

Capítulo 1.- Redes Neuronales y Reconocimiento de Patrones.

Autores: **Dr. Luis Alonso Romero***. **Dr. Teodoro Calonge Cano****.

* Dpto. de Informática y Automática. Universidad de Salamanca- España

** Dpto de Informática. Universidad de Valladolid. España

1.1.- Introducción.

¿Qué es un patrón? Siguiendo la definición de Watanabe [Wat85], un *patrón* es una entidad a la que se le puede dar un nombre y que está representada por un conjunto de propiedades medidas y las relaciones entre ellas (*vector de características*). Por ejemplo, un patrón puede ser una señal sonora y su vector de características el conjunto de coeficientes espectrales extraídos de ella (espectrograma). Otro ejemplo podría ser una imagen de una cara humana de las cuales se extrae el vector de características formado por un conjunto de valores numéricos calculados a partir de la misma. El reconocimiento automático, descripción, clasificación y agrupamiento de patrones son actividades importantes en una gran variedad de disciplinas científicas, como biología, psicología, medicina, visión por computador, inteligencia artificial, teledetección, etc.

Un sistema de reconocimiento de patrones tiene uno de los siguientes objetivos:

- a.- Identificar el patrón como miembro de una clase ya definida (*clasificación supervisada*).
- b.- Asignar el patrón a una clase todavía no definida (*clasificación no supervisada, agrupamiento o clustering*).

El diseño de un sistema de reconocimiento de patrones se lleva a cabo normalmente en tres fases:

- i.- Adquisición y preprocesado de datos.
- ii.- Extracción de características.
- iii.- Toma de decisiones o agrupamiento.

El universo del discurso, o dominio del problema, gobierna la elección de las diferentes alternativas en cada paso: tipo de sensores, técnicas de preprocesamiento, modelo de toma de decisiones, etc. Este conocimiento específico del problema está implícito en el diseño y no se representa como un módulo separado como sucede, por ejemplo, en los sistemas expertos.

Tradicionalmente, el reconocimiento de patrones se ha abordado desde un punto de vista estadístico, dando lugar al llamado reconocimiento estadístico de patrones (REP), del cual se darán unas breves pinceladas en el punto siguiente. No obstante, existe una alternativa que se ha revelado como muy prometedora en algunos casos en que el REP no funciona satisfactoriamente. Dicha alternativa son las Redes Neuronales Artificiales (RNA) que se verán también. Finalmente, presentaremos una visión de los puntos comunes entre ambas técnicas.

1.2.- Reconocimiento Estadístico de Patrones (REP).

El REP es una disciplina relativamente madura hasta el punto de que existen ya en el mercado un cierto número de sistemas comerciales de reconocimiento de patrones que emplean esta técnica. En REP, un patrón se representa por un vector numérico de dimensión n . De esta forma, un patrón es un punto en un espacio n -dimensional (*de características*). Un REP funciona en dos modos diferentes: entrenamiento y reconocimiento. En *modo de entrenamiento*, se diseña el extractor de características para representar los patrones de entrada y se entrena al clasificador con un conjunto de datos de entrenamiento de forma que el número de patrones mal identificados se minimice. En el *modo de reconocimiento*, el clasificador ya entrenado toma como entrada el vector de características de un patrón desconocido y lo asigna a una de las clases o categorías.

El proceso de toma de decisiones en un REP se puede resumir como sigue. Dado un patrón representado por un vector de características

$$\mathbf{x} = (x_1, x_2, \dots, x_n)^T$$

Asignarlo a una de las c clases o categorías, C_1, C_2, \dots, C_c . Dependiendo del tipo de información disponible sobre las densidades condicionales de las clases, se pueden diseñar varias estrategias de clasificación. Si todas las densidades condicionales $p(\mathbf{x} | C_i)$, $i=1,2,\dots,c$ se conocen, la regla de decisión es la de Bayes que establece los límites entre las diferentes clases. Sin embargo, en la práctica las densidades condicionales no se conocen y deben ser estimadas (*aprendidas*) partiendo de los patrones de entrada. Si se conoce la forma funcional de estas densidades pero no sus parámetros, el problema se llama de *toma de decisión paramétrico*. En caso contrario, estamos ante un problema de toma de decisión no paramétrico. Las diferentes dicotomías que aparecen al diseñar un sistema de REP se muestran en la figura 1.1.

