REDES BAYESIANAS Y REDES NEURONALES COMO MODELOS DEL APRENDIZAJE CAUSAL.

¿Cómo pueden aprender un conjunto de neuronas conectadas?

Francisco Javier Rodríguez Mañanes.

Profesor Tutor: José Antonio González del Campo.

Máster en Neurociencias. Instituto de Neurociencias de Castilla Y León (INCYL).

INDICE.

INTRODUCCIÓN.

Sigue siendo algo desconocido cómo el cerebro es capaz, mediante mecanismos de procesamiento de la información, de realizar aprendizajes referidos a la causalidad que muestra el mundo que nos rodea. Tenemos, por un lado, bastantes conocimientos sobre cómo funciona el mundo neuronal. Por otro, los conocimientos empíricos sobre el aprendizaje, en concreto el aprendizaje causal, están siendo cada vez mejor modelizados mediante algoritmos que se implementan en forma de redes, ya sean bayesianas o neuronales, y nos permiten tener pistas sobre los fenómenos matemáticos claves necesarios para que ese procesamiento de información tenga lugar.

No existe, sin embargo, un modelo general (framework) básico que englobe los diferentes algoritmos que se utilizan actualmente en la modelización del aprendizaje causal, que resultan parciales y, en ocasiones, aparentemente diferentes. Sin embargo, en la naturaleza sucede que esos algoritmos son implementados realmente por conjuntos de neuronas conectadas, o lo que hemos dado en denominar en este trabajo, redes neuronales biológicas. ¿Cómo es posible que una red neuronal biológica implemente los fenómenos matemáticos claves que permiten el procesamiento de la información necesaria para el aprendizaje causal?

La pregunta presentada es la motivación y la guía fundamental del presente trabajo. Presentamos pasos necesarios para comenzar la búsqueda de la respuesta, y también varios resultados que muestran que esos pasos se pueden ir dando considerando los conocimientos de la ciencia actual.

FUNCIONAMIENTO COGNITIVO Y FUNCIONAMIENTO NEURONAL EN EL APRENDIZAJE DE LAS RELACIONES CAUSALES: DOS MODOS DE APROXIMACIÓN A UN MISMO PROBLEMA.

Es necesario aclarar primeramente un aspecto sobre terminología referida a distintos conceptos, que se refieren a ámbitos distintos. Se tiene que establecer una distinción entre lo que podemos denominar redes neuronales conceptuales y redes neuronales biológicas.

Con el primer término nos referimos a una red neuronal cuyos nodos están referidos a conceptos cognitivos, y los pesos de conexión entre esos nodos establecen de alguna manera determinada, por diferentes algoritmos, las relaciones de implicación entre esos conceptos. En general, los modelos de redes que se utilizan en la modelización de los procesos de aprendizaje son de este tipo.

Con el término redes neuronales biológicas nos estamos refiriendo a la modelización de redes de neuronas que simulan el funcionamiento real de las neuronas biológicas.

En ocasiones, se mezclan conceptos utilizados en cada uno de los ámbitos, y es necesario ser muy cuidadoso para ver si esa utilización está justificada o es correcta.

Sin embargo, queremos llamar la atención sobre un punto que consideramos importante: en la neurociencia actual el paradigma actualmente aceptado es que los elementos del funcionamiento cognitivo son representaciones mentales, que son a su vez patrones de activación neuronales. Eso conlleva que, en última instancia, el funcionamiento cognitivo se lleva a cabo por un determinado funcionamiento neuronal, interpretado como patrones de activación. Es posible pensar entonces, que el funcionamiento del sistema a nivel de microestructura (neuronal y de conexiones entre las neuronas) pueda transferirse, bajo transformaciones, al nivel de macroestructura o cognitivo, y que ambos niveles de funcionamiento puedan compartir características comunes. Eso podría explicar el hecho, por lo menos llamativo, de que ambos niveles de funcionamiento se estén modelizando con el mismo tipo de herramientas, esto es, con redes neuronales o con redes bayesianas.

MODELOS MATEMÁTICOS UTILIZADOS ACTUALMENTE EN LA MODELIZACIÓN DEL APRENDIZAJE CAUSAL: UNA INTRODUCCIÓN.

En la modelización del aprendizaje causal se emplean diferentes herramientas matemáticas. Cada una de ellas presenta características que las diferencian de las demás.

Muchos autores emplean varias de ellas relacionándolas, en la explicación del aprendizaje causal. Así, por ejemplo, Cheng construye modelos de redes bayesianas, pero relaciona los resultados obtenidos con la relación de contingencia.

Las implementaciones matemáticas que se utilizan actualmente para trabajar en aprendizaje causal son las siguientes:

Relación de Contingencia.

Establece la diferencia de probabilidad de ocurrencia de un suceso ante la presencia o la ausencia de una posible causa potencial. Es, por lo tanto, una medida de la correlación y/o causalidad entre una serie de sucesos.

Modelo de Rescorla-Wagner.

