OpenNMT for Summarization

Scott Chieh-Teng Chang
National Taiwan University

Paul Tardy
Ubiqus
LIUM

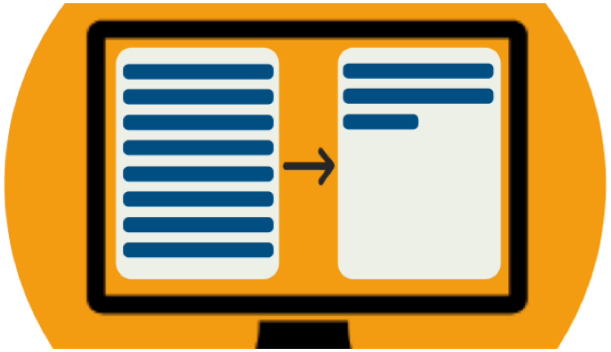
OpenNMT Workshop Paris, March 2018

Introduction

OpenNMT Workshop Paris, March 2018

What is text summarization?

$\mathbf{x}_i \in \{0,1\}^V$ for $i \in \{1, \ldots, M\}$



$\mathbf{y}_i \in \{0,1\}^V$ for $i \in \{1, \ldots, N\}$

$$N < M$$

Extractive

$\mathbf{x}_i \in \{0,1\}^V$ for $i \in \{1,\ldots,M\}$,

$$\operatorname*{arg\,max}_{m \in \{1,\ldots M\}^N} s(\mathbf{x}, \mathbf{x}_{[m_1,\ldots,m_N]})$$

Edmundson et al., (1969), **New Methods in Automatic Extracting**

Abstractive

$\mathbf{x}_i \in \{0,1\}^V$ for $i \in \{1,\ldots,M\}$,

$\mathcal{Y} \subset (\{0,1\}^V, \ldots, \{0,1\}^V)$

$$\arg\max_{\mathbf{y}\in\mathcal{Y}} s(\mathbf{x},\mathbf{y})$$

Rush et al., (2015), **A Neural Attention Model for Abstractive Sentence Summarization**

Evaluation

ROUGE: a word N-gram measure between the model and the reference summary.

ROUGE-1: 1-gram matching
ROUGE-2: 2-gram matching
ROUGE-L: Longest Common Subsequence

ROUGE-1: informativeness
ROUGE-2, ROUGE-L: readability

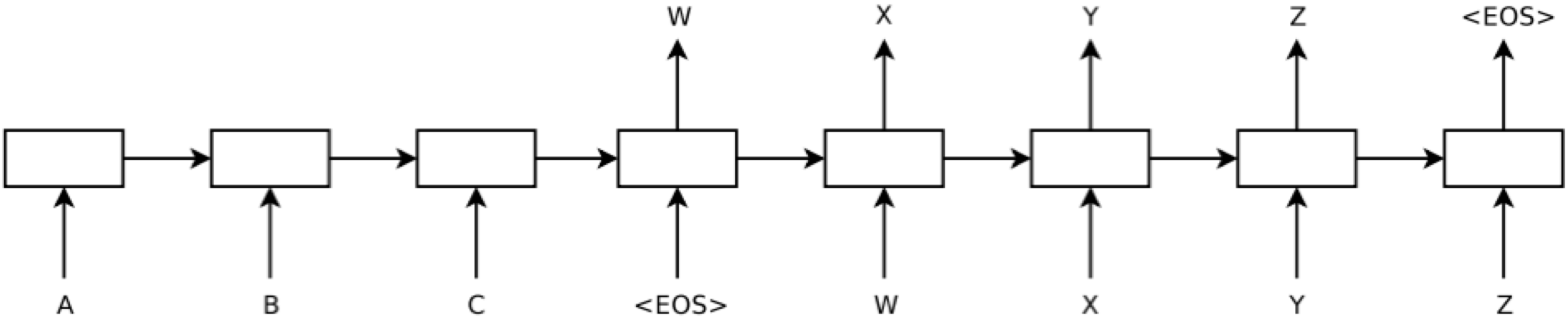
*** The best way of automatic evaluation is still unknown yet.**

