Dialogue Summarization

-- 陈湛一

Content

◆*Meeting Summarization*

数据集：*AMI、ICSI*

模　型：

- ✓ ***HMNet***

  A hierarchical network for abstractive meeting summarization with cross-domain pretraining
- ✓ ***PLM_annotator***

  Language Model as an Annotator: Exploring DialoGPT for Dialogue Summarization
- ✓ ***DDAMS***

  Dialogue discourse-aware graph model and data augmentation for meeting summarization
- ✓ ***Longformer+BART***

  ConvoSumm: Conversation Summarization Benchmark and Improved Abstractive Summarization with Argument Mining

◆*Other Domain*

数据集：*MediaSum*

MEDIASUM: A Large-scale Media Interview Dataset for Dialogue Summarization

◆ *Meeting Summarization ：Datasets*

Meeting Transcript (163 turns)

ME: ... I've done some research. We have we have been doing research in a usability lab where we observed users operating remote controls. we let them fill out a questionnaire. Remotes are being considered ugly. F seventy five percent of the people questioned indicated that they thought their remote were was ugly. and an additional eighty percent
e money on a fancy-looking remote control. Fifty percent of the people indicated
of the buttons on a remote control ...
working design. first about how it works. It's really simple. Everybody knows how
utton. The remote determines what button it is, uses the infrared to send a signal to
rcent of the buttons, we should make very few buttons ...
ip the teletext, because in the world of upcoming internet we think teletext is going
inction we don't need in our remote control...

mote control, you can see, this is quite simple remote control. few buttons but This
. people don't like it, so what I was thinking about was keep the general functions

e possible as. let's see what did we say. More. Should be fancy to, fancy design,
about that. Docking station, LCD. general functions And default materials... And
he corporate image in our product. So it has to be visible in our design, in the way

roject document folder... And we have a lunch-break now.

la and the marketing expert discussed what functions are most relevant on a remote,
what his vision for the appearance of the remote is.
p the idea to include a docking station to prevent the remote from getting lost and

for a user interface with large buttons, a display function, a touchscreen, and the
vices.
e target demographic, the buttons the remote should have, the idea of marketing
a audio signal which can sound if the remote is lost, LCD screens, and language

sign despite the new requirement which indicates that the team is not to work with

he main feature is ugly and ugly.
The remote will only have a few buttons.
The remote will feature a small LCD screen.
The remote will have a docking station.
...

Dataset	MEDIASUM	AMI	ICSI	DiDi	CRD3	MultiWOZ	SAMSum
Source	Transcribed Speech					Written	
Type	Interview	Meeting	Meeting	Customer	Game	Booking	Daily
Real dialogue	✓	✓	✓	✓	✓	✓	×
Open domain	✓	×	×	×	×	×	✓
Public	✓	✓	✓	×	✓	✓	✓
Dialogues	463,596	137	59	328,880	159	10,438	16,369
Dial. words	1,553.7	4,757	10,189	/	31,802.8	180.7	83.9
Summ. words	14.4	322	534	/	2062.3	91.9	20.3
Turns	30.0	289	464	/	2,507.4	13.7	9.9
Speakers	6.5	4	6.2	2	9.6	2	2.2

Table 2: Comparison of dialogue summarization datasets. The number of dialogue words, summary words, turns and speakers are all averaged across all dialogues in the dataset.

A Hierarchical Network for Abstractive Meeting Summarization with Cross-Domain Pretraining

Chenguang Zhu*, Ruochen Xu*, Michael Zeng, Xuedong Huang
Microsoft Cognitive Services Research Group
{chezhu,ruox,nzeng,xdh}@microsoft.com

发表时间： *2020/9/20*
发表期刊： *EMNLP*
论文地址： *https://arxiv.org/pdf/2004.02016.pdf*
GitHub ： *https://github.com/microsoft/HMNet*

本文主要概述

数据集：*AMI*、*ICSI*

模　型：*HMNet*

Meeting Transcript (163 turns)
...
PM: ... another point is we have to skip the **teletext**, because in the world of upcoming internet we think **teletext** is going to be a thing of the past.
ID: ... first about how it works. It?s really simple. Everybody knows how a **remote** works. The user presses a button. The **remote** determines what button it is,
PM: ... Few buttons, we talked about that. **Docking station**, **LCD**. general functions And default materials...
...

Summary from our model (23 sentences)
...
The Project Manager announced that the project would not include a **teletext** feature.
The Industrial Designer gave a presentation of the functions of the **remote**.
The group decided on features to include in the remote, to include an **LCD** screen, and a **docking station** to change the layout of the interface.
...

Table 1: Example excerpt of a meeting transcript and the summary generated by our model in AMI dataset. Keywords are highlighted in colors. PM (program manager) and ID (industrial designer) are roles of the speakers. The meeting transcript contains word errors and grammatical glitches as it is the result from the automatic speech recognition system.

主要贡献：

1. 设计分层的网络架构适应长会议转录：word-level、turn-level；
2. 设计角色向量speaker融入到模型中，提升模型的理解能力
3. 由于meeting数据不足，采用news summary data进行预训练，应用到meeting摘要中

Zhu C, Xu R, Zeng M, et al. A hierarchical network for abstractive meeting summarization with cross-domain pretraining[J]. arXiv preprint arXiv:2004.02016, 2020.

Model： HMNet

- ✓ *Encoder*
  - ✓ *Role Vector*
  - ✓ *Word-Level*
  - ✓ *Turn-Level*
- ✓ *Decoder*

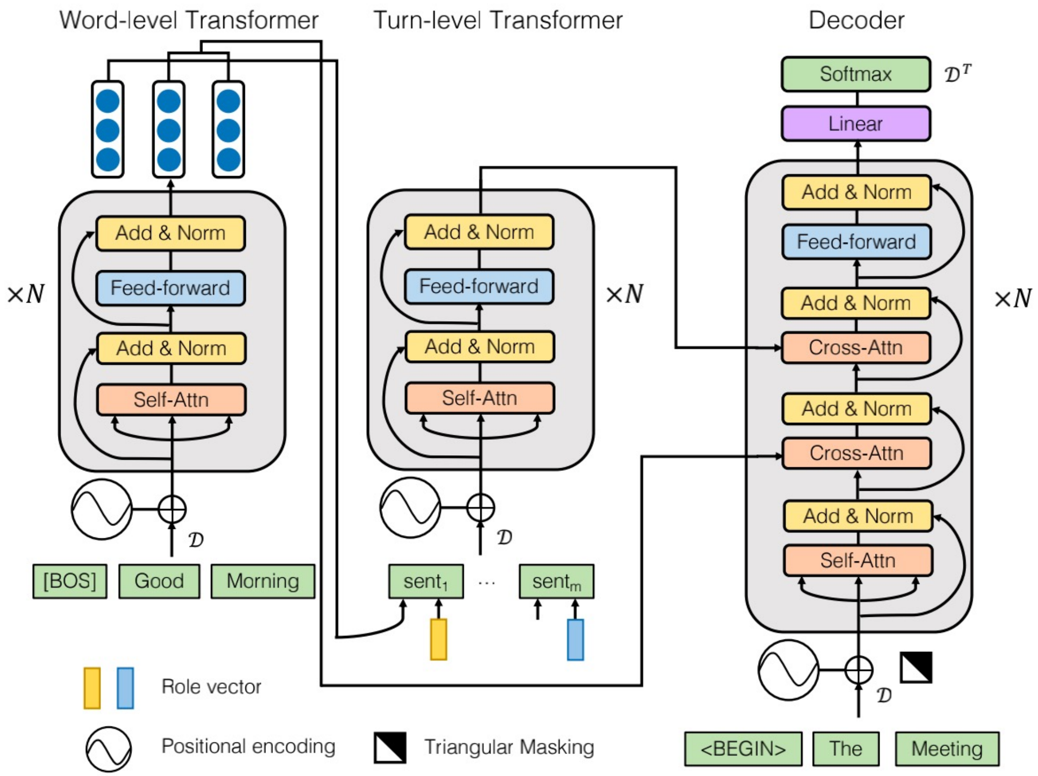


Figure 1: Hierarchical Meeting Summary Network (HMNet) model structure. [BOS] is the special start token inserted before each turn, and its encoding is used in turn-level transformer encoder. Other tokens' encodings enter the cross-attention module in decoder.

