

Análisis señal-ruido de una red neuronal de Hopfield con sinapsis simétricas y distribución de grado arbitraria

Francisco Reynaga¹, Gabriel China¹, Mario Barragán¹ y
Josué Garnica¹

¹ Universidad Tecnológica de Manzanillo,
Camino Hacia las Humedades SN, Col. Salagua, Manzanillo, Colima
Correo electrónico: francisco.reynaga@gmail.com
(Paper received on August 10, 2012, accepted on August 24, 2012)

Resumen. En este trabajo se estudia, usando un método señal-ruido, una red neuronal de Hopfield con sinapsis simétricas y distribución de grado arbitraria. Las ecuaciones obtenidas tienen una muy buena correspondencia con las predicciones realizadas por simulación.

Se analiza la evolución temporal del traslape para redes con tres distribuciones de grado distintas: una distribución Delta, una distribución de Poisson y una distribución con Ley de Potencia. Se muestra que la red con la distribución delta tiene el mejor rendimiento de las tres distribuciones. La distribución con ley de potencia tiene el peor desempeño.

Palabras clave: redes neuronales, máquinas de boltzmann, procesos estocásticos, análisis señal-ruido.

1 Introducción

La arquitectura de las redes neuronales es un tema que ha tomado relevancia en fechas recientes. De especial interés son las arquitecturas que tienen una similitud con las redes biológicas.

En 1982, Hopfield presentó un modelo de red neuronal con memoria autoasociativa, comportamiento determinista y una arquitectura completamente conectada. Este modelo fue generalizado a un modelo estocástico también llamado “Máquinas de Boltzmann”. Estas redes se han aplicado con éxito a problemas de visión artificial (ver por ejemplo [1], [2] y [3]). Algunas de las aplicaciones de visión artificial que contribuyen a la solución de los grandes retos de las Tecnologías de la Información son el registro automático de placas de autos, reconocimiento de rostros, análisis de imágenes de ultrasonidos y rayos X, etc.

El análisis de redes neuronales que recuperan secuencias de patrones con diversas arquitecturas se ha estudiado en [4]. Recientemente se publicó un trabajo donde se estudia, usando un análisis señal-ruido, el comportamiento de las redes neuronales de Hopfield estocásticas con distribución de grado arbitraria para sinapsis asimétricas [5]. En dicho artículo se realizó un análisis comparativo de varias arquitecturas de redes neuronales. En el presente trabajo se extiende el análisis a redes neuronales con sinapsis simétricas. El estudio de redes neuronales con sinapsis simétricas es importante, ya que las Máquinas de Boltzmann también presentan sinapsis simétricas. Por otra parte, el estudio comparativo de arquitecturas de redes permitirá determinar aquellas con mejor desempeño en las Máquinas de Boltzmann.

Para el análisis teórico, se utiliza un método señal-ruido que fue originalmente presentado en [6] para el modelo de Hopfield completamente conectado, el cual es extendido en este trabajo a redes neuronales con distribución de grado arbitraria. Este método genera resultados exactos en modelos con sinapsis simétricas completamente conectados. En las siguientes secciones se presenta una extensión del análisis señal-ruido a redes neuronales con distribución de grado arbitraria. El análisis se realiza en redes neuronales donde la conectividad promedio por neurona es muy grande (infinita, en el análisis matemático). En [7] se estudian redes con una conectividad pequeña (finita, en el análisis matemático) por neurona. Sin embargo es muy difícil obtener la evolución temporal de los parámetros de orden con el método propuesto en dicho trabajo para el caso de sinapsis simétricas. La ventaja del método propuesto en el presente trabajo es que se puede obtener la evolución temporal a tiempos grandes, permitiendo analizar características de la red cuando se va a un estado estacionario, por ejemplo, la cuenca de atracción.

Este trabajo se compone de la siguiente forma: en la sección 2 se presentan las definiciones de la red de Hopfield con distribución de grado arbitraria y su regla de evolución. En la sección 3 se presenta el análisis señal-ruido y las ecuaciones de los parámetros de orden resultantes. En la sección 4 se realiza un análisis numérico usando algoritmos del tipo Monte Carlo y simulaciones numéricas que se comparan con los resultados analíticos. Por último, en la sección 5 se presentan las conclusiones.

