Multi-Party Conversational AI

Jia-Chen Gu

Department of Computer Science

University of California, Los Angeles

Presenter

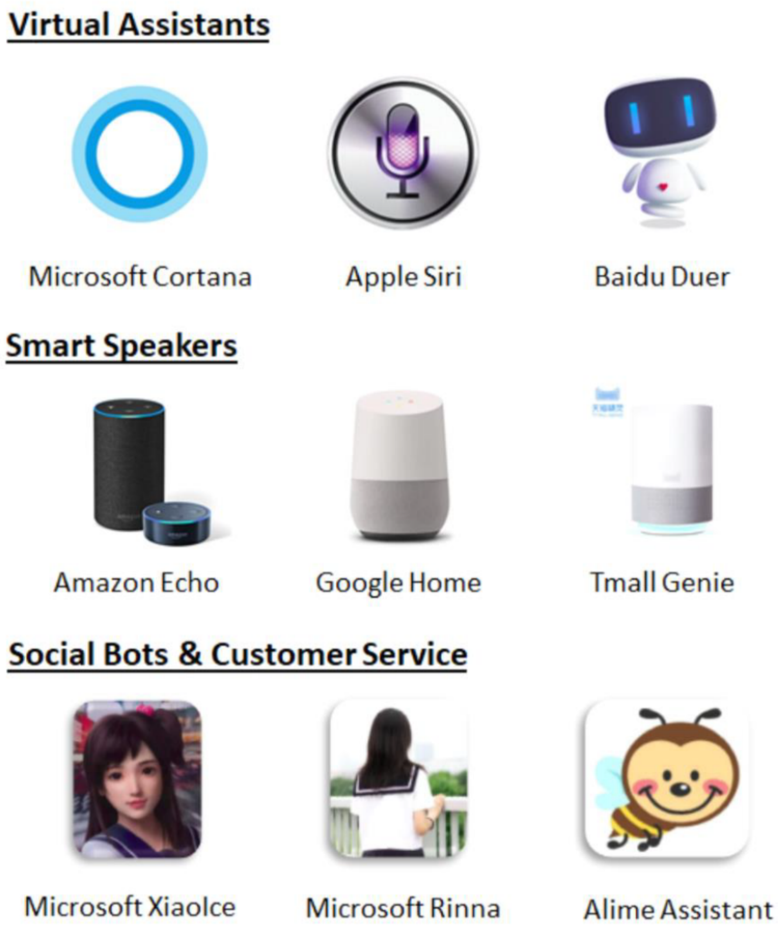
Jia-Chen Gu
Postdoc@UCLA

- 2023 - Best Paper Honorable Mention Award of ACL 2023 (First-author)
- 2022 - Best Paper Award of ACL 2022 DialDoc Workshop (Second-author)
- 2022 - Outstanding Doctoral Dissertation Nomination Award of CIPS
- 2022 - Presidential Scholarship of Chinese Academy of Sciences (Top 1%)
- 2021 - China National Scholarship
- Research intern @Microsoft (2020-2021)
  & Visiting student @Queen's University (2019-2020)

Dialogue Systems are “Hot”

Academia

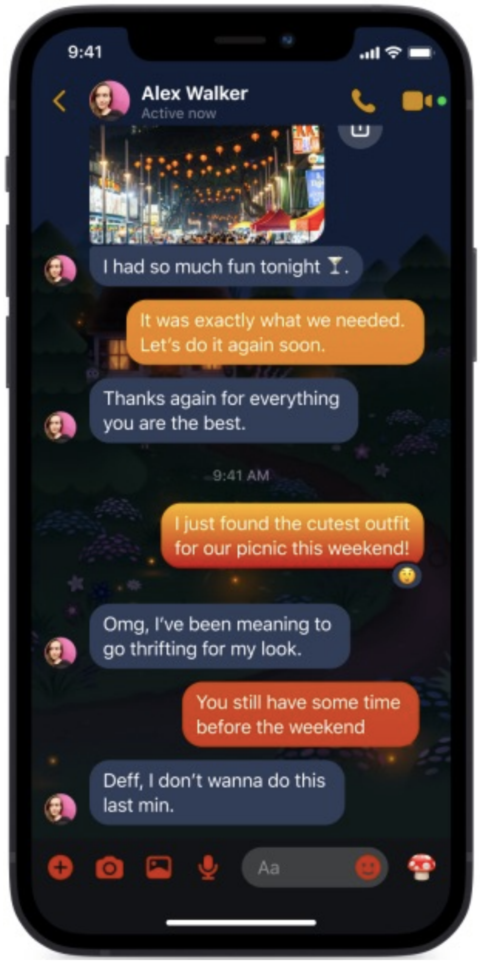
Industry



Chat GPT

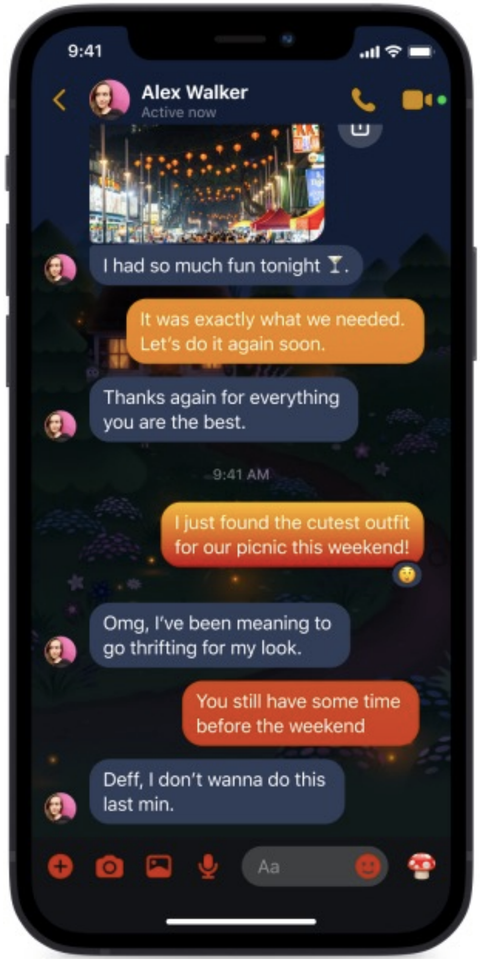
OpenAI

Two-Party Conversations



One-on-one chat
between 2 interlocutors

Two-Party VS. Multi-Party Conversations

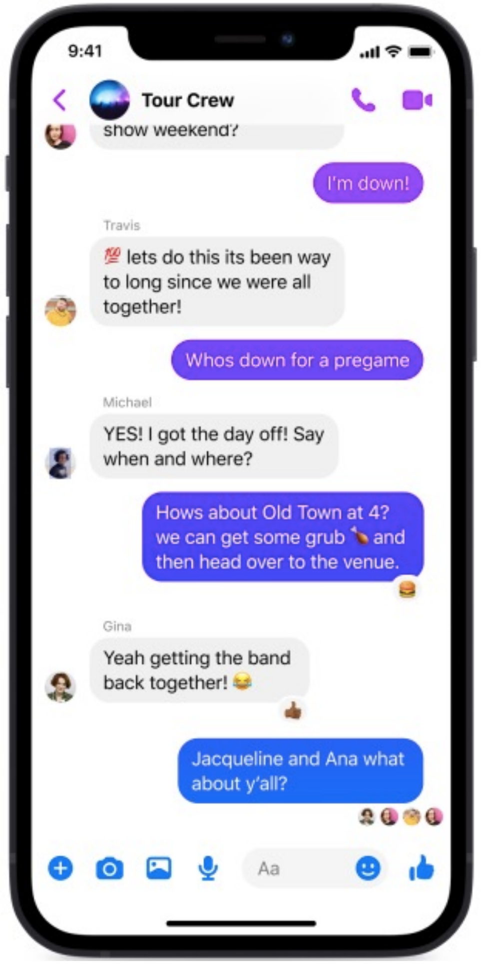


One-on-one chat

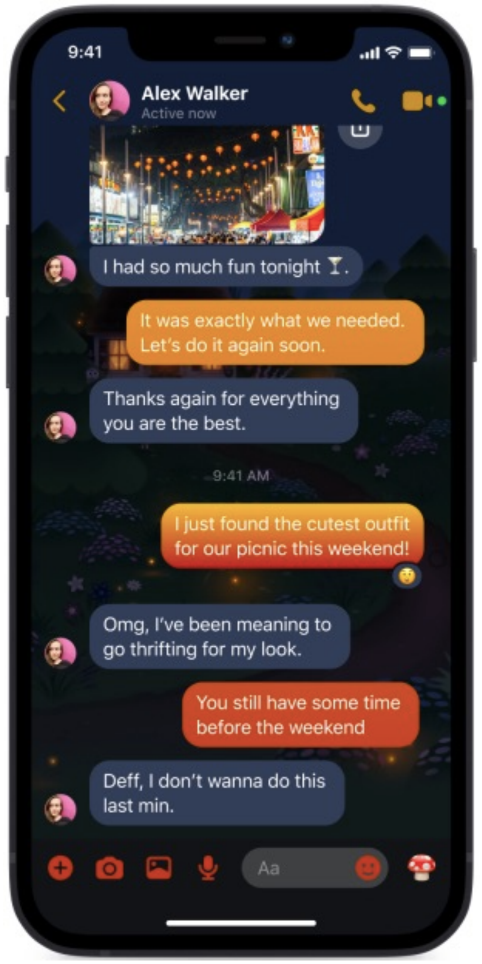
between 2 interlocutors

Group chat

involving 3+ interlocutors



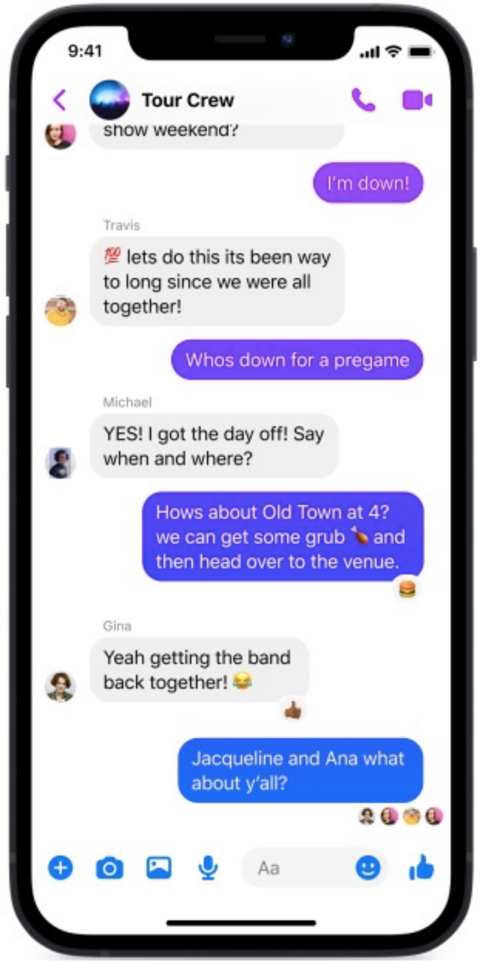
Two-Party VS. Multi-Party Conversations



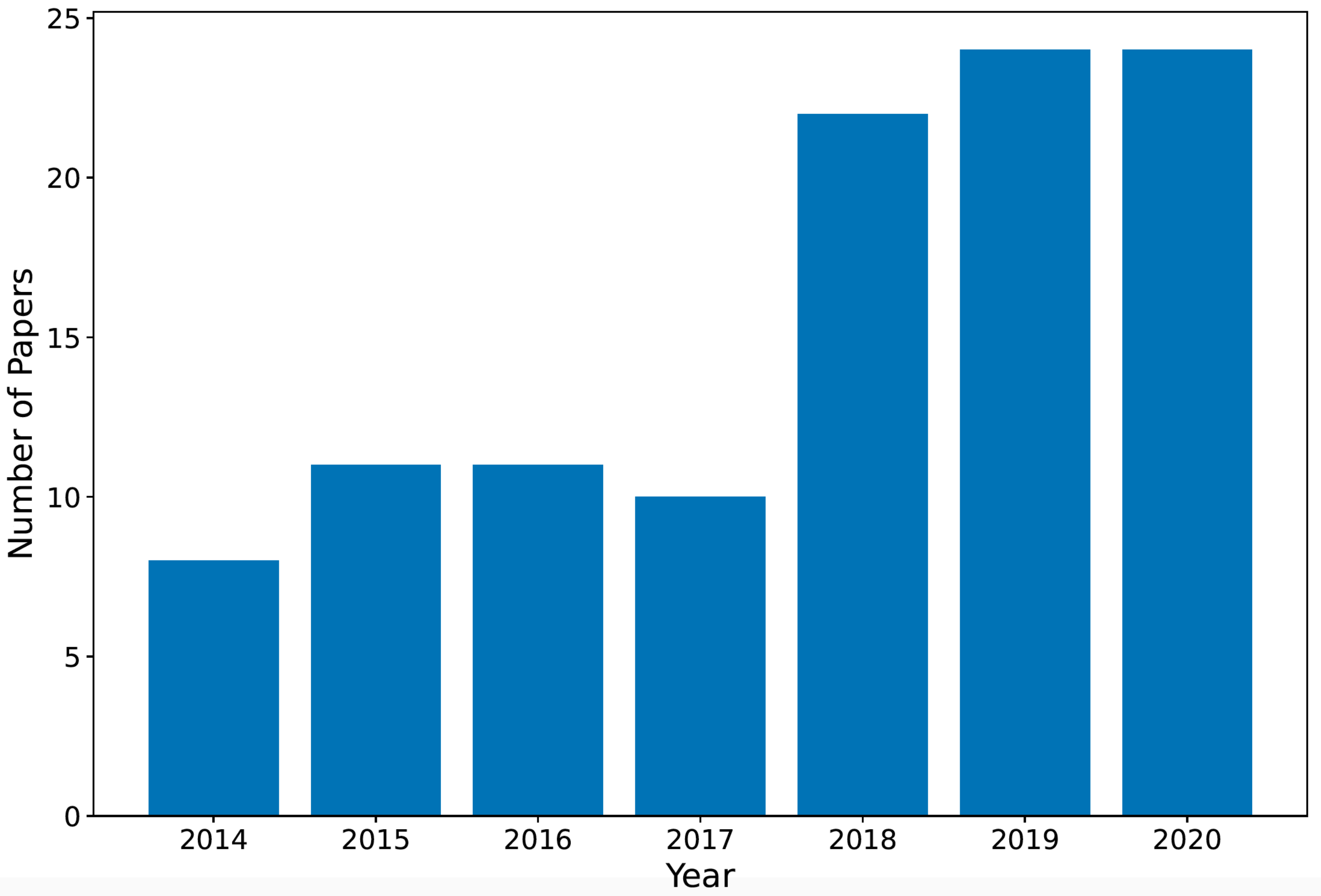
One-on-one chat
between 2 interlocutors

Group chats appear
frequently in daily life!

Group chat
involving 3+ interlocutors



Research Trend on Multi-Party Conversation

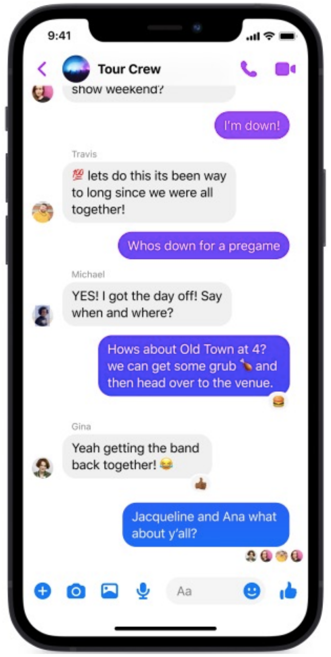


Include the keywords multi-party conversation and its variations, considering papers that appeared at CL and AI venues

Data:https://github.com/khyatimahajan/mpd-references

Why multi-party conversations (MPC)?

