GIFT: Graph-Induced Fine-Tuning for Multi-Party Conversation Understanding

Jia-Chen Gu[1], Zhen-Hua Ling[1], Quan Liu[2,3], Cong Liu[1,3], Guoping Hu[2,3]

[1]National Engineering Research Center of Speech and Language Information Processing, University of Science and Technology of China, Hefei, China

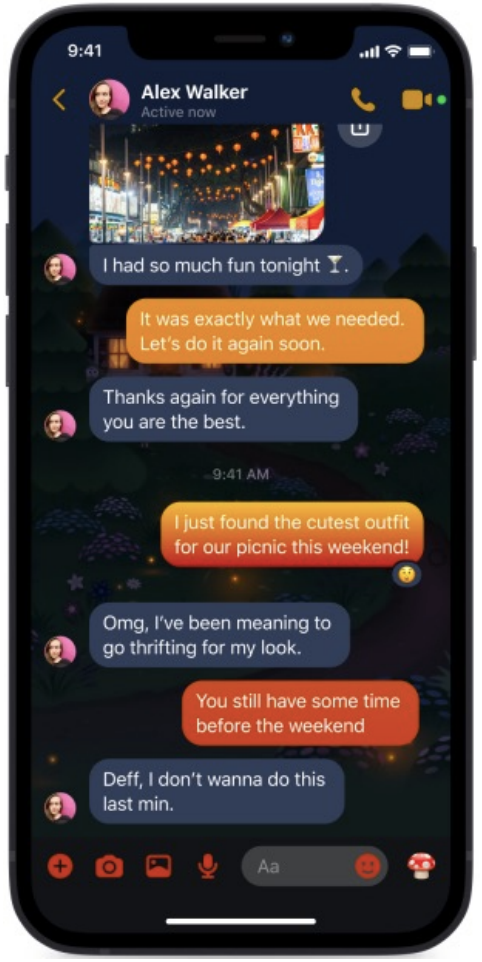
[2]State Key Laboratory of Cognitive Intelligence

[3]iFLYTEK Research, Hefei, China

Outline

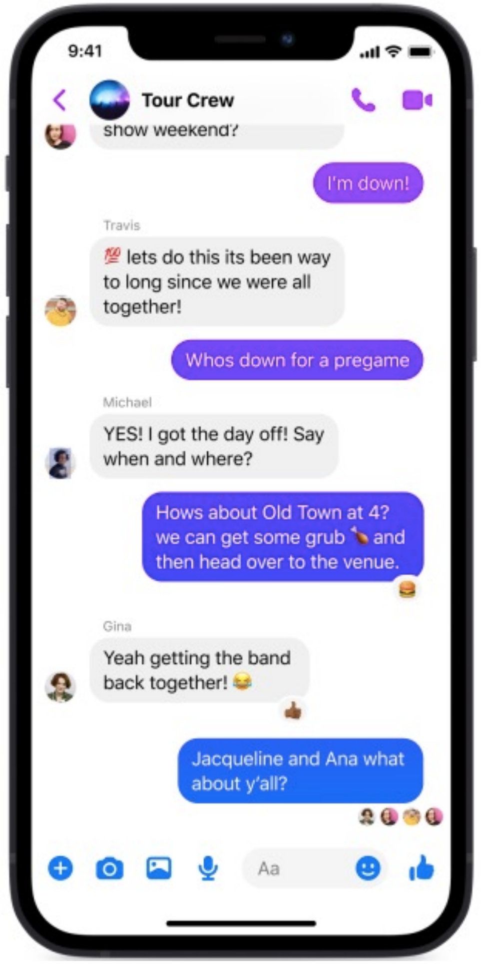
- **Introduction**
- Graph-Induced Fine-Tuning (GIFT)
- Experiments
- Conclusion

