

Universidad del Valle de Guatemala

Campus Central

Facultad de Ingeniería



Algoritmos y Estructura de Datos

PROYECTO SISTEMA DE RECOMENDACIONES

Realizado por:

Gladys De La Roca 15755

Jackeline Hidalgo 15776

Andrea Maybell Peña 15127

Steven Rubio 15044

Catedrático: Lynett Garcia

Sección: 10

Guatemala, 17 de octubre del 2016

ALGORITMOS DE RECOMENDACIÓN

Los algoritmos de recomendación son una subclase de los sistemas de filtrado de información que intentan predecir las preferencias que el usuario puede darle a un producto, estos pueden clasificarse según dos propiedades básicas:

- Tipo de información que utilizan: El contenido del ítem, las preferencias de una comunidad o combinaciones de ambas.
- Cómo se utiliza la información: Que puede ser utilizada directamente (memoria o heurística), o se pueden generar modelos que crean recomendaciones.

Algunos ejemplos de algoritmos de recomendación:

Collaborative filtering

También es conocido como *social filtering*, este algoritmo filtra la información utilizando las recomendaciones de otras personas. Se basa en la idea que es bastante probable que aquellas personas que coincidieron en la evaluación de ciertos elementos en el pasado, coincidan en evaluaciones futuras.

La mayoría de sistemas de filtrado colaborativo utilizan la técnica llamada *neighborhood-based* ("basada en el vecindario"). Para este sistema se toma un grupo de usuarios que han realizado selecciones similares a las del usuario activo. Se realiza una predicción para el usuario activo calculando un promedio ponderado de las calificaciones de los usuarios seleccionados.

Para este caso es muy utilizado el coeficiente de correlación de Pearson, este coeficiente podría definirse como un índice para medir el grado de relación de dos variables siempre y cuando ambas sean cuantitativas.

$$\rho_{X,Y} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y},$$

Donde,

X y Y son dos variables aleatorias sobre una población,

σ_{XY} es la covarianza de (X,Y)

σ_X es la desviación típica de la variable X

σ_Y es la desviación típica de la variable Y

En el filtrado colaborativo suele darse el problema de esparcimiento. Esto ocurre cuando la cantidad de elementos es mucho mayor a la cantidad de usuarios, ya que esto dificulta el realizar las comparaciones entre las selecciones porque hay pocas calificaciones para cada uno de los elementos o ninguna calificación. Algunas soluciones para este problema son: calificaciones implícitas, reducción de la dimensionalidad del espacio de información, descripción del contenido. Esta última solución es una aproximación híbrida entre el filtrado colaborativo y el filtrado basado en contenido.

Content-Based filtering

Este tipo de filtrado se basa en la comparación de contenido de los elementos y el perfil del usuario. Tanto los elementos como el perfil del usuario utilizan palabras clave para su

descripción y estas son utilizadas para poder realizar un análisis de los elementos en comparación con el perfil del usuario y así efectuar las recomendaciones.

Generalmente este tipo de filtrado utiliza documentos de texto para las descripciones. Estos documentos pueden ser representados como vectores con los términos individuales mediante dos técnicas: *vector space model* o *latent semantic indexing*.

Para este método deben considerarse tres aspectos principales. Los términos en las descripciones pueden ser asignados manual o automáticamente. Los términos representativos deben permitir que los elementos y la información de perfil puedan ser comparados en una manera significativa. Debe implementarse un “algoritmo de aprendizaje” que sea capaz de aprender lo que el usuario necesita basado en los elementos que le han interesado y que sea capaz de hacer recomendaciones.

La selección del método de aprendizaje es de suma importancia, ya que este puede determinar la eficiencia del sistema de recomendaciones. La complejidad del algoritmo es un importante indicador de eficiencia, pero también debe tomarse en cuenta la habilidad del método para adaptarse a los cambios en las preferencias del usuario. Un método recomendado es el de *relevance feedback* ya que este tiene una baja complejidad y a la vez necesita poco entrenamiento para adaptarse a cambios.

