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 10](#_Toc53404083)

[2.- Preliminares 11](#_Toc53404084)

[2.1.- Machine Learning 11](#_Toc53404085)

[2.2.- Problemas de Clasificación 11](#_Toc53404086)

[2.3.- Lógica Difusa 11](#_Toc53404087)

[2.3.1- Operadores 13](#_Toc53404088)

[2.3.2.- Implicación 14](#_Toc53404089)

[2.3.3.- Inferencia 15](#_Toc53404090)

[2.3.4.- Reglas Difusas 15](#_Toc53404091)

[2.4.- Tuning Local 16](#_Toc53404092)

[2.5.- Sistemas de Clasificación Basados en Reglas Difusas 16](#_Toc53404093)

[2.5.1.- Método De Razonamiento Difuso 17](#_Toc53404094)

[2.6.- Algoritmos Genéticos 18](#_Toc53404095)

[2.6.1.- Codificación 19](#_Toc53404096)

[2.6.2.- Cruce 20](#_Toc53404097)

[2.6.3.- Mutación 20](#_Toc53404098)

[2.6.4.- Selección 20](#_Toc53404099)

[2.7.- FARC-HD 20](#_Toc53404100)

[2.7.1.- Proceso de aprendizaje 21](#_Toc53404101)

[2.7.2.- Proceso de Clasificación 21](#_Toc53404102)

[2.7.3.- Algoritmo Genético en FARC-HD 22](#_Toc53404103)

[3.- FARC-HD-LOCAL 25](#_Toc53404104)

[3.1.- Tuning local en FARC-HD 25](#_Toc53404105)

[3.2.- Ampliando el rango de búsqueda 25](#_Toc53404106)

[4.- Marco Experimental 26](#_Toc53404107)

[4.1.- Datasets 26](#_Toc53404108)

[4.2.- Configuración de parámetros 26](#_Toc53404109)

[4.3.- Medidas de Rendimiento 27](#_Toc53404110)

[5.- Estudio Experimental 29](#_Toc53404111)

[5.1.- Comparando FARC-HD global y FARC-HD local con las configuraciones originales 29](#_Toc53404112)

[5.2.- Estudiando los modelos FARC-HD local y FARC-HD local incluyendo cambios en los parámetros de búsqueda 29](#_Toc53404113)

[5.3.- Comparando los mejores resultados obtenidos con todos los demás 29](#_Toc53404114)

[6.- Conclusiones 29](#_Toc53404115)

[7.- Líneas Futuras 29](#_Toc53404116)

[Bibliografía 30](#_Toc53404117)

Índice de Ilustraciones

[Ilustración 1. Ejemplo lógica Clásica 12](#_Toc53404118)

[Ilustración 2. Ejemplo Lógica difusa. 12](#_Toc53404119)

[Ilustración 3. Variable difusa con sus etiquetas lingüísticas 13](#_Toc53404120)

[Ilustración 4. Esquema de un SCBRD 17](#_Toc53404121)

[Ilustración 5. Esquema algoritmo evolutivo 19](#_Toc53404122)

[Ilustración 6. Ejemplo cromosoma en codificación binaria 19](#_Toc53404123)

[Ilustración 7. Etiquetas lingüísticas FARC-HD 21](#_Toc53404124)

[Ilustración 8. Desplazamiento de una etiqueta lingüística 22](#_Toc53404125)

[Ilustración 9. Desplazamiento lateral de una etiqueta 23](#_Toc53404126)

[Ilustración 10. Ejemplo codificación genética tuning lateral y selección de reglas 23](#_Toc53404127)

[Ilustración 11. Esquema CHC 24](#_Toc53404128)

Índice de Tablas

[Tabla 1. Descripción de las características de los dataset utilizados 27](#_Toc53404129)

[Tabla 2. Configuración de Parámetros 28](#_Toc53404130)

# 1.- Introducción

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 está 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.

## 2.1.- Machine Learning

En los últimos años empujado por el Big Data el *Machine Learning* a adquirido gran importancia. El *Machine Learning* se define como el aprendizaje de las máquinas y es un subcampo dentro de la Inteligencia Artificial.

Debido a la gran cantidad de datos que se generan diariamente y al potencial de conocimiento que estos datos tienen, surge la necesidad de analizarlos y procesarlos. La cantidad de datos es tal, que resulta imposible tratar los datos manualmente. De aquí nace el *Machine Learning*.

Todos los sistemas del *Machine Learning* generan un modelo y este es el que se encarga de procesar toda la información y tomar decisiones. Este modelo trabaja obteniendo una entrada de datos, procesándolos y obteniendo una salida sobre estos.

## 2.2.- Problemas de Clasificación

Los problemas de clasificación son un subcampo del *Machine Learning*. En este subcampo recibimos ejemplos y tratamos de clasificarlos en la clase a la que pertenecen. Para ello, partimos de un conjunto de ejemplos de los cuales sabemos a qué clase pertenecen. El objetivo es aprender los patrones generales que se repiten para pertenecer a cada clase y así clasificar correctamente ejemplos desconocidos.