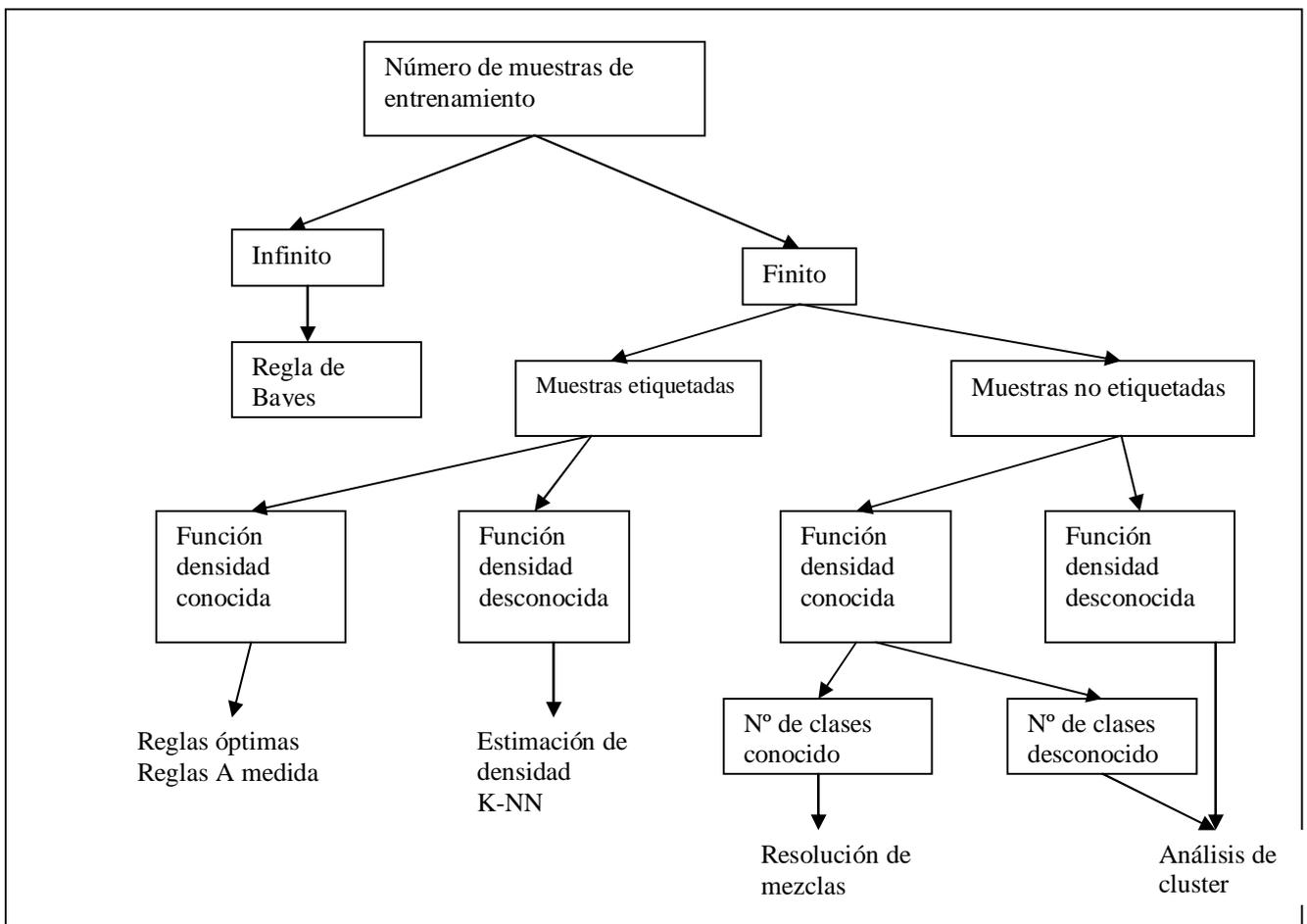


Figura 1.1.- Dicotomías en el diseño de un sistema de REP.

1.3.- Redes Neuronales Artificiales.

La *neurocomputación* es una aportación más al viejo objetivo de crear sistemas *inteligentes*, considerando como tales a máquinas capaces de llevar a cabo tareas que exhiben alguna de las características asociadas a la inteligencia humana. En las dos últimas décadas, los avances en este

campo han sido espectaculares; en particular el desarrollo de las redes neuronales artificiales (RNA). Originalmente, los trabajos en RNA surgen de la idea de que para que las máquinas puedan llevar a cabo dichas tareas inteligentes, sería conveniente que el modelo de computación se asemejara más a la fisiología del cerebro humano que al modelo computacional vigente por aquellas fechas: modelo von Neumann. Sin embargo, el auge que estos sistemas se debe más al éxito obtenido en aplicaciones reales (reconocimiento de patrones, predicción, optimización, etc.) que a la semejanza con el modelo biológico. Por ejemplo, el perceptrón multicapa, que es una de las redes más utilizadas, es criticada por su escaso parecido con el funcionamiento de las neuronas dentro del cerebro humano especialmente en todo lo referente a su algoritmo de aprendizaje.

