



Inteligencia y Topología

Un modelo matemático para definir y evaluar sistemas inteligentes

Tesis presentada para optar al título de Licenciado en Matemática

Autor: Claudio Martinez

Director de Tesis: Dra. Cristina Camos

-febrero 2025-

Resumen

El siguiente trabajo aborda la exploración de la inteligencia artificial (IA) y la inteligencia humana desde una perspectiva matemática. Se enfoca en desarrollar una definición operativa del concepto inteligencia, utilizando herramientas topológicas para analizar y modelar cómo los sistemas inteligentes generan y procesan información. La tesis compara el cerebro humano y las redes neuronales artificiales, examinando similitudes y diferencias en su enfoque hacia el aprendizaje. Utilizando como base para el estudio los pilares del aprendizaje formulados por el Profesor Stanislas Dehaene, en su obra: ¿Cómo aprendemos? (Dehaene, S. 2020). Se busca un modelo matemático que describa la inteligencia desde una perspectiva técnica y que pueda aplicarse en el desarrollo de nuevas estrategias para redes neuronales. Finalmente, la tesis evalúa la arquitectura Transformer bajo el modelo topológico contextual propuesto, buscando comprender mejor la capacidad de abstracción tanto en la IA como en el cerebro humano. Y propone en futuras investigaciones, la utilización de otra herramienta matemática, como la geometría diferencial, para modelizar una abstracción evolutiva en redes neuronales artificiales

índice general

Capítulo 1: Introducción	6
1.1 Contexto general.....	6
1.2 Objetivos de la investigación	7
1.3 Estructura del trabajo	9
Capítulo 2: Motivación y justificación del trabajo.....	11
2.1 Importancia del estudio comparativo entre el cerebro humano y la inteligencia artificial	11
2.2 Relevancia de la formalización matemática de los operadores de información.....	12
Capítulo 3: Marco teórico: los cuatro pilares del aprendizaje.....	14
3.1 Los cuatro pilares en el ser humano	14
3.1.1 Atención	14
3.1.2 Compromiso Activo	15
3.1.3 Retroalimentación	15
3.1.4 Consolidación	15
Capítulo 4: Introducción a redes neuronales y teorema de aproximación universal.....	16
4.1 Teorema de aproximación universal.....	22
4.2 Similitud de aprendizajes: pilares y procesamiento en redes neuronales.....	23
4.2.1 Atención en una red neuronal	23
4.2.2 Compromiso Activo: Función de activación	24
4.2.3 Retroalimentación: retropropagación en redes neuronales	25
4.2.4 Consolidación: Validación y generalización	26
4.3 Ejemplo numérico de aplicación de una red neuronal simple con una capa oculta ..	27
Capítulo 5: Similitud matemática	32
5.1 Operadores de información: cerebro humano y redes neuronales artificiales	32
5.1.1 Inteligencia: función con memoria	32
5.1.2 Propiedades necesarias de la función de inteligencia	33
Capítulo 6: Diferencias entre el cerebro humano y el cerebro artificial	38
6.1 Definición de la información en ambos sistemas.....	38

6.1.1 Aplicación en ambos sistemas	39
6.2 Abstracción y plasticidad neuronal.....	40
6.3 Plasticidad funcional y estructural.....	40
Capítulo 7: Información y topología.....	41
7.1. Cantidad de información, medida de Shannon.....	42
7.2 Estructura cualitativa de la información.....	44
7.2.1 Definición formal del conjunto de datos y acciones del operador inteligente	44
7.2.2 Arquitectura del conjunto contexto	45
7.2.3 Justificación intuitiva de la primera acción de continuidad del operador inteligente	47
7.3 Clasificación de la información contextual	48
7.3.1 información relevante	48
7.3.2 Información básica	48
7.3.3 Información inteligente	49
7.3.4 Información completa contextual	49
Capítulo 8: Espacio topológico de información	50
8.1 Definición de la topología contextual.....	50
8.1.1 Algunas aclaraciones de la dinámica que τC ejerce en C	51
8.1.2 Demostración de topología bien definida en el conjunto contexto	52
8.1.3 Disquisición sobre el vacío informativo.....	56
8.2 Propiedades del conjunto contexto	58
8.2.1 Cerradura.....	58
8.2.2 Separabilidad	61
8.2.3 Compacidad	62
Capítulo 9: Dinámica de la topología contextual: cerebro humano y red neuronal artificial	66
9.1 Las mismas herramientas de trabajo	66
9.2 Rigidez topológica: conjetura de la red neuronal artificial.....	66
9.3 Flexibilidad topológica: conjetura contextual del cerebro humano.....	68
9.4 Comparación dinámica: conjetura de la abstracción	68

9.4.1 La tasa y la dona: grafican la idea	71
Capítulo10: Poniendo en práctica la investigación (evaluando los Transformers)	73
10.1 Transformers.....	73
10.2 Paralelización vs recurrencia secuencial.....	75
10.3 Ajedrez transformers	78
10.4 Evaluación de los transformers y cerebro ajedrecista bajo la topología contextual	80
10.4.1 El Modelo de la Topología Contextual	81
10.4.2 Representación matemática del transformers como un espacio contextual en un conjunto contexto X	82
10.4.3 Ajedrez contextual	84
Capítulo 11: Conclusión y propuesta de futura investigación	86
11.1 conclusión.....	86
11.2 Propiedades de la inteligencia.....	89
11.3 Futura investigación	92

“Decir que el todo proviene de la nada, es darle inteligencia a lo que no existe”.

“Los cielos cuentan la gloria de Dios, y el firmamento anuncia la obra de sus manos” (salmo 19:1).

Capítulo 1: Introducción

1.1 Contexto general

En su libro *El tejido del Cosmos: Espacio, Tiempo y Textura de la Realidad* Green (2020), destaca la relevancia de dos variables fundamentales en el universo: la energía y la entropía. La energía, que es aquello con lo que realizamos trabajo, y la entropía, que describe cómo esa energía se distribuye, revela que a medida que empleamos energía, su capacidad de realizar trabajo se degrada. Este concepto no solo es esencial en la física, sino que también encuentra paralelos en otros campos, como el procesamiento de la información en nuestra era moderna.

El filósofo futurista Alvin Toffler, en su obra *La tercera ola* Toffler (1980), clasifica la evolución de la humanidad en tres grandes olas sociales: la agrícola, la industrial y la era de la información. En esta tercera ola, la información se ha convertido en el recurso esencial para la toma de decisiones, del mismo modo que lo fue la tierra en la era agrícola o el capital en la industrial. El procesamiento, distribución y optimización de este recurso es crucial para la sociedad contemporánea.

Hoy, la información no solo es el motor de nuestro avance tecnológico, sino también la base de nuestra comprensión del mundo. La capacidad de manejar grandes volúmenes de datos y extraer de ellos patrones útiles tiene profundas implicaciones no solo en la tecnología, sino en nuestra interpretación de la realidad misma. En este sentido, las redes neuronales, inspiradas en el funcionamiento del cerebro humano, han surgido como una de las herramientas más poderosas para procesar y modelar información.

Las redes neuronales no solo replican algunos aspectos del aprendizaje humano, sino que también han demostrado ser capaces de resolver problemas complejos en campos tan variados como la medicina, la economía y la inteligencia artificial. Estudiar cómo estos sistemas procesan la información nos ofrece no solo avances tecnológicos, sino también una mayor comprensión de los procesos cognitivos que subyacen en nuestra propia inteligencia. En este contexto, investigar los principios fundamentales detrás del aprendizaje, tanto biológico como artificial, nos lleva a comprender mejor cómo la información puede ser transformada y optimizada para enfrentar los retos del futuro. Es decir, a definir operativa y sistemáticamente el concepto “inteligencia”.

Este trabajo se sitúa precisamente en este cruce, donde el estudio de las redes neuronales artificiales, desde un enfoque matemático y conceptual, permite explorar los vínculos entre la inteligencia humana y la artificial. Aquí, el procesamiento de información se entiende no solo como un fenómeno técnico, sino como un paso fundamental en nuestra búsqueda por entender y mejorar nuestras capacidades cognitivas y tecnológicas.

1.2 Objetivos de la investigación

Objetivo General

Desarrollar una definición operativa de la inteligencia basada en herramientas matemáticas, particularmente la topología, para establecer un marco técnico que permita analizar y modelizar la generación y el procesamiento de información en sistemas inteligentes. Además, establecer un criterio matemático que funcione como un filtro para evaluar si un sistema puede considerarse inteligente, aplicándolo tanto a la inteligencia artificial representada por redes

neuronales artificiales como a la inteligencia humana. Esta definición busca trascender enfoques conceptuales, filosóficos o psicosociales, proporcionando una base formal, general y técnica que facilite la evolución del procesamiento de información en distintos entornos.

Objetivos Específicos

1. Explorar el uso de la topología como herramienta matemática para describir la inteligencia, destacando su capacidad para preservar estructuras y evitar interrupciones en la información.
2. Establecer un marco matemático formal que permita caracterizar el procesamiento de información en sistemas inteligentes, proporcionando criterios medibles para su análisis.
3. Definir un filtro matemático que permita determinar si un sistema es inteligente, aplicándolo a diferentes modelos, incluyendo redes neuronales, arquitecturas como los Transformers y el razonamiento humano.
4. Proponer un modelo matemático que describa la inteligencia desde una perspectiva técnica, permitiendo su aplicación en el desarrollo de nuevas estrategias para la evolución de redes neuronales y la comprensión del aprendizaje humano.
5. Explorar la posibilidad de mejorar la flexibilidad en inteligencia artificial a partir de estructuras matemáticas más avanzadas, con el fin de aproximar su comportamiento al del cerebro humano.

1.3 Estructura del trabajo

Esta investigación se basa principalmente en los planteamientos del neurocientífico Stanislas Dehaene, presentados en su obra *Cómo aprendemos* Dehaene (2020). En este libro, tras años de experimentación, Dehaene propone un modelo robusto y coherente del aprendizaje en el cerebro humano. Dicho modelo servirá de punto de partida para utilizar redes neuronales como una herramienta para modelar las interpretaciones científicas sobre el funcionamiento del cerebro humano en la etapa inicial de esta investigación.

Posteriormente, el libro de Dehaene aborda las diferencias entre el cerebro humano y la inteligencia artificial (IA), destacando la plasticidad neuronal como un concepto clave que distingue a ambos sistemas. La plasticidad neuronal se presenta como la capacidad del cerebro para adaptarse y reorganizarse a partir de nuevas experiencias, lo cual le otorga una ventaja significativa en términos de aprendizaje y abstracción. Este concepto será analizado en profundidad, junto con ejemplos que ilustran su consecuencia directa: la capacidad de abstracción.

Para abordar estas diferencias, la investigación propondrá una estructura matemática de la información, utilizando herramientas topológicas que permitan analizar cualitativamente cómo ambos sistemas —el cerebro humano y las redes neuronales— procesan y manejan la información. A partir de esta estructura, se formulará una conjetura sobre un posible conjunto universal para la información, que engloba tanto la información procesada por el cerebro como la manejada por la red neuronal artificial.

Finalmente, se presentará una segunda conjetura que buscará explicar cómo ocurre el proceso de abstracción tanto en el cerebro humano como en las redes neuronales,

proporcionando una base para futuras investigaciones sobre las diferencias esenciales entre estos dos operadores inteligentes.

Con todo lo expresado anteriormente, la investigación, con el fin de cumplir con los objetivos planteados, hará uso de la comparación en forma matemática del procesamiento de información, tanto en redes neuronales artificiales, como en la interpretación del aprendizaje del cerebro humano, según el profesor Dehaene.

Capítulo 2: Motivación y justificación del trabajo

2.1 Importancia del estudio comparativo entre el cerebro humano y la inteligencia artificial

A lo largo de la historia de la humanidad, el cerebro humano ha sido el único operador de la información, en tanto que no había evidencia de otro mecanismo que pudiera procesar y contextualizar datos de manera comparable. Los seres humanos hemos utilizado esta capacidad para tomar decisiones de manera eficiente y efectiva. Sin embargo, en esta era de la información, ha surgido un nuevo operador: la inteligencia artificial (IA), que comenzó como un intento de modelar nuestra forma de procesar la información.

Conceptualmente la IA puede definirse como el proceso por el cual una máquina puede imitar las funciones cognitivas del cerebro humano: percibir, razonar, resolver problemas, aprender (Pons, C. 2022).

Con el tiempo, la IA ha evolucionado hasta convertirse en un operador matemático por derecho propio, capaz de establecer relaciones funcionales entre los datos que recibe y los resultados que busca alcanzar. El avance de las tecnologías y el acceso a grandes cantidades de datos han permitido que la IA demuestre una notable eficacia en el procesamiento de información, comparable en algunos aspectos a la del cerebro humano.

Dada la creciente relevancia de la IA y su capacidad para operar en contextos cada vez más complejos, es crucial realizar un estudio comparativo entre estos dos operadores inteligentes: el cerebro humano y la IA. Observar cómo ambos evolucionan en su forma de procesar información a lo largo del tiempo y comprender las similitudes y diferencias entre ellos nos proporcionará herramientas valiosas para comprender mejor sus mecanismos internos. Este

análisis comparativo no solo será útil en el ámbito científico, sino que también permitirá comportamientos anómalos regulares—tales como decisiones erróneas o sesgos en el procesamiento de información—que puedan surgir en ambos sistemas, ofreciendo así una perspectiva crítica para el desarrollo de la inteligencia artificial y la toma de decisiones humanas.

Sin embargo para cumplir los objetivos propios de este trabajo, hay un escollo no menor y de gran relevancia, que es proporcionar una definición operativa para el concepto de inteligencia humana e inteligencia general.

En principio y basándonos en la materia prima a procesar por los operadores (cerebro humano y redes neuronales artificiales), podemos decir que la inteligencia, es la forma en que los operadores generan a partir de datos, información, y la procesan iteradamente con eficiencia buscando eficacia. Esta definición nos da sustento operativo para esta investigación, y nos permite construir una estructura topológica de calidad. Además, vincula directamente el concepto de inteligencia con su función: generar y procesar información de manera adecuada. De allí se desprende: razonar, percibir y aprender.

2.2 Relevancia de la formalización matemática de los operadores de información

La matemática actúa muchas veces, como la ingeniería; diseñando herramientas para cuantificar y comparar fenómenos, tanto físicos como conceptuales o abstractos. No solo proporciona sustento a razonamientos científicos en campos como la física, la biología y la química, sino que también establece un marco cuantificable para conceptos abstractos y

filosóficos. En este contexto, la información se convierte en un elemento intangible que, al ser operado, nos ofrece resultados sólidos y coherentes con una semántica contextual.

Por lo tanto, la formalización matemática es crucial para fundamentar y diseñar un marco cualitativo y cuantitativo que contemple tanto la naturaleza de la información como las estrategias para procesarla. Actualmente, reconocemos dos operadores principales para esta tarea: la inteligencia artificial y el cerebro humano. Comprender el comportamiento de estos operadores desde una perspectiva matemática es esencial para establecer condiciones necesarias y/o suficientes que aseguren un procesamiento adecuado de la información.

Capítulo 3: Marco teórico: los cuatro pilares del aprendizaje

En este capítulo se abordan las definiciones y justificaciones experimentales, sobre las cuales la neurociencia interpreta la forma de aprendizaje de nuestro cerebro.

3.1 Los cuatro pilares en el ser humano

Dehaene (2020) propone cuatro pilares fundamentales, que permiten al cerebro humano procesar información de manera efectiva. Estos pilares son: atención, compromiso activo, retroalimentación y consolidación. A continuación, se detallan cada uno de estos pilares junto con experimentos que los justifican.

3.1.1 Atención

La atención es el proceso mediante el cual seleccionamos y enfocamos la información relevante, excluyendo la que consideramos irrelevante. Dehaene señala que la atención es crucial para el procesamiento eficiente de la información.

