

Raumzeitliche Analyse von Twitter Daten

Softwareentwicklungsprojekt I

Niklas Baumbach, Felix Juch und Martin Immel

Inhalt des Projekts

Raumzeitliche
Analyse von
Twitter Daten

Niklas

Baumbach, Felix
Juch und Martin
Immel

Inhalt des
Projekts

Stakeholder

Team-
Organisation

Ziele und
geplante
Funktionalitäten

Verwendete
Technologien

Projektplanung

Aufgetretene
Probleme

Inhalte:

- 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*)
- Implementierung eines Frameworks zum Test des Algorithmus und zur Simulation

Inhalt des Projekts

Clustering

Raumzeitliche
Analyse von
Twitter Daten

Niklas
Baumbach, Felix
Juch und Martin
Immel

Inhalt des
Projekts

Stakeholder

Team-
Organisation

Ziele und
geplante
Funktionalitäten

Verwendete
Technologien

Projektplanung

Aufgetretene
Probleme

Clustering:

- Bestimmung der Elementgruppen eines Datensatzes - Cluster
- Einteilung der Elemente in Gruppen mit ähnlichen Eigenschaften - Clustering
- Zugehörigkeit wird über die Bestimmung eines mehrdimensionalen Distanzmaßes festgelegt

Inhalt des Projekts

FDCA

Raumzeitliche
Analyse von
Twitter Daten

Niklas

Baumbach, Felix
Juch und Martin
Immel

Inhalt des
Projekts

Stakeholder

Team-
Organisation

Ziele und
geplante
Funktionalitäten

Verwendete
Technologien

Projektplanung

Aufgetretene
Probleme

FDCA:

- funktioniert für numerische und kategoriale Daten
- errechnet die Dichte der Datenpunkte
- bestimmen der kleinsten Distanz zu einem Datenpunkt höherer Dichte
- legt Clusterzentren fest (hohe Dichte, große Distanz zu Datenpunkt höherer Dichte)
- sortiert Noise-Punkte aus (geringe Dichte, große Distanz zu Datenpunkt höherer Dichte)

Stakeholder

Raumzeitliche
Analyse von
Twitter Daten

Niklas
Baumbach, Felix
Juch und Martin
Immel

Inhalt des
Projekts

Stakeholder

Team-
Organisation

Ziele und
geplante
Funktionalitäten

Verwendete
Technologien

Projektplanung

Aufgetretene
Probleme



Betreuer: Jens Kersten und Friedericke Klan

Team-Organisation

Raumzeitliche
Analyse von
Twitter Daten

Niklas
Baumbach, Felix
Juch und Martin
Immel

Inhalt des
Projekts

Stakeholder

Team-
Organisation

Ziele und
geplante
Funktionalitäten

Verwendete
Technologien

Projektplanung

Aufgetretene
Probleme

Niklas:

- Entwicklung des Frameworks zur Simulation und Visualisierung
- Erarbeitung von Unittests

Felix und Martin:

- Auswertung des Papers zum FDCA
- Implementierung des FDCA

Investierte Zeit: 10 bis 15 Stunden

Ziele und geplante Funktionalitäten

Raumzeitliche
Analyse von
Twitter Daten

Niklas
Baumbach, Felix
Juch und Martin
Immel

Inhalt des
Projekts

Stakeholder

Team-
Organisation

**Ziele und
geplante
Funktionalitäten**

Verwendete
Technologien

Projektplanung

Aufgetretene
Probleme

- Fertiger Algorithmus zum Clustern von Twitter-Daten (raumzeitlich und kategorisch)
- Exemplarische Anwendung auf Daten von Hurrikan Florence
- Integration in das System des DLR (z.B. „Event Detection Viewer“ oder „Kibana“)

Verwendete Technologien

Raumzeitliche
Analyse von
Twitter Daten

Niklas
Baumbach, Felix
Juch und Martin
Immel

Inhalt des
Projekts

Stakeholder

Team-
Organisation

Ziele und
geplante
Funktionalitäten

Verwendete
Technologien

Projektplanung

Aufgetretene
Probleme



Projektplanung

Raumzeitliche
Analyse von
Twitter Daten

Niklas
Baumbach, Felix
Juch und Martin
Immel

Inhalt des
Projekts

Stakeholder

Team-
Organisation

Ziele und
geplante
Funktionalitäten

Verwendete
Technologien

Projektplanung

Aufgetretene
Probleme

Phase 1 bis 08. Mai:

- Paper über den FDCA auswerten und verstehen
- Geeignete Datasets für Tests finden

Phase 2 bis 06. Juni:

- Implementierung eines Frameworks zur Validierung des Algorithmus und Simulation
- Implementierung des FDCA

Projektplanung

Raumzeitliche
Analyse von
Twitter Daten

Niklas
Baumbach, Felix
Juch und Martin
Immel

Inhalt des
Projekts

Stakeholder

Team-
Organisation

Ziele und
geplante
Funktionalitäten

Verwendete
Technologien

Projektplanung

Aufgetretene
Probleme

Phase 3 bis 20. Juni:

- Evaluierung der Ergebnisse unseres Algorithmus im Vergleich zum Paper
- Anwendung des Algorithmus auf die Twitter-Daten zum Hurrikan Florence
- Visualisierung der Ergebnisse des Clustering

Phase 4 bis 03. Juli:

- Integration des Algorithmus in das System des DLR
- evtl. Implementierung weiterer Clustering-Algorithmen

Aufgetretene Probleme

Raumzeitliche
Analyse von
Twitter Daten

Niklas
Baumbach, Felix
Juch und Martin
Immel

Inhalt des
Projekts

Stakeholder

Team-
Organisation

Ziele und
geplante
Funktionalitäten

Verwendete
Technologien

Projektplanung

Aufgetretene
Probleme

Folgende Probleme sind während des Projekts aufgetreten:

- Sprachliche Unklarheiten im Paper
- Verwendung Funktionen im Paper ohne Definition
- keine Zwischenergebnisse des Algorithmus im Paper
- keine Rückmeldung auf Anfragen an die Autoren
- „lange“ Einarbeitungszeit in Pandas

Noch Fragen?

Raumzeitliche Analyse von Twitter Daten

Niklas
Baumbach, Felix
Juch und Martin
Immel

Inhalt des
Projekts

Stakeholder

Team-
Organisation

Ziele und
geplante
Funktionalitäten

Verwendete
Technologien

Projektplanung

Aufgetretene
Probleme

