Every Rating Matters:

Joint Learning of Subjective Labels and Individual Annotators for Speech Emotion Classification

Huang-Cheng (David) Chou, Chi-Chun (Jeremy) Lee

Department of Electrical Engineering, National Tsing Hua University (NTHU), Hsinchu City, Taiwan

Full Paper: https://ieeexplore.ieee.org/abstract/document/8682170

Slides: https://sigport.org/documents/every-rating-matters-joint-learning-subjective-labels-and-individual-annotators-speech

Overview

Purpose:

Speech emotion classification from acoustic features

- ✓ Task: 4 categories (Neutral, Happiness, Sadness, Anger)

Overview

Purpose:

Speech emotion classification from acoustic features

- ✓ Task: 4 categories (Neutral, Happiness, Sadness, Anger)

Novelty:

A joint learning of **subjective labels** and **individual annotators**, utilizing soft-label and hard-label

- ✓ Use every rating (which are ignored in the previous works)
- ✓ Model individual annotators emotion perception

Overview

Purpose:

Speech emotion classification from acoustic features

✓ Task: 4 categories (Neutral, Happiness, Sadness, Anger)

Novelty:

To joint learning of subjective labels and individual annotators, utilizing **soft-**label **and hard-**label (conventional methods)

✓ Use every rating (which are ignored in the previous works)

✓ Model individual annotators emotion perception

Results:

- Unweighted Accuracy Recall (UAR): 57.12 % → **61.48 %**

Background

Sadness, Anger

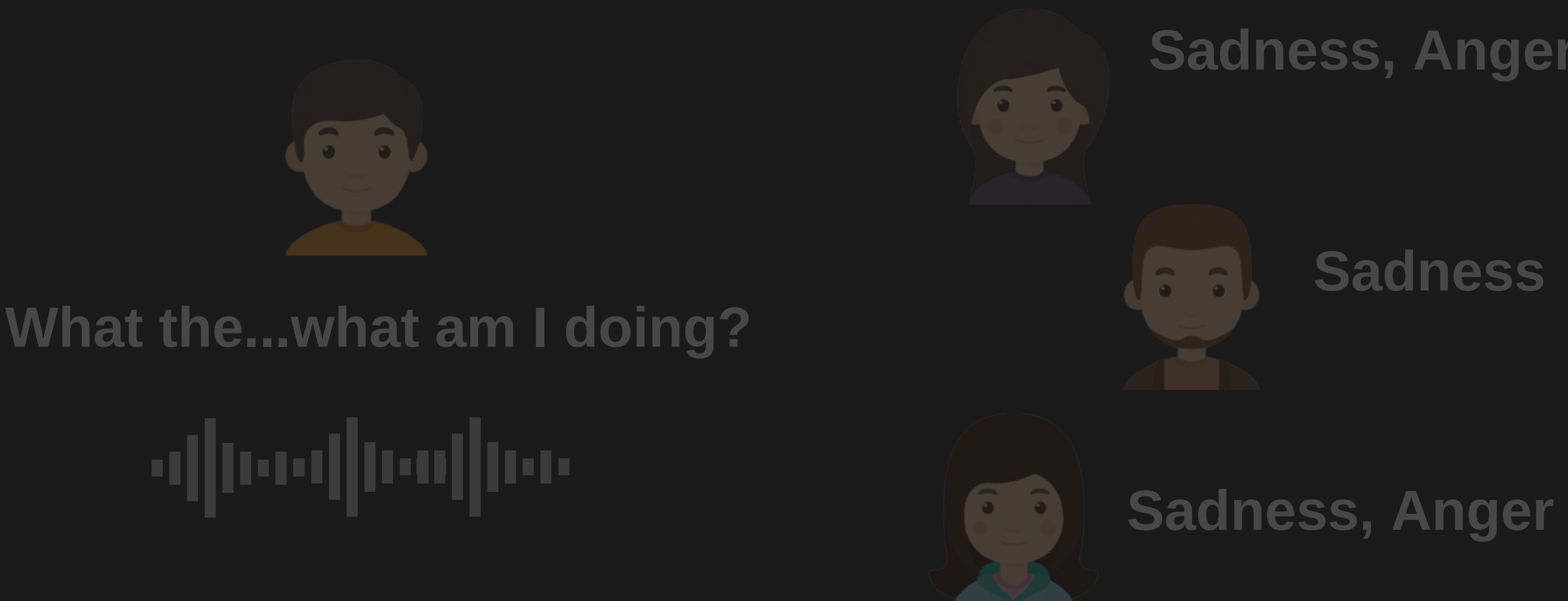
What the...what am I doing?

Sadness

Sadness, Anger

Background

Emotion perception is subjective because the natural bias of human, such as gender, age, and culture



Suzuki, Atsunobu, et al. "Decline or improvement?: Age-related differences in facial expression recognition." Biological psychology 74.1 (2007): 75-84.
Hall, Judith A., and David Matsumoto. "Gender differences in judgments of multiple emotions from facial expressions." Emotion 4.2 (2004): 201.
Matsumoto, David. "American-Japanese cultural differences in the recognition of universal facial expressions." Journal of cross-cultural psychology 23.1 (1992): 72-84.

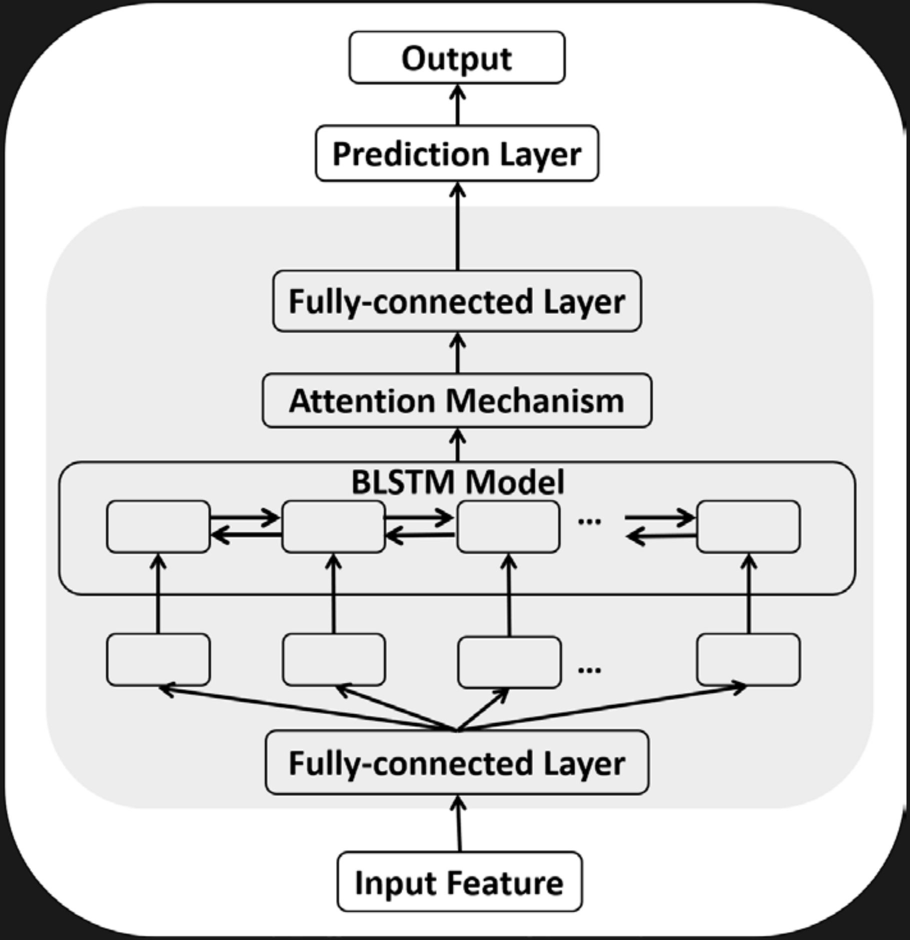
Conventional Method

Frame-level acoustic features + BLSTM-RNNs with Attention

Frame-level Features:

Pitch(F0), MFCCs, energy, loudness, voice probability, zero cross rate, … etc

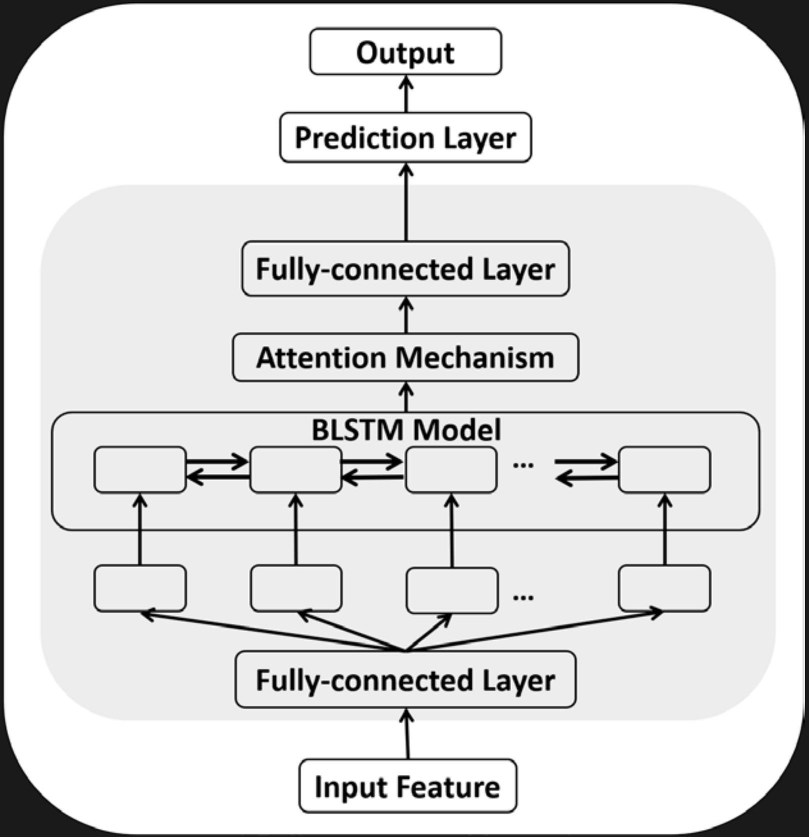
(All features are extracted by openSMILE toolbox)



Mirsamadi, Seyedmahdad, Emad Barsoum, and Cha Zhang. "Automatic speech emotion recognition using recurrent neural networks with local attention." 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2017.

Conventional Method

Use local attention to find specific emotional regions of each utterance



Mirsamadi, Seyedmahdad, Emad Barsoum, and Cha Zhang. "Automatic speech emotion recognition using recurrent neural networks with local attention." 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2017.

Data Label Preprocessing

Consensus (used in conventional method):

✓ **Majority vote of annotations**

=> Train emotion recognizer

Data Label Preprocessing

Working for corporate America? Wow.

Label

Data

Rating Others Sadness Sadness



Majority vote of ratings

Ground Truth Usage

Sadness

Conventional Hard-label Training

Model parameters are update

by cross-entropy loss:

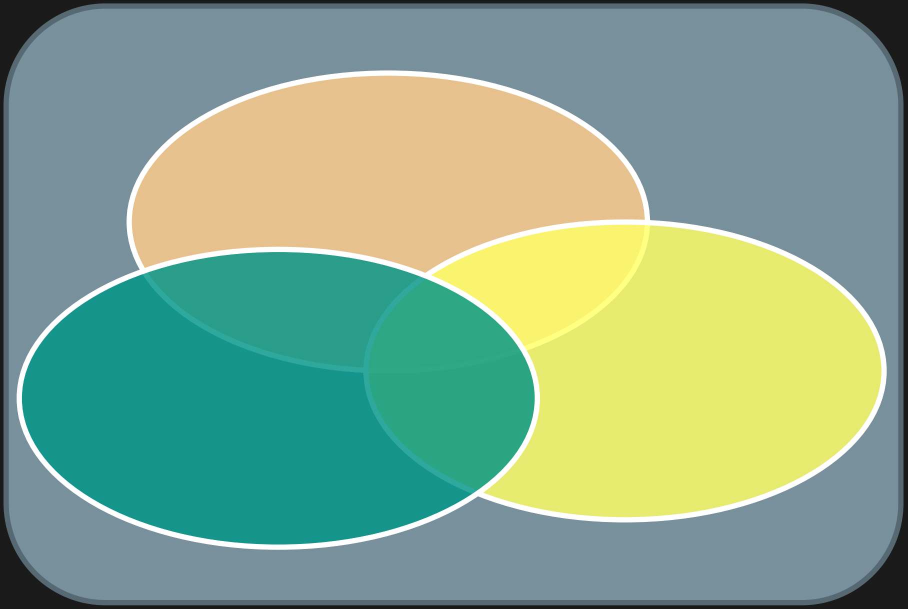
$$\boldsymbol{Loss = -\sum_{k=1}^{N}(pk * logqk),}$$

$$\boldsymbol{q_k = [0, 0, 1, 0]}$$

$$k = Total\ emotion\ classes$$

Conventional Method **Problem**

The boundaries between categories of emotion are fuzzy rather than discrete



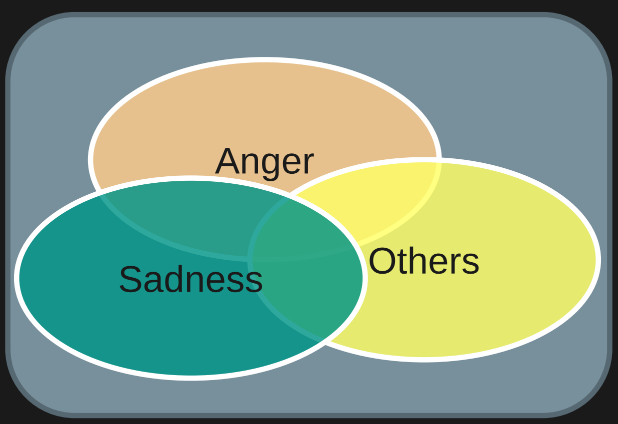
Discrete boundaries

Fuzzy boundaries

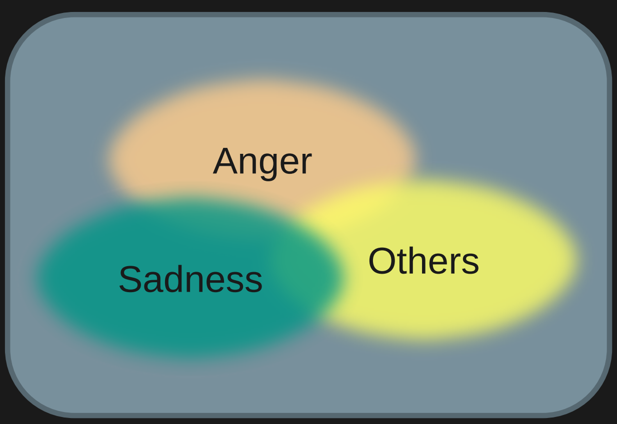
Cowen, Alan S., and Dacher Keltner. "Self-report captures 27 distinct categories of emotion bridged by continuous gradients." Proceedings of the National Academy of Sciences 114.38 (2017): E7900-E7909.