## 2.3.- Lógica Difusa

La Lógica difusa la introdujo Lofti Zadeh [2] en 1965. En la lógica clásica, un elemento solo puede obtener dos valores: verdadero o falso. Es decir, un elemento puede pertenecer o no a un conjunto, pero el grado de pertenencia será absoluto. En la lógica difusa, por lo contrario, cada elemento tiene un grado de pertenencia con un valor real en el rango [0,1] para cada conjunto.

Vamos a explicar el ejemplo clásico que se suele utilizar para entender la lógica difusa. Vamos a determinar si un paciente tiene fiebre o no. Desde la perspectiva de la lógica clásica tendremos fiebre si la temperatura del paciente es mayor que 39º y sino no. Es decir, tendremos fiebre si la temperatura del paciente pertenece al conjunto de “*Fiebre Fuerte*”. Esta situación está representada en la Ilustración 1.

Diagrama

Descripción generada automáticamente

Ilustración 1. Ejemplo lógica Clásica

Por lo contrario, si representamos este caso con la lógica difusa, obtendríamos un valor real entre [0,1] para el conjunto de “*Fiebre Fuerte*”. Este caso, está representado en la Ilustración 2.

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

Ilustración 2. Ejemplo Lógica difusa.

La lógica difusa nos permite representar mejor la realidad, ya que podemos representar conceptos como mucho frio, poco alto, muy gordo…

A la hora de representar conjuntos difusos, tendremos las *variables lingüísticas* como altura, peso, temperatura… y también tendremos *etiquetas lingüísticas* como bajo, medio, alto… Podemos representarlo como en la Ilustración 3.

Diagrama

Descripción generada automáticamente

Ilustración 3. Variable difusa con sus etiquetas lingüísticas

Como estamos tratando con conjuntos, con la lógica difusa también se necesitan operadores de unión, intersección y complementario. Para estas operaciones, existen diferentes métodos llamados *t-normas, t-conormas,* y *negaciones.*

### 2.3.1- Operadores

1. **T-normas**

**Definición:** *Una t-norma es una función T: que satisface:*

* 1. Cota:
  2. Monotonía:
  3. Conmutativa:
  4. Asociativa:

Estos son algunos de los operadores más utilizados para representar la intersección:

* *Mínimo:*
* *Producto:*
* *Lukasiewicz:*

1. **T-conormas**

**Definición:** *Una t-conorma es una función S: que satisface:*

* 1. *Cota:*
  2. *Monotonía:*
  3. *Conmutativa:*
  4. *Asociativa:*

Estos son algunos de los operadores más utilizados para representar la unión:

* *Máximo:*
* *Suma algebraica:*
* *Lukasiewicz:*

1. **Negaciones**

**Definición:**  es una negación difusa si y solo si:

Además, se dice que una negación es estricta si:

* 1. es continua

E involutiva si:

Cuando una negación es estricta e involutiva se dice que es una negación fuerte. Así, dado un conjunto A, su complementario es siendo *c* una negación fuerte.

### 2.3.2.- Implicación

En dos universos U y V una relación difusa es un conjunto dado por el producto cartesiano que se puede definir con una función de pertenencia.

(1)

Una implicación (1) se define como un caso especial de relación difusa, la cual es posible definir mediante una función de pertenencia . Y a partir de esa función podemos construir una matriz R que represente la Regla.

Los operadores de implicación son funciones que satisfacen las propiedades:

Estos son algunos de los operadores de implicación más utilizados:

* *Kleene-Dienes:*
* *Lukasiewicz:*
* *Mamdani:*
* *Zadeh:*
* *Larsen:*

### 2.3.3.- Inferencia

### 2.3.4.- Reglas Difusas

Para construir una relación R que represente una regla, se pueden utilizar operadores de implicación o t-normas. Las t-normas en al practica suelen mejorar el rendimiento, pero no aprovechan parte de la teoría de la lógica clásica que con los operadores de implicación sí.

Para la regla (1) una opción para construir la regla puede ser así:

(2)

Existen distintos tipos de reglas difusas:

1. **Reglas difusas con una clase en el consecuente**

Si es y … y es entonces Y es

Donde , …, son las variables de entrada, , …, son las etiquetas lingüísticas de las variables e Y es la variable de salida que indica a cuál de las clases del conjunto de clases pertenece el ejemplo.

1. **Reglas difusas con una clase y un grado de certeza asociado**

Si es y … y es entonces Y es con grado

Donde , …, son las variables de entrada, , …, son las etiquetas lingüísticas de las variables e Y es la variable de salida que indica a cuál de las clases del conjunto de clases pertenece el ejemplo y es el grado de certeza de la regla .

1. **Reglas difusas con grados de certeza asociados a cada clase del consecuente**

Si es y … y es entonces

Donde , …, son las variables de entrada, , …, son las etiquetas lingüísticas de las variables y el consecuente es un vector de pesos en el que cada denota el grado de certeza de la regla para predecir la clase .