En cualquier caso, lo que se plantea es un modelo computacional alternativo a la máquina von Neumann o a los ordenadores paralelos actuales, los cuales no están dotados de forma global de las siguientes características:

- Masivamente paralelos
- Computación y representación distribuida
- Aprendizaje
- Generalización
- Adaptabilidad
- Procesamiento de información inherente al contexto
- Tolerante a fallos
- Bajo consumo de energía

Estas propiedades son un fiel reflejo de alguna de las que poseen las redes neuronales biológicas. En efecto, una neurona es una célula especializada compuesta de un cuerpo o soma y dos tipos de ramificaciones denominadas dendritas y axón (ver Fig. 1.2). A través de las dendritas recibe señales (impulsos de naturaleza electroquímica) procedentes de otras neuronas. El cuerpo de la neurona, en función de estas señales de entrada, genera una salida que se transmite a través del axón a otras neuronas. La conexión entre las ramificaciones del axón y las de las dendritas se realiza mediante una estructura funcional elemental llamada sinapsis, que a su vez puede ser inhibitoria o excitadora dependiendo del valor del potencial electroquímico de las señales transmitidas. Aunque el mecanismo de funcionamiento es bastante más complejo, lo expuesto es suficiente para entender el origen de las RNA. Finalmente, hay que señalar que el cortex del cerebro humano posee del orden de 10^{11} neuronas, cada una de las cuales tiene conexión aproximadamente con otras 10^3 ó 10^4 neuronas. Esto hace un total de entre 10^{14} y 10^{15} interconexiones. Así pues, si se tiene en cuenta por un lado que el reconocimiento de una cara o de un carácter conlleva unos pocos milisegundos, y, por otro la lentitud de transmisión de las señales electroquímicas respecto a un medio conductor, llevaría a establecer no más de 100 cambios de estados no simultáneamente dentro de la red. Este hecho asegura, por tanto, que la información crítica para llevar a cabo una tarea no es transmitida directamente, sino que es distribuida entre las diferentes interconexiones; de ahí el término *conexionista* con que se califica al modelo introducido por las RNA.

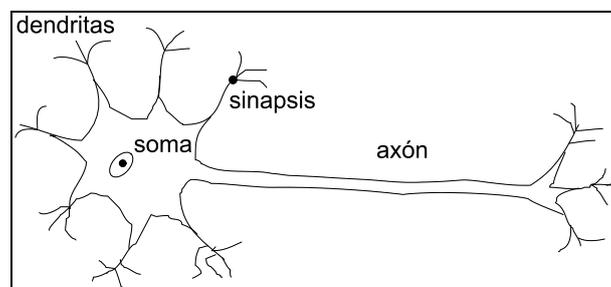


Fig. 1.2: neurona biológica

Cronológicamente, la investigación en RNA ha transcurrido por tres periodos de extensa actividad. Aunque muchos autores se remontan más atrás, es en los años 40 cuando el neurobiólogo McCulloch y el estadístico Pitts (1943) presentan un trabajo pionero en este campo. En él se propone un modelo matemático de tratamiento de información para una neurona artificial basado en su homónimo biológico. El segundo periodo se sitúa en los años 60 con la aparición del teorema de la convergencia del perceptrón (1962) debido a Rosenblatt. Siete años después, tras la euforia de esta aportación, Minsky y Papert publican un trabajo donde se muestran serias limitaciones de la red neuronal propuesta por Rosenblatt, esto es, el perceptrón, cuyos elementos de proceso estaban basados en la neurona de McCulloch y Pitts. Tras un periodo de casi veinte años, en donde las RNA cayeron en el olvido, se dan a conocer varios trabajos que contribuyeron al resurgimiento de estas ideas; de entre ellos destacan dos. El primero es debido a Hopfield donde propone una red neuronal que lleva su nombre; su arquitectura y funcionamiento hacen que sea el sistema que más se asemeja al modelo biológico hasta el momento. Finalmente, Rumelhart y McClelland [Rumel 86] popularizan el algoritmo de aprendizaje de retropropagación para el perceptrón multicapa, que originariamente se atribuye a Werbos [Werb74], quien lo propuso doce años antes.

Desde el punto de vista computacional, una RNA puede ser descrita como un conjunto de autómatas celulares (*neuronas*), por el cual se establece un flujo de información mediante una topología de interconexiones (*sinapsis*).

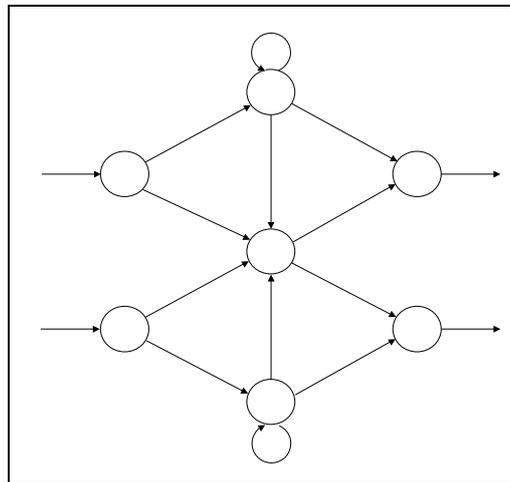


Fig.1.3.: autómatas celulares

En la mayoría de las RNA propuestas, cada neurona procesa la información según el modelo propuesto por McCulloch-Pitts.

