Modelado con selección de estructura automática para sistemas no lineales usando SVM difusos

Julio César Tovar Rodríguez

ESIME Zacatenco, Instituto Politécnico Nacional (IPN),
Av. IPN s/n, Lindavista, Gustavo A. Madero, C.P. 07738, México DF, México
jctovar@ipn.mx

Resumen En este trabajo se estudia la identificación de sistemas dinámicos no lineales usando redes neuronales difusas. Se enfoca en ambas incertidumbres estructurales y paramétrica las cuales han sido exploradas considerablemente en la literatura de identificación de sistemas no lineales. La contribución principal es que la estructura analítica integral se propuso para la selección de estructura de redes neuronales difusas, los parámetros de identificación y el intercambio de histéresis en la red garantiza una ejecución de identificación neuronal. Primeramente, un support vector machine autómata, se propone dentro de un intervalo de tiempo fijo para un criterio dado de construcción de la red. Luego la actualización de los parámetros de la red garantiza un error acotado. La incertidumbre se hace frente con una estrategia de histéresis para habilitar el intercambio del identificador neuro difuso que garantiza la ejecución de la red a lo largo del proceso de intercambio. El análisis teórico y el ejemplo de simulación muestran la eficacia del método propuesto.

Palalbras clave: Red neuronal artificial, *Support Vector Machine* (SVM), lógica difusa, identificación, histéresis.

1. Introducción

Las redes neuro difusas han sido usadas ampliamente en predicción de series de tiempo, modelado de sistemas no lineales y control [1]. Está trabajo se enfoca en el modelado de sistemas no lineales dinámicos con múltiples redes neuro difusas. El problema principal del modelado difuso es la extracción de las reglas, estas se pueden dividir dentro de dos clases [2]: 1) la obtención de las reglas de expertos [3], 2) la obtención de las reglas difusas automáticamente de los datos observados. El método experto usa el criterio imparcial y la técnica de prueba y error, está se aplica fuera de línea.

El proceso de la extracción de reglas difusas para el modelado de sistemas no lineales es llamado estructura de identificación. Un método común es la partición de los datos de entrada y salida, a esto se le conoce como cuadrícula difusa. Estos enfoques requieren que los datos estén listos antes del modelado. Existen pocos métodos de agrupamiento en línea en la literatura.

La idea básica del modelado SVM es el mapeo de las entradas dentro de un espacio característico de dimensión superior, entonces se resuelve la programación cuadrática (QP) con una función de costo apropiada. Solo las soluciones no ceros los cuales son llamados support vectors son útiles para el modelado. A la combinación de multimodelos y redes neuronales difusas se les considera como un enfoque racional para identificación de sistemas no lineales. Se concluye que en sistemas en lazo cerrado son globalmente estables en el sentido que todas las señales involucradas están uniformemente acotadas [4].

Se verá una elección de estructura automatizada se propone primero dentro de un intervalo de tiempo fijo para un criterio de construcción de la red dada. Entonces el algoritmo de actualización de los parámetros de la red se propone con el error de identificación acotado garantizado. Para cubrir con la estructura de incertidumbre, una estrategia de histéresis se desarrolla para habilitar la conmutación del identificador neuro difuso con la ejecución garantizada a lo largo de la conmutación del proceso.

2. Múltiple redes neuronales difusas para identificación de sistemas no lineales

Una planta no lineal discreta puede ser representada por

$$y(k) = f[x(k), \theta] + e(k) \tag{1}$$

donde

$$\begin{aligned} x(k) &= [y(k-1), y(k-2), \cdots, y(k-n_y), \\ &\quad u(k-d), u(k-d-1), \cdots, u(k-d-n_u)]^T \in \Re^{n\times 1} \end{aligned} \tag{2}$$

y $f(\cdot)$ es una ecuación diferencial no lineal desconocida representando las dinámicas de la planta, $u(k)$ y $y(k)$ son escalares medibles entrada y salida, d es el retardo, θ es un vector de parámetros desconocidos asociado con una estructura del modelo apropiada, $e(k)$ es un ruido de observación acotado, n_y y n_u son las longitudes de salida y entrada, $n_y + n_u = n$. De hecho, (1) es un modelo NARX [2].

