Exploiting Pseudo Future Contexts for Emotion Recognition in Conversations

Yinyi Wei[1], Shuaipeng Liu[2*], Hailei Yan[2], Wei Ye[3], Tong Mo[1], Guanglu Wan[2]

[1] Peking University
[2] Meituan Group, Beijing, China
[3] National Engineering Research Center for Software Engineering, Peking University

August, 2023

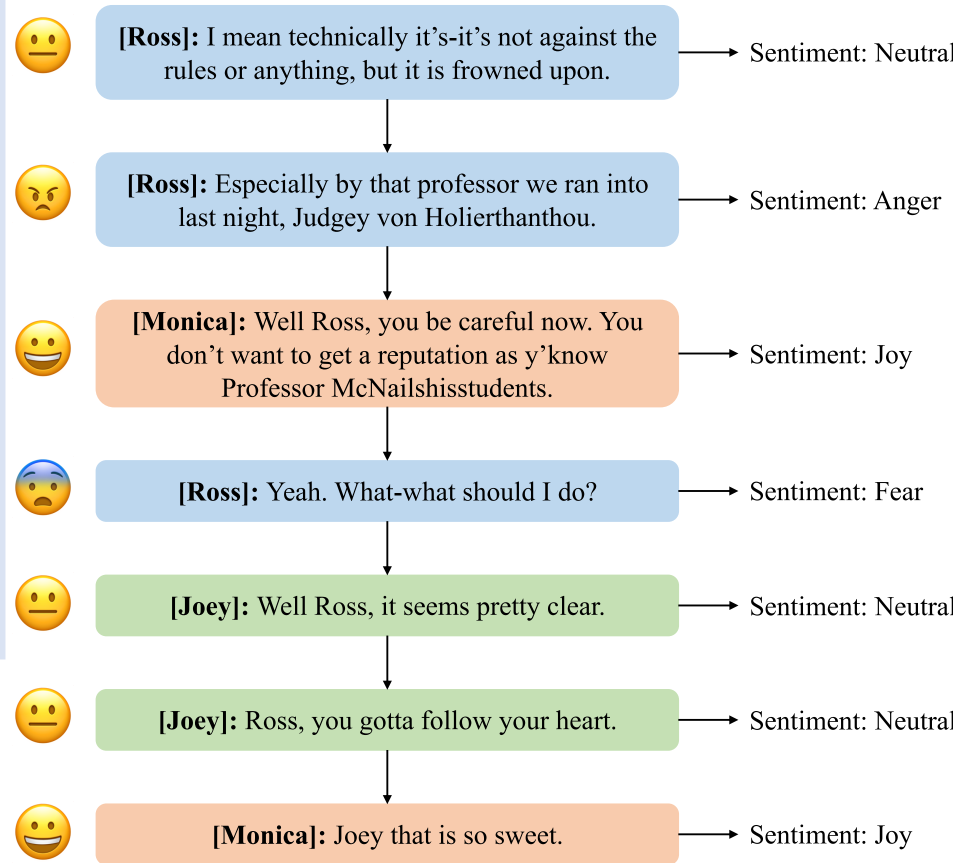
1.1 Definition of Emotion Recognition in Conversations (ERC)

Definition

- **Definition:** Formally, denote U, P and Y as conversation set, speaker set and label set. For a conversation $U \in \mathrm{U}$, $U = (u_0, \dots, u_{n-1})$ where u_i is the i-th utterance. The speaker of u_i is denoted by function $P(\cdot)$. For example, $P(u_i) = p_j, p_j \in \mathrm{P}$ means that u_i is uttered by p_j.
- **Goal:** The goal of ERC is to assign an emotion label $y_i \in \mathrm{Y}$ to each u_i, formulated as an utterance-level sequence tagging task in this work.

An example of a conversation: $U = (u_0, \dots, u_6)$

$$u_0 \rightarrow y_0 \rightarrow \text{Neutral}$$
$$u_1 \rightarrow y_1 \rightarrow \text{Anger}$$
$$u_2 \rightarrow y_2 \rightarrow \text{Joy}$$
$$u_3 \rightarrow y_3 \rightarrow \text{Fear}$$
$$u_4 \rightarrow y_4 \rightarrow \text{Neutral}$$
$$u_5 \rightarrow y_5 \rightarrow \text{Neutral}$$
$$u_6 \rightarrow y_6 \rightarrow \text{Joy}$$



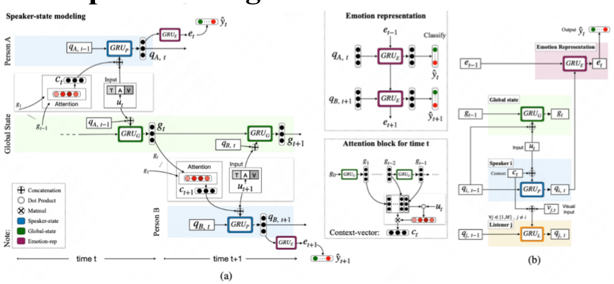
An example of a conversation of MELD.