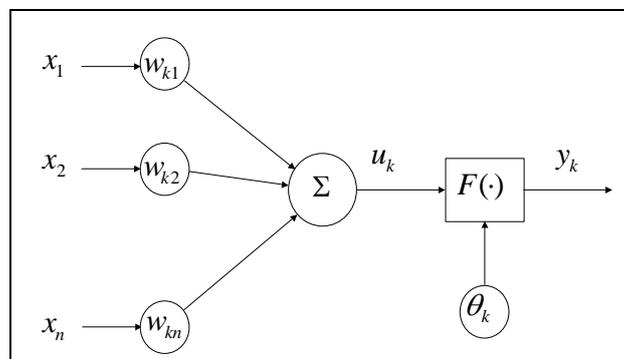


Fig..1.4 : neurona artificial (McCulloch ó Pitts)

En síntesis cada neurona obtiene su salida como la suma ponderada de cada una de sus entradas previa multiplicación por su correspondiente peso. Si dicho peso es positivo, se dirá que la entrada es

excitadora; en caso contrario se hablará de inhibidora de acuerdo con el modelo biológico. Desde el punto de vista matemático, la salida así calculada puede alcanzar cualquier magnitud. Sin embargo, es sabido que en la naturaleza los valores de los potenciales electroquímicos están acotados. Para dar cuenta de este hecho, la salida obtenida es filtrada por una función de activación $F(x)$, que bien podría tratarse de una función salto $[0,1]$ desplazada del origen un cierto umbral θ_k . Esta función estaría en sintonía con los conocimientos fisiológicos acerca del funcionamiento de las neuronas, puesto que cuando una determinada señal nerviosa (electroquímica) no supera un cierto umbral, ésta se convierte en inhibidora y viceversa. A pesar de ello, no es la función salto la más comúnmente empleada sino otra con un perfil parecido: la sigmoide. La razón es de conveniencia numérica, ya que es una función n-diferenciable.

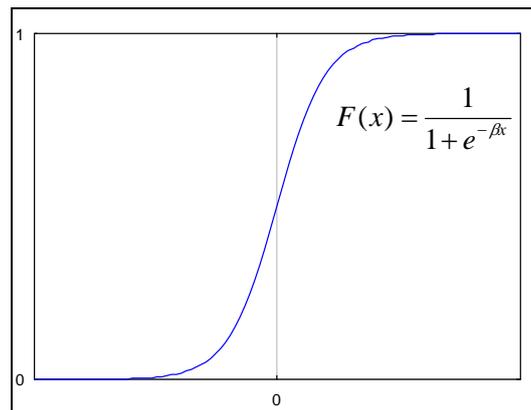


Fig.1.5: función sigmoide

Otra propiedad no menos importante, es el grado de paralelismo inherente en el procesamiento de la información. Tanto en las redes biológicas como en las artificiales, el evento que pone en marcha la obtención de la salida es la disponibilidad de las entradas adecuadas. No obstante, una neurona es 5 ó 6 órdenes de magnitud más lenta a la hora de producir la salida, que cualquiera de las puertas lógicas VLSI empleadas en la fabricación de los más modernos microprocesadores digitales. Por tanto, aunque computacionalmente una neurona sea muy simple y muy lenta, el número de ellas que posee el ser humano y la cantidad que pueden trabajar simultáneamente, es más que suficiente para hacerse una idea de la gran potencia de procesamiento de la información del cerebro humano [Hay94].

Una RNA (ver Fig. 1.3) puede ser considerada como un grafo dirigido y ponderado. Cada uno de sus nodos representa a una neurona artificial. Si el arco es de entrada, llevará asociado en cada momento un valor o peso (grafo ponderado). Dependiendo de las características adicionales del grafo resultante se establecen diferentes clasificaciones. En particular se habla de redes con o sin lazos de realimentación. A continuación, según este criterio se clasificarán algunos de los tipos de redes más conocidas.

