

2022학년도 1학기 컴퓨터언어학

제18강 기계번역과 부호화기-복호화기 모형 (1)

박수지

서울대학교 인문대학 언어학과

2022년 5월 9일 월요일

오늘의 목표

- 1 기계 번역의 역사와 세 가지 유형을 구별해서 설명할 수 있다.
- 2 기계 번역을 평가하는 척도로 Character F-score를 계산할 수 있다.
- 3 Encoder-decoder 모형에서 두 개의 신경망이 어떻게 작동하는지 설명할 수 있다.

기계 번역(machine translation)

출발어(source language)의 문장 x 를 도착어(target language)의 문장 y 로 번역하기

TMI (x, y) 대신 (e, f) 나 (f, e) 를 쓰는 문헌도 많다.

역사 및 유형

1950년대 규칙 기반(Rule-based)

1990년대 통계적(Statistical)

2014년- 신경망(Neural)

규칙 기반 기계 번역 (1950년대)

출발어 문장의 각 단어를 도착어 단어에 일일이 대응시킨다.

1950년대 농담

영어 문장을 러시아어로 번역한 후 다시 영어로 번역한 결과

en The spirit is willing, but the flesh is weak.
[마음은 원이로되 육신이 약하도다 (마태복음 26:41)]

→ **ru** ...

→ **en** The vodka is good, but the meat is rotten.

<https://nytimes.com/1983/04/28/business/technology-the-computer-as-translator.html>

통계적 기계 번역 (1990년대-)

데이터에서 확률 모형을 학습한다.

- 데이터: 병렬 코퍼스(parallel corpus)
- 목표: $\arg \max_y P(y|x) = \arg \max_y P(x|y)P(y)$

x 출발어 문장

y 도착어 문장

$P(x|y)$ 번역 모형: x 의 단어와 구가 얼마나 잘 번역되었는가?

$P(y)$ 언어 모형: 도착어에서 y 가 얼마나 자연스러운 문장인가?

문제

병렬 코퍼스에서 $P(x|y)$ 를 어떻게 계산하는가?

신경망 기계 번역 (2014년-현재)

데이터에서 신경망 모형을 학습한다.

- 데이터: 병렬 코퍼스
- 신경망: seq2seq (= Encoder-Decoder)


seq... 신경망(RNN 등)을 사용하여 출발어 문장 x 를 무엇인가로 인코딩한다.





...2seq 신경망(RNN 등)을 사용하여 무엇인가로부터 도착어 문장 y 를 디코딩한다.


⇒ 두 개의 신경망으로 이루어져 있다.

신경망 기계 번역


파파고 vs. 구글

 웹사이트 번역 GYM | 사전




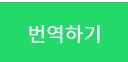
   로그인 


한국어 감지 

↔

잠없는덩. 

5 / 5000





   

영어 

It's not fun.

잇스 낫 펀.

번역 수정 | 번역 평가

신경망 기계 번역

파파고 vs. 구글

The screenshot shows the Google Translate website. At the top, there's a navigation bar with the Google Translate logo, a 'Sign in' button, and a grid icon. Below this, there are tabs for 'Text' and 'Documents'. The main interface has a language selection bar with 'KOREAN' selected on the left and 'ENGLISH' selected on the right. The input text on the left is '잼없는덴.' (jaem-eobsneundeng.), and the output text on the right is 'No jam.'. There are icons for voice input, volume, and a character count (5 / 5,000) on the left. On the right, there are icons for a star, copy, share, and a 'Send feedback' link at the bottom right.

주요 데이터

병렬 코퍼스

- WMT <https://www.statmt.org/>
- 한국어-영어 번역(병렬) 말뭉치 <https://aihub.or.kr/aidata/87>

한국어	영어(초벌)	영어 검수
우리 모두 한국 전통 놀이에 대해서 많은 관심을 가집시다.	Let's get interested in Korean traditional games.	Let's all have more interest in Korean traditional games.
제가 이야기한 것이 인정이 되어 기간을 조율하는 줄 알았습니다.	I thought what I said was accepted, and we are adjusting the period.	I thought what I said was accepted, and we are adjusting the period.
나는 네가 한국어를 정말 열심히 배운다고 생각해.	I think you really study Korean hard.	I think you study Korean really hard.

기계번역 평가 척도: ChrF

ChrF: Character F-score

Character Precision percentage of character 1-grams, 2-grams, ..., k-grams in the hypothesis that occur in the reference, averaged.

Character Recall percentage of character 1-grams, 2-grams,..., k-grams in the reference that occur in the hypothesis, averaged.

Character F $_{\beta}$ $(1 + \beta^2) \frac{\text{ChrP} \cdot \text{ChrR}}{\beta^2 \text{ChrP} + \text{ChrR}}$...ChrP와 ChrR의 조화평균

ChrP 예측한(=기계가 번역한) 문장의 문자 n그램 중에서 실제로 맞춘 비율이 얼마인가?

