Annals of Mathematics and Artificial Intelligence
https://doi.org/10.1007/s10472-021-09772-0

An exploration of combinatorial testing-based approaches to fault localization for explainable AI

Ludwig Kampel[1] · Dimitris E. Simos[1] · D. Richard Kuhn[2] · Raghu N. Kacker[2]

Accepted: 24 August 2021Published online: 20 September 2021

Abstract
We briefly review properties of explainable AI proposed by various researchers. We take a structural approach to the problem of explainable AI, examine the feasibility of these aspects and extend them where appropriate. Afterwards, we review combinatorial methods for explainable AI which are based on combinatorial testing-based approaches to fault localization. Last, we view the combinatorial methods for explainable AI through the lens provided by the properties of explainable AI that are elaborated in this work. We pose resulting research questions that need to be answered and point towards possible solutions, which involve a hypothesis about a potential parallel between software testing, human cognition and brain capacity.

Keywords AI · Explainable AI · Combinatorial testing · Fault localization

Mathematics Subject Classification 2010 05B99 · 94C12 · 68T01

1 Introduction

Artificial intelligence (AI) systems have improved rapidly, with their performance now surpassing human abilities in many or most domains, especially vision and image recognition applications, but also in more safety-critical tasks such as autonomous vehicles [23, 38].

✉ Ludwig Kampel
lkampel@sba-research.org

Dimitris E. Simos
dsimos@sba-research.org

D. Richard Kuhn
kuhn@nist.gov

Raghu N. Kacker
raghu.kacker@nist.gov

[1] SBA Research, Vienna A-1040, Austria

[2] National Institute of Standards & Technology, Gaithersburg, MD, USA