

I.A. para la empresa

Práctica final

10 de Junio de 2024

IMF
Smart Education



1.- ENUNCIADO

La práctica final consiste en crear un sistema de recomendación de libros. Para ellos se va a utilizar el conjunto de datos que aparece dentro de la carpeta data.

Esta carpeta está formada por tres archivos:

- Rating.csv. Donde tenemos los siguientes campos:
 - *User-ID*: Identificador de usuario de la aplicación de venta de libros.
 - *ISBN*: ISBN del libro
 - *Book-Rating*: Rating que ha dado el usuario a dicho libro.
- Users.csv. Donde tenemos los siguientes campos:
 - *User-ID*: Identificador de usuario de la aplicación de venta de libros.
 - *Location*: Información geográfica del usuario
 - *Age*: Edad del usuario.
- Books.csv. Donde tenemos los siguientes campos:
 - *ISBN*: ISBN
 - *Book-Title*: Título del libro
 - *Book-Author*: Autor
 - *Year-Of-Publication*: Año de publicación
 - *Publisher*: Editor
 - *Image-URL-S*: Imagen de la portada en tamaño pequeño
 - *Image-URL-M*: Imagen de la portada en tamaño mediano
 - *Image-URL-L*: Imagen de la portada en tamaño grande.

2.- PREGUNTAS

Para la ejecución de la práctica se pide realizar las siguientes tareas:

1. Crear un filtro colaborativo sencillo basado en memoria. Se puede elegir el que prefiera, item-to-item o user-to-user. Se puede realizar ajuste de hiper-parámetros intentando probar con diferentes variaciones de medida de similitud o forma de ponderar el valor estimado de los ratings. Este modelo nos debería servir como modelo base de partida. **(2 ptos)**
2. Crea una matriz de factorización construida mediante una red neuronal. Se puede hacer ajuste de hiper-parámetros para optimizar la métrica de salida. Elige la métrica y función de pérdida adecuada para entrenar esta red. **(2 ptos)**
3. Prueba a crear un recomendador basado en contenido y en perfil de cliente, sin hacer uso de filtros colaborativos. Utiliza también una red neuronal optimizada mediante hiper-parámetro. **(2 ptos)**
4. Añade al recomendador anterior (si no lo has hecho ya) features basadas en texto para poder extraer información del título del libro. Se recomienda el uso de algún paquete preentrenado de hugging-fase (no hace falta hacer fine-tuning). Elige el o los modelos que consideres más adecuado. **(2 ptos)**
5. Construye un método híbrido que combine todos los métodos (salvo el 1) en una única red neuronal. Prueba a hacer varias combinaciones a ver cual da mejor resultado. Entre ellas se recomienda:
 1. Combinación básica
 2. Deepfm
 3. Two towers.
 4. Deep and wide models.
 5. Cualquier otra combinación que se te ocurra.

¿Qué métodos da mejores resultados. Explica las conclusiones y resultados de tus modelos.? **(2 ptos)**

Nota: Se recomienda usar las buenas prácticas de M.L (Entrenar con conjunto de entrenamiento. Optimización en validación a ser posible con cross-validation, y evaluación final sólo en grupo de test.

2.- PREGUNTAS

Punto extra: Se puede utilizar información de repositorios externos para enriquecer las features de los libros con información como número de página, géneros, descripción o primera frase del libro. Una buena fuente externa es (<https://openlibrary.org/developers/dumps>) . Se recomienda hacer un download de algunas de las base de datos (p.ejemplo work) en vez de usar la API y luego hacer un merge con los datos de nuestra base de datos para extraer solo la información necesaria.

Se valorará con un punto extra aquellos alumnos que utilicen información exterior para enriquecer la base de datos. Sobre todo información en texto libre.

IMF

Smart Education