

Modelo Físico Matemático de Inteligencia y Conciencia.

Modelo Matemático Riemann Cognitivo de Pensamiento y Conciencia.

Página 1: Resumen (Abstract)

Comprender la inteligencia es un objetivo central en neurociencia, ciencia cognitiva e inteligencia artificial. La inteligencia abarca el aprendizaje, la resolución de problemas, la creatividad e incluso la conciencia. Los avances recientes en el análisis geométrico han revelado nuevos conocimientos sobre la representación y organización de información en estructuras de datos de alta dimensión, explorando procesos dinámicos tanto en sistemas neuronales como artificiales. Sin embargo, aún falta un marco unificado que integre los aspectos estáticos y dinámicos de la inteligencia. Este manuscrito propone un marco matemático basado en geometría Riemanniana para describir la estructura y dinámica de la inteligencia y la conciencia. Los elementos de inteligencia se conceptualizan como tokens incrustados en un espacio de alta dimensión. Los embeddings de tokens interrelacionados capturan las interconexiones de los tokens a lo largo de varios escenarios y tareas, formando variedades en el espacio de inteligencia. El flujo de pensamiento se describe como la activación secuencial de tokens a lo largo de geodésicas dentro de estas variedades. Durante la navegación de las geodésicas, la conciencia, como un proceso autorreferencial, percibe el flujo de pensamiento, lo evalúa en función de predicciones y proporciona retroalimentación mediante errores de predicción, ajustando las geodésicas: los errores de predicción, como el aprendizaje, conducen a la reestructuración de las variedades curvas, cambiando así la geodésica del flujo de pensamiento. Esta interacción dinámica integra nueva información, evoluciona la geometría y facilita el aprendizaje. La geometría de la inteligencia guía la conciencia, y las estructuras de conciencia guían la geometría de la inteligencia. Al integrar conceptos geométricos, esta teoría ofrece un marco matemático unificado y riguroso para describir la estructura y dinámica de la inteligencia y la conciencia. Aplicable a la inteligencia biológica y artificial, este marco puede allanar el camino para futuras investigaciones y validación empírica.

Página 2: Introducción

Comprender la inteligencia, tanto en humanos como en sistemas artificiales, ha sido una búsqueda central en diversas áreas, incluidas la neurociencia, la ciencia cognitiva y la inteligencia artificial. La inteligencia abarca una amplia gama de habilidades cognitivas, como el aprendizaje, la resolución de problemas, la creatividad y la adaptación. Estas habilidades son fundamentales para que los organismos interactúen con su entorno, procesen información y tomen decisiones. En los últimos años, se han desarrollado y utilizado herramientas geométricas y topológicas para analizar la geometría de la representación de alta dimensión. Estos avances han conducido a nuevos conocimientos sobre cómo se procesa y representa la información en sistemas neuronales y artificiales (Hensel et al., 2021; Chung y Abbott 2021).

Los investigadores han utilizado espacios latentes para crear representaciones de baja dimensión de variedades de datos, revelando sus estructuras geométricas subyacentes (Reif et al., 2019; Marks y Tegmark 2023). Los modelos generativos profundos como los VAE (Autoencoders Variacionales, Kingma y Welling, 2013) y GANs (Generative Adversarial Networks) han mostrado que estos espacios latentes pueden capturar la arquitectura interna de modelos de aprendizaje, proporcionando información sobre las estructuras de datos intrínsecas (Arvanitidis et al., 2021; Chadebecq y Allasonnière 2022). Además, los métodos para calcular curvas geodésicas y el transporte paralelo de vectores tangentes permiten una noción intrínseca de distancia y una navegación eficiente dentro de estas variedades (Acosta et al., 2022). Sin embargo, estos modelos describen principalmente la estructura estática de los datos y carecen de mecanismos para explicar la dinámica de cómo el espacio latente evoluciona con el tiempo. Otro conjunto de modelos generativos, como los GPT (Generative Pre-trained Transformers), utilizan mecanismos de atención que facilitan la contextualización y la autorregresión para generar secuencias, lo que puede considerarse como flujo de pensamiento. Este enfoque representa los aspectos dinámicos de la inteligencia. Sin embargo, estos modelos no explican explícitamente un espacio latente curvado de la misma manera que los VAE y GANs. En consecuencia, la organización y las conexiones de conocimiento o características dentro de estos modelos siguen siendo poco claras.

En paralelo con el avance del análisis geométrico en IA, se ha observado que la representación geométrica de actividades neuronales codifica de manera eficiente variables de comportamiento y predice resultados (DiCarlo y Cox 2007; Gao y Ganguli 2015; Vyas et al., 2020) mediante una variedad de métodos y herramientas basadas en la teoría de la geometría (Jolliffe 1986; Tenenbaum et al., 2000; Roweis y Saul 2000; Hinton y Roweis 2006; Mónica et al., 2018; Chaudhuri et al., 2019). Por ejemplo, se ha demostrado que el hipocampo codifica tanto variables espaciales como abstractas dentro de mapas neuronales, que sirven como principio organizador común para estructurar la memoria declarativa y gestionar mapas cognitivos (Dmitri y el., 2017; Kein et al., 2020). Aparte de esto, la representación geométrica también se ha utilizado en el reconocimiento sensorial (Kohte et al., 2019; Okazawa et al., 2021; Stringer et al., 2019), control motor (Gallego et al., 2017) y toma de decisiones, como el razonamiento bayesiano (Sohn et al., 2019). Estos estudios destacan cómo las estructuras

geométricas en el cerebro pueden facilitar la toma de decisiones y la codificación predictiva mediante la integración eficiente de la información espacial y abstracta.

