Ming Hsieh Department Modeling, Analysis and Design of Ensembles of

of Electrical Engineering **Multiple Diverse Experts with Signal Processing**



School of Engineering

Applications

Kartik Audhkhasi, Shrikanth S. Narayanan

audhkhas@usc.edu, shri@sipi.usc.edu

Signal Analysis and Interpretation Lab (SAIL)

EXPERT ENSEMBLES ARE EVERYWHERE



• **MODELING:** What is a **realistic model** for ensembles?

BAYES OPTIMAL FUSION RULE FOR GVLC MODEL

• Simple plurality fusion: $\hat{y} = \arg \max_{y} \sum y_{k}^{j}$

where
$$y_k^j = 1$$
 if $y^j = k$ and 0 otherwise

• Variable expert reliability model:

$$\hat{y}_{MAP} = \arg\max_{y} \left[\log\pi_{k} + \sum_{j=1}^{R}\sum_{k_{1}=1}^{K}y_{k}^{j}\log A^{j}(k_{1},k)\right]$$

• GVLC model-based fusion:

$$\hat{y}_{MAP} = \arg\max_{y} \sum_{m=1}^{M} \left[P(z=m)P(x \mid z=m; \Theta_m) P(y=k \mid x; \Lambda) \right]$$

$$\frac{R}{14} \left[\frac{K}{14} A_{j} \left(L - L_{j} \right)^{y_{k_1}^j} \right]$$

- **ANALYSIS:** What is the right decision **fusion rule**?
- **DESIGN:** How to pick the right set of experts **diversity**?

GLOBALLY VARIANT LOCALLY CONSTANT MODEL

- Classical assumption all experts are equally reliable
- Better each expert has different reliability



• Even better – expert reliability varies with data as well





Non-linear fusion rule, other rules special cases

CREATING A DIVERSE MAXENT ENSEMBLE

Collaborators – Abhinav Sethy, Bhuvana Ramabhadran, IBM T. J. Watson Research Lab

"You don't understand anything until you learn it more than one way" – Marvin Minsky

• **Common** view – **ad-hoc training** of diverse models Bagging/boosting – no control on diversity

• Train diverse models in a **principled way with control?**

$$P_{\Lambda}(y \mid x) = \frac{\exp(\sum_{i=1}^{F} \lambda_i f_i(x, y))}{\sum_{y \in Y} \exp(\sum_{i=1}^{F} \lambda_i f_i(x, y))} = \frac{\exp(\sum_{i=1}^{F} \lambda_i f_i(x, y))}{Z(x)}$$

x - input, y - output label, $\{\lambda_i\}_{i=1}^F$ - parameters

• ML estimation:

$$\Lambda^* = \arg \max_{\Lambda} \sum_{n=1}^{N} \left(\sum_{i=1}^{F} \lambda_i f_i(x_n, y_n) - \log Z(x_n) \right) = \arg \max_{\Lambda} L(\Lambda)$$

Reliability constant over homogeneous instances

- Parameter estimation EM with state and true label as hidden variables
- Bayesian version with parameter priors
- Applied to emotion recognition, UCI databases

[1] K. Audhkhasi and S. S. Narayanan, "Data-dependent evaluator modeling and its application to emotional valence classification from speech", Proc. Interspeech, 2010. [2] K. Audhkhasi et al., "Creating ensemble of diverse maximum entropy models", Proc. ICASSP, 2012.

 Introduce diversity constraint w.r.t. reference model Q: $L_{tot}(\Lambda) = L(\Lambda) + \alpha D(P_{\Lambda}, Q_{\Lambda'})$

• Intuitive diversity scores – KL divergence and negative posterior cross-correlation



IN PROGRESS

- Training arbitrary diverse classifier ensembles
- Quantifying diversity benefit in arbitrary expert ensembles

Ming Hsieh Institute

Ming Hsieh Department of Electrical Engineering

• Linking diversity with gradient boosting

This research was supported by NSF, DARPA and NIH