

Dialogue Summarization

Presenter: Wang Chen

Mentor: Piji Li

Outline

- Introduction
 - Task Definition & Applications
 - Taxonomy Based on Data Type
 - Challenges
- Recent Work
 - Keep Meeting Summaries on Topic: Abstractive Multi-Modal Meeting Summarization. ACL short, 2019.
 - Automatic Dialogue Summary Generation for Customer Service. KDD 2019.
- Conclusion

Task Definition & Applications

- **Definition:** Given an input dialogue, the goal is to generate a summary to capture the highlights of the dialogue.

Dialogue
Blair: Remember we are seeing the wedding planner after work
Chuck: Sure, where are we meeting her?
Blair: At Nonna Rita's
Chuck: Can I order their seafood tagliatelle or are we just having coffee with her? I've been dreaming about it since we went there last month
Blair: Haha sure why not
Chuck: Well we both remember the spaghetti pomodoro disaster from our last meeting with Diane
Blair: Omg hahaha it was all over her white blouse
Chuck: :D
Blair: :P
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.

SAMSum [4]

Dialogue:
A: Is this in color.
B: No , it 's black and white.
A: Does it look like an old picture.
B: Yes , i think so.
A: How old do you think the man is.
B: It looks like a young boy and he is what makes me think the picture is older , but the picture is not really old.
A: Do you see more than 1 cow.
B: 2 cows.
A: Is the boy wearing overalls.
B: No , he 's wearing a plaid short sleeved shirt and long pants and regular shoes.
A: How about a hat.
B: No.
A: Has he already started to milk.
B: He is attaching the milking apparatus to the cow.
A: Is he sitting down.
B: He is squatting.
A: What color is the cow.
B: Looks like it would be brown and white.
A: Are they inside a barn.
B: Inside a milking facility.
Description:
a man prepares to milk a dairy cow.

Dial2Desc [3]

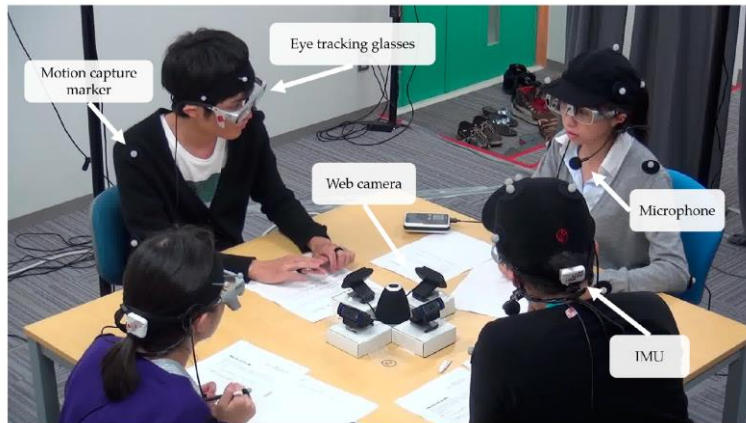
Dialogue
AGENT: Hello, what can I do for you?
USER: What's the standard of electric vehicles for the Express.
AGENT: Do you have a car?
AGENT: Or are you going to buy a car?
USER: I am hesitating which car to buy. One is Jianghuai EV Seven, the other is BYD YUAN.
AGENT: OK, you can fulfill the table in this link (link info) with the type of vehicle you wish to check. We will give you feedback in seven days.
USER: I have not bought yet.
USER: Can you check it now?
AGENT: I am quite sorry for that. A specialist on this issue will check it and call you back.
AGENT: They will give a precise answer for your question.
USER: OK.
AGENT: Thanks for your understanding. What else can I do for you?
USER: Nothing, thanks. Bye.
AGENT: Thank you. Have a nice day.
Summary
The user's question was about the standard of EV car for the Express. He asked the standard to decide which car to buy. I told the user to fill in the type of the cars in our system and we would give feedback in seven days. The user approved the result. The user hung up.
Key point sequence
Question description → Solution → User approval → End

DiDi Customer Service [2]

Task Definition & Applications

• Applications

- Automatic Meeting Summarization
- Medical Conversation Summarization
- Customer Service Summarization
- ...



AMI, Meeting Summarization [7]

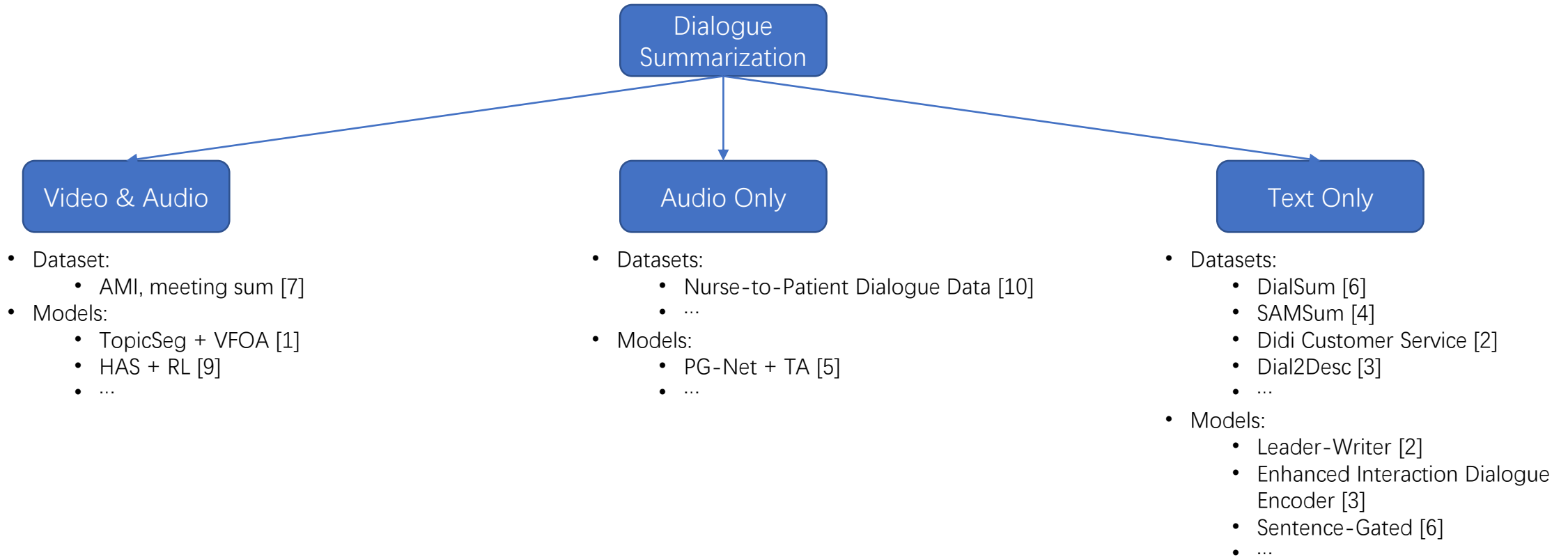


Medical Conversation Summarization [8]

<i>Dialogue</i>
<i>AGENT:</i> Hello, what can I do for you?
<i>USER:</i> What's the standard of electric vehicles for the Express.
<i>AGENT:</i> Do you have a car?
<i>AGENT:</i> Or are you going to buy a car?
<i>USER:</i> I am hesitating which car to buy. One is Jianghuai EV Seven, the other is BYD YUAN.
<i>AGENT:</i> OK, you can fulfill the table in this link (link info) with the type of vehicle you wish to check. We will give you feedback in seven days.
<i>USER:</i> I have not bought yet.
<i>USER:</i> Can you check it now?
<i>AGENT:</i> I am quite sorry for that. A specialist on this issue will check it and call you back.
<i>AGENT:</i> They will give a precise answer for your question.
<i>USER:</i> OK.
<i>AGENT:</i> Thanks for your understanding. What else can I do for you?
<i>USER:</i> Nothing, thanks. Bye.
<i>AGENT:</i> Thank you. Have a nice day.
<i>Summary</i>
The user's question was about the standard of EV car for the Express. He asked the standard to decide which car to buy. I told the user to fill in the type of the cars in our system and we would give feedback in seven days. The user approved the result. The user hung up.
<i>Key point sequence</i>
Question description → Solution → User approval → End

DiDi Customer Service [6]

Taxonomy Based on Data Type



Challenges

- **Logicality**

- The summary should be organized in a readable order.

- **Integrality**

- All the important facts should be covered.

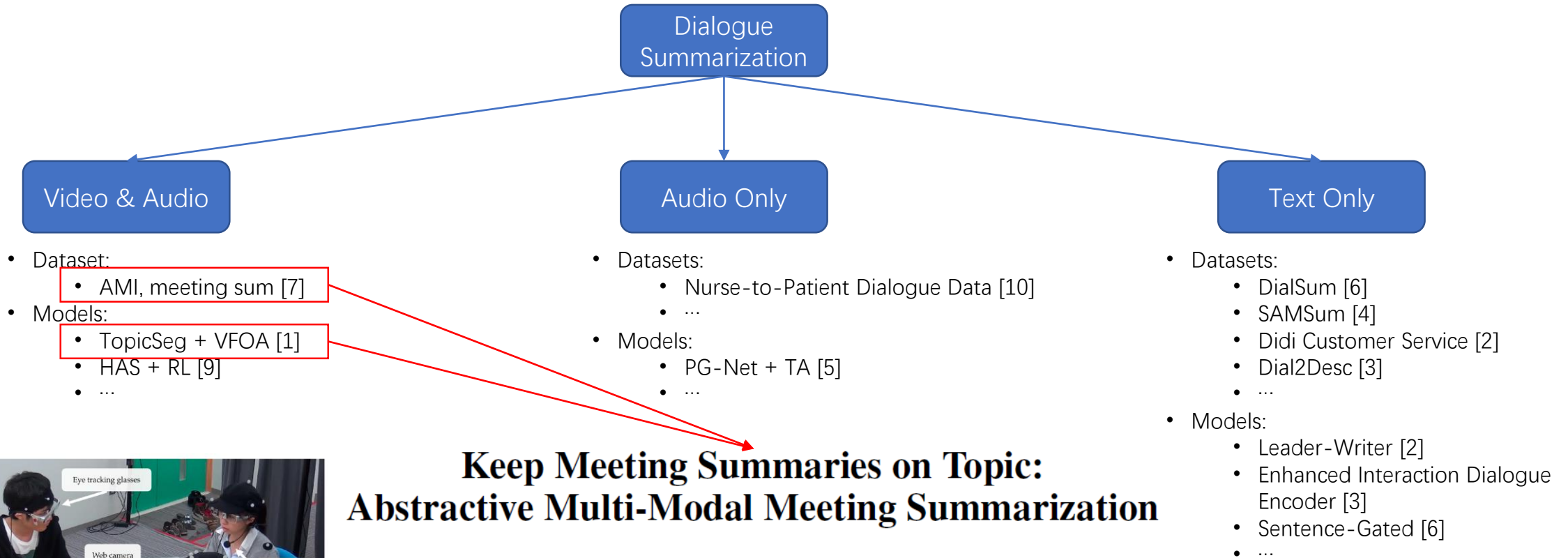
- **Correctness**

- The summary should be consistent with the facts in the dialogue.

