Emotion Recognition in Conversations

Junyin Chen Hanshu Ding

Zoe Fang Yifan Jiang

Roadmap

- ❑ Task description
- ❑ System architecture
- ❑ Core approach
- ❑ Issues and successes
- ❑ Improvements
- ❑ Related reading

Task description

- Primary task
  - Detect Emotion based on English dialogue inputs
- Dataset: MELD (Multimodal EmotionLines Dataset)
  - 13000 utterances from 1433 dialogues from TV-series Friends
  - Multiple speakers
    - 6 main characters and others
  - 7 emotions
    - Joy, surprise, sadness, angry, disgust, fear, neutral

neutral	0.471519
joy	0.174492
surprise	0.120633
anger	0.111022
sadness	0.068375
disgust	0.027130
fear	0.026830

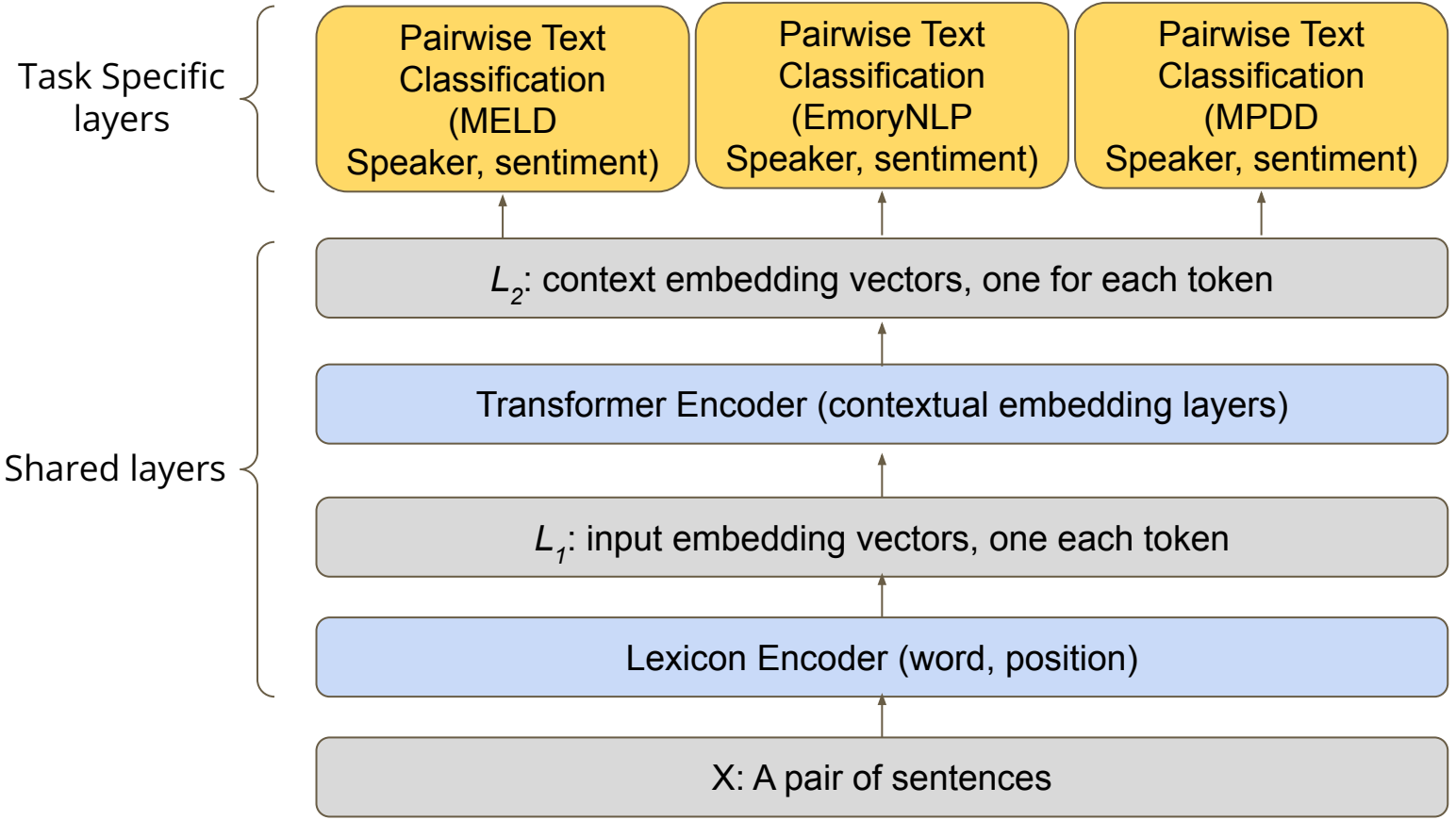
No.,Utterance,Speaker,Emotion,Sentiment,Dialogue_ID,Utterance_ID
1,"Oh my God, he’s lost it. He’s totally lost it.",Phoebe,sadness,negative,0,0
2,What?,Monica,surprise,negative,0,1