Conventional Method **Problem**

It just like the same music brings different sense of emotion feelings to different people



Discrete boundaries



Fuzzy boundaries

Cowen, Alan S., and Dacher Keltner. "Self-report captures 27 distinct categories of emotion bridged by continuous gradients." Proceedings of the National Academy of Sciences 114.38 (2017): E7900-E7909.

Conventional Method **Problem**

Emotion annotation can naturally have

disagreement and be ambiguous

Mower, Emily, et al. "Interpreting ambiguous emotional expressions." *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*. IEEE, 2009.

Conventional Method **Problem**

Emotion annotation can naturally have

disagreement and be ambiguous

The hard label loses

- ➔ **The variability of annotations**
- ➔ **The subjectivity in the emotion perception**

Mower, Emily, et al. "Interpreting ambiguous emotional expressions." *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*. IEEE, 2009.

Conventional Method Problem - **Why**

Working for corporate America? Wow.

Conventional Hard-label

Rating Others Sadness Anger

Ground Truth Usage

Conventional Method Problem - **Why**

Working for corporate America? Wow.

Conventional Hard-label

Rating Others Sadness Anger

Ground Truth Usage

No Consensus

Conventional Method Problem - **Why**

Working for corporate America? Wow.

Conventional Hard-label

Rating Others Sadness Anger

Ground Truth Usage

No use → **Training data limitation**

Soft-label Training

To address training data limitation

$$q(c_k) = \frac{\sum_n h_k^{(n)}}{\sum_{k'} \sum_n h_{k'}^{(n)}}$$

$$h_k^{(n)} = Binary\ label - existence(0/1),$$
$$n - th\ annotator, k - th\ emotion\ class$$

Steidl, Stefan, et al. "" Of all things the measure is man" automatic classification of emotions and inter-labeler consistency [speech-based emotion recognition]." *Proceedings.(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*. Vol. 1. IEEE, 2005.
Fayek, Haytham M., Margaret Lech, and Lawrence Cavedon. "Modeling subjectiveness in emotion recognition with deep neural networks: Ensembles vs soft labels." *2016 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2016.

Soft-label Training

To solve training data limitation

$$q(c_k) = \frac{\alpha + \sum_n h_k^{(n)}}{\alpha K + \sum_{k'} \sum_n h_{k'}^{(n)}}$$

$$\boldsymbol{\alpha \ = \ Smoothing\ coefficient}$$

$$\boldsymbol{k \ = Total\ emotion\ classes}$$

$$h_k^{(n)} = Binary\ label - existence(0/1),$$
$$n - th\ annotator, k - th\ emotion\ class$$

Ando, Atsushi, et al. "Soft-Target Training with Ambiguous Emotional Utterances for DNN-Based Speech Emotion Classification." 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2018.

Conventional Soft-label Method Problem - Why

What the...what am I doing?

Conventional Soft-label



Rating	Sadness, Anger	Sadness	Anger, Sadness
Ground Truth Usage	Sadness	Sadness	Anger

Emotional information lose

Conventional Soft-label Method Problem - Why

What the...what am I doing?

Conventional Soft-label

Rating	Sadness, Anger	Sadness	Anger, Sadness
			
Ground Truth Usage	Sadness	Sadness	Anger

Only used one of them

Conventional Soft-label Method Problem - Why

Use one rating

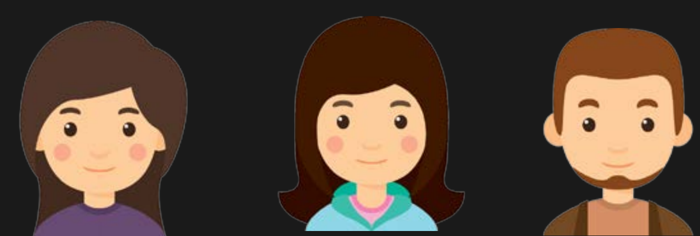
Conventional Soft-label

Rating	Sadness, Anger	Sadness	Anger, Sadness
Ground Truth Usage	Sadness	Sadness	Anger