Posiblemente, el modelo de Rescorla-Wagner pueda ser el modelo más relevante históricamente en la modelización de los fenómenos del aprendizaje. Eso nos puede llevar a pensar que en esencia está implementando fenómenos que pueden estar relacionados con procesos básicos del funcionamiento neuronal. Por ello, el modelo de Rescorla-Wagner puede volver a surgir en otros modelos más actuales. Por otro lado, algunos autores como Pearce construyen redes neuronales en las que los algoritmos que modifican los pesos de conexión entre los distintos nodos lo hacen a la manera del modelo de Rescorla-Wagner, lo que ejemplifica su vigencia.

Redes Bayesianas.

Constituyen uno de los modelos matemáticos que más se emplean en la explicación del aprendizaje causal. Una de sus características principales es que evalúan, de alguna forma, la probabilidad de todas las posibilidades de los sucesos.

Redes Neuronales.

Las redes neuronales constituyen unas de las herramientas más utilizadas en resolución de problemas, tareas de diagnóstico, predicción, etc, en el campo de la Inteligencia Artificial.

Un resultado interesante que hemos obtenido y que expondremos con más detalle más adelante en este trabajo, es que el modelo de Rescorla-Wagner y los algoritmos utilizados en Redes Neuronales no están tan alejados, y podemos demostrar que bajo determinadas circunstancias presentan grandes similitudes (por ejemplo, cuando se trabaja con Redes Neuronales comparables)

CUESTIONES FUNDAMENTALES.

¿Cuáles son los fenómenos matemáticos claves que implementan estos modelos?

El objetivo propuesto con respecto a los modelos matemáticos que se están utilizando actualmente es doble:

Por un lado, necesitamos construir implementaciones informáticas que modelicen dichos algoritmos, de forma que podamos ponerlos a prueba y ver cómo responden ante determinados fenómenos conocidos. En este sentido, a lo largo del presente trabajo se presentarán los resultados de simulaciones realizadas implementando el modelo de Rescorla-Wagner y el modelo de Redes Neuronales con el objetivo de comparar cómo responden ante determinados fenómenos del aprendizaje.

Por otro lado, necesitamos comprender cuáles son los fenómenos matemáticos clave que implementan dichos modelos, y las posibilidades explicativas que poseen potencialmente en función de su construcción. En este sentido, a lo largo del trabajo se irán exponiendo algunos de estos resultados clave.

¿Podemos conseguir un framework básico?

Después de revisar la bibliografía existente sobre el tema en cuestión, y algunos de los artículos actuales más relevantes, vemos que no existe un modelo general que describa el aprendizaje causal. Existen más bien modelos parciales, que describen o modelizan determinados fenómenos observados empíricamente. Cada modelo tiene además, una serie determinada de parámetros y condiciones que no tienen por qué ser comunes o compartidos con el resto de los modelos. No existe, por lo tanto, un modelo teórico globalizador.

La necesidad de este modelo general es compartida por la mayoría de los autores, y esta necesidad es expresada ya tanto por los autores del grupo PDP, como en referencias actuales sobre el tema.

Uno de nuestros objetivos básicos es ir dando pasos en el descubrimiento de las características y, en lo posible, la construcción, de modelos que generalicen los actuales. En este sentido se propondrá un algoritmo que integra el modelo de Rescorla-Wagner y la filosofía de funcionamiento de las Redes Neuronales.

Construcción de redes neuronales biológicas.

Los modelos de redes, tanto neuronales como bayesianas, tratadas en este trabajo, son fundamentalmente modelos de redes conceptuales, aunque varios de sus aspectos parten de las ideas del funcionamiento neuronal.

Otro de los objetivos que también nos planteamos es la construcción de redes neuronales biológicas. En este sentido, buscamos redes neuronales que simulen el funcionamiento neuronal y que sean capaces de asociar determinados patrones de activación como procedimiento de aprendizaje.

¿QUÉ ES EL APRENDIZAJE CAUSAL?

El aprendizaje causal se puede definir como la aprehensión, por parte de un sujeto, de las relaciones causales que observa a su alrededor.

El aprendizaje causal se está mostrando como un concepto globalizador o integrador de diferentes conceptos relacionados con el aprendizaje.

El aprendizaje causal es un concepto que está implicado en diversos ámbitos, de los que expondremos alguno.

Tareas de diagnóstico.

En diferentes profesiones (médicos, ingenieros,…) es de capital importancia el establecimiento, a partir de la observación de unos determinados hechos, de las causas que están posiblemente dando lugar a esos hechos.

En ese sentido, es bastante significativo que las herramientas de las redes neuronales se estén utilizando en muy diversos campos en los que básicamente se tiene que aprender cuándo un conjunto de características o parámetros puedan indicar la ocurrencia o no ocurrencia de un suceso. Podemos señalar, como ejemplo, la utilización de redes neuronales en tareas de diagnóstico médico, en dos tesis doctorales realizadas en Salamanca en 2001.

Método Científico.

Podemos decir que el método científico es el método por excelencia que se ocupa de investigar las relaciones causales entre los sucesos. Cuando se investiga la influencia de un determinado gen sobre algún suceso, se está intentando encontrar las relaciones de causalidad entre ambos.

En este sentido se puede pensar en la labor de investigación como en un establecimiento de relaciones causales entre conceptos de alto nivel.

En este sentido es además plausible, por lo tanto, la siguiente pregunta: ¿el establecimiento de relaciones causales entre conceptos de bajo nivel y el establecimiento de relaciones causales entre conceptos de alto nivel siguen las mismas leyes?

