Apuntes IA Clase 7/10

Gianmarco Oporta Pérez
Ingeniería en Computación
Instituto Tecnológico de Costa Rica
San José, Costa Rica
gooporta@estudiantec.cr

***Abstract*—El presente documento recopila los apuntes de la clase del 7 de octubre, correspondientes al curso de Inteligencia Artificial. Se abordan los lineamientos del Proyecto 1, centrado en redes neuronales convolucionales aplicadas al reconocimiento de voz mediante espectrogramas, así como los fundamentos teóricos de las redes convolucionales y su aplicación en tareas de clasificación y segmentación de imágenes.**

***Index Terms*—Inteligencia Artificial, Redes Neuronales Convolucionales, Clasificación de Audio, PyTorch, Espectrogramas**

I. Resultados del Quiz 5

Se realizó el quiz 5 durante el inicio de la sesión, la cual contiene las siguientes preguntas:

- **Pregunta:** Describa qué es una red totalmente conectada.
  R/ Es una red neuronal en la cual cada neurona está conectada con todas las neuronas de la capa siguiente, de principio a fin.
- **Pregunta:** Mencione tres funciones de activación no lineales.
  R/ ReLU, Sigmoide y Tanh.
- **Pregunta:** Describa los cuatro componentes principales de un agente LLM.
  R/
  - **Perfil:** Define la identidad o personalidad del agente, determinando su tono, estilo de comunicación y comportamiento general.
  - **Memoria:** Permite conservar información relevante de interacciones anteriores o resultados previos, facilitando la continuidad y el contexto en tareas extensas.
  - **Herramientas:** Corresponden a recursos o funciones externas que el agente puede invocar para realizar operaciones específicas o acceder a información adicional.
  - **Planificación o razonamiento:** Consiste en la capacidad del agente para interpretar las instrucciones del usuario, elaborar estrategias y seleccionar la acción más apropiada según el objetivo planteado.
- **Pregunta:** Describa la diferencia entre sistemas de agente único y sistemas multiagentes.
  R/ Un agente único percibe su entorno, toma decisiones y ejecuta acciones de manera independiente, mientras que los sistemas multiagentes están conformados por múltiples agentes que colaboran o compiten entre sí y con su entorno para alcanzar objetivos comunes o individuales.

II. Descripción del Proyecto

El primer proyecto tiene como objetivo aplicar redes neuronales convolucionales (CNN) en la tarea de reconocimiento de voz a partir de espectrogramas. Su propósito es desarrollar un modelo de clasificación multiclase capaz de identificar distintos tipos de sonidos, tales como vocalizaciones humanas y animales, instrumentos musicales o ruidos ambientales.

Para lograrlo, se utiliza un conjunto de datos público orientado a la clasificación de audio, en el cual cada muestra corresponde a una grabación corta etiquetada con su clase correspondiente. Antes del entrenamiento, las señales acústicas son transformadas en representaciones visuales denominadas espectrogramas, las cuales se emplean como entrada a las redes convolucionales.

El proyecto requiere la construcción manual de dos modelos empleando PyTorch sin librerías de alto nivel:

- **Modelo A:** Variante clásica de LeNet-5 adaptada al reconocimiento de audio mediante espectrogramas.
- **Modelo B:** Arquitectura alternativa fundamentada en literatura académica o en un diseño propio justificado teóricamente.

Se deben generar dos versiones del conjunto de datos: una con los audios transformados a imágenes (base) y otra con versiones aumentadas mediante técnicas de *data augmentation* orientadas al dominio del audio. Este proceso busca incrementar la robustez del modelo y su capacidad de generalización.

Durante la fase de entrenamiento se construyen cuatro combinaciones principales: Modelo A/Base, Modelo A/Aumentado, Modelo B/Base y Modelo B/Aumentado. Cada modelo se entrena con diferentes hiperparámetros, evaluando su desempeño con métricas como precisión, pérdida, F1-Score y matriz de confusión. La herramienta *Weights & Biases* se utiliza para monitorear y visualizar los resultados durante el entrenamiento.

