Discriminación y predicción de propiedades de fármacos mediante redes neuronales

Juan Lucas Domínguez Rubio[1] María José Castro Bleda[1] Wladimiro Díaz Villanueva[2]

[1]Dep. de Sistemas Informáticos y Computación — Universidad Politécnica de Valencia
Camino de Vera s/n, 46071 Valencia, España

[2]Dep. de Informática — Universitat de València
Dr. Moliner 50, 46100 Burjassot (Valencia), España

jldoru@hotmail.com mcastro@dsic.upv.es wladimiro.diaz@uv.es

Resumen

En este trabajo se estudia el problema de la *discriminación* y *predicción* de las propiedades farmacológicas de ciertos compuestos moleculares a partir de su topología ("*structure-activity methods*") utilizando perceptrones multicapa. Continuando el trabajo de Cristina Adobes Martín [Ado00] nos centraremos en el diseño de las redes neuronales mediante un barrido por multitud de configuraciones distintas. Entre ellas elegiremos aquellas que puedan formar parte de un *comité de redes* para intentar incrementar la tasa de aciertos en la predicción. Tras realizar pruebas con distintas estrategias internas en los comités, comprobaremos que las predicciones de las redes presentan una fuerte correlación, por lo que las mejoras obtenidas frente a redes individuales son poco significativas.

Palabras clave: Diseño de fármacos, química combinatoria, representación molecular, aprendizaje supervisado, redes neuronales, perceptrón multicapa, comité de redes.

1 Introducción

Como paso previo a la síntesis de un compuesto molecular en el laboratorio es interesante disponer de una herramienta capaz de predecir cuáles serán sus características (físicas, químicas, farmacológicas etc.), a fin de restringir el campo de búsqueda, privilegiando las moléculas más prometedoras. De esta manera se ahorra tiempo y costes en la investigación.

Los problemas que trataremos tienen interés en el campo de la farmacología. Se estudian dos problemas de clasificación (presencia o ausencia de efecto *analgésico* y *antidiabético* en una serie de compuestos moleculares) y dos problemas de predicción cuantitativa (predecir la actividad bactericida –el nivel de concentración mínima inhibitoria, *CIM*– en ciertas quinolonas y el grado de *solubilidad* de una serie de moléculas). En la tabla 1 se listan los problemas que se han abordado y el número de muestras disponibles.

Las moléculas se representan mediante 62 datos topológicos (presencia de ciertos átomos en las moléculas y su posición). Por ejemplo, algunos de estos índices están relacionados con el número total de átomos de un determinado elemento (carbono, nitrógeno, oxígeno,...), el número total de enlaces de un determinado tipo (simples, dobles o triples), distancia de los enlaces, etc.

Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial. No. 18 (2003), pp. 7–16.
ISSN: 1137-3601.

Tabla 1: *Nº de muestras en cada problema.*

C	Clasificación (2 clases: activa, inactiva)
P	Predicción (cuantitativo, sin clases)
+	Muestras positivas (activas)
−	Muestras negativas (inactivas)

		Nº de muestras		
Problema	*Tipo*	+	−	*Total*
'Analgésicos'	C	172	813	985
'Antidiabéticos'	C	180	163	343
'CIM'	P			111
'Solubilidad'	P			92

1.1 Metodología: redes neuronales artificiales

Las redes neuronales artificiales [Bis95] se vienen aplicando en el campo de la química desde hace pocos años [V$^+$96]. Predecir las propiedades farmacológicas de ciertas moléculas es uno de los casos donde se pueden utilizar estos métodos; es decir, deseamos aproximar, mediante una o varias redes neuronales, una función F tal que:

$$\text{Propiedad} = F(\text{Descriptores topológicos})$$

Un punto fundamental en este método es la selección apropiada de la representación de las moléculas. En este trabajo se utiliza un conjunto de índices topológicos que deben ser capaces de capturar de algún modo que es lo que produce las propiedades que nos interesan de las moléculas.

1.2 Antecedentes

Bajo la dirección de Wladimiro Díaz Villanueva, Cristina Adobes Martín presentó el proyecto "Diseño e implementación de herramientas para la predicción de propiedades moleculares" [Ado00, DCA$^+$01] en la Universitat de València, cuyo objetivo era "desarrollar un método para discriminar mediante redes neuronales artificiales las características farmacológicas de ciertas moléculas a partir de las características topológicas de éstas".

Se desarrolló una aplicación que calculaba los ya mencionados índices topológicos y se utilizó después para obtener los índices de una serie de moléculas interesantes en los problemas de clasificación 'Analgésicos' y 'Antidiabéticos'.

