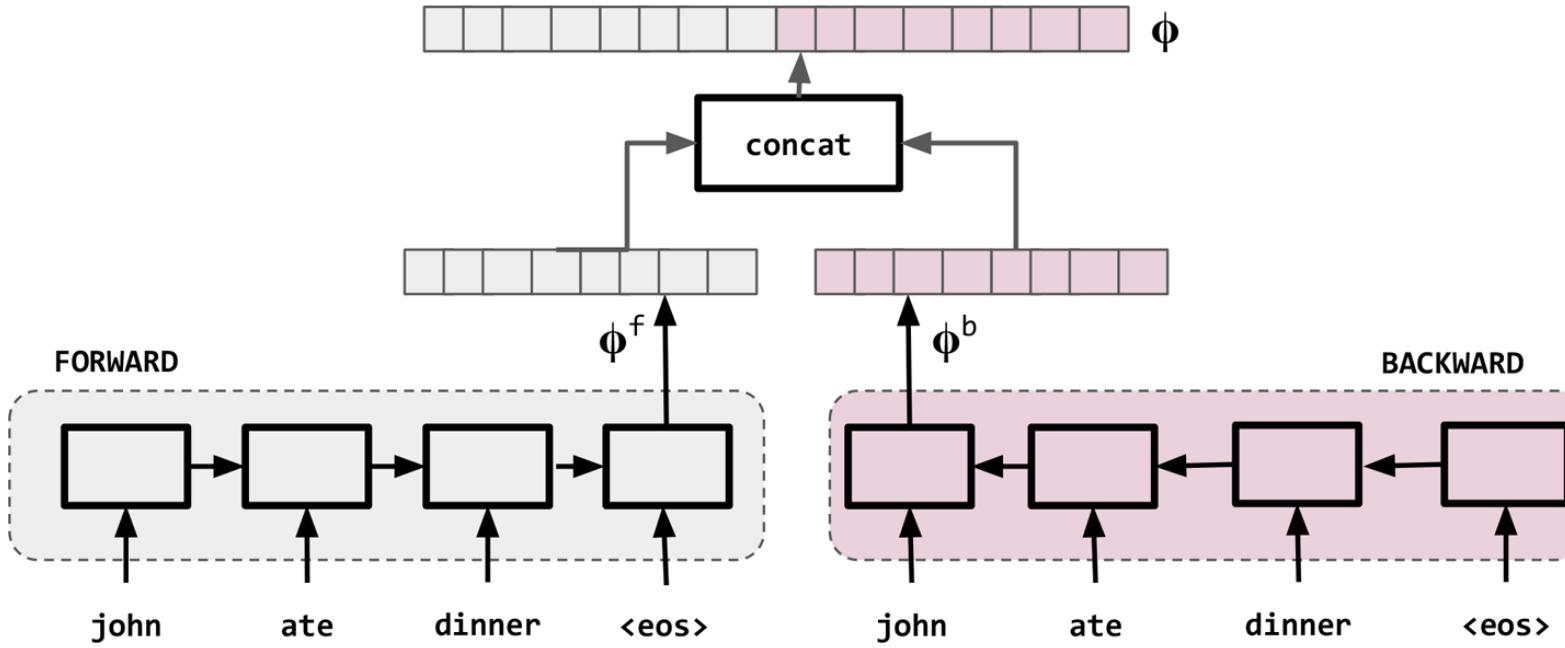
The man who hunts ducks out on the weekends.



Why Bi-Directional Model?

왼쪽→오른쪽으로 문장의 Context를 만들어내는 것하고,
그 반대로 Context를 만들어낼 때 다른 의미를 가질 수 있음.

Bi-directional RNNs



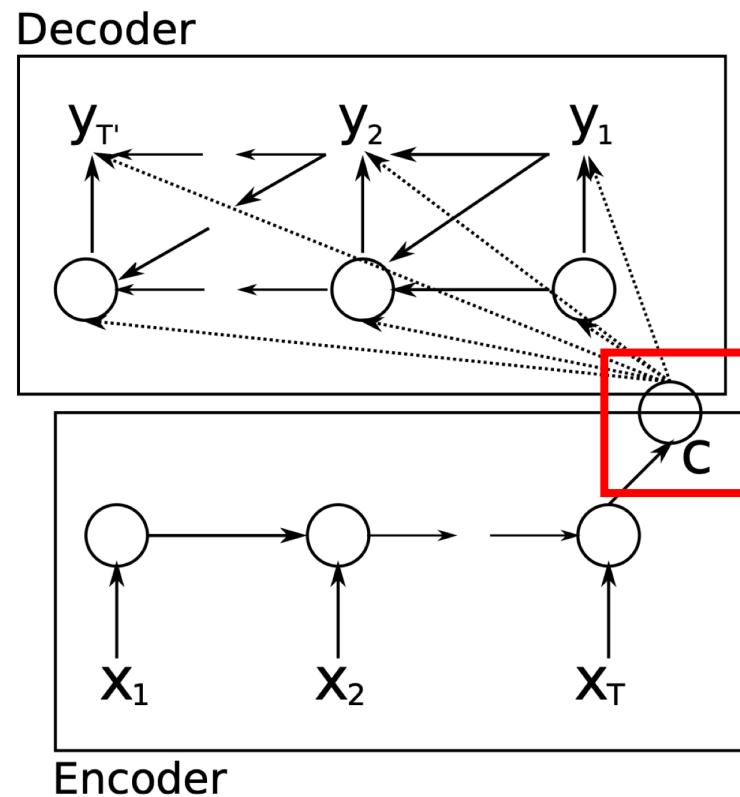
Bi-Directional Sequence to Sequence Model

문장을 순방향 역방향으로 RNN 모델을 입력시킨 뒤,
두 Context Vector를 Concatenate시켜 새 Context로 사용

Bi-directional RNNs

Problems of RNN based Seq2Seq Models Sequence to Sequence 모델에 문제점은 없을까?

