NOVIEMBRE 1996 • VOLUMEN XVII • NÚMERO 1

ISSN 0188-9532

REVISTA MEXICANA DE

INGENIERÍA BIOMÉDICA

PUBLICACIÓN
DE LA SOCIEDAD MEXICANA
DE INGENIERÍA BIOMÉDICA

SUPLEMENTO

MEMORIAS
DEL XIX CONGRESO NACIONAL
DE INGENIERIA BIOMEDICA

CONFERENCIAS MAGISTRALES
RESÚMENES DE TRABAJOS LIBRES

sinc(i) Research Institute for Signals, Systems and Computational Intelligence (http://fich.unl.edu.ar/sinc)
J. Goddard, H. M. Martínez, A. E. Martínez, J. M. Cornejo, H. L. Rufiner & R. C. Acevedo; "Aprendizaje Maquinal en Medicina: Diabetes, un caso de Estudio"
Revista Mexicana de Ingeniería Biomédica. Vol. 17, No. 1, pp. 5-11, Nov. 1996.

REVISTA MEXICANA DE
INGENIERÍA BIOMÉDICA
UAM - Iztapalapa, Edificio AT-212, Av. Michoacán y Purísima, México, D.F., 09340. Tel./fax: (525) 723-6421

Fundador: Carlos García Moreira †

CONSEJO EDITORIAL

EDITOR EN JEFE:
Emilio Sacristán Rock, UAM - Iztapalapa
EDITORES ASOCIADOS
Jaime García Ruiz, Fac. de Ciencias UNAM
Lorenzo Leija Salas, CINVESTAV-IPN
Ruth Mayagoitia Hill, Univ. Iberoamericana
Joaquín Azpiroz Leehan, UAM - Iztapalapa
Genaro Rodríguez Rossini, Inst. Nac. de Cardiología
Ernesto Suaste Gómez, CINVESTAV-IPN
Fernando Prieto Hernández, Hospital General
Hortensia González Gómez, Fac. de Ciencias UNAM
COORDINADORA
Ana María Fernández Rodríguez

SOCIEDAD MEXICANA DE
INGENIERÍA BIOMÉDICA
Juan Badiano No. 1, Col. Sección XVI, Tlalpan 14080, México D.F., Apdo. Postal 22-587 C.P. 14001, D.F. Fax: (525) 573-0926

AFILIADA A:
INTERNATIONAL FEDERATION OF MEDICAL AND BIOLOGICAL ENGINEERING (IFMBE)
FEDERACIÓN DE SOCIEDADES CIENTÍFICAS DE MÉXICO A.C. (FESOCIME)
CONSEJO REGIONAL DE ING. BIOMÉDICA PARA AMERICA LATINA (CORAL)

MESA DIRECTIVA-SOMIB

PRESIDENTE:
Miguel Cadena Méndez
VICEPRESIDENTE:
Adriana Velázquez Berumen
SECRETARIO / TESORERO:
Óscar Infante Vázquez
VOCALES:
Ana María Rule Ruiz
Miguel Ángel Peña Castillo
Ricardo Horta Olivares
Claudia Cárdenas Alanís

Se autoriza la reproducción parcial o total de cualquier artículo a condición de hacer referencia bibliográfica a la Revista Mexicana de Ingeniería Biomédica y enviar una copia a la redacción de la misma.

Diseño: Duotono Diseño
Prol. Miramontes 348/16C
Col. Ejidos de Huipulco
tel./fax: 671 8970

Impresión: Estampa Impresores
Priv. Dr. Márquez No. 53
tel.: 530-5289
fax: 530-9179

sinc(i) Laboratory for Signals and Computational Intelligence (http://fich.unl.edu.ar/sinc)
J. Goddard, F. M. Martínez, A. E. Martínez, J. M. Cornejo, H. L. Rufiner & R. C. Acevedo; "Aprendizaje Maquinal en Medicina: Diabetes, un caso de Estudio"
Revista Mexicana de Ingeniería Biomédica. Vol. 17, No. 1, pp. 5-11, Nov, 1996.

Revista Mexicana de **Ingeniería Biomédica** • Vol XVII • No.1 • ISSN 0188-9532 • Noviembre 1996

sinc(i) Laboratory for Signals and Computational Intelligence (http://fich.unl.edu.ar/sinc)
J. Goddard, F. M. Martínez, A. E. Martínez, J. M. Cornejo, H. L. Rufiner & R. C. Acevedo; "Aprendizaje Maquinal en Medicina: Diabetes, un caso de Estudio"
Revista Mexicana de Ingeniería Biomédica. Vol. 17, No. 1, pp. 5-11, Nov, 1996.

