

Preconceptualización con modelos neuronales lineal-no lineal en el sistema visual.

Dr. Joan Benavent Gomar^{1*}

Dirigido por el Dr. Jesús Malo López

Image Processing Laboratory (UV). ¹jobego2@alumni.uv.es

Máster en neurociencias básicas y aplicadas

Mayo
2018



VNIVERSITAT
ID VALÈNCIA

Palabras clave: sistema visual, normalización, lineal, aprendizaje supervisado.

*PhD. por las Universidades de València, Salamanca, La Laguna, La Coruña, Santiago, Valladolid y CSIC. Ms. Pensamiento filosófico contemporáneo. Ms. Física avanzada. (Astrofísica). Ms. Ciencias y tecnologías de la información. (TIC). Lic. Ciencias físicas. Ing. Informática. Ing. Tec. Informática de sistemas.

Jesús Malo López, doctor en ciències físiques i professor titular de universitat al departament d'òptica de la Universitat de València,

CERTIFIQUE

Que el Dr. **Joan Benavent Gomar**, estudiant del Màster de Neurociències Bàsiques i Aplicades d'aquesta universitat, ha realitzat sota la meua direcció el treball titulat:

Preconceptualización con modelos neuronales lineal-no lineal en el sistema visual

Que constitueix el seu Treball Final de l'esmentat Màster. Igualment puc confirmar que el disseny i execució dels experiments, l'estudi i la confecció de la memòria justifiquen molt sobradament els 12 crèdits ECTS (300 hores de dedicació) de què consta el Treball Final del Màster. La qual cosa signe a Burjassot el dia 8 de maig de 2018,



Índice

1	Resumen	4
2	Introducción	4
3	Material y métodos	7
3.1	Materiales	7
3.2	Métodos	7
3.3	Descripción del proceso de aprendizaje (POP) desarrollado	11
4	Resultados	14
4.1	sobre la hipótesis de convergencia.	15
4.2	sobre el aprendizaje estático supervisado.	16
4.3	sobre el aprendizaje dinámico semi-supervisado.	16
4.4	sobre la “Imaginación”.	17
5	Discusión	21
5.1	del correlato neuronal.	21
5.2	de la hipótesis de convergencia.	21
5.3	del módulo “sistema de aprendizaje de escena” (POP) desarrollado. Breve reflexión.	24
5.4	del invariante como media de las instancias positivas observadas por el sistema sensorial frente al invariante discriminador.	27
5.5	del problema de la labilidad de la memoria. ¿Es necesario recordar todas las instancias o estas pueden comprimirse e incluso olvidarse con el tiempo?	28
5.6	de la puerta del conexionismo.	29
6	Conclusiones	29
7	Agradecimientos	30

1. Resumen

El presente trabajo trata los procesos de normalización en el sistema visual, desde la entrada sensorial en retina, núcleo geniculado lateral (NGL) hasta el área visual V1, con el fin de encontrar un invariante de escena que sea capaz de indentificarla. Entendiendo como procesos de normalización a las transformaciones de las señales sensoriales desde la retina hasta el cortex.

Los tres primeros estados con etapas lineales y no lineales, normalizan la escena ajustando la luminancia, el brillo y el contraste. El cuarto estado transforma la señal y la mapea en el espacio, en frecuencia y en ángulo, según el modelo implementado por Martinez-Garcia et al., (2018). Su correlato neuronal correspondería: los dos primeros estados con la retina, el tercero con el núcleo geniculado lateral (NGL), y el cuarto dentro de V1.

A partir de este estado se propone otro nuevo, coherente con los anteriores, con tres niveles jerárquicos, que es capaz de aprender este invariante con métodos de clasificación lineal. Su correlato neuronal estaría en V1.

Con este fin, se entrena un clasificador lineal con aprendizaje supervisado para el reconocimiento de escenas, y se intenta buscar en sus parámetros el invariante de escena.

Se prueba la eficiencia del nuevo estado en la predicción de escena y en el aprendizaje, tras nuevas evidencias, que se adecúen a la nueva realidad. Se analiza en profundidad este proceso. El hecho de no poder contrastar la predicción con la realidad se ha suplido dando crédito a las predicciones. Este modelo admite aprendizaje no supervisado o semi-supervisado.

A nivel visual, se comprueba la “imaginación” de los invariantes (su versión como imagen), tanto los del modelo como los que no tienen en cuenta la discriminación. Finalmente, se comparan con la entrada visual, concretamente en el estado 3.

El experimento realizado es compatible con la hipótesis de que la normalización en los primeros estados es convergente, apuntando hacia la existencia de un invariante de escena.

2. Introducción

El marco donde se desarrolla este trabajo es la disciplina de neurociencia computacional. Disciplina que enlaza campos diversos como la biofísica, la neurociencia, la ciencia cognitiva, la inteligencia artificial, la informática y las matemáticas. Su característica principal es que enfatiza las descripciones biológicas y funcionales de las neuronas, su fisiología y su dinámica. Suelen seguir los modelos dinámicos de estados para sus descripciones (Ogata, 1987).

Estado de la cuestión

La neurociencia computacional aplicada a los procesos de visión (Rolls & Deco, 2001), trata de modelizar procesos neuronales, con ordenadores, que simulen el comportamiento biológico de las neuronas. La modelización de la visión ha tenido una importancia relevante en este campo.

Actualmente se está tratando de entender el proceso de visión, desde modelos funcionales matemáticos, que tienen en cuenta las perturbaciones que las neuronas vecinas provocan en la respuesta. En el artículo de Carandini & Heeger (2012) los autores proponen la “normalización” de los estímulos como una forma de computación neuronal de los sentidos, entre ellos el de visión. El modelo de normalización está definido por una simple ecuación, la ecuación de normalización. Esta ecuación especifica cómo la respuesta normalizada R_j de una neurona j depende de sus entradas D_k (que no están normalizadas):

$$R_j = \gamma \frac{D_j^n}{\sigma^n + \sum_k D_k^n} \quad (1)$$

El numerador representa la entrada de la neurona D_j en el sistema sensorial, este contenido de entrada proporciona el estímulo conducido a la respuesta. Sus unidades dependen del sistema que se está estudiando. Puede ser en unidades de intensidad de estímulo en un sistema sensorial o, por espigas por segundo, si la entrada es considerada como salida de otra neurona. El denominador es una constante σ más el factor de normalización, que es la suma de todas las entradas D_k del conjunto de normalización. Las constantes γ , σ y n constituyen parámetros libres que son usualmente calculados de forma empírica: γ determina la responsividad general, σ previene de la división por cero y determina cómo responde a la saturación con el incremento de entradas, y n es un exponente que amplifica la entrada individual. Esta normalización tiene una función también en la zona del cortex visual V1 (Coen-Cagli & Schwartz, 2013) que optimiza las estadísticas del entorno natural. El autor citado usa para describir esta función clasificadores lineales.

Además, y centrados en la visión, el modelo de Wilson–Cowan (Cowan et al., 2016) describe ecuaciones que resuelven la dinámica cortical.

Recientemente se ha implementado un diseño modular de capas lineales y no lineales, enfocado a objetos modulares que responderían a mecanismos que, además, se repiten en la biología del sistema nervioso central (Malo & Gutierrez, 2006; Malo & Simoncelli, 2015; Malo et al., 2016; Martinez-Garcia et al., 2018) sugieren un modelo que apunta hacia estas direcciones. En concreto Martinez-Garcia et al., (2018).

Siguiendo la idea de que el procesamiento cortical está optimizado para las estadísticas del entorno natural (Simoncelli & Olshausen, 2001; Zhaoping, 2006), se han estudiado modelos multicapa de estadísticas de imágenes que capturan correlaciones de orden superior (Hinton, 2007; Karklin & Lewicki, 2005; Karklin & Lewicki, 2009; Osindero et al., 2006; Schwartz et al., 2006), para aprender sin supervisión las características intermedias, como los contornos extendidos (Hoyer & Hyvaarinen, 2002), las esquinas (Spratling, 2011), la agrupación en múltiples frecuencias (Hyvaarinen et al., 2005) y orientaciones (Nuding & Zetsche, 2007). Pero no se han encontrado en la literatura modelos que aprendan con supervisión, sería absurdo. El modelo propuesto parte de un aprendizaje semi-supervisado que en las primeras percepciones admite aprendizaje no supervisado. Es el organismo el que valida o refuta las predicciones. Si ha sido percibida como buena o mala por el organismo. Estableciendo las jerarquías de conceptos “a posteriori”.

Por otra parte, aplicando este procedimiento, sería incluso posible obtener descriptores físicos de imagen para tareas de recuperación en bases de datos, diferentes a los utilizados actualmente en ingeniería (Benavent et al., 2010; Benavent et al., 2012; Benavent et al., 2013; Castellanos et al., 2012; Granados et al., 2011; de Ves et al., 2016; Castellanos et al., 2011), que siguen los criterios de Deselaers et al., (2008). Acercándolos a los propuestas de da Fonseca & Samengo, (2016) basados en distancias cromáticas geométricas en humanos, que admiten otro tipo de entrenamiento y otros modelos no lineales.

Hipótesis de trabajo y objetivos

Entendiendo que la función de los procesos de normalización es:

- modelizar las transformaciones que sufre la entrada sensorial desde la retina hasta el cortex visual, y además,
- si estas transformaciones han de tener en cuenta la influencia de las neuronas vecinas sobre la portadora de la señal,

es lícito pensar que las influencias mútuas entre las vecindades neuronales tiendan a una homogeneización de las señales neuronales de una misma percepción. Lo mismo ocurrirá con distintas percepciones de una misma escena. Por tanto, se puede hipotetizar que:

Dado que los procesos de normalización pueden verse como un proceso de convergencia de la entrada sensorial, en el sistema de visión, *es posible que exista un invariante de escena*. Un elemento que la identifique y, además, la discrimine. En definitiva, que nos permita distinguir ante qué escena nos encontramos.

Y si es así, ¿en qué nivel jerárquico situamos al estado que lo elabora? Finalmente, por ser un “constructo” elaborado por el cerebro, ¿podría “imaginarse”, es decir, visualizarse como imagen?

En este sentido, los objetivos del presente trabajo se enfocarán para intentar resolver estas cuestiones.

- *Primero*: Comprobar la convergencia de la entrada sensorial tras los procesos de normalización.
- *Segundo*: Diseñar un estado que explique el aprendizaje y reaprendizaje del invariante ante nuevas entradas sensoriales.
- *Tercero*: Resolver cómo se puede “imaginar” el invariante.

3. Material y métodos

3.1. Materiales

Se ha utilizado un ordenador portátil de doble núcleo y 8 Gb. de memoria RAM. Sistema operativo Linux distribución Ubuntu-14. Programa Matlab (MathWorks versión 9.3.0.713579, revisión R2017b), los paquetes Multilayer-L-NL (Martinez-Garcia et al., 2018) y matlabPyrTools, para cálculos matemáticos. Programas Inkscape (versión 0.48) y Gimp (versión 2.8.10) para gráficos. Texmaker (versión 4.1) para edición. La base de datos (390+9 imágenes) ha sido generada a partir de 6 imágenes sin derechos de autor.