Tabla 2: *Resumen de los resultados previos.*

	Test: tasa de aciertos (%)	
Topología	'Analgésicos'	'Antidiabéticos'
`62-2-1`	82.6	91.6
`62-8-1`	87.8	92.6
`62-16-1`	88.6	87.9
`62-32-1`	85.8	88.8

Para representar los datos de entrada de manera adecuada, se realizó una transformación lineal de los índices topológicos de acuerdo con la fórmula:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}},$$

donde x es el valor original, x' es el valor transformado y $x_{\max}$ y $x_{\min}$ son los valores máximo y mínimo de ese índice para todas las muestras del problema. Lógicamente, de esta manera se obtiene una distribución entre 0 y 1. En cuanto al valor de salida, se codificó de la siguiente manera: molécula activa, 1; molécula inactiva, -1. Para las unidades de salida, la función de activación elegida en los experimentos fue la función tangente hiperbólica, con el fin de ajustar las salidas del perceptrón multicapa a la codificación de las clases. El algoritmo de aprendizaje utilizado fue el *Standard Backpropagation* en varias redes de una capa oculta completamente conectadas hacia adelante sin cortocircuitos. El *software* de redes neuronales utilizado fue el SNNS, de la Universidad de Stuttgart [Z$^+$98].

Tras dividir los conjuntos de muestras en dos bloques (*entrenamiento* y *test*) se realizaron varios entrenamientos para cada problema (con distinto número de nodos en la capa oculta), pero sin un criterio de parada para el entrenamiento, decidiendo a priori el número de ciclos a ejecutar. Para evaluar los resultados, se utilizó el siguiente criterio: si la diferencia entre el valor deseado y el valor obtenido es menor o igual a 0.5, se trata de un acierto; si es mayor que 0.5 y menor o igual a 1, el resultado es indeterminado y si es mayor que 1, se considera un error.

Los resultados finales del proyecto fueron satisfactorios y se muestran en la tabla 2. El objetivo del presente trabajo es profundizar en esta última parte, referente al diseño y evaluación de las redes neuronales, abordando los dos problemas de clasificación ya mencionados ('Analgésicos' y 'Antidiabéticos') y dos nuevos problemas de predicción cuantitativos ('CIM' y

Tabla 3: *Nº de índices topológicos utilizados.*

Problema	*Nº de índices topológicos*
'Analgésicos'	62
'Antidiabéticos'	62
'CIM'	52
'Solubilidad'	51

'Solubilidad'). Utilizaremos de nuevo el paquete SNNS (versión 4.2, de 1998) sobre el sistema operativo `Red Hat Linux 6.2`.

2 Diseño del clasificador-predictor

Utilizaremos un *clasificador estático* para abordar nuestro problema, el perceptrón multicapa (PM), lo que implica disponer de algún criterio para fijar la topología de la red, ya que ésta no variará durante el algoritmo de aprendizaje. También decidiremos el algoritmo de aprendizaje que se debe utilizar y el valor de sus parámetros relevantes. Para ello llevaremos a cabo una búsqueda predeterminada a lo largo de ciertos rangos para los parámetros de los algoritmos y diferentes topologías. El comportamiento en validación será un indicativo de la bondad de la configuración en cuestión.

2.1 Conjuntos de muestras

Los datos de partida, como hemos comentado, son moléculas descritas por 62 índices topológicos, etiquetadas según el tipo de problema: clasificación (activa, 1; no activa, -1) y cuantitativa en los de predicción (un número real). En los problemas de predicción hemos observado que algunos índices no variaban en las muestras, por lo que los hemos eliminado (10 índices en las muestras del problema 'CIM' y 11 índices de las muestras correspondientes a 'Solubilidad'). Estos índices topológicos serán ignorados a partir de este momento, con lo cual utilizaremos para cada problema los índices topológicos que se indican en la tabla 3.

2.1.1 Normalización

La normalización de los datos (tanto de entrada como de salida) se hace por escalado, con un factor igual a $1/x_{\max}$, siendo $x_{\max}$ el mayor valor para ese dato en el conjunto de todas las muestras.

