哈爾濱工業大學
HARBIN INSTITUTE OF TECHNOLOGY

Research Center for Social Computing and Information Retrieval
理解语言 认知社会

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

Xiachong Feng , Xiaocheng Feng , Bing Qin , Xinwei Geng

Research Center for Social Computing and Information Retrieval
Harbin Institute of Technology

Meeting Summarization

- Distill the most important information from a meeting (content selection)
  - General abstract, decisions, actions, problems, task ......
- Convert them into a short textual passage (surface realization)

Parts of the Meeting
$\mathcal{A}$: What if we have a battery charger? $\mathcal{B}$: You can have neat design for it. $\mathcal{C}$: It would increase the cost. $\mathcal{C}$: We have to change the end cost.
Summary
$\mathcal{A}$ asked whether to include a battery charger. $\mathcal{B}$ answered his question. However, $\mathcal{C}$ disagrees with $\mathcal{A}$ since it would increase the final cost.

Background

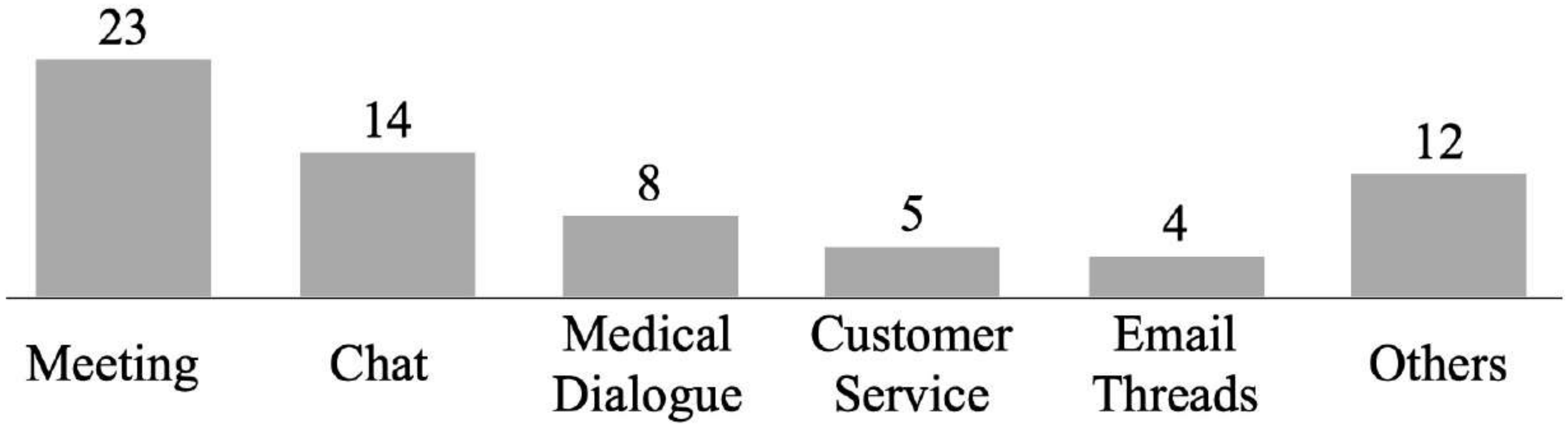
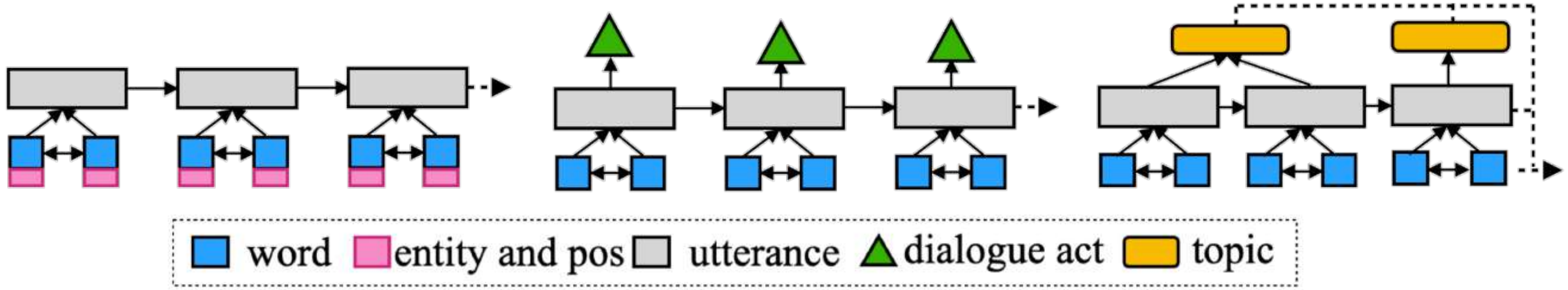


Figure 1: The number of dialogue summarization papers published over the past 5 years for each domain.

