Using big data for adaptive social protection

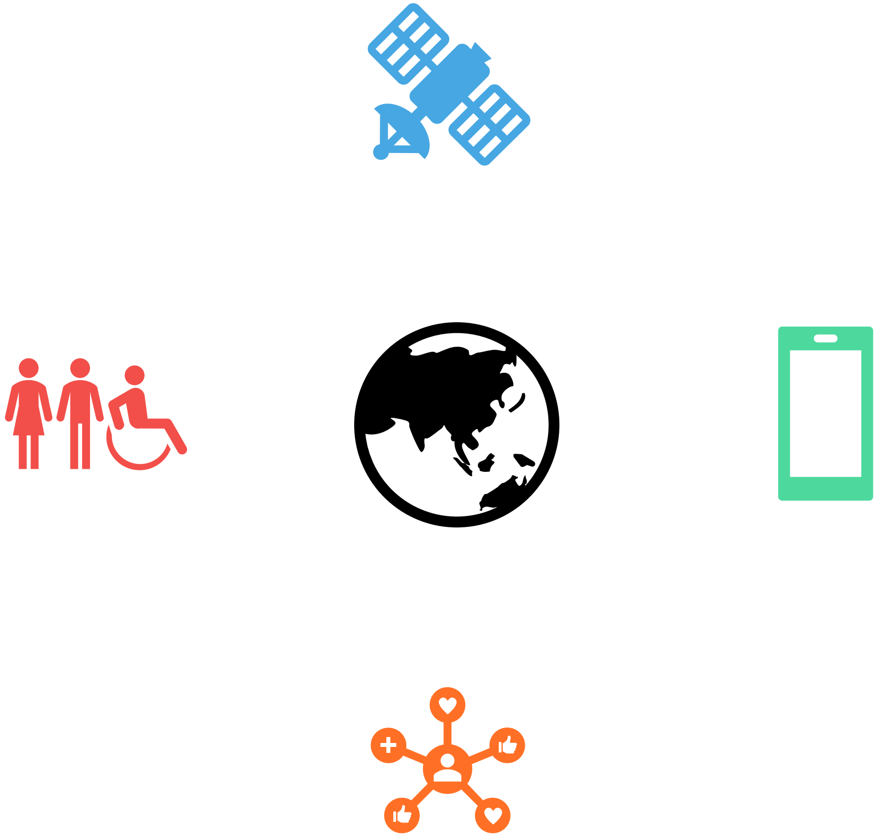
Paul Jasper

Oxford Policy Management

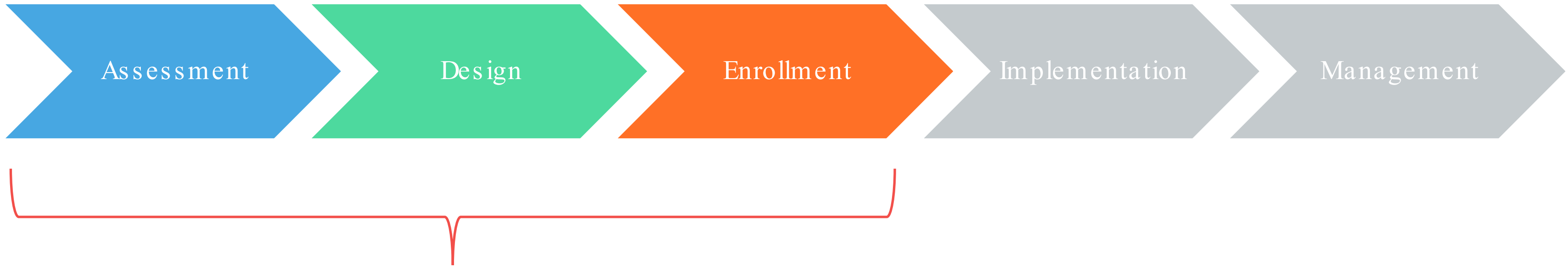
09 March 2023

Content

- Targeting in social protection
- The potential of different types of data:
  - Remote sensing
  - The data exhaust
  - Online data
  - Citizen-reported data
- Opportunities and challenges



Targeting in social protection



New data and new analytical methods, in the context of responses to shocks

"Targeting is the process by which individuals or groups are identified and selected (..), based on their needs and vulnerability"

(Source: Smith et al 2017)

Four examples

Remote sensing

- Remote sensing data includes data from planes, drones, and other flying objects.
- In 2020, about 2,700 functional satellites in orbit. By 2028, over 10,000.
- Some shocks can be ‘observed’ from space, e.g. droughts, floods, conflicts, and can be linked to populations.
- This can inform geographic or household (dwelling)-level targeting.
- Often combined with other geospatial data in machine learning algorithms.

