

LA ESTIMACIÓN MARKOVIANA DIRECTA DE PELIGRO SÍSMICO EN SISTEMAS SISMOGÉNICOS COMO UN CASO PARTICULAR DE APRENDIZAJE NEURONAL TIPO PERCEPTRÓN.

Claudia Herrera y F. Alejandro Nava
Dpto. Sismología, División Ciencias de la Tierra
CICESE

Email: cherrera@cicese.mx
Email: fnava@cicese.mx

RESUMEN

Las probabilidades de transición obtenidas de la aplicación al área del Japón del método estadístico markoviano para la determinación de peligro sísmico, presentado por Herrera (2001) y Nava *et al.* (2004), pueden ser obtenidas también del entrenamiento de una red neuronal artificial tipo perceptrón para algunos valores particulares de coeficientes iniciales y constante de amortiguación. La obtención de probabilidades de transición casi idénticas por métodos completamente distintos, permite interpretar el método de estimación original desde el punto de vista de un proceso de aprendizaje e ilustra algunas de sus limitaciones.

Palabras clave: método markoviano, peligro sísmico, perceptrón.

ABSTRACT

The transition probabilities product of the application to the Japan area of the Markovian method for seismic hazard evaluation presented by Herrera (2001) and Nava *et al.* (2004), can also be obtained from the training of a perceptron type artificial neural network for some particular values of the initial coefficients and learning rate. Obtaining almost identical transition probabilities by different methods allows interpreting the original estimation method from a learning process point of view and points out some of its limitations.

Key words: Markovian method, seismic hazard, perceptron.

INTRODUCCIÓN

Dado que actualmente no es posible predecir adecuadamente la ocurrencia de terremotos, las estimaciones de peligro sísmico, son esenciales para disminuir las pérdidas de vidas y los daños económicos causados por éstos. Definimos como *peligro sísmico* la probabilidad de ocurrencia de uno o más sismos en un área geográfica específica durante intervalos de tiempo y magnitud determinados. Ya que muchos especialistas emplean en forma indistinta los términos de *peligro sísmico* y *riesgo sísmico*, es necesario indicar la diferencia entre ellos. Nuestra definición de peligro sísmico es la ya mencionada, mientras que por *riesgo sísmico* se entiende la probabilidad de ocurrencia, en una región y un intervalo de tiempo dados, de algún efecto, generalmente aceleraciones, pérdidas o daños ocasionados por uno o más sismos.

Las estimaciones de peligro sísmico, están frecuentemente basadas en análisis estadísticos de la historia sísmica (catálogo) del área de interés. Dichas estimaciones se basan en modelos de ocurrencia de terremotos; estos modelos van desde procesos sin memoria, como el asociado con la distribución de

Gutenberg-Richter (Gutenberg y Richter, 1944), hasta variantes probabilísticas de modelos basados en la teoría del rebote elástico (Reid, 1910), que son variantes de los modelos de tiempo predecible y de corrimiento predecible (Shimazaki y Nakata, 1980), y de vacancia (gap) sísmica (Fedotov, 1965).

El método presentado en Herrera (2001) y Nava *et al.* (2004), es un método estadístico que explora heurísticamente la posibilidad de modelar la probabilidad de ocurrencia de patrones de sismicidad en un área geográfica que consta de varias regiones sismogénicas como una cadena de Markov; es decir, como un proceso aleatorio con un número finito de estados y memoria de un paso, esto es, condiciona la probabilidad de ocurrencia de un estado al estado inmediato anterior y solamente a éste. Este método, al que llamaremos aquí *directo* y será explicado brevemente más adelante, da resultados muy satisfactorios (aciertos en la evaluación de probabilidades muy por encima de los resultantes de modelos sin memoria y probabilidades despreciables de obtener dichos resultados al azar) y sugiere varias posibles direcciones para futuras investigaciones.