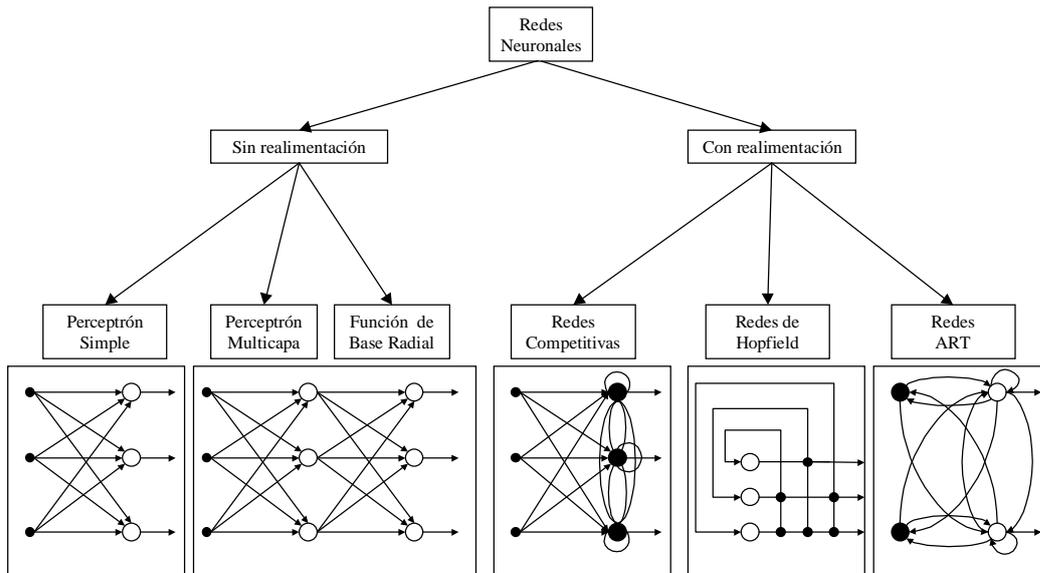


Figura 1.6.- Una taxonomía de RNA

De todo lo expuesto anteriormente se deduce que una neurona biológica puede ser considerada como un dispositivo altamente no lineal, integrado en un sistema masivamente paralelo, dotado de una gran robustez y tolerante a fallos. Aparte de esto, cabe destacar también:

- El aprendizaje mediante la adaptación de sus pesos sinápticos a cambios en el entorno.
- El manejo de información imprecisa, difusa, con ruido, y, en ciertas ocasiones basada en distribuciones de probabilidad.
- La generalización de tareas o ejemplos desconocidos, a partir de otros que sí los son.

El modelo computacional introducido por las RNA trata de acomodarse a todas estas características. De hecho, los algoritmos clásicos basados en una secuencia de instrucciones programadas dan paso a nuevos paradigmas, donde la información se encuentra almacenada en las diferentes conexiones entre neuronas. Teóricamente, con la red y pesos adecuados, según el modelo de McCulloch ó Pitts, es suficiente para realizar cualquier tipo de cómputo.[Pal 96] En la práctica, encontrar la red y los pesos no es un problema fácil. Actualmente, se trabaja con ciertos tipos de redes, los cuales no son universales, sino que funcionan bien para ciertas tareas y para otras no.

Asociado a cada tipo de red existe al menos un algoritmo de aprendizaje. Consiste en un método sistemático para encontrar un valor adecuado de los pesos. En términos generales, se basan en definir una función objetivo implícita o explícitamente, que represente de forma global el estado de la red. A partir de ella, los pesos asignados inicialmente van evolucionando a unos valores que lleven a dicha función a un mínimo (estado estable de la red). El aprendizaje, por tanto, es algo característico del tipo de red.

El *aprendizaje supervisado* maneja pares entrada ó salida deseada. De esta manera, para una determinada entrada se compara la salida obtenida por la red con la deseada. El ajuste de los pesos está dirigido, por tanto, a minimizar en lo posible la diferencia entre dichas salidas. Un ejemplo muy típico de aprendizaje supervisado corresponde al de *retropropagación del error* para el perceptrón multicapa.

En el *aprendizaje no supervisado*, el planteamiento es totalmente diferente. Aunque a la red se le presentan entradas para su entrenamiento (aprendizaje), no existen salidas deseadas, sino que el sistema evoluciona de forma autoorganizada hasta un estado considerado estable.

En cualquier caso, la red aprende utilizando ejemplos, pero lo realmente atractivo de estos sistemas es la capacidad de generalización. Esto se refiere a la calidad de la respuesta ante entradas que no han sido utilizadas en su entrenamiento. Por tanto, cabe distinguir dos modos o fases de funcionamiento de una RNA: *entrenamiento* y *reconocimiento*. Así pues, una vez fijados los pesos en la fase de entrenamiento, la red pasa a la fase de reconocimiento, donde procesa entradas correspondientes a la aplicación real.

