



# A Unified Model for Extractive and Abstractive Summarization using Inconsistency Loss



Wan-Ting Hsu

National Tsing Hua University



Chieh-Kai Lin

National Tsing Hua University

Project page



# Outline

- Motivation
- Our Method
- Training Procedures
- Experiments and Results
- Conclusion

# Outline

- **Motivation**
- Our Method
- Training Procedures
- Experiments and Results
- Conclusion

# Textual Media

“ People spend 12 hours everyday consuming media in 2018.

– eMarketer

<https://www.emarketer.com/topics/topic/time-spent-with-media>

# Textual Media

The screenshot shows the CNN website interface. At the top, there is a navigation bar with the CNN logo and links to various news categories: World, U.S. Politics, Money, Entertainment, Tech, Sport, Travel, Style, Health, Video, and VR. It also indicates 'International Edition' with a search icon and a menu icon.

## Thai cave rescue begins

**LIVE UPDATES**

**Extraction of 12 boys and their soccer coach could take days, officials say**

**Full story:** 13 divers descend into caves to bring out soccer team

**Heartbreaking letter:** Youngest boy dreams of fried chicken

**Photos:** Thai soccer team trapped in cave

**Hero:** Former Thai navy SEAL diver dies in cave

**Watch:** A virtual look inside what cave rescue could look like

**Health:** How a lack of oxygen is affecting the Thai soccer team

**Top stories**

**Pompeo dismisses North Korea's 'gangster' comments**

**Japan floods leave at least 55 dead; 2 million flee homes**

**US destroyers sail through Taiwan Strait**

**Other stories:** Croatia eliminate host Russia in Cup, England beat Sweden to reach first semis, US citizens warned to shelter amid Haiti protests, US soybean farmers say that China tariffs could do 'serious damage', US candidate: 'Outside of Western civilization, there is only barbarism', Washington Post: Trump's personal calls to world leaders leave aides in the dark

consuming media in 2018.

– eMarketer

[www.emarketer.com/topics/topic/time-spent-with-media](http://www.emarketer.com/topics/topic/time-spent-with-media)

# Textual Media



in 2018.  
Marketer  
[spent-with-media](#)

## Textual Media



**CNN** World | U.S. Politics | Money | Entertainment | Tech | Sport | Travel | Style | Health | Video | VR International Edition +

# Thai cave rescue begins

**LIVE UPDATES**

**Extraction of 12 boys and their soccer coach could take days, officials say**

**Full story:** 13 divers descend into caves to bring out soccer team

**Heartbreaking letter:** Youngest boy dreams of fried chicken

**Photos:** Thai soccer team trapped in cave

**Hero:** Former Thai navy SEAL diver dies in cave

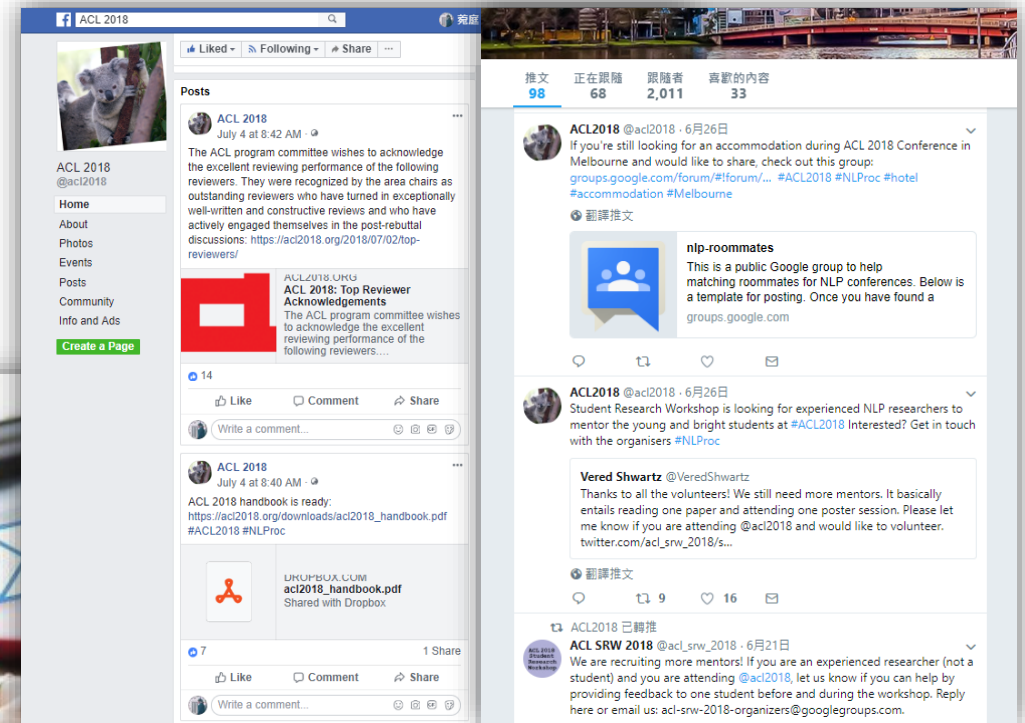
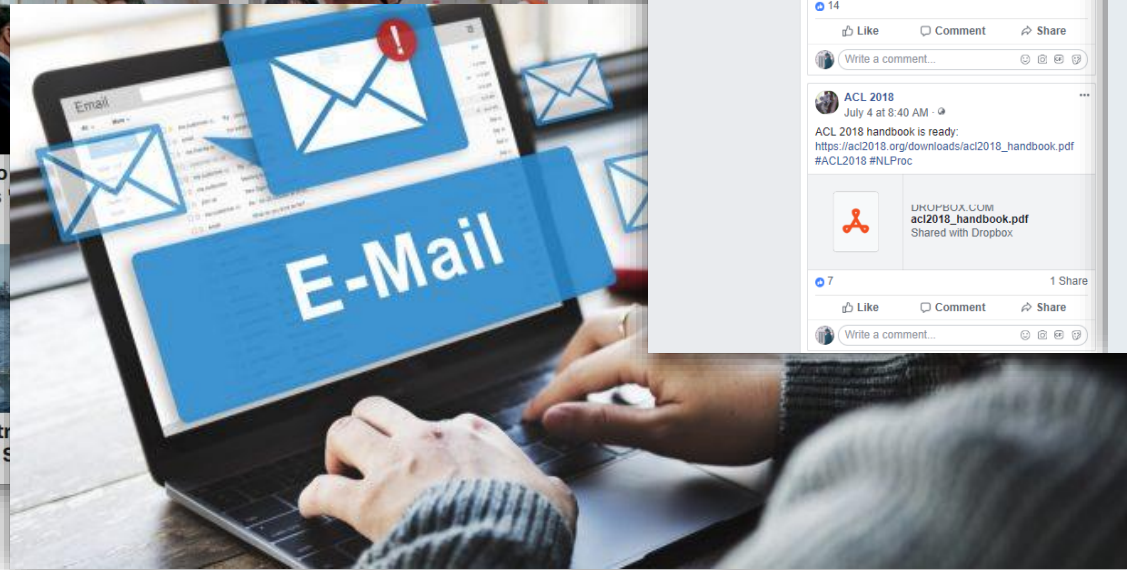
**Watch:** A virtual look inside what cave rescue could look like

**Health:** How a lack of oxygen is affecting the Thai soccer team

Top stories

Pompeo Korea's

US destr Taiwan S



ACL 2018 @acl2018

Posts

ACL 2018 July 4 at 8:42 AM  
The ACL program committee wishes to acknowledge the excellent reviewing performance of the following reviewers. They were recognized by the area chairs as outstanding reviewers who have turned in exceptionally well-written and constructive reviews and who have actively engaged themselves in the post-rebuttal discussions: <https://acl2018.org/2018/07/02/top-reviewers/>

ACL 2018: UKG ACL 2018: Top Reviewer Acknowledgements  
The ACL program committee wishes to acknowledge the excellent reviewing performance of the following reviewers...

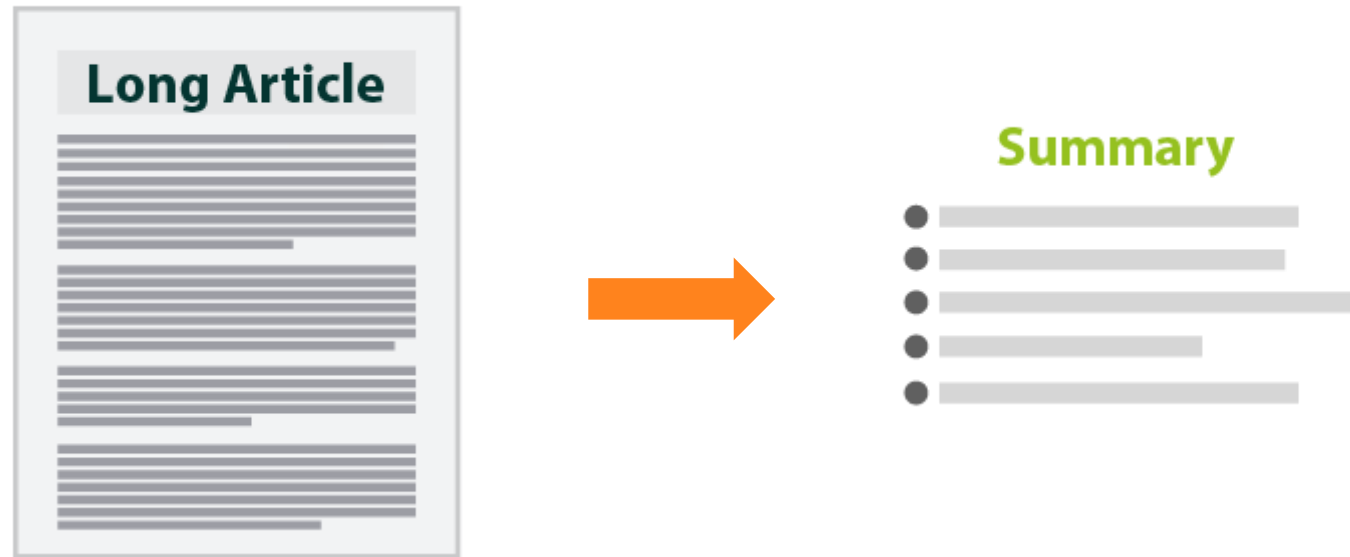
ACL 2018 July 4 at 8:40 AM  
ACL 2018 handbook is ready: [https://acl2018.org/downloads/acl2018\\_handbook.pdf](https://acl2018.org/downloads/acl2018_handbook.pdf) #ACL2018 #NLProc

UKUPBUX.COM acl2018\_handbook.pdf Shared with Dropbox

ACL 2018 已轉推  
ACL SRW 2018 @acl\_srw\_2018 · 6月21日  
We are recruiting more mentors! If you are an experienced researcher (not a student) and you are attending @acl2018, let us know if you can help by providing feedback to one student before and during the workshop. Reply here or email us: [acl-srw-2018-organizers@googlegroups.com](mailto:acl-srw-2018-organizers@googlegroups.com).

# Text Summarization

- To condense a piece of text to a shorter version while maintaining the important points





# Examples of Text Summarization

- Article headlines
- Meeting minutes
- Movie/book reviews
- Bulletins (weather forecasts/stock market reports)

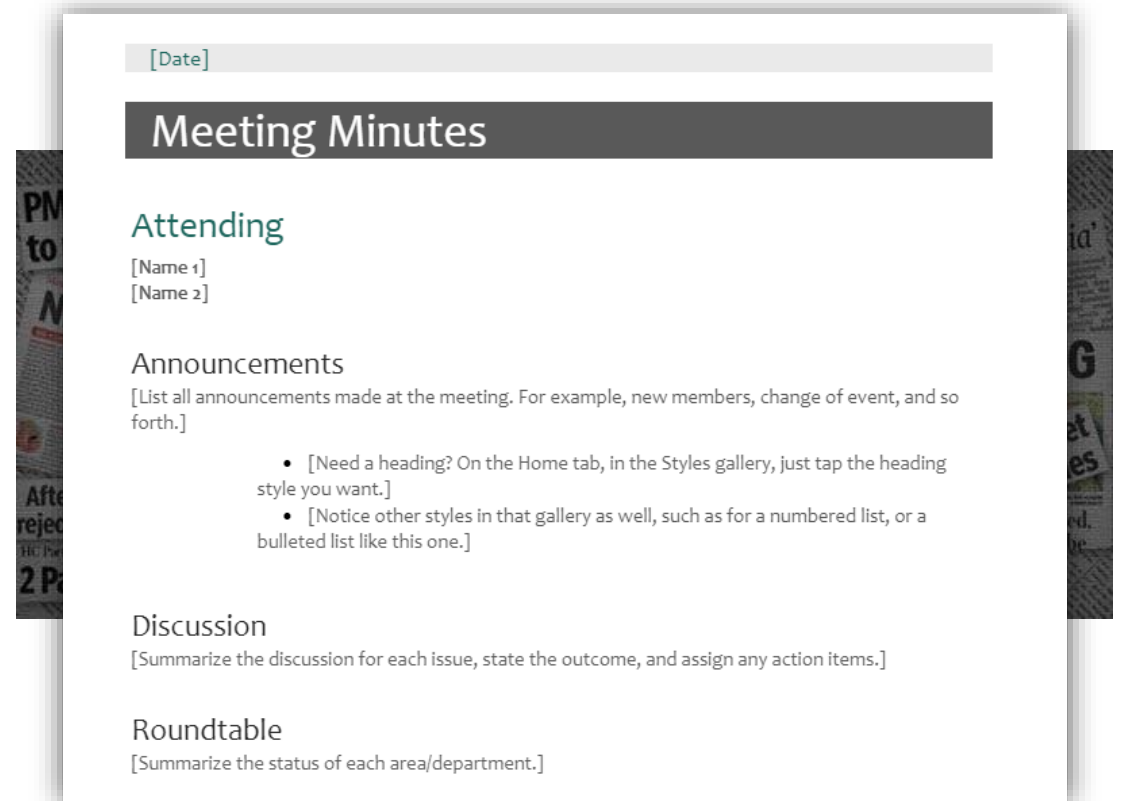
## Examples of Text Summarization

- Article headlines
- Meeting minutes
- Movie/book reviews
- Bulletins (weather forecasts/stock market reports)