Tabla 4: *Rangos de los datos de salida (problemas de predicción cuantitativos).*

	Rango	
Problema	*Inicial*	*Normalizado*
'CIM'	$[0.006, 3.1]$	$[0.00194, 1]$
'Solubilidad'	$[-0.946, 9]$	$[-0.105, 1]$

Tabla 5: *Barrido: tamaño de los conjuntos de entrenamiento y validación.*

	Nº de muestras	
Problema	*Entrenamiento*	*Validación*
'Analgésicos'	689	296
'Antidiabéticos'	240	103
'CIM'	78	33
'Solubilidad'	64	28

Tras la normalización, los valores quedarán, lógicamente, en un rango $[a..1]$, siendo $a = x_{\min}/x_{\max}$ ($x_{\min}$ y $x_{\max}$ son los valores mínimo y máximo para cada índice). Esta normalización no coincide exactamente con la utilizada en [Ado00], ya que para algunos índices no se da el valor 0 en ninguna muestra. Con esta normalización conservamos la proporción entre los valores de los índices: si entre dos valores había una relación $1 : z$, ésta se mantendrá después, ya que es un escalado. La tabla 4 muestra los rangos en los que se encuentran los datos de salida de los problemas de predicción cuantitativos una vez normalizados.

2.1.2 Partición inicial

Para cada uno de los problemas, dividiremos las muestras de manera homogénea en dos bloques: entrenamiento (70 % de las muestras) y validación (30 % de las muestras). Los tamaños de los conjuntos resultantes se muestran en la tabla 5.

2.2 Entrenamiento inicial: Rangos de búsqueda

Las características de las redes fijadas a priori son: número de unidades de entrada determina-

do por el problema (51, 52 ó 62), una única unidad en la capa de salida y tangente hiperbólica como función de activación para todos los nodos de la red. Las características que intentaremos optimizar serán:

Topología de la red: 1 ó 2 capas ocultas con 2, 4, 8, 16, 32 ó 64. Si se trata de un PM con dos capas ocultas, ambas poseerán el mismo número de nodos.

Algoritmo de aprendizaje: *Standard backpropagation*, *Backpropagation-momentum* o *Quick propagation* [DHS01, Z$^+$98].

Parámetros fundamentales asociados al algoritmo de aprendizaje:

- *Standard backpropagation.* Se realizarán pruebas con los siguientes coeficientes de aprendizaje (*learning rate*): 0.1, 0.2, 0.4, 0.7, 0.9, 1.5 y 2.0. La diferencia permitida entre la salida obtenida y la salida deseada se fija a 0 (parámetro $d_{\max}$).
- *Backpropagation-momentum.* Se realizarán pruebas con los siguientes coeficientes de aprendizaje: 0.1, 0.2, 0.4, 0.7 y 0.9, y los siguientes coeficientes momento (*momentum rate*): 0.1, 0.2, 0.4, 0.7 y 0.9. Se probarán todas las combinaciones entre estos valores. Los parámetros c (constante de eliminación de zonas planas) y $d_{\max}$ se fijan a 0.
- *Quick propagation.* Se realizarán pruebas con los siguientes coeficientes de aprendizaje: 0.1, 0.2 y 0.3 y los siguientes coeficientes de crecimiento máximo (*maximum growth parameter*): 1.75, 2.0 y 2.25. Los parámetros ν (*weight decay*) y $d_{\max}$ se fijan a 0.

En total, 492 experimentos generados automáticamente para cada tipo de problema. Para averiguar cuáles son las características óptimas de la red neuronal para cada problema, llevaremos a cabo un entrenamiento para cada configuración diferente de la red. El comportamiento de la red en validación nos indicará la bondad de la configuración probada.

Una vez probadas las configuraciones, disponemos de un listado que podemos ordenar en función del comportamiento (porcentaje de aciertos o ECM, error cuadrático medio). La mejor configuración (con nuestro criterio de validación) para cada uno de los problemas, se muestra en la tabla 6. Consideraremos más adelante las diez mejores configuraciones para cada problema.

Tabla 6: *Barrido: mejor configuración para cada problema.* SB: *Standard Backpropagation;* BM: *Backpropagation Momentum;* LR: *Learning Rate;* MR: *Momentum Rate. En la topología se muestran sólo las capas ocultas. La columna "Validación" muestra el porcentaje de aciertos en los dos primeros problemas y el ECM en los dos últimos.*

		Parámetros			
Problema	*Top.*	*Alg.*	LR	MR	*Validación*
'Analgésicos'	`16`	SB	0.1		89.27 %
'Antidiabéticos'	`4-4`	BM	0.2	0.1	94.30 %
'CIM'	`64`	SB	0.1		0.0282
'Solubilidad'	`32`	SB	0.1		0.0104

3 Comités de redes neuronales

Es habitual, al emplear redes neuronales, entrenar diferentes redes candidatas y escoger la que presenta un mejor comportamiento (en el conjunto de muestras de validación) como el predictor o clasificador definitivo, descartando el resto de redes entrenadas.

