The art of copying in the generation model

Xingwu Lu

East China Normal University

Department of Computer Science and Technology

2018/5/20

Reference

- ACL16-Incorporating Copying Mechanism in Sequence-to-Sequence Learning
- ACL16-Pointing the Unknown Words
- ACL17-Get To The Point: Summarization with Pointer-Generator Networks
- AAAI18-Sequential Copying Networks

Outline

- Background
- Motivation
- Model
- Comparison
- Experiments

Background

Automatic Text Summarization

Automatic summarization is the process of shortening a text document with software, in order to create a summary with the major points of the original document.

By the Input Granularity

- sentence summarization
- single-document summarization
- multi-document summarization

By the Methods

- Extractive methods: assemble summaries exclusively from passages,

You might think of these approaches as like a highlighter.

- Abstractive methods: generate novel words and phrases not featured in the source text, these approaches are like a pen.

Extractive methods

Shortcomings:

- rigid and symbolic

Advantages:

- Ensures baseline levels of grammaticality and accuracy
- use exactly the same words, summary sentence is accurate

Abstractive methods

Shortcomings:

- reproduce factual details inaccurately
- inability to deal with out-of-vocabulary

Advantages:

- generate novel words and phrases, can perform paraphrasing
- can be more diverse
- as a human-written abstract usually does

Motivation

Generation

sequence-to-sequence learning (Seq2Seq):

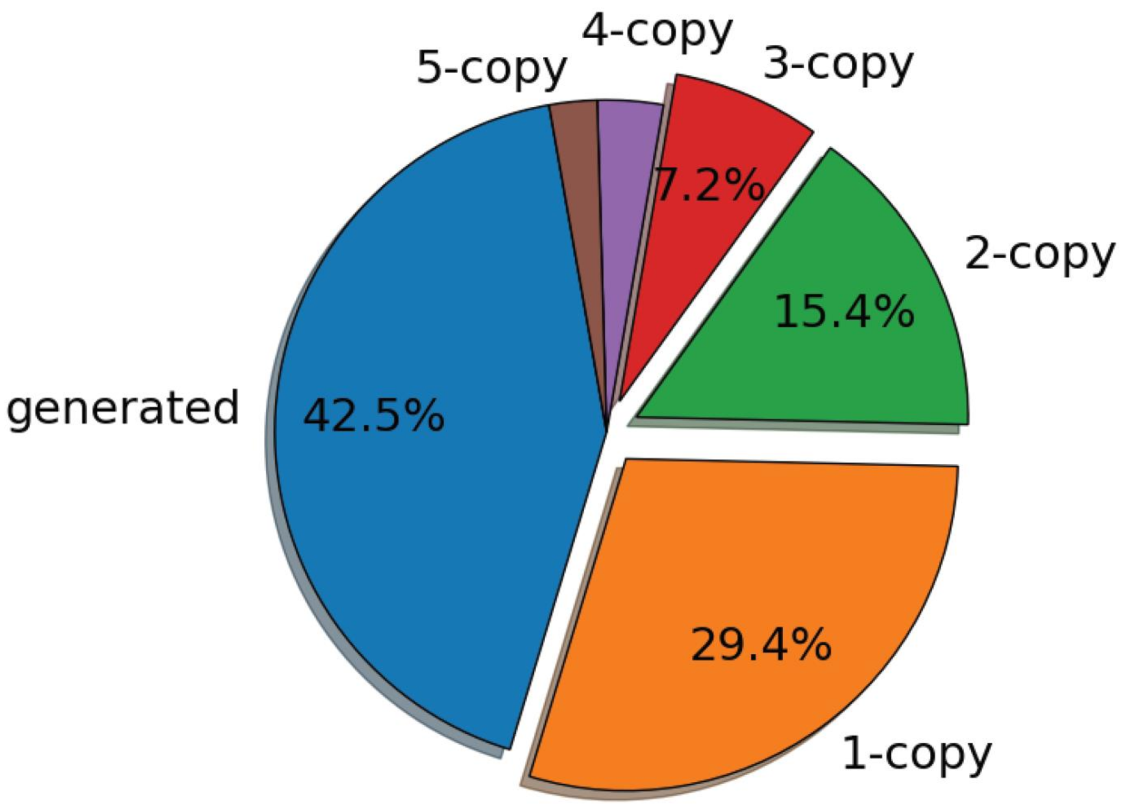
- Machine Translation

  encoder information = decoder information
- Text Summarization

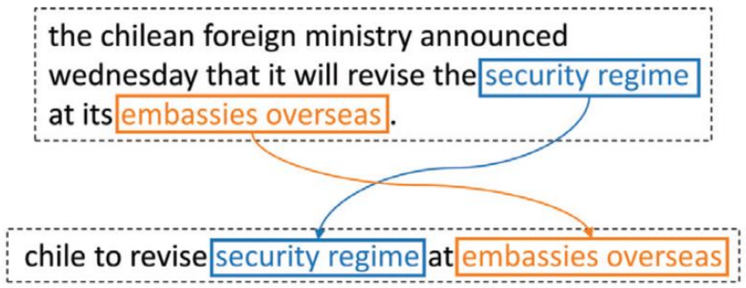
  encoder information > decoder information
- Dialogue Systems

  encoder information ? decoder information

Percentage of generated and copied words



For example, in the Gigaword dataset of abstractive sentence summarization task, about 57.7% words are copied from the input. Moreover, the copied words in multi-word span account for 28.1%, which is also very common.



Solution

combining both extraction (pointing) and abstraction (generating)

- ACL16-Incorporating Copying Mechanism in Sequence-to-Sequence Learning
- ACL17-Get To The Point: Summarization with Pointer-Generator Networks
- AAAI18-Sequential Copying Networks