Situaciones habituales.

El aprendizaje de las relaciones causales del mundo que rodea a las personas es algo habitual.

RELACIÓN DE CONTINGENCIA.

La relación de contingencia es una medida que intenta expresar cuánto es más probable que suceda un efecto en función de la presencia o no de una causa potencial.

Más exactamente, la contingencia es la diferencia de probabilidad de ocurrencia de un suceso (efecto) cuando la potencial causa está presente o no. Esto es,

$$\Delta P = \frac{a}{a+b} - \frac{c}{c+d}$$

donde E simboliza el efecto, y A la causa potencial.

Para ayudar en el cálculo de este parámetro, se suelen utilizar tablas de contingencia, de la forma

	E	$\overline{E}$
C	a	b
$\overline{C}$	c	d

La relación de contingencia es un parámetro que se usa en relación con los modelos del aprendizaje causal, ya que se puede estimar cuál es la relación de contingencia predicha por esos modelos (ver, por ejemplo, Cheng).

MODELO DE RESCORLA-WAGNER.

Como hemos señalado anteriormente, el modelo de Rescorla-Wagner ha tenido una gran influencia histórica a la hora de modelizar diferentes fenómenos del aprendizaje observados empíricamente. Esta gran importancia sería ya de por sí justificación suficiente para su estudio.

Sin embargo, podemos pensar que el modelo de Rescorla-Wagner sigue aún vigente y proporcionando resultados, por lo menos por dos hechos:

Por un lado, diferentes autores, como Pearce, implementan el modelo de Rescorla-Wagner como algoritmo de modificación de los pesos de conexiones en sus redes neuronales.

Por otro lado, como veremos más adelante, el modelo de Rescorla-Wagner y los modelos de Redes Neuronales presentan grandes similitudes.

El modelo de Rescorla-Wagner expresa cómo varían las fuerzas de asociación entre diferentes estímulos y un estímulo resultado o incondicionado. Esta variación viene expresada en forma de un conjunto de ecuaciones en incrementos, de la forma

$$
\begin{aligned}
&V_A^{n+1} = V_A^n + \alpha_A \beta\left(\lambda - \Sigma V_i^n\right)\\
&V_B^{n+1} = V_B^n + \alpha_B \beta\left(\lambda - \Sigma V_i^n\right)\\
&\vdots\\
&V_X^{n+1} = V_X^n + \alpha_X \beta\left(\lambda - \Sigma V_i^n\right)
\end{aligned}
$$

Con respecto al modelo de Rescorla-Wagner, podemos señalar dos resultados que hemos obtenido y que no se suelen explicitar habitualmente:

Se pueden obtener los valores a los que tienden las diferentes fuerzas asociativas cuando realizamos ensayos en los que se presentan los diferentes estímulos condicionados a la vez, y asociados con un estímulo incondicionado, sin necesidad de realizar las sucesivas iteraciones de las ecuaciones.

Por ejemplo, para el caso concreto de dos estímulos, con parámetros de aprendizaje α_A y α_B respectivamente, tendremos que

$$
\lim_{n\to\infty} V_A^n = \frac{\alpha_A}{\alpha_A + \alpha_B} \qquad \lim_{n\to\infty} V_A^n = \frac{\alpha_B}{\alpha_A + \alpha_B}
$$

con lo que habremos obtenido los valores asintóticos de las fuerzas asociativas sin tener que resolver las ecuaciones propias del modelo.

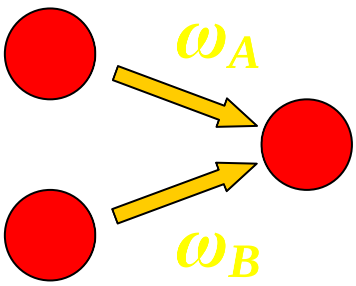
Por otro lado, otro resultado que hemos obtenido es que se puede obtener una ecuación continua que exprese la variación de la fuerza asociativa:

$$f(x) = (1-(1-\alpha))^x$$

donde x es la variable que en principio representa el número de ensayo. De alguna forma, es el primer paso a considerar que en realidad, las variaciones en la fuerza asociativa entre representaciones mentales que establece el sujeto vienen determinadas por diferentes procesos de neuroplasticidad. Es difícil pensar que esos procesos evolucionen de forma discreta, siendo mucho más natural considerar que su evolución se realiza de forma continua.

Podríamos considerar una red bayesiana como un grafo dirigido, en el cual los diferentes nodos se conectan mediante relaciones de implicación que poseen una determinada direccionalidad, y con un algoritmo de modificación de los pesos de conexión basado en el Teorema de Bayes.

Para simplificar, usaremos una red bayesiana muy simple, similar a la utilizada por Cheng como modelo de aprendizaje causal entre un background y una posible causa potencial y un efecto.



Los modelos de redes bayesianas funcionan con dos ideas básicas.

Pero antes necesitamos introducir un concepto preliminar:

Hablaremos de una "máquina" refiriéndonos a redes bayesianas como a un conjunto determinado de sus pesos de conexión. Esto es,

$$\Phi_i \to \left(\omega_{A,i}, \omega_{B,i}\right)$$

Con una máquina determinada, podríamos introducir los valores iniciales y la red nos proporcionaría una salida.

