

Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation

(Seq2Seq, GRU, 2014년 논문)

기계 번역을 위한 자연어 표현 학습

전은영

Abstract

- RNN Encoder-Decoder : 두 개의 RNN으로 구성된 모델 제시
 - 인코더 : 가변 길이의 seq -> 고정된 길이의 벡터
 - 디코더 : 고정된 길이의 벡터 -> (다른 언어) 가변 길이의 seq
 - GRU : 새로운 hidden unit 제시
 - 인코더와 디코더는 같이(jointly) 학습된다.
- RNN Encoder-Decoder로 계산한 번역 확률을 log-linear 모델에 추가 feature로 사용
=> SMT 성능 향상
- 의미적, 문법적 구조를 보존한 자연어 표현(representation) 학습

} seq2seq

Statistical Machine Translation (SMT)

- Decode 과정
- 목적

$$\begin{aligned} & \operatorname{argmax}_f P(f|e) \\ &= \operatorname{argmax}_f P(e|f)P(f) \quad (\because P(f|e) \propto P(e|f)P(f)) \end{aligned}$$

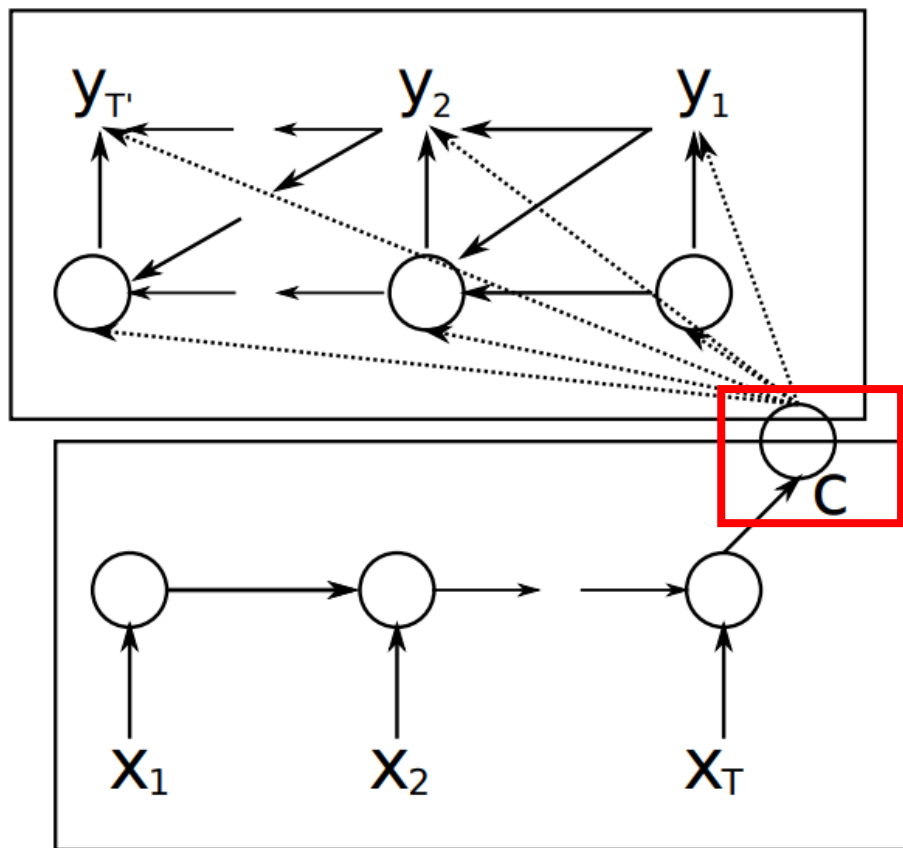
- $P(e|f)$: translation model
- $P(f)$: language model

Statistical Machine Translation (SMT)

- 실제로는 log-linear model로 모델링
 - $\log P(f|e) = \sum_{n=1}^N w_n f_n(f, e) + \log Z(e)$
 - $\{f_n\}_{n=1}^N$: feature 함수의 집합
 - w_n : n 번째 feature 함수의 가중치 (BLEU score 최대화하는 방향으로 학습)
 - $Z(e)$: 일반화 상수
 - 좋은 feature 함수를 찾는 것이 중요!
 - => RNN Encoder-Decoder의 번역 확률을 feature 함수로 사용!

