MPC-BERT: A Pre-Trained Language Model for Multi-Party Conversation Understanding

Jia-Chen Gu[1], Chongyang Tao[2], Zhen-Hua Ling[1], Can Xu[2], Xiubo Geng[2], Daxin Jiang[2]

[1]National Engineering Laboratory for Speech and Language Information Processing,
University of Science and Technology of China, Hefei, China

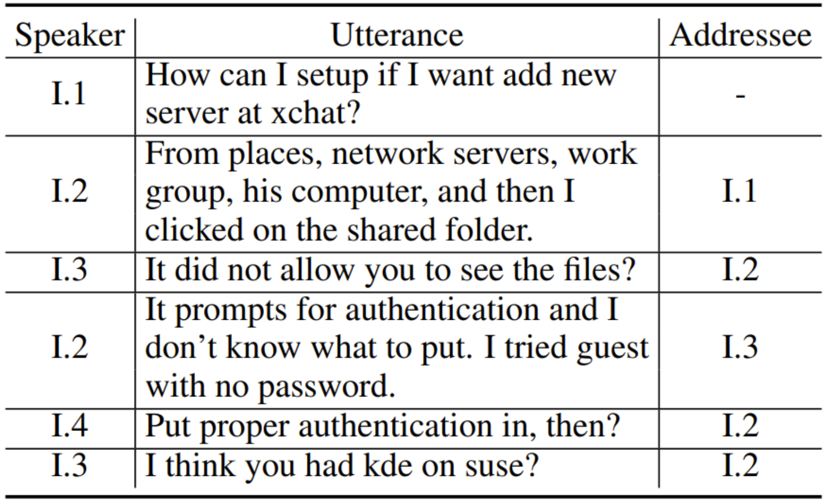
[2]Microsoft, Beijing, China

Outline

- **Introduction**
- MPC-BERT
- Downstream Tasks
- Experiments
- Conclusion

Introduction

- **Two-Party Conversation**: Utterances are posted one by one between two interlocutors sequentially.
- **Multi-Party Conversation**: Each utterance can be spoken by anyone and address anyone else in this conversation.



Speaker	Utterance	Addressee
I.1	How can I setup if I want add new server at xchat?	-
I.2	From places, network servers, work group, his computer, and then I clicked on the shared folder.	I.1
I.3	It did not allow you to see the files?	I.2
I.2	It prompts for authentication and I don't know what to put. I tried guest with no password.	I.3
I.4	Put proper authentication in, then?	I.2
I.3	I think you had kde on suse?	I.2

Table 1: An MPC example in Ubuntu IRC channel. Here, "I." is the abbreviation of "interlocutor".

Related Work

- The representation learning of interlocutors and utterances in MPC are either separate or interactive from two representation spaces.
- Pre-trained language models still overlook the inherent relationships between utterances and interlocutors, such as “address-to”.
- Existing studies design models for each individual task in MPC separately, while neglect the complementary information among these tasks.

Outline

- Introduction
- **MPC-BERT**
- Downstream Tasks
- Experiments
- Conclusion

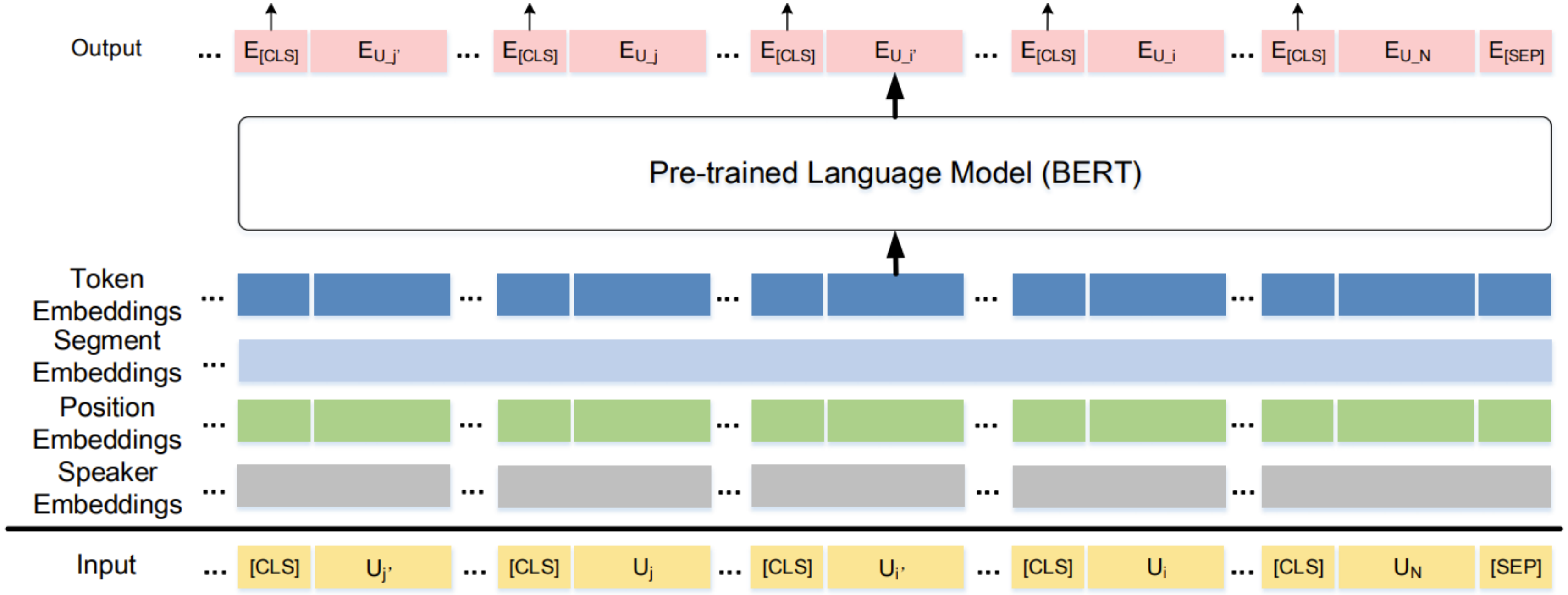
MPC-BERT

Our goal is to build a pre-trained language model for universal MPC understanding. MPC-BERT jointly learns who says what to whom in MPC by designing self-supervised tasks, so that it can produce better interlocutor and utterance representations which can be effectively generalized to multiple downstream tasks of MPC.

- Interlocutor Structure Modeling.
- Utterance Semantics Modeling.

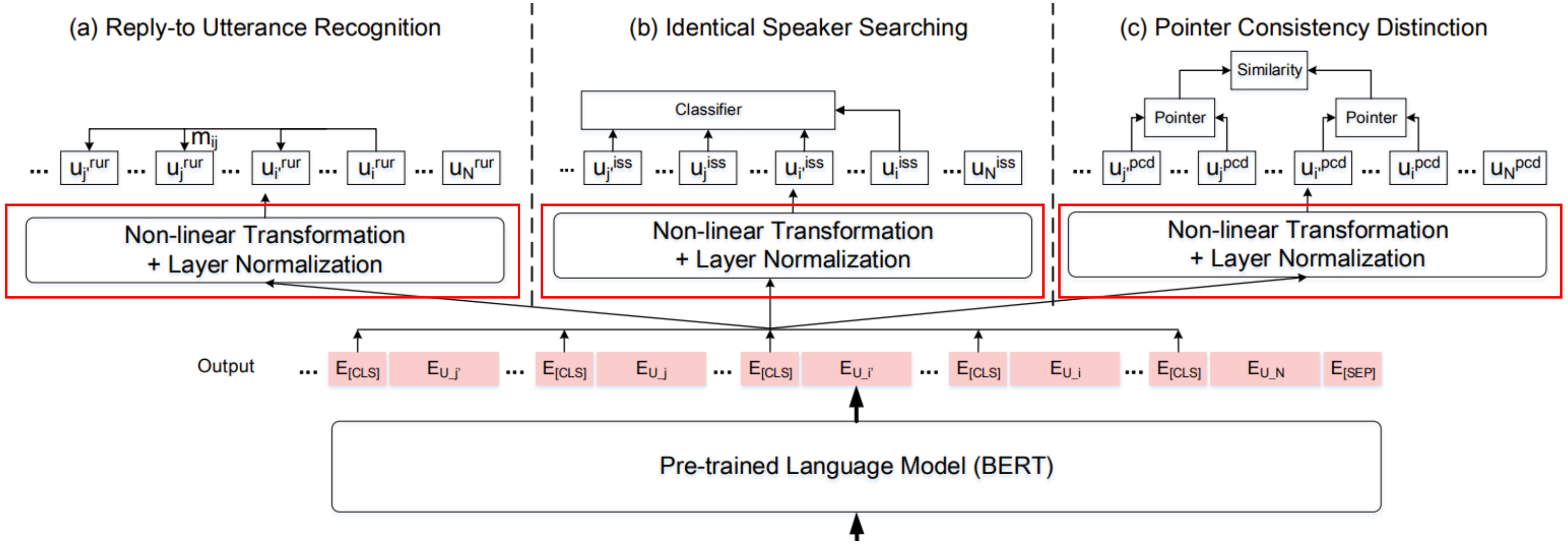
Model overview of MPC-BERT

- A [CLS] token is inserted at the start of each utterance.
- Position-based speaker embeddings are introduced considering that the set of interlocutors are inconsistent in different conversations.



Interlocutor Structure Modeling

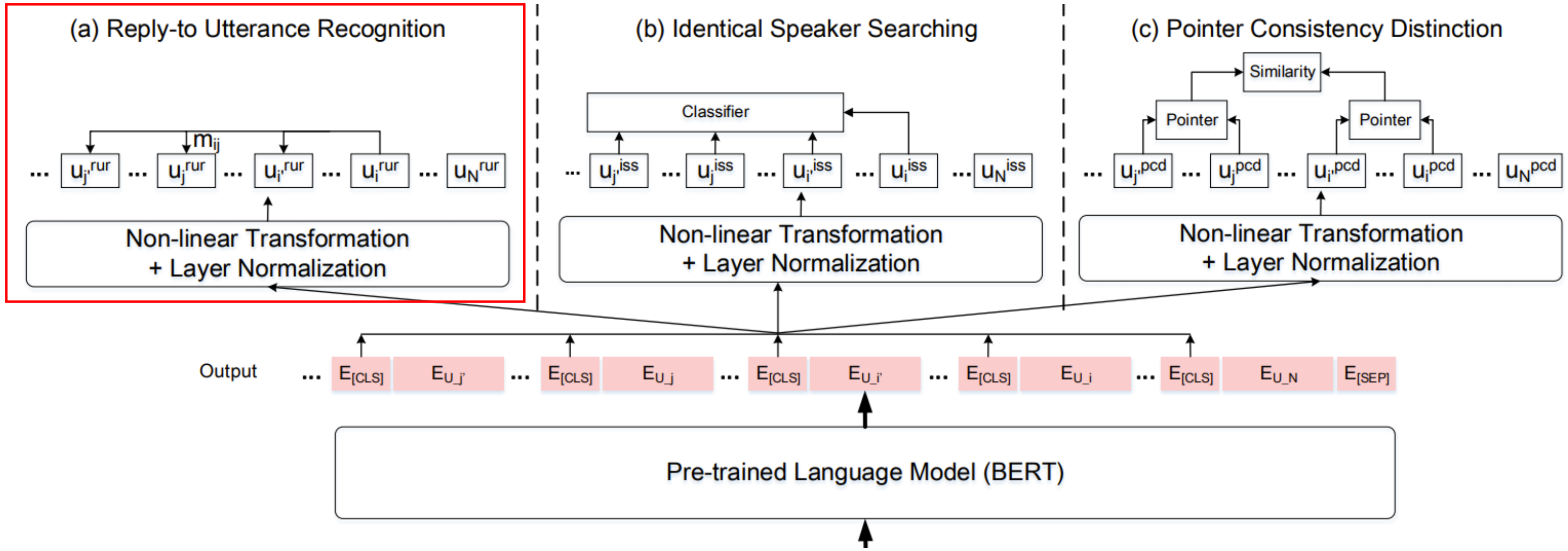
- Extract the contextualized representations for each [CLS] token representing individual utterances.
- A task-dependent non-linear transformation is placed on top of BERT.
- Encoding the input data only once is computation-efficient.



