Tatra Mt. Math. Publ. **45** (2010), 151–159
DOI: 10.2478/v10127-010-0011-z

TATRA MOUNTAINS
Mathematical Publications

DATA MINING AS A TOOL IN PRIVACY-PRESERVING DATA PUBLISHING

Michal Sramka*

ABSTRACT. Many databases contain data about individuals that are valuable for research, marketing, and decision making. Sharing or publishing data about individuals is however prone to privacy attacks, breaches, and disclosures. The concern here is about individuals' privacy—keeping the sensitive information about individuals private to them. Data mining in this setting has been shown to be a powerful tool to breach privacy and make disclosures. In contrast, data mining can be also used in practice to aid data owners in their decision on how to share and publish their databases. We present and discuss the role and uses of data mining in these scenarios and also briefly discuss other approaches to private data analysis.

1. Introduction

There is abundance of data collection about individuals. One of the main reasons for such collection is to use these data to create new useful knowledge. With computers, it became easier to do even complex analysis of data, creating knowledge that is useful for many possible uses. But there is always the possibility of misusing personal or sensitive data. Data collectors need to respect these privacy concerns when sharing, publishing or otherwise releasing the data.

2010 Mathematics Subject Classification: 62–07, 68P99, 97R50.
Keywords: private data analysis, privacy-preserving data publishing, data mining, privacy attacks.
Supported by the grant NIL-I-004 from Iceland, Liechtenstein and Norway through the EEA Financial Mechanism and the Norwegian Financial Mechanism and partly supported by the Spanish Government through projects TSI2007-65406-C03-01 "E-AEGIS" and CONSOLIDER INGENIO 2010 CSD2007-00004 "ARES", by the Government of Catalonia under grant 2009 SGR 01135, and by the Government of Alberta, Canada under "iCore"—Informatics Circle of Research Excellence, now a part of ALBERTA INNOVATES—Technology Futures.
* The author is with the UNESCO Chair in Data Privacy, but he is solely responsible for the views expressed in this paper, which do not necessarily reflect the position of UNESCO nor commit that organization.

1.1. Private data analysis

Data is provided to a data collection by individuals, the custodian of the data collection then becomes the *data owner*. Data analysis can be performed by the data owner or the data owner can outsource the data analysis to other parties. In any case, the privacy concerns of the involved individuals should be addressed and considered at all times. The question of how to obtain valid results and knowledge without learning the underlying private data is studied in *private data analysis*, also referred to as *privacy-preserving data mining* [2], [4]. Private data analysis is achievable in the following ways:

- **Private data analysis over original data**. In this scenario, computations are performed over the original private or even confidential data.
  - Data analysis is performed by the data owner. No other party will learn the data, and the results of the analysis will stay "in house".
  - Data mining is performed over the original data and then the obtained knowledge is published. The published knowledge is protected against privacy leaks in a way that it does not reveal sensitive information about the underlying data. This is achievable by sanitizing the learned knowledge and referred to as the *privacy-preserving knowledge publishing* [7], [3].
  - One or several parties own confidential data and another party performs a computation over them. *Secure multiparty computation* [15], [8] and secure multiparty computation over distributed data sets are fields that study cryptographic tools that allow to compute a function over confidential data without learning anything else than what can be learned from the output of the function.
- **Data analysis over sanitized data**. In this scenario, data is sanitized and then shared or published for analysis. This is referred to as *privacy--preserving data publishing (PPDP)* [1], [6]. Sanitization is usually achieved as a transformation of the data that provides pseudonymity, anonymity, or privacy risk reduction by generalizing, masking, randomizing, or even suppressing some data. The rest of this paper deals exclusively with this scenario.

2. Privacy-preserving data publishing using data mining

We present a framework that allows the owner of a data collection to decide which is the best way to sanitize and publish his/her collection. In the framework, both the disclosure risk and information loss measures are based on data mining concepts, namely data mining utility.

2.1. Sanitization

The collected *original data* DB is transformed into the *sanitized data* $DB^\star$ that can be shared and published. The transformation, called *sanitization*, ensures that privacy of individuals is not compromised and that the data is useful for analytical purposes. Privacy is usually measured using some form of disclosure risk, while the data utility is traditionally measured as information loss between the original data and the transformed sanitized data.

