Incorporating Commonsense Knowledge into Abstractive Dialogue Summarization via Heterogeneous Graph Networks

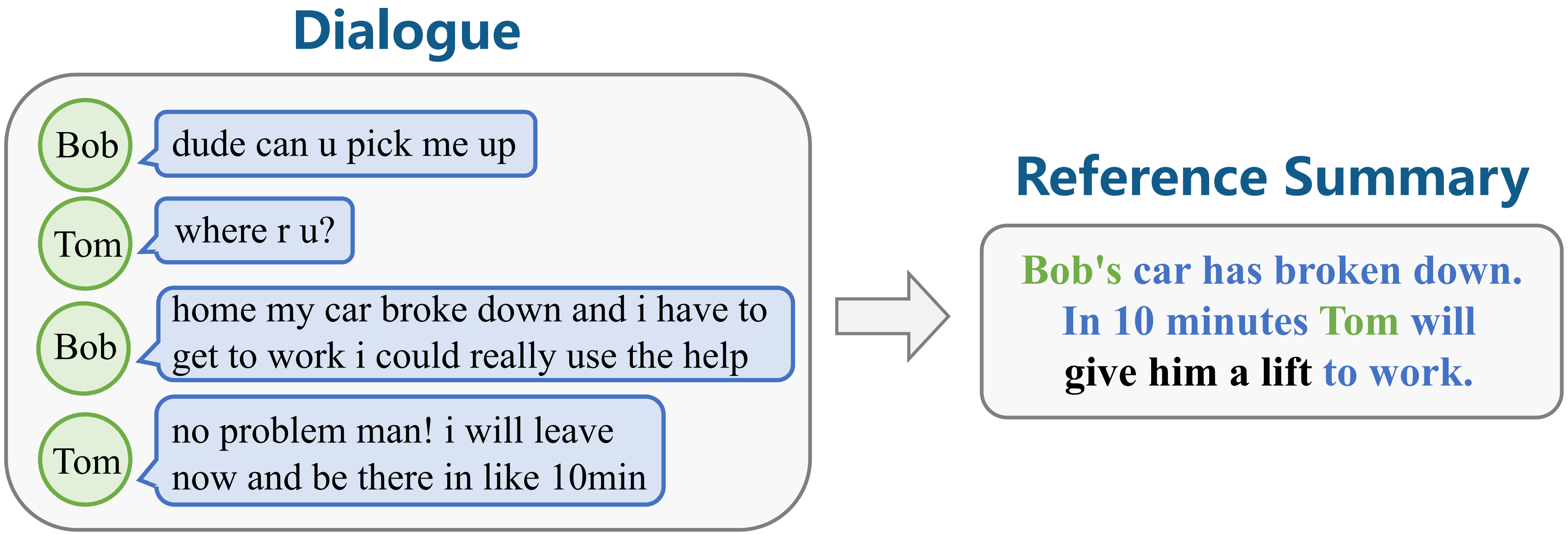
Xiachong Feng[1], Xiaocheng Feng[1,2], Bing Qin[1,2]

[1] Harbin Institute of Technology, [2] Peng Cheng Laboratory

Task Introduction

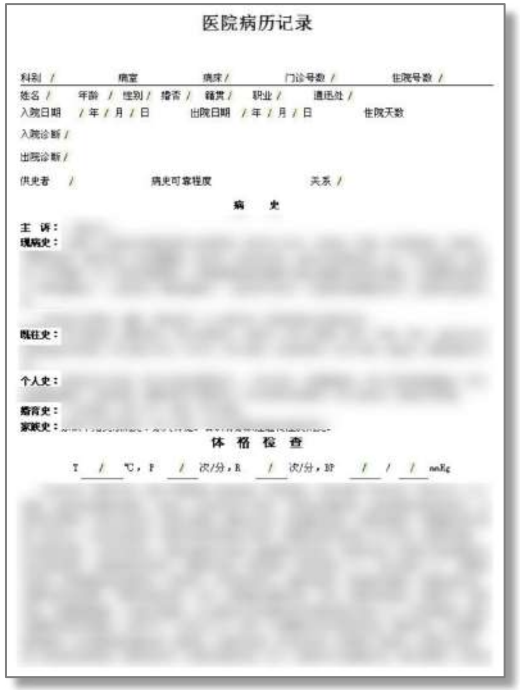
Dialogue Summarization

- **Dialogue summarization** aims to generate a succinct summary while retaining essential information of the dialogue.



Dialogue Summarization


会 议 纪 要


医院病历记录

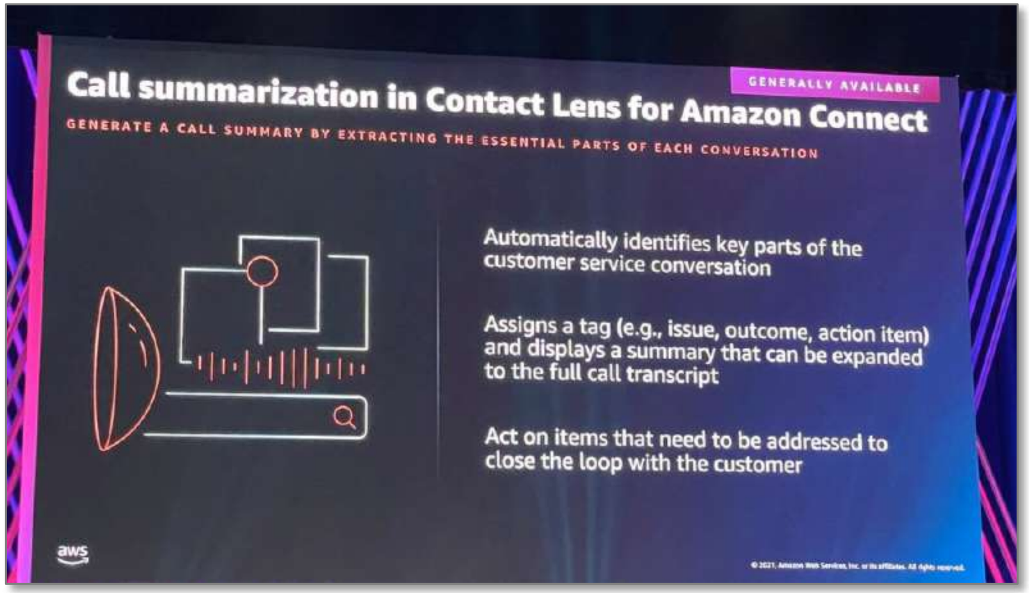
科别 / 病室 病床 / 门诊号数 / 住院号数 /
姓名 / 年龄 / 性别 / 婚否 / 籍贯 / 职业 / 通讯处 /
入院日期 / 年 / 月 / 日 出院日期 / 年 / 月 / 日 住院天数
入院诊断 /
出院诊断 /
供史者 / 病史可靠程度 关系 /

