TEORÍA DE RESONANCIA ADAPTATIVA

INTRODUCCION

La mayoría de las redes tienden a olvidar la información anteriormente aprendida si se intenta añadir nueva información de manera incremental.

Este funcionamiento es adecuado siempre y cuando el dominio del problema para el cual la red se emplea tenga límites bien definidos y sea estable, sin embargo, en muchas situaciones reales, el entorno no está acotado y no es estable.

El objetivo de las redes ART es poder aprender nuevas cosas sin olvidar cosas que se hayan aprendido en el pasado.

A este problema que deben resolver estas redes se le llama **dilema de estabilidad y plasticidad**, es decir, ¿cómo puede un sistema (capaz de aprender) seguir siendo adaptable (plástico) en respuesta a entradas significativas y permanecer estable en respuesta a entradas irrelevantes?, ¿cómo alternar entre los modos plástico y estable?

Una clave para resolver el dilema de estabilidad y plasticidad consiste en añadir un mecanismo de realimentación entre la capa competitiva y la capa de entrada de la red, lo que facilita el aprendizaje de información nueva sin destruir la información anterior.

Para poder entender estos conceptos, primero debemos estudiar cómo es la capa de entrada de la red y a qué se le llama capa competitiva.

Capa de entrada.

Existen n PE en la capa de entrada Vj; donde j = 1,2,....,n.

Cada valor del vector de entrada, Ii, se une a cada nodo Vj, de forma que:

Cuando j=i, la conexión es excitadora

Cuando j≠i, la conexión es inhibidora

A esto se le conoce como **distribución de centro activo, periferia inactiva**.

Como los valores de entrada pueden tomar cualquier valor, el cálculo en la red se ve en conflicto cuando estos valores son muy grandes, pues puede ocurrir un sobreflujo, a este problema se le llama dilema de saturación por ruido. Cuando se realiza la simulación en computadora, los softwares de simulación realizan un preprocesamiento a los valores de entrada para evitar este problema, realizando una normalización o escalamiento. La función de esta capa de entrada es precisamente sustituír ese preprocesamiento implementado por el software de simulación y por lo tanto, normalizar los datos de entrada, haciendo lo siguiente:

Se define un vector o **trama de reflectancias** = θ1, θ2,...θn; donde

$$\theta j = Ij \;/\; (\sum Ii)$$

Por tanto, $\sum \theta j = 1$.

Esta trama es independiente de la magnitud de las entradas y esta trama está normalizada.

Se sabe que lo que se almacena en nuestra memoria y los que se recuerda son tramas de reflectancia y no vectores de entrada, así, nosotros podemos reconocer cosas o personas aunque las condiciones sean distintas, por ejemplo, en un lugar oscuro y en un lugar muy iluminado. En ambos casos, los vectores de entrada son muy distintos entre sí, sin embargo representan lo mismo. Como las tramas de reflectancia no dependen de la intensidad de los elementos de entrada sino de la relación entre ellos, la trama de reflectancia, para ambos casos, resulta igual, por lo tanto, ésta es la que debe almacenarse y recordarse.

Por tanto, la salida de las unidades de la capa de entrada está en función de la trama de reflectancia y no de la entrada original, puesto que a cada PE de la capa de entrada le llega una conexión excitadora y varias inhibidoras, lo cual produce el efecto de evitar que la actividad en el Vi aumente en proporción a Ii.

De la capa de entrada se tiene como respuesta una trama de activación X, sin embargo, la salida real de la capa no es esta tramade activación sino que es una trama de salida S que está en función de la trama de activación, por lo general, $Sj = 1$, si $Xj >= 0$ y viceversa.

Instar.

La capa oculta consta de instars. Un instar es un PE, en el cual se ignora la flecha de salida, obviando su existencia.

La entrada neta en un instar es neta = I W

Y la salida es:

$$Y' = AY + B \text{ neta}$$
$$Y(t) = B/A \text{ neta } (1-e^{-At}) \quad ;t<t0$$
$$Y(t) = Yeq\, e^{-At} \quad ; t>t0$$

Y el valor de equilibrio de Y es:

$$Yeq = B/A \text{ neta.}$$

Lo importante quí es que la salida de equilibrio será mayor cuando neta sea mayor, y para obtener el máximo valor de neta se requiere entonces que IW sea máximo, lo cual sucede cuandoI = W.

Por tanto, para que el instar reponda lo más posible a una entrada lo que se necesita es que le vector de pesos llegue a ser idéntico que I.

Pero los cambios en los pesos se producen más lentamente que los cambios en los demás parámetros.

Para lograr que lo pesos igualen a la entrada se propone que los cambios en los pesos estén dados por:

$$W' = -CW + D\, I\, Y$$

Pero, en ausencia de I, el vector de pesos irá decreciendo hasta llegar a 0, fenómeno que no se desea, pues se considera como un olvido del PE. Para evitar esto, se modifica la forma de aprendizaje de forma que los pesos solo sean cambiados cuando exista un vector I a la entrada (neta>0), de otra forma (neta<0) que los pesos permanezcan igual.

$$W' = (-CW + DI)\, U(\text{neta})$$

Donde:

$$U(\text{neta}) = 1, \text{ si neta}>0$$
$$\qquad 0, \text{ si neta} <0$$

$$W(t+1) = W(t) + \alpha (I – W(t))$$

Donde $\alpha<1$.

Si se tiene un conjunto de vectores relativamente próximos entre sí, a lo que se le llama cúmulo, que podrían representar alguna clase, lo que se desea es que el instar aprenda algún tipo de vector representativo de la clase, como la media.

La forma de aprender es la siguiente:

1. Se selecciona un vector de entrada Ii aleatoriamente, según la distribución de probabilidad del cúmulo.
2. Se calcula $\alpha (I - W)$ y se actualizan los pesos con $W(t+1) = W(t) + \alpha (I – W(t))$.
3. Se repiten lo pasos 1. y 2. Un número de veces igual al número de vectores en el cúmulo.
4. Se repite varias veces el paso 3.

Para saber cuando detener el paso 4 se utiliza como criterio el error medio o se puede ir reduciendo el valor de α conforme avenza el entrenamiento.

Red competitiva.

Varios instars agrupados en una capa conforman una red competitiva.

Cada instar responde de forma máxima a un cierto grupo de vectores (cúmulo) de una región distinta del espacio.

Se puede decir que esta capa clasifica cualquier vector de entrada, porque la instar con mayor respuesta para alguna entrada dada es la que identifica la región del espacio en la cual yace el vector de entrada.

En vez de examinar la respuesta de cada instar para determinar cuál es la mayor, es más fácil si sólo la instar con mayor respuesta tiene una salida no nula y todas las demás tienen una salida nula. Es decir, que las instars compitan entre ellas por el privilegio de activación.

Se requiere la comunicación entre todas las instars para el proceso de decisión de activación.

Lo que se utiliza es el sistema de centro activo, periferia inactiva. La unidad cuyo vector de pesos se asemeja más al vector de entrada y por lo tanto tiene la mayor respuesta es la que manda las señales inhibitorias más fuertes y es la que recibe la mayor realimentación (procedente de ella misma).

Cuando la función de salida es lineal:

Recordemos que para las unidades de la capa de entrada, al desaparecer I, la trama de activación disminuía hasta 0.

Para el caso de las instars, al eliminar la trama de entrada, la trama de actividad (salida de las instars) queda almacenada, es decir, la salida no desaparece, a lo que se le llama efecto de memoria a corto plazo.

