

Introducción a la IA Generativa

Inteligencia artificial

Patricio García Báez

Grado en Ingeniería Informática

Definición

Historia

Modelos

Modalidades

LLM

Prompting

Problemas

Inteligencia Artificial Generativa

- Tipo de sistema de IA capaz de generar texto, imágenes u otros medios
- Suelen generar su salida en respuesta a comandos o *prompts*
- Generan nuevos datos con características similares a sus patrones de entrenamiento
- Aprenden los patrones y las estructuras subyacentes de sus datos de entrenamiento
- Utilizan paradigmas de aprendizaje no supervisado o *autosupervisado*
- Se distinguen de los *modelos discriminativos*, orientados mas hacia la clasificación

Modelo discriminativo vs. modelo generativo

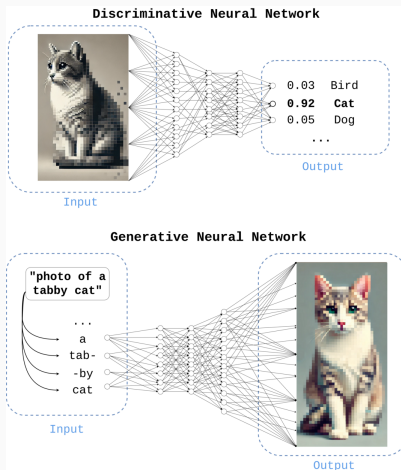


Figura 1: Ejemplo de modelo discriminativo vs. modelo generativo (Lwneal, Wikimedia)

Historia: Inicios

- ~2010: Aprendizaje profundo o *Deep Learnig* (DL)
- 2014: Autocodificador variacional y red generativa adversativa (GAN), primeras redes neuronales DL prácticas capaces de actuar como modelos generativos
- 2017: *Transformer*, por Google, modelo DL base de gran mayoría de IA generativas actuales
- 2018: El primer GPT (*Generative Pre-trained Transformer*) de OpenAI, un tipo de modelo de lenguaje grande (LLM)
- 2019: GPT-2, de OpenAI

Historia: AI Boom

- 2021-22: DALL-E, Midjourney y Stable Diffusion. Modelos generativos de imágenes de alta calidad
- 2021: LaMDA, chatbot LLM Google
- 2021: GitHub Copilot, asistente de código de Visual Studio Code, inicialmente basado en versión de GPT-3 de OpenAI
- 2022: ChatGPT (*Chat Generative Pre-trained Transformer*), chatbot que usa los modelos GPTs
- 2023: LLaMA, LLM de código abierto de Meta
- 2023: Gemini, chatbot de Google DeepMind, sucesor de LaMDA y PaLM
- 2023: GPT-4, de OpenAI, versión temprana de IA fuerte, según Microsoft
- 2023: Microsoft Copilot (inicialmente Bing Chat), chatbot basado en GPT-4

Diferentes modelos de DL se han ido sucediendo, mejorando progresivamente sus capacidades de generación:

- **Autocodificador variacional**: RN con un cuello de botella a su entrada para comprimirla, y luego un decodificador
- **GAN** (Generative Adversarial Network): sistema de dos RNs que compiten mutuamente en juego de suma cero
- **LSTM** (Long Short-Term Memory): RN recurrente que abordar el problema de desvanecimiento de gradiente para aprender dependencias a largo plazo en secuencias de datos
- **Transformer**: RN que incorpora mecanismo de autoatención, dando un peso diferente a cada parte de la entrada
- **LLM** (Large Language Model): Habitualmente variante de Transformer pre-entrenado con un corpus textual enorme

Principales modalidades o tipos de datos de entrenamiento:

- Texto
- Código
- Imágenes
- Moléculas
- Música
- Vídeo
- Acciones
- Modelos 3D

Unimodales o multimodales, según acepten o no distintos tipos de datos a la entrada

