



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



(11) Número de publicación: **2 338 751**

(51) Int. Cl.:

G06N 3/063 (2006.01)

(12)

TRADUCCIÓN DE PATENTE EUROPEA

T3

(96) Número de solicitud europea: **00968107 .3**

(96) Fecha de presentación : **16.10.2000**

(97) Número de publicación de la solicitud: **1224619**

(97) Fecha de publicación de la solicitud: **24.07.2002**

(54) Título: **Componente de una red neuronal.**

(30) Prioridad: **15.10.1999 GB 9924310**

(73) Titular/es: **The University of Manchester
Oxford Road
Manchester M13 9PL, GB**

(45) Fecha de publicación de la mención BOPI:
12.05.2010

(72) Inventor/es: **Furber, Stephen Byrun**

(45) Fecha de la publicación del folleto de la patente:
12.05.2010

(74) Agente: **Carpintero López, Mario**

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

ES 2 338 751 T3

DESCRIPCIÓN

Componente de una red neuronal.

5 La presente invención se refiere a un componente de una red neuronal.

Las redes neuronales son arquitecturas de cómputo inspiradas en la biología para procesar información. Cada vez se usan más para resolver problemas que son difíciles de resolver con programas de algoritmos convencionales que se ejecutan en ordenadores convencionales con programas almacenados. Estos son típicamente problemas de asociación de patrones tales como predicciones del mercado de valores, reconocimiento facial y reconocimiento del habla. Algunas aplicaciones para redes neuronales son comercialmente significativas. Por ejemplo, el touchpad de muchos ordenadores portátiles usa una red neuronal para mejorar la fiabilidad (los sistemas neuronales son relativamente insensibles al ruido).

15 Las neuronas son generalmente dispositivos con entradas múltiples, y una única salida. La fuerza de la señal de salida de una neurona es una función de la suma ponderada de las entradas de dicha neurona y puede representarse mediante la siguiente ecuación:

$$20 \quad Y_i = f_i \left(\sum_j W_{ij} \cdot X_j - T_i \right)$$

25 En donde X_j son las entradas a la neurona (posiblemente desde otras neuronas), W_{ij} son los pesos aplicados a las entradas,

30 la suma $\sum_j W_{ij} \cdot X_j$ es el nivel de activación de la neurona (una medida interna del estado de la neurona), T_i es el umbral de la neurona, f_i es una función de activación (usualmente no lineal), e Y_i es la salida de la neurona i -ésima. Cuando el nivel de activación excede el umbral la neurona generará una salida.

35 Un peso asociado a una entrada dada puede ser positivo, en cuyo caso una señal recibida en dicha entrada provocará un aumento del nivel de activación. Un peso positivo puede por lo tanto considerarse como una entrada excitadora. En algunos casos un peso asociado con una entrada dada puede ser negativo, en cuyo caso una señal recibida en dicha entrada provocará una disminución del nivel de activación. Un peso negativo puede considerarse por lo tanto como una entrada inhibidora.

40 Las conexiones entre las neuronas se refuerzan o se debilitan mediante el ajuste de los valores de los pesos. Por ejemplo, el peso asociado con una entrada concreta de una neurona dada puede aumentarse cada vez que esa entrada recibe una señal. Una señal de entrada recurrente (es decir, una señal recibida varias veces en la misma entrada) aumentará gradualmente el peso asociado con dicha entrada. Una señal recibida en dicha entrada causará por lo tanto un aumento mayor del nivel de activación de la neurona.

45 La función de activación es normalmente la misma para todas las neuronas y es fija; a menudo se usa una función sigmoidea.

50 La actividad de la neurona i -ésima en las redes neuronales conocidas está limitada a ser una función monotónica de sus entradas X_i , determinada por los valores de los pesos W_{ij} . Esto restringe la capacidad de las redes neuronales conocidas para emular comportamientos más complejos, y no monotónicos.

55 La Patente Estadounidense n° 5.278.945 (American Neurological, Inc) describe un aparato procesador neuronal. El aparato procesador neuronal implementa una red neuronal mediante el procesamiento simultáneo de una pluralidad de neuronas usando las mismas entradas sinápticas. Una memoria RAM de pesos externa accede secuencialmente al peso de los datos. Los datos de entrada y los datos de peso pueden configurarse tanto como una única entrada o como una pluralidad de entradas procesadas simultáneamente. El aparato procesador neuronal puede usarse para implementar una red neuronal completa, o puede implementarse usando una pluralidad de dispositivos, implementando cada uno de los dispositivos un número predeterminado de capas neuronales.

60 Un objetivo de la presente invención es proporcionar un componente de una red neuronal que supere la desventaja mencionada anteriormente.

65 De acuerdo con un primer aspecto de la invención se proporciona un componente de una red neuronal, comprendiendo el componente una pluralidad de entradas, al menos un elemento procesador, al menos una salida, y una memoria digital que almacena valores en unas direcciones respectivamente correspondientes con el al menos un elemento procesador, en el cual el al menos un elemento procesador está dispuesto para recibir un valor que comprende una instrucción desde la memoria digital en respuesta a una señal de entrada desde un elemento procesador de un

ES 2 338 751 T3

componente adicional de la red neuronal, y obtiene instrucciones del valor que se recibe desde la memoria digital para ejecutar una de entre una pluralidad de operaciones.

- La invención contrasta con las redes neuronales de la técnica anterior, en las que siempre se suma un valor recibido por un elemento procesador a un nivel de activación del elemento procesador (el valor siempre es un peso). El elemento procesador de la invención es capaz de ejecutar una pluralidad de operaciones, y por lo tanto el valor recibido debe actuar como una instrucción para que el elemento procesador elija una de las instrucciones para su ejecución.

El nivel de activación de un elemento procesador es un valor interno indicativo de la excitación de ese elemento procesador.

Convenientemente, la operación es la suma del valor a un nivel de activación del al menos un elemento procesador.

Convenientemente, la operación es una atenuación de un nivel de activación del al menos un elemento procesador hasta cero, mediante el aumento del nivel de activación si es negativo o la disminución del nivel de activación si es positivo.

Convenientemente, la tasa de atenuación es dependiente del valor recibido desde la memoria digital.

Preferiblemente, la atenuación está controlada por un reloj.

Convenientemente, la operación comprende la generación de una salida por el al menos un elemento procesador, y luego la reposición de un nivel de activación del al menos un elemento procesador a un nivel mínimo predeterminado.