- **Other challenges in generation area**

- Fluency
- Evaluation metrics
- ...

Taxonomy Based on Data Type



Keep Meeting Summaries on Topic: Abstractive Multi-Modal Meeting Summarization

Manling Li¹, Lingyu Zhang², Heng Ji¹, Richard J. Radke²

¹ Department of Computer Science

² Department of Electrical, Computer, and Systems Engineering
Rensselaer Polytechnic Institute



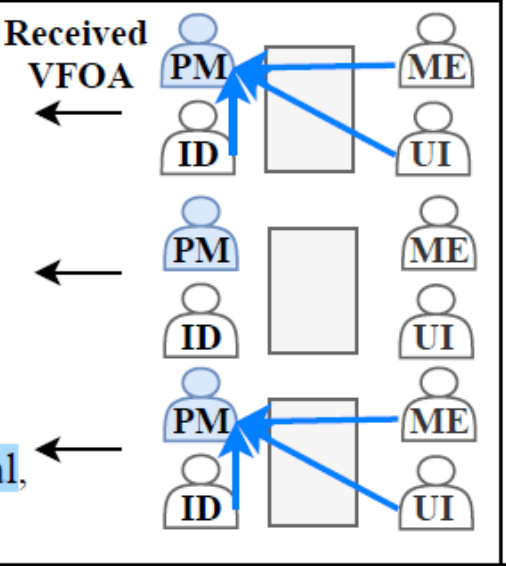
Keep Meeting Summaries on Topic

- **Main Contributions:**

- proposed a novel **hierarchical attention mechanism** across three levels: topic segment, utterance, and word.
- introduced the **multi-modal** feature i.e. Visual Focus of Attention (**VFOA**) to help recognize the important utterances.

Keep Meeting Summaries on Topic

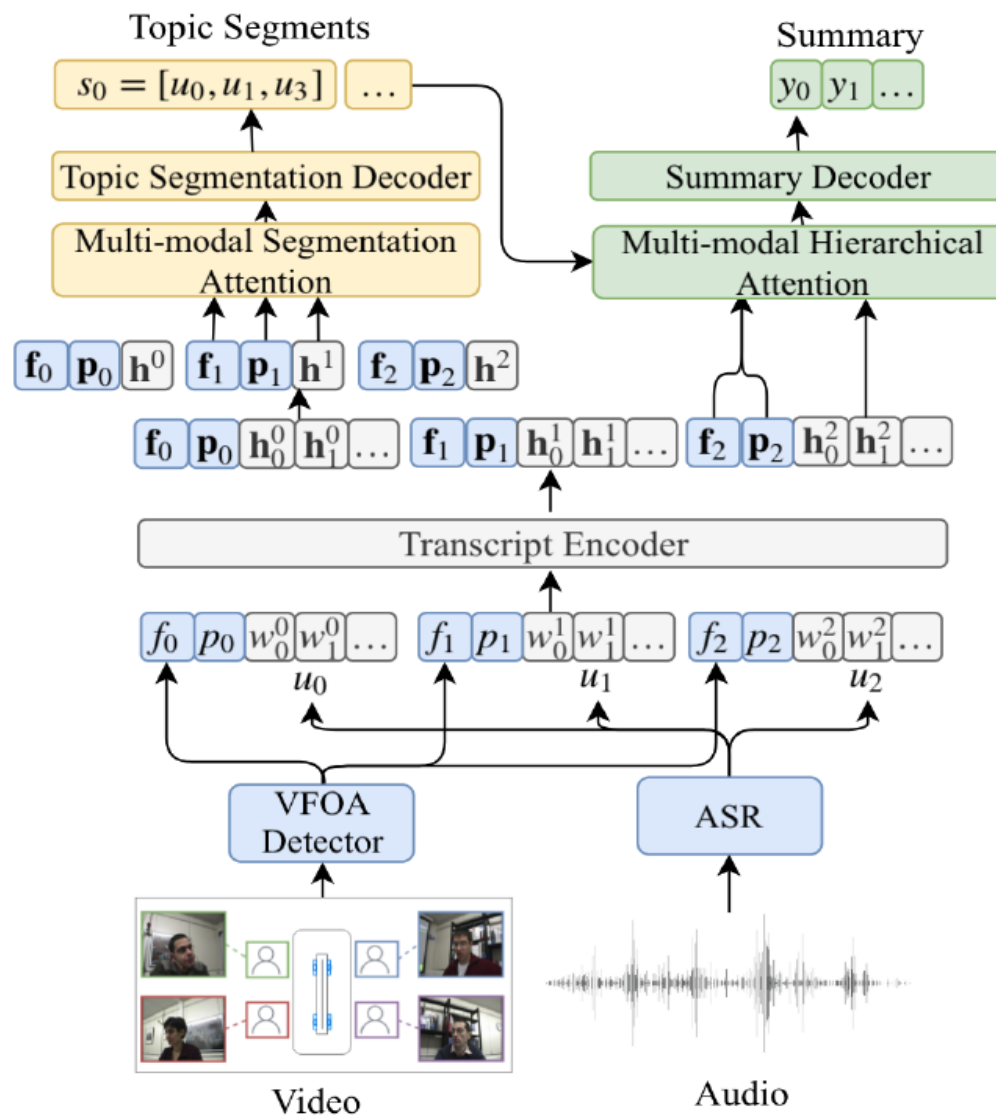
- **Why the Visual Focus of Attention (VFOA) feature is useful?**
 - One widely-accepted **assumption** is that an utterance is **more important if** its speaker receives **more attention**.
 - One data sample from AMI corpus:

Transcript	<p>Um I'm Sarah, the Project Manager and this is our first meeting, surprisingly enough. Okay, this is our agenda, um we will do some stuff, get to know each other a bit better to feel more comfortable with each other.</p> <p>Um then we'll go do tool training, talk about the project plan, discuss our own ideas and everything um and we've got twenty five minutes to do that, as far as I can understand.</p> <p>Now, we're developing a remote control which you probably already know. Um, we want it to be original, something that's uh people haven't thought of, that's not out in the shops, um, trendy, appealing to a wide market, but you know, not a hunk of metal, and user-friendly, grannies to kids, maybe even pooches should be able to use it.</p>	
Manual summary	The project manager gave an introduction to the goal of the project, to create a trendy yet user-friendly remote .	

The **color** indicates the **attention** received by the speaker PM (Project Manager).
Others: ME (Marketing Expert), ID (Industrial Designer), UI (User Interface).

