ANITA

Anonymous big data A project funded by FFG

Architectures for Evaluation in the Virtual Data Lab

Deliverable D3.3

Author(s): Peter Eigenschink

Reviewer(s): Klaudius Kalcher, Olha Drozd

Document version: 0.5
Date: 31.07.2020

WU Wirtschaftsuniversität Wien Vienna University of Economics and Business
MOSTLY AI
George Labs
Statistik Austria Die Informationsmanager
FFG Promoting Innovation.
Federal Ministry Republic of Austria Climate Action, Environment, Energy, Mobility, Innovation and Technology
Programme **"ICT of the Future"** - the initiative of the Austrian Federal Ministry for Climate Action, Environment, Energy, Mobility, Innovation and Technology (BMK)

Disclaimer

This deliverable describes the work and findings of the AI-Based Privacy-Preserving Big Data Sharing for Market Research (Anonymous Big Data (ANITA)) project.

The authors of this document have made every effort to ensure that its content was accurate, consistent and lawful. However, neither the project consortium as a whole nor the individual partners that implicitly or explicitly participated in the creation and publication of this deliverable are responsible for any possible errors or omissions as well as for any results and actions that might occur as a result of using the content of this document.

Table of contents

1 Summary

There are different popular generative deep learning architectures for sequential data, such as Recurrent Neural Networks (RNNs) or Convolutional Neural Networks (CNNs). When creating a model for a certain task, choosing the right architecture often is not an easy decision. The simulation study will provide guidance for the architectural decision of a generative deep learning model for sequential data, depending on the accuracy and privacy requirements of the task at hand.

The Virtual Data Lab (VDL) is an environment designed to evaluate deep learning models in terms of their accuracy and privacy. In the simulation study we will evaluate popular generative deep learning architectures for sequential data in the VDL. We will implement models based on different sequential architectures, namely RNNs, temporal CNNs and attention mechanisms. Models solely based on these three architectures are reported to generate good quality sequential data.

We will also implement models based on more complex architectures. These models will be based on the above-mentioned sequential architectures in combination with an autoregressive architecture or higher-level architectures, namely Generative Adversarial Networks (GANs) or Variational Autoencoders (VAEs).

As the above-mentioned models are not specifically designed to preserve the privacy of the training data, their results will provide a baseline for the evaluation of privacy-preserving techniques. Privacy-preserving techniques are applied to deep learning models to mitigate the risk of leaking sensitive information from the training data. To see the impact these techniques can have on the accuracy and privacy of a model, privacy-preserving versions of the most promising models will also be implemented, optimized with differential private stochastic gradient descend (DP SGD).

Based on the different privacy-preserving models and the non-private baseline models, the simulation study will deliver two important results: (i) a comparison of the accuracy and the privacy of different generative deep learning architectures for sequential data, and (ii) the impact that DP SGD has on the accuracy and the privacy of different generative models.

2 Introduction

The Virtual Data Lab (VDL) is an environment that is designed to evaluate generative models in terms of accuracy and privacy. The evaluation relies solely on properties of the generated synthetic data and the original data. As a result, model creators can evaluate a wide range of models in the VDL independently of the underlying data generating process and the implementation of the model.

In the simulation study we will implement models based on successful generative deep learning architectures for sequential data. These models will be evaluated in the VDL and compared in terms of accuracy and privacy. Privacy-preserving techniques are applied to deep learning models to mitigate the privacy risk. To evaluate their impact on the accuracy and the privacy of non-private models, each model will be compared to a privacy-preserving version of the same model.

In section 3 we introduce successful generative deep learning architectures for sequential data. Based on these architectures we determine the models that will be evaluated as part of the simulation study.

An important part of the simulation study is the evaluation of privacy-preserving techniques and their effect on accuracy and privacy. All architectures selected for evaluation in the VDL in section 3, will also be implemented as a privacy-preserving version by means of a privacy-preserving technique. In section 4 we discuss privacy-preserving techniques and their applicability to different model architectures. Section 5 provides details about models that were selected for the simulation study.

3 Generative Architectures

The mechanism used to generate synthetic data is crucial for the performance of a generative model. Some neural network architectures are able to intrinsically capture the structure of sequential data. Since the focus of the simulation study is sequential data, the evaluated models will be based on such sequential architectures, specifically Recurrent Neural Networks (RNNs), temporal Convolutional Neural Networks (CNNs) and attention mechanisms.