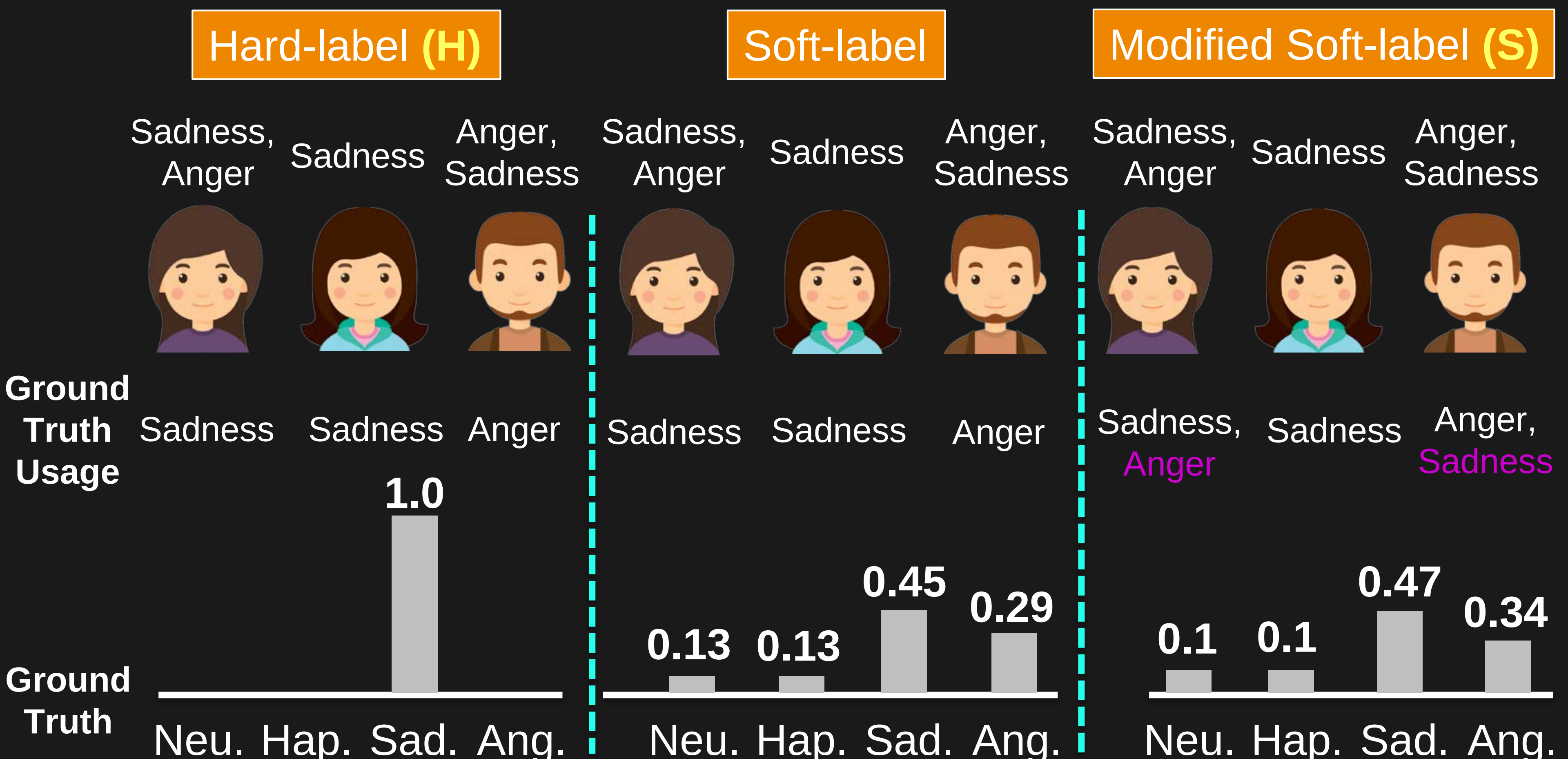
Use **every** rating

Modified Soft-label

Rating	Sadness, Anger	Sadness	Anger, Sadness
Ground Truth Usage	Sadness, **Anger**	Sadness	Anger, **Sadness**

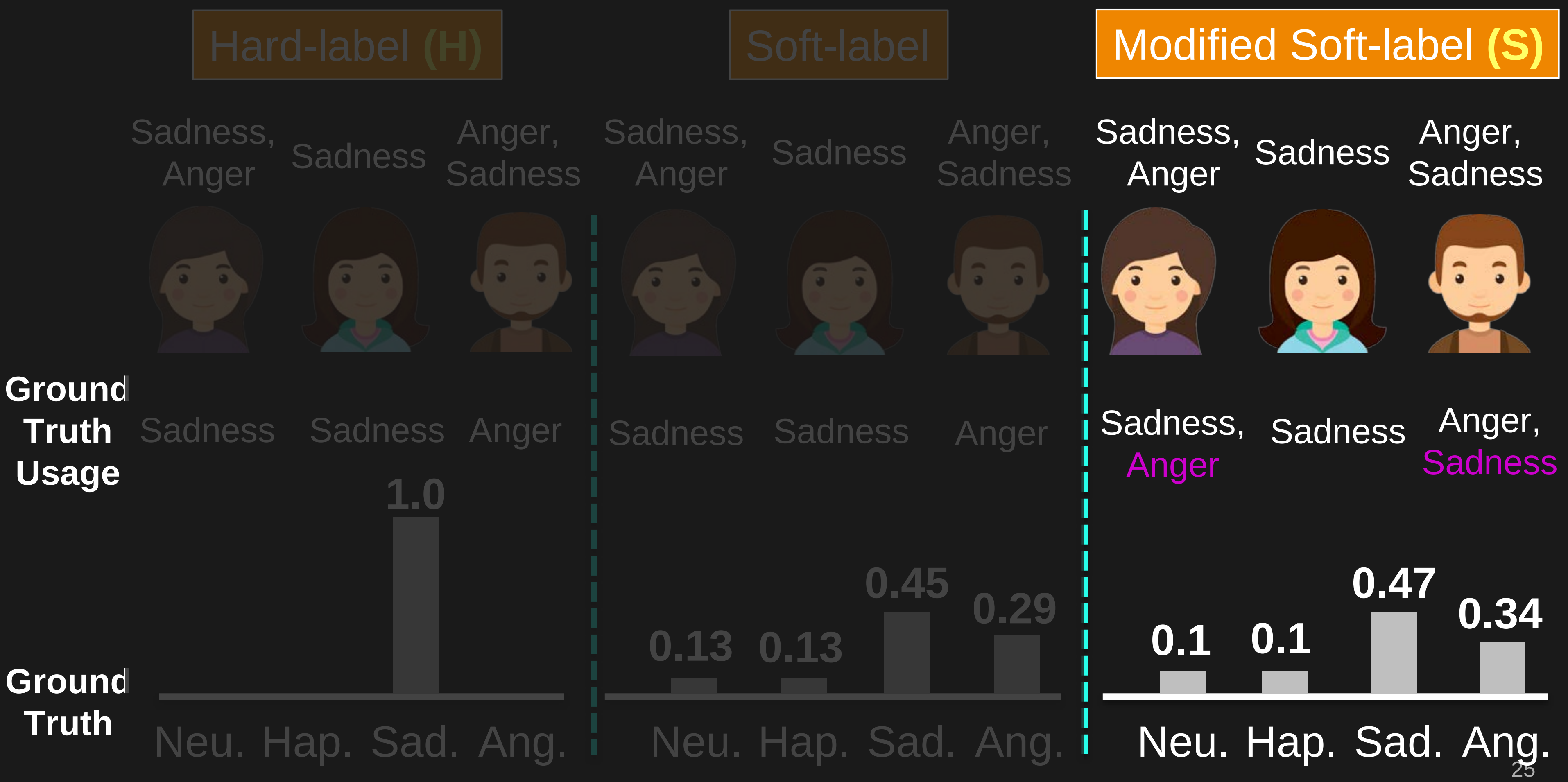


3 Different Method Targets



3 Different Method Targets

Modified soft ground truth is useful to represent subjective emotional *clues*



Model Emotional Sensitivity

- ✓ Use every rating
- ✓ **Build individual annotator’s emotion perception sensitivity model**

Fang, Xia, Gerben A. van Kleef, and Disa A. Sauter. "Revisiting cultural differences in emotion perception between easterners and westerners: Chinese perceivers are accurate, but see additional non-intended emotions in negative facial expressions." *Journal of Experimental Social Psychology* 82 (2019): 152-159.

Fischer, Agneta H., Mariska E. Kret, and Joost Broekens. "Gender differences in emotion perception and self-reported emotional intelligence: A test of the emotion sensitivity hypothesis." *PloS one* 13.1 (2018): e0190712.

Montagne, Barbara, et al. "Sex differences in the perception of affective facial expressions: Do men really lack emotional sensitivity?." *Cognitive processing* 6.2 (2005): 136-141.

Martin, Rod A., et al. "Emotion perception threshold: Individual differences in emotional sensitivity." *Journal of Research in Personality* 30.2 (1996): 290-305.

McCluskey, Ken W., and Daniel C. Albas. "Perception of the emotional content of speech by Canadian and Mexican children, adolescents, and adults." *International Journal of Psychology*16.1-4 (1981): 119-132.