La forma en que se construye la matriz de probabilidades de transición de Markov en el método directo, a partir de la información contenida en el catálogo, puede considerarse como un proceso de aprendizaje. Por lo tanto, consideramos la posibilidad de aplicar otros métodos de aprendizaje al problema de evaluación del peligro sísmico y las redes neuronales eran candidatos naturales para este objetivo. Comenzamos por considerar el tipo más simple de red neuronal, conocido como perceptrón. Esta investigación (aún en proceso) dio un resultado muy interesante que arroja luz sobre algunos aspectos del proceso markoviano y permite utilizar éste como base de referencia para la evaluación de peligro mediante redes neuronales.

Metodología Modelo Directo

La motivación y bases del método directo han sido expuestos en (Herrera, 2001; Nava *et al.*, 2004); aquí haremos un breve resumen, de manera que este artículo pueda entenderse sin necesidad de consultar los trabajos citados.

Consideremos un área geográfica que incluye R regiones sísmogénicas. Dados un catálogo sísmico y un tiempo inicial, durante un intervalo de tiempo específico Δt , el estado de cada región s_r , tiene uno de dos valores: 0 ó 1, que corresponden respectivamente a la ausencia o presencia de sismos con magnitud mayor o igual a un valor umbral M_r . El estado total del sistema es una de las $S = 2^R$ combinaciones posibles de los estados regionales:

$$s = \sum_{r=0}^{R-1} 2^r s_r, \quad (1)$$

y se puede representar en forma binaria por una palabra de bits formada por la concatenación de los estados regionales, representación que permite visualizar fácilmente cuáles regiones son activas y cuáles no. El estado total del sistema va de hasta .

A lo largo del tiempo que abarca el catálogo, se determina el estado del sistema para cada intervalo y, de la sucesión de estados, se forma una matriz de transiciones cuyos elementos corresponden al número de veces que el sistema pasó del estado , observado durante un intervalo número n , al estado. Tras procesar todos los datos del catálogo, se obtiene la matriz de probabilidades de transición con elementos

$$p_{ij} = \frac{\theta_{ij}}{\xi_i}, \quad (2)$$

donde es el número total de transiciones originadas desde el estado . Con base en el estado del sistema observado durante un intervalo específico , esta matriz da estimaciones de peligro para los posibles estados del sistema para el

siguiente Δt .

Modelo Neuronal (Perceptrón)

Las redes neuronales artificiales (RNA) son máquinas computacionales de aprendizaje, que funcionan emulando de alguna manera el comportamiento de las neuronas biológicas (e.g. Hertz *et al.*, 1991; Anderson, 1988). Constan de una entrada, una red de neuronas, generalmente interconectadas entre sí, y una salida producida por éstas con base en dicha entrada. Cada neurona produce una salida determinada por la información que recibe, ya sea directamente de la entrada o a través de otras neuronas. La RNA más sencilla es la tipo *perceptrón* (Rosenblatt, 1958, 1962; Nilsson, 1965; Minsky y Papert 1969) que consta de una serie de entradas $[s] = \{s_i; i = 1, 2, \dots, n\}$, una sola capa de neuronas y un conjunto de salidas $[o] = \{o_j; j = 1, 2, \dots, n\}$ (Fig.1); se caracteriza por ser capaz de resolver problemas linealmente separables.