Finalmente, los modelos seleccionados se comparan para determinar el de mejor rendimiento general. El informe debe presentarse en formato IEEE, acompañado del código fuente y el cuaderno en Jupyter Notebook.

III. Aspectos Prácticos del Proyecto

El desarrollo debe realizarse en PyTorch, construyendo manualmente cada capa de la red. Es necesario registrar todas las métricas relevantes utilizadas en clases anteriores, incluyendo la pérdida, precisión y F1-Score. Dado que los espectrogramas generados pueden ser pesados, se recomienda reducir su resolución a 224 píxeles por lado.

La fecha tentativa de entrega fue establecida para el jueves 30 de octubre. Se espera que los modelos implementados sean completamente reproducibles y que incluyan mecanismos de control de aleatoriedad y registro de resultados.

IV. Fundamentos de Redes Neuronales Convolucionales

Las redes neuronales convolucionales (CNN, por sus siglas en inglés) representan una evolución de las redes neuronales tradicionales orientadas al procesamiento de datos estructurados espacialmente, como imágenes o espectrogramas. A diferencia de las redes completamente conectadas, las CNN aprenden patrones espaciales y jerárquicos de manera eficiente.

Hasta este punto del curso, se ha trabajado con redes que reciben un vector de características como entrada, lo transforman mediante capas ocultas y producen una salida. Sin embargo, este enfoque no considera la estructura espacial de los datos, lo cual puede generar errores al mover, rotar o escalar los objetos dentro de una imagen.

A. Ejemplo: Dataset CIFAR-10

El conjunto de datos CIFAR-10 se utiliza frecuentemente para experimentación en visión por computadora. Contiene imágenes a color de tamaño $32 \times 32 \times 3$, representando diez clases diferentes.

Aunque cada imagen es relativamente pequeña, una versión de mayor resolución, por ejemplo $200 \times 200 \times 3$, aumentaría drásticamente el número de parámetros de entrada, lo que dificultaría la escalabilidad del modelo.

B. Estructura de una Red Convolucional

En una red convolucional, las neuronas se organizan en tres dimensiones: ancho, alto y profundidad. Cada neurona está conectada únicamente a una pequeña región de la capa anterior, en lugar de estar completamente conectada, lo cual reduce la complejidad y mejora la eficiencia computacional.

Las capas principales que componen una CNN son:

- **Capa convolucional:** Calcula las salidas de las neuronas conectadas a regiones locales de la capa anterior.
- **Capa de agrupamiento (pooling):** Reduce el tamaño espacial de las representaciones, manteniendo las características más relevantes.
- **Capa completamente conectada (fully connected):** Transforma la representación final en probabilidades de pertenencia a una clase específica.

A medida que las imágenes avanzan a través de las capas convolucionales y de agrupamiento, se reduce su tamaño espacial, pero aumenta la abstracción de las características aprendidas.

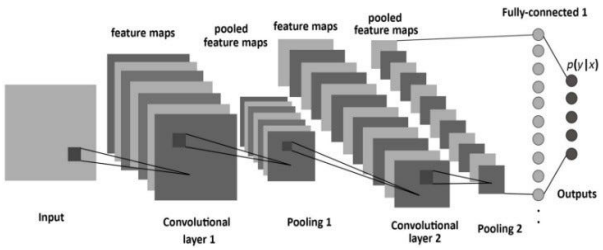

Fig. 1. redes convolucionales

C. Aplicaciones Comunes

Las redes convolucionales se aplican ampliamente en diversas tareas de visión artificial, entre las que destacan:

- Clasificación de imágenes.
- Segmentación de objetos.
- Segmentación de instancias.
- Procesamiento general de imágenes.

Estas arquitecturas han demostrado una gran eficacia en problemas de reconocimiento visual, detección de patrones y procesamiento de señales en el dominio de la visión.

References

[1] S. Pacheco, “Convolutional Neural networks” Presentación,