RescueNet: Joint Building Segmentation and Damage Assessment from Satellite Imagery

Rohit Gupta and Mubarak Shah
Center for Research in Computer Vision
University of Central Florida
Orlando, Florida 32816
Email: rohitg@knights.ucf.edu, shah@crcv.ucf.edu

arXiv:2004.07312v1 [cs.CV] 15 Apr 2020

***Abstract*—Accurate and fine-grained information about the extent of damage to buildings is essential for directing Humanitarian Aid and Disaster Response (HADR) operations in the immediate aftermath of any natural calamity. In recent years, satellite and UAV (drone) imagery has been used for this purpose, sometimes aided by computer vision algorithms. Existing Computer Vision approaches for building damage assessment typically rely on a two stage approach, consisting of building detection using an object detection model, followed by damage assessment through classification of the detected building tiles. These multi-stage methods are not end-to-end trainable, and suffer from poor overall results. We propose RescueNet, a unified model that can simultaneously segment buildings and assess the damage levels to individual buildings and can be trained end-to-end. In order to to model the composite nature of this problem, we propose a novel localization aware loss function, which consists of a Binary Cross Entropy loss for building segmentation, and a foreground only selective Categorical Cross-Entropy loss for damage classification, and show significant improvement over the widely used Cross-Entropy loss. RescueNet is tested on the large scale and diverse xBD dataset and achieves significantly better building segmentation and damage classification performance than previous methods and achieves generalization across varied geographical regions and disaster types.**

I. Introduction

In the wake of natural disasters, resources available to first responders are scarce, and efficient planning and allocation of aid and rescue efforts can help save thousands of lives. Traditionally, response planning has been based on reports and estimates based on ground based assessments. Ground based assessments are risky and potentially impossible to obtain, hence, more recently aerial and satellite imagery has been used for these assessments [2]. While analysis of satellite and aerial imagery by experts is useful for rapid response operations, it still results in time lags that could otherwise be spent on rescue operations, as even large teams can take weeks to completely map out disaster affected areas [3]. Automated methods for analyzing aerial and satellite imagery have been developed, including those relying on handcrafted rules for identifying damaged buildings from LiDAR point clouds [4], segmenting the perimeter of forest fires using deep learning [5], detecting flooded regions [6], detecting collapsed, and damaged buildings using Convolutional Neural Networks [7] [8] and detecting damaged buildings using object detectors [9].

In this paper, we focus on assessing damage levels for buildings, which is relevant to all types of natural disasters, and can have a significant impact on the efficacy of search and rescue operations in their aftermath. State of the art methods for detecting damaged buildings [10] [1] rely on a two stage pipeline, where buildings are detected in the pre-disaster imagery in the first stage, and then detected building are classified into different damage levels by comparing pre and post disaster imagery in the second stage. These multi-stage methods are not end-to-end trainable, and suffer from poor overall results. In contrast, we propose RescueNet, which is an end-to-end trainable, unified model to segment buildings and classify their damage levels in one go. We employ a pixel-level segmentation based approach, and use multi-scale,

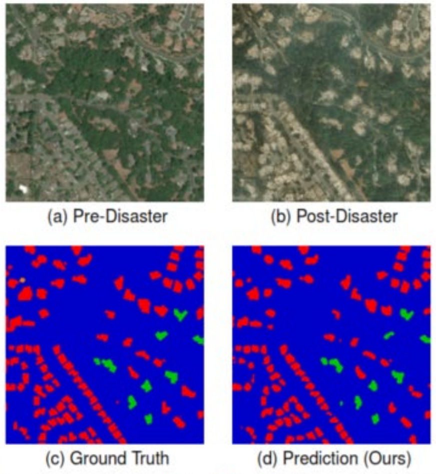


Fig. 1. Samples from the xBD dataset [1] for building damage assessment. (a) and (b) respectively show Pre and Post Disaster images. (c) and shows ground truth and prediction from proposed method. Red labels represent completely damaged buildings, and Green labels represent undamaged buildings. Intermediate damage levels are represented by pink (major damage) and orange (minor damage).