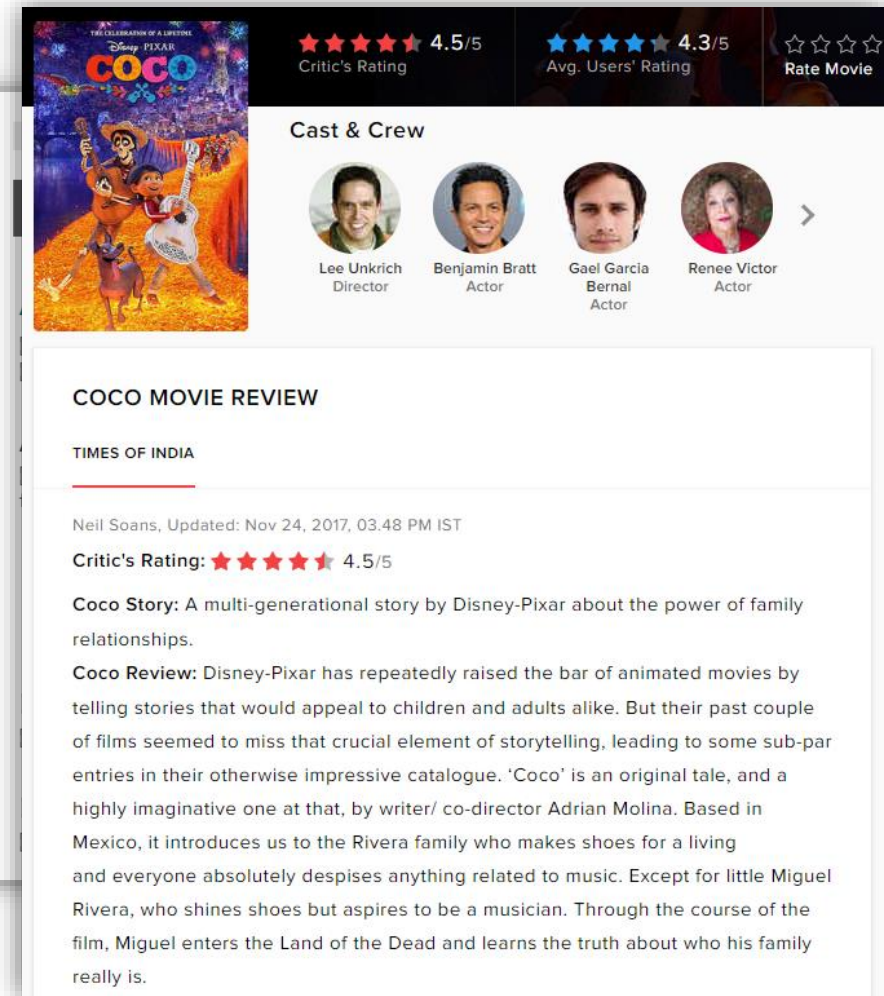
## Examples of Text Summarization

- Article headlines
- Meeting minutes
- Movie/book reviews
- Bulletins (weather forecasts/stock market reports)



## Examples of Text Summarization

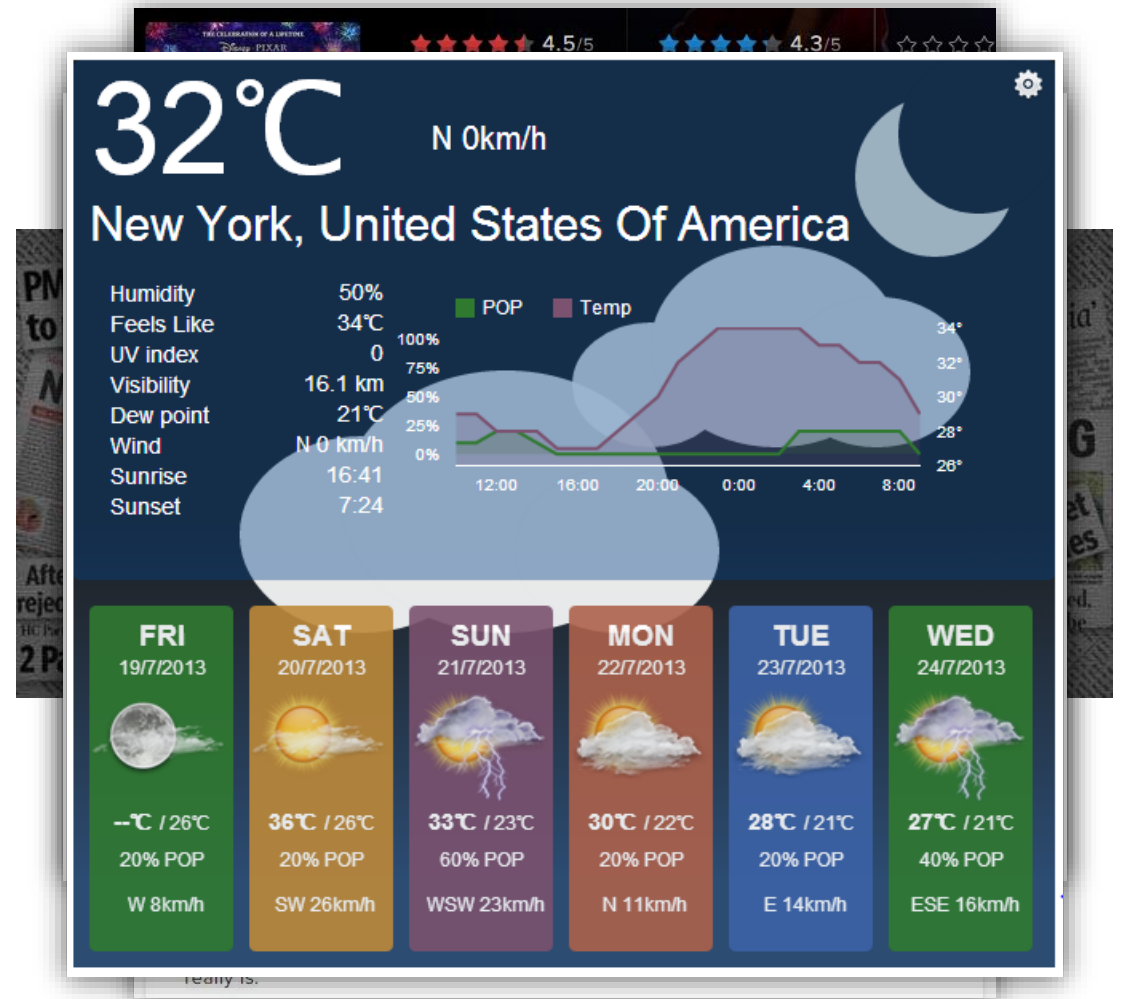
- Article headlines
- Meeting minutes
- Movie/book reviews
- Bulletins (weather forecasts/stock market reports)



The screenshot displays the IMDb page for the movie 'Coco'. At the top, it features the movie poster, a Critic's Rating of 4.5/5 (5 stars), and an Avg. Users' Rating of 4.3/5 (4.3 stars). Below the ratings is the 'Cast & Crew' section, listing Lee Unkrich as Director, Benjamin Bratt as Actor, Gael Garcia Bernal as Actor, and Renee Victor as Actor. The main content is a review titled 'COCO MOVIE REVIEW' from 'TIMES OF INDIA', dated Nov 24, 2017, 03:48 PM IST. The review snippet begins with 'Coco Story: A multi-generational story by Disney-Pixar about the power of family relationships.' and continues with 'Coco Review: Disney-Pixar has repeatedly raised the bar of animated movies by telling stories that would appeal to children and adults alike. But their past couple of films seemed to miss that crucial element of storytelling, leading to some sub-par entries in their otherwise impressive catalogue. 'Coco' is an original tale, and a highly imaginative one at that, by writer/ co-director Adrian Molina. Based in Mexico, it introduces us to the Rivera family who makes shoes for a living and everyone absolutely despises anything related to music. Except for little Miguel Rivera, who shines shoes but aspires to be a musician. Through the course of the film, Miguel enters the Land of the Dead and learns the truth about who his family really is.'

## Examples of Text Summarization

- Article headlines
- Meeting minutes
- Movie/book reviews
- Bulletins (weather forecasts/stock market reports)



# Automatic Text Summarization

- To condense a piece of text to a shorter version while maintaining the important points

## Extractive Summarization



select text from the article

## Abstractive Summarization

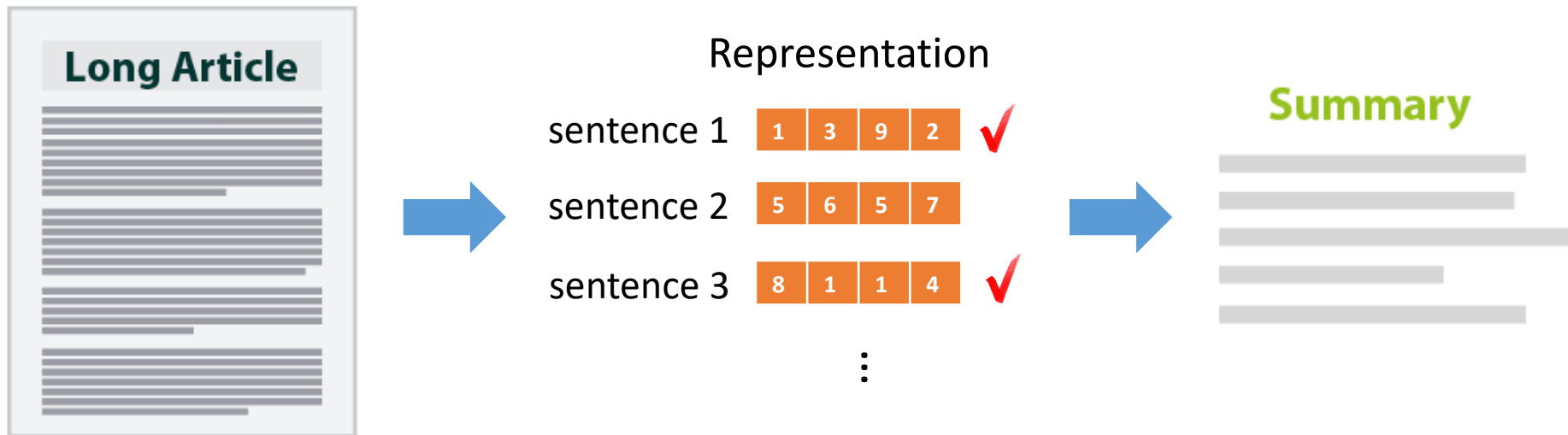


generate the summary word-by-word

# Extractive Summarization



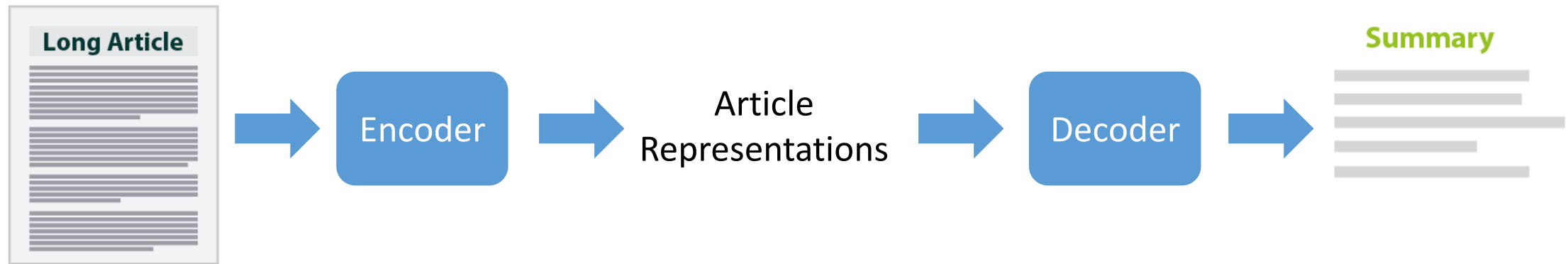
- Select phrases or sentences from the source document



- Shen, D.; Sun, J.-T.; Li, H.; Yang, Q.; and Chen, Z. 2007. Document summarization using conditional random fields. IJCAI 2007.
- Kågebäck, M., Mogren, O., Tahmasebi, N., & Dubhashi, D. Extractive Summarization using Continuous Vector Space Models. EACL 2014.
- Cheng, J., and Lapata, M. Neural summarization by extracting sentences and words. ACL 2016.
- Ramesh Nallapati, Feifei Zhai, and Bowen Zhou. Summarunner: A recurrent neural network based sequence model for extractive summarization of documents. AACL 2017

## Abstractive Summarization

- Select phrases or sentences from the source document



- Alexander M Rush, Sumit Chopra, and Jason Weston. A neural attention model for abstractive sentence summarization. EMNLP 2015.
- Ramesh Nallapati, Bowen Zhou, Cicero dos Santos, Caglar Gulcehre, and Bing Xiang. Abstractive text summarization using sequence-to-sequence rnns and beyond. CoNLL 2016.
- Abigail See, Peter J Liu, and Christopher D Manning. Get to the point: Summarization with pointer-generator networks. ACL 2017.
- Romain Paulus, Caiming Xiong, and Richard Socher. A deep reinforced model for abstractive summarization. ICLR 2018.
- Fan, Angela, David Grangier, and Michael Auli. Controllable abstractive summarization. arXiv preprint arXiv:1711.05217 (2017).



## Motivation

- Extractive summary  
(select sentences):
  - important, correct
  - incoherent or not concise

not concise

Italian artist Johannes Stoetter has painted two naked women to look like a chameleon.

The 37-year-old has previously transformed his models into frogs and parrots but this may be his most intricate and impressive artwork to date.

## Motivation

- Extractive summary  
(select sentences):
  - important, correct
  - incoherent or not concise
- Abstractive summary  
(generate word-by-word):
  - readable, concise
  - may lose or mistake some facts

not concise

Italian artist Johannes Stoetter has painted two naked women to look like a chameleon.

The 37-year-old has previously transformed his models into frogs and parrots but this may be his most intricate and impressive artwork to date.



concise

Johannes Stoetter has previously transformed his models into frogs and parrots but this chameleon may be his most impressive artwork to date.

## Motivation

- Extractive summary  
(select sentences):
  - important, correct
  - incoherent or not concise
- Abstractive summary  
(generate word-by-word):
  - readable, concise
  - may lose or mistake some facts

not concise

Italian artist Johannes Stoetter has painted two naked women to look like a chameleon.

