

Diagnóstico del cáncer de mama empleando clasificador difuso

Juan Antonio Contreras Montes ^{a*} Yuliana Puerta Cruz ^a Roger Misa Llorca ^b

^a*Programa de Ingeniería de Sistemas, Corporación Universitaria Rafael Núñez,
Cartagena, Colombia.*

^b*Instituto Superior Politécnico José Antonio Echeverría, La Habana, Cuba.*

**epcontrerasj@ieee.org;*

(Recibido: Mayo 6 de 2008- Aceptado: Julio 4 de 2008)

RESUMEN

Los sistemas de clasificación difusa generados a partir de datos experimentales presentan una alta precisión pero a costa de sacrificar la integridad semántica del modelo difuso obtenido. En este artículo se presenta un método novedoso para generar sistemas de clasificación difusa a partir de datos, que superan el conflicto entre precisión e interpretabilidad, obteniendo modelos con particiones triangulares de solapamiento 0.5 en sus antecedentes y consecuentes tipo singleton. Para la ponderación de los antecedentes se utiliza un operador de combinación en vez de una T-norma, lo que contribuye a una reducción sustancial en el número de reglas.

PALABRAS CLAVE: Lógica difusa, clasificación, cáncer de mama, interpretabilidad, mínimos cuadrados.

Breast cancer diagnosis using fuzzy classifier

ABSTRACT

Fuzzy Classifiers obtained from experimental data guarantee a good numerical accuracy but the semantic integrity is not considered. A novel method for obtaining fuzzy models from data, with a satisfactory tradeoff between accuracy and interpretability is presented. The antecedent partition uses triangular membership functions with overlap 0.5 and singletons for consequents. Operators Weighted Averaging are used for combining the antecedents.

KEYWORDS: Fuzzy logic, Classification, Breast cancer, Interpretability, Square Means

1. INTRODUCCIÓN

La lógica difusa ha surgido como una importante herramienta para la construcción de sistemas de clasificación. Los clasificadores difusos pueden ser construidos a partir de la información suministrada por un experto, lo cual garantiza una alta interpretabilidad pero con un bajo nivel de precisión, o a partir de datos experimentales. Esta última técnica ha mostrado ser muy eficaz en la obtención de clasificadores con alta precisión pero sacrifica la integridad semántica (interpretabilidad), una de las principales cualidades de la lógica difusa.

Existen tres criterios básicos que definen la interpretabilidad de un modelo difuso [1]-[2]:

- Distinguible: cada etiqueta lingüística debe tener significado semántico y cada conjunto difuso debe tener un rango definido en el universo de discurso. Las funciones de pertenencia deben ser claramente diferentes.
- Número de funciones de pertenencia: no debe exceder de nueve términos distintos
- Número de reglas: debe ser limitado, aunque en este punto no existe consenso sobre un número máximo.

Sin embargo, la gran mayoría de las propuestas encontradas en la literatura científica sobre identificación y clasificación sacrifican la interpretabilidad del sistema en la búsqueda de alta precisión, llegando a obtener particiones con solapamientos mayores a uno, solapamiento de más de dos conjuntos, elevado número de reglas y, en algunos casos, conjuntos no convexos. En [1] se afirma que existen unas tendencias basadas en la precisión de la aproximación o error local y otras basadas en técnicas de agrupamiento borroso. “El resultado de la combinación de esas técnicas es un modelo con buena aproximación numérica pero muy pobre en el contexto de la información lingüística”. Por otro lado, en [3] se afirma que “un costo en que se incurre en la obtención de precisión, es que los modelos resultantes fallan en la interpretación lingüística...”. En [4] se comenta que las técnicas para extracción de reglas difusas a partir de los datos se apoyan principalmente en métodos numéricos de entrenamiento, tales como redes

neuronales o algoritmos evolutivos. “Estos métodos buscan optimizar el desempeño numérico despreocupándose de la interpretabilidad de las reglas inducidas”

No obstante, se aprecia una marcada preocupación en la búsqueda de una metodología que permita extraer, a partir de datos experimentales de entrada y salida, modelos borrosos precisos e interpretables.

