

Modelado de Sistemas Difusos Basado en el Algoritmo Genético Compacto

Miguel Melgarejo¹, Alvaro Prieto¹, Carlos Ruiz¹

¹ Laboratorio de Automática, Microelectrónica e Inteligencia Computacional,
Universidad Distrital Francisco José de Caldas
Bogotá D.C., Colombia
{mmelgarejo, aaprietom, camruizc} @udistrital.edu.co

Resumen. Este artículo presenta dos casos de estudio para el modelado de sistemas difusos basado en el Algoritmo Genético Compacto (CGA). Se estudian los casos de clasificación de la flor Iris de Fisher y la estimación de la función Sinc. Se realiza una comparación del enfoque basado en el CGA contra el enfoque basado en el Algoritmo Genético clásico (GA). Los resultados obtenidos de 100 experimentos para variaciones de los enfoques de CGA y GA, apuntan a que para los casos de estudio, el desempeño del enfoque basado en CGA es mejor dado que el error disminuye, y es más rápido con respecto al número de iteraciones necesarias para converger.

Palabras clave. Sistemas Difusos, Modelado de Sistemas Difusos, Algoritmos Evolutivos, Caminata aleatoria, Clasificador, Estimador.

1. Introducción

Dentro del Laboratorio de Automática, Micro-electrónica e Inteligencia Computacional (LAMIC) de la Universidad Distrital Francisco José de Caldas, se ha venido desarrollando una línea de trabajo en el área de inteligencia computacional. De esta experiencia pedagógica en el pregrado de Ingeniería Electrónica, se han reportado algunos trabajos orientados a la solución de problemas relacionados con sistemas de inferencia difusa (FIS, por sus siglas en inglés) [1], [2]. De estos trabajos, vale la pena destacar uno relacionado con este artículo, y es el realizado por Hostos et al [1]. En este trabajo se reporta el uso de una caja de herramientas para el modelado coevolutivo de FIS, en el problema de la predicción de una serie de tiempo.

El modelado de FIS con el uso de algoritmos evolutivos ha sido utilizado en la solución de problemas en control [2], [3], [4] y reconocimiento de patrones y clasificación [5], [6], [7], por nombrar algunos ejemplos. Con el fin de explorar alrededor del modelado de FIS con una perspectiva evolutiva diferente, se presenta en este artículo una propuesta metodológica basada en el Algoritmo Genético Compacto (CGA). El CGA pertenece a

una clase reciente de algoritmos genéticos conocidos como algoritmos genéticos construidos por medio de modelos probabilísticos (PMBGA, por sus siglas en inglés) [8].

Debido a que la mutación y la recombinación en el algoritmo genético clásico (GA) tienen una naturaleza aleatoria, su compartimiento es nocivo en el proceso evolutivo [8], [9]. Esto se debe a que estos operadores no discriminan las características que tienen una alta contribución al buen desempeño de un individuo, de aquellas que no aportan o empeoran el desempeño del individuo. Las características que aportan al buen desempeño son conocidas como bloques constructores. Es por esta razón que nacen los PMBGA, cuyo objetivo es conservar los bloques constructores de un individuo durante el proceso evolutivo. En la Figura 1 se presenta la propuesta metodológica para el modelado de FIS basada en el CGA.

Hasta la fecha, en la literatura no se reportan documentos en los que se realice el modelado evolutivo de un FIS con un CGA. Dicho algoritmo se ha implementado como algoritmo de optimización en [10], [11], [12]. La validación de dicha propuesta se realizó sobre dos casos de estudio: la clasificación de la flor Iris a partir de los datos de Fisher [13], [14] y la estimación de la función Sinc. Los resultados obtenidos muestran un impacto positivo en el proceso evolutivo cuando se utiliza el CGA como algoritmo de optimización, comparado contra el algoritmo genético clásico GA.