The 37-year-old has previously transformed his models into frogs and parrots but this may be his most intricate and impressive artwork to date.



concise

~~Justin Bieber~~  
Johannes Stoetter has previously transformed his models into frogs and parrots but this chameleon may be his most impressive artwork to date.

## Motivation

- Extractive summary  
(select sentences):
  - important, correct
  - incoherent or not concise
- Abstractive summary  
(generate word-by-word):
  - readable, concise
  - may lose or mistake some facts
- Unified summary:
  - important, correct
  - readable, concise

not concise

Italian artist Johannes Stoetter has painted two naked women to look like a chameleon.

The 37-year-old has previously transformed his models into frogs and parrots but this may be his most intricate and impressive artwork to date.



concise

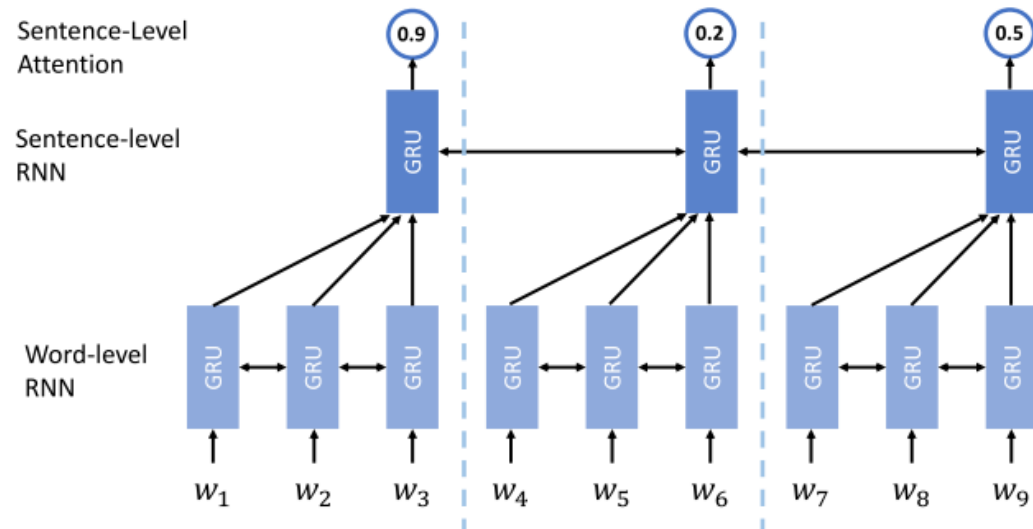
~~Justin Bieber~~  
Johannes Stoetter has previously transformed his models into frogs and parrots but this chameleon may be his most impressive artwork to date.

# Outline

- Motivation
- **Our Method**
- Training Procedures
- Experiments and Results
- Conclusion

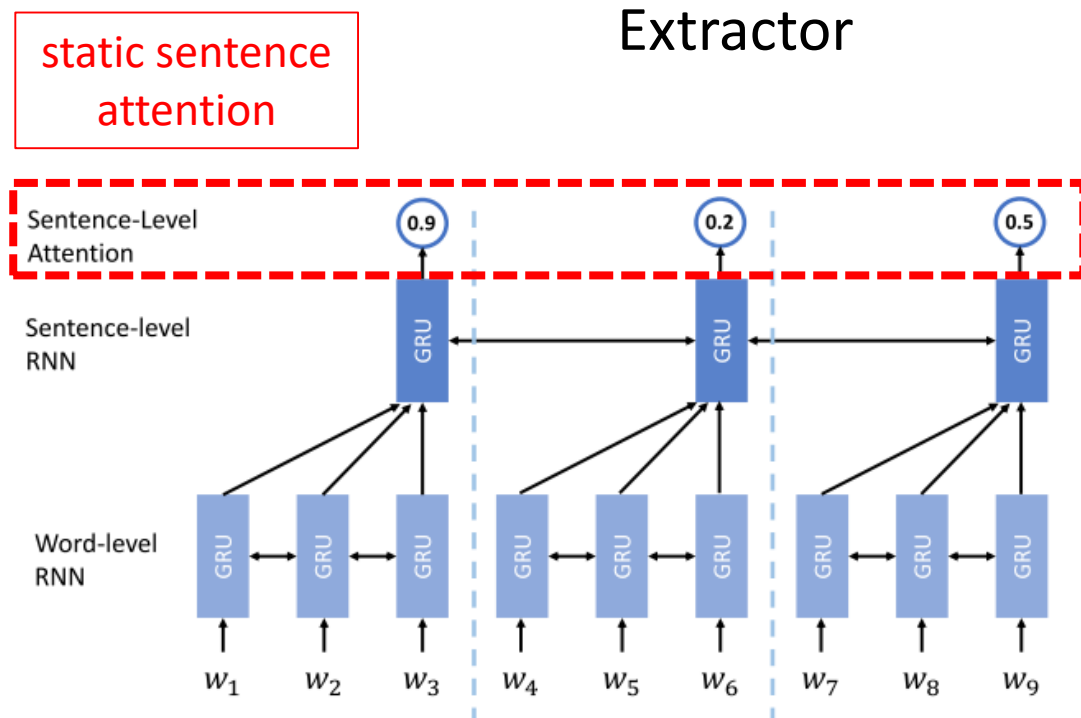
# Models

## Extractor



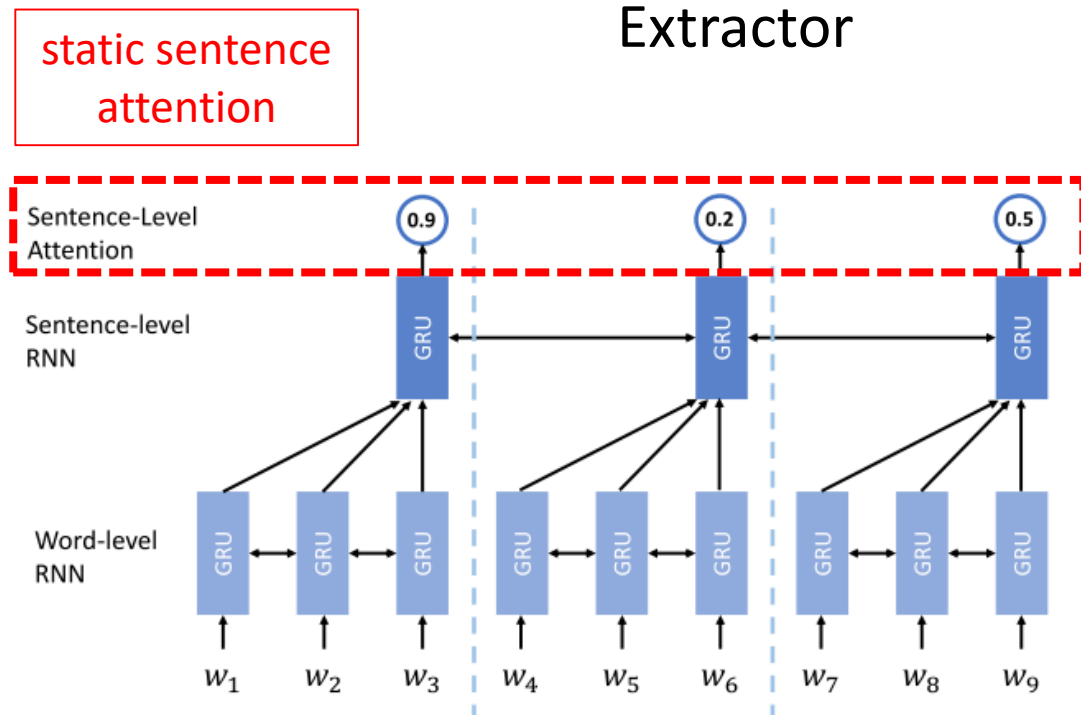
Ramesh Nallapati, Feifei Zhai, and Bowen Zhou. Summarunner: A recurrent neural network based sequence model for extractive summarization of documents. AAAI 2017

# Models

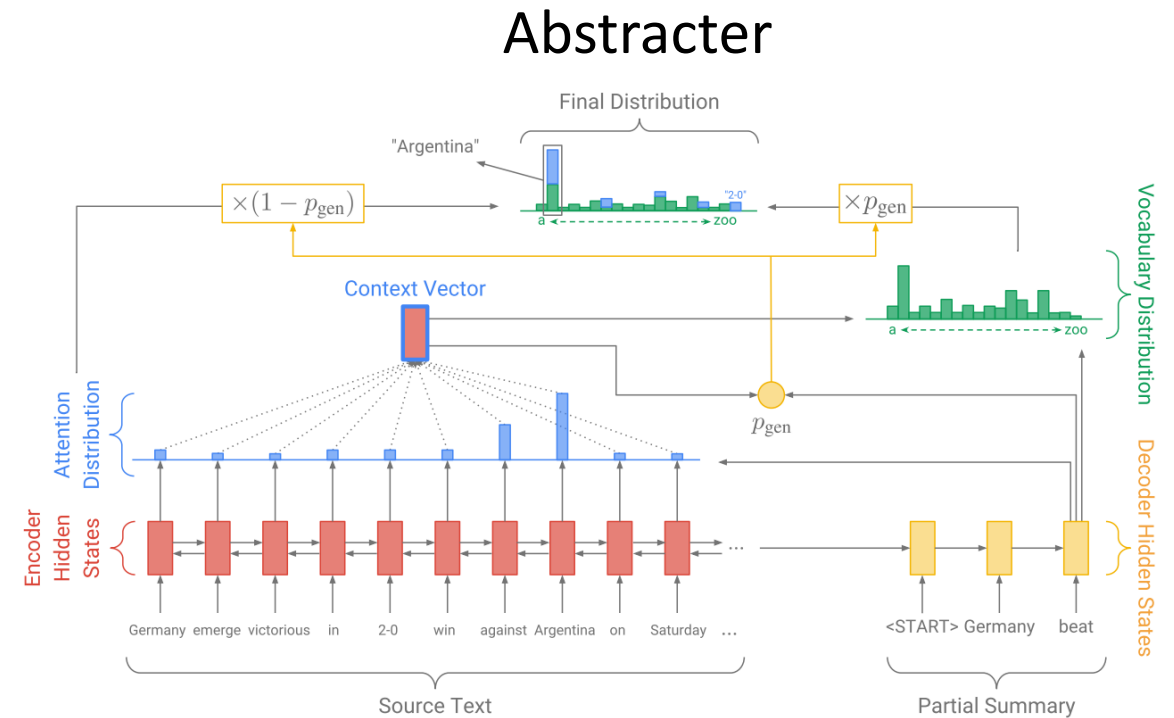


Ramesh Nallapati, Feifei Zhai, and Bowen Zhou. Summarunner: A recurrent neural network based sequence model for extractive summarization of documents. AACL 2017

## Models



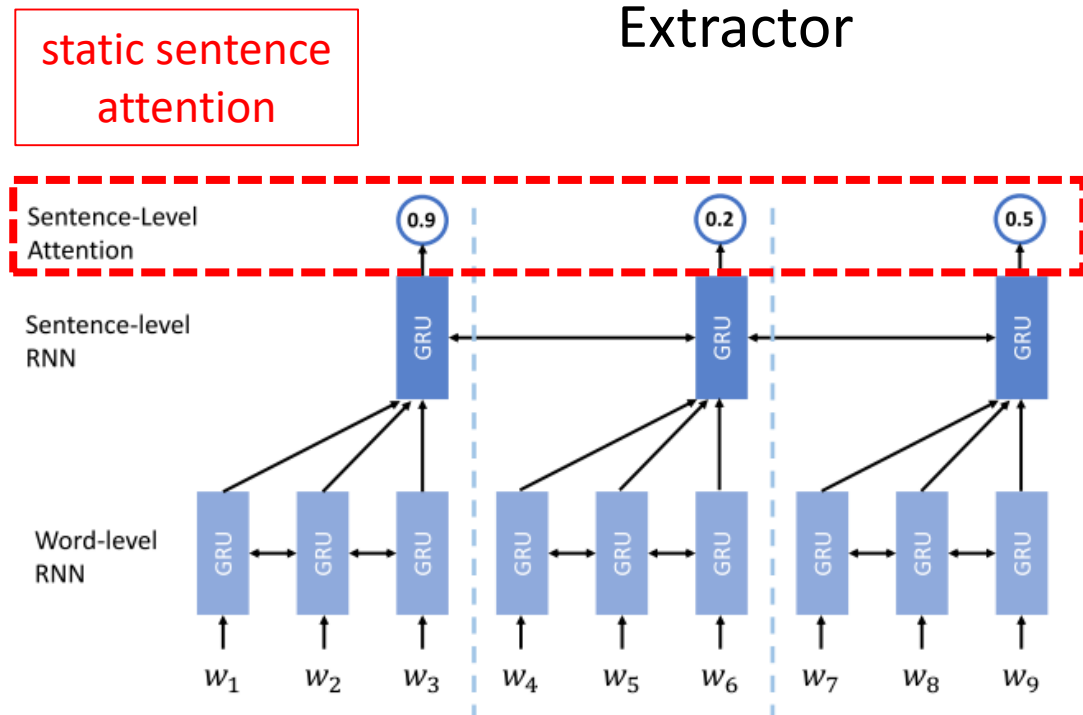
Ramesh Nallapati, Feifei Zhai, and Bowen Zhou. Summarunner: A recurrent neural network based sequence model for extractive summarization of documents. AACL 2017