3.2. Métodos

A partir de imágenes de cinco físicos famosos y uno que **no** lo es, se genera una base de 390 imágenes, con cambios en luminancia, brillo y contraste para entrenamiento de un clasificador lineal. Además, se añaden otras 9 imágenes del físico desconocido, con distorsiones más extremas, como imágenes test.

El número de escenas es ($n = 6$) correspondientes una a cada físico. Para el test, el número de instancias es ($n = 9$). Para el entrenamiento y el análisis de la convergencia ($n = 390$).

Todas las imágenes se transforman en vectores para facilitar su análisis angular en todos los estados.

Se han seguido seis procesos consecutivos:

- **Primero: Normalización.** Siguiendo el modelo implementado por Martinez-Garcia et al., (2018), se someten todas las imágenes a tres estados de normalización (con etapas Lineal+No Lineal en cada uno), que afectan a la luminancia, brillo y contraste. Posteriormente, otro estado, transforma los datos espaciales (1600 por imagen) hasta descriptores (10025 por imagen) con somatotopía pero escalados en frecuencia, ángulo y amplitud con la transformada y filtro wavelet.

- **Estado 1**

- Etapa Lineal

$$y^1 = I * x^0 \quad (2)$$

- Etapa No Lineal

$$x^1 = scale * \left(b^1 + \beta * \left(\frac{ones(d, d)}{d} \right) * (y^1)^{g^1} + I \right) \frac{(y^1)^{g^1}}{b^1 + \beta * \left(\frac{ones(d, d)}{d} \right) * (y^1)^{g^1}} \quad (3)$$

- Parámetros

$$\theta = \begin{cases} N = & 40 \\ d = & N * N \\ I = & Identidad(N * N) \\ scale = & 255 \\ b^1 = & 0,01 \\ \beta = & 0,9 \\ g^1 = & 1,25 \end{cases} \quad (4)$$

- **Estado 2**

- Etapa Lineal

$$y^2 = (I - H^1) * x^1 \quad (5)$$

- Etapa No Lineal

$$x^2 = \frac{y^2}{b^2 + (H^2 * (I - H^1)^{-1}) + y^2} \quad (6)$$

- Parámetros

$$\theta = \begin{cases} \vdots \\ H^1 = & \text{Jacobiano de la respuesta no lineal anterior} \\ H^2 = & \text{Jacobiano de la respuesta no lineal actual} \\ b^2 = & 30 \\ \vdots \end{cases} \quad (7)$$

- **Estado 3**

- Etapa Lineal

$$y^3 = L * x^2 \quad (8)$$

- Etapa No Lineal

$$x^3 = \text{sign}(y^3) \frac{|y^3|^{g^3}}{b^3 + H^3 * |y^3|^{g^3}} \quad (9)$$

- Parámetros

$$\theta = \begin{cases} \lambda - regul - csf & = 0,5 * 10^{-4} \\ H^3 = & \text{Jacobiano de la respuesta no lineal} \\ L = & Hcsf + \lambda * I(N * N) \\ b^3 = & 0,04 \\ g^3 = & 1,5 \\ \vdots & \end{cases} \quad (10)$$

- Estado 4

- Etapa Lineal

$$y^4 = Transformada\ Wavelet(x^3, fs, ns, no, tw, W) \quad (11)$$

- Etapa No Lineal

$$\begin{aligned} x^4 &= conv(gaussian(Hs, Hc), y^3) \\ x^4 &= (y^3 * gaussian(Hs, Hc))(tw) \\ x^4 &= \int_{-\infty}^{+\infty} y^3(\eta) gaussian(tw - \eta) d\eta \end{aligned} \quad (12)$$

- Parámetros

$$\theta = \begin{cases} ns = & 3 \text{ número de escalas} \\ fs^4 = & fs^2 = 64 \text{ frecuencia muestreo c/grado} \\ no = & 4 \text{ número de orientaciones} \\ tw = & 1 \text{ transición en octavas} \\ W = & L = \text{Trans Wavelet} \\ Hs = & 0,24 \text{ anchura grados gaussiana para enmascarar wavelet} \\ Hc = & 1 \text{ Amplitud de gaussiana} \end{cases} \quad (13)$$

- Segundo: **Comprobar la hipótesis de convergencia.** Mediante análisis estadísticos, comprobar si tras el proceso de normalización, las instancias de una determinada escena convergen.
- Tercero: **Entrenamiento.** Aprendizaje supervisado. Mediante un clasificador lineal se genera un modelo de escena para cada físico. Del modelo se extrae un **invariante** de escena. Se pretende que las puntuaciones de las instancias positivas sean mayores que cero, y las de las negativas menores de cero al aplicar la función:

$$f(\vec{x}) = Bias + \vec{x} \cdot \vec{\beta} \quad (14)$$

Siendo \vec{x} las instancias supervisadas. Si el modelo converge, los parámetros del clasificador ($Bias, \vec{\beta}$) cumplirán este requisito.

- Cuarto: **Test.** Comprobar la efectividad en la predicción para todos los modelos con las imágenes test.
- Quinto: **Aprendizaje.** Construcción del modelo POP (Predictor, Observador, Pensador) de aprendizaje no supervisado. Se describe más adelante. Comprobar el aprendizaje del modelo después del test y cómo han cambiado los **invariantes** de escena. La entrada que recibe es la salida del cuarto estado de normalización descrito.

- Etapa **Lineal**

- Pensador

$$yp^5 = (Bias, \vec{\beta}) / \begin{cases} Bias + \vec{X}^4 \cdot \vec{\beta} > 0, \forall \vec{X}^4 \in (\{XP_o\}^k \cup \{XP_m\}^j) \\ \text{instancias Positivas} & \text{(Observadas o Memorizadas)} \\ \wedge \\ Bias + \vec{X}^4 \cdot \vec{\beta} \leq 0, \forall \vec{X}^4 \in (\{XN_o\}^i \cup \{XN_m\}^l) \\ \text{instancias Negativas} & \text{(Observadas o Memorizadas)} \end{cases} \quad (15)$$

- Observador

$$yo^5 = (Obs(ye^5), \vec{x}^4) / \begin{cases} si \text{ } Obs(ye^5) = V \rightarrow \{XP_o\}^k \cup \vec{X}^4 \\ si \text{ } Obs(ye^5) = R \rightarrow \{XN_o\}^i \cup \vec{X}^4 \end{cases} \quad (16)$$

- Predictor o ejecutor

$$ye^5 = Bias + \vec{x}^4 \quad (17)$$

- Etapa **No Lineal**

- Pensador

$$xo^5 = Modelo(\beta)^m = yp^5 + (\{XP\}^i)^m \cup (\{XN\}^j)^m \quad (18)$$

- Observador

$$xe^5 = \begin{cases} (\{XN_o\}^i)^{xe^5} \cup (\{XP_o\}^k)^{xe^5} & si \text{ } k + i > u \\ \emptyset & si \text{ } k + i \leq u \end{cases} \quad (19)$$

- Predictor o ejecutor

$$xe^5 = \begin{cases} \vec{\beta} & si \text{ } ye^5 > 0 \\ \emptyset & si \text{ } ye^5 \leq 0 \end{cases} \quad (20)$$

• **Parámetros:**

$$\begin{cases} u & \text{umbral para repensar el modelo } m \\ V & \text{señal externa de validación} \\ R & \text{señal externa de refutación} \end{cases} \quad (21)$$

- Sexto: **“Imaginación”**. A partir de cada **invariante** de escena, regresar a zonas donde los datos ofrezcan información visual. Se propone el estado 3 para visualizar el invariante imaginable. Se aplican las funciones inversas sobre el invariante. Se pretende comparar visualmente este modelo con modelos más simples como la media de las instancias positivas para una escena.

$$(\vec{x}^3, \vec{y}^3) = (S^4)^{-1} \circ (S^5)^{-1}(\vec{\beta}) \quad (22)$$

El proceso general antes del estado POP puede verse en la figura (1).

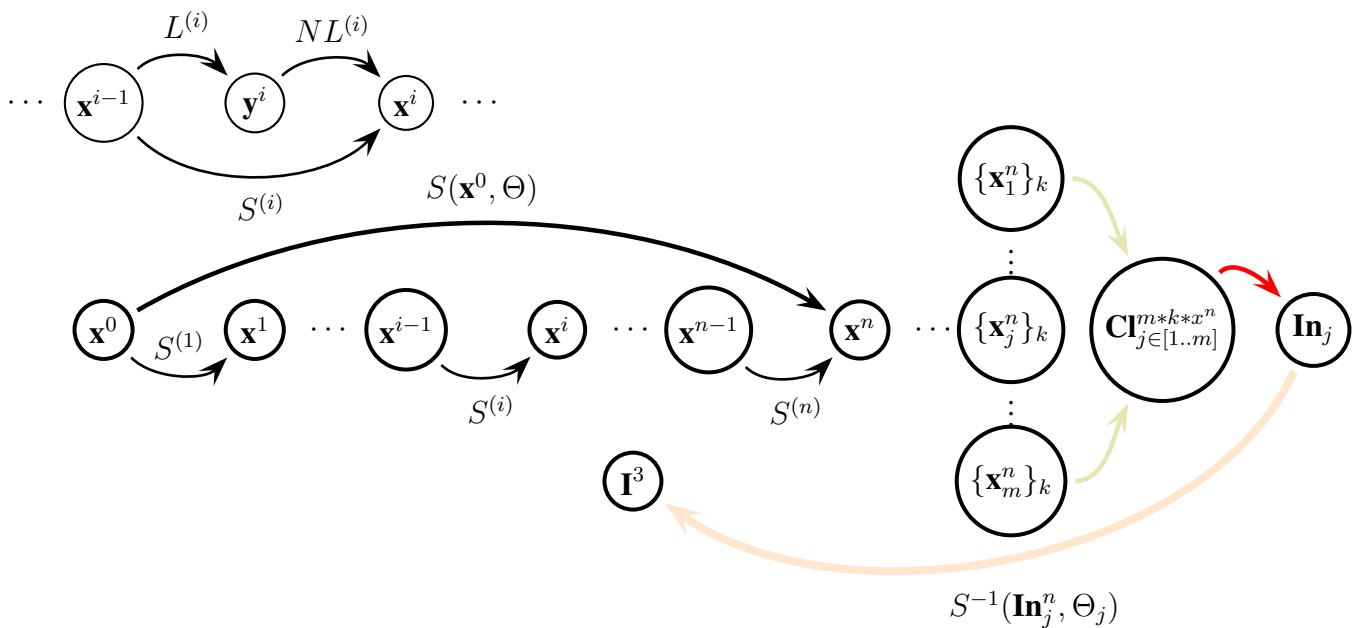


Figura 1: Proceso general.

3.3. Descripción del proceso de aprendizaje (POP) desarrollado

Para poder resolver el problema estático sobre la obtención del invariante, se ha propuesto un modelo de aprendizaje supervisado Russell & Norvig, (2010). En la figura (1) puede observarse la etapa final como un modelo estático de resolución utilizando modelos lineales.

Su transformación a modelo dinámico se ha resuelto como la acción de tres agentes cuyas funciones garantizan el aprendizaje de la experiencia perceptiva.