Models

Background: Neural Models for Seq2Seq

Seq2Seq is essentially an encoder-decoder model

- Encoder

  transform the input sequence to a certain representation

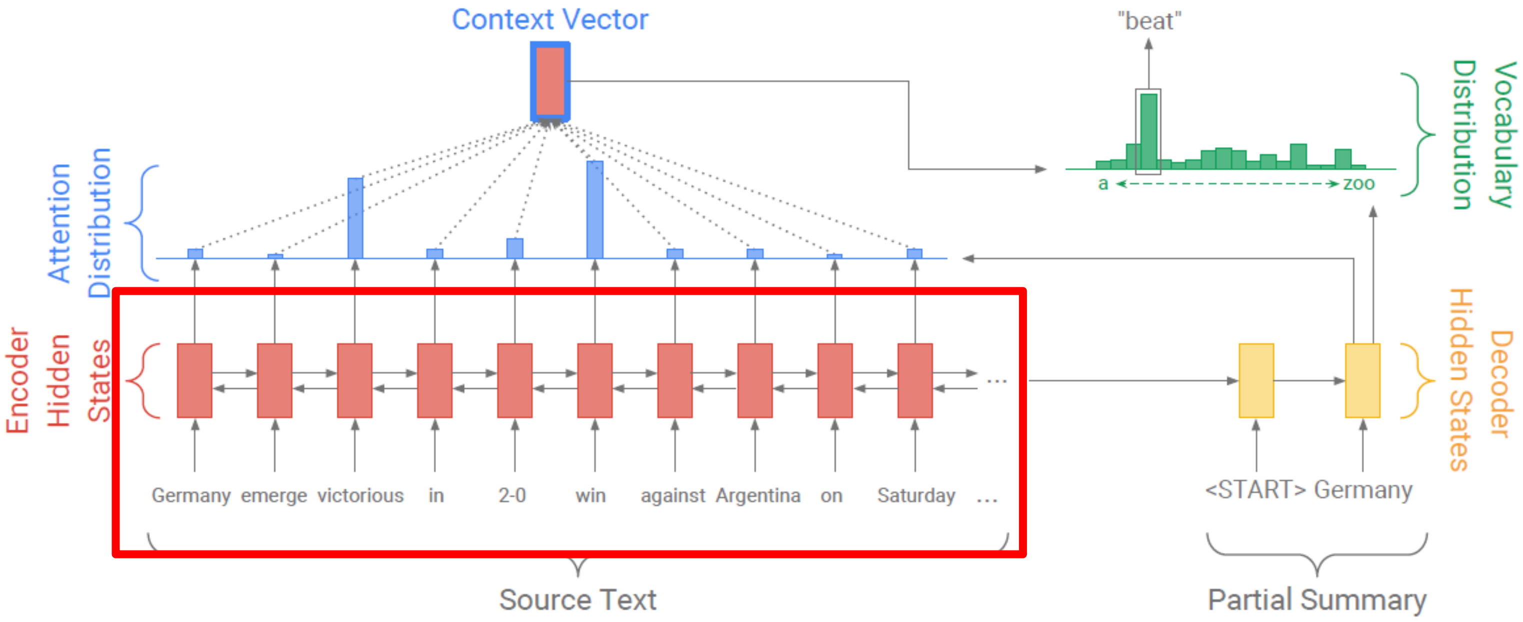
- Decoder

  transform the representation into the output sequence

- Attention mechanism

  revisits the input sequence and dynamically fetches the relevant piece of information

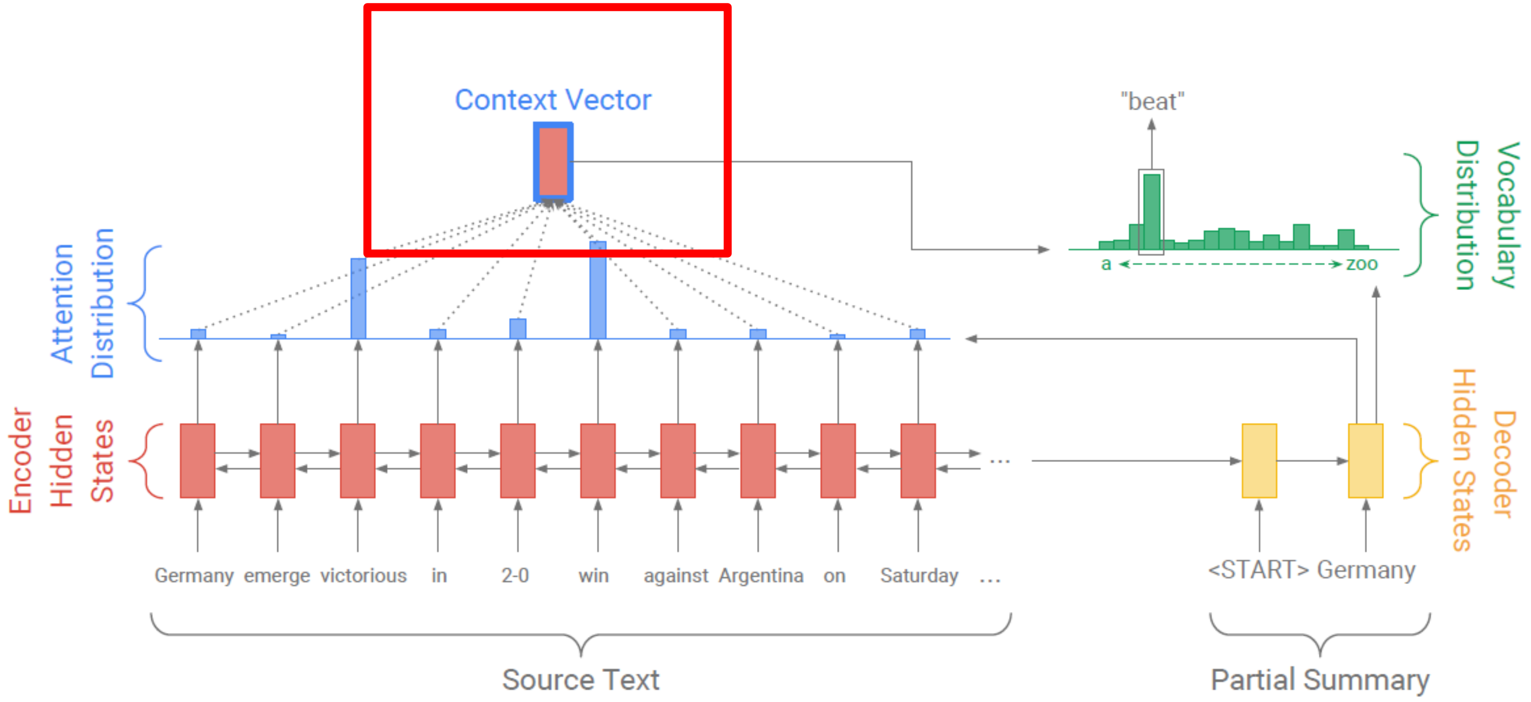
Sequence-to-sequence attentional model



$$z_i = \sigma(\mathbf{W}_z[x_i, h_{i-1}])$$
$$r_i = \sigma(\mathbf{W}_r[x_i, h_{i-1}])$$
$$\widetilde{h}_i = \tanh(\mathbf{W}_h[x_i, r_i \odot h_{i-1}])$$
$$h_i = (1 - z_i) \odot h_{i-1} + z_i \odot \widetilde{h}_i$$

$$\overrightarrow{h}_i = \mathrm{GRU}(x_i, \overrightarrow{h}_{i-1})$$
$$\overleftarrow{h}_i = \mathrm{GRU}(x_i, \overleftarrow{h}_{i+1})$$