Sin embargo, es posible combinar diferentes redes en un *comité* capaz de mejorar las prestaciones. Se demuestra [Bis95] que esto ocurre cuando las redes cometen errores que no presentan correlación entre sí. Intuitivamente, podemos decir que en ese caso las desviaciones o errores se anulan mutuamente y el resultado de combinar las predicciones de las redes (por ejemplo, calculando simplemente la media aritmética, aunque en [PC93, Has94] se muestran métodos más complejos basados en matrices de correlación) puede proporcionar un error inferior al de cualquiera de las redes tomadas individualmente. En [CC00] pueden encontrarse resultados satisfactorios en este sentido, en el campo del reconocimiento del habla.

En nuestro caso, seleccionaremos las redes susceptibles de formar parte de un comité y más tarde las combinaremos de diferentes maneras para intentar mejorar las predicciones.

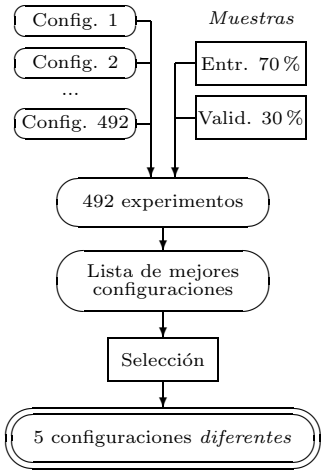

Figura 1: *Esquema de la búsqueda de las mejores configuraciones de las redes para cada problema.*

3.1 Los miembros del comité

De la lista de las diez mejores configuraciones para cada problema según el criterio de validación, debemos elegir cuáles pueden entrar a formar parte de un comité. Como hemos señalado, para que un comité proporcione buenos resultados, es necesario que no exista correlación entre los resultados de sus miembros, así que no parece apropiado incluir en el comité redes que se distingan solamente en el valor de un parámetro. Elegiremos las redes del comité de modo que se diferencien en el algoritmo utilizado, en el número de capas o en el tamaño de las capas. Podrían también incluirse dos redes si tienen parámetros con valores significativamente diferentes.

Así pues, debemos eliminar de las diez mejores configuraciones aquellas que sean parecidas a otras que dan mejores resultados. Nos quedaremos, en concreto, con cinco configuraciones en cada problema (véase figura 1).

3.2 Entrenamiento final

De momento hemos conseguido conocer cuáles son las cinco mejores configuraciones de redes para cada problema. En este punto, haremos una nueva partición de los datos, con un conjunto independiente de muestras de test que nos permitirá evaluar el comportamiento de cada una de las redes individuales y, posteriormente, de los comités. De nuevo obtendremos particiones homogéneas. En la tabla 7 se muestran los tamaños de los conjuntos resultantes.

Tabla 7: *Tamaño de los conjuntos finales de entrenamiento, validación y test.*

	Nº de muestras		
Problema	*Entr.*	*Val.*	*Test*
'Analgésicos'	554	185	246
'Antidiabéticos'	193	64	86
'CIM'	62	21	28
'Solubilidad'	52	17	23

En estos entrenamientos se guarda la red que *maximiza* el porcentaje de clasificación (o minimiza el ECM) sobre las muestras de validación, limita el entrenamiento a $50,000$ iteraciones como máximo y se detiene cuando la evolución del error de validación no permite predecir una mejora.

3.2.1 Resultados de redes individuales

Los resultados obtenidos para cada problema de cada una de las redes entrenadas sobre las muestras de test y con un intervalo de confianza del 95 % se muestran en la tabla 8. El ECM de esta tabla se refiere a los valores obtenidos por el PM comparados con los valores deseados *sin deshacer la normalización*. En la siguiente sección se *desnormaliza* el ECM para estos problemas. El esquema seguido queda representado en la figura 2.

3.3 La estrategia interna del comité

En definitiva, disponemos para cada problema de cinco PMs entrenados para predecir una característica de una misma muestra. Debemos combinar los resultados de los PMs para obtener la predicción del comité. La estrategia inmediata consiste en calcular la media aritmética de esos valores. En el caso de las predicciones cuantitativas ('CIM' y 'Solubilidad') difícilmente podría hacerse de otra manera.