(a) Reply-to Utterance Recognition
(b) Identical Speaker Searching
(c) Pointer Consistency Distinction
Similarity
Classifier
Pointer
Pointer
m_{ij}
... $u_{j'}^{rur}$ ... u_j^{rur} ... $u_{i'}^{rur}$ ... u_i^{rur} ... u_N^{rur}
... $u_{j'}^{iss}$ ... u_j^{iss} ... $u_{i'}^{iss}$ ... u_i^{iss} ... u_N^{iss}
... $u_{j'}^{pcd}$ ... u_j^{pcd} ... $u_{i'}^{pcd}$ ... u_i^{pcd} ... u_N^{pcd}
Non-linear Transformation + Layer Normalization
Non-linear Transformation + Layer Normalization
Non-linear Transformation + Layer Normalization
Output
... $E_{[CLS]}$ $E_{U\_j'}$ ... $E_{[CLS]}$ E_{U_j} ... $E_{[CLS]}$ $E_{U\_i'}$ ... $E_{[CLS]}$ E_{U_i} ... $E_{[CLS]}$ E_{U_N} $E_{[SEP]}$
Pre-trained Language Model (BERT)

Interlocutor Structure Modeling

- **Reply-to Utterance Recognition**: To enable the model to recognize the addressee of each utterance, this task is proposed to learn which preceding utterance the current utterance replies to.

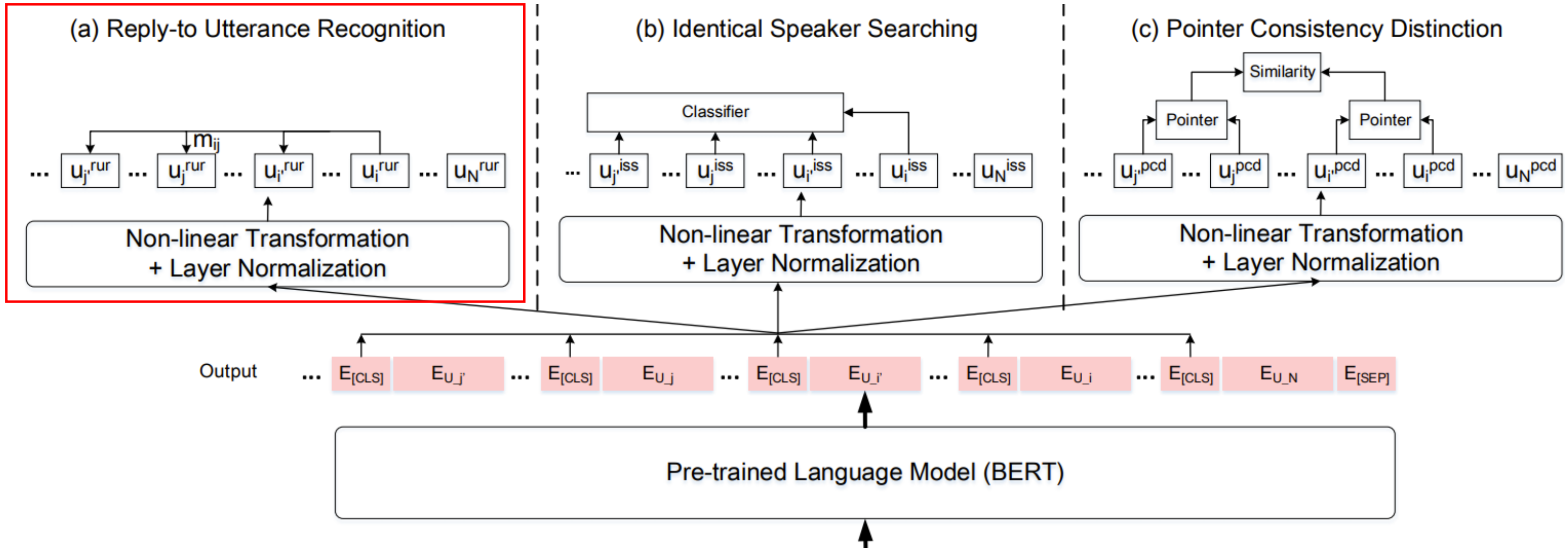

(a) Reply-to Utterance Recognition
m_ij
... u_j'^rur ... u_j^rur ... u_i'^rur ... u_i^rur ... u_N^rur
Non-linear Transformation + Layer Normalization
(b) Identical Speaker Searching
Classifier
... u_j'^iss ... u_j^iss ... u_i'^iss ... u_i^iss ... u_N^iss
Non-linear Transformation + Layer Normalization
(c) Pointer Consistency Distinction
Similarity
Pointer
Pointer
... u_j'^pcd ... u_j^pcd ... u_i'^pcd ... u_i^pcd ... u_N^pcd
Non-linear Transformation + Layer Normalization
Output
... E[CLS] EU_j' ... E[CLS] EU_j ... E[CLS] EU_i' ... E[CLS] EU_i ... E[CLS] EU_N E[SEP]
Pre-trained Language Model (BERT)

Interlocutor Structure Modeling

- **Reply-to Utterance Recognition**: For a specific utterance U$_i$, its matching scores with all its preceding utterances are calculated as

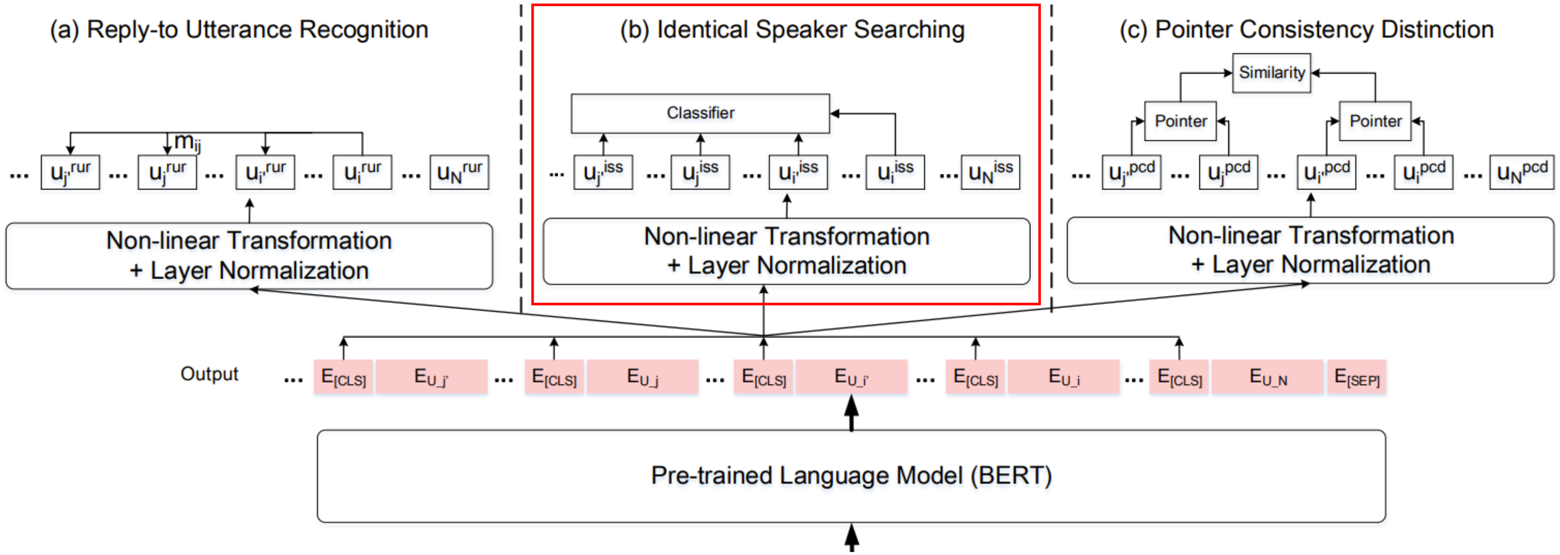
$$m_{ij} = \textbf{softmax}(\mathbf{u}_i^{rur\top} \cdot \mathbf{A}^{rur} \cdot \mathbf{u}_j^{rur})$$

- Dynamic sampling + Cross-entropy loss minimization $\mathcal{L}_{rur} = -\sum_{i \in \mathbb{S}} \sum_{j=1}^{i-1} y_{ij}\, log(m_{ij})$


(a) Reply-to Utterance Recognition
m_{ij}
... $u_{j'}^{rur}$ ... u_j^{rur} ... $u_{i'}^{rur}$ ... u_i^{rur} ... u_N^{rur}
Non-linear Transformation + Layer Normalization
(b) Identical Speaker Searching
Classifier
... $u_{j'}^{iss}$ ... u_j^{iss} ... $u_{i'}^{iss}$ ... u_i^{iss} ... u_N^{iss}
Non-linear Transformation + Layer Normalization
(c) Pointer Consistency Distinction
Similarity
Pointer
Pointer
... $u_{j'}^{pcd}$ ... u_j^{pcd} ... $u_{i'}^{pcd}$ ... u_i^{pcd} ... u_N^{pcd}
Non-linear Transformation + Layer Normalization
Output
... $E_{[CLS]}$ $E_{U\_j'}$ ... $E_{[CLS]}$ E_{U_j} ... $E_{[CLS]}$ $E_{U\_i'}$ ... $E_{[CLS]}$ E_{U_i} ... $E_{[CLS]}$ E_{U_N} $E_{[SEP]}$
Pre-trained Language Model (BERT)

