



Becas colaboración curso 2022/2023

Fecha: 01 Junio 2022

Vicerrectorado de Investigación

Subcomisión de I+D+i

Propuesta del departamento *SISTEMAS INFORMÁTICOS Y COMPUTACIÓN*

Núm Proyecto: 2022/32/00007

Responsable

Román Moltó, José Enrique

E-mail

jroman@dsic.upv.es

Ext.

83520

Responsable

Alvarruiz Bermejo, Fernando

E-mail

fbermejo@dsic.upv.es

Ext

83518

Título proyecto

Uso de software paralelo de álgebra lineal en aplicaciones de minería de datos y análisis de redes complejas

Valoración proyecto

4

Descripción proyecto

La minería de datos y el reconocimiento de patrones tienen cada vez más importancia en numerosas aplicaciones, como por ejemplo el reconocimiento de caras, clasificación de texto manuscrito, búsqueda en texto no estructurado, etc. De entre la multitud de métodos disponibles, las técnicas algebraicas suelen ofrecer una buena precisión, pero habitualmente con un coste computacional elevado. Sin embargo, este inconveniente puede compensarse con el uso de la computación paralela, explotando el paralelismo disponible en los computadores convencionales, o bien haciendo uso de grandes supercomputadores en el caso de colecciones de datos extremadamente grandes. El proyecto se centra en métodos basados en factorizaciones parciales (QR, SVD, GSVD, NMF), que puedan ser aplicados a matrices dispersas de gran dimensión. Algunas de las tareas básicas se pueden realizar con las bibliotecas paralelas PETSc y SLEPc, y el objetivo del proyecto será adaptarlas a problemas más concretos como el análisis discriminante lineal (LDA) o el particionado espectral de grafos. Los desarrollos del proyecto se publicarán en forma de software de código abierto.

Actividades a realizar por el alumno

1. Estudio bibliográfico sobre el tipo de problemas a resolver y los métodos disponibles
2. Implementación de prototipos (en Matlab o python) de los posibles métodos a implementar
3. Familiarización en el uso de las bibliotecas PETSc y SLEPc
4. Implementación de algún método de LDA basado en SVD
5. Implementación del particionado espectral de grafos
6. Recopilación de conjuntos de datos de gran dimensión disponibles en internet
7. Validación de los algoritmos implementados mediante diferentes casos de test
8. Comparación con otras herramientas
9. Estudio de rendimiento paralelo, accediendo a un supercomputador



Becas colaboración curso 2022/2023

Fecha: 01 Junio 2022

Localización de la actividad (Campus)

Campus de Vera

Horario

Horario flexible, a convenir con el alumno