Many techniques and methods for transforming various types of data have been proposed. Here we concentrate mainly on privacy-preserving publishing of *microdata*, i.e., relational and transactional databases. The methods for sanitizing microdata [6] can be divided into two groups: anonymization methods and perturbation methods. Anonymization methods are based on the "safety in number" idea. *Anonymization* generalizes or suppresses information in order to achieve similarities in the data. Individuals are clustered into similarity groups, and each individual is guaranteed anonymity in a group. *Perturbation* methods distort data values, e.g., by adding noise to them, in order to mask the relations between sensitive information and individuals.

2.2. Data mining in PPDP

Data mining (also knowledge discovery or knowledge learning) refers to nontrivial extraction of implicit, unknown, and potentially useful information from data [5]. A *data miner* extracts trends or patterns from data and carefully and accurately transforms them into useful and understandable information. Such information (knowledge), is uncovered through the use of artificial intelligence techniques and is more complex than what is typically retrievable by standard techniques, such as statistics. We model the *data miner* as an algorithm $\mathcal{M}$,

$$x'_i \leftarrow \mathcal{M}(DB, x, i),$$

which takes as an input a (possibly sanitized) database DB, a record x having the same schema as DB, and a field indicator i. Based on the knowledge extracted from DB, it outputs a prediction x'_i for the ith field of x, a value which may have been unknown or masked in x. This model captures various data mining tasks, such as value and trend prediction, clustering, classification, etc. Traditionally, data mining has been used to measure the usefulness of the published sanitized data. Data mining can capture numerous analytical tasks and grade the usefulness of the published sanitized data by measuring and comparing accuracy of prediction over the original and over the sanitized data.

2.3. Data mining utility for PPDP

At first glance, data mining and privacy seem to represent opposing goals, with mining trying to bring all kinds of knowledge "into the open" and privacy being interested in keeping such knowledge "in the dark". But at a closer look, we can

identify some key conditions with regard to both areas that can make the goals of them compatible. The *analysts*—the legitimate users of the published sanitized data want to create knowledge useful for them, which means that the data mining utility of the newly obtained knowledge should be high. On the other side, individuals whose data is included in data collections usually are only concerned with the privacy of some of their data. And so the knowledge obtained by privacy *adversaries* (also users of the published sanitized data albeit with different goals, i.e., malicious intentions) must have low utility. If their utility is not low, there are existing privacy threats and a high risk of disclosures. In this sense, we extended the use of data mining and data mining utility from measuring usefulness (data mining utility for analysts) to also measure privacy [12], [11] (data mining utility for adversaries). Briefly, we proposed to measure the *data mining utility* as

$$\mathcal{U}(DB, \mathcal{M}, i) = \sum_{x \in DB} w(x) \cdot E_i\big(\mathcal{M}(DB, x, i), \ldots\big), \tag{1}$$

where $w(x)$ is a weight function which rates the interest of an analyst or adversary in a record x, and where $E_i(x'_i, \ldots)$ is a function which weights the correctness of the prediction x'_i from either the true value $\bar{x}_i$ or also from the sanitized value $x^\star_i$, depending on the scenario.

A *sanitization is useful* for the field (attribute) i and miner $\mathcal{M}$ if the utility measured over the sanitized data $DB^\star$ is not significantly lower than the utility measured over the original data DB with the same interest weights w and correctness function E_i. The field i, miner $\mathcal{M}$, weights w, and correctness E_i all define a targeted analytical task and so model the analysts and their interests. The reasoning behind this is that the data owner knows for what analytical *purpose* is the data being outsourced. The functions and parameters allow fine-grained modeling, so specific analytical tasks can be captured, e.g., mining over medical data (specialized miner $\mathcal{M}$), predicting whether a patient should be tested for diabetes (field i), concentrating mainly on women over 55 and men over 50 (interest weights w higher for these patients), and possibly doing more tests rather than missing somebody (correctness E_i allowing some false-positives but penalizing true-negatives).

