

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

Inteligencia Artificial

PROYECTO 1:

**Diseño de un sistema de recomendación con
filtrado colaborativo.**

PROFESOR: Estefania Prieto Larios

AYUDANTES:

- **Andrea Fernanda Muñiz Patiño**
- **Edwin Antonio Galván Gámez**
- **Rafael Robles Ríos**

EQUIPO: Comic-OS

INTEGRANTES:

- **González Lara Malinali**
- **Ortiz Medrano Francisco Javier**
- **Serrato Solano Victor Manuel**

INTRODUCCIÓN

En la actualidad, gracias a la existencia de Internet y a su contenido básicamente “infinito” las personas contamos con acceso a información que puede ser (o no) de nuestro interés, sin embargo, esto no siempre fue de esta forma pues antes de la existencia de Internet y de la información a un click desde cualquier lugar las personas se veían limitadas en muchos aspectos, principalmente a medios físicos cercanos a los que tuvieran acceso.

Por ejemplo, una persona podía elegir qué música deseaba escuchar de su propio “catálogo” de CD’S, LP’s o cintas de 8 pistas, esto generalmente se limitaba a una repisa o quizás a un estante con un pequeño número de opciones; con el surgimiento de las plataformas de streaming en Internet ahora tenemos la oportunidad de elegir qué clase de contenido deseamos escuchar de entre millones de canciones de varios miles de artistas, de varios miles de discos... Por lo que si antes una persona se demoraba unos 10 minutos en elegir algo para escuchar de su estante veríamos que ahora ese tiempo no es suficiente para elegir entre millones de opciones. Este ejemplo no solo es aplicable a la música y las múltiples plataformas que podemos elegir, sino también a catálogos de películas y series, o bien, páginas de compras que ofrecen millones de productos.

Todo lo anterior nos lleva a plantearnos una pregunta: ¿cómo elegimos qué tipo de contenido deseamos consumir? (por consumir entiéndase ver, escuchar, etc). En una primera instancia podemos pensar en realizar búsquedas de acuerdo a cierto(s) criterio(s), pero esto no es suficiente. A las grandes empresas que nos ofrecen estos servicios les interesa ganar clientes y mantenerlos por encima de su competencia... entonces, ¿cómo lograr esto? La respuesta es: generando interés en el público a través de recomendaciones, es decir, analizar el contenido que consume una persona “x” con gustos similares y sugerirle a la persona “y” estos contenidos, que tal vez “y” no conozca pero que probablemente le gustaran. Para lograr esto se le permite a la comunidad evaluar cada elemento, y así, construir una inteligencia colectiva que permita al usuario recibir información adaptada a sus gustos.

En el presente trabajo se muestra el trasfondo del funcionamiento de un sistema de recomendaciones basado en filtrado colaborativo así como su implementación.

DEFINICIÓN DEL PROBLEMA

Hablar de un sistema de recomendaciones es hablar de un sistema que puede “crear predicciones”, dichas predicciones deben ser calculadas de alguna manera, no existe una única forma de hacerlo, sino por el contrario, podemos encontrar diferentes maneras de efectuar estas predicciones. Previamente se ha mencionado que en la época actual tenemos acceso a grandes cantidades de información gracias al uso de Internet, sin embargo, esto también conlleva sus desventajas pues en algunas ocasiones al navegar es probable que terminemos divagando y perdiendo el tiempo en información que, quizá no era de nuestro interés.

El problema al que debemos de enfrentarnos derivado de esta situación es que nosotros, como simples humanos, no somos capaces de almacenar, procesar y ordenar toda esa información en nuestra mente de acuerdo a ciertos criterios. Una persona con una excelente memoria podría recordar una parte de la información que revisa pero es imposible que recuerde todo aquello que vio; para muestra planteamos la siguiente pregunta: ¿recuerdas cuál fue la primera búsqueda que realizaste al comenzar el día de hoy? Muy probablemente no, además no es algo que sea relativamente importante de recordar. Es mejor que este tipo de tareas sean realizadas por las mismas computadoras y, en este caso, la solución a esta problemática se centra en el desarrollo de un Sistema de Recomendaciones con filtrado colaborativo.

