

Atencja

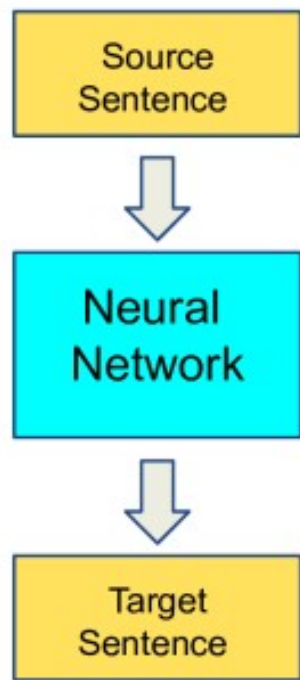
Tomasz Stanisławek (Applica, PhD candidate)

19.11.2018, MI² DataLab

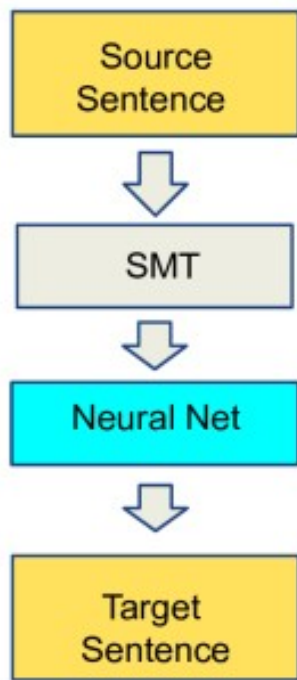
Atencja

1. Omówienie publikacji “Neural machine translation by jointly learning to align and translate, 2014”
 - a) Neural Machine Translation (NMT)
 - b) Atencja w NMT
2. Atencja w innych problemach
3. Podsumowanie

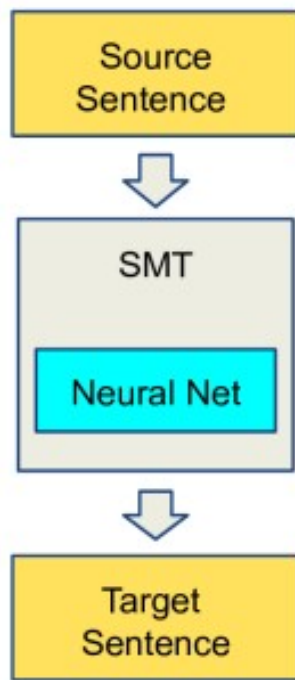
Neural Machine Translation Wprowadzenie



Neural MT



(Schwenk et al. 2006)



(Devlin et al. 2014)