## 2.4.- Tuning Local

## 2.5.- Sistemas de Clasificación Basados en Reglas Difusas

Los sistemas de clasificación basados en Reglas Difusas (SCBRD), se utilizan para resolver problemas de clasificación. La ventaja que tienen es que proporcionan resultados precisos a la vez que interpretables. Esto se logra gracias a los términos lingüísticos utilizados en los antecedentes de las reglas.

El esquema de este tipo de clasificador se muestra en la Ilustración 4. Este tipo de clasificador está formado por dos componentes principales:

1. **Base de Conocimiento (BC):** La base del conocimiento se encarga de almacenar toda la información aprendida a partir de una base de datos de un problema específico. La componen los siguientes elementos:
   1. **Base de Datos (BC):** Contiene la definición de los conjuntos difusos asociados a los términos lingüísticos utilizados por la Base de Reglas.
   2. **Base de Reglas (BR):** Está formada por un conjunto de n Reglas de Difusas de clasificación (en la página 14).
2. **Método de Razonamiento Difuso (MDR):** Utiliza la información contenida en la Base de Conocimiento para determinar la clase a la que pertenece cualquier patrón de datos admisible. Para ello, utiliza el razonamiento aproximado propio de los sistemas difusos.

**Diagrama

Descripción generada automáticamente**

Ilustración 4. Esquema de un SCBRD

### 2.5.1.- Método De Razonamiento Difuso

El método de razonamiento difuso (MRD) es un procedimiento de inferencia que utiliza la información de la BC para predecir una clase ante un ejemplo no clasificado. Al final, indica como aplicar la información contenida en la BR, por ello, es uno de los elementos más importantes de un SCBRD, ya que determinará el rendimiento.

Sea un nuevo ejemplo a clasificar. Generalmente, un MDR utilizará el siguiente método para clasificarlo:

1. *Grado de emparejamiento.* Se calcula el grado de emparejamiento del ejemplo con los antecedentes de cada regla de la BR:

(2.1)

Siendo el grado de pertenencia del ejemplo con el antecedente i-ésimo de la regla, T una t-norma y es el número de antecedentes de la regla .

1. *Grado de asociación.* El grado de asociación del patrón con cada regla de la BR se calcula así:

(2.2)

Donde es el peso de la regla j-ésima, .

1. *Grado de confianza.* Se calcula el grado de confianza para cada clase. Se pueden utilizar diferentes funciones de agregación para calcularlo. Las más típicas son el máximo (MRD de la regla ganadora) y la suma (MRD de combinación aditiva) [3].
2. *Clasificación.* Se predice la clase con grado de confianza más alto.

(2.3)

## 2.6.- Algoritmos Genéticos

Dentro de los modelos de computación bio-inspirados se encuentran, los algoritmos evolutivos, las redes neuronales, los algoritmos basados en enjambres etc. Nosotros vamos a fijarnos en los algoritmos evolutivos.

Los algoritmos evolutivos tienen como objetivo resolver problemas de optimización o búsqueda y lo que los distingue es que el elemento clave de su diseño es algún mecanismo de evolución. Estos algoritmos intentan imitar el funcionamiento de los cromosomas en la vida real.

Los principales componentes de los algoritmos evolutivos son los siguientes:

* Población de individuos. Cada individuo representa una posible solución y se le llama cromosoma.
* Procedimiento de selección. Tiene algún mecanismo de selección basado en la aptitud de los individuos para resolver el problema.
* Procedimiento de transformación. En la evolución del programa se generan nuevos individuos a partir de los anteriores.

Existen distintos tipos de algoritmos evolutivos, pero en todos ellos el funcionamiento general es el siguiente (Ilustración 5):

1. Partimos de una población inicial de cromosomas. El número de cromosomas que conforman la población se mantiene fija.
2. Se evalúa la aptitud de cada cromosoma de la población.
3. Se seleccionan los cromosomas que se utilizaran para crear una nueva generación.
4. Se combinan los cromosomas seleccionados y se generan nuevos cromosomas. Los cromosomas pueden sufrir mutación.
5. Se genera la nueva población. Esta nueva población la completan los mejores cromosomas resultantes del conjunto de cromosomas de la generación anterior y la nueva.
6. Se repite el proceso de evaluación, selección y combinación hasta optimizar el resultado.

Diagrama

Descripción generada automáticamente

Ilustración 5. Esquema algoritmo evolutivo

Los algoritmos genéticos son un tipo de algoritmo evolutivo. Las principales características de un algoritmo genético son 3:

1. Los cromosomas se representan mediante codificación binaria.
2. Como método de selección suelen emplear el método de la ruleta con elitismo.
3. Como método de combinación emplean el cruce y la mutación.

A continuación, vamos a explicar más detalladamente en que se basan o como se realizan los apartados de codificación, cruce, mutación y selección.

