

LA RAZÓN DE SOLAPE COMO MECANISMO DE ADAPTACIÓN PARAMÉTRICA BASADA EN RESTRICCIONES PARA LA OBTENCIÓN DE MODELOS BORROSOS

Miguel Angel Vélez, Omar Sánchez
Ctra. Palos de la Frontera s/n. 21819 Palos de la Frontera. Huelva
{velez}{omar}@uhu.es

Resumen

En el presente artículo se propone un método para seleccionar la razón de solape a partir del grado de pertenencia del solape de funciones de pertenencia adyacentes y la aproximación del soporte en las campanas de Gauss. Se presentan, para cumplir con el anterior objetivo, expresiones analíticas novedosas. Se propone un método de inicialización de funciones de pertenencia basado en la razón de solape. Se demuestra que el ajuste con restricciones que se aplica no deteriora de forma significativa la fase de modelado de funciones no lineales.

1 INTRODUCCIÓN

La lógica borrosa ha demostrado ser una herramienta importante para la obtención de modelos, siendo el método más popular (por su diferenciabilidad) el desborrosificador centro promedio con inferencia producto [1][2].

Los métodos de adaptación paramétrica originalmente aplicados en las redes neuronales se han extendido a la función no lineal equivalente de los sistemas borrosos, ampliándose el campo de aplicación de estos últimos virtualmente a todas las áreas donde las redes neuronales han demostrado ser eficaces (modelado, control adaptativo, control de vehículos robóticos, etc)[3][4][5].

A pesar de la extensión en la aplicación de los sistemas borrosos, la ventaja de los mismos frente a las redes neuronales no queda clara, dado que los conceptos que utiliza un experto para la construcción de un conjunto de reglas se ignoran al sustituir la definición de las funciones de pertenencia (FP) en el universo de discurso por un método alternativo de adaptación paramétrica con autoajuste, donde las redes neuronales han demostrado ser eficientes.

La ventaja fundamental en utilizar los sistemas borrosos como alternativa para el aprendizaje basado en datos de entrada-salida disponibles a través de mediciones (o estimados a través de un predictor) radica en que los parámetros que se adaptan tienen un claro significado físico, son descriptores que definen

a las FP en un universo de discurso definido sobre una variable perteneciente a un conjunto de reglas.

Algunos esfuerzos se han realizado para ajustar los parámetros adaptables de un sistema borroso a las técnicas de diseño sin aprendizaje, por ejemplo en [6] se propone un método con restricciones a la adaptación, pero esta tiene la desventaja de aplicarse a una estructura de inferencia particular denominado α -GEM.

Las ideas propuestas por Marsh en [7] se han extendido en [8] a un sistema borroso adaptable basado en el análisis del solape de las FP, considerándose los siguientes objetivos: (1) Cada punto del universo de discurso debe pertenecer al dominio de al menos una FP, adicionalmente, cada punto no debería pertenecer al dominio de más de dos FP. (2) No deben existir dos FP con el mismo punto de máximo grado de pertenencia. (3) Cuando se solapen dos FP, la suma de los grados de pertenencia para cualquier punto en el solape debe ser menor o igual a 1. (4) El solape no debe cruzar el punto de máximo grado de pertenencia de cualquiera de las FP.

En el presente artículo se incorporan nuevos elementos para la selección de la razón de solape (RS) basado en el grado de pertenencia del solape de FP adyacentes y el soporte de las campanas de Gauss. Se analizan funciones no lineales donde el método de ajuste con restricciones paramétricas no deteriora la fase de obtención del modelo.

2 ÍNDICES DE SOLAPE EN SISTEMAS BORROSOS DEL TIPO TSK

Una base de reglas borrosas se representa por:

$$R^{(l)} : \text{SI } x_1 \text{ es } F_1^l \text{ y...y } x_n \text{ es } F_n^l, \text{ ENTONCES } y \text{ es } G^l \quad (1)$$

donde $l=1,2, \dots, M$, siendo M el número de reglas, F_i^l son conjuntos borrosos en los universos de discurso $U_i \subset R$ de las variables de entrada, $i=1 \dots n$ representa las n variables de entrada, G^l es un subconjunto borroso en el universo de discurso de la

variable de salida $V \subset R$, $\underline{x}=(x_1, \dots, x_n)^T \in U_1 \times \dots \times U_n$ son las variables de entrada e $y \in V$ es la variable de salida.