$$\mathbf{x} = (x_1, \dots, x_{T_x})$$

$$\mathbf{y} = (y_1, \dots, y_{T_y})$$

Neural Machine Translation RNN

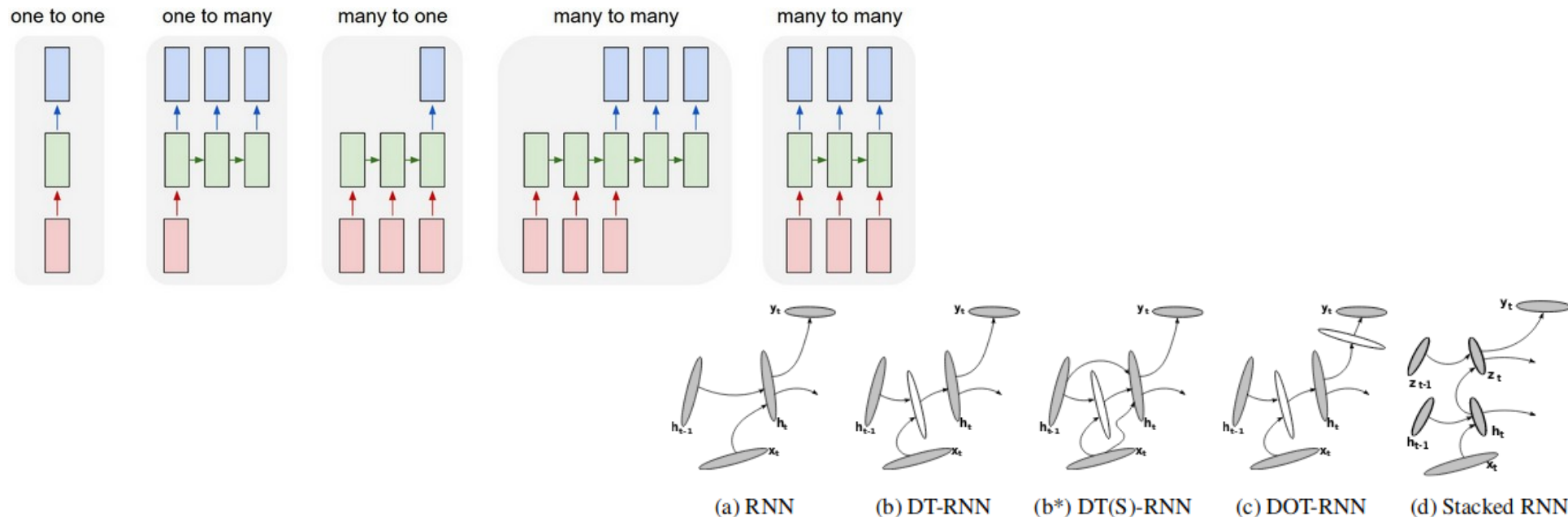
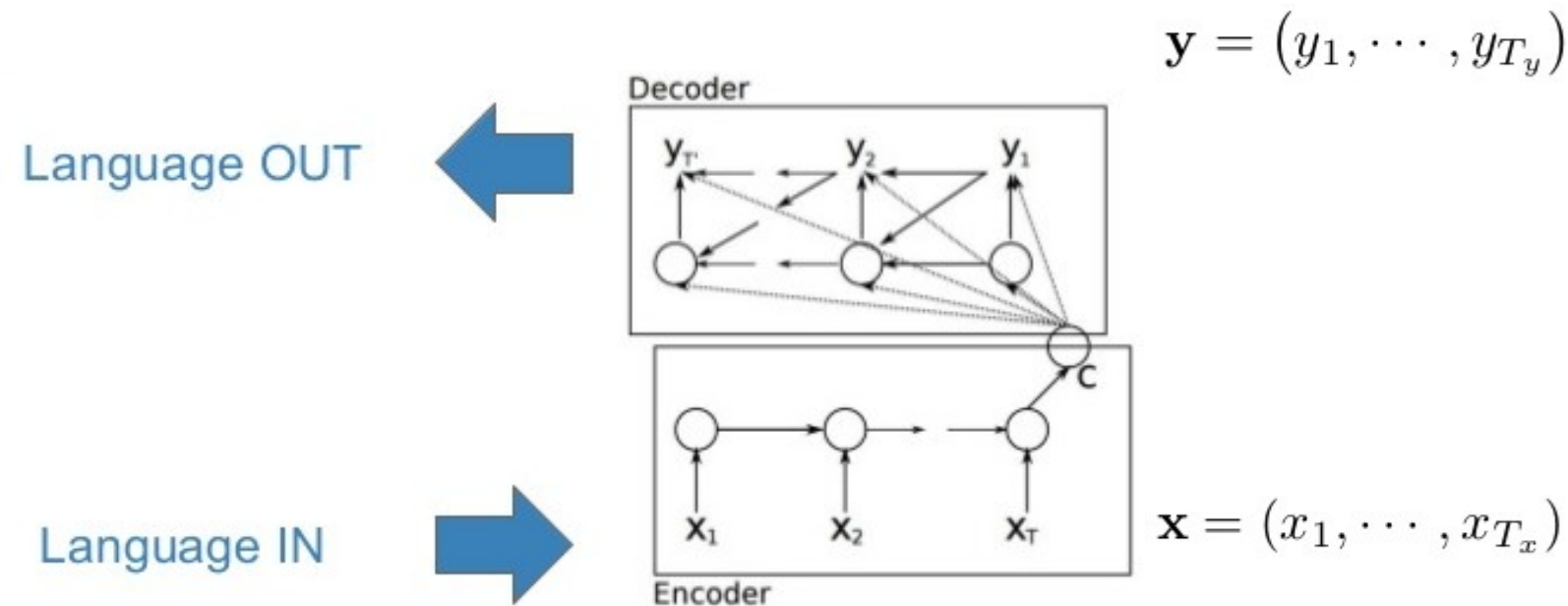


Figure 2: Illustrations of four different recurrent neural networks (RNN). (a) A conventional RNN. (b) Deep Transition (DT) RNN. (b*) DT-RNN with shortcut connections (c) Deep Transition, Deep Output (DOT) RNN. (d) Stacked RNN

