Who Speaks What to Whom

A Brief Introduction to Multi-party Dialogues

Reporter: Yiyang Li

Date: 2022/08/25

上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

目录

1 What are Multi-party Dialogues

2 Problem of *Who*

3 Problem of *To Whom*

4 Problem of *Speaks What*

5 Open Challenges

目录

1 What are Multi-party Dialogues

2 Problem of *Who*

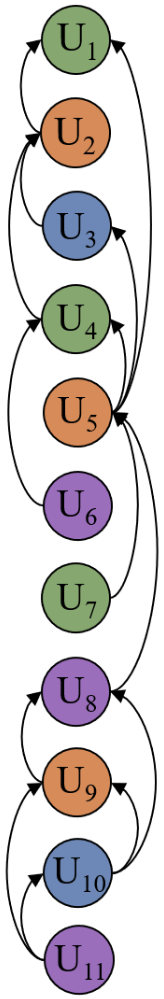
3 Problem of *To Whom*

4 Problem of *Speaks What*

5 Open Challenges

What are Multi-party Dialogues

- Definition:
  - Multi-party dialogues are those dialogues that involve at least three interlocuters, resulting in graph-structured reply-to relations and interleaving information flows.
- Typical scenarios with multi-party dialogues:
  - Group meetings (AMI dataset)
  - Daily conversations (Friends dataset)
  - Group/Forum chatting (Ubuntu/Twitter/Reddit datasets)
  - …
- Related tasks:
  - Response Generation/Selection
  - Discourse Parsing
  - Question Answering
  - …



U_1: [Monica Geller: *Tell him.*]
U_2: [Rachel Green: *No.*]
U_3: [Phoebe Buffay: *Tell him, tell him!*]
U_4: [Monica Geller: *Just… Please tell him.*]
U_5: [Rachel Green: *Shut up!*]
U_6: [**Chandler Bing**: *Tell me what?*]
U_7: [Monica Geller: *Look at you, you won't even look at him.*]
U_8: [**Chandler Bing**: *Oh, come on tell me. I could use another reason why women won't look at me.*]
U_9: [Rachel Green: *All right, all right. Last night, I had dream that, uh, you and I, were…*]
U_{10}: [Phoebe Buffay: *Dating **on this table**.*]
U_{11}: [Chandler Bing: *Wow!*]

Q_1: Who was with Rachel in her dream?
A_1: **Chandler Bing**
Q_2: Where did Rachel and Chandler date?
A_2: **On this table**

目录

1 What are Multi-party Dialogues

2 Problem of *Who*

3 Problem of *To Whom*

4 Problem of *Speaks What*

5 Open Challenges

Problem of *Who*

- This problem is also referred as speaker modeling, where we want to equip the model with the ability to understand who is speaking.
- Two ways of modeling speakers:
  - Explicit modeling:

    Adding speaker embeddings + pre-training [1]; Modeling inputs: #Speaker 1#: blablabla…;
  - Implicit Modeling:

    Pre-training/Multi-task-learning using speaker identification task. [2,3]

References:

[1] Speaker-Aware BERT for Multi-Turn Response Selection in Retrieval-Based Chatbots (CIKM 2020)

[2] MPC-BERT A Pre-Trained Language Model for Multi-Party Conversation Understanding (ACL 2021)

[3] Self- and Pseudo-self-supervised Prediction of Speaker and Key-utterance for Multi-party Dialogue Reading Comprehension (Findings of EMNLP2021)