Problems

- **Sequential text modeling is inadequate.**
  - hinder the exploration of inherently rich interactive relations between utterances.

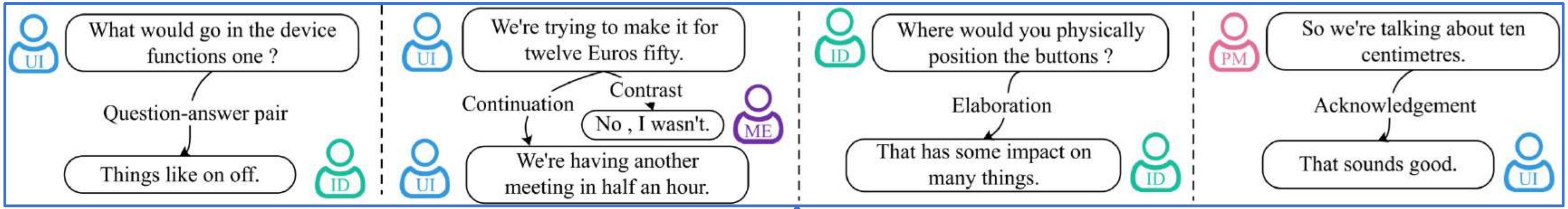


- **Lack of sufficient training data.**
  - hinders the ability of data-hungry neural models.

Dataset	Domain	Train	Valid	Test
CNNDM	News	287227	13368	11490
AMI	**Meeting**	**97**	**20**	**20**
ICSI	**Meeting**	**53**	**25**	**6**

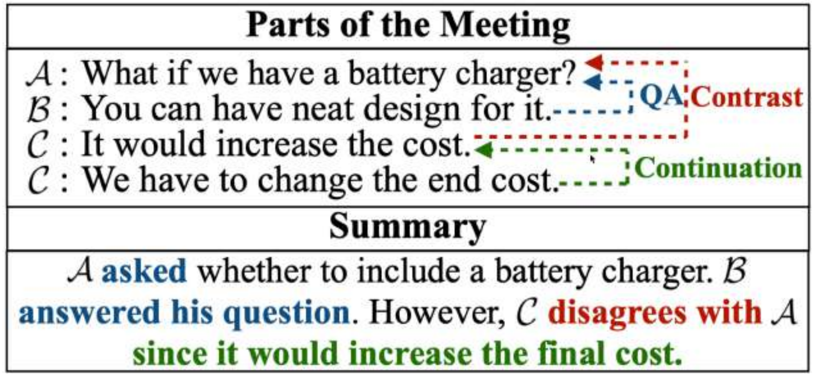
Solution

Dialogue Discourse



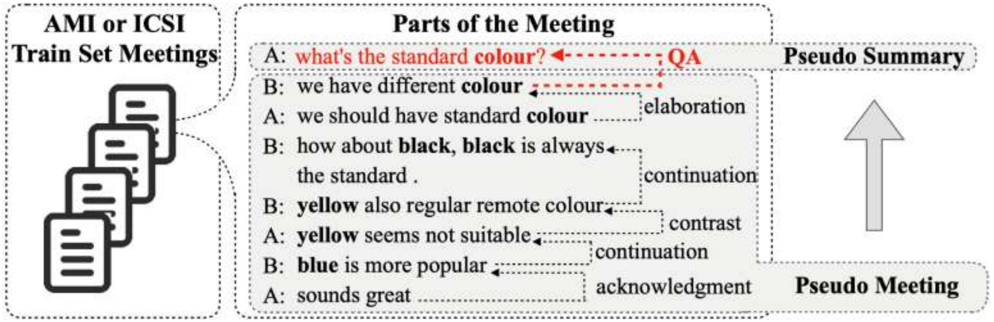
Dialogue Discourse-Aware Meeting Summarizer

