

Tarea 5 NLP

Rubén Manrique

October 2025

Encoders vs Decoders para Clasificación de Texto

Objetivo

El propósito de esta tarea es comparar el desempeño de arquitecturas de tipo **encoder** (modelos BERT ajustados con una cabeza de clasificación) y **decoder** (modelos de lenguaje autoregresivos mediante prompting sin ajuste de parámetros) en tareas de clasificación de texto. Utilizarán los mismos conjuntos de datos empleados en tareas previas: **20Newsgroups** para clasificación multiclase y **Multi-Domain Sentiment Dataset** para análisis de sentimiento. Esta comparación permitirá analizar los límites, ventajas y desventajas de cada enfoque bajo restricciones computacionales razonables.

Instrucciones Generales

- Utilice **Jupyter Notebooks** para el desarrollo de la tarea.
- Incluya docstrings detallados en todas las clases, funciones y métodos.
- Ejecute todos los bloques de código y asegúrese de incluir los resultados en la entrega.
- Presente un informe en PDF que resuma la metodología, resultados y análisis comparativo.
- Pueden realizar en grupos de hasta 4 personas.

Datasets

- **20Newsgroups:** Clasificación en 20 categorías temáticas.
- **Multi-Domain Sentiment Dataset:** Clasificación binaria de sentimiento (positivo/negativo) en cuatro categorías: Books, DVD, Electronics, Kitchen.

Modelos a Utilizar

1. Encoders:

- Utilice al menos dos variantes de modelos BERT preentrenados de código abierto con menos de 200 millones de parámetros (por ejemplo, **bert-base-uncased** y **distilbert-base-uncased**).
- Ajuste (*fine-tuning*) los modelos para cada tarea de clasificación, agregando una cabeza de clasificación (capa densa).
- Justifique la selección de los modelos y documente el proceso de ajuste, hiperparámetros y criterios de parada.

2. Decoders:

- Use al menos dos modelos generativos autoregresivos de código abierto con menos de 7 mil millones de parámetros (**7B**), tales como **GPT-2**, **GPT-Neo-1.3B**, **OPT-1.3B**, **Llama 2 7B** o equivalentes.
- No es obligatorio ajustar los pesos de los modelos; puede utilizar técnicas de **prompting** para resolver las tareas de clasificación.
- Diseñe prompts específicos para cada tarea, justificando su redacción y estructura.
- Ejemplo de prompt para clasificación de sentimiento:

Review: "The battery life of this camera is amazing."
Sentiment (positive/negative):

- Ejemplo de prompt para clasificación de newsgroups:

Text: "The new Linux kernel has improved networking features."
Category:

Metodología

1. Para ambos datasets, divida los datos en conjuntos de entrenamiento, validación y prueba siguiendo los mismos porcentajes que en las tareas anteriores.
2. Para los modelos encoder, ajuste cada modelo en el conjunto de entrenamiento y seleccione buenos hiperparámetros utilizando validación.
3. Para los modelos decoder, diseñe prompts efectivos y utilice las respuestas generadas para inferir la clase correspondiente.
4. Reporte y compare los resultados en términos de **precisión**, **recall**, **F1** (macro y micro), y **accuracy**.

Análisis y Discusión

- Compare el desempeño entre encoders y decoders para cada dataset y tarea.
- Analice la robustez de cada enfoque frente a la variabilidad de los datos.
- Discuta las ventajas y limitaciones de cada paradigma en términos de:
 - Facilidad de uso y ajuste
 - Requerimientos computacionales
 - Interpretabilidad
 - Adaptabilidad a nuevas tareas o dominios
- Justifique la selección de prompts y analice la sensibilidad de los resultados frente a su formulación.
- Incluya un apartado de recomendaciones sobre el uso de encoders y decoders para tareas de clasificación de texto en función de los resultados obtenidos.

Entrega

- Suba todos los notebooks, el informe PDF y los archivos necesarios a la plataforma indicada.
- No envíe la tarea por correo electrónico.

Notas

- Se valorará el rigor experimental, la claridad en la justificación de las decisiones tomadas y la profundidad del análisis comparativo.
- Puede apoyarse en **Transformers** de HuggingFace y librerías asociadas.
- Si utiliza recursos computacionales limitados, puede reducir el tamaño del conjunto de datos o el número de épocas, pero debe justificarlo adecuadamente.