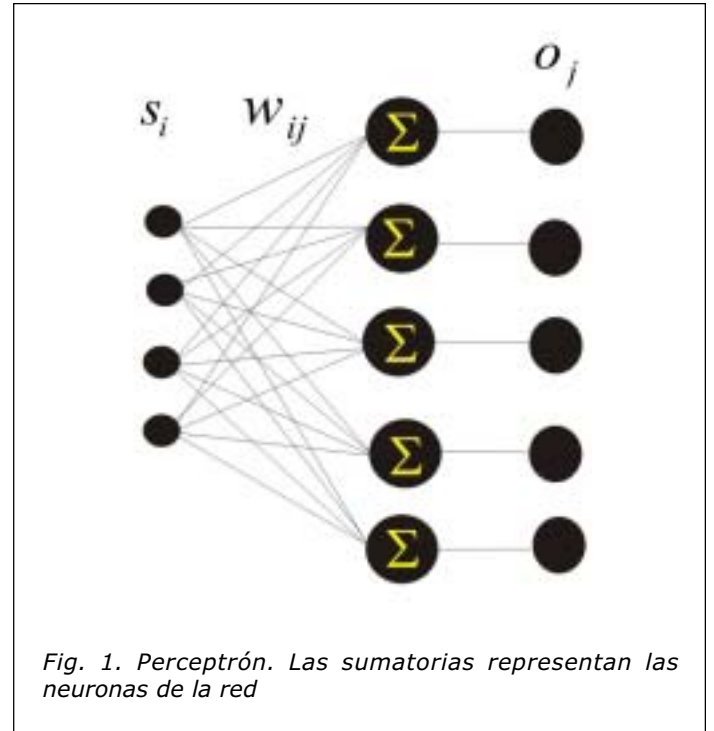


Fig. 1. Perceptrón. Las sumatorias representan las neuronas de la red

A partir de la entrada $[s]$, la neurona j produce la salida

$$o_j = f \left(\sum_{i=1}^n w_{ij} s_i \right), \quad (3)$$

donde las w_{ij} son pesos asignados a las líneas de comunicación de cada entrada a cada neurona, y f es una *función de activación* (usualmente igual para todas las neuronas) que determina cómo es la salida. La función de

activación es escogida de acuerdo con las especificaciones del problema que la neurona tenga que resolver y, en las redes clásicas que emulan los procesos sinápticos de las neuronas, es a menudo una función binaria que determina si la neurona es activada (y propaga la información) o permanece inhibida (salida nula).

Los pesos w_{ij} se eligen de manera que la red produzca la salida deseada $[d]$ para cada entrada $[s]$ dada, pero estos pesos generalmente no se conocen *a priori*. La característica más importante de las RNA es que pueden *aprender*, esto es, ir ajustando los valores iniciales de los pesos, con base en el error cometido. Rosenblatt (1958, 1962) demostró que, presentando a la red un conjunto de patrones de entrada-salida y empleando una simple regla de corrección de los pesos con base en los residuales de las respuestas obtenidas, se puede obtener los parámetros que configuran la respuesta óptima de la red,, en un número finito de pasos. Este procedimiento determina la regla de aprendizaje de la red, conocida como regla de **Hebb**.

$$w_{ij}^{k+1} = w_{ij}^k + \Delta w_{ij}^k$$

$$\Delta w_{ij}^k = \eta (d_j^k - o_j^k) s_i^k, \quad (4)$$

donde el parámetro de amortiguación η , llamado *razón de aprendizaje* o factor de relajación, controla la rapidez con la que son modificados los pesos durante el entrenamiento y toma usualmente valores entre 0 y 1. En algunos casos η puede ser constante; en otros es reducida conforme progresa el entrenamiento hasta alcanzar la estabilidad.

Las entradas y salidas deseadas del perceptrón utilizado aquí corresponden a la historia de estados observada, a lo largo de todo el catálogo, del sistema. Cada par de estados sucesivos (i, j) corresponde respectivamente, a la entrada y a la salida que deseamos proporcione la red. Estos estados están representados por palabras binarias cuyos elementos son todos ceros excepto los elementos i y j , respectivamente, los cuales valen uno.

La función de activación empleada es la función $f(x) = x$; es decir, solamente transmite el promedio pesado de las entradas que recibe cada neurona.

EVALUACIÓN DE DESEMPEÑO

Evaluar el desempeño de diferentes modelos de estimación de peligro, no es inmediato. Una medida natural es la *probabilidad observada* promedio $\langle \hat{p} \rangle$, que es el promedio de la probabilidad de cada transición estimada antes de la ocurrencia de ésta. Esta medida puede interpretarse con respecto a los valores correspondientes a modelos considerados como hipótesis nulas y, sobre todo, con la probabilidad uniforme $u = S^{-1}$.