DDAMS



Dialogue Discourse-Aware Data Augmentation

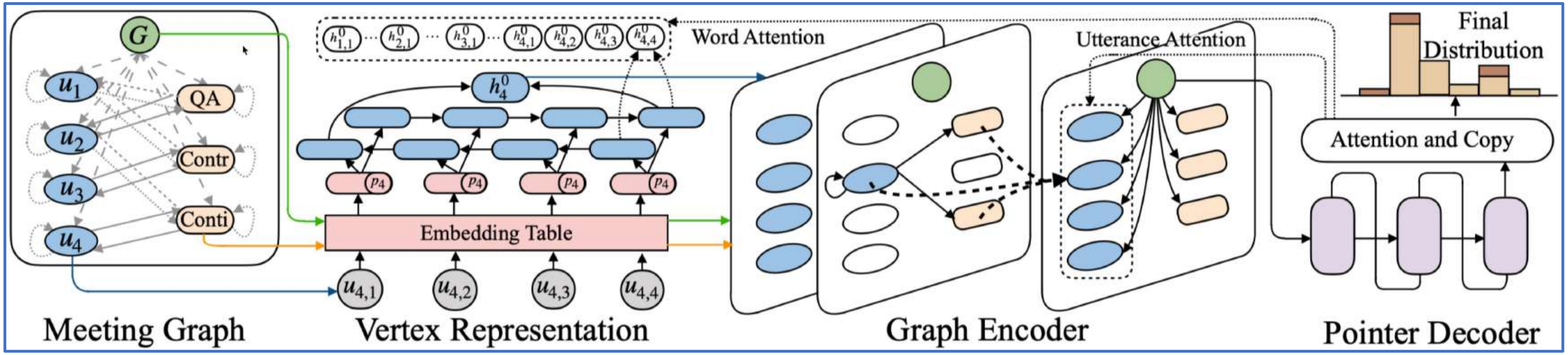
DDADA



Dialogue Discourse-Aware Meeting Summarizer

Gives each type of vertex
an initial representation

Generate the summary

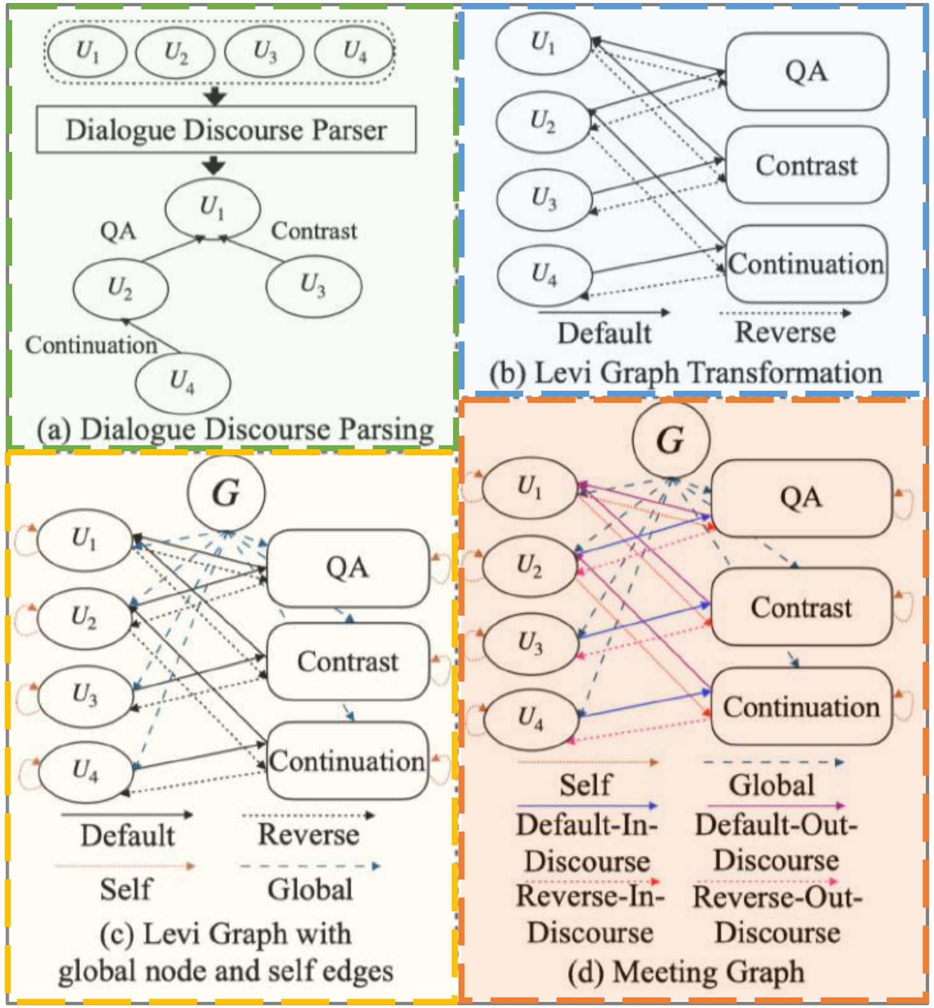


Meeting graph construction

Performs convolutional computation over the meeting graph

Meeting Graph Construction

- SOTA dialogue discourse parser
- 16 discourse relations

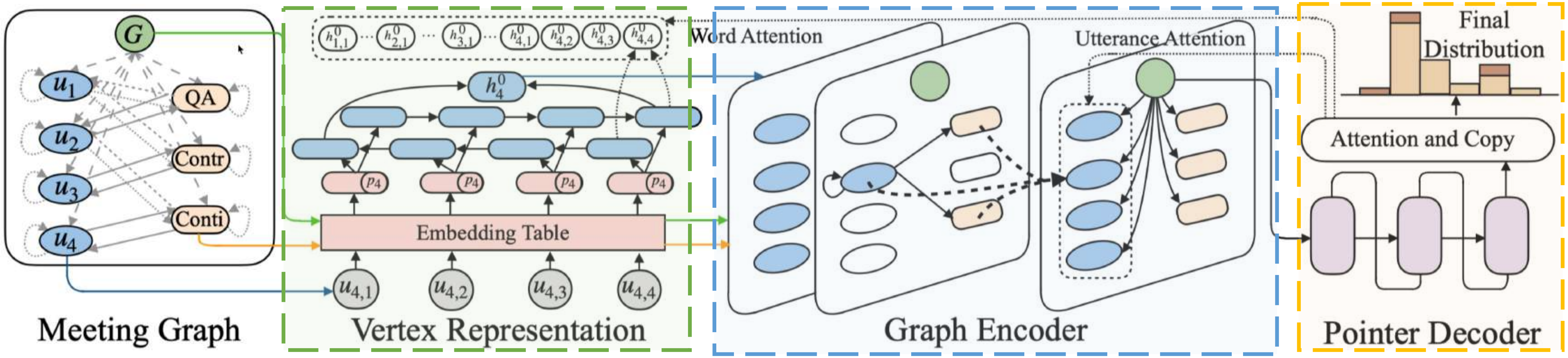


- Levi graph transformation

- Global edges and Self edges

- Reverse edges

Graph2Seq Framework



Training Objective

$$\mathcal{L} = -\sum_{t=1}^{|\mathcal{Y}^*|} \log p\left(y_t^* | y_1^* \ldots y_{t-1}^*, \mathcal{U}\right)$$