Many scenarios involve MPC and require capabilities beyond two-party conversations, e.g., turn-taking, discourse parsing and disentanglement

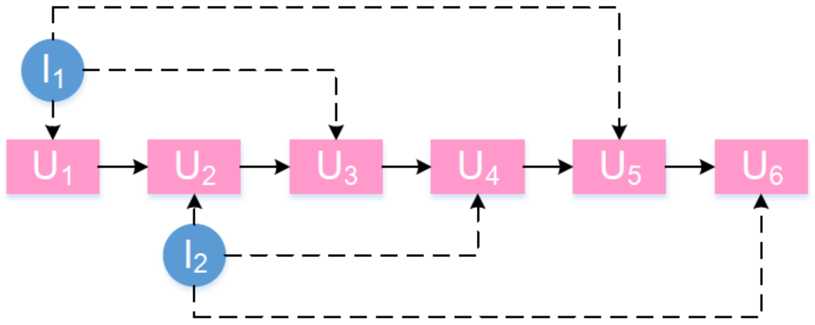


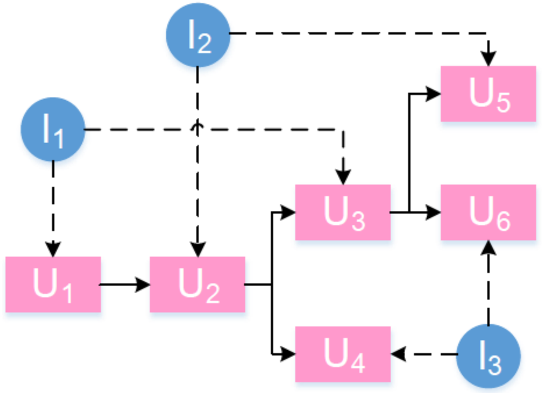
Group Chat

Meeting

Agent Simulacra

Graphical MPC is complicated





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

Challenges (1): WHO speaks

Model the coordination strategies that speakers adopt to acquire or give up the floor, so that an ongoing conversation can go on smoothly (Hawes et al., 2009; Pinhanez et al., 2018; de Bayser et al., 2019)

Speaker	Addressee	Utterance
User 1	-	I have a problem when I install …
Agent	User 1	Did you set initial params?
User 2	User 1	Show the error message, and …
User 1	Agent	How?
User 1	User 2	OK, just a moment!
[Who speak?]		

Should the agent take
the floor to speak or not?

Challenges (2): address WHOM

Understand conversation semantics for the behavior whereby interlocutors indicate to whom they are speaking (Ouchi and Tsuboi, 2016; Le et al., 2019; Gu et al., 2021; Zhu et al., 2023)

Speaker	Addressee	Utterance
User 1	-	I have a problem when I install …
Agent	User 1	Did you set initial params?
User 2	User 1	Show the error message, and …
User 1	Agent	How?
User 1	User 2	OK, just a moment!
Agent	[To whom?]	

User 1?
or
User 2?

Challenges (3): say WHAT

Return an appropriate response which follows the conversation semantics, structures and topic transitions (Zhang et al., 2018; Wu et al., 2020; Wang et al., 2020; Gu et al., 2022; Li et al., 2023)

Speaker	Addressee	Utterance
User 1	-	I have a problem when I install …
Agent	User 1	Did you set initial params?
User 2	User 1	Show the error message, and …
User 1	Agent	How?
User 1	User 2	OK, just a moment!
Agent	User 1	[Say what?]

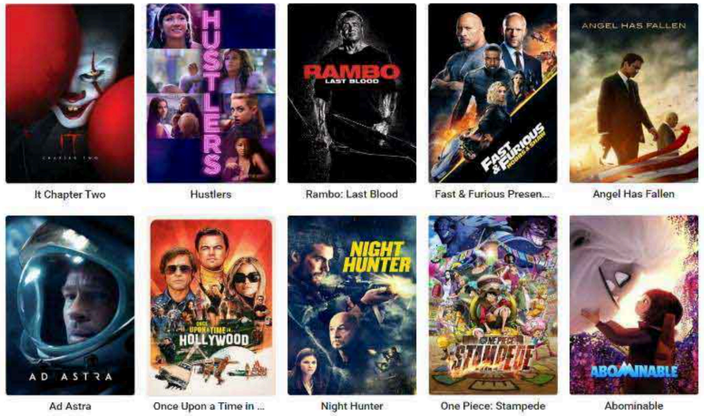
See this URL: http://xxx
or
It’s already in OS

Datasets

- Written corpora: online forums, such as Ubuntu, Reddit …

Datasets

- Written corpora: online forums, such as Ubuntu, Reddit …



- Spoken corpora:
  - ✓scripted refers to planned dialogue, such as TV and movie scripts
  - ✓Unscripted refers to spontaneous and unplanned dialogues, such as meeting

Motivations

Pre-train towards universal MPC understanding?

✓ **Jia-Chen Gu**, Chongyang Tao, Zhen-Hua Ling, et al. 2021. *MPC-BERT: A Pre-Trained Language Model for Multi-Party Conversation Understanding*. In *Proc.* ***ACL***.

Embrace various sources of information in a heterogeneous graph?

✓ **Jia-Chen Gu**, Chao-Hong Tan, et al. 2022. *HeterMPC: A Heterogeneous Graph Neural Network for Response Generation in Multi-Party Conversations*. In *Proc.* ***ACL***.

Introduce graphical structures into various Transformer-based LM encoding?

✓ **Jia-Chen Gu**, Zhen-Hua Ling, et al. 2023. *GIFT: Graph-Induced Fine-Tuning for Multi-Party Conversation Understanding*. In *Proc.* ***ACL***. **(Best Paper Honorable Mention Award)**

Mitigate the common scarcity of addressee labels in MPCs?

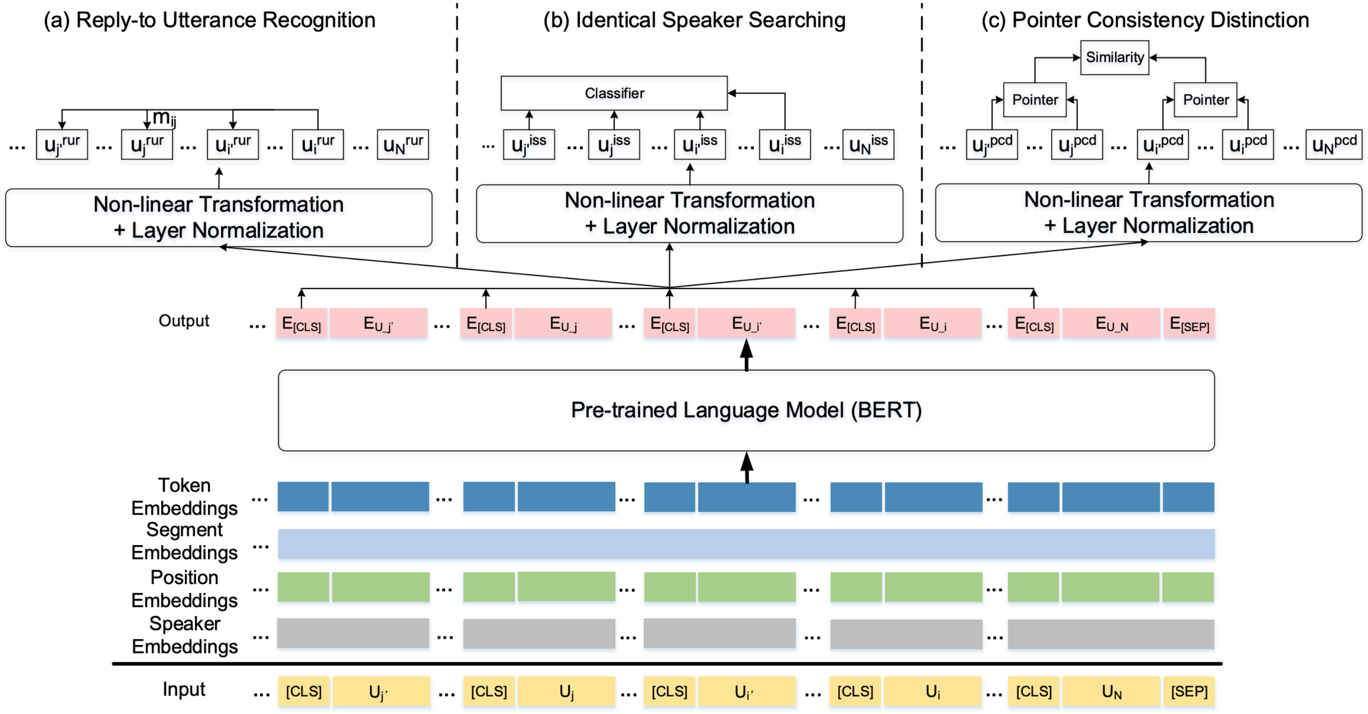
✓ **Jia-Chen Gu**, Chao-Hong Tan, et al. 2023. *MADNet: Maximizing Addressee Deduction Expectation for Multi-Party Conversation Generation*. In *Proc.* ***EMNLP***.

- **Jia-Chen Gu**, Chongyang Tao, Zhen-Hua Ling. 2022. *Who Says What to Whom: A Survey of Multi-Party Conversations*. In *Proc.* ***IJCAI***. (**Tutorial@AACL 2023**)

MPC-BERT for MPC Understanding

MPC-BERT

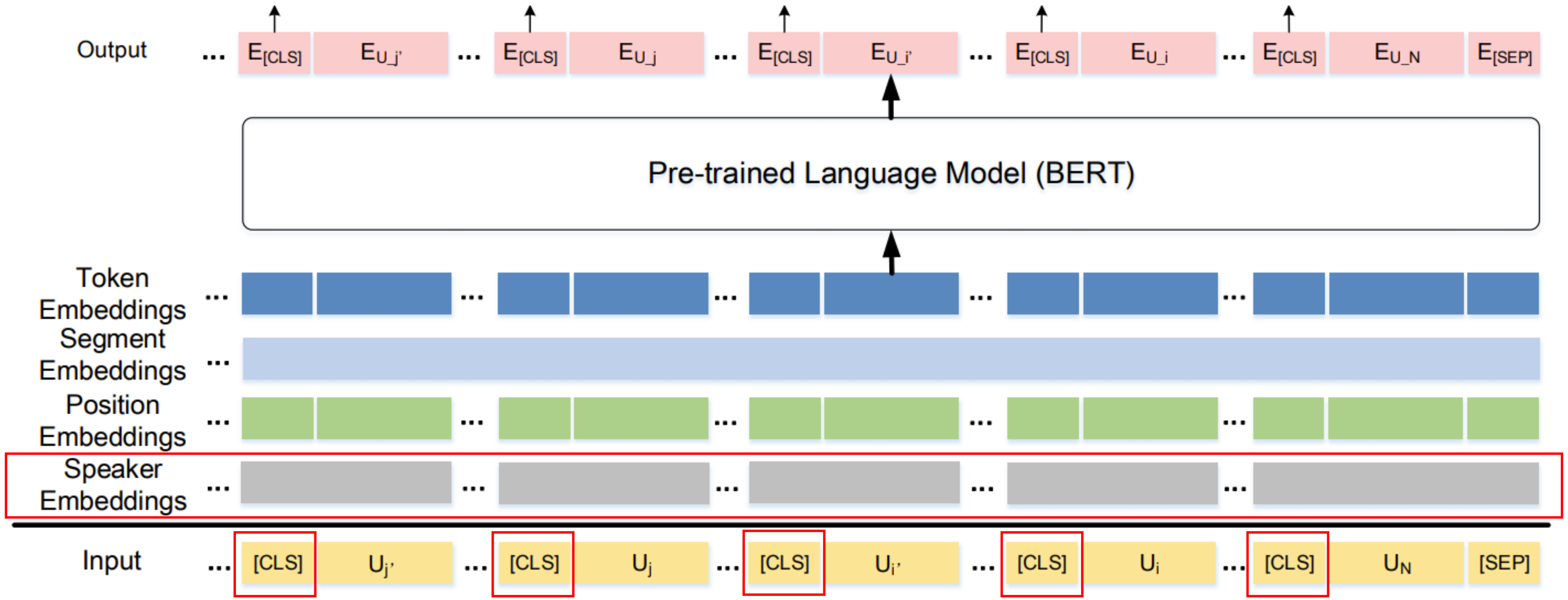
Pretrain BERT with five self-supervision tasks, designed to model the underlying interlocutor structure and utterance semantics, which can be further effectively generalized to multiple MPC downstream tasks



Jia-Chen Gu, et al. *MPC-BERT: A Pre-Trained Language Model for Multi-Party Conversation Understanding*. ACL 2021.

MPC-BERT: model overview

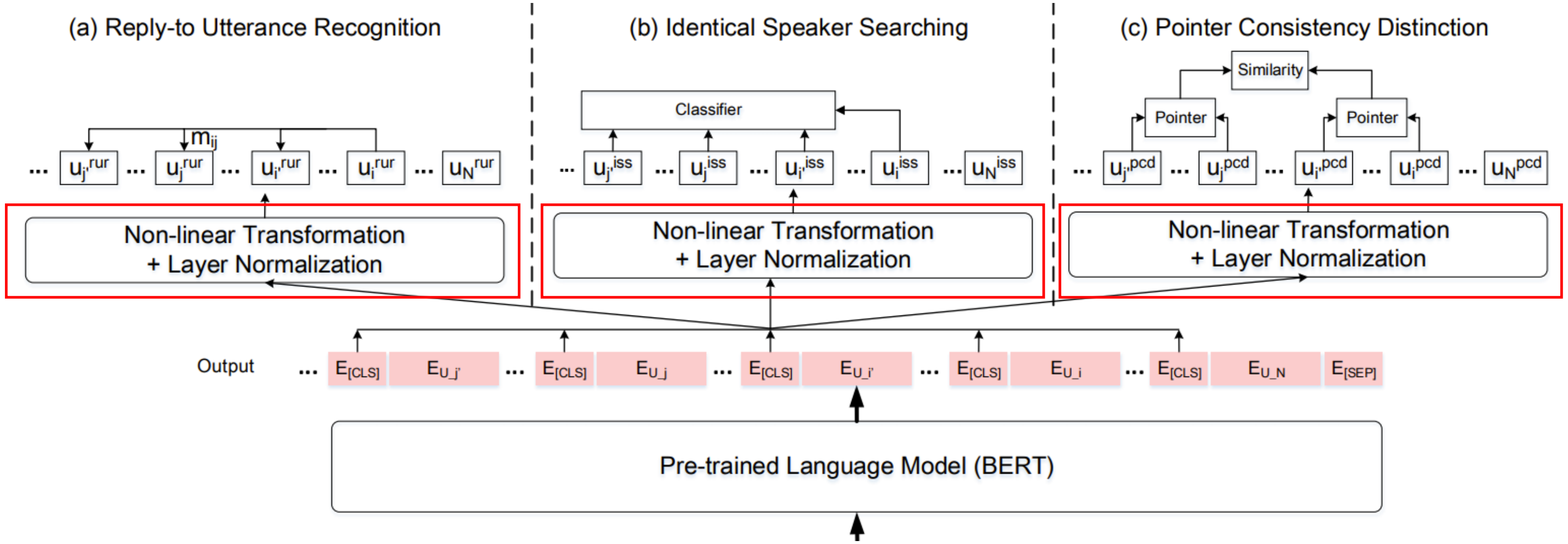
- A [CLS] token is inserted at the start of each utterance
- Position-based speaker embeddings (Gu et al., 2020) are introduced considering that interlocutors are inconsistent in different conversations



Jia-Chen Gu, et al. *Speaker-Aware BERT for Multi-Turn Response Selection in Retrieval-Based Chatbots.* CIKM 2020.