La novedad del método presentado en este documento radica en la consecución de modelos borrosos interpretables con alta precisión y bajo número de parámetros y se basa en el método de error de inferencia planteado por Sala [5]. El método ha sido aplicado de manera exitosa a problemas clásicos como el horno de gas de Box-Jenkins [6] y la serie caótica de Mackey-Glass [7]. En este artículo los autores muestran la capacidad del método que proponen para detección de clases y su aplicación a la solución de problemas clásicos como la detección de cáncer de mama a partir de información extraída de la base de datos WBCD (Wisconsin Breast Cancer Data), la cual está disponible en la Universidad de California, Irvine (URL:<http://www.ics.uci.edu/mllearn/>).

Dentro de la literatura científica sobre sistemas de clasificación difusos aplicado al problema del cáncer de mama (WBCD) se encuentran técnicas diferentes para la extracción de la información [8]-[11], basadas principalmente en redes neuronales, algoritmos genéticos y sistemas neurodifusos.

2. MATERIALES Y MÉTODOS

2.1 Estructura del Modelo difuso

Buscando precisión e interpretabilidad del sistema borroso se deben considerar los siguientes criterios para la selección de los parámetros:

a) *Funciones de pertenencia*: La partición de los universos de las variables de entrada en el proceso de aprendizaje se hará con conjuntos triangulares normalizados con solapamiento específico de 0.5. La razón del uso de este tipo de funciones de pertenencia se debe a que las funciones de pertenencia triangulares permiten la

reconstrucción del valor lingüístico en el mismo valor numérico, luego de aplicar un método de defuzzyficación [12]; además, el solapamiento en 0.5 asegura que los soportes de los conjuntos difusos sean diferentes. Los conjuntos difusos generados para la variable de salida serán tipo singleton.

b) *Distribución de las funciones de pertenencia:* Los conjuntos difusos triangulares de las variables de entrada estarán distribuidos de manera simétrica en cada universo respectivo, de tal manera que cada uno de los elementos de los universos de discurso de cada variable debe pertenecer, al menos, a un conjunto difuso.

c) *Operadores:* Para la combinación de los antecedentes de las reglas se emplearán operadores tipo OWA o de promedio ponderado, el cual promedia los grados de pertenencia de cada antecedente en evaluación.

d) *Método de inferencia:* El método de inferencia empleado está dado por

$$f(x^{(i)}) = \frac{\sum_{j=1}^L \bar{y}^j m_j(x^{(i)})}{\sum_{j=1}^L m_j(x^{(i)})} \quad (1)$$

donde

$$m_j(x^{(i)}) = u_{A_1^j}(x_1^{(i)}) \cdot u_{A_2^j}(x_2^{(i)}) \dots u_{A_n^j}(x_n^{(i)}) \quad (2)$$

es el grado de cumplimiento de los antecedentes de la j -ésima regla de un sistema difuso tipo Sugeno, \bar{y}^j es el valor del singleton correspondiente a la regla j . El denominador siempre arroja un valor igual a 1 cuando se trata de particiones suma 1, el cual es el caso del método de aproximación propuesto en este artículo.

2.2 Algoritmo de Identificación difusa

El algoritmo para generación de sistemas borrosos interpretables a partir de los datos, se basa en la minimización del error de inferencia. El usuario sólo debe introducir los datos de las variables de entrada y salida. El algoritmo determina los rangos de cada variable, distribuye las funciones de pertenencia en sus universos de entrada, ubica los consecuentes tipo singleton en el espacio de salida, determina las reglas y ajusta la ubicación de los consecuentes, empleando mínimos cuadrados

para minimizar el error de aproximación. El algoritmo se detiene cuando se ha alcanzado una métrica de error menor a la requerida por el usuario o cuando el número de conjuntos borrosos por variable de entrada es mayor a 9. La distribución de las funciones de pertenencia en cada universo de entrada se hace de manera uniforme para garantizar que la partición resultante sea suma 1; es decir, la suma de los grados de pertenencia de un dato en una variable de entrada será siempre igual a 1.

