Uncertainty-Aware Multi-Modal Ensembling for Severity Prediction of Alzheimer's Dementia

Utkarsh Sarawgi, Wazeer Zukfikar, Rishab Khincha, Pattie Maes

{utkarshs, wazeer, rkhincha, pattie} @ mit.edu

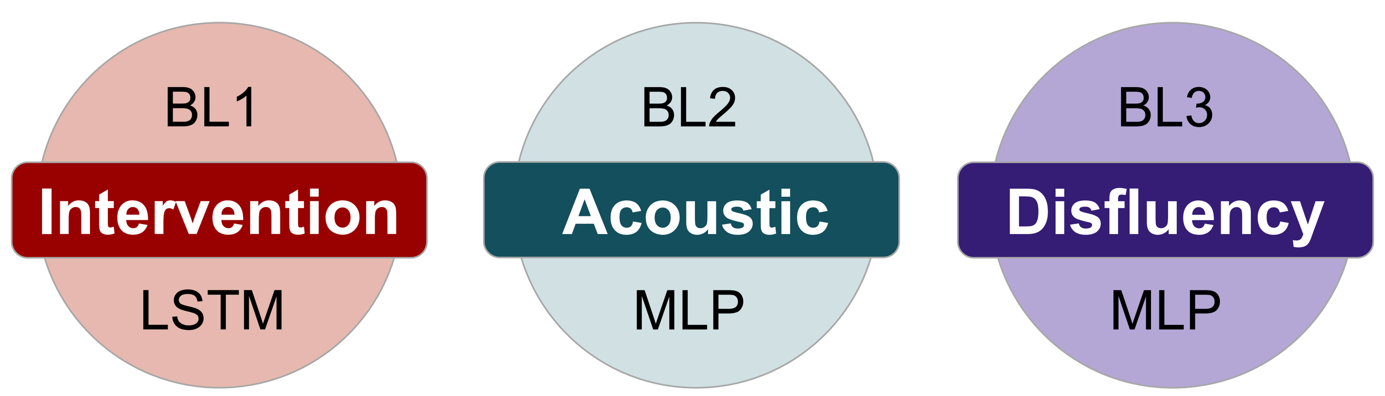
Paper

NEURAL INFORMATION PROCESSING SYSTEMS

Summary

- We present a novel uncertainty-aware (UA) boosting technique for multi-modal ensembling.
- We show how such a boosting method leverages the uncertainty estimates to produce a system that is robust to heteroscedasticity in the data.
- We evaluate this on the benchmark ADReSS dataset to predict the severity of Alzheimer's Dementia (AD).
- We first compare with vanilla boosting and then with other state-of-the-art methods.
- Our UA ensemble outperforms all other methods while also reducing the overall entropy of the system.
- This work aims to encourage fair, aware, and reliable systems and engage in such discussions.

Multi-Modal Base Learners



Training Base Learners

- To estimate uncertainties, we train each of the 3 base learners with a target Gaussian distribution $p_\theta(y_n|\mathbf{x}_n)$ parameterized by mean μ_θ and the standard deviation $\sigma_\theta;\ y_n \sim \mathcal{N}\left(\mu_\theta(\mathbf{x}_n), \sigma_\theta^2(\mathbf{x}_n)\right)$
- Each base learner is trained with their corresponding input features $\mathbf{x}$ and the ground truth values y using a proper scoring rule $l(\theta, \mathbf{x}, y)$, as shown below:

$$-\log p_\theta(y_n|\mathbf{x}_n) = \frac{\log \sigma_\theta^2(\mathbf{x})}{2} + \frac{(y - \mu_\theta(\mathbf{x}))^2}{2\sigma_\theta^2(\mathbf{x})} + \text{constant}$$

Uncertainty-Aware Ensemble

We use boosting to train an ensemble with the 3 base learners.

- **Vanilla ensemble**: This would use RMSE values to weigh the loss while sequentially boosting across the base learners, and then average the predictions of all the base learners for the final prediction.
- **UA ensemble:** We propose an 'uncertainty-aware ensemble' where we use predictive uncertainty quantified by the predicted standard deviation to weigh the loss while sequentially boosting across the base learners, and then average the predictions of all the base learners for the final prediction.
- **UA ensemble (weighted)**: We experiment a variation to the averaging of the predictions. Here, for the final prediction, we take a weighted average of the predictions, where the weights used are the inverse of the respective predictive uncertainty, quantified by the predicted standard deviation, as shown below:

$$P(\mathbf{x}_n) = \frac{\sum_{i=1}^{N} \frac{1}{\sigma_{\theta_i}(\mathbf{x}_n)} \mu_{\theta_i}(\mathbf{x}_n)}{\sum_{i=1}^{N} \frac{1}{\sigma_{\theta_i}(\mathbf{x}_n)}}, \text{ where } \begin{array}{l} P(x_n)\text{: final prediction} \\ N\text{: \# of base learners} \end{array}$$

