

Tareas y ejercicios del Tema 2: Sistemas de Recomendación

Objetivos

1. *Extraer información de datasets basados en elección*
2. *Generar recomendaciones basadas en contenido*
3. *Extraer información de datasets basados en ratings*
4. *Realizar recomendaciones con algoritmos colaborativos*
5. *Evaluar recomendadores y comparar entre diferentes algoritmos*
6. *Utilizar R para analizar y evaluar algoritmos de recomendación.*

Tareas y ejercicios

1. Eres el Científico de Datos de una importante empresa mayorista del sector turístico. La empresa tiene como negocio la creación de paquetes turísticos que luego vende a minoristas, agencias de viaje tanto online como offline. La empresa quiere generar recomendaciones de los paquetes, que se caracterizan por el siguiente conjunto de atributos y valores:
 - a. Calidad del paquete (Num. Estrellas del hotel): 2,3,4, y 5.
 - b. Precio: Valores con rango desde 120 a 870 euros.
 - c. Duración del paquete: 5 o 7.
 - d. Atractivo visual: Alto o Bajo

Se realizó un experimento con una muestra de potenciales clientes. Se mostraron 8 situaciones de elección con 3 paquetes/alternativas en cada elección. El usuario tuvo que realizar una elección para cada situación. Los resultados se recopilaron en el dataset “dataset_eleccion_hotel.xls”. Se pide lo siguiente:

- Calcula el vector de preferencias de los usuarios 1 y 4.
- Suponte que ahora esas 2 personas están buscando un hotel en Santiago. Consulta Booking para conocer las alternativas disponibles. Después, asume el rol de recomendador, y a partir del vector de preferencias aprendido en el punto anterior, selecciona un par de paquetes turísticos a partir del conjunto de alternativas disponibles para recomendarlos.
- Selecciona ahora a un compañero de clase. Como no tienes su perfil de preferencias, tendrás que asignarle uno de los 2 perfiles creados anteriormente. Si a través de un típico proceso de creación de una cuenta online pudieras hacerle 3 preguntas, ¿Cuáles le harías para saber qué perfil le asignarías?
- Asigne uno de los dos perfiles utilizando el método que hayas pensado en el punto anterior, y preséntale los paquetes que diseñaste para recomendar. Pídele que evalúe de 0 a 5 la satisfacción con la recomendación.

2. Ahora toca analizar el histórico de compras del dataset “MO_dataset.compras.xls”. Se pide lo siguiente:
 - Calcula el vector de preferencias de los usuarios 130 y 160.
 - A partir de los vectores anteriores, para cada usuario, crea una oferta de descuento personalizada.
 - A partir de las preferencias aprendidas para estos dos usuarios, y de los datos que observas en el dataset: ¿qué recomendaciones harías a la gerencia del supermercado para organizar la ubicación de los productos?
3. Instalación de R y la librería “recommenderlab”.
 - Instala el entorno de análisis estadístico R. Enlace: <https://cran.r-project.org/>
 - Instala la librería “recommenderlab”. Hay dos opciones para hacerlo. Primera: (1) Selecciona un espejo CRAN en la opción “Paquetes/Seleccionar Espejo Cran”, y (2) Seleccionar instalar la librería “recommenderlab” en la opción “Paquetes/Instalar paquetes”. Segunda: (1) Descarga el fichero ZIP que contiene la librería desde el enlace <https://cran.r-project.org/web/packages/recommenderlab/index.html>, (2) Instala el paquete desde R con la opción “Paquetes/Instalar paquetes desde archivos zip locales”.
 - Carga la librería ejecutando el comando: `library("recommenderlab")`.
 - Instala las librerías adicionales que indica R de las que depende “recommenderlab”
4. Tutorial de “recommenderlab” con el dataset *Jester5k*.
 - Descarga el tutorial sobre la librería “recommenderlab”. Enlace: <https://cran.r-project.org/web/packages/recommenderlab/vignettes/recommenderlab.pdf>
 - Sección 5.4. Inspecciona las propiedades del dataset *Jester5k*.
 - Crea una imagen de las 100 primeras filas y columnas de la matriz con el comando: `image(Jester5k[1:100,1:100])`. ¿Qué usuarios son los que aportan más ratings? ¿Qué chistes han sido más valorados?
 - Sección 5.5. Crea un recomendador siguiendo las indicaciones del tutorial. Crea otra versión del recomendador sólo con los 100 primeros usuarios y vuelve a predecir para los usuarios 1001 y 1002. ¿Detectas diferencias en las recomendaciones?
 - Secciones 5.6 y 5.7. Evalúa las recomendaciones siguiendo el tutorial. ¿Qué diferencia hay entre la evaluación de la sección 5.6 y la 5.7? ¿Cuál te parece más apropiada para un sistema de recomendación?
 - Sección 5.8. Compara los diferentes métodos de recomendación en esta sección. ¿Cómo quedaría el algoritmo de recomendación ítem-based en comparación con los utilizados en esta sección? ¿Cambian los resultados si modificamos el particionamiento utilizado en el procedimiento de evaluación?
5. Recomendación de películas con el dataset *Movielense*.
 - Lee la descripción del dataset Movielens en la descripción de la librería “recommenderlab”(Pg. 14). Enlace: <https://cran.r-project.org/web/packages/recommenderlab/recommenderlab.pdf>

- Carga ahora el dataset Movielense: *data(MovieLense)*
- Inspecciona las propiedades de este dataset. ¿Qué diferencias encuentras respecto al dataset Jester5k?
- Evalúa y compara los resultados de cuatro algoritmos de recomendación en este dataset: random, popular, user-based y ítem-based.