



Becas colaboración curso 2022/2023

Fecha: 01 Junio 2022

Vicerrectorado de Investigación

Subcomisión de I+D+i

Propuesta del departamento *SISTEMAS INFORMÁTICOS Y COMPUTACIÓN*

Núm Proyecto: 2022/32/00002

Responsable

Vidal Oriola, Germán Francisco

E-mail

gvidal@dsic.upv.es

Ext.

73587

Título proyecto

Diseño de casos de prueba para evaluar las técnicas de explicabilidad en inteligencia artificial

Valoración proyecto

4

Descripción proyecto

El proyecto se enmarca en el contexto de la inteligencia artificial explicable (XAI). Más concretamente, hemos diseñado e implementado recientemente una nueva técnica para generar explicaciones a partir de modelos especificados en un lenguaje lógico probabilístico (ProbLog). Dicha técnica se ha materializado en el sistema xgen (<https://github.com/mistupv/xgen>). El objetivo de este proyecto consistirá en identificar casos de prueba representativos (por ejemplo, sistemas de ayuda para la toma de decisiones, de diagnóstico médico, etc), extraer un modelo de los mismos en ProbLog (de manera manual o semiautomática empleando un sistema de extracción de reglas) y, finalmente, evaluar la calidad de las explicaciones producidas mediante el sistema xgen.

Actividades a realizar por el alumno

Las tareas del alumno se centrarán, principalmente, en las primeras fases del proyecto. Concretamente, el alumno deberá participar en las siguientes actividades:

- Búsqueda de casos de prueba. Para ello, el alumno deberá buscar sistemas de IA que estén disponibles públicamente (preferiblemente incluyendo los datos empleados para su entrenamiento). Los sistemas deben ser susceptibles de ser explicados de forma "causal" (esto excluye, por ejemplo, sistemas de reconocimiento de formas a partir de imágenes). Idealmente, se buscarán sistemas de recomendación, de ayuda a la toma de decisiones, de diagnóstico médico, etc, en los que tiene sentido esperar una explicación razonada de las respuestas del sistema. Por ejemplo, si un sistema de diagnóstico médico indica que un paciente tiene cierta enfermedad, sería interesante conocer qué factores (síntomas, valores en una analítica, etc) han motivado dicho diagnóstico. La mayoría de los sistemas de IA actuales operan como una "caja negra", por lo que no suelen ofrecer explicaciones apropiadas.

- Generación de modelos en ProbLog. En este caso, el alumno (en colaboración con el profesor responsable) tratará de generar un modelo en ProbLog para cada caso de prueba identificado en el punto anterior. El modelo se generará normalmente empleando un sistema de extracción de reglas a partir de datos (como, por ejemplo, [probfoil](https://pypi.org/project/probfoil/), <https://pypi.org/project/probfoil/>), aunque también es posible refinar el modelo de forma manual, añadir información contextual que represente el conocimiento del problema, etc.

Localización de la actividad (Campus)

Campus de Vera, Edificio 1F



Becas colaboración curso 2022/2023

Fecha: 01 Junio 2022

Horario

Horario flexible y/o online.