



Taller 1 – Modelos colaborativos

Objetivos

- Desarrollar y evaluar un modelo colaborativo de recomendación de información
- Realizar una práctica sobre un *dataset* real

Prerrequisitos

- Conocimiento básico de programación en los lenguajes sugeridos: Python o Java
- Conocimiento básico de Unix
- Conocimiento del esquema de desarrollo de una aplicación Web

Herramientas

- Máquina virtual Unix
- Ambiente de desarrollo de software
- *Dataset* de pruebas

Enunciado

1) Conocimiento del *dataset* de trabajo

El taller se trabaja sobre un conjunto de datos obtenido de una red social de música, Last.fm.

- a) Ubique los datos en <http://ocelma.net/MusicRecommendationDataset/lastfm-1K.html>. Allí encuentra la información técnica y práctica para el uso de los datos, así como los requerimientos para su uso.
- b) Estudie el formato de los datos, en particular identifique la manera de identificar qué información refleja la interacción entre usuarios e ítems.

2) Pre-procesamiento de datos

- a) Transforme los datos correspondientes a la interacción entre usuarios e ítems, implementando una estrategia para convertir estos datos en unos que sean compatibles con los modelos vistos en clase. Justifique en el informe sus decisiones en este paso.
- b) Tome los datos compatibles con modelos colaborativos y pártalos en dos conjuntos: un grupo de datos le sirve para construir el modelo y el resto para medir sus predicciones. Sepárelos en archivos distintos.

3) (20%) Construcción de modelos colaborativos usuario-usuario

- a) Construya un modelo colaborativo basado en perfiles de usuario con la primera parte de los datos de ratings.
- b) Realice las predicciones de relevancia para los usuarios e ítems que encuentra en la segunda parte de los datos.
 - i) Utilice un modelo basado en Índice de Jaccard
 - ii) Utilice un modelo basado en distancias coseno
 - iii) Utilice un modelo basado en correlación de Pearson.
- c) Compare su predicción de rating con el efectivamente encontrado en el *dataset*. Establezca una forma de evaluar globalmente sus distancias en las predicciones que refleje la calidad de las mismas
- d) Varíe la estrategia de selección de vecinos por umbral de similitud y por número de vecinos. Revise cuál es el impacto al variar estos parámetros.

- e) Revise la estrategia de ponderación por significancia de McLaughlin's [1] (*McLaughlin's significance weighting*) y revise cuál es el impacto al variar los parámetros de esta estrategia.

4) (20%) Construcción de modelos colaborativos ítem - ítem

Realice el mismo proceso anterior para un modelo ítem-ítem

5) (30%) Construya una aplicación Web interactiva sencilla que permita interactuar con sus experimentos.

- Según los resultados obtenidos en la fase de experimentación, implemente el sistema de recomendación que permita obtener las mejores recomendaciones para un usuario activo del sistema. Debe hacer una correcta visualización de los resultados de la recomendación, de manera que se tenga claramente la información sobre el usuario y la información sobre el ítem que se visualiza, de forma que se puedan entender en el dominio de los datos, no códigos o referencias propias a los datos dentro de los archivos
- La aplicación debe permitir loguearse en la aplicación bajo la identidad de cualquier usuario presente en el dataset y obtener recomendaciones, además, debe permitir añadir un nuevo usuario y sus preferencias (por ejemplo, usted o su profesor).
- La aplicación debe permitir visualizar y evaluar la recomendación construida. Debe mostrar muy claramente la información sobre los ratings históricos del usuario activo, del ítem actual, de ítems o usuarios vecinos según el modelo, de tal forma que el usuario pueda entender por qué se están obteniendo esas recomendaciones.
- La aplicación debe estar desplegada en la máquina de publicación de resultados asignada al equipo.**

6) (30%) Análisis de resultados

Elabore un informe en el que explique detalladamente las alternativas contempladas para el procesamiento de datos y por qué se eligió la alternativa implementada. Explique detalladamente qué experimentación realizó sobre los modelos y similitudes propuestas en el enunciado y justifique por qué decidió implementar el sistema sustentándolo en los resultados de la experimentación.

Explique cómo su sistema incorpora nuevos usuarios.

Explique la arquitectura general del sistema implementado (arquitectura, tecnologías utilizadas, modelo de datos)

Incluya en su informe conclusiones generales y lecciones aprendidas.

Usted debe poder siempre explicar los resultados mostrados en la demostración Web.

Metodología

- El taller se desarrolla en los grupos de estudiantes que están previstos para el desarrollo de trabajos prácticos.
- Se espera una participación equitativa de los integrantes del grupo

Entrega de laboratorio

Fecha y hora límite: Lunes 14 de Marzo 13:00 horas.

Archivo de entrega : **<Taller1_NN_login1_login2_login3 >.zip**. NN es el número del grupo y los *logins* corresponden a los integrantes presentes y participantes en el desarrollo.

Contenido del archivo:

- Archivo README.txt que indique integrantes del grupo y contenido del zip.
- Informe de análisis de resultados, en formato `.pdf`, que sigue la plantilla de informes técnicos prevista para el curso. Nombre del archivo: **<Taller1_NN_login1_login2_login3 >.pdf**
- Archivos del código fuente desarrollado en carpeta `code`. El código entregado debe corresponder en su totalidad a la aplicación desplegada en la máquina asignada al equipo.

Recuerde que la sustentación y el cierre hacen parte de la evaluación de su trabajo. Esta entrega de realizarse siguiendo las normas establecidas para presentación, sustentación y cierre del trabajo.

Bibliografía

[1] McLaughlin, Matthew R., and Jonathan L. Herlocker. "A collaborative filtering algorithm and evaluation metric that accurately model the user experience." *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2004.