Keep Meeting Summaries on Topic

- **TopicSec + VFOA**

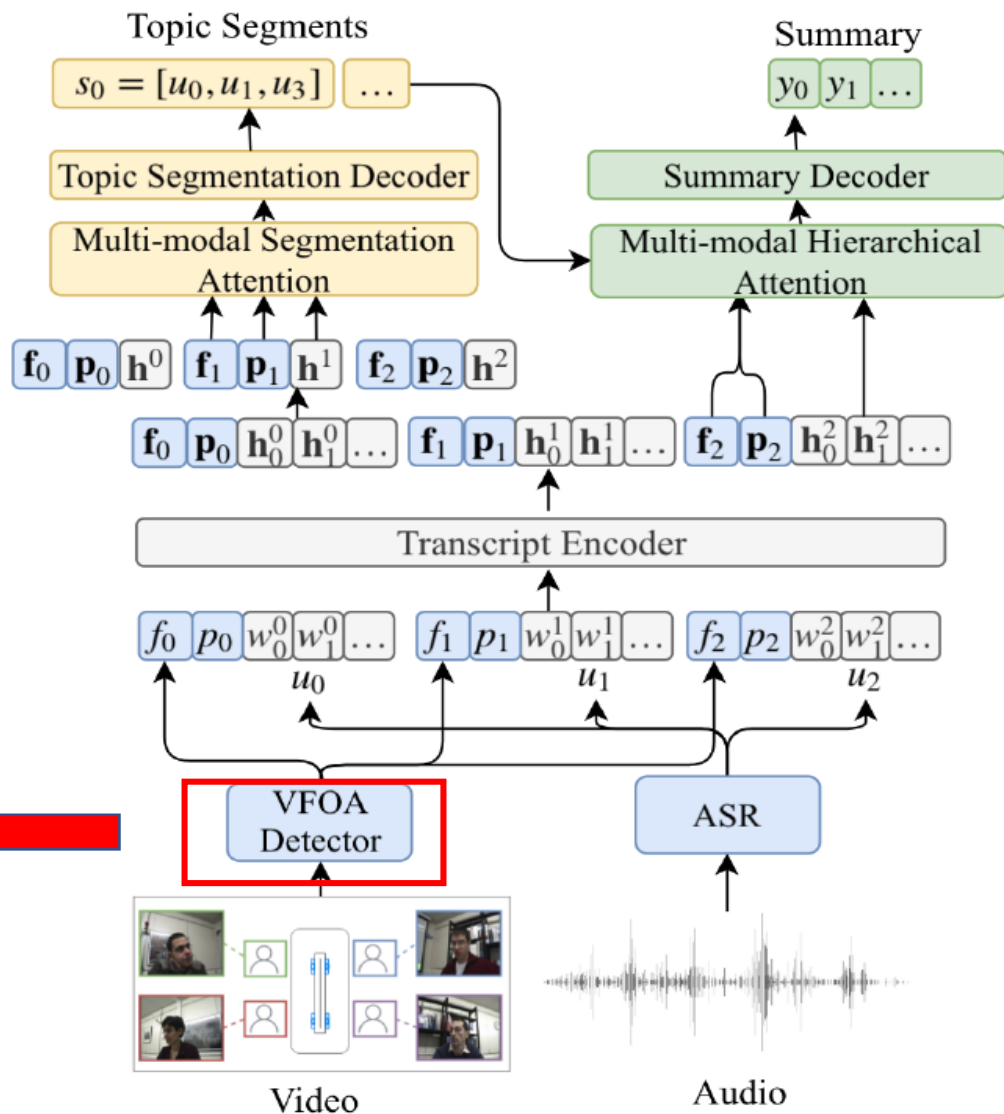
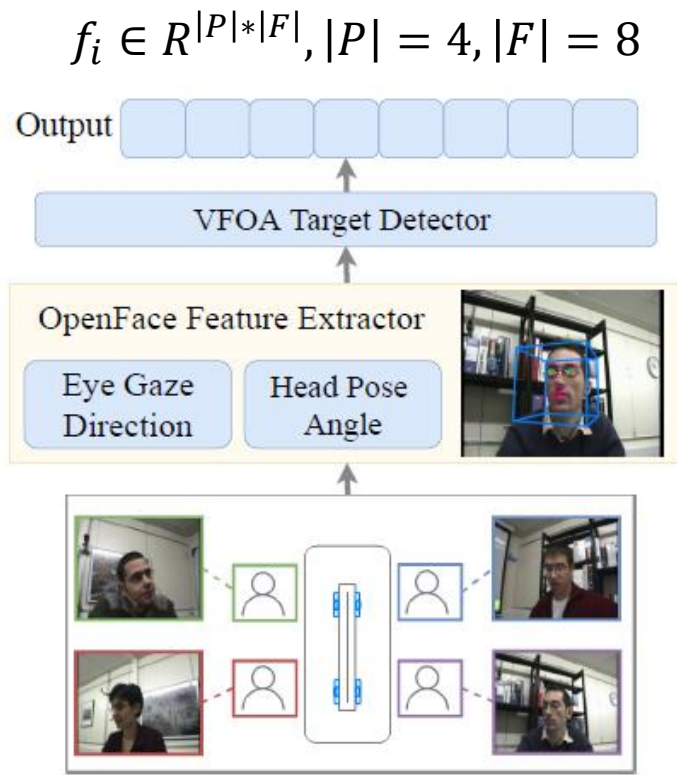


We formulate a meeting transcript as a list of triples $X = \{(p_i; f_i; u_i)\}$

- $p_i \in P$ is the **speaker of utterance** u_i , where P denotes the set of participants.
- $f_i \in R^{|P| \times |F|}$ contains the **VFOA target sequence** over the course of utterance u_i for each participant where $F = \{p_0, \dots, p_{|P|}, \text{table}, \text{whiteboard}, \text{projection screen}, \text{unknown}\}$.
- u_i is a sequence of words.

Keep Meeting Summaries on Topic

• TopicSec + VFOA

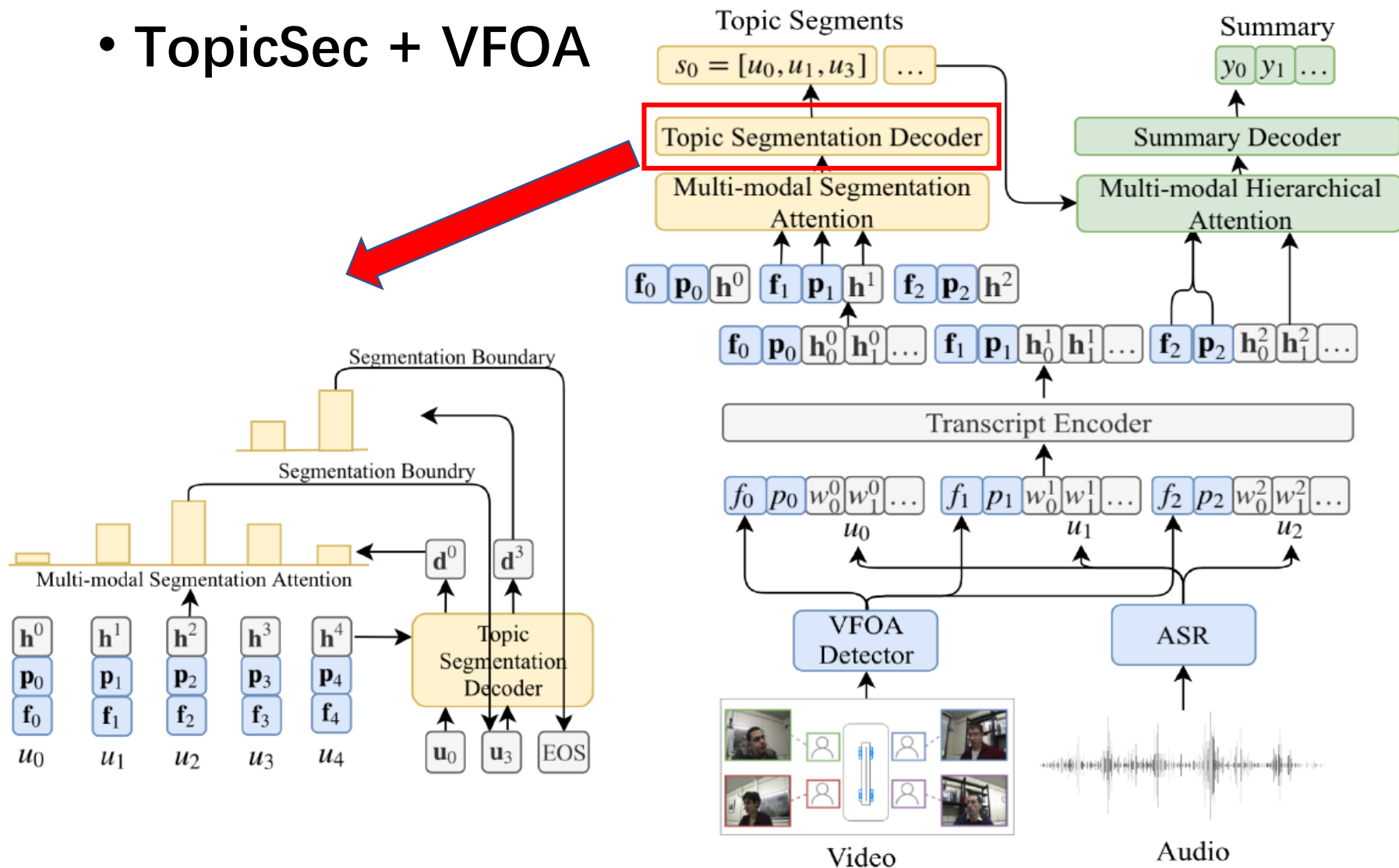


We formulate a meeting transcript as a list of triples $X = \{(p_i; f_i; u_i)\}$

- $p_i \in R^{|P|}$ is the **speaker of utterance u_i** , where P denotes the set of participants. One-hot vector.
- $f_i \in R^{|P|*|F|}$ contains the **VFOA target sequence** over the course of utterance u_i for each participant where $F = \{p_0, \dots, p_{|P|}, table, whiteboard, projection screen, unknown\}$.
- u_i is a sequence of words.

