Neural Machine Translation in 10 Slides

EMA Summer School

Laurent Besacier

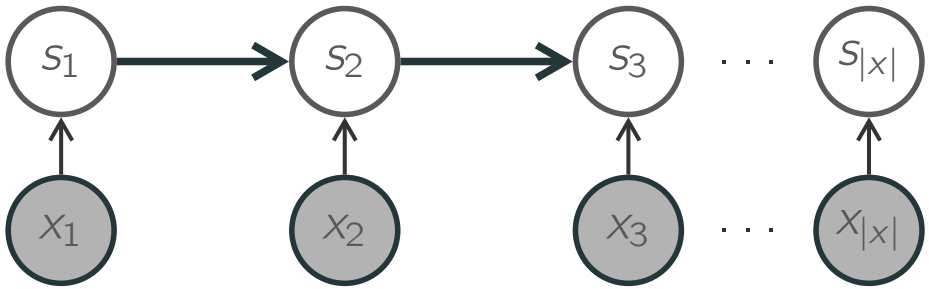
July 11th, 2019

Outline

1. Neural Machine Translation Models

Neural Machine Translation Models

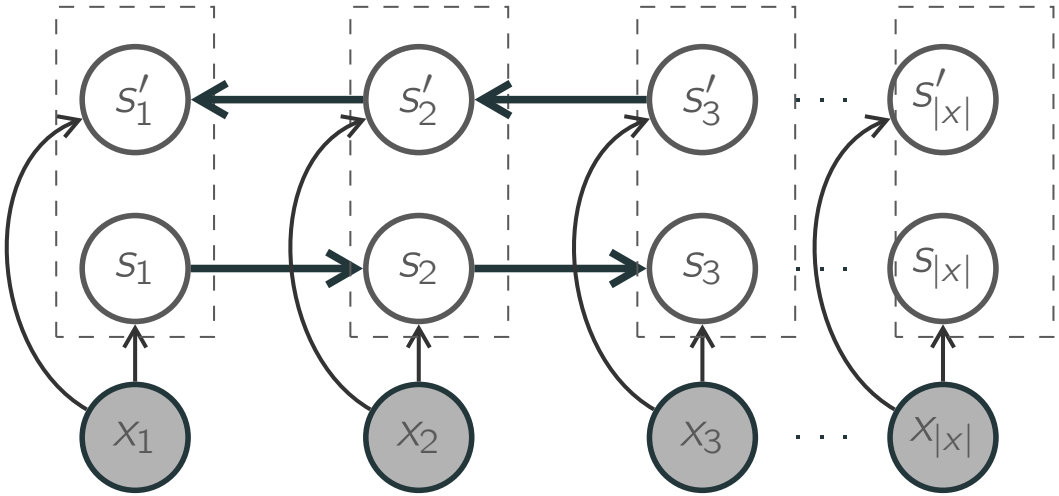
Sequence-to-sequence models | Recurrent networks



Encoder

[Graves, 2013, Sutskever et al., 2014, Cho et al., 2014, Bahdanau et al., 2015]

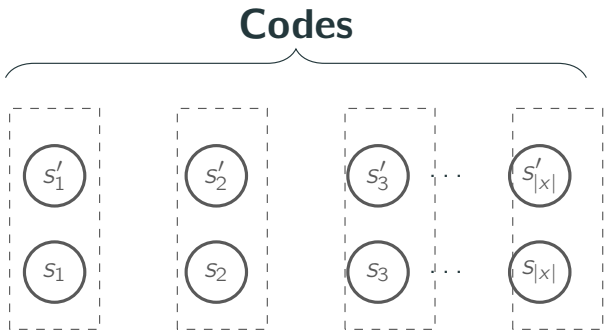
Sequence-to-sequence models | Recurrent networks