Hybrid Recommendation filtering

Este tipo de filtrado surge del hecho que por separado tanto el filtrado colaborativo como el basado en contenido tienen ciertas desventajas. En el filtrado basado en contenido existen problemas como: en algunos dominios es difícil obtener una descripción útil del contenido, es demasiado especializado por lo que no le sugerirá al usuario ningún elemento nuevo, este algoritmo tiene dificultad en distinguir entre la información subjetiva y objetiva. En cuanto al filtrado colaborativo, este presenta los siguientes inconvenientes: este sistema no puede brindar recomendaciones para elementos nuevos, problema de esparcimiento, además, si un individuo no encaja con las preferencias de ningún grupo (neighborhood) recibirá recomendaciones incorrectas.

Hay varias maneras distintas de combinar estos tipos de filtrado para crear un sistema híbrido. A continuación se describe una de ellas:

Colaboración vía contenido

El filtro colaborativo chequea la correlación entre las calificaciones de los usuarios para hacer sus predicciones. Se utilizan tanto las calificaciones de los elementos como su contenido para construir el perfil del usuario. Para realizar la selección de términos claves se usa la estrategia del sistema basado en contenido.

Métodos basados en grafos

En estos se utilizan principalmente tres tipos de datos:

- Grafo bipartido de ítems y usuarios:

Enlaces ponderados respecto a la valoración de un usuarios de un ítem, y el número de accesos.

- Grafo de ítems:

Estos ponderan los enlaces cuando tienen usuarios en común, y dicha ponderación es mayor en función al número de usuarios en común.

- Grafo con información social y de etiquetado:

ETAPAS DE DESIGN THINKING

ENTREVISTAS

Diseño entrevistas:

1. ¿En qué ocasiones usted realiza una recomendación?
2. ¿En base a que hace la recomendación, bajo qué criterios?
3. ¿Cómo cree que Netflix le recomienda una película? / ¿Cómo cree que un sitio de citas como match.com le recomienda una pareja?
4. ¿De qué manera pide recomendaciones y a quién?
5. ¿Qué tan confiables son para usted las recomendaciones?

René Marroquín



Observación, cita o historia (1)

“Suelo hacer una recomendación cuando algo me gusta mucho y considero que a otra persona también le puede gustar. Además, las personas suelen preguntarme sobre las cosas que hago y me gustan, le recomiendo cuando me lo piden”.

Posible Significado (1)

Rene nos indica que cuando algo le gusta y es algo sobre lo que conoce, suele decirle a la gente lo que piensa de ello y hacer una recomendación pues es algo que le agrada y que quisiera que otros también conocieran. A esto también comenta que las personas cuando saben que es un tema del que él es conocedor, ellos mismo piden recomendaciones o referencias.

Observación, cita o historia (2)

“Creo que hago mis recomendaciones en base a experiencia propia, ya que solo recomiendo cosas que haya probado, hecho, o visto antes. Solo recomiendo cosas que en verdad considero que valen la pena, como dirían por ahí, *buenas, bonitas y baratas*”.

Posible Significado (2)

Él nos dice que en sus referencias sólo se basa en su opinión y experiencia con las cosas. Es decir que tuvo que haber experimentado primero y tuvo que ser algo muy bueno para recomendarlo, basándose en calidad, accesibilidad y estética.

Jacqueline Rivera



Observación, cita o historia (1)

“Soy un poco complicada en cuanto a las recomendaciones, cuando pido una suele ser a expertos sobre el tema, necesito buenas referencias. Tiendo a invitar por un café a la persona, para poder analizar personalmente lo que me dice”.

Posible Significado (1)

Jacque nos comenta que para pedirle recomendación a alguien necesita que esta persona de verdad conozco del tema, sentirse confiada en que lo maneja bien y que es importante para ella recibir personalmente la recomendación.

