



PONTIFICIA UNIVERSIDAD CATÓLICA DE
CHILE

PROYECTO SISTEMAS RECOMENDADORES

Recomendación de items basado en ratings que no faltan aleatoriamente

Jacques Hasard Romano Fenzo

Octubre 2018

1. Contexto

En el curso utilizamos algoritmos como Collaborative Filtering en *datasets* que relacionan calificaciones hechas por usuarios sobre *items* cualquiera. Sin embargo, estos algoritmos de recomendación y las pruebas de desempeño como MAE o RMSE asumen que los datos faltantes, es decir, *items* que un usuario no ha calificado, faltan de manera aleatoria.¹ Lo anterior quiere decir que la probabilidad de que un usuario califique o no un *item*, no depende de la percepción que el usuario tiene sobre ese *item*.

Para dejarlo más claro, el supuesto anterior (MAE - Missing At Random) establece que un usuario, ya sea odie, le guste o sea neutral a un item en específico, tendrá la misma probabilidad de calificarlo, lo cual resulta ser un supuesto bastante fuerte.

2. Motivación

Comprobar la mejora en las recomendaciones al utilizar un modelo probabilístico por sobre los algoritmos vistos en clase, especialmente Collaborative Filtering.

3. Objetivos

Implementar algún modelo probabilístico descrito en la literatura para considerar aquella información que con Collaborative Filtering no se está analizando, y que podría mejorar la calidad de las recomendaciones.

4. Solución propuesta

Primero utilizar un *dataset* de calificaciones por usuarios (que no cumple MAR) para entrenar el modelo probabilístico. Luego verificar el desempeño del modelo utilizando un *dataset* de prueba que **si** contenga datos que cumplan con MAR y que se relacionen con los datos utilizados para entrenar.

Al realizar las pruebas con un *dataset* que si cumple con MAR, se está probando la capacidad del modelo de generar recomendaciones considerando calificaciones que no estaban explícitas en el dataset si no que fueron inferidas.

5. Experimentos a realizar

Se propone comparar el desempeño de Collaborative Filtering con el del modelo probabilístico, el cual será en un principio *CPT-v*, descrito en *Collaborative Filtering and the Missing at Random Assumption*, publicación que además fue el motivo de la recaudación del *dataset* que se utilizará en este proyecto.¹

6. Dataset

Yahoo! Music ratings for User Selected and Randomly Selected songs, version 1.0 (1.2 MB)

Este conjunto de datos contiene calificaciones de canciones recopiladas de dos fuentes diferentes. La primera fuente consiste en calificaciones proporcionadas por los usuarios durante la interacción normal con *LaunchCast Radio* y se utilizará para entrenar los modelos. La segunda fuente consiste en calificaciones de canciones seleccionadas al azar recopiladas durante una encuesta en línea realizada por *Yahoo! Research* denominada *LaunchCast Radio Rating Study* y se utilizará para medir el rendimiento de los modelos.

LaunchCast es una radio en la cual los usuarios pueden calificar canciones, alterando el orden en el cual estas son reproducidas.

Características de los datos de interacción normal con *LaunchCast Radio*:

- 1000 canciones
- 15,400 usuarios
- Usuarios con al menos 10 calificaciones
- Usuarios **deciden** qué canciones calificar

Características de los datos relacionados a la encuesta *LaunchCast Radio Rating Study*:

- 1000 canciones
- 5,400 usuarios (subconjunto del *dataset* anterior)
- Cada usuario califica **obligatoriamente** 10 canciones, que son seleccionadas de manera aleatoria

Los datos de la encuesta y las calificaciones de canciones seleccionadas al azar se recopilaron entre el 22 de agosto de 2006 y el 7 de septiembre de 2006. Los datos de interacción normal se recopilaron entre 2002 y 2006. El tamaño de este conjunto de datos es de 1.2 MB.

6.1. LaunchCast Radio Rating Study

6.1.1. Determinar si la preferencia por una canción afecta la decisión de calificarla

La primera parte de la encuesta buscaba determinar si la existencia de un rating sobre una canción por parte de un usuario depende del rating mismo, por ejemplo, si escucha una canción que considera neutral, quizás hay baja

probabilidad de calificarla, mientras que si le encantó, la probabilidad es alta. Para ésto se les hizo responder siete preguntas que se pueden ver en el **Anexo 1**.

6.1.2. Obtener dataset donde los datos faltantes sean aleatorios

A cada uno de los 35,786 usuarios que participaron en la encuesta, se les pidió evaluar exactamente diez canciones. Las canciones fueron seleccionadas aleatoriamente para cada usuario a partir de un set de mil canciones.

Estas mil canciones fueron escogidas aleatoriamente entre el set completo de canciones de LaunchCast Radio que tuviesen al menos 500 ratings.

