E.T.S. de Ingeniería Industrial, Informática  
y de Telecomunicación

Análisis de la eficacia del tuning local en un modelo de clasificación basado en reglas difusas de asociación (FARC-HD)



Grado en Ingeniería Informática

Trabajo Fin de Grado

Alumno: Oier Etxeberria Urrestarazu

Director: José Antonio Sanz

Pamplona, fecha de defensa (I don´t know yet men)



Agradecimientos

Patxi Mario Rodolfo el tercero.

Oier.

Resumen

Hoy en día el campo del *Machine Learning* o Aprendizaje Automático es un campo de investigación que está en auge. Dentro de este campo hay una gran variedad algoritmos y métodos capaces de extraer información. Dentro del subcampo de los algoritmos de clasificación nos encontramos con el FARC-HD.

FARC-HD es un algoritmo muy preciso y para ello en su etapa final aplica un algoritmo evolutivo para realizar un ajuste en las funciones de pertenencia y una selección de las mejores reglas. El ajuste de las funciones de pertenencia se hace de forma global para mantener la interpretabilidad del sistema y el objetivo del trabajo es analizar la eficacia de aplicar un ajuste local a expensas de perder la interpretabilidad del sistema.

Palabras clave

* Sistemas de clasificación
* Reglas de asociación difusas
* FARC-HD
* Tuning lateral
* Tuning local

Summary

Key words

Laburpena

Hitz Klabeak

Índice de Contenido

[1.- Introducción y motivación 9](#_Toc52275223)

[2.- Preliminares 10](#_Toc52275224)

[3.- FARC-HD-LOCAL 11](#_Toc52275225)

[4.- Marco Experimental 12](#_Toc52275226)

[4.1.- Datasets 12](#_Toc52275227)

[4.2.- Configuración de parámetros 12](#_Toc52275228)

[4.3.- Medidas de Rendimiento 12](#_Toc52275229)

[5.- Estudio Experimental 13](#_Toc52275230)

[Bibliografía 14](#_Toc52275231)

Índice de Ilustraciones

Índice de Tablas

[Tabla 1. Descripción de las características de los dataset utilizados 12](#_Toc52278693)

# 1.- Introducción y motivación

La inteligencia artificial es la ciencia que tiene como objetivo hacer que lo ordenadores puedan realizar tareas de humanos tal y como los haría un humano. Dentro de la ciencia de la inteligencia Artificial se encuentra el subcampo del *Machine Learning* o Aprendizaje Automático. Este subcampo se encarga de entrenar a la máquina para que aprenda por sí solo.

El aprendizaje automático tiene como objetivo aprender de una base de datos llena de ejemplos y posteriormente ser capaz de aplicar el aprendizaje en ejemplos desconocidos para la máquina.

Para realizar el aprendizaje podemos empezar de diferentes puntos de partida. Si en los ejemplos proporcionados para aprender sabemos la salida que debemos obtener, estamos ante lo que se llama aprendizaje supervisado. Por otro lado, si no lo sabemos, estamos ante aprendizaje no supervisado. Aparte de estos dos métodos existen más métodos, como el aprendizaje semisupervisado, por refuerzo, multi tarea, transducción etc. En nuestro caso utilizaremos el aprendizaje supervisado.

Este TFG esta basado en el algoritmo de clasificación basado en reglas difusas de asociación FARC-HD (*Fuzzy Rule-Based Classification model for High-Dimensional problems*) [1].

El algoritmo tiene tres fases. En la primera fase, se extraen reglas difusas de asociación. En la segunda fase, se minimizan las reglas preseleccionándolas conforme su calidad. Por último, en la tercera fase, se aplica un algoritmo genético para seleccionar las mejores reglas y ajustar las funciones de pertenencia.

Este trabajo se centra en cambiar el ajuste a las funciones de pertenencia. En el FARC-HD se ajusta de forma global con el fin de mantener la interpretabilidad del sistema. Nosotros estudiaremos el efecto de ajustarnos de forma local para cada regla.

Al ajustarnos de forma local, perdemos la interpretabilidad del sistema, pero logramos ajustarnos mejor a cada regla, por lo tanto, deberíamos mejorar el rendimiento. Por otro lado, para ajustarnos de forma local necesitaremos una base de datos para cada regla, lo que supone un aumento en el tiempo de computación.

# 2.- Preliminares

En este apartado del proyecto explicaremos en detalle el marco teórico necesario para entender el funcionamiento del algoritmo empleado en el TFG. Para ello, explicaremos los conceptos básicos paso a paso.

# 3.- FARC-HD-LOCAL

# 4.- Marco Experimental

En este apartado se presenta el contexto elegido para la realización de los experimentos. En primer lugar, describimos los conjuntos de datos (*datasets*) utilizados. Luego, mostramos los parámetros con los que se han ejecutado las pruebas. Por último, definimos las medidas de rendimiento empleadas para evaluar los resultados.

## 4.1.- Datasets

Hemos seleccionado nueve conjuntos de datos (*datasets*) que también se han utilizado para evaluar el rendimiento de FARC-HD original [1]. Estos conjuntos de datos contienen datos del mundo real obtenidos del repositorio de KEEL [2], disponibles públicamente desde el sitio web del proyecto[[1]](#footnote-1). En la tabla 1 resumimos las características principales de cada conjunto: número de ejemplos (#Ej.), número de atributos (#Atr.), y número de clases (#Clas.).

Tabla 1. Descripción de las características de los dataset utilizados

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **id.** | **Dataset** | **#Ej.** | **#Atr.** | **#Clas.** |
| clv | cleveland | 297 | 13 | 5 |
| eco | ecoli | 336 | 7 | 8 |
| iri | iris | 150 | 4 | 3 |
| mag | magic | 19020 | 10 | 2 |
| pag | pageblocks | 5472 | 10 | 5 |
| pen | penbased | 10992 | 16 | 10 |
| pm | pima | 768 | 8 | 2 |
| wine | wine | 178 | 13 | 3 |
| yst | yeast | 1484 | 8 | 10 |

## 4.2.- Configuración de parámetros

Tabla

## 4.3.- Medidas de Rendimiento

Convergencia, accuracy rate, tiempo.

# 5.- Estudio Experimental

# Bibliografía

1. J. Alcalá-Fdez, R. Alcalá y F. Herrera, «A fuzzy association rule-based classification model for high-dimensional problems with genetic rule selection and lateral tunning.,» IEEE Transactions on Fuzzy Systems, vol. 19, nº 5, pp. 857-872, 2011.
2. J. Alcalá-Fdez, A. Fernandez, J. Luengo, J. Derrac, S. García, L. Sánchez y F. Herrera, «KEEL data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework,» Journal of Multiple-Valued Logic and Soft Computing, Vols. %1 de %217:2-3, pp. 255-287, 2011

Cosas a revisar

* Traducir resumenes
* Poner bien números bibliografía
* Tablas con nombre abreviado?
* Referencia a tablas, ilustraciones, bibliografía..

1. <https://sci2s.ugr.es/keel/datasets.php> [↑](#footnote-ref-1)