

Personalizing Gesture Recognition Using Hierarchical Bayesian Neural Networks

Ajjen Joshi¹, Soumya Ghosh², Margrit Betke¹, Stan Sclaroff¹, Hanspeter Pfister³

¹Boston University, ²IBM T.J. Watson Research Center, ³Harvard University

Problem Statement

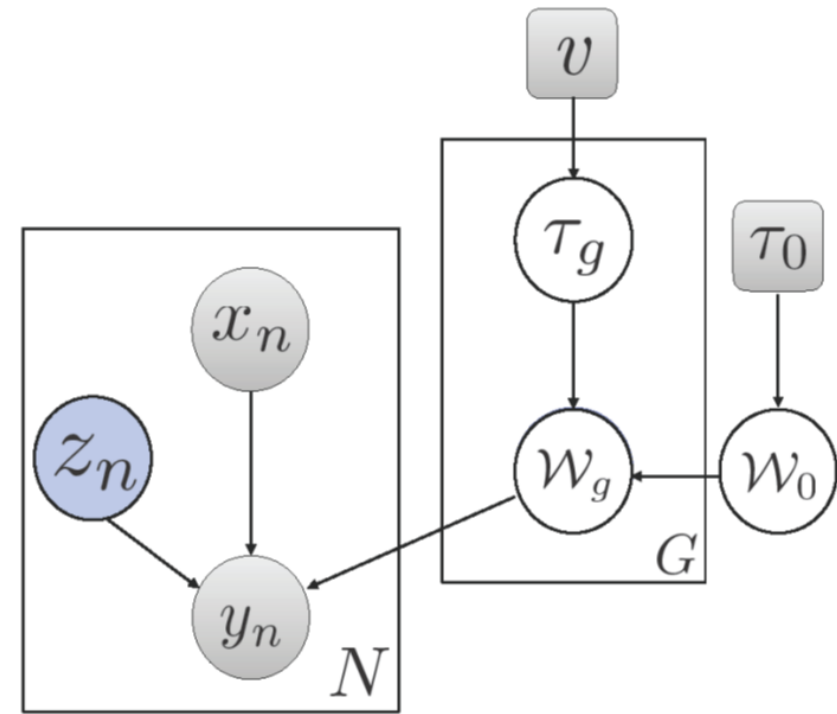
- Building robust classifiers trained on data susceptible to group-specific variations is a challenging problem.
- We develop hierarchical Bayesian neural networks to capture subject-specific variations in gesture performance.
- We propose methods for adapting our model to new subjects and investigate an active learning mechanism to improve personalization in resource-constrained scenarios.

Inference

- We approximate the intractable posterior with a fully factorized approximation,

$$q(\mathcal{W}_0, \mathcal{W}, \mathcal{T} | \phi) = q(\mathcal{W}_0 | \phi_0) \prod_{g=1}^G q(\mathcal{W}_g | \phi_g) q(\tau_g^{-1/2} | \phi_{\tau_g})$$
 - The ELBO is then maximized with respect to the variational parameters using black-box variational Bayes.
- $$\mathcal{L}(\phi) = \mathbb{E}_{q_\phi} [\ln p(\mathcal{W}_0, \mathcal{W}, \mathcal{T}, \mathbf{y} | \mathbf{x}, \mathbf{z}, \tau_0, v)] - \mathbb{E}_{q_\phi} [\ln q(\mathcal{W}_0, \mathcal{W}, \mathcal{T} | \phi)]$$
- In computing the Monte Carlo estimate of the gradients, we use the local reparameterization trick.
 - Unobserved group memberships of held-out data are inferred via a separate inference network.

Hierarchical Bayesian Neural Networks



- Given a dataset $\mathcal{D} = \{x_n, y_n\}_{n=1}^N$ each subject is endowed with its own conditional distribution $p(y_n | z_n = g, f(x_n, \mathcal{W}_g))$.

$$p(\mathcal{W}_g | \mathcal{W}_0, \tau_g) = \prod_{l=1}^L \prod_{i=1}^{V_{l-1}} \prod_{j=1}^{V_l} \mathcal{N}(w_{ij,l}^g | w_{ij,l}^0, \tau_g^{-1})$$

$$p(\mathcal{W}_0 | \tau_0) = \prod_{l=1}^L \prod_{i=1}^{V_{l-1}} \prod_{j=1}^{V_l} \mathcal{N}(w_{ij,l}^0 | 0, \tau_0^{-1})$$

$$p(\gamma_g | v) = \mathcal{N}(\gamma_g | 0, v); \quad \tau_g^{-1/2} = |\gamma_g|$$

- The joint distribution is given by:

$$p(\mathcal{W}_0, \mathcal{W}, \mathcal{T}, \mathbf{y} | \mathbf{x}, \mathbf{z}, \tau_0, v) = p(\mathcal{W}_0 | \tau_0^{-1}) \prod_{g=1}^G p(\gamma_g | v) p(\mathcal{W}_g | \mathcal{W}_0, \tau_g^{-1}) \prod_{n=1}^N \prod_{g=1}^G p(y_n | f(\mathcal{W}_g, x_n))^{1[z_n=g]}$$

