Model Agnostic Artificial Intelligence Explainer

Descrición

This PhD thesis will be developed in the context of the eXplica-IA project of excellence (funded by the "Consellería de Educación, Universidade e Formación Profesional of the Xunta de Galicia" through accreditation ED431F 2018/02), under supervision of Jose M. Alonso. The main objective of this project is to answer the scientific challenge of designing and developing a new generation of algorithms and Artificial Intelligence (AI) techniques which can provide self-explanatory capabilities to Intelligent Systems; that is, with the capacities to explain their behavior and decisions in Natural Language clearly and without linguistic ambiguities, both to expert and non-expert users. More precisely, this PhD thesis is aimed at defining, designing, developing and validating model agnostic explainers in the context of Explainable AI. Main challenges to face are: explaining black-box intelligent systems (e.g., deep learning systems) in terms of grey-box intelligent systems (e.g., decision trees, Bayesian networks, or fuzzy systems); looking for natural human-like explanations supported by analogical, deductive, inductive and abductive reasoning; enriching explanations with common-ground and contextual information; etc.

Perfil

We are looking for an outstanding and highly motivated candidate, with initiative, creativity and team-working ability, including working in interdisciplinary research groups. Candidates should fulfil the following eligibility criteria:

- A Degree in Computer Sciences, Telecommunication Engineering, Data Science, Artificial Intelligence or in another related area within the European Higher Education System.
- It is recommendable that the candidate already has (or is enrolled in) a Master degree or an equivalent University Degree (minimum 300 ECTS) that allows to start a PhD Program at the University of Santiago de Compostela.
- Excellent academic record, minimum of 8.5, will be required to be able to request a predoctoral contract from regional or national Government. A financial line is offered during the time in which the student is not ready to apply for the predoctoral contract (maximum one year, economic conditions depending on the profile of the candidate).
- Good computational skills.
- Very good English level.
- It is desirable basic knowledge on Artificial Intelligence but also on Natural Language Processing and Generation.

Bibliography

- M. T. Ribeiro, S. Singh, C. Guestrin, “Why should I trust you? explaining the predictions of any classifier”, ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), San Francisco, USA, pp.1-10, 2016.
- J. Forrest, S. Sripada, W. Pang, G. M. Coghill, “Towards making NLG a voice for interpretable Machine Learning”, Int. Conference on Natural Language Generation, 2018.
- J. M. Alonso, A. Ramos-Soto, C. Castiello, C. Mencar, “Hybrid data-expert explainable beer style classifier”, IJCAI Workshop on XAI, Stockholm, Sweden, pp. 1–5, 2018.

MORE INFO

For more information about this thesis proposal, start a conversation with us using the Contact webform.

PROGRAMAS CIENTÍFICOS

Tecnoloxías da Linguaxe Natural

Aprendizaxe automática