RNNs are designed specifically to model sequential data. These networks operate on a single step of a sequence at a time and predict the next step of the sequence. A RNN keeps track of all previous steps via an internal state. This state is updated for each step in the sequence, that is passed into the network, and influences the prediction of the next step. Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs) are particularly successful implementations of RNNs and improve the standard RNN architecture for better training and inference.

CNNs are particularly successful in the computer vision domain. These networks are able to abstract from images to "see" higher-level structures, such as edges in an image beyond just single pixels. This information is extracted by analysing squares of adjacent pixels. Similarly, CNNs can also extract higher-level information from multiple steps of a sequence. Temporal CNNs also incorporate the time dependency by working only on steps in the past.

The internal state that RNNs keep to model dependencies within the sequence can make it hard to capture long-term dependencies. Neural networks with attention mechanisms overcome this problem by learning the relative importance of previous steps for a given step in the sequence on their own. Networks with attention mechanisms know how important each step is for the prediction of the next step and predict the next step in a sequence with the previous steps as input.

Models based on RNNs, temporal CNNs and attention mechanisms are able to generate sequences by repeatedly predicting the next step of a sequence (i.e., one step at a time). Such models are able to generate sequential data of good quality and will be evaluated as part of the simulation study. As for RNNs, we will implement models based on the more successful variants LSTM and GRU and also RNNs combined with an attention mechanism.

In many cases, RNNs, temporal CNNs and attention mechanisms are combined with other architectures to further improve the quality of the generated data. These architectures are most commonly either autoregressive or of higher level, such as Generative Adversarial Networks (GANs) or Autoencoders (AEs).

Generative Adversarial Networks (GANs) have become very popular in recent years. They consist of a generator network and a discriminator network working against each other. The generator creates synthetic data and tries to fool the discriminator. The discriminator's objective is to distinguish between real and fake data.

Autoencoders (AEs) aim to reconstruct their input as accurately as possible. They consist of an encoder network and a decoder network. The encoder reduces the dimensionality of the input data and the decoder attempts to reconstruct the original data from that lower-dimensional representation. Variational Autoencoders (VAEs) have become very successful in overcoming issues in the quality of the results. A single datapoint is encoded into a probability distribution instead of a single point. The decoder then samples from that distribution in order to reconstruct the original input.

Both GANs and AEs are very flexible, because the generator, the discriminator, the encoder and the decoder can incorporate features of the data, such as the time dependency in sequential data or the relation between adjacent pixels in images. While simple feed forward networks often achieve good results, reflecting the structure of the data in the model can improve the results significantly. Thus, in order to generate high quality sequential data, GANs and AEs usually incorporate RNNs, CNNs or attention mechanisms into their generator/discriminator or encoder/decoder networks.

For the simulation study, we will implement and evaluate models based on RNNs, CNNs and attention mechanisms to generate sequential data, either on their own or in combination with other architectures. The combination with an autoregressive architecture, GANs or VAEs is the most popular. This way the quality of the synthetic data can be evaluated depending on both the architecture and the complexity of the model.

In the literature, combinations of multiple sequential architectures and, also, of multiple higher-level architectures are applied to specific scenarios. For example, Choi et al. combine a GAN with an AE architecture in [1] to generate synthetic patient records. While these combinations can lead to better results, they can highly increase the complexity, making it hard to interpret the effect of single sequential or higher-level architectures. Thus, for the evaluation in the simulation study, we will prefer models based on either sequential architectures on their own or based on sequential architectures in combination with a single higher-level or autoregressive architecture.

4 Privacy-Preserving Techniques

Deep learning models pose the threat to leak information of the original training data. This risk is particularly relevant when dealing with sensitive data. Privacy-preserving techniques enhance privacy of the deep learning models. However, they can negatively impact the accuracy of a model. In view of this trade-off, it is important to further investigate the impact of those techniques. This will be an essential part of the simulation study in the VDL.

The usual technique to quantify and mitigate the risk of leaking information about the original training data is the introduction of Differential Privacy (DP) into the model. Boulemtafes et al. [2] categorize techniques to introduce DP into deep learning models into three categories (i) differential private model parameters, (ii) differential private input data and (iii) differential private mimic learning. DP via the model parameters is obtained by injecting random noise into the weight parameters of the neurons (either directly or by modifying the optimization procedure). Alternatively, models can be trained on micro-aggregated and perturbed training input data. Lastly, with mimic learning, the final model is never directly exposed to the sensitive training data but only to non-sensitive data labelled by so-called "teacher" models. The details of all three categories are described in the deliverable D3.2 "Privacy-Preserving Techniques for Deep Learning".