CONTENIDO

APRENDIZAJE MAQUINAL EN MEDICINA: DIABETES, UN CASO DE ESTUDIO.

GODDARD CLOSE J.C.*
MARTÍNEZ LICONA A.E.*
MARTÍNEZ LICONA F.M.*
CORNEJO J.M.*
RUFINER DI PERSIA H.L.*#
ACEVEDO ANDRADE R.C. *#

* Departamento de Ingeniería Eléctrica, U.A.M.-Iztapalapa, Av. Purísima y Michoacán s/n, (09340), México D.F., México.
# Facultad de Ingeniería, U.N.E.R., CC. 57, Sucursal 3, Paraná, (3100), E.R., Argentina.

RESUMEN:
La intención del presente trabajo es comparar el uso de diferentes paradigmas de inteligencia artificial (IA) aplicados al diagnóstico médico. En particular, se trata un problema de clasificación utilizando una base de datos sobre diabetes mellitus. Se utilizaron los algoritmos de ID3 y CART para la inducción de reglas de clasificación y el algoritmo de retropropagación aplicado a perceptrones multicapa. También se aprovechó la relación existente entre árboles de decisión y perceptrones multicapas para definir la arquitectura de la red.

PALABRAS CLAVE:
Árboles de decisión, Redes Neuronales, Diabetes.

ABSTRACT:
The intention of the present work is to compare the use of different artificial intelligence paradigms applied to medical diagnosis. Specifically the classification problem here is about a diabetes mellitus database. ID3 and CART algorithms for classification rule induction and backpropagation algorithm were used to train multilayer perceptrons. Furthermore the relation between decision trees and multilayer perceptron was exploited to define the net's architecture.

KEYWORDS:
Decision Trees, Neural Nets, Diabetes.

sinc(*i*) Laboratory for Signals and Computational Intelligence (http://fich.unl.edu.ar/sinc)
J. Goddard, F. M. Martínez, A. E. Martínez, J. M. Cornejo, H. L. Rufiner & R. C. Acevedo; "Aprendizaje Maquinal en Medicina: Diabetes, un caso de Estudio"
Revista Mexicana de Ingeniería Biomédica. Vol. 17, No. 1, pp. 5-11, Nov, 1996.

I. INTRODUCCIÓN

El problema de reconocimiento de patrones tiene una larga historia en el ámbito médico. Aquí se define el problema de manera general siguiendo a Bezdek [1] como *"una búsqueda de estructura en los datos"*. En este contexto, estructura se entiende como la forma en la cual la información contenida en los datos observados puede ser organizada, de manera de identificar las relaciones entre las variables involucradas en el proceso.

Las técnicas que han dominado el reconocimiento de patrones durante mucho tiempo están basadas en la estadística [2]. Sin embargo, en los últimos años se han intentado aplicar otras que han surgido de la IA con diferentes grados de éxito [3].

El área de medicina presenta dificultades que no se encuentran en dominios mas precisos, como la posibilidad de datos inconsistentes, faltantes o inconclusos. En un trabajo previo [4], se utilizó una técnica híbrida que relaciona árboles de decisión y redes neuronales, con mejores resultados que los obtenidos con ambos métodos en forma aislada. Se aplicó la técnica a la clasificación de la planta IRIS, por ser un caso de estudio utilizado de manera universal en la literatura de aprendizaje maquinal.

En el presente artículo se considera una base de datos más compleja del ámbito médico relacionada con el diagnóstico de diabetes mellitus. Más aún, se considera otra técnica para inducir reglas de clasificación conocida como ID3 [5], a manera de presentar otro punto de comparación. Como se verá en la sección de descripción de la base de datos, ésta presenta dificultades no encontradas en el caso de IRIS.

El presente artículo se organiza de la siguiente manera. En las secciones II y III se presentan los paradigmas de inducción de reglas de clasificación y conexionista respectivamente; la relación entre ambos paradigmas se presenta en la sección IV. En la sección V se describe la base de datos utilizada, y los resultados y conclusiones se exponen en las siguientes secciones.

II. INDUCCIÓN DE REGLAS DE CLASIFICACIÓN

Dentro de la amplia variedad de técnicas utilizadas en aprendizaje maquinal la que quizás ha recibido mayor atención, y ha tenido más éxito comercial en su aplicación a sistemas expertos (SE), ha sido el paradigma de aprendizaje por reglas inducidas a partir de un conjunto de ejemplos de entrenamiento.

