SCIENCE · PASSION · TECHNOLOGY

Open Thesis / Project

Differential Private Synthetic Data from HyperNetworks

(Embedded Information Processing Team)

Motivation

In recent years, synthetic data has shown to be a promising direction for privacy preserving data sharing. To generate synthetic data deep learning models are used. Those models are prone to over-fitting and may memorize training examples, which then may be reproduced during data generation. For example, GitHub's OpenAI-powered Copilot seems to be able to suggest Ethereum private keys it found in its training data.

Another shortcoming of current approaches for generating synthetic data is that it usually does not have different levels of privacy. For some use cases the data should be as close to the original as possible (e.g. when the data is studied by a domain expert), while in other use cases data utility is not important, but privacy is (e.g. when the data is used as "test data" during software development.

Differential Privacy is a privacy guarantee that ensures that participating in a data set does not substantially increase the risk of a data subject's privacy as result of participating in a data set. This is usually done by adding noise, the amount of noise is controlled by a value ϵ.

HyperNetworks are a novel type of neural network, which instead of solving a problem directly, predict a neural network to solve a given problem. By doing so, they can predict specific neural networks for solving specifict problems.

In this work you will be combining HyperNetworks and Differential Privacy to predict neural networks for generating synthetic data for a given ϵ.

Interested? Contact us for more details!

Target Group

Students in ICE, Computer Science and Software Development.

Thesis Type

Master Thesis (6 months).

Goals and Tasks

You task will be to implement a HyperNetwork for generate synthetic data based on different levels of ϵ. You won't have to start from zero, we already have some source code, you will be extending. We would also like to have a thorough evaluation of the obtained model. If we are successful, we aim to publish the achieved results (paper writing is not part of the thesis and you will get all our support for this). The project includes the following tasks:

- Literature research on relevant concepts
- Implement a HyperNetwork to predict a Generative Adverserial Network (GAN) for generating differential private synthetic data and evaluate its results.
- Summarize your results in a written report.

Requirements / Skills:

- Good understanding of deep learning modelling, experience training a model on a GPU;
- Interest in data privacy and in working on a previously unsolved problem;
- Programming skills in Python, experience with PyTorch.

Contact

- Franz Papst (papst@tugraz.at)
- Dr. Olga Saukh (saukh@tugraz.at)

Institute for Technical Informatics
Networked Embedded Systems Group

Networked Embedded Systems Group