病 史

主 诉：
现病史：
既往史：
个人史：
婚育史：
家族史：

体 格 检 查

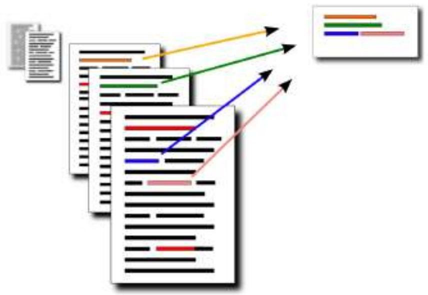
T / ℃，P / 次/分，R / 次/分，BP / / / mmHg



Challenges and Motivation

Challenges

Summarization

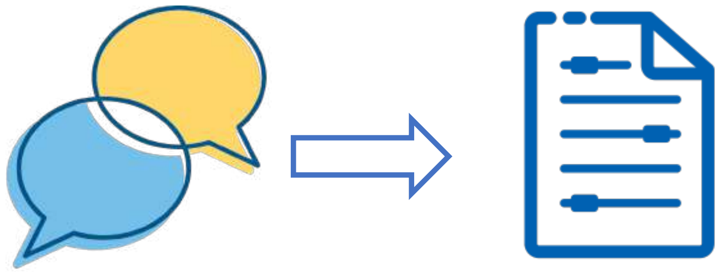


- Key Contents
- Coherence
- Abstractive
- Factual

Dialogue Modeling

- Multi-party
- Structure
- Topic Drift
- Coreference

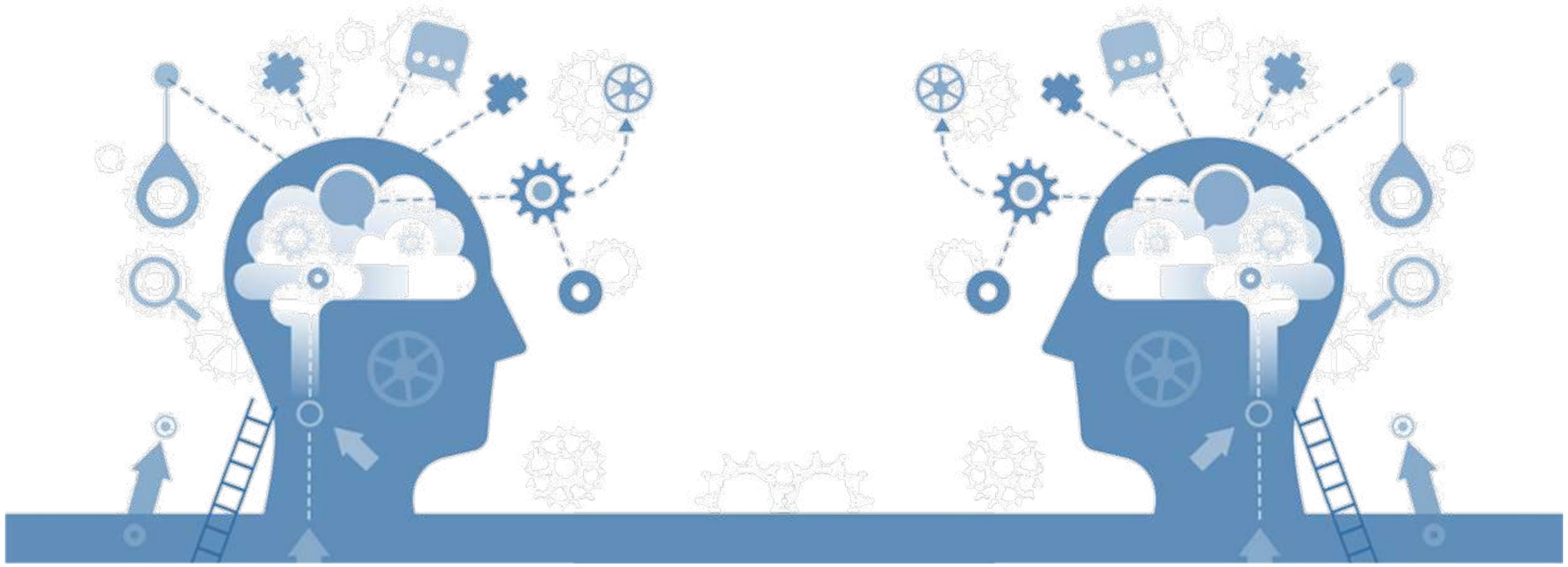
Dialogue Summarization



- Data Resource
- **Dialogue Modeling**
- Domain Specific

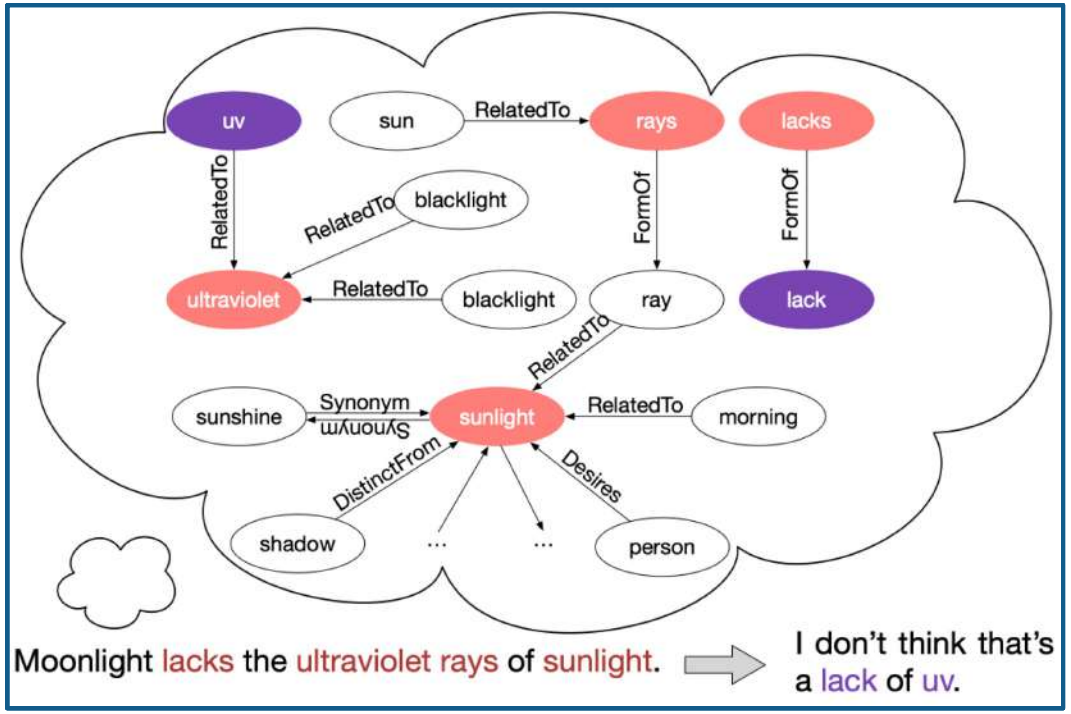
Commonsense Knowledge

- **Commonsense** is sound, practical judgment concerning everyday matters, or a basic ability to perceive, understand, and judge in a manner that is shared by (i.e. *common to*) nearly all people

