Human Mobility Under Disaster Relief with Social Media Mining

Lu Cheng, Jiayong Mo, Linzhen Luo

## Introduction

Natural disasters affect our society in profound ways. According to the 2010 World Disasters Report [1], between 2000 and 2009, disasters resulted in 1 million casualties, affected an additional 2.5 million people, and caused a loss of ~$1Trillion. Consequently, disaster response is considered a grand challenge by the President's Council of Advisors on Science and Technology Problem Description. Today, experts largely agree that data technologies can help in all four phases of disaster management: prevention, preparedness, response, and recovery1. In fact, one of the biggest problems that emergency and disaster managers face today is not a lack of data, but an overload of information. Experts agree that transformative impact in disaster management will arise only through timely information-driven decision making and in the presence of shifting demands of disasters and resources [4-8]. Therefore, data-driven models for disaster preparedness and response are increasingly critical in predicting geo-temporal evolution of disasters and studying human mobility under urgent scenarios. However, with all the accessible data, only a tiny part of it are geo-tagged and the data for each user is extremely sparse. For social media data, because they are usually very noisy, short and informal, advanced data preprocessing techniques are required to tackle the problem

## Goals

1. Develop a robust group-level human mobility model with a multidimensional structure, jointly consider information of location, time and text of each tweet posted by users. Our insight is that the geo-tagged social media data usually contains multiple user groups, where the users within the same group share significant movement regularity. Meanwhile, user grouping and mobility modeling are two intertwined tasks: (1) better user grouping offers better within-group data consistency and thus leads to more reliable mobility models; and (2) better mobility models serve as useful guidance that helps infer the group a user belongs to.
2. Estimate locations for tweets that are not tagged with location information. The motivation is simply because there are a bunch of data that do not have location information while they can provide our human mobility model with significant insights. Also, considering the data sparsity problem, location estimation can significantly augment the data pool.
3. Explore the data we crawled from TweetTracker [9], a system that is specifically created to crawl tweets based on the keywords and locations specified. This step is important because it’s necessary to have a general view of the huge amount of dataset we have.

## Research Questions:

1. What are the intrinsic states underlying people’s movements? To name a few, a state could be working at office in the morning, exercising at gym at noon, or having dinner with family at night. We want each state to provide a unified view regarding a user’s activity: (1) where is the user; (2) what is the user doing; and (3) when does this activity happen.
2. How do people move sequentially between those latent states? For example, where do people working in Manhattan usually go to relax after work? and what are the popular sightseeing routes for a one-day trip in Paris? We aim to summarize people’s transitions between the latent states in a concise and interpretable way.
3. How location estimation can help improve human mobility model development?

# Data

In this project, we are going to use 3 different social media data sets: Hurricane Irma, Hurricane Harvey and Hurricane Sandy. All of them include geo-tagged data as well as a large amount of data without location information.

# Tasks

Human Mobility Model: Lu Cheng

I plan to generate an ensemble of Hidden Markov Models (HMMs) to characterize group-level movement regularity.

Location Estimation: Jiayong Mo

Utilizing the text content and reply information from tweets to infer the location. Try to implement Dynamic Bayesian Network technique and augment the performance.

Data Exporation : Linzhen Luo

Implement the language translator for tweet and develop the data collection tool for the model.

## TimeLine

|  |  |
| --- | --- |
| 09-10 | Literature Review |
| 10-11 | Model development |
| 11-12 | Model Improvement & Evaluation |

## Reference

[1] World Disasters Report 2010: Focus on urban risk, ed. D. McClean, 2010: International Federation of Red Cross and Red Crescent Societies.

[2] President's Council of Advisors on Science and Technology, NlT for Resilient Physical Systems,. President's Council of Advisors on Science and Technology, Executive Office of the President. 2007.

[3] Committee on Environment and Natural Resources, Grand Challenges for Disaster Reduction, 2008, National Science and Technology Council, Executive Office of the President.

[4] “Big Data and Disaster Management A Report from the JST/NSF Joint Workshop” Editors: C. Pu and M. Kitsuregawa. Technical Report No. GIT-CERCS-13-09; Georgia Institute of Technology,

CERCS, 2013.

[5] Mason, B., ed. Community Disaster Resilience: A Summary of the March 20, 2006 Workshop of the Disasters Roundtable. 2006.

[6] Kershaw, P.]. and B. Mason, The Indian Ocean Tsunami Disaster: Implications for U.S. and Global Disaster Reduction and Preparedness- Summary of the june 21, 2005 Workshop of the Disasters Roundtable. 2006.

[7] Committee on Planning for Catastrophe: A Blueprint for Improving Geospatial Data Tools and Infrastructure, Successful Response Starts with a Map: Improving Geospatial Support for Disaster Management, 2007.

[8] Committee on Disaster Research in the Social Sciences: Future Challenges and Opportunities, Facing Hazards and Disasters: Understanding Human Dimensions.

[9] Kumar, S., Barbier, G., Abbasi, M. A., & Liu, H. (2011, July). TweetTracker: An Analysis Tool for Humanitarian and Disaster Relief. In *ICWSM*.