Model Emotional Sensitivity

- ✓ **Build individual annotator's emotion perception sensitivity model**

- ✓ **Emotional sensitivity** is different from person to person because the natural bias of human, like gender, age, and culture

Fang, Xia, Gerben A. van Kleef, and Disa A. Sauter. "Revisiting cultural differences in emotion perception between easterners and westerners: Chinese perceivers are accurate, but see additional non-intended emotions in negative facial expressions." *Journal of Experimental Social Psychology* 82 (2019): 152-159.

Fischer, Agneta H., Mariska E. Kret, and Joost Broekens. "Gender differences in emotion perception and self-reported emotional intelligence: A test of the emotion sensitivity hypothesis." *PloS one* 13.1 (2018): e0190712.

Montagne, Barbara, et al. "Sex differences in the perception of affective facial expressions: Do men really lack emotional sensitivity?." *Cognitive processing* 6.2 (2005): 136-141.

Martin, Rod A., et al. "Emotion perception threshold: Individual differences in emotional sensitivity." *Journal of Research in Personality* 30.2 (1996): 290-305.

McCluskey, Ken W., and Daniel C. Albas. "Perception of the emotional content of speech by Canadian and Mexican children, adolescents, and adults." *International Journal of Psychology*16.1-4 (1981): 119-132.

Model Emotional Sensitivity - Why

People use to understand our own emotional experience also helps us understand the emotions of others

Individual annotator models

Israelashvili, Jacob, et al. "Knowing me, knowing you: Emotion differentiation in oneself is associated with recognition of others' emotions." *Cognition and Emotion* (2019): 1-11.

Experiments

Purpose:

1. Use different types of label (H label/S label) for training
2. Model individual annotators emotion perception
3. Joint all-annotators (**Crowd**) and individual model ($\mathbf{E_N}$)

Busso, Carlos, et al. "IEMOCAP: Interactive emotional dyadic motion capture database." Language resources and evaluation42.4 (2008): 335.

Experiments

Purpose:

1. Use different types of label (**H label & S label**) for training
2. Model individual annotators emotion perception
3. Joint all-annotators (**Crowd**) and individual model ($\mathbf{E_N}$)

Busso, Carlos, et al. "IEMOCAP: Interactive emotional dyadic motion capture database." Language resources and evaluation42.4 (2008): 335.

Experiments

Dataset: **IEMOCAP** Database [Busso+, 08]

- Task: Dyadic emotional interaction (1 male, 1 female)
- Total # of session: 5
- Total # of speakers: 10 (train: 8, test: 2 / per session)
- Average # of annotators / per each utterance: 3 (self, observe)
- # of chose individual annotators (observe): 5

Busso, Carlos, et al. "IEMOCAP: Interactive emotional dyadic motion capture database." Language resources and evaluation42.4 (2008): 335.

Data Usage

Purpose:

1. Use H label and S label for train
2. Model 5 observed annotators emotion perception ($E_1 \sim E_5$)
3. Joint **Crowd and E_N (**will be discussed in the setups**)**

The # of S and H label utterance for each model			
Model	Total	Soft label	Hard label
$Crowd_H$	5531	0	5531
$Crowd_S$	7774	3185	4589
$E1$	5954	44	5910
$E2$	7845	38	7807
$E4$	6429	212	6217
$E5$	422	3	419
$E6$	773	20	753

Data Usage

Purpose:

1. Use H label and S label for train

All-annotators model:
- H: use hard-label
- S: use soft-label

(**Baseline**)

The # of S and H label utterance for each model			
Model	Total	Soft label	Hard label
$Crowd_H$	5531	0	5531
$Crowd_S$	7774	3185	4589
$E1$	5954	44	5910
$E2$	7845	38	7807
$E4$	6429	212	6217
$E5$	422	3	419
$E6$	773	20	753

Data Usage

Purpose:

2. Model 5 observed annotators emotion perception ($E_1 \sim E_5$)

All-annotators model:
- H: use hard-label
- S: use soft-label

(**Baselines**)

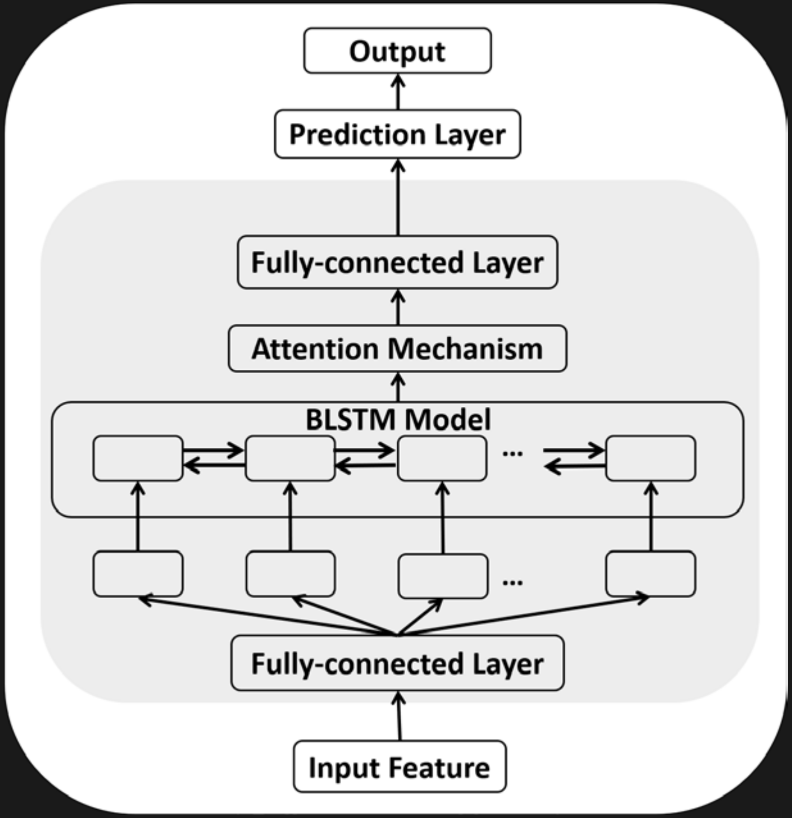
Individual model:
- **Use soft-label**

The # of S and H label utterance for each model			
Model	Total	Soft label	Hard label
$Crowd_H$	5531	0	5531
$Crowd_S$	7774	3185	4589
$E1$	5954	44	5910
$E2$	7845	38	7807
$E4$	6429	212	6217
$E5$	422	3	419
$E6$	773	20	753

Setups

Classifier: BLSTM with attention [Ando + , 2018]

- Main Structure
  - [Dense,256]-[BLSTM with attention,128]- [Dense,256]



Setups

Classifier: BLSTM with attention [Ando + , 2018]

- Main Structure
  - [Dense,256]-[BLSTM with attention,128]- [Dense,256]
- Input: acoustic low level descriptors (LLDs), 45 dims.
  - **12 MFCCs, Δ12 MFCCs, ΔΔ12 MFCCs,**
  - **Loudness, Δ Loudness, ΔΔ Loudness**
  - **Pitch (F0), Δ Pitch (F0), Probability of voicing, Δ Probability of voicing,**
  - **Zero-crossing rate, Δ Zero-crossing rate**

