

Red Multivaluada Estocástica para Optimización. Aplicación al Problema del MaxCut

Domingo López-Rodríguez, Enrique Mérida-Casermeiro, Juan M. Ortiz-de-Lazcano-Lobato

Resumen— El objetivo de este trabajo es presentar la versión estocástica del modelo neuronal multivaluado MREM, el cual ha conseguido muy buenos resultados en muchas aplicaciones, como técnica de optimización. El propósito de esta versión estocástica es el de evitar ciertos mínimos locales de la función objetivo minimizada por la red, esto es, la función de energía. Con este fin, describimos las bases teóricas de este nuevo modelo, que nos garantizan la convergencia de la red a un mínimo local, de una forma rigurosa. Para mostrar la eficiencia de esta versión, el modelo, en ambas versiones, determinista y estocástica, ha sido aplicado para resolver el ya conocido problema de la partición de un grafo, MaxCut. Los experimentos computacionales llevados a cabo nos muestran que el modelo estocástico obtiene mejores resultados que el determinista.

Palabras clave— Red Neuronal, Dinámica Estocástica, Problemas sobre Grafos

I. INTRODUCCIÓN

En la literatura clásica, el problema de MaxCut se define como sigue: Dado un grafo con pesos no dirigido $G = (V, E)$, donde $V = \{v_i\}$ es el conjunto de N vértices y E es el conjunto de n_e arcos, y los pesos de los ejes están dados por la matriz $C = (c_{i,j})_{i,j=1,\dots,N}$ (significando que el peso o coste del arco que une los nodos i y j es $c_{i,j} \geq 0$), encontrar un *corte máximo* de G , es decir, una partición de V en dos conjuntos que maximice el coste total de los arcos cuyos extremos están en diferentes conjuntos de la partición.

Este problema surge en la resolución de muchas situaciones prácticas o teóricas. Algunos ejemplos incluyen: reconocimiento de patrones, clasificación, física estadística y diseño de redes de comunicaciones, y diseño de circuitos [1].

Por tanto, este problema se conoce bastante bien en la literatura. Debido su gran aplicabilidad, muchas variantes suyas han sido formuladas, imponiendo restricciones a la formulación original.

El problema original, con todas las variantes, se sabe que es NP -completo [2], haciendo intratable su resolución computacional, pero el caso de grafos planos pertenece a P , es decir, existe una solución en tiempo polinomial. Por ello, han aparecido muchos algoritmos para resolver MaxCut en el caso general.

Departamento de Matemática Aplicada, Universidad de Málaga, Málaga, España; e-mail: {dlopez,merida}@ctima.uma.es

Departamento de Lenguajes y Ciencias de la Computación, Universidad de Málaga, Málaga, España; e-mail: jnor-tiz@lcc.uma.es

En 1997, Alberti et al. presentaron un modelo neuronal estilo Hopfield para MaxCut [3] pero sus resultados fueron peores que los resultados por Bertoni et al [4]. Takefuyi y sus colegas [5] desarrollaron un modelo neuronal muy potente llamado “máximo” y se demostró que tenía mejores resultados que el resto los algoritmos para resolver un amplio rango de problemas de optimización combinatoria.

En los últimos años, Galán-Marín et al. [6] propusieron un nuevo modelo neuronal llamado OCHOM que obtiene soluciones mucho más eficientes que “máximo”. Además, se puede usar para muchos problemas y tiene también la ventaja de una rápida convergencia a una solución válida sin necesidad de ajustar ningún parámetro. Para hacer que OCHOM escape de mínimos locales, Wang et al [7] han propuesto recientemente una dinámica estocástica para OCHOM, permitiendo descensos temporales de la función objetivo.

Nótese que existen muy pocas referencias bibliográficas para la K -partición (la mayor parte de las referencias se centran en la bipartición).

Recientemente, Mérida et al [8] presentaron un modelo neuronal llamado MREM que puede conseguir K -particiones del grafo, ya que es un modelo multivaluado. Este modelo ha tenido mucho éxito en otros problemas de optimización combinatoria, por ejemplo [9], [10], [11], [12].

El objetivo de trabajo es presentar la versión estocástica del modelo MREM que ayuda a escapar de ciertos mínimos locales, mejorando de esta forma la eficiencia cuando trata con problemas que presentan dificultades, como el que se estudia en este artículo.

II. DESCRIPCIÓN FORMAL DEL PROBLEMA

Sea $G = (V, E)$ un grafo no dirigido sin autoconexiones. $V = \{v_i\}$ es el conjunto de los vértices y E es el conjunto de los n_e arcos. Por cada eje de E hay un peso $c_{i,j} \in \mathbb{R}^+$. Todos los pesos se pueden expresar mediante una matriz real simétrica C , con $c_{i,j} = 0$ cuando no exista un arco con extremos v_i y v_j .

El Problema del Corte Máximo (MaxCut): consiste en encontrar una partición de V en dos subconjuntos A_1 y A_2 , de forma que

$$\sum_{v_i \in A_1, v_j \in A_2, i > j} c_{i,j}$$

sea máximo.

