KAIST Summer Session 2018

Module 3. Deep Learning with PyTorch

Recurrent Neural Network

KAIST College of Business

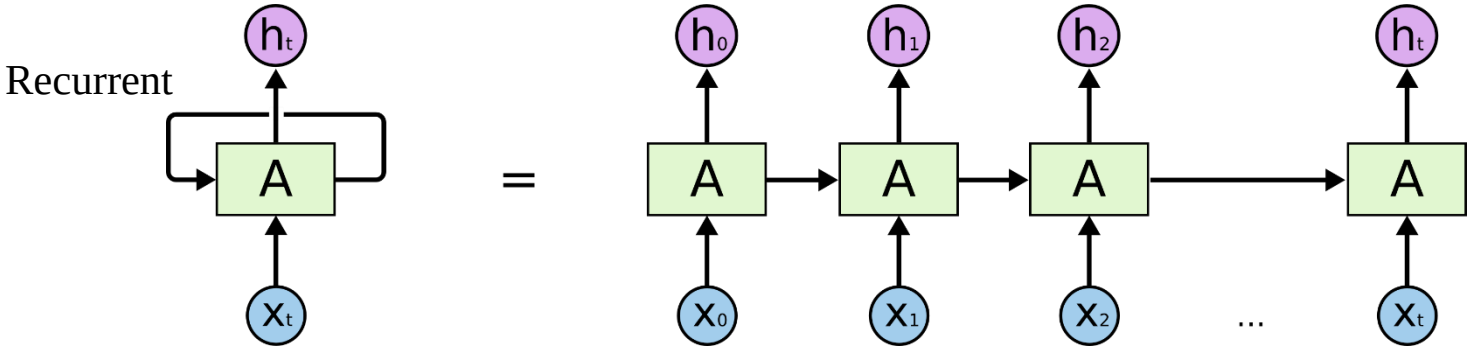
Jiyong Park

23 August, 2018

Review: What is RNN?

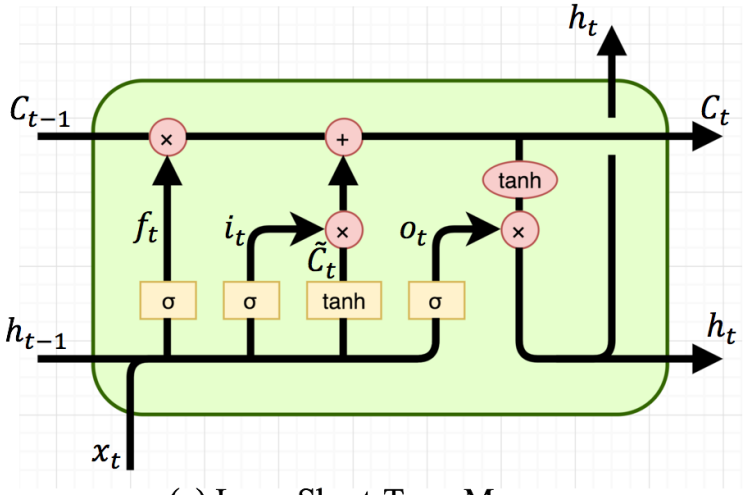
- Unlike CNN, RNN can use past information to learn the present task.
  - Example: Natural Language Processing (NLP)
    - "The clouds are in the ( )."
    - "I grew up in France …… I speak fluent ( )."
  - Vanishing gradient problem
    - As that gap grows, RNN becomes unable to learn to connect the information.

      (the past information would be vanishing or exploding)

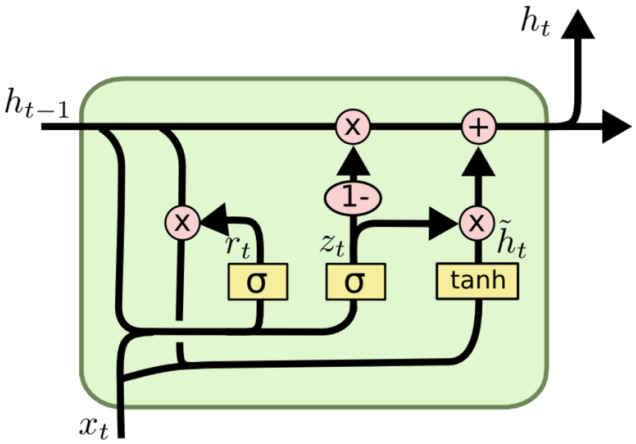


Review: What is RNN?

- Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) allows RNN to learn how much past information would pass to the next.



(a) Long Short-Term Memory



(b) Gated Recurrent Unit

RNN in PyTorch

- Although RNN and GRU yield two states (output and hidden), LSTM yields three states (output, hidden, and cell).

```
output, hidden = self.rnn(input,hidden)
```

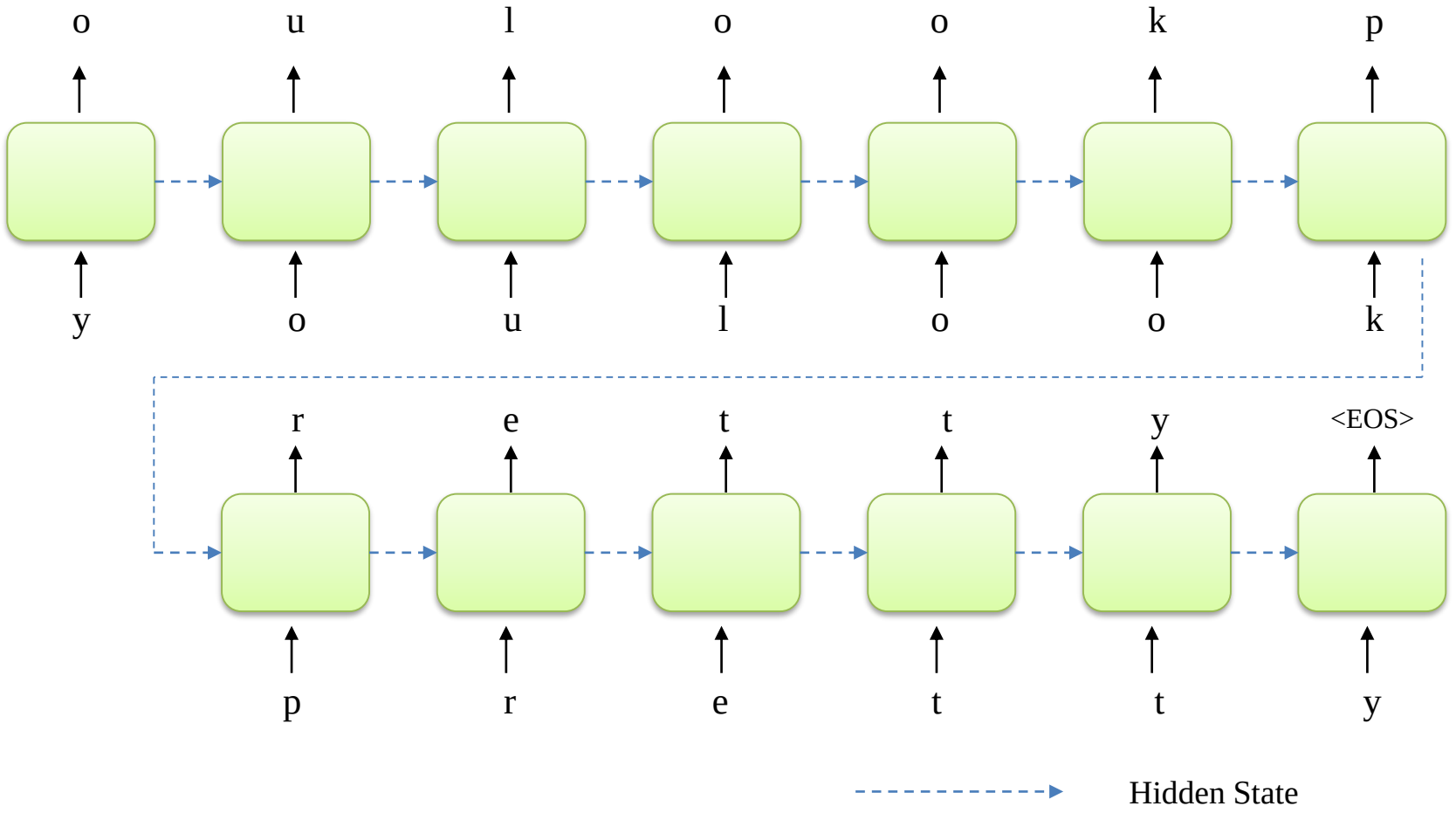
```
output, hidden = self.gru(input,hidden)
```

