達摩院
ALIBABA DAMO ACADEMY

Meeting Action Item Detection with Regularized Context Modeling

Jiaqing Liu, Chong Deng, Qinglin Zhang, Qian Chen, Wen Wang

Speech Lab of DAMO Academy, Alibaba Group

ICASSP 2023

Introduction

達摩院
ALIBABA DAMO ACADEMY

Online Meeting

- ◆ Technological advances & The pandemic
- ◆ More and more common for collaboration and information sharing

Meeting Transcripts

- ◆ ASR (Automatic Speech Recognition)
- ◆ The original record of every detail still needs to be further summarized

Meeting Minutes

- ◆ Human & Machine (extract or generate)
- ◆ Important information such as summaries, decisions, and **action items**

Introduction

达摩院 ALIBABA DAMO ACADEMY

Action Item

- Discussed in the meeting and assigned to participant(s)
- Expected to complete *within a short time window* after the meeting

[267] ***Speaker A***: OK, next time we meet, how about tomorrow?
[268] ***Speaker B***: Okay, we will continue talking about the project tomorrow.
[269] ***Speaker A***: Okay, we'll tentatively schedule at 3 pm, see you tomorrow.

An example of action item. We show the ***Speaker*** and [sentence id], mark the action item.

Action Item Detection

- Sentence-level binary classification task
- Detect sentences containing actionable tasks in meeting transcripts

Reference: Gruenstein, A., Niekrasz, J. and Purver, M., 2005. Meeting structure annotation: Data and tools. In 6th SIGdial Workshop on Discourse and Dialogue.

Data

Action Item Dataset

- Corpus: Far from adequate to evaluate advanced deep learning models
- Annotation: High subjectivity of the action item (ICSI Meeting Kappa=0.36)

Public Meeting Corpora

- **AMI**: 101 annotated AMI meetings with 381 action items (indirect)
- **ICSI**: 75 meetings without publicly available action item annotations

AliMeeting-Action Corpus (AMC-A)

- Corpus: Chinese meeting corpus of **424** meetings
- Annotation: **manual** action item annotations

AMC-A Dataset Link: https://www.modelscope.cn/datasets/modelscope/Alimeeting4MUG/summary

Data

達摩院
ALIBABA DAMO ACADEMY

AMC-A

- **Meeting**: 15-30 minute discussion by 2-4 participants covering certain topics from a diverse set, biased towards work meetings in various industries
- **Annotation**: Each sentence is annotated by three annotators independently following detailed annotation guidelines with sufficient examples

	AMC-A (ours)				AMI
	All	Train	Dev	Test	
Total # Meetings	**424**	295	65	64	101
Total # Utterances	**306,846**	213,235	45,869	47,742	80,298
Total # Action	**1506**	1014	222	270	381
Kappa Coefficient	0.47	0.46	0.49	0.50	/
Avg. # Action per Meeting	3.55	3.44	3.42	4.22	3.77
Std. # Action per Meeting	3.97	3.98	3.35	4.41	1.95

AMC-A Dataset: https://www.modelscope.cn/datasets/modelscope/Alimeeting4MUG/summary