Generalización del Problema MaxCut (K -MaxCut): se busca una partición de V en K subconjuntos disjuntos A_i de forma que la suma de los pesos de los ejes de E que tienen los extremos en diferentes conjuntos de la partición sea máxima. Así, la función a maximizar es

$$\sum_{v_i \in A_m, v_j \in A_n, i > j} c_{i,j} \quad (1)$$

III. EL MODELO MREM ESTOCÁSTICO

Recordemos que el modelo neuronal MREM determinista consta de una serie de neuronas multivalueadas, de forma que el estado de la i -ésima neurona está caracterizado por su salida s_i , que puede tomar cualquier valor en un conjunto finito, denotado por \mathcal{M} .

El estado de la red queda completamente determinado por el vector de estado $\vec{S} = (s_1, s_2, \dots, s_N) \in \mathcal{M}^N$, donde N es el número de neuronas de la red.

Hay una función de energía asociada a cada estado de la red, y se define de la siguiente forma:

$$E(\vec{S}) = -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N w_{i,j} f(s_i, s_j) \quad (2)$$

donde $W = (w_{i,j})$ es una matriz $N \times N$ que representa los pesos de las conexiones entre las diferentes neuronas ($w_{i,j}$ es el peso que la neurona j ejerce sobre la i) y $f : \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R}$ es una función de similitud, es decir, $f(s_i, s_j)$ representa una medida de la similitud entre las salidas de las neuronas i y j .

El propósito de la red determinista es minimizar la función de energía descrita antes. Con este fin, se introduce un estado inicial aleatorio \vec{S}_0 en la red, y en el instante de tiempo t , el vector de estado $\vec{S}(t)$ será cambiado por otro vector de estado $\vec{S}(t+1)$ (definido por la dinámica computacional) si $E(\vec{S}(t+1)) < E(\vec{S}(t))$. Si no hay ningún \vec{S}' en el entorno de $\vec{S}(t)$ tal que $E(\vec{S}') < E(\vec{S}(t))$, entonces $\vec{S}(t)$ es un mínimo local (según la dinámica considerada) y la red para de iterar.

El modelo estocástico MREM tiene la misma arquitectura que el determinista. Está basado en la misma función de energía, dada por la Ecuación (2), pero el descenso de la energía no está garantizado. De hecho, depende de una sucesión $\{T_n\}$, análoga a la sucesión de temperaturas del Enfriamiento Estadístico (Simulated Annealing) [13].

Este nuevo modelo construye una sucesión de vectores de estado $\{\vec{S}_*^{(n)}\}$, donde el superíndice (n) indica que en este estado la temperatura de la red era T_n .

La dinámica consiste en:

- Se genera aleatoriamente $\vec{S}_1^{(1)}$.
- Dado un estado de la red, $\vec{S} = \vec{S}_m^{(n)}$, se muestrea aleatoriamente otro estado \vec{S}' dentro de un entorno $\mathcal{N}_{\vec{S}}$ de \vec{S} .
- Se calcula el incremento de energía correspondiente a actualizar la red de \vec{S} a \vec{S}' : $\Delta E = E(\vec{S}') - E(\vec{S})$.

- Entonces, la red acepta el siguiente estado $\vec{S}_{m+1}^{(n)} = \vec{S}'$ con probabilidad $\mathbb{P}(\Delta E)$, que depende del valor de T_n .

- Si $m+1 = M$, entonces definimos, por simplicidad, $\vec{S}_*^{(n)} = \vec{S}_M^{(n)}$, incrementamos el valor de n y $m = 1$. En otro caso, incrementamos el valor de m .

Se deben satisfacer un par de hipótesis para garantizar la convergencia a estados de energía mínima:

1. $\lim_{n \rightarrow \infty} \mathbb{P}\left(E(\vec{S}_{m+1}^{(n)}) > E(\vec{S}_m^{(n)})\right) = 0$ para todo $m \in \{1, \dots, M-1\}$.

2. La probabilidad de aceptación de \vec{S}' como $\vec{S}_{m+1}^{(n)}$ es de la forma

$$\mathbb{P}(\Delta E) = \begin{cases} 1, & \text{if } \Delta E < 0 \\ g_n(\Delta E) < 1, & \text{if } \Delta E \geq 0 \end{cases}$$

donde $g_n : \mathbb{R}^+ \rightarrow [0, 1)$ y $\Delta E = E(\vec{S}') - E(\vec{S}_m^{(n)})$ (pues, en este caso, $\vec{S} = \vec{S}_m^{(n)}$). Además, debe ser $g_n(\Delta E) > 0$ para todo ΔE . Nótese que la definición de g_n depende de la temperatura T_n .

Se puede probar que si $\{g_n\}$ converge uniformemente a 0, entonces la segunda condición implica la primera.

A continuación describiremos las bases teóricas de este modelo, que garantizan la convergencia a estados de mínima energía:

Lema 1: Se verifica que

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(E(\vec{S}_{m+1}^{(n)}) > E(\vec{S}_1^{(n)})\right) = 0$$

para todo $m \in \{1, \dots, M-1\}$.

