cs224n: MT, Seq2Seq, Attention

자연어 심화세미나

TOBIG'S 20기 박준

Contents

20기 자연어 심화세미나
TOBIG'S 20기 박준

Unit 01

Machine Translation

Unit 01 | Machine Translation

Source Language Sentence x: *L'homme est né libre, et partout il est dans les fers*

Target language Sentence y: *Man is born free, but everywhere he is in chains*

Unit 02 | Statistical Translate Model

20기 자연어 심화세미나
TOBIG'S 20기 박준

$$\text{argmax}_y P(y|x)$$

얼마나 주어진 x의 문장에 대
해서 최고의 y를 찾는 문제.

$$= \text{argmax}_y \underbrace{P(x|y)}\underbrace{P(y)}$$

Translation Model

Models how words and phrases should be translated (*fidelity*). Learned from parallel data.

Language Model

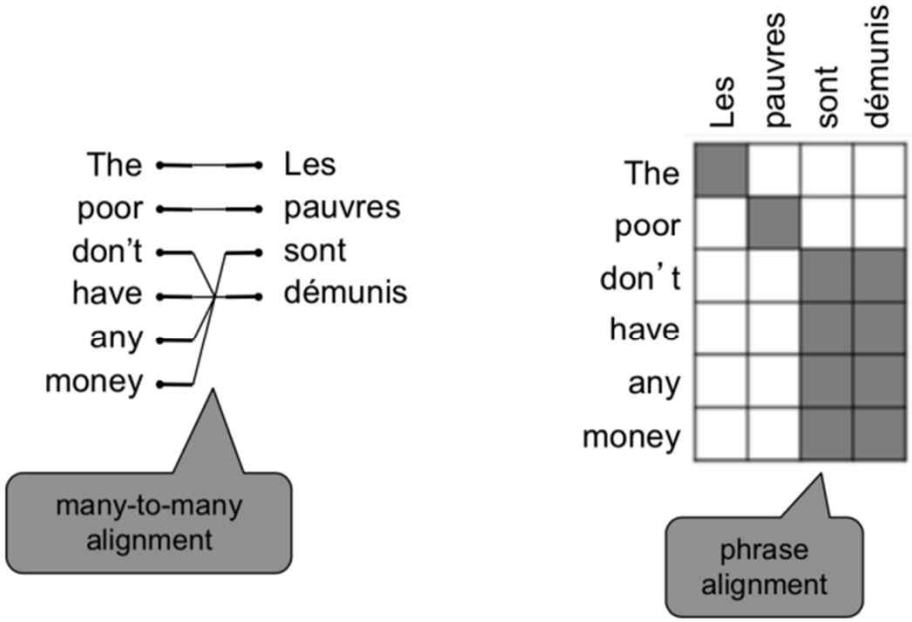
Models how to write good English (*fluency*). Learned from monolingual data.

Bayes rule을 이용해서 두개의
문제로 나누어서 품.

Unit 02 | Statistical Translate Model

$P(x|y)$ → $P(x, a|y)$

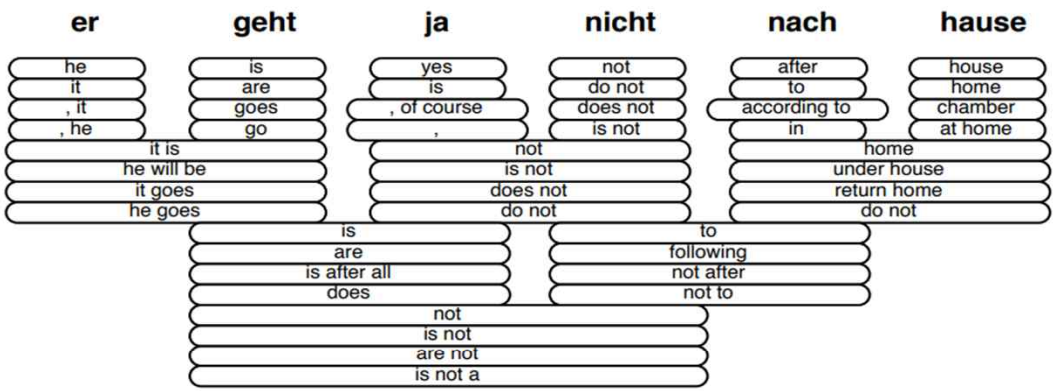
x,y의 문장 사이의 단어의 일치하는 지의 변수 추가.