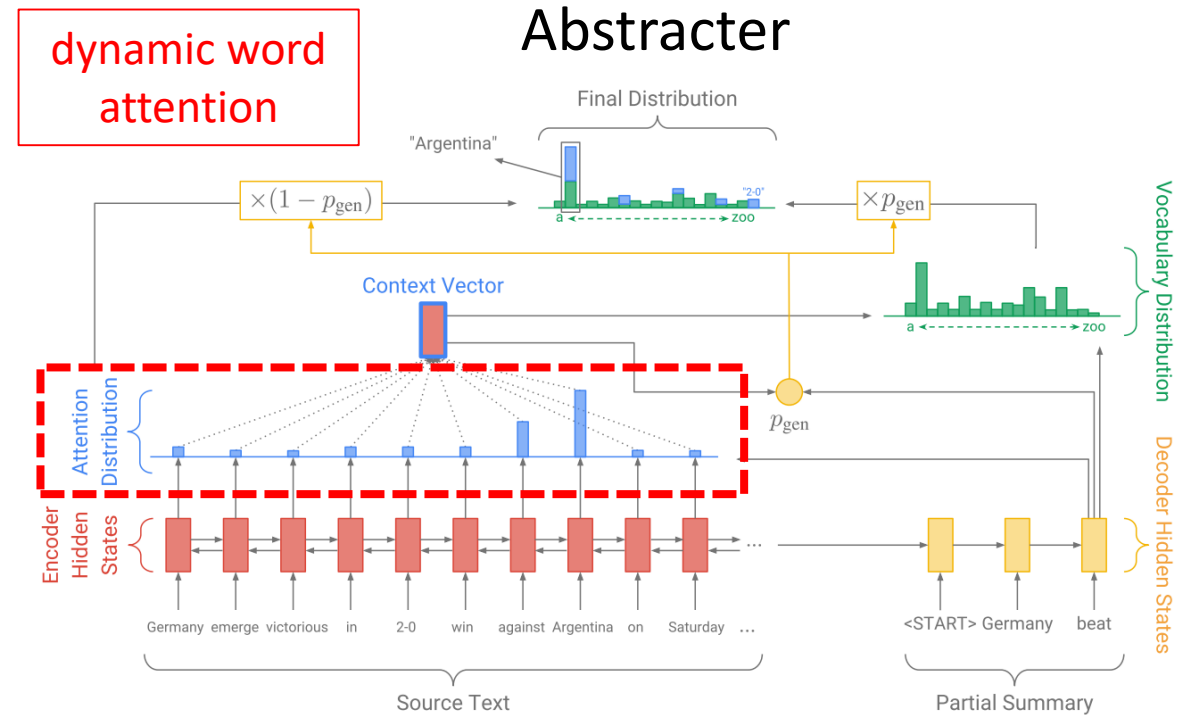
Abigail See, Peter J Liu, and Christopher D Manning. Get to the point: Summarization with pointer-generator networks. ACL 2017



# Models

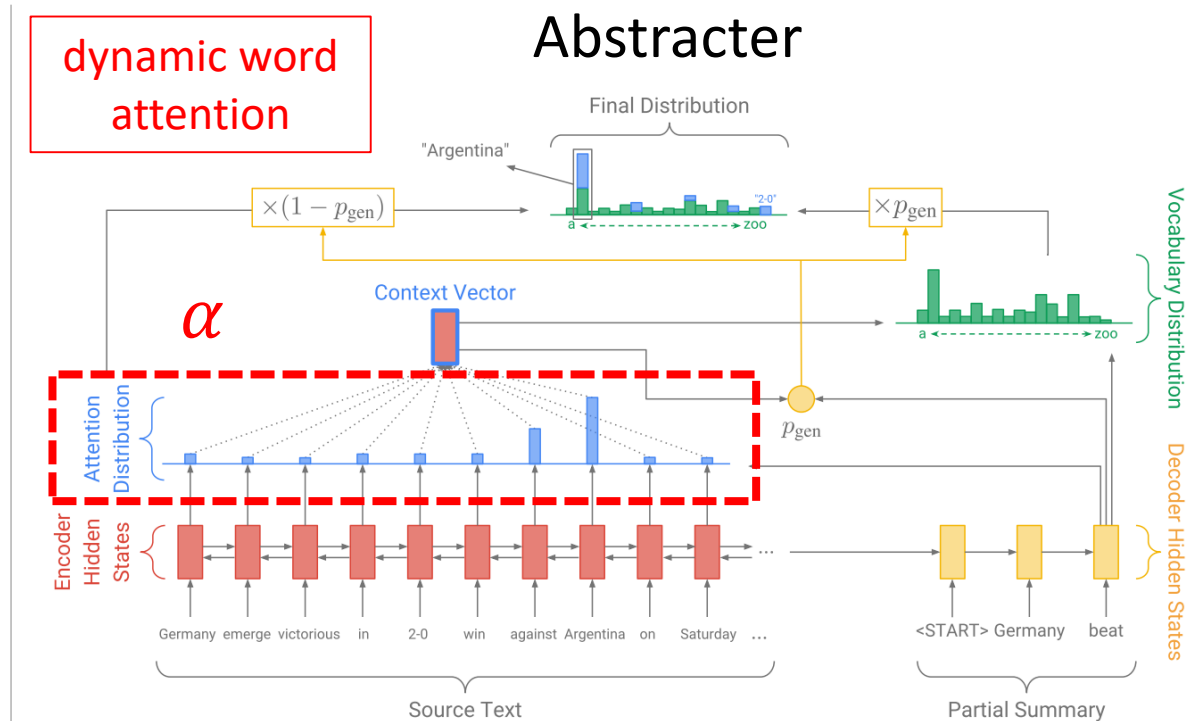
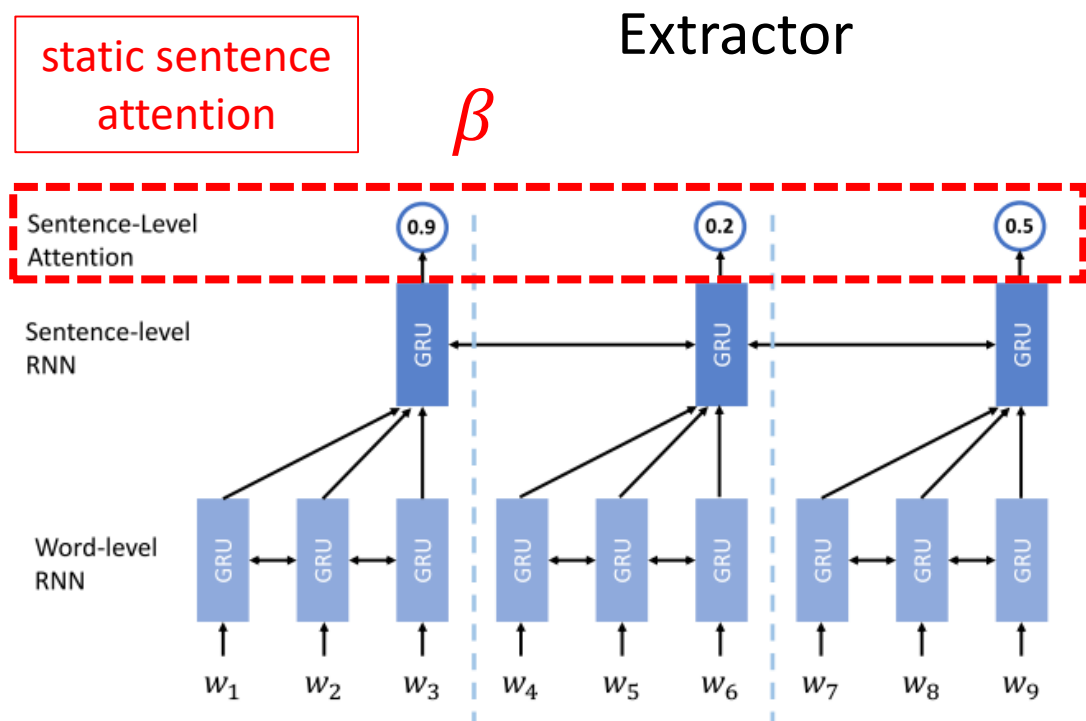


Ramesh Nallapati, Feifei Zhai, and Bowen Zhou. Summarunner: A recurrent neural network based sequence model for extractive summarization of documents. AACL 2017



Abigail See, Peter J Liu, and Christopher D Manning. Get to the point: Summarization with pointer-generator networks. ACL 2017

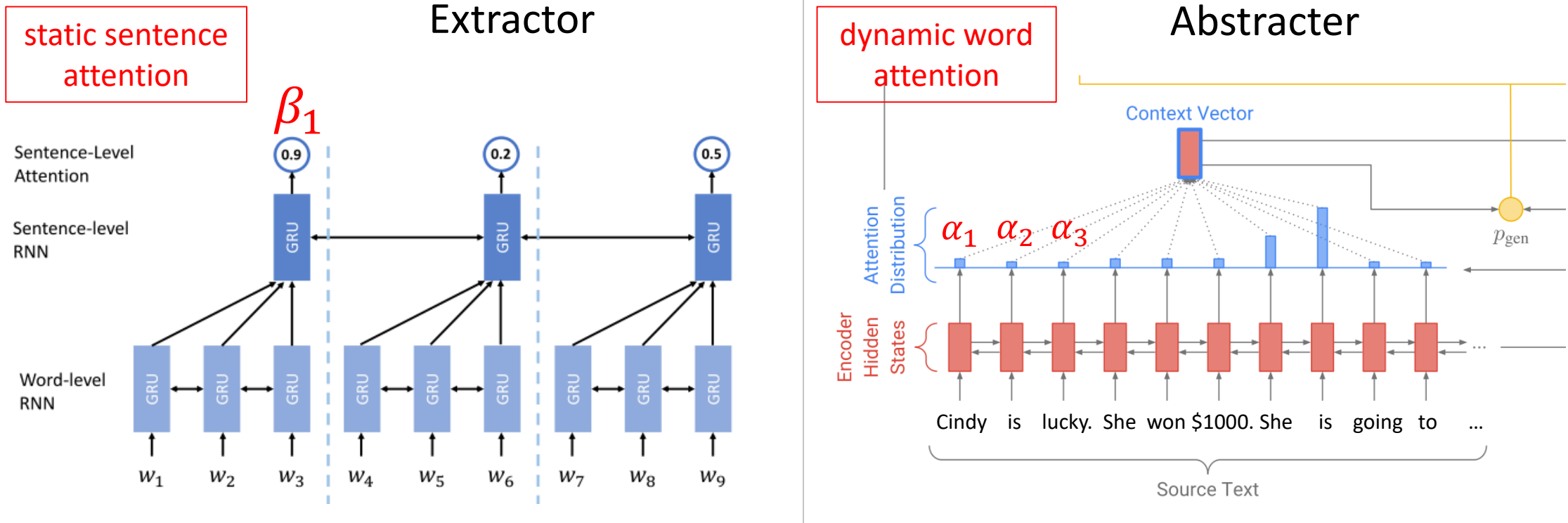
# Combined Attention



$$\hat{\alpha}_m^t = \frac{\alpha_m^t \times \beta_{n(m)}}{\sum_m \alpha_m^t \times \beta_{n(m)}}$$

$m$ : word index  
 $n$ : sentence index  
 $t$ : generated word index

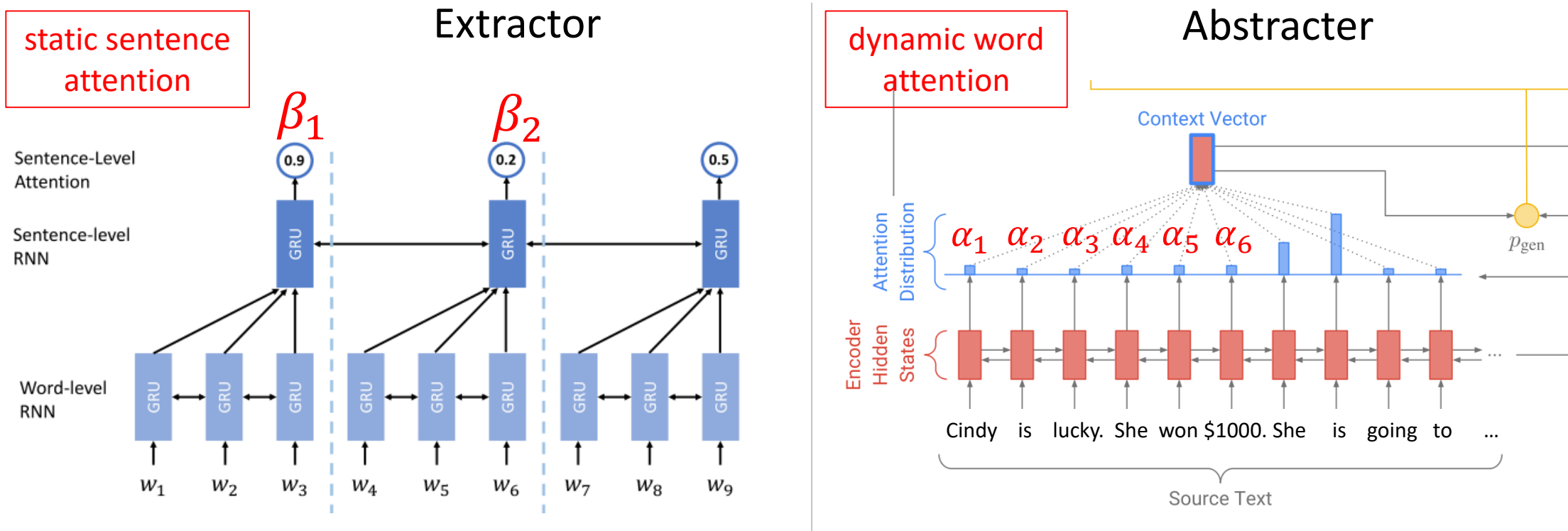
# Combined Attention



$$\hat{\alpha}_m^t = \frac{\alpha_m^t \times \beta_{n(m)}}{\sum_m \alpha_m^t \times \beta_{n(m)}}$$