Neural Network for Summarization

OpenNMT Workshop Paris, March 2018

Sequence-to-Sequence models

$$p(y_t \mid y_{t-1}, \ldots, y_1, \mathbf{x}) = g(h_t^d, y_{t-1}, c)$$



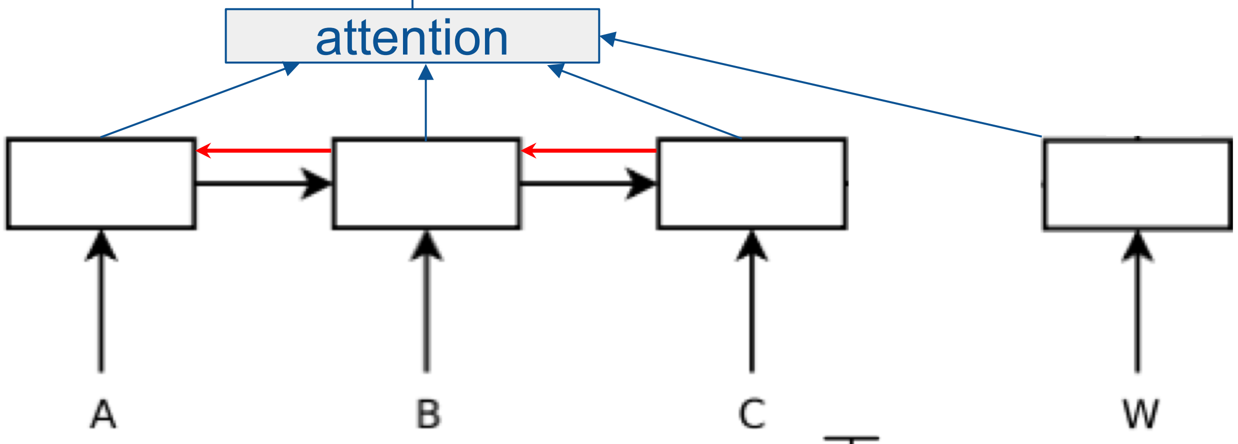
$$h_i^e = f^e(h_{i-1}^e, \tilde{x}_i)$$

$$h_t^d = f^d(h_{t-1}^d, \tilde{y}_{t-1}, c)$$

Cho, et al., (2014), **Learning phrase representations using RNN encoder-decoder for statistical machine translation**
Sutskever et al., (2014), **Sequence to Sequence Learning with Neural Networks**

Attention Mechanism

$$c_t \longrightarrow p(y_t \mid y_{t-1}, \ldots, y_1, \mathbf{x}) = g(h_t^d, y_{t-1}, c_t)$$



$$h_i^e = \left[\overrightarrow{h_i^e}^\top ; \overleftarrow{h_i^e}^\top \right]^\top$$

$$e_{ti} = v_a^\top \tanh(\mathbf{W}_{attn}^e h_t^d + \mathbf{V}_{attn}^e h_i^e)$$