3.3.1 Problemas de predicción cuantitativos: 'CIM' y 'Solubilidad'

En cualquier caso, conviene observar atentamente el comportamiento comparado de las cin-

Tabla 8: *Redes individuales: resultados de test con un intervalo de confianza del 95% (v.c. es 'valor central').*

		Comportamiento en test	
Problema	*Red*	*Aciertos (%)*	*ECM*
'Analgésicos'	1	[81.31..89.94] **v.c. 86.18**	
	2	[81.76..90.29]	
	3	[82.21..90.63]	
	4	[80.41..89.24]	
	5	[79.96..88.89]	
'Antidiabéticos'	1	[87.10..97.49] **v.c. 94.19**	
	2	[84.14..96.00]	
	3	[88.64..98.18]	
	4	[85.60..96.76]	
	5	[85.60..96.76]	
'CIM'	1		0.0281869
	2		0.0281874
	3		0.0284059
	4		0.0284111
	5		0.0287237
'Solubilidad'	1		0.0103847
	2		0.0103955
	3		0.0105003
	4		0.0107689
	5		0.0111014

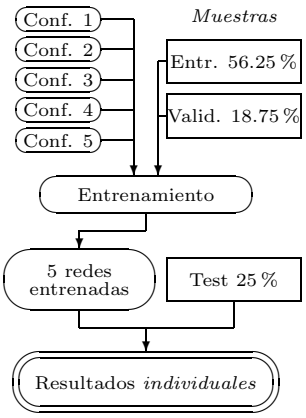

Figura 2: *Esquema para la obtención de resultados individuales de las redes previamente seleccionadas (nueva partición de los datos).*

Tabla 9: *Desnormalización en los problemas de predicción cuantitativos.*

	Comportamiento con muestras de test			
	ECM		*ECM**	
	'CIM'	'Solubilidad'	'CIM'	'Solubilidad'
Red 1	0.04777	0.02158	0.45907	1.74798
Red 2	0.04779	0.02132	0.45926	1.72692
Red 3	0.04618	0.01605	0.44379	1.30005
Red 4	0.04649	0.01480	0.44677	1.19880
Red 5	0.04831	0.02355	0.46426	1.90755
f	0.32258	0.11111		

co redes elegidas para cada problema. En la figura 3 quedan representados los valores que proporciona cada una de las redes, junto con los valores que deberían haberse obtenido para el conjunto de muestras de validación. Se aprecia claramente una fuerte correlación entre los errores cometidos por las redes: el error tiene casi siempre el mismo signo y la diferencia entre las predicciones es pequeña comparada con el error cometido.

Realmente, es innecesario buscar combinaciones de 2, 3, 4 ó 5 redes en un comité, ya que la mejora obtenida será insignificante. Lo que sí debemos mostrar (tabla 9) es el error cometido realmente por estas cinco redes con las muestras de test, una vez deshecha la normalización, siendo f el factor de normalización utilizado:

$$ECM^*(\text{desnormalizado}) = \frac{1}{f^2} ECM$$

3.3.2 Problemas de clasificación: 'Analgésicos' y 'Antidiabéticos'

Respecto a los problemas 'Analgésicos' y 'Antidiabéticos', vale la pena analizar detenidamente el comportamiento de las redes con las muestras de validación. Si nos fijamos en las posiciones que ocupan las muestras de validación mal clasificadas en cada problema y para cada una de las redes candidatas al comité, vemos que hay una fuerte correlación entre los valores que proporcionan las redes: en el problema 'Analgésicos', 125 errores se comprimen en sólo 40 muestras diferentes, es decir, cada una de esas muestras da, en promedio, un error en más de tres redes.

Pensemos en el siguiente ejemplo: una cierta muestra (que representa a una molécula activa

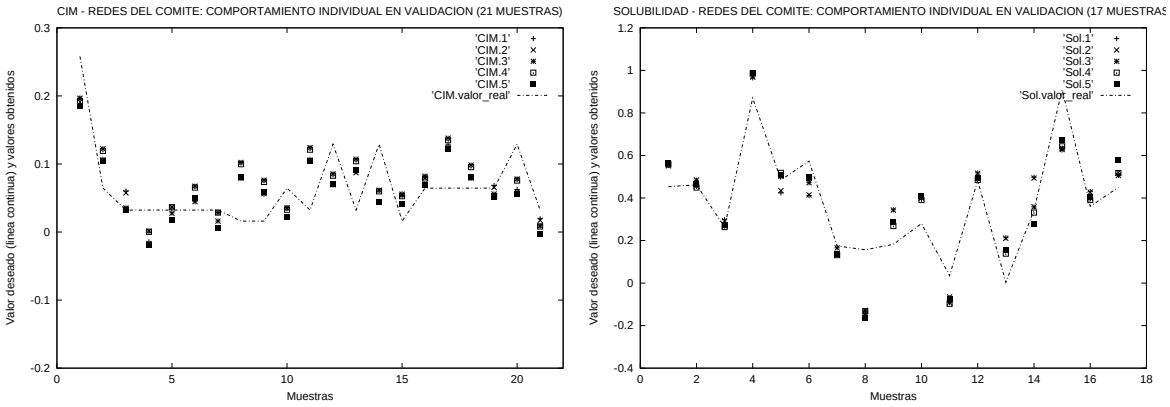

Figura 3: *Validación en 'CIM' y 'Solubilidad': valores obtenidos por redes individuales (puntos) frente a los valores reales de las muestras (línea).*