Cuando la función de salida es de orden diferente a 1:

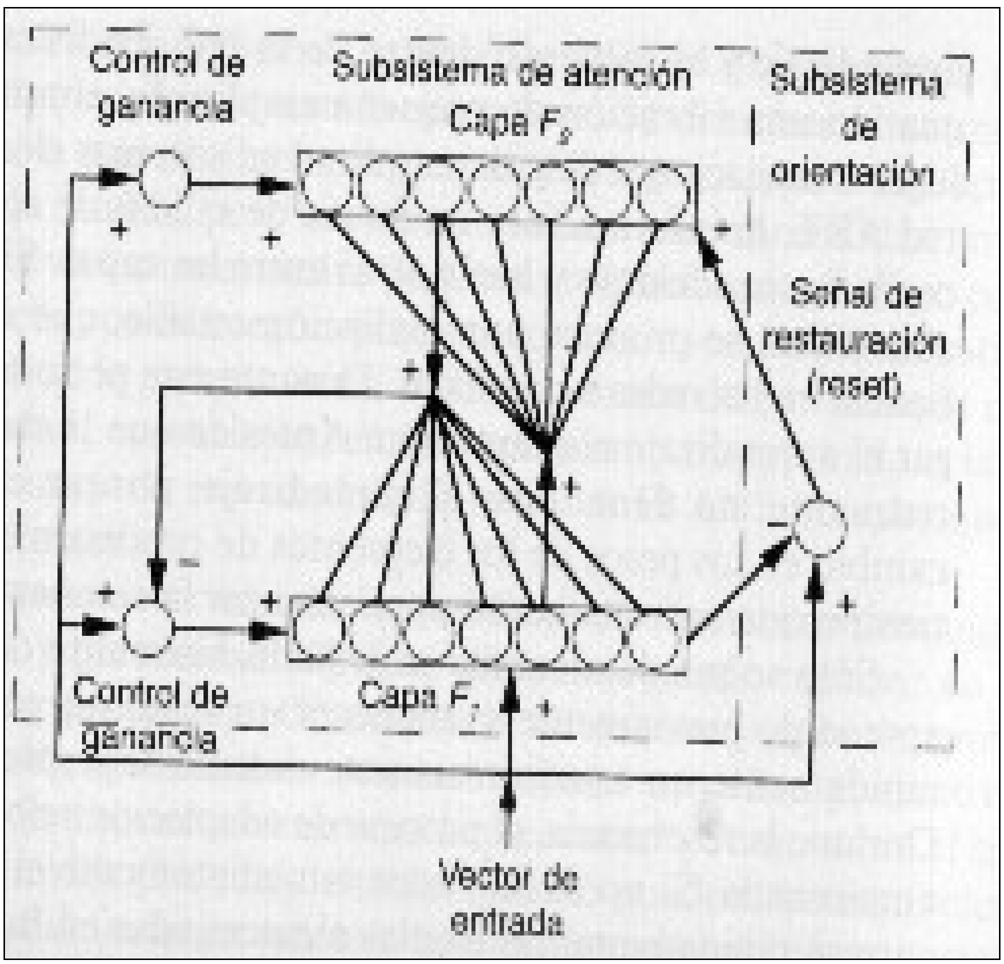
La red tiende a incrementar la actividad de la unidad que posea el mayor valor de respuesta y una vez que se elimina la trama de entrada, sólo la salida de la instar que tenga la mayor respuesta no caerá a cero. Así, esta red supone que todo es ruido excepto la señal mayor.

Cuando la función de salida es sigmoidal:

Combina las características de eliminación de ruido (es decir, se obtiene un umbral de eliminación, donde sólo las unidades cuyas entradas netas se encuentran por encima del umbral, experimentaran un incremento en su actividad) con la capacidad de almacenar una representación precisa del vector de entrada, al tener una función de grado superior a la función lineal para actividades pequeñas y una función lineal para actividades más elevadas.

Si se añade una función con pendiente menor a la lineal para actividades muy grandes, la salida estará acotada para valores muy grandes de actividad.

DESCRIPCIÓN DE LA RED ART.



La red ART recibe su nombre por la forma en que interactúan entre sí el aprendizaje y el recuerdo dentro de la red. En física, la resonancia se produce cuando una vibración de pequeña amplitud y con una frecuencia adecuada da lugar a oscilaciones de gran amplitud en sistemas eléctricos o mecánicos. En una red ART, la información, en forma de salidas de elementos de procesamiento, oscila hacia delante y hacia atrás entre las capas. Si se desarrollan unas tramas adecuadas, se produce una oscilación estable, que es el equivalente de la resonancia en las redes neuronales. Durante este periodo resonante, puede tener lugar el aprendizaje o la adaptación.

Antes de que la red alcance este estado estable o resonante, no se da el aprendizaje, pues el tiempo necesario para que se produzca el aprendizaje es mayor al tiempo en que se alcanza el estado resonante.

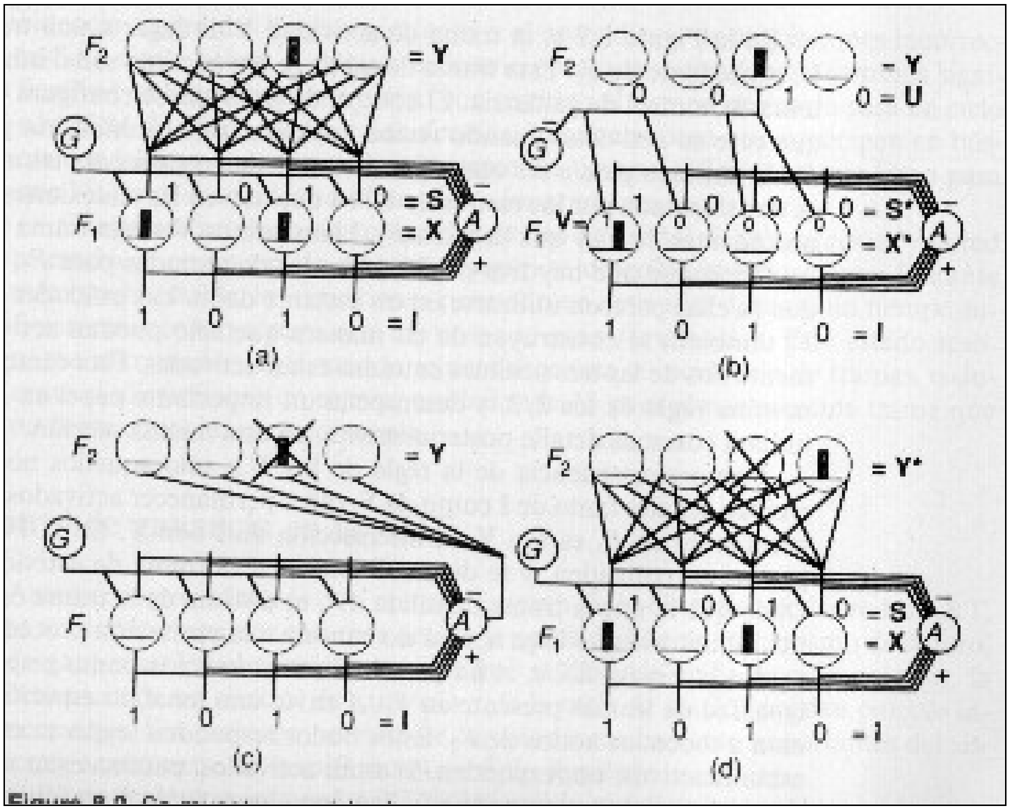
El estado resonante solo se puede alcanzar de dos maneras. Si la red ha aprendido previamente a reconocer un vector de entrada, entonces se alcanzará rápidamente un estado resonante cuando se le presente a la entrada ese vector de entrada y durante este periodo resonante, la red reforzará la memoria de la trama almacenada.Si no se reconoce inmediatamente el vector de entrada, la red buscará a través de todas las tramas almacenadas. Si no encuentra nada, la red entrará en un estado resonante para aprender la trama de la entrada.

Cuanta con un subsistema de atención (el cual tiene dos capas F1 y F2), un control de ganancia y un subsistema de orientación.

Las capas de actividad que aparecen en las capas del subsistema de atención se llaman rastros de memoria a corto plazo, porque sólo existen por asociación a una sola aplicación de un vector de entrada. Los pesos asociados a las conexiones ascendentes y descendentes entre F1 y F2 se llaman memoria a largo plazo, porque codifican información que sigue siendo parte de la red por largos periodos de tiempo.

Reconocimiento de tramas en la red ART.

Se presenta aquí el escenario de una simple operación de reconocimiento de tramas durante la cual una red ART intenta determinar si una cierta trama de entrada se encuentra entre las tramas almacenadas previamente en la red.



Lo que se aprecia en la figura anterior se describe así: en la primera etapa del proceso (a)), una trama de entrada I se presenta en las unidades de la capa F1. Se produce una trama de activación X a través de F1. La capa F1 es la capa de entrada a la red y su funcionamiento está descrito en una sección previa.

El vector de entrada I excita tanto a la capa F1 del subsistema de atención, como al sistema de orientación A y al control de ganancia G.