目录

1 What are Multi-party Dialogues

2 Problem of *Who*

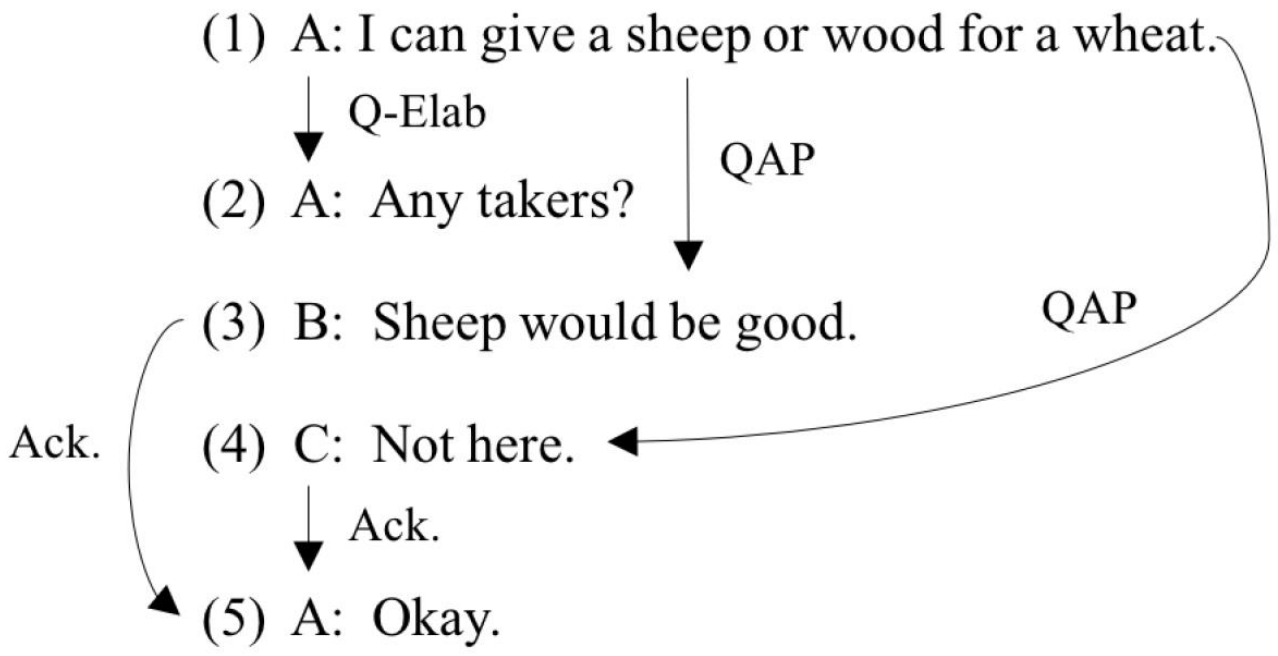
3 Problem of *To Whom*

4 Problem of *Speaks What*

5 Open Challenges

Problem of *To Whom*: Overview

- This problem is also referred as Addressee Prediction or Discourse Parsing, where we want to know the reply-to relations of the whole dialogue.



Problem of *To Whom*: Paper [4]

A Deep Sequential Model for Discourse Parsing on Multi-Party Dialogues

Zhouxing Shi, Minlie Huang*
Dept. of Computer Science & Technology, Tsinghua University, Beijing 100084, China
Institute for Artificial Intelligence, Tsinghua University (THUAI), China
Beijing National Research Center for Information Science and Technology, China
zhouxingshichn@gmail.com; aihuang@tsinghua.edu.cn

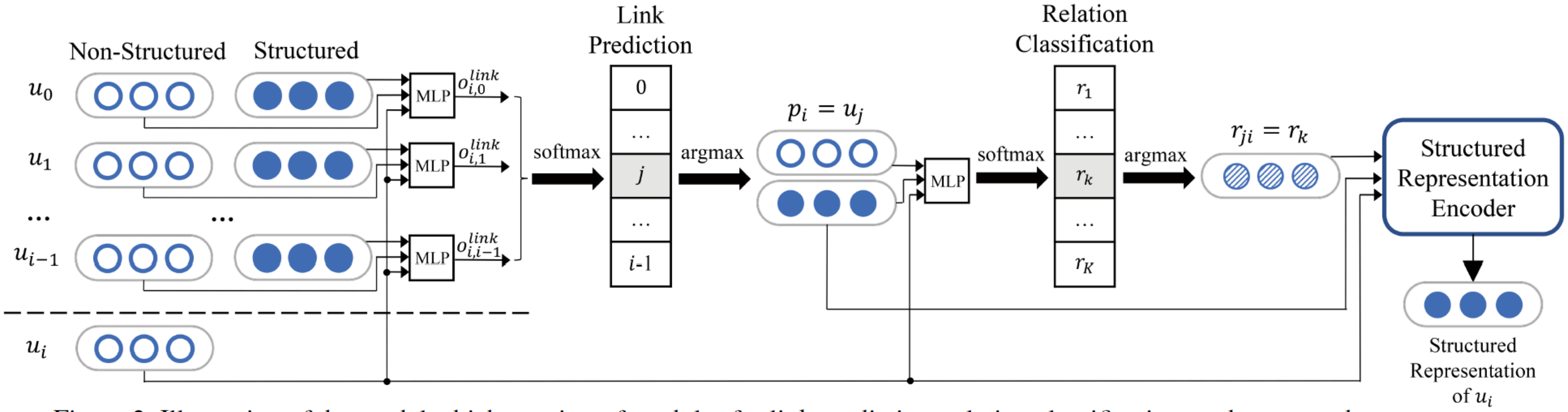


Figure 2: Illustration of the model which consists of modules for link prediction, relation classification, and structured representation encoding. For the current EDU u_i, link prediction estimates a distribution over its preceding EDUs, relation classification estimates a distribution over relation types, and the structured encoder updates the structured representation of u_i using representations of u_i and p_i and the embedding of the predicted relation type r_{ji}. Non-structured representation encoding is performed before the prediction process and is omitted from the illustration.