Sequence-to-sequence attentional model

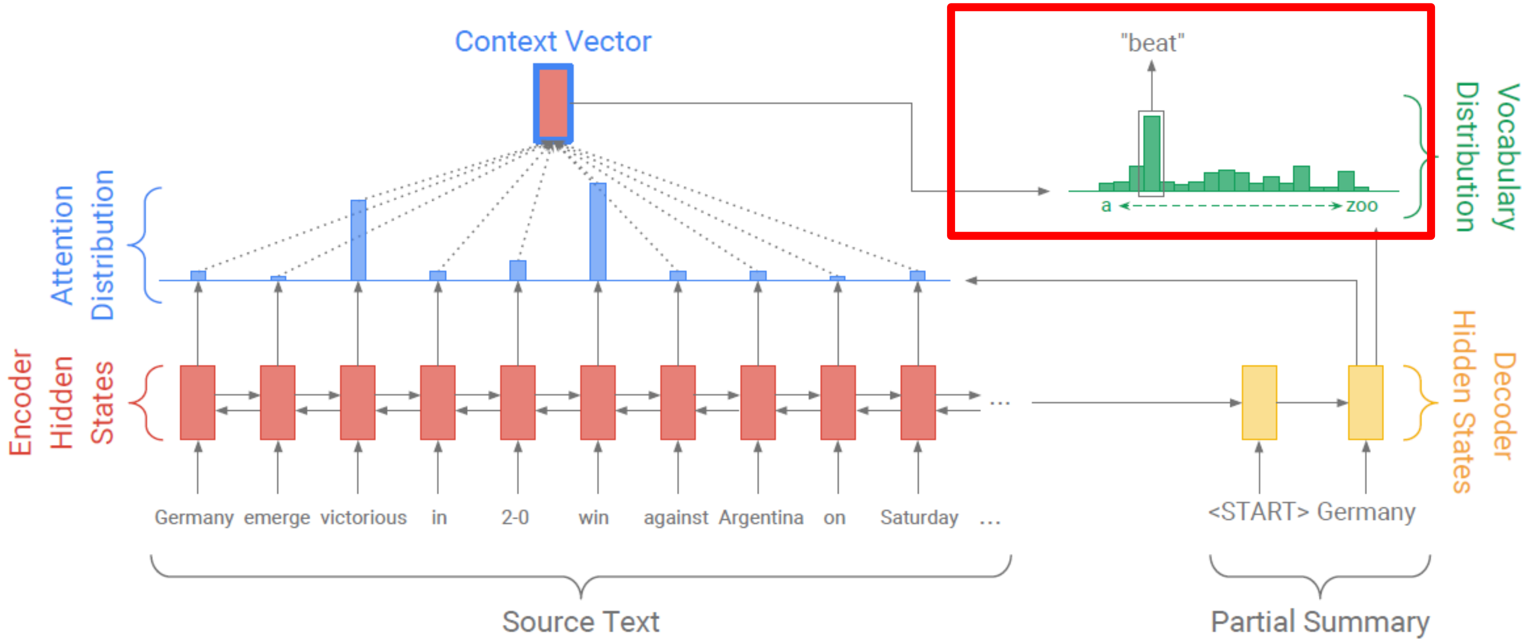


$$e_i^t = v^T tanh(W_h h_i + W_s s_t + b_{attn})$$

$$a^t = softmax(e^t)$$

$$h_t^* = \sum a_i^t h_i$$

Sequence-to-sequence attentional model

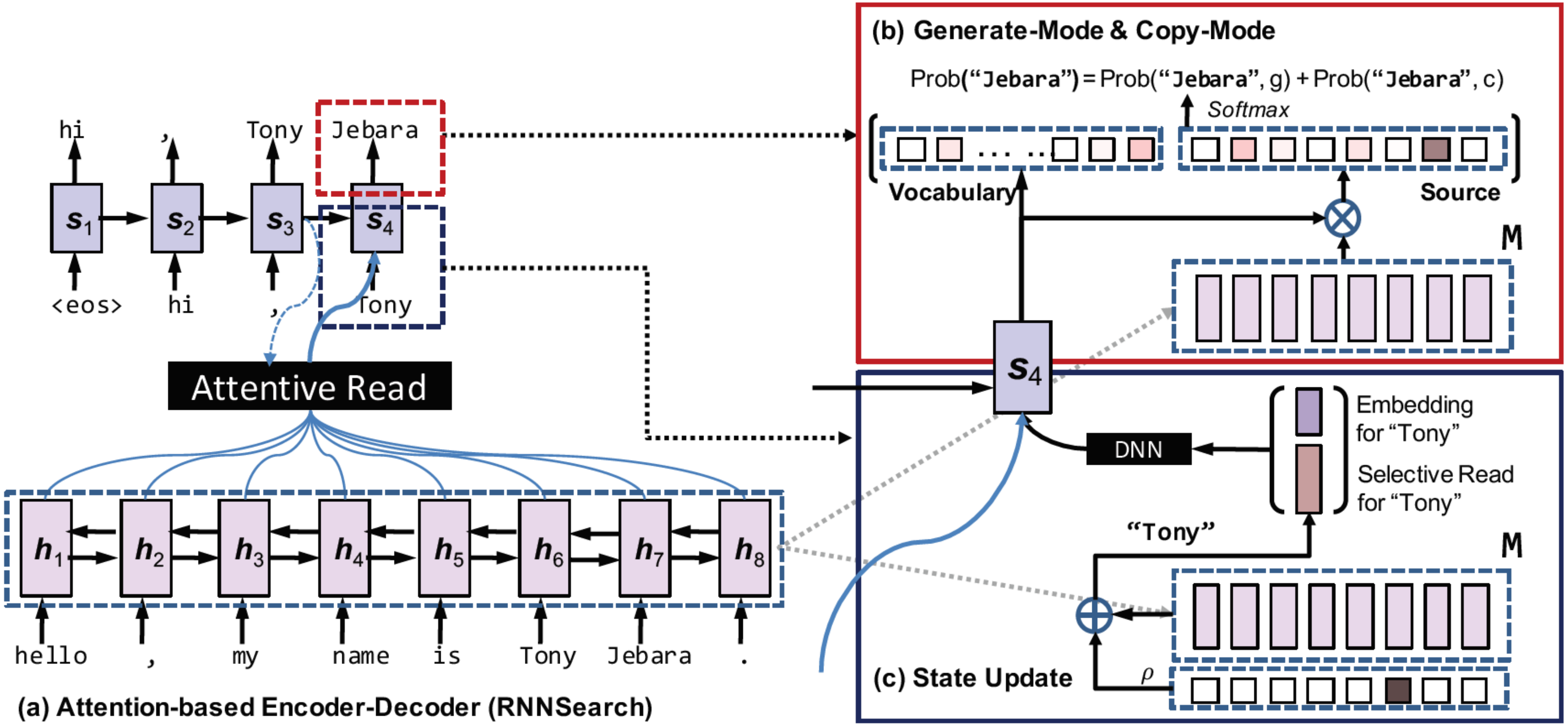


$$P_{vocab} = softmax(V'(V[h_t^*, s_t] + b) + b')$$

$$P(w) = P_{vocab}(w)$$

$$loss_t = -logP(w^*)$$

COPYNET

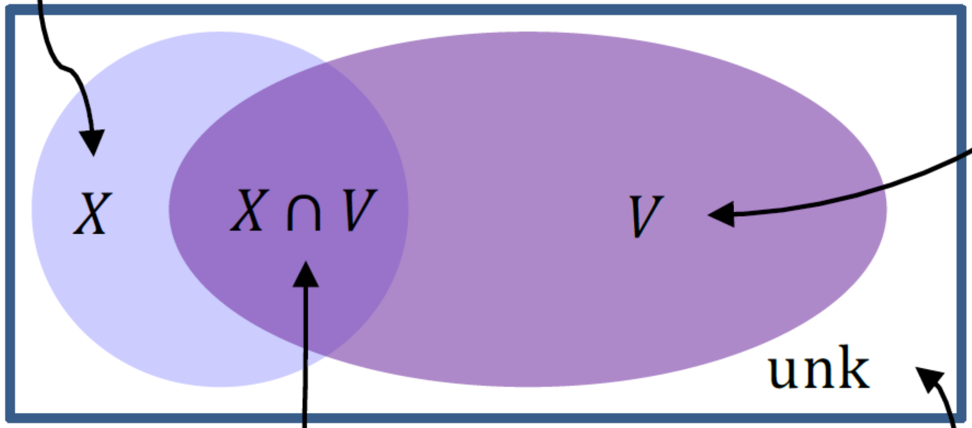


ACL16-Incorporating Copying Mechanism in Sequence-to-Sequence Learning

How to Copy?

$$\frac{1}{Z}\underline{\sum_{x_j}}\exp[\psi_c(x_j)] \mid x_j = y_t$$

$$\frac{1}{Z}\exp[\psi_g(v_i)] \mid v_i = y_t$$



*Z is the normalization term.

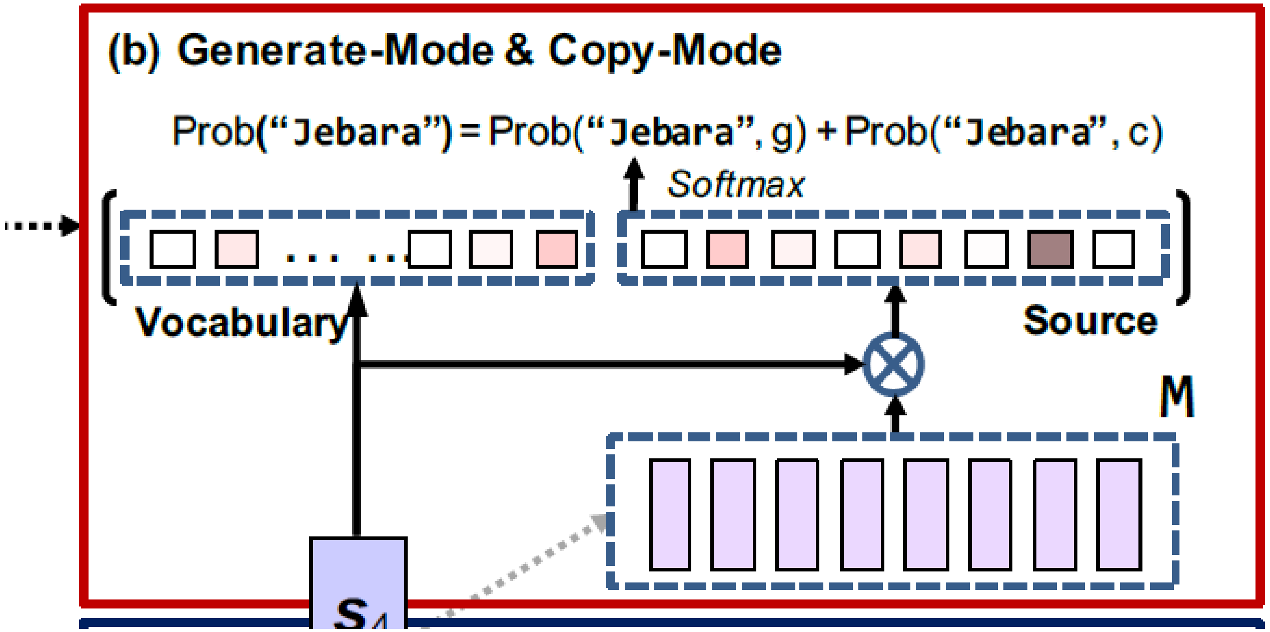
$$\frac{1}{Z}\left(\sum_{x_j}\exp[\psi_c(x_j)] + \exp[\psi_g(v_i)]\right) \mid x_j = y_t, v_i = y_t$$

$$\frac{1}{Z}\exp[\psi_g(\text{unk})]$$

Prediction with Copying and Generation

shared softmax function

$$p(y_t|s_t, y_{t-1}, c_t, M) = p(y_t, g|s_t, y_{t-1}, c_t, M) + p(y_t, c|s_t, y_{t-1}, c_t, M)$$