La trama de activación a través de F1, X, da lugar a una trama de salida S.

La trama de salida de la capa F1, S, da lugar a una señal inhibitoria que también se envía al subsistema de orientación A.

Coma la trama S es igual a la trama de entrada I, esta señal inhibitoria proveniente de S a A cancela el efecto excitatorio procedente de I en A, de tal forma que A permanece inactiva.

Por otro lado, G suministra una señal excitatoria a F1. La misma señal es aplicada a todos los nodos de la capa y por eso se le conoce como señal no específica.

Además, la trama S se envía mediante conexiones a F2. Cada unidad de F2 recibe toda la trama S. Las unidades de F2 calculan su entrada neta en la forma habitual, sumando los productos de los valores de entrada por los pesos de conexión. En respuesta a las entradas provenientes de F1, se produce a través de las unidades de F2 una trama de actividad Y.

F2 es una capa competitiva, cuyo funcionamiento fue descrito también anteriormente.

En la siguiente fase del proceso, (inciso b), se observa que la trama de actividad Y en F2 da lugar a una trama de salida U, la cual se envía como señal inhibitoria al sistema de control de ganancia.

El control de ganancia está configurado de forma tal que cesa su actividad cuando recibe cualquier señal inhibitoria procedente de F2.

U pasa a ser también una segunda entrada para las unidades de F1. Al llegar a F1, U es ponderada con los pesos (memoria a largo plazo) de las conexiones ascendentes de F1, así que U es transformada al multiplicarla por los pesos y al resultado le llamamos V.

Las unidades de F1 y F2 se constituyen de tal forma que sólo pueden activarse si únicamente dos de las tres posibles entradas están activadas. Esta característica se denomina regla de los 2/3, y como consecuencia de ella, sólo aquellos nodos de F1 que reciban señales tanto de I como de V van a permanecer activados. Así, la trama que permanece en F1 es la intersección de I con V.

En la figura se ve que la tramas I y V no coinciden, por lo que se desarrolla una nueva trama de actividad X* en F1 y por tanto también una nueva trama S*, que es distinta a la trama I, por lo que la señal inhibitoria que llega a A desde S* ya no cancela a la excitación procedente de I.

En la tercera figura (inciso c), A se activa como respuesta a la desigualdad presente en F1. A envía una señal no específica de restauración a todos los nodos de F2, los cuales responden según su estado actual. Si están inactivos, no responden, pero si están activados, pasan a estar inactivos y permanecen así durante un intervalo de tiempo considerable. Por tanto, Y desaparece, pues todos los nodos quedan desactivados.

Dado que ya no aparece Y, la salida descendente y la señal inhibitoria que llega al control de ganancia desaparecen también.

En la última figura se observa que la trama original X se vuelve a instaurar en F1, con lo que comienza un nuevo ciclo de búsqueda de igualdad, solo que esta vez aparece una nueva trama Y* en F2. Los nodos participantes en la trama original Y permanecen inactivos debido al efecto a largo plazo de la señal de restauración procedente de A.

Este ciclo de búsqueda de igualdad de tramas se repite hasta que se encuentra una coincidencia o hasta que F2 se queda sin valores almacenados previamente. En cualquiera de ambos casos, la red entra entonces en un estado de resonancia, en donde tiene lugar la modificación de los pesos en las conexiones ascendentes y descendentes de la red.

Si se encuentra una coincidencia, los pesos son modificados para almacenar algún rasgo distintivo de la entrada, así, si la trama almacenada es exactamente la buscada, los pesso solo se refuerzan, y si existe una ligera diferencia, entonces los pesos se ajustan para representar tamién esa característica nueva de la trama de entrada.

Si no se encuentra coincidencia alguna, la red asigna un nodo libre de F2 y empieza a aprender la nueva trama. El aprendizaje tiene lugar a través de la modificación de los pesos.

Este proceso de aprendizaje no comienza ni se detiene, sino que continúa, incluso mientras tiene lugar la búsqueda de igualdades.

Como la capa F2 es una red competitiva, lo que busca cada nodo de la capa (instar) es dar una salida máxima a cada vector de entrada que aparezca, para eso lo que necesita hacer es ir modificando los pesos de forma que éstos se asemejen más al vector de entrada, con lo cual la entrada neta se incrementa y por tato su actividad se incrementa también. Así que la modificación de los pesos tiene lugar incluso en el proceso de búsqueda de igualdades, cada vez que se envían señales a través de las conexiones. Pero la razón para que no existan pérdidas de conocimiento es que el tiempo necesario para que se produzcan cambios significativos en los pesos es muy largo con respecto al tiempo necesario para un ciclo completo de búsqueda de coincidencias . Las conexiones que participan en búsquedas fallidas no están activas un tiempo suficientemente largo para que los pesos asociados resulten afectados gravemente.

Cuando se produce una coincidencia, no hay señal de restauración y la red se establece en un estado resonante. Durante este estado estable, las conexiones permanecen activadas durante un tiempo suficientemente largo para que los pesos se vean fortalecidos.
Esta resonancia sólo puede surgir cuando se produzca una coincidencia de tramas o durante el aislamiento de nuevas unidades de F2 para almacenar una trama que no fuera conocida anteriormente.

Control de ganancia en la red ART.

Puede tenerse un sistema constituido por una jerarquía de redes ART, así que, por ejemplo, la capa F2 podría recibir entradas procedentes de otra capa por enciam de ella, así como de la capa F1 situada debajo de ella. Si la capa F2 fuera estimulada por una capa superior, podría producir una salida descendente y mandar señales a la capa F1. Es posible que la señal descendente llegase a la capa F1 antes de que llegue a F1 una trama de entrada I procedente de la parte inferior. Una señal prematura procedente de F2 podría ser resultado de una anticipación que surgiera en un nivel superior de la jerarquía, lo cual significa que se está indicando lo que se espera que sea la siguiente entrada, antes de que ésta se presente en F1.

Si F1 produjera una salida debido a cualquier entrada que se le presentara, entonces F1 produciría una respuesta en función de esta señal anticipada descendente procedente de F2, con lo cual se produciría el ciclo de búsqueda de coincidencia sin que hubiera llegado a F1 un vector de entrada a buscar. Por esto de incluye en el funcionamiento de la red la regla 2/3 y el control de ganancia.

Como cualquier señal procedente de F2 da lugar a la inhibición de G y G activa de forma no específica a todas las unidades de F1; si está en vigor la regla de los 2/3, la inhibición de G significa que una señal procedente de F2 no puede, por sí misma, desencadenar una salida procedente de F1. Por el contrario, las unidades de F1 quedan sensibilizadas por la trama de entrada I. Al aparecer la trama I en F1, ésta producirá una respuesta ya que está recibiendo entrada de dos de tres entradas posibles; es decir, si está presente una trama descendente procedente de F2, G está inactivo, pero al aparecer I, están entonces dos entradas activas en F1, I y V. Si no hay ninguna trama procedente de F2, entonces, al aparecer I en F1, como I excita también a G, a las unidades de F1 le llegan dos entradas, G e I.

G y la regla de los 2/3 se combinan para permitir que la capa F1 distinga entre una señal de anticipación y un vector de entrada.

ART1

ART1 conserva la misma estructura general mostrada anteriormente. En Art1, la restricción es que todas las entradas a la red deben ser vectores binarios.

Subsistema de atención.

Procesamiento en F1.

Cada elemento de procesamiento, vi, de la capa F1 tiene una actividad x1i y recibe un valor de entrada binario Ii, una señal de excitación, G, procedente del control de ganancia y las señales descendentes, ui, procedentes de F2, las cuales son ponderadas por los pesos zij. Las salida, si, sube hacia F2 y cruza el subsistema de orientación.

Cada unidad de F1 calcula su entrada neta procedente de F2 en la forma:

$$Vi = \sum_{j} uj \; zij$$

Y la función de salida se puede aproximar mediante la función escalón binaria:

$$si = h(x1i) = \begin{cases} 1 & ; x1i > 0 \\ 0 & ; x1i <= 0 \end{cases}$$

La salida del control de ganancia G, depende de las actividades de otras partes de la red.

$$G = \begin{cases} 1 & ; \ I \neq 0 \ \text{y} \ U = 0 \\ 0 & ; \text{en caso contrario} \end{cases}$$

Es decir, si hay un vector de entrada y F2 no está produciendo en ese momento un vector salida, entonces G=1. Cualquier otra combinación de actividad de I y F2 evita prácticamente que el control de ganancia genere su excitación no específica en las unidades de F1.

