Selected Research Topic number: **SRT-d01**
Version: 1.0

Title: Uncertainty evaluation for machine learning

Abstract

There have been huge advances in recent years in the capability of machine learning (ML) (especially neural networks) for building accurate data-driven predictive models. This has been thanks to the availability of increasing volumes of data and advances in computational processing power and has led ML to become a key driver in digital transformation. However, to support this digital transformation, and provide trust in ML model outputs, a metrological framework and uncertainty evaluation for ML models is needed. This requires the ability to (i) reliably evaluate the uncertainty of ML outputs (ii) to quality assure ML models and (iii) to traceably link their outputs to primary measurement standards.

Keywords

Machine learning, artificial intelligence, ML model uncertainty evaluation, uncertainty propagation, neural networks.

Background to the Metrological Challenges

As the analytical engine of Artificial Intelligence (AI), ML is a key driver of digital transformation which has the potential to revolutionise the way that we understand the world and make decisions across all sectors of society, including advanced manufacturing, healthcare, finance, national security, criminal justice, transportation, and smart cities. But to support ML-driven digital transformation, the metrology community needs to incorporate ML into its frameworks, so that ML model outputs can be quality assured and traceable to primary measurement standards. Without such metrological assurances, the risks associated with decisions made on the basis of ML models cannot be properly understood, and limits their value in the context of e.g. conformity assessment, safety and regulation.

Although the evaluation of uncertainties using measurement models is standardised in the “Guide to the Expression of Uncertainty in Measurement” (GUM). The techniques described in the GUM are for fixed measurement models and are not applicable to ML and data-driven models. ML is currently being used within a range of European metrology projects, however to date, there hasn’t been a systematic investigation into ML requirements from a metrological perspective. Further to this, the vast majority of academic research into ML has been focused upon optimising predictive accuracy, rather than evaluating the uncertainty of the output of ML algorithms.

Bayesian uncertainty evaluation has long been established for algorithms based on classical statistical modelling such as Gaussian Processes (GPs). More recently, methods of uncertainty evaluation have been proposed for modern ML algorithms, including variational inference and ensembling of neural networks, and quantile regression forests. But existing methods typically make assumptions which are not valid in metrology applications, e.g. that target distributions are independent Gaussians. In addition, the vast majority of existing methods for ML uncertainty evaluation do not use knowledge about data uncertainties or propagate them through the model. This is a challenging task in which the effects of data uncertainties and model uncertainties must be understood, and the two suitably combined. There has been some work to address this issue, but further work is needed to adapt and enhance methods so that they can satisfy metrological requirements.

If the transformative potential of ML is to be fully realised ML predictions must be accompanied by reliable quantitative assessment of uncertainty and be validated for use in a broad range of high impact industrial applications. Examples include ML models for (i) inferring the state-of-health of lithium batteries from impedance measurements, (ii) the self-calibration of thermocouples, (iii) measuring nanoparticle size, in the production and control of materials and (iv) ensuring the stability of power grids. The European Network of Transmission System Operators for Electricity (ENTSO-E) has recognised a new need to monitor inertia as a critical system stability parameter. ML models of grid measurement data could be used to determine inertia, but the acceptance of the ML outputs by grid operators and standardisation bodies needs uncertainty evaluation.

Objectives

Proposers should address the objectives stated below, which are based on the PRT submissions. Proposers may identify amendments to the objectives or choose to address a subset of them in order to maximise the overall impact, or address budgetary or scientific / technical constraints, but the reasons for this should be clearly stated in the protocol.

The JRP shall focus on metrology research necessary to support digital transformation and the uncertainty evaluation of ML models.

The specific objectives are

1. To develop reliable and scalable methods for propagating data uncertainties through ML models. Method development should include Monte Carlo methods and errors-in-variables models, and the results should enable an extension of the GUM for the inclusion of data-driven models.
2. To develop a metrology framework for the traceability of ML components. This should include (i) combining propagated data uncertainties with model uncertainties, (ii) validating methods for uncertainty evaluation of ML outputs, (iii) ML model reliability and robustness, (iv) reference methods, and (v) the production of good practice guidance. Different methods for evaluating ML model should be considered, e.g. Monte Carlo and Gaussian dropout, variational inference, and deep ensembles.
3. To demonstrate the methods from Objectives 1 and 2, in at least 4 high impact end-user case studies e.g. (i) stability monitoring of power grids, (ii) battery state-of-health modelling, (iii) self-calibrating thermocouples and (iv) nanoparticle characterisation. This requires close collaboration with industrial end users and the deployment of demonstrators within their application domains.
4. To produce open source and quality assured, generic software for the uncertainty evaluation of ML models. The software should be suitable for wide use within the metrology community and accompanied by software instructions and good practice guidance.
5. To facilitate the take up of the technology and measurement infrastructure developed in the project by the measurement supply chain, standards developing organisations (ISO/IEC JTC1, ISO/TC 69/WG12, JCGM/WG1 and committees associated with the EU Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) regulation EC No 1272/2008) and end users (industry, advanced manufacturing, healthcare, finance, national security, transportation and energy sectors).

These objectives will require large-scale approaches that are beyond the capabilities of single National Metrology Institutes and Designated Institutes. To enhance the impact of the research, the involvement of the appropriate user community such as industry, standardisation and regulatory bodies, and other European Partnerships is strongly recommended, both prior to and during methodology development.

Proposers should establish the current state of the art and explain how their proposed project goes beyond this. In particular, proposers should outline the achievements of the EMPIR project(s) 17IND12 Met4FoF 18HLT07 MedalCare, 18HLT05 QUIERO, 18NET05 MATHMET and the EMN MATHMET and how their proposal will build on those.

EURAMET expects the average EU Contribution for the selected JRPs in this TP to be 1.7 M€ and has defined an upper limit of 2.4 M€ for this project.

EURAMET also expects the EU Contribution to the external funded beneficiaries to not exceed 35 % of the total EU Contribution across all selected projects in this TP.

Any industrial beneficiaries that will receive significant benefit from the results of the proposed project are expected to be beneficiaries without receiving funding or associated partners.

Potential Impact

Proposals must demonstrate adequate and appropriate participation/links to the ‘end user’ community, describing how the project partners will engage with relevant communities during the project to facilitate knowledge transfer and accelerate the uptake of project outputs. Evidence of support from the ‘end user’ community (e.g. letters of support) is also encouraged.

You should detail how your JRP results are going to:

- Address the SRT objectives and deliver solutions to the documented needs,
- Feed into the development of urgent documentary standards through appropriate standards bodies,
- Transfer knowledge to industry, advanced manufacturing, healthcare, finance, national security, transportation and energy sectors.

You should detail other impacts of your proposed JRP as specified in the document “Guide 4: Writing Joint Research Projects (JRPs)”

You should also detail how your approach to realising the objectives will further the aim of the Partnership to develop a coherent approach at the European level in the field of metrology and include the best available contributions from across the metrology community. Specifically, the opportunities for:

- improvement of the efficiency of use of available resources to better meet metrological needs and to assure the traceability of national standards
- the metrology capacity of EURAMET Member States whose metrology programmes are at an early stage of development to be increased
- organisations other than NMIs and DIs to be involved in the work.

Time-scale

The project should be of up to 3 years duration.