En este paradigma el algoritmo de aprendizaje busca una colección de reglas que clasifican *"mejor"* los ejemplos de entrenamiento. El término

sinc(i) Laboratory for Signals and Computational Intelligence (http://fich.unl.edu.ar/sinc)
J. Goddard, F. M. Martínez, A. E. Martínez, J. M. Cornejo, H. L. Rufiner & R. C. Acevedo; "Aprendizaje Maquinal en Medicina: Diabetes, un caso de Estudio"
Revista Mexicana de Ingeniería Biomédica. Vol. 17, No. 1, pp. 5-11, Nov, 1996.

mejor se puede definir en términos de exactitud y comprensión de las reglas generadas. Dichas reglas pueden ser representadas como un árbol de decisión (AD). Este puede pensarse como un diagrama de flujo en donde cada nodo representa una prueba y cada rama que sale del nodo representa un resultado posible de dicha prueba.

La generación de un AD puede resumirse de la siguiente manera [6]: si S es el conjunto de ejemplos de entrenamiento se pueden dar dos situaciones:

a) si todos los ejemplos de S pertenecen a la misma clase T, entonces el resultado es un único nodo etiquetado con la clase T.

b) de otra manera se selecciona el atributo con *más información* A, cuyos valores son a_1, .., a_n. De esta forma se particiona S en los subconjuntos S_1, .. S_n de acuerdo a los valores de A.

El resultado final de esto es un AD cuyo nodo raíz es A y las ramas que se desprenden de éste corresponden a los valores a_1, .., a_n del atributo A. En la figura 1 se observa un AD donde T_1, .., T_n son los nodos hojas que representan las clases del problema en cuestión.

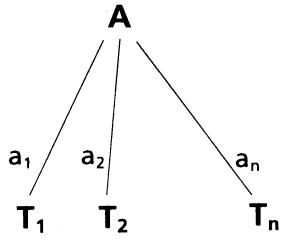

Figura 1 - Árbol de decisión.

Para entender la forma de trabajar de un AD se realizará una breve descripción de un AD binario; el razonamiento es análogo para AD n-arios. En AD binarios se comienza con la función de decisión del nodo raíz, los nodos internos particionan sucesivamente la región de decisión en dos espacios, donde la frontera que los divide está definida por la función del nodo correspondiente. De esta manera se logra una partición jerarquizada del espacio de decisiones. Dependiendo de las características de la función del nodo y del tamaño del árbol, la frontera final de decisión puede ser muy complicada. Una de las funciones más empleadas es la prueba con un cierto umbral para cada atributo, teniendo como resultado la partición del espacio de atributos por medio de hiperplanos paralelos u ortogonales a los ejes coordenados del espacio de atributos. Como característica importante del AD se puede mencionar que a partir de una colección de vectores de patrones etiquetados se configura de manera autónoma, independientemente de cualquier información *a priori* acerca de la distribución de los vectores de patrones en el espacio de decisiones. En lo que concierne con el aprendizaje maquinal, la construcción de AD es sinónimo de la adquisición de conocimiento estructurado en forma de conceptos y reglas para sistemas expertos.

Algoritmo ID3

Este algoritmo genera árboles de decisión n-arios, esto se debe a que particiona el conjunto de ejemplos de entrenamiento en función del *mejor* atributo. La función heurística que utiliza ID3 para determinar el mejor atributo es una medida de la entropía para cada atributo; esta medida está dada por:

$$Entropía = \sum_i w_i \, . \, E_i$$

donde w_i es el peso de la i-ésima rama definida como el número de ejemplos en la rama y dividido por el total de ejemplos en el nodo padre; y E_i es la entropía de la i-ésima rama dada por:

$$E_i = \sum_j p_j \, . \, log_2 \, (p_j)$$

donde p_j es la probabilidad de la j-ésima clase en esa rama estimada de los datos de entreamiento.

Para particionar el nodo bajo análisis se selecciona el atributo que tenga la entropía más baja [5].

Algoritmo C.A.R.T.

La sigla C.A.R.T. deriva de *Classification and Regression Trees;* este algoritmo genera árboles de decisión binarios. Esto se debe a que para particionar el conjunto de ejemplos en un nodo no se selecciona un atributo como en el caso de ID3, sino que se elige el mejor par atributo-valor de acuerdo a un determinado criterio. El criterio utilizado para particionar un nodo es el llamado criterio de Gini.

Para explicar como CART genera un AD se supondrá un nodo A bajo consideración, entonces se define la impureza del nodo A como sigue:

$$I(A) = \sum_{j=} p \, (j/A) \, . \, p(i/A)$$

donde p (i/A) es la probabilidad de la clase *i* en el nodo A. Esta probabilidad puede ser pensada como una frecuencia relativa.