RNN Encoder-Decoder

Decoder



Encoder

- $P(y_1, \dots, y_{T'} | x_1, \dots, x_T)$ 번역 확률 학습

- $X = (x_1, \dots, x_T)$

- $Y = (y_1, \dots, y_{T'})$

- 목적 함수

$$: \operatorname{argmax}_{\theta} \frac{1}{N} \sum_{n=1}^N \log P_{\theta}(Y_n | X_n)$$

representation

- Encoder (요약) : input $\rightarrow C$

- $h_t = f_w(h_{t-1}, x_t)$

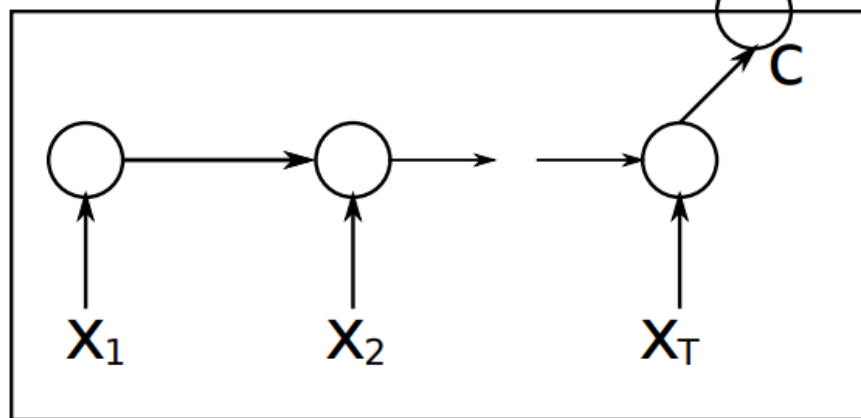
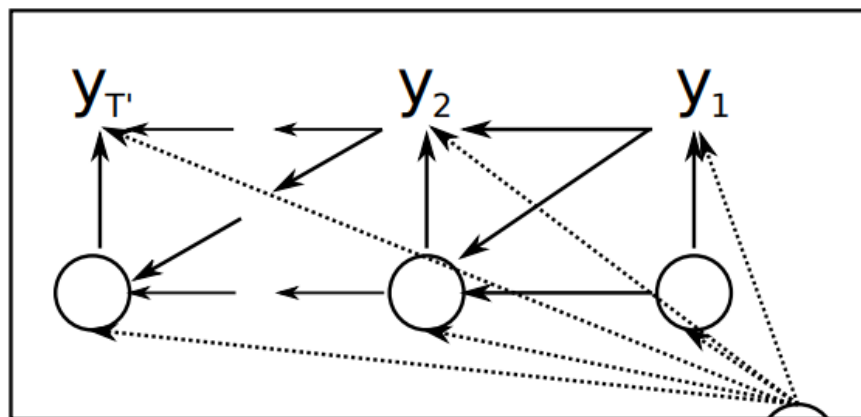
- Decoder (생성) : $C \rightarrow \text{output}$

- $h_t = f_w(h_{t-1}, y_{t-1}, C)$

* C는 고정된 길이

RNN Encoder-Decoder

Decoder



Encoder

- 학습 후 사용하는 두 가지 방법

1. Seq 생성

: 입력 seq가 주어질 때 타겟 seq 생성

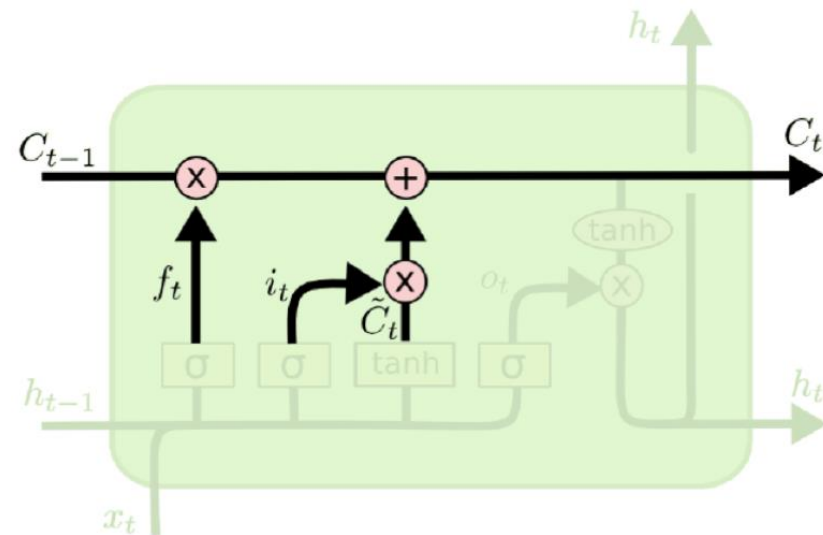
2. 점수 매기기 (Scoring)