Demostración: Probémoslo por inducción:

- Para $m = 1$ lo tenemos por hipótesis.
- Supongámoslo cierto hasta m , es decir,

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(E(\vec{S}_m^{(n)}) > E(\vec{S}_1^{(n)})\right) = 0$$

y veamos que $D_{m+1}^{(n)} = \left\{E(\vec{S}_{m+1}^{(n)}) > E(\vec{S}_1^{(n)})\right\}$ tiene probabilidad cuyo límite es $\lim_{n \rightarrow \infty} \mathbb{P}(D_{m+1}^{(n)}) = 0$.

$D_{m+1}^{(n)}$ está contenido en la unión de tres sucesos:

1. El primero es $A^{(n)} = \left\{E(\vec{S}_m^{(n)}) > E(\vec{S}_1^{(n)})\right\} \cap \left\{E(\vec{S}_{m+1}^{(n)}) > E(\vec{S}_m^{(n)})\right\}$

que es un subconjunto de $\left\{E(\vec{S}_m^{(n)}) > E(\vec{S}_1^{(n)})\right\}$, luego se tiene que $\mathbb{P}(A^{(n)}) \leq \mathbb{P}\left(\left\{E(\vec{S}_m^{(n)}) > E(\vec{S}_1^{(n)})\right\}\right)$ y por tanto

$$0 \leq \lim_{n \rightarrow \infty} \mathbb{P}(A^{(n)}) \leq$$

$$\leq \lim_{n \rightarrow \infty} \mathbb{P}\left(\left\{E(\vec{S}_m^{(n)}) > E(\vec{S}_1^{(n)})\right\}\right) = 0$$

por hipótesis de inducción.

2. El segundo es $B^{(n)} =$

$$\left\{E(\vec{S}_m^{(n)}) > E(\vec{S}_1^{(n)})\right\} \cap \left\{E(\vec{S}_{m+1}^{(n)}) > E(\vec{S}_1^{(n)})\right\}$$

que es un subconjunto de $\left\{E(\vec{S}_m^{(n)}) > E(\vec{S}_1^{(n)})\right\}$, luego llegamos a que su probabilidad es $\mathbb{P}(B^{(n)}) \leq \mathbb{P}\left(\left\{E(\vec{S}_m^{(n)}) > E(\vec{S}_1^{(n)})\right\}\right)$ y por tanto

$$0 \leq \lim_{n \rightarrow \infty} \mathbb{P}(B^{(n)}) \leq$$

$$\leq \lim_{n \rightarrow \infty} \mathbb{P} \left(\{E(\vec{S}_m^{(n)}) > E(\vec{S}_1^{(n)})\} \right) = 0$$

por hipótesis de inducción.

$$3. \text{ Para terminar, el tercero es } C^{(n)} = \\ = \left\{ E(\vec{S}_m^{(n)}) \leq E(\vec{S}_1^{(n)}) \right\} \cap \left\{ E(\vec{S}_{m+1}^{(n)}) > E(\vec{S}_1^{(n)}) \right\}$$

que es a su vez un subconjunto de $\{E(\vec{S}_m^{(n)}) \leq E(\vec{S}_1^{(n)})\} \cap \{E(\vec{S}_{m+1}^{(n)}) > E(\vec{S}_m^{(n)})\}$, luego $\mathbb{P}(C^{(n)}) \leq \mathbb{P}(\{E(\vec{S}_{m+1}^{(n)}) > E(\vec{S}_m^{(n)})\})$ y así

$$0 \leq \lim_{n \rightarrow \infty} \mathbb{P}(C^{(n)}) \leq \\ \leq \lim_{n \rightarrow \infty} \mathbb{P} \left(\{E(\vec{S}_{m+1}^{(n)}) > E(\vec{S}_m^{(n)})\} \right) = 0$$

por hipótesis.

Por tanto,

$$\lim_{n \rightarrow \infty} \mathbb{P}(D_{m+1}^{(n)}) = \lim_{n \rightarrow \infty} \mathbb{P}(A^{(n)}) + \\ + \lim_{n \rightarrow \infty} \mathbb{P}(B^{(n)}) + \lim_{n \rightarrow \infty} \mathbb{P}(C^{(n)}) = 0$$

como queríamos demostrar. \blacksquare

Llamemos $p_n = \mathbb{P}(E(\vec{S}_*^{(n)}) \leq E(\vec{S}_1^{(n)}))$. Entonces, $\lim_{n \rightarrow \infty} p_n = 1$.

Llegamos a uno de los resultados principales de esta sección:

Teorema 1: Con probabilidad 1, existe $L \in \mathbb{R}$ tal que $L = \lim_{n \rightarrow \infty} E(\vec{S}_*^{(n)})$.

Demostración: Sabemos que $\{E(\vec{S}_*^{(n)})\} \subset F(V)$, que es un conjunto compacto, por ser V discreto, luego existe algún punto de acumulación de $\{E(\vec{S}_*^{(n)})\}$. Si esta sucesión sólo posee un punto de acumulación L , entonces éste será su límite: $L = \lim_{n \rightarrow \infty} E(\vec{S}_*^{(n)})$.