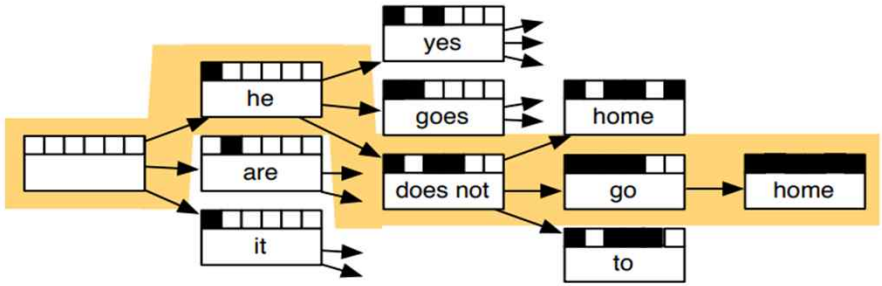


단순히 1:1로 대응되지 않는다. complex

Unit 02 | Statistical Translate Model

$$\text{argmax}_y P(x|y)P(y)$$





Target language의 단어에 대응되는 source language의 단어각자의 대응되는 것들을 찾음.

그 단어들을 가지고 가능성을 판단.

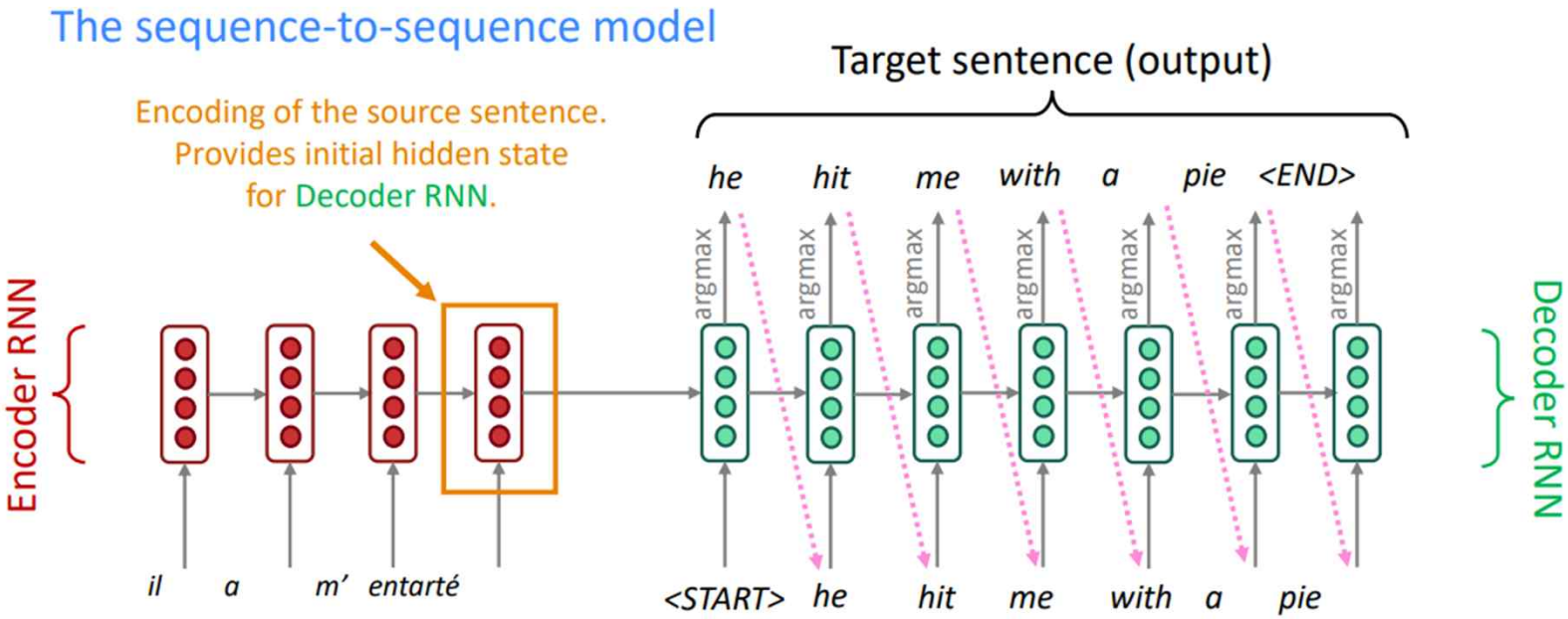
단점

- 너무 복잡하고, 인간이 할 것이 많다.
- 수많은 feature engeenering.
- 계속 쌍들을 최신화 할 필요가 있음.

