



A Social Media Big Data-Based Disaster Assessment Framework for Typhoon-induced Flood: Case Study of Typhoon Lekima

Jiting Tang^{1,2,3}, Saini Yang^{1,2,3}, and Weiping Wang^{4,5,6}

¹Key Laboratory of Environmental Change and Natural Disaster, Ministry of Education, Beijing Normal University, Beijing, China

²State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, Beijing, China

³Academy of Disaster Reduction and Emergency Management, Faculty of Geographical Science, Beijing Normal University, Beijing, China

⁴Institute of Transportation Systems Science and Engineering, Beijing Jiaotong University, Beijing, China

⁵State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China

⁶Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport, Ministry of Transport, Beijing Jiaotong University, Beijing, China

In 2019, the typhoon Lekima hit China, bringing strong winds and heavy rainfall to the nine provinces and municipalities on the northeastern coast of China. According to the Ministry of Emergency Management of the People's Republic of China, Lekima caused 66 direct fatalities, 14 million affected people and is responsible for a direct economic loss in excess of 50 billion yuan. The current observation technologies include remote sensing and meteorological observation. But they have a long time cycle of data collection and a low interaction with disaster victims. Social media big data is a new data source for natural disaster research, which can provide technical reference for natural hazard analysis, risk assessment and emergency rescue information management.

We propose an assessment framework of social media data-based typhoon-induced flood assessment, which includes five parts: (1) **Data acquisition**. Obtain Sina Weibo text and some tag attributes based on keywords, time and location. (2) **Spatiotemporal quantitative analysis**. Collect the public concerns and trends from the perspective of words, time and space of different scales to judge the impact range of typhoon-induced flood. (3) **Text classification and multi-source heterogeneous data fusion analysis**. Build a hazard intensity and disaster text classification model by CNN (Convolutional Neural Networks), then integrate multi-source data including meteorological monitoring, population economy and disaster report for secondary evaluation and correction. (4) **Text clustering and sub event mining**. Extract subevents by BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies) text clustering algorithms for automatic recognition of emergencies. (5) **Emotional analysis and crisis management**. Use time-space sequence model and four-quadrant analysis method to track the public negative emotions and find the potential crisis for emergency management.

This framework is validated with the case study of typhoon Lekima. The results show that social

media big data makes up for the gap of data efficiency and spatial coverage. Our framework can assess the influence coverage, hazard intensity, disaster information and emergency needs, and it can reverse the disaster propagation process based on the spatiotemporal sequence. The assessment results after the secondary correction of multi-source data can be used in the actual system.

The proposed framework can be applied on a wide spatial scope and even full coverage; it is spatially efficient and can obtain feedback from affected areas and people almost immediately at the same time as a disaster occurs. Hence, it has a promising potential in large-scale and real-time disaster assessment.