

Autoregressive model & Non-Autoregressive model

정영석

■ Task 정의

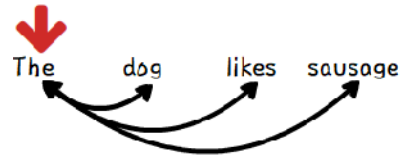
- Sequence to Sequence
 - ✓ Neural Machine Translation (신경망 기반 기계 번역)
 - ✓ Automatic Speech Recognition (음성인식)
 - ✓ Text to Speech Synthesis (음성 생성(?))
 - ✓ OCR (Optical Character Recognition)
- Sequence data processing
 - ✓ Language Model
 - ✓ RNN based model
 - ✓ Text classification model

Autoregressive model?

- Auto Regressive model

- 이전 시점의 결과를 통해 과거의 결과를 예측하는 모델 (Transformer, LAS, Tacotron)

$$Y^* = \prod_{t=1}^M P(y_t | y_{<t}, X; \theta)$$



Autoregressive model?

■ Autoregressive model

- Autoregressive model 장 · 단점!

- ✓ 장점

- 이전 데이터를 참고해 결과를 생성하는 만큼 정확도 ↑
 - 비교적 자연스러운 결과를 생성

- ✓ 단점

- 병렬적으로 결과를 생성하지 못해 추론 · 학습에 필요한 많은 시간이 필요함
 - 이전의 결과에 의존해 결과를 생성하므로 Missing word 같은 문제점을 보임

Non-Autoregressive model?

- Non-Autoregressive model

- 이전 시점의 결과를 고려하지 않고, 모든 시점의 결과를 한번에 만드는 모델(PNAT, FastSpeech, CTC LOSS...)

$$Y^* = \prod_{t=1}^M P(y_t, X; \theta)$$

Non-Autoregressive model?

■ Non-Autoregressive model

• Non-Autoregressive model의 장 · 단점

✓ 장점

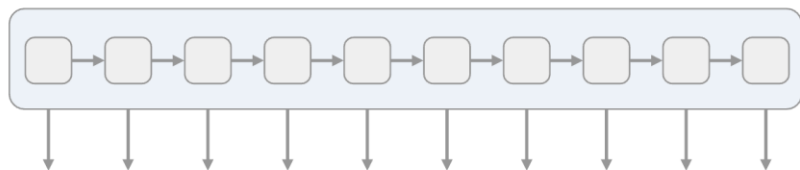
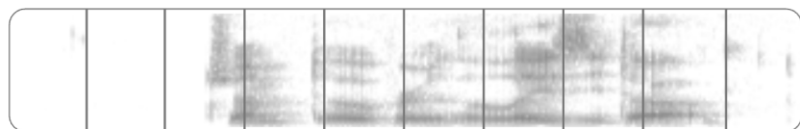
- 병렬적인 연산을 통해 결과를 생성해 추론 · 학습에 필요한 시간을 줄임
- Missing word문제가 적어진다는 것이 몇몇 실험을 통해 검증됨

✓ 단점

- 생성된 결과물의 품질 ↓
- 결과물의 sequence length, word order를 조절할 장치가 필요함.

Non-Autoregressive model?

- Non-Autoregressive model
 - CTC Loss (Speech Recognition)



h	h	h	h	h	h	h	h	h	h
e	e	e	e	e	e	e	e	e	e
l	l	l	l	l	l	l	l	l	l
o	o	o	o	o	o	o	o	o	o
€	€	€	€	€	€	€	€	€	€

h h e € € l l l € l l o

h e € l € l o

h e l l o

h e l l o

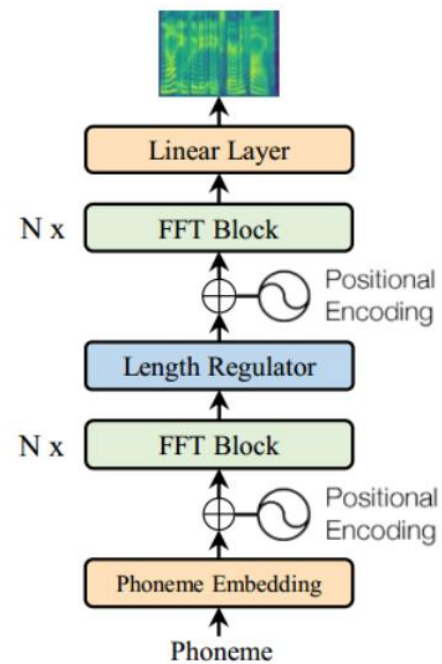
First, merge repeat characters.

