

Automatización de la codificación en tareas de listado de propiedades mediante aprendizaje automático y modelos de lenguaje

Diego Ramos Álvarez

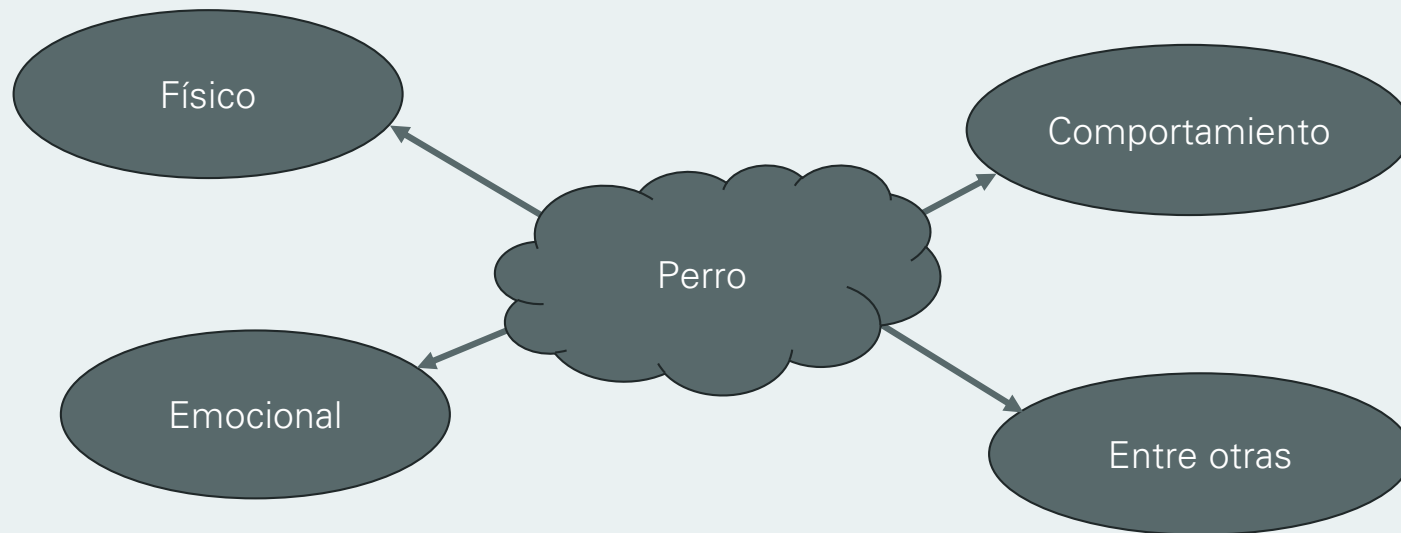
Profesor Guía: Sebastián Moreno
Profesor Co-Guía: Enrique Canessa

Introducción

Contexto del trabajo

Estudio de la Memoria Semántica (MS)

- El estudio de la MS busca entender como la memoria almacena conocimiento generales como palabras, conceptos y otros.



Estudio de la MS

- Para estudiar la MS es necesario recoger opiniones de varias personas.
- En general, los estudios piden que mencionen propiedades (palabras o frases) de un concepto en particular, por ejemplo "que se le viene a la mente cuando se le pregunta: ¿qué es un perro? ".
- El método mas relevante para esta propuesta de tesis es la Property listing task:

Property Listing Task (PLT)

