

Enhanced NLG with Monolingual Corpus

Ki Hyun Kim

nlp.with.deep.learning@gmail.com

In Typical NLG Tasks,

- Encoder-Decoder 구조의 모델들은 parallel corpus가 필요
 - Machine Translation: 한-영 corpus
 - Summarization: 본문-요약 corpus
 - Chatbot: 질문-대답 corpus
- 문제: Parallel corpus는 수집 비용이 매우 비쌈
 - (번역의 경우) 하지만 지천에 깔린 것이 monolingual corpus
- 한정된 parallel corpus로 학습할 경우,
 - encoder는 context embedding 성능이 떨어질 것
 - decoder는 제한된 성능의 언어모델이 될 수 밖에 없음

Language Model Ensemble [Gulcehre et al., 2015]

- Motivations:
 - Parallel corpus는 매우 한정되어 있으며, 이로 인해 decoder 언어모델의 성능이 저하됨
- Main Idea:
 - Monolingual corpus를 통해 학습한 별도의 언어모델을 decoder에 앙상블로 결합하자.

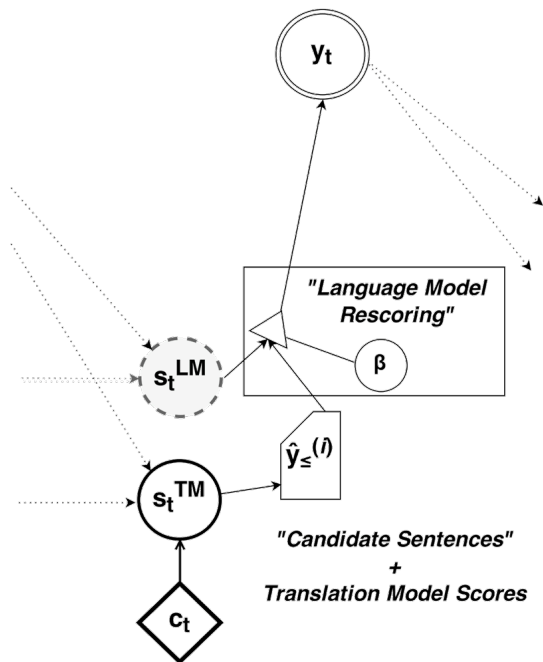
Language Model Ensemble

[Gulcehre et al., 2015]

Shallow Fusion

- Seq2seq와 LM을 interpolation

$$\log p(\mathbf{y}_t = k) = \log p_{\text{TM}}(\mathbf{y}_t = k) + \beta \log p_{\text{LM}}(\mathbf{y}_t = k),$$



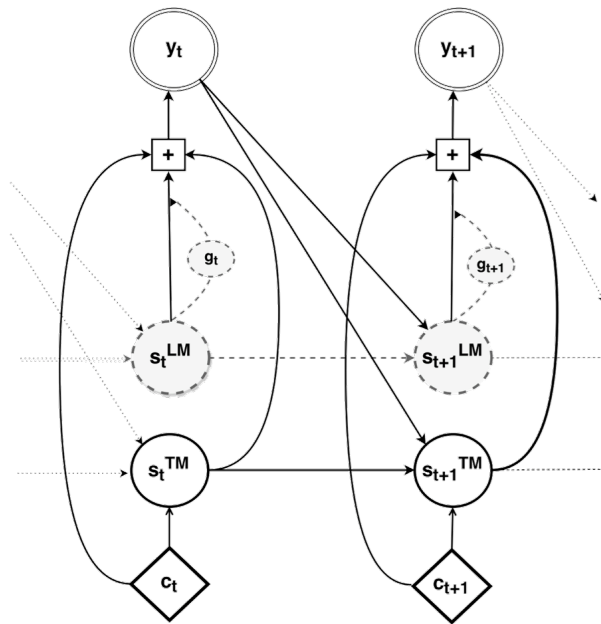
(a) Shallow Fusion (Sec. 4.1)

Deep Fusion

- 디코더와 LM의 hidden layer 결과를 concatenate.

- LM의 상태에 따라 gate를 열고 닫아 정보의 흐름을 컨트롤

$$g_t = \sigma \left(\mathbf{v}_g^\top \mathbf{s}_t^{\text{LM}} + b_g \right)$$



(b) Deep Fusion (Sec. 4.2)

Language Model Ensemble [Gulcehre et al. 2015]

- 단방향 코퍼스를 이용해 만든 언어모델 앙상블을 통해
부족한 코퍼스로 인한 디코더 언어모델을 보완하여 성능 향상

| | De-En | | Cs-En | |
|----------------|--------------|--------------|--------------|--------------|
| | Dev | Test | Dev | Test |
| NMT Baseline | 25.51 | 23.61 | 21.47 | 21.89 |
| Shallow Fusion | 25.53 | 23.69 | 21.95 | 22.18 |
| Deep Fusion | 25.88 | 24.00 | 22.49 | 22.36 |

Table 4: Results for De-En and Cs-En translation tasks on WMT'15 dataset.