$$\overrightarrow{\boldsymbol{h}_{i,j}} = \mathrm{LSTM}_f\left(\overrightarrow{\boldsymbol{h}_{i,j-1}}, \boldsymbol{e}_{i,j}\right)$$

$$\overleftarrow{\boldsymbol{h}_{i,j}} = \mathrm{LSTM}_b\left(\overleftarrow{\boldsymbol{h}_{i,j+1}}, \boldsymbol{e}_{i,j}\right)$$

$$\boldsymbol{h}_{i,j} = [\overrightarrow{\boldsymbol{h}_{i,j}}; \overleftarrow{\boldsymbol{h}_{i,j}}]$$

$$\boldsymbol{h}_i^0 = [\overrightarrow{\boldsymbol{h}_{i,|u_i|}}; \overleftarrow{\boldsymbol{h}_{i,1}}]$$

$$g_j^{(l)} = \mathrm{sigmoid}\left(\boldsymbol{W}_{g,r}^{(l)} \boldsymbol{h}_j^{(l)}\right)$$

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

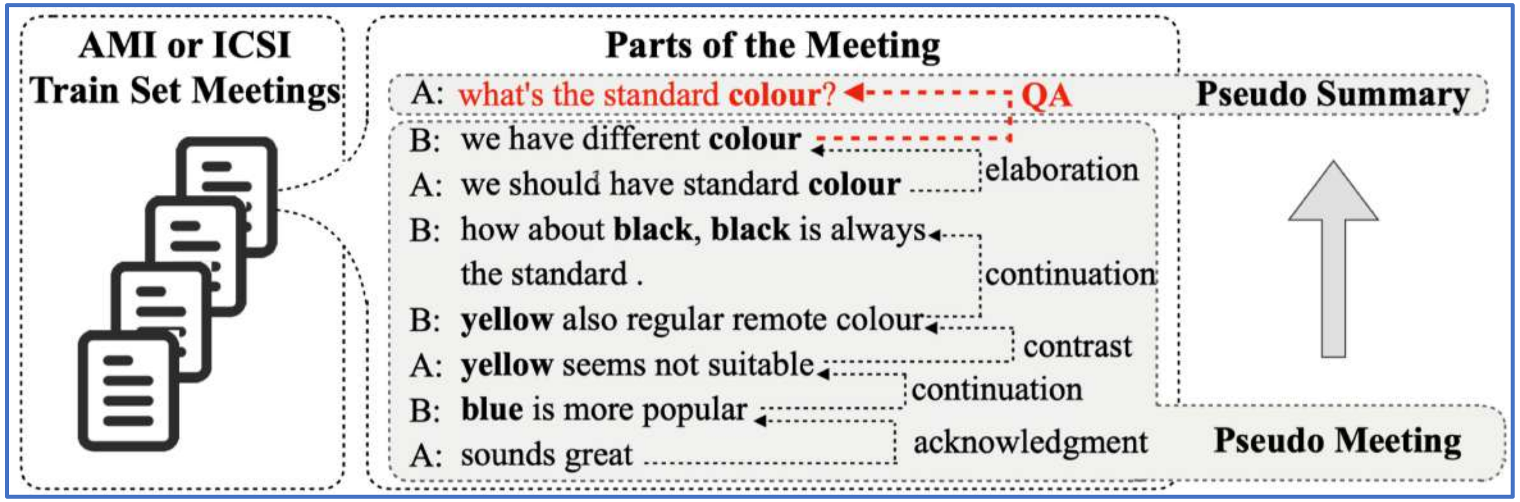
$$e_{i,j}^t = \boldsymbol{s}_t^\top \boldsymbol{W}_a \boldsymbol{h}_{i,j}^0$$