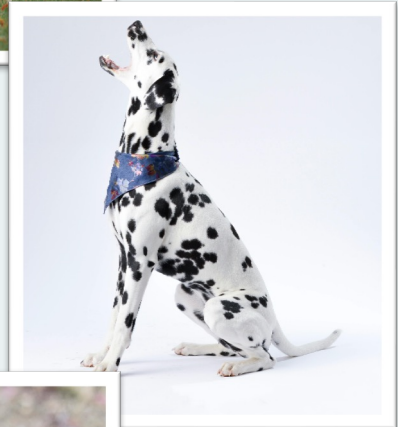
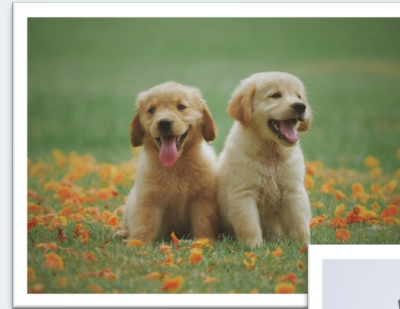
Se le pide a los participantes que mencionen todas las propiedades que asocien al concepto (sin límite de tiempo).

- Otros métodos para estudiar la MS son:

Semantic Fluency Task

Associative Word Task

Ejemplo PLT



Ejemplo PLT



Sujeto 1

Compañero de vida
Cuatro patas
Peludo
...



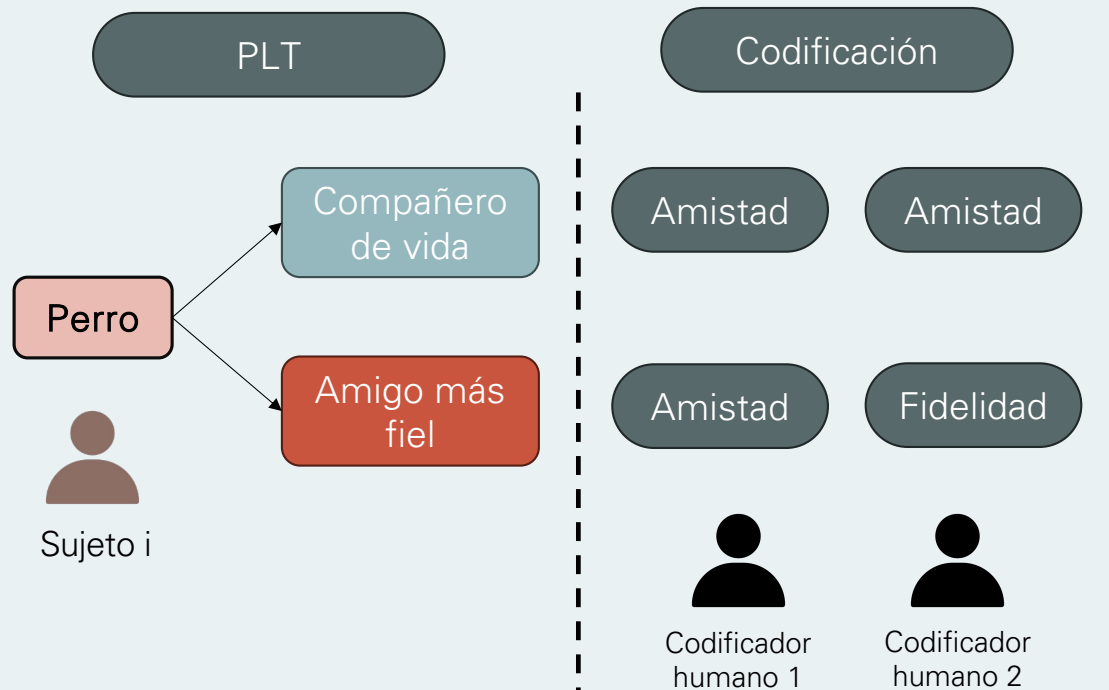
Sujeto 2

Amigo más fiel
Peludo
Ruidoso
...