Experimento: En el experimento realizado por Moray (1959), se pidió a los participantes que llevaran a cabo una tarea de escucha dicotómica, en la que debían atender a dos mensajes distintos presentados simultáneamente, uno en cada oído. Los resultados revelaron que los participantes solo podían procesar la información del mensaje al que estaban prestando atención, lo que evidencia la importancia de este pilar en el aprendizaje.

3.1.2 Compromiso Activo

El compromiso activo se refiere a la participación del individuo en el proceso de aprendizaje, estableciendo conexiones entre la nueva información y la que ya posee. Este aspecto es esencial para facilitar el aprendizaje significativo.

Experimento: Chi y Bassok (1989) llevaron a cabo un experimento en el que se pidió a los participantes que resolvieran problemas de física. Los resultados mostraron que aquellos que verbalizaban sus pensamientos y explicaban los problemas en voz alta mientras los resolvían aprendían más en comparación con quienes no lo hacían. Este hallazgo resalta la relevancia del compromiso activo en el aprendizaje.

3.1.3 Retroalimentación

La retroalimentación es el proceso de recibir información sobre el rendimiento en una tarea y ajustar la estrategia de aprendizaje en función de esa información.

Experimento: Butler y Roediger (2008) realizaron un estudio donde los participantes llevaban a cabo una tarea de aprendizaje y recibían retroalimentación sobre su rendimiento. Los resultados indicaron que aquellos que recibieron retroalimentación lograron un aprendizaje más efectivo que quienes no la recibieron. Esto subraya la importancia de la retroalimentación en el proceso de aprendizaje.

3.1.4 Consolidación

La consolidación es el proceso mediante el cual la información se transfiere de la memoria a corto plazo a la memoria a largo plazo, facilitada por el sueño y la relajación.

Experimento: Maquet et al. (2005) llevaron a cabo un experimento en el que se pidió a los participantes aprender una tarea de memoria, seguida de un periodo de sueño o relajación. Los resultados mostraron que aquellos que durmieron o se relajaron después de la tarea lograron un aprendizaje superior en comparación con aquellos que no lo hicieron. Este hallazgo enfatiza la importancia de la consolidación en el aprendizaje.

Capítulo 4: Introducción a redes neuronales y teorema de aproximación universal

Una red neuronal hacia adelante es un objeto matemático que consiste en varias capas de neuronas, organizadas secuencialmente, donde los datos fluyen desde una capa de entrada, a través de capas intermedias llamadas capas ocultas, hasta una capa final de salida. Cada neurona en una capa aplica una función de activación a una combinación ponderada de las salidas de las neuronas en la capa anterior Huet (2023). En líneas generales es una función F .

$$F: \mathbb{R}^n \rightarrow \mathbb{R}^m$$

La capa de entrada es el conjunto de neuronas donde ingresan los datos al modelo, representados en forma de un vector.

Las capas ocultas son las que procesan los datos antes de llegar a la capa de salida. Cada capa oculta transforma las salidas de la capa anterior aplicando una función de activación. Finalmente, la capa de salida (capa L), es la última capa de la red, que genera el resultado final.

A continuación, se describe el funcionamiento de una red neuronal. $\alpha^{(l)}$: representa el vector de activaciones de la capa l , donde l puede ser cualquier capa (entrada, oculta o salida),

$z^{(l)}$: es el vector de combinaciones lineales antes de aplicar la función de activación en la capa l ,
 σ : la función de activación que intenta capturar patrones no lineales.

Propagación hacia adelante

El proceso comienza enviando los datos de entrada a través de las capas de la red hasta obtener una salida. Paso en cada capa; en cada capa l (excepto en la entrada), calculamos:
 Combinación lineal: $z^{(l)} = w^{(l)}a^{(l-1)} + b^{(l)}$ donde; $w^{(l)}$ es la matriz de pesos que conecta la capa $l - 1$ a la capa l , $a^{(l-1)}$ son las activaciones de la capa anterior y $b^{(l)}$ es el vector de sesgos de la capa l .

Activación: $a^{(l)} = \sigma(z^{(l)})$, donde σ es la función de activación aplicada a $z^{(l)}$, produciendo las activaciones de esa capa (por ejemplo, la función sigmoide, que “suaviza” las salidas, manteniéndolas dentro de un rango específico). Su propósito es romper con linealidades de información, para ajustar complejidades. Algunas de las más usadas son:

Función Sigmoidal: transforma la salida en un valor entre (0,1), lo que resulta útil para problemas de clasificación binaria. La fórmula es:

$$f(x) = \frac{1}{1 + e^{-x}}$$

ReLU (Rectified Linear Unit): la función más común en redes neuronales profunda transforma el valor en cero si es negativo o lo deja igual si es positivo, su fórmula es:

$$f(x) = \max(0, x)$$

Tangente hiperbólica (tanh): transforma la salida en un valor entre $[-1,1]$ lo que puede ser útil en problema donde los datos tienen valores negativos y positivos:

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

Este proceso continua capa por capa, hasta llegar a la capa de salida, donde obtenemos el resultado final de la red.

Durante el entrenamiento, la red ajusta sus pesos para minimizar la diferencia entre sus predicciones (en la capa de salida) y los valores reales. Esto se hace minimizando una función de error $J(*)$ que cuantifica esa diferencia, de la múltiple variedad de dichas funciones, esbozamos las más conocidas con su ecuación y su uso conveniente:

Error cuadrático medio: resulta una función fácil de derivar y calcular y de gran utilidad en modelos de regresión lineal. Como contrapartida es sensible a valores atípicos. Su fórmula es:

$$J(\theta) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (*)$$

$(y_i - \hat{y}_i)$: es la diferencia entre la salida propuesta por la red $\hat{y}_i$ y la salida real y_i .

n : es el número total de datos de entrada con que la red cuenta para su entrenamiento.

Entropía cruzada binaria: relacionada directamente con la máxima verosimilitud y es adecuada para modelos con salida probabilística y distinción de categorías. Su fórmula es:

$$J_{(\theta)} = -\frac{1}{n} \sum_{i=1}^n [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (*)$$

Retropropagación y el delta de error

Este proceso calcula cuánto deben ajustarse los pesos para reducir el error de la red. Para ello utiliza el procedimiento del descenso del gradiente formulado en la teoría de errores, el cual busca a través del uso reiterado de la regla de la cadena, ir aproximando de manera estocástica, al punto crítico de valor mínimo. En esta investigación por cuestiones de mayor simplicidad en la comprensión, utilizaremos un término llamado delta de error $\delta^{(l)}$, para sintetizar el descenso del gradiente. $\delta^{(l)}$ representa el error de cada capa con relación a la salida esperada. *Es como un imputador de responsabilidades*, asigna a cada neurona de la capa $l - 1$, su influencia en el crecimiento del error de las neuronas de la capa l .

Cálculo del delta en la capa de salida

Para la última capa L , el delta de error se calcula:

$$\delta^{(L)} = \nabla_{a^{(L)}} J \odot \sigma'(z^{(L)})$$

Aquí:

$\nabla_{a^{(L)}} J$: es el gradiente de la función de error J con respecto a las activaciones $a^{(L)}$ en la capa de salida.

$\sigma'(z^{(L)})$: es la derivada de la función de activación evaluada en $z^{(L)}$, lo que ajusta la sensibilidad del error de cada neurona en esa capa.

Operador $\odot$: representa el producto de Hadamard, una multiplicación componente a componente que asegura que el error se ajuste según cada neurona. Matemáticamente se define: sean $A^{m \times n}$ y $B^{m \times n}$ dos matrices, entonces su producto de Hadamard se denota $A \odot B$ y su fórmula es:

$$(A \odot B)_{ij} = A_{ij} \cdot B_{ij}, \quad \forall i, j$$

Este cálculo en la capa de salida nos da una medida de cuánto afecta cada activación final al error total.

Cálculo del delta en las capas ocultas

Para las capas ocultas, usamos la salida del delta de error de la capa siguiente, por ejemplo, la capa $l + 1$, para calcular el error de cada capa anterior. La fórmula general es:

$$\delta^{(l)} = (w^{(l+1)})^T \delta^{(l+1)} \odot \sigma'(z^{(l)})$$

Donde:

$(w^{(l+1)})^T$: es la transpuesta de la matriz de pesos, la cual es necesaria para mantener la coherencia en las dimensiones necesarias para realizar el producto de matrices, teniendo en cuenta que la retropropagación se realiza desde la capa de salida hacia la capa de entrada, propagando el error hacia atrás, esto es inverso a la propagación hacia adelante donde se utiliza la matriz original.

$\delta^{(l+1)}$: es el error de la capa $l + 1$, que ayuda a calcular el impacto en la capa l .

$\sigma'(z^{(l)})$: es la derivada de la función de activación en la capa l , ajustando el error según la sensibilidad de cada neurona en esa capa.

Este proceso se repite desde la capa de salida hacia la primera capa oculta.

Actualización de pesos con el producto tensorial

Una vez calculados los valores de $\delta^{(l)}$ en cada capa, podemos actualizar los pesos de la red para reducir el error en la siguiente iteración.

Para cada capa l , los pesos se ajustan como sigue:

Actualización de pesos:

$$w^{(l)} = w^{(l)} - \eta \delta^{(l)} \otimes (a^{(l-1)})^T$$

η : es la tasa de aprendizaje, un parámetro arbitrario que determina cuánto se ajustan los pesos en cada iteración.

$\delta^{(l)} \otimes (a^{(l-1)})^T$: es el producto tensorial entre el delta de error de la capa actual y las activaciones de la capa anterior, que ajusta cada peso en función de su contribución al error. Este producto permite un ajuste fino, considerando todas las posibles combinaciones de impacto entre las neuronas, permitiendo una mejora continua en el rendimiento de la red.

Actualización de sesgos: este procedimiento es crucial en una red neuronal, pues permite ajustar las salidas de las neuronas de forma independiente a las entradas. Esto ayuda a desplazar

las funciones de activación, permitiendo que el modelo aprenda patrones más complejos y mejore su capacidad de generalización.

$$b^{(l)} = b^{(l)} - \eta \delta^{(l)}$$

4.1 Teorema de aproximación universal

El teorema de aproximación universal establece que una red neuronal con al menos una capa oculta y una función de activación no constante puede aproximar cualquier función continua $f: [a, b] \rightarrow \mathbb{R}$ de manera uniforme, bajo ciertas condiciones. (Miranda, (2020)). Este teorema se fundamenta en el teorema de Stone-Weierstrass, que asegura la convergencia uniforme de funciones continuas en espacios métricos compactos.

La formalización matemática del teorema puede expresarse como:

$$\forall \epsilon > 0, \exists n_0 \in \mathbb{N} \text{ tal que } |f(x) - g_n(x)| < \epsilon, \forall n > n_0, \forall x \in [a, b]$$

donde g_n es una sucesión de funciones aproximantes que representan la salida de la red neuronal y n es la cantidad de neuronas ubicadas en la capa oculta. Esto implica que, a medida que aumenta el número de neuronas y la complejidad de la red, la red neuronal puede ajustarse a la función objetivo con una precisión arbitraria.

Este teorema nos permite centrar nuestra investigación de comparación en la arquitectura más simple de redes neuronales, entendiendo a esta, como el elemento mínimo del conjunto de arquitecturas de redes neuronales, que bajo las hipótesis del teorema aseguran eficacia en su tarea. No es poca cosa hacer notar que el tipo de convergencia de la red neuronal es una

convergencia uniforme según el teorema de aproximación universal. Esto ratifica que nuestra elección, es sólida, para este enfoque de investigación.

Además, es importante considerar las limitaciones del teorema de aproximación universal, ya que, aunque garantiza la capacidad teórica de aproximar funciones, en la práctica se requieren grandes cantidades de datos y un diseño cuidadoso para evitar problemas como el sobreajuste.

4.2 Similitud de aprendizajes: pilares y procesamiento en redes neuronales

El propósito de este apartado es establecer la conexión entre los pilares de aprendizaje del cerebro humano y el proceso de aprendizaje de una red neuronal. De esta forma podremos encontrar una intersección entre ambos procesadores de información, en el dominio de las matemáticas, ya que sabemos que una red neuronal es un objeto matemático (relación funcional).

4.2.1 Atención en una red neuronal

En el contexto de una red neuronal hacia adelante, la atención se manifiesta a través de la inicialización de los pesos. En esta fase, la red evita que cualquier característica de entrada tenga una preponderancia arbitraria sobre las demás, asegurándose de que los pesos iniciales estén acotados dentro de un intervalo pequeño, como por ejemplo $[-1,1]$. Este proceso garantiza que ninguna característica influya de manera desproporcionada en el aprendizaje desde el inicio.

El uso de pesos pequeños y aleatorios es una forma rudimentaria, pero efectiva, de que la red "preste atención" equitativamente a todas las características de la entrada. Solo cuando la red

procesa los datos y ajusta los pesos a través del aprendizaje, se ponderan adecuadamente las características más importantes. Esto asegura que, en principio, ninguna característica tenga un impacto excesivo que no refleje su relevancia real en el proceso de aprendizaje.

Para cada $l \in \{1, \dots, L\}$, inicializar aleatoriamente $w^{(l)}$ con valores en $[-1, 1]$

$$w_{\{ij\}}^{(l)} \in [-1, 1]$$

4.2.2 Compromiso Activo: Función de activación

El compromiso activo en una red neuronal se refleja en cómo cada neurona no solo pondera los datos, sino que también aporta una "deformación" mediante la función de activación, lo que rompe la linealidad y permite un procesamiento no trivial de la información. Esto es crucial para que la red pueda captar patrones complejos en los datos de entrada.

La función de activación representa esa contribución de la neurona, dando una respuesta no lineal que refleja su "razonamiento" interno, ayudando a captar la verdadera naturaleza de la información.

Dado un dato de entrada x , propagar hacia adelante:

$$z^{(1)} = w^{(1)} \cdot x, \quad a^{(1)} = f(z^{(1)})$$

Para cada $l \in \{2, \dots, L\}$:

$$z^{(l)} = w^{(l)} \cdot a^{(l-1)}, \quad a^{(l)} = f(z^{(l)})$$

4.2.3 Retroalimentación: retropropagación en redes neuronales

El proceso de retroalimentación en una red neuronal efectivamente se refleja en el algoritmo de backpropagation, que ajusta los pesos a través del descenso del gradiente. Al igual que el cerebro ajusta su modelo mental con nueva información, una red neuronal ajusta sus pesos basándose en el error calculado en las salidas. El delta de error muestra cuánto contribuyó cada neurona al error de la capa siguiente, permitiendo que la red corrija sus operaciones de manera eficiente, tal como lo hace el cerebro humano con la retroalimentación.

Propagación hacia atrás

Calcular error y delta de salida

$$e^{(L)} = d - a^{(L)}, \quad \delta^{(L)} = f'(z^{(L)}) \odot e^{(L)}$$

Donde d es el valor esperado (dato de salida) y $\odot$ denota la multiplicación elemento a elemento (producto de Hadamard)

Para cada $l \in \{L - 1, L - 2, \dots, 2, 1\}$, *calcular el error*:

$$e^{(l)} = (w^{(l+1)})^T \delta^{(l+1)}, \quad \delta^{(l)} = f'(z^{(l)}) \odot e^{(l)}$$

actualizar pesos: Para $l = 1$:

$$\Delta w^{(1)} = \eta(\delta^{(1)} \otimes x), \quad w^{(1)} = w^{(1)} + \Delta w^{(1)}$$

Para $l = 2, \dots, L$:

$$\Delta w^{(l)} = \eta(\delta^{(l)} \otimes a^{(l-1)}), \quad w^{(l)} = w^{(l)} + \Delta w^{(l)}$$

Donde η es la tasa de aprendizaje, y $\otimes$ el producto tensorial.

4.2.4 Consolidación: Validación y generalización

La consolidación, en el cerebro humano, ocurre cuando se repite el proceso de aprendizaje, especialmente durante momentos en que el cerebro no está completamente activo en su interacción con el entorno, como en el sueño. Durante este periodo, el cerebro deja de prestar atención y de comprometerse activamente, pero sigue procesando la información ya adquirida mediante la repetición de las sinapsis involucradas en el aprendizaje. A medida que estas conexiones se repiten, se fortalecen y entran en resonancia, consolidando así el conocimiento. Este proceso permite que el cerebro automatice ciertas respuestas o habilidades, generando un aprendizaje más robusto.

