

Chapter 2

Ethical Data Mining and Social Science Data Exploration and Description: Scope and Limitations in Social Science Research

Arabi U.
Mangalore University, India

ABSTRACT

As data mining is the process of discovering significant, valuable, and interesting relationships in large and complex volumes of data (especially in data-enriched areas of socio-economic domains and in this socio-economic aspect of a society), data mining applications essentially act as effective instruments for providing support for measuring socio-economic pattern in a society. Although social and ethical matters are nowadays concerns to the society of which people are the only elements, in the days of technology innovations, computers are being manipulated with programs to act more like people, and eventually several social and ethical matters come into focus related to computer programming, or artificial intelligence. Researchers from nearly every social science discipline have found themselves in the position of simultaneously evaluating many questions, testing many hypotheses, or comparing many point estimates. In program evaluation, this arises, for instance, when comparing the impact of several different policy interventions; comparing the status of social indicators like test scores, poverty rates, teen pregnancy rates etc. across multiple schools, states, or countries; examining whether treatment effects vary meaningfully across different sub groups of the population; or examining the impact of a program on many different outcomes. Hence, the relevance of positioning of this chapter in a book of ethical data mining applications for socio-economic development of a community, society, or country fits well as the ethical data mining in social science research is crucial as such data information is highly useful in testing many of the hypotheses of economic or socio-economic in nature.

DOI: 10.4018/978-1-4666-4078-8.ch002

INTRODUCTION

The recording, accessing, data mining and dissemination of ethical data information affect in a crucial way the progress of knowledge of mankind. Today a huge amount of data easily accessible of social relevance in electronic form is produced generally through both research and human activity. In the past, collecting data of human activity has been largely obstructed by financial, technological and ethical issues. As a consequence of the difficulties faced in the past in modeling social and economic systems, the mankind in the present century facing many of the grand challenges either of socio-economic nature or involves human factors in a substantial way. Despite the substantial contributions that complexity social science that has already made in this connection, many of its implications have remained largely theoretical so far. This is basically due to the lack of data, the lack of computational power, and the lack of computationally tested institutional designs in the past. In future the task of identification of the success factors of societies and better way of addressing the grand challenges of humanity by the social science researcher could be fulfilled through a variety of means like: (a) by developing strategies to quickly increase the objective knowledge about social and economic systems (b) describing requirements for efficient large-scale scientific data mining of anonymized social and economic data, (c) formulating strategies how to collect stylized facts extracted from large data set, (d) sketching ways how to successfully build up centers for computational social science, (e) proposing plans how to create centers for risk analysis and crisis forecasting and finally, (f) elaborate ethical standards regarding the storage, processing, evaluation, and publication of social and economic data. Thus, currently, the conduct of ethical research is becoming an important element of modern social science research and the computer science and its technological remnants, especially web bases internet technologies touch

nearly every aspect of human life. This call for a social science researcher to conduct and report an overview of their ethical research issues of current relevance of socio-economic development.

As the ethical issues need to be taken even more seriously the need for an in depth research on data mining must be increased to reach the goals of collective benefits through increased opportunities of data mining process. Although the process of recording, accessing, data mining and dissemination of the information affect the progress of knowledge of mankind in a crucial way, a careful examination of the constraints and the limitations in such exercises needed now. In this context, thanks to the developments in information technology, today a huge amount of data easily accessible in electronic form which is produced more generally by both research and human activity. Further, as the social world is increasingly becoming self-documenting and self-archiving, under this environment of using digital traces as resource, researchers who are acquiring record of computer-mediated social interaction require a variety of skills and tools. Some of them might not be familiar to their ability to write scripts to 'scrape' data directly from an online source, and to parse the results into an analytically useful form. Generally data are stored for manipulation and analysis in a relational database. A clear and accessible way to overview of these matters includes relatively straight forward visualisations based on descriptive statistics and 'small multiples'. While a strategy of this kind results in a 'satellite view' of the data, the approach adopted lends itself to integration with more intensive, qualitative analyses. But, much of the social sciences ignore the extent to which they are all embedded in relational contexts; yet, social networks are fundamental to social life, whether online or offline.

Network analysis is considered to be a thriving area of research and the Internet is inherently relational. Therefore network analysts were drawn to the study of online phenomena such as e-mail, web linkages, and social networking sites. Ob-

16 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/ethical-data-mining-social-science/76255

Related Content

Rule-Based Data Mining Cache Replacement Strategy

Ramzi A. Haratyand Joe Zeitouny (2013). *International Journal of Data Warehousing and Mining* (pp. 56-69).

www.irma-international.org/article/rule-based-data-mining-cache/75615

Pattern Mining and Clustering on Image Databases

Marinette Bouet, Pierre Gançarski, Marie-Aude Aufaureand Omar Boussaïd (2008). *Successes and New Directions in Data Mining* (pp. 187-212).

www.irma-international.org/chapter/pattern-mining-clustering-image-databases/29960

A New Similarity Metric for Sequential Data

Pradeep Kumar, Bapi S. Rajuand P. Radha Krishna (2010). *International Journal of Data Warehousing and Mining* (pp. 16-32).

www.irma-international.org/article/new-similarity-metric-sequential-data/46941

Statistical Entropy Measures in C4.5 Trees

Aldo Ramirez Arellano, Juan Bory-Reyesand Luis Manuel Hernandez-Simon (2018). *International Journal of Data Warehousing and Mining* (pp. 1-14).

www.irma-international.org/article/statistical-entropy-measures-in-c45-trees/198971

Understanding the SNN Input Parameters and How They Affect the Clustering Results

Guilherme Moreira, Maribel Yasmina Santos, João Moura Piresand João Galvão (2015). *International Journal of Data Warehousing and Mining* (pp. 26-48).

www.irma-international.org/article/understanding-the-snn-input-parameters-and-how-they-affect-the-clustering-results/129523