Existen palabras que son similares y que se pueden agrupar como una sola, por ejemplo, Compañero de vida y amigo más fiel. **Por esto es necesario codificar las respuestas.**

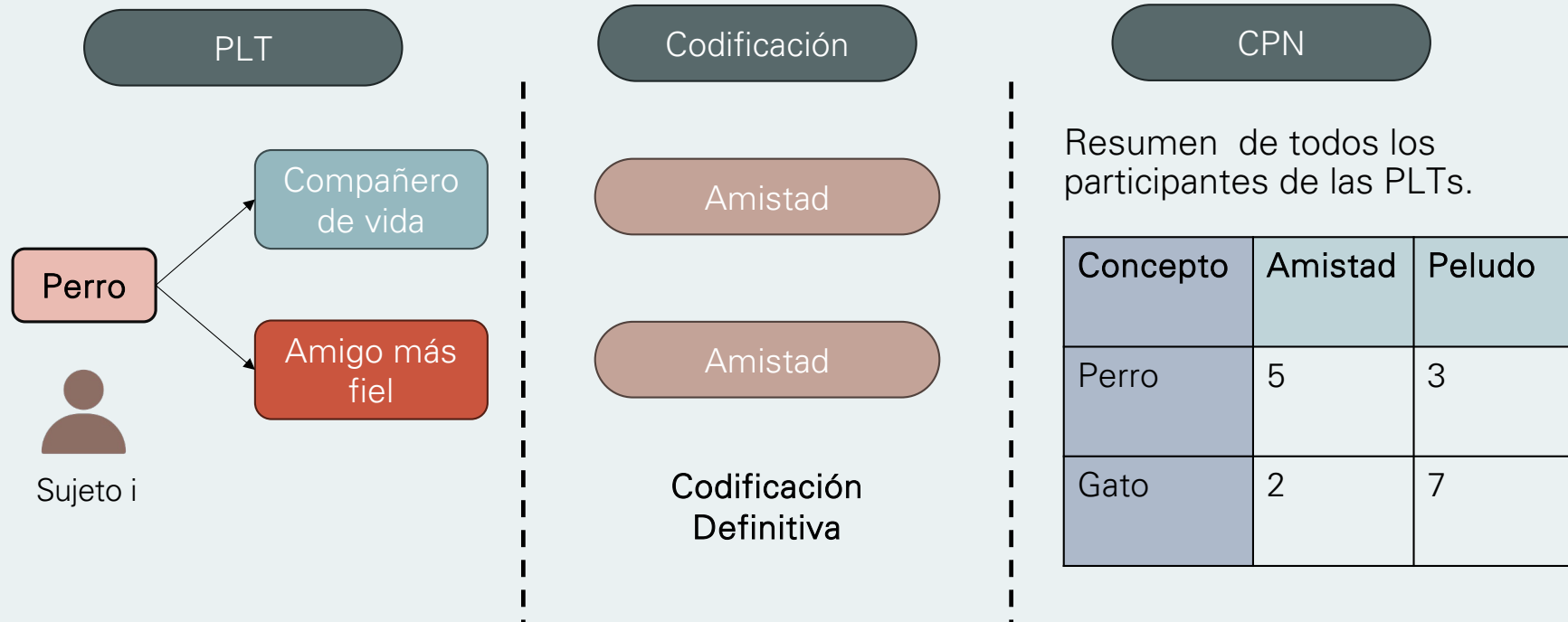
Ejemplo PLT, codificación

La codificación es el proceso de asignar un código único para cada una de las propiedades mencionada por los participantes, agrupando las propiedades que se refieren a la misma idea.



Ejemplo PLT, concept property norm (CPN)

Una vez realizada la codificación definitiva se pasa a una concept property norm (CPN), que es una matriz de frecuencia que se obtiene a partir de las propiedades mencionadas en la PLT.



Problemas de la codificación

- Tenemos dos grandes problemas:



Tiempo

- Un codificador humano entrenado se puede demorar meses en codificar un dataset.
- Un dataset con aproximadamente 4,900 propiedades, demoró cerca de **3 meses** en codificar.