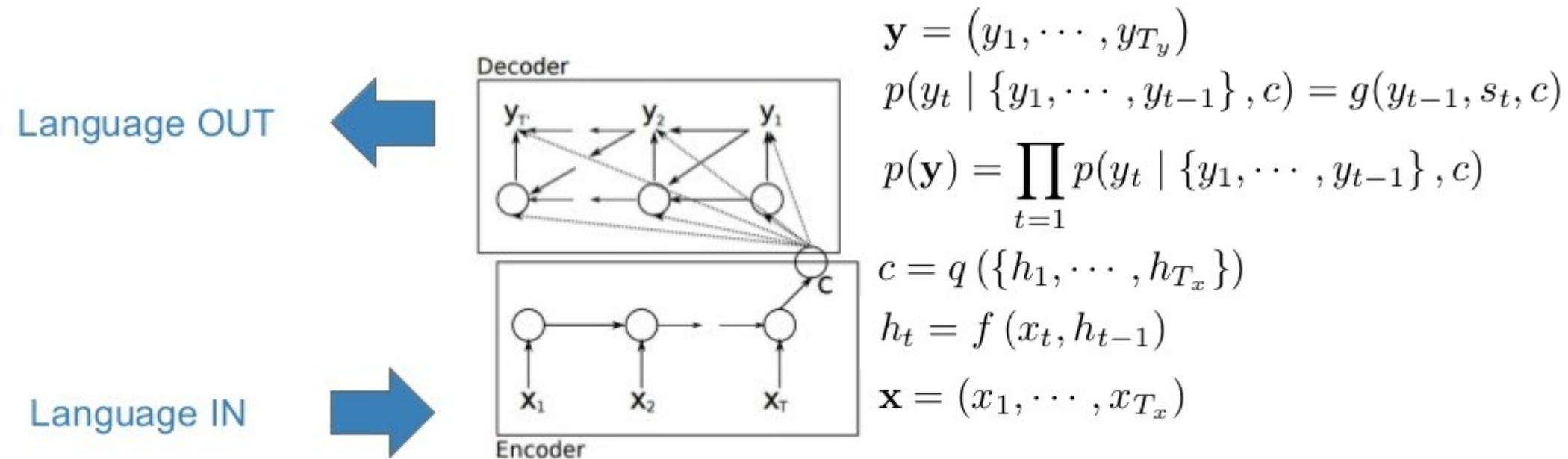
Neural Machine Translation

RNN Encoder - Decoder



Neural Machine Translation

RNN Encoder - Decoder



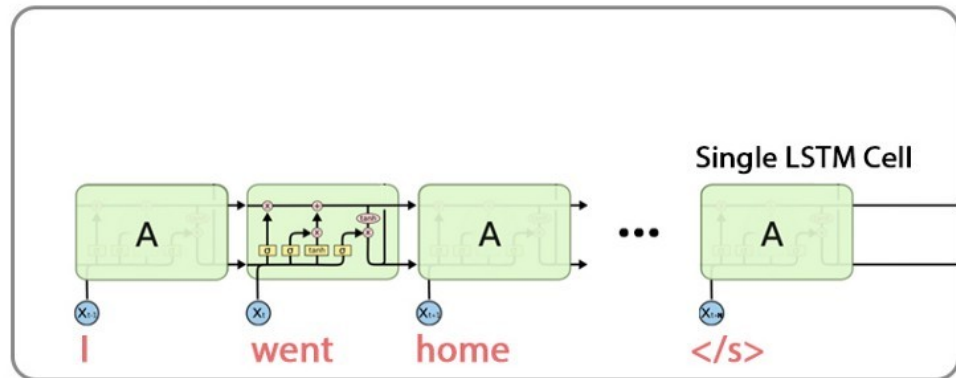
Cho, K. and Others. Learning phrase representations using RNN encoder-decoder for statistical machine translation. (EMNLP 2014)

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

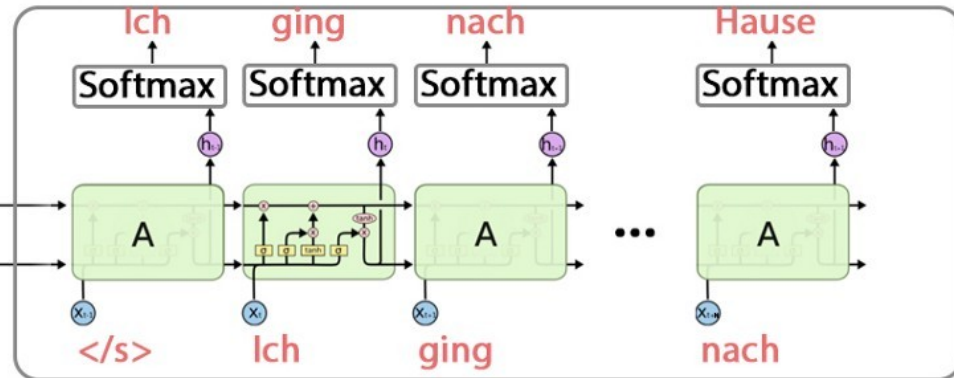
Neural Machine Translation

RNN Encoder - Decoder

Encoder

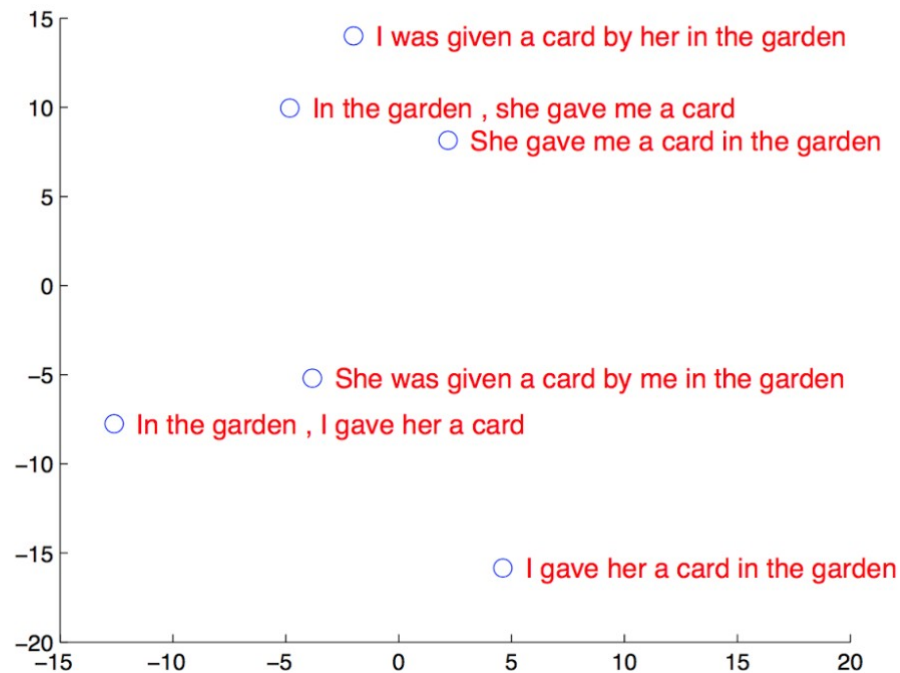
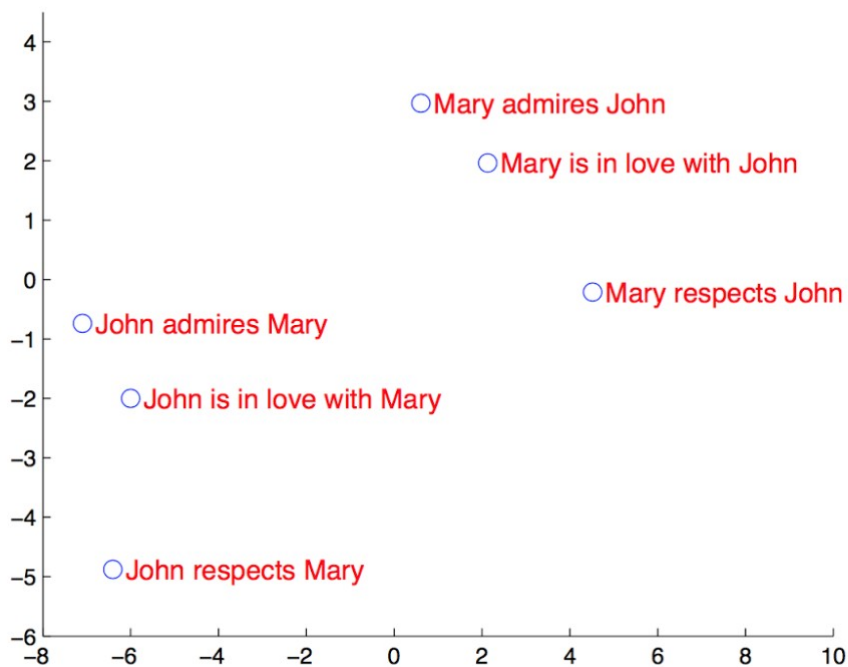


