Data-driven Humanitarian Mapping: Harnessing Human-Machine Intelligence for High-Stake Public Policy and Resilience Planning

Snehalkumar 'Neil' S. Gaikwad*
MIT Media Lab
Massachusetts Institute of Technology
Cambridge, MA, USA

Shankar Iyer
Core Data Science
Facebook Research
Menlo Park, CA, USA

Dalton Lunga
Geospatial Science and Human Security Division
Oak Ridge National Laboratory
Oak Ridge, TN, USA

Elizabeth Bondi
Center for Research on Computation and Society
Harvard University
Cambridge, MA, USA

MOTIVATION

Humanitarian challenges, including natural disasters, food insecurity, climate change, racial and gender violence, environmental crises, the COVID-19 coronavirus pandemic, human rights violations, and forced displacements, disproportionately impact vulnerable communities worldwide. According to UN OCHA, 235 million people will require humanitarian assistance in 2021[1]. Despite these growing perils, there remains a notable paucity of data science research to scientifically inform equitable public policy decisions for improving the livelihood of at-risk populations. Scattered data science efforts exist to address these challenges, but they remain isolated from practice and prone to algorithmic harms concerning lack of privacy, fairness, interpretability, accountability, transparency, and ethics. Biases in data-driven methods carry the risk of amplifying inequalities in high-stakes policy decisions that impact the livelihood of millions of people. Consequently, proclaimed benefits of data-driven innovations remain inaccessible to policymakers, practitioners, and marginalized communities at the core of humanitarian actions and global development. To help fill this gap, we propose the Data-driven Humanitarian Mapping Research Program, which focuses on developing novel data science methodologies that harness human-machine intelligence for high-stakes public policy and resilience planning.

VISION AND GOALS

We envision the invention of trustworthy data science for equitable policy decision-making in humanitarian actions, resilience planning, and sustainable development. Over the last few decades, the world has seen unprecedented growth in data science/AI methodologies and heterogeneous datasets (such as earth observation remote sensing, mobile phone data, online social media, surveys, crowdsourcing datasets, etc.). These analytics tools hold the potential to advance our understanding of the spatiotemporal complexities and risks associated with humanitarian and sustainable development challenges. The Data-driven Humanitarian Mapping Research Program drives the design, engineering, and deployment of data-driven ecosystems through computational, legal, ethical, and policy lenses. Specifically, the program focuses on inventing and critically evaluating (1) novel and ethical ways to collect and validate humanitarian, environmental, and socioeconomic development data; (2) data-driven methods for identifying, mapping, and measuring population densities, food insecurity, socioeconomic development, poverty, natural disasters, at-risk communities, infrastructure damages, malaria prevalence, human displacements, and environmental crises, as a case in point; (3) data and community-driven methodologies to formulate equitable high-stake policies and resilience plans for mitigating and containing humanitarian challenges; and (4) evidence-based actionable insights for humanitarian responders, global development networks, NGOs, planners, and policymakers.

CCS CONCEPTS

• **Computing methodologies** → **Machine learning**; **Modeling and simulation**; **Computer vision**; • **Info systems** → **Spatial-temporal systems**; • **Human-centered computing** → **Collaborative and social computing**; • **Social and professional topics** → **Government technology policy**; **Sustainability**; • **Applied computing** → **Decision analysis, Environmental sciences**.

KEYWORDS

human-centered data science, fair and interpretable machine learning, data-driven humanitarian actions, social computing, computational social science, sustainable development, public policy, algorithmic decision making and ethics, remote sensing.

ACM Reference Format:
Snehalkumar 'Neil' S. Gaikwad, Shankar Iyer, Dalton Lunga, and Elizabeth Bondi. 2021. Data-driven Humanitarian Mapping: Harnessing Human-Machine Intelligence for High-Stake Public Policy and Resilience Planning. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '21), August 14–18, 2021, Virtual Event, Singapore.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3447548.3469461

*Lead & corresponding author, gaikwad@mit.edu
[1]UN Office for the Coordination of Humanitarian Affairs (OCHA) Statistics

KDD '21, August 14–18, 2021, Virtual Event, Singapore

ACM ISBN 978-1-4503-8332-5/21/08.
https://doi.org/10.1145/3447548.3469461

KDD WORKSHOP SERIES

As a part of the initiative, we host the second KDD workshop[2] in order to continue fostering a global community of researchers, policymakers, and practitioners to advance a commonly shared data science research agenda for equitable humanitarian actions, resilience planning, and sustainable development. The workshop provides a global platform to (1) share methodological advances in data science to address the pressing humanitarian challenges; (2) conceptualize novel architecture and frameworks for interpretable machine learning, remote sensing, social media, and edge computing technologies for equitable high-stake decision-making; (3) foster effective cross-disciplinary alliances and address risks of data-driven interventions in humanitarian actions and sustainable development. Each of these goals remains instrumental in advancing fundamental research and practice of knowledge discovery in data science.

The initiative[3] was established by the first workshop organizing committee (Neil Gaikwad, Shankar Iyer, Yu-Ru Lin, Dalton Lunga), led by Neil Gaikwad with support from the Advisory and Program Committees (see the Appendix section). The first workshop included two keynote talks, a panel session, and two dozen research talks. Over 100 participants representing various sectors, including academia, international humanitarian organizations, industry, and government agencies, attended the event. The second workshop builds upon this success.

