

# Modelos de la Asignación de Credito: Rescorla y Wagner

Arturo Bouzas

¿Se comportan los animales cercanos al óptimo?

Nos podemos preguntar si existe una regla alternativa a la contigüidad que capture más verídicamente la estructura estadística o de causalidad en el entorno de los organismos.

Primera familia de asignación de crédito

Modelos estadísticos para detectar relaciones de causalidad entre sucesos del mundo para la reducción de la incertidumbre:

Los organismos adaptables, reducen la incertidumbre acerca de la ocurrencia de algún evento importante, buscando posible predictores. Seleccionando un EC como candidato a predictor, compara la información que este proporciona con la proporcionada por otro posible candidato. Esto quiere decir que para establecer relaciones de causalidad no es óptimo considerar tan solo la contigüidad entre dos estímulos, EC y EI.

Considere el caso más sencillo. Un mundo con dos estímulos, en dos posibles estados cada uno: Un estímulo condicionado presente o ausente (EC, nEC) y un estímulo incondicionado presente o ausente (EI, nEI). La presentación conjunta de los estímulos la podemos representar en una tabla de  $2 \times 2$ .

Cada celda es un suceso conjunto, en nuestro caso la combinación de la presentación o ausencia de nuestros dos estímulos.

Si les es más fácil pueden considerar dos variables discretas, sexo y cabello largo o corto.

Cada celda sería una combinación de sexo (H o M) y largo del cabello (C o L)

Lo que es importante es que en este esquema los estímulos son sucesos discretos.

|     | El | nEl |
|-----|----|-----|
| EC  | a  | b   |
| nEC | c  | d   |

Sumando columnas e hileras obtenemos el número de ocasiones que aparece cada estímulo. Por ejemplo, el estímulo condicionado EC aparece  $a + b$  veces. La probabilidad de la conjunción (cada celda) se obtiene como cualquier otra probabilidad:

$P(EI \cap EC) = \text{Número de elementos en EC y EI} / \text{número total de elementos.}$

$$P(EI \cap EC) = a / N$$

La probabilidad de la ocurrencia de un EC es:

$$P(EC) = (a + b) / N$$



A nosotros nos interesa obtener la probabilidad condicional de la ocurrencia del EI dado la presentación del EC:

$$P(EI | EC) = P(EI \cap EC) / P(EC) = a / (a + b)$$

Para poder concluir que hay una dependencia del EI en el EC, debemos comparar la anterior probabilidad condicional, con la probabilidad condicional de la presentación del EI en la ausencia del EC (nEC) .

$$P(EI | nEC) = P(EI \cap nEC) / P(nEC)$$

$$= c / (c + d)$$

Definimos una medida de la contingencia:

$$\Delta \text{ Contingencia} = P(EI | EC) - P(EI | nEC) \\ = (a / (a + b)) - (c / (c + d))$$

Si graficamos una probabilidad condicional como una función de la otra, obtenemos un espacio de contingencia en el que la diagonal de 45° representa no dependencia. El espacio arriba de la línea resulta en condicionamiento excitatorio y el espacio debajo de la línea resulta en condicionamiento inhibitorio. Conforme nos alejamos del origen la contingencia es más fuerte.

## Aprendizaje: búsqueda de predictores

En términos de optimización el problema de la asignación de crédito se resuelve usando la medida de la contingencia, la que refleja el supuesto de que el condicionamiento se dará solo si el EC comparado con su ausencia (nEC) es un mejor predictor del estímulo incondicionado. La misma regla opera si hay más de un posible estímulo asociado con el EI: el organismo responde a aquel que sea un mejor predictor.

Los resultados estudiados en la clase anterior los presenté informalmente en términos de atención, correlación, y sobre todo predicción. Casi inmediatamente después de la publicación de éstos resultados Rescorla comenta sobre la necesidad de entenderlos desde una perspectiva asociativa formal y en colaboración con Wagner presenta un modelo (ecuación) que determina la dirección del estudio del aprendizaje durante el último cuarto del siglo XX.

# Rescorla y Wagner

El modelo de Rescorla y Wagner incluye dos importantes grandes supuestos:

1. Un Modelo de Refuerzo que establece que el motor del aprendizaje es la reducción en el error de predicción
2. Un modelo de la forma como un organismo percibe estímulos compuestos. En particular el supuesto que los organismos perciben a los estímulos como un conjunto de elementos separables que compiten por la asignación de crédito.

# Modelo de Refuerzo: Un integrador con fuga

Bush y Mosteller (1950)

Los organismos parten el tiempo en momentos discretos en los cuales puede o no ocurrir un suceso  $x$  biológicamente importante.

Los organismos integran en un número el impacto de las experiencias previas con la presentación o ausencia de un estímulo  $x$  a lo largo de los momentos discretos.

A este número se le conoce como la fuerza (asociativa) del estímulo  $x$

$V_x$

El semestre siguiente veremos que un esquema similar se aplica a la fuerza de una respuesta.