Bidirectional encoder

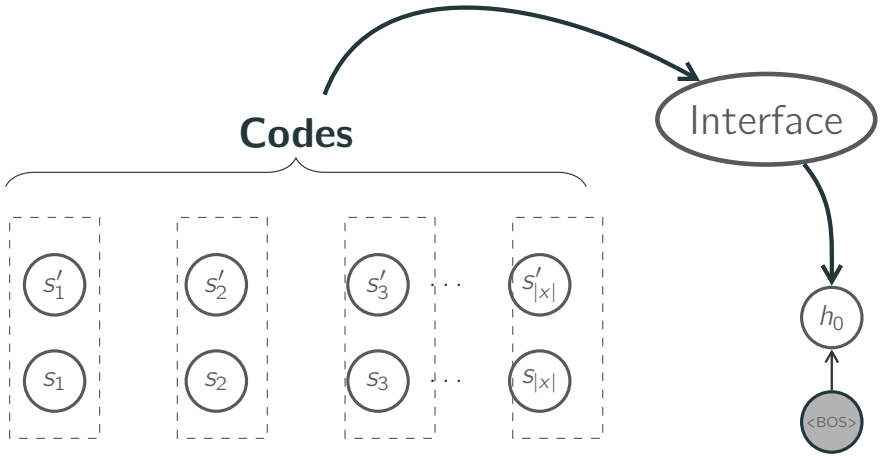
[Graves, 2013, Sutskever et al., 2014, Cho et al., 2014, Bahdanau et al., 2015]

Sequence-to-sequence models | Recurrent networks



[Graves, 2013, Sutskever et al., 2014, Cho et al., 2014, Bahdanau et al., 2015]

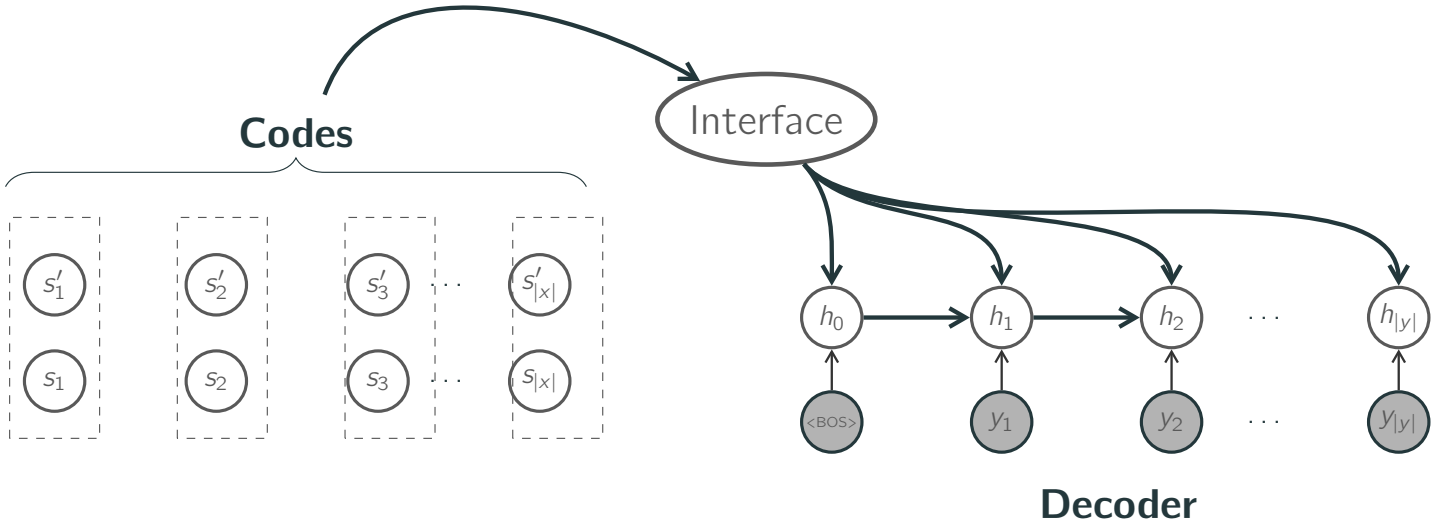
Sequence-to-sequence models | Recurrent networks



Decoder

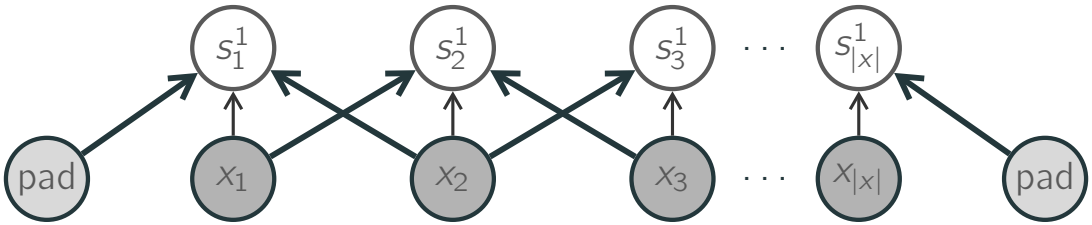
[Graves, 2013, Sutskever et al., 2014, Cho et al., 2014, Bahdanau et al., 2015]