Diversos problemas en computación han sido tratados utilizando total o parcialmente algún tipo de RNA, como por ejemplo: reconocimiento de patrones, predicción, memorias asociativas o control. Ciertamente, no son problemas nuevos, sino que han sido planteados hace mucho tiempo. Para su resolución, se aplicaban ciertas técnicas convencionales, como el REP, que ofrecían buenos resultados en experimentos dentro de entornos controlados pero, en general, la técnica utilizada presentaba malos resultados al cambiar de entorno. Este inconveniente ha desaparecido en muchos casos con la utilización de RNA, debido a la robustez y flexibilidad con que procesa no sólo la información propia sino también la relativa al contexto. Quizás una de las aplicaciones más populares, donde se han aplicado las RNA, sea en reconocimiento de patrones. Dentro de este campo se hace distinción entre patrones estáticos y dinámicos, esto es, aquellos que no contemplan explícitamente la variable tiempo y los que sí. En cualquier caso, lo que se busca es un sistema capaz de asignar lo más correctamente posible cada entrada a una clase o patrón.

El *clustering* o agrupamiento es una tarea conocida también como clasificación no supervisada de patrones. A diferencia del reconocimiento de patrones, no existen etiquetas prefijadas de antemano para las diversas clases. El sistema de por sí es quien extrae las características diferenciadoras entre clases, y, por lo tanto, fija las fronteras entre nubes de puntos. Dichas fronteras no necesariamente se ubican en el espacio de entradas, sino que bien pudieran establecerse en un espacio de características subyacente. Una red especialmente diseñada para este tipo de problemas es el mapa autoorganizado de Kohonen.

La predicción es otro problema para el que se está empleando RNA, aún cuando, formalmente, pueda asimilarse a un problema de reconocimiento de patrones. En este caso se pretende averiguar con un cierto grado de error, cuál sería el siguiente término en una serie temporal. Para ello, la red es entrenada con datos conocidos hasta un cierto instante. A partir de ellos, se espera que el sistema obtenga el valor para instantes posteriores. En la práctica esto puede aplicarse a predicción meteorológica, de valores bursátiles o de demanda de carga en una red de distribución de energía eléctrica. Si las propiedades estadísticas de la serie temporal son estacionarias, las redes estáticas como el perceptrón multicapa dan buenos resultados. En caso contrario, son las redes dinámicas o recurrentes las que dan mejores respuestas.

Las memorias asociativas o direccionables por contenidos constituyen otra aplicación de las RNA. Mediante su uso se pretende recuperar una cierta información almacenada, de la cual se conoce su contenido parcialmente o bien se encuentra distorsionado por alguna causa. En la actualidad, con el desarrollo de grandes bases de datos multimedia esta técnica está implantándose ampliamente.

Finalmente, en el ámbito del control de sistemas, también se están empleando las RNA. Un ejemplo muy típico es la generación de señales de control para que la salida del sistema siga la consigna de referencia o de entrada. Es problema muy conocido, que ha sido resuelto hace varias décadas para sistemas lineales. Sin embargo, donde realmente resulta útil y novedoso el uso de RNA es en sistemas muy complejos y altamente no lineales.

1.4.- Puntos comunes entre REP y RNA.

Las RNA y las técnicas de REP están muy relacionadas y comparten algunos de los problemas de diseño. En la comunidad de Redes Neuronales se considera el reconocimiento de patrones como una de los problemas más desafiantes. Por otro lado, en la comunidad de reconocimiento de patrones, las RNA se consideran un potente complemento a las técnicas clásicas. Es difícil estimar el grado de solapamiento entre las RNA y el REP porque ello requeriría el conocimiento de los límites precisos entre ambas disciplinas. Werbos[Werb 91], estima que un 80% del trabajo que se realiza con Redes Neuronales se hace en el campo del reconocimiento de patrones. Alguno de los puntos comunes entre ambas técnicas se plantean en los campos de representación, extracción de características y clasificadores.

1.4.1.- Representación.

Una buena representación de patrones debería cumplir, al menos, los siguientes requisitos:

- Tasa de compresión de datos alta.
- Buena capacidad discriminatoria.
- Invariancia frente a transformaciones de los datos.
- Robustez frente al ruido.

En la mayoría de los sistema de REP, los esquemas de representación son desarrollados por los diseñadores usando su conocimiento y experiencia en el dominio del problema. Una vez que el sistema de reconocimiento está desarrollado, estos esquemas son inamovibles. En muchas aplicaciones con RNA se sigue el mismo procedimiento de forma que la red neuronal lleva a cabo, en esencia, el proceso de clasificación. Sin embargo, las redes neuronales tienen la propiedad de construir una representación interna de los patrones (*extracción de características*), aunque difícilmente visible. Por esta razón, algunos investigadores alimentan a la red con los datos en bruto (o con un preproceso mínimo, como normalización) y esperan que la propia red extraiga (aprenda) una representación a partir de ellos.

En cualquier caso, una representación adecuada de los datos facilita el proceso de toma de decisión y mejora las tasas de generalización. Sin embargo, el diseño de una buena representación exige un conocimiento profundo de la naturaleza del problema, lo cual no siempre es posible. La forma de aprender un esquema de representación partiendo de un conjunto de datos es todavía un problema abierto.

