

논문 발표 (2) - Neural machine translation by jointly learning to align and translate

📅 Date	@September 13, 2021
📌 Status	

Introduction to Attention

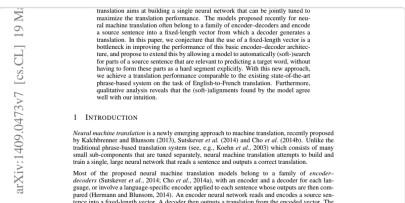


Neural machine translation by jointly learning to align and translate

▼ paper

Neural machine translation by jointly learning to align and translate.pdf

https://drive.google.com/file/d/10A-MGOWtdJXMMuqkgzMi5_Ei1wa7jR3d/view?usp=sharing

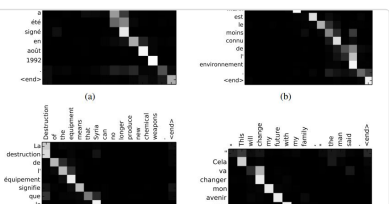


▼ references

NMT By Jointly Learning To Align And Translate 리뷰

5년이나 지났지만 기초부터 다시자는 생각에 이 논문을 시작으로 NLP 관련 논문들을 부족하지만 리뷰해보고자 합니다. 틀린 내용에 대한 이야기는 언제든지 환영합니다. 이 논문 이전의 Machine Translation에서 쓰이던 방법들은 주로 Statistical Machine Translation였

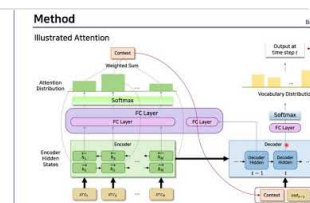
<https://heiwais25.github.io/nlp/2019/06/18/neural-machine-translation-by-jointly-learning-to-align-and-translate/>



[Paper Review] Neural Machine Translation by Jointly Learning to Align and Translate

1. 발표자: 석사과정 김탁영2. 논문: <https://arxiv.org/abs/1409.04733>. 발표자료: https://drive.google.com/file/d/1UQpdZP1C_w_NfL1iHXciWtTfMqFemwTK/view?usp=sharing

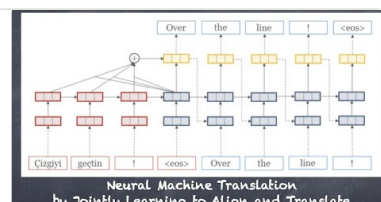
<https://youtu.be/S2msiG9g7Us>



주정현 - Neural Machine Translation by Jointly Learning to Align and Translate

딥러닝논문스터디 - 47번째 자연어처리 주정현님의 'Neural Machine Translation by Jointly Learning to Align and Translate' 입니다.모임 참여 및 문의는 <https://forms.gle/KhJQEWirGfM2G5a68> 으로 ...

<https://youtu.be/I9pWT6BHpj0>



[딥러닝 기계번역] 시퀀스 투 시퀀스 + 어텐션 모델

시퀀스 투 시퀀스 + 어텐션 (seq2seq2 + attention)을 활용한 기계번역에 대해 알아봅니다.

주요 개념1. seq2seq2. attention3. teacher forcing참고자료:NEURAL MACHINE TRANSLATIONBY JOINTLY LEARNING T...

<https://youtu.be/WsQLdu2JMgl>

시퀀스 투 시퀀스 + 어텐션

인공신경망 기계번역

개념 요약

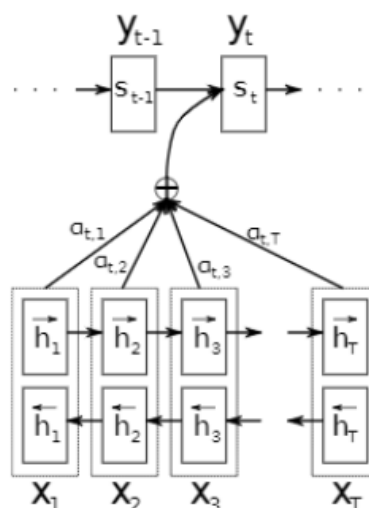
- Bahdanau Attention
- encoder-decoder 모델에서 고정된 길이의 context vector 를 생성하므로 발생하는 bottleneck 문제를 해결하기 위해 Attention(align and translate) 이라는 새로운 기법 도입

Backgrounds

- Statistical Machine Translation을 지나 Neural Machine Translation이 등장하던 시기였음
- Neural Machine Translation의 대표적인 구조가 Encoder-decoder 구조였음
- 하지만 기존의 Encoder-decoder 구조는 Encoder 에서 고정된 길이의 context vector 를 생성해서 사용했기 때문에 bottleneck 문제가 발생했다.
 - 입력 문장의 길이가 길어질수록 충분한 정보를 담을 수 없다.