$m$ : word index  
 $n$ : sentence index  
 $t$ : generated word index

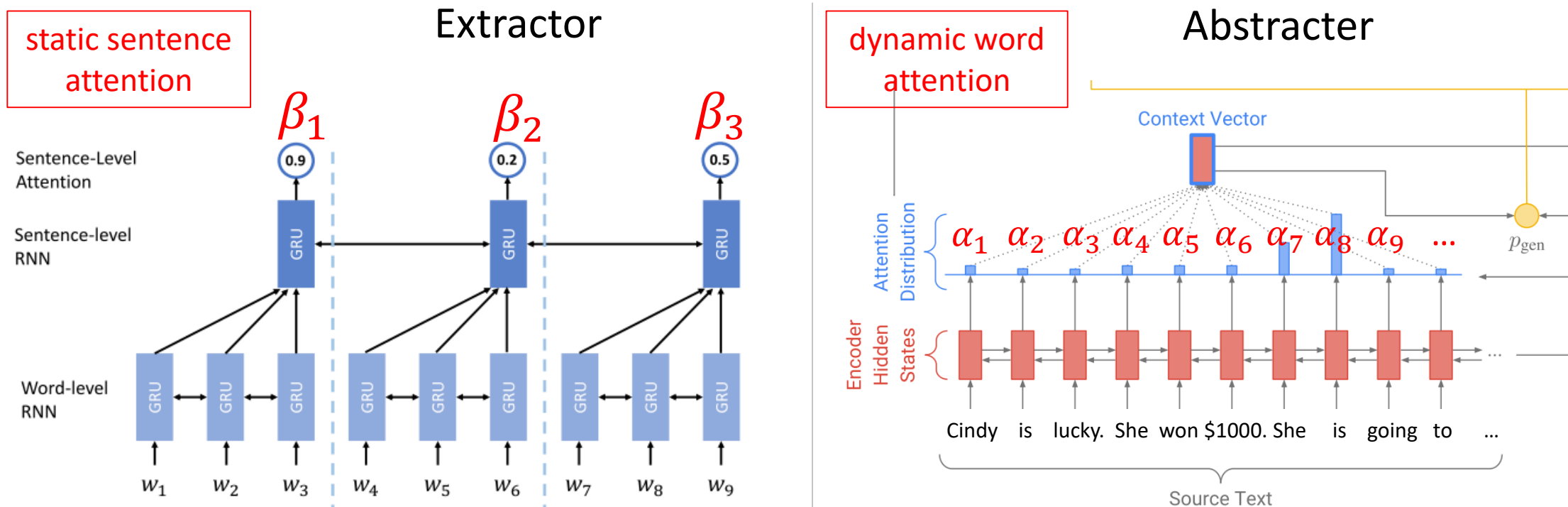
# Combined Attention



$$\hat{\alpha}_m^t = \frac{\alpha_m^t \times \beta_{n(m)}}{\sum_m \alpha_m^t \times \beta_{n(m)}}$$

$m$ : word index  
 $n$ : sentence index  
 $t$ : generated word index

# Combined Attention



$$\hat{\alpha}_m^t = \frac{\alpha_m^t \times \beta_{n(m)}}{\sum_m \alpha_m^t \times \beta_{n(m)}}$$

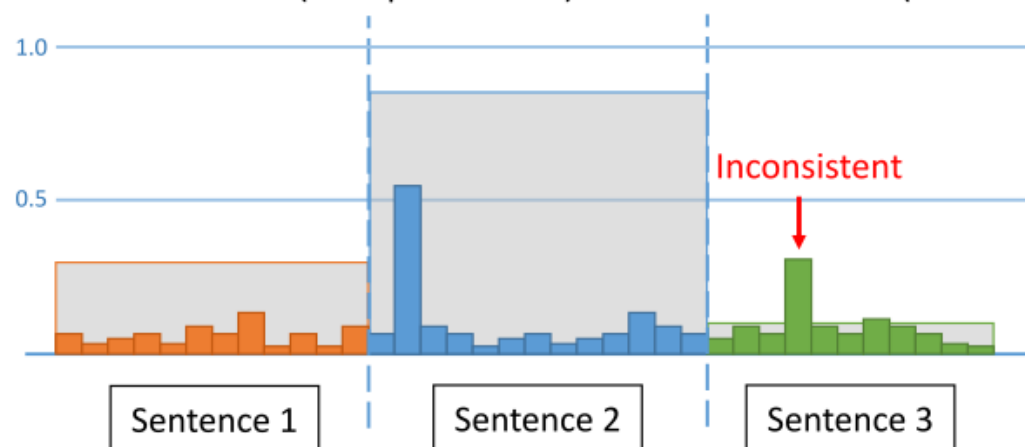
$m$ : word index  
 $n$ : sentence index  
 $t$ : generated word index

# Combined Attention

$$\hat{\alpha}_m^t = \frac{\alpha_m^t \times \beta_{n(m)}}{\sum_m \alpha_m^t \times \beta_{n(m)}}$$

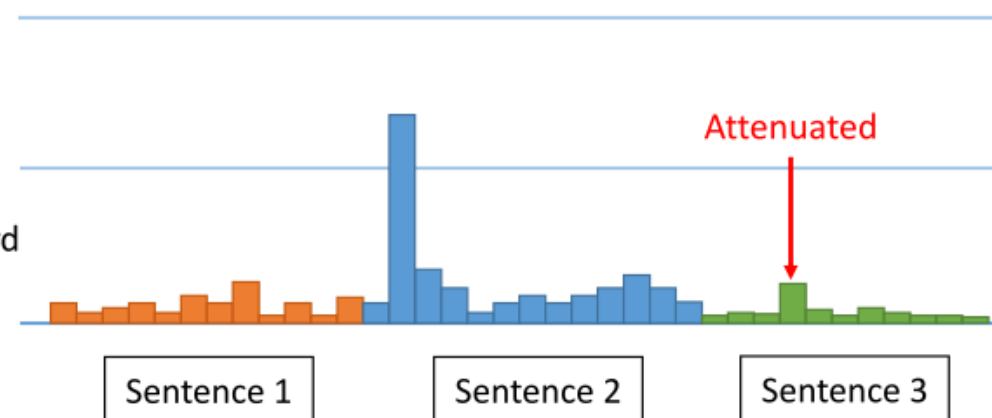
- Our unified model combines **sentence-level** and **word-level attentions** to take advantage of both extractive and abstractive summarization approaches.

Sentence Attention (transparent bars) and Word Attention (solid bars)



Multiplying and  
Renormalizing  
  
 Sentence and Word  
Attentions

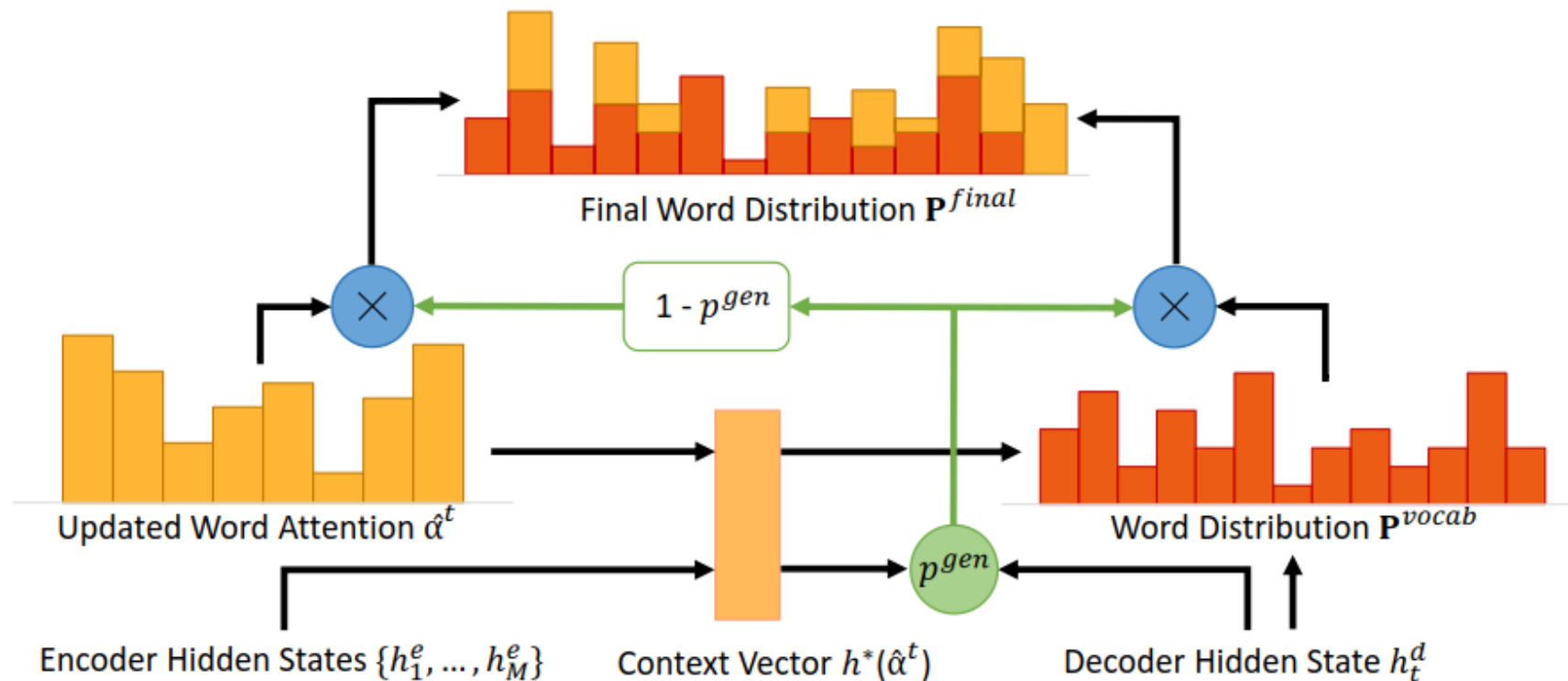
Updated Word Attention



# Combined Attention

$$\hat{\alpha}_m^t = \frac{\alpha_m^t \times \beta_{n(m)}}{\sum_m \alpha_m^t \times \beta_{n(m)}}$$

- Updated word attention is used for calculating the context vector and final word distribution



# Encourage Consistency

- We propose a **novel inconsistency loss function** to ensure our unified model to be mutually beneficial to both extractive and abstractive summarization.

$$L_{inc} = -\frac{1}{T} \sum_{t=1}^T \log\left(\frac{1}{|\mathcal{K}|} \sum_{m \in \mathcal{K}} \alpha_m^t \times \beta_{n(m)}\right)$$

multiplied attention of  
top K attended words

maximize ↑

where  $\mathcal{K}$  is the set of top K attended words

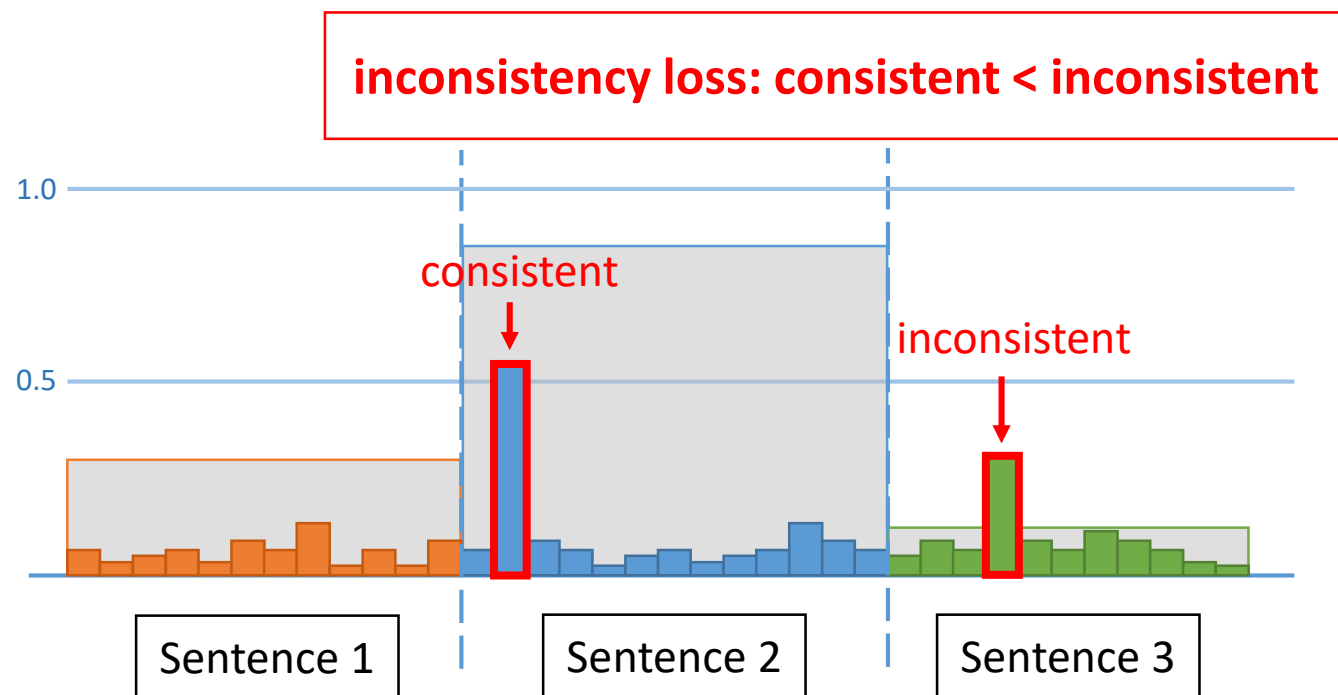


# Encourage Consistency

$$L_{inc} = -\frac{1}{T} \sum_{t=1}^T \log\left(\frac{1}{|\mathcal{K}|} \sum_{m \in \mathcal{K}} \alpha_m^t \times \beta_{n(m)}\right)$$

- encourage consistency of the **top K attended words** at each decoder time step.