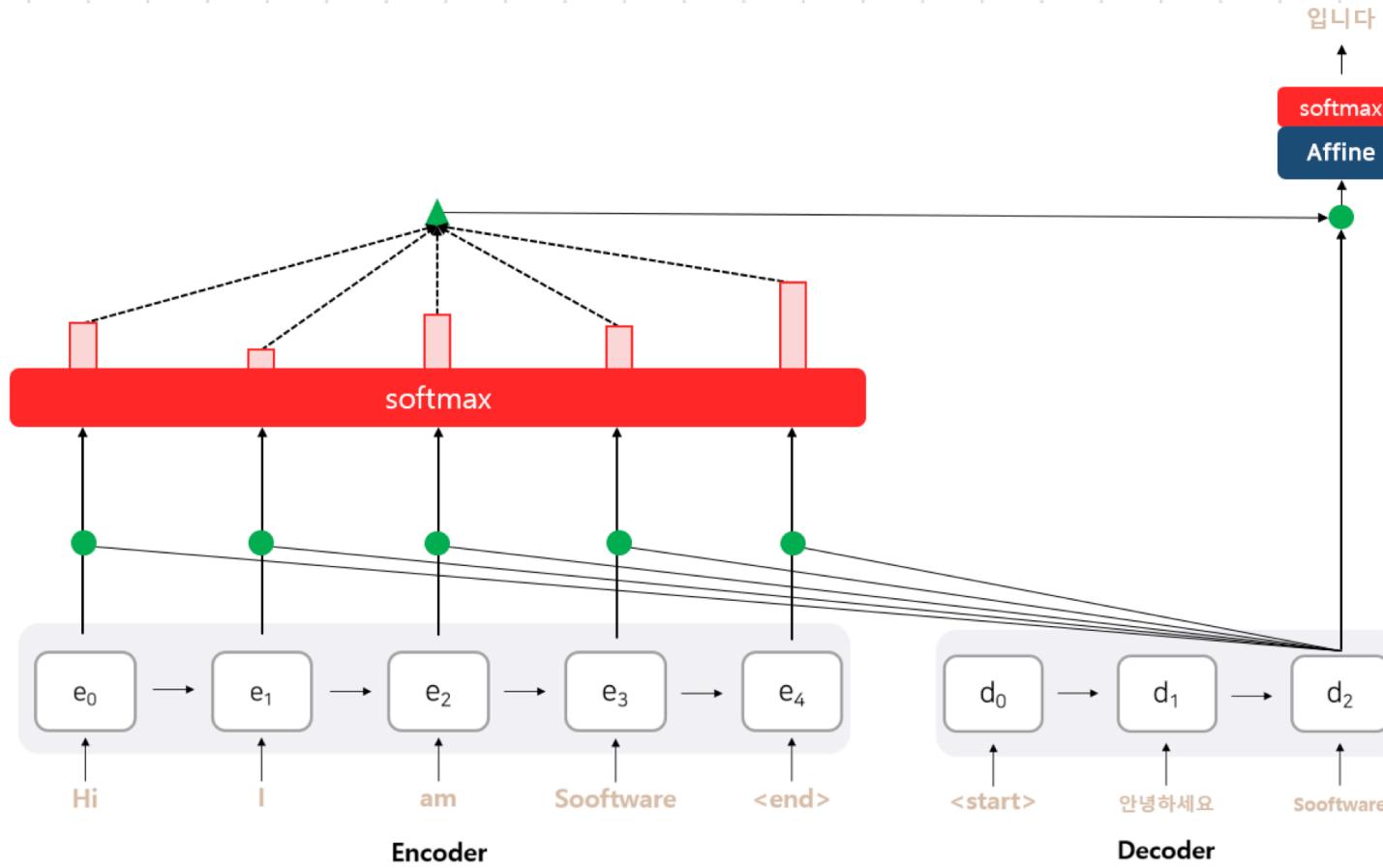
Bi-directional RNNs



Problems of RNN based Seq2Seq Models

주어진 문장을 고정 길이의 Context Vector 하나로만 압축하는 과정에서 **정보손실**이 있으며, 이전 결과를 현재의 Inference에 필요로 하는 RNN의 특성때문에 **연산의 병렬화**가 힘듦.

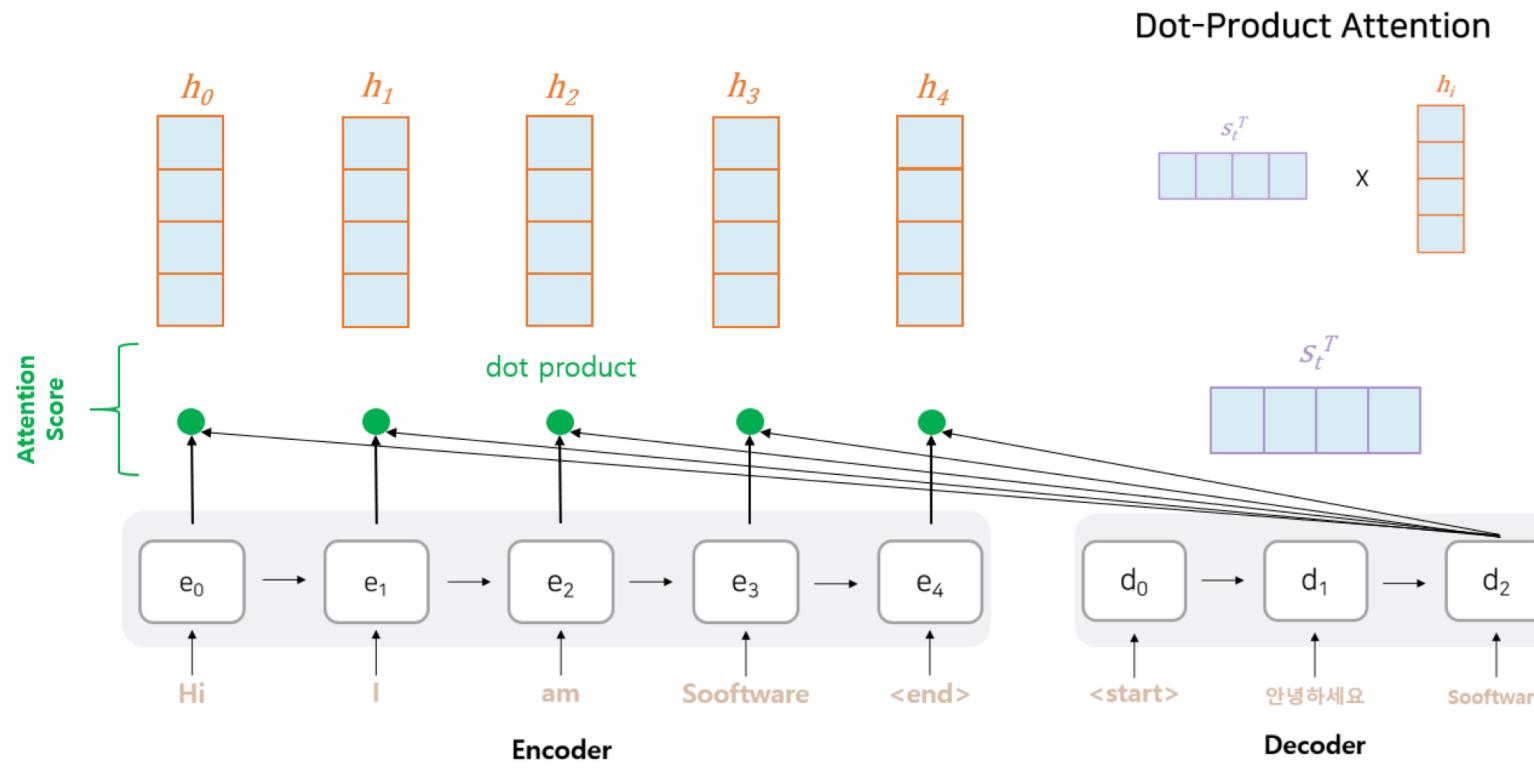
Attention Mechanism



Attention Mechanism

Decoder에서 입력받은 토큰을 바탕으로 Encoder 모든 Step의 Hidden State를 참고하여 각각의 State를 얼마나 참고할지 Weight를 구한 뒤, Weighted Sum으로 Context를 생성