Problem of *To Whom*: Model

- Structured Representation Encoder:

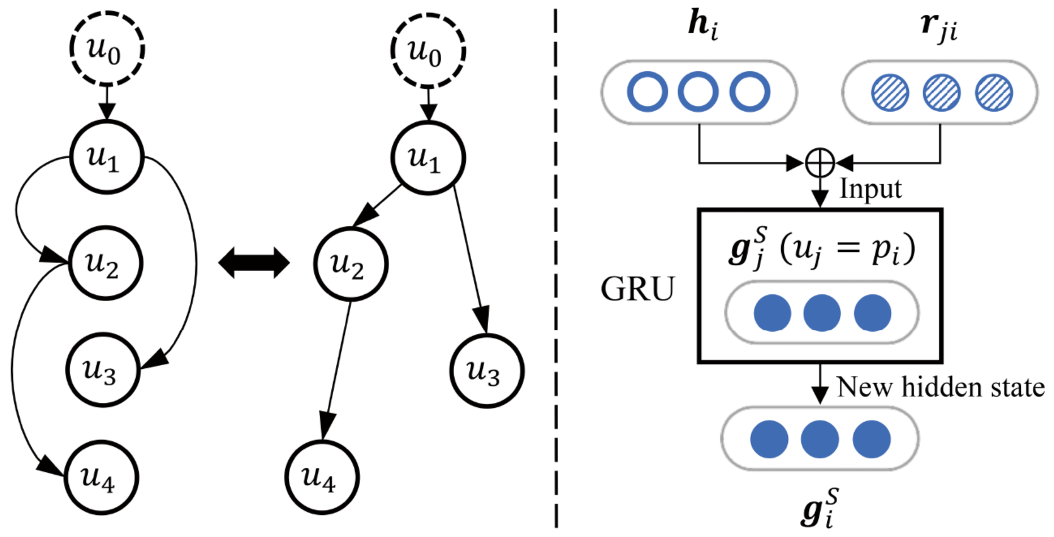


Figure 3: An example dependency tree (left) and the structured encoder (right), where $\boldsymbol{h}_i$ is the local representation of EDU u_i, $\boldsymbol{g}_i^S$ and $\boldsymbol{g}_j^S$ are structured representations, $\boldsymbol{r}_{ji}$ is the relation embedding, and $u_j = p_i$ is the parent of u_i.

- Speaker Highlighted Mechanism:

$$\boldsymbol{g}_{i,a}^S = \begin{cases} \mathbf{0} & i = 0 \\ \mathbf{GRU}_{hl}(\boldsymbol{g}_{j,a}^S, \boldsymbol{h}_i \oplus \boldsymbol{r}_{ji}) & a_i = a, i > 0 \\ \mathbf{GRU}_{gen}(\boldsymbol{g}_{j,a}^S, \boldsymbol{h}_i \oplus \boldsymbol{r}_{ji}) & a_i \neq a, i > 0 \end{cases} \quad (3)$$

where $\oplus$ denotes vector concatenation, **GRU** stands for the functions of a GRU cell, and $\boldsymbol{r}_{ji}$ denotes the embedding vector of relation type r_{ji}, and hl and gen are short for *highlighted* and *general* respectively.

- Fuse Information for Link/Relation Prediction:

For each EDU u_i, the link predictor predicts its parent node p_i and the relation classifier categorizes the corresponding relation type r_{ji} if $p_i = u_j$. For each EDU $u_j (j < i)$ that precedes u_i in the dialogue, we concatenate the representations $\boldsymbol{h}_i, \boldsymbol{g}_i^{NS}, \boldsymbol{g}_j^{NS}, \boldsymbol{g}_{j,a_i}^S$ to obtain an input vector $\boldsymbol{H}_{i,j}$ for link prediction and relation classification:

$$\boldsymbol{H}_{i,j} = \boldsymbol{h}_i \oplus \boldsymbol{g}_i^{NS} \oplus \boldsymbol{g}_j^{NS} \oplus \boldsymbol{g}_{j,a_i}^S \quad (4)$$

Problem of *To Whom*: Model

- Link Prediction:

$$L_{i,j}^{link} = \tanh(\boldsymbol{W}_{link} \cdot \boldsymbol{H}_{i,j} + \boldsymbol{b}_{link}) \quad (5)$$

$$o_{i,j}^{link} = \boldsymbol{U}_{link} \cdot \boldsymbol{L}_{i,j}^{link} + b'_{link} \quad (6)$$

$$P(p_i = u_j | \boldsymbol{H}_{i,<i}) = \frac{exp(o_{i,j}^{link})}{\sum_{k<i} exp(o_{i,k}^{link})} \quad (7)$$

$$p_i = \operatorname*{argmax}_{u_j : j<i} P(p_i = u_j | \boldsymbol{H}_{i,<i}) \quad (8)$$

