Proceedings of International Conference on Technology and Social Science 2018 (ICTSS 2018)
Invited Paper

Towards Privacy Preservation in Data Utilization Based on Patient Information Severity

Osamu Takaki[1, a]

[1]Faculty of Social and Information Studies, Gunma University,

4-2 Aramaki-cho, Maebashi-shi, Gunma, 371-8510, Japan

[a]takaki@gunma-u.ac.jp

Keywords: data utilization, privacy preservation, patient information, k-anonymity, l-diversity

Abstract. This lecture describes a perspective towards privacy preservation in data utilization focused on severity of patients' information. The first half explains a process to generalize data by using an extension of l-diversity that is defined based on formalized severity of patients' information [1], and then discuss how to put the process into practical use. The second half introduces an approach towards estimation of the number of patients with high severity of information.

1. Motivation to Focus on Severity of Patients' Information

The theme of this lecture is preservation of patients' privacy in utilization or administration of medical data. One of the biggest features of medical data from the viewpoint of privacy preservation is that medical data have significantly different impacts of informations. In a usual case, when an administrator of a database in a hospital is required to provide medical data, he/she has to determine an appropriate range and level of detail of the data. Therefore, it is important to anonymize the medical data properly to maintain safety as well as usability of the data.

In order to assess the risk for an adversary to obtain new sensitive information of patients if the adversary has certain information of target patients and he/she obtain a new data, Sweeney proposed k*-anonymity* [2], which was a property to assess the risk of identification of the target patients. Moreover, as an extended property of k-anonymity, Machanavajjhala, et al., [3] introduced l*-diversity*, which is an indicator to assess the risk of identification of sensitive information of the target patients based on Bayesian estimation and information theory. Chunyong, et al., [4] and Xiaoxun, et al., [5] extended k-anonymity and l-diversity respectively based on ranks that are defined by classifications of the set of diseases. On the other hand, LeFevre, et al., [6] proposed an algorithm *Incognito* to generalize a given data in a way where it can conduct the generalization with fewest steps.

Although these researches are all important, there have not yet been realized any research how to anonymize medical data according to severity of patients' information from the semantic viewpoints, despite its importance. In fact, it is challenging to quantify the risk or severity for an adversary to obtain sensitive information of a target patient, and it needs to conceptualize and quantify severities of patients' information from a wider viewpoint than the existing one.

2. Extended l-diversity Based on Risk-Impact Ontology for Patients' Sensitive Information and Generalization Process

In [1], the authors developed an ontology, which was called *Risk-Impact Ontology for Patients' Sensitive Information (RIOPSI)* and in which the authors conceptualized motivations of adversaries and defined criteria of diseases, medicine and medical treatment based on the adversaries' motivations above. And then they formalized severities of patients' information based on RIOPSI and extended l-diversity by using the formalized severities. Moreover, the authors proposed a simple process to

generalize medical data in accordance with the extended l-diversity and demands by a user who requested the medical data.

The first purpose of this lecture is to propose how to put the process above into practical use. To this end, a framework that was introduced by Eman and Arbuckle [7] is employed. In order to combine them, two levels of risk assessments are considered: the first assessment is to calculate possibility of identification of target patients and the second one is to estimate severity to identify the target patients.

3. Approach Towards Estimation of the Number of Patients with High Severity of Information

The second half of this lecture introduce an approach towards estimation of the number of patients with high severity of information. As an example of a group of patients with high severity of information, patients of intractable diseases that are defined in Health, Labour and Welfare Ministry of Japan and cancers that are classified by sites of occurrence are considered. The statistical data about the patients above are published in the sites of Japan Intractable Diseases Information Center [8] and National Cancer Center Japan [9]. As an approach towards estimation of severity of their information, this lecture introduce *a gap by excessing utilization of data* and employs fundamental theory of life insurance medicine.

References

[1] O. Takaki, T. Asao and Y. Seki, “Extensions of l-diversity to reduce the risk of revealing patient severe health conditions,” ***In Proceedings of the 1st International Conference on Mechanical, Electrical and Medical Intelligent System 2017 (ICMEMIS2017)***, 2017.

[2] L. Sweeney, "k-anonymity: a model for protecting privacy, " ***Int. J. Uncertain. Fuzziness Knowl.-Based Syst***. Vol. 10, No. 5, 2002, pp. 557-570.

[3] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, "L-diversity: Privacy beyond k-anonymity," ***TKDD***, Vol. 1, No. 1(3), 2007, pp. 1-52.

[4] Yin, Chunyong, Sun Zhang, Jinwen Xi and Jin Wang, "An improved anonymity model for big data security based on clustering algorithm," ***Concurrency and Computation: Practice and Experience***, Vol 29, 2017, pp. 1-13.

[5] Sun, Xiaoxun and Wang, Hua and Li, Jiuyong and Truta, Traian Marius, "Enhanced p-sensitive k-anonymity models for privacy preserving data publishing," ***Transactions on Data Privacy***, 1 (2). 2008, pp. 53-66.

[6] K. LeFevre, D. J. DeWitt, and R. Ramakrishnan. "Incognito: efficient full-domain K-anonymity," ***In Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data (SIGMOD '05)***, ACM, New York, NY, USA, 2005, pp. 49-60.

[7] K. Eman and L. Arbuckle, “Anonymizing Health Data: Case Studies and Methods to Get You Started,” ***O’Reilly Medica***, 2013.

[8] Japan Intractable Diseases Information Center, http://www.nanbyou.or.jp/.

[9] National Cancer Center Japan, ***Cancer Registration and Statistics***, https://ganjoho.jp/reg_stat/index.html.