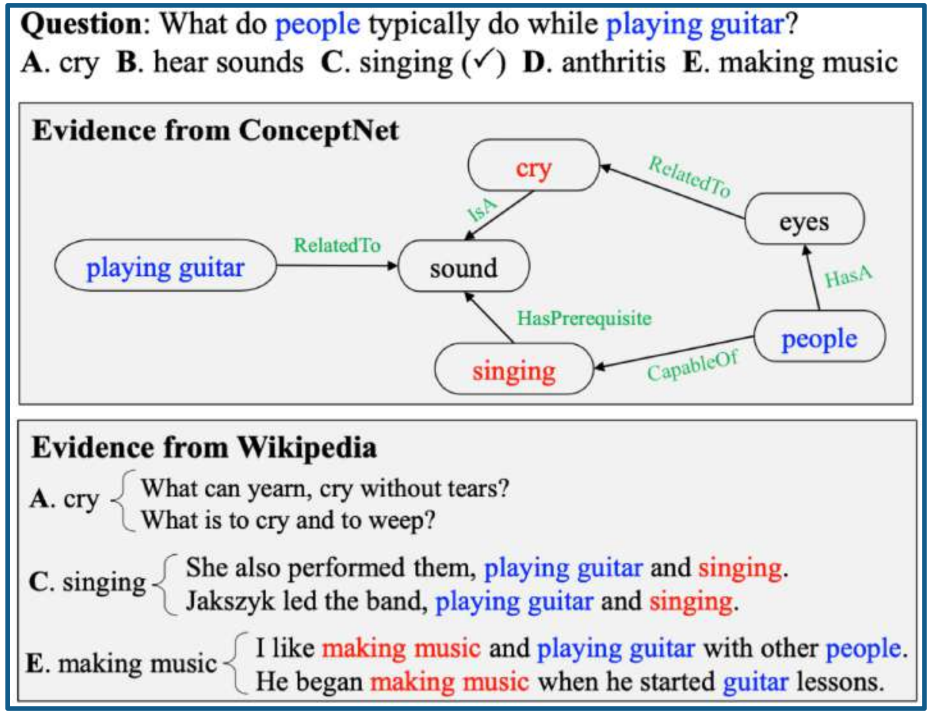


https://en.wikipedia.org/wiki/Common_sense

Commonsense For NLP



Response Generation

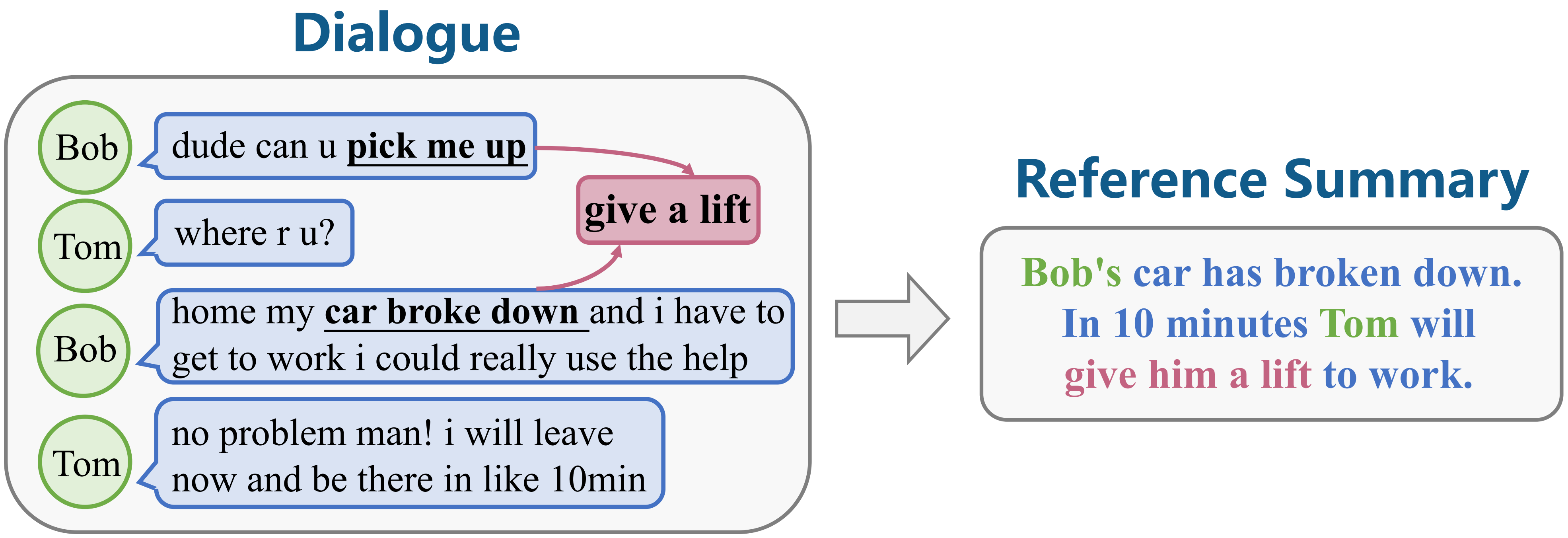


Question Answering

Zhou H, Young T, Huang M, et al. Commonsense knowledge aware conversation generation with graph attention
Lv S, Guo D, Xu J, et al. Graph-based reasoning over heterogeneous external knowledge for commonsense question answering

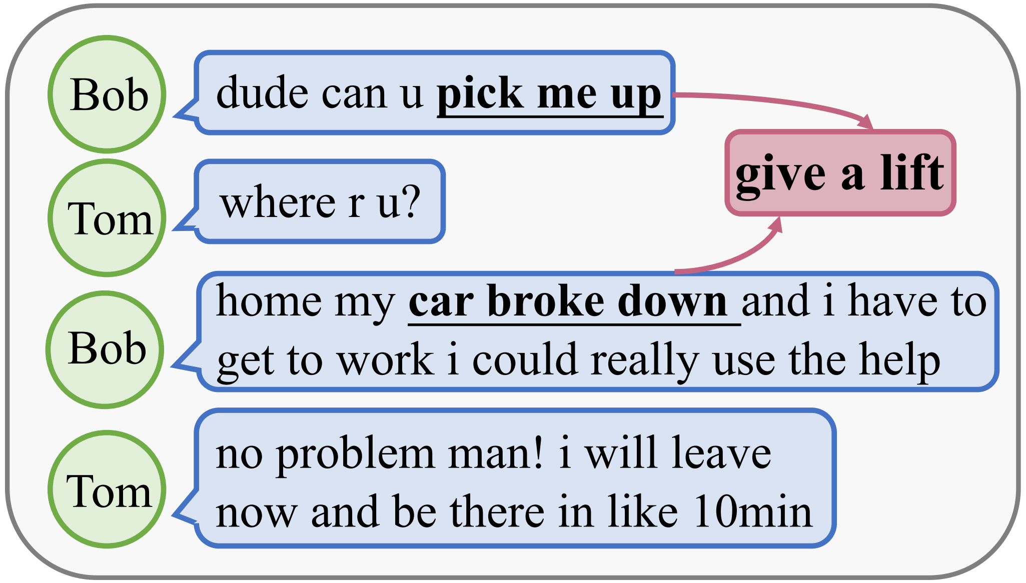