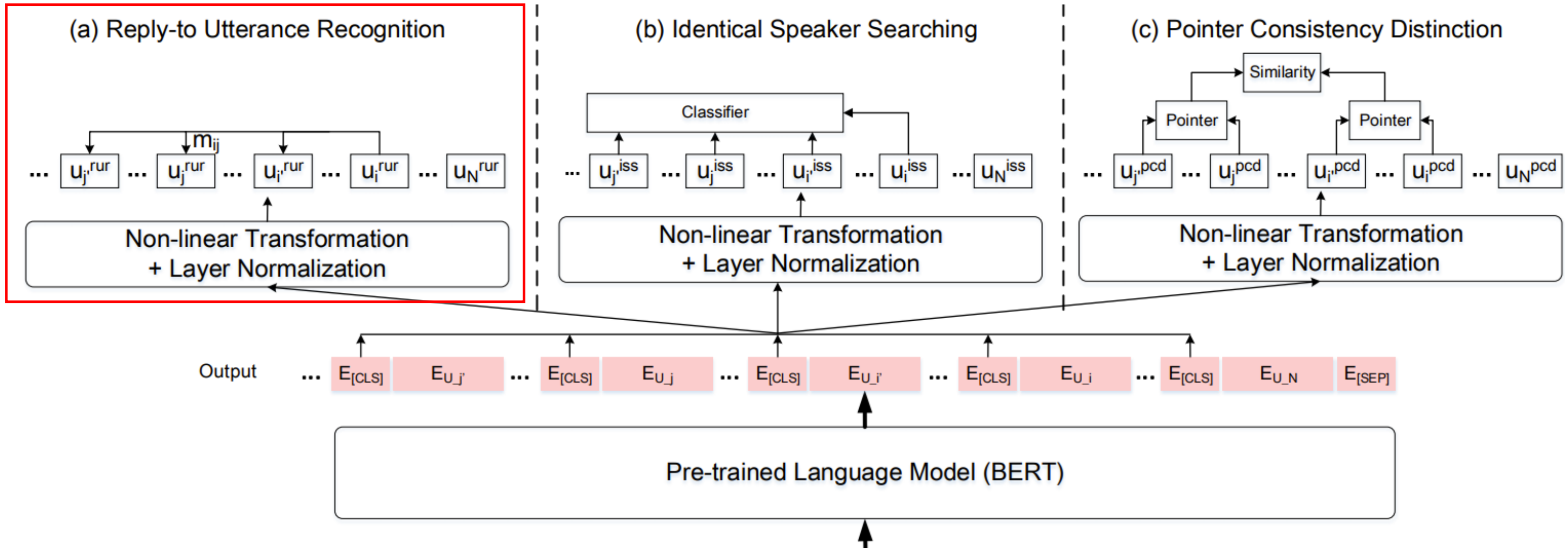
MPC-BERT: interlocutor structure modeling

- Extract the representations for each [CLS] token representing utterances
- Task-dependent non-linear transformations are placed on top of BERT for three self-supervised tasks
- Encoding the input data only once is computation-efficient



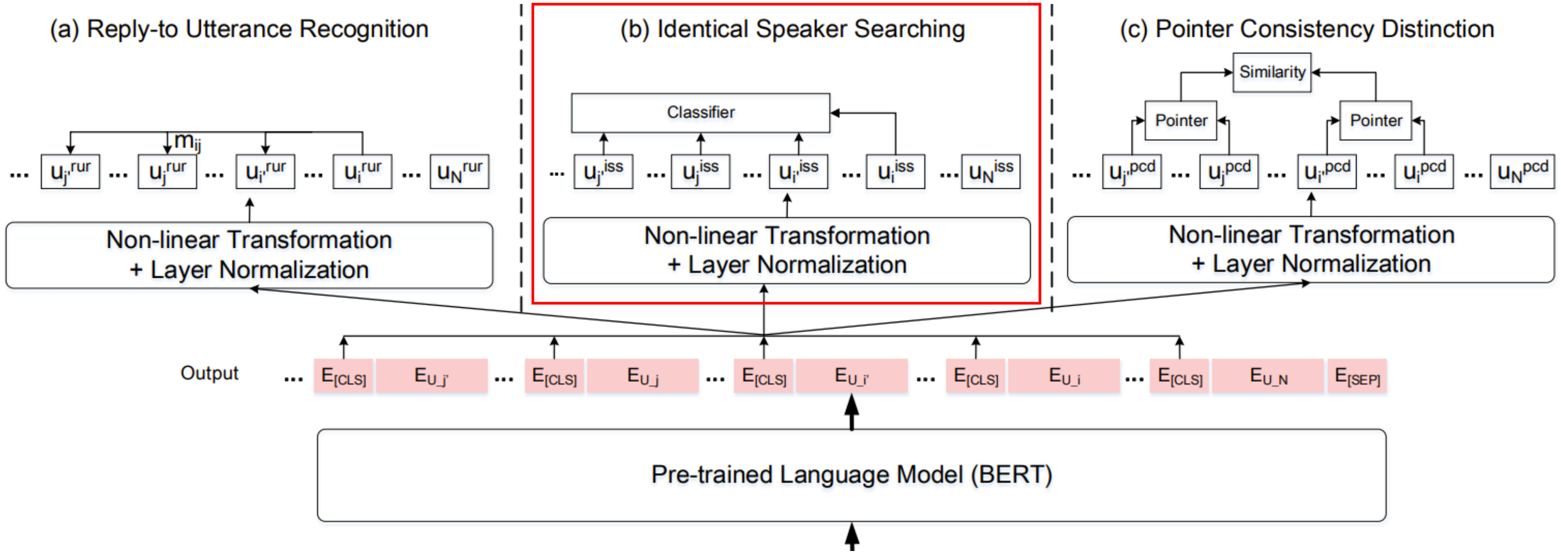
Reply-to Utterance Recognition

- **Motivation**: learn which preceding utterance the current utterance replies to
- **Implementation**: calculate the matching scores with all its preceding utterances for a target utterance



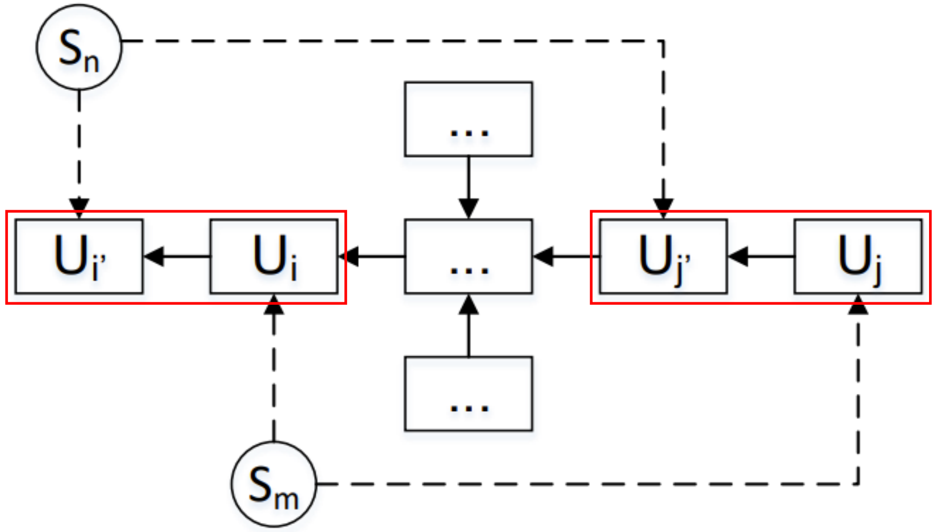
Identical Speaker Searching

- **Motivation**: reformulate as searching for the utterances sharing the identical speaker, since interlocutors varies across conversations
- **Implementation**: mask the speaker embedding of a target utterance, and calculate the probability of utterances sharing the same speaker



Pointer Consistency Distinction

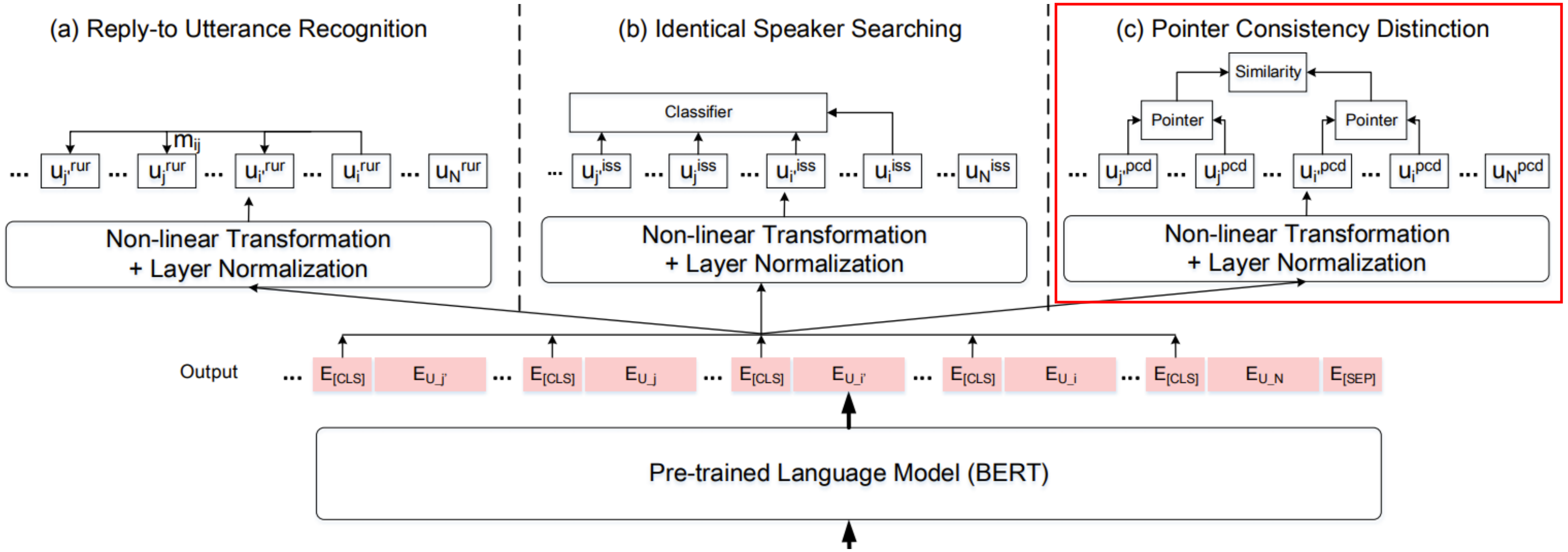
- **Definition**: a speaker-to-addressee pointer is expressed as a pair of utterances representing the "reply-to" relationship
- **Assumption**: the representations of two pointers directing from the same speaker to the same addressee should be consistent



Pointer Consistency Distinction

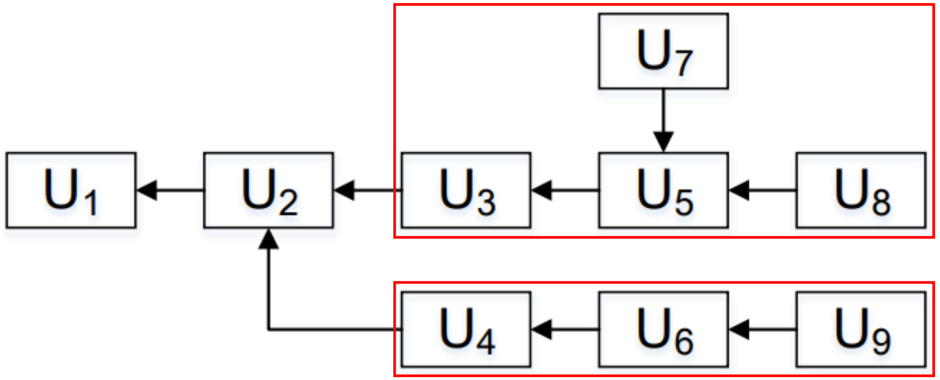
- **Implementation** : a) capture the pointer information contained in each utterance pair

  b) sample a consistent pointer and an inconsistent one from this conversation, and calculate similarities between every two pointers



Utterance Semantics Modeling: Shared Node Detection

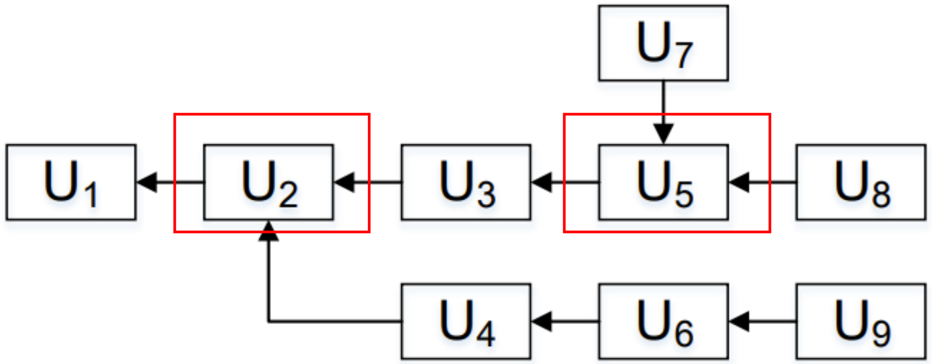
- A full MPC instance can be divided into several sub-conversations, e.g., two sub-conversations {U3, U5, U7, U8} and {U4, U6, U9} share the same parent node U2



- **Assumption**: the representations of sub-conversations under the same parent node tend to be similar

Utterance Semantics Modeling: Masked Shared Utterance Restoration

- A shared utterance is semantically relevant to more utterances in the context than non-shared ones, e.g., U2 and U5

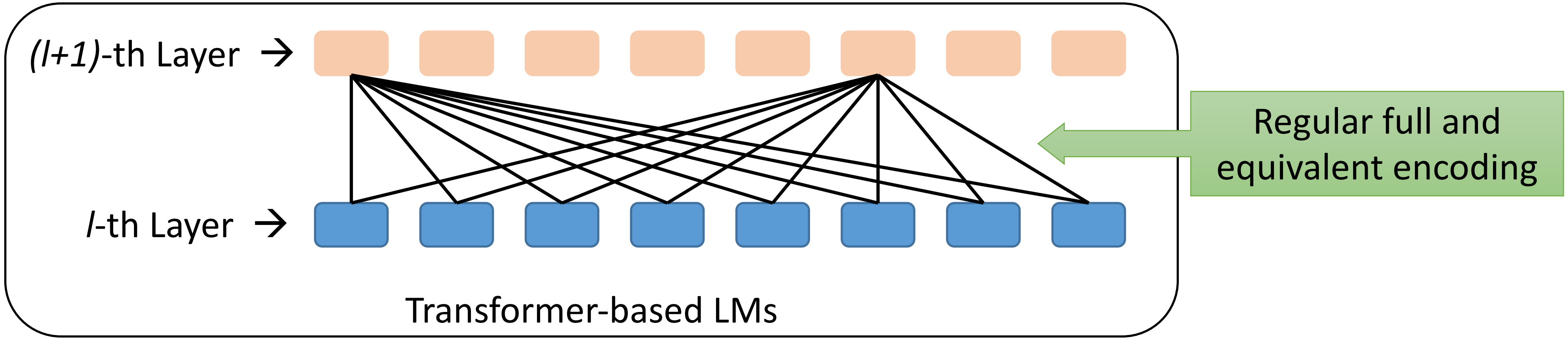


- **Assumption**: masking a sampled shared utterance and enforcing model to restore the masked shared utterance given the rest conversation can enhance the conversation understanding

GIFT for MPC Understanding

GIFT

- **Motivation**: full and equivalent connections among utterance tokens ignore sparse but distinctive dependency of one utterance on another

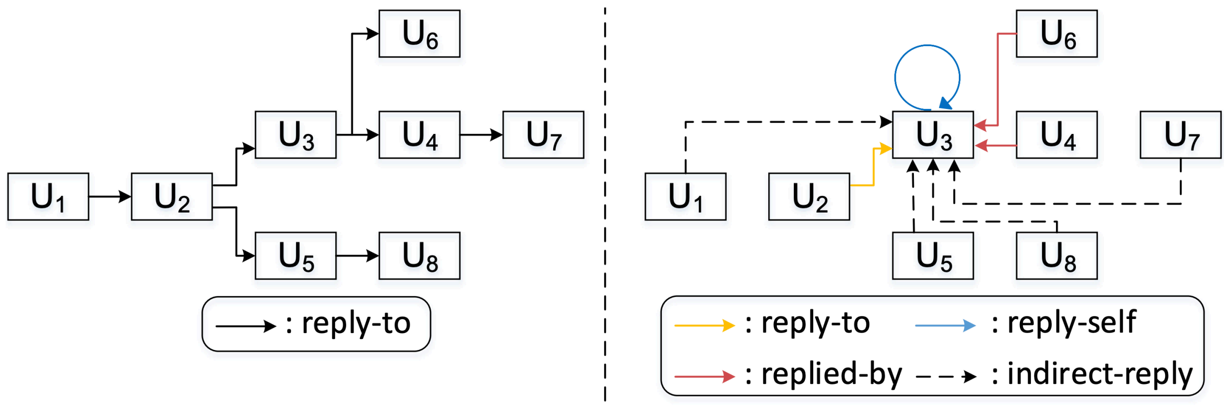


- **Methodology**: distinguish different utterance relationships and model inherent MPC graph structures via graph-induced fine-tuning

Jia-Chen Gu, et al. *GIFT: Graph-Induced Fine-Tuning for Multi-Party Conversation Understanding*. ACL 2023.