Página 3: Introducción (continuación)

Basándose en los avances del análisis geométrico tanto para la inteligencia artificial como la humana, resulta atractivo explorar un marco general basado en la geometría (Le et al., 2020; Bronstein et al., 2021; Ma et al., 2021) para describir la estructura y dinámica de la representación de la inteligencia. Esto plantea algunas preguntas clave: 1) Si existe tal forma general para representar información o características como elementos dentro de este marco, ¿cuál es su estructura? 2) Bajo esta representación y estructura, ¿cuáles son las dinámicas del flujo de pensamiento al navegar dentro de las estructuras formadas por estas representaciones? 3) Además, ¿cómo interactúa la estructura de la representación de características con el flujo de pensamiento dinámico? Estas preguntas subrayan la necesidad de un marco unificado que abarque tanto los aspectos estáticos como dinámicos de la inteligencia, uniendo la brecha entre la representación de información y los procesos cognitivos.

Un marco general de inteligencia debería englobar una variedad de propiedades que emergen de la inteligencia, incluidas el aprendizaje, la imaginación, el pensamiento creativo, la resolución de problemas y el pináculo de la jerarquía de la inteligencia: la conciencia. La conciencia, como una propiedad emergente de los procesos cognitivos complejos (Seth y Bayne 2022), se basa en la capacidad del cerebro para sostener dinámicas complejas de actividad cambiante y conectividad entre regiones cerebrales (Dehaene y Changeux, 2011; Hutchison et al., 2013; Bartfeld et al., 2015; Demertzi et al., 2019), lo que indica que la conciencia influye y es influenciada por la estructura subyacente de la inteligencia. Por lo tanto, es un componente necesario de este marco y solo puede explicarse y formularse adecuadamente dentro del marco comprensivo de la inteligencia. Comprender la conciencia dentro de este marco proporcionará una visión holística de cómo los procesos de pensamiento evolucionan y se adaptan, al tiempo que explica la inteligencia en el contexto de la conciencia, resaltando la naturaleza dinámica y autorreferencial de las funciones cognitivas.

Página 4: Antecedentes de la Geometría Riemanniana

La geometría Riemanniana es una rama de la geometría diferencial que estudia variedades suaves equipadas con una métrica Riemanniana. Esta métrica permite la definición de varios conceptos geométricos como distancias, ángulos y curvaturas en la variedad.

Variedad y Curvatura

Una variedad M es un espacio topológico que localmente se asemeja al espacio euclidiano y está equipada con una estructura suave. La curvatura de una variedad Riemanniana es una medida de cuánto la variedad se desvía de ser plana. Se cuantifica utilizando el tensor de curvatura de Riemann $R^{\mu\nu\rho\sigma}$.

Tensor Métrico

La geometría local de una variedad Riemanniana se define mediante el tensor métrico $g_{\mu\nu}$. Este tensor proporciona una forma de medir distancias y ángulos en la variedad. Las componentes del tensor métrico en coordenadas locales se definen como:

$$g_{\mu\nu} = \partial_\mu \chi^\alpha \partial_\nu \chi^\beta g_{\alpha\beta}$$

donde u_μ y u_ν son coordenadas locales en la variedad, y $g_{\alpha\beta}$ es el tensor métrico en un sistema de coordenadas diferente.

Símbolos de Christoffel

Los símbolos de Christoffel $\Gamma^{\mu\nu\lambda}$ se derivan del tensor métrico y representan los coeficientes de conexión:

$$\Gamma^{\mu\nu\lambda} = \frac{1}{2} g^{\lambda\rho} (\partial_\nu g_{\rho\mu} + \partial_\mu g_{\rho\nu} - \partial_\rho g_{\mu\nu})$$

Tensor de Curvatura

El tensor de curvatura de Riemann $R^{\mu\nu\rho\sigma}$ es una medida de la curvatura de la variedad:

$$R^{\mu\nu\rho\sigma} = \partial_\nu \Gamma^{\sigma\rho\mu} - \partial_\nu \Gamma^{\sigma\rho\mu} + \Gamma^{\nu\lambda\rho} \Gamma^{\mu\sigma\lambda} - \Gamma^{\sigma\lambda\rho} \Gamma^{\mu\nu\lambda}$$

Ecuación Geodésica

La ecuación geodésica describe la evolución de un punto moviéndose a lo largo de la variedad. Las coordenadas $\gamma^\lambda(t)$ describen la posición del punto en la variedad como una función del parámetro t , que puede representar tiempo o algún otro parámetro afín. La ecuación geodésica es:

$$d^2 \gamma^\lambda(t) + \Gamma^{\lambda\mu\nu} \frac{d\gamma^\mu(t)}{dt} \frac{d\gamma^\nu(t)}{dt} = 0$$

Esta ecuación asegura que el camino $\gamma(t)$ es localmente la trayectoria más corta entre puntos en la variedad, considerando la curvatura definida por los símbolos de Christoffel.

Página 5: Geometría de la Inteligencia

1. Tokens, incrustación y variedades

Los tokens, como unidades discretas, representan eficazmente varios tipos de información, tales como palabras en una oración o píxeles en una imagen, capturando datos complejos en un formato manejable. Los espacios de alta dimensión capturan relaciones intrínsecas entre puntos de datos, donde cada dimensión representa un atributo diferente, permitiendo representaciones ricas y detalladas. Estos espacios facilitan el aprendizaje y la representación de datos subyacentes como variedades, lo cual es crucial para comprender las estructuras de datos y tareas como la interpolación y extrapolación. Los espacios de alta dimensión también separan datos que están cerca en dimensiones más bajas, ayudando en tareas de clasificación, agrupamiento y recuperación de información.