Keep Meeting Summaries on Topic

- **TopicSec + VFOA**

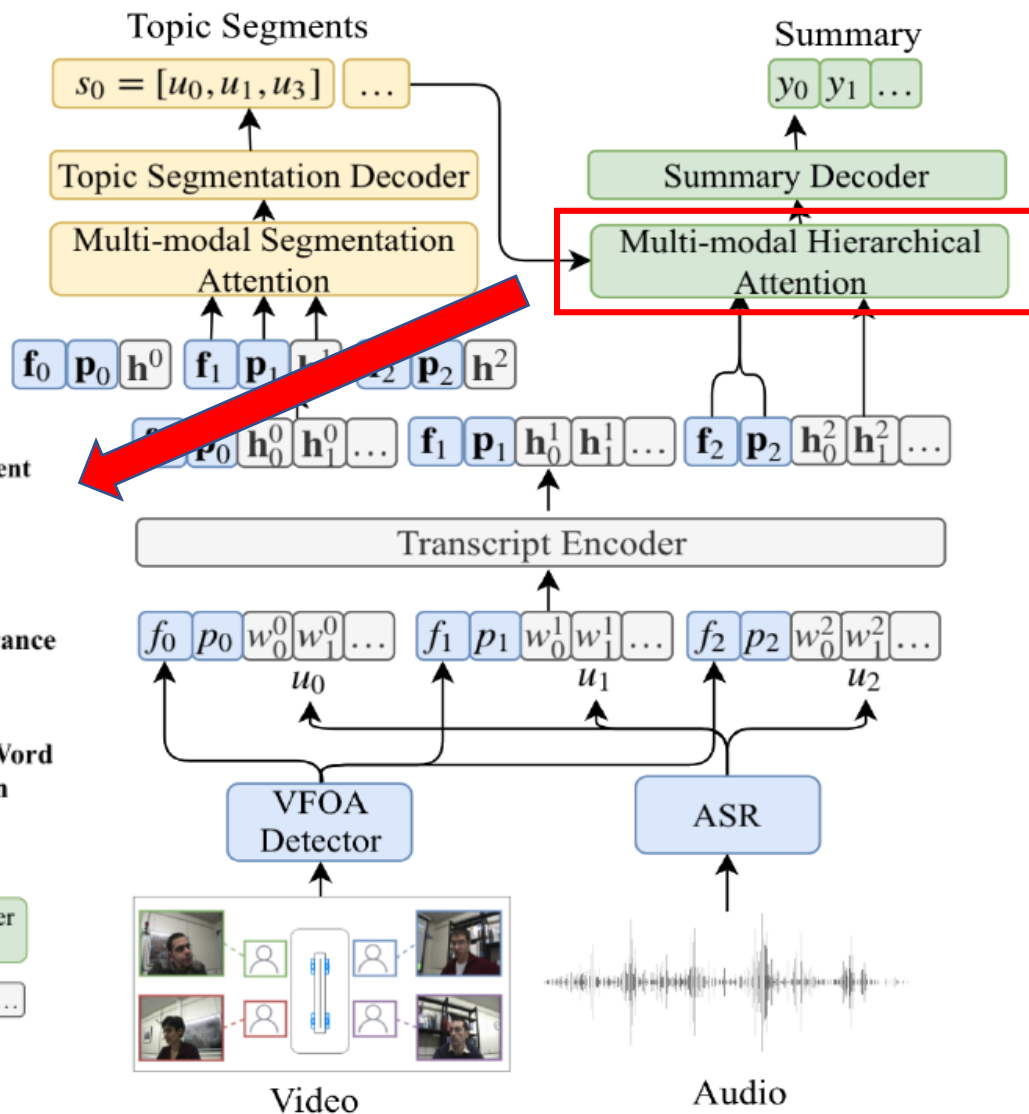


We formulate a meeting transcript as a list of triples $X = \{(p_i; f_i; u_i)\}$

- $p_i \in R^{|P|}$ is the **speaker of utterance u_i** , where P denotes the set of participants. One-hot vector.
- $f_i \in R^{|P| * |F|}$ contains the **VFOA target sequence** over the course of utterance u_i for each participant where $F = \{p_0, \dots, p_{|P|}, \text{table}, \text{whiteboard}, \text{projection screen}, \text{unknown}\}$.
- u_i is a sequence of words.

Keep Meeting Summaries on Topic

- **TopicSec + VFOA**



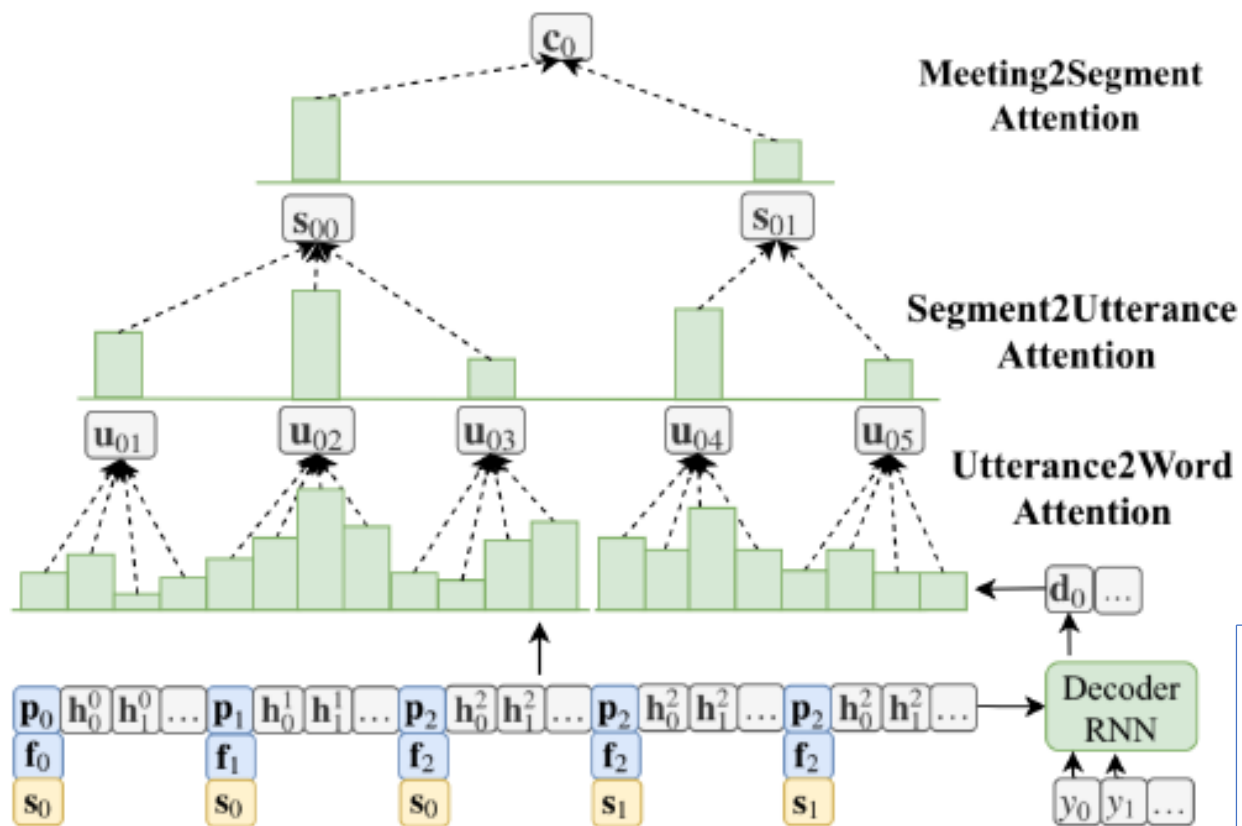
We formulate a meeting transcript as a list of triples

$$X = \{(p_i; f_i; u_i)\}$$

- $p_i \in R^{|P|}$ is the **speaker of utterance u_i** , where P denotes the set of participants. One-hot vector.
- $f_i \in R^{|P| \times |F|}$ contains the **VFOA target sequence** over the course of utterance u_i for each participant where $F = \{p_0, \dots, p_{|P|}, \text{table}, \text{whiteboard}, \text{projection screen}, \text{unknown}\}$.
- u_i is a sequence of words.

Keep Meeting Summaries on Topic

- TopicSec + VFOA



Hierarchical Attention in Summary Decoder

$$e''_{iq} = v_3^\top \tanh(W_{d3}d_i + W_s c_{iq}).$$

$$e'_{ik} = f_k^\top p'_k \left(v_2^\top \tanh(W_{d2}d_i + W_u u_{ik}) \right).$$

$$c_{iq} = \text{Softmax}(e'_{ik}) u_k, u_k \in s_q.$$

$$e_{ij} = v_1^\top \tanh(W_{d1}d_i + W_w w_j + W_p p_j + W_f f_j)$$

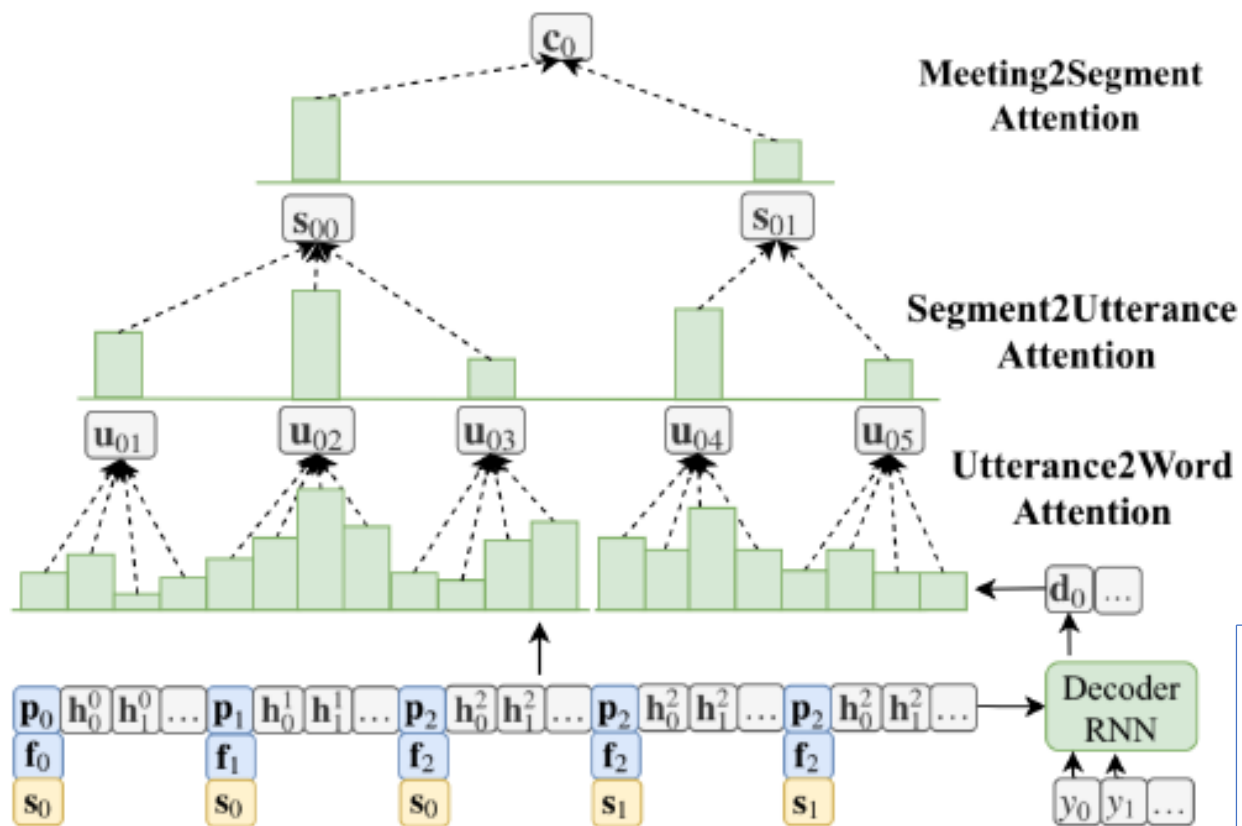
$$u_{ik} = \text{Softmax}(e_{ij}) w_j, w_j \in u_k.$$

$$\alpha_{ij}^{sum} = \frac{\exp(e_{ij} e'_{ik} e''_{iq})}{\sum_{j \in s_q} \exp(e_{ij} e'_{ik} e''_{iq})},$$

The probability of generating y_i follows the decoder in PGNet, and α_{ij}^{sum} is the copying probability.