Razonando por reducción al absurdo, supongamos que existen varios puntos de acumulación distintos. Sin pérdida de generalidad, podemos suponer que sólo existen dos de ellos, $L < L'$ (el caso en que hay más de dos puntos de acumulación se puede probar de una forma similar a ésta).

Por tanto, existen $\{n_k\}$ y $\{n_j\}$ de forma que $\{\vec{S}_*^{(n_k)}\}$ y $\{\vec{S}_*^{(n_j)}\}$ son subsucesiones complementarias de $\{\vec{S}_*^{(n)}\}$, con:

$$\lim_{k \rightarrow \infty} E(\vec{S}_*^{(n_k)}) = L$$

$$\lim_{j \rightarrow \infty} E(\vec{S}_*^{(n_j)}) = L'$$

Sea entonces $\bar{L} = \frac{L+L'}{2}$. Por lo tanto, por lo mencionado antes, existe un natural N tal que si $n_k, n_j \geq N$ entonces

$$E(\vec{S}_*^{(n_k)}) < \bar{L} < E(\vec{S}_*^{(n_j)})$$

Tomemos $n_k \geq N$ con $n_k + 1 = n_j$ para algún n_j . Hemos de notar que existen infinitos $\{n_k\}$ que lo verifican.

Aplicando lo que ya sabemos, llegamos a que se verifica $E(\vec{S}_*^{(n_k+1)}) \leq E(\vec{S}_1^{(n_k+1)}) = E(\vec{S}_*^{(n_k)})$ con probabilidad p_{n_k} , es decir, se tiene $E(\vec{S}_*^{(n_j)}) \leq E(\vec{S}_*^{(n_k)})$ con probabilidad p_{n_k} , pero esto contradice el hecho, mencionado antes, de que $E(\vec{S}_*^{(n_k)}) < E(\vec{S}_*^{(n_j)})$. Luego llegamos a una contradicción con probabilidad de al menos p_{n_k} . Luego la probabilidad de que exista un único punto de acumulación y, por tanto, exista el límite, es mayor o igual que p_{n_k} .

De todo esto se puede concluir por tanto que existe un único punto de acumulación L de $\{E(\vec{S}_*^{(n)})\}$ con probabilidad superior o igual a p_{n_k} para todo k . Pero, $\lim_{k \rightarrow \infty} p_{n_k} = 1$. Así, $L = \lim_{n \rightarrow \infty} E(\vec{S}_*^{(n)})$ con probabilidad 1. \blacksquare

Corolario 2: Si \hat{S} es un punto de acumulación de la sucesión $\{\vec{S}_*^{(n)}\}$, entonces, con probabilidad 1, tenemos $E(\hat{S}) = \lim_{n \rightarrow \infty} E(\vec{S}_*^{(n)})$.

Lema 2: Se tiene la siguiente igualdad

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(E(\vec{S}_m^{(n)}) \geq E(\vec{S}_*^{(n)}) \right) = 1$$

para todo $m \in \{1, \dots, M-1\}$.

Demostración: Definamos el suceso $E_m^{(n)} = \{E(\vec{S}_m^{(n)}) \geq E(\vec{S}_*^{(n)})\}$, para todo $m \in \{1, \dots, M-1\}$.

Probemos el resultado por inducción:

■ Para $m = M-1$, es consecuencia directa de las condiciones que hemos impuesto desde el principio.

■ Supongamos que $\lim_{n \rightarrow \infty} \mathbb{P}(E_{m+1}^{(n)}) = 1$ y veamos que $\lim_{n \rightarrow \infty} \mathbb{P}(E_m^{(n)}) = 1$:

Primero, notemos que el suceso $E_m^{(n)}$ está contenido en la unión de tres subsucesos:

1. El primero es $A^{(n)} =$

$$= \left\{ E(\vec{S}_m^{(n)}) \geq E(\vec{S}_{m+1}^{(n)}) \right\} \cap \left\{ E(\vec{S}_{m+1}^{(n)}) \geq E(\vec{S}_*^{(n)}) \right\}$$

luego $\lim_{n \rightarrow \infty} \mathbb{P}(A^{(n)}) =$

$$= \lim_{n \rightarrow \infty} \mathbb{P} \left(E(\vec{S}_m^{(n)}) \geq E(\vec{S}_{m+1}^{(n)}) \right) \cdot \mathbb{P}(E_{m+1}^{(n)}) = 1$$

debido a las condiciones impuestas al principio y a la hipótesis de inducción.

2. El segundo es $B^{(n)} =$

$$= \left\{ E(\vec{S}_m^{(n)}) \geq E(\vec{S}_*^{(n)}) \right\} \cap \left\{ E(\vec{S}_m^{(n)}) < E(\vec{S}_{m+1}^{(n)}) \right\}$$

y se deduce

$$\lim_{n \rightarrow \infty} \mathbb{P}(B^{(n)}) = \lim_{n \rightarrow \infty} \mathbb{P} \left(E(\vec{S}_m^{(n)}) \geq E(\vec{S}_*^{(n)}) \right) \cdot \\ \cdot \mathbb{P} \left(E(\vec{S}_m^{(n)}) < E(\vec{S}_{m+1}^{(n)}) \right) = 0$$

debido a las condiciones impuestas.