K = 2



# Outline

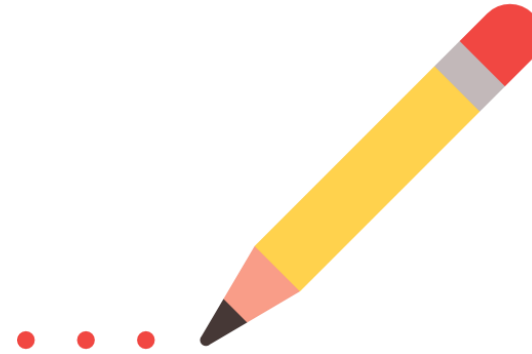
- Motivation
- Our Method
- **Training Procedures**
- Experiments and Results
- Conclusion

## Extractive Summarization



select sentences from the article

## Abstractive Summarization

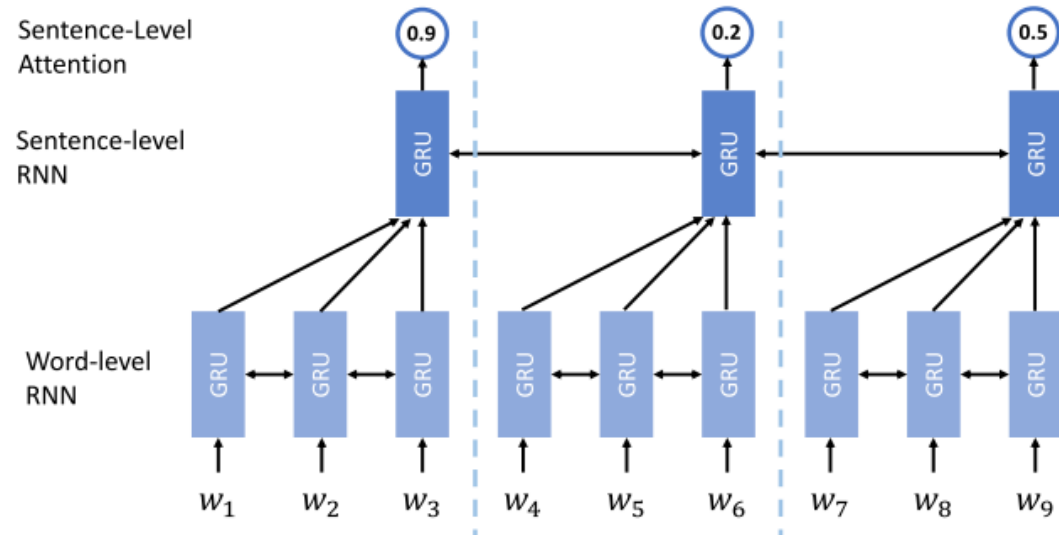


generate the summary word-by-word

- 3 types of loss functions:
  1. extractor loss
  2. abstracter loss  
+ coverage loss
  3. inconsistency loss

- 3 types of loss functions:

1. extractor loss →
2. abstracter loss + coverage loss
3. inconsistency loss

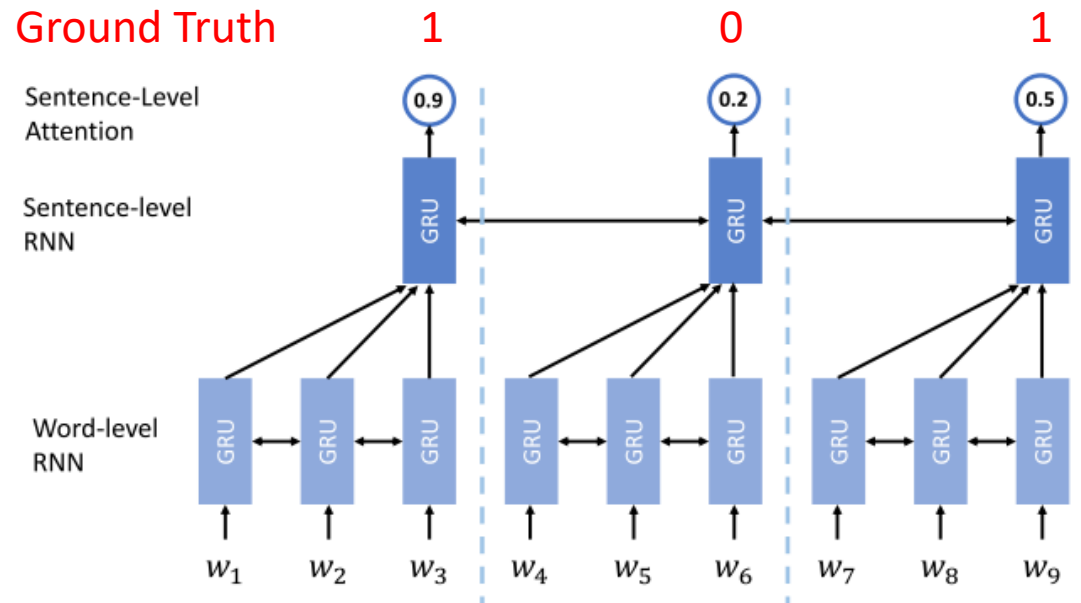


$$L_{ext} = -\frac{1}{N} \sum_{n=1}^N (g_n \log \beta_n + (1 - g_n) \log(1 - \beta_n))$$

where  $g_n \in \{0, 1\}$  is the ground-truth label for the  $n^{th}$  sentence and  $N$  is the number of sentences.

- 3 types of loss functions:

1. extractor loss →
2. abstracter loss + coverage loss
3. inconsistency loss

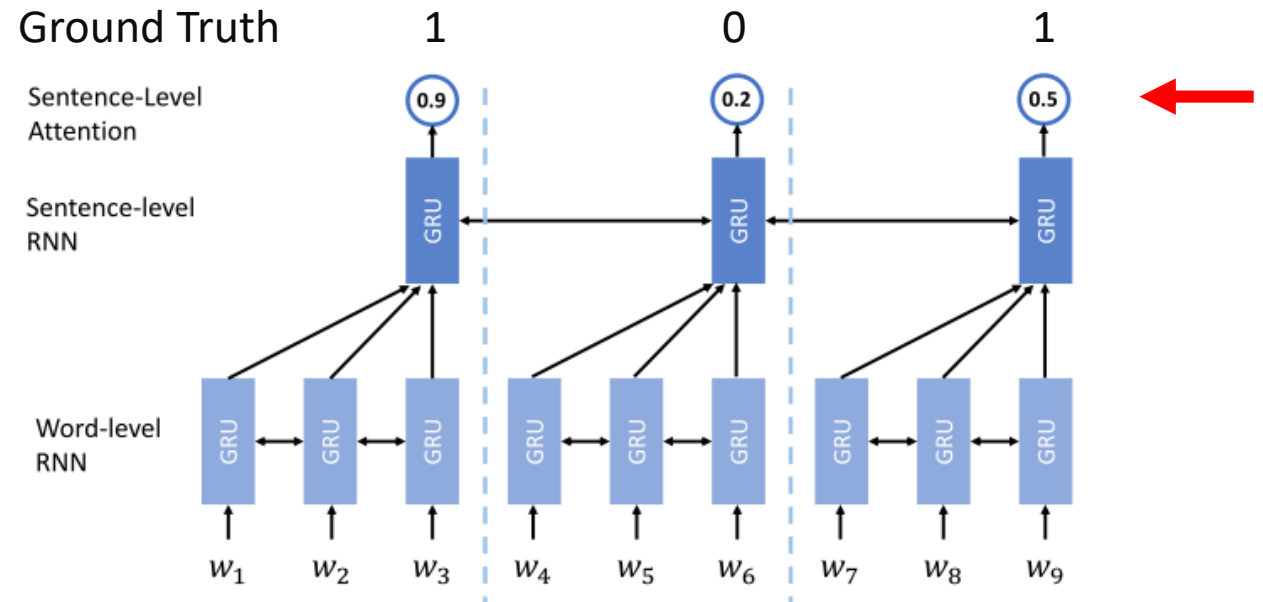


$$L_{ext} = -\frac{1}{N} \sum_{n=1}^N (g_n) \log \beta_n + (1 - (g_n)) \log(1 - \beta_n)$$

where  $g_n \in \{0, 1\}$  is the ground-truth label for the  $n^{th}$  sentence and  $N$  is the number of sentences.

- 3 types of loss functions:

1. extractor loss →
2. abstracter loss + coverage loss
3. inconsistency loss



$$L_{ext} = -\frac{1}{N} \sum_{n=1}^N (g_n \log \beta_n + (1 - g_n) \log(1 - \beta_n))$$

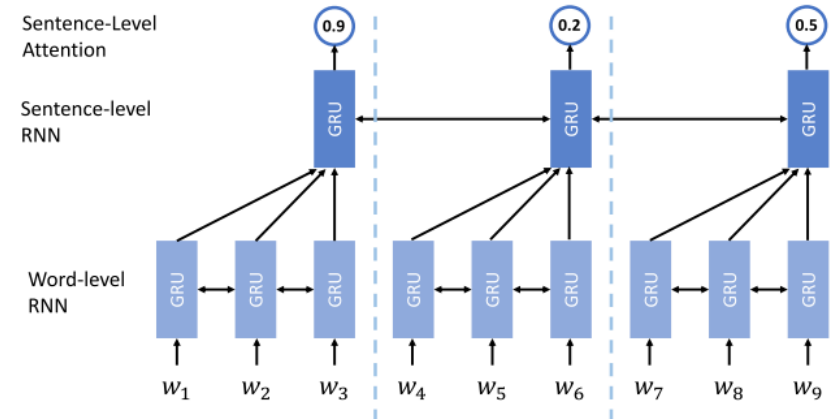
where  $g_n \in \{0, 1\}$  is the ground-truth label for the  $n^{th}$  sentence and  $N$  is the number of sentences.

## Extractor Target

- To extract sentences with **high informativity**: the extracted sentences should contain information that is needed to generate an abstractive summary as much as possible.

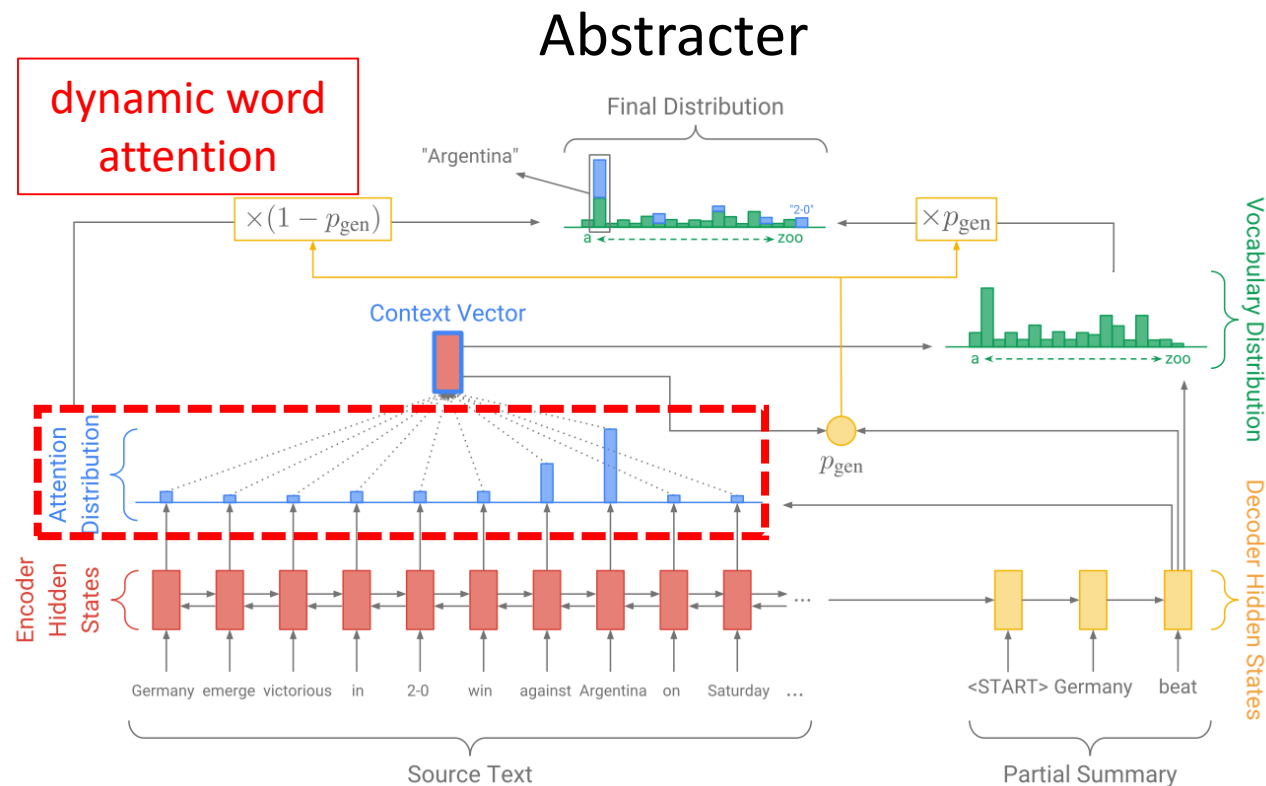
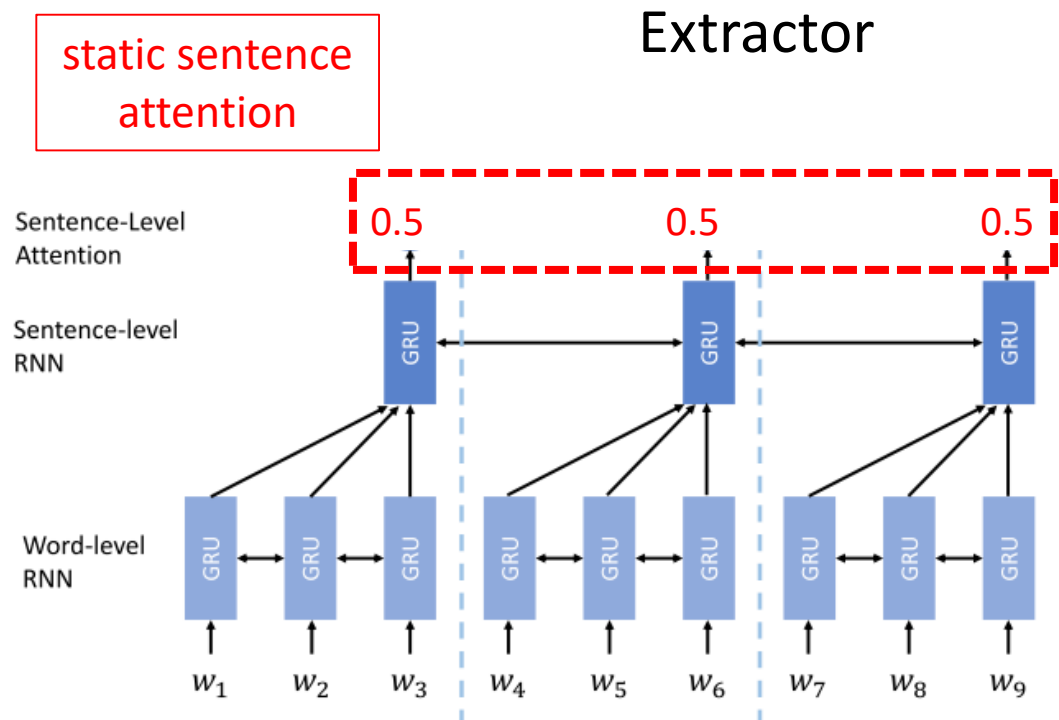
- Ground-truth labels:

- Measure the informativity of each sentence in the article by computing the **ROUGE-L recall score** between the sentence and the reference abstractive summary.
- Select the sentence **in the order of high to low informativity** and add one sentence at a time if the new sentence can increase the informativity of all the selected sentences.





