



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

IIC3633 — Sistemas Recomendadores — 2024-02

## Análisis & Baselines

Nombre:	Fernanda Pérez Hargreaves	Mathias Madsen Sánchez
RUT:	20.428.286-2	20.415.980-7
Numero de alumno:	19640315	19623925

Nombre:	Felipe Olivares Labarca	Beatriz Errázuriz Camus
RUT:	20.075.251-1	20.165.959-0
Numero de alumno:	19642644	19638906

### 1. Descripción del problema y su justificación

Los sistemas de recomendación actuales se basan principalmente en el comportamiento pasado del usuario para generar sugerencias personalizadas. Sin embargo, esta estrategia a menudo conlleva una falta de diversidad en las recomendaciones, lo que limita el descubrimiento de nuevos contenidos o categorías de interés. Por ejemplo, si un usuario ha mostrado preferencia por canciones de rock, el sistema puede enfocarse desproporcionadamente en recomendar canciones de ese género, ignorando otros que también podrían ser de su agrado, como la música clásica. Esta falta de diversidad puede empobrecer la experiencia del usuario, reduciendo la exploración de nuevos contenidos y ocasionando una experiencia repetitiva y predecible.

Este proyecto tiene como objetivo abordar esta limitación mediante el diseño y la implementación de una métrica que permita capturar y medir la diversidad en las recomendaciones. Se busca proporcionar una herramienta útil tanto para desarrolladores como para analistas, optimizando los sistemas de recomendación y mejorando la calidad de las sugerencias generadas. En particular, se pretende evaluar el impacto de las recomendaciones no diversificadas en plataformas populares como Spotify, donde los algoritmos tienden a reforzar patrones de consumo preexistentes.

La métrica propuesta no solo permitirá un análisis más profundo de la diversidad, sino que también proporcionará una base para desarrollar nuevos enfoques que fomenten el descubrimiento de contenido, ofreciendo una experiencia más rica y variada. Esto resulta fundamental en un entorno donde la personalización no debe comprometer la capacidad del usuario para explorar y descubrir nuevas opciones.

Además, se realizará un análisis exhaustivo (survey) de las métricas de diversidad existentes, tanto en dominios generales como específicos de música, para identificar en qué medida la métrica propuesta ofrece mejoras sobre las métricas actuales y cubre áreas que no se estaban midiendo previamente.

### 2. Objetivos del proyecto

#### 2.1. Objetivo principal

Desarrollar una métrica que permita medir la diversidad en las recomendaciones generadas por un sistema de recomendación. El objetivo es evaluar si el sistema ofrece una variedad equilibrada de opciones que representen los múltiples intereses del usuario, en lugar de centrarse exclusivamente en un subconjunto predominante de su historial de preferencias.

## 2.2. Objetivos específicos

- Definir formalmente una nueva métrica de diversidad, considerando diferentes dimensiones de la variedad en las recomendaciones, como la diversidad de contenido, géneros o categorías.
- Implementar la métrica en un sistema de recomendación y probar su rendimiento utilizando un conjunto de datos públicos relevante, como aquellos de música o películas.
- Comparar la métrica propuesta con otras métricas existentes para la evaluación de la diversidad, analizando sus ventajas y limitaciones.
- Analizar el impacto de la nueva métrica en la experiencia del usuario, evaluando si un aumento en la diversidad percibida mejora la satisfacción del usuario y fomenta el descubrimiento de nuevos contenidos.

Este enfoque permitirá explorar cómo las métricas actuales miden la diversidad en las recomendaciones y en qué medida la métrica propuesta ofrece una nueva perspectiva o mejoras. Además, se evaluará el comportamiento de la diversidad a lo largo del tiempo, con el fin de ofrecer un análisis más profundo del impacto en la experiencia del usuario.

## 3. Análisis descriptivo de datos

### 3.1. Dataset MovieLens

El dataset de MovieLens (ml-32m) contiene información sobre calificaciones en una escala de 5 estrellas y etiquetas aplicadas a películas. A continuación, se presenta un análisis de los archivos principales que componen este conjunto de datos:

- **ratings.csv**: Contiene 32,000,204 calificaciones realizadas por 200,948 usuarios sobre 87,585 películas. Los usuarios fueron seleccionados aleatoriamente, y todos han calificado al menos 20 películas. El formato de los datos es el siguiente:
  - **userId**: Identificador único del usuario.
  - **movieId**: Identificador único de la película.
  - **rating**: Calificación otorgada (en una escala de 0.5 a 5.0 estrellas).
  - **timestamp**: Fecha y hora en que se realizó la calificación (en segundos desde el 1 de enero de 1970).

La distribución de calificaciones en este dataset es fundamental para entender las preferencias de los usuarios y la relevancia de ciertos géneros y películas. Un análisis preliminar sugiere que la mayoría de los usuarios tiende a calificar películas con 3 o más estrellas, indicando una inclinación positiva en la evaluación.

