|  |
| --- |
| Universidad veracruzana |
| Sistemas de recomendación en servicios orientados a la salud |
| Diseño de un sistema de recomendación híbrido de servicios profesionales orientados a la salud. |
|  |
| **Mario Hurtado López** |
| **07/05/2019** |

|  |
| --- |
|  |

Tabla de contenido

[Introducción 2](#_Toc6991026)

[1. Sistemas de recomendación 2](#_Toc6991027)

[a. Definición 2](#_Toc6991028)

[b. Técnicas de retroalimentación en sistemas de recomendación 3](#_Toc6991029)

[2. Sistemas de recomendación de filtrado colaborativo 5](#_Toc6991030)

[a. Introducción 5](#_Toc6991031)

[b. Búsqueda y selección de usuarios vecinos 5](#_Toc6991032)

[c. Calculo de similitud entre usuarios 6](#_Toc6991033)

[d. Calificación de ítems y generación de recomendación 7](#_Toc6991034)

# Introducción

La expansión de contenidos multimedia en Internet, el nacimiento del comercio electrónico y el entretenimiento en línea han permitido satisfacer necesidades de los usuarios de este servicio, el Internet ha permitiendo la generación de contenidos de diversas categorías y hacerlos disponibles para todo el mundo a través de la web. Debido a la numerosa cantidad de información que puede ser albergada y creada a cada instante, ha sido necesaria la generación de tecnologías y estrategias que permitan mostrar sugerencias de información a un usuario entre un gran conjunto de datos.

Algunas veces los usuarios buscan ayuda de otros usuarios que han tenido intereses o necesidades similares, o buscan *ítems* que son cercanos a la necesidad que desean satisfacer. Los sistemas de recomendación surgieron con el propósito de mostrar al usuario información relevante a sus necesidades, tratando de evitar una sobre-información entre una gran cantidad de contenido. Empleando técnicas de recuperación de información, los sistemas de recomendación han logrado asistir a los usuarios en la selección de *ítems* basándose en el conocimiento previo de un perfil así como sus intereses.

## Sistemas de recomendación

### Definición

Los sistemas de recomendación son una tecnología que permite ofrecer sugerencias de un producto o servicio que puede ser de potencial utilidad a un usuario, su objetivo es proporcionar información sobre la confiabilidad de un proveedor de servicios. De acuerdo a Petrlic, Lutters, & Sorge. (2014). “Un sistema de recomendación recopila las calificaciones proporcionadas por un socio de transacción permitiendo generar un valor de reputación”. Este tipo de tecnología permite extraer patrones para generar recomendaciones y proveer información útil ayudando en la toma de decisiones.

De acuerdo a Ruchika, Singh, & Sharma, (2017) un sistema de recomendación se define como “un sistema de filtrado de información el cual toma calificaciones de usuarios hacia *ítems* en cuenta y predecir las preferencias de ese usuario*”.* Muchos sitios de comercio en línea generan recomendaciones basándose en retroalimentación implícita y explicita del usuario en busca de un producto o servicio. De acuerdo Lerato, Esan, Ebunoluwa, Ngwira, & Zuva, (2015) “la retroalimentación implícita se basa en patrones de búsqueda del usuario como su historial. Dentro de retroalimentación explicita, la información es obtenida gracias a calificaciones provistas por un usuario”.

Los sistemas de recomendación basan su funcionamiento en el aprendizaje máquina que es una rama de la estadística y la inteligencia artificial, definido de acuerdo a Ruchika, Singh, & Sharma, (2017) como “el proceso en el cual un sistema es capaz de tomar decisiones como un humano”. El aprendizaje es el proceso de ganar conocimiento gracias a estudio o experiencia de funcionamiento.

### Técnicas de retroalimentación en sistemas de recomendación

La retroalimentación es un proceso para recolección de datos generados por un usuario a través de la interacción con una plataforma web, esta a su vez, genera una recomendación en base al historial de interacción de ese usuario. Esta información permite obtener los datos que permitirán una generación de recomendación adecuada y acorde a las preferencias de un usuario.