GIFT 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*}

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

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

Graph-Induced Signals Integration

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

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

  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

Graph-Induced Signals Integration

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

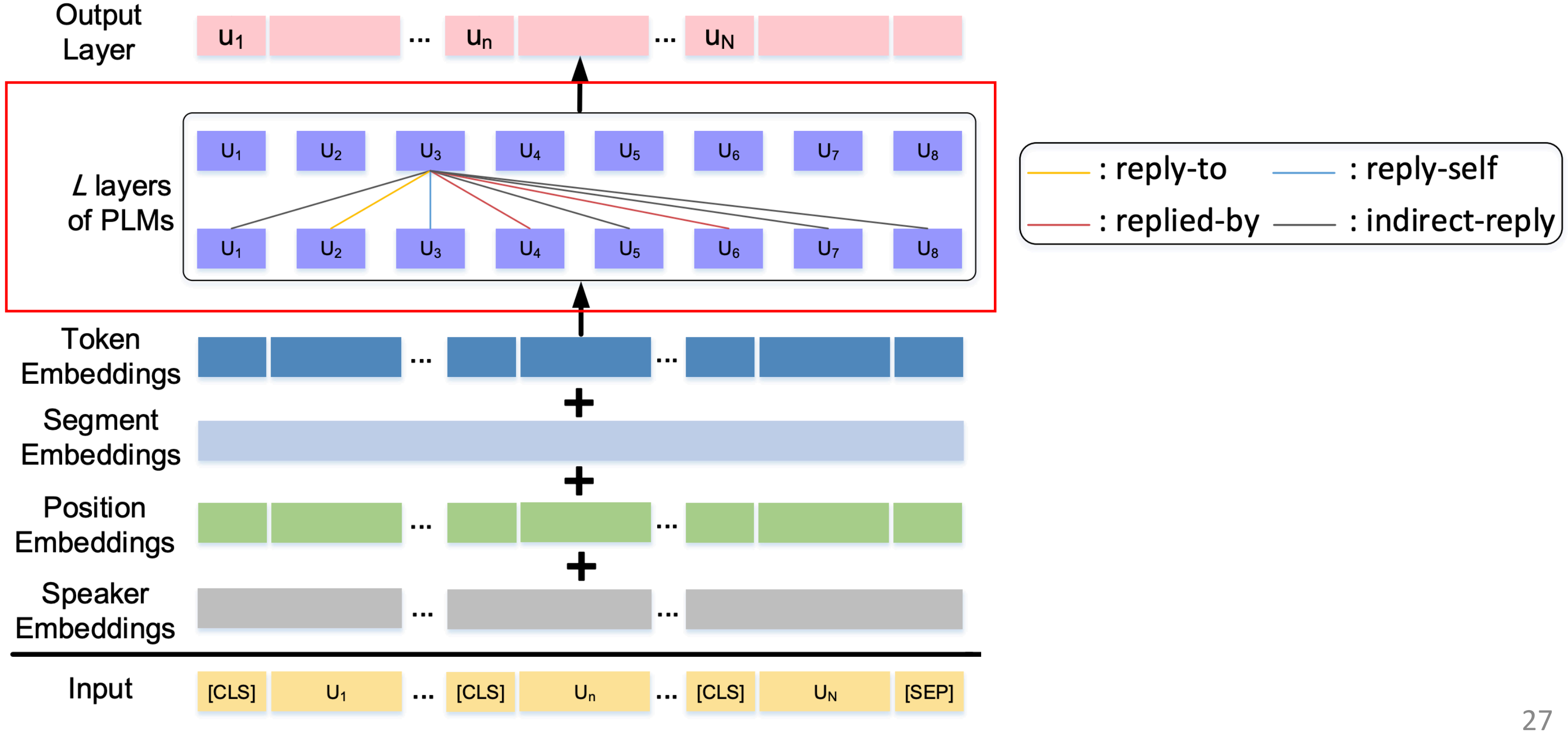
$$\mathrm{Atten}(q,k,v) = \mathrm{softmax}(\phi(e_{q,v})\frac{\mathbf{q}^\top \mathbf{k}}{\sqrt{d}})\mathbf{v}$$

  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

GIFT 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, reflecting graphical conversation flow in Transformer

Downstream Tasks

- **Addressee Recognition**: to recognize the addressees of 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

Datasets

Evaluated 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

Hiroki Ouchi and Yuta Tsuboi. 2016. *Addressee and Response Selection for Multi-Party Conversation.* In *Proc. EMNLP*.
Wenpeng Hu, Zhangming Chan, Bing Liu, et al. 2019. *GSN: A Graph-Structured Network for Multi-Party Dialogues*. In *Proc. IJCAI*.

Results: Addressee Recognition

- MPC-BERT outperforms SA-BERT by margins of 2.56%, 2.22%, 2.40% and 2.14% on these test sets respectively in terms of Precision (P@1)
- GIFT improves BERT by margins of 2.92%, 2.73%, 5.75% and 5.08% respectively

GIFT 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†	82.65†
MPC-BERT w/ GIFT	**90.18**	**85.85**†	**84.13**†	**83.61**†

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

Results: Speaker Identification

- MPC-BERT outperforms SA-BERT by margins of 7.66%, 2.60%, 3.38% and 4.24% P@1
- GIFT improve BERT by margins of 13.71%, 27.50%, 29.14% and 28.82% P@1

  improve SA-BERT by margins of 12.14%, 25.05%, 25.14% and 26.59% P@1

  improve MPC-BERT by margins of 6.96%, 23.05%, 23.12% and 22.99% P@1

	Hu et al. (2019)	Ouchi and Tsuboi (2016)		
		Len-5	Len-10	Len-15
BERT (Devlin et al., 2019)	71.81	62.24	53.17	51.58
SA-BERT (Gu et al., 2020)	75.88	64.96	57.62	54.28
MPC-BERT (Gu et al., 2021)	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**†

Results: Response Selection

- MPC-BERT outperforms SA-BERT by margins of 3.82%, 2.71%, 2.55% and 3.22% $R_{10}@1$
- GIFT improve BERT by margins of 2.48%, 2.12%, 2.71% and 2.34% $R_{10}@1$

  improve SA-BERT by margins of 3.04%, 4.16%, 5.18% and 5.35% $R_{10}@1$

  improve MPC-BERT by margins of 1.76%, 0.88%, 2.15% and 2.44% $R_{10}@1$

	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**†

Ablations of Self-supervised Tasks

	Hu et al. (2019)		Ouchi and Tsuboi (2016)					
			Len-5		Len-10		Len-15	
	P@1	Acc.	P@1	Acc.	P@1	Acc.	P@1	Acc.
Preceding (Le et al., 2019)	-	-	63.50	40.46	56.84	21.06	54.97	13.08
Subsequent (Le et al., 2019)	-	-	61.03	40.25	54.57	20.26	53.07	12.79
DRNN (Ouchi and Tsuboi, 2016)	-	-	72.75	58.18	65.58	34.47	62.60	22.58
SIRNN (Zhang et al., 2018)	-	-	75.98	62.06	70.88	40.66	68.13	28.05
W2W (Le et al., 2019)	-	-	77.55	63.81	73.52	44.14	73.42	34.23
BERT (Devlin et al., 2019)	96.16	83.50	85.95	75.99	83.41	58.22	81.09	44.94
SA-BERT (Gu et al., 2020a)	97.12	88.91	86.81	77.45	84.46	60.30	82.84	47.23
MPC-BERT	**98.31**	**92.42**	**88.73**	**80.31**	**86.23**	**63.58**	**85.55**	**52.59**
MPC-BERT w/o. RUR	97.75	89.98	87.51	78.42	85.63	62.26	84.78	50.83
MPC-BERT w/o. ISS	98.20	91.96	88.67	80.25	86.14	63.40	85.02	51.12
MPC-BERT w/o. PCD	98.20	91.90	88.51	80.06	85.92	62.84	85.21	51.17
MPC-BERT w/o. MSUR	98.08	91.32	88.70	80.26	86.21	63.46	85.28	51.23
MPC-BERT w/o. SND	98.25	92.18	88.68	80.25	86.14	63.41	85.29	51.39

Table 3: Evaluation results of addressee recognition on the test sets. Results except ours are cited from Le et al. (2019). Numbers in bold denote that the improvement over the best performing baseline is statistically significant (t-test with p-value < 0.05).

	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., 2020a)	92.98	75.16	86.53	55.24	87.98	59.27	88.34	60.42
MPC-BERT	**94.90**	**78.98**	**87.63**	**57.95**	**89.14**	**61.82**	**89.70**	**63.64**
MPC-BERT w/o. RUR	94.48	78.16	87.20	57.56	88.96	61.47	89.07	63.24
MPC-BERT w/o. ISS	94.58	78.82	87.54	57.77	88.98	61.76	89.58	63.51
MPC-BERT w/o. PCD	94.66	78.70	87.50	57.51	88.75	61.62	89.45	63.46
MPC-BERT w/o. MSUR	94.36	78.22	87.11	57.58	88.59	61.05	89.25	63.20
MPC-BERT w/o. SND	93.92	76.96	87.30	57.54	88.77	61.54	89.27	63.34

Table 5: Evaluation results of response selection on the test sets. Results except ours are cited from Ouchi and Tsuboi (2016) and Zhang et al. (2018). Numbers in bold denote that the improvement over the best performing baseline is statistically significant (t-test with p-value < 0.05).

	Hu et al. (2019)	Ouchi and Tsuboi (2016)		
		Len-5	Len-10	Len-15
BERT (Devlin et al., 2019)	71.81	62.24	53.17	51.58
SA-BERT (Gu et al., 2020a)	75.88	64.96	57.62	54.28
MPC-BERT	**83.54**	**67.56**	**61.00**	**58.52**
MPC-BERT w/o. RUR	82.48	66.88	60.12	57.33
MPC-BERT w/o. ISS	77.95	66.77	60.03	56.73
MPC-BERT w/o. PCD	83.39	67.12	60.62	58.00
MPC-BERT w/o. MSUR	83.51	67.21	60.76	58.03
MPC-BERT w/o. SND	83.47	67.04	60.44	58.12

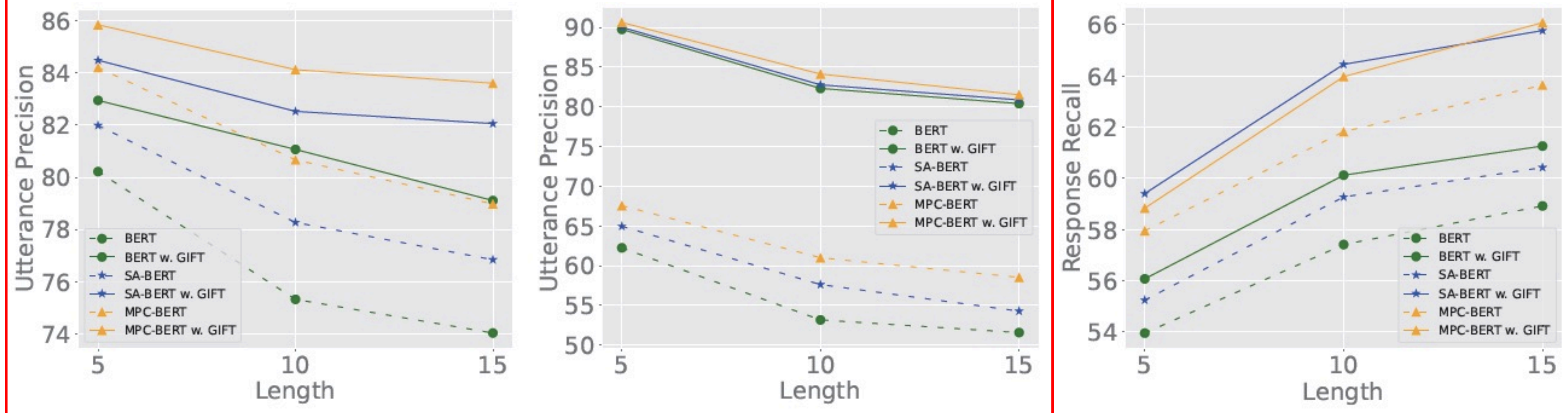
Table 4: Evaluation results of speaker identification on the test sets in terms of P@1. Numbers in bold denote that the improvement over the best performing baseline is statistically significant (t-test with p-value < 0.05).

- Tasks focusing on interlocutor structures contribute more to addressee recognition and speaker identification
- Tasks focusing on utterance semantics contribute more to response selection

Performance Change at Different Lengths

Results: the performance of addressee recognition and speaker identification **dropped** as the session length increased

Analysis: longer sessions always contain more interlocutors which increase the difficulties of predicting interlocutors

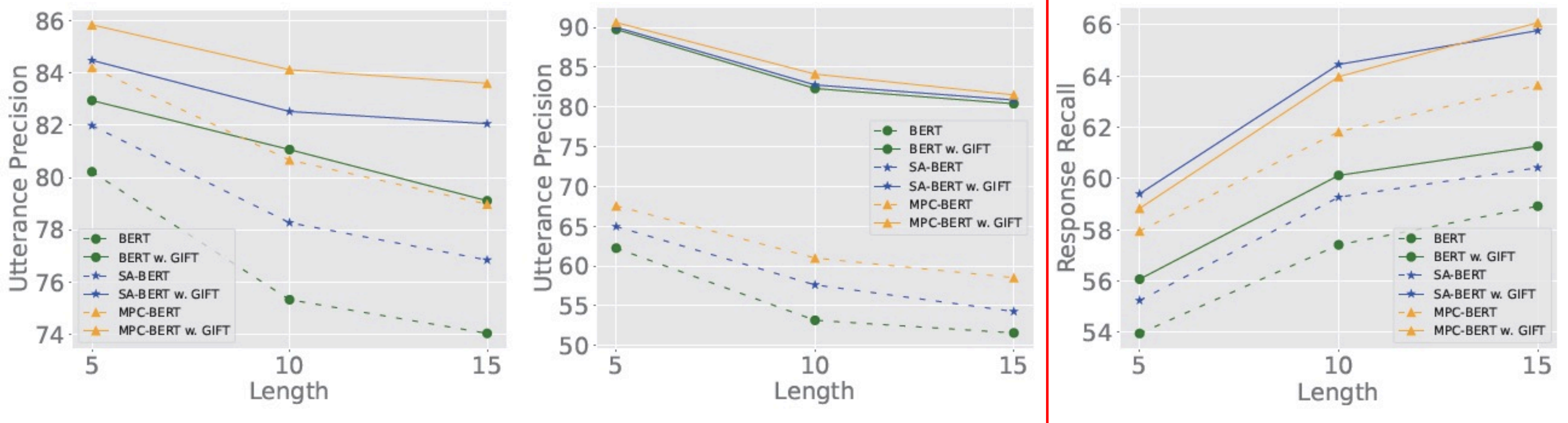


(a) Addressee Recognition (b) Speaker Identification (c) Response Selection

Performance Change at Different Lengths

Results: the performance of response selection was significantly **improved** as the session length increased

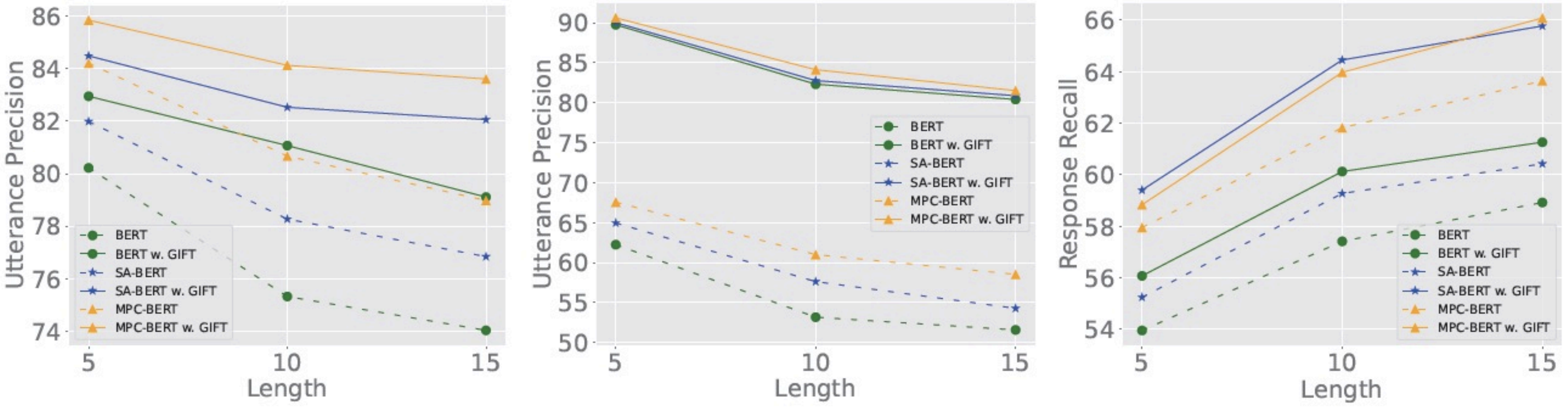
Analysis: longer sessions enrich the representations of contexts with more details which benefit response selection



