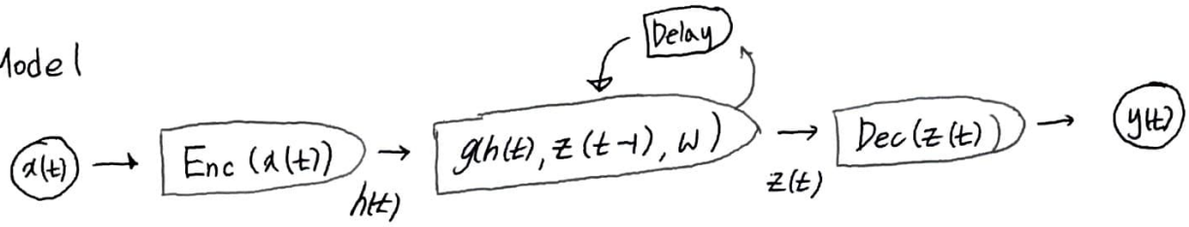
데이터의 시퀀스를 다루기위해

• Recurrent Neural Network : 순환신경망 (신경망에 순환고리를 적용)

→ Basic Model



Encode $h(t) = Enc(x(t))$

$z(t) = g(h(t), z(t-1), w)$ → recurrence formula

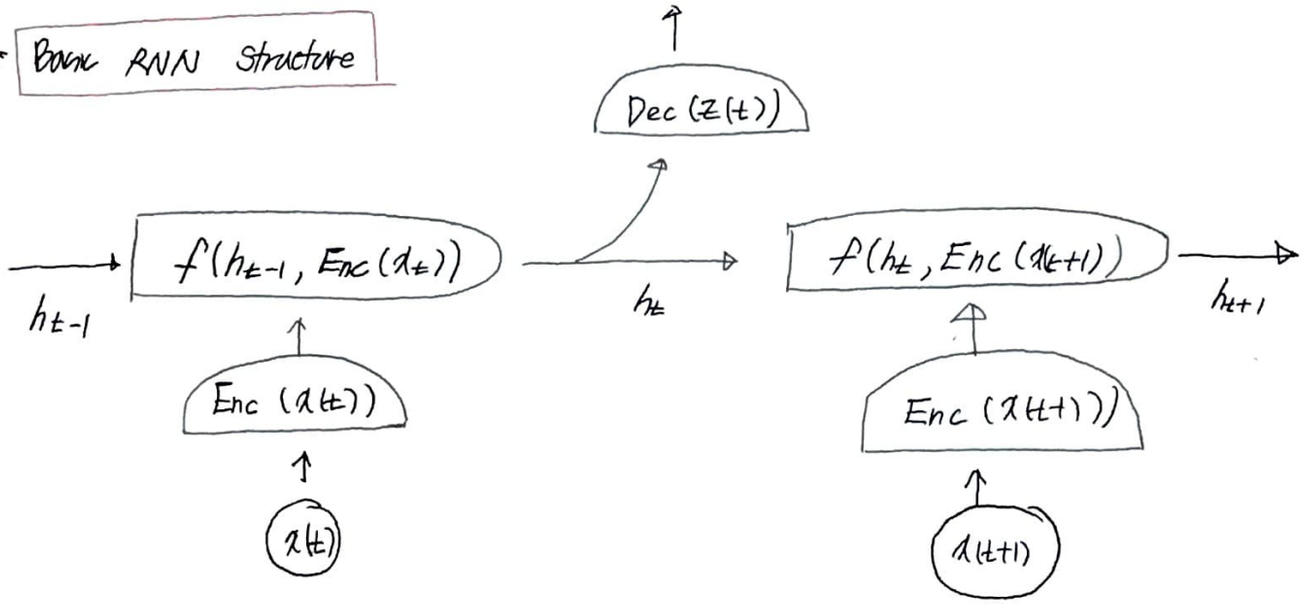
Decode $y(t) = Dec(z(t))$

$h_t = f_w(h_{t-1}, x_t)$ (old state, input vector)