Convenientemente, la operación comprende reposicionar un nivel de activación del al menos un elemento procesador a un nivel mínimo predeterminado sin que el al menos un elemento procesador genere una salida.

Convenientemente, la operación comprende modificar un nivel de activación del al menos un elemento procesador de acuerdo con un algoritmo de aprendizaje automático.

Convenientemente, la operación comprende modificar el valor retenido en la memoria digital de acuerdo con un algoritmo de aprendizaje automático.

Convenientemente, el algoritmo de aprendizaje automático comprende un algoritmo de aprendizaje de Hebbian.

Convenientemente, la salida generada por el componente de una red neuronal comprende un número representativo de la dirección del elemento procesador desde el cual se emitió la salida.

Convenientemente, el componente de una red neuronal está provisto de unos medios de arbitraje y codificación dispuestos para determinar el orden cronológico de los impulsos de señal recibidos desde los elementos procesadores, y para representar cada impulso de señal como un número. Esto es ventajoso porque supera el problema de restricción de pines en la comunicación off-chip, reemplazando por ejemplo 256 pines, si cada salida neuronal tiene su propio pin, por 8 pines (donde se usa codificación de dirección binaria de 8 bits).

Convenientemente, el componente de una red neuronal comprende adicionalmente un procesador convencional con acceso de lectura/escritura a la memoria digital.

Convenientemente, los accesos del procesador convencional a la memoria digital se someten a arbitraje asíncronamente con los accesos requeridos por los elementos procesadores neuronales.

Convenientemente, un gran número de componentes de una red neuronal son supervisados por una red paralela de procesadores convencionales.

Preferiblemente, la memoria digital es una Memoria de Acceso Aleatorio (RAM).

Convenientemente, el componente funciona de manera asíncrona. Alternativamente, el componente de una red neuronal puede funcionar de manera temporizada.

De acuerdo con un segundo aspecto de la invención se proporciona un procedimiento para operar un componente de una red neuronal, comprendiendo el procedimiento almacenar valores en una memoria digital en direcciones que corresponden respectivamente al menos a un elemento procesador, transfiriendo un valor que comprende una instrucción hasta el al menos un elemento procesador en respuesta a una señal de entrada desde un elemento procesador de un componente adicional de la red neuronal y ejecutando una de entre una pluralidad de operaciones en el elemento procesador en respuesta al valor que se recibe desde la memoria digital.

El procedimiento puede incluir cualquiera de las anteriores características convenientes o preferibles del primer aspecto de la invención.

ES 2 338 751 T3

A continuación se describirá una realización específica de la invención, a modo de ejemplo únicamente, con referencia a las Figuras adjuntas, en las cuales:

5 La Figura 1 es una representación esquemática de una realización de un componente de una red neuronal de acuerdo con la invención;

La Figura 2 es una representación esquemática del componente de una red neuronal, incluyendo unos medios de arbitraje y codificación;

10 La Figura 3 es una representación esquemática de los medios de arbitraje y codificación mostrados en la Figura 2;

La Figura 4 es un gráfico ilustrando un mecanismo de aprendizaje de Hebbian; y

15 Las Figuras 5a a 5c son diagramas de flujo que ilustran el funcionamiento del componente de una red neuronal.

Un componente de una red neuronal tal como el que se muestra en la Figura 1 comprende 1024 entradas 1, una Memoria de Acceso Aleatorio (RAM) 2, 256 elementos procesadores 3 y 256 salidas 4.

20 Cada entrada porta una señal desde un elemento procesador de un componente previo de la red neuronal. Cada entrada porta sólo 1 bit de información, en otras palabras está “conectada” o “desconectada”. Las señales de entrada adoptan la forma de impulsos en las entradas.

25 La RAM 2 contiene un conjunto de valores. Cada combinación de entrada y salida tiene adjudicada una dirección diferente en la RAM (el número de direcciones es 1024x256). Refiriéndose a la Figura 1, para ayudar al lector y sólo con propósitos ilustrativos, puede considerarse la RAM como una matriz bidimensional que comprende columnas y filas, conteniendo cada fila todos los valores pertenecientes a una entrada concreta, y conteniendo cada columna todos los valores pertenecientes a un elemento procesador concreto. Tal como se discute más adelante, una proporción significativa de las direcciones RAM puede contener valores nulos.

30 En respuesta a un impulso recibido en una entrada dada, la RAM 2 busca valores localizados en la fila correspondiente a dicha entrada. Cada valor se envía al elemento procesador 3 correspondiente a la columna en la que el valor está localizado. El nivel de activación del elemento procesador 3 (una medida interna del estado del elemento procesador) se modifica de acuerdo con el valor que se recibe. Cada elemento procesador 3 actúa como una neurona, y tiene 1024 entradas y 1 salida.

35 En contraste con la técnica anterior, los valores almacenados en la RAM 2 pueden ser pesos o pueden ser instrucciones. Cuando el valor enviado a un elemento procesador 3 es una instrucción, el elemento procesador 3 lleva a cabo una operación de acuerdo con la instrucción, tal como se describe más adelante. Cuando el valor enviado a un elemento procesador 3 es un peso, dicho peso se suma al nivel de activación del elemento procesador 3. La operación de un elemento procesador “i” puede expresarse matemáticamente como:

$$45 \quad Y_i = f_i \left(\sum_j W_{ij} \cdot X_j - T_i \right)$$

50 En donde X_i son entradas a la neurona, W_{ij} son los pesos, la suma $\sum_j W_{ij} \cdot X_j$ es el nivel de activación de la neurona, T_i es el umbral del elemento procesador, f_i es una función de activación (usualmente no lineal), e Y_i es la salida del elemento procesador i-ésimo. El elemento procesador emite un impulso cuando el nivel de activación excede el umbral (este es de hecho la activación de la neurona).

55 En el presente ejemplo, j , que designa una entrada específica puede estar entre 1 y 1024, e i , que designa una salida específica puede estar entre 1 y 256.

60 Muchas de las redes neuronales de la técnica anterior comprenden neuronas que generan valores de salida analógicos. Por ejemplo, refiriéndose a la ecuación anterior, la salida Y_i de una neurona sería un número que tuviera un valor entre 0 y 1. Esta salida, al entrar en una subsiguiente neurona, sería multiplicada por un peso, y el resultado de la multiplicación sería añadido al nivel de activación de dicha subsiguiente neurona.