Observación, cita o historia (2)

“No confío mucho en las recomendaciones, porque al final no me baso por lo vivido por los demás si no por personal. Y dependería de la persona también, lo tomaría en cuenta si tenemos cosas en común”.

Posible Significado (2)

Ella expresa la desconfianza que tiene hacía las recomendaciones porque considera que la perspectiva de las personas es diferente y prefiere valorar las cosas personalmente, pero cuando toma en cuenta alguna recomendación es porque la persona puede tener gustos y exigencias similares.

Ramon Samayoa



Observación, cita o historia (1)

“Considero que un portal de citas recomienda parejas con algún tipo de algoritmo en el cual se ingresan todos los gustos y características, edad, altura y descripción personal de una persona, y en base a lo que nosotros también ingresamos y por gustos en común y estos datos, obtiene posibilidades y las recomienda”.

Posible Significado (1)

Ramón quiso decir que un portal de parejas toma en cuanto datos sobre la persona como edad, sexo, altura e incluso descripción personal y junto con los datos sobre gustos y preferencias, busca en una base de datos cosas en común o características que una persona tengo que uno haya marcado como gusto, para poder

hacer una recomendación.

Observación, cita o historia (2)

“Pido una recomendación sobre algo como un restaurante, películas o incluso entidades como bancos o aseguradoras, cuando conozco a alguien muy bien y la relación de amistad o profesional me ha permitido ver todos los valores y características que esta persona posee”.

Posible Significado (2)

Él expresa que, para poder confiar en una recomendación, para actividades o incluso preferencia de bancos o aseguradoras, debe conocer muy bien a una persona. Y tomar no solo en cuenta sus gustos sino un conjunto de valores morales que pueda tener esta persona, los cuales conoce de una estrecha relación.

Jorge Suchite



Observación, Cita o Historia

“Recomiendo por valoración, pero siempre que realizó una recomendación las reviso ya que estas son muy imparciales”.
“Una página muy buena de recomendación de películas es Rotten Tomatoes”.

Posible Significado

Las recomendaciones se basan en la valoración que otras personas les dan a las películas.

Enma López



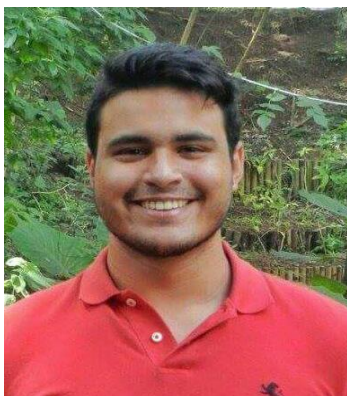
Observación, Cita o Historia

“Netflix recomienda por las películas que miras entonces como ya están categorizadas saben qué características te gustan, lo realizan por medio de base de datos”.

Posible Significado

Las recomendaciones se basan en una base de datos por medio de las películas vistas anteriormente y los géneros parecidos o relacionados con esta.

Alejandro Chinchilla



Observación, Cita o Historia

“Recomiendo películas dependiendo de qué tan populares sean, que tantas vistas y por los actores que aparecen”

Posible Significado

Las recomendaciones se basan en las calificaciones de los críticos y de la categoría de los actores que aparecen en esta.

Jessica Scarlet Peña



Observación, cita o historia

“Pregunto a personas cercanas que compartan mis intereses.”

Posible significado

Para confiar en una recomendación ella necesita que sea de una persona conocida que suele compartir sus mismos gustos para asegurarse que a ella puede gustarle también.

Evelyn de Peña



Observación, cita o historia

“Creo que Netflix hace recomendaciones basándose en los intereses de cada persona y en lo que está de moda.”

Posible significado

Ella cree que las recomendaciones también se basan en lo que está de moda en ese momento, probablemente porque para que algo esté de moda muchas personas deben estar interesadas en ello o considerar que es bueno.