The two keynote talks (*"Interpreting Empty Spaces: How do We Know What We Don't Know?"* by Megan Price and *"21st Century Disaster Response: Data, Methods, and Translational Readiness"* by Caroline Buckee) focused on the role of data-driven humanitarian mapping in armed conflicts and epidemiology (the COVID-19 pandemic), respectively. The panel discussion evolved from a conversation about the intersection of gender and climate change-induced displacements to a broader conversation about how demographics intersect with data analysis in questions of humanitarian relevance. Unifying themes across keynotes, talks, and panel discussions highlighted a wide range of research challenges that currently stand in the way of effectively translating data to measurement development to privacy-preserving mapping to model interpretability to policy engagements and humanitarian actions. We continue to notice a growing interest and need for critical application areas of data science, including sustainability, food security, social inequalities, public policy and urban planning, climate change, and distinctly focused on equitable humanitarian actions and global development.

This year, we will build on the prior success and continue expanding the vibrant community by reaching a diverse demographic of scientists and practitioners. We remain committed to promoting the values of diversity, inclusion, and belonging in data science, computing, and policy research. To foster and sustain an inclusive environment, we have established a Diversity Scholarship for underrepresented groups. Additionally, the workshop guidelines require all participants to adhere to the anti-harassment policy.

[2] 2st KDD Workshop on Data-driven Humanitarian Mapping, Gaikwad et al.,2021
[3] 1st KDD Workshop on Data-driven Humanitarian Mapping, Gaikwad et al.,2020

TOWARDS EQUITABLE DATA SCIENCE AND POLICY DECISION MAKING

The adverse impact of overarching societal and environmental challenges exemplifies the criticality of Data-driven Humanitarian Mapping Research. There remain decisive gaps in the data science field that can be addressed by closer collaboration between cross-disciplinary researchers, practitioners, and at-risk communities. The discipline is at a juncture to convene a broad spectrum of communities to advance fundamental research for informing just public policies for humanitarian actions, resilience planning, and global development. We envision the Data-driven Humanitarian Mapping will bring in new paradigms for equitable data science and policy decision-making while helping create a sustainable world.

APPENDIX

A ADVISORY COMMITTEE 2021 & 2020

Alex 'Sandy' Pentland (MIT), Budhu Bhaduri (Oak Ridge National Laboratory), Carlos Castillo (Universitat Pompeu Fabra), Danielle Wood (MIT), Haishan Fu (World Bank), Hamed Alemohammad (Radiant Earth Foundation), Jie Yin (University of Sydney), Joshua Blumenstock (University of California, Berkeley), Marta Gonzalez (University of California, Berkeley), Megan Price (Human Rights Data Analysis Group), Miho Mazereeuw (MIT), Rahul Panicker (Wadhwani AI), Ramesh Raskar (MIT). Vipin Kumar (University of Minnesota), Yu-Ru lin (University of Pittsburgh).

B PROGRAM COMMITTEE 2021 & 2020

Abigail Horn, Aditya Mate, Afreen Siddiqi, Alex Dow, Alex Pompe, Amanda Coston, Amanda Hughes, Amy Rose, Andrés Abeliuk, Antonio Fernández Anta, Aruna Sankaranarayanan, Ashique Khudabukhsh, Ayesha Mahmud, Bianca Zadrozny, Bryan Wilder, Chintan Vaishnav, Clio Andris, Delia Wendel, Emily Aiken, Eric Shook, Eric Sodomka, Esteban Moro, Eugenia Giraudy, Gautum Thakur, Grigorios Kalliatakis, Hannah Kerner, Helene Benveniste, Hemant Purohit, Jackson Killian, Jigar Doshi, Jill Kelly, Katherine Hoffmann Pham, Kentaro Torisawa, Kevin Sparks, Kiran Garimella, Kostas Pelechrinis, Kristen Altenburger, Kuldeep Kurte, Kurt Luther, Laura Alessandretti, Lijun Sun, Lily Xu, Marie Urban, Marie-Laure Charpignon, Marina Kogan, Meeyoung (Mia) Cha, Meha Jain, Mert Ertugrul, Minoo Rathnasabapathy, Morgan Frank, Muhammad Imran, Nagendra Singh, Niklas Stoehr, Nishant Kishore, Nitin Kohli, Omari Sefu, Paige Maas, Philipe Dias, Rishemjit Kaur, Robert Soden, Sai Ravela, Samira Barzin, Shawn Newsam, Sherrie Wang, Shruti Palaskar, Silvia Danielak, Siqi Zheng, Soroush Vosoughi, Tao Liu, Trivik Verma, Vivek Singh, Xidao Wen, Yan Leng, Yingjie Hu.

C VOLUNTEER TEAM 2021 & 2020

Bhavani Ananthabhotla, Indrayani Deshmukh, Nikhil Behari, Takahiro Yabe, Xiaofan Liang.

D SPEAKERS AND PANELISTS 2020

Caroline Buckee (Harvard University), Dyan Mazurana, (Tufts University), Justin Ginnetti (IDMC), Karen Chapple (University of California, Berkeley), Megan Price, (Human Rights Data Analysis Group), Paige Maas (Facebook Core Data Science).