#### Retroalimentación explícita

De acuerdo a Lerato, Esan, Ebunoluwa, Ngwira, & Zuva, (2015) “la técnica de retroalimentación explicita es el proceso que involucra la asignación de una calificación numérica o *rating* de puntuación para evaluar los *ítems* dados al usuario en una recomendación.” Frecuentemente la retroalimentación explicita se presenta agregada dentro de una escala discreta como *likes* en redes sociales y estrellas en algunos servicios en línea.

Las calificaciones proporcionadas en escalas discretas permiten que una calificación sea procesada estadísticamente permitiendo obtener rangos, promedios, distribuciones y escalas. Este tipo de puntuaciones ayudan a los usuarios a expresar su interés sobre algún *ítem* de manera directa. De acuerdo a Lerato, Esan, Ebunoluwa, Ngwira, & Zuva, (2015) existen 3 enfoques para expresar interés de un usuario en un objeto particular:

Like/dislike: ítems pueden ser clasificados como ‘relevante o irrelevante’ usando una escala de calificación binaria.

Ratings: opinión sobre ítems usando una escala numérica

Comentarios de texto: en este enfoque, los comentarios de un *ítem* son obtenidos y mostrados a los usuarios como ayuda en el proceso de toma de decisiones acerca de un *ítem*, o bien, conocer como una comunidad aprecia ese *ítem* en particular.

Este tipo de retroalimentación permite especificar explícitamente lo que al usuario le gusta y lo que no, así como reforzar las preferencias sobre cierto tipo de *ítems*. Algunas de sus desventajas son la sobre especialización, debido a que refuerza el interés de un usuario sobre un tipo de contenido particular. La intrusión de usuarios es otra de sus desventajas, esta se define de acuerdo a Gunes & Polat, (2016), como “cambio de la predicción de calificaciones de un ítem en particular para afectar la reputación de usuarios genuinos”, lo cual tiende a provocar un mal comportamiento del sistema generador de recomendación.

#### Retroalimentación implícita

La técnica de retroalimentación implícita hace referencia a una recuperación de información sobre el usuario de manera no obstrusiva sobre la valoración acerca de la utilidad para un usuario de un *ítem* en particular. Este tipo de recuperación de información permite la captura e interpretación de información del usuario usando herramientas dependientes del dominio de la aplicación y algunas metodologías.

Está técnica permite la operación y toma de decisiones basándose en las acciones del usuario, este enfoque mide el gusto, conocimiento e interés del usuario sin su consentimiento. Este tipo de retroalimentación puede encontrarse de diversas maneras como el historial web y patrones de búsqueda, incluso, algunas veces puede basarse en el movimiento del mouse sobre diversos í*tems* y la frecuencia en que sucede. De acuerdo a acuerdo Lerato, Esan, Ebunoluwa, Ngwira, & Zuva, (2015) la retroalimentación implícita permite:

Recolección de información a muy bajo costo, el tiempo de uso del usuario permite analizarlo y calificarlo.

Permite la recolección continua de información a través de la interacción del usuario con el sistema, ya que puede ser usada para actualizaciones de sugerencias del usuario.

Permite una calificación efectiva y no intrusiva, difícil de manipular por usuarios.

Este tipo de retroalimentación tiene diversas desventajas como la sensibilidad al contexto de interacción del usuario, esto significa, cada parte del contexto almacenado para generación de recomendación puede ser muy diversa y difícil de interpretar. También esta técnica es menos exacta que la técnica de retroalimentación explicita así como que la retroalimentación será siempre positiva.