Carlos González



Observación, cita o historia

“Para pedir recomendaciones consultó a alguien que sepa del tema que me interesa y a mis amistades”

Posible Significado

Él piensa que una recomendación debe ser realizada por una persona conocedora para que esta sea confiable y que sea de las mejores recomendaciones.

Berta Valdizon



Observación, Cita o Historia

“Creo que una buena recomendación tiene que venir de alguien, o algo, que me conoce bien más que de algo aleatorio”. “Una buena página asocia perfiles o personas con base a gustos o cosas que les interesen”.

Posible Significado

Ella piensa que una recomendación debe realizarse en base a información concreta que un usuario ha realizado sobre sí mismo o sobre la información que se sabe con certeza de él.

Karin Aroche



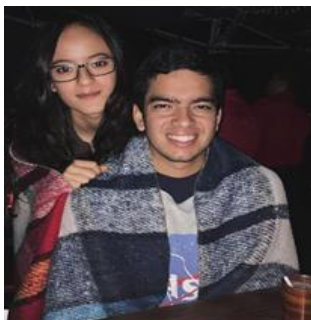
Observación, Cita o Historia

“Yo creo que los sitios realizan recomendaciones en base a lo que las personas dicen de sí mismas o a lo que buscan, pero esto no necesariamente reflejan un dato confiable”, “No creo que un programa sea tan eficiente, ya que se basa en parámetros y características muy específicas y no en información real de la persona”.

Posible Significado

- Ella es algo escéptica en cuanto a cómo trabajan algunos algoritmos de recomendaciones, ya que cree que se hacen en base a muy pocos parámetros que no identifican realmente los gustos de las personas.

Álvaro Mérida



Observación, Cita o Historia

“Hago y sigo recomendaciones todos los días desde: que comer, que deberes, ir a clase o no, etc.”. “Hago una recomendación en base a experiencias pasadas o si me fue bien o no”. “Creo que una página mira que gustos tienes con los datos que llenaste y en base a eso realiza todo tipo de recomendaciones”.

Posible Significado

El creé que el realizar y seguir recomendaciones es algo que te puede ayudar todos los días. Creé que para saber qué tan buena es una recomendación debe ser en base a si en otras ocasiones esta ha resultado o no.

DEFINICIÓN DEL PROBLEMA

Se encontró que existe falta de confianza, por parte de los usuarios, en los algoritmos actuales de recomendaciones. Además, son algoritmos que siguen patrones y toman en cuenta parámetros muy específicos, es decir que las recomendaciones serán posiblemente siempre las mismas, aunque estas sean previamente descartadas.

PROPUESTAS DE IDEAS DE SOLUCIÓN

Una variante del Content-Based filtering, el cual re defina las palabras clave de match en cada entrada y guarde la efectividad de las mismas. También seleccionando las palabras claves que se utilizaran para la recomendación de manera aleatoria, para evitar que siempre sean las mismas recomendaciones tomando en cuenta que pueden ser descartadas.

PROTOTIPOS DE BAJA FIDELIDAD

Durante la etapa de prototipos se realizó, en físico, la interfaz respectiva para el sistema de recomendación de películas, es decir un prototipo en papel para que los usuarios y entrevistados puedan interactuar y ayudar a realizar una retroalimentación de la solución del problema.

Figura 1. Prototipo de baja fidelidad de interfaz del sistema de recomendación de películas.



TESTEO

Los resultados sobre la presentación de los prototipos son los siguientes:

Figura 2 y 3. Testeo del prototipo con personas previamente entrevistadas.