2 Definiciones

Estudiaremos redes que consisten de N neuronas binarias $\sigma_i \in S$ con $S = \{-1, 1\}$. Un estado de la red es descrito por

$$\sigma(t) = \{\sigma_i(t)\} \quad i \in \{1, \dots, N\} \quad (1)$$

y la historia del sistema hasta el tiempo t es descrita por

$$\sigma = \{\sigma_i(t')\} \quad i \in \{1, \dots, N\}, \quad t' \in \{0, \dots, t\}. \quad (2)$$

Se tiene interés en almacenar p patrones $\xi^\mu = \{\xi_i^\mu \in S\}$ con $i \in \{1, \dots, N\}$, $\mu \in \{1, \dots, p\}$. Las componentes de los patrones forman una colección de variables aleatorias idénticamente distribuidas (*vaiid*), tomando los valores ± 1 con probabilidad $1/2$. Cada neurona está conectada con el resto de las neuronas por sinapsis J_{ij} que definen el modelo a estudiar.

El campo local de la neurona i se define por $h_i(t) = \sum_j J_{ij} \sigma_j(t)$. Al

tiempo $t = 0$, todas las neuronas se escogen *vaiid* de acuerdo a la distribución de probabilidad

$$P(\sigma_i(0) = \pm 1) = \frac{1 \pm m_0 \xi_i^1}{2}. \quad (3)$$

Así que, el traslape entre el estado inicial $\sigma(0)$ y el patrón ξ^1 es m_0 en el límite $N \rightarrow \infty$.

En una red con distribución de grado arbitraria, las sinapsis están definidas de la siguiente forma

$$J_{ij} = \frac{c_{ij}}{c} \sum_{\mu=1}^p \xi_i^\mu \xi_j^\mu \quad J_{ii} = 0. \quad (4)$$

donde c representa la conectividad promedio de las neuronas. Definimos el grado k_i de la neurona i (número de sinapsis de dicha neurona) como $k_i = \sum_j c_{ij}$. Así

que, $c = \frac{1}{N} \sum_i k_i$. La capacidad de estas redes está dada por $\alpha = p/c$. Las propiedades de la arquitectura están descritas por el conjunto de variables $\mathbf{c} = \{c_{ij}\}$, $i, j \in \{1, \dots, N\}$, con $c_{ij} \in \{0, 1\}$. Estas variables se escogen aplicando las siguientes probabilidades (con la restricción para los grados mostrada arriba):

$$P(c_{ij}) = \frac{c}{N} \delta_{c_{ij}, 1} + \left(1 - \frac{c}{N}\right) \delta_{c_{ij}, 0}. \quad (5)$$

En nuestro caso $c_{ii} = 0$, que implica ausencia de autointeracciones, y $c_{ij} = c_{ji}$ ya que los modelos en este trabajo tienen sinapsis simétricas. Los grados de conectividad son establecidos mediante una distribución de grado arbitraria $P(k_i = k)$, tal que $c = \sum_k k P(k)$. Escogemos la conectividad c de tal forma que diverge con N , es

dicho, $\lim_{N \rightarrow \infty} \frac{1}{c} = 0$. Sin embargo $\lim_{N \rightarrow \infty} \frac{c}{N} = 0$. En el caso $c = N$, se regresa al modelo de Hopfield completamente conectado. El límite termodinámico se da cuando $c, N \rightarrow \infty$.