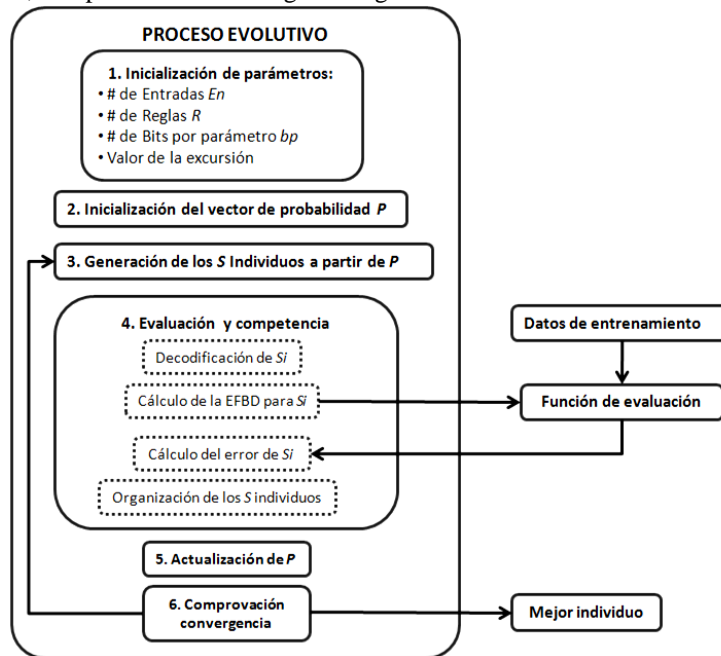


Fig 1. Propuesta para el modelado de un FIS tipo-1 con CGA

El artículo se organiza de la siguiente manera: la sección II introduce algunos principios metodológicos de la propuesta. La sección III presenta la implementación de la metodología sobre el primer caso de estudio: Clasificador de la flor de Iris. La sección IV presenta el segundo caso: El estimador de la función Sinc. Para ambos casos se muestran y discuten los resultados obtenidos de los experimentos. Finalmente, se concluye en la sección V.

2. Consideraciones metodológicas

2.1 FIS

En el bloque 4 de la Figura 1 se observa el cálculo de una Expansión de Funciones de Base Difusa (EFBD). Una EFBD es una forma de representar un FIS, en la cual un vector de entrada \mathbf{x} se relaciona con una salida puntual y de tal forma que $y = f(\mathbf{x})$ [15]. De esta manera se representa de manera compacta el proceso de inferencia en un FIS y la función resultante es un estimador universal [15], [16]. En las ecuaciones (1), (2) y (3) se muestran la representación matemática de una EFBD.

$$y(\mathbf{x}) = f(\mathbf{x}) = \sum_{l=1}^M y_l \phi_l(\mathbf{x}) \quad (1)$$

$$\phi_l(\mathbf{x}) = \frac{\prod_{i=1}^N \mu_{A_i^l}(x_i)}{\sum_{l=1}^M \prod_{i=1}^N \mu_{A_i^l}(x_i)} \quad (2)$$

$$\mu_{A_i^l}(x_i) = e^{-\frac{(x_i - c_i^l)^2}{2(\sigma_i^l)^2}} \quad (3)$$

El FIS representado por (1), (2) y (3) tiene las siguientes características: *Singleton Fuzzification*, T-norma producto, motor de inferencia producto, *defuzzification* por promedio de centros y una función de pertenencia Gaussiana. El índice l hace referencia a la l -ésima regla, siendo M el total de reglas. Por su parte el índice i hace referencia a la i -ésima entrada y N es el total de las mismas. La función de pertenencia $\mu_{A_i^l}(x_i)$ sería, entonces, única para cada entrada en cada regla. De igual manera el centro del conjunto consecuente y_l sería único en cada regla. De esta manera se garantiza una base de reglas completa y cada regla tendría N términos lingüísticos asociados.