Sequence-to-sequence models | Recurrent networks



[Graves, 2013, Sutskever et al., 2014, Cho et al., 2014, Bahdanau et al., 2015]

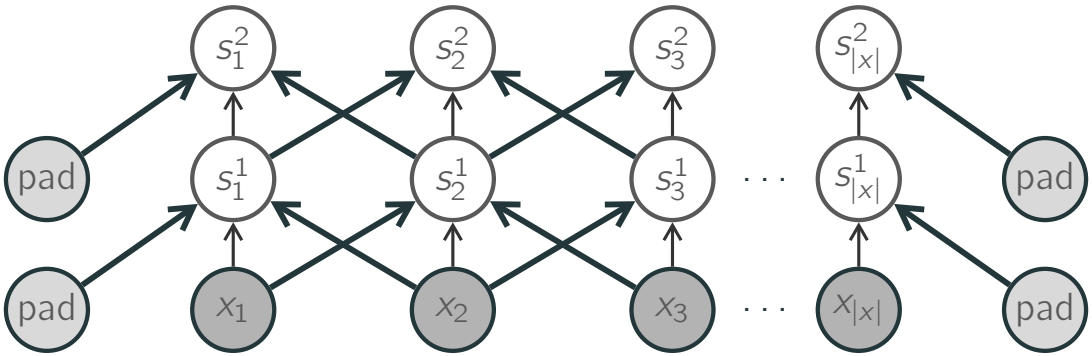
Sequence-to-sequence models | Convolutional networks



Encoder

[Kalchbrenner et al., 2014, Kim, 2014, Gehring et al., 2017]

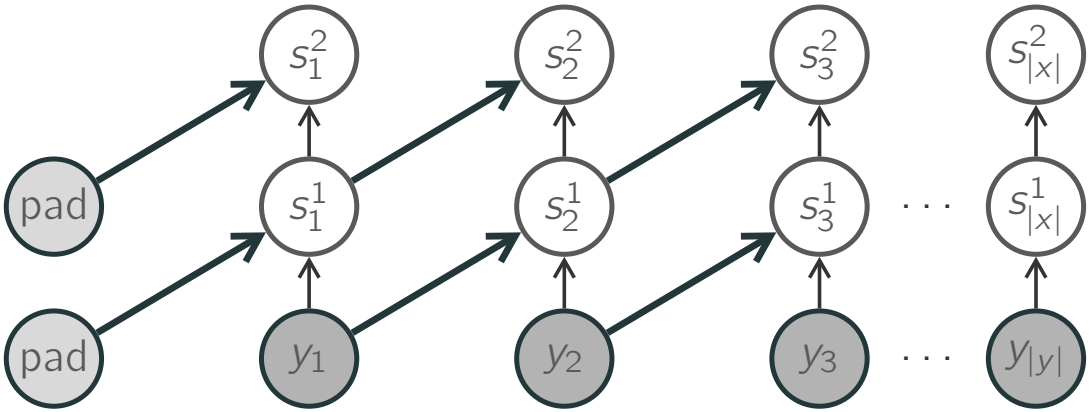
Sequence-to-sequence models | Convolutional networks



Encoder

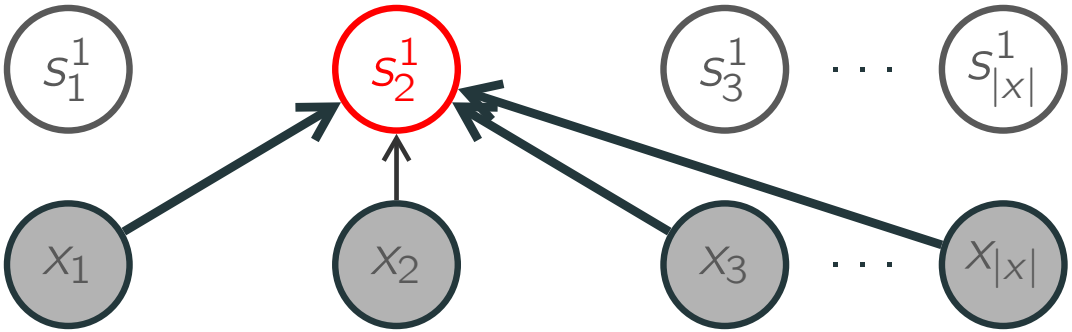
[Kalchbrenner et al., 2014, Kim, 2014, Gehring et al., 2017]