Setups

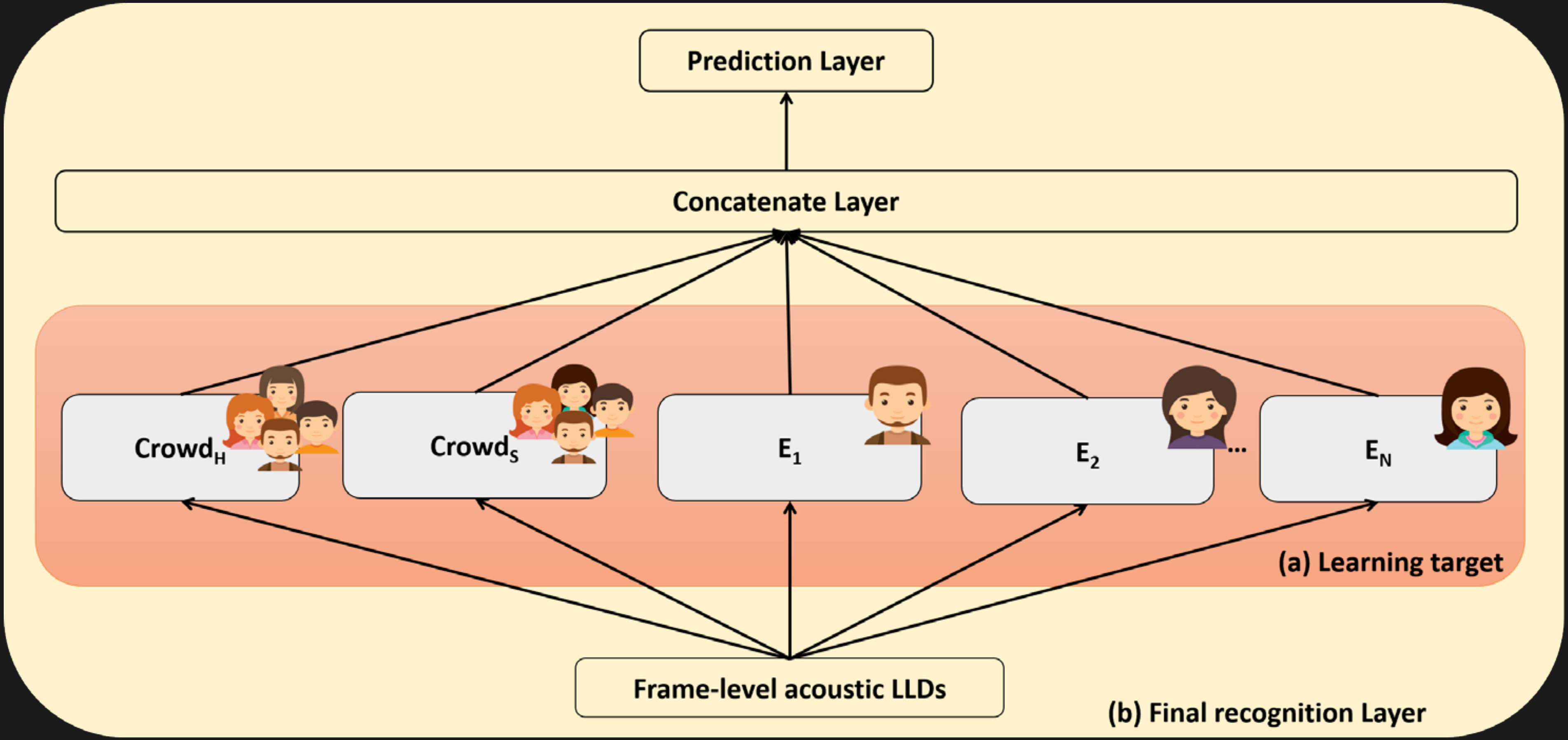
Classifier: BLSTM with attention [Ando + , 2018]

- Main Structure
  - [Dense,256]-[BLSTM with attention,128]- [Dense,256]
- Input: acoustic low level descriptors (LLDs), 45 dims.
  - **12 MFCCs, Δ12 MFCCs, ΔΔ12 MFCCs,**
  - **Loudness, Δ Loudness, ΔΔ Loudness**
  - **Pitch (F0), Δ Pitch (F0), Probability of voicing, Δ Probability of voicing,**
  - **Zero-crossing rate, Δ Zero-crossing rate**
- Target: **Hard-label (one-hot label) when testing**

Evaluation measure: Unweighted Accuracy Recall (UAR)

- Average results of 5 sessions (**Leave-one-session-out**)