Todas las neuronas se actualizan simultáneamente de acuerdo a la siguiente probabilidad de transición

$$P(\sigma_i(t+1) | \sigma_{(i)}(t), \mathbf{c}_{(i)}, \xi) = \frac{1}{2} [1 + \sigma_i(t+1) \tanh(\beta h_i(t))] \quad (6)$$

Donde

$$\mathbf{\sigma}_{(i)}(t) = \mathbf{\sigma}(t) \setminus \{\sigma_i(t)\} \quad (7)$$

$$\mathbf{c}_{(i)} = \mathbf{c} \setminus \{c_{ii}\} \quad i \in \{1, \dots, N\}, \quad (8)$$

$$\xi = \{\xi_j^\nu\} \quad j \in \{1, \dots, N\}, \quad \nu \in \{1, \dots, p\} \quad (9)$$

con $\beta = T^{-1}$. La cantidad T , también llamada temperatura, es una medida del comportamiento estocástico de la red (en el caso $T = 0$ el sistema es completamente determinista).

El traslape está definido de la siguiente forma $m(t) = \frac{1}{N} \sum_{j=1}^N \xi_j^1 \sigma_j(t)$. De las condiciones iniciales (4), $m(0)$ es de orden $O(1)$. Para $t > 0$ introducimos el ansatz condensado, suponemos que el sistema tiene una correlación finita, en el límite termodinámico, solamente con el patrón ξ^1 al cual llamamos condensado. Es decir, se supone que $m(t)$ es de orden $O(1)$. Debido a que el patrón ξ^μ , $\mu \neq 1$, es independiente de ξ^1 , la correlación entre el sistema y el patrón ξ^μ es pequeña para N grande.

3 Análisis señal-ruido

El campo local se separa en dos partes

$$h_i(t) = \xi_i^1 \frac{1}{c} \sum_{j \neq i} c_{ij} \xi_j^1 \sigma_j(t) + \frac{1}{c} \sum_{j \neq i} c_{ij} \sum_{\mu \neq 1}^p \xi_i^\mu \xi_j^\mu \sigma_j(t) \quad (10)$$

El primer término del lado derecho de la ecuación anterior, que tiene al patrón condensado, se interpreta como una señal y el segundo que es ruido. Se puede realizar un análisis señal-ruido como el desarrollado en [6] para el modelo de Hopfield completamente conectado. El resultado del análisis son tres parámetros de orden $m(t)$, $C(t, t')$ y $G(t, t')$ que describen la dinámica de la red estudiada: el traslape

$$m(t) = \frac{1}{N} \sum_{j=1}^N \xi_j^1 \sigma_j(t) \stackrel{\text{Pr}}{=} \langle \xi^1 \sigma(t) \rangle_*, \quad (11)$$

la función de correlación

$$C(t, t') \equiv E \left[\frac{1}{N} \sum_{j=1}^N \sigma_j(t) \sigma_j(t') \right]^{\text{Pr}} = \langle \sigma(t) \sigma(t') \rangle_*, \quad (12)$$

y la función respuesta, que se escribe con la ayuda de un campo externo $\theta_j(r)$ que se introduce al campo local $h_j(r)$, $r \in \{0, \dots, t-1\}$,

$$G(t, t') = \frac{1}{N} \sum_{j=1}^N \left(\frac{\partial \langle \sigma_j(t) \rangle_j}{\partial \theta_j(t')} \Bigg|_{\{\theta_j(r)=0\}} \right) \xrightarrow{\text{Pr}} \frac{\partial \langle \sigma(t) \rangle_*}{\partial \theta(t')} \Bigg|_{\{\theta(r)=0\}}. \quad (13)$$

en todos los casos $\sigma(t)$ representa una neurona efectiva, Pr denota convergencia en probabilidad y E el valor esperado. El promedio efectivo (que aparece en las ecuaciones anteriores) de una función $f(\sigma)$, para $t_f > 0$, está dada por (omitiendo el índice en el patrón condensado) :

$$\langle f(\sigma) \rangle_* = \sum_{\sigma(0) \dots \sigma(t_f)} \sum_k \sum_{\xi} \int d\Phi P(\Phi) P(k) P(\xi) P(\sigma | \xi, \Phi, k) f(\sigma) \quad (14)$$

$$\text{donde } \sigma = (\sigma(0), \dots, \sigma(t_f)), \Phi = (\phi(0), \dots, \phi(t_f - 1)), d\Phi = \prod_{t=0}^{t_f-1} d\phi(t).$$