Main Idea

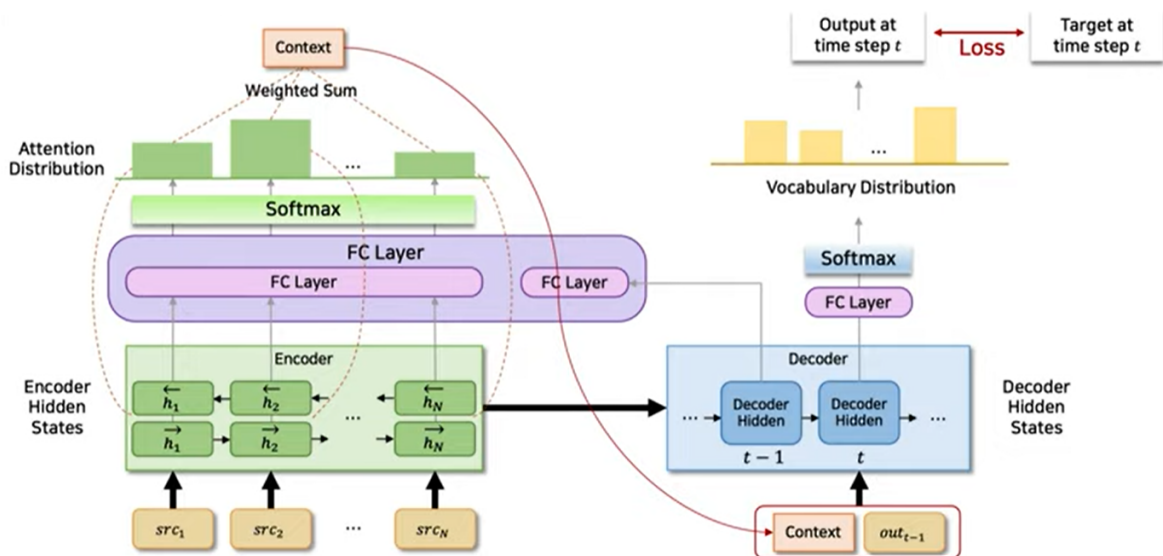
- RNN ENcoder-decoder를 기반으로 하고 있음
- 이를 해결하기 위해 **Alignment** 구조를 제안했다.



- **Alignment**
 - 출력 문장의 단어가 입력 문장의 어느 부분에 주목해야 하는 지에 대한 개념

- 이를 계산하기 위해 Encoder의 각 time step의 hidden state vector와 Decoder의 hidden state vector를 사용한다.