## Combined Attention



$$\hat{\alpha}_m^t = \frac{\alpha_m^t \times \beta_{n(m)}}{\sum_m \alpha_m^t \times \beta_{n(m)}}.$$

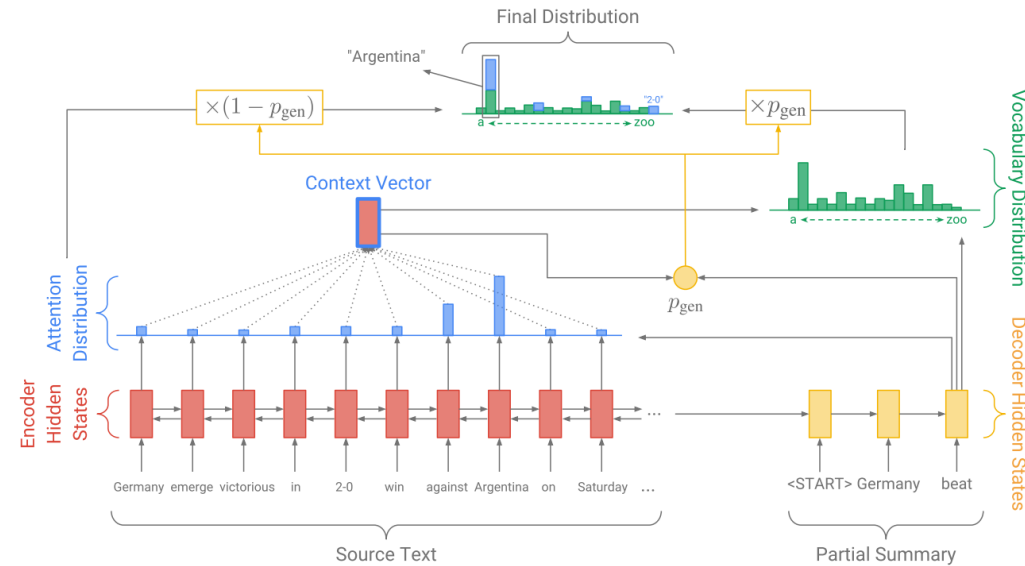
$m$ : word index  
 $n$ : sentence index  
 $t$ : generated word index

- 3 types of loss functions:

1. extractor loss

2. abstracter loss + coverage loss

3. inconsistency loss



$$L_{abs} = -\frac{1}{T} \sum_{t=1}^T \log P_{\hat{y}^t}^{final}$$

$$L_{cov} = \frac{1}{T} \sum_{t=1}^T \sum_{m=1}^M \min(\hat{\alpha}_m^t, c_m^t)$$

$$c^t = \sum_{t'=1}^{t-1} \hat{\alpha}^{t'}$$

- 3 types of loss functions:

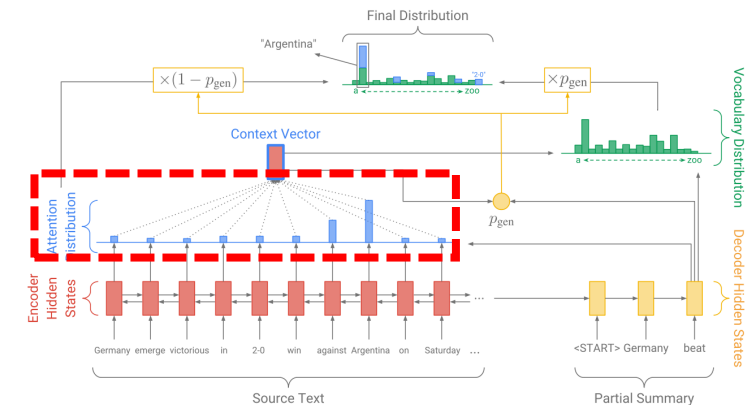
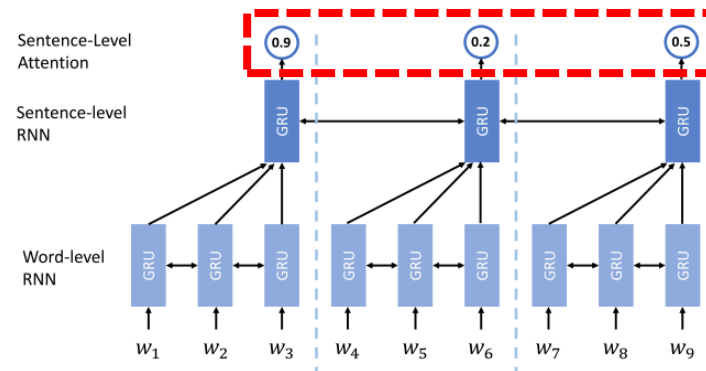
1. extractor loss

2. abstracter loss + coverage loss

3. inconsistency loss ➔

$$L_{inc} = -\frac{1}{T} \sum_{t=1}^T \log\left(\frac{1}{|\mathcal{K}|} \sum_{m \in \mathcal{K}} \alpha_m^t \times \beta_{n(m)}\right)$$

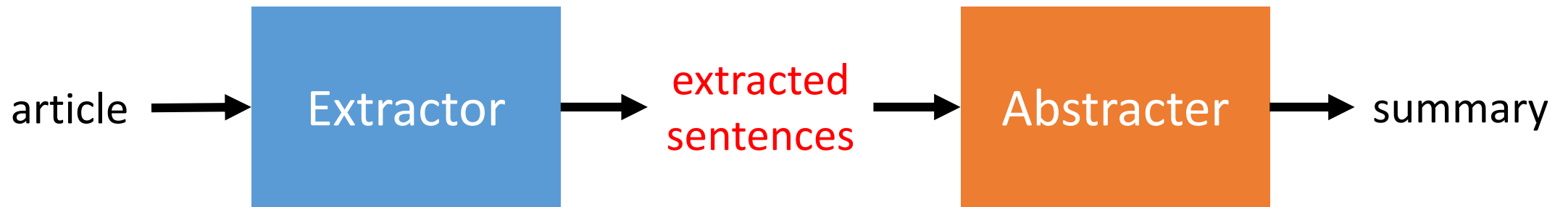
where  $\mathcal{K}$  is the set of top K attended words



1. Two-stages training
2. End-to-end training without inconsistency loss
3. End-to-end training with inconsistency loss

## 1. Two-stages training

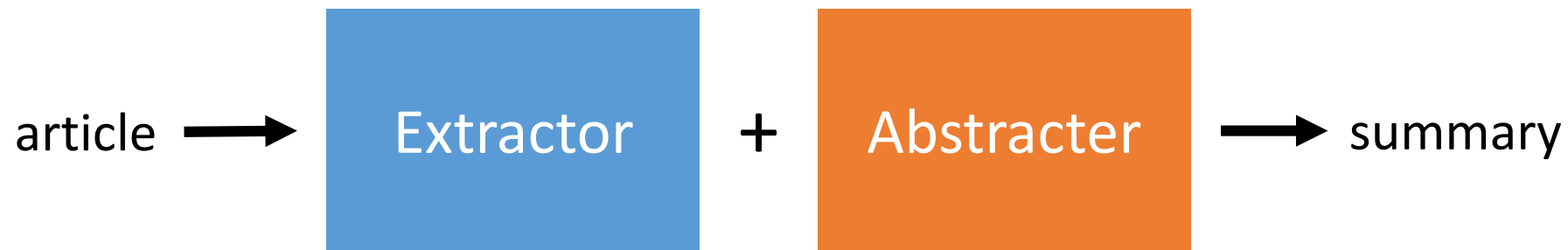
- The extractor is used as a classifier to select sentences with high informativity and output only those sentences. = **Hard attention** on the original article.
- simply combine the extractor and abstracter **by feeding the extracted sentences to the abstracter.**



## 2. End-to-end training **without inconsistency loss**

- the sentence-level attention is **soft attention** and will be combined with the word-level attention
- minimize extractor loss and abstracter loss

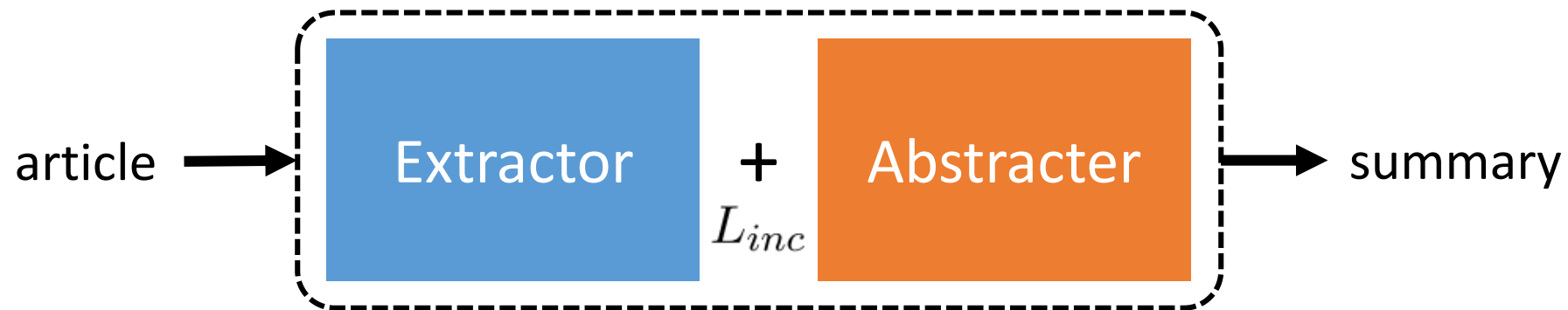
$$L_{e2e} = \lambda L_{ext} + L_{abs} + L_{cov}$$



### 3. End-to-end training **with inconsistency loss**

- the sentence-level attention is **soft attention** and will be combined with the word-level attention
- minimize extractor loss, abstracter loss and **inconsistency loss**:

$$L_{e2e} = \lambda L_{ext} + L_{abs} + L_{cov} + L_{inc}$$



# Outline

- Motivation
- Our Method
- Training Procedures
- **Experiments and Results**
- Conclusion



# Dataset – CNN/DailyMail Dataset

Article  $\approx$  766 words  
Summary  $\approx$  53 words

	Train	Validation	Test
Article-summary pairs	287,113	13,368	11,490



45 CONGRESS SECURITY THE NINE TRUMPAMERICA STATE

## STORY HIGHLIGHTS

Bannon was expected to return at 2 p.m.  
ET Thursday

The postponement follows an exchange  
of terse letters by the House panel and  
Bannon's attorney

**Washington (CNN)** — The House Intelligence Committee has agreed to give former White House chief strategist Steve Bannon until the end of the month to return to the panel and finish his closed-door interview, according to multiple sources with knowledge of the matter.

Bannon's attorney, in a letter to the committee shared with CNN, said the committee asked that his legal team (...)

# Dataset – CNN/DailyMail Dataset

Article  $\approx$  766 words  
Summary  $\approx$  53 words

	Train	Validation	Test
Article-summary pairs	287,113	13,368	11,490

**CNN** politics 45 CONGRESS SECURITY THE NINE TRUMPAMERICA STATE

**Highlight**  
50 words

**STORY HIGHLIGHTS**

Bannon was expected to return at 2 p.m. ET Thursday

The postponement follows an exchange of terse letters by the House panel and Bannon's attorney

**Article**  
700 words

**Washington (CNN)** — The House Intelligence Committee has agreed to give former White House chief strategist Steve Bannon until the end of the month to return to the panel and finish his closed-door interview, according to multiple sources with knowledge of the matter.

Bannon's attorney, in a letter to the committee shared with CNN, said the committee asked that his legal team (...)

# Results – Abstractive Summarization

Method	ROUGE-1	ROUGE-2	ROUGE-L
HierAttn (Nallapati et al., 2016b)*	32.75	12.21	29.01
DeepRL (Paulus et al., 2017)*	39.87	15.82	36.90
→ pointer-generator (See et al., 2017)	39.53	17.28	36.38
→ GAN (Liu et al., 2017)	39.92	17.65	36.71
two-stage (ours)	39.97	17.43	36.34
end2end w/o inconsistency loss (ours)	40.19	17.67	36.68
end2end w/ inconsistency loss (ours)	<b>40.68</b>	<b>17.97</b>	<b>37.13</b>
lead-3 (See et al., 2017)	40.34	17.70	36.57

Table 2: ROUGE F-1 scores of the generated abstractive summaries on the CNN/Daily Mail test set. Our two-stages model outperforms pointer-generator model on ROUGE-1 and ROUGE-2. In addition, our model trained end-to-end with inconsistency loss exceeds the lead-3 baseline. All our ROUGE scores have a 95% confidence interval with at most  $\pm 0.24$ . ‘\*’ indicates the model is trained and evaluated on the anonymized dataset and thus is not strictly comparable with ours.

# Results – Abstractive Summarization

Method	ROUGE-1	ROUGE-2	ROUGE-L
HierAttn (Nallapati et al., 2016b)*	32.75	12.21	29.01
DeepRL (Paulus et al., 2017)*	39.87	15.82	36.90
pointer-generator (See et al., 2017)	39.53	17.28	36.38
GAN (Liu et al., 2017)	39.92	17.65	36.71
two-stage (ours)	39.97	17.43	36.34
end2end w/o inconsistency loss (ours)	40.19	17.67	36.68
end2end w/ inconsistency loss (ours)	<b>40.68</b>	<b>17.97</b>	<b>37.13</b>
lead-3 (See et al., 2017)	40.34	17.70	36.57

Table 2: ROUGE F-1 scores of the generated abstractive summaries on the CNN/Daily Mail test set. Our two-stages model outperforms pointer-generator model on ROUGE-1 and ROUGE-2. In addition, our model trained end-to-end with inconsistency loss exceeds the lead-3 baseline. All our ROUGE scores have a 95% confidence interval with at most  $\pm 0.24$ . ‘\*’ indicates the model is trained and evaluated on the anonymized dataset and thus is not strictly comparable with ours.

