A Study on the Enabling Measure for Secure Data Transfer Based on Privacy

Gee-hee Yun[*] and Kyoung-Jin Kim

Sungshin Women’s University, Seoul, Republic of Korea.
geehee_yun@naver.com, kyongjin@sungshin.ac.kr

Abstract

Our research introduces a matrix for the optimal generation of synthetic data in Mydata services. Through this matrix, we apply differential privacy to safeguard the original data while maintaining its utility. This enables us to assist in providing personalized services to users based on securely protected large-scale data.

Keyword : Privacy Metrix, Data Portability, Privacy Enhancing Technique

1 Introduction

Financial data, including sensitive information such as credit data, is a critical subject of privacy protection. It is also designated as a target of Data Portability through regulations like GDPR, in order to achieve data sovereignty restoration and provide highly personalized services. This enables users to enjoy sophisticated tailored services, while businesses can address customer information imbalances and offer a diverse range of services. However, users must be mindful of the potential risk of personal information leakage [1]. The research that devises optimal strategies for generating synthetic data in machine learning environments, specifically for financial data including time series and tabular data [2], employs metrics such as Euclidean distance and maximum entropy methods for measuring similarity. In this study, we propose an evaluation matrix for synthetic data to enhance service utility and improve privacy while considering the service format and data characteristics prevalent in the financial environment.

2 Privacy Protection Data Portability Model

Table 1 includes a matrix of data utility and privacy protection based on utilization patterns.

Service Category	Example	Realism	Coherence	Diversity
Integrated Single-View	Personal Financial Management	High	High	Low
Converge Industries	Combining Data from Various Industries	Low	Low	High
Trustworthy Assurance	Insurance Enrollment Assessment	High	High	Low
Seamless Connect	Insurance Claim	Low	Low	High
Expand Network	Tax Consultation	High	High	Low

Table 1: Synthetic Data Evaluation Criteria Based on Service Types

The 7th International Conference on Mobile Internet Security (MobiSec’ 23), Dec. 19-21, 2023, Okinawa, JAPAN, Article No. P-28

[*] Corresponding author: Sungshin Women’s University, Seoul, 02844, South Korea, Email: geehee_yun@naver.com

Integration, assurance, and Expand Network patterns are related to service trust. Therefore, similarity metrics like realism and coherence with the original data should be maintained at a high level, while diversity should be included to a minimum extent. Convergen and Connect patterns derive results through the statistical characteristics, including the combination with other data. Hence, it's important to maintain the statistical characteristics in line with the original data, and diversity can be used to enhance privacy protection aspects. Figure 1 illustrates the proposed model. In this framework, Mydata operators collect user information and generate synthetic data based on the set metrics information according to the requested service type. To ensure coherence and realism during synthetic data generation, the data distribution and distances between data points are measured. Additionally, for diversity, the accuracy based on the range of epsilon values is examined to include appropriate epsilon-based noise for differential privacy in the generated synthetic data.

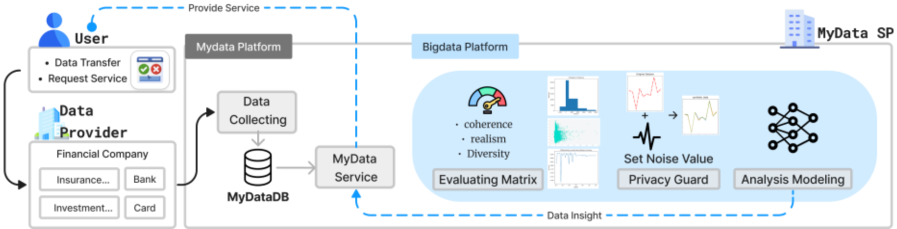

Figure 1: Synthetic Data and Service Delivery Models

3 Conclusion

Ultimately, this study proposes a model that provides benefits through the provision of services based on users' right to data portability and the enhancement of services based on synthetic data in the financial industry. Throughout this process, depending on the type of Mydata service, evaluation metrics such as coherence, realism, and diversity during synthetic data generation reveal the similarity to the original data and the relationship with privacy assurance. The proposed model, utilizing differential privacy in data generation, ensures the protection of the original data while obtaining the same benefits. This allows for the advantages of personalized services based on protected big data and enables both protection and utilization against potential threats when utilizing the data.

Acknowledgments :This work is partly supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government(MSIT) (No. 2022R1F1A1074038), the Korea Institute for Advancement of Technology (KIAT) grant funded by the Korean government(MOTIE) (P0008703, The Competency Development Program for Industry Specialist), and the MSIT under the ICAN (ICT Challenge and Advanced Network of HRD) program (No. IITP-2022-RS-2022-00156310) supervised by the Institute of Information & Communication Technology Planning & Evaluation (IITP).

References

[1] Alorwu, A. Kheirinejad. S, Berkel, N, Kinnula, M. Ferreira, D. Visuri, A. 2021. Assessing MyData Scenarios: Ethics, Concerns, and the Promise. CHI '21, NY, USA, Article 209, 1–11.

[2] Samuel A. Assefa, Danial Dervovic, Mahmoud Mahfouz, Robert E. Tillman, Prashant Reddy, and Manuela Veloso. 2021. Generating synthetic data in finance: opportunities, challenges and pitfalls, ICAIF '20, NY, USA, Article 44, 1–8.