El conjunto de reglas de la forma (1) es equivalente a:

$$\hat{y} = f(\underline{x}) = \frac{\sum_{l=1}^M y^l \left[\prod_{i=1}^n \mu_{F_i^l}(x_i, \beta_i^l, \gamma_i^l) \right]}{\sum_{l=1}^M \left[\prod_{i=1}^n \mu_{F_i^l}(x_i, \beta_i^l, \gamma_i^l) \right]} \quad (2)$$

donde y^l , β_i^l , γ_i^l son el consecuente, el ancho y el centro de las FP de las variables de entrada con $M + M \cdot n + M \cdot n$ parámetros ajustables, donde las FP pueden definirse según:

$$\mu_{F_i^l} = \begin{cases} \text{Gauss} & \left\{ \exp \left[- \left(\frac{x_i - \gamma_i^l}{\beta_i^l} \right)^2 \right] \right. \\ \text{Triangular} & \left. \begin{cases} 1 - \frac{2|x_i - \gamma_i^l|}{\beta_i^l} & \gamma_i^l - \frac{\beta_i^l}{2} < x_i < \gamma_i^l + \frac{\beta_i^l}{2} \\ 0 & x_i \leq \gamma_i^l - \frac{\beta_i^l}{2} \text{ ; } \\ & x_i \geq \gamma_i^l + \frac{\beta_i^l}{2} \end{cases} \right. \end{cases} \quad (3)$$

Si se define una función de coste de la forma:

$$J = \frac{1}{2} (f(\underline{x}) - y)^2 = \frac{1}{2} (\hat{y} - y)^2 \quad (5)$$

la aplicación del gradiente descendente sobre los parámetros adaptables cumple la siguiente regla:

$$\theta^{n+1} = \theta^n - \eta (\hat{y} - y) \frac{\partial (\hat{y})}{\partial \theta^n} \quad (6)$$

Siendo θ el parámetro a adaptar y η el factor de aprendizaje ($0 < \eta < 1$).

El consecuente puede ser un escalon (singleton), del tipo TSK o una función genérica dependiente o no de las variables de entrada.

Para el caso de que se seleccione un consecuente TSK, resulta:

$$y^l = f^l(x_1, x_2, \dots, x_n) = a_0^l + \sum_{i=1}^n a_i^l x_i^l = a_0^l + b^l \quad (7)$$

donde a_i^l constituyen los parámetros adaptables del consecuente TSK. Nótese que si el término $b^l = 0$ entonces el consecuente es singleton, un caso particular de esta representación.

De sustituir (3) en (2), posteriormente en (6) y desarrollando este último, la adaptación de los centros resulta:

$$\gamma_{i, n+1}^l = \gamma_{i, n}^l - \eta (\hat{y} - y) \frac{\sum_{l=1}^M \prod_{i=1}^n \mu_{F_i^l}(x_i, \beta_i^l, \gamma_i^l) \frac{2(x_i - \gamma_i^l)}{\beta_i^{2l}}}{\sum_{l=1}^M \prod_{i=1}^n \mu_{F_i^l}(x_i, \beta_i^l, \gamma_i^l)} \quad (8)$$

Los restantes parámetros se ajustan de forma similar.

La RS entre dos FP adyacentes para un mismo universo de discurso U_i responde a:

$$RS = \frac{\text{Rango de Solape (RnS)}}{\text{Rango de Definición de Funciones Adyacentes (RfA)}} \quad (9)$$

siendo RnS el intervalo del universo de discurso donde dos funciones adyacentes poseen un soporte común y RfA el soporte generado por dos funciones adyacentes.

En la Fig. 1 se puede apreciar la representación de FP triangulares y Gaussianas sobre el mismo universo de discurso.

Si se definen las variables sop_min y sop_max como los límites del soporte inferior y superior de una FP, entonces la RS para dos FP j y $j+1$ adyacentes cumple:

$$RS(j) = \frac{sop_max(j) - sop_min(j+1)}{sop_max(j+1) - sop_min(j)} \quad (10)$$

Los soportes para cada FP en el caso de FP triangulares están claramente identificados:

$$sop_min(j) = \gamma^j - \frac{\beta^j}{2}; \quad sop_max(j) = \gamma^j + \frac{\beta^j}{2} \quad (11)$$

pero para el caso de las FP Gaussianas estos soportes dependen del Grado de Pertenencia (GP) que se considere cero (una FP Gaussiana alcanza un GP cero teóricamente en el infinito).