65 La realización descrita de la invención opera de una forma diferente a las redes neuronales analógicas, en el sentido de que un elemento procesador (neurona) tiene sólo dos posibles salidas, es decir 1 ó 0 (o de manera equivalente “conectada” o “desconectada”). Un impulso de salida es generado por un elemento procesador siempre que el nivel de activación de dicho elemento procesador excede su umbral. Desde el punto de vista de un observador externo,

ES 2 338 751 T3

una salida de impulsos de un elemento procesador puede considerarse como el equivalente de una salida analógica. Por ejemplo, si un elemento procesador generara 23 impulsos por segundo, entonces esto podría considerarse como el equivalente de un valor analógico nociional de 0,23. A partir de ahora la generación de impulsos de los elementos procesador se denominará “codificación de tasa de impulsos”.

5 A primera vista puede parecer que la codificación de tasa de impulsos es muy intensa a nivel computacional cuando se compara con el uso de señales analógicas. Por ejemplo, un peso adjudicado a una entrada dada de una neurona puede ser multiplicado por la señal analógica 0,23 en un solo cálculo, y luego añadido al nivel de activación de la neurona. Por el contrario, cuando se usa la codificación de tasa de impulsos el peso debe sumarse al nivel de
10 activación de la neurona 23 veces (es decir una vez por cada impulso), precisando 23 cálculos separados. Se hace observar sin embargo que a menudo no hay salida desde una neurona dada. Cuando se usa la codificación de tasa de impulsos, en esta eventualidad no se efectúa ningún cálculo ya que no se reciben impulsos de entrada. Por el contrario, en las redes neuronales conocidas, un valor cero de salida es efectivamente un valor de salida que resulta ser cero. Los pesos de las neuronas se multiplican por el valor de salida cero, y el resultado (también cero) se suma a los niveles de
15 activación de las neuronas. Este cálculo, que no tiene efecto sobre los niveles de activación de las neuronas, se produce con mucha frecuencia y es un gran despilfarro a nivel computacional. El uso de la codificación de tasa de impulsos evita el despilfarro de esta multiplicación por cero.

20 El componente de una red neuronal está provisto de 256 elementos procesadores 3. Cada elemento procesador 3 actúa como una neurona, y tiene 1024 entradas y 1 salida. Si hubiera que proporcionar conexiones individuales entre cada elemento procesador y cada componente subsiguiente de una red neuronal, el número de conexiones se volvería rápidamente tan grande que sería impracticable. Para evitar este problema, tal como se muestra en la Figura 2, las salidas de los elementos procesadores están todas dirigidas hacia una única línea 5 de salida de 8 bits. Un impulso de salida desde un elemento procesador dado es convertido en un número binario de 8 bits representativo de dicho
25 elemento procesador, el cual es portado por la línea de salida. Esta conversión, que puede denominarse “codificación direccional de sucesos”, es llevada a cabo por una unidad 6 de arbitraje y codificación.

30 El funcionamiento de la unidad 6 de arbitraje y codificación está ilustrado esquemáticamente en la Figura 3. La unidad 6 tiene dos funciones: determinar el orden cronológico en el que van a enviarse los impulsos hasta la línea 5 de salida, y convertir cada salida en un número indicativo del elemento procesador que generó dicha salida. La unidad 6 consigue ambas funciones simultáneamente. La unidad 6 que se muestra en la Figura 3 es para un conjunto de ocho elementos procesadores.

35 Una primera fila de la unidad 6 comprende cuatro bloques 10-13 de arbitraje y codificación. Un lado de entrada de un primer bloque 10 de arbitraje y codificación está provisto de dos entradas a, b y dos salidas a_{ack} , b_{ack} de confirmación. Cuando la salida de un elemento procesador se pone en alta, ésta se recibe en la entrada a, una primera salida del bloque 10 se pone en alta, y una segunda salida z emite simultáneamente un número de 1 bit representativo de la entrada a (en este caso el número binario es “1”). Entonces la salida de confirmación a_{ack} se pone en alta, indicando por lo tanto al elemento procesador que el bloque 10 ha procesado la salida. Entonces la salida del elemento procesador se pone en baja.
40

45 Si la entrada a y la entrada b se ponen en alta aproximadamente al mismo tiempo, entonces un elemento Seitz de exclusión mutua determina qué entrada ha sido la primera en ponerse en alta (los elementos de exclusión mutua se describen en las páginas 260 y subsiguientes de la “Introducción a los Sistemas VLSI” de Mead y Conway, publicado por Addison Wesley). Por consiguiente se genera la salida desde el bloque 10. Una vez que se ha generado la salida, y se ha devuelto una confirmación al elemento procesador apropiado, el bloque 10 procesa y emite la última de las entradas.

50 Una segunda fila de la unidad 6 comprende una pareja de bloques 14, 15 de arbitraje y codificación. Refiriéndose a un primer bloque 14 de la pareja, dos entradas y, w están conectadas a las salidas de los bloques 10, 11 precedentes. Una primera salida u del bloque 14 se pone en alta en respuesta a una señal de entrada, y una segunda salida v emite simultáneamente un número de 2 bits. El bit más significativo del número de 2 bits representa la entrada y, y el bit menos significativo del número de dos bits representa la entrada a del bloque 10 precedente. Por lo tanto, el número de 2 bits es el número binario “11”.
55

La tercera y última fila de la unidad 6 funciona de la misma manera que las filas anteriores. La salida del bloque 16 individual que comprende la última fila es un número binario de 3 bits, y en este caso el número es “111”.

60 En la descripción de la Figura 1, se dice que la salida del elemento procesador 3 tiene la forma de un impulso. Se observará que en la descripción de la Figura 3 se dice que la salida del elemento procesador está en alta de manera continua y sólo termina una vez que se ha recibido una señal de confirmación desde el elemento procesador. Esta modificación es necesaria para permitir que la unidad 6 de arbitraje y codificación funcione correctamente.

65 Refiriéndose a la Figura 2, se precisan 8 filas de bloques de arbitraje y codificación para generar una salida con un número binario de 8 bits.

Refiriéndose a la Figura 2, hay cuatro líneas 19 de entrada de 8 bits conectadas a los elementos procesadores 3 y a la RAM 2. Se usa una unidad 20 de arbitraje y codificación para arbitrar entre las cuatro líneas 19 de entrada. La

ES 2 338 751 T3

unidad 20 de arbitraje y codificación funciona de la manera representada esquemáticamente en la Figura 2 y descrita anteriormente. En este caso la unidad de arbitraje y codificación incluye dos filas de bloques. La unidad de arbitraje y codificación tiene una salida de 10 bits, siendo usados los dos bits extra para distinguir entre las cuatro líneas 19 de entrada de 8 bits.