- Relation Prediction:

$$\boldsymbol{L}_{i,j}^{rel} = \tanh(\boldsymbol{W}_{rel} \cdot \boldsymbol{H}_{i,j} + \boldsymbol{b}_{rel}) \quad (9)$$

$$P(r | \boldsymbol{H}_{i,j}) = softmax(\boldsymbol{U}_{rel} \cdot \boldsymbol{L}_{i,j}^{rel} + \boldsymbol{b}'_{rel}) \quad (10)$$

- Loss Functions:

We adopt the negative log-likelihood of the training data as the loss function:

$$L_{link}(\Theta) = -\sum_{d \in \mathcal{D}} \sum_{i=1}^{n} \log P(p_i = p_i^* | \boldsymbol{H}_{i,<i}) \quad (11)$$

$$L_{rel}(\Theta) = -\sum_{d \in \mathcal{D}} \sum_{i=1}^{n} \log P(r_{ji} = r_{ji}^* | \boldsymbol{H}_{i,j}, u_j = p_i^*) \quad (12)$$

$$L_{all}(\Theta) = L_{link}(\Theta) + L_{rel}(\Theta) \quad (13)$$

where Θ is the set of parameters to be optimized, $\mathcal{D}$ is the training data, d is a dialogue in $\mathcal{D}$, p_i^* and r_{ji}^* are the golden parent and the corresponding golden relation type respectively.

Problem of *To Whom*: Experiment

- Dataset:

  STAC Corpus (Asher et al. 2016): 1,062 dialogues, a small dataset

- Experimental Results:

Model	Link	Link & Rel
MST	68.8	50.4
ILP	68.6	52.1
Deep+MST	69.6	52.1
Deep+ILP	69.0	53.1
Deep+Greedy	69.3	51.9
Deep Sequential (shared)	72.1	54.7
Deep Sequential	**73.2**	**55.7**

Table 1: F_1 scores (%) for different models. *Link* means link prediction; and *Link & Rel* means that a correct prediction must predict dependency link and relation type correctly at the same time.

Model	Link	Link & Rel
Deep+Greedy	69.3	51.9
Deep Sequential (NS)	71.0	53.7
Deep Sequential (Random)	71.8	53.7
Deep Sequential (w/o SHM)	71.7	54.5
Deep Sequential	**73.2**	**55.7**

Table 2: F_1 scores (%) for different models.

Problem of *To Whom*: Limitations

- Though using previously predicted structure can provide richer information for modeling structures, it can also lead to problems with severe error propagation.
- To alleviate error propagation, Wang et al. [5] adopt an edge-centric graph neural network to update the information between each utterance pair layer by layer, so that expressive representations can be learned without historical predictions.

References:

[4] A Deep Sequential Model for Discourse Parsing on Multi-Party Dialogues (AAAI 2019)

[5] A Structure Self-aware Model for Discourse Parsing on Multi-party Dialogues (IJCAI 2021)

Problem of *To Whom*: Benefits

- The parsing results can be used to enhance multi-party dialogue encoding on both generative and understanding tasks [6, 9, 10].

- This can also give us insights of modeling graph-structured or semi-structured data by using the parsing results.
  - We can enhance a language model using semantic parsing results. [7]
  - We can model programming languages using the parsed AST (Abstract Syntax Tree) obtained from a compiler. [8]
  - …

References:

[6] Multi-Party Empathetic Dialogue Generation: A New Task for Dialog Systems

[7] Semantics-Aware BERT for Language Understanding (AAAI 2020)

[8] GraphCodeBERT: Pre-training Code Representations with Data Flow (ICLR 2021)

目录

1 What are Multi-party Dialogues

2 Problem of *Who*

3 Problem of *To Whom*

4 Problem of *Speaks What*

5 Open Challenges

Problem of *Speaks What*: Papers [9, 10]

- This problem is also referred as Response Generation/Selection for multi-party dialogues.
- Today we focus on response generation, which is the direction I am investigating recently.
- Briefly introduce two papers today.

