Chapter 7.37
Does Protecting Databases Using Perturbation Techniques Impact Knowledge Discovery?

Rick L. Wilson
Oklahoma State University, USA

Peter A. Rosen
University of Evansville, USA

ABSTRACT

Data perturbation is a data security technique that adds noise in the form of random numbers to numerical database attributes with the goal of maintaining individual record confidentiality. Generalized Additive Data Perturbation (GADP) methods are a family of techniques that preserve key summary information about the data while satisfying security requirements of the database administrator. However, effectiveness of these techniques has only been studied using simple aggregate measures (averages, etc.) found in the database. To compete in today's business environment, it is critical that organizations utilize data mining approaches to discover information about themselves potentially hidden in their databases. Thus, database administrators are faced with competing objectives: protection of confidential data versus disclosure for data mining applications. This chapter empirically explores whether data protection provided by perturbation techniques adds a so-called Data Mining Bias to the database. While the results of the original study found limited support for this idea, stronger support for the existence of this bias was found in a follow-up study on a larger more realistic-sized database.

INTRODUCTION

Today, massive amounts of data are collected by organizations about customers, competitors, supply chain partners, and internal processes. Organizations struggle to take full advantage of this data, and discovering unknown bits of knowledge in their massive data stores remains a highly sought after goal.

Database and data security administrators face a problematic balancing act regarding access to organizational data. Sophisticated organizations benefit greatly by taking advantage of their large databases of individual records, discovering previously unknown relationships through the use of data mining tools, and knowledge discovery algorithms (e.g., inductive learning algorithms, neural networks, etc.).

However, the need to protect confidential data elements from improper disclosure is another important issue faced by the database administrator. This protection concerns not only traditional data access issues (i.e., hackers and illegal entry) but also the more problematic task of protecting confidential record attributes from unauthorized internal users.

Techniques that seek to accomplish masking of individual confidential data elements while maintaining underlying aggregate relationships of the database are called **data perturbation techniques**. These techniques modify actual data values to hide specific confidential individual record information.

Recent research has analyzed increasingly sophisticated data perturbation techniques on two dimensions: the ability to protect confidential data and, at the same time, the ability to preserve simple statistical relationships in a database (means, variances, etc.). However, value-adding knowledge discovery and data mining techniques find relationships that are much more complex than simple averages (such as creating a decision tree for classifying customers, etc.). To our knowledge, only one previous study (Wilson & Rosen, 2003) has explored the impact of data perturbation techniques on the performance of knowledge discovery techniques. The present study expands on this initial study, better quantifying possible knowledge losses or the so-called Data Mining bias.

REVIEW OF RELEVANT LITERATURE

Data Protection through Perturbation Techniques

Organizations store large amounts of data, and most may be considered confidential. Thus, security and protection of the data is a concern. This concern applies not just to those who are trying to access the data illegally but to those who should have legitimate access to the data.

Our interest in this area relates to restricting access of confidential database attributes to legitimate organizational users (i.e., data protection). **Data perturbation techniques** are statistically based methods that seek to protect confidential numerical data by adding random noise to the original data elements. Note that these techniques are not encryption techniques, where the data is first modified, then (typically) transmitted, and, on receipt, reverted back to the original form.

The intent of data perturbation techniques is to allow legitimate users the ability to access important aggregate statistics (such as mean, correlations, etc.) from the entire database while protecting the identity of each individual record. For instance, in a perturbed database on sales figures, a legitimate system user may not be able to access original data on individual purchase behavior, but the same user could determine the average of all individual purchasers.

Data perturbation methods can be analyzed using various **bias** measures (see Muralidhar, Parsa & Sarathy, 1999). A data perturbation method exhibits bias when the results of a database query on perturbed (i.e., protected) data produces a significantly different result than the same query executed on the original data. Four types of biases have previously been identified, termed Type A, Type B, Type C, and Type D.