La variable k representa el grado de la neurona efectiva. Las variables aleatorias ϕ tienen distribución Gaussiana con promedio cero y matriz de covarianza $\mathbf{C} = \{C(t, t')\}$. En este promedio efectivo, $P(\Phi)$ es una distribución de probabilidad Gaussiana multidimensional

$$P(\Phi) = \frac{1}{\sqrt{\det(2\pi\mathbf{C})}} \exp \left(-\frac{1}{2} \sum_{t,t'=0}^{t_f-1} \phi(t) \mathbf{C}^{-1}(t, t') \phi(t') \right). \quad (15)$$

Por otra parte, $P(\sigma | \xi, \Phi)$ es la distribución de probabilidad de trayectoria de la neurona efectiva

$$P(\sigma | \xi, \Phi, k) = P(\sigma(0)) \prod_{r=0}^{t_f-1} \frac{1}{2} [1 + \sigma(r+1) \tanh(\beta h(r))], \quad (16)$$

la cual depende del campo efectivo

$$h(t) = \xi \frac{k}{c} m(t) + \alpha \frac{k}{c} \sum_{t'=0}^{t-1} G(t, t') \sigma(t') + \sqrt{\frac{\alpha k}{c}} \phi(t) + \theta(t). \quad (17)$$

El campo efectivo contiene tres términos, el primero se genera a partir de la señal en (10). El segundo y tercer términos se generan a partir del ruido. Como se explica en [6], el segundo término es característico de las redes con sinapsis simétricas y se genera debido a los bucles en la arquitectura de las redes. El tercer término es ruido Gaussiano, el cual se presenta tanto en redes con sinapsis simétricas como asimétricas.

4 Análisis numérico

El proceso para encontrar la evolución temporal de los parámetros de orden se complica desde los primeros pasos temporales. Por ejemplo, al tiempo t , es necesario realizar una integral sobre t variables y una traza sobre t neuronas. Por este motivo, la solución se debe encontrar en forma numérica. Vamos a utilizar el método Eissfeller-Opper [8], el cual es una variante de los métodos del tipo Monte Carlo, para encontrar los parámetros de orden. En este método se genera un número M grande de réplicas del sistema efectivo, las cuales evolucionan formando M trayectorias de las variables involucradas en el promedio efectivo.

Se han obtenido las soluciones de las ecuaciones de los parámetros de orden numéricamente, usando el método Eissfeller-Opper. En nuestros cálculos el número de réplicas es $M = 5 \cdot 10^5$. La evolución del traslape m en los primeros pasos temporales es mostrado en la figura 1 para $T = 0.15$, $\alpha = 0.10$. Las líneas con marcas representan las predicciones teóricas obtenidas con el método Eissfeller-Opper. Las líneas con guiones fueron obtenidas mediante la simulación de una red de $N = 62500$ neuronas promediadas sobre 10 corridas y considerando 10 patrones. En ambos casos se utilizó la distribución delta con conectividad promedio $c = 100$. Los resultados de las simulaciones soportan nuestras predicciones teóricas.

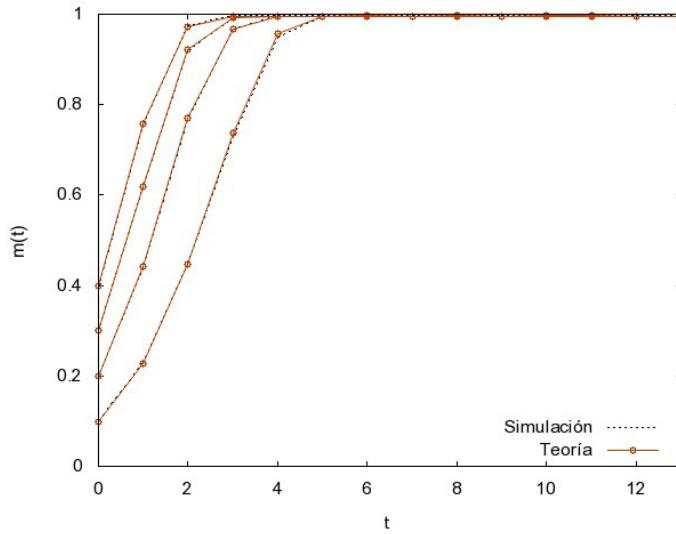


Fig. 1. Evolución temporal del traslape m con $T = 0.15$, $\alpha = 0.10$ y $c = 100$. Las líneas con marcas denotan resultados teóricos y las líneas con guiones los resultados obtenidos por simulación en computadora con $N = 62500$ neuronas promediados sobre 10 corridas.

