

# Proyecto de Simulación e Inteligencia Artificial (SIA)

Autores:

Marco Antonio Ochil Trujillo

Kevin Majim Ortega Alvarez

Lidier Robaina

Problema: Evaluación de un sistema de recomendación de películas simulando la reacción de los usuarios

## **Introducción:**

En este proyecto, nos propusimos mejorar la efectividad de los sistemas de recomendación a través de la simulación de la interacción de usuarios y sistemas. Para lograr este objetivo, hemos diseñado un enfoque integral que combina temas de búsqueda, conocimiento y procesamiento de lenguaje natural.

## **El Problema:**

Se tiene un sistema de recomendación de películas, al cual el usuario, le ingresa a partir de una selección, que películas le han gustado y en base a esos gustos, devuelve otro grupo de películas que recomienda. Dado la salida de ese sistema, se desea simular la reacción del usuario para determinar, cuales películas pueden gustarle y cuales no, para devolver esa reacción al sistema de recomendación como base para su retroalimentación, y para evaluar, que tan bueno es ese sistema de recomendación, puesto que las reacciones de los usuarios son cruciales, para optimizar las recomendaciones dadas por el sistema, y ofrecer una experiencia personalizada y satisfactoria.

Para resolver este problema contamos con una base de datos, que contiene la información de las películas, así como otras dos, que poseen información acerca de los géneros de las películas y los comentarios de los usuarios a ellas.

## **Técnicas utilizadas:**

Creamos un sistema de recomendación básico que se encarga de obtener la entrada del usuario y devolverle una serie de películas basadas en sus gustos. Al inicio teníamos un sistema con un modelo que calculaba la distancia de Jaccard [1]; dado que al ser impartido en el curso de Sistemas de Recuperación de Información este semestre, poseíamos bastante conocimiento del mismo, además de ser un modelo sencillo, pero muy efectivo, por lo que su implementación se nos facilitó, para realizar las primeras pruebas.

Para reaccionar a las recomendaciones de ese sistema, creamos un agente reactivo, basado en creencias (Un agente BDI), el cual toma las películas que el usuario selecciona y le envía al sistema; como sus creencias iniciales, en una de las bases de datos, guardamos cuanto posee de cada género una película. Luego con esa información se generan las creencias del agente, para que dadas las películas que son devueltas por el sistema, creamos una especie de similitud de coseno entre las películas que posee en su base de creencias y las recomendadas, de esta manera el agente reacciona a ella si la película es lo suficientemente similar. Por suficientemente similar nos referimos a, nosotros calculamos la similitud entre las películas que posee en las creencias, y utilizamos como umbral ese valor, si una película supera ese valor entonces es una película lo suficientemente cercana a sus gustos como para gustarle, luego se añade a sus creencias y se vuelve a calcular la misma.

Otro agente que utilizamos para la reacción del usuario es un agente que como base para sus creencias y su reacción utiliza lógica difusa, simulando un sistema experto. Este agente en su base de creencias posee, actores, descripciones y directores de las películas elegidas por el usuario como sus gustos, además de varios chats, con un modelo de lenguaje (Gemini [2]) con el cual, enviándole la información que el posee en su base de creencias verifica que tanto puede gustar la información que esta recomendando el sistema (Esta información son los actores, descripción y directores de las películas). Entonces una vez obtenida la comparación (un valor de 0 a 10), utilizamos inferencias en lógica difusa para verificar si una película es de su gusto.

El sistema difuso, está compuesto por las siguientes variables de entrada, que son las más relevantes en el momento de investigar una película:

- Actores
- Descripción
- Director

Dado que como lo que se desea es formar un sistema experto con esta lógica difusa (Hemos visto que se ha realizado anteriormente [3]), estas son las variables de mayor interés, además de que se está simulando la reacción de una persona, la cual no tiene todo el conocimiento de una base de datos. Como todo buen sistema de lógica difusa, necesita reglas de inferencia, las cuales son proporcionadas por un experto (En este caso un cinéfilo, el mismo que nos sugirió esas variables de entrada). Antes de contar con la ayuda de esta persona, propusimos demasiadas reglas para la inferencia, las cuales pudieron ser redundantes en algunos aspectos y daban resultados poco coherentes, luego filtramos esas reglas y creamos algunas otras. Como variable de salida se encuentra que tanto puede gustar la película a la persona. si el valor defucificado, se encuentra en el rango de le encanta la película entonces se añade automáticamente a sus creencias, si no entonces se añade la descripción y se aumenta en uno los contadores para cada actor y director, si le han gustado 3 o más películas con estos entonces se añaden a las creencias.

Luego del sistema de jaccard decidimos hacer recomendaciones, con algoritmos genéticos (investigamos que se podía hacer [4]) y distancia de jaccard. Primero con la distancia de jaccard se recomiendan muchas películas, todas aquellas películas que tienen cercanía con las películas que ingresó el usuario. Luego con el algoritmo genético, se filtran esas películas, permitiendo quedarnos solo con 10 de ellas. Para el proceso de filtrado hay dos variantes, considerar las opiniones de todos los usuarios o considerar las opiniones de los usuarios que estaban en ese umbral del jaccard (Dado que las películas se recomiendan gracias a los gustos de otros usuarios), en ambos algoritmos, se puede modificar la cantidad de generaciones y la población inicial, aunque posee otros parámetros que puede modificar como son, el índice de mutación y el índice de cruzamiento.

Creamos otro agente, el cual trabaja con las opiniones que poseen los usuarios a una película. Este agente, asocia los comentarios de las mismas a números utilizando un modelo pre-entrenado.

El ultimo agente del proyecto es el encargado de decidir, dado la salida de los tres agentes, cual salida es la más factible. Además de que el mismo es el que inicializa el resto de agentes además de enviarle la información

En cuanto a la evaluación, en la base de datos se encuentran varios usuarios los cuales poseen sus películas favoritas. La idea es tomar 10 de los usuarios de la base de datos y correr la simulación para de esta manera, verificar que tan buena es la simulación y el propio sistema de recomendación.

Los datos recopilados se entregarán y explicaran en la exposición del proyecto.

## **Bibliografía:**

1. Bag Sujoy, Kumar Krishna Sri, Tiwari Kumar Manoj (mayo 2019): An efficient recommendation generation using relevant Jaccard similarity. Url: <https://www.sciencedirect.com/science/article/abs/pii/S0020025519300325>
2. Google Developers: Get started with the Gemini API: Python
3. Nuñez Flores, Vergara Ortiz, Bocanegra Garcia (2014): Sistema experto basado en lógica difusa tipo 1 para determinar el grado de riesgo de preeclampsia. Url: <https://revistascientificas.cuc.edu.co/index.php/ingecuc/article/view/341>
4. Silva B, Tsang Ren, Cavalcanti G, Tsang Jyh (2010): A graph-based friend recommendation system using Genetic Algorithm