

Tarea 01 - Tópicos Avanzados en Ciencia de Datos

Implementación de operaciones matriciales para el análisis de enlaces sobre Big Data.

Objetivos.

Ciertas operaciones básicas sobre matrices, como el producto Matriz-vector o Matriz-Matriz, son esenciales en el cálculo de indicadores muy importantes en el análisis de grafos. Con esta asignación se pretende introducir al estudiante en elementos prácticos asociados con el manejo de matrices gigantes mediante el uso del paradigma MapReduce. El cálculo de operaciones Matriz-Vector permite la obtención de vectores como el "PageRank" que permite jerarquizar la importancia de las páginas web y que es la base del análisis de enlaces. Operaciones similares permiten hacer cálculos de centralidad importantes en la detección de, por ejemplo, influyentes en las redes sociales. El objetivo es, dados una matriz A y un vector x , realizar una función que implemente, con MapReduce, el producto Matriz-Vector cuando el vector x por su gran tamaño, no cabe en la memoria RAM. Para esto le haremos llegar una implementación del producto Ax , en la que se asume que el vector x si cabe en la memoria.

Archivo de datos.

Nombre del archivo	Formato	Uso
prod_Mv_type1.R	R	Método de cálculo de Ax , caso en el que x cabe en la memoria RAM
prod_Mv_type1_verif.R	R	Código en R que demuestra el uso de la función anterior
tblAkv3x3.csv	CSV	Matriz de ejemplo A , de dimensión $[3 \times 3]$
tblxkv3.csv	CSV	Vector de ejemplo x , de dimensión $[3]$
tblAkv10x10.csv	CSV	Matriz de ejemplo A , de dimensión $[10 \times 10]$
tblxkv10.csv	CSV	Vector de ejemplo x , de dimensión $[10]$
tblAkv10x10ident.csv	CSV	Matriz identidad $[10 \times 10]$

Actividades a realizar.

1. Implemente una función para el cálculo de una Matriz por Vector, para el caso en que ni la Matriz ni el Vector caben en la memoria RAM, usando el paradigma MapReduce.
2. Implemente una función para el cálculo de una Matriz por Matriz, usando el paradigma MapReduce.

Condiciones.

- Las acciones propuestas deben implementarse en el lenguaje R.
- El funcionamiento correcto del algoritmo será probado con los datos de prueba entregados en esta tarea.
- Se puede realizar individualmente o en equipos de máximo 2 personas.
- Debe ser entregada hasta el viernes 28 de marzo de 2016.
- Elabore un breve informe con los resultados de las acciones propuestas, el código fuente y sus comentarios.