Las dos ideas fundamentales a las que nos referíamos anteriormente son:

La red sería todo el conjunto de posibles máquinas, cada una con su respectiva probabilidad.

Esto es, una red bayesiana, a diferencia de una red neuronal, no es una única configuración de pesos de conexiones, sino el conjunto de todas las posibles configuraciones o máquinas, asociada cada una con una cierta probabilidad.
En un modelo continuo, quedaría, por ejemplo:

$$\Phi = p_1\Phi_1 + p_2\Phi_2 + \ldots + p_{10}\Phi_{10}$$

Aprender significa modificar las probabilidades de cada máquina.

Esto es, no se modifican en sí mismos los valores de los pesos de conexión entre los nodos. Las diferentes posibilidades siguen existiendo. Lo que se modifican son las diferentes probabilidades de cada máquina.

¿Cómo se modifican estas probabilidades a lo largo del entrenamiento de la red, según se van presentando los diferentes ensayos? La forma de modificar las diferentes probabilidades viene determinada por el Teorema de Bayes:

$$p_i^{n+1} = \frac{p_i^n \Phi_i}{\sum_j p_j^n \Phi_j}$$

Estas diferentes conceptualizaciones entre las redes bayesianas y las redes neuronales (existencia de diferentes máquinas y aprendizaje como variación de las probabilidades de cada máquina en el caso de redes bayesianas y existencia de una máquina concreta y aprendizaje como variación de los parámetros de esa máquina en el caso de redes neuronales) plantean un problema a la hora de buscar un framework básico que englobe a ambas.

De hecho, el funcionamiento de una red bayesiana se puede asociar a una red neuronal determinada. Pero el aprendizaje ya no se puede asociar, porque se perdería información: diferentes redes bayesianas se podrían asociar con la misma red neuronal, y sin embargo aprenderían de forma diferente. Cómo resolver este punto clave podría llevarnos a construir un modelo globalizador entre ambos tipos de redes.

Un último apunte con respecto al proceso implementado por las redes bayesianas es su gran parecido al formalismo matemático de la mecánica cuántica, en la que el estado de un sistema se puede dar como

$$\Psi = a_1\Psi_1 + a_2\Psi_2 + ... + a_n\Psi_n$$

Un sistema en mecánica cuántica opera como un conjunto de estados posibles, asociados cada uno a una determinada probabilidad, y esa es la misma idea subyacente a las redes bayesianas.

Explorar las posibilidades de estas implicaciones excede el propósito de este trabajo, si bien es cierto que diversos autores han propuesto diferentes teorías en las que el funcionamiento cerebral viene explicado con este tipo de conceptos. Y, por otro lado, está surgiendo actualmente un nuevo paradigma en el tratamiento de la información, que es la computación cuántica. Estos hechos implican la necesidad de estar atentos y conocer los diferentes desarrollos que se vayan produciendo en estas áreas de conocimiento.

Una red neuronal es un grafo dirigido, formado por un conjunto de nodos relacionados direccionalmente por relaciones de influencia parametrizadas por una serie de pesos de conexión. Generalmente se utilizan redes con dos o tres capas de nodos, diciendo en este último caso que existe una capa de nodos oculta.

Decimos que una red neuronal aprende cuando modifica sus pesos de conexión para ajustar su salida en función de una serie de datos de entrada y salida determinados.

Con respecto a las redes neuronales vamos a señalar una serie de ideas que son clave para entender qué se está implementando realmente cuando se trabaja con redes neuronales.

Una red neuronal ajusta una superficie n-dimensional a un conjunto de puntos en un espacio (n+1)-dimensional.

Este conjunto de puntos es precisamente el conjunto de entrenamiento que se proporciona a la red en su proceso de aprendizaje.

¿Cómo ajusta una red neuronal su superficie asociada al conjunto de puntos?

Para que una red neuronal ajuste su superficie asociada a un conjunto de puntos de entrenamiento se utiliza un método de mínimos cuadrados.

Esto es, la red neuronal varía sus pesos de conexión para conseguir que la suma de las distancias al cuadrado entre su valor de salida y la salida esperada sea mínima. Esta es en realidad la misma idea subyacente en un procedimiento de regresión lineal.

Con respecto a estas dos ideas básicas, señalaremos algunos aspectos que es importante tener en cuenta a la hora de trabajar con redes neuronales.

Problema de la toma de datos.

En una red neuronal, queremos obtener una salida adecuada a un conjunto de parámetros de entrada. Esos parámetros de entrada son un conjunto de valores

$$\left(x_1, x_2, \ldots, x_n\right)$$

que podrán tomar diferentes valores pertenecientes a lo que podremos llamar un espacio de posibilidades. Lo correcto sería que el conjunto de entrenamiento de la red estuviera constituido por un conjunto de parámetros de entrada representativo del sistema al que se está aplicando. Si no se hiciera así, y se entrenara un conjunto no representativo de todo el espacio de posibilidades de esos parámetros, podríamos estar dejando sin presentar regiones del espacio en que el sistema podría comportarse de otra forma, por lo que la red no podría aprender esas variaciones. Al presentar en ese caso a la red un conjunto test que perteneciera a ese espacio no presentado, la salida de la red no tiene por qué ser correcta.