$$\alpha_{ti} = \frac{\exp(e_{ti})}{\sum_{j=1}^{T_x} \exp(e_{tj})}$$

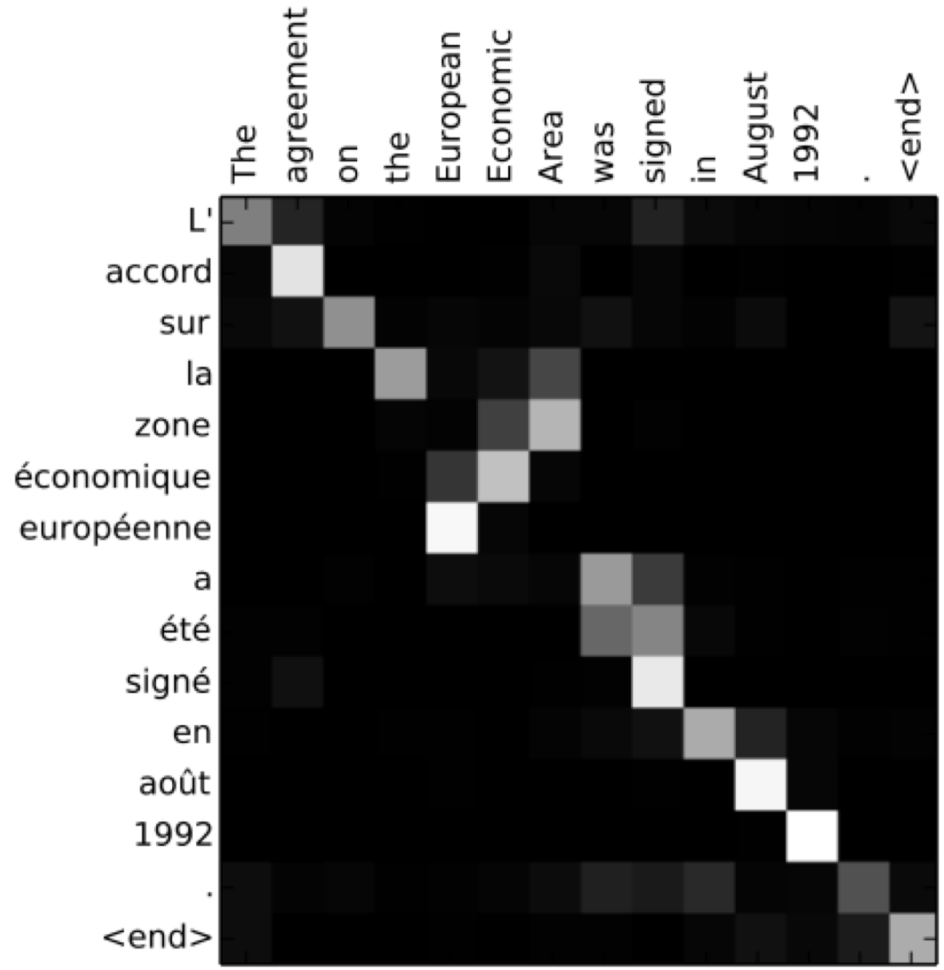
$$c_t = \sum \alpha_{ti} h_i^e$$

Bahdanau et al., (2014) -- **Neural machine translation by jointly learning to align and translate**

Attention Mechanism (2)