- El *agente predictor o ejecutor* con una etapa lineal y no lineal resuelve la predicción de escena ante una nueva entrada.
- El *agente observador*, también con ambas etapas, resuelve la contabilización de aciertos y fracasos de las predicciones del ejecutor e insta al agente filósofo a repensar el modelo.
- El *agente pensador o filósofo* transforma el modelo anterior y proporciona un modelo nuevo adaptado a las observaciones percibidas.

El proceso puede verse como tres niveles de un mismo estado, con etapas lineal, no lineal siguiendo la propuesta de Martínez-García et al., (2018), enriquecido con retroalimentación de algunos de sus módulos. Ver figura(2).

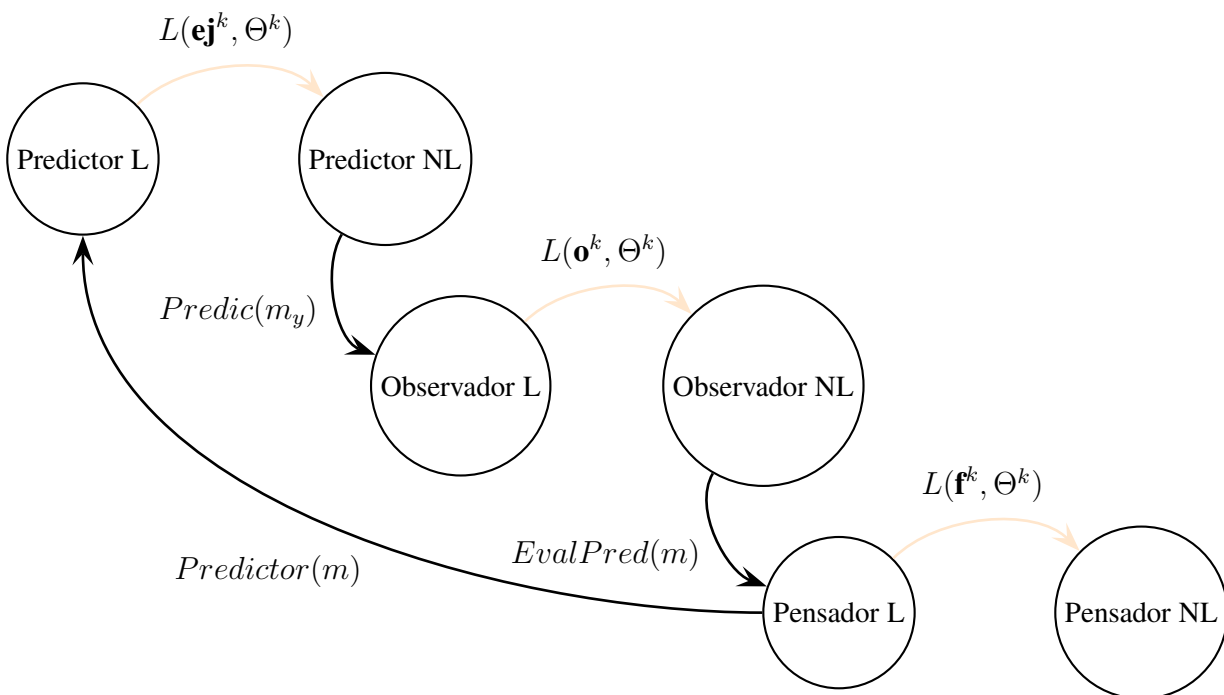


Figura 2: Proceso estado 5. POP.

Para verificar el aprendizaje de este modelo, sobre las entradas de test, el observador ha dado credibilidad a las predicciones del ejecutor (predicador), instando al filósofo (pensador) a repensar el modelo y construir uno nuevo. Se han comprobado los cambios de los invariantes y además su “imaginación” en el estado 3.

Una interpretación más libre de este proceso, solo a efectos didácticos, se muestra en la figura (3). Este modelo no debe leerse literalmente, es una hipótesis sin correlato neuronal en este momento, se muestra para ilustrar el potencial explicativo del modelo POP.

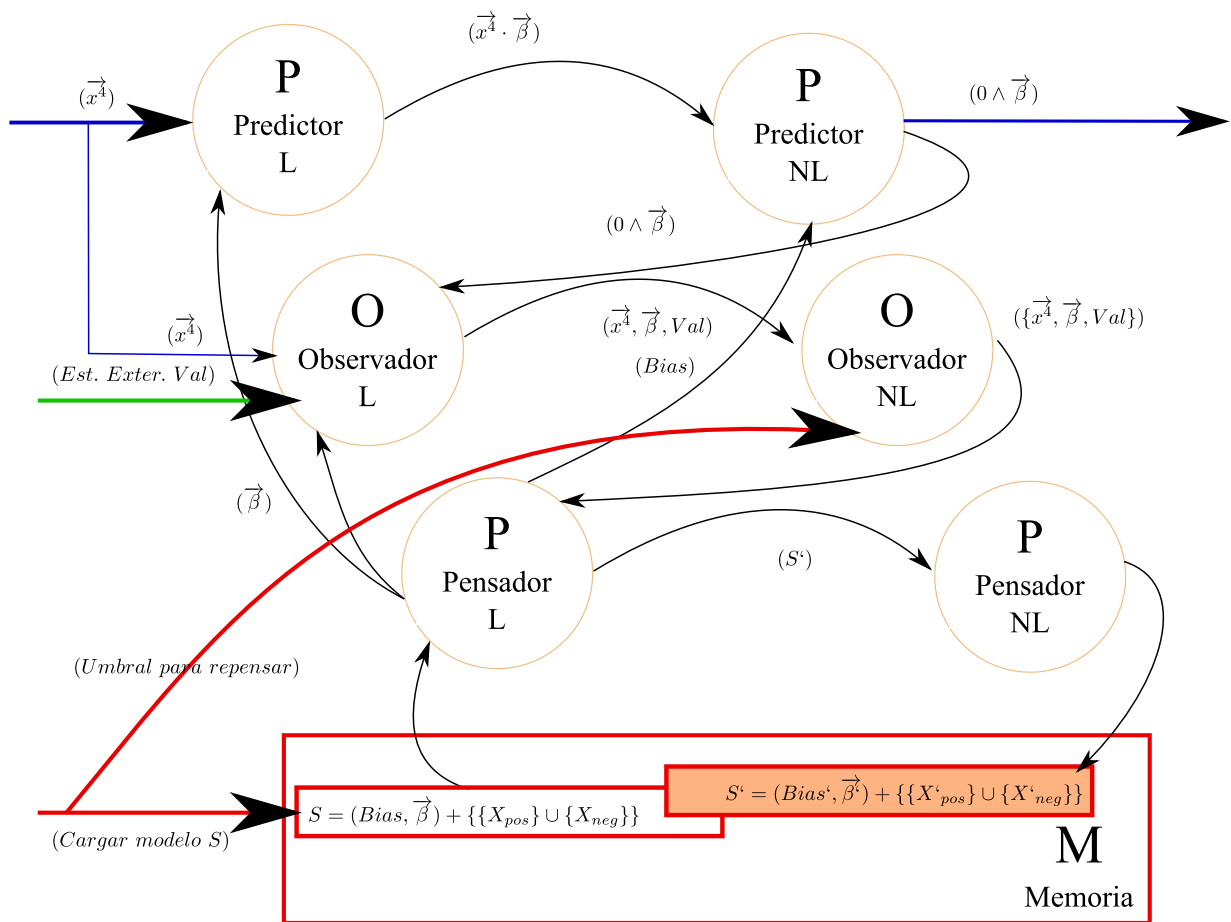


Figura 3: Posible implementación del estado 5. POP.

4. Resultados

Las imágenes de los físicos utilizados para encontrar los invariantes de escena se pueden ver en la figura 8 (8a, 8b, 8c, 8d, 8e, 8f).

Las imágenes positivas para cada escena se han obtenido alterando el contraste y el brillo de la escena original. Véase la figura 4 para observar las alteraciones de la escena del físico Tesla.



Figura 4: Imágenes de entrenamiento de la escena del físico Tesla.

4.1. sobre la hipótesis de convergencia.

Se han analizado los datos desde dos perspectivas diferentes. Un análisis basado en la convergencia angular y un análisis basado en distancias euclídeas. El motivo de este doble análisis radica en la posibilidad de encontrar ángulos iguales entre vectores que muestran diferentes distancias entre ellos y viceversa.

A) Tratamiento angular de las imágenes.

Se han tomado las muestras a la salida de la primera etapa lineal del primer estado (y_1 , retina), a la salida de la segunda etapa no lineal del tercer estado (x_3 , núcleo geniculado lateral NGL) y a la salida de la segunda etapa no lineal del cuarto estado (x_4 , área cortical V1). Los datos corresponden a los vectores de las imágenes de entrenamiento de cada escena, se miden los ángulos (en grados) respecto a un vector hipotético promedio de las instancias de cada escena. Las estadísticas de las imágenes de entrenamiento se muestran en la tabla (1).

Estadísticos de las imágenes de entrenamiento						
	Retina: y_1		NGL: x_3		V1: x_4	
Físico	media	σ	media	σ	media	σ
Desconocido	5,5426	2,623	3,7177	1,8719	3,9545	1,8656
Einstein	4,6663	1,7936	10,9195	3,4378	9,9389	3,3325
Feynman	5,3706	1,6771	5,4991	3,2143	6,9764	3,858
Fermi	7,9068	2,3317	9,185	3,7252	9,7561	4,0978
Tesla	4,4946	1,5301	7,2646	4,2121	7,0963	4,7598
Hawking	9,658	3,7641	8,3893	3,3887	8,7538	3,755

Cuadro 1: Estadísticos de imágenes de entrenamiento. Ángulos en grados.

B) Tratamiento métrico de distancias euclídeas entre las imágenes.

Se toman muestras en los mismos estados que el apartado anterior. Los datos corresponden a los vectores de las imágenes de entrenamiento de cada escena, se miden las distancias con métrica euclídea de cada vector, respecto a un vector hipotético promedio de las instancias de cada escena. Las estadísticas de las imágenes de entrenamiento se muestran en la tabla (2).

Estadísticos de las imágenes de entrenamiento						
	Retina: y_1		NGL: x_3		V1: x_4	
Físico	media	σ	media	σ	media	σ
Desconocido	2,8092	1,5688	2,4453	1,2212	1,0918	0,5135
Einstein	6,8943	2,4665	6,5966	2,0733	2,5946	0,8456
Feynman	4,3738	1,4165	3,6772	2,0972	1,8679	1,02
Fermi	5,421	1,8307	5,8169	2,3156	2,5414	1,0729
Tesla	6,4278	2,1615	4,8942	2,7837	1,9997	1,3003
Hawking	5,1158	2,2303	5,2749	2,1541	2,3069	0,9932

Cuadro 2: Estadísticos de imágenes de entrenamiento. Distancias euclídeas.

4.2. sobre el aprendizaje estático supervisado.