### 2.6.1.- Codificación

La codificación es la forma en la que se representa la información del mundo real en nuestro cromosoma. Recordemos que un cromosoma debe representar una posible solución a nuestro problema, por lo tanto, la codificación es la forma en la que representamos una posible solución en un cromosoma.

Tenemos diferentes formas de codificación como la representación real, codificación entera, alfanumérica, codificación binaria etc. En todos ellos, la codificación se basa en una secuencia de reales, enteros, ceros y unos etc. Los algoritmos genéticos utilizan la codificación binaria (Ilustración 5).



Ilustración 6. Ejemplo cromosoma en codificación binaria

### 2.6.2.- Cruce

Los operadores de cruce se encargan de generar nuevos cromosomas para una nueva generación a partir de cromosomas de la generación anterior. Por ejemplo, a partir de dos cromosomas padres, generamos uno o más descendientes para la siguiente generación. De este modo permitimos evolucionar al algoritmo genético.

Explicar distintos tipos: cruce dos puntos, cruzamiento uniforme. Ejemplos!!!!

### 2.6.3.- Mutación

Basándose en la naturaleza la mutación consiste en modificar ligeramente el cromosoma después de aplicar el cruce. El objetivo de la mutación es dar más flexibilidad de evolución a la población. El algoritmo al final evoluciona gracias a los operadores de cruce y la mutación.

La mutación puede aplicarse de distintos modos, como cambiando un valor del cromosoma, invertir el orden de los valores del cromosoma en una parte o intercambiar valores del cromosoma etc.

### 2.6.4.- Selección

El objetivo principal de los métodos de selección es escoger los individuos mejor adaptados del conjunto de cromosomas de la anterior generación y la nueva generación obtenida a partir del cruce y la mutación. Para realizar la selección existe diferentes métodos como el método de la ruleta…

Ejemplos!!!!

## 2.7.- FARC-HD

Uno de los SCBRDs más interpretables de la literatura es el algoritmo FARC-HD (Fuzzy Association Rule-Based Clasification Model for High-Dimensional problems with Genetic Rule Selection and Lateral Tuning) [1]. El algoritmo utiliza la siguiente estructura de reglas:

Regla : Si es y … y es entonces Clase = con (3.1)

Donde es la etiqueta de la regla j-ésima, representa un ejemplo a través de un vector n-dimensional, representa un conjunto difuso, es la etiqueta de la clase y es el peso de la regla.

El peso , se calcula mediante el valor de confianza o grado de certeza definido en [5]:

(3.2)

done representa el grado de emparejamiento del ejemplo con los antecedentes de la regla difusa calculado mediante la Ecuación 2.1. En este caso, se utiliza la t-norma producto.

A la hora de modelar las etiquetas lingüísticas, en el FARC-HD se modelan mediante funciones de pertenencia triangulares uniformemente distribuidas, tal y como se muestra en la Ilustración 7.

Diagrama

Descripción generada automáticamente

Ilustración 7. Etiquetas lingüísticas FARC-HD

### 2.7.1.- Proceso de aprendizaje

El proceso de aprendizaje que se desarrolla en FARC-HD, está compuesto por tres etapas:

1. **Extracción de reglas difusas para clasificación:** Utilizamos un árbol de búsqueda para generar reglas difusas de asociación. Limitamos la profundidad máxima del árbol.
2. **Filtrado de reglas candidatas:** A causa de que obtenemos demasiadas reglas, aplicamos un filtrado basado en la mejora relativa de la medida de precisión (wWRAcc´) junto al algoritmo APRIORI-SD [7]. De este modo nos quedamos con las reglas más relevantes.
3. **Selección evolutiva de reglas y ajuste de etiquetas:** Por último, se emplea un algoritmo genético CHC [8] que mezcla la selección de reglas con el tuning [9].

### 2.7.2.- Proceso de Clasificación

FARC-HD aplica el método de Razonamiento Difuso llamado *combinación aditiva* [10]. El proceso empleado a la hora de clasificar un ejemplo dado es el descrito en el apartado 2.5.1.- Método De Razonamiento Difuso.

La combinación aditiva se basa en utilizar el método de la suma para obtener el grado de confianza:

(3.2)

### 2.7.3.- Algoritmo Genético en FARC-HD

Ahora explicaremos las características del algoritmo evolutivo empleado en FARC-HD:

1. *Codificación de los Cromosomas:* A la hora de codificar los cromosomas, codificamos dos partes distintas. Por un lado, codificamos el ajuste de las etiquetas y por otro la selección de reglas .

Para la selección de reglas, empleamos una codificación binaria donde el 1 significa que hemos seleccionado la regla y el 0 indica que no. De este modo, obtenemos la siguiente forma , siendo P en número de reglas candidatas.

Para el ajuste de las etiquetas (tuning), empleamos una codificación en valores reales que representan el parámetro α. Este parámetro indica el desplazamiento de las etiquetas lingüísticas tal y como se ve en la Ilustración 8.