the European Economic Area

la zone économique européenne

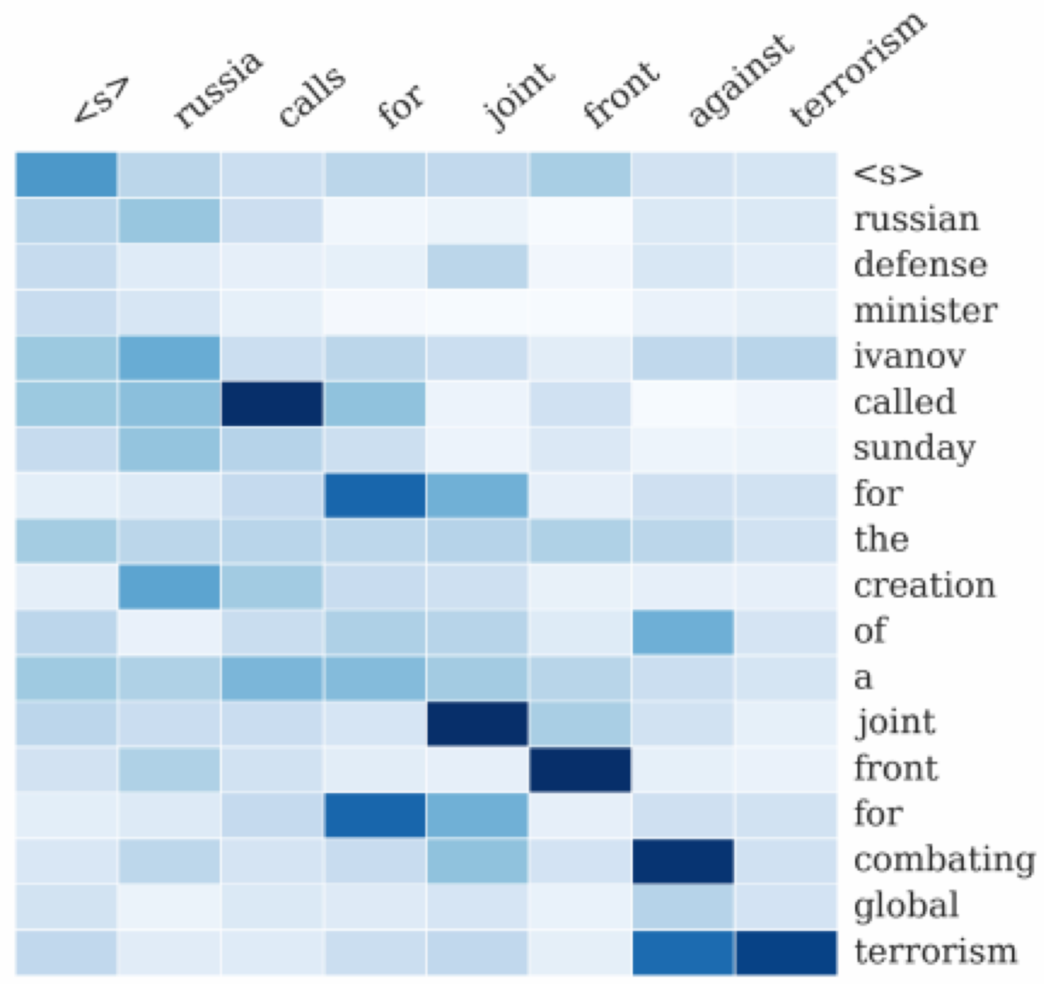


Bahdanau et al., (2014) -- **Neural machine translation by jointly learning to align and translate**

Neural Attention Model for Abstractive Sentence Summarization

Rush, Chopra, Weston, (2015) - **Neural Attention Model for Abstractive Sentence Summarization**

Chopra, Auli, Rush, (2016) - **Abstractive Sentence Summarization with Attentive Recurrent Neural Networks**



Limitations

the ## - year - old son is in stable condition in stable condition , a spokesman says the family 's mother , nancy <unk> , was treated at a hospital , a hospital says the family 's mother , nancy <unk> , was treated at a hospital , and is in stable condition the family is in stable condition and is in stable condition

Unknown tokens

Repetitions

Considering past attention

Temporal Attention

Nallapati et al., (2016)

Lower attention of previously attended inputs

$$\beta_{ti} = \sum_{k=1}^{t-1} \alpha'_{ki} \qquad \alpha_{ti} = \frac{\alpha'_{ti}}{\beta_{ti}}$$

DESTRUCTIVE

Considering past attention

Coverage

See et al., (2017)

Keep memory of previous attention

$$\beta_{ti} = \sum_{k=1}^{t-1} \alpha'_{ki}$$

$$\alpha_{ti} = \text{attention}(h_t^d, h_i^e, \beta_{ti})$$

INFORMATIVE

Considering past attention

Coverage Penalty

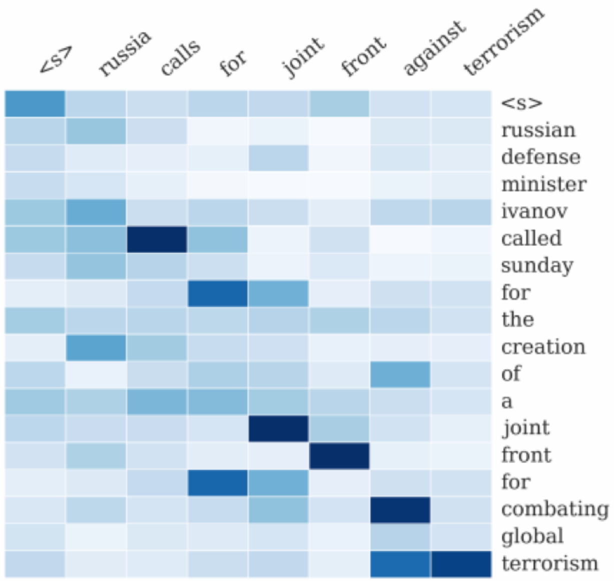
Wu et al., (2016)