Commonsense for Dialogue Summarization

- By introducing commonsense knowledge according to the **pick up** and **car broke down**, we can know that **Bob expects Tom to give him a lift**.
- Commonsense knowledge can serve as a bridge **between non-adjacent utterances** that can help the model better understanding the dialogue.



Two Research Problems

Dialogue



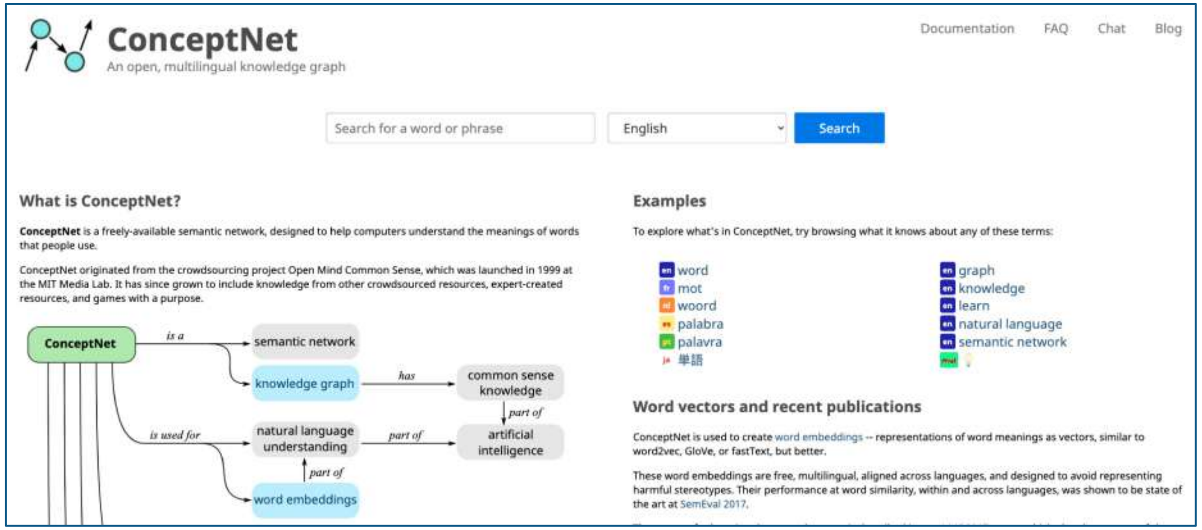
How to incorporate commonsense knowledge?

How to model the dialogue with commonsense?