: input과 output의 점수(번역 확률) 계산

LSTM

: cell state 도입 for vanishing gradient 방지

- Forget gate : f_t
- Input gate : i_t, \tilde{C}_t
- Cell state
: 보존할 정보 계산
- Output gate
: cell state 반영한 h_t 계산



Cell state

정보 얼마나 반영 결정

$$C_t = f_t \times \underbrace{C_{t-1}}_{\text{이전 cell state}} + i_t \times \underbrace{\tilde{C}_t}_{\text{현재 input 정보}}$$

이전 cell state 현재 input 정보

GRU(Gated Recurrent Unit)

- LSTM의 단순 변형
 - f_t, i_t 가 아닌 $z_t, (1 - z_t)$ 로
=> Update gate
 - cell state와 hidden state 합침
=> 별도의 memory cell 가지지 않음

GRU 수식

- Reset gate

$$r_j = \sigma([W_r X]_j + [U_r h_{t-1}]_j)$$

$$\tilde{h}_j^t = \varphi([W X]_j + [U(r \odot h_{t-1})]_j)$$

- Update gate

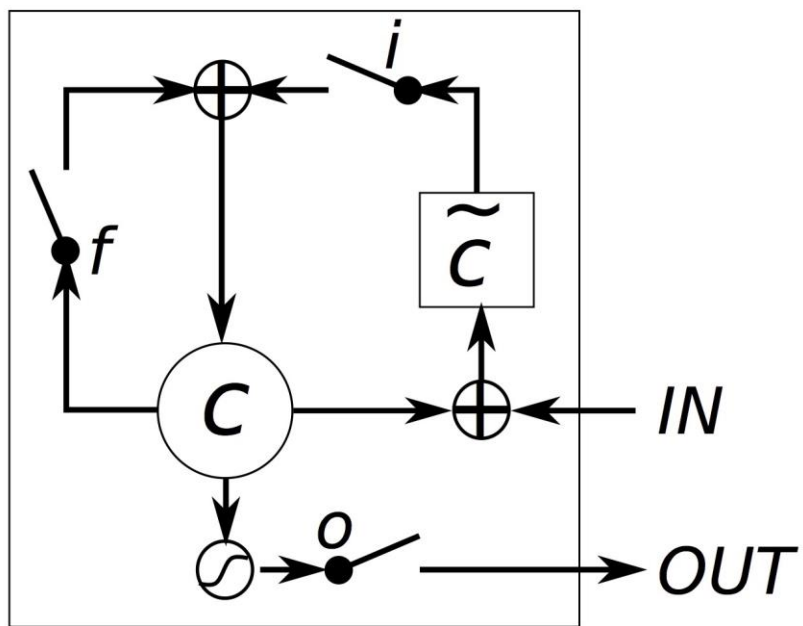
$$z_j = \sigma([W_z X]_j + [U_z h_{t-1}]_j)$$

- Hidden state 계산

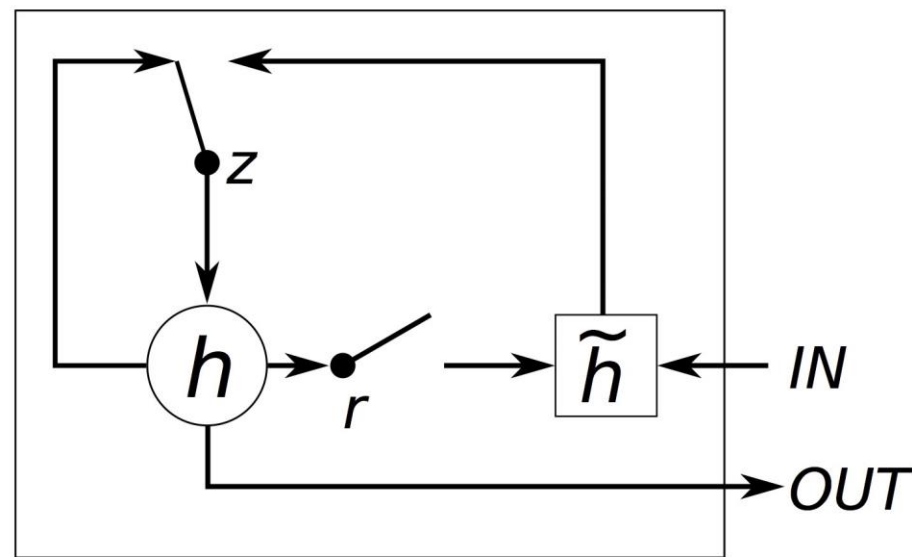
$$h_j^t = z_j h_j^{t-1} + (1 - z_j) \tilde{h}_j^t$$

