

Raumzeitliche Analyse von Twitter Daten

Softwareentwicklungsprojekt I

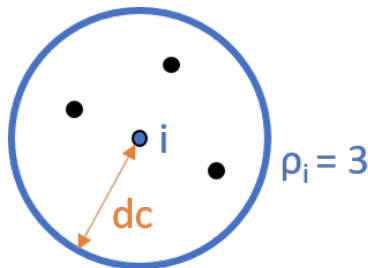
Niklas Baumbach, Felix Juch und Martin Immel

Ziele:

- raumzeitliche Auswertung von Twitter-Daten zur Detektion und Analyse von Naturkatastrophen
- Clustern der geo-lokalisierten Tweets mit Krisenbezug
- Implementierung eines geeigneten Clustering-Algorithmus → FDCA (*Fast Density Clustering Algorithm*)
- Erstellen eines Frameworks zum Test des Algorithmus und zur Simulation

FDCA-Clustering:

Dichte ρ und Grenzdistanz dc :



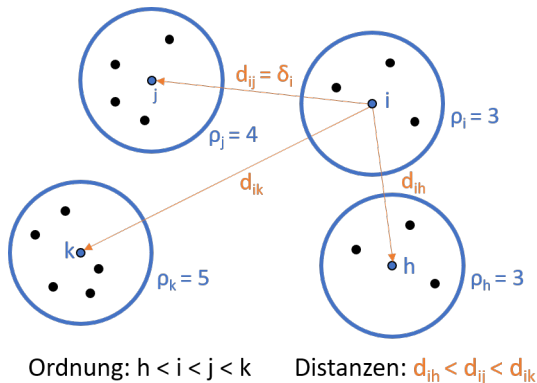
- dc = Grenzdistanz
- ρ = Anzahl Datenpunkte in Grenzdistanz

Inhalt des Projekts

Clustering

FDCA-Clustering:

Delta δ :



- δ = minimaler Abstand zu Punkt höherer Dichte (bei gleicher Dichte: höhere Ordnung)
- Ordnung durch Index

Inhalt des Projekts

Clustering

FDCA-Clustering:

Clusterzentren bestimmen:

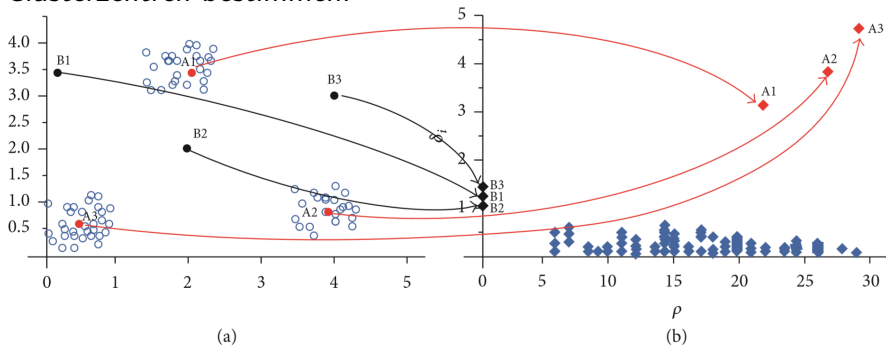


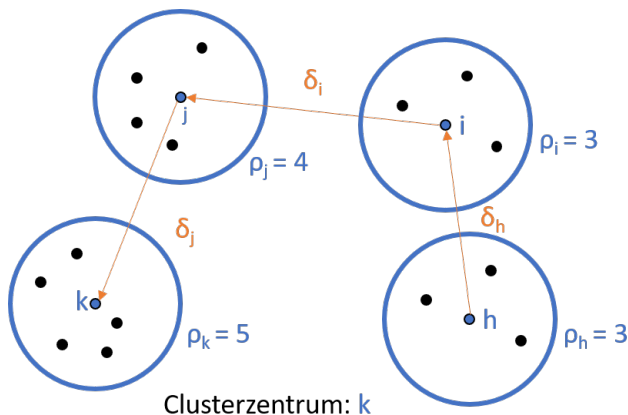
FIGURE 1: The algorithm in two dimensions. (a) Point distribution. (b) ρ and δ distribution of (a).

Inhalt des Projekts

Clustering

FDCA-Clustering:

Zuweisung der Cluster:





Betreuer: Jens Kersten und Friedericke Klan

Wesentliche Leistungen

- Implementierung
- Fehler aus dem Paper aufgelöst
- Aufarbeitung & Vorbereitung der Daten
- Entwicklung geeigneter Darstellungsformen
- Einbindung in Commandline-Programm

Nicht erreichte Ziele

- Exakte Implementierung des FDCA
- Automatische dc-Suche
- Einbindung in das System des DLR

Bilder

Noch Fragen?