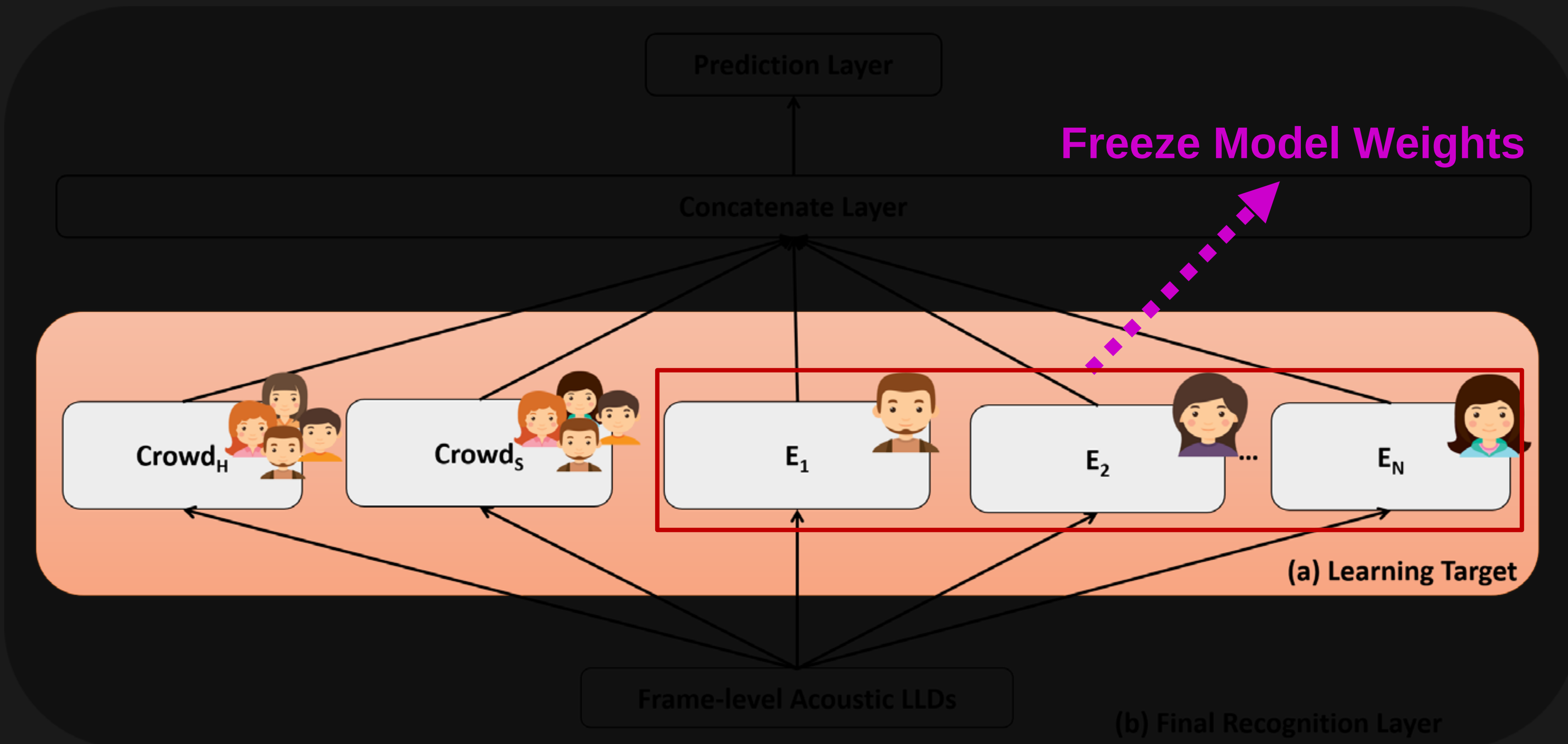
Joint Crowd and E_N



(a) Learning target (b) Final recognition Layer

Joint Crowd and E_N

Stage-1: Train E_N models and Crowd models

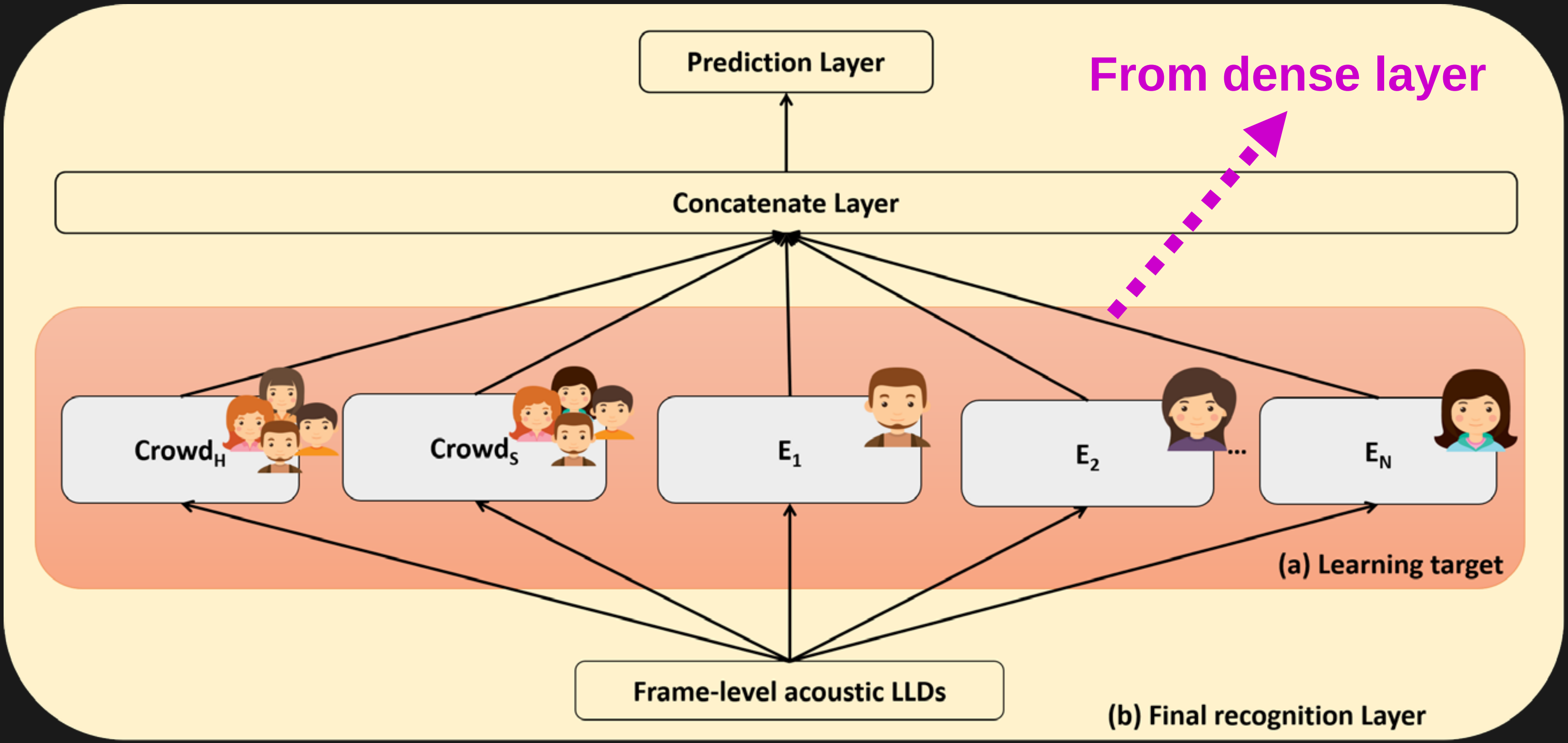


(a) Learning target (b) Final recognition Layer

Joint Crowd and E_N

Stage-1: Train E_N models and Crowd models

Stage-2: Use stage weights, train some epochs; then recognition (Fix E_N weights)



(a) Learning target (b) Final recognition Layer

Results

Our proposed model achieves unweighted average recall (UAR) 61.48%

Model	Overall	Neutral	Anger	Happiness	Sadness
$Crowd_H$	57.45%	55.71%	63.29%	45.02%	65.77%
$Crowd_S$	57.12%	49.70%	62.98%	62.85%	53.14%
$E1$	50.98%	8.04%	61.31%	77.24%	57.34%
$E2$	59.68%	38.78%	64.35%	64.25%	62.61%
$E4$	48.59%	81.29%	45.42%	38.20%	29.44%
$E5$	37.62%	86.89%	47.62%	11.21%	4.75%
$E6$	45.82%	36.85%	40.10%	60.39%	45.95%
$Crowd_{HS}$	58.58%	59.66%	59.31%	53.63%	61.71%
$Proposed$	**61.48%**	**54.55%**	**64.51%**	**60.32%**	**66.56%**

Results

$Crowd_S$ obtains a better recognition rate for happiness compared to $Crowd_H$

Model	Overall	Neutral	Anger	Happiness	Sadness
$\mathbf{Crowd_H}$	57.45%	55.71%	63.29%	**45.02%**	65.77%
$\mathbf{Crowd_S}$	57.12%	49.70%	62.98%	**62.85%**	53.14%
$E1$	50.98%	8.04%	61.31%	77.24%	57.34%
$E2$	59.68%	38.78%	64.35%	64.25%	62.61%
$E4$	48.59%	81.29%	45.42%	38.20%	29.44%
$E5$	37.62%	86.89%	47.62%	11.21%	4.75%
$E6$	45.82%	36.85%	40.10%	60.39%	45.95%
$Crowd_{HS}$	58.58%	59.66%	59.31%	53.63%	61.71%
$Proposed$	61.48%	54.55%	64.51%	60.32%	66.56%

