Zusammenfassung

Thema dieser Dissertationsschrift ist das k-Median Problem unter Verwendung eines Abstandsmaßes D_φ aus der Familie der Bregman-Divergenzen: Zu einer gegebenen Eingabemenge P der Größe n aus dem $\mathbb{R}^d$ ist eine Zentrenmenge C der Größe k gesucht, welche die Zielfunktion $\mathrm{cost}(P,C) = \sum_{p\in P} \min_{c\in C}\big\{\mathrm{D}_\varphi(p,c)\big\}$ minimiert. Dieses Problem ergibt sich in Anwendungen aus den verschiedensten Teilbereichen der Informatik, etwa in der Informationstheorie, in der Statistik, beim Durchsuchen großer Datenbestände oder bei der Verarbeitung von Sprachsignalen.

Das Hauptresultat dieser Arbeit besteht in der Entwicklung einer Sammlung von Algorithmen und Techniken, die sich auf (nahezu) alle Bregman-Divergenzen anwenden lassen. Insbesondere präsentieren wir einen randomisierten Approximationsalgorithmus der Güte $(1+\varepsilon)$ für das Bregman-k-Median-Problem. Dieser Algorithmus berechnet seine Lösung unter Verwendung von maximal $2^{\tilde{\mathcal{O}}(k/\varepsilon)}n$ arithmetischen Operationen, darunter auch Auswertungen des Abstandsmaßes D_φ. Dabei handelt es sich um den ersten für das Bregman-k-Median-Problem anwendbaren Algorithmus, der eine beweisbare Approximationsgüte aufweist. Außerdem präsentieren wir einen effizienten, praktisch relevanten, randomisierten Approximationsalgorithmus, der Lösungen der Güte $\mathcal{O}(\log k)$ berechnet; für spezielle, wohlseparierte Eingabeinstanzen berechnet dieser Algorithmus sogar Lösungen konstanter Güte.

Darüber hinaus untersuchen wir die Anwendung von Kernmengen für das Bregman-k-Median-Problem. Kurz zusammengefasst handelt es sich bei einer Kernmenge um eine kleine (gewichtete) Punktemenge, welche die gleichen Clustering-Eigenschaften wie die ursprüngliche Eingabemenge aufweist. Wir demonstrieren, wie sich klassische Kernmengenkonstruktionen des euklidischen k-Mittelwert-Problems auf eine spezielle Teilmenge der Bregman-Divergenzen verallgemeinern lassen, nämlich auf die Klasse der so genannten Mahalanobis-Distanzen. Wir präsentieren ferner eine neue, praktisch vorteilhafte, randomisierte Kernmengenkonstruktion für das Mahalanobis-k-Median-Problem in niedrigdimensionalen Räumen. Zudem greifen wir das Konzept der schwachen Kernmengen auf und prä-

sentieren damit die erste Kernmengenkonstruktion, die sich für (fast) alle Bregman-Divergenzen anwenden läßt. Unter Anwendung dieser schwachen Kernmengen erhalten wir den derzeit asymptotisch effizientesten $(1+\varepsilon)$-Approximationsalgorithmus für das Bregman-k-Median-Problem. Dieser Algorithmus benötigt maximal $\mathcal{O}(kn) + 2^{\tilde{\mathcal{O}}(k/\varepsilon)} \log^{k+2}(n)$ arithmetische Operationen, darunter auch Auswertungen des Abstandsmaßes D_φ.