→ $h_t = \tanh(W_{hh} h_{t-1} + W_{ah} x_t)$

$= \tanh\left(W \begin{pmatrix} h_{t-1} \\ x_t \end{pmatrix}\right)$

→ Basic RNN Structure



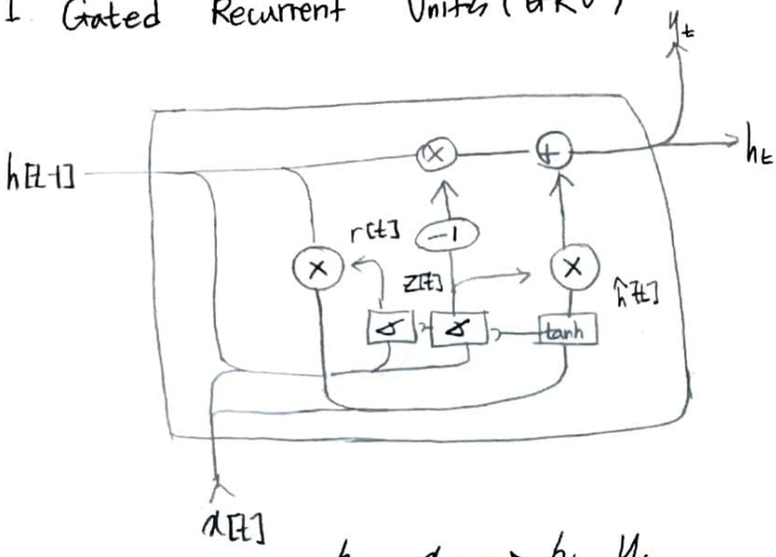
- Back Propagation Through time (BPTT)

P1) 경사소멸) 극복위해 GRU, LSTM 고안!

P2) 경사폭발 ^

Gated Recurrent Network 의 종류인

1. Gated Recurrent Unites (GRU)



h_{t-1} (hidden), $x_t \rightarrow h_t, y_t$

$$z_t = \sigma_g(W_z x_t + U_z h_{t-1} + b_z)$$

$$r_t = \sigma_g(W_r x_t + U_r h_{t-1} + b_r)$$

$$h_t = z_t \odot h_{t-1} + (1 - z_t) \odot \phi_h(W_h x_t + U_h(r_t \odot h_{t-1}) + b_h)$$

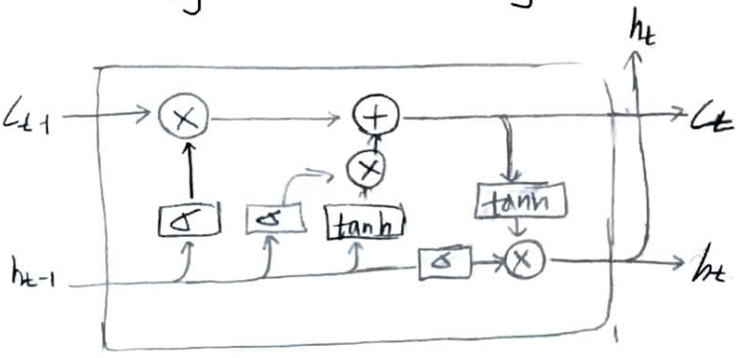
(미래 정보 / 과거 정보)

- r_t : reset gate, 과거정보를 얼마나 잊을지 (0 ~ 1)
- z_t : update gate 과거정보 중 뭘 미래로 보낼지.

미래 정보 z_t

과거 정보 $(1 - z_t) \rightarrow r_t$

2. Long Short-Term Memory (LSTM)



C_{t-1} (cell state), h_{t-1} (hidden), $x_t \rightarrow C_t, h_t$

$$f_t = \sigma_g(W_f x_t + U_f h_{t-1} + b_f)$$

$$i_t = \sigma_g(W_i x_t + U_i h_{t-1} + b_i)$$

$$O_t = \sigma_o(W_o x_t + U_o h_{t-1} + b_o)$$

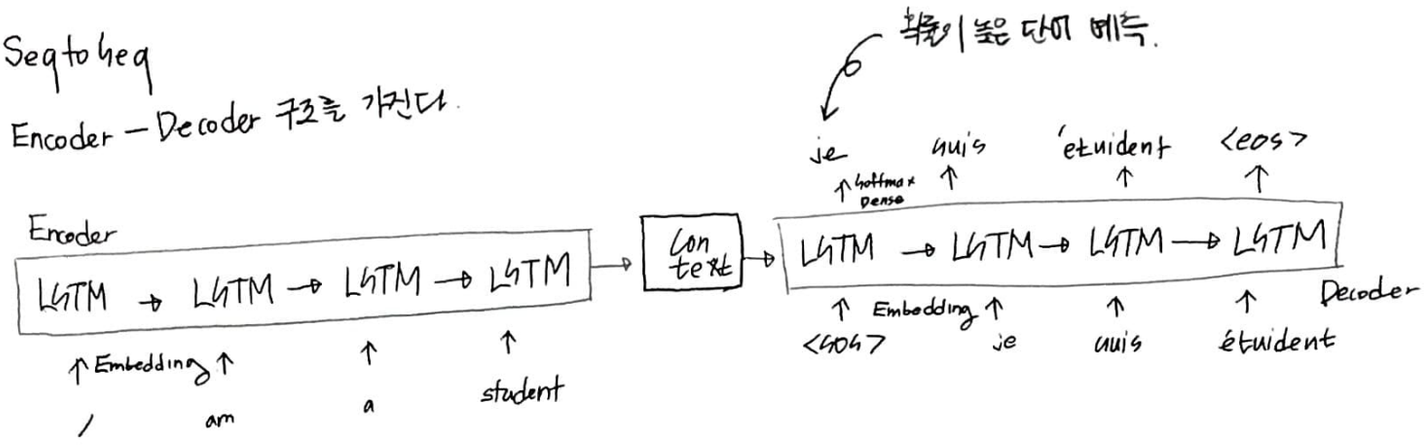
$$C_t = f_t \odot C_{t-1} + i_t \odot \tanh(W_c x_t + U_c h_{t-1} + b_c)$$