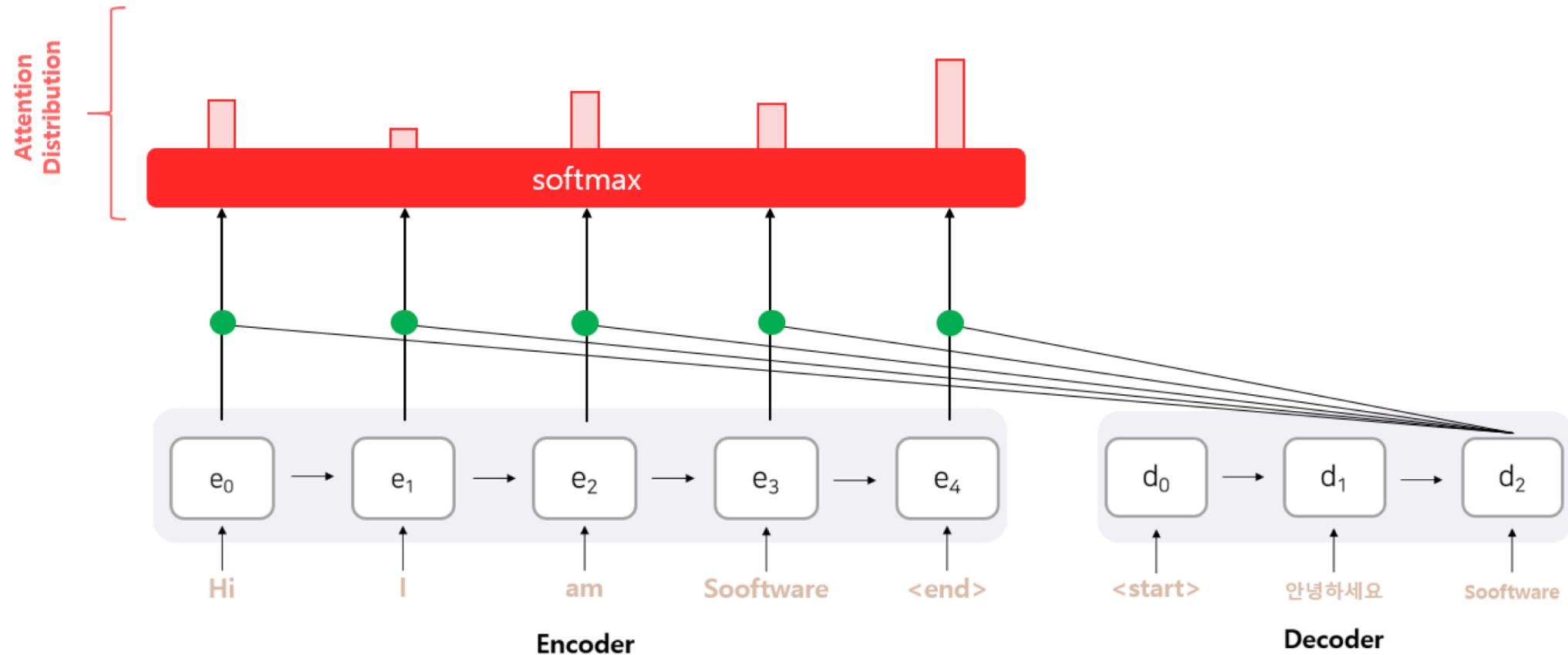
Attention Mechanism



Attention Part 1 : Get Attention Score

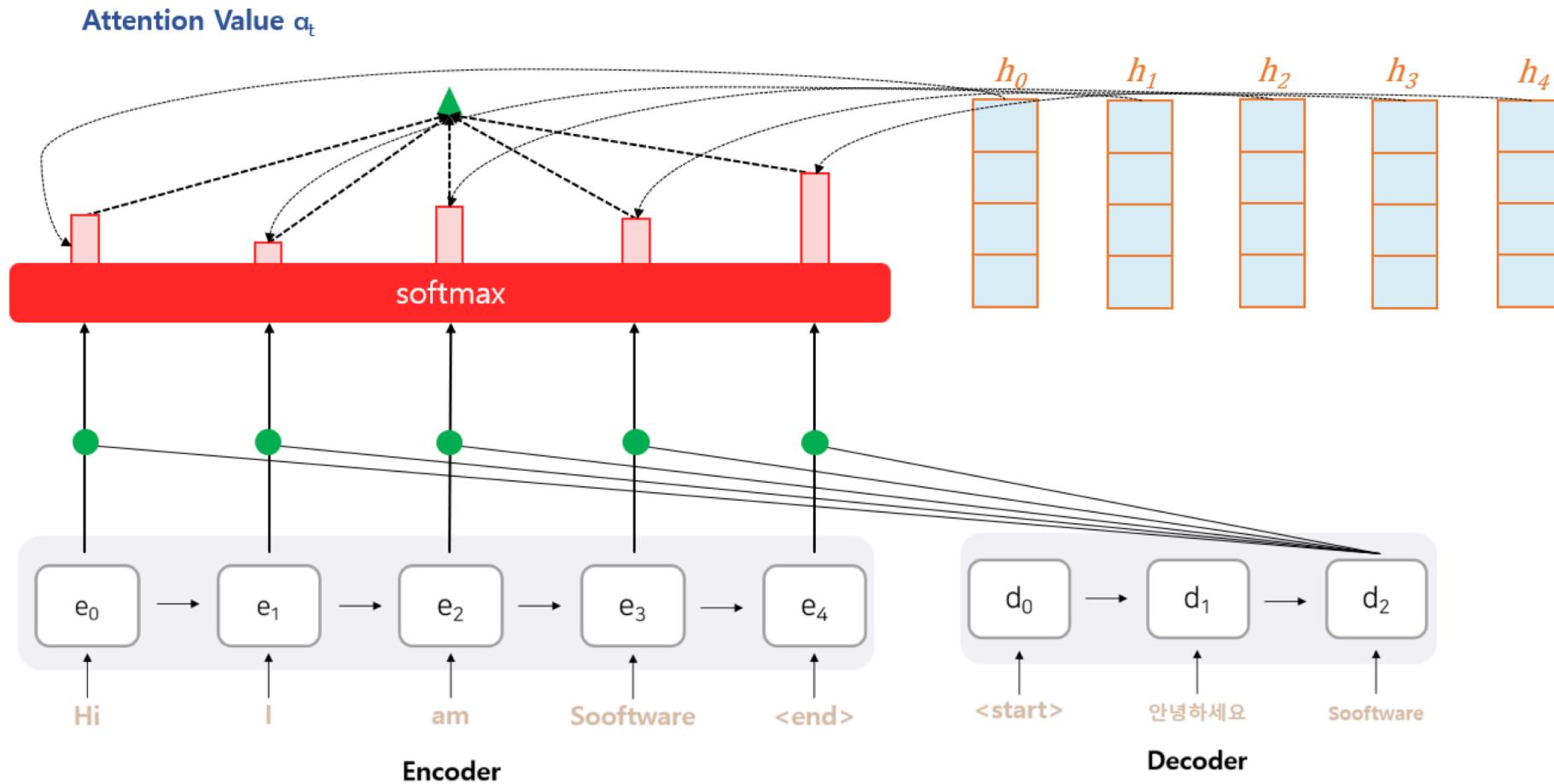
현재(i-step) Decoder의 Hidden State와 Encoder의 모든 Hidden State를 내적하여 각각의 Attention Score를 구한다. (Decoder와 Encoder의 Hidden State 유사도를 계산)

Attention Mechanism



Attention Part 2 : Get Attention Distribution
Softmax를 통해 0~1로 정규화하여 확률의 형태로 만든다.