Task description

- Adaptation task
  - Detect Emotion based on Chinese dialogue inputs
- Dataset: MPDD (Multi-party Dialogue Dataset)
  - 25548 dialogues from 4142 dialogues from TV-series
    - Manual random split
  - Multiple speakers
  - 7 emotions
    - Joy, surprise, sadness, angry, disgust, fear, neutral

No.,Utterance,Speaker,Emotion,Dialogue_ID,Utterance_ID
22377,這個，放在這存一下，我們上那邊買衣服，很快回來，行嗎？ ,亞琳,joy,3313,0
22378,謝謝老闆！,亞琳,joy,3313,1

System Architecture



Core Approach

- Multi-Task Deep Neural Network
  - Shared encoder: **RoBERTa**
  - Task Heads:
    - Main Task: ERC on MPDD
    - Auxiliary Task
      - Speaker Classification
      - Data Augmentation

- Input Preprocessing
  - Speaker information
    - “<speaker name>: “ + utterance
  - Utterance Context
    - past/future utterances

Issues and Successes - D2

Adding context works

D2 Result →

Past utterance	Future utterance	Weighted F1
0	0	60.17
0	6	61.96
6	0	62.46

Accuracy by emotions →

Emotion	Correct count	Total count	Accuracy
neutral	1026	1256	0.8169
joy	247	402	0.6144
anger	156	345	0.4522
surprise	173	281	0.6157
sadness	66	208	0.3173
disgust	5	68	0.0735
fear	1	50	0.02

Issue and Success – D2

<s>Monica:Is that too much to ask after six year?!Monica:I mean, all I’m asking for is just a little emotio! Chandler:And you’re upset because you didn’t make your best friend cry?</s></s>I mean what?</s>

Predict: surprise, Rachel

Truth: anger , Monica

<s>Rachel:Oh, that sounds great.Others:How does that sound?Others:Well, I’ve got a project for you that’s a lot more related to fashion.Others:Well, don’t think I haven’t noticed your potential.Rachel:Oh, you got me.Others:Eh.</s></s>Come on over here, sweetheart.</s>

Predict: neutral, Others

Truth: neutral, Others

D3 Improvements

1. Data Augmentation: EmoryNLP dataset
   - Emotion Detection on TV Show Transcripts with Sequence-based Convolutional Neural Networks
   - 12606 utterances from 897 dialogues from TV-series Friends
   - Multiple speakers
     - 6 main characters and others
   - 7 emotions (Willcox (1982)'s feeling wheels)
     - neutral, joyful, peaceful, sad, mad, scared, powerful
     - (neutral, joy, surprise, sadness, anger, fear, disgust)

Neutral	0.305416
Joyful	0.219851
Scared	0.129354
Mad	0.108315
Peaceful	0.090598
Powerful	0.078921
Sad	0.067546

```
Utterance,Speaker,Emotion,Dialogue_ID,Utterance_ID
Coffee.,Rachel,Neutral,1,0
Thank you.,Joey,Neutral,1,1
Cappuccino.,Rachel,Neutral,1,2
```

D3 Improvements

2. Hyperparameter tuning

- No. of Utterance: 6 -> 10
- Speaker: T -> F
- No. of Epoch: 5 -> 6
- Batch size: 8 -> 16
- Training task: speaker + emotion -> only emotion prediction

```
kwargs = {
    "seed": 42,
    "data_dir": "data/",
    "train_dir": "outputs/multi_task_model",
    "model_file": "outputs/pytorch_model.bin",
    "model_id": "1U6Ek3c75RjxypFAj7_B-yfQ9NyDNk-eS",
    "num_past_utterances": 10,
    "num_future_utterances": 0,
    "speaker_in_context": False,
    "epoch": 6,
    "learning_rate": 1e-5,
    "batch_size": 16,
    "do_train": False,
    "checkpoint": "roberta-base",
    "training": ["Emotion"],
    "evaluation": "Emotion",
    'output_file': 'outputs/predictions.out',
    "result_file": "results/scores.out",
}
```