Gráfico, Esquemático, Gráfico de cajas y bigotes

Descripción generada automáticamente

Ilustración 8. Desplazamiento de una etiqueta lingüística

Las etiquetas lingüísticas, tienen un rango máximo de desplazamiento que es el ± 0.5. Con esto evitamos que las etiquetas se intercambien las posiciones. Un ejemplo de desplazamiento lateral hacia la izquierda la podemos ver en la Ilustración 9.

Diagrama

Descripción generada automáticamente

Ilustración 9. Desplazamiento lateral de una etiqueta

De este modo tenemos una codificación tal que así: . Donde () indica el número de etiquetas por cada una de las n variables.

En total, un cromosoma nos quedaría tal y como se ve en la Ilustración 10. .

Diagrama

Descripción generada automáticamente

Ilustración 10. Ejemplo codificación genética tuning lateral y selección de reglas

1. *Evaluación de los cromosomas:* Para la evaluación de los cromosomas se emplea la siguiente función fitness:

(3.3)

Donde *#Hits* es el número de ejemplos correctamente clasificados, es el número de reglas candidatas, NR el número de reglas seleccionadas y un parámetro que determina la compensación entre el porcentaje de acierto y la complejidad, dado por un experto.

1. *Operadores de cruce:* En cada parte del cromosoma emplea un operador de cruce distinto.

Para el ajuste de las etiquetas (), emplea el operador *Parent Centric BLX* (PCBLX) [11]. El operador se basa en el concepto de vecindad, y permite a los genes resultantes del cruce desplazarse a una zona próxima a la de los progenitores.

Para la selección de reglas (), emplea el operador *Half Uniform Crossover Scheme* (HUX) [12]. El operador intercambia la mitad de los genes que sean diferentes en ambos progenitores.

1. *Modelo Evolutivo CHC:* FARC-HD utiliza el modelo evolutivo CHC [8]. Este modelo, emplea mecanismos de selección elitistas en la población para avanzar en la búsqueda global. Además, también emplea mecanismos de prevención de incesto mediante la distancia *hamming*. El mecanismo de incesto se basa en asegurar que los padres seleccionados para el cruce no son muy parecidos, y para ello solo los cruza si la distancia entre ellos es mayor que un umbral. Por último, el algoritmo cada vez que no se generan nuevos individuos, decrece el umbral, e itera hasta que el umbral es menor que cero. Podemos ver un esquema del funcionamiento en la Ilustración 11.

Diagrama

Descripción generada automáticamente

Ilustración 11. Esquema CHC

# 3.- FARC-HD-LOCAL

La idea principal de este TFG, consiste en aplicar tuning local en vez de tuning global en la última fase del FARC-HD original. Con este cambio se pretende dar más flexibilidad a las reglas para poder ajustarnos mejor y así obtener mejores resultados a costa de perder la interpretabilidad.

En esta sección primero explicaremos en qué consistirá el tuning local y cómo afectará al algoritmo y luego estudiaremos como mejorar la efectividad del tuning local. En el marco experimental se verán los resultados obtenidos para cada configuración y si es efectivo o no.

## 3.1.- Tuning local en FARC-HD

El proceso de entrenamiento del FARC-HD tiene tres fases, tal y como se explica anteriormente en **2.7.3.- Algoritmo Genético en FARC-HD**:

1. Extracción de reglas difusas para clasificación
2. Filtrado de reglas candidatas
3. Selección evolutiva y ajuste de la posición de las etiquetas

Para aplicar el tuning lateral local, nosotros solo modificaremos la última fase del FARC-HD original, en las demás fases no será necesaria ninguna modificación. En esta sección vamos a describir todos los cambios que conlleva aplicar el tuning local, primero explicaremos en que se basa el tuning local y luego como codificamos dichos cambios.

### 3.1.1.- Bases Tuning local

Se le llama tuning al proceso del ajuste de la posición de las etiquetas. En el FARC-HD original, el tuning se hace de forma lateral y global. De tal modo que si desplazamos a la izquierda una etiqueta como en la Ilustración 9, para todas las reglas tenemos el mismo desplazamiento.

El tuning lateral local consiste en ampliar el rango de búsqueda y de las reglas. Para ello, dejamos que cada regla se ajuste independientemente a cada etiqueta. De este modo si por ejemplo tenemos reglas y etiquetas, el desplazamiento de la etiqueta para la regla no es el mismo que el desplazamiento de la etiqueta para todas las demás reglas. Al final esto se traduce en tener lo mismo que tenemos en la Ilustración 9 para cada etiqueta de cada regla.

### 3.1.2.- Codificación tuning local

Como ahora vamos a tener un ajuste independiente para cada etiqueta de cada regla, necesitamos un cromosoma que contenga toda esta información. El cromosoma que vamos a utilizar para codificar el tuning local, consistirá en una matriz G donde cada fila será igual que el del cromosoma original empleado en FARC-HD, más una última celda que representa si seleccionamos la regla o no. Tendremos tantas filas como reglas. La siguiente matriz es una ilustración gráfica:

Donde hay cinco etiquetas *a, b, c, d, e*. Además, tenemos *m* variables y *n* reglas. Por último, la última columna de la matriz representa el del cromosoma.

