



Informe de Propuesta (Análisis & Baselines)

IIC3633 - Sistemas Recomendadores

Grupo 11

Integrantes:

Vicente Correa

Alberto Maturana

Mariela Zambrano

3 de octubre, 2025

Descripción de problema y justificación

Los sistemas recomendadores suelen funcionar para un único usuario. Según sus preferencias se crean sugerencias de productos, servicios o contenidos. Pero en muchos escenarios de la vida real se trabaja con preferencias de grupos enteros, como es el caso de los juegos de mesa, en donde se intenta ser una experiencia agradable para todos.

Por este motivo, nace la necesidad de crear o mejorar un sistema de recomendación capaz de considerar múltiples preferencias de múltiples usuarios, con el objetivo de maximizar la satisfacción grupal. Así, el desafío es integrar y balancear los gustos individuales de un determinado grupo para recomendar un juego maximice la satisfacción de los jugadores.

El problema resulta relevante por varias razones. En primer lugar, en términos prácticos, podemos pensar en los juegos de mesa como un mercado, donde la elección de un título llamativo puede marcar la diferencia al momento de elegir entre los juegos disponibles. En segundo lugar, presenta un desafío en sistemas recomendadores, ya que la recomendación grupal abarca a más de un usuario (lo tradicional). En tercer lugar, los datos disponibles para los juegos de mesa son multimodales, es decir, contienen información numérica (ratings y duración), textual e incluso imágenes. Esto permite explorar métodos adicionales que aporten positivamente el modelo de recomendación. Finalmente, puede ser interesante realizar la comparativa de hacer estas recomendaciones de manera grupal versus una recomendación individual, midiendo los beneficios y las limitaciones de ambos métodos.

Objetivos del proyecto

El objetivo del proyecto a nivel macro es desarrollar un sistema de recomendación grupal de juegos de mesa que integre las preferencias de múltiples usuarios, utilizando el dataset de BoardGameGeek. Esto con el fin de generar sugerencias que maximicen la satisfacción colectiva y comparar su desempeño frente a métodos de recomendación individual.

Se pueden destacar objetivos más específicos como:

- Analizar y preprocesar el dataset de juegos de mesa, caracterizando sus atributos.
- Implementar modelos de referencia que permitan determinar un punto base para compararlo con otros modelos o métodos.
- Agregar preferencias de forma estratégica, tal que permitan combinar la calificación individual en una recomendación grupal.
- Ver y evaluar la multimodalidad de los datos, para analizar el impacto de considerar distintas fuentes de información.
- Comparar el desempeño de la recomendación grupal contra la individual.

Análisis de los datos

El dataset seleccionado (Board Games Database from BoardGameGeek, disponible en Kaggle (Wadkins, 2021)) contiene información relevante de miles de juegos de mesa, incluyendo: calificaciones de usuarios, categorías, reglas, número de jugadores y duración del juego. Estos atributos permiten tanto caracterizar los gustos de los usuarios, como evaluar su compatibilidad con nuevos juegos al estar en grupos.

El dataset contiene 9 tablas con las siguientes características:

- **GAMES** - información básica de 22 mil juegos de mesa. Contiene información como nombre, número de jugadores, calificación promedio.
- **RATINGS DISTRIBUTION** - distribución de ratings para cada juego de mesa. Va de manera discreta desde el 0 al 10 en intervalos de 0.1 puntos.
- **THEMES** - temáticas de cada juego (fantasía, aventura, pelea, entre otros). Se utilizan *flags* binarias para cada indicar la presencia de la temática.
- **MECHANICS** - reglas y estilos de juego. Utiliza *flags* binarias para indicar presencia en cada juego de mesa.
- **SUBCATEGORIES** - subcategorías con *flags* binarias para indicar presencia en cada juego de mesa.
- **ARTISTS REDUCED** - información de los artistas involucrados (con 3 o más juegos de mesa). Se utilizan *flags* binarias para indicar participación.
- **DESIGNERS REDUCED** - información de los diseñadores involucrados (con al menos 3 trabajos). Se utilizan *flags* binarias para indicar la participación.
- **PUBLISHERS REDUCED** - información de los publicadores involucrados (con 3 o más trabajos). Se utilizan *flags* binarias para indicar participación.
- **USER RATINGS** - Contiene todos los ratings hechos por los usuarios a los distintos juegos de mesa. Contiene más de 411 mil usuarios distintos con un total de 19 millones de *ratings* realizados.

Lo más importante de los datos brutos es la distribución de ratings y ids. Curiosamente, como se ve en el Anexo 1, los identificadores para los juegos no están distribuidos del 0 al 22 mil sino que existen identificadores con valores mucho más altos. Creemos que esto se debe a la posibilidad de que existan muchos juegos sin ratings o que se hayan eliminado de los datos. Al momento de crear o añadir nuevos juegos, se le asigna identificadores antiguos que no estaban ocupados.

