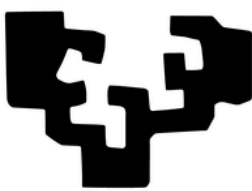


EVALUACIÓN DEL CONOCIMIENTO ESPACIAL DE LOS MODELOS MULTIMODALES

Grado en Ingeniería Informática
Facultad de Informática, Universidad del País Vasco (EHU-UPV)
Trabajo de Fin de Grado

Autora: Silvia Arenales Muñoz - Directores: Gorka Azkune y Ander Salaberria

eman ta zabal zazu



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

INFORMATIKA
FAKULTATEA
FACULTAD
DE INFORMÁTICA

DESCRIPCIÓN

El reconocimiento y la interpretación de relaciones espaciales es una tarea crucial de la inteligencia humana que nos permite comprender el mundo que nos rodea. Es por eso, que creemos que es un aspecto que se debe de estudiar para poder decir que los modelos multimodales son capaces de razonar inteligentemente.

Por tanto, hemos evaluado el modelo multitarea como, OFA utilizando un conjunto de datos adaptado específicamente para nuestro proyecto (VSR-Retrieval), derivado del dataset VSR . La evaluación se ha llevado a cabo mediante diferentes métodos y criterios.



The person is on
the top of the
horse?



The person is
beneath the
horse?



CRITERIOS DE EVALUACIÓN

1.ITM (Image Text Matching)

- 1.1. Generación de "YES"/"NO"
- 1.2. Probabilidades
- 1.2. Probabilidades Normalizadas

2. Image Captioning

- 2.1. Probabilidades medias
- 2.2. Suma Logits

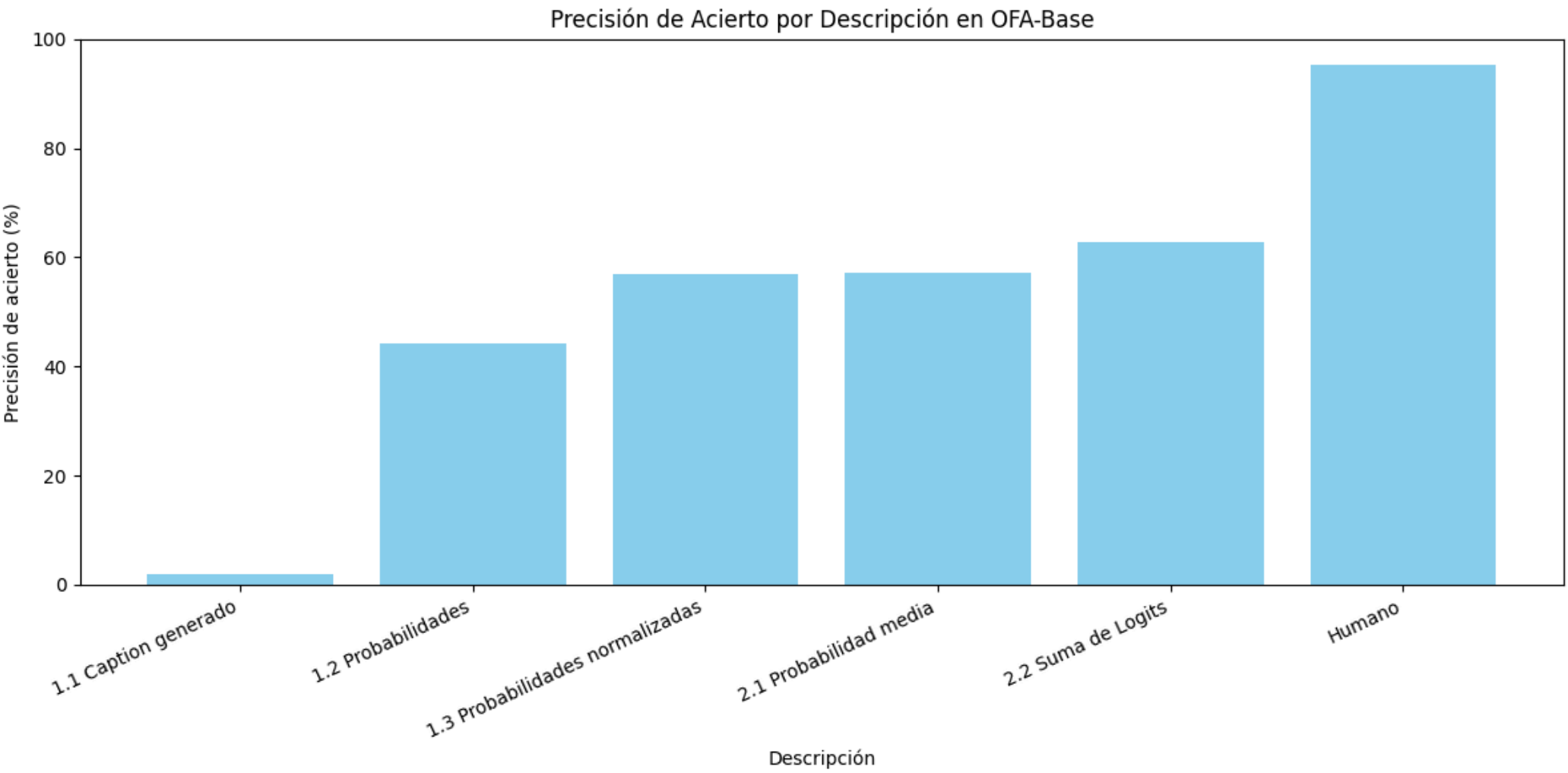
OBJETIVOS

Esta investigación tiene como objetivo principal explorar si los modelos multitarea, OFA, pueden adquirir conocimientos sobre las relaciones espaciales a través de las tareas de visión y texto en las que han sido preentrenadas.

En nuestra investigación, no nos centraremos en el rendimiento bruto del modelo, sino en cómo la forma de evaluación afecta a su desempeño. Esto es fundamental, ya que un modelo multimodal debería ser adaptable y capaz de generalizar su conocimiento a nuevas tareas relacionadas.

RESULTADOS

	BASE	LARGE	HUGE
ITM			
Respuesta generada	1,84%	0,41%	0,71%
Probabilidad	44,24%	49,92%	48,02%
Probabilidad normalizada	56,91%	57,43%	63,99%
GENERATOR			
Probabilidad media	57,25%	47,84%	46,84%
Suma logits 2	62,87%	61,15%	58,25%



CONCLUSIONES

- Gran disparidad en los resultados obtenidos mediante diferentes métodos de evaluación subraya la importancia de utilizar múltiples enfoques para evaluar las capacidades de un modelo. Confiar en un solo método puede dar una visión sesgada de las verdaderas capacidades del modelo.
- Aún están lejos de alcanzar el nivel de precisión humana (95.4%). Esto indica que los modelos actuales todavía tienen un largo camino por recorrer para poder adaptarse a tareas complejas como el razonamiento espacial.
- La incapacidad del modelo para mantener un rendimiento consistente a través de diferentes métodos de evaluación sugiere que los modelos multimodales como OFA aún enfrentan desafíos significativos en términos de adaptabilidad y generalización. La forma en que se introducen y evalúan los datos puede afectar drásticamente el rendimiento del modelo, lo cual es un problema importante que necesita ser abordado en futuras investigaciones.