Prediction with Copying and Generation

shared softmax function

$$p(y_t|s_t, y_{t-1}, c_t, M) = p(y_t, g|s_t, y_{t-1}, c_t, M) + p(y_t, c|s_t, y_{t-1}, c_t, M)$$

$$p(y_t, \mathrm{g}|\cdot) = \begin{cases} \frac{1}{Z} e^{\boxed{\psi_g(y_t)}}, & y_t \in \mathcal{V} \\ 0, & y_t \in \mathcal{X} \cap \bar{V} \\ \frac{1}{Z} e^{\psi_g(\mathrm{UNK})} & y_t \notin \mathcal{V} \cup \mathcal{X} \end{cases} \quad (5)$$

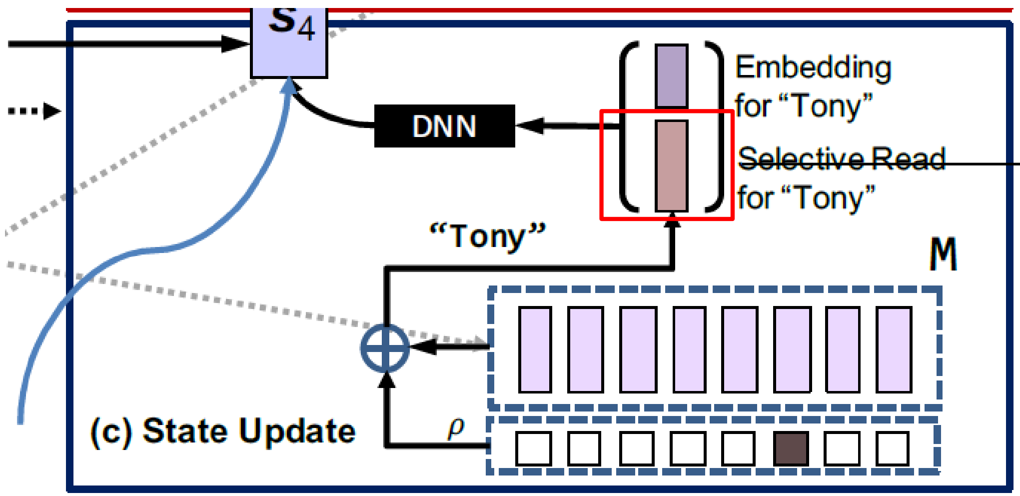
$$\psi_g(y_t = v_i) = \mathbf{v}_i^\top \mathbf{W}_o \mathbf{s}_t, \quad v_i \in \mathcal{V} \cup \mathrm{UNK}$$

$$p(y_t, \mathrm{c}|\cdot) = \begin{cases} \frac{1}{Z} \sum_{j:x_j=y_t} e^{\boxed{\psi_c(x_j)}}, & y_t \in \mathcal{X} \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

$$\psi_c(y_t = x_j) = \sigma\left(\mathbf{h}_j^\top \mathbf{W}_c\right) \mathbf{s}_t, \quad x_j \in \mathcal{X} \quad (8)$$

$$Z = \sum_{v \in \mathcal{V} \cup \{\mathrm{UNK}\}} e^{\psi_g(v)} + \sum_{x \in X} e^{\psi_c(x)}.$$

When to Copy?