Un FIS que se presenta a través de una EFBD se puede parametrizar y de esta manera queda totalmente definido a partir de una colección de parámetros. El número de entradas N y el número de reglas M definirían la cantidad de conjuntos antecedentes y centros de conjuntos consecuentes. Por su parte, la media c y la desviación estándar σ son los parámetros que definen la función de pertenencia Gaussiana. Teniendo en cuenta que por entrada y en cada regla hay dos parámetros del conjunto antecedente y un centro en el consecuente, entonces la cantidad de parámetros que definen el FIS queda definida en (4).

$$Op = (2 * M * N) + M \quad (4)$$

2.2 CGA

Por medio del modelo matemático conocido como paseo aleatorio [9], se puede analizar el crecimiento o caída de un gen (característica) de un individuo en un proceso evolutivo. En el transcurrir de las generaciones, este gen, lucha con sus competidores y prevalecerá dependiendo de las decisiones que el algoritmo tome, entendiéndose como decisión el proceso de selección de los individuos que conforman la siguiente generación. La variable que determina el crecimiento o la caída de un gen es la población n , que representa la cantidad de individuos generados en un proceso evolutivo [9].

En un proceso evolutivo el genoma está conformado por aquellos parámetros o características que definen una solución para el problema dado. La cantidad de dichos parámetros está necesariamente ligada a la naturaleza del problema. En el proceso de selección en un algoritmo genético clásico (GA, por sus siglas en inglés) se le da relevancia al individuo con mejor desempeño sin distinguir aquellas características que aportan más a la solución. Por su parte, el CGA además de dar prevalencia en la selección al individuo con mejor desempeño, también da prevalencia a aquellos genes que son diferentes a los de su competencia y que por tanto son los que aportan, para bien o para mal, en el desempeño. En la Figura 2 se presenta el pseudocódigo del CGA.

En primera instancia, el vector de probabilidad es una representación estadística de toda la población en la cual cada posición representa la probabilidad de que dicha posición en cada uno de los S individuos sea '1'. El tamaño de este vector dependerá de la cantidad de características (genes) que representa una única solución al problema. Cuando se inicializa el algoritmo este vector tiene un valor de 0.5 en todas las posiciones con el objetivo de iniciar la búsqueda en todo el universo de soluciones disponibles.

Dependiendo del problema que se busca solucionar se escoge la función de aptitud con la que se evaluarán los individuos. A partir del valor que se obtiene de la función de aptitud para cada individuo se selecciona el que tiene el mejor desempeño, en otras palabras, aquel individuo que representa la mejor solución. A continuación se realiza la actualización del vector de probabilidad. Dicha actualización se realiza comparando bit a bit (gen a gen) cada uno del resto de individuos perdedores, $S-1$, con respecto al mejor. La actualización se realiza sobre cada posición del vector de probabilidad, y sólo se realiza si en dicha posición el valor del individuo perdedor difiere con respecto al valor del individuo ganador en la misma posición. El valor que se suma o resta para realizar la actualización es el inverso del valor n .

El criterio de convergencia del CGA está ligado al vector de probabilidad. Si el vector de probabilidad en todas las posiciones tiene valores de '0' o '1' significa que la probabilidad de que dichas posiciones tengan un valor de '1' es del 100% o 0%. Por lo tanto todos los individuos generados a partir del vector de probabilidad van a ser iguales al vector de probabilidad y este, al final del proceso evolutivo, representará la mejor solución al problema.

```

1. Generar S individuos a partir del vector de probabilidad
   For i: = 1 to s do
     S[i] := generate(p);
2. Evaluar y ordenar los individuos de tal manera que el mejor quede en
   la posición S [1].
3. Competencia del individuo en S[1] con los demás individuos.
   For j: = 2 to s do
     Begin
       Winner, loser: = compete (S[1], S[j]);
       For i: 1 to L do
         If winner[i] ≠ loser[i] then
           If winner[i] = 1 then P[i]:=P[i] + 1/n
             Else p[i]:=P[i] – 1/n;
4. Revisar si el vector de probabilidad converge
   For i: =1to L do
     If P[i]>0 and P[i] <1 then
       Return to step 2;
5. El vector de probabilidad P representa la solución final

Parámetros del algoritmo genético compacto:
L=tamaño del individuo
n=tamaño de la población

```

Fig 2. Pseudocódigo del CGA. Tomado de [9]

Es importante en este punto recalcar que si los parámetros que conforman el genoma del proceso evolutivo no están representados de manera binaria, es necesario diseñar una codificación acorde con el problema que transforme los parámetros desde su representación original a una representación binaria.