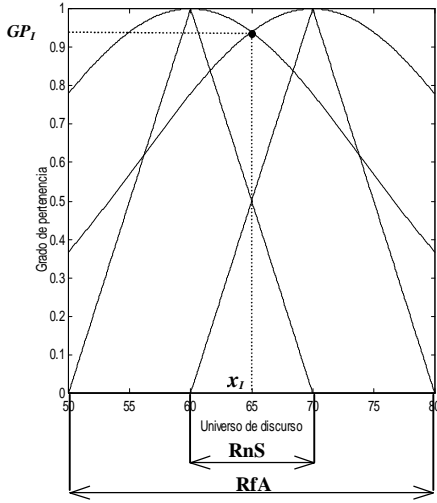


Fig. 1.-Representación de FP triangulares y Gaussianas sobre el mismo universo de discurso definidas sobre anchos $[\beta^j \beta^{j+1}] = [20 \ 20]$ y centros $[\gamma^j \gamma^{j+1}] = [60 \ 70]$.

Para aplicar (10) es necesario definir un soporte, por lo que, si se considera que la FP Gaussiana se hace cero para ψ veces el ancho (β^j), entonces (11) puede expresarse por:

$$sop_min(j) = \gamma^j - \psi \beta^j; sop_max(j) = \gamma^j + \psi \beta^j \quad (12)$$

de sustituir (12) en (10) resulta:

$$RS(j) = \frac{\gamma^j - \gamma^{j+1} + \psi(\beta^j + \beta^{j+1})}{\gamma^{j+1} - \gamma^j + \psi(\beta^j + \beta^{j+1})} \quad (13)$$

La función que nos devuelve el GP para la FP Gaussiana dado un valor ψ se obtendrá de aplicar (3), siendo $x_i = sop_min(j)$ o bien $x_i = sop_max(j)$. De ello resulta:

$$GP_\psi = \exp(-\psi^2) \quad (14)$$

donde:

$$\psi = \sqrt{-\ln(GP_\psi)}; \quad 0 < GP_\psi < 1 \quad (15)$$

Por ejemplo (ver Fig. 2), para un valor $\psi = 2$ el GP_ψ es de 0.0183 (1.83% del GP_{max}) mientras para un valor $\psi = 2.5$ el GP_ψ es de 0.0019 (0.19% del GP_{max}), siendo $GP_{max} = 1$.

Para las FP triangulares representadas en la Fig. 1, $RS=0.3333$ (de aplicar (10) y (11)). En el caso de las FP Gaussianas, para $\psi = 2$ y $\psi = 2.5$, entonces la razón de solape resulta $RS=0.7777$ y $RS=0.8181$ respectivamente (de aplicar (13)).

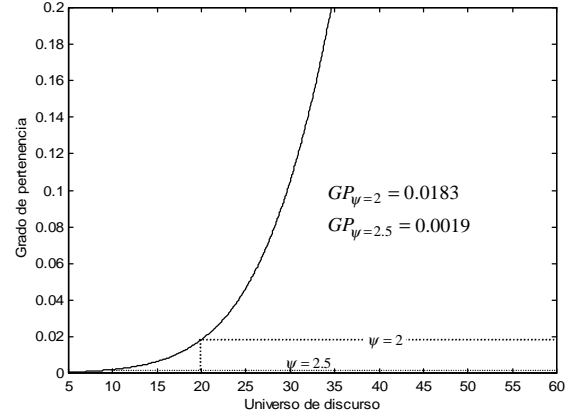


Fig. 2. Representación del GP_ψ con $\psi = 2$ y $\psi = 2.5$ para una FP Gaussiana de $\gamma = 60$ y $\beta = 20$

3 ADAPTACIÓN CON RESTRICCIONES BASADA EN LA RAZÓN DE SOLAPE

Los pasos que se siguen en la estimación de la función $f(x)$ pueden resumirse como sigue:

MIENTRAS (error(y, \hat{y}) > Valor)

- {
- Se adaptan los parámetros (consecuentes, centros y anchos);
- Estimación de la función con los parámetros adaptados;
- **Adaptación con restricciones basada en la razón de solape;**
- Estimación de la función con los parámetros adaptados con restricciones;
- Cálculo del error;
- }

donde *Valor* representa el error máximo permitido en la estimación de la función.