Zhu C, Xu R, Zeng M, et al. A hierarchical network for abstractive meeting summarization with cross-domain pretraining[J]. arXiv preprint arXiv:2004.02016, 2020.

Model：HMNet

Role Vector

对于每一个角色，训练一个固定长度的vector表示

Hierarchical Transformer

✓ *Word-Level*
处理每个轮次内的内容

✓ *Turn-Level*
将每个轮次的内容+role vector 拼接起来

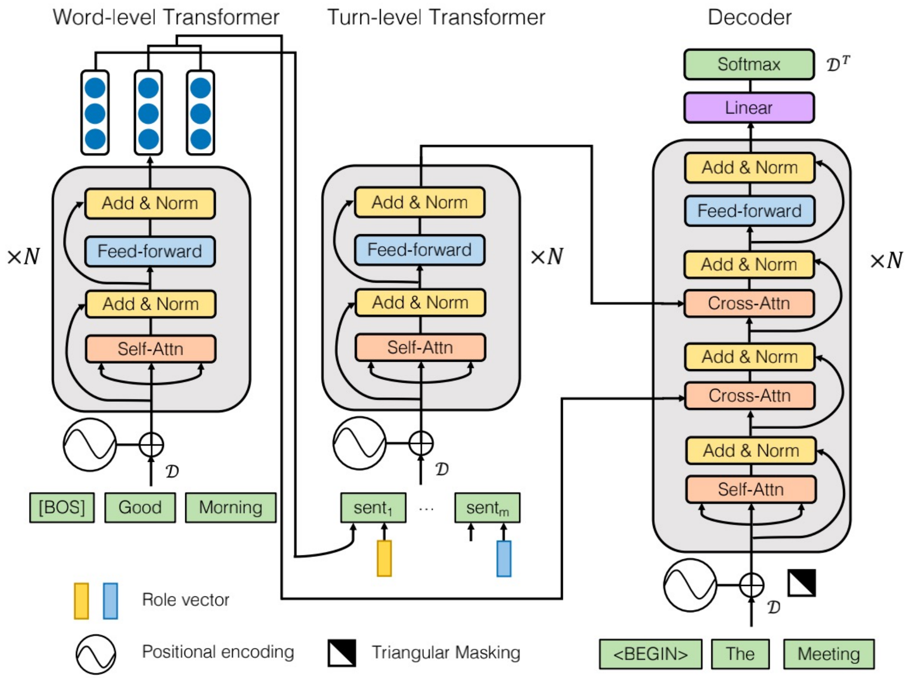


Figure 1: Hierarchical Meeting Summary Network (HMNet) model structure. [BOS] is the special start token inserted before each turn, and its encoding is used in turn-level transformer encoder. Other tokens' encodings enter the cross-attention module in decoder.

Zhu C, Xu R, Zeng M, et al. A hierarchical network for abstractive meeting summarization with cross-domain pretraining[J]. arXiv preprint arXiv:2004.02016, 2020.

HMNet : Experiments

	AMI			ICSI		
Model	ROUGE-1	R-2	R-SU4	ROUGE-1	R-2	R-SU4
Random	35.13	6.26	13.17	29.28	3.78	10.29
Template	31.50	6.80	11.40	/	/	/
TextRank	35.25	6.9	13.62	29.7	4.09	10.64
ClusterRank	35.14	6.46	13.35	27.64	3.68	9.77
UNS	37.86	7.84	14.71	31.60	4.83	11.35
Extractive Oracle	39.49	9.65	13.20	34.66	8.00	10.49
PGNet	40.77	14.87	18.68	32.00	7.70	12.46
Copy from Train	43.24	12.15	14.01	34.65	5.55	10.65
MM (TopicSeg+VFOA)*	**53.29**	13.51	/	/	/	/
MM (TopicSeg)*	51.53	12.23	/	/	/	/
HMNet	53.02	**18.57****	**24.85****	**46.28****	**10.60****	**19.12****

Table 2: ROUGE-1, ROUGE-2, ROUGE-SU4 scores of generated summary in AMI and ICSI datasets. Numbers in bold are the overall best result. * The two baseline MM models require additional human annotations of topic segmentation and visual signals from cameras. ** Results are statistically significant at level 0.05.

Zhu C, Xu R, Zeng M, et al. A hierarchical network for abstractive meeting summarization with cross-domain pretraining[J]. arXiv preprint arXiv:2004.02016, 2020.

Language Model as an Annotator: Exploring DialoGPT for Dialogue Summarization

Xiachong Feng[1], Xiaocheng Feng[1,2*], Libo Qin[1], Bing Qin[1,2], Ting Liu[1,2]
[1]Harbin Institute of Technology, China
[2]Peng Cheng Laboratory, China
{xiachongfeng,xcfeng,lbqin,bqin,tliu}@ir.hit.edu.cn

发表时间： *2021/5/28*
发表期刊： *ACL*
论文地址： *https://arxiv.org/pdf/2105.12544*
GitHub ： *https://github.com/xcfcode/PLM_annotator*

本文主要概述

数据集： *SAMSum* 、*AMI*

模　型： *PLM_annotator*

主要贡献：

--利用DialoGPT作为标注器标注下面三个特征

1. ***Keywords Extraction: DialoGPT$_{KE}$***
2. ***Redundancy Detection: DialoGPT$_{RD}$***
3. ***Topic Segmentation: DialoGPT$_{TS}$***

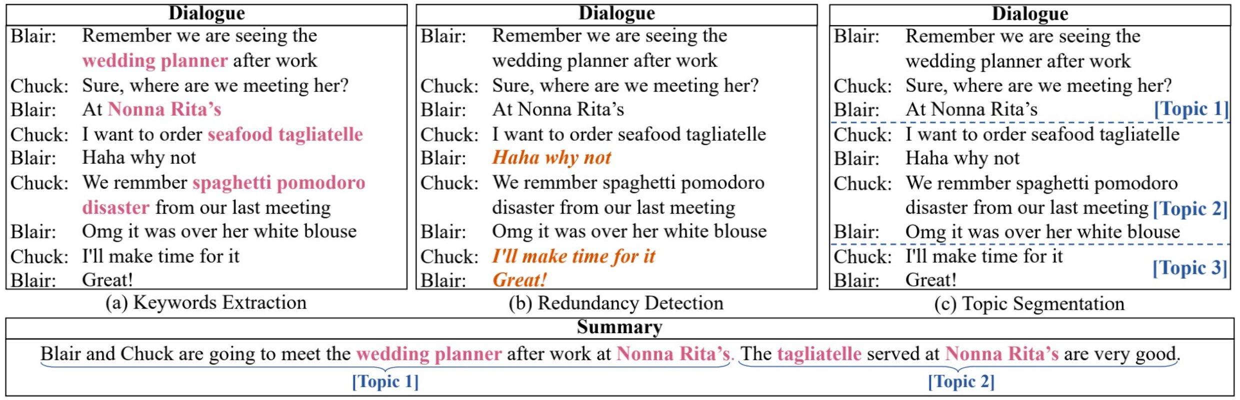


Figure 1: Example dialogue from SAMSum (Gliwa et al., 2019) with the human annotated summary. (a) Keywords extraction aims to extract words that are most important to the dialogue. (b) Redundancy detection aims to detect nonsignificant utterances in the dialogue. (c) Topic segmentation aims to divide the whole dialogue into several fine-grained topics. All three auxiliary information can do good to final summary generation.

Feng X, Feng X, Qin L, et al. Language Model as an Annotator: Exploring DialoGPT for Dialogue Summarization[J]. arXiv preprint arXiv:2105.12544, 2021.

