Contribution ID: **60** Type: **not specified**

Statistical Inference for Trustworthy AI: Cases for xAI and Uncertainty Quantification

Wednesday, 15 May 2024 09:15 (1 hour)

Trustworthy AI is dedicated to the development of methodologies and proofs that demonstrate the "proper" behavior of Artificial Intelligence algorithms, in order to favor their acceptance by users and organizations. By considering explainable AI and Uncertainty Quantification, we will show that defining consistent inferential procedures give systematic, reliable and arguable information for users. Starting from the Shapley Values for Importance Attributions, whose standard computations and interpretations have important limitations, we introduce the concept of Same Decision Probability that permits to identify import local and regional variables, and for which we can derive statistical consistency. Hence, regional measures of importance for variables appears to be a good scale for deriving consistent explanations, such as sufficient explanations, instead of local measures. In a second part, we will discuss the usefulness and potential of conformal prediction for deriving prediction intervals with guaranteed coverage rate. We will insist on the genericity and flexibility of the algorithms that permits to develop distribution-free inference for a large set of AI tasks, hence providing reliability measures useful for interacting with users.

Type of presentation

Invited Talk

Primary author: BRUNEL, Nicolas (LaMME (Laboratoire de Modélisation et Mathématiques d'EVRY), EDMH)

Presenter: BRUNEL, Nicolas (LaMME (Laboratoire de Modélisation et Mathématiques d'EVRY), EDMH)

Session Classification: Keynote

Track Classification: Spring Meeting