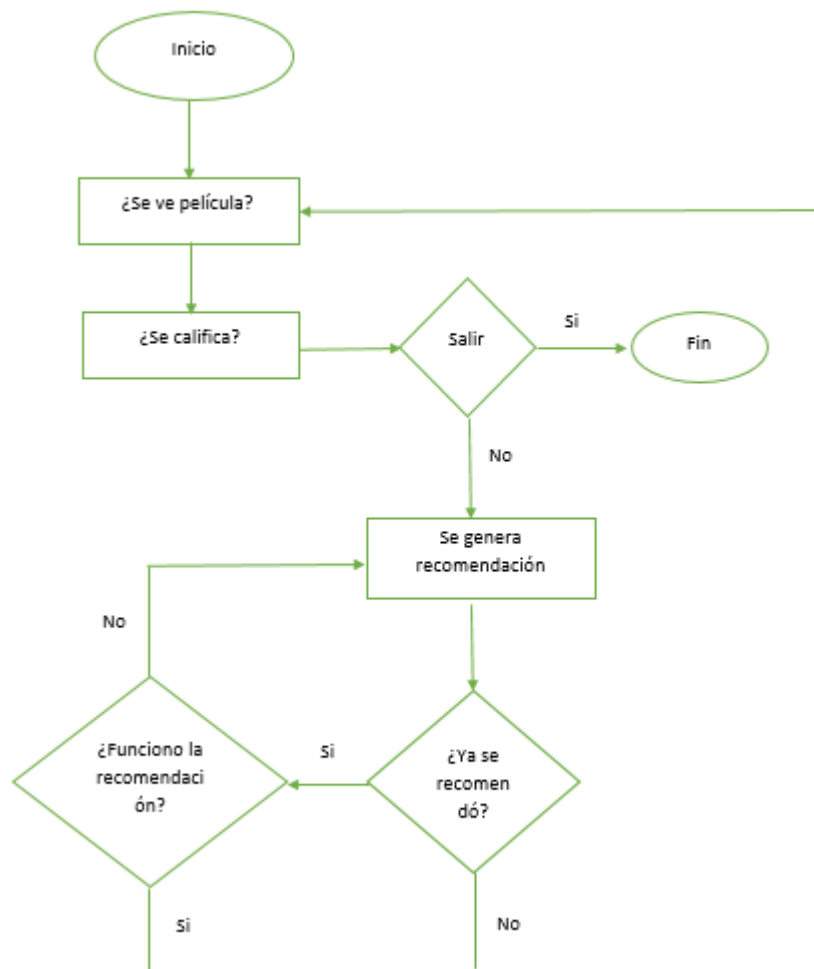
Al presentar el prototipo a las personas previamente entrevistadas se obtuvieron respuestas más significativas que a las personas que solo se les presentó el prototipo. Los previamente entrevistados lograron dar una retroalimentación más extensa, realizando comentarios para mejorar el algoritmo de recomendación, tales como parámetros que podrían usarse y una interfaz con más visibilidad gráfica.

Figura 4, 5 y 6. Testeo del prototipo con personas que no habían sido entrevistadas.



En cuanto a los que solo se les presentó el prototipo los comentarios fueron de mayor aceptación, expresando el gusto por la interfaz y la manera en que se realizaban las recomendaciones.

Diagrama de Flujo



Bases de datos

La base de datos que se utilizará está basada en grafos, es por ello que se realizó en Neo4j que es un software libre orientado a grafos. Debido al tipo de sistema de recomendación elegido, la descripción de las películas y los perfiles de las personas son lo más importante para poder hacer uso de las palabras claves que relacionan a los usuarios con el tipo de películas que les gustan. Para las películas se tienen los siguientes campos para su descripción: Título en español, año de estreno, directores, actores, país y género. Además, en la base de datos se tiene información de personas a las que les ha gustado esta película para que eso nos ayude a armar sus perfiles de usuario conociendo la información antes mencionada de cada película.

Base en siguiente archivo:

[BasePelisP2.xlsx](#)

Referencias:

Recommender-Systems.org (2012) *Recommender Systems* En: <http://recommender-systems.org/content-based-filtering/> [con acceso el 13 de octubre de 2016].