Add selective read to update State S: $[\mathbf{e}(y_{t-1}); \boxed{\zeta(y_{t-1})}]^\top$

Hybrid Addressing of M

$$\zeta(y_{t-1}) = \sum_{\tau=1}^{T_S} \rho_{t\tau}\mathbf{h}_\tau$$

$$\rho_{t\tau} = \begin{cases} \frac{1}{K} p(x_\tau, \mathsf{c}|\mathbf{s}_{t-1}, \mathbf{M}), & x_\tau = y_{t-1} \\ 0 & \text{otherwise} \end{cases} \qquad (9)$$

K: $\sum_{\tau':x_{\tau'}=y_{t-1}} p(x_{\tau'}, c|\mathbf{s}_{t-1}, \mathbf{M})$

When to Copy?

hidden states in M:

- semantics information

  the attentive read is driven more by the semantics

- location information

  the selective read is often guided by the location information.

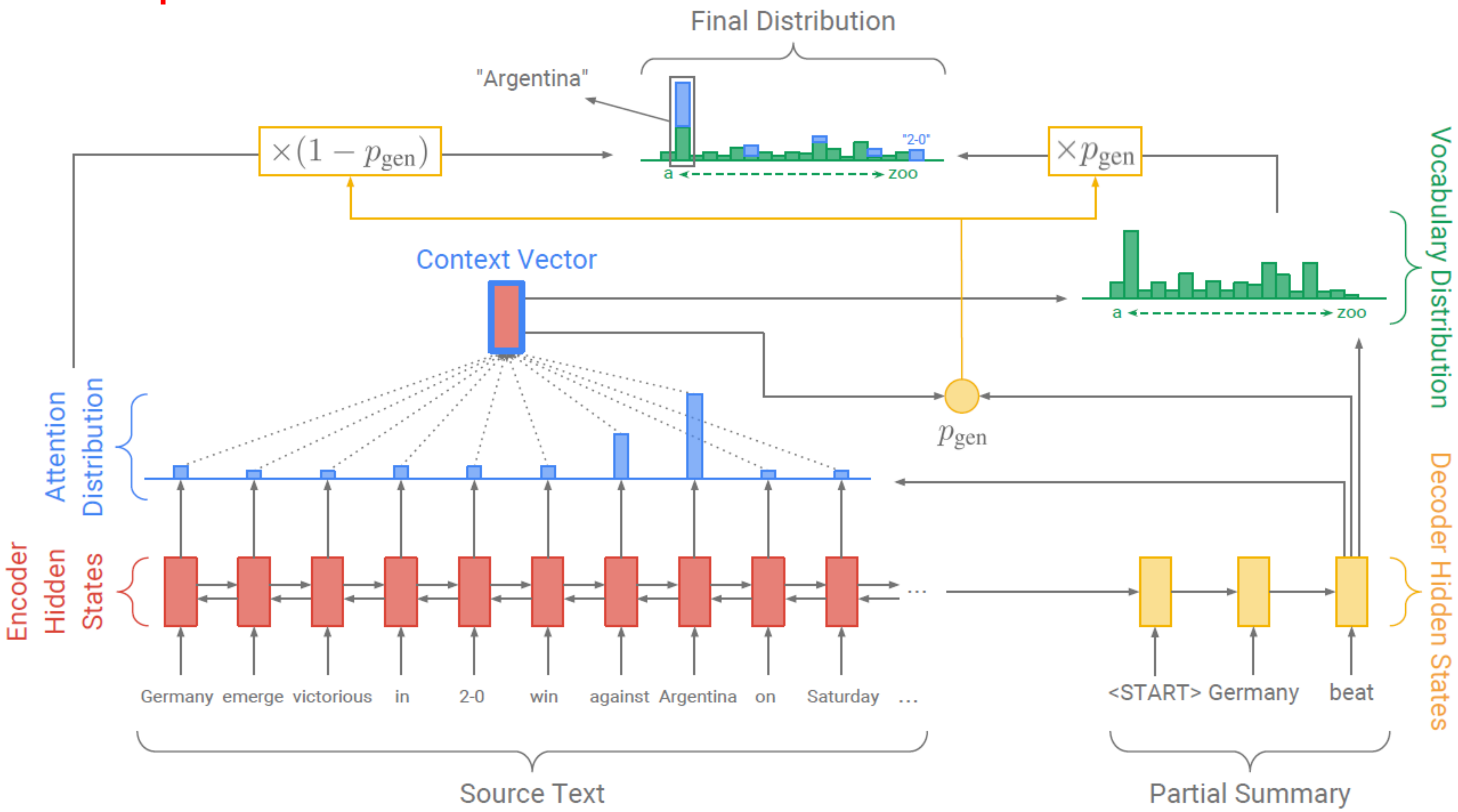
provides a simple way of “moving one step to the right” on X

$$\zeta(y_{t-1}) \longrightarrow update\ s_t \longrightarrow predict\ y_t \longrightarrow sel.\, read\ \zeta(y_t)$$

Location-based Addressing

Pointer-generator network

Simplify this operation



ACL17-Get To The Point: Summarization with Pointer-Generator Networks

Pointer-generator Network

$$p_{gen} = \sigma(w_{h^*}^T h_t^* + w_{s_t}^T s_t + w_{x_t}^T x_t + b_{gen})$$

When to Point

$$P(w) = p_{gen} P_{vocab}(w) + (1 - p_{gen}) \sum_{i:w_i = w} a_i^t$$

How to Point

generation probability: $p_{gen} \in [0, 1]$

Comparison

CopyNet model V.S. Pointer-generator network

1. explicit switch probability p_{gen} V.S shared softmax function
2. copy distribution: recycle the attention distribution V.S two separate distributions