Personalization

- $\{\mathcal{W}_g\}_{g=1}^{G+1}$ are conditionally independent given $\mathcal{W}_0$.
 - Given a model trained on $\mathcal{D}$, we only update $\mathcal{W}_{G+1}$ while keeping everything else fixed.
 - To best utilize limited labeling resources, we adopt the Bayesian Active Learning by Disagreement (BALD) algorithm to adaptively select training instances for the new group.
- $$x_l = \operatorname{argmax}_{x \in X_{pool}} \mathbb{H}[y | x, \mathcal{D}] - \mathbb{E}_{\mathcal{W}_g \sim p(\mathcal{W}_g | \mathcal{D})} \mathbb{H}[y | x, \mathcal{W}_g]$$

Results

We test our method on 3 gesture datasets: MSRC-12, ChaLearn 2013 and NATOPS

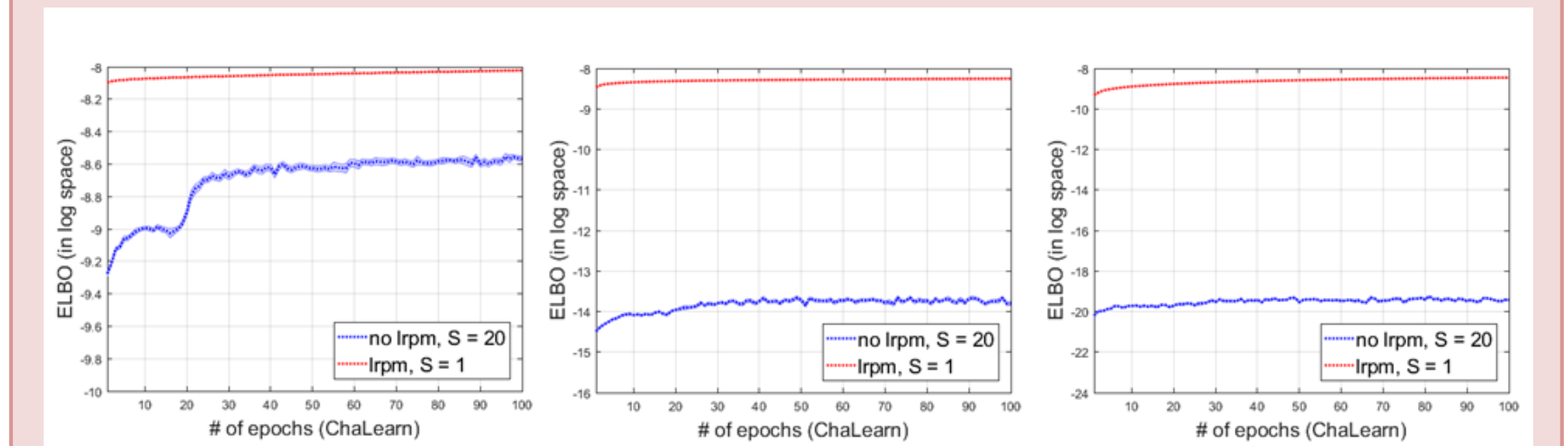


MSRC-12

ChaLearn 2013

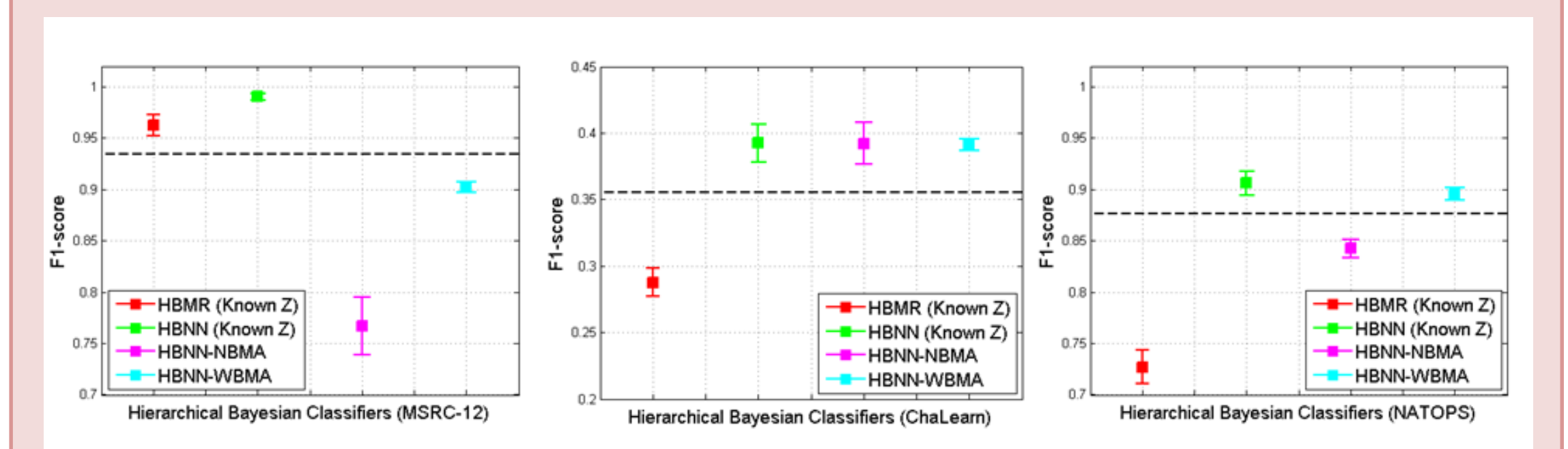
NATOPS

1. Benefits of Local Reparameterization



lrpm: local reparameterization

2. Gesture Recognition



HBMR: Hierarchical Bayesian Multinomial Regression model
HBNN: Hierarchical Bayesian Neural Network model with 2 hidden layers

3. Personalization

