



UNIVERSIDAD DE BUENOS AIRES
FACULTAD DE CIENCIAS EXACTAS Y NATURALES
FACULTAD DE INGENIERÍA

Sistemas de recomendación colaborativos e híbridos

Trabajo Final de Especialización en Explotación de Datos y Descubrimiento del
Conocimiento

Adrian Norberto Marino

Buenos Aires, 2022

RESUMEN

Acá va un resumen del trabajo. Con el resumen se debería poder tener una idea del trabajo en su totalidad. Desde los objetivos y los datos hasta los resultados y conclusiones. Se escribe al final. (aprox. 200 palabras).

Palabras claves: representativas del trabajo, los métodos, los datos (no menos de 5).

Índice general

1..	Introducción	1
1.1.	Tipos de sistemas de recomendación	1
1.2.	Descripción del problema y motivación ($> 1/2$ carilla)	2
1.3.	Trabajos previos (> 1 carilla)	2
1.4.	Objetivos ($> 1/2$ carilla)	2
2..	Materiales y Métodos	4
2.1.	Datos (> 1 carilla)	4
2.2.	Análisis exploratorio	4
3..	Métodos (> 3 carillas)	5
3.1.	Método I	5
3.2.	Método II	5
3.3.	Métricas	5
4..	Experimentos (si los hay)	6
5..	Resultados (> 3 carillas)	7
6..	Conclusiones (> 1 carilla)	8

1. INTRODUCCIÓN

Los sistemas de recomendación tienen por objetivo acercar a sus usuarios información, productos, contenido, etc.. relevantes a sus gustos o necesidades, permitiendo a estos encontrar con mayor facilidad lo que están buscando. Una de las ideas principales es simplificar la búsqueda de información (Textos, videos, productos, comidas, etc..). Formalizando esta definición podemos decir que: Los sistemas de recomendación apuntan a ayudar a un usuario o grupo de usuarios a seleccionar items de forma personalizada dado un conjunto de items de gran extensión o un gran espacio de búsqueda.

Este objetivo puede cambiar según el contexto de cada negocio. Para un e-commerce de delivery de comidas, el objetivo es acercar a los usuarios el tipo de comida que quieren probar en ese mismo momento, a un precio que puedan pagar con tiempo de entrega aceptable. Para un e-commerce de venta de productos, se busca acercar al usuario aquellos productos que este necesitando en ese mismo momento, los cuales tienen un precio que el mismo puede pagar y por otro lado, asegurar una experiencia satisfactoria con el vendedor. En el negocio de visualización de contenido (Ya sea audio o video), el objetivo es acercar al usuario contenido a fin a sus gustos para mejorar su experiencia en la plataforma y así aumentar el engagement de sus usuarios.

Por otro lado, el objetivo de fondo siempre es el mismo, mejorar la conversión. Con esto nos referimos a aumentar el volumen de ventas para un e-commerce de venta de productos, la cantidad de deliveries mensuales, la cantidad de impresiones de publicidad en aplicaciones de visualización de contenido, aumentar el tiempo de permanencia en las plataformas de streaming de audio o video, etc.. Podemos encontrar muchos ejemplos distintos donde el objetivo común es mejorar la conversión y engagement de los usuarios.

Desde un punto de vista mas técnico, los sistemas de recomendación se utilizan para predecir el grado de preferencia de un usuario con respecto a un item. En general, se puede lograr aplicando de un algoritmo de optimización, el cual minimiza la diferencia entre el grado de preferencia esperado versus real. Otros enfoques hacen uso de medidas de distancia para establecer este grado de preferencia.