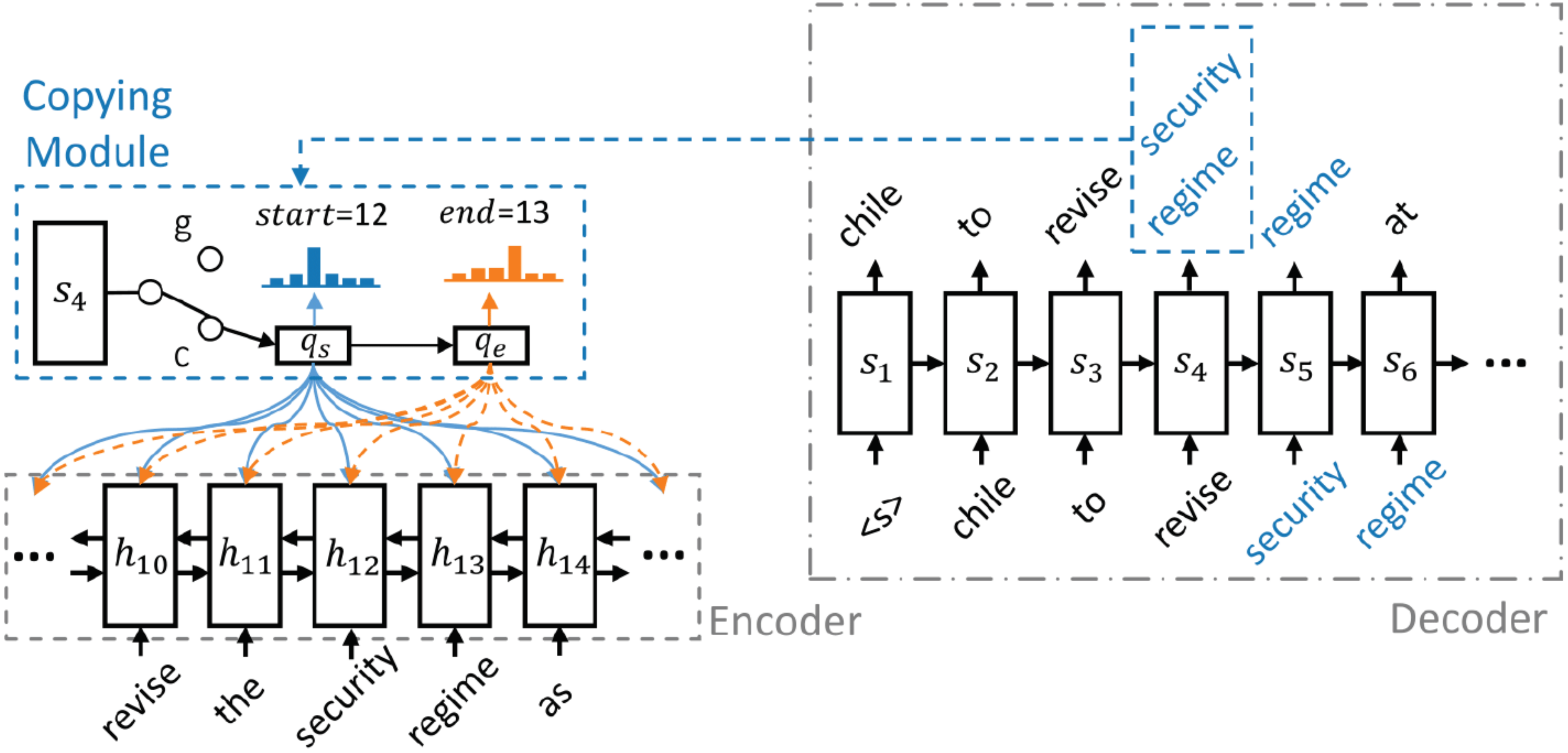
Further thinking

Sequential Copying Networks (SeqCopyNet):

- Problem: this “single word copy” paradigm may introduce errors due to these separate decisions.

- Solution: We copy them once and for all

SeqCopyNet


Copying Module
s_4
g
c
q_s
q_e
$start$=12
end=13
h_{10}
h_{11}
h_{12}
h_{13}
h_{14}
revise
the
security
regime
as
Encoder
s_1
s_2
s_3
s_4
s_5
s_6
<s>
chile
to
revise
security
regime
chile
to
revise
security
regime
regime
at
Decoder

Model Overview

three main components:

- an RNN based sentence encoder
- an attention-equipped decoder
- newly designed copying module:
    - the copy switch gate network:

      used to make decisions of whether to copy according to the current decoding states
    - the pointer network

      used to extract a span from the input sentence
    - copy state transducer

      update the copying state so that the pointer network can predict the end position

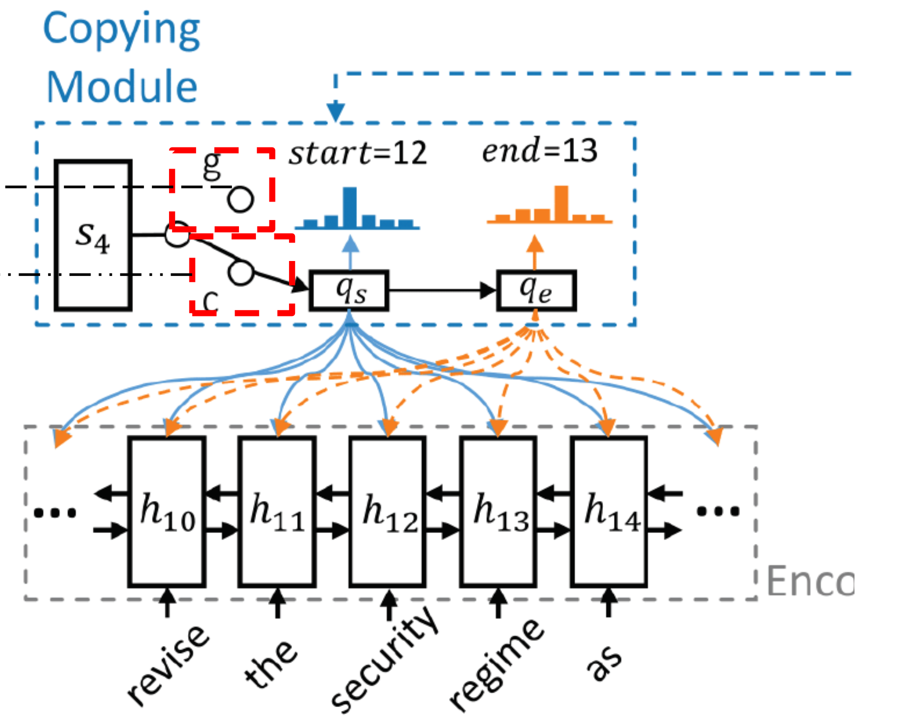
Copy Switch Gate Network

Decoder memory: $m_t = \begin{bmatrix} y_{t-1} \\ s_t \\ c_t \end{bmatrix}$

Copy Switch Network: $\mathcal{G}(x) = \sigma\left(\mathbf{W_2}(\tanh(\mathbf{W_1}x + b_1)) + b_2\right)$