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

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 (2 margins for 3 models on 3 tasks)



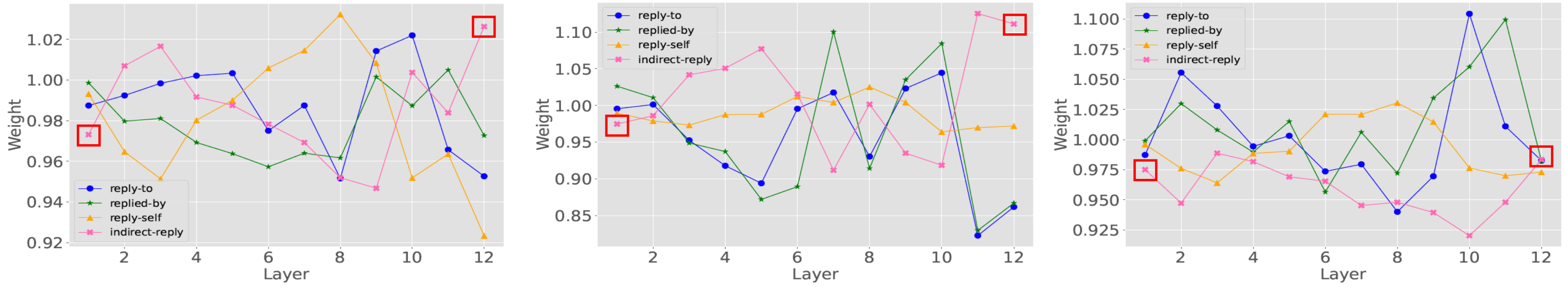
(a) Addressee Recognition (b) Speaker Identification (c) Response Selection

	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 GIFT 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

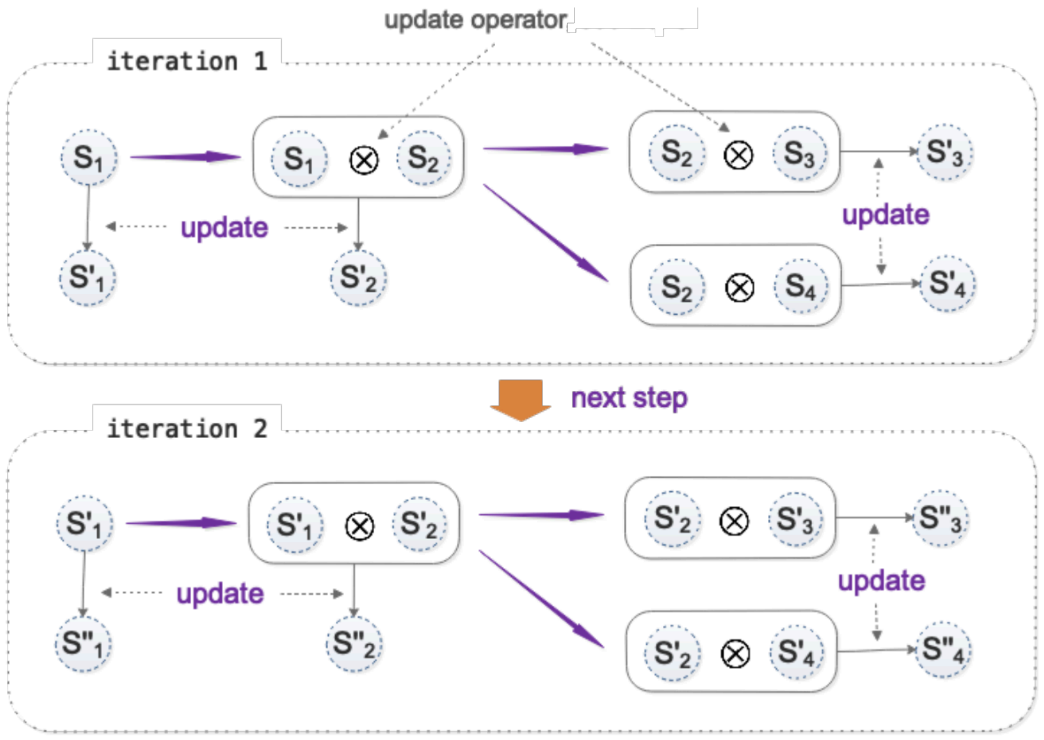


(a) Addressee Recognition (b) Speaker Identification (c) Response Selection

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

HeterMPC for MPC Generation

Previous Work: GSN



Homogeneous graph
composed of only utterances!

Utterance-level graph-based encoder which encodes utterances based on the graph topology rather than the appearance sequence

Each utterance (a node in the graph) accepts information from all its connected utterances (nodes) in each iteration

Wenpeng Hu, et al. *GSN: A Graph-Structured Network for Multi-Party Dialogues*. IJCAI 2019.

Is a homogeneous graph expressive enough to represent an MPC?

Q1: Are there other sources of information in addition to utterances that should be embraced in a unified graph?

Q2: Is it necessary to distinguish the fine-grained and complicated interactions between utterance and interlocutor graph nodes?

HeterMPC: Graph Construction

- *M* utterances and *I* interlocutors → a heterogeneous graph *G(V, E)*

Jia-Chen Gu, et al. *HeterMPC: A Heterogeneous Graph Neural Network for Response Generation in Multi-Party Conversations*. ACL 2022.

HeterMPC: Graph Construction

- M utterances and I interlocutors → a heterogeneous graph $G(V, E)$
- V : a set of $M + I$ nodes, each denoting an utterance or an interlocutor

Jia-Chen Gu, et al. *HeterMPC: A Heterogeneous Graph Neural Network for Response Generation in Multi-Party Conversations*. ACL 2022.

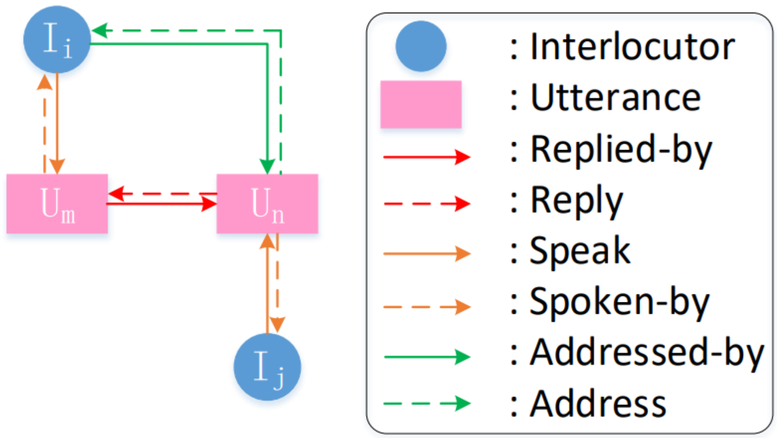
HeterMPC: Graph Construction

- M utterances and I interlocutors → a heterogeneous graph $G(V, E)$
- V : a set of $M + I$ nodes, each denoting an utterance or an interlocutor
- $E = \{e_{p,q}\}_{p,q=1}^{M+I}$: a set of directed edges, each edge $e_{p,q}$ describing the connection from node p to node q

Jia-Chen Gu, et al. *HeterMPC: A Heterogeneous Graph Neural Network for Response Generation in Multi-Party Conversations*. ACL 2022.

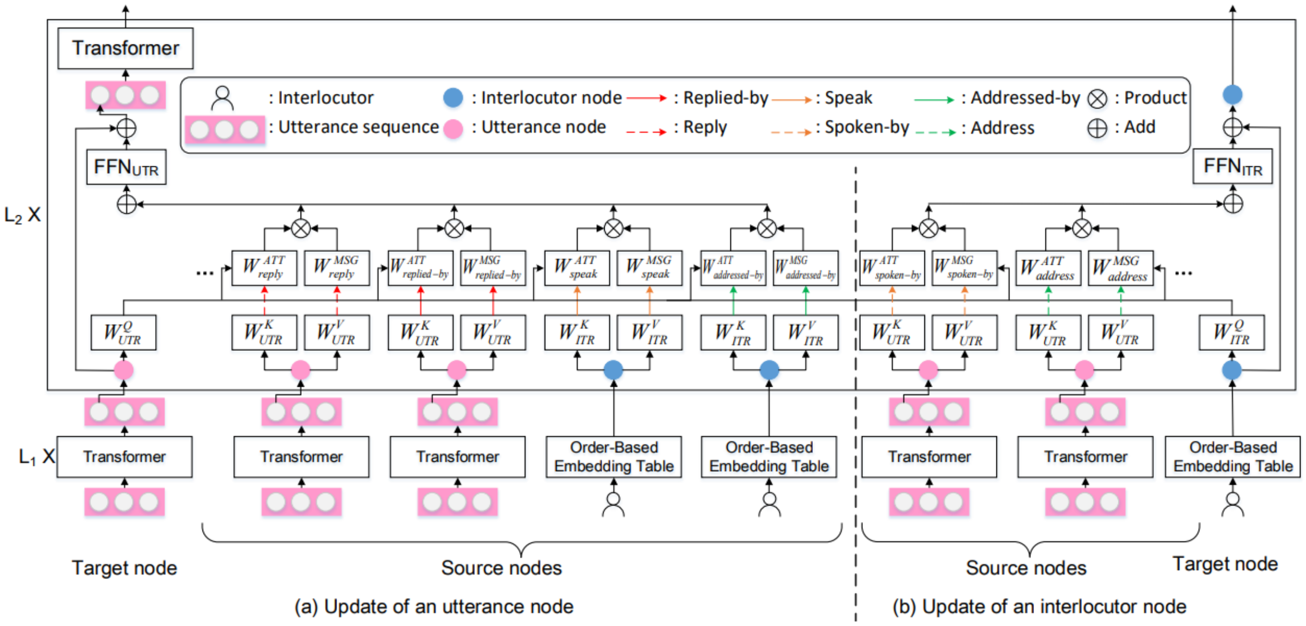
HeterMPC: Graph Construction

- *M* utterances and *I* interlocutors → a heterogeneous graph *G(V, E)*
- *V* : a set of *M + I* nodes, each denoting an utterance or an interlocutor
- $E = \{e_{p,q}\}_{p,q=1}^{M+I}$: a set of directed edges, each edge $e_{p,q}$ describing the connection from node *p* to node *q*
- Six types of meta relations: {*reply, replied-by, speak, spoken-by, address, addressed-by*} to describe directed edges between two nodes



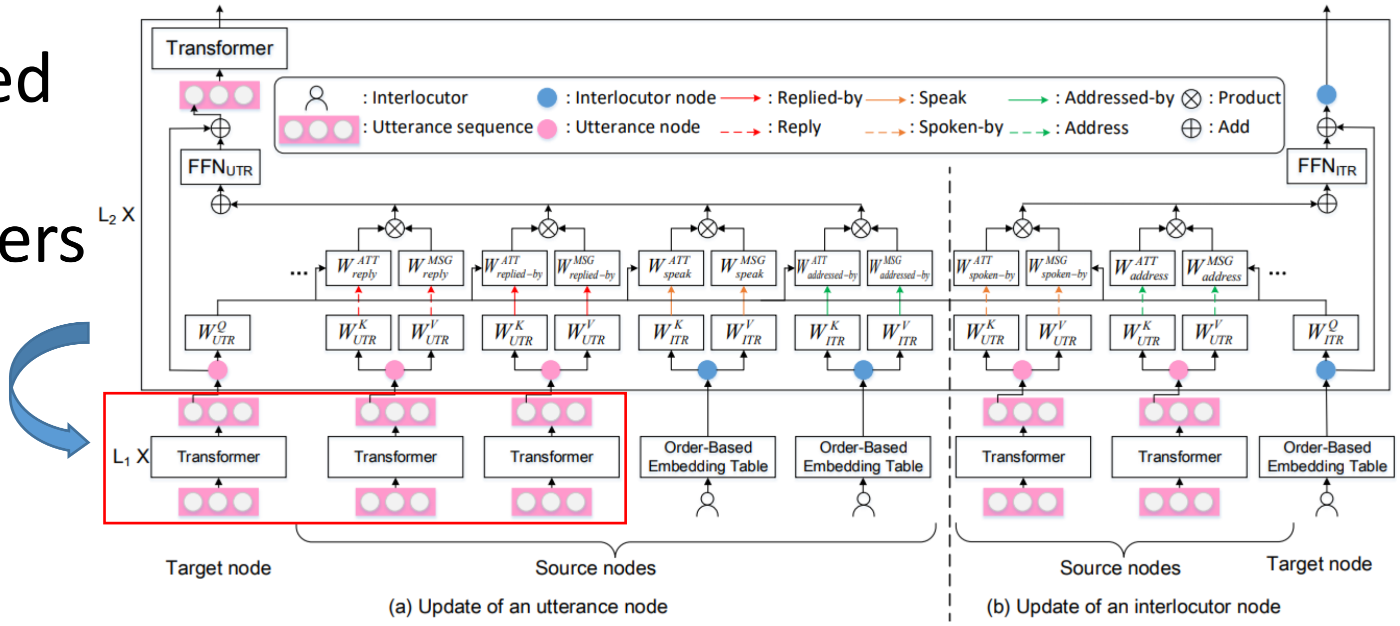
Jia-Chen Gu, et al. *HeterMPC: A Heterogeneous Graph Neural Network for Response Generation in Multi-Party Conversations*. ACL 2022.