Model： Dialogue Annotator

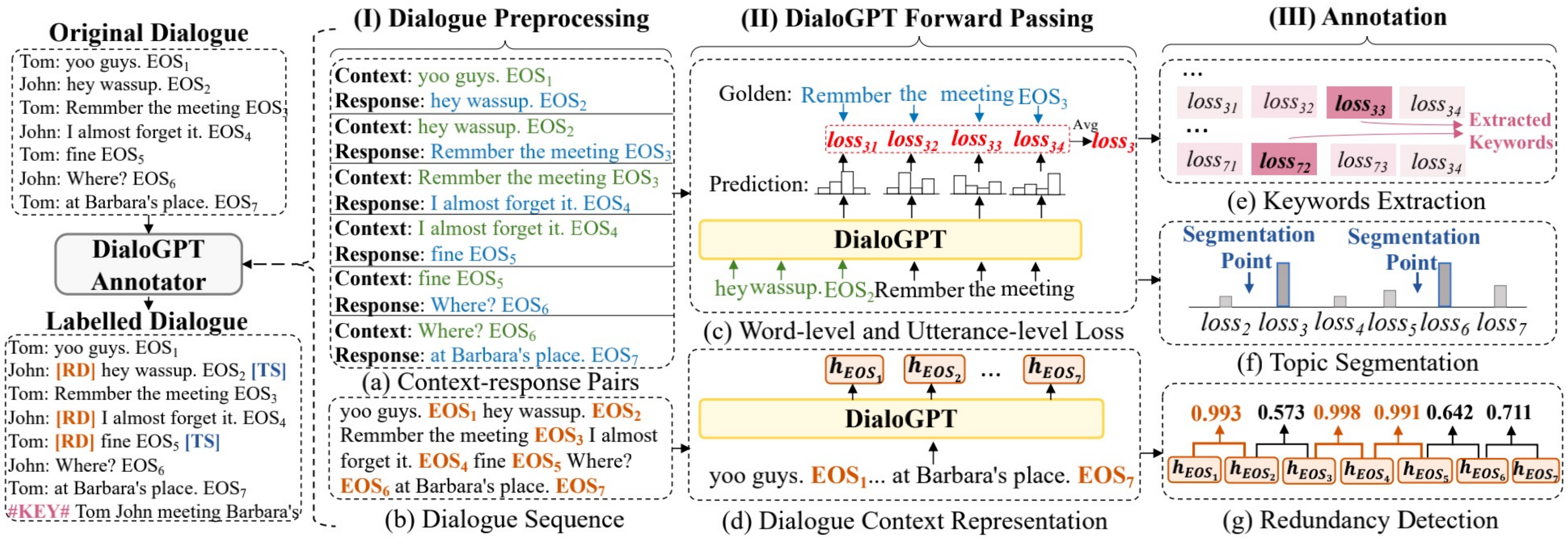
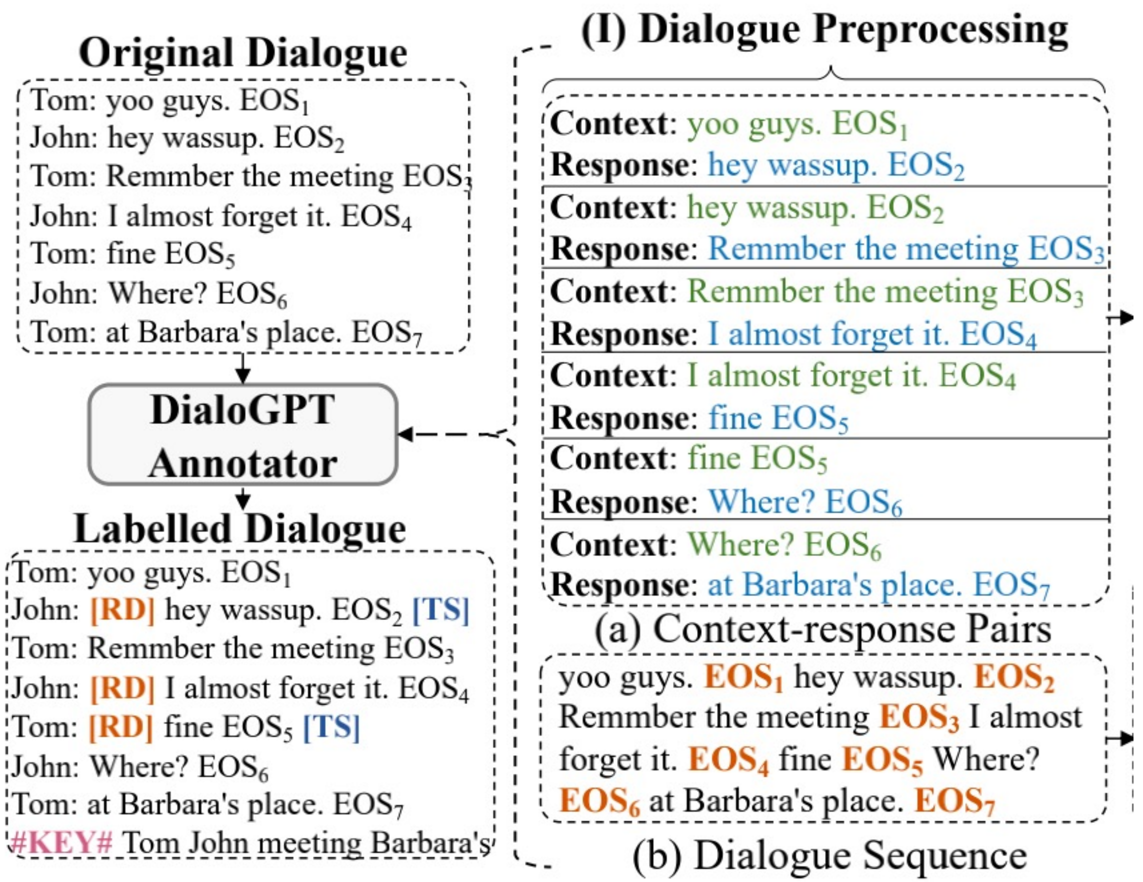


Figure 2: Illustration of our DialoGPT annotator. (I) Given one dialogue, we preprocess it into two formats: context-response pairs and the dialogue sequence. (II) We input them into the DialoGPT, after the forward pass, we can get the word-level and utterance-level predicted losses and representations for dialogue context. (III) We perform three annotation tasks: keywords extraction, redundancy detection and topic segmentation. Finally, we can get a labelled dialogue. #KEY#, [RD] and [TS] are specific tags, which are inserted into the dialogue.

Feng X, Feng X, Qin L, et al. Language Model as an Annotator: Exploring DialoGPT for Dialogue Summarization[J]. arXiv preprint arXiv:2105.12544, 2021.