Two-Party VS. Multi-Party Conversations



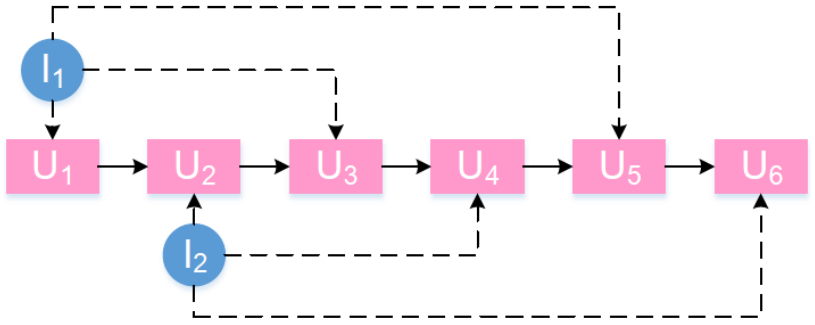
One-on-One Chat

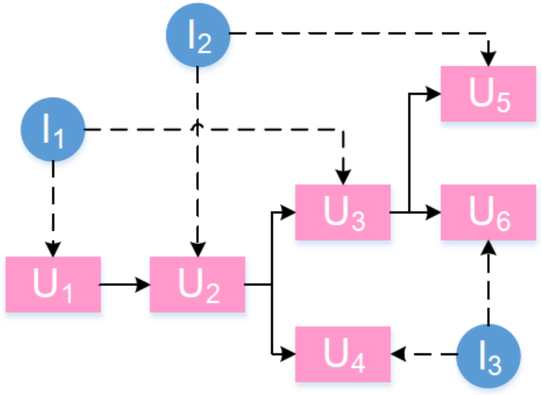
Group chats appear frequently in daily life!



Group Chat

Graphical Multi-Party Conversations





Utterances in a two-party conversation are posted one by one between two interlocutors, constituting a sequential information flow.

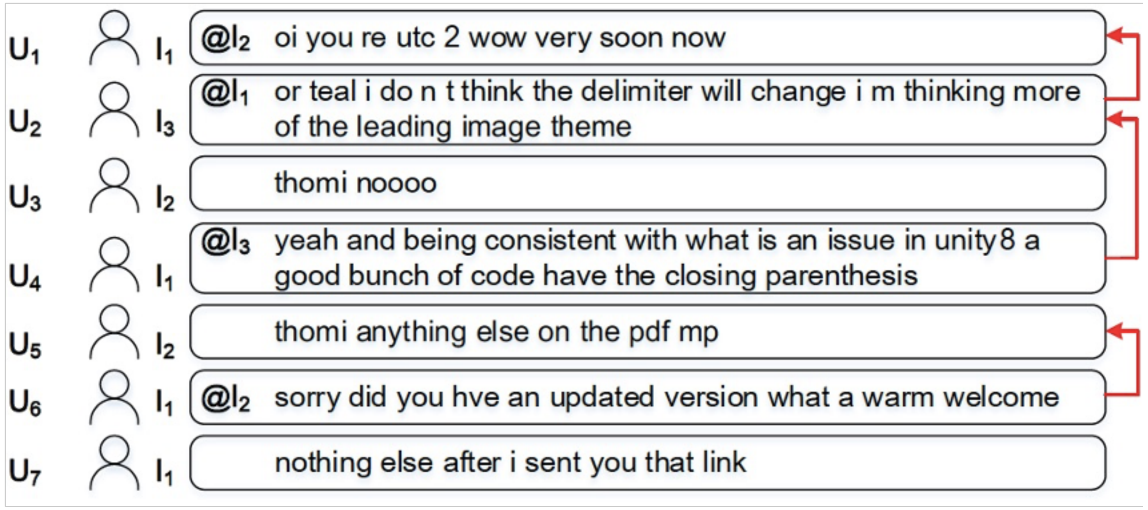
Utterances in a multi-party conversation (MPC) can be spoken by anyone and address anyone else, constituting a graphical information flow.

