

Seq2seq Attention: Neural Machine Translation by jointly learning to align and translate

최성욱

목차

1. Introduction

2. Approach

2.1 Original Encoder–Decoder Model

2.2 Proposed Encoder–Decoder Model

3. Experiments

4. Results

4.1 Quantitative Results

4.2 Qualitative Results

5. Conclusion

1. Introduction

- Why do we need to use this proposed model?

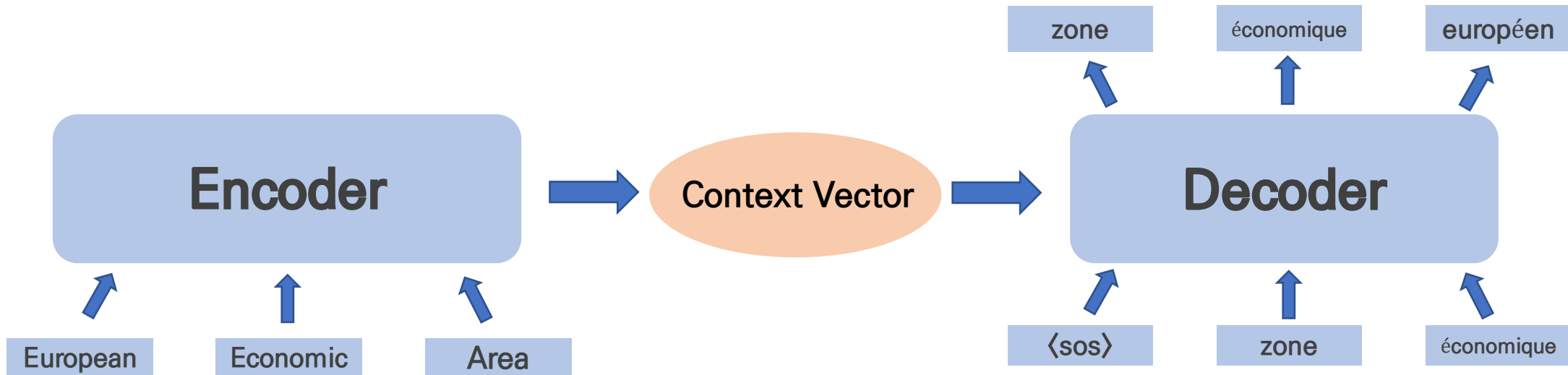
1. 무조건 사용할 정도로 성능이 좋은 것은 아니다.
2. 하지만, Neural Machine Translation에 있어서 큰 발전에 기여할 것이다.



1. 기존 Encoder-Decoder 기반의 Neural Machine Translation 모델에 비해 성능이 좋다.
2. Single Neural Network만을 활용해서 전통적인 Statistical Machine Translation에 버금가는 성능을 달성했다. (기존에는 SMT에 NMT를 더하는 방식으로 SOTA를 달성했다.)

2. Approach

· Original Encoder-Decoder Model



1. Train

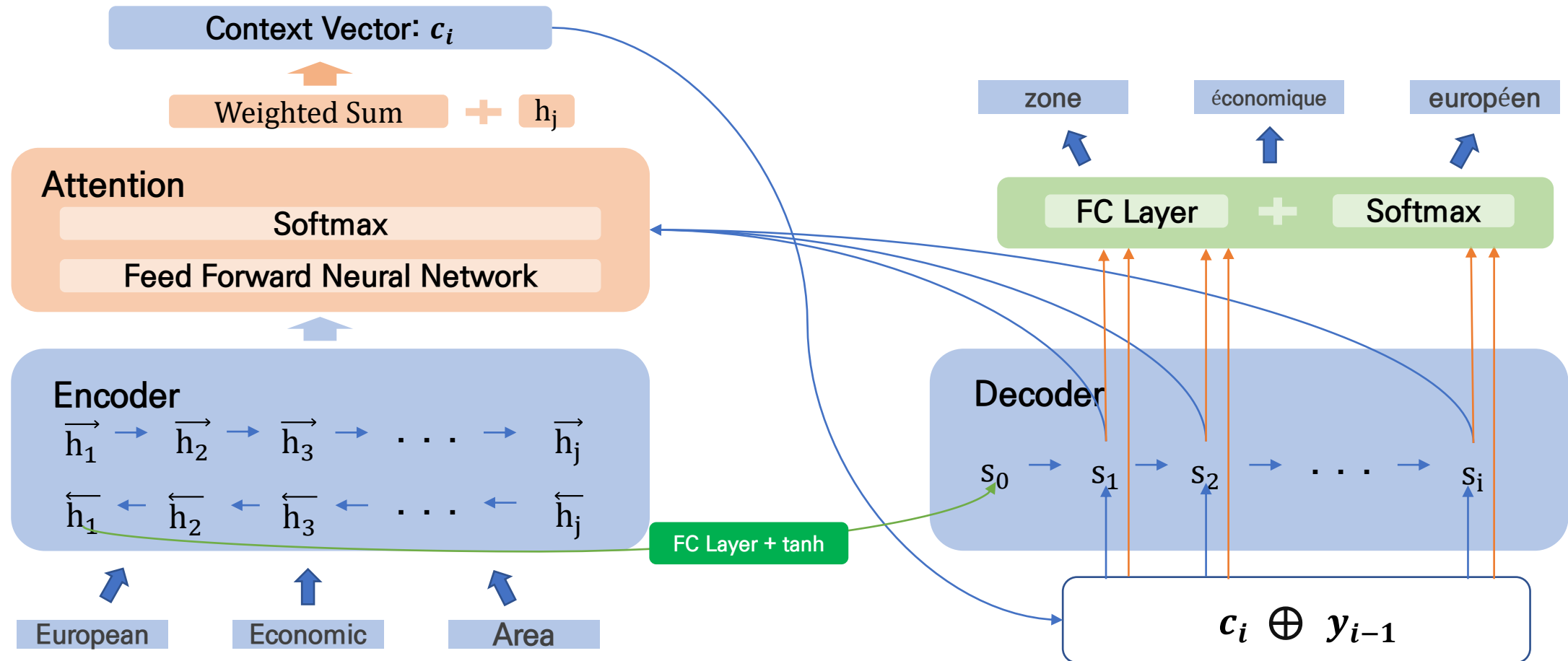
- Source Sentence에 대하여, 올바른 Translation Sentence가 나올 확률을 최대화 하는 방식으로 진행