Dialogue Annotator：Dialogue Processing

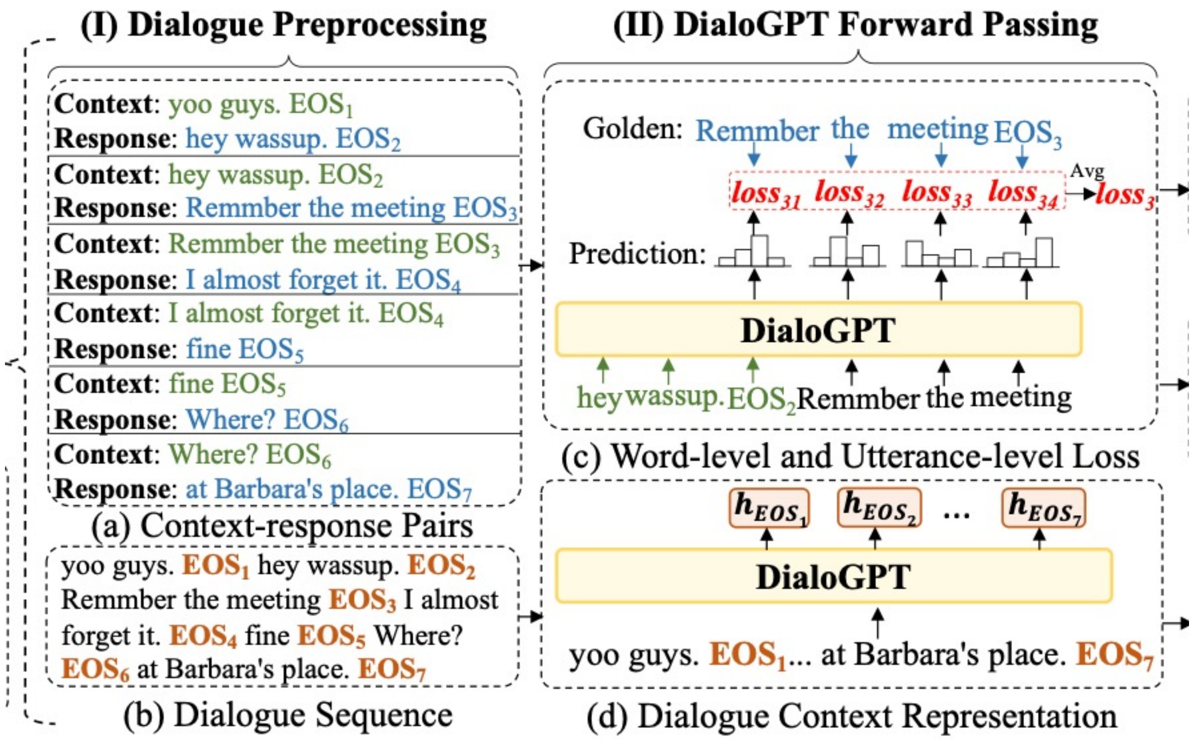


将原始文本处理成两种格式:

1. *context-response pairs*
2. *Dialogue Sequence*

Feng X, Feng X, Qin L, et al. Language Model as an Annotator: Exploring DialoGPT for Dialogue Summarization[J]. arXiv preprint arXiv:2105.12544, 2021.

Dialogue Annotator：DialoGPT Forward Passing



$$loss_{i,t} = -\log p\left(u_{i,t} | u_{i,<t}, u_{i-1}\right)$$

$$loss_i = \frac{1}{|u_i|+1} \sum_{t=1}^{|u_i|+1} loss_{i,t} \quad (1)$$

$$\boldsymbol{h}_{\mathrm{EOS}_1}, \boldsymbol{h}_{\mathrm{EOS}_2}, \ldots, \boldsymbol{h}_{\mathrm{EOS}_{|\mathcal{D}|}} = \boldsymbol{H}(\mathrm{EOS}) \quad (2)$$

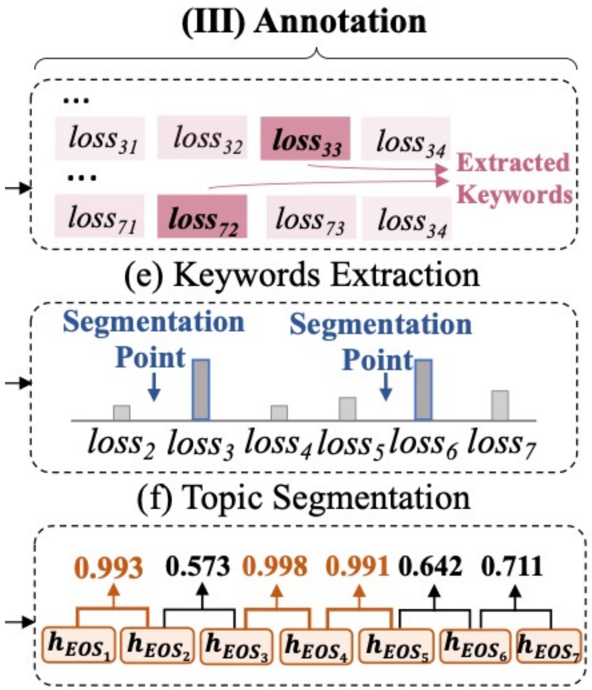
两个目的：

1. *对于context-response pairs，得到word-level、utterance-level的predicted loss、*

2. *对于dialogue sequence，得到每个EOS的表示*

Feng X, Feng X, Qin L, et al. Language Model as an Annotator: Exploring DialoGPT for Dialogue Summarization[J]. arXiv preprint arXiv:2105.12544, 2021.

Dialogue Annotator： Annotation



(e) Keywords Extraction

(f) Topic Segmentation

(g) Redundancy Detection

1. ***Keywords Extraction: DialoGPT$_{KE}$***

   将word中loss较大的词认为是关键词，并添加到句子后。

$$\mathcal{D}_{\text{KE}} = [\underbrace{p_1, u_{1,1}, \ldots,}_{\mathcal{D}} \underbrace{\texttt{\#KEY\#}, \mathbb{P}, \texttt{Key}_1, \texttt{Key}_2, \ldots}_{keywords}].^4$$

2. ***Redundancy Detection: DialoGPT$_{RD}$***

$$\mathcal{D}_{\text{RD}} = [p_1, \texttt{[RD]}, u_{1,1}, \ldots, \texttt{EOS}_1, \ldots, p_{|\mathcal{D}|}, \ldots, \texttt{EOS}_{|\mathcal{D}|}]$$

3. ***Topic Segmentation: DialoGPT$_{TS}$***

   假设上下文很难预测时，一个句子的response可能是另一个主题

$$\mathcal{D}_{\text{TS}} = [p_1, u_{1,1}, \ldots, \texttt{EOS}_1, \texttt{[TS]}, p_2, u_{2,1}, \ldots, \texttt{EOS}_2, \ldots]$$

Feng X, Feng X, Qin L, et al. Language Model as an Annotator: Exploring DialoGPT for Dialogue Summarization[J]. arXiv preprint arXiv:2105.12544, 2021.

Dialogue Annotator：Summarizer

- ✓ 对于SAMSum，采用BART预训练语言模型

$$\boldsymbol{X}^N = \textsc{Encoder}(\boldsymbol{X}^0) \overset{N}{\underset{n=1}{:=}} \textsc{Ffn}\left(\textsc{Att}(\boldsymbol{X}^{n-1})\right)$$
$$\boldsymbol{Y}^M = \textsc{Decoder}(\boldsymbol{Y}^0, \boldsymbol{X}^N)$$
$$\overset{M}{\underset{m=1}{:=}} \textsc{Ffn}\left(\textsc{Att}\left(\textsc{Att}(\boldsymbol{Y}^{m-1}), \boldsymbol{X}^N\right)\right) \tag{3}$$

- ✓ 对于AMI，采用PGN

Seq2Seq attention model 和Pointer-NetWork的混合模型

Feng X, Feng X, Qin L, et al. Language Model as an Annotator: Exploring DialoGPT for Dialogue Summarization[J]. arXiv preprint arXiv:2105.12544, 2021.

Dialogue Annotator : Experiments

Model	R-1	R-2	R-L
	Extractive		
LONGEST-3	32.46	10.27	29.92
TextRank	29.27	8.02	28.78
	Abstractive		
Transformer	36.62	11.18	33.06
D-HGN	42.03	18.07	39.56
TGDGA	43.11	19.15	40.49
DialoGPT	39.77	16.58	38.42
MV-BART	53.42	27.98	**49.97**††
	Ours		
BART	52.98	27.67	49.06
BART($\mathcal{D}_{\mathrm{KE}}$)	**53.43**††	**28.03**††	49.93
BART($\mathcal{D}_{\mathrm{RD}}$)	53.39	28.01	49.49
BART($\mathcal{D}_{\mathrm{TS}}$)	53.34	27.85	49.64
BART($\mathcal{D}_{\mathrm{ALL}}$)	**53.70**†	**28.79**†	**50.81**†

Table 2: Test set results on the SAMSum dataset, where "R" is short for "ROUGE". BART means fine-tuning BART on the original SAMSum. BART($\mathcal{D}_{\mathrm{KE}}$), BART($\mathcal{D}_{\mathrm{RD}}$) and BART($\mathcal{D}_{\mathrm{TS}}$) represent fine-tuning BART on the SAMSum with keywords, redundancy and topic annotation respectively. $\mathcal{D}_{\mathrm{ALL}}$ means the SAMSum with all three annotations. † and †† indicate the first-ranked and second-ranked results respectively.