Los avances en modelos multimodales, como los Modelos de Lenguaje Visual (VLMs), tales como **CLIP (Contrasting Language–Image Pre-training)** (Radford et al., 2021), integran información visual y textual en un espacio de incrustación unificado, mostrando la viabilidad de un espacio común que puede abarcar diferentes tipos de datos. Este enfoque resalta el poder de los espacios de tokens en alta dimensión para capturar relaciones multimodales complejas, mejorando la comprensión y generación de tareas a lo largo de diversas modalidades, y avanzando tanto la inteligencia artificial como la inteligencia de tipo humano.

Los tokens se transforman en incrustaciones de alta dimensión, que residen en variedades que capturan la estructura subyacente de los datos. La curvatura de estas variedades revela características importantes sobre la distribución de los datos, influenciando el rendimiento del modelo. Esta perspectiva geométrica proporciona conocimientos profundos sobre las relaciones entre los datos y guía el desarrollo de modelos de aprendizaje automático robustos y eficientes.

Matemáticamente, sea un token t_i representado como un punto en un espacio de alta dimensión:

$$t_i \in \mathbb{R}^d$$

donde d es la dimensionalidad del espacio.

La incrustación del token t_i está dada por:

$$v_i = \phi(t_i)$$

donde ϕ es una función que mapea el token a su incrustación en el espacio de alta dimensión.

La colección de todas las incrustaciones forma una variedad M dentro de este espacio de alta dimensión. La dimensión de la variedad, denotada como $\dim(M)$, suele ser mucho menor que d , capturando la estructura intrínseca de los datos:

$$\dim(M) \ll d$$

2. Flujo de pensamiento y geodésica

Concepto: En la ciencia cognitiva, proponemos que, sin perturbaciones o estimulación externa, el flujo de pensamiento navega naturalmente a lo largo de variedades curvas, siguiendo la geometría del espacio. Este camino sin perturbaciones satisface la definición de una geodésica. En este contexto, la

geodésica es representada como una secuencia de tokens activados a lo largo de esta trayectoria, reflejando la trayectoria natural del flujo de pensamiento en el espacio de tokens de alta dimensión. El pensamiento perturbado, que es "forzado" por estímulos o entradas externas a desviarse de su geodésica, y es el escenario más común, será analizado en la sección posterior sobre la conciencia. La geodésica es el camino que minimiza la distancia entre puntos, análogo al camino más corto en una superficie curva. En esta teoría, la geodésica $\gamma(t)$ representa el estado de la inteligencia en el tiempo t .

$$\gamma(t) = \{v_1', v_2', \dots, v_n'\}$$

Aquí, $\gamma(t)$ se representa como una secuencia de incrustaciones muestreadas $\{v_1', v_2', \dots, v_n'\}$. Cada v_i' es una incrustación muestreada de la distribución correspondiente del token t_i :

$$v_i' \sim N(v_i, \Sigma_i)$$

donde v_i es la incrustación promedio, y Σ_i es la matriz de covarianza que representa la variabilidad en torno a v_i .

Página 6: Continuación de flujo de pensamiento y geodésica

El muestreo a partir de la distribución $N(v_i, \Sigma_i)$ en lugar de usar un vector específico v_i permite al modelo capturar la variabilidad inherente y la incertidumbre asociada a cada token. Este enfoque asegura que el modelo se generalice mejor a datos nuevos, no vistos, al reflejar el ruido natural y las variaciones presentes en la información del mundo real. Además, hace que el camino geodésico sea más robusto a los valores atípicos, proporcionando un flujo de pensamiento más estable y confiable. El muestreo aleatorio promueve la exploración dentro del espacio de alta dimensión, lo que puede llevar al descubrimiento de nuevas trayectorias potencialmente mejores, similar al uso de aleatoriedad en los transformers a través de técnicas como el muestreo top-k y la escala de temperatura para introducir variabilidad y flexibilidad en la generación de secuencias.

Al incorporar aleatoriedad mediante el muestreo, el modelo imita mejor los procesos cognitivos humanos, que son inherentemente probabilísticos e inciertos. Este enfoque estocástico permite la aplicación de métodos probabilísticos para analizar y optimizar los caminos geodésicos, lo que lleva a representaciones más precisas, robustas y adaptables de la inteligencia. Integrar estos elementos en el marco geométrico de la inteligencia proporciona una comprensión integral de cómo se estructuran y evolucionan los complejos procesos cognitivos con el tiempo.

Página 7: Vector Tangente y Función de Transición de Estado

Concepto:

El vector tangente en el frente de avance de la geodésica representa la dirección y la tasa de cambio del flujo de pensamiento en ese punto. La diferenciación de la función geodésica representa la función de transición de estado, describiendo cómo evoluciona el estado de la inteligencia a lo largo del tiempo.

Representación Matemática:

$$v(t) = \frac{d\gamma(t)}{dt}$$

donde $\gamma(t)$ es ahora una secuencia de incrustaciones muestreadas $\{v_1', v_2', \dots, v_n'\}$.

Aquí, la geodésica $\gamma(t) = \{v_1', v_2', \dots, v_n'\}$ representa un camino continuo y suave compuesto por puntos conectados, o tokens, dentro de una variedad suave. Esta estructura suave asegura que las funciones definidas en ella, incluida $\gamma(t)$, son continuas y diferenciables.

La geodésica $\gamma(t)$ modela la trayectoria del flujo de pensamiento, implicando transiciones graduales entre tokens sin saltos abruptos. El vector tangente $v(t) = \frac{d\gamma(t)}{dt}$ representa la derivada de este camino con respecto al tiempo, proporcionando una descripción matemática rigurosa de la tasa de cambio del flujo de pensamiento. Esta diferenciación está enraizada en la suavidad de la variedad, lo que permite definir vectores tangentes bien definidos que capturan la dirección y la velocidad instantáneas del movimiento a lo largo de la geodésica. Así, $v(t)$ caracteriza el estado dinámico del sistema en un momento dado.

