Reporte: "Convex Optimization for Big Data"

Miguel Castañeda 175840

31 de mayo de 2018

El articulo hace una revisión de los últimos avances en los algoritmos de optimización convexa para "Big Data" que buscan reducir los cuellos de botella en almacenamiento, proesamiento y comuniciación.

El artículo describe los fundamentos de la optimización de big data mediante la formulación:

$$F^* = min_x\{F(x) := f(x) + g(x) : x \in R^n\}$$

donde f y g son funciones convexas.

Los problemas de optimización de big data se baasan en tres pilares:

- Metodos de primer orden. Los métodos de primer orden obtienen soluciones de baja a media precisión numérica , estos métodos tienen proporciones de convergencia de dimensiones independientes lo que los hace ideales para cómputo paralelo y distribuido.
- Aleatorización. Las técnicas de aleatorización se basan en técnicas de aproximación para mejorar la escalabilidad de los métodos de primer orden, las ideas clave incluyen la actualizaciones aleatorias parciales de las variables de optimización reemplazando el gradiente determinista con estimaciones estadísticas aproxiados menos costosas.
- Cómputo Paralelo y Distribuido. Los métodos de primer orden proveen de un marco que permite la distribución de tareas y realizar los cálculos de manera paralela.

Estos tres conceptos se complementan para ofrecer escalabilidad a los problemas de optimización de big data.

Métodos de primer orden

La fuente principal para los problemas de big data tienen una forma lineal en muchos casos

$$y = \Phi X_0 + z$$

donde X_0 es un parámetro desconocido y Φ es una matriz conocida y z representa las perturbaciones o ruido este problema se puede resolver utilizando la formulación del estimado de mínimos cuadrados el cual ya ha sido resuelto de manera eficiente por Krylov, una variante importante es la regularización mediante LASSO el cual incorpora un parámetro λ que controla el valor de la regularización.

Objetivos suaves, un caso importante es cuando la función objetivo F consiste únicamente de una función diferenciable convexa f para este caso la técinca que se debe usar es la del método del gradiente que usa solo el gradiente local de manera iterativa realizando actualizaciones.

Escalamiento vía aleatoriación

Los métodos de primer orden son muy robustos a usar aproximaciones tales como el gradiente como por ejemplo el algoritmo PageRank de Google que mide la importancia de los nodos dado un grafo via una matriz de incidencias estos problemas pueden llegar a creceer considerablemente en dimensiones por lo que hace necesario generar técincas que permitan mediante aproximaciones resolver el problema, como por ejemplo para este caso el descenso en gradiente coordinado

Otra técnica son los métodos de gradiente estocástico los cuales actualizan una sola coordenada a la vez con el valor exacto del gradiente y de manera graduar actualizar todas las coordenadas lo cual hace crucial la seleccion adecuada de los puntos por los cuales debe seleccionar en cada iteración.

Para los problemas de Big data, las operaciones básicas de algebra lineal tales como la descomposición de matrices (eigenvalores, valor singular, Cholesky) y multimplicación de matrices pueden resultar en cuellos de botella debido a la dependencia a las dimesiones.

Pero cuando las matrices tienen bajo rago la eficiencia de estos metodos se mejoran, por ejemplo para el caso de la descoposición SVD de una matriz M esta tienen un costo de $O(pr^2 + r^3)$ flops.

Los problemas de primer orden parecen ser los que mejor se pueden mejorar mediante el uso de sistemas heterogénos de cómputo con mejoras en los sistemas de comunicación, sincronización.