La adaptación con restricciones que se aplica representará un ajuste de los parámetros adaptados por gradiente de forma tal que se distribuyan los mismos para que se cumpla (13).

Considerando lo anterior, el ajuste paramétrico se realiza basándose en la siguiente función de coste:

$$J = \frac{1}{2}(RS - RS^d)^2 \quad (16)$$

donde RS^d es el valor deseado de razón de solape y RS responde a (13). La adaptación de cada parámetro resulta (de (6)):

$$\xi_{i\ n+1}^j = \xi_{i\ n}^j - \eta_{RS}(RS - RS^d) \frac{\partial(RS)}{\partial \xi_{i\ n}^j} \quad (17)$$

siendo $\xi_{i\ n}^j$ cualquiera de los parámetros γ_i^j, β_i^j . η_{RS} es el factor de aprendizaje para la razón de solape y RS^d la razón de solape que deseamos para las FP.

El resultado de aplicar (17) sobre cada parámetro implicado en el ajuste es:

$$\gamma_{i\ n+1}^j = \gamma_{i\ n}^j - \eta_{RS}(RS - RS^d) 2\psi \frac{\beta_i^j + \beta_i^{j+1}}{(\gamma_i^{j+1} - \gamma_i^j + \psi(\beta_i^j + \beta_i^{j+1}))^2} \quad (18)$$

$$\beta_{i\ n+1}^j = \beta_{i\ n}^j - \eta_{RS}(RS - RS^d) 2\psi \frac{\gamma_i^{j+1} - \gamma_i^j}{(\gamma_i^{j+1} - \gamma_i^j + \psi(\beta_i^j + \beta_i^{j+1}))^2} \quad (19)$$

$$\gamma_{i\ n+1}^{j+1} = \gamma_{i\ n}^{j+1} - \eta_{RS}(RS - RS^d) - 2\psi \frac{\beta_i^j + \beta_i^{j+1}}{(\gamma_i^{j+1} - \gamma_i^j + \psi(\beta_i^j + \beta_i^{j+1}))^2} \quad (20)$$

$$\beta_{i\ n+1}^{j+1} = \beta_{i\ n}^{j+1} - \eta_{RS}(RS - RS^d) 2\psi \frac{\gamma_i^{j+1} - \gamma_i^j}{(\gamma_i^{j+1} - \gamma_i^j + \psi(\beta_i^j + \beta_i^{j+1}))^2} \quad (21)$$

Además de ψ otro parámetro crítico para la adaptación con restricciones es la elección de la RS^d . En un universo de discurso normalizado lo óptimo sería que los centros estuviesen uniformemente distribuidos entre -1 y 1 y los anchos fueran tales que la RS entre cada dos FP adyacentes fuese RS^d .

Para M reglas, la distribución de centros óptima sería:

$$\text{Centro}(l) = \frac{2(l-1)}{M-1} - 1; \quad l=1..M \quad (22)$$

y todos los anchos tendrían como valor:

$$\text{Ancho}(l) = \frac{(RS^d + 1)}{\psi(1 - RS^d)(M-1)}; \quad l=1..M \quad (23)$$

El punto de intersección entre dos FP adyacentes debería ser tal que su GP (GP_I , ver Fig. 1) sea inferior o igual a 0.5 para que se cumpla el tercer objetivo en la búsqueda de un adecuado solape de las FP. El punto en el que dos FP adyacentes se

interceptan (x_I , ver Fig. 1) se obtiene de igualar dos FP adyacentes del tipo (3) en las que los centros y anchos serán los correspondientes a aplicar (22) y (23). El punto de intersección dependerá de las dos FP adyacentes que se elijan de las M existentes y del número de reglas M . Considerando lo anterior se cumple:

$$x_I^l = \frac{2l - M}{M - 1} \quad (24)$$

Sustituyendo (22), (23) y (24) en (3) resulta:

$$GP_I = \exp\left(-\left(\frac{\psi(1 - RS^d)}{(RS^d + 1)}\right)^2\right) = \exp\left(-\left(\frac{\sqrt{-\ln(GP_\psi)}(1 - RS^d)}{(RS^d + 1)}\right)^2\right) \quad (25)$$

Si se fija una RS^d se tendrá para cada posible valor de GP_ψ un valor de GP_I . En la Fig. 3 se muestra la evolución del GP_I en función del GP_ψ para distintas RS^d , lo que facilita la elección de la RS^d una vez decidido el GP_ψ y el GP_I .