Este sistema cumplirá la función de actuar como un filtro de toda la información que tenemos disponible al alcance de un clic y así, brindar solo aquella información que ‘probablemente’ le interese a una persona, es decir, se trabaja con predicciones futuras sobre lo que cierta persona podría interesarle; no importa si se trata de una película que se acaba de estrenar, publicidad sobre artículos o las noticias más recientes que le sean relevantes. El objetivo de un sistema de recomendación es sugerir nuevos elementos a un usuario cualquiera con base en sus elecciones previas, así como en las elecciones de otros usuarios que han dado valoraciones parecidas o que cuenta con un historial similar al primer usuario.

Se considera que la implementación de un sistema de este tipo es de vital importancia pues puede ayudar a personalizar la experiencia del usuario en algún sitio web, también puede facilitar las búsquedas que este realice, por lo que la implementación de un buen Sistema de Recomendaciones puede ayudarnos a tener una mayor cantidad de usuarios felices pues tienen a su alcance contenido de su interés y usuarios más felices se convierten en mayores ingresos para nuestra aplicación.

DESCRIPCIÓN DE LA PROPUESTA

Al hablar de un sistema basado en Filtrado Colaborativo podemos encontrarnos con varios algoritmos con los que podemos implementar nuestro sistema, por ejemplo:

- Algoritmos de filtrado colaborativo basados en memoria.
También son conocidos como “algoritmos de vecinos cercanos”; utilizan la base de datos de los elementos y/o usuarios para calcular las similitudes entre ellos este cálculo se usa para realizar predicciones y a su vez generar recomendaciones.

Para realizar este cálculo en primera instancia se emplean técnicas estadísticas para encontrar “vecinos cercanos”, es decir, usuarios con un historial o valoraciones similares a la del usuario actual. Una vez que se tiene la lista de los vecinos se combinan las preferencias de estos para arrojar una nueva lista con los n elementos más recomendables para el usuario en particular.

Una de las principales desventajas radica en que se requiere contar con un número mínimo de usuarios y predicciones para poder efectuar las recomendaciones.

- Algoritmos de filtrado colaborativo basados en modelo.
Primero se desarrolla un modelo de las valoraciones del usuario. El problema se trata como uno de predicción estadística y se calcula el valor esperado para cada ítem de acuerdo a las valoraciones anteriores obtenidas en el modelo. Pueden utilizar una matriz de votaciones para crear el modelo a través del cual establecerán el conjunto de usuarios similares al usuario activo. Algunos ejemplos de estos modelos son los clasificadores bayesianos, las redes neuronales o algoritmos genéticos.

Nosotros proponemos un sistema de recomendación de filtrado colaborativo basado en memoria utilizando **slope-one**. Para realizar el cálculo de la predicción (tal como se ha mencionado antes) de un usuario U se debe de tomar en cuenta la información que se tenga de otros usuarios que tengan como común la votación o valoración que haya realizado el usuario U , así como la información de los demás elementos valorados. Consideramos que esta es la forma más sencilla de implementar un recomendador pues en el fondo se limita a realizar cálculos simples como sumas, multiplicaciones y divisiones.

¿QUÉ RECOMENDAMOS?

Nuestro sistema está pensado para otorgar una recomendación que consta de “las tres mejores opciones” de un tipo de pan que le pueda gustar al usuario, dicha recomendación podrá brindarse de acuerdo a alguna valoración de algún artículo (pan) que este usuario haya realizado previamente.

La base de datos fue creada desde cero por el equipo pero está inspirada en las propuestas que Rafael nos brindó en las clases de laboratorio; mientras que el nos sugirió usar alguna base prediseñada de Comida india, o de ramen, nosotros decidimos recomendar comida pero esta vez con tipos variados de pan, siendo honestos ¿a quién no le gusta el pan? y si un sistema puede hacerte una recomendación sobre qué tipo de pan es más probable que te guste sería maravilloso, más aún, sí al entrar a una nueva panadería vemos esta sugerencia podría ser muy útil, tanto para quien desea vender el pan como para quienes desean probarlo pero no pueden decidirse tan fácil por una sola pieza.

Un sistema de recomendación de pan puede ayudar a pequeños y grandes negocios a incrementar sus ventas pues al contar con un sistema de este tipo se podría atraer a más público por distintos aspectos: aquellos intrigados por el funcionamiento, quienes son indecisos o curiosos de la tecnología que quieran entenderlo o “jugar” con el programa. Con todo esto se podrían crear más registros para la base y eventualmente al contar con un dataset más amplio, el recomendador podría mejorar el porcentaje de aciertos a las recomendaciones.

