Contribution ID: **70** Type: **not specified**

Response Grid Plots for Model-Agnostic Explainable AI

Machine learning (ML) models are used to provide predictive characterization of response functions based on observed or simulated data. Insight on the nature of the approximation to the underlying response function is essential for the model output to be trusted and used by decision makers, a key part of exlainable AI. When a response function is well-behaved, methods that identify marginal effects and interactions for important input variables can provide adequate insight on its nature. Alternatively, interpretable ML models can provide an ensemble of simple models to decompose a complex response into interpretable elements. This talk introduces a model-generated but model-independent visual display of functional information that provides direct insight, not interpreted through model form, model coefficients, or numerical characterizations such as those produced by sensitivity analysis. This is shown to complement existing xAI approaches. Further, the method can be applied directly to system response data if the data are collected from a factorial (grid) design. The value of response grid plots (RGPs) is shown through several examples.

Classification

Mainly methodology

Special/ Invited session

QSR session - alternative track would be trustworthy and explainable AI

Keywords

Interpretable AI, Explainable AI, Graphical Methods

Primary author: BARTON, Russell (Pennsylvania State University)

Presenter: BARTON, Russell (Pennsylvania State University)

Track Classification: Other/special session/invited session