Si sólo hay dos clases, la ecuación anterior se reduce a:

$$I(A) = 2 \, . \, p(1/A) \, . \, p(2/A)$$

donde $p(1/A)$ y $p(2/A)$ son las probabilidades de las clases 1 y 2 en el nodo A respectivamente.

Luego se define el decremento en impurezas como:

$$\Delta I(A) = I(A) - p_l \, . \, I(A_l) - p_r \, (A_r)$$

y se particiona el nodo de acuerdo al par atributo-valor con menor decremento de impurezas [5].

Aunque los AD son intuitivamente atractivos y han tenido aplicaciones exitosas [3], [7] existen algunos problemas que pueden obstaculizar su empleo en casos reales. Problemas tales como el diseño del árbol en sí mismo, la presencia de datos inconclusos, incompletos o ruidosos y el empleo simultáneo de todos los vectores de atributos.

Para solucionar algunos de estos problemas se debe pensar en algún criterio adecuado de terminación, es decir, hasta qué punto se debe hacer crecer el árbol de tal forma que no se obtenga un árbol demasiado complejo, difícil de entender y poco eficiente.

Una manera diferente de atacar el problema de reconocimiento de patrones es con el uso de los perceptrones multicapas (PMC). Éstos tienen la ventaja de que pueden aprender fronteras de decisión arbitrariamente complejas [8]. En la siguiente sección se presentan los aspectos relevantes de los PMC utilizados en el presente trabajo.

III. Perceptron Multicapas

Las redes neuronales artificiales (RNAs) intentan simular, al menos parcialmente, la estructura y funciones del cerebro y sistema nervioso de los seres vivos. Una RNA es un sistema de procesamiento de información o señales compuesto por un gran número de elementos simples de procesamiento, llamados *neuronas artificiales* o simplemente *nodos*. Dichos nodos están interconectados por uniones directas llamadas *conexiones* y cooperan para realizar procesamiento en paralelo con el objetivo de resolver una tarea computacional determinada.

Una de las características atractivas de las RNAs es su capacidad para autoadaptarse a condiciones ambientales especiales cambiando sus fuerzas de conexión llamadas *pesos*, o su estructura.

Muchos modelos de RNAs han sido desarrollados para una variedad de propósitos. Cada uno de ellos difieren en estructura, implementación y principio de operación; pero a su vez tienen características comunes.

El PMC consiste en un arreglo de nodos ubicados en capas, de forma tal que los nodos de una capa están conectados a todos los nodos de la capa anterior y de la siguiente mediante los pesos de conexión. La primera capa se denomina *capa de entrada* y la última se denomina *capa de salida;* las capas que quedan entre estas 2 se denominan *capas ocultas*. No existe un límite para fijar la cantidad de capas de un PMC, pero se ha demostrado que un PMC con una capa oculta y con un número suficiente de nodos es capaz de solucionar casi cualquier problema. Si se agrega una capa oculta más, un PMC soluciona cualquier tipo de problemas y en forma más eficiente que con una sola capa oculta.

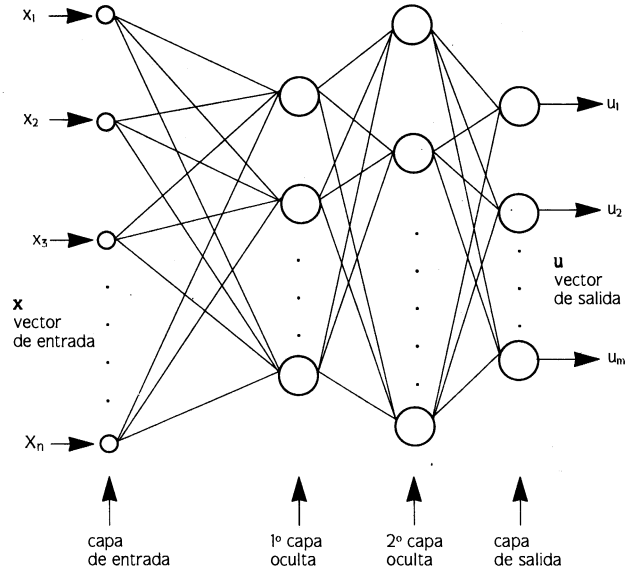

Figura 2 - PMC con 2 capas ocultas

Debido a restricciones analíticas en los algoritmos de entrenamiento, la función de activación de los nodos debe ser diferenciable, y con frecuencia se utilizan sigmoides. Dichas funciones de activación tiene la siguiente forma:

$$u(x) = \frac{1}{1 - e^{-x}}$$

donde x es la sumatoria de las entradas al nodo pesadas por los pesos de conexión y $u(x)$ la salida del nodo correspondiente.