$$\boldsymbol{a}^t = \mathrm{softmax}(\boldsymbol{e}^t)$$

$$\boldsymbol{h}_t^{wl} = \sum_i \sum_j a_{i,j}^t \boldsymbol{h}_{i,j}^0$$

Dialogue Discourse-Aware Data Augmentation

- **Motivation**
  - a **question** often sparks a **discussion** and contains salient terms or concepts expressed in the discussion.



	AMI	ICSI
#	137	59
Avg.Turns	289	464
Avg.Tokens	4,757	10,189
Avg.Sum	322	534

	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

Experiments

- **Datasets:**
  - AMI and ICSI

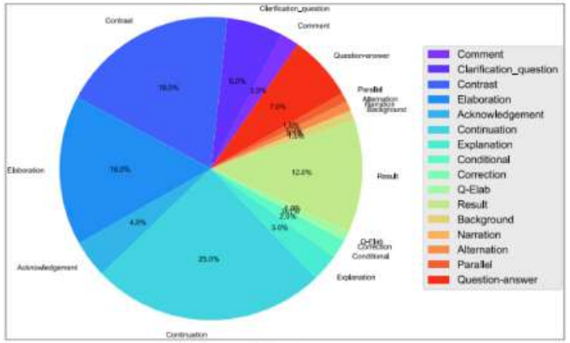
	AMI	ICSI
#	137	59
Avg.Turns	289	464
Avg.Tokens	4,757	10,189
Avg.Sum	322	534

	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

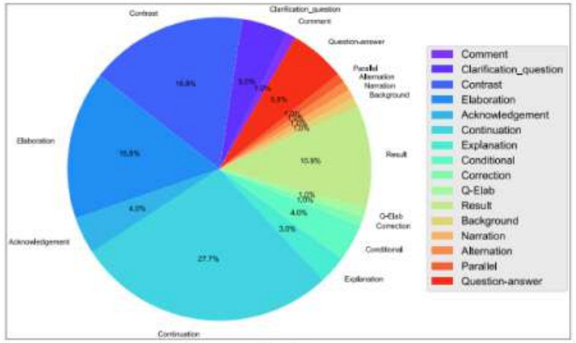
- **Dialogue Discourse Parser**
  - Deep Sequential
- **Evaluation**
  - ROUGE