#### Retroalimentación híbrida

Es la combinación de la técnica de retroalimentación implícita y explicita, este enfoque utiliza calificaciones numéricas y comportamiento humano para predecir interés de los usuarios. Sus ventajas de acuerdo a Lerato, Esan, Ebunoluwa, Ngwira, & Zuva, (2015) son “crear una predicción eficaz y usa un *rating* y expresiones humanas para crear una recomendación”. Sus deficiencias se encuentran en que no es sencillo de implementar y requiere de procesamiento de información intensivo.

## Sistemas de recomendación de filtrado colaborativo

### Introducción

De acuerdo a Shakirova (2017). “un sistema de recomendación tiene el propósito de modificar la interacción de un humano con Internet: incrementar el grado de interactividad y mejora la experiencia de usuario”. La meta principal de un sistema de recomendación es incrementar el uso de una plataforma o ventas de un producto, debe ofrecer productos que se consideren interesantes a un usuario y mostrar *ítems* que no han sido mostrados antes.

El filtrado colaborativo es una técnica que usa información de los usuarios como *likes, ratings* o interacción de un usuario con el sistema, así como comparación con usuarios que tienen gustos similares. Para generar una recomendación, el sistema aprende las preferencias de un usuario o analiza su perfil. La técnica de filtrado colaborativo recolecta y analiza información respecto a experiencias pasadas, comportamiento, actividades y predice las posibles calificaciones que un *ítem* puedeobtenerbasado en la similitud con otros usuarios.

Está técnica es generalmente usada en redes sociales y comercio electrónico, encargándose de recomendar *ítems* que al usuario podrían interesar. Este tipo de recomendación es muy útil dentro de un gran conjunto de datos, debido a que no aprende nueva información y se limita a predecir calificaciones en *ítems* no vistos mediante la similitud con otros usuarios.

Según lo señalado por a Shakirova (2017). Existen dos formas de recomendación en sistemas de filtrado colaborativo, que pueden definirse como filtrado colaborativo basado en usuario, donde se generan recomendaciones a partir de la similitud entre usuarios calculada a través de gustos o calificaciones en común y un enfoque basado en *ítems,* donde se permite la sugerencia y búsqueda de *ítems* relevantes al usuario que mantengan una relación con otros *ítems* con los que el usuario ha tenido interacción sin la necesidad de buscar similitudes entre usuarios.

### Búsqueda y selección de usuarios vecinos

En un sistema de recomendación basado en filtrado colaborativo, la búsqueda y selección de usuarios se define mediante el algoritmo de agrupamiento k- medias*,* definido como un algoritmo de clasificación de objetos que permite agrupar conjuntos de datos (también llamados *clúster*) que mantienen patrones de comportamiento parecidos. De acuerdo a Farooque, Khan, Junaid, & Gupta, (2014) “El tamaño de un conjunto de datos para selección de vecinos depende del tamaño total de registros de usuarios. Idealmente el tamaño de un *clúster* de datos debería rondar de 30 a 50 usuarios”.

Se han estudiado de acuerdo a Farooque, Khan, Junaid, & Gupta, (2014) que en algunas ocasiones el agrupamiento en *clúster*  de usuarios no tiene muchas ventajas dependiendo de la situación y extensión de la base de conocimientos, debido a que puede incrementar el número de cálculos, por lo que se ha optado por métodos como la selección de todos los vecinos (si el conjunto de datos no es extenso) y selección aleatoria.

### Calculo de similitud entre usuarios

#### Coeficiente de correlación de Pearson

Esta medida de similitud permite calcular la similitud entre dos variables, a su vez, se puede definir como un índice utilizado para medir el grado de relación entre dos variables cuantitativas así como su nivel de correspondencia. De acuerdo a Pujahari & Padmanabhan, (2015) el coeficiente de correlación puede arrojar un nivel de relación entre +1 y -1, cuando este valor es menor a cero, la correlación se toma como negativa, cuanto más próximo a uno es más evidente la relación entre ambas variables. Su método de cálculo se explica en la figura 1.