Sequence-to-sequence models | Convolutional networks



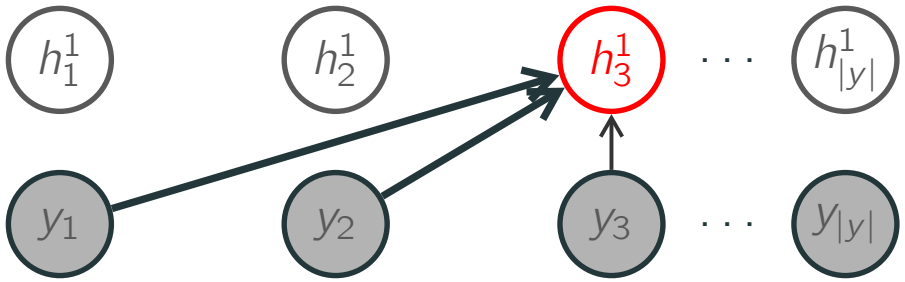
Decoder (Causal convolutions)

[Kalchbrenner et al., 2014, Kim, 2014, Gehring et al., 2017]



Encoder (self-attention)

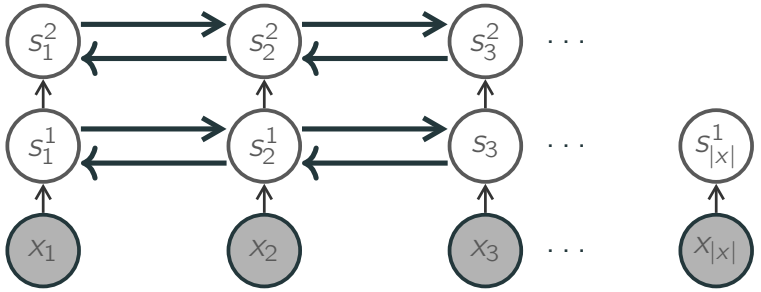
[Vaswani et al., 2017]



Decoder (masked self-attention)

[Vaswani et al., 2017]

Sequence-to-sequence models



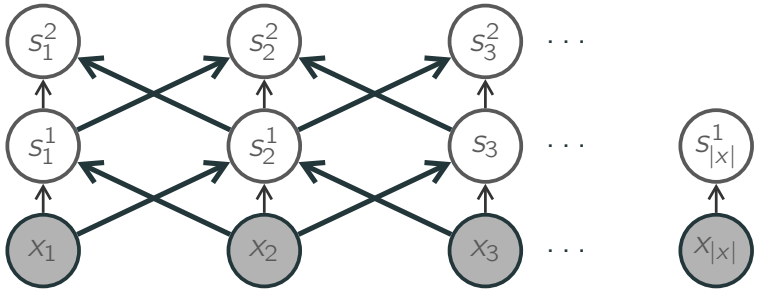
Recurrent

1. Unbounded dependencies.

2. $\mathcal{O}(T)$ sequential steps.

3. Full context.

Sequence-to-sequence models



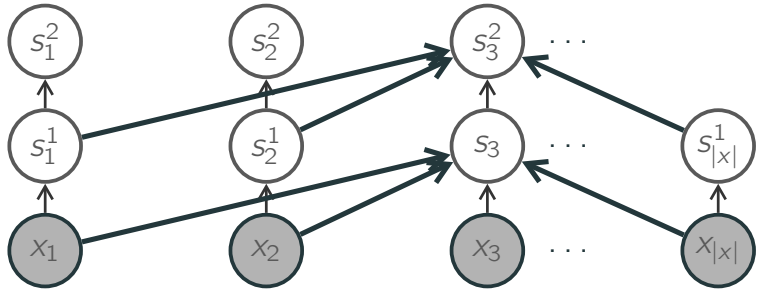
Recurrent

1. Unbounded dependencies.
2. $\mathcal{O}(T)$ sequential steps.
3. Full context.

Convolutional

1. Bounded dependencies.
2. $\mathcal{O}(1)$ sequential steps.
3. Incrementally built context.

Sequence-to-sequence models



Recurrent

1. **Unbounded dependencies.**
2. $\mathcal{O}(T)$ sequential steps.
3. **Full context.**