Capacidad de la red para implementar un suceso.

Ya hemos comentado que una red neuronal implementa una superficie n-dimensional a un conjunto de puntos en un espacio (n+1)-dimensional. Las posibles superficies que puede implementar la red dependen de las funciones que se ejecuten en sus nodos. Eso quiere decir que una red determinada no tiene por qué ser capaz de implementar cualquier forma de superficie. Este hecho presenta una cuestión a tener muy en cuenta cuando se quieran utilizar redes neuronales para simular procesos de aprendizaje (o cualquier proceso en general): si el sistema a estudiar o a modelizar presenta una superficie que la red no puede implementar, o acercarse a ella de alguna forma, la red neuronal no funcionará, en el sentido de que no será capaz de dar respuestas correctas en todos los valores de entrada posibles.

Con respecto a esta forma de pensar en la actuación de las redes neuronales, señalaremos dos aspectos relevantes para el tema que nos ocupa:

Una red neuronal sin funciones intermedias sólo puede implementar planos.

Esto es así porque la salida de la red siempre va a ser una combinación lineal de combinaciones lineales de los parámetros de entrada, que vuelve a ser una combinación lineal. Y una combinación lineal es la ecuación de un plano.
Esto nos lleva a las siguientes conclusiones:

Sería inútil, por lo tanto, añadir capas ocultas.

Si estamos trabajando con un modelo de red neuronal sin funciones en los nodos (o dicho de otra forma, implementando la función identidad), y no se ajusta a los resultados que buscamos, la solución no podrá venir por el hecho de añadir más capas ocultas a la red, por muchas que éstas sean o por muchos nodos que contengan. La red seguirá implementando planos.

Una red sin funciones intermedias no podrá resolver un aprendizaje de configuraciones.

Si definimos un aprendizaje de configuraciones como aquél en el que la no presentación de los estímulos no dé lugar a la respuesta, mientras que la presentación del estímulo A sí, la del estímulo B también, pero la presentación conjunta de AB no dé lugar a la respuesta, un plano no podrá pasar por esos cuatro puntos a la vez, y necesitaríamos una red que implementara otro tipo de superficie.

Como ejemplo de lo que decimos, podemos citar el modelo de Pearce. Este autor construye una red neuronal con nodos internos. La capacidad de esa red para resolver un aprendizaje de configuraciones viene determinada no sólo por la existencia de esos nodos intermedios, sino porque se implementa en ellos una función cuadrática que hace que la superficie implementada sea curva, y se pueda ajustar a los puntos anteriormente dichos.

Cuando la función implementada en los nodos es una función escalón o una sigmoidea aguda, la red delimita regiones del espacio en las que se dará o no el efecto.

Esto es, la función de la red es clasificar regiones del espacio. En algunas de esas regiones se dará el efecto estudiado, y en el resto no. Al aprender, la red lo que hace es delimitar esas regiones.

Esto nos da una idea del tipo de procesos que se podrán estudiar con redes de estas características.

Como ejemplos de aplicación de este tipo de redes, podemos citar dos tesis doctorales publicadas en la Universidad de Salamanca en el año 2001, en las que se entrenaban redes neuronales para realizar tareas de diagnóstico. O ejemplos en los que redes neuronales aprendían, en función de determinados parámetros de entidades bancarias, si éstas irían o no a la quiebra. Es todos estos casos, la red delimita regiones del espacio en las que se dará o no el efecto.

Otro aspecto muy importante de las redes neuronales es que el aprendizaje es una variación sucesiva de los valores de los pesos que forman la red.

Visto así, la red se puede pensar como un sistema dinámico. Podemos imaginar el estado actual de la red como un punto en el espacio de pesos de conexión, y el aprendizaje implica una evolución de ese punto en ese espacio. Normalmente, esta evolución viene expresada por una ecuación en diferencias, del tipo

$$\omega_i^{n+1} = \omega_i^n + f\left(\omega_j^n, x_j^n, t^n\right)$$

en el que los valores del siguiente estado de la red vienen determinados por los valores en el estado actual más una función que depende de esos valores, los parámetros de entrada y la salida buscada o target.

Posiblemente el algoritmo más utilizado en el entrenamiento de redes neuronales sea el de Backpropagation. Mediante ese algoritmo se buscan los mínimos de una función error definida, como hemos dicho anteriormente, como un error cuadrático. Esta función error es el equivalente en sistemas dinámicos a una función energía o, con mayor generalidad, una función de Liapunov que hay que minimizar.