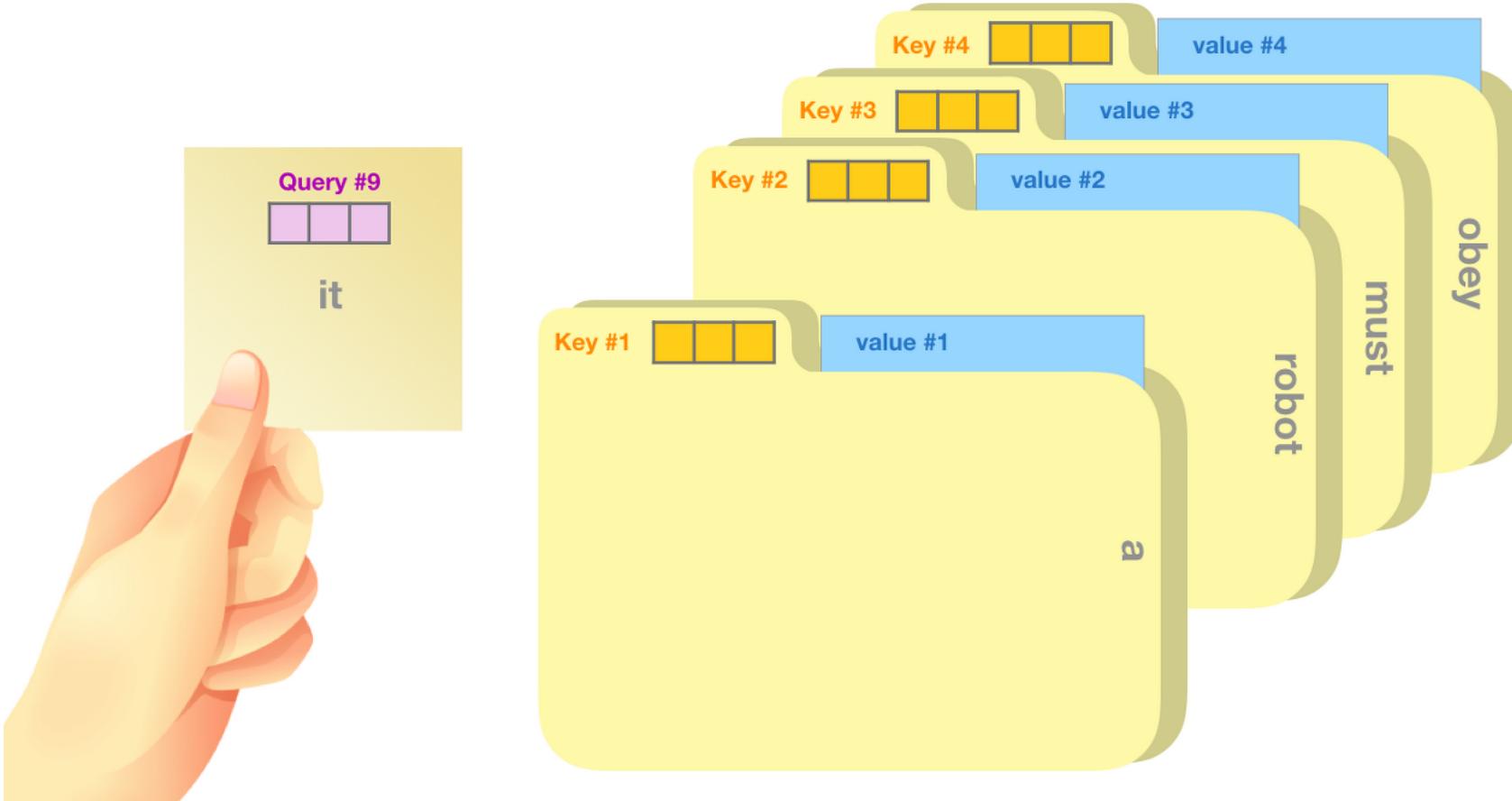
Attention Mechanism



Attention Part 3 : Weighted Sum

구해낸 Attention Distribution으로 Encoder의 Hidden State를 가중치에 따라 합친다.

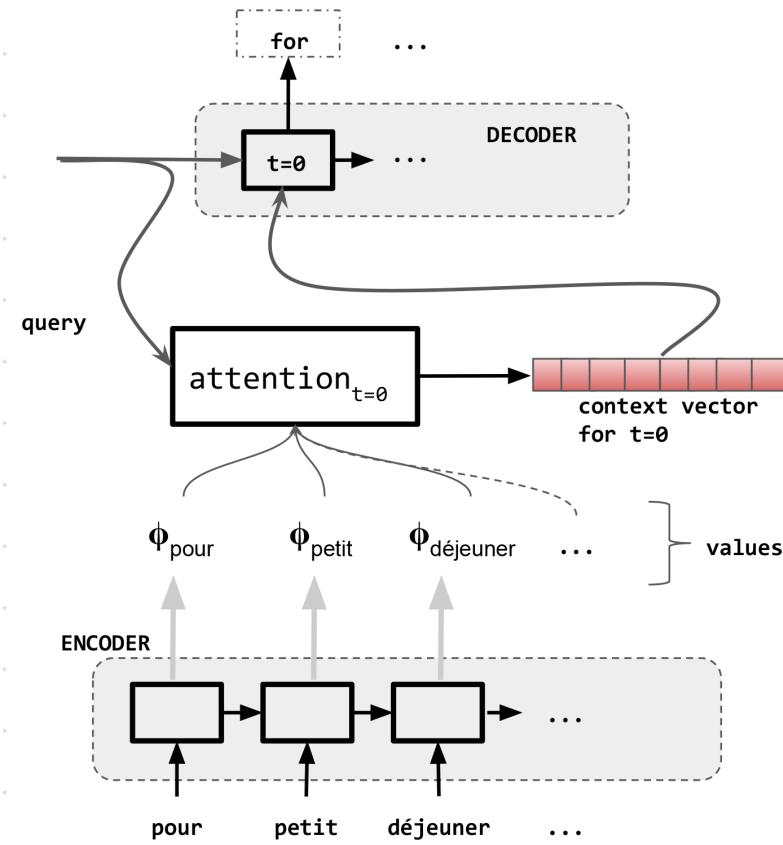
Attention Mechanism



Abstraction of Attention Mechanism

Attention에게 주어진 Context를 질의하여(Query) 주어진 Query와 Attention안의 Key간 유사도(Attention Score)를 구한 뒤, 해당 Key의 값(Value)과 Weighted Sum을 구해낸다.

