Sistemas de Recomendación

Uayeb Caballero Rodríguez

Universidad Nacional Autónoma de Honduras, Ciudad Universitaria, Honduras

July 17, 2017

This section will be for read an abstract in english, it will contain a narrative about what is dimensionality reduction, SVD as applied technique and introducing an example

1 Introduccion

Los sistemas de recomendaciones son herramientas que generan recomendaciones sobre un determinado objeto de estudio, a partir de las preferencias y opiniones dadas por los usuarios. El uso de estos sistemas se está poniendo cada vez más de moda en Internet debido a que son muy útiles para evaluar y filtrar la gran cantidad de información disponible en la Web con objeto de asistir a los usuarios en sus procesos de búsqueda y recuperación de información. En este trabajo realizaremos una revisión de las características y aspectos fundamentales relacionados con el diseño, implementación y estructura de los sistemas de recomendaciones analizando distintas propuestas que han ido apareciendo en la literatura al respecto.

2 Collaborative Filtering

El Filtrado colaborativo (FC) es una técnica utilizada por algunos sistemas recomendadores. En general, el filtrado colaborativo es el proceso de filtrado de información o modelos, que usa técnicas que implican la colaboración entre múltiples agentes, fuentes de datos, etc. [2] Las aplicaciones del filtrado colaborativo suelen incluir conjuntos de datos muy grandes. Los métodos de filtrado colaborativo se han aplicado a muchos tipos de datos, incluyendo la detección y control de datos (como en la exploración mineral, sensores ambientales en áreas grandes o sensores múltiples, datos financieros) tales como instituciones de servicios financieros que integran diversas fuentes financieras, o en formato de comercio electrónico y aplicaciones web 2.0 donde el foco está en los datos del usuario, etc. Esta discusión se centra en el filtrado colaborativo para datos de usuario, aunque algunos de los métodos y enfoques pueden aplicarse a otras aplicaciones.

En el enfoque más reciente, el filtrado colaborativo es un método para hacer predicciones automáticas (filtrado) sobre los intereses de un usuario mediante la recopilación de las preferencias o gustos de información de muchos usuarios (colaborador). El Filtrado colaborativo se basa, en que si una persona A tiene la misma opinión que una persona B sobre un tema, A es más probable que tenga la misma opinión que B en otro tema diferente que la opinión que tendría una persona elegida azar. Por ejemplo, un sistema de recomendación basado en el filtrado colaborativo para televisión podría hacer predicciones acerca de los programas que le gustarían a un usuario a partir de una lista parcial de los gustos de ese usuario (gustos o disgustos). Notese que estas predicciones son específicas para el usuario, pero utilizan la información obtenida de muchos usuarios. Esto difiere del enfoque más simple de otorgarle una puntuación promedio (poco específico) para cada elemento de interés, por ejemplo sobre la base de su número de votos.

2.1 Índice Jaccard

El índice de Jaccard (IJ) o coeficiente de Jaccard (IJ) mide el grado de similitud entre dos conjuntos, sea cual sea el tipo de elementos.
La formulación es la siguiente:
J(A,B) = $|A \cap B| \ / \ |A \cup B|$

Es decir, la cardinalidad de la intersección de ambos conjuntos dividida por la cardinalidad de su unión. Siempre toma valores entre 0 y 1, correspondiente este último a la igualdad total entre ambos conjuntos. En ecología se usa para medir la similitud, disimilitud o distancias que existen entre dos estaciones de muestreo, con una formulación equivalente [1]

2.2 Rango de una matriz

El rango de una matriz es el número máximo de columnas (filas respectivamente) que son linealmente independientes. El rango fila y el rango columna siempre son iguales: este número es llamado simplemente rango de A (prueba más abajo). Comúnmente se expresa como rg(A). El número de columnas independientes de una matriz A de m filas y n columnas es igual a la dimensión del espacio columna de A. También la dimensión del espacio fila determina el rango. El rango de A será, por tanto, un número no negativo, menor o igual que el mínimo entre m y n:

$\mathbb{A} \in M_{mxn} \to 0 \leq rang(\mathbb{A}) \leq min(m,n)$
Ejemplo en R:

```
R code
========
# test
rm(list=ls())
x<-1:10
mean(x)
```

2.3 Eigen valores y eigen espacios

Formalmente, se definen los vectores propios y valores propios de la siguiente manera: Sea $A : V \to V$ n operador lineal en un cierto $\mathbb{K}$-espacio vectorial V y v un vector no nulo en V. Si existe un escalar c tal que

$$\mathbf{A}\mathbf{v} = c\mathbf{v}, \qquad \mathbf{v} \neq \mathbf{0}, c \in \mathbb{K},$$

entonces decimos que v es un vector propio del operador A, y su valor propio asociado es c. Observe que si v es un vector propio con el valor propio c entonces cualquier múltiplo diferente de cero de v es también un vector propio con el valor propio c. De hecho, todos los vectores propios con el valor propio asociado c junto con 0, forman un subespacio de V, el espacio propio para el valor propio c. Observe además que un espacio propio Z es un subespacio invariante de A, es decir dado w un vector en Z, el vector Aw también pertenece a Z.