Then, remove any ϵ tokens.

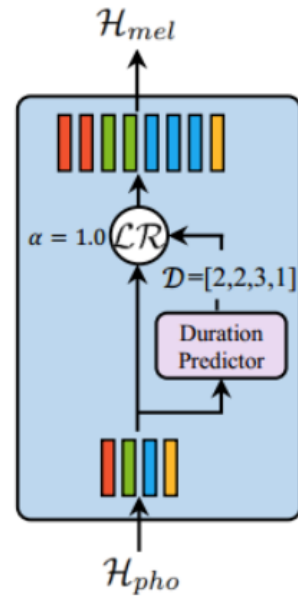
The remaining characters are the output.

Non-Autoregressive model?

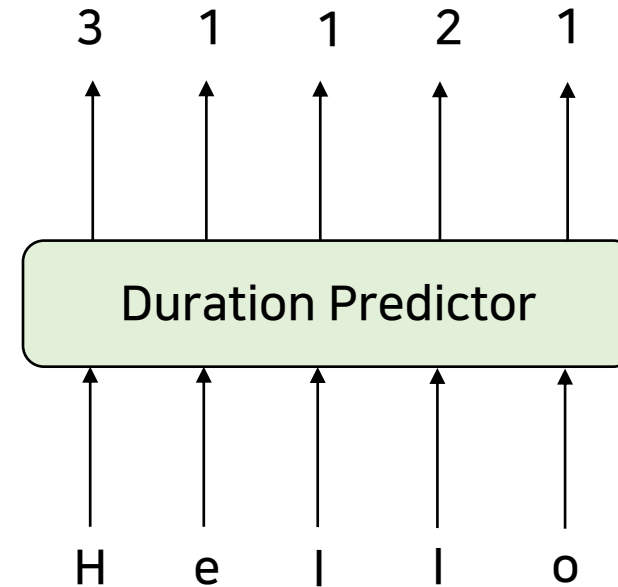
- Non-Autoregressive model
 - Fastspeech [NIPS 2019] (Length Predictor)



(a) Feed-Forward Transformer



(c) Length Regulator

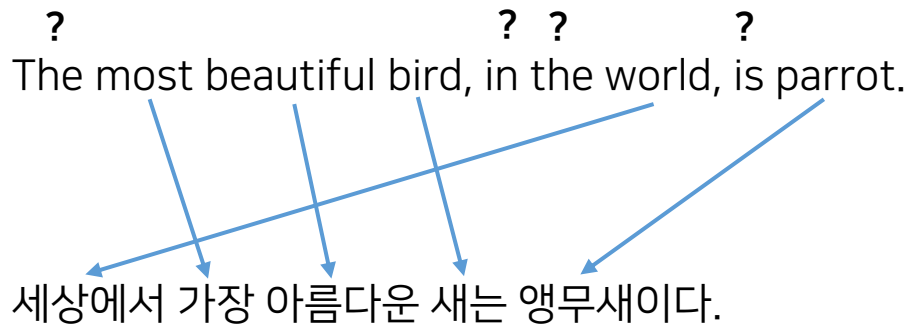


Non-Autoregressive model?

