

CLASIFICACIÓN DE POEMAS COMO HUMORÍSTICOS CON LLM

PABLO RUIZ FABO

Université de Strasbourg / CiTIUS, Universidade de Santiago de Compostela

Los grandes modelos de lenguaje (LLM), dada su capacidad de generar un texto apropiado como respuesta a una tarea arbitraria, para la que no fueron entrenados directamente, tienen aplicaciones prácticas a la anotación automática en Humanidades Digitales impensables hace años. Presentamos experimentos de clasificación automática de poemas como humorísticos o no con LLM, mediante *prompts* sin afinado. El corpus comporta 100 sonetos en español (siglos XVI a XX). Pedimos también al modelo una explicación de la clasificación que propone. Otra instrucción pide identificar al autor o autora y continuar el poema, para comprobar si el modelo memorizó el texto en su entrenamiento, lo que no pareció ser el caso.

Los resultados varían ampliamente según el modelo. Gpt-4o de OpenAI, un modelo con capacidades altas, obtiene 0.88 F1, fallando en casos sutiles. Gpt-3.5-turbo, un modelo más básico de OpenAI, da interpretaciones anacrónicas y elige en exceso la categoría “humorístico”, considerando cómica la dicción poética en poemas de corte clásico. Sin embargo, gpt-4o-mini, modelo básico que reemplazó a 3.5-turbo en julio de 2024, obtiene resultados comparables a los de gpt-4o, con patrones y causas de error similares.

Los modelos Mistral muestran otro comportamiento. Mistral-small (modelo de base) tiene resultados bajos en clasificación y no propone explicaciones adecuadas. Mistral-large (modelo avanzado) muestra “cautela” en la categorización, rechazando clasificar como humorísticos poemas en los que identifica algún elemento de tipo cómico, proponiendo además una interpretación alternativa.

Aunque los resultados muestran variación y es difícil generalizar, el interés de la tarea reside en examinar la visión del humor que modelos ampliamente utilizados por el público promueven. Comentamos también otras limitaciones del estudio, en términos de la anotación del corpus de test y dado el uso de modelos en la nube que dificultan la reproducibilidad de los resultados.