Uncertainty in AI: A Variance-Based Approach

Ruxiao Duan[1], Brian Caffo[2], Harrison Bai[3], Haris Sair[3], and Craig Jones[3,4,5]

[1] *Department of Applied Mathematics and Statistics, Johns Hopkins University*
[2] *Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health*
[3] *Department of Radiology and Radiological Science, Johns Hopkins University School of Medicine*
[4] *Department of Computer Science, Johns Hopkins University*
[5] *Malone Center for Engineering in Healthcare, Johns Hopkins University*

Evidential deep learning [3] has been designed to quantify the predictive uncertainty of deep learning models with a single deterministic network. By assuming the prediction target follows a bi-level probability distribution, the evidential neural network predicts the parameters of the higher-order evidential distribution, from which aleatoric uncertainty and epistemic uncertainty of the prediction can be derived.

Traditionally, the classification uncertainty of a sample is quantified as the Shannon entropy of the expected class probabilities [2]. However, this entropy-based approach can only measure the total uncertainty of a sample, namely the sample-level uncertainty. The individual class-level uncertainties are unfortunately lost. That means the entropy-based approach does not know which classes contribute more to the uncertainty of a sample.

Inspired by deep evidential regression [1], we propose an alternative variance-based approach to calculate the evidential uncertainties of a classification sample. The class prediction is considered as a one-hot label vector, whose covariance matrix can be derived from the distribution assumptions. From the diagonal of the covariance matrix, class-level variance can be obtained, which is regarded as the uncertainty of each class. Based on the law of total covariance, the covariance matrix of the class prediction can be decomposed into an aleatoric component and an epistemic component, the diagonals of which provide class-level aleatoric and epistemic uncertainties. In addition, the class correlation matrix can also be computed from the covariance matrix, giving the between-class correlations of the classification sample.

We evaluate our variance-based uncertainty quantification approach on a downstream application known as active domain adaptation. By optimizing the model with the variance-based evidential uncertainties of target domain data, the domain shift is greatly reduced, significantly improving the domain adaptation performance. In active sampling, the uncertainties also help to accurately identify samples with low confidence, resulting in the state-of-the-art active learning performance on Office-Home dataset. Comparing our variance-based with the previous entropy-based approach, our method not only quantifies sample-level uncertainties but also derives class-level uncertainties and between-class correlations, which the entropy-based approach cannot provide.

References

[1] Alexander Amini, Wilko Schwarting, Ava Soleimany, and Daniela Rus. Deep evidential regression. *Advances in Neural Information Processing Systems*, 33:14927–14937, 2020.

[2] Andrey Malinin and Mark Gales. Predictive uncertainty estimation via prior networks. *Advances in neural information processing systems*, 31, 2018.

[3] Murat Sensoy, Lance Kaplan, and Melih Kandemir. Evidential deep learning to quantify classification uncertainty. *Advances in neural information processing systems*, 31, 2018.