■ Non-Autoregressive model

• Neural Machine Translation

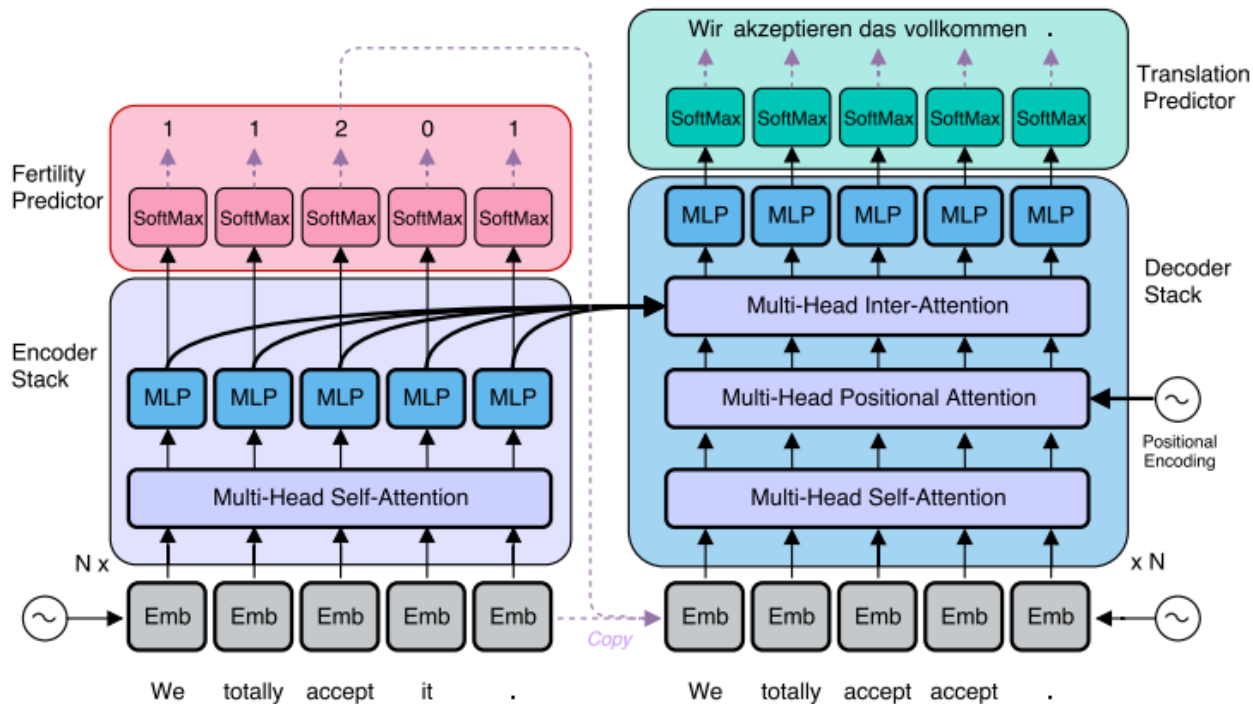
- ✓ 기계 번역에서 앞서 설명한 방법론이 어려운 이유
 - 언어에 따라 어순이 달라짐
 - 단어의 개수가 반드시 병렬적으로 같지 않음



Non-Autoregressive model?

- Non-Autoregressive model

- Non-Autoregressive Transformer [ICLR 2018] (Positional Attention)



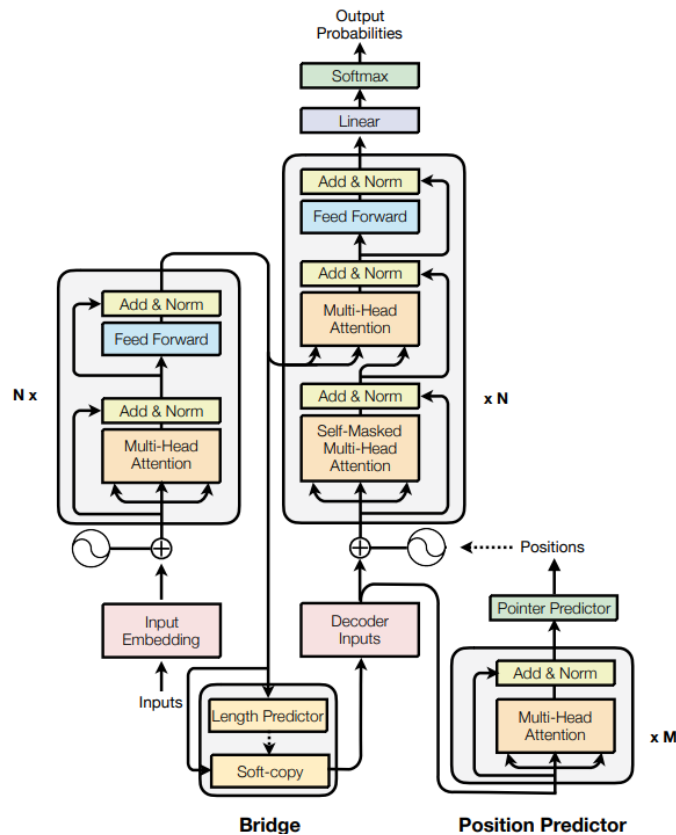
$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_{\text{model}}}}\right) \cdot V$$

Q, K = positional encoding

Non-Autoregressive model?

■ Non-Autoregressive model

• PNAT (Position Predictor)



✓ Bridge(Length Predictor)

- Input Sentence와 Target Sentence 사이의 상대적 거리를 예측 [-20, 20]

✓ Position Predictor

- Input Sentence의 단어를 target sentence의 해당하는 word 위치에 mapping하는 역할을 수행함