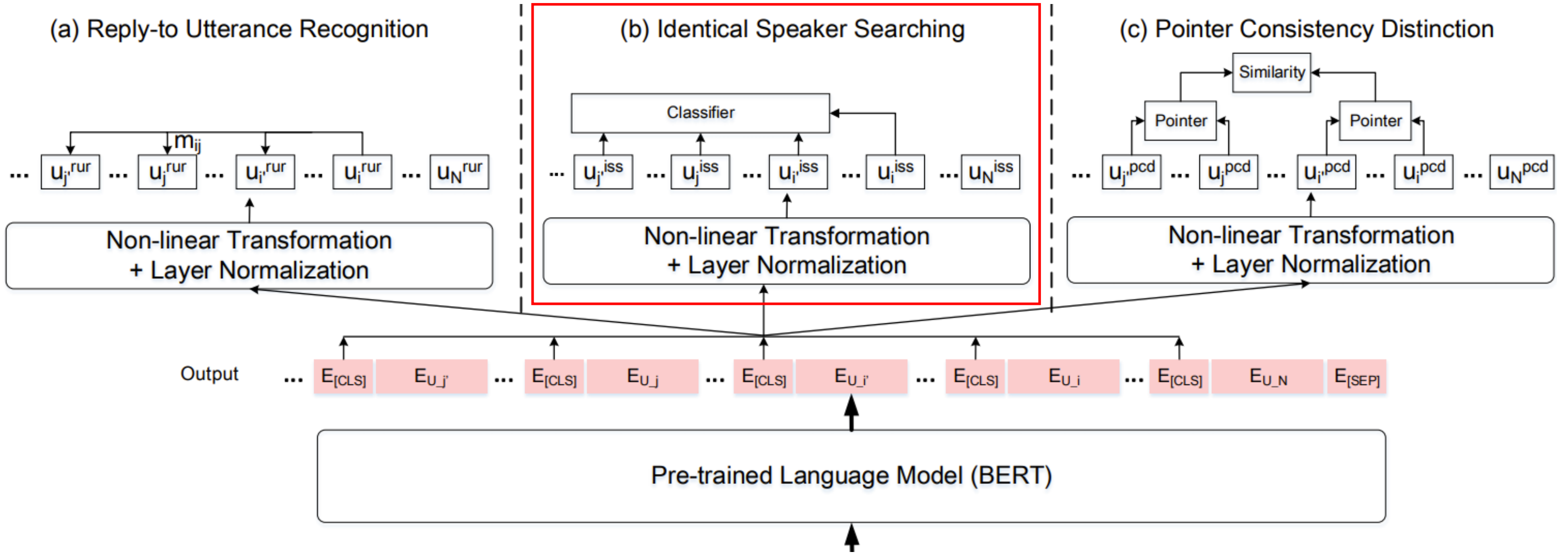
Interlocutor Structure Modeling

- **Identical Speaker Searching:** Since the set of interlocutors vary across conversations, the task of predicting the speaker of an utterance is reformulated as searching for the utterances sharing the identical speaker.



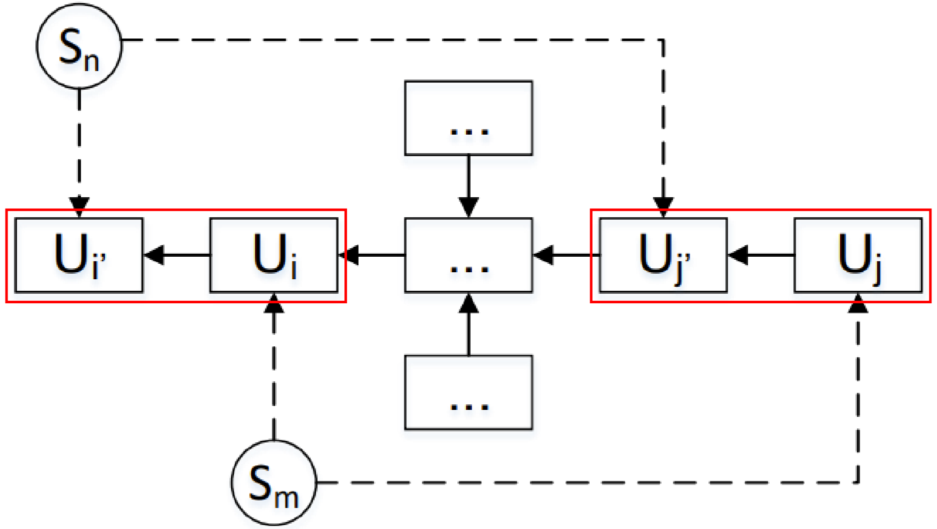
Interlocutor Structure Modeling

- **Identical Speaker Searching:** Mask the speaker embedding of a specific utterance in the input representation, and calculate the probability of two utterances sharing the same speaker.
- Dynamic sampling + Cross-entropy loss minimization


(a) Reply-to Utterance Recognition
(b) Identical Speaker Searching
(c) Pointer Consistency Distinction
Similarity
Pointer
Pointer
Classifier
m_{ij}
u_j^{rur} $u_{j'}^{rur}$ u_i^{rur} $u_{i'}^{rur}$ u_N^{rur}
u_j^{iss} $u_{j'}^{iss}$ u_i^{iss} $u_{i'}^{iss}$ u_N^{iss}
$u_{j'}^{pcd}$ u_j^{pcd} $u_{i'}^{pcd}$ u_i^{pcd} u_N^{pcd}
Non-linear Transformation + Layer Normalization
Non-linear Transformation + Layer Normalization
Non-linear Transformation + Layer Normalization
Output
$E_{[CLS]}$ $E_{U\_j'}$ $E_{[CLS]}$ E_{U_j} $E_{[CLS]}$ $E_{U\_i'}$ $E_{[CLS]}$ E_{U_i} $E_{[CLS]}$ E_{U_N} $E_{[SEP]}$
Pre-trained Language Model (BERT)

Interlocutor Structure Modeling

- **Pointer Consistency Distinction:** A pair of utterances representing the "reply-to" relationship is defined as a speaker-to-addressee pointer.
- We assume that the representations of two pointers directing from the same speaker to the same addressee should be consistent.

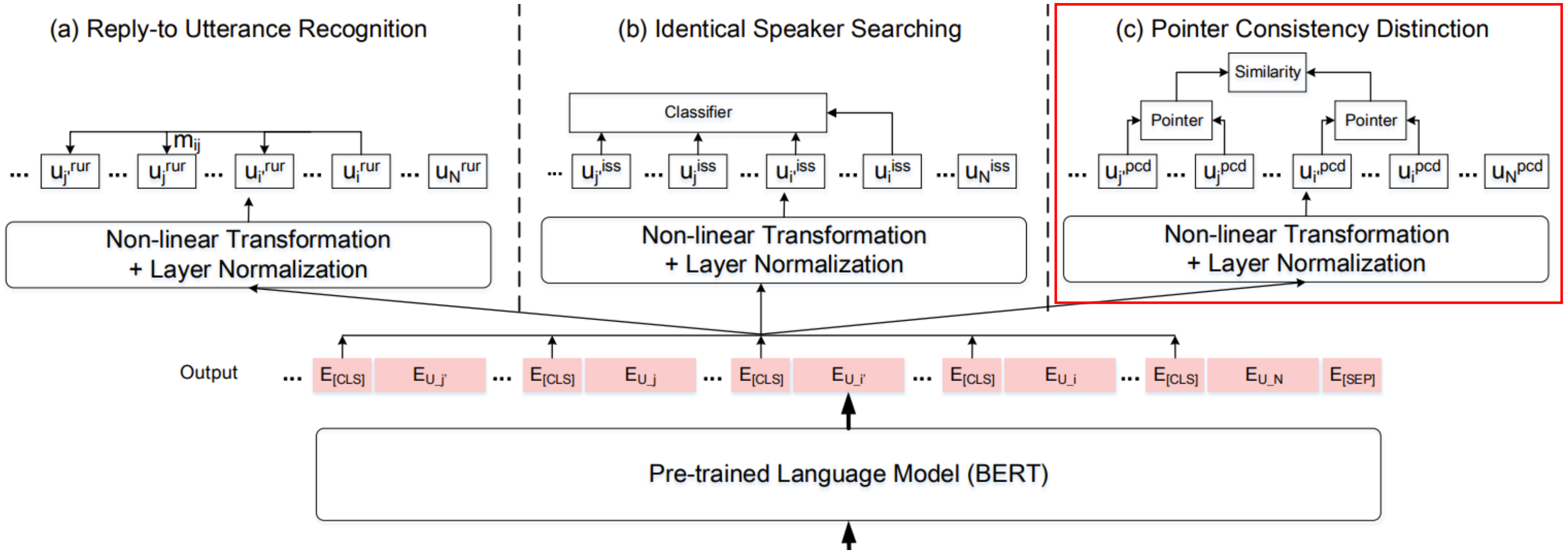


Interlocutor Structure Modeling

- **Pointer Consistency Distinction:** Capture the pointer information contained in each utterance tuple as

$$\mathbf{p}_{ii'} = [\mathbf{u}_i^{pcd} - \mathbf{u}_{i'}^{pcd}; \mathbf{u}_i^{pcd} \odot \mathbf{u}_{i'}^{pcd}]$$

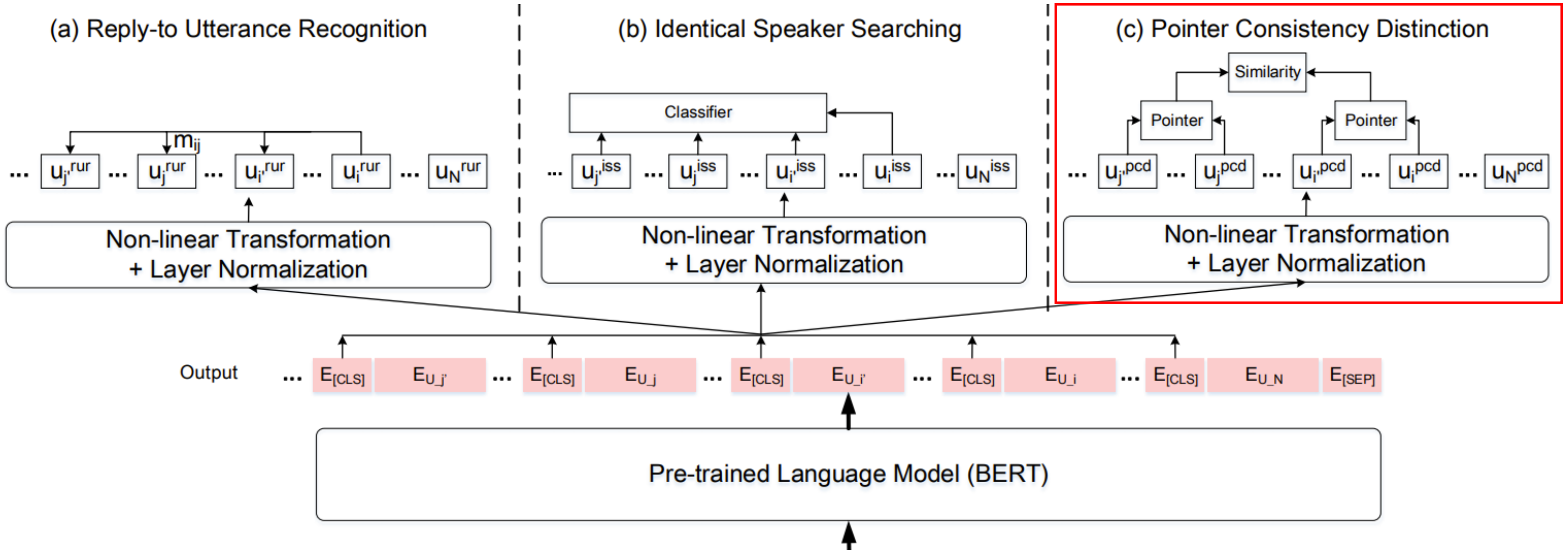
$$\bar{\mathbf{p}}_{ii'} = \mathbf{ReLU}(\mathbf{p}_{ii'} \cdot \mathbf{W}_{pcd} + \mathbf{b}_{pcd})$$