generate probability: $p_g = 1 - p_c$

copy probability: $p_c = \mathcal{G}(m_t)$



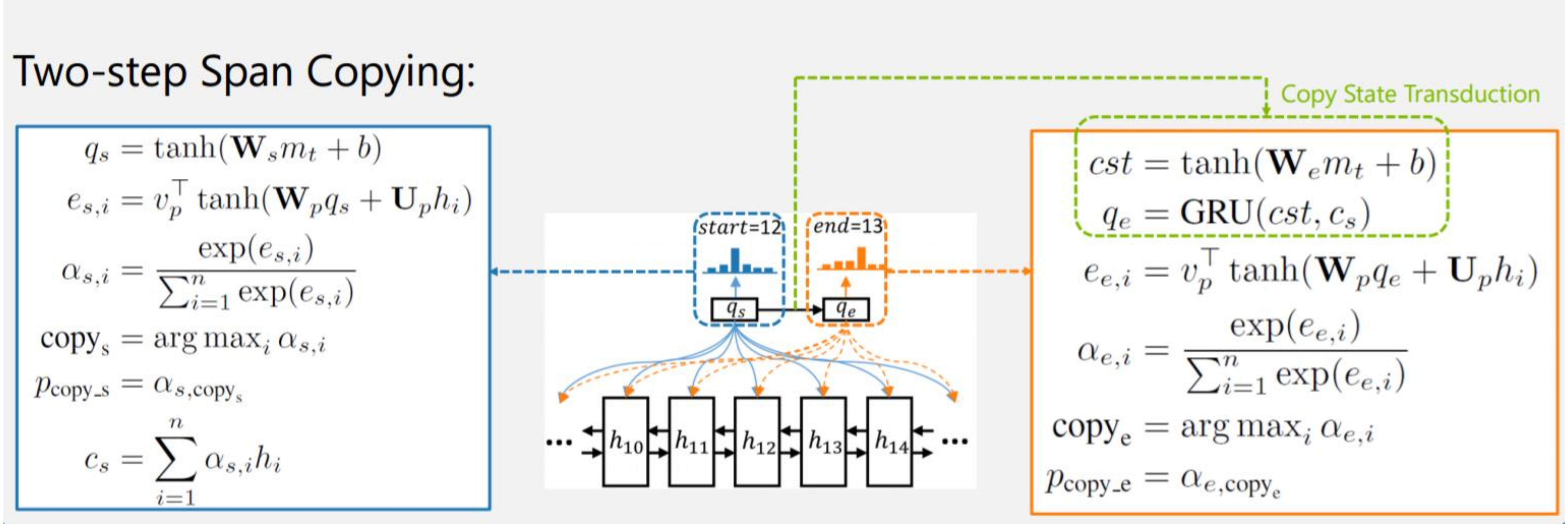
Generate Mode:

$$r_t = \mathbf{W}_r w_{y-1} + \mathbf{U}_r c_t + \mathbf{V}_r s_t$$

$$r'_t = [\max\{r_{t,2j-1}, r_{t,2j}\}]^\top_{j=1,\ldots,d}$$

$$p(y_t|y_{<t}) = \mathrm{softmax}(\mathbf{W}_o r'_t)$$

Pointer Network



Copy Run for Multi-word Span:

adapt itself to smoothly switching between these two modes:

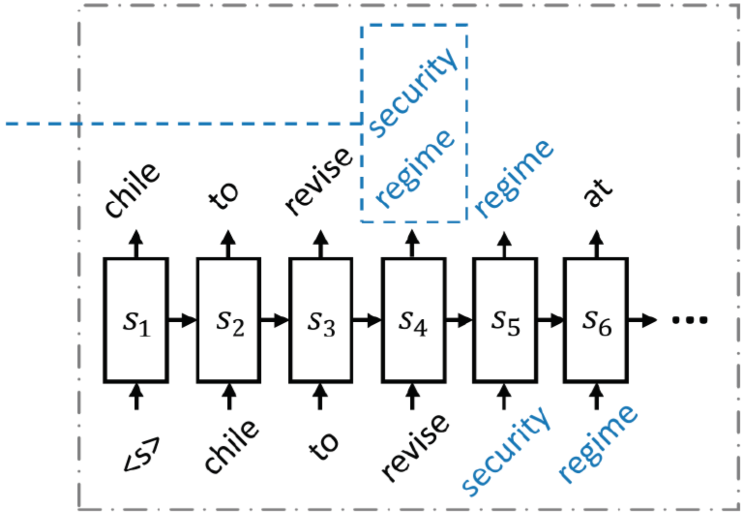
make the decoder GRU keep its state updated when it copies a long sequence

- During training:

  keeping the decoder running over all output words representation

- During testing:

  keeping the decoder running over the copied words



- Objective Function：

$$\mathcal{L} = -\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{t=1}^{T_y}\log p_g p(y_t) + \sum_{span\in C_k}\log p_c p_{\text{start}} p_{\text{end}}\right)$$

Experiments

Task 1: Text Summarization
Dataset: LCSTS dataset (news medias on SinaWeibo1)
Evaluation Metric: ROUGE F1

Models		ROUGE scores on LCSTS (%)		
		R-1	R-2	R-L
RNN	+C	21.5	8.9	18.6
(Hu et al., 2015)	+W	17.7	8.5	15.8
RNN context	+C	29.9	17.4	27.2
(Hu et al., 2015)	+W	26.8	16.1	24.1
COPYNET	+C	**34.4**	**21.6**	**31.3**
	+W	**35.0**	**22.3**	**32.0**

Task 2: Single-turn Dialogue
Dataset: DS-I and DS-II with slot filling on 173 collected patterns.
Evaluation Metric: decoding accuracy

	DS-I (%)		DS-II (%)	
Models	Top1	Top10	Top1	Top10
RNNSearch	44.1	57.7	13.5	15.9
COPYNET	**61.2**	**71.0**	**50.5**	**64.8**

Table 4: The decoding accuracy on the two testing sets. Decoding is admitted success only when the answer is found exactly in the Top-K outputs.

Example Output

Input(1): 今天上午 9 点半， 复旦 投毒案 将 在 上海 二中院 公开审理 。 被害 学生 黄洋 的 亲属 已 从 四川 抵达 上海， 其父 称待 刑事 部分结束 后， 再 提 民事 赔偿， 黄洋 92 岁 的 奶奶 依然 不知情 。 今年 4 月， 在 复旦 上海医学院 读 研究生 的 黄洋 疑遭 室友 林森 浩 投毒， 不幸身亡 。 新民 网

Today 9:30, the Fudan poisoning case will be will on public trial at the Shanghai Second Intermediate Court. The relatives of the murdered student Huang Yang has arrived at Shanghai from Sichuan. His father said that they will start the lawsuit for civil compensation after the criminal section. HuangYang 92-year-old grandmother is still unaware of his death. In April, a graduate student at Fudan University Shanghai Medical College, Huang Yang is allegedly poisoned and killed by his roommate Lin Senhao. Reported by Xinmin

Golden: 林森 浩 投毒案 今日 开审 92 岁 奶奶 尚不知 情

the case of Lin Senhao poisoning is on trial today, his 92-year-old grandmother is still unaware of this

RNN context: 复旦投毒案： 黄洋疑遭室友投毒凶手已从四川飞往上海，父亲命案另有4人被通知家属不治？

CopyNet: 复旦 投毒案 今 在 沪 上 公开审理

the Fudan poisoning case is on public trial today in Shanghai

Input(2): 华谊 兄弟 （ 300027 ） 在 昨日 收盘 后 发布公告 称， 公司 拟 以 自有 资金 3.978 亿 元 收购 浙江 永乐 影视 股份 有限公司 若干股东 持有 的 永乐 影视 51 % 的 股权 。 对于 此项 收购， 华谊 兄弟 董秘 胡明 昨日 表示： " 和 永乐 影视 的 合并 是 对 华谊 兄弟 电视剧 业务 的 一个 加强 。

Huayi Brothers (300027) announced that the company intends to buy with its own fund 397.8 million 51% of Zhejiang Yongle Film LTD's stake owned by a number of shareholders of Yongle Film LTD. For this acquisition, the secretary of the board, Hu Ming, said yesterday: "the merging with Yongle Film is to strengthen Huayi Brothers on TV business".