Con respecto a este punto de vista, queremos señalara dos aspectos importantes:

¿***Busca una red neuronal biológica un mínimo de energía global o hace modificaciones estocásticas y comprueba el error?***

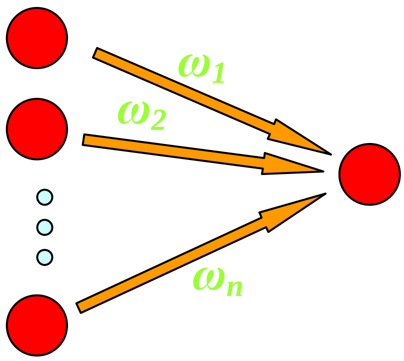
Aparte de esa cuestión a dilucidar, esta perspectiva de las redes neuronales nos abre la posibilidad de implementar algoritmos de búsqueda del mínimo error que sean compatibles con los conocimientos sobre el funcionamiento neuronal. En este sentido, la segunda opción manifestada en la pregunta parece más razonable, y nos debería llevar a la misma solución que las redes empleadas típicamente en Inteligencia Artificial, aunque deberíamos comprobar por un lado si esto es así en todos los casos y, por otro lado, si existen diferencias observables en el proceso de aprendizaje, fenómeno esperable dado que el algoritmo implementado es ciertamente diferente.

En aprendizajes normales (un solo estímulo a condicionar, condicionamiento de varios estímulos en fenómeno de ensombrecimiento,…) el algoritmo de Backpropagation y el de Rescorla-Wagner implementan el mismo conjunto de ecuaciones.

Este es un resultado muy importante a considerar cuando estamos buscando modelos generalizadores de ambos. Si se piensa en el modelo de Rescorla-Wagner como una red de conexiones entre distintos estímulos y una salida, sin capas ocultas intermedias, y se aplica a esa red la filosofía de resolución implementada por algoritmos como el backpropagation, que intentan llegar al mínimo de la función error o energía mediante el gradiente de esa función error, obtenemos el mismo conjunto de ecuaciones que el del modelo de Rescorla-Wagner.

REDES NEURONALES Y MODELO DE RESCORLA-WAGNER.

Uno de los resultados que hemos obtenido es la demostración de lo dicho anteriormente:



Si tenemos una red como la de la figura, tendremos que la salida de la red será,

$$y = \omega_1 x_1 + \omega_2 x_2 + \ldots + \omega_n x_n$$

Si el tipo de ensayos que presentamos a la red es de la forma

$$[1,1,\ldots 1/1]$$

Entonces los valores de los parámetros de entrada son 1, por lo que la función energía quedará

$$E = \frac{1}{2}[1 - (\omega_1 + \omega_2 + \ldots + \omega_n)]^2$$

Buscar el gradiente de esta función energía consiste en derivar con respecto a cada uno de los pesos de conexión, lo que daría para las derivadas parciales

$$\frac{\partial E}{\partial \omega_i} = (\lambda - \Sigma \omega_j) \cdot (-1)$$

Y como en el algoritmo de Backpropagation se implementa que la variación de los pesos viene dada por

$$\omega_i^{n+1} = \omega_i^n - \alpha \frac{\partial E}{\partial \omega_i^n}$$

Nos quedaría entonces la expresión

$$\omega_i^{n+1} = \omega_i^n + \alpha\left(\lambda - \Sigma\omega_j^n\right)$$

que es precisamente el modelo de Rescorla-Wagner.

Esto es, en este tipo de aprendizajes el modelo de redes neuronales y el de Rescorla-Wagner es el mismo. Podremos comprobar este resultado teórico mediante las simulaciones que presentaremos más adelante en este trabajo.

GENERALIZACIÓN RW-RN.

Ya hemos visto que el modelo de Rescorla-Wagner y el de Redes Neuronales se reducen al mismo conjunto de ecuaciones en varios de los tipos de aprendizaje que son habituales.

Sin embargo, es cierto que la idea subyacente a cada uno de ellos es en principio diferente.

En concreto, el modelo de Rescorla-Wagner compara el parámetro λ con la suma de los pesos de conexión establecidos:

$$\omega_i^{n+1} = \omega_i^n + \alpha\left(\lambda - \Sigma\omega_j^n\right)$$

Sin embargo, una red neuronal compara el parámetro λ con la salida de la red, esto es,

$$\omega_i^{n+1} = \omega_i^n + \alpha\left(\lambda - \Sigma\omega_j^n x_j\right)$$

El funcionamiento de cada uno de estos modelos es, en principio, diferente, y pueden estar implementando de alguna forma procesos neuronales distintos.

En nuestra búsqueda de un framework básico a los modelos actuales en aprendizaje causal, un paso sería integrar precisamente estos dos modelos en un modelo más general. Una posibilidad de esta integración sería construir un algoritmo que funcionara en parte a la manera de Rescorla-Wagner, teniendo en cuenta la fuerza asociativa total conseguida, y en parte como una Red Neuronal, en que se tuviera en cuenta precisamente la salida de la red.

Posiblemente el modelo más sencillo que implementa esta idea sería

$$\omega_i^{n+1} = \omega_i^n + \alpha\left(\lambda - \Sigma\omega_j^n\left(1 + \Lambda\left(x_j - 1\right)\right)\right)$$

donde el parámetro Λ nos daría una medida de en qué proporción esta red actúa como el modelo de Rescorla-Wagner o como una Red Neuronal.

Los casos extremos serían

a) $\Lambda = 0$, con lo que quedaría el modelo de Rescorla-Wagner.
b) $\Lambda = 1$, con lo que obtendríamos un modelo de Red Neuronal.