GSN: A Graph-Structured Network for Multi-Party Dialogues

Wenpeng Hu[1,3,*], **Zhangming Chan**[2,3,*], **Bing Liu**[4,†], **Dongyan Zhao**[2,3], **Jinwen Ma**[1] and **Rui Yan**[2,3,†]
[1]Department of Information Science, School of Mathematical Sciences, Peking University
[2]Center for Data Science, Peking University
[3]Institute of Computer Science and Technology, Peking University
[4]Department of Computer Science, University of Illinois at Chicago
{wenpeng.hu,zhangming.chan,zhaody,ruiyan}@pku.edu.cn, liub@uic.edu, jwma@math.pku.edu.cn

HETERMPC: A Heterogeneous Graph Neural Network for Response Generation in Multi-Party Conversations

Jia-Chen Gu[1*†], **Chao-Hong Tan**[1†], **Chongyang Tao**[2], **Zhen-Hua Ling**[1], **Huang Hu**[2], **Xiubo Geng**[2], **Daxin Jiang**[2‡]
[1]National Engineering Research Center for Speech and Language Information Processing, University of Science and Technology of China, Hefei, China
[2]Microsoft, Beijing, China
{gujc,chtan}@mail.ustc.edu.cn, zhling@ustc.edu.cn,
{chotao,huahu,xigeng,djiang}@microsoft.com

References:

[9] GSN: A Graph-Structured Network for Multi-Party Dialogues (IJCAI 2019)

[10] HeterMPC A Heterogeneous Graph Neural Network for Response Generation in Multi-Party Conversations (ACL 2022)

Problem of *Speaks What*: GSN - Overview

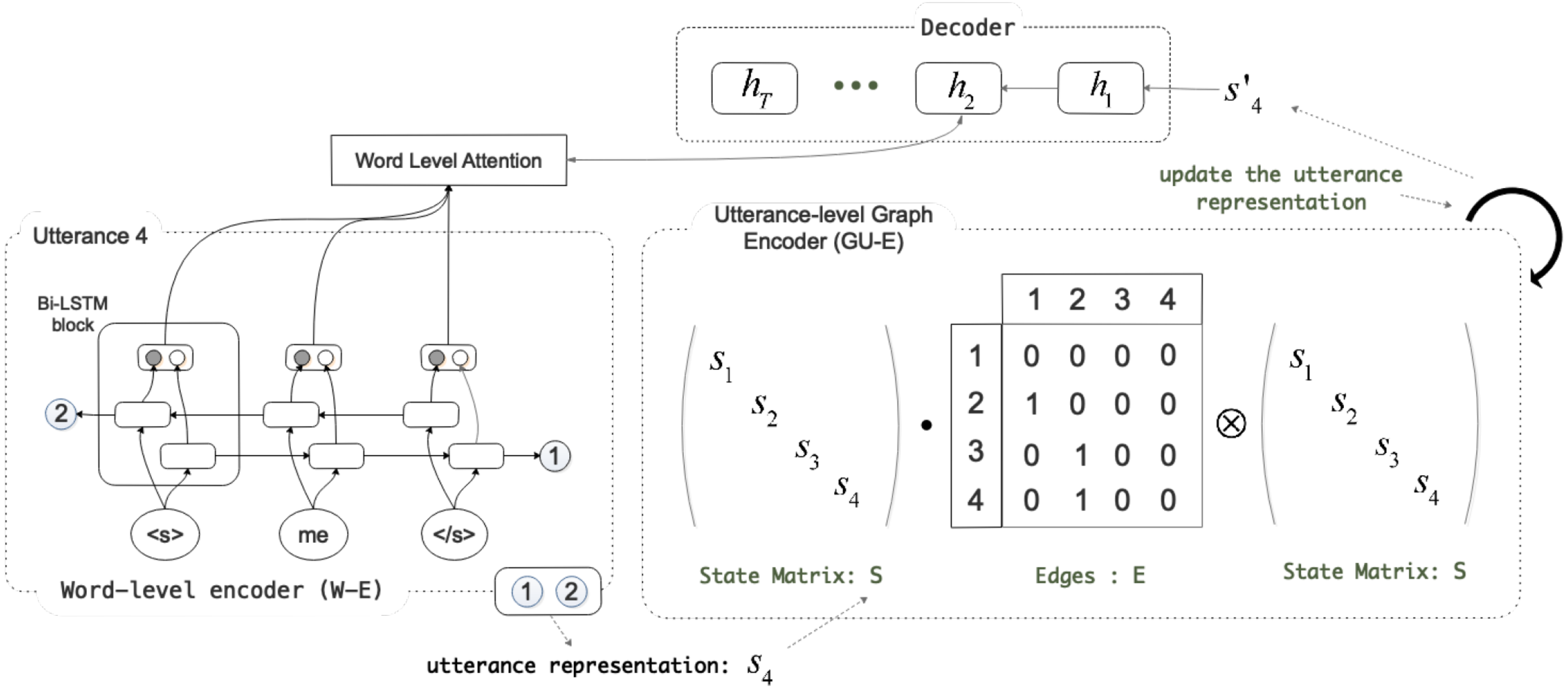


Figure 2: Architecture of GSN.