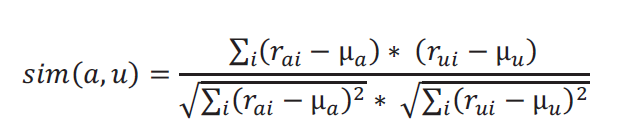


Figura 1 - Fórmula de cálculo de correlación de Pearson

La fórmula de la figura 1 se expresa de la siguiente manera:

* *r =* calificación de un *ítem* por un usuario
* *a =* usuario cuya similitud se va a comparar
* *i* = *ítems* calificados por ambos usuarios ‘a’ y ‘u’ (ítems mutuos)
* *µ =* media o promedio de valoración dado por el usuario

#### Vector de similitud coseno

La similitud coseno puede ser definida de acuerdo a Pujahari & Padmanabhan, (2015) como una medida de similitud existente entre dos vectores en un espacio que posee un producto interior con el que se evalúa el valor del coseno del ángulo entre ambos. Esta función trigonométrica proporciona un valor igual a 1 si el ángulo entre ambos vectores apunta a un mismo lugar. Si ambos vectores apuntan en sentido contrario su valor sería -1, de tal manera que los valores se encuentran en un intervalo cerrado. Su método de cálculo se explica en la figura 2.

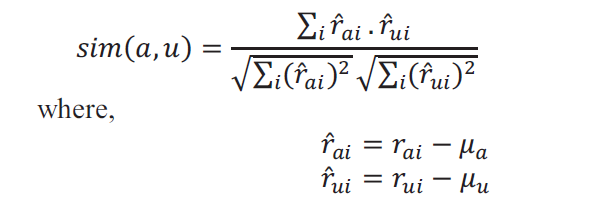


Figura 2 - Fórmula de cálculo de similitud coseno

* *r =* rating e un ítem por un usuario
* *a =* el usuario cuya similitud es comparada
* *u =* usuarios en el vecindario del usuario a
* *i =* ítems calificados por ambos usuarios ‘a’ y ‘u’(ítems mutuos)
* *µ* = media o valoración media por el usuario

#### Diferencia cuadrática media (MSD)

La diferencia cuadrática media es una medida de similitud que ha sido aplicada debido a su gran nivel de precisión, permite adaptarse a diversos entornos de datos (o conjuntos de datos) permitiendo un cálculo basado en la similitud de calificaciones o historial previo de un usuario hacia diversos *ítems* que coincidan con las calificaciones dadas por el usuario que solicita la recomendación.

De acuerdo a Sanchez, Serradilla, Martinez, & Bobadilla, (2008) la métrica de diferencia cuadrática se enfoca en lograr un buen nivel de precisión. Tiene diversas deficiencias como las pocas calificaciones de usuario, cuando un usuario ha calificado pocos elementos muestran un alto nivel de similitud con casi cualquier usuario que coincida con las calificaciones propuestas.

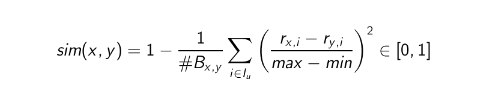


Figura 3 - Fórmula de cálculo de diferencia cuadrática media

* sim(x, y) = similitud que tiene el usuario “x” con el usuario “y”.
* #B x, y = Número de ítems que ambos usuarios han votado.
* r x, i - r y, i = Votos emitidos por los usuarios “x-y” con el ítem “i”.
* max = Calificación máxima que puede ser emitida.
* Min = Calificación que puede ser emitida.

### Calificación de ítems y generación de recomendación

En un sistema de filtrado colaborativo, cuando se han elegido los vecinos para cálculo de similitud y se tienen los valores que representan esa similitud en una escala continua, es necesario generar las posibles calificaciones que el usuario podría asignar a un *ítem* de acuerdo a otros usuarios. De acuerdo a Moya, (2013) “la predicción se puede hacer con la media aritmética de los votos de los vecinos o con la media ponderada de los votos de los vecinos teniendo en cuenta la similitud entre ellos”.