Relation Distribution Statistics

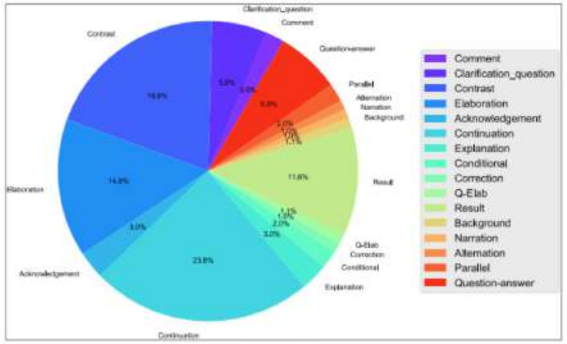
AMI | **ICSI**



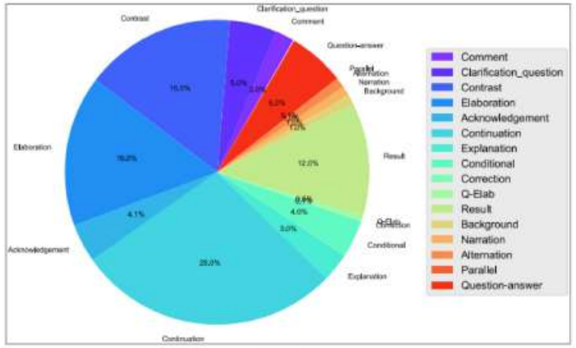
Train



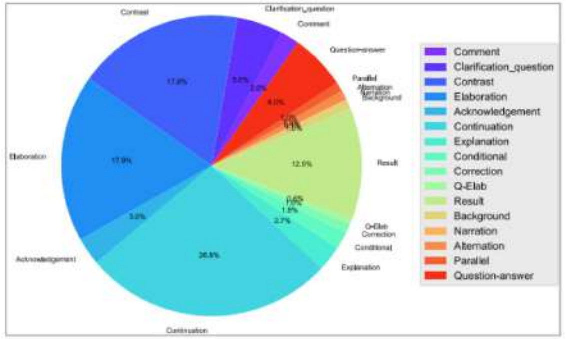
Train



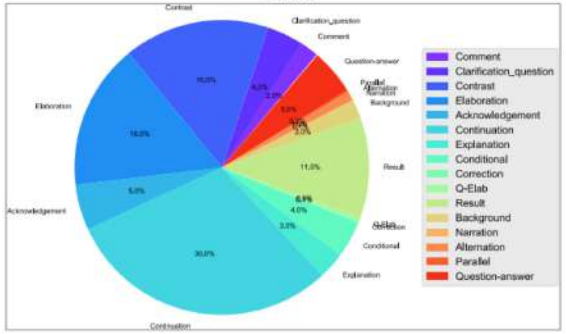
Valid



Valid



Test



Test

Automatic Evaluation

		Model	AMI R-1	AMI R-2	AMI R-L	ICSI R-1	ICSI R-2	ICSI R-L	
Baseline	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**	**SOTA!**
		DDAMS + DDADA (w/o fine-tune)	28.35	4.67	14.92	25.94	4.18	13.92	

Datasets: AMI, ICSI

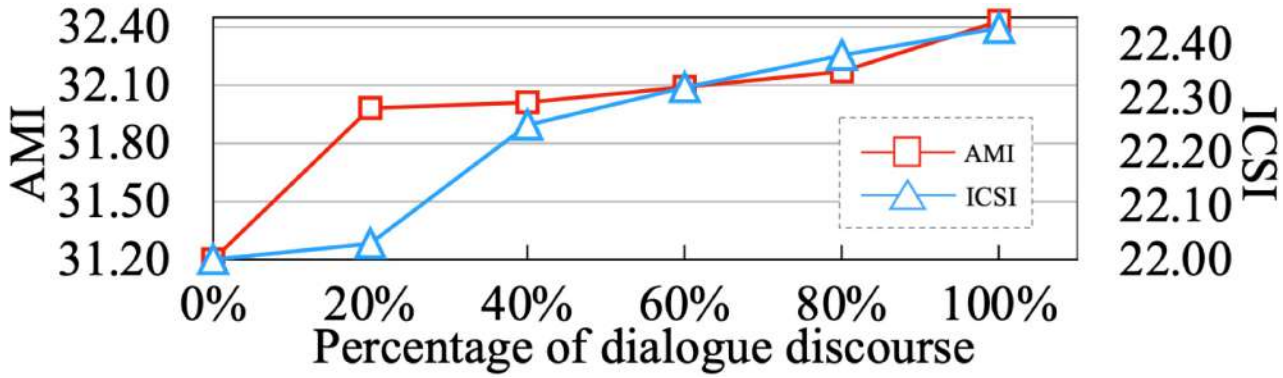
Human Evaluation

- DdaMS+DdaDA achieves higher scores in both relevance and informativeness.
- Ground truth obtains the highest scores compare with generated summaries indicating the challenge of this task.