Keep Meeting Summaries on Topic

- TopicSec + VFOA



Hierarchical Attention in Summary Decoder

$$e''_{iq} = v_3^\top \tanh(W_{d3}d_i + W_s c_{iq}).$$

$$e'_{ik} = \mathbf{f}_k^\top \mathbf{p}'_k \left(v_2^\top \tanh(W_{d2}d_i + W_u u_{ik}) \right).$$

$$c_{iq} = \text{Softmax}(e'_{ik}) u_k, u_k \in s_q.$$

$$e_{ij} = v_1^\top \tanh(W_{d1}d_i + W_w w_j + W_p p_j + W_f f_j)$$

$$u_{ik} = \text{Softmax}(e_{ij}) w_j, w_j \in u_k.$$

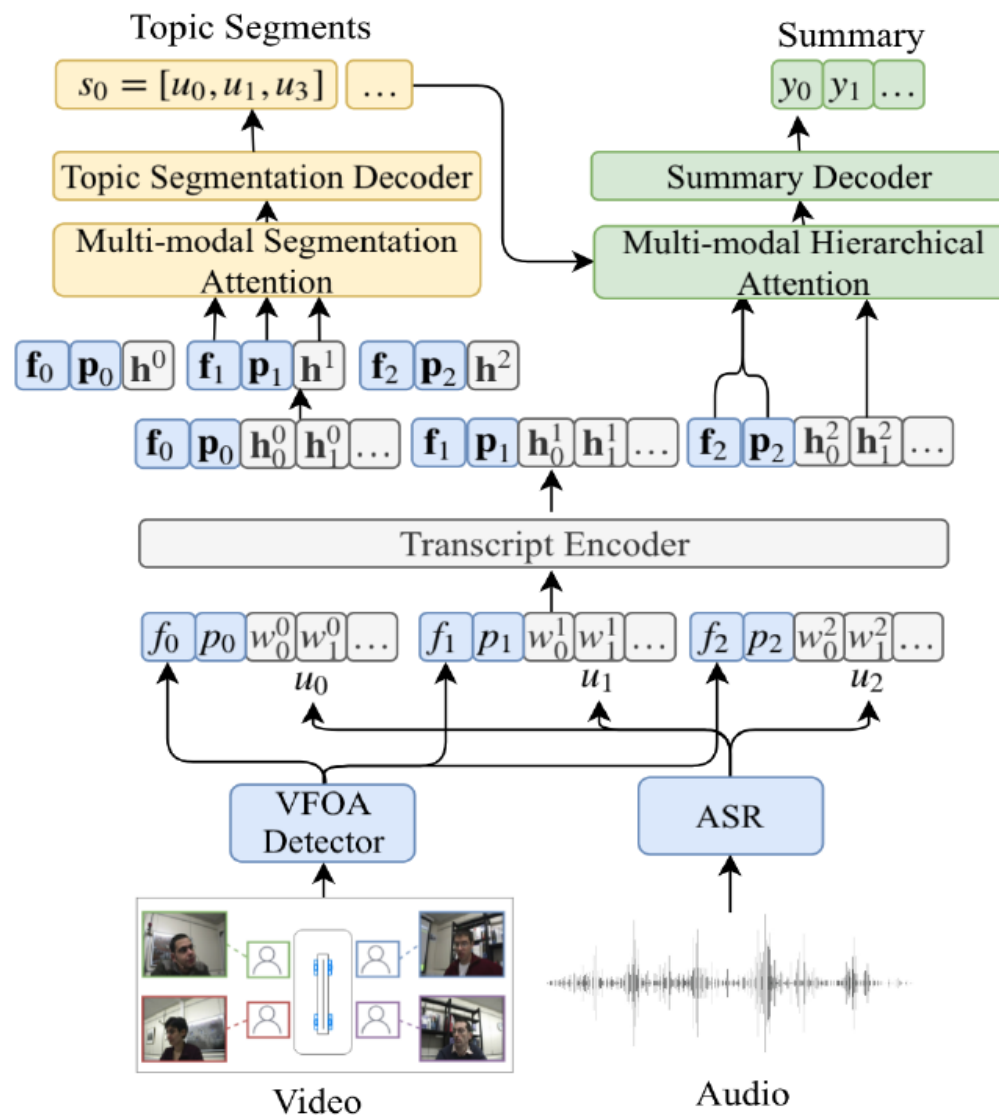
$$\alpha_{ij}^{sum} = \frac{\exp(e_{ij} e'_{ik} e''_{iq})}{\sum_{j \in s_q} \exp(e_{ij} e'_{ik} e''_{iq})},$$

The probability of generating y_i follows the decoder in PGNet, and α_{ij}^{sum} is the copying probability.

Keep Meeting Summaries on Topic

- **TopicSec + VFOA**

$$\begin{aligned} \mathcal{L} &= -\log P(Y, S|X) \\ &= \sum_{y_i \in Y} -\log P(y_i|X) + \sum_{s_j \in S} -\log P(s_j|(p_j, f_j, u_j)) \end{aligned}$$



Keep Meeting Summaries on Topic

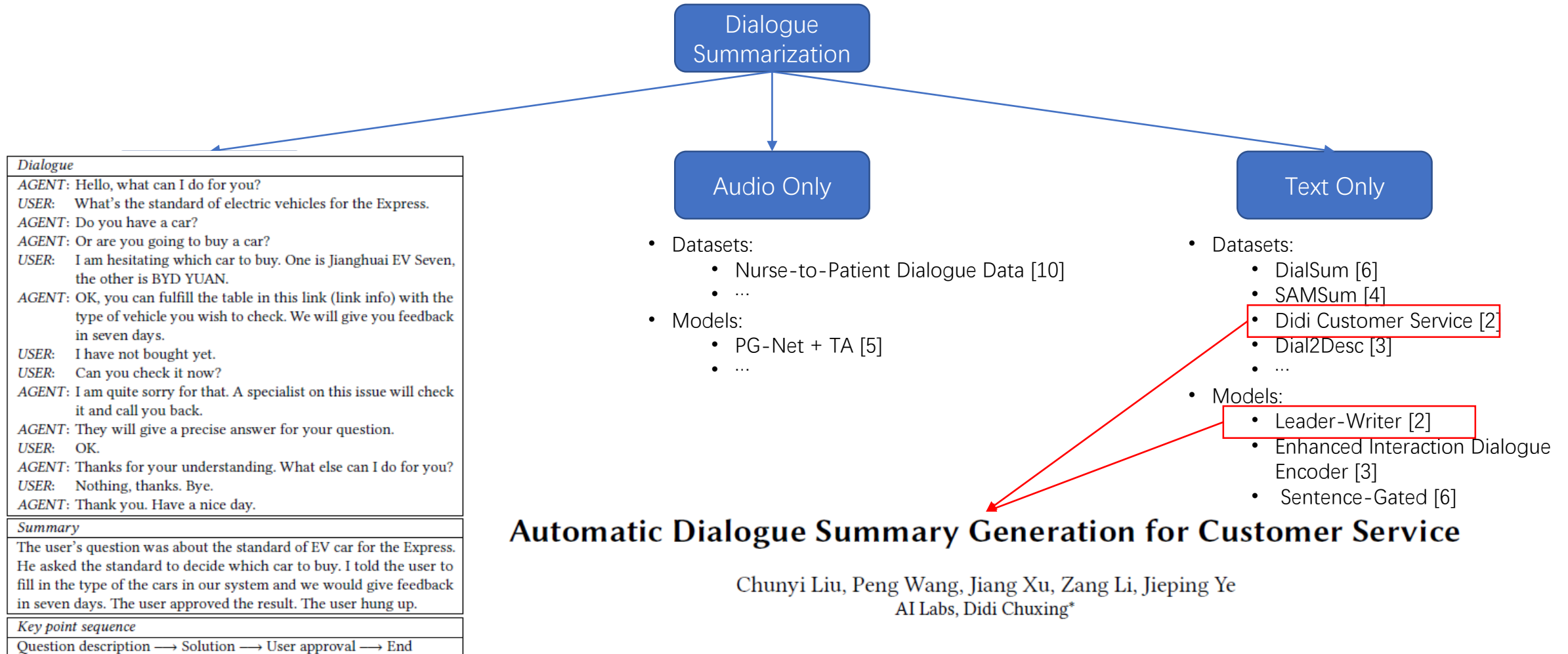
- **Dataset**

- 97 meetings for training; 20 meetings for validation; 20 meetings for testing.
- Each meeting lasts 30 minutes.

- **Experiment Results**

Model	ROUGE			BLEU			
	ROUGE_1	ROUGE_2	ROUGE_L	BLEU_1	BLEU_2	BLEU_3	BLEU_4
CoreRank (Shang et al., 2018)	37.86	7.84	13.72	17.17	6.78	1.77	0.00
PGN (See et al., 2017)	36.75	10.48	23.81	37.89	23.41	12.84	6.92
Our Approach (TopicSeg+VFOA)	53.29	13.51	26.90	40.98	26.19	13.76	8.03
Our Approach (TopicSeg)	51.53	12.23	25.47	39.67	24.91	12.37	7.86

Taxonomy Based on Data Type



Dialogue Summarization for Customer Service

- **Main Contributions:**

- proposed to use **auxiliary key point sequences** to ensure the logic and integrity of dialogue summaries.
- proposed a novel hierarchical decoder architecture, the **Leader-Writer net**, to generate both key point sequences and the summaries.