Dada una colección de datos experimentales de entrada y salida, $\{x_k^{(i)}, y^{(i)}\}$, con $i = 1 \dots N$; $k = 1, \dots, p$, donde $x_k^{(i)}$ es el vector de entrada p -dimensional $x_1^{(i)}, x_2^{(i)}, \dots, x_p^{(i)}$ y $y^{(i)}$ es el vector unidimensional de salida.

a. Organización del conjunto de p variables de entrada y una variable de salida, cada una con N datos.

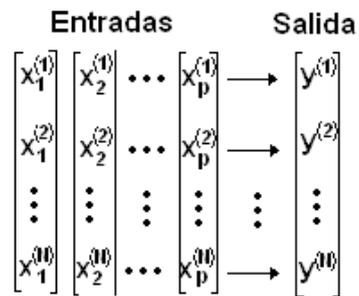


Figura 1. Organización del Conjunto de datos.

b. Determinación de los rangos de los universos de cada variable de acuerdo a los valores máximos y mínimos de los datos asociados

$$[x_k^-, x_k^+], [y^-, y^+]$$

c. Distribución de las funciones de pertenencia triangulares sobre cada universo. Se tiene como condición general que el vértice con valor de pertenencia uno (valor modal) cae en el centro de la región cubierta por la función de pertenencia, mientras que los otros dos vértices, con valor de pertenencia iguales a cero, caen en los centros de las dos regiones vecinas. Para poder aproximar eficientemente los extremos inferior y superior de una función representada por los datos, es necesario que en la partición triangular las

funciones de pertenencia que cubren el inicio y final del universo coincidan sus vértices con valor de pertenencia uno con sus vértices izquierdo y derecho respectivamente, como se aprecia en la figura 2. Se inicia con dos conjuntos triangulares.

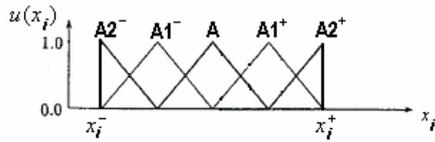


Figura 2. Partición triangular suma 1.

d. Cálculo de la posición de los valores modales de la(s) variable(s) de entrada

$$\text{if } u_{A_k^{(n)}}(x_k^{(i)}) = 1 \tag{3}$$

$$y s_k^{(n)} = y [i]$$

end

donde $y s_k^{(n)}$ corresponde a la proyección sobre el espacio de salida de la evaluación del dato $x^{(i)}$ de la k -ésima variable de entrada en el n -ésimo conjunto de la partición correspondiente. El valor de la salida correspondiente a dicha proyección está dado por el valor de la i -ésima posición del vector de salida y y se constituye en el singleton asociado a ese conjunto. De esta manera se aplica el criterio de error de inferencia nulo empleando consecuentes tipo singleton.

e. Determinación de las reglas. El número máximo de reglas está determinado por el número de conjuntos de cada variable de entrada multiplicado por el número de variables; es decir igual a $n \times k$, que es igual al número de singletons generados en el paso anterior. Cuando existen singletons ubicados en la misma posición, se combinan los antecedentes de acuerdo a (4) en una sola regla, lo que permite reducir el número de reglas. El operador \wedge representa un operador de combinación de la lógica borrosa de la clase promedio ponderado

$$((u_{A_1'}(x_1^{(i)}) \wedge (u_{A_2'}(x_2^{(i)}) \wedge \dots \wedge (u_{A_p'}(x_p^{(i)}))) = u_B(y^{(i)}) \tag{4}$$

f. Validación del modelo empleando el método de inferencia descrito por (1), donde \bar{y}^j es el valor del singleton correspondiente a la regla j .

