Sparse Partial Bayesian Networks: Efficient Uncertainty Quantification in Medical Image Analysis

Efficiently quantifying predictive uncertainty in medical image analysis is essential for improving the reliability of AI-driven diagnostic systems. Bayesian Neural Networks (BNNs) are effective due to their probabilistic approach, offering uncertainty estimates alongside predictions. However, BNNs are hindered by significant computational demands, especially during training, making them impractical for large-scale and real-time medical applications. Although Bayesian approximations such as ensembles have shown promise, they still suffer from high training costs. Existing approaches to reducing computational burden primarily focus on lowering the costs of BNN inference, with limited efforts to improve training efficiency and minimize parameter complexity. This paper introduces a Sparse Bayesian Network that selectively incorporates Bayesian parameters based on first-order gradient sensitivity analysis, drastically reducing the number of Bayesian parameters while maintaining model performance and reducing training costs.

The proposed methodology involves a three-step process (demonstrated in the figure below): (1) Training a deterministic model for the given task; (2) Performing a gradient-based sensitivity analysis to identify the most sensitive parameters; (3) Training the model as a sparse variational Bayesian network (SBNN) by converting the Topk sensitive parameters to Bayesian, while the rest remain deterministic.

The effectiveness of this approach is demonstrated on multi-label classification with the ChestMNIST dataset and segmentation tasks using the ISIC and LIDC-IDRI datasets. In these experiments, the SBNN consistently reduced the Bayesian parameter count by over 95% compared to fully Bayesian models. Despite the drastic reduction in Bayesian parameters, the SBNN maintained competitive classification accuracy, uncertainty (Brier Score) and segmentation quality performance compared to ensembles. The significance of the SBNN approach lies in its ability to maintain task performance and match the predictive uncertainty of ensembles while requiring 80% fewer parameters.

The primary contribution of this work is a cost-effective recipe for enabling uncertainty quantification in DNNs without the prohibitive computational costs typically associated with BNNs and ensembles. By demonstrating that an SBNN approach can achieve similar or better performance with a fraction of the parameters, this work paves the way for more scalable and efficient uncertainty quantification for safety-critical applications.

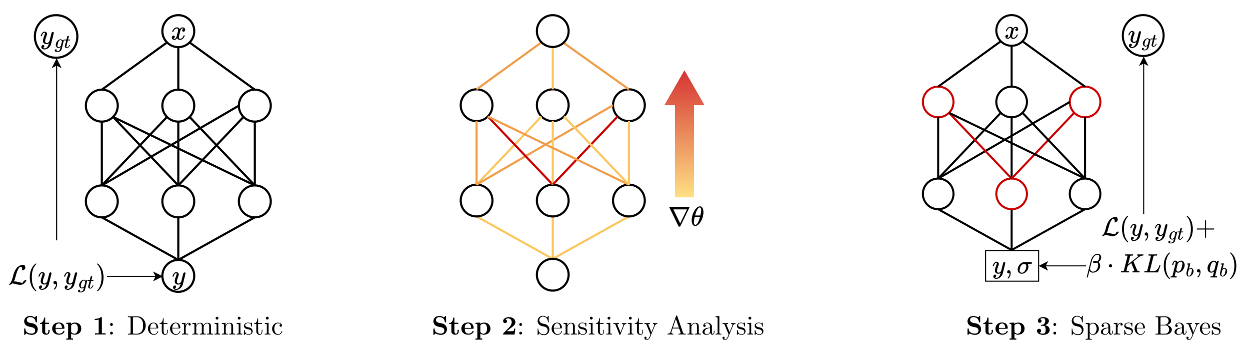

Figure 1: Our proposed training of sparse (partial) Bayesian network. **Step 1**: Train a deterministic model by minimizing the negative log likelihood $\mathcal{L}(y, y_{gt})$ where the parameters are represented as point estimates. **Step 2**: Perform a gradient-based sensitivity analysis, denoted as $\nabla\theta$, and identify the Topk parameters corresponding to the highest magnitude gradients (in red). **Step 3**: Train a sparse (partial) Bayesian model with the Topk parameters as variational Bayesian parameters and the remaining network as deterministic by minimizing the evidence lower bound (ELBO) loss $\mathcal{L}(y, y_{gt}) + \beta \cdot D_{KL}\left(q(\theta_b), p(\theta_b)\right)$, where $p(\theta_b)$ and $q(\theta_b)$ are the prior and posterior distributions for the θ_b Bayesian parameters.