Pero veamos como es la actividad de las unidades de F1 al presentarse las cuatro combinaciones posibles de I y la salida de F2.

1. Primero, en el caso en que no hay vector de entrada y F2 está inactiva, las unidades sin entradas se mantienen en un estado de actividad en su nivel más negativo.
2. Al aplicar un vector de entrada I, sin que exista una entrada procedente de F2, tanto F1 como el control de ganancia reciben señales procedentes del vector de entrada. Dado que F2 está inactiva, G no está inhibida. En este caso, las unidades que reciben un valor de entrada no nulo (y por default positivo) también generan un valor de actividad mayor que cero, y poseen un valor de salida de 1. Las unidades que no reciben una entrada no nula, sin embargo, experimentan una subida de sus actividades hasta el nivel cero, a través de la señal de excitación procedente de G.
3. En el caso en que existen entradas en las unidades de F1 tanto procedentes de F2, como del vector de entrada I, sucede lo siguiente. Si una unidad tiene un valor de entrada positivo, Ii, y una entrada neta positiva grande procedente de la parte superior, Vi, entonces la regla de los 2/3 dice que la actividad de la unidad es mayor que cero.
4. Si F2 está produciendo una salida descendente. Pero todavía no hay un vector de entrada I , la actividad de la unidad asciende hasta algún valor superior al más negativo, pero sigue siendo negativa porque no se desea que la unidad tenga salida no nula basada en entradas descendentes.

Procesamiento en F2.

Cada elemento de la capa F2, vj, tiene una actividad x2j.

Cada elemento de procesamiento de la capa F2 recibe entradas procedentes de la capa F1, del cual recibe toda la trama de salida S. Además recibe entrada del sistema de control de ganancia, G, y del subsistema de orientación, A. Las señales descendentes ascendentes sji, procedentes de F1, son ponderadas por los pesos zji. La trama de salida U se envía de vuelta a la capa F1. Además, cada unidad recibe un término de realimentación positiva procedente de sí misma y envía una señal idéntica, g(x2j), a trvés de una conexión inhibitoria, a todas las demás unidades de la capa F2.

La entrada neta recibida procedente de la capa F1 es:

$$Tj = \Sigma \quad sj \;\; zji$$

i

La entrada inhibitoria a cada unidad es:

$$J_j = \sum_{k \neq j} g(x2k)$$

La capa F2 es una capa competitiva con interacciones de centro activo periferia inactiva.

La forma funcional de g(x) se selecciona de forma que se incremente la actividad del único nodo de F2 que tenga el mayor valor de entrada neta procedente de F1 y las actividades de los demás nodos se reduzcan a cero, por lo tanto esta función debe ser de orden mayor que 1, por lo visto en la sección de red competitiva previa.

La salida del nodo ganador recibe un valor de uno, por lo que la salida de los nodos de F2 se expresa de la forma:

$$u_j = f(x2j) = \begin{cases} 1 & ;\ T_j = \text{max respuesta } T_k \text{, para todo } k \\ 0 & ;\ \text{en caso contrario} \end{cases}$$

Pareciera que se viola la regla de que en las instars solo se tiene una salida por nodo, al presentar como salidas a f(x2j) y a g(x2j), sin embargo se puede tomar como que la salida única es x2j para la unidad vj y que ésta se envía a otros dos elementos de proceso cada uno de los cuales produce una salida distinta: f(x2j) y a g(x2j) respectivamente.

Pesos descendentes o rastros descendentes de LTM.

Son los pesos de las conexiones procedentes de unidades de F2 y con destino a unidades de F1.

Dado que f(x2j) sólo es no nula para un único valor de j (para un nodo de F2, vj)., sólo los pesos en las unidades de F1, que corresponden a esa unidad, vj, en F2, serán modificados. Así, se tienen tres casos:

$$z_{ij} = \begin{cases} \text{tiende a } 1 & ;\ \text{si } v_i \text{ (en F1) y } v_j \text{ (en F2) están activados} \\ \text{tiende a } 0 & ;\ \text{si } v_i \text{ (en F1) no está activado y } v_j \text{ (en F2) está activado} \\ 0 & ;\ \text{si } v_i \text{ (en F1) y } v_j \text{ (en F2) están desactivados} \end{cases}$$

Si F2 está activada, entonces, vi (en F1) sólo puede estar activada si se recibe una entrada Ii, y además una entrada neta suficientemente grande Vi, procedente de F2. Esta restricción para a ser una condición que afecta al peso zij, el cual debe tener un valor mínimo, sino, irá decreciendo hasta cero incluso en el caso en que la unidad vj (en F2) esté activada y vi (en F1) esté recibiendo una entrada Ii.

Es decir, no basta con que la unidad vi en la capa F1 reciba un elemento Ii del vector de entrada y una entrada neta procedente de la capa F2, sino que esta entrada neta debe ser mayor que un cierto umbral, para que la unidad en F1, vi, pueda tener una actividad mayor que cero. Pero como la entrada neta está constituida por la trama de salida de la capa F2 multiplicada por los pesos de las conexiones de F2 a la unidad en F1, y como las salidas de las unidades de la capa F2 serán todas cero, excepto una, la cual tendrá un valor de uno, entonces, la restricción afecta a los pesos que ponderan salida de F2.

Todos los pesos de conexión descendente deben recibir un valor inicial mayor que el mínimo ya comentado, para que pueda tener lugar el aprendizaje en la red, pues de otra forma, eventualmente, todos los pesos de estas conexiones serán nulos y la red inservible.

Pesos ascendentes o rastros ascendentes de LTM

Sólo los pesos del nodo ganador en la capa F2 serán modificados. Los pesos en la unidad de F2 ganadora se modifican según la siguiente fórmula:

$$zji = \begin{cases} L / (L - 1 + |S|) & ; \text{vi está activado} \\ 0 & ; \text{en caso contrario} \end{cases}$$

Donde L es una constante y |S| es el módulo de S, que corresponde al número de salidas no nulas que haya en F1.

Además:

$$S = \begin{cases} I & ; \text{si F2 está desactivado} \\ I \cap V & ; \text{si F2 esrá activado} \end{cases}$$

Por otro lado, una red que posea nodos libres en F2, es decir, nodos que todavía no hayan participado en ningún aprendizaje, es preciso asegurarse de que sus pesos no sean tan grandes que ganen accidentalmente a un nodo que haya aprendido una trama. Pro tanto, es preciso mantener todos los pesos iniciales por debajo de un cierto valor: $L / (L-1 + M)$, donde M es el número de nodos que hay en F1 y por tanto es el número de conexiones ascendentes que llegan a cada nodo de F2.

Subsistema de orientación.

El subsistema de orientación es el responsable de detectar faltas de coincidencia entre las tramas ascendentes y descendentes de la capa F1. Se puede modelar a este subsistema como un único elemento de procesamiento, A, con una salida para cada una de las unidades de la capa F2. Las entradas de A son las salidas de las unidades de F1, S, y el vector de entrada I. Los pesos de las conexiones que proceden del vector de entrada son todos ellos iguales a un cierto valor P, los correspondientes a las coenxiones procedentes de F1 son todos iguales a un valor –Q. La entrada neta a A es P|I| - Q|S|. La salida de A se activa si la entrada neta pasa a ser no nula:

$$P|I| - Q|S| > 0$$

La magnitud P/Q recibe el nombre de parámetro de vigilancia y suele identificarse mediante el símbolo ρ.

Por tanto, en caso de que |S| / |I| >= ρ, la salida de A no se activa.

Cuando |S| = |I|, el subsistema de orientación debe evitar una señal de restauración a F2; por lo tanto , ρ debe ser menor o igual que 1.