FUNCIONAMIENTO DEL SISTEMA DE RECOMENDACIÓN.

El algoritmo Slope-One (Pendiente uno) toma el promedio de la diferencia de cada producto, este es utilizado para los productos no evaluados (los que se van a recomendar). A estas diferencias se le suman los valores del usuario al que se le va a brindar la recomendación y se dividen entre la cardinalidad (el número de diferencias por producto).

Como es “pendiente uno” necesitamos hacer uso de una función $f(x) = mx + b$, donde $m = 1$ (la pendiente). Para esto se define la diferencia entre los elementos i y j entre la cardinalidad de las diferencias entre el producto i y sus diferencias con los productos j . Ej. concha - todos los panes restantes con calificación.

$$dev_{i,j} = \sum_{itemUsuario \in P_{i,j}(X)} \frac{itemUsuario_i - itemUsuario_j}{card(P_{i,j}(X))}$$

De la fórmula anterior tenemos que **itemUsuario** está dentro del conjunto de entrenamiento **P** (producto seleccionado vs productos calificados), y **card(P i , j (X))** es la cardinalidad en conjunto de i y j dentro del conjunto de entrenamiento (diferencia de tuplas) y **(itemUsuario i - itemUsuario j)** la diferencia de los ratings de los usuarios sobre sus productos. Con lo anterior ahora hay que hacer la suma de cada desviación con el dato $\neq 0$ del usuario al que se le va a recomendar, una vez que se han obtenido las sumas de todas las desviaciones este resultado se divide entre la cardinalidad de todas las diferencias obtenidas:

$$rating = \frac{1}{card(T_j)} \sum_{i \in T_j} (dev_{i,j} + u_i)$$

EJEMPLO

A continuación se presenta un ejemplo del funcionamiento de nuestro recomendador:

En la siguiente tabla se muestra una posible muestra de la base de datos.

Usuario	Dona	Concha	Bolillo
Paco	4	5	3
Pepe	0	3	4
Pita	4	3	?

Vamos a predecir el valor marcado con ? donde *Pita* aún no ha probado el bolillo pero si la Concha y la Dona. Hay que obtener el promedio de las diferencias, esto será tomando los ratings de los usuarios que han probado los panes que ya probó *Pita*, es decir, tomamos a *Paco* y el *rating de Bolillo*, esa será nuestra referencia contra la *Dona* y lo mismo con *Pepe*. [(3-5) + (4-3)]. Ahora dividimos sobre la cardinalidad de las diferencias ($card = 2$) [(3-5) + (4-3)]/2.

Se repite lo mismo con la *Dona* pero en este caso, no se podrá evaluar contra la *Dona* de Pepe ya que él no ha otorgado una calificación a este pan, por lo tanto, tendremos una diferencia de la forma [(3-4)]/1.

Tenemos que [(3-5) + (4-3)]/2 = -.5 y [(3-4)]/1 = -1, entonces, a estas diferencias le sumamos los valores de *Pita* a las diferencias correspondientes. (-.5) + 4 = 3.5 y (-1) + 3 = 2, ahora corresponde sumar estos dos datos 3.5 + 2 = 5.5 y dividirlos sobre la cardinalidad de los productos evaluados (1/2)(5.5) = 2.25.

$$Rating = \frac{1}{2} \left(\left(\left(\frac{[(3-5) + (4-3)]}{2} \right) + 4 \right) + \left(\left(\frac{[(3-4)]}{1} \right) + 3 \right) \right)$$

Después de efectuar todos estos cálculos tenemos que el valor **2.5** es nuestro posible rating para el bolillo.

Este resultado para una tabla más grande y sin productos evaluados se les obtiene su rating y se recomiendan los productos con ratings calculados más altos.