(a) Reply-to Utterance Recognition
(b) Identical Speaker Searching
(c) Pointer Consistency Distinction
m_{ij}
u_j^{rur} $u_{j'}^{rur}$ u_i^{rur} $u_{i'}^{rur}$ u_N^{rur}
Classifier
u_j^{iss} $u_{j'}^{iss}$ u_i^{iss} $u_{i'}^{iss}$ u_N^{iss}
Similarity
Pointer
Pointer
$u_{j'}^{pcd}$ u_j^{pcd} $u_{i'}^{pcd}$ u_i^{pcd} u_N^{pcd}
Non-linear Transformation + Layer Normalization
Non-linear Transformation + Layer Normalization
Non-linear Transformation + Layer Normalization
Output
$E_{[CLS]}$ $E_{U\_j'}$ $E_{[CLS]}$ E_{U_j} $E_{[CLS]}$ $E_{U\_i'}$ $E_{[CLS]}$ E_{U_i} $E_{[CLS]}$ E_{U_N} $E_{[SEP]}$
Pre-trained Language Model (BERT)

Interlocutor Structure Modeling

- **Pointer Consistency Distinction:** A consistent pointer representations and an inconsistent one sampled from this conversation are obtained. The similarities between every two pointers are calculated as

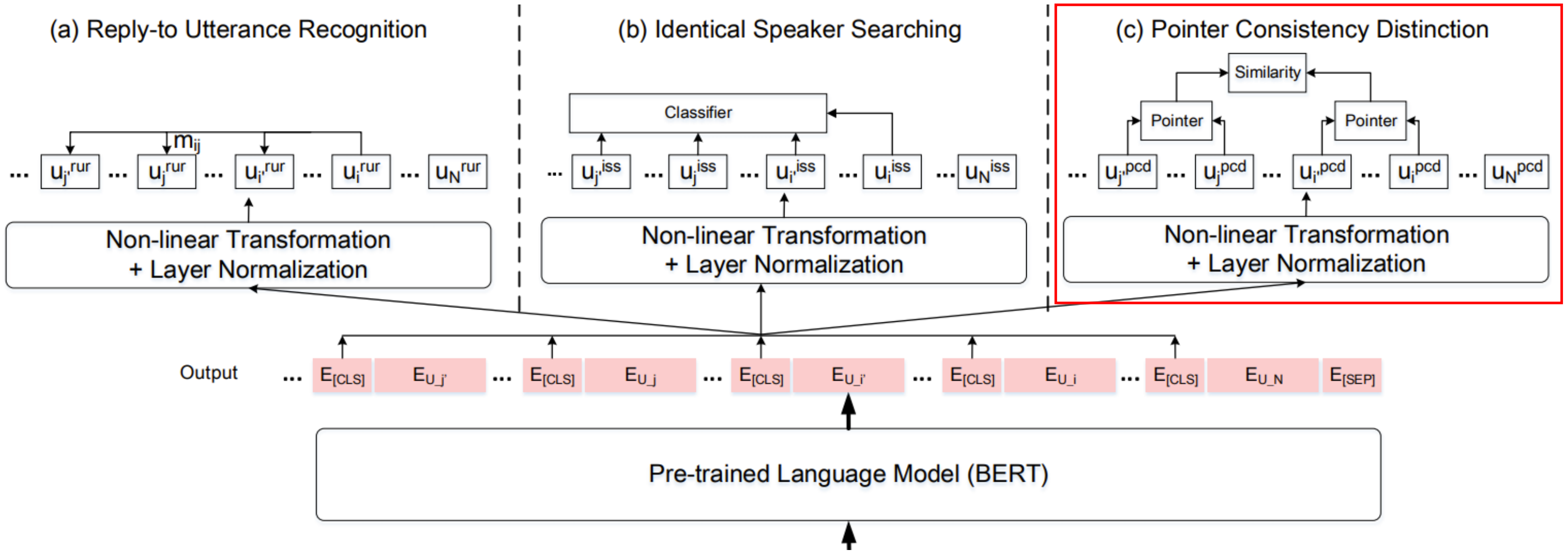
$$m_{ij} = \textbf{sigmoid}(\bar{\mathbf{p}}_{ii'}^{\top} \cdot \mathbf{A}^{pcd} \cdot \bar{\mathbf{p}}_{jj'})$$


(a) Reply-to Utterance Recognition
(b) Identical Speaker Searching
(c) Pointer Consistency Distinction
m_{ij}
$u_{j'}^{rur}$ u_{j}^{rur} $u_{i'}^{rur}$ u_{i}^{rur} u_{N}^{rur}
Classifier
$u_{j'}^{iss}$ u_{j}^{iss} $u_{i'}^{iss}$ u_{i}^{iss} u_{N}^{iss}
Similarity
Pointer
Pointer
$u_{j'}^{pcd}$ u_{j}^{pcd} $u_{i'}^{pcd}$ u_{i}^{pcd} u_{N}^{pcd}
Non-linear Transformation + Layer Normalization
Non-linear Transformation + Layer Normalization
Non-linear Transformation + Layer Normalization
Output
$E_{[CLS]}$ $E_{U\_j'}$ $E_{[CLS]}$ E_{U_j} $E_{[CLS]}$ $E_{U\_i'}$ $E_{[CLS]}$ E_{U_i} $E_{[CLS]}$ E_{U_N} $E_{[SEP]}$
Pre-trained Language Model (BERT)

Interlocutor Structure Modeling

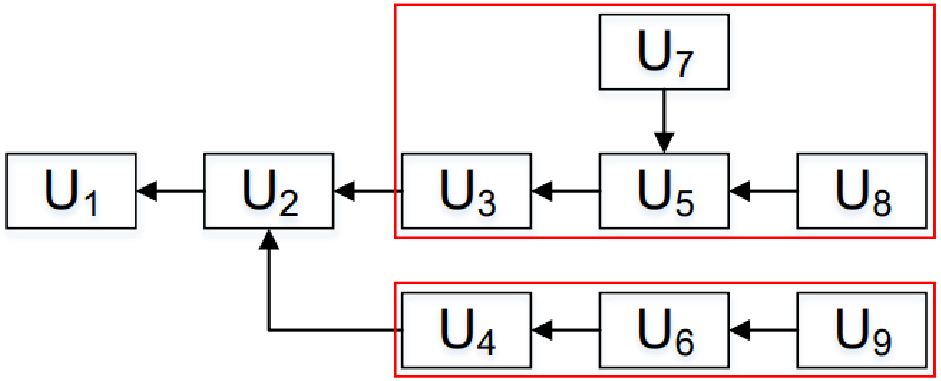
- **Pointer Consistency Distinction:** Minimize the hinge loss which enforces m_{ij} to be larger than m_{ik} by at least a margin Δ as

$$\mathcal{L}_{pcd} = \mathbf{max}\{0, \Delta - m_{ij} + m_{ik}\}$$


(a) Reply-to Utterance Recognition
m_{ij}
... $u_{j'}^{rur}$ ... u_j^{rur} ... $u_{i'}^{rur}$ ... u_i^{rur} ... u_N^{rur}
Non-linear Transformation + Layer Normalization
(b) Identical Speaker Searching
Classifier
... $u_{j'}^{iss}$ ... u_j^{iss} ... $u_{i'}^{iss}$ ... u_i^{iss} ... u_N^{iss}
Non-linear Transformation + Layer Normalization
(c) Pointer Consistency Distinction
Similarity
Pointer
Pointer
... $u_{j'}^{pcd}$ ... u_j^{pcd} ... $u_{i'}^{pcd}$ ... u_i^{pcd} ... u_N^{pcd}
Non-linear Transformation + Layer Normalization
Output
... $E_{[CLS]}$ $E_{U\_j'}$ ... $E_{[CLS]}$ E_{U_j} ... $E_{[CLS]}$ $E_{U\_i'}$ ... $E_{[CLS]}$ E_{U_i} ... $E_{[CLS]}$ E_{U_N} $E_{[SEP]}$
Pre-trained Language Model (BERT)

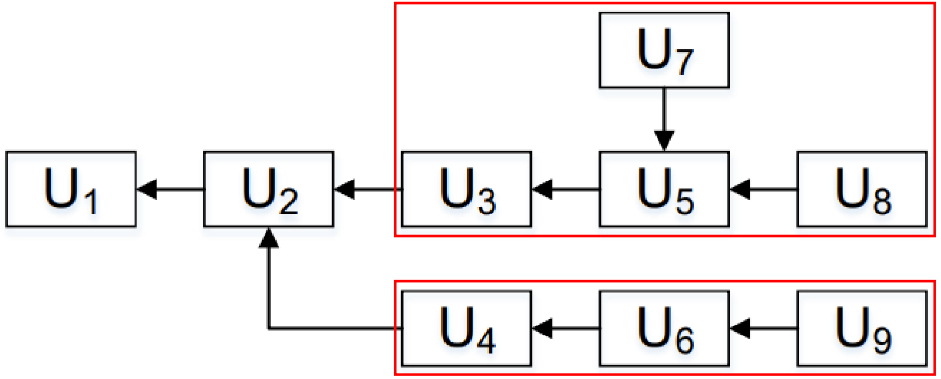
Utterance Semantics Modeling

- **Shared Node Detection:** A full MPC instance can be divided into several sub-conversations and we assume that the representations of sub-conversations under the same parent node tend to be similar.
- For example, two sub-conversations {U3, U5, U7, U8} and {U4, U6, U9} share the same parent node U2.



