

Introducción a Transformers y Modelos Preentrenados

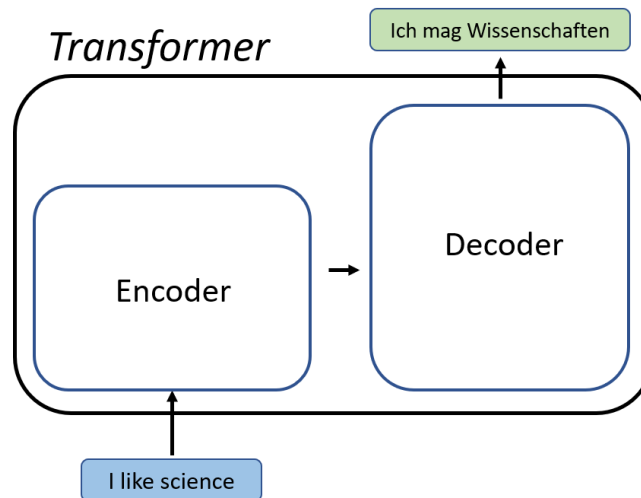
UNMSM – FISI – Unidad de Post Grado

Profesor: Juan Gamarra Moreno

Introducción a los Transformers

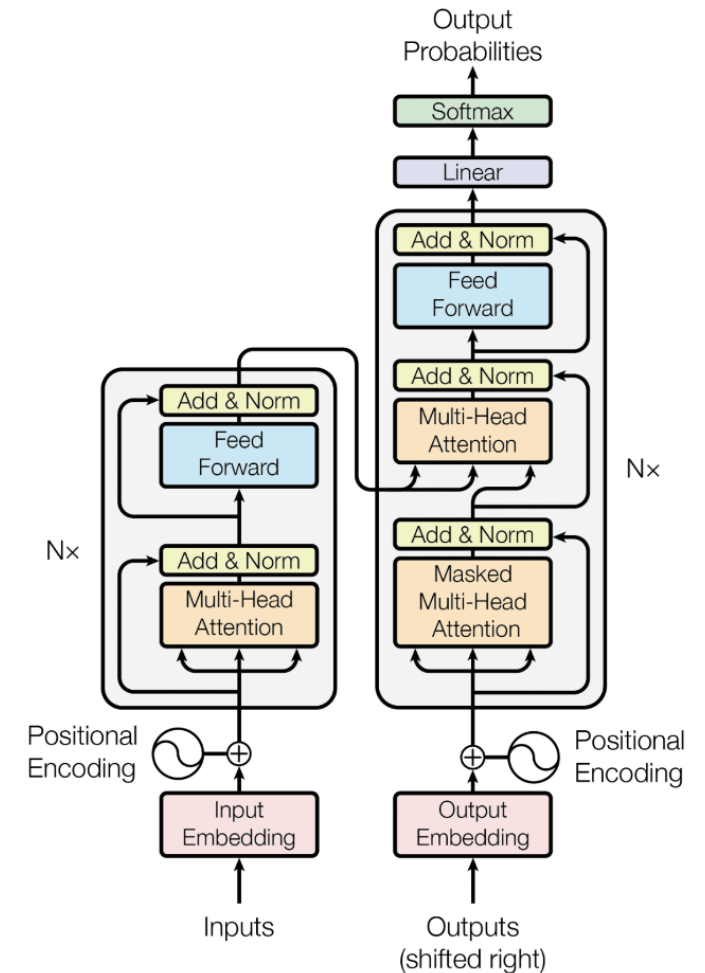
¿Qué es un Transformer?

- Un **Transformer** es una arquitectura de red neuronal desarrollada específicamente para manejar datos secuenciales, como texto, pero que ha demostrado ser útil en una amplia gama de tareas de aprendizaje profundo. Fue introducida en el artículo *"Attention Is All You Need"* (Vaswani et al., 2017).