Unit 03 | Neural Machine Translation

20기 자연어 심화세미나
TOBIG'S 20기 박준

번역을 위해 RNN 두개를 붙인 sequence-to-sequence model



Unit 03 | Neural Machine Translation

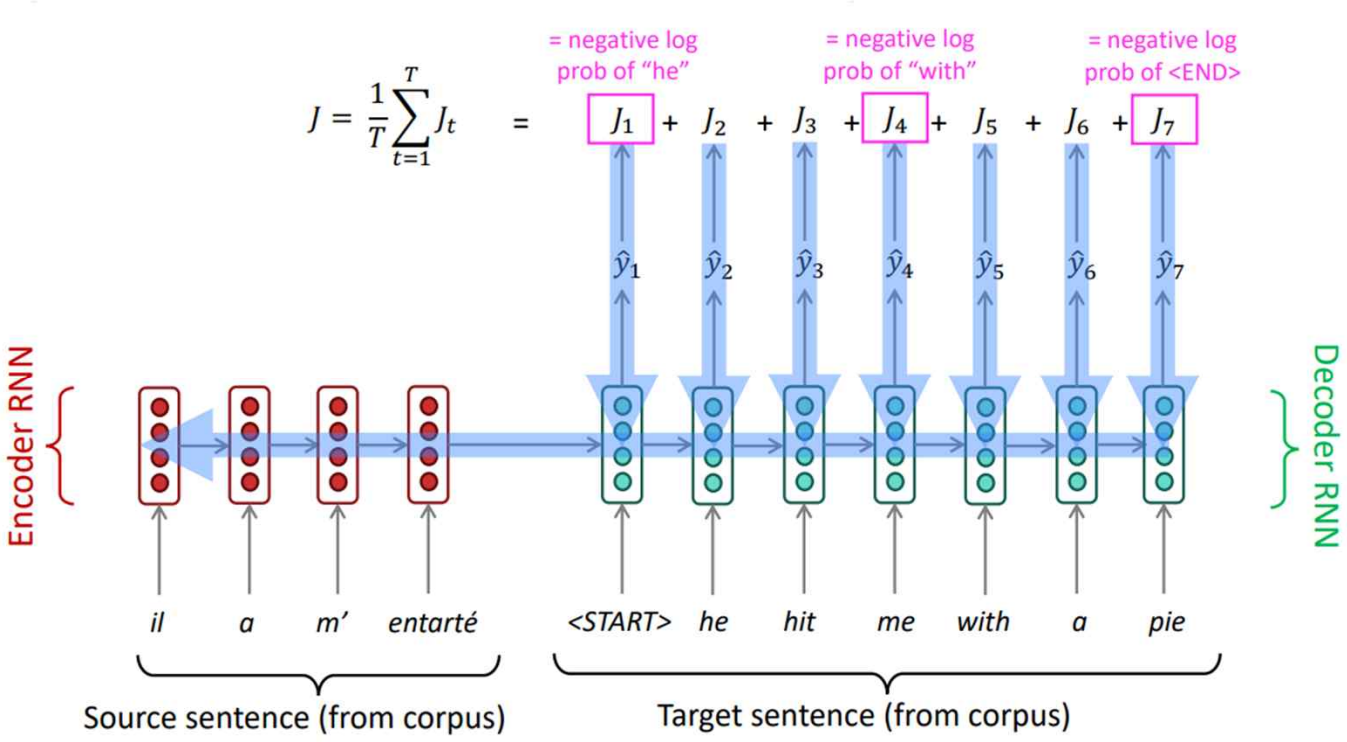
- NMT directly calculates $P(y|x)$:

$$P(y|x) = P(y_1|x)\,P(y_2|y_1,x)\,P(y_3|y_1,y_2,x)\ldots\underbrace{P(y_T|y_1,\ldots,y_{T-1},x)}$$

Probability of next target word, given target words so far and source sentence *x*

직접 P(y/x)를 구하고, 다음 목표 단어의 조건이 주어진
x문장에 그 전 단어들까지 포함됨.