: Interlocutors

: Utterances

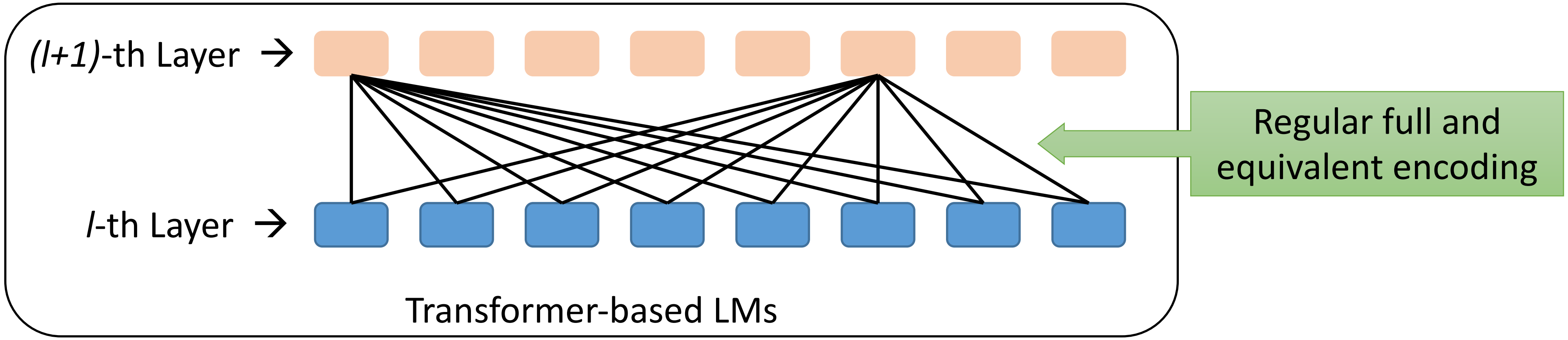
MPC Example

- Reply relationships can be constructed based on “@” labels



Regular Transformer Encoding

- The full and equivalent connections among utterance tokens ignore the sparse but distinctive dependency of one utterance on another
- Overlook the inherent MPC graph structure on various downstream tasks



Outline

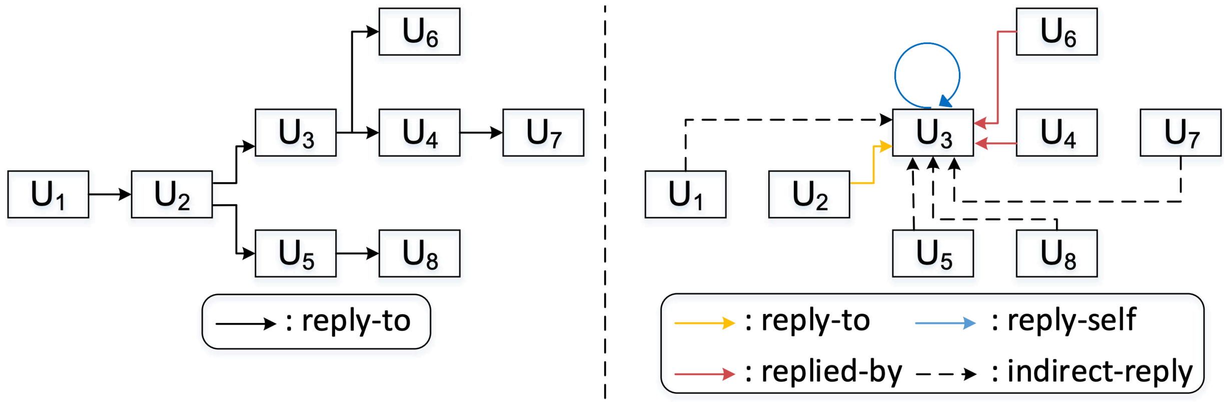
- Introduction
- **Graph-Induced Fine-Tuning (GIFT)**
- Experiments
- Conclusion