Table 1

Model	RMSE
Disfluency	5.71 ± 0.39
Acoustic	6.66 ± 0.30
Interventions	6.41 ± 0.53
Vanilla Ensemble	5.17 ± 0.27
UA Ensemble	$\mathbf{5.05 \pm 0.53}$
UA Ensemble (weighted)	$\mathbf{4.96 \pm 0.49}$

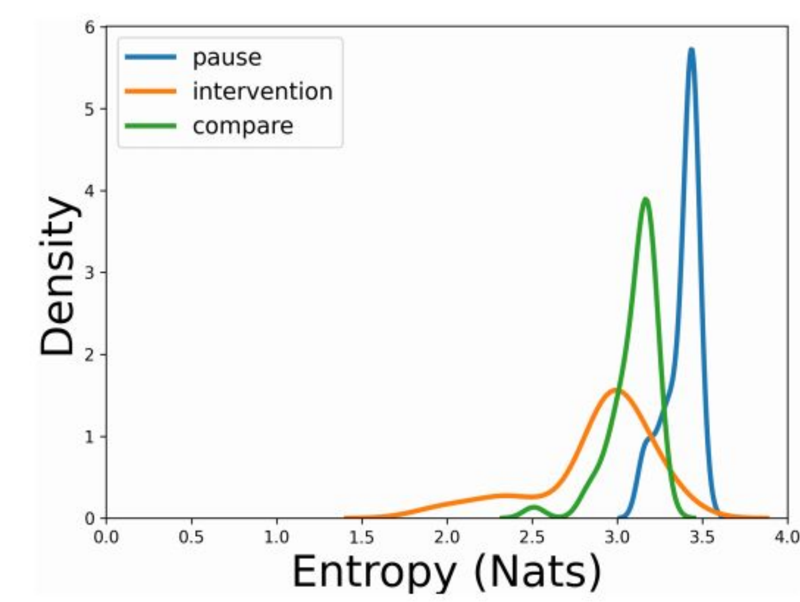
Table 2

Model	RMSE	Model	RMSE
Pappagari et al.	5.37	Rohanian et al.	4.54
Luz et al.	5.20	Sarawgi et al.	4.37
Sarawgi et al.	4.60	**UA Ensemble**	**4.35**
Searle et al.	4.58	**UA Weighted**	**3.93**

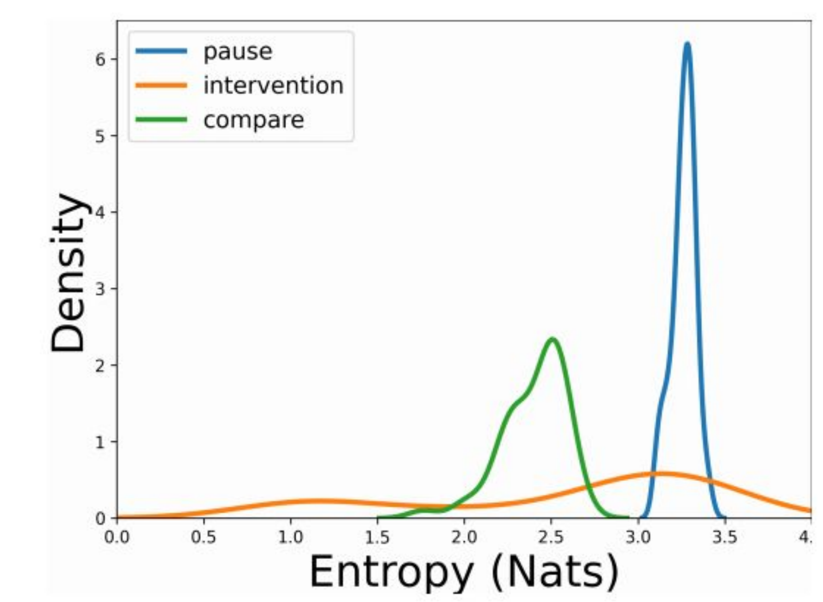
Evaluation and Results

- For robustness, we repeat every evaluation 5 times using random seeds and report the mean and the variance of the RMSE results on the standardized held-out test set.
- We first evaluate each modality i.e. base learner individually and then compare with the vanilla and uncertainty-aware ensemble (Table 1).
- We also compare our uncertainty-aware ensemble methods with current state-of-the-art methods (Table 2).
- To evaluate the robustness of the ensemble with uncertain datapoints/subjects, we evaluate the entropy of the base learners in the ensemble methods while sequentially boosting in the vanilla ensemble and the UA ensemble (Figures below).

Vanilla ensemble



UA ensemble



Discussion

- The propagation of the uncertainty sequentially through the base learners of every modality aids the multi-modal system to decrease the overall entropy in the system, making it more reliable when compared to vanilla ensembles.
- The increased reduction in the entropy as we sequentially move from the first base learner to the last base learner of the ensemble further indicates the significance of introducing uncertainty awareness into the ensemble.
- Further improves the state-of-the-art on the the benchmark ADReSS dataset, thus making important contributions in the severity prediction of Alzheimer's Dementia.