Attention Mechanism



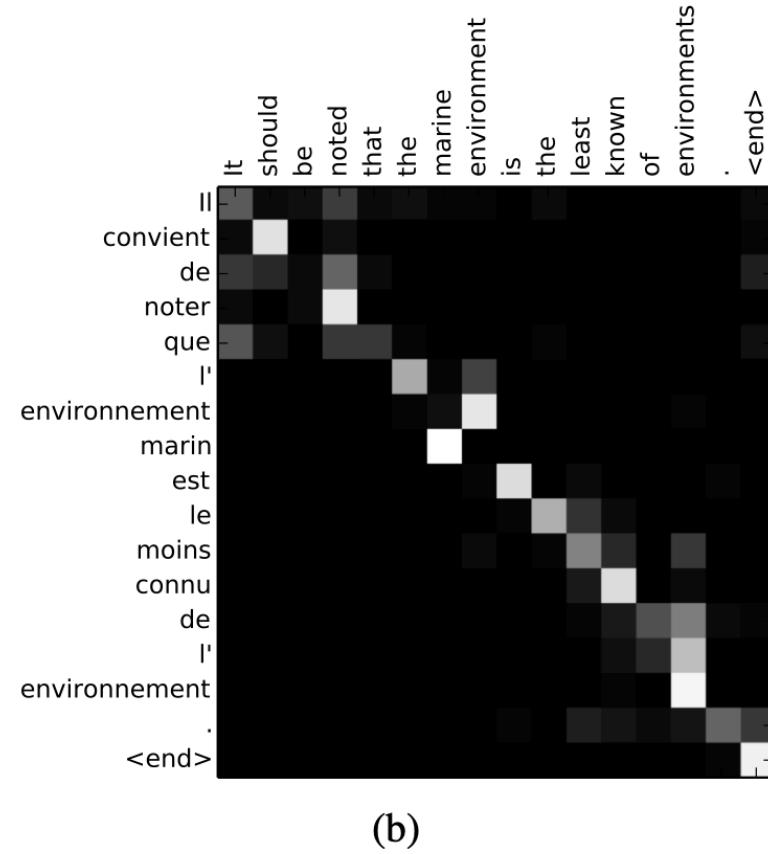
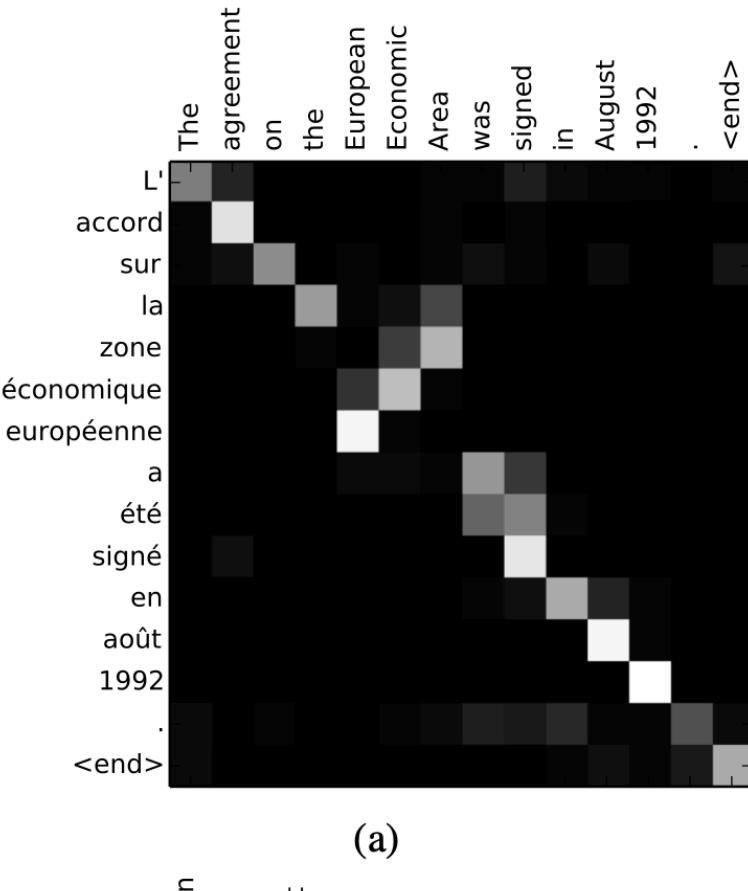
Query Key Value in Attention Machinism

Query : 현재 Step에서 만들어낸 Decoder의 Hidden State

Key : Encoder의 Hidden States

Value : Encoder의 Hidden States

Attention Mechanism



Result of Attention Mechanism

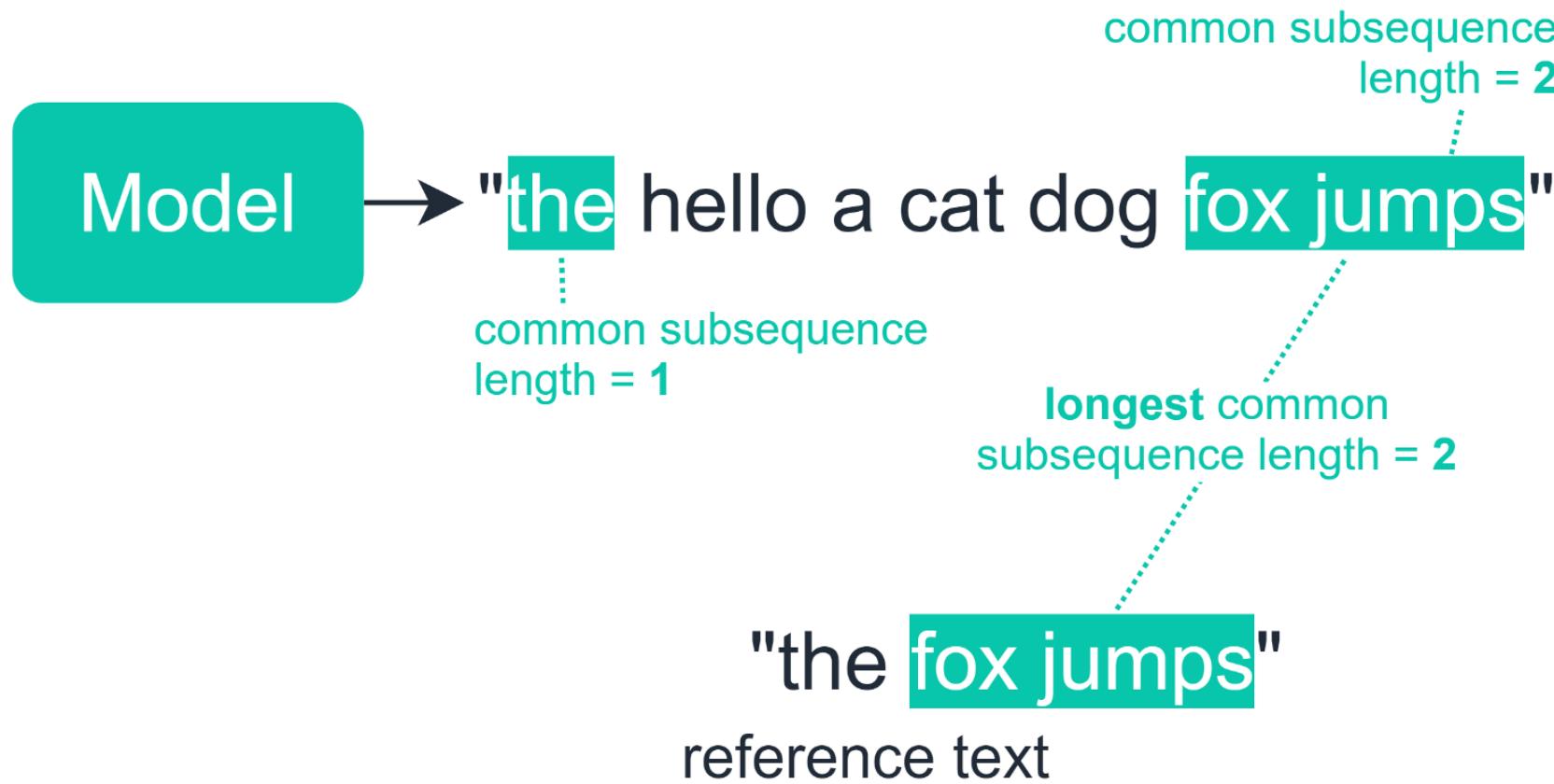
어떠한 토큰을 Inference할 때 Encoder에서 어떤 정보를 참고했는지 Visualize할 수 있음.