Dialogue Summarization for Customer Service

- **What is a key point sequence?**

- A key point is the **theme** of a contiguous set of one or more summary sentences.
- One example

<i>Summary</i>
<u>The user's question was about the standard of EV car for the Express.</u> <u>He asked the standard to decide which car to buy.</u> <u>I told the user to</u> <u>fill in the type of the cars in our system and we would give feedback</u> <u>in seven days.</u> <u>The user approved the result.</u> <u>The user hung up.</u>
<i>Key point sequence</i>
<u>Question description</u> → <u>Solution</u> → <u>User approval</u> → <u>End</u>

- The key point sequence can be used to enhance the **logic** and **integrity** of the generated summary.

Dialogue Summarization for Customer Service

- How to generate a key point sequence for the training dataset?



<i>Summary</i>
The user's question was about the standard of EV car for the Express. He asked the standard to decide which car to buy. I told the user to fill in the type of the cars in our system and we would give feedback in seven days. The user approved the result. The user hung up.
<i>Key point sequence</i>
Question description → Solution → User approval → End

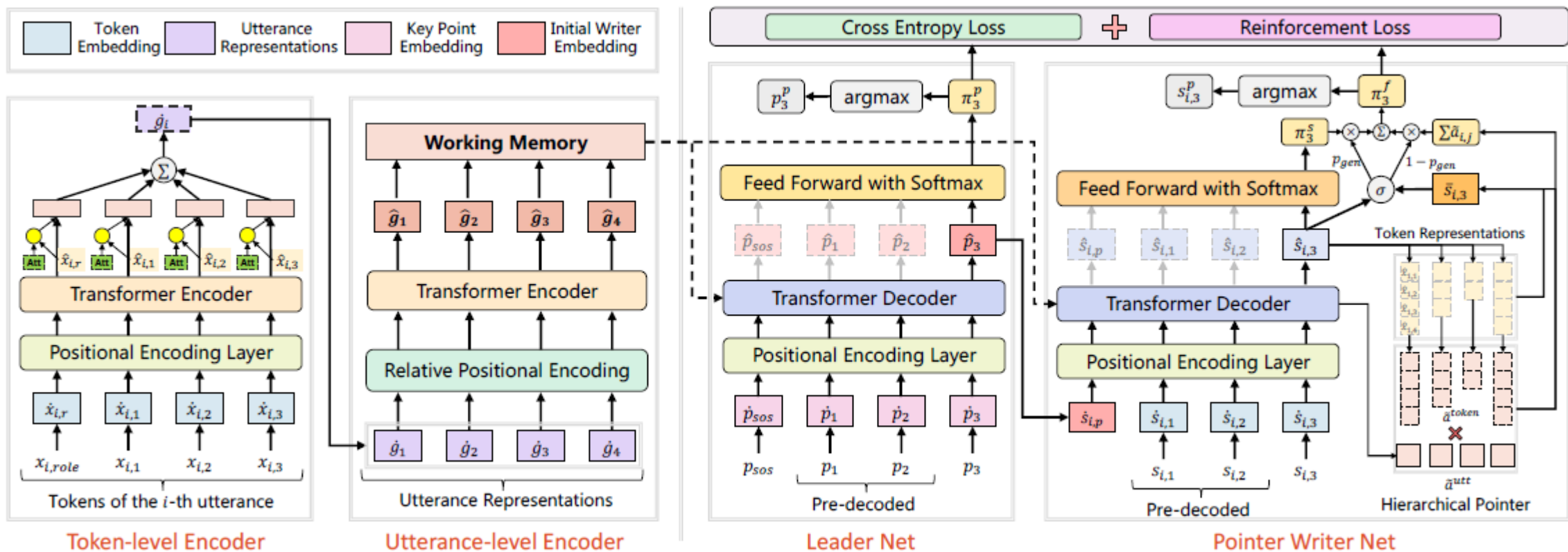
Key point	Corresponding sample summary
Question description	The user's question is about the standard of EV car for the Express drivers.
Solution	I told the user to fulfill the type of the cars in our system and we would give feedback in seven days.
End	User hung up.
User approval	The user approved the result.
User disapproval	The user disapproved the result.
Suggestion	I suggested the user go to the downtown for orders.
Need call back	We should call her back in 30 minutes.
Acknowledgement	The user acknowledged the results.
Escalation	Need senior agent's further investigation.

...

Totally 51 key points

Dialogue Summarization for Customer Service

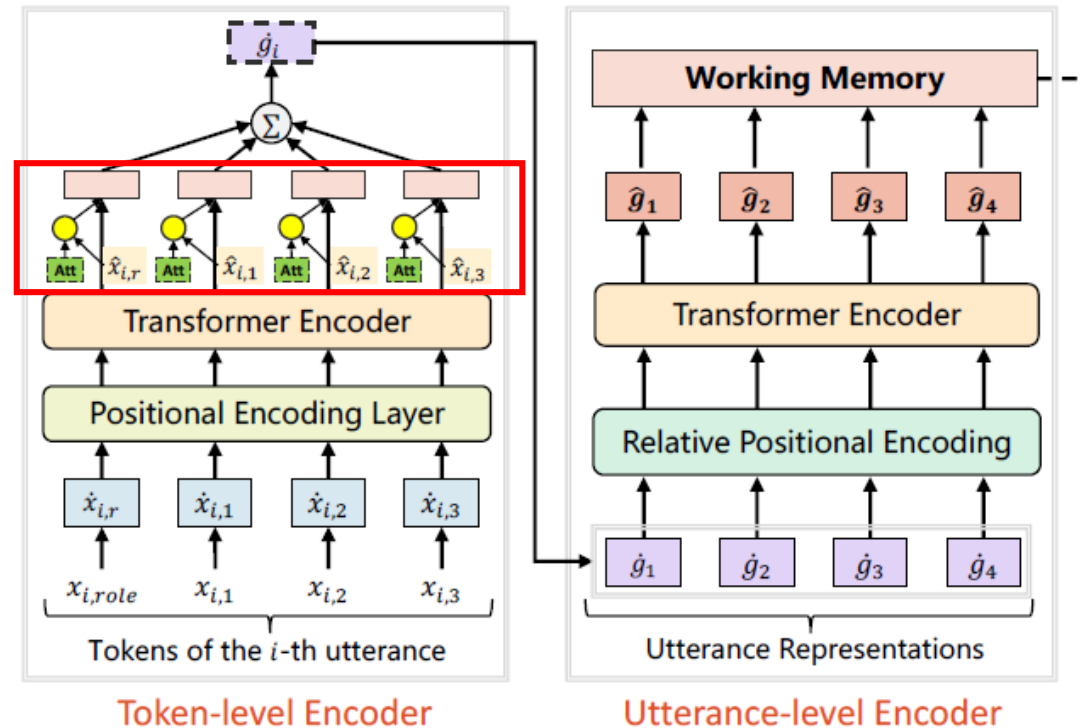
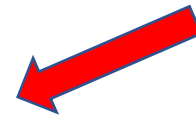
- **Leader-Writer net: Overall architecture**



Dialogue Summarization for Customer Service

- **Leader-Writer net: Hierarchical Encoder**

$$\hat{g}_i = \text{softmax} \left(\text{score} \left(\hat{X}_i \right) \right) \hat{X}_i$$
$$\text{score} \left(\hat{x}_{i,j} \right) = v^T \tanh \left(W^S \hat{x}_{i,j} + b^S \right)$$



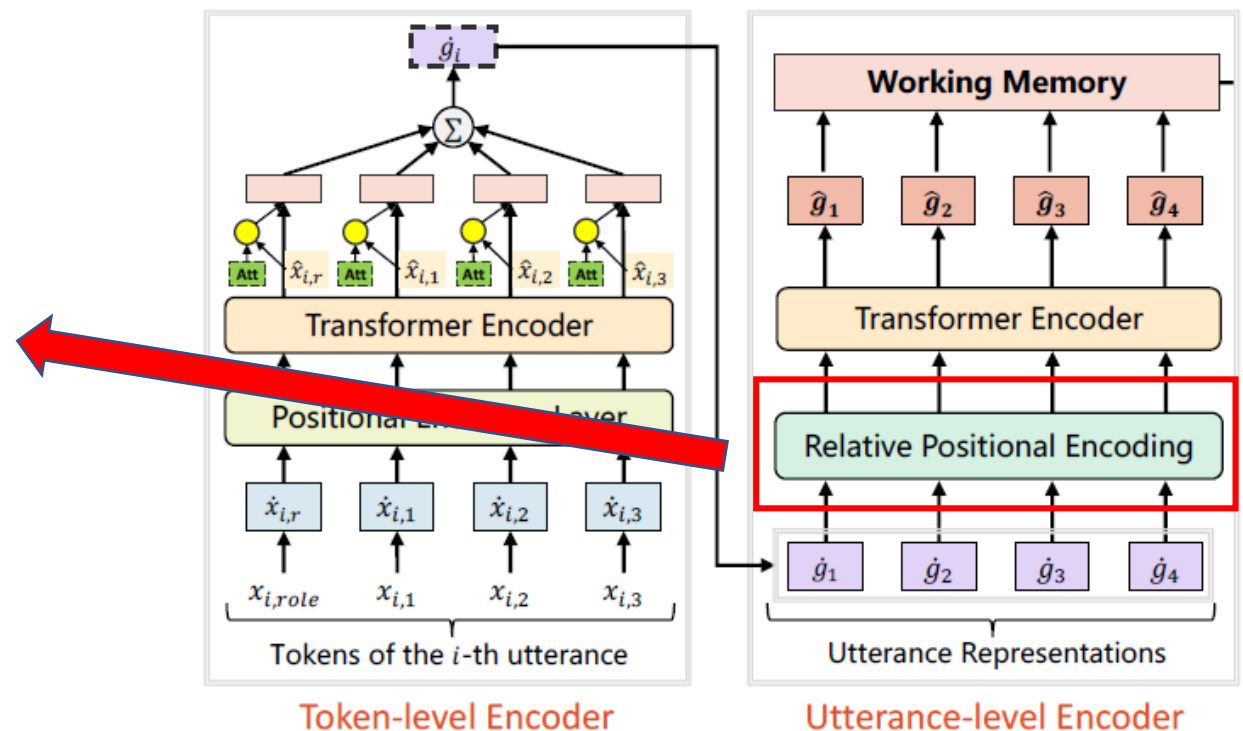
Dialogue Summarization for Customer Service

- **Leader-Writer net: Hierarchical Encoder**

The relative position of i -th utterance:

$$\left\lfloor \frac{iK}{M} \right\rfloor$$

where M is the utterance number of a dialogue and K is the maximum relative position number. K is set to 30 in the experiments.