The data exhaust

- Data that are produced as a by-product of people's interactions with digital services in their daily lives.
- Mobile phones, smartphones, mobile money, airtime, credit transactions.
- Location and movement of populations around shocks can be tracked, to help geographic targeting.
- Mobile usage, transactions, credit, airtime, often combined with machine learning, can help to infer wealth and poverty status.
- Individual user-level data has the potential to improve individual-level targeting.

Article

Machine learning and phone data can improve targeting of humanitarian aid

https://doi.org/10.1038/s41586-022-04484-9
Received: 15 July 2021
Accepted: 25 January 2022
Published online: 16 March 2022
Open access
Check for updates

Emily Aiken[1,5], Suzanne Bellue[2], Dean Karlan[3], Chris Udry[4] & Joshua E. Blumenstock[1,5] ✉

The COVID-19 pandemic has devastated many low- and middle-income countries, causing widespread food insecurity and a sharp decline in living standards[1]. In response to this crisis, governments and humanitarian organizations worldwide have distributed social assistance to more than 1.5 billion people[2]. Targeting is a central challenge in administering these programmes: it remains a difficult task to rapidly identify those with the greatest need given available data[3,4]. Here we show that data from mobile phone networks can improve the targeting of humanitarian assistance. Our approach uses traditional survey data to train machine-learning algorithms to recognize patterns of poverty in mobile phone data; the trained algorithms can then prioritize aid to the poorest mobile subscribers. We evaluate this approach by studying a flagship emergency cash transfer program in Togo, which used these algorithms to disburse millions of US dollars worth of COVID-19 relief aid. Our analysis compares outcomes—including exclusion errors, total social welfare and measures of fairness—under different targeting regimes. Relative to the geographic targeting options considered by the Government of Togo, the machine-learning approach reduces errors of exclusion by 4–21%. Relative to methods requiring a comprehensive social registry (a hypothetical exercise; no such registry exists in Togo), the machine-learning approach increases exclusion errors by 9–35%. These results highlight the potential for new data sources to complement traditional methods for targeting humanitarian assistance, particularly in crisis settings in which traditional data are missing or out of date.

The COVID-19 pandemic has led to a sharp decline in living standards across the world, as policies designed to stop the spread of the disease have disrupted normal economic activity. Economically vulnerable households in low- and middle-income countries have been among the hardest hit, with more than 100 million individuals estimated to have transitioned into extreme poverty since the onset of the pandemic[5].

To offset the most severe consequences of this sudden decline in income, governments and humanitarian organizations around the world have mobilized relief efforts. It has been estimated that more than 3,300 new social assistance programmes have been launched[2] since early 2020, providing more than US$800 billion in cash transfer payments to over 1.5 billion people (roughly one fifth of the world's population).

The overwhelming majority of COVID-19 response efforts—and the majority of cash transfer programmes globally—provide targeted social assistance[3,4]. In other words, specific criteria—typically a proxy for socioeconomic status—are used to determine potential eligibility. In most wealthy nations, governments rely on recent household income data to determine programme eligibility[6]. However, in low- and lower middle-income countries (LMICs), where economic activity is often informal and based on home-produced agriculture, governments typically do not observe income for the vast majority of the population[3]. Other potential sources of targeting data are often incomplete or out of date[7,8]; for example, only half of the poorest countries have completed a census in the past 10 years[9]. In such contexts, data gaps preclude governments from implementing well-targeted social assistance programmes[10,11].

Here we develop, implement and evaluate an approach to targeting social assistance based on machine-learning algorithms and non-traditional 'big data' from satellites and mobile phone networks. This approach leverages recent advances in machine learning that show that such data can help accurately estimate the wealth of small geographic regions[12–16] and individual mobile subscribers[17–19]. It also builds on a rich economics literature on the design of appropriate mechanisms for targeting social assistance[3,20–29]. See Supplementary Discussion, section 1 for a summary of previous work.