5 La salida de 10 bits de la unidad 20 de arbitraje y codificación está conectada a un decodificador 21. El decodificador convierte cada número de entrada de 10 bits en una salida por una correspondiente línea 1 de salida de 1 bit (hay 1024 líneas de salida). Las líneas 1 de salida corresponden a las líneas 1 de entrada representadas en la Figura 1.

10 En el sistema de codificación de tasa de impulsos usado en la realización de la invención, la comunicación entre los componentes de la red neuronal se produce en una escala de tiempo del orden de $1 \mu\text{s}$. Esto parece instantáneo en comparación con la resolución temporal de las neuronas que se activan a una tasa que no supera un orden de 1 kHz.

15 El hecho de que la red neuronal sea digital proporciona un grado de flexibilidad nunca visto en componentes de redes neuronales analógicas. En concreto, puede almacenarse un amplio rango de valores en la RAM 2, y esto permite usar algunos de los valores como “instrucciones” para el elemento procesador 3. El término “instrucción” pretende significar que se instruye al elemento procesador para que lleve a cabo una operación diferente a la de simplemente sumar el valor del nivel de activación. En las redes neuronales conocidas de la técnica anterior, la acción llevada a cabo por una neurona es siempre la misma, es decir sumar un valor al nivel de activación. El uso de instrucciones permite a 20 los elementos procesadores 3 tener mucha más flexibilidad en su funcionamiento.

25 El uso de instrucciones proporciona una red neuronal que tiene propiedades más cercanas a las neuronas biológicas reales. Una de tales propiedades es una integración “con fugas”, en otras palabras un nivel de activación que desciende si no hay actividad de entrada durante un periodo de tiempo. Otra propiedad es la “refracción”, que quiere decir que 30 un elemento procesador 3 ignora todas las entradas durante un cierto periodo de tiempo posterior a la activación. Esto limita la tasa de activación máxima del elemento procesador (es decir la tasa a la que el elemento procesador 3 puede producir impulsos de salida), evitando por lo tanto que la red neuronal se vuelva inestable debido a una actividad neuronal siempre creciente. Tanto la integración con refracción como la integración con fugas implican cierta dependencia del tiempo real.

35 En la realización descrita de la invención se proporciona la integración con refracción y con fugas usando un suceso de referencia en tiempo real, por ejemplo un reloj de 32 kHz (no representado en las Figuras 1 ó 2). Los impulsos del reloj pueden transmitirse como un número binario concreto a través de la entrada de 8 bits (por ejemplo 00000001). Cuando se precisa una integración con fugas, el valor almacenado en cada dirección de RAM correspondiente a dicho 40 número de entrada dará instrucciones al elemento procesador para disminuir su nivel de activación si el nivel de activación es positivo y para aumentar el nivel de activación si es negativo. Distintos valores contenidos en la RAM pueden dar instrucciones para diferentes tasas de integración con fugas, por ejemplo aumentar/disminuir en respuesta a cada impulso del reloj, o aumentar/disminuir en respuesta a cada segundo impulso del reloj.

45 Puede proporcionarse una instrucción que determine si deberá ajustarse el peso o no mediante algoritmos de aprendizaje (por ejemplo algoritmos de aprendizaje de Hebbian). A modo de ejemplo de una neurona con capacidades de aprendizaje, puede modelarse el siguiente comportamiento neuronal, tal como se muestra en la Figura 4.

50 El nivel de activación del elemento procesador opera dentro del rango $\{-L, +T\}$, en donde $-L$ es el nivel refractario y T es el umbral que, al ser alcanzado, hace que el elemento procesador emita un impulso de salida y el nivel de activación sea repuestado a $-L$. $\{-L, 0\}$ es el rango refractario del elemento procesador: si se transmite un peso al elemento procesador en respuesta a un impulso de entrada, no se suma el peso al nivel de activación (el nivel de activación no se cambia), y se disminuye el propio peso. $\{0, F\}$ es el rango débilmente excitado: si se transmite un peso hasta el elemento procesador en respuesta a un impulso de entrada, se sumará al nivel de activación y el mecanismo de aprendizaje no aplicará cambios al propio peso. $\{F, T\}$ es el rango fuertemente excitado: si se transmite un peso hasta el elemento procesador en respuesta a un impulso de entrada, se sumará al nivel de activación y el propio peso aumentará.

55 El funcionamiento de un elemento procesador, dispuesto para funcionar de acuerdo con el modelo representado en la Figura 3, se muestra de forma esquemática mediante un diagrama de flujo en la Figura 5a-c.

Refiriéndose primero a la Figura 5a, se recibe un impulso en una entrada j dada. El valor asociado a dicha entrada j para un elemento procesador dado se recupera desde la RAM y se transmite hasta el elemento procesador. El valor estará dentro de uno de entre tres rangos: un primer rango corresponde a los pesos que han de sumarse al nivel de activación, un segundo nivel corresponde a los pesos que han de sumarse al nivel de activación y además han de dar instrucciones al elemento procesador para que el peso asociado a dicha entrada sea aumentado, y un tercer rango comprende instrucciones distintas de los pesos.

60 El resto de la Figura 5 a trata sobre un valor que está dentro del primer rango. El valor se suma al nivel de activación del elemento procesador (si el peso es negativo el nivel de activación se reducirá). Si el nivel de activación que sigue a la suma del valor es menor que el umbral T entonces no se toma ninguna medida adicional. Si el nivel de activación es mayor que el umbral T , entonces el elemento procesador emite un impulso de salida y el nivel de activación del elemento procesador se reduce a $-L$.

ES 2 338 751 T3

Si el valor entra dentro del segundo rango entonces el funcionamiento del elemento procesador es tal como se muestra en las ramificaciones del diagrama de flujo de la Figura 5b. Si el nivel de activación del elemento procesador es mayor o igual que cero entonces el valor se suma al nivel de activación del elemento procesador. A continuación, si el nivel de activación es mayor que cero entonces el valor guardado en la RAM aumenta.

5

Si el nivel de activación del elemento procesador es menor que cero entonces el valor no se suma al nivel de activación, y el valor guardado en la RAM disminuye.

Si el valor entra dentro del tercer rango, ello indica alguna otra forma de instrucción. Por ejemplo, tal como se muestra en la Figura 5c, el elemento procesador puede ser instruido para emitir un impulso independientemente del nivel de activación, y reposicionar el nivel de activación al nivel refractario.