The perturbation of model parameters seems to be the most generally applicable approach to enforce differential privacy. In particular, the optimization of a model via differential private stochastic gradient descent (DP SGD), as proposed in [3], has already been applied successfully to various use cases. Still, DP SGD has a dependency between the number of optimization steps for the neural network model and the amount of leaked information during the training. Usually, more optimization steps increase the accuracy of a model, but, as a result of the aforementioned dependency, this simultaneously increases the amount of leaked information.

There are other promising techniques, such as the perturbation of the loss function or the Adaptive Laplace Mechanism (AdLM), proposed in [4] and further expanded in [5]. In the AdLM noise is injected into the model parameters relative to their relevance for the output. While this technique has only been applied to classification tasks, an extension to unsupervised learning and GANs or AEs would be interesting, but has not been described in the literature yet.

As in the majority of the literature about privacy in deep learning DP SGD is applied as privacy-preserving technique, our focus in the simulation study will be on the impact of DP SGD on the accuracy and the privacy of a model. Selected generative models will also be implemented as a privacy-preserving version, optimized with DP SGD. Comparing these models to

non-private baseline models in the VDL will allow us to evaluate the impact of privacy-preserving techniques on the accuracy and privacy of generative models.

5 Architectures for the Virtual Data Lab

Privacy-preserving generative deep learning models are of special interest in the medical domain. For obvious reasons, medical data contains highly sensitive data that is subject to strict access regulations. In particular, that applies to sequential data, such as electronic health records, medical texts or blood pressure time series. As a result, many attempts to create privacy-preserving models have their origin in the medical domain. In [6] Esteban et al. consider a GAN with a LSTM generator and a LSTM discriminator. Both, the generator and the discriminator, are trained with DP SGD. Beaulieu-Jones et al. [7] also apply a GAN, trained via DP SGD, to generate synthetic data. Additionally, GANs without any DP mechanisms are also applied in medical applications ([1], [8], [9]). Recent approaches are dominated by GAN architectures, that are often combined with LSTMs and trained with DP SGD.

While privacy-preserving LSTMs and GANS are especially dominant in the medical domain, there are generative architectures that are successful in other domains, while there privacy may not be the main issue. In the simulation study we will compare non-private models based on generative architectures for sequential data, that are popular in different domains, in terms of their privacy and their accuracy. The impact of privacy-preserving techniques on accuracy and privacy will be evaluated by comparing privacy-preserving models with non-private baseline models. In the following paragraphs we describe the model architectures and the applied privacy-preserving technique that will be evaluated in the simulation study.

RNNs, attention mechanisms, and temporal CNNs are popular architectural building blocks of neural networks dealing with sequential data. These architectures are able to generate good quality synthetic sequences. As part of the simulation study we will implement models based on LSTMs, GRUs, temporal CNNs, Transformers (i.e. attention mechanisms on their own) and RNNs with attention mechanisms. These models will be compared against each other in the VDL.

The above-mentioned architectures will also be combined with autoregressive architectures, GANs or VAEs. The resulting models will be compared with the simpler models, solely based on the sequential architectures. This comparison will give insights about the effect of an increased model complexity on the quality of the synthetic data.

Table 1 contains an overview of models that will be evaluated in the VDL. The rows contain the basic sequential architectures. These will either be implemented on their own or in combination with the architectures mentioned in the columns. The cells in the table contain references to published application examples. Empty cells signify architectures for which no relevant publications about their application are available. As most of these published approaches work with completely different

datasets, they will not be directly implemented in the VDL. However, they will serve as templates and points of reference for the implementation in the VDL.

Besides the mere quality of the generated samples, we are also interested in the privacy that the models are able to guarantee. In particular, the trade-off between privacy and accuracy is highly interesting. To evaluate the impact of privacy-preserving techniques on the accuracy and privacy of neural networks, privacy-preserving versions of the most promising models, optimized with DP SGD [3], will also be implemented. In the simulation study, these privacy-preserving models will be compared with their non-private counterparts.