Subjetividad

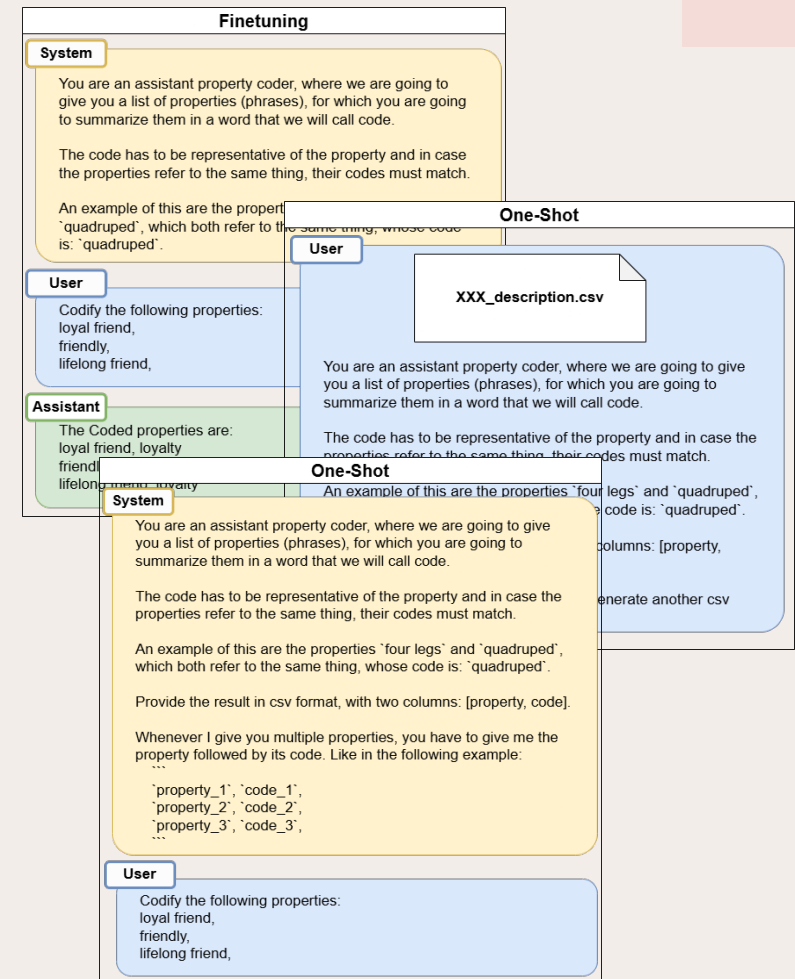
- Las reglas de codificación son interpretativas, la codificación de una propiedad puede variar mucho dependiendo del codificador.
- Ejemplo: "**Amigo más fiel**" se puede codificar como "**fidelidad**" o "**amistad**".

Puede un LLM hacer la codificación

Using Chat-GPT for coding properties in semantic memory studies. Accepted at ICPRS 2025

¿Que fue lo que hicimos?

- Utilizamos Chat-GPT siendo uno de los LLMs más populares.
- Probamos varios métodos:
 1. One-Shot encoding (GPT-3.5 y GPT-4o)
 2. Finetuning (GPT-3.5)
 3. Repetir las instrucciones humanas (GPT-5)



Datos utilizados

Trabajamos sobre dos datasets en **español**: CPN27 y CPN120.