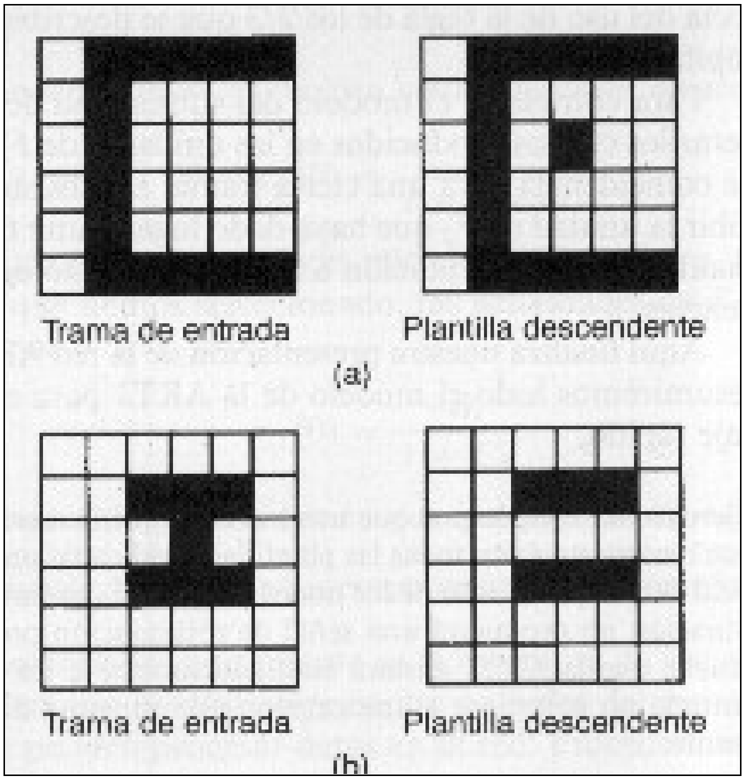
El valor del parámetro de vigilancia mide hasta que grado discrimina el sistema entre distintas clases de tramas de entrada.

El valor de ρ determina la granularidad con la cual son clasificadas las tramas de entrada por parte de la red. Para un cierto conjunto de tramas que haya que clasificar, un valor grande de ρ dará lugar a una discriminación más fina de las clases que la que se tendría con un valor menor de ρ.

Al tener una valor de ρ menor que uno, se permite que la trama descendente que esté codificada en F2 para que representa a una cierta clase pueda cambiar a medida que se le presente a la red nuevos valores de entrada.

Por ejemplo, de la figura se observa que la figura de la derecha (que tiene un punto adicional) había sido codificada por uno de los nodos de F2. La aparición de la trama ascendente sin el punto adiciional no da lugar a una restauración, así que se establece una resonancia entre la capa F1 y el nodo ganador de F2 que haya producido la trama descendente. Durante esta resonancia pueden cambiar los pesos . El nodo de F1 que corresponde a la característica del centro no está activado, por lo que el peso descendente de esa conexión va a desaparecer, y el peso ascendente del nodo de F2 decrecerá.

Esto no tiene problemas con inestabilidades de la red, pues la red ART1 está diseñada para que en el aprendizaje de categorías, la red se estabilice al cabo de una pocas recodificaciones. Esta estabilidad es una consecuencia de la regla de los 2/3.



Cuando se produce una flata de coincidencia para una cierta trama, el subsistema de orientación debe inhibir la unidad de F2 que haya dado lugar a una trama no coincidente y esta inhibición debe mantenerse a lo largo del resto del ciclo de búsqueda de coincidencias.

Resumen del procesamiento en ART1

1- Se aplica un vector de entrada I a F1. Las actividades de F1 se calculan mediante

$$x_{1i} = \frac{Ii}{1 + A1(I_i + B_1) + C_1}$$

2- Se calcula el vector de salida correspondiente a F1

$$s_i = h(x_{1i}) = \begin{cases} 1 & x_{1i} > 0 \\ 0 & x_{1i} \leq \theta \end{cases}$$

3- Se propaga S hacia delante (hasta F2) y se calculan las actividades de acuerdo con

$$T_j = \sum_{i=1} s_i z_{ji}$$

4- Sólo el nodo ganador de F2 posee una salida no nula:

$$u_j = \begin{cases} 1 & T_j = \max_k \{T_k\} \forall k \\ 0 & en\ caso\ contrario \end{cases}$$

supondremos que el nodo ganador es vj.

5- Se propaga la salida de F2 retrocediendo hasta F1. Se calculan las entradas netas procedentes de F2 que llegan a las unidades de F1:

$$V_i = \sum_{j=1} u_j z_{ij}$$

6- Se calculan las nuevas actividades de cuerdo con:

$$x_{1i} = \frac{Ii + D_1 V_1 - B_1}{1 + A1(I_i + D_1 V_1) + C_1}$$

7- Se determinan los nuevos valores de salida, Si, igual que en el paso2

8- Se determina el grado de coincidencia entra la trama de entrada y la plantilla descendente:

$$\frac{|S|}{|S|} = \frac{\sum_{i=1}^{M} S_i}{\sum_{i=1}^{M} I_i}$$

9- Si |S|/|I|<ρ, se marca a vj desactivada, se ponen a cero las salidas de F2 y se vuelve al paso 1 empleando la trama de entrada original. Si |S|/|I| ≥ ρ, continuamos.

10- Se actualizan los pesos ascendentes de vj

$$Z_{Ji} = \begin{cases} \dfrac{L}{L - 1 + |s|} & si\ v_i\ está\ activada \\ 0 & si\ v_i\ está\ desactivada \end{cases}$$

11- Se actualizan solamente los pesos descendentes que provienen de vJ y llegan a todas las unidades de F1:

$$Z_{iJ} = \begin{cases} 1 & \textit{si } v_i \textit{ está activada} \\ 0 & \textit{si } v_i \textit{ está desactivada} \end{cases}$$

12- Se elimina la trama de entrada. Se restauran todas las unidades inactivas de F2. Se vuelve al paso 1 con una nueva trama de entrada.

ART2

Superficialmente, ART2 solamente difiere de ART1 en la naturaleza de las tramas de entrada: ART2 admite componentes analógicas (o de escala de grises) en sus vectores, además de componentes binarios. Esta capacidad implica una significativa mejora del sistema.

Más allá de la diferencia superficial entre ART1 y ART2, existen diferencias de arquitectura que dan a la ART2 su capacidad para tratar tramas analógicas. Estas diferencias son a veces más complejas, y aveces menos, que las estructuras correspondientes de la ART1.

Aparte del hecho evidente consistente en que las tramas binarias y las analógicas difieren en la naturaleza de sus componentes respectivos, la ART2 debe enfrentarse a complicaciones adicionales. Por ejemplo, ART2 debe ser capaz de reconocer la similitud subyacente de tramas idénticas que se superpongan a fondos constantes que posean distintos niveles. Comparadas en un sentido absoluto, dos tramas como éstas pueden tener un aspecto completamente distinto, aun cuando lo cierto es que deberían ser clasificadas como una misma trama.

El precio de esta capacidad adicional es sobre todo un aumento de complejidad en el nivel de procesamiento de F1. El nivel F1 de ART2 consta de varios subniveles y varios sistemas de control de ganancia. El procesamiento en F2 es el mismo. Como compensación parcial de esta mayor complejidad en la capa F1, las ecuaciones LTM son un poco más sencillas para ART2 que lo fueran para ART1.

Los creadores de la arquitectura, Carpenter y Grossberg, han experimentado con varias versiones de la arquitectura de ART2. En el momento de escribir estas líneas, su trabajo sigue adelante. La arquitectura que se va a describir aquí es una de entre varias versiones que han publicado ellos mismos.

Arquitectura de ART2

Tal como se mencionaba en la introducción de esta sección, ART2 posee una semejanza superficial con ART1. Ambas tienen un subsistema de atención y un subsistema de orientación. El subsistema de atención de ambas arquitecturas consta de dos capas de elementos de procesamiento, F1 y F2, de una sistema de control de ganancia.