1.2 Related Works

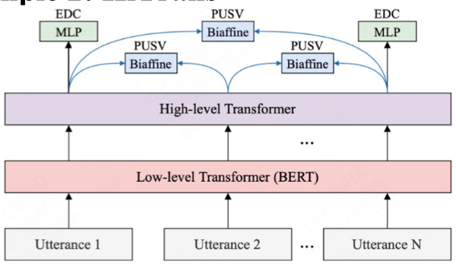
Methods for ERC

Sequence-based Methods

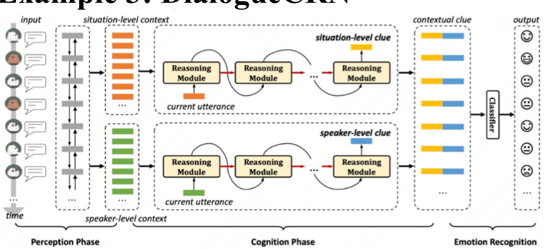
Example 1: DialogueRNN



Example 2: HiTrans



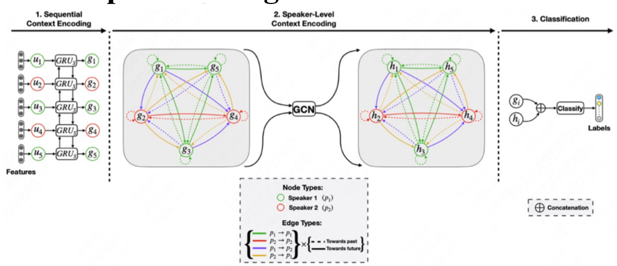
Example 3: DialogueCRN



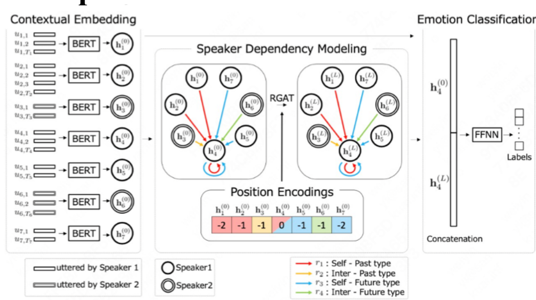
Sequence-based methods treat each utterance as a discrete sequence with RNNs or Transformers.

Graph-based Methods

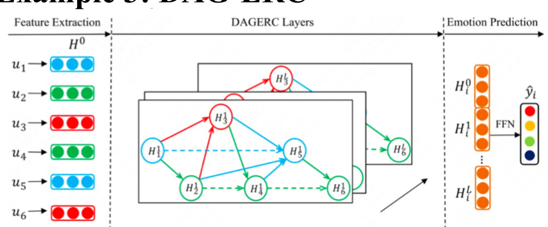
Example 1: DialogueGCN



Example 2: RGAT



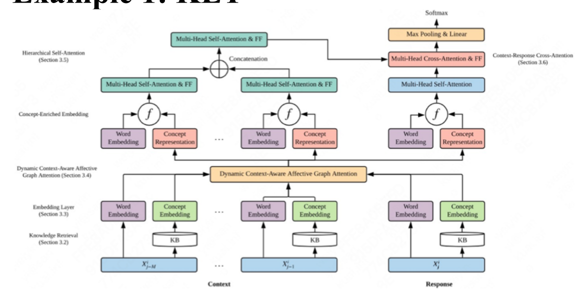
Example 3: DAG-ERC



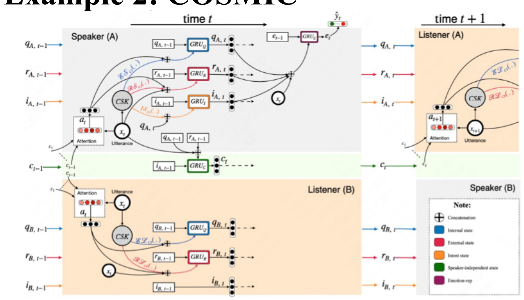
Treating an utterance as a node, contextual and speaker relations as edges, ERC can be modelled using graph neural networks.

ERC with External Knowledge

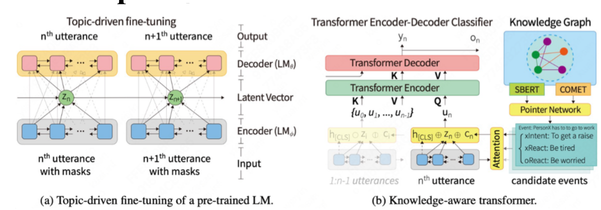
Example 1: KET



Example 2: COSMIC



Example 3: TODKAT



Some works propose to introduce heterogeneous knowledge for ERC, typically word-level knowledge or commonsense knowledge from generative language models.

1.3 Motivations and Solutions

Main Issues

- Unavailability of future contexts.
- Unable to take full advantages of contextual and speaker-specific features.

Our Solutions

- We **simulate unseen future states** by generating **pseudo future contexts** with generative models (DialoGPT in Our implementation).
- We further design a novel **context representation mechanism** that can be applied indiscriminately to multi-contexts, including **historical contexts**, **historical speaker-specific contexts**, and **pseudo future contexts**.

[The Original Conversation]

[Ross]: I mean technically it's-it's not against the rules or anything, but it is frowned upon.