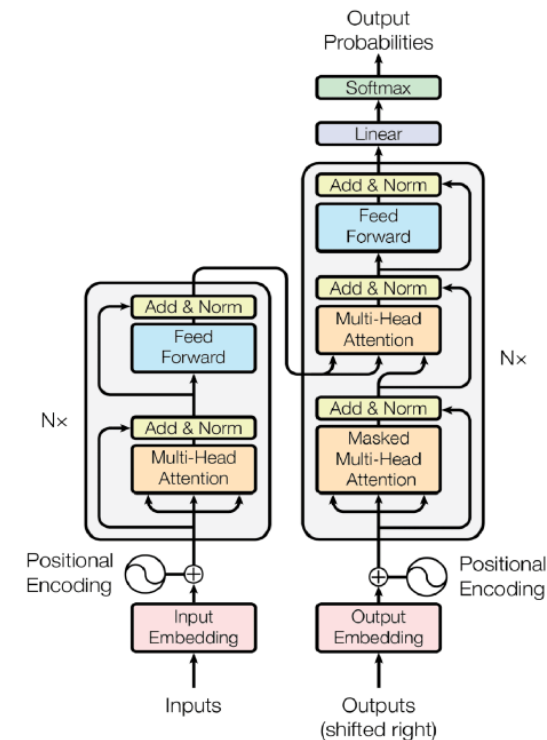
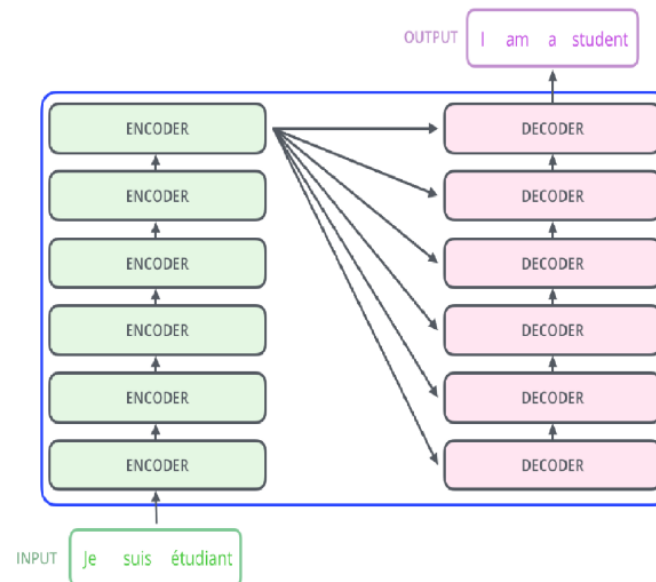
¿Qué es un Transformer?

- El Transformer se basa en un mecanismo llamado **atención** y su variante **auto-atención**, que permite a la red asignar pesos diferentes a las partes más relevantes de la entrada, considerando todo el contexto de la secuencia en paralelo. Esto lo hace más eficiente y preciso que los modelos anteriores como las Redes Neuronales Recurrentes (RNN) o LSTMs, que procesaban los datos de forma secuencial.



¿Para qué sirve un Transformer?

Los Transformers tienen aplicaciones prácticas en muchas áreas, principalmente en **procesamiento de lenguaje natural (NLP)** y otras tareas que involucran secuencias o datos estructurados.



¿Para qué sirve un Transformer?

1. Procesamiento de Lenguaje Natural (NLP):

- **Traducción automática:**
 - Ejemplo: Traducir texto de un idioma a otro (como Google Translate).
- **Generación de texto:**
 - Ejemplo: Crear respuestas en chatbots o generar contenido creativo.
- **Análisis de sentimientos:**
 - Ejemplo: Clasificar reseñas de productos como positivas o negativas.
- **Clasificación de texto:**
 - Ejemplo: Categorización de correos como spam o no spam.
- **Resumen automático:**
 - Ejemplo: Reducir un texto largo a un resumen breve.
- **Extracción de información:**
 - Ejemplo: Identificar entidades (nombres, fechas, ubicaciones) en un texto.

¿Para qué sirve un Transformer?

2. Visión por Computadora:

- **Clasificación de imágenes:**

- Usar Transformers adaptados (Vision Transformers, ViT) para categorizar imágenes.