Un modelo difuso genérico se presenta como una colección de reglas difusas en la siguiente forma

$$\begin{aligned} &R^j : IF\ x_1\ is\ A_1^j\ and\ x_2\ is\ A_2^j\ and \cdots x_n \\ &is\ A_n^j\ THEN\ y\ is\ B^j \end{aligned} \tag{3}$$

Se usa $l(j=1,2,\cdots,l)$ reglas difusas IF-THEN para realizar un mapeo del vector lingüístico de la entrada $x=[x_1,\cdots,x_n]\in\Re^n$ a la salida ŷ. Se sabe, usando el producto inferencia, promedio de centros y el difusificador singleton, la p-ésima salida del sistema lógico difuso se puede expresar como

$$y = \left(\sum_{j=1}^{l} w_j \left[\prod_{i=1}^{n} \mu_{A_i^j} \right] \right) \Big/ \left(\sum_{j=1}^{l} \left[\prod_{i=1}^{n} \mu_{A_i^j} \right] \right) = W\varphi[V, x(k)] \tag{4}$$

Generalmente las redes neuronales difusas (4) no puede coincidir dado un sistema no lineal (1) exactamente, el sistema no lineal (1) se puede representar como:

$$y(k) = W\varphi[V, x(k)] + \varepsilon(k) \tag{5}$$

donde $\varepsilon(k)$ se define como el error de modelado. El sistema no lineal (1) puede ser escrita como

$$y(k) = W^0\varphi[V^0, x(k)] + f(k) \tag{6}$$

donde $\varepsilon(k)$ es el error modelado, V^0 y W^0 son conjuntos de $f(k)$ parámetros conocidos escogidos por el usuario.

De (6) se conoce el error modelado f_t depende de la estructura de la red neuronal difusa. Para algunos procesos no lineales, sus condiciones de operación varían con el tiempo o su entorno de operación es complicado, y un modelo no es suficiente para describir la planta completa. Los múltiples modelos pueden dar una mejor exactitud de identificación. Aunque una simple red neuronal difusa (4) puede identificar cualquier proceso no lineal (caja negra), el error de identificación puede ser grande si la estructura de la red difusa no se escoge apropiadamente. Generalmente hablando, no se puede encontrar la estructura de la red optima representando el sistema (1) bajo todas las condiciones diferentes de operación. Una solución posible es usar varias redes y seleccionar la mejor por el algoritmo de conmutación apropiada. La estructura de los identificadores múltiples neuronales se muestra en la figura 1.

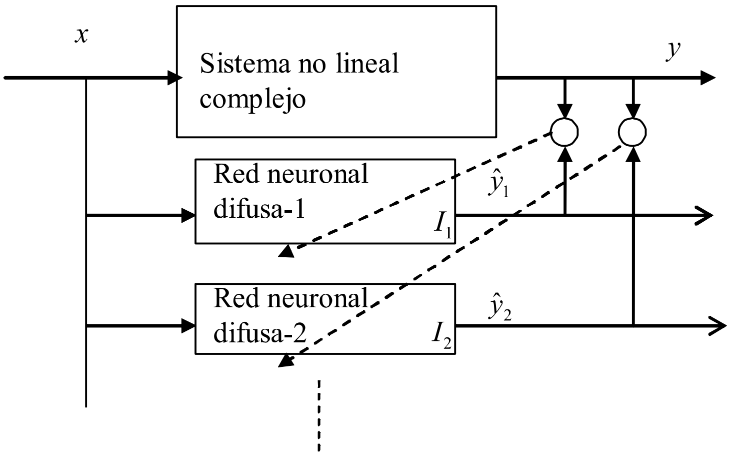

Fig. 1. Identificador múltiple neuro difuso

Aquí I_1, I_2 son los dos identificadores neuro difuso, cuyas salidas son $\hat{y}_1$, $\hat{y}_2$. Se usa una política de conmutación para escoger un identificador I_i tal que el error de identificación entre la salida de esta red neuronal y la planta ($\hat{y}_i(k)$-$y(k)$) se minimiza. Los identificadores múltiples neuro difusos son presentados como