- **tags.csv**: Contiene 2,000,072 etiquetas generadas por los usuarios para describir películas. El formato es el siguiente:
  - **userId**: Identificador único del usuario.
  - **movieId**: Identificador único de la película.
  - **tag**: Etiqueta generada por el usuario.
  - **timestamp**: Fecha y hora en que se aplicó la etiqueta.

Las etiquetas representan una forma valiosa de capturar las impresiones y características subjetivas que los usuarios asocian con las películas. Estas etiquetas pueden proporcionar información útil sobre los gustos de los usuarios más allá de las calificaciones numéricas.

- **movies.csv**: Este archivo contiene información sobre las películas incluidas en el dataset. El formato es el siguiente:

- **movieId**: Identificador único de la película.
- **title**: Título de la película (incluyendo el año de lanzamiento entre paréntesis).
- **genres**: Lista de géneros asociados a la película (separados por barras verticales).

Los géneros más comunes incluyen *Drama*, *Comedy*, y *Action*, lo que refleja la preferencia de los usuarios por estos tipos de contenido. Sin embargo, también se observa una gran cantidad de películas con múltiples géneros, lo que puede sugerir una mayor diversidad en las preferencias de los usuarios.

- **links.csv**: Contiene identificadores que permiten enlazar las películas con bases de datos externas como IMDb y The Movie Database (TMDb). Esto facilita el acceso a información adicional sobre las películas.

### 3.2. Dataset Last.fm

El dataset de Last.fm contiene 166,153 registros que documentan las preferencias musicales de los usuarios entre el 1 y el 31 de enero de 2021. A continuación, se presenta un análisis de los principales atributos:

- **Username**: Identificador único del usuario.
- **Artist**: Nombre del artista cuya música fue escuchada.
- **Track**: Nombre de la canción reproducida.
- **Album**: Nombre del álbum al que pertenece la canción.
- **Date**: Día en que el usuario escuchó la canción (entre el 1 y el 31 de enero de 2021).
- **Time**: Hora exacta en que el usuario escuchó la canción.

Este conjunto de datos permite un análisis detallado de los hábitos de escucha de los usuarios, incluyendo la distribución de artistas y géneros musicales. Además, la información temporal puede proporcionar insights sobre los patrones de consumo a lo largo del día y la semana.

### 3.3. Conclusiones del análisis descriptivo

Ambos datasets ofrecen información rica y variada que permitirá evaluar la diversidad en las recomendaciones generadas por los sistemas de recomendación. En el caso de MovieLens, el análisis de las calificaciones y etiquetas ofrece una visión de las preferencias cinematográficas de los usuarios, mientras que el dataset de Last.fm permite estudiar la diversidad de los hábitos musicales. Estos datos serán clave para la implementación y evaluación de la nueva métrica de diversidad que se desarrollará en este proyecto.

## 4. Implementación de modelos

La implementación de los modelos incluirá una combinación de enfoques baselines y avanzados para evaluar la métrica de diversidad propuesta. Los modelos se implementarán utilizando bibliotecas comunes de sistemas recomendadores como **PyRecLab**, **Surprise** y otros frameworks de aprendizaje automático.

## 4.1. Modelos Baseline

Los siguientes modelos de referencia se utilizarán como puntos de comparación:

- **Random Recommendations:** Este modelo generará recomendaciones completamente aleatorias. Servirá como una referencia para observar cómo se comporta la métrica de diversidad en un entorno donde no hay información ni personalización.
- **Most Popular:** Las recomendaciones se basarán en los ítems más populares entre todos los usuarios. Este modelo es ideal para evaluar cómo la popularidad influye en la reducción de la diversidad en las recomendaciones.
- **Collaborative Filtering (Matrix Factorization):** Este es un enfoque clásico que predice las preferencias de los usuarios basándose en las interacciones de usuarios similares. Se usará para observar cómo maneja la diversidad en comparación con otros modelos más avanzados.

## 4.2. Modelos Avanzados

Además de los baselines, se probarán modelos más complejos y especializados:

- **Context-aware Recommendations:** Modelos que integran el contexto del usuario (como la hora del día, la ubicación geográfica, etc.) en el proceso de recomendación. Estos modelos pueden aumentar la diversidad al adaptarse a los cambios contextuales en las preferencias del usuario.
- **Sequence-aware Models:** Estos modelos tienen en cuenta el orden secuencial en que los usuarios consumen ítems. Modelos como las redes neuronales recurrentes (RNN) se usarán para capturar la dependencia temporal y su impacto en la diversidad de las recomendaciones.
- **Hybrid Systems:** Una combinación de técnicas colaborativas y basadas en contenido. Estos modelos permiten aprovechar tanto las similitudes entre usuarios como las características del contenido para aumentar la diversidad.
- **DeepFM (Deep Factorization Machines):** Un modelo avanzado que combina factorización matricial y redes neuronales profundas. Este enfoque es útil para capturar relaciones complejas entre ítems y usuarios, lo que podría influir en la diversidad de las recomendaciones.

