icassp 2022 Singapore
May 10, 2022
ICASSP
Singapore &
Shenzhen, China

之江实验室 ZHEJIANG LAB

Hierarchical and Multi-View Dependency Modeling Network for Conversational Emotion Recognition

◆ Introduction

1. Background

➢The modeling of conversational context plays an important role for emotion recognition in conversations

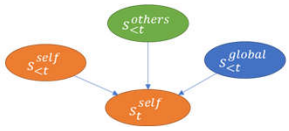
➢The structured information in conversational context are generally complicated, especially for those with multi-turn and multi-speaker

Hierarchical

token/word → utterance → context

The speaker dependency are reflected in two aspects, i.e., the token-level and the utterance-level

Multi-View



2. Our work

➢propose a hierarchical and multi-view dependency modelling network (HMVDM)

➢the HMVDM model has a hierarchical structure with two main modules :

1)token-level dependency modeling module (TDM)

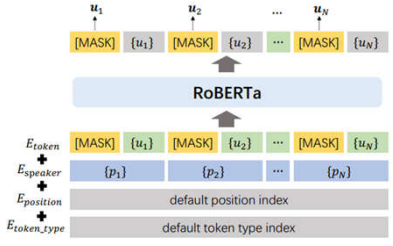
aims to learn the long-range token-level dependency between different utterances

2) utterance-level dependency modelling module (UDM)

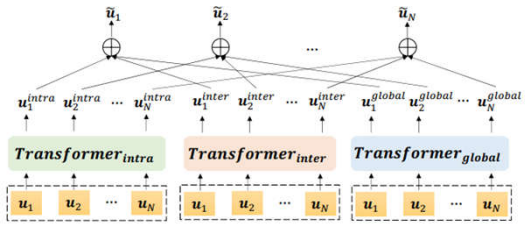
aims to learn the utterance-level dependency from intra-, inter-, and global-speaker(s) view

◆ Methods

1. Token-level dependency modeling



2. Utterance-level dependency modeling



◆ Experiments

1. Datasets

	DD	MELD	IEMOCAP	EmoryNLP
#Dial.	13,118	1,432	151	897
train	11,118	1,038	100	713
dev	1,000	114	20	99
test	1,000	280	31	85
#Utt.	102,979	13,708	7,433	12,606
train	87,170	9,989	4,810	9,934
dev	8,069	1,109	1,000	1,344
test	7,740	2,610	1,623	1,328

2. Overall performance

• Compared with the RNN-based and Graph-based baseline models on four datasets

Models	DailyDialog		MELD		IEMOCAP		EmoryNLP	
	Macro F1	Micro F1	W-Avg. F1	Micro F1	W-Avg. F1	Micro F1	W-Avg. F1	Micro F1
DialogueRNN	-	-	57.03	-	62.75	-	-	-
+ RoBERTa	-	57.32	63.61	-	64.76	-	37.44	-
COSMIC	51.05	58.48	65.21	-	65.28	-	38.11	-
DialogueGCN	49.95	53.71	58.37	56.17	60.85	60.63	34.29	33.13
DialogXL	-	54.93	62.41	-	65.94	-	34.73	-
RGAT	-	54.31	60.91	-	65.22	-	34.42	-
KET	-	53.48	58.18	-	59.56	-	34.39	-
DAG-ERC*	-	59.33	63.65	-	**68.03**	-	39.02	-
TODKAT*	52.56	58.47	**68.23**	64.75	61.33	61.11	**43.12**	42.68
w/o KB	50.03	53.44	63.97	61.11	58.96	57.38	33.79	32.62
HMVDM	**53.48**	**68.42**	65.92	**66.31**	67.96	**67.88**	38.46	**42.91**

3. Ablation studies

• The ablation studies for the TDM module in HMVDM with considering the speaker embeddings and long-range token dependency

	W-Avg. F1	Micro F1
original	67.96	67.88
w/o speaker embedding	64.02	64.43
w/o long-range token dep.	62.00	61.71

• The ablation studies for the UDM module in HMVDM with considering the utterance-level dependency from intra-, inter-, and global-speaker view respectively

	W-Avg. F1	Micro F1
original	67.96	67.88
w/o intra-speaker dep.	64.81	64.73
w/o inter-speaker dep.	64.78	64.67
w/o global-speaker dep.	62.49	62.58

◆ Conclusion

• The long-range token-level dependency modeling is important for the ERC task

• Modeling the utterance-level dependency from intra-, inter-, and global-speaker(s) simultaneously is helpful