Se denominará el *dataset* anterior como '*survey ratings*', el cual representa un *dataset* con calificaciones aleatorias. Este será utilizado para realizar los tests.

Por otro lado, se extrajeron todos los *ratings* que los usuarios tenían sobre esas mil canciones (calificaciones previas a la encuesta).

Se denominará el *dataset* anterior como '*base ratings*', el cual representa un *dataset* con calificaciones sobre canciones que el usuario **escogió** calificar. Este será utilizado para entrenar los modelos.

Finalmente, de la totalidad de encuestados, se eliminaron todos ellos que tuviesen menos de 10 calificaciones registradas en '*base ratings*', resultando en un *dataset* de 5,400 usuarios.

6.2. Análisis de la encuesta

De la primera parte de la encuesta se elabora el **Cuadro 1**, el cual muestra el comportamiento de los usuarios en la calificación de canciones.

Frecuencia de calificacion	Nivel de preferencia				
	Lo odio	No me gusta	Neutral	Me gusta	Me encanta
Nunca	6.76 %	4.69 %	2.33 %	0.11 %	0.07 %
Muy infrecuente	1.59 %	4.17 %	9.46 %	0.54 %	0.35 %
Infrecuente	1.63 %	4.44 %	24.87 %	1.48 %	0.20 %
Frecuente	12.46 %	22.50 %	26.83 %	25.30 %	5.46 %
Muy frecuente	77.56 %	64.20 %	36.50 %	72.57 %	93.91 %

Cuadro 1: Frecuencia de calificación reportada por usuarios según nivel de preferencia

Se observa de manera general un incremento en la frecuencia de calificación a medida que la preferencia por la canción aumenta. Lo anterior implica que probablemente el dataset '*base ratings*' tenderá a tener mayor cantidad de calificaciones de valor 4 y 5, pues las calificaciones bajas como 2 o 3 son menos

probables de ocurrir.

Por otro lado, el 64,85 % de los encuestados asegura que el nivel de preferencia **sí** afecta la decisión de calificar una canción, respaldando lo visto en el cuadro anterior. Adicionalmente, el 81.56 % de los encuestados asegura calificar algunas canciones diariamente.

Finalmente, se procede a analizar las calificaciones entregadas por los usuarios durante la encuesta. Como se puede apreciar en la **Figura 1**, la hipótesis planteada anteriormente es confirmada. Cuando los usuarios pueden decidir que canciones calificar, entonces las calificaciones tienden a ser mayores.

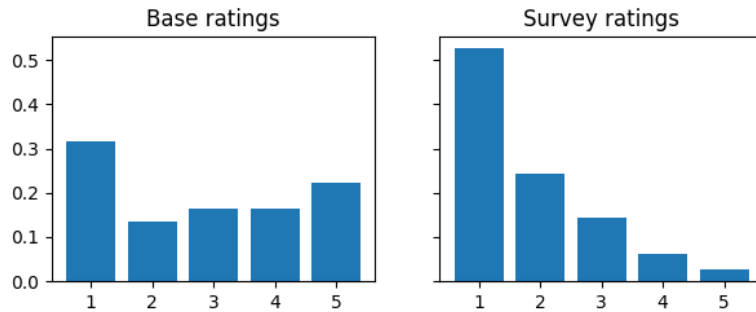


Figura 1: Comparación de frecuencia de calificaciones en *'base ratings'* y *'survey ratings'*

7. Anexos

7.1. Encuesta LaunchCast Radio

Frecuencia en la que califico la música que escucho en LaunchCast Radio

1. Nunca
2. Algunas canciones desde que me registré
3. Algunas canciones al mes
4. Algunas canciones por semana
5. Algunas canciones por día

Frecuencia en la que califico canciones que odio

1. Nunca
2. Muy infrecuente

3. Infrecuente
4. Frecuente
5. Muy frecuente

Frecuencia en la que califico canciones que no me gustan

1. Nunca
2. Muy infrecuente
3. Infrecuente
4. Frecuente
5. Muy frecuente

Frecuencia en la que califico canciones que son neutrales

1. Nunca
2. Muy infrecuente
3. Infrecuente
4. Frecuente
5. Muy frecuente

Frecuencia en la que califico canciones que me gustan

1. Nunca
2. Muy infrecuente
3. Infrecuente
4. Frecuente
5. Muy frecuente

Frecuencia en la que califico canciones que me encantan

1. Nunca
2. Muy infrecuente
3. Infrecuente
4. Frecuente
5. Muy frecuente

Mis preferencias por una canción no afectan la decisión de calificarla

1. Verdadero - Mis preferencias **no** afectan mi decisión
2. Falso - Mis preferencias **si** afectan mi decisión

Referencias

- ¹ Marlin et al. Collaborative filtering and the missing at random assumption. *UAI*, 2007.