Incorporating functional summary information in Bayesian neural networks using a Dirichlet process likelihood approach

Vishnu Raj, Tianyu Cui, Markus Heinonen & Pekka Marttinen

Aalto University, Finland

Overview

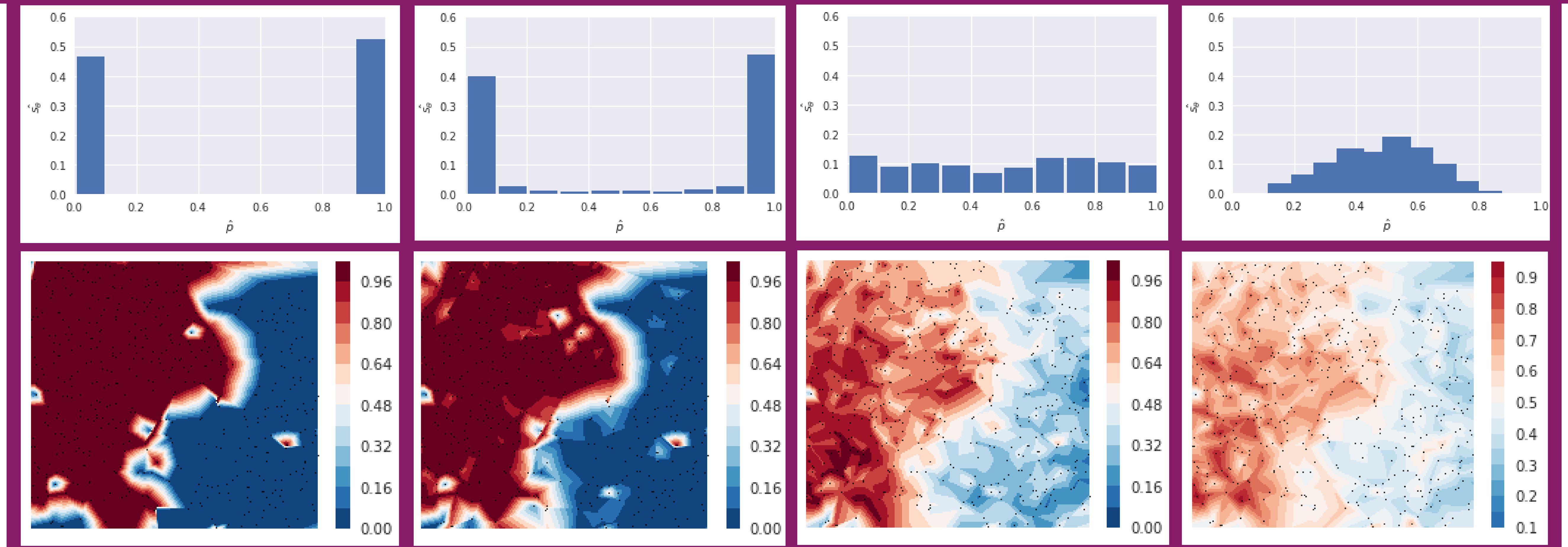
Problem: Prior specified over parameters of large Bayesian Neural Networks rarely reflects true prior knowledge. We present a simple approach to incorporate prior knowledge in BNNs based on external summary information.

Approach: Our approach augment the training data with summary information about the distribution of predicted scores, and models it using Dirichlet Process. We derive Summary-ELBO to train BNNs using Variational Inference.

Results: Experiments show improved accuracy, uncertainty calibration and robustness against corruption on multiple tasks

Key takeaways

1. Incorporating summary information can improve prediction accuracy as well as calibration
2. When we have multiple prior information, we should consider all except one of them as data and model the joint likelihood
3. The proposed summary likelihood modeling can be used in other training settings such as MCMC methods or even with deterministic neural networks



Regular BNN | s_0 = Beta(0.1, 0.1) | s_0 = Beta(1.0, 1.0) | s_0 = Beta(5.0, 5.0)

Posterior distribution of predicted sigmoid scores with different summary statistic observations s_0 in binary classification in MNIST. Regular BNN predicts scores peaked towards 0 or 1, indicating possible overconfidence. By introducing a likelihood term for the summary statistic, the proposed Summary ELBO is able to control how the predicted sigmoid scores are distributed. This is also evident in the decision surface; the regular BNN has a sharp decision boundary with extreme predicted values while the Summary ELBO yields a smoother decision surface.

Approach

We augment training data **(X, Y)** with summary information $\mathbf{s_0}$

$$p(\mathbf{X},\mathbf{Y},\mathbf{s}_0,\theta) = \left[\prod_{i=1}^{N} p(\mathbf{y}_i|\mathbf{x}_i,\theta)\right] \underbrace{p(\mathbf{s}_0|\mathbf{X},\theta)}_{\text{Likelihood of summary information}} p(\theta)$$

Summary information computed with model parameters

$$p(\mathbf{s}_0|\mathbf{X},\theta) = \mathcal{DP}(\mathbf{s}_0|\mathbf{s}_\theta,\alpha).$$

Summary information provided by domain expert

Summary ELBO

$$\mathcal{L}(\phi) = \mathbb{E}_{q_\phi(\theta)} \log p(\mathcal{D}|\theta)) - \mathrm{KL}[q_\phi(\theta))||p(\theta))]$$