Similarly, a *sanitization preserves privacy* for the field j using the miner $\mathcal{M}$ if the utility in the equation (1) is low, with the interest weight function w representing the adversary's interest in records (in individuals whose privacy is being attacked) and the correctness function E_j measuring the success of predicting the original value from sanitized value. Here, the miner $\mathcal{M}$, field j, weights w, and correctness E_j allow to model a real-world privacy adversaries and their intentions, e.g., the miner $\mathcal{M}$ is a privacy attack that discloses a sensitive information about the field j, say age, the weights w represent the interest of the adversary in a selected few famous people, and the correctness E_j captures the adversary's interest in exact or approximate age (exact or partial disclosure).

We note that data mining demonstrates that it is impossible to preserve privacy and usefulness for the same field. Therefore in practice, these fields must be distinct, e.g., field i for usefulness can be prediction of buying power, while another field j for privacy preservation can be age, which the data owner tries to keep private for the individuals.

2.4. Our framework for PPDP

Quantifying the trade-off between usefulness and privacy of published data has been the subject of much research in recent years. The above definition of utility for data mining is a universal measure that can quantify both privacy and usefulness of sanitized data, acting as both disclosure risk and information loss metrics. The utility measure is flexible enough to model the needs of legitimate analysts as well as adversaries' intentions captured as their gain in attacking published data. Our framework allows these differences to be systematically taken into account by modeling analysts and adversaries and using multiple data miners.

Based on the utility definition, we propose a pragmatic framework for evaluating sanitization systems in real-life [12], [11]. Our approach to evaluate sanitization methods is practical and provides a decision support mechanism for data owners who are faced with the question of which sanitization method is more suitable for their "purpose". Using the framework, data owners are able to estimate the advantages and risks involved with each choice.

Using data mining as the primary tool of the adversaries and analysts is a good approximation of how sanitized data is used by the legitimate analysts as well as how it is compromised by the privacy adversaries. Data mining adversaries capture a large class of automated attacks on published data and hence our framework provides a baseline evaluation method for evaluating privacy guarantees afforded by sanitization methods. By expanding the set of data miners that are used for evaluation one may consider a wider class of attacks and thus a privacy guarantee against stronger attackers. In fact, other utility and information loss measures, many existing privacy attacks, and other privacy and disclosure risks measures can be modeled in our framework as data miners.

3. Privacy attacks using data mining

Data mining can be used to evaluate privacy, just by trying to see what knowledge can an adversary obtain from the published sanitized data. For privacy reasons, the data mining utility of this knowledge must be low. On the contrary, if the utility is high, the obtained knowledge can be exploited by adversaries to breach privacy and make sensitive disclosures from the published

data. An attack against statistical disclosure control that looks for private information in different versions of the same data using clustering techniques has been published in [14]. We on the other hand concentrate on employing data mining for a single sanitized version of the original data.

We have explored the idea of predicting original values from anonymized and perturbed data in [10], [9]. The adversary model assumes that the adversary possesses multiple data miners and no background knowledge about individuals. Using the multiple data miners, the adversary extracts knowledge from the published data and makes multiple predictions that are fused together to estimate data values of the original data. In this sense, the adversary is trying to "desanitize" the published data, effectively pushing the transformed sanitized data $DB^\star$ back toward the original data DB using data mining and fusion techniques. Because of the simplicity of running data mining today, the attack is practical and can be launched even by non-expert adversaries.

3.1. A noise removing attack on perturbed data

The attack can be explained on the following example [9]: A data owner wants to release his/her data collection about individuals. The data owner considers the individuals' ages to be a sensitive information. The owner protects the privacy of the ages by adding random noise to all the ages individually, and then publishes the perturbed data. The perturbed data are then available to legitimate analysts as well as malicious adversaries. The adversary launches the attack by applying one or more data mining algorithms to uncover hidden trends and patterns in the data that would allow him/her to partially remove the added noise and obtain quality estimates of the ages.

The attack consists of first obtaining predictions for the ages that have been perturbed. Data miners can provide such predictions. However, multiple data miners produce multiple predictions. Moreover, there may be different predictions, based on individual records in the database, for the same original value of age. Therefore the adversary combines the different predictions obtained from multiple miners using fusion methods. The adversary's intention is to use all these predictions and combinations to obtain reasonable estimates of the ages––estimates that contain much less noise than the perturbed values and therefore estimates that breach privacy.

