Chapter 7
Social Science Data Analysis:
The Ethical Imperative

Anthony Scime
State University of New York, USA

Gregg R. Murray
Texas Tech University, USA

ABSTRACT

Social scientists address some of the most pressing issues of society such as health and wellness, government processes and citizen reactions, individual and collective knowledge, working conditions and socio-economic processes, and societal peace and violence. In an effort to understand these and many other consequential issues, social scientists invest substantial resources to collect large quantities of data, much of which are not fully explored. This chapter proffers the argument that privacy protection and responsible use are not the only ethical considerations related to data mining social data. Given (1) the substantial resources allocated and (2) the leverage these "big data" give on such weighty issues, this chapter suggests social scientists are ethically obligated to conduct comprehensive analysis of their data. Data mining techniques provide pertinent tools that are valuable for identifying attributes in large data sets that may be useful for addressing important issues in the social sciences. By using these comprehensive analytical processes, a researcher may discover a set of attributes that is useful for making behavioral predictions, validating social science theories, and creating rules for understanding behavior in social domains. Taken together, these attributes and values often present previously unknown knowledge that may have important applied and theoretical consequences for a domain, social scientific or otherwise. This chapter concludes with examples of important social problems studied using various data mining methodologies including ethical concerns.

INTRODUCTION

Social and economic developments are driven by human behavior, and the social sciences are the academic and professional disciplines that study society and human behaviors. These disciplines include but are not limited to: anthropology, communication, criminology, economics, education, history, human geography, law, linguistics, political science, psychology, social work, and sociology. Social scientists address such questions as (Giles, 2011): "How can we induce people to look after their health?" "How do societies create effective and resilient institutions, such as governments?" "How can humanity increase its collective wisdom?" "Why do so many female workers still earn less than male workers?" And "Why do social processes, in particular civil violence, either persist over time or suddenly change?" In an effort to understand the issues in which they are interested, social scientists invest vast resources to ask subjects thousands of questions in surveys and observe millions of behaviors that usually are systematically archived but often are not analyzed in a comprehensive manner.

The protection of privacy and responsible implementation are not the only ethical considerations related to data mining social data. Given (1) the substantial resources allocated and (2) the leverage these "big data" give on such weighty issues, this chapter suggests social scientists are, in addition, ethically obligated to conduct comprehensive analysis of their data, an obligation that data mining techniques can facilitate (National Science Foundation, 2009; Rosenthal, 1994). To support this argument, the chapter begins with a background discussion about resources invested in social data and then a review of some major data sources used in the social sciences. This is followed by examples of how data mining has been used in analysis of political science, social work, sociology, health science, education, and criminal justice issues. These examples only scratch the surface of what can be done with the collected data to meet this ethical obligation. The section on future research directions notes the problem of under-analyzed data will get worse, but new technologies will be developed to solve or mitigate the problem. We conclude on the hopeful note that as data mining becomes main stream for social science analysis the backlog of social science data will decline, and, more importantly, more social problems will be resolved.

BACKGROUND

Social scientists allocate substantial resources to data collection. At colleges and universities in the United States more than $4.4 billion was spent on social science research and development in fiscal year 2009 (National Science Board, 2012). Governments and non-government organizations spend billions of dollars more collecting data on society. This investment represents not only a financial expenditure but also the expenditure of countless days, weeks, and months of researcher, participant, and administrator time and effort. The results of this immense investment are often embodied in extensive data sets. Minimal efficiency demands reasonable output from this substantial input in data collection. This is not a new concept. Rosenthal (1994, p. 130) contends there is a larger obligation:

[D]ata are expensive in terms of time, effort, money, and other resources… If the research was worth doing, the data are worth a thorough analysis, being held up to the light in many different ways so that our research participants, our funding agencies, our science, and society will all get their time and their money's worth.

Further, these data sets may contain answers to some of society's pressing issues. Very broadly, the challenge to social investigators as stated in the National Science Foundation's mission statement is lofty: "to promote the progress of science; to

Related Content

Essentiality of Machine Learning Algorithms for Big Data Computation

Manjunath Thimmasandra Narayanapppa, T. P. Puneeth Kumarand Ravindra S. Hegadi (2016). *Big Data: Concepts, Methodologies, Tools, and Applications (pp. 887-898).*

www.irma-international.org/chapter/essentiality-of-machine-learning-algorithms-for-big-data-computation/150198

A Query Language for Mobility Data Mining

Roberto Trasarti, Fosca Giannotti, Mirco Nanni, Dino Pedreschiand Chiara Renso (2011). *International Journal of Data Warehousing and Mining (pp. 24-45).*

www.irma-international.org/article/query-language-mobility-data-mining/49639

Concept of Temporal Pretopology for the Analysis for Structural Changes: Application to Econometrics

Nazha Selmaoui-Folcher, Jannai Tokotoko, Samuel Gorohouna, Laisa Roi, Claire Leschiand Catherine Ris (2022). *International Journal of Data Warehousing and Mining (pp. 1-17).*

www.irma-international.org/article/concept-of-temporal-pretopology-for-the-analysis-for-structural-changes/298004

A Dynamic Privacy Manager for Compliance in Pervasive Computing

Riccardo Bonazzi, Zhan Liu, Simon Ganièreand Yves Pigneur (2013). *Data Mining: Concepts, Methodologies, Tools, and Applications (pp. 793-815).*

www.irma-international.org/chapter/dynamic-privacy-manager-compliance-pervasive/73471

SLOD-BI: An Open Data Infrastructure for Enabling Social Business Intelligence

Rafael Berlanga, Lisette García-Moya, Victoria Nebot, María José Aramburu, Ismael Sanzand Dolores María Llidó (2016). *Big Data: Concepts, Methodologies, Tools, and Applications (pp. 1784-1813).*

www.irma-international.org/chapter/slod-bi/150243