La fuerza asociativa  $V_x$  se emplea para predecir la ocurrencia del estímulo biológicamente importante y ajustar el comportamiento dada esa predicción.

En cada momento en el tiempo  $V_{x_{n+1}}$  se actualiza como una función de dos variables, su valor integrado hasta el momento inmediatamente anterior  $V_{x_n}$  y la ocurrencia o no del evento biológicamente importante (refuerzo)  $R_n$

$$V_{n+1} = f(V_n, R_n)$$

Una función **f** que ha dominado la literatura psicológica es un modelo lineal que recibe el nombre de un integrador con “fuga”, operador lineal o modelo de reforzamiento, el cual describe la fuerza asociativa  $Vx_{n+1}$  como una suma ponderada del valor de  $Vx$  acumulado hasta el momento  $n$  y si ocurre o no un refuerzo

$$Vx_{(n+1)} = (1-a)Vx_n + aR_n \quad (1)$$

donde:  $0 < a < 1$



El parámetro  $a$  determina la importancia de la experiencia acumulada hasta el momento  $n$ , relativa a la ocurrencia o no de el suceso biológico.

Imagine una secuencia de momentos  $n$ , en cada uno de ellos se lanza una moneda al aire y puede caer águila o sol, su tarea es predecir si en el siguiente momento si caerá águila o sol.

1 1 0 1 1 1 0 1 0 1 0 1110

Ud está en el momento 15 y tiene que predecir que ocurrirá en el momento 16,

El parámetro  $a$  determina que tan importante es la evidencia acumulada hasta el momento 15 relativa a lo que ocurrió en el momento 15. Si  $a$  es cercano a uno el que en el momento 15 observará un 0 sería muy importante y lo llevaría a predecir un 0, por el contrario si  $a$  fuese cercano a 0 la fuerza acumulada por haber observado más 1s que 0s lo llevaría a predecir un 1.

El parámetro  $a$  es una especie de ventana temporal que determina que tan atrás vamos en nuestra experiencia para predecir el futuro.

# Aprendizaje como Reducción en el Error de Predicción

Otra forma de interpretar la ecuación del modelo de refuerzo, es como un proceso que le permite al organismo comparar de momento a momento la predicción que hace a partir de la integración en memoria de los eventos encontrados en el pasado con el resultado obtenido en el presente. Este proceso de comparación es de enorme generalidad e implica psicológicamente que respondemos a las diferencias entre lo que esperamos y lo que obtenemos. Para ver este componente con más claridad

considere modelar la magnitud del cambio de un momento a otro

Para ver este componente con más claridad considere modelar la magnitud del cambio de un momento a otro  $\Delta V_{n+1}$

Restándole  $V(t)$  a ambos lados de la ecuación del integrador y dejando que

$$\Delta V = (V_{n+1} - V_n)$$

$$\Delta V = (1-a)V_n + aR_n - V_n$$

$$\Delta V = V_n - aV_n - V_n + aR_n$$

$$\Delta V = aR_n - aV_{(t)}$$

$$\Delta V = a(R - V) \quad (2)$$

La ecuación 2 con la simbología de los libros introductorios es:

$$\Delta V_{n+1} = \alpha(\lambda - V_{n+1}) \quad (3)$$

Si dejamos que  $(\lambda - V_{n+1})$  sea  $\delta$ , que llamaremos error de predicción

Y la expresamos solo en términos de  $V_{n+1}$  nos quedamos con la versión más conocida

$$Vx_{n+1} = Vx_n - \delta \quad (4)$$

Rescorla y Wagner , a partir del modelo de reforzamiento presentado anteriormente, propusieron un modelo que podía explicar los resultados no consistentes con el principio de la contigüidad , pero en el contexto de una teoría asociacionista. Modelo que simultáneamente tenía que dar cuenta de la forma de las curvas de adquisición y extinción (ganancias y pérdidas decrecientes).

Para entender el modelo de RyW consideremos que característica del entorno causal modela. Hasta antes de los años sesenta, los investigadores limitaban sus experimentos a protocolos que presentaban estímulos aislados. Sin embargo, el entorno no consiste de elementos únicos que aparecen en relaciones que difieren en su contigüidad.

Por el contrario, los organismos enfrentan múltiples estímulos que se presentan en forma simultánea y en contiguidad con sucesos biológicamente importantes. Más aún, la experiencia de los organismos incluye contactos con algunos de los estímulos en forma separada, contiguos con reforzadores, o no correlacionados entre sí. Cómo puede un organismo extraer relaciones de “causalidad” en ésta red de diversas experiencias?



En resumen, cuando un organismo enfrenta dos o más estímulos presentes en forma simultánea y contigua con un reforzador, que principios describen que tanto aprenderá sobre cada uno de los elementos (a cual le va a “atribuir” el reforzador), y el efecto que tienen sobre ésta decisión la experiencia previa con cada uno de los elementos. El modelo de RyW describe una solución al problema de la atribución (asignación) de crédito: a cual de una multiplicidad de elementos igualmente contiguos con un reforzador se le atribuye su presentación.