Humanitarian response to COVID-19 in Togo

Our results are based on the design and evaluation of Novissi, a flagship emergency social assistance programme carried out in Togo. The Government of Togo launched Novissi in April 2020, shortly after the first cases of COVID-19 appeared in the country. As economic lockdown orders forced many Togolese to stop working and led to widespread food insecurity (Supplementary Fig. 1), Novissi aimed to provide subsistence cash relief to those most affected (see https://novissi.gouv.tg/). Eligible beneficiaries received bi-weekly payments of roughly US$10. In an effort to minimize in-person contact, Novissi enrolment and payments were implemented

[1]School of Information, University of California, Berkeley, CA, USA. [2]Department of Economics, University of Mannheim, Mannheim, Germany. [3]Kellogg School of Management, Global Poverty Research Lab, Northwestern University, Evanston, IL, USA. [4]Department of Economics, Global Poverty Research Lab, Northwestern University, Evanston, IL, USA. [5]These authors jointly supervised this work: Emily Aiken, Joshua Blumenstock. ✉e-mail: jblumenstock@berkeley.edu

Online data

- Data shared online purposefully: social media, online articles, Wikipedia.
- Less explored than previous data sources in the context of targeting and social protection.
- High-resolution poverty maps use social media data as one input and can inform geographical targeting.
- Location and mobility tracking via social media can help inform programme design.
- As before, analytical data science techniques are often deployed.

Facebook Disaster Maps: Aggregate Insights for Crisis Response & Recovery

Paige Maas Facebook pmaas@fb.com

Shankar Iyer Facebook shankar94@fb.com

Andreas Gros Facebook andreasg@fb.com

Wonhee Park Facebook wonheepark@fb.com

Laura McGorman Facebook lmcgorman@fb.com

Chaya Nayak Facebook cnayak@fb.com

P. Alex Dow Facebook adow@fb.com

ABSTRACT

After a natural disaster or other crisis, humanitarian organizations need to know where affected people are located and what resources they need. While this information is difficult to capture quickly through conventional methods, aggregate usage patterns of social media apps like Facebook can help fill these information gaps.

In this paper, we describe the data and methodology that power Facebook Disaster Maps. These maps utilize information about Facebook usage in areas impacted by natural hazards, producing aggregate pictures of how the population is affected by and responding to the hazard. The maps include insights into evacuations, cell network connectivity, access to electricity, and long-term displacement.

In addition to descriptions and examples of each map type, we describe the source data used to generate the maps, and efforts taken to ensure the security and privacy of Facebook users. We also describe limitations of the current methodologies and opportunities for improvement.

Keywords

crisis mapping, crisis informatics, GIS, social media

INTRODUCTION

As social media and messaging apps continue to be important communication tools in people's everyday lives, they have also come to play an important role in how people prepare for, respond to, and recover from disasters (Palen and Anderson 2016; Castillo 2016). They are used by people affected by a crisis event, people responding to it, and people following and observing the event from afar (Olteanu et al. 2015). They can be used for individual and mass communication, information seeking, and gaining situational awareness. A significant body of research in crisis informatics has focused on studying these behaviors, and on developing techniques and tools for harnessing social media and other data sources for improved crisis response.

Citizen-reported data

- Citizen-reported data, sometimes referred to as crowd-sourced data, that are purposefully submitted by citizens.
- Relatively well-established for mapping in crisis zones.
- Fewer use cases in poverty or social protection related mapping.
- Potential for feedback mechanisms and citizen agency.


Hunt and Specht *Journal of International Humanitarian Action* (2019) 4:1
https://doi.org/10.1186/s41018-018-0048-1
Journal of International Humanitarian Action

RESEARCH ARTICLE Open Access

Crowdsourced mapping in crisis zones: collaboration, organisation and impact

Amelia Hunt and Doug Specht*

Abstract

Crowdsourced mapping has become an integral part of humanitarian response, with high profile deployments of platforms following the Haiti and Nepal earthquakes, and the multiple projects initiated during the Ebola outbreak in North West Africa in 2014, being prominent examples. There have also been hundreds of deployments of crowdsourced mapping projects across the globe that did not have a high profile. This paper, through an analysis of 51 mapping deployments between 2010 and 2016, complimented with expert interviews, seeks to explore the organisational structures that create the conditions for effective mapping actions, and the relationship between the commissioning body, often a non-governmental organisation (NGO) and the volunteers who regularly make up the team charged with producing the map. The research suggests that there are three distinct areas that need to be improved in order to provide appropriate assistance through mapping in humanitarian crisis: regionalise, prepare and research. The paper concludes, based on the case studies, how each of these areas can be handled more effectively, concluding that failure to implement one area sufficiently can lead to overall project failure.

