In the future, we are planning an experiment to verify the effectiveness of the proposed algorithm.

6. Conclusion and Future Works

In this study, we have taken up the issue on how to consider anonymization methods for datasets that contain not only publicly available information, such as SNS data, but also confidential information. With regard to this problem, Section 2 has described the types of data handled, the main methods used in privacy-preserving data mining, and the relationship of this research with privacy-preserving data mining.

In Section 3, we have proposed a method for evaluating the degree of anonymization when only a part of the input data was anonymized. Moreover, we considered the data distribution.

In Section 4, the degree of safety of anonymization for data in which only some data was anonymized was examined experimentally against the bias of data distribution.

Based on the experimental results, it has been confirmed that the non-anonymized attributes include those that render the narrowing down of data easy and those that make it otherwise. This indicates that there are attributes that cannot be anonymized to increase the degree of security. Since attributes that make it easy to narrow down data are those that contain data with a small number of distributions, it is safer to delete such attributes as identifiers when using the anonymization approach.

In Section 5, we introduced privacy protection decision tree learning using randomization. Next, we proposed privacy protection clustering using the proposed randomization.

7. References

[1] R. Agrawal and R. Srikant, “Privacy-preserving data mining,” SIGMOD Rec., vol. 29, no. 2, pp. 439–450, May 2000. http://doi.acm.org/10.1145/335191.335438 (Access Date: 15 January, 2021).

[2] C. C. Aggarwal and P. S. Yu, A General Survey of Privacy-Preserving Data Mining Models and Algorithms. Boston, MA: Springer US, 2008, pp. 11–52. https://doi.org /10.1007/978-0-387-70992-5 2 (Access Date: 15 January, 2021).

[3] J. Sakuma and S. Kobayashi, “Privacy-preserving data mining,” The Japanese Society for Artificial Intelligence, vol. 24, no. 2, pp. 283–294, mar 2009.

[4] A. Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkitasubramaniam., (2006). L-diversity: privacy beyond k-anonymity, in 22nd International Conference on Data Engineering (ICDE’06), April, pp. 24–24.

[5] L. Sweeney., (2002). k-anonymity: A model for protecting privacy, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, vol. 10, no. 05, pp. 557–570. http://www.worldscientific.com/doi/abs/10.1142/S0218488502001648 (Access Date: 15 January, 2021).

[6] R. Cramer, I. Damg˚ard, and J. B. Nielsen, (2001). Multiparty computation from threshold homomorphic encryption, in Advances in Cryptology - EUROCRYPT 2001, B. Pfitzmann, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, , pp. 280–300.

[7] Utd anonymization toolbox. http://www.cs.utdallas.edu/ dspl/cgi-bin/toolbox/index.php (Access Date: 15 January, 2021).

[8] M. Lichman, “UCI machine learning repository, (2013). http://archive.ics.uci.edu/ml(Access Date: 15 January, 2021).

8. Acknowledgements

This research work is supported by JSPS KAKENHI Grant Number JP17K00310.