

Resumen

Muchos problemas procedentes de aplicaciones del mundo real pueden ser modelados como problemas matemáticos con magnitudes no negativas, y por tanto, las soluciones de estos problemas matemáticos solo tienen sentido si son no negativas. Estas magnitudes no negativas pueden ser, por ejemplo, la concentración de los elementos en un compuesto químico, las frecuencias en una señal sonora, las intensidades de los píxeles de una imagen, etc.

Algunos de estos problemas pueden ser modelados utilizando un sistema de ecuaciones lineales sobredeterminado, es decir, un sistema de ecuaciones con más ecuaciones que incógnitas. Cuando la solución de dicho problema debe ser restringida a valores no negativos, aparece un problema llamado problema de mínimos cuadrados no negativos (NNLS por sus siglas en inglés). La solución de dicho problema tiene múltiples aplicaciones en ciencia e ingeniería, concretamente para resolver problemas de optimización con restricciones de no negatividad.

Otra descomposición no negativa importante es la Factorización de Matrices No negativas (NMF por sus siglas en inglés). La NMF es una herramienta muy popular utilizada en varios campos, como por ejemplo: clasificación de documentos, minado de datos, aprendizaje automático, análisis de imagen, análisis químicos o separación de señales sonoras. Esta factorización intenta aproximar una matriz no negativa con el producto de dos matrices no negativas de menor tamaño. Además, esta descomposición matricial suele crear representaciones por partes de los datos en la matriz original.

Los algoritmos diseñados para calcular la solución de estos dos problemas no negativos tienen un elevado coste computacional, y debido a ese elevado coste, estas descomposiciones pueden beneficiarse mucho del uso de técnicas de Computación de Altas Prestaciones (HPC por sus siglas en inglés). Hoy en día existen sistemas computacionales muy potentes capaces de ofrecer mucha potencia de cómputo que son utilizados para resolver problemas extremadamente complejos necesarios en diversos campos de la ciencia y la ingeniería. La potencia de estos sistemas continúa aumentando continuamente desde los modernos computadores multinúcleo a lo último en aceleradores de cálculo (Unidades de Procesamiento Gráfico (GPU), Intel Many Integrated Core (MIC), etc.). Para obtener el máximo rendimiento de estos sistemas de computación de altas prestaciones, los desarrolladores deben utilizar tecnologías software tales como la programación paralela, la vectorización o el uso de librerías de computación

altas prestaciones.

A pesar de que existen diversos algoritmos para calcular la NMF y resolver el problema NNLS, no todos ellos disponen de una implementación paralela y eficiente. Además, es muy interesante reunir diversos algoritmos con propiedades diferentes en una sola librería computacional. Esta tesis presenta una librería computacional de altas prestaciones que contiene implementaciones paralelas y eficientes de los mejores algoritmos existentes actualmente para calcular la NMF. Además la tesis también incluye una comparación experimental entre las diferentes implementaciones presentadas. Esta librería centrada en el cálculo de la NMF soporta múltiples arquitecturas tales como CPUs multinúcleo, GPUs e Intel MIC. El objetivo de esta librería es ofrecer un abanico de algoritmos eficientes para ayudar a científicos, ingenieros o cualquier tipo de profesionales que necesitan hacer uso de la NMF. Debido a que la NMF es una herramienta transversal que puede ser utilizada en múltiples disciplinas, no todos los profesionales que puedan beneficiarse del uso de la NMF tienen los conocimientos para aprovechar al capacidad de cómputo de los sistemas de altas prestaciones actuales. Esta librería trata de resolver dicho problema ofreciendo un librería fácil de utilizar que incluye implementaciones de altas prestaciones de los mejores algoritmos diseñados para resolver la NMF.

Otro problema abordado en esta tesis es la actualización de las factorizaciones no negativas. El problema de la actualización se ha estudiado tanto para la solución del problema NNLS como para el cálculo de la NMF. Existen problemas no negativos cuya solución es próxima a otros problemas que ya han sido resueltos, el problema de la actualización consiste en aprovechar la solución de un problema A que ya ha sido resuelto, para obtener la solución de un problema B cercano al problema A. Utilizando esta aproximación, el problema B puede ser resuelto más rápido que si se tuviera que resolver sin aprovechar la solución conocida del problema A. En esta tesis se presenta una metodología algorítmica para resolver ambos problemas de actualización: la actualización de la solución del problema NNLS y la actualización de la NMF. Además se presentan evaluaciones empíricas de las soluciones presentadas para ambos problemas. Los resultados de estas evaluaciones muestran que los algoritmos propuestos son más rápidos que resolver el problema desde el inicio en todos los casos examinados.