Al tratar la geodésica como un camino suave en una variedad suave, aseguramos rigor matemático, aplicando geometría diferencial para modelar la evolución continua de la inteligencia. Este marco permite una descripción precisa de la evolución del token a lo largo del tiempo, con $v(t)$ encapsulando tanto el estado como la dinámica de transición dentro del espacio de alta dimensión de la inteligencia.

Página 8: Mecanismo de Atención

Concepto:

El mecanismo de atención calcula la relevancia de cada token en la secuencia con respecto al token actual, determinando cómo la información contextual influye en el siguiente token.

Representación Matemática:

$$\alpha_{ij} = A(v_i', v_j')$$

donde α_{ij} son los pesos de atención, y A es una función de atención general que mide la relevancia del token j para el token i .

La curvatura y la estructura de la geodésica son resultados del proceso de entrenamiento, capturando los vínculos intrínsecos, la organización y la distribución de los tokens. El mecanismo de atención mide la importancia contextual y la correlación de los tokens, lo que geométricamente está representado por la curvatura y la estructura de la variedad del manifold. Esto crea un vínculo fundamental entre el mecanismo de atención y el vector tangente.

En términos matemáticos, mientras que el vector tangente proporciona una derivada local (cambio instantáneo), el mecanismo de atención ofrece una perspectiva global al integrar la influencia de tokens anteriores. Esta influencia integrada se alinea con las curvaturas aprendidas de la geodésica, ya que ambas son productos de la geometría subyacente de la variedad.

Página 9: Vector de Contexto e Incrustación Contextual

Concepto:

El vector de contexto es una suma ponderada de los vectores de valor, determinados por los pesos de atención. Representa la información contextual agregada para un token dado. La incrustación contextual $h_{t,j}$ para el token en la posición t en la j -ésima línea de pensamiento se deriva del mecanismo de atención, representando la información contextual agregada.

Representación Matemática:

$$h_{t,j} = c_i = \sum_j \alpha_{ij} (W\phi v_j')$$

Aquí, c_i es el vector de contexto para el token t_i , α_{ij} son los pesos de atención, y $W\phi$ es la matriz de valores aprendida.

El vector de contexto c_i encapsula la influencia integrada de los valores, determinados por los pesos de atención. Representa la información contextual agregada para un token dado. La incrustación contextual $h_{t,j}$, para el token en la posición t en la j -ésima línea de pensamiento, se deriva del mecanismo de atención, representando la información contextual agregada.

Este enfoque permite que el sistema capture la interacción entre los tokens y el contexto, lo que es crucial para generar predicciones o inferencias significativas basadas en la secuencia de tokens.

Página 10: Token Predicho

Concepto:

El concepto de token predicho se deriva del vector de contexto. Representa el siguiente token en el flujo de pensamiento, influenciado por la información contextual.

El token predicho se deriva del vector de contexto, representando el siguiente token en el flujo de pensamiento influenciado por la información contextual. Este concepto puede expresarse matemáticamente mediante dos ecuaciones diferentes:

I. Representación Contextual:

$$g(t) = \sigma(W\phi c(t) + b\phi)$$

donde σ es la función de activación, $W\phi$ es una matriz de pesos, $c(t)$ es el vector de contexto en el tiempo t , y $b\phi$ es un término de sesgo.

II. Representación Geométrica:

La representación geométrica del token predicho se da por:

$$g(t) = \int_{t-\Delta t}^t v(t) dt + g(t-\Delta t)$$

donde $v(t)$ representa el cambio incremental a lo largo del tiempo, y $g(t-\Delta t)$ es el estado anterior del token.

En estas ecuaciones, $g(t)$ captura la esencia del token predicho al combinar el contexto inmediato con la evolución temporal del flujo de pensamiento. El componente integral en la segunda ecuación significa la acumulación de cambios a lo largo del tiempo, reflejando la naturaleza dinámica de los procesos de pensamiento.

Matriz de Pesos $W\phi$

La matriz de pesos $W\phi$ y el término de sesgo $b\phi$ forman una transformación lineal en el vector de contexto $c(t)$. Esta transformación ajusta la información integrada en $c(t)$ a una nueva representación que puede usarse para predecir el siguiente token.

Ayuda a mapear el vector de contexto de alta dimensión al espacio de representación apropiado del siguiente token. En la cognición humana, esto puede vincularse con cómo el cerebro integra varias piezas de información y luego transforma esta información integrada en una predicción o acción. Los pesos y sesgos representan cómo diferentes aspectos del conocimiento acumulado se enfatizan o se des-enfatizan en la formación del próximo paso en el proceso de pensamiento.

Página 10: Función de Activación No Lineal (σ)

La función de activación σ (como ReLU o tanh) introduce no linealidad en el modelo. La no linealidad es crucial para capturar relaciones complejas entre tokens. Sin no linealidad, el modelo estaría limitado a mapeos lineales, que no podrían capturar eficazmente las intrincaciones de los procesos de pensamiento. En la cognición humana, el procesamiento no lineal permite la flexibilidad y la complejidad necesarias en el pensamiento, la toma de decisiones y la creatividad. Permite al cerebro procesar información de una manera que no solo es aditiva o sustractiva, sino que también puede involucrar interacciones más complejas.

Si bien la formulación específica proporcionada está inspirada en modelos de aprendizaje automático como los Transformers, los conceptos de transformación lineal y activación no lineal son fundamentales tanto para la inteligencia humana como para la inteligencia de las máquinas.