Ubiquitous Graph Data Structure

- Hu et al. (2019) and Gu et al. (2022) have indicated that the complicated graph structures can provide crucial interlocutor and utterance semantics
- We are inspired to
  - ✓ view an MPC as a conversation graph where features can be represented by considering available explicit connectivity structures (i.e., graph structures)
  - ✓ refine Transformer-based LMs by modeling graph structures during internal encoding to help establish the sparse but distinctive dependency of an utterance on another

MPC Graph Topology

- Four types of edges (*reply-to, replied-by, reply-self and indirect-reply*) are designed to distinguish different relationships between utterances



(a) A Graphical Information Flow of an MPC

(b) Reply Relationships in a Graph Structure for U_3

* Rectangles (U) denote utterances, and solid lines (⟶) represent the “reply" relationship between two utterances

Graph-Induced Signals Integration

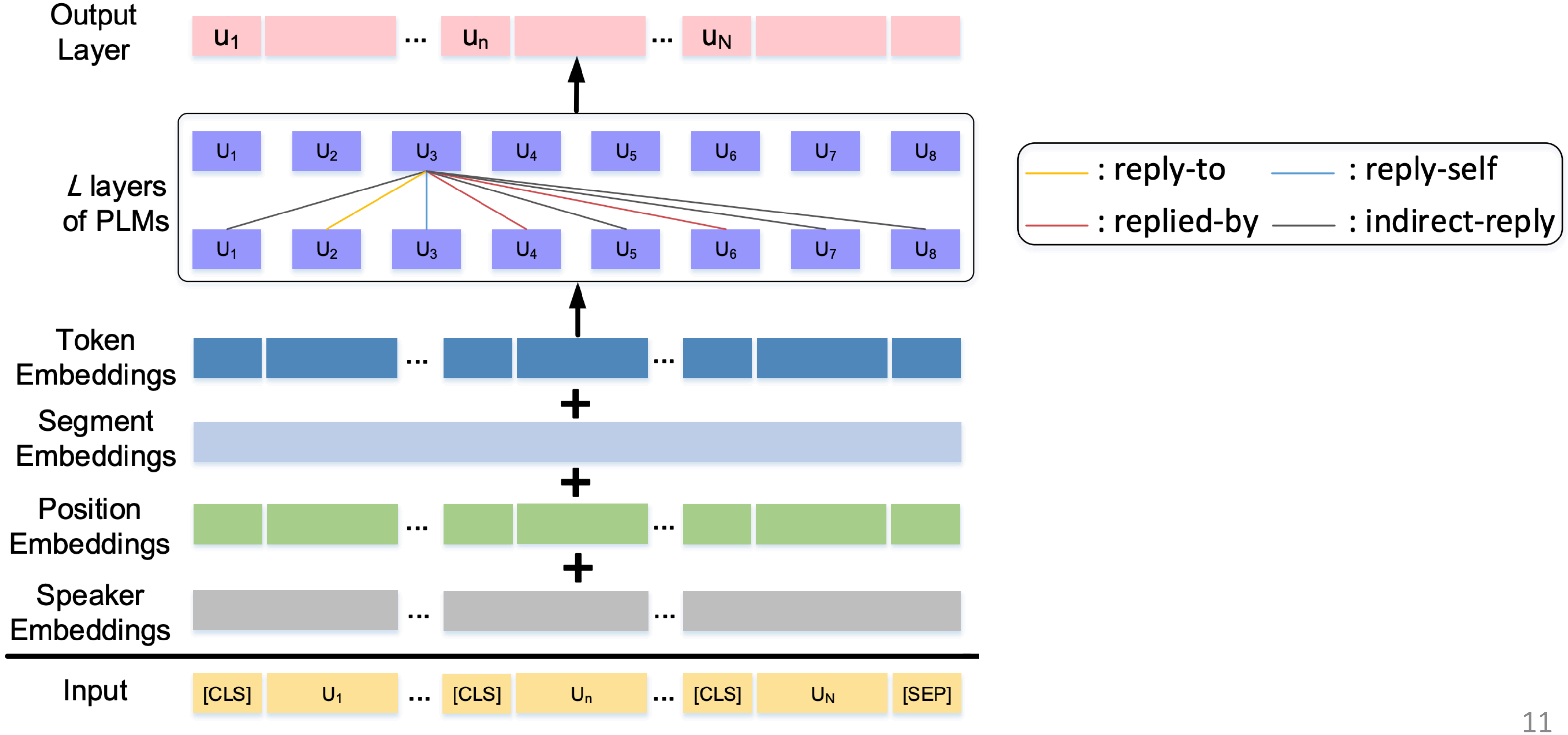
- Integrated in the attention mechanism by utilizing edge-type-dependent parameters to refine the attention weights