Method

達摩院
ALIBABA DAMO ACADEMY

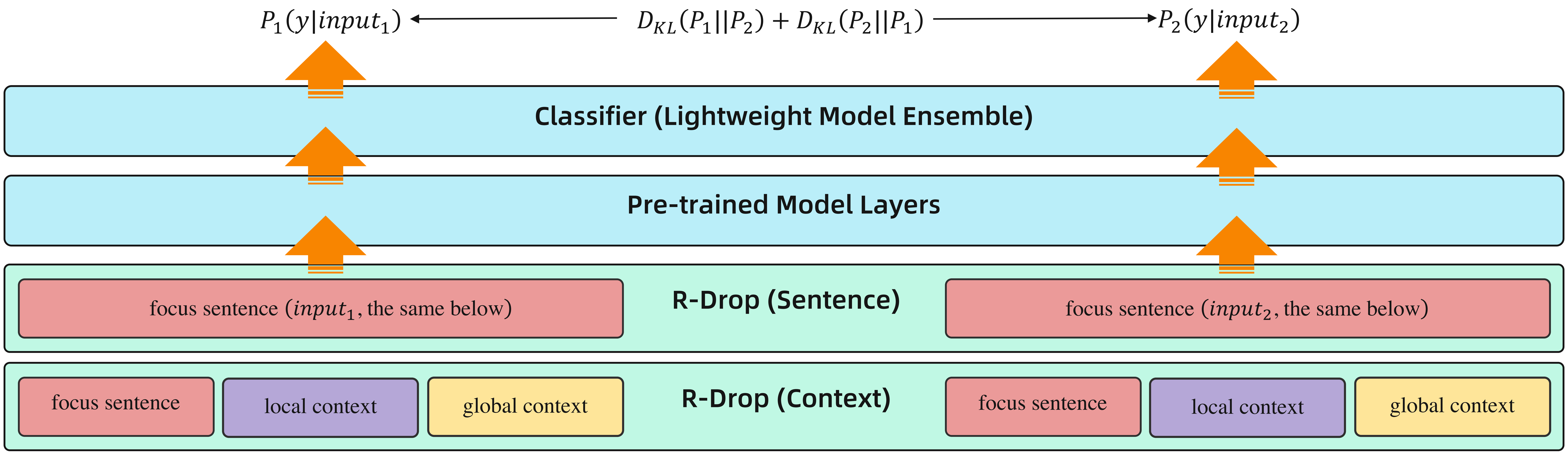
[001] ***Speaker A***: Hello everyone, welcome to the weekly meeting.
[002] ***Speaker A***: Firstly, let's look at *this tourist area development project*.
[003] ***Speaker A***: Tim, could you please tell us about the tourism area? ...
[035] ***Speaker B***: There are some issues with *our tourism development project*.
[036] ***Speaker B***: The positioning of the tourist area is still unclear. ...
[267] ***Speaker A***: OK, next time we meet, how about **tomorrow**?
[268] ***Speaker B***: Okay, we will continue talking about the *project* **tomorrow**.
[269] ***Speaker A***: Okay, we'll tentatively schedule at **3 pm**, see you **tomorrow**.

An example of action item. We show the ***Speaker*** and [sentence id], mark the action item.
The local context provides the **timeframe**. And the global context provides the *task description*.

Context

- Local Context: Adjacent sentences (explored by prior works)
- Global Context: Relevant but non-contiguous sentences (retrieved through context selection method by computing the similarities)

Method

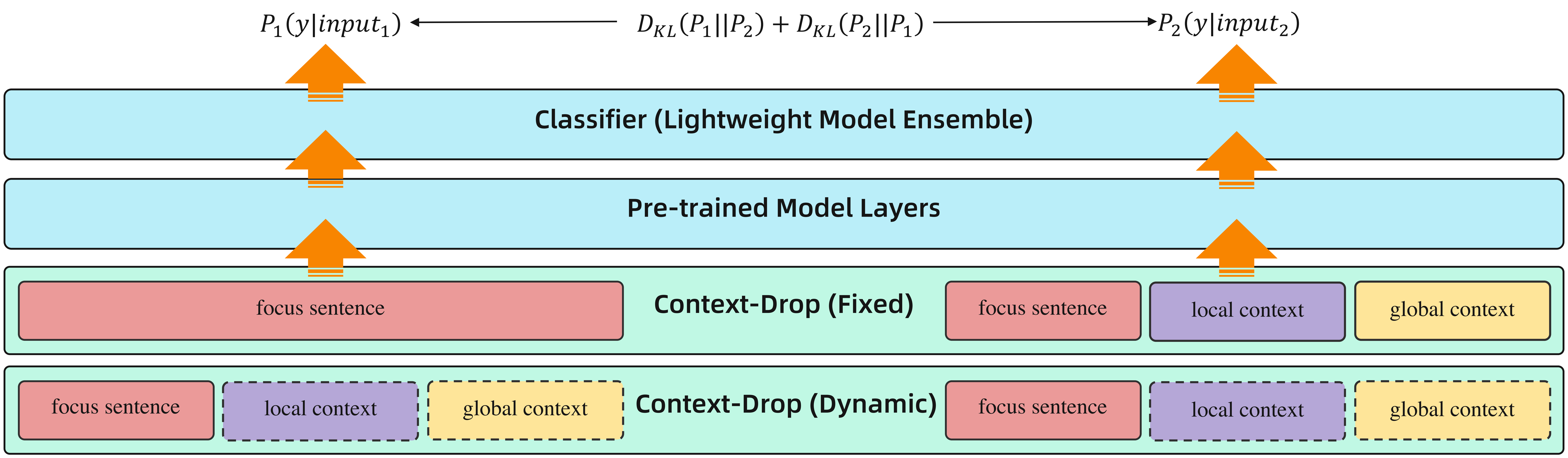


R-Drop

- Context understanding plays a critical role in the action item detection task
- However, local/global contexts may contain irrelevant information (may distract the classifier)