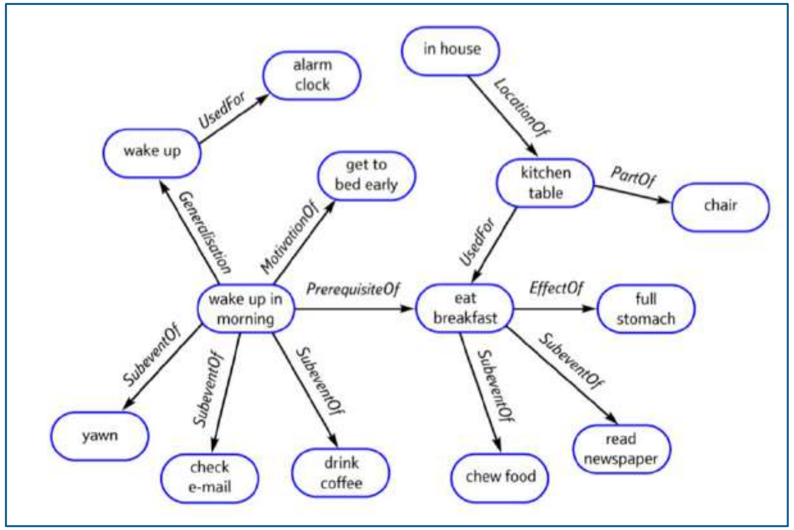
Heterogeneous Dialogue Graph Construction

Background: ConceptNet

- A freely-available large-scale commonsense knowledge base
- Includes **words** and **common phrases.**
- ConceptNet is a knowledge graph that connects words and phrases of natural language **(terms)** with labeled, weighted edges **(assertions).**



Homepage



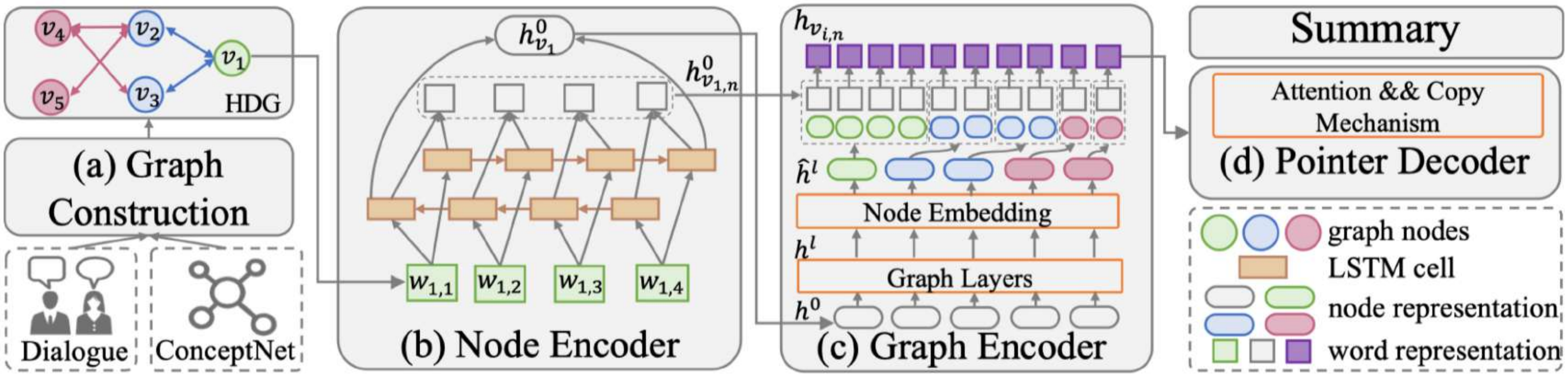
Example

Heterogeneous Dialogue Graph Construction

Dialogue Heterogeneous Graph Network

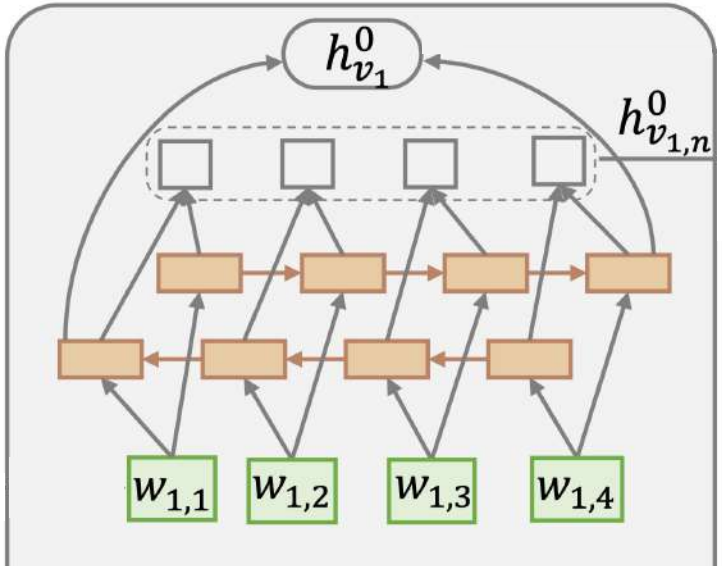
Overall

- **Graph construction** receives a dialogue and ConceptNet and outputs a heterogeneous dialogue graph .
- **Node encoder** receives a sequence of words for a node and produces initial node and word representations.
- **Graph encoder** conducts graph operations for initial node representations.
- **Pointer decoder** either generate summary words from the vocabulary or copy from the input words.



Node Encoder

- The role of **node encoder** is to give each node $v_i \in \mathcal{V}$ an initial representation.