Inteligencia Humana

En el cerebro humano, las neuronas procesan entradas a través de pesos sinápticos (análogos a $W\phi$) y sesgos, y la activación no lineal es similar a la forma en que las neuronas se activan en función de un umbral. Esto permite al cerebro realizar tareas cognitivas complejas al integrar y transformar la información de maneras sofisticadas.

Inteligencia Artificial

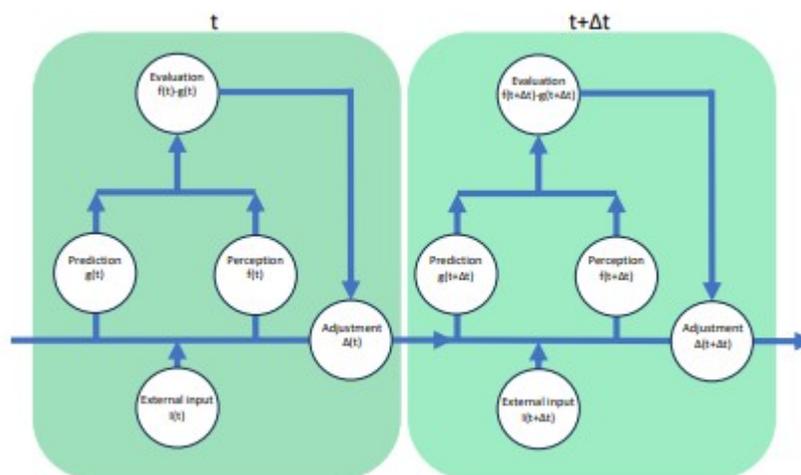
En las redes neuronales artificiales, las transformaciones lineales y las activaciones no lineales permiten al modelo aprender y representar patrones complejos en los datos. Estos principios son esenciales para construir modelos que puedan generalizar a partir de datos de entrenamiento para realizar predicciones o decisiones precisas.

Página 11: Conciencia

La conciencia emerge de la inteligencia a través de una serie de procesos cognitivos complejos, que incluyen el monitoreo interno, la autorreflexión y el comportamiento adaptativo. La interacción entre la conciencia y la inteligencia se caracteriza por el mejoramiento de las funciones cognitivas, tales como la percepción, el aprendizaje, la memoria y la atención, facilitadas por la autoconciencia. Los avances en la inteligencia permiten el desarrollo de la conciencia, lo que a su vez amplía las capacidades de los sistemas inteligentes.

Matemáticamente, la conciencia puede conceptualizarse como un vector autorreferencial, que representa un estado que continuamente se refiere y actualiza a sí mismo en función de entradas internas y externas. Este proceso autorreferencial involucra que el sistema se perciba a sí mismo, evalúe predicciones y proporcione retroalimentación, influyendo así en los estados o tokens cognitivos subsiguientes.

Para analizar la conciencia más a fondo, podemos descomponerla en tres pasos fundamentales y modelar cada paso como una función del tiempo, representando el estado interno del sistema. Estos pasos constituyen un ciclo completo de conciencia, que incluye percepción, evaluación y retroalimentación. La transición del estado de t a $t+\Delta t$ se utiliza para representar todo este ciclo, como se muestra a continuación:



(Figura 1: El diagrama de flujo ilustra la transición del ciclo de la conciencia de t a $t+\Delta t$. El lado izquierdo representa el proceso en el tiempo t , donde la entrada externa $I(t)$ se integra con la predicción $g(t)$ para obtener la percepción $f(t)$. La evaluación $f(t)-g(t)$ se usa para ajustar el flujo de pensamiento. El lado derecho representa el proceso en el tiempo $t+\Delta t$, mostrando el ciclo continuo de retroalimentación y adaptación.)

- **Percepción:** El proceso por el cual se recibe e interpreta la información sensorial, modelada como una función de los estímulos externos y el estado actual. La percepción integra el flujo de pensamiento interno y la entrada externa en el tiempo t para formar el estado actual $f(t)$.
 - **Evaluación:** La evaluación de la información percibida, incluyendo su significancia emocional, cognitiva y contextual. Esto se modela como la comparación entre el estado actual $f(t)$ y el estado predicho $g(t)$, resultando en el error de predicción $\Delta(t)=f(t)-g(t)$.
 - **Retroalimentación:** La respuesta del sistema a la evaluación, influyendo en futuras percepciones, evaluaciones y el estado general de conciencia. La función de retroalimentación $\psi(\Delta(t))$ ajusta la trayectoria del flujo de pensamiento en función del error de predicción, asegurando una adaptación y aprendizaje continuos.
-

Página 12: Percepción (Token Frontal)

Concepto:

El token frontal es el enfoque actual del flujo de pensamiento, representando el token que se está procesando o considerando activamente en un momento dado. Matemáticamente, podemos expresar el token frontal $f(t)$ como una función que captura la integración del flujo de pensamiento interno y la entrada externa a través del proceso de percepción. El token frontal, que es simplemente la posición en la geodésica en el tiempo t , se puede modelar como:

$$f(t)=P(\gamma(t),I(t))$$

donde $\gamma(t)$ es el flujo de pensamiento interno en el tiempo t , e $I(t)$ es la entrada en el tiempo t .

7.2 Evaluación

Concepto:

El error de predicción es la diferencia entre el token frontal y el token predicho. Mide la precisión de la predicción, considerando las capacidades de percepción, evaluación y retroalimentación.