Sin embargo esta medida no toma en cuenta que

el principal objetivo de las estimaciones de peligro sísmico es pronosticar la ocurrencia de terremotos, para que la sociedad pueda responder adecuadamente a estos fenómenos. Una estimación útil también debe tomar en cuenta factores tales como número de éxitos n_{xm} , donde un éxito se define como la ocurrencia de un caso con probabilidad mayor que un umbral dado $p_x = f_x u$, expresado como f_x (*factor de éxito*) veces la probabilidad uniforme, pronósticos múltiples m_i (necesariamente contradictorios), falsas alarmas n_{fa} , error de pronóstico e_r , definido como el número de regiones cuya actividad fue pronosticada erróneamente, etc.

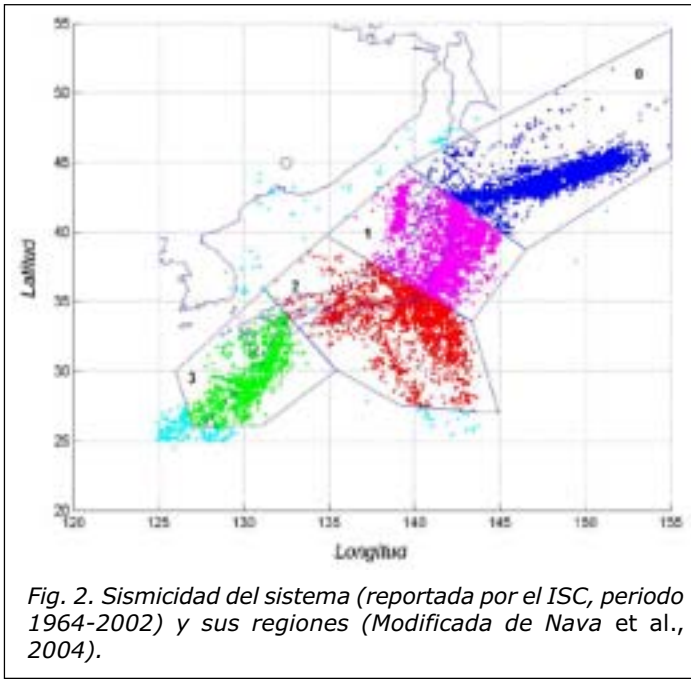
Elegir el mejor método o parámetro f_x no es sencillo, porque usualmente hay una serie de intercambios entre factores deseados y factores no deseados. Por ejemplo un valor bajo de f_x resulta en un gran número de éxitos (deseable), pero también un valor grande de falsas alarmas y error regional (ambos no deseables); mientras, un valor grande de f_x minimiza el número de falsas alarmas, pero también reduce el número de pronósticos y de éxitos.

Por todo ello, una inspección directa de los valores resultantes de las diferentes medidas de desempeño no es práctica para elegir el parámetro óptimo para un determinado modelo, y puede provocar confusión cuando tratamos de comparar distintos métodos. Por lo tanto, para evaluar el desempeño decidimos utilizar discriminantes, i.e. funciones matemáticas cuyo valor incrementa con los resultados deseados y disminuya con los no deseados; con diferentes pesos a cada factor, acordes con sus relativos tamaño e importancia. Esta ponderación por importancia refleja el consenso de deseabilidad para los diferentes factores.

No hay una forma determinada para los determinantes; nosotros usamos varios determinantes distintos, pero ilustraremos los resultados presentados aquí con uno solo:

$$d = 3 \left[\left(\frac{0.2 n_{pred} + 8.0 n_{xm} - 1.0 n_{fa}}{n_t} \right) \left(\frac{2.5 \langle p_{\hat{p}} \rangle}{\langle p \rangle [0.22 + |1 - \langle m \rangle|]} \right) + 0.001 f_x \right], \quad (5)$$

donde n_t es el número de transiciones.



RESULTADOS

Los modelos descritos en la sección anterior fueron aplicados al área de Japón, empleando los datos del catálogo sísmico mundial reportados por el *Internacional Seismological Centre*. Esta región fue elegida por sus condiciones tectónicas apropiadas y por su extenso catálogo sísmico.