3. El tercero es $C^{(n)} =$

$$= \left\{ E(\vec{S}_m^{(n)}) \geq E(\vec{S}_*^{(n)}) \right\} \cap \left\{ E(\vec{S}_{m+1}^{(n)}) < E(\vec{S}_*^{(n)}) \right\}$$

y de aquí deducimos

$$\begin{aligned} & \lim_{n \rightarrow \infty} \mathbb{P}(C^{(n)}) = \\ & = \lim_{n \rightarrow \infty} \mathbb{P} \left(E(\vec{S}_m^{(n)}) \geq E(\vec{S}_*^{(n)}) \right) \cdot [1 - \mathbb{P}(E_{m+1}^{(n)})] = 0 \end{aligned}$$

debido a la hipótesis de inducción.

Por tanto,

$$\begin{aligned} & \lim_{n \rightarrow \infty} \mathbb{P}(E_m^{(n)}) \leq \lim_{n \rightarrow \infty} \mathbb{P}(A^{(n)}) + \\ & + \lim_{n \rightarrow \infty} \mathbb{P}(B^{(n)}) + \lim_{n \rightarrow \infty} \mathbb{P}(C^{(n)}) = 1 \end{aligned}$$

Y así queda demostrado este resultado. \blacksquare

Haciendo uso de este lema, podemos demostrar el siguiente teorema:

Teorema 3: Sea \vec{S} un vector de estado con $E(\vec{S}) < L$, donde

$$L = \lim_{n \rightarrow \infty} E(\vec{S}_*^{(n)})$$

Entonces, la probabilidad de muestrear el vector \vec{S} es 0 para todo $n \geq N$.

Demostración: Sabemos que $L = \lim_{n \rightarrow \infty} E(\vec{S}_*^{(n)})$ con probabilidad 1 si, y sólo si, para todo $\varepsilon > 0$ existe $n_0 \in \mathbb{N}$ tal que si $n \geq n_0$ entonces $|E(\vec{S}_*^{(n)}) - L| < \varepsilon$ con probabilidad 1, es decir, $E(\vec{S}_*^{(n)}) \in (L - \varepsilon, L + \varepsilon)$ con probabilidad 1.

Consideremos $\varepsilon = \frac{L - E(\vec{S})}{2}$. Entonces, $E(\vec{S}) < L - \varepsilon = E(\vec{S}) + \varepsilon$.

Sea $n \geq n_0$.

Supongamos que la probabilidad de muestrear el vector \vec{S} es $\rho > 0$.

Consideremos el suceso $A = \{E(\vec{S}_*^{(n+1)}) \notin (L - \varepsilon, L + \varepsilon)\}$ y calculemos una cota para su probabilidad:

$$\mathbb{P}(A) \geq \mathbb{P}(\text{muestrear } \vec{S}) \cdot \mathbb{P}(\text{aceptar } \vec{S}_m^{(n+1)} = \vec{S}).$$

$$\cdot \mathbb{P} \left(E(\vec{S}_m^{(n+1)}) \geq E(\vec{S}_*^{(n+1)}) \right)$$

La primera de estas 3 probabilidades es igual a ρ . La tercera tiene límite 1 cuando n tiende a ∞ (por el lema anterior), así que existe n_1 tal que si $n \geq n_1$ entonces aquella probabilidad es mayor que $\eta > 0$. La segunda es siempre positiva, ya que impusimos la condición de que fuera $g_n(\Delta E) > 0$ para todo n . En particular, es positiva para todo $n \geq n_1$.

Tomemos $N = \max\{n_0, n_1\}$ y $n \geq N$. Llegamos entonces a la conclusión de que $\mathbb{P}(A) \geq \rho\eta\mu > 0$, donde $\mu = \mathbb{P}(\text{aceptar } \vec{S}_m^{(n+1)} = \vec{S})$. Por tanto, tenemos que la probabilidad $\mathbb{P} \left(E(\vec{S}_*^{(n+1)}) \in (L - \varepsilon, L + \varepsilon) \right) = 1 - \mathbb{P}(A) < 1 - \rho\eta\mu < 1$, lo cual contradice el que $E(\vec{S}_*^{(n)}) \in (L - \varepsilon, L + \varepsilon)$ con probabilidad 1 para todo $n \geq n_0$.

Como conclusión, la probabilidad de muestrear el vector de estado \vec{S} es 0. \blacksquare

Este resultado nos trae dos importantes corolarios que tratan sobre la optimalidad de los puntos de

acumulación de la sucesión $\{\vec{S}_*^{(n)}\}$ y la convergencia de esta sucesión.

Proposición 4: Sea $\hat{\vec{S}}$ un punto de acumulación de la sucesión $\{\vec{S}_*^{(n)}\}$. Entonces, $E(\hat{\vec{S}}) \leq E(\vec{S})$ para todo $\vec{S} \in \mathcal{N}_{\hat{\vec{S}}}$. De esta forma, $\hat{\vec{S}}$ es un mínimo local de E .

