PAMI Seminar SS 19

Focus of Papers/Theme: Explainable AI

Prof. Dr. Nils Bertschinger and Prof. Dr. Visvanathan Ramesh

“Why Should I Trust You?”
Explaining the Predictions of Any Classifier

Marco Tulio Ribeiro
University of Washington
Seattle, WA 98105, USA
marcotcr@cs.uw.edu

Sameer Singh
University of Washington
Seattle, WA 98105, USA
sameer@cs.uw.edu

Carlos Guestrin
University of Washington
Seattle, WA 98105, USA
guestrin@cs.uw.edu

This looks like *that*: deep learning for interpretable image recognition

Chaofan Chen[1*]
cfchen@cs.duke.edu

Oscar Li[1*]
runliang.li@duke.edu

Alina Barnett[1]
abarnett@cs.duke.edu

Jonathan Su[3]
su@ll.mit.edu

Cynthia Rudin[1,2]
cynthia@cs.duke.edu

[1]Department of Computer Science, Duke University, Durham, NC, USA 27708
[2]Department of Electrical and Computer Engineering, Duke University, Durham, NC, USA 27708
[3]MIT Lincoln Laboratory, Lexington, MA 02421-6426[†]

Abstract

Deep Learning for Case-Based Reasoning through Prototypes: A Neural Network that Explains Its Predictions

Oscar Li*[1], Hao Liu*[3], Chaofan Chen[1], Cynthia Rudin[1,2]
[1]Department of Computer Science, Duke University, Durham, NC, USA 27708
[2]Department of Electrical and Computer Engineering, Duke University, Durham, NC, USA 27708
[3]Kuang Yaming Honors School, Nanjing University, Nanjing, China, 210000
runliang.li@duke.edu, 141242059@smail.nju.edu.cn, {cfchen, cynthia}@cs.duke.edu

Transparency by Design: Closing the Gap Between Performance and Interpretability in Visual Reasoning

David Mascharka*[1] Philip Tran[2] Ryan Soklaski[1] Arjun Majumdar*[1]
[1]MIT Lincoln Laboratory[†] [2]Planck Aerosystems[‡]
{first.last}@ll.mit.edu, phil@planckaero.com

Explanations based on the Missing: Towards Contrastive Explanations with Pertinent Negatives*

Amit Dhurandhar[†1], Pin-Yu Chen[†1], Ronny Luss[1], Chun-Chen Tu[2], Paishun Ting[2], Karthikeyan Shanmugam[1] and Payel Das[1]

February 22, 2018

The Mythos of Model Interpretability

Zachary C. Lipton [1]

On the Robustness of Interpretability Methods

David Alvarez-Melis [1] Tommi S. Jaakkola [1]

AI in Education needs interpretable machine learning: Lessons from Open Learner Modelling

Cristina Conati [1] Kaśka Porayska-Pomsta [2] Manolis Mavrikis [2]

Building Machines That Learn and Think Like People

Brenden M. Lake,[1] Tomer D. Ullman,[2,4] Joshua B. Tenenbaum,[2,4] and Samuel J. Gershman[3,4]
[1]Center for Data Science, New York University
[2]Department of Brain and Cognitive Sciences, MIT
[3]Department of Psychology and Center for Brain Science, Harvard University
[4]Center for Brains Minds and Machines

Causality

Introduction to
Judea Pearl's Do-Calculus

Robert R. Tucci
P.O. Box 226
Bedford, MA 01730
tucci@ar-tiste.com

May 24, 2013

Snorkel: Rapid Training Data Creation with Weak Supervision

Alexander Ratner   Stephen H. Bach   Henry Ehrenberg
Jason Fries   Sen Wu   Christopher Ré
Stanford University
Stanford, CA, USA