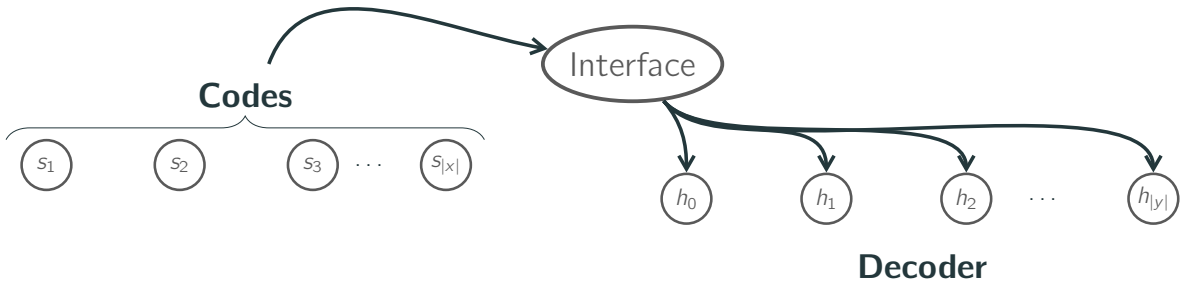
Convolutional

1. Bounded dependencies.
2. $\mathcal{O}(1)$ **sequential steps.**
3. Incrementally built context.

Transformer

1. **Unbounded dependencies.**
2. $\mathcal{O}(1)$ **sequential steps.**
3. **Full context.**

Encoder-decoder interfacing



Single vector:

- context $= s_{|x|}$.
- context $= \frac{1}{|x|} \sum s_i$
- context $= \max s_i$

Only at h_0 or at every time step.

Attention mechanisms:

$$e_i = \text{score}(s_i, h_{t-1}), \forall i$$

$$\alpha = softmax(e_i)_i$$

$$\text{context}_t = \sum_i \alpha_i s_i$$

Illustration: NMT Lab

- You will train your first NMT system using OpenNMT toolkit
- Based on RNNs with attention (but OpenNMT also implements Transformer models)
- Language pair will be French-English (BTEC corpus)
- Training might take a while (be patient)
- Let's start !

References i

Bahdanau, D., Cho, K., and Bengio, Y. (2015).
Neural machine translation by jointly learning to align and translate.
In ICLR.

Cho, K., van Merrienboer, B., Gülçehre, Ç., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y. (2014).
Learning phrase representations using RNN encoder-decoder for statistical machine translation.
In EMNLP.

References ii

Elbayad, M., Besacier, L., and Verbeek, J. (2018).
Pervasive attention: 2d convolutional neural networks for sequence-to-sequence prediction.
In Proceedings of the 22nd Conference on Computational Natural Language Learning, CoNLL 2018, Brussels, Belgium, October 31 - November 1, 2018, pages 97–107.

Gehring, J., Auli, M., Grangier, D., Yarats, D., and Dauphin, Y. (2017).
Convolutional sequence to sequence learning.
In ICML.

References iii


Godard, P., Zanon Boito, M., Ondel, L., Berard, A., Yvon, F., Villavicencio, A., and Besacier, L. (2018).
Unsupervised word segmentation from speech with attention.
In Interspeech.


Graves, A. (2013).
Generating sequences with recurrent neural networks.
CoRR, abs/1308.0850.


Kalchbrenner, N., Grefenstette, E., and Blunsom, P. (2014).
A convolutional neural network for modelling sentences.
In ACL.

References iv


Kim, Y. (2014).
Convolutional neural networks for sentence classification.
In ACL.


Luong, T., Pham, H., and Manning, C. (2015).
Effective approaches to attention-based neural machine translation.
In EMNLP.


Sutskever, I., Vinyals, O., and Le, Q. (2014).
Sequence to sequence learning with neural networks.
In NIPS.

References v

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, L., and Polosukhin, I. (2017).
Attention is all you need.
In NIPS.

Zanon Boito, M., Berard, A., Villavicencio, A., and Besacier, L. (2017).
Unwritten languages demand attention too! word discovery with encoder-decoder models.
In Automatic Speech Recognition and Understanding (ASRU), 2017 IEEE Workshop on. IEEE.