Language Model as an Annotator: Exploring DialoGPT for Dialogue Summarization

ACL 2021

Xiachong Feng, Xiaocheng Feng*, Libo Qin, Bing Qin, Ting Liu

Harbin Institute of Technology, China

❖ Introduction

- **Dialogue Summarization**
  - usually encode the text with semantic features.
- **Problem**
  - obtained via open-domain toolkits or relied on human annotations.
- **Solution**
  - view pre-trained DialoGPT as an unsupervised dialogue annotator to label three features.

Dialogue
Blair: Remember we are seeing the wedding planner after work
Chuck: Sure, where are we meeting her?
Blair: At Nonna Rita's
Chuck: I want to order seafood tagliatelle
Blair: Haha why not
Chuck: We remmber spaghetti pomodoro disaster from our last meeting
Blair: Omg it was over her white blouse
Chuck: I'll make time for it
Blair: Great!

(a) Keywords Extraction

Dialogue
Blair: Remember we are seeing the wedding planner after work
Chuck: Sure, where are we meeting her?
Blair: At Nonna Rita's
Chuck: I want to order seafood tagliatelle
Blair: *Haha why not*
Chuck: We remmber spaghetti pomodoro disaster from our last meeting
Blair: Omg it was over her white blouse
Chuck: *I'll make time for it*
Blair: *Great!*

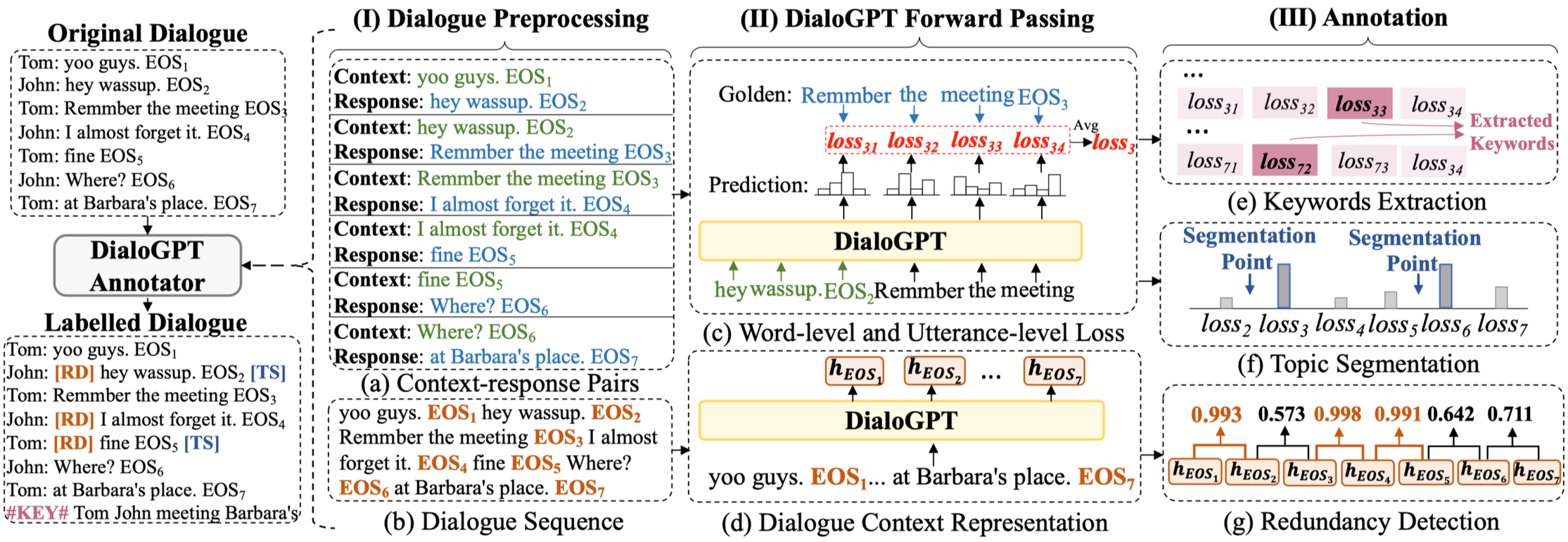
(b) Redundancy Detection

Dialogue
Blair: Remember we are seeing the wedding planner after work
Chuck: Sure, where are we meeting her?
Blair: At Nonna Rita's [Topic 1]
Chuck: I want to order seafood tagliatelle
Blair: Haha why not
Chuck: We remmber spaghetti pomodoro disaster from our last meeting [Topic 2]
Blair: Omg it was over her white blouse
Chuck: I'll make time for it [Topic 3]
Blair: Great!

(c) Topic Segmentation

Summary
Blair and Chuck are going to meet the wedding planner after work at Nonna Rita's. The tagliatelle served at Nonna Rita's are very good.
[Topic 1] [Topic 2]

❖ Method



❖ Experiments

- **We experiment on SAMSum and AMI datasets.**

Code:

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($\mathcal{D}_{\text{KE}}$)	**53.43**††	**28.03**††	49.93
BART($\mathcal{D}_{\text{RD}}$)	53.39	28.01	49.49
BART($\mathcal{D}_{\text{TS}}$)	53.34	27.85	49.64
BART($\mathcal{D}_{\text{ALL}}$)	**53.70**†	**28.79**†	**50.81**†

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

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($\mathcal{D}_{\text{KE}}$)	50.22	17.74	24.11
PGN($\mathcal{D}_{\text{RD}}$)	50.62	16.86	24.27
PGN($\mathcal{D}_{\text{TS}}$)	48.59	16.07	24.05
PGN($\mathcal{D}_{\text{ALL}}$)	50.91	**17.75**††	**24.59**††

- **Intrinsic evaluation for keywords.**

Method	Precision	Recall	F_1
TextRank	47.74%	17.44%	23.22%
Entities	**60.42%**	17.80%	25.38%
DialoGPT$_{\text{KE}}$	33.20%	**29.49%**	**30.31%**

- **Extrinsic evaluation for redundancy.**

Model	R-1	R-2	R-L
SAMSum			
Rule-based	53.00	27.71	**49.68**
DialoGPT$_{\text{RD}}$	**53.39**	**28.01**	49.49
AMI			
Rule-based	50.19	16.45	23.95
DialoGPT$_{\text{RD}}$	**50.62**	**16.86**	**24.27**

- **Extrinsic evaluation for topic.**

Model	R-1	R-2	R-L
SAMSum			
C99			
w/ BERT emb	52.80	27.78	49.50
w/ DialoGPT emb	53.33	**28.04**	49.39
DialoGPT$_{\text{TS}}$	**53.34**	27.85	**49.64**
AMI			
Golden	50.28	19.73	24.45
C99			
w/ BERT emb	48.53	15.84	23.63
w/ DialoGPT emb	**49.22**	**16.79**	23.88
DialoGPT$_{\text{TS}}$	48.59	16.07	**24.05**

Email: xiachongfeng@ir.hit.edu.cn
Homepage: http://xcfeng.net/