Our Code: https://github.com/alibaba-damo-academy/SpokenNLP/tree/main/action-item-detection

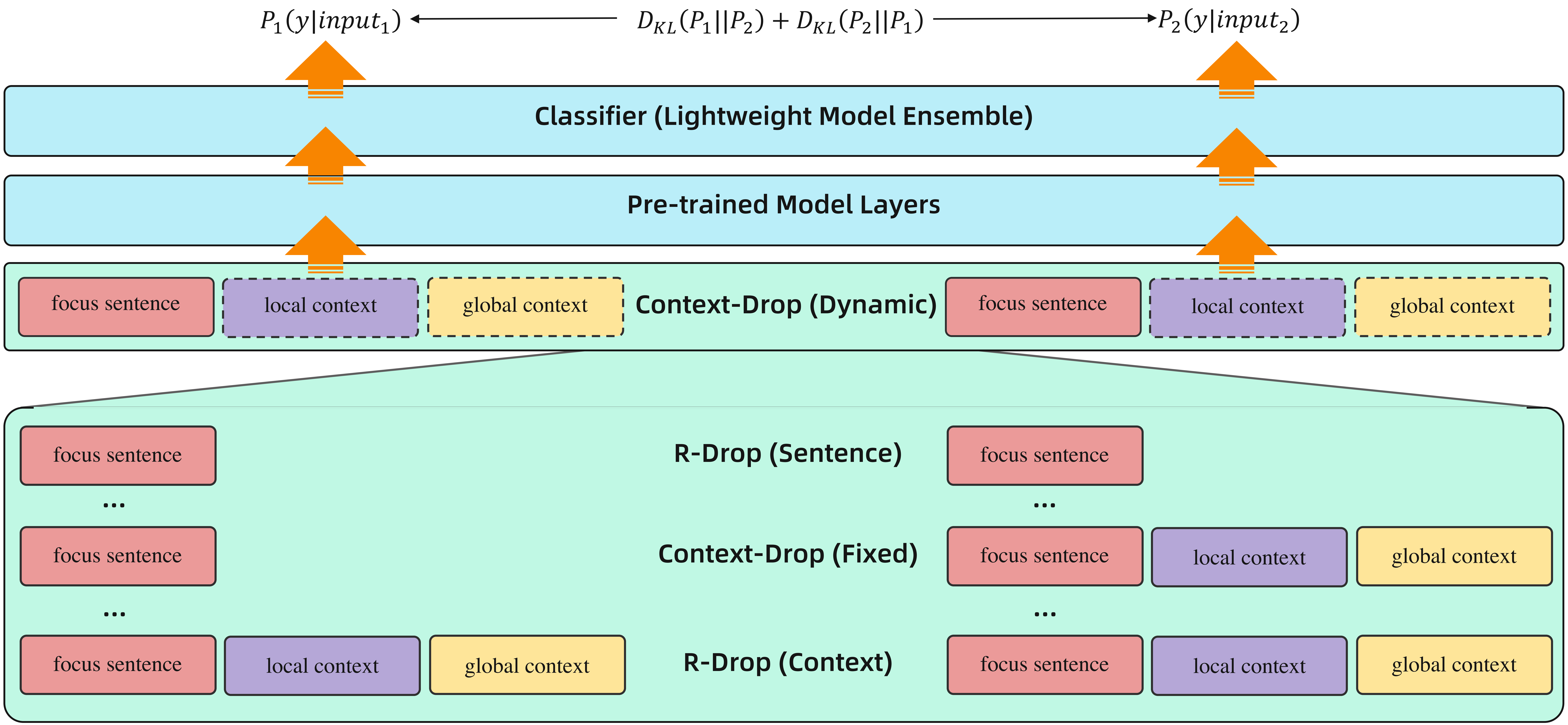
Method



Context-Drop

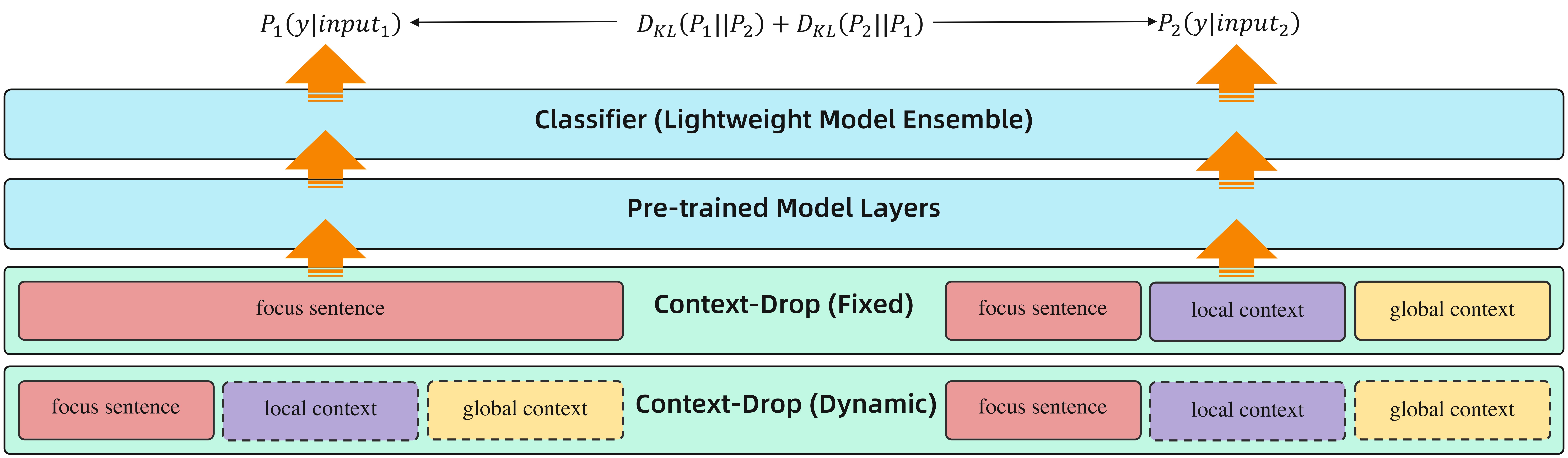
- **Motivation**: Focus more on the current sentence ➔ Better exploit relevant information in context & Be less distracted by irrelevant information in context
- **Context modeling with regularization**: Force the prediction probability distributions of a single sentence and the sentence with its context to be consistent with each other

Method



$P_1(y|input_1)$
$D_{KL}(P_1||P_2) + D_{KL}(P_2||P_1)$
$P_2(y|input_2)$
Classifier (Lightweight Model Ensemble)
Pre-trained Model Layers
focus sentence
local context
global context
Context-Drop (Dynamic)
focus sentence
local context
global context
focus sentence
R-Drop (Sentence)
focus sentence
...
...
focus sentence
Context-Drop (Fixed)
focus sentence
local context
global context
...
...
focus sentence
local context
global context
R-Drop (Context)
focus sentence
local context
global context

Method



Lightweight Model Ensemble

- **Motivation**: During annotation, the majority voting results are usually correct despite the relatively low inter-annotator agreement ➔ Explore model ensemble while preserving inference latency
- **Method**: Initialize encoder layers from one pre-trained model A and initialize pooler layer from another pre-trained model B to integrate knowledge from different pre-trained models

