

El algoritmo del insomnio aprende de nuestras costumbres.

Por: Alfredo Moreno. alai. 11/02/2020

El futuro del internauta se predice por el pasado de quienes se le asemejan. No es necesario clasificar la información a partir de los contenidos de los documentos, de las evaluaciones y juicios emitidos por expertos, del reconocimiento de la comunidad de pertenencia. Se trata de calcular el perfil del usuario a partir de los rastros que dejan sus actividades en la red, desarrollando técnicas de Aprendizaje automático que aproximen al máximo el registro de sus gestos.

Para justificar el desarrollo de estas herramientas, las empresas conocidas como GAFAM (Google, Amazon, Facebook, Apple y Microsoft) promotoras del Big Data no hacen lugar a la sensatez y juicios de los humanos, sosteniendo que los expertos cometen errores en sus predicciones. Las técnicas estadísticas de Aprendizaje automático han logrado penetrar en la existencia cotidiana de los internautas.

El Machine Learning o Aprendizaje automático, es una disciplina de las Ciencias de la Computación en el campo de la Inteligencia Artificial. Las máquinas o computadoras aprenden mediante Algoritmos que se entrenan con datos. Este entrenamiento permite encontrar conocimientos en los datos y realizar predicciones mediante métodos estadísticos expresados en los Algoritmos. La calidad y cantidad de datos que entrenan los Algoritmos, mejoran la predicción. Esta es la base de la Ciencia de datos en el contexto del Big Data.

El gran competidor de Netflix es el sueño

Cada vez pasamos más horas pegados a las pantallas y, en ocasiones, hasta dejamos de dormir para estar conectados. Netflix busca nuestro insomnio y los consumidores pasivos “se dejan llevar”. Netflix: “El negocio es quedarnos con las horas de tú descanso”.

No podemos salir de la redefinición digital del mundo, pero podemos ser consumidores inteligentes de modo de preservar nuestra libertad humana.

La cuestión es... ¿a qué [hora consumimos más contenidos](#) en estas plataformas?

normalmente al regreso de trabajo a casa, con aumento en los fines de semana, sobre todo los viernes por la noche y los domingos. Además según las generaciones se consume el contenido por las [multiplataforma](#) del internet móvil. Al llegar el 5G todo este consumo será muy favorecido en velocidad y calidad de las imágenes que se desplegarán en los dispositivos fijos y móviles.

Cada vez que accedemos al servicio de Netflix, el sistema de recomendaciones intenta “ayudarte” a encontrar una serie o una película de tu agrado de la manera más simple. Para calcular la probabilidad de que te interesaría ver un determinado título del catálogo Netflix, utilizan varios factores, algunos de ellos:

- Interacción con el servicio de video por demanda de la empresa, historial de visualización y calificaciones asignadas a otros títulos.
- Actividad de otros miembros usuarios (clientes) con gustos y preferencias similares.
- Información sobre los títulos, como género, categorías, actores, año de lanzamiento, etc.

Además de saber qué vimos en Netflix, también utilizan los siguientes datos para personalizar las recomendaciones:

- Hora del día en que se ve el contenido.
- Dispositivos que se utilizan para ver Netflix.
- Tiempo de permanencia mirando el contenido.

Todos estos datos sirven para alimentar y procesar el Algoritmo de recomendación.

La búsqueda en el catálogo de contenidos es lo más simple y rápida posible. Cuando escribimos un término de búsqueda, los principales resultados que se muestran se basan en las acciones de otros miembros que hicieron la misma consulta o una similar.

El Sistema de recomendaciones

El servicio de vídeo streaming de Netflix, ha producido un Sistema de recomendación de contenidos basado en algoritmos (propios y privados) que utiliza técnicas de Machine Learning (ML). Las mismas tienen por finalidad ir más allá de las ideas preconcebidas de los usuarios, poder ofrecerles alternativas, que, aunque no las habían considerado al principio, podrían ser de su agrado: el valor de predecir

con buenos datos las recomendaciones a la carta.

La primera capa del sistema de recomendaciones de Netflix analiza qué es lo que ha visto un espectador, el orden y las fechas de reproducción, los programas que ha abandonado sin terminar o los que ha marcado como favoritos, entre otros datos similares.

La segunda capa, analiza esta información de uso y se combina con otros parámetros que ayudan a comprender mejor el contenido de los programas.

