Chapter 5.13
Metaphors and Models for Data Mining Ethics

Peter Danielson
University of British Columbia, Canada

ABSTRACT

Our regulatory institutions, broadly taken, include our moral norms and models and have not fully adapted to significant changes in data mining technology. For example, we suggest that the metaphors — Big Brother and "data mining" itself — commonly used to describe and assess this new technology are deficient, overemphasizing social discipline by the state and the passivity of the so-called data subject. We move from metaphors to a set of models more adequate for building an ethics of data mining, using a framework of informal game theory. We sketch three models of interaction: pure conflict, pure coordination, and a mixed motive cooperation game, with special application to security, health, and commerce, respectively. We recommend these three models as heuristics within a simple account of an ethics of data mining regulated by informed consent.

INTRODUCTION

Many problems in the ethics of technology arise because our regulatory institutions, including our moral norms and mental models, take time to adapt to technological change. Data mining is a good example of this institutional inertia. Twenty years ago, (Clarke, 1988) set out a policy framework for the emerging information technologies that he called 'dataveillance' that now includes data mining (see Key Terms). As Clarke predicted, the growth of information technology in general, and the Internet in particular, has exacerbated the problems he catalogued. Nonetheless, neither the weak regulatory framework nor individual ignorance of common data mining practice has changed in the U.S. We will focus on the U.S. as the extreme case of a democratic society where data mining technology is highly developed and widely used but weakly regulated and poorly understood. (We will discuss the (Turow, Feldman, & Meltzer, 2005) survey data below, as well

as a survey of U.K. physicians attitudes towards privacy, to ground our discussion in what may be local cultural attitudes towards privacy.) Where early U.S. database matching focused on target subjects in government databases – welfare recipients and government employees – today almost everyone in the U.S. economy is a data subject in multiple data bases. For example, "Acxiom gathers and sorts information about 196 million Americans …'Metromail…has a detailed data base on more than 90 per cent of American households'" (Whitaker, 1999, pp. 132 -13 3).

Discussing a related topic, (Danielson, 2005) argued that inappropriate informal models hampered our understanding of surveillance and impeded construction of an adequate ethics for the subject. In that case, metaphors such as Jeremy Bentham's Panopticon supported static and one-sided thinking about surveillance, displacing the more nuanced models needed for ethics. In this chapter we extend this argument from surveillance to data mining, where metaphors and models are even more central to understanding the more abstract technology involved. We begin with the deficiencies in two common metaphors that guide thinking about data mining, beginning with "Big Brother" and moving to "data mining" itself. We suggest moving from metaphors to more precise models, sketching three broad types of interaction in which data mining plays a role: pure conflict, pure coordination, and a mixed motive cooperation game. This chapter deploys a framework of informal game theory to elucidate some of the ethical issues raised by data mining technology. We do not use "games" to diminish the importance of the issues we discuss, but rather to highlight their interactive, strategic, and dynamic aspects. Indeed, a game theoretic approach is particularly suited to the topic of data mining, for two reasons. First, descriptively, the point of morally significant data mining is strategic. In both main uses of the technology, security and commerce, one mines personal data to know more about one's opponent, aiming to change the terms of interaction in one's favor. Therefore, taking a game theory approach does not import or overemphasize strategic considerations. A second and more normative reason to use a game theory approach is to emphasize people as active agents in contrast to the passive data subjects assumed by some data mining practice. This agent-centered approach provides us with a critical perspective, focusing on how choices and alternatives might be better structured by data mining technologies, regulations, and norms, and thus yielding new opportunities for ethics to guide us.

BACKGROUND: MORALS, METAPHORS, & MODELS

Moral Scope

The basic technologies used in data mining—machine learning and automated statistical analysis applied to large data bases—are not themselves morally problematic. They only become so when used on morally significant data, typically information about persons gathered without their full consent (Wahlstrom & Roddick, 2001, p. 23). Contrast the case of bioinformatics – data mining applied to genomic and other biological data – applied to a typical species used in research, such as the c. elegans worm and applied to humans. In the worm case, with no morally significant personal data, the use of data mining tools raises no new ethical issues unlike the latter case of human bioinformatics. Data mining can be used for other morally problematic activities, where the victims are firms, or states, or perhaps animals, but personal data and consent are the core moral problem and the focus of the models introduced in this chapter.

However, the powerful techniques used in data mining can find new meaning by linking otherwise trivial transaction data, putting great