A la hora de programar, también hay que tener en cuenta que como vamos a ajustarnos de forma independiente para cada regla, también necesitamos una base de datos independiente para cada regla.

## 3.2.- Mejorando efectividad del tuning local

El FARC-HD original, emplea distintos parámetros a lo largo de todo el algoritmo, los cuales están optimizados para aprender de la mejor manera posible. Ahora, como hemos cambiado el ajuste de las etiquetas aplicando tuning local, debemos estudiar si los parámetros que afecten de forma directa siguen teniendo un valor óptimo o no. Para ello debemos recordar el **Algoritmo Genético en FARC-HD.**

Para empezar, si nos fijamos en el funcionamiento del modelo evolutivo, también mostrado en la Ilustración 11, vemos que estamos ante un proceso cíclico que depende del *threshold o umbral* y el número de iteraciones.

Modificar el número de iteraciones es sencillo, puede pasar que el tuning local necesite más iteraciones dado que tiene un espacio de búsqueda más amplio. Por lo tanto, puede pasar que localmente vayamos ajustándonos poco a poco, pero sin pausa y que el número de iteraciones predefinido en le FARC-HD original se nos quede corto. Para evaluar la influencia hay que estudiar la convergencia del algoritmo a lo largo de las iteraciones.

Por otro lado, tenemos el *threshold o umbral.* El operador de cruce empleado para realizar el entrecruzamiento de los padres es un operador basado en la vecindad. Ahora, como hemos modificado la codificación del cromosoma también debemos estudiar como se define la vecindad entre distintos cromosomas y comprobar si es mejorable o no.

En el FARC-HD original [1], el mecanismo de incesto se controla mediante un parámetro denominado L. Este parámetro, empieza con un valor grande que representa la distancia entre dos cromosomas, es decir, cuanto mayor sea el valor, más distintos significa que son. Luego, cada vez que itera y no obtiene cromosomas nuevos decrementa el valor de L para forzar el cruce entre cromosomas más parecidos. L originalmente se inicializa con el siguiente valor:

donde *nLabels* es el número de variables por el número de etiquetas; *BITS\_GEN* es una constante que introducimos nosotros y *nReglas* es el número de reglas.

Como ahora tenemos más valores en el cromosoma la distancia que inicialmente utilizábamos puede quedarnos demasiado corta, por lo que interesa estudiar si este parámetro se nos queda demasiado exigente y por lo tanto merece la pena ampliar la distancia para llegar a mejores resultados. En el **Estudio Experimental** estudiamos este caso con la siguiente inicialización:

Por último, también puede pasar que al ajustarnos localmente dejemos algún ejemplo sin cubrir y lo metamos en la clase por defecto. La clase por defecto es la clase que más ejemplos tiene, por lo que lo introducimos ahí por estadística de acierto. Sin embargo, si son muchos los ejemplos que clasificamos de este modo conseguimos un efecto contraproducente y provocamos un aprendizaje menos eficiente. Si se diera este caso, podríamos aplicar una clasificación más eficiente como el basado en los centroides como el KNN por ejemplo. Es por ello por lo que es importante estudiar si obtenemos muchos ejemplos no cubiertos o no.

# 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 veintinueve conjuntos de datos (*datasets*) que son los mismos utilizados en [7] para evaluar el aprendizaje de sistemas de clasificación basadas en reglas difusas. Estos conjuntos de datos contienen datos del mundo real obtenidos del repositorio de KEEL [8], 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.). Los ejemplos con valores nulos como los del dataset *bands, Cleveland* y *Wisconsin* han sido eliminados tal y como en [7].

Para evaluar el rendimiento de los algoritmos, se ha utilizado un modelo de validación cruzada de 5 particiones [9]. Los resultados finales mostrados en el estudio experimental son el promedio del rendimiento *accuracy rate*, que son el porcentaje de ejemplos clasificados correctamente para cada una de las 5 particiones.

## 4.2.- Configuración de parámetros

En este apartado se muestran los parámetros con los que se ha realizado la experimentación. Dichos parámetros son los mismos que se utilizan en el FARC-HD [1] para realizar las pruebas. En el FARC-HD los parámetros que se utilizan son los recomendados por los autores para cada sección del algoritmo global [13]. Los parámetros que hemos utilizado se muestran en la Tabla 2.

