Proceedings of International Conference on Mechanical, Electrical and Medical Intelligent System 2017
Invited Paper

A Design of Privacy-Enhanced Survey System That Can Be Used for Hospital Evaluation by Patients

Atsushi Iwai

Faculty of Social and Information Studies, Gunma University, 4-2 Aramaki-mach,
Maebashi City, Gunma Prefecture, 371-8510, Japan
iwai@gunma-u.ac.jp

Keywords: privacy preservation, hospital evaluation, social survey, personal information, anonymity measure

Abstract. This study presents a design of privacy-enhanced survey system that can be used for hospital evaluation by patients. As technical tools for enhancing privacy, k-anonymity and l-diversity are widely known, but such researches aimed at concealing patients' information from the public. In contrast, the aim of this study is to protect patients' information from survey assessors, or medical staff. Similar to class evaluation by students at school, hospital evaluation by patients is needed for improving the quality of medical service. In this type of survey, privacy from medical staff is important to prevent deterioration of the quality of the obtained data. By expanding on survey systems designed in previously presented studies, this study presents a formal description of a new enhanced survey system design.

1. Introduction

As technical tools for enhancing privacy, k-anonymity by Sweeny ([12]) and l-diversity by Machanavajjhala et al. ([10]) have been widely known. These researches were conducted in medical contexts, and aimed at protecting patients' personal information from the general public.

However, these frameworks fail to address the problem of how to protect personal information from the survey assessors, or medical staff themselves. Similar to class evaluation by students at school, hospital evaluation by patients is needed for improving the quality of medical service. In this type of survey, privacy from medical staff is important to prevent deterioration of the quality of the obtained data.

This study presents a design of privacy-enhanced survey system that can be used for hospital evaluation by patients. By expanding on survey systems designed in previously presented studies, a formal description of the new enhanced survey system design is to be demonstrated.

2. Previous Studies

As a typical survey system design presented in previous studies, this study focuses on the design described in [5], [9]. The basic design approach is as exemplified as follows, for the purpose of illustrating problematic issues. For an example, A course evaluation was conducted in a small class of 3 male and 15 female students, with a single question sheet that contained a question concerning gender and other questions about course evaluation. This would be potentially harmful to male student privacy and could result in a deterioration of the quality of the obtained data. However, if the question sheet was divided into two parts, with one part including only the gender question and the other part only the course evaluation questions, then no privacy problem would arise to compromise the quality of the students' answers (Figure 1).

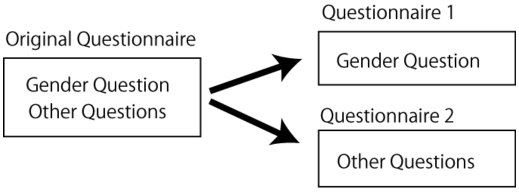

Figure 1 Modification of Questionnaire Design ([9])

The target system processes this division operation after all students have finished responding to the questionnaire, and when it finds problematic questions that can lead to information leakage. The division process is realized as a database operation of modifying the table structure related to the questionnaire design. As the computational process is triggered automatically and perfectly terminates before a lecturer sees the output of the system, no information leakage is possible.

This system design is based on class evaluation context and the hypothesis that all questions on a question sheet can be divided into two categories, X and Y. X is defined as a category of individual attributes, such as gender or age. Y is defined as a category of individual attitudes such as course evaluation. For each Y category question, a cross tabulation of many X category questions is likely to yield special cells where only a small number of respondents exist, and these cells are likely to cause some unintended information leakage.

The prototype system described in [5] and [9] consists of a framework that analyzes the input data to find elements that can lead to information leakage and a mechanism to correct such flaws by modifying the questionnaire design in the database.

Furthermore, in a k-anonymity or l-diversity context, the relationship between basic personal information (like age or gender) and medical data are focused. The key question has been how to disconnect them and how to protect medical data of patients from the public. However, when basic personal information and medical data are already known to medical staff, the problem turns to be how to disconnect the relationship between the both type of data and newly obtained hospital evaluation data from patients. It should be noticed that a k-anonymity or l-diversity framework doesn't protect the former data from medical staff, as medical staff can access their database directly. The target is to establish an electronic survey system that can obtain such evaluation data while guaranteeing anonymity of patients and protect their privacy from medical staff.