Este algoritmo de aprendizaje de Hebbian, que resulta en una red neuronal estable si se eligen unos valores apropiados, depende de un mecanismo para retornar el nivel de activación desde $-L$ hasta 0 que es independiente del procesamiento numérico de los pesos, y por lo tanto depende del uso de una instrucción de “incremento a cero con el tiempo” que puede implementarse tal como se describe a continuación.

Los expertos en la técnica comprenderán que los elementos procesadores pueden ser capaces de ejecutar otras instrucciones, que pueden estar definidas por valores almacenados en la memoria de valores RAM. Por ejemplo, el nivel de activación puede ser repuesto al nivel refractario sin activar la neurona. Esto modela una entrada fuertemente inhibitoria.

Refiriéndose otra vez a la Figura 1, el componente de una red neuronal puede incluir un procesador convencional (no representado) con acceso de lectura/escritura a la memoria RAM 4 de valores. Adicionalmente, el procesador puede tener la capacidad de monitorizar y generar entradas y salidas a los flujos de actividad neuronal, usando dispositivos periféricos con una interfaz adecuada incorporados al componente de una red neuronal.

El procesador accede a la memoria RAM 4 de valores y los flujos de actividad neuronal pueden arbitrarse asincrónicamente hacia la red completamente operativa con una mínima interferencia con el proceso neuronal siempre y cuando la red no esté operando al borde de la saturación en ese momento. Un procesador de monitorización podría supervisar un número de componentes de una red neuronal, pero un sistema muy grande podría precisar de varios de tales procesadores. Una implementación basada en tecnología DRAM de 1 Gbit podría tener un procesador de monitorización en cada chip. Por lo tanto, la red neuronal puede implementarse como una cantidad muy elevada de componentes de una red neuronal supervisados por una red paralela de procesadores convencionales. Un mérito de la invención es la facilidad con la que la arquitectura conecta con máquinas convencionales. El procesador de monitorización extiende ampliamente la flexibilidad de la arquitectura y proporciona un estrecho acoplamiento entre el procesamiento neuronal y una máquina secuencial convencional.

Si la comunicación se implementa usando una lógica temporizada, la información de activación temporal estará cuantizada. Si se usa una lógica asíncrona, se evitará dicha cuantización. El uso de la lógica asíncrona es por lo tanto ventajoso.

Aunque en la realización ilustrada un elemento procesador corresponde a una neurona, se observará que un elemento procesador puede ser responsable de llevar a cabo los cálculos para más de una neurona.

45

50

55

60

65

REIVINDICACIONES

- 5 1. Un componente de una red neuronal, comprendiendo el componente una pluralidad de entradas (1), al menos un elemento procesador (3), al menos una salida (4), y una memoria digital (2) que almacena valores en direcciones respectivamente correspondientes con el al menos un elemento procesador (3), en el cual el al menos un elemento procesador (3) está dispuesto para recibir un valor que comprende una instrucción desde la memoria digital (2) en respuesta a una señal de entrada desde un elemento procesador (3) de un componente adicional de la red neuronal, y obtiene instrucciones del valor que se recibe desde la memoria digital (2) para ejecutar una de entre una pluralidad de operaciones.
- 10 2. Un componente de una red neuronal de acuerdo con la reivindicación 1, en el cual la operación es la suma del valor a un nivel de activación del al menos un elemento procesador (3).
- 15 3. Un componente de una red neuronal de acuerdo con la reivindicación 1, en el cual la operación es una atenuación de un nivel de activación del al menos un elemento procesador (3) hasta cero, mediante el aumento del nivel de activación si es negativo o la disminución del nivel de activación si es positivo.
- 20 4. Un componente de una red neuronal de acuerdo con la reivindicación 3, en el cual la tasa de atenuación es dependiente del valor recibido desde la memoria digital (2).
- 25 5. Un componente de una red neuronal de acuerdo con la reivindicación 3 ó 4, en el cual la atenuación está controlada por un reloj.
- 30 6. Un componente de una red neuronal de acuerdo con la reivindicación 1, en el cual la operación comprende la generación de una salida por el al menos un elemento procesador (3), y luego la reposición de un nivel de activación del al menos un elemento procesador (3) a un nivel mínimo predeterminado.
- 35 7. Un componente de una red neuronal de acuerdo con la reivindicación 1, en el cual la operación comprende reposicionar un nivel de activación del al menos un elemento procesador (3) a un nivel mínimo predeterminado sin que el al menos un elemento procesador (3) genere una salida.
- 40 8. Un componente de una red neuronal de acuerdo con la reivindicación 1, en el cual la operación comprende modificar un nivel de activación del al menos un elemento procesador (3) de acuerdo con un algoritmo de aprendizaje automático.
- 45 9. Un componente de una red neuronal de acuerdo con la reivindicación 1 o la Reivindicación 8, en el cual la operación comprende modificar el valor retenido en la memoria digital (2) de acuerdo con un algoritmo de aprendizaje automático.
- 50 10. Un componente de una red neuronal de acuerdo con la reivindicación 8 o la Reivindicación 9, en el cual el logaritmo de aprendizaje automático comprende un logaritmo de aprendizaje de Hebbian.
- 55 11. Un componente de una red neuronal de acuerdo con cualquier reivindicación precedente, en el cual la salida (4) generada por el componente de una red neuronal comprende un número representativo de la dirección del elemento procesador (3) desde el cual se emitió la salida (4).
- 60 12. Un componente de una red neuronal de acuerdo con la reivindicación 11, en el cual el componente de una red neuronal está provisto de unos medios (6) de arbitraje y codificación dispuestos para determinar el orden cronológico de los impulsos de señal recibidos desde los elementos procesadores, y para representar cada impulso de señal como un número.
- 65 13. Un componente de una red neuronal de acuerdo con cualquier reivindicación precedente, comprendiendo adicionalmente un procesador convencional con acceso de lectura/escritura a la memoria digital (2).
- 55 14. Un componente de una red neuronal de acuerdo con la reivindicación 13, en el cual los accesos del procesador convencional a la memoria digital (2) son sometidos a un arbitraje asíncrono con los accesos requeridos por los elementos procesadores (3) neuronales.
- 60 15. Un componente de una red neuronal de acuerdo con las reivindicaciones 13 ó 14, en el cual un gran número de componentes de una red neuronal son supervisados por una red paralela de procesadores convencionales.
- 65 16. Un componente de una red neuronal de acuerdo con cualquier reivindicación anterior, en el cual la memoria digital (2) es una Memoria de Acceso Aleatorio (RAM) (2).
- 65 17. Un componente de una red neuronal de acuerdo con cualquier reivindicación anterior, en el cual el componente trabaja de manera asíncrona.

