

Normativa

- La práctica se realizará en grupos de 2 alumnos.
- Si se detecta cualquier sospecha de copia conllevará un 0.
- Es obligatorio completar los dos apartados para poder participar en la evaluación final.
- Los resultados obtenidos y presentados deben ser producto original del trabajo de los estudiantes.
- Los notebook deberán contener celdas de documentación.

1. Objetivo de la práctica

En esta práctica se abordarán dos partes diferenciadas pero complementarias:

1. **Fine-tuning mediante LoRA (Low-Rank Adaptation)** sobre modelos preentrenados.
2. **RAG (Retrieval-Augmented Generation)** para integrar recuperación de conocimiento con generación.

Cada bloque tendrá una valoración de **5 puntos**, según la rúbrica especificada en cada apartado.

2. Parte I: Fine-tuning con LoRA (5 puntos)

Descripción general

El objetivo de esta parte es realizar el ajuste fino de modelos de lenguaje utilizando la técnica *Low-Rank Adaptation (LoRA)* sobre un dataset elegido por el alumno.

Tareas a realizar

1. **Selección del dataset:** Elija un dataset adecuado (preferiblemente de texto, no visto en clase) y justifique brevemente su elección.
2. **Selección de modelos:** Elija dos modelos distintos desde Hugging Face a los usando en los notebooks de ejemplo de la asignatura que no hayan sido especializados en la tarea.
3. **Análisis de los modelos:** Compare y analice ambos modelos, considerando:
 - Número de parámetros, arquitectura base, licencia, peso y tamaño del modelo, técnicas de preentrenamiento utilizadas
 - Otros parámetros de interés o destacables.
4. **Fine-tuning con LoRA:** Entrene ambos modelos sobre el dataset seleccionado utilizando la técnica *LoRA*. Puede usar bibliotecas **PEFT** y **transformers**.
5. **Evaluación de resultados:** Compare el rendimiento de ambos modelos tras el ajuste fino. Justifique cuál de los dos obtiene mejores resultados y por qué.

Rúbrica de evaluación

- Dataset original (no visto en clase): **Obligatorio**
- Análisis en profundidad de los modelos (parámetros, licencias, pesos, técnicas, etc.): **1 puntos**
- Entrenamiento LoRA 1 correctamente implementado: **1 puntos**
- Entrenamiento LoRA 2 correctamente implementado: **1 puntos**
- Análisis crítico y justificación de resultados: **2 punto**

3. Parte II: Recuperación Asistida (RAG) (5 puntos)

Descripción general

El objetivo de esta parte es integrar recuperación de información con generación mediante la arquitectura RAG. El alumno deberá implementar y comparar dos configuraciones distintas.

Tareas a realizar

1. **Corpus de conocimiento:** Seleccione un corpus propio o descargado (artículos, documentación técnica, resúmenes, etc.). Elija un dominio de interés y justifique su complejidad.
2. **Implementación 1:** Utilice un **modelo de embeddings A** y un **modelo generador A** para construir un pipeline RAG básico.
3. **Implementación 2:** Cambie ambos modelos (embedding y generador) y repita el experimento con el mismo corpus.
4. **Evaluación:** Realice pruebas de recuperación y generación, analizando las diferencias en precisión y coherencia de las respuestas.

Rúbrica de evaluación

- Complejidad y originalidad del corpus: **1 punto**
- Implementación 1 con dos modelos distintos: **1.5 puntos**
- Implementación 2 con dos modelos distintos: **1.5 puntos**
- Comparación y análisis de resultados: **1 punto**

4. Entrega en Moodle

Cada grupo deberá subir a Moodle:

- Un archivo **.pdf** con los dos análisis y las conclusiones.
- Los **notebooks** empleados en el desarrollo de los entrenamientos y pruebas.

El informe deberá ser claro, lógico y mostrar comprensión profunda de los conceptos de fine-tuning y RAG. La redacción crítica y la justificación de decisiones técnicas serán valoradas positivamente.

Esta obra está bajo una licencia Creative Commons “Atribución-NoComercial-CompartirIgual 3.0 No portada”.