Para entrenar el PMC generalmente se utiliza el algoritmo de retropropagación [8]. Durante el entrenamiento la respuesta de cada nodo en la capa de salida es comparada con la correspondiente respuesta deseada y los errores asociados son calculados para adaptar los pesos de la capa de salida.

Mientras que el PMC presenta las ventajas anteriormente mencionadas, también presenta el inconveniente que debe definirse su estructura en términos de nodos y capas, y en general puede demorar demasiado para llegar a un mínimo aceptable. Por otra parte, por ser entrenado por el método de gradiente descendente es posible que converja a un mínimo local.

En la siguiente sección se expondrá una relación que se puede establecer entre un AD y un PMC que nos da un método para definir la estructura de este último además de tener inicialmente un comportamiento equivalente.

IV. Mapeo de un Árbol de Decisión en un Perceptron Multicapas

Esta tarea implica transformar las reglas de clasificación, representadas por el árbol de decisión, en una red neuronal. Para esto se utilizó un método

sinc(i) Laboratory for Signals and Computational Intelligence (http://fich.unl.edu.ar/sinc)
J. Goddard, F. M. Martínez, A. E. Martínez, J. M. Cornejo, H. L. Rufiner & R. C. Acevedo; "Aprendizaje Maquinal en Medicina: Diabetes, un caso de Estudio"
Revista Mexicana de Ingeniería Biomédica, Vol. 17, No. 1, pp. 5-11, Nov. 1996.

sinc(i) Laboratory for Signals and Computational Intelligence (http://fich.unl.edu.ar/sinc)
J. Goddard, F. M. Martínez, A. E. Martínez, J. M. Cornejo, H. L. Rufiner & R. C. Acevedo; "Aprendizaje Maquinal en Medicina: Diabetes, un caso de Estudio"
Revista Mexicana de Ingeniería Biomédica. Vol. 17, No. 1, pp. 5-11, Nov, 1996.

de mapeo para convertir árboles de decisión binarios en perceptrones multicapas [4], [9], [10].

Para clasificar un patrón desconocido con un AD todas las condiciones a lo largo del camino desde el nodo raíz a una hoja determinada deben satisfacerse. De esta manera cada camino realiza una operación AND sobre un conjunto de subespacios definidos por los nodos intermedios en el camino. Si uno o más nodos resultan en la misma decisión entonces la operación implicada es una OR. De lo anterior se desprenden las reglas que producen la transformación del AD en un PCM:

a) El número de neuronas en la primera capa oculta de la red equivale al número de nodos internos del árbol de decisión (incluida la raíz).

b) Todos los nodos terminales u hojas tienen una correspondiente neurona en la segunda capa oculta donde se implementa la operación AND o de intersección.

c) El número de neuronas en la capa de salida equivale al número de las distintas clases o acciones. Esta capa implementa la operación OR o de unión.

V. Descripción de la Base de Datos

La base de datos utilizada en el presente trabajo fue creada en el *National Institute of Diabetes and Digestive and Kidney Diseases* (EE.UU.) y fue obtenida de [11]. Esta base de datos contiene información sobre mujeres mayores de 21 años descendientes de los indios Pima de Arizona (EE.UU.). Hay 768 casos con información en 8 atributos de entrada que se consideran pertinentes para la clasificación. El diagnóstico consiste en una variable binaria que indica si el paciente muestra signos de diabetes mellitus. De estos 768 casos, se eliminaron cerca de la mitad por tener datos faltantes en los atributos presión diastólica, grosor de la piel y concentración de insulina en la sangre. Este hecho no se mencionaba en la documentación que acompañaba a la base de datos.

Para la creación de los AD y el entrenamiento del PMC se debió separar los datos en dos subconjuntos, uno para el entrenamiento propiamente dicho y el otro para prueba. Se tomaron al azar el 90 % de estos datos para entrenamiento y el 10 % restante para prueba o verificación. Sin embargo por algunos problemas en la convergencia de las redes se terminó utilizando un subconjunto del anterior conjunto de entrenamiento con igual cantidad de sujetos sanos y enfermos (eliminando algunos casos sanos). En las tablas 1 y 2 se muestran las estadísticas de estos ejemplos de entrenamiento.