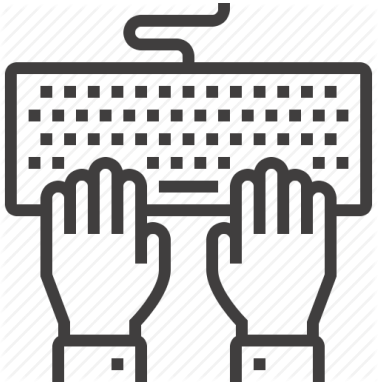
```
output,(hidden, cell) = self.lstm(input,(hidden, cell))
```

Let's Make a Reply Bot (답정너 봇)

Let's Make a Reply Bot (답정너 봇)

- We want to make the Reply Bot_v1 to always reply that "you look pretty."





Reply Bot_v1

using Recurrent Neural Network

M3.6 Recurrent Neural Network_Reply Bot.ipynb

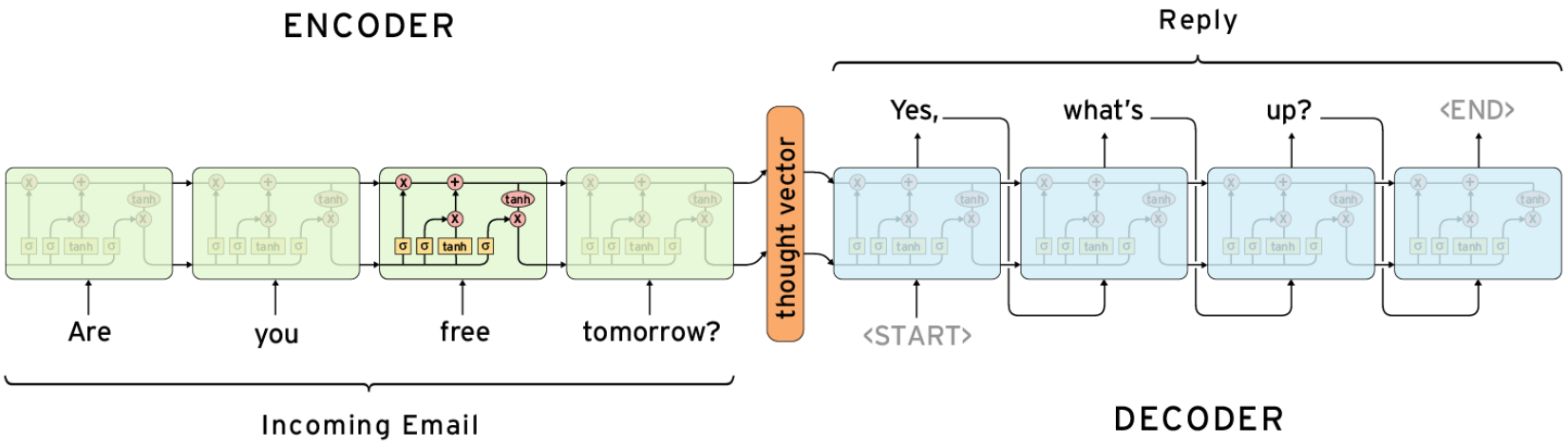
We Want More for the Reply Bot…

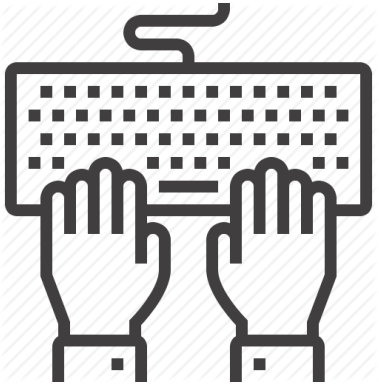
- Our Reply Bot_v1 does not understand the contexts. Let's make our Reply Bot more smart!