Unit 03 | Neural Machine Translation

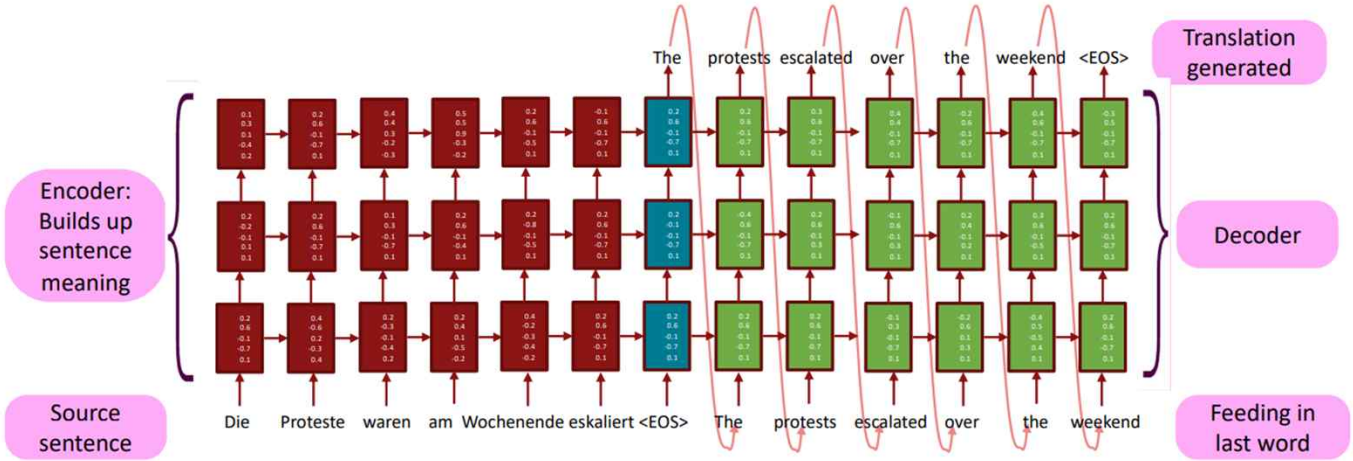


decorder에 있는 함수를 학습하며 역전파를 통해 encoder 끝까지 학습 가능.

Unit 03 | Neural Machine Translation

Multi-layer RNN

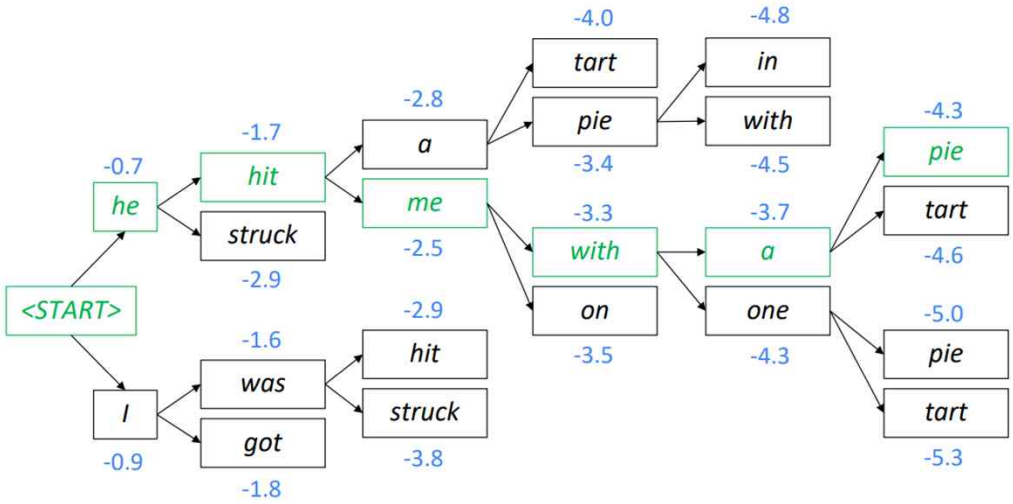
- Rnn을 여러겹 쌓으면서 좀 더 깊게 만듬.
- 더 복잡한 표현 가능->higher level features. 2~4개가 적당.
- 2층이 1층의 rnn보단 훨씬 좋지만, 3층부터는 그 향상이 적다.