Para este proyecto, se realizó una partición de los datos en conjuntos de entrenamiento (60%), validación (20%) y test (20%), tras un proceso de limpieza que incluyó la eliminación de calificaciones duplicadas por un mismo usuario para un mismo ítem. Las estadísticas se pueden ver en el Anexo 2, donde se puede observar que la matriz de interacciones usuario-ítem es dispersa, con una densidad del 0.14% en el conjunto de entrenamiento. Esto es un desafío común en sistemas de

recomendación y requiere modelos capaces de manejar esta falta de datos. Por el lado de los ratings, como se observa en el Anexo 3, vemos que estos tienen ciertos valores que se repiten mucho más que otros, como valores enteros 6, 7 y 8, en vez de valores decimales. Esto es por el comportamiento natural del ser humano al asignar valores discretos y “redondos” (Honda, Kagawa, Shirasuna, 2022). Adicionalmente vemos el sesgo de las calificaciones “buenas” que se gravita sobre los valores 6 7 y 8 y la ausencia de calificaciones de 1, 2 y 3 (Hu, Zhang, Pavlou, 2009).

Implementación de modelos

Se implementaron tres modelos de referencia. Primero se implementó el modelo Random para comparar el modelo de recomendación con uno que recomienda sin ningún tipo de análisis de los datos. Luego, se implementó el modelo Most Popular tomando el promedio de ratings para todos los ítems y seleccionando aquellos con un mejor promedio. Finalmente se implementó el modelo SVD++. Además, cabe agregar que para los modelos Random y Most Popular probamos haciendo sugerencias a grupos de 4. Para validar la recomendación simplemente tomábamos la calificación promedio que le habían dado los usuarios al ítem. Esa se comparaba con nuestra calificación predicha. Creemos que en una futura entrega se puede desarrollar algo para que el “ground truth” sea algo más representativo que un promedio simple. O que las funciones de pérdida para los modelos entrenados consideren este aspecto “grupal”.

De estos 3 modelos obtuvimos métricas simples para comparar su rendimiento. Se adjunta la tabla con los resultados para los modelos con recomendaciones individuales y grupales según corresponda.

Modelo	Random (individual)	Random (grupal)	Most Popular (individual)	Most Popular (grupal)	SVD++ (individual)
MAE	4.0517	3.5	0.9992	0.4803	0.9971
RMSE	4.5735	3.5	1.3201	0.6151	1.3123

Los resultados muestran claras diferencias. El modelo Random presenta un error alto, validando su rol como baseline mínimo. El modelo Most Popular, pese a su sencillez, alcanza un RMSE de 1.3201 con un costo computacional muy bajo, resultando competitivo. En contraste, el SVD++ logra el mejor desempeño, aunque la mejora es marginal frente al Most Popular y demanda un tiempo mucho mayor y casi 9 veces más memoria. Esto evidencia la dificultad del problema y la necesidad de modelos más sofisticados, como híbridos, capaces de capturar interacciones complejas y superar con mayor claridad a los baselines.

Planificación para siguiente etapa

1. Implementación de Modelos Avanzados:

Modelo Híbrido/Basado en Contenido: Se explorará la implementación de un modelo que combine el filtrado colaborativo con características de contenido de los juegos (categorías, mecánicas, etc.), como por ejemplo, un modelo de DeepFM. El objetivo es ver si la información del contenido puede mejorar las predicciones para usuarios con pocas calificaciones.

2. Desarrollo de Estrategias de Agregación Grupal:

Se implementará al menos tres estrategias para combinar las predicciones individuales en una recomendación grupal:

Estrategia de Promedio (Average): Se calculará el rating promedio predicho por el modelo para todos los miembros del grupo.

Estrategia de Mínima Miseria (Least Misery): La calificación del grupo para un ítem será la calificación mínima predicha entre todos los miembros, buscan evitar ítems que disgusten a alguien.

Estrategia de Máximo Placer (Most Pleasure): La calificación del grupo será la calificación máxima predicha, buscando maximizar la satisfacción del miembro más entusiasta.

3. Diseño de la Evaluación Grupal:

Creación de Grupos Sintéticos: Se formarán grupos de usuarios de manera sintética a partir del dataset de validación, buscando usuarios que hayan calificado al menos un ítem en común.

Métricas de Evaluación: Para evaluar la calidad de las listas de recomendación grupales, se utilizarán métricas de ranking como nDCG@10 (Normalized Discounted Cumulative Gain) y Precision@10. Estas métricas son más adecuadas que el RMSE para evaluar listas ordenadas de recomendaciones.

Cronograma Propuesto:

Semana 1 (hasta 10 Oct): Implementación del modelo DeepFM y ajuste de hiperparámetros.

Semana 2 (hasta 17 Oct): Desarrollo y prueba de las tres estrategias de agregación grupal.

