

Discurso de agradecimiento en la ceremonia de entrega de los Premios Sociedad de Estadística e Investigación Operativa – Fundación BBVA

Andrés Alonso y Daniel Peña

Queremos agradecer a todas las personas que han hecho posible este premio: Al presidente de la SEIO, Jesús López Fidalgo, que los propuso, y al Director de la Fundación BBVA, Rafael Pardo, que los apoyó con entusiasmo. Ambos han llevado a buen puerto esta excelente iniciativa. Muchas gracias por un buen trabajo. En segundo lugar, agradecemos al jurado que haya valorado nuestra aportación. Finalmente, gracias a la comprensión de nuestras parejas, Tamara Comas y Jette Bohsen, hemos podido dedicar muchas horas de nuestro ocio común a la investigación premiada. Les dedicamos este premio y nos sentimos afortunados por su cariño y su generosidad.

Uno de los problemas importantes en el análisis de datos masivos, o big data, es la clasificación de datos en grupos, por ejemplo, agrupar pacientes de una enfermedad por su genoma para encontrar nuevos tratamientos. Este problema se ha abordado desde la estadística, con la estimación de mezclas de distribuciones, desde la investigación operativa, con métodos de optimización, como máquinas de vector soporte, y desde la informática, con diversos algoritmos de clasificación. Sin embargo, la mayoría de los desarrollos se refieren a datos independientes, cuando, con frecuencia creciente, los datos automáticos masivos son dependientes, en el espacio y en el tiempo.

La agrupación de series temporales es un campo reciente, con destacada presencia española, basado en las similitudes entre las propiedades dinámicas de cada serie. Sin embargo, no existía un método para clasificar series por la dependencia entre ellas. Por ejemplo, agrupar enfermos por su similar respuesta a un medicamento, es decir, por correlaciones cruzadas altas, con independencia de la autocorrelación. Para ello, hemos definido una función,

15 de febrero de 2022

bautizada como correlación cruzada generalizada, que resume toda la dependencia dinámica entre dos series en distinto retardos. Utilizando esta función para clasificar las series hemos mostrado que el método propuesto es capaz de clasificar correctamente series simuladas y, con datos reales, encuentra grupos razonables y explicables, que no se detectan al agrupar por sus propiedades univariantes. En un trabajo posterior, realizado con la colaboración adicional de Pedro Galeano, hemos mostrado como nuestro método es capaz de descubrir grupos de series creados por un modelo factorial dinámico.

Para concluir, queremos felicitar a todos los premiados en estas dos primeras ediciones y manifestar el honor que nos supone compartir con amigos y compañeros estos premios, a los que deseamos una muy larga vida.