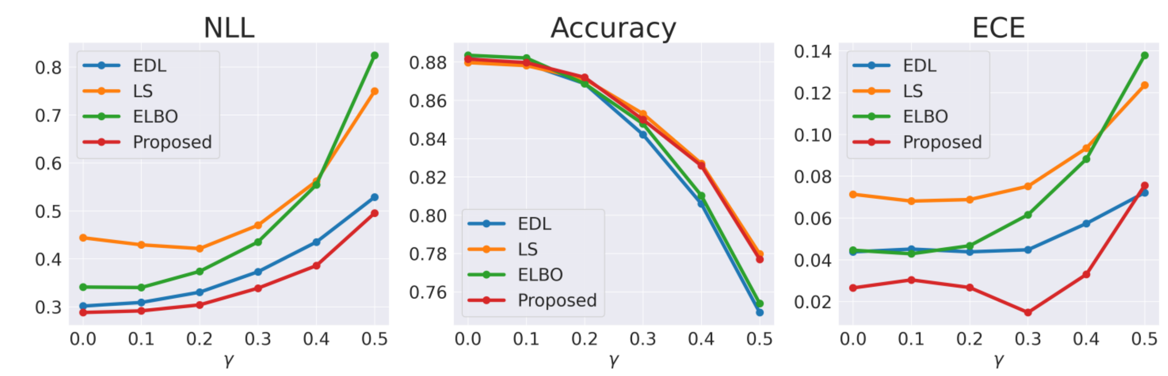
$$\approx \frac{1}{M}\sum_{i=1}^{N}\sum_{j=1}^{M} \log \mathrm{Cat}\left(\mathbf{y}_i|f(\mathbf{x}_i,\theta_j)\right)$$

$$+ \boxed{\frac{1}{M}\sum_{j=1}^{M} \log \mathcal{DP}\left(\mathbf{s}_0|\mathbf{s}_{\theta_j},\alpha\right)} - \mathrm{KL}\left[\mathcal{N}_\theta(\mu_\phi,\mathbf{\Sigma}_\phi)\,||\,\mathcal{N}_\theta(\mathbf{0},\sigma^2 I)\right]$$

Incorporating summary information into optimization objective

Experiments

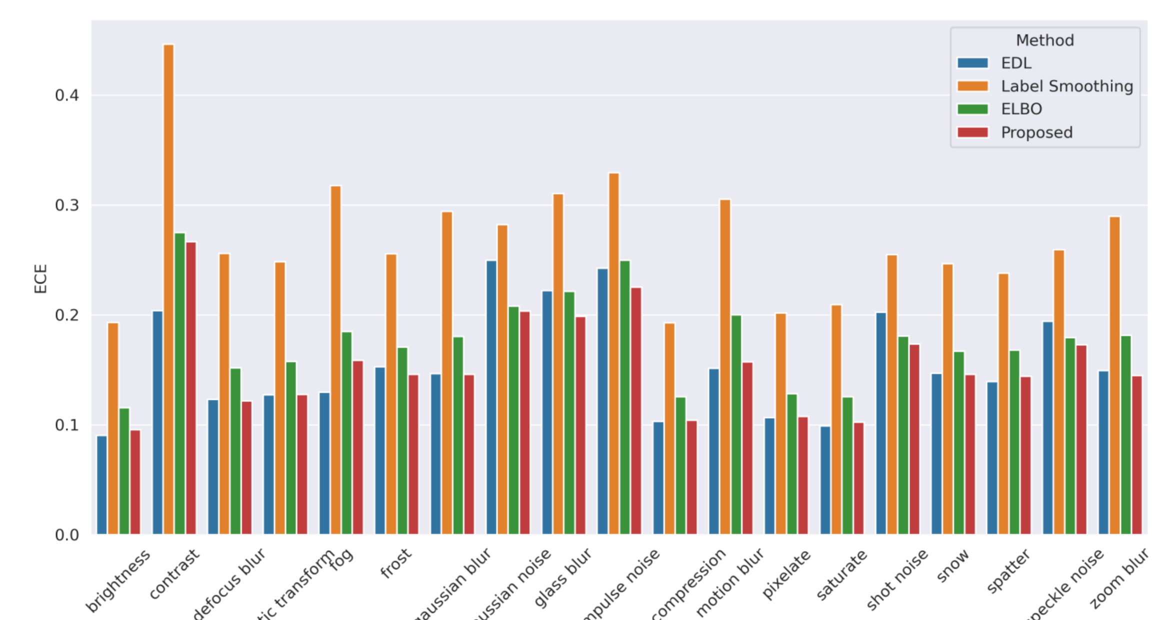
1. Sentiment Analysis task



Method	NLL↓	ECE↓
ELBO	0.341 ± 0.024	0.045 ± 0.009
LS($\epsilon = 0.05$)	0.444 ± 0.029	0.071 ± 0.004
EDL	0.301 ± 0.001	0.044 ± 0.004
Proposed ($\alpha = 1000$)	$\mathbf{0.288 \pm 0.002}$	$\mathbf{0.026 \pm 0.001}$

2. Image Classification

	In-domain testset		
Method	NLL↓	Accuracy↑	ECE↓
ELBO	0.76 ± 0.01	$\mathbf{0.82 \pm 0.00}$	0.10 ± 0.00
LS	1.93 ± 0.02	0.79 ± 0.00	0.17 ± 0.00
EDL	0.78 ± 0.01	$\mathbf{0.82 \pm 0.00}$	$\mathbf{0.08 \pm 0.00}$
Proposed	$\mathbf{0.68 \pm 0.01}$	$\mathbf{0.82 \pm 0.00}$	$\mathbf{0.08 \pm 0.00}$



This work was supported by the Academy of Finland (Flagship programme: Finnish Center for Artificial Intelligence FCAI, and grants 336033, 352986) and EU (H2020 grant 101016775 and NextGenerationEU).