Dado las dificultades que aparecieron, sobre todo en la convergencia de las redes, se realizaron diversos análisis sobre los datos para tratar de explicarlas. En el caso de la aplicación de técnicas de agrupamiento se observó que existían varios grupos de ejemplos que eran muy similares y sin embargo pertenecían a clases opuestas. Esto indica la dificultad implicada en su clasificación. En el último conjunto de entrenamiento se notó una tendencia más marcada de la concentración de glucosa a influir en la clasificación, manifestándose esto, entre otras formas, como un aumento del coeficiente de correlación entre este atributo y el diagnóstico.

atributo	mín.	máx.	v. medio	desv. stand.	descrip.
Num_Emb	0	17	3.38	3.32	Embarazos
Gluc_Plas	68	198	127.62	32.69	Glucosa en plasma
Pres_Diast	24	110	71.25	13.11	Presión diastólica
Piel_Triceps	7	63	30.28	10.55	Grosor de la piel
Insulina	14	846	166.47	121.49	Insulina en sangre
Masa_Corp	18.2	67.1	33.75	7.23	Masa corporal
Pedigree	0.085	2.42	0.56	0.37	Pedigree
Edad	21	81	31.99	10.59	Edad

Tabla 1 - Estadísticas de los ejemplos de entrenamiento.

clases	entrenamiento	prueba	total
Enfermos	116 (50 %)	14 (35.9 %)	130 (47.97 %)
Sanos	116 (50 %)	25 (64.1 %)	141 (52.03 %)

Tabla 2 - Distribución de los ejemplos en los archivos de entrenamiento y prueba.

VI. Resultados

Con el 90 % de los ejemplos sin datos faltantes se procedió a la construcción de los AD, sin embargo el PMC parecía tener problemas en la convergencia sobre todo para estructuras pequeñas (3 capas y menos de 10 unidades por capa). A partir del árbol generado por CART (ver figura 4) y podado con Chi^2 0.5 % se construyó un PMC mediante el método de mapeo descrito anteriormente. El desempeño inicial de este fue igual al del árbol que le dio origen (según se desprende del procedimiento de mapeo), luego se continuó el entrenamiento y el PMC comenzó a olvidar lo aprendido a través del mapeo. Se debe notar que a pesar de que el desempeño inicial fue idéntico (igual porcentaje de reconocimiento para cada clase y para cada archivo de datos) esto no implica un Error Cuadrático Medio (ECM) bajo ya que el PMC realiza la clasificación asignando la clase de la salida más alta. En particular en este caso el ECM normalizado fue de 0.48, siendo este valor relativamente alto.

Se notó que uno de los problemas con el PMC era que existían muchos más ejemplos (casi 70 %) de individuos sanos. Por lo tanto el PMC lograba porcentajes de reconocimiento relativamente altos pero en realidad clasificaba casi siempre los patrones mostrados como sanos y casi nunca acertaba con los enfermos. Al parecer la influencia de ejemplos de una clase crea mínimos locales muy fuertes en el ECM sobre el archivo de entrenamiento, pero de valor (error) relativamente alto y por lo tanto resultan en malas soluciones.

Figura 3 - AD generado por ID3

Figura 4 -AD generado por CART

sinc(i) Laboratory for Signals and Computational Intelligence (http://fich.unl.edu.ar/sinc)
J. Goddard, F. M. Martínez, A. E. Martínez, J. M. Cornejo, H. L. Rufiner & R. C. Acevedo; "Aprendizaje Maquinal en Medicina: Diabetes, un caso de estudio"
Revista Mexicana de Ingeniería Biomédica. Vol. 17, No. 1, pp. 5-11, Nov, 1996.

sinc(i) Laboratory for Signals and Computational Intelligence (http://fich.unl.edu.ar/sinc)
J. Goddard, F. M. Martínez, A. E. Martínez, J. M. Cornejo, H. L. Rufiner & R. C. Acevedo; "Aprendizaje Maquinal en Medicina: Diabetes, un caso de Estudio"
Revista Mexicana de Ingeniería Biomédica. Vol. 17, No. 1, pp. 5-11, Nov, 1996.

construc. a partir de	unid. 1° capa oculta	unid. 2° capa oculta	convergencia
Árbol podado c/chi² 1%	10	9	no
Árbol podado c/chi² 2%	16	15	no
Árbol podado c/chi² 5%	25	24	no
Árbol sin podar	33	32	si (498 épocas)
Árbol sin podar	50	50	si (1280 épocas)

Tabla 3- Estructuras de los PCM utilizados.

Para remediar este problema se eliminaron algunos ejemplos sanos de manera de igualar la cantidad de sujetos sanos y enfermos en el archivo de entrenamiento (ver tabla 3). Con este archivo se volvieron a entrenar los PMC convergiendo éstos sin dificultad y mostrando una buena generalización. Los árboles construidos con estos datos también fueron más compactos y lograron generalizar mejor (inclusive sobre todo el conjunto de ejemplos original).