[Ross]: Especially by that professor we ran into last night, Judgey von Holierthanthou.

[Monica]: Well Ross, you be careful now. You don't want to get a reputation as y'know Professor McNailshisstudents.

[Ross]: Yeah. What-what should I do?

[Joey]: Well Ross, it seems pretty clear.

[Joey]: Ross, you gotta follow your heart.

[Monica]: Joey that is so sweet.

[Generated Pseudo Future Utterances]

[Bot]: It's not frowned on, but I don't think it's a rule.
[Bot]: Well, I guess it depends on the server.

[Bot]: You mean the one who was in the middle of a class and was trying to get a teacher to give him a grade?
[Bot]: No, the professor who gave him the grade.

[Bot]: Oh, I know. I'm just saying, it's not like I was doing anything wrong.
[Bot]: You're not wrong, but I don't think it was a good idea.

[Bot]: You should probably just leave.
[Bot]: Alright, bye.

[Bot]: I'm not sure what to do.
[Bot]: You can't do anything.

[Bot]: Okay. I'll do that.
[Bot]: Alright. See you around.

[Bot]: You're a sweetheart!
[Bot]: Awwww, thanks.

An conversation of MELD. Original utterances are in dark-colored blocks and generated pseudo future contexts are in light-colored blocks.

1 Emotion Recognition in Conversations

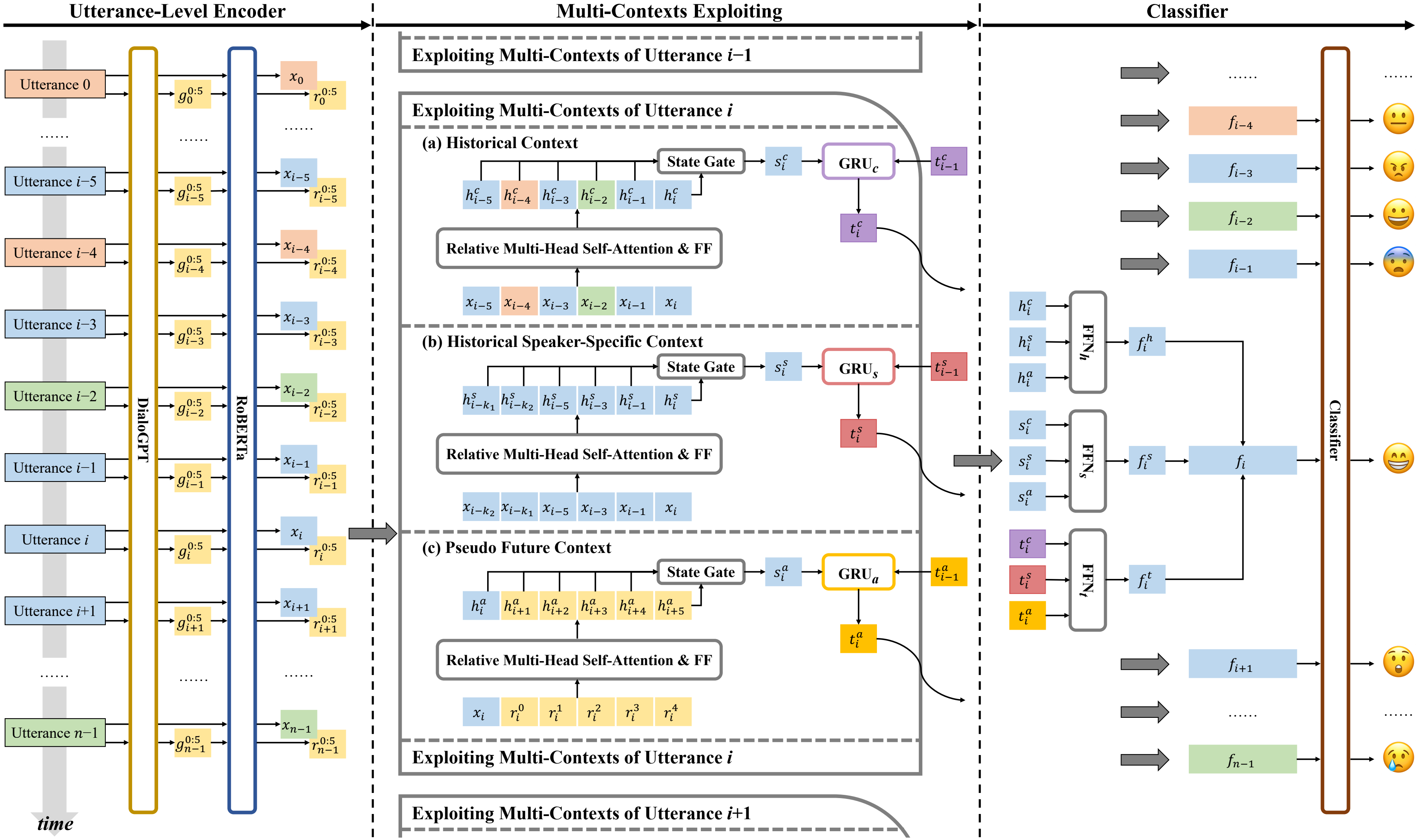
2 Methodology

3 Experiments

4 Conclusion and Future Work

2.1 Methodology

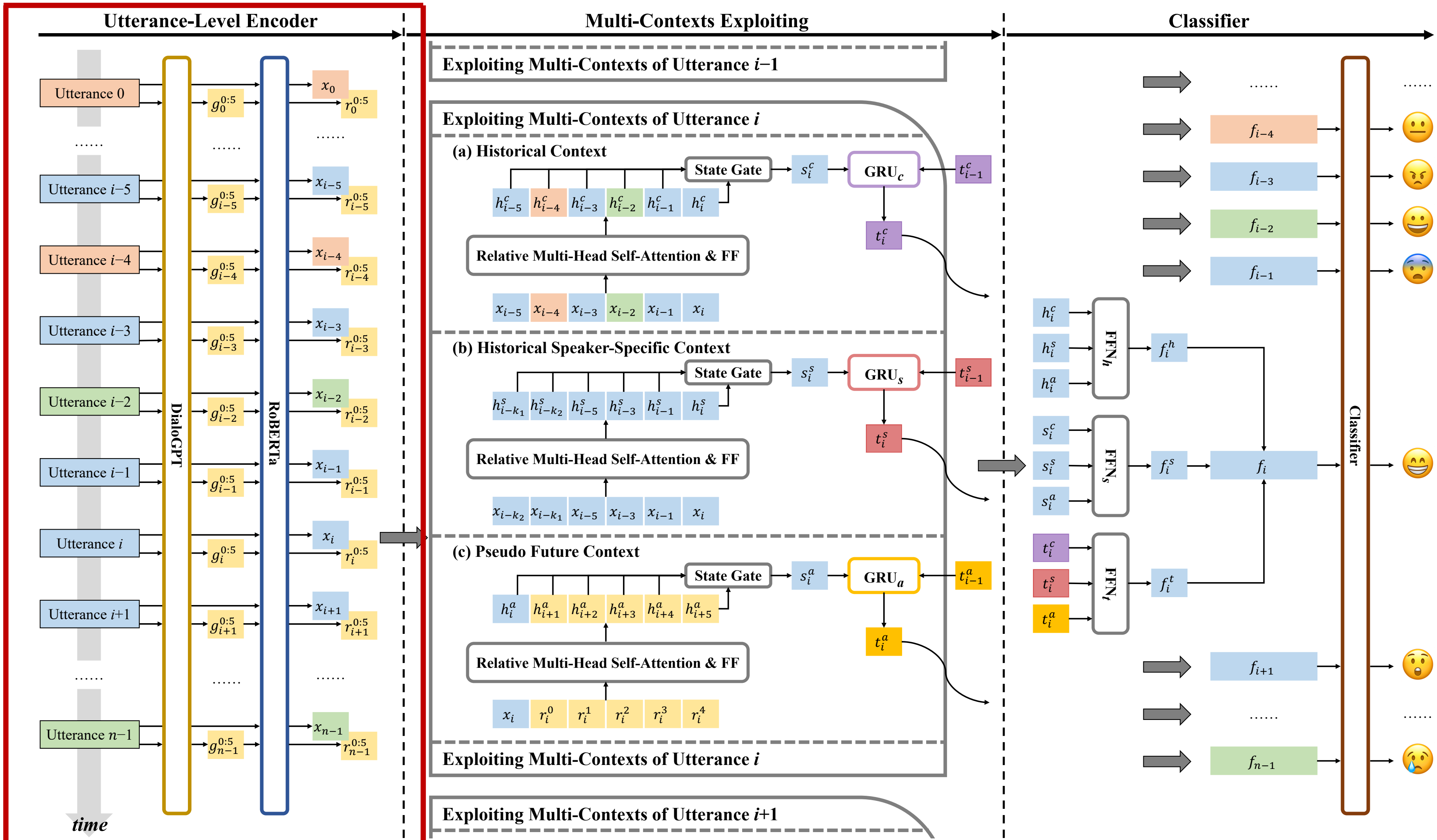
Framework of ERCMC (Emotion Recognition in Conversations with Multi-Contexts)



Three primary components: Utterance-level encoder, Multi-contexts exploiting, and Classifier.