---

favor outputs that fully cover the source sentence according to the attention module

$$\arg\max_{Y} \sum_{i=1}^{|X|} \log(\min(1.0, \sum_{j=1}^{|Y|} \alpha_{ij}))$$

INFERENCE ONLY



Considering past attention

Intra-encoder temporal + intra-decoder

Paulus et al., (2017)

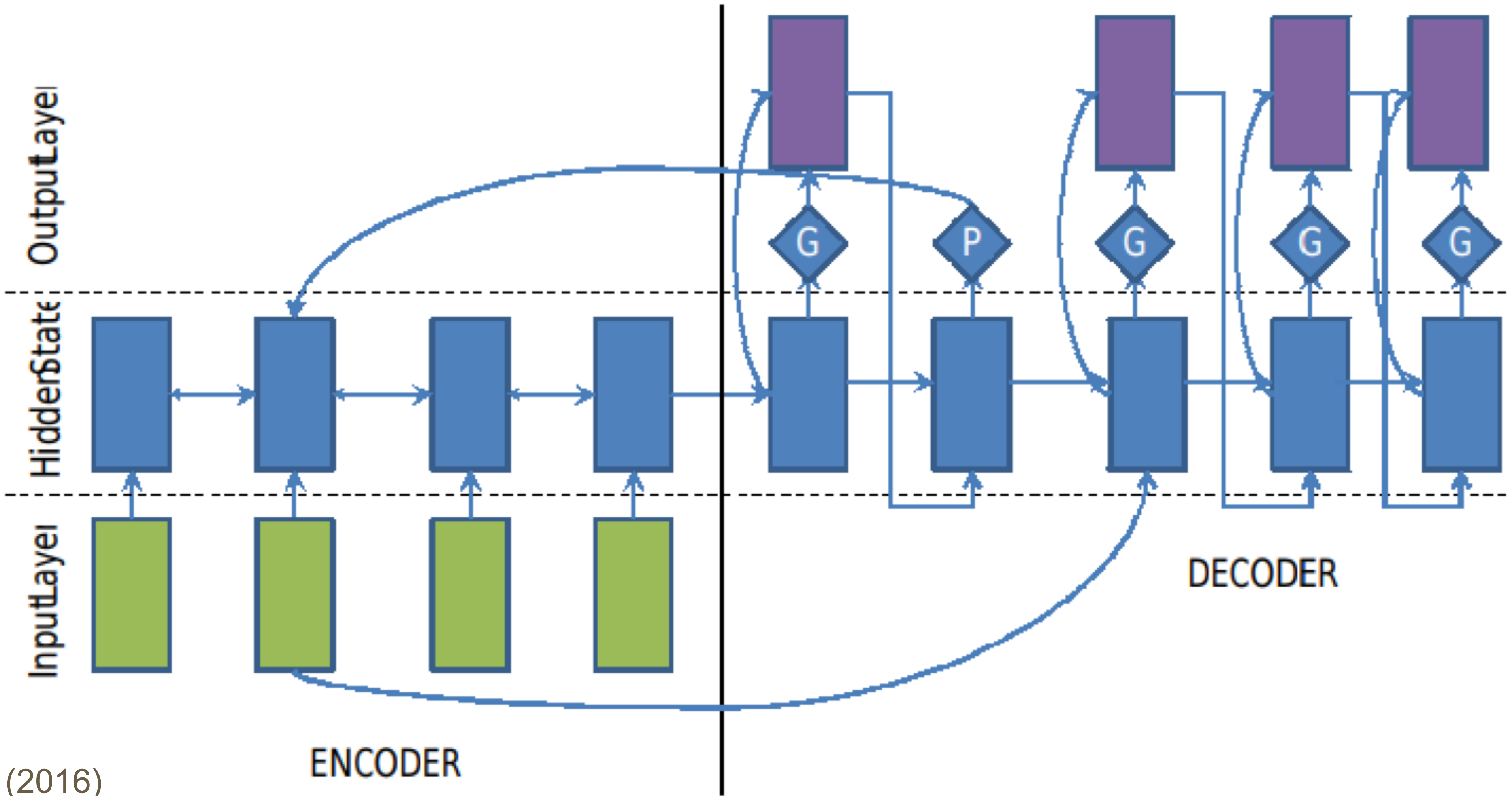
Use attention over past decoding state along with temporal attention over input

$$\alpha^e_{ti} = \text{temporal attention}(h^d_t, h^e_i, \beta_{ti}) \quad \text{for } i \in [1, \ldots, M]$$

$$\alpha^d_{tt'} = \text{attention}(h^d_t, h^d_{tt'}) \quad \text{for } t' \in [1, \ldots, t-1]$$