DESCRIPCIÓN DEL AGENTE Y SUS REAS

Proponemos el uso de un agente basado en conocimiento, cuyo funcionamiento se centra en la “base de conocimientos”, decimos que existen dos acciones, en primer lugar el programa le “dice” a la base de conocimiento (la base de datos de usuarios-panes) lo que ha percibido y a continuación le “pregunta” a esta misma base qué acción debe ejecutar. Es en este segundo paso de responder a la pregunta, donde es necesario realizar un razonamiento extenso acerca del estado actual del “mundo”, es decir, cómo se encuentran las condiciones actuales en el programa, así como los efectos que pueden tener las posibles acciones. Una vez que la acción ha sido elegida, el agente la graba mediante un “decir” y procede a ejecutarla, con este acto es posible saber que la “acción hipotética” ahora es una realidad que se ha ejecutado.

Con base en lo anterior consideramos que las REAS de dicho agente son las que se definen a continuación:

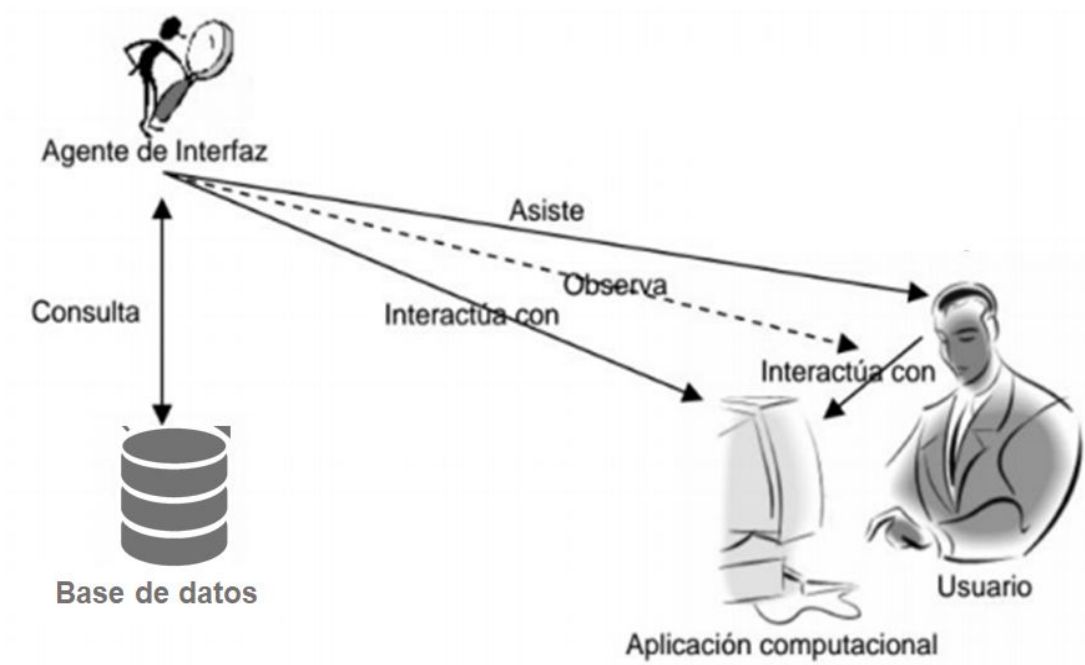
- *Medidas de rendimiento:* Para medir el rendimiento de nuestro sistema de recomendaciones tenemos claro que este debe de entregar una recomendación acertada sobre el tipo de pan a un usuario (cliente), sin embargo, ningún sistema de este tipo cuenta con un 100% de certeza al realizar una sugerencia, por lo que nosotros hemos decidido que una buena medida de rendimiento sería otorgar una certeza del 70%, es decir, que 7 de cada 10 usuarios obtendrán una recomendación adecuada de acuerdo a datos que ingresen y los parámetros que ellos mismos elijan, por lo tanto nuestro margen de error será del 30%.
- *Entorno:* El entorno en el que se desenvolverá nuestro agente será en una panadería con los Clientes que acudan a esta, dichos clientes serán los usuarios por lo que requerimos conocer algunos datos de los usuarios como: nombre, edad, forma de pago, género y la compra del pan que prefieren.
- *Actuadores:* Como actuadores tenemos el mismo programa en sí, pues es este quien se encarga de generar la recomendación, así como la base de datos pues es con esta información previamente almacenada que podemos obtener datos para compararlos y obtener la respuesta final.
- *Sensores:* Como sensores utilizaremos un teclado para el ingreso de los datos solicitados al usuario, también será necesaria el uso de una pantalla o monitor para visualizar los datos (tanto los ingresados como los presentados por el programa).