Sequence-to-Sequence Model

- Our understanding of contexts
  - (Listening/Reading) → (Understanding) → (Speaking/Writing)

- Understanding of sequence-to-sequence models
  - (Encoding) → (Thought Vector) → (Decoding)





Reply Bot_v2

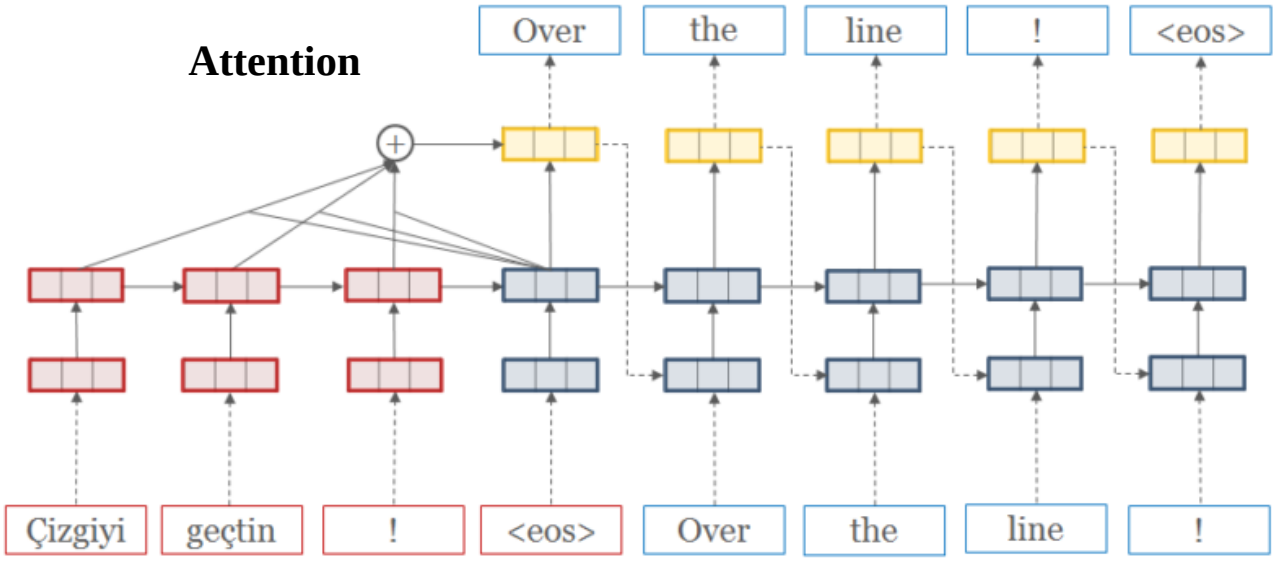
using Sequence-to-Sequence

M3.6 Sequence-to-Sequence_Reply Bot.ipynb

Projects for RNN

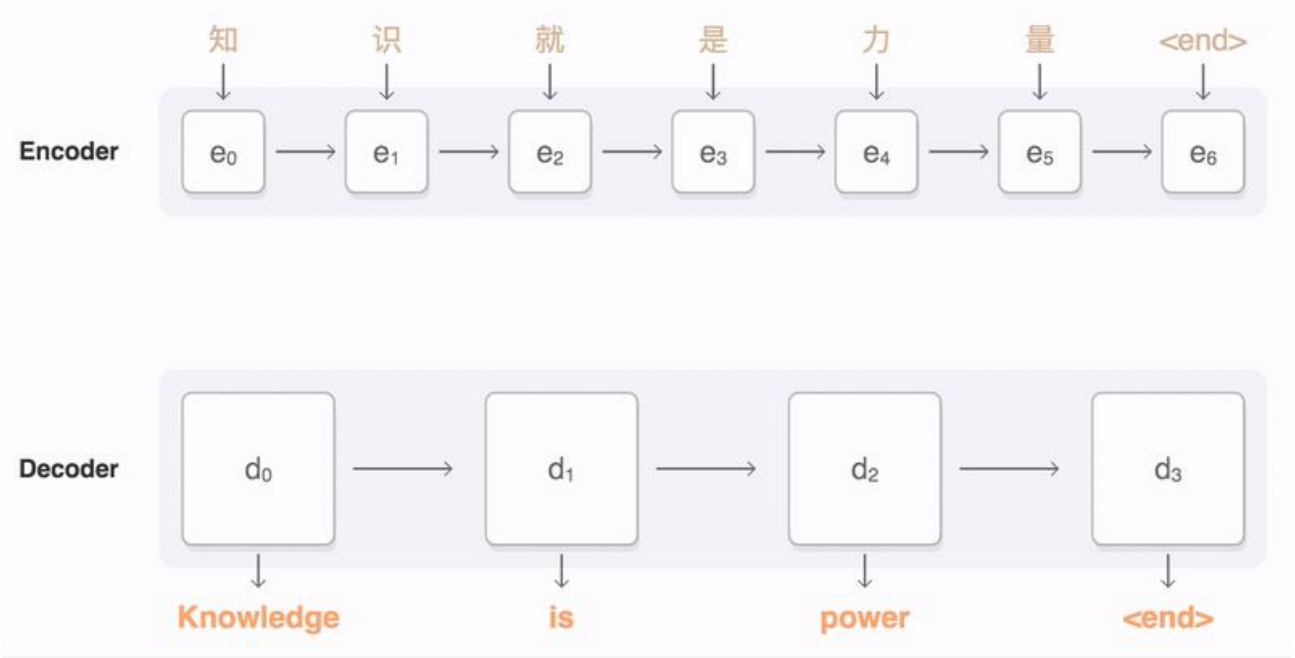
Neural Machine Translation

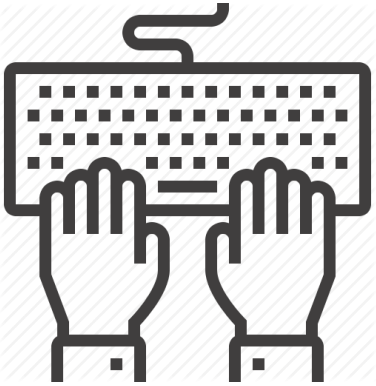
- Sequence-to-sequence models lie at the core of machine translation (e.g., Google Translate), as well as other end-to-end tasks (e.g., chat bot)
  - (Encoding) → (Thought Vector) → (Decoding)



Neural Machine Translation

- Sequence-to-sequence models lie at the core of machine translation (e.g., Google Translate), as well as other end-to-end tasks (e.g., chat bot)
  - (Encoding) → (Thought Vector) → (Decoding)

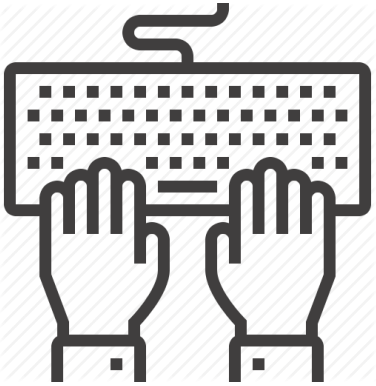




Neural Machine Translation using Sequence-to-Sequence with Attention

M3.6 Sequence-to-Sequence with Attention

_Machine Translation.ipynb



Simple Chat Bot using Sequence-to-Sequence with Attention

M3.6 Sequence-to-Sequence with Attention

_Chat Bot.ipynb

End of Document