Automated Reasoning for EXplainable Artificial Intelligence

Maria Paola Bonacina

Dipartimento di Informatica
Università degli Studi di Verona
Verona, Italy, EU

First ARCADE Workshop
"Automated Reasoning: Challenges, Applications, Directions, Exemplary Achievements"
Satellite of the 26th Int. Conf. on Automated Deduction (CADE-26)
Gothenburg, Sweden, EU

6 August 2017

Reasoning and learning

- Automated reasoning and machine learning in artificial intelligence
- AI between imitation and creation
- Machine learning as AI in contemporary general culture
- Automated reasoning between AI and theory
- Automated reasoning as enabling, background technology?

Only one way?

- Big data, better hw/sw, training tricks:
  great expectations for machine learning
- Automated reasoning as a source of big data
- Application of machine learning to automated reasoning
- Application of automated reasoning to the theory of machine learning
- How about applying automated reasoning to machine learning?

Challenges

- Machine learning for all kinds of decision making
- Black-box approach:
  - Amplification of biases
  - Prediction without explanation
- EXplainable AI (XAI):
  - What is explanation?
  - More than transparency (e.g., how?)
  - At least say what could go wrong if following the prediction
  - Attribution problem: input-based explanation
  - New: explain prediction based on training data?

Challenges for automated reasoning

- Explanation in AR: abduction, explanation of conflicts
- Could explanations for machine learning be computed by automated reasoning?
- How to bridge the gap between statistical and logical inference?
- How to bridge the gap of abstraction levels?
- Take-home message:
  - AR for XAI as long-term challenge
  - At least do not take for granted or immutable the current state of things