



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA
DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN

IIC3633 — Sistemas Recomendadores — 2' 2021

Propuesta de Proyecto

Integrantes:

- Benjamín Farías
- Benjamín Lepe
- Juan Romero

Contexto

Hoy en día, la industria de los videojuegos ha crecido enormemente, resultando en uno de los hobbies más comunes en el mundo. El catálogo más diverso actualmente corresponde al de los juegos de computador, surgiendo así una comunidad inmensa de personas apasionadas por este gran arte. Dentro de este mundo se encuentra *Steam*, una de las plataformas de distribución de videojuegos más grandes del mundo, en la que se puede encontrar una gran cantidad de información tal como las preferencias de los usuarios, géneros y comentarios sobre distintos juegos que se encuentran disponibles, entre otros. Para este proyecto trabajaremos con datasets de *Steam*, los que contienen datos sobre las horas dedicadas a los juegos comprados por cada uno de los usuarios, características sobre los videojuegos y sus respectivas reviews realizadas por los mismos usuarios de la plataforma. La razón por la que escogimos este tema es que, en primer lugar, tenemos bastante conocimiento sobre videojuegos, lo que hace más fácil comprobar si las recomendaciones realizadas tienen algún sentido, y en segundo lugar, tiene una amplia variedad de datos, lo que nos permite implementar distintos tipos de algoritmos, desde recomendaciones *Content Based* hasta recomendaciones basadas en *Collaborative Filtering*.

Problema

Ante el catálogo increíblemente amplio y variado de juegos de la tienda de Steam, los usuarios se enfrentan constantemente al problema de encontrar juegos que les interesen y entretengan, lo que puede consumir bastante tiempo de forma innecesaria. Además, dadas las recomendaciones mismas de la plataforma (especialmente durante temporada de ofertas), al usuario le gustaría poder determinar aquellos juegos que vale más la pena comprar. Es por esto que este proyecto se centrará en probar distintos métodos de recomendación que permitan apoyar a los usuarios de esta plataforma al momento de comprar juegos.

Objetivos

Los objetivos del sistema recomendador de este proyecto son:

- **Ranking:** Dado un usuario y una colección de juegos, el software debe ser capaz de ordenar dicho conjunto en base a la preferencia que les daría el usuario.
- **Rating:** Dado un usuario y un juego, el software desarrollado debe ser capaz de responder la pregunta: ¿Le gustaría dicho juego al usuario?
- **Relacionados:** Dado un determinado juego y una colección de juegos, seleccionar aquellos que sean similares a dicho juego.

Solución Propuesta

La información de los datasets será procesada para poder ser manipulada por los algoritmos que eventualmente se usarán para la recomendación. En particular, uno de los principales objetivos en esta etapa es obtener un rating para cada interacción entre usuarios y juegos, el que se puede obtener de manera explícita (reviews) e implícita (horas jugadas) con los datos disponibles. Una vez obtenidos los ratings, se aplicarán recomendaciones basadas en técnicas de filtrado colaborativo, contenido y posiblemente *Deep Learning*. Cabe mencionar que la representación de los ratings, así como de los usuarios e ítems, va a variar según la técnica probada.

Entre las técnicas a aplicar se encuentran:

- **Colaborative Filtering:** Algoritmos de filtrado colaborativo, tales como *KNN*, *Matrix Factorization* (*ALS* y *BPR*), *Funk SVD*, entre otros, los que nos permitirán sacar provecho a la gran cantidad de datos sobre interacciones entre usuarios e ítems.
- **Content Based:** Algoritmos de recomendación basados en contenido. Se deberán obtener los vectores para representar a los juegos y a los perfiles de usuario, los que se pueden generar a partir de las características de los videojuegos que ha consumido un jugador. Posteriormente, se pueden implementar técnicas basadas en *Deep Learning*, *Decision Trees* y *KNN*.

Se omitirá la realización de recomendaciones basadas en imágenes, pues esta información es poco representativa respecto al contenido del juego (y en varios casos engañosa). En cuanto a las recomendaciones basadas en texto, se omitirán los comentarios de las reviews debido a que se requiere un sistema de análisis de sentimientos y semántica que funcione en varios idiomas, lo que se escapa del alcance del curso.

Finalmente, una vez realizados los ajustes antes mencionados, se experimentará con distintos algoritmos y se seleccionará el o aquellos que hayan presentado un mejor desempeño al momento de aplicarlo sobre particiones representativas de los datasets. A partir de los resultados, veremos también si es posible crear algún tipo de ensamble o sistema híbrido que supere el rendimiento de las técnicas individuales.

Experimentos

Para entrenar y evaluar las distintas técnicas se generarán particiones de training, validation y testing a partir de los datasets originales. Para obtener una interpretación del rendimiento de nuestros modelos, se calcularán métricas tales como *MAP*, *nDCG* y *RMSE*.

A continuación se indican experimentos atinentes a cada objetivo:

- **Ranking:** Se pueden combinar métodos de *Matrix Factorization* con representaciones basadas en contenido, permitiendo balancear entre juegos que le podrían interesar al usuario (en base a las interacciones) y juegos similares a los que ya ha jugado (en base al contenido).
- **Rating:** Con alguno de los algoritmos de *Matrix Factorization* se puede obtener un vector de largo N con un cierto número de juegos recomendados, y si un ítem en específico no se encuentra dentro de esta recomendación se puede predecir que no está entre las preferencias del usuario. Otra forma puede ser utilizando métodos basados en **Deep Learning**, en los que a partir de un ítem en particular se entregue un output binario (2 neuronas de salida).
- **Relacionados:** Estrictamente relacionado al contenido de los juegos. Se probarán representaciones *Content Based* que permitan aplicar *KNN* y *Clustering* sobre los ítems, agrupándolos según diversos criterios de similaridad.

Una vez realizados los experimentos según técnica y objetivo, se probarán combinaciones entre las que mejor rindieron, buscando mejorar el rendimiento individual de cada método.

Bibliografía

- **Dataset de Reviews y Usuarios de Steam:**
https://cseweb.ucsd.edu/~jmcauley/datasets.html#steam_data
- **Dataset de Videojuegos de Steam:**
<https://www.kaggle.com/nikdavis/steam-store-games>
- **Comportamiento de los Usuarios de Steam:**
<https://dl.acm.org/doi/10.1145/2987443.2987489>
- **Sistemas Recomendadores para Steam:**
<https://towardsdatascience.com/steam-recommendation-systems-4358917288eb>