Model	R-1	R-2	R-L
	Extractive		
TextRank	35.19	6.13	15.70
SummaRunner	30.98	5.54	13.91
	Abstractive		
UNS	37.86	7.84	13.72
TopicSeg	**51.53**††	12.23	**25.47**†
HMNet	**52.36**†	**18.63**†	24.00
	Ours		
PGN	48.34	16.02	23.49
PGN($\mathcal{D}_{\mathrm{KE}}$)	50.22	17.74	24.11
PGN($\mathcal{D}_{\mathrm{RD}}$)	50.62	16.86	24.27
PGN($\mathcal{D}_{\mathrm{TS}}$)	48.59	16.07	24.05
PGN($\mathcal{D}_{\mathrm{ALL}}$)	50.91	**17.75**††	**24.59**††

Table 3: Test set results on the AMI dataset. PGN($\mathcal{D}_{\mathrm{KE}}$), PGN($\mathcal{D}_{\mathrm{RD}}$) and PGN($\mathcal{D}_{\mathrm{TS}}$) represent training PGN on the AMI with keywords, redundancy and topic annotation respectively.

SAMSum		AMI	
Model	**BS**	**Model**	**BS**
BART	86.91	PGN	80.51
MV-BART	88.46	HMNet	82.24
BART($\mathcal{D}_{\mathrm{ALL}}$)	**90.04**	PGN($\mathcal{D}_{\mathrm{ALL}}$)	**82.76**

Table 4: Test set results on the SAMSum and AMI. "BS" is short for BERTScore.

Feng X, Feng X, Qin L, et al. Language Model as an Annotator: Exploring DialoGPT for Dialogue Summarization[J]. arXiv preprint arXiv:2105.12544, 2021.

Dialogue Discourse-Aware Graph Model and Data Augmentation for Meeting Summarization

Xiachong Feng , Xiaocheng Feng , Bing Qin , Xinwei Geng
Harbin Institute of Technology, China
{xiachongfeng, xcfeng, bqin, xwgeng}@ir.hit.edu.cn

发表时间： *2021/5/19*
发表期刊： *IJCAI*
论文地址： *https://www.ijcai.org/proceedings/2021/0524.pdf*
GitHub ： *https://github.com/xcfcode/DDAMS/*

本文主要概述

数据集：*AMI*、*ICSI*

模　型：*DDAMS*

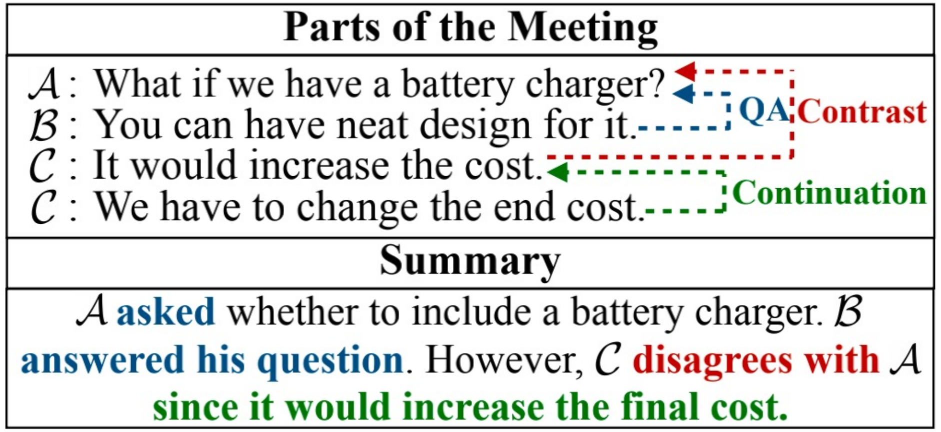


Figure 1: An example of a meeting with its corresponding summary. *QA*, *Contrast* and *Continuation* are dialogue discourse relations, which explicitly show the interaction between utterances.

主要贡献：

1. 在对话摘要中尝试使用对话结构信息建模；
2. 设计一个数据增强策略，将meeting dialogue中的*query*作为摘要、*answer*作为对话文本，数据扩充了约20倍。

Feng X, Feng X, Qin B, et al. Dialogue discourse-aware graph model and data augmentation for meeting summarization[J]. Dialogue, 2020, 1: U2.

Model： DDMAS

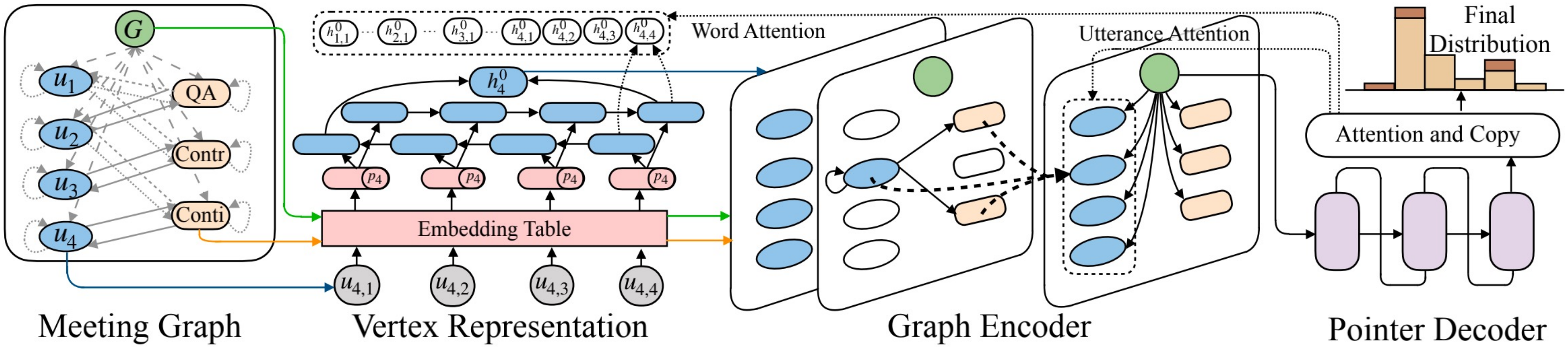


Figure 3: Illustration of our DDAMS model. (1) First, we construct our meeting graph consisting of three types of vertices: global vertex, utterance vertex and discourse relation vertex. (2) Then, the vertex representation module gives each type of vertex an initial representation. (3) Further, the graph encoder performs convolutional computation over the meeting graph based on the relational graph convolutional network. (4) Finally, the pointer decoder attends to the updated utterance representations and the word representations to generate the summary words either from the fixed-length vocabulary or copy from the input.

Feng X, Feng X, Qin B, et al. Dialogue discourse-aware graph model and data augmentation for meeting summarization[J]. Dialogue, 2020, 1: U2.