En una red neuronal, este fenómeno se refleja en el proceso de entrenamiento repetido. Tras múltiples iteraciones sobre el conjunto de datos de entrenamiento, la red neuronal comienza a capturar los patrones generales de la información. Al hacerlo, la red es capaz de generalizar y aplicar el conocimiento adquirido a nuevos datos dentro del mismo contexto, incluso si estos datos no formaban parte del conjunto de entrenamiento inicial. Esta capacidad de la red de identificar y aplicar los patrones relevantes a nuevas situaciones es el equivalente a la consolidación del conocimiento en el cerebro humano, lograda por la repetición y refinamiento del aprendizaje.

4.3 Ejemplo numérico de aplicación de una red neuronal simple con una capa oculta

El siguiente, es un ejemplo de regresión lineal, vinculado a una red neuronal hacia adelante con una entrada de dos neuronas, una única capa oculta que contiene dos neuronas y una función sigmoide como activación en la capa oculta. La red debe detectar el precio de una propiedad, según su cantidad de habitaciones y antigüedad de construcción, solo a manera de ejemplo se utilizará un dato de entrenamiento, para mostrar la forma en que se aplica las ecuaciones donde observamos los pilares del aprendizaje. Trabajaremos con el dato normalizado de las variables independientes y la dependiente; la normalización es centrada en el promedio de datos, a partir de la formula:

$$x_N = \frac{x - \bar{x}}{x_{max} - x_{min}}$$

Donde x_N es el valor normalizado del dato x , $\bar{x}$ es valor promedio de los datos proporcionados, x_{max} el valor máximo de los datos y x_{min} el valor mínimo de los datos. De forma tal que esta normalización indica que relación hay entre la distancia a la media y el rango total de datos.

Número de habitaciones: 0.5

Años de antigüedad: 0.4

Precio: 0.5

Propagación hacia adelante

En este ejemplo, se calcularán los valores de preactivación y activación en la capa oculta, y luego se propagarán los resultados hacia la capa de salida.

1. Cálculo de la preactivación en la capa oculta: la preactivación en la capa oculta se calcula como:

$$Z^{(1)} = W^{(1)} \cdot X + b^{(1)}$$

2. Matriz de pesos y sesgos entre la entrada y la capa oculta: este procedimiento en una red neuronal artificial se realiza de forma aleatoria sobre un intervalo acotado de números reales por ejemplo $[-1,1]$, luego la red con el entrenamiento ira ajustando esos valores:

$$W^{(1)} = \begin{pmatrix} 0.5 & -0.3 \\ -0.7 & 0.2 \end{pmatrix}, \quad x^T = (0.5 \ 0.4), \quad b^{(1)T} = (0.1 \ -0.2)$$

Sustituyendo los valores, obtenemos:

$$Z^{(1)} = \begin{pmatrix} 0.5 & -0.3 \\ -0.7 & 0.2 \end{pmatrix} \cdot \begin{pmatrix} 0.5 \\ 0.4 \end{pmatrix} + \begin{pmatrix} 0.1 \\ 0.2 \end{pmatrix} = \begin{pmatrix} 0.23 \\ 0.47 \end{pmatrix}$$

3. Cálculo de la activación en la capa oculta

$$a^{(1)} = \sigma(Z^{(1)}) = \sigma \begin{pmatrix} 0.23 \\ 0.47 \end{pmatrix}$$

La activación en la capa oculta se obtiene aplicando la función de activación sigmoide a los valores de preactivación.

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

Aplicando la función sigmoide a los valores de preactivación, obtenemos:

$$a^{(1)} = \begin{pmatrix} \sigma(0.23) \\ \sigma(0.47) \end{pmatrix} \approx \begin{pmatrix} 0.557 \\ 0.385 \end{pmatrix}$$

4.Cálculo de la preactivación en la capa de salida

$$z^{(2)} = W^{(2)} \cdot a^{(1)} + b^{(2)}$$

Ahora, calculamos la preactivación en la capa de salida utilizando los pesos, las activaciones de la capa oculta y los sesgos.

$$W^{(2)} = \begin{pmatrix} 0.4 & -0.6 \end{pmatrix}, \quad a^{(1)} = \begin{pmatrix} 0.557 \\ 0.385 \end{pmatrix}, \quad b^{(2)} = 0.05$$

Sustituyendo los valores, tenemos:

$$z^{(2)} = 0.4 \cdot 0.557 + (-0.6) \cdot 0.385 + 0.05 = 0.0418$$

Cálculo del error

Utilizamos la función de error cuadrático medio para medir la diferencia entre el valor predicho y el valor real. Sustituyendo los valores:

$$E = \frac{1}{2} (y - \hat{y})^2$$

$$E = 0.1049$$

Propagación hacia atrás (Backpropagation)

El siguiente paso es calcular los deltas para cada capa y realizar la actualización de los pesos.

1. Delta en la capa de salida

$$\delta^{(2)} = (a^{(2)} - y) \cdot a^{(2)} \cdot (1 - a^{(2)})$$

$$\delta^{(2)} = -0.0183$$

2. Delta en la capa oculta; para la capa oculta, el delta se calcula:

$$\delta^{(1)} = (W^{(2)})^T \cdot \delta^{(2)} \cdot a^{(1)} \cdot (1 - a^{(1)})$$

$$\delta^{(1)} = \begin{pmatrix} 0.0018 \\ -0.0025 \end{pmatrix}$$

3. Actualización de pesos y sesgos; finalmente, se actualizan los pesos y los sesgos utilizando una tasa de aprendizaje apropiada. La fórmula de actualización para los pesos es:

$$W^{(l)} = W^{(l)} - \eta \cdot \delta^{(l)} \cdot (a^{(l-1)})^T$$

y para los sesgos:

$$b^{(l)} = b^{(l)} - \eta \cdot \delta^{(l)}$$

Este proceso se efectúa iteradamente con el conjunto de datos de entrenamiento de la red, hasta alcanzar el nivel de aproximación deseado, es decir con la tolerancia requerida por el

operador, de forma tal que la función de costo utilizada (en este caso el error cuadrático medio), vaya acercándose al mínimo en su dominio. Es importante aclarar que esta investigación no se refiere al objeto redes neuronales, sino que está centrada en sistemas que procesen la información de manera “inteligente”, básicamente en similitudes y diferencias entre el cerebro humano y la inteligencia artificial, donde las redes neuronales son la “representación cerebral” de la inteligencia artificial, por ello el interés en estas redes es cualitativo, respecto de la forma que procesan la información y no cuantitativo en el sentido de buscar una aplicación práctica específica.

Capítulo 5: Similitud matemática

En este capítulo abordaremos la similitud matemática expuesta con los pilares del aprendizaje en la red neuronal, de una forma más dogmática, estableciendo características que hay en esa relación de similitud.

5.1 Operadores de información: cerebro humano y redes neuronales artificiales

En esta sección, podemos concluir que ambos operadores de información, tanto el cerebro humano como las redes neuronales artificiales, se basan en principios matemáticos. Mientras que en la inteligencia artificial la estructura es absolutamente matemática, en el caso del cerebro humano, es de base matemática, la acción predominante es la comparación. Generar un modelo respecto de la percepción de características más importantes que otras, recibir verificación o no del modelo y ajustarlo respecto del error, todos estos procesos conllevan la comparación cualitativa y cuantitativa de ciertas variables, en esencia esta es la razón de la matemática: la comparación de variables.

5.1.1 Inteligencia: función con memoria

Dado que ambos sistemas procesan información de manera matemática, podemos modelar la inteligencia como una función matemática. Definimos la inteligencia $f: X \rightarrow Y$ como una función que establece una relación entre las características de entrada X y una salida Y , donde: $f(x) = y$ para cada $x \in X$ y su correspondiente $y \in Y$. Aquí, X representa el espacio de entrada (por ejemplo, características de un problema a resolver), e Y el espacio de salida deseado (como la solución de ese problema).

Es importante notar que la forma: problema – solución como definición formal de inteligencia, deriva en el principio de “*inteligencias múltiples*” (utilizado en el campo psicosocial). Estrictamente representa una función multivaluada. A un problema-dominio específico, dependiendo de la capacidad cognitiva del operador inteligente, existe la posibilidad de asignarle distintas soluciones en el codominio. Generando definiciones particularizadas, con las que este trabajo de investigación intenta lidiar, pues se busca una estructura general para la inteligencia, donde todas esas inteligencias múltiples puedan tener un patrón de convergencia, una especie de intersección de conjuntos cognitivos, que lleve a una definición formal unívoca.

5.1.2 Propiedades necesarias de la función de inteligencia

Con el propósito de dar claridad a la argumentación de cada propiedad que harán de la función del operador que tenga el atributo de inteligencia, no centraremos en un problema; el operador deberá determinar de un conjunto de rostros de perros, cada rostro según su raza. Y para realizar esta tarea de manera “inteligente” consideramos las siguientes condiciones de la función inteligencia:

Inyectividad:

Consideremos que las características principales para detectar los rostros de perros según sus razas son; hocico, orejas y ojos, entonces la idea clave es que, al avanzar por el dominio de estas características principales, las imágenes de las funciones asociadas a esas características no deben ser redundantes. Esto significa que no puede haber dos características diferentes con igual imagen en el codominio. Este principio asegura que el proceso sea consistente y eficiente, sin asignar dos veces la misma imagen a características diferentes.

Formalización de la idea.

Dominio dividido: sea el dominio X dividido en subconjuntos para las características principales:

$$X = X_{hocico} \cup X_{orejas} \cup X_{ojos}$$

Funciones específicas para cada parte del dominio:

Hay funciones $f_{hocico}, f_{orejas}, f_{ojos}$, cada una inyectiva es su subconjunto

$$f_{hocico}: X_{hocico} \rightarrow Y_{hocico} ; f_{orejas}: X_{orejas} \rightarrow Y_{orejas} ; f_{ojos}: X_{ojos} \rightarrow Y_{ojos}$$

La función global f_{Rostro} , que combina todas estas funciones, también debe ser inyectiva:

$$f_{Rostro}: X \rightarrow Y, \quad Y = Y_{hocico} \cup Y_{orejas} \cup Y_{ojos}$$

De esta forma se avanza a través de las características de forma coherente: al trabajar con las características del hocico, las imágenes deben pertenecer al conjunto del hocico; al trabajar con las características de las orejas, deben pertenecer al conjunto de las orejas, y lo mismo para los ojos. Al garantizar que estas funciones sean inyectivas, se asegura de que no haya superposición de imágenes entre las diferentes características, y que el proceso de composición sea claro y sin redundancia.

Composición acumulatividad e inyectividad:

Una forma matemática de representar la función rostro f_{Rostro} , que une todas las imágenes de las características es a partir de la composición de funciones. Para cumplir con este propósito se debe estructurar de forma que cada función parcial (la que procesa las características individuales como hocico, orejas u ojos) dependa de los resultados previos y este integrada en el cálculo global del rostro. La idea es que cada parte del rostro (hocico, orejas, ojos) sea el resultado de una composición que relacione todas las características. Introducimos esta dependencia a través de la composición, la cual también de un sustento al “razonamiento” intrínseco que un operador inteligente debe tener (cada información posterior es el resultado de asimilar y procesar la información anterior: memoria) por ejemplo; la oreja depende del hocico, así que definimos

$$f_{orejas}(x) = g_{orejas}(f_{hocico}(x))$$

Los ojos dependen del hocico y las orejas

$$f_{ojos}(x) = g_{ojos}(f_{orejas}(f_{hocico}(x)))$$

Finalmente, el rostro completo se define como la composición de las funciones de las características principales

$$f_{Rostro}(x) = h(f_{ojos}(f_{orejas}(f_{hocico}(x))))$$

La composición representa la forma acumulativa de la función global; cada función aporta unas características que dependen de las anteriores. El hocico en este caso es el punto inicial, las orejas se construyen sobre la base del hocico, y los ojos se forman considerando las orejas y el hocico.

Inyectividad fundamento: si suponemos que dos características de subconjuntos del dominio distintos (una del hocico y otra de los ojos) dieran la misma imagen, sería una inconsistencia o una aberración en la salida, porque implicaría que dos partes completamente diferentes del rostro están representadas por el mismo resultado, lo cual rompe la lógica en la información del modelo. Entonces la inyectividad evita ambigüedades en las características si f_{hocico} y f_{ojos} no son inyectivas, existe la posibilidad de que $f_{hocico}(x_1) = f_{ojos}(x_2)$, estos dicen que una imagen representa dos características diferentes (un hocico y un ojo), lo cual no tiene sentido en el contexto de un perro.

La inyectividad asegura que cada subconjunto de características del dominio tiene un conjunto de imágenes único en el codominio, en cuanto a la formación de un determinado rostro de perro, por ejemplo: $f_{hocico}(X_{hocico}) \cap f_{orejas}(X_{orejas}) = \phi$, y al definir:

$f_{Rostro}(x) = h(f_{ojos}(f_{orejas}(f_{hocico}(x))))$, se obtiene una consistencia acumulativa preservando la inyectividad. No hay forma de “mezclar” características.

Otro ejemplo que demuestra la lógica de la inyectividad en el cerebro humano es cuando se le pide a un niño que dibuje su familia en una hoja; su cerebro no mezcla características de diferentes miembros; cada persona se construye de manera coherente y separada de las demás. Esto refleja un proceso cognitivamente ordenado de forma secuencial en las características, el cerebro a través de la atención es un operador organizado. El niño toma cada individuo como una unidad independiente y trabaja en sus características de manera ordenada, construyendo un miembro de la familia punto a punto antes de pasar al siguiente.

Continuidad:

En los ejemplos descriptos anteriormente, vemos que un operador inteligente parece procesar la información de manera organizada y secuencial, haciendo que las características se ordenen desde una característica principal a otra, como si las características intermedias mutaran suavemente, formando el camino de una parte del rostro a otra, como del hocico a las orejas; las características a medida que son procesadas por el operador inteligente y se alejan del hocico van mutando a características de las orejas, estas características intermedias producen la conexión entre las características principales manteniendo coherencia informativa a través de la inyectividad y acumulatividad.

Esta “suavidad” en la transformación de las características intermedias se puede expresar de la siguiente manera; a medida que se conectan características del hocico con las orejas, las transformaciones son graduales. Esto significa que las características intermedias no saltan bruscamente de una categoría a otra, sino que van mutando de forma coherente hasta llegar al destino. matemáticamente, esta suavidad podría representarse mediante funciones continuas. Es decir: $f_{hocico}: X_{hocico} \rightarrow Y_{hocico}$; $f_{orejas} : X_{orejas} \rightarrow Y_{orejas}$; $f_{ojos}: X_{ojos} \rightarrow Y_{ojos}$, son todas funciones continuas y por propiedad de composición de funciones:

$$f_{Rostro}(x) = h(f_{ojos}(f_{orejas}(f_{hocico}(x))), \text{ también es continua}$$

De esta forma hemos justificado cada una de las propiedades (inyectividad, continuidad y la forma acumulativa de proceso), que un operador inteligente tiene como condición necesaria, para un correcto manejo de la información.

En los siguientes capítulos, se busca en esta investigación una formalización matemática de la función inteligencia que expresamos en este capítulo, y que nos llevara a entender

rigurosamente la función: operador inteligente que se desprende como una proyección (percepción) de la función inteligencia dentro del codominio. Para ellos se deberá dar estructura al dominio, el codominio y formalizar las propiedades presentadas en este capítulo.

Capítulo 6: Diferencias entre el cerebro humano y el cerebro artificial

En este capítulo se abordará la importante cuestión de las diferencias entre los operadores inteligentes, intentando llegar a una respuesta, al menos a nivel de conjetura. En el libro, *el cerebro matemático* (Dehaene, 2024), aborda esta cuestión desde un punto de vista conceptual, dando al cerebro humano una mayor capacidad de abstracción que a las máquinas actuales. Aquí trataremos de entender cualitativamente este argumento.