Después de varias pruebas fueron elegidas 4 regiones: Islas Kuriles, Japón Central, Japón SE e Islas Ryukyu, mostradas en la Fig. 2. Los posibles estados de las 4 regiones definen 16 posibles estados del sistema (Tabla 1). La elección del intervalo temporal Δt y de la magnitud umbral M_r , fue hecha empíricamente, a manera de obtener, dentro de las limitaciones del catálogo, óptimas definición y cobertura para el método directo (Nava et al., 2004).

Tras procesar todos los datos del catálogo, para un total de 36 años, $\Delta t = 0.10$ años y $M_r = 5.5$, obtuvimos 364 transiciones. Con base en esta historia de transiciones se calculó la matriz de probabilidades de transición de Markov (con 256 elementos) para el método directo según la ecuación (2). Aún despreciando el posible error en la determinación de los estados (Nava et al., 2004), el número de transiciones no es lo suficientemente grande, comparado con el número de incógnitas, para obtener estimaciones completamente confiables; sin embargo las estimaciones presentaron una estabilidad razonable y, lo más importante, los resultados de la aplicación de probabilidades estimadas al análisis del catálogo fueron mucho mejores que los obtenidos para hipótesis cero (modelos uniforme y poissoniano). Se obtuvo una probabilidad observada $\hat{p} = .159$ (más del doble de la uniformemente aleatoria);

y, para el factor de éxito óptimo $f_x = 3.3$: $n_{pred} = 133$, $\langle e_r \rangle = 1.67$, $n_{xm} = 33$, $n_{fa} = 117$ y $\langle m_i \rangle = 1.22$, que resultan en $d = 1.231$; la probabilidad binaria de obtener el número de aciertos al azar es $\sim 10^{-12}$. Los correspondientes valores óptimos para el modelo poissoniano son $d = 0.2571$ para $f_x = 2.6$ y para el modelo uniforme $d = -0.011$ para $f_x = 0.99$.

Tabla 1. Posibles estados del sistema.

		Regiones			
Estados		3	2	1	0
0	=	0	0	0	0
1	=	0	0	0	1
2	=	0	0	1	0
3	=	0	0	1	1
4	=	0	1	0	0
5	=	0	1	0	1
6	=	0	1	1	0
7	=	0	1	1	1
8	=	1	0	0	0
9	=	1	0	0	1
10	=	1	0	1	0
11	=	1	0	1	1
12	=	1	1	0	0
13	=	1	1	0	1
14	=	1	1	1	0
15	=	1	1	1	1

La misma historia de transiciones fue utilizada para entrenar al perceptrón, según la ecuación (4), para varios valores iniciales de los pesos y diferentes valores de amortiguamiento.

Al probar diferentes valores iniciales de los pesos, así como valores varios de amortiguamiento y normalizando los renglones de la matriz W , a manera de obtener una matriz estocástica susceptible de ser interpretada de manera probabilística, se obtuvo el siguiente resultado interesante: para el caso particular de función de activación lineal, pesos iniciales muy pequeños o nulos y fuerte amortiguamiento, la matriz de pesos resultante del aprendizaje del perceptrón es casi igual a la de probabilidades de transición del método directo. La máxima similitud se obtiene para valores iniciales de $w_{ij} = 0 \quad \forall i, j$ y $\eta = 0.002747 (=364^{-1})$, con máxima diferencia absoluta entre valores correspondientes $\delta_{\max} = \text{Max}(|w_{ij} - p_{ij}|; i, j = 0, \dots, 15) = 0.005$

y diferencia absoluta promedio

$$< \delta > = \frac{1}{16^2} \sum_{i,j=0}^{15} |w_{ij} - p_{ij}| = 0.0006. \text{ La similitud entre}$$

los resultados indica que la misma estimación probabilística puede obtenerse por dos métodos con filosofías diferentes; dado que el tratamiento hecho con el perceptrón de activación lineal equivale a realizar una estimación de mínimos cuadrados (Bishop, 1995), resulta particularmente interesante que el perceptrón dé valores similares a los de un modelo que simplemente contabiliza las ocurrencias.