Golden: 华谊 兄弟 拟 收购 永乐 影视 5 1 % 股权

Huayi Brothers intends to acquire 51% stake of Zhejiang Yongle Film

RNN context: 华谊兄弟收购永乐影视51%股权：与永乐影视合并为“和唐”影视合并的“UNK”和“UNK”的区别？

CopyNet: 华谊 兄弟 拟 3.978 亿 收购 永乐 影视 董秘 称 加强 电视剧 业务

Huayi Brothers is intended to 397.8 million acquisition of Yongle Film secretaries called to strengthen the TV business

Example Output

Input(1): 我出生在上海。 /I was born in Shanghai		Golden:	真巧，我也来自上海，老乡啊！	/What a coincidence, I also come from Shanghai, we are fellow townsman!
	✗	RNNSearch:	真巧，我也来自海南，老乡啊！	/What a coincidence, I also come from Hainan, we are fellow townsman!
	○	CopyNet:	真巧，我也来自上海，老乡啊！	/What a coincidence, I also come from Shanghai, we are fellow townsman!
Input(2): 今天去爬黄山吗？ /Shall we climbing Huangshan?		Golden:	去爬泰山吧，黄山没意思！	/Climb Taishan. Huangshan is boring!
	✗	RNNSearch:	去爬长白山吧，衡山没意思！	/Climb Changbaishan, Hengshan is boring!
	○	CopyNet:	去爬嵩山吧，黄山没意思！	/Climb Songshan, Huangshan is boring!
Input(3): 你喜欢喝雪碧还是喜欢可乐？ /Would you like to drink Sprite or Coke?		Golden:	都不喜欢，谢谢！	/Not at all, thanks!
	✗	RNNSearch:	我喜欢喝蜂蜜！	/I like to drink honey!
	○	CopyNet:	我喜欢喝可乐！	/I like to drink coke!
Input(4): 我大青岛四季宜人 /Our great Qingdao has a good weather!		Golden:	壮我大青岛～～～	/Our Qingdao is great~~~
	✗	RNNSearch:	壮我大哈尔滨～～～	/Our Harbin is great~~~
	○	CopyNet:	壮我大青岛～～～	/Our Qingdao is great~~~
Input(5): 女儿会游蝶泳 /Daughter can swim butterfly?		Golden:	我连游泳都不会T^T	/I cannot swim at all T^T
	○	RNNSearch:	我连游泳都不会T^T	/I cannot swim at all T^T
	○	CopyNet:	我连游泳都不会T^T	/I cannot swim at all T^T
Input(6): 苏有朋是谁? /Who is Su You Peng?		Golden:	苏有朋是一个男明星。	/Su You Peng is a male star.
	✗	RNNSearch:	吴亦凡是一个男明星。	/Wu Yifan is a male star.
	✗	CopyNet:	苏有是一个男明星。	/Su You is a male star

Observations

- Most words are from copy-mode, but the summary is usually still fluent
- COPYNET tends to cover consecutive words in the original document, but it often puts together segments
- COPYNET handles OOV words really well

Task 1: Question Generation give a sentence and its desired answer
Dataset: Question Generation dataset based on SQuADoriginal
Evaluation Metric: BLEU-4

Model	Dev set	Test set
PCFG-Trans‡	9.28	9.31
s2s+att‡	3.01	3.06
NQG‡	10.06	10.13
NQG+‡ (single copy)	12.30	12.18
SeqCopyNet	**13.13**	**13.02**

Task 2: Abstractive Sentence Summarization

Dataset: English Gigaword

Evaluation Metric: ROUGE F1

	Test set in Zhou et al. (2017b)			Our internal test set		
Models	**RG-1**	**RG-2**	**RG-L**	**RG-1**	**RG-2**	**RG-L**
ABS[‡]	37.41^{-}	15.87^{-}	34.70^{-}	-	-	-
s2s+att (greedy)	46.21	24.02	43.30	45.46	22.83	42.66
s2s+att (beam)	47.08	25.11	43.81	46.54	24.18	43.55
NMT + UNK_PS (greedy)	45.64	23.38	42.67	45.21	23.01	42.38
NMT + UNK_PS (beam)	47.05	24.82	43.87	46.52	24.41	43.58
SEASS (greedy)[‡]	45.27	22.88	42.20	-	-	-
SEASS (beam)[‡]	46.86	24.58	43.53	-	-	-
SeqCopyNet (greedy)	46.51	24.14	43.20	46.08	23.99	43.26
SeqCopyNet (beam)	**47.27**	25.07	**44.00**	**47.13**	**24.93**	**44.06**

Example Output

Input:	david ortiz homered and scored three times , including the go-ahead run in the eighth inning , as the boston *red sox* beat the toronto *blue jays 10-9* in the american league on tuesday .
Reference:	david ortiz helps red sox beat blue jays 10-9
SingleCopy:	ortiz homers as red sox beat blue jays
SeqCopyNet:	[*red sox*] beat [*blue jays 10-9*]
Input:	*guyana 's president cheddi jagan* , a long-time marxist turned free - marketeer , died here early thursday , an embassy spokeswoman said . he was 78 .
Reference:	guyana 's president cheddi jagan marxist turned marketeer dies at 78
SingleCopy:	guyana 's president jagan dies at 78
SeqCopyNet:	[*guyana 's president cheddi jagan*] dies at 78
Input:	china topped myanmar 's marine *product exporting countries annually* in the past decade among over 40 's , the local voice weekly quoted the marine products producers and exporters association as reporting sunday .
Reference:	china tops myanmars marine product exporting countries in past
SingleCopy:	china tops myanmar 's marine product export
SeqCopyNet:	china tops myanmar 's marine [*product exporting countries annually*]

Table 3: Examples of generated summaries. The highlighted italic words in brackets are copied as a sequence by SeqCopyNet.

Observations

- SeqCopyNet can copy long spans from input sentence
- SeqCopyNet is good at detecting boundaries:
    - named entities
    - noun phrases
- After copying a long span, the decoder can still generate well

THANKS FOR WATCHING

Quisque velit nisi, pretium ut lacinia in, elementum id enim. Cras ultricies ligula sed magna dictum porta. Quisque velit nisi, pretium ut lacinia in.