Analyzing the Impact of Natural Disasters:
A Concept for Spatio-temporal Analyses of Social Media Data

Jens Kersten1(jens.kersten@dlr.de), Anna Kruspe1, Friederike Klan1
1German Aerospace Center (DLR), Jena, Germany

Introduction

Problem:
• Analysis of Twitter messages for crisis response activities
• Analyzing the dynamics of reactions caused by disasters taking into account space and time to reveal insights regarding global event impacts on population and environment

Contributions:
• Identification of suitable approaches for spatio-temporal analyses of social media data
• General workflow (Fig. 1) for spatio-temporal social media data
• Experiments with Twitter data recorded during hurricane Florence (September 2018)
• Formulation of future work directions

Proposed Workflow

Introduction
• 600,000 geo-located Tweets
• Hurricane Florence, Carolinas, USA (Fig. 2)
• Record period September 12-19, 2018
• 30,700 crisis-related tweets with geo-location after CNN-based filtering [1]

Dataset

Spatio-temporal Analyses

A. Local hot-spot detection (Getis-Ord G* [3])
• Hot and cold-spot analysis (Fig. 3)
• Clustering of high and low tweet occurrences (crisis-related) on county level
• Occurrences normalized by population density [6]

B. Space-time kernel density estimation, ST-KDE [4]
• Non-parametric approach to compute a density function
• Describes the intensity of geographic events’ distribution
• High activity areas (reflecting population densities): Charlotte, Raleigh and other cities (Fig. 4)
• Higher activities after landfall (Sep. 14)

C. STDBSCAN [5]
• Spatio-temporal extension of DBSCAN (Density-Based Spatial Clustering of Applications with Noise)
• Density associated to a point is obtained by counting the number of surrounding points
• Discovers clusters with arbitrary shapes
• Estimates the number of clusters

Results and Cluster Analysis

G* statistics
• High population density around Charlotte ➔ low relative number of crisis-related tweets ST-KDE
• Spatio-temporal density visualization
• Daily activity patterns and increased activity after the landfall

STDBSCAN
• Densely populated areas tend to produce clusters covering more than one day (Fig. 5)
• Analysis of cluster contents: classification of tweets into 7 information classes [2] (Fig. 6)
• Large clusters ➔ Similar distributions covering all classes
• Similar keyword maps (Fig. 7)
• Peak of tweets related to affected individuals on Sept. 15
• More tweets related to donations and volunteering after the landfall

Conclusions and Outlook

Workflow for spatio-temporal analysis and visualization of crisis-related tweets
• State-of-the-art CNNs used for filtering as well as tweet classification
• Case study: Hurricane Florence, Sep. 2018
• First qualitative results with G* statistics, ST-KDE, and STDBSCAN
• Visualization of hot-spots and spatio-temporal patterns

Open problems and possible solutions
• Data sparsity using geo-located Tweets ➔ Additional acquisition and use of tweets from keyword-based search
• Choice of method parameters ➔ Systematic analyses required
• Large STDBSCAN clusters with similar contents ➔ Incorporation of class labels for clustering + parameter tuning
• Validation of results ➔ Benchmark data for testing + checking against (sub-) events reported in the news

References