(b) Node Encoder

$$\overrightarrow{h_n} = \mathrm{LSTM}_f\left(x_n, \overrightarrow{h_{n-1}}\right)$$

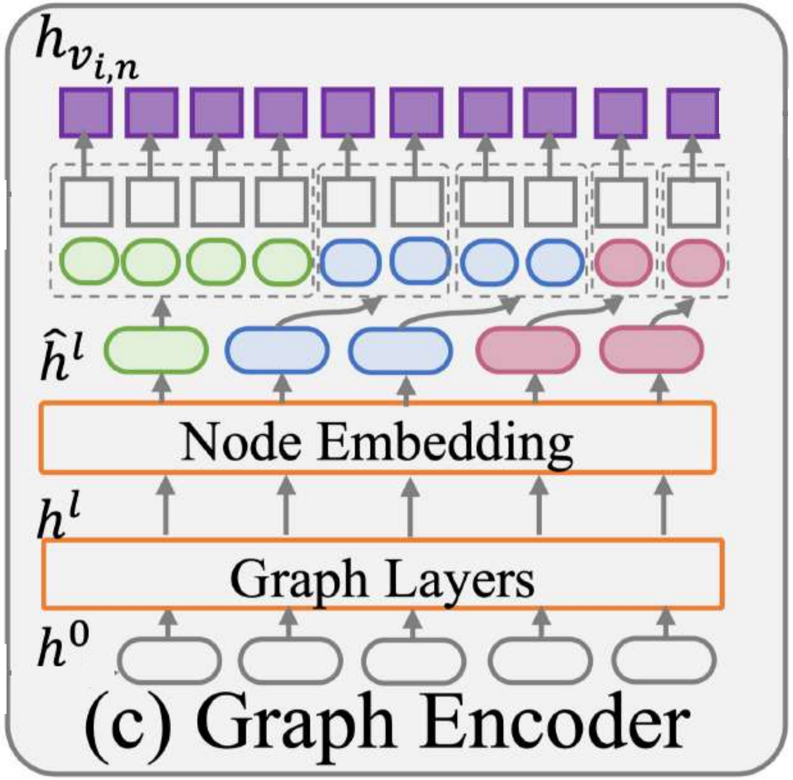
$$\overleftarrow{h_n} = \mathrm{LSTM}_b\left(x_n, \overleftarrow{h_{n+1}}\right)$$

$$h^0_{v_i} = [\overrightarrow{h_{|v_i|}}; \overleftarrow{h_1}]$$

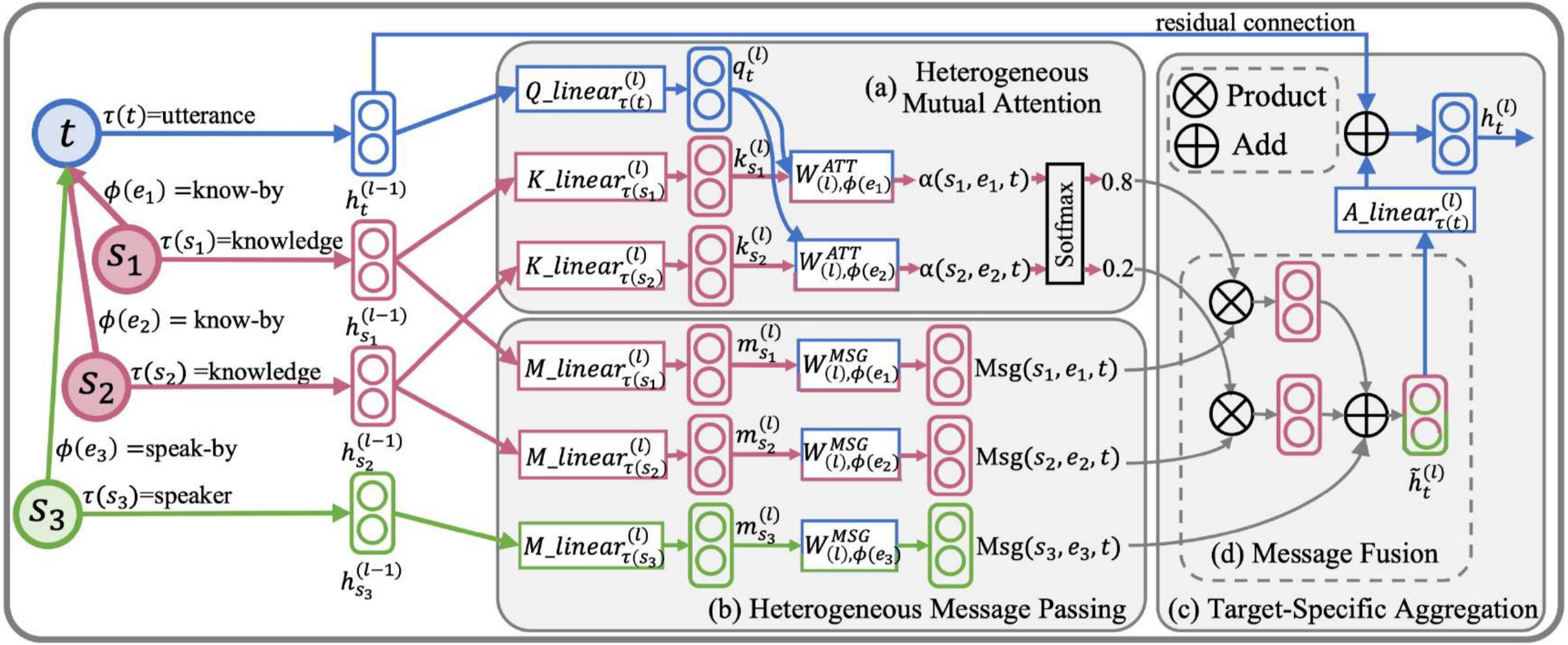
$$h^0_{v_i,n} = [\overrightarrow{h_n}; \overleftarrow{h_n}]$$

Graph Encoder

- **Graph encoder** is used to digest the structural information and get updated node representations. We employ *Heterogeneous Graph Transformer* (Hu et al., 2020) as our graph encoder.