ES 2 338 751 T3

18. Un procedimiento para operar un componente de una red neuronal, comprendiendo el procedimiento almacenar valores en una memoria digital (2) en direcciones respectivamente correspondientes a al menos un elemento procesador (3), transfiriendo un valor que comprende una instrucción hasta el al menos un elemento procesador (3) en respuesta a una señal de entrada (1) desde un elemento procesador (3) de un componente adicional de la red neuronal y ejecutando una de entre una pluralidad de operaciones en el elemento procesador (3) en respuesta al valor que se recibe desde la memoria digital (2).
- 5

10

15

20

25

30

35

40

45

50

55

60

65

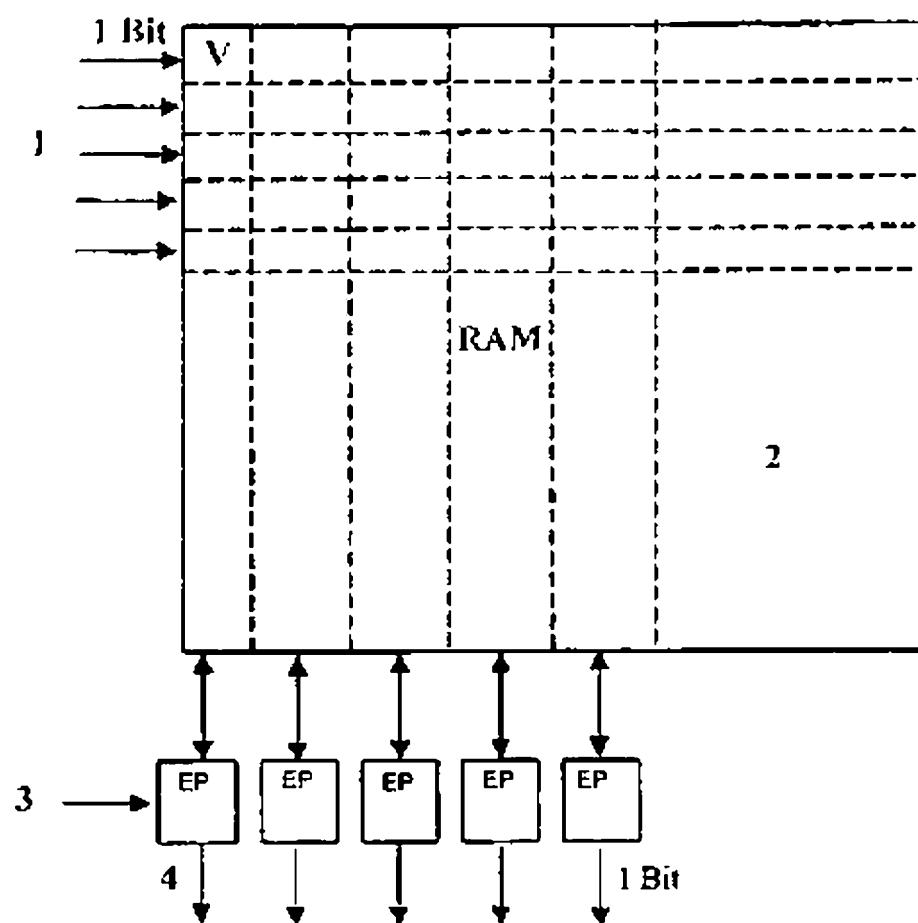
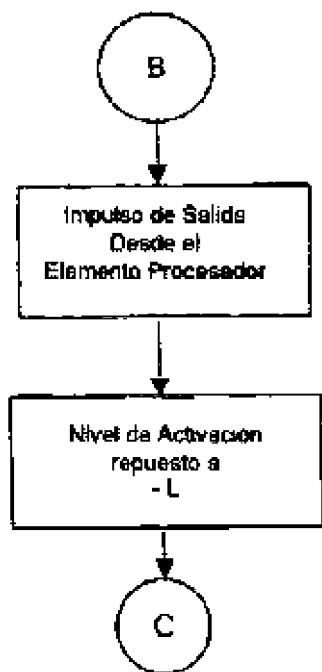
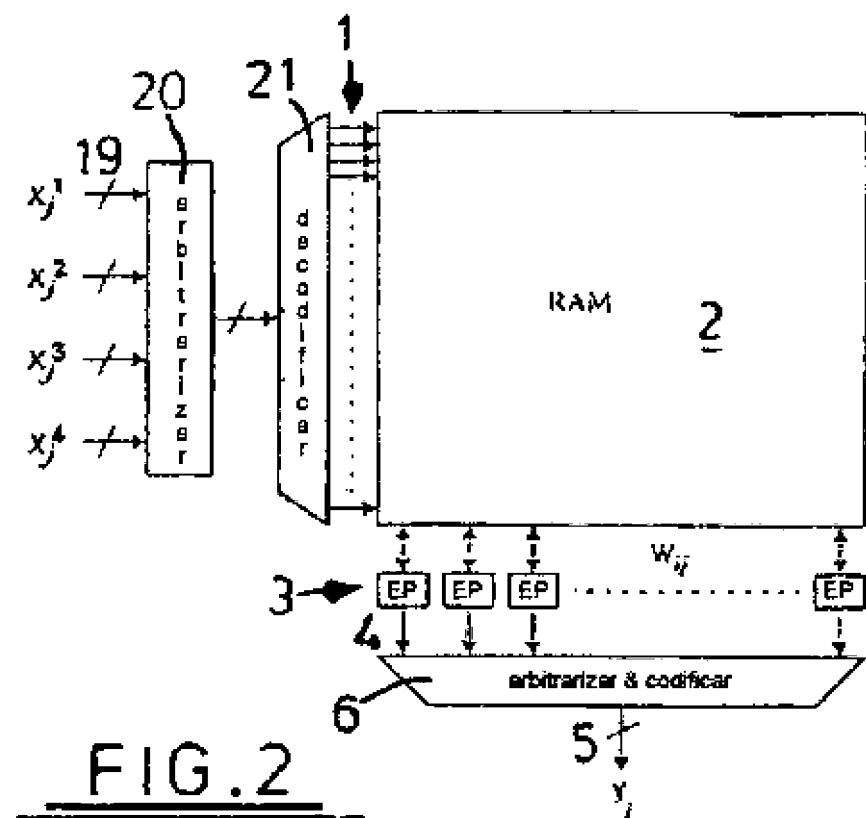


