Authors: M.Sc. Malte Blattmann
Date: 27.04.2023

Creation of evaluation pipeline for generative models and synthetic data

Type	Master's Thesis
Title	On the evaluation of generative models

In recent years, data driven approaches from the fields of machine learning or generally spoken, artificial intelligence, as well as statistical inference became a fundamental part of healthcare applications. Many of these models need to get trained on a vast amount of medical data. However, a lot of medical datasets are not publicly available due to privacy issues and legal barriers. One way to overcome this problem is the generation of synthetic data. Among other approaches, generative models, like variational autoencoder, can be used to create synthetic data.

Currently, we developed different generative models for the creation of medical datasets, whose performance needs to be evaluated. This means, that we have to measure, if the generated data is realistic, often using a restricted set of real world medical data for comparison.

There exist several approaches leading to different evaluation metrics. Often we think of the real world data as samples from an unknown probability density function, which can get estimated by different techniques such as kernel density estimation or vector quantization. This way the synthetic data corresponds with some probability score in a canonical manner, quantifying its quality. There also exist more formal approaches motivated from the mathematical field called topology.

The main task of this work will be to implement different evaluation techniques for generative models and to elaborate the advantages and disadvantages of these respective metrics based on basic test datasets, that need to get created. Depending on the personal interests of the candidate, either formal or evidence-based approaches can get selected for implementation. Motivated candidates can try to evolve existing concepts to gain some first experience with scientific research.

This work generally requires some adequate literature research and scientific papers need to be read to get familiar with the basic concepts. Close collaboration with the person in charge of our research team is possible. Thus, there will be support regarding any issues with the theoretical background, as well as the technical implementation. Active participation from the side of the student by stating proper ideas is welcome.

Qualifications

- Fluent in at least one of the following languages: German / English / Spanish
- At the end of some Master's degree course in a field such as computer science / engineering / mathematics
- Experience with the programming language Python
- Motivation to read scientific literature apart from the practical programming parts

Contact

Universität Leipzig
Innovation Center Computer Assisted Surgery (ICCAS)
M.Sc. Malte Blattmann
E-Mail: Malte.Blattmann@medizin.uni-leipzig.de
Web: www.iccas.de