$$h_t = O_t \odot \tanh(C_t)$$

- f_t : forget gate
- i_t : input/update gate
- O_t : output gate
- C_t : cell state gate 셀정보유지, 제어

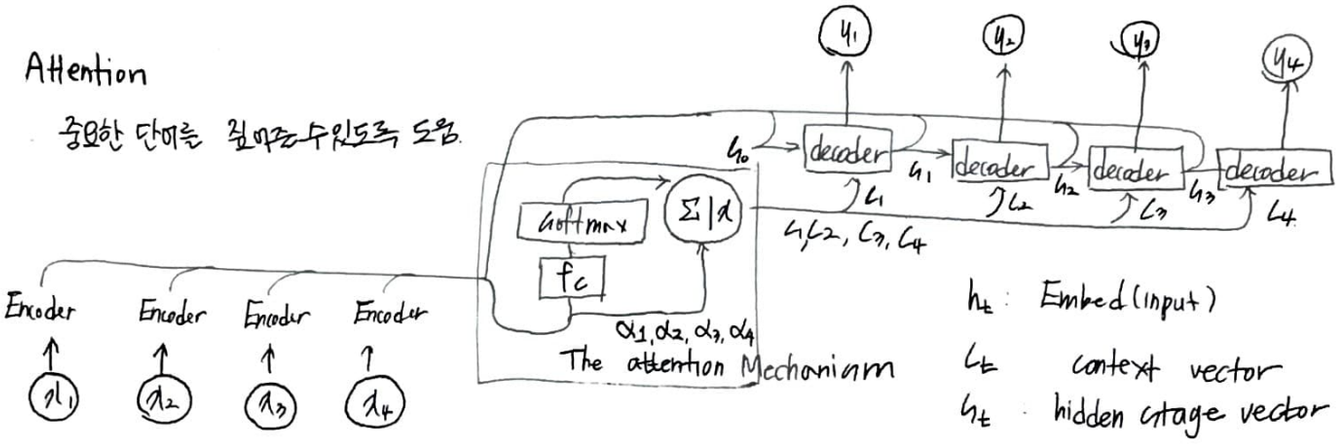
3 Seq to Seq

: Encoder – Decoder 구조를 가진다.



4. Attention

중요한 단어를 집중할 수 있도록 도움



h_t : Embed(input)