El personal interno o externo de la plataforma visualiza minuto a minuto todos los contenidos y los etiqueta, los taguea, es decir les define Tags. Este trabajo global realizado por expertos de manera pormenorizada, crea las mejores condiciones para entrenar el Algoritmo del insomnio, y crear las condiciones para el trabajo fantasma

Se analizan los Tags y los datos de comportamiento de los usuarios y sobre esta base aplican algoritmos de ML para conocer lo más importante, las preferencias y los rechazos o abandonos.

Mediante este proceso, la plataforma etiqueta a cada espectador en un tipo de gusto entre los 100.000 que tienen tipificados, y en función de esta categoría, el sistema de recomendaciones muestra unos contenidos u otros. Si bien las etiquetas que utilizan los algoritmos de ML son las mismas para todo el mundo, hay un subconjunto más pequeño que alimenta directamente en la interfaz del usuario, distinta para cada país, idioma y contexto cultural. Respecto a los datos que alimentan los algoritmos, pueden ser explícitos (los que los usuarios facilitan de manera directa) o implícitos (que se pueden inferir de los datos de comportamiento).

Es así como el usuario trabaja gratuitamente para Netflix, alimentando el sistema con sus datos. Nada de esto está explicitado en la letra de contratación del servicio de video por demanda. Los ciudadanos motorizados por el deseo de consumo, avanzamos rápidamente a los sillones de nuestras casas a disfrutar del entretenimiento.

Un sistema de recomendaciones es mucho más que un algoritmo o un filtro que selecciona productos con más o menos acierto. Podemos dividir los Sistemas de Recomendación en 4 partes: la base de conocimiento (la información, los datos), el procesamiento de la base de conocimientos (tecnología, algoritmos, filtros), el

análisis y control de negocio (medir todo, estrategia de negocio) y finalmente la interfaz del usuario.

El peso de cada parte puede estimarse base de conocimiento 25%, procesamiento de la base de conocimientos 5%, analítica y control de negocio 20% e interfaz del usuario 50%.

Netflix aprende de nuestras costumbres

La personalización permite a Netflix ofrecer a cada uno de sus más de 120 millones de suscriptores, la programación que le gusta.

El algoritmo se alimenta de la información que recibe a diario de millones de usuarios, tiene mucha más información de la que pensamos: desde en qué dispositivo se ve Netflix, a qué hora del día se conecta el usuario, cuánto tiempo pasa en la plataforma...

La personalización consiste en eliminar la saturación, adaptando los contenidos de acuerdo al gusto del usuario. Para lograrlo se vale del Big Data y de la Inteligencia Artificial. Netflix usa un algoritmo que sugiere una programación a los consumidores de acuerdo al historial de horas vistas y a sus preferencias. Su plan de expansión global se sustenta en soluciones de “business intelligence” aplicada a los Big Data y a la personalización del consumo.

No existe un “usuario promedio”, sino que cada espectador es único y por eso las opciones cambian cada 24 horas. Para trabajar el perfil del usuario, no importa el género, sino el contenido que el usuario ha visto y si le ha interesado. A partir de ese criterio se crean grupos con gustos similares, pero cada grupo cambia de composición a diario. Para estos agrupamientos se utiliza la técnica de Clustering.

Clustering es una técnica de minería de datos (data mining) dentro de la disciplina de Inteligencia Artificial que identifica de forma automática agrupaciones o clústeres de elementos de acuerdo a una medida de similitud entre ellos. El objetivo fundamental de las técnicas de clustering consiste en identificar grupos o clústeres de elementos tal que:

- La similitud medida entre elementos del mismo clúster sea alta. Similitud intra-clúster alta.
- La similitud medida entre elementos de distintos clústeres sea baja. Similitud

inter-clúster baja.

La identificación de clústeres o grupos de elementos se basa en una medida de similitud. Diferentes medidas de similitud dan lugar a diferentes clústeres.

Según las métricas de la compañía el 80% de los descubrimientos de Netflix surge de las recomendaciones. Esto significa que los suscriptores encuentran nuevos programas y películas favoritas gracias a las herramientas de personalización. Utilizan la tecnología Deep Learning para impulsar su algoritmo. El Deep Learning es una técnica de IA sustentada en la teoría de reconocimiento de patrones.

La técnica se basa en la construcción de conexiones neuronales artificiales, redes neuronales, que trabajan en diferentes computadoras, procesamiento de alta capacidad en paralelo, y analizan grandes cantidades de datos. Estos procesos permiten al Algoritmo “aprender” cada vez entrenado con más datos. A mayor entrenamiento, mejor predicción automática.

Cada cuenta puede contar con varios perfiles porque cada persona es muy distinta en cuanto a gustos. El algoritmo aprende de las costumbres de consumo que permite determinar qué es lo que gusta y así diferenciar los contenidos que hay que mostrar en cada momento.

