

Convocatoria de Contrato Predoctoral para la formación de Doctores vinculado a Proyecto de Investigación

CONTRATO PARA REALIZACIÓN DE TESIS DOCTORAL EN EL

Instituto de Investigación en Matemáticas (IMUVA). Universidad de Valladolid

Proyecto de investigación: Desarrollo de metodologías de clasificación y validación para datos complejos

Referencia: PID2021-128314NB-I00

Investigador Principal: Eustasio del Barrio, IMUVA

Luis Ángel García Escudero, IMUVA

DESCRIPCIÓN DEL PROYECTO.

Este proyecto está basado en las líneas de trabajo del G.I.R. en Probabilidad y Estadística Matemática de la Universidad de Valladolid, avaladas por una fructífera trayectoria de investigación iniciada hace ya más de 30 años. Se pretende lograr adaptaciones y ampliaciones de estas líneas, orientándolas hacia el desarrollo de metodologías de clasificación y validación aplicables ante datos complejos. La propuesta se centra en el avance del conocimiento, justificado por el necesario análisis de los conceptos y metodologías que se proponen, enfocados a su aplicabilidad en el ámbito de la Ciencia de Datos. Partiremos, como es habitual en nuestro equipo de investigación, de aspectos formales y teóricos de la Probabilidad y Estadística Matemática, con una perspectiva con fuerte componente de aplicabilidad.

Las técnicas de clasificación y validación son herramientas básicas para muchas disciplinas. Constituyen la base de numerosos procedimientos englobados bajo el epígrafe general de Ciencia de Datos y que necesitan acondicionarse para el abordaje de nuevos e importantes retos, consecuencia del fenómeno "Big Data" y de la mayor complejidad en los tipos de datos que aparecen. Las técnicas de clasificación y de validación están íntimamente relacionadas con el "aprendizaje automático", entendido como herramienta para "aprender" de los datos. Puede decirse que se "aprende" cuando el desempeño de alguna habilidad mejora con la experiencia, es decir, contando con más datos. Por ello, los modelos o algoritmos deben ser capaces de generalizar comportamientos e inferencias para conjuntos más amplios (potencialmente infinitos) de datos y, por tanto, pueden ser vistos como un intento de automatizar ciertas partes del método científico mediante técnicas con importantes componentes estadísticas, matemáticas y computacionales. Uno de los motivos fundamentales del notable desarrollo actual de las técnicas de clasificación y validación reside en el vertiginoso avance de la Informática y las Tecnologías de la Información, que proporcionan herramientas eficientes de adquisición, almacenamiento y tratamiento de cantidades ingentes de datos. No resulta sorprendente que muchas innovaciones recientes en las técnicas de clasificación y validación se hayan producido de forma paralela y combinada a importantes avances

en Informática, aunque también resulta patente que gran parte de las técnicas aplicadas tienen sus raíces en problemas típicamente abordados por la Probabilidad y la Estadística Matemática clásica. Resulta por tanto esencial que los grupos de investigación en Probabilidad y Estadística Matemática se impliquen en el abordaje de los nuevos retos que surgen en este entorno y una parte crucial es el tratamiento de conjuntos de datos con creciente nivel de complejidad.

Es muy común disponer de datos con muy diversas estructuras, alta dimensionalidad, datos anómalos y faltantes, así como con otros condicionantes que también deben ser tenidos en cuenta. Es bien sabido que la presencia de (incluso muy pocos) datos anómalos puede producir serios desajustes en los procedimientos de clasificación y validación clásicos, siendo su detección incluso un objetivo en sí mismo, por su interés, a nivel científico o de oportunidad de negocio, si se logra entender la razón que motiva su comportamiento diferencial. El estudio y desarrollo de procedimientos capaces de tratar estas nuevas complejidades resultan fundamentales por su gran repercusión tanto científica como social.

Nuestro punto de partida para abordar estos problemas, con el apoyo de los diferentes enfoques y técnicas ya existentes, será la aplicación y adaptación de principios relativos a las técnicas de recorte y métricas probabilísticas, que suponen el hilo conductor de la trayectoria del equipo. Así, por ejemplo, para tratar con la presencia de datos atípicos, objetivo de la Estadística Robusta, nuestra aproximación se ha realizado desde la óptica de los recortes, a menudo vinculados a métricas probabilísticas. Estas métricas formalizan la idea de proximidad de los objetos característicos de los problemas de naturaleza estocástica, especialmente de las distribuciones de probabilidad. En nuestro caso, merecen especial mención las métricas de Wasserstein, asociadas al problema del transporte óptimo, cuyo estudio ha sido una constante en nuestras investigaciones y que en los últimos años están siendo redescubiertas y utilizadas en todo tipo de aplicaciones. A su vez, los recortes, siendo la herramienta más antigua destinada a evitar el impacto de observaciones alejadas del resto, permiten analizar problemas de estabilidad o robustez en una gran variedad de problemas estadísticos, manteniendo la componente de intuición necesaria para la comprensión de los efectos que producen. También hay que señalar la trayectoria del equipo en la investigación de técnicas de remuestreo, que a menudo permiten aproximaciones computacionalmente factibles a problemas de gran complejidad.

RECEPCIÓN DE SOLICITUDES

Los interesados deberán contactar con uno de los Investigadores Principales del Proyecto en la dirección eustasio.delbarrio@uva.es o lagarcia@uva.es, adjuntando su curriculum vitae, copia de su expediente académico de grado y máster y, al menos, una carta de recomendación de un investigador o profesor universitario con experiencia en temas relacionados.

La cumplimentación y presentación telemática de la solicitud deberá realizarse obligatoriamente a través de los medios electrónicos habilitados para ello en la sede electrónica del Ministerio de Economía y Competitividad. El plazo de presentación de las solicitudes de participación será del 12 al 26 de enero de 2023 a las 14:00 horas (hora peninsular española).

Se puede obtener más información sobre la convocatoria en:

<https://www.aei.gob.es/convocatorias/buscador-convocatorias/ayudas-contratos-predoctorales-formacion-doctoresas-2022>