DDAMS：Meeting Graph

a) Dialogue Discourse Parsing
   得到对话文本中的篇章关系

b) Levi Graph Transformation
   将边之间的关系转换为额外的顶点

c) Levi Graph with global node and self edges
   增加 global node 和 self edges 增强图中的信息

d) Meeting Graph
   将图中关系最终转换为六种关系，得到最终的 meeting graph

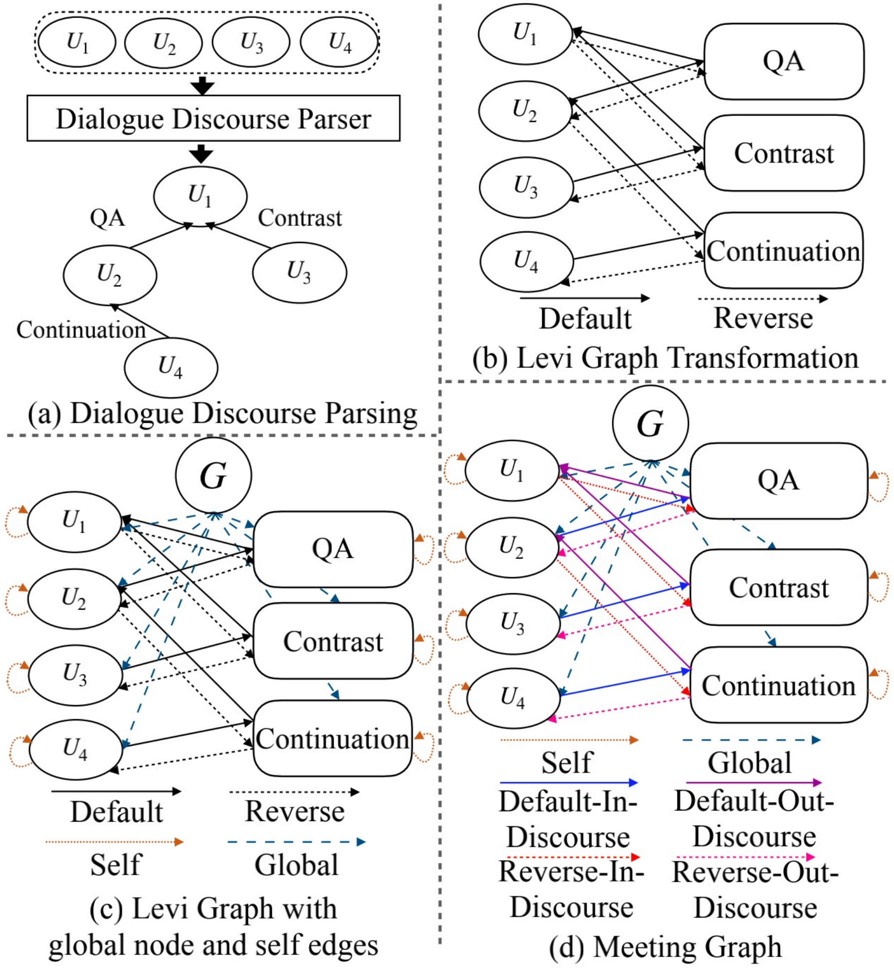


Figure 2: Illustration of meeting graph construction process.

Feng X, Feng X, Qin B, et al. Dialogue discourse-aware graph model and data augmentation for meeting summarization[J]. Dialogue, 2020, 1: U2.

DDAMS：Vertex Representation

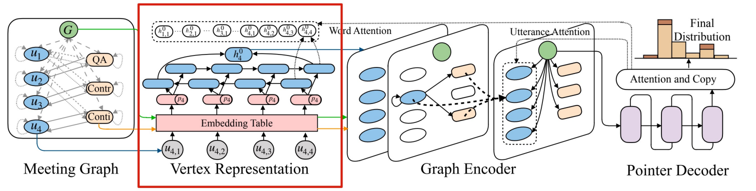


Figure 3: Illustration of our DDAMS model. (1) First, we construct our meeting graph consisting of three types of vertices: global vertex, utterance vertex and discourse relation vertex. (2) Then, the vertex representation module gives each type of vertex an initial representation. (3) Further, the graph encoder performs convolutional computation over the meeting graph based on the relational graph convolutional network. (4) Finally, the pointer decoder attends to the updated utterance representations and the word representations to generate the summary words either from the fixed-length vocabulary or copy from the input.

三种Vertex需要初始化：Global、Relation 、Utterence

- **Global Vertex & Relation Vertex**

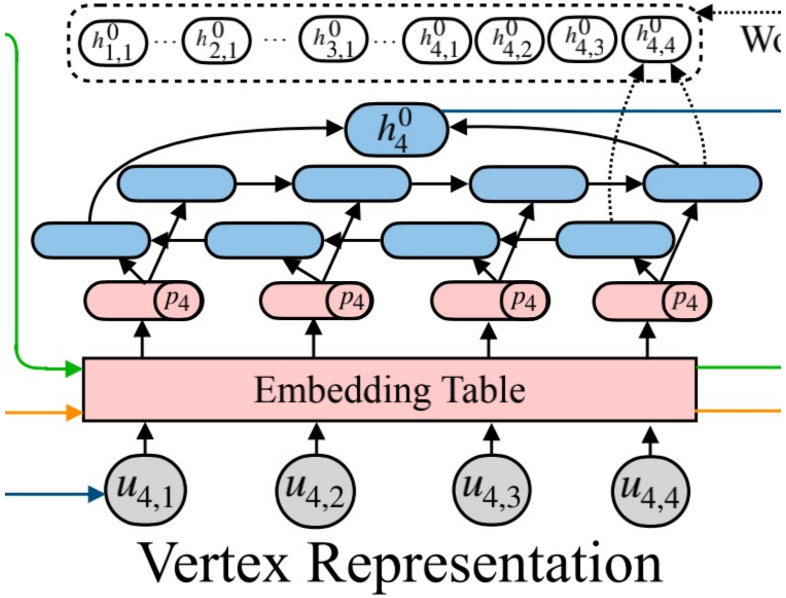
  通过embedding table 进行初始

- **Utterence vertex**

  使用BiLSTM对Utterence vertex进行前向和后向编码

- **Speaker information**

  将Speaker 信息整合到graph中



Feng X, Feng X, Qin B, et al. Dialogue discourse-aware graph model and data augmentation for meeting summarization[J]. Dialogue, 2020, 1: U2.

DDAMS : *Graph Encoder & Pointer Decoder*

➢ *Graph Encoder*

$$\boldsymbol{h}_i^{(l+1)} = \mathrm{ReLU}\left(\sum_{r \in \mathbb{R}_M} \sum_{v_j \in \mathbb{N}_i^r} \frac{1}{|\mathbb{N}_i^r|} \boldsymbol{W}_r^{(l)} \boldsymbol{h}_j^{(l)}\right) \quad (1)$$

$$\boldsymbol{h}_i^{(l+1)} = \mathrm{ReLU}\left(\sum_{r \in \mathbb{R}_M} \sum_{v_j \in \mathbb{N}_i^r} g_j^{(l)} \frac{1}{|\mathbb{N}_i^r|} \boldsymbol{W}_r^{(l)} \boldsymbol{h}_j^{(l)}\right) \quad (3)$$

➢ *Pointer Decoder*

$$\begin{aligned} e_{i,j}^t &= \boldsymbol{s}_t^\top \boldsymbol{W}_a \boldsymbol{h}_{i,j}^0 \\ \boldsymbol{a}^t &= \mathrm{softmax}(\boldsymbol{e}^t) \\ \boldsymbol{h}_t^{wl} &= \sum_i \sum_j a_{i,j}^t \boldsymbol{h}_{i,j}^0 \end{aligned} \quad (4)$$

$$\boldsymbol{h}_t^* = [\boldsymbol{h}_t^{wl}; \boldsymbol{h}_t^{ul}],$$

Feng X, Feng X, Qin B, et al. Dialogue discourse-aware graph model and data augmentation for meeting summarization[J]. Dialogue, 2020, 1: U2.