Decoder



Neural Machine Translation

RNN Encoder - Decoder

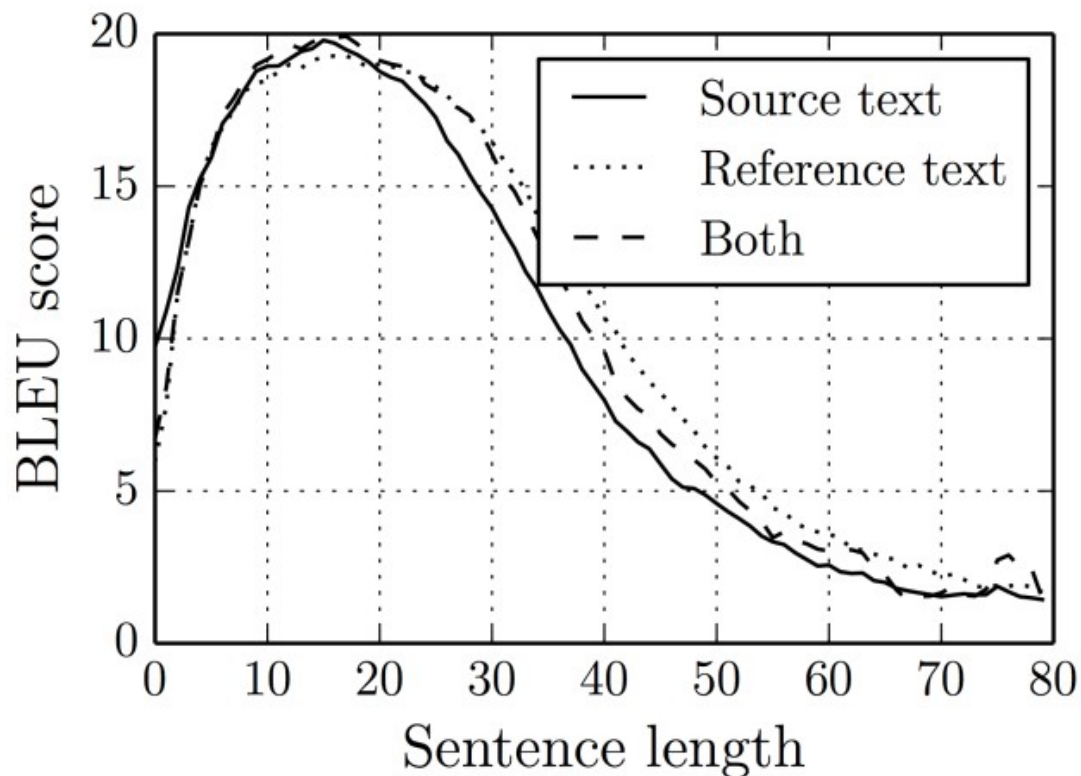


Cho, K. and Others. Learning phrase representations using RNN encoder-decoder for statistical machine translation. (EMNLP 2014)

