CAPITULO 1

INTRODUCCIÓN

Durante más de 40 años las últimas generaciones humanas han visto como la computadora, en sus más variadas formas, ha ido ocupando espacios de la vida cotidiana. Muchas aplicaciones son visibles en forma directa, tales como la calculadora o computadoras personales, los cajeros automáticos, etc. Tantas otras están más ocultas aunque no por eso son menos importantes e incluyen aplicaciones a instrumental médico, supervisión y seguridad de automóviles, comunicación digital, etc. Además, el uso masivo de las computadoras, controladores inteligentes y robots ha permitido que la producción industrial bajara sus costos, haciendo que productos tradicionalmente caros sean ahora de consumo masivo.

La intención aquí no es hacer un análisis sociológico sobre la mejora de la calidad de vida con la computadora, sino, someramente, justificar el gran interés de la comunidad científica del área por hacer computadoras más potentes, versátiles y económicas.

El desarrollo de las ciencias de la computación ha seguido la linea arquitectónica del matemático John von Neumann que en 1952 diseñó una estructura basada en un procesamiento secuencial de datos e instrucciones. Esta estructura sigue rigurosamente un programa secuencial almacenado en memoria. La arquitectura de von Neumann se basa en la lógica de procedimientos que comúnmente utilizamos, hallando soluciones parciales a un problema, y que luego son utilizadas para lograr la solución final.

Contrariamente, una visión del cerebro a distancias microscópicas no se condice con estructuras tipo von Neumann ni con programas almacenados. En cambio, existe una computación masiva de unidades concurrentes y redundancia de tareas y conexiones que proporcionan robustez. El cerebro está especialmente orientado al procesamiento de información sensorial compleja y ruidosa. El ejemplo clásico y más ilustrativo es el de la visión. En dicho

proceso intervienen más de 10 mil millones de neuronas. Ninguna computadora secuencial es capaz de alcanzar esta potencia de cómputo. Además, desde el punto de vista algorítmico, no existe ningún sistema artificial que pueda emular esta función con resultados similares. En otras palabras, el cerebro posee cualidades indiscutibles para la realización de tareas cognitivas, algorítmicamente complejas y con señalización ruidosa.

Si bien el futuro de las computadoras tradicionales parece cada vez más promisorio y complementario de las habilidades humanas, los sistemas neuronales han atraído la atención de un gran número de científicos, entre ellos biólogos e ingenieros, por diferentes motivos. La ciencia neurobiológica se halla interesada en entender el funcionamiento del cerebro y las estructuras y procedimientos involucrados. En cambio, la ingeniería eléctrica y ciencias de computación están interesadas en obtener paradigmas computacionales con raíces en la neurobiología, para su aplicación en problemas tradicionales tales como reconocimiento temporal y espacial de patrones, visión, voz, control, comunicaciones, etc.

El espectro de investigaciones es muy amplio. Si bien en muchos casos, biólogos, físicos, matemáticos e ingenieros comparten un mismo proyecto, en otros, y en particular en nuestra área de interés, se utilizan modelos neuronales artificiales que no respetan al pie de la letra los postulados biológicos, de los cuales solo toman algunas pocas características beneficiosas. A este respecto podemos decir, que aun teniendo en cuenta la gran diversidad de modelos neuronales artificiales y clasificaciones existentes, existen dos características fundamentales y unificadoras: paralelismo masivo y aprendizaje. Estos temas serán discutidos en profundidad en varios puntos del presente trabajo.

Si miramos a los modelos neuronales artificiales como un conjunto de herramientas para solucionar problemas de ingeniería, como los que ya han sido mencionados, deben considerarse tres aspectos fundamentales:

- representación
- aprendizaje

- implementación

El problema de la representación está íntimamente ligado al desarrollo de la teoría y de los modelos neuronales. Se refiere a las clases de funciones que pueden representarse con una red neuronal. En general, el estudio de una red neuronal puede plantearse como un problema de aproximación, asociación, clasificación, predicción o control. En la teoría de aproximación de funciones, las redes neuronales estáticas pueden verse como interpoladores nolineales que representan la mejor aproximación de una hipersuperficie sobre un dominio dado de muestras, según una norma de medida determinada. Las memorias asociativas y los clasificadores también dependen de la estructura del modelo neuronal para calcular la capacidad o el rango de representación. En el caso particular del control, es fundamental el estudio de la estabilidad no solo de la red neuronal misma, en el caso que ésta tenga dinámica, sino del sistema controlado.

El problema del aprendizaje consiste en encontrar un algoritmo que genere un conjunto óptimo de parámetros de una función neuronal. Estos parámetros pueden ser hallados a través del cálculo de una ecuación cerrada, o utilizando un algoritmo iterativo que modifica sucesivamente los valores de éste vector de parámetros hasta llegar a un punto de equilibrio. El punto de equilibrio existe siempre y cuando el algoritmo de aprendizaje sea estable y converja a un vector de parámetros en tiempo finito o en forma asintótica. En cualquier caso, se considera que el sistema ha aprendido cuando la medida de un criterio de error se halle por debajo del valor mínimo aceptado. Algunos métodos utilizados son análogos a los usados para calcular filtros adaptivos o controladores adaptables.

