

Chapter 1.22

Moral Foundations of Data Mining

Kenneth W. Goodman
University of Miami, USA

INTRODUCTION

It has become a commonplace observation that scientific progress often, if not usually, outstrips or precedes the ethical analyses and tools that society increasingly relies on and even demands. In the case of data mining and knowledge discovery in databases, such an observation would be mistaken. There are, in fact, a number of useful ethical precedents, strategies, and principles available to guide those who request, pay for, design, maintain, use, share, and sell databases used for mining and knowledge discovery. These conceptual tools—and the need for them—will vary as one is using a database to, say, analyze cosmological data, identify potential customers, or run a hospital. But these differences should not be allowed to mask the ability of applied ethics to provide practical guidance to those who work in an exciting and rapidly growing new field.

BACKGROUND

Data mining is itself a hybrid discipline, embodying aspects of computer science, artificial intelligence, cryptography, statistics, and logic. In greater or lesser degree, each of these disciplines has noted and addressed the ethical issues that arise in their practice. In statistics, for instance, leading professional organizations have ratified codes of ethics that address issues ranging from safeguarding privileged information and avoiding conflicts of interest or sponsor bias (International Statistical Institute, 1985) to “the avoidance of any tendency to slant statistical work toward predetermined outcomes” (American Statistical Association, 1999).

It is in computer ethics, however, that one finds the earliest, sustained, and most thoughtful literature (Bynum, 1985; Johnson & Snapper, 1985; Ermann, Williams & Gutierrez, 1990;

Forester & Morrison, 1994; Johnson, 1994) in addition to ethics codes by professional societies (Association for Computing Machinery, 1992; IEEE, 1990). Traditional issues in computer ethics include privacy and workplace monitoring, hacking, intellectual property, and appropriate uses and users. The intersection of computing and medicine has also begun to attract interest (Goodman, 1998a).

What is clear about this landscape is that terms we attach to issues—"privacy," for instance—can mask significant differences according as one uses a computer to keep track of warehouse stock or arrest records or real estate transactions or sexually transmitted diseases. Moreover, ethical issues take on somewhat different aspects depending on whether a computer and its data storage media are used by an individual, a business, a university, or a government. Atop this is the general purpose to which the machine is put: science, business, law enforcement, public health, or national security. This triad—content, user, and purpose—frames the space in which ethical issues arise.

MAIN THRUST

One can identify a suite of ethical issues that arise in data mining. All are tethered in one way or another to issues encountered in computing, statistics, and kindred fields. The question of whether data mining, or any discipline for that matter, presents unique or unprecedented issues is open to dispute. Issues between or among disciplines often vary by degree more than by kind. If it is learned or inferred from a database that Ms. Garcia prefers blue frocks, it might be the case that her privacy has been violated. But if it is learned or inferred that she has HIV, the stakes are altogether different. The ability to make ever-more-fine-grained inferences from very large databases increases the importance of ethics in data mining.

Privacy, Confidentiality, and Consent

It is common to distinguish between privacy and confidentiality by applying the former to humans' claim or desire to control access to themselves or information about them, and the latter, more narrowly, to specific units of that information. Privacy, if you will, is about people; confidentiality is about information. Privacy is broader, and it includes interest in information protection and control.

An important correlate of privacy is consent. One cannot control information without being asked for permission, or at least informed of information use. A business that is creating a customer database might collect data surreptitiously, arguably infringing on privacy. Or it might publicly—though not necessarily individually—disclose the collection. Such disclosures are often key components of privacy policies. Privacy policies sometimes seek permission in advance to obtain and archive personal data or, more frequently, disclose that data are being collected and then provide a mechanism for individuals to opt out. The question whether such opt-out policies are adequate to give individuals opportunities to control use of their data is subject to widespread debate. In another context, a government will collect data for vital statistics or public health databases. Such uses, at least in democratic societies, may be justified on grounds of the implied consent of those to whom the information applies and who would benefit from its collection.

It is not clear how much or what kind of consent would be necessary to provide ethical warrant for data mining of personal information. The problem of adequate consent is complicated by what may be hypothesized to be widespread ignorance about data mining and its capabilities. As elsewhere, some solutions to this ethical problem might be identified or clarified by empirical research related to public understanding of data-mining technology, individuals' preference for (levels of) control over use of their information, and similar

5 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/moral-foundations-data-mining/23093

Related Content

An Efficient, Anonymous and Unlinkable Incentives Scheme

Milica Milutinovic, Andreas Putand Bart De Decker (2015). *International Journal of Information Security and Privacy* (pp. 1-20).

www.irma-international.org/article/an-efficient-anonymous-and-unlinkable-incentives-scheme/148300

Maturing an Information Technology Privacy Program: Assessment, Improvement, and Change Leadership

Mike Gregoryand Cynthia Roberts (2021). *Research Anthology on Privatizing and Securing Data* (pp. 2008-2022).

www.irma-international.org/chapter/maturing-an-information-technology-privacy-program/280269

Blockchain-Based Educational Management and Secure Software-Defined Networking in Smart Communities

Bin Fang (2022). *International Journal of Information Security and Privacy* (pp. 1-20).

www.irma-international.org/article/blockchain-based-educational-management-and-secure-software-defined-networking-in-smart-communities/308314

Impact of Financial Risk Ratios on Profitability of Multinational vs. Domestic Pharmaceuticals in India

Kaushik Chakraborty (2014). *International Journal of Risk and Contingency Management* (pp. 54-68).

www.irma-international.org/article/impact-of-financial-risk-ratios-on-profitability-of-multinational-vs-domestic-pharmaceuticals-in-india/115819

Dynamic Warnings: An Eye Gaze-Based Approach

Mini Zeng, Feng Zhuand Sandra Carpenter (2022). *International Journal of Information Security and Privacy* (pp. 1-28).

www.irma-international.org/article/dynamic-warnings/303662