



Datos Masivos

Creación de Sistemas de Recomendación para una plataforma de videojuegos

Licenciatura en Ciencia de Datos
FES Acatlán

Alumnos:

Sevilla Gallardo Saúl Sebastián
Moreno Cruz Jimena

Índice

● Introducción	3
● Descripción de la información	3
1. Juegos	
2. Reseñas	
3. Rankings	
● Modelado e Implementación	4
1. Filtrado Colaborativo	
2. Recomendación Demográfica	
3. Recomendación Basada en Contenido	
● Comparación y uso combinado	6
● Interpretación de Resultados	4
1. Filtrado Colaborativo	
- Interpretación	
- Limitaciones del RMSE	
2. Recomendación Demográfica y Basada en Contenido	
- Recomendación Demográfica	
- Recomendación Basada en Contenido	
● Conclusiones	7
1. Impacto del Sistema	
2. Limitaciones y Futuras Mejoras	
● Bibliografía	8

Introducción

El crecimiento exponencial de plataformas de distribución digital, como Steam, ha generado una cantidad masiva de datos relacionados con juegos, usuarios y sus interacciones. Esto ha hecho imprescindible el desarrollo de sistemas de recomendación que permitan a los usuarios descubrir contenido relevante de manera eficiente, mejorando su experiencia en la plataforma.

En este proyecto, se diseñó un sistema de recomendación de juegos que aprovecha técnicas avanzadas de análisis de datos para entender y predecir las preferencias de los usuarios. Utilizando información como descripciones de juegos, reseñas, géneros y rankings, se implementaron tres enfoques principales:

1. **Filtrado Colaborativo:** Basado en patrones compartidos entre usuarios con gustos similares.
2. **Recomendación Demográfica:** Centradas en tendencias generales y popularidad de los juegos.
3. **Recomendación Basada en Contenido:** Enfocada en las similitudes entre los juegos, considerando atributos como género y estilo.

Este sistema no solo ayuda a los usuarios a encontrar juegos que podrían disfrutar, sino que también ofrece a desarrolladores y publicadores una herramienta valiosa para identificar tendencias y optimizar sus estrategias. En las siguientes secciones, se describe detalladamente el proceso, los datos utilizados y el desempeño de cada modelo de recomendación.

Descripción de la información

El sistema de recomendación se desarrolló utilizando tres conjuntos de datos principales, cada uno enfocado en diferentes aspectos relacionados con los juegos, las reseñas de los usuarios y las métricas de popularidad. A continuación, se describen las características de cada conjunto:

1. Juegos

Este conjunto incluye información detallada sobre los juegos disponibles en la plataforma. Los atributos recopilados son:

- **Nombre:** Identificador principal del juego.
- **Descripción:** Resumen tanto corto como largo del juego.
- **Géneros:** Categorías que describen el tipo de juego, como acción, estrategia, simulación, etc.
- **Requerimientos técnicos:** Especificaciones mínimas y recomendadas para ejecutar el juego.
- **Desarrollador y publicante:** Empresas responsables de la creación y distribución del juego.
- **Rating general:** Valoración promedio basada en las reseñas de los usuarios.

- **Fecha de lanzamiento:** Momento en que el juego estuvo disponible en la plataforma.
-

2. Reseñas

Este conjunto recoge datos sobre las interacciones de los usuarios con los juegos. Los atributos incluidos son:

- **Reseña:** Opinión escrita por los usuarios sobre su experiencia con el juego.
 - **Horas jugadas:** Tiempo total que el usuario ha dedicado al juego.
 - **¿Es útil? y ¿Es divertida?:** Indicadores basados en el feedback de otros usuarios sobre la reseña.
 - **Fecha:** Momento en que se realizó la reseña.
 - **Nombre del juego:** Identificación del juego reseñado.
 - **Nombre de usuario:** Identificador del usuario que realizó la reseña.
-

3. Rankings

Este conjunto proporciona información sobre la popularidad y clasificación de los juegos según distintas métricas. Los atributos son:

- **Nombre del juego:** Título del juego.
- **Género:** Clasificación del juego por su categoría principal.
- **Tipo de ranking:** Métrica utilizada para determinar la popularidad del juego (ejemplo: ventas, calificaciones, horas jugadas).
- **Ranking:** Posición del juego en relación con otros de su género o globalmente.