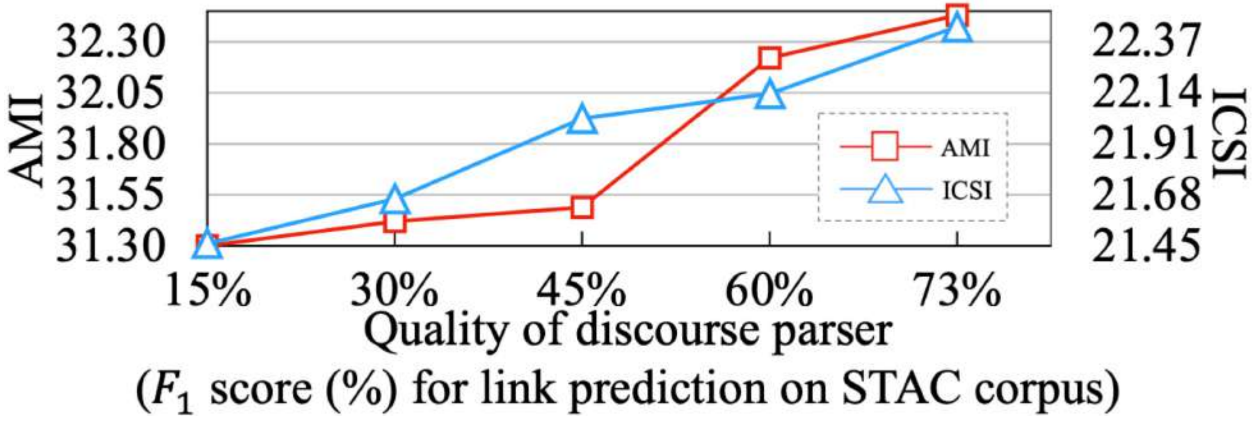
	Model	Relevance	Informativeness
AMI	Ground-truth	4.60	4.56
	Sentence-Gated	3.16	3.60
	HMNet	3.60	3.72
	DdaMS	3.80	3.76
	DdaMS +DdaDA	**3.84**	**3.88**
ICSI	Ground-truth	4.76	4.48
	Sentence-Gated	3.32	3.48
	HMNet	3.80	3.52
	DdaMS	3.76	3.28
	DdaMS +DdaDA	**3.84**	**3.60**

Analyses

- Effect of the number of dialogue discourse.

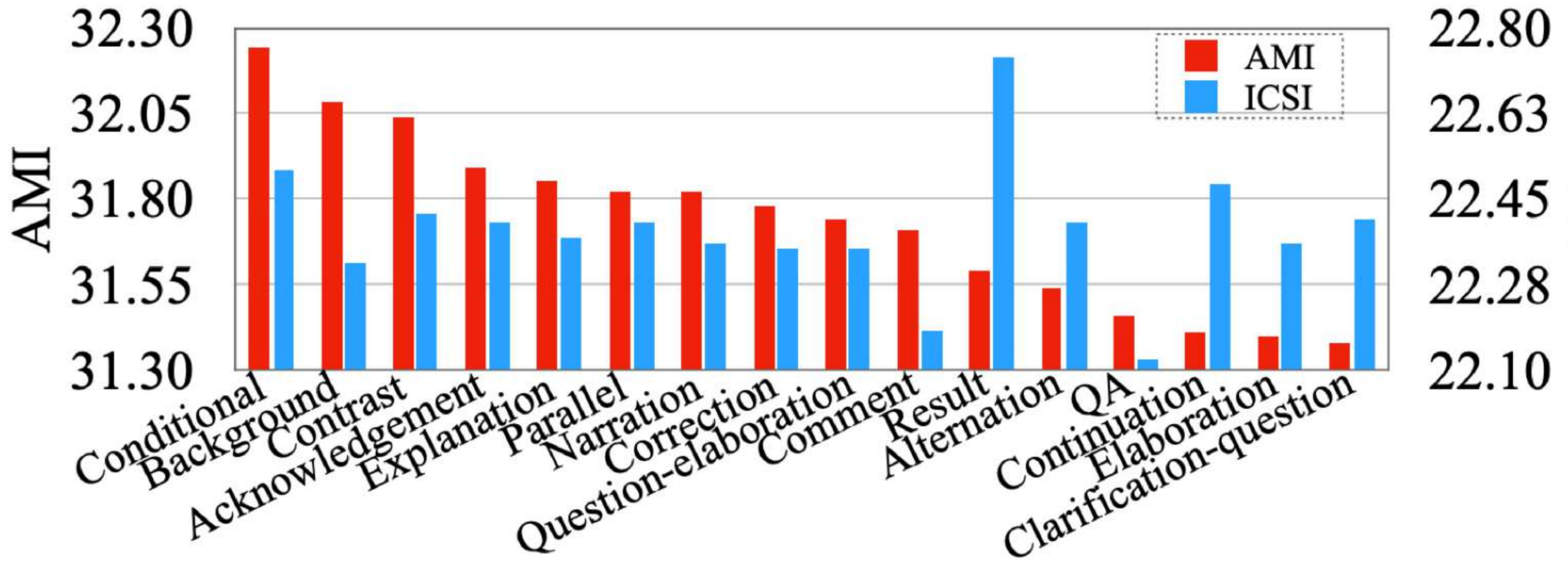


- Effect of the quality of dialogue discourse.



