An Empirical Study of Prior-Data Conflicts in Bayesian Neural Networks

Alexander Marquard A.Marquard@campus.lmu.de
Julian Rodemann Julian.Rodemann@stat.uni-muenchen.de
Thomas Augustin thomas.augustin@stat.uni-muenchen.de
Department of Statistics, LMU Munich, Germany

Imprecise Probabilities (IP) allow for the representation of incomplete information. In the context of Bayesian statistics, this is achieved by generalized Bayesian inference, where a set of priors is used instead of a single prior [1, Chapter 7.4]. The latter has been shown to be particularly useful in the case of prior-data conflict, where evidence from data (likelihood) contradicts prior information. In these practically highly relevant scenarios, classical (precise) probability models typically fail to adequately represent the uncertainty arising from this conflict. Generalized Bayesian inference by IP, however, was proven to handle these prior-data conflicts well when inference in canonical exponential families is considered [3].

Our study [2] aims at accessing the extent to which these problems of precise probability models are also present in Bayesian neural networks (BNNs). Unlike traditional neural networks, BNNs utilize stochastic weights that can be learned by updating the prior belief with the likelihood for each individual weight using Bayes' rule. In light of this, we investigate the impact of prior selection on the posterior of BNNs in the context of prior-data conflict. While the literature often advocates for the use of normal priors centered around 0, the consequences of this choice remain unknown when the data suggests high values for the individual weights. For this purpose, we designed synthetic datasets which were generated using neural networks (NN) with fixed high-weight values. This approach enables us to measure the effect of prior-data conflict, as well as reduce the model uncertainty by knowing the exact weights and functional relationship. We utilized BNNs that use the Mean-Field Variational Inference (MFVI) approach, which has not only seen an increasing interest due to its scalability but also allows analytical computation of the posterior distributions, as opposed to simulation-based methods like Markov Chain Monte Carlo (MCMC). In MFVI, the posterior distribution is approximated by a tractable distribution with a factorized form.

In our work [2, Chapter 4.2], we provide evidence that exact priors centered around the exact weights, which are known from the neural network (NN), outperform their inexact counterparts centered around zero in terms of predictive accuracy, data efficiency and reasonable uncertainty estimations. These results directly imply that selecting a prior centered around 0 may be unintentionally informative, as previously noted by [4], resulting in significant losses in prediction accuracy and data requirement, rendering uncertainty estimation impractical. BNNs learned under prior-data conflict resulted in posterior means that were a weighted average of the prior mean and the likelihood highest probability values and therefore exhibited significant differences from the correct weights while also exhibiting an unreasonably low posterior variance, indicating a high degree of certainty in their estimates. Varying the prior variance yielded similar observations, with models using priors with data conflict exhibiting overconfidence in their posterior estimates compared to those using exact priors.

To investigate the potential of IP methods, we are currently conducting the effect of expectation- valued interval-parameter, to generate resonable uncertainty predictions. Overall, our preliminary results show that classical BNNs produce overly confident but erroneous predictions in the presence of prior-data conflict. These findings motivate using IP methods in Deep Learning.

References

[1] Thomas Augustin, Gero Walter, and Frank Coolen. Statistical inference. In T. Augustin, F. Coolen, G. de Cooman, and M. Troffaes, editors, *Introduction to Imprecise Probabilities*, pages 135–189. Wiley, 2014.

[2] Alexander Marquard. Eine empirische Analyse von Prior-Daten Konflikten in Bayesianischen neuronalen Netzen. Bachelor's Thesis, Ludwig-Maximilians-Universität München, 2023.

[3] Gero Walter and Thomas Augustin. Imprecision and prior-data conflict in generalized bayesian inference. *Journal of Statistical Theory and Practice*, 3(1):255–271, 2009.

[4] Florian Wenzel, Kevin Roth, Bastiaan S Veeling, Jakub Świątkowski, Linh Tran, Stephan Mandt, Jasper Snoek, Tim Salimans, Rodolphe Jenatton, and Sebastian Nowozin. How good is the Bayes posterior in deep neural networks really? *arXiv preprint arXiv:2002.02405*, 2020.