Modelado e Implementación

Para diseñar un sistema de recomendación efectivo, se implementaron tres modelos principales que aprovechan distintas estrategias para generar recomendaciones personalizadas. Cada modelo tiene un enfoque único que permite abordar diferentes aspectos de las preferencias y comportamientos de los usuarios. A continuación, se describen los modelos y su implementación:

1. Filtrado Colaborativo

El modelo de **Filtrado Colaborativo** se basa en identificar patrones de comportamiento entre usuarios o ítems. En este caso, busca recomendar juegos a los usuarios basándose en las calificaciones o interacciones de otros usuarios con preferencias similares.

Implementación:

- **Técnica utilizada:** Se empleó un enfoque basado en matrices factorizadas (SVD - Singular Value Decomposition) para representar las relaciones entre usuarios y juegos en un espacio latente.
 - **Datos requeridos:** Se utilizó la matriz *usuario-juego*, donde cada celda representa la valoración de un usuario para un juego específico.
 - **Resultados:** Este modelo es altamente efectivo para generar recomendaciones personalizadas, como listas de juegos sugeridos para cada usuario bajo el lema *"Nuestra selección para ti"*.
-

2. Recomendación Demográfica

Este modelo realiza recomendaciones generales basadas en tendencias populares o características demográficas de los usuarios, como la región geográfica.

Implementación:

- **Técnica utilizada:** Se identificaron los juegos más populares a partir de métricas globales como el número de valoraciones positivas, horas jugadas, o rankings.
 - **Datos requeridos:** Rankings globales y métricas de popularidad asociadas a los juegos.
 - **Resultados:** Este enfoque es ideal para listas como *"Popular en México"* o *"Lo mejor del mes"*, destacando juegos de amplio reconocimiento que son relevantes para grandes grupos de usuarios.
-

3. Recomendación Basada en Contenido

Este modelo utiliza las características de los juegos (géneros, descripciones, desarrolladores, etc.) para encontrar similitudes entre ellos y generar recomendaciones en función de las preferencias históricas de un usuario.

Implementación:

- **Técnica utilizada:** Se emplearon técnicas de similitud como el coseno o la distancia euclidiana sobre los atributos de los juegos. También se pueden usar modelos más avanzados, como TF-IDF en descripciones textuales, para medir la relación entre juegos.
 - **Datos requeridos:** Información descriptiva de los juegos (géneros, descripciones, desarrolladores, etiquetas, etc.).
 - **Resultados:** Este modelo genera recomendaciones como *"Porque jugaste Grand Theft Auto, te podría gustar Cyberpunk 2077"*, basándose en similitudes temáticas o mecánicas.
-

Comparación y uso combinado

En la implementación final, estos tres enfoques se combinan para cubrir diferentes necesidades de los usuarios:

- **Filtrado Colaborativo:** Ideal para recomendaciones personalizadas.
- **Demográfico:** Útil para destacar tendencias y juegos ampliamente aceptados.
- **Basado en Contenido:** Perfecto para explorar juegos relacionados temáticamente.

Esta combinación asegura un sistema robusto y versátil que maximiza la satisfacción del usuario.

El desempeño del sistema de recomendación se evaluó principalmente a través del modelo de **Filtrado Colaborativo**, dado que es el único modelo con una métrica cuantitativa asociada. Los otros modelos (Recomendación Demográfica y Basada en Contenido) se analizaron en función de su relevancia y aplicabilidad práctica, sin una métrica de ajuste específica.

Interpretación de Resultados

1. Filtrado Colaborativo

Error del Filtrado Colaborativo: RMSE = 0.34

El **Root Mean Square Error (RMSE)** obtenido para el modelo de Filtrado Colaborativo fue de **0.34** en una escala de 0 a 1. Esto indica que, en promedio, las predicciones del modelo difieren de las calificaciones reales en aproximadamente un **34% del rango total de valores posibles**.

Interpretación

- Este error es relativamente bajo, lo que sugiere que el modelo tiene un buen ajuste y es capaz de capturar adecuadamente las preferencias de los usuarios.
- Un RMSE menor a 0.5 en una escala de 0 a 1 suele considerarse un desempeño aceptable en sistemas de recomendación, ya que refleja que las predicciones están razonablemente cerca de las valoraciones reales.

