Zusammenfassung

Das Thema dieser Dissertation ist die exakte Nächster-Nachbar-Suche. Das Problem besteht darin, effiziente Datenstrukturen und Algorithmen zu entwickeln, die zu einer festen Menge von n Punkten $P \subset \mathbb{R}^d$ und einem beliebigen Anfragepunkt $q \in \mathbb{R}^d$, den oder die nächsten Nachbarn aus P zu q finden. Der Abstand wird meist in einer der Minkowski-Metriken $L_1, L_2, \ldots, L_\infty$ definiert. Die Dimension d des Suchraumes ist in vielen Anwendungen dieses Problems sehr groß; sie liegt in der Grössenordnung von Hunderten bis Tausenden. Die bekannten Algorithmen für das exakte Nächster-Nachbar-Problem benötigen eine in d exponentielle Laufzeit oder einen in d exponentiellen Speicherplatz. Sie sind daher ineffizient für hochdimensionale Anwendungen, und können mit der brute-force Methode nicht konkurrieren, welche alle n Distanzen zu dem Anfragepunkt berechnet und den Punkt mit minimaler Distanz auswählt.

In dieser Dissertation werden effiziente Algorithmen für die exakte Nächster-Nachbar-Suche in einem hochdimensionalen $(\mathbb{R}^d, L_\infty)$ Raum entwickelt. Wir untersuchen die erwartete Laufzeit unserer Algorithmen unter der Voraussetzung, dass die Punkte aus P gleichverteilt aus dem d-dimensionalen Einheitswürfel sind.

Die Algorithmen sind einfach zu implementieren und verbessern wesentlich die brute-force Methode bezüglich der erwarteten Laufzeit.

Kapitel 2 stellt Methoden für das exakte Nächster-Nachbar Problem vor, die keine Vorverarbeitung und nur Speicherplatz für die n Punkte benötigen. Deren erwartete Laufzeit ist $O(\frac{nd}{\ln n} + n)$.

Wir entwickeln in Kapitel 3 Algorithmen, die auf der Basis einer einfachen Vorverarbeitung eine effizientere Suche ermöglichen. Der in dem Abschnitt 3.1.2 entwickelte Suchalgorithmus hat eine erwartete Laufzeit von $O\left(n \ln\left(\frac{d}{\ln n} + 1\right) + n\right)$, einen linearen Speicherbedarf und $O(nd \ln n)$ Vorverarbeitungszeit. Im Abschnitt 3.2 geben wir einen Suchalgorithmus an, der eine in der Vorverarbeitung berechnete Zerlegung der Punktmenge P benutzt. Die Zerlegung besteht aus Folgen von Punkten, die monoton in $\mathbb{R}^d$ sind. Die erwartete Laufzeit beträgt $O(\sqrt{d} n^{1-\frac{1}{\sqrt{d}}} \ln n)$ für Dimensionen $d < \left(\frac{\ln n}{\ln \ln n}\right)^2$.

Die vorgestellten Methoden und deren Laufzeitanalyse werden in Kapitel 4 für das k-Nächste-Nachbarn-Problem verallgemeinert. Außerdem werden die erwarteten Laufzeiten der entwickelten Methoden bei Zugrundelegung von anderen Wahrscheinlichkeitsverteilungen analysiert. Weiterhin betrachten wir das Externer-Speicher-Berechnungsmodell und entwickeln einen externen Algorithmus für die L_∞-Nächster-Nachbar-Suche.

In Kapitel 5 wird eine Methode entwickelt, die einen Tradeoff zwischen dem erwarteten Speicherbedarf der Datenstruktur und der erwarteten Laufzeit des dazugehörigen Suchalgorithmus erlaubt.