Tras la primera fase de aprendizaje estático, el sistema obtiene los invariantes de escena de cada físico y además, clasifica correctamente las instancias de test, ver figura (9), con una efectividad del 100 %. El reto matemático planteado en el experimento era simple. La figura (5) muestra las puntuaciones de cada instancia del test para cada físico. Los ángulos entre cada uno de los invariantes se muestran en la tabla (3). Las instancias de test se obtienen tras aplicar diferentes distorsiones sobre el físico desconocido. La figura (9) muestra las imágenes de las distorsiones de saturación (9a, 9b, 9c), la de brillo (9d), la de convolución (9e), la de dilatación (9f), la de erosión (9g), la de óleo (9h) y la de bordes (9i).

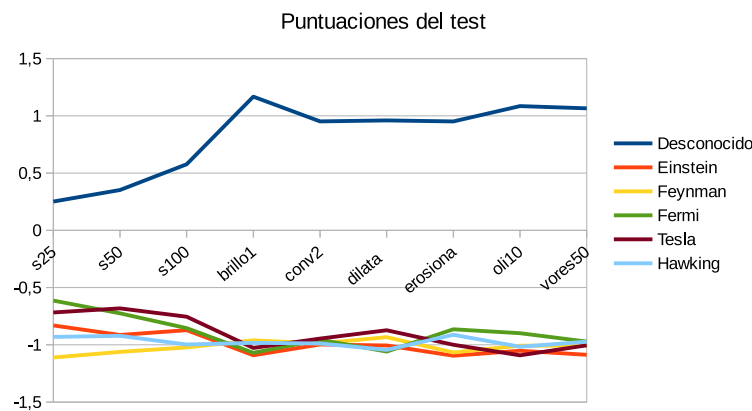


Figura 5: Puntuaciones

Ángulos entre invariantes tras el entrenamiento						
Físico	Desconocido	Einstein	Feynman	Fermi	Tesla	Hawking
Desconocido	0	91.8348	88.6791	95.6051	133.0996	100.5658
Einstein		0	102.0550	107.3804	99.2938	107.9364
Feynman			0	115.6542	93.2869	96.5370
Fermi				0	96.2761	96.8900
Tesla					0	97.7684
Hawking						0

Cuadro 3: Ángulos (en grados) entre invariantes tras entrenamiento.

4.3. sobre el aprendizaje dinámico semi-supervisado.

Tras la segunda fase de aprendizaje dinámico con el proceso POP, el sistema reaprende el “pre-concepto” de invariante de escena de cada físico y se puede observar cómo los vectores invariantes se han transformado de tal manera que acomodan mejor las nuevas observaciones. Los parámetros de los modelos lineales y los cambios de ángulo (en grados) respecto al invariante anterior se muestran en la tabla (4). Los ángulos entre cada uno de los invariantes se muestran en la tabla (5).

Este cambio afecta a los ángulos relativos de las instancias test respecto a los invariantes. En el caso del físico no famoso se aprecia un menor ángulo (más cercano a cero, dentro de $|\pi/2|$), y en los

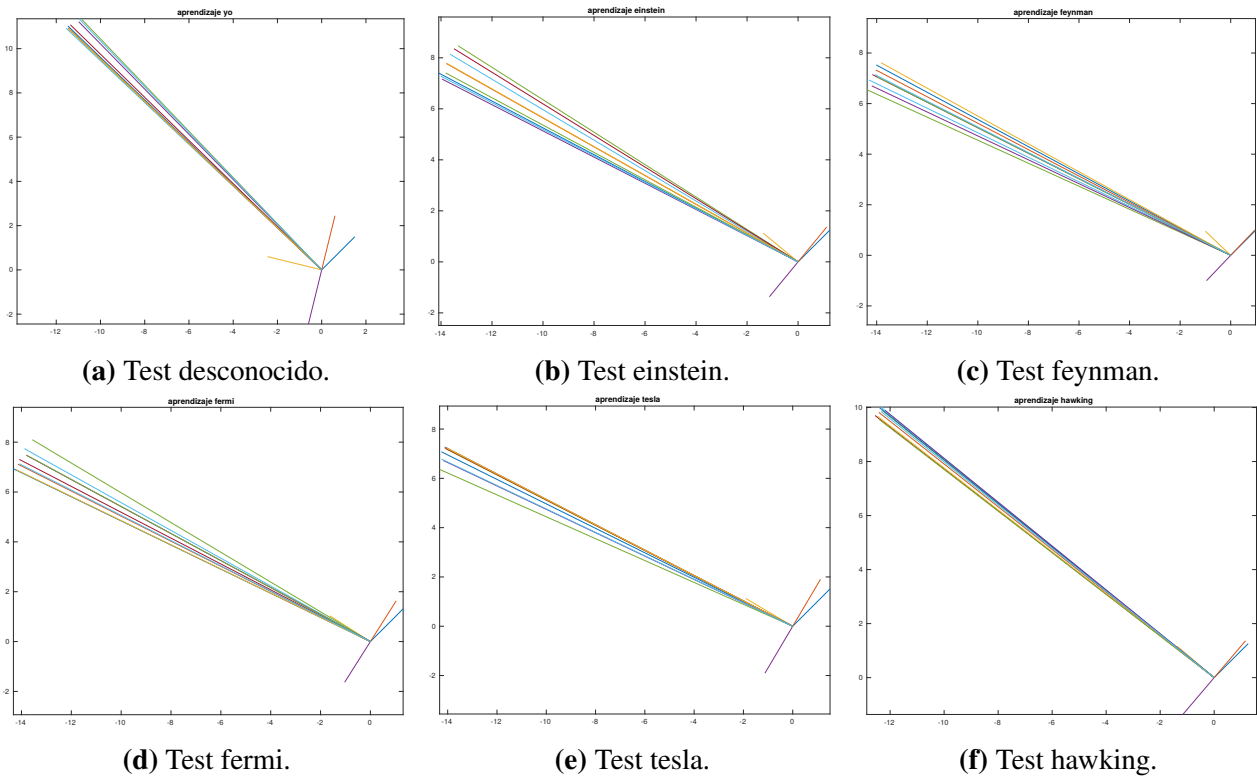
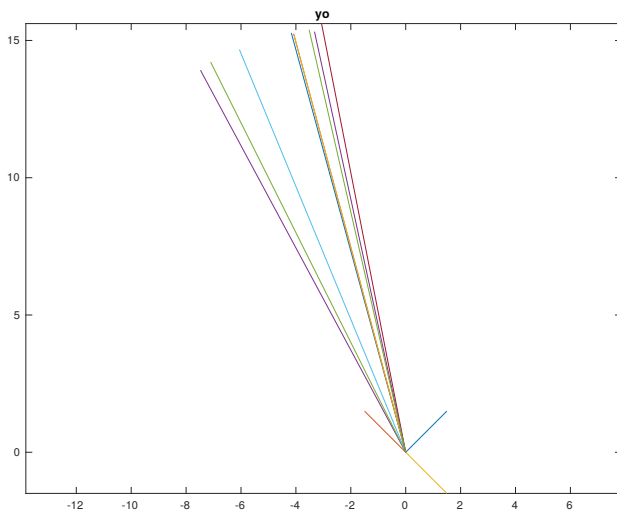


Figura 6: Vectores del test sobre invariantes, antes y después de aprendizaje. Se ha marcado los ángulos π y $\pi/2$ relativos al invariante aprendido. El vector azul es el invariante entrenado, el rojo el aprendido. Los ángulos de los vectores del test son respecto al vector aprendido, una proyección sobre el plano que forma cada vector con el invariante. Por tanto, los vectores representados oscilan en $10025 - 2$ grados de libertad.

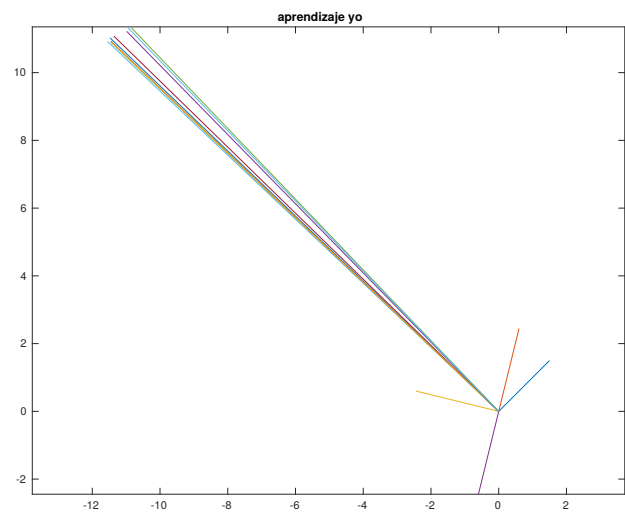
casos de físicos famosos, una tendencia hacia $|\pi/2|$ o mayor. Se puede observar, en la figura (6) las imágenes (6a, 6b, 6c, 6d, 6e, 6f), todos los modelos. Con más detalle, se puede observar los efectos en los ángulos para el físico desconocido y para Tesla, antes y después de aprendizaje, en la figura (7) las imágenes (7a, 7b) y (7c, 7d) respectivamente. Los vectores azules con ángulo de $\pi/4$ corresponden al invariante de cada escena después del entrenamiento. El vector rojo corresponde al invariante después del aprendizaje. En las gráficas se han insertado vectores a π y $\pi/2$ del invariante de escena, en el caso de entrenamiento, y del invariante aprendido en los demás casos. Se pretende visualizar los ángulos que forman los vectores test con los invariantes rápidamente. Hay que tener en cuenta que los ángulos representados en estos planos están proyectados de las 10025 dimensiones del hiperespacio de los descriptores, por tanto, hay que verlos como un hipercono de $10025 - 2$ dimensiones.

4.4. sobre la “Imaginación”.

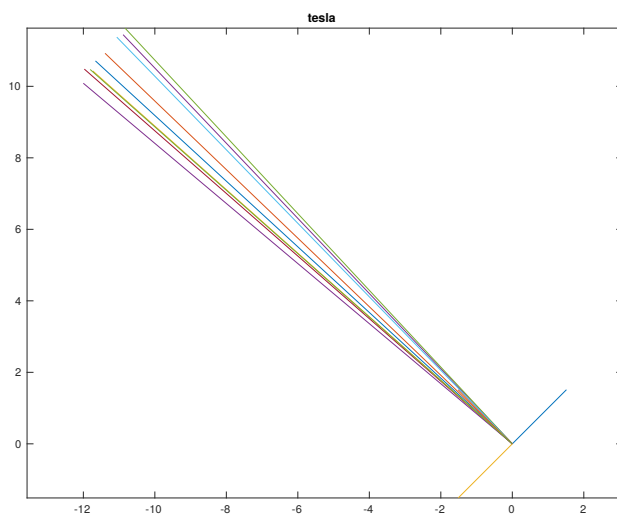
Tras la tercera fase, se observa la “imaginación” del invariante antes y después del aprendizaje dinámico, en el estado 3. También las diferencias con otros modelos que no utilizan discriminación como las medias de las imágenes. Véase la figura (8), imágenes (8h, 8i, 8j) respectivamente. Se pueden apreciar sus diferencias, y las diferencias de estas con la imagen modelo normalizada de la vía de aferencias del sistema visual en la figura (8), imagen (8g), se muestran las salidas de la etapa lineal y no lineal. Véase la figura (8), imágenes (8a, 8b, 8c, 8d, 8e, 8f) para comparar estas diferencias



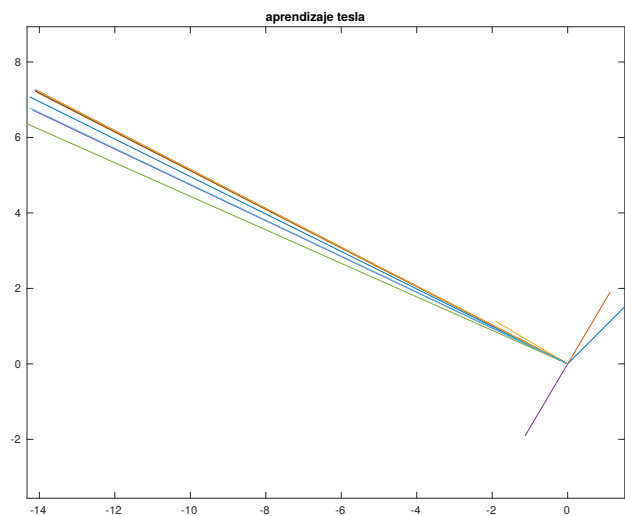
(a) Test desconocido antes de aprendizaje.



