Privacy Preservation and Data Mining

by XXXXXXXXX
10/27/2014

Continued growth in the amount of information that databases need to store and process presents an ever increasing problem of how to protect the privacy of personal data while still being able to utilize the information collected through various data mining techniques. Over the years there have been several approaches to solving this important problem. Most approaches have involved hiding or distorting the information in the database so that when the data are mined the distortion masks individual results while still maintaining accuracy in returned statistical results. The goal is to accomplish this while still preventing an adversary from being able to reverse engineer the hiding process and make discoveries about an individual's protected data.

Agrawal and Srikant [1] approached this problem by trying to determine if the current data mining technique of decision-tree classification can adequately pull data from a database while protecting privacy. They looked only at numerical attributes and tried to determine an effective amount of information that could be distorted in a data set without compromising the original information. They explored two methods of data hiding; Value-Class Membership, where the values of an attribute are divided into specific intervals, and Value Distortion using either a Uniform or a Gaussian distribution to distort the original data. Taking the data prepared by these methods, they then looked at three algorithms in an attempt to reconstruct the original data. The algorithms (Global, ByClass, and Local) differ in how and when reconstruction takes place.[1] The authors compared the ratio of the confidence level of their reconstructed data with the amount of privacy that was preserved in the reconstruction process. The result of their experiment was that the ByClass and Local algorithms performed equally well and both algorithms performed much better than the Global algorithm. Their process of first distorting then reconstructing the data allows for search engines and other data mining software to be developed using these algorithms, providing valuable information to those seeking it while still protecting the privacy of the individual.

---

[1] The three algorithms are defined as follows:

Global: Reconstruct the distribution for each attribute once at the beginning using the complete perturbed training data. Induce decision tree using the reconstructed data.

ByClass: For each attribute, first split the training data by class, then reconstruct the distributions separately for each class. Induce decision tree using the reconstructed data.

Local: As in ByClass, for each attribute, split the training data by class and reconstruct distributions separately for each class. However, instead of doing reconstruction only once, reconstruction is done at each node. To avoid over-fitting, reconstruction is stopped after the number of records belonging to a node becomes small.[1, p445]

Work done by Agrawal and Aggarwal [2] is a continuation of the research done in Agrawal and Srikant [1]. To compare the proficiency of different algorithms, the authors first develop a way to compare the quantification of privacy and information loss in a manner that is independent of any particular type of data mining. This technique allows for a fair comparison in varying methods. The authors then developed a new algorithm, called the Expectation Maximization (EM) algorithm, which reconstructs the distorted data with little cost to both accuracy and privacy. They are then able to compare the EM algorithm with what they term the AS algorithm[2] used in [1] with this new quantification method. The results were objectively equivalent except in extreme situations where the data were particularly distorted. This was due to the fact that the EM algorithm converges more accurately with the original data than does the AS algorithm. The end result is that for "very large data sets, the EM reconstruction algorithm can provide very high privacy guarantees for almost no information loss." [2, p254]

Evfimievski et. al. [3] add an additional component to Agrawal and Srikant [1] and Agrawal and Aggarwal [2] by exploring both sensitive numerical data and categorical data. In addition to looking at the use of randomization in general to protect privacy, they also looked at how to protect privacy while mining association rules from the database. First the authors looked at existing randomization methods and showed that despite their current effectiveness there was still information that leaked through the protective measures. In choosing the data to be randomized, this team focused on randomization for individual transactions. This eliminated concerns with security risks in switching data among users to mask the sensitive data. They then took the data that slipped through and calculated the support for the discoverable data in relation to the association rules. Combining both the randomization algorithms and the support calculations for association rules they were able to develop new mining algorithms that were more effective in preserving privacy while executing the data mining of association rules.

Aggarwal et. al. [4] address the issue of data security from the perspective of finding the minimal amount of data that needs to be hidden so that adversarial data mining cannot be used to derive the protected data. Past research, such as Agrawal and Srikant [1] and Agrawal and Aggarwal [2], focused on data distortion and ignored the possibility of data being reconstructed from association rules. Evfimievski et.al. [3] looked at association rules and uses them to modify existing randomization algorithms. The goal of Aggarwal et. al. [4] is to find an approach that can predict which fields need to be hidden so that the already secure fields cannot be predicted by the remaining data. Rules that predict sensitive data using certain fields to derive them are called adversarial data mining rules. To prevent adversarial data mining Aggarwal et. al. [4] first determined all adversarial rules that could affect the protected fields.

[2] AS stands for the last names of the authors, Agrawal and Srikant in [1].

They then used this information to derive which additional fields need to be hidden. By protecting additional fields association rules can no longer be used to discover the protected data.

Aggarwal et. al. [4] show there are several ways in which adversarial rules can be determined. Using a depth-first search of an enumeration tree the authors mine all possible association rules. They then use several pruning methods to eliminate all rules that are not related to the sensitive data fields. For example, if the data that must be protected do not appear, the association rule can be eliminated. If rules do influence the protected data field, their support needs to be determined to see what influence they might have on data discovery. If support is low then the rule may not be statistically noteworthy enough to be included. The authors developed an algorithm called FAiR (Finding Adversarial Rules) to accomplish this. The last step hides different data fields, checks the relevant adversarial rules that affects these fields, and determines if the rule can be used to unmask that field. For example, will hiding particular data elements cause the rule to be invalidated or marginalized? For this purpose they created the GRaDeS (Generating Derived Set) algorithm, which weights the contribution of the blanked out field. They then determined the best solution that balances maintaining data accuracy and efficiency in protecting the sensitive data. They concluded that this approach is practical from both a processing point of view and the fact that information-loss is minimal.

There seems to be a general change in researching privacy protection. It started with the use of data distortion algorithms to mask the data while still being able to obtain accurate statistical information during data mining processes. The research then changes to the use of association rules to discover weaknesses in these distortion algorithms. Most recently researchers are using these adversarial rules directly to determine the best data fields to hide for the most efficient way to preserve both the integrity of the data and personal privacy. There is also a change in the approach to these problems. Agrawal and Srikant [1] and Agrawal and Aggarwal [2] looks more at what can be done to data mining techniques simply to protect privacy, where beginning with Evfimievski et. al. [3] and more unequivocally in Aggarwal et. al. [4] the approach is more from an “adversarial” point of view. In other words, what would a malicious attacker do to discover the protected data? There have been some significant improvements in privacy preservation over the years. As one can see, many of these methods still contain weaknesses that can be exploited. Adversarial research should be pursued more closely. In real life situations, when protecting immense amounts of data, it is usually the innovation of attackers that helps data analysts discover data system flaws.

References

[1] R. Agrawal and R. Srikant, "Privacy-Preserving Data Mining," in *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, Dallas, TX, 2000, pp. 439-450.

[2] D. Agrawal and C. C. Aggarwal, "On the Design and Quantification of Privacy Preserving Data Mining Algorithms," in *Proceedings of the Twentieth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems* , Santa Barbara, CA, 2001, pp. 247-255.

[3] A. Evfimievski, R. Srikant, R. Agrawal, and J. Gehrke, "Privacy Preserving Mining of Association Rules," in *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* , Edmonton, Alberta, Canada, 2002, pp. 217-228.

[4] A. C. Aggarwal, J. Pei, and B. Zhang, "On Privacy Preservation Against Adversarial Data Mining," in *Proceedings of the 12th ACM SIGKDD International Conference On Knowledge Discovery and Data Mining*, Philadelphia, PA, 2006, pp. 510-516.