GRU vs LSTM

- GRU가 계산량 적고, 구현하기 쉬움



(a) Long Short-Term Memory



(b) Gated Recurrent Unit

Scoring Phrase Pairs

- Phrase table

- ① Rescore : 존재하는 phrase table에 new score 추가

- ② Replace : 존재하는 phrase table 교체 => expensive

- 학습 시 phrase pair의 빈도수 무시

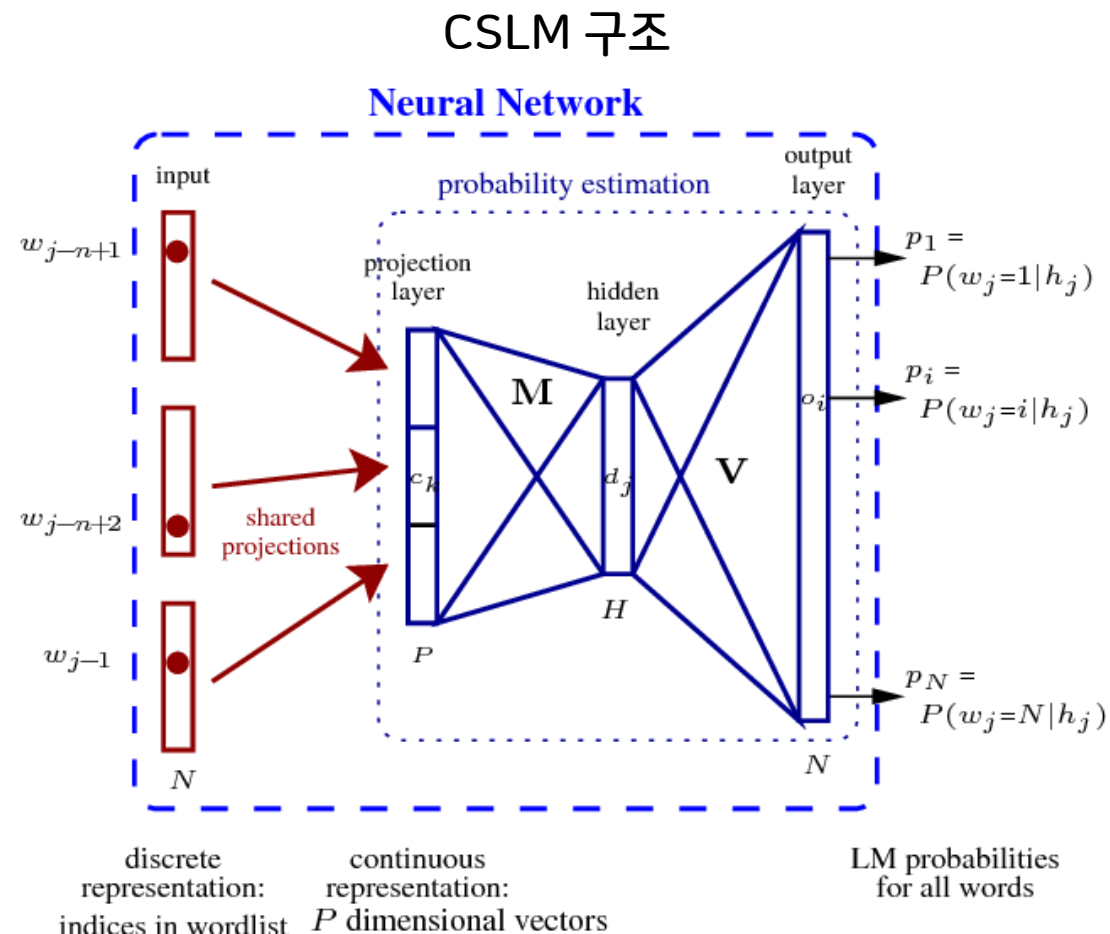
- 계산량 줄이기 위해
 - 단순히 빈도수만으로 학습하지 않다는 것을 보이기 위해
 - 존재하는 번역 확률이 이미 빈도수를 반영

Data and Baseline

- WMT'14 English to French 사용
- Data selection
- baseline phrase-based SMT : Moses 기본 세팅