Keywords: Crowdsourced mapping, Organisational structures, Networks, Humanitarianism, Crisis mapping, Volunteering

Introduction

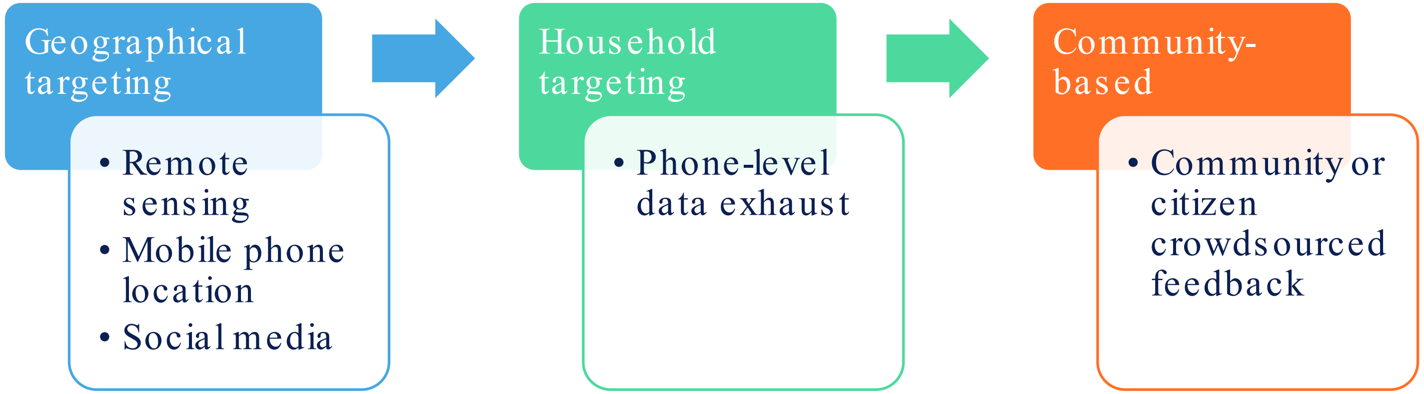
The concept of crowdsourced crisis mapping is perhaps best defined as the provision of services by an international and/or online community, who gather, analyse and map critical information related to disaster-affected populations. Online digital responders often work as part of Volunteer and Technical Communities (V&TCs) which offer free, technical services during, and outside of, humanitarian activations (Capelo et al. 2012). While there have been several explorations of crowdsourced mapping (Walker and Rinner 2013; Meek et al. 2014) and digital humanitarianism (Burns 2015; Meier 2015), this paper considers the current landscape of crowdsourced crisis mapping and the relationship between V&TCs and formal humanitarian organisations. With crowdsourced crisis mapping becoming an ever more prevalent feature of emergency humanitarian response, the need for further research in this field is imperative. Thus far, existing literature has predominantly discussed the technology driving response mechanisms (Meek et al. 2014), with little detail on how technology has been adopted, nor the organisational strategies required to facilitate this. This paper focuses on the organisational nature of these projects, with attention paid to the collaboration between the technical and humanitarian fields, and the changing personal and organisational identities brought about by this global response mechanism.

This paper draws upon a series of high-level interviews to construct a broad impression of how these technologies, and their networks, have been mobilised across 6 years of crisis intervention (2010–2016). This approach seeks to understand the organisational structures required to effectively implement crowdsourced crisis mapping and highlight past points of failure. This research is predicated on two hypotheses, that (i) *the context of crisis plays a more significant role than that assigned to it by V&TCs* and that (ii) *the current level of collaboration across organisations remains inconsistent and inadequate due to poor preparedness strategies for analysing and utilising crowdsourced data.* The paper, firstly, provides context to the emerging field of crisis mapping, before examining existing literature. The methodology is then introduced in more depth, before

* Correspondence: d.specht@westminster.ac.uk
University of Westminster, London, UK

Opportunities and challenges

- Many applications are still in 'pilot' or 'test' stage.
- No single method offers the 'perfect' technology for targeting.
- Potential derives from combinations of a variety of data sources and combining different targeting mechanisms.
- This means they should be complements to each other and to traditional approaches.
- Challenges abound: privacy, data ownership, exclusion of most vulnerable, black box AI, cost, capacity…



Thank you

For further reading:

AI in social protection

New tech to measure poverty

Oxford Policy Management

www.opml.co.uk

OPMglobal