- **Detección de objetos:**

- Identificar y localizar objetos en una imagen.

3. Audio y Video:

- **Reconocimiento de voz:**

- Convertir voz en texto con modelos como Whisper.

- **Procesamiento de video:**

- Capturar dependencias temporales entre cuadros para tareas como clasificación de videos.

¿Para qué sirve un Transformer?

4. Ciencia y Modelado de Datos:

- **Predicción de series temporales:**
 - Aplicado en datos financieros, meteorología y demanda de productos.
- **Modelado de proteínas:**
 - Ejemplo: AlphaFold usa Transformers para predecir estructuras de proteínas.

Introducción a la arquitectura Transformer

Problema que resuelve

- **Limitaciones de RNNs y LSTMs:**

- Dificultad para procesar secuencias largas debido a la naturaleza secuencial del procesamiento.
- Problemas de desvanecimiento y explosión de gradiente que afectan el aprendizaje de relaciones a largo plazo.
- Baja eficiencia computacional debido a la dependencia temporal en los cálculos.

Mecanismo de atención (Attention Mechanism)

- **Atención:**

- Introducción del concepto: Identificación de las partes más relevantes de la entrada para enfocarse en ellas.
- Ejemplo simple: En traducción, prestar más atención a ciertas palabras del texto original.

- **Ventaja del mecanismo de atención:**

- Permite procesamiento en paralelo.
- Mejora la captura de dependencias a largo plazo.

Auto-atención (Self-Attention)

- **Concepto:**

- Cada palabra o token en la secuencia se relaciona consigo misma y con otros tokens.
- Explicación matemática:
 - Cada token genera tres vectores: **Query (Q)**, **Key (K)** y **Value (V)**.
 - Cálculo de puntuación de atención:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Donde d_k es la dimensionalidad de K.

- El resultado es una representación ponderada de los tokens.