실습 : Machine Translation

Evaluation Metrics for Sequence Generation

문장은 정답이 여러개일 수 있어서 Precision, F1-Score와 같은 Metric은 사용하기 힘듦.

실습 : Machine Translation



Reference Output

모델의 Output과 미리 정해둔 Reference Output을 비교하여 정량적 수치로 표현

실습 : Machine Translation

BLEU

BLEU(Bilingual Evaluation Understudy) score란 성과지표로 데이터의 X가 순서정보를 가진 단어들(문장)로 이루어져 있고, y 또한 단어들의 시리즈(문장)로 이루어진 경우에 사용되며, 번역을 하는 모델에 주로 사용된다. 3가지 요소를 살펴보자.

- n-gram을 통한 순서쌍들이 얼마나 겹치는지 측정(precision)
- 문장길이에 대한 과적합 보정 (Brevity Penalty)
- 같은 단어가 연속적으로 나올때 과적합 되는 것을 보정(Clipping)

$$BLEU = \min(1, \frac{output\ length(\text{예측 문장})}{reference\ length(\text{실제 문장})})(\prod_{i=1}^4 precision_i)^{\frac{1}{4}}$$

BLEU

N-gram을 기반으로 Reference Output과 비교하여 수치값을 내놓는다.

실습 : Machine Translation

$$\begin{aligned} \text{PPL}(w_1, w_2, \dots, w_n) &= P(w_1, w_2, \dots, w_n)^{-\frac{1}{n}} \\ &= \sqrt[n]{\frac{1}{P(w_1, w_2, \dots, w_n)}} \end{aligned}$$

Perplexity

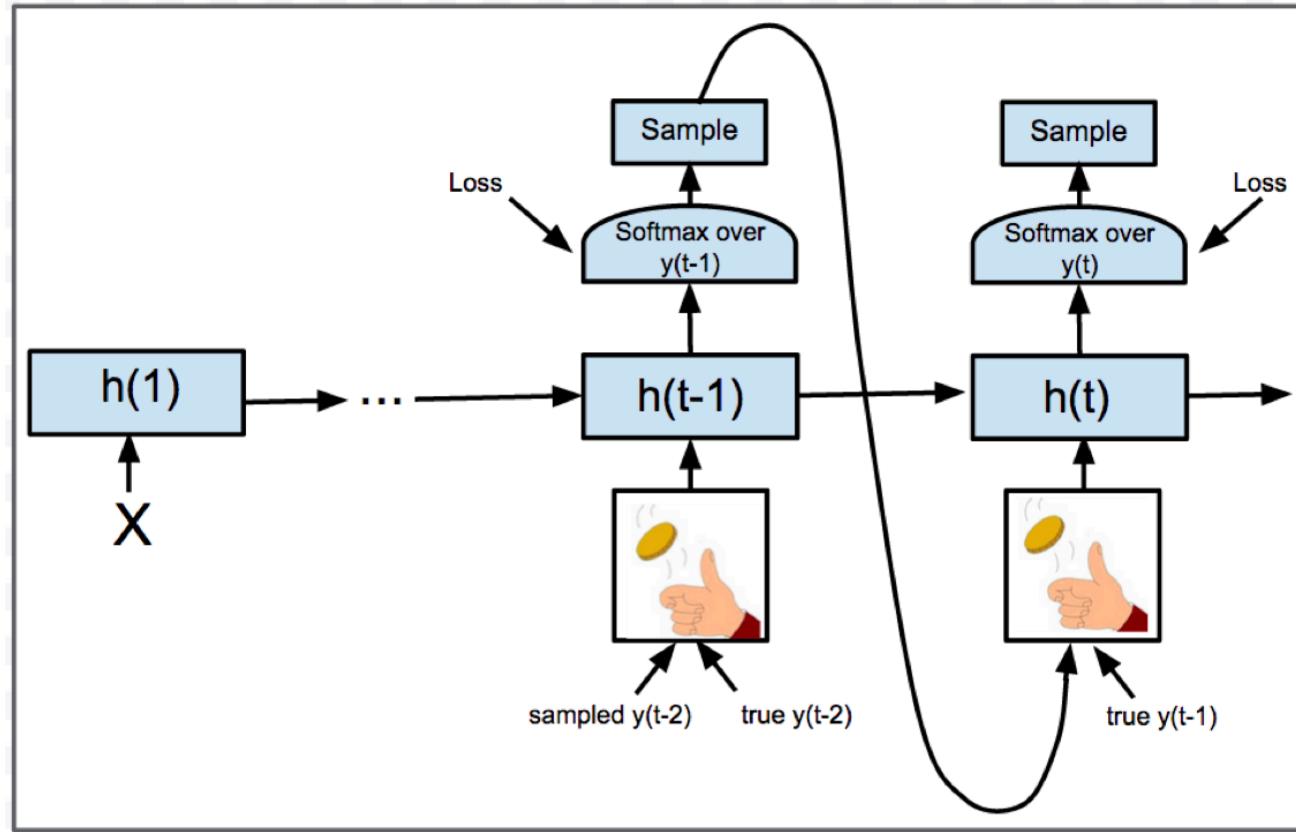
‘모델이 정답을 헷갈려하는 정도’의 의미로 이해해도 된다. (확률의 역수)

실습 : Machine Translation

Problems in Inference

모델 훈련때와는 달리, 실제 Inference에서는 타깃 시퀀스를 알 수 없어,
테스트 상황과 훈련 상황의 괴리가 발생할 수 있다.

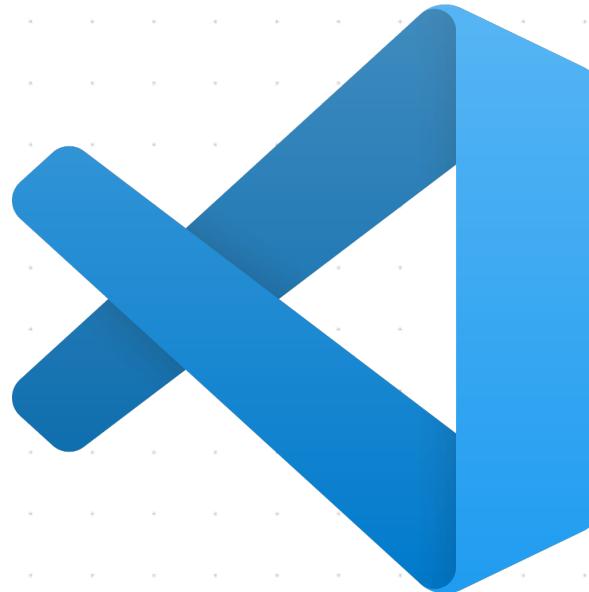
실습 : Machine Translation



Scheduled Sampling

훈련시에도 모델의 출력 분포를 바탕으로 샘플링을 진행한 결과를 Target Sequence로 활용

실습 : Machine Translation



실습 : Machine Translation

실습 1. Scheduled Sampling을 사용하지 않음.