Unit 03 | Neural Machine Translation

탐색 방법

1. Greedy search.-> 그때마다의 최선을 선택. 하지만 한번 생성후에는 돌이킬 수 없기 때문에 한번 잘못 출력하면 문장 전체에 영향을 줄 수 있음.
2. 전체 모든 가능한 y를 전부 계산-> 계산양이 엄청나게 크다.
3. Beam search-> 각 단계에서 score가 높은 것들을 beam size만큼 유지하며 따라감.



$$\text{score}(y_1, \ldots, y_t) = \log P_{\text{LM}}(y_1, \ldots, y_t | x) = \sum_{i=1}^{t} \log P_{\text{LM}}(y_i | y_1, \ldots, y_{i-1}, x)$$

마지막 score에서 1/t를 곱해주며 정규화를 시켜준다. 길이에 따라 달라지기 때문에.

장점

- 성능이 더 좋다.
- 한번에 계산할 수 있다.
- 인간의 할 일을 줄일 수 있다. Feature engeenering.

단점

- 결과에 대한 해석이 어렵다.(이것을 관리하기와 개선하기 힘들다).

Unit 03 | How to Evaluate

BLEU(BiLingual Evaluation Understudy)

- 인간의 번역과 기계의 번역을 비교하여 그것의 유사도를 계산.
- N-gram precision + too short penalty 사용.

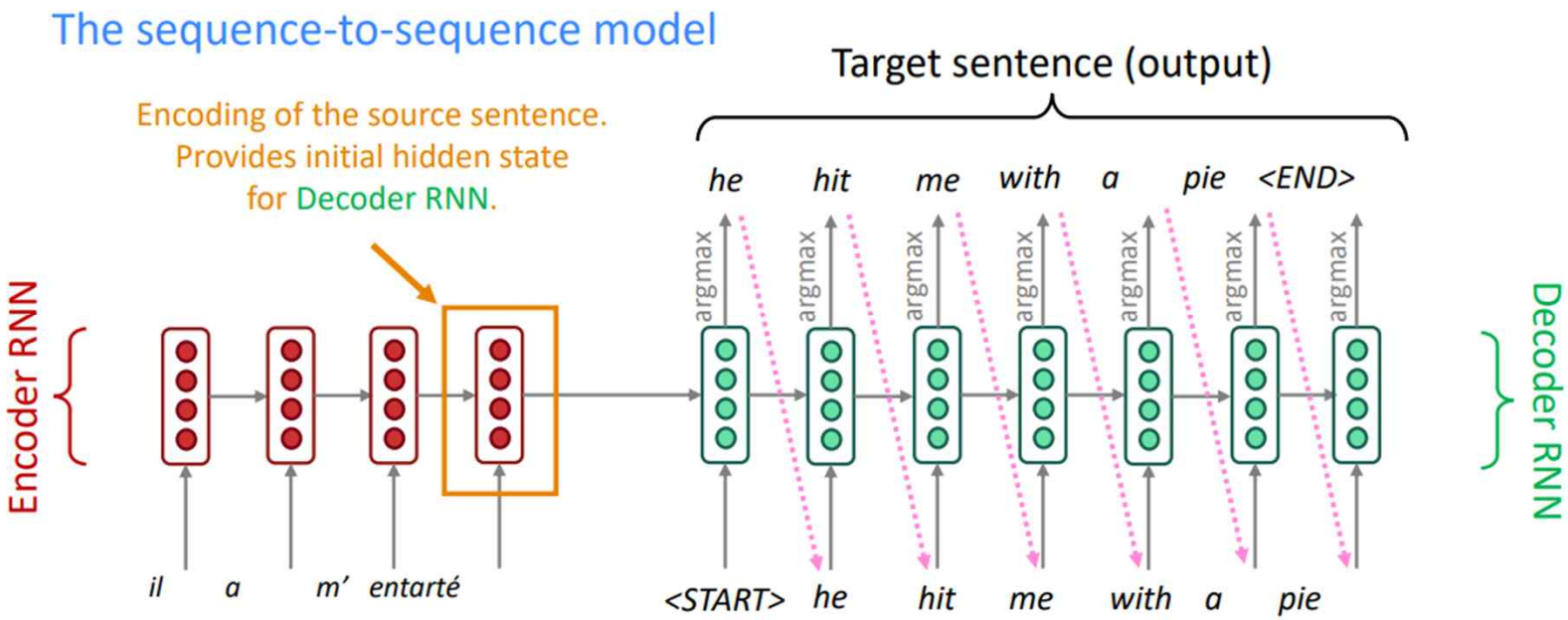
$$\mathrm{BP} = \begin{cases} 1 & \text{if } c > r \\ e^{(1-r/c)} & \text{if } c \leq r \end{cases}$$

