

TRABAJO PRÁCTICO: DECISIÓN TRANSFORMER PARA RECOMENDACIONES

Grupo 19 -Bobbiesi, Sofía - Maurino, Micaela

OP8: Aprendizaje por Refuerzos - Diplomatura en Ciencia de Datos

1.Introducción

En este trabajo analizamos cómo un modelo llamado DT puede aprender a recomendar ítems (productos, películas, artículos, etc.) basándose en el historial de cada usuario.

Para estudiarlo, tratamos el problema como si fuera un juego por etapas, donde el sistema observa lo que el usuario hizo antes, decide qué recomendar después, y recibe una “puntuación” según si la recomendación fue útil o no.

Los objetivos principales fueron:

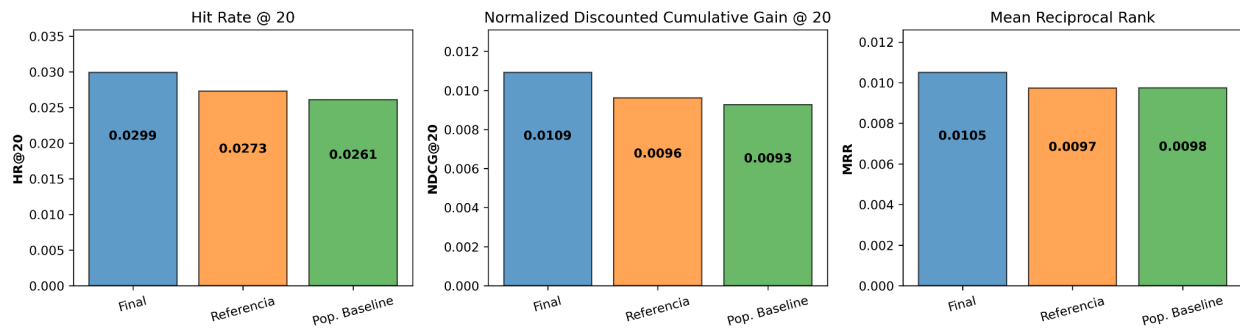
- Construir el modelo desde cero
- Entrenarlo correctamente
- Compararlo contra métodos simples como recomendar lo más popular
- Evaluar si realmente aprende a recomendar mejor

2. Dataset y Preprocesamiento

Usamos un conjunto de datos con interacciones reales entre usuarios e ítems (por ejemplo, clics o selecciones). Antes de entrenar el modelo hubo que:

- Ordenar los datos por tiempo
- Eliminar interacciones sueltas o poco útiles
- Dividir las secuencias muy largas
- Procesar los valores de recompensa para que el modelo los pueda aprender
- También analizamos estadísticas básicas como la cantidad de usuarios, los ítems existentes y las interacciones por usuario.

- El MDP se formula considerando el historial del usuario como estado, las recomendaciones como acciones, y las interacciones (clics) como recompensas.



3. Implementación

El DT es un modelo simple que aprende por secuencia de acciones, que en este caso son ejecutadas por las recomendaciones. Este modelo lee el historial del usuario, compara la experiencia previa y aprende patrones para futuras recomendaciones.

Para entrenarlo se usó un optimizador moderno que ajusta los parámetros automáticamente, y se eligieron valores de entrenamiento basados en trabajos previos. En este proceso se presentaron algunos desafíos, como por ejemplo, evitar que el modelo memorice los datos.