HeterMPC: Node Initialization



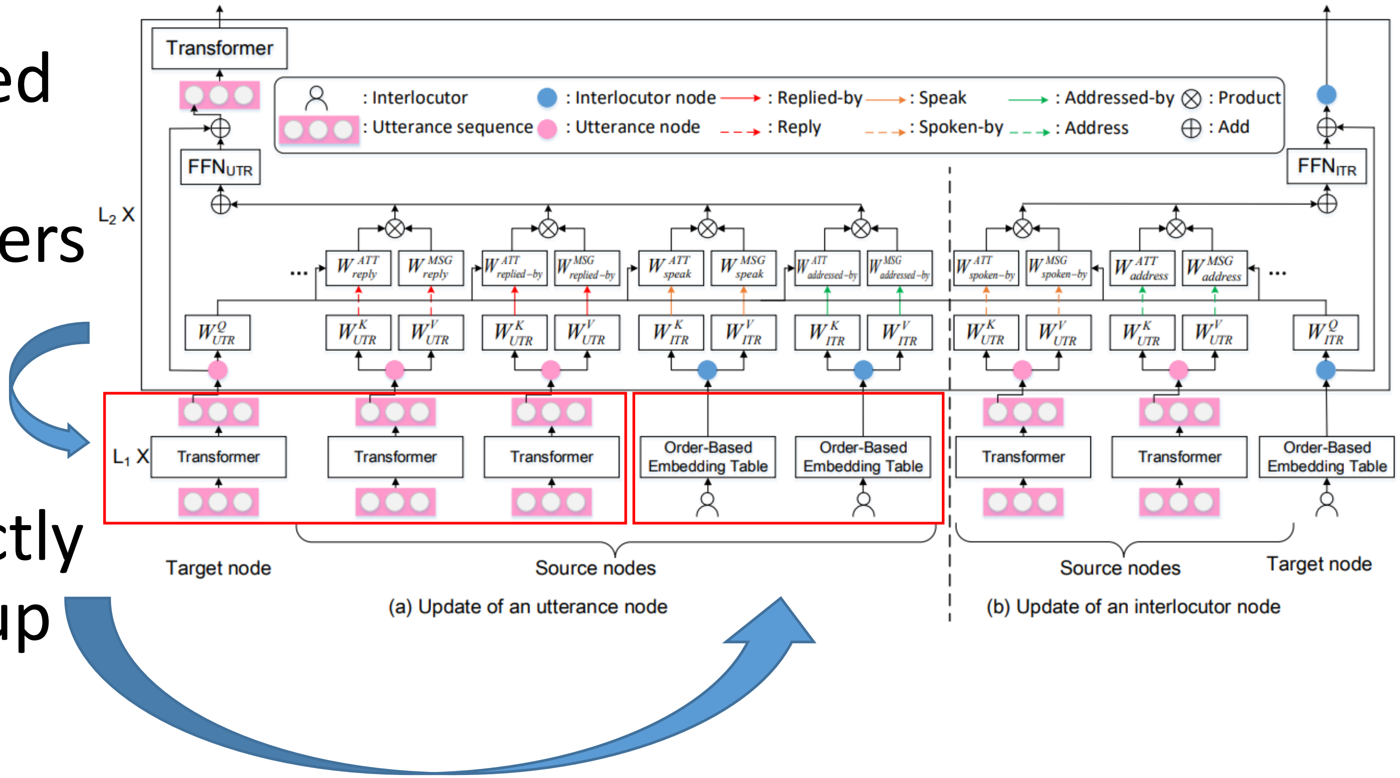
HeterMPC: Node Initialization

- Each utterance is encoded individually by stacked Transformer encoder layers



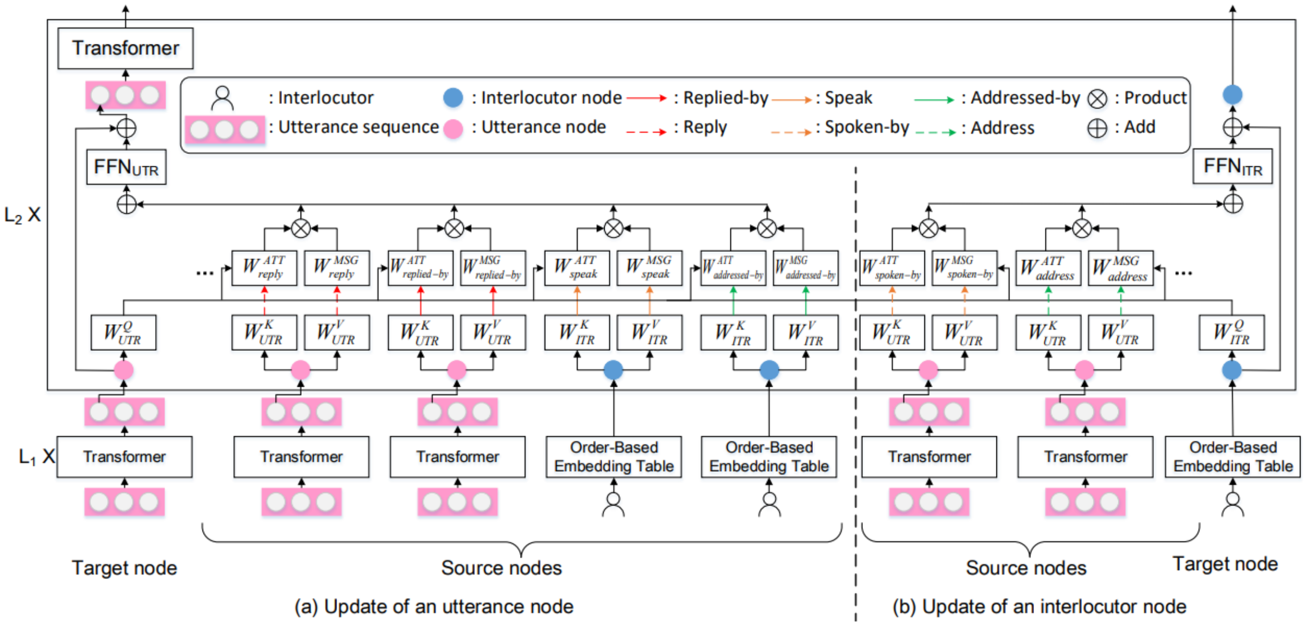
HeterMPC: Node Initialization

- Each utterance is encoded individually by stacked Transformer encoder layers
- Each interlocutor is directly represented by looking up a position-based interlocutor embedding table



HeterMPC: Node Updating

Introduce parameters to model heterogeneity via



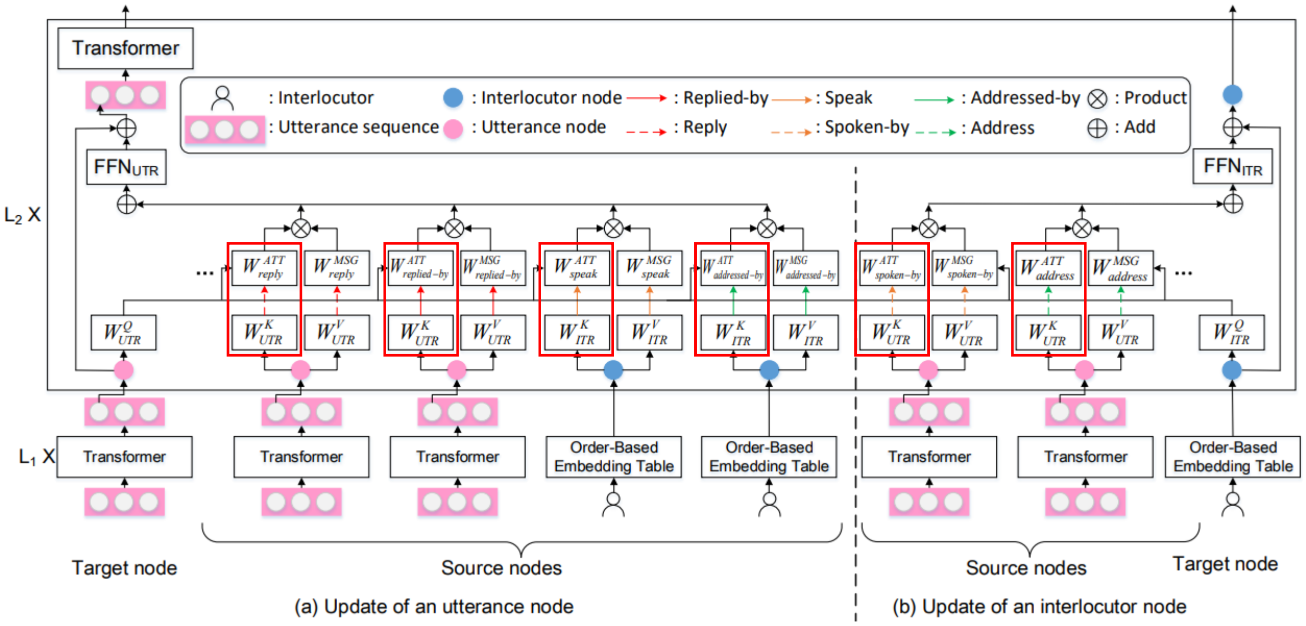
Ziniu Hu, et al. *Heterogeneous Graph Transformer*. WWW 2020.

HeterMPC: Node Updating

Introduce parameters to model heterogeneity via

- attention weights

$$k^l(s) = h_s^l W_{\tau(s)}^K + b_{\tau(s)}^K,$$
$$q^l(t) = h_t^l W_{\tau(t)}^Q + b_{\tau(t)}^Q,$$
$$w^l(s,e,t) = k^l(s) W_{e_{s,t}}^{ATT} q^l(t)^T \frac{\mu_{e_{s,t}}}{\sqrt{d}}.$$



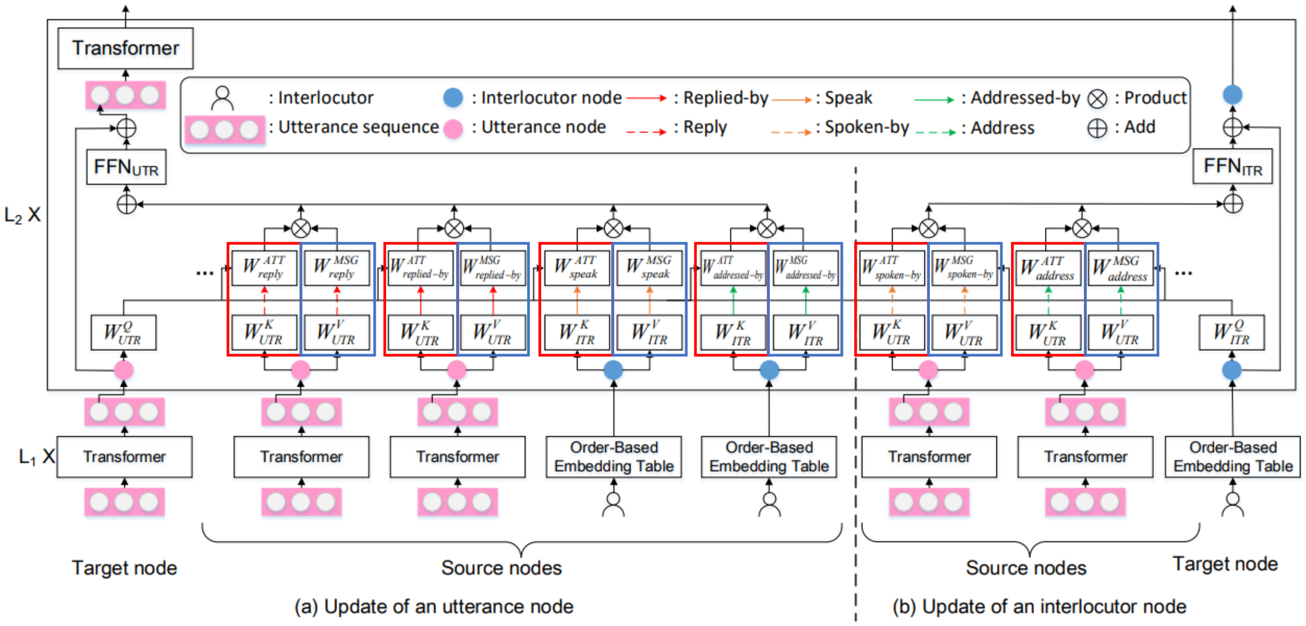
Ziniu Hu, et al. *Heterogeneous Graph Transformer*. WWW 2020.

HeterMPC: Node Updating

Introduce parameters to model heterogeneity via

- attention weights
- message passing

$$\boldsymbol{v}^l(s) = \boldsymbol{h}_s^l \boldsymbol{W}_{\tau(s)}^V + \boldsymbol{b}_{\tau(s)}^V,$$
$$\bar{\boldsymbol{v}}^l(s) = \boldsymbol{v}^l(s) \boldsymbol{W}_{e_{s,t}}^{MSG},$$



Ziniu Hu, et al. *Heterogeneous Graph Transformer*. WWW 2020.

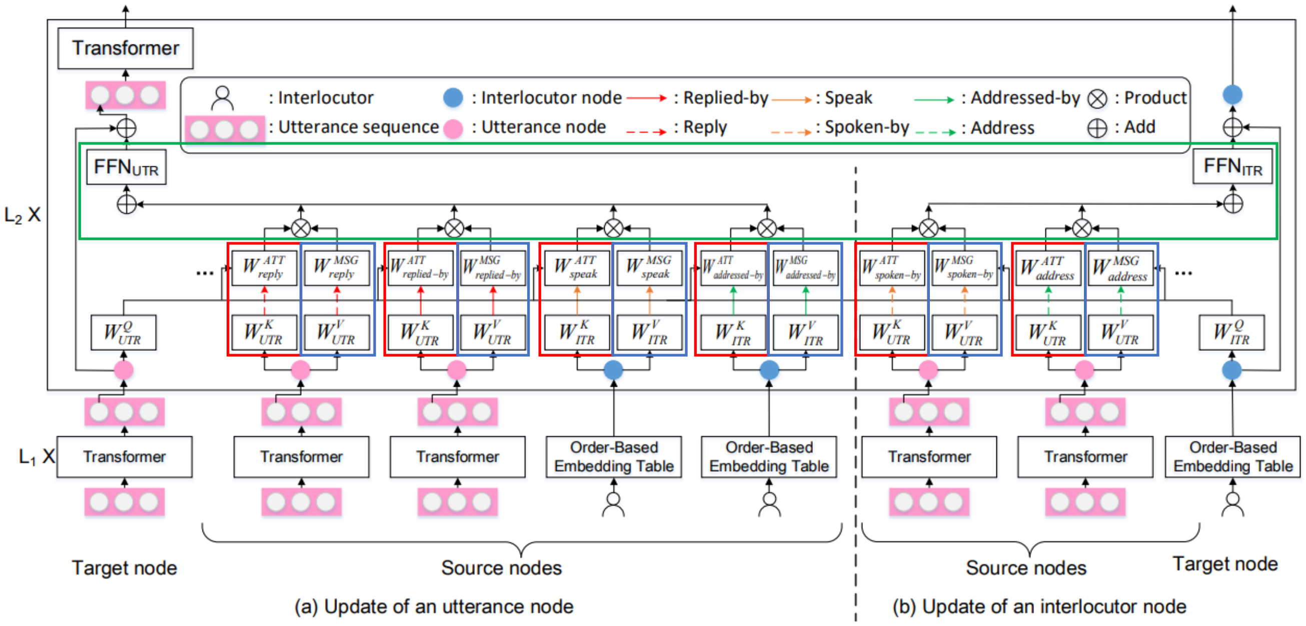
HeterMPC: Node Updating

Introduce parameters to model heterogeneity via

- attention weights
- message passing
- information aggregation

$$\bar{\boldsymbol{h}}_t^l = \sum \mathrm{softmax}(w^l(s,e,t))\bar{\boldsymbol{v}}^l(s),$$

$$\boldsymbol{h}_t^{l+1} = FFN_{\tau(t)}(\bar{\boldsymbol{h}}_t^l) + \boldsymbol{h}_t^l,$$

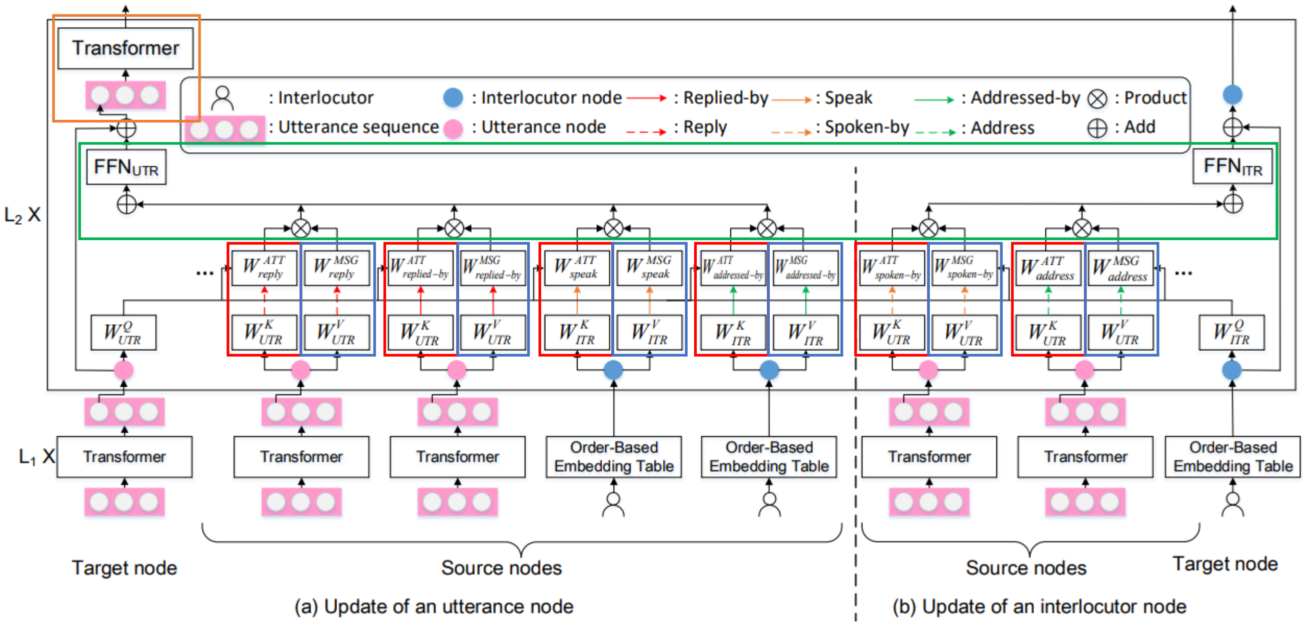


Ziniu Hu, et al. *Heterogeneous Graph Transformer*. WWW 2020.