Una vez que Netflix conoce los gustos de un usuario, el algoritmo le sugerirá lo que puede ver en función de subgéneros que están muy especializados (series de suspenso para hacer maratón, largometrajes familiares de fantasía, series de TV sobre la realeza, por ejemplo).

A veces parece que nos conoce mejor que nosotros mismos.

Netflix hace visibles nuestras preferencias

En la presentación de su contenido, la plataforma no sólo sugiere el contenido que le gusta al usuario. También las imágenes de portada para cada película o serie están determinadas por sus preferencias.

Las ilustraciones de cada contenido, se ofrecen de acuerdo a los hábitos de visualización. Un ejemplo: la película Pulp Fiction. Cómo primera opción para promocionar el film, aparece la imagen oficial con Uma Thurman. Adicionalmente se ofrece otra alternativa que muestra a John Travolta quien también trabajó en la

película de Tarantino. La primera imagen le aparecerá en el menú a las personas quienes tienen en su registro de búsquedas otras películas donde trabajó Thurman. En cambio, la imagen de Travolta será sugerida a quienes prefirieron las películas donde participó el protagonista de Saturday Night Fever.

Netflix no solo selecciona qué serie te sugiere, sino que también cambia la imagen previa entre cuatro o cinco que tiene para elegir. Así la hace más atractiva al cliente según sus gustos, ya que seleccionan la que creen que es más acorde a ellos. Por ejemplo, de Black Mirror tiene fotogramas muy diferentes, más o menos terroríficos.

Como podemos ver en nuestra pantalla de inicio de Netflix, tenemos varias recomendaciones agrupadas por temas según lo que hemos visto en los últimos días. Saben que otros usuarios se han enganchado en cierta serie con otra muy parecida de temática o estilo similar. Así de previsibles somos. La red neuronal se expande asociativamente.

Los consumidores/usuarios pueden **valorar** las series o películas de Netflix con un pulgar hacia arriba o con un pulgar hacia abajo. Dependiendo de la versión de la aplicación que tengas, puedes calificarlo con estrellas, de una a cinco estrellas. Un sistema sencillo, pero también polémico.

Puedes dar pulgares arriba o abajo sin ver la película. Y en función de esto te hará sugerencias más adecuadas. Y si la película no le gusta a la gente, y tiene muchas calificaciones negativas, será eliminada del catálogo.

Para analizar las preferencias y las búsquedas de sus usuarios, se investiga estos datos:

- Cuando pausamos, rebobinamos o adelantamos la reproducción
- Qué días vemos cada cosa
- En qué fechas
- A qué horas
- Nuestra zona geográfica
- Qué dispositivos usamos
- Qué contenidos se abandonan más
- Las puntuaciones que otorgamos
- Las búsquedas
- El comportamiento del usuario en la navegación de los contenidos

Con los algoritmos de análisis de búsquedas, en un mes después de su estreno, *House of Cards* ya era la serie más popular de todo su catálogo.

Gracias a sus algoritmos comprobaron que los trabajos de Kevin Spacey eran muy buscados, así como películas del cineasta David Fincher y que una antigua serie inglesa del mismo nombre, que hablaba de la oscura trastienda del Parlamento, tenía mucho éxito dentro de su público objetivo.

Usando el mismo método que en el caso de *House of Cards*, en años posteriores aparecieron un buen montón de series de producción propia: *Orange Is The New Black*, *Narcos*, *The Get Down*, *Luke Cage*. Series a la carta, los algoritmos ya están entrenados.

El robo del sueño

El sueño es un problema de salud pública, dice la investigadora del Conicet Giannina Bellone, “Tenemos un problema y es de salud pública, como lo decretaron ya entidades como la Asociación de Medicina norteamericana. No se trata sólo de lo que haga el chofer: el Estado y las empresas tienen que comprometerse a mejorar” estas condiciones de trabajo de quienes recorren las rutas de manera profesional, señaló la especialista [ii](#).

El sueño es un problema de salud pública, mientras esperamos las regulaciones de Estado podemos empezar por recuperar nuestras pequeñas libertades y decidir cuando desconectarnos. La libertad nos demanda un esfuerzo a cambio nos permite el buen descanso para mejor uso de nuestras redes neuronales.

Alfredo Moreno

Computador Científico

Profesor de TIC en Universidad Nacional de Moreno

Integrante de www.ticdata.com.ar

@ticdata2

[LEER EL ARTÍCULO ORIGINAL PULSANDO AQUÍ](#)

Fotografía: alai.

Fecha de creación

2020/02/11