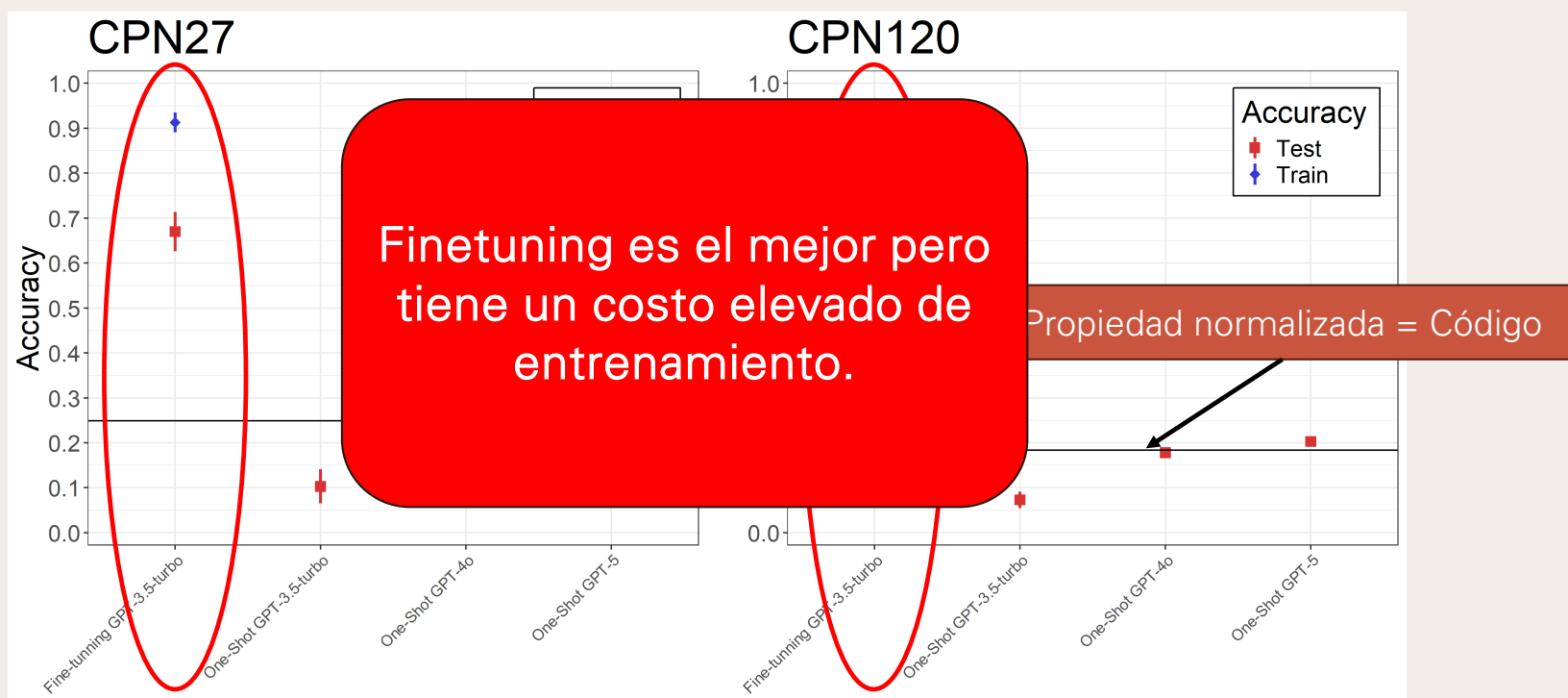
CPN27

- 27 Conceptos Concretos:
- 100 participantes
 - Promedio de edad 21.99
 - 4,941 Propiedades
 - 1,296 Codificaciones

CPN120

- 60 Conceptos Concretos y 60 Abstractos:
- 157 participantes
 - Promedio de edad 23.70
 - 27,138 Propiedades
 - 1,780 Codificaciones

Resultados de Chat-GPT



- Comparamos el accuracy de cada modelo para imitar al codificador humano.