Por ejemplo, para un $GP_\psi = 0.013$ y un $GP_I = 0.45$, la RS^d será de 0.4, tal y como se muestra en la Fig. 3.

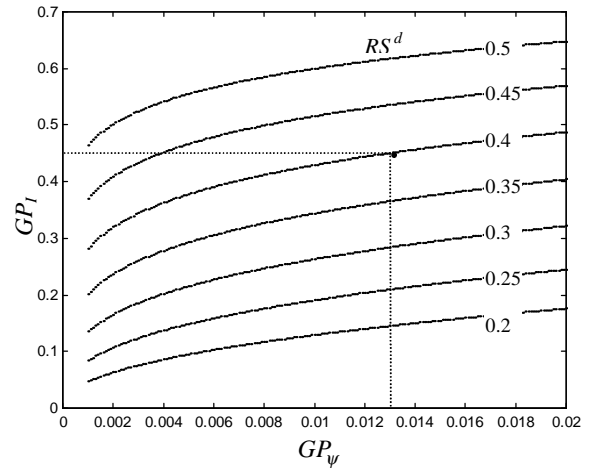


Fig. 3. Gráfica para determinar la RS^d para cada dos FP adyacentes dado un GP_ψ y un GP_I .

4 APLICACIÓN

Ejemplo 1: El modelo serie-temporal de Box y Jenkins que se obtiene a partir de datos de un horno de gas responde a:

$$y_t = 0.55y_{t-1} - 0.52u_{t-3} - 0.4u_{t-4} - 0.51u_{t-5} + N_t \quad (26)$$

donde:

$$N_t = 1.53N_{t-1} - 0.63N_{t-2} + a_t \quad (27)$$

siendo a_t un ruido blanco, N_t el ruido en el sistema, y_t la concentración de dióxido de carbono y u_t el flujo de metano en el instante t .

El sistema puede representarse por un modelo serie-paralelo borroso de la forma:

$$\hat{y} = f(y_{t-1}, u_{t-4}) \quad (28)$$

Al modelar sin restricción paramétrica utilizando consecuente TSK, un número de reglas $M=4$, $\eta=0.001$ y un número de épocas igual a 40, se obtiene el resultado que se muestra en la Fig. 4, con un error de 0.03, para la variable de entrada u_{t-4} .

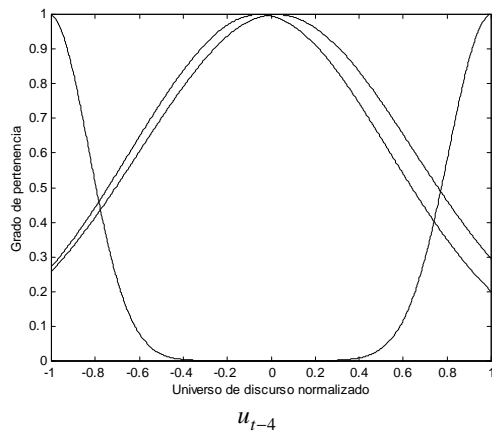


Fig. 4. Representación de las FP para la variable u_{t-4} en el modelo de Box y Jenkins.

Si se aplica la adaptación con restricciones basado en la razón de solape para $M=4$, $\eta=0.001$, factor de aprendizaje para la razón de solape $\eta_{RS}=0.6$, una razón de solape deseada $RS^d=0.4$, un $\psi=2$ y buscando un error ≤ 0.03 , se obtiene una distribución de FP que se muestra en la Fig. 5 para 41 épocas.

Nótese la diferencia entre las distribuciones de las FP para las Fig. 4 y 5. En la primera, los puntos en los que dos FP alcanzan su GP_{max} están muy próximos. Adicionalmente, en la totalidad de la zona de solape la suma de los grados de pertenencia es mayor a 1 y existen puntos del dominio en los cuales están definidos más de dos FP. De lo anterior se deduce que tres de las cuatro condiciones para un adecuado solape son violadas. El resultado de una adaptación paramétrica con restricciones (Fig. 5) permite que las

exigencias al solape entre las FP en el universo de discurso se cumplan.

