Artificial Intelligence in the Life Sciences 2 (2022) 100033

Contents lists available at ScienceDirect

Artificial Intelligence in the Life Sciences

journal homepage: www.elsevier.com/locate/ailsci

Understanding uncertainty in deep learning builds confidence

Jürgen Bajorath

Department of Life Science Informatics and Data Science, B-IT, LIMES Program Unit Chemical Biology and Medicinal Chemistry, Rheinische Friedrich-Wilhelms-Universität, Friedrich-Hirzebruch-Allee 5/6, D-53115 Bonn, Germany

ARTICLE INFO

Keywords:
Machine learning
Deep learning
Predictions
Uncertainty quantification

In a contribution to Volume 1 of *AILSCI*, Lazic & Williams address the issue of uncertainty in machine learning (ML) [1] that has so far only been little considered in interdisciplinary research and drug discovery. In standard ML for compound classification or regression, only a single output value is produced for a test instance, with no additional information concerning the confidence of the prediction or the level of uncertainty associated with it.

Assessing the confidence or uncertainty of predictions adds another layer of information to ML that becomes particular important for judging its results in interdisciplinary settings. Moreover, if ML supports clinical decisions such as the prioritization of treatment strategies –one of the for AI in medicine– uncertainty assessment becomes essential. Hence, going forward, quantifying the uncertainty of predictions is an important topic for ML and especially deep learning (DL). Together with approaches for rationalizing ML/DL decisions, i.e., interpretable or explainable AI (XAI), uncertainty information also aids in model interpretation, decreases the black box character of ML/DL, and increases its acceptance in interdisciplinary research settings [2–5].

In ML/DL, two different categories of uncertainty are distinguished including epistemic and aleatoric uncertainty, which result from model-inherent factors and data variance (including experimental inaccuracies), respectively. Epistemic and aleatoric uncertainty can be separately evaluated. For method development, model-dependent uncertainty provides a natural focal point. There are different ways in which uncertainty of ML/DL predictions can be estimated, as further discussed in the following. In their thoughtfully designed and well presented analysis, Lazic & Williams concentrate on probabilistic modeling and highlight the information that is gained when single output values of models are replaced by probability distributions, providing immediate access to uncertainty of individual predictions and making it possible to quantify potential errors (see Fig. 1 in [1] for an instructive example). Furthermore, the authors discuss different origins of model-based uncertainty. Another characteristic of the paper by Lazic & Williams is that it is written for both experts and non-experts in ML or statistics, making it a must-read for experimental investigators interested in ML/DL for their projects.

For uncertainty quantification, Bayesian deep neural networks (DNNs) are particularly attractive [1,6–8] because they produce the complete posterior probability distribution for predictions. However, the general use of full Bayesian DNNs is restricted by high computational demands for larger data sets. Of note, some ML approaches making use of value distributions such as Gaussian process modeling [9,10], one of the preferred methods for regression tasks, provide inherent uncertainty estimates for their predictions.

Bayesian DNNs mark one end of the spectrum of ML methods yielding uncertainty information that range from probabilistic approaches of varying complexity to ensemble methods. Fig. 1 illustrates different types of approaches falling into this methodological spectrum [11,12]. Ensemble methods quantify deviations in predictions between different versions of the same ML/DL model, preferably trained on bootstrapped data, to derive statistical measures of model uncertainty [13]. As such, computational costs of ensemble assessment also tend to be high. In mean-variance estimation, only the output layer of a DNN is modified to obtain a Gaussian distribution and predict the mean and standard deviation for a given endpoint. This type of modification is also applicable to add a single Bayesian output layer to a DNN, representing a simplification of full Bayesian DNNs with reduced computational demands. Alternatively, computational requirements of Bayesian DNNs might also be reduced through Bayesian inference by using a representative or weighted data subset for learning instead of a complete data sets. In similarity-based approaches, uncertainty of the prediction of a test instance is regarded as inversely proportional to its similarity to training samples. Gaussian process models quantify similarity between samples using kernel functions. Union-based approaches use the output