- Word-level Encoder:
  - Just a Bi-LSTM.
  - Last hidden states as utterance representations.
- Utterance-level Graph Encoder:
  - A graph neural network with a weighted updating mechanism.
- Decoder:
  - A GRU with cross attention to the output of the encoder

Problem of *Speaks What*: GSN - Graph Encoder

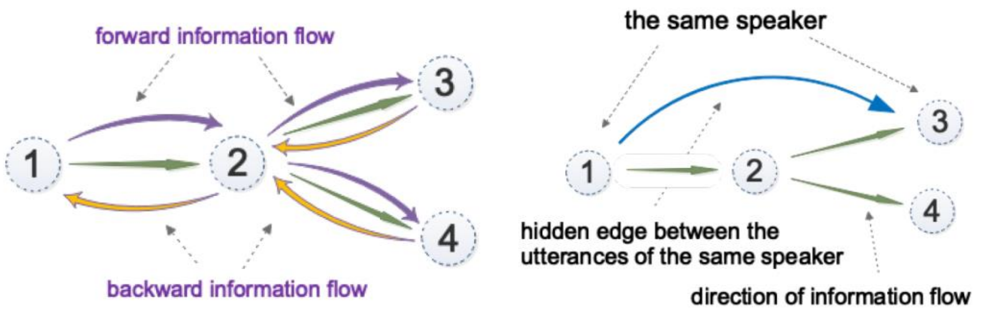
UG-E's basic operation is illustrated in Figure 3. For example, given a session $\mathbf{S} = (\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \mathbf{s}_4)$, in the l-th iteration, the state of the i-th utterance can be calculated by:

$$\begin{aligned} \mathbf{s}_i^l &= \mathbf{s}_i^{l-1} + \eta \cdot \Delta\mathbf{s}_{I|i}^{l-1} \\ \Delta\mathbf{s}_{I|i}^{l-1} &= \sum_{i' \in \varphi} \Delta\mathbf{s}_{i'|i}^{l-1} \end{aligned} \tag{3}$$

where φ is the collection of preceding nodes of the current node i in the direction of the information flow; $\Delta\mathbf{s}_{I|i}^{l-1}$ is the updating information, which is calculated by Eq. 5 below

We use a non-linear "squashing" function (i.e., SQH$(\cdot)$) to give vectors with a small norm a weight close to α, but a large norm a weight close to 1:

$$\eta = \mathrm{SQH}(\Delta\mathbf{s}_{I|i}^{l-1}) = \frac{\alpha + ||\Delta\mathbf{s}_{I|i}^{l-1}||}{1 + ||\Delta\mathbf{s}_{I|i}^{l-1}||} \tag{4}$$



(a) Bi-directional information flow. (b) Speaker information modeling.

Figure 4: Information flow.

$$\Delta\mathbf{s}_{i'|i}^{l-1} = \mathbf{s}_{i'}^{l-1} \otimes \mathbf{s}_i^{l-1} \tag{5}$$

where '$\otimes$', the *update operator*, computes the updating information. Inspired by the updating operation hidden in Gated Recurrent Units (GRU) [Cho *et al.*, 2014], $\otimes$ is defined as:

$$\begin{aligned} \Delta\mathbf{s}_{i'|i}^{l-1} &= (1 - \mathbf{x}_i) * \mathbf{s}_{i'}^{l-1} + \mathbf{x}_i * \tilde{\mathbf{h}}_i \\ \tilde{\mathbf{h}}_i &= \tanh(\mathbf{W} \cdot [\mathbf{r}_i * \mathbf{s}_{i'}^{l-1}, \mathbf{s}_i^{l-1}]) \\ \mathbf{x}_i &= \sigma(\mathbf{W}_x \cdot [\mathbf{s}_{i'}^{l-1}, \mathbf{s}_i^{l-1}] \\ \mathbf{r}_i &= \sigma(\mathbf{W}_r \cdot [\mathbf{s}_{i'}^{l-1}, \mathbf{s}_i^{l-1}] \end{aligned} \tag{6}$$

Problem of *Speaks What*: GSN - Experiments

- Dataset:
  - The Ubuntu IRC Benchmark, constructed by extracting all utterances with response relations indicated by the "@" symbol in the corpus.
  - 370k dialogues for training, 5k for validation and testing, respectively.
- Experimental Results:

Model	BLEU 1	BLEU 2	BLEU 3	BLEU 4	METEOR	ROUGE$_L$
seq2seq	10.45	4.13	2.08	1.02	3.43	9.67
seq2seq W-speaker	10.70	4.98	2.20	1.55	3.92	9.42
Seq2seq (last utte)	9.85	3.04	1.38	0.67	3.98	8.34
HRED [Serban *et al.*, 2016]	10.80	4.60	2.54	1.42	4.38	10.23
HRED W-speaker	11.23	4.82	3.06	1.64	4.36	10.98
GSN No-speaker (1-iter)	9.42	3.05	1.61	0.95	3.74	7.63
GSN No-speaker (2-iter)	12.06	4.87	2.80	1.70	4.32	10.09
GSN No-speaker (3-iter)	12.77▲	5.37▲	3.17	1.99▲	4.53	10.80
GSN W-speaker (1-iter)	10.31	4.06	2.34	1.45	3.88	9.96
GSN W-speaker (2-iter)	12.77	4.93	2.61	1.46	4.79	11.34
GSN W-speaker (3-iter)	**13.50**▲	**5.63**▲	**3.24**▲	**1.99**▲	**4.85**▲	**11.36**▲

Human	HRED	No-speaker		W-speaker	
		1-iter	3-iter	1-iter	3-iter
3.01	1.91	1.89	1.98	2.23▲	**2.37**▲

Table 4: Human evaluation results. ▲denotes p-value < 0.01 in paired t-test against HRED. The perfect score is 4.

Problem of *Speaks What*: HeterMPC - Model

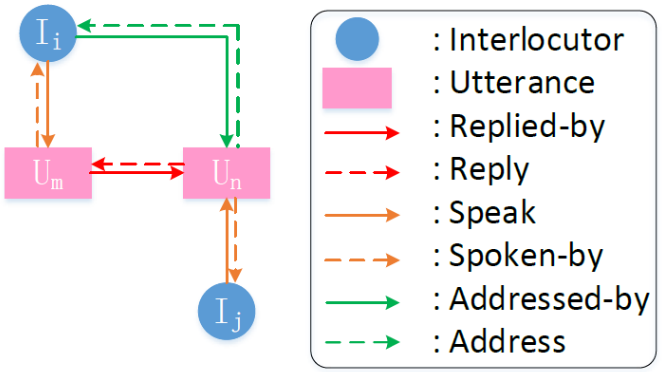


Figure 2: Illustration of the two types of nodes and six types of edges in a heterogeneous conversation graph. This example demonstrates that I_j speaks U_n replying U_m that is spoken-by I_i.