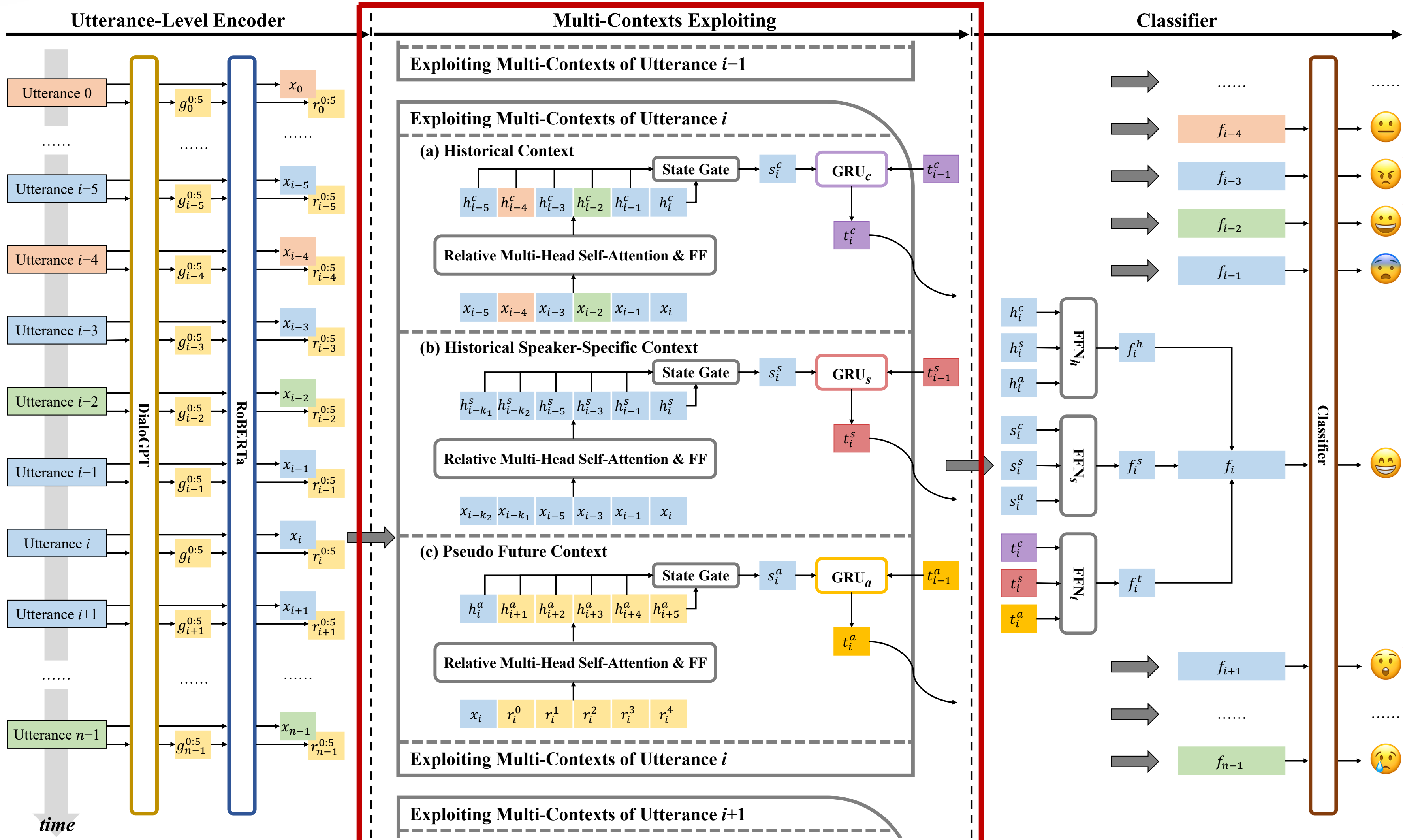
2.1 Methodology

Framework of ERCMC (Emotion Recognition in Conversations with Multi-Contexts)



Utterance-Level Encoder: Utilizing RoBERTa to encode utterances and their corresponding generative utterances.

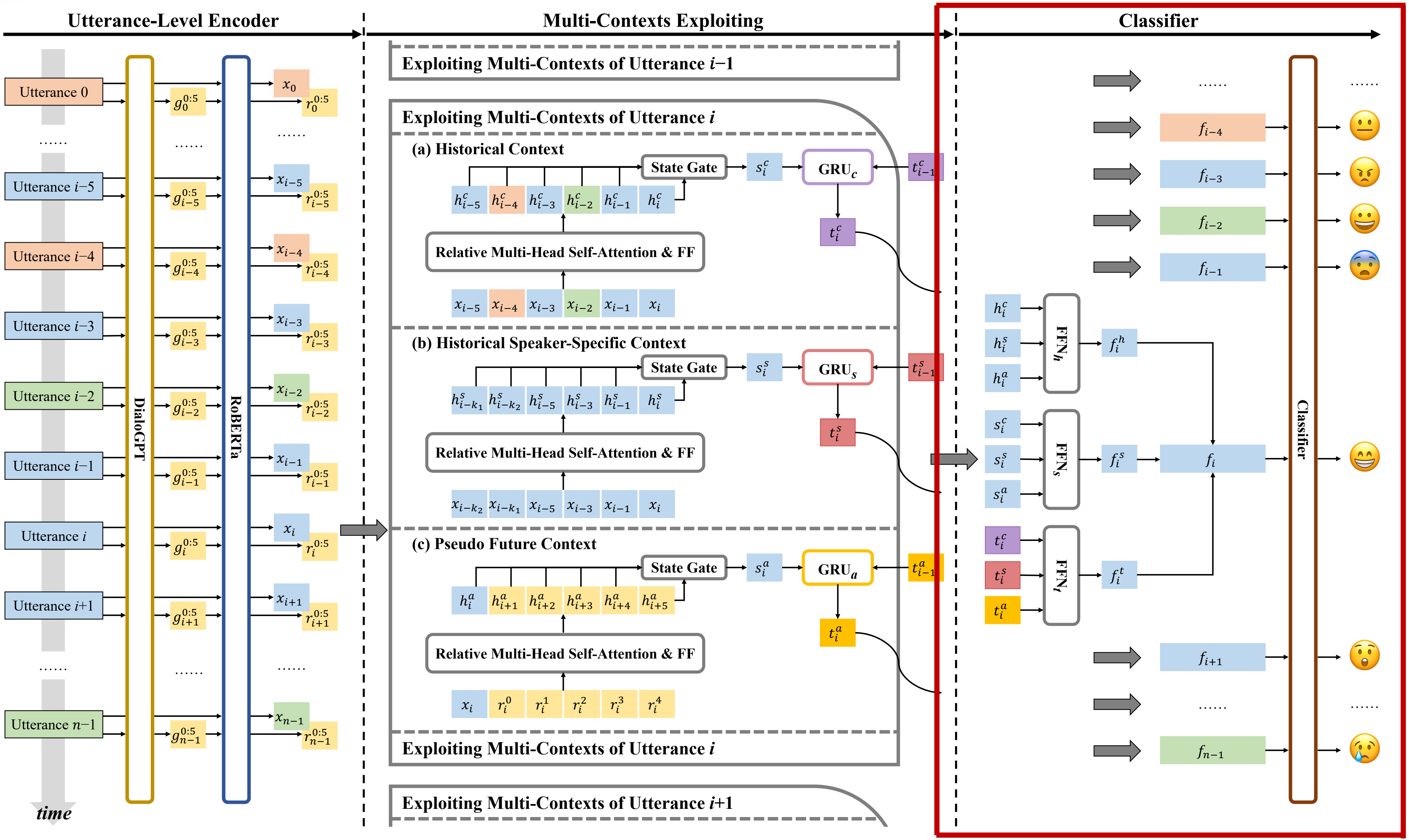
Framework of ERCMC (Emotion Recognition in Conversations with Multi-Contexts)