$$\mathrm{Atten}(q, k, v) = \mathrm{softmax}(\phi(e_{q,v})\frac{\mathbf{q}^\top \mathbf{k}}{\sqrt{d}})\mathbf{v}$$

  where $e_{q,v} \in$ {*reply-to, replied-by, reply-self, indirect-reply*}
- *reply-to*: what the current utterance should be like given the prior utterance it replies to
- *replied-by*: how the posterior utterances amend the modeling of the current utterance
- *reply-self*: how much of the original semantics should be kept
- *indirect-reply*: connect the rest of the utterances for contextualization

Model Overview

- Input data following MPC-BERT that (1) inserts [CLS] tokens at the start of each utterance, and (2) introduces position-based speaker embeddings to distinguish the speakers of utterances



Why These Edges Work?

- Consider both semantic similarity and structural relationships between two utterance tokens
- Distinguish different relationships between utterances, and model utterance dependency following the graph-induced topology for better contextualized encoding
- Characterize fine-grained interactions during LM internal encoding
- Reflect graphical conversation structure and flow in Transformer

Outline

- Introduction
- Graph-Induced Fine-Tuning (GIFT)
- **Experiments**
- Conclusion

Downstream Tasks

- **Addressee Recognition**: to recognize the addressees of the last utterances from the set of all interlocutors that appear in this conversation

- **Speaker Identification**: to identify the speaker of the last utterance in a conversation from the interlocutor set

- **Response Selection**: to measure the similarity between the given context and a response candidate, and then rank a set of response candidates

Setup

- Datasets

  We evaluated the proposed method on two Ubuntu IRC benchmarks

Datasets		Train	Valid	Test
Hu et al. (2019)		311,725	5,000	5,000
Ouchi and Tsuboi (2016)	Len-5	461,120	28,570	32,668
	Len-10	495,226	30,974	35,638
	Len-15	489,812	30,815	35,385

- Baselines

  GIFT was implemented into three Transformer-based PLMs including BERT, SA-BERT and MPC-BERT, which is plug-and-play

Results: Addressee Recognition

- GIFT improves the performance of BERT by margins of 2.92%, 2.73%, 5.75% and 5.08% on these test sets respectively in terms of Precision (P@1)

improves SA-BERT by margins of 1.32%, 2.50%, 4.26% and 5.22% respectively

	Hu et al. (2019)	Ouchi and Tsuboi (2016)		
		Len-5	Len-10	Len-15
Preceding (Le et al., 2019)	-	55.73	55.63	55.62
SRNN (Ouchi and Tsuboi, 2016)	-	60.26	60.66	60.98
SHRNN (Serban et al., 2016)	-	62.24	64.86	65.89
DRNN (Ouchi and Tsuboi, 2016)	-	63.28	66.70	68.41
SIRNN (Zhang et al., 2018)	-	72.59	77.13	78.53.
BERT (Devlin et al., 2019)	82.88	80.22	75.32	74.03
SA-BERT (Gu et al., 2020)	86.98	81.99	78.27	76.84
MPC-BERT (Gu et al., 2021)	89.54	84.21	80.67	78.98
BERT w/ GIFT	85.80†	82.95†	81.07†	79.11†
SA-BERT w/ GIFT	88.30†	84.49†	82.53†	82.06†
MPC-BERT w/ GIFT	**90.18**	**85.85**†	**84.13**†	**83.61**†