ChrR 정답인(=사람이 번역한) 문장의 문자 n그램 중에서 맞게 예측한 비율이 얼마인가?

β 재현도(Recall)를 정밀도(Precision)의 몇 배로 가중할 것인가?

기계번역 평가 척도: ChrF

ChrF2: Character F2-score

$$\frac{5 \cdot \text{ChrP} \cdot \text{ChrR}}{4\text{ChrP} + \text{ChrR}}$$

예시 ($k = 2, \beta = 2$)

REF 은수에게 돈을 빌렸다.

HYP1 나는 은수에게 돈을 빌렸다.

HYP2 은수에게 빌렸다.

Unigrams($n = 1$) in REF	은 수 에 게 돈 을 빌 렸 다 .	(10)
----------------------------	---------------------	------

Bigrams($n = 2$) in REF	은수 수에 에게 게돈 돈을 을빌 빌렸 렸다 다.	(9)
---------------------------	----------------------------	-----

Unigrams($n = 1$) in HYP1	나 는 은 수 에 게 돈 을 빌 렸 다 .	(12)
-----------------------------	-------------------------	------

Bigrams($n = 2$) in HYP1	나는 는은 은수 수에 에게 게돈 돈을 을빌 빌렸 렸다 다.	(11)
----------------------------	----------------------------------	------

기계번역 평가 척도: ChrF

예시 ($k = 2, \beta = 2$)

REF 은수에게 돈을 빌렸다.

HYP1 나는 은수에게 돈을 빌렸다.

	1-grams in HYP1	not in HYP1
1-grams in REF 은 수 에 게 돈 을 빌 렸 다 . (10)		— (0)
not in REF 나는 (2)		— (0)

$$\text{1-gram Precision} = \frac{10}{10 + 2} = \frac{5}{6}$$

$$\text{1-gram Recall} = \frac{10}{10 + 0} = 1$$

기계번역 평가 척도: ChrF

예시 ($k = 2, \beta = 2$)

REF 은수에게 돈을 빌렸다.

HYP1 나는 은수에게 돈을 빌렸다.

	2-grams in HYP1	not in HYP1
2-grams in REF	은수 수에 에게 게돈 돈을 을빌 빌렸 렸다 다. (9)	— (0)
not in REF	나는 는은 (2)	— (0)

$$\text{2-gram Precision} = \frac{9}{9 + 2} = \frac{9}{11}$$

$$\text{2-gram Recall} = \frac{9}{9 + 0} = 1$$

기계번역 평가 척도: ChrF

예시 ($k = 2, \beta = 2$)

REF 은수에게 돈을 빌렸다.

HYP1 나는 은수에게 돈을 빌렸다.

$$\text{ChrP} = \frac{1}{k} \sum_{n=1}^k (\text{k-gram Precision}) = \frac{1}{2} \left(\frac{5}{6} + \frac{9}{11} \right) = \frac{109}{132}$$

$$\text{ChrR} = \frac{1}{k} \sum_{n=1}^k (\text{k-gram Recall}) = \frac{1}{2} (1 + 1) = 1$$

$$\text{ChrF2} = \frac{5\text{ChrP} \cdot \text{ChrR}}{4\text{ChrP} + \text{ChrR}} = 5 \cdot \frac{109}{132} \cdot 1 / \left[4 \cdot \frac{109}{132} + 1 \right] \approx 0.9595$$

	입력	출력
순방향 신경망 (FFNN)	1개 벡터	1개 벡터
합성곱 신경망 (CNN)	n개 벡터 (이미지)	1개 벡터
순환 신경망 (RNN)	n개 벡터 (시퀀스)	1개 또는 n개의 벡터

부호화기-복호화기 모형 (Encoder-decoder model, sequence-to-sequence)

- 입력 시퀀스를 받아서 **임의의 길이**의 시퀀스를 생성할 수 있다.
- 활용: 기계번역, 요약, 질의응답, 대화 등 (cf. 품사 태깅)

Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation (Cho et al., EMNLP 2014) <https://aclanthology.org/D14-1179/>

부호화기-복호화기 모형의 구성요소

- 1 부호화기(encoder) 신경망: $x_1^n \mapsto h_1^n$
 x_1^n 입력 시퀀스...
 h_1^n ...에 대응하는 문맥화된 표상(contextualized representation)들의 시퀀스
- 2 문맥 벡터(context vector): $h_1^n \mapsto c$
 c 입력 시퀀스의 정보를 담고 있는 벡터
- 3 복호화기(decoder) 신경망: $c \mapsto h_1^m \mapsto y_1^m$ ($n \neq m$)
 h_1^m 임의의 길이의 은닉 상태들의 시퀀스...
 y_1^m ...에 대응하는 출력 시퀀스

주의

- 부호화기·복호화기로는 LSTM, CNN, 트랜스포머 등이 모두 가능하다.