Procesamiento en F1

La actividad de las unidades de la subcapa F1 esta gobernada por una ecuación de la forma

$$\varepsilon \dot{x}_k = -Ax_k + (1 - Bx_k)J_k^+ - (C + Dx_k)J_k^-$$

en donde A, B, C y D son constantes, La ecuación es casi idéntica a la de ART1. La única diferencia es la aparición del factor multiplicativo en el primer término del lado derecho de la ecuación. Para el modelo ART2 que se presenta aquí, se hará que B y C sean iguales a cero. Al igual que en ART1, J_k^+ y J_k^- representan factores excitatorios e inhibitorios netos, respectivamente. De manera similar, sólo nos va a interesar la solución asintótica, así que

$$x_k = \frac{J_k^+}{A + DJ_k^-}$$

Magnitud

Capa	A	D	J_k^+	J_k^-
W	1	1	$I_i + au_i$	0
X	e	1	wi	||w||
U	e	1	vi	||v||
V	1	1	$f(x_i) + bf($	0
P	1	1	$u_i + \sum_j g(y$	0
Q	e	1	pi	||p||
R	e	1	ui+cpi	||u||+cp||

Los valores de las magnitudes individuales de la ecuación varían según la subcapa que se esté considerando. Por comodidad, se ha construido la Tabla que muestra todas las magnitudes adecuadas para casa subcapa de F1, así como la capa r del subsistema de orientación. Basándose en la tabla, las actividades de cada una de seis capas de F1 se pueden resumir mediantes las ecuaciones siguientes:

$$w_i = I_i + au_i$$

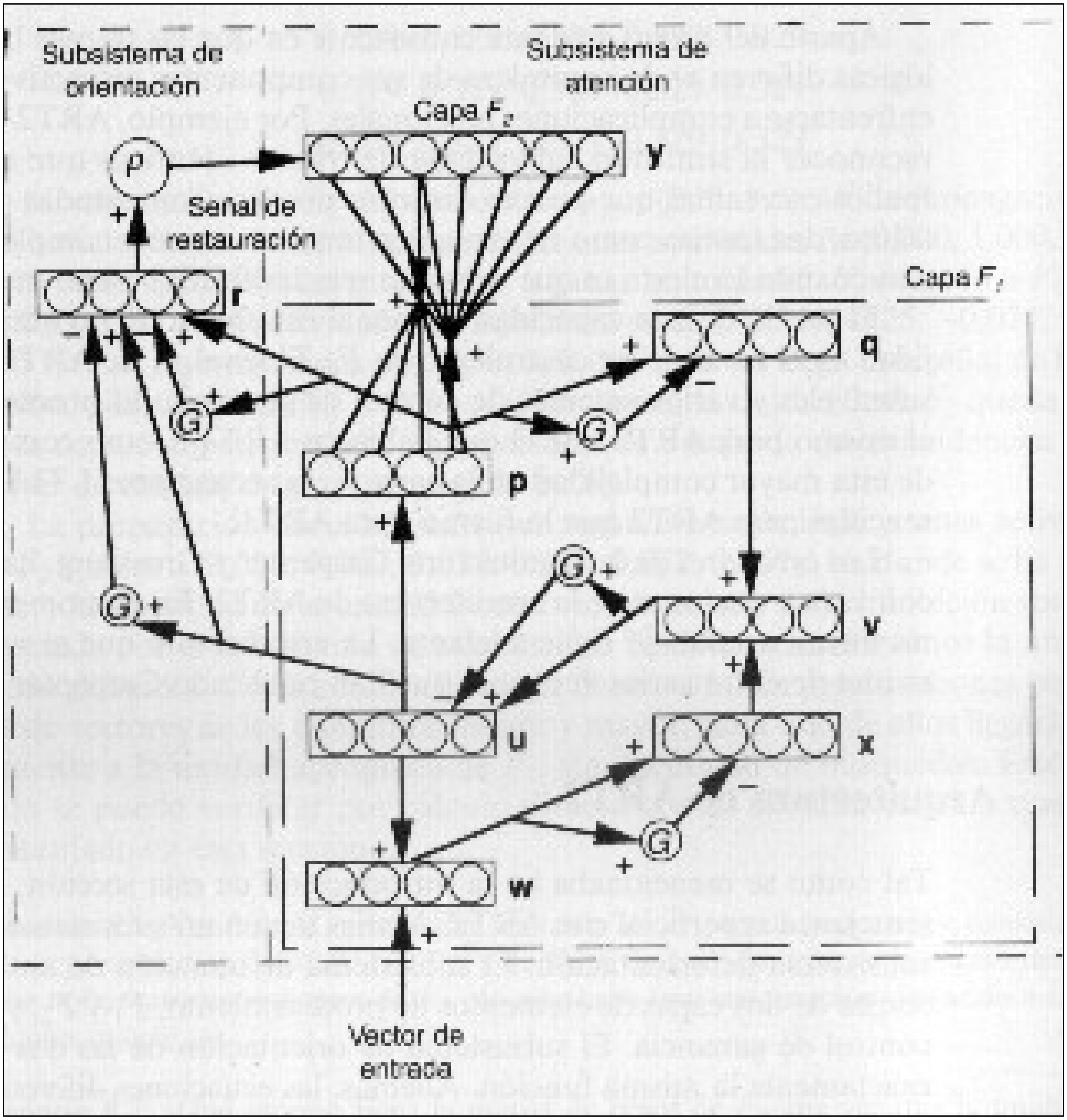
$$x_i = \frac{w_i}{e + \|w\|}$$

$$v_i = f(x_i) + bf(q_i)$$

c

$$p_i = u_i + \sum_j g(y_i) z_{ji}$$

$$q_i = \frac{p_i}{e + \|p\|}$$



Discutiremos dentro de poco la capa r del subsistema de orientación. El parámtero e recibe típicamente un valor positivo y considerablemente menor que

1. Tiene efecto de mantener finitas las activaciones cuando no está presente ninguna entrada en el sistema. No se necesita la presencia de e para nuestro tratamiento, así que se hará e=0 para todo el resto del tema.

Las tres unidades de control de ganancia de F1 inhiben de manera no específica a las subcapas x, u y q. La señal inhibitoria es igual al módulo del vector de entrada que llega a esas capas. El efecto es que las actividades de las tres capas son normalizadas a la unidad por las señales de control de ganancia. Este método es una alternativa para el sistema de centro de activo y periferia inactiva que se presenta después para normalizar las actividades.

La forma de la función f(x) determina la naturaleza de la mejora de contraste que tiene lugar en F1. La elección lógica para esta función podría ser una sigmoide, pero nos quedaremos con la opción de Carpenter:

$$f(x) = \begin{cases} 0 & 0 \le x \le \theta \\ x & x > \theta \end{cases}$$

en donde θ es una constante positiva y menos que 1. En los ejemplos que sigues emplearemos θ=0,2.

Procesamiento en F2

El procesamiento en F2 en la ART2 es idéntico al que se utiliza en ART1.

$$T_j = \sum_i p_i z_{ij}$$

La competencia de F2 da lugar a una mejora de contraste, en el cual se selecciona un único nodo ganador, una vez más de cuerdo con la ART1. La función de salida F2 esta dada por:

$$g(y_i) = \begin{cases} d & T_j = \max_k \{T_k\} \forall k \\ 0 & \textit{en caso contrario} \end{cases}$$

Esta ecuación supone que el conjunto{Tk} contienes únicamente aquellos nodos que no hayan sido restaurados recientemente por el subsistema de orientación. Ahora se puede reescribir la ecuación de procesamiento de la subcapa p de F1 en la forma:

$$p_i = \begin{cases} u_i & si\ F2\ estadesactivada \\ u_i + dz_{ij} & si\ el\ \ j-esimo\ nodo\ de\ F2\ está\ desactivado \end{cases}$$

Ecuaciones de LTM

Las ecuaciones de LTM en ART2 son significativamente menos complicadas que las de ART1. Tanto las ecuaciones ascendentes como las descendentes tienen la misma forma

$$\dot{z}_{ji} = g(y_i)(p_i - z_{ji})$$

para los pesos ascendentes desde vi en F1 hasta vj en F2, y

$$\dot{z}_{ji} = g(y_i)(p_i - z_{ji})$$

para los pesos descendentes que van desde vj en F2 hasta vi en F1. Si vJ es el nodo ganador de F2, entonces se puede utilizar la acuación anterior y demostrar que

$$\dot{z}_{Ji} = d(u_i + dz_{iJ} - z_{Ji})$$

El subsistema de orientación de ART2

Basándonos en la tabla anterior se puede construir la ecuación de las actividades de los nodos de la capa r del subsistema de orientación:

$$r_i = \frac{u_i + cp_i}{\| u \| + \| cp \|}$$

en donde se ha supuesto una vez más, que e=0. La condición para que se produzca la restauración es:

$$\frac{\rho}{\| r \|} > 1$$

en donde ρ es el parámetro de vigilancia, igual que en ART1.