Demostración: La demostración es una consecuencia inmediata del teorema previo. \blacksquare

Para el siguiente resultado, necesitamos que $\{g_n\}$ converja uniformemente a 0, es decir, necesitaremos que

$$\lim_{n \rightarrow \infty} \|g_n\|_\infty = 0$$

donde $\|g_n\|_\infty = \sup_{t \in \mathbb{R}^+} |g_n(t)|$.

Proposición 5: Si los mínimos locales de E son estrictos, $\{g_n\}$ converge a 0 uniformemente, y $\hat{\vec{S}}$ es un punto de acumulación de la sucesión $\{\vec{S}_*^{(n)}\}$, entonces

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\vec{S}_*^{(n+1)} = \hat{\vec{S}} \mid \vec{S}_*^{(n)} = \hat{\vec{S}} \right) = 1$$

Demostración: Consideremos n tal que $\vec{S}_*^{(n)} = \hat{\vec{S}}$ (existe un número infinito de tales n que verifican esa condición, ya que $\hat{\vec{S}}$ es un punto de acumulación de la sucesión $\{\vec{S}_*^{(n)}\}$). Calculemos la probabilidad de $\{\vec{S}_*^{(n+1)} \neq \hat{\vec{S}}\}$:

$$\mathbb{P}(\vec{S}_*^{(n+1)} \neq \hat{\vec{S}}) = \mathbb{P}(\text{muestrear } \vec{S} \in \mathcal{N}_{\hat{\vec{S}}})$$

$$\cdot \mathbb{P}(\text{aceptar } \vec{S} = \vec{S}_2^{(n+1)}).$$

$$\cdot \mathbb{P}(\text{no regresar a } \hat{\vec{S}}) \leq \mathbb{P}(\text{aceptar } \vec{S} = \vec{S}_2^{(n+1)})$$

La probabilidad de aceptar $\vec{S} = \vec{S}_2^{(n+1)}$ comenzando desde $\hat{\vec{S}} = \vec{S}_1^{(n+1)}$ es

$$\mathbb{P}(\Delta E) = g_n(\Delta E) \leq \|g_n\|_\infty$$

cuyo límite es 0. Por tanto, $\mathbb{P}(\vec{S}_*^{(n+1)} \neq \hat{\vec{S}} \mid \vec{S}_*^{(n)} = \hat{\vec{S}})$ tiende a 0.

De este hecho podemos concluir que

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\vec{S}_*^{(n+1)} = \hat{\vec{S}} \mid \vec{S}_*^{(n)} = \hat{\vec{S}} \right) =$$

$$= 1 - \lim_{n \rightarrow \infty} \mathbb{P} \left(\vec{S}_*^{(n+1)} \neq \hat{\vec{S}} \mid \vec{S}_*^{(n)} = \hat{\vec{S}} \right) = 1$$

y la demostración está completa. \blacksquare

De esta forma hemos demostrado que el modelo MREM estocástico, con el esquema de aceptación-rechazo desarrollado en esta sección, tiene la capacidad de converger a un mínimo local de la función de energía E .

Además, hemos probado que esta convergencia no depende de la velocidad de convergencia de la sucesión de temperatura $\{T_n\}$, ya que sólo depende de la convergencia de $\|g_n\|_\infty$ a 0.

IV. APLICACIÓN DEL MODELO AL PROBLEMA DEL MAXCUT

Para poder resolver el problema del MaxCut con esta red neuronal, necesitamos tantas neuronas como el número de nodos N en el grafo. Cada neurona tomando un valor $s_i \in \mathcal{M} = \{1, 2, \dots, K\}$ señala al subconjunto de la partición al cual se asigna el nodo i -ésimo.

La función de coste del problema del K -MaxCut, dada por la Ecuación (1), debe ser identificada con la función de energía de la Ecuación (2). De esta forma, para MaxCut, es $w_{i,j} = c_{i,j}$, y $f(x, y) = \delta_{x,y}$ (la función delta de Krönecker), también válido para K -MaxCut, ya que es equivalente maximizar el coste de los ejes cortados por la partición y minimizar el coste de los ejes cuyos extremos estén en el mismo subconjunto de la partición.

En este trabajo, una dinámica simple, llamada “best-2”, ha sido implementada.

best-2: consiste en obtener el mayor descenso de la función de energía tan sólo cambiando el estado de dos neuronas a la vez. Así, se debe definir un conjunto de estados adyacentes a uno dado. Si las neuronas que se van a actualizar son p y q , llamaremos a este conjunto $\mathcal{N}_{p,q}$. Entonces, si $\vec{S}(t)$ es el estado de la red en el instante de tiempo t , $\vec{S}(t+1)$ será el vector de uno de los $\mathcal{N}_{p,q}$ que maximice el descenso de la energía, $-\Delta E$. En el caso de este problema, el entorno $\mathcal{N}_{p,q}$ de \vec{S} incluye todos los estados posibles de \mathcal{M}^N que difieren de \vec{S} sólo en las salidas de las neuronas p o q (o ambas). De esta forma, habrá K^2 vectores en $\mathcal{N}_{p,q}$.