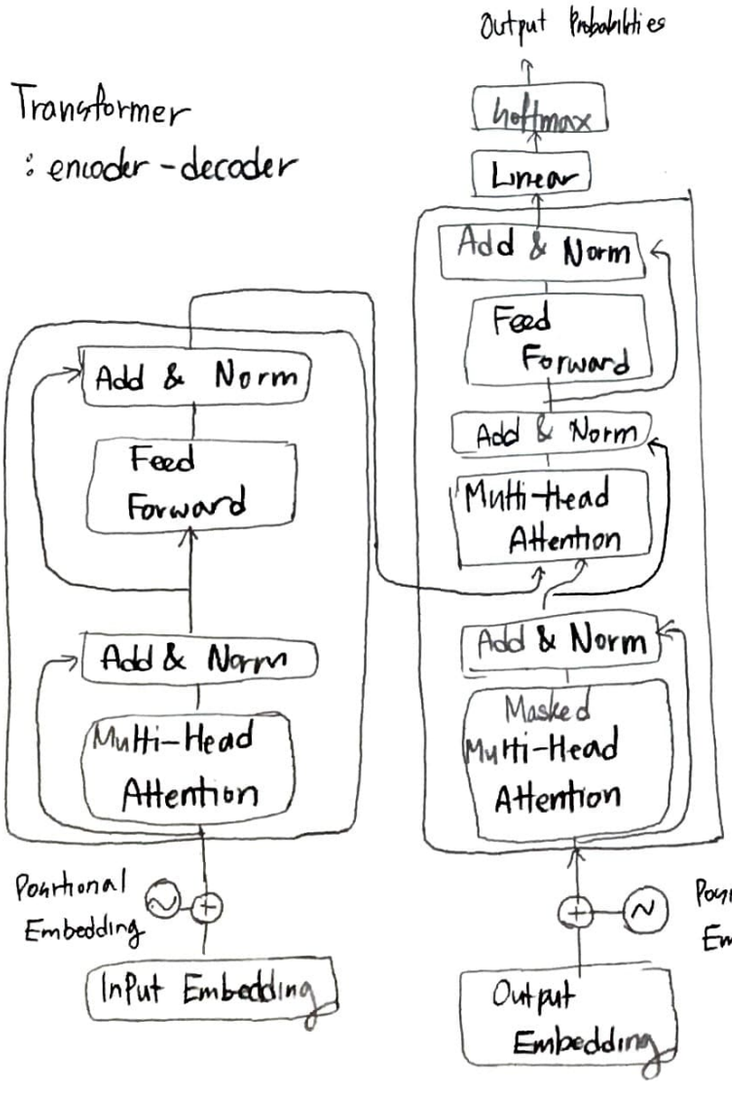
c_t : context vector

s_t : hidden stage vector

α_t : attention

5. Transformer

: encoder - decoder



Encoder : input sequence 압축

Decoder : 타겟 시퀀스 생성

- Multi-Head attention : self attention을 병렬적으로 사용
- Masked Multi-Head attention : 현재 position 후의 position에 attention ×
- self attention : 자기자신에 수행하는 attention
- Cross Attention

* 자연어 처리를 위한 비지도 학습

→ Word2Vec : 단어들에 대한 벡터공간모델 배우기

how? 몇몇 단어를 마스킹하고 공백을 채우기 위해 이웃 단어 사용 (linear Projection).

→ GPT (왼쪽 맥락만!)

Pretrain : 다음 단어 예측. 고전적 언어모델 : 각 타임스텝마다 단어 예측

Fine-tune → ex. 명사, 동사, 형용사, 감성분석

→ ELMo (양쪽 맥락 모두!)

Pretrain : 다음 단어 예측 but 양쪽 맥락 모두 고려.

1) 왼→오 2) 오→왼 3) (1)+2)

→ BERT 마스크처리된 단어 예측

Word to Vec + Transformer

* 자연어 처리를 위한 Self-supervised Learning.

→ XLNet

→ SpanBERT

→ ELECTRA

→ ALBERT

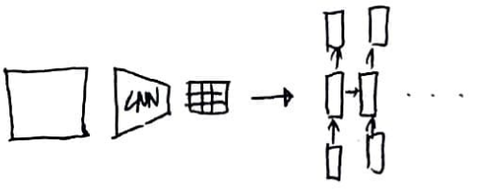
→ XLM

* RNN & CNN

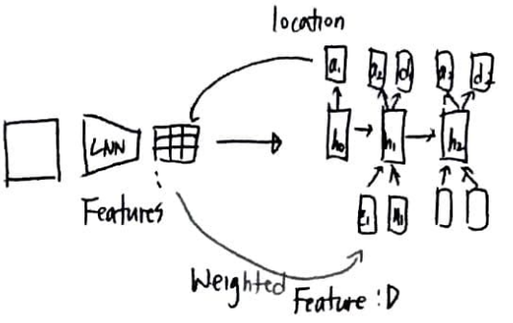
ex 1) Image captioning.

Image to Sentence

CNN → RNN

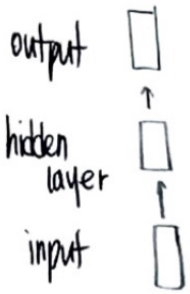


+ Image Captioning with Attention

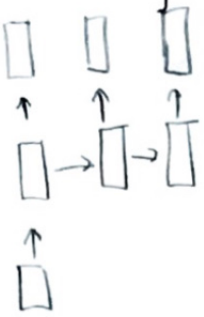


• Recurrent Neural Network : 순환신경망 (신경망에 순환고리)

one to one



one to many

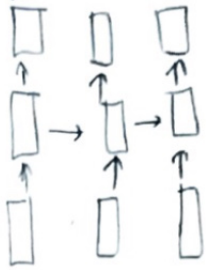


ex image captioning (decode)

many to one

sentiment classification (encode)

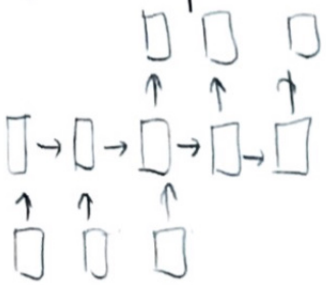
many to many



video classification on frame level

sequence 처리

many to many



ex Machine Translation

+ Non-sequence data → sequential processing

ex 이미지

1) =