3. A Design of Privacy-Enhanced Survey System

This section details the design of a privacy-enhanced survey system that can be used for hospital evaluation by patients. The design can be obtained by simply arranging the specification about sets of respondents and questions presented in previous studies.

The formal description of the system design is as follows:

[Sets of Respondents and Questions]

P denotes the set of all respondents, or patients. Using the respondent number $i(i \leq i \leq N)$, we can define it as follows:

$P = \{1,2,\ldots,N\}$

Q represents the set of all questions in a question form. Each element of Q is classified into X items and Y items. Each element of X is classified into BPI and MD.

X items ($\mathrm{x}_1,\mathrm{x}_2\ldots,\mathrm{x}_\mathrm{n}$) represent individual attributes that are observable by medical staff. Y items ($\mathrm{y}_1,\mathrm{y}_2\ldots,\mathrm{y}_\mathrm{m}$) represent individual attitudes that are not observable even by medical staff. BPI items ($\mathrm{bpi}_1,\mathrm{bpi}_2\ldots,\mathrm{bpi}_\mathrm{i}$) represent basic personal information and MD items ($md_1,\mathrm{md}_2\ldots,\mathrm{md}_\mathrm{d}$) medical data.

The index numbers of X items $\mathrm{x}_1,\mathrm{x}_2\ldots,\mathrm{x}_\mathrm{n}$ denote the priority (an item with a relatively large index number is eliminated faster in the database).

$$BPI = \{b\mathrm{pi}_1,bpi_2,\ldots,bpi_i\}$$
$$MD = \{md_1,md_2,\ldots,md_d\}$$

$$X = \{x_1,x_2,\ldots,x_n\}$$
$$Y = \{y_1,y_2,\ldots,y_m\}$$

$$X = BPI \cup MD$$
$$Q = X \cup Y$$

[Respondents' Answers]

For $\forall q \in Q$, $D(q)$ represents the domain of the answer to the question q. The 3-tuple (i,q,a) indicates that a respondent $i \in P$ selected the answer $a \in D(q)$ for the question $q \in Q$. T_0 denotes the set of all such 3-tuples and contains the information of all the answers provided by all the respondents. For $\forall q \in Y$, $D_C(q)$ is defined as the set of all $D(q)$ elements that are sensitive alternatives requiring their selectors to be concealed. That is, $D_C(q)$ represents the set of the alternatives of a negative evaluation.

I have omitted some of the following formal descriptions. However, the process is simple, and it is equivalent with the process described in [5] and [9]. ([9] is accessible online.)

Now, the main routine to enhance privacy can be described as follows:

For each $AB_j(1 \le j \le M)$, perform the following procedure:

STEP 1)

If $Project(AB_j,2) \cap X = \phi$ or $Flag(AB_j) = 0$, then go to STEP 2.

If $Project(AB_j,2) \cap X \ne \phi$ and $Flag(AB_j) = 1$, $AB_j = Delete(AB_j,2,\{x_{|Project(AB_j,2)\cap X|}\})$ and do STEP 1 again.

STEP 2)

Perform the following procedure:

i) $A_{Attributes} = Random(Delete(T_0,2,Y),1)$

ii) For $k(0 \le k \le n = |X|)$,

$$A_k = Random(\bigcup_{|Project(AB_j,2)\cap X|=k}(AB_j),1)$$

iii) Delete data except of $A_{Attributes}, A_0, A_1, A_2, \ldots, A_n$

iv) Output $A_{Attributes}, A_0, A_1, A_2, \ldots, A_n$

Here, $Project(S, j)$ denotes a function that returns the set of all j-th elements of the 3-tuples that belong to S. Similarly, $Random(S, j)$ represents a function that swaps the j-th elements of all 3-tuples of S and returns the set that can be obtained as a result of the calculation. $Flag(AB_j)$ denotes the function to determine whether the operation of modifying the database is needed for protecting privacy with AB_j. $Delete(S_1, j, S_2)$ returns a subset of the 3-tuple set S_1. This calculation eliminates the 3-tuple of S_1, when the j-th element of the 3-tuple belongs to set S_2. (For more details, see p.313 of [9].)