improves MPC-BERT by margins of 0.64%, 1.64%, 3.46% and 4.63% respectively

Table 1: Evaluation results of addressee recognition on the test sets in terms of P@1. Results except ours are cited from Ouchi and Tsuboi (2016) and Zhang et al. (2018). Numbers marked with † denoted that the improvements after implementing GIFT were statistically significant (t-test with p-value < 0.05) comparing with the corresponding PLMs. Numbers in bold denoted that the results achieved the best performance.

Results: Speaker Identification

- GIFT improves the performance of BERT by margins of 13.71%, 27.50%, 29.14% and 28.82% on these test sets respectively in terms of P@1

improves SA-BERT by margins of 12.14%, 25.05%, 25.14% and 26.59% respectively

	Hu et al. (2019)	Ouchi and Tsuboi (2016)		
		Len-5	Len-10	Len-15
BERT	71.81	62.24	53.17	51.58
SA-BERT	75.88	64.96	57.62	54.28
MPC-BERT	83.54	67.56	61.00	58.52
BERT w/ GIFT	85.52†	89.74†	82.31†	80.40†
SA-BERT w/ GIFT	88.02†	90.01†	82.76†	80.87†
MPC-BERT w/ GIFT	**90.50**†	**90.61**†	**84.12**†	**81.51**†

Table 2: Evaluation results of speaker identification on the test sets in terms of P@1. Results except ours are cited from Gu et al. (2021).

improves MPC-BERT by margins of 6.96%, 23.05%, 23.12% and 22.99% respectively

Results: Response Selection

- GIFT improves the performance of BERT by margins of 2.48%, 2.12%, 2.71% and 2.34%, of SA-BERT by margins of 3.04%, 4.16%, 5.18% and 5.35%, and of MPC-BERT by margins of 1.76%, 0.88%, 2.15% and 2.44% on these test sets respectively in terms of Recall ($R_{10}@1$)

	Hu et al. (2019)		Ouchi and Tsuboi (2016)					
			Len-5		Len-10		Len-15	
	$R_2@1$	$R_{10}@1$	$R_2@1$	$R_{10}@1$	$R_2@1$	$R_{10}@1$	$R_2@1$	$R_{10}@1$
DRNN (Ouchi and Tsuboi, 2016)	-	-	76.07	33.62	78.16	36.14	78.64	36.93
SIRNN (Zhang et al., 2018)	-	-	78.14	36.45	80.34	39.20	80.91	40.83
BERT (Devlin et al., 2019)	92.48	73.42	85.52	53.95	86.93	57.41	87.19	58.92
SA-BERT (Gu et al., 2020)	92.98	75.16	86.53	55.24	87.98	59.27	88.34	60.42
MPC-BERT (Gu et al., 2021)	94.90	78.98	87.63	57.95	89.14	61.82	89.70	63.64
BERT w/ GIFT	93.22†	75.90†	86.59†	56.07†	88.02†	60.12†	88.57†	61.26†
SA-BERT w/ GIFT	94.26†	78.20†	**88.07**†	**59.40**†	**89.91**†	**64.45**†	90.45†	65.77†
MPC-BERT w/ GIFT	**95.04**	**80.74**†	87.97	58.83†	89.77†	63.97†	**90.62**†	**66.08**†

Table 3: Evaluation results of response selection on the test sets. Results except ours are cited from Ouchi and Tsuboi (2016), Zhang et al. (2018) and Gu et al. (2021).