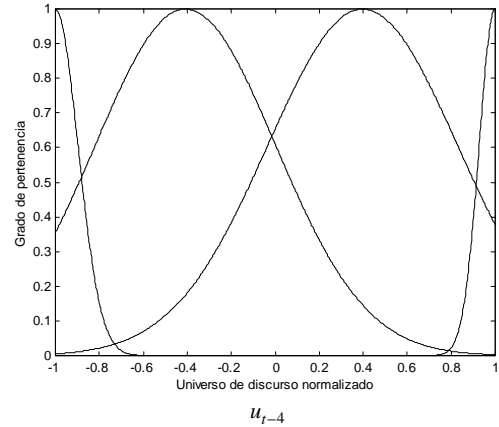


Fig. 5. Representación de las FP para la variable u_{t-4} en el modelo de Box y Jenkins usando restricción paramétrica basada en la razón de solape.

En la Fig. 6 se puede apreciar la comparación entre la serie temporal de Box y Jenkins y la estimación realizada de la función con la aplicación de restricción paramétrica basada en la razón de solape. Nótese que la estimación puede ser válida cuando en el modelado se aplica adaptación con restricciones, como se demuestra en la Fig. 6.

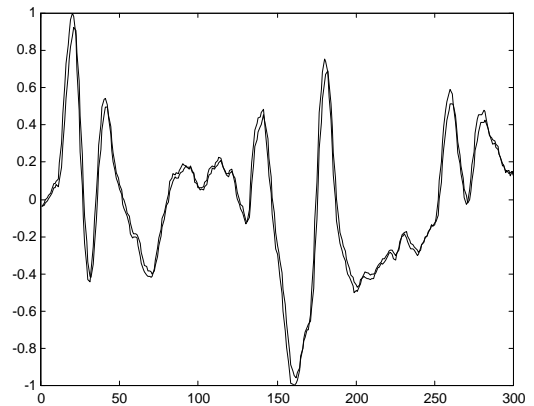


Fig. 6. Representación de la función de Box y Jenkins y de la función estimada mediante restricción paramétrica basada en la razón de solape.

Ejemplo 2: Para realizar una comparación entre la evolución del error utilizando restricción mediante razón de solape y sin usarla, se toman como referencia cuatro funciones no lineales, que se describen en la Tabla 1.

Sistema	y=
I	$2x_1^2 + 4x_2^3$
II	$4\text{sen}(\pi x_1) + 2\text{cos}(\pi x_2)$
III	$(3\exp(3x_1) + 2\exp(-4x_2))^{-0.5}$
IV	$\frac{\text{sen}(x_1)}{x_1} \frac{\text{cos}(x_2)}{x_2}$

Tabla 1. Funciones no lineales utilizadas como referencia para el modelado

Para analizar la influencia de la incorporación de la adaptación con restricciones basada en la razón de solape en la obtención del modelo, se usan los mismos parámetros para todas las funciones: 40 épocas, $M=4$ reglas, $\eta = 0.005$, $\eta_{RS} = 0.1$, $RS^d = 0.4$ y $\psi = 2.5$.

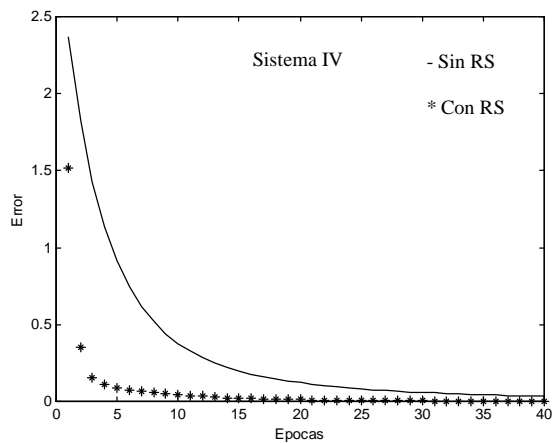
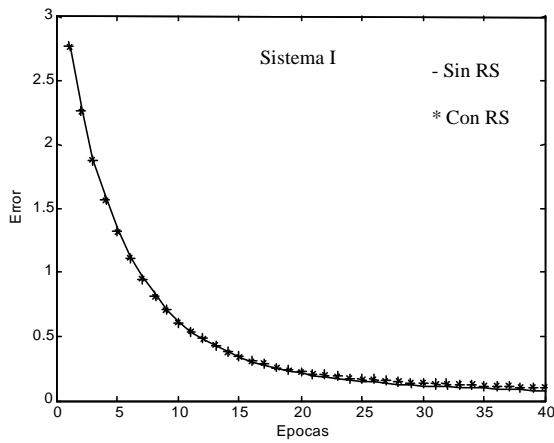


Fig. 7 y 8.- Comparación de la evolución del error aplicando y sin aplicar adaptación paramétrica con restricciones basada en la razón de solape para los sistemas I y IV.