If the questionnaire of the focused hospital has n items of attribute questions, the number of final output tables is expected to be n+2.

4. Discussion

Although the implementation of the system described in the last section is still an ongoing task, it is expected to be feasible, as a prototype system described in [9] was already implemented and evaluated in the study.

However, it would be questionable whether the following calculation method, which the above formal description inherited from the previous studies, is suitable for measuring privacy level in the new framework.

$$L(AB_j, x, q) = \log(\frac{|DS|!}{|DS - DSc|! \times |DSc|!})$$

This calculation method is based on combinatorial approach in measuring anonymity level ([1, 2, 3, 4, 5, 6, 7, 8, 9]), and relates to Shannon's theory ([11]). If it is replaced with another measure such as k-anonymity or l-diversity itself, the whole system might be more consistent. This is an area for new examination and a task for the next stage of this study.

5. Concluding Remark

This study presented the design of a privacy-enhanced survey system that can be used for hospital evaluation by patients. By expanding a survey system designed in previously presented studies, a formal description of the new system was demonstrated.

One original contribution of this study is to have linked two different types of privacy-preserving techniques. That is to say, it seems to be possible to establish a consistent private preserving survey system which guarantees both protecting patients' information from the general public and protecting patients' information from survey assessors or medical staff.

References

[1] R. Bagai, H. Lu, R. Li and B. Tang, "An Accurate System-Wide Anonymity Metric for Probabilistic Attacks", ***Proceedings of the 11th Privacy Enhancing Technologies Symposium (PETS-2011)***, pp. 117-133, 2011.

[2] M. Edman, F. Sivrikaya and B. Yener, "A combinatorial approach to measuring anonymity". In ***Intelligence and Security Informatics***, IEEE, pp. 356–363, 2007.

[3] B. Gierlichs, C. Troncoso, C. Díaz and B. Preneel, "Revisiting a Combinatorial Approach Toward Measuring Anonymity", in ***Proceedings of WPES,*** pp. 111-116, 2008.

[4] A. Iwai, "Tohyô Kôi ni okeru Tokumeisei Gainen no Keishikika", ***Preprints of the 35th Conference of the Japanese Association for Mathematical Sociology***, pp. 92-93, 2003.

[5] A. Iwai, "A Framework of Social Survey System that Prevents Personal Information Leakage by Automatic Modification of Questionnaire Design", in ***Proceedings of 18th symposium on socio-information systems***, pp. 127-132, 2012.

[6] A. Iwai, "Evaluation of an Anonymity Measure as an Index of Voting Privacy", ***Journal of Socio-Informatics***, Vol.5, No.1, pp11-25, 2012.

[7] A. Iwai, "Development of a Robust Course Evaluation System That Prevents Individual Information Leakage By Employing Input Data Analysis", Kagaku Kenkyuhi Josei Jigyo (Gakujutsu Kenkyu Josei Kikin Joseikin) Kenkyu Seika Houkokusho (KAKENHI 23650526), 2013. https://kaken.nii.ac.jp/pdf/2012/seika/F-19/12301/23650526seika.pdf

[8] A. Iwai, "Seisaku Kettei no tameno Kofuku Shihyo wa Jitsugen Suruka", ***Synergy Shakai Ron*** (T. Imada and Y. Tateoka ed.), Tokyo University Press, pp.73-85, 2014.

[9] A. Iwai, "Reviewing privacy-enhanced social survey system that employs combinatorial anonymity measure", ***IMECS2016 (International Multi-Conference of Engineering and Computer Scientist 2016) proceedings***, pp.311-316, 2016.

[10] A. Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkitasubramaniam, "ℓ-diversity: Privacy beyond k-anonymity", ***Proceedings of IEEE International Conference on Data Engineering (ICDE)***, pp. 24–35, 2006.

[11] C. E. Shannon, "A Mathematical Theory of Communication", ***The Bell System Technical Journal***, Vol.27, pp379-423, pp. 623-656, 1948.

[12] L. Sweeney, "k-Anonymity: A Model for Protecting Privacy", ***International Journal of Uncertainty Fuzziness and Knowledge Based Systems***, Vol. 10, No. 5, pp. 557-570, 2002.