EmotionFlow: Capture the Dialogue Level Emotion Transitions

IEEE ICASSP 2022

Xiaohui Song, Liangjun Zang, Rong Zhang, Songlin Hu, Longtao Huang

Institute of Information Engineering
Chinese Academy of Sciences
Alibaba Group

May 7, 2022

Motivation

What is ERC task

An example:

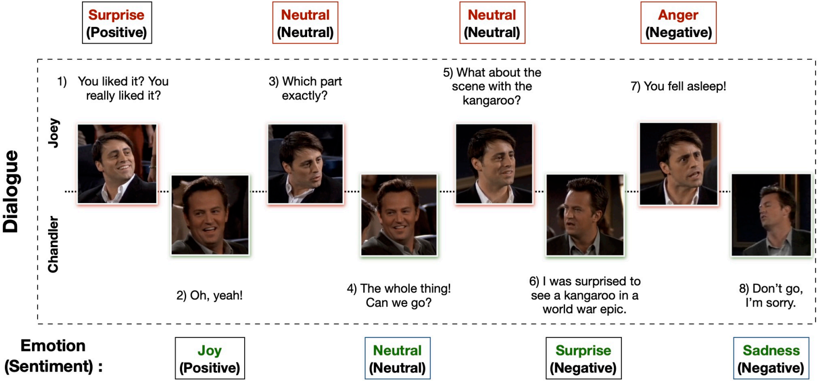

Figure: Example of ERC task.[1]

[1] https://github.com/declare-lab/awesome-emotion-recognition-in-conversations

What is the focus of this paper

Problem settings

1. **single modal**
   text only
2. **real-time**
   only use the past utterances to predict the emotion of current utterance
3. **multi-party**
   a conversation can contains more than 2 speakers
4. **non-anonymous**
   speakers' habits can be learned

Focus

The spread effect of emotions at dialogue level.

Next Emotion

Current Emotion						
Anger	0.49	0.04	0.05	0.17	0.11	0.15
Disgust	0.22	0.25	0.05	0.11	0.16	0.21
Fear	0.23	0.03	0.2	0.23	0.13	0.19
Joy	0.12	0.04	0.04	0.55	0.08	0.17
Sadness	0.16	0.05	0.04	0.16	0.37	0.22
Surprise	0.19	0.05	0.05	0.28	0.13	0.3

Figure: The transition probability between emotions of current turn and next turn.

ICASSP 2022 SINGAPORE

Model

How we design our model

Semantic Context Modeling

1. Roberta as encoder
2. QA-style input construction to capture speaker-specific features
   $X_t = [\texttt{<s>}, s_{t-k}, u_{t-k}, s_{t-k+1}, \ldots, s_t, u_t, \texttt{</s>}, Q]$
   Q = "How does s_t feel now?"
3. supervised signal training on CLS token
   ensure that the outputs of this module are "probability" distributions.

Emotion Sequence Modeling

A linear-chain CRF layer upon the outputs of the Semantic Context Modeling module.

22-27 MAY 2022
ICASSP 2022
SINGAPORE

Model Arch

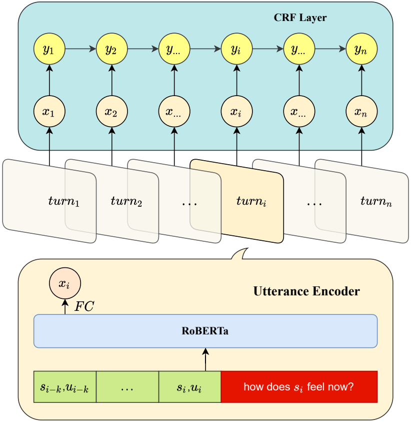

Figure: The overview of EmotionFlow.

- full dialogue as input
- each turn of dialogue is encoded separately then calculate the probability x_i
- feed $x_1 \ldots x_n$ into CRF layer.

Experiments

Datasets

Multimodal EmotionLines Dataset (MELD)

	Train	Dev	Test	total
Conversations	1038	114	280	1432
Utterances	9989	1109	2610	13708
Speakers	260	47	100	274
Speakers >100	6	6	6	6

Table: Statistics of MELD.

There are only 6 speakers that appeared more than 100 times in the dataset, which is good for the model to learn speakers' features.

Main Results

Model	External Knowledge	Weighted F1
DialogueGCN	✗	58.10
RGAT	✗	60.91
HiTrans	✗	61.94
DialogXL	✗	62.41
DAG-ERC	✗	63.65
TODKAT w/o KB	✗	63.97
EmotionFlow(Ours)	✗	**65.05**
KAIMTL	✓	58.97
KET	✓	58.18
COSMIC	✓	65.21
TODKAT	✓	**65.47***

Figure: Performance comparisons on the MELD testset.

22-27 MAY 2022
ICASSP 2022
SINGAPORE

Ablation Study

Model	Weighted F1
EmotionFlow	65.05
EmotionFlow w/o CRF	63.70
EmotionFlow w/o QA	63.55
EmotionFlow w/o [CRF,QA]	62.35

Figure: Ablation study on MELD test set.

Conclusion and Future Work

ICASSP 2022
SINGAPORE
22-27 MAY 2022

Conclusion and Future Work

Conclusion

1. A novel model that can capture the spread effect of emotions via a CRF layer
2. QA-style input construction helps the model to learn speaker-specific features
3. A new state-of-the-art result on a widely used benchmark.

Future Work

Linear-chain CRF $\rightarrow$ probabilistic graphical model

Thanks for your attention!