El análisis matemático considera una red neuronal con $c, N \rightarrow \infty$. Por otra parte en la simulación, $c = 100$ y $N = 62500$, la correspondencia con las simulaciones se mantiene aún para conectividades y número de neuronas menores. Para conectividades muy pequeñas es necesario utilizar el método propuesto en [7]. Sin

embargo, como se ha mencionado anteriormente, el seguimiento temporal de los parámetros de orden es muy complicado con este método. Sería interesante realizar un análisis de los efectos del tamaño finito de la red en la simulación.

Queremos investigar la influencia de la distribución de grado en el rendimiento de la red neuronal. Por esta razón, comparamos en la figura 2 la evolución temporal del traslape para modelos con una distribución de grado con una Ley de Potencia $P(k) = Ak^{-3}$, donde A es una constante de normalización, una de Poisson

$$P(k) = \frac{e^{-c} c^k}{k!} \text{ y una Función Delta dada por } P(k) = \delta(k - c).$$

La distribución con

Ley de Potencia tiene especial importancia debido a que se ha observado en redes biológicas.

La evolución del traslape m , en los primeros pasos temporales, se muestra en la figura 2 para $T = 0.15$, $\alpha = 0.150$ y $c = 133$. Se han obtenido las soluciones de las ecuaciones de los parámetros de orden numéricamente, usando el método Eissfeller-Opper. En nuestros cálculos el número de réplicas es $M = 5 \cdot 10^5$. Las líneas con guiones representan las predicciones obtenidas para la Función Delta, las líneas sólidas el caso de la Función de Poisson y las líneas con marcas la Función con Ley de Potencia. En el recuadro, en la figura 2, se puede ver con más detalle el comportamiento de la red neuronal en cada arquitectura.

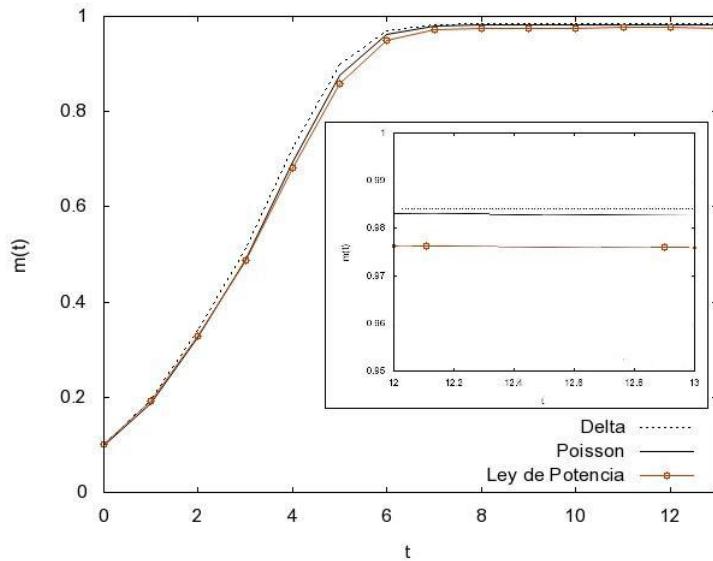


Fig. 2. Evolución temporal del traslape m con $T = 0.15$, $\alpha = 0.150$. Las líneas con guiones representan las predicciones obtenidas para la función delta, las líneas sólidas el caso de la función de poisson y las líneas con marcas la función con ley de potencia. En el recuadro se ve de forma más detallada el comportamiento diferente entre las tres arquitecturas.