Encoder-decoder

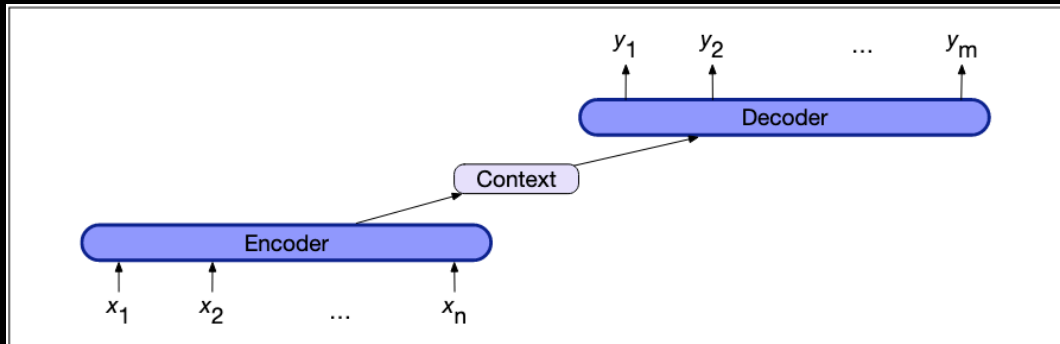


Figure 10.3 The encoder-decoder architecture. The context is a function of the hidden representations of the input, and may be used by the decoder in a variety of ways.

Encoder-decoder with RNNs

부호화기와 복호화기가 모두 순환 신경망인 경우

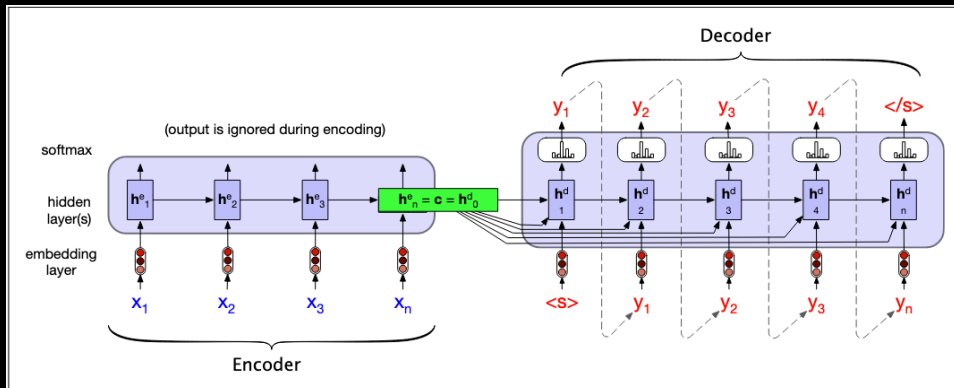


Figure 10.5 A more formal version of translating a sentence at inference time in the basic RNN-based encoder-decoder architecture. The final hidden state of the encoder RNN, h^e_n , serves as the context for the decoder in its role as h^d_0 in the decoder RNN.

Encoder-decoder with RNNs

부호화기와 복호화기가 모두 순환 신경망인 경우

복호화기 d의 구성

$$\vec{c} = \vec{h}_n^e$$

(부호화기 e의 n번째=마지막 은닉 상태)

$$\vec{h}_0^d = \vec{c}$$

$$\vec{h}_t^d = g(\hat{y}_{t-1}, \vec{h}_{t-1}^d, \vec{c})$$

$$\hat{y}_t = \arg \max_w P(w|x, y_1, \dots, y_{t-1})$$

$$\vec{z}_t = f(\vec{h}_t^d)$$

$$y_t = \text{softmax}(\vec{z}_t)$$

Encoder-decoder with RNNs

부호화기와 복호화기가 모두 순환 신경망인 경우

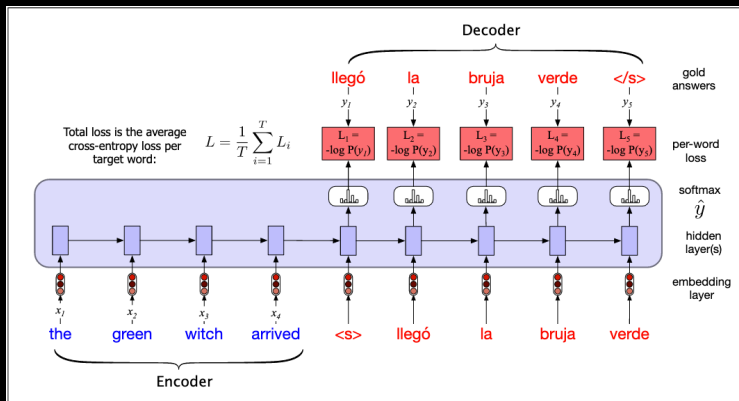


Figure 10.7 Training the basic RNN encoder-decoder approach to machine translation. Note that in the decoder we usually don't propagate the model's softmax outputs \hat{y}_t , but use **teacher forcing** to force each input to the correct gold value for training. We compute the softmax output distribution over \hat{y} in the decoder in order to compute the loss at each token, which can then be averaged to compute a loss for the sentence.