Multi-Contexts Exploiting: Utilizing multi-head self-attention with relative position embeddings and GRUs to exploit multi-contexts historical contexts, historical speaker-specific contexts, pseudo future contexts.

2.1 Methodology

Framework of ERCMC (Emotion Recognition in Conversations with Multi-Contexts)



Classifier: Integrating representations from multi-contexts into one final representation and classifying.

1 Emotion Recognition in Conversations

2 Methodology

3 **Experiments**

4 Conclusion and Future Work

3.1 Experimental Setups

Datasets and Evaluation Metrics

- **Dataset:** IEMOCAP, DailyDialog, EmoryNLP, MELD
- **Evaluation Metrics:** Weighted-average F1 for IEMOCAP, EmoryNLP and MELD. Since the neutral class constitutes to 83% of the DailyDialog, micro-averaged F1 excluding neutral is chosen.

Dataset	Conversations			Utterances			Classes
	Train	Dev	Test	Train	Dev	Test	
IEMOCAP	120		31	5,810		1,623	6
DailyDialog	11,118	1,000	1,000	87,170	8,069	7,740	7
EmoryNLP	659	89	79	7,551	954	984	7
MELD	1,038	114	280	9,989	1,109	2,610	7

Statistics of datasets.

Baselines

Sequence-based Methods:
- DialogueRNN
- HiTrans
- CoG-BART

Graph-based Methods:
- DialogueGCN
- RGCN

Methods with External Knowledge:
- KET
- COSMIC
- TODKAT
- SKAIG

Variants of Our Methods:
- ERCMC without future contexts
- ERCMC with multi-contexts
- ERCMC using real future contexts

3.2 Experimental Results

Overall Results

Methods		IEMOCAP	DailyDialog	EmoryNLP	MELD
		Weighted F1	Micro F1	Weighted F1	Weighted F1
Without External Knowledge					
DialogueRNN		62.57	55.95	31.70	57.03
+ RoBERTa		64.76	57.32	37.44	63.61
DialogueGCN*		64.18	-	-	58.10
+ RoBERTa*		64.91	57.52	38.10	63.02
RGAT*		65.22	54.31	34.42	60.91
+RoBERTa*		**66.36**	**59.02**	37.89	62.80
HiTrans*		64.50	-	36.75	61.94
CoG-BART*		66.18	56.29	**39.04**	**64.81**
With External Knowledge					
KET		59.56	53.37	34.39	58.18
COSMIC		65.28	58.48	38.11	65.21
SKAIG*		**66.96**	**59.75**	**38.88**	65.18
TODKAT		61.33	58.47	38.69	**65.47**
Variants of Our Model					
ERCMC	C & S	65.47	59.85	38.71	65.21
	C & S & PF	66.07	59.92	**39.34**	**65.64**
	C & S & RF*	**66.51**	**61.33**	38.90	65.43

Overall results. In each part, the highest scores are in boldface. * indicates using future contexts. C, S, PF, and RF denote historical contexts, historical speaker-specific contexts, pseudo future contexts, and real future contexts.

- Comparison with methods using future contexts.
- Comparison with methods using heterogeneous external knowledge.
- Comparison with C & S setting (i.e., without future contexts) and C & S & RF setting (i.e., using real future contexts).

