

# Un grupo de científicos granadinos gana un concurso internacional de 'big data'

El cartamen es uno de los más prestigiosos del mundo y se impusieron en la final a equipos de universidades del Reino Unido y Australia

:: **ANDREA G. PARRA**

**GRANADA.** Científicos de la Universidad de Granada, pertenecientes al grupo de investigación 'Soft Computing y Sistemas de Información Inteligentes' (SCI2S), han ganado la 'ECBDL'14 Big Data Competition', un concurso celebrado este verano en Vancouver (Canadá), en el marco del congreso internacional GECCO-2014. Este certamen, uno de los más prestigiosos del mundo, premia los mejores trabajos relacionados con los 'big data', conjuntos de datos de un elevado tamaño cuyo volumen, diversidad y complejidad requieren el uso de nuevas arquitecturas, técnicas, algoritmos y análisis para gestionar y extraer el valor y conocimiento oculto en ellos.

La 'ECBDL'14 Big Data Competition' se ha centrado en esta edición en un problema de clasificación en bioinformática. En concreto, los participantes debían trabajar sobre un conjunto de datos del campo de la



Los responsables del trabajo con el diploma que les acredita como ganadores del concurso. :: IDEAL

predicción de estructuras de proteínas, en el que se pretendía conseguir un predictor para distinguir un conjunto de estructuras a partir de las ya conocidas, especialmente la detección de contactos residuo-residuo en las proteínas.

El conjunto de entrenamiento utilizado en la competición constaba de dos clases, con alrededor de 32 millones de instancias con 631 atributos ocupando 56,7 Gigabytes de datos. Para validar la utilidad de los métodos de la competición se ha considerado un conjunto de test con unos 2,8 millones de ejemplos que se al-

macenan aproximadamente en 5 Gigabytes de datos.

El equipo de la UGR que ha ganado la competición ha propuesto una combinación de técnicas de preprocesamiento de datos (sobremuestreo de alta ratio sobre la clase minoritaria y selección de características basada en pesos) y multclasificadores basados en árboles de decisión utilizando MapReduce, extendiendo las ideas publicadas en la revista 'Information Sciences'. En segundo lugar quedó la Universidad de Newcastle (Reino Unido), y en tercero la Universidad de Nueva Gales del Sur (Aus-

tralia). Como explica el director del grupo de investigación 'Soft Computing y Sistemas de Información Inteligentes' de la UGR, Francisco Herrera, «los desarrollos tecnológicos en torno al 'big data' y el análisis inteligente de datos han dado lugar recientemente al término de Ciencia de Datos (Data Science), definido como un área emergente de trabajo relacionada con la preparación, análisis, visualización, gestión y mantenimiento de grandes colecciones de datos para la obtención de conocimiento que genere ventajas de negocio.