E-mail address: bajorath@bit.uni-bonn.de

https://doi.org/10.1016/j.ailsci.2022.100033
Received 16 February 2022; Accepted 5 March 2022
Available online 6 March 2022

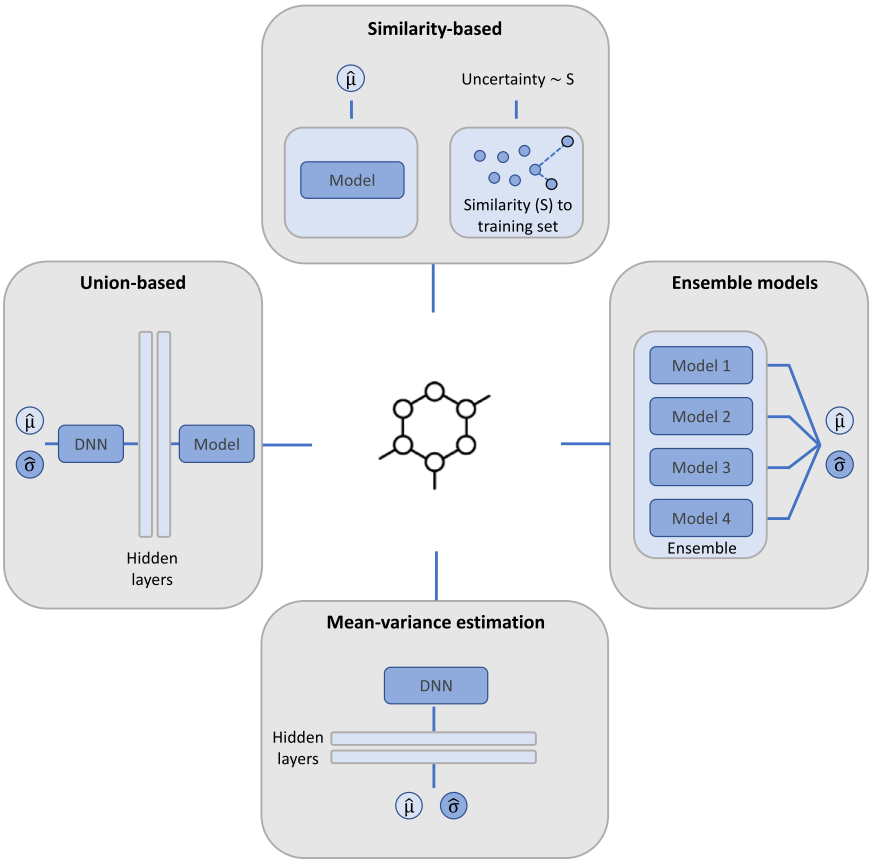

Fig. 1. Uncertainty estimation. Different approaches for quantifying uncertainty of ML/DL predictions are schematically illustrated. The figure was adopted from [11].

of a trained DNN as an input for another ML model to predict uncertainty.

In addition, the conformal prediction framework is an established approach for deriving error bounds when predicting individual instances, without the need for prior probabilities, which differentiates this approach from Bayesian modeling [14]. Conformal predictions are based upon the assumptions of data randomness and exchangeability, which often leads to approximate values [15]. Furthermore, the concept of evidential DL (EDL) [16] has recently been adapted for uncertainty analysis [17]. EDL is related to Bayesian DNNs and derives higher-order distributions for likelihood parameters defining a probability distribution. This so-called evidential distribution is defined by higher-order parameters, which are learned to yield the uncertainty of associated probabilistic predictions.

Hence, taken together, various approaches are available for uncertainty quantification of ML/DL predictions, ranging from established methods such as probabilistic modeling or conformal predictions to promising new concepts such as EDL. However, recent benchmark studies revealed that uncertainties quantified using different methods did not overall accurately correlate with absolute DL regression errors, at least in compound property prediction [12,18]. Relative performance often varied depending on the prediction task. In some cases, DNN ensembles and bootstrapping reached higher performance levels than dropout sampling techniques [18]. In others, message passing DNNs combined with Gaussian process models or message passing DNNs with mean-variance estimation performed best [12]. Furthermore, EDL achieved better correlation between estimated uncertainties and regression errors than ensemble or dropout methods [17].