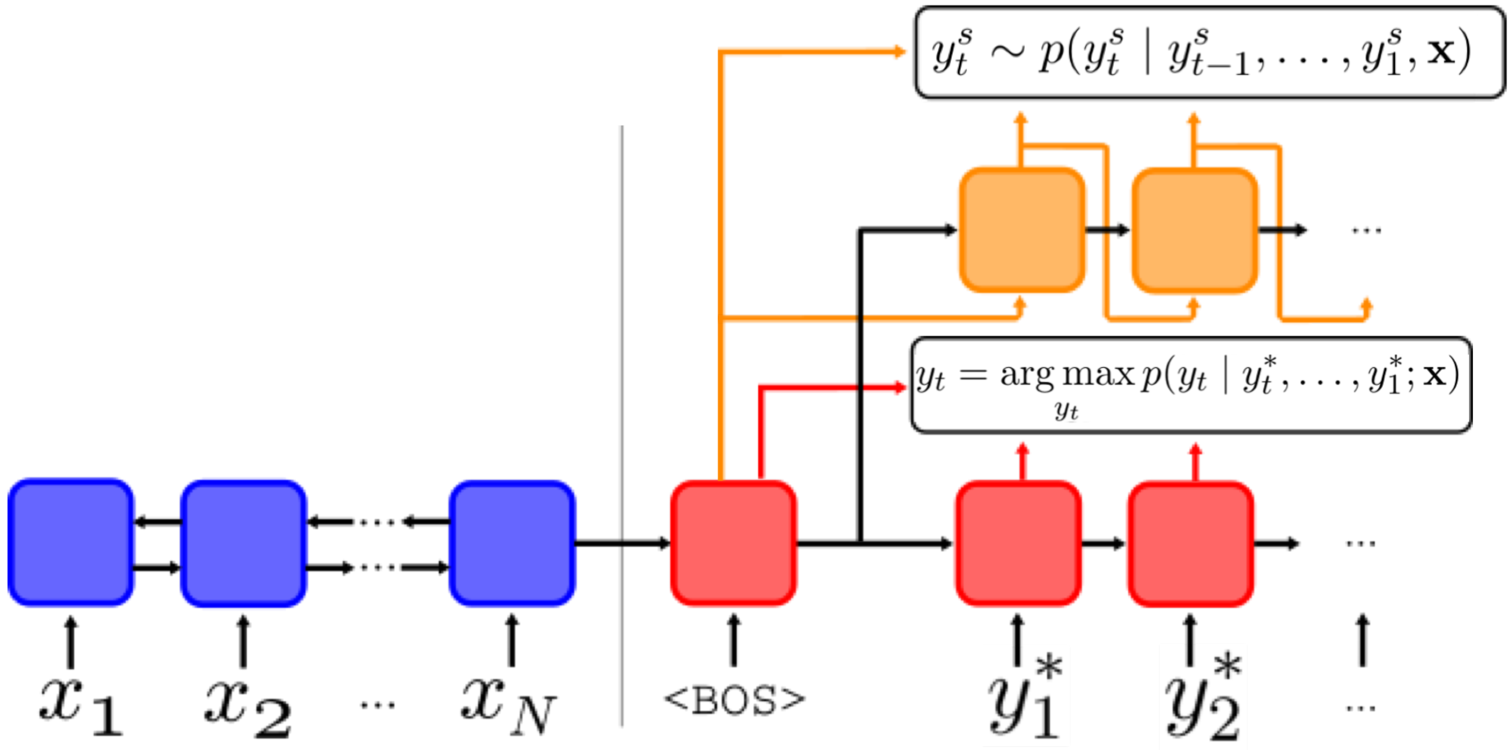
DESTRUCTIVE + INFORMATIVE

Pointing source word with attention
Nallapati et al., (2016)