Combination

- Baseline
: Moses 기본 세팅
- Baseline + RNN
: RNN Encoder-Decoder 추가 적용
- Baseline + CSLM + RNN
: CSLM 추가 적용 (NN의 전통적인 방법)
- Baseline + CSLM + RNN + Word Penalty
: 모르는 단어에 penalty 적용



Quantitative Analysis

| Models | BLEU | |
|-----------------|-------|-------|
| | dev | test |
| Baseline | 30.64 | 33.30 |
| RNN | 31.20 | 33.87 |
| CSLM + RNN | 31.48 | 34.64 |
| CSLM + RNN + WP | 31.50 | 34.54 |

- CSLM + RNN이 best !
=> 두 방법의 기여가 많이
연관되지는 않는다
- WP 적용 시 dev 성능 올랐지만,
test 성능은 떨어짐

Qualitative Analysis - 1

- 짧은 phrases를 더 선호

=> BLEU Score 짧은 phrase에서 높게 나옴

| Source | Translation Model | RNN Encoder–Decoder |
|--|---|---|
| , Minister of Communications and Transport | [Secrétaire aux communications et aux transports :] [Secrétaire aux communications et aux transports] | [Secrétaire aux communications et aux transports] [Secrétaire aux communications et aux transports :] |
| did not comply with the | [vestimentaire , ne correspondaient pas à des] [susmentionnée n' était pas conforme aux] [présentées n' étaient pas conformes à la] | [n' ont pas respecté les] [n' était pas conforme aux] [n' ont pas respecté la] |
| parts of the world . | [© gions du monde .] [régions du monde considérées .] [région du monde considérée .] | [parties du monde .] [les parties du monde .] [des parties du monde .] |
| the past few days . | [le petit texte .] [cours des tout derniers jours .] [les tout derniers jours .] | [ces derniers jours .] [les derniers jours .] [cours des derniers jours .] |
| on Friday and Saturday | [vendredi et samedi à la] [vendredi et samedi à] [se déroulera vendredi et samedi ,] | [le vendredi et le samedi] [le vendredi et samedi] [vendredi et samedi] |

(b) Long, rare source phrases

Qualitative Analysis - 2

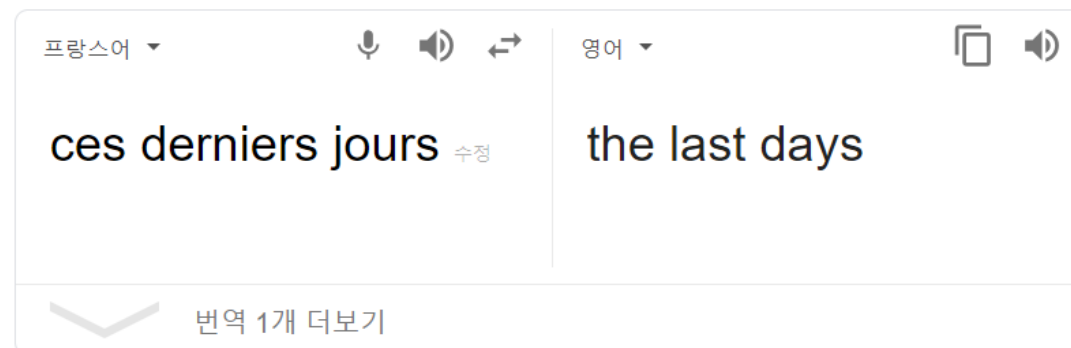
- 단순히 빈도수만으로 학습하는 것 아님

| | | |
|---------------------|---|---|
| the past few days . | [le petit texte .] [cours des tout derniers jours .] [les tout derniers jours .] | [ces derniers jours .] [les derniers jours .] [cours des derniers jours .] |
|---------------------|---|---|



Google 번역에서 열기

사용자 의견



Google 번역에서 열기

사용자 의견

Qualitative Analysis - 3

- 실제 phrase table 없이도 target phrase 잘 만든다.

| Source | Samples from RNN Encoder-Decoder |
|---|--|
| , Minister of Communica- tions and Transport | [, ministre des communications et le transport] (×13) |
| did not comply with the | [n' tait pas conforme aux] [n' a pas respect l'] (×2) [n' a pas respect la] (×3) |
| parts of the world . | [arts du monde .] (×11) [des arts du monde .] (×7) |
| the past few days . | [quelques jours .] (×5) [les derniers jours .] (×5) [ces derniers jours .] (×2) |
| on Friday and Saturday | [vendredi et samedi] (×5) [le vendredi et samedi] (×7) [le vendredi et le samedi] (×4) |

