UNIVERSIDAD EAFIT
ST0263: TÓPICOS ESPECIALES EN TELEMÁTICA
Proyecto Final – Opción 3: Procesamiento Paralelo - distribuido: Sistemas de recomendación basado en la Correlación de Pearson

Integrantes:
Alejandro Gomez Londoño **cod:** 201010001010
Santiago Palacio Gomez **cod:** 201110021010
Pablo Velásquez Manrique **cod:** 201110059010

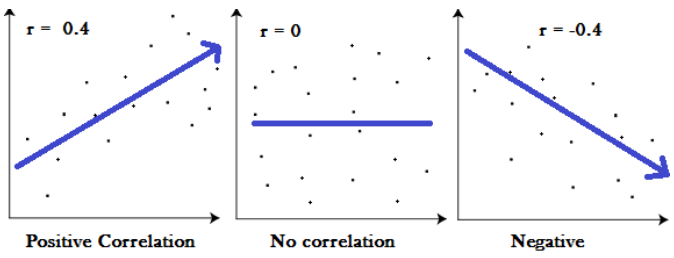
El presente documento da cuenta del diseño e implementación de un Sistema de recomendación basado en la Correlación de Pearson, mediante un algoritmo paralelo que puede ser ejecutado en un clúster MPI, y que permita disminuir el tiempo de procesamiento para la generación de la matriz SR.

En primer lugar, la correlación en conjuntos de datos es medida mediante como están relacionadas. La medida más común de correlación en estadística es la correlación de Pearson. Esta muestra la relación lineal entre dos conjuntos de datos. Y se usa una ρ para representar población y r para una muestra.

$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[\, n\Sigma x^2 - (\Sigma x)^2 \,][\, n\Sigma y^2 - (\Sigma y)^2 \,]}}$$

Los resultados iran de -1 a 1, pero es raro obtener exactamente 0, -1 o 1. Generalmente se obtienen números entre estos. Mientras más cercano sea el valor de r a 0, mayor será la varianza de los datos alrededor de la relación lineal.

Se puede diferenciar las correlaciones de la siguiente manera:
Alta: de 0.5 a 1.0 ó de -0.5 a -1.0
Media: de 0.3 a 0.5 ó de -0.3 a -0.5
Baja: de 0.1 a 0.3 ó de -0.1 a -0.3



Uno de los elementos más relevantes del procesamiento paralelo en contraposición al procesamiento serial, es la disminución de tiempo. en este aspecto fue implementada la solución tanto serial como paralela, arrojando los siguientes resultados:

Solución	**Tiempo de procesamiento (ms)**
Serial	
Paralela	