Analyses

- Effect of the type of dialogue discourse.



Analyses

- Effect of the type of dialogue discourse.
  - filter out N useless relations.

	Model	R-1	R-2	R-L
AMI	DDAMS	51.42	**20.99**	**24.89**
	filter-useless-3	51.28	19.68	23.84
	filter-useless-5	**51.44**	20.26	24.11
ICSI	DDAMS	39.66	**10.09**	**17.53**
	filter-useless-3	**39.71**	9.64	17.46
	filter-useless-5	39.21	9.52	17.33

Analyses

- Effect of meeting graph
  - taking the type of vertices into consideration, our model DDAMS can get better results.

	Model	R-1	R-2	R-L
AMI	DDAMS	51.42	**20.99**	**24.89**
	DDAMS (w/ Levi graph)	**51.46**	20.75	24.31
ICSI	DDAMS	**39.66**	**10.09**	**17.53**
	DDAMS (w/ Levi graph)	39.20	9.54	17.48

Analyses

- Effect of attention mechanisms.

	Model	R-1	R-2	R-L
AMI	DDAMS	**51.42**	**20.99**	24.89
	w/o *utter-attn*	51.22	20.57	**25.02**
	w/o *word-attn*	50.27	19.81	23.91
ICSI	DDAMS	**39.66**	**10.09**	**17.53**
	w/o *utter-attn*	39.59	9.90	17.24
	w/o *word-attn*	38.96	9.61	17.40

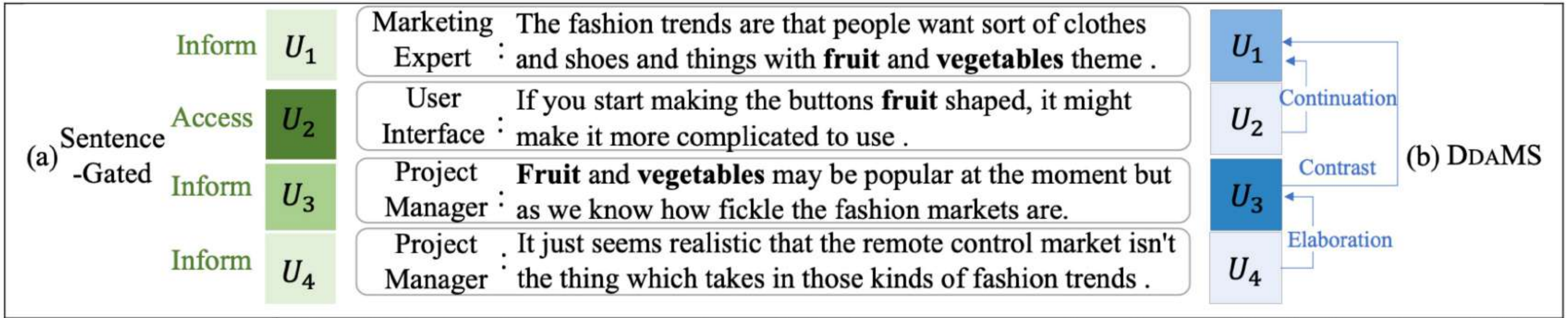
Analyses

- Effect of pseudo-summarization data.
  - pretraining on pseudo-summarization data constructed based on RBDA still achieves a better result, which indicates the rationality of our pretraining strategy.

	Model	R-1	R-2	R-L
AMI	DDAMS	51.42	20.99	24.89
	+ RBDA	52.94	21.96	25.05
	+ DDADA	**53.15**	**22.32**	**25.67**
ICSI	DDAMS	39.66	10.09	17.53
	+ RBDA	39.42	10.60	18.19
	+ DDADA	**40.41**	**11.02**	**19.18**

Case Study

- Utterance 1 and 3 are both related to two utterances, which make them the core nodes of our graph.



Ground-truth	The Marketing Expert presented trends in the remote control market and the **fruit** and **vegetable** and spongy material trends in fashion.
Pointer-Generator	They discussed the possibility of a **fruit or fruit and fruit**.
Sentence-Gated	The need to incorporate a **fruit** theme into the design of the remote.
DDAMS	The buttons will be included in a **fruit** and **vegetable** theme into the shape of the remote control.

Conclusion

- We make the first attempt to successfully explore dialogue discourse to model the utterances interactions for meeting summarization.
- We devise a dialogue discourse-aware data augmentation strategy to alleviate the data insufficiency problem.
- Extensive experiments show that our model achieves SOTA performance.

Thanks~