en nuestro problema, es decir, su salida debe ser 1) es analizada por un comité de tres redes, obteniéndose las salidas 0.8, 0.9 y -0.5 en cada una de ellas. Si calculamos la media aritmética para obtener la predicción del comité, tenemos:

$$\frac{1}{3} \times (0.8 + 0.9 - 0.5) = \mathbf{0.4} < 0.5$$

con lo cual el comité falla, 'arrastrado' por la tercera red. Por desgracia, esto ocurrirá un gran número de veces. Como ejemplo, en la figura 4 se muestran los resultados de test del problema 'Analgésicos' obtenidos con las redes individuales (`Ana.miembros.1`) junto con los de todos los comités posibles de 2, 3, 4 y 5 redes (`Ana.miembros.2`, ..., `Ana.miembros.5`). El esquema seguido es el de la figura 5. Los resultados de los comités se han obtenido calculando la media aritmética y se han ordenado de peor a mejor. Tal y como se aprecia en la figura 4, los resultados más favorables son los de las redes individuales. Para intentar evitar este efecto de 'arrastre', tendremos que recurrir a alguna estrategia para el comité más compleja que el promediado.

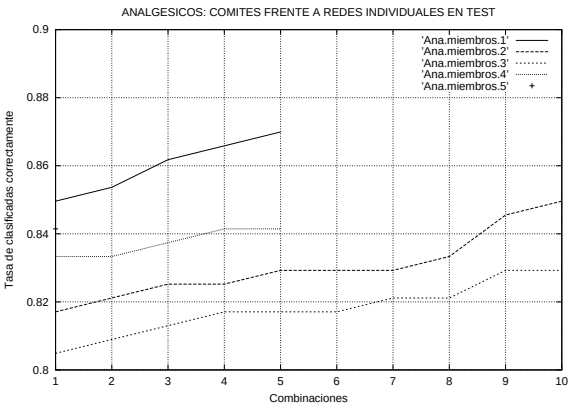

Figura 4: *'Analgésicos': comparación de resultados de las muestras de test con todos los posibles comités (con media aritmética de los valores).*

Media aritmética y umbrales

Cabe esperar que cuando una red acierta al clasificar una determinada muestra, el valor devuelto tiene un módulo cercano a 1, y que cuando una red proporciona un resultado erróneo, raramente proporcionará un valor de módulo cercano a 1 (lógicamente, con el signo erróneo) sino más bien valores en torno a 0. Si esto fuera así, cabría la esperanza de mejorar el rendimiento del comité estableciendo un umbral previo: sólo se tomarían en consideración las redes

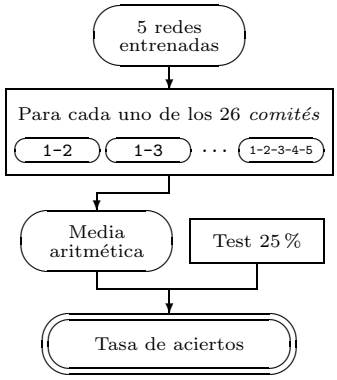

Figura 5: *Esquema de la búsqueda de comités mediante media aritmética.*

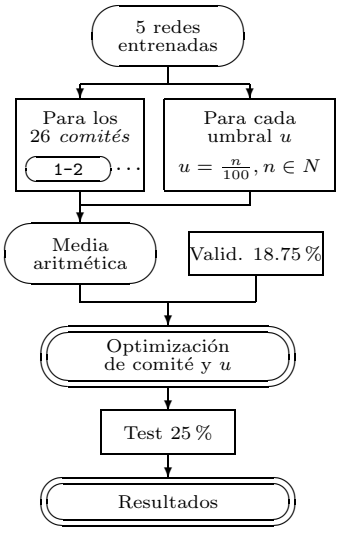

Figura 6: *Esquema de la búsqueda de comités mediante media aritmética y umbral.*

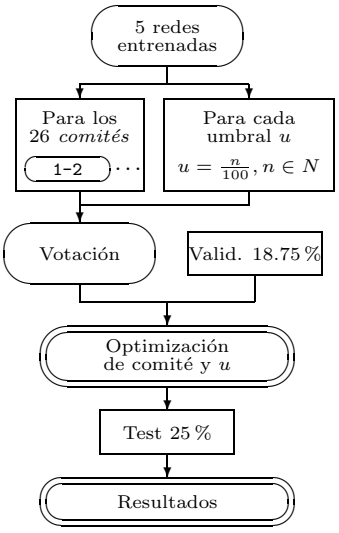

Figura 7: *Esquema de la búsqueda de comités mediante votación y umbral.*

que, para una muestra en particular, devolvieran una predicción con un módulo *superior* a ese umbral u. A continuación, se calcularía la media aritmética de los valores que cumplen ese requisito, y el resultado sería considerado acierto o error según el criterio habitual.