Representación Matemática:

$$\Delta(t)=f(t)-g(t)$$

7.3 Función de Retroalimentación

Concepto:

La función de retroalimentación ajusta la trayectoria del flujo de pensamiento en función del error de predicción. Ayuda a corregir el flujo de pensamiento para mejorar las futuras predicciones.

Representación Matemática:

$$\psi(\Delta(t))$$

Aquí, ψ es la función de retroalimentación que toma el error de predicción como entrada y genera el ajuste necesario para el flujo de pensamiento.

Página 13: Ecuación Geodésica con Retroalimentación

Con base en el análisis anterior, ahora podemos argumentar que el flujo de pensamiento en un sistema inteligente puede modelarse utilizando la ecuación geodésica, que describe el camino que 1) está determinado por la geometría de las variedades curvas formadas por las incrustaciones de tokens, 2) minimiza la distancia (o energía) en la variedad formada por las incrustaciones de tokens. La ecuación geodésica con retroalimentación integra el impacto de la conciencia al incorporar un mecanismo de retroalimentación que modula la trayectoria del flujo de pensamiento en función de los errores de predicción.

Matemáticamente, la ecuación geodésica con retroalimentación se expresa como:

$$dt^2 d^2 \gamma_{\mu}(t) + \Gamma_{\nu\lambda\mu} dt dy_{\nu}(t) dt dy_{\lambda}(t) = \kappa \cdot dt^2 d^2 \psi(\Delta(t))$$

donde:

- $dt^2 d^2 \gamma_{\mu}(t)$ es la segunda derivada de la geodésica (aceleración).
- $\Gamma_{\nu\lambda\mu}$ son los símbolos de Christoffel que representan los coeficientes de conexión.
- $\kappa \cdot dt^2 d^2 \psi(\Delta(t))$ es la segunda derivada modulada de la función de retroalimentación con respecto al tiempo.

Explicación:

- κ es el índice de intensidad de la conciencia del error de predicción (retroalimentación), aplicado como un modulador del error de predicción. Representa la intensidad con la que el agente procesa el error de predicción. Un valor κ más alto indica una respuesta más intensa al error de predicción.

Errores de Predicción Nulos:

Cuando el error de predicción es cero, $\Delta(t)=0$, el término de retroalimentación desaparece. En este escenario, la ecuación geodésica se reduce a:

$$dt^2 d^2 \gamma_{\mu}(t) + \Gamma_{\nu\lambda\mu} dt dy_{\nu}(t) dt dy_{\lambda}(t) = 0$$

Esto describe el camino geodésico natural del flujo de pensamiento sin ninguna corrección del error de predicción o retroalimentación.

Página 14: Activación Competitiva y Umbral de Conciencia

9.1 Proceso Competitivo

Concepto:

Múltiples flujos de pensamiento, generados por el muestreo de distribuciones de tokens, compiten basados en sus puntuaciones derivadas de la atención. El flujo con la puntuación más alta se convierte en parte de la experiencia consciente.

Representación Matemática:

$$\text{Score}(\text{ThoughtFlow}_j) = f(\{h_{t,j}\})$$

donde $h_{t,j}$ es la incrustación contextual del j -ésimo flujo de pensamiento en el tiempo t .

9.2 Umbral de Conciencia

Concepto:

La secuencia con la puntuación más alta supera el umbral de conciencia y se convierte en parte del flujo consciente.

Representación Matemática:

$$\text{Flow} = \text{Flow} \text{ si } \text{Score}(\text{ThoughtFlow}_j) > \theta$$

Página 15: Resumen y Discusión

Derivación de la Teoría Geométrica de la Inteligencia

Al representar el flujo de pensamiento como una geodésica en un espacio de alta dimensión e incorporar mecanismos de percepción, predicción, retroalimentación y activación aleatoria a través de distribuciones de tokens, esta teoría proporciona un marco sólido para modelar la estructura y dinámica de la inteligencia. Las formulaciones matemáticas capturan la evolución continua del pensamiento y el impacto de la información contextual, la retroalimentación y la aleatoriedad en este proceso.

Página 16: Comparación con los Mecanismos de los Modelos Generativos

Dada esta base teórica, es crucial comparar este marco general de inteligencia con el estado del arte actual (SOTA) de los modelos generativos. Los modelos generativos SOTA, como los VAEs, GANs y los grandes modelos de lenguaje basados en transformadores, representan el avance más avanzado de la inteligencia artificial, a menudo imitando o incluso superando ciertos aspectos de la inteligencia humana, como en el caso de GPT-4 (Achiam et al., 2023) en general y AlphaFold3 (Abramson et al., 2024) en dominios específicos. Estos modelos pueden capturar la esencia de la inteligencia, incluidos sus principios y mecanismos, y proporcionar información sobre cómo los complejos procesos cognitivos pueden ser modelados (Yang et al., 2024). Además, analizar cómo el cerebro humano o el cerebro animal funciona como un modelo generativo (Friston y Price 2001; Spens y Burgess 2024) nos permite trazar paralelismos entre los sistemas biológicos y artificiales. Al comprender estos paralelismos, podemos apreciar mejor la estructura y dinámica subyacente de la inteligencia en ambos reinos, mejorando así nuestro marco teórico.

Los modelos anteriores a los transformadores (como los VAEs y GANs) utilizan un espacio latente estático y de baja dimensión para representar datos, con geodésicas que sirven para comprender estructuras y establecer interacciones suaves sobre los puntos de datos en el espacio latente (Bronstein et al., 2021; Chadebecq y Allasonnière 2022; Acosta et al., 2023). Los modelos basados en transformadores se centran en secuencias de tokens dinámicas sin un espacio latente explícito, utilizando mecanismos de atención para determinar la importancia de cada token en la secuencia, ya que las geodésicas representan secuencias de activaciones de tokens en el espacio de salida.