Obsérvese que hay dos subcapas de F1, p y u, que participan en el proceso de búsqueda de coincidencias. A medida que cambian los pesos descendentes en la capa p durante el aprendizaje, la actividad de las unidades de la capa p también va cambiando. La capa u permanece estable durante este proceso, así que su inclusión en el proceso de búsqueda de coincidencias evita que tenga lugar la restauración mientras está teniendo lugar el aprendizaje de una nueva trama.

Supóngase ahora que F2 si tiene una salida, procedente de alguna unidad ganadora, y que la trama de entrada tiene que ser aprendida o codificada por la unidad F2. Tampoco se desea una restauración en ese caso. A partir de la ecuación anterior, se ve que p=u+dzij, en donde la j-ésima unidad de F2 es la ganadora. Si damos valores iniciales nulos a todos los pesos descendentes entonces la salida inicial de F2 no tendrá efecto sobre el valor de p; esto es, p seguirá siendo igual a u.

Al igual que en ART1, una falta de coincidencia suficiente entre el vector de entrada ascendente y la plantilla descendente da lugar a una restauración. En ART2, la trama ascendente se toma en el subnivel u de F1 y la plantilla descendente se toma en p.

Iniciación de LTM ascendente.

Hemos esta describiendo la modificación de los rastros de LTM, o pesos, en el caso de aprendizaje rápido. Examinemos la conducta dinámica de los pesos ascendentes durante un intento de aprendizaje. Supóngase que un cierto nodo de F2 ha codificado anteriormente un vector de entrada tal que zji=ui/(1-d) es el vector de pesos ascendente del j-ésimo nodo de Fs. Supóngase que ese mismo nodo es el ganador para una trama ligeramente distinta, una para la cual el grado de falta de coincidencia no sea suficiente para producir una restauración. Entonces los pesos ascendentes serán recodificados para que coincidan con el nuevo vector de entrada. Durante este proceso dinámico de recodificación, ||zj|| puede disminuir antes de volver al valor. Durante este período decreciente,, ||r|| también irá disminuyendo. Si hay otros nodos cuyos pesos recibido valores iniciales tales que ||zj(0)||>1/(1-d), entonces la red podría cambiar de ganador en medio del proceso de aprendizaje.

Resumen del procesamiento en ART2

1- Se da un valor inicial igual al vector nulo a las salidas de todas las capas y subcapas, y se da el valor uno a un contador de ciclos.

2- Se aplica una trama de entrada, I, a la capa w de F1. La salida de esta capa es:

$$w_i = I_i + au_i$$

3- Se hace una propagación hacia delante, hasta la subcapa x.

$$x_i = \frac{w_i}{e + \|w\|}$$

4- Se hace propagación hacia adelante, hasta la capa v.

$$v_i = f(x_i) + bf(q_i)$$

5- Se hace una propagación hasta la subcapa u.

$$u_i = \frac{v_i}{e + \|v\|}$$

6- Se hace propagación hasta la subcapa p

$$p_i = u_i + \sum_j g(y_i) z_{ji}$$

7- Se hace una propagación hasta la subcapa q.

$$q_i = \frac{p_i}{e + \|p\|}$$

8- Se repiten los pasos 2 al 7 cuantas veces sea necesario para estabilizar los valores de F1.

9- Calcular la salida de la capa r.

$$r_i = \frac{u_i + cp_i}{\| u \| + \| cp \|}$$

10- Se determina si está indicada una restauración. Si ρ/(e+||r||)>1, entonces de envía una señal de restauración a F2. Se marcan todos los posibles nodos activos de F2 como no aptos para la competición, se vuelve a poner a 1 el contador de ciclos, y se vuelve al paso 2. Si no hay restauración, y el contador de ciclos esta a 1, se incrementa el contador de ciclos y se sigue con el paso 11. Si no hay restauración pero en contador de ciclos es mayor que 1, se salta hasta el paso 14, puesto que se ha establecido la resonancia.

11- Se propaga la salida de la capa p hasta la capa F2. Se calculan las entradas netas a F2

$$T_j = \sum_{i=1} p_i z_{ji}$$

12- Sólo el nodo ganador de F2 tiene la salida no nula

$$g(T_j) = \begin{cases} d & T_j = \max_k \{T_k\} \forall k \\ 0 & en\ caso\ contrario \end{cases}$$

13- Se repiten todos los pasos del 6 al 10.

14- Se modifican los pesos ascendentes de la unidad ganadora de F2.

$$z_{Ji} = \frac{u_i}{1-d}$$

15- Se modifican los pesos descendentes de provienen de la unidad ganadora de F2.

$$z_{iJ} = \frac{u_i}{1-d}$$

16- Se elimina el vector de entrada. Se restauran todas las unidades inactivas de F2. Se vuelve al paso 1 con una nueva trama de entrada.

Aplicaciones del modelo ART.

La utilización de la red ART suele estar relacionada con tareas de reconocimiento de patrones, ya que se trata de una red especialmente útil en aquellas aplicaciones que requieran el establecimiento automático de categorías para la clasificación de datos binarios (ART1) y continuos (ART2).
También se ha utilizado para el diseño de sistemas de control y diagnóstico adaptativos.

Se presenta un ejemplo de reconocimiento de imágenes, en el cual se tiene un aprendizaje de dígitos, para el cual se diseña una red para clasificar figuras de dígitos representados mediante 7x6=42 pixeles. Esta red consta de 42 neuronas de entrada a través de las cuales se recibe la figura que debe clasificar. La red debe responder indicando a qué clase, de las establecidas hasta el momento, pertenece el dígito. Si no pertenece a ninguna, deberá crear una nueva. Se fija una valor de 0.9 para el parámetro de vigilancia; es decir, se exije una semejanza mayor del 90% para considerar que dos figuras son de la misma clase.

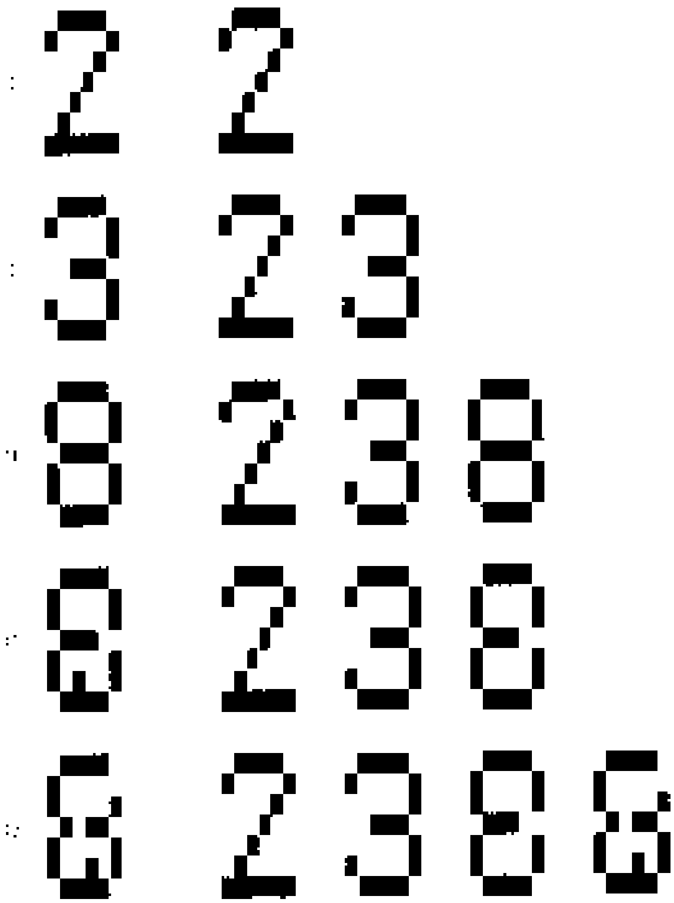


Figura 6.6 Categorías establecidas por la red cuando ρ=0.9

El primer dígito que se aplica a la red es el 2. Como se trata de la primera información, se realiza el aprendizaje completo de la misma, actualizándose los pesos de las conexiones para almacenar este primer ejemplar, que será el representante o prototipo de la primera categoría establecida por la red.