De esta manera, se hace evidente la similitud entre la propuesta presentada en la Figura 1 y el pseudocódigo de la Figura 2. Un individuo generado dentro del CGA corresponde a una descripción binaria de los parámetros que conforman un FIS completo. Para este caso particular el genoma está conformado por las medias c y desviaciones estándar σ de los conjuntos antecedentes, y por los centros y de los conjuntos consecuentes.

3 Primer caso de estudio

3.1 Problema

El problema de clasificación de la flor Iris es un problema tipo juguete bien conocido en el reconocimiento de patrones [13] [14], el cual consiste en la clasificación de tres tipos de flores Iris a partir de mediciones hechas del sépalo y pétalo de la flor. La base de datos

está compuesta por 50 instancias para cada flor, y cada instancia contiene la longitud y ancho tanto del sépalo como del pétalo. De esta manera se tienen 4 entradas al FIS.

Se propone un enfoque para el clasificador tipo controlador, como el presentado en [13], que tiene como objetivo que el FIS aproxime una función escalera con una cantidad de peldaños equivalente al número de clases existentes. Así pues, se establecen unos umbrales que al final determinan a qué clase pertenece cada instancia. Entonces, tomando como punto de referencia el trabajo desarrollado en este problema en [13], se propone que el primer escalón de la función escalera corresponda a la clase *Setosa*, el segundo a la clase *Vericolor* y el tercer escalón sea *Virginica*.

Una de las partes fundamentales de un proceso evolutivo y que es inherente al problema es la función de evaluación. La función de evaluación propuesta está compuesta por dos partes: Una función que mide el error medio cuadrático *Frmse* y otra que mide el porcentaje de instancias mal clasificadas, *Fc*. Esta función de evaluación busca que el FIS se aproxime lo mejor posible a una función escalera y además que el error de clasificación del sistema sea mínimo. En la ecuación (5) se muestra la definición de la función de evaluación *F*. *Frmse* se define en la ecuación (6)

$$F = \begin{cases} Frmse * Fc, & \text{si } Fc \neq 0 \\ Frmse, & \text{si } Fc = 0 \end{cases} \quad (5)$$

$$Frmse = \sqrt{\frac{1}{n} \sum_{i=1}^n (f_i - y_i)^2} \quad (6)$$

El subíndice *i* de la ecuación (6) hace referencia al *i*-ésimo valor que se evaluará, y el valor real que se busca y *f* el valor que se obtiene del FIS. De esta manera, la salida del FIS que se encuentre en el intervalo (0,1) pertenece a la clase 1 *Setosa*, si está en el intervalo (1,2) pertenece a la clase 2 *Vericolor* y en el intervalo (2,3) a la clase 3 *Virginica*.

3.2 Experimentos

En las siguientes tablas se compilan los valores de los parámetros que limitan los experimentos realizados. En la Tabla 1 se presenta el número de reglas y el número de entradas para los experimentos. En primera instancia el valor del número de entradas se define por el número de características definidas en la base de datos. El número de reglas se definió teniendo en cuenta el trabajo realizado en [13].

