



UGR | Universidad
de Granada



Propuesta TFG. Curso 2015-16

Departamento
Ciencias de la Computación e
Inteligencia Artificial

1. DATOS DEL TFG OFERTADO

Título del trabajo: Análisis del Conocimiento obtenido sobre Conjuntos de Datos Relacionados con la Biología mediante el Uso de Métodos de Clasificación Asociativa

Resumen (máx 250 palabras) estructurado en Objetivos y Plan de trabajo. Se debe incluir en folio adjunto.

Palabras clave: Data mining, Clasificación Asociativa, KEEL.

Número de alumnos por trabajo ofertado (máximo 3): 1

Ofertado por:

1. Profesor del Departamento
2. Profesor del Departamento junto con Empresa ó Institución
3. Propuesto por alumno (*)

(*) En el caso de TFG propuesto por alumno, por favor completar la siguiente información:

Nombre y apellidos del alumno:
e-mail institucional:

2. MODALIDAD

1. Trabajo bibliográfico
2. Trabajo experimental (*)
3. Informe o proyecto de naturaleza profesional (*)

(*) En el caso de trabajos experimentales e informes o proyectos de naturaleza profesional desarrollados en empresas u otras instituciones ajenas a la Universidad de Granada, por favor, completar la siguiente información

Nombre de la empresa/institución:
Domicilio social:
Teléfono/ e-mail de contacto:

3. DATOS DEL TUTOR DE LA UGR Y TUTOR DE LA EMPRESA O INSTITUCIÓN (en su caso) DEL TFG OFERTADO

Nombre y apellidos del tutor/a UGR: JESÚS ALCALÁ FERNÁNDEZ

Teléfono: 958240429

e-mail: jalcala@decsai.ugr.es

Nombre y apellidos del tutor/a de la empresa o institución:

Empresa o Institución:

Teléfono:

e-mail:

Resumen

Introducción

Los sistemas de clasificación han sido empleados en muchas aplicaciones del mundo real, tales como biología, medicina, etc. Sin embargo, en muchas de estas áreas los datos de los que disponemos tienen un número muy elevado de ejemplos y/o variables. Esto dificulta el aprendizaje de los sistemas de clasificación debido al crecimiento exponencial que experimenta el espacio de búsqueda, dando lugar a problemas de escalabilidad y complejidad.

El descubrimiento de asociaciones es una de las técnicas de minería de datos más utilizadas para extraer conocimiento interesante a partir de grandes conjuntos de datos. Debido a ello, durante los últimos años se han hecho grandes esfuerzos para utilizar sus ventajas para la clasificación, lo que ha dado lugar a un nuevo enfoque denominado clasificación asociativa. El descubrimiento de asociaciones tiene como objetivo encontrar relaciones interesantes entre los diferentes elementos de una base de datos, mientras que la clasificación tiene como objetivo descubrir un modelo a partir de un conjunto de datos de entrenamiento que pueda ser utilizado para predecir la clase a la que pertenecen ejemplos nuevos. Tanto el descubrimiento de asociaciones como la clasificación son esenciales en las aplicaciones de minería de datos y su integración resulta en un mayor ahorro y comodidad para el usuario.

Objetivos

El objetivo del presente TFG es familiarizar al alumno con la clasificación asociativa mediante el uso de algunas de las propuestas que existen en algunos problemas relacionados con la biología.

Plan de trabajo

Para ello se propone el siguiente plan de trabajo:

- El alumno debe primero estudiar, comprender y resumir el funcionamiento de los algoritmos básicos más conocidos para la clasificación asociativa. Se proponen los siguientes 4 algoritmos a estudiar: CBA, CPAR, CMAR, y FARCHD.
- Después, el alumno debe familiarizarse con la herramienta KEEL (<http://www.keel.es/>), para aplicar los 4 algoritmos nombrados en los siguientes 8 conjuntos de datos que puede obtener del repositorio KEEL-dataset (<http://www.keel.es/datasets.php>): Ecoli, Zoo, Abalone, Yeast, Breast, Bupa, newThyroid e Iris.
- Por último, el alumno debe realizar un análisis comparativo de los distintos algoritmos para determinar cuál es el más adecuado para cada problema. Adicionalmente, deberá analizar desde el punto de vista de un usuario final los conjuntos de reglas obtenidos por dicho algoritmo, mostrando que tiene capacidad para interpretar las reglas obtenidas.

Tabla de actividades y dedicación horaria estimada	
Planteamiento, orientación y supervisión	30 horas
Formación en la clasificación asociativa	20 horas
Análisis de distintas técnicas de la clasificación asociativa	50 horas
Formación en la herramienta software KEEL	30 horas
Análisis de los resultados obtenidos y selección de los más interesantes	80 horas
Preparación de la memoria	60 horas
Exposición del trabajo	30 horas
TOTAL (12 ECTS)	300 horas