Para reducir el coste computacional del modelo, proporcionamos aquí una expresión para el descenso de energía. Supongamos que vamos a actualizar la salida de las neuronas p y q , y que denotamos $s_i(t) = s_i$ y $s_i(t+1) = s'_i$ para todo i . Entonces, el descenso de energía está dado por

$$U_{p,q} = -\Delta E = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N w_{i,j} (f(s_i, s_j) - f(s'_i, s'_j)) = \sum_{i=1}^N (\Delta_{i,p} + \Delta_{i,q}) - \Delta_{p,q} \quad (3)$$

(debido a la simetría de la función f), donde $\Delta_{i,j} = w_{i,j} (f(s_i, s_j) - f(s'_i, s'_j))$.

Así, la dinámica “best-2” se puede resumir como sigue:

1. Se asigna un estado a la red inicialmente de forma aleatoria.
2. Repetir hasta que no haya ningún cambio en el vector de estado de la red:

a) El controlador (scheduling) selecciona un valor $d \in \{1, \dots, \lfloor \frac{N}{2} \rfloor\}$. Para $d > \lfloor \frac{N}{2} \rfloor$, se repetirían los siguientes cálculos, y así podemos reducir el esfuerzo computacional.

b) Lo siguiente se puede hacer en paralelo: cada neurona p estudia todas las posibilidades de cambiar

las salidas de las neuronas p y $q = (p+d) \bmod (N)$, con $0 < q \leq N$, es decir, p calcula el potencial asociado a los posibles cambios, y éste se almacena como un vector \vec{u}_p cuyas componentes representan el descenso de energía asociado a los vectores de $\mathcal{N}_{p,q}$, mediante la aplicación de la Ecuación (3).

c) La neurona p calcula $\vec{\alpha}(p) = \text{máx} \vec{u}_p$, asociado a un estado $\vec{S}_{p,q} \in \mathcal{N}_{p,q}$.

d) El scheduling selecciona el siguiente estado de la red, $\vec{S}(t+1) = \vec{S}_{p,q}$ para el que $p = \arg \text{máx} \vec{\alpha}$.

Algunos resultados experimentales para la dinámica aquí propuesta se muestran en la siguiente sección.

V. RESULTADOS EXPERIMENTALES

En esta sección mostraremos los resultados experimentales de comparar ambas versiones del modelo MREM, determinista y estocástica.

Se ha construido un conjunto de prueba formado por 240 grafos aleatorios, dependiendo de dos parámetros: $N \in \{50, 100, 150, 200\}$ (el cardinal del conjunto de vértices), y $\rho \in \{0'05, 0'15, 0'25, 0'5, 0'75, 0'90\}$ (la densidad de arcos en el grafo, significando que $n_e \approx \rho \frac{N(N-1)}{2}$). Los pesos de los ejes son enteros elegidos aleatoriamente en el conjunto $[0, 10]$. Así, para que el conjunto de prueba fuera completo, se escogieron los valores de los parámetros de forma que cubriesen un amplio rango de grafos.

Para cada grafo, se realizaron 10 ejecuciones independientes. Para MREM se usó la dinámica llamada best-2, y para sMREM (*stochastic MREM*) se usó la versión estocástica de best-2, siendo

$$g_n(\Delta E) = \exp\left(\frac{-|\Delta E|}{T_n}\right)$$

la función de aceptación definida en una sección anterior, y $M = 20$ iteraciones por cada n .

Observemos que la sucesión $\{g_n\}$ converge a 0 uniformemente en el conjunto $[\varepsilon, \infty)$ (con $\varepsilon > 0$), y esto basta para asegurar los resultados de convergencia ya vistos.

El siguiente estado de la red se muestrea con una probabilidad proporcional a $\exp(\frac{-\Delta E}{T_n})$, siendo ΔE el incremento de energía que supone dicho cambio.

En este caso, hemos considerado tan sólo un número finito de temperaturas T_1, \dots, T_{n_a} ($n_a = 5$), decreciendo linealmente desde $T_1 = 1$ a $T_{n_a} = 0$.

En la Tabla I se presentan los resultados de estos experimentos. Se puede comprobar cómo sMREM obtiene mejores resultados de media que la versión determinista en todos los casos, mientras que mejora la solución óptima final en un 86 % de los casos. Por tanto, supone una mejora con respecto a ella. Su única pega es su alto consumo de tiempo de cálculo.

VI. CONCLUSIONES

En este trabajo hemos presentado la versión estocástica del modelo neuronal multivaluado llama-

TABLA I
COMPARACIÓN DE RESULTADOS ENTRE MREM Y sMREM PARA EL PROBLEMA 2-MAXCUT.