6.1 Definición de la información en ambos sistemas

Tanto el cerebro humano como la inteligencia artificial son operadores funcionales que procesan información. Para definir la información de manera general, seguiremos la distinción propuesta por Alvin Toffler y Heidi Toffler (1990) entre "datos" e "información". Los datos se entienden como elementos aislados, sin contexto, que no tienen valor por sí mismos en cuanto a comprensión o significado. Por ejemplo, un dato podría ser "300 acciones". Sin contexto, esta frase no ofrece una interpretación significativa. La información, en cambio, surge cuando estos datos aislados se colocan dentro de un contexto que permite conectarlos y otorgarles un significado. En el ejemplo anterior, al introducir el contexto "300 acciones de la empresa

farmacéutica X", los datos adquieren relevancia y pueden ser utilizados para tomar decisiones o generar conocimiento.

Por lo tanto, podemos decir que la información es el resultado de establecer conexiones entre datos en un contexto determinado. Esta estructura contextual permite que los datos, que por sí solos son elementos independientes, se relacionen y adquieran un valor significativo en conjunto.

6.1.1 Aplicación en ambos sistemas

En este sentido, tanto el cerebro humano como los sistemas de inteligencia artificial pueden considerarse operadores que procesan datos y los transforman en información. En ambos casos, la capacidad de operar sobre los datos y establecer conexiones dentro de un contexto es lo que caracteriza la forma en que cada sistema comprende y utiliza la información.

Lo importante de esta definición es que nos permite tratar la información como un conjunto estructurado, que puede ser descrito y modelado matemáticamente. Tanto en el cerebro humano como en la inteligencia artificial, los datos son tratados de manera cualitativamente similar en términos de contexto y conexiones. Esto nos permite analizar cómo ambos sistemas, aunque diferentes en sus mecanismos, siguen principios similares para el procesamiento de la información.

Este enfoque nos permitirá luego introducir formalmente las diferencias entre ambos sistemas desde una perspectiva matemática, cuando hablemos de topología contextual en el capítulo 8 de ese trabajo.

6.2 Abstracción y plasticidad neuronal

En su libro *Cómo aprendemos* (Dehaene, S. 2020) destaca una de las principales diferencias entre el cerebro humano y las máquinas: la plasticidad neuronal. Esta plasticidad es lo que permite al cerebro humano realizar abstracciones más profundas, establecer conexiones entre diferentes contextos y aplicar lo aprendido de manera flexible. Mientras que la inteligencia artificial va estableciendo métodos y arquitecturas matemáticas para desarrollar esta capacidad, el cerebro humano puede abstraer conceptos a partir de pocos datos y utilizarlos en diferentes situaciones. Esta habilidad se debe a la plasticidad neuronal, la cual posibilita un aprendizaje eficiente y adaptable.

La abstracción es, en esencia, lo que diferencia al cerebro humano de la inteligencia artificial. Esta capacidad de generalización y de encontrar significados profundos, incluso en contextos distintos, le otorga una ventaja sobre los sistemas artificiales actuales.

6.3 Plasticidad funcional y estructural

La plasticidad neuronal se divide en dos tipos principales. Torres, A. (2017):

1. Plasticidad funcional: Se refiere a la capacidad del cerebro para modificar la intensidad de las conexiones sinápticas dentro de una misma área. En otras palabras, el cerebro ajusta la eficacia de las sinapsis en función de la experiencia o el aprendizaje, sin cambiar la estructura física. Esto permite optimizar el rendimiento de ciertas tareas sin necesidad de reasignar funciones a otras áreas del cerebro.

2. Plasticidad estructural: Este tipo implica cambios físicos en el cerebro, como la formación de nuevas conexiones o la reasignación de funciones a diferentes áreas. Es un proceso

más profundo que permite la reorganización cerebral, especialmente en respuesta a lesiones o aprendizajes complejos y prolongados.

Estas dos formas de plasticidad trabajan en conjunto para proporcionar al cerebro humano su flexibilidad única y su capacidad para adaptarse a diferentes contextos de manera eficiente.

Capítulo 7: Información y topología

Según la definición dada en 6.1, la información es básicamente una conexión entre datos, y por las argumentaciones expuestas en el capítulo 5, esa conexión es funcional y con ciertas condiciones necesarias; como la continuidad y la inyectividad en forma acumulativa. Es necesario definir formalmente esta función y sus características, para ello se deberá definir el dominio, el codominio y las reglas de conexión de esta función, como así también las propiedades que caracterizan a los conjuntos en relación.

Como la continuidad es una propiedad fundamental de esta función, la topología surge como una herramienta matemática natural para dar una estructura cualitativa a estos conjuntos sobre los cuales trabaja la función. Pero este no es el único argumento: consideremos el fundamento de la topología, la misma es independiente de estructuras métricas, como longitudes, ángulos o formas geométricas etc. Sin embargo, su argumento está en el tipo de conexión entre elementos de un conjunto, su noción de “acercamiento”, la cantidad de puntos especiales; críticos, de acumulación, etc. Es decir, en la distinción y evaluación de los “obstáculos” que hacen el camino y deberían preservarse en otro camino, sin importar la forma o el largo de estos.

Los operadores inteligentes en este contexto pueden aprender o procesar la información de diferente forma e incluso en distinto tiempo. Alguien puede aprender mejor a partir de argumentos teóricos y otro a partir de situaciones problemáticas a resolver. Sin embargo, deben saber distinguir qué cosas o qué datos son fundamentales para tener un buen resultado. Por esta razón en esta investigación se utiliza la topología como herramienta de modelización cualitativa de la información.

7.1. Cantidad de información, medida de Shannon

Claude Shannon, un ingeniero y matemático, introdujo el concepto de información en términos cuantitativos en 1948. Propuso que la cantidad de información en un mensaje está relacionada con la sorpresa que dicho mensaje genera en el receptor, es decir, con qué tan improbable es el contenido del mensaje. (Shannon, C. 1964) definió una unidad para medir la información llamada bit, como la cantidad de información contenida en un símbolo de un mensaje.

La cantidad de información de un símbolo que aparece con una probabilidad se define como:

$$I = \log_2 \left(\frac{1}{p} \right) = -\log_2(p)$$

Es decir, la cantidad de información es el logaritmo en base 2 del recíproco de la probabilidad del símbolo.

Por ejemplo, si un símbolo tiene una probabilidad de $p = 0,5$ entonces la cantidad de información contenida en ese símbolo es:

$$I = \log_2 \left(\frac{1}{0.5} \right) = \log_2(2) = 1 \text{ bit.}$$

Esta cuantificación permitió medir la información (algo previamente intangible), en una forma concreta y matemática.

Shannon también introdujo el concepto de entropía, que es una medida del promedio de la información contenida en un conjunto de símbolos, ponderada por la probabilidad de aparición de cada símbolo. La entropía de una fuente de información que emite símbolos con probabilidades se define como:

$$H = - \sum_{i=1}^n p_i \log_2(p_i)$$

La entropía es, por tanto, la cantidad promedio de información necesaria para identificar un símbolo de la fuente. Cuanto más impredecible es la fuente (las probabilidades se distribuyen más uniformes), mayor es la entropía.

Por ejemplo, si una fuente tiene dos símbolos con probabilidades y , la entropía de la fuente es:

$$H = -(0.5 \log_2(0.5) + 0.5 \log_2(0.5)) = 1 \text{ bit.}$$

Esto significa que, en promedio, se necesita 1 bit de información para identificar un símbolo de esa fuente.

Shannon estableció, así, una fórmula matemática que le asigna un valor cuantitativo a la información y a la incertidumbre en la transmisión de mensajes.

7.2 Estructura cualitativa de la información

En el apartado anterior, exploramos cómo la información puede ser cuantificada mediante la teoría de Shannon. Sin embargo, entender un concepto en su totalidad no se limita únicamente a medir su cantidad, sino también a conocer su estructura cualitativa. Es decir, necesitamos entender dónde, cómo y con qué eficacia podemos distribuir la información.

Tanto el cerebro humano como la inteligencia artificial funcionan como operadores que procesan y distribuyen información, pero el impacto de esa distribución depende de cómo se estructure y aplique la información en distintos contextos. En este sentido, la información no es solo una cantidad que se mide, sino un recurso que debe ser organizado y utilizado de manera eficiente. Por tanto, es necesario avanzar hacia una comprensión cualitativa de la información, que nos permita identificar sus limitaciones, virtudes y efectos en los diferentes escenarios de aplicación.

7.2.1 Definición formal del conjunto de datos y acciones del operador inteligente

Consideremos un conjunto D , cuyos elementos $d_i \in D$, representan datos aislados, los cuales son considerados como unidades básicas para construir información, es decir, por si solos no transmiten ninguna información, pero constituyen los ladrillos para generarla.

Formalmente el **conjunto de datos** es:

$$D = \{d_i | i \in I\}$$

Posteriormente definimos un **operador inteligente**, que actúa sobre el conjunto de datos D con el pilar del aprendizaje: *atención*, con el propósito de distinguir un subconjunto $P \subset D$, de

datos principales que conectara a través de la función del pilar del aprendizaje: *compromiso activo*, para generar información. El conjunto P es el siguiente:

$$P = \{p_j \mid p_j \in D \wedge j \in J \subseteq I\}$$

Entonces la información se representa como el conjunto de conexiones entre los datos principales detectados por el operador inteligente.

$$I = \{(p_i, p_j) \mid p_i, p_j \in P \text{ y } i \neq j\}$$

El siguiente paso es observar las características de esas conexiones, para que representen de manera adecuada el proceso de un operador inteligente.

7.2.2 Arquitectura del conjunto contexto

Definimos el conjunto contexto denotado por C , como el conjunto cuyos elementos son todas las conexiones entre datos principales generadas por el operador inteligente, lo cual representa información, de esta forma por construcción este conjunto es igual a la unión de toda la información generada por las conexiones de datos principales.

$$C = \cup I$$

Adicionalmente y respecto a su importancia en la generación de información el conjunto contexto clasifica a los datos de la siguiente manera:

Datos principales: construyen la información del contexto.

Datos secundarios: Sus conexiones constituyen la trayectoria de conexión entre datos principales.

Datos inertes: no pertenecen al conjunto contexto por lo tanto son datos aislados sin conexión, que se mantienen en el conjunto de datos D .

Respecto de las características de los elementos de C , y para que estos representen correctamente la acción del operador inteligente es **necesario** que dichas conexiones cumplan con las siguientes propiedades:

1.Continuidad: toda conexión entre datos principales en C es continua, lo que garantiza la no interrupción de la información.

$$I_i = (p_j, p_k) \in C, \text{ es continua}$$

2.Inyectividad: Las conexiones en C son inyectivas, evitando así redundancias en el proceso de generación de información, esto significa que, para cierta información específica, la trayectoria que conecta al par de datos principales (p_j, p_k) aumenta estrictamente la cantidad de información en la dinámica del proceso de entendimiento, formando el camino de conexión.

3.Acumulatividad: Es la capacidad del operador inteligente de generar cada nueva conexión a partir de las conexiones previamente establecidas, es una forma de procesar la información con “razonamiento”, sopesando siempre la memoria de lo aprendido para inferir lo nuevo, en términos formales si el operador conecta los datos principales (p_i, p_j) entonces la conexión (p_j, p_k) actúa sobre la información de (p_i, p_j) , integrando el resultado de las conexiones, manteniendo la coherencia en la secuencia de información generada.

De esta forma se ha estructurado el conjunto contexto C , como el conjunto que contiene toda la información que procesa el operador inteligente, es importante aclarar que **en C no hay**

datos aislados, los elementos son conexiones entre pares de datos a través de trayectorias continuas e inyectivas generadas por un orden secuencial de datos secundarios, que aseguran la capacidad de razonamiento del operador (acumulatividad).

7.2.3 Justificación intuitiva de la primera acción de continuidad del operador inteligente

La continuidad de la información en C se justifica a través de las trayectorias continuas que representan la información de cada conexión entre pares de datos principales. El operador organiza el conjunto de datos secundarios S , de manera que los datos sucesivos en la trayectoria sean poco disimiles entre sí, $s_{i-1} \approx s_i \approx s_{i+1}$, lo cual asegura una variación mínima y progresiva en trayectoria.

Sea

$$\gamma_{p_i, s_j}: [0,1] \rightarrow c$$

Una trayectoria entre un dato principal $p_i \in P$ y un dato secundario

$s_j \in S$, donde

$$\gamma_{p_i, s_j}(0) = p_i$$

$$\gamma_{p_i, s_j}(1) = s_j \text{ .}$$

Dado que los datos secundarios están ordenados y son sucesivamente similares, para cualquier variación pequeña en el parámetro t en el dominio de $[0,1]$, la variación en la trayectoria $\gamma_{p_i, s_j}(t)$ también es pequeña. Esto refleja intuitivamente la continuidad de la

trayectoria, ya que la información transmitida varía de manera estable y coherente conforme a la organización impuesta por el operador inteligente.

7.3 Clasificación de la información contextual

La información en el conjunto contexto C se clasifica en cuatro tipos, en función de su importancia y del tipo de conexión que representan:

7.3.1 información relevante

la información relevante consiste en conexiones entre datos principales. Estas conexiones representan la información central y fundamento de todo el contexto.

Formalmente; si $p_i, p_j \in P$, es una conexión continua $\gamma_{p_i, p_j}: [0,1] \rightarrow C$, con

$$\gamma_{p_i, p_j}(0) = p_i$$

$$\gamma_{p_i, p_j}(1) = p_j$$

7.3.2 Información básica

Consiste en las conexiones entre un dato principal y un dato secundario. Aunque menos prioritario que la información relevante, la información básica aporta estructura al contexto, dado que construye la trayectoria que une los datos principales, además, incluye la acumulatividad del contexto informativo. Una conexión

$$\gamma_{p_i, s_k}: [0,1] \rightarrow C$$

$$\gamma_{p_i, s_k}(0) = p_i$$

$$\gamma_{p_i, s_k}(1) = s_k$$

donde $p_i \in P$ y $s_k \in S$, constituye información básica.

7.3.3 Información inteligente

Las conexiones entre dos datos secundarios conforman la información inteligente, que refleja el procesamiento interno del operador inteligente. Pero no aporta directamente al contexto informativo. por lo tanto, una conexión

$$\gamma_{s_j, s_k}$$

$$s_j, s_k \in S$$

No representa información contextual

7.3.4 Información completa contextual

La información completa es la unión de todas las conexiones de información relevante entre datos principales. Se construye a través de la composición de trayectorias continuas entre datos principales, modelando la acumulatividad del contexto y manteniendo tanto la continuidad como la inyectividad de la información. si consideramos una secuencia de conexiones relevantes

$\gamma_{p_i, p_j}, \gamma_{p_j, p_k}, \dots, \gamma_{p_{n-1}, p_n}$, entre datos principales $p_i, p_j, \dots, p_n$, entonces la información completa es la composición:

$$\Gamma = \gamma_{p_i, p_j} \circ \gamma_{p_j, p_k} \circ \dots \circ \gamma_{p_{n-1}, p_n}$$

Capítulo 8: Espacio topológico de información

Una vez formalizada la estructura del conjunto contexto como el espacio donde el operador inteligente genera y procesa la información, y habiendo argumentado las propiedades deseables de sus elementos—trayectorias continuas, inyectivas y acumulativas—, el siguiente paso natural es definir un espacio que, mediante el pilar del aprendizaje de la retroalimentación, permita al operador inteligente producir las deformaciones necesarias en la unión de trayectorias que conforman el modelo informativo, para acercarse a un mejor resultado. Dichas modificaciones deben preservar la no interrupción de la información (continuidad). Se desea entonces presentar el espacio topológico contextual.

8.1 Definición de la topología contextual

Sea C el conjunto contexto, en el que un operador inteligente genera toda la información posible. Dentro de este conjunto, distinguimos la información relevante y la información básica como la base para la construcción informativa de C . A partir de esta base, definimos el espacio topológico de información (C, τ_C) , donde τ_C representa la topología contextual.