Se puede apreciar en las Fig. 7 y 8 como el error cuando se usa adaptación con restricciones basada en la razón de solape evoluciona favorablemente, aunque consiguiendo generalmente peores resultados transcurridos un determinado número de épocas, pero ese será el coste para conseguir unas FP con un grado de solape más adecuado.

Como se aprecia en las Fig. 7 y 8, la incorporación de la adaptación basada en la razón de solape no incorpora una diferencia significativa en la obtención del modelo, en el caso del sistema IV incluso lo mejora.

En todos los casos la evolución del error no se deteriora, tal y como muestra la Fig. 9.

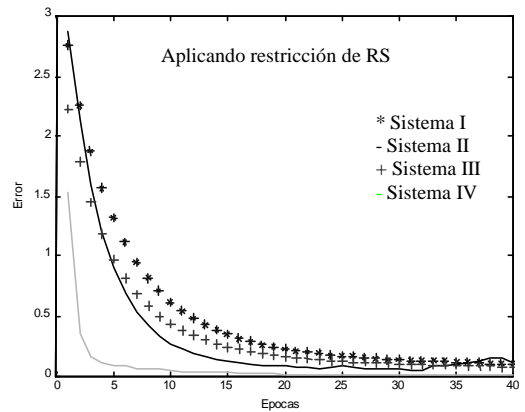


Fig. 9.- Comparación de la evolución del error para los cuatro sistemas del ejemplo aplicando restricción paramétrica basada en la razón de solape

5 CONCLUSIONES Y TRABAJOS FUTUROS.

Se ha propuesto un método para la selección de la razón de solape basado en el grado de pertenencia del solape de FP adyacentes y el soporte variable de las campanas de Gauss, se presentan expresiones novedosas para el anterior análisis; adicionalmente, se propone un método para establecer las condiciones iniciales de los parámetros a adaptar basado en la RS. Se ha presentado el resultado de simulación en la fase de modelado de cinco funciones no lineales sin que exista un deterioro significativo con respecto al modelado sin restricciones paramétricas basadas en la RS.

La incorporación del ajuste paramétrico permite incorporar conceptos de un diseñador heurístico a la fase de adaptación por gradiente.

En trabajos futuros se pretende:

1.- Extender el ajuste paramétrico a los consecuentes de tipo TSK considerando el carácter singleton de la formulación no lineal.

2.- Aplicar el concepto de robustez del solape, con restricciones paramétricas basada en el área de funciones adyacentes.

3.- Incorporar los dos tipos de formulaciones anteriores a las restricciones aquí expuestas.

6 REFERENCIAS

[1] Wang, L.X.: "Adaptive fuzzy system and control". Prentice Hall. 1996.

[2] Jang, J.; Sun, C.; Mizutani, E.: "Neurofuzzy and soft computing". Prentice Hall. 1997.

[3] Hsu, Y.; Chen, G.: "Fuzzy dynamical modeling techniques for nonlinear control systems and their applications to MIMO systems". Fuzzy Theory Systems: Techniques and Applications", Vol. 1. Academic Press. Págs. 48-84. 1999.

[4] Ying, T.; Lee, C.: "Fuzzy Model-Reference Adaptive Control". IEEE Transactions on Systems, Man and Cybernetics, vol.25, No. 12, págs. 1606-1615. 1995.

[5] Sánchez, O.; Ollero, A. Heredia, G.: "Hierarchical Fuzzy Path Tracking and Velocity Control of Autonomous Vehicles". Integrated Computer-Aided Engineering. No. 6 Vol. 4. págs 289-301. 1999.

[6] Uehara, K.; Hirota, K.: "Parallel fuzzy inference based on α -level sets and generalized means". Information Sciences, Vol. 100, No. 1-4. págs. 165-206. Elsevier Science. 1997.

[7] Motorola: "Fuzzy Logic Program 2.0". 1994.

[8] Sánchez, O.; Ollero, A.; Aracil, J.: "Ajuste de Funciones de Pertenencia para la Obtención de Modelos Borrosos de Sistemas no Lineales". XIX Jornadas de Automática. Madrid. 1998.