

Guía de Laboratorio Nro. 1

Sistemas de Recomendación

y

1. Objetivo:

- Aplicar las métricas de distancia en sistemas de recomendación grandes y multidimensionales.

2. Conceptos básicos:

- a) Distancia de Manhattan

$$D_M = |x_1 - x_2| + |y_1 - y_2|$$

- b) Distancia Euclidiana

$$D_E = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

- c) Distancia de Minkowski

$$D(x, y) = \left(\sum_{k=1}^n |x_k - y_k|^r \right)^{\frac{1}{r}}$$

r=1 -> Distancia de Manhattan

r=2 -> Distancia de Euclidiana

3. Materiales:

- Software de Libre elección: c++, Python, etc.

4. Indicaciones de Desarrollo

4.1. Pensando en N-Multidimensionalidad

Suponga que los usuarios pueden calificar las bandas de música en una escala de 1 a 5. La siguiente tabla muestra 8 usuarios y sus calificaciones para 8 bandas

| | Angelica | Bill | Chan | Dan | Hailey | Jordyn | Sam | Veronica |
|------------------|----------|------|------|-----|--------|--------|-----|----------|
| Blues Traveler | 3.5 | 2 | 5 | 3 | - | - | 5 | 3 |
| Broken Bells | 2 | 3.5 | 1 | 4 | 4 | 4.5 | 2 | - |
| Deadmau5 | - | 4 | 1 | 4.5 | 1 | 4 | - | - |
| Norah Jones | 4.5 | - | 3 | - | 4 | 5 | 3 | 5 |
| Phoenix | 5 | 2 | 5 | 3 | - | 5 | 5 | 4 |
| Slightly Stoopid | 1.5 | 3.5 | 1 | 4.5 | - | 4.5 | 4 | 2.5 |
| The Strokes | 2.5 | - | - | 4 | 4 | 4 | 5 | 3 |
| Vampire Weekend | 2 | 3 | - | 2 | 1 | 4 | - | - |

Los guiones indican que un usuario no calificó a una banda en particular. Para las métricas de distancias tratadas en la práctica, solo se considerarán aquellas que fueron calificadas por ambos usuarios.

4.2. Calcule las siguientes distancias (Usando lápiz y papel)

- Distancia de Manhattan entre Hailey y Verónica
- Distancia Euclidiana entre Hailey y Jordyn
- Distancia de Minkowski entre Angelica y Bill (con $r=0$ y $r=1$)

4.3. Implemente las métricas de distancia del sistema de recomendación

Considere que la estructura de datos que emplee debe ser escalable, tanto en dimensionalidad como en tamaño.

5. Entregables

Al finalizar el estudiante deberá:

5.1. Para la implementación: mostrar el resultado de los valores obtenidos

Fecha de entrega : 20/04/2020

5.2. Crear un github de la implementación del sistema de recomendación y compartir el enlace del mismo.

6. Referencias

Zacharski, R. (2015). A Programmer's Guide to Data Mining: The Ancient Art of the Numerati.