Chapter 1
Ethical Issues of 'Morality Mining':
Moral Identity as a Focus of Data Mining

Markus Christen
University of Zürich, Switzerland & University of Notre Dame, USA

Mark Alfano
Princeton University, USA & University of Oregon, USA

Endre Bangerter
Bern University of Applied Sciences, Switzerland

Daniel Lapsley
University of Notre Dame, USA

ABSTRACT

When data mining aims to disclose information about the moral competences and values of individuals or groups – an undertaking we call 'morality mining' –, novel ethical problems emerge. These are only partially covered by the current debate on ethical data mining focusing on privacy with respect to discrimination, threats to autonomy, misuse of data, and the consequences of erroneous information. An ethics of morality mining is of particular relevance for research in social science and psychology that increasingly relies on data emerging from social networks, media portals, etc., where people act from or at least in accordance with their own values. In this conceptual contribution, we outline the basic idea of morality mining, explain why we believe that morality mining is associated with novel ethical problems, and suggest ways to address these problems that could potentially help to resolve various socio-economic problems a society or community faces.

INTRODUCTION

Today, countless processes in the social and economic lives of individuals rely on the use of technological systems (Internet, cell phones, GPS, credit card pay systems, etc.) that generate large amounts of data. These data are increasingly the target of data mining (Fayyad et al., 1996; Nisbett et al., 2009; Witten et al., 2011) that intends to uncover "hidden patterns" in large data sets. The results of data mining can be used in various practical ways, e.g., for market segmentation, customer profiling, recommender systems, credit rating, and fraud detection. Furthermore, they are increasingly used for understanding socio-economic developments. A well-elaborated ethical debate on privacy violations through data mining deals with issues such as discrimination, threats to autonomy, misuse of data, and the consequences of erroneous information (Custers et al. 2013, Vaidya et al., 2006; Zarksky, 2003). These issues are, however, not the focus of our contribution.

Our main interest is in an undertaking we call 'morality mining', which occurs when data mining aims to disclose information on the moral competences and values of individuals or groups. We claim that in this case novel ethical issues emerge, resulting from the fact that moral beliefs and convictions are central components of people's identities (Narvaez & Lapsley, 2009). A person with a moral identity constructs his or her self around moral categories, beliefs, and convictions that are chronically accessible for interpreting the interpersonal landscape. Presumably, the sensibilities and preferences of moral people would be on display in their interactions in virtual or Web-based environments, so social scientists and those with political or economic interests have strong incentives to collect data on the moral foundation of the behavior of individuals. The availability of "Big Data" (vast archives of digital text, speech, and video, along with new analysis technology and inexpensive computation) (Williford & Henry, 2012) for psychologists, sociologists, and ethicists

working on moral issues offers novel opportunities to investigate the connection between the moral identity of a person and their behavior. For example, one may investigate mismatches between digital reputation in networks designed for professionals (like LinkedIn) and failures to comply with (moral) norms in work environments; one may identify basic ethical orientations (Narvaez, 2008) that have a predictive value for people's behavior; or one may track people's potentials for moral hypocrisy (e.g., cheating behavior) based on their behavior in social networks. As political opinions are related to specific "moral worldviews" (Haidt, 2012), this knowledge may become a tool for understanding people's political behavior (e.g., in political science), but might also open up new avenues for manipulation according to hidden agendas (e.g., by providing biased information on personalized news portals that are tailored to the moral psychology of the reader).

Big science projects such as the EU flagship proposal FuturICT (www.futurICT.eu) aim to perform large-scale social data mining to forecast socio-economic crises. As "the lack of data, the lack of computational power, and the lack of computationally tested institutional designs" (Helbing & Balietti, 2011, p. 4) have been identified as major obstacles to scientifically addressing socio-economic crises, it seems plausible that governments, corporations, and social scientists will end up collecting information on both the moral values people hold (in order to identify desired states of individuals, groups, or societies) and the moral competences they have (in order to evaluate whether desired system states are achievable). Thus, seen from this very general point of view, it is probable that moral values and competences will become a focus of data mining activities; in particular, if they refer to the understanding of socio-economic development.

The increasing interest in morality mining raises novel ethical issues that go beyond the current discussion of the ethics of data mining. We mention just a few examples: A first issue

Related Content

Estimating the Number of Clusters in High-Dimensional Large Datasets
Xutong Zhuand Lingli Li (2023). *International Journal of Data Warehousing and Mining (pp. 1-14).*
www.irma-international.org/article/estimating-the-number-of-clusters-in-high-dimensional-large-datasets/316142

Forecast Analysis for Sales in Large-Scale Retail Trade
Mirco Nanniand Laura Spinsanti (2010). *Data Mining in Public and Private Sectors: Organizational and Government Applications (pp. 219-244).*
www.irma-international.org/chapter/forecast-analysis-sales-large-scale/44291

A Dynamic and Semantically-Aware Technique for Document Clustering in Biomedical Literature
Min Song, Xiaohua Hu, Illhoi Yooand Eric Koppel (2011). *Integrations of Data Warehousing, Data Mining and Database Technologies: Innovative Approaches (pp. 307-319).*
www.irma-international.org/chapter/dynamic-semantically-aware-technique-document/53081

Automated Integration of Heterogeneous Data Warehouse Schemas
Marko Banek, Boris Vrdoljak, A Min Tjoaand Zoran Skocir (2008). *International Journal of Data Warehousing and Mining (pp. 1-21).*
www.irma-international.org/article/automated-integration-heterogeneous-data-warehouse/1815

Visual Mobility Analysis using T-Warehouse
A. Raffaetà, L. Leonardi, G. Marketos, G. Andrienko, N. Andrienko, E. Frentzos, N. Giatrakos, S. Orlando, N. Pelekis, A. Roncatoand C. Silvestri (2011). *International Journal of Data Warehousing and Mining (pp. 1-23).*
www.irma-international.org/article/visual-mobility-analysis-using-warehouse/49638