Dialogue Summarization for Customer Service

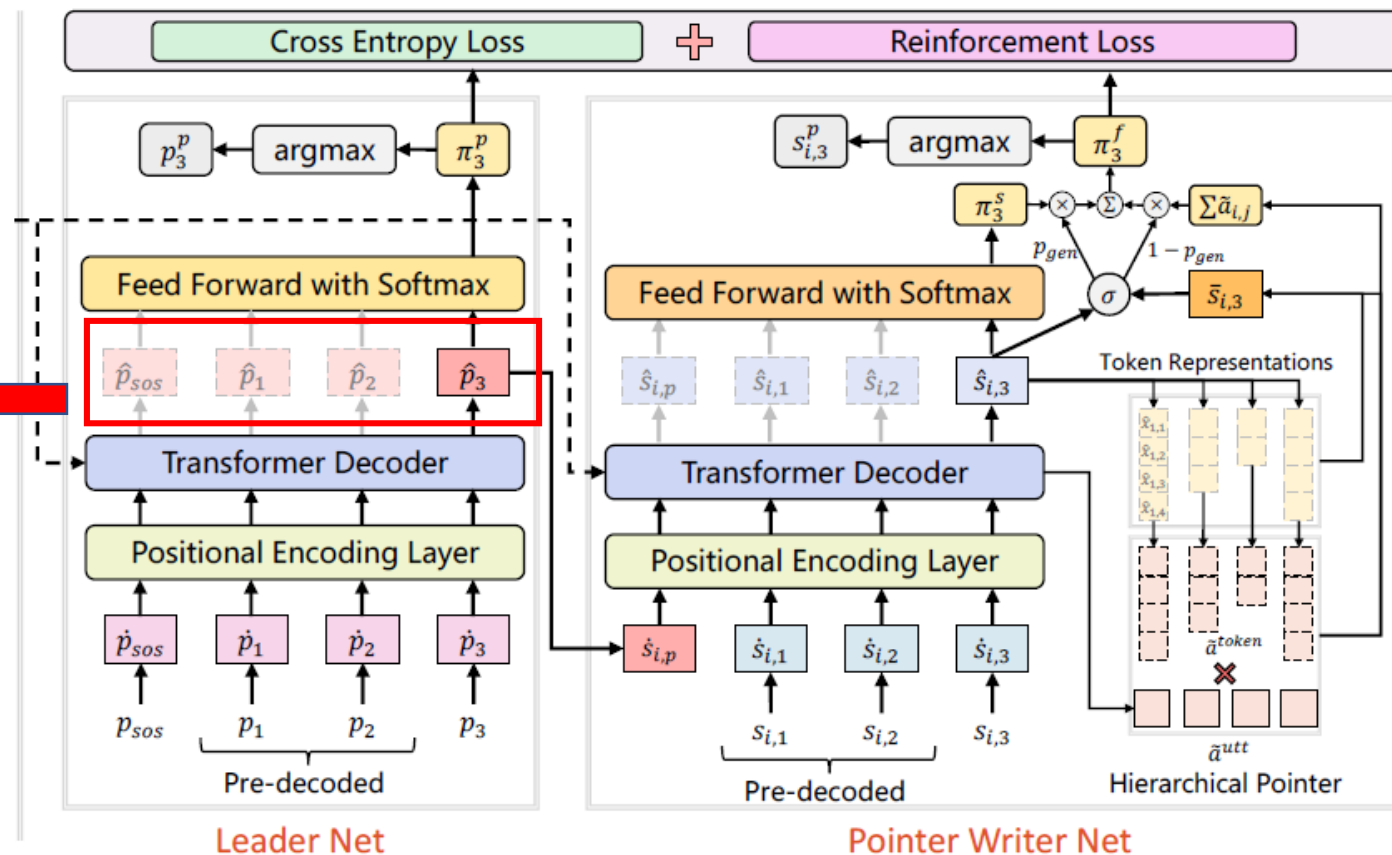
- **Leader-Writer net: Hierarchical Decoder**

The **hidden state** of the key point decoder is used to guide the generation of the sub-summary **instead of the predicted key point**.

The reason is to avoid the exactly the same initial states for two sub-summaries under the same key point in one key point sequence.

e.g.,

[· · · , Solution, User disapproval, Solution, User approval, · · ·]



Dialogue Summarization for Customer Service

- **Leader-Writer net: Training**

- **Cross-entropy loss**

$$L_{ce}^p = -\frac{1}{n} \sum_{t=1}^n \log \pi_t^p(p_t)$$

$$L_{ce}^s = -\frac{1}{\sum_{i=1}^n n_i} \sum_{i=1}^n \sum_{j=1}^{n_i} \log \pi_{i,j}^f(s_{i,j})$$

- **Reinforcement loss**

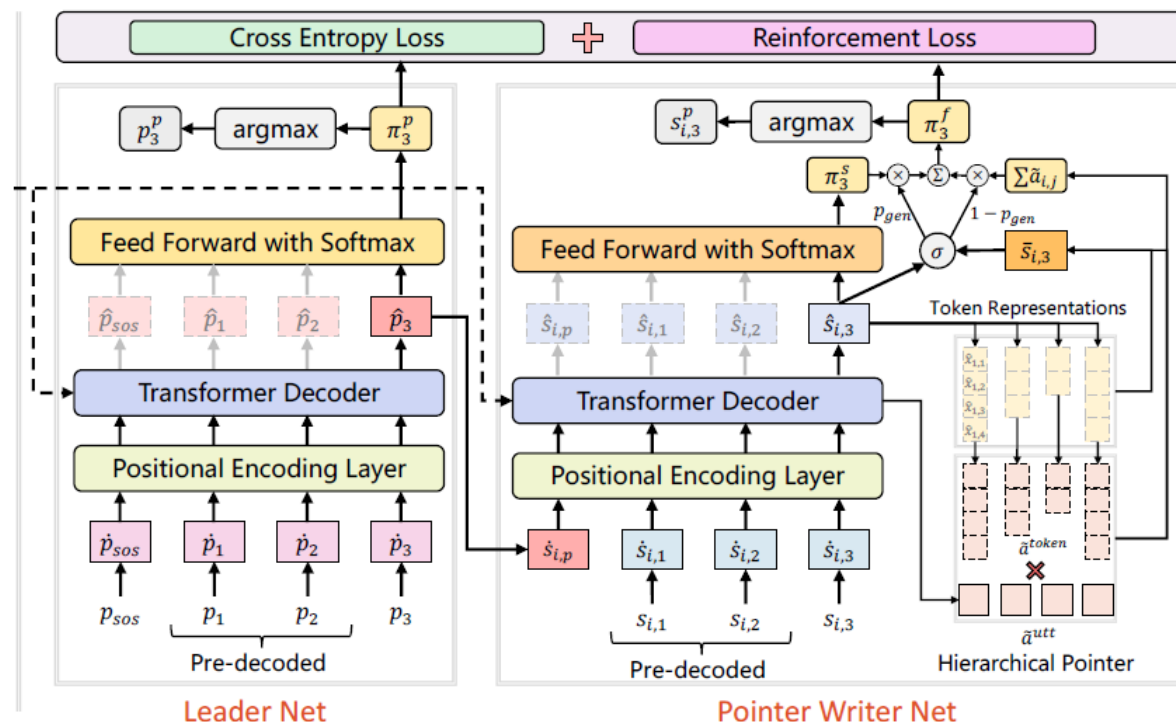
(use ROUGE-L as reward)

$$L_{rl}^p = \frac{1}{n} (R(P^p) - R(P^r)) \sum_{t=1}^n \log \pi_t^p(p_t^r)$$

$$L_{rl}^s = \frac{1}{\sum_{i=1}^n n_i} \sum_{i=1}^n (R(S_i^p) - R(S_i^r)) \sum_{j=1}^{n_i} \log \pi_{i,j}^f(s_{i,j}^r)$$