La cuestión fundamental es, evidentemente, fijar ese umbral y averiguar qué comité (de entre dos y cinco miembros) ofrece un mejor resultado. Para ello, realizaremos pruebas con todos los umbrales de la forma $n/100$, siendo n un número entero entre 0 y 100, y todos los comités posibles con nuestras cinco redes. En concreto, se pueden formar 26 comités diferentes: `1-2`, `1-3`, `1-4`, ..., `4-5`, `1-2-3`, ..., `1-2-3-4-5`, donde "`1-3`" representa el comité formado por las redes 1 y 3, según los identificadores establecidos anteriormente. Un umbral igual a 0 equivale a calcular simplemente la media aritmética de los valores que proporcionan *todos* los miembros del comité. Así pues, este barrido contempla también ese caso (véase figura 6).

Utilizando las muestras de validación, e incluyendo las redes individuales para poder comparar los resultados, obtenemos varias tablas como resultado de la búsqueda. En la tabla 10 sólo se incluyen las cuatro (al menos) mejores combinaciones sobre los datos de validación para los dos problemas de clasificación, 'Analgésicos' y 'Antidiabéticos'. Existe un rango de valores para el umbral dentro del cual el comité (o red individual) presenta el mismo comportamiento sobre las muestras de validación. El umbral que se muestra en la tabla 10 es el valor central de ese rango. El test se realizó con ese valor y se ha calculado el intervalo para el porcentaje de aciertos con un nivel de confianza del 95 %.

Así pues, establecer un umbral antes de calcular la media ha sido un fracaso: como mucho se iguala el resultado de la mejor red individual. Esto quiere decir que los errores *sí* tienen un módulo cercano a 1, además del signo equivocado, de modo que los errores en la mayor parte de los casos *se acumulan* y no *se anulan*.

Votaciones y umbrales

Este mecanismo consiste en realizar un recuento de la 'opinión' de los miembros del comité y la opción más votada se toma como la decisión final. En caso de empate en los comités formados por un número par de redes, realizaremos la suma de los valores de todas las redes para llegar a la decisión final en función del signo. Ya que hemos implementado el mecanismo del umbral previo, lo mantendremos y realizaremos una búsqueda similar a la anterior (véase figura 7). Los resultados se muestran en la tabla 11.

En este caso sí hemos logrado mejorar ligeramente la tasa de aciertos con las muestras de validación respecto a los resultados obtenidos con redes individuales. En los casos de varios rangos para el umbral, se toma sólo el primero

Tabla 10: *Problemas de clasificación: comités con media aritmética y umbral. Resultados de validación y test, éstos últimos con un intervalo de confianza del 95% (v.c. es 'valor central').*

'Analgésicos'			
Comité		*% Acierto*	
o red	*Umbral*	*Validación*	*Test*
`3`	0.38	87.57	[82.21..90.63] **v.c. 86.99**
`3-5`	0.85	87.03	[82.21..90.63]
`2`	0.25	86.49	[81.76..90.29]
`5`	0.28	86.49	[79.96..88.89]

'Antidiabéticos'			
Comité		*% Acierto*	
o red	*Umbral*	*Validación*	*Test*
`1`	0.31	92.19	[87.10..97.49] **v.c. 94.19**
`3`	0.44	90.63	[88.64..98.18]
`5`	0.32	90.63	[85.60..96.76]
`1-3`	0.83	90.63	[87.10..97.49]
`1-3-5`	0.94	90.63	[87.10..97.49]

Tabla 11: *Problemas de clasificación: comités con votación y umbral. Resultados de validación y test, éstos últimos con un intervalo de confianza del 95% (v.c. es 'valor central').*

'Analgésicos'			
Comité		*% Acierto*	
o red	*Umbral*	*Validación*	*Test*
`3-4-5`	0.76	89.19	[82.21..90.63] **v.c. 86.99**
`3-4`	0.61	88.65	[82.21..90.63]
`1-4-5`	0.45	88.65	[83.12..91.32]
`1-3-5`	0.45	88.65	[83.58..91.67]
`3-5`	0.79	88.65	[83.12..91.32]
`2-3-5`	0.73	88.65	[83.58..91.67]

'Antidiabéticos'			
Comité		*% Acierto*	
o red	*Umbral*	*Validación*	*Test*
`1`	0.47	92.19	[87.10..97.49] **v.c. 94.19**
`3`	0.44	90.63	[88.64..98.18]
`5`	0.44	90.63	[87.10..97.49]
`2-5`	0.44	90.63	[87.10..97.49]
`3-5`	0.44	90.63	[90.24..98.81]

de ellos y se muestran los resultados de test sólo para los cinco primeros comités.