HeterMPC: Node Updating

Introduce parameters to model heterogeneity via

- attention weights
- message passing
- information aggregation

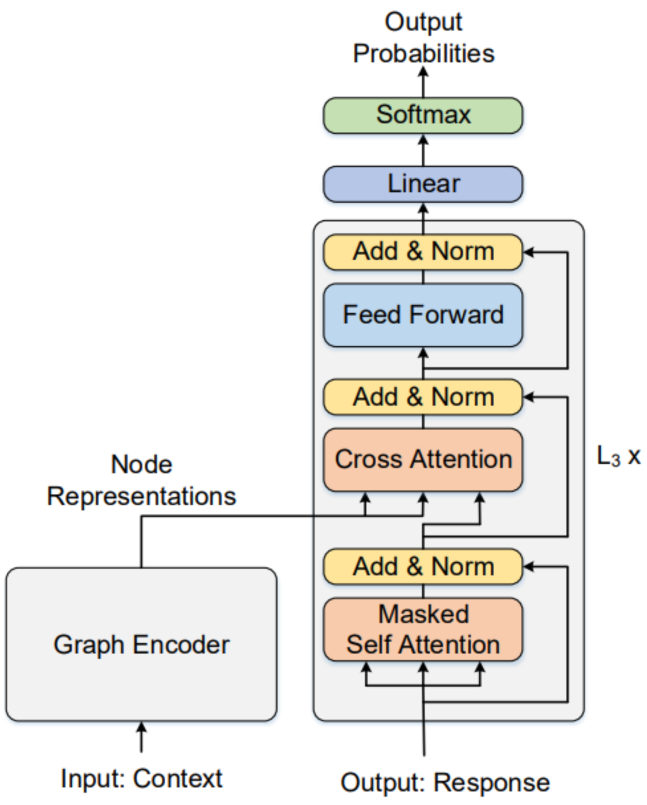


- Specifically, the context information in an utterance node is shared with other tokens in this utterance through another layer of intra-utterance Transformer encoding

Ziniu Hu, et al. *Heterogeneous Graph Transformer*. WWW 2020.

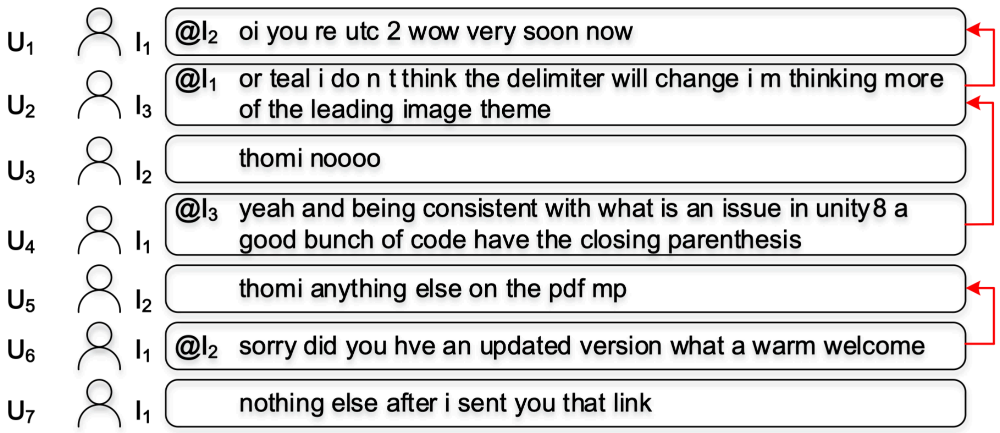
HeterMPC: Decoder

- Standard implementation of Transformer decoder
- A cross-attention operation over the node representations of the graph encoder output is performed to incorporate graph information

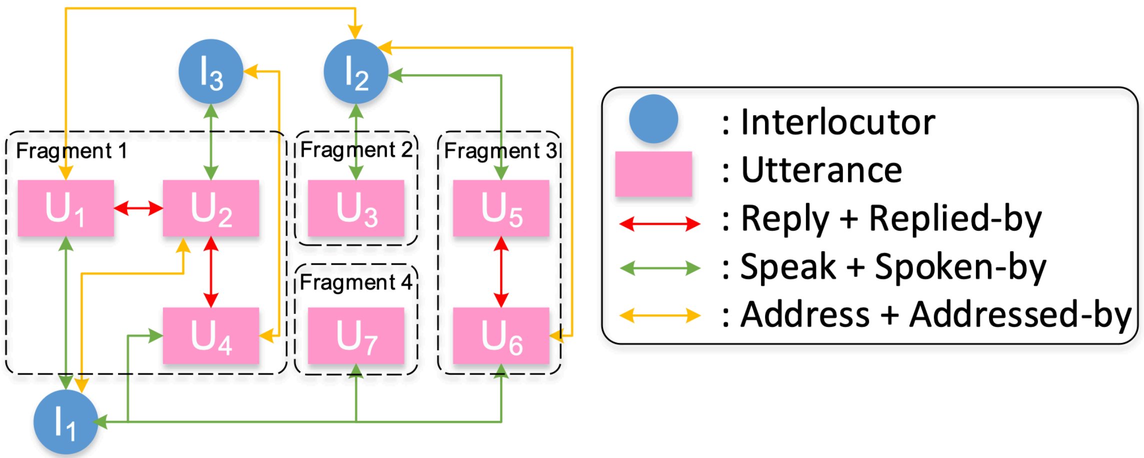


MADNet for MPC Generation

Missing Addressee Labels



An MPC instance with a few addressee labels (@) missing



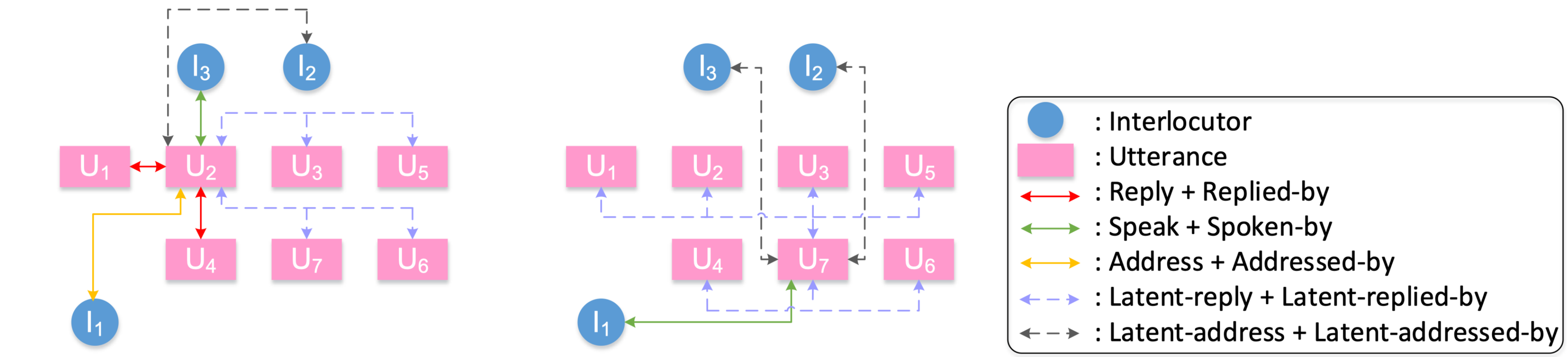
The graphical information flow and fragments established in HeterMPC (Gu et al., 2022)

Nodes without direct connections cannot exchange information between each other through one-hop message passing

55% addressee labels missing in Ouchi and Tsuboi (2016)

MADNet: Fully-Connected Graph

Design four additional types of latent edges {*latent-reply, latent-replied-by, latent-address, latent-addressed-by*} to build a consecutively connected conversation graph



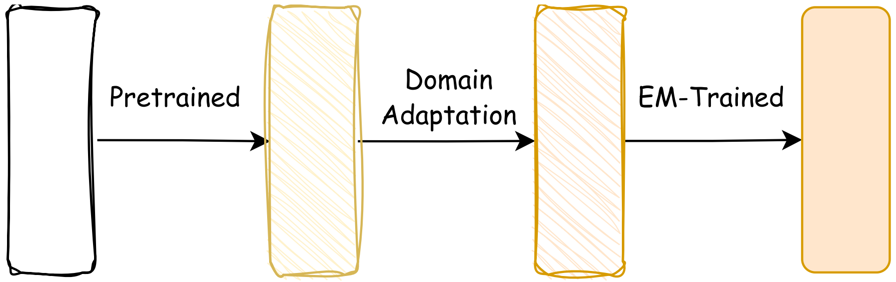
(a) Illustration of edges for utterances with addressee labels, e.g., U_2

(b) Illustration of edges for utterances without addressee labels, e.g., U_7

Jia-Chen Gu, et al. *MADNet: Maximizing Addressee Deduction Expectation for Multi-Party Conversation Generation*. EMNLP 2023.

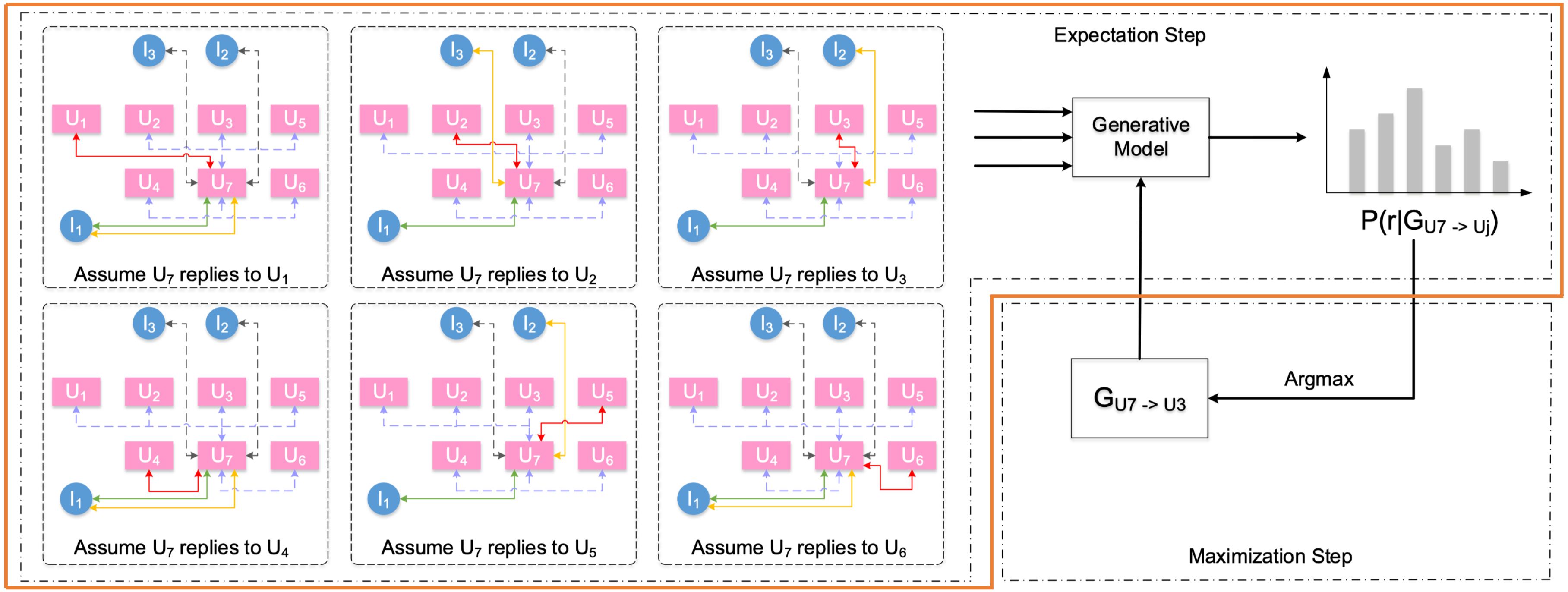
MADNet: EM for Addressee Deduction

- Initialized with PLMs followed by domain adaptation based on the fully-connected graph, since better initialization of addressee labels helps converge to optimal model parameters



- E steps consider the addressee as a discrete latent variable and iteratively generate silver addressee labels
- M steps selects the one with the highest probability from the addressee distribution and optimize the generative model

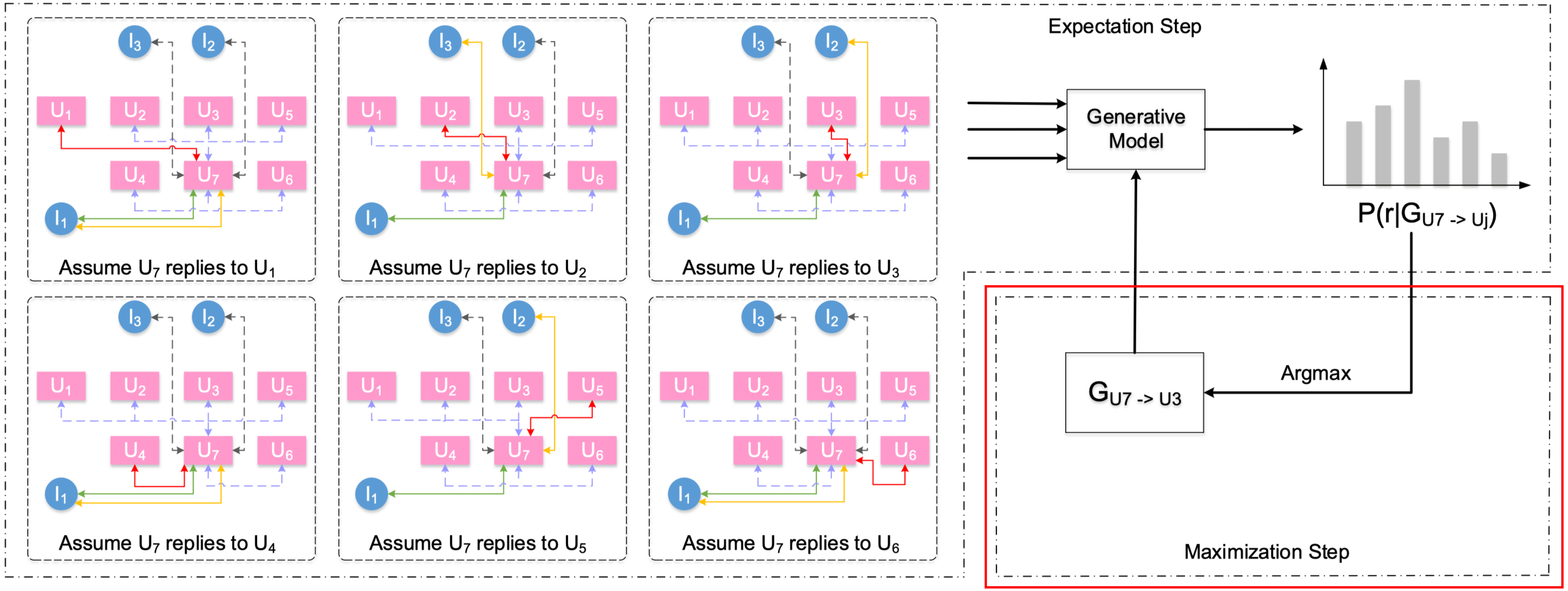
Addressee Deduction: E Steps



- Iteratively generate silver addressee labels by considering the addressee as a discrete latent variable
- The latent addressee distribution is estimated by applying Bayes' rule as:

$$P(\mathbb{G}_{U_i \to U_j} | \boldsymbol{c}, \boldsymbol{r}; \boldsymbol{\theta}) = \frac{P(\boldsymbol{r} | \mathbb{G}_{U_i \to U_j}, \boldsymbol{c}; \boldsymbol{\theta})}{\sum_{k=1}^{i-1} P(\boldsymbol{r} | \mathbb{G}_{U_i \to U_k}, \boldsymbol{c}; \boldsymbol{\theta})}$$

Addressee Deduction: M Steps



- Selects the addressee with the highest probability and optimize the generative dialogue model
- Select the addressee $\bar{U}_j = \underset{U_j}{\text{argmax}}\, P(\mathbb{G}_{U_i \to U_j}|\boldsymbol{c}, \boldsymbol{r}; \boldsymbol{\theta}),\ j < i$
- The maximization step is approximated as $\log P(\boldsymbol{r}, \mathbb{G}_{U_i \to \bar{U}_j}|\boldsymbol{c}; \boldsymbol{\theta})$

Setup

- Dataset: two Ubuntu IRC benchmarks where addressee labels for
  - ✓part of history utterances were missing (Ouchi and Tsuboi, 2016)
  - ✓all history utterances were complete (Hu et al., 2019)

- Baselines
  - ✓Non-graph-based: RNN-based Seq2Seq, Transformer, GPT-2, BERT and BART
  - ✓Graph-based: GSN

- Metrics
  - ✓Automated: BLEU-1 to BLEU-4, METEOR and ROUGE_L
  - ✓Human: relevance, fluency and informativeness

Results

- Evaluation on Ouchi and Tsuboi (2016)