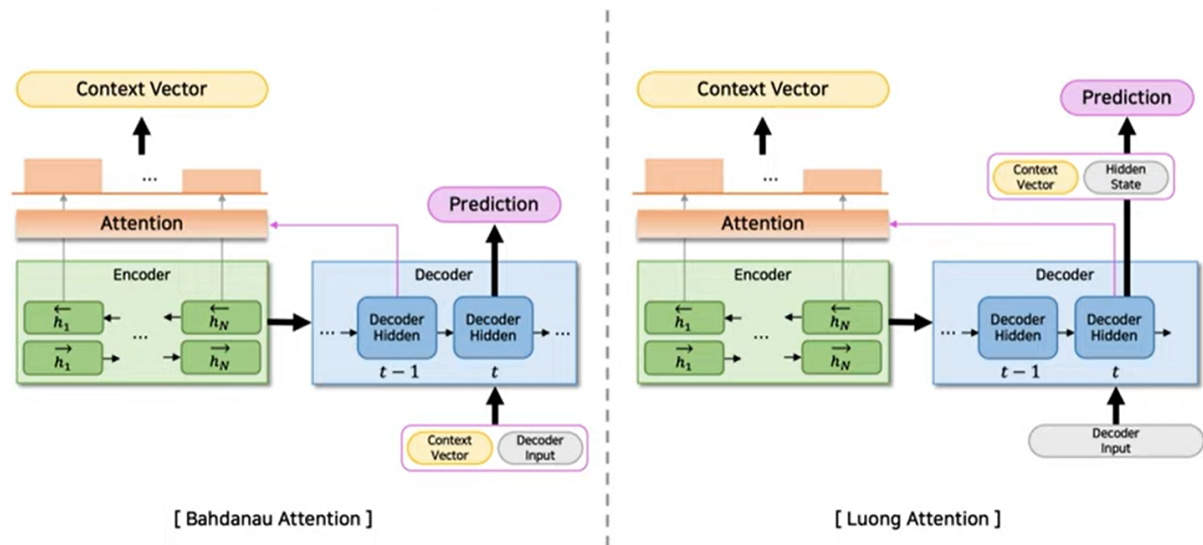
Bahdanau Attention Process



<https://youtu.be/S2msiG9g7Us>

- Decoder의 t 시점의 출력 단어를 생성하기 위해 Decoder의 $t-1$ 시점의 hidden state vector와 Encoder의 hidden state vector들을 이용한다.
- 이들을 각각 FC Layer를 통과시킨 다음 이 둘을 더해준 다음 tanh 를 적용한 후 한번 더 FC Layer를 통과시켜 Attention scores 를 산출한다.
- Attention scores를 softmax 를 적용하여 Attention Distribution을 만든다.
- 이 Attention Distribution을 이용하여 Encoder의 hidden state vector들을 가중합(Weighted Sum)하여 Context vector를 생성한다.
- 이렇게 생성한 context vector를 Decoder의 t 시점의 입력으로 사용한다.

cf) Luong Attention vs Bahdanau Attention



<https://youtu.be/S2msiG9g7Us>

Luong Attention

- Decoder에서 t 시점의 단어를 예측하기 위해 t 시점의 hidden state vector를 사용
- Attention scores 계산 시 dot product 이용
- Context vector가 Decoder의 t 시점의 출력과 결합하여 예측

Bahdanau Attention

- Decoder에서 t 시점의 단어를 예측하기 위해 $t-1$ 시점의 hidden state vector를 사용
- Attention scores 계산 시 FC Layer 이용
- Context vector가 Decoder의 t 시점의 입력으로 사용하여 예측


Experiments

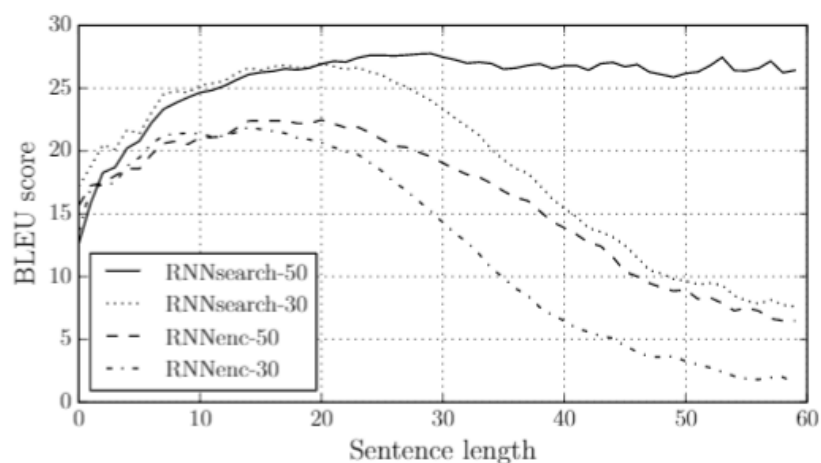
Dataset

- WMT' 14 (English \rightarrow French)
- 영어를 불어로 번역하는 task
- All : 모든 문장을 대상으로 실험
- No UNK : unknown token 이 없는 문장들을 대상으로 실험

Results

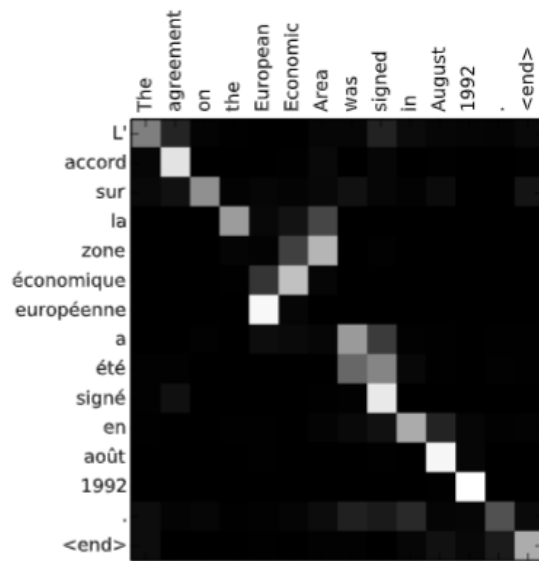
Model	All	No UNK ^o
RNNencdec-30	13.93	24.19
RNNsearch-30	21.50	31.44
RNNencdec-50	17.82	26.71
RNNsearch-50	26.75	34.16
RNNsearch-50*	28.45	36.15
Moses	33.30	35.63