g. Ajuste de los consecuentes reubicando los singletons de salida mediante el empleo del método de mínimos cuadrados. La ecuación (1) puede expresarse de la forma

$$f(x^{(i)}) = \sum_{j=1}^L w_j(x^{(i)}) \bar{y}^j \tag{5}$$

donde

$$w_j(x^{(i)}) = \frac{m_j(x^{(i)})}{\sum_{j=1}^L m_j(x^{(i)})} = w_j^i \tag{6}$$

La ecuación (5) puede expresarse en forma matricial como $Y = W\theta + E$, donde Y representa los valores de salida $f(x^{(i)})$ arrojados por el modelo borroso, W es la matriz de grados de pertenencia obtenida de (6), θ representa el vector de consecuentes y E es el error de aproximación que debe ser minimizado. Es decir:

$$\begin{bmatrix} y^1 \\ y^2 \\ \vdots \\ y^L \end{bmatrix}_Y = \begin{bmatrix} w_1^1 & w_2^1 & \dots & w_L^1 \\ w_1^2 & w_2^2 & \dots & w_L^2 \\ \vdots & \vdots & \ddots & \vdots \\ w_1^n & w_2^n & \dots & w_L^n \end{bmatrix}_W \begin{bmatrix} \bar{y}^1 \\ \bar{y}^2 \\ \vdots \\ \bar{y}^L \end{bmatrix}_\theta + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}_E \tag{7}$$

Empleando la norma del error cuadrático se tiene

$$E^2 = (Y - W\theta)^2 = (Y^2 - 2YW\theta + (W\theta)^2) \tag{8}$$

La solución a este problema de mínimos cuadrados está dada por

$$\frac{\partial E^2}{\partial \theta} = 0 = -2YW + 2W^T W \theta \quad (9)$$

de donde se obtiene

$$\theta = \frac{Y^T W}{W^T W} = (W^T W)^{-1} Y^T W \quad (10)$$

Esta solución es válida si $(W^T W)$ es no singular, lo que quiere decir que todas las reglas deben recibir suficiente excitación durante el entrenamiento. En la práctica esto no es siempre posible, por lo que es recomendable recurrir a la aplicación de mínimos cuadrados recursivos, buscando garantizar que la adaptación sólo afecte las reglas excitadas.

Terminar, si la medida del error cuadrático medio MSE es menor a una medida previamente establecida o si el número de conjuntos por variable de entrada es mayor a 9. De otra manera, incrementar en 1 el número n de conjuntos de la variable de entrada y volver al paso 3. Este aparte tiene por objeto proporcionar información suficiente para que un lector competente pueda repetir el estudio. En él se describe el diseño de la investigación y se explica la forma como se realizó la práctica, justificando la elección de procedimientos y técnicas. Así mismo, se describen los detalles del equipo instrumental, los materiales y las condiciones del experimento.

3. RESULTADOS

La base de datos del WBC contiene 699 patrones agrupados en dos clases: maligno y benigno. Cada patrón tiene asociado 9 variables de entrada (atributos):

- Espesor del grup.
- Uniformidad del tamaño de la célula
- Uniformidad de la forma de la célula
- Adhesión marginal

- Tamaño de la célula epitelial sencilla
- Núcleo
- Cromatina blanda
- Nucléolo normal
- Mitosis

Esos atributos miden el cambio de la apariencia externa e interna del cromosoma en nueve escalas diferentes mediante valores enteros en el rango [1,10]. Las dos clases existentes, benigno (no canceroso) y maligno (canceroso), se representan numéricamente mediante los valores 2 y 4, respectivamente. Debido a la existencia de valores erróneos o incompletos, en algunos datos se consideraron 683 patrones, 444 benignos y 239 malignos. De los 683 patrones se tomaron 342 para el entrenamiento y 341 para la validación.

