



**UNIVERSIDAD
DE GRANADA**



Propuesta TFGB. Curso 2017-18

**DEPARTAMENTO: Ciencias de la
Computación e Inteligencia Artificial**

CÓDIGO DEL TFG: CCIA-04

1. DATOS DEL TFG OFERTADO:

Título: Implementación y Mejora de Algoritmos de REGRESIÓN para su Aplicación sobre Conjuntos de Datos Relacionados con la Biología

Resumen (máx 250 palabras, estructurado en Introducción, Objetivos y Plan de trabajo):

Introducción

El diseño de un modelo de REGRESIÓN a partir de datos, desde el punto de vista del aprendizaje supervisado, consiste en la obtención de modelos (sistemas basados en reglas difusas, sistemas basados en reglas intervalares, sistemas basados en árboles, sistemas basados en redes neuronales, etc.) que permitan predecir valores futuros de una variable objetivo (variable continua) a partir de mediciones o valores disponibles para otras variables de un sistema. En algunos casos, caso de los modelos basados en reglas, el modelo puede ser interpretado desde el punto de vista del usuario final que pretende obtener un sistema regresor para un problema dado, y en cualquier caso estos modelos pueden ser analizados y utilizados para inferir el posible valor de la variable objetivo (normalmente desconocida y/o difícil de medir para ciertos problemas). El manejo y comprensión de éstas herramientas de la IA es de especial interés en el área de la biología, la química, la medicina, etc....

Objetivos

El objetivo del presente TFG es familiarizar al alumno/a con dicho tipo de técnicas mediante su implementación (programación en Java o en R) y su uso en algunos problemas relacionados con la biología. Para ello se propone el siguiente plan de trabajo:

- El alumno/a debe estudiar, comprender y resumir el funcionamiento de los algoritmos básicos más conocidos de extracción de reglas para regresión. Se proponen los siguientes 7 algoritmos a estudiar: WM-R, M5Rules-R, CART-R, Ensemble-R, EPSILON_SVR-R, NU_SVR-R, RadomForest-R y WM+GFS-LLRS-T.
- El alumno/a debe familiarizarse con la herramienta KEEL (<http://www.keel.es/>), para aplicar los algoritmos nombrados en los siguientes conjuntos de datos (o similares) para regresión: Relative location of CT slices, KEGG Metabolic Relation Network (directed and undirected), Breast Cancer Wisconsin (Prognostic), Physicochemical Properties of Protein Tertiary Structure y Fertility. También debe buscar en los repositorios públicos disponibles conjuntos de datos relacionados con la biología adicionales, que le permitan analizar estadísticamente (si es posible) el comportamiento de dichos algoritmos.
- El alumno/a debe realizar un análisis comparativo de los distintos algoritmos para determinar cuál o cuales son los más interesantes.
- El alumno/a debe implementar en Java, R o lenguaje de programación equivalente los dos algoritmos que considere más interesantes de la anterior lista con el objetivo de mostrar que es capaz de entender y reproducir el funcionamiento de dichos algoritmos. Este paso lo/la preparará para poder modificar/adaptar y mejorar dichos algoritmos para un mejor funcionamiento en problemas particulares relacionados con la Biología.
- De entre todos los problemas (conjuntos de datos analizados), el alumno/a debe seleccionar uno que le resulte suficientemente interesante e intentar mejorar el funcionamiento de alguno de los dos algoritmos implementados (o ambos) en dicho problema. Debe proponer cambios o mejoras en dichos algoritmos y mostrar los resultados obtenidos respecto a los obtenidos por las versiones originales de dichos algoritmos.

Con todo ello se pretende que el alumno/a sea capaz de enfrentarse a un problema de regresión que se le pudiese plantear en su futuro desempeño como Biólogo/a, estudiando los algoritmos existentes, seleccionando el/los mejor/es, e incluso mejorándolos específicamente para obtener los mejores resultados posibles.

Conocimientos previos recomendados

Tener conocimientos de programación en Java, R o similar.

Tabla de actividades y dedicación estimada:

Planteamiento, orientación y supervisión	30 horas
Exposición del trabajo	1 horas
Desarrollo del trabajo	209 horas
Preparación de la memoria	60 horas
TOTAL (12 ECTS)	300 horas

OFERTADO POR:

Profesor/a del Departamento

Profesor/a del Departamento junto con Empresa o Institución

Propuesto/Acordado por estudiante y profesor/a (*)

(*) En este caso, por favor completar la siguiente información sobre el estudiante:

Apellidos:

Nombre:

e-mail institucional:

2. **MODALIDAD:** Trabajo bibliográfico



3. DATOS DEL TUTOR/A UGR:

Apellidos: ALCALÁ FERNÁNDEZ
Teléfono: 958 240427

Nombre: RAFAEL
e-mail: alcala@decsai.ugr.es

**En el caso de trabajos desarrollados en Empresas u otras Instituciones ajenas a la Universidad de Granada, por favor completar la siguiente información:

TUTOR/A DE LA EMPRESA O INSTITUCIÓN:

Apellidos:
Empresa/Institución:
Teléfono:

Nombre:
e-mail: