|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | |  | | --- | |  | | **Universidad Iberoamericana de Puebla** | | ***Agentes Inteligentes***  **Diego Pacheco Valdez**  **Gabriel Pablo Torres** | |  | |  | |  | |  | | **“Actividad de Investigación**  **Transformers”** | | Noviembre 2024 | |  | |  | |  |

**Introducción a los Transformers**

Los transformers son una arquitectura de redes neuronales que ha revolucionado el campo del procesamiento de lenguaje natural (NLP) y otros dominios donde se procesan secuencias de datos. A diferencia de las redes recurrentes (RNNs) y las redes convolucionales (CNNs), los transformers no dependen de la secuencialidad para procesar datos, lo cual permite un procesamiento paralelo y una eficiencia significativamente mayor. Esto se logra mediante un mecanismo de autoatención, que les permite centrarse en diferentes partes de una secuencia al procesar cada elemento.

**Principales Aportes de *Attention Is All You Need***

El artículo *Attention Is All You Need*, publicado por Vaswani et al. en 2017, introdujo el modelo transformer, que reemplaza los mecanismos tradicionales de recurrencia con un enfoque completamente basado en la atención. En lugar de procesar secuencias en pasos lineales, la atención permite que el modelo analice todas las posiciones de una secuencia simultáneamente, capturando dependencias a largo plazo e forma efectiva.

El modelo propuesto en el artículo se compone de capas de **encoder** y **decoder**. El encoder transforma la entrada en una representación de alto nivel, mientras que el decoder utiliza esta representación para generar la salida final. Cada capa de encoder y decoder incluye mecanismos de autoatención y subcapas de redes neuronales feed-forward, permitiendo que el modelo preste atención tanto a palabras cercanas como lejanas en un contexto.

**Impacto y Aplicaciones**

La introducción de los transformers y el mecanismo de autoatención ha impulsado avances significativos en NLP, especialmente en tareas como la traducción automática, el resumen de texto y la generación de lenguaje. Modelos derivados de esta arquitectura, como BERT y GPT, han establecido nuevos estándares en precisión y han ampliado las capacidades de los sistemas de inteligencia artificial en tareas de lenguaje. El enfoque de atención ha demostrado ser clave no solo en NLP, sino también en áreas como visión por computadora, debido a su capacidad para manejar grandes volúmenes de datos y relaciones complejas.

**Conclusión**

El artículo *Attention Is All You Need* sentó las bases para una era de modelos basados en transformers, cambiando el paradigma del procesamiento de lenguaje y estableciendo nuevas posibilidades en múltiples áreas de la inteligencia artificial. Su enfoque en la autoatención y el procesamiento paralelo no solo mejoró el rendimiento, sino que abrió la puerta a arquitecturas más complejas y potentes, consolidando a los transformers como una pieza central en el desarrollo de IA moderna.