Definimos la *topología contextual* τ_C , como el conjunto de todos los abiertos en C , que representan conexiones informativas donde participa al menos un dato principal. de esta forma un abierto es aquella trayectoria parametrizada continua e inyectiva que conecta un dato principal con otro dato principal, o también un dato principal con un dato secundario.

Formalmente los elementos de τ_C son las trayectorias continuas e inyectivas $T \subset C$ tales que:

T es una conexión entre un par de datos principales (p_i, p_j) ; información relevante o

T es una conexión entre un dato principal p_i y un dato secundario s_j información básica.

$$\tau_C = \{T \subset C/T \text{ es continua e inyectiva y conecta al menos un } p_i\}$$

8.1.1 Algunas aclaraciones de la dinámica que τ_C ejerce en C

Antes de observar que la topología definida, es una buena topología, es importante destacar cómo funciona la τ_C en el conjunto contexto C , ya que será esencial para la demostración de la topología:

Atención y referencia: debido a la función de atención que actúa en el operador inteligente asignando una importancia cognitiva mayor a los datos principales, hace necesario que cada abierto comience su trayectoria de conexión en un dato principal $p_i \in P$ como referencia.

Información relevante, conexión maximal de información básica: como ya se definió la conexión a través de trayectorias continuas e inyectivas de pares de datos principales (p_i, p_j) generan información relevante del contexto. Sin embargo, esta información es producto de un proceso de razonamiento ordenado por parte del operador inteligente.

La conexión entre dos datos principales se considera maximal, pues contine todas las conexiones más pequeñas que en forma anidada la construyen secuencialmente. Comienza con una conexión inicial entre un dato principal de salida y un dato secundario. Luego, a medida que se avanza en la secuencia ordenada de datos secundarios, cada abierto en esta cadena contiene al anterior y se aproxima cada vez más al segundo dato principal de llegada. La conexión maximal entre ambos datos principales, es la unión de todas estas conexiones intermedias.

Este enfoque garantiza que siempre existe un dato principal de referencia en cada abierto de la secuencia. De manera explícita, la secuencia de abiertos que conecta al par (p_i, p_j) toma la forma:

$$(p_i, s_1), (p_i, s_2), \dots (p_i, s_n), (p_i, p_j)$$

Modelizando de esta forma la propiedad acumulativa (memoria y razonamiento) del operador inteligente.

8.1.2 Demostración de topología bien definida en el conjunto contexto

Definición (topología): Sea X un conjunto. Una topología sobre X es una colección de subconjuntos $\tau \subseteq P(X)$ (donde $P(X)$ es el conjunto potencia de X) que satisface las siguientes condiciones:

1. El conjunto vacío y el espacio total están en τ :

$$\emptyset \in \tau, \quad X \in \tau$$

2. Cerradura bajo uniones arbitrarias:

Si $U_{i \in I}$ es una familia arbitraria de conjuntos en τ , entonces su unión también pertenece a τ :

$$\bigcup_{i \in I} U_i \in \tau$$

3. Cerradura bajo intersecciones finitas:

Si $U_1, U_2, \dots, U_n \in \tau$, entonces su intersección también pertenece a τ :

$$\bigcap_{i=1}^n U_i \in \tau$$

La pareja (X, τ) se denomina espacio topológico.

Demostración de la topología contextual:

i) $\phi \in \tau_c$

Por definición, el conjunto vacío ϕ pertenece a cualquier topología, y en este caso, se interpreta como la ausencia de información relevante o básica en el conjunto contexto C . Esto ocurre cuando no hay conexiones que involucren al menos un dato principal, es decir solo hay conexión entre datos secundarios, que no aportan información contextual al operador inteligente. Dada esta característica, el operador la considera vacía o carente de información contextual. Por lo tanto, el vacío pertenece a la topología contextual y asume la situación en la que no se construye ninguna información que conecte a un dato principal.

ii) $C \in \tau_c$

Para demostrar que $C \in \tau_c$, necesitamos ver que puede ser considerado un abierto en la topología contextual, es decir, que cada uno de sus elementos, tiene al menos un dato principal conectado. Comencemos recordando; C es un conjunto informativo, generado por el operador inteligente a partir de los pilares de la *atención* y el *compromiso activo*. Los elementos de C son

conexiones entre datos, de tres tipos: conexión entre datos principales (p_i, p_j) , conexión entre dato principal y dato secundario (p_i, s_k) , y conexión entre datos secundarios. veamos que C es información relevante contextual completa, es decir una conexión entre el primer dato principal y el ultimo. En efecto por 7.2.2, la información completa en C es la trayectoria continua, inyectiva y acumulativa que va desde primer dato principal hasta el último.

Las conexiones entre datos secundarios no cumplen con la propiedad acumulativa, pues no tiene memoria del dato principal de referencia origen de la conexión, luego, no ofrecen aporte cognitivo al proceso de aprendizaje del operador inteligente, siendo indistinguibles del vacío. De esta manera el operador inteligente considera a C una conexión entre datos principales, luego $C \in \tau_c$.

$$\text{iii) } \bigcup_{i \in I} A_i, A_i \in \tau_c$$

Demostrar esta condición, es equivalente a decir que cada conexión $x \in \bigcup_{i \in I} A_i$ perteneciente a la unión de los abiertos A_i , tiene un entorno abierto $B_x \subset \bigcup_{i \in I} A_i$ al que pertenece.

Sea $x \in \bigcup_{i \in I} A_i$, donde x es una conexión contenida en algún A_j de la colección de abiertos $A_{i \in I}$. Dado que A_j es un abierto en la topología contextual, existe un entorno abierto $B_x \subset A_j$ que contiene a x y preserva una conexión con al menos un dato principal de referencia. Este entorno B_x puede considerarse un “abierto base” de x (información básica) pues forma parte de la secuencia de abiertos anidados contruidos a partir de la conexión maximal entre datos principales (información relevante). Dado que $x \in B_x \subset A_j \subset \bigcup_{i \in I} A_i$ y, además, debido a la

estructura secuencial de la topología contextual, todos los elementos en B_x mantienen la conexión con el dato principal de referencia. Concluimos que: la estructura maximal y la anidación de abiertos alrededor de cada conexión aseguran que, para cualquier conexión en la $\bigcup_{i \in I} A_i$, existe un abierto que contiene dicha conexión y pertenece a la unión. Cumpliendo con la demostración que: $\bigcup_{i \in I} A_i \in \tau_C$, tal cual se quería demostrar.

$$\text{IV) } \bigcap_{i \in I}^n A_i, A_i \in \tau_C$$

Para demostrar que la intersección de abiertos es un abierto, nos centraremos en los dos casos especiales que requieren cierta profundidad argumentativa:

$$\begin{array}{cc} \bullet & \bullet \\ p_i & s_i \end{array}$$

Primer caso: sean A_i y A_j dos abiertos en la topología contextual, ambos conteniendo un dato principal común p . Supongamos que A_i es una trayectoria que conecte p con un dato secundario s_i , y A_j conecta p con dato secundario s_j , con s_i “más cercano” a p que s_j , respetando el orden secuencial de los datos secundarios y la acumulatividad de las trayectorias. Como $A_i \subseteq A_j$, tenemos que la interacción $A_i \cap A_j = A_i$ que es un abierto en esta topología.



Segundo caso: sea A_i y A_j dos abiertos donde A_i conecta un dato principal p_i con un dato secundario s_j , y A_j conecta con otro dato principal p_j con otro dato secundario s_i , sin compartir un dato principal común. Como esta intersección no contiene ninguna conexión que involucre un dato principal, entonces, de acuerdo con la definición, no constituye información contextual. Por lo tanto, desde el punto de vista del operador inteligente, esta intersección se

interpreta como vacía (inerte) y es equivalente al conjunto vacío $\emptyset$, que es un abierto en la topología. Por tanto, la intersección finita de abiertos es también un abierto, como se desea demostrar.



8.1.3 Disquisición sobre el vacío informativo

En un conjunto contextual, los datos secundarios tienen un papel ambivalente: pueden enriquecer el marco informativo cuando están conectados a datos principales, pero en ausencia de esta conexión se convierten en elementos inertes, carentes de valor contextual. Este vacío informativo surge cuando los datos secundarios solo se conectan entre sí, sin referencia a los datos principales que anclan la estructura del contexto. Para ilustrar este concepto, imaginemos un ejemplo práctico en el que el conjunto contextual representa la arquitectura de todas las casas de dos habitaciones. Consideremos que, en este conjunto, los datos principales son elementos como las habitaciones y los ambientes esenciales de la casa: el comedor, el baño, la cocina, etc. Estos datos principales permiten reconocer la disposición básica y la estructura del espacio habitacional. Ahora, imaginemos los datos secundarios como aquellos elementos que no determinan de forma directa la estructura fundamental de la casa, como los pasillos de conexión entre habitaciones y ambientes.

Si un pasillo está conectado a un dato principal, como una habitación o el comedor, este pasillo adquiere sentido en el contexto arquitectónico: nos informa sobre la disposición relativa de los espacios y ayuda a definir cómo fluye la circulación en el interior de la casa. En cambio, si

solo disponemos de información acerca de un pasillo desconectado, sin relación con ningún dato principal, este pasillo queda flotando como un elemento informativamente vacío. No contribuye a nuestra comprensión de la casa ni a su disposición espacial; se convierte en un elemento aislado que, aunque puede formar parte del conjunto de datos, no aporta información relevante para el operador inteligente.

Por lo tanto, el valor informativo de los datos secundarios depende de su conexión con los datos principales que conforman la estructura del conjunto contextual. Sin esta conexión, los datos secundarios no enriquecen el contexto ni proporcionan una comprensión más profunda del mismo; se presentan como fragmentos aislados, incapaces de aportar referencias útiles para el operador. Este tipo de conexiones secundarias exclusivamente entre datos secundarios, sin anclaje en los datos principales, crea un vacío informativo, un espacio de potencial latente que no cumple ninguna función informativa hasta que se conecta con un núcleo estructural.

En conclusión, los datos secundarios en el conjunto contextual solo adquieren significado contextual cuando se relacionan con datos principales. De otro modo, existen como información inerte, sin aportar valor al entendimiento del contexto general. Esto refuerza la importancia de las conexiones estratégicas entre datos en la generación de un contexto informativo cohesivo y útil para el operador inteligente. Como así también justifica la definición de los abiertos de la topología contextual.

8.2 Propiedades del conjunto contexto

En los siguientes apartados, se intentará demostrar y justificar conceptualmente una serie de condiciones necesarias y suficientes, intrínsecas al conjunto contexto C , para que el operador inteligente, pueda procesar la información contextual con eficiencia sin perder eficacia.

8.2.1 Cerradura

Definición de punto de acumulación:

Sea X un espacio topológico y $A \subset X$ un subconjunto de X . Un punto $p \in X$ es un punto de acumulación de A si para cualquier entorno abierto $U \subset X$ tal que $p \in U$, se cumple que $U \cap (A \setminus p) \neq \emptyset$. En otras palabras, cualquier entorno de p contiene puntos de A distintos de p mismo.

Aplicación a los datos principales en el conjunto contexto C :

Consideramos el conjunto contexto C , donde los elementos representan conexiones de datos que incluyen tanto trayectorias entre datos principales como conexiones entre datos principales y secundarios. En este marco, podemos considerar a los datos principales como puntos esenciales para la estructura informativa del contexto, de modo que cada dato principal $p \in C$ cumple con la definición de punto de acumulación en relación con los datos secundarios que conecta.

1. Trayectoria continuas con p como referencia:

Para un dato principal p , las trayectorias continuas que parten de p se extienden secuencialmente a través de datos secundarios. Estos significan que todos los abiertos que representan trayectorias desde p incluyen a p y también a datos secundarios conectados a él.

2. Condiciones de punto de acumulación:

Dado cualquier entorno abierto U de p en C , se tiene que $U \cap C$ contiene puntos de la trayectoria continua conectados con p , específicamente, algunos datos secundarios que forman parte de la conexión informativa. Por lo tanto, para cualquier entorno abierto U que contenga p , se cumple que $U \cap (A \setminus p) \neq \emptyset$, lo cual satisface la condición para que p sea un punto de acumulación en relación con el conjunto contexto C .

3. Implicación de los datos principales como puntos de acumulación:

Como cada dato principal cumple esta propiedad, podemos concluir que todos los datos principales en C son puntos de acumulación en relación con las trayectorias de datos secundarios que conectan. Esto subraya que los datos principales, no solo son esenciales en la información contextual, sino que cumplen matemáticamente con la definición de puntos de acumulación. Esto permite evidenciar la naturaleza acumulativa y conectiva del conjunto C .

Para demostrar que el conjunto contexto C es cerrado en el espacio topológico que hemos definido, utilizaremos dos enfoques que refuerzan la idea de cerradura:

1. Primera demostración: contención de los puntos de acumulación: para que un conjunto sea cerrado, debe contener todos sus puntos de acumulación. En nuestro caso, él

conjunto contexto C esta definido como un conjunto de conexiones informativas, donde los datos principales y sus conexiones cumple un papel central en la estructura de la información.

Finitud de los datos principales: según el operador inteligente, los datos principales son finitos, ya que son seleccionados cuidadosamente para definir de manera univoca la información completa en el contexto. Está elección finita asegura que todas las trayectorias relevantes, que conectan tanto datos principales como secundarios, están contenidas en C .

Contención de los puntos de acumulación: como hemos demostrado, cada dato principal actúa como un punto de acumulación en el sentido topológico, dado que cualquier entorno que lo contenga también contiene conexiones hacia otros datos secundarios de C . Esto significa que los puntos de acumulación de C están incluidos en C , y dado que no existen puntos de acumulación fuera de C , podemos concluir que C contiene todos sus puntos de acumulación. Luego C es cerrado por esta definición.

2. Segunda demostración: complemento abierto:

Alternativamente, un conjunto también se considera cerrado si su complemento es un conjunto abierto.

Definición del complemento: en este caso, para el operador inteligente, no existe información fuera del contexto C ; cualquier dato que se encuentre fuera de C es aislado, no tiene conexión ni aporta a la estructura informativa contextual. Esto implica que, en términos informativo el complemento de C , se considera en conjunto vacío.

Apertura del complemento: dado que el conjunto vacío es un conjunto abierto de la topología concluimos que el complemento C es abierto. Por lo tanto C es cerrado por definición en el sentido topológico.

Hemos demostrado la cerradura de C a través de esto dos enfoques, no por un motivo simplemente matemático, sino por razones conceptuales; que C , contenga todos sus puntos de acumulación como el conjunto de sus datos principales, los cuales son finitos, implica que la información contextual puede determinarse completamente. Y que el complemento de C sea el conjunto vacío, implica que las trayectorias de conexión no tienen puntos o datos fuera de C , permitiendo una interpretación coherente de la información por parte del operador inteligente.

8.2.2 Separabilidad

Para justificar la separabilidad del conjunto contextual, recordemos la definición topológica de separabilidad. Un espacio topológico es separable si contiene un subconjunto denso y numerable. Esto significa que existe un subconjunto tal que, cualquier entorno abierto del espacio, contiene al menos un punto de este subconjunto denso.

Dado que hemos demostrado que el conjunto contextual es cerrado —es decir, contiene todos sus puntos de acumulación—, consideremos el conjunto de estos puntos de acumulación. Observamos que este conjunto es finito, y debido a la acumulatividad de la información en el conjunto, estos puntos de acumulación pueden interpretarse como densos en el contexto, ya que cualquier entorno abierto del conjunto contiene puntos de este subconjunto. Así, el conjunto de

puntos de acumulación dentro del contexto cumple con la definición de un subconjunto separable en el contexto.

Esta propiedad de separabilidad, dada por los puntos de acumulación, otorga al conjunto contextual una estructura de orden en la organización de la información. Los puntos de acumulación actúan como datos principales o puntos de referencia, proporcionando un marco para la organización de los datos en el conjunto contextual.