Dialogue Discourse-Aware Data Augmentation

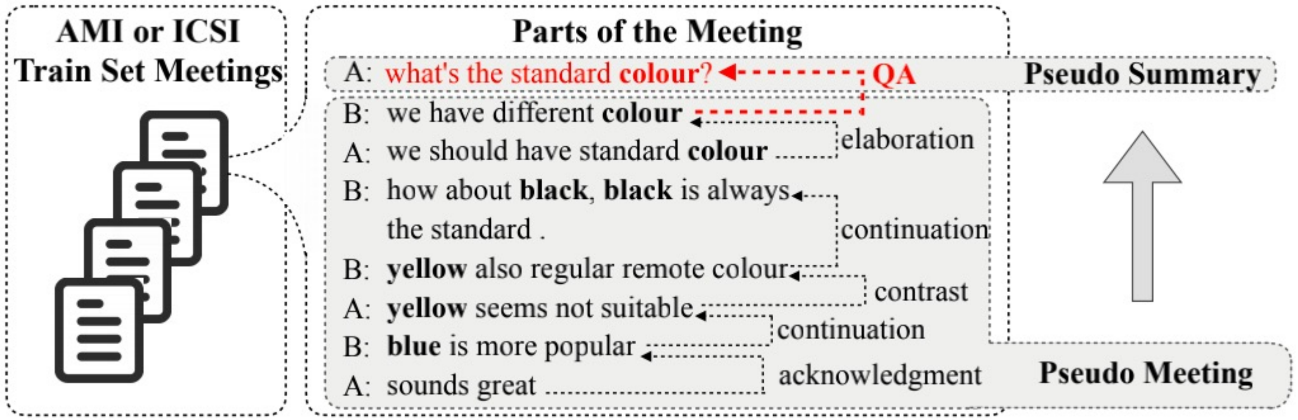


Figure 4: Illustration of how to construct a pseudo meeting-summary pair. Given a meeting from the original meeting train set, we use *QA* discourse relation to identify the question in the meeting. Then, the subsequent discussion with discourse relations becomes a pseudo meeting and the question becomes a pseudo summary.

	AMI Pseudo Corpus	ICSI Pseudo Corpus
# of Original Data	97	53
# of Pseudo Data	1539	1877
Avg.Tokens	124.44	107.44
Avg.Sum	13.18	11.97

Table 1: Pseudo-summarization corpus statistics. "# of Original Data" means the number of original meetings in the train set, "# of Pseudo Data" means the number of pseudo meeting-summary pairs, "Avg.Tokens" means the average length of pseudo meetings and "Avg.Sum" means the average length of pseudo summaries.

1. 将原文中的一个question作为Pseudo Summary，对应的answer作为Pseudo Meeting；
2. 排除不包含名词和形容词的问题。

Feng X, Feng X, Qin B, et al. Dialogue discourse-aware graph model and data augmentation for meeting summarization[J]. Dialogue, 2020, 1: U2.

DDAMS：Experiments

	Model	AMI R-1	AMI R-2	AMI R-L	ICSI R-1	ICSI R-2	ICSI R-L
Extractive	TextRank [Mihalcea and Tarau, 2004]	35.19	6.13	15.70	30.72	4.69	12.97
	SummaRunner [Nallapati *et al.*, 2017]	30.98	5.54	13.91	27.60	3.70	12.52
Abstractive	UNS [Shang *et al.*, 2018]	37.86	7.84	13.72	31.73	5.14	14.50
	Pointer-Generator [See *et al.*, 2017]	42.60	14.01	22.62	35.89	6.92	15.67
	HRED [Serban *et al.*, 2016]	49.75	18.36	23.90	39.15	7.86	16.25
	Sentence-Gated [Goo and Chen, 2018]	49.29	19.31	24.82	39.37	9.57	17.17
	TopicSeg [Li *et al.*, 2019]	51.53	12.23	25.47	-	-	-
	HMNet [Zhu *et al.*, 2020]	52.36	18.63	24.00	**45.97**	10.14	18.54
Ours	DDAMS	51.42	20.99	24.89	39.66	10.09	17.53
	DDAMS + DDADA	**53.15**	**22.32**	**25.67**	40.41	**11.02**	**19.18**
	DDAMS + DDADA (w/o fine-tune)	28.35	4.67	14.92	25.94	4.18	13.92

Table 2: Test set results on AMI and ICSI Datasets, where "R-1" is short for "ROUGE-1", "R-2" for "ROUGE-2", "R-L" for "ROUGE-L". The DDAMS represents the model that is trained only on the meeting dataset. The DDAMS +DDADA means the model that is pre-trained using pseudo-summarization data and then fine-tuned on meeting dataset. DDAMS +DDADA (w/o fine-tune) means the model that is pre-trained using pseudo-summarization data and without fine-tuning on the meeting dataset.[7]

	Model	Relevance	Informativeness
AMI	Ground-truth	4.60	4.56
	Sentence-Gated	3.16	3.60
	HMNet	3.60	3.72
	DDAMS	3.80	3.76
	DDAMS +DDADA	**3.84**	**3.88**
ICSI	Ground-truth	4.76	4.48
	Sentence-Gated	3.32	3.48
	HMNet	3.80	3.52
	DDAMS	3.76	3.28
	DDAMS +DDADA	**3.84**	**3.60**

Table 3: Human evaluation results.

Feng X, Feng X, Qin B, et al. Dialogue discourse-aware graph model and data augmentation for meeting summarization[J]. Dialogue, 2020, 1: U2.

ConvoSumm: Conversation Summarization Benchmark and Improved Abstractive Summarization with Argument Mining

Alexander R. Fabbri[†] **Faiaz Rahman[†]** **Imad Rizvi[†]** **Borui Wang[†]**
Haoran Li [‡] **Yashar Mehdad[‡]** **Dragomir Radev[†]**
[†] Yale University [‡] Facebook AI
{alexander.fabbri, faiaz.rahman, imad.rizvi,
borui.wang, dragomir.radev}@yale.edu
{aimeeli, mehdad}@fb.com

发表时间： *2021/6/1*
发表期刊： *ACL*
论文地址： *https://arxiv.org/pdf/2106.00829.pdf*
GitHub ： *https://github.com/Yale-LILY/ConvoSumm*

本文主要概述

数据集：NYT、Reddit、Stack、Email

模　型：longformer + bart

主要贡献：

1. 设计一个annotation protocol标注四个数据集：基于issues-viewpoints-assertion

2. 通过图的构建进行argument mining，过滤掉一些噪声输入

Fabbri A R, Rahman F, Rizvi I, et al. ConvoSumm: Conversation Summarization Benchmark and Improved Abstractive Summarization with Argument Mining[J]. arXiv preprint arXiv:2106.00829, 2021.

Argument Graph Summarization

Graph construction：

1. *Argument extraction*
2. *Relationship type classification*
3. *Major claim dection*
4. *Graph contraction*

Fabbri A R, Rahman F, Rizvi I, et al. ConvoSumm: Conversation Summarization Benchmark and Improved Abstractive Summarization with Argument Mining[J]. arXiv preprint arXiv:2106.00829, 2021.

Argument Graph： *argument extraction*

- Argument unit 是句子，从句子中识别出哪些是claims，哪些是premise
  Claims 可以判断哪些事情是真的，premise 用于得出结论的命题，claims和premise作为information-nodes(I-nodes)

- 移除non- argumentative units

- 使用Chakrabarty et al. (2019) 提出的方法训练一个三分类用于argument extraction

---

Fabbri A R, Rahman F, Rizvi I, et al. ConvoSumm: Conversation Summarization Benchmark and Improved Abstractive Summarization with Argument Mining[J]. arXiv preprint arXiv:2106.00829, 2021.

Argument Graph : *Relationship Type Classification*

- 使用entailment决定文档中论元(argument unit)的关系 ;
- *把所有的边简化为support edges*
- *每一个premise和一个claim绑定，利用RoBERTa对其进行分数评定，若大于0.33，则创建边*

Fabbri A R, Rahman F, Rizvi I, et al. ConvoSumm: Conversation Summarization Benchmark and Improved Abstractive Summarization with Argument Mining[J]. arXiv preprint arXiv:2106.00829, 2021.