Inicialmente las clases son tratadas como una función que sólo toma valores de salida de 2 y de 4 para cada combinación de los nueve atributos o variables de entrada. Por lo tanto, el algoritmo procede, en el proceso de entrenamiento, a generar un modelo borroso que aproxime a dicha función. Cada variable de entrada culmina, después del entrenamiento, con una partición suma 1 conformada por 3 conjuntos triangulares plenamente identificables: Bajo (B), Medio (M) y Alto (A). Estos conjuntos tienen sus valores modales ubicados en los valores mínimo, medio y máximo del rango de cada variable de entrada, como muestra la figura 3. Los consecuentes están ubicados, antes del ajuste por mínimos cuadrados, en los valores 2, que representa la clase “Benigna” y 4 que representa la clase “Maligna”.

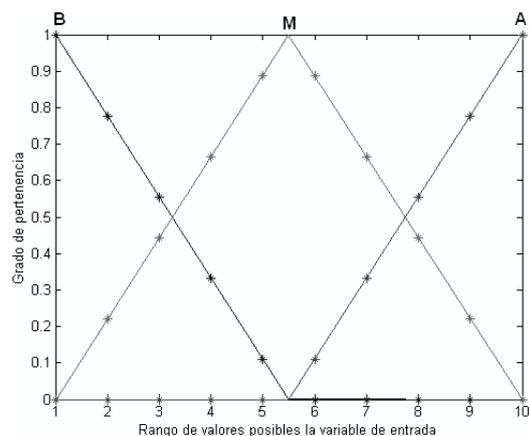


Figura 3. Funciones de pertenencia para cada atributo o variable.

El algoritmo de entrenamiento inicialmente asigna una regla por cada conjunto de cada variable de entrada. En este caso específico existen nueve variables de entrada y tres conjuntos por cada variable, lo cual genera 27 reglas. La figura 4 muestra las reglas generadas en una matriz donde se presentan las nueve variables de entrada, X1-X9, y las clases benigna (2) y maligna (4).

	X1	X2	X3	X4	X5	X6	X7	X8	X9
B	2	2	2	2	2	2	2	2	2
M	2	4	4	2	4	4	4	4	2
A	4	4	4	4	4	2	4	4	4

Figura 4. Base de reglas representada en forma matricial.

De las celdas de la matriz presentadas se pueden extraer las reglas de forma lingüística, así:

- Regla 1: Si X1(Espesor del grupo) es Baja (B), entonces no existe riesgo de cáncer (2)
- Regla 2: si X1 es Media (M), entonces no exist riego de cáncer (2)
- Regla 3: Si X1 es Alto (A), entonces existe riesgo de cáncer (4)
- Regla 4: Si X2(Uniformidad del tamaño de la célula) es Baja (B), entonces no existe riesgo de cáncer (2)
- Regla 4: Si X2 es Media (M), entonces existe riesgo de cáncer (4)

Las reglas pueden reducir a sólo 2, combinando los antecedentes de las reglas con consecuentes iguales mediante un operador de combinación, así:

- Regla1: Si X1 es Baja, X1 es Media, X2 es Baja X3 es Baja,..., X8 es Baja, X9 es Baja y X9 es Media, Entonces “Benigno”.
- Regla2: Si X1 es Alta, X2 es Media, X2 es Alta X3 es Baja,..., X8 es Media, X8 es Alta y X9 es Alta, Entonces “Maligno”.

El porcentaje de aciertos es del 96.78%. La tabla I muestra una comparación con los resultados obtenidos por otros investigadores

Tabla 1. Comparación con los resultados alcanzados por otros investigadores

Método	Desempeño %
Método de Setiono y Liu [13]	93.9
Método de Lee et al [14]	94.7
Método de Wang y Lee [15]	96.3
Método propuesto	96.78

4. DISCUSIÓN

Se presentó una nueva metodología para identificación de modelos borrosos interpretables a partir de datos experimentales de entrada y salida, utilizándola como un sistema clasificador.