- **Joint loss**

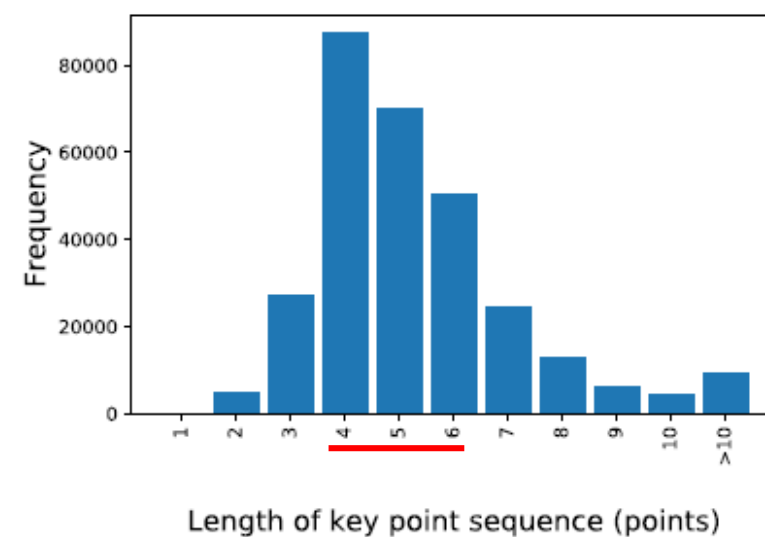
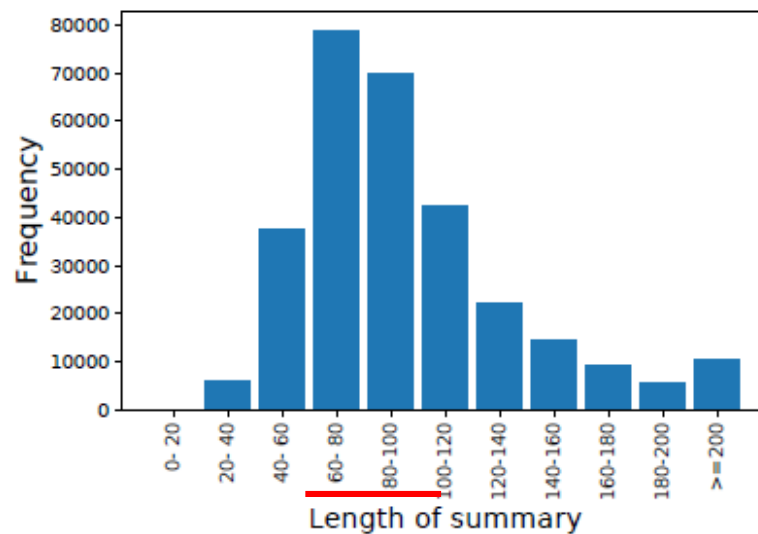
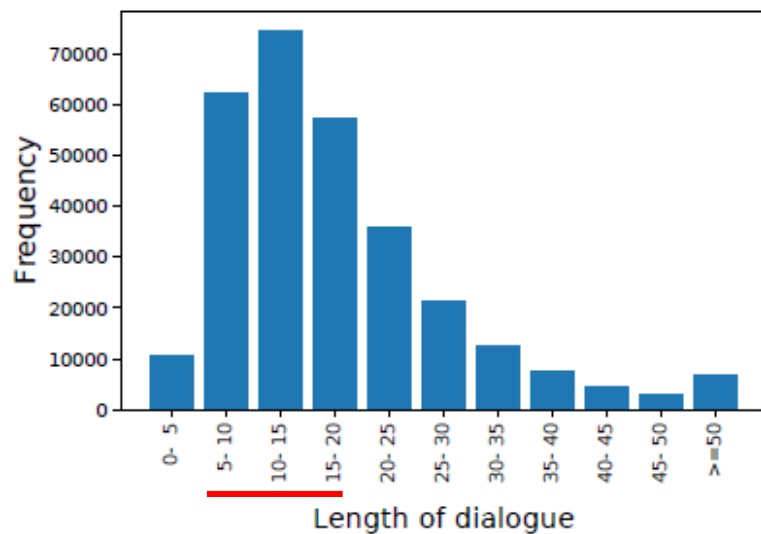
$$L = \alpha_1 L_{ce}^p + \alpha_2 L_{ce}^s + \alpha_3 L_{rl}^p + (1 - \alpha_1 - \alpha_2 - \alpha_3) L_{rl}^s$$



Dialogue Summarization for Customer Service

- **Dataset**

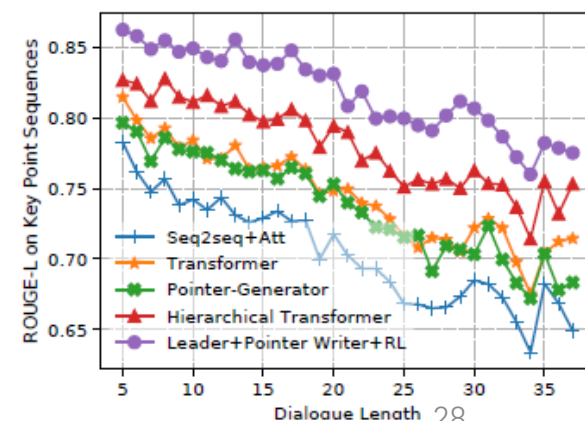
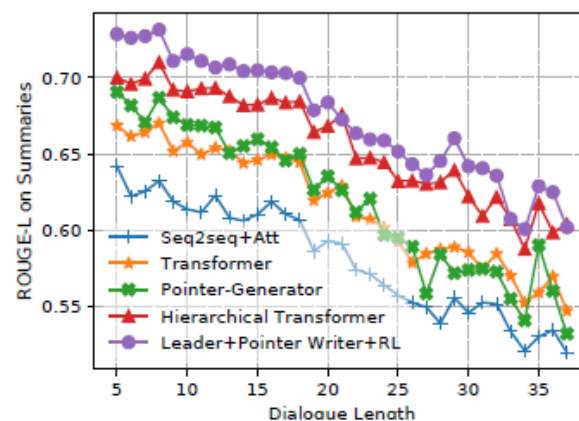
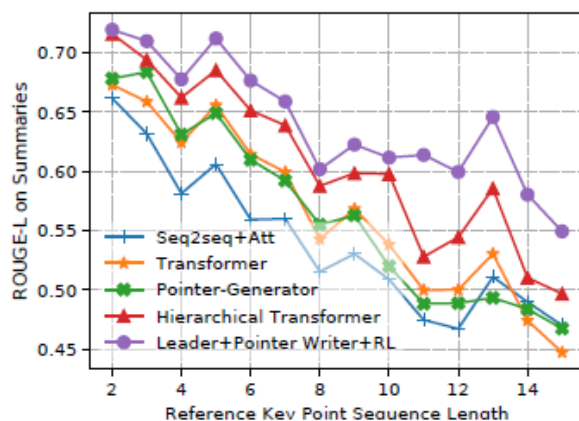
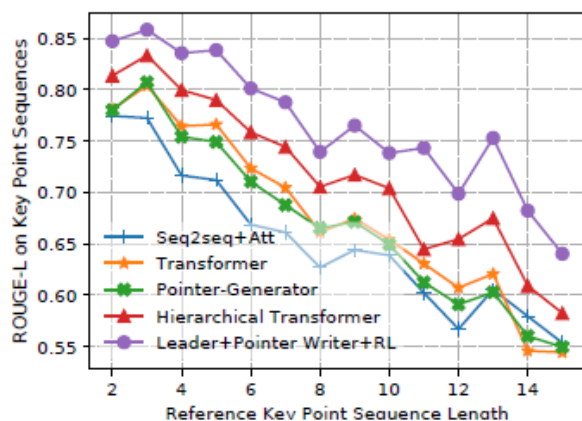
# of <dialogue, summary> pairs in training set	296,263
# of <dialogue, summary> pairs in developing set	2,963
# of <dialogue, summary> pairs in testing set	29,654
# of word vocabulary for dialogues and summaries	23,950
# of key point set	51



Dialogue Summarization for Customer Service

• Experiments

Models	Parameter size	Summary				Key point sequence			
		BLEU	ROUGE-1	ROUGE-2	ROUGE-L	BLEU	ROUGE-1	ROUGE-2	ROUGE-L
Seq2seq	16M	37.7	49.4	17.2	54.7	25.7	86.7	61.4	66.7
Seq2seq+Att	17M	44.2	58.1	32.0	59.8	35.7	91.5	73.3	72.2
Transformer	14M	48.1	57.3	36.1	63.4	42.7	92.6	77.7	76.1
Pointer-Generator	17M	51.9	83.1	61.6	64.2	40.4	97.0	78.2	75.4
Hierarchical Transformer	16M	52.3	58.6	38.4	66.9	48.4	93.7	80.1	79.2
Leader+Writer	17M	52.9	72.7	49.5	67.1	55.0	97.0	86.4	81.7
Leader+Pointer Writer	17M	55.3	78.7	54.3	68.8	56.9	97.0	83.2	83.0
Leader+Pointer Writer+RL	17M	55.9	79.3	54.9	68.9	57.0	97.2	83.0	83.2



Conclusions

- A **brief introduction** for dialogue summarization is given.
- Two recent papers are introduced
 - One is in **meeting summarization** area, which introduced the **VFOA features** to enhance the performance.
 - The other is in customer service summarization, which introduced the **key point sequence** to improve the logicality and integrity of the generated summary.

Q&A

References

- [1]. Li, Manling, et al. "Keep meeting summaries on topic: Abstractive multi-modal meeting summarization." Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019.
- [2]. Liu, Chunyi, et al. "Automatic Dialogue Summary Generation for Customer Service." Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ACM, 2019.
- [3]. Pan, Haojie, et al. "Dial2Desc: End-to-end Dialogue Description Generation." arXiv preprint arXiv:1811.00185 (2018).
- [4]. Gliwa, Bogdan, et al. "SAMSum Corpus: A Human-annotated Dialogue Dataset for Abstractive Summarization." arXiv preprint arXiv:1911.12237 (2019).
- [5]. Liu, Zhengyuan, et al. "Topic-aware Pointer-Generator Networks for Summarizing Spoken Conversations." arXiv preprint arXiv:1910.01335 (2019).
- [6]. Goo, Chih-Wen, and Yun-Nung Chen. "Abstractive Dialogue Summarization with Sentence-Gated Modeling Optimized by Dialogue Acts." 2018 IEEE Spoken Language Technology Workshop (SLT). IEEE, 2018.
- [7]. Carletta, Jean, et al. "The AMI meeting corpus: A pre-announcement." International workshop on machine learning for multimodal interaction. Springer, Berlin, Heidelberg, 2005.
- [8]. Espejel, Jessica López. "Automatic summarization of medical conversations, a review."
- [9]. Zhao, Zhou, et al. "Abstractive Meeting Summarization via Hierarchical Adaptive Segmental Network Learning." The World Wide Web Conference. ACM, 2019.
- [10]. Zhengyuan Liu et al. "Fast prototyping a dialogue comprehension system for nurse-patient conversations on symptom monitoring". NAACL 2019.
- [11]. Liu, Zhengyuan, et al. "Topic-aware Pointer-Generator Networks for Summarizing Spoken Conversations." arXiv preprint arXiv:1910.01335 (2019).