The experimental results confirm that this attack is practical and presents a significant privacy risk to published perturbed data. The results show that up to 93 % of the noise added during perturbation can be effectively removed using general-purpose readily-available data miners. Interestingly, the higher the aimed privacy, the higher the percentage of noise can be removed. This suggests that adding more noise does not always increase the real privacy.

3.2. A fusion attack on sanitized data

Similar to the previous scenario, this attack [10] employs multiple data miners and fusion methods to provide predictions about anonymized, perturbed or possibly otherwise sanitized data. The predictions are consequently combined using some fusion method (e.g., average, weighted average, voting) in order to increase the success chance of breaching privacy of individuals. The fusion attack is practical and provides a powerful method of breaching privacy for both anonymized and perturbed data. In summary, the fusion attack:

- makes partial disclosures as well as exact disclosures (predictions of the exact value in the original data purely from the published sanitized data);
- provides an effective way of approximating predictions of the best miner (a miner that provides the best results among all considered miners) even when this miner cannot be determined. And we note that in general the adversary has no means of identifying this miner;
- closely approximates the success of the ideal perfect attacker in the case of perturbed data; and
- as a baseline evaluation, it is better than a simple guessing strategy.

4. Open problems: impact of data mining on privacy

The intuition is that a sanitization method should allow the patterns and trends observed in the original data to be observable in the sanitized data, and so to provide successful prediction. We have demonstrated that it is possible to make useful predictions about the sanitized medical data when rules discovered from the original unsanitized medical data are used [13]. The usefulness is established in comparison with the case where no sanitization takes place. This suggests that prediction rules discovered from unsanitized data can be used to make prediction about sanitized data and hence the considered sanitization methods do generate useful data. But it remains an unanswered question *whether knowledge discovered from unsanitized data have negative impact on privacy of sanitized data.*

From the perspective of an individual whose personal data is included in the data collections, the knowledge that organizations can gain from these collections can have positive or negative effects on the individual. Quantifying what comprises positive or negative, wanted or unwanted, or sensitive information is a personal feeling of an individual, and therefore hard to measure or even define. In the rest of the section, suppose that the knowledge obtained from mining over (sanitized) data can be used to predict some specific information about individuals.

From the perspective of the end user—legal analyst or adversary—a correct prediction is almost always considered useful. After all, that is precisely why they preform the analysis. On the other side, from the viewpoint of the individual, each prediction can be privacy-impacting for the individual, no matter whether it is correct or incorrect. Indeed, even an incorrect prediction can have a negative impact on an individual. For example, due to an incorrect prediction an individual is wrongly classified. The individual may loose his/her social status (e.g., embarrassment, bashing, gossips, marriage) or economic opportunities (e.g., job loss, inaccessibility to new opportunities, loans).

In general, the degree of the impact of a prediction differs, for example based on who is doing the prediction. Consider the following examples:

(1) A hospital predicts that a patient has a disease. The prediction is private but useful to the individual.

(2) An insurance company predicts the same disease for the same person. The prediction is useful to the individual (treatment can start). It is also private, embarrassing, and threatening (insurance premiums can be raised). The prediction may be based on discriminatory information or the prediction itself may lead to discrimination.

(3) A bank predicts a customer as a "high credit risk". The prediction is not useful, but private and embarrassing to the individual. Furthermore, it may be based on discriminatory information or itself may lead to discrimination.

Often, in these examples, it does not matter whether the prediction is based on private data or published data. But in the case of published data, an important question still remains: *Under what circumstances is a prediction private and/or discriminating from the perspective of an individual?*

Acknowledgements. The author would like to thank to all his collaborators without whom this research would not be possible.

REFERENCES

[1] ADAM, N. A.—WORTMAN, J. C.: *Security-control methods for statistical databases*, ACM Comput. Surv. **21** (1989), 515–556.

[2] *Privacy-Preserving Data Mining: Models and Algorithms* (C. C. Aggarwal, P. S. Yu, eds.), Springer, New York, NY, USA, 2008.