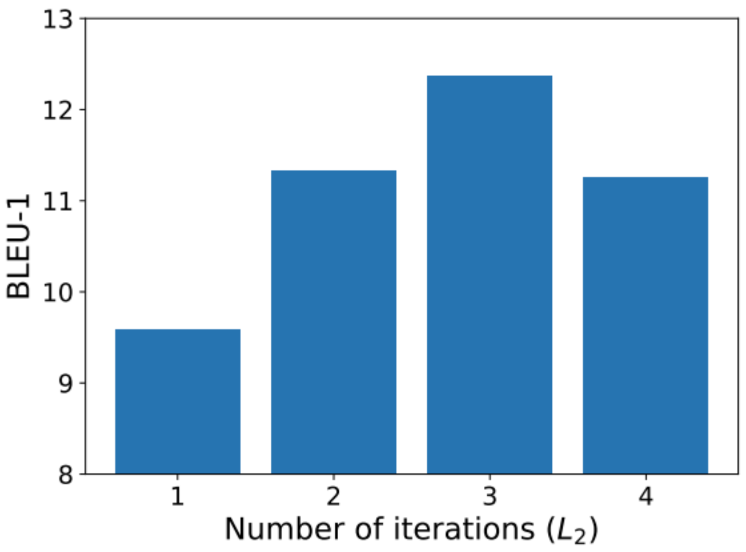
Models \ Metrics	BLEU-1	BLEU-2	BLEU-3	BLEU-4	METEOR	ROUGE$_L$
GSN (Hu et al., 2019)	6.32	2.28	1.10	0.61	3.27	7.39
GPT-2 (Radford et al., 2019)	9.12	3.40	1.93	1.39	3.28	8.92
BART (Lewis et al., 2020)	11.13	3.95	2.11	1.44	4.45	10.20
HeterMPC (Gu et al., 2022)	11.40	4.29	2.43	1.74	4.57	10.44
MADNet	**11.82**†	**4.58**†	**2.65**	**1.91**	**4.90**†	**10.74**†
MADNet w/o. EM for addressee deduction	11.62	4.48	2.59	1.88	4.80	10.63
MADNet w/o. latent-reply and latent-replied-by	11.76	4.43	2.47	1.74	4.83	10.67
MADNet w/o. latent-address and latent-addressed-by	11.54	4.44	2.57	1.87	4.72	10.52

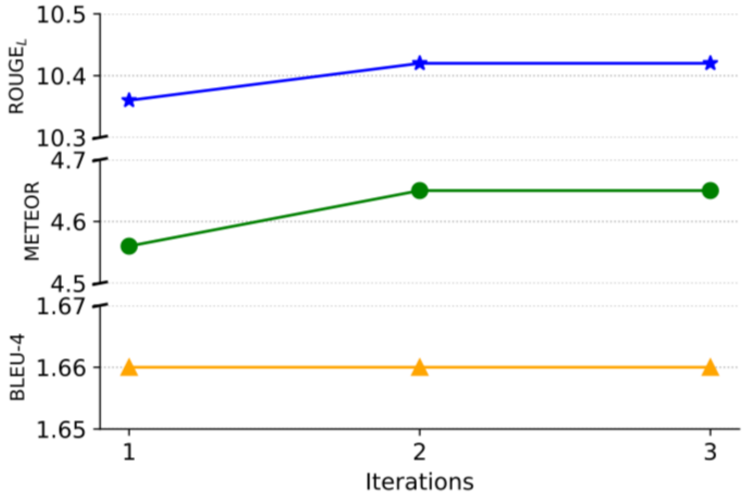
Models \ Metrics	Score
Human	2.09
GSN (Hu et al., 2019)	1.20
BART (Lewis et al., 2020)	1.54
HeterMPC (Gu et al., 2022)	1.62
MADNet	1.79

- Evaluation on Hu et al., (2019)

Models \ Metrics	BLEU-1	BLEU-2	BLEU-3	BLEU-4	METEOR	ROUGE$_L$
GSN (Hu et al., 2019)	10.23	3.57	1.70	0.97	4.10	9.91
GPT-2 (Radford et al., 2019)	10.37	3.60	1.66	0.93	4.01	9.53
BART (Lewis et al., 2020)	11.25	4.02	1.78	0.95	4.46	9.90
HeterMPC (Gu et al., 2022)	12.26	4.80	2.42	1.49	4.94	11.20
MADNet	**12.73**†	**5.12**†	**2.64**	**1.63**	**5.31**†	**11.74**†
MADNet w/o. latent-reply and latent-replied-by	12.54	4.91	2.53	1.59	5.20	11.60
MADNet w/o. latent-address and latent-addressed-by	12.45	4.92	2.52	1.55	5.18	11.60

Analysis





Performance was significantly improved with more node iterations at the beginning. Then, the performance was stable and dropped slightly

Performance was improved with more EM iterations. Then, the performance was stable

Accuracy of Addressee Deduction

Comparing methods:

- HeterMPC
- $\text{HeterMPC}_{\text{rand}}$: each utterance whose addressee label was masked was randomly assigned a previous utterance as its reply-to utterance
- $\text{HeterMPC}_{\text{prec}}$: assigned its preceding utterance as its reply-to utterance
- MADNet
- MADNet with the oracle addressee labels

Models \ Metrics	Accuracy	BLEU-4	METEOR	ROUGE_L
HeterMPC	-	1.33	5.03	11.35
HeterMPC_{rand}	37.4	1.29	4.94	11.23
HeterMPC_{prec}	44.8	1.32	4.96	11.32
MADNet	50.1	1.51	5.17	11.65
MADNet_{orac}	100.0	1.63	5.31	11.74

- ✓ The prediction of addressees significantly affects the performance of MPC generation
- ✓ Seriously wrong predictions might even hurt performance

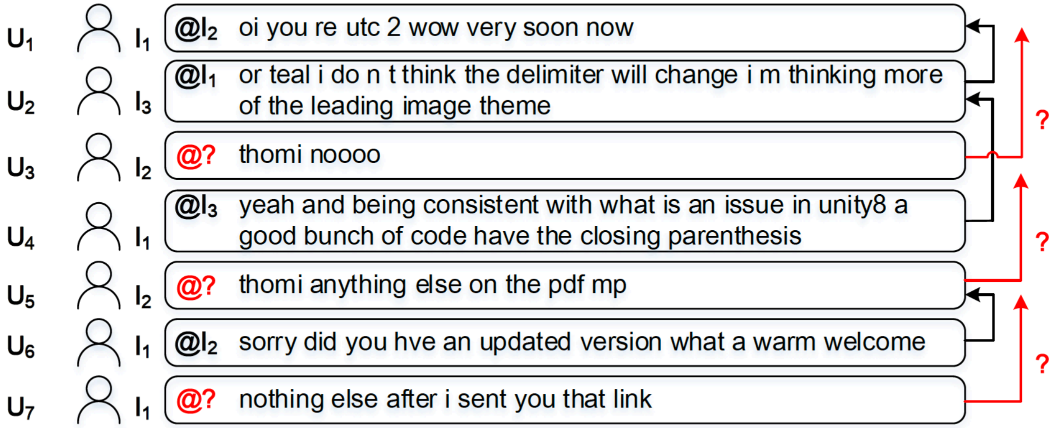
Case Study

- Other system can only generate generic responses such as “i m not sure ...”
- For MADNet, the missing addressee label of the fourth utterance was deduced as I.3
- Given the deduced addressee label, the message of “**phased update**” in the third utterance can be passed to the fourth utterance

Speaker	*Utterance*	*Addressee*
I.1	perhaps but not everyone uses that	-
I.2	i ll ask him for his history log i think	-
I.3	for people who do n t the phased update percentages are n t considered ok 0	I.1
I.1	true	I.3 (Deduced)
I.3	i first thought it might be related to https launchpad net ubuntu source unity scopes api 0 6 19 15 **(Human)**	I.1
	i do n t know how to do that but i m not sure what you want to do with the **(GSN)**	
	i m not sure if you can get a silo for that but i m not aware of any other **(BART)**	
	i m not sure if you can get that to work for you but i think it s a good **(HeterMPC)**	
	i think it s a bit of a corner case for people who do n t have the phased update **(MADNet)**	

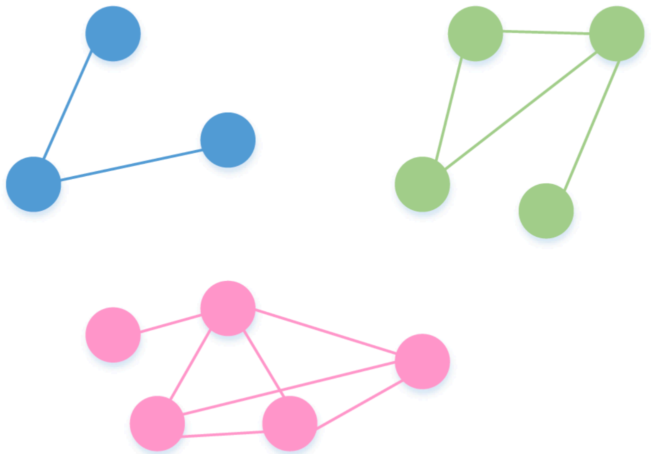
Challenge (1): Scarcity of addressee labels

Reduce the heavy dependency on the necessary addressee labels



(a) An MPC instance with a few addressee labels (@) missing

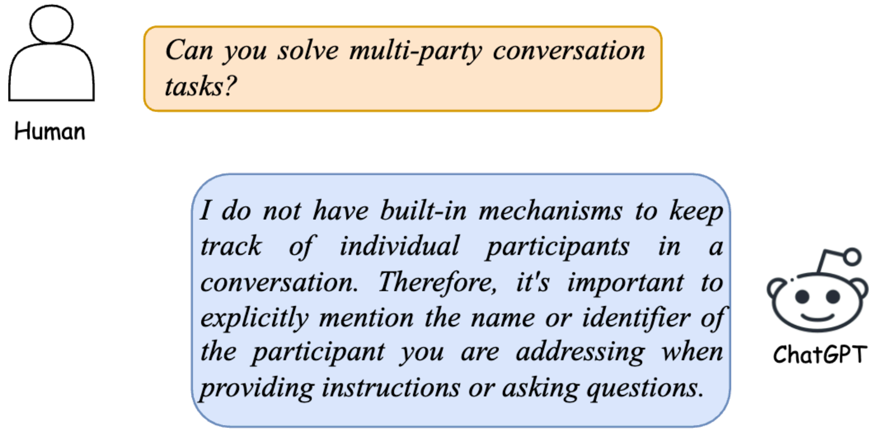
Addressees of 55% of the utterances in the Ubuntu dataset are not specified



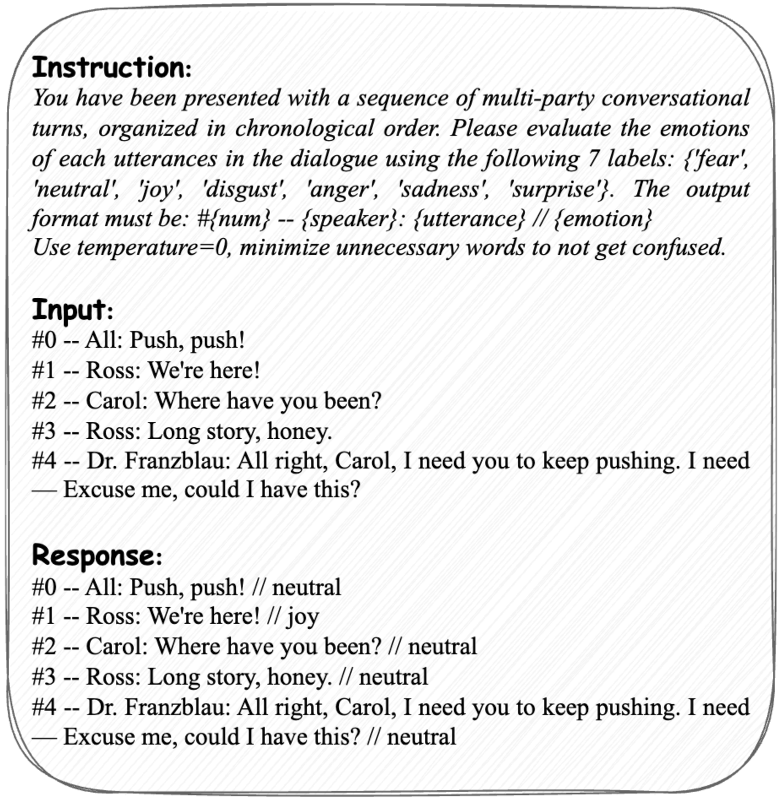
Fail to build a consecutively connected conversation graph, but only a few separate conversation fragments instead

Challenge (2): ChatMPC Solver

Is ChatGPT a Good Multi-Party Conversation Solver?



ChatGPT's performance on MPC tasks leaves much to be desired, whilst GPT-4's results portend a promising future


Instruction:
You have been presented with a sequence of multi-party conversational turns, organized in chronological order. Please evaluate the emotions of each utterances in the dialogue using the following 7 labels: {'fear', 'neutral', 'joy', 'disgust', 'anger', 'sadness', 'surprise'}. The output format must be: #{num} -- {speaker}: {utterance} // {emotion}
Use temperature=0, minimize unnecessary words to not get confused.

Input:
#0 -- All: Push, push!
#1 -- Ross: We're here!
#2 -- Carol: Where have you been?
#3 -- Ross: Long story, honey.
#4 -- Dr. Franzblau: All right, Carol, I need you to keep pushing. I need — Excuse me, could I have this?

Response:
#0 -- All: Push, push! // neutral
#1 -- Ross: We're here! // joy
#2 -- Carol: Where have you been? // neutral
#3 -- Ross: Long story, honey. // neutral
#4 -- Dr. Franzblau: All right, Carol, I need you to keep pushing. I need — Excuse me, could I have this? // neutral

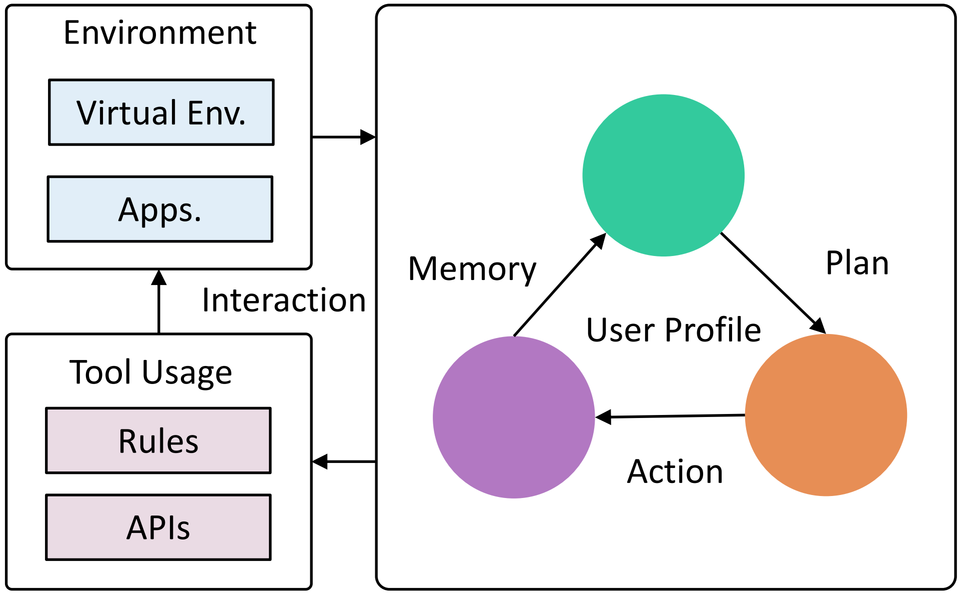
Chao-Hong Tan, Jia-Chen Gu, et al. *Is ChatGPT a Good Multi-Party Conversation Solver?* Findings of EMNLP 2023.

Challenge (3): Multi-modal MPC

Incorporate the face and speech interactions in conversations

Challenge (4): Virtual agent simulacra

Instantiate generative agents and empower multi-party interactions to simulate believable human behavior, e.g., collaborate, debate, and evolve



MPC in a Multi-Agent Perspective

Joon Sung Park, et al. *Generative Agents: Interactive Simulacra of Human Behavior*. arXiv 2023.

Open Questions

- Still don't know yet how to better model the core issues of interlocutor and conversation structure
- How to make use of the complicated and complementary interactions between interlocutors and utterances for universal MPC understanding
- Why performance drops significantly when its comes to topic transition and long context
- How to design communitive agents: multimodal environment interaction, multi-agent collaboration
- Lack of high-quality MPC datasets

Thanks! Q&A

Homepage: https://JasonForJoy.github.io

Contact: gujc@ucla.edu

Github: https://github.com/JasonForJoy