FIG. 1



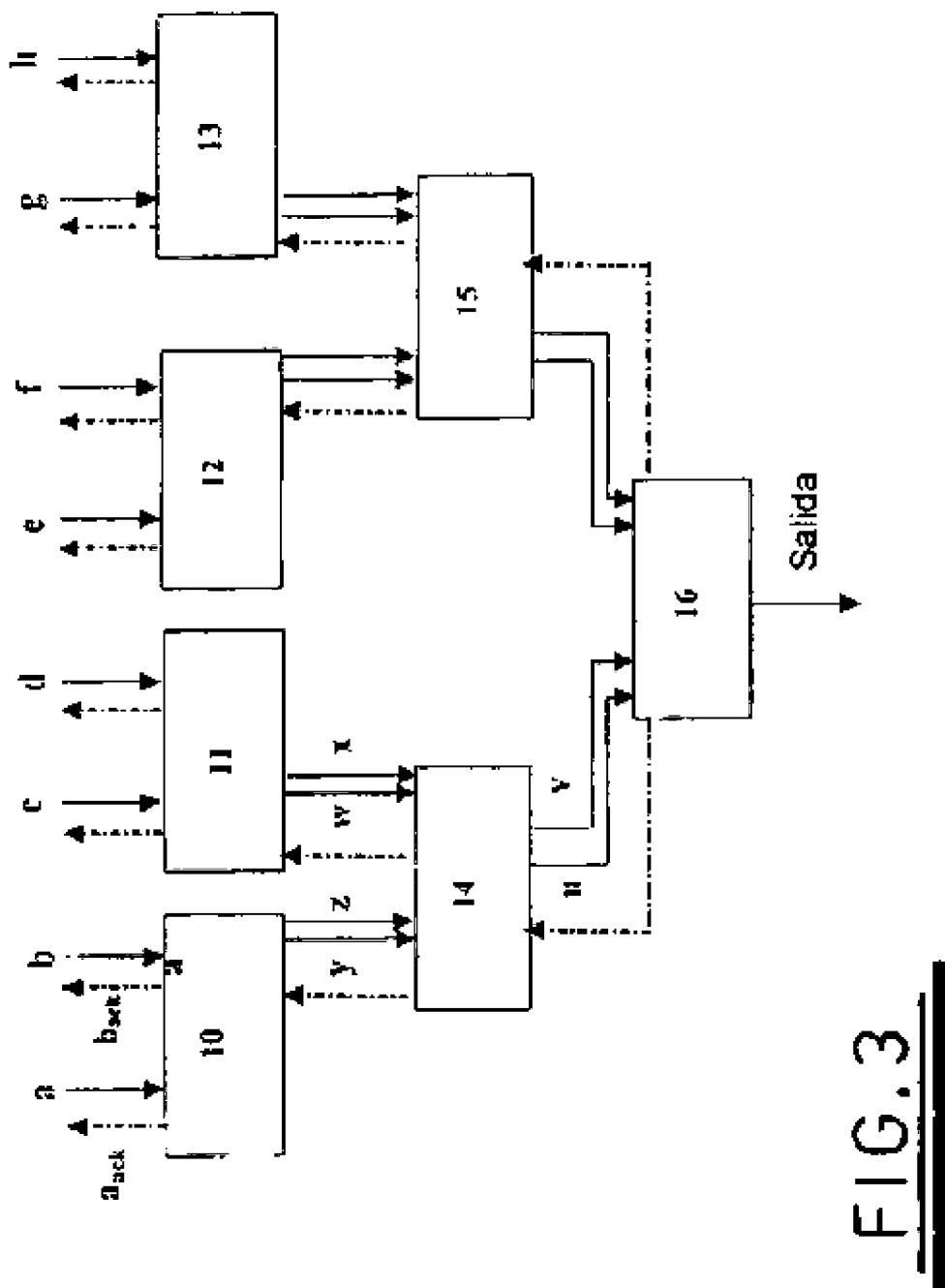


FIG. 3

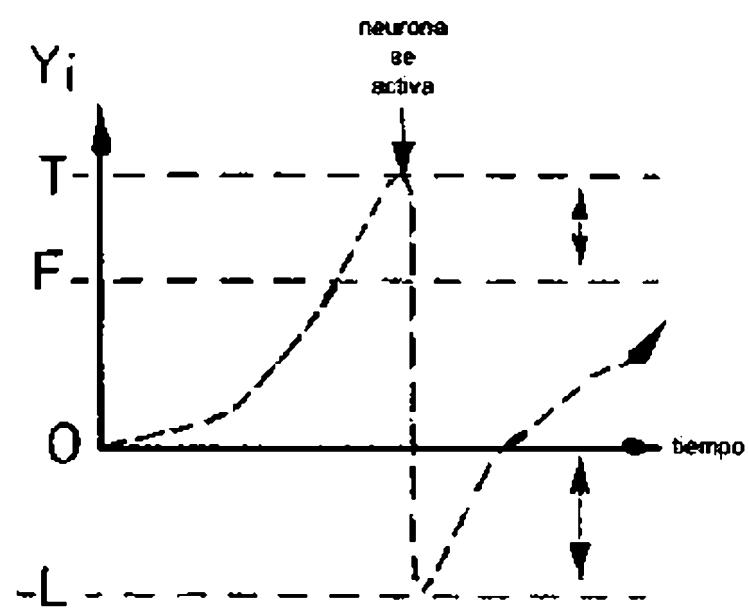


