

Reporte de lectura: Cevher, et al (2014) Convex Optimization for Big Data

Fernando Felipe Briseño Martínez

31 de mayo de 2018

Resumen

El artículo revisa los avances más recientes en optimización convexa que buscan reducir los cuellos de botella de almacenamiento, cómputo y comunicación para los datos de gran escala. El artículo da un panorama de las técnicas como los métodos de primer orden, aleatorización para escalabilidad, y el cómputo en paralelo y distribuido.

Estructura del artículo

El artículo consta de cuatro secciones y una introducción:

1. Métodos de primer orden para la optimización convexa suavizada y no suavizada
2. Gran escala a través de aleatorización
3. El papel de la computación paralela y distribuida
4. Perspectivas para la optimización convexa

En general, se explican en la introducción los elementos básicos y definiciones de métodos de primer orden, aleatorización y computación paralela y distribuida.

Argumentación

Los autores plantean que hay un renovado interés por las formas convexas y la optimización debido a los avances teóricos relacionados con la realidad estructurada y disminución de rango, así como el uso cada vez más común de modelos de aprendizaje estadístico como las máquinas de soporte vectorial.

Este interés pone gran presión para poder tener algoritmos que permitan el trabajo sobre grandes conjuntos de datos, por lo que el límite de algoritmos clásicos a menudo se queda en la discusión teórica sobre la tratabilidad de los problemas de optimización.

Los autores proponen que el problema de optimización en datos a gran escala usa la formulación compuesta de dos funciones. Esta, por ejemplo, es la que se usa en métodos bayesianos en la que se tiene una función de verosimilitud y una a priori.

Para la optimización a gran escala de este tipo de funciones, los autores describen tres grandes pilares interconectados entre sí:

Métodos de primer orden:

Son soluciones numéricas de baja o media precisión usando información de primer orden, tal como estimaciones de gradientes.

Los autores hacen un ejemplo de cómo el uso de descenso gradiente estocástico puede mejorar la eficiencia de la estimación LASSO basada en el algoritmo clásico de álgebra lineal basado en productos interiores.

Aleatorización

Las técnicas de aleatorización permiten mejorar la escalabilidad de métodos de primer orden, ya que se puede controlar su comportamiento esperado. Algunas ideas para aplicaciones son el uso de actualizaciones parciales aleatorias de parámetros a optimizar, el uso de estimadores estadísticos para sustituir cálculos complejos, y usar aleatorización para acelerar el álgebra lineal básica.

Computación paralela y distribuida

Con este tema, los autores describen tanto el potencial de escalar el cómputo a través del cálculo en paralelo, como el uso de algoritmos asíncronos y descentralizados a través del cómputo distribuido.

Perspectivas sobre el artículo

El artículo concluye que los problemas de gran escala necesitan un cambio fundamental de cómo se diseñan los algoritmos de optimización convexa y sugieren elecciones computacionales no convencionales, pues los recursos computacionales han crecido de forma modesta comparados con los requerimientos de los problemas actuales.

La principal conclusión de los autores es que las nuevas herramientas de aproximación adaptarán los algoritmos de optimización convexa a la heterogeneidad de las plataformas computacionales. Asimismo, los autores proponen que el uso de problemas compuestos, junto con sus principios de mapeo permitirá enfrentarse al ruido y otros problemas y serán parte de los desarrollos futuros en la optimización convexa a gran escala.

Los problemas planteados por los autores son de frontera, y por lo tanto, en relativamente poco tiempo surgirán soluciones a estos problemas, mientras que al mismo tiempo aparecerán otros nuevos. Las herramientas aquí planteadas nos permitan reenfocar la energía a problemas viejos usando herramientas nuevas, lo que sienta las bases para grandes expectativas de lo que sucederá en este campo.