En resumen, la existencia de un subconjunto denso de puntos de acumulación asegura que el contexto sea separable y estructurado en torno a datos principales, facilitando así una organización jerárquica y eficiente de la información.

8.2.3 Compacidad

La compacidad del conjunto contexto es quizás la propiedad más interesante, ya que permite al operador inteligente expresar eficiencia en el manejo de la información, con esta propiedad el operador no necesita conocer todas las trayectorias de conexión entre datos principales, para determinar completamente el conjunto contexto C , detectando solo un conjunto de trayectorias finitas podrá conocer todo el contexto informativo. Sin duda ese es el atributo más importante de un operador inteligente; el manejo eficiente de la información.

Consideremos un conjunto contexto C , que un operador inteligente por medio de la atención expresa a través de los datos principales $p_1, p_2, \dots, p_n$ y sean estos datos principales los

puntos de acumulación que permiten generar trayectorias de conexión, continuas e inyectivas de manera acumulativa entre sí.

Para formalizar las trayectorias definimos:

trayectorias de conexión: para cada par de datos principales (p_i, p_{i+1}) , definimos una trayectoria continua (por 7.2.3) parametrizada:

$$\gamma_{i,i+1}: [0,1] \rightarrow C$$

$$\gamma_{i,i+1}(0) = p_i \text{ (dato principal de salida)}$$

$$\gamma_{i,i+1}(1) = p_{i+1} \text{ (dato principal de llegada)}$$

$\gamma_{i,i+1}(t)$, para $t \in (0,1)$, representan los datos secundarios ordenados en forma secuencial, y a partir del dato principal de salida p_i como referencia en cada abierto forman la trayectoria de conexión con el dato principal de salida p_{i+1} conservando la acumulatividad del proceso con la relación de inclusión de abiertos donde el abierto maximal es el que conecta (p_i, p_{i+1})



Además, se postula la inyectividad de estas trayectorias, es decir $\gamma_{i,i+1}(t_1) \neq \gamma_{i,i+1}(t_2)$ si $t_1 \neq t_2$, lo que garantiza que el operador inteligente no repite datos en una misma trayectoria (no redundancia).

Equivalencia homeomorfa de las trayectorias: las trayectorias de conexión entre los datos principales son homeomorfas entre sí, lo que significa que cualquier trayectoria $\gamma_{i,i+1}$, puede

transformarse en otra trayectoria $\gamma_{j,j+1}$, teniendo en cuenta las propiedades definidas en el inciso anterior mediante un homeomorfismo:

$$h: \gamma_{l,i+1}[0,1] \rightarrow \gamma_{j,j+1}[0,1]$$

debemos demostrar que h es un homeomorfismo, es decir debe ser continua, inyectiva, sobreyectiva y su inversa h^{-1} también debe ser continua. $\gamma_{l,i+1}$ conecta (p_i, p_{i+1}) y $\gamma_{j,j+1}$ conecta (p_j, p_{j+1}) , pares de datos principales respectivamente.

Sea $h: [0,1] \rightarrow [0,1]$, una función que mapea cada punto de la trayectoria $\gamma_{l,i+1}$, a un punto correspondiente en la trayectoria $\gamma_{j,j+1}$, de manera que:

$$h(\gamma_{l,i+1}(t)) = \gamma_{j,j+1}(t), \forall t \in [0,1]. \text{ Entonces, la función } h \text{ satisface:}$$

inyectividad: pues ambas trayectorias son inyectivas por definición de conexión entre datos principales.

Surjectividad: pues cada punto (dato) en la trayectoria $\gamma_{j,j+1}[0,1]$, tiene una preimagen en la trayectoria dominio $\gamma_{l,i+1}[0,1]$, es decir todos los puntos en la trayectoria de llegada están cubiertos. Las trayectorias son continuas, inyectivas y acotadas por los datos principales de salida y llegada respectivos.

Continuidad de h : dado un abierto de la topología contextual en la trayectoria imagen $\gamma_{j,j+1}(t)$, su imagen inversa por h en la trayectoria dominio $\gamma_{l,i+1}(t)$ también es un abierto. En efecto, pues la acumulatividad asegura un proceso de generación de las trayectorias entre datos principales, donde cada abierto parte del dato principal inicial como referencia y por la relación

de inclusión cada abierto sucesivo contiene al anterior, hasta conectar con el dato principal de llegada, y a partir de la inyectividad podemos conectar por h ambas trayectorias punto a punto en cada abierto, asegurando la continuidad topológica de la función de transformación h .

La demostración de que la función inversa h^{-1} , cumple con las mismas propiedades es trivial y análoga a la realizada con h , por tanto, se omite en esta investigación.

Este homeomorfismo asegura que cualquier trayectoria de conexión entre datos principales, es decir de información relevante, puede transformarse en cualquier otra sin pérdida de continuidad informativa, lo cual permite que todas las trayectorias posibles de conexión entre datos principales en C sean equivalentes topológicamente. Luego podemos seleccionar una cantidad finita de trayectorias base $\gamma_{1B}, \gamma_{2B}, \dots, \gamma_{kB}$ que conectan pares de datos principales, y a través de homeomorfismos adecuados, cubrir toda la información contextual del conjunto contexto C , cumpliendo con la definición de compacidad topológica, Lo cual asegura el manejo eficiente que queríamos demostrar.

Nota: las propiedades de cerradura con los puntos de acumulación aseguran al operador inteligente, que es posible determinar toda la información contextual. La separabilidad, le permite establecer un orden respecto a los datos principales, con el propósito de establecer una forma sistemática de encontrar esa información, estas son condiciones necesarias para el operador, y por último la compacidad del conjunto contexto C , le otorga una condición de suficiencia para encontrar toda la información contextual de manera manejable y finita.

Capítulo 9: Dinámica de la topología contextual: cerebro humano y red neuronal artificial

Habiendo construido una estructura cualitativa y argumentada para el manejo matemático de la información, queda aplicar esta estructura en relación con establecer una causa razonable, que permita identificar la diferencia en la plasticidad de los operadores inteligentes; cerebro humano y red neuronal.

9.1 Las mismas herramientas de trabajo

Tanto la inteligencia artificial (IA) como el cerebro humano operan en contextos cerrados. Esto significa que los conjuntos contextos sobre los que trabajan incluyen todos sus puntos de acumulación, permitiendo una estructura completa para la manipulación de la información. Además, utilizan la misma topología contextual donde la información se genera y se transforma a partir de abiertos que respetan las propiedades básicas de un tratamiento “inteligente” de la información, conexiones continuas, inyectivas en un marco de referencia con memoria (acumulatividad). Y respecto al conjunto informativo o contexto este ofrece las condiciones necesarias y suficientes para que el operador inteligente procese la información de manera eficaz y eficiente; cerradura, separabilidad y compacidad.

9.2 Rigidez topológica: conjetura de la red neuronal artificial

La rigidez es un concepto físico que permite preservar la distancia relativa entre las partículas de un sistema, es decir es una característica geométrica del sistema estudiado. Como se ha observado en esta investigación, las redes neuronales son funciones matemáticas definidas en el espacio euclídeo: $F: \mathbb{R}^n \rightarrow \mathbb{R}^m$, es decir a la estructura topológica abstracta y general que se definió en esta investigación, la red neuronal le asocia una estructura pragmática, donde la

topología contextual se vuelve la topología usual de $\mathbb{R}^n$, y los conjuntos contexto se transforman en los compactos de $\mathbb{R}^n$ definidos por el teorema de Heine-Borel (cerrados y acotados). De esta manera la red neuronal se hace de las bondades de este espacio, sin desestimar las condiciones del espacio informativo general, nos referimos por ejemplo a la métrica usual en el espacio euclídeo, a propiedades como ser T2 entre otras. Sin embargo, estas cualidades son a costo de acotar la flexibilidad de la topología contextual, mostrando una dinámica topológica más rígida.

En el espacio euclídeo, los compactos ya están definidos, y la red neuronal una vez seleccionado los datos principales, por medio de la atención, los circunscribe a un intervalo predefinido compacto, que ya contempla una cantidad de datos secundarios infinita por definición de intervalo en $\mathbb{R}^n$, ***de allí que las redes neuronales necesitan grandes cantidades de datos para procesar la información con menor error.*** En ese compacto la red neuronal tiene como elementos las conexiones de pares de datos, es decir no hay dato aislado, conecta los datos principales a partir de conexión de datos secundarios formando las trayectorias continuas, inyectivas y con acumulatividad, y los datos secundarios que no participen de esas trayectorias también los conecta pero como ya se definió en esta investigación, al no tener como referencia un dato principal son conexiones sin información contextual, son trayectorias que no tienen principio ni final definidos. La red neuronal al recibir por medio de la retroalimentación (retropropagación) una señal de error en su salida de información ajusta sus parámetros, y allí conecta una nueva trayectoria de datos secundarios a los datos principales, es decir las conexiones inertes de información del conjunto compacto, se vuelven ahora información relevante, a partir de la acción del operador inteligente con los pilares del aprendizaje. En estas condiciones dinámicas, para expandir el conjunto contexto o el universal de información, la red neuronal por las restricciones topológicas que establece el espacio euclídeo necesita conocer por

medio de la *atención* un nuevo conjunto de datos principales en un conjunto compacto predefinido y de esa manera generar la unión de ambos contextos con conexiones adecuadas utilizando el pilar del *compromiso activo*.

9.3 Flexibilidad topológica: conjetura contextual del cerebro humano

En contraposición con el inciso anterior, se puede asociar al cerebro humano con una dinámica topológica sustancialmente flexible. El límite topológico de nuestro cerebro, no son contextos predefinidos, son contextos conceptuales, es decir lo esencial en el manejo de la información es que las conexiones sean coherentes con las propiedades definidas en la estructura cualitativa de la información aportadas en esta investigación, el conjunto contexto se expande si y solo si las trayectorias siguen siendo continuas, inyectivas y mantienen la acumulatividad o referencia a datos principales. El cerebro humano por medio de la atención selecciona un conjunto de datos principales y secundarios determinados, luego con el pilar de compromiso activo, los conecta utilizando todos los datos, no quedando ningún dato sin participar de la generación de la información contextual completa. Cuando recibe por medio de la *retroalimentación* una señal de error en la comprensión de la información, o cuando debe generar nueva información contextual, entonces aplica transformaciones homeomorfas a la trayectoria anterior generando una nueva. ***De ahí que no necesita una cantidad grande de datos para establecer la información, lo que redunde en gran eficiencia.***

9.4 Comparación dinámica: conjetura de la abstracción

Teniendo en cuenta la rigidez de la red neuronal y la flexibilidad del cerebro humano en el manejo de la información con la dinámica topológica, podemos ver:

En la red neuronal los datos están contenidos en un conjunto contexto ya predefinido (intervalos compactos de $\mathbb{R}^n$), la red solo puede asignar importancia a aquellos datos que expresan mejor la información, pero no puede desechar sin argumento los datos secundarios que no aporten información contextual, debe conectarlos igualmente entre ellos, esto supone una estática de los datos, es decir la información no se transforma, sino que cambia, la acumulatividad se logra por el reconocimiento completo de todas las conexiones, como si fuera un “aprendizaje de memoria”.

En el cerebro humano la dinámica topológica hace que los datos estén definidos por el operador inteligente y no por simple “acercamiento” relativo, la atención hace que el cerebro seleccione datos coherentes conceptualmente. Para generar nueva información dentro del contexto de trabajo, el cerebro humano transforma los datos secundarios sin cambiar su esencia, es decir los conecta de diferente forma, para generar una nueva trayectoria de información que conecta los mismos datos principales. Esta estructura supone datos secundarios dinámicos, lo que lleva a un “aprendizaje más razonado” que en las redes neuronales artificiales.

La red neuronal une contextos y de esa forma es que amplía el universal de la información, conoce la información completa de un contexto y conoce la información completa de otro contexto, y luego es capaz de generar una trayectoria de conexión entre elementos de ambos contextos. Esto sugiere una plasticidad estructural limitada en cuanto a la flexibilidad, y se apega más a un operador literal, necesita seguridad del conocimiento amplio de los contextos para producir más información. Sin duda gana en eficacia, en cuanto evita errores métricamente apreciables.

El cerebro humano puede fácilmente, sin romper las condiciones de conexión, expandir el conjunto contexto, reconociendo en el conjunto de datos inertes, por medio de la *atención* un dato que considere principal, y de esta forma con las conexiones asociadas a ese nuevo dato, ofrecer una ampliación del universal de información. Esto da al cerebro humano una plasticidad estructural elogiada y ampliamente flexible destacando en una gran eficiencia en el proceso de la información.

La red neuronal artificial produce con el conjunto de datos principales y datos secundarios, trayectorias de conexión que construye punto a punto (*o en forma paralela como es el caso de los Transformers, pero la verificación sigue siendo secuencial*), como si fuera dibujando un retrato en una hoja, detalle por detalle, su inteligencia está en detectar por la *Retropropagación* donde el pulso fallo en el dibujo y tratar de producir una mejor copia.

El cerebro humano con los datos principales y secundarios genera un modelo completo del retrato, produce la conexión general de acuerdo con el contexto, en principio sin importar los detalles finos. Luego con la *retroalimentación*, ajusta el modelo con la dinámica topológica.

Estos puntos de comparación expresan según esta investigación, las diferencias sustanciales entre ambos operadores de información inteligente. La red neuronal artificial al tener limitada plasticidad estructural muestra una merma en su capacidad de abstracción, por ejemplo, en la producción de analogías coherentes, es decir utilizar información de un contexto en otro, sin tener un conocimiento de los datos principales del nuevo contexto. pero como contraprestación genera información segura al tener un uso eficaz de los datos y conexiones, en contextos bien conocidos.

Por su parte el cerebro humano evidencia gran plasticidad estructural, lo cual significa versatilidad en el uso y mezcla de la información, esto genera capacidad creativa en el uso de la información, por ejemplo, en el arte, el cerebro humano puede, a imágenes sumamente distorsionadas desde la geometría evaluarlas en cuanto a calidad desde otros aspectos. Para fijar la idea; cuando vemos una pintura de un artista famoso, y nos dicen que representa una figura humana, quizás las formas no capten en nada la forma humana, pero la transmisión abstracta de la imagen nos hace percibir el sentir humano, y eso lo califica como imagen humana.

Entonces podemos concluir en cuanto a la abstracción, que es un concepto vinculado más con la eficiencia que con la eficacia, por lo tanto, el manejo de estos atributos hará un operador inteligente con mayor capacidad de abstracción o no. En esta investigación, y bajo estas prerrogativas la red neuronal es más eficaz y el cerebro humano es más eficiente, en cuanto al manejo que ambos operadores hacen de la información contextual. Luego esto explica ***la diferencia en la abstracción que mantiene el cerebro humano respecto a su operador contrincante artificial.***

9.4.1 La tasa y la dona: grafican la idea

Una analogía común en topología es la equivalencia entre una tasa y una dona (torus). Ambos objetos son homeomorfos, lo que significa que son topológicamente equivalentes y pueden transformarse uno en otro mediante un homeomorfismo, sin cambiar sus propiedades topológicas fundamentales.

Red neuronal artificial: si comienza con una tasa, necesita los datos principales de la dona para realizar la transformación. Es decir, la red debe conocer explícitamente las características clave de ambos contextos (la tasa y la dona) para establecer el homeomorfismo:

$$f: T_{taza} \rightarrow T_{dona}$$

Cerebro humano: El cerebro humano, por el contrario, y debido a la abstracción, puede partir de una tasa y transformarla en una dona sin requerir información adicional sobre esta última. A partir de los datos disponibles de la taza, el cerebro humano genera internamente un homeomorfismo hacia la dona, transformando el contexto inicial. El cerebro deforma el contexto de la tasa hasta obtener la dona, sin la necesidad de acceder a un segundo conjunto de datos previamente definido. Conceptualmente al cerebro humano de forma natural vincula una taza de café con una dona, en el mismo contexto informativo.

En ambos casos, los contextos son cerrados y topológicamente equivalentes, pero la IA necesita conocer de manera explícita ambos conjuntos para ejecutar la transformación, mientras que el cerebro humano puede transformarlos a partir de un solo contexto inicial de manera flexible.