(b) Test desconocido después de aprendizaje.



(c) Test tesla antes de aprendizaje.



(d) Test tesla después de aprendizaje.

Figura 7: Detalle. Vectores del test sobre invariantes, antes y después de aprendizaje. Se ha marcado los ángulos π y $\pi/2$ relativos al invariante aprendido, o al invariante entrenado, si todavía no ha habido aprendizaje. El vector azul es el invariante entrenado, el rojo el aprendido. Los ángulos de los vectores del test son respecto al vector aprendido, una proyección sobre el plano que forma cada vector con el invariante. Por tanto, los vectores representados oscilan en $10025 - 2$ grados de libertad.



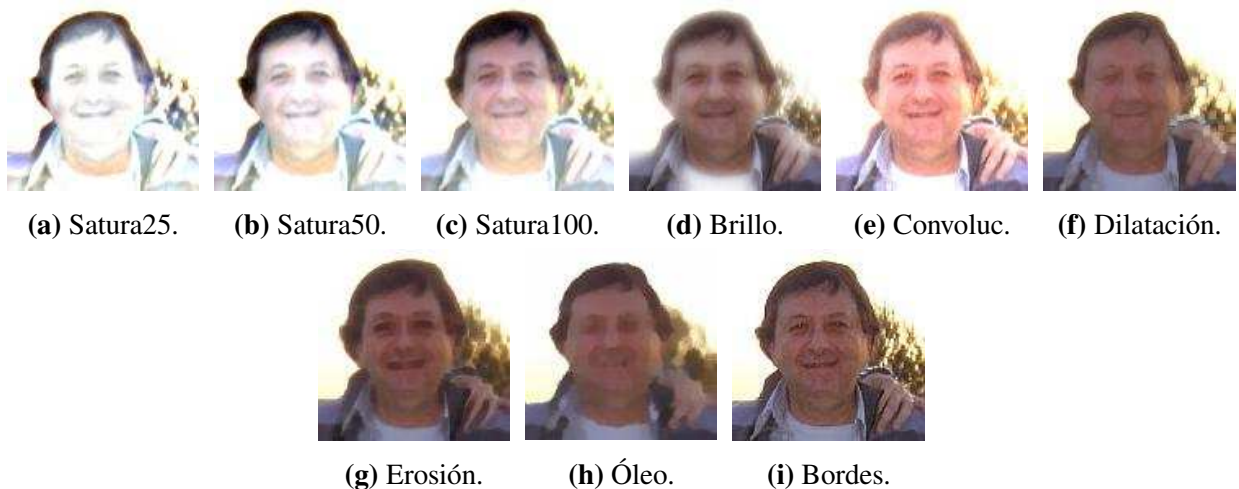
Figura 8: Físicos. Observado e imaginado en el tercer estado.

con las entradas originales de las escenas.

Parámetros de los modelos lineales					
Físico	Entrenamiento		Δ Ángulo	Aprendizaje	
	Bias	Norma		Bias	Norma
Desconocido	-0,7361	0,212	31,2304	-1,0116	0,2519
Einstein	-0,6382	0,1763	5,5524	-0,6427	0,177
Feynman	-0,4182	0,1385	1,3618	-0,4321	0,1385
Fermi	-0,9441	0,1857	12,6546	-0,9004	0,1933
Tesla	-0,8254	0,2141	14,3706	-0,768	0,2214
Hawking	-0,9161	0,1782	4,612	-0,9187	0,1791

Cuadro 4: Parámetros.

Ángulos entre invariantes tras el aprendizaje						
Físico	Desconocido	Einstein	Feynman	Fermi	Tesla	Hawking
Desconocido	0	94.7494	91.6644	103.7709	132.0919	101.3827
Einstein		0	101.8718	104.6913	95.7644	106.7143
Feynman			0	95.9838	92.6830	95.9838
Fermi				0	90.9758	94.9122
Tesla					0	97.7684
Hawking						0

Cuadro 5: Ángulos (en grados) entre invariantes tras aprendizaje.**Figura 9:** Imágenes test. Distorsiones del físico desconocido.

5. Discusión

5.1. del correlato neuronal.

Las diferentes partes o estados del presente trabajo tienen un correlato neuronal asociado, discutido en los diferentes artículos referenciados en la introducción. No se discute su contenido, pero se extraen los correlatos neuronales de las funciones expuestas en este trabajo:

- **Retina.** Estados 1 y 2, normalización de luminancia, brillo y contraste local.
- **Núcleo Geniculado Lateral (NGL).** Estado 3, filtro y máscara de contraste. Visión normalizada.
- **Área V1.** Estados 4 y 5. Campo receptivo visual. Aprendizaje del invariante.
- **Núcleo Geniculado Lateral (NGL).** Visión normalizada. “Imaginación” del invariante.

El proceso global puede verse en el esquema de la figura (10).

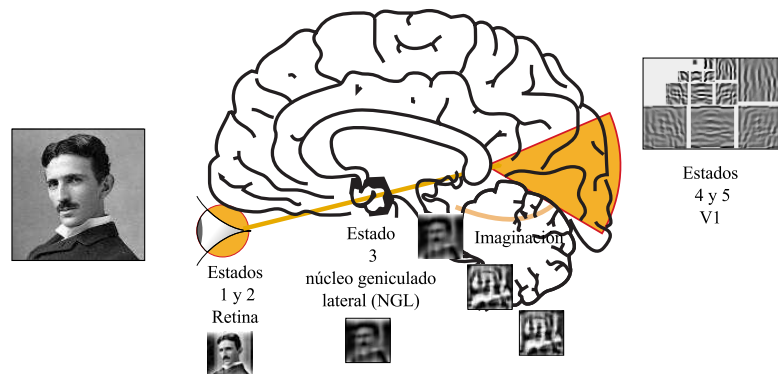


Figura 10: Ruta y correlato neuronal.

En la figura 11, subfigura (11b) se pueden apreciar los campos receptivos de V1 calculados para la escena del físico Tesla. Las 3 escalas y los 4 ángulos utilizados en el cálculo (0, 45, 90, 135 en grados). En la figura 11, subfigura (11a), se muestran los campos receptivos reales de un macaco publicados por Ringach, (2002).

5.2. de la hipótesis de convergencia.

Tras el análisis angular y de distancias métricas, no podemos aceptar la hipótesis nula de la igualdad de varianzas, por lo que en las imágenes de entrenamiento hay algunas varianzas diferentes ($p = 0$ en el test de igualdad de varianza de Barlett para todas las muestras).

Cuando se analiza por escena, para la del físico desconocido, con los datos a la salida de $y1$, $x3$ y $x4$ (que cubre toda la vía retina-cortex V1) rechazamos la hipótesis nula de igualdad de varianzas con un ($p = 0,0062$) para ángulos y ($p = 0$) para distancias, con las salidas $x3$, $x4$ (NGL-V1) no

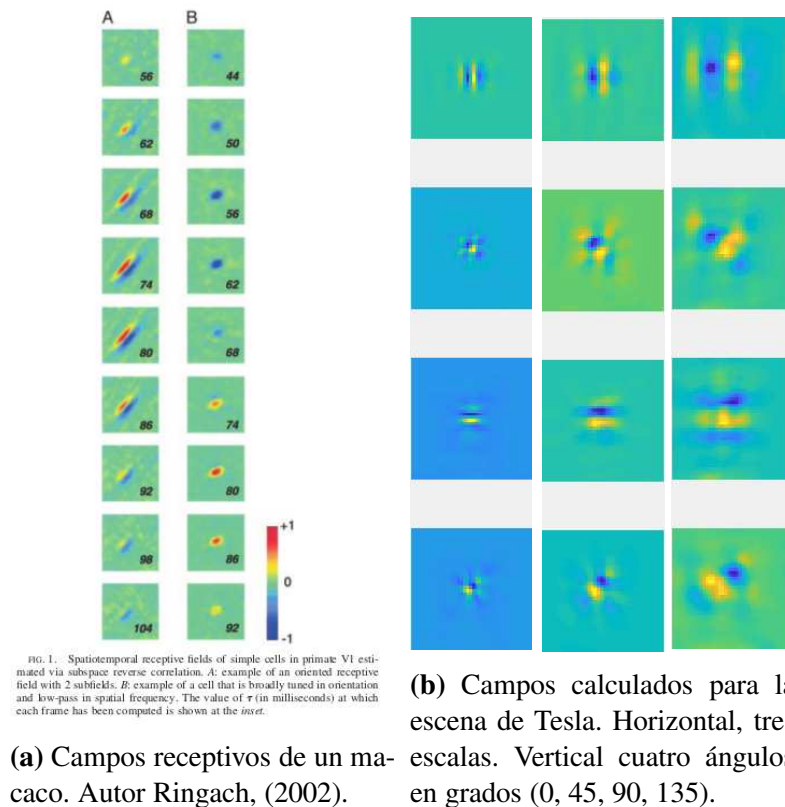


Figura 11: Campos receptivos en V1. Reales en macaco (a). Calculados para la escena de Tesla (b).

podemos rechazar la hipótesis nula con un ($p = 0,979$) para ángulos y se rechaza ($p = 0$) para distancias, con las salidas $y1, x3$ (retina-NGL), varianzas diferentes ($p = 0,0082$) para ángulos e iguales ($p = 0,0488$) para distancias, con las salidas $y1, x4$ (retina-V1) varianzas diferentes ($p = 0,0076$) para ángulos y ($p = 0$) para distancias.

Para la escena del físico Einstein, no podemos rechazar que las salidas $x3, x4$ tengan la misma varianza ($p = 0,8058$) para ángulos y sí ($p = 0$) para distancias, las demás muestran varianza diferente, no se puede aceptar la hipótesis nula con significación ($p = 0$) para ángulos. Para distancias, varianzas iguales ($p = 0,1708$) en las salidas $y1, x3$.