[3] ATZORI, M.—BONCHI, F.—GIANNOTTI, F.—PEDRESCHI, D.: *Anonymity preserving pattern discovery*, VLDB J. **17** (2008), 703–727.

[4] DOMINGO-FERRER, J.—SAYGIN, Y.: *Recent progress in database privacy*, Data Knowl. Eng. **68** (2009), 1157–1159.

[5] FAYYAD, U. M.—PIATETSKY-SHAPIRO, G.—SMYTH, P.: *From data mining to knowledge discovery: an overview*, in: Advances in Knowledge Discovery and Data Mining (U. M. Fayyad, et al., eds.), AAAI, Menlo Park, CA, USA, 1996, pp. 1–34.

[6] FUNG, B.—WANG, K.—CHEN, R.—YU, P. S.: *Privacy-preserving data publishing: a survey on recent developments*, ACM Comput. Surv. **42** (2010), 1–53.

[7] KANTARCIOĞLU, M.—JIN, J.—CLIFTON, C.: *When do data mining results violate privacy?* in: Proc. of the 20th ACM SIGKDD Internat. Conf. on Knowledge Discovery and Data Mining—KDD '04, Seattle, WA, USA, 2004, ACM Press, New York, NY, USA 2004, pp. 599–604.

[8] LINDELL, Y.—PINKAS, B.: *Privacy preserving data mining*, J. Cryptology **15** (2002), 177–206.

[9] SRAMKA, M.: *A privacy attack that removes the majority of the noise from perturbed data*, in: Proc. of the Internat. Joint Conf. on Neural Networks—IJCNN '10, Barcelona, Spain, 2010, IEEE Computer Soc., Washington, DC, USA, 2010, pp. 356–363.

[10] SRAMKA, M.—SAFAVI-NAINI, R.—DENZINGER, J.: *An attack on the privacy of sanitized data that fuses the outputs of multiple data miners*, in: Proc. of the 9th IEEE Internat. Conf. on Data Mining Workshops—ICDMW '09: Internat. Workshop on Privacy Aspects of Data Mining—PADM '09, Miami Beach, FL, USA, 2009, IEEE Computer Soc., Washington, DC, USA, 2009, pp. 130–137.

[11] SRAMKA, M.—SAFAVI-NAINI, R.—DENZINGER, J.—ASKARI, M.: *A practice-oriented framework for measuring privacy and utility in data sanitization systems*, Transaction on Data Privacy, 2010 (to appear).

[12] SRAMKA, M.—SAFAVI-NAINI, R.—DENZINGER, J.—ASKARI, M.: *A Practice--oriented framework for measuring privacy and utility in data sanitization systems*, in: Proc. of the 3th Internat. Conf. on Extending Database Technology Workshops––EDBT '10: 3rd Internat. Workshop on Privacy and Anonymity in the Inform. Soc.––PAIS '10, Lausanne, Switzerland, 2010, ACM, New York, NY, USA, 2010, pp. 1–10.

[13] SRAMKA, M.—SAFAVI-NAINI, R.—DENZINGER, J.—ASKARI, M.—GAO, J.: *Utility of knowledge extracted from unsanitized data when applied to sanitized data*, in: Proc. of the 6th Annual Conf. on Privacy, Security and Trust—PST '08, Fredericton, New Brunswick, CA, 2008, IEEE Computer Soc., Washington, DC, USA, 2008, pp. 227–231.

[14] VALLS, A.—TORRA, V.—DOMINGO-FERRER, J.: *Semantic based aggregation for statistical disclosure control*, Internat. J. Intel. Syst. **18** (2003), 939–951.

[15] YAO, A. C. C.: *Protocols for secure computations (extended abstract)*, in: Proc. of the 23rd Annual Symposium on Foundations of Computer Science—FOCS '82, Chicago, IL, USA, 1982, IEEE Computer Soc., Washington, DC, USA, 1982, pp. 160–164.

Received June 26, 2010

Department of Computer Engineering and Mathematics
UNESCO Chair in Data Privacy
Universitat Rovira i Virgili
Av. Paisos Catalans 26
43007 Tarragona
CATALOINA (SPAIN)
E-mail: michal.sramka@urv.cat