Issue and Success – D3 with speaker

<s>Joey:You know, I think I was sixteen. Monica:Please, just a little bit off the back. Phoebe:I'm still on "no." </s></s>Uh, morning. Do you guys think you could close your eyes for just a sec?</s>

Predict: neutral, Monica

Truth: neutral, Rachel

<s>Rachel:It's not a purse! It's a shoulder bag. Joey:It looks like a women's purse. Rachel:No Joey, look. Trust me, all the men are wearing them in the spring catalog. Look. See look, Joey:See look, </s></s>Exactly! Unisex!</s>

Predict: joy, Rachel

Truth: neutral, Rachel

Issue and Success – D3 without speaker

<s>You know, I think I was sixteen. Please, just a little bit off the back. I'm still on "no." </s></s>Uh, morning. Do you guys think you could close your eyes for just a sec?</s>

Predict: neutral

Truth: neutral

<s>It's not a purse! It's a shoulder bag. Joey:It looks like a women's purse. No Joey, look. Trust me, all the men are wearing them in the spring catalog. Look. See look, See look, </s></s>Exactly! Unisex!</s>

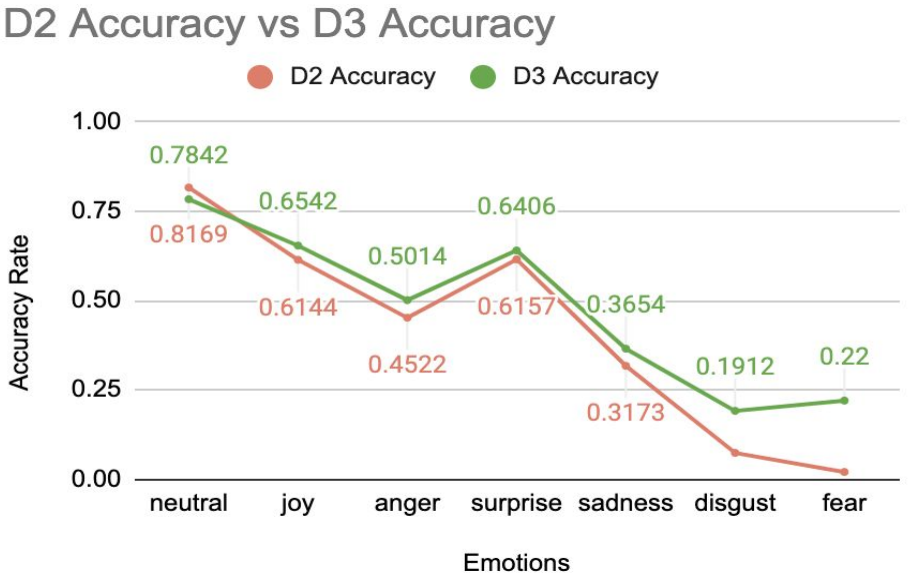
Predict: joy

Truth: neutral

Issue and Success – Comparison

Weighted-F1: 62.46 -> 64.74
Leaderboard: about No.10

Neutral to non-neutral: ⬆
Non-neutral to neutral: ⬇



Related readings

- Yi-Ting Chen, Hen-Hsen Huang, and Hsin-Hsi Chen. 2020. MPDD: A multi-party dialogue dataset for analysis of emotions and interpersonal relationships. n *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 610–614, Marseille, France. European Language Resources Association.

- Chao-Chun Hsu, Sheng-Yeh Chen, Chuan-Chun Kuo, Ting-Hao Huang, and Lun-Wei Ku. 2018. Emotion-Lines: An emotion corpus of multi-party conversa-ions. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

- Taewoon Kim and Piek Vossen. 2021. EmoBERTa: Speaker-Aware Emotion Recognition in Conversation with RoBERTa. *arXiv e-prints, page arXiv:2108.12009.*

- Xiaodong Liu, Pengcheng He, Weizhu Chen, and Jianfeng Gao. 2019. Multi-task deep neural networks for natural language understanding. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 4487–4496, Florence, Italy. Association for Computational Linguistics.*

- Soujanya Poria, Devamanyu Hazarika, Navonil Majumder, Gautam Naik, Erik Cambria, and Rada Mihalcea. 2019. MELD: A multimodal multi-party dataset for emotion recognition in conversations. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 527–536, Florence, Italy. Association for Computational Linguistics.*
- Sayyed M. Zahiri and Jinho D. Choi. 2017. Emotion detection on TV show transcripts with sequence-based convolutional neural networks. *arXiv e-prints, page arXiv:1708.04299*

Thank you!