

Sprint 3 - Recomendación

Interacción Sociotecnológica

Daniel Gómez-Marín

En este ejercicio vamos a implementar un sistema muy simple de recomendación de personas y de características usando el algoritmo K Nearest Neighbours¹ (KNN). Este algoritmo es bastante común en sistemas de interacción socio tecnológica para encontrar usuarios similares basados en sus gustos. Lo que necesitamos para implementar el algoritmo es una lista de usuarios (filas) y sus gustos o características (columnas). Cada característica va en una columna diferente y, por simplicidad, está calificada en la misma escala (i.e. de 0 a 1). La idea es que, a través de las características, podamos encontrar personas similares. La hipótesis básica es: “dos usuarios son similares en la medida en que sus características son similares” En este ejercicio vamos a revisar algunas ideas nuevas:

- usuario
- características
- vector de características
- distancia
- vecindario de usuarios
- fuerza del vínculo

1. Crear entre todo el grupo una lista de ingredientes de pizza que estén en un menú de un restaurante famoso de la ciudad. Todos los miembros del salón llenamos una encuesta de ingredientes dándole un valor a todos los ingredientes entre 1 y 10. 1= No me gusta, 10= me encanta.

2. Una vez que todos llenemos la encuesta, creamos un programa de computador para calcular la Distancia entre un usuario y los demás. Para esto típicamente se usa la distancia Coseno, (realizada en el sprint anterior) que consiste en usar el vector de características (todas las respuestas de las características de cada usuario) para medir la distancia entre dos usuarios.

3. Luego, seleccionamos un valor arbitrario de K vecinos (usualmente la raíz cuadrada del número total de usuarios) y con este valor encontramos los K vecinos más cercanos a cada usuario. Es decir escogemos los K usuarios que están a menor distancia del usuario seleccionado. Finalmente, presentamos el usuario seleccionado, sus K vecinos más cercanos y la distancia entre el usuario y sus vecinos cercanos.

¹ https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm

4. La entrega de este ejercicio es un algoritmo funcional cuyas entradas son: el nombre P de una persona del grupo y el valor de K. La salida es el “vecindario de usuarios” o el grupo de K usuarios más cercanos a la persona P de toda la base de datos, indicando el valor de distancia o la “fuerza del vínculo”.

5. Reflexionar qué influencia tiene el algoritmo KNN en un sistema de recomendación en el contexto pospuesto de una pizzería. Identificar por lo menos un contexto nuevo en el cual este algoritmo pueda ser utilizado provechosamente para realizar decisiones para grupos. Para esta reflexión debemos pensar que P, dependiendo de la base de datos, no necesariamente es una persona, también puede ser un producto o un servicio.

Bonus: Quien entregue un ejercicio en donde la fuerza del vínculo entre P y los miembros de K se manifieste de manera espacial y no (solo) numérica, tiene 0.5 más en el ejercicio final de la unidad 3 (que vale el 40% del curso).