

Algorithmen für dynamische geometrische Datenströme

Dissertation von Gereon Frahling

Oktober 2006

Die zunehmende Vernetzung moderner Computersysteme erzeugt oftmals sehr große Mengen von Daten, die nicht gespeichert werden können. In dieser Arbeit entwickeln wir daher Datenstromalgorithmen, die sich Daten sequenziell ansehen. Gespeichert wird dabei nur eine kleine Skizze der Daten, die später benutzt werden kann, um Anfragen über den Inhalt der Daten approximativ zu beantworten. Speziell beschäftigen wir uns mit dem sogenannten Modell der dynamischen geometrischen Datenströme, in dem ein Datenstrom gegeben ist, der aus m Einfüge- und Löschooperationen von Punkten eines d -dimensionalen Raumes in eine Menge M besteht. Wir entwickeln Datenstromalgorithmen, die höchstens $\text{poly}(\log m)$ Speicherbits benutzen:

Nachdem wir gezeigt haben, wie man effizient Stichproben der Menge M in dynamischen Datenströmen aufrechterhalten kann, entwickeln wir eine Kernmengen-Technik, die benutzt werden kann, um $(1 + \epsilon)$ -approximative k -Median-, k -Means- und MaxCut-Clusterings der Punktmenge unter den Einfüge- und Löschooperationen effizient aufrechtzuerhalten. Für die Kernmengen-Technik werden daraufhin viele weitere Anwendungen abgeleitet und gezeigt, wie sie verwendet werden kann, um iterative Clustering-Algorithmen in der Praxis zu beschleunigen.

Im letzten Kapitel der Arbeit stellen wir schließlich Datenstromalgorithmen vor, um Minoren wie Dreiecke oder bipartite Cliques in großen, als Datenstrom gegebenen Graphen wie dem Webgraph approximativ zu zählen.