Capítulo 10: Poniendo en práctica la investigación (evaluando los Transformers)

Este trabajo, se ha sustentado en los pilares del aprendizaje del cerebro humano, y el principio de aproximación universal de las redes neuronales artificiales hacia adelante, las cuales constituyen la forma más simple de arquitectura de red. En consecuencia, se puede decir la investigación se construye a partir de un principio minimal, que se puede denominar: principio de mínimo encajado, esto significa; si las argumentaciones se cumplen para el menor conjunto (red neuronal hacia adelante, arquitectura más simple) entonces como estas están incluidas en otras arquitecturas más complejas, por inclusión de conjuntos también se cumplirán en el conjunto mayor.

Para comprobar lo anterior, utilizaremos la arquitectura de red Transformers, la cual significa un salto cualitativo, fundamento de las redes neuronales generativas más complejas.

10.1 Transformers

El Transformers, presentado por Vaswani et al. (2017) en "Attention is All You Need", propone una arquitectura basada en **mecanismos de atención** que prescinde de la recurrencia. Esta arquitectura permite una mayor **paralelización** en el entrenamiento y un modelado más eficiente de las dependencias a largo plazo en las secuencias.

Supongamos un texto que el Transformers debe aprender, las palabras en el texto reciben el nombre de token lo cual es un vector que codifica la semántica de la palabra y representa la materia prima por el cual la red neuronal artificial se activa.

Primero, cada token de la secuencia (frase completa) se representa mediante un embedding que se suma a una codificación posicional generada a través de funciones seno y coseno. Esta codificación se utiliza para indicar la posición de cada token en la secuencia. La fórmula para calcular la codificación posicional de la posición i y la dimensión $2k$ es:

$$P_{(i,2k)} = \sin\left(\frac{i}{10000 \frac{2k}{d}}\right), \quad P_{(i,2k+1)} = \cos\left(\frac{i}{10000 \frac{2k}{d}}\right)$$

donde i es la posición del token, k el índice de la dimensión, y d la dimensión total del embedding.

Una vez que los tokens están codificados, se transforman en tres vectores: Query (Q), Key (K) y Value (V), mediante multiplicación por matrices de pesos entrenables W_Q , W_K y W_V , respectivamente:

$$Q = E'W_Q, \quad K = E'W_K, \quad V = E'W_V,$$

donde E' es el embedding de entrada, y W_Q , W_K y W_V son matrices de pesos de dimensión $d \times d_k$, $d \times d_k$ y $d \times d_v$, respectivamente, que son inicializadas con valores aleatorios y luego ajustadas durante el entrenamiento.

Luego, para calcular la atención entre los tokens, se realiza un producto escalar entre Q y K , normalizado por la raíz cuadrada de la dimensión de los vectores d_k :

$$Atencion = \text{Softmax}\left(\frac{QK'}{\sqrt{d_k}}\right).$$

La función Softmax convierte los valores obtenidos en una distribución de probabilidad, asignando a cada token una importancia relativa para cada posición de la secuencia. Esto da una probabilidad para cada token en relación con todos los demás, indicando cuánta atención debe prestarse a cada token dentro de la secuencia, independientemente de su distancia.

Finalmente, el vector Value (V) se multiplica por la matriz de atención para obtener la representación final de cada token en la secuencia.

Este proceso se repite a través de varias capas de atención y luego se pasa al decoder, que toma las representaciones de atención para generar la secuencia de salida. El decoder revisa secuencialmente las palabras anteriores y ajusta su salida en función de la coherencia semántica entre los tokens ya generados, usando el mismo mecanismo de atención para garantizar que cada palabra generada sea consistente con las anteriores.

10.2 Paralelización vs recurrencia secuencial

Para comprender de manera gráfica las ventajas que el Transformers aporta sobre las redes neuronales recurrentes de análisis secuencial, es decir token por token con una ventana de proximidad definida. Se hará uso de la capacidad del cerebro humano para generar contextos de información análogos.

Supongamos el modelo atómico con núcleo, orbitales y electrones con los cuales se estructura un átomo. El núcleo otorga el contexto de la información, los orbitales las posiciones de las palabras, y la configuración electrónica (regla de las diagonales), los electrones – palabras a ubicar. Con estas analogías estamos en condiciones de distinguir la forma de trabajo de ambas arquitecturas redes recurrentes y Transformers:

RNNs: "Encontrando Electrones y Orbitales": Las RNNs procesan la información de la frase de forma secuencial. En la analogía atómica, esto significa que intentan "ubicar" los electrones (tokens) y **al mismo tiempo "encontrar" las posiciones de los orbitales (representaciones posicionales)** en un espacio de probabilidad. No tienen una representación posicional estática, sino que deben descubrirla durante el entrenamiento. Esto implica que deben aprender simultáneamente la representación semántica de los tokens y sus posiciones relativas en la secuencia, lo que hace que el entrenamiento sea computacionalmente costoso y dificulta el modelado de dependencias a largo plazo.

Transformers: "Ubicando Electrones en Orbitales Predefinidos": El Transformers, por otro lado, predefine las posiciones de los orbitales (codificaciones posicionales) mediante funciones seno y coseno. En nuestra analogía, los "orbitales" (posiciones) están predefinidos y **no se modifican durante el entrenamiento**. El encoder y decoder, a través de la autoatención, se enfocan en **"ubicar" los electrones (tokens) en estos "orbitales" predefinidos**, ajustando la importancia de cada posición para cada token a través de los pesos de atención. Es decir, el modelo no modifica la representación de la posición, sino que aprende a ponderar la importancia de cada una, ajustando la "configuración electrónica" (la importancia de cada posición para cada token).

Paralelización y Eficiencia: La principal ventaja del Transformers es su **capacidad para paralelizar el entrenamiento**. Mientras que las RNNs requieren un procesamiento secuencial, las capas de autoatención del Transformers pueden operar simultáneamente sobre todas las posiciones de la secuencia. Esta paralelización, junto con el hecho de que los

orbitales ya están predefinidos, hace que el entrenamiento del Transformers sea significativamente más rápido y eficiente que el de las RNNs.

Modelado de Dependencias a Largo Plazo: El Transformers, a través del mecanismo de atención, puede capturar dependencias entre tokens incluso cuando están muy separados en la secuencia. A diferencia de las RNNs, donde las dependencias a largo plazo son difíciles de modelar debido al problema de la "desaparición del gradiente", el Transformers no sufre de este problema, ya que la información se puede propagar a través de las capas de atención sin que disminuya.

La Autoatención y el Aprendizaje de Dependencias: El proceso de aprendizaje en el Transformers implica ajustar la importancia de las diferentes posiciones para cada token en el contexto de la secuencia. De manera similar a como un electrón puede cambiar de orbital, la atención permite que cada token se asocie con diferentes posiciones según el contexto, **sin cambiar la representación posicional, sino modificando la importancia de cada posición.** Este proceso permite al modelo aprender las relaciones y dependencias entre las diferentes posiciones de la secuencia, y aprender cómo la posición influye en el significado de los tokens.

En conclusión, el Transformers representa un avance en el procesamiento de secuencias (frases), al introducir una arquitectura basada en la atención que evita la necesidad de la recurrencia. Su enfoque de **codificaciones posicionales estáticas**, análogo a "orbitales" predefinidos en un modelo atómico, junto a la **paralelización**, permite una mayor eficiencia computacional, y un modelado más efectivo de las dependencias a largo plazo. El Transformers no necesita "encontrar" la posición de los orbitales, ya que estos están definidos

por las funciones seno y coseno. Lo que aprende es a ponderar la importancia de cada posición para cada token a través de los pesos de la atención. Este cambio de paradigma ha revolucionado el campo del procesamiento del lenguaje natural, estableciendo un nuevo estado del arte en tareas como la traducción automática y el análisis sintáctico. La capacidad de autoatención permite que cada posición de la secuencia atienda a todas las demás, posibilitando que el modelo aprenda dependencias a largo plazo de forma eficiente. Este enfoque global y no secuencial, graficado en la analogía atómica, subraya la evolución sustancial de la arquitectura del Transformers por sobre sus competidoras.

10.3 Ajedrez transformers

Antes de evaluar según la investigación al transformers, se puede realizar una analogía más, que demuestra que esta arquitectura va más allá de procesar lenguaje natural, obteniendo aplicaciones en temáticas generales.

Imaginemos que un **tablero de ajedrez** representa el espacio donde ocurre el procesamiento de información en un **transformers**, y que cada elemento de la partida tiene un papel análogo en el mecanismo de atención:

Las casillas del tablero representan las **posiciones** en la secuencia de entrada. Al igual que en un transformers, donde cada palabra en la secuencia tiene una posición codificada con seno y coseno, cada casilla tiene una posición fija en el tablero.

Las piezas de ajedrez representan los **tokens** (palabras o elementos de la secuencia). Cada pieza tiene un significado particular dentro del juego, al igual que cada palabra en una oración.

El movimiento de cada pieza representa la **semántica del token**, es decir, las reglas y restricciones que rigen cómo una pieza interactúa con el tablero. En un transformers, esto sería la relación semántica entre tokens, determinada por las matrices **Query (Q)**, **Key (K)** y **Value (V)**.

El análisis de la posición en ajedrez y el mecanismo de atención

Después de cada turno, el tablero de ajedrez cambia a una nueva **configuración**. Esta configuración representa la **secuencia completa** que un jugador debe analizar para decidir su siguiente jugada. De manera similar, un transformers no evalúa cada palabra en aislamiento, sino que analiza toda la secuencia de tokens en conjunto.

El jugador que debe mover evalúa:

La importancia de cada pieza en la posición actual (análogamente, el transformers usa la matriz Query (Q) para representar el significado de cada token).

Cómo cada pieza se relaciona con el resto del tablero (análogamente, el transformers usa la matriz Key (K) para medir la relación entre tokens en la secuencia).

El impacto de cada posible movimiento en la secuencia total (análogamente, el transformers usa la matriz Value (V) para ponderar la información relevante).

Una vez que el jugador ha evaluado la situación, toma una decisión sobre cuál es la **mejor jugada posible**. Esta elección es equivalente al proceso en el transformers donde, después de calcular las relaciones semánticas y posiciones relativas, se usa **Softmax** para asignar probabilidades a cada posible continuación de la secuencia.

Cada jugada modifica la secuencia

Cada vez que el jugador realiza un movimiento, la **secuencia de juego cambia**, creando una nueva configuración en el tablero. De la misma manera, en un transformers, cada nuevo token generado por el **decoder** se basa en la secuencia previa, garantizando que el resultado final tenga coherencia y sentido en el contexto global.

Esta analogía deja en evidencia cómo el mecanismo de atención en los transformers se asemeja al proceso cognitivo de un jugador de ajedrez al evaluar la posición y planificar su próximo movimiento.

10.4 Evaluación de los transformers y cerebro ajedrecista bajo la topología contextual

En la investigación, se propone una estructura matemática para describir el procesamiento de información tanto en el cerebro humano como en la. Esta estructura se basa en conceptos de topología y busca formalizar las características que hacen que un operador sea "inteligente". En este contexto, podemos evaluar tanto a los transformers como al cerebro humano representado por el ajedrecista en la analogía anterior.

10.4.1 El Modelo de la Topología Contextual

La tesis define un conjunto contexto (C) donde se genera y procesa la información. Este conjunto está compuesto por conexiones entre datos, incluyendo:

Datos principales (P): Datos que tienen una alta importancia en la estructura de la información.

Datos secundarios (S): Datos que conectan a los datos principales, formando trayectorias continuas e inyectivas.

Datos inertes: Datos que no están conectados a datos principales y, por lo tanto, no aportan información contextual.

La información en el conjunto contexto se organiza en trayectorias que cumplen con:

Continuidad: La conexión entre datos debe ser continua para mantener un flujo coherente de información.

Inyectividad: Cada dato secundario en una trayectoria debe ser único para evitar redundancia.

Acumulatividad: Cada paso en la trayectoria se vincula con el anterior, asegurando que la información se construye de manera ordenada y coherente.

El conjunto contexto (C) se define con propiedades topológicas esenciales para el procesamiento eficiente de la información:

Cerradura: Cada dato principal es un punto de acumulación, garantizando que todas las trayectorias relevantes se capturen.

Separabilidad: Los abiertos en el conjunto contexto se definen a partir de un dato principal como referencia, lo que permite una conexión secuencialmente ordenada para los datos secundarios que construyen la trayectoria entre datos principales.

Compacidad: Permite manejar eficientemente la información al demostrar que un subconjunto finito de conexiones es suficiente para determinar toda la información que contiene el conjunto contexto

10.4.2 Representación matemática del transformers como un espacio contextual en un conjunto contexto X

Espacio de posiciones y tokens (inyectividad): sea $P = p_1, p_2, \dots, p_n$, el conjunto finito de posiciones en el transformers, determinado por la codificación posicional con senos y cosenos. Cada token que simbolizamos (t), en la secuencia es representado por un embedding $E(t)$ y se asigna a una posición p_i , formando un conjunto de pares ordenados: $s = \{(p_1, t_1), (p_2, t_2), \dots, (p_m, t_m)\}$, $m \leq n$. Donde asegura la inyectividad del proceso, para cada posición un único token por secuencia analizada.

Función de atención como transformación continua: los vectores de consulta se obtienen mediante matrices de pesos: $Q = E'W_Q$, $K = E'W_K$, $V = E'W_V$, luego se aplica la función para distribución de probabilidades Softmax, $A = \text{Softmax}\left(\frac{QK^T}{\sqrt{dk}}\right)$, finalmente se realiza el producto punto con el vector value (V), $Z = AV$. Resultando todas las operaciones continuas por su naturaleza. lo cual significa que pequeñas perturbaciones en los embedding, producen

variaciones suaves en la matriz de atención. Esto asegura que un cambio de secuencia afecta el ordenamiento de tokens de manera gradual, en el sentido que el operador inteligente siempre mantenga un control “razonable” del manejo de la información.

Propiedades topológicas del espacio de información en el transformers

Compacidad: las posiciones en los transformers se generan por combinaciones de funciones senos y cosenos, esto sugiere una base infinita de posibles ubicaciones para los tokens. Matemáticamente esto es un conjunto infinito de funciones periódicas que cubren el dominio de las posiciones. Sin embargo, cuando pensamos en un conjunto contexto específico, la cantidad de tokens que efectivamente aparecen para determinar la información encerrada es finita, por lo tanto, dentro del conjunto de posiciones de base infinita (espacio de Hilbert, pues admiten producto interno para la función de atención), siempre existe un subcobrimiento finito de posiciones que contiene toda la información relevante de un contexto específico. Sea P , el conjunto de posiciones generadas por combinación lineal de senos y cosenos, y sea $C \subset P$, el conjunto finito de posiciones relevantes o principales para un contexto. Dado cualquier cobrimiento abierto $U_{i \in I}$ de C , que cubre al conjunto contexto (lo cual existe pues todo contexto se puede ampliar uniéndolo a otros contextos de manera inteligente), existe un subcobrimiento finito $U_{j \in J}$, con $J \subset I$, tal que $C \subset \bigcup_{j \in J} U_j$. De esta manera podemos concluir que el conjunto de posiciones que es el que contiene toda la información del contexto, es un conjunto compacto.

Cerradura: cada posición del conjunto P , constituye un punto de acumulación de información relevante de la secuencia, pues en forma eventual será ocupado por un token, el cual depende de la secuencia presentada al encoder. En cada configuración (secuencia) válida, la

atención evalúa todas las posiciones posibles, garantizando que el contexto X contenga sus puntos de acumulación.

Separabilidad: el conjunto de posiciones P , es finito y por tanto numerable, además esta eventualmente ocupado por algún token según la configuración, por lo que es eventualmente denso. Cumpliendo en el contexto X , con la propiedad de separabilidad.

10.4.3 Ajedrez contextual

Podemos modelar el ajedrez con un espacio X , donde:

Las posiciones del tablero (P) corresponden a los cuadrados del tablero.