Tabla 1 Parámetros del FIS

Parámetro	Valor
Número de reglas	3, 6
Número de entradas	4

Tabla 2. Parámetros del CGA

Parámetro	Valor
Numero de experimentos	100
Tamaño de población (N)	100,500,1000
Número de individuos generados (S)	30
Método de codificación binaria	Signo-magnitud
Cantidad de bits	8 bits
Criterio de parada	Convergencia de vector de probabilidad y repetición de error
Máximo número de generaciones	10000
Función de desempeño	Función de evaluación compuesta

Tabla 3. Parámetros del GA

Parámetro	Valor
Numero de experimentos	100
Numero de generaciones	3000
Tamaño de población (N)	30
Número de individuos intermedios	60
Probabilidad de cruce	0.7
Probabilidad de mutación	0.03
Criterio de parada	Numero de generaciones
Función de desempeño	Error medio cuadrático

3.3 Resultados

Con el fin de medir porcentualmente la relación existente entre proceso evolutivo con el CGA que obtiene los mejores resultados versus el proceso evolutivo con CGA que obtiene los segundos mejores resultados y el mejor proceso evolutivo con el GA, se define (7):

$$\frac{\text{Resultado } X}{\text{Resultado del mejor CGA}} * 100\% \quad (7)$$

Donde X representa el resultado obtenido por el segundo mejor proceso evolutivo con CGA, o el mejor resultado obtenido por el proceso evolutivo con GA.

En la tabla 4 presentada a continuación se presentan los resultados obtenidos de los experimentos realizados. Esta compila los valores de la media del error de clasificación F_c , la desviación del error F_c , media y desviación estándar del número de iteraciones para converger. Y finalmente, unas medidas de efectividad. Dichas medidas corresponden: al mejor y peor resultado de error de clasificación F_c obtenido de los 100 experimentos realizados, el porcentaje de experimentos en el cual el mejor individuo posee un error de clasificación que se encuentra por debajo de la media del error.

Tabla 4. Resultados del CGA para el clasificador de la flor Iris

T1-FLS para el clasificador de la flor Iris sintonizado con CGA						
	3 reglas			6 reglas		
	n=100	n=500	n=1000	n=100	n=500	n=1000
Media Error Fc	6.3333333	1.6666666	2.0444444	3.1777777	2.2888888	2.4222222
Desviación Estándar Error Fc	9.9900187	1.5593988	1.74949443	5.0363933	1.6549909	1.5822623
Media Iteraciones	61.68	917.51	1936.99	72.3	1123.82	2401.34
Desviación Estándar Iteraciones	12.511715	135.34575	139.90166	12.86330	74.027483	180.40242
Mejor resultado	0	0	0	0	0	0
Peor resultado	42.22	6.66	13.33	35.55	6.66	8.88
% bajo media	73	39	24	74	77	84

Tabla 5. Resultado del GA para la clasificación de la flor Iris

T1-FLS para el clasificador de la flor Iris sintonizado con GA		
	3 reglas	6 reglas
Media del error Fc	5.97777778	4.71111111
Desviación Estándar del error Fc	8.270854448	7.22523463
Mejor resultado	0	0
Peor resultado	35.55555556	40
% bajo media	70	75

Los experimentos apuntan a que el mejor proceso evolutivo es el caso en el que se tiene una población n de 500 y tres reglas (CGA-1). Con respecto al segundo mejor, n de 1000 y tres reglas (CGA-2), la media del error para este proceso es 1.2266 veces (122.66%) la media del error del proceso evolutivo con CGA-1. De igual manera la desviación estándar del CGA-2 es 1.1219 veces (112.19%) la desviación de los resultados obtenidos por CGA-1. Esto implica, que de manera general, los resultados obtenidos por el CGA-2 están más dispersos que los del CGA-1. Además, la media del número de iteraciones necesarias para que el CGA-2 converja es 2.11 veces (211.11%) la media del mejor proceso. Por lo tanto, el mejor proceso es más rápido con respecto al número de iteraciones necesarias para converger.

De acuerdo con los resultados obtenidos para el GA, en la tabla 5, el mejor proceso evolutivo es el caso de un FIS con 6 reglas (GA). De esta manera, si se compara con