Como ejemplo, un valor de $\Lambda = 0.45$ estaría implementando una red que funcionaría en un 45% como un modelo de Rescorla-Wagner y en un 55% como una Red Neuronal (usando un lenguaje porcentual quizás poco riguroso pero muy clarificador con respecto a la idea que estamos intentando transmitir).

¿Qué resultados diferenciales podremos obtener con este modelo?

Algunas de las formas de actuación de este modelo se podrán deducir teóricamente.

Por ejemplo, es de esperar que en los casos en los que tanto el modelo de Rescorla-Wagner y el de Red Neuronal proporcionen los mismos resultados, el modelo propuesto también proporcionará esos resultados, con lo que no se obtendrán diferencias. Sin embargo, sí que es esperable encontrar resultados diferentes cuando ambos modelos originales no proporcionen los mismos resultados.

Sin embargo, en todos los casos se pueden implementar simulaciones de todos estos modelos para poder estudiar su comportamiento bajo determinadas condiciones.

¿Qué tipo de procesos neuronales (diferentes tipos de procesos de neuroplasticidad) implementa cada modelo?

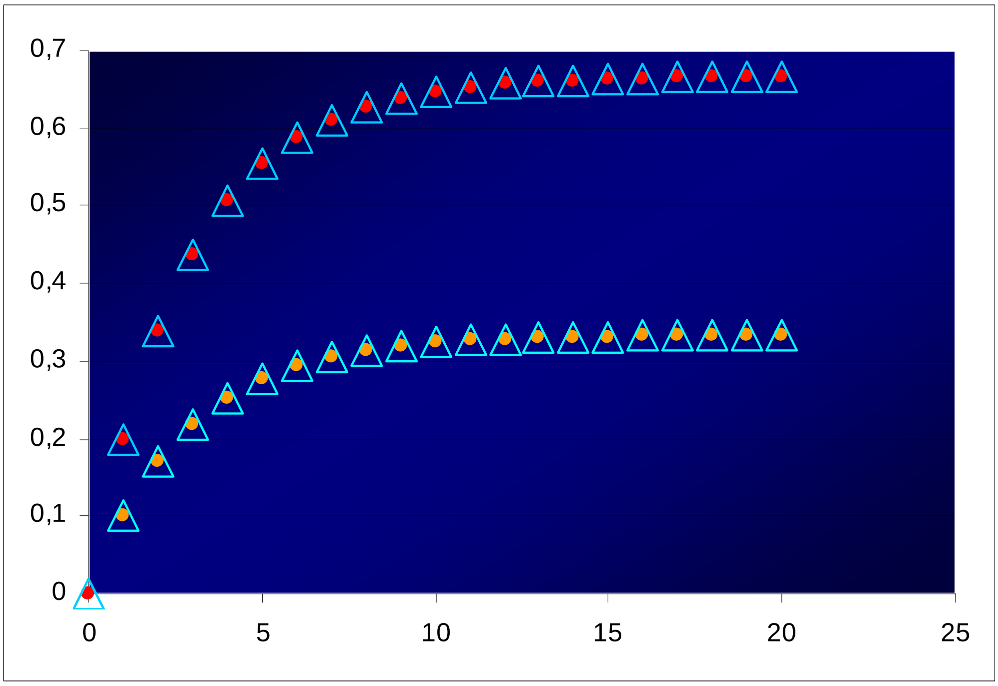
Ya hemos señalado que la idea subyacente al modelo de Rescorla-Wagner y al de Redes Neuronales es en principio diferente. Esto puede estar implicando diferentes procesos de neuroplasticidad en el aprendizaje.

Con el modelo que hemos propuesto podemos controlar cuánto de cada tipo de proceso de neuroplasticidad queremos que esté presente en un determinado proceso de aprendizaje.

SIMULACIÓN: COMPARACIÓN ENTRE EL MODELO RW Y RN EN EL FENÓMENO DE ENSOMBRECIMIENTO.

Hemos realizado simulaciones en las que hemos comparado cómo pueden responder tanto el modelo de Rescorla-Wagner como el modelo de Redes Neuronales ante un fenómeno de ensombrecimiento.

En el caso que aquí presentamos, se realizaron 20 ensayos en los que se asociaban dos estímulos condicionados A y B a un estímulo incondicionado. Como parámetros de aprendizaje se utilizaron $\alpha_A = 0,1$ y $\alpha_B = 0,2$.