# Results – Abstractive Summarization

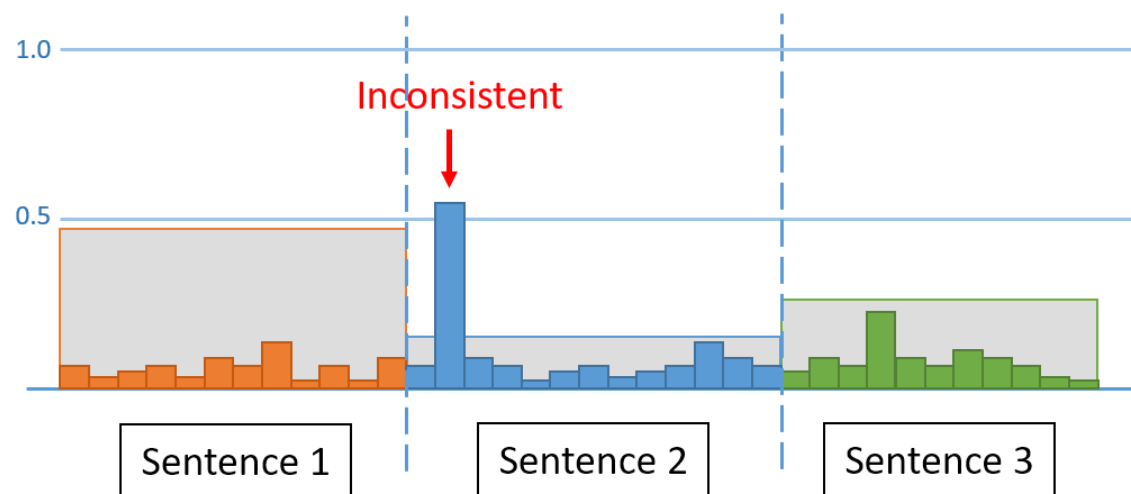
Method	ROUGE-1	ROUGE-2	ROUGE-L
HierAttn (Nallapati et al., 2016b)*	32.75	12.21	29.01
DeepRL (Paulus et al., 2017)*	39.87	15.82	36.90
pointer-generator (See et al., 2017)	39.53	17.28	36.38
GAN (Liu et al., 2017)	39.92	17.65	36.71
two-stage (ours)	39.97	17.43	36.34
end2end w/o inconsistency loss (ours)	40.19	17.67	36.68
end2end w/ inconsistency loss (ours)	<b>40.68</b>	<b>17.97</b>	<b>37.13</b>
lead-3 (See et al., 2017)	40.34	17.70	36.57

Table 2: ROUGE F-1 scores of the generated abstractive summaries on the CNN/Daily Mail test set. Our two-stages model outperforms pointer-generator model on ROUGE-1 and ROUGE-2. In addition, our model trained end-to-end with inconsistency loss exceeds the lead-3 baseline. All our ROUGE scores have a 95% confidence interval with at most  $\pm 0.24$ . ‘\*’ indicates the model is trained and evaluated on the anonymized dataset and thus is not strictly comparable with ours.

# Results – Inconsistency Rate $R_{inc}$

**inconsistency step  $t_{inc}$ :**

$$\beta_{n(\text{argmax}(\alpha^t))} < \text{mean}(\beta)$$



sentence attention and word attention in time step  $t$

**inconsistency rate:**

$$R_{inc} = \frac{\text{Count}(t_{inc})}{T}$$

where  $T$  is the length of the summary.

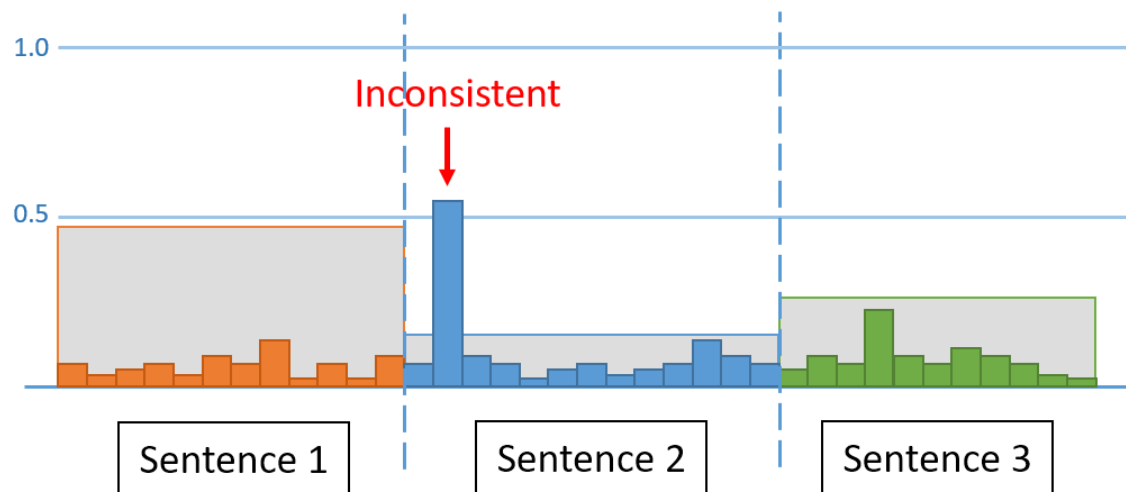
Method	avg. $R_{inc}$
w/o incon. loss	0.198
w/ incon. loss	0.042

Table 3: Inconsistency rate of our end-to-end trained model with and without inconsistency loss.

# Results – Inconsistency Rate $R_{inc}$

**inconsistency step  $t_{inc}$ :**

$$\beta_{n(\text{argmax}(\alpha^t))} < \text{mean}(\beta)$$



sentence attention and word attention in time step  $t$

**inconsistency rate:**

$$R_{inc} = \frac{\text{Count}(t_{inc})}{T}$$

where  $T$  is the length of the summary.

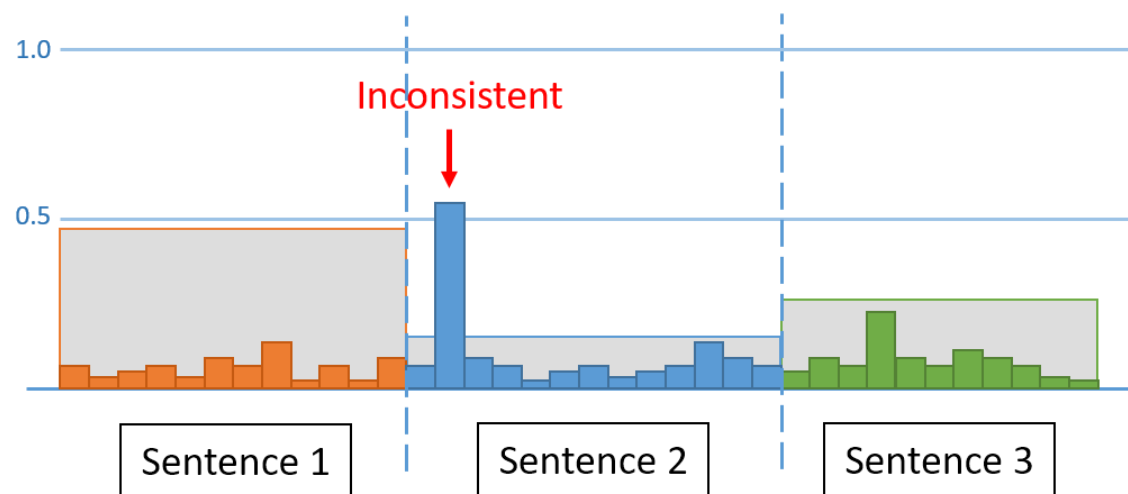
Method	avg. $R_{inc}$
w/o incon. loss	0.198
w/ incon. loss	0.042

Table 3: Inconsistency rate of our end-to-end trained model with and without inconsistency loss.

# Results – Inconsistency Rate $R_{inc}$

inconsistency step  $t_{inc}$ :

$$\beta_{n(\text{argmax}(\alpha^t))} < \text{mean}(\beta)$$



sentence attention and word attention in time step  $t$

inconsistency rate:

$$R_{inc} = \frac{\text{Count}(t_{inc})}{T}$$

where  $T$  is the length of the summary.

Method	avg. $R_{inc}$
w/o incon. loss	0.198
w/ incon. loss	0.042

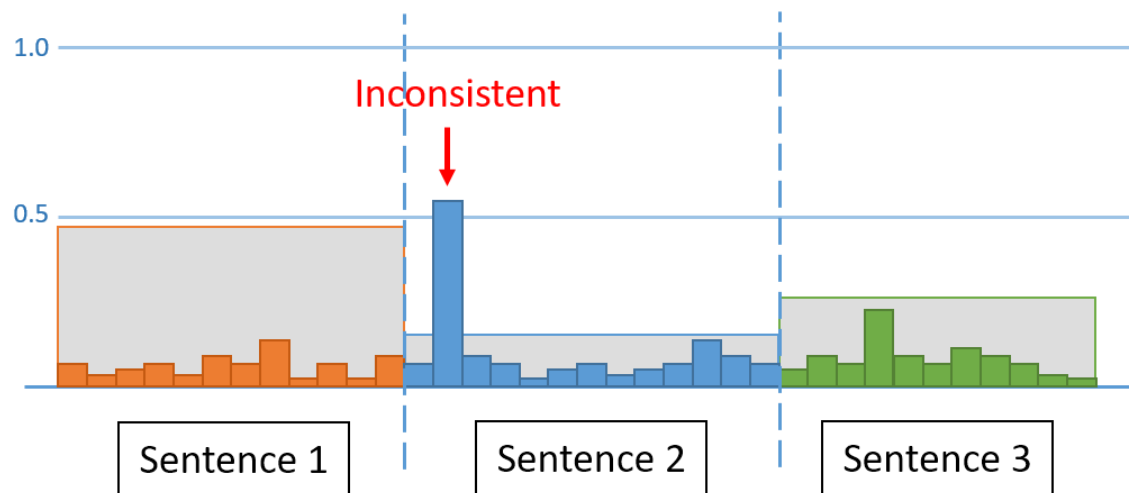
Table 3: Inconsistency rate of our end-to-end trained model with and without inconsistency loss.



# Results – Inconsistency Rate $R_{inc}$

**inconsistency step  $t_{inc}$ :**

$$\beta_{n(\text{argmax}(\alpha^t))} < \text{mean}(\beta)$$



sentence attention and word attention in time step  $t$

**inconsistency rate:**

$$R_{inc} = \frac{\text{Count}(t_{inc})}{T}$$

where  $T$  is the length of the summary.

Method	avg. $R_{inc}$
w/o incon. loss	0.198
w/ incon. loss	0.042

Table 3: Inconsistency rate of our end-to-end trained model with and without inconsistency loss.

## Results – Human Evaluation on MTurk

- **Informativity:**  
how well does the summary capture the important parts of the article?
- **Conciseness:**  
is the summary clear enough to explain everything without being redundant?
- **Readability:**  
how well-written (fluent and grammatical) the summary is?

Summary 1						Summary 2						Summary 3					
Henrik Larsson was forced to play his 42-year-old kit man in goal. The emergency stopper kept a clean sheet as Helsingborg drew 0-0. Helsingborg manager Henrik Larsson said : ' it was a scenario that I never could have prepared myself for.						Daniel Andersson, Helsingborg's 42-year-old kit man, kept a clean sheet. The emergency stopper played in season opener against Kalmar. Henrik Larsson's first-choice goalkeepers were both out injured. The former goalkeeper earned one cap for Sweden back in 2001.						Henrik Larsson was forced to play Daniel Andersson with goalkeepers Par Hansson and Matt Pyzdrowski out injured. The emergency stopper kept a clean sheet as Helsingborg drew 0-0 against Kalmar in the Isvenskan season opener. Helsingborg manager Henrik Larsson was forced to play 42-year-old kit man Daniel Andersson in goal.					
	1	2	3	4	5		1	2	3	4	5		1	2	3	4	5
Informativity	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Informativity	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Informativity	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Conciseness	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Conciseness	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Conciseness	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Readability	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Readability	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Readability	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Summary 4						Summary 5						Summary 6					
Helsingborg manager Henrik Larsson was forced to play his 42-year-old kit man in goal on saturday. The former Celtic and Barcelona striker had no option but to play Daniel Andersson with goalkeepers Par Hansson and Matt Pyzdrowski out injured. Andersson made 130 appearances for the club between 2004 and 2009 and also spent a season with Scottish club Hibernian.						Henrik Larsson was forced to play his 42-year-old kit man in goal. The Celtic and Barcelona striker had no option but to play Daniel Andersson. Helsingborg manager Henrik Larsson was to play with goalkeepers Par Hansson and Matt Pyzdrowski out injured.						A new survey found seven in 10 people end up injured while doing DIY. Poll of 2,000 people found 68% say they or their partner have ended up hurt. Two in five said they injured their back and one in five had cut themselves.					
	1	2	3	4	5		1	2	3	4	5		1	2	3	4	5
Informativity	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Informativity	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Informativity	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Conciseness	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Conciseness	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Conciseness	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Readability	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Readability	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Readability	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

trap

# Results – Human Evaluation

- **Informativity**: how well does the summary capture the important parts of the article?
- **Conciseness**: is the summary clear enough to explain everything without being redundant?
- **Readability**: how well-written (fluent and grammatical) the summary is?

Method	informativity	conciseness	readability
DeepRL (Paulus et al., 2017)	3.23	2.97	2.85
pointer-generator (See et al., 2017)	3.18	3.36	3.47
GAN (Liu et al., 2017)	3.22	3.52	3.51
Ours	<b>3.58</b>	3.40	<b>3.70</b>
reference	3.43	<b>3.61</b>	3.62

Table 3: Comparing human evaluation results with state-of-the-art methods.

# Outline

- Motivation
- Our Method
- Training Procedures
- Experiments and Results
- **Conclusion**

# Conclusion

- We propose a **unified model** combining the strength of extractive and abstractive summarization.
- A novel **inconsistency loss function** is introduced to penalize the inconsistency between two levels of attentions. The inconsistency loss enables extractive and abstractive summarization to be mutually beneficial.
- By end-to-end training of our model, we achieve the **best ROUGE scores** while being the **most informative and readable** summarization on the CNN/Daily Mail dataset in a solid human evaluation.

# Acknowledgements



Min Sun  
Wen-Ting Tsu  
Chieh-Kai Lin  
Ming-Ying Lee

Kerui Min  
Jing Tang

# Q & A



## **Project page**

- **Code**
- **Test output**
- **Supplementary material**

[https://hsuwanting.github.io/unified\\_summ/](https://hsuwanting.github.io/unified_summ/)