2.4 Proceso de ortogonalización de Gram-Schmidt

El método de Gram-Schmidt se usa para hallar bases ortogonales (Espacio Euclideo no normalizado) de cualquier base no euclídea. En primer lugar tenemos que:

$$\mathbf{v} - \frac{\langle \mathbf{v}, \mathbf{u} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle}\mathbf{u} = \mathbf{v} - \mathrm{proy}_{\mathbf{u}}(\mathbf{v})$$

Es un vector ortogonal a $\mathbf{u}$. Entonces, dados los vectores $\mathbf{v}_1, \ldots, \mathbf{v}_n$, se define:

$$\mathbf{u}_1 = \mathbf{v}_1,$$

$$\mathbf{u}_2 = \mathbf{v}_2 - \frac{\langle \mathbf{v}_2, \mathbf{u}_1 \rangle}{\langle \mathbf{u}_1, \mathbf{u}_1 \rangle}\mathbf{u}_1,$$

$$\mathbf{u}_3 = \mathbf{v}_3 - \frac{\langle \mathbf{v}_3, \mathbf{u}_1 \rangle}{\langle \mathbf{u}_1, \mathbf{u}_1 \rangle}\mathbf{u}_1 - \frac{\langle \mathbf{v}_3, \mathbf{u}_2 \rangle}{\langle \mathbf{u}_2, \mathbf{u}_2 \rangle}\mathbf{u}_2,$$

Generalizando en k:

$$\mathbf{u}_k = \mathbf{v}_k - \sum_{j=1}^{k-1} \frac{\langle \mathbf{v}_k, \mathbf{u}_j \rangle}{\langle \mathbf{u}_j, \mathbf{u}_j \rangle}\mathbf{u}_j\mathbf{u}_k = \mathbf{v}_k - \sum_{j=1}^{k-1} \frac{\langle \mathbf{v}_k, \mathbf{u}_j \rangle}{\langle \mathbf{u}_j, \mathbf{u}_j \rangle}\mathbf{u}_j$$

2.5 Descomposición en valores singulares

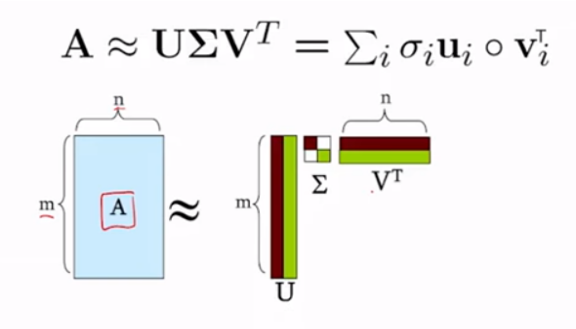
En álgebra lineal, la descomposición en valores singulares de una matriz real o compleja es una factorización de la misma con muchas aplicaciones en estadística y otras disciplinas.

Dada una matriz real $A \in \Re^{m \times n}$, los autovalores de la matriz cuadrada, simétrica y semidefinida positiva $A^T A \in \Re^{n \times n}$ son siempre reales y mayores o iguales a cero. Teniendo en cuenta el producto interno canónico vemos que:

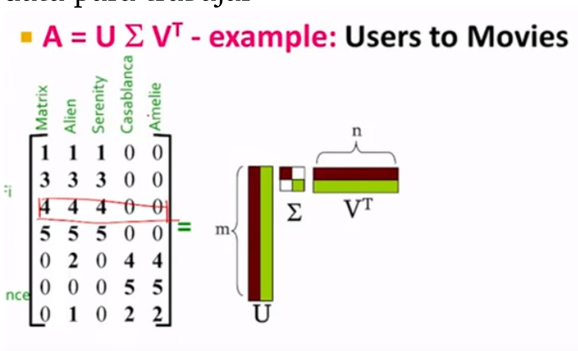
$\left(A^T A\right)^T = A^T \left(A^T\right)^T = A^T A$. O sea que es simétrica
$(Ax, Ax) = x^T A^T Ax = ||Ax||^2 \geq 0$ es decir $A^T A$ es semidefinida positiva, es decir, todos sus autovalores son mayores o iguales a cero.

Si λ_i es el i-ésimo autovalor asociado al i-ésimo autovector, entonces $\lambda_i \in \Re$. Esto es una propiedad de las matrices simétricas.

3 Planteamiento del problema



data para trabajar



References

[1] Real, R., Vargas, J. M. (1996). The probabilistic basis of Jaccards index of similarity. Systematic biology, 45(3), 380-385

[2] Terveen, Loren; Hill, Will (2001). Beyond Recommender Systems: Helping People Help Each Other