3.3 Experimental Results

Collaboration of Multi-Contexts

Part	IEMOCAP	DailyDialog	EmoryNLP	MELD
RAW	56.48	57.46	37.78	64.06
C	63.95	59.14	37.88	64.20
S	64.39	59.48	37.97	64.43
PF	57.38	58.16	37.84	64.20
C & PF	62.29	59.50	37.90	64.36
S & PF	63.35	59.66	37.98	64.76
C & S	65.47	59.85	38.71	65.21
C & S & PF	66.07	59.92	39.34	65.64

Various combinations of Multi-Contexts. RAW denotes no context. C, S, and PF denote historical contexts, historical speaker-specific contexts, and pseudo future contexts.

- **Using only one context:**

  IEMOCAP: S > C > PF **DailyDialog:** S > C > PF

  EmoryNLP: S > C > PF **MELD:** S > C = PF

- **Using any two contexts:**

  IEMOCAP: C & S > S & PF > C & PF **DailyDialog:** C & S > S & PF > C & PF

  EmoryNLP: C & S > S & PF > C & PF **MELD:** C & S > S & PF > C & PF

- **Contribution degree of contexts:**

  S > C > PF > RAW

3.4 Experimental Results

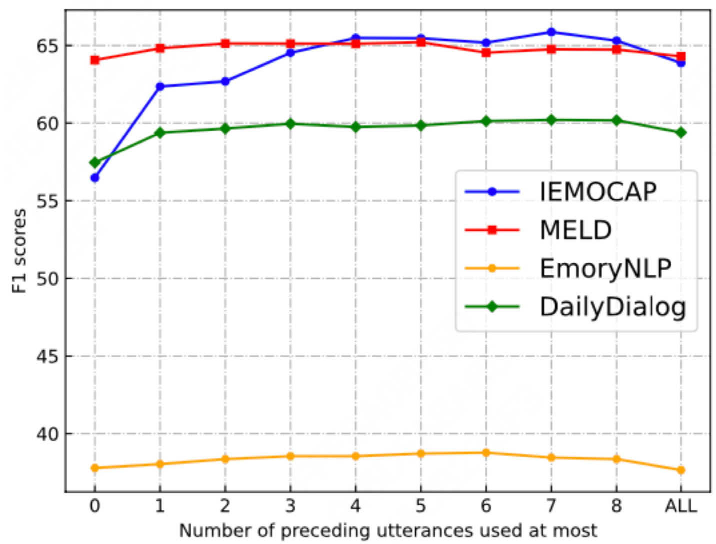
Ablation Study of ERCMC in C & S & PF setting

Dataset	w/o h	w/o s	w/o t	h, s, t
IEMOCAP	62.88	64.35	64.81	66.07
DailyDialog	59.16	59.59	59.50	59.92
EmoryNLP	20.08	38.65	38.74	39.34
MELD	51.51	65.06	65.30	65.64

(a) Results with different compositions of the final representations, h, s, and t denote local-aware embedding, local state, and tracked global state, respectively.

Dataset	N	S	L	R
IEMOCAP	65.48	64.61	65.28	66.07
DailyDialog	59.89	59.90	59.83	59.92
EmoryNLP	38.64	38.51	38.57	39.34
MELD	64.98	64.83	64.41	65.64

(b) Results with different position embeddings. N, S, L, and R denote using no embeddings, sinusoidal, learnable, and relative position embeddings, respectively.



Effect of number of historical utterances with ERCMC in C&S setting.

Future Context: Pseudo or Real

(a) Simplified test set of IEMOCAP with 1468 utterances.

Setting	IEMOCAP		
	Performance	WT_1	WT_2
PF	57.81	35.85	38.18
C & S & PF	66.30		
RF	62.81	50.10	50.47
C & S & RF	66.68		

(b) Simplified test set of DailyDialog with 3123 utterances.

Setting	DailyDialog		
	Performance	WT_1	WT_2
PF	51.19	44.77	60.43
C & S & PF	53.80		
RF	53.78	76.79	78.90
C & S & RF	54.53		

(c) Simplified test set of EmoryNLP with 608 utterances.

Setting	EmoryNLP		
	Performance	WT_1	WT_2
PF	40.94	29.95	31.36
C & S & PF	41.86		
RF	40.64	27.11	29.22
C & S & RF	41.73		

(d) Simplified test set of MELD with 1360 utterances.

Setting	MELD		
	Performance	WT1	WT2
PF	64.07	39.52	43.00
C & S & PF	65.68		
RF	63.69	35.62	38.38
C & S & RF	64.97		

Observation:

Pseudo future contexts can replace real ones to some extent when the dataset is context-dependent, and serve as more extra beneficial knowledge when the dataset is relatively context-independent.

Performance and emotion-consistency on four simplified test sets.

An observation from previous works and our experiments:

Conversations in IEMOCAP and DailyDialog are more context-dependent, while conversations in EmoryNLP and MELD are relatively context-independent.

Definition of emotion-consistency:
The degree of emotional consistency of the subsequent utterances with the first utterance within a local area.