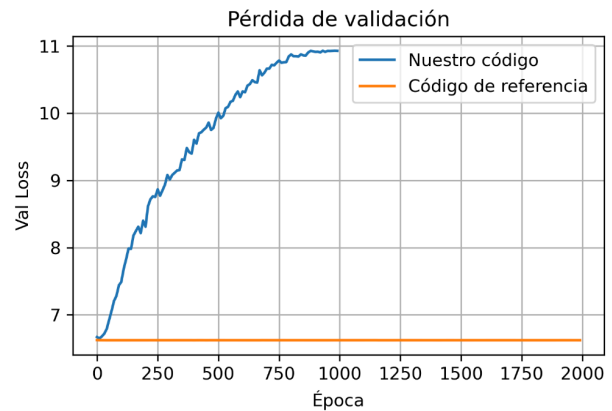
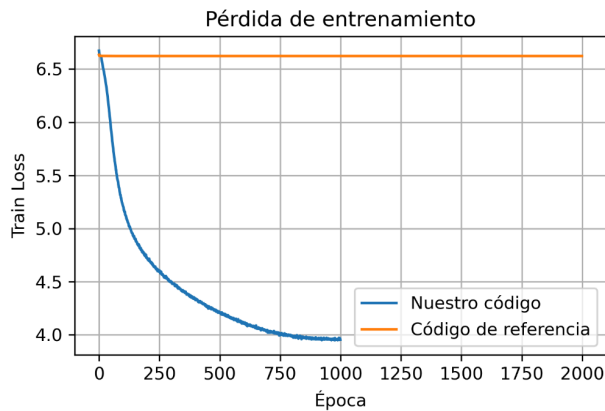
4. Resultados

Los resultados muestran que el modelo **Final (Decision Transformer)** tiene mejores números que los modelos simples, especialmente cuando las recomendaciones son más difíciles.

A demás, las métricas principales (HR@20, NDCG@20, MRR) indican que el DT recomienda más ítems “correctos” y en posiciones más altas.

Al mismo tiempo se evaluó Return conditioning (cómo cambia el rendimiento según la “meta” que se le pide al modelo) y **Cold-start** (qué tan bien recomienda a usuarios con poca información)

1	Modelo	Recursos humanos a 5	NDCG@5	Recursos humanos a 10	NDCG@10	Recursos humanos a 20 años	NDCG@20	MRR
2	Final	0.008766334250343878	0.0050865748177245365	0.016509628610729025	0.0075757337736665304	0.029924346629986245	0.010918955731486192	0.010501870247359463
3	Referencia	0.006830601092896175	0.0040279496990429245	0.01366120218579235	0.006207048457339813	0.0273224043715847	0.00961785290858459	0.009725504096144495
4	Línea base de población	0.0068775790921595595	0.004115932420520283	0.012379642365887207	0.005851512731215769	0.026134800550206328	0.009266422636801804	0.009751004152887259



5. Análisis y Discusión

El modelo fue aprendiendo patrones como qué ítems suelen aparecer consecutivamente, como es el historial que se indica en cada usuario activo y cómo responder cuando el usuario tiene interacciones previas.

¿Cuándo el modelo funciona bien y cuando falla? ¿Cuáles son sus limitaciones y mejoras?

- Cuando tiene suficiente historial ✓
- Cuando la recompensa es clara y consistente ✓
- En usuarios totalmente nuevos (sin historial) ✗
- Con datos muy ruidosos o poco variados ✗
- Consume más recursos que modelos simples y requiere muchos datos ⚠
- Puede sobre ajustarse si no se entrena correctamente ⚠
- Agregar información adicional (categorías, texto) ↑
- Usar pre-entrenamiento en secuencias más grandes ↑
- Implementar regularización para evitar sobreajuste ↑

5. Conclusiones

En este trabajo analizamos cómo el DT puede utilizarse para mejorar sistemas de recomendación aprovechando el historial de interacción de los usuarios. Observamos que el modelo es capaz de aprender patrones temporales y de comportamiento, lo que le permite generar recomendaciones más precisas que métodos tradicionales como la popularidad o los modelos de referencia básicos.

Además, el uso del “return conditioning” demostró ser una herramienta útil para guiar al modelo hacia mejores decisiones, especialmente en situaciones donde el historial del usuario es limitado. A pesar de sus ventajas, el DT presenta desafíos como su mayor demanda computacional y la necesidad de grandes cantidades de datos para rendir al máximo. Sin embargo, sus resultados muestran un potencial claro para aplicaciones reales de recomendación.

En adelante, trabajos futuros podrían incluir más información contextual, explorar variantes del Transformer o aplicar técnicas de pre entrenamiento para mejorar su rendimiento y eficiencia. En conjunto, la experiencia evidencia que este enfoque no solo es prometedor, sino también una evolución natural hacia modelos de recomendación más inteligentes y adaptativos.