

Experimental Study on Generating Multi-modal Explanations of Black-box Classifiers in terms of Gray-box Classifiers

Título Experimental Study on Generating Multi-modal Explanations of Black-box Classifiers in terms of Gray-box Classifiers

Autores [Jose M. Alonso](#), [Javier Toja-Alamancos](#), [Alberto Bugarin](#)

Tipo Comunicación para congreso

Fonte  [IEEE World Congress on Computational Intelligence](#), Glasgow (Scotland), IEEE, pp. 1-8 , 2020.

Abstract Artificial Intelligence (AI) has become a first class citizen in the cities of the 21st century. New applications are including features based on opportunities that AI brings, like medical diagnostic support systems, recommendation systems or intelligent assistance systems that we use every day. Also, each day, people are more concerned regarding the security and reliability of those AI-based systems. Moreover, trust, fairness, accountability, transparency and ethical issues are becoming main issues regarding AI-based systems. Institutions begin to issue regulations and to sponsor projects to promote AI transparency and to ensure that every decision made by an AI-based system can be convincingly explained to humans. In this context, Explainable AI (XAI), has become a hot topic within the research community. In this paper we have conducted an experimental study with 15 datasets to validate the feasibility of using a pool of gray-box classifiers to automatically explain a black-box classifier.

Palabras clave Explainable Artificial Intelligence, Interpretable Machine Learning, Classification, Open Source Software

DESCARGAS

 Referencia BibTex

PROXECTOS DE INVESTIGACIÓN

eXplica-IA: Diseñando Sistemas Inteligentes Explicables que Interaccionan con Personas y Generan Explicaciones en...

PROGRAMAS CIENTÍFICOS

Tecnoloxías da Linguaxe Natural