Results

$Crowd_H$ works better for neutral and sadness than $Crowd_S$

Model	Overall	Neutral	Anger	Happiness	Sadness
$\mathbf{Crowd_H}$	57.45%	**55.71%**	63.29%	45.02%	**65.77%**
$\mathbf{Crowd_S}$	57.12%	**49.70%**	62.98%	62.85%	**53.14%**
$E1$	50.98%	8.04%	61.31%	77.24%	57.34%
$E2$	59.68%	38.78%	64.35%	64.25%	62.61%
$E4$	48.59%	81.29%	45.42%	38.20%	29.44%
$E5$	37.62%	86.89%	47.62%	11.21%	4.75%
$E6$	45.82%	36.85%	40.10%	60.39%	45.95%
$Crowd_{HS}$	58.58%	59.66%	59.31%	53.63%	61.71%
$Proposed$	61.48%	54.55%	64.51%	60.32%	66.56%

Results

$\boldsymbol{E1}$ and $\boldsymbol{E2}$ models are good at telling anger and happiness

Model	Overall	Neutral	Anger	Happiness	Sadness
$Crowd_H$	57.45%	55.71%	63.29%	45.02%	65.77%
$Crowd_S$	57.12%	49.70%	62.98%	62.85%	53.14%
$\mathbf{E1}$	50.98%	8.04%	**61.31%**	**77.24%**	57.34%
$\mathbf{E2}$	59.68%	38.78%	**64.35%**	**64.25%**	62.61%
$E4$	48.59%	81.29%	45.42%	38.20%	29.44%
$E5$	37.62%	86.89%	47.62%	11.21%	4.75%
$E6$	45.82%	36.85%	40.10%	60.39%	45.95%
$Crowd_{HS}$	58.58%	59.66%	59.31%	53.63%	61.71%
$Proposed$	61.48%	54.55%	64.51%	60.32%	66.56%

Results

$E4$ and $E5$ models are good at telling neutral

Model	Overall	Neutral	Anger	Happiness	Sadness
$Crowd_H$	57.45%	55.71%	63.29%	45.02%	65.77%
$Crowd_S$	57.12%	49.70%	62.98%	62.85%	53.14%
$E1$	50.98%	8.04%	61.31%	77.24%	57.34%
$E2$	59.68%	38.78%	64.35%	64.25%	62.61%
E4	48.59%	**81.29%**	45.42%	38.20%	29.44%
E5	37.62%	**86.89%**	47.62%	11.21%	4.75%
$E6$	45.82%	36.85%	40.10%	60.39%	45.95%
$Crowd_{HS}$	58.58%	59.66%	59.31%	53.63%	61.71%
$Proposed$	61.48%	54.55%	64.51%	60.32%	66.56%

Results

$\boldsymbol{E6}$ model is sensitive to Happiness

Model	Overall	Neutral	Anger	Happiness	Sadness
$Crowd_H$	57.45%	55.71%	63.29%	45.02%	65.77%
$Crowd_S$	57.12%	49.70%	62.98%	62.85%	53.14%
$E1$	50.98%	8.04%	61.31%	77.24%	57.34%
$E2$	59.68%	38.78%	64.35%	64.25%	62.61%
$E4$	48.59%	81.29%	45.42%	38.20%	29.44%
$E5$	37.62%	86.89%	47.62%	11.21%	4.75%
$\mathbf{E6}$	45.82%	36.85%	40.10%	**60.39%**	45.95%
$Crowd_{HS}$	58.58%	59.66%	59.31%	53.63%	61.71%
$Proposed$	61.48%	54.55%	64.51%	60.32%	66.56%

Conclusion

Summary:

- ✓ Purpose: speech emotion classification from acoustic LLDs
- ✓ Approach: Utilizing every rating to model subjective labels and individual annotators
- ✓ Method: Soft-label and hard-label joint learning
- ✓ Results: Performances were improved
  - 57.45% [$Crowd_H$] → 61.48% (**3.18%**)
  - 57.12% [$Crowd_S$] → 61.48% (**4.36%**)

Future works:

- ✓ Evaluations by other language emotion dataset, such as **NNIME** database [Chou+, 2017]
- ✓ Test on personalized emotion perception recognition

Chou, Huang-Cheng, et al. "NNIME: The NTHU-NTUA Chinese interactive multimodal emotion corpus." 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII). IEEE, 2017.

Reviewers’ Questions

Potential Issues

- ✓ Why use soft-label for training but evaluate on hard-label ?

Potential Issues

- ✓ Why soft-label training improves model performance?
- Because the training data increased, we get the same finding with previous works [Ando+, 18] and [Kim+, 18].

Ando, Atsushi, et al. "Soft-Target Training with Ambiguous Emotional Utterances for DNN-Based Speech Emotion Classification." 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2018.

Kim, Yelin, and Jeesun Kim. "Human-Like Emotion Recognition: Multi-Label Learning from Noisy Labeled Audio-Visual Expressive Speech." 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2018.

Potential Issues

- ✓ How is the robustness for modeling individual annotators? If we remove 1 or 2 annotator from training process, does this model can still work?

Ando, Atsushi, et al. "Soft-Target Training with Ambiguous Emotional Utterances for DNN-Based Speech Emotion Classification." 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2018.
Kim, Yelin, and Jeesun Kim. "Human-Like Emotion Recognition: Multi-Label Learning from Noisy Labeled Audio-Visual Expressive Speech." 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2018.

Annotation distribution (ratio)

$Note$: if two (or more) ratings for one data from annotator, we will calculate by 2 (or more).

Model	Neutral	Anger	Happiness	Sadness
$Crowd_H$	30.88%	19.94%	29.58%	19.60%
$Crowd_S$	29.33%	17.77%	35.79%	17.10%
$\boldsymbol{E1}$	8.49%	21.21%	**49.67%**	20.64%
$\boldsymbol{E2}$	22.45%	26.58%	**31.35%**	19.62%
$\boldsymbol{E4}$	**52.88%**	12.41%	23.76%	10.95%
$\boldsymbol{E5}$	**69.88%**	15.29%	8.94%	5.88%
$\boldsymbol{E6}$	26.73%	15.76%	**43.38%**	14.22%

Results (Only E_N Model)

$\boldsymbol{E}_N$ model is sensitive to Happiness, Anger, and Sadness. Instead, Crow model has good recognition rate for Neutral

Model	Overall	Neutral	Anger	Happiness	Sadness
$Crowd_H$	57.45%	55.71%	63.29%	45.02%	65.77%
$Crowd_S$	57.12%	49.70%	62.98%	62.85%	53.14%
$E1$	50.98%	8.04%	61.31%	77.24%	57.34%
$E2$	59.68%	38.78%	64.35%	64.25%	62.61%
$E4$	48.59%	81.29%	45.42%	38.20%	29.44%
$E5$	37.62%	86.89%	47.62%	11.21%	4.75%
$E6$	45.82%	36.85%	40.10%	60.39%	45.95%
$Crowd_{HS}$	58.58%	59.66%	59.31%	53.63%	61.71%
$\boldsymbol{E}_N$	**60.24%**	**49.64%**	**63.64%**	**61.48%**	**66.19%**
$Proposed$	**61.48%**	54.55%	64.51%	60.32%	66.56%

Thank You

Full Paper: https://ieeexplore.ieee.org/abstract/document/8682170

Slides: https://sigport.org/documents/every-rating-matters-joint-learning-subjective-labels-and-individual-annotators-speech

Question ?

Full Paper: https://ieeexplore.ieee.org/abstract/document/8682170

Slides: https://sigport.org/documents/every-rating-matters-joint-learning-subjective-labels-and-individual-annotators-speech