

Tarea 1

Profesor Denis Parra
Ayudante Hans Findel
19 de Agosto de 2014

Indicaciones

- Fecha de Entrega: 7 de Septiembre de 2014 a las 23:59
 - Debes entregar la tarea en tu repositorio git privado, que debes notificar al equipo docente.
 - Cada hora de atraso descuenta 1.5 puntos de la nota que obtengas.
 - La tarea es *individual*. La copia será sancionada con una nota 1.1 en el la tarea, además de las sanciones disciplinarias correspondientes.
-

Objetivos

Los principales objetivos de esta tarea es que aprendas a:

- Utilizar distintas técnicas propias de los sistemas recomendadores.
- Predecir una evaluación.
- Elaborar una lista de recomendaciones.

Adicionalmente se espera que seas capaz de:

- Evaluar distintas métricas
- Realizar una comparación entre los métodos

Instrucciones

La tarea tiene dos partes. En ambas partes debe utilizar y comparar los algoritmos que se mencionan en la sección de *Algoritmos a utilizar*. Adicionalmente para cada una de éstas se entregará un archivo con la información específica de los usuarios sobre los que se desea realizar la tarea.

- La primera parte consiste en realizar recomendaciones específicas a ciertos ítems para cada usuario ¹, más específicamente, predecir ratings.
- La segunda parte tiene por objetivo realizar una lista con 10 (diez) recomendaciones para cada uno de los usuarios presentes en la lista.

Algoritmos a utilizar

- *Popularidad*
- *User-based collaborative filtering*
- *Item-based collaborative filtering*
- *Slope one*

Datos base

Para realizar la tarea debes descargar el archivo comprimido de la primera tarea que se encuentran en el siding. Al descomprimirlo encontrarás:

- *ratings.csv*: Es un listado con todos los ratings de distintos usuarios sobre un set de ítems. En cada línea del archivo encontrarás tres valores, que corresponden al id de usuario, id del ítem y su evaluación (*rating*) respectivamente.
- *prediction.csv*: Es un listado en el que se solicita una predicción para algunos ítems para algunos usuarios. El archivo tiene una lista de ids de usuarios con los respectivos id de los ítems anidados. ²
- *list-top-n.txt*: Es un listado de ids de usuarios para los cuales tendrás que elaborar una lista de recomendaciones.

¹Los ítems pueden cambiar entre usuarios tanto en su id como en la cantidad de los mismos.

² Ejemplo de tupla en *prediction.txt*:

user-id, item-id, rating

Archivo que debes entregar

- Breve análisis de los datos entregados ³
- Implementación de los algoritmos ⁴
- Resultados para cada parte en archivos similares a los entregados ^{5, 6}.
- Análisis/comparación de los métodos utilizados ⁷

Metodología de corrección

Las entregas serán evaluadas por un software, por lo que el formato de la entrega debes seguirlo con exactitud. Cada parte de la entrega tiene una métrica de evaluación distinta, las cuales se detallan a continuación:

Parte 1

Dado que solo se entregan valores, se medirá el $RMSE = \sqrt{\frac{\sum_i^n (r_{iu} - \hat{r}_{ui})^2}{n}}$ (Pueden entregar valores reales.) En base a este puntaje, se ordenarán las entregas en un ranking.

Parte 2

Para la evaluación de esta parte de la entrega, se medirán las respuestas bajo la siguiente lista de métricas:

- Precision@10 = Recall@10 ⁸
- MAP (mean average precision)
- nDCG

Luego, con estos valores se obtendrá un promedio con el cual se realizará una evaluación de las entregas.

Habrà un punto adicional para el alumno con el mejor desempeño en cada parte de la tarea.

³Considera incluir tabla con estadísticas de usuarios, ítems y densidad del dataset. Agrega gráficos de distribución.

⁴Incluye las referencias de sus implementaciones.

⁵Para la primera parte debes reemplazar los 0 (cero) por el rating que corresponda.

⁶Para la segunda debes incluir el listado de los ids de los ítems para cada id de usuario.

⁷Respecto a su implementación, tiempo de ejecución, procesamiento y memoria requeridos

⁸Paolo Cremonesi, Yehuda Koren, and Roberto Turrin. 2010. Performance of recommender algorithms on top-n recommendation tasks. <http://doi.acm.org/10.1145/1864708.1864721>