

Data anonymization and synthesis

problem submitted by Desjardins

Context

Issues surrounding the protection of personal data are garnering more and more attention in society. Machine learning requires big data as well as granular data: thus it involves challenges, especially the protection of personal data and the transformation of data so that they cannot be traced to individuals. Machine learning applications entail risks when one attempts to valorize data and they may hinder collaborations where data sharing is needed between parties (businesses, suppliers, research centres, academic researchers, etc.).

Goal

We wish to investigate best practices for anonymizing or synthesizing data. Of course these techniques must allow one to retain as many "original data" features as possible: this is required to develop good predictive models. Our goal is to establish research collaborations in data valorization where data will not be used in a detrimental fashion.

More precisely the anonymized or synthesized data

- (a) must be such that they cannot be traced to individuals and
- (b) will allow predictive or explanatory models to "learn" and yield results as close as possible to those of equivalent models trained with the original data.

The following characteristic is very important for us.

- (c) We should be able to use the descriptive analysis of a collaborator who has worked with the modified data. This means that we should be able to determine the "original" variables corresponding to the modified ones.

Data

The first experiments will be carried out on structured data such as socio-demographic data and transactional data.

Applications

The collaborator receiving the secure (i.e., modified) data sets may use them for supervised or non supervised learning. It should be possible to carry out descriptive work on the modified data sets.

Potential approaches

To anonymize data one may remove features allowing the identification of individual persons.

The production of synthesized data could make use of Variational Autoencoders (VAEs) or Generative Adversarial Networks (GANs).

This list of approaches is not exhaustive.