Chapter 33
Algorithms for Selecting the Optimum Dataset While Providing Personalized Privacy and Compensation to its Participants

Rajeev Kumar
Department of Business Administration, College of Business, Kutztown University, Kutztown, PA, USA

ABSTRACT

The privacy preserving microdata sharing literature has proposed several techniques that allow a database administrator to share a dataset in a privacy preserving manner. This paper considers the implications of adding a market layer to that setting. In this setting, individuals (data providers) can receive a market-determined compensation in exchange for their information while they also receive a personalized privacy protection. The computational burdens of satisfying a variety of privacy requirements of individuals (sellers) and dataset requirements of the data receiver (buyer) are analyzed in this paper. The author presents a polynomial time reformulation procedure that proves that the "optimum information product" creation problem reduces to multiple-choice knapsack problem, which is a weakly NP hard problem. The problem of various instance sizes is solved using FICO Xpress 7.0 optimization software. The insights presented in the paper can be utilized for creating a market of individual information in different settings.

INTRODUCTION

Privacy laws provide a baseline level of "privacy protection" and due diligence for only some types of individual level information that are being utilized by government agencies and private businesses. For example, there are laws in the US such as the Fair Credit Reporting Act of 1970, the Family Education

DOI: 10.4018/978-1-7998-8954-0.ch033

Rights and Privacy Act of 1978, the Privacy Protection Act of 1980, the Cable Communications Policy Act of 1984, and the Video Privacy Protection Act of 1988 that aim to protect individual information about communications between people, credit data, education, cable, and retail video industries. However, privacy laws do not exist for many other types of individual information such as driving records, rental histories, retail purchases, social security earnings, unlisted phone numbers, etc. Many scholars in history, sociology, business, law, and political science have criticized the existing legislative approach to individual privacy on the grounds that it does not acknowledge individuals as the legitimate owner of their data (Blackwell, 2008; Gavison, 1980; Laudon, 1996; Paul 2004). Some scholars have advocated for a stronger definition of privacy protection for individuals, which is called the "property rights" approach. In this approach, personal information of individuals is considered their private property and can only be exchanged by creating a private contract between an individual (seller) and a data receiver (buyer).

Chellappa and Shivendu (2007) discuss the property rights approach to privacy for regulating choices in the online personalization context. Xu et al. (2010), in the context of a location-based service, show that individual compensation is likely to increase consumers' judgments of the benefits of information disclosure. Garfinkel et al. (2006) consider a market setting where individuals have the option of selling inexact sensitive information about themselves while receiving a level of privacy protection. Buyers in this setting can only purchase inexact answers to the queries of their choices. For instance, a buyer can demand the average salary of a certain demographic of her choice and receive an inexact interval answer for the query. Li and Raghunathan (2014) consider a setting where an organization sells a dataset about its individuals' sensitive information to another organization. In this setting, the organization sharing the data takes the sole responsibility of providing privacy protection to individuals and selecting a pricing model for the market. Individuals do not play any role in the pricing of their data and the privacy protection mechanism of the market.

This paper presents a setting where individuals (data providers), whose data is being considered for sharing/selling, can themselves decide the level of privacy protection in terms of the level of anonymity that they want to receive from the buyer (data receiver). Specifically, sellers can choose a personalized version of privacy protection (anonymity), as discussed in the privacy preserving microdata sharing literature (Machanavajjhala et al., 2007; Sweeney, 2002). Moreover, sellers can also specify the minimum compensation requirement (reservation price) for their information, which allows them to be participants in the price discovery process of the market. Buyers, on the other hand, can buy individual level information, not just an answer to a query. In this market context, this paper discusses the computational complexities of the problems associated with the creation of the optimal information product (optimum dataset) for a buyer's dataset request. The algorithms presented in this paper can be used by the market maker for effectively creating the optimum datasets for buyers.

LITERATURE REVEW

From the perspective of individual privacy, personal information can be broadly categorized into two categories: (1) nonsensitive information, and (2) sensitive information. Attributes such as zip code of residence, gender, etc., whose values about an individual are typically publicly known, are considered nonsensitive. Attributes such as disease, etc., are considered sensitive, as individuals generally do not want the values of these attributes to be publicly known. A vector of individual level information over a set of attributes is called a microdata vector (MDV) in this paper. For example, Zip Code = 18349,