Generative Approach Using Soft-Labels to Learn Uncertainties in Predicting Emotional Attributes

Kusha Sridhar, Carlos Busso

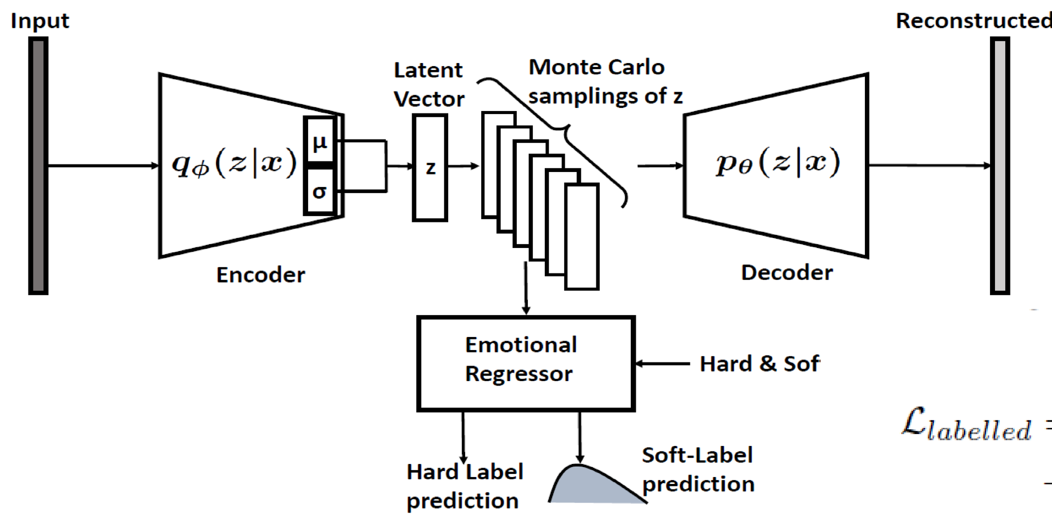
UT Dallas MSP Multimodal Signal Processing Laboratory

THE UNIVERSITY OF TEXAS AT DALLAS

Affective Computing & Intelligent Interaction (ACII) 2021
Virtual Event
28 September - 1 October 2021

Proposed Framework

- Exploring annotator level uncertainties
- Generative modeling approach with soft-labels of emotional attributes
- Variational Autoencoder (VAE) with an Emotional Regressor (ER) attached to the bottleneck layer
  - Multiple Monte Carlo from the latent space of the VAE to learn prediction uncertainties
  - Soft-labels to train ER. Hard-labels to constrain the latent space of VAE



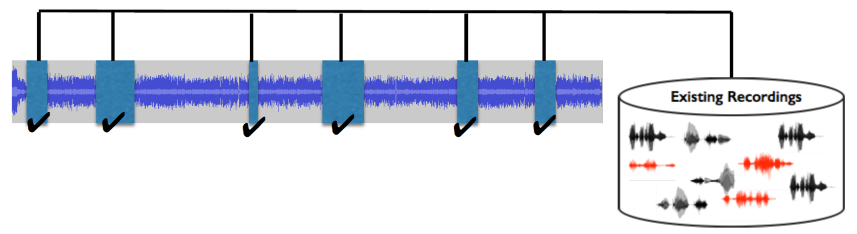
$$\mathcal{L}_{unlabelled} = -E_{q_\phi(z|x)} log(p_\theta(x|z))$$

$$\mathcal{L}_{labelled} = \alpha \cdot (NLL)_{VAE} + \beta \cdot (1 - CCC)_{consensus} + \gamma \cdot (KLD)_{soft}$$

Database and Features

The MSP-Podcast Corpus

- Emotionally rich speaking turn from speaker appearing in various podcasts (2.75s to 11s in length)
- Annotated for categorical and attribute-based emotion labels on Amazon Mechanical Turk
- Version 1.6: Train = 34,280 sentences
  Test = 10,124 sentences from 50 speakers
  Validation = 5,958 sentences from 40 speakers
- 42,567 sentences with speaker ID (1,078 speakers)



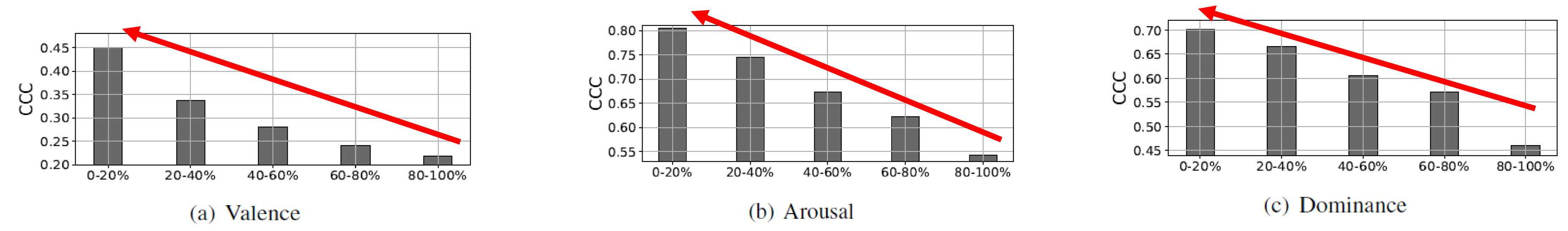
Acoustic Features

- Interspeech 2013 ComParE feature set extracted using Opensmile toolkit
- 65 LLDs and 6,373 HLDs

