Chapter 6
Big Data Dilemmas:
The Theory and Practice of Ethical Big Data Mining for Socio-Economic Development

Debashis "Deb" Aikat
University of North Carolina at Chapel Hill, USA

ABSTRACT

This chapter delineates the theory and practice of ethical big data mining for socio-economic development in four parts. The first part enunciates the ethical role of big data mining for socio-economic development by theorizing big data as a 20th Century phenomenon and its surging significance in the 21st Century digital era. The second part elucidates ethical values relating to big data mining with particular emphasis on the interplay of theory and practice. The third part connects classical theories of ethics to propose a code of conduct that relates to core ethical values such as privacy, confidentiality, objectivity, transparency, conflict of interest, and common good. The fourth and final part identifies privacy as a major challenge of ethical big data mining and postulates needed research directions. This chapter also features a list of additional reading and big data terms with concise definitions explicating their relevance to big data mining for socio-economic development.

Man is about to be an automaton; he is identifiable only in the computer. As a person of worth and creativity, as a being with an infinite potential, he retreats and battles the forces that make him inhuman.

– United States Supreme Court Justice William O. Douglas, in "Points of Rebellion" (Douglas, 1970, pp. 32-33)

Ethics grow out of the same root as world- and life-affirmation, for ethics, too, are nothing but reverence for life. That is what gives me the fundamental principle of morality, namely, that good consists in maintaining, promoting, and enhancing life, and that destroying, injuring, and limiting life are evil.

– Nobel Laureate and Theologian Albert Schweitzer in "The Philosophy of Civilization," (Schweitzer, 1959, p. 79)

These quotations signify words of wisdom that explicate complex ethical issues surrounding big data mining for socio-economic development and its related research processes that enhance human knowledge. While big data[1] may uncover patterns of human behavior and help predict social, economic and political trends, unethical use of big data may disrupt the data mining landscape. For instance, some corporations monopolize big data for illicit profits and unscrupulous government administrators may use big data to manipulate their citizens' rights to privacy to foster a digital surveillance state.

From a socio-economic development perspective, people in all modern societies are surrounded by big data surveillance with meticulously-designed computer algorithms that sift through big data to reveal important insights about people, products and perspectives. Ethical big data mining has the unique potential to measure and monitor trends that are critical for socio-economic development. For instance, ethical big data mining may include analyses of population census data to identify low income regions that need economic aid, insights into consumer behavior based on web-browsing trails of online shopping trends, warnings about natural disasters predicted by weather sensor signals, studies to improve public transit with global positioning system (GPS) tracking, development of healthcare intervention based on disease trends, and unification of social network messages to predict changes in public opinion, crime and job opportunities among other indicators of socio-economic development.

A commitment to ethical values and best practices sustains big data mining (Chhabra & Rahman, 2011; Rahman, 2009a; 2009b; 2009c). By situating big data mining within a larger intellectual context of ethics, this chapter presents big data dilemmas and analyzes the theory and practice of ethical big data mining for socio-economic development at the international level. Drawing upon a meta-analysis of the theory and practice of big data mining, this chapter delineates in four parts some big data dilemmas for ethical big data mining for socio-economic development.

The first part enunciates the ethical role of big data mining for socio-economic development by delineating the evolution of big data as a 20th century phenomenon and their surging significance in the 21st century digital era. The second part elucidates ethical values relating to big data mining with particular emphasis on the interplay of theory, concept and practice. The third part connects classical theories of ethics with the central purpose of big data mining to propose a code of conduct for ethical big data mining. This code of conduct relates to core ethical values such as privacy, confidentiality, objectivity, transparency, conflict of interest, and common good. The fourth and final part identifies privacy as a major challenge of ethical big data mining and postulates needed research directions by exploring theoretically-grounded implications that advance the field of ethical data mining. This chapter also features a list of additional reading and big data terms with concise definitions explicating their relevance to big data mining for socio-economic development.

INTRODUCTION

Empowered by big data applications, data miners can now identify statistically significant trends in human behavior, shopping habits and other preferences. Big data are transforming data mining research (Ratner, 2003; Koops, 2011; Boyd & Crawford, 2012; Kundra, 2012; Oboler, Welsh, & Cruz, 2012; Mohin, 2012). Big data mining blends large, diverse, complex, longitudinal, and distributed data sets generated from instruments, sensors, Internet transactions, email, video, click streams, and all other myriad sources (Bollier, 2010; Brynjolfsson et al., 2011; Franks, 2012; Davis & Patterson, 2012). Big data originate from "everywhere: sensors used to gather climate information, posts to social media sites, digital

Related Content

The Model-Driven Architecture for the Trajectory Data Warehouse Modeling
Noura Azaiezand Jalel Akaichi (2020). *International Journal of Data Warehousing and Mining (pp. 26-43).*
www.irma-international.org/article/the-model-driven-architecture-for-the-trajectory-data-warehouse-modeling/265255

Sentiment Analysis of Game Review Using Machine Learning in a Hadoop Ecosystem
Arvind Panwarand Vishal Bhatnagar (2022). *Research Anthology on Implementing Sentiment Analysis Across Multiple Disciplines (pp. 463-483).*
www.irma-international.org/chapter/sentiment-analysis-of-game-review-using-machine-learning-in-a-hadoop-ecosystem/308503

A Highly Scalable and Adaptable Co-Learning Framework on Multimodal Data Mining in a Multimedia Database
Zhongfei (Mark) Zhang, Zhen Guo, Christos Faloutsosand Jia-Yu Pan (2013). *Data Mining: Concepts, Methodologies, Tools, and Applications (pp. 567-586).*
www.irma-international.org/chapter/highly-scalable-adaptable-learning-framework/73458

Case Study: Efficient Faculty Recruitment Using Genetic Algorithm
Amit Verma, Iqbaldeep Kaur, Dolly Sharmaand Inderjeet Singh (2019). *Extracting Knowledge From Opinion Mining (pp. 299-311).*
www.irma-international.org/chapter/case-study/211566

Graph-Based Modelling of Concurrent Sequential Patterns
Jing Lu, Weiru Chenand Malcolm Keech (2012). *Exploring Advances in Interdisciplinary Data Mining and Analytics: New Trends (pp. 110-127).*
www.irma-international.org/chapter/graph-based-modelling-concurrent-sequential/61171