Exploiting Annotators' Typed Description of Emotion Perception to Maximize Utilization of Ratings for Speech Emotion Recognition

Huang-Cheng Chou\textsuperscript{1,2}, Wei-Cheng Lin\textsuperscript{1}, Chi-Chun Lee\textsuperscript{2}, Carlos Busso\textsuperscript{1}

\textsuperscript{1} Multimodal Signal Processing (MSP) lab, Department of Electrical and Computer Engineering, University of Texas at Dallas (UTD), USA

\textsuperscript{2} Behavioral Informatics & Interaction Computation (BIIC) lab, Department of Electrical Engineering, National Tsing Hua University (NTHU), Taiwan
Emotion Perception

Emotional stimulus  Emotion perception  Emotion decoding  Annotation

Speech Emotion Recognition

Different emotional experiences!
Example of Annotation in the MSP-Podcast corpus

Annotations for primary emotion (single-choice):

- **MSP-PODCAST_0004_0073.wav**
  - Rater 1: Neutral
  - Rater 2: Neutral
  - Rater 3: Happy
  - Rater 4: Other (accusatory)
  - Rater 5: Other (Pleased)

Disagreement: Neutral

Discarded

Never used

Consensus label: Neutral

---

Where Typed Words Come From

Is any of these emotions the primary emotion in the audio? If not, select Other and specify the emotion.

- Angry
- Sad
- Happy
- Surprise
- Fear
- Disgust
- Contempt
- Neutral
- Other

**Single-choice**
(d) Primary emotion

Please pick all the emotional classes that you perceived in the audio (Include the primary emotions selected in previous question)

- Angry
- Sad
- Happy
- Amused
- Neutral

- Frustreated
- Depressed
- Surprise
- Concerned

- Disgust
- Disappointed
- Excited
- Confused

- Annoyed
- Fear
- Contempt
- Other

**Multi-choice**
(e) Secondary emotion

Primary emotion example:
MSP-PODCAST_0004_0073.wav
1. W0002117; Other (Pleased)
2. W0000060; Neutral
3. W0003012; Other (accusatory)
4. W0002999; Neutral
5. W0003011; Happy

Example: Four-class SER task
Neutral (N), Angry (A), Sad (S), Happy (H)

SER

Single-label Task

Sad
Sad
Sad
Sad

Angry
Angry

For training

Multi-label Task

Sad, Angry
Sad, Angry

Emotion Coexistence
Ambiguity
Human-like Emotion Perception

N A S H (0.0, 0.0, 1.0, 0.0)
1.0
0.5
0.0

N A S H (0.0, 0.4, 0.6, 0.0)
1.0
0.5
0.0

N A S H (0.0, 1.0, 1.0, 0.0)
1.0
0.5
0.0

N A S H (0.0, 0.4, 0.6, 0.0)
1.0
0.5
0.0

(0.0, 0.4, 0.6, 0.0)
We aim to utilize all emotional annotations to improve the prediction of primary and secondary emotions!
Methodology

Purpose:
- Explore the benefits of using the typed words provided by evaluators when they selected the class "other" in the primary or secondary emotions for improving performance of SER systems

Method:
- Propose a three-dimensional (3D) polarity label (positive, negative, and ambiguous emotion words) with all emotional annotations
  - Include all typed words
  - Include primary and secondary emotions
  - Polarity obtained with Linguistic Inquiry and Word Count (LIWC) 2015

Emotion Corpus: MSP-PODCAST version 1.9

Audio sentences:
- Train set: 55,283
- Validation set: 9,546
- Test set: 16,570

Emotional Annotations:
- Crowdsourcing platform: Amazon Mechanical Turk
- Every sentence has more than 5 annotators
- 8-class Primary emotion (P) (Single-choice):
  - anger, sadness, happiness, surprise, fear, disgust, contempt, neutral, and other
- 16-class Secondary emotion (S) (Multi-choice):
  - Primary emotions
    - amusement, frustration, depression, concern, disappointment, excitement, confusion, and annoyance and other