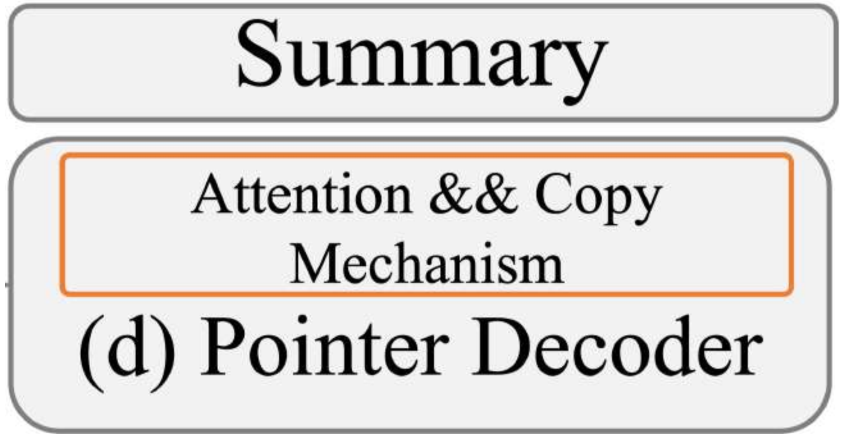


Graph Layer


residual connection
t
$\tau(t)$=utterance
$\phi(e_1)$ =know-by
s_1
$\tau(s_1)$=knowledge
$\phi(e_2)$ = know-by
s_2
$\tau(s_2)$ =knowledge
$\phi(e_3)$ =speak-by
s_3
$\tau(s_3)$=speaker
$h_t^{(l-1)}$
$h_{s_1}^{(l-1)}$
$h_{s_2}^{(l-1)}$
$h_{s_3}^{(l-1)}$
$Q\_linear_{\tau(t)}^{(l)}$
$q_t^{(l)}$
(a) Heterogeneous Mutual Attention
$K\_linear_{\tau(s_1)}^{(l)}$
$k_{s_1}^{(l)}$
$W_{(l),\phi(e_1)}^{ATT}$
$\alpha(s_1, e_1, t)$
$K\_linear_{\tau(s_2)}^{(l)}$
$k_{s_2}^{(l)}$
$W_{(l),\phi(e_2)}^{ATT}$
$\alpha(s_2, e_2, t)$
Sotfmax
0.8
0.2
$M\_linear_{\tau(s_1)}^{(l)}$
$m_{s_1}^{(l)}$
$W_{(l),\phi(e_1)}^{MSG}$
Msg(s_1, e_1, t)
$M\_linear_{\tau(s_2)}^{(l)}$
$m_{s_2}^{(l)}$
$W_{(l),\phi(e_2)}^{MSG}$
Msg(s_2, e_2, t)
$M\_linear_{\tau(s_3)}^{(l)}$
$m_{s_3}^{(l)}$
$W_{(l),\phi(e_3)}^{MSG}$
Msg(s_3, e_3, t)
(b) Heterogeneous Message Passing
Product
Add
$h_t^{(l)}$
$A\_linear_{\tau(t)}^{(l)}$
$\tilde{h}_t^{(l)}$
(d) Message Fusion
(c) Target-Specific Aggregation

Pointer Decoder

- We employ a LSTM with attention and copy mechanism to generate summaries.



$$s_0 = \text{Average}\left(\sum_{v_i \in G} \sum_{n \in [1,|v_i|]} h_{v_i,n}\right)$$

Training

- We minimize the negative log-likelihood of the target words sequence.

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

Experiments

Datasets

- **SAMSum** is a human generated dialogue summary dataset, which contains dialogues in various scenes of the real-life.
- **Argumentative Dialogue Summary Corpus (ADSC)** is mainly around debate topics.

Dataset	Split	#	Coverage	Average Know
SAMSum	Train	14732	94.43%	19.60
	Valid	818	95.72%	18.23
	Test	819	93.89%	19.77
ADSC	Full	45	100%	6.50

Automatic Evaluation

Type	Model	Know.	Heter.	Utter.	RL	R-1	R-2	R-L
Extractive	LONGEST-3	✗	✗	✗	✗	32.46	10.27	29.92
	TextRank	✗	✗	✗	✗	29.27	8.02	28.78
	SummaRunner	✗	✗	✗	✗	33.76	10.28	28.69
Abstractive	Transformer	✗	✗	✗	✗	36.62	11.18	33.06
	PGN	✗	✗	✗	✗	40.08	15.28	36.63
	HRED	✗	✗	✗	✗	40.39	16.13	37.65
Pipeline	Abs RL	✗	✗	✓	✓	40.96	17.18	39.05
	AbsRL Enhance	✗	✗	✓	✓	41.95	18.06	39.23
Ours	D-GCN	✓	✗	✗	✗	41.33	16.98	38.70
	D-GAT	✓	✗	✗	✗	41.08	16.89	38.61
	D-RGCN	✓	✗	✗	✗	41.36	17.07	38.93
	D-HGN	✓	✓	✗	✗	**42.03**	**18.07**	**39.56**

Table 2: Test set results on the SAMSum Dataset, where "R-1" is short for "ROUGE-1", "R-2" for "ROUGE-2", "R-L" for "ROUGE-L". "Know.", "Heter.", "Utter." and "RL" indicate whether knowledge, heterogeneity modeling, utterance-level extraction labels and reinforcement learning are used or not.

effectiveness of graph modeling

effectiveness of heterogeneity modeling

Human Evaluation

- Compared with D-HGN, D-HGN(w/o knowledge) gets a lower score in abstractiveness, which indicates knowledge incorporation can help our model express deeper meanings.
- D-HGN(w/o speaker) performs worse than D-HGN in correctness, which shows effectiveness of heterogeneity modeling by viewing speakers as heterogeneous data.
- AbsRL Enhance performs worst in correctness, which may due to the utterances extraction will break the coherence of dialogue contexts.

Model	Abstractiveness	Informativeness	Correctness
PGN	2.70	2.68	2.49
AbsRL Enhance	2.94	3.23	2.43
D-HGN	**3.26**	**3.25**	**2.92**
w/o *knowledge*	3.09	3.16	2.80
w/o *speaker*	3.23	3.21	2.60

Table 3: Human evaluation results.