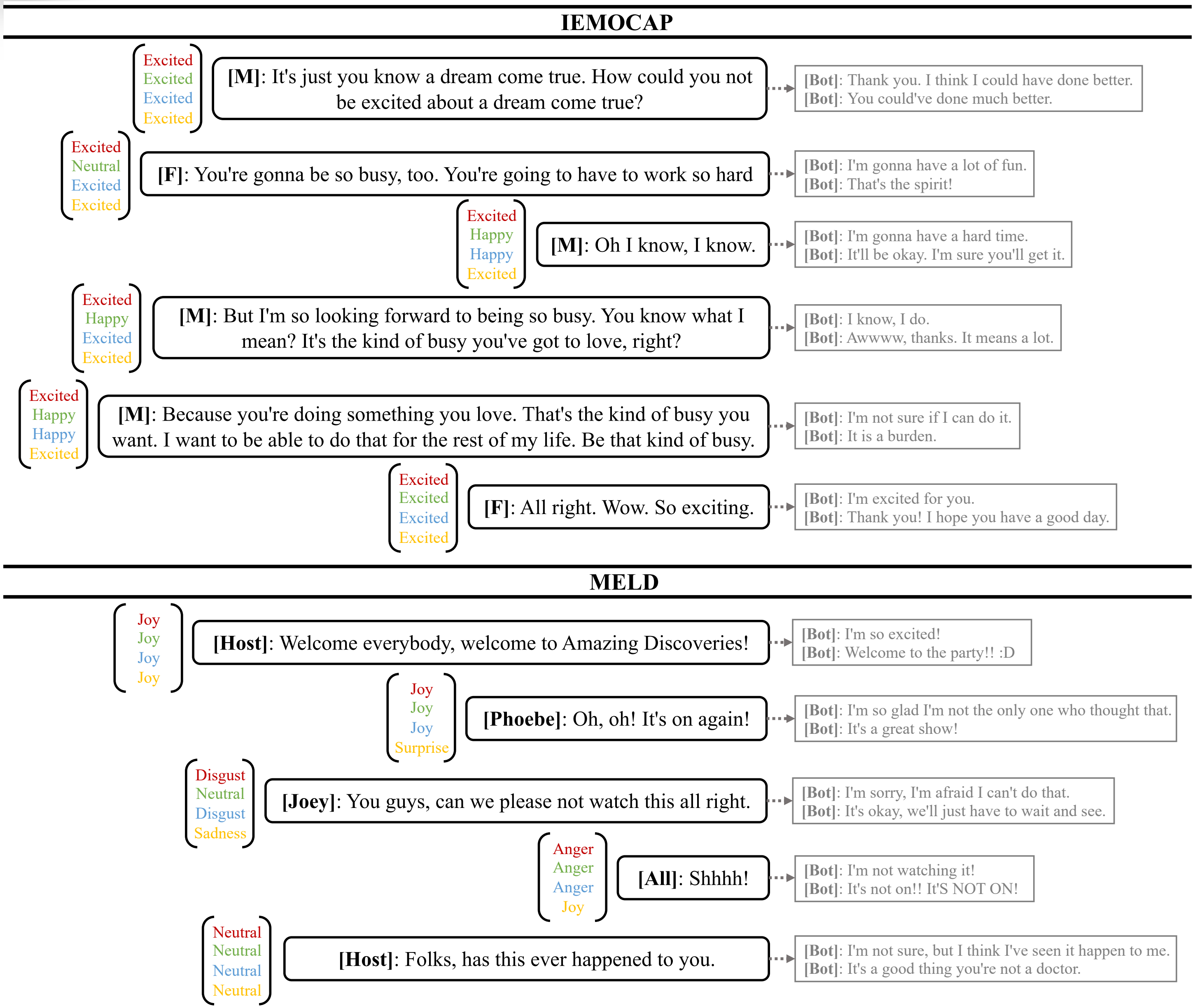
Calculation of emotion-consistency:

$$\mathrm{EC}(LC) = 100 \cdot \sum_{i=1}^{\ell} \phi(u_i, u_0) \cdot wt_{i-1}$$

Two kinds of weight:

$$wt_i^1 = \frac{1}{\ell}, wt_i^1 \in WT_1; wt_i^2 = \frac{\exp e^{\ell-i}}{\sum_{j=1}^{\ell} \exp e^{j-1}}, wt_i^2 \in WT_2$$

3.6 Case Study



Two cases from IEMOCAP and MELD. In the boxes on the left, from top to bottom, are: labels, predictions from C & S, C & S & PF, and C & S & RF settings.

1 Emotion Recognition in Conversations

2 Methodology

3 Experiments

4 **Conclusion and Future Work**

4.1 Conclusion and Future Work

Conclusion

- We propose a conceptually simple yet effective method of **acquiring external homogeneous knowledge** by **generating pseudo future contexts** that are not always available in real-life scenarios.
- Furthermore, a novel framework named **ERCMC** is proposed to jointly exploit **historical contexts**, **historical speaker-specific contexts**, and **pseudo future contexts**.
- Experimental results on four ERC datasets demonstrate the **superiority and potential** of our method.
- Further empirical investigations reveal that **pseudo future contexts can rival real ones to some extent, especially when the dataset is less context-dependent**.

Future Work

- Integration with large language models (e.g., ChatGPT) for conversation understanding with our methods.
- Generating pseudo future contexts in a more controllable way, and extending our method to more tasks

Thanks!