Se entrenaron diferentes arquitecturas de PMC tomadas del árbol que generó CART luego de podarlo con diferentes coeficientes. Los resultados se muestran en la tabla 4. En el último caso se entrenó un PMC tratando de que fuera bastante más grande que el obtenido a partir del árbol sin podar. Se puede observar que el PMC generado a partir del árbol sin podar converge y lo hace en una cantidad de épocas menor que la estructura más grande. Esto indicaría que esta estructura es la óptima entre las utilizadas. Se experimentó también generando un PMC a partir del árbol de ID3 (ver figura 3) por ser en este caso binario debido a que todos los atributos fueron del tipo numérico. Este experimento arrojó resultados muy similares a los de la estructura derivada de CART.

Los resultados de las distintas técnicas, en términos de desempeño y tamaño, se pueden observar en la tabla 5. Los árboles generados por CART e ID3 fueron bastante similares aunque CART generó árboles más compactos. De los algoritmos inductivos CART obtuvo mejores porcentajes de reconocimiento, aunque el PMC fue el mejor en este sentido. El tiempo de entrenamiento fue muy superior en el caso del PMC, siguiendo ID3 y luego CART como el más rápido (todos corridos en la misma plataforma). Se debe notar que el desempeño de estas técnicas superó al 76 % reportado con anterioridad para el algoritmo ADAP y los mismos datos [12]. Este algoritmo es una rutina de aprendizaje adaptativa que genera y ejecuta análogos digitales de dispositivos tipo perceptron.

VII. Discusión y Conclusiones

Como principal conclusión se puede mencionar el hecho de que los porcentajes de diagnóstico acertado para todas las técnicas son altos especialmente para los sujetos diabéticos. Esto ilustra que la aplicación de estas técnicas en la práctica médica merecen ser consideradas. Sin embargo para elegir una técnica en particular se debe pesar no solo su exactitud sino cuestiones como si puede justificar sus respuestas, el tiempo de entrenamiento, la complejidad computacional, etc. El peso de estos factores depende del destino final del sistema.

En cuanto a la justificación de un diagnóstico, las técnicas inductivas llevan una ventaja sobre el PMC por representar el conocimiento explícitamente (en base a las reglas). Se debe mencionar que hay investigación reciente sobre la interpretación del conocimiento contenido en un PMC [5].

El trabajo interdisciplinario en este tipo de campo puede enriquecer mucho las conclusiones y mejorar el desempeño de los métodos planteados. En este sentido se está comenzando a explorar la interacción con el médico en preguntas como ¿qué pasa si se parte de distintos atributos en la raíz?, ¿cómo se interpretan las reglas?, ¿qué se puede concluir sobre los ejemplos mal clasificados?, ¿cuáles son los ejemplos más representativos?, etc. Es de notar que ambos métodos inductivos eligieron la concentración de glucosa en plasma como el atributo más importante, lo que coincide con el criterio de la Organización Mundial de la Salud. Este tipo de resultados sugiere también el empleo

	algoritmo ID3		algoritmo CART		perceptron		perceptron mapeado	
Clase	entrenam.	test	entrenam.	test	entrenam.	test	entrenam.	test
Enfermo	94.8	84.0	99.1	92.9	100.0	85.0	87.2	92.9
Sano	98.3	78.5	92.2	76.0	100.0	78.0	72.4	50.0
Total	89.2	81.2	95.7	82.1	100.0	81.5	79.8	71.4

Tabla 4- Porcentaje de ejemplos clasificados correctamente

algoritmo ID3		algoritmo CART		perceptron		perceptron mapeado	
nodos int.	nodos hojas	nodos int.	nodos hojas	1° capa oculta	2° capa oculta	1°capa oculta	2° capa oculta
38	39	32	33	32	33	8	7

Tabla 5- Estructuras de los AD y de los PMC

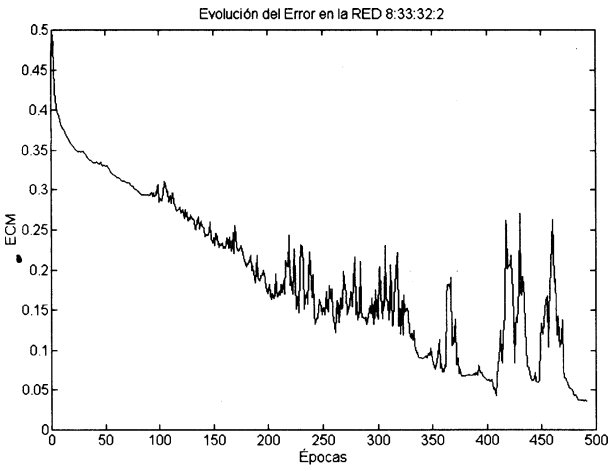

Figura 5 -Evolución del ECM para el PCM.

de técnicas de extracción de características como las basadas en algoritmos genéticos [13].