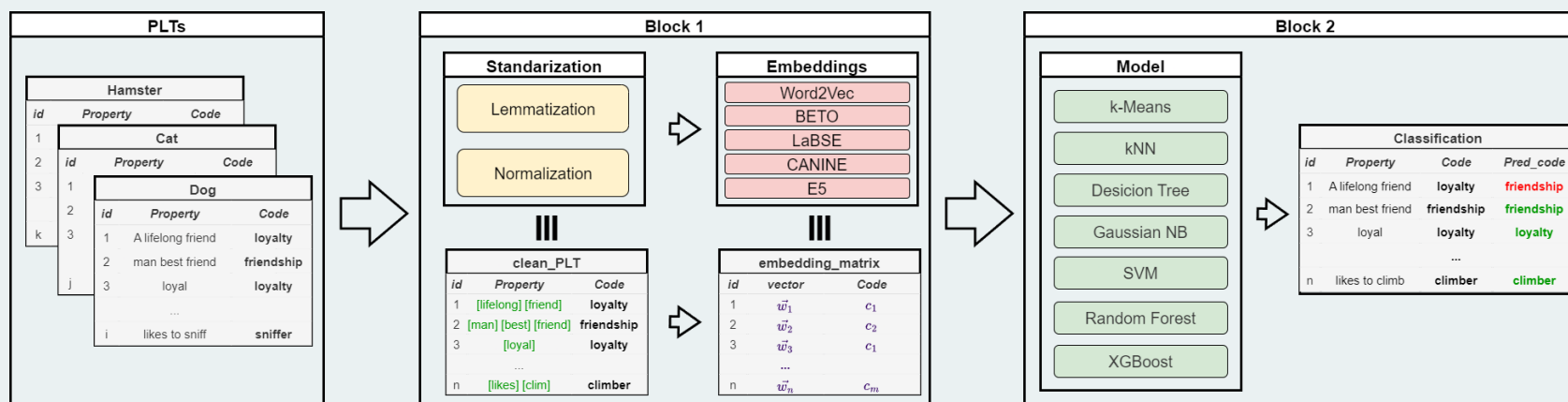
Puede un modelo de ML superar a Chat-GPT?

Ramos, D., Moreno, S., Canessa, E. *et al.* AC-PLT: An algorithm for computer-assisted coding of semantic property listing data. *Behav Res* 56, 3366–3379 (2024). <https://doi.org/10.3758/s13428-023-02260-9>

Ramos, D., Moreno, S., Canessa, E. *et al.* Towards scalable and reliable coding of semantic property norms: ChatGPT vs. an improved AC-PLT. *Behav Res* 57, 302 (2025). <https://doi.org/10.3758/s13428-025-02838-5>

¿Qué hicimos en este paper?

- Anteriormente creamos un framework llamado AC-PLT para codificación (Ramos et. al., 2024 y Ramos et. al., 2025)
- El framework se compone de 2 bloques:
 - Bloque 1: Estandarización y Embedding.
 - Bloque 2: Clasificación