Después se presenta el dígito 3. Es ente caso, la red comprobaría el grado de semejanza con respecto al único prototipo almacenado en la red (el 2). Al ser informaciones binarias, y suponiendo que un pixel negro se representa mediante un bit 1 en el vector de entrada (Ek), y uno blanco se representa mediante un bit 0. Se puede comprobar el grado de semejanza aplicando la fórmula donde en el denominador se indica el número total de pixeles negros, que en este caso son 17, y en el numerador se indica el número de negros coincidentes al superponer el dígito de entrada con el prototipo con el que se compara (operación AND entre ambas figuras). En este caso se puede comprobar que hay 11 pixeles coincidentes, por tanto la relación de semejanza es 11/17 = 0.65, menor que el parámetro de vigilancia, que se había fijado a 0.9, por tanto la red determinaría que el dígito 3 no pertenece a la categoría

del 2, por lo que caería en los pesos del nuevo prototipo (exactamente igual a la figura del 3) para esta nueva categoría.
Cuando se introduce el dígito 8, se compara con los dos prototipos almacenados (2 y 3) obteniéndose una relación de semejanza respecto a 2 de 11/20 = 0.55 < 0.9, y respecto a 3 de 17/20 = 0.85 < 0.9. Pro tanto, se considera como prototipo de una nueva categoría.
Cuando se presenta la cuarta figura, un 8 distorsionado, al comparar con el prototipo 8 se obtiene una relación de 19/20 = 0.95 > 0.9, por lo que se integra en la categoría del 8, ajustando entonces la red los pesos para adaptar el prototipo de esa categoría, incorporando alguna de las características de la nueva figura. En realidad, el nuevo prototipo se obtiene aplicando una operación and entre los dos números.
Finalmente, cuando se presenta el 8 con otra distorsión, la semejanza respecto al prototipo anterior es de 17/20 = 0.89 < 0.9, con lo que no se considera de la misma categoría, creándose una nueva de la que esta figura pasa a ser su representante.
Para comprobar el efecto del parámetro de vigilancia, puede analizarse lo que ocurriría si su valor fuese menor. En ese caso, se estaría relajando la condición de vigilancia, de tal forma que formarán parte de la misma categoría figuras que difieran en una mayor proporción que en el caso anterior. Si, por ejemplo, el parámetro de vigilancia es 0.8, cuando se presente la imagen del 8, la relación de semejanza con el 3 es de 17/20 = 0.85, que es mayor que el parámetro, con lo que se consideran de la misma clase, perdiéndose, por tanto, el dígito 8, al establecer el nuevo prototipo con la operación AND entre ambos patrones. Lo que ocurre al presentar el resto de la imágenes puede observarse en la misma figura.

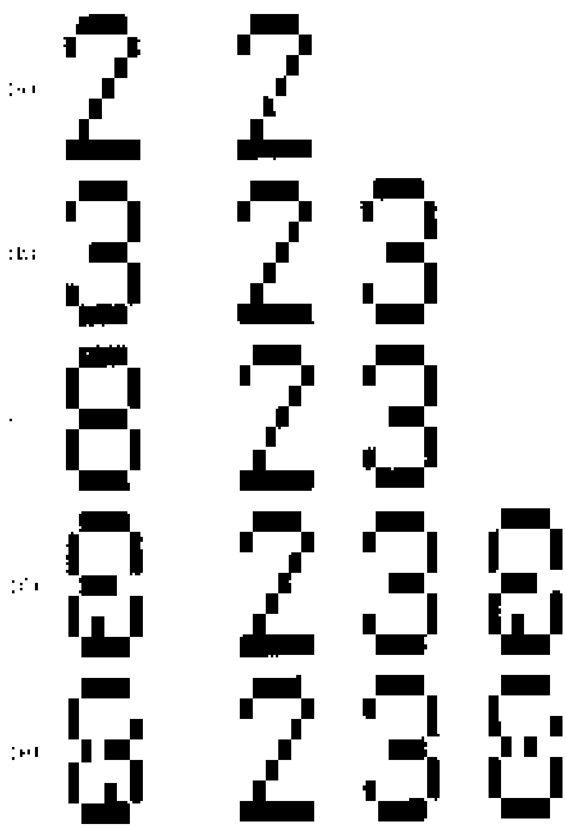


Figura 6.7 Categorías establecidas por la red cuando p=0.8

Un ejemplo más complejo que el anterior es un sistema basado en ART2 que contiene además una etapa de preprocesamiento. En esta etapa previa se realizan tres funciones: a) recepción de la imagen radar, b) detección, marcado y completado del contorno de la imágenes mediante una red neuronal denominada; c) filtrado de la imagen independiente de su posición, rotación o tamaño. La red ART2 recibe la imagen transformada y establece la categoría a la que pertenece.
En la figura se muestra la correcta clasificación de 24 imágenes de 4 tipos de vehículos en cuatro categorías. En este ejemplo se trabaja con un 10% de ruido en las imágenes.

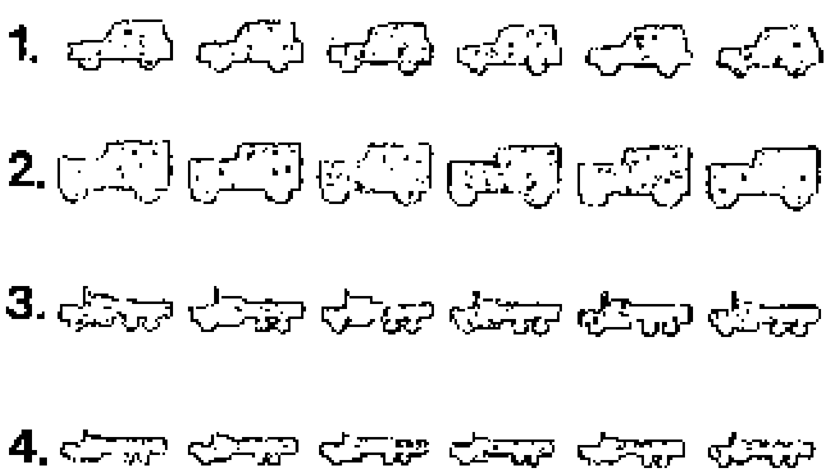


Figura 6.8 Clasificación de 24 imágenes en 4 categorías, realizada por una red ART2 (Carpenter 87)

Se pueden realizar modificaciones espaciales de las figuras (rotación, traslación y cambio de tamaño) y repetir el proceso anterior con las nuevas imágenes. En la figura siguiente se muestra el resultado de la clasificación de 32 figuras, con un 5% de ruido, en 5 categorías.

Se ha producido una división en 2 clases (3,5) de las imágenes que originalmente eran del tipo 3. Este problema puede solucionarse mediante una etapa de postprocesamiento en otra red ART.

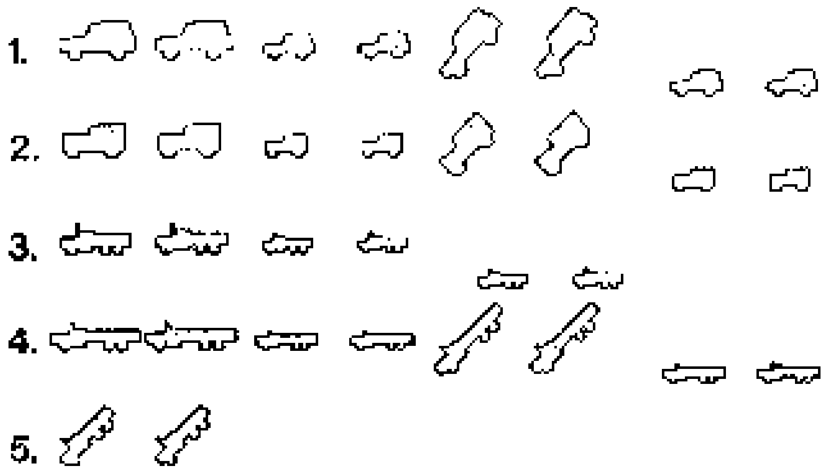


Figura 6.9 Clasificación invariante de 32 imágenes con una red ART-2 [Carpenter 87]

Bobliografia:

Redes neuronales
James A. Freeman
Addison Wesley Iberoamericana

Redes neuronales artificiales
Fundamentos, modelos y aplicaciones.
José R. Hilera
Victor J. Martínez
Addison Wesley Iberoamericana