Utterance Semantics Modeling

- **Shared Node Detection:** Given a full MPC, the two sub-conversations under the top shared node (most utterances) form a positive pair empirically. Replace one sub-conversation with another one randomly sampled from the training corpus to form a negative pair.
- Sequence-pair prediction with the representation of the [CLS] token.
- Cross-entropy loss minimization.



Utterance Semantics Modeling

- **Masked Shared Utterance Restoration:** There are usually several utterances replying-to a shared utterance in MPC. A shared utterance is semantically relevant to more utterances in the context than non-shared ones.
- All tokens in a sampled shared utterance are masked with a [MASK] token and the model is enforced to restore the masked utterance given the rest conversation. (Utterance-level Language Model)

Multi-task Learning

- The tasks of masked language model (MLM) and next sentence prediction (NSP) in original BERT pre-training are also adopted, which have been proven effective for incorporating domain knowledge.
- MPCBERT is trained by performing multi-task learning that minimizes the sum of all loss functions as

$$\begin{aligned}\mathcal{L} = \mathcal{L}_{rur} + \mathcal{L}_{iss} + \mathcal{L}_{pcd} + \mathcal{L}_{msur} \\ + \mathcal{L}_{snd} + \mathcal{L}_{mlm} + \mathcal{L}_{nsp}\end{aligned}$$

Outline

- Introduction
- MPC-BERT
- **Downstream Tasks**
- Experiments
- Conclusion

Downstream Tasks

- To measure the effectiveness of these self-supervised tasks and to test the generalization ability of MPC-BERT, we evaluate MPC-BERT on three downstream tasks including addressee recognition, speaker identification and response selection, which are three core research issues of MPC.

Addressee Recognition

- In this paper, we follow the more challenging setting in Le et al. (2019) where addressees of all utterances in a conversation are asked to recognized.
- Given $\{(s_n, u_n, a_n)\}_{n=1}^{N} \setminus \{a_n\}_{n=1}^{N}$, models are asked to predict $\{\hat{a}_n\}_{n=1}^{N}$ where $\hat{a}_n$ is selected from the interlocutor set in this conversation.

*a, u, s and / denote addressee, utterance, speaker and exclusion respectively.

Speaker Identification

- This task aims to identify the speaker of the last utterance in a conversation, where the identified speaker is selected from the interlocutor set in this conversation.
- Given $\{(s_n, u_n, a_n)\}_{n=1}^{N} \setminus s_N$, models are asked to predict $\hat{s}_N$, where $\hat{s}_N$ is selected from the interlocutor set in this conversation.

Response Selection

- This tasks aims to measure the similarity between a context and a response, and then rank a set of response candidates, which is an important retrieval-based approach for chatbots.
- This task asks models to select $\hat{u}_N$ from a set of response candidates given the conversation context $\{(s_n, u_n, a_n)\}_{n=1}^{N} \setminus u_N$.

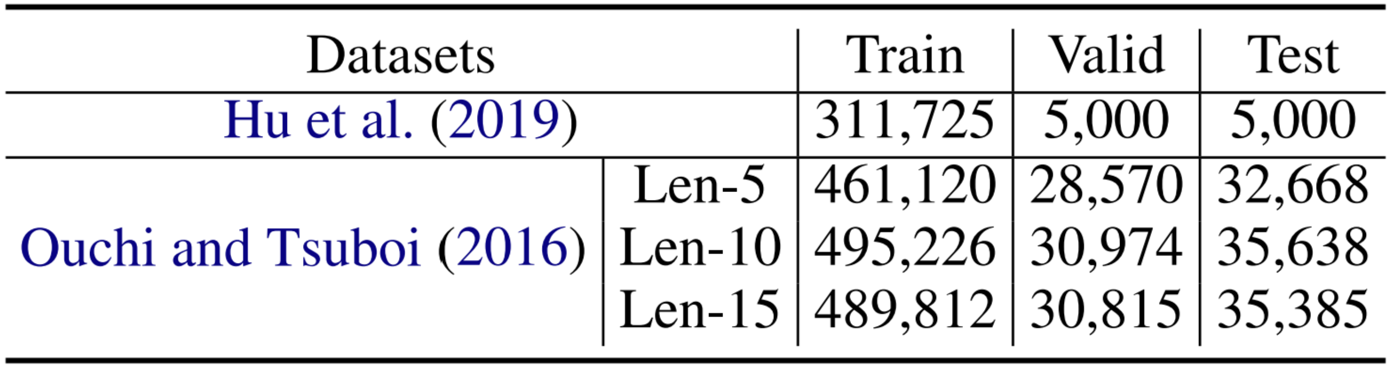
Outline

- Introduction
- MPC-BERT
- Downstream Tasks
- **Experiments**
- Conclusion

Experiments

- Datasets

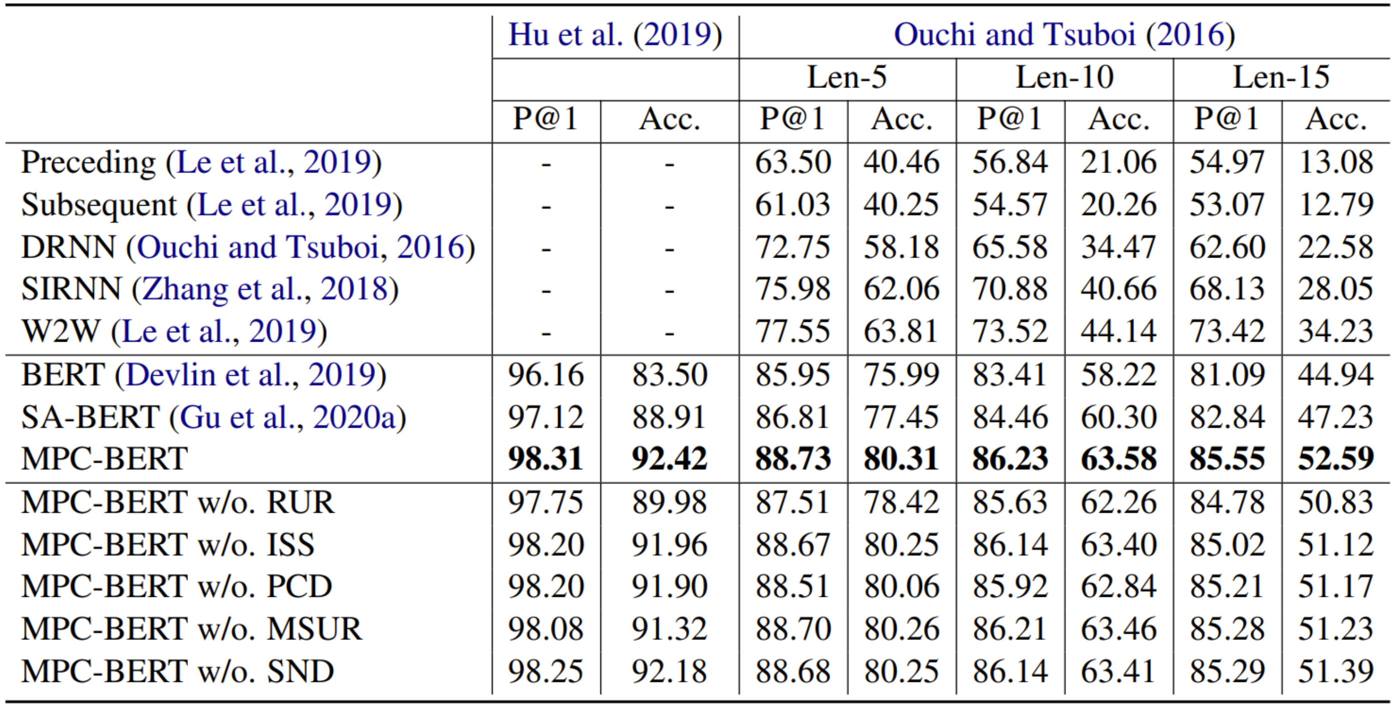
  We evaluated MPC-BERT on two Ubuntu IRC benchmarks.



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

Addressee Recognition

- Precision@1 (P@1) to evaluate each utterance with ground truth. Accuracy (Acc.) to evaluate a session if all addressees are recognized.

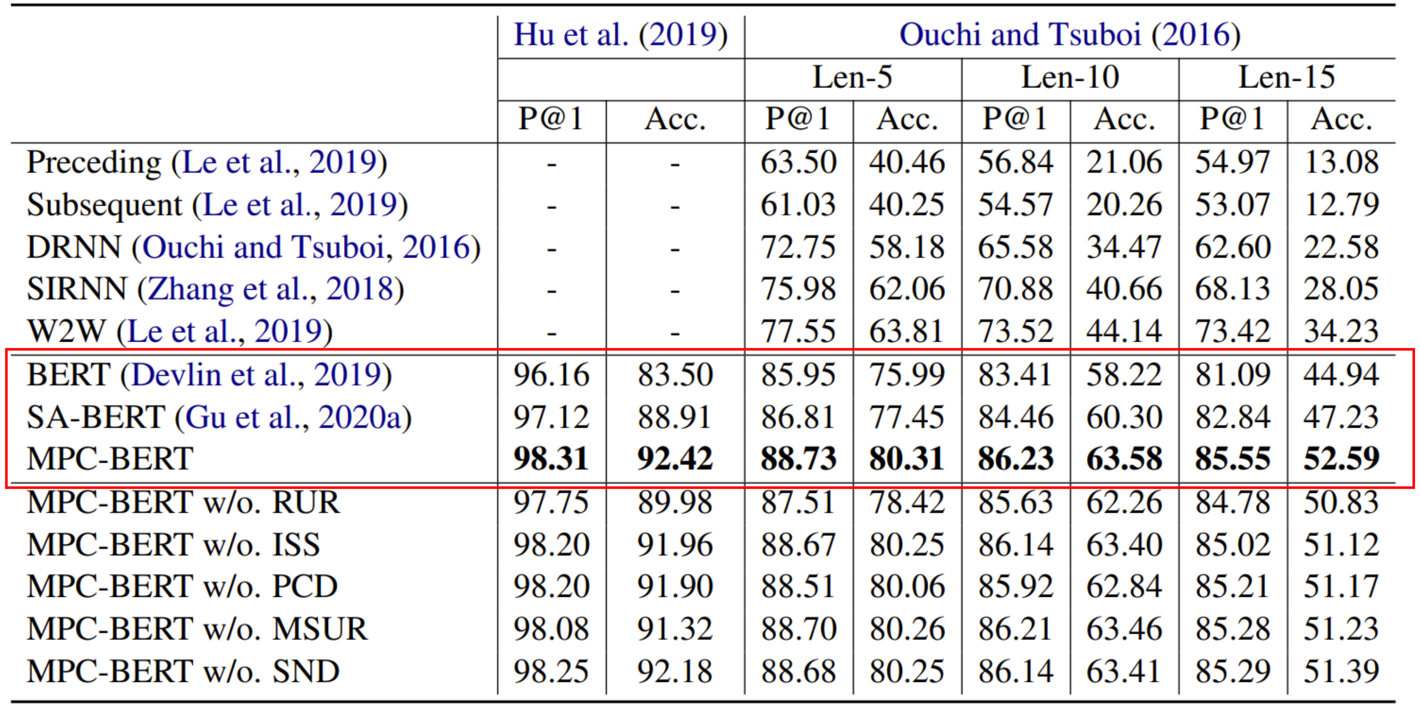


	Hu et al. (2019)		Ouchi and Tsuboi (2016)					
			Len-5		Len-10		Len-15	
	P@1	Acc.	P@1	Acc.	P@1	Acc.	P@1	Acc.
Preceding (Le et al., 2019)	-	-	63.50	40.46	56.84	21.06	54.97	13.08
Subsequent (Le et al., 2019)	-	-	61.03	40.25	54.57	20.26	53.07	12.79
DRNN (Ouchi and Tsuboi, 2016)	-	-	72.75	58.18	65.58	34.47	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).