Tabla 1. Descripción de las características de los dataset utilizados

|  |  |  |  |
| --- | --- | --- | --- |
| **Dataset** | **#Ej.** | **#Atr.** | **#Clas.** |
| banana | 5300 | 2 | 2 |
| bupa | 345 | 6 | 2 |
| cleveland | 297 | 13 | 5 |
| ecoli | 336 | 7 | 8 |
| glass | 214 | 9 | 6 |
| haberman | 306 | 3 | 2 |
| ion | 351 | 33 | 2 |
| iris | 150 | 4 | 3 |
| magic | 1902 | 10 | 2 |
| newthyroid | 215 | 5 | 3 |
| pageblocks | 547 | 10 | 5 |
| penbased | 1099 | 16 | 10 |
| phoneme | 5404 | 5 | 2 |
| pima | 768 | 8 | 2 |
| ring | 740 | 20 | 2 |
| satimage | 643 | 36 | 7 |
| shuttle | 5800 | 9 | 7 |
| tae | 151 | 5 | 3 |
| titanic | 2201 | 3 | 2 |
| twonorm | 740 | 20 | 2 |
| vehicle | 846 | 18 | 4 |
| wine | 178 | 13 | 3 |
| wisconsin | 683 | 11 | 2 |
| bands | 365 | 19 | 2 |
| spectfheart | 267 | 44 | 2 |
| wdbc | 569 | 30 | 2 |
| yeast | 1484 | 8 | 10 |
| winequality-red | 1599 | 11 | 11 |
| balance | 625 | 4 | 3 |

## 4.3.- Medidas de Rendimiento

Para evaluar el rendimiento de las pruebas, utilizamos el grado de precisión o *accuracy rate*. Este parámetro calcula el porcentaje de ejemplos clasificados correctamente. Por otro lado, con tal de aportar soporte estadístico a los resultados, utilizamos la prueba de rangos con signo Wilcoxon [14].

Tabla 2. Configuración de Parámetros

|  |
| --- |
| **Parámetros** |
| Número de etiquetas lingüísticas por variable: 5 |
| Soporte mínimo: 0.05 |
| Confianza mínima: 0.8 |
| Profundidad máxima Árbol: 3 |
| Parámetro K: 2 |
| Número máximo de evaluaciones: 20000 |
| Número de individuos: 50 |
| Parámetro α: 0.02 |
| Bits por gen: 30 |
| Inferencia: 1 (Combinación aditiva) |

# 5.- Estudio Experimental

En este capítulo, se muestran y analizan los resultados obtenidos por los dos métodos de tuning utilizados, con distinta configuración de parámetros. Para ello, hemos dividido el estudio experimental en distintas etapas:

1. Se presentan los resultados obtenidos con el FARC-HD global y FARC-HD local en sus configuraciones, por defecto y se analizan.
2. Se comparan los resultados obtenidos en FARC-HD global y FARC-HD local con las distintas configuraciones de parámetros.
3. Se comparan todas las configuraciones y sus respectivos resultados.

En las tablas el nombre de los algoritmos esta codificado con su configuración. La primera letra indica si es global (g) o local (l). El siguiente número indica el número de iteraciones; 20000 (20) o *num\_etiquetas* \* *num\_variables* \* 5000 (var). Por último, el último número indica si el parámetro L esta modificado (2) o no (4). Tal y como se explica en **Mejorando efectividad del tuning local**.

## 5.1.- Comparando FARC-HD global y FARC-HD local con las configuraciones originales

En esta sección, primero mostramos los resultados obtenidos en los dos modelos con la misma configuración de parámetros con los que se ejecuta el FARC-HD [1]. Para obtener dichos resultados se utilizan los **Datasets** definidos previamente. Podemos observar los resultados en la Tabla 3.

Los resultamos muestran que localmente sin modificar ningún parámetro, no obtenemos el efecto de mejor ajuste que estábamos buscando ya que el tuning global logra resultados ligeramente mejores.