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

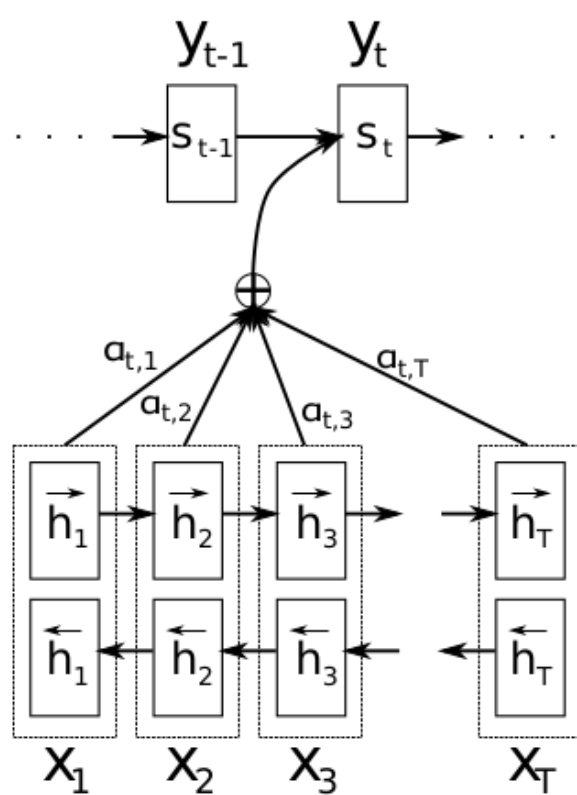
Neural Machine Translation

RNN Problems



Neural Machine Translation

RNN + Atencja

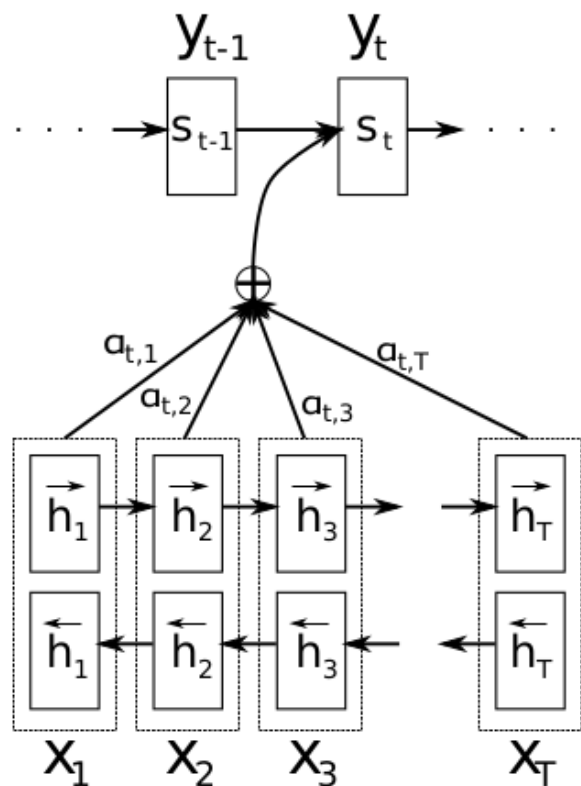


$$\mathbf{y} = (y_1, \dots, y_{T_y})$$

$$\mathbf{x} = (x_1, \dots, x_{T_x})$$

Neural Machine Translation

RNN + Atencja



$$\mathbf{y} = (y_1, \dots, y_{T_y})$$

$$p(y_i | y_1, \dots, y_{i-1}, \mathbf{x}) = g(y_{i-1}, s_i, c_i)$$

$$s_i = f(s_{i-1}, y_{i-1}, c_i)$$

$$c_i = \sum_{j=1}^{T_x} \alpha_{ij} h_j$$

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{T_x} \exp(e_{ik})}$$

$$e_{ij} = a(s_{i-1}, h_j)$$

$$h_j = \left[\vec{h}_j^\top; \overleftarrow{h}_j^\top \right]^\top$$

$$\mathbf{x} = (x_1, \dots, x_{T_x})$$

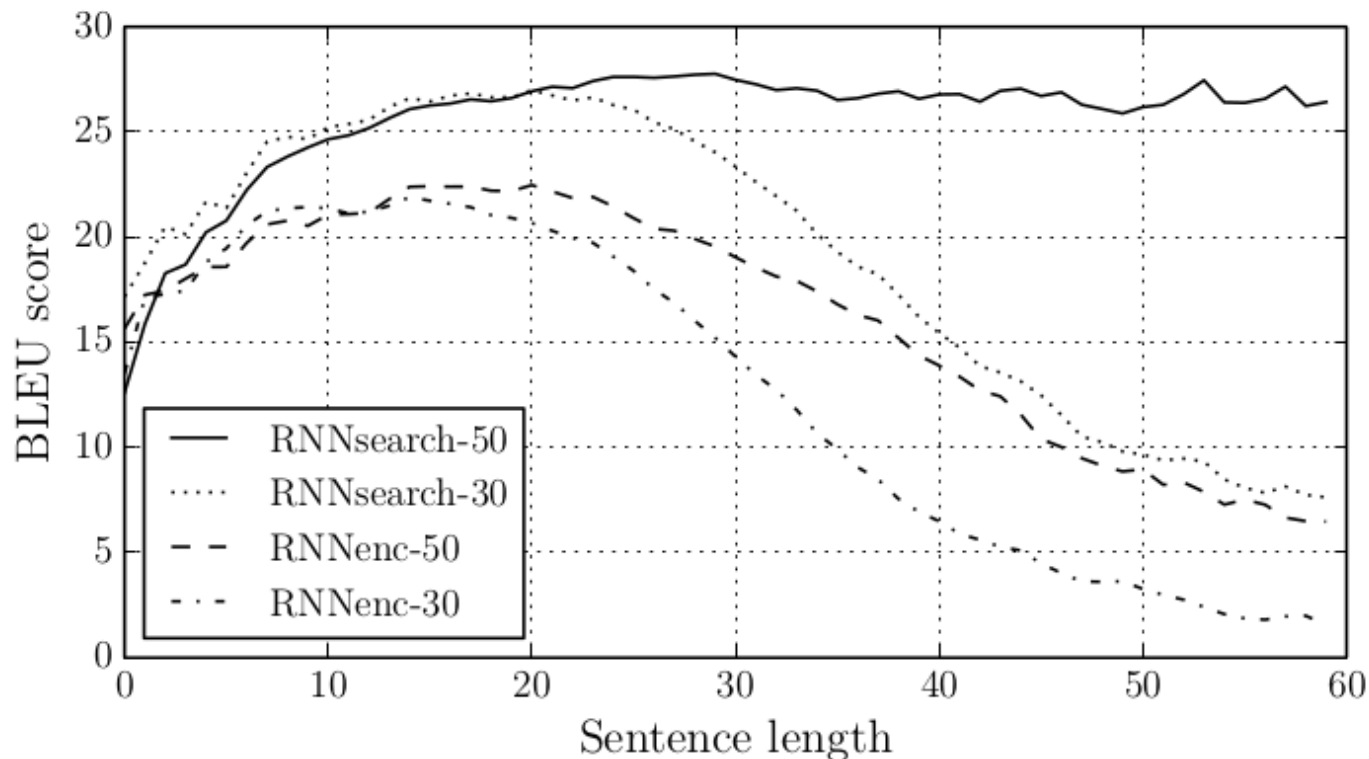
Neural Machine Translation

RNN + Atencja

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

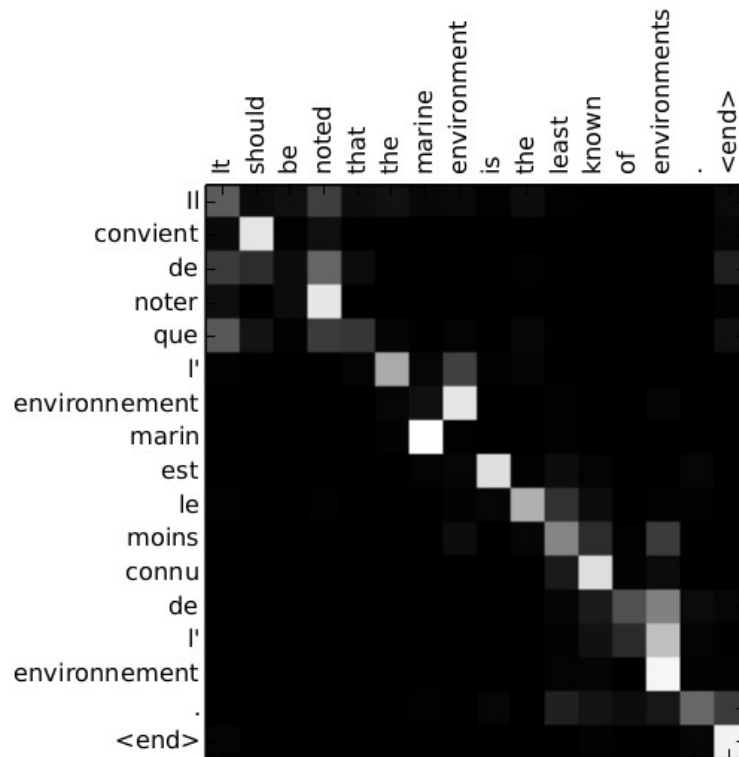
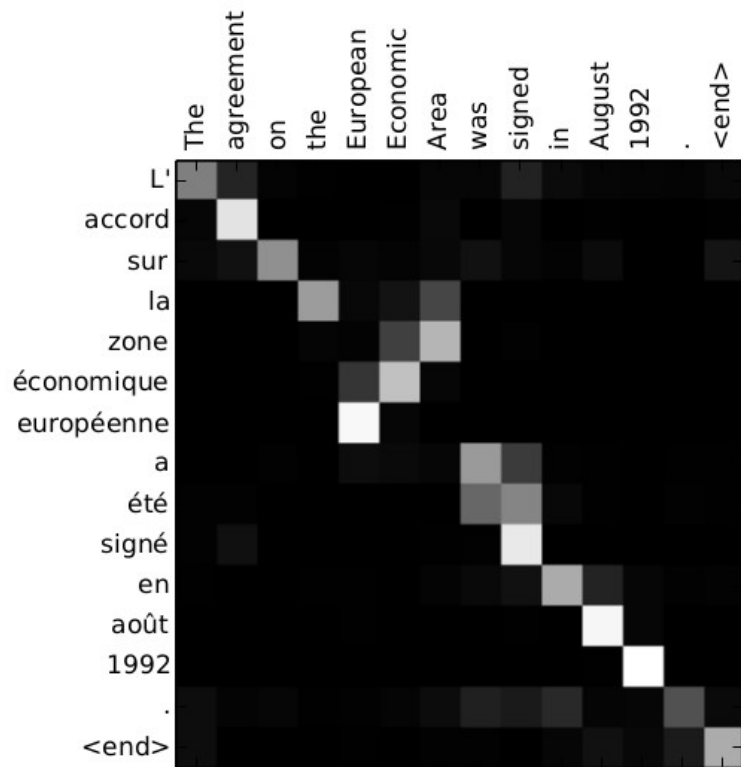
Neural Machine Translation