Ablation

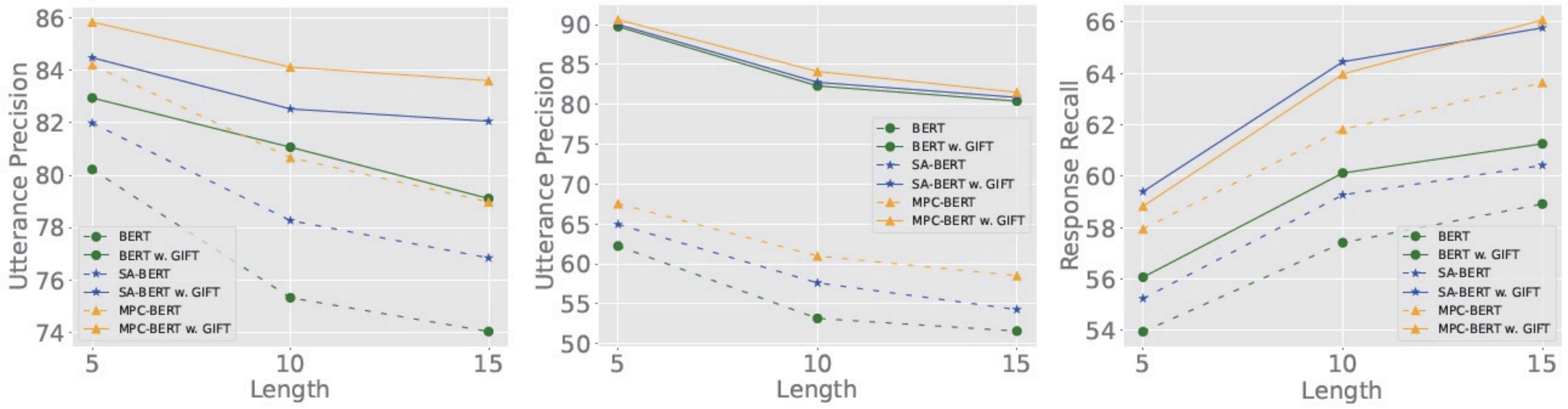
- Merge reply-to and replied-by edges with in-direct edges
- Merge reply-to or replied-by edges together without distinguishing bidirectionality
- Merge reply-self with in-direct edges with in-direct edges

	AR (P@1)	SI (P@1)	RS (R_{10}@1)
BERT w/ GIFT	86.24	86.50	75.26
w/o reply-to and replied-by	84.38	70.67	72.30
w/o reply-to or replied-by	85.72	85.67	74.00
w/o reply-self	85.72	85.92	74.72
SA-BERT w/ GIFT	88.88	89.32	78.80
w/o reply-to and replied-by	86.90	77.07	77.50
w/o reply-to or replied-by	88.44	88.87	78.22
w/o reply-self	88.42	89.05	78.32
MPC-BERT w/ GIFT	90.78	91.72	81.08
w/o reply-to and replied-by	90.38	84.32	79.60
w/o reply-to or replied-by	90.52	90.90	80.22
w/o reply-self	90.46	91.10	80.02

Table 5: Evaluation results of the ablation tests on the validation set of Hu et al. (2019) on the tasks of addressee recognition (AR), speaker identification (SI), and response selection (RS).

Performance Change at Different Lengths

As the session length increased, the performance of models with GIFT dropped more slightly on addressee recognition and speaker identification, and enlarged more on response selection, than the models without GIFT in most 14 out of 18 cases



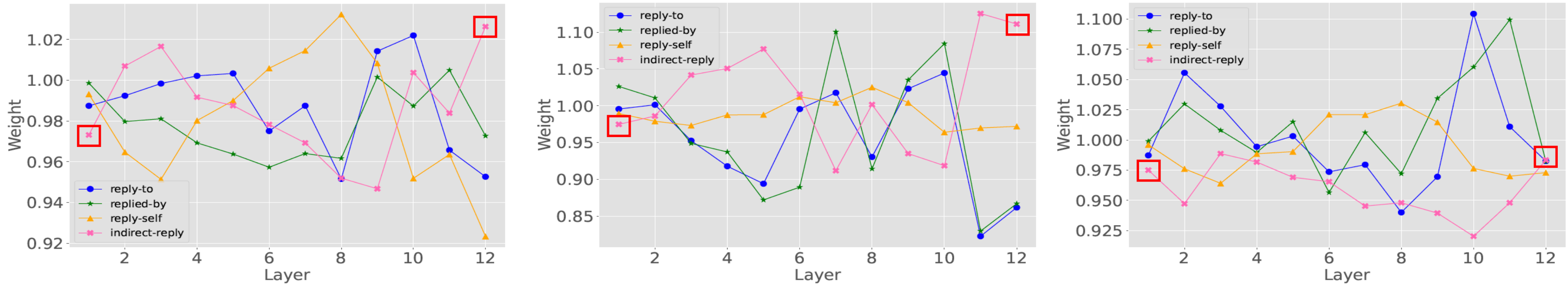
(a) Addressee Recognition (b) Speaker Identification (c) Response Selection