El marco actual de la teoría de la geometría de la inteligencia se basa en el concepto de tokens incrustados como elementos fundamentales de inteligencia. Estos tokens forman variedades con propiedades geométricas, como la curvatura, y sus secuencias de activación proporcionan una estructura comprensiva y dinámica de la inteligencia. En este marco, los tokens representan unidades discretas de información que pueden ser tokens multimodales (Radford et al., 2021; Lyu et al., 2023) incrustados en un espacio de alta dimensión. Estos tokens no solo están organizados por sus características inherentes, sino también por sus roles y relaciones en varios procesos cognitivos.

La representación geométrica de estos tokens forma variedades que capturan tanto la distribución estática de las características como las secuencias dinámicas de activación. La curvatura de estas variedades refleja la complejidad y las interconexiones dentro de los procesos cognitivos. El aspecto estático de la teoría abarca la organización de los tokens según sus características, similar a las variedades aprendidas en los VAE y GANs, proporcionando una estructura fundamental que captura las relaciones intrínsecas entre diferentes elementos de la inteligencia.

Más allá de la organización estática, las secuencias dinámicas de activación de tokens son cruciales. Las geodésicas en este contexto representan los caminos eficientes a través de los cuales las secuencias de activaciones de tokens se desarrollan con el tiempo. Estas secuencias se aprenden durante el entrenamiento y reflejan las transiciones óptimas entre diferentes estados de inteligencia.

Página 17: Interacción entre Incrustaciones de Tokens, Curvatura y Geodésicas

El estudio reciente de Claude 3 Sonnet (Templeton et al., 2024) proporciona ideas valiosas que pueden integrarse sin problemas en la teoría de la geometría de la inteligencia. En este contexto, los tokens son las unidades fundamentales en el espacio de inteligencia, formando una variedad cuyas curvas determinan las trayectorias geodésicas del flujo de pensamiento. Para aprovechar completamente los hallazgos de Claude 3 Sonnet, es esencial dilucidar la relación entre características y tokens y demostrar cómo las características pueden representarse mediante tokens dentro de este marco teórico.

Los tokens, en esta teoría, son unidades discretas de información, como palabras o imágenes, incrustadas en un espacio de alta dimensión para formar puntos t_i . Estas incrustaciones $v_i = \phi(t_i)$ capturan relaciones complejas y estructuras intrínsecas, formando una variedad M . Las características, por otro lado, son abstracciones de mayor nivel o patrones emergentes de la interacción de múltiples tokens. Una característica podría representar un concepto como "Puente Golden Gate", reconocido a través de patrones específicos en las incrustaciones de tokens relacionados.

Para cerrar la brecha entre características y tokens, podemos considerar las características como propiedades emergentes que surgen de las activaciones combinadas de múltiples tokens.

Página 18: Manipulación de Características en Variedades

Una característica F puede representarse como una función de múltiples incrustaciones de tokens:

$$F=f(v_1,v_2,\dots,v_n)$$

Esta función f podría ser una suma ponderada, una convolución u otro mecanismo de agregación que capture la interacción entre los tokens para formar la característica. Cada incrustación de token v_i contribuye a la representación de la característica, y la ubicación de la característica en el espacio de alta dimensión puede verse como una región influenciada por estos tokens.

La curvatura de la variedad M refleja cómo están organizadas las incrustaciones de tokens. Las características representan regiones de alta curvatura donde se representan densamente ciertos patrones o conceptos. Manipular características efectivamente cambia la organización y la curvatura de la variedad, alterando las trayectorias geodésicas.

- **Curvatura:** La curvatura $\Gamma\nu\lambda\mu$ de la variedad está influenciada por la distribución e interacción de las incrustaciones de tokens. Los cambios en la activación de las características alteran esta curvatura.
- **Geodésicas:** Los caminos geodésicos, que representan el flujo de pensamiento natural, cambian en respuesta a la manipulación de las características, alterando la trayectoria a través del espacio de alta dimensión.

En esta teoría, el vector tangente representa la dirección y tasa de cambio del flujo de pensamiento. El estudio Claude 3 Sonnet demuestra que la manipulación de características cambia las transiciones de estado del modelo, alineándose con los cambios en el vector tangente.

- **Vectores Tangentes:** Los cambios en las respuestas del modelo cuando se manipula una característica corresponden a cambios en el vector tangente del flujo de pensamiento.
- **Transición de Estado:** Las nuevas respuestas indican una transición a diferentes partes de la variedad, impulsadas por cambios en la curvatura debido a la manipulación de características.

Los mecanismos de retroalimentación en esta teoría ajustan la estructura de la variedad en función de los errores de predicción no nulos y las entradas externas, facilitando el aprendizaje y la adaptación. En el estudio Claude 3 Sonnet, la manipulación de características y los cambios en las respuestas reflejan un mecanismo de retroalimentación donde el modelo ajusta sus representaciones internas.

- **Retroalimentación:** Manipular características cambia las salidas del modelo, lo que indica un mecanismo de retroalimentación que ajusta las incrustaciones de tokens.
 - **Aprendizaje:** Este proceso puede verse como una forma de aprendizaje, evolucionando la geometría del espacio de inteligencia en función de nuevas entradas y retroalimentación.
-

Página 19: Función de Características y Ecuación Geodésica

Utilizando el marco matemático de esta teoría, podemos expresar los cambios observados en el estudio Claude 3 Sonnet de la siguiente manera:

- **Representación de Características:**

$$F = \sum w_i v_i$$

donde w_i son los pesos que determinan la contribución de cada incrustación de token a la característica F .