Uncertainty Analysis and Reject Options

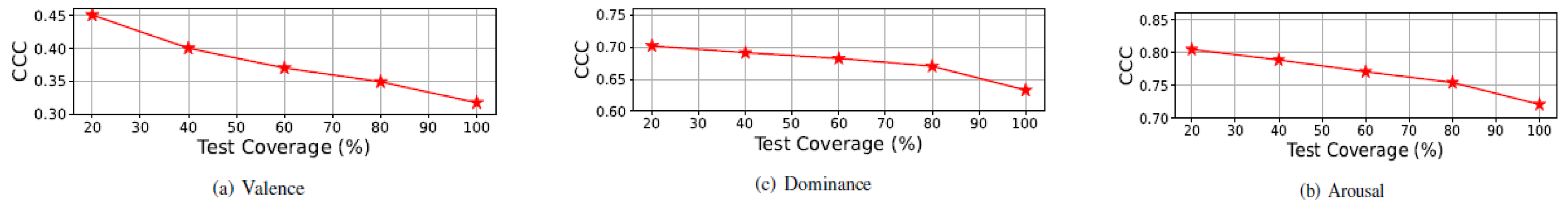
Predicted attribute scores in terms of CCC as a function of prediction uncertainties

- Entropy to quantify uncertainty



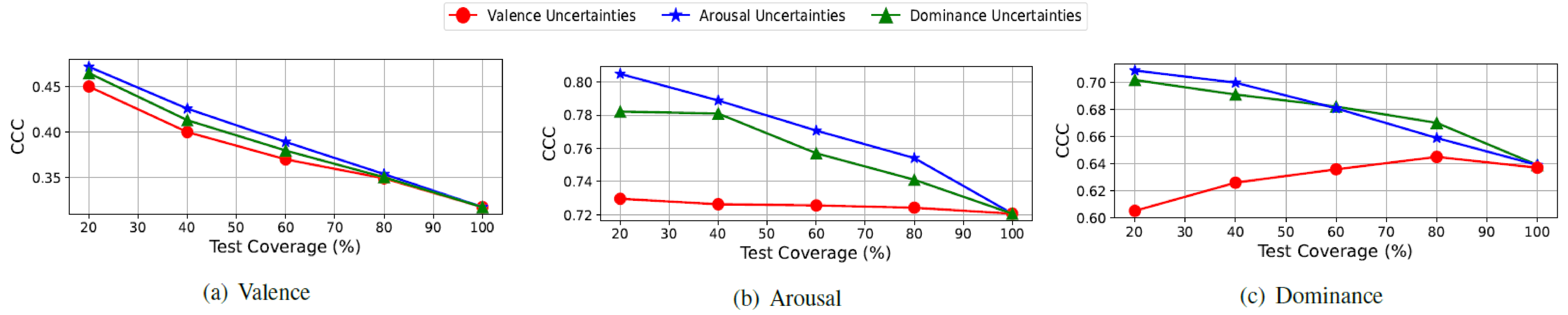
Application in Reject Options for SER

- At 60% test coverage, relative gains in CCC up to 16.85% for valence, 7.12% for arousal and 8.01% for dominance



Uncertainty Transfer Learning (UTL)

- **Can information from uncertainty prediction on one emotional attribute be transferred to another emotional attribute?**
- **Arousal and dominance uncertainties improve valence recognition performance but the vise versa is not true**
  - Transferred learned uncertainties lead to higher performance gains than self-learned uncertainties



Gains up to 19.55% at 60% test coverage for valence

Model Generalization

Cross-corpus experiments on the IEMOCAP corpus

- Comparing proposed approach with other Monte Carlo Dropout (MCD) based approaches used for uncertainty modeling
- Experiments with UTL in application to reject options

Attribute – Uncertainty Pairs	Approach	CCC at different coverage			
		80%	60%	40%	20%
Val-Val	Proposed	**0.5489**	**0.5710**	**0.6122**	0.6433
	AE-MCD	0.5380	0.5561	0.5918	0.6450
	MCD MTL	0.5400	0.5650	0.5888	**0.6492**
	MCD Soft	0.5391	0.5691	0.5820	0.6000
	MCD	0.5340	0.5620	0.5725	0.6181
Val-Aro	Proposed	**0.5500**	**0.5929**	**0.6292**	**0.6558**
	AE-MCD	0.5420	0.5650	0.5900	0.6329
	MCD MTL	0.5322	0.5690	0.5899	0.5980
	MCD Soft	0.5480	0.5710	0.5902	0.5981
	MCD	0.5280	0.5315	0.5428	0.5750
Val-Dom	Proposed	**0.5498**	**0.5717**	**0.6073**	**0.6490**
	AE-MCD	0.5411	0.5650	0.5881	0.6126
	MCD MTL	0.5325	0.5350	0.5510	0.5738
	MCD Soft	0.5450	0.5680	0.5838	0.5811
	MCD	0.5255	0.5364	0.5427	0.5833

Conclusion

- **Novel generative modeling approach using soft-labels of emotional attributes in uncertainty predictions**

Experiments	Merits of the proposed VAE-ER approach
Uncertainty Predictions	Using soft-labels
Reject Options using self-learned uncertainties	At 60% coverage, gains in CCC up to: 16.85% for valence 7.12% for arousal 8.01% for dominance
Uncertainty Transfer Learning	Works best with valence 19.55% gains in CCC
Model Generalization	Cross-corpus results on IEMOCAP: Generalizes better than MCD based approaches
Computational complexity at inference	74.36% faster than MCD based approaches

This work was funded by NSF career award IIS-1453781