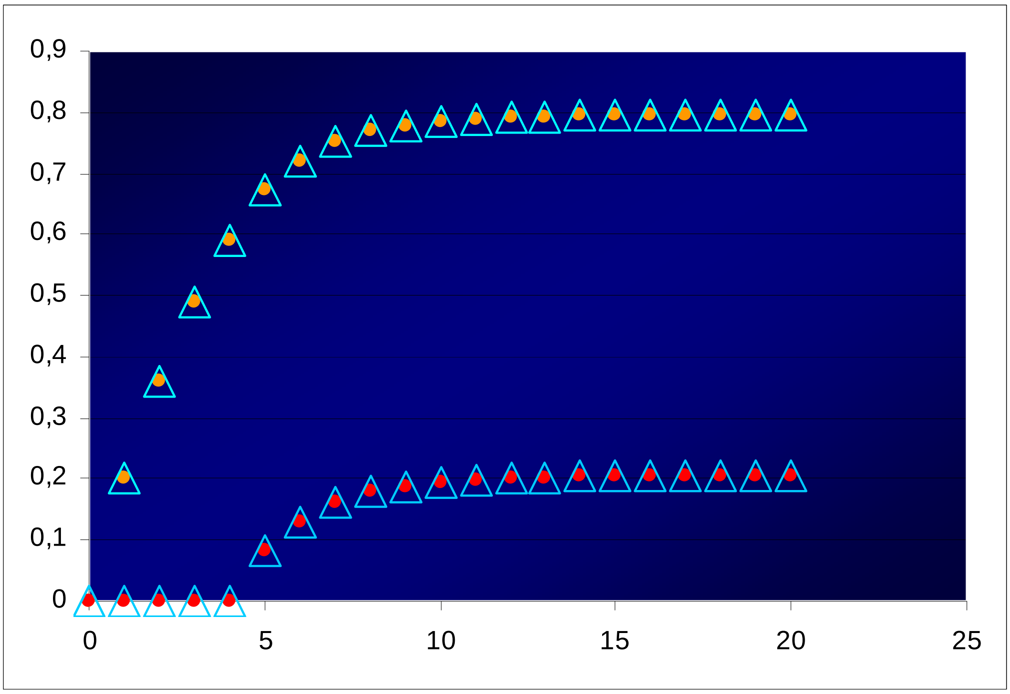


Como se puede observar, ambos modelos aprendieron de la misma forma. Este resultado está de acuerdo con la deducción realizada anteriormente en la que se mostraba cómo ambos modelos llevan, en determinados fenómenos de aprendizaje, al mismo conjunto de ecuaciones.

SIMULACIÓN: COMPARACIÓN ENTRE EL MODELO RW Y RN EN EL FENÓMENO DE BLOQUEO.

Hemos simulado también el fenómeno de bloqueo, con la intención de comprobar cómo se comportaban el modelo de Rescorla-Wagner y el modelo de Redes Neuronales.

En este caso presentamos cuatro ensayos en los que se condicionaba solamente el estímulo A al estímulo incondicionado. Después se presentaron dieciséis ensayos en los que se presentaban ambos estímulos condicionados, A y B, junto con el incondicionado. Los parámetros de aprendizaje fueron $\alpha_A = 0.2$ y $\alpha_B = 0.2$. Los resultados obtenidos se muestran en la siguiente gráfica:



donde los círculos se refieren al modelo de Rescorla-Wagner, y los triángulos al de Red Neuronal. Como se puede apreciar, en este caso también el aprendizaje se realiza de la misma forma.

REDES NEURONALES BIOLÓGICAS.

Son construcciones matemáticas de redes de neuronas que implementan el funcionamiento de las neuronas reales.

Hemos implementado un modelo de red neuronal biológica. El funcionamiento de este modelo no se basa en la asociación de cada neurona con un número, como en las redes neuronales utilizadas en Inteligencia Artificial, sino que cada neurona presenta una cierta probabilidad de disparo de potenciales de acción.

Hemos implementado esta probabilidad de disparo de potenciales de acción. En la red se puede apreciar la evolución temporal de esta activación.

Hemos implementado el concepto de patrón de activación como una mayor frecuencia de disparo de potenciales de acción en un conjunto determinado de neuronas.

Hemos asociado determinados estímulos con la activación de diferentes patrones de activación de neuronas. En concreto, simulamos la presencia de una luz con la pulsación de la tecla “l” del teclado, y la presencia de comida como la tecla “c” del teclado. La pulsación de cualquiera de esas teclas da lugar a la generación de su patrón de activación respectivo.

Hemos implementado diferentes pesos de conexión entre determinadas neuronas.

El objetivo es simular los procesos de aumento y disminución de la conexión sináptica, e intentar reproducir procesos de asociación entre diferentes patrones de activación neuronales.

BIBLIOGRAFÍA.

Freeman, J.A. y Supura, D.M. (1993). *Redes neuronales. Algoritmos, aplicaciones y técnicas de programación.* Addison-Wesley.

Hongjing Lu, Alan L. Yuille, Mimi Liljeholm, Patricia W. Cheng, and Keith J. Holyoak. (2008). Bayesian Generic Priors for Causal Learning. *Psychological Review.* Vol. 115, No. 4, 995-984.

Kruschke, J.K. (2006). Locally bayesian learning with applications to retrospective revaluation and highlighting. *Psychological Review.* Vol. 113. No 4. 677-699.

Martín del Brío, B. y Sanz Molina, A. (2001). *Redes Neuronales y Sistemas Borrosos.* Editorial RA-MA.

Pearce, J.M. (1994). Similartity and discrimination: a selective review and a connectionist model. *Psychological Review*. Vol 101, No. 4, 587-607.

Rummelhart D.E. & McClelland J.L. (Eds.) (1988). *Parallel distributed processing.* Vol 1 y 2. Cambridge, MA: MIT Press.

Schwartz, B,., Wasserman, E. A. y Robbins, S. J. (2001). *Psychology of Learning and Behavior* (Capítulo 15). W. W. Norton & Company.

Shanks, D.R. (2006). Bayesian associative learning. *Trends in Cognitive Sciences.* Vol. 10 No. 11.