$$I_\sigma : y_\sigma(k) = W^\sigma(k)\varphi[V^\sigma(k)x(k)] \tag{7}$$

donde $\sigma=\{1,2,\cdots r\}$ y r son el número total de identificadores neuro difusos.

El objetivo de la identificación múltiples redes neuro difusas es diseñar redes neuro difusas apropiadas y la política de conmutación tal que un índice de ejecución es minimizado y la conmutación infinita no ocurre.

3. Selección de estructura automática

Un SVM puede separar los datos dentro de dos clases con un hiperplano de margen máximo [4]. Si el entrenamiento es separable por el hiperplano, la función es escogida como $f(x)=(w \cdot x)+b$. El margen se define como la distancia mínima desde una muestra hasta la superficie de resolución. El margen en turno se puede medir por la longitud del vector $\boldsymbol{w}$, tal que los puntos cercanos al hiperplano satisface $|(w \cdot x)+b|=1$.

Para N muestras $\{x(k),y(k)\}^{N}_{k=1}$, se usa SVM para aproximar una función no lineal. Considerar la regresión en un conjunto de funciones no lineales

$$f(x) = w^T f(x) + b \tag{8}$$

donde el kernel $K(x,x_k)=\varphi(x)^T\varphi(x_k)$. Existen muchas posibilidades para escoger el kernel $K(x,x_k)$, el cual requiere la condición de Mercer [4]. Por ejemplo kernel lineal $K(x,x_k)=x^T{}_k x$, kernel MLP $(x,x_k)=tanh(k_1 x^T{}_k x+k_2)$, kernel RBF $K(x,x_k)=exp(-||x-x_k||^2/\sigma^2)$. En esta sección, se usará el kernel difuso el cual se define como $K(x,x_k)=\{u(x_i,x_j)\},\ \ i,j=1,\cdots N$

donde u_i es una función de membresía. Se usa la función Gaussiana:

$$u(x_i,x_j) = \exp\left(\frac{\left\|x_i - x_j\right\|^2}{2\sigma^2}\right) \tag{9}$$

La regresión es resolver el problema siguiente

$$\begin{aligned} minJ_p &= \frac{1}{2} w^T w + C\sum_{k=1}^{n} s_k \xi_k \\ sujeto\ a &\left| y_k - (w^T \varphi(x_k) + b) \right| \le \varepsilon \end{aligned} \tag{10}$$

donde $\xi_k \geq 0$, $k=1,2,\ldots,N$, C es constante, ξ_i es la variable débil. s_k es un factor difuso. Este denota el grado de importancia de la muestras x_i para el aprendizaje del hiperplano óptimo en SVM. Se selecciona si como la función forma de campana ver Figura 2. Así el s_k más pequeño es, el más pequeño efecto que la muestra x_k.

4. Selección de modelos

Se define el índice del error de desempeño $J_p(k)$ para el p--ésimo identificador neuronal como

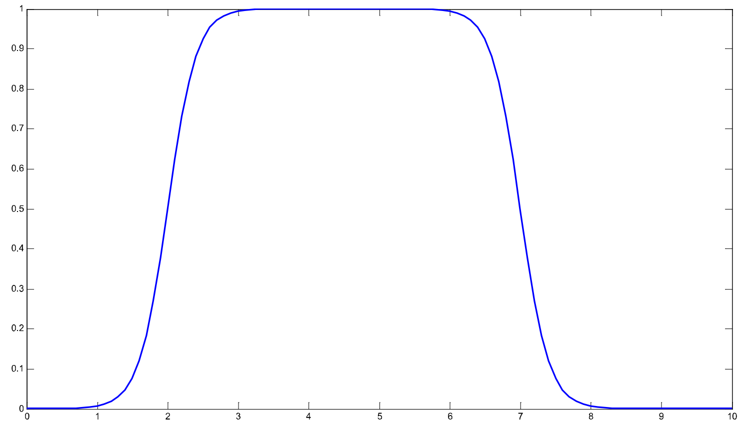

Fig. 2. Función de membresía en forma de campana generalizada.

