Separation of Aleatoric and Epistemic Uncertainty in Deterministic Deep Neural Networks

Denis Huseljic, Bernhard Sick, Marek Herde, Daniel Kottke
{dhuseljic | bsick | marek.herde | daniel.kottke}@uni-kassel.de
Intelligent Embedded Systems Group, University of Kassel, Germany

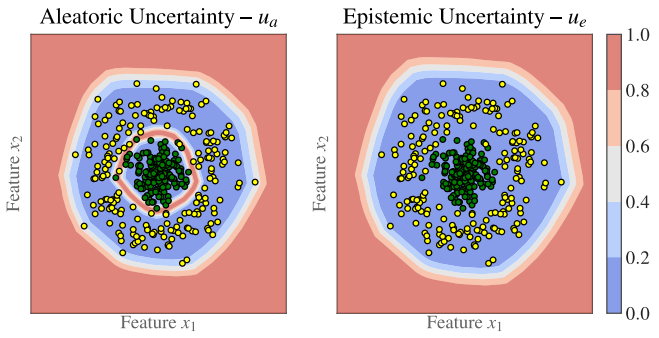
Motivation

Situation: Standard Deep Neural Networks (DNN) lack in the estimation of uncertainty associated with predictions.

Problem: In autonomous driving, for example, predictions must be reliable to avoid life-threatening situations.

Task: Provide models that measure epistemic uncertainty that describes the reliability of a prediction, and aleatoric uncertainty which describes the risk of a predicted class while maintaining proper generalization capabilities.

Approach: AE-DNN, a method for separating **Aleatoric** and **Epistemic** uncertainty in **DNN**.



(a) Low epistemic uncertainty, low aleatoric uncertainty: ideal as predictions are reliable.
(b) Low epistemic uncertainty, high aleatoric uncertainty: predictions are reliable, but not distinct (e.g., due to sensor noise).
(c) High epistemic uncertainty: predictions are unreliable (aleatoric uncertainty can be disregarded).

Novelty

AE-DNN combines two objective functions – the first optimizing on in-distribution (ID) samples, the second on out-of-distribution (OOD) samples – into one by means of a convex combination.

Characteristics:

- At run-time, AE-DNN allows for a detection of samples that were never seen during training (referred to as OOD detection) or an estimate of the risk coming with a decision.
- For computational efficiency, the inference avoids multiple forward passes through DNN as needed in ensembles, for instance.
- The inference is deterministic in the sense that the same input always leads to the same output (in contrast to Bayesian NN or Monte Carlo dropout).
- OOD samples are generated by means of Generative Adversarial Networks (GAN). As a result, AE-DNN does not require explicitly provided OOD data sets.
- The hyperparameter within the convex combination (OOD vs. ID) allows for a control of the degree of desired certainty in a concrete application.

A comparison to related techniques for uncertainty modeling in DNN is given below. We denote optimization criteria and mark benefits and flaws by $+$ and $-$.

	Epistemic Uncer.	Aleatoric Uncer.	Inference Time	Training Time	ID Optim. Criterion	OOD Optim. Criterion
Ordinary	$-$	$+$	$++$	$++$	MLE	N/A
Ensembles	$+$	$++$	$-$	$--$	MLE	N/A
Dropout	$+$	$+$	$--$	$+$	MLE	N/A
EDL	$--$	$--$	$++$	$++$	Bayes-risk + KL	N/A
PN	$++$	$--$	$++$	$+$	KL	KL
AE-DNN	$++$	$+$	$++$	$+$	MLE	KL

The most similar approach to AE-DNN is Prior Networks (PN) which differs regarding the ID optimization criterion (Kullback-Leibler (KL) in PN and Maximum Likelihood (MLE) in AE-DNN).

Method

Idea: Inspired by the ideas of (1) using a Dirichlet distribution as the target distribution to be optimized and (2) the ability of GAN to generate OOD samples, we propose AE-DNN which combines both ideas and allows for an intuitively understandable separation of aleatoric and epistemic uncertainty.

The optimization of the model parameters in our approach is based on a convex-combination of two different, but complementary objective functions. We can summarize our method as follows:

A. For ID samples, we optimize the parameters of the DNN such that its output defines a Dirichlet-Categorical distribution over the classes.
B. For OOD samples, we optimize the parameters of the DNN to enforce the Dirichlet distribution over the class probabilities, which is part of the above Dirichlet-Categorical distribution, to be a uniform distribution over the simplex of possible values.
C. Since we obtain a Dirichlet distribution for every sample, we can derive measures describing the heteroscedastic aleatoric and epistemic uncertainty.