Ablation Study

Model	ROUGE-1	ROUGE-2	ROUGE-L
D-HGN	**42.03**	**18.07**	**39.56**
w/o *message fusion*	41.29	17.09	38.74
w/o *node embedding*	41.99	17.85	38.89

Table 4: Ablation Study for Two Modules

without taking position information(w/o node embedding) or message fusion module(w/o message fusion) into account, our model will lose some performance.

without knowledge integration (w/o knowledge), the model suffers the performance drop

Model	ROUGE-1	ROUGE-2	ROUGE-L
D-HGN	**42.03**	**18.07**	**39.56**
w/o *knowledge*	41.52	17.38	38.76
w/o *speaker*	41.06	17.17	38.92

Table 5: Ablation Study for Different Types of Nodes

Zero-shot Setting

Model	ROUGE-1	ROUGE-2	ROUGE-L
PGN	28.69	4.77	22.39
AbsRL Enhance	30.00	4.87	22.27
D-GAT	32.90	5.46	22.47
D-HGN	**33.55**	**5.68**	**22.75**

Table 6: ROUGE F_1 results on the Argumentative Dialogue Summary Corpus.

The homogeneous model D-GAT that uses knowledge can get better results than other baselines.

The D-HGN gets the best score.

Knowledge can help our models better understand the dialogue in the new domain.

Visualization

- We apply **t-SNE** to these vectors.
- **D-HGN** can generate more discrete and easily distinguishable representations.
- **D-GAT** also tends to separate representations of different types of nodes

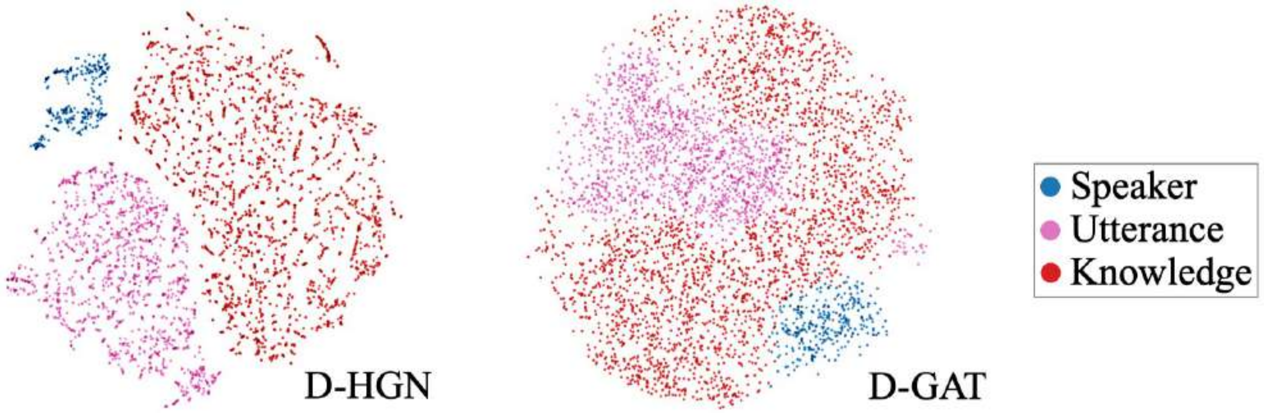
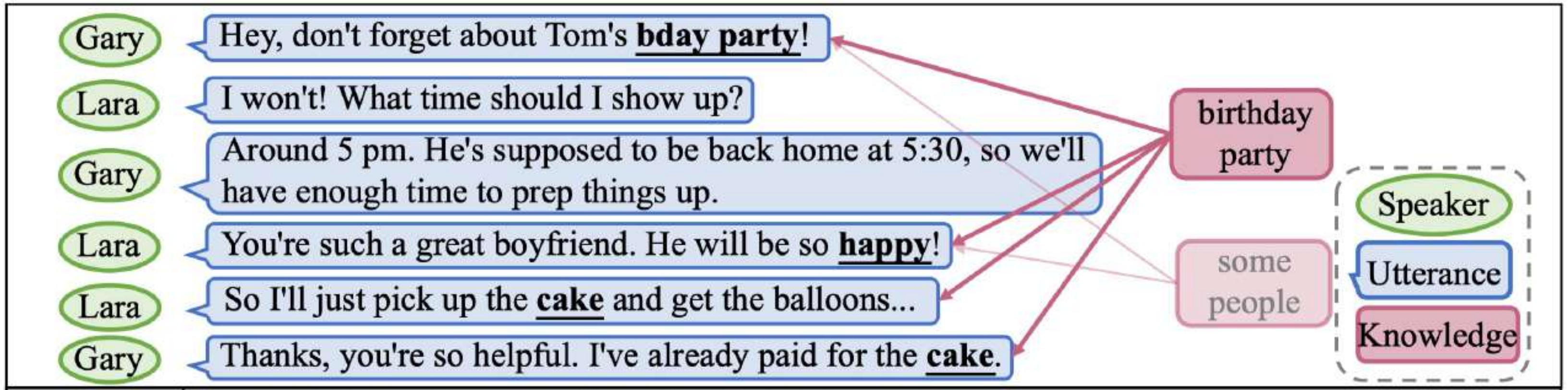


Figure 5: Visualization of node representations generated by the last graph layer of D-HGN and D-GAT.

Case Study



PGN	Gary and Lara will meet at 5 pm for Tom's **bday party.**
AbsRL Enhance	Gary and Lara are supposed to be back Tom's home at 5 pm for his **bday party.**
D-HGN	Gary and Lara are going to Tom's **birthday party** at 5 pm. Lara will pick up the **cake**.
Golden	It's Tom's **birthday**. Lara and Gary will come to Tom's place about 5 pm to prepare everything. Gary has already paid for the **cake** Lara will pick it.

Figure 6: Example summaries generated by different models for one dialogue.

Conclusion

Conclusion

- We are the first to improve abstractive dialogue summarization by incorporating commonsense knowledge.
- We introduce knowledge from the ConceptNet and present a Dialogue Heterogeneous Graph Network.
- Experiments on the SAMSum dataset show the effectiveness of our model. Zero-shot setting experiments show that our model can better generalized to the new domain.

哈爾濱工業大學
HARBIN INSTITUTE OF TECHNOLOGY

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

鹏城实验室
PCL
Peng Cheng Laboratory

Thanks!

Paper

Code