# Supuestos del Modelo de Rescorla y Wagner

1. La fuerza asociativa de un estímulo condicionado es su capacidad para activar en memoria una representación del estímulo incondicionado. A ésta representación la podemos interpretar como lo que el organismo espera obtener dado la presentación del estímulo condicionado. La fuerza asociativa de un estímulo se representa con una variable ( $V$ ), la cual puede tomar valores positivos (excitatoria) y negativos (inhibitoria).

Competencia entre elementos separables de un estímulo compuesto

2. Los estímulos en compuesto están conformados por elementos (estímulos) separables. Una cara no sería un estímulo integrado, sino un conjunto de estímulos (ojos, boca, orejas, nariz, etc), cada uno con su propia fuerza asociativa.

3. La ecuación que proponen computa para cada uno de los elementos de un compuesto, una fuerza asociativa ( $V$ ). Si el compuesto incluye dos estímulos A y B, tendremos que computar dos fuerzas asociativas:  $V_A$  y  $V_B$ . La fuerza de la representación del estímulo incondicionado (que esperamos obtener) es el agregado (la suma) de las fuerzas asociativas de todos los elementos individuales que integran el estímulo compuesto. Si son dos estímulos, entonces

$$V_{\text{total}} = V_A + V_B.$$

# Aprendizaje como reducción de error

La ecuación de RyW captura la noción de que el aprendizaje (los cambios en la fuerza asociativa de cada uno de los estímulos) es proporcional a la discrepancia entre lo que se obtiene ( $\lambda$ ) y lo que se espera obtener dado la suma de las fuerzas asociativas de todos los estímulos presentes simultáneamente ( $V_{total}$ )

$$\Delta Vx_{n+1} = \alpha(\lambda - V_{total\ n})$$

Equivalentemente

$$Vx_{n+1} = Vx_n + \alpha(\lambda - V_{total\ n})$$

$\alpha$  es un parámetro de aprendizaje que determina la importancia relativa del pasado vs el error de predicción.

Si  $\alpha$  es cercana a 0 cuenta principalmente el pasado integrado, si es cercana a 1 cuenta principalmente la discrepancia entre lo obtenido y lo predicho (el error de predicción).

5. Es claro que la ecuación es una instancia de un integrador, que especifica como motor del aprendizaje la reducción del error en la predicción. La parte novedosa de la ecuación de Recorla y Wagner, consiste en tomar como predicción la suma del valor predictivo de todos los estímulos presentes  $V_{total} = V_A + V_B$ . Esta consideración genera el último supuesto,

6. Competencia. Cada reforzador permite una cierta asociabilidad representada por el parámetro  $\lambda$ , que es también el equilibrio de la ecuación (3) (la asíntota). Recuerde que  $\Delta V$  es cero cuando  $\lambda - V = 0$ . Por ello, podemos hablar que los estímulos separados compiten para convertirse en el mejor predictor del reforzador.



Considere el caso del protocolo del bloqueo. Para el grupo que recibe en la fase 1 experiencia con el tono seguido de la descarga, para el final de esa fase la fuerza asociativa del tono es igual a  $\lambda$  y consecuentemente  $\lambda - V_{\text{tono}} = 0$ . En lenguaje menos técnico, el tono predice perfectamente a la descarga eléctrica. Cuando en la fase 2 se presenta al tono junto con una luz, contiguos con la descarga eléctrica, la  $V$  total es igual a  $\lambda$  y consecuentemente no se aprende nada acerca de la luz. En otras palabras, el organismo ya esperaba la descarga eléctrica y no existe ninguna discrepancia entre lo esperado y lo obtenido.

La ecuación de R y W hace predicciones contra intuitivas. Si dos estímulos A y B se entrenan hasta asíntota, y posteriormente se les presenta en compuesto contiguos al mismo reforzador, R y W predicen que habrá una sobreexpectación del reforzador que resulta en que cada elemento individual pierde, en vez de ganar, fuerza asociativa. Existe evidencia a favor de la predicción.

Un reto importante para la ecuación de R y W es dar cuenta de los resultados de los experimentos de Rescorla, en los que se manipula la relación de contingencia (la probabilidad de la presentación del EI dada la presencia o la ausencia del EC). La solución fue asumir que el protocolo experimental presenta en realidad dos estímulos uno que es el contexto X solo y otro que es el compuesto del contexto con el EC explícito AX. En el caso del procedimiento con igual probabilidad del EI en la presencia y la ausencia del EC, la ecuación de R y W interpreta el experimento como uno de bloqueo en el que el contexto X es el mejor predictor del EI.

Una forma de evaluar su comprensión del modelo de R&W es considerar su predicción en experimentos con protocolos no correlacionados, en los cuales los reforzadores durante el intervalo entre ensayos son señalados con otro estímulo. En este caso, los estímulos presentados durante el intervalo entre ensayos es predicho por la nueva señal y el contexto no gana control y no compite con el EC explícito.