Addressee Recognition

- MPC-BERT outperforms SA-BERT by margins of 3.51%, 2.86%, 3.28% and 5.36% on these test sets respectively in terms of Acc.

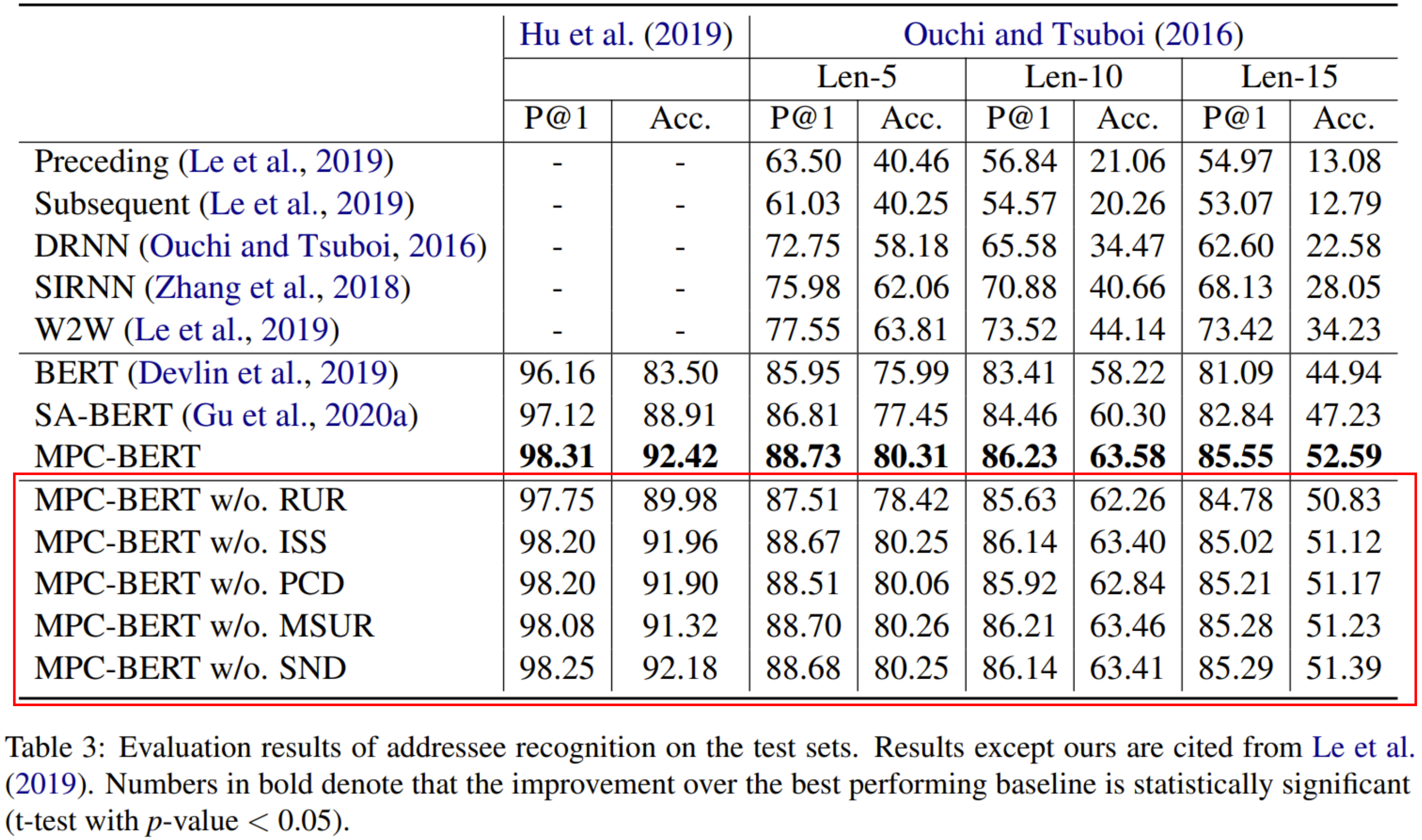


	Hu et al. (2019)		Ouchi and Tsuboi (2016)					
			Len-5		Len-10		Len-15	
	P@1	Acc.	P@1	Acc.	P@1	Acc.	P@1	Acc.
Preceding (Le et al., 2019)	-	-	63.50	40.46	56.84	21.06	54.97	13.08
Subsequent (Le et al., 2019)	-	-	61.03	40.25	54.57	20.26	53.07	12.79
DRNN (Ouchi and Tsuboi, 2016)	-	-	72.75	58.18	65.58	34.47	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).

Addressee Recognition

- RUR contributes the most, and the tasks modeling interlocutor structure contribute more than those for utterance semantics.

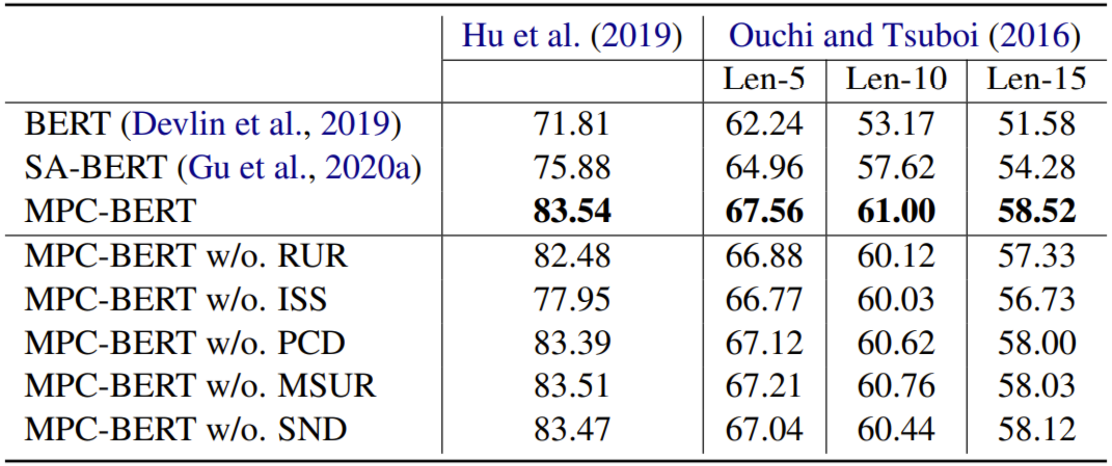


	Hu et al. (2019)		Ouchi and Tsuboi (2016)					
			Len-5		Len-10		Len-15	
	P@1	Acc.	P@1	Acc.	P@1	Acc.	P@1	Acc.
Preceding (Le et al., 2019)	-	-	63.50	40.46	56.84	21.06	54.97	13.08
Subsequent (Le et al., 2019)	-	-	61.03	40.25	54.57	20.26	53.07	12.79
DRNN (Ouchi and Tsuboi, 2016)	-	-	72.75	58.18	65.58	34.47	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).

Speaker Identification

- Precision@1 (P@1) to evaluate the last utterance of a conversation.
- MPC-BERT outperforms SA-BERT by margins of 7.66%, 2.60%, 3.38% and 4.24% respectively in terms of P@1.
- ISS and RUR contribute the most.