Clearly, there currently is no consensus regarding the relative performance and consistency of uncertainty quantification methods, and no generally preferred gold standard is available. Moreover, studies that prospectively correlate ML/DL prediction uncertainties with experimental outcomes are currently missing, which provides many opportunities for future investigations with high relevance for interdisciplinary research. Accordingly, high-quality contributions such as the one by Lazic & Williams are important for the field to raise more awareness of the still under-investigated area of ML/DL uncertainty estimation and trigger further studies and prospective applications.

AILSCI explicitly welcomes such investigations to further advance DL across the life sciences and increase its utility for experimental design.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Lazic SE, Williams DP. Quantifying sources of uncertainty in drug discovery predictions with probabilistic models. Artif Intell Life Sci 2021;1:100004.
[2] Abdar M, Pourpanah F, Hussain S, Rezazadegan D, Liu L, Ghavamzadeh M, Fieguth P, Cao X, Khosravi A, Acharya UR, Makarenkov V. A review of uncertainty quantification in deep learning: techniques, applications and challenges. Inf Fusion 2021;76:243–97.
[3] Murdoch WJ, Singh C, Kumbier K, Abbasi-Asl R, Yu B. Definitions, methods, and applications in interpretable machine learning. Proc Natl Acad Sci U S A 2019;116:22071–80.
[4] Linardatos P, Papastefanopoulos V, Kotsiantis S. Explainable AI: a review of machine learning interpretability methods. Entropy 2021;23:18.
[5] Bajorath J. Second-generation artificial intelligence approaches for life science research. Artif Intell Life Sci 2021;1:100026.
[6] Lampinen J, Vehtari A. Bayesian approach for neural networks - review and case studies. Neural Netw 2001;14:257–74.
[7] Ryu S, Kwon Y, Kim WY. A Bayesian graph convolutional network for reliable prediction of molecular properties with uncertainty quantification. Chem Sci 2019;10:8438–46.
[8] Wang H, Yeung DY. A survey on Bayesian deep learning. ACM Comput Surveys 2020;53:1–37.

[9] Deringer VL, Bartók AP, Bernstein N, Wilkins DM, Ceriotti M, Csányi G. Gaussian process regression for materials and molecules. Chem Rev 2021;121:10073–141.
[10] Hie B, Bryson BD, Berger B. Leveraging uncertainty in machine learning accelerates biological discovery and design. Cell Syst 2020;11:461–77.
[11] Miljković F, Rodríguez-Pérez R, Bajorath J. Impact of artificial intelligence on compound discovery, design, and synthesis. ACS Omega 2021;6:33293–9.
[12] Hirschfeld L, Swanson K, Yang K, Barzilay R, Coley CW. Uncertainty quantification using neural networks for molecular property prediction. J Chem Inf Model 2020;60:3770–80.
[13] Lakshminarayanan B, Pritzel A, Blundell C. Simple and scalable predictive uncertainty estimation using deep ensembles. Adv Neural Inf Process Syst 2017;30:6402–13.
[14] Shafer G, Vovk V. A tutorial on conformal prediction. J Mach Learn Res 2008;9:371–421.
[15] Krstajic D. Critical assessment of conformal prediction methods applied in binary classification settings. J Chem Inf Model 2021;61:4823–6.
[16] Sensoy, M., Kaplan, L., Kandemir, M.. Evidential deep learning to quantify classification uncertainty. arXiv preprint arXiv:1806.01768, 2018.
[17] Soleimany AP, Amini A, Goldman S, Rus D, Bhatia SN, Coley CW. Evidential deep learning for guided molecular property prediction and discovery. ACS Cent Sci 2021;7:1356–67.
[18] Scalia G, Grambow CA, Pernici B, Li Y-P, Green WH. Evaluating scalable uncertainty estimation methods for deep learning based molecular property prediction. J. Chem Inf Model 2020;60:2697–717.