## 4.3. Evaluación de la Métrica de Diversidad

La métrica de diversidad se implementará como una evaluación adicional junto con las métricas de precisión tradicionales como RMSE y MAE. Se utilizarán diferentes datasets con metadatos variados (género musical, popularidad, etc.) para medir cómo la nueva métrica captura la diversidad en comparación con las métricas existentes, como la **entropía de Shannon**, **long tail**, y **novedad**.

La evaluación se realizará en al menos dos datasets: **MovieLens** para películas y **Last.fm** para música. Se analizarán diferentes dimensiones de diversidad, como:

- **Diversidad de géneros** (musicales o cinematográficos).
- **Diversidad en la popularidad** (long tail).
- **Diversidad temporal** (cambios en las recomendaciones a lo largo del tiempo).

## 5. Planificación para la etapa Midterm

Esta de forma de tabla en el Cuadro 1, el cual establece fechas y descripción de lo que se hará, después seguimos con los criterios de evaluación y los métodos de seguimiento. Se realizarán reuniones semanales online para revisar el progreso, discutir avances, y resolver problemas. Estas reuniones asegurarán que el equipo esté alineado y se sigan los tiempos de entrega planificados.

## 5.1. Criterios de evaluación

Los criterios de evaluación para este proyecto incluyen:

- **Definición de la métrica:** Claridad en la definición y justificación comparada con métricas actuales.
- **Implementación de los modelos baseline:** Modelos implementados de manera correcta y probados con resultados preliminares.
- **Funcionamiento de la métrica de diversidad:** Métrica correctamente integrada y mostrando diferencias claras con respecto a métricas tradicionales.
- **Evaluación de modelos avanzados:** Comparación adecuada entre los modelos baseline y avanzados en términos de precisión y diversidad.
- **Detección y ajuste de problemas:** Capacidad de identificar problemas y proponer ajustes pertinentes.
- **Calidad del informe Midterm:** El informe debe incluir un análisis claro de los resultados obtenidos y justificación de los ajustes propuestos.

## 6. Bibliografía relevante

- Adomavicius, G., & Kwon, Y. (2012). Improving aggregate recommendation diversity using ranking-based techniques. *IEEE Transactions on Knowledge and Data Engineering*.
- Castells, P., Hurley, N. J., & Vargas, S. (2015). Novelty and diversity in recommender systems. *Recommender Systems Handbook*.
- Kaminskas, M., & Bridge, D. (2017). Diversity, serendipity, novelty, and coverage: A survey and empirical analysis of beyond-accuracy objectives in recommender systems. *ACM Transactions on Interactive Intelligent Systems*.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*.
- Ziegler, C. N., McNeely, S. M., Konstan, J. A., & Lausen, G. (2005). Improving recommendation lists through topic diversification. *Proceedings of the 14th International Conference on World Wide Web*.
- Hu, Y., Koren, Y., & Volinsky, C. (2008). Collaborative filtering for implicit feedback datasets. *IEEE International Conference on Data Mining*.
- Johnson, J., & Ranganathan, K. (2019). DeepFM: A factorization-machine based neural network for CTR prediction. *Proceedings of the 26th International Conference on Neural Information Processing*.

Semana	Fechas	Título asociado al Midterm	Descripción de las tareas
<b>Semana 1-2</b>	30 de octubre - 5 de noviembre	<b>Progreso en el desarrollo de la solución propuesta</b>	Definir la métrica de diversidad, realizar un <i>survey</i> sobre métricas actuales y preparar datasets con metadatos diversos.
<b>Semana 3</b>	6 - 11 de noviembre	<b>Progreso en el desarrollo de la solución propuesta</b>	Implementar los primeros modelos baseline (Random, Most Popular, Collaborative Filtering).
<b>Semana 4-5</b>	12 - 19 de noviembre	<b>Experimentación realizada y evaluación intermedia</b>	Implementar la métrica de diversidad y realizar pruebas preliminares con los modelos baseline.
<b>Semana 6</b>	20 - 26 de noviembre	<b>Análisis preliminar de los resultados obtenidos</b>	Evaluar los modelos adicionales (Matrix Factorization, Context-aware, Sequence-aware, DeepFM) y realizar un análisis comparativo de los resultados en términos de diversidad y precisión.
<b>Semana 7</b>	27 de noviembre - 3 de diciembre	<b>Problemas identificados durante el proceso</b>	Detectar posibles problemas con la métrica o los modelos durante la experimentación y planificar ajustes.
<b>Semana 8 (Midterm)</b>	4 - 8 de diciembre	<b>Revisión del plan propuesto en etapa anterior y justificación de ajustes</b>	Preparar el informe para el Midterm con análisis de resultados preliminares y justificación de los ajustes propuestos en la métrica o los modelos.

Cuadro 1: Planificación del proyecto con títulos asociados al Midterm.