N	ρ	MREM			sMREM		
		Mejor	Media	t	Mejor	Media	t
50	0'05	276'8	256'28	0'05	276'8	256'8	1'31
50	0'15	672'8	631'56	0'06	687'2	637'2	1'37
50	0'25	1013'2	970'84	0'06	1020'4	971'24	1'32
50	0'50	1778'8	1724'08	0'06	1774'4	1729'26	1'29
50	0'75	2663'6	2475'48	0'05	2646'4	2476'92	1'34
50	0'90	2941'8	2876'18	0'06	2949'2	2883'94	1'33
100	0'05	990'2	917'72	0'15	971'2	920'68	5'58
100	0'15	2384'4	2323'9	0'16	2408'6	2340'46	5'54
100	0'25	3719'2	3620'9	0'14	3720'8	3646'9	5'28
100	0'50	6711'6	6637'08	0'13	6745'6	6683'9	5'41
100	0'75	9816'2	9524'1	0'14	9813'4	9581'22	5'21
100	0'90	11348'8	11215'06	0'14	11434'4	11254'66	5'29
150	0'05	2009'8	1933'6	0'26	2020'2	1939'86	14'48
150	0'15	5136	5014'62	0'26	5120'4	5066'62	13'85
150	0'25	7990	7807'16	0'26	8068'2	7868'72	12'84
150	0'50	14701'4	14531'06	0'24	14818	14661'4	12'61
150	0'75	21126'2	20899'94	0'22	21308'6	21018'06	12'58
150	0'90	24926	24589'62	0'22	25056	24729'68	12'05
200	0'05	3411'4	3321'84	0'38	3443'8	3340'26	27'58
200	0'15	8765'4	8653'52	0'42	8873'6	8726'9	25'51
200	0'25	13741	13533'9	0'35	13959'4	13707'9	24'24
200	0'50	25750'8	25500'18	0'34	25967'2	25740'64	22'41
200	0'75	37038'6	36789'2	0'32	37217'8	37063'42	22'04
200	0'90	43584'8	43296'26	0'33	43854'8	43527'38	23'17

do MREM, muy útil en muchos problemas de optimización combinatoria, con el objetivo de ayudar a MREM a evitar ciertos mínimos locales de la función de energía.

Hemos propuesto las bases teóricas de este nuevo modelo, basadas en resultados que prueban su convergencia a mínimos de la función objetivo.

Para poder mostrar la efectividad de esta versión estocástica, hemos aplicado los dos modelos al problema del MaxCut, bastante conocido de la literatura especializada gracias a sus aplicaciones, ya que recientemente MREM ha conseguido los mejores resultados obtenidos por un modelo neuronal. Estos experimentos han demostrado que en la inmensa mayoría de los casos sMREM sobrepasa a MREM en calidad de la solución obtenida, pero usando más tiempo de cálculo. La reducción del tiempo computacional usado por sMREM es un tema a estudiar en próximos trabajos.

REFERENCIAS

- [1] F. Barahona, M. Grotschel, M. Junger, and G. Reinelt, "An application of combinatorial optimization to statistical physics and circuit layout design," *Operat. Research*, vol. 36, pp. 493 – 513, 1988.
- [2] M.R. Garey and D.S. Johnson, *Computers and Intractability. A guide to the theory of NP-Completeness*, W. H. Freeman and Company, 1979.
- [3] A. Alberti, A. Bertoni, P. Campadelli, G. Grossi, and R. Posenato, "A neural algorithm for max-2sat: performance analysis and circuit implementation," *Neural Networks*, vol. 10, no. 3, pp. 555–560, 1997.
- [4] A. Bertoni, P. Campadelli, and G. Grossi, "An approximation algorithm for the maximum cut problem and its experimental analysis," in *Algorithms and Experiments*, 1998, vol. 9, pp. 137 – 143.
- [5] Y. Takefuyi and J. Wang, *Neural computing for optimization and combinatorics*, vol. 3, World Scientific, 1996.
- [6] G. Galán-Marín and J. Muñoz-Pérez, "Design and analysis of maximum hopfield networks," *IEEE Trans. on Neural Networks*, vol. 12, no. 2, pp. 329 – 339, 2001.
- [7] J. Wang and Z. Tang, "An improved optimal competitive hopfield network for bipartite subgraph problems," *Neurocomputing*, vol. 61, pp. 413 – 419, 2004.
- [8] E. Mérida-Casermeiro and D. López-Rodríguez, "Graph partitioning via recurrent multivalued neural networks," *Lecture Notes in Computer Science*, vol. 3512, pp. 1149 – 1156, 2005.
- [9] E. Mérida-Casermeiro, G. Galán-Marín, and J. Muñoz Pérez, "An efficient multivalued hopfield network for the travelling salesman problem," *Neural Processing Letters*, vol. 14, pp. 203 – 216, 2001.
- [10] E. Mérida-Casermeiro, J. Muñoz Pérez, and R. Benítez-Rochel, "Neural implementation of dijkstra's algorithm," *Lecture Notes in Computer Science*, vol. 2686, pp. 342 – 349, 2003.
- [11] E. Mérida-Casermeiro and D. López-Rodríguez, "Multivalued neural network for graph maxcut problem," in *Lecture Series on Computer and Computational Sciences 2004*, 2004, vol. 1, pp. 375 – 378.
- [12] D. López-Rodríguez and E. Mérida-Casermeiro, "Matrix bandwidth minimization: A neural approach," in *Lecture Series on Computer and Computational Sciences 2004*, 2004, pp. 324 – 327.
- [13] S. Kirkpatrick, C. D. Gellat, and M.P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, pp. 671 – 680, 1983.