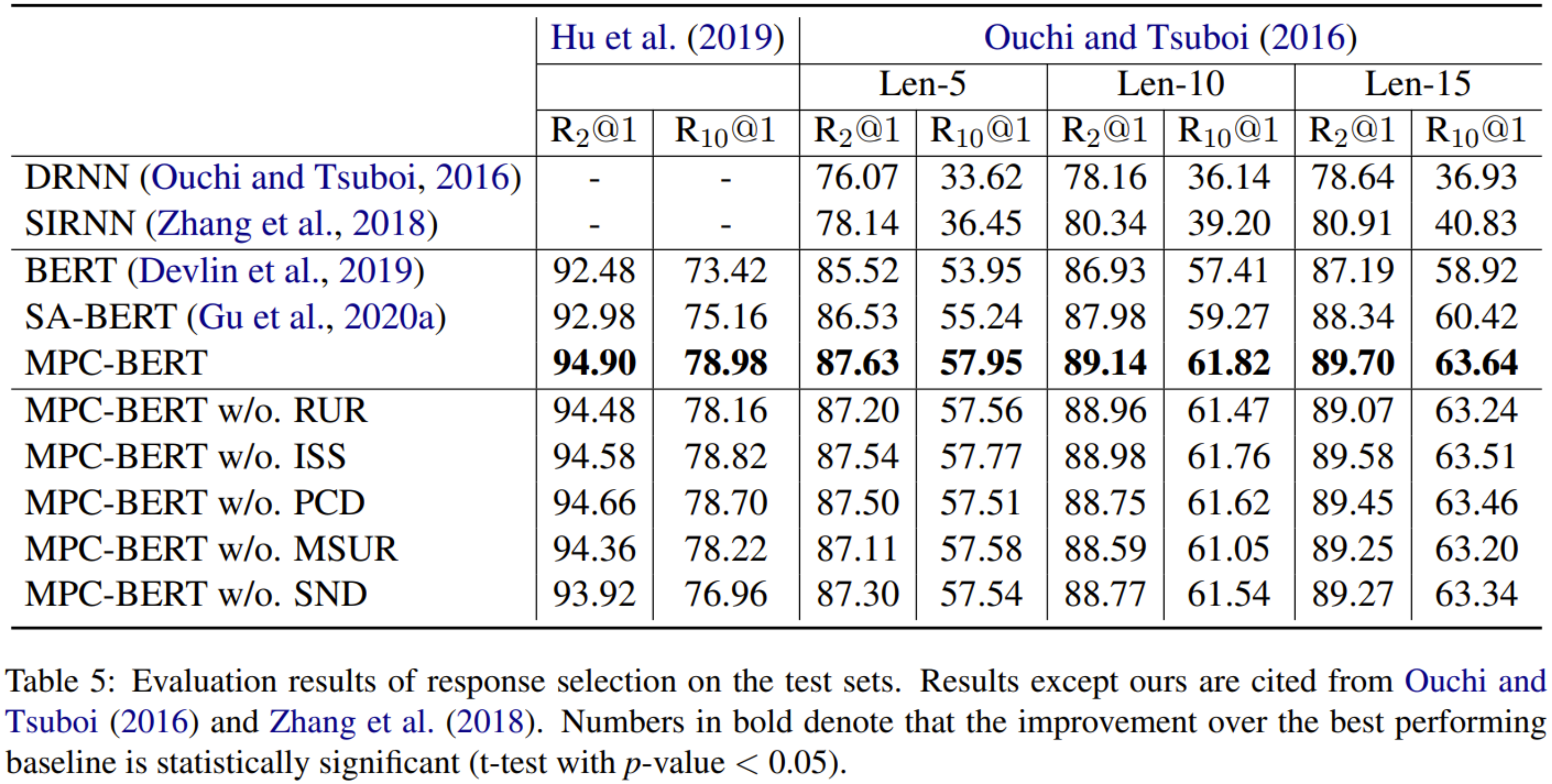


	Hu et al. (2019)	Ouchi and Tsuboi (2016)		
		Len-5	Len-10	Len-15
BERT (Devlin et al., 2019)	71.81	62.24	53.17	51.58
SA-BERT (Gu et al., 2020a)	75.88	64.96	57.62	54.28
MPC-BERT	**83.54**	**67.56**	**61.00**	**58.52**
MPC-BERT w/o. RUR	82.48	66.88	60.12	57.33
MPC-BERT w/o. ISS	77.95	66.77	60.03	56.73
MPC-BERT 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).

Response Selection

- $R_n@k$ to evaluate top-*k* selected responses from *n* available candidates. Two settings of $R_2@1$ and $R_{10}@1$ were followed.

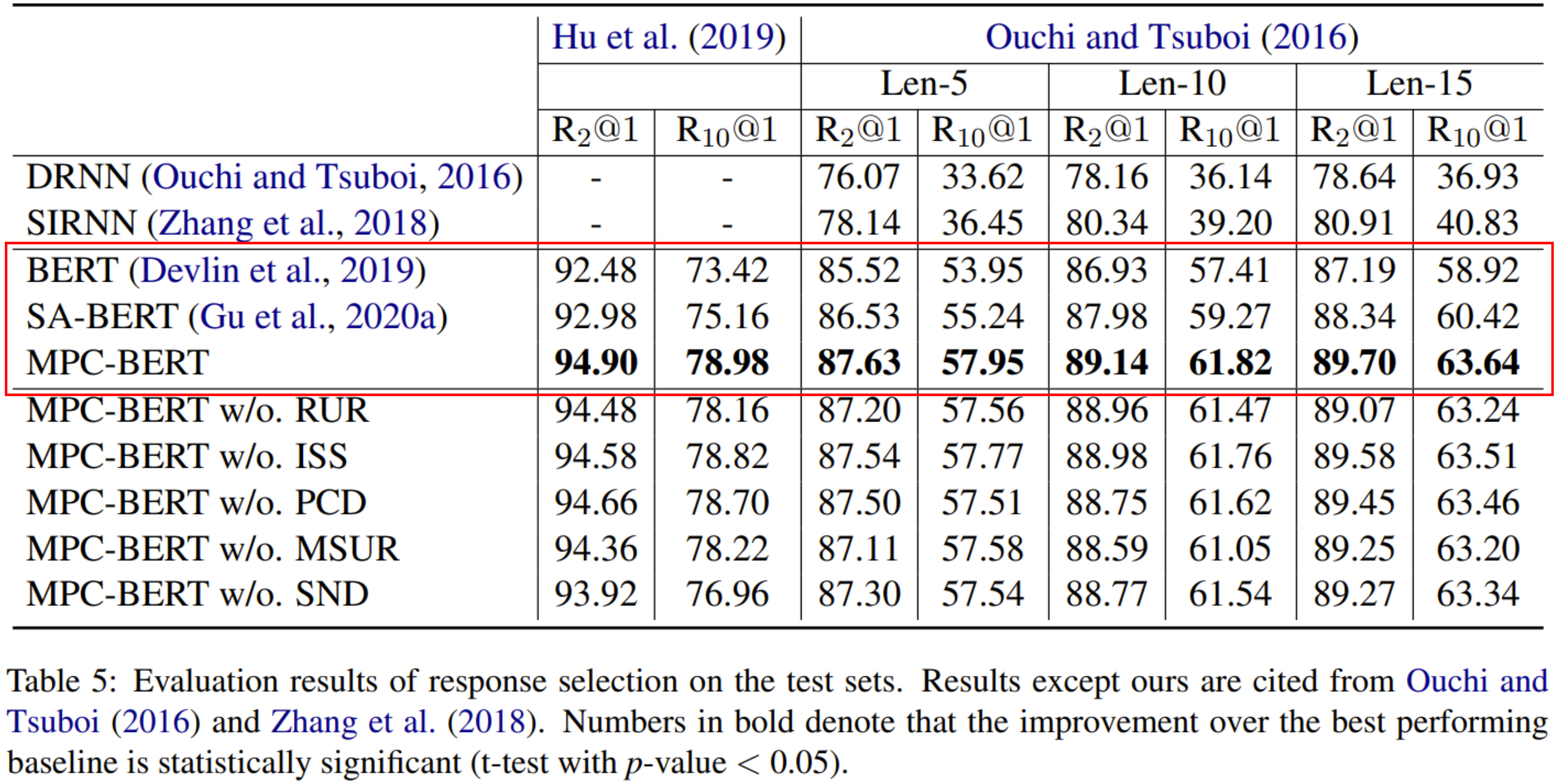


	Hu et al. (2019)		Ouchi and Tsuboi (2016)					
			Len-5		Len-10		Len-15	
	$R_2@1$	$R_{10}@1$	$R_2@1$	$R_{10}@1$	$R_2@1$	$R_{10}@1$	$R_2@1$	$R_{10}@1$
DRNN (Ouchi and Tsuboi, 2016)	-	-	76.07	33.62	78.16	36.14	78.64	36.93
SIRNN (Zhang et al., 2018)	-	-	78.14	36.45	80.34	39.20	80.91	40.83
BERT (Devlin et 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).

Response Selection

- MPC-BERT outperforms SA-BERT by margins of 3.82%, 2.71%, 2.55% and 3.22% respectively in terms of $R_{10}@1$.

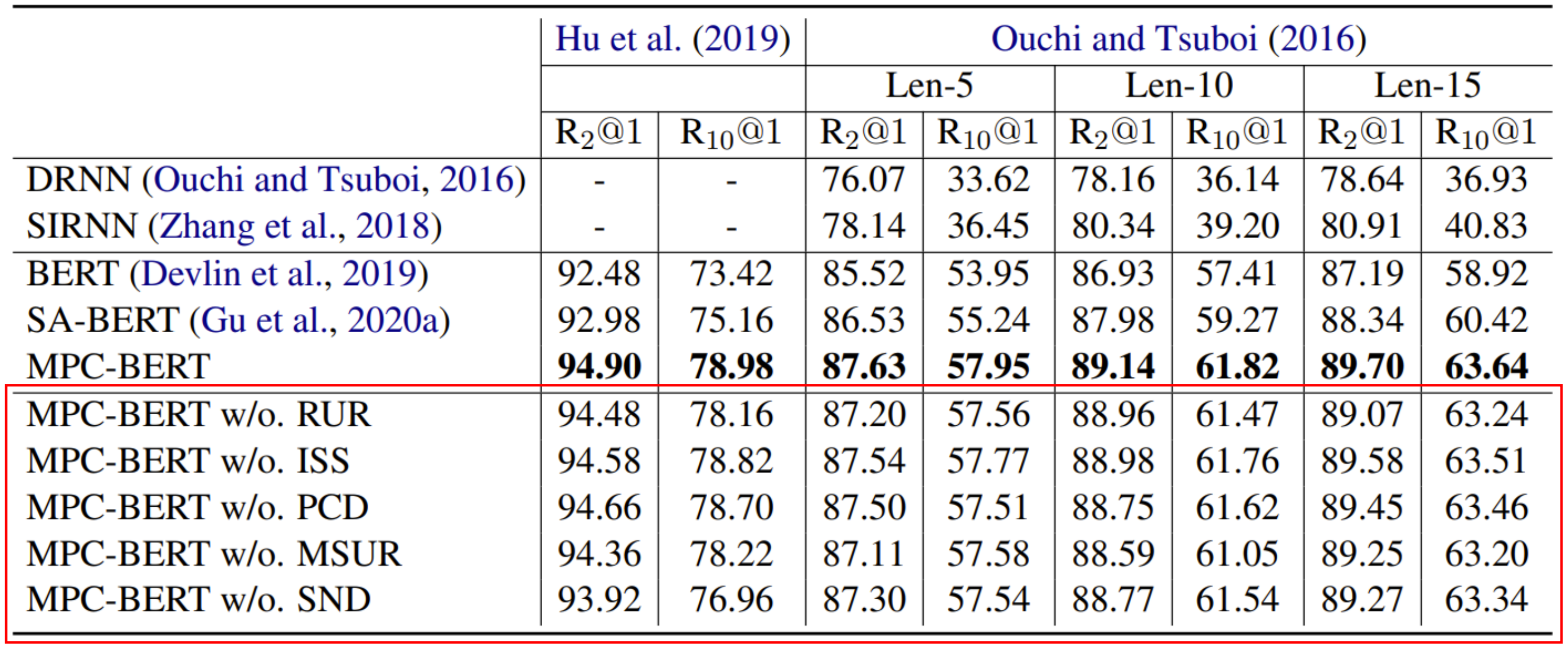


	Hu et al. (2019)		Ouchi and Tsuboi (2016)					
			Len-5		Len-10		Len-15	
	$R_2@1$	$R_{10}@1$	$R_2@1$	$R_{10}@1$	$R_2@1$	$R_{10}@1$	$R_2@1$	$R_{10}@1$
DRNN (Ouchi and Tsuboi, 2016)	-	-	76.07	33.62	78.16	36.14	78.64	36.93
SIRNN (Zhang et al., 2018)	-	-	78.14	36.45	80.34	39.20	80.91	40.83
BERT (Devlin et al., 2019)	92.48	73.42	85.52	53.95	86.93	57.41	87.19	58.92
SA-BERT (Gu et al., 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).

Response Selection

- SND contributes the most, and the two tasks modeling the utterance semantics contribute more than those for the interlocutor structures.