FIG. 4

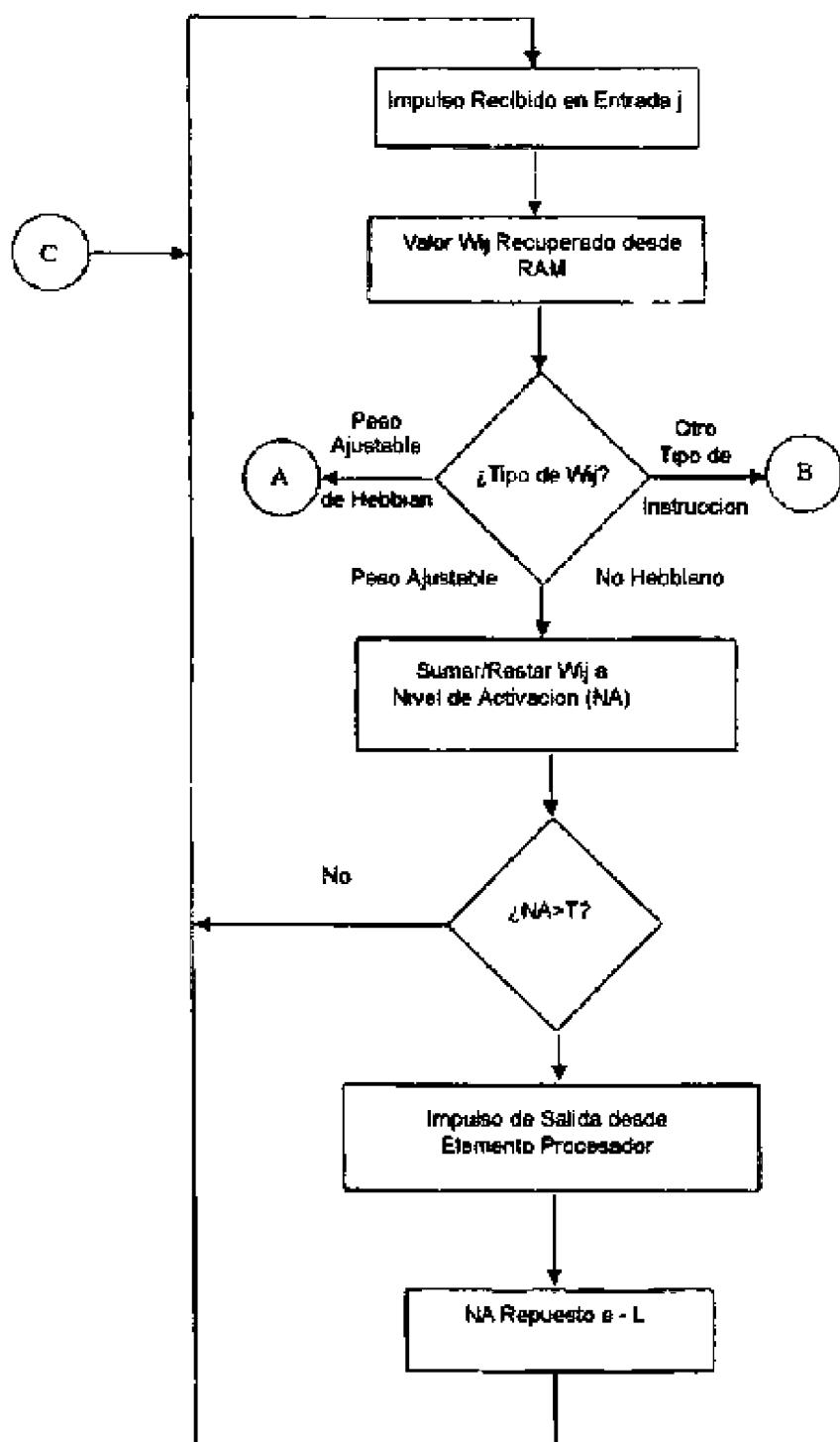


FIG. 5a

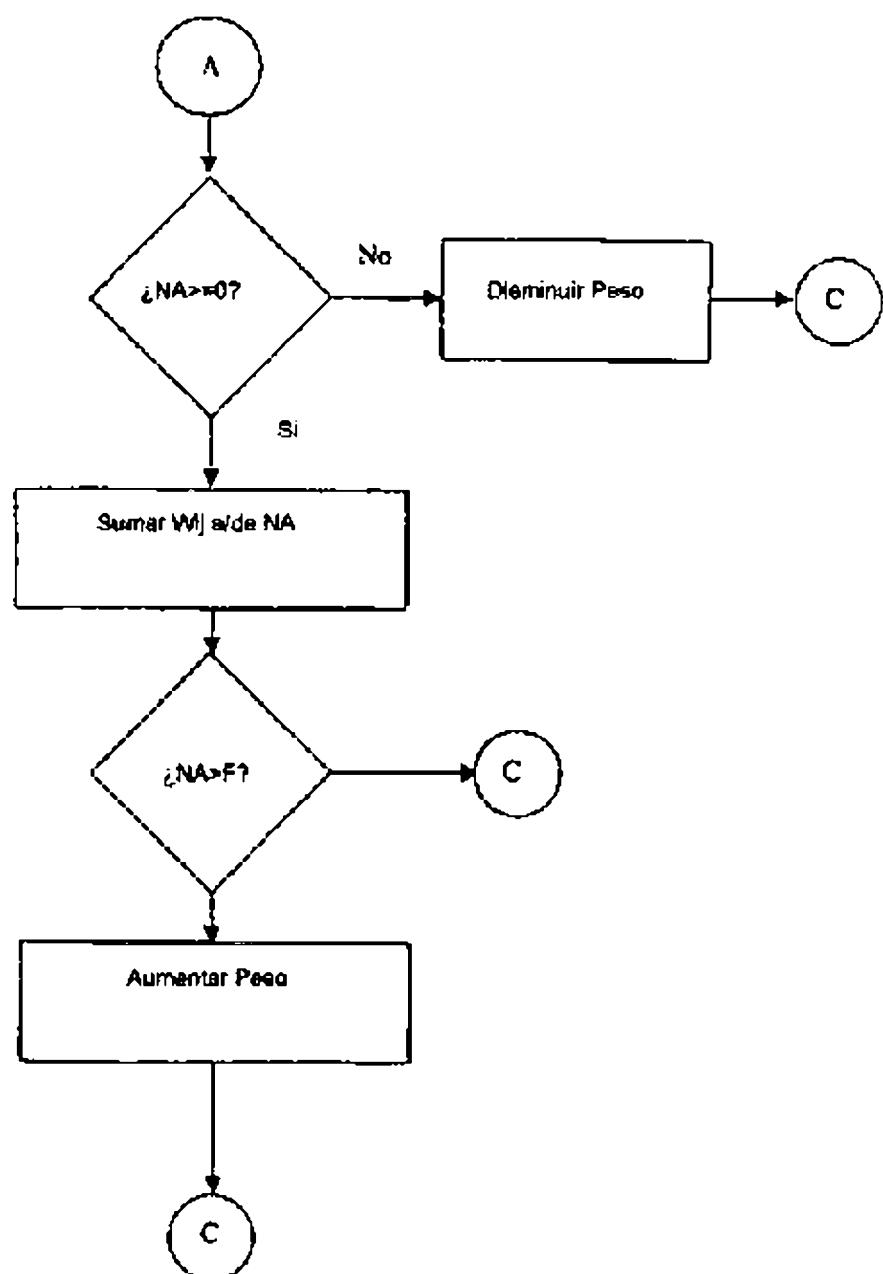


FIG.5b