RNN + Atencja



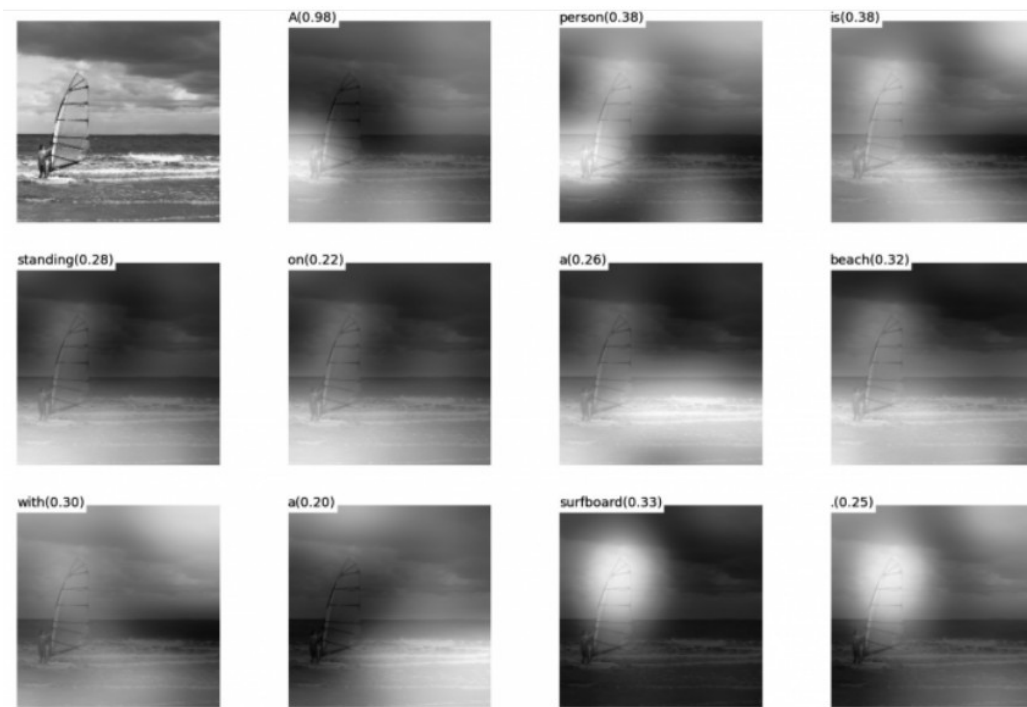
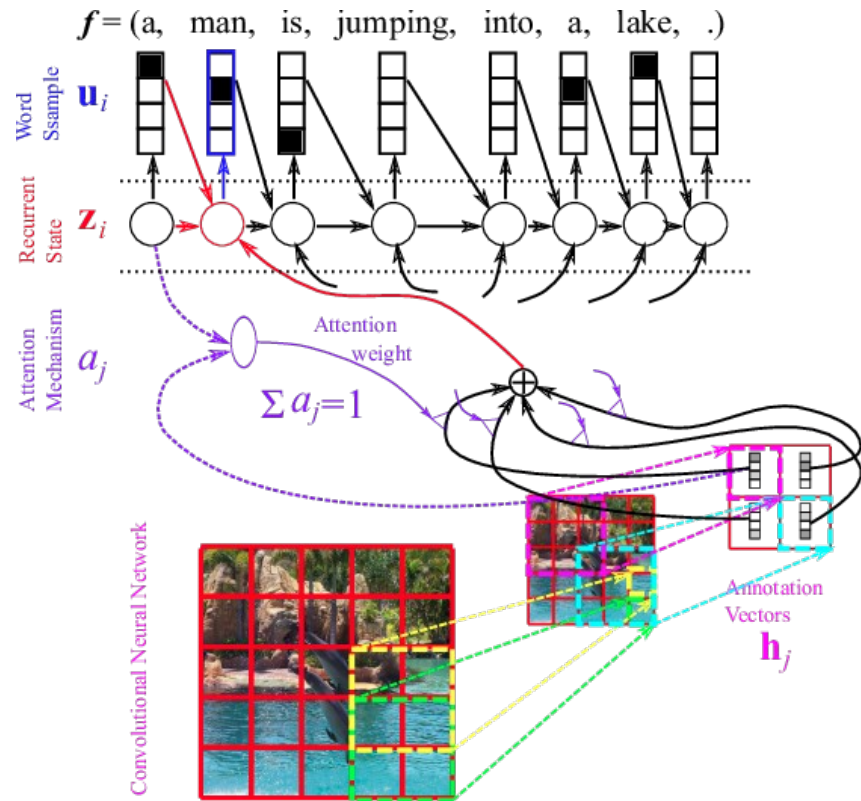
Neural Machine Translation

RNN + Atencja



Atencja w innych problemach

Generowanie opisu z obrazka



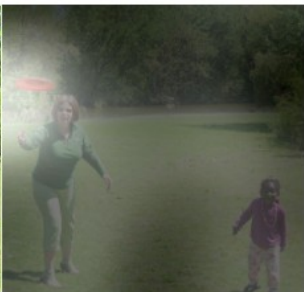
(b) A person is standing on a beach with a surfboard.

Atencja w innych problemach

Generowanie opisu z obrazka



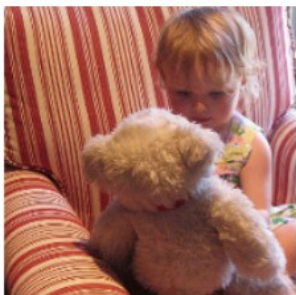
A woman is throwing a frisbee in a park.



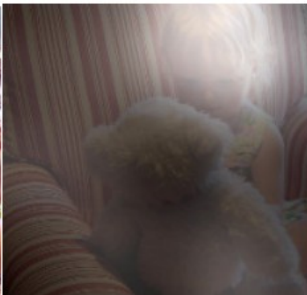
A dog is standing on a hardwood floor.



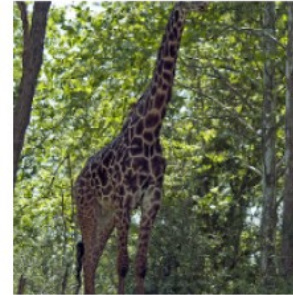
A stop sign is on a road with a mountain in the background.



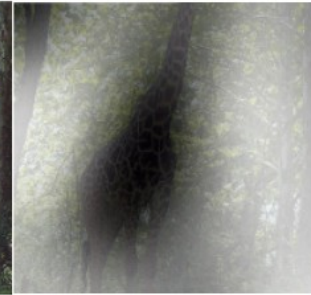
A little girl sitting on a bed with a teddy bear.



A group of people sitting on a boat in the water.



A giraffe standing in a forest with trees in the background.