$$J_p(k) \square ae^2(k) + (1-\alpha) \sum_{t=k-N_p}^{k} e^2(t) \tag{11}$$

donde α es un parámetro de diseño para ambos errores de pesos término corto y término largo, N_p es la longitud del p--ésimo identificador neuro difuso. La decisión de la estructura de conmutación (cambio) está dirigido por el monitoreo del índice de desempeño (11). Para prevenir un cambio rápido arbitrariamente debido a las perturbaciones, un algoritmo de decisión de histéresis es necesario. En este trabajo, se cambiara el modelo de la red neuro difusa solo cuando las funciones de membresía son casi convergentes. Se define la función de cambio como (k>N)

$$\begin{gathered}\omega(k) \square \frac{1}{2}\left| tr(W^T(k)W(k)) - \frac{1}{k-N}\sum\nolimits_{t=N}^{k} tr(W^T(t)W(t)) \right| \\ \frac{1}{2}\left| tr(V^T(k)V(k)) - \frac{1}{k-N}\sum\nolimits_{t=N}^{k} tr(V^T(t)V(t)) \right|\end{gathered} \tag{12}$$

donde $V^T(k)=[c_{ji}(k),\sigma_{ji}(k))]$. El algoritmo de decisión de histéresis puede ser formulado como

$$\begin{cases} r+1 \text{ si } J(k) > J(k-1) + h \text{ y } \omega(k) \le L \\ r \quad \text{otra manera} \end{cases} \tag{13}$$

donde $h>0$ es la constante de histéresis, L es el umbral para el cambio de pesos. $r=r+1$ define la estructura de la r--ésima red neuronal no es apropiada para el dato entrante, se debe comenzar una red neuronal, ahora $\sigma=\{1,2,\cdots r,r+1\}$. Se escoge L tal que el algoritmo de conmutación (13) pueda trabajar después de los parámetros no afecten significativamente la identificación neuro difusa.

En general incluso para sistemas lineales conmutados invariantes en el tiempo, la estabilidad de cada componente del sistema en lazo conmutado no garantiza la estabilidad del sistema conmutado entero bajo leyes de conmutación arbitrarios. Algunas condiciones sobre la política de conmutación es muy especial: solo pueden conmutar desde el modelo r hasta el modelo $r+1$ en el eje de tiempo. La conmutación es una secuencia. El teorema asegura que el error de identificación en cada subsistema es estable. Puesto que los subsistemas son conmutados uno por uno, se discutirá la convergencia de nuestro esquema de identificación multi neuronal.

Lema 1. *(a) Para cada tiempo finito k_1, existe por lo menos un modelo $r_0 \in \sigma$ tal que el índice de desempeño $Jr_0(k)$ en (11) está acotado en $[1,k_1)$*

(b) Para cada $i \in \sigma$ el índice de desempeño $Ji(k)$ en (11) tiene un límite (puede ser infinito) como $k \to k_1$. Sea f una función está definida sobre un intervalo abierto conteniendo a c (excepto posiblemente a c.) EL argumento "Límite Infinito" $\lim_{x \to c} f(x)=\infty$ significa que para cada $M>0$ existe una δ tal que $f(x)>M$ siempre que $0<|x-c|<\delta$.