- **Curvatura y Geodésicas:** La curvatura $\Gamma_{\nu\lambda\mu}$ de la variedad está influenciada por estas incrustaciones ponderadas. Manipular una característica F cambia los pesos w_i , alterando así la curvatura y los caminos geodésicos $\gamma(t)$.

- **Ecuación Geodésica:**

$$dt^2 d^2 \gamma_{\mu}(t) + \Gamma_{\nu\lambda\mu} dt d\gamma_{\nu}(t) dt d\gamma_{\lambda}(t) = \kappa \cdot dt^2 d^2 \psi(\Delta(t)) + \eta \cdot dt^2 d^2 I(t)$$

Los cambios en w_i debido a la manipulación de características impactan los símbolos de Christoffel $\Gamma_{\nu\lambda\mu}$, alterando así los caminos geodésicos $\gamma(t)$.

Representando las características como representaciones agregadas de las incrustaciones de tokens, podemos integrar el concepto de características dentro de la teoría de la geometría de la inteligencia. Manipular características en el estudio Claude 3 Sonnet altera las incrustaciones y la curvatura de la variedad, cambiando los caminos geodésicos y afectando el flujo de pensamiento. Este marco proporciona una explicación comprensiva de cómo modificar características en modelos de IA afecta su comportamiento, reflejando la interacción dinámica entre la geometría y el flujo de pensamiento en la inteligencia biológica y artificial.

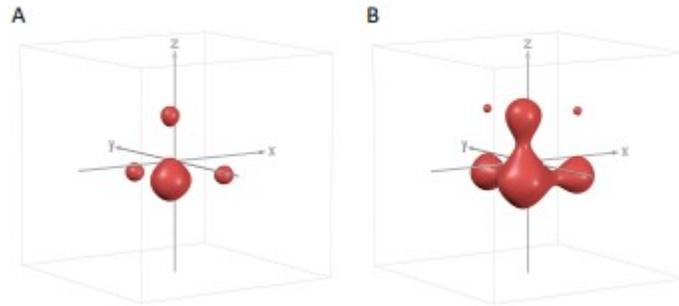
Página 20: Explicación de Funciones Avanzadas y Propiedades de la Inteligencia

La Geometría de la "Comprensión"

La comprensión es un proceso cognitivo crucial que permite la asimilación coherente de nueva información, habilitando un funcionamiento y aprendizaje cognitivo suave. En el contexto de nuestro marco geométrico, la comprensión involucra la integración y estabilización de nueva información dentro de la variedad de la inteligencia. Al comprender la geometría de la "comprensión" y la "falta de comprensión", podemos mejorar nuestros modelos de entrenamiento, mejorando la comprensión y abordando la falta de comprensión de manera más efectiva. En nuestro marco basado en la geometría Riemanniana, estos conceptos se esclarecen a través de la integración de tokens, la navegación geodésica y las dinámicas de curvatura.

Los tokens son unidades discretas de información incrustadas en un espacio de alta dimensión. La comprensión ocurre cuando los nuevos tokens se integran de manera coherente con los tokens existentes, lo que lleva a variedades suaves y estables que aseguran una incorporación perfecta de nueva información. La falta de comprensión surge cuando los nuevos tokens se integran de manera

deficiente, lo que resulta en incrustaciones dispersas o desconectadas que forman cúmulos aislados que carecen de coherencia, caracterizados por alta curvatura, singularidades o regiones no diferenciadas, como se muestra en la Figura 2. Esto conduce a geodésicas inestables o erráticas, donde el flujo de pensamiento muestra cambios abruptos, lo que indica una falta de coherencia.



(Figura 2: Representación simplificada de las variedades en el espacio de inteligencia. A. En el espacio de inteligencia, los tokens se integran de manera coherente para formar variedades suaves y estables. Algunas variedades permanecen aisladas, lo que indica que no son explicables por otras. No existen caminos geodésicos entre estas variedades aisladas. B. A medida que el aprendizaje progresa, las variedades se expanden con la adición de nuevos tokens. Consecuentemente, más variedades se interconectan, formando estructuras más grandes, suaves y estables que facilitan la navegación geodésica más prolongada y compleja a lo largo de ellas.)

El análisis de curvatura puede identificar regiones propensas a representar mala comprensión o falta de comprensión, permitiéndonos enfocarnos en esas áreas y mejorar la integración de los tokens. En casos de falta de comprensión, los bucles de retroalimentación fallan en converger, lo que lleva a un flujo constante. Por el contrario, la comprensión indica bucles de retroalimentación estables, lo que indica información bien integrada. Las transiciones erráticas en los estados indican falta de comprensión, mientras que las transiciones suaves indican comprensión estable. Visualizar la variedad utilizando técnicas como t-SNE (Van der Maaten y Hinton 2008) o UMAP (McInnes et al., 2018) ayuda a identificar estas transiciones y corregir las áreas problemáticas.

Página 21: Creatividad y Resolución de Problemas

Pensamiento Creativo

El pensamiento creativo es el proceso de generar nuevas ideas, originales y soluciones. A menudo se describe como una recombinação de conocimiento existente de maneras novedosas, facilitada por el pensamiento divergente y la flexibilidad cognitiva (Beaty et al., 2016). Dentro de la teoría de la geometría de la inteligencia, el pensamiento creativo puede modelarse como la navegación de geodésicas que conectan ideas diversas y previamente no vinculadas en la variedad cognitiva.

Resolución de Problemas

La resolución de problemas es el proceso cognitivo de encontrar soluciones a problemas complejos o desafiantes. En el contexto de la geometría de la inteligencia, la resolución de problemas implica navegar por la variedad para identificar y recorrer caminos geodésicos que conducen a soluciones efectivas.