A Privacy-Preserving Framework for Generative Model-driven Synthetic Datasets

Debalina R Padariya

School of Computer Science and Informatics
De Montfort University
Leicester, United Kingdom
p2723446@my365.dmu.ac.uk

Abstract

Despite the advancement of generative model-based synthetic datasets, several challenges, such as privacy attacks and limitations of current privacy-preserving approaches, undermine the trust in this field. This research attempts to alleviate these challenges by developing a novel privacy-preserving framework that will contribute to the practical advancements of synthetic data generation across industry and the public sector.

Introduction

Over the past few years, generative models (GMs) have gained widespread success across different fields for generating synthetic data, such as image applications, tabular data synthesis, time-series data generation, audio/video synthesis, natural language processing, and many more (Feuerriegel et al. 2024). The most popular deep generative models, such as Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs), have successfully produced realistic synthetic samples due to their potential to learn high-dimensional complex data distributions. Synthetic data generation (SDG), which aims to resemble real-world data closely, offers a promising solution to preserving privacy while ensuring sufficient usefulness for future purposes.

However, in recent years, widespread concerns have emerged about the potential threats of GMs. The possible privacy attacks attempt to infer sensitive information about the target model at different levels, such as training data, attributes, model, or identification-based (Sun et al. 2023). A critical SDG consideration is its reliability in the current regulatory landscape, such as re-identification risks in the European Union's GDPR (Council 2016) or de-identification requirements in CPRA (California Privacy Rights Act, an update to CCPA) (Blesch 2023).

Earlier approaches, e.g., anonymization, struggle to find a formal definition of privacy that matches users' expectations. A popular and formal mathematical approach is differential privacy (DP), which holds great promise for disclosure control and quantifying the privacy risk of synthetic data (Abadi et al. 2016). However, researchers have investigated the implications of DP in GMs, such as reducing task accuracy, optimal selection of privacy budget, and the challenging applicability of DP in some applications (Ganev, Oprisanu, and Cristofaro 2022; Yoon, Drumright, and van der Schaar 2020). Therefore, despite the considerable progress, the development of GMs can still pose challenges regarding privacy protection benchmarks.

Research Questions

My research aims to build a comprehensive picture of privacy and utility for existing GMs that offer privacy guarantees with the state-of-the-art DP framework yet remain susceptible to attacks on training data. This aim has been realized through the following three research questions that support my thesis:

- **What are the approaches for quantifying privacy in GMs?**

  The possible risks of information leakage in synthetic data have led researchers to focus on protecting privacy in GMs. While attack-based privacy metrics are a generic approach that measures the adversarial success rate (Sun et al. 2023), the proportion of overfitting in generalization can estimate information leakage (Chen et al. 2021). However, training GMs without privacy guarantees can lead to the model's overfitting, allowing the adversaries to mount attacks. The differential privacy-based mechanism provides rigorous privacy guarantees by incorporating noise-adding strategies, i.e., training the model with the DPSGD (Differentially Private Stochastic Gradient Descent) (Abadi et al. 2016). While DP offers rigorous statistical assurance to counter privacy attacks, careful consideration is needed while calibrating the noise.

- **What are the utility benchmarks for assessing synthetic data?**

  Utility metrics are crucial for measuring the effectiveness of synthetic data, where suitable metrics are chosen depending on relevant purposes. Some utility metrics measure the similarity between real and synthetic samples considering individual or distributional distance/similarity measures (Yoon, Drumright, and van der Schaar 2020), known as fidelity-based metrics. Others can estimate the performance of synthetic data for specific tasks, such as downstream classification or regression tasks

(Ganev, Oprisanu, and Cristofaro 2022), called specific-utility-based metrics.

- **How can we develop optimized metrics suites to evaluate the privacy/utility trade-offs in synthetic data effectively?**

  Metrics suites, a collection of different metrics, are important for evaluating specific aspects of synthetic data, such as privacy and usefulness. Optimizing metrics suites involves composition, selecting the most relevant metrics, and weighting, assigning weight to individual metrics in a suite. A well-optimized metric suite can be relevant to specific use cases of synthetic data and ensures the appropriate balance between privacy/utility trade-offs while considering interpretability and computational resources.

