

Title: Exploration of the Effect of Task and User Role on the Evaluation of Interpretability Techniques

Task: BPP5.2

Co-Authors:

Cardiff University: Daniel Harborne, Alun Preece, Harri Taylor, Liam Hiley

BAE Systems: Chris Willis

IBM UK: Dave Braines, Richard Tomsett

IBM US: Supriyo Chakraborty

UCL: Simon Julier, Amy Widdicombe

UCLA: Moustafa Alzantot

Contact Author:

Dan Harborne

harborned@cardiff.ac.uk

Abstract:

As the effectiveness and accuracy of machine learning techniques has increased over time, the role they play in offering decision support in many domains also has increased. In most of these domains, the ability to understand the reasoning behind the outputs offered by these machine learning based systems is a crucial part of the validity of adopting them in real world scenarios.

In recent work, we have proposed that the nature of the task and the role of the user within that task has an impact on which forms of explanation offered by machine learning systems provide the most relevant insights to the user.

This demo is a step towards the exploration of this task- and user role- dependent impact on the value of different explanation types. By presenting an interface that allows the user to swap and configure the dataset, the model and the explanations offered, we create a platform to build intuition for the properties of various interpretability techniques within the scope of different tasks.