Tabla 3. FARC-HD global vs FARC-HD local

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | g20-4 | | l20-4 | |
|  | Train | Test | Train | Test |
| banana | 86,9811321 | 86,1886792 | 86,7641509 | 85,8867925 |
| bupa | 79,4202899 | 66,0869565 | 80,0724638 | 62,6086957 |
| cleveland | 89,8996561 | 56,8926554 | 88,8887707 | 58,2485876 |
| ecoli | 92,1880375 | 82,1378402 | 92,1874827 | 81,8481124 |
| glass | 81,3096695 | 67,2978959 | 82,0121039 | 64,5182724 |
| haberman | 81,2084309 | 74,4949762 | 81,8618267 | 71,2268641 |
| ion | 98,8607016 | 88,8933602 | 98,6474326 | 88,889336 |
| iris | 98,5 | 94,6666667 | 98,6666667 | 94,6666667 |
| magic | 84,2533484 | 80,9650504 | 84,5030804 | 81,280978 |
| newthyroid | 99,1860465 | 95,8139535 | 99,1860465 | 95,8139535 |
| pageblocks | 97,0796018 | 93,9749791 | 97,0340437 | 93,6113428 |
| penbased | 98,4772727 | 92,1818182 | 98,6136364 | 91,4545455 |
| phoneme | 83,2809352 | 81,513482 | 83,5075697 | 81,3658992 |
| pima | 83,8212971 | 74,0853917 | 83,7890946 | 75,1243528 |
| ring | 97,0945946 | 90,5405405 | 97,2635135 | 91,7567568 |
| satimage | 85,5754599 | 81,0222868 | 85,4974878 | 79,934593 |
| shuttle | 97,1494253 | 96,9655172 | 96,5862069 | 96,2758621 |
| tae | 76,4917355 | 59,6129032 | 76,822314 | 57,6344086 |
| titanic | 79,0663427 | 78,8733251 | 79,0663427 | 78,8733251 |
| twonorm | 97,8040541 | 90,2702703 | 97,6689189 | 90 |
| vehicle | 80,4074275 | 69,2655761 | 80,2006765 | 68,5638705 |
| wine | 100 | 96,047619 | 100 | 93,8095238 |
| wisconsin | 98,8286424 | 96,9246458 | 98,6090631 | 96,9246458 |
| bands | 89,1095801 | 69,6973839 | 88,8351384 | 69,6748401 |
| spectfheart | 93,3522004 | 78,6373166 | 92,8844719 | 79,0216632 |
| wdbc | 98,5937922 | 96,6604565 | 98,6378446 | 95,9587021 |
| yeast | 64,2185829 | 58,490536 | 64,6059301 | 58,3563109 |
| winequality-red | 65,9162432 | 60,9112461 | 65,4473343 | 60,4715909 |
| balance | 92,16 | 85,76 | 92,16 | 86,72 |
|  |  |  |  |  |
| Media | **88,6287759** | **80,857701** | 88,6213659 | 80,3627756 |

## 5.2.- Estudiando los modelos FARC-HD local y FARC-HD local incluyendo cambios en los parámetros de búsqueda

**Mejorando efectividad del tuning local**

## 5.3.- Comparando los mejores resultados obtenidos con todos los demás

# 6.- Conclusiones

# 7.- Líneas Futuras

* Revisar reglas no cuviertas -> KNN
* Optimizar ajuste local (cambios de parametros) ¿?
* Amplitud en el tuning local y global

# 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. L. A. Zadeh, «Fuzzy Sets,» Information and Control, vol. 8, nº 3, pp. 338-353, 1965
3. O. Cordón, M. del Jesus y F. Herrera, «A proposal on reasoning methods in fuzzy rulebased classification systems",» International Journal of Approximate Reasoning, vol. 20, nº 1, pp. 21-45, 1999
4. M. Mitchell, Introduction to Genetic Algorithms, 1998.
5. H. Ishibuchi y T. Yamamoto, «Rule weight specification in fuzzy rule-based classification systems,» IEEE Transactions on Fuzzy Systems, vol. 13, nº 4, pp. 428- 435, 2005.
6. D. E. Goldberg, Genetic algorithms in search, optimization and machine learning, 1989.
7. B. Kavsek and N. Lavrac, “Apriori-sd: Adapting association rule learning to subgroup discovery,” Applied Artificial Intelligence, vol. 20, no. 7, pp. 543–583, 2006.
8. L. Eshelman, «The chc adaptative search algorithm: How to have safe search when engaging in nontraditional genetic recombination,» de Foundations of Genetic Algorithms, G. Rawling, Ed. Morgan Kaufmann, 1991, pp. 265-283.
9. R. Alcala, J. Alcal ´ a-Fdez, and F. Herrera, “A proposal for the genetic ´ lateral tuning of linguistic fuzzy systems and its interaction with rule selection,” IEEE Transactions on Fuzzy Systems, vol. 15, no. 4, pp. 616– 635, 2007.
10. O. Cordón, M. del Jesus y F. Herrera, «A proposal on reasoning methods in fuzzy rulebased classification systems",» International Journal of Approximate Reasoning, vol. 20, nº 1, pp. 21-45, 1999.
11. L. Eshelman y J. Schaffer, «Real-coded genetic algorithms and interval schemata,» de Foundations of Genetic Algorithms, vol. 2, D. Whitley, Ed. Morgan Kaufmann, 1993, pp. 187-202.
12. M. Lozano, F. Herrera, N. Krasnogor y D. Molina, «Real-coded memetic algorithms with crossover hill-climbing,» Evolutionary Computation, vol. 12, nº 3, pp. 273-302, 2004
13. ASOC
14. 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*, vol. 17, no, 2-3, pp. 255-287, 2011
15. T. Wong, P. Yeh, Reliable accuracy estimates from k-fold cross validation, IEEE Transactions on Knowledge and Data Engineering 32 (8) (2020) 1586– 805 1594.
16. J. Alcala-Fdez, L. Sánchez, S. García, M. del Jesus, S. Ventura, J. Garrell, J. Otero, C. Romero, J. Bacardit, V. Rivas, J. Fernandez, and ´ F. Herrera, “KEEL: A software tool to assess evolutionary algorithms to data mining problems,” Soft Computing, vol. 13, no. 3, pp. 307–318, 2009
17. F. Wilcoxon, «Individual comparisons by ranking methods,» Biometrics, vol. 1, nº 6, pp. 80-83, 1945.

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)