Limitaciones del RMSE

- No refleja necesariamente la calidad percibida de las recomendaciones por parte de los usuarios, ya que no mide la relevancia subjetiva de los juegos sugeridos.
 - Se centra únicamente en la precisión de las predicciones numéricas, sin considerar factores como la diversidad o la novedad de las recomendaciones.
-

2. Recomendación Demográfica y Basada en Contenido

Estos modelos no cuentan con métricas de ajuste cuantitativas, ya que su desempeño depende más de su capacidad para ofrecer recomendaciones relevantes que de predicciones numéricas.

Recomendación Demográfica

- Basado en la popularidad global, este modelo es útil para usuarios nuevos sin historial previo en la plataforma (el problema del *cold start*).
- Su éxito radica en identificar juegos ampliamente reconocidos y populares, pero puede carecer de personalización.

Recomendación Basada en Contenido

- Este modelo es efectivo para usuarios que buscan juegos similares a los que ya les gustan, utilizando atributos como género, descripciones y etiquetas.
- Es menos dependiente de datos históricos de interacción entre usuarios, pero podría limitarse si los atributos disponibles no capturan completamente las preferencias individuales.

Conclusiones

El desarrollo del sistema de recomendación para juegos de Steam demostró la capacidad de los modelos de recomendación para mejorar la experiencia del usuario al facilitar el descubrimiento de contenido relevante. A través de la implementación de tres enfoques distintos, se logró cubrir diversas necesidades y comportamientos de los usuarios.

1. Filtrado Colaborativo:

Este modelo fue el más robusto en términos de personalización, logrando un bajo error de predicción ($RMSE = 0.34$) en una escala de 0 a 1. Su capacidad para identificar patrones compartidos entre usuarios permitió ofrecer recomendaciones altamente ajustadas a las preferencias individuales.

2. Recomendación Demográfica:

Al basarse en la popularidad global de los juegos, este modelo fue especialmente útil para usuarios nuevos sin historial previo de interacción, abordando el problema del *cold start*. Aunque carece de personalización, proporciona recomendaciones confiables y ampliamente aceptadas.

3. Recomendación Basada en Contenido:

Este enfoque destacó por su capacidad para conectar juegos a través de atributos descriptivos como géneros, descripciones y etiquetas, permitiendo explorar contenido similar al ya conocido por el usuario. Fue particularmente efectivo para usuarios que buscan afinidades temáticas.

1. Impacto del Sistema

El sistema de recomendación tiene el potencial de aumentar la satisfacción del usuario al proporcionar contenido relevante, diverso y accesible. Al combinar los tres modelos, se garantiza un enfoque integral que:

- Responde a las preferencias individuales.
 - Ayuda a usuarios nuevos o con poca actividad previa.
 - Fomenta la exploración de nuevos juegos basados en similitudes.
-

2. Limitaciones y Futuras Mejoras

A pesar de los buenos resultados, se identificaron áreas para mejorar:

- **Evaluación cualitativa:** Incorporar retroalimentación de usuarios reales para evaluar la percepción subjetiva de las recomendaciones.
- **Diversidad y novedad:** Explorar estrategias que prioricen sugerencias novedosas para evitar recomendaciones redundantes.
- **Datos adicionales:** Incluir más atributos o fuentes de datos, como precios, descuentos o modos de juego, para enriquecer las recomendaciones.

En conclusión, este sistema representa un paso significativo hacia una experiencia más personalizada y enriquecedora para los usuarios de Steam, con oportunidades claras para seguir optimizando su desempeño y aplicabilidad.

Bibliografía

- Fronkon Games. (n.d.). *Steam Games Dataset* [Dataset]. Kaggle. Recuperado de <https://www.kaggle.com/datasets/fronkongames/steam-games-dataset/data>
- Ricci, F., Rokach, L., Shapira, B., & Kantor, P. B. (Eds.). (2015). *Recommender Systems Handbook*. Springer.
- Python Software Foundation. (n.d.). *Python Language Reference*. Recuperado de <https://www.python.org>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., & others. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830.