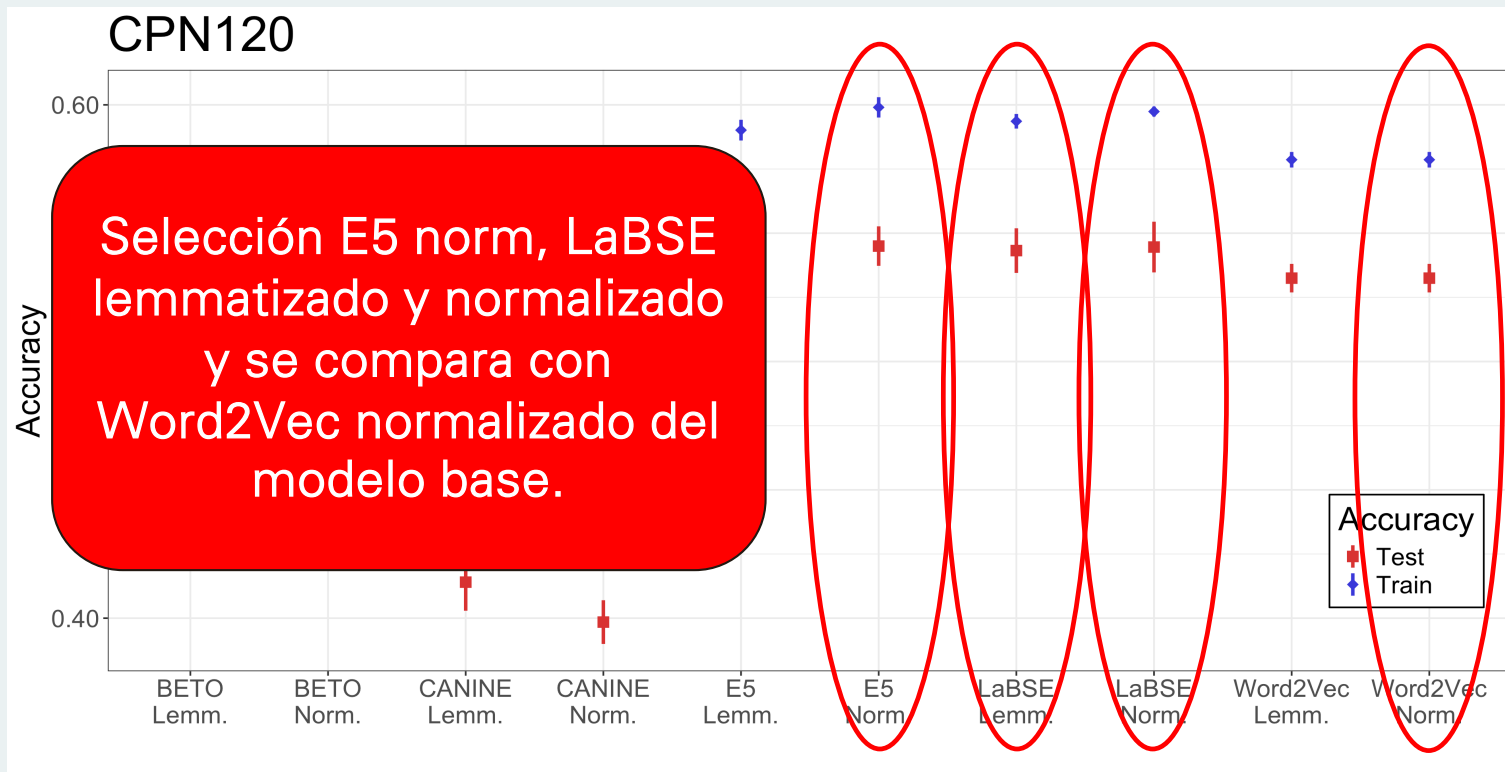
15

“Assisted Coding for Property Listing Task (AC-PLT)”

¿Qué hicimos en este paper?

- Originalmente AC-PLT, utiliza normalización, Word2Vec y k-Means, pero se realizó un ablation study y probamos varias combinaciones en cada bloque:
- Probamos hacer varias combinaciones en cada bloque:
 1. Combinaciones de estandarización (normalización y lematización) y de embedding (Word2Vec, BETO, LaBSE, CANINE y E5) con **k-Means** para clasificación.
 2. Se selecciona la mejor combinación del paso 1 y se analiza otros modelos de clasificación (**k-Means**, **kNN**, **naïve Bayes**, **SVM**, **random forest** y **XGBoost**).

Resultados Experimento 1 (CPN120)



- CPN27 obtiene un comportamiento similar.

Resultados Experimento 2 (CPN120)