- RNNencdec
 - 기존의 RNN Encoder-decoder 구조의 모델
 - 뒤의 숫자(30, 50)은 sequence length를 의미
- RNNsearch
 - Bahdanau Attention 모델
 -  가 붙은 모델은 다른 모델보다 더 오래 학습시킨 모델 (성능 향상이 없을 때 까지)
- Moses
 - PBSMT: Phrase-Based Statistical Machine Translation
- baseline인 RNN Encoder-decoder 구조의 모델보다 Bahdanau Attention 모델이 전반적으로 좋은 성능을 보인 것을 볼 수 있다.
- 더 긴 길이의 시퀀스를 사용한 RNNencdec-50 보다 짧은 시퀀스를 사용한 RNNsearch-30 이 더 좋은 성능을 보인 것 또한 확인할 수 있다.

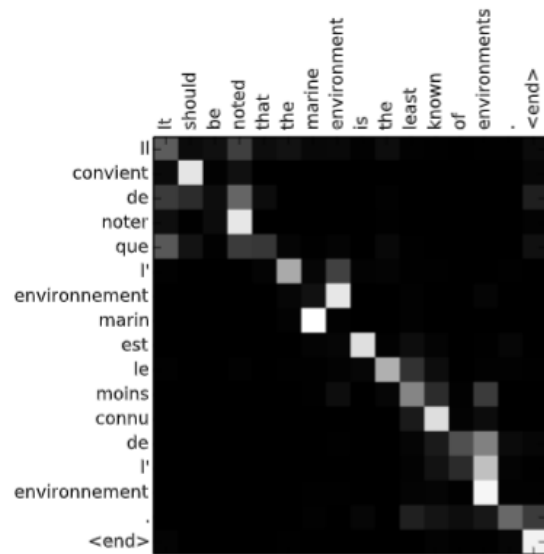


- RNNsearch-50 이 시퀀스의 길이가 계속 길어져도 좋은 성능을 보이고 있다.

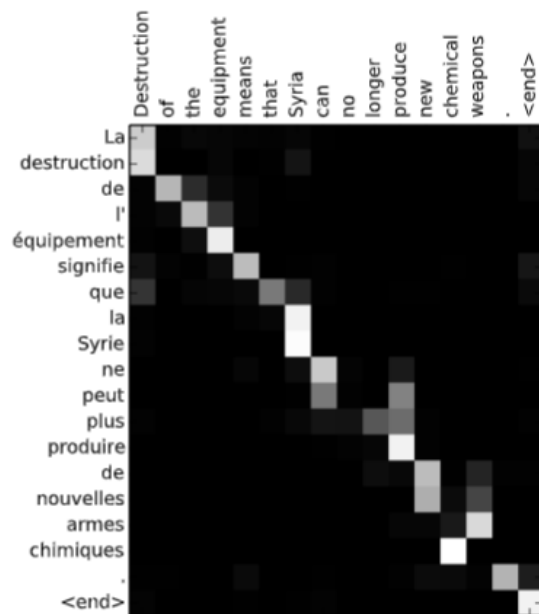
Qualitive Analysis



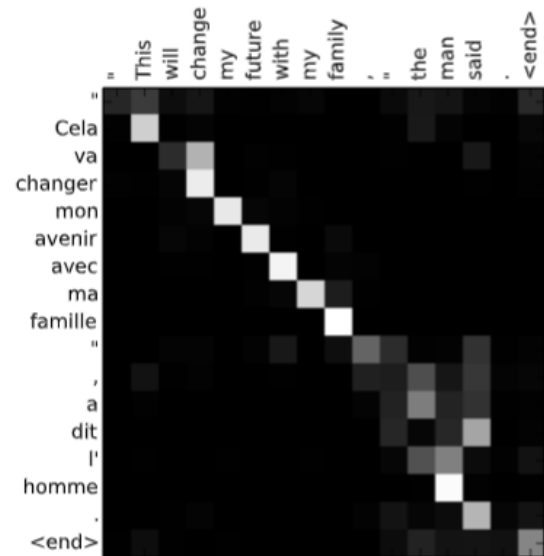
(a)



(b)



(c)



(d)

- RNNsearch-50 으로 실험한 결과
- 영어 문장을 불러 문장으로 변환할 때 불러 문장의 특정 단어 생성 시 영어 문장의 어떤 단어를 참고했는 지를 시각화하여 보여준다.
- 어순의 차이도 고려해서 예측하는 모습을 볼 수 있다.