El modelo borroso obtenido tiene alta precisión e interpretabilidad y no requiere del empleo de otras técnicas, como las redes neuronales, para el entrenamiento. Es decir, no es un sistema híbrido, es un sistema borroso ciento por ciento.

La comparación con otros métodos permite verificar el buen desempeño de la metodología propuesta. En la literatura relacionada no se incluye mucha información sobre las estructuras de los sistemas borrosos obtenidos pero se puede apreciar en gran parte de ella una permanente violación a las condiciones de interpretabilidad, así como un elevado número de reglas.

5. REFERENCIAS BIBLIOGRÁFICAS

[1] J. Espinosa, J. Vandewalle, *Constructing fuzzy models with linguistic integrity from numerical data-afreli algorithm*, IEEE Trans. Fuzzy Systems, vol. 8, pp. 591 – 600, 2000.

[2] R. P. Paiva, A. Dourado, *Interpretability and learning in neuro-fuzzy systems*, Fuzzy Sets and System, 147, pp. 17-38, 2004.

[3] T. Sudkamp, A. Knapp, and J. Knapp, *Model generation by domain refinement and rule*

- reduction*, IEEE Trans. on Systems, Man and Cybernetic, vol. 33, No. 1, pp.45-55, 2003.
- [4] S. Guillaume, B. Carnomordic, *Generating an interpretable family of fuzzy partitions from data*, IEEE Trans. Fuzzy Systems, vol. 12, No. 3, pp. 324–335, Jun, 2004.
- [5] A. Sala, *Validación y aproximación funcional en sistemas de control basados en lógica borrosa*, Tesis Doctoral, Universidad Politécnica de Valencia, 1998.
- [6] J. Contreras, R. Misa, J. Paz, *Building interpretable fuzzy systems: A new methodology*, in Proceedings of Electronics, Robotics and Automotive Mechanics Conference CERMA 2006, IEEE Computer Society, Pp. 172-178, Nov, 2006
- [7] J. Contreras, R. Misa, J. Paz, L. Murrillo, J. Sarabia, *Efficient fuzzy identification based on inference error*, Revista Colombiana de Tecnologías Avanzadas, Vol 1, No.9, pp. 62-67, 2007
- [8] J. Soto, J. Egea, A. Flores Sintas, *Clasificación difusa como una herramienta: una aplicación en el cáncer de mama*, IX Conferencia Española de Biometría, pp. 33-36, 2003.
- [9] F. Peter Pach and J. Abonyi, *Association rule and decision tree based methods for fuzzy rule base generation*, Proceedings of World Academy of Science, Engineering and Technology, Vol. 13, pp. 45-50, 2006
- [10] T. Z. Tan, G. S. Ng and C. Queck, *Complementary learning fuzzy neural network: an approach to imbalanced dataset*, Proceedings of International Joint Conference on Neural Networks, Orlando, Florida, 2007.
- [11] C. Peña-Reyes, *Coevolutionary fuzzy modeling*, Tesis Doctoral, École Polytechnique Fédérale de Lausanne, 2002.
- [12] W. Pedriycz, *Why triangular membership functions?*, IEEE Trans. Fuzzy Sets and System, vol. 64, pp.21-30, 1994
- [13] R. Setiono and H. Liu, *Neural network feature selector*, IEEE Trans. Neural Network, vol. 88, pp.654-662, 1997.
- [14] H. M. Lee, C. M. Chen, J. M. Chen, and Y. L. Jou, *An efficient fuzzy classifier with feature selection based on fuzzy entropy*, IEEE Trans. Syst. Man and Cybern, vol. 31, pp.426-432, 2001
- [15] J. S. Wang and C. S. G. Lee, *Self adaptive neuro-fuzzy inference system for classifications application*, IEEE Trans. Fuzzy Systems, vol. 10, pp.790-802, 2002.