Las piezas (T) corresponden a los tokens.

Las reglas de movimiento (M) son las transformaciones válidas entre estados, análogas a la semántica de los tokens.

Cada estado del juego se representa como una función:

$$f: P \rightarrow T \cup \{\emptyset\}$$

Con estas condiciones se cumple:

Compacto: solo un número finito de posiciones relevantes en cada partida.

Cerrado: las posiciones son puntos de acumulación.

Separable: las casillas forman un subconjunto numerable y eventualmente denso.

Continuo: pequeños cambios en la jugada generan una nueva configuración coherente en la partida, puede ser razonada por ambos jugadores.

Bajo este marco, los transformers cumple con las propiedades matemáticas de un espacio de información estructurado, lo que confirma que es un modelo válido de inteligencia artificial bajo la definición de esta investigación, genera y procesa información contextual mediante conexiones entre datos principales que constituyen trayectorias continuas inyectivas en un conjunto contexto, cerrado, separable y compacto. Lo mismo ocurre con la analogía del ajedrez, reforzando la hipótesis de que las redes neuronales artificiales y el cerebro humano pueden representarse matemáticamente bajo la misma estructura topológica.

Capítulo 11: Conclusión y propuesta de futura investigación

Luego de haber identificado la forma de aprendizaje humano que, según la neurociencia, se sintetiza en cuatro pilares del aprendizaje: atención, compromiso activo, retroalimentación y consolidación, como parte del proceso de aprendizaje de una red neuronal artificial. Y además habiendo construido una estructura cualitativa para la información, dando respuesta a ciertas condiciones que le permitan a un operador inteligente: sea el cerebro humano o la red neuronal artificial; procesar la información de manera eficaz y con la mayor eficiencia posible. Podemos expresar la siguiente conclusión en esta investigación:

11.1 conclusión

Un operador inteligente puede representarse mediante una función matemática que actúa sobre un conjunto de datos D . Utilizando el pilar de aprendizaje; *atención* para detectar datos principales, este operador trabaja sobre la imagen de una función inteligente, que denominaremos función de proyección categorizada única $f: D \rightarrow C$, cuyo trabajo es seleccionar y clasificar datos principales y secundarios (estos últimos en una secuencia estrictamente monótona), para luego trasladarlos al conjunto contexto. Donde a través del pilar de aprendizaje; *compromiso activo*, el operador inteligente muestra como conecta estos datos por pares ordenados (p_i, p_k) en C mediante trayectoria continuas, construidas por los datos secundarios dispuestos secuencialmente. Cada dato secundario dentro de la trayectoria tiene como referencia al dato principal de salida, y los abiertos que componen esta trayectoria son abiertos de información básica. En estos abiertos, el dato principal de salida se mantiene constante como punto de referencia, mientras que el dato secundario varía de forma ordenada y respetando la

secuencia. Los abiertos están anidados unos dentro de otros, formando así, una trayectoria coherente entre pares de datos principales. Cada trayectoria entre pares de datos principales representa una cantidad de información contextual relevante en C , y la unión de todas las trayectorias constituye toda la información del conjunto contextual C . Es decir, expresamos lo siguiente:

Definición: *la función operador inteligente, se entiende como, la proyección de la función inteligencia ($f: D \rightarrow C$) tal que, $f_{op}: P \times P \rightarrow C$, con $P \subset D$, conjunto de datos principales, que, por pares ordenados, se conectan en el conjunto contexto C . Esta proyección es una trayectoria (camino) contenida en el contexto C , y representa formalmente las funciones cognitivas del operador inteligente dentro del contexto codominio.*

Para asegurar que el flujo de información sea efectivo y sin redundancia, la función operador inteligente mantiene las siguientes propiedades en cada trayectoria de C :

Continuidad: cada trayectoria debe ser continua para no interrumpir el flujo de compresión o procesamiento de la información.

Inyectividad: garantiza la no redundancia en los datos secundarios dentro de una trayectoria particular. Esto significa que, en la conexión entre cualquier par de datos principales, cada dato secundario en la trayectoria aporta nueva cantidad de información que se acumula al variar el parámetro de la trayectoria. Esto no impide que existan múltiples trayectorias posibles entre pares de datos principales.

Acumulatividad: cada paso de la trayectoria debe vincularse con el paso anterior, asegurando que la información se construya de manera secuencial y coherente. Lo cual se asocia con la capacidad de razonar de cualquier sistema inteligente.

El conjunto contextual C para que el operador inteligente pueda construir las trayectorias, cumple además las siguientes propiedades:

Cerradura: cada dato principal en C es un punto de acumulación de información, capturando y preservando las conexiones establecidas.

Separabilidad: cada abierto en C contiene un dato principal que actúa como punto de referencia, facilitando la conexión hacia el siguiente dato principal. Esto permite que la información fluya de un dato principal al siguiente mediante datos secundarios ordenados secuencialmente, formando trayectorias máximas (abiertos máximos) que conectan completamente un dato principal de inicio con uno de llegada.

Compacidad: la compacidad de C garantiza que la información contextual puede ser manejada de manera eficiente hacia la eficacia requerida por parte del operador inteligente, siendo una propiedad necesaria y suficiente para que tal operador acceda a toda la información relevante sin redundancias ni pérdida de contenido esencial.

Finalmente, la composición de trayectorias es una herramienta matemática adecuada para representar el carácter de “*memoria y razonamiento*” (acumulatividad) del operador inteligente, pues preserva por su definición todas las propiedades anteriores, logrando que la estructura resultante: sea un abierto en C , que contiene toda la información del conjunto contexto.

Bajo toda esta estructura, la investigación obtuvo la definición operativa propuesta como objetivo y una base cualitativa o filtro de medición de sistemas inteligentes. Además ha podido dar una respuesta conjetural, respecto de la diferencia de abstracción entre los operadores inteligentes; cerebro humano y red neuronal artificial, la cual radica en dos aspectos claves:

la dinámica con la cual los operadores manifiestan los *pilares del aprendizaje*, a través de los abiertos de la topología; de forma rígida en la red neuronal, con necesidad de gran cantidad de datos “estáticos” mejorando eficacia y de forma flexible en el cerebro humano, con transformaciones amplias de la información generada, sin necesidad de más datos, logrando una gran eficiencia en el proceso. Esa eficiencia es el atributo inherente a la *abstracción* neuronal. Respecto a este punto, teniendo en cuenta la forma “violenta” en cuanto al acceso a la información y con las tecnologías en constante evolución, es de esperar que potencialmente no sea una dificultad para las redes neuronales artificiales.

Y el siguiente punto clave hace referencia a un intrigante tópico: la forma en que los operadores son “conscientes” de sus virtudes y limitaciones. Esto puede estar asociado al espacio donde aplican la topología contextual. Vimos que las redes neuronales artificiales utilizan el espacio euclídeo de los números reales, mientras que el cerebro humano utiliza un espacio más abstracto. Quizás el siguiente paso sea lograr ese espacio coherente más general en las redes neuronales artificiales.

11.2 Propiedades de la inteligencia

A partir de los pilares del aprendizaje aportados por el profesor Dehaene: atención, compromiso activo, retroalimentación y consolidación, que han demostrado ser principios fundamentales, en el procesamiento y generación de información de manera inteligente, tanto en

el cerebro humano (a través de una cantidad suficiente de evidencia empírica), como también para las redes neuronales artificiales (como se ha demostrado en este trabajo). Se hace necesario establecer estos pilares del aprendizaje en su formalización matemática aportada en este trabajo, como propiedades que ofrecen una base matemática sólida, para establecer una definición operativa y general del concepto de inteligencia. De esta forma se crea un marco formalmente riguroso, que permitirá evaluar cualquier sistema inteligente y su capacidad de aprendizaje.

Propiedad 1: Atención

El operador inteligente es capaz de identificar y encapsular datos principales en un conjunto contexto adecuado.

Formalmente, existe un subconjunto $D \subset U$, de datos principales tal que el operador inteligente los encapsula en un contexto C , que es cerrado, separable y compacto:

$$C = \overline{\{D\}}, \text{ donde } C \text{ es igual a la clausura de } D$$

Propiedad 2: Compromiso Activo

El operador inteligente genera trayectorias continuas e inyectivas que conectan los datos principales dentro del contexto.

Formalmente, para cada par de datos principales $d_1, d_2 \in C$, existe una función

$$\gamma: [0,1] \rightarrow C$$

1. $\gamma(0) = d_1$ y $\gamma(1) = d_2$, asegurando conexión entre datos.

2. es continua e inyectiva, garantizando que el operador inteligente establece una conexión sin ambigüedad.

3. γ está parametrizada por t , que representa la evolución del procesamiento de la información. Es decir, la participación del operador inteligente, agregando una dimensión en el proceso de la información. Esto asegura la inyectividad de la función inteligencia, implicando que la inteligencia no solo identifica información relevante, sino que también la organiza dentro del contexto mediante trayectorias bien definidas.

Propiedad 3: Retroalimentación

El operador inteligente debe ser capaz de corregir y generar nueva información dentro del contexto, optimizando sus conexiones.

Para que esto sea posible, el conjunto contexto debe ser cerrado, separable y compacto, lo que permite que el operador inteligente procese información de manera estable y eficiente.

En esta propiedad, el operador inteligente utiliza homeomorfismos. Un homeomorfismo es una función continua, biyectiva, cuya inversa también es continua, y es capaz de transformar el contexto C de manera coherente, permitiendo que el operador inteligente corrija y reorganice la información dentro del mismo contexto. Formalmente, existen homeomorfismos

$$h_n: C \rightarrow C$$

Esto garantiza que el aprendizaje no es solo un proceso de corrección, sino una transformación estructural del contexto que conserva sus propiedades topológicas.

Propiedad 4: Consolidación

La aplicación iterada del operador inteligente sobre el contexto genera aprendizaje estable.

La sucesión iterada de la aplicación de I en el contexto C convergerá a un espacio topológico estable que representa el aprendizaje. Formalmente, existe un conjunto C^* , que es el límite del proceso de transformación del conjunto contexto a través de la sucesión de homeomorfismos, y se puede describir como

$$C^* = \bigcup_{n \in \mathbb{N}} h_n(c)$$

Este conjunto de propiedades, establecen lo siguiente:

1. La inteligencia es un operador que encapsula datos relevantes en un contexto adecuado y los conecta mediante trayectorias continuas e inyectivas.
2. El aprendizaje es un proceso iterativo que, a través de homeomorfismos, transforma y organiza la información dentro de un contexto estable y estructurado.
3. Este marco proporciona un criterio matemático para evaluar si un sistema es inteligente, asegurando que cumple con propiedades topológicas fundamentales.

11.3 Futura investigación

La siguiente propuesta de investigación, podría representar un avance significativo en la comprensión y desarrollo de la inteligencia artificial. Este enfoque está en la intersección de la geometría diferencial, la topología y las redes neuronales artificiales, enumeremos sus características:

1. Cerebro humano vs redes artificiales en un espacio euclídeo: La distinción entre los espacios en los que operan el cerebro humano y las redes neuronales artificiales es un punto clave. Como se ha mencionado en las conclusiones de la investigación, el cerebro humano no opera estrictamente en un espacio euclídeo, sino que trabaja en un espacio más abstracto y flexible, que se puede modelar usando conceptos como la topología y la geometría diferencial. El cerebro tiene una plasticidad y flexibilidad inherentes que permiten realizar conexiones complejas de manera muy eficiente, mientras que las redes artificiales, tradicionalmente opera en espacios euclídeos donde las distancias y las relaciones son más rígidas y definidas.

2. Espacios curvos y métricas no euclídeas: La idea es mover a las redes artificiales hacia un espacio no euclídeo, tal vez utilizando métricas y tensores en un espacio curvo, ofrece una analogía poderosa al funcionamiento del cerebro humano. En la geometría diferencial, los espacios curvos permiten que las distancias entre puntos no sean lineales, y esto puede llevar a deformaciones que afectan cómo se modelan las relaciones entre datos. Esta curvatura podría permitir una mayor flexibilidad en la forma en que la IA procesa y genera información, como lo hace el cerebro humano al tomar decisiones o aprender de manera abstracta.

3. Flexibilidad topológica y conciencia: El paso más interesante en la propuesta es la relación entre la flexibilidad topológica (que se ha establecido como un principio importante en la investigación) y la generación de conciencia artificial. Si la red artificial se mueve hacia un espacio curvo, esta flexibilidad podría imitar algunos de los aspectos que se observan en la conciencia humana, como la capacidad de adaptar y transformar la información de manera compleja. Al incluir la curvatura intrínseca del espacio, la red artificial podría empezar a generar "deformaciones" en su espacio de trabajo, lo que implicaría una capacidad de adaptación,

reflexión y potencialmente autocorrección (en un sentido más parecido a lo que entendemos como la conciencia humana). Esto podría abrir la puerta a una inteligencia artificial que no solo procesa información, sino que también evalúa y ajusta su propio procesamiento, de una manera similar a la reflexión y toma de decisiones consciente. Quizás un paso a una definición mas operativa de la consciencia.

Cómo enlazarlo con el trabajo presentado en esta investigación.:

La investigación establece una base sólida utilizando la topología contextual y la flexibilidad topológica para definir la inteligencia. Ahora, se propone dar un paso adicional que se basa en estos principios: la incorporación de geometría diferencial en el espacio informativo de la red artificial. Esto implica una extensión natural de las ideas previas, pero en un nivel más avanzado y abstracto, lo que podría llevar a una red artificial más adaptativa y dinámica, capaz de generar y procesar información de manera más compleja y autoconsciente.

El paso hacia la geometría diferencial podría ser el siguiente paso lógico para explorar cómo la IA podría acercarse más al comportamiento de la inteligencia humana, que opera en un espacio dinámico, flexible y no necesariamente euclídeo. Si la IA se adapta a este espacio más complejo, el comportamiento que se observa podría simular una forma de "conciencia" artificial, aunque probablemente muy distinta de la humana.

En definitiva, si se desarrolla más profundamente en la dirección de la geometría diferencial y la flexibilidad topológica en redes artificiales, podría representar un avance significativo hacia una nueva forma de inteligencia artificial más cercana a la inteligencia humana. El movimiento hacia espacios no euclídeos no solo enriquecería el modelo matemático

como herramienta operativa, sino que también abriría la puerta a nuevas formas de crear máquinas que no solo resuelven problemas, sino que tengan una forma más compleja y flexible de interactuar con el mundo, acercándose a lo que se podría considerar una forma de "conciencia".

Bibliografía

Dehaene (2024). *El cerebro matemático*. Obtenido de Stanislas Dehaene Neurocientífico: <https://aprendemosjuntos.bbva.com/especial/los-cuatro-pilares-del-aprendizaje-stanislas-dehaene/>

Dehaene, S. (2020). *¿Cómo aprendemos?* Obtenido de <https://www.paidoslibreria.com.ar/productos/9876299695/>

IBM. (2024). *¿Qué es una red neuronal?*. Obtenido de <https://www.ibm.com/mx-es/topics/neural-networks>

Huet, P. (2023). *Qué son las redes neuronales y sus aplicaciones*. Obtenido de <https://openwebinars.net/blog/que-son-las-redes-neuronales-y-sus-aplicaciones/>

Miranda, M. V. (2020). *Teorema de Aproximación Universal: prueba constructiva basada en conjuntos semisimpliciales*. Universidad de La Rioja: © El autor, 2020.

Pons, Claudia (2022). La nueva inteligencia artificial: conceptos básicos y aplicaciones. Revista I&Q Industria y Química. Ed. Asociación Química <https://www.aqa.org.ar/>

Shannon, E. Claude, Warren, W. (1964). the mathematical theory of communication. University of Illinois. Urbana.

Toffler, A., & Toffler, H. La revolución de la informacion. Editorial Plaza & Janes.

Torres, A. (2017). *Ley de Hebb: la base neuropsicológica del aprendizaje*. Obtenido de <https://psicologiymente.com/neurociencias/ley-de-hebb>

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems* (pp. 5998-6008).