Atencja w innych problemach

Generowanie opisu z video



+Local+Global: Someone is frying a fish in a pot

+Local: Someone is frying something

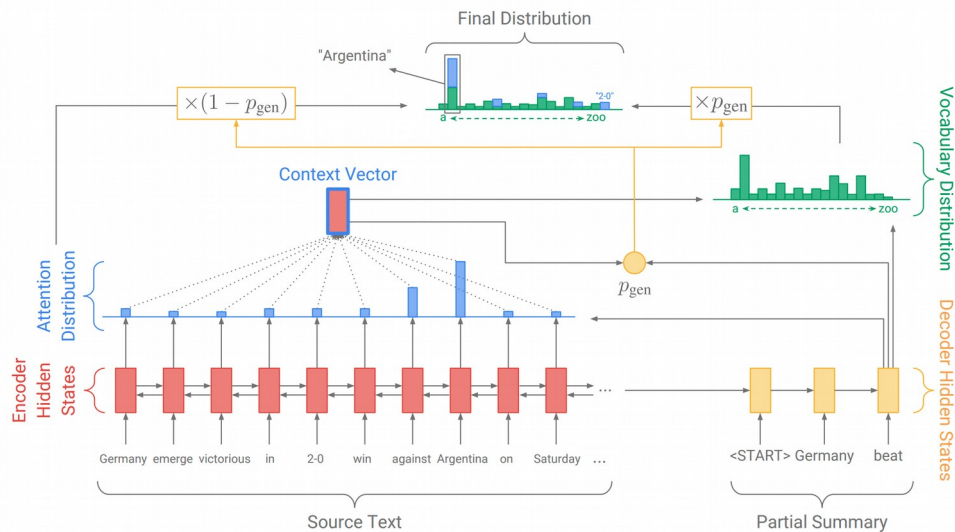
+Global: The person is cooking

Basic: A man cooking its kitchen

Ref: A woman is frying food

Atencja w innych problemach

Streszczanie

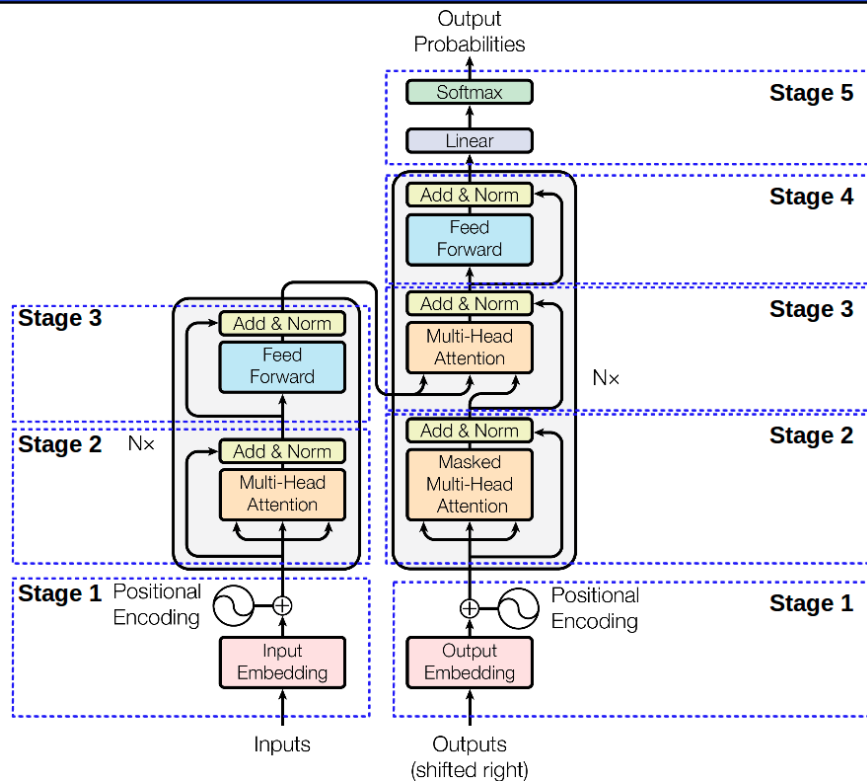


Source Text: Germany emerge victorious in 2-0 win against Argentina on Saturday

Summary: Germany beat Argentina

Atencja w innych problemach

Neural Machine Translation



Model	BLEU		Training Cost (FLOPs)	
	EN-DE	EN-FR	EN-DE	EN-FR
ByteNet [18]	23.75			
Deep-Att + PosUnk [39]		39.2		$1.0 \cdot 10^{20}$
GNMT + RL [38]	24.6	39.92	$2.3 \cdot 10^{19}$	$1.4 \cdot 10^{20}$
ConvS2S [9]	25.16	40.46	$9.6 \cdot 10^{18}$	$1.5 \cdot 10^{20}$
MoE [32]	26.03	40.56	$2.0 \cdot 10^{19}$	$1.2 \cdot 10^{20}$
Deep-Att + PosUnk Ensemble [39]		40.4		$8.0 \cdot 10^{20}$
GNMT + RL Ensemble [38]	26.30	41.16	$1.8 \cdot 10^{20}$	$1.1 \cdot 10^{21}$
ConvS2S Ensemble [9]	26.36	41.29	$7.7 \cdot 10^{19}$	$1.2 \cdot 10^{21}$
Transformer (base model)	27.3	38.1	$3.3 \cdot 10^{18}$	
Transformer (big)	28.4	41.8	$2.3 \cdot 10^{19}$	

Ashish Vaswani, and Others (Łukasz Kaiser). Attention Is All You Need, 2017

Łukasz Kaiser, Attention is all you need attentional neural network models, 2017, <https://www.youtube.com/watch?v=rBCqOTefxvg>

Podsumowanie

1. Intuicja za mechanizmem atencji
2. Istnieją różne rodzaje atencji
3. Stosuje się ją do rozwiązywania różnych problemów

Koniec

Demo: <https://distill.pub/2016/augmented-rnns/>