#### Media aritmética

La media aritmética es el valor obtenido de la suma de un conjunto de datos y la división entre el número total de datos. Su método de cálculo se explica en la figura 4.

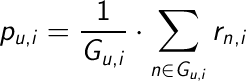


Figura 4 - Cálculo de la media aritmética

* Pu, i = Es la calificación obtenida de la media aritmética.
* Gu, i = Es el valor de K que indica el número de vecinos cercanos utilizados para la predicción de la calificación.
* Rn, i = Es la calificación del usuario “n” en el ítem “i”, siendo los ítems que el usuario actual no ha visto.

#### Media ponderada

La media ponderada es un método de que consiste en dar a cada dato de la evaluación un peso, esto indica la importancia de cada uno de los valores que serán evaluados para generar una calificación de un *ítem.*

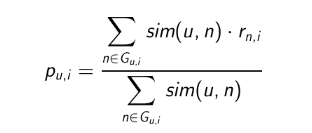


Figura 5 - Cálculo de la media ponderada

* Pu, i = Es la calificación obtenida de la media ponderada.
* Sim(u,n) = Similitud del usuario ‘u’ respecto al usuario ‘n’.
* Rn, i = Es la calificación del usuario “n” en el ítem “i”, siendo los ítems que el usuario actual no ha visto.

**Referencias:**

Petrlic, R., Lutters, S., & Sorge, C. (2014). Privacy-preserving reputation management. En Proceedings of the 29th Annual ACM Symposium on Applied Computing - SAC ’14 (pp. 1712-1718). Gyeongju, Republic of Korea: ACM Press. <https://doi.org/10.1145/2554850.2554881>

Gunes, I., & Polat, H. (2016). Detecting shilling attacks in private environments. *Information Retrieval Journal,* 19(6) 547-572. Recuperado de: <https://doi.org/10.1007/s10791-016-9284-4>

Ruchika, Singh, A. V., & Sharma, M. (2017). Building an effective recommender system using machine learning based framework. 2017 International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions) (ICTUS), 215-219. <https://doi.org/10.1109/ICTUS.2017.8286008>

Lerato, M., Esan, O. A., Ebunoluwa, A.-D., Ngwira, S., & Zuva, T. (2015). A survey of recommender system feedback techniques, comparison and evaluation metrics. 2015 International Conference on Computing, Communication and Security (ICCCS), 1-4. <https://doi.org/10.1109/CCCS.2015.7374146>

Sanchez, J. L., Serradilla, F., Martinez, E., & Bobadilla, J. (2008). Choice of metrics used in collaborative filtering and their impact on recommender systems. 2008 2nd IEEE International Conference on Digital Ecosystems and Technologies, 432-436. <https://doi.org/10.1109/DEST.2008.4635147>

Farooque, U., Khan, B., Junaid, A. B., & Gupta, A. (2014). Collaborative Filtering based simple restaurant recommender. 2014 International Conference on Computing for Sustainable Global Development (INDIACom), 495-499. <https://doi.org/10.1109/IndiaCom.2014.6828187>

Pujahari, A., & Padmanabhan, V. (2015). Group Recommender Systems: Combining User-User and Item-Item Collaborative Filtering Techniques. 2015 International Conference on Information Technology (ICIT), 148-152. <https://doi.org/10.1109/ICIT.2015.36>

Shakirova, E. (2017). Collaborative filtering for music recommender system. 2017 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus), 548-550. <https://doi.org/10.1109/EIConRus.2017.7910613>

Moya, R.. (2013, septiembre 24). Sistemas de Recomendación basados en Filtrado Colaborativo (K-Vecinos). Recuperado 23 de abril de 2019, de Jarroba website: <https://jarroba.com/sistemas-de-recomendacion-basados-en-filtrado-colaborativo-k-vecinos/>