2. Cons

- Context Vector가 Fixed Length Vector
- 이로 인해, 모든 정보를 고정된 길이의 벡터로 압축하면서 정보의 손실 발생
- 학습 Corpus보다 긴 문장이 들어왔을 경우, 성능이 좋지 않다.

2. Approach

- Proposed Encoder-Decoder Model



3. Experiments

· Dataset

1. WMT 14 – Total 850M words
2. **Data Selection Method – Actual Using Words 348M**
3. News-test-2012, News-test-2013이 각각 Validation & Test Set
4. **Vocab Size: 가장 빈번하게 나온 단어 기준 30000개**
5. **Not Specialized Preprocessing**

· Model

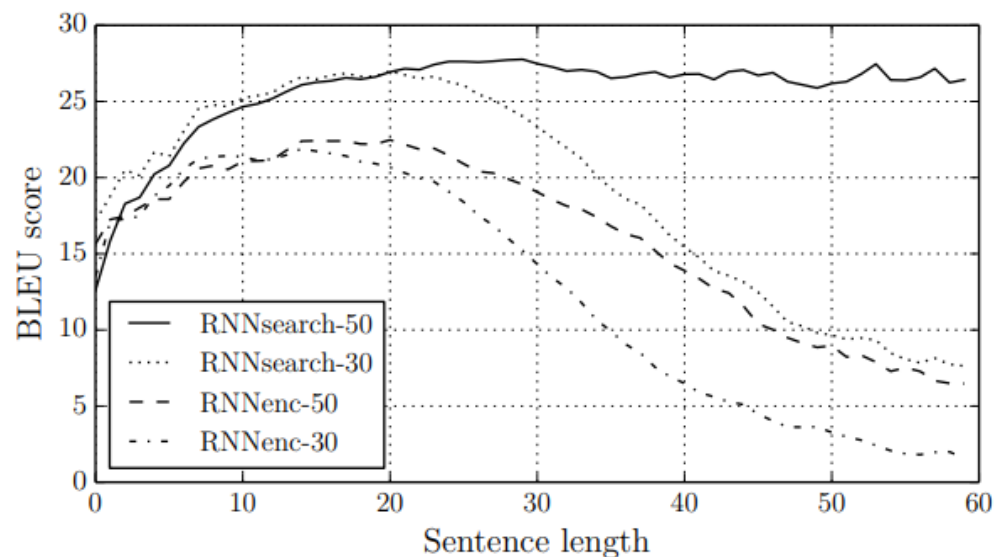
1. RNNencdec Model과 비교실험
2. **문장의 길이를 30단어와 50단어로 구분 지어 비교실험**
3. **Single Maxout Hidden Layer 활용**
4. Optimizer: Adadelata
5. Mini Batch with 80 Sentence
6. **번역 결과의 조건부 확률을 최대화 하기 위하여 Beam Search 활용**

4. Results

Quantitative Results

1. RNNsearch (Proposed Model) 의 성능이 전체적으로 더 좋다.
2. UNK token이 없는 문장들만 고려했을 경우, **Phrases-Based 모델의 성능에 버금간다.**
3. Moses의 경우, Monolingual Corpus 418M을 사용한 것을 고려할 때, RNNsearch의 성능은 더욱 유의미하다.
4. **A Fixed Length Vector**를 활용한 RNNencdec의 경우, 문자의 길이가 길어질수록 성능이 급격히 감소한다.
5. RNNsearch의 경우, 문장의 길이가 50단어보다 더 길어져도 성능의 악화를 크게 보이지 않는다.