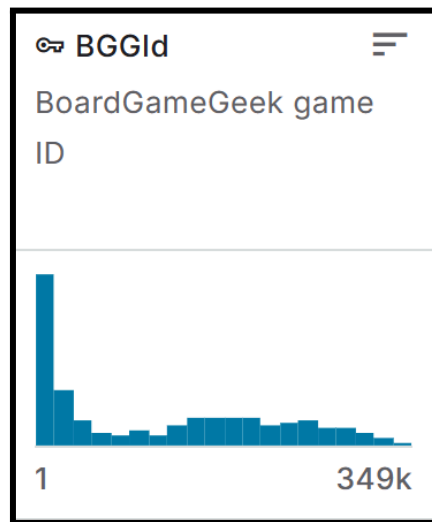
Semana 3 (hasta 24 Oct): Ejecución de los experimentos finales y recolección de métricas (RMSE, nDCG@10).

Semana 4 (hasta 30 Oct): Análisis de resultados, conclusiones y redacción del informe intermedio.

Bibliografia

- Honda, H., Kagawa, R., & Shirasuna, M. (2022). On the round number bias and wisdom of crowds in different response formats for numerical estimation. *Scientific reports*, 12(1), 8167. <https://doi.org/10.1038/s41598-022-11900-7>
- Hu, N., Zhang, J., & Pavlou, P. A. (2009). Overcoming the J-shaped distribution of product reviews. *Communications of the ACM*, 52(10), 144–147. <https://doi.org/10.1145/1562764.1562800>
- Wadkins, J. (2021). *Board Games Database from BoardGameGeek* [Data set]. Kaggle. <https://www.kaggle.com/datasets/threnjen/board-games-database-from-boardgamegeek>

Anexos

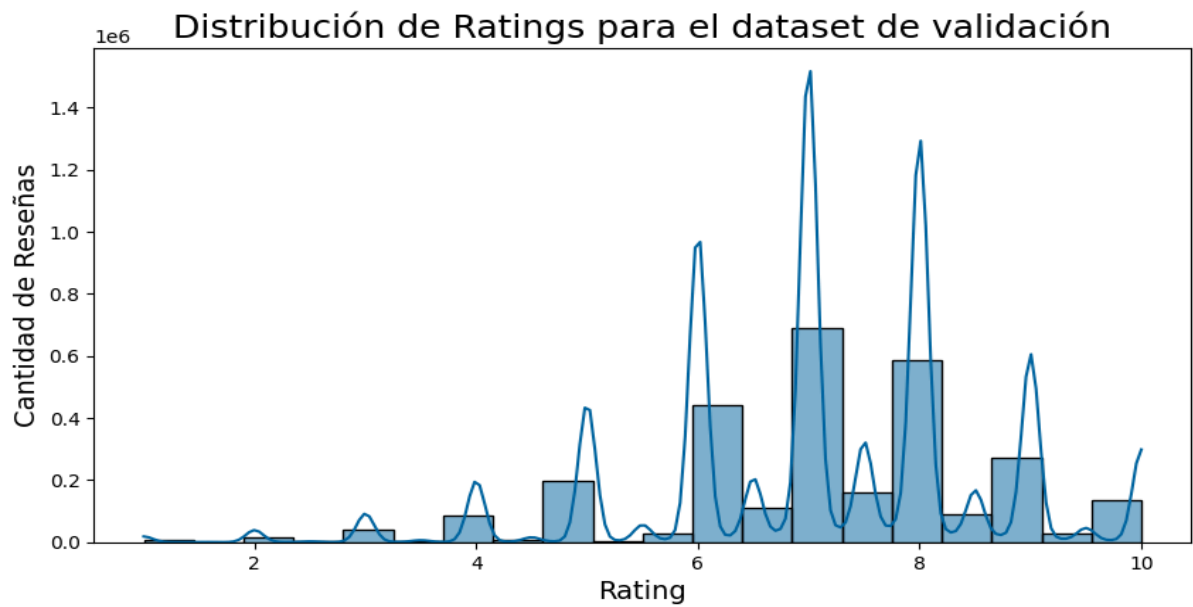


Anexo 1: Distribución Ids de juegos de mesa.

	Estadística	Full Dataset	Train	Validation
0	Número de Usuarios	411374.00	411374.00	260470.00
1	Número de Ítems	21925.00	21925.00	21900.00
2	Total de Reseñas	15306854.00	12390406.00	2901744.00
3	Promedio de Reseñas por Usuario	37.21	30.12	11.14
4	Promedio de Reseñas por Ítem	698.15	565.13	132.50
5	Promedio de Rating	7.14	7.15	7.10
6	Desviación Estándar del Rating	1.55	1.55	1.53
7	Máximo Número de Reseñas por un Usuario	5195.00	4147.00	1039.00
8	Máximo Número de Reseñas para un Ítem	87843.00	71852.00	15998.00

9	Mínimo Número de Reseñas por un Usuario	1.00	1.00	1.00
10	Mínimo Número de Reseñas para un Ítem	7.00	7.00	1.00
11	Densidad (%)	0.17	0.14	0.05

Anexo 2: Tabla con estadísticas descriptivas de los datos.



Anexo 3: Gráfico Cantidad de reseñas vs Rating.