REPRESENTACIÓN DEL CONOCIMIENTO.

En el esquema que se muestra a continuación podemos observar la interacción que se lleva a cabo entre un usuario y el sistema recomendador:

El usuario interactúa con el programa directamente al escribir datos que el programa utiliza o al visualizar los resultados desplegados en pantalla.

De igual forma el agente interactúa con el usuario al recibir los datos proporcionados por este, también lo “observa” pues es necesario que el agente conozca ciertos datos para poder efectuar la recomendación; finalmente lo “asiste” pues la recomendación proporcionada por este le ayuda a tomar una decisión (en este caso a elegir su pan). No hay que olvidar que el agente debe consultar información predispuesta en una base de datos, pues es justo con esta información previamente recabada y con la nueva proporcionada por el usuario que podrá brindar una recomendación fiable.



La representación del conocimiento se realiza mediante un enfoque en la construcción del sistema, es por esto que se trata de un enfoque declarativo.

Inicialmente, el programa del agente (antes de que empiece a recibir cualquier tipo de percepciones), se construye mediante la adición de una a una de las sentencias que representan el conocimiento del mundo. El diseño del lenguaje de representación permite expresar este conocimiento mediante sentencias simplificando el problema de la construcción de dicho agente.

CARACTERÍSTICAS DEL PERFIL DE USUARIO.

Para realizar una recomendación requerimos que el usuario nos proporcione algunos datos:

- * Nombre.
- * Compra del producto (el tipo de pan).
- * Calificación otorgada al producto.

El nombre es utilizado como un punto de referencia para tener conocimiento del usuario y de esta forma poder darle un trato más personal. Mientras que los datos de sus compras y calificaciones son requeridos para poder efectuar los cálculos mencionados en la parte de **FUNCIONAMIENTO DEL SISTEMA DE RECOMENDACIÓN**.

En este momento no podemos brindar recomendaciones a un usuario completamente nuevo pues no contamos con un punto de comparación, para que la recomendación pueda ser efectuada el usuario debe de contar con al menos una calificación previa de un producto. El objetivo es recomendar productos (panes) de acuerdo a los “gustos” del cliente y para poder determinar esos gustos es necesario saber si alguno de los productos disponibles es o no de su agrado.

De igual forma no es posible almacenar la información generada en la base de datos, esto solo se genera en tiempo de ejecución, es decir, una vez que el programa es cerrado ya no es posible recuperar la información, razón por la cual no podemos decir que nuestro sistema de recomendaciones tenga la capacidad de aprender.

CONCLUSIONES

Los sistemas de recomendación ya son usados ampliamente en Internet y debido al increíble crecimiento de la información que cada año aumenta más, su uso en el futuro será imprescindible.

Si bien es cierto que ya existen múltiples algoritmos que nos permiten brindar recomendaciones certeras, también es cierto que aún no existe uno que sea 100% efectivo, uno de los criterios que podemos aplicar para mejorar los resultados es contar con más atributos de los usuarios en vez de limitarnos solo a estadísticas, pues al tener más puntos de comparación es posible obtener una mejor sugerencia; además la técnica de filtrado colectivo es una de las mejores formas de implementar un recomendador.

BIBLIOGRAFÍA

- Galán Nieto, Sergio Manuel. 2007. Filtrado colaborativo y sistemas de recomendación.
- Xiaoyuan Su y Taghi M. Khoshgoftaar. 2009. A Survey of Collaborative Filtering Techniques.
- Stuart Russell, Peter Norvig. 2004. Inteligencia artificial un enfoque moderno. 2da. Edición.
- Lemire, D., & Maclachlan, A. (2005, April). Slope One Predictors for Online Rating-Based Collaborative Filtering. In SDM (Vol. 5, pp. 1–5).
- Pablo Sardina. 2016. ¿Cuánta Información hay en Internet en el mundo? <https://www.sarpanet.es/es/cuanta-informacion-hay-en-el-mundo/>
- Sección: Agentes de asistencia, pág. 22. <https://www.ridaa.unicen.edu.ar/xmlui/bitstream/handle/123456789/1346/Ramovecchi%2c%20Hern%c3%a1n%20y%20Garc%c3%ada%2c%20Mar%c3%ada%20Sol.PDF?sequence=1&isAllowed=y>