Teorema 1. *Si se usan las multi redes neuronales como en (7) y el algoritmo de conmutación de histéresis como en (13), entonces (a) existe un tiempo finito $k*$ después de que para todo $k > k*$, $Jr(k)$ está acotada en $[1,K)$, $K > k*$, (b) el proceso de conmutación de los modelos neuronales paran, i.e., r es constante.*

5. Resultados

La planta a identificar es

$$y(k) = \frac{y(k-1)y(k-2)[y(k-1)+2.5]}{(1+y(k-1)^2+y(k-2)^2)} + u(k-1) \quad (14)$$

La señal de entrada entrenada es seleccionada como números aleatorios en el intervalo [0,1].

$$\begin{aligned} X(k) &= [y(k-1), y(k-2), u(k-1)]^T \\ &= [x_1(k), x_2(k), x_3(k)]^T \end{aligned}$$

Se selecciona $\alpha=0.4$. Como conocido a priori, se conocen los cambios máximos en la entrada y la salida son alrededor de 3 y 1, $\alpha\|x_{max}-x_{min}\|+(1-\alpha)\|y_{max}-y_{min}\|=1.8\alpha$. Así L podría ser escogida tal que $L<1.8$, en esta aplicación se selecciona $L=1.5$

El resultado utilizando el enfoque multimodelo y el intercambio de cada grupo se muestra en la figura 3.

6. Conclusión

En este artículo se propuso un nuevo kernel para la estructura de los support vector machines estándar, y se observó que en el proceso de extracción de reglas difusas para el modelado de sistemas no lineales es llamado estructura de identificación. A la combinación de multimodelo y redes neuronales difusas se les considera como un enfoque racional para identificación de sistemas no lineales. Se muestra un rango de aprendizaje variante en el tiempo para uso común del algoritmo retro-propagación (*backpropagation*), se prueba que los errores de identificación están acotados, el cual cambiará el modelo de la red neuro difusa solo cuando las funciones de membresía son casi iguales.

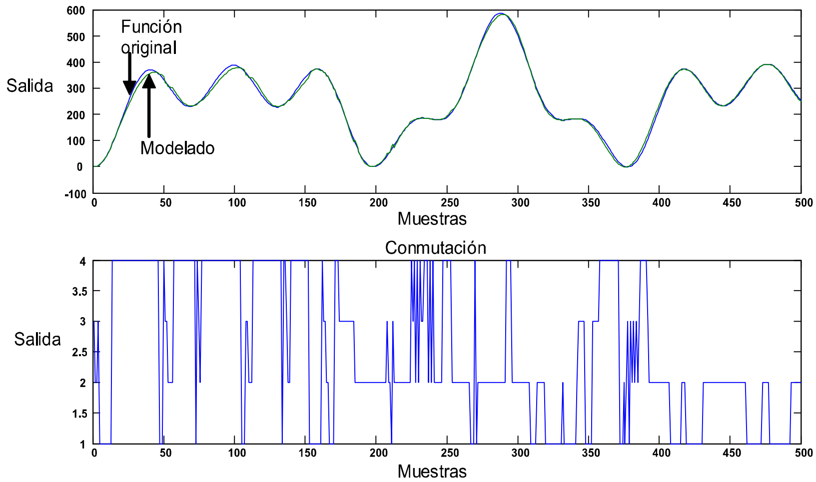

Fig. 3. Enfoque multimodelo del sistema no lineal

Referencias

1. Wang, L.X.: Adaptive Fuzzy Systems and Control. Englewood Cliffs NJ: Prentice-Hall (1994)
2. Leski, Jacek M.: TSK-Fuzzy Modeling Based on ε-Insensitive Learning. IEEE Trans. on Fuzzy System, vol. 13, no. 2, pp181-193 (2005)
3. Rivals I. and Personnaz, L.: Neural-network construction and selection in nonlinear modeling. IEEE Transactions on Neural Networks, Vol.14, No.4, 804-820 (2003)
4. Tovar, J., Wen Yu, Ortiz, F., Román Mariaca, C., and Rubio, J.J.: Modeling via on-line clustering and fuzzy support vector machines for nonlinear system. In: 50th IEEE Conference on Decision and Control and European Control Conference, December 12-15, Orlando, FL, USA (2011)