Para la escena del físico Feynman, no podemos rechazar que las salidas $x3, x4$ tengan la misma varianza ($p = 0,1501$) para ángulos. Tampoco las salidas $y1, x4$ para distancias con ($p = 0,0101$) siendo estrictos ($\alpha = 1\%$).

Para la escena del físico Fermi, no podemos rechazar que las salidas $x3, x4$ tengan la misma varianza ($p = 0,4514$) para ángulos. Tampoco las salidas $y1, x3$ con ($p = 0,0645$) para distancias.

Para la escena del físico Tesla, no podemos rechazar que las salidas $x3, x4$ tengan la misma varianza ($p = 0,3344$) para ángulos. Tampoco las salidas $y1, x3$ con ($p = 0,0466$) para distancias.

Para la escena del físico Hawking, no podemos rechazar que las salidas $y1, x3, x4$ tengan la misma varianza ($p = 0,6446$) para ángulos, y por tanto, cualquiera de los pares, las salidas $x3, x4$ con ($p = 0,4174$), las salidas $y1, x3$ con ($p = 0,4066$) y las salidas $y1, x4$ con ($p = 0,9848$). Tampoco las salidas $y1, x3$ con ($p = 0,7835$) para distancias.

Las varianzas, o mejor su raíz, las desviaciones típicas, que observamos diferentes son:

- Para la escena del físico desconocido, considerando el análisis angular, las de y_1, x_3 e y_1, x_4 . Pasan de 2, 623 (y_1) a 1, 8719 (x_3) y 1, 8656 (x_4). Además, las medias de los ángulos respecto al centro de la distribución de cada estado pasan de 5, 5426 grados (y_1) a 3, 7177 (x_3) y 3, 9545 (x_4). Sí que se puede inferir un mayor grado de convergencia en los estados finales.

Para el análisis de distancias, las de x_3, x_4 e y_1, x_4 . Pasan de 2, 8092 (y_1) a 2, 4453 (x_3) y 1, 0918 (x_4). Además, las medias de los ángulos respecto al centro de la distribución de cada estado pasan de 1, 5688 (y_1) a 1, 2212 (x_3) y 0, 5135 (x_4). También se puede inferir un mayor grado de convergencia en los estados finales.

- Para la escena del físico Einstein, considerando el análisis angular, las de y_1, x_3 e y_1, x_4 . En (y_1) 1, 7936, en (x_3) 3, 4378 y en (x_4) 3, 3325. Las medias de los ángulos pasan de 4, 6663 grados (y_1) a 10, 9195 (x_3) y 9, 9389 (x_4). Se observa una dispersión en los datos.

Para el análisis de distancias, las de y_1, x_3 e y_1, x_4 . En (y_1) 2, 4665, en (x_3) 2, 0733 y en (x_4) 0, 8456. Las medias de las distancias pasan de 6, 8943 grados (y_1) a 6, 5966 (x_3) y 2, 5946 (x_4). Se observa convergencia en las distancias.

- Para la escena del físico Feynman, considerando el análisis angular, las de y_1, x_3 e y_1, x_4 . En (y_1) 1, 6771, en (x_3) 3, 2143 y en (x_4) 3, 858. Las medias de los ángulos pasan de 5, 3706 grados (y_1) a 5, 4991 (x_3) y 6, 9764 (x_4). También se observa una dispersión en los datos.

Para el análisis de distancias, las de y_1, x_3 e y_1, x_4 . En (y_1) 1, 4165, en (x_3) 2, 0972 y en (x_4) 1, 02. Las medias de las distancias pasan de 4, 3738 grados (y_1) a 3, 6772 (x_3) y 1, 8679 (x_4). Se observa convergencia en las distancias.

- Para la escena del físico Fermi, considerando el análisis angular, las de y_1, x_3 y las de y_1, x_4 . En (y_1) 2, 3317, en (x_3) 3, 7252 y en (x_4) 4, 0978. Las medias de los ángulos pasan de 7, 9068 grados (y_1) a 9, 185 (x_3) y 9, 7561 (x_4). Se observa una dispersión en los datos.

Para el análisis de distancias, las de y_1, x_4 y x_3, x_4 . En (y_1) 1, 8307, en (x_3) 2, 3156 y en (x_4) 1, 0729. Las medias de las distancias pasan de 5, 421 grados (y_1) a 5, 8169 (x_3) y 2, 5414 (x_4). Se observa convergencia en las distancias, si tenemos en cuenta que y_1 y x_3 muestran varianzas iguales en el test de Barlett.

- Para la escena del físico Tesla, considerando el análisis angular, las de y_1, x_3 y las de y_1, x_4 . En (y_1) 1, 5301, en (x_3) 4, 2121 y en (x_4) 4, 7598. Las medias de los ángulos pasan de 4, 4946 grados (y_1) a 7, 2646 (x_3) y 7, 0963 (x_4). Se observa una dispersión en los datos.

Para el análisis de distancias, las de y_1, x_4 y x_3, x_4 . En (y_1) 2, 1615, en (x_3) 2, 7837 y en (x_4) 1, 3003. Las medias de las distancias pasan de 6, 4278 grados (y_1) a 4, 8942 (x_3) y 1, 9997 (x_4). Se observa convergencia en las distancias.

- Para la escena del físico Hawking, considerando el análisis angular, todas son desviaciones típicas iguales, las de y_1, x_3, x_4 , las de y_1, x_3 , las de x_3, x_4 y las de y_1, x_4 . En (y_1) 3, 7641, en

(x_3) 3, 3887 y en (x_4) 3, 755. Las medias de los ángulos respecto al centro de la distribución de cada estado pasan de 9, 658 grados (y_1) a 8, 3893 (x_3) y 8, 7538 (x_4). Se observa una pequeña convergencia, no significativa, tanto en la desviación típica como en la media de los ángulos.

Para el análisis de distancias, las de y_1 , x_4 y x_3 , x_4 . En (y_1) 2, 2303, en (x_3) 2, 1541 y en (x_4) 0, 9932. Las medias de las distancias pasan de 5, 1158 grados (y_1) a 5, 2749 (x_3) y 2, 3069 (x_4). Se observa convergencia en las distancias.

En resumen, para el análisis de distancias, sí se observa una convergencia con el avance de los estados para todas las escenas. Mientras, en el análisis angular, para la escena del físico desconocido, se aprecia una convergencia angular del primer al último estado, para Hawking igualdad de varianzas y para Einstein, Feynman, Fermi y Tesla una dispersión generalizada. En consecuencia, dado que los análisis de distancias contienen información tanto del ángulo como del módulo de los vectores, aportan más información que el análisis angular. Y, aunque los análisis angulares no confirman la hipótesis de la convergencia, sí lo hacen los análisis de distancias tras los cuatro primeros estados lineales+no lineales.

Esta convergencia del análisis de distancias implica que las entradas sensoriales de una determinada escena, a medida que avanzan por el sistema óptico (retina, NGL, V1), las excitaciones de las poblaciones neuronales van pareciéndose más entre ellas, son más similares, las distancias entre ellas son más cortas, y su dispersión menor.

El camino hacia el invariante de escena parece ser la normalización. Su forma más sencilla sería tomarlo como la media de las entradas de una determinada escena, pero, ¿qué significado tiene si no es frente a otra escena? En este sentido, se ha desarrollado un estado posterior también con etapas lineal y no lineal que resuelva el invariante de escena, tomando como datos de entrada la salida del último estado de normalización, convertido a espacio de direcciones y frecuencias. Este estado tiene en cuenta el significado de la escena frente a otras escenas.

5.3. del módulo “sistema de aprendizaje de escena” (POP) desarrollado. Breve reflexión.

Ya se introdujo de forma funcional el modelo POP en (v. supra subsección 3.3, p. 11). Además se propuso una posible implementación del modelo en (v. supra figura 3, p. 13). Ahora se argumentará su oportunidad en el escenario del presente TFM.

En primer lugar, hay bibliografía científica que avala los modelos lineales en V1, ya se citó en la introducción a (Coen-Cagli & Schwartz, 2013). En este sentido, el modelo POP se basa en modelos de clasificación lineal similares, pero organizados de forma que permiten aprendizaje dinámico.

Y en segundo lugar, se describirán las bases que sustentan el modelo, empezando por la estrategia de Allen, siguiendo con la descripción funcional del POP y terminando con la extracción de elementos del modelo determinantes para la discusión que nos ocupa, el invariante de escena.

Para poder atribuir capacidad conceptual a un ente “O”, Allen propone una estrategia general

consistente en tres principios.

“A un organismo “O” podemos razonablemente atribuir el concepto X siempre que:

- (i) “O” discrimina sistemáticamente algunas instancias X de las no-X,
- (ii) “O” es capaz de detectar algunos de sus propios errores de discriminación entre X y no-X, y
- (iii) “O” es capaz de aprender a discriminar mejor X de no-X como consecuencia de esta capacidad. “

Allen, (1999, p. 37)

El módulo desarrollado cumple (i) e (iii), pero es incapaz de detectar por medios propios sus errores. Sería absurdo. Para poder detectarlos, debería cometer errores a propósito (i), para poder autodetectarlos (ii) con sus propios medios. La apuesta es que la confirmación se ha de realizar por señales externas al módulo que verifiquen o refuten la predicción, pero además, ha de indicarse la fuerza de esta confirmación. Es el primer escalón hacia el aprendizaje.

¿Qué tipo de mecanismos subyacen en este módulo? A nivel funcional podemos distinguir tres bloques:

- (i) *uno ejecutivo*, consistente en lanzar una predicción, dada una instancia de entrada y un modelo sobre el que evaluarla. Es el *ejecutor* de la predicción. Su función y nombre (P) de predictor.
- (ii) *uno observacional*, consistente en recoger evidencias externas que confirmen o refuten la predicción, además, con grado de relevancia para el organismo. Es el *observador*, independiente en sus decisiones. Su función y nombre (O) de observador.
- (iii) *uno filosófico*, consistente en meditar las consecuencias acarreadas por la predicción y repensar el modelo y el sistema donde yace el modelo. Es el *filósofo* el que decide ¿qué cambiar? y ¿cómo cambiar?, tanto el modelo como el sistema. El nuevo sistema se pondrá a disposición del ejecutor para realizar nuevas predicciones. Su función y nombre (P) de pensador.

El predictor que el filósofo (P) entrega al ejecutor es simple, para que su decisión sea rápida. No es admisible que ante una determinada entrada, el ejecutor o predictor (P) dude sobre su clasificación, su predicción va a ser la base de las posibles acciones que el sistema emprenda, y hay algunas de ellas que son cruciales para la supervivencia del organismo. Ejecutada una predicción, el observador (O) ha de encargarse de confirmar o refutar su veracidad (entendida esta como satisfacción), además de la relevancia de la predicción para el sistema y el organismo. Finalmente, con el modelo de sistema anterior (S) y las nuevas observaciones realizadas, el filósofo o pensador (P) elabora un nuevo modelo (S'), capaz de recoger las evidencias facilitadas por el observador (O), que realice mejores predicciones, tanto del “preconcepto” como del sistema en general, que apunten hacia una mayor satisfacción del organismo.

