



RECENT ADVANCES IN AI FOR NATIONAL SECURITY

ScatterUQ: Interactive Uncertainty
Visualizations for Multiclass Deep
Learning Problems – **Harry Li**

Poster Abstract

Helping machine learning end users gauge a model's confidence is important for understanding model uncertainty and deciding follow-up courses of action. We present ScatterUQ, which constructs scatterplots to visually compare the salient features of test samples with the model's training examples to explain why the model predicts a test example to be 1) in-distribution and of a particular class, 2) in-distribution but unsure of the class, and 3) out-of-distribution. We demonstrate the effectiveness of ScatterUQ to explain model uncertainty for multiclass classification neural networks trained on image and cyber datasets. Our open-source code is available at <https://github.com/mit-ll-responsible-ai/equine-webapp>.

About the Author(s)

Harry Li is an associate staff member in the AI Technology and Systems Group. He specializes in prototyping visual tools for cyber and explainable AI applications. Harry received his BS in computer engineering from Rutgers University and MS in computer science from Tufts University through the Lincoln Scholars Program.

For more information, please contact raains@ll.mit.edu

DISTRIBUTION STATEMENT A. Approved for public release. Distribution is unlimited.

This material is based upon work supported by the Department of the Air Force under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Department of the Air Force.

© 2023 Massachusetts Institute of Technology.

Subject to FAR52.227-11 Patent Rights - Ownership by the contractor (May 2014) Delivered to the U.S. Government with Unlimited Rights, as defined in DFARS Part 252.227-7013 or 7014 (Feb 2014). Notwithstanding any copyright notice, U.S. Government rights in this work are defined by DFARS 252.227-7013 or DFARS 252.227-7014 as detailed above. Use of this work other than as specifically authorized by the U.S. Government may violate any copyrights that exist in this work.