Uncertainty Measures: Based on the Dirichlet parameters estimated by the output of our DNN, we are able to define separate measures for aleatoric and epistemic uncertainty, both yielding values in the unit interval:

The *aleatoric uncertainty* $u_a \in [0,1]$ of a sample $\mathbf{x}^*$ is given by

$$u_a = \frac{\mathbb{H}\left[\mathbf{y}^* | \boldsymbol{\alpha} = \mathbf{f}^{\boldsymbol{\omega}}(\mathbf{x}^*) + \mathbf{1}\right]}{\ln K},$$

where $\mathbb{H}[\cdot]$ denotes the entropy, $\mathbf{f}^{\boldsymbol{\omega}}(\mathbf{x}^*)$ is the model output for sample $\mathbf{x}^*$, K is the number of classes, and the one-hot-encoded label $\mathbf{y}^*$ is distributed according a Dirichlet-Categorical distribution. Intuitively, the aleatoric uncertainty u_a tends to be high if the class probabilities derived from the Dirichlet-Categorical are close to a uniform distribution (and vice versa).

The *epistemic uncertainty* $u_e \in [0,1]$ of a sample $\mathbf{x}^*$ is given by

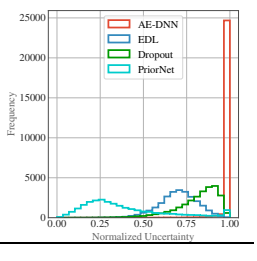
$$u_e = \frac{K}{\left\|\mathbf{f}^{\boldsymbol{\omega}}(\mathbf{x}^*) + \mathbf{1}\right\|_1}.$$

It approaches one if the output $\mathbf{f}^{\boldsymbol{\omega}}(\mathbf{x}^*)$ is small (i.e., has a small 1-norm). In contrast, if the model's output $\mathbf{f}^{\boldsymbol{\omega}}(\mathbf{x}^*)$ has a high 1-norm, the epistemic uncertainty becomes small (i.e., approaches 0). That is, the epistemic uncertainty is high for samples that are not represented in the training dstirbution.

Experimental Evaluation

We evaluate our method on different image data sets by considering train, validation, and test splits. To evaluate epistemic and aleatoric uncertainty, we use common measures such as the *Area Under Receiver Operating Characteristic Curve* (AUROC), *Uncertainty Histogram* (UH), *Expected Calibration Error* (ECE), *Negative Log-Likelihood* (NLL), and *Brier Score* (BS). Since the generalization of a DNN is of crucial importance, we additionally report the accuracy.

The table below summarizes exemplary the results for SVHN (as ID) vs CIFAR10 (as OOD). For further details and additional results, we refer to our implementation which is available at `https://github.com/hsljc/ae-dnn`.

Data Sets	Methods	Generalization	Aleatoric Uncertainty			Epistemic Uncertainty	
		Accuracy (↑)	ECE (↓)	NLL (↓)	BS (↓)	AUROC (↑)	UH
SVHN vs. CIFAR10	Ordinary	0.875±0.009	0.012±0.010	0.440±0.031	0.018±0.001	0.850±0.028	
	Ensembles	0.900±0.004	0.046±0.004	0.361±0.015	0.015±0.001	0.913±0.004	
	Dropout	0.881±0.010	0.015±0.008	0.400±0.026	0.017±0.001	0.921±0.009	
	EDL	0.196±0.000	0.089±0.007	2.291±0.011	0.090±0.000	0.615±0.017	
	PN (OOD gen.)	0.840±0.038	0.107±0.036	0.589±0.134	0.025±0.006	0.933±0.046	
	AE-DNN (OOD gen.)	0.859±0.014	0.014±0.009	0.485±0.038	0.021±0.002	0.970±0.017	
	PN (OOD av.)	0.882±0.009	0.101±0.031	0.468±0.032	0.019±0.001	0.993±0.002	
	AE-DNN (OOD av.)	0.879±0.011	0.019±0.005	0.427±0.033	0.018±0.001	0.997±0.001	