Experiments

Model	Modeling Task	AMC-A F1
BERT	sentence classification	64.76±0.98
Longformer	sequence labeling	65.35±1.33
StructBERT	sentence classification	**67.84**±1.20

StructBERT

- The word structural pretraining objective of StructBERT reconstructs tokens in the correct order from the shuffled trigrams
- This pre-training objective could improve its robustness to disordered sentences, which is quite common in spoken languages

StructBERT Model Link: https://modelscope.cn/models/damo/nlp_structbert_backbone_base_std

Experiments

達摩院
ALIBABA DAMO ACADEMY

Input Method	AMC-A F1	AMI F1
sentence	67.84±1.20	38.67±1.25
w/ R-Drop	68.77±0.82	39.26±1.70
sentence + local context	68.50±1.21	41.03±1.42
w/ R-Drop	68.79±0.42	42.72±0.74
w/ Context-Drop (fixed)	69.15±0.91	**43.12**±0.74
w/o KL loss	68.23±1.11	40.71±1.78
w/ Context-Drop (dynamic)	69.53±0.75	42.05±0.31
w/o KL loss	67.97±0.53	41.44±2.29

Input Method	AMC-A F1	AMI F1
sentence + global context	67.99±1.86	35.82±1.11
w/ R-Drop	69.80±1.14	37.88±1.04
w/ Context-Drop (fixed)	69.07±0.57	39.23±0.73
w/ Context-Drop (dynamic)	70.48±0.63	41.25±1.76
sentence + local & global context	69.09±1.23	41.31±1.51
w/ R-Drop	68.72±1.04	40.75±1.28
w/ Context-Drop (fixed)	69.28±0.95	38.66±0.77
w/ Context-Drop (dynamic)	**70.82**±1.33	41.50±1.52

