Online Ethics Center
FOR ENGINEERING AND SCIENCE

Case: Big Data & Public Health

Author(s)

Valerie Racine

Year

2017

Description

A computer science graduate has an opportunity to work managing large data sets from smartphone applications designed to track individuals' personal health information. His father, an epidemiologist, encourages him to consider the project's use of personal data, as well as the ability to design algorithms that could accurately predict and track outbreaks based on this kind of data.

Body

Zhang Kar-wai, a graduate from Stanford University’s department of computer science, has recently accepted his first paid position at a start-up company in San Francisco. The company’s main products consist of a line of smartphone applications designed to track individuals’ personal health information, including their medical records, their seasonal illnesses, their blood pressure and blood glucose levels, their eating habits, their sleep cycles, and even their weight and reproductive health. The products are designed primarily to help individuals reach health-related goals, and enable users to manage their overall health and wellbeing.

Zhang was hired to join the team responsible for managing the large data sets generated by the users of these apps. One of his responsibilities will be to develop

algorithms and analytic tools that can track the outbreak and spread of infectious diseases in real-time using data gathered from individuals using their applications. Their goal is to improve on the traditional methods used by the US's Center for Disease Control and Prevention (CDC) and the UN's World Health Organization (WHO).

Zhang loved developing algorithms as a student and he is looking forward to participating in the team's project. He was excited to tell his father, a professor of epidemiology at Stanford's School of Medicine, about his assignment. However, when Zhang told his father about his responsibilities at his new job, his father's reaction was not what he expected. Zhang's father expressed concern about the project's use of personal data, as well as their ability to design algorithms that could accurately predict and track outbreaks based on their data.

Zhang explained to his father that the data will be aggregated and the algorithms will undoubtedly sometimes fail, as all models do, but they will be continuously tested and upgraded. So, he assured his father that he was up to the task. His father remained hesitant to share in Zhang's enthusiasm and warned him to think more about the ethical implications of his project, and not just about whether his algorithms will succeed or not. Should Zhang take his father's concerns more seriously?

Discussion Questions

1. Is it ethically permissible to use data from Internet search engines or applications for national public health purposes? If so, does this mean it is also permissible to use privately collected data in global public health contexts?
2. Should users of online applications or search engines be notified about the potential use of their personal data (even in aggregated form) for public health measures? Why/why not?
3. What are some of the ethical risks of the proposed big data analytics or algorithms, which can sometimes lead to false positives in their efforts to identify outbreaks and/or predict outbreak trajectories? What might be done to mitigate such risks?

Bibliography

Butler, Declan. “When Google got flu wrong.” *Nature* 494, no. 7436 (2013): 155.

Brownstein, John S., Clark C. Freifeld, and Lawrence C. Madoff. “Digital disease detection—harnessing the Web for public health surveillance.” *New England Journal of Medicine* 360, no. 21 (2009): 2153-2157.

Cook, Samantha, Corrie Conrad, Ashley L. Fowlkes, and Matthew H. Mohebbi. “Assessing Google flu trends performance in the United States during the 2009 influenza virus A (H1N1) pandemic.” *PloS one* 6, no. 8 (2011): e23610.

Ginsberg, Jeremy, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski, and Larry Brilliant. “Detecting influenza epidemics using search engine query data.” *Nature* 457, no. 7232 (2009): 1012-1014.

Ives, Mike. “When Epidemics go Viral.” *The Atlantic*. October 18, 2016. https://www.theatlantic.com/health/archive/2016/10/when-epidemics-go-viral/504503/. Accessed November 2, 2016.

Lazar, David, and Ryan Kennedy. “What can we learn from the epic failure of Google Flu Trends.” *Wired*. October 1, 2015. https://www.wired.com/2015/10/can-learn-epic-failure-google-flu-trends/ Accessed November 2, 2016.

Lazer, David, Ryan Kennedy, Gary King, and Alessandro Vespignani. “The parable of Google flu: traps in big data analysis.” *Science* 343, no. 6176 (2014): 1203-1205.

Mikal, Jude, Samantha Hurst, and Mike Conway. “Ethical issues in using Twitter for population-level depression monitoring: a qualitative study.” *BMC medical ethics* 17, no. 1 (2016): 1.

Salathe, Marcel, Linus Bengtsson, Todd J. Bodnar, Devon D. Brewer, John S. Brownstein, Caroline Buckee, Ellsworth M. Campbell et al. “Digital epidemiology.” *PLoS Comput Biol* 8, no. 7 (2012): e1002616.

Vayena, Effy, Marcel Salathé, Lawrence C. Madoff, and John S. Brownstein. “Ethical challenges of big data in public health.” *PLoS Comput Biol* 11, no. 2 (2015): e1003904.

Vayena, Effy, Anna Mastroianni, and Jeffrey Kahn. “Ethical issues in health research with novel online sources.” *American journal of public health* 102, no. 12 (2012): 2225-2230.

White, Michael. “The Ethical Risks of Detecting Disease Outbreaks with Big Data.” *Pacific Standard*. February 24, 2015. https://psmag.com/the-ethical-risks-of-detecting-disease-outbreaks-with-big-data-37ba15345aa7#.5cwzsw7b8. Accessed November 2, 2016.

Notes

The author wishes to acknowledge the contributions of Karin Ellison, OEC - Life and Environmental Sciences Editor, and Joseph Herkert, OEC Engineering co-Editor. They provided valuable input in selecting topics and crafting the resources.

Contributor(s)

Valerie Racine
Karin Ellison
Joseph Herkert

Rights

Resource Type

Case Study / Scenario

Parent Collection

Big Data in the Life Sciences Collection

Topics

Big Data
Controversies
Emerging Technologies
Privacy and Surveillance
Social Responsibility

Discipline(s)

Life and Environmental Sciences
Public Health