Nallapati., (2016)

Reinforcement Learning for Abstractive Summarization - Paulus et al., (2017)



$$L_{rl} = (r(y) - r(y^s)) \sum_{t=1}^{n'} \log p(y_t^s | y_1^s, \dots, y_{t-1}^s, x)$$

$$L_{ml} = -\sum_{t=1}^{n'} \log p(y_t^* | y_1^*, \dots, y_{t-1}^*, x)$$

$$L_{mixed} = \gamma L_{rl} + (1 - \gamma) L_{ml}$$

English Summarization Tasks

OpenNMT Workshop Paris, March 2018

English Gigaword

Training pairs	Source length	Target length
~4 million pairs	31.3 words	8.3 words

Model	ROUGE-1	ROUGE-2	ROUGE-L
RAS-Elman Chopra et al., (2016)	33.78	15.97	31.15
words-lvt5k-1sent Nallapati et al., (2016)	35.30	16.64	32.62
ONMT-py - baseline	33.60	16.29	31.45
ONMT-py - copy_attn	**35.51**	**17.35**	**33.17**

English Gigaword - Example

Source:
a us judge denied on tuesday a request for the extradition of former mexican deputy attorney general mario ruiz massieu to mexico where he is charged with a cover-up in his brother 's murder investigation .

Target:
us judge denies mexico 's extradition request for massieu

Output:
us judge denies request for ruiz extradition to mexico

Source:
president bill clinton said thursday he would propose a new plan to congress to reopen many government operations and end the budget impasse , but repeated his intention to veto a republican budget bill .

Target:
clinton congress offer plans to end budget impasse

Prediction:
clinton to propose new plan to end budget impasse

CNN / DailyMail - Results

Training pairs	Source length	Target length
~300k	766 words - 29.74 sentences	53 words - 3.72 sentences

Model	ROUGE-1	ROUGE-2	ROUGE-L
words-lvt2k-temp-att Nallapati et al., (2016)	35.46	13.30	32.65
pointer-generator + coverage See et al., (2017)	39.53	17.28	36.38
ML+RL, with intra-attention Paulus et al., (2017)	**39.87**	15.82	**36.90**
ONMT-py copy_attn + coverage penalty	35.51	**17.35**	33.17
ONMT-py ML+RL + intra-attn	38.27	16.57	35.29

CNN / DailyMail - Example

celtic 's ambitions of persuading manchester city to release jason denayer for another loan deal at parkhead next season are well documented . but what you wish for and what you get are often oceans apart and so may be the case regarding denayer after the pronouncements from city on thursday . put simply , manager manuel pellegrini has confessed that his club 's young academy players are not yet ready to step into his first-team squad . therefore the likes of denayer may soon find themselves returning from loan deals to fill the gap . jason denayer has impressed while playing for celtic on loan this season from manchester city . celtic are keen to retain the services of the 19-year-old belgian defender , but face an uphill struggle to do so . city manager manuel pellegrini has confirmed that the club are looking to invest in homegrown talent . pellegrini 's comments confirm that city will be required to invest heavily in ` homegrown ' talent this summer . current premier league regulations require clubs to carry a minimum eight homegrown players , a group which can include youngsters who have spent three of their formative years at an english club . denayer falls into this category despite hailing from belgium - making him all the more attractive to city next season . pellegrini said : ` i think about the academy , we must be patient . this club build all these things because young players are important . ` maybe it 's not so easy for young players , especially at big clubs , to play in the first team , with the professional squad . ` i 'm sure in the future we will have very important players coming from the academy , they will be part of the squad . ` but it 's important to be patient , important to have years of working the same way . but you never know . maybe some talented young players can do it before . ' denayer would obviously fit the bill . a number of young city professionals , including psv 's karim rekik and marcos lopes with lille , are on loan and may be brought back to plug some of the holes next season . celtic pair virgil van dijk -lrb- left -rrb- and jason denayer -lrb- right -rrb- arrive in milan ahead of the europa league tie .