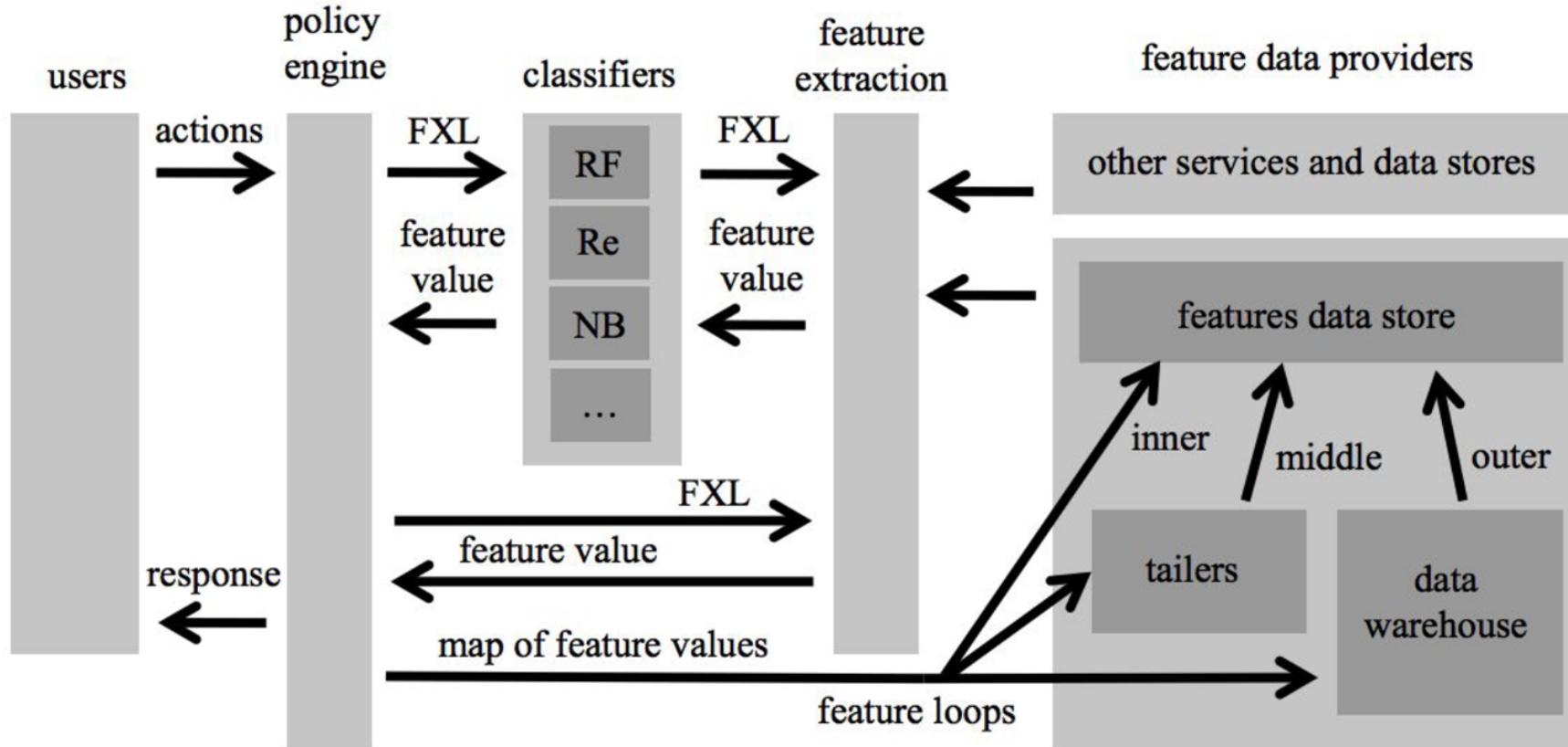
실습 2. Scheduled Sampling을 사용함.

| 고전 모델, 최신 모델, 더 배울 것들

Design Patterns for NLP System

제품수준의 NLP 시스템은 어떻게 만들 수 있을까요?

