EDITORIAL

Big Data for Social Good

Charlie Catlett,^{1,2,*} and Rayid Ghani^{2,*}

The use of large-scale and new, emerging sources of data to make better decisions has been increasing over the past several years. Corporations have been the first to act on this potential in industries such as search, advertising, finance, surveillance, retail, and manufacturing. At the same time, organizations focused on social good are realizing the potential as well but face several challenges as they seek to become more data-driven. The biggest challenge they face is a paucity of examples and case studies on how data can be used for social good. This special issue of *Big Data* is targeted at tackling that challenge and focuses on highlighting some exciting and impactful examples of work that uses data for social good. The special issue is just one example of the recent surge in such efforts by the data science community. Some examples of these efforts include:

- The 2014 ACM's annual Conference on Knowledge Discovery and Data Mining (KDD) designating "Data Science for Social Good" as the conference theme and including a full-day workshop on "Data Science for Social Good."
- University of Chicago creating the Eric & Wendy Schmidt Data Science for Social Good Summer Fellowships in 2013, which is entering its third year and getting a large number of data science students trained and passionate about making a social impact.
- Nonprofits such as Datakind and Bayes Impact expanding efforts to get more people involved in helping nonprofits and governments.

This special issue solicited case studies and problem statements that would either highlight (1) the use of data to solve a social problem or (2) social

challenges that need data-driven solutions. From roughly 20 submissions, we selected 5 articles that exemplify this type of work. These cover five broad application areas: international development, health-care, democracy and government, human rights, and crime prevention.

"Understanding Democracy and Development Traps Using a Data-Driven Approach" (Ranganathan et al.) details a data-driven model between democracy, cultural values, and socioeconomic indicators to identify a model of two types of "traps" that hinder the development of democracy. They use historical data to detect causal factors and make predictions about the time expected for a given country to overcome these traps.

"Targeting Villages for Rural Development Using Satellite Image Analysis" (Varshney et al.) discusses two case studies that use data and machine learning techniques for international economic development—solar-powered microgrids in rural India and targeting financial aid to villages in sub-Saharan Africa. In the process, the authors stress the importance of understanding the characteristics and provenance of the data and the criticality of incorporating local "on the ground" expertise.

In "Human Rights Event Detection from Heterogeneous Social Media Graphs," Chen and Neil describe efficient and scalable techniques to use social media in order to detect emerging patterns in human rights events. They test their approach on recent events in Mexico and show that they can accurately detect relevant human rights-related tweets prior to international news sources, and in some cases, prior to local news reports, which could potentially lead to more timely, targeted, and effective advocacy by relevant human rights groups.

¹Argonne National Laboratory, Lemont, Illinois.

²University of Chicago, Chicago, Illinois.

*Address correspondence to: Charlie Catlett or Rayid Ghani, University of Chicago, 5735 South Ellis Ave., Chicago, IL 60637, E-mail: cec@uchicago.edu or rayid@uchicago.edu

“Finding Patterns with a Rotten Core: Data Mining for Crime Series with Core Sets” (Wang et al.) describes a case study with the Cambridge Police Department, using a subspace clustering method to analyze the department’s full housebreak database, which contains detailed information from thousands of crimes from over a decade. They find that the method allows human crime analysts to handle vast amounts of data and provides new insights into true patterns of crime committed in Cambridge.

Our intentions and hopes for including these examples in this special issue are to (1) highlight the

contributions of the data science community toward social good, (2) to motivate current and future data science researchers and practitioners to direct more attention and effort toward these problems, and (3) to encourage organizations working toward social good to collaborate with data scientists and apply data-driven methods to increase their impact.

Cite this article as: Catlett C, Ghani R (2015) Big data for social good. *Big Data* 3:1, 1–2, DOI: 10.1089/big.2015.1530.