Prediction:

jason denayer has impressed while playing for celtic this season . city manager manuel pellegrini has confirmed that the club are looking to invest in homegrown talent . city boss manuel pellegrini has confessed that his young academy players are not yet ready to step into his first-team squad .

Target:

jason denayer has impressed for celtic while on loan this season . the parkhead outfit are keen to keen to sign the youngster permanently and hope that parent club manchester city will release him . however , city boss manuel pellegrini has confirmed that the club are looking to strengthen their homegrown talent pool . denayer , 19 , entered city 's youth academy in 2013 and fits the bill .

Chinese Summarization Tasks

OpenNMT Workshop Paris, March 2018

Chinese Character vs. Word

In English: I am very happy today

Sentence: 我今天很開心

I am very happy today

Character

我	今	天	很	開	心
I	now	day god weather	very	Open develop boiling	mind heart inside

Word

我	今天	很	開心
I	today	very	happy

Chinese Dataset - LCSTS

LCSTS: A Large Scale Chinese Short Text Summarization Dataset

Source: Sina Weibo, a Chinese social media

Training data: 2, 400, 591 (Part I)

Testing data: 725 (Part III)

【江西高考被曝替考 有关考生已被警方控制】人民日报记者吴齐强消息，江西高考被曝光替考，7日中午江西省教育厅发布消息称，接到有人组织替考的举报后，江西省教育厅、江西省教育考试院立即部署南昌市教育考试院，联合南昌市警方开展调查核实，有关考生已被警方控制。调查进展情况将及时向社会公布。

HU, Baotian; CHEN, Qingcai; ZHU, Fangze(2015). **LCSTS: A large scale chinese short text summarization dataset**

Benchmark on LCSTS

NMT

Model	encoder	decoder	ROUGE-1	ROUGE-2	ROUGE-L
RNN+Context (Hu et al., 2015)	char / 3000	char / 3000	29.9	17.4	27.2
CopyNet (Gum et al., 2016)	word / 3000	word / 10000	35	22.3	32
Distraction (Chen et al., 2016)	char / 4000	char / 4000	35.2	22.6	32.5
DGRD (Li et al., 2017)	char / -	char / -	36.99	24.15	34.21
MRT (Ayana et al., 2016)	char / 3500	char / 3500	37.87	25.43	35.33

A Hybrid Word-Character Model for Abstractive Summarization

HWC1	char / 10599	char / 8248	38.81	26.01	35.95
HWC3	**word / 500000**	**char / 8248**	**46.1**	**33.61**	**43.46**

Chang, et al., (2018), **A Hybrid Word-Character Model for Abstractive Summarization**
https://arxiv.org/abs/1802.09968

Thank you for your attention

OpenNMT Workshop Paris, March 2018

backup slide

OpenNMT Workshop Paris, March 2018

Experiments

Method		Encoder		Decoder		Evaluation		
		based	vocab size	based	vocab size	ROUGE-1	ROUGE-2	ROUGE-L
RNN	(Hu et al., 2015)	char	4000	char	4000	21.5	8.9	18.6
		word	50000	word	50000	17.7	8.5	15.8
RNN+context		char	3000	char	3000	29.9	17.4	27.2
		word	10000	word	10000	26.8	16.1	24.1
CopyNet	(Gum et al., 2016)	char	3000	char	3000	34.4	21.6	31.3
		word	10000	word	10000	35	22.3	32
Distraction	(Chen et al., 2016)	char	4000	char	4000	35.2	22.6	32.5
DGRD	(Li et al., 2017)	char	-	char	-	36.99	24.15	34.21
MRT	(Ayana et al., 2016)	char	3500	char	3500	37.87	25.43	35.33
HWC	HWC_1	char	10599	char	8248	**38.81**	**26.01**	**35.95**
	HWC_2	word	50000	char	8248	40.95	28.58	38.34
	HWC_3	word	500000	char	8248	**46.1**	**33.61**	**43.46**
	HWC_4	word	961213	char	8248	44.79	32.49	42.12

Table 2: ROUGE-F1 on LCSTS dataset.