1.1. Tipos de sistemas de recomendación

Al menos existe dos formas de clasificar a los sistemas de recomendación:

La clasificación basada en contenido toma en cuenta los atributos de items (por ejemplo, las palabras de un libro, género, etc) , y agrupa con medidas de similitud entre ellos, asumiendo que un usuario desea consumir items parecidos. Finalmente, el filtrado colaborativo utiliza perfiles de usuarios y/o items en base a los datos para recomendar basado en medidas de distancia

- Considerando los datos a utilizar: Estos dependen de que datos utilicemos para construirlos.
 - Basados en reglas: Son modelos que hallan un árbol de decisión en base a los datos. Estos árboles podemos verlos como las reglas a utilizar para realizar una recomendación. Como principal ventaja, tienen el mayor grado de explicabilidad posible, ya que a simple vista se puede apreciar las reglas que el modelo ha

encontrado en base a los datos. Como ejemplo, una posible regla hallada podría ser: Si el usuario 1 compro el item A, entonces comprara el item B.

- Basados en contenido: Estos modelos utilizan los atributos o features de los items a recomendar. Por ejemplo, si estamos recomendando películas, los atributos a elegir podrían ser: título, género, director, actores de reparto, fecha de estreno, idioma original, etc... Estos modelos en general codifican estos features usando distintas técnicas para codificar los features (One-hot encoding, embeddings, TF-IDF, etc..), en algunas casos buscan bajar la dimensionalidad de los mismos sin perder información. De esta forma, generan vectores que representan a cada item y finalmente usando alguna medida de distancia establecen el grado de cercanía entre estos vectores. Finalmente, resuelve problemas del tipo: dato un item cuales son los items mas parecidos o cercanos.
 - Basados en filtros colaborativos
- Considerando el modelo: Esta referido a la forma en la que se realiza la predicción.
- Basados en memoria: Este tipo de modelos mantiene sus datos en memoria y recorrer todos los datos(full scan) cada vez que necesita realizar un inferencia o predicción. Un ejemplo es el algoritmo de k vecinos cercanos (KNN), el cual mantiene una matriz rara de distancias en memoria, la cual recorre completamente para compara las distancias entre filas o columnas usando alguna media de distancia como puede ser la Distancia Coseno, Distancia Coseno Ajustada, Distancia de Manhattan, etc...
 - Basados en modelos: Estos están constituidos por un modelo basados en los datos el cual se reutiliza cada vez que se desea realizar un inferencia o predicción (Por ejemplo: Single Value Descomposition)

1.2. Descripción del problema y motivación (> 1/2 carilla)

Ejemplo de ecuación:

$$L = - \sum_{p \in P} \sum_{y \in Y} w_{py} \sum_{t \in T_{py}} y_t \log \hat{y}_t + (1 - y_t) \log(1 - \hat{y}_t) \quad (1.1)$$

1.3. Trabajos previos (> 1 carilla)

En la imagen 1.1 hay un ejemplo de como poner una figura:

1.4. Objetivos (> 1/2 carilla)



Fig. 1.1: Acá va una explicación de la figura. Recuerden usar colores solo si son explicativos, no olvidar nombrar correctamente los ejes y poner una explicación en el caption.

2. MATERIALES Y MÉTODOS

2.1. Datos (> 1 carilla)

Descripción de los datos / obligatorio poner la fuente.

2.2. Análisis exploratorio

3. MÉTODOS (> 3 CARILLAS)

Describir los métodos utilizados (a utilizar). Es bligatorio agregar un parrafo especificando a que materias se relacionan.

3.1. Método I

3.2. Método II

3.3. Métricas

4. EXPERIMENTOS (SI LOS HAY)

Pueden ser comparaciones entre los métodos utilizados, por ejemplo.
En la 4.1 se ve un ejemplo de tabla:

Phone	Total	% Errors	F1	EER
EY	441	13.83	0.80	0.12
JH	178	39.89	0.81	0.18
AY	1040	5.96	0.45	0.21
R	1298	18.34	0.53	0.27

Tab. 4.1: Describir la tabla como para que alguien que agarra el trabajo no dependa del cuerpo del texto para entenderla

5. RESULTADOS (> 3 CARILLAS)

Pueden ser preliminares. Es bligatorio tener algo hecho.

6. CONCLUSIONES (> 1 CARILLA)

Pueden ser preliminares. Si no hay mucho hecho se pueden discutir las dificultades a futuro.