1.4.2.- Extracción de características.

El término "*extracción de características*" tiene un doble significado. Por un lado, se refiere al proceso de extraer alguna medidas numéricas de los datos en bruto de los patrones (representación inicial). Por otro lado, se define también como el proceso de formar un conjunto de características (de dimensión n) partiendo de los datos de entrada (de dimensión $m > n$).

La extracción de características y la proyección de datos multivariantes son aspectos importantes en REP. La extracción de características puede evitar el problema de la dimensionalidad, mejorar la tasa de generalización y reducir los requisitos computacionales del clasificador. La proyección de datos nos permite visualizar en 2D o 3D datos de dimensión mayor para facilitar el análisis. Existen alguna técnicas clásicas para extraer y proyectar datos, que se pueden clasificar en cuatro grupos:

- Lineales no supervisadas: análisis de componentes principales (PCA).
- Lineales supervisadas: análisis discriminante lineal.

- No lineales, no supervisadas: análisis discriminante no paramétrico.
- No lineales, supervisadas: algoritmo de Sammon.

Para extracción de características y proyección de datos multivariantes existen multitud de algoritmos basados en RNA, que pueden clasificarse en dos grupos:

- Nuevas RNA diseñadas explícitamente con este fin: SOM, NP-SOM, [Koh 94]
- Extracción de reglas de formación partiendo de los pesos de una RNA entrenada. Esta es una línea de investigación muy activa que trata de encontrar formas de representar explícitamente el conocimiento almacenado en los pesos de la red. Con ello se combinarían las ventajas inherentes a las RNA (adaptabilidad, rapidez en la fase de reconocimiento, etc) con las de las técnicas clásicas (conocimiento explícito del problema).

1.4.3.- Clasificación.

En REP, el objetivo es asignar un patrón de entrada \mathbf{x} , de dimensión n a una de las c clases o categorías, C_1, C_2, \dots, C_c . Si las densidades *condicionales* $p(\mathbf{x} | C_i)$ y las c densidades *a priori*, $p(C_i)$ se conocen, la estrategia óptima de clasificación es la regla de Bayes: asignar \mathbf{x} a la clase C_k si se cumple la condición:

$$p(C_k | \mathbf{x}) > p(C_i | \mathbf{x}), \quad \forall i = 1, 2, \dots, c, i \neq k$$

donde $p(C_k | \mathbf{x})$ se llama la probabilidad *a posteriori* y se calcula por la ley de Bayes. En REP una regla de decisión se puede formular a menudo en forma de una función discriminante $g_i(\mathbf{x})$, $i=1\dots c$. Se asigna \mathbf{x} a la clase k si $g_k(\mathbf{x}) > g_j(\mathbf{x})$, $\forall j \neq k$. Algunas RNA muy usadas, como el perceptrón multicapa o las funciones de base radial, en realidad, calculan funciones discriminantes no lineales. En REP la forma de calcular las funciones discriminantes es bastante más complicada e implica establecer estimaciones de las densidades condicionales.

El perceptrón multicapa es un clasificador óptimo en el sentido de Bayes si las densidades condicionales de las clases son gaussianas con matrices de covarianza idénticas.[Pao 89]. Si estas premisas no se cumplen, el funcionamiento del perceptrón puede ser muy pobre y es necesario tomar algunas medidas adicionales para mejorarlo. En muchos problemas de REP no se conoce ni siquiera la forma de las densidades condicionales. En estos casos o bien se postula una forma, o se diseñan métodos no paramétricos como los clasificadores de los k vecinos. En el caso de las RNA, existen una serie de redes que pueden abordar este problema con ciertas garantías de éxito: las redes neuronales probabilísticas (PNN), la máquina de Boltzmann que aprende las densidades condicionales [Hin 86] con una serie de variantes (máquina de Boltzmann determinista, por ejemplo).

En los clasificadores de árbol la decisión final se alcanza por una secuencia de decisiones intermedias (nodos no terminales) siguiendo un camino desde un nodo raíz a un nodo hoja terminal. En cada nodo, se usa solamente un subconjunto de las características para decidir. Se ha demostrado que un clasificador de árbol equivale a un perceptrón de tres capas, aunque no se ha encontrado evidencia de cual de ellos funciona mejor [Set 90].

En las técnicas de agrupamiento (clustering) se ha demostrado que las RNA empleadas para tal fin (SOM, ART, LVQ) son equivalentes a técnicas de REP (k medias, conductor secuencial, etc). No obstante, quedan todavía muchos puntos difíciles sin resolver. En particular la relación entre el tipo de métrica (distancia) empleada y las prestaciones del clasificador.