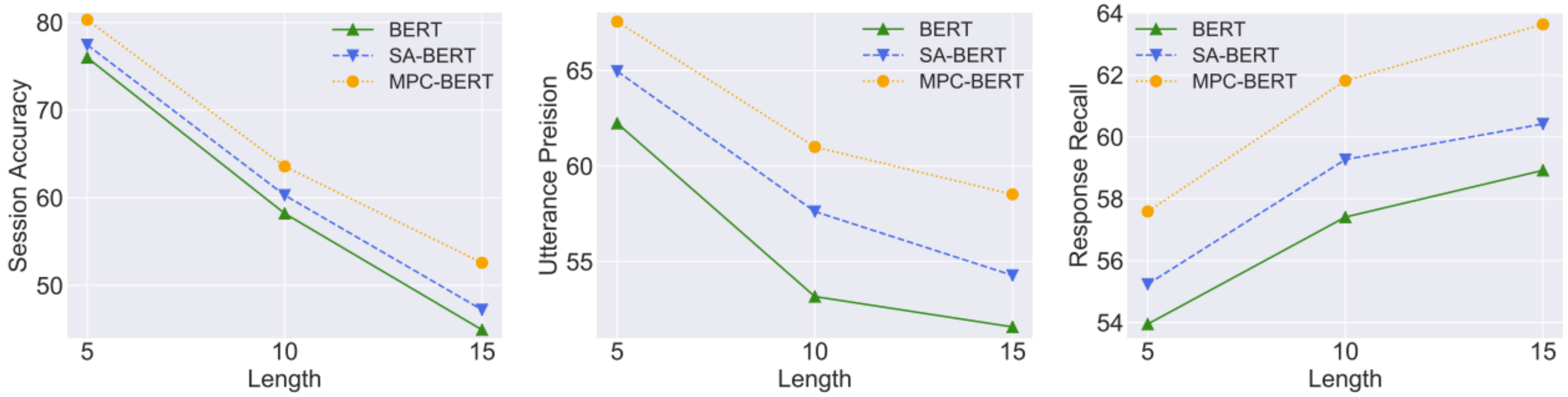


	Hu et al. (2019)		Ouchi and Tsuboi (2016)					
			Len-5		Len-10		Len-15	
	R_2@1	R_{10}@1	R_2@1	R_{10}@1	R_2@1	R_{10}@1	R_2@1	R_{10}@1
DRNN (Ouchi and Tsuboi, 2016)	-	-	76.07	33.62	78.16	36.14	78.64	36.93
SIRNN (Zhang et al., 2018)	-	-	78.14	36.45	80.34	39.20	80.91	40.83
BERT (Devlin et al., 2019)	92.48	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).

Discussions

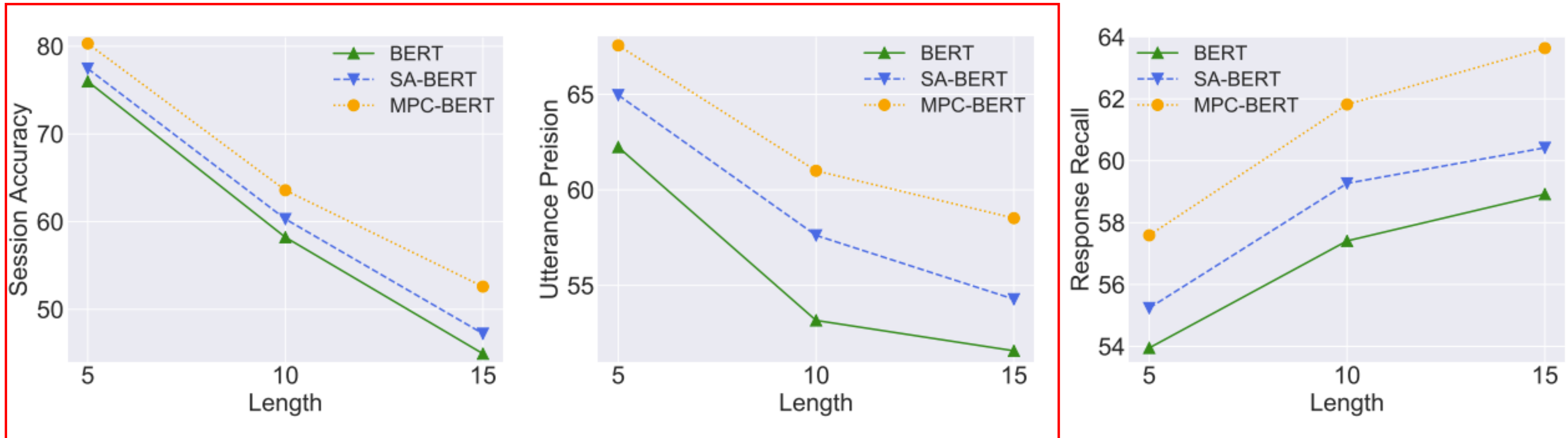
- How the performance of BERT, SA-BERT and MPC-BERT changed with respect to different session lengths on the test sets of Ouchi and Tsuboi (2016).



(a) Addressee recognition (b) Speaker identification (c) Response selection

Discussions

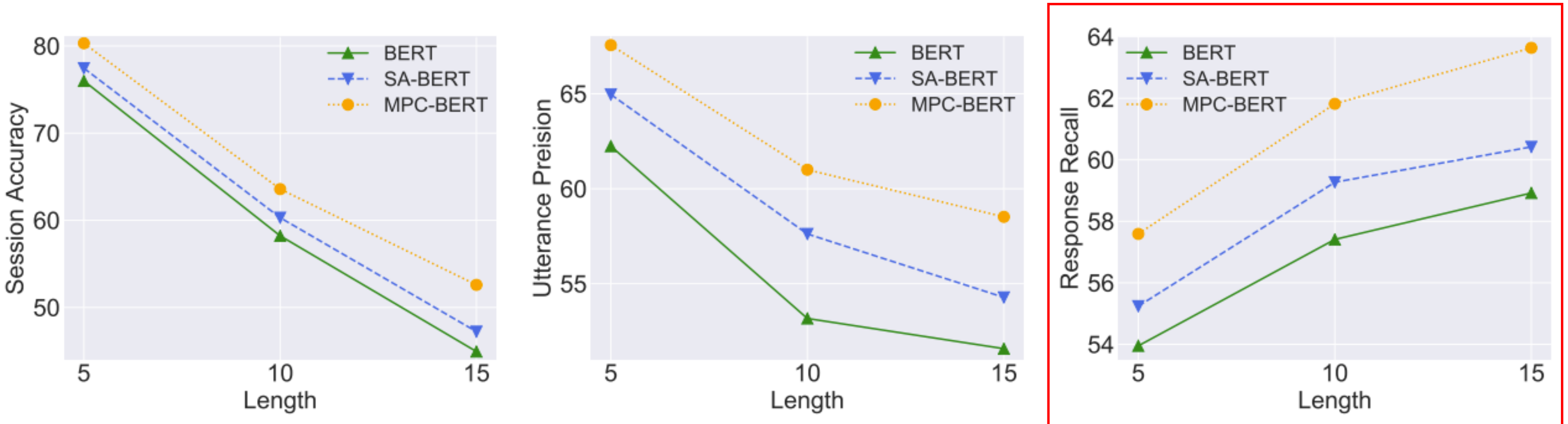
- The performance of addressee recognition and speaker identification dropped as the session length increased.
- The reason might be that longer sessions always contain more interlocutors which increase the difficulties of predicting interlocutors.



(a) Addressee recognition (b) Speaker identification (c) Response selection

Discussions

- The performance of response selection was significantly improved as the session length increased.
- It can be attributed to that longer sessions enrich the representations of contexts with more details which benefit response selection.



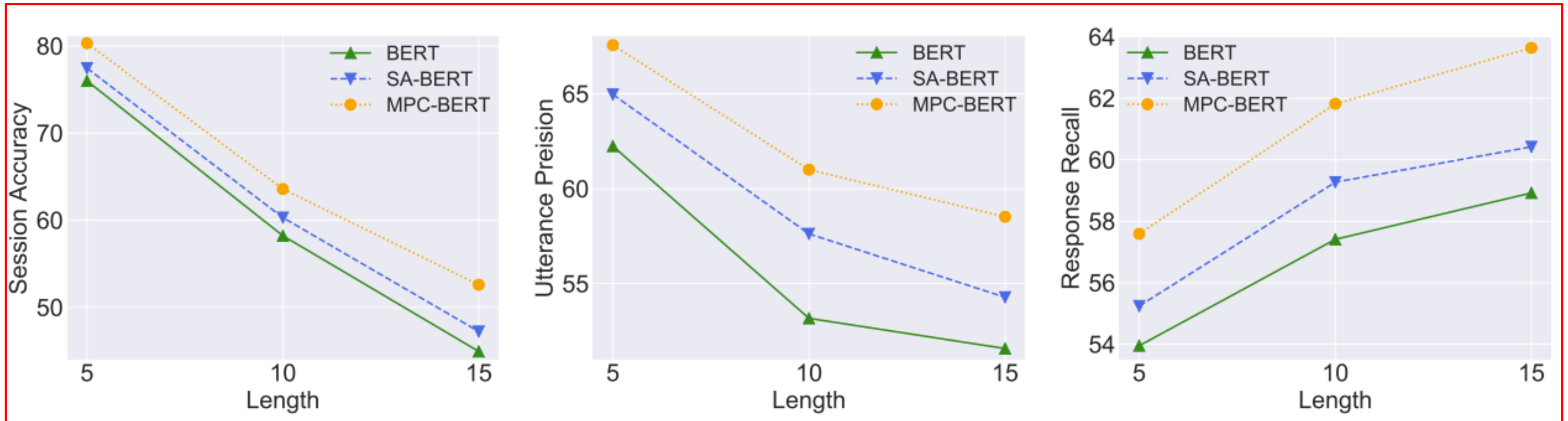
(a) Addressee recognition

(b) Speaker identification

(c) Response selection

Discussions

- As the session length increased, the performance of MPC-BERT dropped more slightly than that of SA-BERT on addressee recognition and speaker identification, and the R_{10}@1 gap between MPC-BERT and SA-BERT on response selection enlarged from 2.71% to 3.22%.
- Imply superiorities of MPC-BERT on modeling complicated structures.



(a) Addressee recognition

(b) Speaker identification

(c) Response selection

Outline

- Introduction
- MPC-BERT
- Downstream Tasks
- Experiments
- **Conclusion**

Conclusion

- We present MPC-BERT, a pre-trained language model with five self-supervised tasks for MPC understanding. These tasks jointly learn who says what to whom in MPCs.
- Experimental results on three downstream tasks show that MPC-BERT outperforms previous methods by large margins and achieves new state-of-the-art performance on two benchmarks.

Jia-Chen Gu

Chongyang Tao

Zhen-Hua Ling

Can Xu

Xiubo Geng

Daxin Jiang

Thanks! Q&A

Paper : https://arxiv.org/pdf/2106.01541.pdf

Code: https://github.com/JasonForJoy/MPC-BERT