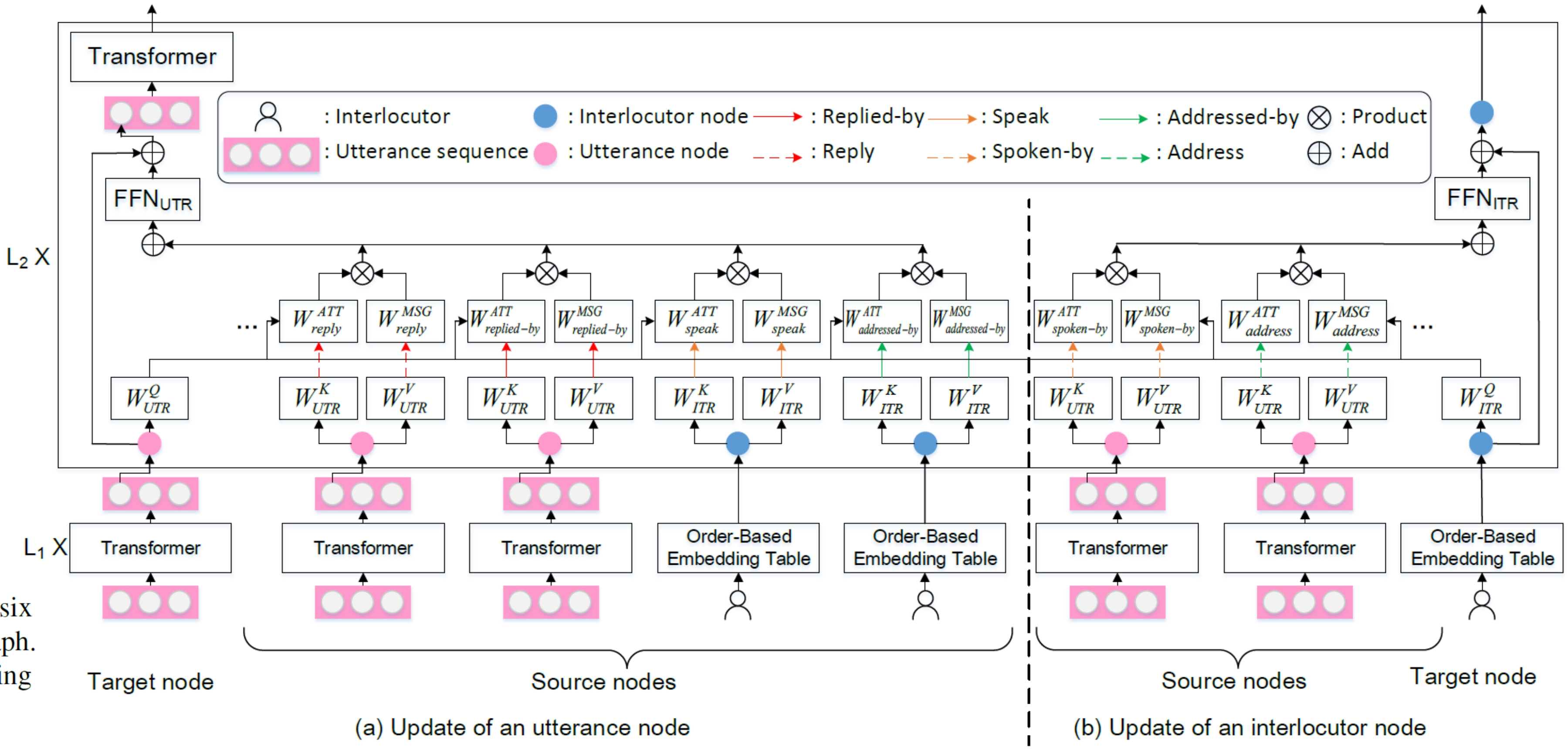


Figure 3: Model architecture of HeterMPC for (a) update of an utterance node and (b) update of an interlocutor node. "UTR" and "ITR" are abbreviations of "utterance" and "interlocutor" respectively. The set of W^*_* denotes the node-edge-type-dependent parameters.

Problem of *Speaks What*: HeterMPC - Experiments

Models \ Metrics	BLEU-1	BLEU-2	BLEU-3	BLEU-4	METEOR	ROUGE_L
Seq2Seq (LSTM) (Sutskever et al., 2014)	7.71	2.46	1.12	0.64	3.33	8.68
Transformer (Vaswani et al., 2017)	7.89	2.75	1.34	0.74	3.81	9.20
GSN (Hu et al., 2019b)	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
BERT (Devlin et al., 2019)	10.90	3.85	1.69	0.89	4.18	9.80
HeterMPC_{BERT}	**12.61**	**4.55**	**2.25**	**1.41**	**4.79**	**11.20**
HeterMPC_{BERT} w/o. node types	11.76	4.09	1.87	1.12	4.50	10.73
HeterMPC_{BERT} w/o. edge types	12.02	4.27	2.10	1.30	4.74	10.92
HeterMPC_{BERT} w/o. node and edge types	11.60	3.98	1.90	1.18	4.20	10.63
HeterMPC_{BERT} w/o. interlocutor nodes	11.80	3.96	1.75	1.00	4.31	10.53
BART (Lewis et al., 2020)	11.25	4.02	1.78	0.95	4.46	9.90
HeterMPC_{BART}	**12.26**	**4.80**	**2.42**	**1.49**	**4.94**	**11.20**
HeterMPC_{BART} w/o. node types	11.22	4.06	1.87	1.04	4.57	10.63
HeterMPC_{BART} w/o. edge types	11.52	4.27	2.05	1.24	4.78	10.90
HeterMPC_{BART} w/o. node and edge types	10.90	3.90	1.79	1.01	4.52	10.79
HeterMPC_{BART} w/o. interlocutor nodes	11.68	4.24	1.91	1.03	4.79	10.65

Table 1: Performance of HeterMPC and ablations on the test set in terms of automated evaluation. Numbers in bold denote that the improvement over the best performing baseline is statistically significant (t-test with p-value < 0.05).

Models \ Metrics	Score	Kappa
Human	2.81	0.55
GSN (Hu et al., 2019b)	2.00	0.50
BERT (Devlin et al., 2019)	2.19	0.42
BART (Lewis et al., 2020)	2.24	0.44
HeterMPC_{BERT}	2.39	0.50
HeterMPC_{BART}	2.41	0.45

Table 2: Human evaluation results of HeterMPC and some selected systems on a randomly sampled test set.

目录

1 What are Multi-party Dialogues

2 Problem of *Who*

3 Problem of *To Whom*

4 Problem of *Speaks What*

5 Open Challenges

Open Challenges

- Shortage of Addressee Labels:
  - The current ways of modeling multi-party dialogues, especially those that utilize the reply-to relations to construct graphs, require explicit addressee annotations. However, these annotations are hard to obtain in most multi-party datasets.
  - Under this circumstance, the pre-training of both generative and understanding tasks of multi-party dialogues is hindered.
  - How to subtly solve the shortage of addressee labels remains an open question.

- Universal Multi-party Dialogue Understanding:
  - Design better supervised or self-supervised tasks to equip the model with more abilities to model the (speaker, addressee, utterance) triplets of multi-party dialogues.
  - Design better model architectures that can effectively and efficiently capture the intrinsic characteristics of multi-party dialogues.

Thank you for listening

Q&A