

S10. Metaheurísticas en Ciencia de Datos y Aprendizaje Automático

La investigación en Minería/Ciencia de datos y Aprendizaje Automático es por naturaleza de carácter multidisciplinar, teniendo un nexo muy relevante con la investigación en optimización (meta)heurística. Así, muchos de los problemas que aparecen durante el pre-procesamiento de los datos, el descubrimiento de modelos y el post-procesamiento de los mismos, son problemas de optimización combinatoria (selección de variables, generación de reglas de clasificación, selección de modelos en ensembles) o numérica (pesos en redes neuronales, intervalos en discretización, definición de conjuntos difusos, etc.). En la literatura podemos encontrar aplicaciones de metaheurísticas tanto de carácter local (simulated annealing, búsqueda tabú, búsqueda en entorno variable, ...), global (algoritmos genéticos, algoritmos de estimación de distribuciones, colonias de hormigas, ...) e híbrido (scatter search, algoritmos meméticos,...) que han demostrado ser herramientas poderosas en la resolución de problemas en el campo del aprendizaje automático y la minería de datos, obteniendo a menudo soluciones de muy buena calidad, si bien a cambio de un mayor coste computacional. Un nuevo reto para estas técnicas es su aplicación a problemas de BigData, que requieren del diseño de soluciones (implementaciones) en plataformas específicas.

El objetivo de esta sesión especial es reunir un conjunto de trabajos que representen una muestra del estado del arte del uso de metaheurísticas en minería/ciencia de datos, presentando algunos de los últimos desarrollos realizados en el área.

La lista de temas de interés para la sesión contiene (pero no se limita a) los que implican el uso de metaheurísticas en:

- Pre-procesamiento de los datos (selección de variables, selección de instancias, discretización,...)
- Aprendizaje de sistemas basados en reglas
- Aprendizaje de clasificadores
- Aprendizaje de sistemas basados en reglas difusas
- Aprendizaje de redes neuronales y sistemas neuro-difusos
- Clustering
- Análisis de asociaciones (reglas asociativas, redes bayesianas, redes de dependencias, etc.)
- Post-procesamiento (ajuste de pesos, simplificación/eliminación de reglas, ...)
- Aspectos de escalabilidad (evaluación rápida de fitness, paralelismo, etc ...), en problemas de alta dimensionalidad y/o gran número de muestras.
- Aplicaciones a problemas reales.

Organizadores de la sesión

- María José del Jesus. Universidad de Jaén. mjjesus@ujaen.es
- José A. Gámez. Universidad de Castilla-La Mancha. Jose.gamez@uclm.es
- José M. Puerta. Universidad de Castilla-La Mancha. Jose.puerta@uclm.es