LLM: Definición

- **Modelo de Lenguaje (ML)**: modelo que trata de predecir la siguiente palabra en un documento
 - "Muchos años después, frente al pelotón de fusilamiento, el coronel Aureliano Buendía había de recordar aquella tarde remota en que su padre lo llevó a conocer el _____"
- **LLM**: ML con muchos parámetros (millones) y entrenado con grandes cantidades de texto sin etiquetar
- Puede capturar gran parte de sintaxis y semántica del lenguaje humano
- Almacena un conocimiento general considerable sobre el mundo y es capaz de "memorizar" una gran cantidad de hechos
- Al crecer es posible la aparición de *habilidades emergentes*

- **Tokenización:** Las palabras de entrada se transforman en *tokens* (números), una secuencia de ellos constituirá la entrada. La salida será una distribución de probabilidad sobre el vocabulario de tokens
- **Entrenamiento:** mediante *preentrenamiento generativo*, dado un conjunto de datos de entrenamiento de billones de tokens de texto, el modelo predice los tokens en el conjunto de datos
- **Tareas posteriores:**
 - Ajuste fino: Entreno supervisado posterior para una tarea específica
 - Prompting: Incluyendo técnicas de "incitación" tipo *zero-shot* o proporcionando ejemplos de resolución

Lista de ejemplos de LLM:

- **BERT**: Google, 2018. 340 mill. de parámetros
- **GPT-2**: OpenAI, 2019. 1500 mill. de parámetros
- **GPT-3**: OpenAI, 2020. 175000 mill. de parámetros
- **LaMDA**: Google, 2022. 137000 mill. de parámetros
- **PaLM**: Google, 2022. 540000 mill. de parámetros
- **LLaMA**: Meta, 2023. 65000 mill. de parámetros
- **GPT-4**: OpenAI, 2024. ~ 1 bill. de parámetros
- **PanGu- Σ** : Huawei, 2023. 1,085 bill. de parámetros

Prompting

- *Prompt*: texto en lenguaje natural que describe la tarea que debe realizar una IA
- *Prompt engineering*: proceso de estructurar un prompt
- Aprendizaje en contexto: capacidad de un modelo para aprender temporalmente a partir de indicaciones (*mesa-optimización*)
- Puede adaptar la forma de:
 - Consulta directa
 - Orden específica
 - Declaración más larga, que incluya contexto

Prompting: Algunas técnicas

- *Cadena de pensamiento (CoT)*: guía al LLM para que explique cómo resuelven un problema paso a paso, antes de dar la respuesta final
- *Solicitud de conocimiento generado*: guía al LLM pidiéndole que primero cree información relacionada con la consulta y luego la use para dar la respuesta
- *Indicación de menor a mayor*: guía al LLM para que resuelva el problema por pasos, primero los mas simples y acabando con los mas complejos
- *En texto a imagen*: indicaciones breves y descriptivas, que incluyan descripción, medio deseado, estilo, iluminación, color y textura. Posible uso de *indicaciones negativas*

- Perdida de empleos
- *Deepfakes*
- Cibercrimen
- Costo de los entrenamientos
- Calidad de contenidos (*alucinaciones*)
- Propiedad intelectual
- Regulación



Figura 2: Ejemplos generados con Midjourney, toma buena y falsa (Porras L., "Estos niños no existen")

Esta presentación está bajo una licencia Creative Commons Attribution-ShareAlike 4.0 International License.





Riedl, Mark.

A very gentle introduction to large language models without the hype.

<https://mark-riedl.medium.com/a-very-gentle-introduction-to-large-language-models-without-the-hype-2023>.

[Online; accessed 19-November-2024].



S. Russell and P. Norvig.

Artificial Intelligence: A Modern Approach (4rd Edition).

Prentice Hall, 2021.



Wikipedia.

Inteligencia artificial generativa — wikipedia, la enciclopedia libre.

`https:`

`//es.wikipedia.org/w/index.php?title=Inteligencia_artificial_generativa&oldid=163178143, 2024.`

[Internet; descargado 23-octubre-2024].



Wikipedia contributors.

Generative artificial intelligence — Wikipedia, the free encyclopedia.

`https:`

`//en.wikipedia.org/w/index.php?title=Generative_artificial_intelligence&oldid=1257969785, 2024.`

[Online; accessed 19-November-2024].