La capacidad de generalización de los clasificadores, clásicos o basados en RNA, depende básicamente de tres factores:

- Número de patrones de entrenamiento.
- Las densidades condicionales presentes en los datos.
- La complejidad del clasificador.

Las relaciones son más o menos conocidas si las distribuciones son gaussianas. En caso contrario, el problema de determinar las tasa de generalización de un clasificador es intratable y el método de prueba y error es el único disponible. Intuitivamente, parece obvio que cuantas más muestras de entrenamiento se utilicen mejor será el clasificador. Pero no siempre se puede disponer de muchas muestras y por este motivo se están desarrollando técnicas que mejoren las tasas de generalización de una red neuronal, que son similares a las empleadas para seleccionar el mejor clasificador de un conjunto dado en REP [Rau 91]. Estas técnicas se basan en algunas de las siguientes categorías:

- Compartición de conexiones locales y pesos.
- Podado adaptativo de la red
- Crecimiento adaptativo de la red
- Regularización.

Los métodos empleados son bastante variados. Algunos se basan en mecanismos biológicos, otros son métodos heurísticos, otros emplean algoritmos genéticos, etc. Las técnicas de regularización proporcionan un procedimiento para inyectar en la red conocimiento a priori sobre el problema en cuestión con el objetivo de reducir el número de parámetros de la red [Mao 93] y conseguir una mejor tasa de generalización.

1.5.- Conclusión.-

Las RNA proporcionan una batería de tratamientos nuevos, o complementarios, para abordar el problema de reconocimiento de patrones. Pero además se han desarrollado arquitecturas computacionales que pueden emplearse para la implementación de alguno de los algoritmos de REP. La adaptabilidad de las RNA es crucial para los problemas de reconocimiento no solamente para mejorar la tasa de generalización sino también para permitir un buen comportamiento frente a cambios en el entorno o frente a datos de entrada incompletos o ruidosos. En este sentido, los métodos basados en RNA son mejores que los clásicos de REP. Sin embargo, las RNA pueden también beneficiarse de algunos resultados bien conocidos en REP, especialmente en el campo de representación de patrones, donde se encuentran actualmente los desafíos importantes en el modelado neuronal. Muy posiblemente ambos campos, RNA y REP, deberán llegar tarde o temprano a una especie de sinergia que combine las fortalezas de los dos tratamientos a la hora de abordar problemas de reconocimiento de patrones muy complejos. En esta sinergia no sólo son importantes los componentes individuales de extracción de características, toma de decisión y agrupamiento sino que es esencial una buena metodología de integración de todos ellos. Este punto de vista coincide con el expuesto por Kanal [Kan 93] y por Minsky [Min 91], y es una de las líneas en que el reconocimiento automático de patrones evolucionará en el futuro.

BIBLIOGRAFÍA.-

[Hay 94] Haykin S, "Neural Networks - A comprehensive foundation", IEEE Press - Macmillan College Publishing Company, Inc. 1994.

[Hin 86] Hinton, Sejnowsky, "Parallel Distributed Processing", MIT Press 1986

[Kan 93] Kanal, L.L. "On pattern, categories and alternate realities", Pattern recognition, Marzo 93

[Koh 94] Kohonen T. "Self Organization and Associative Memory", 4th Edition. Springer 1994

- [Mao 93] Mao, Jain. "Regularization techniques in ANN", WCNN, 1993
- [Min 91] Minsky, M. " Logical versus analogica os symbolic versus connectionist or neat versus scruffy", AI Magazine, Vol 12, 1991
- [Pal 96] S.K.Pal and P.K.Srrimani. òNeurocomputing: motivation, models and hybridizationö, Computer Magazine, IEEE Computer Society, Vol. 29, No. 3, March 1996, pág. 24.
- [Pao 89].- Pao, Y.H.- Adaptive Pattern Recognition and Neural Networks. Addison- Wesley 1989.
- [Rau 91]. Raudys, Kain.- Small sample side problems in designing ANN.- IEEE Trans Pattern Analysis, Marzo 91
- [Rumel86] D.E.Rumelhart and J.L. McClelland, òParallel Distributed Processing: Exploration in the Microstructure of Cognitionö, MIT Press, Cambridge, Mass. 1986.
- [Set 90] Sethi, I.K.- Entropy Nets: from decision trees to neural networks.- Proc IEEE, Oct 1990.
- [Wat85] S.Watanabe.- Pattern Recognition: Human and Mechanical.- Wiley, New York 1985
- [Werb 74] P.J. Werbos, òBeyond Regression: New Tools for Prediction and Analysis in the Behavioural Sciencesö, PhD thesis, Dept. of Applied Mathematics, Harvard University, Cambridge, Mass., 1974.
- [Werb 91] P.J.Werbos. "Links between ANN and statistical pattern recognition".- En "Artificial Neural Networks and Pattern Recognition", Sethi, Jain.- Elsevier 1991