Polarity Label Processing

Step 1: Pre-processing
- Lowercase and spell correction
- Check if secondary emotions (S) includes primary emotions (P) based on the rater-level

Step 2: Check variants of options
- Check if typed emotions are variants of list of emotions

Step 3: Classify polarity of emotional terms
- Linguistic Inquiry and Word Count (LIWC)
- Ambiguous emotion: LIWC does not provide a class

Step 4: Generate the final polarity label (Po)

<table>
<thead>
<tr>
<th>Primary emotion (P):</th>
<th>Secondary emotion (S):</th>
</tr>
</thead>
<tbody>
<tr>
<td>(W1) Other(Excited), (W2) Happy, (W3) Other(Pleased), (W4) Neutral, (W5) Angry</td>
<td></td>
</tr>
<tr>
<td>(W1) Other(Excited), (W2) Happy, (W3) Other(Pleased), (W4) Neutral, (W5) Excitement, Other(interested, curiosity, energetic), Neutral</td>
<td></td>
</tr>
</tbody>
</table>

P: Other(_excited), happy, Other(pleased), neutral, angry
S: Other(excited), happy, Other(pleased), neutral, + angry, Excitement, Other(interested, curiosity, energetic), neutral

S: Other(excited) → excitement, happy, Other(pleased), neutral, angry, excitement, Other(interested, curiosity, energetic), neutral

Positive emotion: happy, pleased, interested, curiosity, energetic, excitement, excitement
Ambiguous emotion: neutral, neutral
Negative emotion: angry

Po = (Neg, Amb, Pos) = (0.1, 0.2, 0.7)
Example of Sentiment Words in the MSP-Podcast Corpus

Positive Emotion Words
- Trust
- Joyful
- Relax

Negative Emotion Words
- Anxious
- Argue
- Aggression

Ambiguous Emotion Words
- Question
- Very
- Informal

Self-Related Words
- Assured
- Trusted
- Confident
- Inspired
- Trusted

Speech Emotion Classification (SEC) Model:
- Chunk-level SER model with the RNN-AttenVec chunk-level attention\(^1\)
- Same hyperparameters as the original paper\(^1\)

Acoustic feature extraction:
- Extract 512-dimensional wav2vec feature vector\(^2\) inspired by the analysis of Keesing et al. [2021]\(^3\)
- Features are z-normalized:
  - The parameters for the mean and standard deviation are estimated from the train set

---


Goal: Investigate the benefits of the proposed polarity label in the predictions of primary or secondary emotions

- **Single-task:** Primary emotion (P), Secondary emotion (S), Polarity label (Po)
- **Multi-task:** (P+Po), (S+Po), (S+P), (S+P+Po)

Experiment Setup (Multi-task SER)

- One single-task
- Two multi-tasks
- Three multi-tasks
Learning Objective Function and Evaluation Metric

Objective functions (Loss):
- Cross-entropy (CE) (softmax)
- Binary cross-entropy (BCE) (sigmoid)
- Kullback–Leibler divergence (KLD) (softmax)

Evaluation metric:
- Macro F1-score (maF1)
  - Binarize threshold: $1/K$, where $K$ is the number of class in the classification task
    - $1/8$ for primary emotion recognition task ($P$)
    - $1/16$ for secondary emotion recognition task ($S$)
Visualization of Improvement for the Prediction of P and S

The macro-F1 scores for primary emotion recognition (P) The macro-F1 scores for secondary emotion recognition (S)

Po: polarity label
P: primary emotion label
S: secondary emotion label
Conclusion

Contribution:
- Utilize annotators’ typed words of emotion perception to maximize the utilization of ratings for Speech Emotion Recognition (SER)

Method:
- Propose a 3D polarity label (positive/ambiguous/negative) to improve the prediction of primary and secondary emotion

Result:
- **8-class** Primary emotion classification: +6.4% performance gain
- **16-class** Secondary emotion classification: +16.56% performance gain

Findings:
- Typed words in the “Other” class have valuable information
- The SER task can be defined as a **multi-label task**