Hay que comentar que, si elegimos nuestro modelo de comité en función del comportamiento con las muestras de validación, sólo el primer resultado de test posee rigor estadístico, ya que esa configuración del comité fue la que ocupó el primer lugar con las muestras de validación. Lógicamente, habrá algunos comités que presenten un comportamiento mejor con las muestras de test, pero esos casos no pueden considerarse resultados rigurosos.

4 Conclusiones

En este trabajo se ha demostrado la viabilidad de la utilización de redes neuronales artificiales en la discriminación y predicción de propiedades farmacológicas de moléculas representadas con un conjunto reducido de índices topológicos que describen la estructura molecular. Los mejores resultados obtenidos para cada problema se muestran en la tabla 12.

Tabla 12: *Mejores resultados finales para los problemas de clasificación ('Analgésicos', 'Ana', y 'Antidiabéticos','Ant') y los problemas de predicción cuantitativa ('CIM' y 'Solubilidad', 'Sol').*

Prob.	*% acierto o ECM*	*Predictor-clasificador*
'Ana'	86.99 %	Comité formado por 3 redes en el que se aplica el método de la *votación* con un umbral de descarte previo $u = 0.76$
'Ant'	94.19 %	Red individual de topología `62-4-4-1` entrenada mediante el algoritmo *Backpropagation Momentum*.
'CIM'	0.45907	Red individual de topología `52-64-1` entrenada con el algoritmo *Standard Backpropagation*.
'Sol'	1.74798	Red individual de topología `52-32-1` entrenada con el algoritmo *Standard Backpropagation*.

La fuerte correlación entre los resultados proporcionados por cada red neuronal individual no permite mejorar los resultados mediante la combinación de varias redes, aunque el uso de umbrales previos de 'descarte' puede resultar de ayuda en algunos casos.

Por último, los resultados sugieren que para profundizar en la resolución de estos problemas sería necesario combinar modelos conexionistas diversos como, por ejemplo, memorias asociativas. Otra estrategia podría implicar el uso herramientas alternativas (como las que se mencionan en [Ado00]) en combinación con los modelos vistos aquí.

Agradecimientos

Agradecemos la colaboración del Dr. D. Facundo Pérez y la Dra. Dña. María Teresa Salabert, del Departamento de Química Física de la Facultad de Farmacia de la Universitat de València, en la elaboración de este trabajo, especialmente por suministrarnos las muestras utilizadas en la experimentación. Asimismo, agradecemos a Cristina Adobes las herramientas desarrolladas para calcular los índices topológicos de las moléculas.

Referencias

[Ado00] Cristina Adobes. Diseño e implementación de herramientas para la predicción de propiedades moleculares. Proyecto Final de Carrera, Facultad de Informática, Universitat de València, 2000.

[Bis95] C. M. Bishop. *Neural networks for pattern recognition*. Oxford University Press, 1995.

[CC00] M. J. Castro and F. Casacuberta. Committees of MLPs for Acoustic Modeling. En *Proceedings of 5th Iberoamerican Symposium on Pattern Recognition*, páginas 797–807, Lisboa (Portugal), Septiembre 2000.

[DCA+01] Wladimiro Díaz et al. Discriminación de la actividad farmacológica utilizando técnicas conexionistas. En *Actas de la IX Conferencia de la Asociación Española para la Inteligencia Artificial*, volumen I, páginas 233–241, Gijón (España), Noviembre 2001.

[DHS01] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley and Sons, 2001.

[Has94] Sherif Hashem. Optimal Linear Combinations of Neural Networks. Technical report, Pacific Northwest Laboratory, Richland, 1994.

[PC93] Michael P. Perrone and Leon N. Cooper. When Networks Disagree: Ensemble methods for Hybrid Neural Networks. En R. J. Mammone, editor, *Neural Networks for Speech and Image Processing*. Chapman and Hall, 1993.

[V+96] V. Venkatasubramanian et al. *Genetic Algorithms in Molecular Modeling*, capítulo Computer-Aided Molecular Design Using Neural Networks and Genetic Algorithms. 1996.

[Z+98] A. Zell et al. *SNNS: Stuttgart Neural Network Simulator. User Manual, Version 4.2*. Institute for Parallel and Distributed High Performance Systems, University of Stuttgart, Germany, 1998.