		Higher-level architecture		
		Auto-regressive	**GAN**	**VAE**
Sequential architecture	**LSTM**	[10] [11]	[6]	[12]
	GRU	[10] [11]	[6]	[12]
	Transformer	[13]	[14]	
	RNN /w attention	[15]		
	CNN	[15]	[17]	

Table 1: Architectures that will be evaluated in the Virtual Data Lab. The cells contain references to published applications of models, which are based on the two architectures mentioned in the corresponding row and column.

6 Conclusion

The models that will be evaluated in the Virtual Data Lab are based on the most popular sequential architectures. Among them RNNs, attention mechanisms and temporal CNNs are particularly interesting. In the simulation study models based on these three architectures will be evaluated, either on their own or in combination with an autoregressive architecture, GANs or VAEs., In the Table 1 we provide an overview of models that will be evaluated in the VDL. For each cell in the Table 1 one model will be implemented, incorporating the architectures in the corresponding row and column. As these baseline models will only be able to ensure a minimal amount of privacy, for the most promising of them another privacy-preserving version, trained with DP SGD, will be implemented. In the simulation study these models will be compared in terms of their accuracy and privacy.

7 References

[1] Edward Choi et al., "Generating Multi-label Discrete Patient Records using Generative Adversarial Networks," in Proceedings of the 2nd Machine Learning for Healthcare Conference, vol. 68, Boston, 2017, pp. 286-305.

[2] Adam Santoro et al., "Relational recurrent neural networks," in Advances in Neural Information Processing Systems 31, 2018, pp. 7299-7310.

[3] Martin Abadi et al., "Deep Learning with Differential Privacy," in Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, 2016.

[4] NhatHai Phan, Xintao Wu, Han Hu, and Dejing Dou. (2017) arXiv. [Online]. http://arxiv.org/abs/1709.05750

[5] T. Adesuyi and B. Kim, "A layer-wise Perturbation based Privacy Preserving Deep Neural Networks," in 2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), 2019, pp. 389-394.

[6] Cristóbal Esteban, Stephanie L. Hyland, and Gunnar Rätsch. (2017) Real-valued (Medical) Time Series Generation with Recurrent Conditional GANs. [Online]. https://arxiv.org/abs/1706.02633

[7] Brett K. Beaulieu-Jones et al., "Privacy-Preserving Generative Deep Neural Networks Support Clinical Data Sharing ," Circulation: Cardiovascular Quality and Outcomes, vol. 12, no. 7, July 2019.

[8] Alexandre Yahi, Rami Vanguri, Noémie Elhadad, and Nicholas P. Tatonetti. (2017) Generative Adversarial Networks for Electronic Health Records: A Framework for Exploring and Evaluating Methods for Predicting Drug-Induced Laboratory Test Trajectories. [Online]. https://arxiv.org/abs/1712.00164

[9] Zhengping Che, Yu Cheng, Shuangfei Zhai, Zhaonan Sun, and Yan Liu. (2017) Boosting Deep Learning Risk Prediction with Generative Adversarial Networks for Electronic Health Records. [Online]. https://arxiv.org/abs/1709.01648

[10] Ziheng Lin et al., "Deep generative models of urban mobility," in IEEE Transactions on Intelligent Transportation Systems, 2017.

[11] Soroush Mehri et al., "SampleRNN: An unconditional end-to-end neural audio generation model," in 5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings, 2017, pp. 1-11.

[12] Adam Roberts, Jesse Engel, Colin Raffel, Curtis Hawthorne, and Douglas Eck, "A hierarchical latent vector model for learning long-term structure in music," in 35th International Conference on Machine Learning, ICML 2018, vol. 10, 2018, pp. 6939-6954.

[13] Ashish Vaswani et al., "Attention is All you Need," in Advances in Neural

Information Processing Systems 30, 2017, pp. 5998-6008.

[14] Weili Nie, Nina Narodytska, and Ankit Patel, "RelGAN: Relational Generative Adversarial Networks for Text Generation," in International Conference on Learning Representations, 2019.

[15] Aaron van den Oord et al. (2016, Sep.) WaveNet: A Generative Model for Raw Audio. [Online]. http://arxiv.org/abs/1609.03499

[16] Sander Dieleman, Aaron van den Oord, and Karen Simonyan, "The challenge of realistic music generation: modelling raw audio at scale," in Advances in Neural Information Processing Systems 31, 2018, pp. 7989-7999.

[17] Sergey Tulyakov, Ming Yu Liu, Xiaodong Yang, and Jan Kautz, "MoCoGAN: Decomposing Motion and Content for Video Generation," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2018, pp. 1526-1535.