Naturalmente, el desempeño de este perceptrón es idéntico al del método directo para pronosticar el estado de sismicidad del sistema. Inmediatamente nos preguntamos si esta combinación de valores iniciales y constante de amortiguamiento es la que resulta en resultados óptimos para el perceptrón. La respuesta es un categórico no; otros modelos de perceptrón con diferentes combinaciones de valores iniciales y amortiguamiento, así como de funciones de activación no lineales, que serán reportados en detalle en otro trabajo, dan resultados bastante mejores; por ejemplo $w_{ij} = 0.0001 \quad \forall i, j$ inicial y $\eta = 0.02$ resultan en $d = 1.866$ para el perceptrón simple considerado aquí, mientras que otras variantes («cruzadas») del método directo y del perceptrón (no discutidas aquí) alcanzan valores $d = 39.667$ y $d = 45.085$, respectivamente, ambas para $f_x = 5.3$.

CONCLUSIONES

La obtención de probabilidades de transición casi idénticas por métodos distintos, indica que, efectivamente, el método directo puede considerarse como un método de aprendizaje y un caso particular del método de perceptrones.

Sin embargo, la comparación con otros modelos de perceptrón indica que, si bien el método directo da resultados bastante buenos, no es el método óptimo para aprovechar la información contenida en los catálogos para la evaluación de peligro sísmico en sistemas sísmogénicos.

AGRADECIMIENTOS

Agradecemos cumplidamente las atinadas observaciones y sugerencias de dos árbitros anónimos. También mucho apreciamos los comentarios de Ewa Glowacka, Luis Munguía, Cecilio Rebollar y Javier Helenes.

REFERENCIAS

- Anderson, J.A. and Rosenfeld, E., 1988. Neurocomputing: Foundations of research, M.I.T. Press, Cambridge, MA., 125 pp.
- Bishop, C.M., 1995. Neural networks for pattern recognition, Clarendon Press Oxford, UK, 482pp.
- Fedotov, S., 1965. Regularities of the distribution of strong earthquakes in Kamchatka, the Kuril Islands, and northeast Japan, Trudy Isnt. Fiz. Zemli. Acad. Nauk. SSSR. 36: 66-94.
- Gutenberg, B. and C. Richter, 1944. Frequency of earthquakes in California. Bulletin of the Seismological Society of America, 34: 185-188.
- Herrera, C., 2001. Determinación de peligro sísmico mediante cadenas de Markov. Tesis de Maestría, División de Ciencias de la Tierra, CICESE. 86 pp.
- Hertz J., Aanders Krogh and Palmer, R.G., 1991. Introduction to the theory of neural computation, Addison Wesley, 327 pp.
- Minsky, M.L., and Papert, S., 1969. Perceptrons: An introduction to computational geometry, M.I.T. Press, Cambridge, MA, 275 pp.
- Nava, F., Herrera, C., Frez, J., and Glowacka, E., 2004. Seismic hazard evaluation using Markov chains; Application to the Japan area. Sometido a PAGEOPH.
- Nilsson, N., 1965. Learning machines, McGraw-Hill, New York, 138 pp.
- Reid, H. F., 1910. The mechanism of the earthquake, The California earthquake of April 18, 1906. Report of the Research Senatorial Commission. Carnegie Institution, Washington D.C. 2:16-18.
- Rosenblatt, F., 1958. The perceptron: a probabilistic model for information storage and organization in the brain. Psychological Review, 65:386-408. Reprinted in: Anderson & Rosenfeld (1988), p.92-114.
- Rosenblatt, F., 1962. Principles of neurodynamics. New York: Spartan, 215 pp.
- Shimazaki, K. and T. Nakata, 1980. Time predictable recurrence model for large earthquakes, Geophysical Research Letters, 7: 279-282.

Recepción del manuscrito: 4 de febrero, 2005

Recepción del manuscrito corregido: 28 de febrero, 2005

Aceptación del manuscrito: 3 de marzo, 2005