Reporte:Singular Value Decomposition on GPU using CUDA

Oliab Herrera Coria 107863

Contents

Métodos de Primer órden. **1**
Regularizaciones suaves. 1
Composite objectives 2

Escalamiento de Grandes Datos vía Aleatorización **2**
Gradiente estocástico. 2

Computo Distribuido y en paralelo **2**

Conclusiones **2**

En los últimos años, dado el gran escalamiento en cantidad de datos y la necesidad de analizarlos, se han popularizado algorítmos, como aquellos utilizados en aprendizaje estadístico, que utilizan la optimización convexa para analizar y encontrar soluciones a problemas de terabites de información.

Actualmente, los algoritmos de opimización en grandes datos tienen tres vertientes:

- Métodos de primer orden.
- Aleatorización.
- Cómputo paralelo y distribuido.

Métodos de Primer órden.

Estos métodos utilizan indormación de primer orden de la función objetivo tal como los gradientes o aproximaciones para transformar el problema inicial en uno similar pero que tenga mejores "propiedades" para el cálulo de sus soluciones. Estos métodos son robustos y tienen tasas de convergencia que no dependen de la dimensión de los datos.

En el articulo el autor utiliza el *proximal gradient framework*, ya que usan descenso gradiente como base del problema, para demostrar que las regularizaciones no-suaves (Lasso) pueden ser resuetas de manera tan eficiente como las regularizaciones suaves.

Regularizaciones suaves.

El autor empueza hablando del método de descenso gradiente que es dado por la ecuacion:

$$x^{k+1} = x^k - \alpha_k \nabla f(xk)$$

Aunque este método no es el más rápido, el autor plantea un algoritmo (El metodo de Nesterov) que nos da el mínimo de pasos para llegar a una solución con un grado ϵ de error.

Otra opción es convertir un problema *convexo* en un problema *estrictamenteconvexo* agregando un termino de regularización cuadrático $\frac{\lambda}{2}||x||^2$ al problema

Composite objectives

Esto consiste en utilizar aproximaciones cuadráticas a la función objetivo para convertir el problemea en uno estrictamente convexo

Escalamiento de Grandes Datos vía Aleatorización

Su mayor diferencia con los otros métodos de aproximación es que se tiene un contro del comportamiento esperado. Las implementaciones de este tipo incluyen el cambiar calculos exactos que son caros computacionalmente por aproximaciones, acelerar rutinas de algebra vía aleatorización.

Gradiente estocástico.

El método de gradiente estocástico utiliza todos los datos disponibles pero usa gradientes aproximados:

El algoritmo es el siguiente: - Escoje $j_k \in \{1, ..., n\}$ de manera que sea aleatorio uniforme. - Calcula el descenso gradiente $x^{k+1} = x^k - \alpha_k \nabla f_{jk}(x^k)$

Computo Distribuido y en paralelo

Gracias a a los avances tecnologícos el poder de computo y procesamiento de las máquinas han dado un boost a la capacidad de realizar cálculos que son utilizados en problemas de optimización convexa. Para poder manejar la cantidad de datos que tenemos actualmente hay que utilizar técnicas de computo distribuido y en paralelo.

Existen dos elementos que pueden dificultar el uso de estas tecnologías.
- Comunicación entre máquinas o procesos internos pueden disminuir la eficiencia de los algoritmos.
- Sincronización, la sincronización entre procesos puede alentar los algoritmos, como lo vimos en clase con los threads.

Esto para adaptar métodos que ya existían, algunos enunciados arriba, de manera que se aprovechen las ventajas recnologicas. pero que a la hora de ser implementados tomen en cuenta la comunicación y sincronización para poder ejecutarse de manera eficiente.

Conclusiones

El autor nos describe diferentes maneras de solucionar los problemas que vienen con el análisis de grandes datos. Desde opciones más tradicionales como el descenso gradiente hasta el uso de computo distribuido, tenemos en nuestras manos herramientas muy poderosas que nos permiten enfrentar estos problemas, pero sin perder de vista que son problemas matemáticos que usan la computación como herramienta