(b) Long, rare source phrases

프랑스어 ▾
⇒
영어 ▾

quelques jours ×

14 / 5000

변역하기

a few days
어 퓨 데이즈

12 번역 수정

🔊

📄

🌟

🔗

프랑스어 ▾
⇒
영어 ▾

les dernier jours ×

17 / 5000

변역하기

the last days
더 래스트 데이즈

12 번역 수정

🔊

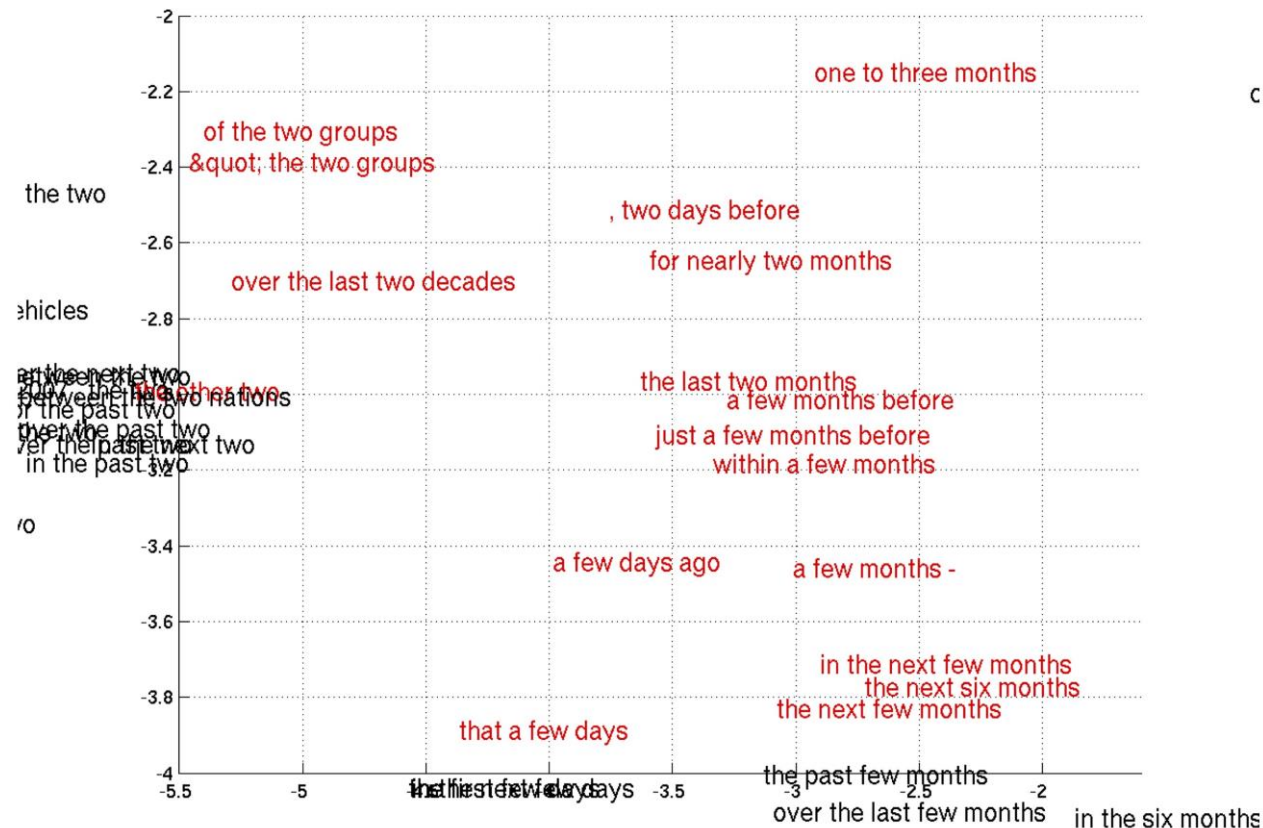
📄

🌟

🔗

Qualitative Analysis - 4

- 의미적, 문법적으로 잘 표현



Conclusion

- 가변 길이의 seq -> 가변 길이의 seq 학습
- Scoring a pair of seqs, Generating seq 가능
- 새로운 hidden unit 제시 (GRU)
- 언어의 규칙성 포착 => representation으로 활용 가능
- SMT에 사용된 기존 NN과 다소 orthogonal
- SMT 대체 가능성

감사합니다.