El modelo es compatible con la “labilidad del sistema de memoria” y su importancia respecto al aprendizaje. En este sentido, para la meditación del filósofo se debe utilizar, y por tanto extraer de la memoria, el modelo de sistema (S) anterior, para que, junto con las nuevas observaciones facilitadas, se elabore un nuevo modelo de sistema (S’) que será guardado en sustitución del antiguo (S). Véase (supra figura 3, p. 13).

El módulo desarrollado sigue la filosofía propuesta por autores actuales en este campo, entre otros, Martínez-García et al., (2018), en los dos sentidos que proponen:

- Ha de contener una fase lineal y una no lineal para completar un estado.
- Ha de ser una computación simple y extrapolable a otros sistemas neuronales que precisen esta función.

Por tanto, el módulo es un estado que, recibe las aferencias de la etapa anterior, pero además recibe aferencias de estados posteriores que imponen al filósofo o pensador (P) y en consecuencia al ejecutor o predictor (P) qué modelos debe utilizar para realizar sus predicciones. También recibe aferencias de otros sistemas hacia el observador (O): *primero*, para valorar la veracidad de la predicción así como su relevancia para el organismo, y *segundo*, el umbral de observaciones acumuladas para que avise al filósofo y este pueda meditar sobre los datos de memoria del sistema. El estado elabora eferencias hacia estados superiores de distinto orden. *Primero* una respuesta dicotómica sobre si la instancia que está mirando “es” o “no es” un determinado escenario, para su posible utilización en estados de acción o procesamiento posteriores. *Segundo* una eferencia hacia la memoria con el nuevo modelo de sistema (S’) cuando proceda.

La fase lineal se encarga de:

- (i) a nivel *ejecutivo*, puntuar la instancia de entrada sobre el modelo de escena.
- (ii) a nivel *observacional*, puntuar las predicciones con las aferencias de módulos externos. Decidir si la predicción del ejecutor ha sido “acertada” o no para el organismo, así como su peso.
- (iii) a nivel *filosófico*, reelaborar nuevos modelos lineales que satisfagan mejor las necesidades y preferencias del organismo. Recuperar modelos en memoria.

La fase no lineal se encarga de:

- (i) a nivel *ejecutivo*, a partir de la puntuación obtenida por el predictor en la fase lineal, elaborar una respuesta dicotómica sobre si la entrada sensorial “es” o “no es” una instancia del modelo de escena.
- (ii) a nivel *observacional*, tras un umbral, facilitar la información de las observaciones al agente filósofo.

- (iii) a nivel *filosófico*, elaborar el mecanismo de decisión dicotómico que utilizará el ejecutor en el predictor de cada modelo. Guardar modelos en memoria.

Concluyendo, sobre la hipótesis de trabajo utilizada que predecía la existencia de un invariante de escena en el sistema visual, este módulo es capaz de reconocer y aprender la escena, y en este sentido, contiene el invariante.

El modelo de escena se ha implementado como un clasificador lineal, por tanto, los parámetros del vector beta, el factor de escala y el bias lo describen por completo.

En primer lugar, el factor de escala aporta información intrínseca al sistema para adecuar la salida dentro de unos límites preestablecidos, es lineal a todos los parámetros beta, por lo que no aporta más información sobre la estructura de la escena. *En segundo lugar*, el bias es una translación en las puntuaciones de salida para modular el paso por cero del modelo, es decir, ayuda a determinar cuándo debe considerarse una puntuación suficiente para realizar una predeción positiva o negativa. Tampoco aporta información relevante sobre la configuración en las entradas al sistema. Y, *en tercer lugar*, el vector de coeficientes beta interacciona directamente sobre los parámetros de entrada de la instancia en el sistema, en concreto, la interacción sirve para puntuar el grado de similitud de la instancia con el modelo, en forma de producto escalar. Es aquí donde encontramos un invariante de escena con información relevante:

- *Primero*: Por las características del producto escalar, el vector de entrada y el vector beta obtendrán una puntuación de similitud máxima cuando el ángulo que los separa sea mínimo.
- *Segundo*: Con el mismo argumento anterior, obtendrán una similitud irrelevante cuando el ángulo de separación sea lo más próximo a $|\pi/2|$. Incluso negativa si superan estos ángulos.
- *Tercero*: Como consecuencia, podemos ver que un hipotético vector de entrada con ángulo cero y módulo distinto de cero sería el que mejor representaría la escena en el hiperespacio de los descriptores de entrada. Este vector puede escalarse linealmente en módulo hasta que su representación en estados anteriores sea significativo y válido para ofrecer una representación visual aceptable, antes de las etapas de transformación sufridas por los descriptores, sin pérdida de generalidad.

5.4. del invariante como media de las instancias positivas observadas por el sistema sensorial frente al invariante discriminador.

¿Se puede hablar de escenas puras, o lo son respecto a otras escenas? Si se habla de escenas puras, el invariante, visto como asociación de las percepciones sobre esa escena (por ejemplo una media) nos dará una imagen más nítida, no sobre lo que es, sino sobre lo que percibimos que es. Si se habla de escenas frente a otras escenas, el invariante debe contener tanto lo que percibimos que es, como lo que no es. Su fundamento es discriminador. Si se toma el invariante como la media de las percepciones catalogadas positivas para esta escena, ante cualquier entrada necesitaremos de

medidas de distancias psicológicas para saber su grado de similitud, como las distancias son siempre positivas, también lo es su inverso, y, en consecuencia, cualquier entrada puede ser catalogada como perteneciente a una escena con una mayor o menor similitud. En cambio, si el invariante se toma del modelo de clasificación lineal, su medida de similitud discrimina la pertenencia de una instancia a una escena con mayor precisión. Sobretodo para las escenas a las que se enfrenta en la construcción del clasificador.

Este debate puede extenderse sin ganadores ni perdedores. Simplemente se puede considerar que se guardan los datos y se reconstruye el clasificador yuxtaponiendo las escenas que, en un determinado momento, el organismo desee evaluar. Pero no nos podemos librar de preguntar ¿qué observador, y en base a qué ha decidido la veracidad de una predicción? Y si el modelo es de clasificación lineal es fácil responder, pero si el modelo es una simple media de las instancias ¿qué veracidad se obtiene de la observación de la predicción? ¿Realmente hay aprendizaje?

Una posible solución adoptaría los dos enfoques y los utilizaría dinámicamente en función del entorno donde deban ser ejecutados. El modelo POP propuesto puede soportar sin ambigüedad estos enfoques. Es decir, se pueden extraer las instancias positivas del clasificador y obtener su media para visualizar un invariante que no dependa de otras escenas.

5.5. del problema de la labilidad de la memoria. ¿Es necesario recordar todas las instancias o estas pueden comprimirse e incluso olvidarse con el tiempo?

Si se considera tanto el modelo de media como el de discriminación, en ambos hay que guardar datos derivados de la percepción. Se podría pensar que el modelo de media no necesita guardar todas las instancias percibidas, si se guarda la media y el número de instancias, se puede recalcularse la media ante otra instancia de entrada. Pero las medidas de dispersión necesitan de toda la información percibida para recalcularse ante una nueva entrada. Por tanto, si se requiere no sólo la media sino el modelo con medidas de dispersión, no se está ganando en compresión del volumen de datos. Ambos modelos son equivalentes en este sentido.

Pero el modelo de discriminación debe, además, guardar las instancias negativas. Se podría argumentar que necesita más volumen de memoria en este sentido. Este problema tiene solución sin incrementar los costes de memoria. Simplemente, se reconstruye el modelo yuxtaponiendo las instancias positivas y negativas dinámicamente, es decir, con cada invocación del sistema.

Ante una entrada de percepción, el modelo, cualquiera que sea, debe predecir la escena, y esperar a que el observador valide o refute la predicción. Ante esta situación, si el observador insta al filósofo a repensar el modelo, el modelo que se guarda ya no es el mismo que el que se tenía. Este hecho, como ya se comentó, es compatible con la labilidad de la memoria y con su modificación dentro de una ventana temporal, marcada esta por el tiempo que pasa desde la entrada sensorial hasta la remodelación realizada por el filósofo. Si el observador no insta al filósofo a remodelar el sistema porque no es necesario, el proceso de labilidad no será evidenciable. Una posible implementación de

este modelo se mostró en (v. supra figura 3, p. 13), donde se puede apreciar la relación de POP con la memoria.

Por otra parte, el peso específico de una instancia utilizada en el modelo es proporcional al número de instancias de que consta el modelo. Una instancia percibida en épocas tempranas de la construcción del modelo, habrá ejercido un peso relativo alto en la predicción pasada, frente a una instancia de la misma importancia percibida en épocas posteriores, tras un incremento considerable de instancias verificadas por el observador. ¿Olvidar es tener menos peso específico en la construcción del modelo? Entonces el inverso del peso específico es compatible con el olvido.

5.6. de la puerta del conexionismo.

Del conexionismo (Rumelhart et al., 1986) al simbolismo. Procesamiento simbólico en área o sistema de Wernicke y de Broca.

Clásicamente se han utilizado modelos simbólicos para explicar el funcionamiento de las áreas o sistemas que procesan el lenguaje, áreas o sistemas de Wernicke y de Broca. Estos modelos suponen una abstracción de las entradas y salidas sensoriales de forma conceptual. Además, el lenguaje tiene sus propias reglas de producción, independientes incluso de la semántica de su contenido. Pero, ¿cómo dotar de sentido a una producción lingüística? No todas las producciones tienen sentido aunque se respeten las normas de construcción. Además, ¿cuál es la semántica de una construcción con sentido? ¿debe poder “imaginarse”?

Aquí es donde el modelo propuesto cobra sentido explicativo. Ante un conjunto ordenado de conceptos ofrecidos por un constructo con sentido del lenguaje, realizado como un proceso simbólico, debe “imaginarse” su contenido. El modelo propuesto es capaz de “imaginar” con procesamiento conexionista, un concepto aprendido. Además, no sólo su forma pura, aquella obtenida por el modelo de medias, sino, aquella que representa lo que es frente a lo que no es, como los modelos discriminativos propuestos. El modelo discriminativo ofrece una riqueza conceptual mucho mayor que el modelo de medias que posibilita la “imaginación” con muchos matices de los conceptos del lenguaje.

6. Conclusiones

Se ha podido verificar la hipótesis de convergencia en los estados de normalización, con todas las cautelas discutidas.

Por tanto, se propone al **vector beta**, o su posible variación en escala, como el invariante de escena que permite el reconocimiento de esta ante una entrada visual en un determinado modelo de sistema de clasificación lineal. Este invariante no pierde generalidad por el hecho de cambiar ante un nuevo proceso de aprendizaje.

El módulo “*sistema de aprendizaje de escena POP*”, diseñado como un estado L-NL, implementado siguiendo a (Martinez-Garcia et al., 2018), resuelve satisfactoriamente los problemas de

crecimiento de los modelos de clasificación estáticos y posibilita el aprendizaje.

Adicionalmente, se muestra cómo la transformación inversa de este invariante hacia las primeras etapas de entrada permite “imaginar”, es decir, visualizar sus características como imagen, de forma más intuitiva. Recoge información de lo que es frente a lo que no es. Su forma pura, solo lo que es, como media de las instancias positivas, también puede extraerse del modelo.