Past and Current Progress

For the first phase of the research, we comprehensively reviewed the literature on privacy-preserving GMs. We have proposed a systematic literature survey that offers an in-depth analysis of an exhaustive list of publications to map the current landscape. Additionally, we have designed novel taxonomies to categorize the privacy and utility metrics of GMs, allowing us to understand the similarities and differences of different metrics. Our literature survey revealed that the time series domain remains underexplored since it leads to complex models due to its sequential structure and high dimensionality.

To address this, in the second phase, we developed a methodology focused on privacy-preserving synthetic data generation for time-series applications. We have designed an extensible framework by integrating the components and the interfaces by realizing different candidate modules, such as generative models, attacks, and defenses. We rely on defenses that provide state-of-the-art differential privacy guarantees. I performed an extensive evaluation to assess the performance of synthetic data for downstream classification and regression tasks across different privacy budgets, demonstrating its effectiveness with the SOTA (state-of-the-art) model.

Future Research Plan

Having established this groundwork, my research will focus on integrating suitable privacy and utility metrics and proposing optimized metrics suites to evaluate the privacy/utility trade-offs.

- **Individual Metrics and Metrics Suites:** Since measuring different aspects of privacy and utility is important, we will test our proposed model's resistance to attacks (another aspect of privacy), the plausibility of generated data, and the similarity of synthetic real with real data (fidelity-based utility metrics). We will also apply optimization methods to optimize the composition of metrics suites. (in progress - late 2024)
- **Privacy/Utility Trade-offs** For this scenario, we will evaluate the privacy/utility trade-offs that allow consistent comparison of privacy and utility values with different SDG configurations. (to do - early 2025)
- **Deployment** To make our framework usable by the practitioners, we will develop a usable library that integrates the framework with a user-facing interface. This contribution will enable users to input their datasets and customize framework components to obtain a comprehensive privacy risk assessment. (to do - mid 2025)

Anticipated Thesis Contribution

My thesis contribution will be to design a privacy-preserving framework for time-series applications, incorporating a differential privacy-based mechanism to protect the model from privacy attacks. Furthermore, due to the lack of standardized evaluation metrics for time-series data, we will develop optimized metrics suites to analyze these complex datasets effectively, providing researchers and industry practitioners with actionable insights.

Acknowledgements

This work is supported by the Alan Turing Institute under the Turing/Accenture strategic partnership grant R-AST-040.

References

Abadi, M.; Chu, A.; Goodfellow, I.; McMahan, H. B.; Mironov, I.; Talwar, K.; and Zhang, L. 2016. Deep Learning with Differential Privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, 308–318. Vienna Austria: ACM. ISBN 9781450341394.

Blesch, W. 2023. The GDPR's Anonymization versus CCPA/CPRA's De-identification.

Chen, J.; Wang, W. H.; Gao, H.; and Shi, X. 2021. PAR-GAN: Improving the Generalization of Generative Adversarial Networks Against Membership Inference Attacks. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, KDD '21, 127–137. New York, NY, USA: Association for Computing Machinery. ISBN 9781450383325.

Council, E. a. 2016. Art. 9 GDPR – Processing of special categories of personal data.

Feuerriegel, S.; Hartmann, J.; Janiesch, C.; and Zschech, P. 2024. Generative AI. *Business & Information Systems Engineering*, 66(1): 111–126.

Ganev, G.; Oprisanu, B.; and Cristofaro, E. D. 2022. Robin Hood and Matthew Effects: Differential Privacy Has Disparate Impact on Synthetic Data. In *Proceedings of the 39th International Conference on Machine Learning*, 6944–6959. PMLR.

Sun, H.; Zhu, T.; Zhang, Z.; Jin, D.; Xiong, P.; and Zhou, W. 2023. Adversarial Attacks Against Deep Generative Models on Data: A Survey. *IEEE Transactions on Knowledge and Data Engineering*, 35(4): 3367–3388.

Yoon, J.; Drumright, L. N.; and van der Schaar, M. 2020. Anonymization Through Data Synthesis Using Generative Adversarial Networks (ADS-GAN). *IEEE Journal of Biomedical and Health Informatics*, 24(8): 2378–2388.