Estructura de las capas del Transformer

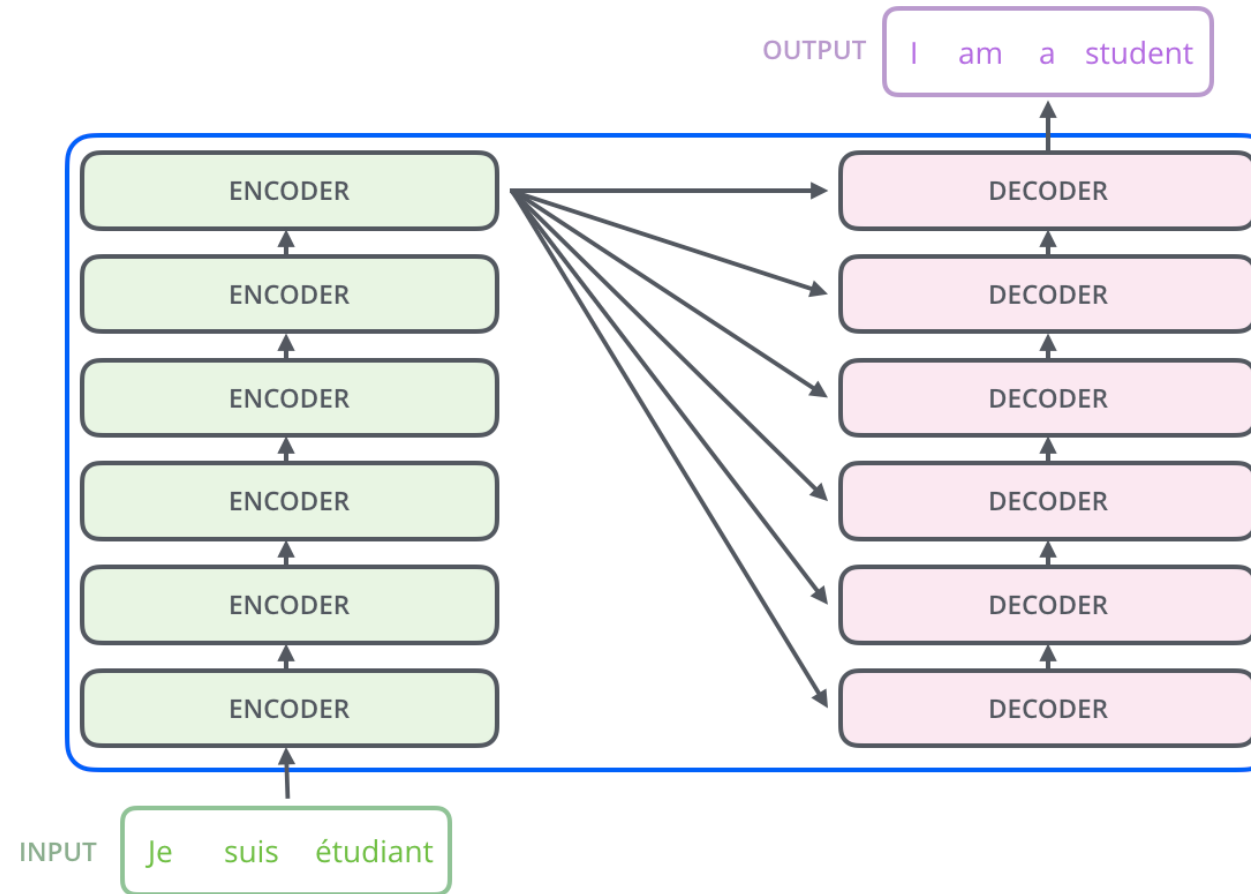
- **Codificador (Encoder):**

- Múltiples capas con mecanismos de atención y redes feedforward.
- Representación rica del contexto de la secuencia de entrada.

- **Decodificador (Decoder):**

- Similar al codificador, pero incluye atención al contexto generado por el codificador.
- Se utiliza para tareas de generación, como traducción o generación de texto.

Estructura de las capas del Transformer



Transformers en la actualidad

Ventajas de los Transformers

Procesamiento en paralelo:

- RNN/LSTM procesan tokens uno por uno, mientras que Transformers procesan la secuencia completa en paralelo, lo que reduce el tiempo de entrenamiento.

Limitaciones superadas:

- **Desvanecimiento del gradiente:** Resuelto gracias a la eliminación de dependencias recurrentes.
- **Escalabilidad:** Mejor aprovechamiento de hardware moderno, como GPUs y TPUs.

Resultados en benchmarks:

- Traducción automática (WMT2014): Transformers superan a RNNs/LSTMs.
- Clasificación de texto: Mejor rendimiento en tareas como análisis de sentimientos (GLUE benchmark).

Modelos clave preentrenados basados en Transformers

- **BERT:**

- **Bidireccionalidad:** Captura contexto antes y después de cada palabra.
- Ejemplo: Clasificación de preguntas como respuesta o no respuesta.

- **GPT:**

- Modelo autoregresivo: Genera texto palabra por palabra, prediciendo el siguiente token.
- Ejemplo: Generación de historias.

- **T5:**

- Traduce todas las tareas en problemas de texto a texto.
- Ejemplo: Resumen de textos largos.

Proceso de preentrenamiento y ajuste fino

- **Preentrenamiento:**

- Corpus grande y generalizado.
- Tareas: Máscara de palabras (BERT), predicción de la siguiente palabra (GPT).

- **Ajuste fino:**

- Dataset específico, como análisis de sentimientos o clasificación.
- Adaptación rápida a problemas particulares.

Aplicaciones prácticas

Clasificación de texto:

- Tarea: Clasificar comentarios en positivos o negativos.
- Ejemplo: Análisis de reseñas de productos.

Generación de texto:

- Aplicaciones: Redacción creativa, respuestas en chatbots.
- Ejemplo: GPT generando una historia a partir de una línea inicial.

Traducción automática:

- Mejora la calidad de traducción frente a modelos estadísticos tradicionales.
- Ejemplo: Traducción inglés-español con Transformer.

Análisis de sentimientos:

- Ejemplo práctico: Clasificación de tweets en positivos, negativos o neutros.
- Uso en marketing digital para entender la percepción de marcas.