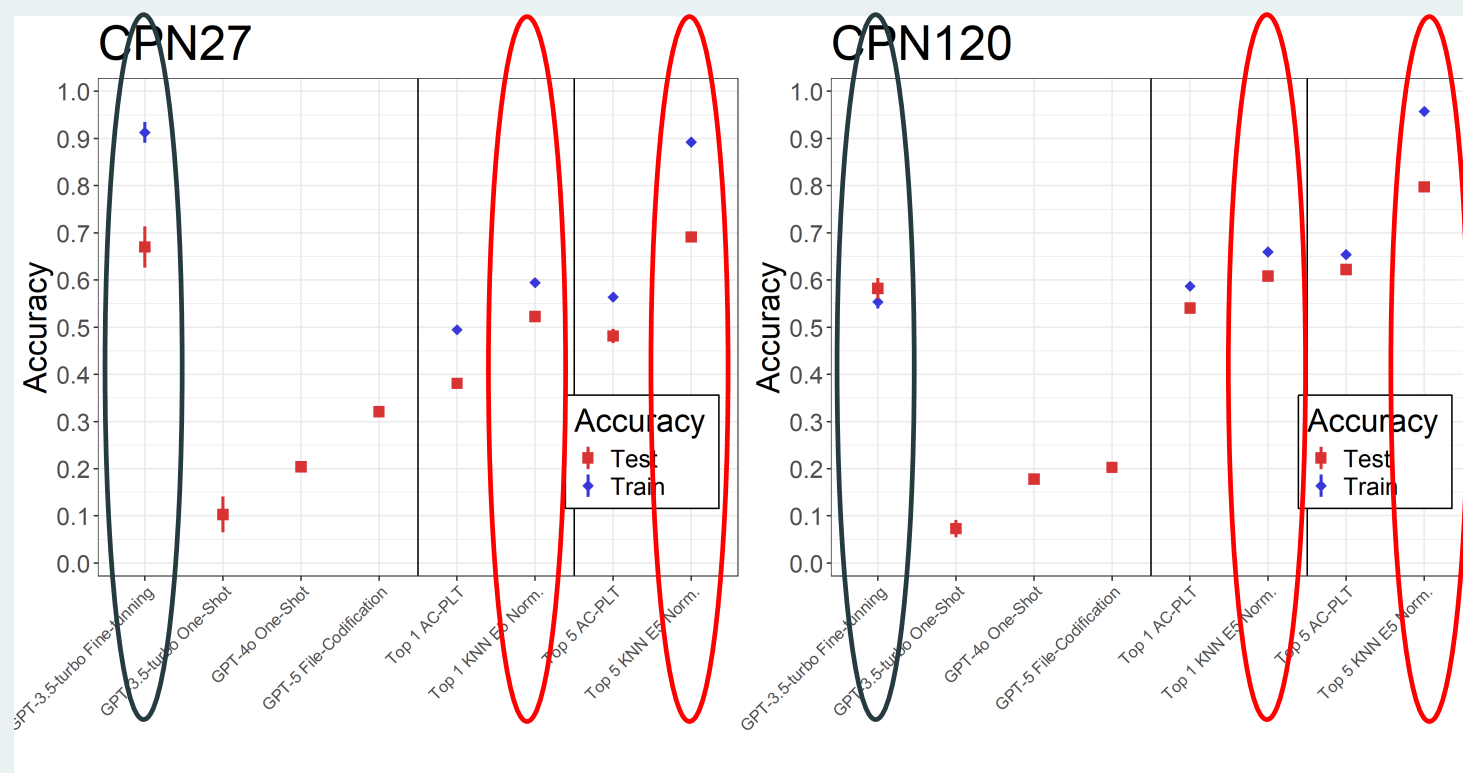


- CPN27 obtiene un comportamiento similar.

Resultados AC-PLT mejorado

- Se entrenó y evaluó el modelo con **5-Fold Cross Validation** y con **top 1 y 5 accuracy**.
- Dentro de los modelos E5 es el que tiene el mejor rendimiento con un $k=10$ para ambos casos:
 - CPN27 Top-1=0.523 y Top-5=0.691
 - CPN120 Top-1=0.608 y Top-5=0.797
- E5 con kNN fue el más consistente para ambos datasets y en ambas métricas.
- Ahora podemos comparar los LLM con el AC-PLT original y Mejorado.

Comparación con Chat-GPT



Conclusiones y Trabajos Futuros

Conclusiones

- A pesar de los avances, no podemos replicar todavía lo que es un codificador humano.
- Solo usamos datos en español, ¿qué pasará en otros idiomas?
- ¿Qué tan bien generaliza los datos un LLM?
- ¿Hay otro modelo de LLM que pueda tener mejores resultados?
- ¿Los LLM realmente entienden el lenguaje humano?
- ¿Los codificadores humanos son realmente el *gold standard*?

Gracias por su
tiempo