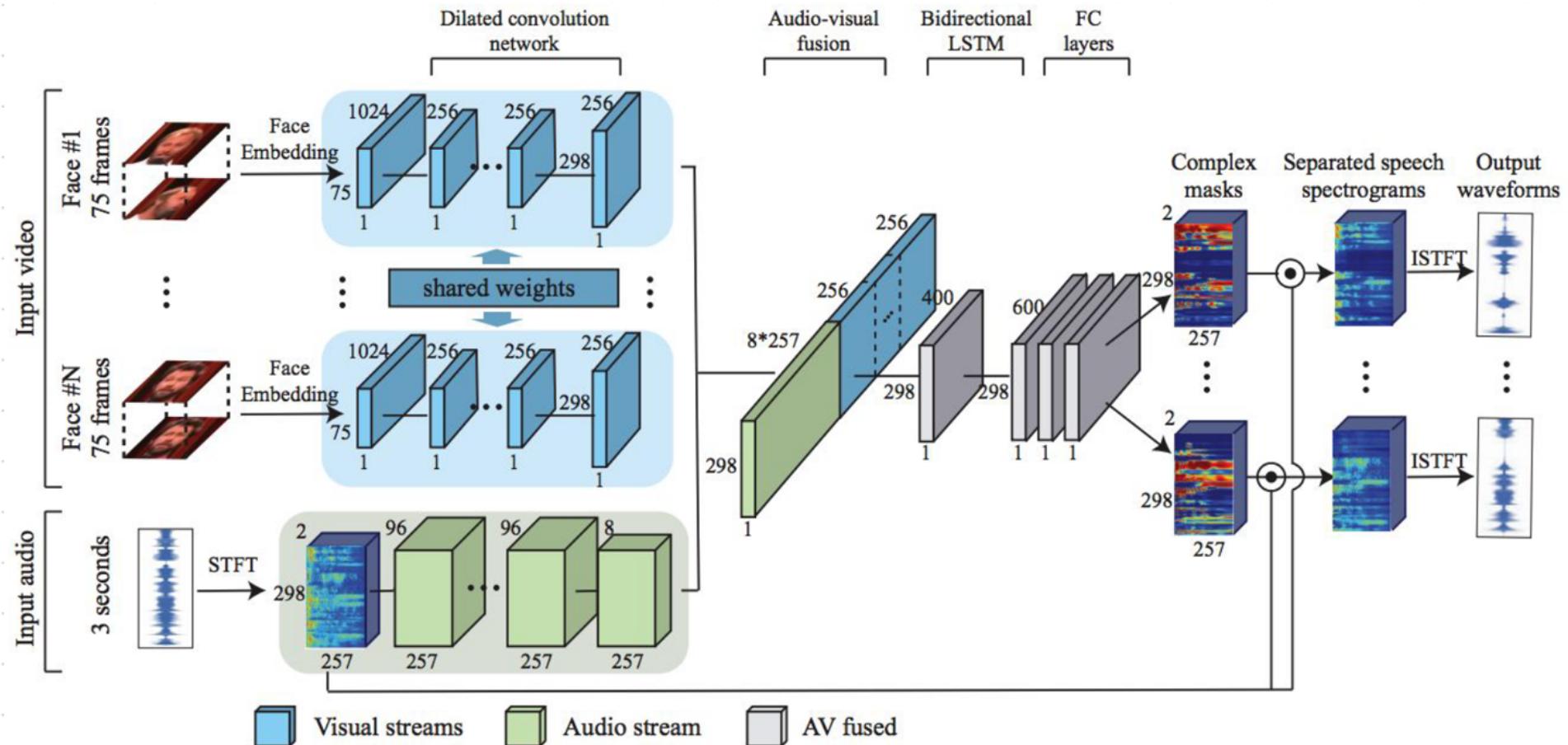
고전 모델, 최신 모델, 더 배울 것들



Online / Offline System

실시간으로 사용자의 요청을 처리 / Batch 단위로 자연어처리 작업을 처리

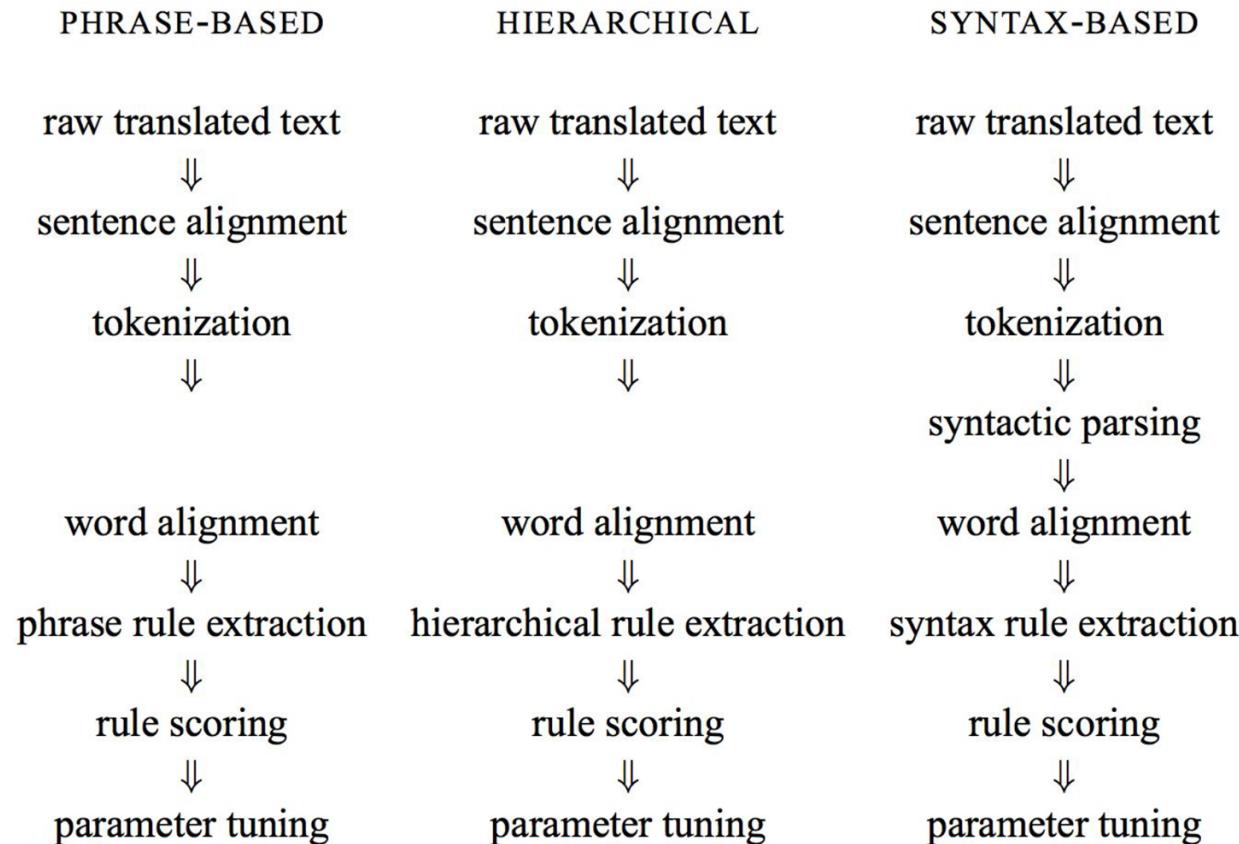
고전 모델, 최신 모델, 더 배울 것들



Unimodal / Multimodal

글 이외 사진, 음성, 영상등의 데이터도 모델링할 것인가?

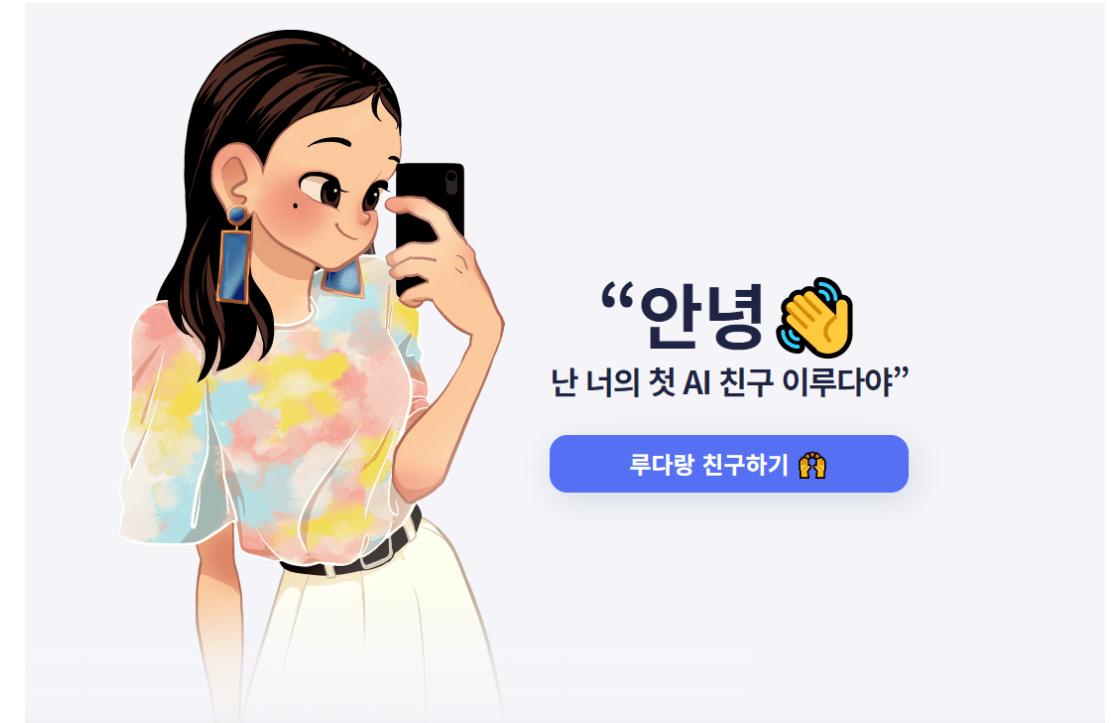
고전 모델, 최신 모델, 더 배울 것들



End-to-End / Piecewise

하나의 파이프라인으로 문제를 해결 / 각 문제를 작은 소문제로 나누어 해결

고전 모델, 최신 모델, 더 배울 것들



Closed-Domain / Open-Domain

한 도메인에서만 학습 / 범용 사용을 위한 여러 도메인에서의 학습

고전 모델, 최신 모델, 더 배울 것들



Monolingual / Multilingual

한 언어에서만 학습 가능하도록 / 여러 언어에서 학습 가능하도록



Thank You