Finalmente, este sencillo pero eficaz modelo, puede presentar problemas de convergencia en hiperespacios reducidos, o, ante grandes cantidades de instancias positivas y/o negativas, por su posible divergencia. En el caso que nos ocupa, el hiperespacio es suficientemente grande (10025 dimensiones) para retrasar estos efectos. Si, aún así presentara estos problemas, el filósofo (P) debería plantearse nuevas escenas de una escena previa, enriqueciendo el “universo conceptual” de su conocimiento, o reclamar la actuación de módulos superiores que sean capaces de resolver estos conflictos, utilizando métodos similares a los aquí expuestos, pero a una escala de abstracción superior. El nivel donde se crea un problema no suele contemplar mecanismos de solución para resolverlo, sería absurdo. Tal solución, si existe, suele ser visible en niveles superiores.

No se debe dejar escapar la oportunidad de considerar las reflexiones del físico y filósofo Mario Bunge, destacado impulsor de la neurociencia. En su libro “El problema mente-cerebro” (Bunge, 1985, 1988, 2002), en concreto su capítulo 1, propone un mecanismo emergentista por el cual la mente es una emergencia de los procesos físicos cerebrales. Además esta misma mente emergida interactúa con los procesos físicos cerebrales. En este sentido podemos ver los invariantes como creaciones del cerebro que emergen hacia conceptos, constructos que, además, interactúan con el cerebro para poder imaginarse.

7. Agradecimientos

A la Universitat de València, en concreto al programa de máster en neurociencias básicas y aplicadas del año 2017-2018 y al departamento de “Image Processing Lab.” por la oportunidad de exponer públicamente mis aportaciones en este campo científico.

Referencias

- [1] Allen, C. (1999). Animal concepts revisited: The use of self-monitoring as an empirical approach. *Erkenntnis*, 51(1), 33–40. 25
- [2] Benavent, J., Benavent, X., de Ves, E., Granados, R., & García-Serrano, A. (2010). Experiences at imageclef 2010 using cbir and tbir mixing information approaches. *CLEF 2010 LABs and Workshops, Notebook Papers*. 6
- [3] Benavent, J., Castellanos, A., Benavent, X., de Ves, E., & García-Serrano, A. (2012). Visual concept features and textual expansion in a multimodal system for concept annotation and re-

- trieval with flickr photos at imageclef2012. *CLEF 2012 Evaluation Labs and Workshop, Online Working Notes*. 6
- [4] Benavent, X., García-Serrano, A., Granados, R., Benavent, J., & de Ves, E. (2013). Multimedia information retrieval based on late semantic fusion approaches: Experiments on a wikipedia image collection. *IEEE Transactions on Multimedia*, *15*(8)(1), 2009–2021. 6
- [5] Bunge, M. (1985, 1988, 2002). *El problema mente-cerebro. Un enfoque psicobiológico*. Oxford: Tecnos gr. Anaya SA, second, reimpresión (2002) ed. 30
- [6] Carandini, M., & Heeger, D. (2012). Normalization as a canonical neural computation. *Nature Rev Neurosci.*, *13*(1), 51–62. 5
- [7] Castellanos, A., Benavent, J., Benavent, X., García-Serrano, A., & de Ves, E. (2012). Using visual concept features in a multimodal retrieval system for the medical collection at imageclef2012. *CLEF 2012 Evaluation Labs and Workshop, Online Working Notes*. 6
- [8] Castellanos, A., Benavent, X., Benavent, J., & García-Serrano, A. (2011). Uned-uv at medical retrieval task of imageclef 2011. *CLEF 2011 Labs and Workshop, Notebook Papers*. 6
- [9] Coen-Cagli, R., & Schwartz, O. (2013). The impact on midlevel vision of statistically optimal divisive normalization in v1. *Journal of Vision*, *13*(8). 5, 24
- [10] Cowan, J., Neuman, J., & van Drongelen, W. (2016). Wilson–cowan equations for neocortical dynamics. *Journal of Mathematical Neuroscience*, *6*(1), 1–24. 5
- [11] da Fonseca, M., & Samengo, I. (2016). Derivation of human chromatic discrimination ability from an information-theoretical notion of distance in color space. *Neural Computation*, *28*(12), 2628–2655. 6
- [12] de Ves, E., Benavent, X., , Coma, I., & Ayala, G. (2016). A novel dynamic multi-model relevance feedback procedure for content-based image retrieval. *Neurocomputing*, *208*, 99–107. 6
- [13] Deselaers, T., Keysers, D., & Ney, H. (2008). Features for image retrieval: An experimental comparison. *Information Retrieval*, *11*, 77–107. 6
- [14] Granados, R., Benavent, J., Benavent, X., de Ves, E., & García-Serrano, A. (2011). Multimodal information approaches for the wikipedia collection at imageclef 2011. *CLEF 2011 Labs and Workshop, Notebook Papers*. 6
- [15] Hinton, G. E. (2007). Learning multiple layers of representation. *Trends in Cognitive Sciences*, *11*, 428–434. 5
- [16] Hoyer, P., & Hyvaarinen, A. (2002). A multi-layer sparse coding network learns contour coding from natural images. *Vision Research*, *42*(12), 1593–1605. 5

- [17] Hyvaarinen, A., Gutmann, M., & Hoyer, P. (2005). Statistical model of natural stimuli predicts edge-like pooling of spatial frequency channels in v2. *BMC Neuroscience*, 6(1), 12–24. 5
- [18] Karklin, Y., & Lewicki, M. S. (2005). A hierarchical bayesian model for learning nonlinear statistical regularities in nonstationary natural signals. *Neural Computation*, 17, 397–423. 5
- [19] Karklin, Y., & Lewicki, M. S. (2009). Emergence of complex cell properties by learning to generalize in natural scenes. *Nature*, 457(7225), 83–86. 5
- [20] Malo, J., & Gutierrez, J. (2006). V1 non-linear properties emerge from local-to-global non-linear ica. *Network: Computation in Neural Systems*, 17(1), 85–102. 5
- [21] Malo, J., Kane, D., & Bertalmio, M. (2016). The maximum differentiation competition depends on the viewing conditions. *J Vision*, 16(12), 822–822. 5
- [22] Malo, J., & Simoncelli, E. (2015). Geometrical and statistical properties of vision models obtained via maximum differentiation. *International Society for Optics and Photonics*, (pp. 93940L–93940L). 5
- [23] Martinez-Garcia, M., Cyriac, P., Bertalmio, M., & Malo, J. (2018). Derivatives and inverse of cascaded linear+nonlinear neural models. *arXiv:1711.00526*. 4, 5, 7, 12, 26, 29
- [24] Nuding, U., & Zetsche, C. (2007). Learning the selectivity of v2 and v4 neurons using non-linear multi-layer wavelet networks. *Biosystems*, 89(1-3), 273–279. 5
- [25] Ogata, K. (1987). *Sistemas Dinámicos*. México: Prentice-Hall Iberoamericana SA. 4
- [26] Osindero, S., Welling, M., & Hinton, G. E. (2006). Topographic product models applied to natural scene statistics. *Neural Computation*, 18(2), 381–414. 5
- [27] Ringach, D.-L. (2002). Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. *J Neurophysiol*, 88, 455–463. 21, 22
- [28] Rolls, E., & Deco, G. (2001). *Computational neuroscience of vision*. New York: Oxford University Press. 4
- [29] Rumelhart, D., McClelland, J., & the P.D.P research group (1986). *Parallel distributed processing, explorations in microstructure of cognition, 1 Foundations*. Cambridge: MIT Press. 29
- [30] Russell, S., & Norvig, P. (2010). *Artificial Intelligence. A modern approach (3^a ed.)*. New Jersey: Pearson Education Inc. 11
- [31] Schwartz, O., Sejnowski, T. J., & Dayan, P. (2006). Soft mixer assignment in a hierarchical generative model of natural scene statistics. *Neural Computation*, 18(11), 2680–2718. 5
- [32] Simoncelli, E., & Olshausen, B. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience*, 24(1), 1193–1216. 5

- [33] Spratling, M. (2011). Unsupervised learning of generative and discriminative weights encoding elementary image components in a predictive coding model of cortical function. *Neural Computation*, (pp. 60–103). 5
- [34] Zhaoping, L. (2006). Theoretical understanding of the early visual processes by data compression and data selection. *Network: Computation in Neural Systems*, 17, 301–334. 5

Índice de figuras

1	Proceso general.	11
2	Proceso estado 5. POP.	12
3	Posible implementación del estado 5. POP.	13
4	Imágenes de entrenamiento de la escena del físico Tesla.	14
5	Puntuaciones	16
6	Vectores del test sobre invariantes, antes y después de aprendizaje. Se ha marcado los ángulos π y $\pi/2$ relativos al invariante aprendido. El vector azul es el invariante entrenado, el rojo el aprendido. Los ángulos de los vectores del test son respecto al vector aprendido, una proyección sobre el plano que forma cada vector con el invariante. Por tanto, los vectores representados oscilan en $10025 - 2$ grados de libertad.	17
7	Detalle. Vectores del test sobre invariantes, antes y después de aprendizaje. Se ha marcado los ángulos π y $\pi/2$ relativos al invariante aprendido, o al invariante entrenado, si todavía no ha habido aprendizaje. El vector azul es el invariante entrenado, el rojo el aprendido. Los ángulos de los vectores del test son respecto al vector aprendido, una proyección sobre el plano que forma cada vector con el invariante. Por tanto, los vectores representados oscilan en $10025 - 2$ grados de libertad.	18
8	Físicos. Observado e imaginado en el tercer estado.	19
9	Imágenes test. Distorsiones del físico desconocido.	20
10	Ruta y correlato neuronal.	21
11	Campos receptivos en V1. Reales en macaco (a). Calculados para la escena de Tesla (b).	22

Índice de tablas

1	Estadísticos de imágenes de entrenamiento. Ángulos en grados.	15
2	Estadísticos de imágenes de entrenamiento. Distancias euclídeas.	15
3	Ángulos (en grados) entre invariantes tras entrenamiento.	16
4	Parámetros.	20
5	Ángulos (en grados) entre invariantes tras aprendizaje.	20

Índice de ecuaciones

1	Normalización	5
2	Estado 1. Etapa lineal.	7
3	Estado 1. Etapa no lineal.	7
4	Estado 1. Parámetros.	8
5	Estado 2. Etapa lineal.	8
6	Estado 2. Etapa no lineal.	8
7	Estado 2. Parámetros.	8
8	Estado 3. Etapa lineal.	8

9	Estado 3. Etapa no lineal.	8
10	Estado 3. Parámetros.	9
11	Estado 4. Etapa lineal.	9
12	Estado 4. Etapa no lineal.	9
13	Estado 4. Parámetros.	9
14	Función de puntuación lineal.	9
15	Estado 5. Pensador. Etapa lineal.	10
16	Estado 5. Observador. Etapa lineal.	10
17	Estado 5. Predictor. Etapa lineal.	10
18	Estado 5. Pensador. Etapa no lineal.	10
19	Estado 5. Observador. Etapa no lineal.	10
20	Estado 5. Predictor. Etapa no lineal.	10
21	Estado 5. Parámetros.	11
22	Imaginación del invariante.	11