$$BLEU = \mathrm{BP} \cdot \exp\left(\sum_{n=1}^{N} w_n \log p_n\right)$$

Unit 03 | Neural Machine Translation

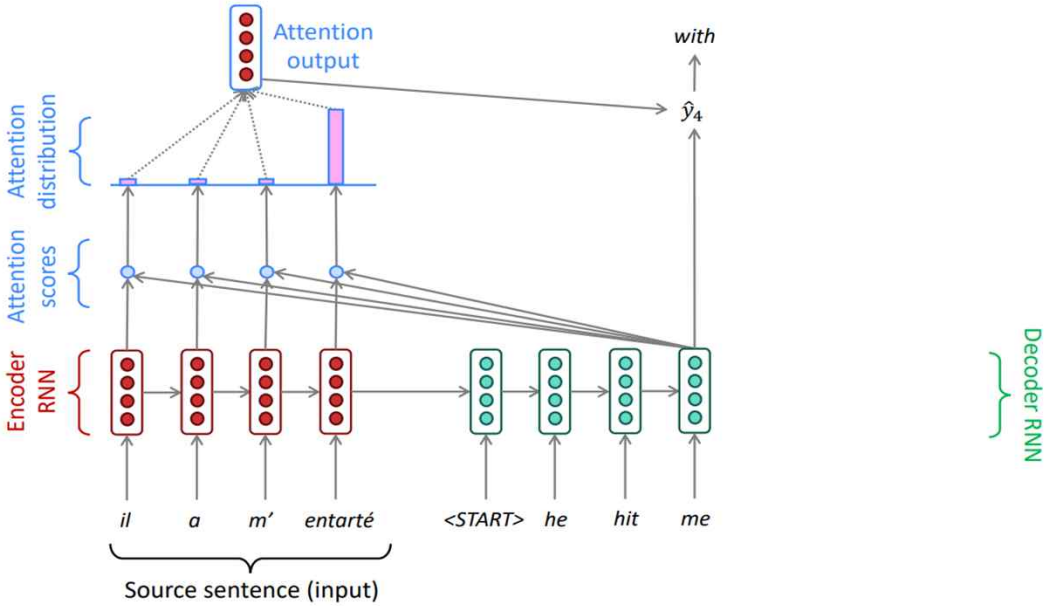
문제점

- Bottleneck problem(수 많은 정보들이 하나의 벡터에 저장됨)



해결책

- 단순히 하나의 벡터로 연결되어 있는 것이 아니라 decoder에서의 한 부분과 모든 encoder의 부분 중 어느 부분이 이 부분과 유사한지 파악.



Unit 04 | Attention

수식

$$h_1, \ldots, h_N \in \mathbb{R}^h$$ 인코더의 hidden state

$$s_t \in \mathbb{R}^h$$ t시점의 decoder hidden state

$$\boldsymbol{e}^t = [\boldsymbol{s}_t^T \boldsymbol{h}_1, \ldots, \boldsymbol{s}_t^T \boldsymbol{h}_N] \in \mathbb{R}^N$$ attention score 소스 문장과 같은 길이의 벡터

$$\alpha^t = \mathrm{softmax}(\boldsymbol{e}^t) \in \mathbb{R}^N$$ attention distribution

$$\boldsymbol{a}_t = \sum_{i=1}^{N} \alpha_i^t \boldsymbol{h}_i \in \mathbb{R}^h$$ attention output(weighted sum) 인코더의 hidden state와 같은 크기의 벡

$$[\boldsymbol{a}_t; \boldsymbol{s}_t] \in \mathbb{R}^{2h}$$ $\hat{y}$

Unit 04 | Attention

Attention의 종류

- Basic dot product attention $e_i = s^T h_i \in \mathbb{R}$
- Multiplicative attention $e_i = s^T W h_i \in \mathbb{R}$
- Addictive attention $e_i = v^T \tanh(W_1 h_i + W_2 s) \in \mathbb{R}$

장점.

- 더 좋은 성능을 보이고, bottleneck 문제를 극복.
- Vanishing gradient problem에서도 좋은 효과.

Reference

Stanford CS224N NLP with Deep Learning | Winter 2021 | Lecture 7 - Translation, Seq2Seq, Attention

: https://www.youtube.com/watch?v=wzfWHP6SXxY&list=PLoROMvodv4rOSH4v6133s9LFPRHjEmbmJ&index=7

https://ladun.tistory.com/71 bleu

***All Images without clarified source are retrieved on the above reference.**

감사합니다