De acuerdo a los resultados numéricos, el mejor desempeño lo tiene el modelo con la Función Delta y el peor el que usa la Función con Ley de Potencia. Este comportamiento se puede explicar usando (17). Las neuronas con una conectividad pequeña sufren más perturbaciones del ruido que presenta el campo, en comparación a las neuronas con mayor conectividad. En el caso de la Función con Ley de Potencia, se tiene un número importante de neuronas con conectividad pequeña, lo cual afecta el rendimiento general de la red. Resultados similares se presentan en el modelo con sinapsis asimétricas estudiado en [5]. Este comportamiento podría ser un elemento constante en los modelos de Hopfield, debido a que, como se ha mencionado anteriormente, el ruido Gaussiano se encuentra tanto en modelos con sinapsis simétricas como asimétricas.

5 Conclusiones

En este trabajo se obtuvieron, usando un método señal-ruido, las ecuaciones de la dinámica (expresadas en términos de parámetros de orden) de una red neuronal de Hopfield con sinapsis simétricas y distribución de grado arbitraria cuando el número de neuronas $N \rightarrow \infty$. La importancia del estudio de redes neuronales con sinapsis simétricas radica en que las Máquinas de Boltzmann presentan este tipo de sinapsis. Las arquitecturas presentadas en este estudio pueden ser implementadas en problemas reales (por ejemplo, en visión artificial) con Máquinas de Boltzmann. Una tarea aún por resolver es encontrar un algoritmo de aprendizaje eficaz para este tipo de arquitecturas.

Las ecuaciones de los parámetros de orden se resolvieron numéricamente usando el método Eissfeller-Opper. Se realizaron simulaciones numéricas en un sistema finito. Los resultados de las simulaciones tienen una muy buena correspondencia con las predicciones teóricas.

Se han comparado varias arquitecturas de redes, determinando las características óptimas para las Máquinas de Boltzmann. Se analizó la evolución temporal del traslape de tres modelos de redes neuronales. El mejor rendimiento lo obtuvo la red con distribución de grado Delta y el peor lo obtuvo el modelo con la distribución de Ley de Potencia. Las neuronas con poca conectividad generan este comportamiento, disminuyendo el rendimiento general de la red neuronal. Se observa que este comportamiento es consistente en las redes neuronales de Hopfield, ya que también se ha reportado para redes con sinapsis asimétricas.

En una versión extendida se planea hacer un análisis comparativo de la cuenca de atracción de las distintas arquitecturas mostradas en este trabajo.

Bibliografía

1. Hinton, G. E.: Training Products of Experts By Minimizing Contrastive Divergence. *Neural Computation* 14, 1771 (2002).
2. Salakhutdinov, R., Hinton, G.: Deep Boltzmann Machines. *Proceedings of the*

- 12o International Conference on Artificial Intelligence and Statistics (2009).
- 3. Ranzato M., R., Krizhevsky A., Hinton, G.: Factored 3-Way Restricted Boltzmann Machines For Modeling Natural Images. *Proceedings of the 13o International Conference on Artificial Intelligence and Statistics* (2010).
 - 4. Chen Y., Zhang, P., Yu L., Zhang S.: Transient dynamics for sequence processing neural networks: effect of degree distributions, *Phys. Rev. E* **77**, 016110 (2008).
 - 5. Zhang, P., Chen, Y.: Transient dynamics of sparsely connected Hopfield neural networks with arbitrary degree distributions, *Physica A* **387**, 1009 (2008).
 - 6. Reynaga F.: Signal-to-noise analysis of Hopfield neural networks with a formulation of the dynamics in terms of transition probabilities, *Physica A* **388**, 4872 (2009).
 - 7. Mimura, K., Coolen. A.C.C.: Parallel dynamics of disordered Ising spin systems on finitely connected directed random graphs with arbitrary degree distributions, *J. of Phys. A: Math. Theor.* **42**, 415001 (2009).
 - 8. Eissfeller, H., Opper, M.: New method for studying the dynamics of disordered spin systems without finite-size effects, *Phys. Rev. Lett.* **68**, 2094 (1992).