La implementación de las redes neuronales es el tema principal de esta tesis. Si bien las redes neuronales son interesantes desde un punto de vista teórico, y ya han producido resultados relevantes en el área de reconocimiento de patrones (Haykin, 1994) y el control (Partasarathy, 1992), el fin último de las mismas es plasmar esa enorme potencia de cómputo en implementaciones

paralelas. Las redes neuronales son particularmente atractivas cuando se aplican a problemas complejos, para los cuales las soluciones de los métodos tradicionales son insatisfactorias. Las características adaptivas de una red neuronal juegan un papel preponderante. Pero justamente, estos problemas complejos resueltos con métodos neuronales necesitan una gran potencia de cómputo. Una de las críticas importantes de la que las redes neuronales han sido objeto es, justamente, el largo tiempo que lleva simular y entrenar sus modelos en computadoras digitales que trabajan en forma secuencial. Una computadora orientada a la computación neuronal deberá poseer múltiples unidades computacionales independientes, con un funcionamiento local y alto nivel de interconectividad.

El desarrollo de nuevas técnicas en neurocomputación obliga a estudiar los sistemas neuronales tanto desde el punto de vista de la unidad elemental de cómputo como de la integración que estas unidades deben tener para conformar una estructura eficiente de computación paralela. La interrelación entre la unidad de cómputo y el sistema es escencial dado que cada unidad encierra todo lo necesario para realizar la función de procesamiento elemental. No existen módulos funcionales especializados como en las computadoras convencionales, tales como memoria, unidades aritméticas, interfases, etc. En consecuencia, en esta tesis se realiza un estudio sobre las posibilidades de la electrónica para modelar unidades y sistemas neuronales. En particular, se proponen implementaciones de redes neuronales analógicas. La primera a nivel microelectrónico, comprende el diseño de una neurona computacional con función de activación gaussiana. Primeramente, se demostrará que las unidades computacionales con función de activación gaussiana forman una base para aproximadores universales de funciones continuas y discontinuas medibles. La segunda abarca la realización de una neurocomputadora basada en un dispositivo neuronal analógico. Este sistema aprovecha toda la potencia de cómputo de un procesador neuronal capaz de procesar 300 mil patrones analógicos por segundo. El ambiente desarrollado permite que dicho procesador se convierta en el corazón de una computadora neuronal, dentro de una computadora personal tradicional. Para ello, se desarrolla el hardware de interfaz y procesamiento necesario y el software para comunicación entre

procesadores, conversión análogo-digital de patrones y de manejo del sistema computador.

1.1 ORGANIZACIÓN DE LA TESIS

El capítulo 2 en su primera parte introduce el tema general de las redes neuronales artificiales y centra su atención en un tipo particular, las redes estáticas *feedforward*. En este capítulo se analizan los temas centrales de este tipo de redes, aprendizaje, capacidad y generalización. Se analizan los teoremas fundamentales que aseguran la suficiencia de dos capas en una red neuronal como mejor aproximador de funciones continuas. Como modelos neuronales se utilizan dos tipos diferentes, los perceptrones multicapa y las redes de base radial (RBF) con función de activación gaussiana. Por último se propone el diseño de una red neuronal en un caso típico de aproximación de una curva en un mapa bidimensional.

El capítulo 3 plantea el problema de la generación de modelos neuronales en sistemas electrónicos. Estos sistemas dan lugar a elementos y sistemas de cómputo neuronales. Se detallan las características de técnicas analógicas, digitales e híbridas. Se realiza un análisis comparativo, destacando las ventajas y desventajas en cada caso. Este análisis es utilizado para tomar decisiones en virtud del desarrollo de modelos para neuronas y computadoras neuronales.

El capítulo 4 plantea el uso de dispositivos microelectrónicos en la generación de funciones representativas de modelos neuronales. Primeramente se analizan los bloques constitutivos básicos y algunos problemas y limitaciones de los transistores. A continuación se propone un modelo para una neurona RBF gaussiana y se presentan los resultados obtenidos.

En el capítulo 5 se propone una neurocomputadora a la cual se la ha denominado Netmap. Este sistema permite construir redes neuronales estáticas de una o dos capas de procesamiento. Se describe el sistema

neuronal, la arquitectura multiprocesador utilizada, y los métodos para la optimización temporal en el procesamiento de patrones y modificación de pesos durante el aprendizaje.

El capítulo 6 está dedicado a la programación de tareas dentro de la neurocomputadora. Se analiza la funcionalidad del código dentro del sistema y la comunicación con otros módulos. Se presentan, también, las simulaciones relizadas en función de solucionar el problema de la programación de los pesos de la red neuronal. Se propone un algoritmo de modificación de pesos que incluye al dispositivo neuronal como parte del lazo de aprendizaje.

En el capitulo 7 se hace una discusión sobre los resultados alcanzados y futuras realizaciones en los temas de esta tesis.