Un aspecto importante en el diagnóstico clínico es la necesidad de una precisión mayor en una de las clases. Por ejemplo, no es lo mismo clasificar un paciente como sano cuando en realidad está enfermo que a la inversa. En el primer caso el resultado podría ser fatal, mientras que en el segundo, si bien no deseable, sería menos grave. Si esto no se tiene en cuenta el sistema puede tener un buen desempeño global pero no tener aplicación práctica. Como una posible solución a explorar se podría ponderar de manera distinta los errores cometidos en cada clase.

Se debe destacar la importancia que tiene la elección de los ejemplos para entrenamiento, distintas elecciones al azar pueden dar resultados muy diferentes. Para intentar resolver este problema se podrían utilizar algoritmos genéticos para escoger los conjuntos de acuerdo con algún criterio de aptitud. Este enfoque será objeto de análisis en un trabajo futuro.

Por último, se determinó la importancia del análisis realizado por los AD en la determinación de la estructura del PMC, concluyendo que dan una buena aproximación a la arquitectura óptima. En [4] se mostró que un PMC mapeado a de un AD podado producía mejor generalización que el árbol sin podar y que PMCs más grandes. Sin embargo esto fue cierto para datos relativamente sencillos de clasificar. En este caso los PMC mapeados de árboles podados no convergieron. Esto resalta la necesidad de un análisis más cuidadoso de las condiciones bajo las cuales es válido el podado en este contexto.

Referencias

1. Bezdek, "Pattern Recognition with Fuzzy Objetive Function Algorithms", New York, 1981.
2. Duda, R.O., & Hart, P.E. 1973 "Pattern Classification and Scene Analysis".(Q327.D83) John Wiley & Sons. ISBN 0-471-22361-1. pag. 218.
3. Quinlan, J.R."C4.4", Morgan Kaufmann 1993.
4. Goddard J., Rufiner L., Acevedo R., Martínez F., Martínez A.y Cornejo J., "Redes Neuronales y Árboles de Decisión: un Enfoque Híbrido", Memorias del Simposio Internacional de Computación, IPN, noviembre 1995.
5. Sestito, Dillon, "Automated Knowledge Acquisition", Prentice Hall 1994.
6. Bratko, "Prolog", Addison-Wesley, 1990
7. Breiman, L. Friedman J. H.,. Olshen R. A, Stone C. J., "Classification and Regression Trees", Wadsworth. Int. 1984.
8. Lippmann, R. P. "An introduction to computing with neural nets" IEEE ASSP Magazing, Apr. 1987.
9. Sethi, I.K."Decision tree performance enhancement using an artificial neural network implementation", Artificial Neural Networks and Statistical Pattern Recognition Old and New Connections, Elsevier Science Publishers B.V., pp 71-88, 1991.
10. Brent, R.P."Fast Training Algorithms for Multilayer Neural Nets", IEEE Trans. Neural Networks, vol. 2, nº 3, May 1991.
11. UCI Repository of Machine Learning Databases and Domain Theories (ics.uci.edu: pub/machine-learning -databases).
12. Smith J. W., Everhart J.E., Dickson W. C., Knowler W. C., Johannes R. S., "Using the ADAP learning algorithm to forecast the onset of diabetes mellitus", Proceedings of the Symposium on Computers Applications and Medical Care, pp. 261-265, IEEE Computer Society Press, 1988.
13. Kermani, B.G. "Feature Extraction by Genetic Algorithms in Breast Cancer", 1995 IEEE Eng. in Medicine & Biology 17th Annual Conference & 21st Canadian Medical and Biological Engineering Conference, Canadá, 1995.

sinc(i) Laboratory for Signals and Computational Intelligence (http://fich.unl.edu.ar/sinc)
J. Goddard, F. M. Martínez, A. E. Martínez, J. M. Cornejo, H. L. Rufiner & R. C. Acevedo; "Aprendizaje Maquinal en Medicina: Diabetes, un caso de Estudio"
Revista Mexicana de Ingeniería Biomédica. Vol. 17, No. 1, pp. 5-11, Nov. 1996.