MEDIASUM: A Large-scale Media Interview Dataset for Dialogue Summarization

Chenguang Zhu*, Yang Liu*, Jie Mei, Michael Zeng
Microsoft Cognitive Services Research Group
{chezhu,yaliu10,jimei,nzeng}@microsoft.com

发表时间： *2021/5/28*
发表期刊： *NAACL*
论文地址： *https://arxiv.org/pdf/2103.06410.pdf*
GitHub ： *https://github.com/zcgzcgzcg1/MediaSum/*

MediaSum

Statistics	NPR	CNN
Dialogues	49,420	414,176
Avg. words in dialogue	906.3	1,630.9
Avg. words in summary	40.2	11.3
Turns	24.2	30.7
Speakers	4.0	6.8
Novel summary words	33.6%	24.9%

Table 1: Data statistics of NPR and CNN transcripts and summaries.

National Public Radio (NPR)：

- ✓ *每个采访内容的转录作为对话文本，overview作为摘要，共49.4K*
- ✓ *过滤掉word超过200的摘要*

CNN：

- ✓ *采访内容作为对话文本，主题描述作为摘要，共414.2K*
- ✓ *存在多个主题，进行主题分割，一个采访可能拆分成几个摘要*


```
{
  "id": "NPR-11",
  "program": "Day to Day",
  "date": "2008-06-10",
  "url": "https://www.npr.org/templates/story/story.php?storyId=91356794",
  "title": "Researchers Find Discriminating Plants",
  "summary": "The ‘sea rocket’ shows preferential treatment to plants that are its kin. Evolutionary plant ecologist Susan Dudley of McMaster University in Ontario discusses her discovery.",
  "utt": [
    "This is Day to Day. I’m Madeleine Brand.",
    "And I’m Alex Cohen.",
    "Coming up, the question of who wrote a famous religious poem turns into a very unchristian battle.",
    "First, remember the 1970s? People talked to their houseplants, played them classical music. They were convinced plants were sensuous beings and there was that 1979 movie, ‘The Secret Life of Plants.’",
    "Only a few daring individuals, from the scientific establishment, have come forward with offers to replicate his experiments, or test his results. The great majority are content simply to condemn his efforts without taking the trouble to investigate their validity.",
    ...
    "OK. Thank you.",
    "That’s Susan Dudley. She’s an associate professor of biology at McMaster University in Hamilt on Ontario. She discovered that there is a social life of plants."
  ],
  "speaker": [
    "MADELEINE BRAND, host",
    "ALEX COHEN, host",
    "ALEX COHEN, host",
    "MADELEINE BRAND, host",
    "Unidentified Male",
    ...
    "Professor SUSAN DUDLEY (Biology, McMaster University)",
    "MADELEINE BRAND, host"
  ]
}
```


Table 5: Example dialogue and summary from MEDIASUM. The number of strings in *utt* and *speaker* fields are the same.

MediaSum:Data stastics

1. Dialogue topics:

politics (26.3%), international news (13.3%), crime (12.7%), economy (12.5%) and US news (11.7%).

2. 基本数据 (average)

30turns, 6.5 speaker, 1553.7 words, 14.4 words/summary

3. Positional Bias

- ✓ *CNN和NPR在文本开头都包含了更多的summary words*
- ✓ *NPR在文本结尾处也出现了一些summary words*

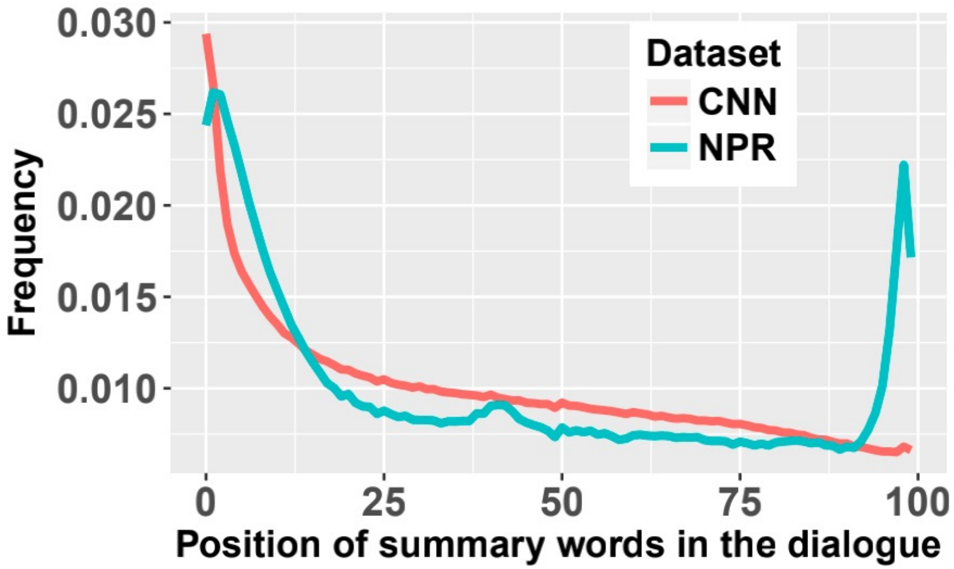


Figure 1: The frequency of the non-stop summary words appearing at different positions of the dialogue. The positions are normalized to [0, 100].

Experiments

Model	R-1	R-2	R-L
LEAD-3	14.96	5.10	13.29
PTGen	28.77	12.24	24.18
UniLM	32.70	17.27	29.82
BART	**35.09**	**18.05**	**31.44**

Table 3: ROUGE-1, ROUGE-2 and ROUGE-L F1 scores for models on MEDIASUM test set.

Model	R-1	R-2	R-L
CNN			
LEAD-3	13.36	4.37	11.10
PTGen	27.54	11.47	23.45
BART	34.07	17.57	31.36
UniLM	31.97	16.97	29.88
UniLM_{Com}	31.88	16.97	29.79
NPR			
LEAD-3	28.39	11.21	19.90
PTGen	35.86	16.01	24.46
BART	43.55	21.99	32.03
UniLM	41.42	20.73	30.65
UniLM_{Com}	41.58	21.25	31.24

Table 7: ROUGE-1, ROUGE-2 and ROUGE-L F1 scores on the CNN and NPR partitions of the test data. All models are trained on the corresponding partition of the training data, except UniLM_{Com}, which is trained on the entire MEDIASUM.

Experiments : transfer learning

Model	**R-1**	**R-2**	**R-L**
AMI			
UniLM	50.61	**19.33**	25.06
UniLM+MEDIASUM	**51.90**	**19.33**	**25.58**
ICSI			
UniLM	42.91	9.78	17.72
UniLM+MEDIASUM	**43.65**	**10.13**	**18.59**
SAMSum			
UniLM	50.00	26.03	42.34
UniLM+MEDIASUM	**50.55**	**26.39**	**42.68**

Table 4: Results on AMI, ICSI and SAMSum by using MEDIASUM as a dataset for transfer learning.

Thanks !

References

1. *Zhu C, Xu R, Zeng M, et al. A hierarchical network for abstractive meeting summarization with cross-domain pretraining[J]. arXiv preprint arXiv:2004.02016, 2020.*
2. *Feng X, Feng X, Qin L, et al. Language Model as an Annotator: Exploring DialoGPT for Dialogue Summarization[J]. arXiv preprint arXiv:2105.12544, 2021.*
3. *Feng X, Feng X, Qin B, et al. Dialogue discourse-aware graph model and data augmentation for meeting summarization[J]. Dialogue, 2020, 1: U2.*
4. *Fabbri A R, Rahman F, Rizvi I, et al. ConvoSumm: Conversation Summarization Benchmark and Improved Abstractive Summarization with Argument Mining[J]. arXiv preprint arXiv:2106.00829, 2021.*
5. *Zhu C, Liu Y, Mei J, et al. MediaSum: A Large-scale Media Interview Dataset for Dialogue Summarization[J]. arXiv preprint arXiv:2103.06410, 2021.*
6. *Feng X, Feng X, Qin B. A survey on dialogue summarization: Recent advances and new frontiers[J]. arXiv preprint arXiv:2107.03175, 2021.*