Global Context

- sentence + local & global context performs better than sentence + local context on both Chinese AMC-A and English AMI meeting corpora
- Global context provides complementary information to local context and **combination of global & local context achieves further improvement**

Experiments

達摩院
ALIBABA DAMO ACADEMY

Input Method	AMC-A F1	AMI F1
sentence	67.84±1.20	38.67±1.25
w/ R-Drop	68.77±0.82	39.26±1.70
sentence + local context	68.50±1.21	41.03±1.42
w/ R-Drop	68.79±0.42	<u>42.72</u>±0.74
w/ Context-Drop (fixed)	69.15±0.91	**43.12**±0.74
w/o KL loss	68.23±1.11	40.71±1.78
w/ Context-Drop (dynamic)	69.53±0.75	42.05±0.31
w/o KL loss	67.97±0.53	41.44±2.29

Input Method	AMC-A F1	AMI F1
sentence + global context	67.99±1.86	35.82±1.11
w/ R-Drop	69.80±1.14	37.88±1.04
w/ Context-Drop (fixed)	69.07±0.57	39.23±0.73
w/ Context-Drop (dynamic)	<u>70.48</u>±0.63	41.25±1.76
sentence + local & global context	69.09±1.23	41.31±1.51
w/ R-Drop	68.72±1.04	40.75±1.28
w/ Context-Drop (fixed)	69.28±0.95	38.66±0.77
w/ Context-Drop (dynamic)	**70.82**±1.33	41.50±1.52

Context-Drop

- Context-Drop performs better than baseline on both AMC-A and AMI corpora
- Context-Drop: Focus more on the current sentence & Exploit relevant information in context & Be less distracted by irrelevant information in context
- Reduction in the standard deviations ➔ Improvement of model stability and robustness

Experiments

Input Method	AMC-A F1	AMI F1
sentence	67.84±1.20	38.67±1.25
w/ R-Drop	68.77±0.82	39.26±1.70
sentence + local context	68.50±1.21	41.03±1.42
w/ R-Drop	68.79±0.42	42.72±0.74
w/ Context-Drop (fixed)	69.15±0.91	**43.12**±0.74
w/o KL loss	68.23±1.11	40.71±1.78
w/ Context-Drop (dynamic)	69.53±0.75	42.05±0.31
w/o KL loss	67.97±0.53	41.44±2.29

Input Method	AMC-A F1	AMI F1
sentence + global context	67.99±1.86	35.82±1.11
w/ R-Drop	69.80±1.14	37.88±1.04
w/ Context-Drop (fixed)	69.07±0.57	39.23±0.73
w/ Context-Drop (dynamic)	70.48±0.63	41.25±1.76
sentence + local & global context	69.09±1.23	41.31±1.51
w/ R-Drop	68.72±1.04	40.75±1.28
w/ Context-Drop (fixed)	69.28±0.95	38.66±0.77
w/ Context-Drop (dynamic)	**70.82**±1.33	41.50±1.52

Context-Drop & Ablation Analysis

- Context-Drop (dynamic) performs best in most cases ➔ This flexible and dynamic contrastive learning method can achieve better performance
- Ablation analysis: w/o regularization loss of KL divergence (KL loss) degrades the performance ➔ **Contrastive learning is important for the gains**

Experiments

Model Layers	Pooler Layer	AMC-A F1
StructBERT	StructBERT	67.84±1.20
	RoBERTa	**68.36**±0.93
RoBERTa	RoBERTa	66.87±0.44
	StructBERT	**67.25**±0.93

Lightweight Model Ensemble

- Lightweight Model Ensemble (initializing from different pre-trained models) performs better than initializing from one pre-trained model
- This method could integrate knowledge from different models and achieve better performance without increasing the number of parameters

達摩院
ALIBABA DAMO ACADEMY

Download AMC-A

Our Code

Conclusion and Future Work

Conclusion

- **AMC-A**: The first Chinese meeting corpus with action item annotations ➔ Alleviate the scarcity of resources and prompt research on meeting action item detection
- **Context-Drop**: Improve context modeling of both local and global contexts with regularization ➔ Achieve improvement in accuracy and robustness of action item detection for both Chinese and English meeting corpora
- **Lightweight Model Ensemble**: Integrate knowledge from different pre-trained models ➔ Achieve improvement in accuracy while preserving inference latency

Future Work

- Refine Lightweight Model Ensemble and investigate its efficacy on other tasks
- Combine the Context-Drop and Lightweight Model Ensemble methods

達摩院
ALIBABA DAMO ACADEMY

Thanks

Q & A