	Len 5 → Len 10	Len 10 → Len 15
	AR (P@1)	
BERT	-4.90	-1.29
BERT w. GIFT	-1.88‡	-1.96
SA-BERT	-3.72	-1.43
SA-BERT w. GIFT	-1.96‡	-0.47‡
MPC-BERT	-3.54	-1.69
MPC-BERT w. GIFT	-1.72‡	-0.52‡
	SI (P@1)	
BERT	-9.07	-1.59
BERT w. GIFT	-7.43‡	-1.91
SA-BERT	-7.34	-3.34
SA-BERT w. GIFT	-7.25‡	-1.89‡
MPC-BERT	-6.56	-2.48
MPC-BERT w. GIFT	-6.49‡	-2.61
	RS (R_{10}@1)	
BERT	+3.46	+1.51
BERT w. GIFT	+4.05‡	+1.14
SA-BERT	+4.03	+1.15
SA-BERT w. GIFT	+5.05‡	+1.32‡
MPC-BERT	+3.87	+1.82
MPC-BERT w. GIFT	+5.14‡	+2.11‡

Table 6: Performance change of models as the session length increased on the test sets of Ouchi and Tsuboi (2016). For models with GIFT, numbers marked with ‡ denoted larger performance improvement or less performance drop compared with the corresponding models without GIFT.

Visualization of Weights

- The changing trends of reply-to and replied-by edges were roughly the same, while the values of these two edges were always different
- The values of the indirect-reply edge were always the minimum at the beginning, and surprisingly became the maximum in the last layer:
    - ✓ less attention to irrelevant utterances to themselves at first glance
    - ✓ after comprehending the most relevant utterances, turn to indirectly related ones in context for fully understanding the entire conversation



(a) Addressee Recognition (b) Speaker Identification (c) Response Selection

Figure 4: The weights of four types of edges in different encoding layers of MPC-BERT trained on Hu et al. (2019).

Outline

- Introduction
- Graph-Induced Fine-Tuning (GIFT)
- Experiments
- **Conclusion**

Conclusion

- We present graph-induced fine-tuning (GIFT) for multi-party conversation understanding, which is
  - ✓plug-and-play: adapt various Transformer-based LMs, e.g., BERT, SA-BERT and MPC-BERT
  - ✓lightweight: add only 4 additional parameters per encoding layer
  - ✓universal: show effectiveness on 3 downstream tasks, e.g., addressee recognition, speaker identification and response selection
- Experimental results on three downstream tasks show that GIFT significantly helps improve the performance of three PLMs and achieves new state-of-the-art performance on two benchmarks

Challenges

- Reduce the heavy dependency on the necessary addressee labels, while the scarcity of addressee labels is a common issue in MPCs (55% missing in Ubuntu)
- Extend to multi-modal MPCs, including face and speech interactions
- Data-centric dataset construction for MPCs

Jia-Chen Gu

Zhen-Hua Ling

Quan Liu

Cong Liu

Guoping Hu

Thanks! Q&A

Contact: gujc@ustc.edu.cn

Homepage: http://home.ustc.edu.cn/~gujc

Code: https://github.com/JasonForJoy/MPC-BERT