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



4. Results

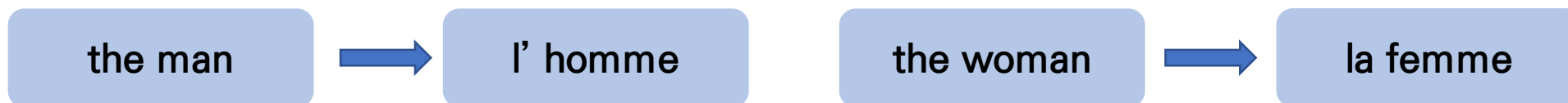
· Qualitative Results

1. Alignment

1. **Source Sentence**와 **Target Sentence** 사이의 **attention**을 직관적으로 볼 수 있다.
2. English & French 사이의 Alignment가 단조롭다.
3. Hard alignment가 아니라 **Soft alignment**를 함으로서, 단어를 건너 뛰어가며 전체 문장을 번역해 나갈 수 있다.
4. Source Sentence와 Target Sentence 사이의 길이의 차이에 대하여 **NULL**로 매핑되지 않는다.

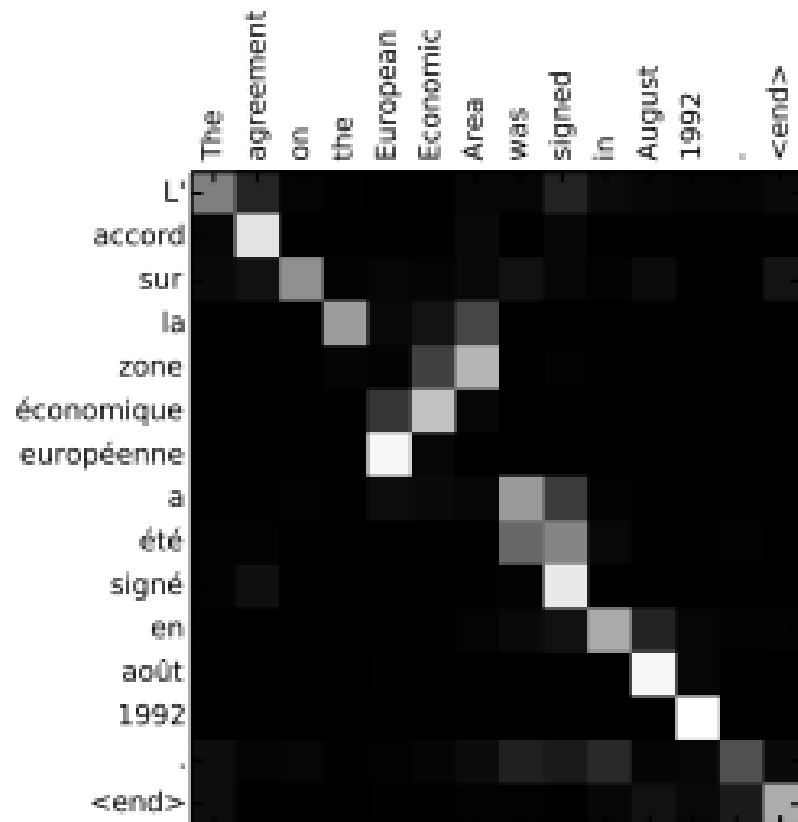
2. Long Sentences

1. 긴 문장을 고정된 길이의 Vector로 encoding할 필요가 없다.
2. **Proposed Model**이 **Basic Model**보다 긴 문장에 대하여 번역을 잘 한다는 가설을 확인할 수 있다.



4. Results

Qualitative Results



Input Sentence

An admitting privilege is the right of a doctor to admit a patient to a hospital or a medical centre to carry out a diagnosis or a procedure, based on his status as a health care worker at a hospital.

RNNencdec

Un privilège d'admission est le droit d'un médecin de reconnaître un patient à l'hôpital ou un centre médical d'un diagnostic ou de prendre un diagnostic en fonction de son état de santé.

based on his state of health

RNNsearch

Un privilège d'admission est le droit d'un médecin d'admettre un patient à un hôpital ou un centre médical pour effectuer un diagnostic ou une procédure, selon son statut de travailleur des soins de santé à l'hôpital.

5. Conclusion

- ✓ 긴 문장을 번역하는데 있어, 고정된 길이의 **Context Vector**를 사용하는 것은 성능 저하의 원인이라고 추측
- ✓ 이러한 Issue에 대하여, **Target Word**를 생성할 때, **Annotation** 집합을 검색하도록 확장.
- ✓ 기존 **RNNencdec** 모델을 능가
- ✓ 문장의 길이와 상관없이 훨씬 더 긴 문장에 대하여 **Robust**하다.
- ✓ **Phrase-based SMT Model**의 성능에 버금가는 성능을 이루어냈다.
- ✓ **Unknown & Rare word**를 더 잘 다루는 것은 향후 연구 과제이다.

References

- Seq2seq Attention: <https://arxiv.org/pdf/1409.0473.pdf>
- Data Selection Method: <https://aclanthology.org/D11-1033.pdf>
- RNNencdec: <https://arxiv.org/pdf/1409.3215.pdf>
- Maxout hidden layer:
<https://m.blog.naver.com/PostView.naver?isHttpsRedirect=true&blogId=laonple&logNo=220836305907>